TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS 1979

DIRECTORIO DE PROFESORES

DR. FLORENCIO ABOYTES GARCIA
INGENIERO ESPECIALISTA EN OPTIMIZACION
DE SISTEMAS
DEPARTAMENTO DE METODOLOGIA
COMISION FEDERAL DE ELECTRICIDAD
AV. TOLUCA Y CALLE DON MANUELITO S/N
COL. OLIVAR DE LOS PADRES
MEXICO 20, D.F
TEL: 595.54.00 exts. 232 a la 235

DR. EDUARDO ARREOLA VALDES
JEFE DE LA OFICINA DE METODOLOGIA DE
SISTEMAS ELECTRICOS
COMISION FEDERAL DE ELECTRICIDAD
RODANO No. 14 sala 405 piso 4
MEXICO 5, D.F.
TEL: 553.71.33 ext. 2676

M. EN I. RAFAEL CRISTERNA OCAMPO
JEFE DEL DEPARTAMENTO DE METODOLOGIA
COMISION FEDERAL DE ELECTRICIDAD
RODANO No. 14 sala 405 piso 4
MEXICO 5, D.F.
TEL: 553.71.33 ext. 2677

DR. SERGIO A. MOLINA GARCIA
AUXILIAR DE LA GERENCIA GENERAL DE OPERACION
COMISION FEDERAL DE ELECTRICIDAD
AV. TOLUCA Y CALLE DON MANUELITO S/N
COL. OLIVAR DE LOS PADRES
MEXICO 20, D.F.
TEL: 595.54.00 ext. 232 a la 235

M. EN I. COSME URDAIBAY ZUBILLAGA
JEFE DE LA OFICINA DE METODOLOGIA DE
INVESTIGACION DE OPERACIONES
COMISION FEDERAL DE ELECTRICIAD
RODANO No. 14 sala 405 piso 4
MEXICO 5, D.F.
TEL: 553.71.33 ext. 2084

DR. ALBERTO MAYER SASSON
INGENIERO EN JEFE DE APLICACIONES
DE COMPUTADORAS A SISTEMAS ELECTRICOS DE POTENCIA
AMERICAN ELECTRIC POWER
NUEVA YORK, U.S.A.

'pmc.8/I/79

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
(del 15 al 19 de enero de 1979)

FECHA	HORARIO	TEMARIO	PROFESORES
15 de enero	9:00 h	Registro de participantes y entrega de notas	
	10:00 h	Inauguración del curso.	Ing. Javier Jiménez Espriú
	10:15 h	Introducción.	Dr. Eduardo Arreola Valdés
	11:00 h	Conceptos de Algebra Matricial	Dr. Eduardo Arreola Valdés
	14:00 h	Solución de Sistemas de Ecuaciones Lineales usando Técnicas de Dispersidad.	Dr. Sergio A. Molina García
	16:00 h	Conceptos de Algebra Matricial	Dr. Sergio A. Molina García
	17:45 h	Sesión de preguntas y comentarios.	
16 de enero	9:00 h	Metodología de Estimación de Estado para Sistemas Eléctricos de Potencia.	Dr. Alberto Mayer Sasson
	11:00 h	Implantación de un Estimador de Estado y Experiencia de Operación	" " " "
	14:00 h	Uso de Equivalentes Externos para Análisis de Contingencias	" " " "
	16:00 h	Selección Automática de Contingencias.	" " " "
	17:45 h	Sesión de preguntas y comentarios	

FECHA	HORARIO	TEMARIO	PROFESORES
17 de enero	9:00 h	Diseño Funcional de Centros de Control de Energía	Dr. Alberto Mayer Sasson
	11:00 h	Equipo de Cómputo e Interfaz Hombre Máquina en un Centro de Control de Energía.	" " " "
	14 a 17:30 h	MESA REDONDA	
18 de enero	9:00 h	Conceptos Básicos de Optimización	M. en I. Rafael Cristerna Ocampo
	11:00 h	Programación No-Lineal	M. en I. Rafael Cristerna Ocampo
	14:00 h	Programación Lineal	M. en I. Cosme Urdaibay Zubillaga
	16:00 h	Programación Entera y Mixta	M. en I. Coste Urdaibay Zubillaga
	17:45 h	Sesión de preguntas y comentarios	
19 de enero	9:00 h	Despacho Económico con Restricciones I	Dr. Florencio Aboytes García
	11:00 h	Despacho Económico con Restricciones II	Dr. Florencio Aboytes García
	14:00 h	Desarrollo Optimo del Sistema de Generación	M. en I. Rafael Cristerna Ocampo
	16:00 h	Localización de Plantas y Lineas de Transmisión	Dr. Florencio Aboytes García
	17:45 h	Sesión de Preguntas y Comentarios	
	19:00 h	CLAUSURA	

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA I INTRODUCCION

DR. EDUARDO ARRIOLA VALDES

ENERO, 1979.

I N T R O D U C C I O N

Los avances tecnológicos logrados en el diseño y fabricación de computadoras digitales, han puesto a la disposición de la comunidad científica en general una poderosa herramienta de cálculo que ha dado pauta al desarrollo de nuevas técnicas de solución consideradas con anterioridad como imposibles por su complejidad y laboriosidad. Estos avances a su vez han exigido por parte del ingeniero y del científico un conocimiento más profundo de los problemas y de las bases matemáticas necesarias para su planteamiento y solución.

El sector eléctrico ha sido tradicionalmente uno de los usuarios de computadoras de mayor magnitud y se cuenta entre los que requieren mayor sofisticación en sus técnicas de modelado y de cálculo. Lo anterior es comprensible cuando se toma en consideración el costo del equipo eléctrico de potencia y los severos requerimientos de operación, necesarios para suministrar energía eléctrica en forma contínua y de buena calidad.

La planeación, el diseño y la operación de los sistemas eléctricos de potencia requieren de un análisis detallado y contínuo que permita evaluar las condiciones de funcionamiento del sistema, y estudiar las diferentes alternativas de expansión con objeto de garantizar, hasta donde sea posible, la mejor utilización del capital disponible.

En adición a su uso en los diversos aspectos de la planeación, diseño y operación, las computadoras digitales se utilizan cada vez con mayor intensidad en aplicaciones de tiempo real del sistema de potencia, a saber :

a) adquisición de datos en línea.

b) monitoreo del sistema.

c) análisis en tiempo real.

d) control en tiempo real.

Estas aplicaciones aunadas al desarrollo de los nuevos centros de control han sido quizás las tendencias mas significativas en el campo de los sistemas eléctricos de potencia en los últimos diez años. El trabajo desarrollado en este lapso ha consistido en parte en la adaptación de técnicas originalmente utilizadas en el area de planeación y operación a funciones de tiempo real del sistema de potencia, pero ciertamente las exigencias de este tipo de aplicaciones han demandado el desarrollo de nuevas técnicas y la utilización de tecnología moderna para satisfacerlas adecuadamente.

El curso tiene como objetivo el de transmitir las experiencias adquiridas en el desarrollo, implantación y utilización de nuevos modelos matemáticos; empleados unos en funciones de control en tiempo real de una red eléctrica y otros en la planeación de la expansión de la misma, con lo cual se propone dar una amplia visión de

la problemática de análisis que presenta un sistema eléctrico de potencia en diferentes fases de su estudio.

Para facilitar la comprensión de los diferentes modelos de red y las técnicas empleadas en su solución, se han incluido, a modo introductorio, temas de algebra matricial, probabilidad y estadística y optimización que dan al curso una mayor solidez permitiendo a los participantes obtener un mejor aprovechamiento del mismo.

Eduardo Arriola Valdés

Enero 1979.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA II CONCEPTOS BASICOS DE ALGEBRA MATRICIAL APLICABLES AL ANALISIS DE SISTEMAS ELECTRICOS DE POTENCIA

DR. EDUARDO ARRIOLA VALDEZ

ENERO, 1979.

CONCEPTOS BASICOS DE ALGEBRA MATRICIAL APLICABLES AL ANALISIS DE SISTEMAS ELECTRICOS DE POTENCIA

INTRODUCCION

Las técnicas de algebra matricial fueron inicialmente utilizadas en el análisis de sistemas eléctricos de potencia por el simple hecho de proporcionar un método conciso y sistemático para la formulación y solución de sistemas de ecuaciones. Entre las principales ventajas que presenta la utilización de estas técnicas se pueden mencionar las siguientes:

a) Reducen la probabilidad de cometer errores al efectuar cálculos manuales.

b) Permiten a través de expresiones concisas la formulación e interpretación de modelos matemáticos de sistemas complejos.

En adición a estas ventajas que en sí hacen atractivo el uso de técnicas de algebra matricial, es necesario mencionar que dada la importancia que ha adquirido la computadora digital en el análisis de sistemas de potencia, el estudio de estas técnicas se ha convertido para el ingeniero de sistemas de potencia en una imperiosa necesidad ya que el análisis con computadora de los problemas que presentan las grandes redes eléctricas involucran invariablemente operaciones sobre arreglos de números.

DEFINICIONES Y NOTACION

Una matriz es un arreglo rectangular de números, llamados elementos, colocados de manera ordenada y sistemática en " m " renglones y " n " columnas. Estos elementos pueden ser números reales o números complejos y se utiliza una notación con doble subíndice. a_{ij} para identificarlos. El primer subíndice i indica el renglón y el segundo subíndice indica la columna dando así una localización única a cada elemento. Usualmente se utilizan letras mayúsculas entre paréntesis cuadrados, p.e. $[A]$, para denotar una matriz.

Sea por ejemplo una matriz rectangular de la forma

$$[A] = \begin{bmatrix} a_{11} & a_{12} \cdots & a_{1n} \\ a_{21} & a_{22} \cdots & a_{2n} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ a_{m1} & a_{m2} \cdots & a_{mn} \end{bmatrix}_{mxn}$$

cuya dimensión se define como el número total de renglones y el número total de columnas contenidos en la misma. Esta se indica en la parte inferior derecha del arreglo como un producto m x n. y se dice entonces que la matriz es de dimensión " m por n ".

Una matriz consistente de un solo renglón o una sola columna se conoce con el nombre de vector. Un vector renglón es

una matriz de un solo renglón, es decir es una matriz de dimensión 1 x n. Un vector columna es una matriz de una sola columna, es decir es una matriz de dimensión m x 1. Algunos ejemplos de matrices y vectores con sus dimensiones son :

$$\begin{bmatrix} 1 & 4 & 0 \\ 2 & 3 & 2 \\ 1 & 6 & 5 \end{bmatrix}_{3x3} \qquad \begin{bmatrix} 1 & 8 \\ 6 & 0 \\ 4 & 2 \end{bmatrix}_{3x2} \qquad \begin{bmatrix} 1 & 3 & 5 & 7 \end{bmatrix}_{1x4} \qquad \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}_{4x1}$$

Una matriz aumentada se forma anexando una matriz (o vector) de dimensiones adecuadas a la derecha o a la izquierda de otra. Por ejemplo, si

$$A = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}_{2x3} \qquad y \qquad B = \begin{bmatrix} 7 & 8 \\ 6 & 9 \end{bmatrix}_{2x2}$$

la matriz aumentada A, B será :

$$[A, B] = \left[\begin{array}{ccc|cc} 1 & 3 & 5 & 7 & 8 \\ 2 & 4 & 6 & 6 & 9 \end{array}\right]$$

Se dice que una matriz es dispersa cuando un porcentaje significativo de sus elementos son igual a cero. Es usual que cuando se escribe una matriz, los elementos de la misma con valor cero sean simplemente dejados en blanco. Por ejemplo :

$$[A] = \begin{array}{|c|c|c|c|} \hline 1 & & & 5 \\ \hline 2 & 4 & & \\ \hline & & 7 & \\ \hline & 3 & 1 & 8 \\ \hline \end{array}$$

La matriz nula es aquella en la que todos sus elementos son igual a cero.

La transpuesta de una matriz A, indicada como A^T, se encuentra intercambiando los renglones y las columnas de A de tal manera que cada renglón i de A se convierta en la columna i de A^T. Si

$$A = \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array}_{2x3} \quad ; \text{ entonces } \quad A^T = \begin{array}{|c|c|} \hline 1 & 2 \\ \hline 3 & 4 \\ \hline 5 & 6 \\ \hline \end{array}_{3x2}$$

La matriz conjugada de una matriz compleja, indicada como A^*, se encuentra reemplazando cada elemento de A por su complejo conjugado.

$$A = \begin{array}{|c|c|} \hline 2 + j\,3 & -j\,5 \\ \hline -j\,8 & 6 - j\,2 \\ \hline \end{array} \qquad A^* = \begin{array}{|c|c|} \hline 2 - j\,3 & j\,5 \\ \hline j\,8 & 6 + j\,2 \\ \hline \end{array}$$

MATRICES CUADRADAS

Una matriz cuadrada es una matriz de dimensión m x m. Los elementos a_{ij} en que $i = j$ reciben el nombre de elementos diagonales y aquellos en que $i \neq j$ elementos no-diagonales o elementos fuera de la diagonal.

a_{11}	a_{12}	a_{13}
a_{21}	a_{22}	a_{23}
a_{31}	a_{32}	a_{33}

a_{ii} elementos diagonales

a_{ij} elementos no-diagonales

Una matriz diagonal es una matriz cuadrada cuyos elementos no-diagonales son todos cero.

3	
	1

a_{11}		
	a_{22}	
		a_{33}

Un caso especial de matriz diagonal es la llamada matriz unitaria o matriz identidad y en la que todos los elementos diagonales son unos.

1			
	1		
		1	
			1

esta matriz se denota usualmente por el símbolo I .

Una matriz simétrica es una matriz cuadrada con la propiedad de que $a_{ij} = a_{ji}$ para toda i y toda j .

2	-3
-3	4

1	-2	0
-2	2	7
0	7	3

Una matriz triangular es una matriz cuadrada cuyos elementos en el triángulo superior (elementos a_{ij} con $j > i$) o en el triángulo inferior (elementos a_{ij} con $i < j$) son todos igual a cero.

Una matriz triangular superior contiene elementos diferentes de cero unicamente en su triángulo superior, mientras que una matriz triangular inferior contiene elementos diferente de cero unicamente en su triángulo inferior.

2	-4	5	9
	8		-5
		6	2
			1

Triangular Superior

1			
7	4		
-4	2	3	
1	-7	0	8

Triangular Inferior

OPERACIONES CON MATRICES

Equivalencia .- Se dice que las matrices A y B son equivalentes o iguales si y solo sí:

a) Son de la misma dimensión

b) $a_{ij} = b_{ij}$ para toda i y toda j

Adición.- La operación de suma o resta matricial se puede efectuar unicamente si las matrices son de la misma dimensión. La matriz C se define como la suma de las matrices A y B a tra-

vés de la ecuación

$$A + B = C$$

y los elementos de C están dados por :

$$a_{ij} + b_{ij} = c_{ij} \quad \text{para toda} \quad i \quad \text{y toda} \quad j$$

$$\begin{array}{|c|c|c|} \hline 2 & 4 & 6 \\ \hline 3 & 5 & 7 \\ \hline \end{array} + \begin{array}{|c|c|c|} \hline 1 & 2 & 3 \\ \hline 4 & 5 & 6 \\ \hline \end{array} = \begin{array}{|c|c|c|} \hline 3 & 6 & 9 \\ \hline 7 & 10 & 13 \\ \hline \end{array}$$

Las leyes asociativa y conmutativa se cumplen para la suma matricial, esto es, para matrices A, B y C de la misma dimensión se cumple que :

$$A + B = B + A$$

y $$A + (B + C) = (A + B) + C$$

MULTIPLICACION POR UN ESCALAR

El producto de un escalar k y una matriz A se encuentra multiplicando todos y cada uno de los elementos de A por el escalar k.

$$k \begin{array}{|c|c|} \hline a_{11} & a_{12} \\ \hline a_{21} & a_{22} \\ \hline \end{array} = \begin{array}{|c|c|} \hline k\,a_{11} & k\,a_{12} \\ \hline k\,a_{21} & k\,a_{22} \\ \hline \end{array}$$

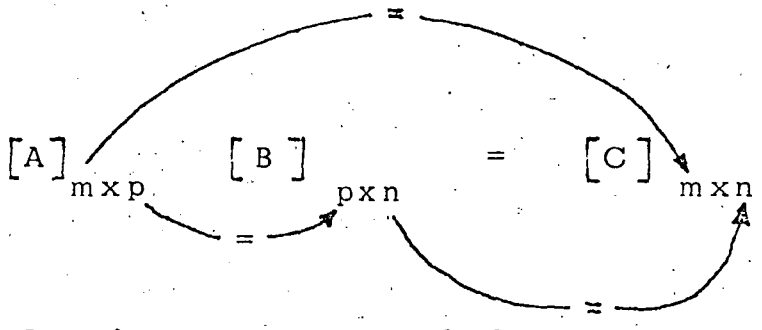
También en este caso de multiplicación por un escalar se cumplen las leyes conmutativa y distributiva

$$k A = A k$$

$$k (A + B) = k A + k B$$

MULTIPLICACION MATRICIAL

El producto de la multiplicación de dos matrices A x B esta definido únicamente cuando el número de columnas de A es igual al número de renglones de B. Cuando se satisface esta condición se dice que las matrices son conformables. El producto esta dado por :



Los elementos c_{ij} se calculan sumando los productos de los elementos correspondientes del renglón i de A y la columna j de B mediante la ecuación

$$c_{ij} = \sum_{k=1}^{p} a_{ik}\, b_{kj} \; ; \quad i = 1,2,\ldots,m \qquad j = 1,2,\ldots,n$$

Ejemplos :

$$\begin{bmatrix} 3 & 0 \\ 1 & 2 \end{bmatrix}_{2\times 2} \begin{bmatrix} 5 \\ 4 \end{bmatrix}_{2\times 1} = \begin{bmatrix} 3\times 5 + 0\times 4 \\ 1\times 5 + 2\times 4 \end{bmatrix} = \begin{bmatrix} 15 \\ 13 \end{bmatrix}_{2\times 1}$$

$$\begin{bmatrix} 2 & 4 & 6 \end{bmatrix}_{1\times 3} \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}_{3\times 1} = 2\times 1 + 4\times 3 + 6\times 5 = 44$$

Al efectuar el producto $A \times B = C$ se acostumbra a decir que la matriz B es premultiplicada por la matriz A. También se dice que la matriz A es postmultiplicada por la matriz B. La razón por la que se hace esta diferencia entre pre y post multiplicación es muy importante ya que la multiplicación matricial no es conmutativa, es decir

$$A \times B \neq B \times A \qquad \text{en general}$$

Sin embargo, la multiplicación matricial si es asociativa y distributiva con respecto a la suma

$$(A \times B)\, C = A\, (B\, C)$$

$$(A + B)\, C = A \times C + B \times C$$

La transposición de un producto matricial es igual al producto de las matrices transpuestas en orden invertido es decir :

$$(A\, B)^T = B^T\, A^T$$

$$(A\, B\, C)^T = C^T\, B^T\, A^T$$

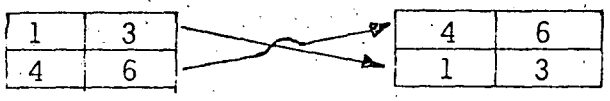
Cuando se pre o post multiplica una matriz cuadrada por una matriz identidad de dimensiones conformables el resultado es la matriz original

$$I\, A = A\, I = A \; ; \; A \text{ matriz cuadrada}$$

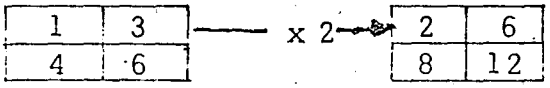
OPERACIONES ELEMENTALES SOBRE RENGLONES Y COLUMNAS.

Las manipulaciones efectuadas sobre una matriz conocidas como Operaciones Elementales sobre renglones pueden ser de 3 tipos, a saber :

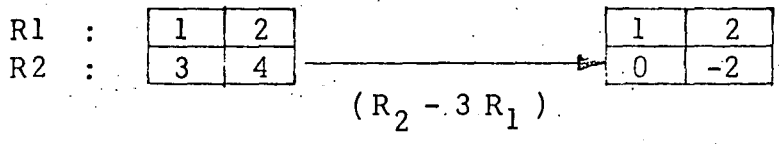
a) Intercambio del renglón " k " y el renglón " l " de una matriz



b) Multiplicación del renglón k por una constante diferente de cero



c) Suma al renglón " k " del contenido del renglón " l " multiplicado por " c " siendo c una constante diferente de cero.



Estas operaciones pueden ser aplicadas también sobre las columnas de la matriz utilizando la denominación de operaciones elementales sobre columnas.

DETERMINANTES.-

Asociado a toda matriz cuadrada existe un escalar denominado determinante y cuyo valor se denota como $|A|$. El cálculo del determinante de una matriz se calcula por técnicas de expansión. Por ejemplo el determinante de una matriz de 2 x 2 y de 3 x 3 se calcula de la siguiente manera :

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}\ a_{22} - a_{12}\ a_{21}$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\ a_{22}\ a_{33} + a_{12}\ a_{23}\ a_{31} + a_{21}\ a_{32}\ a_{13} - a_{13}\ a_{22}\ a_{31} - a_{12}\ a_{21}\ a_{33} - a_{11}\ a_{23}\ a_{32}$$

Una observación interesante en la expansión de $|A|$ es el hecho de que en cada término de la expansión aparece un elemento de cada columna y uno de cada renglón.

El cálculo de los determinantes se facilita con la introducción de nuevos elementos como son los llamados menores y cofactores de una matriz. El menor del término a_{ij} de la matriz A se define como el determinante de la matriz resultante de eliminar de la matriz original el renglón " i " y la columna " j ". Por ejemplo :

$$\text{si } A = \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline 3 & 2 & 8 \\ \hline \end{array} \quad ; \quad M_{23} = \begin{vmatrix} 1 & 3 \\ 3 & 2 \end{vmatrix} = 2 - 9 = -7$$

El co-factor de un elemento es el menor del mismo con el signo adecuado, esto es :

$$C_{ij} = (-1)^{(i+j)} M_{ij}$$

Estos elementos se utilizan en el cálculo del determinante de una matriz cuadrada a través de las siguientes expresiones :

$$|A| = \sum_{i=1}^{n} a_{ij} C_{ij} \quad \text{(suma de productos en la columna j)}$$

$$\text{ó} \quad |A| = \sum_{j=1}^{n} a_{ij} C_{ij} \quad \text{(suma de productos en el renglón i)}$$

en donde n es el número de renglones de la matriz A y C_{ij} es el cofactor del término a_{ij} . Ejemplo. Evaluar el determinante de una matriz de (3 x 3) usando cofactores :

$$A = \begin{array}{|c|c|c|} \hline 1 & 2 & 4 \\ \hline 5 & 6 & 7 \\ \hline 3 & 8 & 6 \\ \hline \end{array}$$

si seleccionamos el renglón 1 para el cálculo tenemos :

$$|A| = \begin{vmatrix} 1 & 2 & 4 \\ 5 & 6 & 7 \\ 3 & 8 & 6 \end{vmatrix} = 1 \begin{vmatrix} 6 & 7 \\ 8 & 6 \end{vmatrix} + (-1)\, 2 \begin{vmatrix} 5 & 7 \\ 3 & 6 \end{vmatrix} + 4 \begin{vmatrix} 5 & 6 \\ 3 & 8 \end{vmatrix}$$

$$= 1\,(36 - 56) - 2\,(30 - 21) + 4\,(40 - 18) = 50$$

Observe que el valor de cualquier determinante de dimensión finita puede ser calculado por la aplicación sucesiva del uso de los cofactores.

PROPIEDADES DE LOS DETERMINANTES

a) Si se intercambian dos renglones o dos columnas de un determinante, el valor del determinante cambia de signo

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = -2 \; ; \quad \begin{vmatrix} 3 & 4 \\ 1 & 2 \end{vmatrix} = 2$$

b) Si todos los elementos de un renglón o todos los elementos de una columna se multiplican por una constante k, el valor del determinante resultante es k A

$$\begin{vmatrix} ka_{11} & ka_{12} \\ a_{21} & a_{22} \end{vmatrix} = ka_{11}\, a_{22} - k a_{12}\, a_{21}$$

$$= k\,(a_{11}\, a_{22} - a_{12}\, a_{21})$$

c) Si se añade un múltiplo de una línea (renglón o columna) a una línea paralela el valor del determinante no se altera

$$\begin{array}{l} R1 \\ R2 \\ R3 \end{array} \begin{vmatrix} 3 & 0 & 6 \\ 5 & 1 & 2 \\ 2 & 6 & 7 \end{vmatrix} = 153 \; ; \quad \begin{vmatrix} 3 & 0 & 6 \\ 5 & 1 & 2 \\ 2 & 6 & 7 \end{vmatrix} \longleftarrow R2 + 2R1 \longrightarrow$$

$$\begin{vmatrix} 3 & 0 & 6 \\ 5+6 & 1+0 & 2+12 \\ 2 & 6 & 7 \end{vmatrix} = 153$$

Si una matriz tiene dos renglones idénticos o dos columnas idénticas, su determinante cero.

Para demostrar esto, observe que si restamos una de esas líneas idénticas (renglón o columna) de la otra, obtenemos como resultado una línea de ceros dando como resultado un valor cero para el determinante.

Una matriz cuyo determinante es cero recibe el nombre de matriz singular.

El determinante de una matriz diagonal o de una matriz triangular se evalúa tomando el producto de los elementos sobre la diagonal principal.

$$\begin{vmatrix} 2 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 5 \end{vmatrix} = 2 \times 8 \times 5 = 80$$

$$\begin{vmatrix} 4 & 5 & 6 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{vmatrix} = 4 \times 1 \times 3 = 12$$

Esta característica se puede explotar para desarrollar una técnica computacionalmente eficiente, que consiste en llevar una matriz dada a forma triangular a través de operaciones elementales sobre los renglones y columnas de la misma y una vez logrado esto efectuar el producto de los elementos de la diagonal principal.

MATRIZ ADJUNTA

Si cada elemento de una matriz cuadrada es reemplazado por su cofactor y se transpone la matriz resultante, obtenemos la matriz adjunta la cual se denota por A^+, esto es:

$$A^+ = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \\ C_{31} & C_{32} & C_{33} \end{bmatrix}^T$$

MATRIZ INVERSA

La inversa de una matriz cuadrada, representada por A^{-1}, se define de la sigiente manera :

$$A^{-1} A = A A^{-1} = I$$

Observe que la inversión de una matriz es análoga a la operación algebraica de división.

La inversa de una matriz se puede obtener de diferentes formas siendo la mas explícita la siguiente :

$$A^{-1} = \frac{A^{+}}{|A|}$$

de donde se puede observar que no existe inversa de una matriz singular (su determinante es cero).

Otro método para calcular la inversa de una matriz involucra el uso de operaciones elementales y consiste en anexar una matriz unidad de dimensión n x n al lado derecho de la matriz A cuya inversa se desea obtener. A continuación se efectúan operaciones elementales sobre los renglones de la matriz aumentada, con el objetivo de forzar a la matriz A a convertirse en una matriz unidad de dimensión n x n. Cuando se logra este objetivo, la parte derecha de la matriz aumentada es la matriz inversa de A.

Ejemplo : Encontrar la inversa de la matriz:

$$A = \begin{array}{|c|c|} \hline 2 & 3 \\ \hline 5 & 4 \\ \hline \end{array}$$

El primer paso es aumentar A con $I_{2 \times 2}$

2	3	1	0
5	4	0	1

efectuando a continuación las operaciones elementales indicadas en la siguiente secuencia :

R1
R2

2	3	1	0
5	4	0	1

$\xrightarrow{1/2\ R1}$

1	3/2	1/2	0
5	4	0	1

$\xrightarrow{R2-5\ R1}$

1	3/2	1/2	0
0	-7/2	-5/2	1

$\xrightarrow{-2/7\ R2}$

1	3/2	1/2	0
0	1	5/7	-2/7

$\xrightarrow{R1\ -\ 3/2\ R2}$

1	0	-4/7	3/7
0	1	5/7	-2/7

Como la mitad izquierda de la matriz aumentada es igual a la matriz unitaria, la mitad derecha es la deseada matriz inversa, - esto es :

$$A^{-1} = \begin{array}{|c|c|} \hline -4/7 & 3/7 \\ \hline 3/7 & -2/7 \\ \hline \end{array}$$

ALGUNAS PROPIEDADES DE LA INVERSA

La inversa de un producto matricial es igual al producto de las matrices inversas en orden contrario

$$(A\ B\ C)^{-1} = C^{-1}\ B^{-1}\ A^{-1}$$

La inversa de una matriz diagonal es otra matriz diagonal cuyos elementos son los inversos de los elementos de la matriz original.

$$\begin{bmatrix} 2 & & \\ & -3 & \\ & & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 1/2 & & \\ & -1/3 & \\ & & 1/4 \end{bmatrix}$$

Una matriz cuadrada compuesta por bloques diagonales puede ser invertida tomando las inversas de las submatrices respectivas :

$$\begin{bmatrix} A & & \\ & B & \\ & & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & & \\ & B^{-1} & \\ & & C^{-1} \end{bmatrix}$$

SOLUCION DE SISTEMAS DE ECUACIONES LINEALES POR METODOS MATRICIALES.

La aplicación principal del algebra matricial al análisis de sistemas de potencia (y en general al análisis ingenieril) es la solución de conjuntos de ecuaciones lineales de la forma

$$\begin{aligned} a_{11}\,x_1 + a_{12}\,x_2 + \ldots + a_{1n}\,x_n &= b_1 \\ a_{21}\,x_1 + a_{22}\,x_2 + \ldots + a_{2n}\,x_n &= b_2 \\ \vdots \qquad\qquad\qquad\qquad\qquad &\vdots \\ a_{n1}\,x_1 + a_{n2}\,x_2 + \ldots + a_{nn}\,x_n &= b_n \end{aligned} \qquad (1)$$

Este conjunto de ecuaciones puede escribirse en notación matricial como :

$$A \quad x \quad = \quad b$$

en donde :

A matriz cuadrada de coeficientes

b vector de constantes

x vector de incógnitas

El valor del vector de incógnitas x se puede encontrar premultiplicando ambos lados de la ecuación por la inversa de A (suponiendo que dicha inversa existe, es decir la matriz A es no singular),

$$A^{-1} A x = A^{-1} b$$

ó

$$x = A^{-1} b$$

En la práctica los sistemas grandes de ecuaciones no se resuelven por inversión directa, sino que se utilizan técnicas de dispersidad y alguno de los procesos de eliminación Gaussiana.

Un conjunto de ecuaciones lineales se puede resolver mediante operaciones elementales sobre los renglones. El objetivo de estas operaciones es el de transformar la matriz de coeficientes en una matriz triangular superior, con lo cual es posible obtener la solución por sustitución hacia atrás. Si cada operación sobre los renglones de A se efectúa también sobre los elementos correspondientes del vector b , el nuevo conjunto de ecuaciones $A x = b$ tendrá el mismo vector de solución x del sistema original. En la práctica, las operaciones elementales se efectúan sobre la matriz aumentada $[A , b]$ hasta que la matriz A es convertida a forma triangular. Una vez logrado esto el vector x se obtiene fácilmente por sustitución directa, como se ve a continuación :

Sea el sistema de ecuaciones lineales

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

El proceso consiste en llevar la matriz aumentada

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{array}\right]$$

a la forma :

$$\left[\begin{array}{cccc|c} 1 & a'_{12} & \cdots & a'_{1n} & b'_1 \\ 0 & 1 & \cdots & a'_{2n} & b'_2 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 1 & b'_n \end{array}\right]$$

mediante operaciones elementales sobre los renglones. La solución para x de este nuevo conjunto de ecuaciones se obtiene por sustitución hacia atrás de la siguiente manera

$$x_n = b'_n$$

$$x_{n-1} = b'_{n-1} - a'_{n-1,n} x_n$$

$$x_{n-2} = b'_{n-2} - a'_{n-2,n} x_n - a'_{n-2,n-1} x_{n-1}$$

$$\vdots$$

$$x_1 = b'_1 - \sum_{j=2}^{n} a'_{1j} \; x_j$$

La fórmula recursiva para el proceso de sustitución hacia atras se puede entonces escribir como :

$$x_i = b'_i - \sum_{j=i+1}^{n} a'_{ij} \; x_j \; ; \quad i = n, n-1, n-2, \ldots, 2, 1$$

Si se desea resolver un sistema de ecuaciones lineales $A \; x = b$ para varios valores del vector b, resulta mas conveniente llevar la matriz A a la forma :

$$A = L \; U$$

en donde L es una matriz triangular inferior, y U una matriz triangular superior. Para ilustrar este procedimiento conocido como factorización triangular $L \; U$, utilizemos la matriz de 3×3 :

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \quad \begin{bmatrix} U_{11} & U_{12} & U_{13} \\ 0 & U_{22} & U_{23} \\ 0 & 0 & U_{33} \end{bmatrix}$$

De esta igualdad surgen las siguientes relaciones entre los elementos de A L y U :

$$a_{11} = U_{11}$$

$$a_{21} = l_{21} \; U_{11}$$

$$a_{31} = l_{31} \; U_{11}$$

de donde podemos calcular los elementos de la primera columna de U y L. A continuación tenemos

$$a_{12} = U_{12}$$

$$a_{22} = l_{21} \; U_{12} + U_{22}$$

$$a_{32} = l_{31} \; U_{12} + l_{32} \; U_{22}$$

relaciones que nos permiten calcular los elementos de la segunda columna de U y L. Finalmente para nuestra matriz de 3 x 3 tenemos las relaciones

$$a_{13} = U_{13}$$

$$a_{23} = l_{21} \; U_{13} + U_{23}$$

$$a_{33} = l_{31} \; U_{13} + l_{32} \; U_{23} + U_{33}$$

de las cuales se encuentran los valores de la tercera columna de las matrices L y U.

Una vez lograda la transformación la solución del sistema de ecuaciones se obtiene de :

$$A\,x = L\,U\,x = L\,y = b \qquad (3)$$

resolviendo primero para y por sustitución hacia adelante la ec.

$$L\,y = b \qquad (4)$$

y a continuación por sustitución hacia atrás

$$U\,x = y \qquad (5)$$

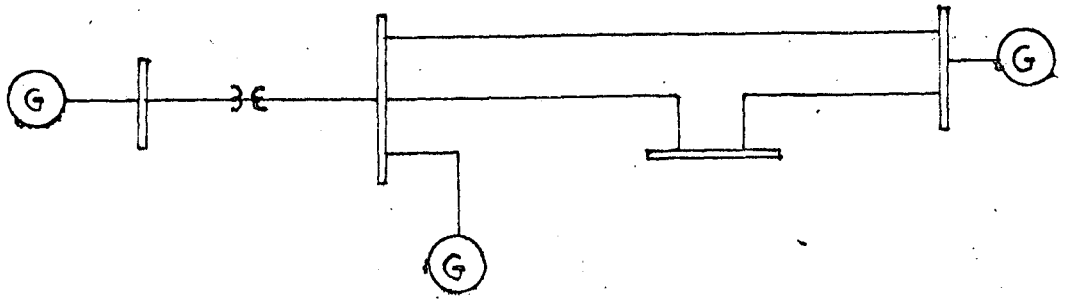
en donde la solución de las ecuaciones (4) y (5) es bastante simple debido a la estructura triangular de las matrices L y U.

GRAFICAS Y MATRICES DE INCIDENCIA

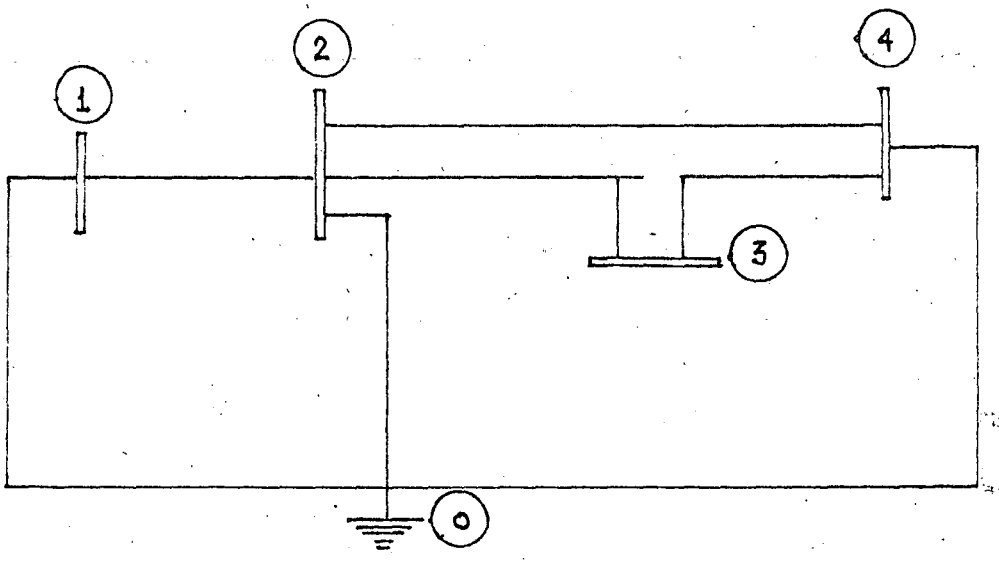
Para describir la estructura geométrica de una red eléctrica es suficiente reemplazar los componentes de la red por segmentos de línea independientemente de las características de dichas componentes. Estos segmentos de línea reciben el nombre de elementos y sus terminales se conocen por nodos. Un nodo y un elemento son incidentes si el nodo es una terminal del elemento. Los nodos pueden ser incidentes a uno o más elementos.

Una gráfica muestra la interconección geométrica de los elementos de una red y se dice que es conexa si y solo si, existe la posibilidad, a través de la conección de varios elementos, de llegar partiendo de cada uno de los nodos a todos los demás. Se dice además que la gráfica es orientada si se asigna una dirección a cada uno de sus elementos. Un ejemplo de representación de una red electrica y su gráfica orientada se presenta en la figura 1

a).-



b).-



c).-

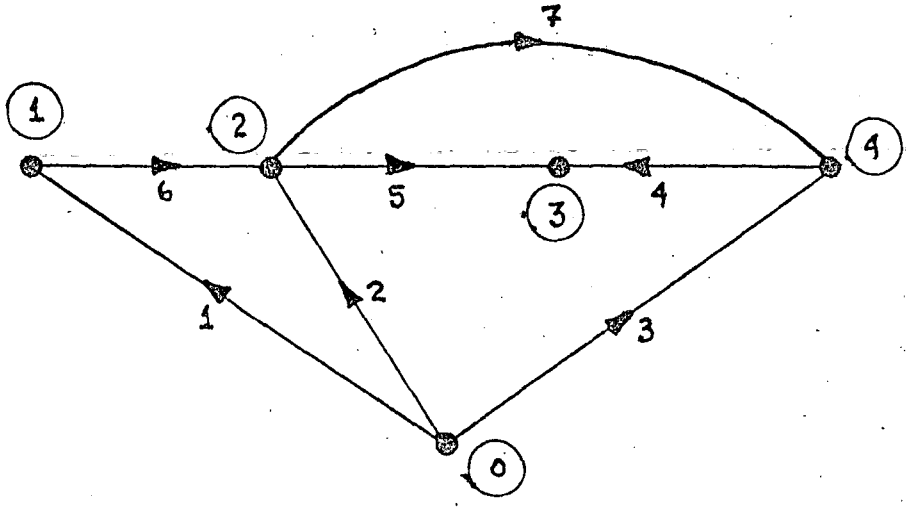


Figura 1 . Representación del sistema de Potencia
- a) Diagrama Unifilar
- b) Diagrama de Secuencia Positiva
- c) Gráfica Conexa Orientada

MATRIZ DE INCIDENCIA ELEMENTO-NODO

La información de la incidencia de elementos a nodos en una gráfica conexa se puede colectar en una matriz de incidencia - Elemento - Nodo $\hat{A}$, en la que sus componentes estan dadas por :

$a_{ij} = 1$ si el i - esimo elemento es incidente a y orientado saliendo del j - esimo nodo.

$a_{ij} = -1$ si el i - esimo elemento es incidente a y orientado hacia el j - esimo nodo.

$a_{ij} = 0$ si el i - esimo elemento no incide en el nodo i.

La dimensión de esta matriz es e x n , en donde e es el número de elementos y n es el número de nodos en la gráfica.

La matriz de incidencia elemento - nodo para la gráfica mostrada en la figura 1 es :

$\hat{A} =$

e \ n	⓪	①	②	③	④
1	1	-1			
2	1		-1		
3	1				-1
4				-1	1
5			1	-1	
6		1	-1		
7			1		-1

Por el hecho de que

$$\sum_{j=0}^{4} a_{ij} = 0 \qquad i = 1, 2, \ldots, e$$

se dice que las columnas de $\hat{A}$ son linealmente dependientes.

MATRIZ DE INCIDENCIA MEDIDOR - NODO

La matriz de incidencia Medidor - Nodo tiene aplicación en las técnicas de estimación de estado, y por ser estas parte del curso se incluye en estas notas. Para formar esta matriz los elementos los constituyen los medidores de flujos de potencia colocados sobre las líneas o medidores de voltaje y/ o inyecciones de potencia nodales. Para el caso de medidores de flujo de potencia en las líneas la gráfica se orienta de tal manera que si el medidor esta colocado del lado del nodo " i " en la línea que conecta " i " con " j" entonces la orientación es de i hacia j. En el caso en que existan medidores colocados en ambos extremos de la línea i - j la gráfica presentará dos elementos en paralelo conectando los nodos i - j . Si en alguna de las líneas digamos k - m no existe medición alguna, en la gráfica no aparecerá conexión entre los nodos k - m aunque la línea este en operación.

Para ilustrar estos conceptos se presenta en la figura 2 una pequeña red con los medidores representados en forma rectangular y su respectiva gráfica en donde los elementos son los medidores.

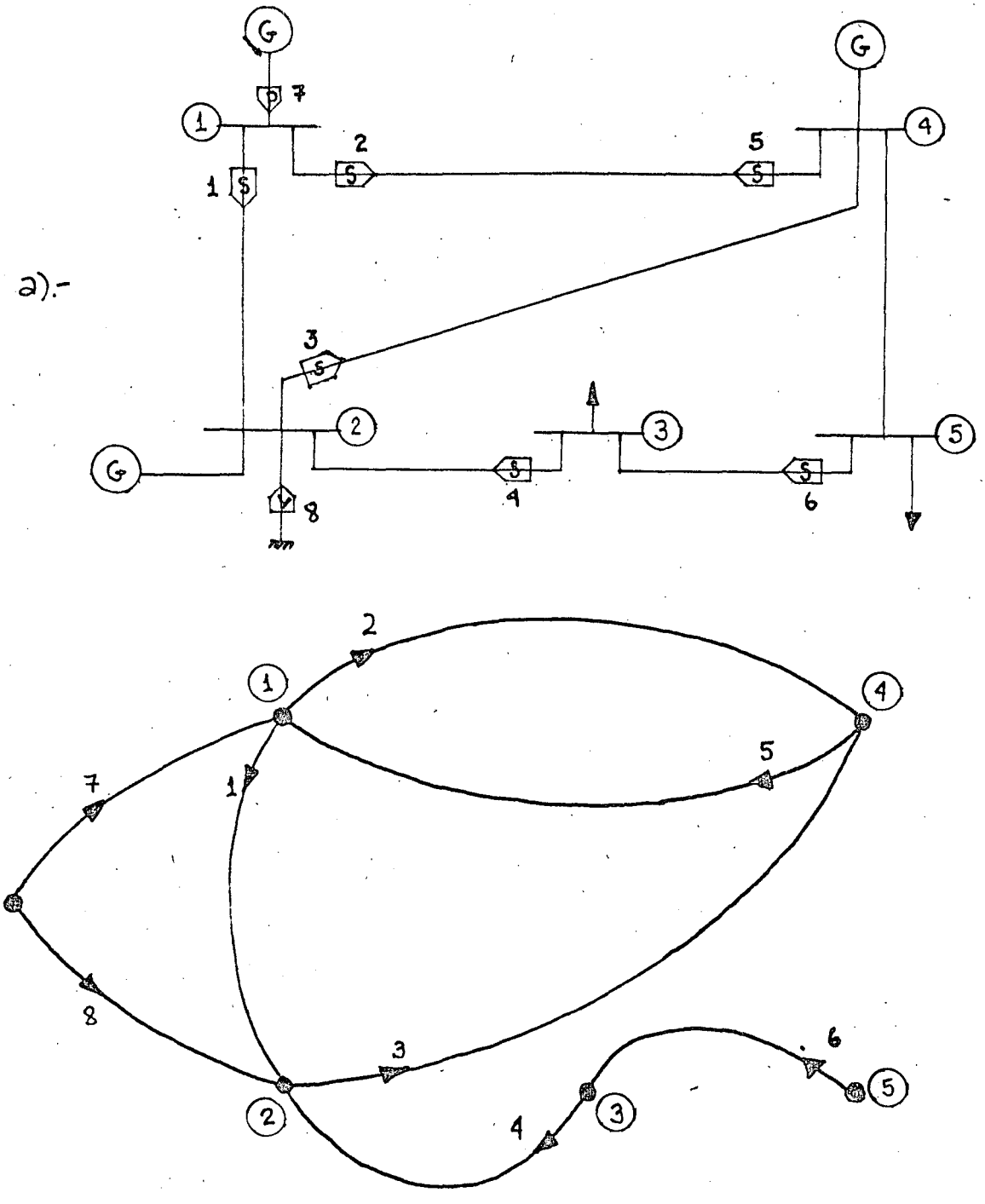


Figura 2. -Representación de una red eléctrica con medidores de Flujo de potencia en líneas [S] de inyección de potencia [P] y magnitud de voltaje [V]

a) Diagrama unifilar
b) Gráfica conexa orientada

La matriz de incidencia medidor - nodo de la gráfica presentada en la figura 2 es :

A =

m \ n	⓪	①	②	③	④	⑤
1		1	-1			
2		1			-1	
3			1		-1	
4			-1	1		
5		-1			1	
6				-1		1
7	1	-1				
8	1		-1			

B I B L I O G R A F I A

1.- G.W. Stagg y A.H. El - Abiad, " Computer Methods in Power System Analysis ", Mc. Graw Hill Book Co., 1968.

2.- A. Ralston, . " A First Course in Numerical Analysis ", Mc. Graw Hill Book Co., 1965.

3.- G. Strang, " Linear Algebra and its Applications ", Academic Press, 1976.

4.- Apuntes del Curso " Modeling and Analisys of Modern Power Systems " Energy Systems Research Center, University of Texas at Arlingt on.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA III MATRICES DISPERSAS

DR. SERGIO A. MOLINA GARCIA

ENERO, 1979.

NOTAS SOBRE MATRICES DISPERSAS.*

INTRODUCCION

D e f i n i c i ó n.

La primer idea asociada con matrices que tienen pocos elementos, es decir, pocos elementos diferentes de cero, es tratar de definirlas en términos cuantitativos por la razón de los - elementos de la matriz. Sin embargo, es mejor decir que una matriz con elementos dispersos es aquella en la cual, se puede aprovechar la proporción y/o la distribución de elementos igual a cero.

En términos cuantitativos, la densidad de una matriz está definida como el porcentaje del número de elementos diferentes de cero al número total de elementos de la matriz.

Existen varias ventajas asociadas a los sistemas descritos -- por matrices dispersas. La más evidente se tiene en el manejo de información, su almacenamiento y recuperación. Este ahorro que es considerable permite el análisis de mayores sistemas y modelos más exactos. Las mayores ganancias son obtenidas al - realizar operaciones sobre estructuras "mantenidas" en forma compacta. La computación (el proceso de cómputo) deberá estar organizada de tal forma que: Se eviten las operaciones con elementos iguales a cero (o unos), reducir al mínimo los costos -

* Estas notas forman parte de la tesis de maestría en ingeniería eléctrica, de la sección de graduados, IPN, de Ricardo - Mota Palomino.

de organización y las operaciones aritméticas involucradas. - Cuando se utilizan métodos directos para la solución del sistema de ecuaciones lineales, es importante el diseño de los algoritmos para preservar al máximo el número de elementos iguales a cero del sistema original.

P r o b l e m a.

El problema a resolver es :

La solución del conjunto de ecuaciones

$$A \underline{x} = \underline{b} \qquad (1)$$

Donde : A es una matriz no-singular con elementos diferentes de cero dispersos, de orden n, $\underline{x}$ y $\underline{b}$ son vectores n-dimensionales.

La mayor parte de los métodos directos para la solución de sistemas de ecuaciones lineales son variantes de la eliminación - gausiana. Aunque estas variantes tienden a comportarse de manera similar en el caso de la matriz A siendo prácticamente llena, la eficiencia de su ejecución con matrices dispersas es bastante variada y frecuentemente el algoritmo escogido puede afectar el esquema de dispersión.

Forma de Eliminación de la Inversa.

La descomposición LU de una matriz A, es el producto de una matriz triangular inferior L y una matriz triangular superior U. Esto lo podemos escribir

$$A = L\ U \qquad (2)$$

Las columnas de L pueden ser consideradas como los factores -- (elementos dispersos) de la matriz L^{-1} que es una matriz normalmente llena. El sistema de ecuaciones (1), puede ser resuelto en forma sencilla por la eliminación hacia adelante

$$L \underline{y} = \underline{b} \qquad (3)$$

seguido por la substitución hacia atrás

$$U \underline{x} = \underline{y} \qquad (4)$$

Ambos pasos son la solución de un sistema triangular de ecuaciones y es tal que la estructura de dispersión de las matrices L y U no es alterada por el proceso. El número de operaciones -- aritméticas involucradas es justamente el número combinado de - elementos diferentes de cero en L y U.

Es conveniente hacer notar que por motivos de estabilidad numérica y de mantener las características de dispersión originales, las matrices de permutación P y Q deberán ser incluídas en (2)

$$P A Q = L U \qquad (5)$$

En general, podemos decir que la mayor parte de los sistemas de ecuaciones a resolver obtenidos de los modelos matemáticos en - los sistemas de energía eléctrica, son numéricamente estables y para propósitos de la explicación siguiente, omitiremos las matrices P y Q.

CAPITULO 1.

Almacenamiento de Matrices Dispersas.

1.1 Introducción.

En compiladores comunes,el almacenamiento de arreglos matriciales de dos dimensiones se efectua en localidades de memoria contiguas ; por ejemplo,para una matriz de orden tres almacenada por columnas:

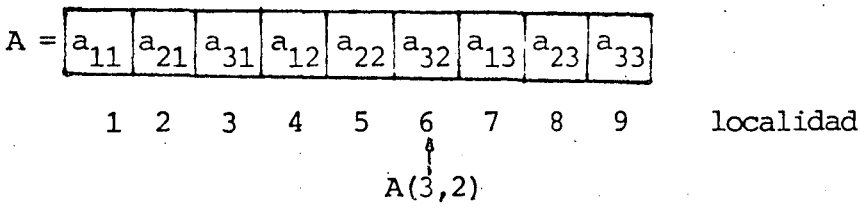


Figura 1.1 Almacenamiento de un arreglo matricial en localidades de memoria contiguas.

De modo que cualquier elemento a_{ij} del arreglo es interpretado por el compilador como : A(I,J)=A(MJ+I-M),donde M es el número de columnas para un arreglo de orden (N x M). Lo anterior significa que al requerir,por ejemplo ,un elemento A(3,2),en realidad nos estamos refiriendo a un elemento A(3x2 + 3 - 3)=A(6) como se ilustra en la figura 1.1 .

La regla de expansión anterior (i,j)=(Mx j +i -M), es construida en el compilador,obteniéndose la información en forma automática al referirse a un elemento del arreglo.

En el caso de matrices dispersas en las cuales no se desea almacenar los elementos nulos , se requiere diseñar un esquema de almacenamiento en arreglos vectoriales(unidimensionales) que nos permita referirnos a cualquier elemento (i,j) del arreglo matricial en forma sencilla. Esta técnica es lo que se conoce como empaquetamiento de matrices. Al empaquetar una matriz dispersa,a cambio de la reducción de localidades de memoria necesarias,

se tendrá que aportar una lógica de programación mas sofisticada,ya que los arreglos utilizados para almacenar la matriz,no corresponderán visualmente a un arreglo de dos dimensiones como los utilizados en lenguajes de programación convencionales.

1.2 Almacenamiento de matrices dispersas en forma de vectores.

Una vez conocida la naturaleza de los elementos de la matriz ó arreglo por empaquetar,la información no nula puede ser almacenada en arreglos unidimensionales reconociendo los atributos que definen a cada elemento del arreglo,esto es:

I número de fila del elemento considerado.

J número de columna del elemento considerado.

V valor del elemento.

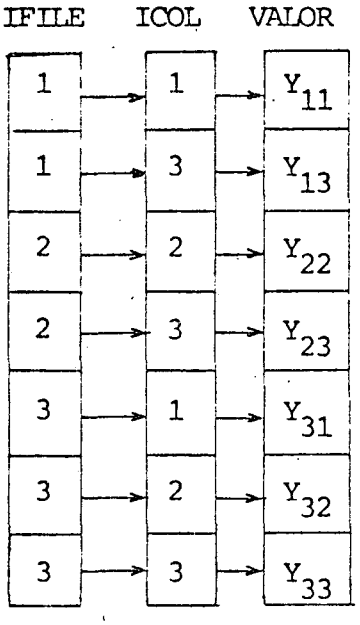
En un arreglo matricial convencional,se almacenarían únicamente los elementos V,pero el hecho de manejar explícitamente (I,J) obliga a tener una gran cantidad de elementos cero almacenados.

Se pueden en base a lo anterior proponer diferentes esquemas para almacenar una matriz dispersa,variando uno de otro por la forma en que se manejan los atributos enlistados y se ilustra con dos de ellos,utilizables prácticamente, ilustrando su aplicación al almacenar la matriz:

$$Y = \begin{array}{|c|c|c|} \hline Y_{11} & Y_{12} & Y_{13} \\ \hline Y_{21} & Y_{22} & Y_{23} \\ \hline Y_{31} & Y_{32} & Y_{33} \\ \hline \end{array}$$

1.2.1 Primer esquema de empaquetamiento.

En este primer esquema ,se almacenan los tres atributos mencionados en la sección anterior,únicamente para cada elemento diferente de cero del arreglo, en tres vectores como se ilustra:



Si la matriz que se empaqueta es una matriz de admitancias de nodo de una red eléctrica,el orden de los vectores que se requieren para empaquetarla serán,dadas las características de simetría de la matriz,si esta se empaqueta completa:

IFILE = N + 2B (Vector entero).

ICOL = N + 2B (Vector entero).

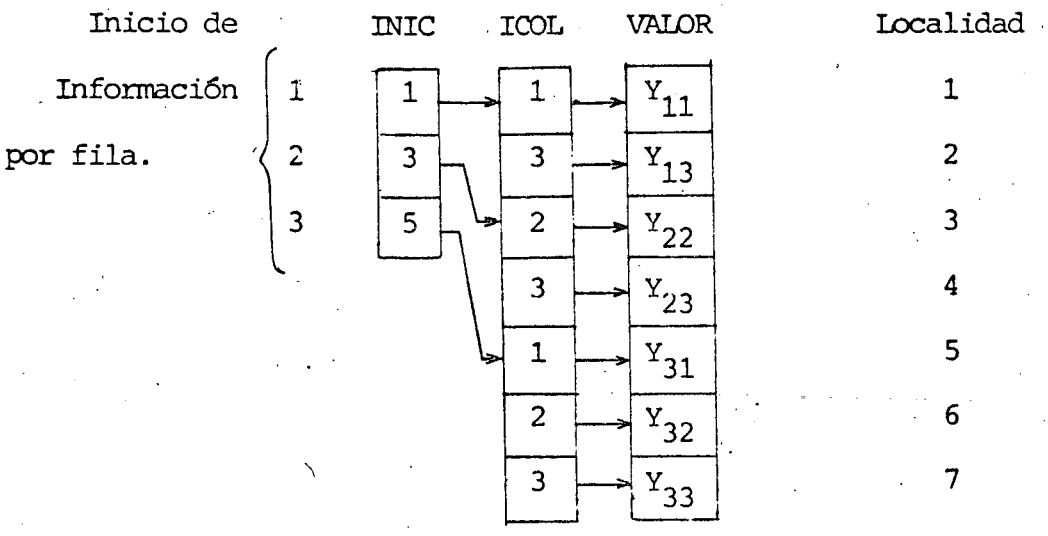
VALOR = N + 2B (Vector complejo).

donde N es el número de nodos y B el número de ramas de la red.

Un inconveniente de este esquema elemental sería que al ahorrar localidades de memoria ,se sacrifique flexibilidad por requerirse de una busqueda exahustiva en los vectores IFILE e ICOL para obtener algún elemento de interés del arreglo.

1.2.2. Segundo esquema de empaquetamiento.

En el esquema anterior se puede observar que en el vector IFILE se almacena información redundante,la cual se puede simplificar almacenando la información ya sea por fila ó por columna,suprimiéndose así el vector IFILE ,y siendo necesario introducir un arreglo que informe la localidad en que se inicia la información de la fila ó columna I en los vectores ICOL,VALOR y que denominaremos INIC.



El orden de los vectores utilizados será ahora:

INIC = N (Vector entero).

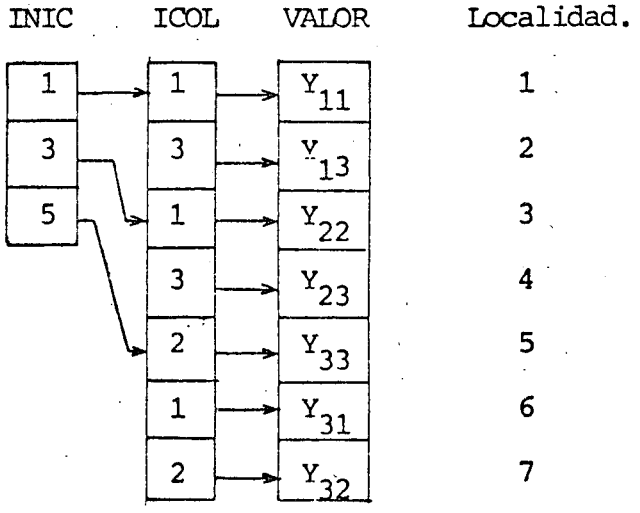
ICOL = N+2B (Vector entero).

VALOR =N+2B (Vector complejo).

En este esquema ,la inconveniencia apuntada en el esquema anterior es disminuida ya que al requerirse un elemento del arreglo,la busqueda de este se reduce a un número limitado de elementos de ICOL,ya que de INIC se conoce el rango de localidades en que se puede encontrar alojado el elemento buscado.

1.2.3. Tercer esquema de empaquetamiento.

La información almacenada mediante el esquema anterior puede ser mejorada en contenido si se coloca en la primera localidad de cada fila ó columna I lo siguiente: En VALOR el elemento diagonal de la fila y en la posición correspondiente de ICOL el número de elementos existentes fuera de la diagonal como sigue:



El orden de los vectores utilizados en este esquema es el mismo que en el esquema anterior pero por la información contenida es mas sencilla la busqueda de un elemento de interés,utilizando INIC e ICOL.

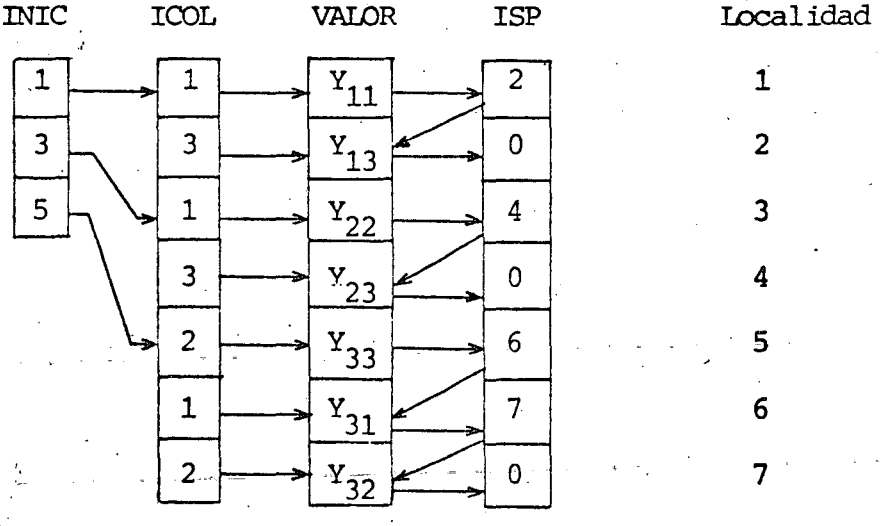
Es conveniente observar que en el caso de matrices simétricas,cualquiera de los esquemas anteriores puede ser modificado a fin de almacenar únicamente una triangular de ella,pero a medida que se compacta mas la información,va resultando mas complicado el acceso a los elementos del arreglo empaquetado.

Los esquemas anteriores dan solo una idea de la técnica para almacenar matrices dispersas y en general el esquema utilizado en una aplicación debe ser tal que se obtenga un buen compromiso entre los requerimientos de memoria y la eficiencia con que se tiene acceso a la información empaquetada.

1.3. Listas encadenadas.

La posibilidad de empaquetar matrices aun en localidades no contiguas de memoria se amplia grandemente mediante la utilización de las llamadas "listas encadenadas". El encadenamiento consiste en agregar a la información de cada elemento del arreglo empaquetado,un apuntador que indique la localidad en que se encuentra otro elemento con alguna característica común al considerado,por ejemplo en este caso el pertenecer a una misma fila ó columna.

Para el tercer empaquetamiento, se tendría ahora una lista encadenada como la siguiente:



El encadenamiento de la información puede ser seguido para la fila 3 como

ejemplo. En la localidad INIC(3) se tiene un 5 almacenado,que como lo indica la flecha correspondiente,apunta a la localidad en que se inicia la información de la fila 3 en los vectores ICOL,VALOR e ISP. En ICOL(5) se almacena un 2 que significa que existen dos elementos fuera de la diagonal en la fila tres de la matriz Y. En VALOR(5) se almacena el elemento diagonal Y_{33}. En ISP(5) se almacena un 6 que indica que en la localidad 6 de ICOL,VALOR e ISP existe otro elemento de la misma fila. En ICOL(6) está almacenado un 1 que indica que en la primera columna de la fila 3 existe un elemento diferente de cero. En VALOR(6) se almacena el valor de este elemento Y_{31} y en ISP(6) se almacena un 7 que apunta a la localidad 7 de ICOL,VALOR e ISP indicando que en esas localidades existe información de la fila 3. En ICOL(7) se almacena un 2 indicando que en la columna 2 de la tercera fila existe un elemento diferente de cero ;el valor de este está almacenado en VALOR(7) (Y_{32});finalmente en ISP(7) se almacena un 0,que significa que esa es la última posición en los vectores ISP,VALOR e ICOL en que existen elementos de la fila 3.

Se observa que para almacenar una matriz dispersa como una lista encadenada,se requiere además de los vectores mencionados en cualquiera de los esquemas de empaquetamiento presentados,de un vector de encadenamiento conocido como de "siguiente posición" ISP de dimensión N+2B y entero.

La desventaja de requerir un vector adicional de apuntadores se justifica ampliamente por la gran flexibilidad que se logra para efectuar modificaciones en la información empaquetada.

Se debe mencionar que la utilización de listas encadenadas para almacenar matrices dispersas es solo una aplicación particular,siendo posible emplearlas en otras áreas dentro de aplicaciones computacionales.

1.3 Modificación de matrices empaquetadas.

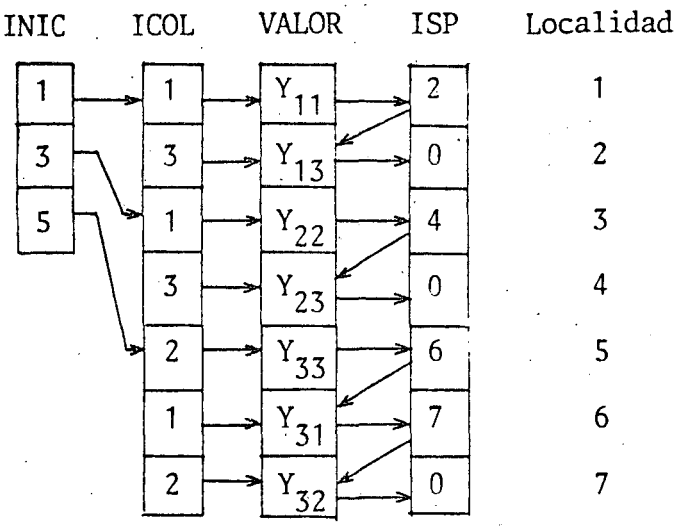
La modificación de una lista encadenada utilizada para almacenar una matriz dispersa se ilustra utilizando la matriz de la sección 1.2 .

1.3.1. Adición de elementos.

Supóngase que la matriz Y está empaquetada como se muestra:

$Y =$

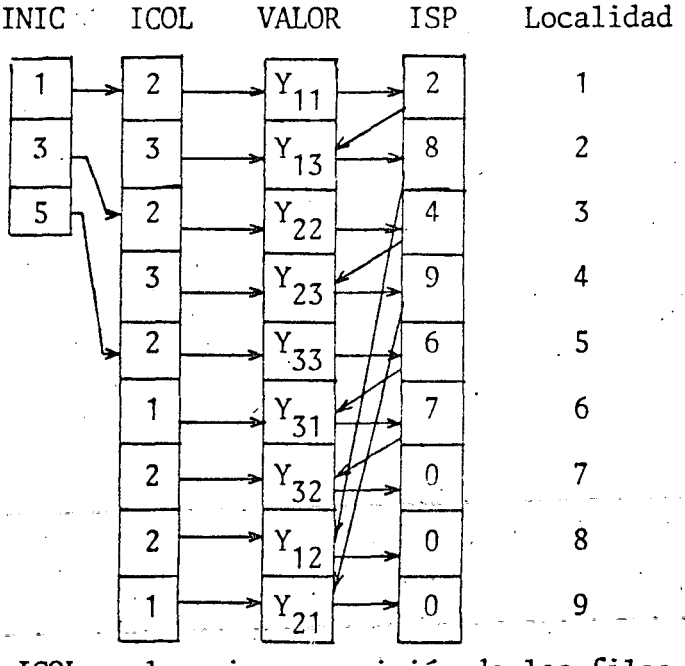
Y_{11}	0	Y_{13}
0	Y_{22}	Y_{23}
Y_{31}	Y_{32}	Y_{33}



El arreglo anterior se va a modificar para incluir los elementos Y_{12} y Y_{21}. Si el esquema de almacenamiento utilizado fuera secuencial(en localidades de memoria contiguas,se necesitarían localidades de memoria auxiliares para recorrer toda la información de los vectores ICOL y VALOR a fin de insertar los elementos que se desea agregar en las posiciones adecuadas. Lo anterior resulta impráctico en el caso de matrices grandes. Sin embargo utilizando listas encadenadas las modificaciones resultan sumamente sencillas. Esto se logra agregando Y_{12} y Y_{21} y en general cualquier elemento que se agregue,al final de los vectores correspondientes y encadenando la información a la de la fila en que se agregan ,mediante el vector ISP como se muestra:

$Y =$

Y_{11}	Y_{12}	Y_{13}
Y_{21}	Y_{22}	Y_{23}
Y_{31}	Y_{32}	Y_{33}



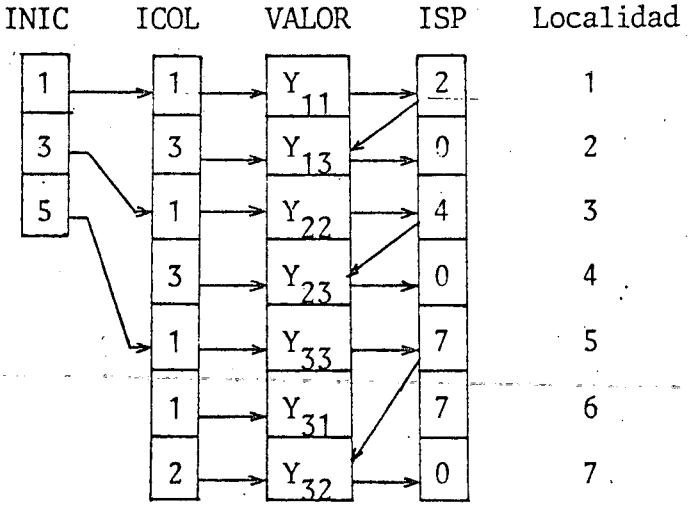
Nótese que se modifica el vector ICOL en la primera posición de las filas 1 y 2

para reflejar el aumento en número de elementos fuera de la diagonal de esas filas. El vector de encadenamiento ISP se modifica en la posición que contenía un 0 indicando el fin de información de la fila 1 y apuntando ahora a la posición 8 que es donde se localiza el elemento insertado Y_{12}. En ISP(8) se coloca un cero indicando que finaliza en esas posiciones la información de la fila. Análogamente ISP(4) que contenía un cero indicando el fin de los elementos de la fila 2 se modifica para apuntar a la posición 9 en que se inserta el elemento Y_{21}, colocándose en la localidad correspondiente ISP(9) un cero para indicar el fin de la información de la fila 2.

1.3.2. Remoción de elementos.

En la misma matriz empaquetada en el ejemplo anterior, supóngase que se desea eliminar el elemento Y_{31}. Lo anterior se logra eliminando el encadenamiento del elemento a la información de la fila como se muestra:

$$Y = \begin{bmatrix} Y_{11} & 0 & Y_{13} \\ 0 & Y_{22} & Y_{23} \\ 0 & Y_{32} & Y_{33} \end{bmatrix}$$



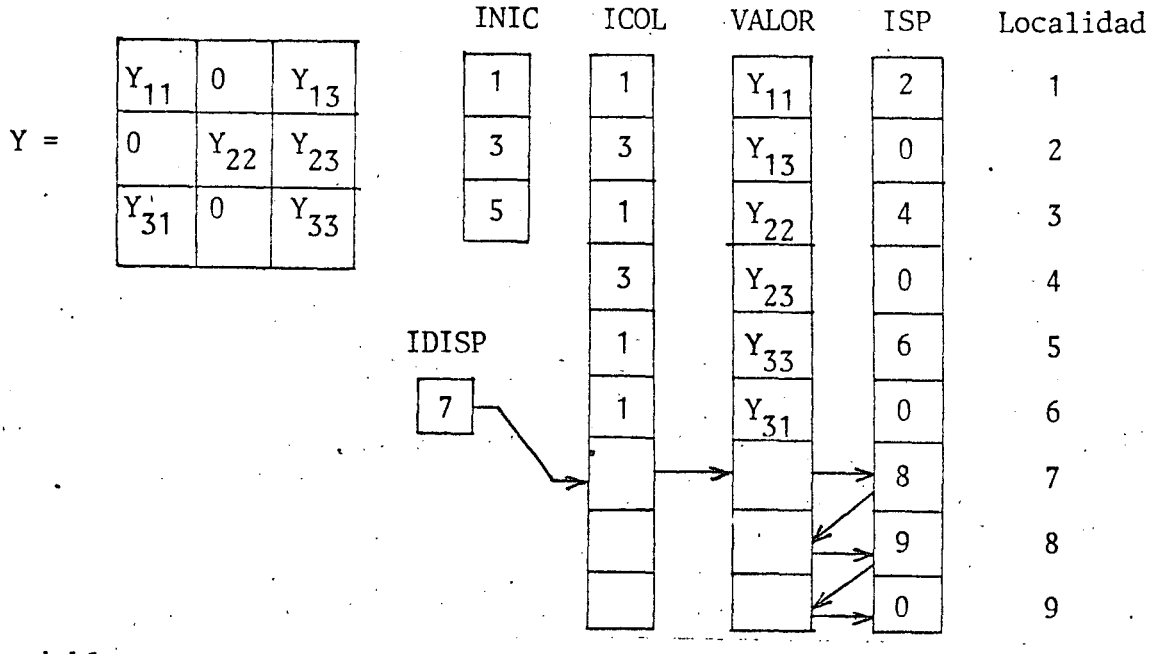
El elemento de ISP que apunta a la localidad 6 de VALOR que contiene el elemento que se desea eliminar (Y_{31}), se modifica para hacerlo ahora apuntar hacia la posición en que se localiza el siguiente elemento de la fila. También el elemento diagonal de ICOL se modifica para reflejar la reducción en el número de elementos fuera de la diagonal de la fila.

1.3.3. Condensación de listas ó "Recolección de Basura".

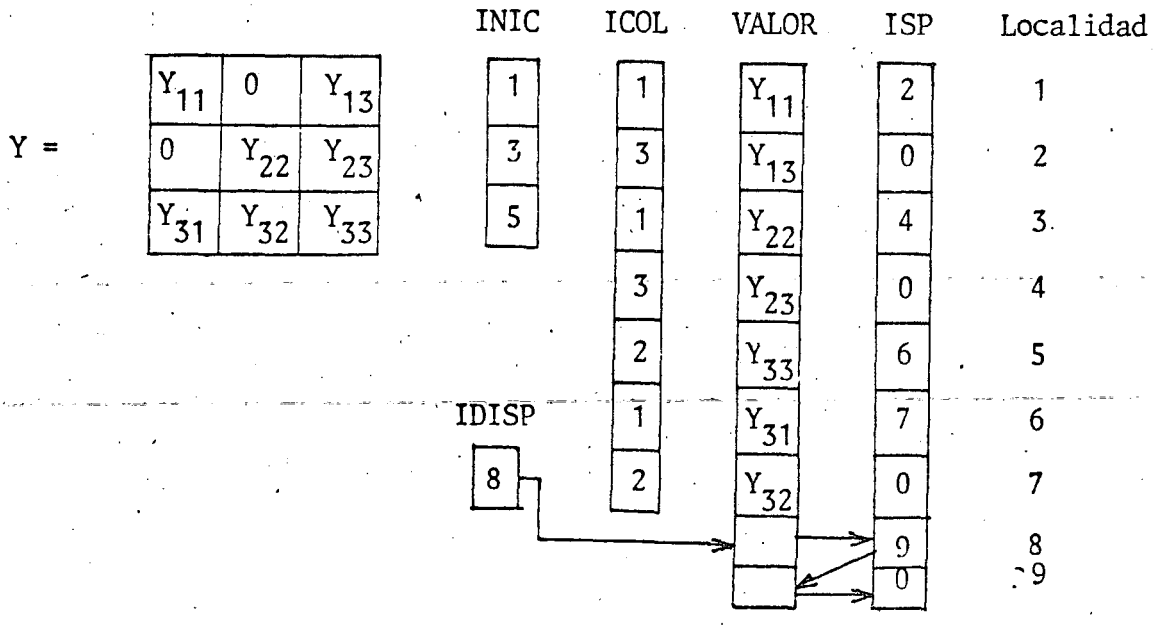
En el caso de procesos de factorización matricial en que se presenta el caso de adición ó eliminación de elementos en el arreglo empaquetado, las localidades dispo

nibles para realizar el procedimiento pueden ser mejor aprovechadas manteniendo una lista encadenada de localidades disponibles para alojar los elementos que se crean durante el proceso. En el esquema hasta ahora utilizado para ilustrar el empleo de la listas encadenadas,esta lista de localidades disponibles puede ser mantenida en el mismo vector de siguiente posición ISP ,utilizando un indicador que apunte a la primera posición disponible para alojar un nuevo elemento creado. El mantenimiento de esta lista se ilustra en el siguiente ejemplo:

Sea la matriz empaquetada:

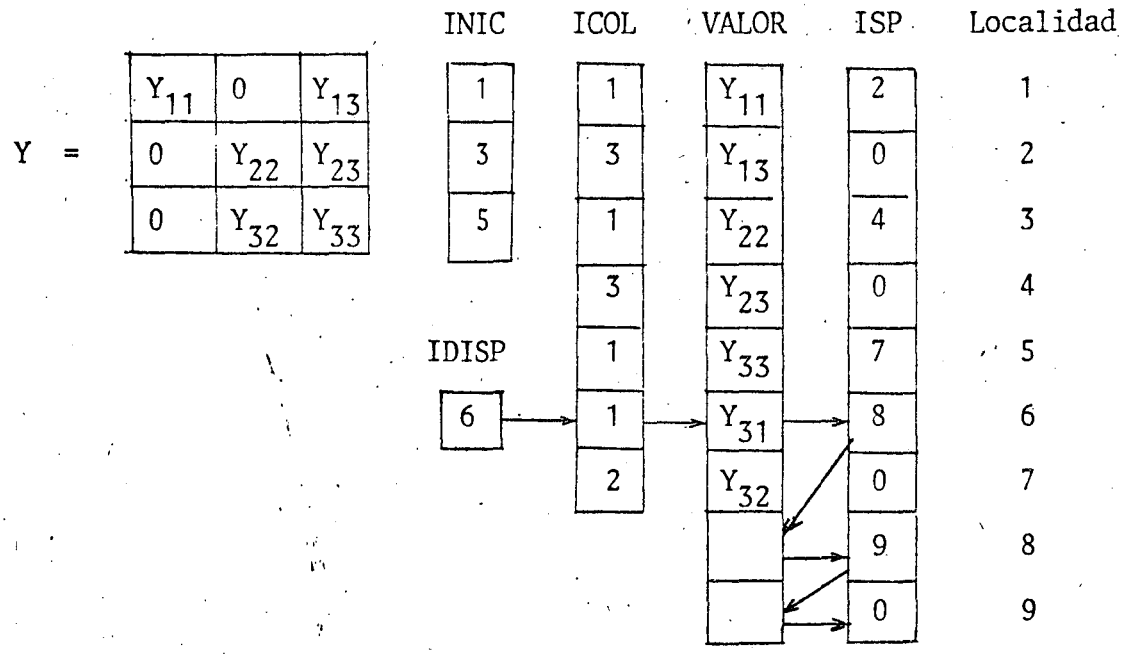


La variable entera IDISP apunta a la localidad 7 de ICOL,VALOR e ISP que son las localidades en que puede ser introducido algún elemento nuevo en el arreglo. Supóngase que se agrega el elemento Y_{32}; los vectores se modifican como:



Los vectores del arreglo se modifican en una forma análoga a la ilustrada en la sección 1.3.1. Ahora la variable IDISP apunta a la localidad 8,que es la nueva primera localidad disponible para alojar un elemento. El elemento correspondiente ISP(8) apuntará a la siguiente posición disponible y así sucesivamente hasta encontrar un cero que indicará el fin de las posiciones disponibles para alojar nuevos elementos en el arreglo.

El caso de remoción de elementos se ilustra suponiendo que en el arreglo anterior se desea eliminar el elmento Y_{31}. Las modificaciones en los vectores serían:



Se observa que la localidad que ocupaba la información del elemento Y_{31} ahora ha sido encadenada en la lista de elementos disponibles,ya que ahora IDISP apunta hacia esas localidades de ICOL,VALOR e ISP. El elemento correspondiente de IDISP se actualiza para continuar el encadenamiento hacia la siguiente localidad libre mientras que los vectores ICOL,VALOR e ISP se modifican para reflejar la eliminación del elemento de acuerdo al procedimiento de la sección 1.3.2 .

Mediante el procedimiento ilustrado para mantener la lista de posiciones de memoria disponibles,siempre la primera de estas posiciones será la colocada mas arriba en los vectores en que se almacena el arreglo.

CAPITULO 2.

Esquemas de Ordenamiento.

2.1. Introducción.

Las matrices de coeficientes que resultan del análisis de SEP,poseen generalmente dispersas,simétricas y con dominancia diagonal. Resulta importante tratar de mantener las características anteriores durante la solución de algún problema en particular.

Cualquier procedimiento utilizado para mantener la dispersidad de la matriz de coeficientes durante el proceso de factorización para resolver sistemas de ecuaciones se reduce,dadas las características de dominancia diagonal,a seleccionar el elemento pivote ,en métodos derivados del básico de eliminación de Gauss,únicamente entre los elementos diagonales de la matriz,obteniéndose implícitamente una buena precisión numérica y manteniéndose la simetría de la matriz de coeficientes durante el proceso.

2.2. Matrices de Incidencia y Gráficas de flujo.

Una manera simple de analizar la efectividad de ordenar las ecuaciones de un sistema a fin de preservar la dispersidad de la matriz de coeficientes,es mediante la aplicación de conceptos topológicos de gráficas de flujo y matrices de incidencia

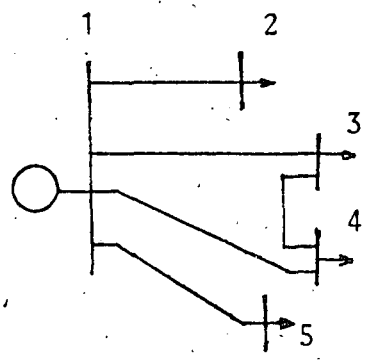
La estructura de una matriz Y de orden N se puede representar por una matriz de incidencia B cuyos elementos serán:

$b_{ij} = 1$ Si el elemento y_{ij} de la matriz de coeficientes es diferente de cero.

$b_{ij} = 0$ Si el elemento y_{ij} de la matriz de coeficientes es cero.

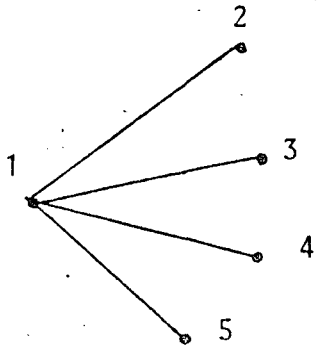
A esta matriz de incidencia le corresponderá una gráfica de flijo que se puede construir incluyendo un número de nodos igual al orden de la matriz de incidencia y donde a cada uno de los nodos le corresponde una fila y/o columna de la matriz. Estos nodos son unidos por segmentos de recta si las intersecciones entre filas y columnas correspondientes a ellos contienen elementos diferentes de cero.

A continuación se muestra un SEP elemental ,la matriz de incidencia de su matriz de admitancias de nodo y la correspondiente gráfica de flujo:



Y_{11}	Y_{12}	Y_{13}	Y_{14}	Y_{15}
Y_{21}	Y_{22}	0	0	0
Y_{31}	0	Y_{33}	Y_{34}	0
Y_{41}	0	Y_{43}	Y_{44}	0
Y_{51}	0	0	0	Y_{55}

1	1	1	1	1
1	1	0	0	0
1	0	1	1	0
1	0	1	1	0
1	0	0	0	1



a) b) c)

Figura 2.1. a)SEP elemental. b)Matriz de incidencia correspondiente a su matriz de admitancias de nodo. c)Gráfica de flujo.

Se debe notar que en el caso de SEP y de una red eléctrica en general,la gráfica de flujo obtenida de la matriz de incidencia de la red ,derivada de su matriz de admitancias de nodo,corresponderá a la gráfica lineal inherente a la propia red,lo que significa que existirá una correspondencia entre red-matriz de admitancias de nodo y gráfica de flujo-matriz de incidencia del sistema,que se explota en la obtención de equivalentes reducidos de las redes .

La utilidad de estos conceptos para evaluar la eficiencia de ordenar las ecuaciones en un proceso de factorización matricial derivados del método de eliminación de Gauss estriban en el hecho de que al eliminar una variable del sistema de ecuaciones que se resuelve,se obtendrá una matriz de coeficientes reducida que puede a su vez ser representada por una gráfica de flujo reducida. Esta gráfica puede obtenerse sin necesidad de realizar el proceso de eliminación de la variable algebráicamente,simplemente eliminando el nodo que corresponde a la ecuación de la variable que se elimina e interconectando todos los nodos que incidian al nodo eliminado. En el caso de que dichos nodos no estuvieran conectados antes de eliminar un nodo,se introducirán ramas. Lo anterior podría ser interpretado como debido a la necesidad de que el estado de los nodos de la gráfica que se conectan a través del nodo que se elimina,debe de permanecer invariable ante la eliminación.

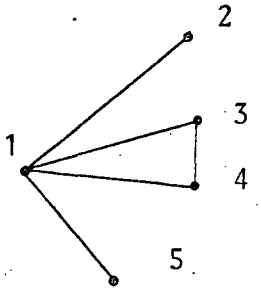
2.2. Efecto de reordenar las ecuaciones de un sistema.

En una red eléctrica,el efecto de reordenar la numeración de los nodos de la red,origina un reordenamiento en las ecuaciones de los nodos (en una formulación nodal).

Esto se reflejará en la disposición de los elementos cero de la matriz de admitancias

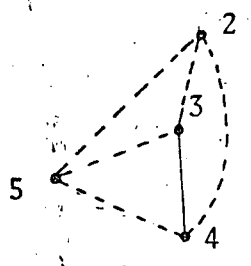
de nodo,y consecuentemente en la matriz de incidencia. Sin embargo la gráfica de flujo resultará la misma que antes de reordenar.

Usando la matriz de incidencia y gráfica de flujo del SEP de la figura 2.1,se pue de simular la eliminación de variables de las ecuaciones de nodo,eliminando nodos de la gráfica de flujo como se mencionó en la sección anterior,en el orden en que están numerados los nodos:



1	1	1	1	1
1	1	0	0	0
1	0	1	1	0
1	0	1	1	0
1	0	0	0	1

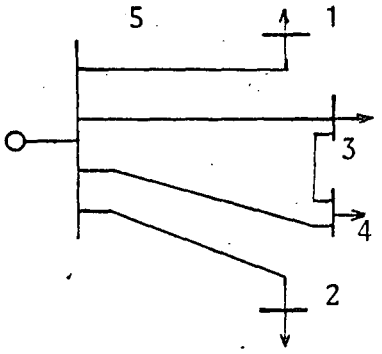
Al eliminar el nodo 1 de la gráfica,se obtiene una gráfica reducida,que corresponde a una matriz de incidencia reducida,que a su vez se derivaría de eliminar la variable de la primera ecuación de nodo de la red,como la siguiente:

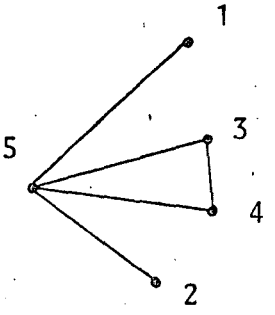


1	1	1	1	1
1	1	Ø	Ø	Ø
1	Ø	1	1	Ø
1	Ø	1	1	Ø
1	Ø	Ø	Ø	1

Las lineas punteadas en la gráfica de flujo reducida así como los Ø en la matriz de incidencia representan los elementos que se introducen al eliminar algebráicamente la primera variable del sistema.En este caso todos los nodos han quedado interconectados resultando también una matriz de incidencia reducida llena. El proceso de eliminación de nodos puede continuarse en cualquier orden sin afectar la estructura de la matriz reducida.

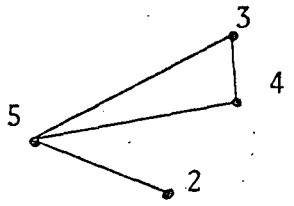
Supóngase que ahora los nodos de la red se reenumeran de acuerdo a la figura siguiente:





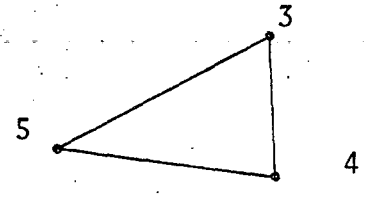
1	0	0	0	1
0	1	0	0	1
0	0	1	1	1
0	0	1	1	1
1	1	1	1	1

Al eliminar la primera variabledel sistema de ecuaciones de nodo de la red,estructuralmente sería equivalente a eliminar al nodo 1 de la gráfica de flujo resultando la matriz de incidencia y gráfica reducidas como se muestra:



1	0	0	0	1
0	1	0	0	1
0	0	1	1	1
0	0	1	1	1
1	1	1	1	1

El eliminar la segunda variable sería equivalente a eliminar el segundo nodo de la gráfica de flujo resultando:



1	0	0	0	1
0	1	0	0	1
0	0	1	1	1
0	0	1	1	1
1	1	1	1	1

Igualmente se puede proceder a eliminar el resto de las variables(3 y 4),momento en que algebráicamente se llegaría al problema de resolver una ecuación en una incógnita.

El ejemplo anterior ilustra el hecho de que el simple reordenamiento de los nodos de un SEP afecta en forma determinante la estructura dispersa de la matriz de coeficientes que describe su comportamiento,al aplicarse un método de factorización matricial para resolverlo.

En el caso del ejemplo se observa que en el segundo orden de eliminación no se introducen nuevos elementos en el proceso de eliminación,es decir se preserva la estructura

de dispersidad de la matriz de coeficientes original. En el primer orden de eliminación en cambio,al eliminar la primera variable,esta estructura de dispersidad se destruye dando lugar a una matriz llena.

2.3. Esquemas de Ordenamiento.

Al aplicar un método de ordenamiento de ecuaciones para resolver un sistema lineal,se debe de considerar que esta técnica es sólo una herramienta para mejorar la eficiencia del método de solución utilizado,de modo que su aplicación deberá emplear un tiempo de ejecución razonable ó recuperable al final de una aplicación. Aún cuándo actualmente se trata de obtener un algoritmo que proporcione un orden que minimize el número de elementos introducidos durante el proceso de solución ,actualmente el obtener dicho orden para cada red en particular resulta prohibitivo en tiempo de procesamiento lo que a causado que se empleen métodos que proporcionan resultados subóptimos pero que emplean recursos computacionales moderados. Se presentan tres esquemas de ordenamiento subóptimo que han mostrado proporcionar resultados satisfactorios en aplicaciones de SEP,aunque las modificaciones a realizar en la presentación básica para el caso de matrices no simétricas o con elementos diagonales iguales a cero ha sido presentada .

Esquema 1. Este esquema consiste simplemente en ordenar en forma ascendente las ecuaciones de acuerdo al número de elementos diferentes de cero fuera de la diagonal que contengan,ó en otras palabras de acuerdo al número de ramas que incidan al nodo correspondiente de la gráfica de flujo. Se dice que este esquema es de preordenamiento estático porque únicamente reordena ecuaciones de acuerdo a la información de la matriz de coeficientes original. En este esquema si dos o mas variables cumplen con la condición de tener el mínimo número de elementos fuera de la diagonal,se selecciona la variable de cualquiera de ellas como siguiente a eliminar .

Esquema 2. Selecciona como variable a eliminar,aquella cuya ecuación contiene el menor número de elementos fuera de la diagonal ó el menor número de ramas incidiendo en el nodo correspondiente en la gráfica de flujo reducida en cada paso del proceso de eliminación. Si dos ó mas variables cumplen con la condición anterior,se selecciona cualquiera de ellas.

Esquema 3. Selecciona como variable a eliminar aquella que al ser eliminada introduce el menor número de elementos en la matriz de coeficientes reducida ó menos ramas

en la gráfica de flujo correspondiente en cada paso del proceso de eliminación.

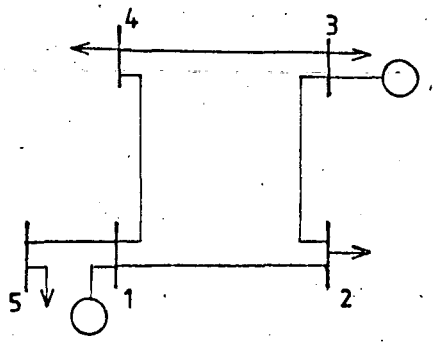
Se dice que los dos últimos esquemas son dinámicos porque toman en cuenta la información de la matriz de coeficientes o gráfica reducidas y no simplemente la información de la matriz de coeficientes original.

Los tres esquemas se listan en orden creciente de complejidad y recursos computacionales requeridos para implementarlos. El esquema 1 es el mas simple y requiere de un tiempo mínimo de ejecución,pero al ser aplicado a sistemas grandes produce un número de elementos creados durante el proceso de eliminación de variables grande. El esquema 2 es el mas ampliamente utilizado por proporcionar un buen compromiso entre tiempo de ejecución y resultados obtenidos en dispersidad. El esquema 3 requiere de mayor tiempo de ejecución así como lógica de programación,pero los resultados obtenidos respecto al esquema 2 no son superiores consistentemente.

2.4. Ejemplo de aplicación de los esquemas de ordenamiento.

A continuación se muestra la aplicación de las matrices de incidencia y gráficas de flujo,aunadas a los métodos de empaquetamiento presentados en el Capítulo 1 para aplicar los métodos de ordenamiento subóptimos.

Considerando la red de la figura 2.2:



X	X	0	X	X
X	X	X	0	0
0	X	X	X	0
X	0	X	X	0
X	0	0	0	X

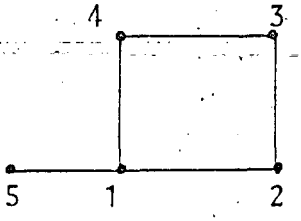
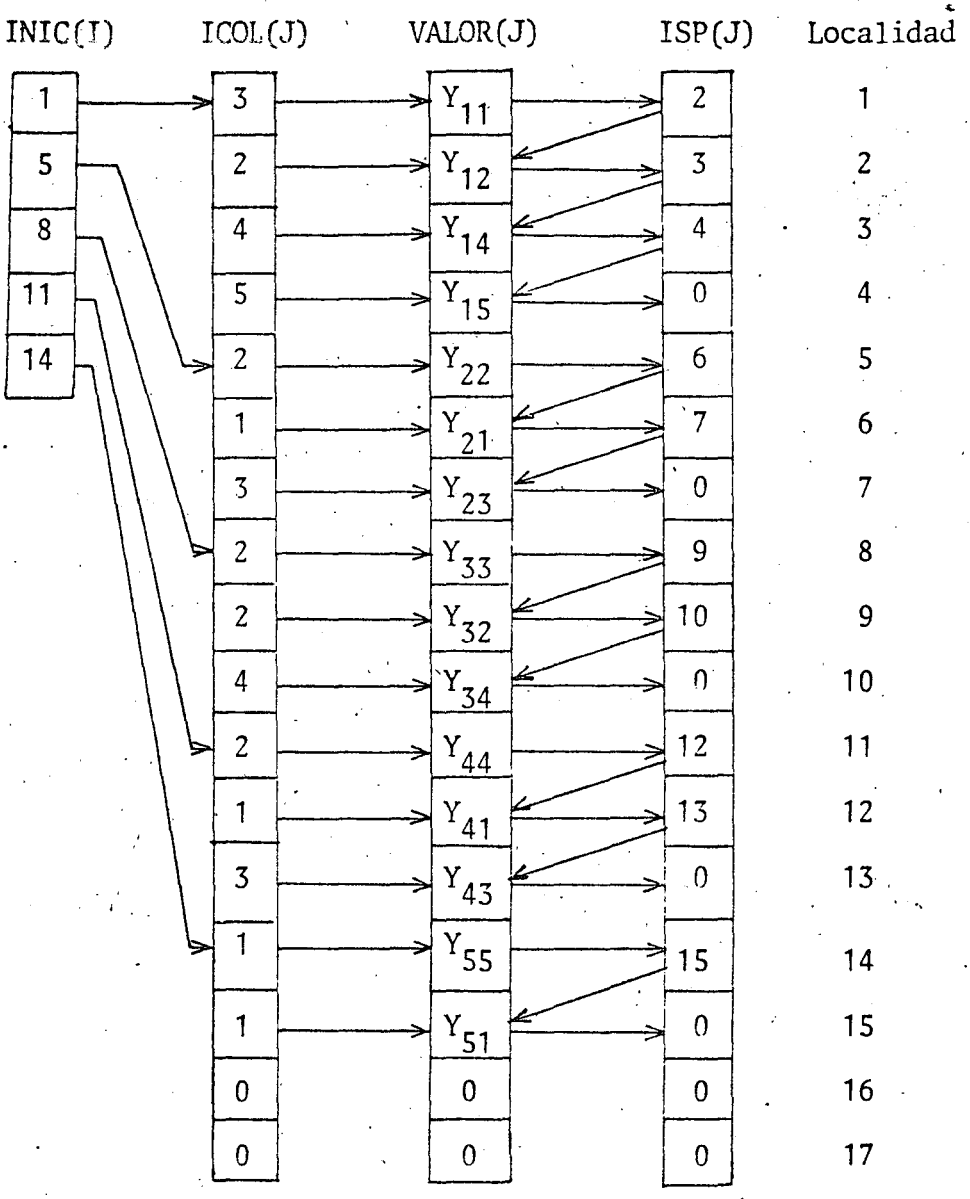


Figura 2.2. a)SEP elemental,su matriz de incidencia y gráfica de flujo.

La información de la matriz de admitancias de nodo del sistema anterior podría ser almacenado en arreglos unidimensionales como se muestra:



Un orden de eliminación de variables desde el punto de vista de dispersidad se determina, según el esquema 1 de ordenamiento ,simplemente comparando el número de elementos que incide. a cada nodo de la gráfica de flujo. En los arreglos en que se empaqueta la matriz de admitancias de nodo,esta información está en las primeras posiciones de la información de cada nodo ó fila de la matriz en el vector ICOL(J), posiciones a las que apuntan los elementos de INIC(I) . El nuevo orden de eliminación de nodos ó variables se podría almacenar en un vector NORDEN(I) paralelo al vector INIC(I) como se muestra:

Número de elementos que inciden al nodo I.

ICOL(INIC(I))

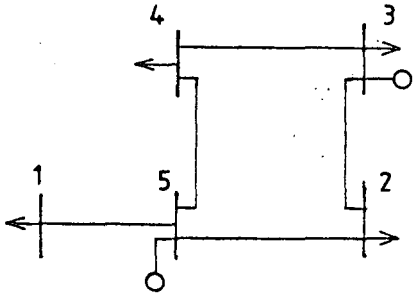
3
2
2
2
1

Vector de nuevo orden de eliminación de nodos .

NORDEN(I)

5
2
3
4
1

El orden de eliminación determinado indica que los nodos del sistema deberán ordenarse como se indica:



X	0	0	0	X
0	X	X	0	X
0	X	X	X	0
0	0	X	X	X
X	X	0	X	X

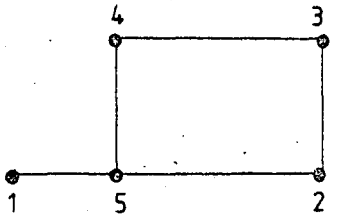
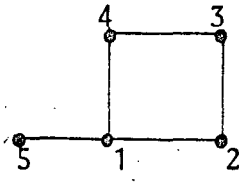
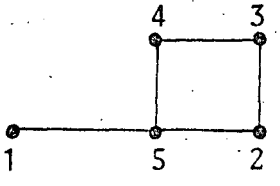


Figura 2.2. SEP elemental. Nodos reordenados de acuerdo al primer esquema.

Para ver el efecto del reordenamiento,se puede observar la eliminación de nodos en las gráficas de flujo correspondientes a la red ordenada y sin ordenar:

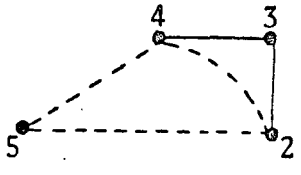


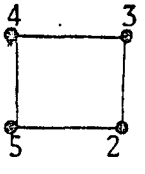
Gráfica de la red original.



Gráfica de la red ordenada.

Eliminando el primer nodo en ambas gráficas se obtienen las siguientes gráficas reducidas:

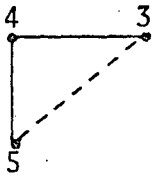


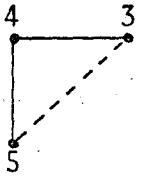


En la red sin ordenar al eliminar el primer nodo,se introducen tres nuevas ramas,

que representarían seis nuevos elementos en la matriz de incidencia reducida de la red. En la gráfica ordenada,no se introducen elementos.

Eliminando el segundo nodo de la gráfica reducida resulta:





Al eliminar este nodo en ambos casos se introduce una rama en la gráfica de flujo reducida resultante. Finalmente la eliminación de cualquiera de los nodos restantes en las gráficas reducidas no introducen nuevos elementos en ninguno de los dos casos.

El aspecto de las matrices de incidencia en uno y otro caso al final del proceso de eliminación sería:

1	1	0	1	1
1	1	1	Ø	Ø
0	1	1	1	Ø
1	Ø	1	1	Ø
1	Ø	Ø	Ø	1

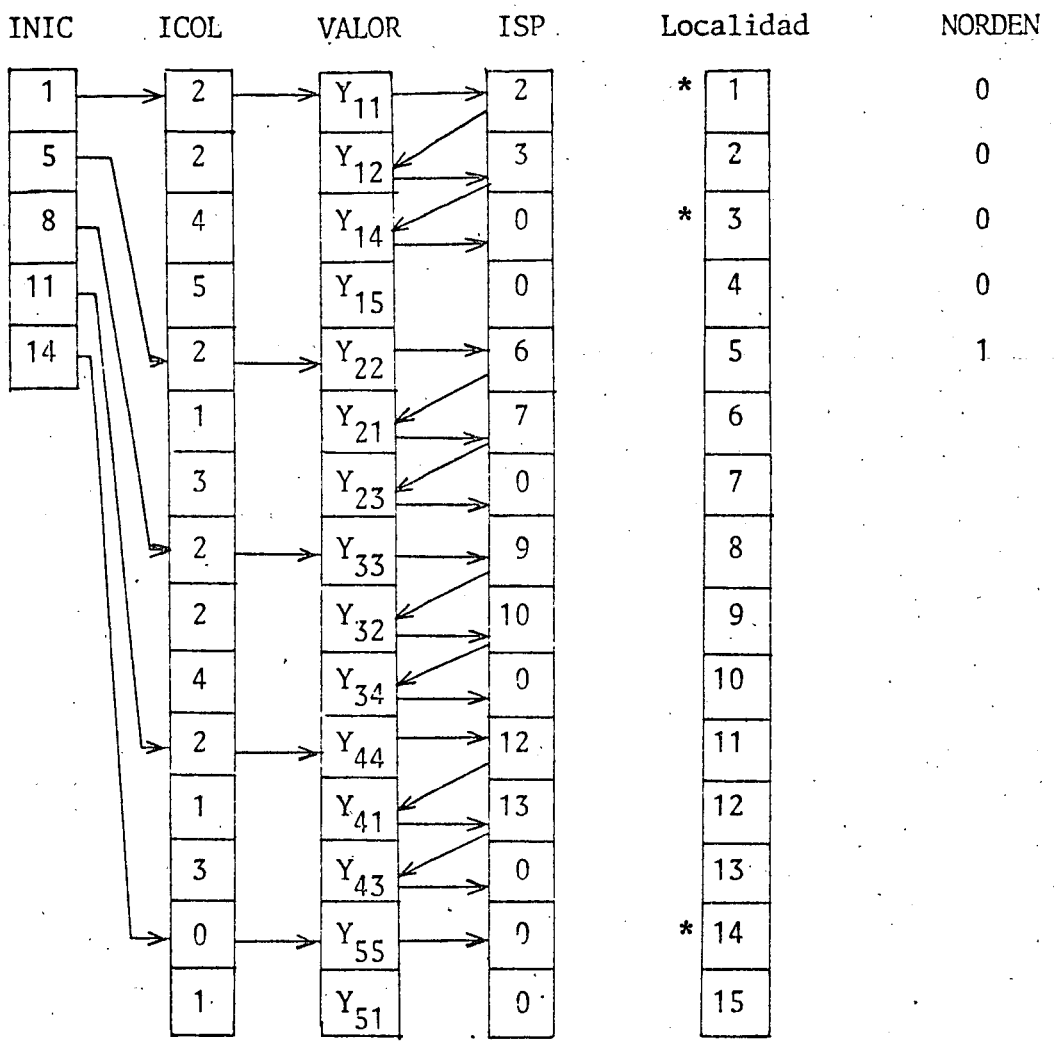
1	0	0	0	1
0	1	1	0	1
0	1	1	1	Ø
0	0	1	1	1
1	1	Ø	1	1

Una vez mas,Ø indica elementos que se introducen en el proceso de eliminación y que de alguna manera necesitan ser almacenados y procesados en algún momento.

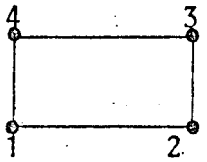
En el esquema anterior se observa que los arreglos en que se almacena la matriz de admitancias de nodo de la red no se modifica en ningun momento. Si se aplica el esquema 2 ó el esquema 3,se tendría que simular el proceso de eliminación de nodos de la gráfica de flujo para ir determinando el siguiente nodo a eliminar en el proceso. A continuación se ilustra la aplicación del segundo esquema a la red del caso anterior.

En este caso,de la información almacenada en ICOL se seleccionaría inicialmente como nodo a eliminar,el nodo que tuviera el menor número de elementos conectados a él. A continuación se simula la eliminación de este nodo,desconectando las ramas correspondientes de acuerdo al procedimiento descrito en la sección 1.3.2. En ICOL se debe de checar la conectividad del nodo que se elimina así como la de los nodos a él conectados ,a fin de determinar las ramas que se introducen. Al eliminar el primer nodo que sería el 5 de la

figura 2.1.,los arreglos se modificarían como se muestra:



Los (*) indican posiciones en las que ha habido modificaciones en ICOL e ISP para simular la eliminación del nodo 5. La gráfica de flujo resultante sería:

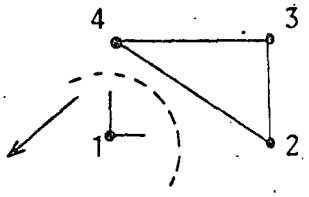


Y no se ha creado ningún elemento al eliminar el nodo. Para determinar el siguiente nodo a eliminar,se inspecciona el número de elementos que inciden a los nodos de la gráfica reducida ;esta información se encuentra en ICOL(INIC(I)).

Todos los nodos tienen ahora dos ramas conectadas y se selecciona el 1 como siguiente a eliminar.

Se procede a simular la eliminación del nodo 1. Como los nodos 2 y 4 están conectados al nodo 1 pero no están interconectados entre sí,se crea una rama entre ellos al elimi-

nar el nodo 1 ,como se muestra en la gráfica reducida:

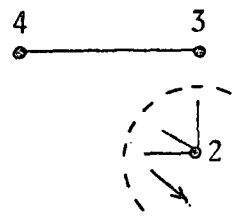


Y los elementos de los arreglos que contienen la información de la grafica de flujo reducida quedaría n como se muestra:

	INIC	ICOL	ISP	Localidad	NORDEN
	1	0	0	1	2
	5	2	3	2	0
	8	4	0	3	0
	11	1	0	4	0
	14	2	7	5	1
		1	7	6	
		3	16	7	
		2	9	8	
		2	10	9	
		4	0	10	
		2	13	11	
		1	13	12	
		3	17	13	
		0	0	14	
		1	0	15	
*		4	0	16	
*		2	0	17	

* Elementos creados al eliminar el nodo 1.

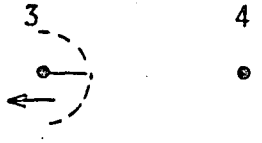
Una vez más se inspecciona el número de ramas que inciden a los nodos aún no eliminados en ICOL(INIC(I)). Todos los nodos sin ordenar tienen el mismo número de conexiones(2) seleccionandose el 2 como siguiente nodo a eliminar. La gráfica de flujo resultante sería:



Quedando los arreglos como:

INIC	ICOL	ISP	Localidad	NORDEN
1	0	0	1	2
5	2	3	2	3
8	4	0	3	0
11	1	0	4	0
14	0	0	5	1
	1	0	6	
	3	7	7	
	1	10	8	
	2	10	9	
	4	0	10	
	1	13	11	
	1	13	12	
	3	0	13	
	0	0	14	
	1	0	15	
	4	0	16	
	2	0	17	

Nuevamente se inspecciona el número de elementos conectados a los nodos aún no eliminados,resultando que los nodos 3 y 4 tienen el mismo número de elementos (1) conectados. Se selecciona como siguiente nodo a eliminar el 3 ,resultando una gráfica como la siguiente:

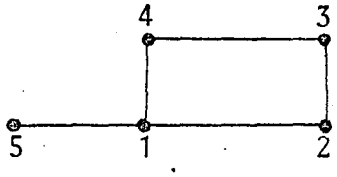


Finalmente el nuevo orden de eliminación proporcionado por el segundo esquema de ordenamiento sería:

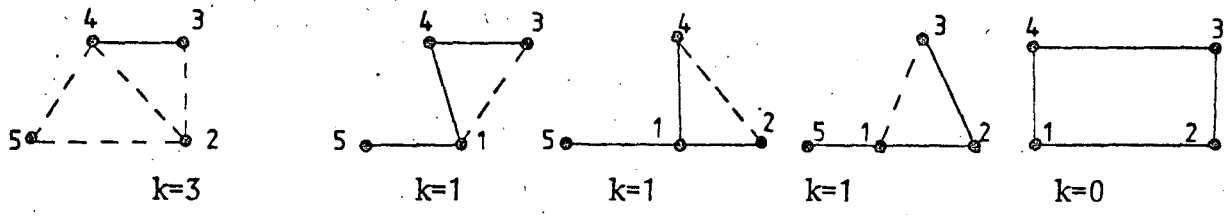
NORDEN

2
3
4
5
1

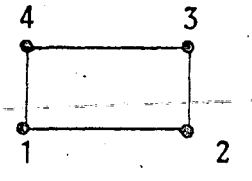
Si se maneja análogamente la información de la estructura de la matriz de admitancias de nodo almacenada en INIC,ICOL e ISP,se puede aplicar el tercer esquema de ordenamiento. La diferencia respecto al segundo esquema sería que habría que eliminar en cada paso del proceso de eliminación,los nodos aún no ordenados y observar cual de ellos introduce el menor número de ramas al ser eliminado. Analizando la gráfica de flujo del ejemplo:



La eliminación de cada uno de los nodos produciría cada una de las siguientes gráficas reducidas,respectivamente:



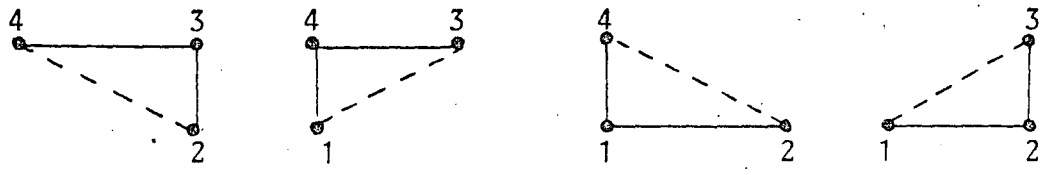
k es un contador de ramas introducidas al eliminar cada uno de los nodos. Se observa que al eliminar el nodo 5 no se crean elementos y por lo tanto se selecciona como primer nodo a eliminar , resultando:



NORDEN

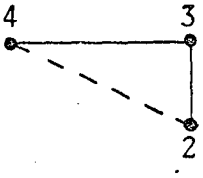
0
0
0
0
1

Eliminando cada uno de los nodos que restan por eliminar se obtendrían las gráficas de flujo reducidas siguientes:



k=1 k=1 k=1 k=1

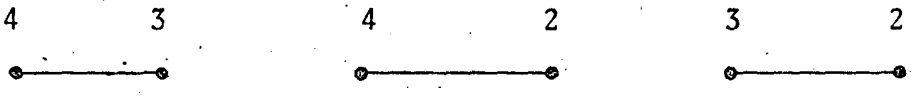
Al eliminar cualquiera de los nodos se introduce una rama. Se selecciona el nodo 1 como siguiente a eliminar,resultando:



NORDEN

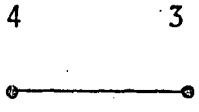
2
0
0
0
1

La eliminación de cada uno de los nodos que restan,producirían las gráficas de flujo siguientes,respectivamente:



k=0 k=0 k=0

Seleccionando como siguiente nodo a eliminar el 2 resultando:



NORDEN

2
3
0
0
1

Finalmente al eliminar cualquiera de los dos nodos restantes no introduciría nuevos elemento,de modo que se pueden ordenar arbitrariamente resultando:

NORDEN

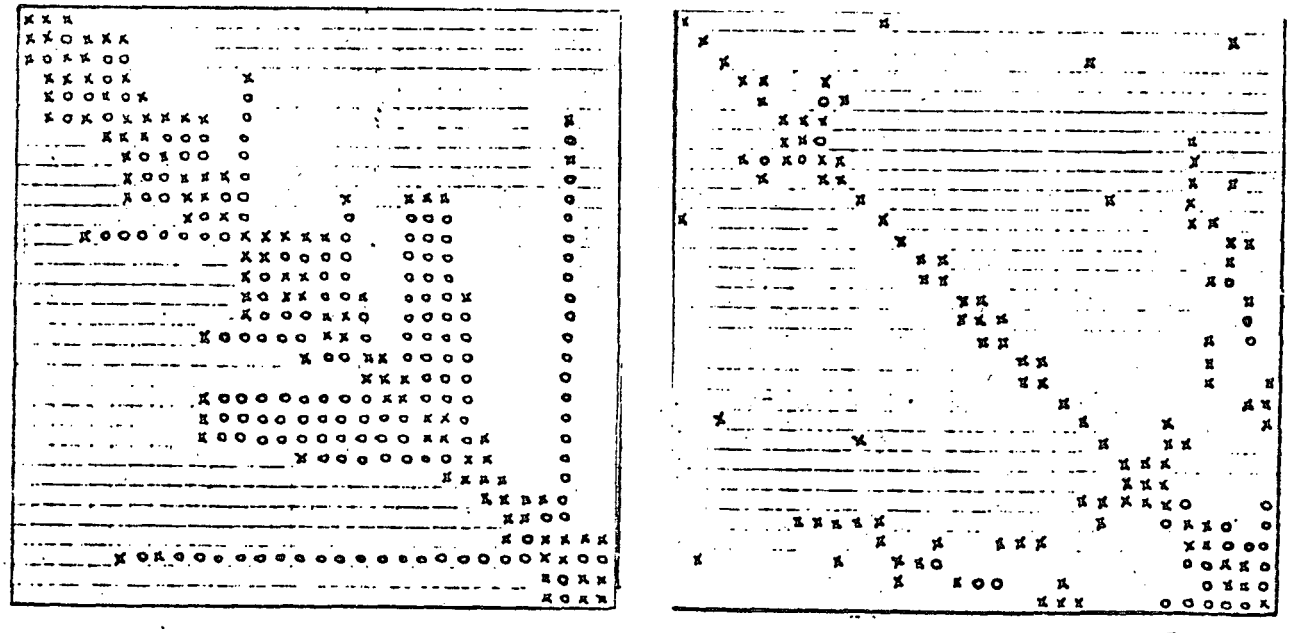
2
3
4
5
1

Aunque en este ejemplo los esquemas 2 y 3 han obtenido iguales resultados,es conveniente

te notar lo siguiente: a) En el ejemplo del segundo esquema, se muestra que al aplicar computacionalmente un esquema dinámico ó estático de ordenamiento, implica manejar únicamente la información de la estructura de la matriz de coeficientes del sistema que se resuelve, información que se empaqueta en vectores enteros, y que se puede simular el proceso de eliminación en forma sencilla manipulando dicha información. b) Existe un grado creciente de complejidad entre esquema y esquema. La diferencia existente en dicha complejidad entre el segundo y tercer esquema y la diferencia en resultados obtenidos han favorecido la mas amplia utilización del esquema 2. c) El concepto de gráficas de flujo es una herramienta para evaluar visualmente el comportamiento de diferentes esquemas de ordenamiento y son la base para simular eliminaciones nodales o factorizaciones matriciales.

2.4.1. Resultados de la aplicación de esquemas de ordenamiento.

El efecto dramático de aplicar métodos de ordenamiento en la dispersidad de una matriz al ser factorizada, se muestra en la siguiente figura para la estructura de la matriz de admitancias de nodo del sistema de prueba de 30 nodos del IEEE , aplicando el esquema 2 de ordenamiento al triangularizar:



(a) (b)

Matriz de incidencia del sistema de prueba del IEEE(30 nodos). a)Matriz triangularizada sin ordenar los nodos de la red. b)Matriz triangularizada después de aplicar el segundo esquema de ordenamiento.

En las matrices anteriores las X son los elementos originales de las matrices,mientras que los 0 son elementos introducidos al triangularizar.

En la siguiente tabla se muestra el resultado de aplicar el segundo esquema de ordenamiento a diferentes SEP de prueba; el coeficiente de dispersidad indicado es la relación entre elementos cero de cada matriz de admitancias de nodo y el número de elementos totales de las mismas.

Sistema	Matriz original Coeficiente de Dispersidad	Elementos creados al Triangularizar Ordenando	Sin ordenar	Matriz triangularizada Coeficiente de dispersidad Ordenando	Sin ordenar
IEEE. 14 nodos 30 ramas	0.7248	8	44	.68367	.5
IEEE 30 nodos 41 ramas	0.87555	28	174	.8444	.68222
IEEE 57 nodos 78 ramas	0.93444	118	790	.89812	.69128
CFE SIN. 98 nodos 111 ramas	0.96668	33	198	.9598	.85089
IEEE 118 nodos 179 ramas	0.96581	172	1692	.95346	.84429

En la tabla anterior se puede observar la forma tan acelerada en que crece el número de elementos diferentes de cero si el orden de eliminación de variables se hace al azar así como la forma en que la dispersidad de la matriz de coeficientes original se mantiene mediante la aplicación del esquema de ordenamiento.

Obviamente la no aplicación de un esquema de ordenamiento tendría como consecuencia un aumento sustancial de la capacidad de memoria requerida para procesar una matriz de coeficientes ,así como del tiempo de ejecución empleado en procesar todos los elementos que se introducen durante el proceso de solución del sistema de ecuaciones. Se observa también en la tabla que resulta más aplicable e imprecindible un método de ordenamiento a medida que el sistema a analizar crece en dimensiones.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA IV CONCEPTOS DE PROBABILIDAD Y ESTADISTICA

DR. SERGIO A. MOLINA GARCIA

ENERO, 1979

NOTAS SOBRE PROBABILIDAD Y ESTADISTICA.

Describiremos la teoría de la probabilidad en términos intuitivos.

Iniciamos nuestro trabajo con la noción de que los juicios probabilísticos contienen ciertas propiedades objetivas del fenómeno involucrado de acuerdo con nuestro conocimiento empírico. El problema probabilístico lo enfocaremos de manera que cuantitativamente podamos estimar la probabilidad de la ocurrencia de algunos fenómenos que es razonable que se presenten bajo ciertas condiciones completamente definidas. Esto significa que no podemos establecer ninguna probabilidad de un evento aislado. Así, por ejemplo, una medición de energía eléctrica.

Ahora, si consideramos a priori que existe una cierta probabilidad de que el error en el valor que asignemos esté expresado como una función

$$p(x) = af(x) ,$$

necesitaremos calcular a, la cual caracteriza el valor de la magnitud del error. Entonces, podemos preguntar ¿en qué medida depende la magnitud del error de otras condiciones, como son: la magnitud de la cantidad medida, condiciones del equipo de medición, etc.?

En este punto podemos continuar con la suposición de que a cada conjunto de condiciones externas le corresponde algún valor definido de a. En todas las aplicaciones de la teoría de la probabilidad, "probabilidad" significa la probabilidad de que algún evento, llamémosle A, ocurrirá bajo la realización de un cierto conjunto de condiciones S, las cuales en principio son reproducibles un número infinito de veces.

Denominamos experimento a cualquier evento cuyo resultado no puede ser predicho con certeza.

Llamaremos espacio de la muestra S, de un experimento, al conjunto de todos los resultados posibles del experimento.

Un evento es un subconjunto del espacio muestra. Cada subconjunto es un evento.

Decimos que un evento ocurre si cualquiera de sus elementos es el resultado del experimento.

Aquellos puntos que son de interés determinarán el espacio muestra a definir. Al mismo tiempo, alguna indicación del número de veces que puede ocurrir cierto resultado puede derivarse de consideraciones analíticas del experimento realizado y los mecanismos inherentes.

Un evento de un solo elemento A, es un subconjunto de S, el cual tiene solamente un elemento que pertenece a S, ésto es, si ahí existe solamente una $x \in S$ tal que $x \in A \subset S$, entonces A es un evento con un solo elemento.

Un evento cuya ocurrencia es inevitable siempre y cuando el conjunto de condiciones S esá presente, se le llama cierto . Si el evento A no puede ocurrir cuando el conjunto de condiciones S está presente, se le llama imposible.

Un evento que puede no ocurrir cuando el conjunto de condiciones S está presente, se le llama aleatorio.

Principio No. 1.- La probabilidad de que el evento A ocurra, siempre y cuando el conjunto de condiciones S esté presente, es igual a p. Las leyes de este tipo son denominadas leyes probabilísticas o estocásticas.

Por ejemplo: la probabilidad de la pérdida de una línea de transmisión en un sistema eléctrico de potencia durante un cierto intervalo de tiempo. El conjunto de condiciones está formado por el nivel de carga del sistema, de frecuencia, voltaje, atmosféricas, etc. El evento A es en sí el hecho de la pérdida de la línea durante el período de tiempo considerado.

Es posible encontrar muchos fenómenos en la vida real, en cualquier campo del conocimiento,de tal modo que podemos hablar de algún juicio "bien definido" de un evento, el cual tiene una mayor o menor probabilidad de ocurrencia. Esto, históricamente llevó a la construcción axiomática de la teoría de la probabilidad.

Definamos U como un conjunto de eventos elementales. Consideremos una familia F de subconjuntos del conjunto U. Llamemos a los elementos de F eventos aleatorios.

Las siguientes 3 condiciones deberán ser satisfechas:

1- F contiene el conjunto U como uno de sus elementos.

2- Si los subconjuntos A y B del conjunto U son elementos de F, entonces los conjuntos A+B, AB, A y B son también elementos de F. Si F es cerrado bajo estas tres operaciones elementales de los eventos, designamos esta familia de eventos F un campo de eventos. El conjunto F formado de esta manera es denominado en la literatura, un campo de eventos de Borel[1] .

3- Si los subconjuntos A_1, A_2, ..., A_n, ... del conjunto U son elementos del conjunto F, entonces las sumas $A_1 + \ldots + A_n + \ldots$ de estos subconjuntos son también elementos de F.

Axiomas.

I. Para cada evento aleatorio A en un campo de eventos F, hay un número no negativo p(A) asociado a él, denominado su probabilidad.

II. $$p(U) = 1$$

III. Si los eventos A_1, A_2, ..., A_n,... son mutuamente excluyentes uno al otro*, entonces

$$p(A_1+A_2+\ldots+A_n) = p(A_1) + p(A_2) + \ldots + p(A_n) \ldots$$

$$= \sum_{i=1}^{n} p(A_i)$$

Propiedades.

i) $P(V) = 0$; $P(V) + P(U) = 1$, donde $V=\emptyset$, el conjunto vacío,

ii) para cualquier evento A, $P(\overline{A})=1-P(A)$

iii) $0 \leq P(A) \leq 1$

iv) si $A \Rightarrow B$, entonces
$P(A) \leq P(B)$

v) $P(A+B) = P(A)+P(B)-P(AB)$
del axioma I, $P(A+B) \leq P(A)+P(B)$

* Si dos eventos aleatorios A y B son tales que los elementos de U pertenecientes a A son distintos de los elementos de B, entonces los denominamos mutuamente excluyentes Definimos U y $\overline{U}=V$, como los eventos cierto e imposible.

Extensión del axioma de la suma (III).

Si el evento A es equivalente a la ocurrencia de por lo menos uno de los eventos mutuamente excluyentes - $A_1, \ldots, A_n, \ldots$ entonces

$$P(A) = P(A_1)+P(A_2)+ \ldots+P(A_n)+\ldots$$

podemos reemplazarlo por el axioma de continuidad. - Si la secuencia de eventos $B_1, B_2, \ldots,$ es el evento imposible, entonces $P(B_n) \to 0$ cuando $n \to \infty$. La demostración de lo anterior puede encontrarse en la referencia 2.

Si la única restricción impuesta al calcular P(A) - es el conjunto de condiciones S, entonces esta probabilidad es llamada incondicional.

En muchos problemas el cálculo de la probabilidad se reduce a contar el número de elementos en S y el número de elementos en los eventos que son de interés. Las probabilidades están entonces dadas por la razón de estas cantidades. Esto es importante cuando estamos construyendo modelos matemáticos para describir la probabilidad de algún evento.

La probabilidad condicional de que ocurra B, dado - que A ha ocurrido, P(B/A), está dada por la expresión siguiente:

$$P(B/A) = P(B \cap A)/P(A) \text{ , donde } P(A) > 0$$

si $P(A) = 0$, definimos

$$P(B/A) = 0$$

Decimos que A y B son mutuamente excluyentes si y - solo si $A \cap B = \emptyset$.

Decimos que 2 eventos, A y B son <u>independientes</u> si y solo si

$$P(A \cap B) = P(A)P(B)$$

Nota.- Si A y B son independientes, entonces

$$P(A/B) = P(AB)/P(B)$$

$$= P(A)P(B)/P(B) = P(A)$$

Un experimento está formado de n eventos independientes, si y solo si:

1) S es el producto cartesiano* de n conjuntos $S_1, S_2, \ldots, S_n$ y
2) la probabilidad de la ocurrencia de un evento de un solo elemento A S, es el producto de las probabilidades de la ocurrencia de los eventos apropiados de un solo elemento $A_i \ S_i$, i=1,2,...n, ésto es, $P(A) = P_1(A_1)P_2(A_2) \ldots P_n(A_n)$, donde A S, $A_i \ S_i$, i=1,2,..,n y $A, A_1, \ldots, A_n$, son cada uno eventos con un solo elemento.

Supongamos n exhaustivas, mutuamente excluyentes e igualmente posibles ocurrencias $A_1, A_2, \ldots, A_n$, de las cuales m son favorables al evento A, k al evento B, y r al evento AB ($r \leq k$ y $r \leq m$)

$$P(A/B) = r/k = (r/n)/(k/n)$$

$$= P(AB)/P(B)$$

* El producto cartesiano AxB de A y B es el conjunto de todos los pares posibles (x_1, x_2) donde $x_1 \in A$ y $x_2 \in B$, o sea,

$$AxB = \{(x_1, x_2) : x_1 \in A, x_2 \in B\}.$$

De la misma manera podemos escribir

$$P(B/A) = P(AB)/P(A).$$

Las últimas 2 expresiones nos llevan al denominado - Teorema de Multiplicación:

$$P(AB) = P(A)P(B/A) = P(B)P(A/B)$$

el cual podemos expresar en palabras como sigue: la probabilidad del producto de 2 elementos es igual al producto de la probabilidad de uno de ellos por la - probabilidad condicional del otro, dado que el primero ha ocurrido.

Un evento A es estocásticamente independiente —o simplemente independiente— de un evento B si la relación

$$P(A/B) = P(A)$$

es válida.
De ello se deriva

$$P(A)P(B/A) = P(A)P(B)$$

$$P(B/A) = P(B)$$

Así, la ocurrencia del evento B no afecta la probabilidad del evento A. La independencia es una relación simétrica. Si los eventos A y B son independientes,

$$P(AB) = P(A)P(B)$$

Un espacio muestra discreto es aquel que tiene un número finito o un número contable infinito* de elementos.

* Un conjunto S contiene una infinidad contable de - elementos si es posible establecer una correspondencia uno a uno entre los elementos de S y el conjunto de los enteros positivos.

Un espacio muestra continuo (unidimensional) es aquel que tiene como elementos todos los puntos en algún intervalo definido sobre la línea de los reales.

Se puede demostrar que para espacios muestra continuos, debemos especificar una regla para asignar las probabilidades a subintervalos continuos más que especificar probabilidades para todos los eventos de un solo elemento. Es sencillo demostrar que la probabilidad de algún punto particular en un espacio contínuo es cero.

Podemos asignar probabilidades para espacios muestra contínuos unidimensionales, dada la longitud de S, $L(S)$, $S = \{ x: a \leq x \leq b \}$ y dado algún subintervalo $A = \{ x: c \leq x \leq d \}$, siempre y cuando $d > c$, $d \leq b$ y $a \leq c$, entonces

$$P(A) = L(A)/L(S) \quad ; \quad A \subset S$$

El enunciado anterior satisface los axiomas de la teoría de probabilidad.

Cuando llevamos a cabo una medición, por ejemplo en un sistema de potencia, tenemos un gran número de efectos de tal manera que podemos decir que el valor de la cantidad medida variará en forma aleatoria alrededor de cierto valor. Si repetimos el experimento en las mismas condiciones (siempre y cuando ésto sea posible), aún en este caso obtendríamos valores diferentes cada vez. Esta variabilidad de los resultados es descrita adecuadamente por la teoría de la probabilidad, por variables aleatorias. Podemos decir que todo el proceso de medición es un proceso aleatorio, puesto que la variabilidad de los valores medidos depende de tantos factores que no podríamos definirla de una manera determinística.. Así, podemos decir que una va

riable cuyo valor está determinado por el resultado de un experimento del tipo antes mencionado es denominada variable aleatoria.

Una variable aleatoria x es una función con valores reales de los elementos de un espacio muestra S.

Una variable aleatoria x es <u>discreta</u> si su rango forma un conjunto discreto (contable) de números reales. Una variable aleatoria x es continua si su rango forma un conjunto continuo (no contable) de números reales, y la probabilidad de que x sea igual a un valor específico en su rango es cero. La función de probabilidad para X es una función, representada por P(x), de una variable real x y será definida por

$$P(x) = P(X(w)=x) \quad \text{para toda x real; donde } X(w) \text{ es una descripción funcional de X.}$$

La función de probabilidad puede ser derivada intuitivamente de la medida de la probabilidad de los subconjuntos del espacio muestra y puede ser usada para producir juicios acerca de la variable aleatoria.

La función de distribución para una variable aleatoria X, expresada por $F_X(t) = P(X \leq t)$, es una función de una variable real t, tal que: 1) El dominio de definición de F_X es la línea recta, y 2) para cualquier recta real t:

$$F_X(t) = P(X,t)$$

Para una variable aleatoria discreta X, podemos escribir

$$F_X(t) = \sum_{x \leq t} p(x)$$

La función de distribución probabilística para una variable aleatoria continua Y, definida por $f_Y(y)$ es una función de una variable real y tal que: 1) el dominio de f_Y es la línea recta y 2) para cualquier número real t

$$F_Y(t) = \int_{-\infty}^{t} f_Y(y)\, dy$$

Se puede pensar que $f_Y(t)$ está midiendo la razón de acumulación de la probabilidad en cualquier punto t.

Podemos calcular la probabilidad de que Y se encuentre en el intervalo comprendido entre a y b con la expresión siguiente :

$$P(a< Y< b) = P(a\leq Y\leq b) = \int_a^b f_Y(y)\, dy$$

Antes de continuar con las propiedades numéricas de las variables aleatorias, vamos a presentar en forma breve, una fórmula importante para la teoría de estimación, la fórmula de Bayes.

Fórmula de Bayes. La probabilidad de A_i dado que B ha ocurrido

$$P\ (A_i/B) = P\ (A_i)\ P\ (B/A_i)\ /\ P\ (B),$$

donde $P\ (B) = \sum_{j=1}^{n} P\ (A_j)\ P\ (B/A_j)$

La mayor parte de las veces, cuando hablamos de un evento, hay algunos resultados con una mayor probabilidad de ocurrir y especialmente hay uno más esperado, en términos numéricos hay un punto, donde el valor promedio puede ser encontrado. A este valor se le denomina, el valor esperado. Las dos características más importantes de una variable aleatoria son su valor esperado y su variancia o dispersión.

Valor esperado o esperanza matemática.

Si X es una variable aleatoria discreta con función de probabilidad $p_X(x)$, el valor esperado de H (x), expresado por $E\{H\ (X)\}$ está definido por

$$E\{H\ (X)\} = \sum_{\substack{\text{rango} \\ \text{de } x}} H\ (x)\ p_X\ (x)$$

Siempre y cuando la suma sea absolutamente convergente.

Si X es una variable aleatoria continua con función de probabilidad $f_X(x)$, el valor esperado de H (X) está definido por

$$E\{H(X)\} = \int_{-\infty}^{\infty} H(x)\, f_X(x)\, dx$$

Siempre y cuando la integral sea absolutamente convergente. Si la integral o la suma no son absolutamente convergentes, decimos simplemente que el valor esperado no existe.

El valor esperado de X, le denominamos la media o el valor promedio de X y lo representamos por μ_X esto es, $\mu_X = E\{X\}$.

El valor medio de la variable aleatoria localiza el centro de la función de probabilidad en el sentido del centro de gravedad.

La media de una variable aleatoria proporciona una medida del centro de la distribución probabilística de la variable aleatoria.

La medida de la variabilidad de una variable aleatoria frecuentemente utilizada es el promedio del cuadrado de la distancia entre la variable aleatoria y su media μ_X. Si X varía relativamente poco, la mayor parte de las veces estará cercano a su valor medio y su distancia promedio de la medial al cuadrado,- será relativamente pequeña; si X varía bastante, entonces una buena parte de las veces diferirá por un valor relativamente grande y su distancia promedio de la media al cuadrado será relativamente grande. A esta desviación de la media al cuadrado se le llama dispersión o variancia de la variable aleatoria.

La variancia o dispersión de una variable aleatoria X, representada por σ_X^2 está definida por $\sigma_X^2 = E\left\{(x-\mu_X)^2\right\}$;

Su raíz cuadrada positiva le llamamos la desviación estándar de X. Así, $\sigma_X = \sqrt{\sigma_X^2}$

Si hay unidades asociadas a la variable aleatoria X, la dispersión será medida con el cuadrado de esas unidades. La desviación estándar reviste un especial interés puesto que está medida en las mismas unidades que las de la variable aleatoria.

La dispersión puede ser expresada:

$$\sigma_X^2 = E\left\{(X-\mu_X)^2\right\} = E\left\{X^2\right\} - \mu_X^2$$

Es fácil ver, que σ_X^2 puede ser calculada como el promedio de los cuadrados de variable X menos el cuadrado del valor promedio de X.

El momento k-ésimo de una variable aleatoria X es expresado por:

$$m_k = E\left\{X^k\right\} .$$

La media o valor esperado y la dispersión o variancia de una variable aleatoria X son funciones de los primeros dos momentos de X. El primer momento, m, esta expresado por:

$$m_1 = E\left\{X^1\right\} = E\left\{X\right\}, \text{ ésto es, } m_1 = \mu_X$$

El segundo momento m_2 está expresado por:

$$m_2 = E\left\{X^2\right\}, \text{ entonces } m_2 = \sigma_X^2 + \mu_X^2 .$$

El conocimiento de m_1 y m_2 de una variable aleatoria nos permitirá calcular la media y la variancia de la variable aleatoria.

Estos dos momentos proporcionan información acerca de la parte media de la ley de probabilidad y el grado de variación alrededor del valor promedio.

La función generadora de momentos, o función característica, expresada por $m_X(t)$ de una variable aleatoria X está definido por:

$$m_X(t) = E\left\{e^{tx}\right\}$$

Para ver por qué a $m_X(t)$ se le denomina una función generadora de momentos, veamos sus derivados con respecto a t, calculadas

para t=0. Como se puede ver, por ejemplo:

$$m_X^{(1)}(0) = E\left\{Xe^{\circ}\right\} = E\left\{X\right\} = m_1$$

La expresión $E\left\{(X - \mu_X)^k\right\} = \mu_k$ se denomina el momento k-ésimo con respecto a la media de X, donde k=1,2,...

Para: $k=1, \quad \mu_1=0$;

$k=2, \quad \mu_2= \sigma_X^2$

Teorema:

Para dos variables aleatorias X y Y con funciones generadoras de momentos $m_X(t)$ y $m_Y(t)$, respectivamente

Entonces, $m_X(t) = m_Y(t)$ si y solo si $F_X(t) = F_Y(t)$ para cualquier real t.

Teorema:

Para una variable aleatoria X con función de distribución $F_X(t)$, definimos $Y = a + bX$, donde $b > 0$. Entonces,

$$F_Y(t) = F_X\left(\frac{t-a}{b}\right)$$

$$F_Y(a+bt) = F_X(t), \text{ para toda t.}$$

La forma estándar de una variable aleatoria X es la variable aleatoria $Z = (X - \mu_X)/\sigma_X$.

Es fácil ver que: $\mu_Z = -\mu_X/\sigma_X + \mu_X(1/\sigma_X) = 0$ y

$$\sigma_Z^2 = (1/\sigma_X^2)\,\sigma_X ;$$

así, la forma estándar de una variable aleatoria siempre tendrá una valor medio igual a cero y su variancia igual a 1.

Teorema:

Supongamos que X es una variable aleatoria continua con función de distribución $F_X(t)$. Si $Y = aX^2$, $a > 0$, entonces la función de distribución de Y es

$$F_Y(t) = 0, \qquad t < 0$$
$$= F_X\left(\sqrt{\frac{t}{a}}\right) - F_X\left(-\sqrt{\frac{t}{a}}\right), \quad t \geq 0$$

y la función de probabilidad de Y es:

$$F_Y(t) = \frac{1}{2\sqrt{at}}\left[f_X\left(\sqrt{\frac{t}{a}}\right) + f_X\left(-\sqrt{\frac{t}{a}}\right)\right], \quad t > 0$$
$$= 0 \quad \text{para cualquier otro valor de t.}$$

Teorema:

Consideremos una variable aleatoria continua X, y F(o) = 0. Si y = X, entonces:

$$F_Y(t) = 0 \qquad \text{para} \quad t < 0$$
$$= F_X(t^2) \qquad \text{para} \quad t \geq 0$$

y

$$F_Y(t) = 0 \qquad \text{para} \quad t < 0$$
$$= 2t\, f_X(t) \qquad \text{para} \quad t \geq 0$$

A continuación presentamos algunos teoremas que expresan - propiedades del valor esperado y la variancia de una variable aleatoria.

Las demostraciones pueden ser encontradas en las referencias.

Teorema:

El valor esperado de una constante es la misma constante:

$$E\{c\} = \int_{-\infty}^{\infty} c f_c(x)\, dx = c \cdot 1 = c$$

Teorema:

El valor esperado de la suma de dos variables aleatorias es igual a la suma de los valores esperados de cada una.

$$E\{\alpha + \beta\} = E\{\alpha\} + E\{\beta\}$$

Corolario: El valor esperado de la suma de un número finito de variables aleatorias es igual a la suma de valores espera-

dos de cada una de ellas.

$$E\left\{\alpha_1 + \alpha_2 + \ldots + \alpha_n\right\} = E\left\{\alpha_1\right\} + \ldots + E\left\{\alpha_n\right\}$$

Teorema:

El valor esperado del producto de dos variables aleatorias independientes α y β es igual al producto de los valores esperados de cada una.

$$E\left\{\alpha\beta\right\} = E\left\{\alpha\right\} \; E\left\{\beta\right\}$$

Corolario 1: El valor esperado del producto de un número finito de variables aleatorias independientes (todas entre sí) es igual al producto de los valores esperados de cada uno de ellas.

$$E\left\{\alpha_1 \alpha_2 \alpha_3 \ldots \alpha_n\right\} = E\left\{\alpha_1\right\} E\left\{\alpha_2\right\} \ldots E\left\{\alpha_n\right\}$$

Corolario 2: Un factor constante puede ser sacado del símbolo de valor esperado:

$$E\left\{c\alpha\right\} = cE\left\{\alpha\right\}$$

Teorema:

La variancia de una constante es cero:

$$E\left\{\left[c - E\left\{c\right\}\right]^2\right\} = 0$$

Teorema:

Si c es una constante, y si $\beta = \alpha c$,

$$E\left\{\left[\beta - E\left\{\beta\right\}\right]^2\right\} = c^2 \, Var(\alpha)$$

$$\sigma_\beta^2 = Var(\beta) = c^2 \; \sigma_\alpha^2$$

Teorema:

La variancia de la suma de dos variables aleatorias independientes α y β es igual a la suma de sus variancias:

$$Var\,(\alpha + \beta) = Var(\alpha) + Var(\beta)$$

$$\sigma_{\alpha+\beta}^2 = \sigma_\alpha^2 + \sigma_\beta^2$$

Corolario: La variancia de la suma de un número finito de variables aleatorias, tales que cada una es independiente del resto, es igual a la suma de las variancias de cada una:

$$\text{Var}(\alpha_1 + \alpha_2 + \ldots + \alpha_n) = \text{Var}\ \alpha_1 + \text{Var}\ \alpha_2 + \ldots + \text{Var}\ \alpha_n$$

Funciones de distribución más comunes:

Una variable aleatoria está completamente caracterizada por su función de distribución. Una función de distribución indica simultáneamente qué valores puede tomar la variable aleatoria y con qué probabilidad. En la práctica, sólo necesitamos una descripción breve de la variable aleatoria de que se trate y ésto puede obtenerse de su valor esperado, su variancia o dispersión y los momentos de varios órdenes.

Llamamos una prueba de Bernoulli a un experimento que solo tiene dos posibles resultados, generalmente denominados éxito y falla.

Sea X el número total de éxitos en n pruebas de Bernoulli independientes con probabilidad de éxito p para cada prueba. A X se le denomina como la variable aleatoria binominal con parámetros n y p, entonces:

$$p_X(x) = \binom{n}{x} p^x q^{n-x}; \quad x = 0,1,2,\ldots,n$$

$$= 0 \quad \text{para cualquier otro valor de } x.$$

El valor esperado y la dispersión, están dados por:

$$\mu_X = n\,p\ ; \quad \sigma_X^2 = npq$$

Consideremos que X es una variable aleatoria binominal con parámetros n = 1 y p. Entonces a X se le llama variable aleatoria de Bernoulli con parámetro p.

La variable aleatoria de Bernoulli es simplemente el número de éxitos que observamos en una sola prueba de Bernoulli y tiene como función de probabilidad:

$$p_X(x) = p \qquad \text{para } x=1$$
$$= q \qquad \text{para } x=0$$
$$= o \qquad \text{para cualquier otro valor.}$$

En muchas ocasiones observamos eventos discretos en un intervalo continuo. Este tipo de eventos está definido como sigue:

En un proceso de Poisson con parámetro λ, eventos discretos son producidos en un intervalo continuo (por ejemplo, tiempo), de la siguiente manera:

1) Consideremos un intervalo suficientemente pequeño, de longitud h, tal que:

   i) la probabilidad de una sola ocurrencia del evento en el intervalo es aproximadamente λh, y

   ii) la probabilidad de 2 o más ocurrencias en el intervalo es aproximadamente 0.

2) La ocurrencia de un evento en un intervalo de longitud h no tiene efecto sobre la ocurrencia o no-ocurrencia en otro intervalo no entrelazado de longitud h. Esto significa que las ocurrencias del evento son estadísticamente independientes una de la otra.

Si observamos un proceso de Poisson durante una unidad de tiempo, el número de eventos que ocurren es una variable aleatoria.

Un proceso de Poisson con parámetro λ es observado durante s unidades de tiempo.
Sea X el número de eventos que ocurren.
Entonces, llamamos a X la variable aleatoria de Poisson con parámetro λs.

Teorema:

Si X es una variable aleatoria de Poisson con parámetro λs,

entonces,

$$p_X(k) = (\lambda s)^k e^{-\lambda s}/k! \ , \ k = 0,1,2,\ldots,\ldots$$
$$= 0 \qquad \text{para cualquier otro valor de k.}$$

El valor esperado y su variancia están dados por:

$$\mu_X = \lambda s \ ; \qquad \sigma_X^2 = \lambda s$$

La variable aleatoria continua más simple es la variable aleatoria uniforme.

X es una variable aleatoria uniforme en el intervalo (a,b) si:

1) El rango de X es el intervalo (a,b).

2) Todos los puntos en el intervalo tienen la misma probabilidad de ocurrir con el valor X (X está distribuida uniformemente en (a,b)).

Teorema:

Si X está distribuida uniformemente en (a,b), entonces

$$f_X(x) = 1/(b-a), \qquad a < x < b$$
$$= 0 \qquad \text{para cualquier otro valor de x.}$$

Su valor esperado y su variancia están dados por:

$$\mu_X = \frac{b+a}{2} \quad \text{y} \quad \sigma_X^2 = \frac{(b-a)^2}{12}$$

La figura 1 ilustra la función de densidad de probabilidad y la función de distribución de la variable aleatoria uniforme.

Dado un proceso de Poisson con parámetro λ, designamos como cero al tiempo en que iniciamos la observación del proceso.
Sea T el tiempo que transcurre hasta que se presenta el primer evento. A T se le denomina la variable aleatoria

exponencial con parámetro λ.

Teorema:

Consideremos a T una variable aleatoria exponencial con parámetro λ. Entonces,

$$F_T(s) = 0 \qquad s < 0$$

$$= 1 - e^{-\lambda s} \qquad s \geq 0, \quad y$$

$$f_T(s) = e^{-\lambda s} \qquad s > 0$$

$$= 0 \qquad \text{para cualquier otro valor de } s.$$

Su valor esperado y variancia están dados por:

$$\mu_T = 1/\lambda \quad ; \qquad \sigma_T^2 = 1/\lambda^2$$

En la figura 2 se muestran la función de densidad y la función de distribución para una variable aleatoria exponencial.

La función de densidad probabilística más común es la distribución normal. La variable aleatoria normal se presenta frecuentemente en problemas prácticos. Además, proporciona una aproximación prácticamente exacta a un número considerable de otras leyes probabilísticas.

Una variable aleatoria X se dice que es normalmente distribuída si y solo si su función de densidad de probabilidad es:

$$f_X(x) = \left(1/(\sigma\sqrt{2\pi})\right)\left[e^{(-(x-\mu)^2/2\sigma^2)}\right] \quad \text{para}$$

toda x real.

El parámetro μ puede ser cualquier número real y el parámetro σ debe ser positivo.

La figura 3 muestra varias gráficas de funciones de densidad para dos valores de μ y tres valores de σ. El pará-

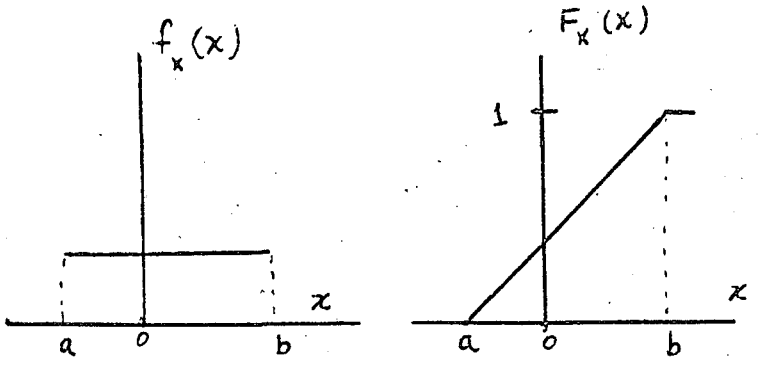


Fig. 1.- Densidad de distribución y Función de distribucion de una variable aleatoria uniforme

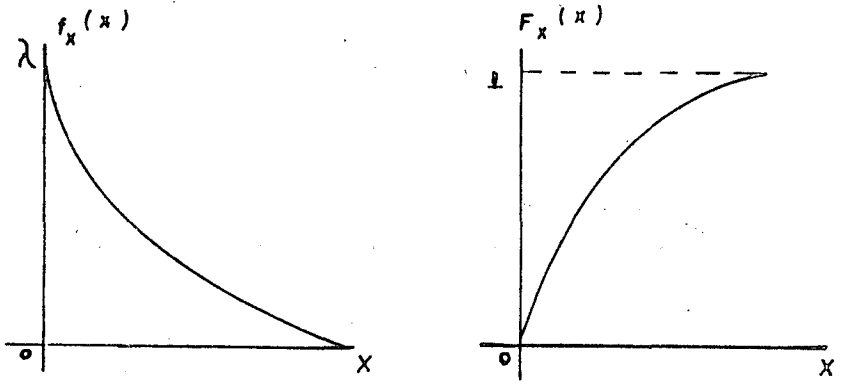


Fig. 2.- Densidad de distribución y Función de distribución de una variable aleatoria exponencial.

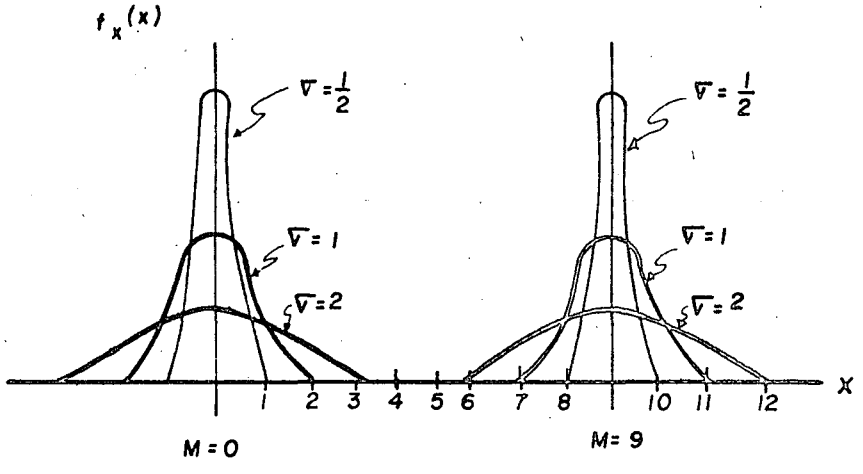


Fig. 3.- Densidades de distribución y Funciones de distribucion de varias variables aleatorias normales

metro μ define la posición de la curva "acampanada" y σ la hace más o menos aplanada. Para el mismo valor de μ y reduciendo σ, la función de densidad se hace más aguda, lo cual significa que aumenta la probabilidad de que cualquier valor de x sea cercano a μ.

Al aumentar σ, se aplana la curva de la función de densidad, produciendo una disminución de la probabilidad de que cualquier valor de x sea cercano a μ.

Teorema:

Si X es una variable aleatoria normal con la función de densidad de probabilidad definida arriba, entonces la función generadora de momentos para X es:

$$m_X(t) = e^{((t\mu + t^2 \sigma^2)/2)}$$

El parámetro μ, en la función de densidad, es el valor esperado de la variable aleatoria X y el parámetro σ es la variancia de X.

La función de distribución para una variable aleatoria normal X es:

$$F_X(t) = \int_{-\infty}^{t} \left(1/(\sigma\sqrt{2\pi})\right) e^{(-(X-\mu)^2/(2\sigma^2))}\, dx$$

Si Z es una variable aleatoria normal con $\mu = 0$, $\sigma^2 = 1$, entonces a Z se le denomina la variable aleatoria normal estándar. Su función de densidad es:

$$n_Z(z) = 1/\sqrt{2\pi}\ e^{-(-(z^2/2))}$$

y su función de distribución es

$$N_Z(t) = \int_{-\infty}^{t} 1/\sqrt{2\pi}\ e^{(-(z^2/2))}\, dz$$

La función de distribución normal estándar puede ser utilizada para calcular la función de distribución para una variable aleatoria normal con valores arbitrarios de μ y Z.

Teorema:

Si X es una variable aleatoria normal con media μ_X, y variancia σ_X^2, entonces,

$$F_X(t) = N_Z\left(\frac{t-\mu}{\sigma}\right)$$

Hasta el momento, hemos estado viendo los conceptos básicos de probabilidad para variables aleatorias unidimensionales. La extensión al caso de variables aleatorias multidimensionales está fuera del propósito de las presentes notas. Sin embargo, lo fundamental está descrito y puede ser aplicado, no en forma sencilla, pero para componentes independientes entre sí de una variable aleatoria multidimensional, una gran parte de las propiedades descritas son válidas.

En el Apéndice B se incluyen algunos conceptos de cálculo matricial, variancia y valor esperado para el caso multivariable.

Antes de proseguir con las distribuciones para muestras pequeñas, haremos una breve descripción simplificada de procesos estocásticos[2].

El estado verdadero de un sistema, para cualquier instante de tiempo, no se conoce exactamente debido a la interacción de una gran cantidad de eventos producidos al azar o en algunos casos que parecen ser al azar, que actúan sobre el sistema.

Podríamos considerar a los procesos aleatorios como la suma de un sistema dinámico determinista y una variable aleatoria cuya función de densidad cambia con el tiempo.

Definimos un proceso estocástico de orden n como la historia en el tiempo de una variable aleatoria, la cual tiene un valor medio, una variancia y todos los otros momentos hasta el orden n, en el intervalo de tiempo de interés.

De una manera formal, definimos un proceso estocástico como sigue:

Tenemos un conjunto de eventos elementales, U, y un parámetro contínuo t.

Ahora, si tenemos alguna relación entre el estado del sistema y el parámetro t, digamos, la función de dos argumentos:

$$\alpha(t) = \varphi(e,t) \; ; \quad e \in U$$

se le denomina un proceso estocástico.

Para cada valor del parámetro t la función $\varphi(e,t)$ es una función de e solamente y, consecuentemente, es una variable aleatoria.

Un proceso estocástico puede ser interpretado como una colección de variables aleatorias $\alpha(t)$ dependiendo del parámetro t.

Podemos describir un proceso estocástico de dos maneras:

a) Considerar el sistema en un instante dado, implicando que en un cierto intervalo de tiempo el sistema no cambia, y

b) considerar que el sistema está cambiando, aún para intervalos de tiempo relativamente pequeños.

Los procesos aleatorios pueden ser clasificados como discretos y continuos.

Los primeros pueden ser representados por un conjunto contable de valores discretos en el tiempo, de una variable

aleatoria.

Los segundos pueden ser descritos por una serie en el tiempo, de secuencia aleatoria continua, de una variable aleatoria.

Se puede elaborar una descripción matemática de la siguiente manera:

Sea $\underline{x}$ un vector aleatorio cuyos componentes son variables aleatorias y están representadas por $(x_1, x_2, \ldots, x_m)$, entonces $\underline{x}(t)$ es un proceso o secuencia de orden n de un vector aleatorio si y sólo si el vector esperado $\mu_x(t)$ y los productos de los momentos

$$E\left\{\left[\underline{x}(t) - \mu_x(t)\right]^i \left[\underline{x}^T(t) - \mu_x^T(t)\right]^j\right\}$$

existen para i y j tales que $i+j=m$.

Para nuestros objetivos sólo estamos interesados en procesos aleatorios de orden 2, o sea, en el vector del valor esperado y matriz de covariancia del proceso estocástico.

Muestra y Población.

En el análisis de un grupo de objetos, individuos, resultados de un experimento, proceso industrial, etc., si el número de elementos que lo integran es muy grande y desde un punto de vista práctico no es posible analizar las propiedades de interés de todo el conjunto, será necesario disponer de un método que nos permita de una manera significativa realizar el estudio deseado. Este tipo de análisis se puede realizar por medio de muestreo.

Llamamos población al conjunto universal de los objetos de interés.

Una muestra de una población es una parte de los elementos de dicho grupo, la cual representa a todo el conjunto.

Muestra aleatoria. Un conjunto de observaciones $x_1, x_2, \ldots, x_n$, forma una muestra aleatoria de tamaño n de una población cuya densidad de probabilidad es $\varphi(x)$, si cada x_i es una variable aleatoria cuya densidad de probabilidad es $\varphi(x)$ y estas n variables aleatorias son independientes.

Distribución de probabilidad de un experimento.

Sea x una variable aleatoria.
Al realizar un experimento observamos n valores de x, de los cuales obtuvimos f_1 iguales a x_1, f_2 iguales a x_2, .. f_m son iguales a x_m.
Es fácil ver que $f_1 + f_2 + \ldots + f_m = n$

Podemos definir las frecuencias relativas:

$$h_1 = \frac{f_1}{n} \; ; \quad h_2 = \frac{f_2}{n} \; ; \quad \ldots \quad h_m = \frac{f_m}{n}$$

Supongamos una variable aleatoria y que toma los valores

$y_1 = x_1$, $y_2 = x_2$, ..., $y_m = x_m$ con probabilidades h_1, h_2, ..., h_m.

Al conjunto de valores h_i $(i=1,...,m)$ se le llama una distribución de probabilidad empírica de la variable aleatoria x.

Tabla de frecuencias. Al arreglo en forma de tabla de los resultados por intervalos de un experimento, junto con sus frecuencias de cada intervalo, se le llama tabla de frecuencias.

En la figura 4 se muestra la gráfica de una tabla de frecuencias. A esta representación se le llama histograma.

Densidad de probabilidad del valor esperado de una muestra.

Consideremos inicialmente que:

$$\hat{x} = \frac{1}{n} \sum_{i}^{n} x_i .$$

donde $\hat{x}$ es el valor medio de la muestra.

Teorema.

Si se obtiene una muestra aleatoria de tamaño n de una población cuyo valor esperado es μ_x y su variancia σ_x^2, entonces $\hat{x}$ es un valor de una variable aleatoria cuya distribución tiene media μ_x. Si la población es infinita la variancia de esta distribución es σ_x^2/n.

El hecho de que la variancia de la variable aleatoria $\hat{x}$ sea igual a σ_x^2/n para muestras aleatorias extraídas de una población infinita es muy importante.

Conviene describir en este punto una de las relaciones más importantes de la teoría de la probabilidad: la desigualdad de Tchebichev.

HISTO

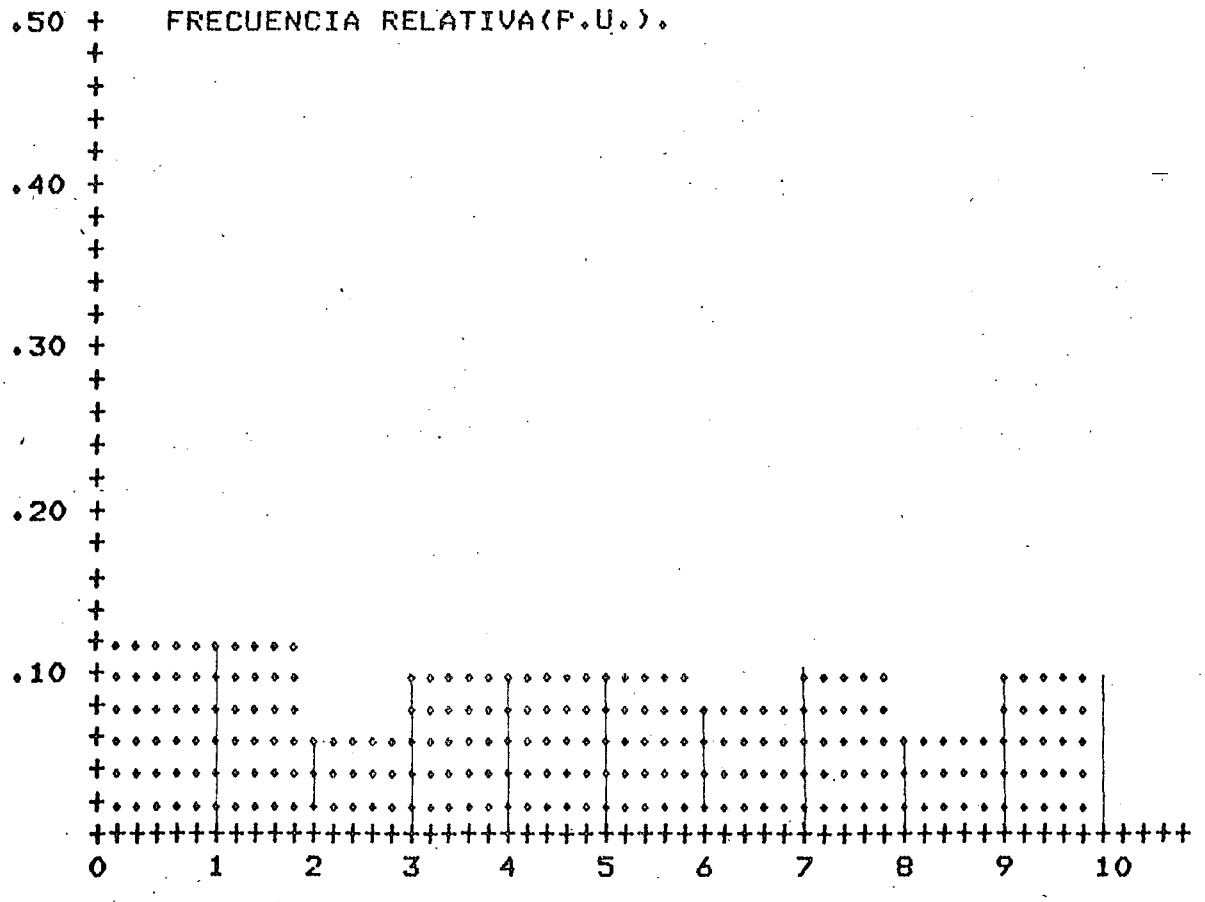
RESIDUAL ACUMULADO CON RESPECTOA 1.-.001

DISTRIBUCION DE FRECUENCIAS

INTERVALO	OCURRENCIA	FRECUENCIA RELATIVA	FRECUENCIA ACUMULADA
0 - 1	57	.116	.116
1 - 2	64	.131	.247
2 - 3	30	.061	.308
3 - 4	51	.104	.412
4 - 5	53	.108	.52
5 - 6	56	.114	.634
6 - 7	40	.082	.716
7 - 8	50	.102	.818
8 - 9	36	.073	.891
9 - 10	53	.108	.999

QUIERES EL HISTOGRAMA ? SI

HISTOGRAMA DE LA MUESTRA? TEST



NOTA.- EL ORIGEN DE LA GRAFICA CORRESPONDE AL VALOR MINIMO (0) Y EL VALOR MAXIMO(.995769) AL INTERVALO MAXIMO.
LA LONGITUD DEL INTERVALO ES: .995769E-1

Fig. 4.- Tabla de frecuencias e histograma de una muestra .

Desigualdad de Tchebychev.

Sea x una variable aleatoria con valor esperado μ_x y σ_x su desviación estándar. La expresión:

$$P\{|x - \mu_x| \geq t\,\sigma_x\} \leq 1/t^2$$

muestra que si t aumenta, la probabilidad de una desviación mayor o igual que $t\,\sigma_x$ disminuye al menos en la relación $1/t^2$. A la expresión anterior se le denomina la desigualdad de Tchebychev.

Ahora, si en la desigualdad de Tchebychev substituimos x por $\hat{x}$ y σ_x por $\sigma_x/\sqrt{n}$ y hacemos $\varepsilon = t\,\sigma_x/\sqrt{n}$, podemos escribir:

$$P\left\{|\hat{x} - \mu_x| \geq \varepsilon\right\} \leq \frac{\sigma_x^2}{n\varepsilon^2}$$

Lo anterior significa que para cualquier $\varepsilon > 0$ por pequeño que sea, la probabilidad de que $\hat{x}$ sea diferente de μ_x en una cantidad mayor o igual que ε puede hacerse tan pequeña como se desee si se escoge n lo suficientemente grande.

Frecuentemente se expresa la confiabilidad de $\hat{x}$ como estimación de μ_x, a través de la expresión $\sigma_x/\sqrt{n}$, lo cual se denomina error estándar de la media.

Estimación puntual.

Sea x una variable aleatoria cuya función de densidad $f_x(x,\theta)$ depende de un parámetro desconocido θ y sean $x_1, x_2, \ldots, x_n$, los valores observados en una muestra aleatoria de tamaño n, de la población asociada a la variable aleatoria x.

Un estimador puntual del parámetro θ es una función:

$$\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$$

que hace corresponder a cada muestra aleatoria de tamaño n un valor estimado de θ. Obsérvese que el estimador $\hat{\theta}$ es a su vez una variable aleatoria que para cada muestra tiene un valor determinado.

Una propiedad importante de una estimación es que sus resultados produzcan un valor estimado lo más cercano posible al valor verdadero de una manera consistente. Esto es, que la dispersión de $\hat{\theta}$ sea lo más pequeña posible.

Lo anterior significa que un estimador debe tener las siguientes propiedades:

I) $\hat{\theta}$ - no-sesgado

II) $\sigma^2_{\hat{\theta}}$ - mínima

La propiedad I establece que:

$$E\left\{\hat{\theta}\right\} = \theta$$

Esto quiere decir que la variable aleatoria $\hat{\theta}$ tiene una densidad de probabilidad cuyo valor esperado es igual a θ.

La propiedad II establece que si hay otro estimador, digamos $\bar{\theta}$, que sea no-sesgado, ésto es,

$$E\left\{\bar{\theta}\right\} = E\left\{\hat{\theta}\right\} = \theta$$

Deseamos que

$$E\left\{(\hat{\theta} - \theta)^2\right\} < E\left\{(\bar{\theta} - \theta)^2\right\}$$

y por lo tanto

$$\sigma^2_{\hat{\theta}} < \sigma^2_{\bar{\theta}}$$

Intervalos de confianza.

El valor estimado del parámetro θ correspondiente a una muestra de tamaño n, es sólo un punto, y puesto que la probabilidad de que el valor estimado $\hat{\theta}$ sea exactamente igual a θ es cero, en casos prácticos lo normal es que θ no sea conocido. Es posible sin embargo determinar un intervalo dentro del cual, con un cierto grado de confianza, se encuentre el valor desconocido de θ. Así, se dice que el intervalo (a,b) es un intervalo del $(1-\alpha)\%$ de confianza del parámetro θ, si:

$$P(a \leq \theta \leq b) = 1 - \alpha$$

en donde los extremos a,b del intervalo son variables aleatorias puesto que dependen de la muestra que se tome. Esto es, el intervalo (a,b) es a su vez una variable aleatoria.

Así, suponiendo que un experimento se repite 100 veces, en cada una de las cuales se toma una muestra de tamaño n y se calcula (a,b), se tendrá en promedio que de cada 100 intervalos así obtenidos, $(1-\alpha)\%$ de ellos contienen el parámetro θ.

Ahora, en realidad se toma una sola muestra de tamaño n y se afirma que el intervalo obtenido contiene el parámetro θ. Esto puede ser falso o verdadero. Sin embargo, se tiene una probabilidad igual a $(1-\alpha)\%$ de estar en lo correcto.

Los límites del intervalo de confianza para el valor esperado están dados por:

$$\hat{\theta} \pm \beta \frac{\sigma_\theta}{\sqrt{n}}$$

en donde β depende del nivel de confianza deseado para la distribución de $\hat{\theta}$.

Para el caso en que la variancia de la población es desconocida, se estima comúnmente calculando la variancia s_θ^2 de la muestra, con lo cual los intervalos de confianza para la media θ se expresan:

$$\hat{\theta} - \beta s_\theta / \sqrt{n} < \theta < \hat{\theta} + \beta s_\theta / \sqrt{n}$$

Pruebas de hipótesis.

Definición. Una hipótesis estadística es una suposición acerca de la densidad de distribución de una variable aleatoria.

Las hipótesis pueden ser simples o compuestas. Si una hipótesis especifica los valores de todos los parámetros de una densidad de probabilidad de una variable aleatoria, se le denomina una hipótesis simple. De lo contrario, compuesta.

Las hipótesis admisibles pueden agruparse de la siguiente manera:

H_o : hipótesis bajo prueba

H_1 ; hipótesis alterna

Si bajo la suposición de que H_o es verdadera se encuentra que los resultados observados en una muestra aleatoria difieren notablemente de los obtenidos basándose en la teoría de la probabilidad y el muestreo, se dice que las diferencias observadas son significativas, y habrá cierta razón para rechazar la hipótesis, o cuando menos no aceptarla.

A los procedimientos que permiten decidir entre la aceptación o rechazo de una hipótesis se les llama pruebas de hipótesis.

Se llama región crítica de una prueba de hipótesis a aquella parte del espacio de muestra en donde se rechaza la hipótesis H_o. La figura 5 ilustra la definición.

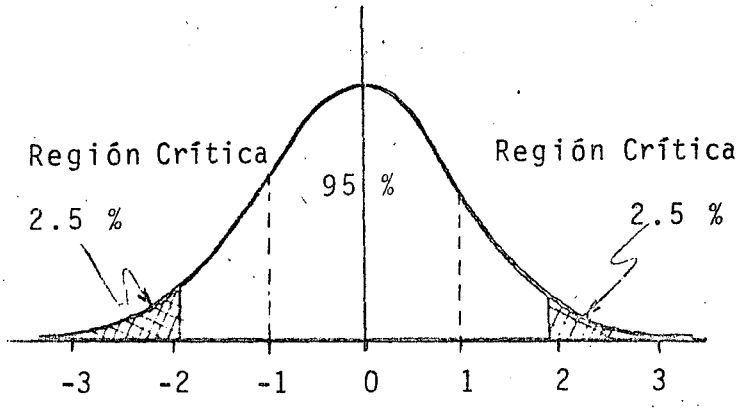


Fig. 5.- Región crítica para H_o con 95% de confianza.

Al diseñar una prueba de hipótesis se elige un nivel de significación, la máxima probabilidad de rechazar a H_o cuando ésta es verdadera.

Si, por ejemplo, se toma un nivel del 5%, existen aproximadamente 5 posibilidades en 100 de que se rechace la hipótesis cuando debía ser aceptada.

Veamos un ejemplo con una variable aleatoria normal estandarizada z. Esto es, $\mu_z = 0$; $\sigma_z^2 = 1$, que corresponde a la figura 5.

Como se indica en la figura, se tiene el 95% de confianza en que si la hipótesis es verdadera el valor de z caerá entre -1.96 y 1.96.
Si al extraer una muestra aleatoria se encuentra que el valor de z cae fuera del intervalo, se dice entonces que la hipótesis se rechaza con un nivel de significación del 5%.

Se podrían elaborar las siguientes reglas de decisión:

a) Rechazar la hipótesis, con un 5% de nivel de significación, si el valor de z cae fuera del intervalo (-1.96,1.96).

b) Aceptar la hipótesis en cualquier otro caso.

Si se desea otro nivel de significación, digamos del 1%, bastaría desplazar la región crítica a los extremos del intervalo (-2.58,2.58). Como estamos interesados en ambos extremos, a estas pruebas se les llama pruebas de dos extremos.

Para continuar con el estudio de las muestras, vamos a describir brevemente las distribuciones asociadas a variables aleatorias normales.

Supongamos que $x_i (i=1,2,\ldots,n)$ son variables aleatorias independientes normales. Además, el valor esperado de cada una de ellas es igual a cero y su desviación estándar es igual a la unidad. En este caso, la suma de los cuadrados de estas variables:

$$y^2 = \sum_{i=1}^{n} x_i^2$$

tiene la densidad de probabilidad χ^2 con k=n grados de libertad.

La densidad de probabilidad de la distribución es:

$$f(y_k) = 0 \qquad \text{para } y \leq 0$$

$$f(y_k) = (1/[2^{k/2} . \Gamma(k/2)])\ e^{-y/2}\ y^{k/2 -1}$$

$$\text{para } y \not\leq 0$$

donde $$\Gamma(w) = \int_0^\infty t^{w-1} e^{-t} dt$$

es la función gamma. En particular,

$$\Gamma(n+1) = n!$$

Como se puede ver, la distribución "ji cuadrada" se determina por un parámetro: el número de grados de libertad k.

Al aumentar el número de grados de libertad, la distribución se aproxima lentamente a la normal.

La figura 6 muestra varias curvas con diferentes grados de libertad k para la distribución "ji cuadrada".

Distribución t de Student.

Supongamos que z es una variable aleatoria normal estandarizada, o sea, $\mu_z = 0$, $\sigma_z = 1$ y v una variable aleatoria independiente de z, con una densidad de probabilidad χ^2 con k grados de libertad. En este caso la variable

$$t = z/\sqrt{v/k}$$

tiene una densidad de probabilidad llamada t o distribución t de Student.

Así, la relación entre la variable aleatoria normal estandarizada y la raíz cuadrada de la variable aleatoria independiente v, dividida por k, se distribuye según la ley de Student, con k grados de libertad.

Al aumentar el número de grados de libertad la distribución t se aproxima rápidamente a la normal. La figura 7 ilustra un ejemplo.

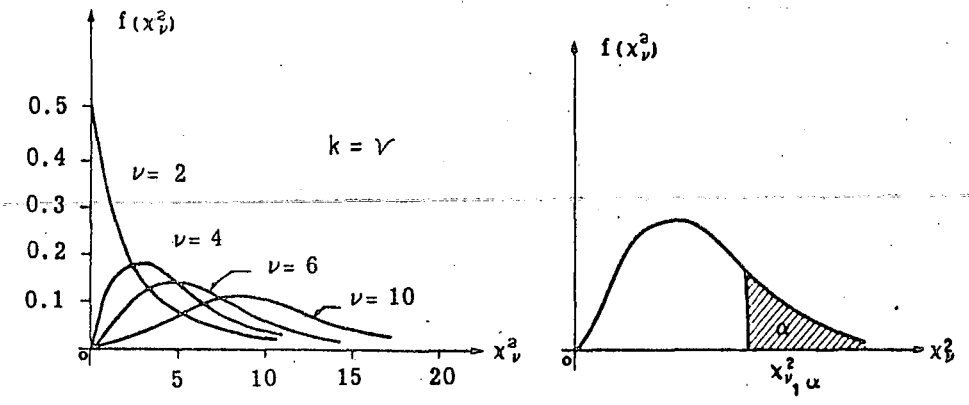


Fig. 6.- Densidad de distribución "ji cuadrada".

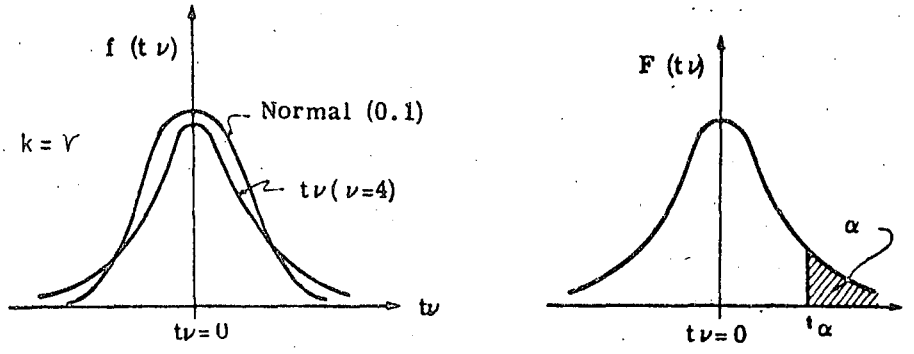


Fig. 7.- Densidad de distribución t de Student.

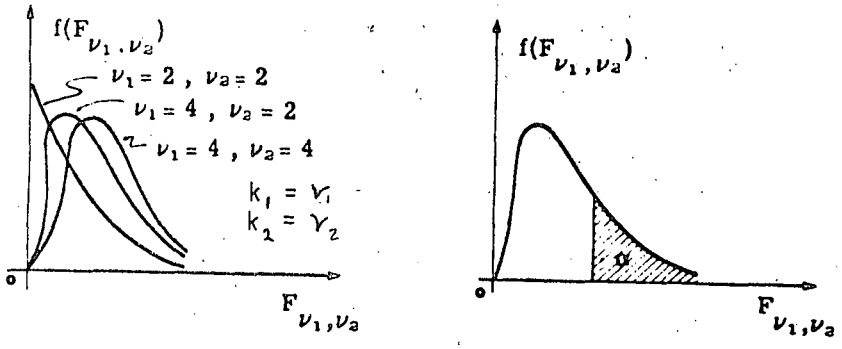


Fig. 8.- Densidad de Distribución F .

Distribución F de Fisher-Snedecor.

Si u y v son variables aleatorias independientes, distribuidas por la ley χ^2 con grados de libertad k_1 y k_2 respectivamente, la magnitud

$$F = \frac{u/k_1}{v/k_2}$$

tiene una distribución llamada distribución F de Fisher-Snedecor con grados de libertad k_1 y k_2. Su función de densidad está dada por:

$$f(x) = 0 \qquad \text{para } x \leq 0$$

$$= C_o \frac{x^{(k_1-2)/2}}{(k_2+k_1x)^{k_1+k_2/2}} \qquad \text{para } x > 0$$

donde

$$C_o = \frac{\Gamma(k_1+k_2/2) \cdot k_1^{k_2/2} \cdot k_2^{k_2/2}}{\Gamma(k_1/2) \quad \Gamma(k_2/2)}$$

Referencias

1.- Papoulis, A., "Probability, Random Variables and Stochastic Processes", McGraw-Hill, 1965.

2.- Gnedenko, V.B., "The Theory of Probability", New York, Chelsea, 1962.

3.- Deutsch, R., "Estimation Theory", New Jersey, Prentice-Hall, 1965.

4.- Spiegel, R.M., "Theory and Problems of Statistics", New York, Schaum, 1961.

5.- Moreno, B.A. y Jauffred, M.F.J., "Elementos de Probabilidad y Estadística", México, Representaciones y Servicios de Ingeniería S.A., 1969.

6.- Gmurman, V.E., "Teoría de las Probabilidades y Estadística Matemática", URSS, Editorial Mir, 1974.

APENDICE A

Least Squares Estimator (3)

Let us have a sequence of measurements $y_1, y_2, \ldots, y_m$, where m is the number of measurements. The measurements can be written as: measurement = true value + noise. Or:

$$\underline{y} = \underline{y}_t + \underline{\eta} \tag{A1-1}$$

and the expected value:

$$E\{\underline{y}\} = \underline{y}_t \tag{A1-2}$$

Now, the feature of the linear estimator is that we have a linear relationship between the state unknowns $\underline{x}$ which are to be estimated.

$$\underline{y}_t = A\underline{x}_t \tag{A1-3}$$

We also want that our linear estimates be unbiased; this means:

$$E\{\underline{\eta}\} = \underline{0} \tag{A1-4}$$

and its covariance matrix:

$$R = E\left\{\underline{\eta}\,\underline{\eta}^T\right\} \tag{A1-5}$$

The theory states (3) that the least squares linear unbiased estimate of $\underline{x}$ is defined as the vector $\underline{x}$, which minimizes the quadratic risk function with respect to $\underline{x}$:

$$J(\underline{x}) = (\underline{y} - A\underline{x})^T(\underline{y} - A\underline{x}) \tag{A1-6}$$

We have to note that the matrix A^TA has to be non-singular. To minimize (A1-6), differentiate with respect to $\underline{x}$, and equating the derivative to zero, we have:

$$\nabla_{\underline{x}} J(\underline{x})\Big|_{\underline{x}=\hat{\underline{x}}} = -2A^T\underline{y} + 2A^TA\hat{\underline{x}} = \underline{0} \tag{A1-7}$$

and

$$\hat{\underline{x}} = (A^TA)^{-1}A^T\underline{y} \tag{A1-8}$$

To show that it is a minimum point, we write the risk function in

the form of the squared norm:

$$d = \left\| \underline{y} - A\underline{x} \right\|^2 \qquad \text{(A1-9)}$$

If $\underline{y} = A\underline{\hat{x}}$,

$$d = \left\| \underline{y} - A\underline{x} \right\|^2 \geq \left\| \underline{y} \right\|^2 - \left\| A\underline{\hat{x}} \right\|^2 \qquad \text{(A1-10)}$$

Then J has a minimum if and only if $\underline{x} = \underline{\hat{x}}$, where $\underline{\hat{x}}$ is given by (A1-7).

The next step is to show that $\underline{\hat{x}}$ is an unbiased estimate.

Let us write:

$$\begin{aligned} E\left\{\underline{\hat{x}} - \underline{x}_t\right\} &= E\left\{(A^TA)^{-1}A^T\underline{y}\right\} - \underline{x}_t \\ &= (A^TA)^{-1}A^T E\left\{\underline{y}\right\} - \underline{x}_t \\ &= (A^TA)^{-1}A^TA\underline{x}_t - \underline{x}_t \\ &= \underline{x}_t - \underline{x}_t = 0 \end{aligned}$$

This means that $\underline{\hat{x}}$ approaches $\underline{x}_t$ and we do not have any deviation of this value. Therefore, $\underline{\hat{x}}$ is an unbiased estimate of $\underline{x}$.

The covariance matrix of $\underline{\hat{x}}$ can be written as follows:

$$\begin{aligned} \text{Cov}\,(\underline{\hat{x}}) &= E\left\{(\underline{\hat{x}} - \underline{x}_t)(\underline{\hat{x}} - \underline{x}_t)^T\right\} \\ &= E\left\{\underline{\hat{x}}\underline{\hat{x}}^T\right\} - \left[E\left\{\underline{\hat{x}}\right\}\right]^2 \\ &= E\left\{(A^TA)^{-1}A^T\underline{y}\,\underline{y}^TA\,(A^TA)^{-1}\right\} - \underline{x}_t\underline{x}_t^T \end{aligned}$$

Define

$$B = (A^TA)^{-1}A^T \quad ; \quad B^T = A(A^TA)^{-1}$$

$$\text{Cov}\,(\underline{\hat{x}}) = E\left\{B\underline{y}\,\underline{y}^TB^T\right\} - \underline{x}_t\underline{x}_t^T$$

$$E\left\{B\,\underline{y}\,\underline{y}^TB^T\right\} = BE\left\{\underline{y}\,\underline{y}^T\right\}B^T = B\left[A\underline{x}_t\underline{x}_t^TA^T + E\left\{\underline{\eta}\underline{\eta}^T\right\}\right]B^T$$

$$= BA\underline{x}_t\underline{x}_t^TA^TB^T + BE\left\{\underline{\eta}\underline{\eta}^T\right\}B^T$$

$$= I\,\underline{x}_t\,\underline{x}^T I + BE\left\{\underline{\eta}\,\underline{\eta}^T\right\} B^T \qquad \text{(A1-11)}$$

Then

$$\text{Cov}(\underline{\hat{x}}) = \underline{x}_t\,\underline{x}_t^T + BE\left\{\underline{\eta}\,\underline{\eta}^T\right\} B^T - \underline{x}_t\,\underline{x}_t^T$$

$$= (A^TA)^{-1}A^TE\left\{\underline{\eta}\,\underline{\eta}^T\right\} A(A^TA)^{-1} \qquad \text{(A1-12)}$$

Weighted Least Squares Linear Estimator

Most of the times we wish to assign a relative importance to the measurements. Then, we assign some weighting factors to the risk function associated with the different measurements. Thus, we can write:

$$J(\underline{x}) = (\underline{y} - A\underline{x})^T W^{-1} (\underline{y} - A\underline{x}) \qquad \text{(A1-13)}$$

In practical situations, we generally have that W^{-1} is a symmetric and positive-definite matrix.

The estimate $\underline{\hat{x}}$, which produces a minimum of the risk function, can be established by operating on J, as follows:

$$\nabla_{\underline{x}} J(\underline{x}) = \underline{0} \qquad \text{(A1-14)}$$

$$\nabla_{\underline{x}} J(\underline{x})\Big|_{\underline{x}=\underline{x}} = \nabla_{\underline{x}} (\underline{y}^T - \underline{x}^T A^T) W^{-1} (\underline{y} - A\underline{x})\Big|_{\underline{x}=\underline{x}}$$

$$= -2A^TW^{-1}\underline{y} + 2A^TW^{-1}A\underline{x} = \underline{0}$$

$$A^TW^{-1}A\underline{\hat{x}} = A^TW^{-1}\underline{y} \qquad \text{(A1-15)}$$

$$\underline{\hat{x}} = (A^TW^{-1}A)^{-1}A^TW^{-1}\underline{y} \qquad \text{(A1-16)}$$

The proof that (A1-16) yields an absolute minimum requires some mathematical manipulations that are out of our present objectives. Reference (3) deals with it in detail.

The covariance matrix of the estimate is:

$$\text{Cov}(\underline{\hat{x}}) = E\left\{\underline{\hat{x}}\,\underline{\hat{x}}^T\right\} - \left[E\left\{\underline{\hat{x}}\right\}\right]^2 \qquad \text{(A1-17)}$$

$$E(\underline{\hat{x}}) = \underline{x}_t \quad ; \quad E\left[\left\{\underline{x}\right\}\right]^2 = \underline{x}_t\,\underline{x}_t^T \qquad \text{(A1-18)}$$

$$E\left\{\hat{\underline{x}}\,\hat{\underline{x}}^T\right\} = E\left\{(A^T W^{-1} A)^{-1} A^T W^{-1}\, \underline{y}\,\underline{y}^T W^{-1} A (A^T W^{-1} A)^{-1}\right\}$$

By analogy with (A1-11) and writing

$$C = (A^T W^{-1} A)^{-1} A^T W^{-1} \quad ,$$

$$E\left\{\hat{\underline{x}}\,\hat{\underline{x}}^T\right\} = C\left[A\,\underline{x}_t\,\underline{x}_t^T A^T + E\left\{\eta\,\eta^T\right\}\right] C^T \qquad \text{(A1-19)}$$

Substituting (A1-17) and (A1-18) into (A1-19):

$$\begin{aligned} \text{Cov}(\hat{\underline{x}}) &= (A\,\underline{x}_t\,\underline{x}_t^T A^T C^T + C E\left\{\eta\,\eta^T\right\} C^T - \underline{x}_t\,\underline{x}_t^T \\ &= I\,\underline{x}_t\,\underline{x}_t^T I + C E\left\{\eta\,\eta^T\right\} C^T - \underline{x}_t\,\underline{x}_t^T \\ &= C E\left\{\eta\,\eta^T\right\} C^T \\ &= (A^T W^{-1} A)^{-1} A^T W^{-1} E\left\{\eta\,\eta^T\right\} W^{-1} A (A^T W^{-1} A)^{-1} \end{aligned} \qquad \text{(A1-20)}$$

We now choose $W = E\left\{\eta\,\eta^T\right\}$, which will provide the minimum variance estimator (3). In other words, any other weighting matrix would give us a greater covariance matrix for the estimated values. Then, substituting $W = E\left\{\eta\,\eta^T\right\}$ into (A1-20) we have:

$$\begin{aligned} \text{Cov}(\hat{\underline{x}}) &= (A^T W^{-1} A)^{-1} A^T W^{-1} W W^{-1} A (A^T W^{-1} A)^{-1} \\ &= (A^T W^{-1} A)^{-1} A^T W^{-1} I A (A^T W^{-1} A)^{-1} \\ &= (A^T W^{-1} A)^{-1} (A^T W^{-1} A)(A^T W^{-1} A)^{-1} \\ &= (A^T W^{-1} A)^{-1} I \end{aligned}$$

$$\text{Cov}(\hat{\underline{x}}) = (A^T R^{-1} A)^{-1} \qquad \text{(A1-21)}$$

Now, if $W_1 \neq R$, we state, without proof, that the covariance matrix is:

$$\text{Cov}(\hat{\underline{x}}_1) = C_1 R^{-1} C_1^T \qquad \text{(A1-22)}$$

where

$$C_1 = (A^T W_1^{-1} A)^{-1} A^T W_1^{-1}$$

is greater than the matrix (A1-21). Then, we write:

$$C_1 R^{-1} C_1^T \geq (A^T R^{-1} A)^{-1}$$

Thus, our choosing of $W = R$ gives us the minimum variance estimate. This expression is of more theoretical than practical use, because in practical applications the exact values of the elements of the covariance matrix of the errors are seldom known. However, it gives us the possibility to state that we use some values as close as possible to the true but unknown elements of the covariance matrix.

REFERENCIAS RELEVANTES SOBRE MATRICES DISPERSAS EN SISTEMAS ELECTRICOS DE POTENCIA.

Tinney, W.F. and Walker J.W."Direct solutions of sparse network equations by optimally ordered triangular factorization", IEEE, Proceedings, Nov. 1967. PP 1801-1809.

Tinney, W.F. and Hart, C.E. "Power flow solution by Newton's Method". IEEE, PAS - VOL 86 1967, PP 1449-1460.

Zollenkopf, K. "B1-Factorisation-Basic Computational algorithm and programming techniques". In Reid, J.K. (ed.)" "Large sparse sets of linear equations". Academic Press, 1971.

Brameller, A., Allan, R.N. and Hamam, Y.M. "Sparsity", Pitman Publishing, London 1976.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA IV CONCEPTOS DE PROBABILIDAD Y ESTADISTICA

DR. SEGIO A. MOLINA GARCIA

ENERO, 1979.

APENDICE B

Conceptos Básicos de Cálculo matricial .

La derivada de una matriz A, cuyos elementos son funciones escalares de la variable t, está definida como la matriz cuyos elementos son las derivadas de los elementos de la matriz-original A.

$$\frac{dA}{dt}[m*n] = \begin{bmatrix} \frac{da_{11}}{dt} & \cdots & \frac{da_{1n}}{dt} \\ \vdots & \ddots & \\ \frac{da_{m1}}{dt} & \cdots & \frac{da_{mn}}{dt} \end{bmatrix} \qquad (1)$$

$$\frac{d(A+B)}{dt} = \frac{dA}{dt} + \frac{dB}{dt} \qquad (2)$$

$$\frac{d(AB)}{dt} = \frac{dA}{dt}B + A\frac{dB}{dt} \qquad (3)$$

$$\frac{dA^2}{dt} = \frac{dA}{dt}A + A\frac{dA}{dt} \qquad (4)$$

Si A es cuadrada

$$\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1} \qquad (5)$$

Definimos al operador de derivadas parciales matricial, como el vector

$$\nabla_q = \nabla_{[n*1]} = \begin{bmatrix} \frac{\partial}{\partial q_1} \\ \frac{\partial}{\partial q_2} \\ \vdots \\ \frac{\partial}{\partial q_n} \end{bmatrix} \qquad (6)$$

como este operador está definido como un vector, solamente puede ser aplicado a un vector transpuesto,

$$\nabla_{q\,[n*1]}\; x^{T}_{[1*m]} = \begin{bmatrix} \frac{\partial}{\partial q_1} \\ \frac{\partial}{\partial q_2} \\ \vdots \\ \frac{\partial}{\partial q_n} \end{bmatrix} \begin{bmatrix} x_1, x_2, \ldots, x_m \end{bmatrix} \quad (7)$$

$$\nabla_q\, x^T = B_{[n*m]}$$

a menos que ∇_q sea del orden de x, las reglas de diferenciación similares a (2), (3), (4) y (5), simplemente no son validas. Por ejemplo $\nabla_q (A_{[1*n]} B_{[n*m]})$ no puede ser expresada como la forma (3). Hay un caso especial en el cual la regla del producto es válida, el caso de $A_{[1*n]} B_{[n*1]}$; entonces

$$\nabla_q\, AB = (\nabla_q A) B + \nabla_q B^T A^T \qquad (8)$$

Si $B_{[n*1]}$ es un vector cuyos elementos no son funciones de q, entonces

$$\nabla_q\, B^T q = B \qquad (\text{un vector}) \quad (9\text{-}a)$$

$$\nabla_q\, q^T D = D \qquad (\text{Una matriz}) \quad (9\text{-}b)$$

Ahora, en el caso de formas cuadráticas, tenemos

$$Q = \underline{a}^T(x)\, \underline{\Phi}\, \underline{a}(x) \qquad (10)$$

Si Φ es una matriz simétrica positiva-semidefinida, Φ puede ser escrita como sigue

$$\underline{\Phi} = S^T S$$

APENDICE B

Conceptos Básicos de Cálculo matricial .

La derivada de una matriz A, cuyos elementos son funciones escalares de la variable t, está definida como la matriz cuyos elementos son las derivadas de los elementos de la matriz-original A.

$$\frac{dA}{dt}[m*n] = \begin{bmatrix} \frac{da_{11}}{dt} & \cdots & \frac{da_{1n}}{dt} \\ \vdots & \ddots & \\ \frac{da_{m1}}{dt} & \cdots & \frac{da_{mn}}{dt} \end{bmatrix} \qquad (1)$$

$$\frac{d(A+B)}{dt} = \frac{dA}{dt} + \frac{dB}{dt} \qquad (2)$$

$$\frac{d(AB)}{dt} = \frac{dA}{dt}B + A\frac{dB}{dt} \qquad (3)$$

$$\frac{dA^2}{dt} = \frac{dA}{dt}A + A\frac{dA}{dt} \qquad (4)$$

Si A es cuadrada

$$\frac{dA^{-1}}{dt} = -A^{-1}\frac{dA}{dt}A^{-1} \qquad (5)$$

Definimos al operador de derivadas parciales matricial, como el vector

$$\nabla_q = \nabla_{[n*1]} = \begin{bmatrix} \frac{\partial}{\partial q_1} \\ \frac{\partial}{\partial q_2} \\ \vdots \\ \frac{\partial}{\partial q_n} \end{bmatrix} \qquad (6)$$

como este operador está definido como un vector, solamente puede ser aplicado a un vector transpuesto,

$$\nabla_{q\,[n*1]}\; x^{T}_{[1*m]} = \begin{bmatrix} \frac{\partial}{\partial q_1} \\ \frac{\partial}{\partial q_2} \\ \vdots \\ \frac{\partial}{\partial q_n} \end{bmatrix} \begin{bmatrix} x_1, x_2, \ldots, x_m \end{bmatrix} \quad (7)$$

$$\nabla_q\, x^T = B_{[n*m]}$$

a menos que ∇_q sea del orden de x, las reglas de diferenciación similares a (2),(3),(4)y(5), simplemente no son validas. Por ejemplo $\nabla_q (A_{[1*n]} B_{[n*m]})$ no puede ser expresada como la forma (3). Hay un caso especial en el cual la regla del producto es válida, el caso de $A_{[1*n]}B_{[n*1]}$; entonces

$$\nabla_q\, AB = (\nabla_q A)B + \nabla_q B^T A^T \qquad (8)$$

Si $B_{[n*1]}$ es un vector cuyos elementos no son funciones de q, entonces

$$\nabla_q\, B^T q = B \qquad (\text{un vector}) \quad (9\text{-}a)$$

$$\nabla_q\, q^T D = D \qquad (\text{Una matriz}) \quad (9\text{-}b)$$

Ahora, en el caso de formas cuadráticas, tenemos

$$Q = \underline{a}^T(x)\, \Phi\, \underline{a}(x) \qquad (10)$$

Si Φ es una matriz simétrica positiva-semidefinida, Φ puede ser escrita como sigue

$$\Phi = S^T S$$

Entonces, $Q = \underline{a}^T(x) S^T S \, \underline{a}(x)$

Definimos: $A = \underline{a}^T(x) S^T = \underline{a}^T S^T$; $B = S\,\underline{a}(x) = S\underline{a}$

de (8) $\nabla Q = \nabla(AB) = (\nabla A) B + (\nabla B^T) A^T$

Pero $A^T = S\,\underline{a} = B$

$\nabla Q = 2\,(\nabla A)\, B$

$\nabla Q = 2(\nabla\, \underline{a}^T S^T) S\,\underline{a}$

$\nabla Q = 2(\nabla\, \underline{a}^T)\, \Phi\, \underline{a}$

$\nabla Q = 2(\nabla\, \underline{a}^T(x))\, \Phi\, a(x)$ (11)

Un caso especial es cuando $\underline{a}(x) = \underline{x}$

Entonces, escribimos

$Q = \underline{x}^T \Phi\, \underline{x}$

$\nabla Q = 2\nabla(\underline{x}^T)\, \Phi\, \underline{x}$

$= 2\Phi\, \underline{x}$ (12)

Otro caso es cuando tenemos una matriz $B(x) = B\underline{x}$; $B_{[n*n]}$

$Q = \underline{x}^T B^T \Phi\, B\, \underline{x}$

$\nabla Q = 2\, B^T \Phi\, B\, \underline{x}$ (13)

El operador integral puede ser aplicado como se ilustra

$$\int A(t)\,dt = \begin{bmatrix} \int a_{11}(t)\,dt & \cdots & \int a_{1m}(t)\,dt \\ \vdots & & \vdots \\ \int a_{n1}(t)\,dt & \cdots & \int a_{nm}(t)\,dt \end{bmatrix} \quad (14)$$

El valor esperado de un vector cuyos componentes son variables aleatorias, puede ser expresado en la forma siguiente

$$E\{\underline{x}\} = \begin{bmatrix} E\{x_1\} \\ \vdots \\ E\{x_n\} \end{bmatrix} = \begin{bmatrix} \int x_1 f(x_1)\,dx_1 \\ \vdots \\ \int x_n f(x_n)\,dx_n \end{bmatrix} \quad (15)$$

El valor esperado para la matriz de momentos de segundo orden del vector aleatorio $\underline{x}$ es

$$\Phi = E\{\underline{x}\,\underline{x}^T\} = E\left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} x_1, & x_2, & \ldots, & x_n \end{bmatrix} \right\}$$

$$\Phi = \begin{bmatrix} E\{x_1 x_1\} & \cdots & E\{x_1 x_n\} \\ \\ E\{x_n x_1\} & \cdots & E\{x_n x_n\} \end{bmatrix} = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1n} \\ \\ \mu_{n1} & & \mu_{nn} \end{bmatrix}$$

donde los momentos $\mu_{ij} = E\{x_i \cdot x_j\}$

Transformaciones lineales de un vector aleatorio $\underline{x}$.

Sea $$\underline{y} = H\,\underline{x} \qquad (17)$$

con $E\{\underline{x}\} = \underline{0}$, entonces la matriz de covariancia de $\underline{y}$ es (podemos sacar H fuera del operador de esperanza matemática si no está en función de las variables aleatorias)

$$R = E\{\underline{y}\,\underline{y}^T\} = E\{H\underline{x}\,\underline{x}^T H^T\} = H[E\{x\,\underline{x}^T\}]H^T$$

$$R = H\,\Phi\,H^T \qquad (18)$$

Si $E\{\underline{x}\} \neq \underline{0}$, la matriz de covariancia R, se puede demostrar que, tiene la misma forma que (18).

Si $\underline{y}$ esta dado por la expresión (19) y $\underline{x}$ tiene el valor esperado $\underline{\mu}_x = E\{x\}$. Calculamos la covariancia de $\underline{y}$, donde $\underline{z}$ y A son matrices con elementos constantes.

$$\underline{y} = A\,\underline{x} + \underline{z} \qquad (19)$$

$$Cov(\underline{y}) = E\{\underline{y}\,\underline{y}^T\} - \underline{\mu}_y\,\underline{\mu}_y^T \qquad (20)$$

$$E\{\underline{x}\,\underline{x}^T\} = Cov(\underline{x}) + \underline{\mu}_x\,\underline{\mu}_y^T \qquad (21)$$

$$E\{\underline{y}\,\underline{y}^T\} = E\{(A\underline{x} + \underline{z})(A\underline{x} + \underline{z})^T\} \qquad (22)$$

$$\underline{\mu}_y \underline{\mu}_y^T = E\{A\underline{x} + \underline{z}\}[E\{A\underline{x} + \underline{z}\}]^T \qquad (23)$$

Usando algebra matricial y aplicando el operador de esperanza matemática, se obtiene la siguiente expresión.

$$Cov(\underline{y}) = A \; cov(\underline{x}) \; A^T \qquad (24)$$

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

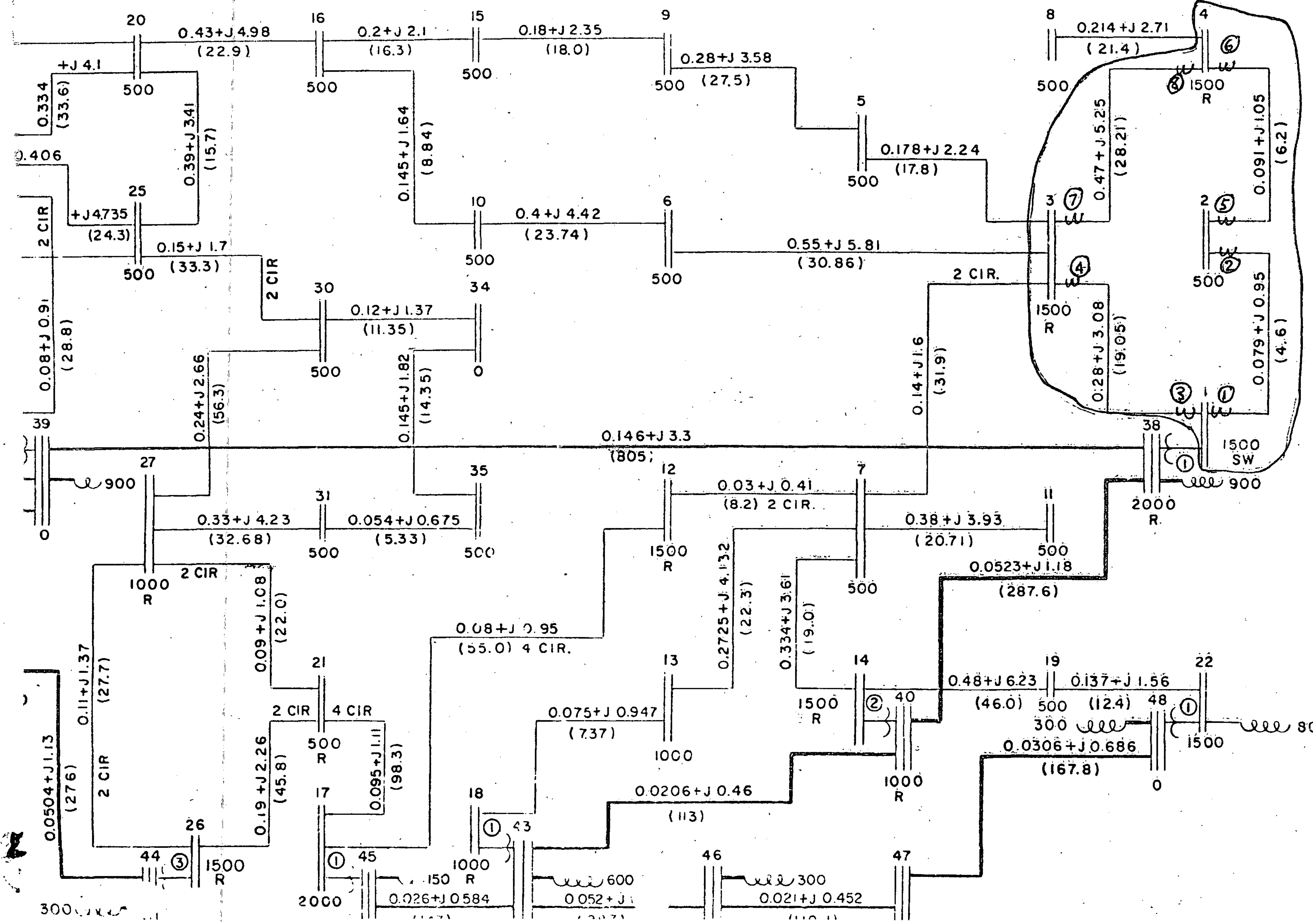
TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA V METODOLOGIA DE ESTIMACION DE ESTADO PARA SISTEMAS ELECTRICOS DE POTENCIA

DR. ALBERTO MAYER SASSON

ENERO, 1979.

(1)


20
0.43+J 4.98
(22.9)
16
0.2+J 2.1
(16.3)
15
0.18+J 2.35
(18.0)
9
0.28+J 3.58
(27.5)
8
0.214+J 2.71
(21.4)
4
6
+J 4.1
500
500
500
500
500
1500
R
8
0.334
(33.6)
0.39+J 3.41
(15.7)
0.406
0.145+J 1.64
(8.84)
5
500
0.178+J 2.24
(17.8)
0.47+J 5.25
(28.21)
0.091+J 1.05
(6.2)
25
+J 4.735
(24.3)
10
0.4+J 4.42
(23.74)
6
3
7
2
5
2 CIR
0.15+J 1.7
(33.3)
500
500
500
0.55+J 5.81
(30.86)
2 CIR.
4
2
500
2 CIR
30
0.12+J 1.37
(11.35)
34
1500
R
500
0
0.08+J 0.91
(28.8)
0.24+J 2.66
(56.3)
0.145+J 1.82
(14.35)
0.14+J 1.6
(31.9)
0.28+J 3.08
(19.05)
0.079+J 0.95
(4.6)
3
1
39
0.146+J 3.3
(805)
38
1500
SW
1
27
900
35
12
0.03+J 0.41
(8.2) 2 CIR.
7
11
900
2000
R.
0
31
0.33+J 4.23
(32.68)
0.054+J 0.675
(5.33)
500
500
0.38+J 3.93
(20.71)
1500
R
500
0.2725+J 4.132
(22.3)
0.334+J 3.61
(19.0)
500
0.0523+J 1.18
(287.6)
1000
R
2 CIR
0.09+J 1.08
(22.0)
0.08+J 0.95
(55.0) 4 CIR.
0.11+J 1.37
(27.7)
21
13
14
19
22
0.48+J 6.23
(46.0)
0.137+J 1.56
(12.4)
2
40
500
48
1
1500
R
300
1500
2 CIR
4 CIR
0.075+J 0.947
(7.37)
500
R
1000
0.0306+J 0.686
(167.8)
0.0504+J 1.13
(276)
2 CIR
0.19+J 2.26
(45.8)
0.095+J 1.11
(98.3)
17
18
1000
R
0
0.0206+J 0.46
(113)
26
1
43
44
3
1500
R
1
45
150
1000
R
600
46
300
47
2000
0.026+J 0.584
0.052+J
0.021+J 0.452
300

(5) Number of busses = 4 ; Number of Measurements = 8 — CASE A — True flows perturbed in the 2 % ra[illegible] with random member generator
Ref. [illegible]s = Bus No. 1 ; Ref. Voltage Manitude = 1.02

Meas. #	Bus to Bus	True Flows P	Q	"Measured" Flows P	Q	.35% F.S.	Random #S	Weight, W^{-1}
1	1 2	514.8	+j 50.9	504.5	+j 49.9	.052	-.9994	$.7895 \times 10^{-2}$
2	2 1	-512.8	-j 35.9	-523.0	-j 36.6	.052	.9969	$.8254 \times 10^{-2}$
3	1 3	237.3	-j 0.1	239.4	-j 0.1	.052	.4260	$.3358 \times 10^{-2}$
4	3 1	-235.8	-j 22.4	-235.1	-j 22.3	.052	-.1426	$.3316 \times 10^{-2}$
5	2 4	383.7	-j 13.3	+379.8	-j 13.2	.052	-.5097	$.6504 \times 10^{-2}$
6	4 2	-382.4	+j 15.7	-379.6	+j 15.6	.052	-.3591	$.5503 \times 10^{-2}$
7	3 4	30.8	-j 31.2	31.2	-j 31.6	.052	.5754	$.1256 \times 10^{-2}$
8	4 3	- 30.8	-j 25.8	- 31.2	-j 26.1	.052	.5937	$.1225 \times 10^{-2}$

Optimal order is equal to original order using Tinney Scheme 1.

busses (1 2 3 4) / measurements (1 2 3 4 5 6 7 8)

$$= \begin{bmatrix} -1 & & & \\ 1 & & & \\ & -1 & & \\ & 1 & & \\ 1 & & & -1 \\ -1 & & & 1 \\ & 1 & -1 & \\ & -1 & & 1 \end{bmatrix} ; \text{Diag. elemts of } D = \begin{bmatrix} .1394 \times 10^{7} \\ .1333 \times 10^{7} \\ .3113 \times 10^{6} \\ .3153 \times 10^{6} \\ .1636 \times 10^{7} \\ .1636 \times 10^{7} \\ .2866 \times 10^{6} \\ .2937 \times 10^{6} \end{bmatrix} ; A^t DA = \begin{bmatrix} .599 \times 10^{7} & 0 & -.3272 \times 10^{7} \\ & .1207 \times 10^{7} & -.5803 \times 10^{6} \\ & & .3852 \times 10^{7} \end{bmatrix}$$

$$\text{Triang} \left\{ A^t DA \right\} = \begin{Bmatrix} 1 & 0 & -0.5454 \\ & 1 & -0.4808 \\ & & 1 \end{Bmatrix}$$

(3)

ITERATIONS

	1		2		3		4	
lt. Bus 2	1.0200	+j0.0000	1.0120	-j0.0476	1.0108	-j0.0475	1.0108	-j0.0476
lt. Bus 3	1.0200	+j0.0000	1.0101	-j0.0710	1.0076	-j0.0709	1.0075	-j0.0710
lt. Bus 4	1.0200	+j0.0000	1.0101	-j0.0868	1.0063	-j0.0869	1.0063	-j0.0870
$\mathcal{V}$m=1	0.0090	+j0.0466	0.0090	+j0.0466	0.0090	+j0.0466	0.0090	+j0.0466
$\mathcal{V}$m=2	-0.0090	-j0.0485	-0.0094	-j0.0485	-0.0094	-j0.0485	-0.0094	-j0.0485
$\mathcal{V}$m=3	-0.0125	+j0.0717	0.0125	+j0.0717	0.0125	+j0.0717	0.0125	+j0.0717
$\mathcal{V}$m=4	-0.0072	-j0.0709	-0.0123	-j0.0708	-0.0124	-j0.0709	0.0124	-j0.0709
$\mathcal{V}$m=5	0.0027	+j0.0392	0.0046	+j0.0392	0.0046	+j0.0393	0.0046	+j0.0393
$\mathcal{V}$m=6	-0.0011	-j0.0393	-0.0045	-j0.0393	-0.0045	-j0.0394	-0.0046	-j0.0394
$\mathcal{V}$m=7	0.0003	+j0.0162	0.0012	+j0.0162	0.0011	+j0.0163	0.0011	+j0.0163
$\mathcal{V}$m=8	0.0002	-j0.0162	-0.0014	-j0.0162	-0.0015	-j0.0163	-0.0015	-j0.0163
D$\mathcal{V}$m)1	$-.1566*10^5$	$-j\ .1227*10^4$	$-.1019*10^5$	$-j\ .9976*10^3$	$-.1014*10^5$	$-j\ .7617*10^3$	$-.1013*10^5$	$-j\ .7612*10^3$
D$\mathcal{V}$m)2	$-.6164*10^4$	$-j\ .3531*10^5$	$-.7054*10^4$	$-j\ .3522*10^5$	$-.7060*10^4$	$-j\ .3525*10^5$	$-.7061*10^4$	$-j\ .3525*10^5$
D$\mathcal{V}$m)3	$-.6253*10^4$	$-j\ .1377*10^6$	$-.1557*10^5$	$-j\ .1379*10^6$	$-.1565*10^5$	$-j\ .1382*10^6$	$-.1566*10^5$	$-j\ .1382*10^6$

Meas.#	Bus to Bus		Calculated Flows P	Calculated Flows Q	Meas-Calc. Residuals P	Meas-Calc. Residuals Q	True-Calc. Residuals P	True-Calc. Residuals Q
1	1	2	515.3	51.0	-10.79	-1.09	-0.5	-0.1
2	2	1	-513.3	-36.0	- 9.73	-0.69	0.5	0.1
3	1	3	237.0	- 0.1	2.36	0.05	0.3	0.0
4	3	1	-235.5	-22.4	0.34	0.07	0.3	0.0
5	2	4	380.9	-13.3	- 1.13	0.15	2.8	0.0
6	4	2	-379.6	15.5	- 0.02	0.09	-2.8	0.2
7	3	4	80.6	-31.2	0.62	-0.40	0.2	0.0
8	4	3	- 30.5	-25.9	- 0.63	-0.26	0.3	0.1

J = 2.92 ; $\chi^2_{10\ d.f.}^{99.5\ \%} = 25.2$

Detection Test is Passed

Voltage Results =

1	1.020 ∠ 0.000
2	1.012 ∠ -2.694
3	1.010 ∠ -4.032
4	1.010 ∠ -4.940

CASE B Same as Case A except that a bad measurement of 0. + j0. was assumed for Meas. 1.

Changes With Respect to CASE A

W_1^{-1} =	$0.9187 * 10^{-3}$	$(A^t DA)_{11}$ =	$0.1658 * 10^{8}$
D_{11} =	$0.1198 * 10^{8}$	Triag. $(A^t DA)_{13}$ =	-0.1973

RESULTS OF CASE B (4 iterations)

Meas. #	Bus to Bus		Calculated Flows P	Calculated Flows Q	Meas.- Calc. Residuals P	Meas.- Calc. Residuals Q
1	1	2	62.0	4.1	- 62.00	- 4.07
2	2	1	- 62.0	-13.3	-461.03	-23.38
3	1	3	175.1	8.8	64.97	8.77
4	3	1	-174.3	-21.4	- 60.88	- 0.90
5	2	4	415.7	- 9.2	- 35.90	- 3.92
6	4	2	-414.1	13.9	34.52	1.72
7	3	4	- 8.5	-33.9	39.69	2.28
8	4	3	8.5	-24.2	- 39.66	- 1.97

$J = 3540 ; \chi^{99.5\ \%}_{10\,d.f.} = 25.2$

Detection Test is Failed

Voltage Results =

1	1.020	$\angle$ + 0.000
2	1.019	$\angle$ - 0.321
3	1.013	$\angle$ - 2.974
4	1.016	$\angle$ - 2.739

Largest Normalized Residual = Meas. # 1 = -3.78 -j 2.48 ; $\left| \tau \right|^{99.5\ \%}_{10\ d.f.}$ = 3.2

Measurement # 1 is Identified as Bad.

(15)

{CASE B} Continued

$$(A^t DA)^{-1} = \begin{bmatrix} .7360 * 10^{-7} & .9075 * 10^{-6} & .6740 * 10^{-7} \\ & & .1642 * 10^{-6} \\ & & .3416 * 10^{-6} \end{bmatrix}$$

Normalized Flow Residuals

P	Q
-3.78	- j .248
-2.85	- j .144
.702	+ j .0957
- .668	- j .0099
- .355	- j .0388
.340	+ j .0171
.698	+ j .0397
- .709	- j .0351

Clearly, Measurement No. 1 , is the largest and is the only one that fails the Student - t Test.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA V METODOLOGIA DE ESTIMACION DE ESTADO PARA SISTEMAS ELECTRICOS DE POTENCIA

J. F. DOPAZO, S. T. HERMANN, A. M. SASSON
Y L. S. VANSLYCK

ENERO, 1979.

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F.

THE AEP STATE ESTIMATION MONITORING AND SECURITY SYSTEM

J.F. DOPAZO, S.T. EHRMANN, A.M. SASSON and L.S. VAN SLYCK

American Electric Power Service Corporation, 2 Broadway, New York, N. Y., 10004 (USA)

SUMMARY

The AEP state estimation monitoring system has been satisfactorily functioning on line. The importance and justification of monitoring a power system through state estimation calculations is discussed in this paper. An intuitive introduction to state estimation concepts serves as a background to the formal presentation of the AEP state estimation algorithm. As an aid to those implementing similar projects in the near future, the paper presents many details on the programming system structure, the meaning of redundancy, the choice of the measurement system and the implications behind the solution of the equations in the AEP algorithm. Some suggestions are also made on procedures to test and debug the real-time programming system. Finally the present state and plans in the security area are detailed with emphasis on contingency calculations and voltage level correction procedures.

INTRODUCTION

The AEP state estimation monitoring system [1-11] is fully implemented and working satisfactorily. A small number of companies or countries have already initiated similar projects while many others are now considering the installation

of new computerized control centers that will include the monitoring function.

What are the main considerations behind this trend? All power systems are monitored to various degrees, whether manually or simi-automatic. As systems grow and become more complex, their operation becomes more involved. Security considerations demand careful analysis before planned switching operations. Loading of facilities and voltage levels have to be seen to be within certain limits. In case of unforseen outages or abnormal system changes, it is increasingly vital to rapidly assess the ability of the system to perform properly under the new present conditions and to be able to initiate corrective actions as quickly as possible if necessary. Furthermore, security or reliability considerations make it highly desirable to assess system performance under contingent conditions.

In order to fulfill better the above tasks a greater number of measurements have to be brought into the control center. Conventional individual meter displays are clearly inadequate if operators have to visually communicate with large numbers of them. In addition, there is the need for checking limiting or abnormal conditions which the operator would only be able to do in a slow manner by following each meter one by one. Furthermore, as it is frequently the case, it is not economical to measure all quantities that are desired, such as, all line flows and currents, all voltages and all injections. While some of these quantities are expensive to get, such as voltage phase angles with a common reference, others may require the installation of current or potential transformers at a prohibitive cost in some system locations. The need for computing some quantities based on other quantities being measured becomes apparent.

The justification of a real-time computer can be made on the grounds of interfacing between an expanded measurement system and operators, presenting system conditions in a

meaningful manner. Further justification is based on the capacity of the computer to perform computations that will automatically check operating limits and compute non-measured quantities and thus extend and complete the measurement system.

The following question arises. What if one or more measurements are bad, that is, outside the accuracy range. Clearly, if such measurements are displayed or used to compute others, an erroneous system condition is communicated to the operator. It would be frustrating to conclude that after all the expense of installing an expanded measurement system, together with its associated and costly communications system and that of a computerized control center, an erroneous system condition may be represented because a single measurement or communication device has become faulty. The basic requirement of insuring the validity of the system conditions communicated to the operator is the central task of state estimation. While it is important to know when some measurements arriving at the control center are faulty it is equally important to be able to present system conditions to the operator without the effect of the bad measurements. This is also accomplished by state estimation. In short, through computations on incoming measurements, state estimation quantities communicated to the operator are guaranteed to be correct even if some measurements were bad. State estimation will not only detect and isolate bad measurements but will provide the means for compensating for such measurements, such that in most cases, the presence of bad measurements becomes transparent to the operator. Bad measurements are announced only for the purpose of making them evident so they can be fixed.

The point has been implied, that for proper monitoring, a measurement system consisting of high quality on site devices, high speed simultaneous communication, a real-time computer, state estimation calculations and efficient displays

are needed. All these pieces have to work together and all are within present day technology. The AEP state estimation monitoring system has been designed with the above considerations in mind and has been performing as planned.

Using the knowledge of present conditions provided through state estimation, the security or reliability of operation can be studied by performing contingency calculations. An important problem here is the representation of external or internal non-monitored systems as their reaction to internal contingencies must be taken into account. This important problem is under study now. Another important security function is the correction of abnormal voltage levels.

CONCEPTS IN STATE ESTIMATION

Consider the measurement of a flow in a transmission line. If a single measurement is made there is no assurance that it is correct for a number of reasons. The meter may have been connected wrong or it may be faulty, in which case a bad measurement results. In any case, measurement devices are not absolutely perfect and measurement error is always present. The device could be calibrated every time a measurement is to be taken which is clearly impractical when frequent measurements are required. In practice, the performance of the device is derived from the precision information given by the manufacturer. He will test a sample of instruments manufactured, to verify that they conform to design goals. His tests will produce the plot of Fig. 1, where μ is the known mean value of the measured test quantity. The curve most likely will represent a Gaussian or normal distribution, where most of the devices will produce measurements close to μ, some will be further out and few will measure even further. The measurements themselves are random events. The average or expected value of all measurements will tend towards μ the more measurements in the sample. Clearly it is important to know the

spread of these measurements. The deviations of each measurement from μ provides this information. A quantity σ can be computed as follows,

$$\sigma = \sqrt{\text{Average}[(\text{Measurement} - \mu)^2]}$$

For a normal curve 68% of measurements fall within $\pm\sigma$ of μ, 95% within $\pm 2\sigma$ and 99% within $\pm 3\sigma$. The term σ is called the standard deviation, and its square the variance.

Considering our original measurement of a transmission line flow, the manufacturer test performance of the instrument actually used is not known as only samples were tested. All that is known is that the device used belongs to a class of instruments whose sample test results produced a certain standard deviation σ. The conclusion can be reached that with a 99% certainty, the true value of the measurement lies within $\pm 3\sigma$ of the measured value.

To improve on the above result, repetition of the measurement is necessary. Assuming the device has a digital readout, the human error of reading a scale is eliminated. For the repetition to be significant, different devices should be used because the precision of a single device does not change from one moment to the next. Each new measurement provides additional or redundant information. Their average will be much closer to the true value the more measurements are taken. It can happen that some of these measurements are indeed closer to the true value than the average. This is not important. What is of importance is that there is a greater certainty that the average is closer to the true value than each individual measurement. In fact, the standard deviation of the average is less than σ. It is thus seen that redundancy will tend to reduce or filter normal measurement error.

It has been tacitly assumed that during the repetition of the measurements the flow in the line has not changed. In other

words, the measurements are simultaneous in the sense that they represent the same system conditions.

From a practical point of view different devices will not be used to measure each flow in each line when a power system is to be monitored. The concept of redundancy and filtering enters in a different, although related way. Consider that a calculation will be performed with the measurement, as for example, to calculate a bus voltage. In that case the equivalent circuit impedances of the measured line are needed. Alternatively to saying that a calculation is performed with the flow measurement, it can be said that a bus voltage "measurement" is made with a "device" that performs the physical flow measurement and the computation. From this point of view, redundancy in the "voltage measurement" can be obtained by measuring the flow in several lines. It is clearly seen that information on impedance and on system configuration must be used. This information is called the model of the system. Note that the model only includes system information used in the calculations. The measurements on the various lines must be taken simultaneously for the computations to be valid.

In the power system monitoring application the "voltage measurement" concept is used. While more than one flow measurement is used redundantly for each voltage, it is also the case that each flow measurement is used in the computation of more than one voltage. In fact, each flow measurement is used in the computation of the voltages at both ends of the line. In a power system, the nodal voltages are the minimum number of variables that relate to all measurements. Such a set of variables are called state variables wherefrom the term state estimation is derived. When all the equations relating the physical measurements to the voltages through the model are written down, a redundant set of equations result, with an equation for each measurement and each equation relating the two line-end voltages. In the simultaneous solution of these equations to compute the voltages, all measurements influence

the value of each voltage. The error component of each measurement is filtered out in the averaging process implied in the solution of the redundant equations. In the same way that the average of a repeated measurement differs from, and has greater certainty of being closer to the true value than each component measurement, it is also the case that the line flows computed from the resulting bus voltages have a greater certainty of being closer to the true line flows than the isolated flow measurement taken on each line. In this statement it is being inferred that flow measurements are used to calculate voltages, which in turn are used to calculate individaul line flows. This last calculation is a direct one.

Why does the solution of the redundant equations imply an averaging process? As was said, each equation relates a single measurement to voltages through the model. If a trial by error solution is followed, a set of voltages is guessed from which, in each equation, a flow is computed and compared to the measured value. From this comparison a better guess of voltages could be made. It is known before hand that a voltage set that exactly matches all measurements does not exist because each measurement contains some independent error component. If there are n voltages, a perfect match could be made with n flows at the expense of large deviations for the rest of the redundant flows. Clearly this is not desirable. The solution sought is the one that averages out the deviations among all flows. Mathematically, one possible solution is the one that produces the minimum sum of squares of deviations between measured and calculated flows. Thus the name <u>least squares</u> for this solution.

It is frequently the case that measuring devices used are not all of the same quality. The manufacturer stated standard deviation will vary accordingly. In a power system current and potential transformers, meters and analog to digital converters are used to get a single measurement. Each of these components introduce error into the measurement. The case is

that in power system monitoring the error component of measurements are not all the same. This information can be taken into account by tending to believe more, or to put greater weight, to the more accurate measurements. This can be done by making the averaging process a weighted averaged one. In the sum of squares to be minimized, different coefficients or weights can be given to each term to reflect the various degrees of certainty in each measurement. The least squares solution will then tend to produce voltages whose computed flows are closer to those taken with more accurate devices. One way of choosing weights is to make them be inversely proportional to the variances of the measurements. This makes each individual term to be as follows,

$$\left[\frac{\text{Measured Value} - \text{Calculated Value}}{\sigma}\right]^2$$

The calculated value is an approximation to the average, mean or true value. For example, consider two measurements taken with devices of unequal precision. A least squares solution without weights would treat each equally and thus would consider the less accurate measurement to be as good as the more accurate one. The weighted least squares would treat them unequally in favor of the more accurate one. Notice that in the weighted least squares case for both measurements the above term would be equal to 1 if each measurement had fallen at its own 1σ value.

The above term can also be seen as a transformation. If for each point of Fig. 1 the mean value is substracted and the difference is divided by the standard deviation, a curve results having zero mean and a variance of one. If the Fig. 1 distribution is a normal one, a unit normal distribution is obtained. In the unit normal, a comparison for measurements taken with devices with different accuracies can be made with

the same scale.

Because the calculated value is an approximation to the mean value, the weighted sum of squares is also a point in a curve of sums of unit normals. Such a curve is called a chi-squared distribution. It is interesting to note that in a problem with no redundancy the weighted least squares solution exactly matches the measurements independently of the weights. Furthermore the sum of squares would be zero. There is no averaging effect taking place, the flow measurements producing voltages that in turn produce calculated flows equal to the measured ones. Only when there is redundancy does the averaging or filtering effect takes place and the sum of squares is positive. The redundancy, that is the excess of measurements over voltages, is given the name degrees of freedom. Keeping the concept of degrees of freedom in mind, the statement made previously that the unit normal has a σ equal to one is analyzed in relation to the chi-squared distribution. From the definition of the standard deviation, it can be seen that the unit normal squared has an expected value equal to 1. The following intuitive reasoning will show that the expected value of a chi-square distribution is equal to its degrees of freedom. With n voltages or state variables, it has been stated that the weighted sum of squares or chi-squared distribution has a value of zero, independent of the measurement values, when there are n measurements. When the degrees of freedom is one, the first contribution to the sum of squares is made. The contribution of that single unit normal squared term has an average value of 1. Thus the average of the term sum of squares is also 1 and it can be said that the expected value of a chi-squared distribution with 1 degree of freedom is equal to 1. Every additional term increases both the degrees of freedom and the average by 1. There is thus a family of chi-squared distributions, one for each degree of freedom, as shown in Fig. 2.

The chi-square distribution plays an important role in the

detection and identification of bad measurements, that is measurements that grossly differ from the true value by many σ values of the device used. Consider the case where one such bad measurement exist. That measurement by definition does not belong to the 3σ range of the normal distribution of the device:

Knowing the degrees of freedom of the particular problems at hand, a chi-square distribution is fixed. The value of the chi-square (the horizontal axis on Fig. 2) that includes 99% of the area below the curve is theoretically the maximum value or boundary attainable by the weighted sum of squares if all measurements were within 3σ of their particular standard variation. The presence of the bad measurement will produce calculated values that for some terms will differ from the measured value by more than their individual 3σ, thus making the sum of squares exceed the boundary for a 99% confidence. Notice that it has not been said that sum of squares term corresponding to the bad measurement is necessarily large. Indeed, it may be that the calculated and measured values at the bad measurement location are close. However, this is at the expense of other terms becoming large. It is this characteristic that makes the chi-squared bad data detection test so powerful.

Having detected the presence of bad data, the problem remains of identifying which is the bad measurement. It has been mentioned that the expected value of a chi-squared distribution is equal to its degrees of freedom. Consider that the calculated value of the sum of squares is taken as an approximation to the expected value. Then the ratio of the sum of squares to the degrees of freedom is a measure of how much the sum of squares differed from the expected value. A ratio somewhat near to one implies that the individual standard deviations used with each measurement corresponded to the normal range of its measured minus calculated values. A large ratio suggests that at least for one measurement this

correspondence did not occur. If all measurement variances are increased by this ratio, then correspondence will have been forced to exist.

The term measured minus calculated value is called a measurement residual. That residual is itself also a random variable. If different complete sets of measurement devices were used to gather sets of measurements, and each set was used to solve the redundant equations by minimizing the weighted sum of squares, a set of residuals corresponding to each measurement would be obtained. Each of these residuals describe a normal distribution with an expected value of zero. The standard deviation of this distribution can be computed from theoretical considerations without the need of obtaining sets of measurements, extracting this information from the redundancy present in the measurement system. The following transformation from normal to unit normal can be made:

$$\frac{\text{Measurement Residual} - \text{Zero}}{\text{Standard Deviation of Measurement Residual}}$$

The zero is the mean value of the residual. In the same way that measurement standard deviations were all increased by the ratio obtained from the chi-squared distribution, the measurement residual standard deviation also becomes increased by this same factor. Theoretically the effect of the factor is to increase the standard deviation of measurement residuals such that the residuals of good measurements are enclosed within the 3σ range around the mean of zero. If the effect of the factor is omitted and the results of the above transformation ranked from large to small values, the residual of the bad measurement would tend to be at the top of the list. Intuitively this can be inferred considering that the effect of a bad measurement is so strong, that the normal possible range of all the other good measurements within their individual 3σ will result in a smaller residual standard deviation for the

bad measurement than for the good ones. As in the above transformation the standard deviation appears in the denominator, it will push the bad measurement term towards the top of the list, even if its actual residual was small. With the inclusion of the factor, the whole list is scaled down such that in the unit normal transformation the good measurement terms will tend to be smaller than 3 with a 99% confidence. In practice, it happens that the factor cannot be computed exactly from the chi-squared distribution as only one point of that distribution is available and the approximation was made to consider that point to be the expected value. The error in the factor makes the transformation be inexact and instead of transforming into a unit normal it can be shown that it transforms into a very similar distribution called the Student-t distribution. In fact, for degrees of freedom larger than about 30, the unit normal and the Student-t are practically identical. The effect of the Student-t is that the 99% confidence instead of being at a level of 3 is at a number somewhat smaller than 3. The Student-t test to determine which of the transformation terms is the largest of those above the 99% confidence level, is called the bad data identification test. The importance of the factor is that it makes the bad measurement identification more clear cut. Occasionally it may happen that several measurements are closely beyond their 3σ levels such that the chi-square detection test flags the existence of bad data. The Student-t identification test finds that the largest term in its ranked list is smaller than the 99% confidence level and concludes that no clear cut bad data exists.

It was mentioned that the standard deviation of the residual distribution can be computed from theoretical considerations. Similarly, the standard deviation of the state variables, the voltages, and that of the flows themselves can be also be computed. Assuming that the calculated value is an approximation to the mean, that value plus or minus 3σ gives a range in which the true value lies with a 99% confidence,

assuming that the voltage and flow distributions are normal. These 3σ ranges are called confidence limits. A small confidence limit implies high accuracy.

THE AEP STATE ESTIMATION ALGORITHM

In this section the equations of the AEP state estimation algorithm will be summarized relating their presentation to the concepts discussed previously.

Line flows are measured at both ends of most lines. The following equation can be written:

$$\bar{S}_m = \bar{S}_c\ (\bar{E}) + \bar{\epsilon} \tag{1}$$

where S_m is a measured flow, S_c is a calculated line power flow as a function of the state variables, the complex model voltages E and ϵ is the measurement error. It is assumed that $\bar{\epsilon}$ belongs to a distribution having zero mean and a diagonal covariance matrix W^{-1}, each diagonal term being the variance of the particular error. Eqs. (1) are a redundant set of equations. Its solution is obtained by minimizing the weighted sum of squares

$$J(\bar{E}) = (\bar{S}_m - \bar{S}_c)^{*t}\ W\ (\bar{S}_m - \bar{S}_c) \tag{2}$$

Eq. (2) can be algebraically transformed into the form

$$J(\bar{E}) = (\bar{v}_m - \bar{v}_c)^{*t}\ D\ (\bar{v}_m - \bar{v}_c) \tag{3}$$

where v is the element voltage across a line, D a diagonal matrix with

$$D_j = \frac{W_j |E_j|^2}{|z|^2} \tag{4}$$

and z is the impedance of the line where measurement j has been taken. The relation between Eqns. (2) and (1) infer that the minimization of J(E) of Eq. (3) implies the solution of the following redundant set,

$$\dot{\bar{v}}_m = \bar{v}_c(\bar{E}) + \bar{\mathcal{V}} \tag{5}$$

where v_m is taken as an element voltage "measurement" computed from the corresponding physical measurement S_m, the impedance characteristics of the line and the nodal voltage E at the measured end of the line. The term v_c is the difference of line end nodal voltages E, which in terms of an incident matrix A can be written as

$$\bar{v}_m = A\,\bar{E} + \bar{\mathcal{V}} \tag{6}$$

In the above equations simultaneity of flow-measurements are assumed. It is to be noticed that the algebraic tansformation performed transforms the nonlinear problem (1) into a linear problem (6) in which the nonlinearity is buried into the "measurement"

The minimization of J(E) of Eq. (3) with respect to rectangular components of E is accomplished by solving

$$\frac{\partial J(\bar{E})}{\partial E} = A^t\, D\, (\bar{v}_m - A\bar{E}) = 0 \tag{7}$$

where the term $(\bar{v}_m - A\bar{E})$ is an element voltage residual vector and v_m has been assumed as constant and independent of E. As this is not strictly true, the solution of Eq. (7),

$$\bar{E} = (A^t\, D\, A)^{-1}\, A^t\, D\, \bar{v}_m \tag{8}$$

has to be iteratively repeated after v_m has been updated for the new values obtained for E. One bus voltage is assumed known and used as reference.

In Eq. (8) no actual matrix inversion takes place. Triangular Gaussian factorization techniques and sparsity are used to solve these linear equations,

$$(A^t\ D\ A)\ \bar{E} = A^t\ D\ \bar{v}_m \qquad (9)$$

The importance of the algebraic transformation performed to get Eq. (3) is now apparent. The $(A^t\ D\ A)$ matrix has the properties:

1. Sparse, permitting the use of sparsity techniques.
2. Symmetrical, reducing by half storage requirements.
3. Constant, allowing the triangularization to be made once for all iterations.
4. Real, decoupling Eq. (9) when rectangular components are used and reducing storage requirements by another half.

Once convergence is obtained, then the chi-squared detection test described in the previous section can be performed to see if J(E) satisfies the inequality

$$J(\bar{E}) < \chi^2_{k,b/2} \qquad (10)$$

where $\chi^2_{k,b/2}$ is the theoretical value of a chi-squared distribution for k degrees of freedom and a probability of confidence b. If inequality (10) is not satisfied this is taken to mean that the existence of bad data has been detected. The degrees of freedom is twice the difference between the number of complex measurements and the number of complex nodal voltages to be estimated.

Once bad data has been detected it is identified by performing the bad data identification tests on each measurement j, and finding the largest one for which,

$$\frac{\bullet\,(\bar{v}_m - \bar{v}_c)_j - \text{zero}}{s\sqrt{D_j^{-1} - \text{diag}[A\ (A^t\ D\ A)^{-1}\ A^t]_j}} > t_{k,b/2} \qquad (11)$$

where $t_{k,b/2}$ is the theoretical value of a Student-t distribution for k degrees of freedom and a probability of confidence b. The term s is the factor mentioned in the previous section in relation to the property that expected value of a chi-squared distribution is its degrees of freedom,

$$s = \frac{J(\bar{E})}{k} \qquad (12)$$

The confidence limits of voltages E is given by

$$C.L.(\bar{E}) = s\sqrt{\text{diag}[(A^t\ D\ A)^{-1}]} \qquad (13)$$

and those of flows by

$$C.L.(\bar{S}) = s\sqrt{\text{diag}[K(A^t\ D\ A)^{-1}K^t]} \qquad (14)$$

where K is the Jacobian of P with respect to E for C.L. (P), or K is the Jacobian of Q for C.L. (Q).

PROGRAMMING SYSTEM STRUCTURE

Up to now, the mathematical considerations of state estimation have been mentioned. A real-time process control computer must be available in the control center to perform these calculations and its programming system developed. There must be programs which interface with incoming measurements and programs which define the problem models before any state estimation computations may be done. Furthermore, after state estimation there must be programs that communicate results to operators. In the AEP control center this is done through

CRT's and through lights on a mimic board. The AEP programming system structure is presented in Fig. 3.

The teleprocessing programs interface with incoming measurements and periodically schedule the other modules unless a breaker status has changed in which case the scheduling is done immediately. The Configurator I programs analyze breaker status to determine open lines and separated substations. The Configurator II programs determines a basic configuration consisting of energized lines whose flow measurements are available and then analyzes the connectivity of the system through that configuration to determine separated subsystems containing a reference voltage measurement. The Validator performs a pseudo-detection test by comparing the theoretical chi-squared bound to a sum of squares in which the calculated values are the most recent past results and the measured values are the new ones just read. If the test is passed, there is no need to process the new measurements as the past results are still valid. The Estimator I programs perform all the non-iterative portions of the least squares solution algorithm, specially the formation and triangularization of the $(A^t D\ A)$ matrix and the optimal ordering. The Estimator II module performs the iterative calculations described by Eq. (9) until convergence is obtained. Line flows are then computed in Flows followed by the performing of the Detector tests. In case bad data is detected, the Identifier programs are used. If there is no bad data, then Limits computes the confidence limits described in Eqs. (13) and (14) for output purposes. The Screen programs interface with the CRT's and drives them upon request entered via the keyboard. Similarly, the Board programs interface with and drives the lights of a large mimic board.

The basic interfaces between modules shown on Fig. 3 are intended to show the adaptive characteristics of the programming system, by which only those modules that are strictly required for the particular conditions at hand are scheduled. Impressive gains in efficiency are made this way and justify

the needed complexity of the programming structure and interfaces.

OTHER TOPICS AND CONCEPTS IN A MONITORING SYSTEM

Global vs. local redundancy

Redundancy has been shown to be the key concept for filtering, detection and identification. The degrees of freedom definition given is a measure of the redundancy of the whole measuring system, the global redundancy. However this global figure does not guarantee proper estimator performance everywhere in the system since there may be localized sections where the redundancy is nonexisting, as for example, in a radial line with only one flow measurement. Locally, that measurement and the voltage to be estimated at the radial end of the line balance out leaving no local redundancy. In fact, the radial end voltage will adjust itself through the solution of the equations such that the computed flow on the line exactly matches the measurement, with no filtering. Detection and identification of bad data on that measurement is impossible. Ref. 15 discusses this in more depth.

Solution process

To gain a deeper insight into the meaning of the solution process and its relation to the measurement system, Eq. (7) should be analyzed. These are the set of simultaneous equations solved, one equation per nodal voltage state variable. In words Eq. (7) says that the algebraic (A), sum of all the weighted, (D), element voltage residuals, $(\bar{v}_m - A\bar{E})$, corresponding to all flow measurements on lines connected to a bus, (A), is equal to zero. Considering the radial line with one measurement, the equation of the radial end bus has only one term, a $\pm$ D, multiplied by an element voltage residual, equated

to zero. This forces the residual to be zero and defines the radial end voltage to correspond exactly with the flow measurement. If the radial line has two flow measurements, then there are two terms in Eq. (7) and the radial end voltage is adjusted to correspond to the weighted average of the two flows. This averaging of two measurements is a filtering process. If one of these measurements is bad, it will be detected mainly due to the large flow residuals that will be obtained on both measurements after the averaging. However, identification is not possible because there is not a third piece of independent information to decide on a two out of three basis which is the good one. In student-t identification test, this case will show as two equal and largest terms on the ranked list, and the action taken is the removal of both measurements, the good with the bad one. Notice that this leaves the line without measurements making impossible the estimation of any flows on the line or the radial end bus voltage. That voltage would not appear in any of the Eqns. (7). Another conclusion that can be reached by studying the solution of Eq. (7) is that Kirchoff's current law forcing currents or flows around a node to add to zero is not used anywhere in the formulation. Kirchoff's voltage law forcing the difference in element voltages around a loop to add to zero, while not directly used, is inferred by Eq. (7) when several equations are added together. For a detailed treatment of the solution process Ref. 15 is recommended.

Measurement system

From the above remarks on local redundancy and its relation to filtering, detection and identification, it is clear that the location, number and type of measurements to be installed has to be chosen judiciously. If there are many loops in the system, flow measurements at both ends of most lines is sufficient. On a radial string flow measurents on both ends of

every line is a must for proper detection but insufficient for identification of a bad measurement without removing a good one along with it and leaving the rest of the radial string hanging in the air as far as the estimator equations are concerned. More measurements are required on the radial string. As flow measurement locations have been exhausted, this leaves two other types of measurements possible: bus voltage or bus injection measurements. Of these two, the voltage one is to be preferred as it will preserve the efficient four characteristics of the solution process mentioned in relation to Eq. (9). These voltage measurements should be processed as recommended in Ref. 12. Injection measurements reduce the efficiency of the algorithms. For a good treatment of the processing of injection measurements with the AEP algorithm Ref. 14 is suggested, complemented by Ref. 15.

Model and parameter uncertainties

For proper estimation it is necessary that the correct configuration model is obtained and the correct impedance and off-nominal tap parameters are used. An interesting property of the AEP algorithm is introduced by the algebraic transformation used between Eqns. (2) and (3). A v_m quantity is computed from the flow measurement, line impedance and tap. The resulting v_m is used as if it were a measurement. Obviously, v_m could be bad due to any of its three possible components. Thus, when a measurement is identified as bad, it may not mean that the actual flow measurement was bad. Impedance and tap errors were found this way during the final check-out tests of the AEP estimator.

Model errors are not identified precisely. Only deenergized lines that are assumed to be in service cause problems. This is usually unlikely to occur if the line is taken as open if deenergized at any end. That is, bad breaker information must be received from both sides. Even then, the Detector

flags bad data and, what is very useful, the identificator will tend to flag as bad measurements that are on lines in the same loop as the incorrectly modeled one. The effect of this is to break the loop leaving the line in question as part of a radial string where its damage is isolated to only that string. In effect, the model error has been neutralized. More on model and parameter errors and their identification can be found in Refs. 13 and 15.

Debugging and implementation strategy

For an efficient implementation of a real-time programming system as the one for monitoring a power system, it is necessary to proceed in certain ordered steps. First, certain essential programs are written and independently tested under simulated data for which results are known. Next, these programs can be interfaced and again tested under simulation. At that time limited, open loop real time testing can start slowly adding the various other modules of the programming system. Finally, full real-time, closed loop, testing can take place. This ordered approach was found very useful in implementing the AEP system. No program is really complete and no interface is final until the whole programming system is debugged. A strategy is recommended by which changes in the design of programs are not frozen at any intermediate step in the implementation process because much is learned about the performance of the programming system during all phases of implementation.

RESULTS

Fig. 4 presents actual results of the AEP state estimation monitoring system on June 4, at 10:10 a.m. A map of the monitored system consisting of approximately 60 busses and 80 lines is prepared off-line and displayed on-line on a cluster of four CRT's. Numerical results are superimposed on the map

in various modes. The one presented in Fig. 4 displays line flows (P° and Q) with a dot signalling the P value alongside the line, voltage magnitudes inside the substation box and voltage phase angles near the box. The magnitudes are in percent and the angles in degrees, with one decimal always implied. When measurements are missing or bad, and they can not be determined from other quantities, MM or BB is displayed instead of the numerical value. If a substation is separated, SS is displayed in place of the voltage magnitude and no phase angle is displayed. Nothing is displayed on lines and the busses they are connected to when their valid measurements could not be used because their corresponding lines were in an isolated subsystem with no reference voltage measurement available.

Fig. 5 presents numerical results corresponding to the same case shown on Fig. 4. These printed output results are not considered normal means of output, they are intended more for checking purposes.

MONITORING AND SYSTEM SECURITY

The enhancement of the security of operation through monitoring was discussed in the introduction of this paper. In this section some of the concepts involving contingency studies and voltage level corrections using the information available through real-time results will be discussed. A proposed contingent situation is a non-existent one and thus cannot be monitored. Through monitoring reliable precontingent system operation information is obtained. For an island system, the last state estimation results may be sufficient as a data base from which contingency analysis may be made. The data precision information derived from the confidence limits of resultant estimated quantities may be used as input to a stochastic load flow [16] contingency analysis. For interconnected systems, the vital problem of the reaction of external systems to contingencies must be considered. This problem is under study

now and preliminary results available at the time of this writing encouragingly point to a satisfactory solution. Basicly, while external system configuration and loadings are not known, the known reaction of the external system in past time as viewed from the monitored system, contains information that may make possible the proper modeling of its future reaction under contingent situations. The stochastic load flow also plays a role here.

Finally, on the question of voltage level correction, the role of optimal load flows, or operation security load flows as they can be called, together with reliable data from the monitoring system, appears to hold much promise. A precise answer to the voltage correction problem is not needed in one step since the monitoring makes it possible to follow the effect of a sequence of corrections step by step.

CONCLUSIONS

In this paper, the importance and justification of centralized computer monitoring of power systems through state estimation has been presented as viewed by the authors. Statistical concepts behind the state estimation have been introduced in intuitive terms to serve as a basis for the presentation of the AEP algorithm and programming system. Other concepts that compliment and extend the understanding of monitoring was discussed. The conclusion can be reached that the AEP experience has shown that:

1. State estimation monitoring is feasible with present technology. This has been confirmed by months of on-line real-time experience at AEP.
2. Such items as number and location of measurements, programming system structure and display systems have to be judiciously designed for the overall success of state estimation monitoring. The problem is indeed much more complex than that of the solution algorithm itself, although this

plays a central role in overall performance. Much needs to be done computationally in the related security area, although it seems clear to the authors that the approach described in this paper appears technically and practically feasible in the near future. At AEP work in this direction is in progress at this time. Through the combination of monitoring and security calculations in a centralized computer, an impressive step forward is taken in assuring a reliable operation of power systems.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the many contributions made by Oscar A. Klitin, a former member of our group, and the programming efforts of Sara Elam and Lewis Nowitz in the Teleprocessing and Display modules respectively.

REFERENCES

1. G. W. Stagg, J. F. Dopazo, O. A. Klitin and L.S. Van Slyck, "Techniques for the Real-Time Monitoring of Power System Operations" IEEE PAS-89, pp. 545-555, April, 1970.

2. J. F. Dopazo, O. A. Klitin, G. W. Stagg and L. S. Van Slyck, "State Calculation of Power Systems from Line Flow Measurements", IEEE PAS-89, pp. 1698-1708, September/October, 1970.

3. J. F. Dopazo, O. A. Klitin and L. S. Van Slyck, "State Calculation of Power Systems from Line Flow Measurement, Part II", IEEE PAS-91, pp. 145-151, January/February, 1972.

4. J. Dopazo, O. Klitin, A. Sasson, L. Van Slyck, "Real-Time Load Flow for the AEP System", Paper No. 3.3/8, 4th Power Systems Computation Conference Proceedings, Grenoble, France, September, 1972.

5. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin and A. M. Sasson, "Justification of the AEP Real-Time Load Flow Project", IEEE PAS-92, pp. 1501-1509, September/October, 1973.

6. A. M. Sasson, S. T. Ehrmann, P. Lynch and L. S. Van Slyck, "Automatic Power System Network Topology Determination", IEEE PAS-92, pp. 610-618, March/April, 1973.

7. A. M. Sasson, S. T. Ehrmann, "Topological Considerations for Real-Time Monitoring of Power Transmission Networks", Proceedings of the IEEE International Symposium on Circuit Theory, Toronto, April, 1973.

8. L. S. Van Slyck and J. F. Dopazo, "Conventional Load Flow Not Suited for Real-Time Power System Monitoring", Proceedings of the 8th IEEE PICA Conference, June, 1973.

9. J. F. Dopazo, O. A. Klitin and A. M. Sasson "State Estimation for Power System: Detection and Identification of Gross Measurement Errors", Proceedings of the IEEE PICA Conference, June, 1973.

10. J. F. Dopazo and A. M. Sasson, "AEP Real-Time Monitoring Computer System", Symposium on Implementation of Real-Time Power System Control by Digital Computer, Imperial College of Science and Technology, London, September, 1973.

11. J. F. Dopazo and A. M. Sasson, "Least Square Estimation Applied to Power Systems", Seminar on Computer Methods of Power System Analysis and Control, Bucharest, Romania, June, 1974.

12. H. L. Fuller and T. A. Hughes, "State Estimation for Power Systems with Mixed Measurements", IEEE Paper No. C74 363-8, Summer Power Meeting, Anaheim, Cal., 1974.

13. F. Aboytes and B. J. Cory, "Identification of Measurement, Parameter and Configuration Errors in Static State Estimation", Proceedings of the 9th IEEE PICA Conference, June, 1975.

14. S. Ionescu and L. Radu, "State Estimation from Line Flow Measurements and Zero Injection Pseudomeasurements", Seminar on Computer Methods of Power System Analysis and Control, Bucharest, Romania, June, 1974.

15. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin, A. M. Sasson and L. S. Van Slyck, "Implementation of the AEP Real-Time Monitoring System", to be submitted for the IEEE Winter Power Meeting, 1976.

16. J. F. Dopazo, O. A. Klitin and A. M. Sasson, "Stochastic Load Flows", IEEE PAS-94, pp. 299-309, March/April 1975.

17. J. F. Dopazo, S. T. Ehrmann, A. F. Gabrielle, A. M. Sasson and L. S. Van Slyck, "The AEP Real-Time Monitoring and Control Computer System", to be submitted for the IEEE Winter Power Meeting, 1976.

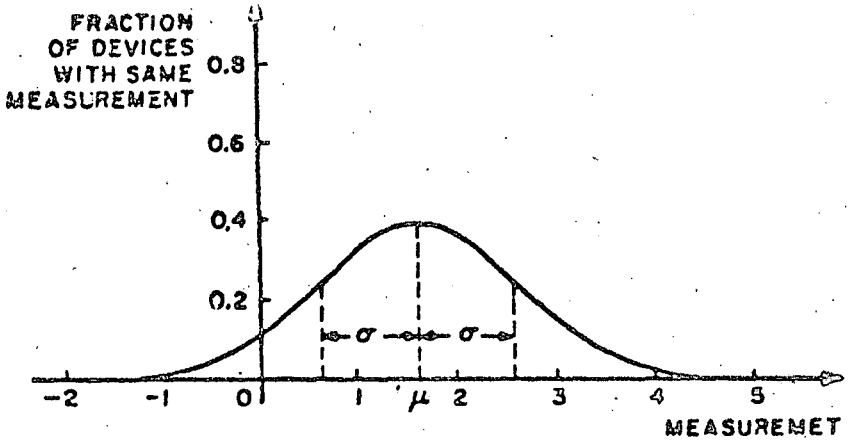


FIGURE 1
AN ACCURACY TEST ON A SAMPLE OF DEVICES

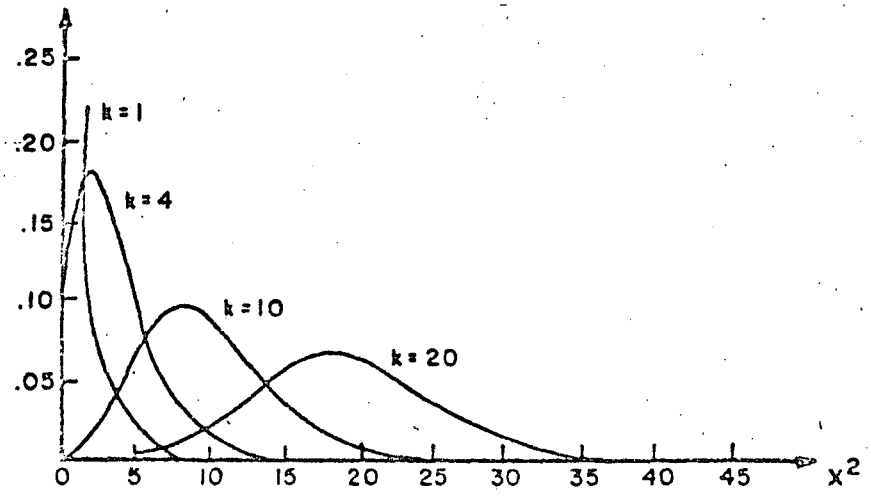


FIGURE 2. CHI-SQUARED DISTRIBUTION

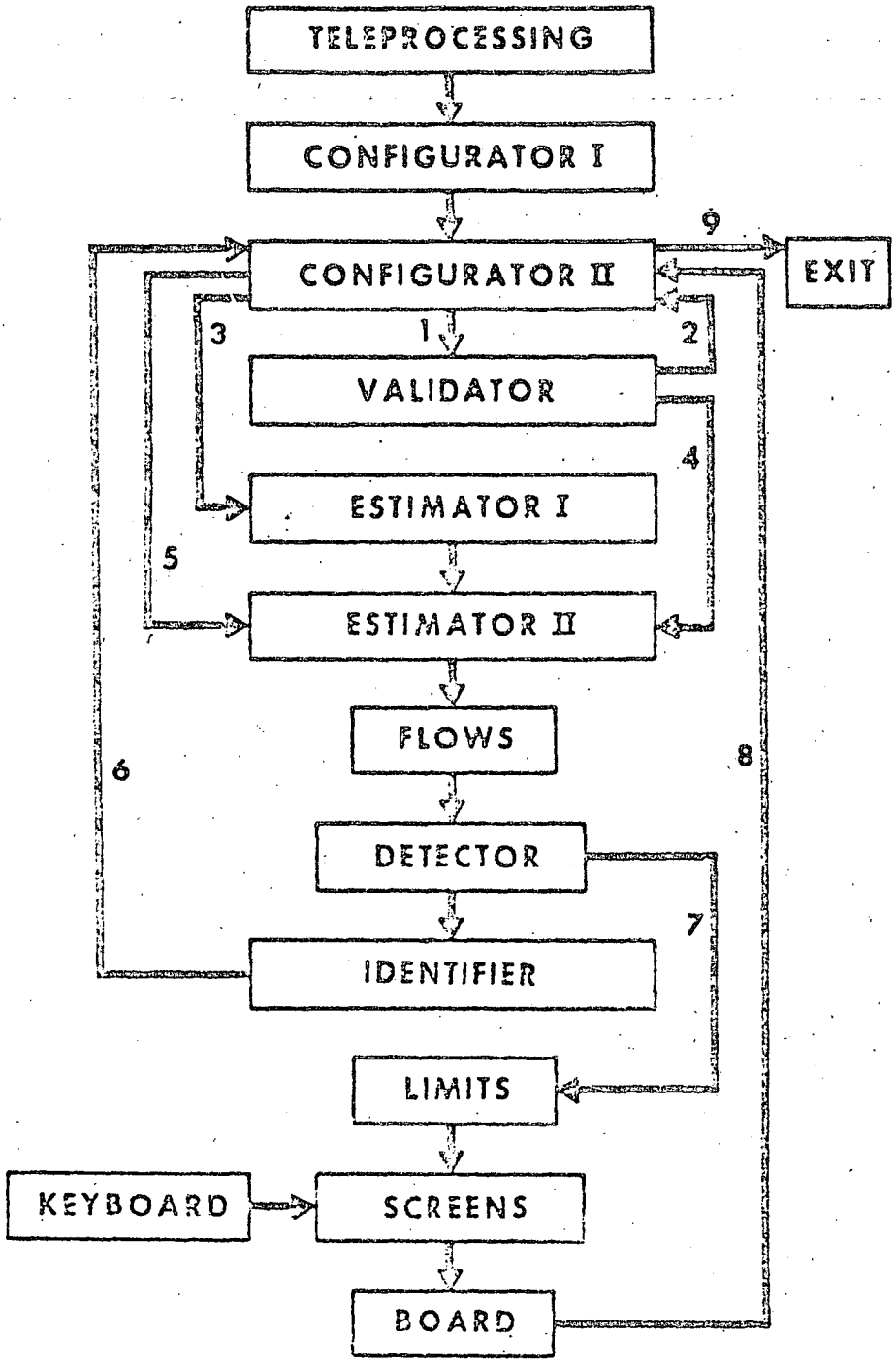


FIGURE 3. PROGRAMING SYSTEM

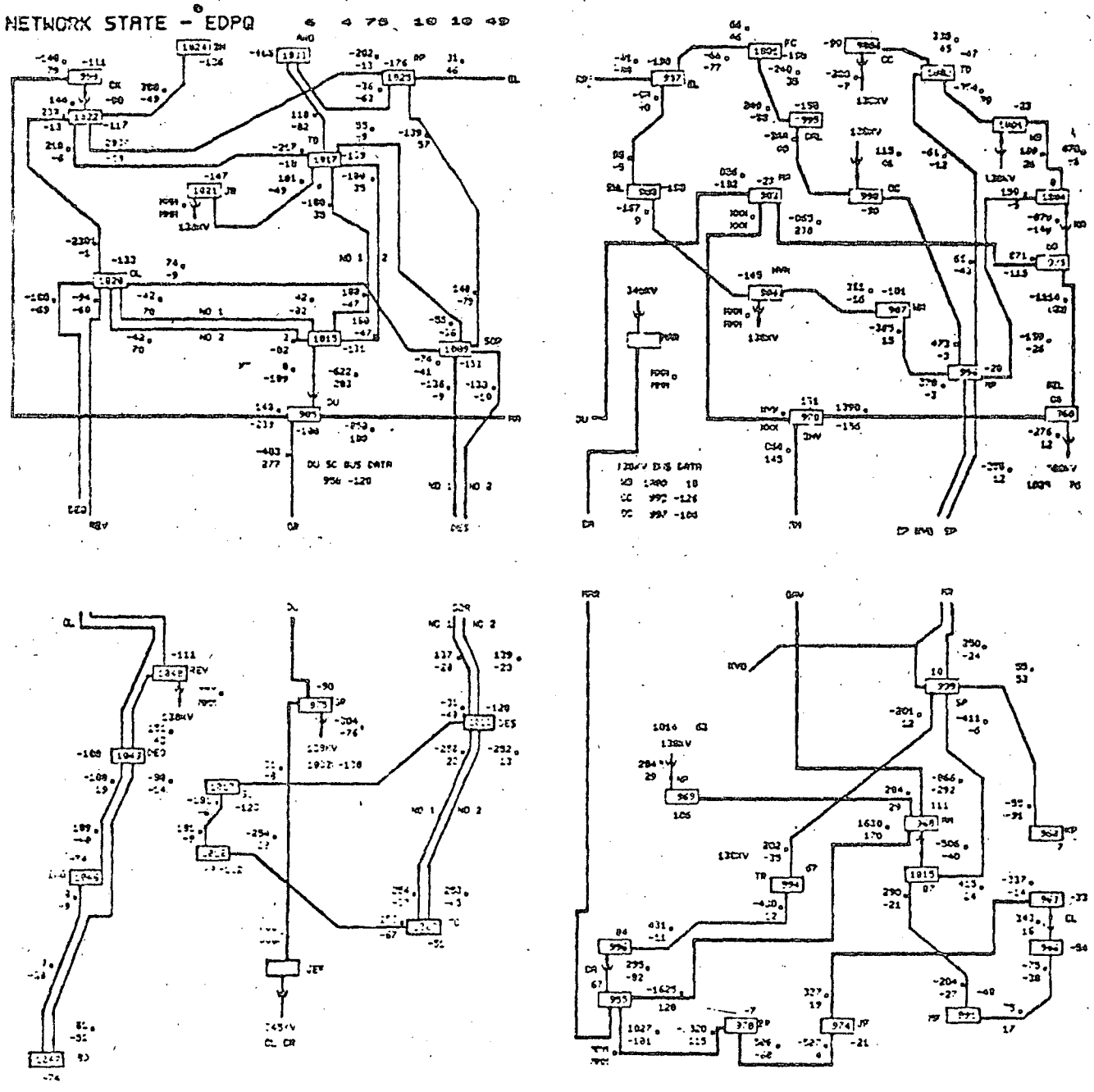


FIGURE 4. STATE ESTIMATION RESULTS ON CLUSTER OF FOUR CRT'S

BUS	VOLTAGE		CONFIDENCE LIMITS	
	MAG	ANGLE	MAG	ANGLE
2	96.7	0.72	0.23	0.13
32	102.1	-14.69	0.34	0.19
50	100.0	1.78	1.01	0.58
52	103.4	7.58	0.17	0.10
53	99.6	-10.45	0.22	0.12
54	99.2	-12.57	0.13	0.07
56	95.7	-12.77	0.27	0.15
57	103.2	-13.01	0.34	0.19
61	102.3	-13.55	0.29	0.16
62	101.3	6.32	0.83	0.47
1	99.5	8.44	0.59	0.34
3	99.4	6.74	0.59	0.33
6	96.9	-0.71	0.50	0.28
7	97.3	-2.10	0.50	0.29
8	99.0	-4.79	0.59	0.33
9	98.6	-5.40	0.58	0.33
15	97.7	13.05	0.40	0.23
16	98.2	-2.34	0.39	0.22
20	98.0	-9.91	0.42	0.24
21	101.7	-11.16	0.69	0.39
23	99.4	-15.49	0.64	0.36
24	100.1	-19.03	0.63	0.36
26	98.8	-18.97	0.64	0.37
27	98.4	-14.51	0.79	0.45
28	98.7	-10.06	0.47	0.27
34	98.9	-11.09	0.37	0.21
38	97.4	-9.02	0.23	0.13
40	104.2	-7.35	0.36	0.20
41	104.5	-7.36	0.25	0.14
43	104.0	-11.10	0.35	0.20
45	100.6	-5.10	0.31	0.18
46	96.2	-3.31	0.52	0.29
47	101.6	-12.28	0.46	0.26
49	96.8	10.62	0.09	0.05
55	98.8	-12.80	0.02	0.01
60	103.3	-16.52	1.04	0.59
12	97.1	5.01	0.30	0.17
13	95.4	6.71	0.43	0.24
14	101.4	8.69	0.40	0.23
18	100.1	-2.27	0.10	0.06
19	100.1	-4.65	0.20	0.11
22	99.0	-8.97	0.42	0.24
25	98.7	-19.76	0.63	0.36
36	101.4	-13.05	0.68	0.39
37	98.4	-10.79	0.38	0.22
44	101.1	-12.76	0.16	0.09
48	97.9	8.36	0.36	0.20
4	96.7	11.13	0.40	0.23
5	99.8	1.83	0.37	0.21
29	102.4	-17.63	0.59	0.34
30	100.8	-15.28	0.63	0.36
42	104.8	-10.01	0.35	0.20
11	99.6	-2.77	0.28	0.16
31	101.7	-13.93	0.65	0.37
33	102.1	-11.67	0.17	0.10
35	102.0	-13.25	0.71	0.40

FIGURE 5a. NUMERICAL RESULTS CORRESPONDING TO CASE OF FIG. 4

AREA	BUSSES		NORMALIZED RESIDUAL	NORMALIZED RESIDUAL X100	UNNORMALIZED RESIDUAL	UNNORMALIZED RESIDUAL MW	CONFIDENCE LIMITS	CONFIDENCE LIMITS MW
2	13	6					21.8	21.8
3	13	4					37.0	37.7
4	1	13					60.2	59.5
5	1	3					11.3	11.2
6	2	5	-0.3	17.6	-0.0	0.9	11.1	10.9
7	3	1	-0.0	0.0	-0.0	0.0		
8	3	5	-76.7	-173.5	-3.2	-7.7	10.9	10.5
9	4	49	-0.0	-0.0	-0.0	-0.0		
10	4	13	-29.3	-11.2	-4.4	-1.7	37.3	37.3
11	4	15	-73.8	-100.2	-5.1	-7.0	27.1	26.9
12	14	4	-384.6	-163.9	-23.0	-9.9	34.6	34.9
13	14	8	134.0	39.1	7.4	2.1	6.0	6.0
14	14	5	59.0	1.9	2.6	0.0	10.2	10.2
15	5	3	-91.2	-171.9	-3.6	-6.8	10.8	10.6
16	5	14	-192.8	-58.1	-8.3	-2.4	10.0	10.1
17	5	2	-3.3	87.0	-0.1	3.2	11.1	11.4
18	5	11	150.6	84.3	6.5	3.7	9.9	9.7
19	6	7	-1.9	-51.4	-0.1	-4.6	22.6	22.3
20	6	13	47.7	7.6	4.4	0.6	21.5	22.3
21	7	44	-27.8	-8.9	-2.5	-0.8	22.3	22.3
22	7	6	35.2	-46.3	3.2	-4.2	22.3	22.4
23	8	9					10.4	10.4
24	8	14					5.8	5.8
25	46	7	22.9	-2.7	2.1	-0.2	22.2	22.2
26	9	8	-28.8	-1.1	-1.2	-0.0	10.4	10.3
27	46	9	-159.7	-24.1	-7.7	-1.1	30.0	30.0
28	25	29	0.4	-82.5	0.0	-2.3	7.0	6.6
29	25	26	1.6	-4.8	0.0	-0.2	10.7	10.7
30	25	24	35.7	22.1	1.4	0.9	10.6	10.4
31	26	27	-0.0	-0.0	-0.0	-0.0		
32	26	25	1.6	-4.7	0.0	-0.2	10.7	10.7
35	16	12					20.8	22.2
36	16	37					27.0	26.3
38	28	27					36.2	36.1
39	24	11					10.8	11.1
40	11	17	111.2	59.4	7.3	3.9	9.1	9.1
41	11	5	-107.6	39.7	-5.0	1.8	9.8	9.8
42	11	28	0.0	0.0	0.0	0.0		
43	11	22	-74.1	-311.3	-1.2	-5.1	15.0	15.0
44	11	19	79.2	51.8	4.3	2.8	6.7	6.5
45	22	53	0.0	-0.0	0.0	-0.0		
46	17	12	294.8	245.4	23.1	19.2	31.6	31.6
47	12	48	271.2	91.1	22.3	7.4	49.8	50.0
48	12	16	4.3	72.1	0.4	6.9	21.1	20.9
49	17	18	-30.6	-24.7	-1.0	-0.8	12.3	12.1
50	17	11	-101.7	-69.7	-4.6	-3.1	9.2	9.2
51	19	20	-0.0	0.0	-0.0	0.0		
52	19	18	33.7	21.6	1.5	1.0	19.9	20.1
53	19	11	-50.8	-38.4	-2.7	-2.0	6.6	6.6
54	50	18	0.0	0.0	0.0	0.0		
55	52	48	0.0	-0.0	0.0	-0.0		
56	54	20	0.0	0.0	0.0	0.0		
57	23	22	136.3	293.7	8.7	17.9	20.0	20.6
58	23	24	-14.5	-59.3	-1.2	-5.1	13.1	12.8
59	24	23	67.8	96.8	2.2	3.1	12.9	13.1
60	24	25	-18.2	-72.7	-0.8	-3.1	10.6	10.7
61	15	4	12.2	-70.7	2.2	-12.7	27.2	27.3
62	15	48	-284.1	-131.8	-23.7	-11.9	49.7	48.9
64	49	62	0.0	0.0	0.0	0.0		
65	33	61	0.0	0.0	0.0	0.0		
66	34	37	-10.8	-66.4	-1.6	-9.7	37.5	37.6
67	34	33	70.7	221.7	5.7	19.0	50.0	49.4
68	33	31	-46.6	27.5	-4.4	2.6	36.1	36.1
69	33	35	4.1	7.7	0.9	1.6	62.2	62.1
70	33	29	-43.4	-112.0	-2.6	-6.7	11.8	11.7
71	31	32	0.0	0.0	0.0	0.0		
72	31	60	-54.6	-21.2	-2.8	-1.1	22.2	22.2
73	31	30	-3.4	2.7	-0.1	0.1	6.3	6.2
74	31	36	-25.3	-7.4	-1.5	-0.4	10.9	10.7
75	31	36	-25.3	-7.4	-1.5	-0.4	10.9	10.7
76	31	33	-142.2	-48.1	-14.6	-4.5	35.4	35.4
78	56	55	0.0	-0.0	0.0	-0.0		
79	37	16	-41.6	-120.9	-7.5	-21.8	26.6	27.2
80	37	38	-0.0	-0.0	-0.0	-0.0		
81	37	34	10.2	-6.6	1.5	-1.0	37.6	37.1
82	36	55	-0.0	-0.0	-0.0	-0.0		
83	36	35	11.2	2.2	0.7	0.1	10.0	10.3
84	36	35	11.3	2.2	0.7	0.1	10.0	10.3
85	36	31	-47.6	-6.6	-2.9	-0.4	10.9	10.5
86	36	31	-47.6	-40.7	-2.9	-2.5	10.9	10.5
87	35	33	4.3	-7.9	0.9	-1.7	61.9	62.1
88	35	30	68.3	83.3	2.0	2.5	5.9	5.8
89	35	36	-7.0	-10.8	-0.4	-0.6	10.0	10.4
90	35	36	-7.0	-10.8	-0.4	-0.6	10.0	10.4
91	35	43	-12.6	8.6	-0.6	0.4	17.0	16.4
92	35	42	23.0	-16.6	0.8	-0.5	14.1	13.3
93	29	33	49.7	-70.8	3.0	-4.2	11.6	11.6
94	29	60	54.4	17.4	4.4	1.4	15.3	15.3
95	29	30	-51.4	-57.7	-1.4	-1.5	7.5	7.4
96	29	25	91.4	70.3	2.4	2.0	7.0	7.1
97	40	41	7.7	-291.2	0.3	-13.2	10.3	10.2
98	40	42	364.7	74.0	11.5	2.3	6.5	6.3
99	41	42	287.1	-128.5	8.4	-3.5	17.4	17.2
100	41	40	-46.2	-142.8	-2.0	-6.4	10.3	10.3
101	57	38	-0.0	-0.0	-0.0	-0.0		
102	30	44					20.2	20.2
103	30	44					19.9	20.0
104	30	35					5.9	5.9
105	30	31					6.2	6.2
106	30	29					7.6	7.2
107	44	47	-0.0	-0.0	-0.0	-0.0		
108	44	45	20.8	-4.6	0.6	-0.1	5.1	5.1
109	44	45	9.7	-83.3	0.3	-2.7	5.1	5.1
110	44	30	-5.1	-5.9	-0.3	-0.4	20.4	20.4
111	44	30	5.3	6.0	0.3	0.4	20.1	20.1
112	42	43	11.2	-6.3	0.2	-0.1	11.5	11.5
113	42	40	415.5	-113.8	12.1	-3.3	6.5	6.4
114	42	41					17.2	17.4
116	21	45					14.6	15.2
117	21	47					76.5	76.3
118	47	21					76.3	76.3
119	47	44					17.0	16.9
121	45	21	0.0	0.0	0.0	0.0		
122	45	44	115.1	144.3	3.9	4.8	5.3	4.9
123	45	46	-62.3	-280.5	-2.1	-9.4	5.2	4.9

FIGURE 5b. NUMERICAL RESULTS CORRESPONDING TO CASE OF FIG. 4

IEEE Transactions on Power Apparatus and Systems, Vol. PAS-95, no. 5, September/October 1976

THE AEP REAL-TIME MONITORING AND CONTROL COMPUTER SYSTEM

by

J. F. DOPAZO, S. T. EHRMANN, A. F. GABRIELLE, A. M. SASSON AND L. S. VAN SLYCK

AMERICAN ELECTRIC POWER SERVICE CORPORATION

NEW YORK, N. Y. 10004

ABSTRACT

This paper presents design criteria and performance results of the AEP real-time monitoring and control computer system. Included is a description of the computer system hardware and the associated application software. The various display modes are described. The paper also includes real-time results and computer system availability performance. Immediate future developments are outlined.

INTRODUCTION

In 1969, management at American Electric Power approved a new real-time control computer project for its energy control and dispatching center in Canton, Ohio. The objectives of the new system were to monitor the EHV transmission network to provide system operators with reliable, real-time information about the state of the network and the status of its components. In addition, the system would perform the economic dispatch and logging functions then operating on a second generation computer, (1) which was installed in 1964.

All the real-time functions have been implemented on one moderate size process control computer. This paper will describe the hardware components and software design of the computer system. It will discuss the display system-operator interface and present cost figures and results. Future developments regarding the solution of the system security problem will be mentioned.

SYSTEM DESIGN OBJECTIVES

Overall design objectives for the new control system were:

1. Limiting the scope of the project to real-time functions which include:
   - creation of a real-time data base.
   - the economic dispatch of power plants.
   - the monitoring of the transmission network by computational techniques that would produce results of assured accuracy and reliability for decision making purposes.
   - the acquisition of data for regulatory reporting, intercompany billing, and system operator records.
2. Utilizing a single process control computer for the real-time functions and interconnecting it to a general purpose host computer for the off-line functions.
3. Creating a modular design that would permit handling of future problems, such as contingency analysis and system security, with tomorrow's solutions as well as tomorrow's hardware costs.
4. Developing a staff of engineering personnel to handle the interactive process of problem definition, mathematical formulation and the selection of solution techniques. This staff is essential to the continuing support and solution of power system operation problems as viewed from the real-time laboratory of the control center.

Based on the design objectives, research was initiated in 1969 pertaining to the solution of the real-time generation/transmission operating reliability problem. The implementation of this system embraced analysis, design, simulation, programming, and testing and was carried out during the 1970-1975 period by a staff that ranged between 4 to 7, depending upon the project stage.

HARDWARE

CENTRAL COMPUTER AND PERIPHERALS

The AEP control computer is a dedicated, real-time system which provides the functions of state estimation of the EHV network, study-mode and control-mode economic dispatch, and megawatt hour logging through a variety of display interfaces to system operators. To provide all these functions, the computer required a number of input-output interfaces such as process I/O terminals to interface with existing analog Load Frequency control equipment; external interrupt structure to interface with modern display devices such as CRT's, CRT keyboards, and a Dynamic Wall Display Board; and binary synchronous communication adapters to interface with other computers at remote locations. Secondly, a high degree of reliability was necessary because a second computer to provide a back-up function was not planned. Finally, a simple-to-use, proven, and relatively maintenance-free operating system was required. Figure 1 presents the hardware interfaces required for the control computer.

Five different computers were evaluated and an IBM 1800 with a 64 thousand word (16 bit) core, utilizing the MPX real-time multiprogramming executive was selected as meeting all the above criteria. Two floating point processors complete with software support for FORTRAN or Assembler Language programs were added to improve calculation speed. Additional hardware includes 1.5 Million words of disk storage, a line printer, a card reader/punch, 3 console typewriters, and process I/O such as analog inputs, digital inputs and outputs, and external interrupt structure.

The average monthly availability of the system for the 19 month period between January 1974 and July 1975 was above 99 per cent. The system has been on site since January 1972 and required only one complete operating system generation since then.

DISPLAY SYSTEM

There are two basic display devices used at the control center: a cluster of 4 CRT's and a Dynamic Wall Display Board. The CRT's and Keyboards were supplied by Data Disc, Inc. and interfaced through external process interrupts and digital output terminations. A Gould electrostatic printer is connected to the CRT system for hard copy and can reproduce any display diagram. The Display

Paper F 76 086-9, recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE PES Winter Meeting and Tesla Symposium, New York, N.Y., January 25-30, 1976. Manuscript submitted September 2, 1975; made available for printing October 31, 1975.

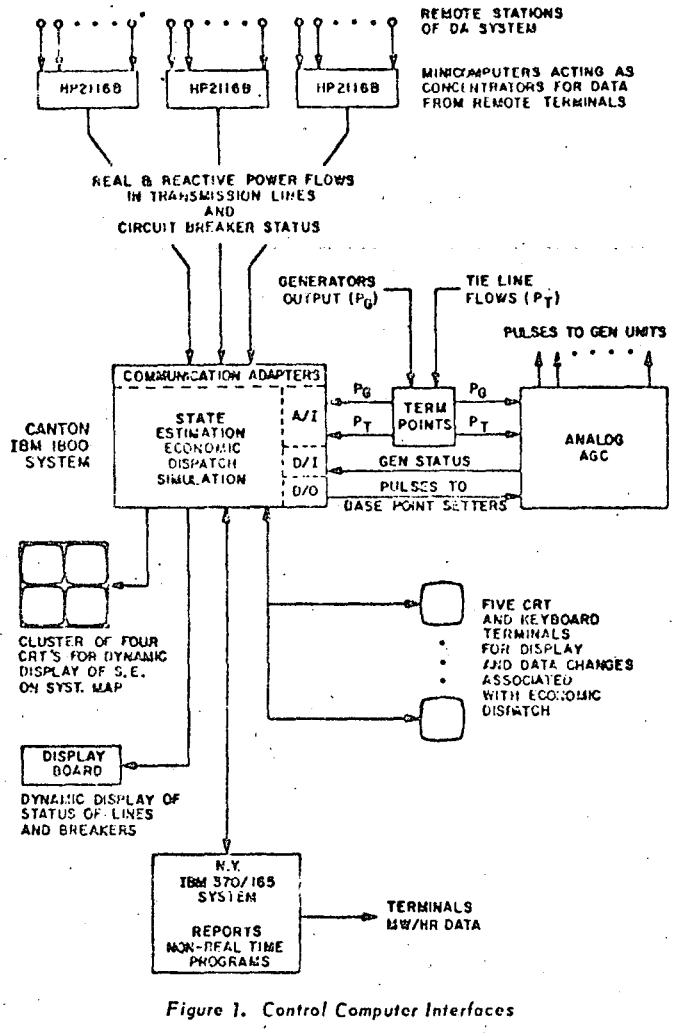


Figure 1. Control Computer Interfaces

Board System was supplied by Mavell and Quindar and is interfaced through external process interrupts and a digital output register.

A map of the EHV transmission system largely following geographical lines is impressed on the 4 CRT's. For each measured line end, one of the following display quantities may appear:

- numerical real and reactive computed line flows.
- numerical current and MVA computed values.
- xxx indicating a line is open at that end.
- (blank) indicating that flow calculations for the line were not possible even though the measured flow was received in the computer because the line was either electrically or mathematically disconnected from a path to a reference bus.
- mmm indicating that the measurements were missing.
- bbb indicating that the measurement has been identified as bad.

In the case of missing or bad data, it is usually possible to calculate numerical flow values from the voltages at the terminals of the line. These are then displayed instead of MMM or BBB.

For each electrical node or substation, the voltage magnitude in per cent and angle in degrees are displayed. There are three possible display modes for each substation:

- numerical values of voltage magnitude and angle.
- (blank) indicating that the node is electrically or mathematically disconnected from the voltage reference and numerical calculations were impossible.
- sss indicating that the node has separated into two or more nodes as a result of circuit breaker operations.

The cluster of 4 CRT's is updated only on request by the operator via a keyboard, allowing him to study a static map for any length of time. He has the ability to obtain a hard copy of the screen contents on a Gould electrostatic printer via another keyboard request. Thus, there is logic to insure that the contents of all screens refer to the same state estimation case. Figure 2 presents a hard copy of the CRT cluster contents for conditions derived from 123 measurements on June 4, 1975 at 10 hours, 19 minutes, and 47 seconds. All line flows are adjacent to the darkened squares with the real flow above the reactive flow and a positive quantity indicating a direction "away from" the node. Voltage magnitudes (63) are within the node box and angles are near the box. Figure 3 presents additional information about the calculated line flows such as confidence limits, and the normalized and unnormalized residuals. Figure 4 presents the confidence limits for the bus voltages and angles and displays the reference bus availability and its voltage. If a station separates, its new configuration is printed. Other than the new configuration print out, information on Figures 2 and 3 is used for diagnostic purposes and is not normally available to the system operator. It is planned to have a confidence limit display in the future.

Outside of the control room, there are single CRT's that can display on request any of the four screens described above. Remote CRT displays at regional control centers can be made available.

A large Dynamic Wall Display Board is driven by the control computer. Appropriate signals are sent to activate lights on the Board to reflect the following conditions:

- open-close position of circuit breakers.
- de-energized, energized or overloaded condition of a transmission line or transformer.
- direction of power flow in a line or transformer.

A memory of the current state of the Display Board is kept in the control computer in order to drive the Board only on changes from current conditions.

The Board is driven directly from the computer without operator intervention, in contrast to the CRT displays that require an operator request for an update. The normal use of the Display System is for the operators to view the Mimic Board to obtain a broad, nonnumeric overview of system conditions. When breaker operations, line outages, overloads or change of flow directions occur, the operator can use the CRT's to study certain numerical values. Before giving clearances for planned outages, the CRT's provide the detailed numerical information required for decision making. In addition, the effect of the clearance can be followed by making a CRT request once the Board update reveals the outage has occurred.

SOFTWARE

The algorithms developed to estimate network voltages from line flow measurements, to establish network configuration and connectivity from circuit breaker and measurement status, and to detect and identify bad measurements have been previously reported. This section will discuss the software operation and design.

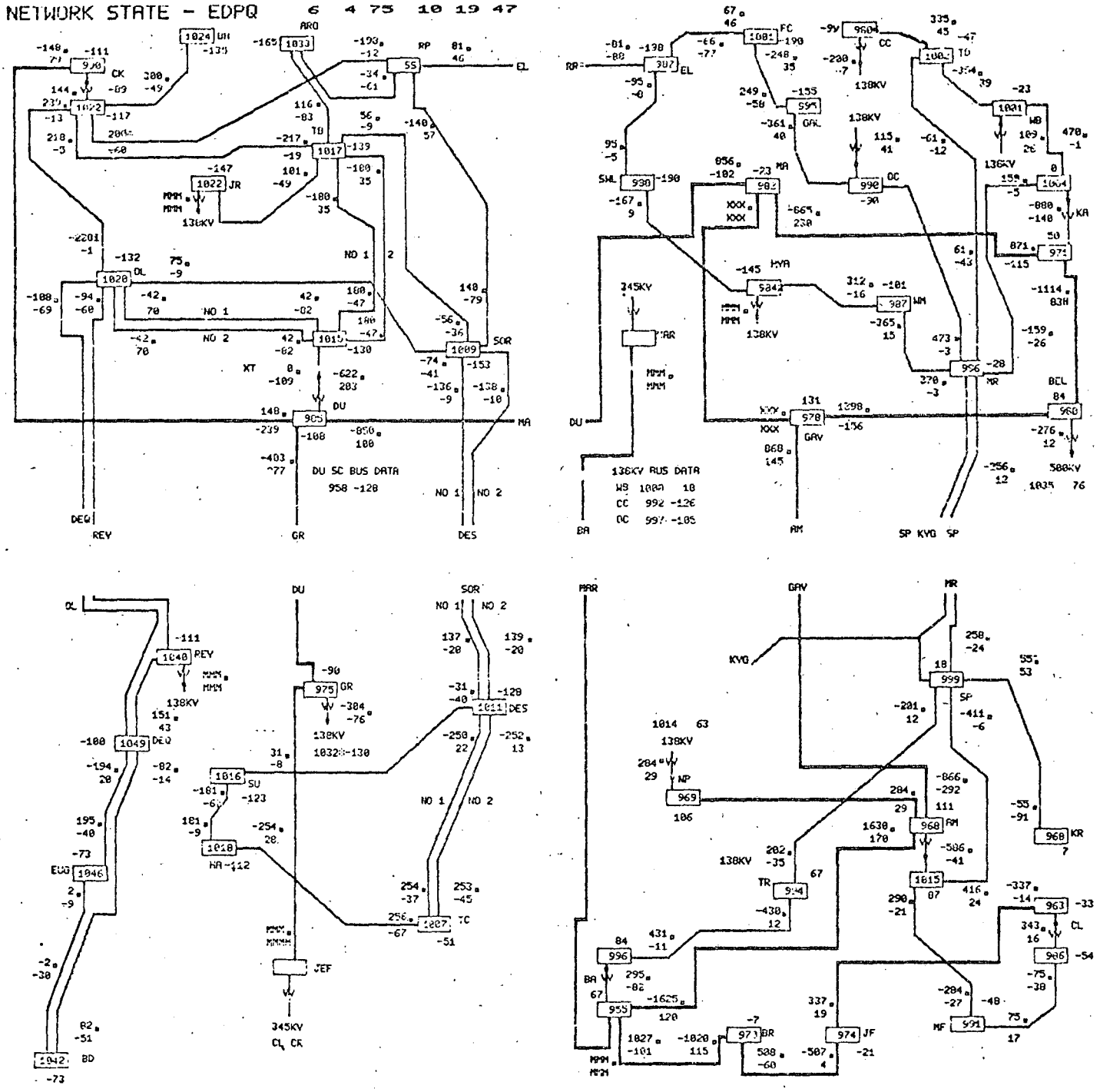


Figure 2. State Estimation Results on cluster of Four CRT's

The MPX operating system supports multiprogramming with a fixed number of partitions. It allows overlay program structures and also supports the scheduling of programs into one partition from another partition. These two features were used extensively in our system design. A functional core map of our system is presented in Figure 5.

All core sizes are expressed in 16 bit words. The Common Area is shared by all partitions and contains the following types of data: raw measurements and breaker status transmitted from the Data Acquisition System, the vector of calculated and measured line flows, all the analog inputs from the Load Frequency Control equipment, input buffers from all display devices, a change list to drive the Dynamic Wall Display Board, a number of system connectivity matrices used by the State Estimator, and many other quantities. The idea is to use this area to transfer data between partitions when speed is essential.

The analog input area is dedicated to continuously reading analog inputs from the Load Frequency control equipment and storing them in the Common Area for use by economic dispatch and display programs.

The CRT and Keyboard interface provide the function of character echo-back to the CRT's, cursor movement on the CRT's, interrogating CRT requests and scheduling the appropriate display application program into Dynamic Region II, and filling interface buffers in the Common Area with data base changes from the keyboards. This partition is dedicated to these routines to insure fast response on the CRT system.

The Data Acquisition System (DAS) communication routines transmit requests for data and receive data from three mini-computer data concentrators over three 50,000 bit per second binary synchronous communication links. The communication routine operates in the following manner. Every second, a request for a status indication is transmitted in parallel to each mini-computer. The data transmitted to the 1800 is a series of 1/0 flags indicating one or more of the following states:

- no change since last transmission.
- a circuit breaker has changed status.
- megawatt hour values have been read into the mini and are available to be transmitted to the 1800 upon request.
- data base update from the 1800 to the mini-computer has been accomplished successfully.
- line flow measurements have been read into the mini and are available to be transmitted.
- data base update from the 1800 is too large.

The mini-computers are polling the substations two or three times every second, searching for a breaker change. If a change occurs, the 1800 is notified via the status indication and then proceeds to read the breaker status information from the mini-computer and transmits a request to all mini's to initiate reading all flow measurements in preparation for a new state estimate. All readings are available in the mini's within 3 seconds and are transmitted to the 1800 when requested. Flow measurements that are received are placed in the Common Area and Configuration and State Estimation routines are scheduled into Dynamic Region III provided that additional circuit breakers have not changed status. A companion paper describes the operation of these routines in more detail.[5] If no

Figure 3. Numerical Results Corresponding to Case of Figure 2 →

MEAS	BUSSES		NORMALIZED RESIDUAL	NORMALIZED RESIDUAL X100	UNNORMALIZED RESIDUAL	UNNORMALIZED RESIDUAL MW	CONFIDENCE LIMITS	CONFIDENCE LIMITS MW
2	13	6					22.2	22.1
3	13	4					37.7	38.4
4	1	13					61.2	60.5
5	1	3					11.5	11.4
6	2	5	-0.3	17.2	-0.0	0.5	11.3	11.1
7	3	1	-0.0	0.0	-0.0	0.0	11.4	11.4
8	3	5	-75.4	-170.5	-3.2	-7.7	11.1	10.7
9	4	49	-0.0	0.0	-0.0	0.0		
10	4	13	-28.8	-11.0	-4.4	-1.7	38.0	37.9
11	4	15	-72.2	-98.2	-5.0	-7.0	27.6	27.3
12	14	4	-377.2	-160.0	-23.0	-9.8	35.2	35.2
13	14	8	131.8	38.5	7.4	2.1	6.1	6.1
14	14	5	57.4	1.4	2.6	0.0	10.4	10.3
15	5	3	-89.6	-168.9	-3.6	-6.8	11.0	11.0
16	5	14	-188.9	-56.7	-8.2	-2.4	10.2	10.2
17	5	2	-3.2	85.7	-0.1	3.2	11.3	11.6
18	5	11	147.3	82.4	6.4	3.7	10.1	9.9
19	6	7	-1.9	-50.5	-0.1	-4.6	22.8	22.7
20	6	13	46.9	7.5	4.4	0.6	21.9	22.7
21	7	46	-27.3	-8.8	-2.5	-0.8	22.7	22.6
22	7	6	34.6	-45.5	3.2	-4.2	22.7	22.8
23	8	9					10.6	10.6
24	8	14					5.9	5.9
25	46	7	22.5	-2.6	2.1	-0.2	22.6	22.6
26	9	8	-28.3	-1.1	-1.2	-0.0	10.5	10.5
27	46	9	-157.1	-23.8	-7.7	-1.1	30.6	30.6
28	25	29	-2.3	-82.9	-0.0	-2.4	7.1	6.8
29	25	26	1.5	-6.7	0.0	-0.2	10.9	10.9
30	25	24	36.9	21.9	1.5	0.9	10.8	10.6
31	26	27	-0.0	0.0	-0.0	0.0		
32	26	25	1.6	-6.7	0.0	-0.3	10.9	10.9
35	16	12					21.2	22.6
36	16	37					27.5	26.8
38	28	27					39.0	38.8
39	28	11					10.9	11.3
40	11	17	110.3	58.9	7.4	3.9	9.4	9.3
41	11	5	-105.1	39.4	-5.0	1.8	10.0	10.0
42	11	28	0.0	-0.0	0.0	-0.0		
43	11	22	-89.8	-343.8	-1.3	-5.2	15.4	15.4
44	11	19	78.4	51.3	4.3	2.8	6.8	6.6
45	22	53	0.0	-0.0	0.0	-0.0		
46	17	12	292.5	243.5	23.3	19.4	32.1	32.1
47	12	48	265.5	88.5	22.2	7.3	50.7	50.4
48	12	16	5.8	72.3	0.5	7.0	21.5	21.3
49	17	18	-30.3	-24.4	-1.0	-0.8	12.5	12.5
50	17	11	-101.4	-69.4	-4.6	-3.2	9.4	9.4
51	19	20	-0.0	-0.0	-0.0	-0.0		
52	19	18	33.4	21.3	1.6	1.0	20.2	20.5
53	19	11	-50.5	-38.1	-2.7	-2.1	6.8	6.8
54	50	18	0.0	0.0	0.0	0.0		
55	52	48	0.0	0.0	0.0	0.0		
56	54	20	-0.0	0.0	-0.0	0.0		
57	23	22	155.7	325.5	9.3	18.0	21.4	22.2
58	23	24	-15.7	-59.0	-1.3	-5.2	13.4	13.2
59	24	23	71.2	97.9	2.4	3.1	13.2	13.4
60	24	25	-19.4	-72.0	-0.8	-3.2	10.8	10.9
61	15	4	11.9	-69.6	2.1	-12.7	27.7	27.8
62	15	48	-278.3	-128.4	-23.6	-11.7	50.6	49.8
65	49	62	0.0	-0.0	0.0	-0.0		
66	33	61	0.0	-0.0	0.0	-0.0		
66	34	37	-11.5	-63.6	-1.7	-9.7	38.2	38.3
67	34	33	76.1	221.6	6.2	19.3	50.9	50.3
68	33	31	-50.9	24.1	-5.0	2.3	36.9	36.9
69	33	35	0.0	5.7	0.0	1.2	63.5	63.4
70	33	63	17.2	-87.0	0.9	-4.9	13.4	13.4
71	31	32	0.0	0.0	0.0	0.0		
72	31	60	-26.9	-9.5	-1.2	-0.4	24.7	24.4
73	31	30	-15.9	-2.4	-0.8	-0.1	6.6	6.6
74	31	36	-25.1	-7.3	-1.5	-0.4	11.0	10.8
75	31	36	-25.1	-7.3	-1.5	-0.4	11.0	10.8
76	31	33	-132.5	-44.4	-12.9	-4.3	36.7	36.7
78	56	55	-0.0	-0.0	-0.0	-0.0		
79	37	16	-44.3	-120.6	-8.2	-22.1	27.1	27.7
80	37	38	0.0	-0.0	0.0	-0.0		
81	37	34	10.8	-5.9	1.6	-0.9	38.2	37.8
82	36	55	0.0	0.0	0.0	0.0		
83	36	35	10.9	2.1	0.7	0.1	10.1	10.5
84	36	35	10.9	2.1	0.7	0.1	10.1	10.5
85	36	31	-46.4	-6.3	-2.9	-0.4	11.0	10.7
86	36	31	-46.4	-39.8	-2.9	-2.4	11.0	10.7
87	35	33	8.2	-6.1	1.8	-1.1	63.2	63.4
88	35	30	46.4	72.8	1.4	2.2	5.8	5.8
89	35	36	-6.6	-10.5	-0.4	-0.6	10.1	10.5
90	35	36	-6.6	-10.5	-0.4	-0.6	10.1	10.5
91	35	43	-12.3	8.4	-0.6	0.4	17.4	16.7
92	35	42	22.5	-16.2	0.8	-0.5	14.3	13.8
93	63	33	-8.8	-93.0	-0.5	-5.2	13.2	13.1
94	63	60	26.3	7.8	1.9	0.5	18.7	18.4
95	29	30	-27.9	-49.5	-0.7	-1.3	7.5	7.7
96	29	25	92.8	70.1	2.7	2.0	7.2	7.3
97	40	41	50.4	-293.1	2.3	-13.5	10.5	10.4
98	40	42	312.4	79.0	9.8	2.5	6.6	6.4
99	41	42	78.0	-96.8	2.4	-2.8	17.8	17.6
100	41	40	-88.3	-133.8	-4.0	-6.1	10.5	10.5
101	57	38	-0.0	-0.0	-0.0	-0.0		
102	30	44					22.7	22.7
103	30	44					22.4	22.4
104	30	35					5.8	5.7
105	30	31					6.6	6.4
106	30	29					7.6	7.4
107	44	47	-0.0	-0.0	-0.0	-0.0		
108	44	45	20.5	-4.7	0.6	-0.1	5.2	5.2
109	44	45	9.5	-82.0	0.3	-2.7	5.2	5.2
110	44	30	-4.4	-5.1	-0.3	-0.4	22.9	22.8
111	44	30	4.6	5.2	0.3	0.4	22.6	22.5
112	42	43	10.9	-6.1	0.2	-0.1	11.7	11.7
113	42	40	458.0	-123.2	13.6	-3.6	6.6	6.6
114	42	41					17.6	17.8
116	21	45					14.8	15.4
117	21	47					77.9	77.6
118	47	21					77.6	77.7
119	47	44					17.3	17.2
121	45	21	-0.0	0.0	-0.0	0.0		
122	45	44	113.1	190.7	3.9	6.0	5.3	5.0
123	45	44	-61.2	-275.7	-2.1	-9.8	5.3	5.0

changes in circuit breaker status occur during a 30 second interval, a flow measurement scan is initiated to check if a new state estimate may be required due to load changes or changes in status of unmonitored components in the network.

Megawatt-hour values are read automatically by the mini at ½ hour or 1 hour intervals and notification is transmitted to the 1800. The values are read in, placed into the Common Area, and communication routines are scheduled into Dynamic Region II to transmit them to the N.Y. IBM 370 system. Data Base revisions for the mini-computer system are under control of the 1800 computer to insure that no incompatibility exists between the three mini's and the 1800. Any revisions to the hardware monitoring system are entered into the 1800 which updates the mini-computers data base and its own data bases and transmits the revised data base to the respective mini. All the communication routines are resident.

Dynamic Region I has programs scheduled there for some megawatt hour logging, communication restarts, and Wall Display Board. Dynamic Region II handles all application displays and display updates, study-mode economic dispatch and transaction evaluation routines, some data base routines, and megawatt hour logging routines. Dynamic Region III handles all State Estimation routines, Control-mode Economic dispatch, and the computer functions such as compiles, assemblies, and disk management functions.

The dynamic regions were established to use the 64 thousand words of the machine in an optimal fashion. All routines are scheduled, execute, and exit when complete. Displays are not refreshed from the 1800 once placed onto the CRT's. They can be requested again if necessary. State Estimation results are stored on disk and transferred to the CRT's when requested. All Economic Dispatch displays can be placed on CRT's within a 2-3 second period. Study mode Economic Dispatches can be run and the results displayed within an 8-10 second period. State Estimation results can be displayed on a single CRT within 12 seconds and the 4 CRT cluster can be updated within a 15 to 20 second period. It requires approximately 1 second per point to update the Wall Display Board.

The programs are efficient because many algorithms were designed to work by exception only. Stations where circuit breakers had changed were the only ones analyzed by the configuration routines. A memory of the Wall Board State was kept in the 1800 and only those components that had changed were driven by the 1800. Estimation algorithms were executed only if the Chi-Square test failed.[4] We believe that it is the efficient design of the computational algorithms and the utilization of the Dynamic Regions that allowed us to use one moderate sized computer for the above functions. It is estimated that at present the system includes 150 application programs totalling approximately 90,000 lines of code.

COSTS

The hardware and development costs were as follows:

1. Hardware

   a. IBM 1800 process control computer, 64k core storage, floating point hardware, communication adapters, disks, printer and miscellaneous peripherals. $491,000

   b. Limited Graphic, Non-color CRT Display System (Data Disc Corp.). $ 54,000

   c. High-speed printer for CRT Displays (Gould Mfg. Co.). $ 19,000

   $564,000

BUS	VOLTAGE MAG	VOLTAGE ANGLE	CONFIDENCE LIMITS MAG	CONFIDENCE LIMITS ANGLE
2	96.7	0.72	0.23	0.13
32	102.1	-14.68	0.34	0.19
50	100.0	1.78	1.03	0.59
52	103.4	7.58	0.18	0.10
53	99.6	-10.45	0.22	0.12
54	99.2	-12.57	0.14	0.08
56	95.7	-12.76	0.28	0.16
57	103.2	-13.00	0.35	0.20
61	102.4	-13.53	0.30	0.17
62	101.3	6.32	0.85	0.48
1	99.5	8.43	0.60	0.34
3	99.4	6.73	0.60	0.34
6	96.9	-0.71	0.50	0.29
7	97.3	-2.10	0.51	0.29
8	99.1	-4.79	0.60	0.34
9	98.6	-5.40	0.59	0.33
15	97.7	13.05	0.41	0.23
16	98.2	-2.34	0.40	0.23
20	98.0	-9.91	0.43	0.24
21	101.7	-11.17	0.70	0.40
23	99.4	-15.50	0.71	0.41
24	100.0	-19.04	0.72	0.41
26	98.8	-18.99	0.74	0.42
27	98.4	-14.53	0.88	0.50
28	98.7	-10.06	0.48	0.27
29	102.4	-17.65	0.72	0.41
34	99.0	-11.08	0.38	0.22
38	97.4	-9.01	0.23	0.13
40	104.2	-7.30	0.36	0.21
41	104.5	-7.27	0.26	0.15
43	104.0	-11.09	0.36	0.20
45	100.6	-5.10	0.32	0.18
46	96.2	-3.31	0.53	0.30
47	101.6	-12.28	0.47	0.26
49	96.8	10.62	0.09	0.05
55	98.8	-12.80	0.02	0.01
60	103.3	-16.48	1.01	0.58
63	102.5	-17.51	0.71	0.41
12	97.1	5.01	0.31	0.17
13	95.4	6.71	0.44	0.25
14	101.4	8.69	0.41	0.23
18	100.1	-2.27	0.10	0.06
19	100.1	-4.65	0.21	0.12
22	99.0	-8.97	0.44	0.25
25	98.6	-19.77	0.72	0.41
36	101.4	-13.04	0.70	0.40
37	98.5	-10.77	0.39	0.22
44	101.1	-12.76	0.16	0.09
48	97.9	8.36	0.37	0.21
4	96.7	11.13	0.41	0.23
5	99.8	1.83	0.38	0.22
30	100.8	-15.29	0.71	0.41
42	104.8	-10.00	0.36	0.20
11	99.6	-2.77	0.29	0.16
31	101.7	-13.92	0.66	0.38
33	102.1	-11.66	0.18	0.10
35	102.0	-13.24	0.72	0.41

BUS	AVAILABILITY	VALUE
9	1	97.9
17	0	100.3
36	0	102.6

SEPARATED STATION CONDITION

STATION 29 SEPARATED INTO
NODES 29 63
NEW CONFIGURATION AT STATION 29

25	29
29	30
29	25
30	29
33	63
63	33
63	60

Figure 4. Numerical Results Corresponding to Case of Figure 2

Monthly Cost (7 yr. full-payout lease)

a. Hardware $8,000/mo.

b. Maintenance $1,000/mo.
$9,800/mo.

2. Development

a. General purpose computer usage for simulation, formulation and analysis of solution techniques. $200,000

b. Advance installation of IBM 1800 computer for program development. $180,000

c. Personnel (approximately 5000 man-days) $400,000

d. Education $6,000
$876,000

3. Investment Totals

a. Hardware $564,000

b. Development Effort $786,000
$1,350,000

FUTURE DEVELOPMENTS

The monitoring function is an important application by itself. By providing a picture of current system conditions, operators can watch for impending abnormal conditions, authorize clearances and power interchanges and correct abnormal situations basing their judgments on reliable information. However, it is clear that other functions are needed for further enhancements of reliability of operation. Two such security oriented functions are currently under development: the contingency evaluation and the corrective action determination function.

At AEP, the main problem in contingency evaluation is to represent the effect of the reaction of external systems to contingencies in a sufficiently accurate manner. Research has been oriented towards the development of an adaptive equivalence of external system reaction that could be used for stochastic load flow calculations. It now appears that an equivalence, together with its accuracy information, is possible to be developed from prior state estimation results on the internal monitored system. State estimation results on current system conditions provide data on system loads and voltages and their accuracies. The stochastic load flow will combine all this information to calculate probable range of power flows and voltages throughout the system for a specific contingency.

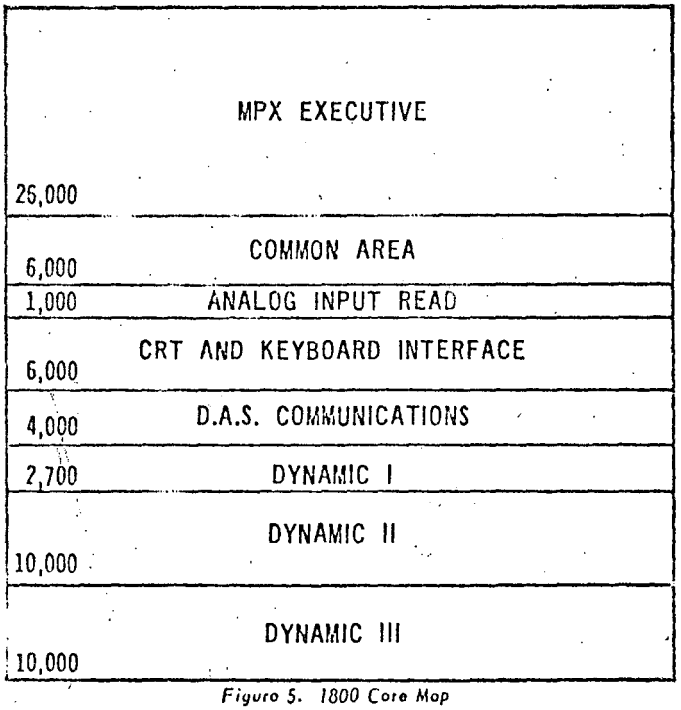


Figure 5. 1800 Core Map

Once the contingency evaluation function is ready, it is expected that it will be manually actuated by an operator request. Certainly, the effect of a planned outage will be studied with this function. During other conditions, such as heavy system-loading, the operator may test certain specific outages for the purpose of being alert and ready to take action should such contingencies occur.

The corrective action determination function will be a security load flow that will derive control actions required to alleviate a current problem by an optimization technique. Correction of overloads or abnormal voltage levels are typical problems that will be treated in this way. State estimation calculations not only provide the initial system conditions, but will make possible the tracking of the effect of control action thereby providing feedback to compensate for model inaccuracies. As in the contingency evaluation case, operator assistance in initializing the corrective action determination function and executing its control results will be required.

CONCLUSIONS

The paper has described the design of the AEP control computer. It has presented data that shows costs can be minimized without sacrificing response time or availability. Design concepts that may be useful to others who are planning a control computer installation are discussed.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the many contributions made by Oscar A. Klitin, a former member of our group, and the programming efforts of Sara Elam, Lewis Nowitz, and Janet Williams.

REFERENCES

1. G.W. Stagg, J.F. Dopazo, M. Watson, J.M. Crawley, G.R. Bailey and E.F. Alderink, "A time-sharing on-line control system for economic operation of a power system," Proc. Instrum. Soc. Am., October, 1966.

2. J.F. Dopazo, O.A. Klitin, G.W. Stagg, and L.S. VanSlyck, "State Calculation of Power Systems from Line Flow Measurements," IEEE PAS-89, pp. 1698-1708, September/October, 1970.

3. A.M. Sasson, S.T. Ehrmann, P. Lynch and L.S. VanSlyck, "Automatic Power System Network Topology Determination," IEEE PAS-92, pp. 610-618, March/April, 1973.

4. J.F. Dopazo, O.A. Klitin, and A.M. Sasson "State Estimation for Power Systems: Detection and Identification of Gross Measurement Errors," Proceedings of the IEEE PICA Conference, June, 1973.

5. J.F. Dopazo, S.T. Ehrmann, O.A. Klitin, A.M. Sasson and L.S. VanSlyck, "Implementation of the AEP Real-Time Monitoring System," to be submitted for the IEEE Winter Power Meeting, 1976.

For Combined Discussion, see page 1625

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA VI: IMPLANTACION DE UN ESTIMADOR DE ESTADO Y EXPERIENCIA DE OPERACION

DR. ALBERTO MAYER SASSON

ENERO, 1979.

IEEE Transactions on Power Apparatus and Systems, Vol. PAS-95, no. 5, September/October 1976

IMPLEMENTATION OF THE AEP REAL-TIME MONITORING SYSTEM

J. F. DOPAZO, S. T. EHRMANN, O. A. KLITIN, A. M. SASSON AND L. S. VAN SLYCK
AMERICAN ELECTRIC POWER SERVICE CORPORATION
NEW YORK, N. Y. 10004

ABSTRACT

During the final implementation of the AEP real-time monitoring system a considerable theoretical and practical experience was obtained which has resulted in a deeper insight and understanding of its performance. The paper presents the various modules and interfaces of the programming system and discusses concepts related to radial lines, redundancy, identification, separated stations, weights, convergence, model and parameter uncertainties and zero injection measurements. The final part of the paper presents hints on debugging and implementation strategies. The information presented would be of use to both researchers and to those with responsibilities in implementing estimation monitoring systems.

INTRODUCTION

Our previous papers have presented and discussed the justification and theoretical aspects of the AEP real-time monitoring system.[1-11] During the final implementation and checkout phases, a considerable experience was obtained which resulted in a deeper insight and understanding of both theoretical and practical considerations of the AEP monitoring system. This paper will present some of this experience which may be useful to both researchers and those with responsibilities in implementing state estimation monitoring systems.

The paper will first summarize the AEP state estimation and configurator algorithms to provide a basis for some of the topics discussed later. This will be followed by a block diagram presentation of the monitoring system and its interfaces. The central portion of the paper presents topics related to the experience and recent insights gathered during implementation. The paper ends by presenting debugging and implementation hints.

AEP STATE ESTIMATION ALGORITHM

Line flow information of real and reactive power, Sm, is received for one or both ends of most EHV 345 and 765 KV lines. The following equations can be written,

$$\bar{S}_m = \bar{S}_c(\bar{E}) + \bar{\varepsilon} \tag{1}$$

when $\bar{S}_c(\bar{E})$ is the calculated line power flow as a function of the complex nodal voltages $\bar{E}$. The measurement errors $\bar{\varepsilon}$ are assumed to be zero mean with diagonal covariance matrix W^{-1}. The solution of the redundant set (1) is accomplished by minimizing the weighted sum of squares

$$J(\bar{E}) = (\bar{S}_m - \bar{S}_c)^{*T} W (\bar{S}_m - \bar{S}_c) \tag{2}$$

which can be algebraically reduced to the form

$$J(\bar{E}) = (\bar{v}_m - \bar{v}_c)^{*T} D (\bar{v}_m - \bar{v}_c) \tag{3}$$

where $\bar{v}$ is the element voltage across the line, D a diagonal matrix with

$$D_j = \frac{W_j |E_j|^2}{|z|^2} \tag{4}$$

and z is the impedance of the line where measurement j has been taken. The relation between Eqns (2) and (1) infer that the minimization of J(E) of Eq. (3) implies the solution of the following redundant set,

$$\bar{v}_m = \bar{v}_c(\bar{E}) + \bar{\nu} \tag{5}$$

where $\bar{v}_m$ is taken as an element voltage "measurement" computed from the corresponding physical measurement $\bar{S}_m$, the impedance characteristics of the line and the nodal voltage $\bar{E}$ at the measured end of the line. The term $\bar{v}_c$ is the difference of nodal voltages $\bar{E}$, which in terms of an incident matrix A can be written as

$$\bar{v}_c = A\bar{E} + \bar{\nu} \tag{6}$$

The minization of $J(\bar{E})$ of Eq. (3) with respect to rectangular components of $\bar{E}$ is accomplished by solving

$$\frac{\partial J(\bar{E})}{\partial E} = A^T D(\bar{v}_m - A\bar{E}) = 0 \tag{7}$$

where the term $(\bar{v}_m - A\bar{E})$ is an element voltage residual vector and v_m has been assumed as constant and independent of $\bar{E}$. As this is not strictly true, the solu- of Eq. (7),

$$\bar{E} = (A^t DA)^{-1} A^t D \bar{v}_m \tag{8}$$

has to be iteratively repeated after $\bar{v}_m$ has been updated for the new values obtained for $\bar{E}$. One bus voltage is assumed known and used as reference.

BAD DATA DETECTION, IDENTIFICATION AND CONFIDENCE LIMITS

In a previous paper, it was discussed that at its minimum, $J(\bar{E})$ should satisfy the inequality

$$J(\bar{E}) < X^2_{k,\,b/2} \tag{9}$$

where $X^2_{k,\,b/2}$ is the theoretical value of a chi squared distribution for k degrees of freedom and for a probability of confidence b. If inequality (9) is not satisfied this is taken to mean that the existence of bad data has been detected. The degrees of freedom is twice the difference between the number of complex measurements and the number of complex nodal voltages to be estimated.

Once bad data has been detected, it is identified as those measurements j where the largest value of

$$\frac{(\bar{v}_m - \bar{v}_c)_j - \text{zero}}{s\sqrt{Dj^{-1} \cdot \text{diag}\,[A(A^t DA)^{-1} A^t]_j}} > t_{k,b/2} \tag{10}$$

occurs. The value of $t_{k,b/2}$ is that of a student-t distribution and that of s^2 is $J(\bar{E})/k$. Since the $(\bar{v}_m - \bar{v}_c)$ term is an element voltage residual, the expression (10) is a normalized residual with the denominator being the standard deviation of the residual.

The confidence limits of voltages $\bar{E}$ is given by

$$\text{C.L.}(\bar{E}) = s\sqrt{\text{diag}\,[(A^t DA)^{-1}]} \tag{11}$$

and those of the flows by

$$\text{C.L.}(\bar{S}) = s\sqrt{\text{diag}\,[K(A^t DA)^{-1} K^t]} \tag{12}$$

where K is the Jacobian of P with respect to E for C.L. (P), or K is the Jacobian of Q for C.L. (Q).

NETWORK CONFIGURATOR

The basic function of the network configurator is to define subsystems internally connected by measured energized lines and containing a reference voltage measurement. As measurements and breaker status data is received the configurator performs the following tasks:

1. Analyzes breaker status changes to determine internal connections of lines and busses at substations.
2. Determines if there are deenergized lines.
3. Determines if there are separated substations
4. Determines if there are lines whose measurements are not available.

Paper F 76 087-7, recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE PES Winter Meeting and Tesla Symposium, New York, N.Y., January 25-30, 1976. Manuscript submitted August 28, 1975; made available for printing October 31, 1975.

5. Determines a basic system configuration containing only measured energized lines.

6. Analyzes the basic configuration to determine its connectivity.

7. Determines connected subsystems containing at least one available reference voltage measurement.

The network configurator design concepts presented in Refs. 6 and 7, have proven in the real-time environment to be adequate to cope with all the above tasks.

PROGRAMMING SYSTEM STRUCTURE

The programming system structure presented in Figure 1 was designed to achieve efficiency from the points of view of running time and core requirements. As each set of measurements is to be processed, analysis is made to determine the optimum path through the various steps. This adaptive characteristic of the programming system increases its complexity, to be able to achieve the flexibility of deciding on-line which programming modules are necessary to process the incoming data.

The various modules and some of their interfaces will now be described. Many modules have data bases prepared by off-line programs and these are not shown in Figure 1.

MODULES

1. Teleprocessing

This name summarizes programs that interface with the communication system, receive and store incoming data, determine which data is missing and schedule the monitoring process on a time basis or when breaker changes have occurred. Parts of this module are continuously resident in the computer. In addition it has special functions during system initialization where there is no previous history.

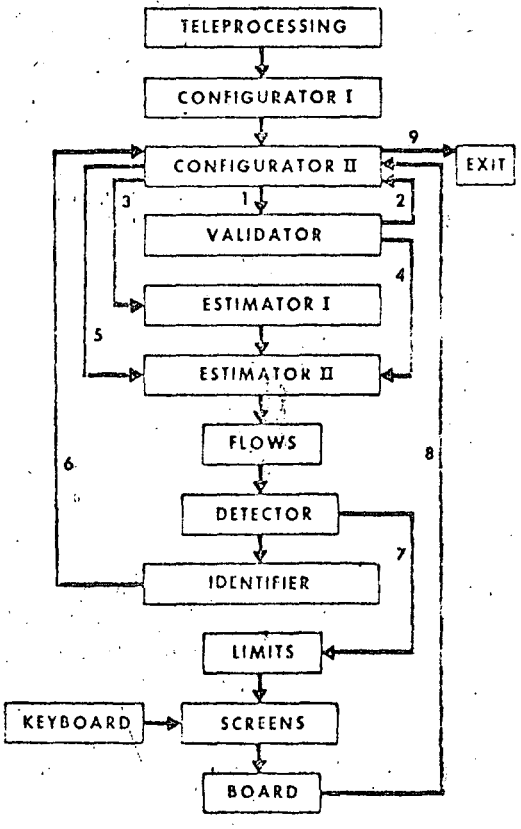


Fig. 1. Programming System Structure

2. Configurator I

The programs in Configurator I analyze breaker status changes, and their effect on deenergizing lines or separating stations. The module is by-passed if there are no breaker changes.

3. Configurator II

This module determines which lines are measured and energized, studies the connectivity of these lines and determines the number and composition of separated subsystems containing a voltage magnitude measurement. Most of this module is by-passed if there are no changes in lines being measured and energized.

4. Validator

Although Eq. (9) presents the chi-square test to detect bad data, a similar test is made in the Validator to determine whether the incoming data is statistically comparable with the previous results.

5. Estimator I

Once a subsystem has been defined Estimator I proceeds with all the non-iterative portions of the estimator algorithm. Basically, it defines matrix D, optimally orders the equations contained in (8), calculates and triangularizes the (A^t DA) matrix and performs a series of preparatory functions to increase the efficiency of the iterative section.

6. Estimator II

This is the iterative section, where Eq. (8) is solved using sparsity techniques.

7. Flows

The output quantities of line flows and voltages are calculated in this module based on the results of Estimator II.

8. Detector

The chi-square test described by Eq. (9) is performed here to detect the presence of bad data.

9. Identifier

Once bad data has been detected, this module performs the student-t test described in Eq. (10) and removes identified bad data.

10. Limits

This module computes the confidence limits of voltages and flows described in Eqs. (11) and (12) for output purposes.

11. Screens and Keyboard

The screen programs prepare the output that will be displayed on CRT's upon requests entered through the keyboard.

12. Board

The display board programs analyze output data and drive a large mimic board by actuating various lights.

INTERFACES

1. If there are no changes in subsystem configurations, Configuration II interfaces with the Validator.
2. When the Validator chi-square test is successful, it interfaces back to Configurator II to pick next subsystem. Previous output remains valid.
3. If there are changes in subsystem configuration, new measurements must be processed and Configurator II interfaces with Estimator I by-passing the Validator.
4. When the Validator chi-square test fails, measurments must be processed. The Validator interfaces with Estimator II by-passing Estimator I.
5. When there are no changes in subsystem configuration coming via 6.
6. When bad data are identified and removed in the Identifier it interfaces with Configurator II to determine the effect of this removal upon subsystem configuration before repeating the estimation process.
7. When Detector does not detect bad data it interfaces with Limits to provide output for subsystem.
8. After completion of all output, Board interfaces with Configurator II to process next subsystem.
9. If all subsystems have been processed the programming system exits from core.

STORAGE OF INFORMATION

A real time process such as the present one requires a careful organization of the transfer of information between modules and the treatment of information within modules. Specifically, a careful balance must be decided upon on two issues:

1. Quantities that will be maintained in their original state.
2. Quantities that will be transferred into a packed state.

In both cases a price is paid. In the first one, a series of pointers may have to be constructed in real time to selectively access information whose organization will remain static. Additionally, a series of checks may have to be made to determine what is the pertinent information for the current condition.

In the second case, information must either be transferred or formed to get it into a compact state for processing. A series of pointers may have to be constructed during this operation to be able to interpret results in their original state which is the physically meaningful one.

Both of these issues burden the interfaces between modules especially when they treat information differently. A process in which the problem size and content can vary in real time, but within a certain fixed framework, is particularly difficult to design.

In addition to the above issues from the point of view of current problem definition, there is another different level which also must consider the same two issues, this is the concept of sparsity programming and its related issue of optimal ordering. The question here is whether certain information should be packed or not sparsity wise, and should it be organized in original or optimal ordering. Again on this level a price is paid either way and a careful program design must be made.

When the above two issues are properly resolved in its two levels, an efficient program is obtained from the point of view of both core requirements and speed of execution. Both are vital in a real time process.

RADIAL LINES

For a deeper understanding of the estimation, detection and identification processes, some simple situations can be examined in detail. During implementation, such close examination of isolated cases proved to be invaluable in defining the final function and structure of many programs and in locating programming errors. This section will examine closely the radial line. The next section will extend the concepts derived from the radial case to lines belonging to loops.

Behavior of Estimator

It is important to understand the behavior of radial lines in the estimation, detection and identification processes. Frequently lines become radial due to the breaking of a loop due to the opening of some other line or to measurements being either missing or previously removed as bad. The effect of the estimator solution process on the measurements of a radial line can be predicted when the following analysis is made.

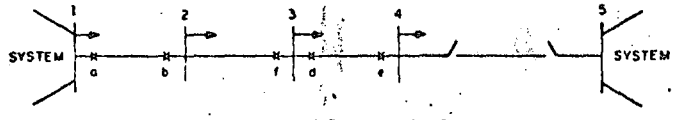


Fig. 2. Radial Portion of a System

Consider Figure 2 where line 4-5 is open thus breaking a loop leaving a radial string. Line 2-3 has one measurement and lines 1-2 and 3-4 have two measurements each. From the point of view of Eq. 7 the complex equations solved by the Estimator include the following for busses 2, 3 and 4.

$$-D_b R_{12} + D_b R_{21} - D_f R_{32} = 0 \quad (13)$$

$$D_f R_{32} + D_d R_{34} - D_e R_{43} = 0 \quad (14)$$

$$-D_d R_{34} + D_e R_{43} = 0 \quad (15)$$

where D are the v_m variances and

$$R_{ij} = v_{m,ij} \quad v_{c,ij} \quad (16)$$

and

$$v_{c,ij} = E_i - E_j \quad (17)$$

it is clear that in Eq. (15)

$$v_{c,34} = - v_{c,43} \quad (18)$$

thus,

$$-D_d v_{m,34} + D_e v_{m,43} - D_d (E_4 - E_3) - D_e (E_4 - E_3) = 0 \quad (19)$$

and

$$E_4 - E_3 = \frac{-D_d v_{m,34} + D_e v_{m,43}}{D_d + D_e} \quad (20)$$

If E_3 was known, Eq. (21) shows that E_4 is only dependent on weighted effect of the two measurements on line 3-4. This is a general statement for a radial line with two measurements. Another characteristic that can be derived from Eq. (15) is that if $D_d = D_e$, then $R_{34} = R_{43}$ and Eq. (20) becomes an average.

$$E_4 - E_3 = \frac{-v_{m,34} + v_{m,43}}{2} \quad (21)$$

A comparison between Eqs. (14) and (15) shows that after (15) has been satisfied, (14) becomes $D_f R_{32} = 0$ which implies that $R_{32} = 0$ independently of D_f:

$$E_3 - E_2 = v_{m,32} \quad (22)$$

If E_2 was known, E_3 is directly determined by Eq. (22). Finally, a comparison of Eqs. (14) and (13) after convergence of (15) and (14), shows that (13) becomes the same type of equation as was (15). Eq. (20), with subscrips 4, 3, d and e substituted for 2, 1, a and b, also holds.

The voltage at bus 1 is not affected by measurements or conditions in the radial string, from the point of view of the estimator convergence process. That voltage is determined by conditions imposed by Kirchoff's voltage law on the system side and involving the reference voltage measurement. This can be seen clearly by realizing that given E_1 at any value, Eq. (13) determines E_2, Eq. (14) provides E_3 and Eq. (15) fixes E_4, so that the radial string does not place any conditions on E_1.

In relation to Eq. (22), it is true that E_3 is directly determined from this equation at every iteration of the estimation process. However, $R_{32} = 0$ only after convergence. The term $v_{m,32}$ was calculated using the flow measurement and the voltage from the previous iteration. Thus $v_{m,32}$ changes as soon as a new voltage is computed except, of course, at convergence.

Redundancy

From the previous paragraph it could be erroneously concluded that there is ample redundancy for bad data identification on the radial string being examined, 5 measurements and only 3 voltages to determine. This global determination of the redundancy of the string is not meaningful. The examination of the solution process made on Eqs. (15), (16) and (17) reveals the more valid concept of local redundancy. Whatever measurement f is, good or bad, Eqs. (16) or (24) will produce an $R_{32} = 0$, that is, a perfect match, adjusting E_3 accordingly. Thus, there is no way that a bad measurement f can be detected or identified. Its local redundancy is zero, one measurement affecting one voltage. In lines 1-2, or 3-4, there are 2 measurements to determine one voltage in each case. In these cases, a single bad measurement will be detected because the residuals will be large. However, it is not possible to identify which is the bad one of the two. If any of the two were to be removed, the residual would become zero as in line 2-3. The only step that could be taken is the removal of both measurements, the good and the bad one, paying the price that from that line outwards, the string is lost.

Variance and Bad Data Identification

To see mathematically how the Bad Data Identifier would remove both measurements of a radial line containing two measurements, one of these being bad, the following example is enlightening:

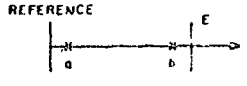


Fig. 3. Two Bus Sample System

One end has been chosen as the reference bus with no loss of generality according to the reasons given in relation to E_1, in Fig. 3. As in the previous example, only one of the rectangular components will be analized.

$$A = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad D = \begin{bmatrix} D_a & \\ & D_b \end{bmatrix}$$

calling α the denominator of Eq. (10),

$$\alpha_a = s^2 \frac{D_b / D_a}{D_a + D_b} \tag{23}$$

$$\alpha_b = s^2 \frac{D_a / D_b}{D_a + D_b} \tag{24}$$

From Eq. (15), it can be shown that

$$R_a = \frac{D_b}{D_a} R_b \tag{25}$$

The normalized residual used in the Student-t test is, according to Eq. (10).

$$R_a^n = \frac{R_a}{\alpha_a^{1/2}} = \frac{R_a}{s \left(\frac{D_b / D_a}{D_a + D_b}\right)^{1/2}} \tag{26}$$

Substituting Eq. (25) into (26),

$$R_a^n = \frac{R_b}{s \left(\frac{D_a / D_b}{D_a + D_b}\right)^{1/2}} = \frac{R_b}{\alpha_b^{1/2}} = R_b^n \tag{27}$$

demonstrating that both normalized residuals will be identical. As it has been shown that the bad data will affect only its own line residuals in the radial case, R_a and R_b will be large. However, the normalized residuals R_a^n and R_b^n will be equal and both measurements will be simultaneously removed as being the largest ones that fail the student-t test.

Effect of Additional Voltage Measurements

If a system is predominately radial, some thought should be given to adding voltage magnitude measurements and processing these in the manner suggested in Ref. 12. Such a measurement will not change the advantageous characteristic of the Estimator in terms of sparsity, core and speed. A voltage measurement essentially closes a loop by providing a voltage difference between the measured bus and the reference bus. This is a loop in the sense of Kirchoff's voltage law and not of network configuration.

Consider the effect of adding a voltage measurement E as shown in Fig. 3. Matrices A and D become,

$$A = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad D = \begin{bmatrix} D_a & & \\ & D_b & \\ & & D_v \end{bmatrix}$$

where Dv is the weight of the voltage measurement. Eqs. (23) - (27) become,

$$\alpha_a = s^2 \frac{D_b / D_a + D_v / D_a}{D_a + D_b + D_v} \tag{28}$$

$$\alpha_b = s^2 \frac{D_a / D_b + D_v / D_b}{D_a + D_b + D_v} \tag{29}$$

$$R_a = \frac{D_b}{D_a} R_b + \frac{D_v}{D_a} R_v \tag{30}$$

$$R_a^n = \frac{R_a}{s \left(\frac{D_b / D_a + D_v / D_a}{D_a + D_b + D_v}\right)^{1/2}} \tag{31}$$

$$R_a^n = \frac{R_b + R_v D_v / D_b}{s \left(\frac{D_a / D_b + D_a D_v / D_b^2}{D_a + D_b + D_v}\right)^{1/2}} \neq R_b^n \tag{32}$$

Eq. (32) concludes that the normalized residuals of a and b are no longer equal making it possible for bad data identification. For a simpler analysis, consider a case in which $D_a = D_b = D_v = D$,

$$R_a^n = \frac{R_a^n}{s\,(2/3D)^{1/2}} = R_b^n + \frac{R_v}{s\,(2/3D)^{1/2}} \tag{33}$$

$$R_b^n = \frac{R_b}{s\,(2/3D)^{1/2}} = R_a^n - \frac{R_v}{s\,(2/3D)^{1/2}} \tag{34}$$

and Eq. (30) becomes,

$$R_a = R_b + R_v \tag{35}$$

Consider that b becomes a bad measurement. Then, Eq. (35) will force the absolute values of R_a and R_v to increase to offset a large Rb. As in the case analyzed, the denominator of R_a^n and R_b^n are equal, Eqs. (33) and (34) show that if $R_b^n > R_a^n$ then $R_b^n > R_a^n$ and proper identification is made.

In relation to Fig. 2, if a voltage measurement is added at bus 4, it would then be possible to identify a bad measurement on any of the five measurements, including f the single measurement on line 2-3, However, if measurement d were to be missing leaving a single measurement on two branches of the radial string, a voltage measurement at 4 will provide bad data detection if f or e were bad but it would not be possible to identify which of these two is bad. For proper identification between two measurements, it must be possible for the good one to be removed and the effect of the bad one continue to be felt. In the case considered, if either f or e is removed, the residual of the remaining one would become zero making both detection and identification impossible. An additional voltage measurement at bus 2 would not help because it has the effect of correlating with measurements a and b which are good ones. An additional one at 3, however, will provide proper detection and identification.

REDUNDANCY IN LOOPS

The concept of redundancy in estimation refers to the difference between the number of measurements and states. This is a global concept involving all measurements and states. There also exists the concept of local redundancy when a subgroup of measurements and states are related only to each other. Such a case is discussed in the section on radial lines. However, local redundancy can exist with a network loop. Consider the example of Fig. 4,

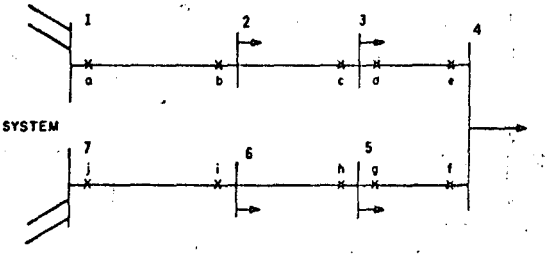


Fig. 4. Loop Portion of a System

where the portion of a loop that has no boundaries with other loops is shown. Lines 2-3 and 5-6 each have only one measurement. If one of these two becomes bad, bad data will be detected because residuals all over the loop will be large. An analysis similar to the one made for radial lines would reveal that the normalized residuals at c and h would be equal and also the largest, making identification between the two impossible. As in the radial line case discussed, if the good measurement were to be removed, the effect of the bad one disappears because the loop would be broken leaving two radial strings. This appears to be a valid test for determining the possibility of bad data identification.

If a system has many loop sections as the one of Fig. 4, serious consideration should be given to the inclusion of voltage measurements for bad data identification purposes. For the example just discussed, a voltage measurement at busses 3, 4 or 5 would make possible bad data identification of either c or h. A voltage measurement at busses 2 or 6 would not provide this condition.

In the absence of voltage measurements in the loop of Fig. 4, when a bad data occurs at c or h, both measurements would be removed. Conversely, if any of the other measurements becomes bad, the loop and the remaining good measurement on the same line provide the necessary redundancy for identification.

IDENTIFICATION WITH FLOWS OR ELEMENT VOLTAGES

It has been mentioned that line flow measurements are the physical quantities that are transmitted to the monitoring computer. However, once bad data has been detected, there are two alternative quantities that theoretically can be used for bad data identification:

1. Line power flow measurements, $\bar{S}_m$
2. Derived element voltage measurements, $\bar{v}_m$

The second alternative uses Eq. (10) for the identification test. In the first alternative, the following relation must be used:

$$\frac{(\bar{S}_m - \bar{S}_c)_j}{s \sqrt{W_j^{-1} - \text{diag}\,[K(A^t DA)^{-1} K^t]_j}} > t_{k\,b/2} \tag{36}$$

when K is the Jacobian of S with respect to E.

It must be remembered that in deriving Eqs. (12) or (36) the approximation was made that $\bar{v}_m$ is a constant, per iteration, computed from $\bar{S}_m$ and the best nodal voltage values available. This linearizes equations, enhances sparsity and decouples equations from the point of view of the rectangular components of nodal voltages. The decoupled state covariance matrix introduces a small error in the computation of the variance of $\bar{S}$, which is normally negligible. When computing the variance of a flow that is the only measurement on a radial line, the result should be exactly the input measurement variance. This leads to a zero normalizing term for the denominator of Eq. (36) for the Student-t bad data identification test. Practical real time experience has shown that the small error in the computation of the variance of $\bar{S}$ is sufficient to sometimes provide a variance slightly larger than the input one, resulting in a negative term, within the square root of Eq. (36). This is not so serious because this condition can be checked before taking the square root and the bad data identification skipped. When the error is such that the variance is smaller than the input one its difference is sometimes a large enough number that it becomes difficult to distinguish it from those that are indeed different from zero in order to skip bad data identification.

All the above problems can be avoided by computing the variance of v_m using alternative 2. Real time experience has shown that the differences obtained are several orders of magnitude smaller than the ones that should indeed be different from zero, allowing proper discrimination.

However, for the computation of confidence limits of $\bar{S}$, the small error mentioned is completely negligible.

EFFECT OF SEPARATED STATIONS ON CALCULATIONS

It has been mentioned that the configurator modules will alter the system configuration when breaker operation result in a separated station. All data files are dynamically changed to reflect the existing condition. It is clear that a station separation will substantially alter flows and voltages around that station. In testing the ability of the programming system to cope with separated stations, cases where provoked by arbitrarily considering open breakers that were closed in the power system. Theoretically, if all measurements and weights were perfect, such a non-existent separation should not alter any of the results around that station, and the separated busses should turn out to have identical voltages. In practice some differences will be encountered due to the change introduced in local redundancy.

Consider the Eq. (7) corresponding to the station in question. Normally, when station i is one electrical node, Eq. (7) is as follows,

$$\sum_{j=1}^{m} (A^t D)_{ij} R_j = 0 \qquad (37)$$

considering that there are m measurements on lines connected to bus i. After a separation in two parts i_1 and i_2, Eq. (7) becomes,

$$\sum_{j=1}^{m_1} (A^t D)_{i_1 j} R_j = 0 \qquad (38)$$

$$\sum_{j=1}^{m_2} (A^t D)_{i_2 j} R_j = 0 \qquad (39)$$

where $m_1 + m_2 = m$. It is clear that to force m terms to add to zero is not the same as to force the m_1 and m_2 terms to separately meet the same condition.

WEIGHTING FACTORS

In previous published work it was advocated use of measurement variances which would be calculated as a function of the magnitude of the flow and as a function of the full scale component as follows:

$$W_j = \frac{1}{[.02 * MVA_j + .0052 * FS_j]^2} \qquad (40)$$

Experience with the estimation algorithm has shown that possibly such computation of the weight may be undesirable.

For instance, consider a bad line flow measurement of 0 + j0. More often than not, such cases are few physically and such a reading may be a bad measurement. In that case a very large weight would be computed from equation (40). When used in the algorithm, such a weight comparatively makes that measurement more accurate than others with smaller weights making identification more difficult. The same argument applies in reverse, if the measurement is relatively close to the full scale value, and is faulty, when processed by the estimator and in relation to others which may be, say, near the mean of the range.

Since these considerations could eventually affect real-time performance of identification of bad measurements, the weights were changed to be constants depending only on the full scale of the measurement assuming a reading at the mean of the scale.

$$W_j = \frac{1}{[.0152 * FS_j]^2} \qquad (41)$$

Additionally, if measurements are not faulty, that is, if residuals are small, the effect of weights on estimator performance are of second order.

For bad data detection and identification it is essential to have the correct weight level. If all weights are larger than their true level, the Detector will see an abnormally large sum of squares in Eq. (2) leading to the conclusion that bad data exist. The identification may not find any bad data. If weights are smaller than they should be, the risk of not detecting bad data exists. It has been found that there is room for a range of values around the true weights between these two extreme conditions.

Concerning implementation, when a measurement is missing or bad, the measurement on the other end of the same line is used in its place. The corresponding weight must also be used in Eq. (8) in two places: the matrix $(A^t DA)$ and the term $A^t D \bar{v}_m$. The weight is included in D as given in Eq. (4). An exception to the ideal choice of this weight was made when the missing or bad measurement occurs without the configuration changing at the same time. To avoid forming and triangularizing matrix $A^t DA$ again to reflect the change in weight, the term $A^t D \bar{v}_m$ must also use the weight of the missing or bad measurement. The measurement value used will be the one at the other end of the line. Eventually, when the configuration changes due to breaker operation or measurement unavailability, the $A^t DA$ matrix is updated, at which time the right weight is used everywhere.

CONVERGENCE AND NUMERICAL CONSIDERATIONS

Estimator convergence characteristics are generally good and convergence can be expected, from flat start, in 6 to 9 iterations with a rectangular voltage tolerance of 0.000005. Additional iterations may cause slight improvements but are not warranted.

Such a statement can aid those who may be implementing this programming system to recognize abnormalities or errors in programming or data. Considerable debugging time was spent in the AEP final implementation because the above statement was violated in some instances of the day in an apparent random fashion. Small errors were eventually located and corrected.

There are two criteria that could be followed for testing convergence. Eq. (7) could be computed at the end of every iteration and used as a mismatch tolerance test. This requires additional computation. For this reason, the change in the voltage of two successive iterations, as computed by Eq. (8), was used. Various tolerance levels were tested. The one chosen was a good compromise between accuracy and number of iterations.

At some stages of the implementation certain problems arose which appeared to point towards some numerical round-off problems in some computations. Although later it was found that this was not the case, some efforts were conducted in testing the need for double precision computations. It was concluded that single precision round-off affected the sixth and seventh digit of the triangularized $A^t DA$ matrix without affecting the solution of the equations in a noticeable manner.

Strictly speaking, Eq. (8) computes the nodal voltages with respect to the reference bus voltage. The reference voltage must be added to each computed one for output purposes or for further computations involving voltages, once convergence has been obtained. An alternative approach could be followed in which Eq. (8) is rewritten as:

$$\bar{E} = (A^t DA)^{-1} A^t D (\bar{V}m - Ag \bar{E}g) \qquad (42)$$

where the elements of Ag are ± 1 for measurements on lines connected to the reference and 0 for other measurements, and Eg is the reference voltage. The computed voltage would be automatically with respect to ground. The numerical round-off properties of this alternate formulation are not the same as with the one implemented and should be kept in mind as a possibility in the event that numerical problems are encountered.

MODEL AND PARAMETER UNCERTAINTIES

It was previously mentioned that bad measurements can be detected and identified. For this to happen, it must be assumed that adequate parameter values and model are available. Parameters refer to transmission line impedance and transformer taps and impedances. By model it is referred to the network configuration. These two factors, parameters and model, affect the process in re-

lated but different manners. A recent paper addresses this problem similarly as presented here. [13]

Parameter Errors

Eq. 7 is the equation solved by the Estimator. A $\bar{v}_m$ term can be said to be

$$\bar{v}_m = f(\bar{s}_m, \bar{z}, \bar{a}) \quad (43)$$

where $\bar{z}$ are the impedances and $\bar{a}$ the transformer tap setting. As far as $\bar{v}_m$ is concerned, it can be in gross error due to bad values in any of its three functional components. The theoretical conclusion can be reached that errors in $\bar{z}$ and/or $\bar{a}$ will be detected and identified as bad data without the identification process being able to discriminate between any of the three sources.

Gross errors in $\bar{z}$ will occur only during the checkout period when the programs are first being used or when a new line or transformer is added to the monitored system. Such errors are first taken to be measurement errors. Only after these are confirmed to be correct the suspicion arises that the impedance related to the line-end where the bad flow measurement is being identified is the real cause. This psychological process was experienced under both simulation and real-time checkout phases. In every case that the measurement was reasonably checked to be correct a further check on the impedance showed that it was indeed in error.

These same comments are applicable to transformer tap settings as they affect both the reactance and the equivalent circuit. Wrong tap settings can happen in real time operation. Its effect will be detected, identified and removed as if it were a bad measurement. During checkout it was found that errors were committed in selecting the correct equivalent circuit shunt element of an off-nominal tap transformer. Care has to be exercised in determining whether the tap is on the measurement side or not. This is because only the measurement side shunt branch is used with a flow measurement to calculate $\bar{v}_m$ in Eq. 8.

From the theoretical considerations and practical experiences just mentioned, the conclusion was reached that with the estimation method used and the impedance values that a Line Constant Program produced, there was no need to do parameter estimation. We have said that serious errors in impedances were readily identified. Small inaccuracies in their values produce the effect of altering the precision of the flow measurements. If the estimator results are acceptable to the Bad Data Detector, then the impedance inaccuracy is of no serious consequence.

Model Errors

Model errors are produced by incorrect breaker status information. Only those errors that lead to the conclusion of introducing a deenergized line into the model have to be considered. The incorrect removal of an energized line only has the effect of reducing measurement redundancy and will not introduce further errors in estimated results.

In the AEP system most substations are of a breaker-and-a-half scheme, there thus being two breakers associated with each line-end. For an open line to be considered closed, wrong breaker status must be received from at least one breaker on each end. As a safety feature, a line is considered open if the breakers on one end are open. Even then, the Detector flags bad data and, what is very useful, the Identifier will tend to flag measurements that are on lines in the same loop as the incorrectly modeled one. The effect of this is to break the loop leaving the line in question as part of a radial string where its damage is isolated to only that string. In effect, the model error has been neutralized.

ZERO INJECTION MEASUREMENTS

The estimation, detection and identification processes depend on having proper global and local redundancy. Redundancy, of course, increases by having more measurements. However, measurements are expensive in data acquisition equipment imposing economical limitations to their number. There is one type of measurement which is free from data acquisition expense, and that is a zero injection measurement at a bus where there is no load or generation. However, what is free in one aspect can become expensive in another.

Ref. 14 presents a detailed analysis of the treatment of the zero injection measurement in the AEP algorithm as was originally suggested in the Closure to Ref. 5. There are two alternative methods:

1. Consider an $\bar{I} = \bar{Y}\bar{E}$ equation for each measurement of this type with $\bar{I} = \bar{0}$. This maintains the linear aspect and some other characteristics of the estimator. However, the price is paid in substantially increasing storage requirements as Y becomes part of matrix A and $(A^t DA)$ will have a complex Hessian like structure, considerably more dense than the present admittance like structure.
2. The unconstrained minimization process behind the estimation process can be expanded to an equality constrained minimization one, imposing the zero injection as a constraint to be handled Lagrangially. This augments the state but conserves sparsity. However, it has the additional serious consideration, that it changes the identification process.

If a system has a serious redundancy limitation, the alternative of using zero injection measurements should be given consideration. Otherwise, the price of using them may not be worth the extra care, time and programming complication.

The point can also be made that the two techniques described above can also be used for non-zero injection measurements. All that has to be considered is that the current is computed from the injection power and the voltage available at the iteration, obtaining the correct current at convergence.

DEBUGGING AND IMPLEMENTATION STRATEGY

After the initial programming of Configurator and Estimator modules it is helpful to have a plan for debugging the evolving real-time programming system. Such a plan is presented in Fig. 5. The arrows indicate possible debugging strategies that were found from experience to be effective. Each box in Fig. 5 has been numbered.

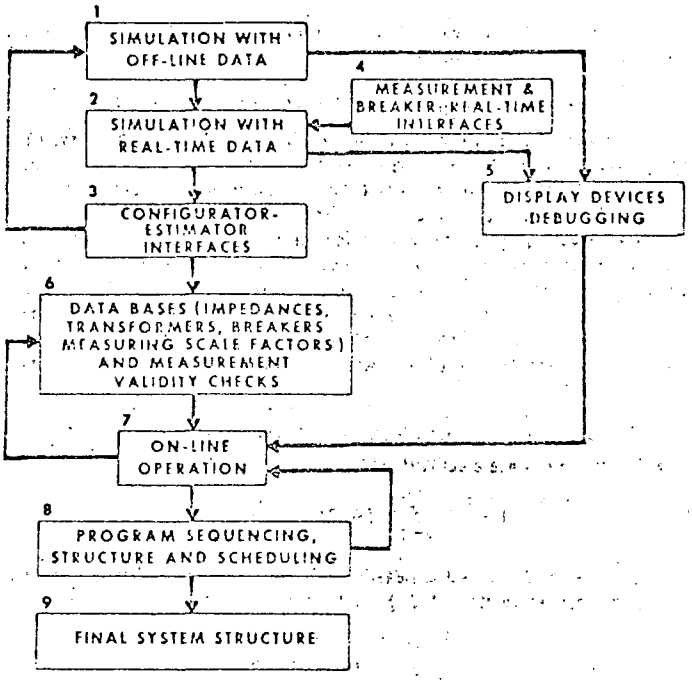


Fig. 5. Debugging and Implementation Strategy

1. Simulation with off-line data

The initial Configurator and Estimator modules should be thoroughly checked with simulated data before further implementation be carried out. This is one of the most time consuming but perhaps more useful steps, requiring about 50% to 60% of the debugging effort.

2. Simulation with real-time data

Once there is enough confidence in off-line performance, real-time simulation can start. Real-time measurements and conditions can be obtained for an instant of time and saved to provide a controlled testing environment. This need not be for the whole system.

3. Configurator-Estimator Interfaces

With the help of item 2, Configurator-Estimator interfaces could then be tested. This is one of the most critical interfaces in the system. The level of abstraction is such that it may be necessary to retrace back to items 1 and 2 to establish firm foundations on some details previously overlooked.

4. Measurement and Breaker real-time interfaces

Temporary programming of this function may be initially necessary for item 2. Final programming and testing can proceed after item 3 has been completed.

5. Display devices debugging

At this stage production programming and testing can proceed on display modules.

6. Data bases and measurement validity checks

After leaving the items 1, 2 and 3 loop, full system parameter checkout should be done. Impedances, transformers, breakers and measuring scale factors data bases must be finalized at this stage. The Detector-Identifier team of programs was found most useful in detecting and isolating many data problems. Psychologically, the confidence in the capabilities of the whole process is built here.

7. On-line Operation

Closed-loop on-line testing can start at this stage and be used in conjunction with item 6 for data base checking. Interfaces with display modules and their testing can be carried out. The full implication of on-line operation will become quickly evident as numerous details originally overlooked will have to be coped with. The amount of debugging effort involved in this step is of the order of 30% of total debugging time. Many of the problems that are encountered are solved by making changes in many modules. Given the complexity of the programming system being tested, it is in many cases difficult to relate symptoms to causes. This is especially so for problems that are related to specific system conditions that do not exist at all times making it difficult, if not impossible to duplicate for further testing. These problems appear sometimes to be random. The possibility always exists that it is not a programming problem but a hardware one, and special tests may have to be designed to answer this question.

8. Program sequencing, structure and scheduling

In performing the tests of item 7, much is learned from the operation of the system in relation to the organization of the various modules and the sequencing of programs for specific system conditions to achieve added efficiency. This new knowledge translates into changes in various programs requiring further verification.

9. Final system structure

Item 8 slowly converges to the final system structure as the appearance of problems start being less frequent. Changes in interfaces for extracting added flexibility and small changes in design derived from on-line operation experience are the last implementation and debugging chores.

CONCLUSIONS

The paper has summarized the experience gained by the authors in implementing the AEP real-time monitoring system. Although the main algorithms had been developed and reported in the literature, the final implementation produced a wealth of knowledge that is summarized here. This includes many helpful hints for others who have or will be faced with responsibilities in implementing state estimation monitoring systems. The original algorithms have proven to be adequate for the function they were designed to perform.[15-16]

ACKNOWLEDGEMENTS

The authors wish to acknowledge the excellent programming efforts of Sara Elam and Lewis Nowitz in the Teleprocessing and Display modules respectively. In addition, Dr. Hyde M. Merrill is gratefully acknowledged for many helpful comments made during the implementation phase.

REFERENCES

1. G. W. Stagg, J. F. Dopazo, O. A. Klitin and L. S. Van Slyck, "Techniques for the Real-Time Monitoring of Power System Operations" IEEE PAS-89, pp. 545-555, April, 1970.

2. J. F. Dopazo, O. A. Klitin, G. W. Stagg and L. S. Van Slyck, "State Calculation of Power Systems from Line Flow Measurements", IEEE PAS-89, pp. 1698-1708, September/October, 1970.

3. J. F. Dopazo, O. A. Klitin and L. S. Van Slyck, "State Calculation of Power Systems from Line Flow Measurement, Part II", IEEE PAS-91, pp. 145-151, January/February, 1972.

4. J. Dopazo, O. Klitin, A. Sasson, L. Van Slyck, "Real-Time Load Flow for the AEP System", Paper No. 3-3/8, 4th Power Systems Computation Conference Proceedings, Grenoble, France, September, 1972.

5. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin and A. M. Sasson, "Justification of the AEP Real-Time Load Flow Project", IEEE PAS-92, pp. 1501-1509, September/October, 1973.

6. A. M. Sasson, S. T. Ehrmann, P. Lynch and L. S. Van Slyck, "Automatic Power System Network Topology Determination", IEEE PAS-92, pp. 610-618, March/April, 1973.

7. A. M. Sasson, S. T. Ehrmann, "Topological Considerations for Real-Time Monitoring of Power Transmission Networks", Proceedings of the IEEE International Symposium on Circuit Theory, Toronto, April, 1973.

8. L. S. Van Slyck and J. F. Dopazo, "Conventional Load Flow Not Suited for Real Time Power System Monitoring", Proceedings of the 8th IEEE PICA Conference, June, 1973.

9. J. F. Dopazo, O. A. Klitin and A. M. Sasson "State Estimation for Power System: Detection and Identification of Gross Measurement Errors", Proceedings of the IEEE PICA Conference, June, 1973.

10. J. F. Dopazo and A. M. Sasson, "AEP Real-Time Monitoring Computer System", Symposium on Implementation of Real-Time Power System Control by Digital Computer, Imperial College of Science and Technology, London, September, 1973.

11. J. F. Dopazo and A. M. Sasson, "Least Square Estimation Applied to Power Systems," Seminar on Computer Methods of Power System Analysis and Control, Bucharest, Romania, June, 1974.

12. H. L. Fuller and T. A. Hughes, "State Estimation for Power Systems with Mixed Measurements," IEEE Paper No. C 74 363-8, Summer Power Meeting, Anaheim, California, 1974.

13. F. Aboytes and B. J. Cory, "Identification of Measurement, Parameter and Configuration Errors in Static State Estimation," Proceedings of the 9th IEEE PICA Conference, June, 1975.

14. S. Ionescu and L. Radu, "State Estimation from Line Flow Measurements and Zero Injection Pseudomeasurements, "Seminar on Computer Methods of Power System Analysis and Control, Bucharest, Romania, June, 1974.

15. J. F. Dopazo, S. T. Ehrmann, A. F. Gabrielle, A. M. Sasson and L. S. Van Slyck, "The AEP Real-Time Monitoring and Control Computer System," to be submitted for the IEEE Winter Meeting, 1976.

16. J. F. Dopazo, S. T. Ehrmann, A. M. Sasson and L. S. Van Slyck, "The AEP State Estimation and Security System," to be presented at Computerized Operation of Power Systems Conference, Sao Carlos, Brasil, August, 1975, (To be published in book form by Elsevier, Amsterdam, 1976).

Combined Discussion[1,2]

G. F. Wollenberg (Power Technologies, Inc., Schenectady, New York): I would like to thank the authors of this paper for their willingness to publish some facts and figures relating to the development of the AEP real time monitoring and control computer system. This was a real innovative project which pushed the state-of-the-art considerably in regard to the software development and I commend the authors on their achievement.

To me, the value of this particular paper lies in what it reveals about the process of developing real time systems. More specifically, it is very useful to compare the figures presented to those which might have been presented had the entire system been purchased rather than developed internally in AEP. Certain factors become immediately apparent and I will try to summarize them under the terms cost, schedules, and risks.

Cost: The personnel cost of $400,000 for 5000 man-days as quoted by the authors results in a personnel cost rate of $80 per man-day. This figure is probably in the range of twenty-five to fifty percent of the rate which might be charged by a supplier for such work. Clearly, a different method of accounting is used to obtain the $80 per man-day figure. Differences in accounting between utilities and suppliers are to be expected, of course, but they must be factored in when comparing costs for internal versus external development work. It is also apparent that only the direct costs of the equipment needed specifically for this phase of the system's development were included in the paper. I mention this because there is a tendency in purchasing such systems from suppliers to purchase the entire system at one time. Thus, in comparing costs one should bear in mind that the cost of remotes and mini computer data concentrators are not included.

Schedules. Scheduling of complex systems projects is a very difficult task. This is especially true when one is pushing the state-of-the-art as was the case with the authors' project. Very often today a system supplier takes on a project with a fixed time schedule. The contract for the project may include a system of incentives and penalties which are a function of how closely the schedule is met. Without doubt, the members of any group carrying out a similar development within a utility are asked to meet time schedules. I rather suspect, however, that compliance with such an internal schedule has far less weight than it would were the schedule imposed upon an outside supplier. The 1970-1975 period does not seem unreasonable when one considers that power system state estimation was in its infancy in 1970. However, if such a project had been awarded to an equipment supplier, the supplier would probably have been expected to meet a much tighter schedule.

Risk: As indicated above, power system state estimation was in its infancy in 1970 when the authors started this project. This means that there was considerable risk involved as to whether their ultimate objectives would be met. In fact, it would not be at all surprising to learn that the details of the project objectives changed radically over time as new developments opened new opportunities for enhancing the system. For example, did the authors foresee having a complete bad measurement detection and identification scheme when they started? This is an important point when purchasing such a system from an outside supplier who has to minimize risks if he is to meet a fixed schedule at a fixed price. When such risks do exist they are best handled by carefully identifying them and treating them differently from other functions as to their schedule and cost.

In summary, let me again commend the authors on their work, the system described is one of the most advanced systems operating at this time. Comparison of the figures presented here and those which may come from an outside supplier cannot be made without adjusting for differences in cost accounting, scheduling, and risks involved.

Manuscript received February 17, 1976

H. H. Happ (General Electric Company, Schenectady, New York): To my knowledge, this is the first implementation of State Estimation in a Real Time Computer. It represents a real achievement.

Costs for implementation both for Hardware as well as what the authors call Development were included and will be read with interest by people who are considering a similar installation.

Since the figures given may include the costs required to develop the basic technology used in state estimation, it would be helpful if the authors could estimate what it may require in developmental costs starting from the present state of the art to implement the software indicated.

In the Section entitled Future Developments, the authors indicate that they plan to develop a stochastic load flow for contingency evaluation to be executed upon operator request. They also plan to develop a security load flow for corrective strategy. Two questions come to mind to which the authors may wish to respond:

1. Is it correct to assume that a deterministic contingency evaluation program[6] using the data base generated by the state estimator is presently not being planned.

2. Is the security load flow as presently envisioned limited to corrective strategy, or will it be used as an optimal load flow also.

Manuscript received February 13, 1976.

[1] J. F. Dopazo, S. T. Ehrmann, A. F. Gabrielle, A. M. Sasson, and S. S. Van Slyck, The AEP Real-Time Monitoring and Control Computer System, this issue, pp. 1612.

[2] J. F. Dopazo, S. T. Ehrmann, O. A. Klitin, A. M. Sasson, and L. S. Van Slyck, Implementation of the AEP Real-Time Monitoring System, this issue, pp. 1618.

REFERENCE

[6] "Economy Security Functions in Power System Operations", a committee report, IEEE Special Publication, 75 CHO 969-6 PWR.

J. B. Ring (Leeds & Northrup Company, North Wales, PA): This description of the design concepts and real-time operation results of AEP's control system is most interesting and educational. Some additional information in some areas would have been helpful in evaluating the performance of the system. For instance, does the state estimator run on a periodic basis or upon a status or load change detection? If the former, what is the period? If the latter, what is the average number of state estimator runs in a 24-hour period? Indeed, a summary of the 150 application programs giving the functions performed and the period of execution would be desirable.

Concerning the authors' comments on their use of the 64K of 16 bit core, additional data concerning data base sizing is needed. It appears from the results given in the paper that the state estimator is currently sized for 57 buses and 115 lines. Other real-time digital systems have been implemented which, for instance, run 500 bus, 1000 line load flows in 48K of 32 bit core, and 1100 bus, 1667 line load flows in 80K of 32 bit core.

Finally, is it anticipated to replace the analog AGC with a digital system?

Manuscript received February 12, 1976.

H. Siemaszko (Instytut Energetyki, Warsaw, Poland): The authors have presented very interesting paper dealing with practical aspects of the AEP-Estimator being implemented and tested for something like three years.

This "hardware-software measurement tool" cost "approximately 5000 man-days" which means nearly nothing in comparison with the facilities and possibilities of the AEP-Estimator as far as more exact power systems operations is concerned.

It would be interesting to obtain some information as to further development of the AEP-Estimator. There is no doubt that the existing "measurement tool" is operation orient[illegible]one but is it possible to establish with the aid of this tool perhaps [illegible]oriented one for the results and calculations concerning the short term planning of the basic system elements like: the more exact dates of individual lines construction, the more exact dates of sub's construction more exact voltage regulation parameters and as far as phase shifting is concerned as well. If yes it could bring significant savings in investment costs.

Manuscript received February 11, 1976.

E. Handschin (University of Dortmund, 46 Dortmund; FRG): The authors of this paper and the companion paper (F 76 086-9) are to be commended for a distinct contribution to the problem of on-line estimation and system monitoring. This kind of paper is especially important because it does not describe what could be achieved with modern theoretical results but what has been realized with a sound combination of theory and practical engineering knowledge. When implementing a state estimation scheme – independent whether it is based on a weighted least-squares criterion or on the line-flow-measurements-only-method – a very important practical question is concerned with the accuracy of the measurements. Indeed, not all the measurements are taken with the same type of instrument and transmission ways may differ from one switchyard to another. Any comments concerning this problem would be appreciated.

A very important practical question deals with the problem whether the estimator is able to find a unique solution or not. However it is believed that the meter location problem must not only be treated from this point of view but also with respect to a high probability for detecting bad data. Ref. [1] describes a method where a systematic procedure

Manuscript received February 17, 1976.

is given to find an optimal meter configuration. Is it possible to generalize the examples given in the paper for radial lines and loops?

The AEP computer system is based on only one computer. In case of a failure caused by hardware or software errors what are the quantities displayed on the display board? It is felt that the display of line flow measurements only is not sufficient for an emergency dispatch solution. On the other hand, the display of infeeds or loads to the operator when the computer is not working, may require these measurements. What are the operators' experiences with line flow measurements only when the computer system is not available.

REFERENCE

[1] E. Handschin and C. Bongers: "Theoretical and Practical Considerations in the Design of State Estimators for Electric Power Systems", in S.C. Savulescu (editor): 'Computerized Operation of Power Systems', Elsevier, Amsterdam, 1976

H. Duran (Sistecom LTDA, Bogota): One aspect of the AEP detection and identification scheme that has been criticized in discussions to earlier papers is the use of the scaling "constant" σ^2 which multiplies the error covariance matrix W^{-1}[1]. It is apparent from equations 10, 11 and 12 that the authors still think that this procedure plays an essential role in bad data identification. As far as detection it seems that there is no longer such a claim.

It appears that the σ^2 procedure is not essential to the identification scheme either, since the measurement with the largest value given by expression 10 does not change if all the normalized residuals are divided by the same constant s. There-fore if the factor s is ignored the data identified as bad would be the same provided there is at least one measment which satisfies the corresponding inequalities.

It must be observed at this point that the test given by 10 should be changed to.

$$\frac{(\bar{V}_m - \bar{V}_c)_j}{\sqrt{D_j^{-1} - \text{diag}\,[A(A^T DA)^{-1}A^T]_j}} \rightarrow \eta 6/2$$

if the error covariance matrix is assumed known and constant. η 6/2 is the value of a normal variate with zero mean and unit-standard deviation and for a probability of confidence b. It is well known that $t_{k,b/2} > \eta_{b/2}$ for all values of k; also $s>1$ if bad data has been detected. Hence if inequality 10 is satisfied, the inequality above is also satisfied, showing the equivalence of the identification schemes.

The "line-flow only" scheme for state estimation is justified by the resulting simplicity of the computational algorithm and the consequently lower computer time and storage requirements. The price paid by this method is its lack of flexibility to accommodate other types of measurements, such as injections, which, in some cases are instrumental in reaching the minimum level of local redundancy necessary to identify bad data. This is particularly true in the case of radial networks, as the authors point out. With the trend towards lower communications and hardware marginal costs it is worth asking how do the trade-offs between increased flexibility and functionality and the corresponding bigger computational requirements balance out today. Could the authors comment on this?

REFERENCE

[1] F. C. Schweppe, E. Handschin, V. Kohlas, "Discussion of" State Estimation for Power Systems: Detection and Identification of Gross Measurements Errors", IEEE Trans PAS vol. PAS-93, p. 21, 1974.

Manuscript received February 11, 1976.

Paul H. Haley (Westinghouse Research, Pittsburgh, PA): The authors' work is valuable not only for the knowledgeable application of statistical and mathematical analysis to power system monitoring and control, but also for the very useful and informative results about their experiences in applying these methods. This discusser would like to add to this paper by elaborating on a few topics given in the paper.

1) In reference 1, the advantages of interpreting the solution of the equation

$$\bar{E} = (A^t DA)^{-1} A^t D \bar{v}_m \qquad (1)$$

in terms of an equivalent network of resistors with branch elements having conductances equal to the elements of D, with an element-node incidence matrix A and with series branch voltages $\bar{v}_m$ are discussed. These advantages are applicable here. Briefly, the line flow measurement system of Figure 1 induces the equivalent circuit system of Figure 1b

Manuscript received February 11, 1976.

Setting $[g_{ij}] = D$, that is giving the conductance elements values equal to the weights, we see that:

a) branch currents, $\bar{i}_g = [\bar{g}_{ij}]\,\bar{v}_m \qquad (2)$

b) Residuals, $\bar{R} = \bar{v}_m - \bar{v}_c = [g_{ii}]^{-1}\,\bar{i}_g \qquad (3)$

where $[\bar{g}_{ij}] = [g_{ii}] - [g_{ii}]\,A(A^t[g_{ii}]A)^{-1}\,A^t\,[g_{ii}] \qquad (4)$

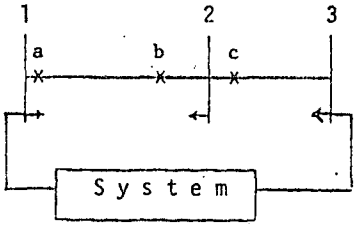


Fig. 1a Portion of Measurement System

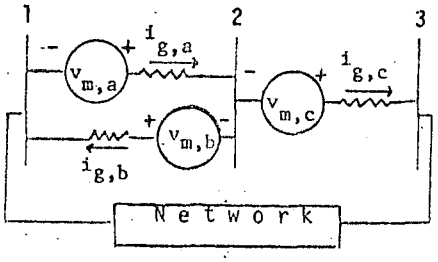


Fig. 1b Portion of Equivalent Circuit

The branch currents $\bar{i}_g$ are a function of the errors in $\bar{v}_m$ and are unrelated to the line currents of the actual system. If $\bar{v}_m$ is a consistent set of measurements $\bar{i}_g = 0$. A simple circuit interpretation exists for the normalized residuals. Let

$$[\sigma_{ii}^2] = \text{diagonal}\,(D^{-1} - A(A^t DA)^{-1}A^t) \qquad (5)$$

and

$$[\bar{g}_{ii}] = \text{diagonal}\,([g_{ii}] - [g_{ii}]A(A^t[g_{ii}]A)^{-1}A^t[g_{ii}]) \qquad (6)$$

then

$$[\bar{g}_{ii}/g_{ii}^2] = [\sigma_{ii}^2] \qquad (7)$$

and

normalized residuals, $\bar{R}^n = [\sigma_{ii}]^{-1}(\bar{v}_m - \bar{v}_c) = [\bar{g}_{ii}]^{-1/2}\,\bar{i}_g \qquad (8)$

or

$$\bar{R}^n = [\hat{g}_{ij}]\,([\bar{g}_{ii}]^{1/2}\bar{v}_m) \qquad (9)$$

where

$$[\hat{g}_{ij}] = [\bar{g}_{ii}]^{-1/2}\,[\bar{g}_{ij}]\,[\bar{g}_{ii}]^{-1/2} \qquad (10)$$

The diagonal element $\bar{g}_{ii}$ is the driving point admittance seen by the corresponding voltage source when all other voltage sources are shorted. The normalized matrix $[g_{ij}]$ is such that

$$\hat{g}_{ii} = 1 \qquad \text{for all } i \qquad (11)$$

$$|\hat{g}_{ij}| \leq 1 \qquad \text{for all } i \neq j \qquad (12)$$

since $[\bar{g}_{ij}]$ is positive semi-definite. The detection and identification of bad data is easily understood in terms of the circuit equations and Fig.2.

[illegible] we may assume that all voltages are zero with the exception of one voltage, $v_{m,a}$. By use of equations (9), (11) and (12) we see that

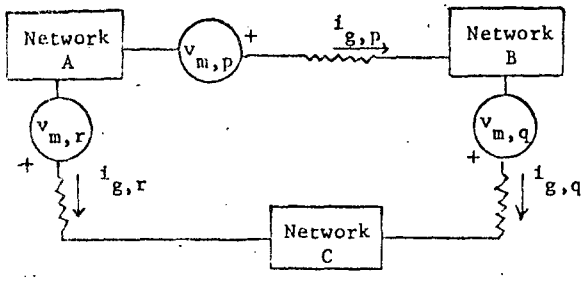
$$R^n_a = (\bar{g}_{aa})^{1/2} v_{m,a}$$

and

$$R^n_i = \hat{g}_{ai} (\bar{g}_{aa})^{1/2} v_{m,a} \quad \text{for } i \neq a$$

Hence

$$|R^n_a| \geq |R^n_i| \quad \text{since } |\hat{g}_{ai}| \leq 1.$$



$$\bar{g}_{pp} = \bar{g}_{qq} = \bar{g}_{rr}$$

$$|i_{g,p}| = |i_{g,q}| = |i_{g,r}|$$

$$|R^n_p| = |R^n_q| = |R^n_r|$$

Fig. 2 Measurement Elements in Common Series Loop

The equality sign holds when series loops exist such that the current in one branch equals in magnitude the current in other branch elements as illustrated in Figure 2 and expressed by equation (8). Obviously the driving point admittance seen by each of the voltage sources is the same; hence the magnitude of the residuals is the same. In such a case one can detect the existence of bad data but not isolate or identify its location. By reiterating the estimation equations with low weight values for the high residual measurements, one can get voltage estimates which are consistent within each measurement "island" but, the relations between "islands" will be corrupted by the error. The radial line with two measurements is a special case of a loop. Bad data can be detected but the bad line measurement cannot be identified.

If measurements are added to eliminate series paths, then the strict inequality holds in equation (11). From circuit considerations it is easy to show that if branch a is not in any common loops, then

$$\bar{g}_{aa}\,\bar{g}_{ii} > (\bar{g}_{ai})^2 \quad \text{for } i \neq a$$

and thus

$$|\hat{g}_{ai}| = |\bar{g}_{ai}/(\bar{g}_{aa})^{1/2}(\bar{g}_{ii})^{1/2}| < 1$$

2) By a casual look at the expression

$$D_j = W_j |E_p|^2 / |Z_j|^2 \qquad (13)$$

one might conclude that the weights D_j are wildly varying functions of the series line impedance Z_j. However if the full scale reading is expressed as a proportionality factor times $|Z_j|^{-1}$ then equation (41) of the authors' paper shows

$$W_j = k^2_j / |Z_j|^2$$

and thus

$$D_j = k^2_j |E_p|^2$$

The significant question is, "How much do the proportionality factors k_j or the weights D_j vary within a group of network elements, i.e. from line to line, from transformer to transformer?" By setting all of the weights D_j equal to the same constant a simple suboptimal detection scheme may be implemented. When the weights are equal, the unnormalized residuals R may be used to detect bad data. With only a single bad measurement, the largest residual will correspond to the bad measurement[1]. The bad data may be removed one by one iteratively and on the final iteration the exact value of the weights, given by equation (13), used for the good data. This method avoids the need to recompute the scaling factors σ_{ij} after a change in the estimation equation due to the elimination of a measurement. The one disadvantage to this scheme is that wider limits that are effectively nonuniform among the measurements must be used in deciding whether to accept or reject a measurement. Assuming that the gross errors have a much broader distribution than the gaussian components, the percent of gross errors not detected by this suboptimal scheme is slight.

Perhaps the authors could comment on whether the k_j's are narrowly or broadly distributed for their system and whether the assumption of a broad distribution for the gross errors is valid. Answers to these questions will help decide the viability of the proposed suboptimal bad data detection scheme.

REFERENCE

[1] Paul H. Haley and Mark Enns, "Power System State Estimation: Generalizations of the AEP Algorithm with Improved Bad-Data Suppression" Paper C 73 479-3 IEEE PES Summer Meeting: Vancouver, B.C. Canada; July, 1973. Abstract in IEEE PAS-92 November/December 1973, p. 1811.

I. S. T. Despotović (Electrical Research Institute Nikola Tesla, Belgrade, Yugoslavia): The paper is very useful and interesting from the theoretical point of view as well as from the practical point of view. It has been the result of a considerable theoretical and practical experience of the authors during the final implementation of the AEP real-time monitoring system. The theoretical and practical information and experience as well as the original algorithms presented in the paper, it is necessary to point out, will be of a great use to both researchers and those who have or will be faced with responsibilities in implementing state estimation monitoring systems.

The authors are to be congratulated for an excellent and very useful paper.

2. The AEP Real-Time. . . . . Paper Number F 76 086-9.

S. T. Despotović (Electrical Research Institute Nikola Tesla, Belgrade, Yugoslavia): The paper has given design criteria of the AEP real-time monitoring and control computer system, a description of the computer system hardware and software as well as real-time results and computing system availability performance. Given design concepts and future developments may be very useful to researchers and designers, as well as to companies which are planning a control computer system.

The authors are to be commended for contributing the valuable and useful paper in the area of the design of the real-time monitoring and control computer system.

Manuscript received February 11, 1976.

George Gross, Ronald P. Thompson, and Richard H. Webster (Pacific Gas and Electric Company, San Francisco, CA): The sharp increases in production costs experienced recently by utilities coupled with the imposition of considerably tighter constraints in the operation of power systems have resulted in increasing emphasis on economic production scheduling. A prime requirement for the optimal economic dispatch of a power system is the availability of an accurate real-time data base. The authors are performing a valuable service to the engineering community by imparting their experiences in the implementation of the AEP real-time monitoring and control system.

The state estimation program of the AEP system uses the line flow measurements (real and reactive) taken at the ends of each transmission line. Reliance on this limited set of measurements results in a rather low level of redundancy. In general, the quality of the estimates of a state estimator increases with the level of redundancy. Large errors in estimating the state variables can be caused by missing and bad data. This is particularly true for systems with a minimal number of parallel paths when line flows are the only data available. If, on the other hand, additional measurements such as real and reactive power injections and voltage magnitudes at each node were available, processing of this more redundant data would result in an improved state estimate. Also, the problems associated with radial portions of the system would be less pronounced. Once the communication links are established, the costs associated with the acquisition of these additional measurements can be justified in terms of the improvement in the accuracy of the state estimation. Of course, we realize that certain algorithmic modifications

Manuscript received February 10, 1976.

would become necessary; while those associated with the use of the injection measurements would be quite substantial, the changes associated with using the nodal voltage measurements would be rather small. Has such additional data at AEP not been utilized because of the relatively small dedicated computer used?

In the so-called emergency mode a power system is subjected to a variety of rapidly varying conditions. How was the state estimator used during periods of rapid changes in the power system network configuration? What uses have been made of the state estimator for post mortem analysis?

A. Le Roy (Electricite de France, Paris, France): I have read with great interest the two papers on state estimation. I compliment the authors for their new job and, more generally, for their significant contribution in the development of real-time static estimation.

However, I have some questions and comments. First, I am surprised that network topology and estimated quantities are not refreshed once displayed on the CRT's. In our systems we always display updated data, in accordance with the cycle of acquisition for measurements and the cycle of calculation programs for computed quantities. Secondly, it is mentioned that when a measurement is missing or bad, the measurement on the other side of the same line is used in its place. This means that the same measurement is used twice and this is a questionable procedure. Thirdly, the proposed benefits of having additional voltage magnitude measurements is not confirmed by our experience. We think voltage measurements are not accurate enough to be used in the same way as reactive power components. If instead of having DA=Db=Dv, it was 10Dv=Da=Db the conclusions of the paper would not be so clear.

At last I have noticed some typist mistakes:

– Eq (33) $$R_a^n = \frac{R_a^n}{s(2/3)\ D^{1/2}}$$

Instead of $$R_a^n = \frac{R_a}{s(2/3)\ D^{1/2}}$$

some lines further " if $R_b^n > R_a^n$ then $R_b^n > R_a^n$"

instead of " if $R_b > R_a$ then ..."

– on fig. 4 node numbers are omitted.

Manuscript received April 21, 1976.

J. F. Dopazo, S. T. Ehrmann, A. F. Gabrielle, A. M. Sasson, L. S. Van Slyck, and O. A. Klitin: The authors are pleased with the positive response and suggestions made by the discussors. We will comment on the discussions in the order they appear.

Dr. Wollenberg inquires about the accounting method used to arrive at our cost figures. The personnel costs of $80/per man-day – cited in the paper – reflect the composite direct salaries of the personnel involved during the 1970-1975 time span. Without question it does not equate with charges that a supplier would need to charge to cover risk, profit and overhead.

As mentioned by Dr. Wollenberg, the scheduling of complex state-of-the-art systems is a very difficult task. The project was approved on the basis of its economic, technical and operational feasibility. However, much system analysis and engineering remained to be carried out at the time of approval. In addition, the project encompassed the replacement of existing economic dispatch, data logging and interconnection transaction study mode programs and installing them in the new computer system. This of course required some systems analysis to take advantage of new techniques and a completely new system design. The effort to achieve all these tasks – including inventing the future – when carried out by adequately qualified people in a motivating environment results in the best schedule attainable.

Dr. Wollenberg is correct that the research aspects of this project involve risks which a supplier could not be expected to meet at a fixed price and with a firm schedule. However, those same risks had to be met internally in terms of price since our feasibility study committed us to management that the project could be successfully achieved at a fixed capital cost for data acquisition and a small scale process control computer and associated display peripherals. However, we were allowed some leeway in terms of schedule so long as the energy control center personnel were not impacted.

Manuscript received April 21, 1976.

At the heart of Dr. Wollenberg's perceptive questions concerning cost, schedules and risk is the issue of inside or outside development of computer projects. However computer projects are similar to other power company projects. If a power company makes a committment to carry out its own full range of engineering services, it follows that it should have its own staff of engineers devoted to the application of computer analytic techniques, since engineering and the computer are integrated.

Dr. Happ asked for an estimate of costs to others now that our research work has been published and implemented. While the development of the algorithms is an important step that can now be by-passed, we feel that there is a considerable gap between having them and having a finished implemented system. It is not possible to condense in papers all the experience we obtained while carrying out the project. A prospective group implementing a similar project will find that a deeper understanding of all the issues involved and their interrelations will develop slowly as their work progresses. Their savings in relation to our costs might be in the 10% to 30% range.

A recent paper[1] describes some of our current thoughts on contingency evaluation and the related subject of external system equivalence which are based in non-deterministic formulations. Our security load flow is of an optimal load flow type.

Mr. Ring inquires about the periodicity of running the state estimation programs. At present, the programs are scheduled in two ways, (1) every time a circuit breaker status change occurs and (2) every five minutes. If in the former case the breaker change does not change network configuration and the validator find the previous results are still valid, a state estimation calculation is by-passed even if the programs had been scheduled.

In the latter case, the validator program may also inhibit a state estimation calculation. When a calculation is made, timings normally vary between 50 and 100 seconds depending on which modules are called upon. Data base dimensions are currently for 100 busses, 160 line flow measurements and 320 circuit breakers and programs fit in a 10K partition with some room left for future expansions. We have no plans at present to replace the analog AGC with a digital system.

Dr. Siemaszko's comments are appreciated. Indeed the state estimation monitoring results are being used for calibrating off-line load flows for short term planning studies and for general information on the performance of our system. These applications were not part of the original project plans.

Dr. Handschin refers to his excellent paper on the question of meter placement from the point of view of bad data detection and identification. He presents a scheme whereby the covariance matrix of the measurement residuals is analyzed and shows that the effect of bad data does not spread very far from its physical location. He then determines the probability of identifying bad data and how it can be increased by better meter placement, which usually means a more even spread of meters throughout the system. We have observed similar behavior on the effects of bad data. Our studies led to defining the concept of local redundancy as one that is more meaningful than global redundancy. Meter placement then reduces to the simple criteria that to detect a bad data at least one other measurement must exist nearby that would interact and prevent the residual at the bad data point from reducing to zero. To identify properly, at least two additional measurements must exist that would verify each other and thus signal at the bad data one. A knowledge of the estimation procedure is sufficient to determine which measurements interact with each other under radial or loop configurations. On the point of measurement weights, the D weights are normally in the 10^6 and 10^8 range for our lines and voltage levels. At lower voltage levels these numbers will become smaller. Finally, Dr. Handschin comments on the need for injection measurements, specially for display as raw measurements when there are computer software or hardware problems. Based on more recent on-line experiences, we can say that the greater than 99% availability of our control computer over a 27 month period has indicated that the line flow plus voltage measurement system is an adequate one. Moreover, because of the frequent existence of one or a few more bad data points, it was considered improper to display raw data.

Dr. Duran comments on the scaling constant σ^2. This gives us the opportunity to clarify our position on this item that has been somewhat controversial. When there is bad data present, the expression in his inequality yields large numbers, much greater than η or t. The value of σ^2 is also large, and thus its effect is one of scaling down the expression. Under certain marginal circumstances or when some types of model errors are present, the detection test is failed but all measurements pass the identification test. We take note of this situation, request a new set of measurements and, if it persists, consider the system to be under some software failure or measurement system problem condition. This occurred frequently during the implementation testing phase and was considered valuable in diagnosing problem areas. Under real time operation, we have rarely encountered this situation and when it has occurred, it has not presisted, thus recovering normal operation when the next set of measurements came in. On the question of injections, it is known

their use. To their reduced accuracy they are not very helpful in the estimation process but there is no doubt that they can be useful for bad data identification. We would recommend that if their use can be avoided and still you are able to identify bad data properly, then their impact on the algorithm efficiency should not be allowed.

Dr. Haley presents an interesting scheme for interpreting the estimation concepts with an equivalent circuit analogy. We are sure readers of our paper will find it enlightening. On the range of his scaling k values, they fall in the 10^3 to 10^5 range for our measurements.

We are very grateful to Dr. Despotovic for his remarks.

Messrs. Gross, Thompson and Webster refer also to the question of injection measurements. The decision on their use should be based on the ability of identifying bad data without them. It had, in our case, nothing to do with the size of our computer or partition. On our use of the estimator during periods of rapid system changes in network configuration, we point out that it was not meant to do that by design. After a breaker operates a delay of several seconds is introduced before flow measurements are taken. If during this period, or while measurements are being read, there are further breaker operations a new delay is introduced. The end result is a higher guarantee of measurement compatibility with each other and with system configuration. The hard copy provisions allow cases to be saved for future analysis.

Dr. LeRoy comments on the fact that we do not update the CRT screens after each state estimate. We on purpose wanted to avoid changing screen information at the moment it is being viewed by an operator and to allow him to obtain a hard copy for further study. However, the mimic board, with its non-numeric display, is updated dynamically and alerts the operator to serious conditions for which he then requests the most recent CRT display. We are aware that the substitution of a measurement at one end for a missing or bad one at the other end does not contribute to accuracy or redundancy. However, it does have the property of avoiding the formation and inversion of the AUDA matrix on many occasions. We agree with the comments on the weights of voltage measurements. If in a particular system the voltage weights are much smaller than those of other measurements then their effect on accuracy is negligible. However, their effect on bad data identification can be a very useful one. We appreciate and agree with the typographical errors spotted by Dr. LeRoy.

The authors wish to point out a typographical error in the paper "The AEP Real-Time Monitoring and Control Computer System". In the section entitled costs, the total monthly cost in subsection 1 should read $9,000 rather than $9,800 and the total development cost in subsection 2 should read $786,000 rather than $876,000.

In closing, may we extend our appreciation to all the discussors for their kind remarks on our work and their helpful comments.

REFERENCE

[1] J. F. Dopazo, M. H. Dwarakanath, J. J. Li and A. M. Sasson, "An External System Equivalent Model Using Real-Time Measurements for System Security Evaluation", submitted for the IEEE Summer Meeting, 1976.

J.F. DOPAZO, S.T. EHRMANN, A.M. SASSON, L.S. VANSLYCK

AMERICAN ELECTRIC POWER SERVICE CORPORATION - 2 BROADWAY, NEW YORK, NEW YORK 10004, U.S.A.

INTRODUCTION

In 1969, management at American Electric Power approved a new real-time control computer project for its energy control and dispatching center in Canton, Ohio. The objectives of the new system were to monitor the 765-345 kV EHV transmission network to provide system coordination operators with reliable, real-time information about the state of the network and the status of its components. In addition, the system would perform the economic dispatch and logging functions then operating on a second generation computer, (1-3) which was installed in 1964.

CONTROL COMPUTER SYSTEM

Figure 1. shows the control computer interface scheme. Automatic generation control is a digitaly directed analog function. The IBM 1800 computer is a data buffer and concentrator for the non-real-time logging/reporting billing functions performed in the New York IBM 370/168 computer. The state estimation EHV monitoring functions are accomplisned in the IBM 1800. They process the information received through the data acquisition system (DAS) which includes the three (HP) Hewlett Packard 2116 B minicomputers located in Roanoke (Virginia), Fort Wayne (Indiana) and Canton (Ohio). State estimation output is accomplished through ten black and white CRT's, six associated keyboards, a printer, copier and a dynamic display board. The keyboards and six of the CRT's are also used for economic dispatch functions. The printer/copier is part of the keyboard CRT system and provides hard copy identical to a CRT image. The 2.7m by 6m display board is driven by the IBM 1800 through an HP 2100 A minicomputer to reflect the following conditions:

- –open/closed status of circuit breakers.
- –open/closed or overloaded condition of a transmission line or transformer.
- –direction of watt flow in a line or transformer.

The computer system configuration together with its costs are more fully described in Reference 4. The dispatch and logging functions are not greatly different than described in (1-3). The state estimation algorithms and interfaces are described in (5) and the references listed therein.

Figure 2 shows the control center functional relationship. The state estimation system comprised the functions of network configuration, estimation, detection and identification of errors, and display. The network configurator functions are to develop the contiguity of monitored circuit elements by analysing circuit breaker and measurement status (6-10). The display board is driven after every state calculation.

THE STATE ESTIMATION ALGORITHM

In the AEP algorithm the state vector is the complex nodal voltages $|E| \angle |E|e^{j\delta}$, and the measurement vector is the complex line flows $S_m = P_m + j\,Q_m$. Calculated line flows S_c are functions of the state vector E, $S_c = S_c(E)$. The state estimate is the value of E which minimizes the scalar residual:

$$J(E) = (S_m - S_c)^{*T} W (S_m - S_c) \quad \ldots\ldots (1)$$

where:

- (–) overbar indicates a vector quantity
- (*) indicates complex conjugate
- (T) indicates matrix transpose
- W is a diagonal weighting matrix

The algorithm is extensively discussed in previous papers, as mentioned, and it was said that at its minimum J (E) should satisfy the inequality:

$$J(E) < X^2\, k,\, b/2 \quad \ldots\ldots (2)$$

where the right hand side is the value of a chi squared distribution for k degrees of freedom and a probability of confidence b. We use a b value corresponding to the 3σ point, i.e. three times standard deviation:

and $k = 2m - 2(n-1)$

where

- m = number of complex flow measurements, the dimension of S_m
- n-1 = number of complex bus voltages to estimate: the dimension of E. δ is defined at one bus, a reference, where $|E|$ is measured).

Subject to the vicissitudes of maintenance, equipment malfunction and telemeter failure, m has a changing value through time, and when several measurements are missing from a portion of the network n also changes because voltages may not be computable in the local area. Moreover, inequality (2) may not always be satisfied. This is taken to mean that bad data has been detected. Thus the DAS has two types of infidelity:

- –Some data missing, time variant where m varies with time.
- –Bad data, one or more components of S_m is grossly in error.

This infidelity of the DAS is the raison d'etre for state estimation. The monitoring system must function in the presence of these vicissitudes with as great a veracity as possible. The accuracy enhancement realized, namely:

$$\left|\mathrm{ROIM}\,(S_c - S_{true})_i\right| < \left|\mathrm{ROIM}\,(S_m - S_{true})_i\right| \quad \ldots\ldots (3)$$

where, ROIM means corresponding "real or imaginary part", is merely a filtering by-product of the state estimator and not a justification for it. (Of course, no one knows S_{true}, and the fact of the latter inequalities, (3) can only be inferred from simulation studies).

From a physical S_m, a "measured" branch voltage for each line flow measurement is obtained.

$$\nu_m = z\left(\frac{S_m^*}{E^*} - y\,E\right) \quad \ldots\ldots (4)$$

where:

- S is an element of S_m
- E is an element of E, at the bus where S is taken
- z is the series impedance of the line PI section, and
- y is the shunt admittance of the line PI section at the measurement side. (The two shunt are of different value for transformer equivalents with off-nominal tap.)

For each measurement there is also a computed branch voltage, being the difference of two nodal voltages. This can be written in terms of a measurement node incidence matrix A as:

Once bad data has been detected, it is identified as those measurements where the largest value of normalized residual branch voltage:

$$\frac{|\text{ROIM}\ (v_m - v_c)_i|}{s\sqrt{D^{-1}_i - \text{diag}\ [A(A^T DA)^{-1} A^T]_i}} > t_{k,\,b/2} \quad \ldots\ldots(6)$$

occurs. (5)

Where: $t_{k,\,b/2}$ is a Student - t distribution:

$$s^2 = J(E)/k$$

$$D_i = W_i |E_i|^2 / z^2 \quad \ldots\ldots(7)$$

The numerator of (eq. 6) is the ROIM of the branch voltage residual and the denominator is its standard deviation, a normalizing factor. The identified bad data is removed from S_m and E is then recomputed from the reduced S_m.

THE ESTIMATOR SIZE AND TIMING

The AEP data acquisition system has had dimensions as shown in Table 1 for the past year:

DATE	BUSSES	MEASUREMENTS
March '75	62	121
September '75	63	124
December '75	65	134

Table 1. DAS Size

Every second the IBM 1800 polls the three HP 2116 B minicomputers for a status indication. (4) If no breakers have changed status, flow measurements are read into the 1800 every five minutes. If the network configuration has changed, flow measurements are then read after an 8 second delay for transients to die out. This long wait is because A-D converter and multiplexor equipment has a surge protection circuit with a two second time constant (6-7). If network configuration has not changed from the prior state estimate, Sc (old) is compared with Sm (new), using (eq. 1), to obtain a residual. If this residual satisfies (eq. 2) the prior state estimate is deemed to still be valid and no further computations are carried out. It may be necessary, however, to update the map board with one or more breaker changes which did not affect network contiguity.

Thus, the state estimator is run following any network change after the system transient has died, or at five minute intervals if the prior solution is no longer valid. The five minute interval can easily be changed, but our experience has shown it to be completely satisfactory.

Figure 3 shows our experience in passing the residual test with new measurements versus the prior state estimate. As one would expect it shows that the system is most steady in the early morning hours, but becomes rather dynamic by 6 to 8 a.m. most probably due to transmission switching in preparation for line and equipment maintenance. The drop of the curve prior to midnight is generally due to load drop after a 9 to 9:30 evening peak followed by a characteristic change in system energy sales.

Recent energy transfer experience has been a daily transfer of about a 1000 MW to the south daytime and 1000 MW north nightime.

Figure 3 shows aggregate November data. The passing percentage was improved by about a four-to-one ratio due to tuning of the weighting factors. This will be discussed later in the paper.

Figure 3 shows about 240 trials per hour for a 30 day period; as the residual test is being made on an average of about 8 times per hour. Since a state estimate is iniated about every five minutes, of course, when system contiguity has changed.

The AEP system covers a geographical area about 600 Km by 500 Km. Therefore, a certain amount of data is often missing due to communication anomalies and/or maintenance. The network configurator must determine a network model appropriate to this missing data. This model may be one, two or three separated systems each including at least one of three possible voltage reference busses in order that the system state may be computed in each subsystem. As a result of missing data, certain line quantities may be computed without redundancy which would guarantee absence of errors. Therefore, a feature has been implemented to flag nonredundant results of the state estimator output. (5)

BAD DATA EXPERIENCE

Figure 4 shows bad data experience for 29 days of November 1975. This month is typical of our past years experience with DAS. It should be noted that if either watts or vars are bad we say the measurement is bad. We always try to determine the cause for bad data and we have found a full spectrum of reasons. In the beginning, naturally, there were many bad data due to miswiring, watt/var address reversals, etc. The bad data identifier logic in the state estimator was of great assistance in identifing and correcting these expected problems. (5)

When a measurement is removed as bad, we have found that it is better to leave it out for a period of time rather than expecting it to be only a momentary failure. An expedicious method to accomplish this has been to leave it flagged as bad until the end of the current hour. At the end of the hour then, we look at the flags on all measurements and log those that are "on". This log is printed once a day which lists all the measurements so detected and the number of times each had been flagged. Therefore, the maximum flag count would be 24 for any measurement.

Figure 5 shows our experience with bad measurement durati also for November 1975. There is no regular feedback to meter maintenance personnel in regard to bad measurements until the log is inspected the following morning. As an expediency while doing certain maintenance work, field personnel may allow bad data to temporarily exist knowing that the state estimator will handle it properly. This would account for some bad data up to several hours duration. In regard to particular measurements, many have very high fidelity while a few are particular "bad actors" requiring repeated maintenance. Whether this will continue through the second years' experience and beyond, we can't predict, but it has been true to date. Certainly our bad data experience alone has been enough to justify state estimation versus a read-display scheme. Bad data does not impair the veracity of state estimator results. We feel a read-display scheme with the same infidelity level would have made system operators very frustrated and dubious about using its output.

WEIGHTING FACTORS

Many of our previous papers have included a discussion of weighting factors, "W" in eq. (1). Statistically the correct weight to use is:

$$W_i = \left[(.02|S|_{true} + .0052\ FS)/3\right]^{-2} \quad \ldots\ldots(8)$$

which is the standard deviation of the measurement varience. The 2% factor is due to c.t. and potential device error as a percentage of Mva. The 0.52% factor is due to a/d converter and transducer error as a percentage full scale (FS) value. These percentages correspond to the 3σ error level of the measurement hardware.

FS, Full scale is the Mva corresponding to rated transducer output. As previously stated (5), this made bad data identification difficult so:

$$W_i = \left[.0152\ FS/3\right]^{-2} \quad \ldots\ldots(9)$$

was proposed. This resulted from feigning a constant $|S|_{true}$ equal

to .5 FS. This alleviated the identification problem caused by assigning improper weights to bad data where Sm differed greatly from S_{true}. However, it introduced a different undesirable effect.

Because the weighting factors are crucial to proper bad data identification Sm could not be substituted for S_{true}. But the computed state vector E is also a function of the weights. Therefore, hen (9) was used the solution was degraded to a certain extent (not enough to invalidate the results for transmission monitoring; but enough that occasionally, (once or twice per week) the inequality (2) would fail, yet no bad datum could be found via eq. (6)). Also, we seldom passed the residual test with the prior state estimate.

Having gained this experience the compromise was apparent. The S_{true} term in (8) must be replaced by a "base case" or "system normal" value. That is, a constant value would solve the identification problem, and a typical value would solve the detection problem. It was not necessary that S_{true} be accurate, annual review would seem sufficient. This change was made in October 1975 and it about doubled the likelihood of passing the residual test with the prior state estimate. No bad data identification failures have occurred since.

APPLICATION SITUATIONS

Two particular system conditions have made the state estimator an important tool for system operators: High voltage alerts and line loading alerts. On several occasions at both 765 kv and 345 kv levels, transformers voltages have been within 0.5% of contingency limits at a few locations throughout the system. This necessitated continuous monitoring of voltage levels. The state estimator was important for this application and system operators exhibited a great deal of confidence in it.

The line loading alerts were that heavy loading on an EHV circuit caused a concern that if that circuit were lost the loading on a 138 kv line would exceed its 20 minute contingency loading thereby requiring immediate action. Again the state estimator was the best tool for closely monitoring the alert.

System conditions projected for some three to five years in the future are for generally tighter generation capacity and heavier line loading. Security alert and system reliability concerns will make the estimator an almost indispensible tool under these conditions.

CONCLUSIONS

Bad data suppression and availability of system data when measurements are missing make the state estimator an invaluable tool for network monitoring. While the state estimator is justified because of its infidelities it also provides system data of enhanced accuracy. This is particularly important for voltage level monitoring.

State estimation together with economic dispatch and some other functions can be accomplished in a relatively small computer main frame. The CPU availability of our system has exceeded 99% for the past two years.

REFERENCES

1. G. W. Stagg, J. F. Dopazo, M. Watson, J. M. Crawley, G. R. Bailey and E. F. Alderink, "A Time-Sharing On-Line Control System for Economic Operation of a Power System," Proc. Instrum. Soc. Am., Oct., 1966.

2. J. H. Kinghorn, G. H. McDaniel and C. P. Zimmerman, "Development of Coordination and Control of Generation and Power Flow on the AEP System," Proc. Am. Pwr. Conf., 1965.

3. W. S. Morgan, A. H. Willenmar, N. Cohn and C. Nichols, "Facilities for the AEP System Power Production Control Center," Proc. Am. Pwr. Conf., 1965.

4. J. F. Dopazo, S. T. Ehrman, A. F. Gabrielle, A. M. Sasson and L. S. VanSlyck, "The AEP Real-Time Monitoring and Control Computer System," IEEE PES Winter Meeting and Tesla Symposium. paper F 76 086-9. New York, N.Y., Jan. 1976.

5. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin, A. M. Sasson and L. S. VanSlyck, "Implementation of the AEP Real-Time Monitoring System." IEEE meeting as in Ref. 4, paper F76o87-7.

6. K. A. Clements, O. J. Dennison and K. J. Ringlee, "Effects of Measurement Non-Simultaneity, Bias, and Parameter Uncertainty on Power System State Estimation," 1973 PICA Proc. IEEE, Abstract pp. 5 PAS-93, Jan-Feb 1974.

7. L. S. VanSlyck, discussion of Ref. 6 IEEE PAS-93, Jan-Feb 1974, pp. 24.

8. J. F. Dopazo, O. A. Klitin and A. M. Sasson, "State Estimation of Power Systems: Detection and Identification of Gross Measurement Errors." Proc. IEEE PICA Conf., 1973. Discussions and closures, IEEE PAS-93, Jan-Feb 1974 pp. 20.

9. A. M. Sasson, S. T. Ehrman, P. Lynch and L. S. VanSlyck, "Automatic Power System Network Topology Determination," IEEE Pas-92, pp. 610-618, Mar-Apr 1973.

10. A. M. Sasson, S. T. Ehrman, "Topological Considerations for Real-Time Monitoring of Power Transmission Networks," Proc. IEEE Int. Symp. Cir. Th., Toronto, April, 1973.

CAPTION AREA

ILLUSTRATION

CAPTION AREA

TERMINATION OF COLUMN 2

COLUMN 1 TEXT START

ILLUSTRATION

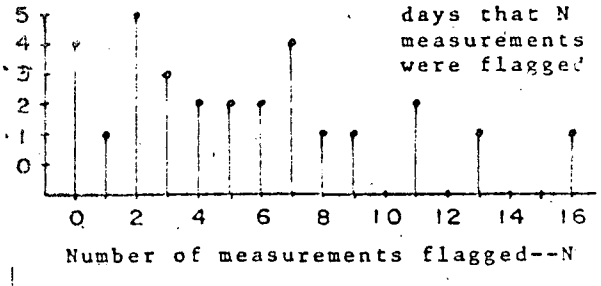


Figure 4. Bad Measurement Experience (November 1975 data).

COLUMN 2 TEXT START

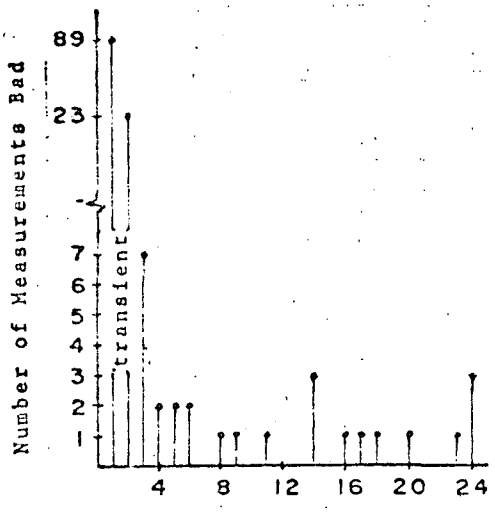


Figure 5. Bad Measurement Duration (Aggregate data for November 1975)

ILLUSTRATION

ILLUSTRATION

TION AREA

CAPTION AREA

TERMINATION OF COLUMN 2

COLUMN 1 TEXT START

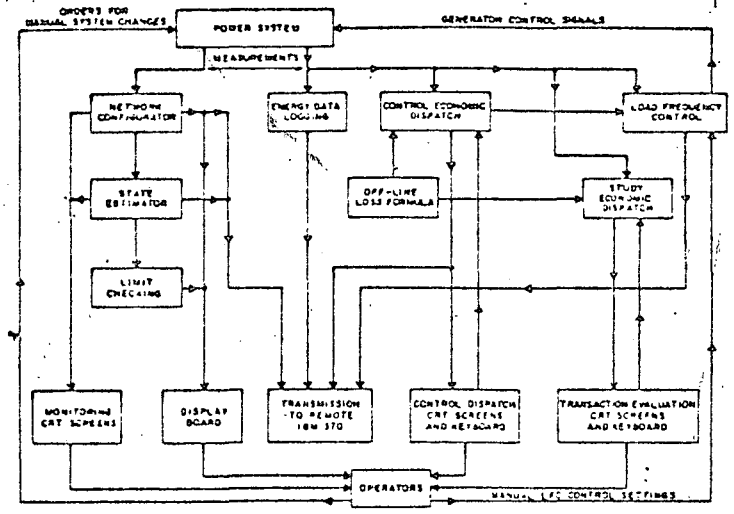


Figure 1. Functional Flow Chart of the AEP Control Center (1976)

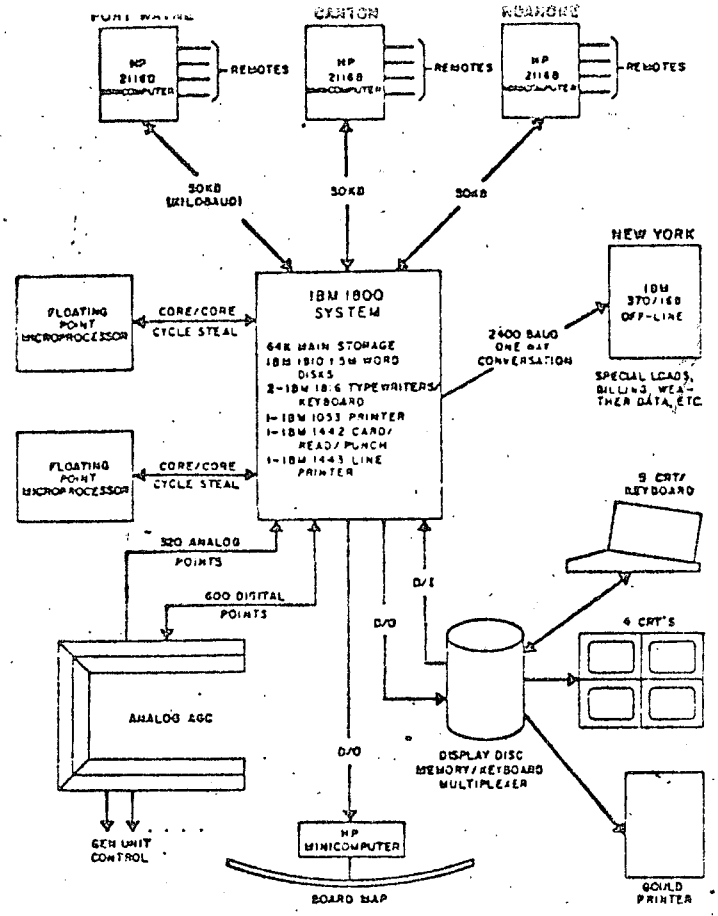
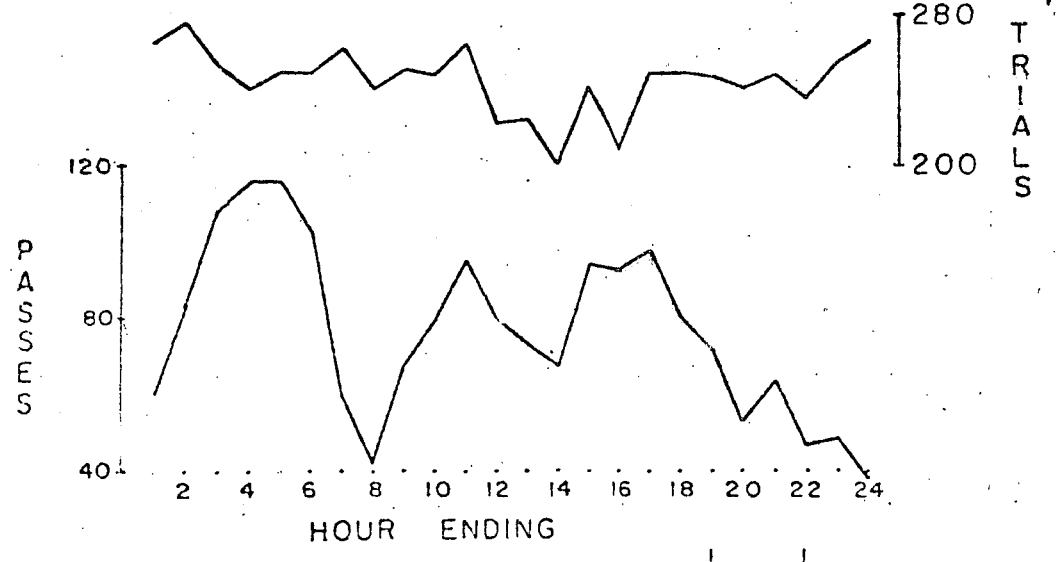


Figure 2. Control Center and DAS Hardware



ILLUSTRATION

Figure 3. Residual Test Experience of Prior State Estimate (Aggregate Data for November 1975).

CAPTION AREA

TERMINATION OF COLUMN 2

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA VI.1 USO DE EQUIVALENTES EXTERNOS PARA ANALISIS DE CONTINGENCIAS

W. F. FINNEY

ENERO, 1979.

SYSTEMS & CIRCUITS SYMPOSIUM

REI NETWORK REDUCTION FOR POWER SYSTEM APPLICATIONS

W. F. Tinney

Bonneville Power Administration
P.O. Box 3621 Portland, Oregon 97208

ABSTRACT

The REI approach to power network equivalents largely overcomes certain difficulties encountered with other equivalents and permits a more favorable trade-off between accuracy and computer requirements. In the REI approach a group of functionally related active nodes is replaced by an equivalent fictitious node. In the resulting equivalent the fictitious node partially preserves the identity and closely approximates the effect of the active nodes it replaces. It also closely approximates the original power input-output relationships. The paper outlines the basic ideas, shows how to make the REI equivalent compatible with sparsity and suggests topics for further investigation.

INTRODUCTION

Better power network equivalents are needed for steady-state and dynamic simulation applications. Although a number of different power network equivalents have been proposed, the most widely used equivalents are based on straightforward network reduction[1, 2]. This paper explains the little known REI approach which is an extension of network reduction. The designation REI, which is derived from the words Radial, Equivalent and Independent, is used by Dr. Paul Dimo of Romania for the power network equivalent scheme he developed in the early 1960's[3, 4, 5]. His work has only recently gained attention in the United States. At Bonneville Power Administration (BPA) REI equivalents are now being used in power flow and transient stability studies[6]. The REI approach offers significant advantages over conventional network reduction for power network equivalents in accuracy, computer requirements and dependability.

Power transmission systems in normal steady-state operation can be analyzed as networks composed of linear admittance elements. Power system network problems, however, are usually nonlinear because load and generation are specified as complex power instead of current and voltage. Therefore, equivalents for power network problems are approximations whose accuracies can be determined only by comparing solutions obtained with and without their use.

Because of the large difference between the power frequency and the much lower frequencies of electromechanical oscillations associated with transient and dynamic stability, steady-state phasor analysis can be used for the transmission network in most power system dynamic simulation problems. Therefore, improvements in steady-state equivalents such as the REI approach are also relevant for these dynamic problems.

The largest interconnected power transmission networks now being studied have from five to ten thousand buses (or nodes); systems of two thousand buses are common. Network problems of this size can be solved, but they tax the largest computer systems. Often only a portion of a large network is of interest. For these situations, efficient, accurate, convenient and dependable methods for reducing the problem to an equivalent representing only the area of interest would be of great benefit. Equivalents also can be useful in overcoming other difficulties associated with system size such as the lack or uncertainty of data. Present methods for power system equivalents fall short of meeting the ideal requirements for most applications.

In the REI approach a group of functionally related active nodes of a network is replaced by a single fictitious equivalent node. The equivalent node is connected through a lossless fictitious network to the group of active nodes which it replaces. The real and reactive power injections at the equivalent node are the algebraic sums of the real and reactive injections of the group of nodes. After the connection of the fictitious network, the formerly active nodes are passive and the fictitious network supplies the former injections through its connections to the original network. The passive nodes may then be eliminated by ordinary network reduction leaving only the equivalent node to represent their effect on the remainder of the original network.

In a typical application the nodes that are not needed in the output are formed into a number of groups based on their functional similarity and each group is replaced by an equivalent REI node as outlined. The resulting network equivalent is exact at a chosen operating point and a good approximation at neighboring operating points. The variables at the equivalent REI nodes are physically meaningful and can be used to monitor and control solutions obtained with the equivalent. In contrast to conventional network reduction schemes the power input, output, and losses are approximately the same for the REI equivalent as for the original network.

Early work on the REI approach was concerned mainly with reduction in number of nodes in a network and ignored its

effects on sparsity. Since computer programs for all large power system problems now employ sparse matrix techniques, it is necessary to modify the REI approach to cope with sparsity.

This paper presents the REI approach from a different viewpoint than taken by Dimo. It explains the effect of the REI approach on sparsity and gives guidelines for strategies to mitigate the associated sparsity problem. A more generalized form of the REI equivalent is suggested. Possibilities for further investigation which should be of interest to circuit theorists are pointed out.

NOTATION

Upper case letters are used for complex quantities, lower case letters for real quantities. Complex conjugate is denoted by *. All matrices and vectors are enclosed in brackets [ ]

$[Y]$	: nodal admittance matrix
$[V]$	: node voltage vector
$[I]$	: node current vector
$S_i = p_i + jq_i$	: complex power injection at node i
I_i	: complex current injection at node i
V_i	: complex voltage at node i
Y_i	: complex physical admittance of branch i, (not to be confused with transfer admittance)
Y_{ii}	: complex driving point admittance of $[Y]$
Y_{ij}	: complex transfer admittance of $[Y]$

NETWORK REDUCTION FOR POWER NETWORK EQUIVALENTS

Since the REI approach is essentially an extension of and a substitute for conventional network reduction, it is helpful to briefly review network reduction and its deficiencies in power system applications.

The relationship between node-to-ground voltages and nodal current injections in a network is

$$[Y]\,[V] = [I] \tag{1}$$

In discussing network reduction it is convenient to divide the nodes of the network into three sets as follows:

(*A*) Nodes to be retained that have no branches connecting to nodes of set *C*.

(*B*) Nodes to be retained that have one or more branches connecting to nodes of set *C*.

(*C*) Nodes to be eliminated.

With these definitions Eq. (1) can be written in partitioned form as

$$\begin{bmatrix} Y_{AA} & Y_{AB} & \\ Y_{BA} & Y_{BB} & Y_{BC} \\ & Y_{CB} & Y_{CC} \end{bmatrix} \begin{bmatrix} V_A \\ V_B \\ V_C \end{bmatrix} = \begin{bmatrix} I_A \\ I_B \\ I_C \end{bmatrix} \tag{2}$$

Eliminating V_C *gives*

$$\begin{bmatrix} Y_{AA} & Y_{AB} \\ Y_{AB} & Y'_{BB} \end{bmatrix} \begin{bmatrix} V_A \\ V_B \end{bmatrix} = \begin{bmatrix} I_A \\ I'_B \end{bmatrix} \tag{3}$$

where

$$\left[Y''_{BB}\right] = \left[Y_{BB}\right] - \left[Y_{BC}\right]\left[Y_{CC}^{-1}\right]\left[Y_{CB}\right] \tag{4a}$$

$$\left[I''_B\right] = \left[I_B\right] - \left[Y_{BC}\right]\left[Y_{CC}^{-1}\right]\left[I_C\right] \tag{4b}$$

Introducing new symbols, Eqs. (4a) and (4b) can be written more compactly as

$$\left[Y'_{BB}\right] = \left[Y_{BB}\right] - \left[Y''_{BB}\right] \tag{5a}$$

$$\left[I'_B\right] = \left[I_B\right] - \left[I''_B\right] \tag{5b}$$

The elimination of the nodes of set *C* introducts fictitious branches between and fictitious current injections into the nodes of set *B*, known as the boundary nodes. Set *A* is unaffected by the elimination of *C* because it is isolated from set *C* by set *B*.

Eqs. (1) through (5) outline linear network reduction in which the injections are currents. In power networks the given or known injections are complex power instead of current. The injection S_i at node i is

$$S_i = V_i I_i^* \tag{5}$$

If S_i and V_i are known, I_i can be determined. Therefore, various adaptations of the linear reduction scheme can also be used for power network equivalents. However, it is first necessary to obtain a solution to the nonlinear system of power flow equations in order to determine the current injections. Since the set of power injections chosen for this solution will affect the resulting equivalent, the choice of the injections for the base case must take into account the problems to be solved with the equivalent.

Assuming a suitable base case solution for the entire network is known, a power network equivalent can be derived from the linear reduction scheme. Following the procedure indicated by Eq. (4) produces an equivalent nodal admittance matrix and a set of fictitious current injections $\left[I''_B\right]$. For most power system applications these current injections must be converted to equivalent complex power injections which can be combined with the given power injections at the boundary nodes. Since the voltages at the boundary nodes are known from the base case solution, the fictitious injection S''_i at each node i of the boundary can be computed from

$$S''_i = V_i I_i^* \tag{6}$$

where V_i is from $\left[V_B\right]$ and I_i^* is from $\left[I''_B\right]$. The augmented injection S'_i at each boundary node is

$$S'_i = S_i - S''_i \tag{7}$$

where S_i is the actual injection at the boundary node i.

An important alternative of this procedure is to convert the current injections $[I_C]$ to equivalent shunt admittances before performing the reduction. This step modifies the diagonal elements of $[Y_{CC}]$ of Eq. (2) and makes $[I_C] = [0]$. Each diagonal element Y_{ii} of $[Y_{CC}]$ is changed to Y'''_{ii} as follows:

$$Y'''_{ii} = Y_{ii} - I_i/V_i \tag{8}$$

where I_i and V_i are the i-th elements of $[I_C]$ and $[V_C]$ respectively.

If the shunt admittance alternative of Eq. (8) is used, there are no fictitious injections at the boundary nodes but $[Y''_{BB}]$ is a function of the base case solution. If the distributed current scheme is used, $[Y''_{BB}]$ is independent of the chosen base case. Intermediate schemes in which some nodes of C are treated in one way and some in the other way are also used. Similarly, a fraction of the injections at each node can be treated each way. All of these alternatives influence the accuracy and other properties of the equivalent, but knowledge of the effects is largely empirical.

The power flow problem is representative of most power network applications of equivalents; therefore, the ensuing discussion will be based on it.

In the power flow problem nodes are of three basic types[7].

(1) VO nodes where $|V_i|$ and θ_i are specified.

(2) PQ nodes where p_i and q_i are specified.

(3) PV nodes where p_i and $|V_i|$ are specified.

The relationships between injections p_i and q_i and the dependent variables V_i are

$$\text{real} \left\{ V_i \sum_j Y^*_{ij} V^*_j \right\} = p_i \tag{9a}$$

$$\text{imag} \left\{ V_i \sum_j Y^*_{ij} V^*_j \right\} = q_i \tag{9b}$$

In the polar formulation each PQ node requires Eqs. (9a) and (9b), and each PV node requires only Eq. (9a). The VO nodes do not appear explicitly in the system of equations but their effect is represented in Eqs. (9a) and (9b).

The nonlinear power flow equations can be solved by several iterative methods[8] but the performance of these methods is problem dependent in ways that are not well understood and the conditions necessary for the existence of a power flow solution are not known[7]. For a given problem there may be no solution or many solutions, only one of which is physically acceptable. Problem conditions also affect the rate of convergence of solution methods. In general, it is more difficult to solve power flow problems after a network reduction than before. The following factors seem to be responsible for some of the difficulties:

(1) The elimination of critical PV nodes in set C.

(2) The great diversity of magnitudes of fictitious injections $[I''_B]$.

(3) The abnormal values of $[Y'_{BB}]$.

In general, the equivalent does not represent the effect of the eliminated system over a wide enough range of conditions with consistent accuracy.

Difficulties in power flow solutions with conventional equivalents are manifested in the following ways:

(1) The solution method fails to converge although a solution exists.

(2) The method converges to a physically unacceptable solution.

(3) The method converges to the desired solution but it takes more than the usual number of iterations.

(4) The accuracy of the solution is unacceptable.

These difficulties can be mitigated to some extent by transferring certain critical PV nodes from set C to sets A or B. If too many PV nodes are retained, it largely defeats the purpose of having the equivalent; if too few PV nodes are retained, the difficulties persist. Identification of the critical PV nodes that should be retained depends on experience.

In certain applications it would be possible to eliminate many more nodes of any of the three types if it were possible to monitor the effects of the subsequent solutions upon them or to modify their given conditions in accordance with problem requirements. This cannot be done with the conventional equivalent because the identity of eliminated nodes is lost. Accuracy is poor because the equivalent injections at the boundary nodes do not adequately reflect the influence of the eliminated system.

Another defect of conventional equivalents is that they do not preserve the power input-output relationships for the total network. In many applications it is desirable to be able to determine the total system losses as the algebraic sum of the complex power injections. Although the power flows and losses in the branches of $[Y_{AA}]$ and $[Y_{BB}]$ are the same in the equivalent as in the full system for the base case, the losses in the fictitious branches of $[Y''_{BB}]$ are not the same as in the eliminated network. The identity of generation and load is lost in the equivalent.

REI EQUIVALENT

a. Basic Development

The REI concept can be explained with schematic diagrams. Fig. (1a) shows a network at a known operating point with a set of N active nodes, having complex power injections S_i, to be made into an REI equivalent. The first step is shown in Fig. (1b) where an REI network is connected to the N nodes. The REI network has one active node R, with injection S_R, in addition to the N nodes connecting it to the original network. At this point the REI network has no specific internal structure, but it is assumed to be composed of passive linear elements with no connection to ground. The injection S_R must be equal to the algebraic sum of the N known injections and the REI network must have zero real and reactive power losses. The power flows from the REI network into the N nodes of the original network must match the N original injections; the node voltages V_i must also be the same as before. Since the N connecting nodes, as well as any hidden internal nodes, are passive, they can all be eliminated as shown in Fig. (1c) without affecting the conditions

at remaining nodes of the original network. Since S_R replaces the N original injections, the original power input-output relationship is preserved.

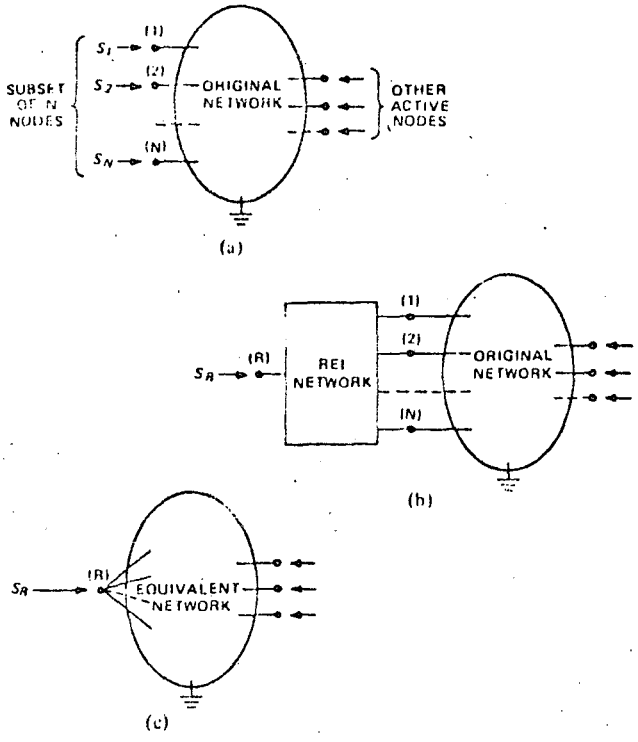


Fig. 1 (a) Original network showing group of nodes to be made into equivalent. (b) REI network connected to original network. (c) Equivalent network after elimination of passive nodes.

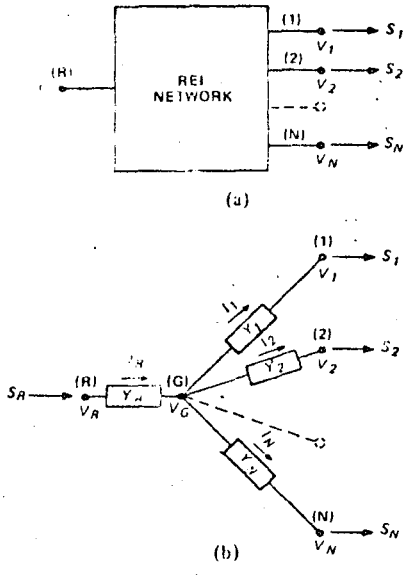


Fig. 2. (a) REI network with known terminal values. (b) Star configuration of REI network showing branch admittances, branch currents with assumed polarities and other derived quantities.

Except for some degenerate cases of no practical importance, it is always possible to establish an REI network that will satisfy the foregoing conditions. Fig. (2a) shows the general case for an REI network of N nodes whose power injections (direction reversed) and node voltages are obtained from the solution of a specific network problem. The REI network must satisfy these known conditions as its terminals. The star configuration of Fig. (2b) fulfills the requirements. It has one internal passive node G whose voltage can be assigned an arbitrary value. Dimo always lets $V_G = 0$, but a more general form of the equivalent can be obtained by allowing V_G to assume other values.

Each current injection I_i into node i must be

$$I_i = S_i^*/V_i^* \tag{10}$$

The current I_R entering node R must be

$$I_R = \sum_{i=1}^{N} I_i \tag{11}$$

The complex power injection S_R is required to be

$$S_R = \sum_{i=1}^{N} S_i \tag{12}$$

The voltage V_R must then be

$$V_R = S_R/I_R^* \tag{13}$$

If $V_G = 0$, the branch admittances are

$$Y_i = -I_i/V_i = -S_i^*/|V_i|^2 \tag{14a}$$

$$Y_R = I_R/V_R = S_R^*/|V_R|^2 \tag{14b}$$

If $V_G \neq 0$, the branch admittances are

$$Y_i = I_i/(V_G - V_i) \tag{15a}$$

$$Y_R = I_R/(V_R - V_G) \tag{15b}$$

Note that Eqs. (10) through (13) do not depend on the topology or branch admittances of the configuration chosen. Eqs. (14) and (15), of course, apply only to the star configuration. Branch currents do not have to be computed explicitly.

b. Properties

Since the node G is passive, it can be eliminated without changing the terminal conditions. Its elimination results in a mesh network in which the N connecting nodes and node R are fully interconnected. If N is large this star-to-mesh conversion can greatly increase the number branches in the equivalent. As will be explained, in some applications there is no reason to eliminate the node G unless or until it is advantageous in the strategy for conserving sparsity. For other applications it may be necessary or desirable to eliminate node G. If the REI equivalent is based on $V_G = 0$, V_G will always be close to zero when problems are solved using the equivalent. Since some iterative solution methods depend on node voltages being near to their nominal values, the retention of node G for applications using these methods may cause difficulties. In the examples in this paper node G is eliminated.

Infinitely many REI equivalents could be derived from the star configuration by the freedom of choice of V_G. One case of possible interest is when V_G is set equal to V_R. If this is done, the branch RG has infinite admittance and, in effect, vanishes. Any advantages of this special case have not been investigated.

Infinitely many topological configurations would also fulfill the REI requirements. But since all configurations would be reducible to an $N+1$ node mesh equivalent, there is no reason to consider configurations more complicated than this mesh. Whether any mesh that would not be the result of eliminating node G of the star would be of any use has not been investigated. In this paper the star configuration is assumed.

Since the admittances of the REI network are functions of the operating point of the original network and the choice of V_G, $\left[Y''_{BB}\right]$ of Eq. (5a) becomes a function of these same factors when the REI approach is used. The accuracy of the REI equivalent depends in part on how closely the computed values of its admittances approximate the values that would give exact results in each problem for which the equivalent is used. It would be desirable to know how the choice of V_G affects the performance of the equivalent. This has been examined to some extent experimentally but not analytically.

Experiment as well as reason indicates that $V_G = 0$ is probably the optimum choice for most power network equivalents. When $V_G = 0$, the values of the branch admittances Y_i of the REI network are the same as those of the shunt admittances which replace node injections. This can be seen by comparing Eqs. (8) and (14a). It is well known that the equivalent shunt admittance is a good approximation for the nodal power injection because it is sensitive only to changes in node voltage magnitude, not angle, and the voltage magnitudes remain relatively constant in most power network problems. This same reasoning applies to the choice of $V_G = 0$ for the REI equivalent. If $V_G = 0$, the branch admittances of the REI network are functions of voltage magnitudes only; if $V_G \neq 0$, they are also functions of voltage angles which tend to change from case to case more than the magnitudes. Nevertheless, there may be application or problems where the choice of $V_G \neq 0$ is better.

The other quantity of importance is V_R which is a current weighted average of the N node voltages to which the REI network is connected. This voltage is physically meaningful and in using the equivalent in subsequent problems it can be monitored or adjusted in much the same way as the voltage of an actual node.

c. Application

The REI equivalent can be used in many different power system applications some of which have already been implemented but most of which are still speculative. Only a few suggestions can be given here.

The nodes to be eliminated are organized into groups according to some criteria and each group is combined into an REI node. As an extreme example, all of the unneeded active nodes in an entire network could be combined into one REI equivalent. More typically the number of REI equivalents in a power network might range between 10 and 100. The grouping criteria would normally be based on functional similarity of nodes. For example, in the simulation of electromechanical dynamics, nodes representing machines whose oscillations were known to be approximately coherent would be combined into one REI node. In a power flow problem nodes of the same type would be grouped together. As will be shown, sparsity effects must also be taken into account along with functional similarity and other problem related factors in grouping nodes. Proper grouping of nodes for REI equivalents is an important subproblem requiring more study.

REI equivalents can be used in conjunction with conventional network reduction for elimination of active nodes, some nodes being handled by the REI scheme, others by the conventional scheme. However, when active nodes are eliminated by the conventional scheme, their injections will be distributed to the R nodes as well as to the normal boundary nodes of the equivalent. If the G nodes are retained, the injections will be distributed to them instead of to the R nodes. In either case these distributed injections must be taken into account properly in using the equivalent. Also, when combined with the conventional scheme, some of the advantages of the REI approach are diminished.

It is possible to have one group of nodes connected to two or more REI networks each having a different influence or function in the problems to be solved with the equivalent. For example, the real power could be supplied through one REI network and the reactive power through the other. In a decoupled solution scheme[7] the equivalents for each decoupled part would not have to be the same. In some dynamic problems where the load behavior at each node is divided into parts, the parts could be segregated to separate REI equivalents. These are only suggestions of possibilities; their practical advantages have not been explored.

Although the R nodes are fictitious, they are physically meaningful. In subsequent problems the dependent variables of the R nodes can be monitored in much the same way as the individual eliminated nodes which they collectively represent. The independent variables at the R nodes can be changed according to conditions monitored during the problem solution or according to problem requirements prior to solution. The ability to control the independent variables at the R nodes and thereby approximate the same control at the eliminated nodes is a unique advantage of the approach.

SPARSITY ASPECTS OF REI APPROACH

When conventional matrix methods are to be used, the objective of network reduction is to eliminate as many nodes as possible. When sparse matrix methods are to be used, the objective of network reduction is to enhance the sparsity exploitation of the resulting equivalent. Practical suboptimal strategies for sparsity enhancement in network reduction have been developed and used for several years[9]. Network reduction itself is implemented with sparsity techniques and the matrix inversions indicated in Eq. (4) are used only for symbolic convenience. Since sparsity-oriented network reduction has been covered elsewhere[9], it will be reviewed only briefly here.

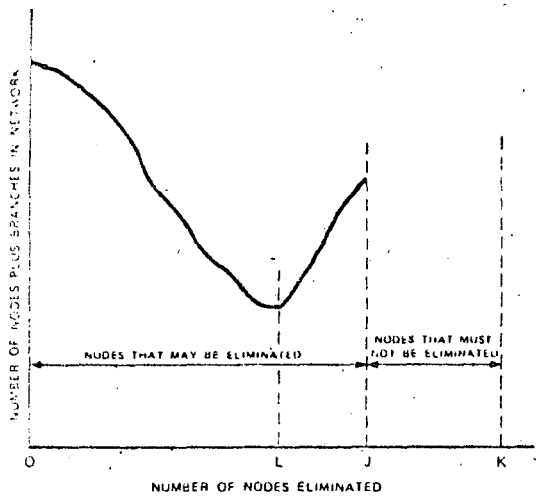


Fig. 3. Effect of network reduction on number of nodes plus branches in remaining network.

Before performing a sparsity-oriented network reduction it is necessary to divide the nodes of the network into two sets; (1) nodes that must be retained and (2) nodes that may be, but do not necessarily have to be, eliminated. If the nodes of set (2) are eliminated in a sparsity-directed order, the effect is characteristically as indicated in Fig. (3) in which the number of nodes plus branches in the network is plotted as a function of the number of nodes eliminated. As the reduction proceeds, the total number of nodes and branches, which is a measure of the computational complexity of the network, decreases until it reaches a minimum (point L), then it increases until all of set (2) is eliminated (point J). If the network reduction is stopped at the minimum, the resulting equivalent will be much better suited for sparsity methods. Nodes remaining in set (2) (between L and J) at the minimum point are retained along with the nodes of set (1) (between J and K) to enhance sparsity.

Since the REI networks become a part of the total network before reduction, the way in which nodes are grouped to establish the REI networks can have a significant influence on the sparsity of the final equivalent. After the REI networks have been added, the network reduction is no different than for any other problem. Therefore, the strategy for sparsity enhancement of the REI approach must be applied in the grouping of nodes for the REI equivalents. In grouping the nodes, functional similarity and sparsity effects must both be considered. In some cases these two factors can be considered independently, but more often they conflict and a compromise must be established.

From the standpoint of sparsity the grouping of nodes for REI equivalents should be based on topologically weakly coupled subnetworks according to the same principles that should be observed in ordinary sparsity-oriented reduction[9]. Each weakly connected subnetwork recognized by the sparsity strategy can be connected to one or more REI networks, but an REI network should not be established which would connect two otherwise weakly connected subnetworks. This is particularly important if the G nodes are to be eliminated.

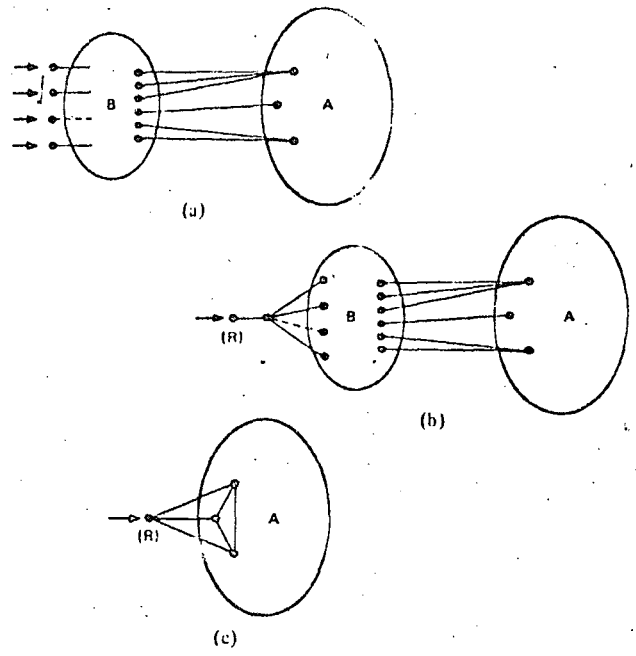


Fig. 4. (a) Main network A with subnetwork B which can be eliminated. (b) Connection of REI network to selected active nodes of B. (c) Situation after elimination of passive nodes.

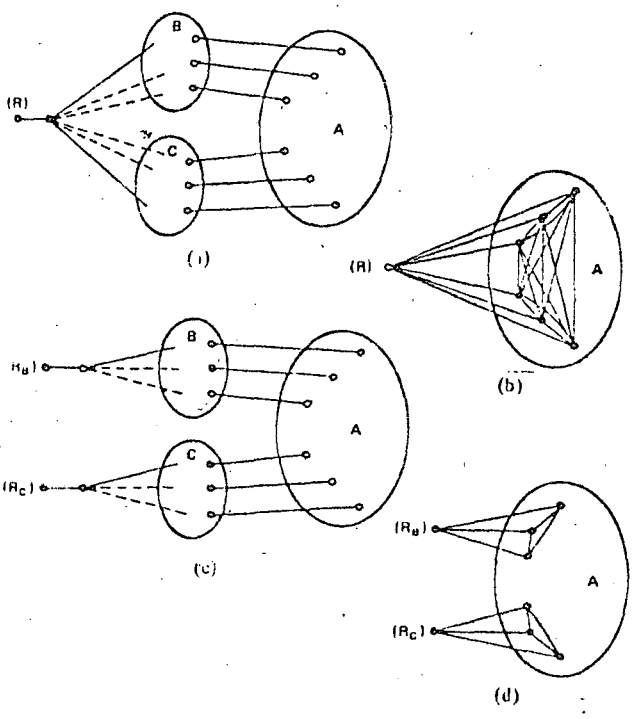
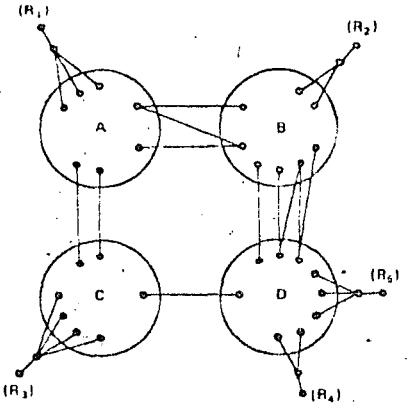


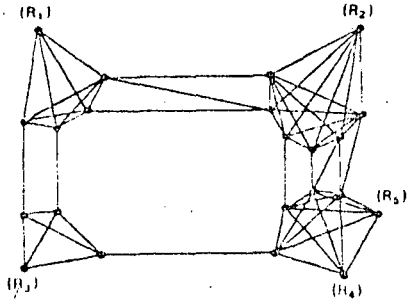
Fig. 5. (a) Main network A with radially connected subnetworks B and C connected together through an REI network. (b) Situation after elimination of all nodes of B and C. (c) Same situation as (a) but with each subnetwork connected to a separate REI network. (d) Situation of (c) after elimination of all nodes of B and C.

The idea can be shown schematically. Fig. (4a) shows a subnetwork A, which is to be retained, and a subnetwork B, all nodes of which may be eliminated. Subnetwork A is connected to B at only three boundary nodes. Fig. (4b) shows a selected group of nodes of B connected to an REI network. Fig. (4c) shows the result after eliminating all nodes of subnetwork B and retaining only node R. If the nodes of B had been eliminated by the conventional method of Eq. (3), there would have been added, at most, only three equivalent branches interconnecting the three boundary nodes. With the REI approach three more branches are needed to connect node R to the boundary nodes. Also there is the additional node R.

The way in which sparsity can be sacrificed by establishing an REI equivalent that connects two subnetworks is shown in Fig. (5). Fig. (5a) shows two subnetworks B and C each connected to three different boundary nodes of network A. If nodes from B and C are connected to one REI network, there is a potential fill-in of 21 branches as shown in Fig. (5b). If a separate REI network is established for each subnetwork, as shown in Fig. (5c), the potential fill-in, as shown in Fig. (5d), is only 12 branches. Almost the same conservation of sparsity could have been achieved for the single REI equivalent if its node G has not been eliminated.



(a)



(b)

Fig. 6. (a) Four subnetworks with REI networks connected to selected nodes of each. (b) Situation after elimination of all but the boundary nodes and R nodes of the REI networks.

As shown in Ref. [9], if all nodes of a subnetwork are eliminated, the potential fill-in is an equivalent branch connecting each boundary node to every other boundary node. For M boundary nodes the maximum fill-in is $(M^2 - M)/2$, but some of these branches may already exist so the net fill-in is usually less. If an REI network is connected to a subnetwork that has been properly identified by the sparsity logic, the boundary node fill-in is the same as for conventional network reduction, but there will be M additional branches connecting the M boundary nodes to the node R or G. Note that the number of branches to R or G in the final equivalent is M, the number of boundary nodes, not N, the number of original connecting nodes. Note also that whether the node G is eliminated or not makes very little difference unless the REI network involves two or more subnetworks. For each node G that is not eliminated there will be one additional branch RG and one additional node R. Since these branches are radial, they will cause no fill-in in the sparse triangular factorization of the equivalent. There are situations in which retaining node G enhances sparsity and others in which it is better to eliminate it.

The aim of the strategy for selecting REI nodes should be to avoid joining of subnetworks as much as possible. A simplified example is shown in Fig. (6). Fig. (6a) shows four subnetworks with five REI networks. Fig. (6b) shows the situation after elimination of all active and passive interior nodes (not shown), connecting nodes and G nodes; only the R nodes and boundary nodes of the subnetworks are retained. Conventional network reduction has been combined with the REI approach. The G nodes were eliminated because there was no advantage from the standpoint of sparsity in retaining them.

Logic for automatic grouping of nodes for REI equivalents according to functional similarity is straightforward. But the requirement to make grouping compatible with sparsity is more difficult because the problem of identifying weakly connected subnetworks has not been solved. A sparsity-directed ordering of the nodes prior to establishing the REI networks can be helpful in identifying subnetworks. In the BPA applications the input data is always coded by control areas which approximately conform to the criterion of weak topological connectivity. In establishing the REI groupings, the logic restricts each functional grouping to only one control area. For most power system problems the normal data coding provides sufficient clues to the network topology for the purposes of sparsity exploitation.

DISCUSSION AND CONCLUSIONS

The merits of the REI approach have been clearly established in large-scale practical applications in Europe and the United States. It appears to offer advantages as a replacement for conventional network reduction schemes now widely used in most power system applications. Furthermore, because of its unique properties it could open the way for new applications of equivalents.

Critical examination of the REI approach reveals it is not as simple as it first seems to be. Present knowledge concerning its performance is largely empirical and based on relatively few of its

many possible applications. Its present applications in power flow and transient stability have not been perfected and significant improvements are still expected.

The urgent need for better power network equivalents and the favorable results obtained with the REI equivalent indicate that more research and development work should be devoted to it. The REI equivalent should be studied both theoretically and experimentally and compared with alternative approaches to power network equivalents. The paper suggests some of the topics that should be considered in such investigations. Study of the REI equivalent could be undertaken as an independent research problem or treated as a subproblem of a particular application.

REFERENCES

1. J. B. Ward, "Equivalent circuits for power flow studies," AIEE Trans, Power Apparatus and Systems, vol. PAS-68, pp. 819-823, 1949.

2. W. L. Snyder, Jr., "Load flow equivalent circuits–an overview," C 72 172-1, paper given at IEEE Winter Power Meeting, New York, Jan. 30, - Feb. 4, 1972.

3. P. Dimo, "REI Models," Computer methods of power system analysis and control, Joint Research Seminar, National Science Foundation USA, National Council for Science and Technology of Romania, Ministry of Electric Energy of Romania, pp. 1-17, Bucharest, June 3-6, 1974.

4. P. Dimo, L. Groza, S. Ionescu, B. Ungereanu, I. Petcu, "Research concerning the generalized utilization of a single equivalent to the power system states analysis," Ibid, pp. 18-36.

5. P. Dimo, "Nodal analysis of power systems," Abacus Press, Tunbridge Wells, Kent, England, 1975.

6. W. F. Tinney, W. L. Powell, "The REI approach to power network equivalents," paper submitted for IEEE PICA Conference, Toronto, May 24-27, 1977.

7. B. Scott, "Review of load-flow calculations," Proceedings of IEEE, pp. 916-929, July, 1974.

8. F. D. Galiana, "Power-voltage limitations Imposed by the Network structure of a power system," IEEE PICA Conference Proceedings, New Orleans, pp. 356-363, May, 1975.

9. W. F. Tinney, W. L. Powell, W. M. Peterson, "Sparsity-oriented network reduction," IEEE PICA Conference Proceedings, Minneapolis pp. 384-390, July, 1973.

10. W. F. Tinney, J. W. Walker, "Direct solutions of large sparse networks optimally-ordered triangular factorization," Proceedings of IEEE, Vol. 55, pp. 1801-1809, Nov., 1967.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA VII: USO DE EQUIVALENTES EXTERNOS PARA ANALISIS DE CONTINGENCIAS

DR. ALBERTO MAYER SASSON

ENERO, 1979.

REAL-TIME EXTERNAL SYSTEM EQUIVALENT FOR ON-LINE CONTINGENCY ANALYSIS

J. F. DOPAZO, FELLOW, IEEE G. IRISARRI, MEMBER, IEEE A. M. SASSON, SENIOR MEMBER, IEEE

AMERICAN ELECTRIC POWER SERVICE CORPORATION
2 BROADWAY
NEW YORK, N.Y. 10004

Abstract - An equivalent model for external systems to be used in on-line contingency studies is presented. The equivalent uses Dimo's REI network reduction technique together with a method for tracking on-line injection and network changes in the external system by modifying the parameters of the equivalent. This important characteristic results in a dependable equivalent with excellent response at all operating points.

Another important aspect of this paper is the fact that for the first time the testing of on-line equivalents is carried out in a real-time environment rather than under simulated conditions. The response of the equivalent model proposed herein to the outage of a major 765KV circuit in the AEP System is reported and the results provided by the equivalent are compared to the results provided by the AEP State Estimator. The results obtained with the model presented in this paper are superior to the results obtained with other equivalent models.

INTRODUCTION

A great deal of attention has recently been given to the problem of obtaining equivalents of external systems to be used in the on-line static security analysis of interconnected power systems. The need for these equivalents results from the fact that at a particular power system computer monitoring facility or control center, very little - if any - real-time data from the external system is available. Although the recent tendency in the electric utility industry has been to favor a more active interchange of data between neighboring companies, it does not seem feasible that complete data bases - such as those required by, for instance, a state estimation based monitoring system - will be available to a particular control center in the near future. System security problems, nevertheless, are being felt with more and more intensity imposing a tremendous burden on both the system and the operator. Consequently, it seems that the only viable alternative available today to the control center designer is to forgo the possibility of basing the design on the use of real-time information from the external systems and concentrate on developing equivalent models of the external system that respond to both internal and external network and injection changing conditions with sufficient accuracy for decision making.

Two approaches have been suggested for the determination of external equivalents. One consists of the so-called topological equivalents of the Ward type [1,2] or the REI equivalent type [3-6], whereby the equivalents are obtained, off-line, using standard network reduction techniques and later coupled to the internal system to be used for on-line calculations. The second approach consists of the on-line determination of an equivalent of the Ward type by using the real-time measurements as inputs to a parameter estimation algorithm [7,8]. The former method requires the knowledge of at least a base case system condition for the external network. In that approach, since the equivalents are calculated for a particular base case operating condition, they fail to reflect the continuous changes in both injections and network configuration that take place at all times in the power system. The on-line equivalent model, on the other hand, has the advantage of attempting to reflect current operating conditions that are manifest in the available measurements; however, this type of equivalent, being of the Ward type, fails to provide an adequate measure of reactive support to the network, aside from the inherent computational difficulties of the parameter estimation algorithms so far proposed [7,8].

In the light of the above discussion, an improved external equivalent modeling technique that overcomes many of the limitations of previously reported methods is presented in this paper. The new method essentially benefits from the advantages of both the topological and on-line equivalencing approaches. It combines an off-line topological equivalent of the Dimo REI-type with on-line techniques for updating the parameters of the equivalent. In this way it is possible to reflect both external system injection and network changes as well as any modeling errors that may be present in the base case load flow conditions considered in the calculation of the off-line equivalent.

The procedure for calculating the equivalent consists of the following steps:

1. Obtain a base case load flow condition of the interconnected system for peak conditions for the present day.
2. Define various external areas according to a certain criterion such as ownership boundaries, etc.
3. Calculate a Dimo equivalent for each area with two equivalent nodes in each, one for area generation and one for area load.
4. Determine unbalances between real-time boundary conditions and those given by the equivalents.
5. Adjust Dimo node voltages to minimize these unbalances.
6. Adjust equivalent transmission network parameters to further minimize boundary unbalances.
7. Test for acceptance of equivalents by examining the level of unbalances.

Steps 4 through 7 are repeated every time a state estimation solution is obtained.

Results obtained with the equivalent under a real-time testing environment are presented in the paper. It was observed that the equivalent predicted with remarkable accuracy the effects of outaging a major 765 KV line in the AEP System. Also, by adjusting the parameters of the equivalent as indicated in steps 5 and 6 above the equivalencing technique is able to track changing load and generation conditions as well as any network changes or modeling inaccuracies that do not warrant recalculation of the equivalent. Another important characteristic of the equivalencing technique is the ability to detect any major injection or network changes in the external system. If a particular equivalent fails the acceptance test, the corresponding company or companies comprising the equivalized area may be contacted for information pertaining to possible configuration or severe injection changes. The equivalent is then modified to reflect the change.

In the following sections of the paper the equivalencing technique will be described in detail and the results of the real-time tests performed with the equivalent will be presented.

EQUIVALENT MODEL

The equivalencing technique described in the introduction consists of, essentially, a two part method. In the first part a topological equivalent of the external system of the Dimo REI-type is obtained, off-line, from a base case system condition. The second part consists mainly of on-line tuning procedures to correct modeling errors and reflect current operating conditions not included in the base case data.

Given a power system as illustrated in Fig. 1., the following data is available from the base case load flow solution: a) the complex voltages in magnitude and angle, and the amount of generation and/or load at every node and; b) a topological description of the network. With this information it is possible to determine, off-line, a Dimo REI-type equivalent of the external system. The external system can be divided in as many areas as desired to reflect criteria such as external company boundaries, interchange areas or geographical groupings, etc., or if desired, the external system nodes can be grouped according to sparsity considerations [4]. In any case, a Dimo REI equivalent is obtained for each area which consists of a generation node and a load node plus the interconnecting lines to each of the boundary nodes and to the other Dimo nodes. Fig. 2 shows the

Price: Members $2.50 Nonmembers $3.00
At Meeting $2.00

F 78 674-4. A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE PES Summer Meeting, Los Angeles, CA, July 16-21, 1978. Manuscript submitted February 1, 1978; made available for printing April 25, 1978.

equivalent that would be obtained for the power system shown in Fig. 1. After the network reduction, the following information is available: a) the complex voltages in magnitude and angle at every node including the Dimo generation and load nodes; b) the amount of injection at every node, where the Dimo injections are equal to the sum of the external area wide generation and load for the base case load flow condition, and; c) the topological description of both the equivalent and the retained internal system. The calculations necessary to obtain the equivalent can be performed off-line since they only require the information given by the base case load flow. The procedure described so far corresponds to steps 1 to 3 of the seven step logic to obtain the equivalent given earlier. Notice that the boundary mismatches are, of course, equal to zero for the base case conditions.

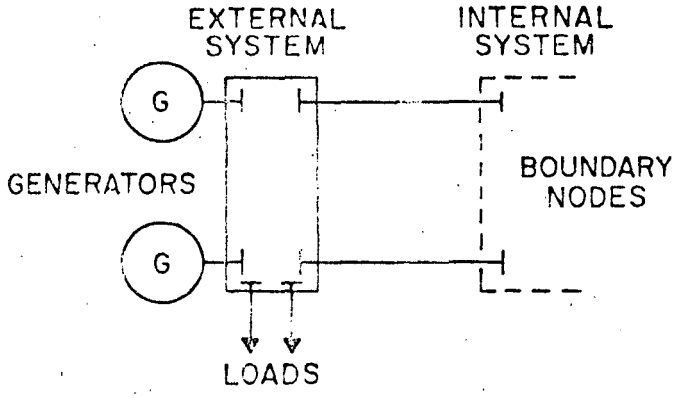


Fig. 1 Interconnected Power System

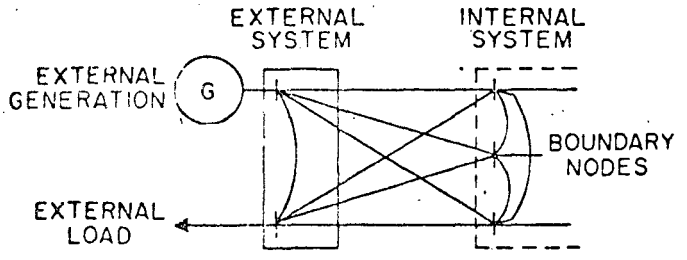


Fig. 2 Dimo REI Equivalent of External System

The second part of the equivalencing procedure consists of the coupling of the topological REI equivalent to the real-time transmission network monitoring system. This system consists of the data acquisition and state estimation monitoring facility. The data provided by the state estimator consists of the latest complex voltages and node injections at every internal node (including, of course, the boundary nodes), and the corresponding network configuration for the internal system. Since system conditions at a particular time (both internal and external) may not agree with the base case conditions for which the equivalent was calculated, mismatches will appear at every boundary bus as shown in Fig. 3. These mismatches are a manifestation of the fact that the power interchanges between the internal system and the equivalized external system are not in agreement because the injection conditions in the external systems are the assumed peak load conditions. It is clear, then, that in order to reduce the boundary mismatches as close to zero as possible, the node voltages and, as a consequence, the node injections for the REI equivalent have to be different than those obtained for the base case conditions.

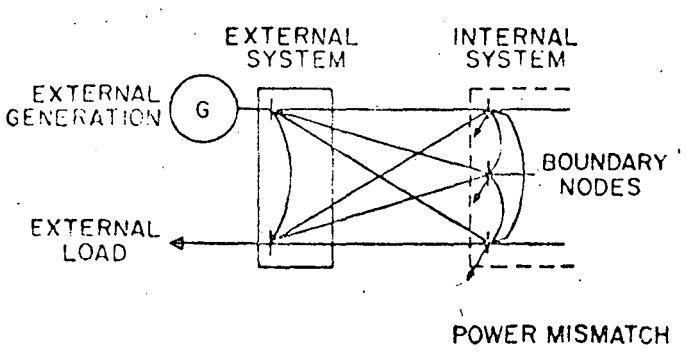


Fig. 3 Boundary mismatches at boundary nodes of equivalent

Redundant Load Flow

To calculate the new REI voltages and injections, a state estimation like calculation will be performed where the "measured" quantities are the boundary complex voltages and boundary injections and the "state" would be given by the REI generation and load nodes at each of the Dimo equivalents. This problem is thus formulated as a "redundant" load flow where there are twice as many equations as there are boundary nodes and there are only four unknowns per Dimo equivalent (the complex voltages at the generation and load nodes of the equivalent). In the experimental work representing actual conditions for the AEP EHV system late 1977, there were 45 boundary nodes and only 5 Dimo area-equivalents. Therefore, there were 90 equations and only 20 unknowns.

The redundant load flow problem can be solved in exactly the same way as the Newton load flow, the only difference being that at each iteration instead of solving a linear system of equations where there are as many equations as there are unknowns, an overdetermined system has to be solved. The solution, in this case, can be obtained as the solution to a linear least-squares problem. The redundant load flow constitutes successive least squares solutions of the overdetermined jacobian-matrix equation

$$\begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} = \underbrace{\begin{bmatrix} H & N \\ L & M \end{bmatrix}}_{J} \begin{bmatrix} \Delta\delta \\ \Delta|V| / |V| \end{bmatrix} \qquad (1)$$

where the vector $[\Delta P \quad \Delta Q]^T$ is the vector of boundary mismatches, the vector $[\Delta\delta \quad \Delta|V|/|V|]^T$ is the vector of angle and voltage magnitude corrections for the REI node voltages and the J is the jacobian matrix. The solution to the system of equations (1) is given by the system of normal equations for the linear least-squares problem [9]:

$$\begin{bmatrix} \Delta\delta \\ \Delta|V| / |V| \end{bmatrix} = (J^T J)^{-1} J^T \begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} \qquad (2)$$

The redundant load flow iterative least-squares solution process is repeated until the maximum voltage magnitude and angle corrections are smaller than a specified tolerance. Notice that this does not imply that at the solution point the boundary mismatches are smaller than a maximum prespecified nodal mismatch as is usually required in the load flow problem.

After the new REI nodal voltages have been determined, the revised REI injections, i.e., new generation and load values at each area equivalent, can be calculated. These injection values correspond to the aggregated external system injections at the new operating point. With this procedure, it is thus possible to "track", on-line, the external system injection changes without the need for receiving real-time measurements from the external systems.

Equivalent Transmission Network Adjustment

After the calculation of the current operating point REI node voltages has been completed, the new boundary mismatches can be recalculated. These boundary mismatches are partly due to transmission modeling inaccuracies. These inaccuracies might result from modeling errors in the base case load flow network data. For instance, lines may have been included in the base case data although these lines were out of service in the system. Also, during a typical day there are a variety of line switchings that could not possibly be included in the base case data and that most certainly will affect the parameters of the equivalent. Another obvious reason is that both the base case data and the load flow network modeling are not exactly equal to the real values in the system. All these errors manifest themselves as boundary mismatches.

Since the boundary mismatches are real-time quantities, they convey certain amount of information concerning the network modeling errors. It thus seems

appropriate to use this information to correct the parameters of the equivalent network. This can be accomplished very efficiently by using system identification algorithms such as those described in references [8,10]. Notice that this does not imply that all configuration errors can be corrected with this method. Rather, the parameters of the equivalent are modified so as to reduce the boundary mismatches that result from configuration errors.

At each of the boundary nodes the mismatch can be expressed as:

$$\Delta I_i = \sum_j \Delta Y_{ij} V_j \tag{3}$$

where ΔI_i is the complex current mismatch, ΔY_{ij} is the amount of correction to the equivalent bus admittance matrix and V_j is the voltage at boundary bus j as given by the state estimator.

Equation (3) can be written in matrix form as:

$$\Delta I = \Delta Y V \tag{4}$$

or, with $\Delta I = (\Delta I_R + j\,\Delta I_M)$, $\Delta Y = (\Delta G + j\Delta B)$ and $V = (V_R + j\,V_M)$, equation (4) can be rewritten as:

$$\Delta I_R(k) = \Delta G(k) V_R(k) - \Delta B(k) V_M(k) \tag{5.a}$$
$$\Delta I_M(k) = \Delta B(k) V_R(k) + \Delta G(k) V_M(k) \tag{5.b}$$

where the time dependance of the different quantities in equation (4) has been made evident by the inclusion of the counting variable k.

The problem reduces to that of calculating the parameter corrections $\Delta G(k)$ and $\Delta B(k)$ given the real-time "measurements" $\Delta I_R(k)$, $\Delta I_M(k)$, $V_R(k)$ and $V_M(k)$. The initial condition on the parameter matrices is, of course, $\Delta G(0)=0$ and $\Delta B(0)=0$. The method of solution consists of determining the real correction $\Delta G(k)$ given $\Delta B(k)$ using equation (5.a) and then, using the calculated value for $\Delta G(k)$ in equation (5.b), determine the imaginary correction $\Delta B(k)$. In general, each of these subproblems can be expressed as

$$z(k) = A(k)\,u(k) \tag{6}$$

The algorithm for "estimating" the matrix A(k) given the measurement vectors z(k) and u(k) is given by [10]:

$$A(k) = A(k-1) + [z(k) - A(k-1)u(k)]\,u^T(k)S(k) \tag{7}$$

$$S(k) = S(k-1) - \frac{S(k-1)u(k)u^T(k)S(k-1)}{1 + u^T(k)S(k-1)u(k)} \tag{8}$$

with initial conditions A(0)=0 and S(0) any suitable positive definite weighting matrix reflecting confidence in the estimate A(0). The transmission network adjustment algorithm is schematically represented in Fig. 4.

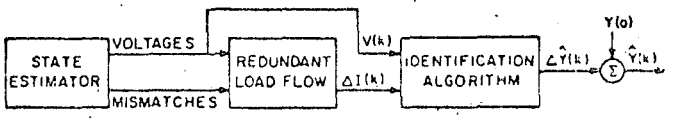


Fig. 4 Transmission network adjustment scheme

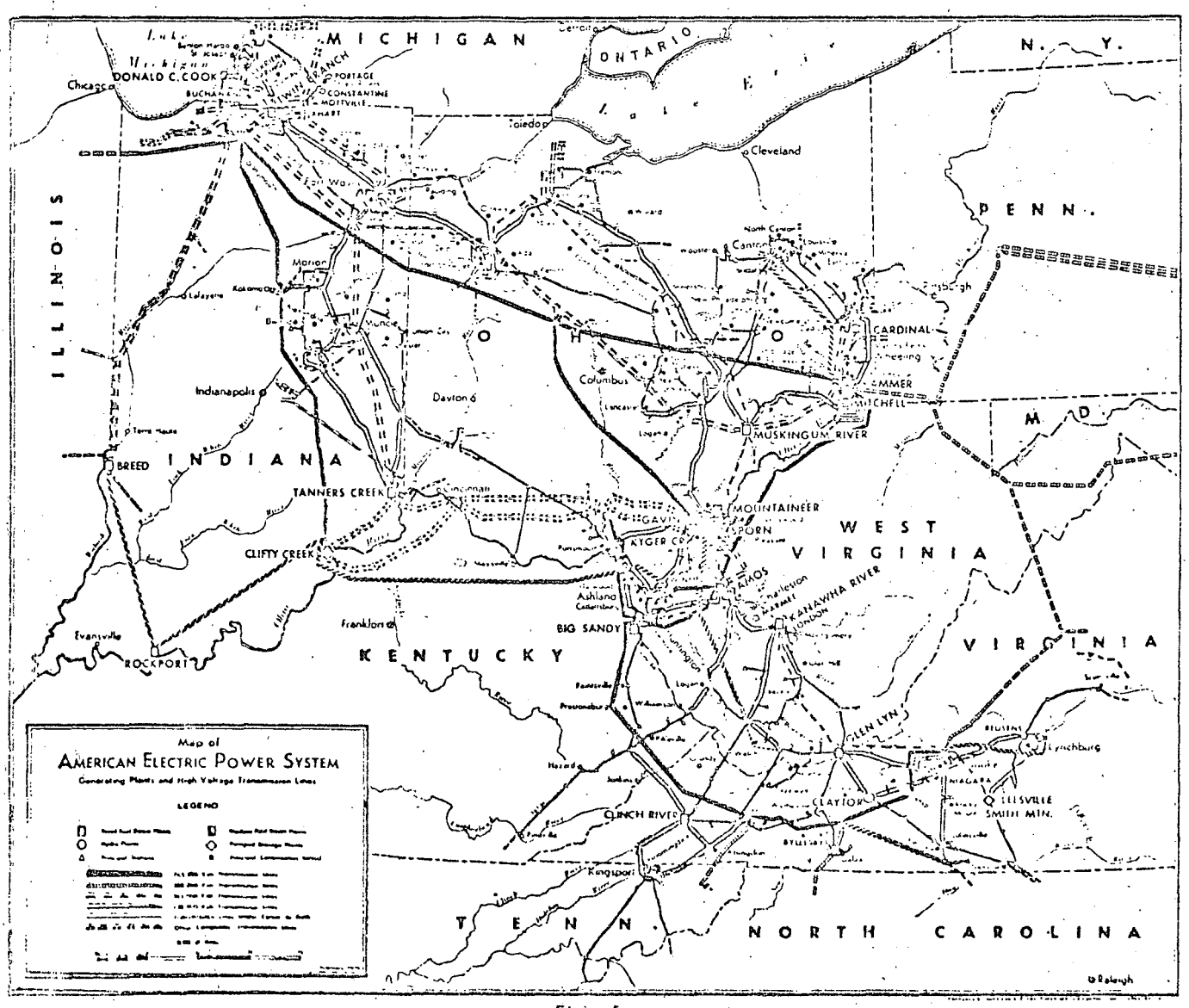


Figure 5

The procedure just described for "tracking" the external system operating point conditions and adjusting the equivalent network parameters on-line to reflect modeling and other inaccuracies, is applied every time a new state estimate is calculated at the transmission system monitoring facility. The method is very efficient and, more importantly, highly reliable.

An important aspect of the equivalencing technique presented in this paper is the ability to detect any major injection or network changes in the external systems. The on-line tuning procedures described above would, essentially, give a "smoothed" version of the original boundary mismatches. Consequently, if there is an important change in the external systems - that is, a change which is statistically "different" from the history of observed values -, it will be possible, by "observing" the boundary mismatches, to assess that an important change in the external system did occur and, in particular, in what area-equivalent the change took place. The testing performed so far with this important aspect of the equivalencing technique has produced satisfactory results.

REAL-TIME TEST OF EQUIVALENT

For the purpose of testing the validity of the steady-state real-time equivalencing technique described in the preceeding paragraphs, the planned outage of the AEP Dumont-Marysville 765KV line was used as a contingency. (See Figure 5). State estimation cases before and after the outage of the line were saved for the purpose of comparing the results of the on-line contingency simulation to the actual results as given by the state estimator.

As has been indicated before, the equivalencing technique consists of an off-line network reduction technique followed by the on-line tuning algorithms. A separate presentation for each part will be given next.

Off-line Equivalent Calculation

The topological equivalent of the Dimo REI-type was calculated off-line from the peak loading conditions base case load flow prepared by System Operations personnel for the day of the outage. The system and component data for the case is given below:

System Data	MW	MVAR
Generation	184,273	34,016
Load	181,030	45,330
Losses	3,239	33,339
Mismatch	1	4
Cap/Reactors		3,945
Line charging		40,711

Components	
Nodes	3,561
Lines	6,003
Transformers	757
Cap/Reactors	379
Regulated nodes	477
Systems	42
Tie lines	469
Meter jumpers	0

The external systems (including the non-monitored AEP 138 Kv system) were divided into five (5) external areas according to a "geographical proximity" criterion as shown in Fig. 6. A Dimo REI equivalent was calculated for each of these areas. The total area generation and area load were aggregated into one Dimo generation and one Dimo load node respectively. The totals per area are given in Table I. The passive external network (including the additional REI nodes and lines) was reduced, using standard network reduction techniques, from a total of 3581 nodes and 8833 lines to 72 nodes (45 boundary nodes, 10 Dimo nodes and 17 internal nodes) and 1647 lines (the equivalent admittance matrix is a full matrix). The equivalent matrix and the additional node data given in Table I comprise the off-line Dimo equivalent of the external systems.

On-line Tuning Procedure

The on-line calculations described earlier are repeated every time a new state estimate of the internal system is obtained. For the purpose of showing the effect of both the redundant load flow and the network parameter correction algorithm, 19 state estimation cases were processed prior to the outage of the

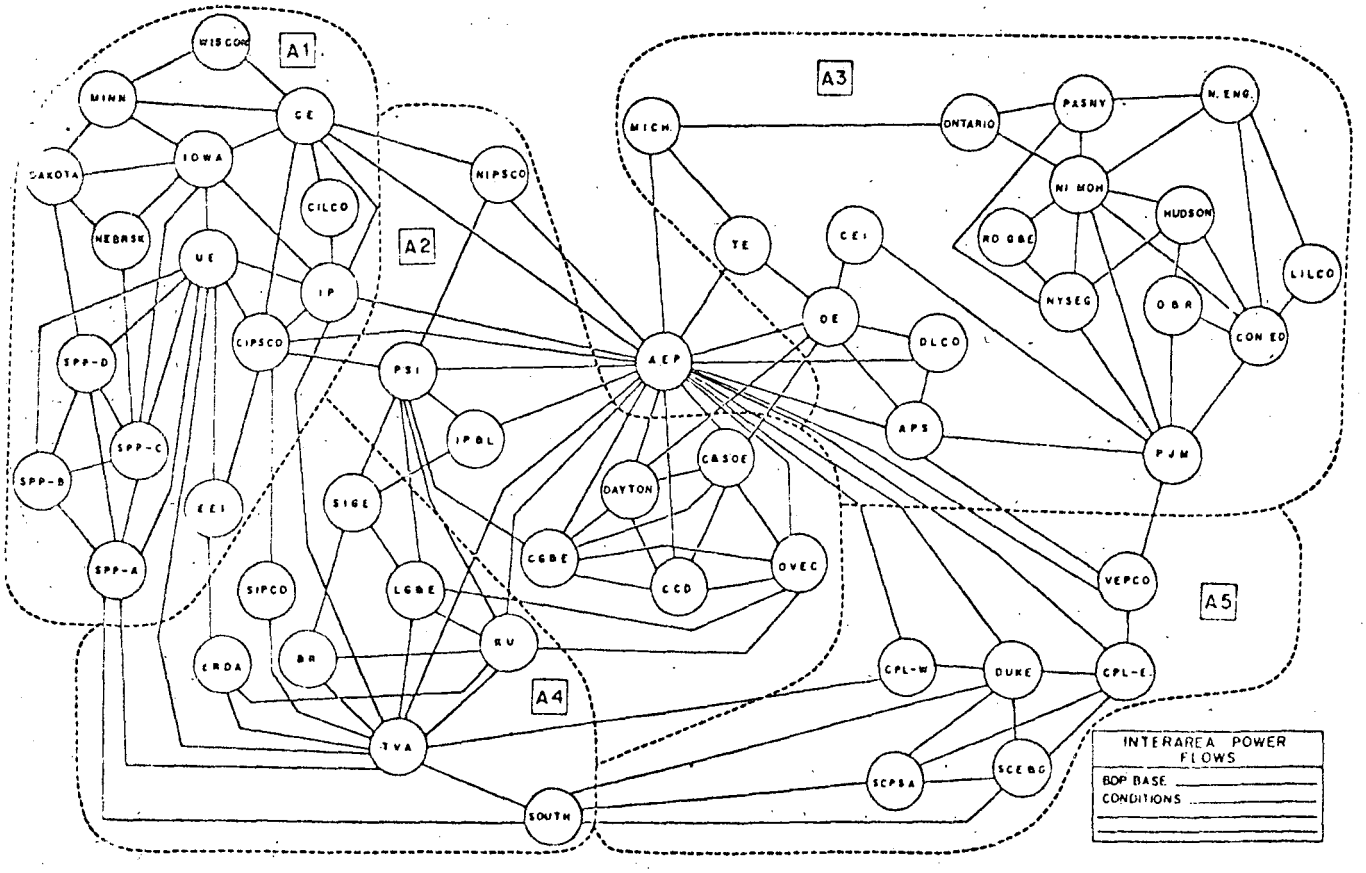


Fig. 6 External system areas for equivalencing

Area	REI Node	Generation		Load		Voltage	
		MW	MVAR	MW	MVAR	Mag.	Angle(rad)
1	Gen.	27,479	3,939	-	-	1.060	-0.127
	Load	-	-	29,567	6,308	1.040	-0.338
2	Gen.	13,521	1,788	-	-	1.045	-0.033
	Load	-	-	17,625	4,100	0.995	-0.239
3	Gen.	32,423	6,042	-	-	1.071	-0.152
	Load	-	-	31,793	6,835	1.007	-0.403
4	Gen.	37,412	6,482	-	-	1.067	0.235
	Load	-	-	33,905	10,596	1.010	0.020
5	Gen.	19,309	3,845	-	-	1.036	-0.038
	Load	-	-	19,158	5,524	0.987	-0.264

Table I. REI Node Voltages and Injections

Dumont-Marysville 765 KV line on October 18, 1977. The first case calculated by the state estimator corresponds to 5.542 am and the last case immediately before the contingency was calculated at 7.057 am. The outage of the line took place immediately after the last case. Figures 7 (a-e) present plots of the total area active and reactive mismatches. The total mismatch is equal to the algebraic sum of the boundary mismatches for all boundary nodes in a particular area. In each of the figures, the total mismatch is shown for a case in which the 19 state estimation cases were processed without the network parameter adjustment algorithm and with the algorithm.

To show the effect of configuration errors in the equivalent, the Dimo off-line calculation was repeated without including the 765 KV tie line Dumont (AEP) - Wilton Ctr (CE). The same state estimation cases were processed with this error on the area equivalent. Figures 8 (a-e) give the total area mismatches for this case. Notice that only the total mismatch for area equivalent 1 shows a larger than "expected" value. This immediately suggests that a large configuration change has occurred in area one. Therefore, analyzing the individual boundary changes, it is thus possible to narrow down the location of the change to perhaps a few neighboring companies that can subsequently be contacted for information pertaining the possible configuration changes.

Contingency Simulation

The most important aspect of any equivalencing method, of course, is the ability to accurately predict the effects of a contingency. To test the method presented herein, the outage of a major 765 KV line in the AEP system was used as a contingency. A load flow case simulating the outage of the line was run with the system conditions given by the AEP state estimator just prior to the outage of the line. Since state estimation results after the outage were saved, it was possible to compare the results of the load flow with the system conditions corresponding to the actual values observed in the power system. Tables II and III summarize these results. Only the values in a few nodes and lines in the immediate neighborhood of the outage are given. The results in nodes and lines further apart from the outage are, of course, even better.

For comparison purposes, similar results obtained with the Extended Ward Equivalent (EW) of reference [2] and the Ward equivalent of reference [1] are also given in Tables II and III.

Discussion of results

Several important characteristics of the equivalencing technique are made evident by the results presented in this section.

1) The plots shown in Fig. 7 (a-b) demonstrate the importance of both the redundant load flow and the network parameter adjustment algorithm. The redundant load flow will calculate the REI equivalent nodal voltages that minimize the boundary mismatches. By letting these voltages be adjustable quantities, the injection level of the equivalent will change in agreement with the system load curve. This, in effect, is the "tracking" effect of the redundant load flow. In addition, the parameter identification algorithm modifies the admittances of the equivalent to further reduce the boundary mismatches due to various modeling errors, as evidenced by the plots. It is interesting to notice that the parameter identification algorithm will cease adjusting the transmission parameters of the equivalent after a few cases are processed unless a new change takes place in the external system.

2) The plots of Fig. 8 (a-e) show the ability of the equivalencing methodology to detect external system changes. It is clearly shown in the plots that only the mismatch in the area where the transmission change occurred responded to the change. In this way it would be possible to identify the change and proceed to correct the equivalent. If the effect of the change is minor, the network parameter adjustment technique corrects the equivalent without the need of recalculating the Dimo off-line equivalent.

3) The results presented in Tables I-III strongly indicate the satisfactory performance of the equivalent in predicting the effects of a major transmission contingency. It is also shown in the tables that the equivalencing technique proposed in this paper responds more accurately than the Ward or the Extended Ward equivalents.

CONCLUSIONS

In this paper, the first - to the authors' knowledge - successful real-time test of an external network equivalent technique for on-line contingency analysis has been presented. The results obtained with the new method to the outage of a major 765 KV line in the AEP system indicate that the equivalent has excellent response characteristics.

The essence of the new technique consists of the combination of a reliable off-line network reduction method - Dimo's REI method - with on-line tuning procedures that adjust the REI equivalent to the changing conditions in the external systems without requiring the transfer of real-time information between different systems. Only real-time information from the internal system is used in the equivalent adjustment algorithms. Although the results presented herein are quite encouraging, further testing of the equivalencing methodology is presently being conducted at AEP. For instance, the following tests are being planned for the immediate future:

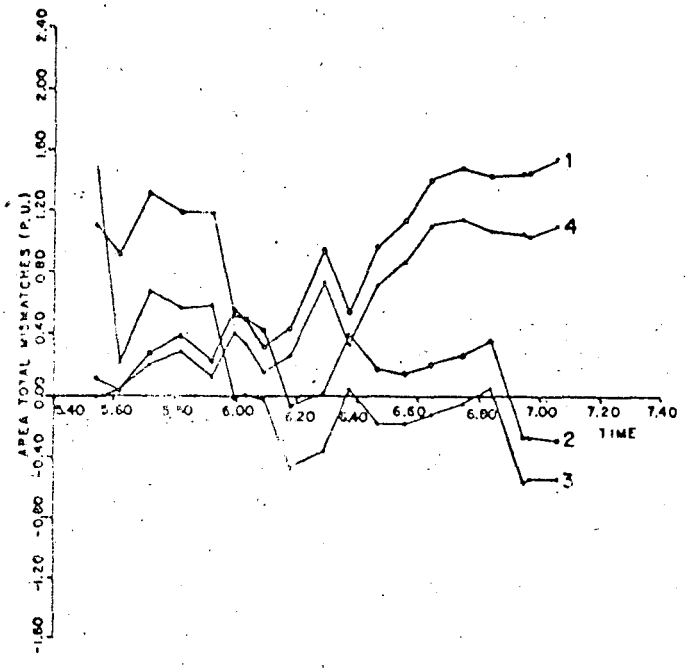
1) testing of several additional transmission contingencies.

Node Name	SE - before		SE - after		AEP Method		Ext. Ward		Ward	
	Mag.	Angle	Mag.	Angle	Mag.	Angle	Mag.	Angle	Mag.	Angle
COOK3	1.026	-.055	1.028	-.154	1.026	-.153	1.026	-.151	1.026	-.153
COOK7	0.997	-.032	1.004	-.143	0.998	-.143	0.998	-.139	0.998	-.141
DUMONT7	0.994	-.021	1.002	-.138	0.995	-.138	0.995	-.134	0.994	-.136
DUMONT3	1.021	-.059	1.022	-.162	1.021	-.162	1.021	-.159	1.020	-.160
OLIVE3	1.026	-.057	1.027	-.156	1.026	-.156	1.026	-.153	1.024	-.154
OLIVEC	1.022	-.064	1.023	-.163	1.023	-.163	1.022	-.160	1.021	-.161
TWINBRA3	1.025	-.076	1.026	-.172	1.026	-.171	1.025	-.169	1.024	-.171
SYLMA3	1.000	-.091	0.994	-.104	0.996	-.101	0.998	-.103	0.992	-.104
MARYSV3	1.028	-.075	1.017	-.095	1.019	-.098	1.023	-.097	1.018	-.097
MARYSV7	0.978	-.107	0.967	-.139	0.969	-.144	0.972	-.143	0.969	-.143
KAMMER7	0.980	-.214	0.978	-.222	0.980	-.228	0.980	-.226	0.980	-.227
BELMONT7	0.990	-.219	0.987	-.226	0.990	-.232	0.990	-.230	0.989	-.231
GAVIN7	0.984	-.228	0.981	-.237	0.984	-.240	0.984	-.238	0.984	-.238
AMOS7	0.976	-.190	0.[illegible]72	-.192	0.976	-.193	0.976	-.192	0.976	-.192

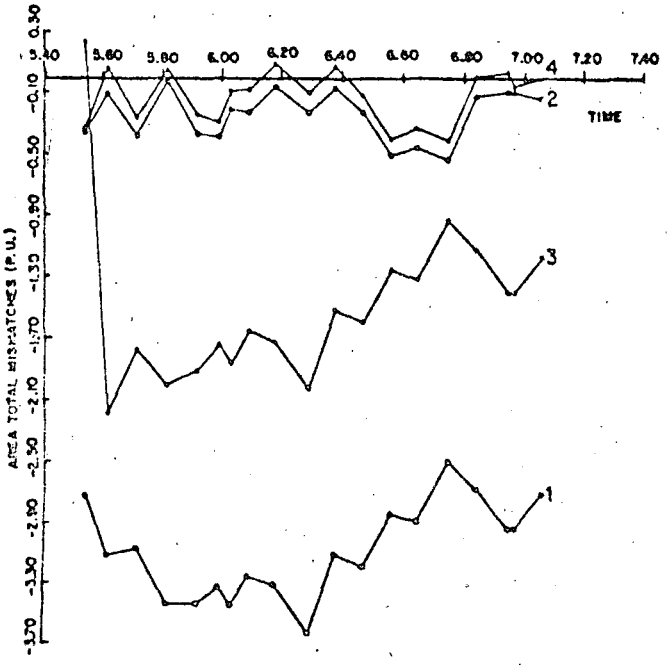
Table II. Complex node voltages in internal system

Line Flows		SE - before		SE - after		AEP Method		Ext. Ward		Ward	
From	To	MW	MVAR	MW	MVAR	MW	MVAR	MW	MVAR	MW	MVAR
COOK3	COOK7	- 320.6	43.8	- 143.9	- 15.7	- 142.2	28.7	- 153.0	31.4	- 154.2	38.1
	OLIVE3	16.9	- 15.1	26.0	- 3.8	22.6	- 12.8	13.7	- 11.3	13.5	- 2.6
COOK7	COOK3	320.9	- 36.0	144.0	17.3	142.3	- 27.1	153.2	- 29.5	154.4	- 36.1
	DUMONT7	- 326.2	36.3	- 144.0	- 16.7	- 147.5	26.2	- 158.3	28.6	- 159.5	35.3
DUMONT7	MARYSVI7	739.6	161.1	-	-	-	-	-	-	-	-
	COOK7	326.4	-197.0	144.0	-149.4	147.5	-189.9	158.4	-192.7	- 159.6	-198.6
DUMONT3	OLIVE3	- 40.6	- 71.2	- 81.2	- 65.8	- 84.0	- 56.2	- 89.0	- 63.1	- 89.2	- 66.5
	OLIVEC	63.0	- 32.1	14.5	- 23.4	12.1	- 25.2	14.9	- 27.9	14.9	- 25.0
	TWINBRA3	366.6	-140.0	221.8	-128.2	224.0	-140.2	239.8	-130.4	240.2	-128.2
OLIVE3	COOK3	- 16.9	- 5.7	- 26.0	- 17.1	- 22.6	- 8.1	- 13.7	- 9.6	- 13.5	- 18.3
	DUMONT3	40.7	58.6	81.3	53.5	84.1	43.9	89.1	51.0	89.3	54.4
OLIVEC	DUMONT3	- 63.0	19.5	- 14.5	- 10.4	- 12.1	12.3	- 14.9	15.0	- 14.9	12.2
MARYSVI7	DUMONT7	744.0	-193.8	-	-	-	-	-	-	-	-
	GAVIN7	-1016.5	55.5	- 811.4	- 41.7	- 805.5	- 45.5	- 804.8	- 29.4	- 801.8	- 53.4
	MARYSVI3	950.7	54.5	1300.9	31.4	1331.1	43.9	1330.3	14.0	1325.4	58.7
	KAMMER7	- 733.2	106.8	- 558.0	22.9	- 579.0	25.3	- 579.1	38.8	- 577.1	18.8
GAVIN7	MARYSVI7	1021.0	182.5	814.0	236.2	808.4	239.1	807.6	222.4	804.6	246.5
	AMOS7	993.9	99.0	1161.6	109.4	1218.8	94.5	1216.0	94.6	1219.1	94.5
	BELMONT7	144.9	-242.2	185.8	-246.7	131.5	-245.4	135.0	-240.7	135.0	-238.8
AMOS7	GAVIN7	- 991.8	-240.7	-1158.7	-235.5	-1215.7	-217.3	-1212.9	-217.6	-1215.9	-217.2
MARYSVI3	MARYSVI7	- 949.7	- 23.4	-1299.7	27.5	-1329.8	17.4	-1328.0	46.9	-1324.2	2.3
	SWLIMA3	495.6	65.4	582.7	60.6	584.6	64.8	590.4	66.6	587.8	74.1
SWLIMA3	MARYSVI3	- 489.5	- 21.5	- 574.3	16.9	- 576.1	13.0	- 581.8	12.6	- 579.1	5.9
TWINBRA3	DUMONT3	- 366.0	111.8	- 221.6	100.4	- 223.8	112.6	- 239.6	103.0	- 240.0	101.0
BELMONT7	GAVIN7	- 144.8	- 30.2	- 185.7	- 23.2	- 131.4	- 27.2	- 134.9	- 31.8	- 135.0	- 33.7
	KAMMER7	127.5	98.8	106.3	98.9	91.0	103.4	83.8	97.7	83.5	95.3
KAMMER7	BELMONT7	- 127.3	- 26.0	- 106.2	- 26.7	- 90.9	- 30.8	- 83.7	- 25.2	- 83.4	- 22.9
	MARYSVI7	737.0	- 31.3	560.2	19.5	581.5	20.4	581.5	6.8	579.5	26.6

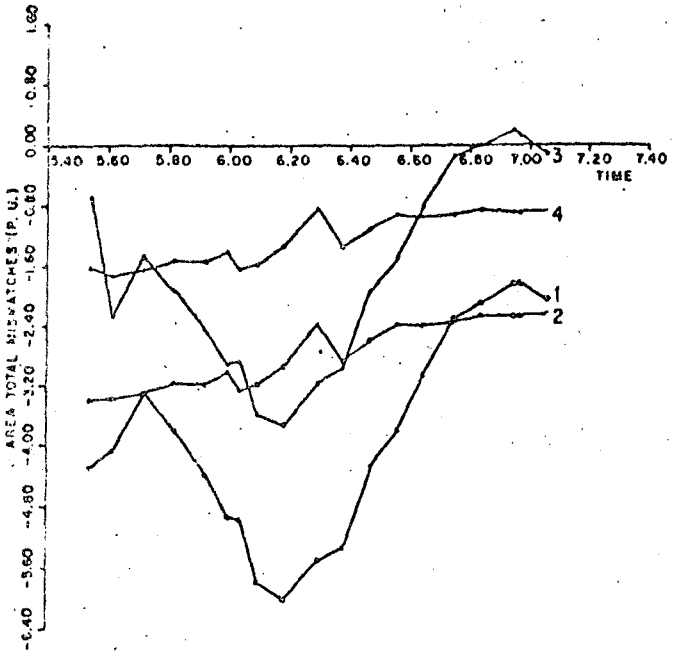
Table III. Line flows in the vicinity of the Dumont-Marysville line



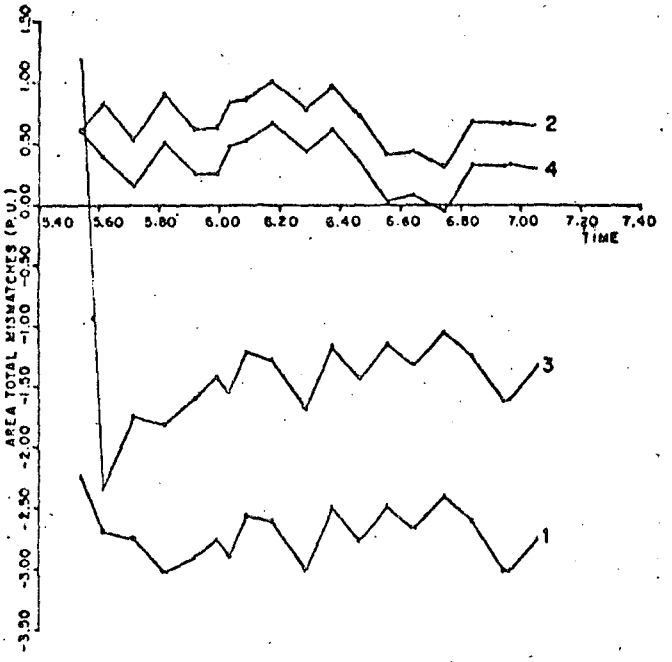
7a) Area 1



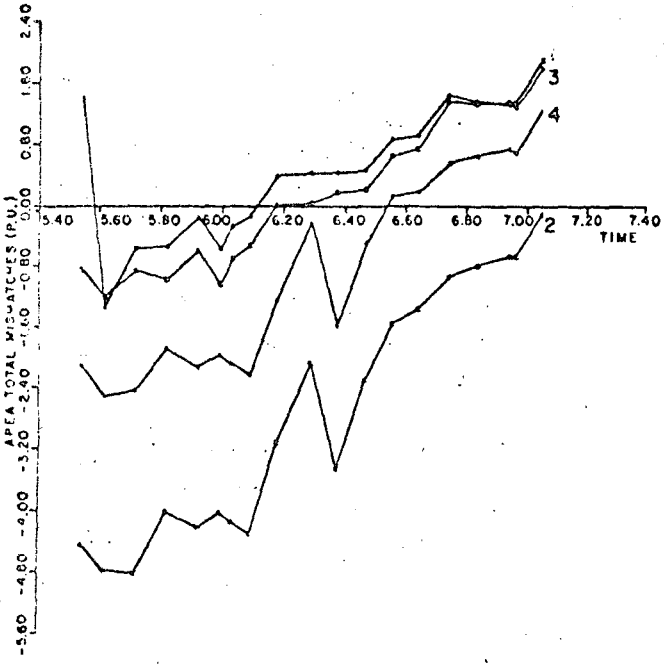
(7b) Area 2



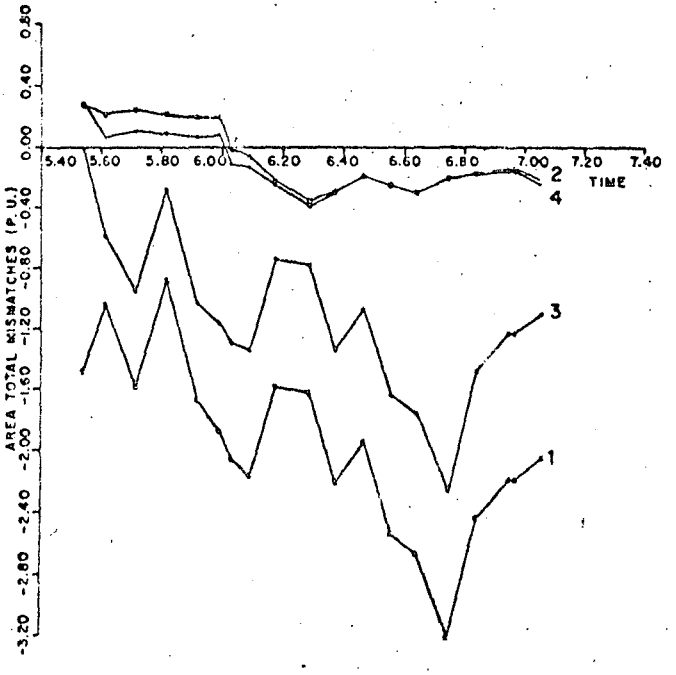
(8a) Area 1



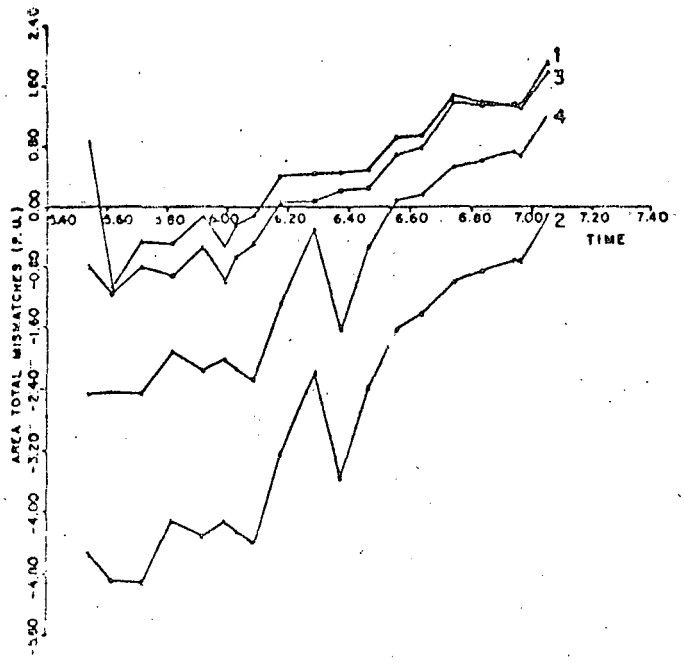
(8b) Area 2



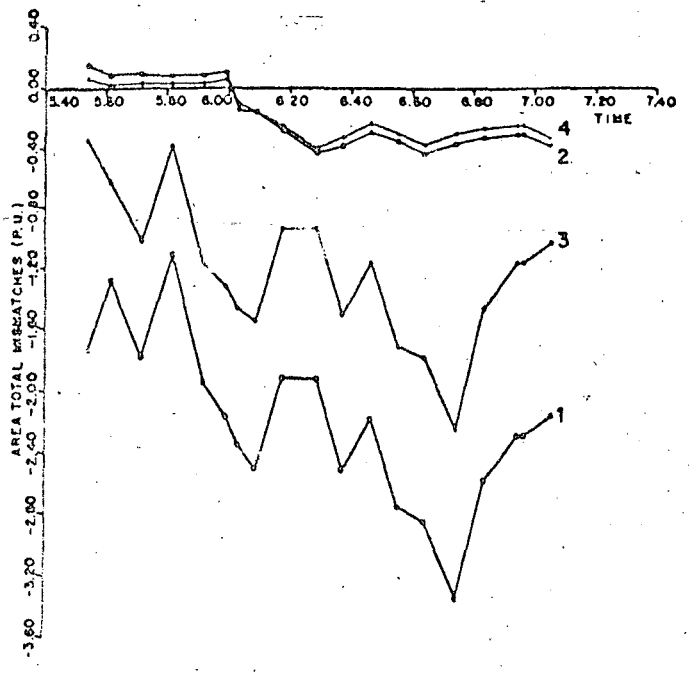
(8c) Area 3



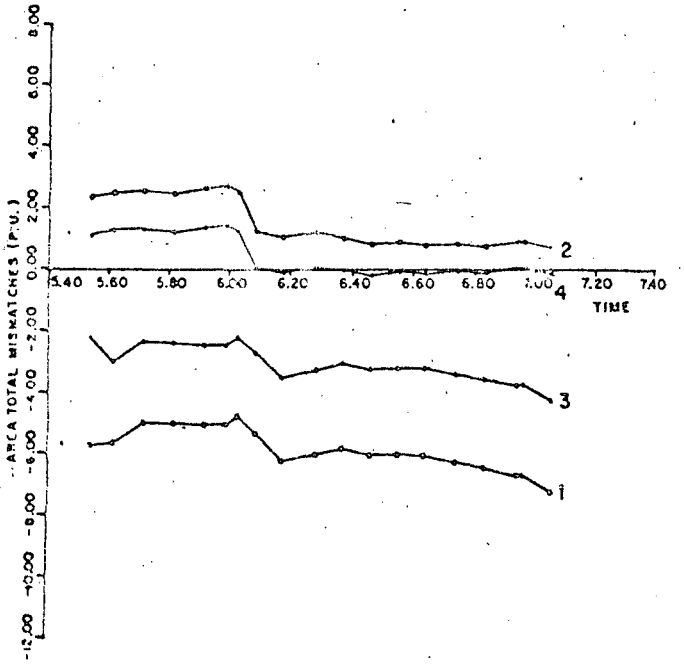
(8d) Area 4



(7c) Area 3

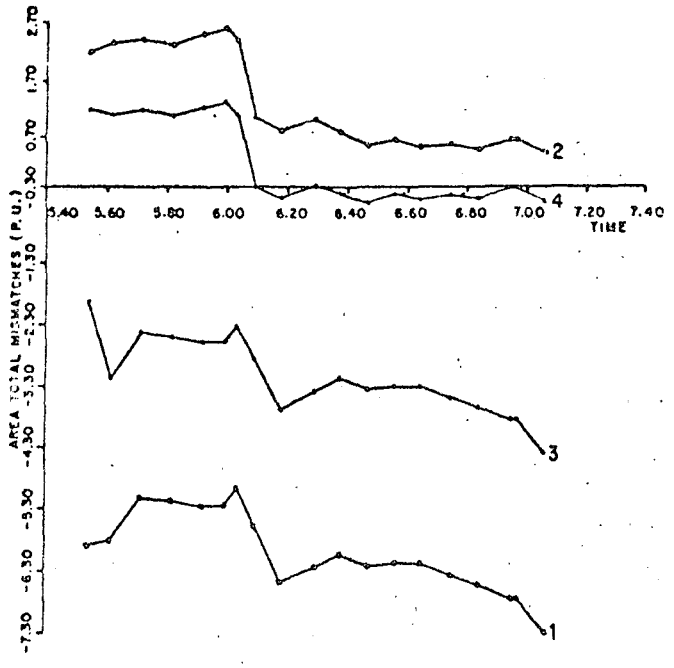


(7d) Area 4



(7a) Area 5

Fig. 7 Total area mismatches. 1) Real mismatch no identification; 2) Reactive mismatch no identification; 3) Real mismatch with identification; 4) Reactive mismatch with identification.



(8a) Area 5

Fig. 8 Total area mismatch (error in off-line equivalent). 1) Real mismatch no identification; 2) Reactive mismatch no identification; 3) Real mismatch with identification; 4) Reactive mismatch with identification.

2) studing the ability of the equivalent to predict the effect of generation and/or load contingencies.

3) determination of the response of the equivalencing model to inter-system energy transfers.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the many contributions made by Dr. M. H. Dwarakanath, a former member of our group, and the valuable help provided by Mrs. Julia J. Li in preparing some of the real-time data. Special appreciation is conveyed to L. S. VanSlyck and to T. L. Overly and his group in Canton, Ohio, for their help in coordinating the preparation of the data for the cases studied.

REFERENCES

(1) Ward J. B., "Equivalent Circuits for Power-Flow Studies", AIEE Trans., Vol. 68, pp. 373-382, 1949.

(2) Monticelli A., Deckman S., Garcia A., and Stott B., "Real-Time External Equivalents for Static Security Analysis", F78 087-9 presented at the 1978 IEEE PES Winter Meeting, New York, Feb. 1978.

(3) Dimo P., **Nodal Analysis of Power Systems**, Abacus Press, England, 1975.

(4) Tinney W. F. and Powell W. L., "The REI Approach to Power Network Equivalents", PICA Conference, Toronto, May 1977.

(5) Dimo P., Groza L., Ionescu S., Ungureanu B., and Petcu I., "Research Concerning the Generalized Utilization of a Single Equivalent to the Power System State Analysis", Computer Methods of Power System Analysis and Control, Joint Research Seminar between Romania and USA, Bucarest, Romania, June 1974.

(6) DyLiacco T. E., Savulescu S. C., and Ramarao, K. A., "An On-line Topological Equivalent of a Power System", F 77 523-4, presented at the 1977 IEEE PES Summer Meeting, Mexico City, July 1977.

(7) Deville T. G., and Schweppe F. C., "On-line Identification of Interconnected Network Equivalents from Operating Data", C72 464-6 presented at the 1972 IEEE PES Summer Meeting, San Francisco, July 1972.

(8) Dopazo, J. F., Dwarakanath, M. H., Li, J. J. and Sasson, A. M., "An External System Equivalent Model Using Real-Time Measurements for System Security Evaluation", IEEE TPAS, Vol. PAS-96, March/April 1977.

(9) Stewart G. W., **Introduction to Matrix Computations**, Academic Press, New York, 1973.

(10) Kashyap, R. L. and Rao, A. R., **Dynamic Stochastic Models from Empirical Data**, Academic Press, New York, 1976.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE NALISIS DE REDES ELECTRICAS

TEMA VIII SELECCION AUTOMATICA DE CONTINGENCIAS

DR. ALBERTO MAYER SASSON

ENERO, 1979.

F 79 212-2

AUTOMATIC CONTINGENCY SELECTION FOR ON-LINE SECURITY ANALYSIS - REAL-TIME TESTS

G. IRISARRI, MEMBER, IEEE

D. LEVNER, MEMBER, IEEE

A. M. SASSON, SENIOR MEMBER, IEEE

AMERICAN ELECTRIC POWER SERVICE CORPORATION
2 BROADWAY
NEW YORK, N.Y. 10004

Abstract - Results of various real-time tests performed on the AEP-EHV System with the automatic contingency selection algorithm proposed in [1] are reported.

The first-order performance index sensitivities used in the algorithm for line contingency ranking are rederived in a straight forward manner and, moreover, the ranking algorithm is extended to accommodate second-order performance index sensitivities. Many tests, however, indicate that - for the AEP-EHV Network operating conditions - the automatic contingency ranking algorithms do not perform reliably.

As an alternative, it is shown that a DC-Load Flow, although computationally more demanding than the first-order automatic contingency selection algorithm (but less computationally complex that the second-order extension), is a reliable and efficient way of ranking line contingencies.

INTRODUCTION

The on-line steady-state contingency analysis packages presently being used in energy control centers simulate transmission line contingencies by either

- selecting the outages to be studied from a pre-prepared list of meaningful contingencies, or,
- by operator selection depending on current system conditions as given by a state estimator or any other power system monitoring facility.

The former approach is an off-line planning practice applied to the on-line situation. This method fails to recognize that system conditions assumed for the preparation of the contingency lists may be far from real time conditions. In fact, what is understood to be a single contingency in a planning study may very well be a double or even a triple contingency in a real time situation. The second method - i.e. operator selection - may prove to be burdensome even to experienced operators. As operating conditions in most systems get more stringent, the number of contingency choices gets larger and no amount of experience guarantees proper selection of troublesome contingent situations.

As a consequence, it is necessary to develop methods for automatic selection of meaningful contingency cases. A first approach in this direction is the method proposed by Ejebe-Wollenberg [1] (see also reference [2]). In their paper the authors propose a methodology for ranking transmission line contingencies by evaluating the normalized sensitivities of a system wide performance index with respect to line outages. The authors show that through the use of Tellegen's theorem the first-order sensitivities of the performance index with respect to changes in line admittance can be readily determined and efficiently implemented. The normalized sensitivities (or first-order changes in the performance index), according to Ejebe-Wollenberg, can then be used to rank contingencies from the most important (largest positive performance index change) to the least important (largest negative performance index change).

Tests performed on the AEP-EHV system reveal, however, that the automatic contingency ranking method of Ejebe-Wollenberg is unreliable. In fact, for the case of real power flow performance index the method may produce extensive misclassification of transmission line contingencies resulting not only in false alarms (i.e. unimportant outages classified high in the list) but, more critically yet, meaningful contingencies may be classified as unimportant (low in the list). The main reason for this unpredictable behaviour of the automatic contingency selection method is that the performance index suggested in [1] is not a monotonic function of the susceptance (DC-Load Flow formulation) of the lines. This monotonicity condition seems to be at least a necessary condition for the trustworthy performance of the method.

To improve the performance of the method, the calculation of the normalized sensitivities of Ejebe-Wollenberg was extended to include second-order effects. Even in this case, however, the automatic contingency selection algorithm was not found to be sufficiently reliable.

However, the use of a DC-Load Flow calculation was determined to be highly reliable and less computationally demanding than the second-order extension considered in this paper. It is concluded, then, that use of a DC-Load Flow model may be a desirable alternative to the automatic method of Ejebe-Wollenberg for contingency selection.

In the following sections of the paper the first-order and second-order sensitivities of the real power flow performance index will be developed. The approach presented herein for the development of these formulas is different from that suggested by Ejebe-Wollenberg. Rather than using Tellegen's theorem, it is shown that the same results can be obtained in a straight forward manner by the use of elementary differential calculus. Also, results of the tests performed with the first-order, second-order and DC-Load Flow methods are compared to the "correct" ranking as given by full AC-Load Flow outage simulations.

CONTINGENCY SELECTION ALGORITHM

According to Ejebe-Wollenberg [1] (see also [4-5]), an intuitively appealing index for quantifying the extent of line overloads may be defined in terms of a real power performance index

$$J_{MW} = \sum_{\ell=1}^{NL} \frac{W_\ell}{2n} \left(\frac{|P_\ell|}{P_\ell^{max}} \right)^{2n} \qquad (1)$$

where

- P_ℓ megawatt flow in line ℓ
- P_ℓ^{max} megawatt capacity of line ℓ
- W_ℓ real non-negative weighting coefficient (assumed equal to unity in the sequel)
- NL number of lines in the system
- n specified exponent (n = 1 used in the sequel)

The DC-flow in line ℓ (between nodes m and n) is

$$P_\ell = - B_\ell \theta_\ell \qquad (2)$$

where B_ℓ is the susceptance of line ℓ and θ_ℓ is the angle across the line ($\theta_\ell = \delta m - \delta n$).

Price: Members $2.50 Nonmembers $3.00
At Meeting $2.00

79 212-2 A paper recommended and approved by the IEEE Power System Engineering Committee of the IEEE Power Engineering Society for presentation at the IEEE PES Winter Meeting, New York, NY, February 4-9, 1979. Manuscript submitted September 1, 1978; made available for printing November 6, 1978.

Moreover, if it is assumed that the megawatt capacity of line ℓ (P_ℓ^{max}) is a linear function of the susceptance of the line, that is,

$$P_\ell^{max} = K_\ell \, |B_\ell| \tag{3}$$

then, substituting eqs. (2) and (3) in (1), results in the performance index

$$J_{MW} = \sum_{\ell=1}^{NL} \frac{1}{7} \left(\frac{|\theta_\ell|}{K_\ell} \right)^2 \tag{4}$$

The assumption of eq. (3), although not mentioned explicitely, was also made by Ejebe-Wollenberg in their paper [1]. In this paper results for both performance indeces (eqs. (1) and (4)) will be given.

The performance indeces of eqs. (1) and (4) are normalized functions of the line flows (eq. (1)) and angles across the lines (eq. (4)) respectively. In both cases, the performance indeces have a small value when all the line flows are less than the capability of the respective lines, and a high value whenever there are line overloads.

The purpose of the automatic contingency selection method is to efficiently identify the contingencies (line outages) that warrant further study by means of a full AC-Load Flow solution. This is achieved by calculating the changes in the MW-flow performance index to changes in susceptance of the lines in the network. These changes to the performance index can be obtained from the sensitivities of J_{MW} with respect to the susceptance of the lines in the network. The first-order and second-order sensitivity expressions are given next (see appendix I for details).

First-Order Sensitivities

The first-order sensitivity of the MW-performance index of eq. (4) is (similar formulas hold for the performance index of eq. (1)),

$$\frac{\partial J_{MW}}{\partial B_\ell} = - \hat{\theta}_\ell \, \theta_\ell \, , \ell = 1, 2, \ldots, NL \tag{5}$$

where

θ_ℓ angle across line ℓ (between nodes m and n) for the base case condition

$\hat{\theta}_\ell$ angles across line ℓ in the "adjoint" network (Note: the same notation of [1] is used herein although the concept of adjoint network as used in that paper is not used in the present paper.)

The adjoint angle across line ℓ can be expressed as

$$\hat{\theta}_\ell = \underline{M}_\ell^T \, \hat{\underline{\delta}} \tag{6}$$

where

$\underline{M}_\ell$ column vector (NBx1) which is null accept for elements m and n which are equal to 1 and -1 respectively (NB = number of nodes)

$\hat{\underline{\delta}}$ vector of adjoint nodal angles. (NBx1)

In eq. (6) the vector of adjoint nodal angles can be obtained from

$$\underline{B} \, \hat{\underline{\delta}} = \hat{\underline{P}} \tag{7}$$

where

$\underline{B}$ susceptance matrix (NBxNB)

$\hat{\underline{P}}$ vector of adjoint injections (NBx1)

The adjoint injections vector can be calculated from

$$\hat{\underline{P}} = \sum_{j=1}^{NL} \left\{ \frac{\partial J_{MW}}{\partial |\theta_j|} \, \text{sign}(\theta_j) \, \underline{M}_j \right\} \tag{8}$$

where $\partial J_{MW} / \partial |\theta_j|$ can be calculated directly from eq. (4).

Notice that the angles θ_ℓ of eqs. (5) - (8) correspond to angles differences across the lines for the base case condition. This base case, for on-line applications, corresponds to the current angles at the nodes of the network given by, for example, a state estimation monitoring facility. The first-order sensitivity method only requires one forward and backward solution (once B is triangulated) as evidenced by eq. (7).

Second-Order Sensitivities

The second-order sensitivity of the MW-flow performance index of eq. (4) can be shown to be (see appendix I)

$$\frac{\partial^2 J_{MW}}{\partial B_\ell^2} = 2 \, (\underline{M}_\ell^T \, \underline{B}^{-1} \, \underline{M}_\ell) \, \hat{\theta}_\ell \, \theta_\ell + \underline{M}_\ell^T \, \underline{B}^{-1} \, \hat{\underline{B}} \, \underline{B}^{-1} \, \underline{M}_\ell \, \theta_\ell^2 \tag{9}$$

where

$\hat{\underline{B}}$ "adjoint" network susceptance matrix

The adjoint susceptance matrix of eq. (9) is given by

$$\hat{\underline{B}} = \sum_{j=1}^{NL} \left\{ \frac{\partial^2 J_{MW}}{\partial |\theta_j|^2} \, \underline{M}_j \, \underline{M}_j^T \right\} \tag{10}$$

For both cases - i.e. first-order and second-order sensitivities - the change in the performance index ΔJ_ℓ is calculated using the Taylor series expansion of J_{MW} as a function of the susceptance B_ℓ of line ℓ ($\ell = 1, 2, \ldots, NL$). That is,

$$\Delta J_\ell = \frac{\partial J_{MW}}{\partial B_\ell} (- B_\ell) + \frac{1}{2} \frac{\partial^2 J_{MW}}{\partial B_\ell^2} (- B_\ell)^2 \tag{11}$$

where the derivatives $\partial J_{MW} / \partial B_\ell$ and $\partial^2 J_{MW} / \partial B_\ell^2$ are given by eqs. (5) and (9) respectively. Notice that the "change" in susceptance is precisely equal to the susceptance of the line so as to simulate a line outage. For the first-order effects only the first term in the right hand side of eq. (11) is needed. Notice that the second-order method requires NL forward-backward solutions (see eq. (9)).

DC-Load Flow

The MW-flow in the lines of the network after a change in susceptance ΔB_ℓ for line ℓ are given by (see appendix I)

$$P_k^\ell = \begin{cases} P_k + b_\ell \, (\underline{M}_k^T \, \underline{B}^{-1} \, \underline{M}_\ell) \, B_k \, \theta_\ell \, , \; k \neq \ell \\ (P_\ell - \Delta B_\ell \, \theta_\ell) \, [1 - b_\ell \, (\underline{M}_\ell^T \, \underline{B}^{-1} \, \underline{M}_\ell)] \, , \; k = \ell \end{cases} \tag{12}$$

where P_k^ℓ is the new flow in line k for a change in susceptance in line ℓ and

$$b_\ell = (\Delta B_\ell^{-1} + \underline{M}_\ell^T \, \underline{B}^{-1} \, \underline{M}_\ell)^{-1} \tag{13}$$

Therefore, the effect of the outage can be easily simulated by calculating the approximate line flows from eq. (12) and substituting in the performance index of eq. (1). Notice that the computational complexity of this approach is of the same order as that of the second-order sensitivity method as NL forward-backward solutions are required.

TEST CASES AND RESULTS

The automatic contingency selection algorithm described above was tested under real-time conditions on the AEP-EHV system. The base case conditions corresponded to the system conditions given by the AEP State Estimator at 6 AM on October 18, 1977. The reaction of the external system to changes in the AEP-EHV internal system was considered by coupling to the internal system an updated REI equivalent of the external system [3]. A Ward equivalent (constant current) was also used to determine the effect of the equivalent on the selection algorithm. The results obtained with this type of equivalent were inferior to the results obtained with the REI equivalent.

Results with both performance indeces (eqs. (1) and (4)) for the first-order sensitivity, second-order sensitivity and DC-Load Flow methods are given in the sequel. The "correct" ranking, as given by a full AC-Load Flow, is used as reference for comparison. Thus, the normalized sensitivities and approximate DC-Load Flow performance index values are ordered according to the ranking given by the AC-Load Flow.

TABLE I: CONTINGENCY RANKING FOR PERFORMANCE INDEX OF EQ. (1)

Line p	Line q	AC-Load Flow	First Order Sensitivity	Sec.-Order Sensitivity	DC-Load Flow
7	8	0.84406E+01	0.35523E+00	0.71958E+00	0.86371E+01
11	10	0.81172E+01	0.11892E+00	0.25049E+00	0.85319E+01
5	4	0.76232E+01	-0.11607E+00	-0.13657E+00	0.77251E+01
10	9	0.75269E+01	-0.87468E-01	-0.63054E-02	0.76032E+01
14	9	0.74670E+01	0.70404E-01	0.13016E+00	0.75049E+01
31	7	0.74506E+01	0.33356E-01	0.14318E+00	0.74665E+01
31	7	0.74506E+01	0.31032E-01	0.14507E+00	0.74666E+01
49	33	0.74243E+01	0.26069E+00	0.44455E+00	0.77241E+01
30	31	0.74206E+01	0.17424E+00	0.27030E+00	0.74586E+01
9	21	0.74004E+01	-0.13945E+00	-0.16912E+00	0.75072E+01
14	15	0.73720E+01	0.20059E+00	0.29771E+00	0.74445E+01
11	12	0.73393E+01	-0.55354E-01	-0.84866E-01	0.73931E+01
10	19	0.73370E+01	-0.94446E-01	0.20769E-01	0.78772E+01
43	44	0.72934E+01	0.10426E+00	0.14971E+00	0.72641E+01
31	32	0.72519E+01	0.67503E-01	0.11781E+00	0.72300E+01
21	22	0.72477E+01	0.14570E-01	0.46449E-01	0.72666E+01
20	21	0.72075E+01	-0.92726E-01	-0.13446E+00	0.72789E+01
29	30	0.72069E+01	0.49955E-01	0.86273E-01	0.72679E+01
55	25	0.72062E+01	0.93056E-01	0.12432E+00	0.72866E+01
55	25	0.72032E+01	0.92792E-01	0.12353E+00	0.72855E+01
25	5	0.72119E+01	0.74389E-01	0.10640E+00	0.72163E+01
8	59	0.72100E+01	0.14557E-01	0.29611E-01	0.72505E+01
25	1	0.71930E+01	0.89399E-01	0.11869E+00	0.72197E+01
14	13	0.71924E+01	0.59447E-01	0.83199E-01	0.71845E+01
42	49	0.71845E+01	0.62595E-01	0.91675E-01	0.72024E+01
42	49	0.71845E+01	0.62575E-01	0.91675E-01	0.72024E+01
32	33	0.71701E+01	0.62379E-01	0.87572E-01	0.71894E+01
69	59	0.71601E+01	0.55968E-02	0.11187E-01	0.72797E+01
35	37	0.71449E+01	0.16917E-01	0.26295E-01	0.71460E+01
10	20	0.71356E+01	0.30485E-02	0.83189E-02	0.71336E+01
29	28	0.71321E+01	0.19436E-01	0.32363E-01	0.71449E+01
32	34	0.71291E+01	0.21984E-01	0.34120E-01	0.71344E+01
34	35	0.71279E+01	0.17699E-01	0.24183E-01	0.71149E+01
34	35	0.71271E+01	0.17459E-01	0.23773E-01	0.71142E+01
13	25	0.71102E+01	0.11001E-01	0.17663E-01	0.71149E+01
37	38	0.71023E+01	0.42589E-02	0.69791E-02	0.70939E+01
13	58	0.71014E+01	0.70012E-02	0.10392E-01	0.70999E+01
35	36	0.71003E+01	-0.39132E-02	-0.33840E-02	0.70843E+01
35	36	0.70986E+01	-0.38655E-02	-0.33407E-02	0.70843E+01
39	38	0.70974E+01	0.22834E-02	0.33602E-02	0.70943E+01
42	43	0.70969E+01	0.42921E-02	0.77304E-02	0.71036E+01
37	40	0.70907E+01	0.10769E-02	0.19343E-02	0.70923E+01
45	46	0.70899E+01	-0.15026E-02	-0.64304E-03	0.70897E+01
49	34	0.70874E+01	0.32654E-02	0.43041E-02	0.70918E+01
48	50	0.70800E+01	-0.37575E-01	-0.41432E-01	0.70944E+01
7	6	0.70791E+01	-0.22947E+00	-0.19744E+00	0.75462E+01
6	59	0.70789E+01	-0.54482E-02	-0.86293E-02	0.70789E+01
49	48	0.70753E+01	-0.24219E-01	-0.21754E-01	0.70734E+01
42	41	0.70715E+01	-0.23519E-02	-0.51833E-02	0.70821E+01
39	38	0.70704E+01	-0.13255E-01	-0.18022E-01	0.70664E+01
15	16	0.70711E+01	-0.53230E-02	-0.91143E-02	0.70585E+01
41	34	0.70708E+01	-0.15376E-01	-0.14235E-01	0.70814E+01
3	2	0.70652E+01	-0.73650E-02	-0.10934E-01	0.70729E+01
57	13	0.70639E+01	-0.82317E-02	-0.12427E-01	0.70704E+01
32	50	0.70578E+01	-0.83433E-02	-0.12735E-01	0.70727E+01
48	43	0.70512E+01	-0.19804E-01	-0.25179E-01	0.70677E+01
20	23	0.70492E+01	-0.12012E+00	-0.21289E+00	0.70012E+01
12	13	0.70486E+01	-0.37555E-01	-0.43120E-01	0.70657E+01
54	47	0.70363E+01	-0.27090E-01	-0.39538E-01	0.70422E+01
45	44	0.69866E+01	-0.89342E-01	-0.11172E+00	0.69700E+01
46	44	0.69839E+01	-0.74783E-01	-0.11035E+00	0.69466E+01
5	6	0.69528E+01	-0.76950E-01	-0.11893E+00	0.69825E+01
1	2	0.69303E+01	-0.87714E-01	-0.11790E+00	0.69587E+01
43	47	0.67086E+01	-0.17640E+00	-0.21566E+00	0.68949E+01
24	23	0.68253E+01	-0.19120E+00	-0.23116E+00	0.68834E+01
27	26	0.68377E+01	-0.31343E+00	-0.41344E+00	0.68834E+01
17	16	0.68275E+01	-0.10655E+00	-0.13662E+00	0.68831E+01
18	17	0.68270E+01	-0.61632E-01	-0.11321E+00	0.67855E+01
27	28	0.68214E+01	-0.10666E+00	-0.16737E+00	0.68536E+01
19	18	0.68148E+01	-0.38955E-01	-0.71692E-01	0.71154E+01
44	47	0.68132E+01	-0.11185E+00	-0.18582E+00	0.68026E+01
51	48	0.68077E+01	-0.18334E+00	-0.26929E+00	0.68402E+01
52	51	0.68051E+01	-0.81906E-01	-0.14395E+00	0.68500E+01
61	4	0.68006E+01	-0.33175E+00	-0.51053E+00	0.68131E+01
26	56	0.67977E+01	-0.17343E+00	-0.26633E+00	0.67814E+01
9	8	0.67972E+01	-0.15727E+00	-0.26412E+00	0.70206E+01
48	53	0.68269E+01	-0.35833E+00	-0.54612E+00	0.67709E+01
1	4	0.69437E+01	-0.20687E+00	-0.42653E+00	0.64698E+01
25	26	0.68170E+01	-0.39742E+00	-0.12167E+01	0.62844E+01

TABLE II: CONTINGENCY RANKING FOR PERFORMANCE INDEX OF EQ.(4)

Line p	Line q	AC-Load Flow	First-Order Sensitivity	Second-Order Sensitivity
7	8	0.87901E+01	0.61241E+00	0.11185E+01
11	10	0.82872E+01	0.36905E+00	0.77301E+00
5	4	0.79739E+01	0.10160E+01	0.21343E+01
10	9	0.77740E+01	0.38803E+00	0.77703E+00
14	9	0.76912E+01	0.20216E+00	0.40054E+00
9	21	0.76599E+01	-0.71413E-02	0.89425E-01
31	7	0.76517E+01	0.42261E+00	0.68558E+00
31	7	0.76516E+01	0.42545E+00	0.69213E+00
30	31	0.75369E+01	0.17635E+00	0.27333E+00
14	15	0.76141E+01	0.20414E+00	0.39228E+00
49	33	0.75910E+01	0.23981E+00	0.41505E+00
11	12	0.75802E+01	0.76755E+00	0.17051E+01
10	19	0.75381E+01	0.79741E+00	0.15221E+01
43	44	0.74692E+01	0.95305E-01	0.13814E+00
20	21	0.74643E+01	0.47570E+00	0.10335E+01
31	32	0.74576E+01	0.74695E-01	0.12711E+00
21	22	0.74448E+01	0.23710E+00	0.42392E+00
55	25	0.74268E+01	0.96784E-01	0.12891E+00
29	30	0.74259E+01	0.48405E-01	0.83691E-01
55	25	0.74231E+01	0.95290E-01	0.12805E+00
8	59	0.74223E+01	0.53540E-01	0.11025E+00
25	5	0.74079E+01	0.78099E-01	0.10737E+00
14	13	0.73976E+01	0.61844E-01	0.85353E-01
25	1	0.73964E+01	0.87995E-01	0.11686E+00
42	49	0.73750E+01	0.59208E-01	0.86945E-01
42	49	0.73750E+01	0.59208E-01	0.86945E-01
32	33	0.73592E+01	0.57664E-01	0.81429E-01
69	59	0.73571E+01	0.19684E-01	0.45459E-01
10	20	0.73420E+01	0.37657E-01	0.73582E-01
35	37	0.73394E+01	0.14894E-01	0.25691E-01
29	28	0.73293E+01	0.18730E-01	0.31195E-01
34	35	0.73218E+01	0.16568E-01	0.22693E-01
34	35	0.73210E+01	0.16344E-01	0.22299E-01
32	34	0.73210E+01	0.19598E-01	0.30655E-01
13	25	0.73076E+01	0.11569E-01	0.18572E-01
13	58	0.72995E+01	0.78154E-02	0.11270E-01
37	38	0.72931E+01	0.38441E-02	0.63205E-02
35	36	0.72950E+01	-0.28353E-02	-0.20590E-02
35	36	0.72943E+01	-0.28903E-02	-0.20433E-02
39	38	0.72927E+01	0.20330E-02	0.35181E-02
42	43	0.72899E+01	0.34418E-02	0.63540E-02
45	46	0.72894E+01	-0.30920E-02	-0.32295E-02
37	40	0.72868E+01	0.91564E-03	0.16454E-02
49	34	0.72934E+01	0.30653E-02	0.40469E-02
6	59	0.72807E+01	0.87505E-02	0.19357E-01
48	50	0.72730E+01	0.11065E+00	0.20166E+00
49	48	0.72705E+01	0.13119E-01	0.20305E-01
42	41	0.72695E+01	0.19545E-01	0.46575E-01
7	6	0.72692E+01	0.20136E+00	0.52972E+00
39	36	0.72685E+01	-0.11687E-01	-0.15817E-01
41	34	0.72680E+01	0.21131E-02	0.89955E-02
15	16	0.72666E+01	-0.59422E-02	-0.95191E-02
3	2	0.72649E+01	-0.68313E-02	-0.10148E-01
57	13	0.72612E+01	-0.85949E-02	-0.12977E-01
32	50	0.72547E+01	-0.92294E-02	-0.13343E-01
48	43	0.72544E+01	0.12193E-01	0.24545E-01
12	13	0.72480E+01	-0.35650E-01	-0.40506E-01
54	47	0.72403E+01	0.10222E-01	0.23217E-01
45	44	0.72045E+01	-0.85077E-01	-0.10538E+00
46	44	0.71831E+01	-0.69835E-01	-0.10287E+00
20	23	0.71911E+01	0.74960E+00	0.17572E+01
5	6	0.71792E+01	-0.73462E-01	-0.11395E+00
43	47	0.71340E+01	-0.27767E-01	-0.91518E-02
1	2	0.71327E+01	-0.87234E-01	-0.11721E+00
44	47	0.70780E+01	0.15934E+00	0.36853E+00
27	26	0.70697E+01	0.35367E+00	0.73553E+00
24	23	0.70592E+01	0.40874E-01	0.85872E-01
27	28	0.70531E+01	-0.10937E+00	-0.17270E+00
52	51	0.70497E+01	-0.32178E-01	-0.35494E-01
17	16	0.70467E+01	0.92637E-01	0.25026E+00
18	17	0.70464E+01	0.46561E-01	0.13451E+00
51	48	0.70429E+01	0.14037E+00	0.31639E+00
19	18	0.70145E+01	-0.44297E-01	-0.60725E-01
61	4	0.69649E+01	0.88379E+00	0.16965E+01
26	56	0.69642E+01	0.19899E+00	0.42312E+00
48	53	0.68833E+01	0.71561E+00	0.15183E+01
9	8	0.68549E+01	0.66599E-01	0.22379E+00
1	4	0.66604E+01	-0.10613E-02	0.72421E-01
25	26	0.63935E+01	0.52910E+00	0.12275E+01

Table I gives the results obtained for the performance index of eq. (1), while Table II gives corresponding results obtained for the performance index of eq. (5). The results obtained with the DC-Load Flow formulation are shown in Table I and they are the same for both performance indeces. The base case value of the performance index for the results of Table I is 7.087 while for the results of Table II is 7.2829.

From these results it is clear that although the ranking of contingencies according to the first-order and second-order sensitivity methods is not totally incorrect (as evidenced by the "clustering" of contingencies according to importance), some of the possible line outages were misclassified. For instance, for the performance index of eq. (1), the outage of lines (5-4), (10-9), (9-21), etc., that appear to be important from the ranking given by the AC-Load Flow, would have been classified as unimportant contingencies by both the first-order and second-order sensitivity methods. Also, for the results of Table II, line (9-21) would have been classified as unimportant based on first-order sensitivities. Use of second-order sensitivities corrects this particular case (it does not, however, provide extensive correction in other cases). Cases such as these are particularly dangerous and any automatic contingency selection method should avoid such situations. On the other hand, false alarms (unimportant lines classified as important) are not as dangerous except if the misclassification is extensive. Indeed, if some unimportant cases are classified high in the list, they may overshadow important cases which otherwise would have been considered when running the full AC-Load Flows to further study the outages. This situation is also to be avoided. Examples of false alarms can be seen from Table II. For instance, lines (61-4), (25-26), (48-53) would have been classified high by both the first-order and second-order sensitivity methods although the AC-Load Flow flagged them as unimportant.

For the purpose of demonstrating the reasons behind the unreliable behaviour of the sensitivity methods, plots of J_{MW} versus B_ℓ were made for a few lines. From the plots it was discovered that the performance indeces (both eqs. (1) and (4)) are not always monotonic functions of B_ℓ. Figure 1, for instance, is a plot of J_{MW} (eq. (4)) versus B_ℓ for line (19-18). Notice that for this particular line J_{MW} is not a monotonic function of B_ℓ. Moreover, the operating point (OP) is a point of negative slope and, within the region of interest (from OP to the ordinate axis), the slope of the performance index function J_{MW} changes sign. Therefore, if the first-order sensitivity (and, in this case, even the second-order sensitivity because J_{MW} is not a second-order function of B_ℓ) is used to "predict" the value of J_{MW} after outaging the line ($B_\ell = 0$), the prediction would be far from the "true" value given by the AC-Load Flow. This behaviour of the performance index function is, of course, system and operating point dependent and may or may not occur. This inconsistent behaviour is, precisely, what makes the method of automatic selection unreliable and, to some extent, unattractive.

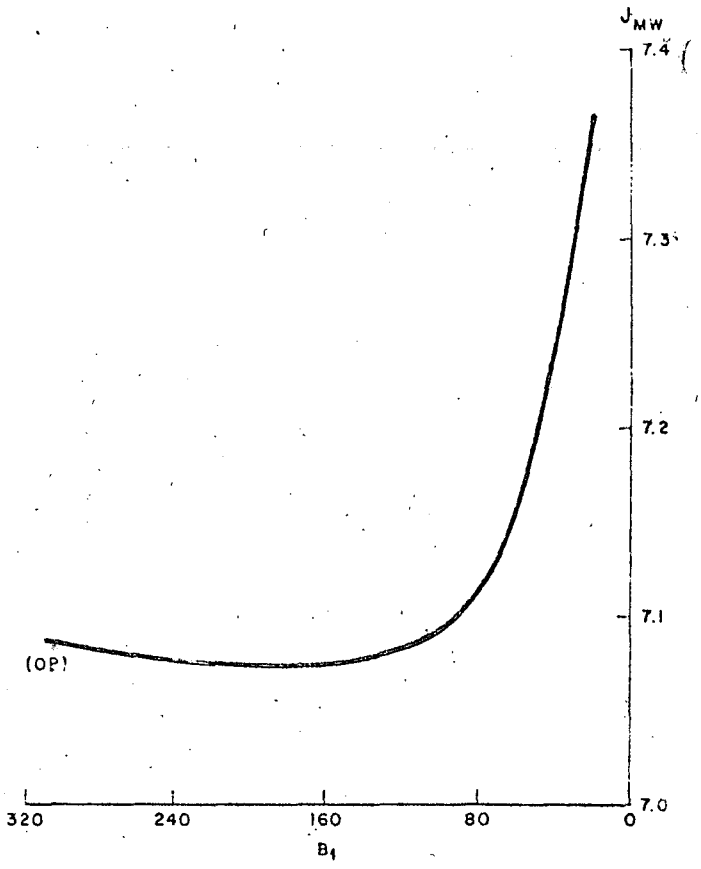


Figure 1: Plot of the Performance Index of Eq. (4) vs. Susceptance (B_ℓ) for line (19 - 18)

Another obvious reason for the unpredictable behaviour of the method is that the approximation to the performance index function as given by eq. (11) is, of course, valid only in the neighborhood of the "operating point", and may not hold true (even with second-order effects) when the susceptance change is large, as is precisely the case when simulating a line outage.

A way of overcoming these difficulties is provided by the use of the DC-Load Flow model of eq. (12). The results given in Table I for the DC-Load Flow model (which also apply for the performance index of eq. (4)) show that this approach is highly reliable and accurate. Minor misclassification resulting in false alarms (as is the case of lines (10-19), (7-6), (19-18), and (9-8)) may occur as shown in the results. It is expected, however, that these errors will be more infrequent in cases where there is more separation between the important and unimportant contingencies. The DC-Load Flow methods is, of course, more demanding, computationally, than the first-order sensitivity method. It is, however, less so than the second-order sensitivity method.

CONCLUSIONS

The results of applying the automatic contingency selection method of Ejebe-Wollenberg [1] to the real-time AEP-EHV system conditions were presented in this paper. It was found that this method, although conceptually highly appealing, is unreliable and, under certain circumstances, may give incorrect results. Two types of anomalous behavior were identified,

- classification of unimportant line outages as important (false alarms), and,
- classification of important outages as unimportant.

The former type of error, although not extremely dangerous, may increase the computational load of running full AC-Load Flow cases for further studies. It may even mask important lines that otherwise would appear higher in the ranking than the misclassified cases. The latter case - misclassification of important outages - is, of course, highly unwanted in an on-line automatic environment.

It was shown in the paper that the unreliable behaviour of the automatic contingency selection method is due to the fact that the performance index function is not a monotonic function of transmission line susceptances. This is, in fact, at least a necessary condition for the acceptable performance of the method. Even if the performance index is monotonic, a high degree of misclassification may result depending on the degree of convexity or concavity of the performance index function. This, however, is highly system and operating point dependent and cannot be ascertained a priori. All these factors make the method, at best, unreliable.

For the purpose of enhancing the performance of the automatic contingency selection, the method was extended to include second-order effects. Although the results obtained with the second-order extension were better than those obtained with the first-order method, the improvement was not significant. It was shown that, not only the first-order sensitivity but the second-order sensitivity formulas can be readily developed using elementary differential calculus providing thus an alternative, and clearer way of developing the equations, than the use of Tellegen's theorem.

It was also shown in the paper that the DC-Load Flow provides a competitive alternative to the automatic contingency selection method. It is apparent from the results obtained with the DC-Load Flow, that this method is accurate and reliable. Therefore, for on-line contingency analysis, serious consideration should be given to the possibility of using a DC-Load Flow model for the selection of meaningful transmission line contingencies.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the various suggestions made by Dr. B. Wollenberg of Power Technologies, Inc., and Dr. R. Campo and Mr. P. Ruiz of American Electric Power Service Corporation.

REFERENCES

[1] Ejebe G. C. and Wollenberg, B. F., "Automatic Contingency Selection", paper F78 229-9, presented at the IEEE-PES Winter Meeting, New York, NY, January 1979.

[2] Clements K. A. and Ringlee R. J., Final Report - Research on Adequacy Assessment of Interconnected Electric Power Systems, Worcester Polytechnic Institute, Electrical Engineering Department, December 1977.

[3] Dopazo J. F., Irisarri G., and Sasson A. M. "Real-Time External System Equivalent for On-Line Contingency Analysis", paper F78 674-4, presented at the IEEE-PES Summer Meeting, Los Angeles, CA, July 1978.

[4] Fischl R. and Puntel W. R., "Efficient Method for Computing Electric Power Transmission Network Sensitivities", paper C72 167-0, IEEE Winter Power Meeting, New York, NY, January 1972.

[5] Fischl R. and Puntel W. R., "Computer-Aided Design of Electric Power Transmission Networks," paper C72 168-8, ibid.

APPENDIX I

First-Order Sensitivities

The gradient of the performance index of eq. (4) with respect to changes in susceptance of the lines in the network is

$$(\nabla_B J_{MW})^T = \left[\frac{\partial J_{MW}}{\partial B_1}, \ldots\ldots, \frac{\partial J_{MW}}{\partial B_{NL}} \right] \tag{I.1}$$

where the ℓth element is

$$\frac{\partial J_{MW}}{\partial B_\ell} = \left[\frac{\partial J_{MW}}{\partial |\underline{\theta}|} \right]^T \left[\frac{\partial |\underline{\theta}|}{\partial B_\ell} \right] \tag{I.2}$$

with $|\underline{\theta}|$ a vector of angle differences across the lines in the network. Notice that the elements of vector $[\partial J_{MW}/\partial |\underline{\theta}|]$ in eq. (I.2), can be readily obtained from eq. (4). On the other hand, element jth of $[\partial |\underline{\theta}|/\partial B_\ell]$ is equal to

$$\frac{\partial |\theta_j|}{\partial B_\ell} = \text{sign}(\theta_j) \frac{\partial \theta_j}{\partial B_\ell}, \quad \theta_j \neq 0 \tag{I.3}$$

To calculate the derivative $\partial \theta_j / \partial B_\ell$ in eq. (I.3), consider the DC-Load Flow formulation

$$-\underline{B}\,\underline{\delta} = \underline{P} \tag{I.4}$$

where $\underline{\delta}$ is the vector of nodal angles and P is the vector of nodal injections. From eq. (I.4) it is clear that

$$\frac{\partial \underline{\delta}}{\partial B_\ell} = -\frac{\partial \underline{B}^{-1}}{\partial B_\ell}\,\underline{P} \tag{I.5}$$

but the derivative of the inverse susceptance matrix is equal to

$$\frac{\partial \underline{B}^{-1}}{\partial B_\ell} = -\underline{B}^{-1}\,\frac{\partial \underline{B}}{\partial B_\ell}\,\underline{B}^{-1} \tag{I.6}$$

using eq. (I.6) and noticing that $\partial \underline{B}/\partial B_\ell = \underline{M}_\ell \underline{M}_\ell^T$ it can be easily shown that

$$\frac{\partial \underline{\delta}}{\partial B_\ell} = -\underline{B}^{-1}\,\underline{M}_\ell\,\theta_\ell \tag{I.7}$$

and from eqs. (I.3) and (I.7) it is finally seen that

$$\frac{\partial |\theta_j|}{\partial B_\ell} = -\text{sign}(\theta_j)\,(\underline{M}_j^T\,\underline{B}^{-1}\,\underline{M}_\ell)\,\theta_\ell \tag{I.8}$$

To calculate the elements of the gradient vector, eq. (I.2) can be written as

$$\frac{\partial J_{MW}}{\partial B_\ell} = \sum_{j=1}^{NL} \frac{\partial J_{MW}}{\partial |\theta_j|}\,\frac{\partial |\theta_j|}{\partial B_\ell} \tag{I.9}$$

and using eq. (I.8) in (I.9) and the fact that $\underline{M}_j^T \underline{B}^{-1} \underline{M}_\ell = \underline{M}_\ell^T \underline{B}^{-1} \underline{M}_j$, it is seen that

$$\frac{\partial J_{MW}}{\partial B_\ell} = -\underline{M}_\ell^T\,\underline{B}^{-1}\left[\sum_{j=1}^{NL} \frac{\partial J_{MW}}{\partial |\theta_j|}\,\text{sign}(\theta_j)\,\underline{M}_j\right]\theta_\ell \tag{I.10}$$

Let

$$\hat{\underline{P}} = \sum_{j=1}^{NL}\left[\frac{\partial J_{MW}}{\partial |\theta_j|}\,\text{sign}(\theta_j)\,\underline{M}_j\right] \tag{I.11}$$

and

$$\hat{\underline{\delta}} = \underline{B}^{-1}\,\hat{\underline{P}} \tag{I.12}$$

Using eq. (I.12) in (I.10) and letting $\hat{\theta}_\ell = \underline{M}_\ell^T \hat{\underline{\delta}}$, it is finally obtained

$$\frac{\partial J_{MW}}{\partial B_\ell} = -\hat{\theta}_\ell\,\theta_\ell \tag{I.13}$$

which is the ℓth (ℓ = 1,2, ..., NL) element of the gradient vector of eq. (I.1).

Second-Order Sensitivities

From eq. (I.9) it can be shown that the second order derivative $(\partial J_{MW}/\partial B_k\,\partial B_\ell)$ is given by

$$\frac{\partial^2 J_{MW}}{\partial B_k \partial B_\ell} = \sum_{j=1}^{NL}\left\{\frac{\partial J_{MW}}{\partial |\theta_j|}\,\frac{\partial^2 |\theta_j|}{\partial B_k \partial B_\ell} + \frac{\partial^2 J_{MW}}{\partial |\theta_j|^2}\,\frac{\partial |\theta_j|}{\partial B_k}\,\frac{\partial |\theta_j|}{\partial B_\ell}\right\} \tag{I.14}$$

In this equation the only term that needs further elaboration is the derivative $(\partial^2 |\theta_j|/\partial B_k\,\partial B_\ell)$. The other terms can be obtained from previous results.

It can be easily seen that

$$\frac{\partial^2 |\theta_j|}{\partial B_k \partial B_\ell} = \text{sign}(\theta_j)\,\frac{\partial^2 \theta_j}{\partial B_k \partial B_\ell}, \quad \theta_j \neq 0 \tag{I.15}$$

From eq. (I.8), and using similar manipulation as that used for the first-order case, it can be shown that eq. (I.15) can be written

$$\frac{\partial^2 |\theta_j|}{\partial B_k \partial B_\ell} = \text{sign}(\theta_j)\,(\underline{M}_\ell^T \underline{B}^{-1} \underline{M}_k)\,(\underline{M}_j \underline{B}^{-1})\,(\underline{M}_\ell \theta_k + \underline{M}_k \theta_\ell) \tag{I.16}$$

Notice that this is a general formula that can be used to simulate the effects of double contingencies. However, for the case of single contingencies, eq. (I.16) reduces to

$$\frac{\partial^2 |\theta_j|}{\partial B_\ell^2} = 2\,\text{sign}(\theta_j)\,(\underline{M}_\ell^T \underline{B}^{-1} \underline{M}_\ell)\,(\underline{M}_j^T \underline{B}^{-1} \underline{M}_\ell) \tag{I.17}$$

Using eq. (I.17) in (I.14) (with $k = \ell$) it is obtained

$$\frac{\partial^2 J_{MW}}{\partial B_\ell^2} = 2\,(\underline{M}_\ell^T \underline{B}^{-1} \underline{M}_\ell)\,(\underline{M}_\ell^T \underline{B}^{-1})\,\hat{\underline{P}}\,\theta_\ell$$

$$+ (\underline{M}_\ell^T \underline{B}^{-1})\left\{\sum_{j=1}^{NL} \frac{\partial^2 J_{MW}}{\partial |\theta_j|^2}\,\underline{M}_j \underline{M}_j^T\right\}(\underline{B}^{-1} \underline{M}_\ell)\,\theta_\ell^2 \tag{I.18}$$

Let

$$\hat{\underline{B}} = \sum_{j=1}^{NL}\left\{\frac{\partial^2 J_{MW}}{\partial |\theta_j|^2}\,\underline{M}_j \underline{M}_j^T\right\} \tag{I.19}$$

Substituting eq. (I.19) into (I.18) and simplifying it is finally obtained that

$$\frac{\partial^2 J_{MW}}{\partial B_\ell^2} = 2\,(\underline{M}_\ell^T \underline{B}^{-1} \underline{M}_\ell)\,\hat{\theta}_\ell\,\theta_\ell + \underline{M}_\ell \underline{B}^{-1} \hat{\underline{B}}\,\underline{B}^{-1} \underline{M}_\ell \theta_\ell^2 \tag{I.20}$$

which can be used, together with eq. (I.13), to calculate the changes in the performance index to changes in the susceptance of line ℓ using eq. (11).

DC-Load Flow

From eq. (I.1) the nodal angles, after a change in susceptance for line ℓ of ΔB_ℓ, are given by

$$\underline{\delta}^\ell = -(\underline{B} + \Delta B_\ell \underline{M}_\ell \underline{M}_\ell^T)^{-1} \underline{P} \tag{I.21}$$

Using the well known matrix inversion lemma it can be shown that

$$\underline{\delta}^\ell = \underline{\delta} - b_\ell \underline{B}^{-1} \underline{M}_\ell \theta_\ell \tag{I.22}$$

where b_ℓ is given by eq. (13) and $\underline{\delta}$ is the vector of nodal angles corresponding to the base case conditions.

The new line flows are

$$P_k^\ell = \begin{cases} -B_k \theta_k^\ell & , k \neq \ell \\ -(B_\ell + \Delta B_\ell)\theta_\ell^\ell & , k = \ell \end{cases} \tag{I.23}$$

Substituting θ_k^ℓ in eq. (I.23) from eq. (I.21) it is finally obtained

$$P_k^\ell = \begin{cases} P_k + b_\ell (\underline{M}_k^T \underline{B}^{-1} \underline{M}_\ell) B_k \theta_\ell , & k \neq \ell \\ (P_\ell - \Delta B_\ell \theta_\ell)[1 - b_\ell (\underline{M}_\ell^T \underline{B}^{-1} \underline{M}_\ell)], & k = \ell \end{cases} \tag{I.24}$$

The new flows given by eq. (I.24) can be substituted in eq. (1) to calculate the value of the performance index.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA IX: DISEÑO FUNCIONAL DE CENTROS DE CONTROL DE ENERGIA

DR. ALBERTO MAYER SASSON

ENERO, 1979.

AN OVERVIEW OF POWER SYSTEM CONTROL CENTERS

Tomas E. Dy Liacco, Fellow, IEEE

The Cleveland Electric Illuminating Company
Cleveland, Ohio

ABSTRACT - This paper describes the objectives, functions, and elements of power system control centers. The distinguishing feature of modern control centers is the addition of monitoring and control functions related to security of system operation. The four major elements of control center design - data acquisition and control, computers, man-machine interface, software, and people - are discussed briefly.

INTRODUCTION

Throughout the electric utility industry in the world today, traditional dispatcher's offices are giving way to modern system control centers [1]. This change from the old to the new is not merely one of modernization of dispatching and supervisory equipment, although there are indeed many new centers which provide little more than what used to be done with old-style equipment. What is significant is the change from a limited concept of generation dispatching or supervisory control to a more comprehensive and integrated approach to monitoring and controlling a power system. This broader concept which we shall refer to as the "security control" concept [2], has stimulated ongoing development of the hardware and software functions which characterize the current trends in power system control center design.

One may ask why a modern control center is needed at all. Or rather, why should a control center be designed to do more than generation and supervisory control? The answer, of course is that there are many difficult operating problems which are outside the purview of these traditional controls. The nature, complexity, and severity of these operating problems vary from one company to another. But whatever the company and its specific conditions, the resolution of these day-to-day and even minute-to-minute operating problems had historically been left solely to the human operator. The operator had generally been left to his own resources to make operating decisions with little relevant information about the nature of a problem and of the implications of each decision alternative that he could consider. Ironically, in cases of really serious and complex situations the operator would be flooded with too much information which tended to hamper or confuse rather than assist his decision-making process.

Over the years, improvements were being made in the traditional central controls. By the end of the 1960's, the state-of-the-art in generation control had evolved from an analog system to a digitally-directed analog and eventually to a full digital control system. Similarly, supervisory control systems had evolved from one hardwired master per remote to one hardwired master for several remotes and eventually to a digital computer master. Thus, by the end of the 1960's there were in service two types of digital computer control installations -- the dispatch computer and the supervisory control computer -- using small computers and requiring no more data than what was essential for either generation dispatch or for supervisory control. Digital telemetry was also coming into use to replace analog telemetry. Up to this time the man-machine interface consisted of strip-chart recorders, loggers, indicating lights, annunciators, console pushbutton panels, thumbwheels, and other special-purpose hardware. Later, black and white or color CRT's were introduced in the design of these traditional centers but the monitoring and control objectives still remained limited to generation dispatch and supervisory control.

The first significant step in expanding the scope of control centers was made with the addition of system security consideration to the generation control and supervisory control requirements. This factor caused radical changes in the real-time data requirements, the amount and sophistication of information processing, the computer configuration, and in the design of the man-machine interface. We can correctly say that the difference between a modern control center and a traditional one is the incorporation of functions related to security.

SYSTEM OPERATION AND SECURITY CONTROL

The goal of system control center design is the implementation of security control.

Security control requires the proper integration of both automatic and manual control functions, i.e., a total systems approach with the human operator being an integral part of the control system design. Security control requires that all conditions of operation be recognized and that control decisions by the man-computer system must be made not only when the power system is operating normally, but also when it is operating under abnormal conditions.

The nature of security control in terms of a design organization or structure was presented at the 1966 IEEE Summer Power Meeting in the paper, "The Adaptive Reliability Control System" [3]. The concepts advanced in this initial paper were further developed in a succeeding work, "Control of Power Systems via the Multi-Level Concept" published in 1968 [4]. I will briefly discuss the basic ideas originally presented in these two references.

The Overall Operating Problem

The power system may be assumed as being operated under two sets of constraints: load constraints and operating constraints.

The load constraints impose the requirement that the load demands must be met by the system. The operating constraints impose maximum or minimum operating limits on system variables and are associated with both steady-state and stability limitations. Mathematically, the load constraints can be expressed in the form of the familiar load flow equations. The operating constraints can be expressed in the form of inequalities, such as on equipment loadings, bus voltage, phase angle differences, generator real and reactive powers, etc.

The conditions of operation can then be categorized

into three operating states -- normal (or preventive), emergency, and restorative.

A system is in the normal state when the load and operating constraints are satisfied. It is reasonable to assume that in the normal state the power system is in a quasi-steady-state condition. For any given time, the intersection of the load constraints and the operating constraints defines the space of all feasible normal operating states. The power system may be operated anywhere in this space.

A system is in the emergency state when the operating constraints are not completely satisfied. Two types of emergency may be noted. One is when only steady-state operating constraints are being violated, e.g., an equipment loading limit is exceeded or the voltage at a bus is below a given level. The other is when a stability operating constraint is violated and as a result of which the system cannot maintain stability. The first type of emergency may be called "steady-state emergency" and the second type, "dynamic emergency." For the moment, however, we shall not distinguish between the two types of emergency.

A system is in the restorative state when the load constraints are not completely satisfied. This means a condition of either a partial or a total system shutdown. In case of a partial shutdown the reduced system may be in an emergency state. This is the start of a cascading situation, and if uncorrected, would lend to a further deterioration of the system.

The concept of three operating states breaks up the complex overall operating problem into three operating sub-problems with different control objectives. Of primary interest and of major impact on the design of system control centers is the control done in the normal state. It is basically the development and implementation of functions in this area that represent the state-of-the-art in system control centers. Emergency and restorative controls are needed for a complete security control system, but so far their implementation at control centers has been very limited in scope and in ingenuity.

The effectiveness of security control depends heavily on the control done during the normal operating state. If a system could be controlled so that it remains normal 100% of the time, then all the load constraints would be met without any problem and there would exist the maximum opportunity for realizing the full economic benefits of sound operation. The objective of security control may therefore be restated as follows: to keep the power system operating in the normal operating state, i.e., to prevent or to minimize the departures from normal state into either the emergency or the restorative state. To realize an effective strategy for carrying out this objective let us look more closely into the concept of system security.

The Concept of System Security

System security may be considered as the ability of a power system in normal operation to undergo a disturbance without getting into an emergency condition. The system is then said to be "secure." On the other hand, a normal operating system would be "insecure" if there were a disturbance which could bring about an emergency operating condition. If one considers all possible disturbances it would be impossible to find a secure power system. In practice system security is determined with reference to an arbitrary subset of the complete disturbance set. This subset is called the "next-contingency" set. The choice of the composition of the next-contingency set is dictated by the probability of occurrence of the contingency within the next short period of time (in the order of minutes) and the consequences to the system should the contingency occur. In most power systems the next-contingency set includes, as a minimum, the following types of disturbances:

1) any circuit out;
2) any generating unit out;
3) Any phase-to-phase or 3-phase short circuit.

Other types of disturbances may be added. The more disturbances included in the next-contingency set the more stringent the system-security requirements become.

For a given next-contingency set, the set of all normal operating states may be partitioned into two disjoint subsets--secure and insecure. That is, a normal operating system is either secure or insecure. We see then that for security control to accomplish its objective of preventing or minimizing departures from the normal state it would be highly desirable to be able to identify, firstly, whether the system is normal or not, and secondly, if normal, whether the system is insecure or not, and thirdly, if insecure, what corrective action may be recommended to make the system secure. This leads us to the functions of security monitoring and security analysis which we shall discuss a little later.

General Characteristics of Security Control

We should recall the previously stated objective of system control center design as the implementation of security control in the broad sense of integrating all required automatic and manual functions for all conditions of operation. From this prospective we can see the general patterns in which system control centers have been developing in the recent years.

The necessity for integration has brought together the previously separately implemented functions of generation control and transmission control into one system. For geographically small power systems the integration is carried out in the system control center. For large systems or systems with existing regional or area control centers this integration is accomplished by linking the centers at various levels into a computer hierarchy.

The new requirement of security monitoring alone has necessitated the collection of a large volume of real-time system data every few seconds and has brought about the use of filtering and state estimation techniques.

In addition, the integration of automatic and manual functions is being manifested in the form of advanced display devices and techniques. The CRT with limited graphics has become the universal man-machine interface for system control centers.

Operating decisions by the human operator are being supported by the presentation of more complete and coherent information about the power system than was ever done before.

If we review the modern system control centers which have been placed in service throughout the world since the start of the 1970's and also those which are in the process of implementation we come up with a totality of real-time features and functions which includes the following:

1. Hierarchical structures consisting of several levels of computer systems.
2. Dual real-time processors or multi-processors plus redundant peripherals.
3. High-speed digital telemetry and data-acquisition equipment.
4. System-wide instrumentation of electrical quantities and device status.
5. Color CRT's with graphics for interactive display.
6. Dynamic wallboard group display.
7. Automatic generation control.
8. Economic dispatch calculation.
9. Automatic voltage (Var) control.
10. Supervisory control (breakers, capacitors,

transformer taps, generating unit startup and shutdown).

11. Security monitoring.
12. State estimation.
13. On-line load flow.
14. Steady-state security analysis.
15. Optimum power flow.
16. Automatic system trouble analysis.
17. On-line abort-circuit calculation.
18. Emergency control -- automatic load shedding, generator shedding, line tripping.
19. Automatic circuit restoration.

There is no control system that has all of the functions just enumerated. This is to be expected. Operating problems differ due to different networks and generation resources. Operating philosophy and the structure of operating responsibilities are not the same for all companies. A few centers have adopted an evolutionary approach, adding something new to existing control equipment and telemetry. Finally there is the significant time gap between the testing of a new idea on paper and its implementation in a real-time control system.

FUNCTIONS OF A CONTROL CENTER

In this section we will summarize only the real-time functions that are generally of widespread concern to system operation. There are of course other functions which belong to the structure of security control but which are usually run off-line or in a batch processing mode. This is not to say that these latter functions are not important, but we recognize that the difficult part of system control center design lies in the implementation of functions which run in the real-time environment.

Automatic Generation Control (AGC)

The function of Automatic Generation Control (AGC) is to determine the generation required to meet the actual system load and to allocate this generation among the regulating units, coordinating the requirements of regulation with the desired base operating point of each unit. The last part of this definition identifies an important interface between AGC and some other function which calculates the desired base points or settings. Traditionally the base settings are determined by the economic dispatch function. But in our concept of security control this need not always be the case. During certain operating conditions, other functions such as security analysis or emergency control could establish the desired base operating points.

The basic AGC algorithms, i.e., the calculation of area control error and the assignment of regulation to each unit recognizing the desired base points, are well-known. To apply these algorithms in a system control center requires the addition of modules which in effect interface with the real-time environment. These modules should initialize the AGC function, coordinate all information from other programs which affect AGC, prepare and hand off to the data-acquisition subsystem the signals to be sent to the plants, and communicate with the display subsystem.

The use of plant computers communicating with the system control center offers felxibility for carrying out the AGC function. An example of this application is at the Cleveland Electric Illuminating Company (CEI). The AGC software at CEI's system control center sends desired mw signals for each regulating unit to the plant computers. The plant computers act as local closed-loop controllers for each unit [5]. The control algorithms at the plant computers recognize the individual rate of response of each unit. Over the same data links the plant computers report to the system control center every second the control status of each unit and its short-term raise and lower capability. This information is used by the AGC algorithm such that the desired mw requested is within the dynamic capability of the unit. The computer-to-computer link also handles special requests by a unit operator to place a unit off or on regulation or to change a unit's operating limits.

Economic Dispatch Calculation (EDC)

Economic dispatch calculation is performed every few minutes using the set of coordination equations which requires that the incremental cost of delivered power from each generating unit to an arbitrary reference point be the same for each unit. The incremental cost of delivered power to a given point from a generating unit is equal to the incremental cost of generated power multiplied by a penalty factor. Traditionally the penalty factors are calculated using transmission loss B-constants.

In present day control centers, B-constants are usually calculated off-line and are updated very infrequently. There is an economic advantage to be gained in updating B-constants on-line especially in these times of high fuel costs.

In centers where a real-time load flow is required for other reasons, it would be possible to calculate the penalty factors on line by adding a real power optimization routine thus obtaining an optimum power flow [5],[6]. At CEI, every time there is a network change or when the system load has changed significantly in magnitude or in relative distribution between areas, the optimum power flow runs automatically and a new set of penalty factors is passed on to the EDC. The penalty factor calculation takes less than 40 seconds on the Sigma 5 computer. This is the total response time and includes network configuration update, 3 to 4 fast decoupled load flows [7], Jacobian calculation at the optimum solution point, calculation and transfer of new penalty factors to the data base.

Although EDC should be made only for those units which are regulating, it is desirable to make another calculation including all the other units on local control. This second-pass EDC is made everytime the regular EDC is run. The results of the second-pass EDC are displayed to the operator so that he may manually direct the units on local control to be moved closer to their optimum generating points. Considerable additional economy may be realized this way.

Supervisory Control (SBC, SVC)

Supervisory control is not a new operating function. Its integration into a system control center is new. Since supervisory control is a manual function it is exercised via the man-machine interface or the display subsystem.

The integration of supervisory control of circuit breakers (SBC) is not always straightforward in the case of a control hierarchy where, historically, supervisory control had been exercised at a lower level, such as a district or regional office. The common approach has been to retain this function at the lower level and morely report the breaker status to the central or higher level. Such a structure will need re-examination of the interfaces in the event that there would be a requirement for breaker control from the higher level. An example of this would be some form of emergency control such as load-shedding or system splitting.

When there is a need, the control center may also perform supervisory control of voltage regulating devices (SVC), such as tap-changers, capacitors, generator voltage regulators.

Automatic Voltage/Var Control (AVC)

The automatic control of system voltage and of var allocation is not yet in wide use even by those companies who feel they need it, primarily due to the absence of an efficient on-line optimization algorithm. In Japan, however, AVC has been in use for many years now.

The AVC regulates the voltage profile and also minimizes losses due to reactive power flow [8]. The control variables are generator reactive powers, transformer taps, shunt capacitors, and shunt reactors. The control is a two-step operation. Voltages and var flows are checked periodically and when there is any deviation beyond certain tolerances the voltage profile control calculation is initiated. At less frequent intervals the minimum loss calculation and control is executed.

In 1976, the Potomac Electric Power Company placed in service their new control center [9] and one of its unique features is a closed-loop voltage control [10] of distribution bus voltages, the first of its kind in the world.

Security Monitoring (SM)

Security monitoring (SM) is the on-line identification and the display of the actual operating conditions of the power system. This one function has made the difference between the traditional dispatch center and the modern system control center. SM requires a systemwide instrumentation on a greater scale and variety than that required by a center without SM. The types of measurements include: MW and MVAR flows, branch currents, bus voltages, bus MW and MVAR injections, frequencies, energy readings, circuit breaker status or operation counts, manual switch positions, protective relaying operations, transformer tap positions, and miscalleneous substation status and alarms.

The SM function, in general, checks the analog values against limits basically to determine whether the system is close to, or at, the emergency state. The limit-checking also allows some kind of data validation and the rejection of incongruous data. Limit-checking is done as often as the data is brought in which is usually in the order of every one to a few seconds.

The display required for SM entails the use of CRT's and a large number of display formats. The dynamic wall display is also used for SM. Part of the SM function is the on-line determination of the network topology [11], [12]. In most cases it is sufficient to determine the network configuration. In centers where there is a direct responsibility for transmission switching and safety is a paramount factor, the SM function should include an identification of the electrical status (energized or de-energized) or every physically isolatable segment.

Static State Estimation (SE)

State estimation (SE) may be defined as a mathematical procedure for calculating, from a set of system measurements, a "best" estimate of the vector of bus voltage magnitudes and phase angles of the network.

The measurement set is understood to contain an adequate degree and spread of redundancy to allow the statistical correlation and correction of the measurements, detect and preferably identify bad data, and yield calculated values for non-telemetered quantities.

An excellent summary of SE and its methods is given in the 1974 Proceedings of the IEEE paper by Schweppe and Handschin [13].

Although there are just a few control centers with SE in operational use the value to operation of this function is becoming more widely acknowledged. Recent specifications for control centers include SE as part of the software requirements. As presently practiced, SE is used for the following purposes:

- bad data identification
- calculation of non-telemetered or missing data
- provide inputs to security monitoring function
- provide vector of bus injections for an on-line load flow, security analysis, and bus load forecasting.

On-Line Load Flow (OLF)

By "on-line load flow" I do not mean a load flow that is made available to the operator for planning or study purposes. However such a load flow is run, either by conventional batch processing or interactively, it is still an off-line load flow. An on-line load flow (OLF) is one which is used for real-time functions such as security monitoring, security analysis, and penalty factor calculation, and can also be used for study purposes. OLF makes use of real-time data.

The OLF requires a vector of bus injections. In the general case, the bus injections are calculated from statistical data obtained on-line and some off-line historical information. The bus injections may also be obtained from the results of a state estimation program. These injections may be used as they are or normalized to produce a set of load distribution factors. These distribution factors may be projected to a future time for predictive purposes.

The on-line load flow is a necessary function for system control centers. It should not be interpreted, however, as supplanting state estimation. As we have seen, these two functions serve different needs. Since the on-line load flow uses bus injections which are statistical in origin, the ultimate OLF should give results with some kind of statistical interpretation, i.e., an stochastic load flow. We are not yet there with the present state-of-the art. However, the basic formulation of the OLF for penalty factor calculation, for establishing the base case of security analysis, as an alternative method for performaing contingency evaluation is of value now at system control centers.

Steady-State Security Analysis (SA)

The first function of security analysis (SA) is to determine whether the normal system is secure or insecure. The second function is to determine what corrective action strategy should be taken when the system is insecure.

The first function is commonly known as contingency evaluation since by definition, the security of a system is determined with reference to a set of next-contingencies. In present state-of-the-art, only steady-state contingency evaluation is done at system control centers. That is, the emergency condition that is to be avoided is overloading of equipment or poor bus voltages. There is still nothing in the way of dynamic security analysis.

The earliest method used for contingency evaluation is the distribution factor method derived from elements of the bus reactance or bus impedance matrix [14], [15]. This method is used at several control centers. The same approach is used for determining a feasible, though not necessarily optimal, corrective action.

Load flow methods are also in use for security analysis. Among these techniques are: DC load flow, Gauss-Seidel, Newton-Raphons, linearized AC [16], and Stott's Fast Decoupled Load Flow. The last mentioned method has the advantage of having the same algorithm useable to obtain either an approximate solution or a full AC solution. The approximate solution is comparable in speed to the method of distribution factors but it is more reliable and accurate in that the voltage profile is taken into account. The first iteration of the Fast Decoupled Load Flow yields the approximate solution. If a full AC solution is desired further iterations are run until the mismatch requirement is satisfied.

A recent survey of security analysis methods is

given in [17].

Security analysis as presently modelled requires an up-to-date equivalent of the external interconnection. So far, the only equivalent available and used at control centers has been the traditional Ward equivalent which has several recognized shortcomings. There is now a revived interest in equivalents for security analysis. Two basic types are emerging: topological and non-topological. Topological equivalents, like the Ward equivalent, are derived from prior knowledge of the detailed external system. Non-topological equivalents require no physical network information but are derived from real-time measurements via stochastic approximation techniques. A recently developed topological equivalent [18] based on Dimo's REI method [19] has features for on-line application not available with the Ward equivalent. Work on non-topological equivalents is continuing and initial results have been reported in the literature [20],[21].

As discussed previously, the space of feasible normal states may be partitioned into secure and insecure regions. This, of course, is a dynamic situation. As the system generation, load, and topology change so does the space of normal states and so does the boundary between secure and insecure regions. In fact, either region could be a null subspace. Clearly, as system conditions change the contingencies in the next-contingency set which yields insecure operating points also change. If at times the system is very strong that no contingency in the next-contingency set can cause an emergency, the insecure region is null and contingency evaluation is not required. At other times only certain contingencies need be evaluated. This leads us to the idea that we should have a more scientific or systematic way of determining on-line whether there is any need to do security analysis and if so, which contingencies we should be looking at. Presently, we do not have any techniques for accomplishing this. We are thus compelled to use a fixed list of contingencies, perhaps with some spare room for operator-specified contingencies. Since the security analysis routines could impose a large computational burden, in certain centers the next-contingency list is pared down to a small number of items. This is not always possible. There could still be enough contingencies to cause loading problems of computer resources. Part of the problem is the requirement that security analysis be run periodically, 24 hours a day. An alternative approach would be to use the Security Monitoring function to determine whether or not there is a need for SA. This could be based on arbitrary levels of line loadings.

A detail sometimes overlooked in control center specifications is the fact that in many power system networks there are multi-terminal lines, such as lines with a tap for a transformer connection. For a 3-terminal line, a line outage would mean an outage of three load flow branches and the isolation of one node. This fact is often lost sight of by a software designer with little power system background. The contingency evaluation program gets erroneously developed on the basis of a line outage being a branch outage in the load flow sense.

DESIGN CRITERIA

The successful performance of control center functions such as those described in the preceding section depends on the adherence to certain design and performance criteria. The three most important design and performance criteria for a system control center are: system response, system availability, and system maintainability.

System response is measured in terms of the time it takes from the instant a function is requested until the instant the outputs from that function are available. The response time requirement depends upon the nature of the function in question. The actual response time obtained depends upon the speed of the hardware reliability or software reliability, or both combined. The measure of system availability also depends upon system response. For any given function, the availability, A, is given by:

$$A = \frac{\text{Available Time}}{\text{Period of Interest}}$$

$$= 1 - \frac{\text{Unavailable Time}}{\text{Period of Interest}}$$

Unavailable time is the total time during the period of interest when the function is not available. Any time beyond the maximum prescribed response of a function is also considered as unavailable.

It is an extreme design requirement to specify the same availability for all functions, critical and non-critical. An overall system availability is difficult to define let alone measure. Besides, an overall system availability requirement does not necessarily ensure a good response and availability of a critical function. It would therefore be more reasonable and preferable to specify two availability figures, one for critical functions and another for non-critical functions. A critical function which should have a high availability is the man-machine interface. The availability of this function could be used as a simple, readily measurable criterion for system design and performance. The thinking behind this is that as long as the man-machine interface is available the operator is not completely helpless. Even if other operating functions were not working, the operator could do something manually if the interface were there to provide some information and to permit manual corrections.

Achievement of good response and high availability should be pursued from the very start of system design, through implementation, and during the life of the system. Very much a factor in this achievement is system maintainability. The levels of response and availability are obviously affected by hardware and software maintenance. The repair times following hardware or failures depend upon the maintenance capability, diagnostic aids, and equipment that are available to maintenance personnel. Preventive maintenance, system debugging, corrections, updates, tests, and enhancements are on-going activities which have to be performed using the computer system facilities. The system design from the very beginning should provide for this type of work to be done at any time with little or no impact on the performance of the real-time system.

Software maintenance could be a serious problem in a system control center. Even with an adequate staff of trained people, it is highly advisable to have enough computerized maintenance and testing aids in order to reduce significantly the time required to do maintenance. Just like system response and availability, system maintenance must be designed into the system at the start and not as an after-thought.

CONTROL SYSTEM COMPONENTS

A system control center consists of the following elements or subsystems: data-acquisition and control; communications; computers; display; software; uninterruptible power supply; the building; and people. The communication channels, the power supply and the building facilities design are all important to the proper functioning of a control center will not be discussed here. Their requirements are not as intimately woven in with the control design problem as are those of the data-acquisition, the computer, the display, and software subsystems, and people.

The Data-Acquisition and Control Subsystem

The data-acquisition and control subsystem consists of: remote terminal equipment for interfacing

with power system instrumentation and control devices; interfaces with communication channels; and master station equipment for interfacing with the system control center. In some centers a dedicated channel is assigned to each remote station. In others there are less channels than remote stations requiring more than one remote to share a channel. Analog data is scanned periodically in the order generally of 1 second to a few seconds. Each scan is triggered by the system control center at the prescribed interval by using a request to all remote stations to send in data. Data is received at the master equipment in a random order. The hardware equipment which converts the bit-serial data into a bit-parallel word does error-checking and raises an interrupt to the computer for each word received. There are two approaches to this: one is to have a single interrupt for all channels; the other is to have one hardware interrupt for each channel. The single-interrupt method requires polling by a software routine to find out where the data word came from. The multiple interrupt approach results in a much better response time due to the very fast interrupt processing.

Status data is also processed in the same way as analog data except that there are two ways of reporting status changes. The first way is to send in all status information from all remotes at the required intervals regardless of whether or not there has been a change. This approach requires a software routine at the system control center to check each new status with the old status to determine any changes. Considering the very large number of status points that is monitored in a power system this approach represents a sizable burden on the central processor at the control center. The second way is to send status data from the remote only when there has been an actual change of status. Since normally the system is quiescent and since, if there are any status changes, only a certain number of stations are involved, the second method results in a better overall system response for the same amount of computer resources. There are, however, many systems in service which use the continuous status scan approach and which apparently are not bothered by this processing overhead. At least, now yet. Having a not-so-frequent scan helps. Assigning data-acquisition to a front-end computer also helps.

The use of front-end computers for the data-acquisition function is a desirable option as it off-loads the main computers which would be doing the rest of the real-time functions. In some applications the front-end computer serves only as a message switcher. This does not help system response.

The data link procedures and word structures are different with each data-acquisition equipment manufacturer and sometimes with different models from the same manufacturers. While this situation creates a constraint on the expansion of an existing system it should not lock in a utility company to an obsolescent model when, due to the fast-changing technology, better and more cost-effective equipment may be available. Microprocessors would resolve this industry problem by making it easier and less expensive to convert from one data format to another. There is now a trend to so-called programmable remotes which would inecitably become microcomputers or minicomputers. Eventually the data-acquisition subsystem would be a computer network using a standard data link format and control.

The data-acquisition software besides managing the collection of data and placing them in computer memory, also performs: error-checking; conversion to engineering units; limit-checking; and interfacing with application programs. For fast response, the data-acquisition software must be: resident in main memory; of the highest hardware priority of all application software; as independent of the operating system as possible, making use of hardware interrupts for scheduling its own I/O's. The real-time database must also be resident in main memory.

One should look askance at a system design in which the real-time data base is not resident in main memory. Such a design is handicapped from the start since already a large amount of traffic is being built into the I/O channel. Since the cost of main memory is no longer very high, the extra cost incurred to make the data base resident is a small price to pay to ensure good performance. One should not readily yield to the argument that one's requirements are not too great anyhow and that the design would be adequate. Control centers have a habit of expanding from very humble beginnings. If it turns out that the reason the database cannot be made resident in main memory is that there is no more expandability, then the computer being offered is not the right one to have.

THE COMPUTER SUBSYSTEM

Real-Time Computer Characteristics

A system control center is a real-time system and the computers selected for this application must be designed for real-time. Essentially this means that the computer must have outstanding hardware features and must have a proven and efficient real-time operating system. Some of the hardware features that have been found to be important at control centers are the following:

-- memory cycle times of microsecond or lower

-- multiple external interrupt structure with a fair number of interrupts

-- fast access disk in the order of less than 20 milliseconds access time and transfer rate of better than 250 kbytes per second

-- multiport memory banks with provision for interleaving

-- memory expandability to, at least, 64K 32 bit words or equivalent

-- direct memory access (DMA) with multiplexer for several peripherals sharing the DMA channel

-- floating point hardware

-- internal interrupts for various trap conditions

-- internal real-time clocks

-- watchdog timer.

The actual performance of a computer system for the same hardware depends upon the configuration, the operating system, and the software design.

Computer Configuration

The design criteria of response, availability, and maintainability dictate the use of more than one processor. Placing all functions -- real-time monitoring and control, background processing, software maintenance and testing--in a single processor makes it extremely difficult or impractical to obtain a high level of response and availability. One would have to limit the scope of functions assigned to the digital computer (such as use analog for AGC) and also accept a degraded security monitoring function. But this is not representative of the new breed of system control centers that we are investigating, where functions such as AGC,

EDB, SM, SA, SE, OLF, SBC are being integrated into one system and CRT response times of 1 second or better are common.

There are some control centers with single processors. All of these have analog AGC controllers either as part of the system or available as backup. In such centers, the man-machine interface is evidently not considered critical as it is unavailable whenever the single processor is down.

In the majority of the control centers, the computer configuration used is the "dual" computer system. This is shown in greatly simplified form in Figure 1a. A and B are basically identical computer systems, each consisting of a central processor, main memory, and auxiliary memory.

There are several ways in which functions may be assigned to A and B. This depends primarily on the availability requirements. One way would be to say that all functions, critical and non-critical, as well as some types of background processing, must be fully supportable on one computer. This takes us back to the response problem of a single processor. However, in this case the availability would be much better since there is a second computer in stand-by. The more common practice is to share the functions between the two computers. Critical real-time functions would be assigned, say to A, which would be called "primary", and non-critical real-time functions, off-line functions, and background processing would be assigned to B, designated "secondary".

In case of failure of the primary computer, the secondary can assume the critical real-time functions by manual or automatic failover of the real-time interfaces, such as the data-acquisition equipment and operators' consoles, and initializing itself to the real-time environment. Figure 1b with the dashed line between computers, represents the failover arrangement.

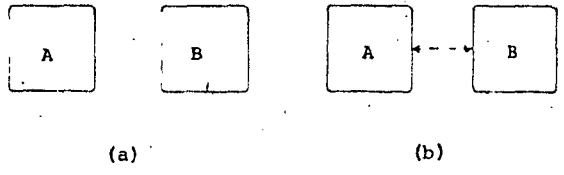


Fig. 1 - Dual Computer Configuration

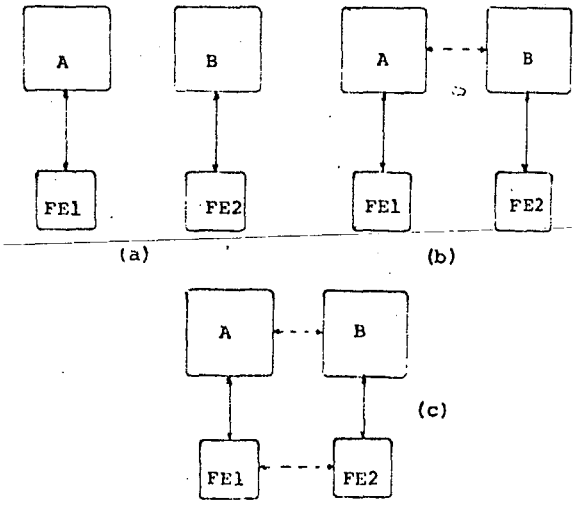


Fig. 2 - Dual Computer Configuration with Front-End Computers

In actual operating experience, a computer system will recover from most failures by simply starting it over again. Doing this automatically is a desirable design feature. Actually, automatic restart is more beneficial to system availability than is automatic failover. The latter function entails some hardware and software complication and works only if the other computer system is available. Failover also takes longer. A few control centers have both automatic restart and automatic failover. Most have automatic failover only. Definitely worth considering in design is automatic restart only plus manual failover.

The automatic restart routine is designed for a certain number of re-tries. Prior to each trial preassigned areas of main memory would be dumped on disk or on magnetic tape for later diagnosis. A minimum uptime is also observed such that, after a restart, if the computer does not stay up for so many minutes, retrials would not be attempted. Obviously if the failure is in the auxiliary memory automatic restart would not be initiated. Failover must be resorted to.

After a failover from A to B, there are some design options as to what functions should be assumed by B. B could take on only the real-time critical functions that A had been doing and abandon all of the other functions. Or B could have all functions that were originally assigned to it plus all of A's, as long as we accept a reduced response and availability. Now this second option is not so bad if you have enough main memory in each computer. This ensures good response and availability most of the time but gives you a slightly degraded performance during the times that the system is down to one computer.

The system design should make it possible to operate one of the computers in a stand-alone mode for maintenance, testing, or large program development. It should also be possible to operate a computer in a pseudo-real-time mode with one console and a data-acquisition channel or two attached to it so that a program change or new program may be tested in a real-time environment.

Each half of the dual configuration need not consist only of one computer. As discussed previously, front-end computers could be used for data-acquisition thus enhancing the response times of the main computer. Figure 2a shows a dual configuration with front-end computers FE1 and FE2. There are two possible schemes for failover. The simple one, shown in Fig. 2b is to consider FE1 as an extension or slave of A and FE2 as a slave of B. The other failover scheme, shown in Fig. 2c allows the front-end computers to be switched to either one of the main computers.

Typically, the front-end computers are 16-bit minicomputers of a size large enough so that either one can handle all of the data acquisition channels. The other would be a purely redundant backup. If the number of channels (or remotes) increase beyond the expandability of the minicomputer it would be time for a re-design with a larger front-end capability. One could at the outset divide the channels between FE1 and FE2 so that both are sharing the work of data-acquisition. Such an arrangement is shown in Figure 3a where both FE's now have links with A and B. The failover scheme is shown in Figure 3b. Since one minicomputer is not quite big enough to handle the entire data-acquisition load then on a failure of FE1 or FE2, the remaining mini would have to do everything in a degraded mode.

There are other possible configurations such as those using multi-processors with shared memories or redundant high-speed data paths. Redundant I/O channels are also possible, in fact desirable. And the concept of distributed processing in the sense of one processor for one function has yet to be explored for control center application. With regard to distributed processing, one should keep in mind the trade-off in complexity of hardware and software for automatic failover.

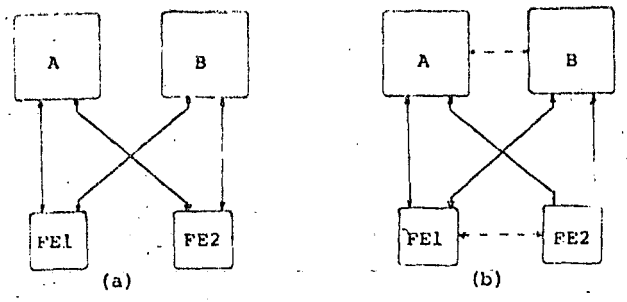


Fig. 3 - Dual Computer Configuration with Task-Sharing Front-End Computers

In spite of the many possibilities in computer configuration, the first-time user can only choose from what the system suppliers have to offer. The suppliers cannot afford to deviate too much from a standard configuration that they have already developed. Hence, even before a new set of specifications are released to vendors one can generally predict the various computer configurations that would be proposed. Most of these configurations are based on a pre-conceived grouping of critical and non-critical functions. There is nothing wrong with this as long as the user has the same grouping in mind. Unfortunately this is not necessarily true. For instance, if one subscribed to the often-held argument that for real-time monitoring and control we must first of all have a reliable data base, one would consider state estimation to be a function as critical as data-acquisition itself and would require that state estimation have the same availability as data acquisition, AGC, and the man-machine interface. None of the currently offered configurations guarantee this kind of availability. One utility engineer has lamented the fact that although his company's modern control center had been authorized on the basis of improved system security, the security-related functions including state estimation had been relegated to the non-critical category.

The Real-Time Operating System

Superior hardware features in a computer system do not necessarily guarantee good performance unless used effectively and to the overall system advantage by the real-time operating system. One of the more difficult problems of control center design and operation is obtaining a computer with a well-designed, field proven, real-time operating system.

Historically computer software has lagged behind hardware development by at least a year. This is still true with some of the new computers that are coming out on the market today. Also some computers, not necessarily new, are not intended by the manufacturers for real-time application and therefore are not supported by any real-time operating system.

The industry picture so far shows: special operating systems developed by vendors for an initial project and offered on subsequent projects; standard real time operating systems with slight modifications and with some enhancements to provide additional real-time features; real-time operating systems operating as a job under a general-purpose operating system; and time sharing operating systems modified to handle real-time processing. The situation has improved recently. There are now some standard real-time operating systems with practically all of the features desired for a system control center.

For a system control center, what is a good real-time operation? I would say one that carried out its functions without unduly impairing system response. The less overhead the operating system takes, the better for the response.

Overhead is incurred and hence response is affected wherever there is a shared resource. Main memory, the I/O channel, subroutines and files are shared resources. Operating systems differ in the way they manage these resources. These are key features of an operating system which should also be considered in software design.

THE MAN-MACHINE SUBSYSTEM

In system control centers the man-machine interface or the display subsystem consists of CRT's, dynamic wall displays, trend recorders, loggers, and alarm devices.

The CRT Display

The cathode-ray tube (CRT) display has become a universal feature of system control centers. This is because the CRT display can provide practically all of the needed interaction with the human operator. While other forms of display may also be necessary in a system control room, these devices do not have anywhere near the utility or the wide-ranging power of the CRT as a device via which the operator can observe, analyze and control the power system.

The CRT display should be designed to meet the needs for man-machine interactions in: power systems operation; control system diagnostics; software development; and control system maintenance. Detailed discussions of these needs plus other CRT design considerations may be found in [22] and [23]. Specific CRT applications at various control centers are described in [5], [6], [24-28].

The CRT idsplay hardware consists basically of a color monitor, display generator with refresh memory, parallel interface unit to the computer system, cursor control devices, and standard keyboard.

To insure fast response, the rate of data transfer through the parallel interface should be in the order of 300 kilobytes per second. The computer architecture should have the facility for this I/O to take place concurrently with CPU processing.

The use of special-purpose function keys should be held to a bare minimum. It is highly desirable to adhere to the standard general-purpose keyboard and to one or more of the standard cursor control options.

The design of interactive procedures between the operator and the computer should be aimed at minimizing the amount of manual inputs required of the operator. This is particularly essential in designing the input procedures for support programs such as the load flow or security analysis. The use of one-line diagrams both for input and output purposes is highly desirable. It should be possible for the operator to indicate specific network configuration changes by simple operations with the light pen on network diagrams. To be an effective and satisfactory man-machine interface, the display subsystem must have a response time of about 2 seconds on the average, from the instant the operator makes a CRT selection to the time requested format is completely on view. In the worst case, it is not unreasonable to require a maximum response time of 5 seconds. If the power system is quiescent, response times of 1 second or less should be readily obtainable.

The frequency of update of dynamic data on a CRT display depends on the purpose of the specific picture on view. For displays used by the system operator analog data may be updated every 10 seconds on the average. Status data should be updated immediately after a change occurs. For displays used by programmers and maintenance personnel the frequency of update could be as often as the data is scanned. Most of the dynamic data is real-time data and since this is assumed to be resident in main memory updating faster than every 10

seconds should not create a serious problem in overhead. The situation would be quite different if the real-time data base were in bulk memory off the I/O channel.

It is desirable to be able to exercise all of the man-machine interface functions at any one of the redundant consoles. When the functions are broken up into dedicated, special-purpose consoles, problems of backup become more difficult to handle and the provisions for redundancy become more costly. Having all consoles alike and each capable of performing all functions in the interface repertoire will make it possible to operate the system even if only one console were in working condition.

The Dynamic Wall Display

The dynamic wall display is intended to give an overview of the power system. The overview concept is best accomplished by a simplified representation preserving as much as possible the geographical orientation of the system. Details on the wallboard not only spoil the overview perspective but, actually, are best left to the CRT's. For example, there is no point in representing all the substation details including manual switches on a wall display when they can all be presented more effectively on the CRT.

There are differences of opinion on the need for dynamic wall display at a control center. My own recommendation is to exclude the dynamic wall display from the control center requirements. It may be considered, as an option, if the cost-benefit ratio could be more attractive through simplification and a reduction of the size to moderate dimensions so that geographical orientation could be perserved. I would, however, recommend consideration of an electrostatic plotter/printer to provide in a few seconds, whenever wanted, a real-time snapshot of the entire system in graphic form. If, in addition to this diagram, something on the wall is still desired, a completely static representation, should be adequate. Again, it should be of moderate size, as trying to cover an entire wall from floor to ceiling distorts the representation to a grossly elongated rectangle.

THE SOFTWARE SUBSYSTEM

The software in a control system may be divided into three categories:

1. System software consists of: the real-time operating system; processors for assembly; compiling, loading and overlay structures, file management, system generation, utility routines for debugging and testing. The system software is usually supplied by the computer manufacturer.

2. Application software includes all the programs which performs tasks for the operation of the power system, such as data-aquisition software, display software, software to implement various control functions, and operation planning programs.

3. Support software are programs used by support personnel for computer system monitoring, real-time diagnostics and debugging, maintenance and testing.

Application Software Development

The software development cycle consists of three parts: analysis and design; coding and debugging; checkout and test.

Analysis and design is the most important phase of software development and should be carried out as thoroughly as possible. Problems in coding, debugging, checkout and testing, and in maintenance, are largely attributable to a poor analysis and design. This initial phase of software development specifies the individual program modules, the inputs, outputs, tables used, tables updated, and the algorithms. Design decisions are made on the database structure and access method, table and file structures, data update requirements, and backup requirements. All hardware-to-software and software-to-software interfaces are spelled out. Initialization procedures, CRT and logger messages, maintenance requirements, test procedures, and acceptance criteria are all specified. All this must be done, reviewed, and agreed upon before one line of code is put on paper. All too often the natural tendency to get going and start coding gets the better of the software designers and the analysis and design effort gets hurried through. This is an invitation to disaster.

The coding and debugging phase typically should take less than 25 percent of the entire software development work. Most of the real-time application software for control centers have been written in assembly language. However, the capabilities of Fortran for real-time use have improved and there is a growing trend toward more use of this language. Whatever language is used careful attention must be given to those program features which affect real-time linkages, response times, and program reliability.

The real-time linkages of a program consist of: accesses to the database; I/O requests to use and/or update files; I/O requests for CRT display and logger messages; program execution requests. Good response means anefficient code, a minimum I/O, effective use of subroutines, and proper use of interrupt control. Program reliability means use of fail-safe logic, proper initialization on system startup, avidance of timing problems, invulnerability to bad parameters, control of possible arithmetic overflow, and proper handling of error returns on I/O and program execution requests.

The checkout and test phase takes up the rest of the software development work. The individual program is tested in the foreground in as complete a real-time environment as possible. The real-time environment is built up gradually as each program is integrated into the system. When the entire system is put together, hardware and software, the individual program tests are repeated as part of the system checkout. The acceptance tests complete the test cycle. The test drivers written for the individual program tests are kept for later use in system maintenance.

Support Software

There are two groups of support software. The first group consists of foreground diagnostic programs which are run on-line to monitor and control the performance of the control system. These programs provide the following functions:

1. Summary and control of remote station status: - On request, the status of the data-acquisition subsystem is displayed on the CRT. The display, which id dynamically updated, shows the scan conditions of each remote. Using this display it should be possible to place any remote off or back on the scan.

2. Display of all data received from a remote station: - On the CRT, all data being received from a selected station may be viewed. This is a dynamic display and the data is seen as it changes from one scan to the next. Color is used to indicate when a piece of data is not updated or when it is out-of-limits. This function is very useful for checking data and for trouble-shooting programs which use the database.

3. Summary of data link errors: - A summary of all types of data link errors is kept on file. The summary shows the remote station, the type of error, the number of times the error has occurred, and the times of the first occurrence and of the last occurrence. This summary is viewable on the CRT. It is periodically printed out on the logger for review by maintenance personnel.

4. Dynamic display of activity in computer overlay area: - This is a CRT display which shows dynamically what programs are in memory executing or waiting for I/O, what programs called them, and what programs are waiting in the queue.

5. Display of main or auxiliary memory and the ability to patch any memory locations: - With this function an area of main or auxiliary memory starting with a specified address may be display on the CRT. In this manner, tables or segments of program code may be examined. The patch capability is useful for on-line debugging, program testing, and an immediate correction of an erroneous condition. The patch on the on-line system is intended to an interim measure until a permanent system revision can be made.

6. On-line measurement of computer system performance: - This on-line function, sometimes known as "software accounting", gathers statistics at specified time intervals about CPU utilization, CPU idle time, I/O waits, number of I/O transfers, what programs have run and how often, etc. This function may run automatically or at operator's request. The statistics are useful for evaluating system loading and performance. The impact of adding new programs may also be measured by taking before-and-after statistics.

7. Dump of real-time data on tape: - Via the CRT the operator can initiate the dumping of selected real-time data at specified time intervals on magnetic tape. The operator specifies on the CRT the data to be dumped and the time interval. The dump continues until the operator stops it by a CRT entry. The magnetic tape is later printed out on the line printer by an editing program on the secondary computer.

Several control centers have some or all of the support functions mentioned above. Actually these programs are of tremendous value during the implementation of the control system and not just after the control center is placed in service. The system design should therefore include such on-line support programs. These should be completed by the system supplier early in the implementation period so that the development work could be speeded up.

The second group of support software consists of programs for system maintenance. These are off-line programs which are used for updating files to match changes or additions in the power system. A necessary file maintenance subsystem is one that updates all files related to the man-machine interface. This includes a "picture compiler" which would allow a maintenance programmer to compose a CRT picture at a console and then run the compiler to generate the CRT code of the picture and store it in the correct file location. In an interactive mode the maintenance program will step through whatever operator inputs are required to update all other tables related to the new picture. A good maintenance program design should minimize the amount of operator inputs. Other file maintenance subsystems may be designed for other families of programs, such as the SE, SA, and OLF group or the AGC, EDC group. Although it is not necessary to have one overall file maintenance system to update all affected files in one operation, some control centers have such a support program.

An interesting program is in service at General Public Utilities which allows updates of files on-line in the primary computer system [26]. This interactive program also updates application programs which are affected by the power system change or addition.

People or the In-House Project Team

The last, but certainly not the least, of the elements of system control center design which we shall discuss in this paper is people. The correctness of a control center design and its successful implementation obviously depends on the people assigned to the project. It is absolutely necessary that this group include an "in-house" project team assigned full-time to work on the project from the time the project is authorized by management up to time that the control center has been turned over to the system operators and has been rid of major problems.

The in-house team should have a combined background in power system engineering, digital hardware, computer application, and system operation. The team should have enough software specialists who can understand and develop real-time software. The team would be charged with the mission of becoming thoroughly familiar with the control system and of making sure that it carries out the functions according to specifications. Specifications, no matter how well written, cannot always be interpreted unequivocally. The individuals assigned by the vendor to translate the specifications into a detailed design may not necessarily have a full appreciation of what the utility really has in mind for certain functions. Or the vendor may feel certain it has understood an item in the specifications and on that assumption goes ahead with the design. Numerous problems arise in software design and implementation. Real-time software has to be broken down into small modules and each module is assigned to an individual programmer. Hence, very rarely does a programmer know the full context or significance to system operation of the module he is working on. The purpose of the in-house team is to review very thoroughly, check, and approve all system interfaces and all detailed design specifications. In addition to checking the correctness of the design with regard to the functions required, the in-house test would also check testing procedures, maintenance features, and the design adaptability to future needs.

Close collaboration between the system contractor and the in-house project team is essential during the software coding and debugging phase. Interaction between the vendor and the team must be encouraged at the programmer level and not just at the supervisory level. Such a rapport is readily achieved if the in-house team were doing some portion of the programming. I strongly recommend that the in-house team reserve for some of its members some of the programming effort. This not only gives valuable training but also gives the team members the ability to be of assistance to the vendor's programmers.

In organizing the in-house team, consideration must be given to the future work of system maintenance and enhancement. Specialists earmarked for key positions in this future function should be part of the team from its inception.

SUMMARY

What has been presented in this paper is a des-

cription of what a power system control center is intended to do, how it should perform, what it consists of, and how it should be designed.

As of this writing there are about 60 power system control centers in service or under development which have as a minimum, automatic generation control and security monitoring. About 20 installations in service have supervisory breaker control, less than 15 do security analysis. State estimation is in service in 8 control centers. For security analysis, the distribution factor method is used in the older installations. Newer centers with security analysis use or plan to use an on-line load flow method. The fast decoupled load flow is in use at one center. There are signs that this approach will be used in more future centers. Recent specifications for control centers now recognise a need for state estimation. Whereas in the early investigations of this function the approach favored meter placement to fit known algorithms, the trend is now to find the proper algorithm to fit the actual metering already installed or planned.

Currently two methods are in use in medium to large systems: the AEP or "lines-only" method and the basic weighted least squares. The Kalman filter or sequential approach is in use for a very small network and has not as yet been installed for a large system. Overall, the implementation of advanced application software is proceeding at a slow pace.

Innovations in the hardware area are taking place at a little faster rate. There is a trend in the increased use of minicomputers, new configurations with shared memory are appearing, and computer-communication networks are being studied. Although full-graphic color CRT's have been available there is still no significant movement towards their application.

A control center should be designed and built for system operator. Admittedly, a control center has prestige value as well. And a utility is entitled to exploit this aspect as liberally as possible. Still in the final analysis one should stand back from the impressive physical features and technical statistics of a control center and ask what actually does the control center do for operation.

If a control center is inadequate to the needs of operation or performs very poorly when it is most needed, the system operator can take no comfort in the fact that he is housed in an architectural showpiece or that redundant computers are gathering thousands of data every so many seconds, when, in fact, little useful information is being produced and it is taking forever to get a needed picture on the CRT screen.

It is not easy to design, build, and maintain a system control center that will do the job well. But it can be done. With a well thought-out design, hard work, smart work, an in-house project team, and a bit of luck, it can be done rather smoothly.

REFERENCES

[1] T. E. Dy Liacco, "System Control Center Design", Proceedings Engineering Foundation Conference on "Systems Engineering for Power: Status and Prospects", Henniker, N. H. August 1975.

[2] T. E. Dy Liacco, "The Emerging Concept of Security Control", Proceedings 1970 Symposium on Power Systems, Purdue University, Lafayette, May 1970.

[3] T. E. Dy Liacco, "The Adaptive Reliability Control System", IEEE Transactions, PAS-86, May 1967.

[4] T. E. DyLiacco, "Control of Power Systems via the Multi-Level Concept", Systems Research Center, Case Western Reserve University, SCR-68-19, Cleveland, 1968.

[5] T. E. DyLiacco, B. F. Wirtz, D. A. Wheeler, "Automation of the CEI System for Security", IEEE Transactions, Vol. PAS-91, May/June 1972.

[6] J. C. Russell, W. Rades, "The Wisconsin Electric Power Company Energy Control System," Power Industry Compyter Applications Conference, Minneapolis, June 4-6, 1973.

[7] B. Stott, O. Alsac, "Fast Decoupled Load Flow", IEEE Trans. Vol. PAS-92, May/June 1974.

[8] K. Kumai, K. Ode, "Power System Voltage Control by Using a Process Control Computer", IEEE Transactions, Vol. PAS-87, December 1968.

[9] W. A. Johnson, S. H. Bouchey, D. R. Penders, "The Potomac Electric Power Company's Consolidated Control Center", C74 366-1, IEEE Summer Power Meeting, Anaheim, California, July 1974.

[10] S. A. Arafeh, R. E. Kilmer, J. H. Rumbaugh, "Closed-Loop Computer Control of a System of Radial Load Busses, Using Transformers, Capacitors and Reactors", C75 032-8, IEEE Winter Meeting, New York, January 27-31, 1975.

[11] A. M. Sasson, S. T. Ehrmann, P. Lynch, L. S. Van Slyck, "Automatic Power System Network Topology Determination", IEEE Transactions, Vol. PAS-92, March/April 1973.

[12] T. E. DyLiacco, K. A. Ramarao, A. W. Weiner, "Network Status Analysis for Real-Time Systems", IEEE TP XII-A, Power Industry Computer Applications Conference, Minneapolis, June 4-6, 1973.

[13] F. C. Schweppe, E. J. Handschin, "Static State Estimation in Electric Power Systems", Proceedings of the IEEE, July 1974.

[14] W. S. Ku, R. Van Olinda, "Security and Voltage Applications of the Public Service Dispatch Center", Power Industry Computer Applications Conference, Denver, May 18-21, 1969.

[15] H.D. Limmer, "Techniques and Applications of Security Calculations Applied to Dispatching Computers", Proceedings, Power Systems Computations Conference, Rome, June 23-27, 1969.

[16] N. M. Paterson, W. F. Tinney, D. W. Bree, Jr., "Iterative Linear AC Power Flow Solution for Fast Approximate Outage Studies", IEEE Transactions, Vol. PAS-91, Sept./Oct. 197 .

[17] A. S. Debs, A. R. Benson, "Security Assessment of Power Systems", Proceedings Engineering Foundation Conference on "Systems Engineering for Power: Status and Prospects", Henniker, N. H. Aug. 1975.

[18] T. E. Dy Liacco, S. C. Savulescu, K. A. Ramarao "An On-Line Topological Equivalent of a Power System", IEEE paper submitted for presentation at 1977 Summer Power Meeting.

[19] P. Dimo, "Nodal Analysis of Power Systems", Abacus Press, Kent, England.

[20] J. F. Dopazo, M. H. Dwarakanth, A. M. Sasson, "An External System Equivalent Model Using Real-Time Measurements for System Security Evaluation", F76 370-1, IEEE Summer Power Meeting, Portland, 197 .

[21] T. Contaxis, A. S. Debs, "Identification of External System Equivalents in Steady-State Security Assessment", IEEE Paper to be presented at 1977

Summer Power Meeting.

[22] T. E. Dy Liacco, D. L. Rosa, "Designing Effective Man-Machine Interface for Power System Control", IEEE TP X-B, Power Industry Computer Applications Conference, New Orleans, June 2-4, 1975.

[23] T. E. Dy Liacco, "Design Elements of the Man-Machine Interface for Power System Monitoring and Control", Computerized Operation of Power Systems edited by S. C. Savulescu, Elsevier, 1976.

[24] J. W. Dillow, U. G. Knight, D. G. Taylor, "System Control in the Central Electricity Generating Board", IFAC World Congress, Boston, August 24-30, 1975.

[25] T. C. Cihlar, "New Control System with an Advanced Man/Machine Interface for Commonwealth Edison Company's System Security", IFAC World Congress, Boston, August 24-30, 1975.

[26] C. F. Doll, C. B. Woodward, M. H. Ewer, L. A. Schmidbaner, R. P. Whitesell, J. R. Evans, W. A. Johnson, "Multi-Company Electric Power System Control on the Pennsylvania-New Jersey-Maryland Interconnection", IFAC World Congress, Boston, August 24-30, 1975.

[27] A. J. Harris, J. W. Shelly, "Ontario Hydro's Data-Acquisition and Control System", IFAC World Congress, Boston, August 24-30, 1975.

[28] I. Gustafsson, J. Lindqvist, "The New Information System for the National Dispatching Centre at the Swedish State Power Board", IFAC World Congress, Boston, August 24-30, 1975.

Tomas E. Dy Liacco was born in Naga City, Phillippines on November 12, 1920. He obtained the BSEE (cum laude) and the BSME degree from the University of the Philippines, the MSEE degree from Illinois Institute of Technology, and the Ph.D. degree from Case Western Reserve University.

He is Principal Systems Engineer at The Cleveland Electric Illuminating Company as well as Adjunct Professor in Electrical Engineering at Case Western Reserve University.

Dr. Dy Liacco is Chairman of the IEEE Working Group on Power System Control Centers, Vice-Chairman of the EPRI Task Force on System Planning and Operation, and Executive Chairman of PICA-79. He has written numerous IEEE papers on power system control In 1975 he was commissioned by ERDA to prepare a state-of-the-art report on "System Control Center Design". He has given talks on control centers at universities and technical conferences throughout the world.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA X: EQUIPO DE COMPUTO E INTERFAZ HOMBRE-MAQUINA EN UN CENTRO DE CONTROL DE ENERGIA

DR. ALBERTO MAYER SASSON

ENERO, 1979.

POWER SYSTEM SECURITY

J.F. Dopazo

American Electric Power Service Corporation
2 Broadway
New York, New York

ABSTRACT

The object of this paper is to introduce the reader to basic concepts in the area of Power System Security from the operation viewpoint. Emphasis has been placed on concepts rather than derivation of equations which can be found in the publications included as references. The following topics are covered:

- Power System Monitoring
- Instrumentation
- Redundancy
- State Estimation
- Network Configuration
- Contingency Analysis
- Corrective Strategies

Terminology and simple concepts of Probability and Statistics are included throughout the text.

None of the topics mentioned can be completely covered in this single paper; therefore, the reader is encouraged to study the publications listed in the references.

INTRODUCTION

Power System Security begins at its planning stage. A poorly planned system cannot achieve a high level of security regardless of how well it is operated. On the other hand, because of economic considerations, the amount of built-in security has to be limited. Thus, regardless of the strength planned into a power system, its day to day operation must be such that maximum economy and security is attained within the limitations of the system. To accomplish this objective, the operation of power systems is coordinated from control centers.

In the early days, the coordination for economic operation was performed through telephone communications between system and plant operators. At the same time, for performing the security function, the system operator depended on telephone communications with plant and substation operators for monitoring the various facilities of the power system.

Generally, system operators had direct access to the values of frequency and voltage at their location. Knowledge of these variables permitted them to detect some system abnormalities. However, to identify the abnormal condition, the system operator had to rely on information from personnel at substations where the abnormal conditions could be monitored.

The relatively small amount of data at each substation could be analyzed by its operator. Thus, the information received by the system operator was sufficiently reliable to be synthesized for taking corrective action. The experience and good judgement of substation operators played an important role in power system security.

Also, in those days, the number of strong interconnections was relatively small. As a consequence, the interaction amongst the various portions of the network was not as pronounced as it is today.

As power systems expanded in size, voltage class, major concentration of generation and interconnections, their operation became more complex. Furthermore, the advent of remote telemetering, although a progressive step, triggered a trend toward unattended substations. As a consequence, the experience of substation operators became a missing link in the process, and the system operator was burdened with substantial amounts of raw data.

Under these conditions, in addition to synthesis of information, analysis of a large number of individual pieces of data was transferred to a higher level in the hierarchy of system operators.

During the early 1960's, process control digital computers became a valuable tool for system operations. However, the main emphasis at that time was placed on the economic operation function as well as in relieving the system operator from the work load associated with the preparation of operation reports.

In 1965, the northeast blackout created a major concern amongst electric utilities and government authorities. System security considerations became an issue of the highest priority and the electric industry focused on developing methods to ensure reliable operation of plant and transmission facilities.

Since the late 1960's and early 1970's major efforts have been placed in the following broad areas:

Power System Monitoring to improve knowledge of current system conditions

Contingency Analysis to determine the effects of outages of system facilities.

Corrective Strategies to provide the system operator with real-time guidelines for eliminating undesirable system conditions

Substantial progress has been made in Power System Monitoring. A number of utilities have new control centers in which some kind of improved monitoring system is operational. Also, for contingency analysis, a number of control centers have access to computer programs to simulate power system outages.

However, it is the author's opinion that, expecially in the areas of contingency analysis and corrective strategies, additional work is necessary to satisfy the real-time requirements of power systems security.

POWER SYSTEM MONITORING

Power System Monitoring has always been a basic function for ensuring a secure system.

The development of modern Data Acquisition Systems coupled with communication networks and digital computers, permitted the automatic collection of large amounts of real-time data to be displayed at central locations. The ability of transmiting any amount and type of data to a central location raised the question of which quantities should be measured.

Representatives of some utilities favored the philosophy of measuring those quantities which permit monitoring of some key facilities of the power system. Others have followed the route of gathering the necessary data to perform conventional Load Flow calculations in real-time for monitoring all system components. A third group of utilities have used state estimation techniques in their monitoring schemes, to take into account the reliability of the measuring system. To the author's knowledge, this approach has been followed by five or six power systems of various sizes around the world. However, others are showing interest in the approach. The AEP state estimator became operational in February 1975 and has proved to be a valuable tool for security of system operation.

The philosophy of monitoring key facilities, in the author's opinion, yields an incomplete monitoring system and appears to be conflicting with good planning practices. It is reasonable to expect that any facility of a power system may become a key one depending on the existing operating conditions at a particular time. On the other hand, if under normal conditions certain facilities can be selected as needing to be monitored more closely than others of the same category, this situation should be remedied by better planning.

The approach of performing a conventional Load Flow in real-time permits the System Operator to have access to measured quantities as well as to other quantities of interest which can be obtained from calculations. This approach evolved from the experience of using the Load Flow in power system simulation for planning purposes. However, for the real-time function, it has several limitations:

(a) In the conventional Load Flow, the input variables are restricted to the complex powers at load buses and the real power and scheduled voltage at regulated buses. In real-time however, this constraint presents a serious limitation because conveniently available measurements of other variables cannot be used.

(b) Since the Load Flow formulation consists of a set of independent equations, a colution cannot be obtained if one of the input variables becomes unavailable.

(c) If a piece of input data is incorrect, the results might be rendered useless.

(d) It is not possible to determine the level of confidence that can be placed on displayed quantities. This severely impacts the System Operator in his decision making function.

The justification for using state estimation techniques has been based on the facts that a certain amount of error is inherent in any measurement acheme and that individual measurements can become grossly incorrect or missing.

State Estimation provides the ability for coping with measurement error, detects and identifies incorrect or missing data and ensures the validity of the information displayed to the system operator, including quantities whose measurements have been missing or identified to be incorrect.

INSTRUMENTATION

Adequate instrumentation is a basic requirement for monitoring any physical process. An instrumentation system, in order to be adequate, must permit extracting the proper amount and quality of information from the process such that the monitoring system is trustworthy. It would be a paradox if the monitoring of a process were less reliable than the process itself.

The main quantities of interest in assessing the overall performance of a power system are: complex voltages at the network buses and complex powers and current flows in the various facilities of the system.

In modern instrumentation systems, there are a number of devices which collect information at a remote location and transmit it to a control center. These devices include: instrument transformers, sensors and analog to digital converters.

Any one of these devices can fail and none of them is perfect, i.e. a certain amount of error is inherent in its performance.

The amount of error in a particular device is, in general, unknown. However, from experiments in a controlled environment, the manufacturer can provide information on the statistical behaviour of a device that belongs to a specific precision category.

These experiments consist of taking a sample from the entire population of the devices in question and, with a fixed input, the output of each device is observed. In the process of repeating the experiment, the outcome varies from trial to trial in a random fashion. Thus, the quantity of interest is said to be a RANDOM VARIABLE.

A plot of the various outputs vs the number of devices associated with each output is likely to be as shown in Fig. 1.

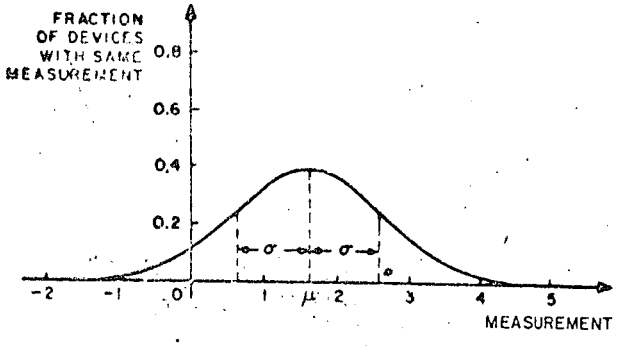


FIGURE I
AN ACCURACY TEST ON A SAMPLE OF DEVICES

This plot is called a GAUSSIAN or NORMAL probability density function. The value μ is the MEAN or EXPECTED VALUE of the various outcomes and provides an index of the tendency of most of the outputs being close to a certain "average" number. An indication of the spread of the various outputs is provided by the VARIANCE, which is denoted by σ^2 and is approximately equal to the average of the sum of the squares of the deviations of the various outputs with respect to μ, i.e.

$$\sigma^2 \approx \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_i)^2$$

The square root of the VARIANCE σ^2 is the STANDARD DEVIATION σ. The significance of σ is that in a GAUSSIAN probability density function the following conditions are satisfied:

68% of the outputs fall within $\mu \pm \sigma$
95% of the outputs fall within $\mu \pm 2\sigma$
99% of the outputs fall within $\mu \pm 3\sigma$

The STANDARD DEVIATION then provides valuable information on the precision of the various devices in the instrumentation system. It permits not only estimating the amount of error which is likely to be associated with any individual measurement but, what is more important, it allows discriminating incorrect measurements from acceptable ones. Also, it permits mixing measurements taken with devices of different precisions. This is done by placing more weight in the information content of higher precision devices.

REDUNDANCY

It has been mentioned that the error ontent of an individual measurement is unknown. Thus, the true value of a quantity of interest cannot be obtained. Then, in order to discriminate between correct and incorrect measurement as well as to improve the accuracy in the values of the measured quantities (error filtering effect) if it is required, REDUNDANCY is necessary, i.e. more than one measurement of the quantity of interest must be taken.

The amount and type of redundancy requirements depend on factors such as: quality of the measurement system, requirements for detection and identification of incorrect measurements, feasibility of modeling the process in some mathematical fashion, efficiency of solution techniques, cost, etc.

In summary, adequacy of redundancy cannot be determined from specific rules, but rather from a good understanding of the process and the requirements of the monitoring system.

To illustrate the train of thought in the approach to the redundancy problem, lets consider the simple case where the temperature of a particular process must be monitored. Obviously, a good quality thermometer could be used. However, if this temperature happens to be a critical quantity, a single thermometer might not be sufficient for ensuring the reliability of the monitoring system.

A second thermometer, for instance, would provide the means of detecting malfunctioning of the monitoring system when the two readings differ by an amount larger than the precision of the instruments in question. Although this two to one redundancy ratio permits detecting an incorrect measurement, it is still inadequate for identifying which one of the two measurements is incorrect. Thus, for this purpose, a third piece of information is needed. This additional redundancy can be obtained from a third termperature reading or from a pressure already available if a mathematical model can be formulated to relate the two variables. Then, the monitoring system of this process requires a minimum redundancy ratio of three to one to be adequate.

In a power system, a mathematical model can be formulated and, in general, duplicate measurements are not taken. The model together with STATE ESTIMATION techniques relates most of the quantities of interest. In the process, each measurement contributes to the estimation of more than one quantity and each quantity is estimated from more than one measurement.

STATE ESTIMATION

State Estimation techniques provide the means of processing a set of redundant information to obtain an ESTIMATE of the STATE VARIABLES of the system. Once the STATE VARIABLES are determined, other quantities of interest can be obtained. A fundamental property of the solution process is that it determines the averages or MEAN values of the quantities of interest. Thus, the calculated values, in general, do not match any one of the measurements. Instead, the solution is reached by a BEST FIT of the entire set of input data.

The procedure consists of minimizing a function of the STATE VARIABLES. One possible objective function is the sum of the squares of the deviations between measured and calculated values. To take into account the accuracy of each measurement, each of the terms in the summation is weighted in inverse proportion to the VARIANCE of the associated measurement. Thus, the solution criterion is the weighted least squares i.e.

$$\text{Minimize } J(x) = \sum_{i=1}^{N} \frac{[z_i - f_i(x)]^2}{\sigma^2}$$

where x is the set of state variables
z is the set of measured quantities
$f_i(x)$ is the functional relationship of the ith measured variable with respect to the state variables
N is the number of measurements

The difference between the number of state variables and the number of measurements is the redundancy and is given the name DEGREES OF FREEDOM.

At the solution point, the calculated values of the measure variables are approximately the MEAN values of these variables.

Inspection of the objective function J(x) reveals that, at the solution, each of its terms is approximately equal to

$$\left(\frac{z_i - u_i}{\sigma_i}\right)^2$$

It was mentioned that the measured value Z_1 has the GAUSSIAN or NORMAL probability density function shown in Fig. 1. If the MEAN value is subtracted from Z_1 and the result is divided by the STANDARD DEVIATION σ_1, a new probability density function is obtained which has a MEAN value equal to zero and a VARIANCE equal to one. The new curve is a UNIT NORMAL. Thus, the objective function becomes a sum of squares of UNIT NORMALS and is denoted as a CHI-SQUARE distribution. A CHI-SQUARE distribution has the property that its expected value is equal to the number of degrees of freedom. Fig. 2 shows a family of CHI-SQUARE distributions.

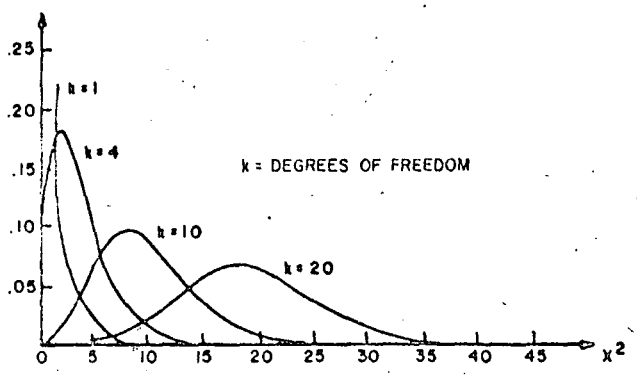


FIGURE 2. CHI-SQUARED DISTRIBUTION

The chi-square distribution plays an important role in the detection and identification of bad measurements, that is measurements that grossly differ from the true value by many σ values of the device used. Consider the case where one such bad measurement exists. That measurement by definition does not belong to the 3σ range of the normal distribution of the device.

Knowing the degrees of freedome of the particular problems at hand, a chi-square distribution is fixed. The value of the chi-square (the horizontal axis on Fig. 2) that includes 99% of the area below the curve is theoretically the maximum value or boundary attainable by the weighted sum of squares if all measurements were within 3 of their particular standard variation.

The presence of a bad measurement will produce calculated values that for some terms will differ from the measured value by more than their individual 3σ; thus making the sum of squares exceed the boundary for a 99% confidence. Notice that it has not been said that the sum of squares term corresponding to the bad measurement is necessarily large. Indeed, it may be that the calculated and measured values at the bad measurement location are close. However, this is at the expense of other terms becoming large. It is this characteristic that makes the chi-squared bad data detection test so powerful.

Having detected the presence of bad data, the problem remains of identifying which is the bad measurement. It has been mentioned that the expected value of a chi-squared distribution is equal to its degrees

of freedom. Consider that the calculated value of the sum of squares is taken as an approximation to the expected value. Then the ratio of the sum of squares to the degrees of freedom is a measure of how much the sum of squares differed from the expected value.

A ratio somewhat near to one implies that the individual standard deviations used with each measurement corresponded to the normal range of its measured minus calculated values. A large ratio suggests that at least for one measurement this correspondence did not occur. If all measurement variances are increased by this ratio, then correspondence will have been forced to exist.

The term measured minus calculated value is called a measurement residual. That residual is itself also a random variable. If different complete sets of measurement devices were used to gather sets of measurements, and each set was used to solve the redundant equations by minimizing the weighted sum of squares, a set of residuals corresponding to each measurement would be obtained. Each of these residuals is a normally distributed random variable with an expected value of zero. The standard deviation of this distribution can be computed from the theoretical considerations without the need of obtaining sets of measurements. This information can be extracted from the redundancy present in the measurement system. The following transformation from normal to unit normal can be made:

$$\frac{\text{Measurement Residual - Zero}}{\text{Standard Deviation of Measurement Residual}}$$

The zero is the mean value of the residual. In the same way that measurement standard deviations were all increased by the ratio obtained from the chi-squared distribution, the measurement residual standard deviation also becomes increased by this same factor. Theoretically the effect of the factor is to increase the standard deviation of measurement residuals such that the residuals of good measurements are enclosed within the 3σ range around the mean of zero. If the effect of the factor is omitted and the results of the above transformation ranked from large to small values, the residual of the bad measurement would tend to be at the top of the list. Intuitively this can be inferred considering that the effect of a bad measurement is so strong, that the normal possible range of all the other good measurements within their individual 3σ will result in a smaller residual standard deviation for the bad measurment than for the good ones. As in the above transformation the standard deviation appears in the denominator, it will push the bad measurement term towards the top of the list, even if its actual residual was small. With the inclusion of the factor, the whole list is scaled down such that in the unit normal transformation the good measurement terms will tend to be smaller than 3σ with a 99% confidence. In practice, it happens that the factor cannot be computed exactly from the chi-squared distribution as only one point of that distribution is available and the approximation was made to consider that point to be the expected value. The error in the factor makes the transformation inexact and instead of transforming into a unit normal, it can be shown that it transforms into a very similar distribution called the Student-t distribution. In fact, for degrees of freedom larger than about 30, the unit normal and the Student-t are practically identical. The effect of the Student-t is that the 99% confidence instead of being at a level of 3σ is at a number somewhat smaller that 3σ. The Student-t test to determine which of the transformation terms is the largest of those above the 99% confidence level, is called the bad data identification test. The importance of the factor is that it makes the bad measurement identification more clear cut. Occasionally it may happen that several measurements are closely beyond their 3 levels such that the chi-square detection test flags the existence of bad data. The Student-t identification test finds that the largest term in its ranked list is smaller than the 99% confidence level and concludes that no clear cut bad data exists.

It was mentioned that the standard deviation of the residual distribution can be computed from theoretical considerations. Similarly, the standard deviation of the state variables, the voltages, and that of the flows themselves can also be computed. Assuming that the calculated value is an approximation to the mean, that value plus or minus 3σ gives a range in which the true value lies with a 99% confidence, assuming that the voltage and flow distributions are normal. These 3σ ranges are called confidence limits. A small confidence limit implies high accuracy.

NETWORK CONFIGURATION

An important aspect in power system security is the knowledge of the status of the various system facilities.

These facilities are interconnected by means of circuit breakers which may operate at any time. Thus, monitoring of circuit breaker status in real-time is essential to determine the present operating condition of the system as well as for the analysis of system contingencies.

There are three basic effects of circuit breaker operations:

1) Circuits may or may not be disconnected.
2) Substations may include one or more electrical nodes. Thus, the total number of system nodes can be variable.
3) The power system may split into two or more separated areas.

Also, knowledge of the status of the various measurements is required for state estimation purposes because the number of equations in the system model depends on the number of measurements available.

Therefore, in any monitoring system, a Network Configurator should be included. The function of the Network Configurator is to analyze the status of circuit breakers as well as measurements and to automatically determine the current model of the power system.

A possible Network Configurator is the one developed at AEP which, although conceptually simple, is general in the sense that it can handle any breaker scheme at the various substations.

The logic to determine which facilities are connected at a substation is the same as that to determine if the power system has split into two or more areas.

An important complement of the configurator logic is a Data Base in which every circuit breaker is defined in terms of substation and facilities connected at its two terminals. It also includes identification of measurements with associated facilities. This Data Base is maintained current by operations personnel in accordance with any structural changes or additions at the various substations.

The Network Configurator interfaces with the Teleprocessing System, the State Estimator and the Display System.

The Teleprocessing System continuously scans the status of circuit breakers and in the event of a change, the configurator is informed. The Configurator logic then traces the paths provided by closed breakers at the various substations where changes have occurred. Lists associated with each closed path then include all those facilities which are incidental to common nodes.

Once the status of the network facilities is determined, the same logic is used in tracing the paths provided by facilities energized at their terminals. Appearance of more than one list of interconnected facilities means two or more separated areas.

At this point, the configurator can pass to the Estimator an updated version of the network model.

When faulty measurements are detected and identified, the configurator function is to remove the measurements in questions, analyze the effect on the model and return the proper model to the Estimator.

Finally, the configurator, through his interface with the Display System, drives the proper indications and alarms.

CONTINGENCY ANALYSIS

The analysis of system contingencies has been a function performed for system planning purposes since the days of the Network Analyzer. This function consists of simulating outages of generating units and transmission facilities to study their effect on bus voltages, power flows, and the transient stability of the power system as a whole.

With the advent of digital computers, programs such as the Load Flow and Transient Stability were developed in the late 1950's and early 1960's to perform the contingency analysis function.

Substantial progress has been made in improving the quality of these programs in terms of their calculation speed as well as their ability to simulate more precisely the various system components. For instance, the early Load Flow programs using the Bus Admittance Matrix for its formulation and Gauss Seidel as the numerical solution, evolved through the Newton Raphson solution, exploiting the sparsity of the Jacobian matrix and up to a decoupled technique. In this technique, the Jacobian matrix is assumed constant and the variables of the problem become decoupled i.e. real powers are related only to the voltage angles and reactive powers to voltage magnitute.

This substantially improved the calculation speed of the Load Flow program. In the Transient Stability area improved models for synchronous machines, governors, excitation systems, etc. have been developed.

For operation security, most control centers have access to these programs in an off-line mode and some are adapting them to the real-time environment. In this case, the System Operator has been provided with means such as keyboards, lightpens and CRT's to simplify the input data requirements and the selection of the contingency cases to be studied.

Emphasis has been placed in adapting the Load Flow program for the analysis of steady state contingencies. Also, various techniques have been developed in which distribution factors are used to screen, from all probable outages, those outages which are likely to result in unacceptable loading conditions. These outages then, are studied more rigorously using a conventional A.C. Load Flow program.

There are some fundamental differences between the contingency analysis to be performed at a control center and that for planning purposes:

1) Current System Conditions

In real-time, the current state of the power system must be known in order to predict a future state resulting from the outage of one of the system facilities. Also, the selection of the contingency cases to be studied is strongly governed by current operating conditions. State Estimation and System Configuration programs provide the starting point on which contingency analysis can be based.

2) Contingency Selection

In planning studies, outages of system facilities are simulated in accordance with a single, double or higher order contingency criterion. In real-time, depending on current conditions, which may already include one or more outages, a subsequent single

contingency case might correspond to a higher order one in planning studies.

Planning studies focus on outages of generation and transmission facilities. In real-time, the status of circuit breakers at the various substations is important for contingency selection. For instance, due to construction, maintenance, temporary arrangements, etc., the breaker status at a substation might be such that the operation of another breaker, which normally results in the outage of a single facility, might this time result in several facilities becoming out of service.

Proper selection of contingency cases is a difficult problem. Pattern recognition techniques appear to be promising in this area.

3) External System

External System is defined as that portion of an interconnected network which is outside the reach of the direct monitoring system of the network of interest.

The operating condition as well as the reaction of the external system, affects the results of contingency analysis of the monitored portion.

This basic problem of insufficient information could be solved by real-time interchange of data amongst the utilities of the interconnected network and/or extending the range of monitoring systems. However, the interchange of real-time data is not a simple problem, and various procedures to cope with this fact either are in use or under investigation.

(a) Conventional Network Reduction

This technique has been used in system planning studies where a portion of a network is replaced by its equivalent for reducing the problem size. However, it does not take into account the varying operating conditions of the replaced portion of the network.

(b) Network Identification

Several papers have reported theoretical attempts to identify external networks solely from measurements taken in the internal system. However, no approach has been announced that has been shown to work under all operating conditions of the external system.

(c) Stochastic Equivalent

In this approach, conventional reduction techniques are complemented by internal measurements in an attempt to compensate for the assumptions made in conventional reduction techniques. A procedure that appears promising is as follows:

The portion of the external system consisting of all lines and modes directly connected to internal boundary buses, is retained as part of the internal system. The remaining portion of the external system is then replaced by an equivalent using conventional reduction techniques and assumptions on loads, generation and configuration.

State Estimation is then extended to the new boundary buses.

Equivalent injections at the new boundary buses are computed from the algetraic sum of the total flow into these nodes from the internal system and the flows in the equivalent lines.

The equivalent injections are tested using past history to verify their validity. These tests also detect if unreported configuration and/or injection changes have occurred in the external system.

The above procedure is being tested at AEP at the time of this writing. So far, results have been encouraging.

4) Stochastic Load Flow

Once the external equivalent is available the problem remains of solving the network equations with the simulated contingency case. However, to account for the uncertainties in the input data it appears that the Stochastic Load Flow can be suitable. The Stochastic Load Flow is a conventional Load Flow with a post-processor in which the variances of the input quantities are considered to determine the variances of the results.

CORRECTIVE STRATEGIES

The monitoring system keeps track of current steady state operating conditions whereas results from the contingency analysis provides conditions which are likely to occur in the event of system outages. Either one of these conditions may violate some specified limits; therefore, the system operator should be provided with corrective strategies to bring the system back to normal.

The principal means available to the operator to take corrective action are: power generation schedule, switching of reactive sources, transformer taps, voltage schedules and power interchange with interconnected networks.

At present, corrective strategies are based on operator experience and guidelines from off-line studies. With increased complexity of power systems, it appears that this procedure may become inadequate.

Since the 1960's, analytical techniques have been developed in the areas of optimal load flows. These techniques, in contrast with conventional load flows, permit handling additional constraints such as: allowed range of voltage levels as well as capabilities of generation and transmission facilities. Therefore, such techniques appear suitable to the problem of determining corrective strategies.

Various methods have been proposed; however, to the author's knowledge, none has been used in a real-time environment at a control center.

The purpose of these methods is to optimize some system objective function, such as production cost, losses, etc. subject to physical limiting constraints on facilities and the observation of the network laws.

Following are brief description of three proposed methods:

1) Reduced Gradient

In this method, a Lagrangian approach is followed, i.e.:

> The objective function is augmented by the constraints multiplied by their associated lagrangian multipliers.
>
> The partial derivatives of the augmented function with respect to the control variables is forced to be zero at the solution point.

For the process, the control variables are corrected in the direction along the reduced gradient. The reduced gradient is the gradient of the augmented function computed at the point where the dependent variables force quality constraints to be met. In other words, the direction of travel is along the equality constraints and the amount of correction is determined by adjusting the step size. This adjustment is the most critical part of the process.

The following simple example permits a geometrical interpretation of the process.

Consider the economic dispatch of two generating units, one of which operates at a limit.

The objective function to be minimized is

$$f = C_1(P_1)+C_2(P_2)$$

subject to the constraint $h = (P_o-P_1-P_2)=0$

where P_o is the total generation required.

The augmented function is

$$F = f+\lambda h$$

Taking the partial derivatives,

$$\vec{\nabla}F = \vec{\nabla}f+\lambda\vec{\nabla}h$$

where $\vec{\nabla}$ denotes gradient vector whose entries are the partial derivatives of the associated function with respect to the control variables.

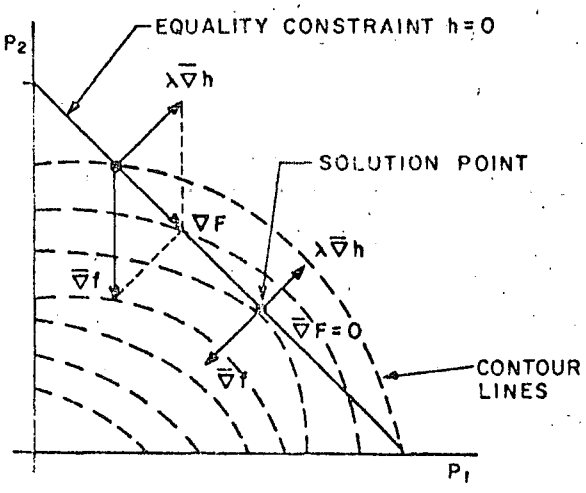


FIG. 3
GEOMETRICAL INTERPRETATION OF LAGRANGIAN APPROACH

2) Hessian Matrix

> The objective function is augmented by the square of each constraint equation multiplied by a weighting factor.
>
> The resulting augmented function is expanded into a Taylor series up to the first order term.
>
> The partial derivatives of the expanded function with respect to the state variables are equated to zero. This results in an equation containing a matrix of second partial derivatives, the Hessian Matrix.
>
> This last equation is iteratively solved to successively correct the state variables until a minimum of the augmented function is found. During this iterative process, the weighting factors of those constraint equations that persist in remaining outside limits is increased to force the process to concentrate on these constraints.
>
> Similarly to the case of the Reduced Gradient in which the adjustment of the step size was critical, the adjustment of the weighting factors is the most critical part of the Hessian Matrix approach.

3) Hybrid Approach

This method is currently being developed at AEP in an attempt to cope with the problembs of step size and weighting factors. The method uses a sensitivity relationship to predict the effect on the constraints when the control variables are corrected. This mechanism permits adjusting the step size and weighting factors in a stable

manner. This method is reported in a paper presented at PICA 1977.

CONCLUSIONS

Proper planning is an essential prerequisite for security of system operation, but planning studies cannot cover all the conditions which will occur in real-time.

In the real-time environment, the system operator has to face the fact that the full capabilities of the system are seldom available. Nevertheless he has to do his best in optimizing the secure and economic operation of the system. A basic requirement to perform this function is a trustworthy monitoring system.

The complexity of present power systems requires that means for real-time contingency analysis be available at the control center.

Techniques for contingency analysis should address the problem at the substation level. This is in contrast with the approach followed in system planning in which outages of generation and transmission facilities are studied. A major problem in this area is the proper inclusion of the effect, upon the network of interest, of those portions of the network which are outside the reach of the monitoring system.

In the area of corrective strategies, additional efforts should be made for applying optimal load flow techniques in the real-time environment. These techniques are suitable for arriving at real-time guidelines for eliminating undesirable conditions such as overloading of system facilities and violations of scheduled voltage levels.

Finally, although this paper deals with the subject of power system security, the problem should be approached as a part of the overall functions of a control center. Figure 4 shows a diagram which attempts to indicate these functions and their interactions.

REFERENCES

1. G.W. Stagg, J.F. Dopazo, M. Watson. J.M. Crawley, G.R. Bailey and E.F. Alderink, "A Time-Sharing On-Line Control System for Economic Operation of a Power System," Pro. Instrum. Soc. Am., Oct., 1966.

2. F.C. Schweppe and J. Wildes, "Power system static-state estimation - Part I: Exact model," IEEE Trans. PAS-89, pp. 120-125, 1970.

3. F.C. Schweppe and D.B. Rom, "Power system static-state estimation - Part II: Approximate model," IEEE Trans. PAS-89, pp. 125-130, 1970.

4. F.C. Schweppe, "Power system static-state estimation - Part III: Implementation," IEEE Trans. PAS-89, pp. 130-135, 1970.

5. J.F. Dopazo, O.A. Klitin, G.W. Stagg and L.S. VanSlyck, "State Calculation of Power Systems from Line Flow Measurements," IEEE PAS-89, pp. 1698-1708, September/October 1970.

6. J.F. Dopazo, O.A. Klitin and L.S.VanSlyck, "State Calculation of Power Systems from Line Flow Measurement, Part II," IEEE-PAS-91, pp. 145-151, January/February, 1972.

7. J.F. Dopazo, O.A. Klitin, A.M. Sasson, L.S. VanSlyck, "Real-Time Load Flow for the AEP System," Paper No. 3.3/8, 4th Power Systems Computation Conference Proceedings, Grenoble, France, September, 1972.

8. B. Poretta and R. S. Dhillon, "Performance evaluation of state estimation from line flow measurements on Ontario Hydro power system," IEEE Winter Power Meeting, T 73 086-6, 1973.

9. J. F. Dopazo, S. T. Ehrmann, O. A. Klitin and A. M. Sasson, "Justification of the AEP Real-Time Load Flow Project," IEEE PAS-92, pp. 1501-1509, September/October, 1973.

10. A. M. Sasson, S. T. Ehrmann, P. Lynch and L.S. VanSlyck, "Automatic Power System Network Topology Determination," IEEE PAS-92, pp. 610-618, March/April, 1973.

11. L.S. VanSlyck and J.F. Dopazo, "Conventional Load Flow Not Suited for Real-Time Power System Monitoring," Proceedings of the 8th IEEE PICA Conference, June, 1973.

12. J. F. Dopazo, O. A. Klitin and A. M. Sasson, "State Estimation of Power Systems: Detection and Identification of Gross Measurement Errors." Proc. IEEE PICA Conf. 1973. Discussions and closures, IEEE PAS-93, Jan-Feb 1974 pp.20.

13. J.F. Dopazo and A.M. Sasson, "AEP Real-Time Monitoring Computer System," Symposium on Implementation of Real-Time Power System Control by Digital Computer, Imperial College of Science and Technology, London, September, 1973.

14. J.F. Dopazo and A.M. Sasson, "Least Square Estimation Applied to Power Systems," Seminar on Computer Methods of Power System Analysis and Control, Bucharest, Romania, June, 1974.

15. K.A. Clements, O.J. Dennison and K.L. Ringlee, "Effects of Measurement Non-Simultaneity, Bias and Parameter Uncertainty on Power System State Estimation,"1973 PICA Proc. IEEE, Abstract pp.5 PAS-93, Jan-Feb 1974.

16. F. Aboytes and B.J. Cory, "Identification of Measurement, Parameter and Configuration Errors in Static State Estimation," Proceedings of the 9th IEEE PICA Conference, June, 1975.

17. J.F. Dopazo, O.A. Klitin and A.M. Sasson, "Stochastic Load Flows," IEEE PAS-94, pp. 299-309, March/April 1975.

18. J.F. Dopazo, S.T. Ehrmann, O.A. Klitin, A.M. Sasson and L.S. VanSlyck, "Implementation of the AEP Real-Time Monitoring System," IEEE Winter Power Meeting, 1976.

19. J.F. Dopazo, S.T. Ehrmann, A.F. Gabrielle, A.M. Sasson and L.S. VanSlyck, "The AEP Real-Time Monitoring and Control Computer System," IEEE Winter Power Meeting, 1976.

20. W.L. Snyder Jr. and A.M. Sasson, "Security Load Flow Employing a Modified Decoupled Hessian Technique," paper submitted for the 1977 PICA Conference, Toronto, May, 1977.

21. M. Flam and A.M. Sasson, "Stochastic Load Flow - Decoupled Implementation," paper to be presented at the 1977 Summer Power Conference, Mexico, July, 1977.

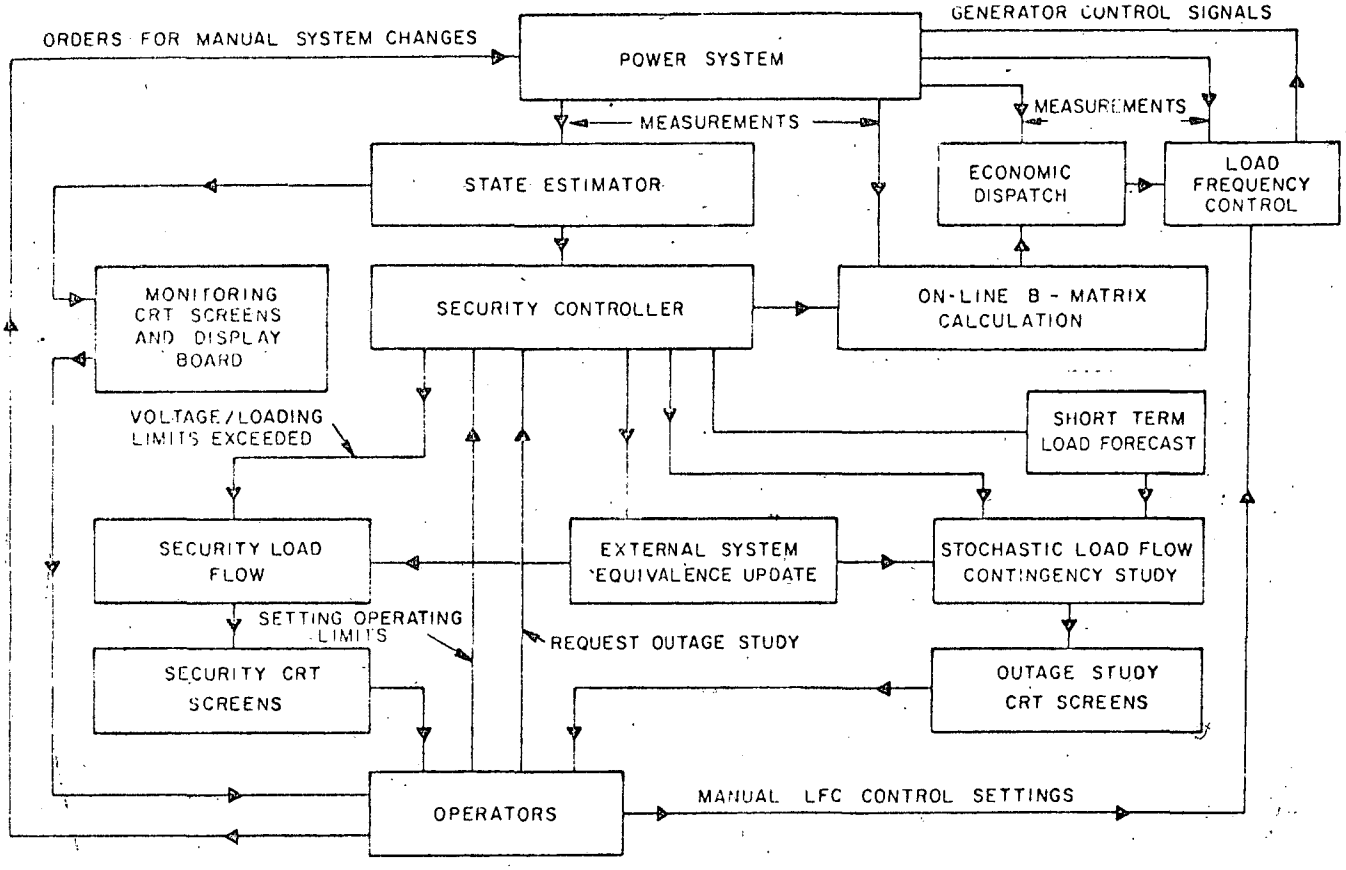


FIG. 4. FUNCTIONAL FLOW CHART OF THE EHV SYSTEM OPERATION SECURITY PROJECT

Jorge F. Dopazo is currently Head, Engineering and Management Systems/Computer Applications, American Electric Power Service Corporation. He is responsible for directing the research, development and implementation of major computer applications in the areas of engineering and management science for the AEP System.

In 1950 he received the Diploma of Electrical Engineer from the University of Havana, Cuba; graduated from the Advanced Power Systems Engineering courde at General Electric Co., Schenectady, 1957-58 and did additional graduate work at the Polytechnic Institute of Brooklyn, 1970-71.

After graduation he became an engineer with the Cuban Electric Co. responsible for starting and testing all electrical equipment in new power plants and transmission substations. From 1953 to 1955 he worked on system operation and system planning problems. In 1955 he was promoted to Head of System Operations, and during the period 1958-1961 he was Head of System Operations and System Planning.

In 1961 he joined the American Electric Power Service Corp. as an engineer responsible for the development of analytical methods and computer programs for the solution of power systems problems. In 1968 he was promoted to Head of Control Computer Systems where he was responsible for directing the research, development and implementation of analytical methods and programming systems for the application of control computers to power system operation. A major project in this area was the implementation of the first computer system in the country which used state estimation techniques for real-time monitoring of a power system. In 1974 he became Head, Engineering Systems where, in addition to control computers, his responsibilities expanded to all engineering computer applications. In 1976 he was promoted to his present position.

Mr. Dopazo has published some thirty technical papers in the areas of network analysis, transmission losses, optimization techniques and real-time monitoring of power systems. He has lectured on Computer Methods for Power System Analysis and Real-Time Control and Monitoring of Power Systems at a number of universities in the USA and abroad. He has also lectured on these subjects at various IEEE educational meetings.

Mr. Dopazo has been a member of IEEE technical committees and is a member of the Working Group on Control Centers of the Computer Applications Subcommittee, Power System Engineering Committee, Power Engineering Society. He is a member of Sigma Chi and a Fellow of IEEE.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA XI Y XII

NOTAS DE: CONCEPTOS BASICOS DE OPTIMIZACION Y PROGRAMACION NO-LINEAL

M. EN I. RAFAEL CRISTERNA OCAMPO

ENERO, 1979.

N O T A S D E P R O G R A M A C I O N N O - L I N E A L

RAFAEL CRISTERNA OCAMPO
DIVISION DE ESTUDIOS SUPERIORES
FACULTAD DE INGENIERIA
UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

ESTAS NOTAS PRETENDEN INTRODUCIR EN FORMA BREVE ALGUNOS CONCEPTOS DE OPTIMIZACION DE PROBLEMAS NOLINEALES.

EL PROBLEMA DE PROGRAMACION NOLINEAL CONSISTE EN SELECCIONAR LOS VALORES DE CIERTAS VARIABLES, GENERALMENTE NO NEGATIVAS, DE MANERA QUE SE MAXIMICE O MINIMICE UNA FUNCION DADA, SUJETA A UN CONJUNTO DE RESTRICCIONES DE IGUALDAD Y/O DE DESIGUALDAD.

EN GENERAL, LA FUNCION POR OPTIMIZAR PUEDE REPRESENTAR UN BENEFICIO O UN COSTO DEBIDO AL DESARROLLO DE ALGUNA(S) ACTIVIDAD(ES) MEDIANTE LO CUAL SE BUSCA CUMPLIR ALGUN OBJETIVO O ALCANZAR ALGUNA META O DEFINIR EL PROCESO DE OPERACION DE ALGUN SISTEMA, ETC.

ANALITICAMENTE, EL PROBLEMA DE PROGRAMACION NO LINEAL PUEDE EXPRESARSE COMO:

$$\min_{z} F(z)$$

$$\text{sujeto a: } g(z) = b \qquad (1)$$

$$\text{generalmente: } h(z) \leq c$$

$$z \geq 0$$

ESCRITO EN FORMA DETALLADA:

$$\min_{z_1,\ldots,z_n} F(z_1, z_2, \ldots\ldots, z_n)$$

$$\text{sujeto a:} \quad \begin{aligned} g_1(z_1,\ldots\ldots z_n) &= b_1 \\ g_2(z_1,\ldots\ldots z_n) &= b_2 \\ &\cdots\cdots\cdots \\ g_m(z_1,\ldots\ldots z_n) &= b_m \end{aligned} \qquad \begin{aligned} h_1(z_1,\ldots\ldots,z_n) &\leq c_1 \\ h_2(z_1,\ldots\ldots z_n) &\leq c_2 \\ &\cdots\cdots\cdots \\ h_p(z_1,\ldots\ldots,z_n) &\leq c_p \end{aligned} \qquad (1')$$

GENERALMENTE: $z_1 \geq 0,\ z_2 \geq 0, \ldots\ldots\ldots z_n \geq 0$

LAS n VARIABLES $z_1, z_2, \ldots\ldots z_n$ SE DEFINEN COMO LAS n COMPONENTES REALES DEL VECTOR COLUMNA Z. LA FUNCION OBJETIVO F(Z) REPRESENTA AL CRITERIO PARA EL CUAL SE REQUIERE ENCONTRAR SU VALOR MINIMO. LAS FUNCIONES DE RESTRICCION $g_1(z), g_2(z), \ldots g_m(z)$ SON REPRESENTADAS POR ELVECTOR COLUMNA $g(z)$ ASI COMO LAS $h_1(z), h_2(z), \ldots h_p(z)$ INVOLUCRADAS EN LAS RESTRICCIONES DE DESIGUALDAD SE REPRESENTAN POR EL VECTOR COLUMNA $h(z)$. LAS CONSTANTES $b_1, b_2, \ldots\ldots b_m$ Y $c_1, c_2, \ldots\ldots c_p$ SE DENOMINAN LAS CONSTANTES O TERMINOS DERECHOS DE LAS RESTRICCIONES Y SE REPRESENTAN POR LOS VECTORES COLUMNA b Y c RESPECTIVAMENTE.

SE ACEPTA QUE LAS $m+p+1$ FUNCIONES, $F(Z), g_1(z), \ldots g_m(z), h_1(z), \ldots h_p(z)$ SON FUNCIONES DADAS, CONTINUAMENTE DIFERENCIABLES, QUE NO CONTIENEN ELEMENTOS ALEATORIOS, b Y c ESTAN FORMADOS POR NUMEROS REALES Y QUE Z PUEDE SER CUALQUIER VECTOR CON n COMPONENTES REALES, SUJETO UNICAMENTE A LAS $m+p+n$ RESTRICCIONES DEFINIDAS EN (1). n, m Y p SON FINITAS.

ES COMUN DESIGNAR A LAS VARIABLES Z COMO LOS "INSTRUMENTOS" DEL PROCESO QUE SE REQUIERE OPTIMIZAR. LAS RESTRICCIONES DE DESIGUALDAD REPRESENTAN GENERALMENTE LIMITACIONES FISICAS O DE OPERACION DE ALGUN ELEMENTO DENTRO DEL SISTEMA O LIMITACIONES DEL PROCESO. CUANDO SE TRATA DE UN SISTEMA FISICO, POR EJEMPLO, APARECEN COMUNMENTE RESTRICCIONES DE IGUALDAD, LAS QUE NORMALMENTE REPRESENTAN LEYES FISICAS DEL COMPORTAMIENTO DEL SISTEMA, EN EL CASO DE UN SISTEMA ELECTRICO, ESTAS ULTIMAS PUEDEN SER LAS LEYES DE KIRCHHOFF Y LAS PRIMERAS PUEDEN SER VOLTAJES O CAPACIDADES DE GENERADORES. CUANDO SE INVOLUCRAN EN EL PRO-

BLEMA RESTRICCIONES DE IGUALDAD (LEYES FISICAS O DE COMPORTAMIENTO), EN OCASIONES ES CONVENIENTE SUBDIVIDIR LAS VARIABLES Z EN 2 TIPOS DE VARIABLES: VARIABLES DE CONTROL Y VARIABLES DE ESTADO DEL SISTEMA, YA QUE EN CONDICIONES NORMALES ES POSIBLE EJERCER ALGUN TIPO DE CONTROL SOBRE EL SISTEMA PARA CONDUCIRLO A UN ESTADO DETERMINADO. ES CONVE-- NIENTE NOTACIONALMENTE EN CIERTAS OCASIONES, SUSTITUIR LA VARIABLE Z POR EL PAR DE VECTORES (X,U), CON LO CUAL EL PROBLEMA (1) QUEDA COMO SIGUE:

$$\min_{x,u} F(x,u)$$

$$\text{sujeto a:} \quad g(x,u) = b \qquad (2)$$

$$h(x,u) \leq c$$

$$\text{generalmente:} \quad x \geq 0, u \geq 0$$

NOTESE QUE EL CONJUNTO DE RESTRICCIONES $g(x,u)=b$ Y $h(x,u)\leq c$ ES LA INTERSECCION DE LOS CONJUNTOS (X,U) PARA LOS CUALES SE CUMPLEN SI MULTANEAMENTE EL CONJUNTO DE RESTRICCIONES:

$$g_i(x,u) = b_i \qquad i = 1,2,\ldots,m$$

$$h_j(x,u) \leq c_j \qquad j = 1,2,\ldots,p$$

ESTE CONJUNTO (X,U) SE DENOMINA EL CONJUNTO DE SOLUCIONES FACTIBLES O CONJUNTO DE OPORTUNIDADES DEL PROBLEMA.

OPTIMIZACION CLASICA

CON EL FIN DE INICIAR EN FORMA SENCILLA EL ANALISIS DE LAS - CONDICIONES NECESARIAS O REQUISITOS PARA QUE (X,U) SEA UN PUNTO OPTIMO SE SUPONDRA QUE $m=p=0$, ES DECIR QUE NO EXISTEN RESTRICCIONES PARA Z, SINO QUE UNICAMENTE SE DESEA ENCONTRAR UN VECTOR Z REAL QUE HAGA MI NIMA A LA FUNCION F(Z). A ESTE CASO Y AL CASO CON RESTRICCIONES DE I-- GUALDAD SE ACOSTUMBRA LLAMARLOS COMO AL PROBLEMA DE OPTIMIZACION CLASI

6A. LA SOLUCION SE ENCUENTRA AL OBTENER EL PUNTO Z^*, TAL QUE PARA CUALQUIER CAMBIO ΔZ EN EL VECTOR Z^* SE CUMPLE:

$$F(z^*) \leq F(z^* + \Delta z) \tag{3}$$

(EN TODO LO QUE SIGUE SE SUPONDRA QUE F ES DOBLE DIFERENCIABLE CONTINUAMENTE Y CON DERIVADAS FINITAS Y CONTINUAS). DESARROLLANDO EN SERIE DE TAYLOR ALREDEDOR DE Z^*:

$$F(z^*+\Delta z) = F(z^*) + \left[\frac{\partial F}{\partial z}\right]'_{z^*} \Delta z + \frac{1}{2}\Delta z' \left[\frac{\partial^2 F(z^*+\theta\Delta z)}{\partial z^2}\right] \Delta z \tag{4}$$

$$0 < \theta < 1$$

ENTONCES, SE OBTIENE DE (2) Y (3):

$$\left[\frac{\partial F}{\partial z}\right]'_{z^*} \Delta z + \frac{1}{2}\Delta z' \left[\frac{\partial^2 F(z^*+\theta\Delta z)}{\partial z^2}\right] \Delta z \geq 0 \tag{5}$$

EN DONDE

$\left[\frac{\partial F}{\partial z}\right]$ ES EL VECTOR GRADIENTE DE F(Z)

Y

$\left[\frac{\partial^2 F}{\partial z^2}\right]$ ES LA MATRIZ HESSIANA O DE SEGUNDAS DERIVADAS.

ESTA DESIGUALDAD SE ACOSTUMBRA LLAMAR "DESIGUALDAD FUNDAMENTAL", LA CUAL DEBE CUMPLIRSE PARA CUALQUIER VARIACION ARBITRARIA ΔZ, EN PARTICULAR, SI SE TOMA LA COMPONENTE i DEL VECTOR ΔZ NEGATIVA Y TODAS LAS DEMAS NULAS, TOMANDO LIMITES SE IMPLICA $\frac{\partial F}{\partial z_i} \leq 0$. SI POR EL CONTRARIO SE TOMA $\Delta z_i > 0$ Y TODAS LAS DEMAS CERO, TOMANDO LIMITES SE IMPLICA $\frac{\partial F}{\partial z_i} \geq 0$ POR LO TANTO:

$$\frac{\partial F}{\partial z_i} = 0 \tag{6}$$

AHORA, USANDO (6), LA DESIGUALDAD FUNDAMENTAL IMPLICA:

$$\Delta z' \left[\frac{\partial^2 F}{\partial z^2}\right]_{z^*} \Delta z \geq 0 \qquad (7)$$

ES DECIR, LA MATRIZ HESSIANA DEBE SER POSITIVA SEMIDEFINIDA PARA - QUE z^* SEA UN PUNTO EXTREMO (MINIMO O PUNTO DE INFLEXION) O ESTRICTAMENTE POSITIVA DEFINIDA PARA GARANTIZAR UN MINIMO LOCALMENTE EN LA VECINDAD DE z^*.

OPTIMIZACION CON RESTRICCIONES DE IGUALDAD

EL PROBLEMA CONSISTE EN ENCONTRAR z^* TAL QUE SE TENGA $\min F(z)$ Y AL MISMO TIEMPO SE CUMPLAN LAS RESTRICCIONES DE IGUALDAD $g(z) = b$ ES DECIR:

$$\begin{aligned} &\min F(z) \\ &\text{s.a. } g(z) = b \end{aligned} \qquad (8)$$

PARA QUE EXISTA SOLUCION A ESTE PROBLEMA, SE REQUIERE IMPONER ALGUNAS CONDICIONES MAS FUERTES QUE EN EL CASO ANTERIOR, LAS CUALES PERMITEN TOMAR EN CUENTA LAS CARACTERISTICAS DE LAS RESTRICCIONES DE IGUALDAD. LA CONDICION FUNDAMENTAL REQUERIDA, ES QUE SEA POSIBLE APLICAR EL "TEOREMA DE LA FUNCION IMPLICITA" AL CONJUNTO DE RESTRICCIONES DE IGUALDAD. ESTO IMPLICA QUE SI EXISTEN m RESTRICCIONES, ES POSIBLE EFECTUAR UNA PARTICION DE LAS n VARIABLES O COMPONENTES DE Z, EN UN VECTOR X CON m COMPONENTES Y UN VECTOR U CON $n-m$ COMPONENTES; Y QUE, ADEMAS SEA POSIBLE RESOLVER PARA X A PARTIR DE LAS - m RESTRICCIONES EN LA "VECINDAD" DE LA SOLUCION $z^* = (x^*, u^*)$ ES DECIR:

$$x = x(u) \qquad (9)$$

EN OTRAS PALABRAS, LAS RELACIONES FUNCIONALES (9) SON EQUIVALENTES A

LAS RESTRICCIONES (8)[1]. SUSTITUYENDO ENTONCES (9) EN LA FUNCION OBJETIVO POR MINIMIZAR, EL PROBLEMA PUEDE ESCRIBIRSE COMO:

$$\min_{z} F(z) = \min_{x,u} F(x,u) = \min_{u} F(x(u),u) = \min_{u} H(u) \qquad (10)$$

ESTE ULTIMO PROBLEMA EXPRESADO EN (10) ES UN PROBLEMA SIN RESTRICCIONES QUE IMPLICITAMENTE INVOLUCRA A LAS RESTRICCIONES $g(z)=b$ Y QUE SU ESPACIO DE SOLUCIONES SE HA REDUCIDO AL ESPACIO DE LAS $n-m$ VARIABLES DE CONTROL U. ENTONCES LAS CONDICIONES DEL OPTIMO PUEDEN OBTENERSE EN FORMA SENCILLA COMO SIGUE: LA CONDICION NECESARIA PARA UN MINIMO LOCAL ES:

$$\frac{\partial H'}{\partial u} = \frac{\partial F(x,u)'}{\partial u} = \frac{\partial F'}{\partial u} + \frac{\partial F'}{\partial x}\frac{\partial x}{\partial u} \qquad (11)$$

PUESTO QUE LAS RESTRICCIONES $g(x,u)=b$ SE PUEDEN ESCRIBIR COMO UNA IDENTIDAD:

$$g(x,u) \equiv b \qquad (12)$$

DERIVANDO:

$$\frac{\partial g}{\partial u} + \frac{\partial g}{\partial x}\frac{\partial x}{\partial u} = 0 \qquad (13)$$

COMO LA MATRIZ $\left[\frac{\partial g}{\partial x}\right]$ ES NO SINGULAR, SE PUEDE RESOLVER PARA LA MATRIZ $\frac{\partial x}{\partial u}$:

$$\frac{\partial x}{\partial u} = -\left[\frac{\partial g}{\partial x}\right]^{-1}\left(\frac{\partial g}{\partial u}\right) \qquad (14)$$

Y LA CONDICION (11) SE PUEDE ESCRIBIR COMO:

$$\frac{\partial H'}{\partial u} = \frac{\partial F'}{\partial u} - \frac{\partial F'}{\partial x}\left[\frac{\partial g}{\partial x}\right]^{-1}\left(\frac{\partial g}{\partial u}\right) = 0 \qquad (15)$$

OBVIAMENTE:

$$\frac{\partial F'}{\partial x} - \frac{\partial F'}{\partial x}\left[\frac{\partial g}{\partial x}\right]^{-1}\left(\frac{\partial g}{\partial x}\right) = 0 \qquad (16)$$

[1]NOTA: PARA GARANTIZAR LA EXISTENCIA DE (9), ES NECESARIO QUE EL JACOBIANO DE LAS RESTRICCIONES CON RESPECTO A X SEA DE RANGO m.

SE PUEDE DEFINIR EL VECTOR λ LLAMADO DE MULTIPLICADORES DE LAGRANGE COMO SIGUE:

$$\lambda' = [\lambda_1, \lambda_2, \ldots \lambda_m] = -\frac{\partial F}{\partial x}' \left[\frac{\partial g}{\partial x}\right]^{-1} \qquad (17)$$

POR LO TANTO, LAS CONDICIONES NECESARIAS (15) Y (16) SE PUEDEN ESCRIBIR COMO:

$$\frac{\partial F}{\partial u}' + \lambda' \frac{\partial g}{\partial u} = 0 \qquad (15')$$

$$\frac{\partial F}{\partial x}' + \lambda' \frac{\partial g}{\partial x} = 0 \qquad (16')$$

O EN FORMA GLOBAL CON Z EN LUGAR DE (X, U):

$$\frac{\partial F}{\partial z}' + \lambda' \frac{\partial g}{\partial z} = 0 \qquad (18)$$

SI SE OBSERVA LA CONDICION (18), PUEDE OBSERVARSE QUE LAS CONDICIONES NECESARIAS JUNTO CON LAS RESTRICCIONES ORIGINALES $g(z) = b$ PUEDEN OBTENERSE DERIVANDO LA FUNCION $F(z) + \lambda'[g(z) - b]$ CON RESPECTO A LAS VARIABLES Z Y λ. ESTE ULTIMO RESULTADO CORRESPONDE A LA TECNICA DE LOS MULTIPLICADORES DE LAGRANGE APLICADA AL PROBLEMA GENERAL DE OPTIMIZACION CLASICA. ESTA TECNICA CONSISTE EN LA APLICACION DE LOS 3 PASOS SIGUIENTES:

1.- SE INTRODUCE UN NUEVO VECTOR DE VARIABLES λ CON m COMPONENTES

2.- SE DEFINE LA FUNCION DE LAGRANGE FORMADA POR LA SUMA DE LA FUNCION OBJETIVO F Y EL PRODUCTO INTERNO DEL VECTOR λ DE LOS MULTIPLICADORES DE LAGRANGE POR LAS RESTRICCIONES DE IGUALDAD $g(z) - b = 0$.

$$L(z) = F(z) + \lambda'[g(z) - b] \qquad (19)$$

O

$$L(x, u) = F(x, u) + \lambda'[g(x, u) - b] \qquad (19')$$

3.- SE ENCUENTRA EL PUNTO (z^*,λ^*) ó (x^*,u^*,λ) PARA EL CUAL SE ANULAN TODAS LAS DERIVADAS PARCIALES DE PRIMER ORDEN, ES DECIR:

$$\frac{\partial L'}{\partial z} = \frac{\partial F'}{\partial z} + \lambda' \frac{\partial g}{\partial z} = 0 \qquad n \text{ ECUACIONES}$$

$$\frac{\partial L}{\partial \lambda} = g(z) - b = 0 \qquad m \text{ ECUACIONES} \tag{20}$$

AL ANALIZAR ESTAS ULTIMAS CONDICIONES DEBE NOTARSE QUE LAS PRIMERAS n CONDICIONES IMPLICAN QUE EL GRADIENTE DE LAFUNCION OBJETIVO EVALUADO EN EL PUNTO OPTIMO, ES UNA COMBINACION LINEAL DE LOS GRADIENTES DE LAS FUNCIONES DE RESTRICCION, EN LA CUAL LOS COEFICIENTES CONSTITUYEN LOS MULTIPLICADORES DE LAGRANGE. LAS ULTIMAS m CONDICIONES SIMPLEMENTE REPRESENTAN A LAS RESTRICCIONES DE IGUALDAD. POR LO TANTO, LAS CONDICIONES (20) IMPLICAN QUE "Z^* ESTA EN EL CONJUNTO FACTIBLE DE LAS RESTRICCIONES O CONJUNTO DE OPORTUNIDADES DEL PROBLEMA, Y QUE LA DIRECCION PREFERENTE DE VARIACION PARA LA FUNCION OBJETIVO ES UNA COMBINACION LINEAL DE LOS VECTORES NORMALES (GRADIENTES) A LAS CURVAS DE RESTRICCIONES". ESTA INTERPRETACION GEOMETRICA PUEDE OBSERVARSE A PARTIR DE LA DIFERENCIAL DE LAS ECUACIONES DE RESTRICCION $g_i(z) - b_i = 0$ YA QUE:

$$dg_i = \sum_k \frac{\partial g_i}{\partial z_k} dz_k = 0 \tag{21}$$

Y PUESTO QUE LOS dz_k ESTAN EN LA DIRECCION TANGENTE A LA CURVA, EL VECTOR $\frac{\partial g_i}{\partial z_k}$ ES NORMAL A LA CURVA $g_i(z) - b_i = 0$.

LAS CONDICIONES DE 2DO. ORDEN ESTABLECEN QUE LA MATRIZ HESSIANA DEL LAGRANGEANO DEBE SER POSITIVA DEFINIDA CUANDO SE EVALUA EN EL PUNTO MINIMO LOCAL SUJETA A LAS CONDICIONES DE QUE LA DIRECCION DE EVALUACION SE ENCUENTRE SOBRE EL HIPERPLANO TANGENTE A LAS SUPERFICIES DE RESTRICCION, LO CUAL ANALITICAMENTE PUEDE EXPRESARSE SEGUN (22) Y (23):

$$\Delta z^{1} \begin{pmatrix} \frac{\partial^2 L}{\partial z_1^2} & \frac{\partial^2 L}{\partial z_1 \partial z_2} & \cdots \\ \frac{\partial^2 L}{\partial z_2 \partial z_1} & \frac{\partial^2 L}{\partial z_2^2} & \cdots \\ \frac{\partial^2 L}{\partial z_n \partial z_1} & \cdots & \frac{\partial^2 L}{\partial z_n^2} \end{pmatrix} \Delta z \qquad (22)$$

SUJETA A:

$$\Delta g = \left[\frac{\partial g}{\partial z}\right] \Delta z = 0 \qquad (23)$$

CUANDO LA MATRIZ HESSIANA ES POSITIVA DEFINIDA, SUJETA A LAS CONDICIONES (23), LAS CONDICIONES (20) SON SUFICIENTES. NOTESE QUE LAS CONDICIONES DE 2DO. ORDEN NO IMPLICAN QUE LA MATRIZ HESSIANA DE LA FUNCION OBJETIVO SEA POSITIVA DEFINIDA, SINO QUE, ESTO DEBE CUMPLIRSE PARA LA MATRIZ HESSIANA DE LA FUNCION DELAGRANGE.

INTERPRETACION DE LOS MULTIPLICADORES DE LAGRANGE

SE HA VISTO QUE LA SOLUCION DEL PROBLEMA DE OPTIMIZACION NOS PROPORCIONA ADEMAS DE LOS VALORES DE LAS VARIABLES Z, LOS VALORES DE LOS MULTIPLICADORES DE LAGRANGE λ, QUE TIENEN MUCHA IMPORTANCIA, YA QUE PROPORCIONAN UNA MEDIDA DE LA SENSITIVIDAD DEL VALOR OPTIMO DE LA FUNCION OBJETIVO F^* A PEQUEÑAS VARIACIONES EN LAS CONSTANTES b Y c DE LAS RESTRICCIONES, ES DECIR:

$$\lambda^* = \frac{\partial F^*}{\partial b} \qquad (24)$$

PARA PROBAR LO ANTERIOR, DEBE PROBARSE ANTES QUE SI SE TRATAN LAS b's COMO VARIABLES, ES POSIBLE RESOLVER A PARTIR DE LAS CONDICIONES DE PRIMER ORDEN (20), PARA LAS VARIABLES Z Y λ EN FUNCION DE LAS VARIABLES b. PARA ESTO, SE PUEDEN CONSIDERAR LAS CONDICIONES (20) COMO SI-

GUE:

$$g(z) - b = 0$$

$$\frac{\partial F'}{\partial z} + \lambda' \frac{\partial g}{\partial z} = 0 \qquad (20')$$

LAS QUE FORMAN $m+n$ ECUACIONES CON $2m+n$ VARIABLES (b, λ, z), ENTONCES LA MATRIZ JACOBIANA DEL SISTEMA ES:

$$\begin{bmatrix} -I & 0 & \frac{\partial g}{\partial z} \\ 0 & \frac{\partial g}{\partial z} & \frac{\partial^2 L}{\partial z^2} \end{bmatrix} \qquad (25)$$

LA CUAL, ACEPTANDO LA NOSINGULARIDAD IMPUESTA A $\frac{\partial g}{\partial z}$, ES DE RANGO $m+n$ POR LO TANTO, POR EL TEOREMA DE LA FUNCION IMPLICITA ES POSIBLE RESOLVER PARA λ Y z EN FUNCION DE b:

$$\lambda = \lambda(b)$$
$$z = z(b) \qquad (26)$$

SUSTITUYENDO EN LA FUNCION DE LAGRANGE:

$$L(b) = F(z(b)) + \lambda'(b)[g(z(b)) - b] \qquad (27)$$

DERIVANDO RESPECTO A b:

$$\frac{\partial L}{\partial b} = \frac{\partial F'}{\partial z}\frac{\partial z}{\partial b} + \lambda' \frac{\partial g}{\partial z}\frac{\partial z}{\partial b} + [g(z) - b]' \frac{\partial \lambda}{\partial b} - \lambda$$

$$= \left[\frac{\partial F'}{\partial z} + \lambda' \frac{\partial g}{\partial z}\right]\frac{\partial z}{\partial b} + [g(z) - b]' \frac{\partial \lambda}{\partial b} - \lambda \qquad (28)$$

$$= -\lambda^* \leftarrow \text{"En la solución óptima"}$$

($\frac{\partial F'}{\partial z} + \lambda' \frac{\partial g}{\partial z} = 0$ en la solución óptima; $g(z) - b = 0$)

EN EL PUNTO DE LA SOLUCION OPTIMA (z^*, λ^*) LOS DOS PRIMEROS TERMINOS SE ANULAN, POR LO TANTO:

$$\frac{\partial L(z^*, \lambda^*)}{\partial b} = \frac{\partial F^*}{\partial b} = -\lambda^* \qquad (29)$$

DEBIDO A (29), AL OBTENER LA SOLUCION OPTIMA A UN PROBLEMA DE LA FORMA DE (8), SE OBTIENE ADEMAS UNA MEDIDA DE LA SENSITIVIDAD, YA QUE LOS MULTIPLICADORES DE LAGRANGE INDICAN QUE TAN SENSIBLE ES EL VALOR OPTIMO DE LA FUNCION OBJETIVO A LOS CAMBIOS EN LAS CONSTANTES DE LAS RESTRICCIONES. ESTAS CONSTANTES GENERALMENTE REPRESENTAN LA CANTIDAD DE RECURSOS DISPONIBLES O REQUERIDOS PARA LA OPERACION DE UN SISTEMA. EL SIGNIFICADO ANTERIOR DE LOS MULTIPLICADORES TAMBIEN ES VALIDO PARA LAS CONSTANTES C DE LAS RESTRICCIONES DE DESIGUALDAD EN EL PROBLEMA (I).

PROBLEMA GENERAL DE PROGRAMACION NO LINEAL

ESTE CASO ESTA REPRESENTADO POR EL PROBLEMA (I), QUE ADEMAS DE m RESTRICCIONES DE IGUALDAD CONTIENE p RESTRICCIONES DE DESIGUALDAD. EL ESTABLECIMIENTO DE LAS CONDICIONES QUE SE DEBEN CUMPLIR EN EL PUNTO OPTIMO O EN LA SOLUCION, REQUIERE DE UNA GENERALIZACION DE LA APLICACION DE LOS MULTIPLICADORES DE LAGRANGE, LA CUAL CONDUCE A LA OBTENCION DEL TEOREMA DE KUHN Y TUCKER, EL QUE NOS INTRODUCE AL PROBLEMA GENERAL DE PROGRAMACION NO LINEAL EN VARIABLES REALES (O EN EL ESPACIO EUCLIDEANO n-DIMENSIONAL). EXISTEN DIVERSAS TEORIAS SOBRE ESTE PROBLEMA, LAS CUALES SON ENCAMINADAS A LA OBTENCION DE PROPOSICIONES GENERALES Y DEFINITIVAS DE LAS CONDICIONES NECESARIAS Y SUFICIENTES EN LA SOLUCION. CABE NOTAR QUE TALES RESULTADOS NO SE PRETENDA OBTENERLOS MEDIANTE UNA SIMPLE EXTENSION DE LA TEORIA APLICABLE A LOS PROBLEMAS QUE CONTIENEN UNICAMENTE RESTRICCIONES DE IGUALDAD. DADO EL CARACTER INTRODUCTORIO DE ESTAS NOTAS, LA INCORPORACION DE LAS RESTRICCIONES DE DESIGUALDAD SE HARA EN FORMA SENCILLA PARA FACILITAR EL DESARROLLO DEL TEMA. SI SE OBSERVA EL PROBLEMA (I), LAS RESTRICCIONES DE DESIGUALDAD PUEDEN

CONVERTIRSE A RESTRICCIONES DE IGUALDAD MEDIANTE LA ADICION DE UN TERMINO POSITIVO EN CADA RESTRICCION DE DESIGUALDAD DEFINIDO COMO:

$$s_i^2 = c_i - h_i(z) \tag{30}$$

DEBE NOTARSE QUE PARA QUE EL VECTOR Z ESTE CONTENIDO EN EL CONJUNTO FACTIBLE, SE DEBE CUMPLIR QUE $s_i^2 \geq 0$, YA QUE $h_i(z) \leq c_i$, ES DECIR, LAS COMPONENTES s_i DEBEN SER REALES.

ENTONCES, EL PROBLEMA (1) PUEDE EXPRESARSE COMO:

$$\begin{aligned} &\min F(z) \\ &\text{s.a:} \quad g(z) = b \\ &\quad h_i(z) + s_i^2 = c_i \qquad i = 1, 2, \ldots, p \end{aligned} \tag{31}$$

ESTE PROBLEMA CON p VARIABLES ADICIONALES DE HOLGURA CONTIENE UNICAMENTE RESTRICCIONES DE IGUALDAD, POR LO TANTO SE PUEDE APLICAR LA TECNICA DE LOS MULTIPLICADORES DE LAGRANGE. LOS PROBLEMAS (1) Y (31) SON EQUIVALENTES AUN CUANDO ESTE ULTIMO CONTIENE p VARIABLES ADICIONALES.

SE FORMA LA FUNCION DE LAGRANGE:

$$L(z) = F(z) + \lambda'[g(z) - b] + \sum_{i=1}^{p} \mu_i [h_i(z) + s_i^2 - c_i] \tag{32}$$

CUYAS VARIABLES SON (z, λ, μ, s) ó $(x, \mu, \lambda, \mu, s)$

LAS CONDICIONES DE PRIMER ORDEN SON:

$$\frac{\partial L'}{\partial z} = \frac{\partial F'}{\partial z} + \lambda' \frac{\partial g}{\partial z} + \sum \mu_i' \frac{\partial h_i}{\partial z} = 0 \qquad \text{n ECS.} \tag{33}$$

$$\frac{\partial L}{\partial \lambda} = g(z) - b = 0 \qquad \text{m ECS.} \tag{34}$$

$$\frac{\partial L}{\partial \mu_i} = h_i(z) + s_i^2 - c_i = 0 \qquad \text{p ECS.} \tag{35}$$

$$\frac{\partial L}{\partial s_i} = 2\mu_i s_i = 0 \qquad p \text{ ECS.} \qquad (36)$$

ELIMINANDO LA VARIABLE s_i DE LAS CONDICIONES (35) Y (36), ESTAS SE PUEDEN ENGLOBAR EN UNA SOLA CONDICION, LLAMADA DE EXCLUSION O DE HOLGURA COMPLEMENTARIA, LA CUAL SE PUEDE EXPRESAR COMO:

$$\mu_i [h_i(z) - c_i] \qquad i = 1, 2, \ldots, p \qquad (37)$$

ES POSIBLE PROBAR QUE LOS MULTIPLICADORES μ_i DE KUHN Y TUCKER ASOCIADOS A LAS RESTRICCIONES DE DESIGUALDAD ESTAN RESTRINGIDOS EN SU SIGNO MEDIANTE CONDICIONES DE NO NEGATIVIDAD. TOMANDO EN CUENTA LA CONDICION (33), LA ECUACION (23) Y LA DESIGUALDAD FUNDAMENTAL, ASI COMO EL HECHO DE QUE PARA UNA RESTRICCION DE DESIGUALDAD ACTIVA (EN SU LIMITE) SE CUMPLE QUE PARA UN MOVIMIENTO FACTIBLE Δz:

$$\Delta h_i = \frac{\partial h_i}{\partial z} \Delta z \leq 0 \qquad (38)$$

SE OBTIENE EN EL PUNTO OPTIMO:

$$\frac{\partial F(z^*)}{\partial z} \Delta z = -\mu_i \frac{\partial h_i(z^*)}{\partial z} \Delta z = -\mu_i \Delta h_i^* \geq 0 \qquad (39)$$

DE (38) Y (39) SE CONCLUYE QUE:

$$\mu_i \geq 0 \qquad (40)$$

RESUMIENDO LOS RESULTADOS ANTERIORES SE PUEDE ENUNCIAR EL:

TEOREMA DE KUHN Y TUCKER

SEA z^* UN PUNTO MINIMO LOCAL REGULAR[1] DEL PROBLEMA (1), ENTONCES EXISTE UN VECTOR λ, m-DIMENSIONAL Y UN VECTOR $\mu \geq 0$, p-DIMENSIONAL TAL QUE:

$$\frac{\partial F(z^*)}{\partial z} + \lambda' \frac{\partial g(z^*)}{\partial z} + \mu' \frac{\partial h(z^*)}{\partial z} = 0 \qquad (41)$$

$$\mu' [h(z^*) - c] = 0 \qquad (42)$$

[1] LA REGULARIDAD DE z IMPLICA LA NO SINGULARIDAD DE LA MATRIZ CUYOS RENGLONES SON LOS GRADIENTES DE h y g.

SI SE COMPARAN (41) Y (42) Y LA FORMULACION DEL PROBLEMA (I) CON LAS CONDICIONES DE LAGRANGE, NOTESE LA SIMILITUD EXISTENTE A TRAVES DE LAS CONDICIONES (41), YA QUE ESTAS PUEDEN OBTENERSE A PARTIR DE LA FUNCION DE - KUHN Y TUCKER DEFINIDA COMO:

$$T(z,\lambda,\mu) = F(z) + \lambda'[g(z)-b] + \mu'[h(z)-c] \qquad (43)$$

SIN EMBARGO, EXISTE UNA DIFERENCIA MUY IMPORTANTE A TRAVES DE LAS CONDICIONES (42), ASI COMO LA NO NEGATIVIDAD DE LOS MULTIPLICADORES μ DE KUHN Y TUCKER, YA QUE ESTO NO SE PRESENTA EN EL CASO DE LAGRANGE.

EL METODO DEL GRADIENTE

EL CONCEPTO DE DIRECCION

CUALQUIER VECTOR n-DIMENSIONAL PUEDE SERVIR COMO UNA DIRECCION. DADO UN PUNTO X, SE PUEDE GENERAR UNA LINEA RECTA QUE PASA POR X SI SE APLICA UNADIRECCION d (VECTOR CON n COMPONENTES $d_1, d_2, \ldots, d_n$) Y UN ESCALAR τ TAL QUE $-\infty \leq \tau \leq \infty$, ES DECIR, SI

$$y = x + \tau d \tag{44}$$

y RECORRE LA LINEA RECTA EN LA DIRECCION $\pm d$, QUE PASA POR X CUANDO $\tau = 0$.

PUEDE DEMOSTRARSE QUE SI NO ES NULO, EL GRADIENTE MISMO APUNTA EN UNA DIRECCION TAL, QUE UN PEQUEÑO MOVIMIENTO EN ESA DIRECCION AUMENTA A LA FUNCION; SEA:

$$f(x+\tau d) = f(x) + \nabla f(x)\,\tau d + O(\tau^2) \tag{45}$$

SI LA DIRECCION d SE SELECCIONA IGUAL AL GRADIENTE:

$$f(x+\tau \nabla f(x)) = f(x) + \tau |\nabla f(x)|^2 + O(\tau^2) \tag{46}$$

ENTONCES EN LA VECINDAD DE X PARA $\tau > 0$:

$$f(x+\tau \nabla f(x)) > f(x) \tag{47}$$

ESTE ULTIMO RESULTADO SUGIERE UN PROCEDIMIENTO PARA LA BUSQUEDA DE LA SOLUCION OPTIMA AL PROBLEMA DE PROGRAMACION NO LINEAL SIN RESTRICCIONES. SI SE TRATA DE UN PROBLEMA EN EL QUE SE REQUIERE ENCONTRAR EL MAXIMO, DADO UN PUNTO EN LA VECINDAD DEL OPTIMO SE EFECTUA UNA CORRECCION ΔZ-TAL QUE:

$$z + \Delta z = z + \tau \nabla F(z) \tag{48}$$

SE APROXIME AL PUNTO OPTIMO Z^* MEDIANTE LA SELECCION DE UN VALOR ADECUA

DO DE τ, EL CUAL SE OBTIENE DE:

$$\max_{\tau} F[z + \tau \nabla F(z)] \qquad (49)$$

EN CAMBIO, COMO $-\nabla F(z)$ APUNTA EN LA DIRECCION EN QUE LA FUNCION F(Z) DISMINUYE, SI SE TRATA DE ENCONTRAR UN MINIMO, SE EFECTUA UNA CORRECCION Δz EN LA DIRECCION $-\nabla F(z)$ SELECCIONANDO EL VALOR ADECUADO DE τ PARA OBTENER:

$$\min_{\tau} F[z - \tau \nabla F(z)] \qquad (50)$$

COMO ESTA CORRECCION SE HACE A LO LARGO DE UNA LINEA RECTA Y LA FUNCION F(Z) QUE NOS INTERESA ES NO LINEAL, SE GENERA UN PROCESO ITERATIVO PARA LA BUSQUEDA DE LA SOLUCION OPTIMA z^*. ESTE PROCEDIMIENTO SE CONOCE COMO EL METODO DEL GRADIENTE Y SE PUEDE RESUMIR COMO SIGUE:

PASO 0 SE ASIGNAN VALORES DE ARRANQUE $z = z_0$

PASO 1 SE CALCULA EL GRADIENTE $\nabla F(z)$. SI LA MAGNITUD DEL GRADIENTE TIENDE A CERO, EL PROCESO TERMINA Y LA SOLUCION ES $z^* = z$

PASO 2 SE OBTIENE τ^* TAL QUE: $\tau^* = \{\tau ; \min_{\tau} F[z - \tau \nabla F(z)]\}$

PASO 3 SE CALCULA UN NUEVO VALOR DE Z CON LA EXPRESION

$$z_{nueva} = z + \Delta z = z - \tau^* \nabla F(z)$$

Y SE REPITE EL PROCESO APLICANDO DE NUEVO LOS PASOS 1,2 Y 3 HASTA LOGRAR LA CONVERGENCIA.

METODO DEL GRADIENTE REDUCIDO

ESTE METODO ES DIRECTAMENTE APLICABLE AL PROBLEMA DE PROGRAMACION NO LINEAL CON RESTRICCIONES DE IGUALDAD. SU NOMBRE PROVIENE DE LA FORMA PARTICULAR QUE TOMA LA EXPRESION (15), YA QUE AL REPRESENTAR A LAS VARIABLES Z COMO VARIABLES X Y U (DE ESTADO Y DE CONTROL RESPECTIVAMENTE), IMPLICITAMENTE LAS VARIABLES DE CONTROL U TOMAN EL PAPEL DE VARIABLES DEL PROBLEMA Y ESTE REDUCE SU DIMENSIONALIDAD A $n-m$ VARIABLES.

BAJO OTRO PUNTO DE VISTA, ESTE ENFOQUE TAMBIEN PUEDE VISUALIZARSE COMO UNA TECNICA DE DESCOMPOSICION NO LINEAL.

EL PROCEDIMIENTO PUEDE VISUALIZARSE FACILMENTE A PARTIR DEL METODO DEL GRADIENTE YA ESTABLECIDO Y DE LAS RELACIONES (15) Y(16) O (15') Y (16'), ASI COMO DEL CUMPLIMIENTO DE LAS RESTRICCIONES EN EL PUNTO OPTIMO (CONDICIONES DE LAGRANGE).

ESQUEMA DEL METODO DEL GRADIENTE REDUCIDO PARA OBTENER EL MINIMO DE F(Z)

PASO 0 SE SUPONE UN CONJUNTO DE VALORES PARA U (VALORES DE ARRANQUE)

PASO 1 SE CALCULAN LOS VALORES DE X A PARTIR DE:

$$g(x,u) - b = 0$$

PASO 2 SE CALCULAN LOS VALORES DE λ A PARTIR DE LA EC. (16'):

$$\lambda' = -\left(\frac{\partial F}{\partial x}\right)\left(\frac{\partial g}{\partial x}\right)^{-1}$$

PASO 3 SE CALCULAN LOS VALORES DEL GRADIENTE REDUCIDO $\frac{\partial L}{\partial u}$ QUE EN GENERAL NO SERAN NULOS, A PARTIR DE (15) O (15'):

$$\nabla L_u = \nabla F_u' + \lambda' \nabla g_u = \frac{\partial F'}{\partial u} + \lambda' \frac{\partial g}{\partial u}$$

PASO 4 SE TOMA LA DIRECCION DEL GRADIENTE REDUCIDO CON SIGNO NEGATIVO Y SE CALCULA EL ESCALAR τ^* TAL QUE:

$$\tau^* = \{\tau;\ \min_{\tau} L(x+\tau\Delta x, u+\tau\Delta u)\}$$

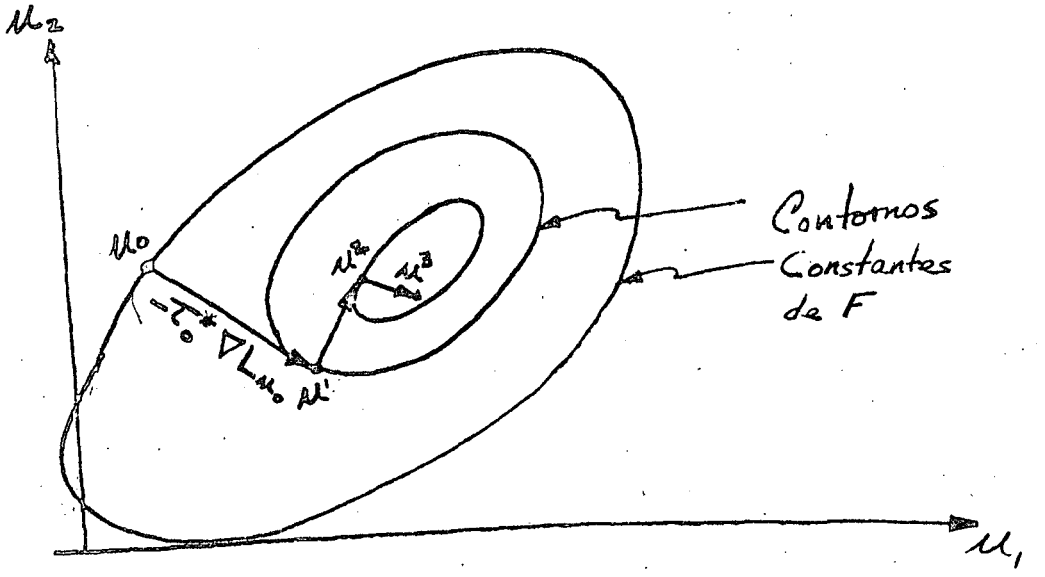
CON $\Delta u = -\nabla L_u$ Y $\Delta x = \left(\frac{\partial g}{\partial x}\right)^{-1}\left(\frac{\partial g}{\partial u}\right)\left(\frac{\partial L}{\partial u}\right)$

PASO 5 SE CALCULAN LOS NUEVOS VALORES DE U APLICANDO EL VALOR DE τ^*

$$u_{nueva} = u + \tau^* \Delta u = u - \tau^* \nabla L_u$$

PASO 6 SE PRUEBA LA CONVERGENCIA DEL PROCESO: SI LA MAGNITUD DEL GRADIENTE REDUCIDO TIENDE A CERO, EL PROCESO TERMINA. EN CASO CONTRARIO SE REPITE EL PROCESO CON LOS NUEVOS VALORES DE U A PARTIR DEL PASO 1.

GEOMETRICAMENTE, EN EL ESPACIO DE LAS VARIABLES DE CONTROL U, EL PROCESO PUEDE VISUALIZARSE COMO SE MUESTRA EN LA FIGURA SIGUIENTE.



REPRESENTACION GEOMETRICA DEL PROCESO DE DISMINUCION DE LA FUNCION F(Z)

METODO DE PENALIZACION

UN METODO COMUNMENTE USADO PARA EL MANEJO TANTO DE LAS RESTRICCIONES DE IGUALDAD COMO DE DESIGUALDAD, ES EL CONOCIDO COMO EL METODO DE LA FUNCION DE PENALIZACION, CUYA IDEA ES MUY SENCILLA. SUPONGASE QUE SE QUIERE RESOLVER EL PROBLEMA (1); LA IDEA DEL METODO DE LA FUNCION DE PENALIZACION CONSISTE EN REEMPLAZAR EL PROBLEMA (1) POR UN PROBLEMA NO RESTRINGIDO DE LA FORMA:

$$\min F(z) + \gamma P(z) \qquad (51)$$

EN DONDE γ ES UNA CONSTANTE POSITIVA Y P(Z) ES UNA FUNCION REAL DE n VARIABLES REALES QUE SATISFACE:

i) P(Z) ES CONTINUA

ii) $P(Z) \geq 0$ PARA TODA Z CON COMPONENTES REALES

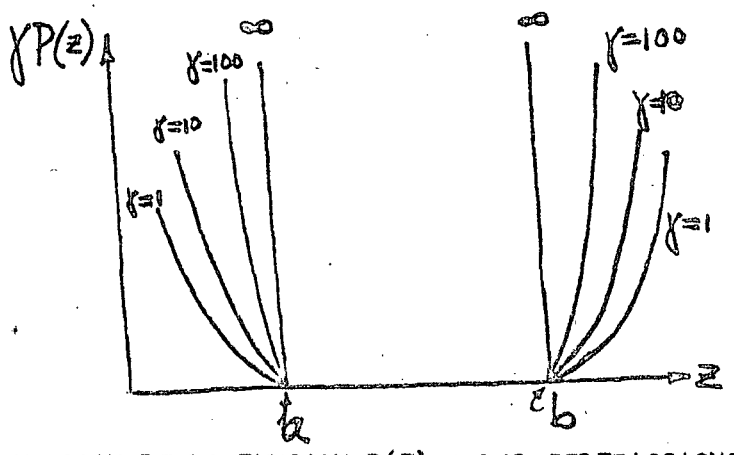
iii) $P(Z) = 0$ SI Y SOLO SI Z ESTA EN LA REGION FACTIBLE.

EN EL CASO DEL PROBLEMA (1), UNA FUNCION DE PENALIZACION UTIL ES:

$$P(z) = \sum_{i=1}^{m} \left(q_i(z) - b_i\right)^2 + \sum_{i=1}^{p} \left(\max\left[0, h_i(z) - \epsilon_i\right]\right)^2 \qquad (52)$$

LA SIGUIENTE FIGURA MUESTRA $\gamma P(z)$ PARA EL CASO UNIDIMENSIONAL DE Z, EN

EL QUE SE TIENEN 2 RESTRICCIONES DE DESIGUALDAD $z-b\leq 0$, $a-z\leq 0$



APROXIMACION DE LA FUNCION P(Z) A LAS RESTRICCIONES VERDADERAS

PARA VALORES GRANDES DE γ ES CLARO QUE EL PUNTO MINIMO DEL PROBLEMA (51) SE LOCALIZARA EN UNA REGION EN DONDE P(Z) SE HAGA PEQUEÑA, -- POR LO QUE PARA VALORES CRECIENTES DE γ SE ESPERA QUE LOS PUNTOS CORRESPONDIENTES A LA SOLUCION SE APROXIMARAN A LA REGION FACTIBLE, LO CUAL TENDERA A MINIMIZAR A LA FUNCION F(Z).

ES FACIL VISUALIZAR QUE PARA EL CASO DE RESTRICCIONES DE IGUALDAD, EN CASO DE EXISTIR UN MINIMO FACTIBLE, LA FUNCION DE PENALIZACION - DEL TIPO CUADRATICA TENDERA A SATISFACER LAS RESTRICCIONES DE IGUALDAD, YA QUE:

$$\min \left[g_i(z)-b_i\right]^2 = 0 \qquad (53)$$

EXISTEN TECNICAS QUE PERMITEN GENERAR UNA SECUENCIA CRECIENTE DE VALORES PARA γ , LOS CUALES SE ASOCIAN A UNA SECUENCIA DE PROBLEMAS CUYA SOLUCION TIENDE A LA SOLUCION DE (1).

EL METODO EXTENDIDO DEL GRADIENTE REDUCIDO

ES POSIBLE ADAPTAR LA TECNICA DEL GRADIENTE REDUCIDO PARA MANIPULAR LAS RESTRICCIONES DE DESIGUALDAD MEDIANTE 2 FORMAS DISTINTAS: a)A-

GREGANDO VARIABLES DE HOLGURA (TERMINOS POSITIVOS) A CADA RESTRICCION DE DESIGUALDAD PARA CONVERTIRLA EN IGUALDAD, LO QUE EN GENERAL NO ES EFICIEN TE YA QUE AUMENTA EL NUMERO DE VARIABLES Y AUMENTA EL NUMERO DE RESTRIC- CIONES DE IGUALDAD O) MANEJANDO LAS VARIABLES DE CONTROL EN FORMA DIR RECTA EN SUS RESTRICCIONES DE COTAS SUPERIORES O INFERIORES Y MANEJANDO MEDIANTE TECNICAS DE PENALIZACION A LAS VARIABLES DE ESTADO Y LAS RES-- TRICCIONES DE DESIGUALDAD ASOCIADAS A ELLAS. ESTA ULTIMA OPCION HA SIDO UTILIZADA EN FORMA SIMILAR CON CIERTO EXITO EN LA OPTIMIZACION DE LA O- PERACION, EN EL CALCULO DE FLUJOS OPTIMOS Y EN LA PLANEACIONDE SISTE-- MAS ELECTRICOS. ES SENCILLO VISUALIZAR ESTA EXTENSION AL PROCEDIMIENTO DEL GRADIENTE REDUCIDO Y A CONTINUACION SE PRESENTA UN POSIBLE ESQUEMA & PARA LA BUSQUEDA DE LA SOLUCION, EN EL CUAL, SI SE COMPARA CON EL ESQUE- MA ANTERIOR, PUEDEN OBSERVARSE LAS DIFERENCIAS IMPORTANTES ENTRE AMBOS.

ESQUEMA EXTENDIDO DEL METODO DEL GRADIENTE REDUCIDO PARA MINIMO DE F(Z)

PASO 0 SE SUPONE UN CONJUNTO FACTIBLE DE VALORES DE ARRANQUE PARA U

PASO 1 SE CALCULAN LOS VALORES DE X A PARTIR DE:

$$g(x,u)-b=0$$

PASO 2 SE DETECTAN LAS RESTRICCIONES DE DESIGUALDAD $h(x,u)-c\leq 0$ QUE SE VIOLAN Y SE INCLUYEN EN LA FUNCION OBJETIVO COMO FUNCIONES DE PENALIZACION:

$$F_p(x,u)=F(x,u)+\sum_j \gamma_j\left[h_j(x,u)-c_j\right]^2$$

SE CALCULAN LOS VALORES DE λ APLICANDO LA NUEVA FUNCION $F_p(X,U)$

$$\lambda'=-\left(\frac{\partial F_p}{\partial x}\right)\left(\frac{\partial g}{\partial x}\right)^{-1}$$

PASO 3 SE CALCULAN LOS VALORES DEL GRADIENTE REDUCIDO APLICANDO LA NUEVA FUNCION PENALIZADA $F_p(X,U)$

$$\nabla L_{P_u}=\frac{\partial L_p}{\partial u}=\frac{\partial F_p}{\partial u}+\lambda'\frac{\partial g}{\partial u}$$

PASO 4 SE TOMA LA DIRECCION DEL GRADIENTE REDUCIDO Y SE CALCULA EL ESCALAR τ^* TAL QUE:

$$\tau^*=\left\{\tau;\ \min_{\tau} L_p(x+\tau\Delta x,u+\tau\Delta u)\right\}$$

$$\text{CON}\quad \Delta u=-\frac{\partial L_p}{\partial u}=-\nabla L_{P_u}\ \text{Y}\ \Delta x=\left(\frac{\partial g}{\partial x}\right)^{-1}\left(\frac{\partial g}{\partial u}\right)\nabla L_{P_u}$$

PASO 5 SE CALCULAN LOS NUEVOS VALORES DE U APLICANDO EL VALOR DE τ^*, SOSTENIENDO LAS COMPONENTES EN SUS LIMITES SI ES QUE LA CORRECCION TIENDE A VIOLAR SUS COTAS. SE CALCULAN LOS VALORES DE X EN LA MISMA FORMA QUE EN EL PASO 1.

PASO 6 SE REALIZAN LAS SIGUIENTES PRUEBAS: SI EL PROCESO OBTIENE CONVERGENCIA MEDIANTE LA MAGNITUD DEL GRADIENTE PERO LAS RESTRICCIONES DE DESIGUALDAD VIOLADAS NO SE APROXIMAN A SUS LIMITES, SE RE ESTABLECEN LOS FACTORES DE PENALIZACION, Y SE REPITE EL PROCESO A PARTIR DEL PASO 2. CUANDO SE CUMPLEN AMBAS TOLERANCIAS, EL PROCESO TERMINA. EN CASO CONTRARIO, EL PROCESO SE REPITE A PARTIR DEL PASO 2 CON LOS NUEVOS VALORES DE X.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA XIII NOTAS DE PROGRAMACION LINEAL

M. EN I. COSME URDAIBAY ZUBILLAGA

ENERO 1979.

NOTAS DE PROGRAMACION LINEAL

Las presentes notas describen en forma general, qué es un problema de programación líneal y muestran, aprovechando conceptos geométricos, cuales son los principios del método simplex.

Mucha gente considera a la programación lineal entre uno de los más importantes avances científicos de estos últimos años, su impacto desde 1950 ha sido extraordinario. Actualmente es una herramienta de uso estándar que ha ahorrado muchos miles de millones de dólares en todo el mundo. Docenas de libros sobre el tema se han estrito,y se han publicado una gran cantidad de artículos. De hecho, la compañía IBM en 1970 estimó que el 25 por ciento de todo el calculo científico en computadoras usa programación líneal y otras técnicas estrechamente relacionadas.

Veamos ahora que es la programación líneal. Brevemente, la programación líneal trata típicamente problemas de ubicar recursos limitados entre actividades competitivas de la mejor manera posible, es decir, de menera óptima. Este problema de ubicación de recursos se tiene cuando uno trata de definir el nivel al que han de desarrollarse ciertas actividades que compiten por

cierto tipo de recursos escasos, necesarios para poder efectuar dichas actividades. La variedad de situaciones donde es aplicable esta descripción es sumamente diversa, pueden ser situaciones tales como ubicación de facilidades de producción, selección de portafolios, selección de rutas de envío de productos etc.

La programación líneal usa un modelo matemático para describir el problema. El adjetivo líneal, significa que todas las funciones en el modelo deben ser lineales, y la palabra programación tiene un sentido de planeación es decir la programación líneal es la planeación de actividades para obtener un resultado óptimo, en otras palabras, un resultado que cumple las metas especificadas de la mejor manera posible de acuerdo al modelo matemático, entre todos los resultados factibles.

Para entender con más claridad lo que es la programación líneal, usaremos el siguiente ejemplo:

Hay una compañía que produce varios artículos de vidrio de alta calidad incluyendo entre ellos puetas y ventanas, tiene tres plantas; en la primera se tiene la forja para el aluminio; en la segunda se hacen los marcos de madera y en la tercera se produce vidrio y se ensamblan los productos.

Debido a que han empezado a declinar los ingresos, la Gerencia de la compañía ha decidido descontinuar algunos productos, y dejar la capacidad sobrante para producir uno o dos artículos nuevos que han sido solicitados. El primer producto es una puerta de vidrio con marco de aluminio, el segundo es una ventana con marco de madera. El departamento de mercado ha concluido que la compañía podría vender cualquier cantidad de cada producto que pueda producir con la capacidad de producción de la compañía, sin embargo, debido a que ambos productos deben competir por la misma capacidad de producción en la planta 3, no es claro que mezcla de productos es la más conveniente.

Despues de una pequeña investigación se determinó el porcentaje de capacidad disponible en cada planta, la utilidad que puede dejar cada producto y los requerimientos en porcentaje, de capacidad de producción que cada producto tiene por minuto en cada una de las plantas, Esta información se resume en la tabla 1. Debe notarse que cuando la capacidad de producción en la planta 3 es usada por alguno de los productos, ya no puede ser usada por el otro

Producto / Planta	Capacidad usada por unidad de producción		Capacidad Disponible
	Producto 1	Producto 2	
1	1	0	4
2	0	2	12
3	3	2	18
Utilidad Unitaria	3	5	

Tabla 1. Características de la Producción.

Para formular el modelo matemático de programación lineal para este problema, llamemos x_1 , el número de unidades del primer producto que debemos producir y x_2 el número de unidades del segundo producto, y llamemos Z la contribución por minuto a la utilidad. Entonces x_1 y x_2 son las variables de decisión del modelo y hay que seleccionar sus valores de modo de obtener el máximo de

$$Z = 3x_1 + 5x_2$$

sujeto a las limitaciones impuestas por las capacidades de producción disponibles en cada planta. En la tabla 1 se ve que cada unidad producida de producto 1 por minuto requiere 1% de la capacidad de la planta 1 y que sólo hay 4% de la capacidad disponible, esta restricción se

expresa matemáticamente como

$$x_1 \leq 4$$

Similarmente, la planta 2 impone la restricción

$$2x_2 \leq 12$$

El porcentaje de uso de la planta 3 seleccionando los valores x_1 y x_2 como tasas de producción es $3x_1 + 2x_2$, entonces la ecuación matemática de la capacidad en la planta 3 es

$$3x_1 + 2x_2 \leq 18$$

y finalmente debido a que las tasas de producción no pueden ser negativas, se requiere que $x_1 \geq 0$, $x_2 \geq 0$.

En resumen el problema puede plantearse como

$$\text{Minimizar} \quad Z = 3x_1 + 5x_2$$

sujeto a

$$\begin{aligned} x_1 &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \\ x_1 \geq 0, \quad x_2 &\geq 0 \end{aligned}$$

El problema anterior es muy pequeño, pues sólo tiene dos variables, y por consiguiente, sólo dos " dimensiones ", de modo que se puede usar un procedimiento gráfico para resolverlo. Este procedimiento involucra la construcción de una gráfica en dos dimensiones con x_1 y x_2 en los ejes. Lo primero es identificar los valores que las restricciones permiten tomar a x_1 y x_2

Los valores permisibles para (x_1, x_2) se muestran en la figura 1

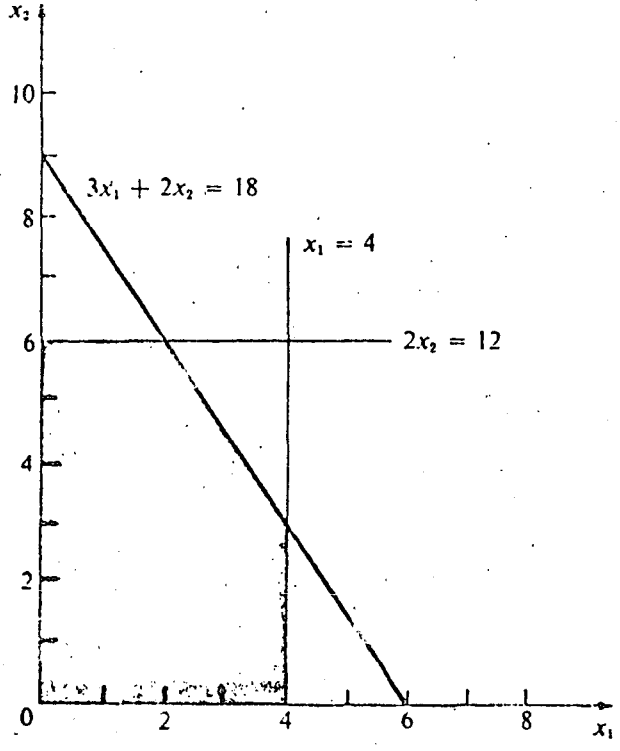


Figura 1. Valores Permisibles para (x_1, x_2)

Finalmente hay que seleccionar el punto dentro de la región de la figura 1 que maximiza el valor de $Z = 3x_1 + 5x_2$. Primero podemos buscar por ejemplo si hay algún valor que haga que la función objetivo valga 10, ello lo podemos lograr dibujando en la región de puntos factibles la línea $3x_1 + 5x_2 = 10$; como vemos en la figura 2 hay una gran cantidad de puntos sobre la línea que dan una utilidad igual a 10.

Ahora podemos intentar un mayor valor de Z, por ejemplo $Z = 20 = 3x_1 + 5x_2$. Nuevamente vemos

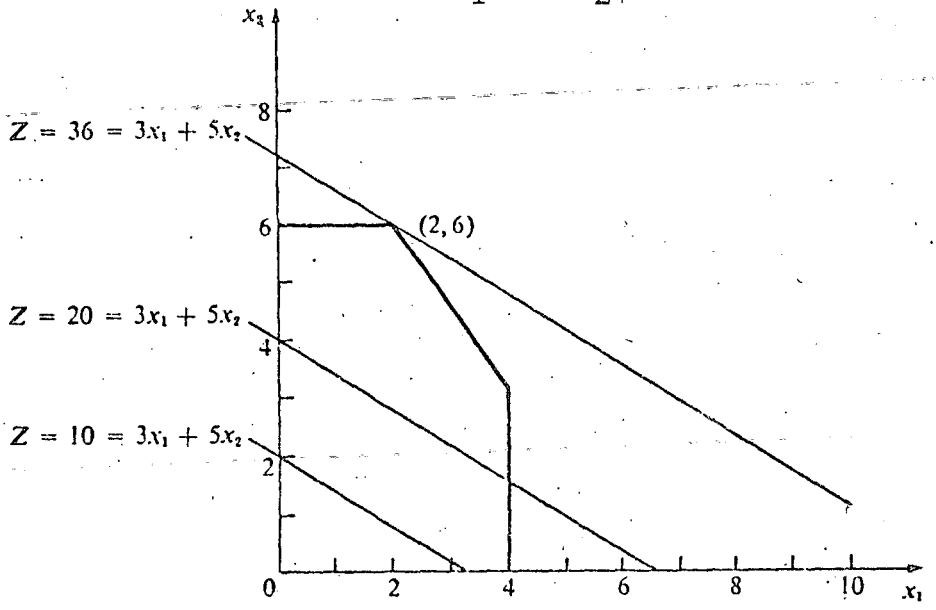


Figura 2. Búsqueda de la Solución Optima.

que hay muchos puntos que producen una utilidad igual a 20.

Nótese que la recta que produce el valor $Z = 20$ está arriba de la que produce el valor $Z = 10$ y que ambas rectas son paralelas. Es decir este procedimiento de prueba

y error no es otra cosa sino el dibujar una familia de líneas paralelas que contengan al menos un punto dentro de la región factible, y seleccionar la línea que proporciona el mayor valor de Z. En nuestro ejemplo la línea pasa por el punto (2, 6), es decir, la solución al problema es $x_1 = 2$, $x_2 = 6$, que indica que la compañía deberá producir el producto 1 a una tasa de dos por minuto y el producto 2 a una tasa de seis por minuto, con un resultado de una utilidad de 36 por minuto; más aun, no hay otra mezcla de productos, de acuerdo al modelo, que sea tan buena como esta.

Ahora podemos generalizar el problema anterior, supongamos que hay un total de m recursos limitados de cualquier clase, que deben ser distribuidos entre n actividades competitivas, también de cualquier clase. Sean x_j el nivel de la actividad j, Z la medida de efectividad total, c_j el incremento en Z que resulta de producir una unidad adicional de la actividad j. Además sea b_i la cantidad disponible de recurso i para ser distribuido y finalmente, sea a_{ij} la cantidad de recurso i consumido por cada unidad de actividad j.

Procediendo como en el ejmplo anterior, podemos formular el problema de programación líneal, el cual consiste

en seleccionar los valores x_1 , x_2 , ... , x_n de modo de maximizar $Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n$

sujeto a:
$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots a_{1n}x_n &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots a_{2n}x_n &\leq b_2 \\ \vdots \quad \vdots \qquad\qquad \vdots \quad &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots a_{mn}x_n &\leq b_m \\ x_1, x_2, \ldots, x_n &\geq 0 \end{aligned}$$

Esta última formulación del problema se conoce como forma canónica del problema de programación lineal.

Un problema más general de programación lineal puede incluir restricciones con signo de igualdad, es decir, por ejemplo

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

y además no todas las variables x_j tienen que ser restringidas a tomar valores no negativos.

A continuación veremos los principios en los que se basa el método simplex para la solución de problemas de programación lineal, usaremos para ello el ejemplo desarrollado anteriormente.

En primer lugar introduciremos alguna terminología básica en programación líneal

- Una SOLUCION FACTIBLE es una solución que satisface todas las restricciones.
- Una SOLUCION OPTIMA es una solución factible que produce el mejor valor en la función objetivo. El mejor valor será el menor o el mayor, dependiendo si se trata de un problema de mínimización o maximización.

Nótese que un problema de programación líneal puede tener una única solución óptima, como es el caso de nuestro ejemplo; puede también no tener solución, como sería el caso si al problema agregamos la restricción de que deseamos obtener una utilidad de cuando menos 50 por minuto ($3x_1 + 5x_2 \geq 50$) o, finalmente, puede tener un número infinito de soluciones óptimas, como sería el caso en el que la utilidad del producto 2 fuera igual a dos, pues en este caso la función objetivo y la restricción serían paralelas.

Una SOLUCION BASICA FACTIBLE es una solución factible que no puede representarse como combinación convexa de cualesquiera otras soluciones, es decir, que no esta en el

interior de ningún segmento de línea que conecte cualesquiera otras dos soluciones. Las soluciones básicas factibles son precisamente las soluciones que están en las esquinas del conjunto de puntos factibles. Dado que en más de tres dimensiones no es posible identificar geométricamente las soluciones básicas factibles, es conveniente identificarlas algebraicamente: las soluciones básicas factibles en un problema de programación línal con una matriz de restricciones A de dimensiones mxn, se obtienen de la solución simultenea de un sistema de ecuaciones de tamaño mxm, haciendo igual a cero las variables que no intervienen en las ecuaciones del sistema reducido.

Debe tomarse en cuenta que no todas las soluciones básicas así obtenidas son soluciones factibles.

A continuación mencionaremos tres propiedades importantes sobre las soluciones básicas factibles, ellas constituyen los principios del método simplex.

Propiedad 1 Si existe sólo una solución óptima ésta debe ser una solución básica factible; si hay multiples soluciones óptimas entonces al menos dos deben ser soluciones básicas factibles adyacentes.

Propiedad 2. Sólo hay un número finito de soluciones básicas factibles.

Propiedad 3. Si una solución básica factible medida mediante la función objetivo, es mejor que las soluciones básicas factibles adyacentes a ella, entonces es mejor que todas las otras soluciones básicas factibles.

Ahora para finalizar, resumimos el método simplex en forma de algoritmo

PASO INICIAL. Empezar en una solución básica factible.

PASO ITERATIVO Moverse a una mejor solución básica factible adyacente.

REGLA PARA FINALIZAR Parar cuando la actual solución básica factible sea mejor que todas las soluciones básicas factibles adyacentes.. Esta última solución es óptima

REFERENCIAS BIBLIOGRAFICAS.

1. - LINEAR PROGRAMMING

M. Simonnard

Prentice-Hall, 1966

2. - LINEAR PROGRAMMING AND EXTENSIONS

G. B. Dantzig

Princeton University Press, 1962

3. - LINEAR PROGRAMMING: METHODS AND APPLICATIONS

S. I. Gass.

Mc. Graw-Hill , 1958.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA XIV PROGRAMACION ENTERA Y MIXTA

M. EN I. COSME URDAIBAY ZUBILLAGA

ENERO, 1979.

Palacio de Minería · Calle de Tacuba 5, primer piso. México 1, D. F.

COMISION FEDERAL DE ELECTRICIDAD

GERENCIA GENERAL DE ESTUDIOS E INGENIERIA PRELIMINAR

SUBGERENCIA DE ESTUDIOS ELECTRICOS

DEPARTAMENTO DE METODOLOGIA

C. Urdaibay Octubre de 1978

NOTAS DE PROGRAMACION ENTERA

AJBBA/78/37 26 Páginas

RESUMEN: Se presentan una notas de programación en enteros a nivel intruductorio, para una plática de aproximadamente 90 minutos de duración. se ilustran los principales métodos de solución de problemas en enteros.

NOTAS DE PROGRAMACION ENTERA .

La programación entera es una rama de la programación matemática . Un problema general de programación matemática puede establecerse como

$$\min f(x), \quad x \in S \qquad (1)$$

el conjunto S es llamado conjunto de restricciones y f es llamada función objetivo .

Cada $x \in S$ es llamada solución factible al problema (1), y si existe un $x^\circ \in S$ tal que $-\infty < f(x^\circ) \leq f(x)$ para toda $x \in S$, entonces x° es llamada solución óptima para el problema (1) . El propósito de la programación matemática es establecer cuándo existe una solución óptima y encontrar una o tal vez todas ellas .

Un problema de programación entera es un problema de programación matemática en el que además se pide como restricción que los valores que el vector x pueda tomar sean enteros . Un problema de programación entera mixta , es un problema en el cual al menos una, pero no todas, de las componentes del vector x se restringen a tomar valores enteros .

PROGRAMACION LINEAL ENTERA .

El desarrollo de la programación entera se ha encaminado fundamentalmente a la busqueda de algorit -

mos cuando el problema es un problema lineal, es decir, cuando

$$f(x) = cx$$

y

$$S = \{ x \mid Ax = b,\ x \geq 0 \}$$

en este caso plantearemos el problema de programación entera como

$$\min\ z = cx$$

sujeto a

$$\begin{aligned} Ax &= b \\ x &\geq 0 \end{aligned} \qquad (2)$$

$$x_j \in N = \{ 0, 1, 2, \ldots \}\ \forall\ j \in J$$

donde $J \subset \Gamma = \{ 1, 2, \ldots n \}$

Nótese que si en el problema (2) $J = \emptyset$ tenemos un problema de programación lineal; si $J \subset \Gamma$ y $J \neq \Gamma$, tenemos un problema de programación lineal mixta y si $J = \Gamma$ tenemos un problema puro en enteros.

Un caso particular muy importante del problema (2) es en el que $x_j \in \{0, 1\}$ para toda j, es decir, cuando las variables son "binarias" o "lógicas", ello se debe a que muchos problemas se plantean exclusivamente en función de variables binarias, por otro lado, los algoritmos de solución son en general más eficientes para este tipo de problemas, y además cualquier problema en enteros se puede transformar a un problema en variables binarias, siempre y cuando las variables estén acotadas por arriba, por ejemplo, si $x_j \in N$ y $x_j \leq 13$, sustituimos x_j por

$$x_j = y_{1j} + 2y_{2j} + 4y_{3j} + 8y_{4j}$$

con

$$y_{ij} \in \{0,1\}, \quad i = 1,2,3,4$$

cualquier valor entre 0 y 13 puede expresarse como la suma anterior, substituyendo los valores adecuados de 0 y 1 para las y_{ij}, Si es necesario podemos agregar la restricción

$$\sum_{i=1}^{4} 2^{i-1} y_{ij} \leq 13$$

REDONDEO DE LA SOLUCION AL PROBLEMA LINEAL ASOCIADO .

El primer intento de resolver un problema, en enteros es redondear la solución obtenida al resolver el problema lineal asociado, sin embargo, la solución redondeada y la solución real pueden ser drásticamente diferentes, como puede verse en los dos ejemplos que siguen

Ejemplo 1 .

$$\min \; 10X_1 + 10X_2 + 10X_3 + 10X_4 + 9X_5 + 9X_6$$

sujeto a :

$$\begin{aligned}
X_1 + X_2 + X_3 \quad\quad + X_5 \quad\quad &\geq 1 \\
X_2 + X_3 + X_4 + X_5 \quad\quad &\geq 1 \\
X_1 \quad + X_3 + X_4 \quad\quad + X_6 &\geq 1 \\
X_1 + X_2 \quad + X_4 \quad\quad + X_6 &\geq 1
\end{aligned}$$

$$X_j \in N, \quad J = \Gamma$$

La solución del problema lineal asociado, es decir, reemplazando la restricción $X_j \in N \; \forall_j$ por $x \geq 0$, es

$$x_1 = x_2 = x_3 = x_4 = 1/3, \quad x_5 = x_6 = 0$$

el valor de la función objetivo es

$$Z^* = 13\frac{1}{3}$$

La solución óptima del problema en enteros es

$$x_1 = x_2 = x_3 = x_4 = 0, \quad x_5 = x_6 = 1$$

con un valor en la función objetivo de

$$Z^* = 18$$

Nótese que las variables que son diferentes de cero son radicalmente diferentes , lo cual en términos prácticos puede significar políticas drásticamente diferentes y que además el redondeo de la solución del problema lineal no proporciona una solución factible si se redondea a cero y si se redondea a uno, el valor de la función objetivo sería $Z^* = 40$, que es más del doble del valor óptimo

Ejemplo 2

$$\max \; 8x_1 + 8.5x_2$$

sujeto a

$$6x_1 + 7x_2 \leq 56$$
$$5x_1 + 2x_2 \leq 35$$
$$x_1, x_2 \in N$$

La solución del problema lineal asociado es

$$x_1 = 133/23, \quad x_2 = 70/23, \quad Z^* = 1659/23$$

mientras que la solución al problema discreto es

$$X_1 = 0, \quad X_2 = 8, \quad Z^* = 68$$

Obsérvese que nuevamente la solución óptima está muy lejos del óptimo del problema lineal asociado.

El problema se ilustra geométricamente en la Figura 1.

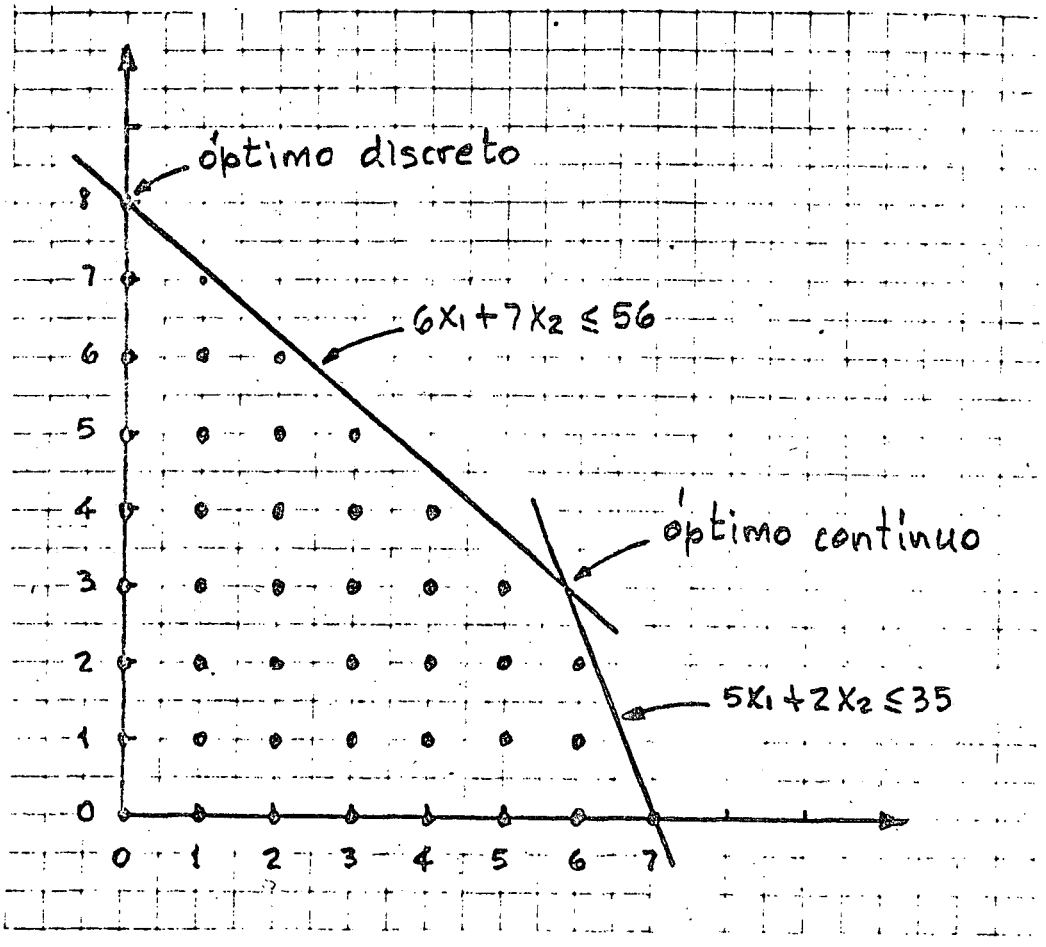


Figura 1. Región Factible del Ejemplo 2

MODELADO DE PROBLEMAS.

El problema del redondeo discutido anteriormente, no es sin embargo lo que da la mayor importancia al

problema de programación de enteros, quizá la mayor importancia del problema estriba en el hecho de poder modelar condiciones lógicas . A continuación ilustraremos unos ejemplos :

Ejemplo 3 . El problema del cargo fijo .

Este problema es de mucha importancia en la práctica, ya que hay multitud de situaciones donde, al contratar o aceptar una actividad, se debe pagar un cargo fijo, además de incurrir en costos creciente, según el nivel al que se realice la actividad . Es decir, la función de costos es como se muestra en la figura 2 .

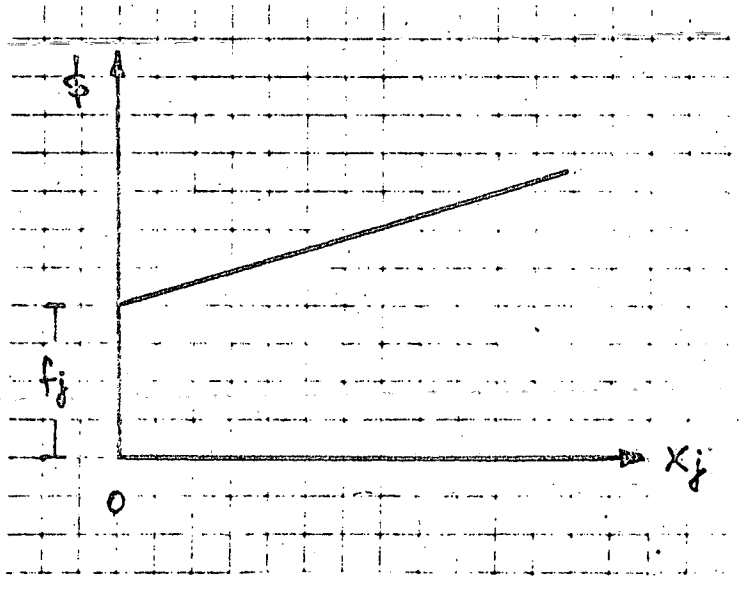


Figura 2. Problema del Cargo Fijo

En este caso , la condición lógica es: pagamos el cargo fijo f_j, si realizamos la actividad j , y no se paga

en caso contratio .

Matemáticamente el problema lo po

demos formular como sigue :

$$\min Z = \sum_{j=1}^{n} (c_j x_j + f_j \delta_j)$$

sujeto a

$$AX = b$$
$$x \geq 0$$
$$x_j \leq u_j \delta_j$$
$$\delta_j \in \{0,1\}$$

se ve claramente que si

$$\delta_j = 0 \Rightarrow x_j = 0$$

y si

$$\delta_j = 1 \Rightarrow x_j \leq u_j$$

donde u_j es una cota superior adecuada para la variable x_j .

Ejemplo 4 . La variable menor .

En este caso se desea usar solamen

te la variable que tenga el valor mínimo (o máximo) de entre dos cua-

lesquiera . Consideraremos la variable

$$u = \min \{ x_1, x_2 \}$$

y supongamos que x_1 , x_2 están acotadas .

Para modelar la variable menor requerimos de dos variables binarias δ_1, δ_2 y de tres variables continuas y_1, y_2 y y_3, entonces

$$\delta_1 = \begin{cases} 1 & \text{si } x_1 \geq x_2 \\ 0 & \text{en caso contrario} \end{cases}$$

$$\delta_2 = \begin{cases} 1 & \text{si } x_2 > x_1 \\ 0 & \text{en caso contrario} \end{cases}$$

hacemos

$$\delta_1 + \delta_2 = 1$$
$$y_1 \leq M \delta_1$$
$$y_2 \leq M \delta_2$$

finalmente

$$x_1 = y_1 + y_3$$
$$x_2 = y_2 + y_3$$
$$U = y_3$$

Para comprobar que este grupo de variables y restricciones resuelvan el problema, veamos el proceso inverso :

(a) $\delta_1 = 1 \Rightarrow \delta_2 = 0 \Rightarrow y_1 \leq M \; ; \; y_2 \leq 0 \Rightarrow y_2 = 0$

$\therefore \; x_2 = y_3 \Rightarrow x_1 = y_1 + x_2 \geq x_2 \quad \text{si } y_1 \geq 0$

$\therefore \; U = x_2$

(b) $\delta_2 = 1 \Rightarrow \delta_1 = 0 \Rightarrow y_2 \leq M\,;\; y_1 \leq 0 \Rightarrow y_1 = 0$

$\therefore \quad x_1 = y_3 \Rightarrow x_2 = y_2 + x_1 \geq x_1 \quad \text{si } y_2 \geq 0$

$\therefore \quad u = x_1$

(c) Finalmente supóngase que $x_1 < x_2$, entonces

$$y_1 + y_3 < y_2 + y_3 \Rightarrow y_1 < y_2 \Rightarrow \delta_2 = 1 \Rightarrow \delta_1 = 0$$

Ejemplo 5 . Alternativas discretas .

Supóngase que se tiene el problema siguiente :

$$\text{MIN } cx$$

sujeto a

$$Ax = b$$
$$x \geq 0$$

y $u^t x \leq \beta u$ ó $v^t x \leq \beta v$ pero no ambas .

Supóngase además que se tienen estas L_u y L_v tales que

$$\left.\begin{aligned} u^t x - \beta u &\leq L_u \\ v^t x - \beta v &\leq L_v \end{aligned}\right\} \forall x \text{ -} \ni \text{- } \{Ax = b,\, x \geq 0\}$$

entonces sea

$$\delta = \begin{cases} 1 & \text{si la restricción v debe ser respetada} \\ 0 & \text{en caso contrario} \end{cases}$$

La formulación del problema queda como

$$\min \; c\,x$$

sujeto a :

$$Ax = b$$

$$x \geq 0$$

$$u^t x - \delta L u \leq \beta u$$

$$v^t x - (1-\delta) L v \leq \beta v$$

$$\delta \in \{0,1\}$$

entonces para Lu y Lv suficientemente grandres, si $\delta = 1$ se tiene que

$u^t x \leq \beta u + Lu$ no es restrictiva, mientras que

$v^t x \leq \beta v$ si es restrictiva .

Ejemplo 6 . La función "y"

Dadas $x_1, x_2 \in \{0, 1\}$

definimos

$$\delta = \begin{cases} 1 & \text{si } x_1 \text{ y } x_2 = 1 \\ 0 & \text{en caso contrario} \end{cases}$$

entonces el problema se puede modelar como

$$\delta \leq x_1$$

$$\delta \leq x_2$$

$$\delta \geq x_1 + x_2 - 1$$

Ejemplo 7. La Función Negación

Dado $x \in \{0,1\}$ definimos

$$\delta = \begin{cases} 1 & \text{si } x=0 \\ 0 & \text{si } x=1 \end{cases}$$

entonces

$$\delta = 1 - x$$

Ejemplo 8 La Función "o"

Dados $x_1 , x_2 \in \{0,1\}$ definimos

$$\delta = \begin{cases} 1 & \text{si } x_1, x_2 \text{ o ambas valen uno} \\ 0 & \text{en caso contrario} \end{cases}$$

entonces

$$\delta \geq x_1 \quad , \quad \delta \geq x_2 \quad , \quad \delta \leq x_1 + x_2$$

Métodos de Solución

A continuación se mencionarán brevemente los principales métodos de solución de los problemas de programación lineal entera

a) Unimodularidad

Considérese el problema de programación lineal en enteros

$$\min \; cx$$

sujeto a

$$Ax = b$$
$$x \geq 0$$
$$x_j \in N, \; j \in J$$

nótese que si podemos garantizar que, para cualquier submatriz no singular B formada por m columnas de A, que $B^{-1}b$ es un vectos de enteros; entonces la restricción de integralidad es superflua es decir, podemos resolver el problema en enteros simplemente resolviendo el proble-- ma líneal asociado. Para problemas de optimización asociados con gráficas, hay varios casos en los que esto se cumple, debido a que la matriz A, que está íntimamente ligada con la metriz de incidencia de la gráfica asociada al problema tiene la propiedad de unimodularidad

Una matriz cuadrada B es llamada unimodular si $D = |\det B| = 1$. Una matriz A de dimensiones m x n es totalmente unimodular si cada submatriz cuadrada no singular es unimodular; es decir, si A es totalmente unimodular, cada base B es unimodular, y cada solución básica

$$(X_B, X_R) = (B^{-1}b, 0)$$

es entera.

Nótese que una condición necesaria para que una matriz A sea totalmente unimodular es que

$$a_{ij} \in \{-1, 0, 1\}$$

Una forma rigurosa de probar si una matriz es totalmente unimodular, se enuncia en el siguiente teorema

(la prueba puede verse en la referencia 1 . .)

Teorema

Una matriz A entera, con elementos $a_{ij} \in \{-1, 0, 1\}$ para toda i y toda j es totalmente unimodular si

1. Cada columna tiene no más de dos elementos diferentes de cero

2. Los renglones pueden ser particionados en dos subconjuntos Q_1 y Q_2 tales que

   (a) Si una columna tiene dos elementos diferentes de cero con el mismo signo, un elemento está en cada subconjunto.

   (b) Si una columna tiene dos elementos diferentes de cero con signo opuesto, los dos elementos pertenecen al mismo conjunto.

Un problema particular muy conocido en programación líneal, que tiene la propiedad de unimodularidad es el problema del transporte, el cual tiene la siguiente estructura:

$$\min \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \qquad (3)$$

Sujeto a

$$\sum_{i=1}^{m} x_{ij} = b_j \qquad j = 1,2,\ldots,n \qquad (4)$$

$$\sum_{j=1}^{n} x_{ij} = a_i \qquad i = 1,2,\ldots,m \qquad (5)$$

$$x_{ij} \geq 0 \qquad \forall i,j \qquad (6)$$

donde c_{ij} representa el costo de enviar una unidad de un bien desde el origen i hacia el destino j, entonces la función objetivo (3) busca el costo mínimo de envío de la totalidad de los bienes; la restricción (4) significa que el total de bienes enviados desde todos los origenes hacia el destino j debe ser igual a la demanda b_j del destino j; finalmente la restricción (5) significa que lo que se envía a todos los destinos desde el origen i debe ser igual a la disponibilidad a_i en el origen i.

b) Enumeración Exhaustiva.

El primer intento para resolver un problema de programación entera es hacer una enumeración exhaustiva de todas las posibles soluciones, y tomar como solución óptima aquella que mínimice el valor de la función objetivo y satisfaga todas las restricciones. Pensemos en un problema con variables binarias, y si se tienen dos variables, el número de posibles soluciones es $2^2 = 4$; si el número de variables es 3, las posibles soluciones son $2^3 = 8$; en general con n variables se tendrán 2^n posibles soluciones, muchas de ellas tal vez no satisfagan las restricciones, pero sin embargo habrá que analizarlas. En el caso de que el problema no sea en variables binarias, habrá muchas más soluciones posibles a analizar

A pesar de ser posible resolver este tipo de problemas por enumeración explicita, sólo es atractivo para problemas muy pequeños, por lo que se hace necesaria una técnica de enumeración implícita para problemas grandes.

c) Enumeración Implícita .

En este procedimiento para resolver problemas binarios, todas las soluciones (factibles y no factibles)

son enumeradas, pero la gran mayoría son numeradas implícitamente, sólo unas cuantas son explícitamente enumeradas. La noción de enumeración implícita es simple y puede ilustrarse claramente con un ejemplo.

Considerese el problema

$$\min f = 3X_1 + 8X_2 + X_3 + 16X_4 + X_5 \qquad (7)$$

Sujeto a

$$g_1(x) = -X_1 + 2X_2 + 6X_3 - 2X_4 - 3X_5 \geq 0 \qquad (8)$$

$$g_2(x) = -X_1 + 3X_3 + 2X_4 - 2X_5 \geq 2 \qquad (9)$$

$$g_3(x) = -X_1 + 5X_2 - 4X_3 + X_4 + 2X_5 \geq 5 \qquad (10)$$

$$X_i \in \{0,1\} \quad i = 1,2,\ldots,5 \qquad (11)$$

Hay 32 posibles soluciones, incluyendo las factibles y no factibles.

Dado que deseamos minimizar la función objetivo f, y que todos los coeficientes en f son positivos, lo primero que podemos hacer es igualar todas las variables a cero - llamemos $x^\circ = (0,0,0,0,0,)$ a esta solución; ello hará el valor de f lo mas pequeño posible tomando en cuenta sólo las restricciones (1). Sin embargo, a menos que el problema sea trivial, violará una o más restricciones.

En seguida buscamos las restricciones que viola esta primer solución . Es inmediato que g_2 y g_3 están violadas, pues

$$g_2(x^\circ) = 0$$
$$g_3(x^\circ) = 0$$

Analizaremos primero la restricción 9, si g_2 debe ser mayor o igual a 2, entonces una o mas variables en g_2, que tienen coeficiente positivo, deben tomar valor 1. Las variables en g_2 con coeficiente positivo son x_3 y x_4. Haciendo $x_3 = x_4 = 1$, llegamos a que g_2 valga 5 y sólo es necesario que sea mayor o igual a 2, entonces, dado que pretendemos minimizar el valor de f, debemos intentar llevar al valor uno el menor número de variables.

Hagamos entonces $x_3 = 1$ y conservemos $x_4 = 0$

Llamemos $X^1 = (0,0,1,0,0,)$

En este caso $g_2(x^1) = 3$ y ya no esta violada.

La restricción 8 no es violada por x^1, pero la restricción 10 continúa violada. Las variables con coeficiente positivo en g_3 son x_2, x_4 y x_5. Nótese que si $x_3 = 1$ y todas las variables con coeficiente positivo en g_3 valen uno, la restricción g_3 seguirá violada, ello implica que ninguna solución con $x_3 = 1$ puede ser factible, por lo que podemos desde ahora eliminar todas las soluciones que incluyen $x_3 = 1$ y podemos decir que hemos ennumerado implícitamente todas estas soluciones.

Regresemos ahora a las restricciones; con $x_3 = 0$, g_2 y g_3 son violadas, a menos que otras variables tomen valor 1 entonces, tomando en cuenta la restricción 9, x_4 debe valer 1, pues es la única variable sobrante en g_2 que tiene coeficiente positivo.

Con la solución $x^2 = (0,0,0,1,0,)$ tenemos que $g_2(x^2) = 2$, pero $g_3(x^2) = 1$, por lo que alguna combinación de x_2 y x_5 debe tomar valores uno, dado que son las variables

remanentes con coeficiente positivo. El hacer solamente $x_5 = 1$ no es suficiente para no violar g_3, y no es necesario hacer que ambas x_2 y x_5 valgan uno, entonces analicemos

$$x^3 = (0, 1, 0, 1, 0,)$$

vemos que x^3 no viola a ninguna restricción, entonces tenemos una solución factible, es además fácil darse cuenta que ésta es la solución óptima, pues el hacer cualquiera otra variable igual a uno sólo aumenta el valor de la función objetivo

c) Ramificación y Acotamiento

El método de Ramificación y Acotamiento (Branch and Bound) es una generalizacióndel método de enumeración implícita , y es útil para programación entera mixta, no sólo para programación cero-uno.

Para entender el método de ramificación y acotamiento se debe entender que un problema de programación lineal entera es un problema de programación lineal " más restringido".

Si añadimos a un problema de programación lineal la restricción de que alguna o todas las variables sean enteras, no aumentamos el tamaño de la región factible. El valor de la función objetivo asociada con la solución entera puede ser menos atractiva que la solución asociada con la solución óptima no entera, pero nunca será más atractiva.

Entonces el valor de la función objetivo asociada con la solución óptima no entera, siempre será una cota inferior para todas las soluciones " más restringidas " por el requerimiento de integralidad de algunas de las variables.

Similarmente, para un problema de maximización , la solución óptima no entera será una cota superior a todas las soluciones más restringidas.

En la solución de problemas por el método de ramificación y acotamiento se explotan las propiedades anteriores, se hace una partición de las variables, formando un árbol, se busca entre aquellos subconjuntos con la mejor cota inferior o superior, dependiendo de si se minimiza o maximiza .

Resumiendo las características antes mencionadas tenemos

1. - Si la solución a un problema lineal es una solución con todas las variables tomando un valor entero, esta solución entera es una solución óptima al correspondiente problema de programación lineal entera.

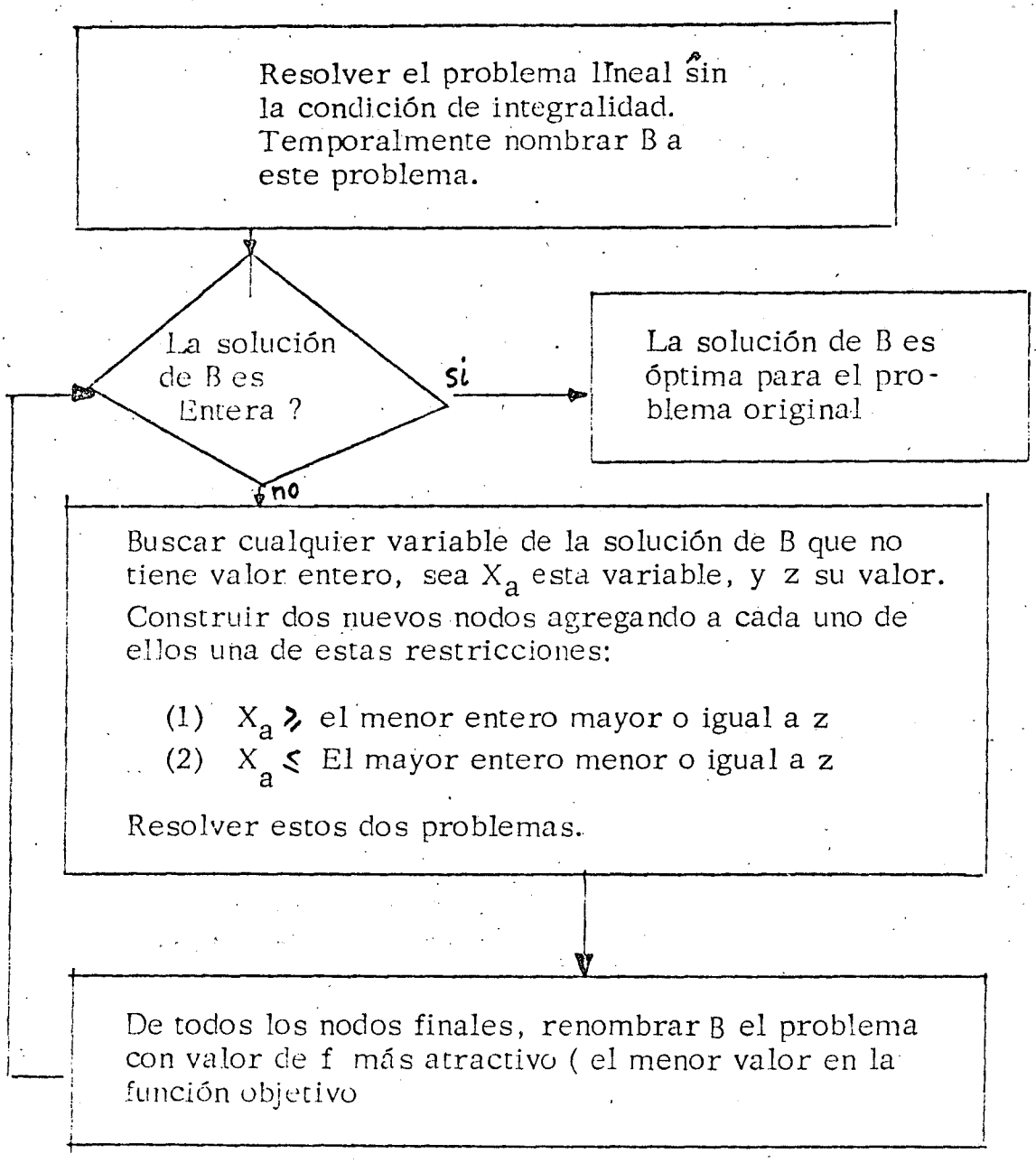
2. - Si dividimos la región factible de un problema de programación lineal con una restricción del tipo, ya sea $x_3 \leq 4$ ó $x_3 \geq 5$, reducimos la región factible del problema de programación lineal, pero se deja invariante la región factible del problema lineal entero correspondiente, pues no hay enteros entre 4 y 5. Ahora podemos incorporar este tipo de restricciónes, resolviendo dos problemas lineales, un problema con cada una de las restricciones agregada al previo conjunto de restricciones.

   Entonces seleccionaremos como solución, aquella con valor más atractivo en la función objetivo. Este procedimiento de agregar restricciones para formar un nuevo par de problemas se llama ramificación. Llamaremos al problema al final de cada rama, un nodo.

3. - El valor de la función objetivo en la solución del problema líneal de cada nodo es una cota para el valor de la función objetivo de los problemas que se deriven de ese nodo.

Habremos encontrado la solución al problema entero original cuando encontremos una solución entera cuyo valor en la función objetivo sea al menos tan atractivo como la cota en cada nodo terminal del árbol.

A continuación se da un diagrama del algoritmo de ramificación y acotamiento.

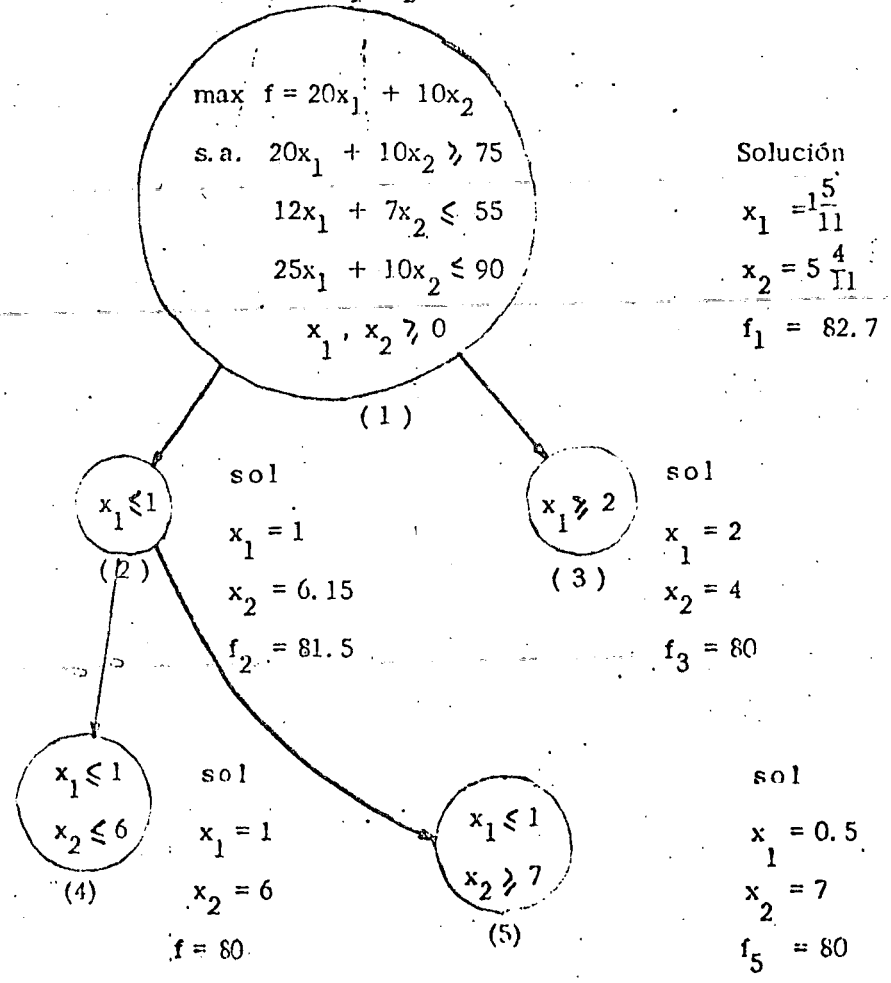

Resolver el problema lineal sin la condición de integralidad. Temporalmente nombrar B a este problema.
La solución de B es Entera ?
si
La solución de B es óptima para el problema original
no
Buscar cualquier variable de la solución de B que no tiene valor entero, sea X_a esta variable, y z su valor.
Construir dos nuevos nodos agregando a cada uno de ellos una de estas restricciones:
(1) $X_a \geq$ el menor entero mayor o igual a z
(2) $X_a \leq$ El mayor entero menor o igual a z
Resolver estos dos problemas.
De todos los nodos finales, renombrar B el problema con valor de f más atractivo (el menor valor en la función objetivo

Ejemplo;

$$\text{Max} \quad f = 20x_1 + 10x_2$$

Sujeto a

$$20x_1 + 10x_2 \geq 75$$
$$12x_1 + 7x_2 \leq 55$$
$$25x_1 + 10x_2 \leq 90$$
$$x_1, x_2 \geq 0$$
$$x_1, x_2 \text{ enteros}$$



Solución

Optima

REFERENCIAS BIBLIOGRAFICAS.

1. - INTEGER PROGRAMMING

R. S. Garfinkel and G. L. Neunhauser

John Wiley 1972

2. - DISCRETE OPTIMIZATION. INTEGER PROGRAMMING AND NETWORK ANALYSIS FOR MANAGEMENT DECISIONS

D. R. Plane and C. Mc. Millan, Jr.

Prentice-Hall 1971

3. - INTEGER PROGRAMMING AND NETWORK FLOWS

T. C. Hu

Addison Wesley 1969

3. - LINEAR AND INTEGER PROGRAMMING

S. Zionts

Prentice-Hall 1974

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA XV Y XVI DESPACHO ECONOMICO CON RESTRICCIONES

I Y II

ENERO, 1979.

Palacio de Minería Calle de Tacuba 5, primer piso. México 1, D. F.

N O T A S

SOBRE DESPACHO ECONOMICO

PARA EL CURSO INTENSIVO

"TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS"

Preparados por:

F. Aboytes
B. Vidrio
H. Torres

Sección de Graduados
ESIME-IPN

Enero 1979

ESTUDIO DE FLUJO OPTIMO

El estudio de flujos convencional se puede plantear matematicamente, como la solución de un sistema de ecuaciones, no-lineales.

$$g_i(\underset{\sim}{x},\underset{\sim}{y}) = 0 \qquad i = 1, n$$

donde $\underset{\sim}{x}$ vector de variables dependientes (estado)

$\underset{\sim}{y}$ vector de variables independientes

Una especificación típica de las variables es como sigue:

$$\begin{bmatrix} \underset{\sim}{x} \end{bmatrix} = \begin{bmatrix} V \\ \theta \end{bmatrix} \text{ en nodos } P,Q \quad ; \quad \begin{bmatrix} \theta \end{bmatrix} \text{ en nodos } P,V$$

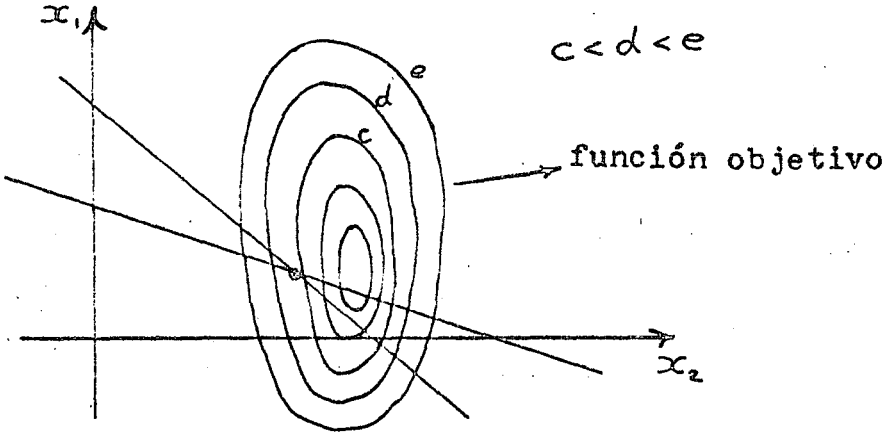
$$\begin{bmatrix} \underset{\sim}{y} \end{bmatrix} = \begin{bmatrix} V \\ \theta \\ P \\ Q \\ P \\ V \end{bmatrix} \begin{matrix} \text{nodo} \\ \text{slack} \\ \text{nodos } P,Q \\ \\ \text{nodos } P\ V \\ \end{matrix}$$

Para encontrar el vector $\underset{\sim}{x}$ se selecciona un número de ecuaciones de $\underset{\sim}{y}$ igual al número de incognitas.

$$\underset{\sim}{g}(\underset{\sim}{x}, \underset{\sim}{y}) = \begin{bmatrix} \text{ec. } P \\ \text{ec. } Q \\ \text{ec. } P \end{bmatrix} \begin{matrix} \text{en nodos} \\ P,Q \\ \text{en nodos } PV \end{matrix}$$

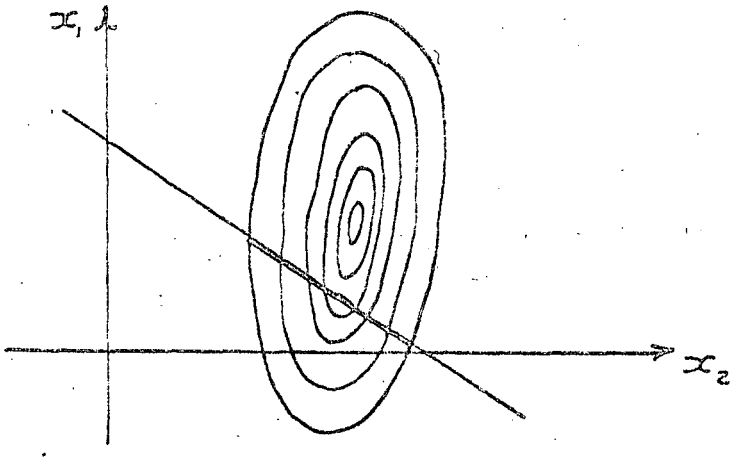
la característica importante es que las variables independientes tienen un valor fijo (potencias y voltajes especificados) durante el proceso de solución. Así, se tendrá una solución única.

En el caso de 2 variables se puede ilustrar como sigue:



En la figura anterior se observa que si se tiene una función objetivo a optimizar al existir solo un punto que cumple con las restricciones esa es la solución optima del problema. Es importante notar que las restricciones impiden escoger una mejor solución de acuerdo a la función objetivo, en este caso se dice que no se tienen grados de libertad.

Si solo se tuviera una restricción de igualdad en el problema, es decir solo una variable independiente tomará un valor fijo, entonces



Se tiene libertad para escoger el punto óptimo de acuerdo a la función objetivo, existe un grado de libertad. La linea doble en la figura ilustra el conjunto de puntos que satisfacen la restricción con diferentes valores de la función objetivo.

En el problema convencional de flujos al tener las variables independientes fijas se forza a tener una solución sin tomar en consideración ningún criterio económico o de funcionamiento.

En la práctica el problema de flujos tiene un conjunto de restricciones extras que hacen más dificil su solución ya que se requiere control sobre las variables independientes. Así se manejan:

1. límites de reactivos de generadores
2. cambio automático de taps en transformadores
3. intercambios de potencia activa entre areas
4. control de voltaje en nodos de carga

Una forma de resolver el problema es utilizar un procedimiento de prueba y error en el cual las variables independientes del problema se cambian entre iteraciones para tratar de satisfacer las restricciones impuestas. Generalmente los criterios que se usan para determinar el valor de las variables independientes son diseñados en forma aislada para cada restricción que se desea satisfacer, lo cual en algunos casos crea problemas de convergencia ya que existen conflictos entre las variables independientes que se manejan. Por otro lado es importante notar que en la solución de flujos ajustada lo que se desea es obtener una solución que cumpla con las restricciones impuestas sin tener en consideración si es la mejor en algún sentido.

El estudio de flujos óptimo permite formular el problema de flujos sujeto a un conjunto de restricciones. La solución que se obtiene no solo cumple con las restricciones impuestas sino que es optima en algún sentido. Para lograrlo se requiere que algunas de las variables independientes controlen el problema para obtener la mejor solución.

Funciones Objetivo

En la formulación del problema de flujos óptimos - algunas de las funciones objetivo más comunes son:

Costo de generación

Pérdidas mínimas

Costo compensación de reactivos

Factibilidad de flujos

Costo de corte de carga

Generación de reactivos

Errores de potencia nodales

La solución óptima se obtiene encontrando las variables de control que cumplan con las restricciones impuestas y que al mismo tiempo minimizen la función objetivo.

Una de las funciones más comunmente usadas en la operación y planeación de los sistemas eléctricos de potencia es el costo de generación y el problema de flujos óptimos correspondiente recibe el nombre de despacho económico.

Determinación de Curvas de Costo de Generación

Los componentes del costo de generación más importantes son:

Costo por combustible

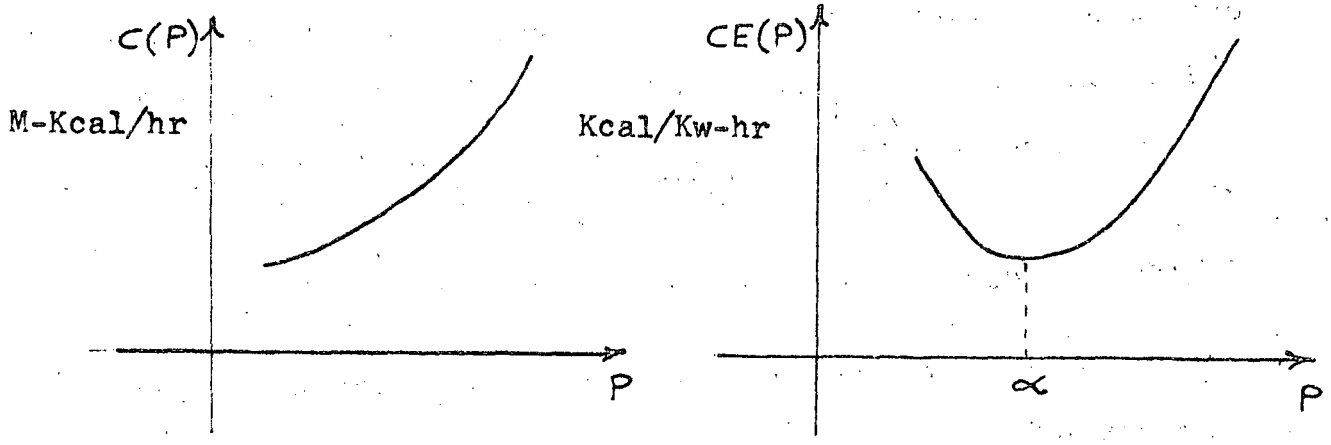
Costo de personal de operación

Una diferencia importante entre los costos anteriores es que el costo por combustible se ve afectado por el nivel de ge-

neración que se tenga, en cambio el costo de personal se mantiene casi independiente de la generación.

Por lo tanto se considera como un cargo fijo que no afecta la distribución de generación, una vez que se a decidido que máquinas están en servicio.

El costo de combustible se obtiene por medio de pruebas en las plantas generadoras. Existen dos curvas que se determinan en la práctica una que relaciona Potencia Térmica con Potencia Eléctrica (curva de entrada - salida) y otra que relaciona la eficiencia de la conversión de la energía con la potencia de salida (curva de consumo específico).



La relación entre las curvas en la siguiente

$$CE(P) = \frac{C(P)}{P}$$

Es decir un punto en la curva de consumo específico se obtiene dividiendo la potencia térmica (Kcal/hr) entre la potencia eléctrica correspondiente.

La determinación de los puntos de las curvas requiere de controlar el proceso térmico y permitir su estabilización. La curva de consumo específico tiene un mínimo a cuyo valor de generación (α) el proceso de conversión de energía es más eficiente.

Para obtener una función analítica de las curvas se asume un polinomio que ajuste adecuadamente los puntos obtenidos en las pruebas.

Tipicamente se utiliza una función cuadrática o cúbica para aproximar la curva de entrada salida. Para una curva cuadrática se tiene

$$C(P) = a P^2 + b P + c$$

y para el consumo específico

$$CE(P) = a P + b + c/P$$

Para determinar las constantes (a,b,c) se requieren por lo menos tres mediciones de consumo específico. Sin embargo en la práctica se requieren más valores para determinar en forma adecuada las constantes debido a los errores en las mediciones de consumo específico.

Para n mediciones se tienen las siguientes ecuaciones

$$\begin{bmatrix} CE_1 \\ CE_2 \\ \vdots \\ CE_n \end{bmatrix} = \begin{bmatrix} P_1 & 1 & \frac{1}{P_1} \\ P_2 & 1 & \frac{1}{P_2} \\ & & \\ P_n & 1 & \frac{1}{P_n} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \\ \varepsilon_n \end{bmatrix}$$

En algunos casos, con las mediciones anteriores se obtienen curvas con el punto de máxima eficiencia en un valor negativo o muy pequeño de potencia lo cual no es aceptable.

Para orientar la curva se requiere de una restricción extra. En el punto de máxima eficiencia se tiene un consumo específico mínimo

$$\left.\frac{dCE}{dP}\right|_{P_{ef}} = a - \frac{c}{P_{ef}^2} = 0$$

Generalmente se tiene

$$\alpha = \frac{P_{ef}}{P_{max}} = 85 - 100\%$$

de donde se obtiene la restricción

$$a \alpha P_{max} - \frac{c}{\alpha P_{max}} = 0$$

El conjunto de ecuaciones completo es

$$\begin{bmatrix} 0 \\ CE_1 \\ CE_2 \\ \vdots \\ CE_n \end{bmatrix} = \begin{bmatrix} \alpha P_{max} & 0 & -\frac{1}{\alpha P_{max}} \\ P_1 & 1 & \frac{1}{P_1} \\ P_2 & 1 & \frac{1}{P_2} \\ & & \\ P_n & 1 & \frac{1}{P_n} \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

En forma compacta la ecuación anterior se puede expresar como

$$\underset{\sim}{CE} = H \underset{\sim}{x} + \underset{\sim}{\varepsilon}$$

La suma de cuadrados de errores ponderados se puede expresar como

$$F = (\underset{\sim}{CE} - H \underset{\sim}{x})^t W(\underset{\sim}{CE} - H \underset{\sim}{x})$$

donde

W matriz diagonal de factores de ponderación

Si se minimiza F se obtiene una estimación de las constantes del polinomio

$$\underset{\sim}{x^*} = \begin{bmatrix} a^* \\ b^* \\ c^* \end{bmatrix} = (H^t WH)^{-1} H^t W \underset{\sim}{CE}$$

Es importante notar que el procedimiento de ajuste es general y en el caso de tener un polinomio cúbico se tendrán cuatro constantes por determinar.

Valores típicos de consumos específicos para plantas del país.

Unidad 1 Salamanca

Capacidad 158 MW

Generación %	Consumo Esp. Kcal./Kw-hr.	Consumo M-Kcal/hr
100	2442	385.7
87.5	2459	339.5
81.25	2473	316.7
75	2489	297.0
68.75	2512	273.9
62.5	2538	250.7
50	2630	208.3

Unidad 2 Tula

Capacidad 300 MW

Generación %	Consumo Esp. Kcal/Kw-hr.	Consumo M-Kcal/hr.
100	2427	728
87.5	2363	618.8
81.25	2349	575.2
75	2359	531.5
68.75	2395	495.1
62.5	2459	458.7
52.7	2657	419.8

Despacho económico simplificado

Este caso particular se caracteriza por no considerar la configuración del sistema de transmisión, sobre el cual se lleva a cabo el estudio.

Esto equivale a suponer que, tanto la generación como la carga se encuentran concentradas en un mismo nodo;

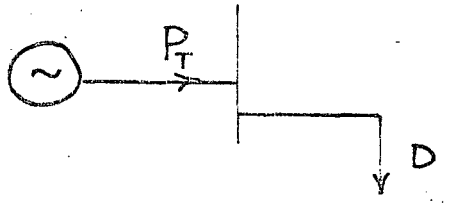


Figura 1

La consecuencia de esta suposición es no considerar flujos en líneas, con lo cual se excluye el efecto que las pérdidas ejercerían sobre los resultados del Estudio de Despacho económico, siendo ésta la causa de su simplificación.

Obviamente de la Fig. 1, se deduce que los costos usados, , estan referidos en terminales de generación.

La formulación matemática del problema de despacho económico, para el caso de 2 generadores alimentando una carga D, con funciones de costo de generación cuadráticas es como sigue;

minimizar

$$F = C_1 + b_1 P + a_1 P_1^2 + C_2 + b_2 P_2 + a_2 P_2^2$$

sujeto a:

$$P_1 + P_2 = D$$

Antes de analizar las posibles formas de solución se debe verificar la existencia de grados de libertad (no. de restricciones de igualdad < no. de incognitas). En éste caso existe 1 grado de libertad, ya que se tienen dos incógnitas (P_1 y P_2) y una sola restricción ($P_1 + P_2 = D$).

El algoritmo de solución es:

1. Despejar P_2 en función de P_1 ($P_2 = D - P_1$)
2. Substituir $P_2 = f(P_1)$ en la función objetivo.
3. Encontrar el mínimo de la función objetivo con respecto a P_1 y sin restricciones.
4. Encontrar P_2 con el valor de P_1 obtenido en el paso anterior.

En forma desarrollada se tiene

1. $P_2 = D - P_1$
2. $F = C_1 + b_1 P_1 + a_1 P_1^2 + C_2 + b_2 (D - P_1) + a_2 (D - P_1)^2$
3. $\frac{dF}{dP_1} = b_1 + 2a_1 P_1 + b_2 + 2a_2 (D - P_1)(-1) = 0$

de donde:

$$P_1 = \frac{b_2 - b_1 + 2a_2 D}{2(a_1 + a_2)}$$

4. $P_2 = D - P_1$

Este procedimiento aunque sencillo en el caso de 2 variables se complica cuando se tienen más variables y restricciones - al despejar variables en función de otras.

Ejemplos numéricos sencillos:

caso I:

Las funciones de costo de los generadores son

$$C_1 = \frac{1}{2} P_1^2 + 3.89 P_1 + 0.406$$

$$C_2 = \frac{1}{2} P_2^2 + 3.51 P_2 + 0.444$$

La demanda es D = 2.6

$$P_1 + P_2 = 2.6$$

$$P_2 = 2.6 - P_1$$

$$C_t = C_1 + \frac{1}{2}(2.6 - P_1)^2 + 3.5(2.6 - P_1) + 0.444$$

$$\frac{\partial C_t}{\partial P_1} = 2P_1 - 2.22 = 0$$

$$P_1 = 1.11$$
$$P_2 = 2.6 - 1.11 = 1.49$$

caso II.

Se tiene además la restricción

$$P_1 = 2P_2$$
$$P_1 + P_2 = 2.6$$

resolviendo simultáneamente:

$$2P_2 + P_2 = 2.6$$
$$P_2 = \frac{2.6}{3}$$
$$P_1 = 2(\frac{2.6}{3})$$

No se requiere evaluar derivadas por no existir grados de libertad.

Multiplicadores de Lagrange

El problema de optimización con restricciones se convierte en un problema sin restricciones incluyendo un multiplicador de Lagrange por cada restricción de igualdad.

La función de costo aumentada (Lagrangiano) se obtiene como

$$L = F + \lambda (P_1 + P_2 - D)$$

donde

λ es el multiplicador de Lagrange y es una variable más del problema.

La condición necesaria para que ésta función sea un mínimo es que el gradiente de la función con respecto a las variables sea cero.

$$\frac{\partial L}{\partial P_1} = 0, \quad \frac{\partial L}{\partial P_2} = 0, \quad \frac{\partial L}{\partial \lambda} = 0$$

Derivando:

$$\begin{aligned} \frac{\partial L}{\partial P_1} &= \frac{\partial F}{\partial P_1} + \lambda = 0 \\ \frac{\partial L}{\partial P_2} &= \frac{\partial F}{\partial P_2} = 0 \\ \frac{\partial L}{\partial \lambda} &= P_1 + P_2 - D = 0 \end{aligned} \tag{1}$$

Se puede observar que $\frac{\partial F}{\partial P_1} = \frac{dF_1}{dP_1}$ ya que la variable P_1 solo está contenida en F_1; y lo mismo se puede decir para la variable P_2, por lo que $\frac{\partial F}{\partial P_2} = \frac{dF_2}{dP_2}$

Entonces

$$\frac{\partial L}{\partial P_1} = \frac{dF_1}{dP_1} + \lambda = 0 \tag{2}$$

$$\frac{\partial L}{\partial P_2} = \frac{dF_2}{dP_2} + \lambda = 0 \tag{3}$$

de donde se deduce que

$$\frac{dF_1}{dP_1} = -\lambda \tag{4}$$

$$\frac{dF_2}{dP_2} = -\lambda \tag{5}$$

La $\frac{dF_i}{dP_i}$ se conoce como costo incremental del generador i, y representa el incremento de costo debido a un incremento en la potencia generada.

Utilizando el método de los multiplicadores de Lagrange el problema consiste en la solución de un sistema de ecuaciones simultáneas, el cual se logra en forma directa o en forma secuencial, según sea el grado de las funciones de costo de los generadores.

Funciones de costo lineales de la forma bi Pi + Ci

Del conjunto de ecuaciones (1) se tiene que

$$\frac{\partial L}{\partial Pi} = bi + \lambda = 0$$
$$\frac{\partial L}{\partial \lambda} = \sum Pi - D = 0 \qquad (6)$$

El conjunto de ecuaciones (6) no se satisfacen simultáneamente, de donde surge la necesidad de considerar límites en las potencias de generación ($P_{i\,min} \leq P_i \leq P_{i\,max}$) debido a que los costos incrementales de cada generador son en general diferentes.

En éste caso todos los generadores se cargan inicialmente a su límite mínimo y secuencialmente los de menor costo incremental se cargan a su límite máximo.

El algoritmo de solución para este caso es:

1. Ordenar los coeficientes bi de menor a mayor
2. Cargar todos los generadores en su límite mínimo
3. Cargar los generadores en su límite máximo de acuerdo al orden del inciso 1, hasta satisfacer $\sum Pi = D$

Nota. Obviamente al último generador se le asignará solo la potencia necesaria para satisfacer $\sum P_i = D$.

Ejemplo numérico

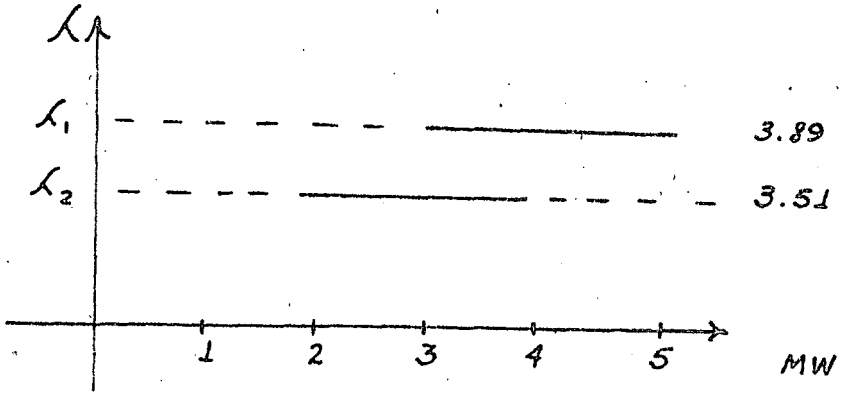
En el caso de dos generadores cuyas funciones de costo son

$$C_1 = 3.89\ P_1 + 0.406$$
$$C_2 = 3.51\ P_2 + 0.444$$

con límites

$$3 \leq P_1 \leq 5$$
$$2 \leq P_2 \leq 4$$

y demanda $D = 8$



Paso 1

	b_i
2	3.51
1	3.89

Paso 2

$$P_1 = 3$$
$$P_2 = 2$$

Paso 3

$$P_2 = 4$$
$$P_1 = D - P_2 = 8-4 = 4$$

Funciones de costo cuadráticas de la forma $a_i P_i^2+b_iP_i+C_i$

Se observa que en este caso el conjunto de ecuaciones (1) queda

$$\frac{\partial L}{\partial P_i} = 2 a_i P_i + b_i + \lambda = 0 \qquad (7)$$
$$\frac{\partial L}{\partial \lambda} = \sum P_i - D = 0$$

La forma directa de resolver el conjunto de ecuaciones simultáneas (7) para los P_i y λ (que es una variable más) incluyendo además límites, se logra mediante el siguiente algoritmo.

1. Escribir el conjunto de ecuaciones (7) en forma matricial y resolverlo.

$$\begin{bmatrix} 2a_1 & \cdots & 0 & \cdots & 0 & 1 \\ \vdots & & \vdots & & \vdots & \\ 0 & \cdots & 2a_i & & \vdots & 1 \\ \vdots & & & & \vdots & \\ 0 & \cdots & 0 & \cdots & 2a_n & 1 \\ 1 & & 1 & & 1 & 0 \end{bmatrix} \begin{bmatrix} P_i \\ \hline \lambda \end{bmatrix} = \begin{bmatrix} -b_i \\ \hline D \end{bmatrix}$$

2. Verificar límites

Si existen violaciones, sustituir la ecuación de costo incremental por la del límite violado, en la siguiente forma:

$$\begin{bmatrix} 2a_1 & 0 & 0 & 0 & 1 \\ 0 & 2a_2 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & & 2a_n & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ P_i \\ P_n \\ \lambda \end{bmatrix} = \begin{bmatrix} -b_i \\ P_i \text{ violada} \\ -b_n \\ D \end{bmatrix}$$

Resolver éste sistema y repetir el paso 2 hasta que no existan violaciones de límites.

Ejemplo

Para el caso de dos generadores cuyas funciones de costo son:

$$C_1 = \frac{1}{2} P_1^2 + 3.89\ P_1 + 0.406$$

$$C_2 = \frac{1}{2} P_2^2 + 3.51\ P_2 + 0.444$$

y con demanda

$$D = 8$$

y límites

$$3 \leq P_1 \leq 5$$

$$2 \leq P_2 \leq 4$$

Paso 1

$$\frac{\partial L}{\partial P_1} = P_1 + 3.89 + \lambda = 0$$

$$\frac{\partial L}{\partial P_2} = P_2 + 3.51 + \lambda = 0$$

$$P_1 + P_2 = 8$$

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} -3.89 \\ -3.51 \\ 8 \end{bmatrix}$$

La solución de éste sistema es

$$\lambda = -7.7$$

$$P_1 = 3.81$$

$$P_2 = 4.19$$

Paso 2

De acuerdo a los resultados del paso 1, el límite má-

ximo de P_2 se ha violado, por lo que el sistema de ecuaciones queda:

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \lambda \end{bmatrix} = \begin{bmatrix} -3.89 \\ 4 \\ 8 \end{bmatrix}$$

Para este sistema de ecuaciones la solución es

$$\begin{aligned} \lambda &= -7.89 \\ P_1 &= 4 \\ P_2 &= 4 \end{aligned}$$

Funciones de costo cúbicas de la forma $a_iP_i^3+b_iP_i^2+C_iP_i+d$

En éste caso el cunjunto de ecuaciones (1) queda

$$\begin{aligned} \frac{\partial L}{\partial P_i} &= 3a_iP_i^2 + 2b_iP_i + C_i + \lambda = 0 \\ \frac{\partial L}{\partial \lambda} &= \sum P_i - D = 0 \end{aligned} \qquad (8)$$

Para resolver éste sistema de ecuaciones cuadráticas se usa el método de Newton.

$$\begin{bmatrix} J \end{bmatrix} \begin{bmatrix} \Delta x \end{bmatrix} = - \begin{bmatrix} g \end{bmatrix}$$

donde

$\begin{bmatrix} J \end{bmatrix}$ es el Jacobiano de (8)

$\begin{bmatrix} \Delta x \end{bmatrix}$ es el vector de incógnitas

$\begin{bmatrix} g \end{bmatrix}$ es el vector independiente, cuyos elementos son

en un principio el conjunto (8) evaluado para condiciones iniciales.

El algoritmo de solución es:

1. Seleccionar valores iniciales P_i°, λ°
2. Calcular los elementos de $[J]$ y $[g]$
3. Formar el siguiente sistema de ecuaciones

$$\begin{bmatrix} \frac{\partial^2 C_i}{P_i^2} & & & 1 \\ & \ddots & & \vdots \\ & & \frac{\partial^2 C_n}{\partial P_n^2} & 1 \\ & & & \vdots \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \Delta P_1 \\ \vdots \\ \Delta P_n \\ \vdots \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} \lambda^\circ & & \frac{\partial C_i}{\partial P_i}\Big|^\circ \\ & \vdots & \\ \sum P_i^\circ & - & D \end{bmatrix}$$

y resolver para $\Delta P_i, \Delta\lambda$

4. Corregir los valores de P_i, λ
5. Continuar iterativamente con los pasos 3 y 4 hasta que $\Delta P_i, \Delta\lambda$ sea menor o igual que una tolerancia especificada.

Despacho económico con Pérdidas

La inclusión de las pérdidas en el estudio de Despacho económico de potencia activa (distribución óptima de generación) da lugar a formulaciones que incluyen su efecto,

unas veces en forma aproximada y otros en forma exacta.

Se describen aquí los siguientes métodos:

I. fórmula de pérdidas (constantes B)

II. solución de las ecuaciones de coordinación utilizando el método desacoplado en el cálculo de pérdidas incrementales. (solución explícita)

III. solución implicita (gradiente)

IV. programación lineal

V. método de la matriz Hessiana

I. Fórmula de Pérdidas

Determinación de constantes B. Las pérdidas son una función de las potencias de generación $P_L = f(P_i)$

Expansión en serie de Taylor de las pérdidas

$$P_L = P_L^{\circ} + \sum_i \frac{\partial P_L}{\partial P_i} \Delta P_i + \frac{1}{2} \sum_i \sum_j \Delta P_i \frac{\partial^2 P_L}{\partial P_i \partial P_j} \Delta P_j \qquad (9)$$

Las incógnitas son los coeficientes

$$\frac{\partial P_L}{\partial P_i}, \quad \frac{\partial^2 P_L}{\partial P_i \partial P_j}$$

Para determinarlos se requiere por lo menos el mismo número de ecuaciones.

Algoritmo:

1. Partiendo de un caso base de flujos se determina P_L^o

2. Se calculan casos de flujos para diferentes incrementos de los P_i (ΔP_i) alrededor del caso base P_i^o; su número no debe ser menor al número de incógnitas.

3. Se calculan las pérdidas totales P_{L_i} para cada caso, y se forma el sistema de ecuaciones

casos de flujos

$$\left\{\left[\begin{array}{c|cc} \Delta P & \Delta P & \Delta P^t \end{array}\right]\right. \left[\begin{array}{c} \frac{\partial P_L}{\partial P_i} \\ \\ \frac{\partial^2 P_L}{\partial P_i \partial P_j} \end{array}\right] = \left[\begin{array}{c} P_{L_i} - P_L^o \end{array}\right]$$

transformación lineal de la forma $A\underset{\sim}{x} = \underset{\sim}{b}$

4. Utilizando la técnica de mínimos cuadrados obtener un valor estimado para las incógnitas ($\underset{\sim}{x}$) del inciso anterior.

minimizar el error:

$$F = (A\underset{\sim}{x} - \underset{\sim}{b})^t (A\underset{\sim}{x} - \underset{\sim}{b})$$

solución

$$A^t A \underset{\sim}{x} = A^t \underset{\sim}{b}$$

5. Se determinan las constantes B como sigue:

$$P_L = B_o + \sum_i B_i P_i + \sum_i \sum_j B_{ij} P_i P_j \qquad (10)$$

comparando la ecuación (9 y 10)

$$B_o = P_{L_o} - \sum_i P_i^o \frac{\partial P_L}{\partial P_i} + \sum_i \sum_j P_i^o \frac{\partial^2 P_L}{\partial P_i \partial P_j} P_j^o$$

$$B_i = \frac{\partial P_L}{\partial P_i} - 2 \sum_j P_j^o \frac{\partial^2 P_L}{\partial P_i \partial P_j}$$

$$B_{ij} = \frac{\partial^2 P_L}{\partial P_i \partial P_j}$$

en forma matricial

$$P_L = B_o + B_G^t \quad P + P^t \; B \; P$$

Desventajas del cálculo de pérdidas con constantes B :

1. Se obtienen para condiciones particulares y es difícil su generalización
2. Se requieren diferentes grupos de constantes para diversas condiciones de operación.
3. Su inexactitud aumenta para condiciones de operación no comprendidas en su cálculo.

Nota. Se toma en todos los cálculos la carga equivalente del sistema como variable dependiente,

Aplicación de las constantes B

$$\text{Min.} \sum C_i(P_i)$$

sujeto a:

$$\sum P_i = D + P_L(P_i)$$

$$P_i^{min} \leq P_i \leq P_i^{max}$$

Utilizando multiplicadores de Lagrange

$$L(P_i, \lambda) = \sum C_i(P_i) + \lambda (D + P_L(P_i) - \sum P_i)$$

Las condiciones necesarias para la existencia de un óptimo son:

$$\frac{\partial L}{\partial P_i} = \frac{dC_i}{dP_i} + \lambda \left(\frac{\partial P_L}{\partial P_i} - 1 \right) = 0 \qquad 11a$$

(11)

$$\frac{\partial L}{\partial \lambda} = (D + P_L(P_i) - \sum P_i) = 0 \qquad 11b$$

El conjunto de ecuaciones 11a son ecuaciones de coordinación.

La derivada parcial $\frac{\partial P_L}{\partial P_i}$ son las pérdidas incrementales; significa el incremento de las pérdidas dado un incremento de P_i.

$$FP_i = \left(\frac{1}{1 - \frac{\partial P_L}{\partial P_i}} \right)$$

$$\left(\frac{dC_i}{dP_i} \right) FP_i = \lambda \qquad i \neq nc$$

$$\frac{dC_{nc}}{dP_{nc}} = \lambda$$

Se obtiene el óptimo con costos incrementales iguales que a diferencia del Despacho Económico simplificado no se obtienen en terminales del generador, sino en un punto común a donde se refiere la carga equivalente (nodo compensador). Debido a la derivada $\frac{\partial P_L}{\partial P_i} = B_i + 2 \sum B_{ij}P_j$ el conjunto de ecuaciones 11 será no-lineal.

Nota. Si $\frac{\partial P_L}{\partial P_i} = 0$ i=1,NG el conjunto de ecuaciones se reduce al conjunto 1 (caso sin pérdidas).

En adelante se hará notar la similitud que existe entre el multiplicador de Lagrange de cada tipo de formulación.

Ejemplo numérico

Las características de costo incremental de producción de dos unidades generadoras de vapor están dadas por

$$\frac{dC_1}{dP_1} = P_1 + 2$$

$$\frac{dC_2}{dP_2} = P_2 + 1.5$$

donde P está expresada en por unidad a una base de 100 MVA. Los límites de operación máximo y mínimo para ambas unidades son 20 MW y 100MW respectivamente. Los coeficientes de la fórmula de pérdidas en por unidad a una base de 100MVA están

dados por:

$$[B] = \begin{bmatrix} 0.10 & -0.05 \\ -0.05 & 0.20 \end{bmatrix}$$

a) Calcule la distribución óptima de generación para satisfacer una demanda de 112.11MW dentro de una tolerancia de 1 MW.

b) Calcule las pérdidas por transmisión P_L

Solución

$$P_L = \begin{bmatrix} P_1 & P_2 \end{bmatrix} \begin{bmatrix} 0.10 & -0.05 \\ -0.05 & 0.20 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

$$= \begin{bmatrix} P_1 & P_2 \end{bmatrix} \begin{bmatrix} 0.10P_1 & -0.05P_2 \\ -0.05P_1 & +0.20P_2 \end{bmatrix}$$

$$= P_1(0.10P_1 - 0.05P_2) + P_2(-0.05P_1 + 0.20P_2)$$
$$= 0.10P_1^2 - 0.05P_1P_2 - 0.05P_1P_2 + 0.20P_2^2$$
$$= 0.10P_1^2 - 0.10P_1P_2 + 0.20P_2^2$$

$$\frac{\partial P_L}{\partial P_1} = 0.2\ P_1 - 0.1\ P_2$$

$$\frac{\partial P_L}{\partial P_2} = 0.4P_2 - 0.1\ P_1$$

La función de costo aumentada es

$$L = C_1 + C_2 + \lambda(D + P_L - P_1 - P_2)$$

Las ecuaciones de coordinación son:

$$\frac{\partial L}{\partial P_1} = \frac{dC_1}{dP_1} + \lambda\left(\frac{\partial P_L}{\partial P_1} - 1\right)$$

$$= P_1 + 2 + \lambda(0.2P_1 - 0.1P_2 - 1) = 0$$

$$\frac{\partial L}{\partial P_2} = \frac{dC_2}{dP_2} + \lambda\left(\frac{\partial P_L}{\partial P_2} - 1\right)$$

$$= P_2 + 1.5 + \lambda(0.4P_2 - 0.1P_1 - 1 = 0$$

$$\frac{\partial L}{\partial \lambda} = D + P_L - P_1 - P_2 = 1.12 + (0.1P_1^2 - 0.1P_1P_2 + 0.2P_2^2) - P_1 - P_2 = 0$$

Para resolver éste sistema de ecuaciones cuadráticas se emplea el método de Newton $[J][\Delta x] = -[g]$ donde los elementos de $[J]$ son

$$J(1,1) = \frac{\partial^2 L}{\partial P_1^2} = 1 + 0.2\lambda$$

$$J(1,2) = \frac{\partial^2 L}{\partial P_1 \partial P_2} = -0.1\lambda$$

$$J(1,3) = \frac{\partial^2 L}{\partial P_1 \partial \lambda} = 0.2P_1 - 0.1P_2 - 1$$

$$J(2,1) = J(1,2)$$

$$J(2,2) = \frac{\partial^2 L}{\partial P_2^2} = 1 + 0.4\lambda$$

$$J(2,3) = \frac{\partial^2 L}{\partial P_2 \partial \lambda} = 0.4P_2 - 0.1P_1 - 1$$

$$J(3,1) = J(1,3)$$

$$J(3,2) = J(2,3)$$

$$J(3,3) = \frac{\partial^2 L}{\partial \lambda^2} = 0$$

y los elementos del vector independiente $[g]$ son

$$g(1) = P_1(1 + 0.2\lambda) - 0.1\lambda P_2 - \lambda + 2$$

$$g(2) = P_2(1 + 0.4\lambda) - 0.1\lambda P_1 - \lambda + 1.5$$

$$g(3) = 0.1P_1^2 + 0.2P_2^2 - 0.1P_1P_2 - P_1 - P_2 + 1.12$$

El sistema de ecuaciones por resolver es:

$$\begin{bmatrix} J_{11} & J_{12} & J_{13} \\ J_{21} & J_{22} & J_{23} \\ J_{31} & J_{32} & J_{33} \end{bmatrix} \begin{bmatrix} \Delta P_1 \\ \Delta P_2 \\ \Delta \lambda \end{bmatrix} = - \begin{bmatrix} g_1 \\ g_2 \\ g_3 \end{bmatrix}$$

Seleccionando valores iniciales de $P_1^o = .5$, $P_2^o = .5$, $\lambda^\circ = 2.5$ el sistema de ecuaciones queda

$$\begin{bmatrix} 1.5 & -.25 & -.95 \\ -.25 & .2 & -.85 \\ -.95 & -.85 & 0. \end{bmatrix} \begin{bmatrix} \Delta P_1 \\ \Delta P_2 \\ \Delta \lambda \end{bmatrix} = \begin{bmatrix} -.125 \\ .125 \\ -.1711 \end{bmatrix}$$

de donde

$$\Delta P_1^o = .0517$$

$$\Delta P_2^o = .1435$$

$$\Delta \lambda^\circ = .1754$$

Los valores corregidos son

$$P_1^1 = P_1^o + \Delta P_1^o = .5517$$

$$P_2^1 = P_2^o + \Delta P_2^o = .6435$$

$$\lambda' = \lambda^\circ + \Delta\lambda^\circ = 2.67$$

Con estos nuevos valores se evalúan $[J]$ y $[g]$ y el sistema por resolver en ésta iteración es:

$$\begin{bmatrix} 1.535 & -.2675 & -.954 \\ -.2675 & .2070 & -.7977 \\ -.9540 & -.7977 & 0 \end{bmatrix} \begin{bmatrix} \Delta P_1 \\ \Delta P_2 \\ \Delta\lambda \end{bmatrix} = \begin{bmatrix} .7 \times 10^{-3} \\ -.91\times10^{-2} \\ -.36\times10^{-2} \end{bmatrix}$$

La solución de éste sistema es

$$\Delta P_1^1 = .47 \times 10^{-2}$$
$$\Delta P_2^1 = -.10 \times 10^{-2}$$
$$\Delta\lambda' = .71 \times 10^{-2}$$

de donde

$$P_1^2 = .5564$$
$$P_2^2 = .6424$$
$$\lambda^2 = 2.68$$

el gradiente valuado en este punto es muy pequeño con lo cual se tiene la solución, todos los componentes son menores que 10^{-5}.

Despacho Económico de Potencia Activa

II. Método exacto (solución explicita)

El problema de Despacho Económico se formula como sigue

$$\text{Min} \sum_{i} C_i(P_i)$$

Sujeto a:

$$\sum_{i} P_i = D + P_L(P_i)$$

$$P_i^{min} \leq P_i \leq P_i^{max}$$

Utilizando multiplicadores de Lagrange

$$L(P_i, \lambda) = \sum_{i} C_i(P_i) + \lambda(D + P_L(P_i) - \sum_{i} P_i)$$

Las condiciones necesarias para la obtención de un óptimo son:

$$\frac{\partial L}{\partial P_i} = \frac{dC_i}{dP_i} + \lambda\left(\frac{\partial P_L}{\partial P_i} - 1\right) = 0$$

$$\frac{\partial L}{\partial \lambda} = (D + P_L - \sum_{i} P_i) = 0 \qquad (12)$$

Este método utiliza un algoritmo que difiere del anterior (constantes B), en los siguientes puntos:

1) Calcula flujos en cada ciclo, manteniendo el balance nodal al cumplir con las ecuaciones de la red $g(\underset{\sim}{x}, \underset{\sim}{u}) = 0$

$\underset{\sim}{x}$ variables de estado

$\underset{\sim}{u}$ potencias de generación (P_i)

2) Calcula en forma exacta P_L (pérdidas) y $\frac{\partial P_L}{\partial P_i}$ (pérdidas incrementales) en cada ciclo

3) El conjunto de ecuaciones (12) es lineal si las C_i son cuadráticas y es posible obtener una solución simultánea para P_i y λ ; si las funciones de costo C_i son de orden > 2 existe la necesidad de linealizar el conjunto 12 de la misma manera que 11, solo que las pérdidas en este caso, se mantendrán constantes.

4) Las cargas permanecen en su situación real, tomándose una de las generaciones como dependiente del resto (nodo compensador).

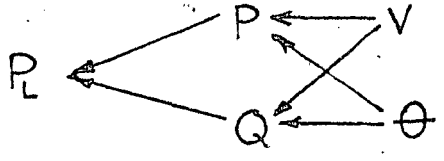
Nota: Si se excluyen las pérdidas y por lo tanto $\frac{\partial P_L}{\partial P_i}$ (pérdidas incrementales), el conjunto de ecuaciones 12, resulta el Despacho Económico simplificado.

Se derivan expresiones para el cálculo de las pérdidas incrementales

como $P_L = f(P_1, P_2, \ldots P_n, Q_1, Q_2, \ldots Q_n)$

y $P = f(V, \theta)$, $Q = f(V, \theta)$

por la règla de la cadena



en forma compacta

$$\begin{bmatrix} \frac{\partial P_L}{\partial \theta} \\ \frac{\partial P_L V}{\partial V} \end{bmatrix} = \begin{bmatrix} H & N \\ J & L \end{bmatrix}^t \begin{bmatrix} \frac{\partial P_L}{\partial P} \\ \frac{\partial P_L}{\partial Q} \end{bmatrix} \tag{13}$$

Utilizando el principio de desacoplamiento (relación x/r) grande

$$\frac{\partial P_L}{\partial \theta} = H^t \frac{\partial P_L}{\partial P} \tag{14}$$

$$\frac{\partial P_L}{\partial V} V = L^t \frac{\partial P_L}{\partial Q} \tag{15}$$

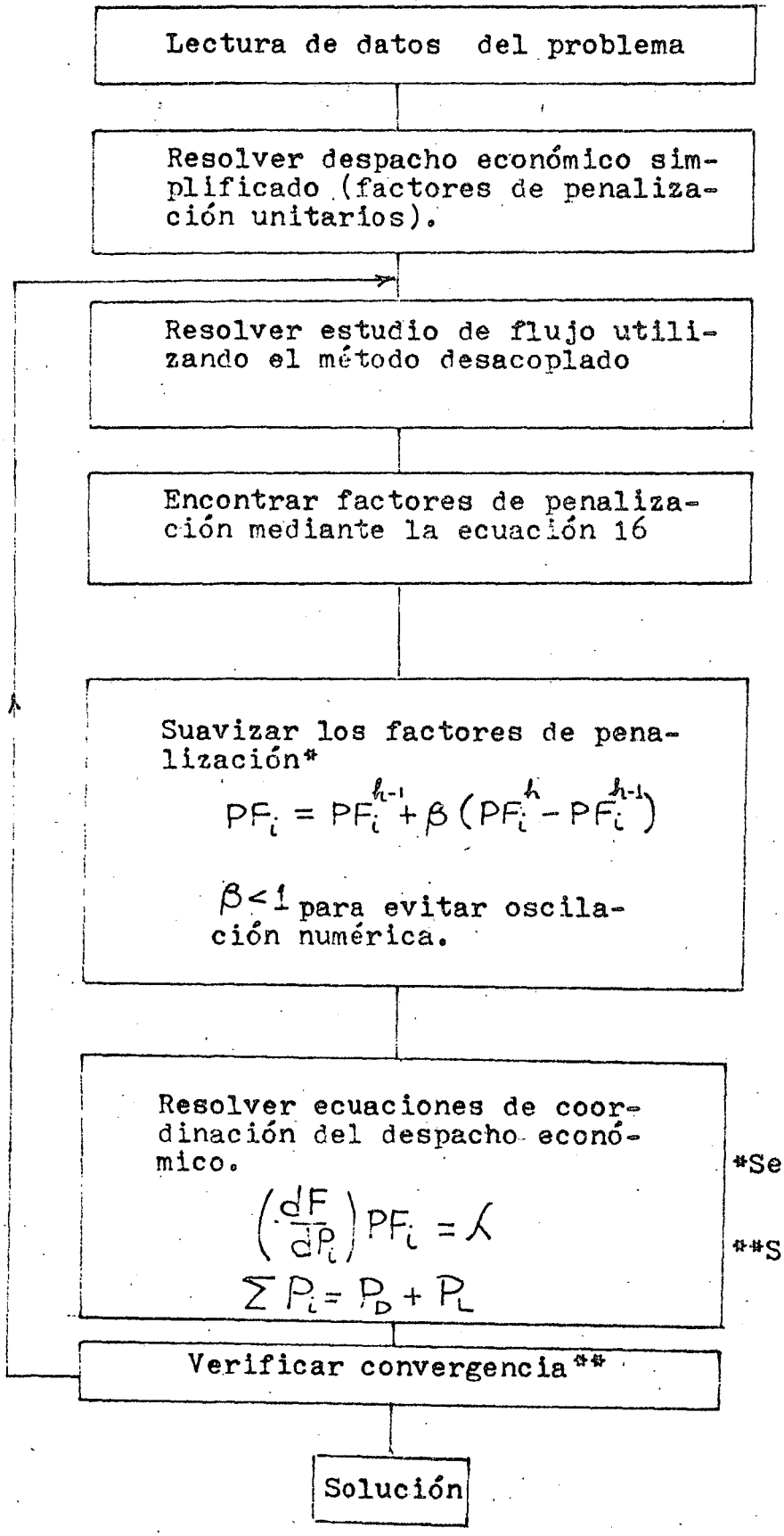
Simplificando 14 con base al algoritmo del método desacoplado para solución de flujos.

$$\left[\frac{\partial P_L}{\partial \theta}\right][V]^{-1} = [B]\left[\frac{\partial P_L}{\partial P}\right] \tag{16}$$

El término independiente (vector) de la ecuación 16 se evalúa con la fórmula siguiente:

$$\frac{\partial P_L}{\partial \theta_i} \frac{1}{V_i} = 2 \sum_{j \in i} Sen(\theta_i - \theta_j) G_{ij} \tag{17}$$

DIAGRAMA DE FLUJO DEL ALGORITMO



*Se usó $\beta = 0.6$

**Se usó $\varepsilon = .005$

EJEMPLO ILUSTRATIVO

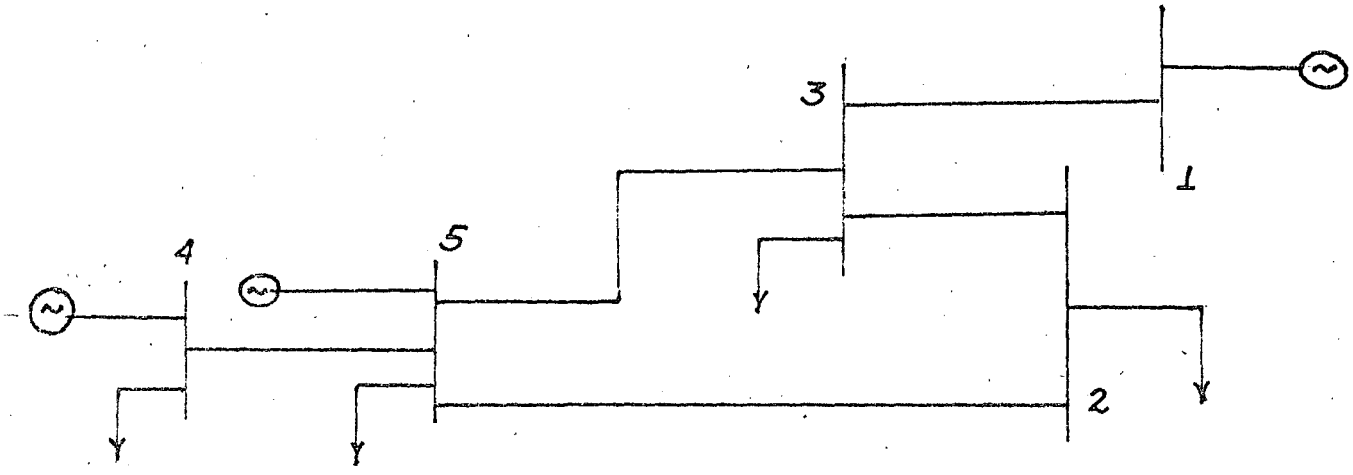


Figura 2

Sistema de prueba

Ecuaciones de costo utilizadas

Generador	1	$50P_1^2$	+	351.	P_1	+ 44.0
Generador	4	$50P_4^2$	+	245.0	P_4	+ 105.0
Generador	5	$50P_5^2$	+	389.	P_5	+ 40.6

DATOS DE LINEAS

BUS	BUS	IMPEDANCIA
4	5	0.030+j0.103
2	5	0.080+j0.262
3	5	0.105+j0.347
2	3	0.033+j0.118
3	1	0.106+j0.403

DATOS DE CARGAS

BUS	MW	MVARS
1	0	0
2	30	12
3	70	3
4	86	20
5	80	10

Se muestra el algoritmo a seguir y tablas de resultados de cada una de las iteraciones. En la Tabla I, se listan los resultados cuando se modelan como nodos (P,V) 1, 4 y 5, y como nodos (P,Q) 2 y 3. En la Tabla II, todos los nodos son (P,V).

T A B L A I

Resultados del ejemplo de la Fig. 2 (MW)

	ITERACION			
BUS	1	2	3	4
1	53.0	53.9	53.10	53.10
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	176.4	172.2	171.7	171.4
5	44.2	46.0	47.2	47.5
COSTO TOT DE GEN. $	1,159.2	1,153.55	1,153.6	1,153.3

MAGNITUD VOLTAJES p.u.

BUS	1	2	3	4
1	1.05	1.05	1.05	1.05
2	0.9845	0.9865	0.9863	0.9864
3	0.9853	0.9887	0.9884	0.9885
4	1.05	1.05	1.05	1.05
5	1.04	1.04	1.04	1.04

FACTORES DE PENALIZACION

Generador	1	2	3	4
1	1.0	1.0	1.0	1.0
4	0.9586	0.9704	0.9697	0.9707
5	0.9324	0.9307	0.9264	0.9259

T A B L A II

RESULTADOS DEL EJEMPLO DE LA FIGURA 2 (MW)

BUS	1	2	3	4
1	54.1	54.7	54.1	54.1
2	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0
4	175.6	171.7	171.2	170.8
5	43.5	45.5	46.6	46.9
COSTO TOT. DE GEN. $	1,157.2	1.152.4	1,152.4	1,152.15

FACTORES DE PENALIZACION

GENERADOR	1	2	3	4
1	1.0	1.0	1.0	1.0
4	0.963	0.973	0.973	0.974
5	0.936	0.933	0.930	0.929

VECTOR DE MAGNITUD DE VOLTAJE (p.u.)

BUS	MAGNITUD
1	1.05
2	1.0153
3	1.0158
4	1.05
5	1.04

III. Método implícito

En éste método la función de costo a minimizar es $f(\underset{\sim}{x},\underset{\sim}{u})$, donde:

$x_i = \theta_i$, $i \neq$ N.C. son las variables de estado

$u_i = P_i$, $i \neq$ N.C. son variables de control.

$g_i(\underset{\sim}{x}, \underset{\sim}{u})$ son las restricciones de igualdad. (ecuaciones de red).

La función de costo aumentada o Lagrangiano es

$$L = f(\underset{\sim}{x}, \underset{\sim}{u}) + \underset{\sim}{\lambda}^t \underset{\sim}{g}(\underset{\sim}{x}, \underset{\sim}{u})$$

Las condiciones necesarias para la existencia de un óptimo (mínimo) son:

$$\frac{\partial L}{\partial \underset{\sim}{x}} = \frac{\partial f(\underset{\sim}{x},\underset{\sim}{u})}{\partial \underset{\sim}{x}} + \left(\frac{\partial g(\underset{\sim}{x},\underset{\sim}{u})}{\partial \underset{\sim}{x}}\right)^t \underset{\sim}{\lambda} = 0 \qquad (18)$$

$$\frac{\partial L}{\partial \underset{\sim}{u}} = \frac{\partial f(\underset{\sim}{x},\underset{\sim}{u})}{\partial \underset{\sim}{u}} + \left(\frac{\partial g(\underset{\sim}{x},\underset{\sim}{u})}{\partial \underset{\sim}{u}}\right)^t \underset{\sim}{\lambda} = 0 \qquad (19)$$

$$\frac{\partial L}{\partial \lambda} = g(\underset{\sim}{x},\underset{\sim}{u}) = 0 \qquad (20)$$

donde

$\frac{\partial g(\underset{\sim}{x},\underset{\sim}{u})}{\partial x} = J$ es un Jacobiano.

De la ecuación (18), $\left(\frac{\partial g(x,u)}{\partial x}\right)^t \lambda = -\frac{\partial f(x,u)}{\partial x}$

la cual en forma extendida queda

$$\begin{bmatrix} \frac{\partial P_2}{\partial \theta_2} & \frac{\partial P_3}{\partial \theta_2} & \cdots & \frac{\partial P_n}{\partial \theta_2} \\ \vdots & & & \\ \frac{\partial P_L}{\partial \theta_n} & \frac{\partial P_3}{\partial \theta_n} & \cdots & \frac{\partial P_n}{\partial \theta_n} \end{bmatrix} \begin{bmatrix} \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = - \begin{bmatrix} \frac{\partial f}{\partial \theta_2} \\ \vdots \\ \frac{\partial f}{\partial \theta_n} \end{bmatrix} \qquad (21)$$

de la ecuación (21) se deduce que las unidades dimensionales de λ deben ser:

$$\underset{\sim}{\lambda} = - \begin{bmatrix} \frac{\partial f_1}{\partial P_2} \\ \vdots \\ \frac{\partial f_1}{\partial P_n} \end{bmatrix} = \frac{\partial f_1}{\partial P_1} \begin{bmatrix} -\frac{\partial P_1}{\partial P_2} \\ \vdots \\ -\frac{\partial P_1}{\partial P_n} \end{bmatrix} \qquad (22)$$

Por otro lado se sabe que en todo sistema se debe cumplir el balance de potencia

$$\sum_i P_i - D - P_L = 0, \quad i = 1,N$$

Tomando una generación como variable dependiente (nodo compensador $i=1$)

$$P_1 + \sum_i P_i - D - P_L = 0, \quad i \neq 1$$

de donde

$$P_1 = P_L + D - \sum_i P_i \qquad i \neq 1$$

Derivando con respecto a las variables independientes:

$$\frac{\partial P_1}{\partial P_i} = \frac{\partial P_L}{\partial P_i} - 1 , \quad i \neq 1$$

$$- \frac{\partial P_1}{\partial P_i} = 1 - \frac{\partial P_L}{\partial P_i} , \quad i \neq 1 \qquad (23)$$

sustituyendo (23) en (22) se tiene:

$$\underset{\sim}{\lambda} = \frac{df_1}{dP_1} \begin{bmatrix} 1 - \frac{\partial P_L}{\partial P_2} \\ \vdots \\ 1 - \frac{\partial P_L}{\partial P_n} \end{bmatrix} \qquad (24)$$

y sustituyendo (24) en (19) se tiene

$$\frac{\partial f}{\partial P_i} + \left(\frac{\partial g}{\partial P_i}\right)^t \underset{\sim}{\lambda} = 0 , \quad i \neq 1$$

En forma desarrollada y sabiendo que $\left(\frac{\partial g}{\partial P_i}\right) = \begin{bmatrix} -1 & & 0 \\ & \ddots & \\ 0 & & -1 \end{bmatrix}$

el sistema de ecuaciones anteriores queda

$$\begin{bmatrix} \frac{df_2}{dP_2} \\ \vdots \\ \frac{df_n}{P_n} \end{bmatrix} - \frac{df_1}{dP_1} \begin{bmatrix} 1 - \frac{\partial P_L}{\partial P_2} \\ \vdots \\ 1 - \frac{\partial P_L}{\partial P_n} \end{bmatrix} = 0 \qquad (25)$$

Diferencias con las formulaciones anteriores

i) El multiplicador de Lagrange (λ) calculado con la ecuación (18), es $\lambda\,(-\frac{d\,1}{dP_1})$ de las formulaciones anteriores afectada por el factor de penalización, el cual está implícito en el cálculo.

ii) El conjunto de ecuaciones (25) no se resuelve en forma analítica para las P_i, sino que se le da un significado de gradiente reducido de la forma:

$$\underset{\sim}{\nabla F} = \frac{\partial F}{\partial u} + \left(\frac{\partial g}{\partial u}\right)^t \underset{\sim}{\lambda} \qquad (26)$$

El cual se resuelve en forma numérica para obtener incrementos en las variables $\underset{\sim}{u}$ que conducen al óptimo deseado

$$u^{n} = u^{a} + \Delta u$$

$$\Delta u = - c \nabla F \qquad (27)$$

El signo negativo de (27) indica proceso de minimización (dirección opuesta al gradiente), c es un factor que optimiza el Δu evitando oscilaciones alrededor del mínimo.

Se llega al óptimo (mínimo) cuando $\nabla F = 0$. En la práctica es usual especificar una tolerancia, $|\nabla F| \leq \epsilon$

En el caso de que las variables $\underset{\sim}{u}$ (de control) esten restringidas

$$u_i^{min} \leq u_i \leq u_i^{max}$$

Las componentes de ∇F tendrán el siguiente comportamiento en el óptimo.

$$\frac{\partial F}{\partial u_i} = 0 \qquad u_i^{min} \leq u_i \leq u_i^{max}$$

$$\frac{\partial F}{\partial u_i} \leq 0 \qquad u_i = u_i^{max}$$

$$\frac{\partial F}{\partial u_i} \geq 0 \qquad u_i = u_i^{min}$$

iii) Las restricciones de igualdad de la función aumentada

$$L(\underset{\sim}{x}, \underset{\sim}{u}, \underset{\sim}{\lambda}) = f(\underset{\sim}{x}, \underset{\sim}{u}) + \underset{\sim}{\lambda}\ g(\underset{\sim}{x}, \underset{\sim}{u})$$

$g\ (\underset{\sim}{x}, \underset{\sim}{u})$ son las ecuaciones de red

En las formulaciones anteriores la restricción de igualdad es:

$$g = P_L + P_d - \sum P_i = 0$$

Algoritmo de Solución

1. Suponer un conjunto de variables de control $\underset{\sim}{u}$
2. Resolver un estudio de flujos convencional para determinar una solución factible (se utiliza el método desacoplado).
3. Resolver la ecuación (18) para $\underset{\sim}{\lambda}$
4. Sustituir $\underset{\sim}{\lambda}$ del paso 3 en la ecuación (19) para calcular el gradiente ∇F

$$\nabla F = \frac{\partial f}{\partial u} + \left(\frac{\partial g}{\partial u}\right)^t \lambda$$

5. Obtener incrementos

$$\Delta u = -c\, \nabla F$$

donde

$$c = \frac{u^{h-1} - u^h}{\nabla F_i^{h-1} - \nabla F_i^h}$$

6. Calcular u_i^h

$$u_i^h = \begin{cases} u_i^{max} & si \quad u_i^{h-1} + \Delta u_i > u_i^{max} \\ u_i^{min} & si \quad u_i^{h-1} + \Delta u_i < u_i^{min} \\ u_i^{h-1} + \Delta u_i & u_i^{min} < u_i^{h-1} + \Delta u_i < u_i^{max} \end{cases}$$

7. Verificar convergencia

Si $$\left| u_i^h - u_i^{h-1} \right| \leq \varepsilon$$

el mínimo se ha determinado, de otra manera, regresar al paso 2.

Condiciones de Kuhn-Tucker

Minimizar la función objetivo con respecto a las variables de control u . f es función de variables dependientes (x)y de control (u).

$$\text{Min} \quad f(\underset{\sim}{x}, \underset{\sim}{u})$$

Sujeto a restricciones de la red, balance de potencias nodal

$$g(\underset{\sim}{x}, \underset{\sim}{u}) = 0$$

y sujeto a restricciones de desigualdad en las variables de control

$$u - u^{max} \leq 0$$

$$u^{min} - u \leq 0$$

Condiciones necesarias para el mínimo

$$\nabla L = 0 \qquad \text{con respecto a } (\underset{\sim}{u}, \underset{\sim}{x}, \underset{\sim}{\lambda})$$

y

$$\begin{aligned} \mu^{max}(u - u^{max}) &= 0 \\ \mu^{min}(u^{min} - u) &= 0 \\ \mu^{max} \geq 0 \qquad \mu^{min} &\geq 0 \end{aligned} \tag{28}$$

Si la variable de control está entre límites

$$u^{min} \leq u \leq u^{max}$$

entonces

$$\mu^{max} = \mu^{min} = 0$$

Si la variable de control está en un límite

$$\mu^{max} \geq 0 \quad o \quad \mu^{min} \geq 0$$

El Lagrangiano de la función aumentada es

$$L = f(\underset{\sim}{x}, \underset{\sim}{u}) + \lambda^t g(\underset{\sim}{x}, \underset{\sim}{u}) + \mu^{max}(u - u^{max}) + \mu^{min}(u^{min} - u)$$

μ^{max}, μ^{min} multiplicadores asociados a las desigualdades en las variables de control.

$$\nabla L_x = 0 = \frac{\partial f}{\partial x} + \left(\frac{\partial g}{\partial x}\right)^t \lambda = 0$$

$$\nabla L_u = 0 = \frac{\partial f}{\partial u} + \left(\frac{\partial g}{\partial u}\right)^t \lambda + \mu = 0$$

tambien puede escribirse como

$$\nabla L_u = \nabla_u f + \mu = 0$$
$$\mu = -\nabla_u f \qquad (29)$$

si el gradiente = 0 $\Rightarrow$ las variables de control están entre límites

$$u^{min} \leq u \leq u^{max}$$

si $\mu > 0$ implica que se tiene la variable de control en el máximo y el gradiente es negativo $\frac{\partial f}{\partial u} < 0$ si se libera la restricción la función de costo disminuye.

si $\mu < 0 \Longrightarrow$ se tiene la variable de control en el mínimo y el gradiente es positivo $\frac{\partial f}{\partial u} > 0$

Como las condiciones de Kuhn-Tucker se deben cumplir, no se permite que las variables de control violen límites y las condiciones de exclusión se deben de cumplir en el óptimo

$$\begin{aligned} \nabla_u f = 0 \qquad & u^{min} \leq u \leq u^{max} \\ \nabla_u f \leq 0 \qquad & u = u^{max} \\ \nabla_u f \geq 0 \qquad & u = u^{min} \end{aligned} \qquad (30)$$

Métodos para obtener cambios en las variables de control.

Sea $f(\underset{\sim}{x},\underset{\sim}{u}) = f_x(\underset{\sim}{x},\underset{\sim}{u})\ \Delta x + f_u\ (\underset{\sim}{x},\underset{\sim}{u})\Delta u\ + f(\underset{\sim}{x}^o, \underset{\sim}{u}^o)$

$$\Delta f(\underset{\sim}{x},\underset{\sim}{u}) = f_x(\underset{\sim}{x},\underset{\sim}{u})\ \Delta x + f_u\ (\underset{\sim}{x},\underset{\sim}{u})\ \Delta u$$

las variables (x,u) están relacionadas mediante

$$g(\underset{\sim}{x},\underset{\sim}{u}) = 0$$

$$g_x(\underset{\sim}{x},\underset{\sim}{u})\ \Delta x + g_u\ (\underset{\sim}{x},\underset{\sim}{u})\ \Delta u\ = 0$$

$$\Delta x = -\ g_x^{-1}(x,u)\ g_u\ (x,u)\ \Delta u$$

por lo tanto

$$\Delta f(\underset{\sim}{x},\underset{\sim}{u}) = f_x(\underset{\sim}{x},\underset{\sim}{u}) \left\{ -g_x^{-1}(\underset{\sim}{x},\underset{\sim}{u}) g_u\ (\underset{\sim}{x},\underset{\sim}{u})\ \Delta u \right\}$$

$$+ f_u\ (\underset{\sim}{x},\underset{\sim}{u})\ \Delta u$$

$$\Delta f(\underset{\sim}{x},\underset{\sim}{u}) = \nabla_u f\ \Delta u \qquad (31)$$

el problema es determinar los cambios Δu que minimizan $f(\underset{\sim}{x},\underset{\sim}{u})$

$\nabla_u f$ gradiente reducido de f

1. Método descenso rápido

si $\Delta u = -c\ \nabla_u f$ (32)

c = constante

luego $\Delta f(\underset{\sim}{x},\underset{\sim}{u}) = -(\nabla_u f \cdot \nabla_u f)c$

lo cual ciertamente minimiza $f(\underset{\sim}{x},\underset{\sim}{u})$ sin embargo hay que determinar el rango de validez de $\nabla_u f$ y el valor C, esto requiere una busqueda unidimensional (Fibonacci, intervalo medio, busquedas dicotomica, etc.) Generalmente se emplea una interpolación, la convergencia generalmente es lenta y oscilatoria.

2. Método de programación lineal

$$Min \quad \nabla_u f \, \Delta u \tag{33}$$

S.a.

$$\Delta u^{min} \leq \Delta u \leq \Delta u^{max}$$

$$\Delta x^{min} \leq \Delta x(\Delta u) \leq \Delta x^{max}$$

estas últimas restricciones se obtienen por renglones de

$$\Delta x = - g_x^{-1}(\underset{\sim}{x},\underset{\sim}{u})\, g_u\,(\underset{\sim}{x},\underset{\sim}{u})\, \Delta u$$

3. Método de Segundo Orden

Si se hace una expansión hasta el segundo término de la función objetivo.

$$f(\underset{\sim}{x},\underset{\sim}{u}) = f(\underset{\sim}{x}^{\circ},\underset{\sim}{u}^{\circ}) + \nabla_u f^t \Delta u + \tfrac{1}{2}\, \Delta u^t H_u \, \Delta u \tag{34}$$

$$\nabla_u f + H_u \, \Delta u = 0$$

$$\Delta u = - H_u^{-1} \nabla_u f. \qquad (35)$$

Aquí el problema es determinar la inversa de la matriz Hessiana H_u, una aproximación es despreciar los elementos de fuera de la diagonal y obtener los de la diagonal mediante desplazamientos

$$H_{ii} = \frac{\Delta_i(\text{gradiente})}{\Delta u_i}$$

Como la matriz Hessiana debe ser positiva definida cualquier elemento negativo se considera como cero.

Esta aproximación es equivalente a tener una constante para cada variable de control multiplicada por el gradiente.

4. Método Mixto

Se combina el método de descenso rápido con el método de segundo orden, primero se usa el método de descenso rápido calculando el factor "C" en forma optima. El método de segundo orden se usa cuando algún componente del gradiente ha cambiado de signo (de una iteración (h-1) a otra (h))

$$H_{ii} = \frac{\left(\nabla_{u_i} f\right)^{h-1} - \left(\nabla_{u_i} f\right)^{h}}{u_i^{h-1} - u_i^{h}} \qquad (36)$$

siempre y cuando esta "segunda derivada" sea positiva.

5. Programación cuadrática

$$Min \quad \Delta f = \nabla_u f + \frac{1}{2} \Delta u^t H_u \Delta u \qquad (37)$$

s.a.

$$\Delta u^{min} \leq \Delta u \leq \Delta u^{max}$$

$$\Delta x^{min} \leq \Delta x(\Delta u) \leq \Delta x^{max}$$

el problema aquí es determinar Hu

sea $S = - g_x^{-1} g_u$

luego $H_u = L_{uu} + S^t L_{xx} S + 2 S^t L_{xu}$

aunque esta expresión se puede evaluar, requiere de mucho trabajo, por lo tanto es mejor evaluar H_u a traves de desplazamiento alrededor de un punto de operación.

Sea $f(\underset{\sim}{x},\underset{\sim}{u})$ una función cuadrática

$f(\underset{\sim}{x},\underset{\sim}{u}) = f(\underset{\sim}{x}^o,\underset{\sim}{u}^o) + a^t h(\underset{\sim}{x},\underset{\sim}{u}) + h^t(\underset{\sim}{x},\underset{\sim}{u}) A \; h(\underset{\sim}{x},\underset{\sim}{u})$

un cambio Δu produce $\Delta h(\underset{\sim}{x},\underset{\sim}{u})$

por lo tanto

$$\Delta f = \left(a^t + 2 h^t(\underset{\sim}{x},\underset{\sim}{u}) A\right) \Delta h(\underset{\sim}{x},\underset{\sim}{u}) + \Delta h^t(\underset{\sim}{x},\underset{\sim}{u}) A \, \Delta h(\underset{\sim}{x},\underset{\sim}{u})$$

el cambio $\Delta h(\underset{\sim}{x},\underset{\sim}{u})$ producido por el Δu se puede obtener como

$$\Delta h(\underset{\sim}{x},\underset{\sim}{u}) = S_h \Delta u$$

donde la matriz S_h es una matriz de sensitividad

$$\Delta f = \left(a^t + 2 h^t(\underset{\sim}{x},\underset{\sim}{u}) A\right) S_h \Delta u + \Delta u^t \left(S_h^t A S_h\right) \Delta u \qquad (38)$$

el primer término $\left(a^t + 2h^t(\underset{\sim}{x},\underset{\sim}{u}) A\right) S_h$
es el gradiente reducido $\nabla_u f$

el segundo término $S_h^t A S_h$ es la matriz Hessiana H_u

La matriz S_h se obtiene mediante desplazamientos Δu y viendo su efecto en Δh, cada elemento requiere de una substitución hacia adelante y hacia atrás con un Jacobiano constante.

Restricciones Funcionales

En este tipo de restricciones se involucran variables dependientes

$$\phi(x,u) \leq 0 \qquad (39)$$

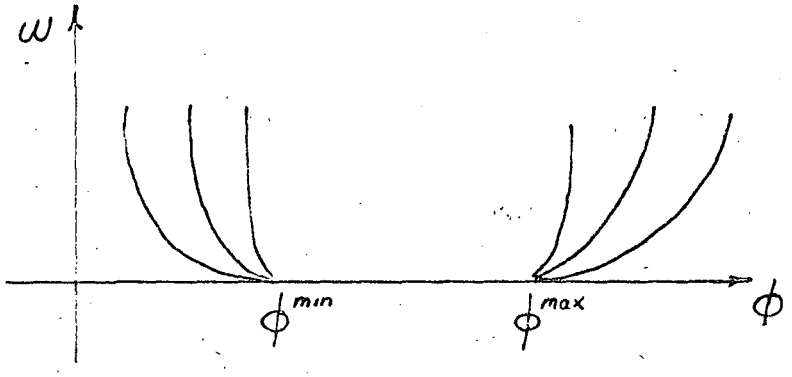
Como no se tiene ningún control sobre las variables dependientes, una forma de incluir estas restricciones es a través de penalizaciones cuando la restricción se viola

$$f = f(x,u) + \sum \omega_j$$

donde las penalizaciones son del tipo

$$\omega_j = S_j \left(\phi_j - \phi_j^{max} \right)^2 \qquad \omega_j = S_j \left(\phi_j - \phi_j^{min} \right)^2 \qquad (40)$$

donde S_j es una constante que modifica la penalización de acuerdo al proceso



generalmente se empieza en una S_j calculada de manera que la función objetivo y las penalizaciones sean del mismo orden, después los valores de S_j se van modificando en forma arbitraria de acuerdo al problema. Ejemplo, si un término de penalización no disminuye en un factor de 4 la constante se multiplica por 5.

Este es uno de los aspectos débiles del método ya que requiere ajustes empéricos.

Otra forma más efectiva de incluir restricciones fun-

cionales es expresarlos en función de las variables de control a través de relación de sensitividad

$$\phi(\underset{\sim}{x}, \underset{\sim}{u}) = \left[S_{\phi,u} \right] \left[\Delta u \right] \qquad (41)$$

esto se puede hacer mediante calcular analíticos, o bien mediante desplazamientos (estudios de flujos).

Al incluir las restricciones funcionales de esta manera se calculan los Δu las variables de control ni las restricciones funcionales.

IV. Programación lineal

En esta formulación se requieren curvas de costo lineales o en segmentos lineales y restricciones lineales. El atractivo es que la solución se obtiene en un número finito de pasos, y no se tienen problemas de convergencia de las formulaciones no-lineales.

La función objetivo a minimizar será el costo de generación total

$$F = \sum_i C_i = \sum_i a_i + \sum_i b_i P_{G_i}$$

En forma alternativa se puede minimizar el cambio de costo a partir de un punto de operación especificado $(P_{G_i})_o$

$$\Delta F = \sum_i b_i \Delta P_{G_i} \qquad (42)$$

Las restricciones que se deben satisfacer en la minimización son:

i. la demanda en todos los nodos de la red

ii. la generación de las unidades generadoras debe estar entre un máximo y un mínimo.

iii. el flujo de potencia activa en líneas de transmisión debe ser menor que un máximo especificado en cualquier dirección.

Las restricciones (i) son las ecuaciones de flujos de la red.

Las restricciones (ii) pueden expresarse matematicamente como

$$P_{G_i}^{min} - P_{G_i} \leq \Delta P_{G_i} \leq P^{max} - P_{G_i} \qquad (43)$$

Las restricciones (iii) se pueden expresar como

$$\Delta P_{ij} \leq P_{ij}^{max} - P_{ij} \qquad (44)$$

para lineas que han violado el límite

De la ecuación (23) se tiene que el cambio de potencia en el nodo compensador se expresa como

$$\Delta P_s = \sum_{\substack{i=1 \\ i \neq s}} \frac{\partial P_s}{\partial P_{G_i}} \Delta P_{G_i} \qquad (45)$$

Esta es una restricción adicional al problema la cual puede ser manejada como restricción o bien se substituye en la función objetivo. Siguiendo la segunda opción la función objetivo se transforma a

$$\Delta F = \sum_{\substack{i=1 \\ i \neq s}} \left(b_i + \frac{\partial P_s}{\partial P_{G_i}} b_s\right) \Delta P_{G_i} \qquad (46)$$

la cual solo es función de las variables independientes. Los términos dentro del paréntesis son componentes del gradiente reducido de la función (ver ecuacion (25))

Los cambios en los flujos de potencia activa también pueden expresarse en función de las generaciones independientes

$$\Delta P_{ij} = \sum_{\substack{i=1 \\ i \neq s}} \frac{\partial P_{ij}}{\partial P_{G_i}} \Delta P_{G_i} \qquad (47)$$

De las ecuaciones (46) y (47) se observa que se requiere la evaluación de $\frac{\partial P_s}{\partial P_{G_i}}$ y $\frac{\partial P_{ij}}{\partial P_{G_i}}$, la cual se puede efectuar mediante un cálculo analítico o bien por desplazamientos a partir de un punto de operación.

Para el cálculo por desplazamientos se usa el Jacobiano del problema de flujos

$$H \underset{\sim}{\Delta x} = \underset{\sim}{\Delta u} \qquad (48)$$

donde

- H es el Jacobiano de la potencia activa con respecto a los ángulos de fase.
- $\underset{\sim}{\Delta x}$ es el vector de cambios en los ángulos de fase.
- $\underset{\sim}{\Delta u}$ es el vector de cambios en las variables independientes.

$\underset{\sim}{\Delta u}$ es un vector de ceros excepto el elemento con respecto al cual se calcula la sensitividad. Con el $\underset{\sim}{\Delta x}$ producido por el $\underset{\sim}{\Delta u}$ se calculan los cambios en la generación del nodo compensador (ΔP_s) y el cambio en la potencia activa de las líneas sobre cargadas (ΔP_{ij}). Los coeficientes de sensitividad se obtienen como

$$\frac{\partial P_s}{\partial P_{G_i}} = \frac{\Delta P_s}{\Delta P_{G_i}} \qquad \frac{\partial P_{ij}}{\partial P_{G_i}} = \frac{\Delta P_{ij}}{\Delta P_{G_i}}$$

El tiempo de cálculo es mínimo ya que solo involucra una substitución hacia adelante y hacia atrás en el Jacobiano triangularizado. Este procedimiento se repite para cada variable independiente.

El cálculo analítico hace uso de las condiciones necesarias para el óptimo (ecuaciones (18)-(20))

$$\frac{\partial L}{\partial x} = H^t \lambda + \frac{\partial f}{\partial x} = 0$$

$$H^t \lambda = - \frac{\partial f}{\partial x} \qquad (49)$$

Como la única potencia de generación dependiente es la del nodo compensador se tiene

$$H^t \lambda = - \frac{\partial f_s}{\partial x} \qquad (50)$$

De la solución de este sistema de ecuaciones se obtiene λ que se puede expresar como

$$\lambda = - \frac{\partial f_s}{\partial P}$$

o bien

$$\lambda = - \frac{\partial P_s}{\partial P} b_s \qquad (51)$$

que corresponde al segundo término dentro del paréntesis de la ecuación (46)

Para el cálculo de coeficientes de sensitividad de

líneas se utiliza la regla de la cadena

$$\left(\frac{\partial P}{\partial \theta}\right)^t z = \frac{\partial P_{ij}}{\partial \theta} \qquad (52)$$

donde

$$z = \frac{\partial P_{ij}}{\partial P}$$

La ecuación (52) se resuelve para cada línea que viole su límite de operación. Es importante notar que en las ecuaciones (50)y(52) se utiliza el Jacobiano transpuesto, el cual para mejorar la eficiencia se aproxima por términos constantes (ecuacion (16)). La matriz resultante es simétrica y solo se requiere una triangularización.

Al tener una función objetivo lineal en las variables independientes, se requieren límites para efectuar el despacho económico. El método de programación lineal permite manejar en forma eficiente todos los límites y restricciones del problema.

De acuerdo a las condiciones es necesarios para el óptimo.

1. Las ecuaciones de flujos se resuelven en forma independiente ($\frac{\partial L}{\partial \lambda} = 0$).

2. La parcial ($\frac{\partial L}{\partial x} = 0$) se utiliza para calcular coeficientes de sensitividad.

3. La parcial ($\frac{\partial L}{\partial u} = 0$) (gradiente reducido) se resuelve mediante la programación lineal.

V. Método de la Matriz Hessiana

En este enfoque la minimización se efectúa en función de las variables dependientes y se usa la técnica de penalización. La función objetivo del problema se penaliza con términos cuadráticos de las violaciones de restricciones. En el problema de flujos no existe función objetivo pero las restricciones de igualdad del problema siempre se encuentran violadas, excepto en la solución, por lo tanto

$$F = \sum_i (g_i(\underset{\sim}{x},\underset{\sim}{y}))^2 \qquad (53)$$

El mínimo de este problema se conoce que es cero al tener satisfechas simultaneamente todas las restricciones. Linealizando F con respecto a x se tiene

$$F = \sum_i \left(g_i(\underset{\sim}{x}^o,\underset{\sim}{y}) + \sum_j \frac{\partial g}{\partial x_j} \Delta x_j\right)^2$$

en forma matricial

$$F = (g_i(\underset{\sim}{x}^o,\underset{\sim}{y}) + J\,\Delta x)^t (g(\underset{\sim}{x}^o,\underset{\sim}{y}) + J\,\Delta x) \qquad (54)$$

Minimizando con respecto a x

$$\nabla_x F = 0 = 2\,J^t J\,\Delta x + 2\,J^t g(\underset{\sim}{x}^o,\underset{\sim}{y})$$

$$J^t\,J\,\Delta x = -\,J^t\,g(\underset{\sim}{x}^o,\underset{\sim}{y}) \qquad (55)$$

Es interesante notar que al ser J una matriz cuadra-

da no-singular la ecuación (55) se reduce a

$$J \Delta x = - g(\underset{\sim}{x}^0, \underset{\sim}{y}) \qquad (56)$$

que es el resultado que se obtiene con el método de Newton.

En el problema de flujos óptimos existe una función objetivo sujeta a restricciones de igualdad y desigualdad

$$F = f(\underset{\sim}{x})$$

sujeto a

$$\begin{aligned} g_i(\underset{\sim}{x}, \underset{\sim}{y}) &= 0 \qquad i=1, \ldots n \\ g_j(\underset{\sim}{x}, \underset{\sim}{y}^{lim}) &\geqslant 0 \qquad j=1, \ldots m \end{aligned} \qquad (57)$$

Las restricciones de igualdad siempre estarán violadas por lo tanto se incluirán en la función penalizada, en cambio las restricciones de desigualdad solo se incluirán cuando estén violadas

$$F = f(\underset{\sim}{x}) + \sum_i (g_i(\underset{\sim}{x}, \underset{\sim}{y}))^2 - \sum_{k \in \alpha} (g_k(\underset{\sim}{x}, \underset{\sim}{y}^{lim}))^2 \qquad (58)$$

α conjunto de restricciones de desigualdad violadas.

Si se expande la función objetivo hasta el segundo término de la serie de Taylor

$$F = f(\underset{\sim}{x}_0) + \frac{\partial f}{\partial x}^t \Delta x + \frac{1}{2} \Delta x^t \frac{\partial^2 f}{\partial x^2} \Delta x$$

$$= (g(\underset{\sim}{x}^0, \underset{\sim}{y}) + G \Delta x)^t W (g(\underset{\sim}{x}^0, \underset{\sim}{y}) + G \Delta x) \qquad (59)$$

donde

W es una matriz diagonal de ponderación

G es el Jacobiano de restricciones violadas

Minimizando con respecto a x se tiene

$$\left(\frac{\partial^2 f}{\partial x^2} + 2\,G^t W G\right)\Delta x = -2\,G^t W\,g(x^\circ, y) - \frac{\partial f}{\partial x}\Big|^\circ \qquad (60)$$

la corrección en cada iteración es

$$x^i = x^{i-1} + \beta \Delta x^i \qquad (61)$$

La característica importante del método es que considera la matriz de segundas derivadas de la función de costo y maneja en forma sencilla cualquier restricción.

Es importante notar que la técnica de penalización se puede utilizar en el método de gradiente para controlar las variables dependientes o funciones de variables dependientes.

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA XVII DESARROLLO OPTIMO DEL SISTEMA DE GENERACION

M. EN I. RAFAEL CRISTERNA OCAMPO

ENERO, 1979.

Comisión Federal de Electricidad, México.

Gerencia General de Planeación y Programa

Sub-Gerencia de Programa y Control

Departamento de Metodología

Año 1974.

Ing. Rafael Cristerna Ocampo.

APLICACION DEL PRINCIPIO DEL MAXIMO DE PONTRYAGIN A LA SELECCION DE EQUIPO PARA EL DESARROLLO DE UN SISTEMA DE POTENCIA. (FORMULACION SIMPLIFICADA DEL MNI PARA DESCRIBIR EL PROCESO DE CALCULO).

RESUMEN: Se enfoca como una aplicación del máximo de PONTRYAGIN a la selección de equipo para el desarrollo de un sistema de potencia. El problema se visualiza como un problema de control óptimo con el fin de obtener una descomposición en el tiempo. El proceso de solución se basa en la aplicación del método del gradiente a cada uno de los subproblemas en forma iterativa, los que se coordinan a través de las ecuaciones adjuntas, hasta la obtención de la convergencia.

APLICACION DEL PRINCIPIO DEL MAXIMO DE PONTRYAGIN A LA SELECCION DE EQUIPO PARA EL DESARROLLO DE UN SISTEMA DE POTENCIA.

(FORMULACION SIMPLIFICADA PARA DESCRIBIR EL PROCESO DE CALCULO).

R. CRISTERNA.

Enero 1974.

SISTEMAS CON ETAPAS MULTIPLES.

Un sistema de esta clase puede esquematizarse según la figura 1.

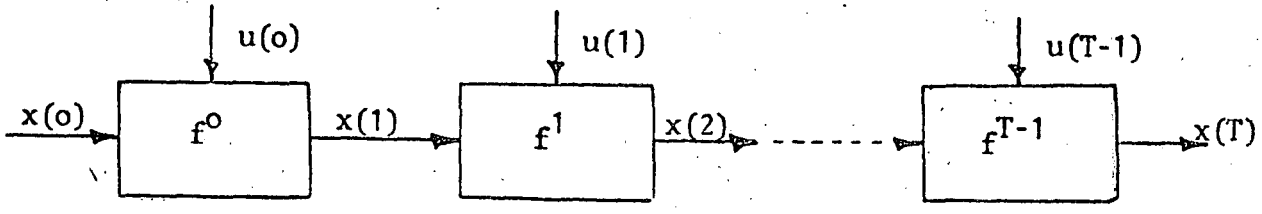


Fig. 1 Sistema con etapas múltiples.

La obtención de los controles u(o), u(1),---u(T-1) para llevar el sistema desde el estado x(o) al x(T), de tal forma -- que se minimice un cierto índice de comportamiento $J = J\left[x(t),u(t)\right]$, consiste en la solución de un problema de control óptimo discreto en el tiempo, el cual se enuncia como sigue:

"Para un sistema con etapas mútiples descrito por -- las ecuaciones del comportamiento (ecuaciones de diferencias no li - neales) secuencial o ecuaciones de estado de evolución de sistema:

$$x(t+1) = f^t\left[x(t), u(t)\right] \quad , \quad t=0,1,---T-1 \qquad (1)$$

$$\underline{u}(t) \leq u(t) \leq \bar{u}(t)$$

Con x(o) dado, x(t) una secuencia de n vectores de-estado, que son definidos por u(t), una secuencia de m vectores de-control. Si existe un índice de comportamiento global del sistema -- que sea de la siguiente forma (separable).

$$J = \phi\left[x(T)\right] + \sum_{t=o}^{T-1} L^t\left[x(t),u(t)\right] , \qquad (2)$$

el problema consiste en obtener la secuencia u(t) que maximiza (o minimiza) J".

FORMULACION PARA LA SOLUCION.

Las ecuaciones (1) pueden verse como un conjunto secuencial de restricciones de igualdad, las cuales se pueden adjuntar al índice J a través de una secuencia de multiplicadores $\psi(i)$ para obtener una nueva función $\bar{J}$:

$$\bar{J} = \phi\left[(x(T)\right] + \sum_{t=0}^{T-1} \left\{ L^t\left[x(t),u(t)\right] + \psi^*(t+1) \left\{ f^t\left[x(t),u(t)\right] - x(t+1) \right\} \right\} \qquad (3)$$

Se definirá una secuencia escalar H^t (Hamiltoniano):

$$H^t = L^t\left[x(t),u(t)\right] + \psi^*(t+1)\, f^t\left[x(t),u(t)\right] , \quad t=0,1,\ldots,T-1 \qquad (4)$$

Entonces, se puede escribir:

$$\bar{J} = \phi\left[x(T)\right] - \psi(T)x(T) + \sum_{t=1}^{T-1} \left[H^t - \psi^*(t)x(t)\right] + H^o \qquad (5)$$

Si se considera un cambio diferencial en $\bar{J}$, debido a cambios diferenciales u(t):

$$d\bar{J} = \left[\frac{\partial \phi}{\partial x(T)} - \psi^*(T)\right] dx(T) + \sum_{t=1}^{T-1} \left\{ \left[\frac{\partial H^t}{\partial x(t)} - \psi^*(t)\right] dx(t) + \frac{\partial H^t}{\partial u(t)} du(t) \right\}$$

$$+ \frac{\partial H^o}{\partial x(o)} dx(o) + \frac{\partial H^o}{\partial u(o)} du(o) \qquad (6)$$

Entonces, las condiciones para obtener un punto extremo quedan:

$$\frac{\partial H^t}{\partial x(t)} - \psi^*(t) = 0 \Rightarrow \psi^*(t) = \frac{\partial H^t}{\partial x(t)} = \frac{\partial L^t}{\partial x(t)} + \psi^*(t+1) \frac{\partial f^t}{\partial x(t)}$$

$t = o,1,\ldots,T-1$ (7)

$$\psi^*(T) = \frac{\partial \phi}{\partial x(T)} \qquad (8)$$

$$\frac{\partial H^t}{\partial u(t)} = 0 \quad t=o,1,\ldots,T-1 \qquad (9)$$

como x(o) es dado, dx(o) = o

Es decir, para maximizar el índice de comportamiento global es necesario maximizar el Hamiltoniano H^t en cada etapa, Ec. (9) resolviendo simultaneamente a las ecuaciones adjuntas (7), tomando en cuenta las condiciones finales (8) y las ecuaciones del comportamiento dinámico del sistema (1).

EXPANSION OPTIMA GLOBAL DE UN SISTEMA DE POTENCIA.

El problema consiste en construir un modelo que permita determinar las políticas de selección de equipo a lo largo de un cierto período de tiempo tratando de minimizar el costo total actualizado:

Costo de la expansión = Inversión + Operación + Falla. (10)

FORMULACION DEL PROBLEMA.

El sistema se caracteriza por un cierto conjunto de medios de producción (generación)

Nuclear

Térmica convencional

Térmica de pico

Hidráulica

en el que para cada medio de los anteriores, pueden existir varios tipos.

Los costos de operación en un año dado, para las plantas de un cierto tipo, son función de los conjuntos de las plantas de otros tipos disponibles en ese año.

El período de vida elevado de las plantas (30 o mas años) define que los estudios se hagan a largo plazo.

La formulación de este problema se hará en base al problema de control óptimo establecido en las ecuaciones (1) a (9).

El sistema se caracteriza por las variables de estado x_i^t que representan la cantidad de potencia instalada hasta antes del año t del medio de generación del tipo i, (i=1,2,...,n)

El flujo de inversión queda directamente definido por la cantidad de potencia puesta en servicio el año t del tipo i, representada por el vector de control u_i^t, (i=1,2,.....n).

Entonces, las ecuaciones de estado, que caracterizan la evolución del sistema de generación se escriben a partir de la Ec. (1).

$$x_i^{t+1} = x_i^t + u_i^t \,, \qquad (i=1,2,\ldots,n) \tag{11}$$

con las restricciones $u_i^t \geq 0$, $\underline{u}_i^t \leq u_i^t \leq \bar{u}_i^t$ (12)

FUNCION OBJETIVO.

Se debe escribir una expresión que defina el costo - global actualizado, de acuerdo con la Ec. (10).

El costo actualizado de inversión en el año t, es:

$$I^t = \sum_{i=1}^{n} I_i^t \, u_i^t \tag{13}$$

en donde I_i^t es el costo unitario actualizado de inversión del equipo de tipo i en el año t, (costo/MW)

Se definirán:

$G^t\left[x^t,u^t\right]$ = Costo actualizado de operación en el año t,

$F^t\left[x^t,u^t\right]$ = Costo actualizado de falla (o de la energía no suministrada) en el año t.

Los dos conceptos anteriores son completamente probabilísticos, por lo tanto deben considerarse como esperanzas matemáticas de los costos.

Entonces, el problema de selección óptima de equipos de generación puede escribirse como sigue:

"Encontrar las capacidades instaladas u_i^t (y los flujos de inversión $I_i^t \, u_i^t$) óptimos, de manera que el costo global actualizado

$$CGA = \sum_{t=1}^{T} \left\{ \sum_{i=1}^{n} I_i^t \, u_i^t + G^t(x^t,u^t) + F^t(x^t,u^t) \right\} \tag{14}$$

sea mínimo para el sistema que evoluciona de acuerdo con las ecuaciones de estado

$$x_i^{t+1} = x_i^t + u_i^t \qquad (i=1,2,..,n) \qquad (11)$$

con las condiciones iniciales dadas x_i^o.

CONDICIONES DEL OPTIMO.

El Hamiltoniano de la Ec. (3) queda, para el caso de minimización:

$$H^t = -\left\{ \sum_{i=1}^{n} I_i^t u_i^t + G^t\left[x^t,u^t\right] + F^t\left[x^t,u^t\right] \right\} + \psi^{t+1}\left[x^t+u^t\right] \qquad (15)$$

Aplicando las Ecs. (7), (8) y (9), se obtiene:

$$\frac{\partial H^t}{\partial x_i^t} = \psi_i^t = -\frac{\partial G^t}{\partial x_i^t} - \frac{\partial F^t}{\partial x_i^t} + \psi_i^{t+1} \quad \Longrightarrow \qquad (16)$$

$$\psi_i^{t+1} = \psi_i^t + \frac{\partial G^t}{\partial x_i^t} + \frac{\partial F^t}{\partial x_i^t} \quad (i = 1,2,...,n) \qquad (16^1)$$

$$\frac{\partial H^t}{\partial u_i^t} = -I_i^t - \frac{\partial G}{\partial u_i^t} - \frac{\partial F^t}{\partial u_i^t} + \psi_i^{t+1} \qquad (17)$$

Sustituyendo (16^1) en (17):

$$\frac{\partial H^t}{\partial u_i^t} = -I_i^t - \frac{\partial G^t}{\partial u_i^t} - \frac{\partial F^t}{\partial u_i^t} + \psi_i^t + \frac{\partial G^t}{\partial x_i^t} + \frac{\partial F^t}{\partial x_i^t} \qquad (18)$$

puesto que:

$$G^t(x^t,u^t) = G^t(x^t+u^t) \tag{19}$$

$$F^t(x^t,u^t) = F^t(x^t+u^t) \tag{20}$$

$$\frac{\partial G^t}{\partial x_i^t} = \frac{\partial G^t}{\partial u_i^t} \quad ,y \quad \frac{\partial F^t}{\partial x_i^t} = \frac{\partial F^t}{\partial u_i^t} \tag{21}$$

Sustituyendo en (18), el gradiente queda:

$$\frac{\partial H^t}{\partial u_i^t} = - I_i^t + \psi_i^t \quad \text{ó}$$

$$\boxed{\frac{\partial H^t}{\partial u_i^t} = \psi_i^t - I_i^t} \tag{22}$$

ALGORITMO DE SOLUCION.

Se utiliza el método del gradiente de 1er órden (máximo descenso).

Paso 0 Se asigna una evolución inicial u_i^t para toda i y t.

Paso 1 Se calculan las potencias instaladas x_i^t con la Ec.-(11) a partir de t=1

$$x_i^{t+1} = x_i^t + u_i^t$$

Paso 2 Se calcula desde la frontera T, los valores ψ_i^t con la Ec. (16).

$$\psi_i^t = \psi_i^{t+1} - \frac{\partial G^t}{\partial x_i^t} - \frac{\partial F^t}{\partial x_i^t}$$

Paso 3 Se mejora el control u_i^t, corrigiendo en la dirección del gradiente de Π^t.

$$u_i^t(\theta) = o \quad si \quad u_i^t = o \quad y \quad \psi_i^t - I_i^t < o$$

$$u_i^t(\theta) = u_i^t + \theta K^t(\psi_i^t - I_i^t) \;, u_i^t \leq u_i^t \leq \bar{u}_i^t$$

Paso 4 Si $\left| \frac{\psi_i^t - I_i^t}{I_i^t} \right| < \varepsilon, \ \forall i,t \ \text{con} \ u_i^t > o$

el proceso termina, en caso contrario se regresa al paso 1.

CURVA DE DEMANDA.

La demanda se puede introducir al modelo, utilizando-la curva de duración de carga del sistema, la cual puede representarse con un polinomio de séptimo grado.

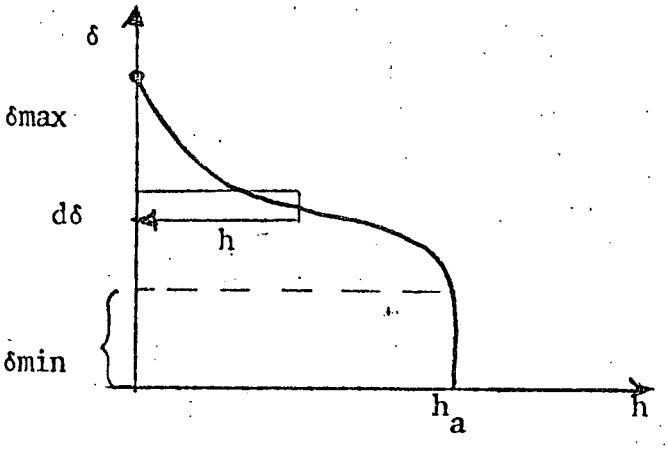
$$\delta(h) = \sum_{i=o}^{7} a_i h^i$$

en donde:

$\delta(h)$ = Demanda de potencia

h = Tiempo

La curva de duración de carga tiene la forma de la figura siguiente.



δmax. = Demanda máxima del sistema.

δmin = Demanda mínima del sistema.

h_a = Tiempo total del período considerado.

El área bajo la curva representa la energía E requerida por el sistema.

$$E = \delta min\ h_a + \int_{\delta min}^{\delta max} h d\delta(h)$$

$$g_i(x_i) = \int_{\delta_{i-1}}^{\delta_i = \delta_{i-1} + (1-f_{pi})\, f_{mi}\, x_i} h\,d\delta(h) + f_{pi}\, f_{mi}\, x_i\, h_a$$

5.- El costo de la generación suministrada por cada grupo es:

$$G_i(x_i) = \gamma_i \int_{\delta_{i-1}}^{\delta_i} h(\delta)d\delta + \beta_i\, f_{pi}\, f_{mi}\, x_i\, h_a$$

En donde:

γ_i = Costo de la generación complementaria del grupo i.

β_i = Costo de la generación del grupo i, cuando trabaja en el límite inferior de capacidad.

CALCULO DEL COSTO DE FALLA.

Se supone que la generación no suministrada por falla en el sistema es

$$G_f = A e^{\alpha(D-X)} \qquad (1)$$

D = Demanda máxima anual

X = Capacidad instalada

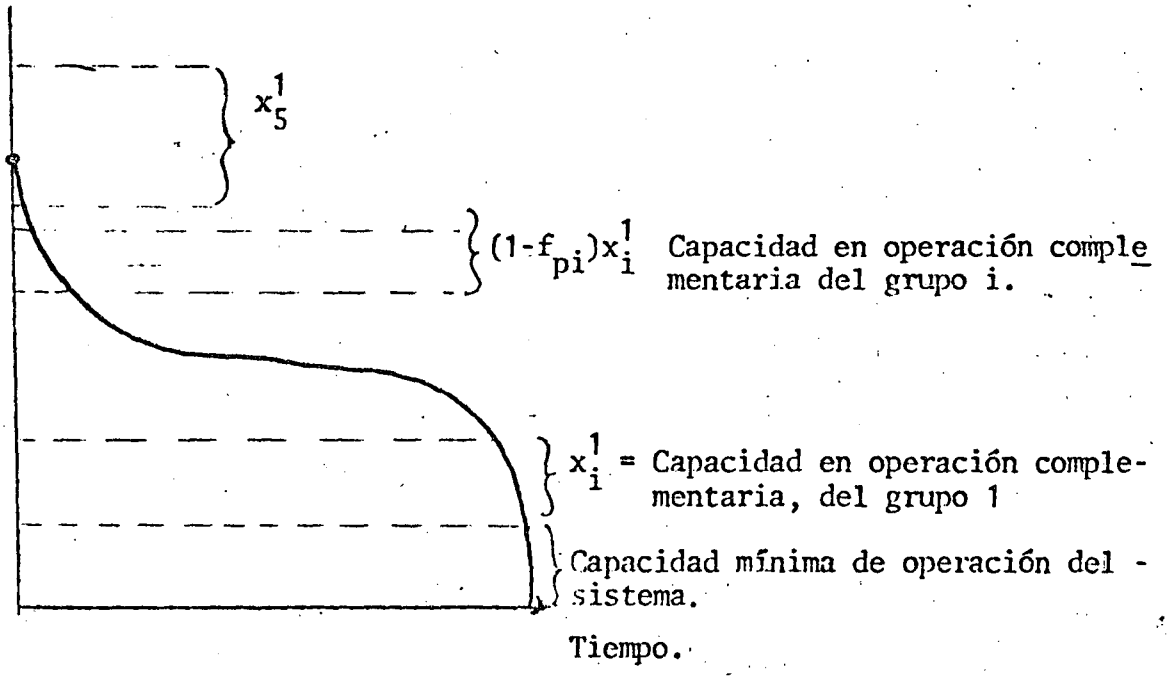
G_f = Generación no suministrada por falla

$X \leq D$

CALCULO DEL COSTO DE OPERACION.

El costo de operación dependerá del reparto de generación entre los distintos grupos de producción para satisfacer la demanda.

El reparto se hará en la forma clásica de asignación por órden de eficiencia bajo la curva de duración de carga. Si se tienen 5 grupos de producción, su operación queda asignada bajo la curva de duración de carga según la figura.



Sea x_i = Capacidad instalada del tiempo i,

f_{mi} = Factor de disponibilidad por mantenimiento del grupo i,

$$f_{mi} \leq 1.0$$

f_{pi} = Factor de operación de potencia mínima para plantas de tipo i.

p_i = Probabilidad de servicio para plantas de tipo i.

$\delta(h)$ = Ordenada de la curva de duración de carga (potencia) en función de la abscisa h que representa el tiempo.

Debido al mantenimiento, la capacidad en operación - de cada grupo es

$$x_i^1 = f_{m_i} \; x_i$$

El proceso de asignación de generación es el siguiente:

1.- A partir de la base $\delta = 0$, se asigna la capacidad mínima de operación del sistema.

$$x^1_{min} = \Sigma f_{pi} \cdot x_i^1 = \Sigma \; f_{pi} \; f_{mi} \; x_i$$

2.- A partir de la ordenada $\delta = x^1 min$, se asignan las capacidades de operación complementarios de cada grupo $(1 - f_{pi}) \; x_i^1$, en órden de- mayor a menor eficiencia.

3.- Se obtienen así, los niveles superiores de operación para cada grupo, que son

$$\delta_i = \delta_{i-1} + (1 - f_{pi}) \; x_i^1$$

$$= \delta_{i-1} + (1-f_{pi}) \; f_{mi} \; x$$

4.- La generación que suministra cada grupo es entonces:

Obtención de los parametros A y α

a) Si x=o, la energía no suministrada es igual a la Energía requeri_ da por el sistema.

$$E = Ae^{\alpha D} \qquad (2)$$

b) Si la capacidad en operación esta constituida por n grupos:

$$x = \Sigma x_i \qquad (3)$$

y cada grupo tiene una probabilidad de servicio asociada p_i, la capacidad esperada en servicio es:

$$\bar{x} = \Sigma p_i x_i \qquad (4)$$

Tomando como punto particular de operación el cual -- x = D, se puede calcular la generación no suministrada cuando se tiene la capacidad en servicio correspondiente $\bar{x}$, o sea:

$$G_{f_{(x=D)}} = Ae^{\alpha(o)}$$

$$A = G_f(x=D), \text{ para } \bar{x} < x \qquad (5)$$

De (2)

$$\boxed{\alpha = \frac{1}{D} \ln \frac{E}{A}} \qquad (6)$$

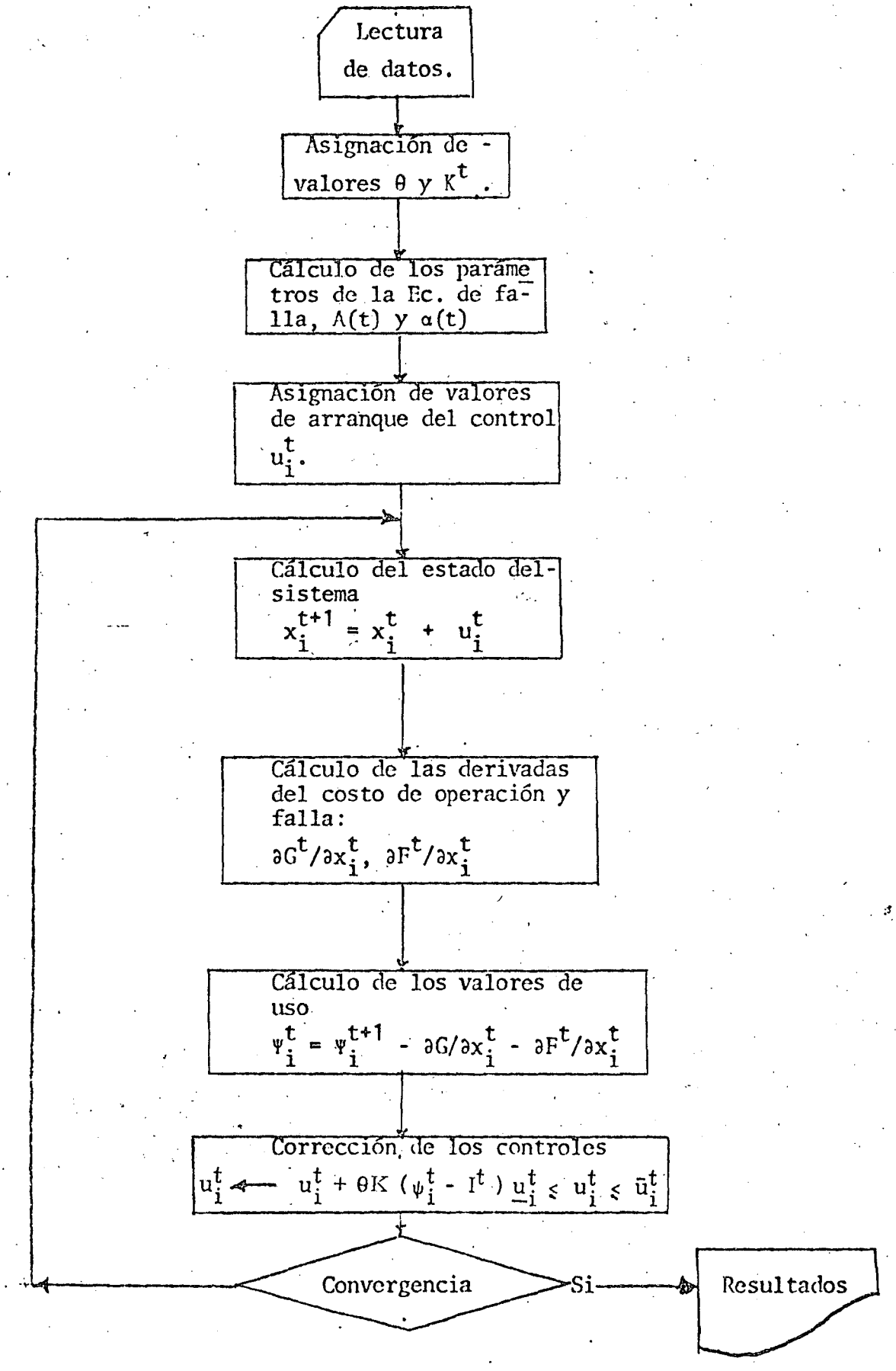
Es decir, para cada intervalo de tiempo, los parámetros A y α dependen de la demanda máxima D, la energía requerida total E y la energía no suministrada en un punto particular de operación .

Si ω es el costo del GWH no suministrado, el costo de

la falla es:

$$F(x) = \omega A e^{\alpha(D-x)}$$

DIAGRAMA DE BLOQUES - ADICION OPTIMA DE CAPACIDAD



centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

TEMA XVIII. LOCALIZACION DE PLANTAS Y LINEAS DE TRANSMISION

DR. FLORENCIO ABOYTES GARCIA

ENERO, 1979.

N O T A S

SOBRE LOCALIZACION DE PLANTAS Y

LINEAS DE TRANSMISION

PARA EL CURSO INTENSIVO

"TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS"

Preparadas por:

F. Aboytes
Comisión Federal de
Electricidad.

Enero 1979

LOCALIZACION DE LINEAS DE TRANSMISION Y PLANTAS GENERADORAS

INTRODUCCION

La expansión del sistema eléctrico de potencia requiere de la localización de plantas generadoras y líneas de transmisión para satisfacer la demanda de energía eléctrica. El problema requiere para su solución del uso de técnicas de optimización para obtener la alternativa óptima. La formulación involucra un número de variables y restricciones muy considerable , esto se debe a la característica dinámica del problema que abarca varios períodos de tiempo. Por otro lado se requiere manejar tamaños estandar de líneas y plantas generadoras para obtener una solución realista y poder aprovechar la economía de escala.

La solución óptima requiere de una complejidad matemática formidable por lo que se opta por hacer algunas consideraciones que simplifiquen el problema y se obtengan soluciones subóptimas.

FORMULACION

OBJETIVO

Localizar dinámicamente, en varios períodos de tiem-

po, dentro de un horizonte especificado.

i) medios de generación en tamaño estándar
ii) líneas de transmisión de capacidad estándar
iii) demanda industrial reubicable.

de manera que

i) se minimize una función objetivo
ii) se satisfagan restricciones impuestas

MODELO CURVA DURACION DE CARGA

El balance de potencia que se debe cumplir en cada nodo requiere de simplificar el modelo de la curva de duración de carga.

La curva de duración de carga del sistema determina la duración de los modos de operación del sistema, así para dos escalones del sistema se tienen los modos de operación.

i) en el pico
ii) en la base

La forma de la curva de duración de carga en cada nodo puede ser diferente, solo se respeta la duración de en

cada modo de operación. Las figuras 1-3 muestran las curvas de duración de carga de sistema y de dos nodos para las mismas duraciones de los modos de operacion

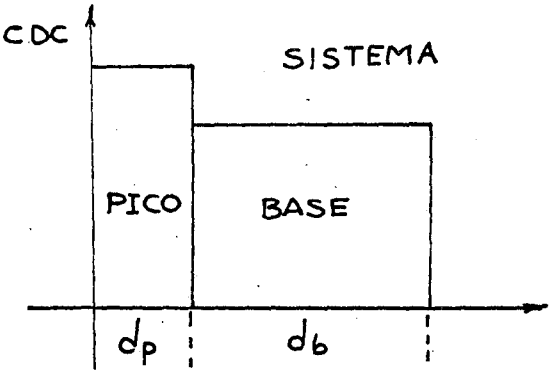


Fig 1.

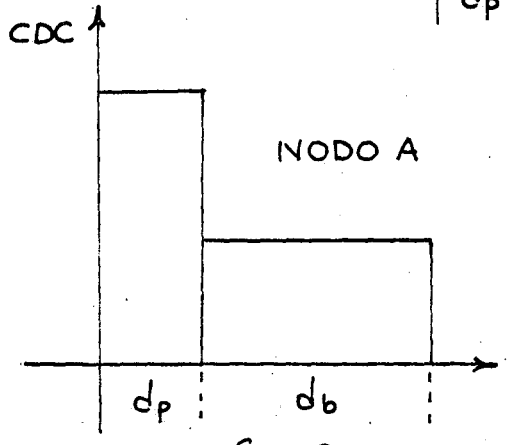


Fig. 2.

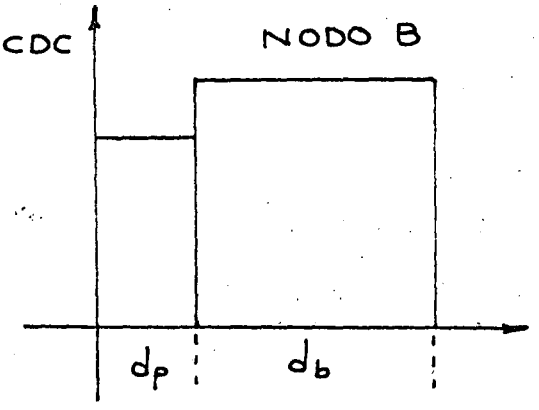


Fig. 3

MODELO DE PLANTAS GENERADORAS

Se deben considerar distintos tipos de plantas

1. nuclear
2. carbón
3. térmicas convencionales
4. turbina de gas
5. geotérmicas
6. hidráulicas

Y aún dentro de cada categoría se pueden considerar varios tamaños estándar para aprovechar la economía de escala. Por ejemplo térmica de 150MW y 300MW.

La utilización de la planta se controla mediante el uso de factores de carga superiores e inferiores, de manera de que las inversiones en generación sean efectivas.

Se debe tener la flexibilidad de manejar varias unidades del mismo tipo, esto resulta crítico cuando se manejan tamaños estándar.

Se utiliza el costo de inversión por nodo y tipo de generación, además el costo de operación se utiliza por nodo, tipo y modo de operación.

MODELO DE ENLACES DE TRANSMISION

En cada enlace de transmisión se requiere determinar el nivel de voltaje a usar. En general se deben tener alternativas en el voltaje de transmisión. Se recomienda tener dos alternativas de voltajes de transmisión por enlace.

La representación de las pérdidas se efectúa en forma aproximada, esta se obtiene mediante el producto del flujo en las líneas por la longitud, todo afectado por un factor de pérdidas.

Se utiliza el costo de inversión y costo de pérdidas por enlace de transmisión.

RESTRICCIONES

Las restricciones a considerar son las siguientes:

1. balance nodal de potencia
2. demanda industrial reubicable
3. utilización efectiva de generadores
4. límite de localización de generación
5. capacidad enlaces de transmisión
6. reserva de generación de demanda máxima
7. energía hidráulica disponible
8. potencia hidráulica máxima

FUNCION OBJETIVO

Se forma con el valor presente de

1. costo de operación de generación
2. costo de pérdidas de transmisión
3. costo de inversión en generación
4. costo de inversión en sistema de transmisión

DESCRIPCION DE RESTRICCIONES

BALANCE NODAL DE POTENCIA

Se requiere que para todos los nodos de la red en cada modo de operación y cada período se logre un equilibrio de potencia.

potencia entrada = potencia salida
nodo/modo/período

$$\sum_{\text{Tipo}} GR\ n,t,m,p + \sum FL\ n,n,m,p = DF(n,m,p) + DI\ n,m,p$$

DEMANDA INDUSTRIAL REUBICABLE

En cada modo de operación y para cada período se desea localizar en el sistema una demanda reubicada especificada.

suma demanda reubicable en nodos modo/período = demanda reubicable por sistema

$$\sum_{\text{Nodos}} DI\ n,m,p = DTR\ m,p$$

UTILIZACION EFECTIVA DE PLANTAS GENERADORAS

Se requiere que cada generador existente o que el modelo localize, sea utilizado en cada modo y cada período entre los límites impuestos por el factor de carga superior e inferior correspondiente.

(fci)(capacidad) $\leq$ generación $\leq$ (fcs)(capacidad)

nodo/tipo/modo/período

fcs factor de carga superior

fci factor de carga inferior

$$GR\ n,t,m,p \geq FACI(n,t,m,p) \left\{ GN\ n,t,p * AB + CAGI(n,t) \right\}$$

$$GR\ n,t,m,p \leq FACS(n,t,m,p) \left\{ GN\ n,t,p * AB + CAGI(n,t) \right\}$$

LIMITE DE LOCALIZACION DE GENERACION

En cada nodo del sistema y para cada período se desea tener un límite de capacidad de generación a instalar.

total de capacidad a instalar en un nodo $\leq$ límite especificado

nodo/período

$$\sum_{\text{Tipo}} GN\ n,t,p * AB \leq GLIM(n,p)$$

CAPACIDAD DE ENLACES DE TRANSMISION

Se requiere exista capacidad de enlace para transmitir el flujo de potencia en líneas en cada modo de operación y para cada período.

flujo de potencia por enlace $\leq$ capacidad de transmisión del enlace

enlace/modo/período

$$FL\ n,n,m,p \leq \sum_{p} \sum_{\text{Voltaje}} Ln,n,v,p * BA+CAPEL(n,n)$$

RESERVA GENERACION EN DEMANDA MAXIMA

En cada período se desea tener una reserva de generación de sistema para la demanda máxima del sistema.

generación total del sistema en demanda máxima. $\leq$ (factor de reserva) (capacidad de generación instalada en el sistema)

período

$$\sum_{\text{Nodos}} \sum_{\text{Tipos}} GR\ n,t,m,p \leq$$

$$FACRE \left\{ \sum_{\text{Nodos}} \sum_{\text{tipos}} \sum_{p} GN\ n,t,p * AB\text{-}CAGI(n,t) \right\}$$

ENERGIA HIDRAULICA DISPONIBLE

De acuerdo con los recursos hidroeléctricos disponibles, se tiene un límite para efectuar inversiones en plantas hidroeléctricas.

energía hidráulica utilizada $\leq$ límite especificado

nodo/período

POTENCIA HIDRAULICA MAXIMA

La potencia hidráulica utilizada debe ser menor que la instalada.

potencia hidráulica en demanda máxima $\leq$ potencia hidráulica máxima

DESCRIPCION FUNCION OBJETIVO

COSTO OPERACION

C_o = (potencia generada)(duración)(costo energía generada)

nodo/tipo/modo/transmisión

C_o = COOP(n,t,m)*DURA(m) * GRn,t,m,p/DE

COSTO PERDIDAS TRANSMISION

C_p = (costo energía)(duración)(flujo de potencia)(factor de pérdidas)(distancia)

enlace/modo/período

C_p = COPE(m)* DLON(k) * DURA(m) * FL n,n,m,p/ DE

COSTO INVERSION EN GENERACION

C_g = (costo por unidad de potencia instalada.) (número de unidades) (capacidad estándar a instalar)

nodo/tipo/período

C_g = CAPCO(n,t) * AB * GN n,t,p / DE

COSTO INVERSION EN LINEAS

C_ℓ = (costo por unidad de longitud) (longitud del enlace) (número de líneas a instalar)

enlace/voltaje/período

C_ℓ = COSTOL(k,v) * DLON(k) * L n,n,v,p/ DE

METODO DE SOLUCION

El problema se formula con:

i) La función objetivo y restricciones lineales

ii) La localización de generadores y líneas con variables enteras 1/0

iii) La generación y el flujo de potencia en líneas con variables contínuas.

En la solución se utiliza un método de:

Programación Lineal Mixta

CARACTERISTICAS DEL MODELO

Se ha diseñado para tener flexibilidad en la formulación de diferentes condiciones del problema.

Las variables de decisión para inversiones en generación y/o sistema de transmisión se pueden designar enteras o contínuas a elección del usuario.

La especificación de planes de desarrollo en construcción o próximos a realizarse se logra fijando variables en el valor decidido.

Se designa por excepción en que nodos puede existir generación y de que tipos.

Se especifican los enlaces de transmisión factibles y los voltajes adecuados por enlace.

CALCULO DE LOS COSTOS REGIONALES

Para simplificar la exposición del cálculo de los costos regionales, supongamos que el modelo de programación lineal mixta descrito anteriormente tiene la siguiente estructura

$$\text{Min} \quad c_p^t x_p + c_e^t x_e \qquad (1)$$

Sujeto a

$$A_p x_p + A_e x_e = b \qquad (2)$$

$$x_p, x_e \geq 0 \qquad (3)$$

donde:

- x_p es el vector de variables asociadas al costo por potencia (Inversión en generación e inversión en líneas)
- x_e es el vector de incógnitas asociadas al costo por energía (Costo de operación y costo de pérdidas)
- c_p es el vector de costos asociado a x_p
- c_e es el vector de costos asociado a x_e
- $A = [A_p, A_e]$ es la matriz de restricciones, donde la submatriz A_p toma en cuenta los coeficientes asociados al vector x_p y la submatriz A_e los asociados a x_e

Nótese que en la formulación anterior se omiten las restricciones de integralidad de algunas variables, ello se hace solo para simplificar la exposición y no implica ninguna pérdida de generalidad.

En función de la formulación anterior es posible escribir la i-ésima restricción del problema como

$$(a_p)_{i.} x_p + (a_e)_{i.} x_e = b_i$$

donde $(a_p)_{i.}$ y $(a_e)_{i.}$ son respectivamente el i-ésimo renglón de la submatriz A_p y A_e .

En programación lineal se cumple que la variable dual μ_i asociada a la i-ésima restricción mide marginalmente el incremento sufre el valor óptimo de la función objetivo por el aumento de una unidad en b_i, es decir, indica el valor que la función objetivo tomaría en un problema que fuera exactamente igual al definido en las ecuaciones (1), (2) y (3), excepto que el i-ésimo elemento del vector b tuviera una unidad más:

$$(a_p)_{i.} x_p + (a_e)_{i.} x_e = b_i + 1$$

Aprovechando esta propiedad de las variables duales es posible cuantificar el costo que para el sector eléctrico tiene el suministrar un KWh adicional en un nodo n, en un modo de operación - (pico y fuera de pico) m y un período p, pues si la i-ésima restricción representa la satisfacción de la demanda del nodo n en el modo de operación m, y el período de p, entonces la variable dual μ_i que podemos denotar más claramente como $\mu_{nmp,}$ cuantifica el costo de la satisfacción de esta unidad adicional.

El valor μ_{npm} cuantifíca el costo de la satísfacción de un KWh adicional en el nodo n, modo m y período p, sin embargo, el satisfacer en KWh adicional implica erogaciones en dos conceptos diferentes que definen la estructura tarifaria actual: es necesario hacer erogaciones por inversión en generación y en líneas, estas erogaciones se consideran imputables al costo por potencia de ese KWh adicional a satisfacer; y además es necesario hacer erogaciones adicionales motivadas por el costo de operación y las pérdidas en transmisión, estos últimos gastos sé consideran asociados al costo por energía del KWh adicional, es decir es necesario dividir la variable μ_{nmp} en dos partes que cuantifiquen en forma separada el costo imputable a la potencia y el costo imputable a la energía motivados por la satisfacción de un KWh adicional, y que estas dos componentes adicionadas den como resultado el valor de la variable μ_{nmp}.

En el siguiente párrafo se menciona la manera de hacer esta división.

Las variables duales se calculan con la fórmula matricial siguiente:

$$\mu = c_B B^{-1}$$

donde B^{-1} es la inversa de la base del método simplex y c_B el costo en la función objetivo de los vectores básicos.

Como ya se mencionó anteriormente el vector de costos de la función objetivo (1) está compuesto de dos partes, el costo imputable a la potencia y el costo imputable a la energía, de modo que si la i-ésima variable dual μ_i está dada por

$$\mu_i = \sum_i c_{B_i} \beta_{.i}$$

donde $\beta_{.i}$ es la i-ésima columna de B^{-1} este valor se puede descomponer en dos partes, en la primera acumular todos los productos $c_{B_i} \beta_{.i}$ en los que c_{B_i} es un elemento del vector c_p, llamemos a esta suma π_i; en la segunda parte que llamaremos η_i se acumulan todos los productos $c_{B_i} \beta_{.i}$ en los que c_{B_i} es elemento del vector c_e. Es obvio que

$$\mu_i = \pi_i + \eta_i$$

Usando la misma notación anterior, π_{nmp} mide el costo imputable de la potencia y η_{nmp} mide el costo imputable a la energía que para el sector tiene una demanda de un KWh adicional en el nodo n, nodo m y período p.

Una vez que se han calculado los valores π_{nmp} y η_{nmp}, es necesario analizarlos, pues no son valores que puedan ser usados de inmediato para definir un costo regional de la energía, es posible que por las características propias de la programación lineal, los valores π y η para un mismo nodo y un mismo modo tengan cambios bruscos de un período a otro, por lo que habrá de "suavizarlos" para que tengan un comportamiento adecuado en el horizonte de estudio desde el punto de vista de su evolución.

Una vez "suavizadas" las variables es necesario convertirlar a pesos, ya que los costos marginales no pueden ser usados di-

rectamente para tarificar, pues si el sistema está subequipado, los costos marginales serán excesivamente altos, y serán sumamente bajos, o tal vez valgan cero si el sistema esta sobre-equipado

Cuando los costos han sido convertidos a pesos es necesario adicionarles los costos que el modelo no considera, éstos son por personal, por materiales y mantenimiento y otros. Posteriormente se considerará la manera de agregar estos costos.

Como observación final en cuanto a las variables duales, hay que mencionar que son útiles cuando el sistema se encuentra en el óptimo que indica el modelo, y que si se permite al modelo hacer un desarrollo óptimo en función de los datos que se le alimenten, este desarrollo puede ser diferente al que se menciona en el POISE o en los estudios a más largo plazo desarrollados con el MNI-PROLOG-LOG.

Por otra parte, "forzar" al modelo mediante restricciones adicionales para que obtenga un plan de expansión dado altera el valor de las variables duales y se pierden su significado.

OBTENCION DEL COSTO TOTAL DEL KW Y DEL KWH REGIONAL.

La asignación de costos en forma regional para el KW y KWH en el nivel de consumidores industriales, se puede hacer en forma práctica tomando en cuenta las distintas componentes que ocasionan dicho costo. Por ejemplo, si C_p y C_e representan los costos por KW y los costos por KWH respectivamente:

$$C_p = C_{vp} + C_{fp}$$

$$C_e = C_{ve}$$

en donde los índices p, e, v y f son los índices de potencia, energía, costos variables y costos fijos respectivamente.

A su vez, los costos variables están definidos por 4 componentes que son:

C_c = Costo de combustible \$/KWH
C_j = Costo por pérdidas \$/KWH
C_{ip} = Costo de inversión en plantas \$/KW
C_{il} = Costo de inversión en líneas \$/KW

Entonces:

$$C_{vp} = C_{ip} + C_{il}$$

"Costos Variables"

$$C_{ve} = C_c + C_j$$

En estas condiciones, es facil identificar que las 4 - componentes de costos anteriores forman los términos correspon - dientes de la función objetivo, para los cuales, a través del proce - so de optimización, es necesario encontrar su distribución entre to dos los nodos.

Por otro lado, los costos fijos incurridos en el suministro son:

C_{sp} = Costo del personal (sueldos y prestaciones) \$/KW.

C_{mp} = Costo por materiales y mantenimiento, \$/KW.

C_o = Costo por otros conceptos, \$/KW.

Entonces:

$$C_{fp} = C_{sp} + C_{mp} + C_o$$

En estas condiciones, es necesario estimar los costos fijos y distribuirlos uniformemente en todos los nodos, para adicionarlos a los costos variables obtenidos en los procesos de optimización, separación y suavizado de costos marginales ya descrito.

Por lo tanto, el ajuste proporcional de los costos marginales suavizados π_i y η_i de potencia y energía respectivamente, para obtener los costos totales α_{nmp} y β_{nmp} del KW y del KWH, se hace como sigue:

$$\alpha_{nmp} = K_{1p}\ \pi_{nmp} + C_{fp}$$

$$\beta_{nmp} = K_{2p}\ \eta_{nmp}$$

en donde K_{1p} y K_{2p} se obtienen de las relaciones siguientes:

$$K_{1p} = \frac{\text{Costo Total de Inversión en Plantas y Líneas en el período p}}{\sum_{m=1}^{M} \sum_{n=1}^{N} \pi_{nmp}\ P_{nmp}}$$

$$K_{2p} = \frac{\text{Costo Total de Combustible en el período p}}{\sum_{m=1}^{M} \sum_{n=1}^{N} \eta_{nmp}\ E_{nmp}}$$

P_{nmp} y E_{nmp} son las potencias y energías suministradas al sistema en el nodo n, modo de operación m y período p.

NOMENCLATURA

Indices:

Nodos	n
Modos	m
Períodos	p
Tipos de generación	t
Voltajes de transmisión	v
Líneas	k

Datos y variables

Número de nodos	NN
Número de modos	NM
Número de períodos	NP
Número de tipos de generación	NT
Número de voltajes de transmisión/enlace	NV
Número de enlaces	NE

Costo de operación por generador	COOP(n,t,m)	$/(MW-HR)
Duración del modo de operación	DURA(m)	HRS.
Potencia activa generada	GRn,t,m,p	MW
Costo de pérdidas	COPE(m)	$/(MW-HR)
Longitud de la línea	DLON(k)	U.L.
Flujo de potencia en líneas	FLn,m,p	MW
Costo de inversión en generación	CAPCO(n,t)	$/MW

Capacidad estándar de generación	AB	MW
Costo de inversión de líneas	COSTOL(k,v)	$/KM
Variable de decisión para instalar líneas	Ln,n,v,p	
Demanda fija	DF(n,m,p)	MW
Demanda reubicable	DIn,m,p	MW
Demanda total reubicable	DTR(m,p)	MW
Factor de carga superior para generador	FACS(n,t,m,p)	
Factor de carga inferior para generador	FACI(n,t,m,p)	
Límite de capacidad de generación a instalar	GLIM(n,p)	MW
Capacidad de Generación existente	GAGI(n,t)	MW
Capacidad de Transmisión existente	CAPEL(n,n)	MW
Variable de decisión para instalar generación	GN n,t,m	
Factor de descuento	DE	

centro de educación continua
división de estudios superiores
facultad de ingeniería, unam

TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS

"PROGRAMACION LINEAL"

Ing. Cosme Urdaibay Z.

enero, 1979

and

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad i = 1, 2, \ldots, m$$

b. Minimize

$$cX \tag{1.1}$$

subject to

$$X \geq O \tag{1.2}$$

and

$$AX = b \tag{1.3}$$

where $c = (c_1, c_2, \ldots, c_n)$ is a row vector, $X = (x_1, x_2, \ldots, x_n)$ is a column vector, $A = (a_{ij})$, $b = (b_1, b_2, \ldots, b_m)$ is a column vector, and O is an n-dimensional null column vector.

c. Minimize

$$cX$$

subject to

$$X \geq O$$

and

$$x_1P_1 + x_2P_2 + \cdots + x_nP_n = P_0$$

where P_j for $j = 1, 2, \ldots, n$ is the jth column of the matrix A and $P_0 = b$.

2. PROPERTIES OF A SOLUTION TO THE LINEAR-PROGRAMMING PROBLEM

In this section we shall state a number of standard definitions and describe the more important characteristics of a solution to the general linear-programming problem. Much of this material and that in Chap. 4 is contained in Dantzig [17] and Charnes, Cooper, and Henderson [12].

Definition 1. A *feasible solution* to the linear-programming problem is a vector $X = (x_1, x_2, \ldots, x_n)$ which satisfies conditions (1.2) and (1.3).

Definition 2*a*. A *basic solution* to (1.3) is a solution obtained by setting $n - m$ variables equal to zero and solving for the remaining m variables, provided that the determinant of the coefficients of these m variables is nonzero. The m variables are called *basic variables*.

Definition 2*b*. A *basic feasible solution* is a basic solution which also satisfies (1.2); that is, all basic variables are nonnegative.

Definition 3. A *nondegenerate basic feasible solution* is a basic feasible solution with *exactly* m positive x_i; that is, all basic variables are positive.

Definition 4. A *minimum feasible solution* is a feasible solution which also minimizes (1.1).

Unless otherwise stated, when we refer to a solution, we shall mean any feasible solution.

Definition 5. A *linear functional* $f(X)$ is a real-valued function defined on an n-dimensional vector space such that, for every vector $X = \alpha U + \beta V$, $f(X) = f(\alpha U + \beta V) = \alpha f(U) + \beta f(V)$ for all n-dimensional vectors U and V and all scalars α and β. For example, let $U = (9,3)$, $V = (6,6)$, $\alpha = \frac{1}{3}$, $\beta = -\frac{2}{3}$, and $f(X) = 2x_1 + x_2$. We have $\alpha U + \beta V = (-1,-3)$ and $f(\alpha U + \beta V) = -5$; $\alpha f(U) = 7$ and $\beta f(V) = -12$.

We note that the objective function (1.1) is a linear functional for those X satisfying (1.2) and (1.3).

Theorem 1. *The set of all feasible solutions to the linear-programming problem is a convex set.*

Proof. We need to show that every convex combination of any two feasible solutions is also a feasible solution. (The theorem is true, of course, if the set of solutions has only one element.) Assume there are at least two solutions X_1 and X_2. We have

$$AX_1 = b \qquad \text{for } X_1 \geq O$$

and

$$AX_2 = b \qquad \text{for } X_2 \geq O$$

For $0 \leq \alpha \leq 1$, let $X = \alpha X_1 + (1 - \alpha)X_2$ be any convex combination of X_1 and X_2. We note that all the elements of the vector X are nonnegative; that is, $X \geq O$. We then see that X is a feasible solution, for we have

$$AX = A[\alpha X_1 + (1 - \alpha)X_2] = \alpha AX_1 + (1 - \alpha)AX_2$$
$$= \alpha b + b - \alpha b = b$$

In a similar manner, one can prove that the sets of solutions to the inequalities (4.4) and the equalities (5.1) of Chap. 2 are convex sets.

We shall denote the convex set of solutions to the linear-programming problem by K. Since K is determined by the intersection of the finite

set of linear constraints (1.2) and (1.3), the boundary of K (if K is not void) will consist of sections of some of the corresponding hyperplanes. K will be a region of E_n and can either be void, a convex polyhedron, or a convex region which is unbounded in some direction. If K is void, then our problem does not have any solutions; if it is a convex polyhedron, then our problem has a solution with a finite minimum value for the objective function; and if K is unbounded, the problem has a solution, but the minimum *might* be unbounded. Valid linear-programming models should yield K's of the second or possibly the third type, as they are models of situations that have a number of possible solutions. By Theorem 1, if a problem has more than one solution, it has, in reality, an infinite number of solutions. Out of all these solutions, it is our task to determine the one which minimizes the corresponding objective function. This work is somewhat simplified by the results of Theorem 2 below. Before proceeding with this theorem, we should note the following: If K is a convex polyhedron, then K is the convex hull of the extreme points of K. That is, every feasible solution in K can be represented as a convex combination of the extreme feasible solutions in K. (By definition, a convex polyhedron has a finite number of extreme points.) An unbounded K also has a finite number of extreme points, but not all points of K can be represented as convex combinations of these extreme points. For ease in discussion, we can assume that all our problems have a K that is a convex polyhedron. As will be shown in later chapters, computational devices exist that determine whether K is void or whether a problem has an unbounded minimum. (For computational purposes, we always assume that K is a convex polyhedron.)

With the assumption that K is a convex polyhedron, we can surmise from the above discussion that we need only to look at the extreme points of the convex polyhedron in order to determine the minimum feasible solution. We prove this with the following theorem:

Theorem 2. *The objective function* (1.1) *assumes its minimum at an extreme point of the convex set* K *generated by the set of feasible solutions to the linear-programming problem. If it assumes its minimum at more than one extreme point, then it takes on the same value for every convex combination of those particular points.*

Proof. Since we have assumed K to be a convex polyhedron, K has a finite number of extreme points. In two dimensions, K might look like Fig. 3.1. Let us denote the objective function by $f(X)$, the extreme points by $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p$, and the minimum feasible solution by X_0. This means that $f(X_0) \leq f(X)$ for all X in K. If X_0 is an extreme point, the first part of the theorem is true. Suppose

X_0 is not an extreme point (as indicated in Fig. 3.1). We can then write X_0 as a convex combination of the extreme points of K, that is,

$$X_0 = \sum_{i=1}^{p} \alpha_i \bar{X}_i$$

for $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$. Then, since $f(X)$ is a linear functional, we have

$$f(X_0) = f\left(\sum_{i=1}^{p} \alpha_i \bar{X}_i\right) = f(\alpha_1 \bar{X}_1 + \alpha_2 \bar{X}_2 + \cdots + \alpha_p \bar{X}_p)$$
$$= \alpha_1 f(\bar{X}_1) + \alpha_2 f(\bar{X}_2) + \cdots + \alpha_p f(\bar{X}_p) = m \tag{2.1}$$

where m is the minimum of $f(X)$ for all X in K.

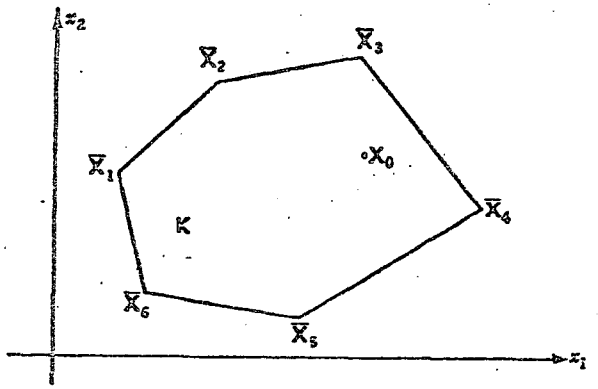


FIGURE 3.1

Since all $\alpha_i \geq 0$, we do not increase the sum (2.1) if we substitute for each $f(\bar{X}_i)$ the minimum of the values $f(\bar{X}_i)$. Let $f(\bar{X}_m) = \min_i f(\bar{X}_i)$. Substituting in (2.1) we have, since $\sum_i \alpha_i = 1$,

$$f(X_0) \geq \alpha_1 f(\bar{X}_m) + \alpha_2 f(\bar{X}_m) + \cdots + \alpha_p f(\bar{X}_m) = f(\bar{X}_m)$$

Since we assumed $f(X_0) \leq f(X)$ for all X in K, we must have

$$f(X_0) = f(\bar{X}_m) = m$$

Therefore, there is an extreme point, $\bar{X}_m$, at which the objective function assumes its minimum value.

To prove the second part of the theorem, let $f(X)$ assume its minimum at more than one extreme point, say at $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_q$. Here we have $f(\bar{X}_1) = f(\bar{X}_2) = \cdots = f(\bar{X}_q) = m$. If X is any

convex combination of the above $\bar{X}_i$, say

$$X = \sum_{i=1}^{q} \alpha_i \bar{X}_i$$

for $\alpha_i \geq 0$ and $\sum_i \alpha_i = 1$, then

$$\begin{aligned} f(X) &= f(\alpha_1 \bar{X}_1 + \alpha_2 \bar{X}_2 + \cdots + \alpha_q \bar{X}_q) \\ &= \alpha_1 f(\bar{X}_1) + \alpha_2 f(\bar{X}_2) + \cdots + \alpha_q f(\bar{X}_q) = \sum_i \alpha_i m \\ &= m \end{aligned}$$

The proof is now completed. By making the obvious changes, the theorem can be proved for the case where (1.1) is to be maximized. By Theorem 2, we need only to consider the extreme points of K in our search for a minimum feasible solution to the linear-programming problem.

Recall that a feasible solution is a vector $X = (x_1, x_2, \ldots, x_n)$, with all $x_i \geq 0$, such that

$$x_1 P_1 + x_2 P_2 + \cdots + x_n P_n = P_0$$

Assume we have found a set of k vectors that is linearly independent and that there exists a nonnegative combination of these vectors that is equal to P_0. Let this set of vectors be $P_1, P_2, \ldots, P_k$. We then have the following theorem:

Theorem 3. *If a set of $k \leq m$ vectors $P_1, P_2, \ldots, P_k$ can be found that is linearly independent and such that*

$$x_1 P_1 + x_2 P_2 + \cdots + x_k P_k = P_0$$

and all $x_i \geq 0$, then the point $X = (x_1, x_2, \ldots, x_k, 0, \ldots, 0)$ is an extreme point of the convex set of feasible solutions. Here X is an n-dimensional vector whose last $n - k$ elements are zero.

Proof. Suppose X is not an extreme point. Then, since X is a feasible solution, it can be written as a convex combination of two other points X_1 and X_2 in K. We have $X = \alpha X_1 + (1 - \alpha) X_2$ for $0 < \alpha < 1$. Since all the elements x_i of X are nonnegative and since $0 < \alpha < 1$, the last $n - k$ elements of X_1 and X_2 must also equal zero; that is,

$$X_1 = (x_1^{(1)}, x_2^{(1)}, \ldots, x_k^{(1)}, 0, \ldots, 0)$$
$$X_2 = (x_1^{(2)}, x_2^{(2)}, \ldots, x_k^{(2)}, 0, \ldots, 0)$$

Since X_1 and X_2 are feasible solutions, we have

$$AX_1 = b$$

and

$$AX_2 = b$$

Rewriting these equations in vector notation, we have

$$x_1^{(1)} P_1 + x_2^{(1)} P_2 + \cdots + x_k^{(1)} P_k = P_0$$

and

$$x_1^{(2)} P_1 + x_2^{(2)} P_2 + \cdots + x_k^{(2)} P_k = P_0$$

But $P_1, P_2, \ldots, P_k$ are linearly independent, and hence P_0 can be expressed as a *unique* linear combination in terms of $P_1, P_2, \ldots, P_k$.† This implies that $x_i = x_i^{(1)} = x_i^{(2)}$. Therefore, X cannot be expressed as a convex combination of two distinct points in K and must be an extreme point of K.

Theorem 4. *If $X = (x_1, x_2, \ldots, x_n)$ is an extreme point of K, then the vectors associated with positive x_i form a linearly independent set. From this it follows that, at most, m of the x_i are positive.*

Proof. Let the nonzero coefficients be the first k coefficients, so that $\sum_{i=1}^{k} x_i P_i = P_0$. We prove the main part of the theorem by contradiction. Assume that $P_1, P_2, \ldots, P_k$ are linearly dependent. Then there exists a linear combination of these vectors which equals the zero vector,

$$d_1 P_1 + d_2 P_2 + \cdots + d_k P_k = 0 \tag{2.2}$$

with at least one $d_i \neq 0$. From the hypothesis of the theorem, we have

$$x_1 P_1 + x_2 P_2 + \cdots + x_k P_k = P_0 \tag{2.3}$$

† This can be shown as follows: Let $P_1, P_2, \ldots, P_k$ be a set of linearly independent vectors and assume that we can represent the vector P_0 in terms of these vectors by two different linear combinations, for example,

$$e_1 P_1 + e_2 P_2 + \cdots + e_k P_k = P_0$$

and

$$f_1 P_1 + f_2 P_2 + \cdots + f_k P_k = P_0$$

To show that these combinations have to be identical, we subtract the second from the first to obtain

$$(e_1 - f_1) P_1 + (e_2 - f_2) P_2 + \cdots + (e_k - f_k) P_k = 0$$

By the definition of linear independence we must have each $e_i - f_i = 0$, which implies that $e_i = f_i$.

For some $d > 0$, we multiply (2.2) by d and add and subtract the result from (2.3) to obtain the two equations

$$\sum_{i=1}^{k} x_i P_i + d \sum_{i=1}^{k} d_i P_i = P_0$$

$$\sum_{i=1}^{k} x_i P_i - d \sum_{i=1}^{k} d_i P_i = P_0$$

We then have the two solutions to (1.3) (note that they might not be feasible solutions):

$$X_1 = (x_1 + dd_1, x_2 + dd_2, \ldots, x_k + dd_k, 0, \ldots, 0)$$

and

$$X_2 = (x_1 - dd_1, x_2 - dd_2, \ldots, x_k - dd_k, 0, \ldots, 0)$$

Since all $x_i > 0$, we can let d be as small as necessary, but still positive, to make the first k components of both X_1 and X_2 positive. Then X_1 and X_2 are feasible solutions. But $X = \frac{1}{2}X_1 + \frac{1}{2}X_2$, which contradicts the hypothesis that X is an extreme point. The assumption of linear dependence for the vectors $P_1, P_2, \ldots, P_k$ has thus led to a contradiction and hence must be false; i.e., the set of vectors $P_1, P_2, \ldots, P_k$ is linearly independent.

Since every set of $m + 1$ vectors in m-dimensional space is necessarily linearly dependent, we cannot have more than m positive x_i. For assume that we did. Then the above proof of the main part of the theorem would imply that there exist vectors $P_1, \ldots, P_m, P_{m+1}$ that are linearly independent.

We can, without any loss of generality, assume that the set of vectors $P_1, P_2, \ldots, P_n$ of the linear-programming problem always contains a set of m linearly independent vectors. If this property is not evident when a particular problem is being solved, the original set of vectors is augmented by a set of m linearly independent vectors, and we then seek a solution to the extended problem. This procedure will be explained in detail in the succeeding chapters.

Corollary 1. *Associated with every extreme point of* K *is a set of m linearly independent vectors from the given set* $P_1, P_2, \ldots, P_n$.

Proof. Theorem 4 has shown that there are $k \leq m$ such vectors. For $k = m$, the corollary is proved. Assume that $k < m$ and that we can find only additional vectors $P_{k+1}, \ldots, P_r$ such that the set

$$P_1, \ldots, P_k, P_{k+1}, \ldots, P_r$$

for $r < m$ is linearly independent. Then this implies that the remaining $n - r$ vectors are dependent on $P_1, \ldots, P_r$. But this contradicts the assumption that we always have a set of m linearly independent vectors in the given set of $P_1, \ldots, P_n$. Therefore, there must be m linearly independent vectors $P_1, \ldots, P_m$ associated with every extreme point, such that

$$\sum_{i=1}^{k} x_i P_i + \sum_{i=k+1}^{m} 0 P_i = P_0$$

We can sum up the preceding theorems by the following:

Theorem 5. $X = (x_1, x_2, \ldots, x_n)$ *is an extreme point of* K *if and only if the positive* x_j *are coefficients of linearly independent vectors* P_j *in*

$$\sum_{j=1}^{n} x_j P_j = P_0$$

As a result of the assumptions and theorems of this section, we have:

1. There is an extreme point of K at which the objective function takes on its minimum.
2. Every basic feasible solution corresponds to an extreme point of K.
3. Every extreme point of K has m linearly independent vectors of the given set of n associated with it.

From the above we can conclude that we need only investigate extreme-point solutions and hence only those feasible solutions generated by m linearly independent vectors. Since there are at most $\binom{n}{m}$ sets of m linearly independent vectors from the given set of n, the value $\binom{n}{m}$ is the upper bound to the number of possible solutions to the problem.[1] For large n and m it would be an impossible task to evaluate all the possible solutions and select one that minimizes the objective function. What is required is a computational scheme that selects, in an orderly fashion, a small subset of the possible solutions which converges to a minimum solution. The *simplex procedure*, devised by G. B. Dantzig, is such a scheme.[2] This procedure finds an extreme point and determines whether it is the minimum. If it is not, the procedure finds a neighboring extreme point[3] whose corresponding value of the objective

[1] See Saaty [88] and Quandt and Kuhn [85c] for discussions of the upper bound to the number of possible solutions.

[2] The name simplex method is due to the use of the equation $\sum_j x_j = 1$ as a constraint in a geometric interpretation of this procedure, as described in Dantzig [17].

[3] Two extreme points are said to be neighbors if they are joined by a boundary of the convex polyhedron.

function is less than or equal to the preceding value. In a finite number of such steps (usually between m and $2m$), a minimum feasible solution is found. The simplex method makes it possible to discover whether the problem has no finite minimum solutions or no feasible solutions. It is a powerful scheme for solving *any* linear-programming problem.

Before going into the validity and full computational aspects of the simplex procedure, we wish to introduce the following computational element of the procedure.

3. GENERATING EXTREME-POINT SOLUTIONS

Here we assume that an extreme-point solution in terms of m vectors $\mathbf{P}_j$ of the original set of n vectors is known. We can let this set of m linearly independent vectors be the first m, and let

$$\mathbf{X} = (x_1, x_2, \ldots, x_m, 0, \ldots, 0)$$

be the solution vector. We then have

$$x_1\mathbf{P}_1 + x_2\mathbf{P}_2 + \cdots + x_m\mathbf{P}_m = \mathbf{P}_0 \tag{3.1}$$

where all $x_i \geq 0$. With this information, the problem is to determine, in a computationally efficient manner, a new extreme-point solution. (We shall, of course, assume that a different extreme solution exists.) Since the vectors $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_m$ are linearly independent, they form a basis in m-dimensional vector space. We can then express every vector of the given n as a linear combination of these basis vectors. We can write

$$\sum_{i=1}^{m} x_{ij}\mathbf{P}_i = \mathbf{P}_j \qquad j = 1, \ldots, n$$

Assume that some vector not in the given basis, say $\mathbf{P}_{m+1}$, has at least one element $x_{i,\,m+1} > 0$ in the expression

$$x_{1,\,m+1}\mathbf{P}_1 + x_{2,\,m+1}\mathbf{P}_2 + \cdots + x_{m,\,m+1}\mathbf{P}_m = \mathbf{P}_{m+1} \tag{3.2}$$

Let θ be any number, and multiply (3.2) by θ and subtract the result from (3.1) to obtain

$$(x_1 - \theta x_{1,\,m+1})\mathbf{P}_1 + (x_2 - \theta x_{2,\,m+1})\mathbf{P}_2 + \cdots + (x_m - \theta x_{m,\,m+1})\mathbf{P}_m + \theta\mathbf{P}_{m+1} = \mathbf{P}_0 \tag{3.3}$$

The vector $\mathbf{X}' = (x_1 - \theta x_{1,\,m+1}, x_2 - \theta x_{2,\,m+1}, \ldots, x_m - \theta x_{m,\,m+1}, \theta)$ is a solution to the problem, and if all the elements of $\mathbf{X}'$ are nonnegative, $\mathbf{X}'$ is a feasible solution. Since we want $\mathbf{X}'$ to be a feasible solution different from $\mathbf{X}$, we restrict θ to be greater than zero.[1] With this restriction, all the elements of $\mathbf{X}'$ that have a negative or zero $x_{i,\,m+1}$ will also have a nonnegative $x_i - \theta x_{i,\,m+1}$. We need only concern ourselves with those elements having a positive $x_{i,\,m+1}$. We wish to find a $\theta > 0$ such that

$$x_i - \theta x_{i,\,m+1} \geq 0 \tag{3.4}$$

for all $x_{i,\,m+1} > 0$.

From (3.4) we have

$$\frac{x_i}{x_{i,\,m+1}} \geq \theta$$

and hence any θ for which

$$0 < \theta \leq \min_i \frac{x_i}{x_{i,\,m+1}}$$

will give a feasible solution for (3.3). However, as we are looking for an extreme-point solution, we know by our theorems in Sec. 2 that we cannot have all the $m + 1$ elements of $\mathbf{X}'$ positive. We then must force at least one of the elements of $\mathbf{X}'$ to be exactly equal to zero. We see that, if we let

$$\theta = \theta_0 = \min_i \frac{x_i}{x_{i,\,m+1}}$$

for $x_{i,\,m+1} > 0$, then the element in $\mathbf{X}'$ for which this minimum is attained will reduce to zero. Let this element be the first, that is,

$$\theta_0 = \min_i \frac{x_i}{x_{i,\,m+1}} = \frac{x_1}{x_{1,\,m+1}}$$

We have now obtained a new feasible solution

$$x_2'\mathbf{P}_2 + x_3'\mathbf{P}_3 + \cdots + x_m'\mathbf{P}_m + x_{m+1}'\mathbf{P}_{m+1} = \mathbf{P}_0$$

where

$$x_i' = x_i - \theta_0 x_{i,\,m+1} \qquad i = 2, \ldots, m$$

$$x_{m+1}' = \theta_0$$

[1] The reader should note that this restriction causes us to assume (for illustrative purposes) that associated with an $x_{i,\,m+1} > 0$ is an $x_i > 0$. This, in general, is not the case and will be true only if nondegeneracy is assumed for all basic feasible solutions. As discussed in Chap. 4, a value of $\theta = 0$ is an acceptable one; i.e., some $x_i = 0$ for an $x_{i,\,m+1} > 0$. With $\theta = 0$, the transformation retains the old extreme point but selects a new feasible basis. In sum, the simplex procedure allows for all values of $\theta \geq 0$. The reader should review the applicability of the above discussion to the case $\theta = \min_i (x_i/x_{i,\,m+1}) = 0$ for $x_{i,\,m+1} > 0$.

If all the $x_{i,\,m+1}$ had been equal to or less than zero, then we would not have been able to select a positive θ that would have eliminated at least one of the vectors $\mathbf{P}_1, \ldots, \mathbf{P}_m$ from the basis. For this situation we obtain for *any* $\theta > 0$ a non-extreme-point feasible solution associated with the $m + 1$ vectors $\mathbf{P}_1, \ldots, \mathbf{P}_m, \mathbf{P}_{m+1}$. As will be shown in Chap. 4, this situation indicates that the problem does not have a finite minimum solution.

To show that $X' = (x'_2, \ldots, x'_m, x'_{m+1})$ is an extreme point, we have to prove $\mathbf{P}_2, \ldots, \mathbf{P}_m, \mathbf{P}_{m+1}$ are linearly independent. Assume they are linearly dependent. We can then find (from the definition of linear dependence)

$$d_2\mathbf{P}_2 + d_3\mathbf{P}_3 + \cdots + d_m\mathbf{P}_m + d_{m+1}\mathbf{P}_{m+1} = \mathbf{0} \tag{3.5}$$

where not all $d_i = 0$. Since any subset of a set of linearly independent vectors is also a set of linearly independent vectors, $\mathbf{P}_2, \ldots, \mathbf{P}_m$ are linearly independent. This implies that $d_{m+1} \neq 0$. From (3.5) we have

$$e_2\mathbf{P}_2 + e_3\mathbf{P}_3 + \cdots + e_m\mathbf{P}_m = \mathbf{P}_{m+1} \tag{3.6}$$

where

$$e_i = \frac{-d_i}{d_{m+1}}$$

Subtracting (3.6) from (3.2), we obtain

$$x_{1,\,m+1}\mathbf{P}_1 + (x_{2,\,m+1} - e_2)\mathbf{P}_2 + (x_{3,\,m+1} - e_3)\mathbf{P}_3 + \cdots + (x_{m,\,m+1} - e_m)\mathbf{P}_m = \mathbf{0} \tag{3.7}$$

Since $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_m$ are linearly independent, all the coefficients in (3.7) must equal zero. But $x_{1,\,m+1}$ was assumed to be positive. Hence the assumption of linear dependence for $\mathbf{P}_2, \ldots, \mathbf{P}_{m+1}$ has led to a contradiction, and these vectors must be linearly independent.

In order to continue this process of obtaining new extreme feasible solutions, we need the representation of any vector not in the new basis $\mathbf{P}_2, \mathbf{P}_3, \ldots, \mathbf{P}_{m+1}$ in terms of this basis. From (3.2) we have

$$\mathbf{P}_1 = \frac{1}{x_{1,\,m+1}}(\mathbf{P}_{m+1} - x_{2,\,m+1}\mathbf{P}_2 - \cdots - x_{m,\,m+1}\mathbf{P}_m) \tag{3.8}$$

Let

$$\mathbf{P}_j = x_{1j}\mathbf{P}_1 + x_{2j}\mathbf{P}_2 + \cdots + x_{mj}\mathbf{P}_m \tag{3.9}$$

be any vector not in the new basis. Substitute the expression (3.8) for $\mathbf{P}_1$ in (3.9) to obtain

$$\mathbf{P}_j = \left(x_{2j} - \frac{x_{1j}}{x_{1,\,m+1}}x_{2,\,m+1}\right)\mathbf{P}_2 + \left(x_{3j} - \frac{x_{1j}}{x_{1,\,m+1}}x_{3,\,m+1}\right)\mathbf{P}_3 + \cdots + \left(x_{mj} - \frac{x_{1j}}{x_{1,\,m+1}}x_{m,\,m+1}\right)\mathbf{P}_m + \frac{x_{1j}}{x_{1,\,m+1}}\mathbf{P}_{m+1}$$

The reader will note that the formulas for complete elimination required in Exercise 8, Chap. 2, are equivalent to the transformation that describes $\mathbf{P}_0$ and $\mathbf{P}_j$ in terms of the new basis. The procedure for obtaining new extreme-point solutions is that of selecting a new variable to be introduced into the system, determining which variable has to be removed from the solution in order to preserve feasibility, and applying the complete elimination formulas to obtain the new solution and the new representations of the vectors not in the basis. The criterion used to determine which variable is to be introduced into the solution is a feature of the simplex procedure and will be considered in Chap. 4.

Example. We are given the following set of equations:

$\mathbf{P}_1$	$\mathbf{P}_2$	$\mathbf{P}_3$	$\mathbf{P}_4$	$\mathbf{P}_5$	$\mathbf{P}_6$	$\mathbf{P}_0$
$3x_1 -$	$x_2 +$	$2x_3 +$	x_4			$= 7$
$2x_1 -$	$4x_2$			$+ x_5$		$= 12$
$-4x_1 -$	$3x_2 +$	$8x_3$			$+ x_6$	$= 10$

We have as an initial extreme-point solution $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, $x_4 = 7$, $x_5 = 12$, $x_6 = 10$, which in vector notation is given by

$$7\mathbf{P}_4 + 12\mathbf{P}_5 + 10\mathbf{P}_6 = \mathbf{P}_0 \tag{3.10}$$

Here the basis vectors $\mathbf{P}_4$, $\mathbf{P}_5$, $\mathbf{P}_6$ are unit vectors. We wish to introduce vector $\mathbf{P}_1$ to obtain another extreme-point solution. The representation of $\mathbf{P}_1$ in terms of the basis vectors is simply

$$3\mathbf{P}_4 + 2\mathbf{P}_5 - 4\mathbf{P}_6 = \mathbf{P}_1 \tag{3.11}$$

that is,

$$x_{41} = 3 \qquad x_{51} = 2 \qquad x_{61} = -4$$

If we multiply (3.11) by θ and subtract the result from (3.10), we have

$$(7 - 3\theta)\mathbf{P}_4 + (12 - 2\theta)\mathbf{P}_5 + (10 + 4\theta)\mathbf{P}_6 + \theta\mathbf{P}_1 = \mathbf{P}_0 \tag{3.12}$$

Since $x_{41} = 3$ and $x_{51} = 2$ are both positive, we determine θ_0 by evaluating, for these positive x_{i1},

$$\theta = \theta_0 = \min \frac{x_i}{x_{i1}} = 7/3\dagger$$

Substituting this value in (3.12), we eliminate $\mathbf{P}_4$ from the basis to obtain

$$2\tfrac{2}{3}\mathbf{P}_5 + 5\tfrac{8}{3}\mathbf{P}_6 + \tfrac{7}{3}\mathbf{P}_1 = \mathbf{P}_0$$

† We form the ratios for those i in the current solution. Here $i = 4, 5, 6$.

or the extreme-point solution $x_1 = 7/3, x_2 = 0, x_3 = 0, x_4 = 0, x_5 = 22/3, x_6 = 58/3$.

If, instead of P_1, we tried in a similar manner to obtain an extreme solution with P_2, where

$$-P_4 - 4P_5 - 3P_6 = P_2$$

we would have developed the following expression for P_0 in terms of P_4, P_5, P_6, P_2:

$$(7 + \theta)P_4 + (12 + 4\theta)P_5 + (10 + 3\theta)P_6 + \theta P_2 = P_0 \qquad (3.13)$$

From (3.13) we see that any $\theta > 0$ yields a feasible solution $x_1 = 0$, $x_2 = \theta$, $x_3 = 0$, $x_4 = 7 + \theta$, $x_5 = 12 + 4\theta$, $x_6 = 10 + 3\theta$. Here, since all $x_{i2} < 0$, we do not obtain a new extreme-point solution.

A more efficient way of interpreting the problem is as a transformation accomplished by the elimination procedure. Here we detach the coefficients of the equations and set up the following tableau:

P_1	P_2	P_3	P_4	P_5	P_6	P_0	θ
③	−1	2	1	0	0	7	$7/3 = \theta_0$
2	−4	0	0	1	0	12	6
−4	−3	8	0	0	1	10	

As we want to introduce P_1 into the basis, we again form the ratios x_i/x_{i1} for $x_{i1} > 0$. Since $\theta_0 = 7/3$ is the minimum of these ratios, we let the element 3 of P_1 be the pivot element of the elimination procedure, as denoted by the circle. That is, we shall eliminate x_1 from all the equations except the first. If we carry out the elimination transformation, we obtain a new tableau. Here $x_1 = 7/3$, $x_5 = 22/3$, $x_6 = 58/3$, and $x_2 = x_3 = x_4 = 0$.

P_1'	P_2'	P_3'	P_4'	P_5'	P_6'	P_0'	θ
1	$-1/3$	$2/3$	$1/3$	0	0	$7/3$	
0	$-10/3$	$-4/3$	$-2/3$	1	0	$22/3$	
0	$-13/3$	$32/3$	$4/3$	0	1	$58/3$	

We now have a basis of P_1, P_5, P_6, with P_2, P_3, P_4 explicitly given in terms of these basis vectors, that is,

$$-1/3P_1 - 10/3P_5 - 13/3P_6 = P_2$$
$$2/3P_1 - 4/3P_5 + 32/3P_6 = P_3$$
$$1/3P_1 - 2/3P_5 + 4/3P_6 = P_4$$

Hence if we wanted to obtain an extreme-point solution with P_3 in the basis, we could start with the second tableau and determine θ_0 as before and transform this tableau by the elimination formulas. The resulting tableau will yield the representation of the vectors not in the basis in terms of the new basis vectors.

EXERCISES

1. Find all basic feasible solutions for the equations

$$2x_1 + 6x_2 + 2x_3 + x_4 = 3$$
$$6x_1 + 4x_2 + 4x_3 + 6x_4 = 2$$

and determine the associated general convex combination of extreme-point solutions.

2. Construct and graph linear-programming problems in three variables with a unique extreme-point optimum solution, three extreme-point optimum solutions, and four extreme-point optimum solutions.

3. In the discussion in Sec. 3, the extreme-point solution (3.1) can be degenerate. Discuss the computation of θ and the transformation to a new basic feasible solution if some $x_i = 0$. Can there be any assurance that $\theta_0 > 0$? Also discuss and interpret the situation where the selection of $\min (x_i/x_{ij})$ for $x_{ij} > 0$ is not unique.

4. The following set of equations has a given extreme-point solution $X = (x_1, x_2, x_3) = (4,3,6)$. By the algebraic procedure of Sec. 3, obtain two basic solutions for the bases P_1, P_2, P_4 and P_2, P_3, P_4. In each case, determine the expression for P_5 and the vector eliminated in terms of the new basis.

P_1	P_2	P_3	P_4	P_5	P_0
x_1			$+2x_4$	$- x_5 = 4$	
	x_2		$- x_4$	$+ x_5 = 3$	
		x_3	$+3x_4$	$-2x_5 = 6$	

5. Do Exercise 4 by setting up the tableau and applying the general formulas developed for the complete elimination procedure in Exercise 8 of Chap. 2. For both bases, the variable x_4 is eliminated from all the equations except one.

6. For the 3×3 matrix formed by vectors P_1, P_2, P_4 of Exercise 4 compute its inverse by means of the adjoint matrix as described in Chap. 2. Do the same for vectors P_2, P_3, P_4.

7. Given the following set of equations:

$$x_1 + 4x_2 - x_3 = 3$$
$$5x_1 + 2x_2 + 3x_3 = 4$$

Determine the basic feasible solution involving x_1 and x_3. Do basic feasible solutions exist for x_1 and x_2 and for x_2 and x_3?

Discuss the graphical solution to this problem if each column of the matrix is assumed to be a vector in two-dimensional space.

CHAPTER 4 / THE SIMPLEX COMPUTATIONAL PROCEDURE

We shall next discuss and prove the validity of the basic elements of the simplex procedure and related computational algorithms. By the simplex procedure, we can, once any basic (extreme-point) feasible solution has been determined, obtain a minimum feasible solution in a finite number of steps. These steps, or iterations, consist in finding a new feasible solution whose corresponding value of the objective function is less than the value of the objective function for the preceding solution. This process is continued until a minimum solution has been reached. From the discussion of Chap. 3, we have that all extreme-point solutions, and especially the minimum solution, have m linearly independent vectors associated with them. We then limit our search to those solutions that are generated by m linearly independent vectors. We note that there are a finite number of such solutions.

1. DEVELOPMENT OF A MINIMUM FEASIBLE SOLUTION

We assume that the linear-programming problem is feasible, that every basic feasible solution is nondegenerate, and that we are given a basic feasible solution.[1] These assumptions, as will be discussed later, are made without any loss in generality. Let the given solution be $\mathbf{X}_0 = (x_{10}, x_{20}, \ldots, x_{m0})$† and the associated set of linearly independent vectors be $\mathbf{P}_1, \mathbf{P}_2, \ldots \mathbf{P}_m$. We then have

$$x_{10}\mathbf{P}_1 + x_{20}\mathbf{P}_2 + \cdots + x_{m0}\mathbf{P}_m = \mathbf{P}_0 \tag{1.1}$$

$$x_{10}c_1 + x_{20}c_2 + \cdots + x_{m0}c_m = z_0 \tag{1.2}$$

[1] The definitions of these terms are given in Sec. 2 of Chap. 3.

† In order to generalize the simplex transformations we now denote the solution vector $\mathbf{X} = (x_1, x_2, \ldots, x_m)$ by the vector $\mathbf{X}_0 = (x_{10}, x_{20}, \ldots, x_{m0})$. We should note that the remaining $n - m$ values of the solution vector have arbitrarily been set equal to zero.

where all $x_{i0} > 0$, the c_i are the cost coefficients of the objective function, and z_0 is the corresponding value of the objective function for the given solution. Since $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_m$ are linearly independent, we can express any vector from the set $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_n$ in terms of $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_m$. Let $\mathbf{P}_j$ be given by

$$x_{1j}\mathbf{P}_1 + x_{2j}\mathbf{P}_2 + \cdots + x_{mj}\mathbf{P}_m = \mathbf{P}_j \qquad j = 1, \ldots, n \tag{1.3}$$

and *define*

$$x_{1j}c_1 + x_{2j}c_2 + \cdots + x_{mj}c_m = z_j \qquad j = 1, \ldots, n \tag{1.4}$$

where the c_i are the cost coefficients corresponding to the $\mathbf{P}_i$.

Theorem 1. *If, for any fixed j, the condition $z_j - c_j > 0$ holds, then a set of feasible solutions can be constructed such that $z \leq z_0$ for any member of the set, where the lower bound of z is either finite or infinite. (z is the value of the objective function for a particular member of the set of feasible solutions.)*

Case I. If the lower bound is finite, a new feasible solution consisting of exactly m positive variables can be constructed whose value of the objective function is less than the value for the preceding solution.

Case II. If the lower bound is infinite, a new feasible solution consisting of exactly $m + 1$ positive variables can be constructed whose value of the objective function can be made arbitrarily small.

The following analysis applies to the proof of both cases:

Multiplying (1.3) by some number θ and subtracting from (1.1), and similarly multiplying (1.4) by the same θ and subtracting from (1.2), for $j = 1, 2, \ldots, n$ we get

$$(x_{10} - \theta x_{1j})\mathbf{P}_1 + (x_{20} - \theta x_{2j})\mathbf{P}_2 + \cdots + (x_{m0} - \theta x_{mj})\mathbf{P}_m + \theta\mathbf{P}_j = \mathbf{P}_0 \tag{1.5}$$

$$(x_{10} - \theta x_{1j})c_1 + (x_{20} - \theta x_{2j})c_2 + \cdots + (x_{m0} - \theta x_{mj})c_m + \theta c_j = z_0 - \theta(z_j - c_j) \tag{1.6}$$

where θc_j has been added to both sides of (1.6). If all the coefficients of the vectors $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_m, \mathbf{P}_j$ in (1.5) are nonnegative, then we have determined a new feasible solution whose value of the objective function is, by (1.6), $z = z_0 - \theta(z_j - c_j)$. Since the variables $x_{10}, x_{20}, \ldots, x_{m0}$ in (1.5) are all positive, it is clear, from our discussion in Sec. 3 of Chap. 3, that there is a value of $\theta > 0$ (either finite or infinite) for which the coefficients of the vectors in (1.5) remain positive. From the assumption that, for a fixed j, $z_j - c_j > 0$, we have

$$z = z_0 - \theta(z_j - c_j) < z_0$$

for $\theta > 0$. We see that in either event a new feasible solution can be obtained whose corresponding value of the objective function is less than the value for the preceding solution.

The proof of Case I follows:

If, for the fixed j, at least one $x_{ij} > 0$ in (1.3) for $i = 1, 2, \ldots, m$, the largest value of θ for which all coefficients of (1.5) remain nonnegative is given by

$$\theta_0 = \min_i \frac{x_{i0}}{x_{ij}} > 0 \tag{1.7}$$

for $x_{ij} > 0$ (see Sec. 3 of Chap. 3). Since we assumed that the problem is nondegenerate, i.e., that all basic feasible solutions have m positive elements, the minimum in (1.7) will be obtained for a unique i. If θ_0 is substituted for θ in (1.5) and (1.6), the coefficient corresponding to this unique i will vanish. We have then constructed a new basic feasible solution consisting of P_j and $m - 1$ vectors of the original basis. This new basis can be used as the previous one. If a new $z_j - c_j > 0$ and a corresponding $x_{ij} > 0$, another solution can be obtained which has a smaller value of the objective function. This process will continue either until all $z_j - c_j \leq 0$, or until, for some $z_j - c_j > 0$, all $x_{ij} \leq 0$. If all $z_j - c_j \leq 0$, the process terminates.

For Case II we have:

If at any stage we have, for some j, $z_j - c_j > 0$ and all $x_{ij} \leq 0$, then there is no upper bound to θ and the objective function has a lower bound of $-\infty$. We see for this case that, for any $\theta > 0$, all the coefficients of (1.5) are positive. We then have a feasible solution consisting of $m + 1$ positive elements. Hence, by taking θ large enough, the corresponding value of the objective function given by the right-hand side of (1.6) can be made arbitrarily small.

Theorem 2. *If for any basic feasible solution* $\mathrm{X} = (x_{10}, x_{20}, \ldots, x_{m0})$ *the conditions* $z_j - c_j \leq 0$ *hold for all* $j = 1, 2, \ldots, n$, *then* (1.1) *and* (1.2) *constitute a minimum feasible solution.*[1]

[1] This optimality criterion is sometimes varied. For a minimization problem we could have computed, instead of the $z_j - c_j$, the numbers $c_j - z_j$ and selected as the vector to be introduced into the basis the one corresponding to the min $(c_j - z_j)$. An optimum has been reached when all $c_j - z_j \geq 0$. If the problem was originally to be maximized, we could use the following criterion instead of changing to a minimization problem: Compute the $z_j - c_j$ and select a new vector corresponding to min $(z_j - c_j)$; an optimum solution has been found when all $z_j - c_j \geq 0$. Or compute the $c_j - z_j$ elements with the new vector corresponding to max $(c_j - z_j)$, the procedure stopping when all $c_j - z_j \leq 0$. It is much more efficient, however, especially in developing a computational procedure for an electronic computer, to select one criterion to be used in solving all problems. See page 71 for the discussion on the criterion used to select a vector to be introduced into the basis.

Proof. Let

$$y_{10}\mathrm{P}_1 + y_{20}\mathrm{P}_2 + \cdots + y_{n0}\mathrm{P}_n = \mathrm{P}_0 \tag{1.8}$$

and

$$y_{10}c_1 + y_{20}c_2 + \cdots + y_{n0}c_n = z^* \tag{1.9}$$

be any other feasible solution with z^* the corresponding value of the objective function. We shall show that $z_0 \leq z^*$. (Note that the nondegeneracy assumption is not required for this theorem.)

By hypothesis, $z_j - c_j \leq 0$ for all j, so that replacing c_j by z_j in (1.9) yields

$$y_{10}z_1 + y_{20}z_2 + \cdots + y_{n0}z_n \leq z^* \tag{1.10}$$

For each j we substitute the expression for P_j given by (1.3) into (1.8), to obtain

$$y_{10}\left(\sum_{i=1}^{m} x_{i1}\mathrm{P}_i\right) + y_{20}\left(\sum_{i=1}^{m} x_{i2}\mathrm{P}_i\right) + \cdots + y_{n0}\left(\sum_{i=1}^{m} x_{in}\mathrm{P}_i\right) = \mathrm{P}_0$$

or, by regrouping terms,

$$\left(\sum_{j=1}^{n} y_{j0}x_{1j}\right)\mathrm{P}_1 + \left(\sum_{j=1}^{n} y_{j0}x_{2j}\right)\mathrm{P}_2 + \cdots + \left(\sum_{j=1}^{n} y_{j0}x_{mj}\right)\mathrm{P}_m = \mathrm{P}_0 \tag{1.11}$$

Similarly, for each j we substitute the expression for z_j given by (1.4) into (1.10) to obtain

$$\left(\sum_{j=1}^{n} y_{j0}x_{1j}\right)c_1 + \left(\sum_{j=1}^{n} y_{j0}x_{2j}\right)c_2 + \cdots + \left(\sum_{j=1}^{n} y_{j0}x_{mj}\right)c_m \leq z^* \tag{1.12}$$

Since the set of vectors $\mathrm{P}_1, \mathrm{P}_2, \ldots, \mathrm{P}_m$ is linearly independent, the coefficients of the corresponding vectors in (1.1) and (1.11) must be equal,[1] and hence (1.12) becomes

$$x_{10}c_1 + x_{20}c_2 + \cdots + x_{m0}c_m \leq z^*$$

or, by (1.2), $z_0 \leq z^*$.

The results of Theorems 1 and 2 enable us to start with a basic feasible solution and generate a set of new basic feasible solutions that converge to the minimum solution or determine that a finite solution does not exist.

The nondegeneracy assumption was invoked to ensure the convergence to the minimum solution. If we did not make this assumption, it

[1] See footnote of Theorem 3, Chap. 3, for a proof of the equality of the coefficients of (1.1) and (1.11).

would be possible to have at least one of the m x_{i0} of the given solution equal to zero. If this were the case, then θ_0 could equal zero and the value of the objective function for the new solution would then equal the value of the old solution. This lack of improvement in the solution could continue for a number of successive steps. For this situation, the procedure could conceivably repeat a basis and hence keep returning to this basis. The simplex procedure is then said to have *cycled*, and the computational routine for determining the minimum solution breaks down. During actual computation, the phenomenon of degeneracy is reflected by basic solutions with less than m positive x_{i0} and/or by more than one i yielding $\theta_0 = \min_i (x_{i0}/x_{ij})$ for $x_{ij} > 0$. If the i is not unique, some of the x_{i0} in the new solution will equal zero.

Dantzig, Orden, and Wolfe [32] and Charnes, Cooper, and Henderson [12] have resolved degeneracy from both the theoretical and computational points of view. Computational experience, however, does not warrant incorporating their "degeneracy techniques" into the standard simplex procedure. Out of the many linear-programming problems considered by investigators in the field, only three have been known to cycle. These were artificially constructed by Hoffman [58] and Beale [4] to demonstrate that cycling could occur. What is normally done in a computation is to treat degenerate solutions as nothing unusual and to compute with $\theta_0 = 0$ whenever the procedure yields such a value for θ. If ties occur when determining θ_0, the usual rule is to select $\theta_0 = \min_i (x_{i0}/x_{ij})$ for the smallest index i. Degeneracy techniques and a cycling problem of Beale are discussed in Chap. 7.

2. COMPUTATIONAL PROCEDURE

In this section we assume either that (1) we have selected m linearly independent vectors that yield a feasible solution and have expressed all other vectors in terms of this basis or that (2) our problem matrix contains m vectors that can be explicitly arranged to form a unit matrix of order m.

For the first case let the m linearly independent vectors be $\mathbf{P}_1$, $\mathbf{P}_2$, . . . , $\mathbf{P}_m$ and denote the $m \times m$ matrix $(\mathbf{P}_1\mathbf{P}_2 \cdots \mathbf{P}_m)$ by $\mathbf{B}$. The matrix $\mathbf{B}$ is termed an *admissible basis*. To compute the corresponding solution vector $\mathbf{X}$ and the representation of the other vectors in terms of the basis, we must first compute $\mathbf{B}^{-1}$. Since

$$\mathbf{B}\mathbf{X}_0 = \mathbf{P}_0$$

we have

$$\mathbf{X}_0 = \mathbf{B}^{-1}\mathbf{P}_0$$

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
DEL 15 AL 19 DE ENERO DE 1979

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
1.	BENJAMIN ACOSTA SEGURA José Ma. Morelos No. 42 Col. Concepción México 20, D.F. Tel. 568-53-78	DELEGACION ALVARO OBREGON Fco. I. Madero Esq. Av. Revolución Col. Sn. Angel México 20, D.F. Tel. 559-25-33 Ext. 125
2.	ENRIQUE ACHA DAZA Medicina No. 65 Dpto. 5 Col. Copilco México, D.F. Tel.	
3.	ING. SALVADOR ACHA DAZA Bernal Díaz del Castillo No. 275 Col. Chapultepec Sur Morelia, Mich. Tel.	ESCUELA DE ING. ELECTRICA (UNIVERSIDAD MICHOACANA) Morelia, Mich. Tel. 277-76
4.	GONZALO ARROYO A. Apdo. 22-210 Col. Vergel de Coy. México 22, D.F. Tel. 677-31-06	COMISION FEDERAL DE ELECTRICIDAD Apdo. 19099 Col. Olivar de los Padres México, D.F. Tel.
5.	ING. AUGUSTO AVALOS REYNA Berlín No. 305 Col. Campestre Gómez Palacio,Dgo. Tel. 4-23-93	COMISION FEDERAL DE ELECTRICIDAD Boulevard Alemán y Gunaceví Gómez Palacio, Dgo. Tel. 4-28-87 4-39-31
6.	ING. JOSE MA. ALVAREZ M. José Ma. Iglesias No. 62 Col. Constitución Hermosillo, Son. Tel. 4-14-78	COMISION FEDERAL DE ELECTRICIDAD Juárez y San Luis Potosí Hermosillo, Son. Tel. 3-27-13

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
DEL 15 AL 19 DE ENERO DE 1979

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
7.	MANUEL AVILA OÑATE Asturias No. 62 Col. Alamos México, D.F. Tel. 519-07-48	S I C A R T S A Lazaro Cardenas, Mich.
8.	ING. JAIME BARRIOS DIAZ Córdoba No. 480 Col. Valle Dorado Tlanepantla, Edo. de Méx. Tel. 379-21-61	COMISION FEDERAL DE ELECTRICIDAD Rodano No. 14 Col. Cuauhtémoc México 5, D.F. Tel. 553-71-33 Ext. 2182
9.	ING. LUIS CABEZA RESENDEZ Clarín No. 13 Col. Las Arboledas Atizapan, Edo. de Méx. Tel. 379-00-01	INST. DE INVESTIGACIONES ELECTRICAS Shakespeare No. 6-5o. Piso Col. Ansures México, D.F. Tel. 511-34-74
10.	JOSE LUIS CALDERON GARCIA Presa Sta. Teresa No. 177-201 Col. Irrigación México 10, D.F. Tel. 557-68-35	COMISION FEDERAL DE ELECTRICIDAD Av. Toluca Esq. Don Manuelito Col. Olivar de los Padres México 20, D.F. Tel. 595-55-66
11.	ING. JUAN MANUEL CALVILLO ZAMUDIO Taxqueña No. 1818 casa K-12 Col. Coyoacan México 21, D.F. Tel. 549-00-69	COMISION FEDERAL DE ELECTRICIDAD Calle Don Manuelito Av. Toluca Col. Olivar de los Padres México 20, D.F. Tel. 595-54-00
12.	ING. ARTURO CAMARGO RAYON M. Ocampo No. 171-222 Col. Anáhuac México 17, D.F. Tel. 546-58-39	CIA. DE LUZ Y FUERZA M. Ocampo No. 156-3er. Piso Col. Sn. Rafael México, D.F. Tel. 518-00-80 Ext. 117

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
DEL 15 AL 19 DE ENERO DE 1979

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
13.	ING. RAYMUNDO CAMPOS MILAN 8 Norte Priv. C No. 2609 Col. Rincón del Bosque Puebla,Pue. Tel. 41-93-42	COMISION FEDERAL DE ELECTRICIDAD Antiguo Camino Resurrección Km.41/2 Puebla, Pue. Tel. 41-79-64
14.	ING. JOSE CARRASCO MORALES Faro No. 79 Depto. 11 Col. Vallejo la Patera México 14, D.F. Tel. 535-19-13	DELEGACION CUAUHTEMOC D.D.F. Aldama y Mina Col. Buena Vista México 3, D.F. Tel. 535-19-13
15.	JESUS CHI INFANTE EDIF. 136 Dpto. 1 Unidad Kennedy Col. Jardín Balbuena México 17, D.F. Tel. 552-53-55	DEPARTAMENTO DEL DISTRITO FEDERAL Tlalpan México, D.F. Tel. 577-00-55
16.	ING. ANTONIO DE LA ROSA SORIA Palmira Apto. Postal 475 Cuernavaca, Mor. Tel. 4-13-71	INSTITUTO TECNOLOGICO REG. DE LA LAGUNA
17.	ING. PAULINO DIAZ MEDINA Puebla No. 353-504 Col. Roma México 7, D.F. Tel. 553-71-33 Ext. 2161	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 14-406 Col. Cuauthémoc México 5, D.F. Tel. 53-71-33 Ext. 2161
18.	ISAIAS ELIZARRARAZ ALCARAZ A. Serdán No. 789-B Morelia, Mich. Tel.	ESC. DE ING. ELECTRICA (UNIVERSIDAD MICHOACANA) Morelia, Mich. Tel. 2-77-76

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS (DEL 15 AL 19 de ENERO DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
19.	ING. ADOLFO EQUIHUA TAPIC Gral. José Frontera No. 8 Col. Mercurio Querétaro, Qro. Tel: 2-44-22	INSTITUTO TECNOLOGICO REGIONAL DE QUERETARO Av. Tecnologico Querétaro, Qro.
20.	ING. WOLF FAINSILBER Chilpancingo 14-702 Col. Condesa México 11, D. F. Tel: 5-64-76-71	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 14 Col. Anzurez México 5, D. F. Tel: 5-53-71-33 Ext. 2084
21.	ING. ARSENIO FERNANDEZ AGRAZ Sto. Santiago 3548 Jardines de los Arcos Guadalajara, Jal. Tel: 21-59-70	COMISION FEDERAL DE ELECTRICIDAD Km. 12.5 Carretera a Morelia
22.	ING. JAIME FLORES BOTELLO Carolota 175 Int. 7 Col. Guadalupe Tepeyac México 14, D. F. Tel: 5-17-02-34	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel:5-53-71-33 Ext. 2400
23.	ING. ANGEL FLORES LOPEZ Paseo de la Reforma No. 730-1501 Edif. Zacatecas Tlatelolco México 3, D. F. Tel: 5-53-71-33 Ext. 2161	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-71-33
24.	MIGUEL FLORES ORTEGA Corona Boreal 92 Bis Col. Prado Churubusco México 13, D. F. Tel: 5-81-21-12	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-71-33 Ext. 2677
25.	ING. ARMANDO FRAUSTO SOLIS Hda. Olivar del Conde 36 Fraccionamiento Prados del Rosario México 16, D. F. Tel: 5-46-58-39	CIA. DE LUZ Y FUERZA DEL CENTRO,S.A. Melchor Ocampo 171-222 Col. Anáhuac México, D. F. Tel: 5-46-58-39
26.	CESAR FUENTES ESTRADA Circ. Musicos Gustavo Campa No. 11 Cd. Satélite, Edo. de México	ESIME-SECCION DE GRADUADOS E INVESTIGACION Unidad Profesional Zacatenco Col. Lindavista México 14, D. F.

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS (DEL 15 AL 19 DE ENERO DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
27.	AGUSTIN GALVAN MALDONADO Herreras 117 Col. 20 de Noviembre México 2, D. F. Tel: 7-89-08-74	SISTEMA DE TRANSPORTE COLECTIVO "METRO" Delicias 67 Edif. P.C.C. 5o. Piso México 1, D. F. Tel: 5-12-76-86
28.	ING. ADOLFO GARCIA GARCIA Arboleda de la Hacienda 68 Col. Arboledas Atizapan Edo. de México Tel: 3-79-26-91	CIA. DE LUZ Y FUERZA DEL CENTRO,S. A. Melchor Ocampo 171 Col. Anáhuac México, D. F. Tel: 5-35-16-10
29.	ING. GUILLERMO GARIBAY BORREGO San Marcos 11 Edif. AMALFI 102 Sta. Teresa Contreras México 20, D. F. Tel: 5-68-71-56	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-71-33 Ext. 2616
30.	ING. J. JESUS GONZALEZ HERNANDEZ Ret. 3 de Manuel Rivera Cambas No. 28 Col. Jardín Balbuena México 9, D. F. Tel: 5-71-42-06	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-66-24
31.	ING. ENRIQUE GUERRA VAZQUEZ Paseo del Reloj 47 M.-12-L-14 Ojo de Agua, Edo. de México Tel: 80148	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-66-24.
32.	J. AGUSTIN GUERRERO FLORES Edif. 4 Depto. 3 Zona Hab. La Orilla Lazaro Cárdenas, Michoacan Tel: 2-03-33 Ext. 1367	SIDERURGICA LAZARO CARDENAS"LAS TRUCHAS, S. A. " Domicilio Conocido Lazaro Cárdenas, Michoacan Tel: 2-03-33
33.	ING. CARLOS GUTIERREZ R.	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F.
34.	ING. ANIBAL HERNANDEZ LOPEZ Golf No. 175 Country Club México 21, D. F. Tel: 5-49-94-26	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-65-08

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS (DEL 15 AL 19 DE ENERO DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
35.	ING. RAUL HERRERIAS CORZO Castillo Velasco No. 44 Col. Periodista México 10, D. F. Tel: 5-57-30-23	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Rio Ródano México 5, D. F. Tel: 5-53-71-33 Ext. 2410
36.	ING. MARIO LARA CARMONA Emilio Castelar 44-104 Col. Polanco México 5, D. F. Tel: 5-31-15-76	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 México 14, D. F.
37.	ING. FAUSTINO LARA NUÑEZ Rio Panuco No. 137 Col. Cuauhtémoc México 5, D. F. Tel: 5-28-86-17	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-53-71-33 Ext. 2677
38.	ING. ALBERTO F. LINDSAY Colima No. 4 Hermosillo, Son. Tel: 3-33-86	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F.
39.	RENE LOPEZ GUZMAN Coahuila 156-38 Col. Roma México 7, D. F. Tel" 5-25-67-93	INSTITUTO DE INVESTIGACIONES ELECTRICAS Shallespeare No. 6-3er. Piso Col. Anzures México, D. F.
40.	ING. CESAR LOPEZ ALDRETE Río Janeiro 100-A Col. Alta Vista Monterrey, N. L. 58-17-19	COMISION FEDERAL DE ELECTRICIDAD Carretera Antigua a Roma Km.6 1/2 Tel: 52-50-14
41.	ING. ALEJANDRO LOZANO G. San Marcos No. 102-6 La Joya Tlalpan México 22, D. F. Tel: 5-11-34-74	INSTITUTO DE INVESTIGACIONES ELECTRICAS Shakespeare No. 6-5o. Piso Col. Anzurez México 5, D. F. Tel: 5-11-34-74
42.	ING. J. JESUS MALDONADO LOPEZ M-5 L-14 Lomas Santiaguito Morelia, Mich. Tel; 2-33-44	UNIVERSIDAD MICHOACAN,ESCUELA DE INGENIERIA ELECTRICA Morelia, MIch. Tel: 2-77-76

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS (DEL 15 AL 19 DE ENERO DE 1979)

NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
43. ING. WILFRIDO G. MARTINEZ FUENTES Sierra Diamantina No. 8 Col. Real de las Lomas México 20, D. F.	COMISION FEDERAL DE ELECTRICIDAD Don Manuelito Esq. Av. Toluca Col. Olivar de los Padres México 10, D. F. Tel: 5-95-54-00
44' LUCIO MAYA ARADILLAS Av. 565 - 147 Unidad Aragón México 14, D. F. Tel: 5-60-34-75	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuauhtémoc México 5, D. F. Tel: 5-73-71-33 Ext. 2161
45. EDUARDO MOLERO RAMIREZ Milayo 11 Mayorazgo del Bosque Edo. de México Tel: 3-79-58-63	CIA. DE LUZ Y FUERZA DEL CENTRO,S.A. Melchor Ocampo 171 Ofina No. 407 Col. Anáhuac México, D. F. Tel: 5-18-00-80 Ext. 221
46. ING. RICARDO MOTA PALOMINO División del Norte 3000 Atlántida, Coyoacán México 21, D. F. Tel: 5-49-10-25	COMISION FEDERAL DE ELECTRICIDAD Calle Don Manuelito Esq. Av. Toluca Olivares de los Padres México 22, D. F. Tel: 5-95-54-22
47. ING. ALFREDO NAVA SEGURA Chiapas No. 135-804 Col. Roma México 7, D. F. Tel: 5-86-27-38	IPN-SECCION DE GRADUADOS ESIME Unidad Profesional de Zacatenco Col. Lindavista México 14, D. F. Tel: 5-86-27-38
48. SALVADOR NUÑEZ CASTILLO Calle Industria No. 7 Col. Del Recreo Azcapotzalco México, D. F. Tel: 5-61-84-22	SISTEMA DE TRANSPORTE COLECTIVO"METRO" Ernesto Pugibet No. 8-6o. Piso México 1, D. F. Tel: 5-21-86-20
49. ING. HECTOR NUÑEZ DIAZ Gardenia No. 18 Col. Jardín México 21, D.F. Tel. 549-51-45	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 18 Col. Cuauhtémoc México 5, D.F. Tel. 286-20-94

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
(DEL 15 AL 19 DE ENERO DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
50.	MAXIMINO OCHOA RIVAS Colina de dos Rocas No. 52 México 5, D.F. Tel. 286-20-94	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14-413 Col. Cuauhtémoc México 5, D.F. Tel. 286-20-94
51.	GUILLERMO ORTEGA RODRIGUEZ Gustavo Sotelo No. 1795-1A Col. Héroes de Nacozari Mexicali, B.C.	COMISION FEDERAL DE ELECTRICIDAD Benito Juárez y Lazaro Cárdenas Mexili, B.C. Tel. 878-01
52.	ING. JOSE RUPERTO PEÑA MORENO Ebano No. 18 Col. Santa María la Ribera México 4, D.F. Tel. 541-15-19	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 14-4o. Piso Col. Cuauhtémoc México 5, D.F. Tel. 553-66-24
53.	ING. ARTURO I. PEON ZAPATA Las Flores No. 60 Col. La Candelaria Coy. México 21, D.F. Tel. 549-67-30	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 14-405 Col. Cuauhtémoc México 5, D.F. Tel. 553-71-33 Ext. 2677,2143,2084
54.	GUILLERMO JOSE LUIS PERALTA SOLORIO Cali No. 323-A Col. Valle Dorado Tlalnepantla, Edo. de Méx. Tel. 379-34-16	COMISION FEDERAL DE ELECTRICIDAD Av. Toluca y Calle Don Manuelito Col. Olivar de los Padres México 20, D.F. Tel. 595-54-00
55.	ING. RODOLFO PIMENTEL C. Hda. Santiago No. 82 Col. Prado del Rosario México 16, D.F. Tel. 561-71-37	CIA. DE LUZ Y FUERZA DEL CENTRO Tlaloc No. 90-4o. Piso Col. Anáhuac México 16, D.F. Tel. 546-11-55
56.	ING. LORENZO REYES OSORIO Valle Dmieper No. 8 Col. Valle de Aragón Edo. de Méx. Tel.	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14-403 Col. Cuautémoc México 5, D.F. Tel.553-71-33 Ext. 2781
57.	ING. VICTOR MANUEL RODRIGUEZ LOPEZ Ayuntamieto No. 114 G-501 Col. Tlalpan México 22, D.F.	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 14 Col. Cuauhtémoc México 5, D.F.

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
(DEL 15 AL 19 DE ENERO DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
58.	ING. RIKICA ROMANO Aldama No. 9-303 Col. Buenavista México 3, D.F. Tel.	INST. DE INVESTIGACIONES ELECTRICAS Shakespeare No. 6-5 Col. Anzures México 5, D.F. Tel. 511-34-74
59.	ING. DAVID ROMERO ROMERO Edificio Libertad B-2-207 Col. Centro México, D.F. Tel. 529-13-09	COMISION FEDERAL DE ELECTRICIDAD Calle Don Manuelito y Av. Toluca Col. Olivar de los Padres México 22, D.F. Tel. 595-55-11
60.	ING. HECTOR G. SARMIENTO Cerro Coporo No. 46 Col. C. Churubusco México, D.F. Tel. 549-10-46	INST. DE INVESTIGACIONES ELECTRICAS Internado Palmira Apdo. Postal 475 Cuernavaca, Mor. Tel. (731) 413-71
61.	ING. GERARDO SORDO LINARES Cuzco No. 881 Col. Lindavista México 14, D.F. Tel. 586-85-68	CIA. DE LUZ Y FUERZA DEL CENTRO Melchor Ocampo No. 171 Col. San Rafael México 5, D.F. Tel. 518-00-80 Ext. 213
62.	ING. VIRGILIO VALDEZ CERVANTES Salazar No. 827-A Pte. Monterrey, N.L. Tel. 74-34-47	COMISION FEDERAL DE ELECTRICIDAD Anaya y Universidad Monterrey, N.L. Tel. 52-50-14
63.	ING. RICARDO VARGAS SALAS Priv. del Rosal No. 37 Cuernavaca, Mor. Tel. 265-67	INST. DE INVESTIGACILES ELECTRICAS Cuernavaca,Mor. Tel.
64.	ING. CESAR M. VILCHIS CERVANTES Guerrero No. 387-0-713 Col. Nonoalco Tlatelolco México 3, D.F. Tel. 597-45-46	COMISION FEDERAL DE ELECTRICIDAD Ródano No. 14-4o. Piso Col. Cuauhtémoc México 5, D.F. Tel. 553-71-33 Ext. 2003

DIRECTORIO DE ASISTENTES AL CURSO: TEMAS SELECTOS DE ANALISIS DE REDES ELECTRICAS
(DEL 15 AL 19 DE ENERO DE 1979)

	NOMBRE Y DIRECCION	EMPRESA Y DIRECCION
65.	ING. DANIEL ZAVALA MALDONADO Av. San Jeronimo No. 1363-8 Col. San Jerónimo México 20, D.F. Tel. 595-58-30	COMISION FEDERAL DE ELECTRICIDAD Calle Don Manuelito Av. Toluca Col. Olivar de los Padres México 20, D.F. Tel. 595-55-44 Ext. 224
66.	C. ALBERTO BLANCO MONTIEL Calz. Niño Perdido No. 580 Depto. 302 Col. Alamos México 13, D.F. Tel. 590-32-77	COMISION FEDERAL DE ELECTRICIDAD Río Ródano No. 14 Col. Cuautémoc México 5, D.F. Tel. 553-71-33 Ext. 2189
67.	ING. HORACIO CRISTIAN BUITRON SANCHEZ Hernández No. 64 Col. Tepeyac Ins. México 14, D.F. Tel. 577-94-86	COMISION FEDERAL DE ELECTRICIDAD Av. Toluca Esq. Don Manuelito Col. Olivar de los Padres México 20, D.F. Tel. 595-54-00
68.	RODOLFO CALDERON GARCIA Clarín No. 13 Col. Las Arboledas Atizapan, Edo. de Méx.	INST. DE INVESTIGACIONES ELECTRICAS Shakespeare No. 6-5o. Piso Col Anzures México, D.F.