



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

IMPLEMENTACIÓN Y COMPARACIÓN DE
ALGORITMOS DE ESTIMACIÓN DE
PARÁMETROS PARA PREDICCIÓN Y FILTRADO

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Ingeniero Eléctrico Electrónico

PRESENTA:

Rodrigo Moreno Carrillo

DIRECTOR DE TESIS:

Dr. Rogelio Alcántara Silva



México, D.F., 2015

*A la Facultad de Ingeniería y a la Universidad, por la formación que me han dado.
Pero principalmente a mis padres Ricardo y Norma a quienes debo todo lo que tengo
y todo lo que soy y por lo que les dedico este gran logro de mi vida.*

En verdad, gracias.

Rodrigo Moreno Carrillo.

Reconocimientos

En esta sección, quisiera reconocer y más que nada agradecer a todas las personas que me apoyaron a lo largo de mis estudios universitarios, desde los profesores que compartieron conmigo sus conocimientos, pasando por los compañeros con los que conviví a lo largo de este proceso, hasta los grandes amigos que me apoyaron en todas las situaciones en los que los necesité, con los que compartí grandes momentos y a quienes espero seguir frecuentando durante toda mi vida.

Agradezco también enormemente al Dr. Rogelio Alcántara Silva por su voluntad de transmitir nuevos conocimientos y experiencias, por su vocación para *fortalecer* y *motivar* las habilidades de sus alumnos y principalmente por el apoyo incondicional que recibí de su parte.

Por otra parte, me siento enormemente agradecido y afortunado, por haber cursado mis estudios en una institución tan grande en todos los sentidos como lo es la *Universidad* y en específico en la Facultad de Ingeniería, de la cual me siento orgulloso de haber sido alumno.

Y guardando la mejor parte para el final, quiero agradecer con todo mi corazón a mi familia de quien he recibido siempre apoyo, ayuda, motivación y sobre todo cariño. A mi tía y a mis abuelos por estar siempre a mi lado siendo siempre una figura a seguir, a mis padres Ricardo y Norma por brindarme siempre todo lo necesario, sin pedir nada a cambio, para completar mis estudios, por motivarme y hacer que siempre diera lo mejor de mi.

Quiero brindar un agradecimiento especial a mi mamá, quien siempre ha estado a mi lado en cada momento de esta etapa, en cada día de la universidad y cada momento de mi vida. Gracias por tu apoyo, tu manera de presionarme para ser siempre el mejor, por despertarme cada mañana, por tu ayuda en todo lo que podías apoyarme. Sin encontrar más palabras para expresar mi gratitud sólo puedo decir: *Gracias por estar siempre conmigo.*

Declaración de autenticidad

Por la presente declaro que, salvo cuando se haga referencia específica al trabajo de otras personas, el contenido de esta tesis es original y no se ha presentado total o parcialmente para su consideración para cualquier otro título o grado en esta o cualquier otra Universidad. Esta tesis es resultado de mi propio trabajo y no incluye nada que sea el resultado de algún trabajo realizado en colaboración, salvo que se indique específicamente en el texto.

Rodrigo Moreno Carrillo. México, D.F., 2015

Resumen

Este trabajo desarrolla las bases teóricas, principios y conceptos, necesarios para comprender el funcionamiento de los filtros, en especial el filtro de Wiener y el filtro de Kalman, desde un enfoque principalmente probabilístico. De esta manera se puede comprender los efectos que causan perturbaciones de naturaleza aleatoria, causadas principalmente por la presencia de ruido común en prácticamente todos los sistemas sin importar la viabilidad o calidad del modelado matemático, así como comportamientos no modelados debido a su dificultad.

Posteriormente, una vez entendidas las ideas planteadas anteriormente, se brinda una breve ejemplificación de la funcionalidad de las aplicaciones de filtros a señales ruidosas para, finalmente dadas, las ventajas presentadas por el filtro de Kalman desarrollar una aplicación más elaborada en la que se puedan demostrar de una manera más tangible, y fácil de ver, los resultados y los beneficios obtenidos al utilizar un filtro de Kalman para mejorar la precisión de la estimación de la posición de un robot móvil diferencial.

Índice general

1. Introducción	1
1.1. Objetivo	1
1.2. Planteamiento del problema	2
1.3. Metodología	2
1.4. Motivación	3
1.5. Contexto histórico	3
1.6. Estructura de la tesis	5
2. Modelos y estimación paramétrica en filtrado lineal	7
2.1. Modelos entrada-salida	7
2.2. Modelado paramétrico	8
2.2.1. Modelo autoregresivo (AR)	10
2.2.2. Modelo de promedio móvil (MA)	10
2.2.3. Modelo Autoregresivo de promedio móvil (ARMA)	11
2.3. Filtros lineales	11
2.3.1. Definición de filtro	11
2.3.2. Operaciones básicas	11
2.3.3. Definición del filtro lineal	12
2.3.4. Planteamiento y Solución al problema del filtrado lineal	12
2.4. Estructuras de filtros	13
2.4.1. Estructuras de Filtros FIR	14
2.4.2. Estructuras de Filtros IIR	16
3. Filtro de Wiener	19
3.1. Consideraciones y restricciones	19
3.2. Filtro lineal óptimo de Wiener	20
3.2.1. Criterio de optimización	21
3.2.2. Planteamiento del problema del filtrado de Wiener	22
3.3. Principio de Ortogonalidad	22
3.4. Ecuaciones de Wiener-Hopf	27
3.4.1. Solución de ecuaciones de <i>Wiener-Hopf</i>	28

ÍNDICE GENERAL

3.5. Superficie de desempeño	29
3.6. Problema planteado	31
3.6.1. Caracterización del problema	31
3.6.2. Resolución del problema	32
3.7. Conclusiones	40
4. Filtro de Kalman	43
4.1. Filtros Adaptables	43
4.1.1. Enfoque de operación basado en la teoría del filtro de Kalman	44
4.2. Filtro de Kalman	45
4.2.1. ¿Por qué el filtro de Kalman?	47
4.3. Problema del filtrado de Kalman	48
4.3.1. Modelo en espacio de estados	48
4.3.2. Proyecciones ortogonales	48
4.3.3. Solución al problema de filtrado	50
4.4. Algoritmo para el filtrado de Kalman	54
4.4.1. Predicción	55
4.4.2. Corrección	56
4.4.3. Parámetros del Filtro de Kalman	58
4.5. Problema planteado	59
4.5.1. Caracterización del problema	59
4.5.2. Resolución del problema	61
4.5.3. Simulaciones	62
5. Seguimiento de trayectorias	65
5.1. Planteamiento del problema	65
5.2. Robot diferencial	66
5.2.1. Modelo dinámico	67
5.2.2. Modelo discreto	67
5.3. Planificación de trayectoria	70
5.3.1. Trayectoria ideal	70
5.3.2. Trayectoria con ruido	72
5.4. Estimación óptima de trayectoria	72
5.4.1. Predicción	75
5.4.2. Corrección	76
5.5. Resultados	76
6. Conclusiones	81

A. Sistemas y Señales en tiempo discreto	83
A.1. Señal	83
A.1.1. Señales continuas y discretas	84
A.2. Señales en tiempo discreto	84
A.2.1. Señales elementales	87
A.2.2. Clasificación de las señales discretas en el tiempo	92
A.3. Sistemas en tiempo discreto	94
A.3.1. Clasificación	95
A.4. Ecuaciones en diferencias	99
B. Procesos Aleatorios	103
B.1. Variables aleatorias	103
B.1.1. Caracterización	105
B.1.2. Momentos de una variable aleatoria	109
B.2. Variables aleatorias multidimensionales	114
B.2.1. Caracterización	115
B.2.2. Momentos	116
B.2.3. Correlación e Independencia	118
B.3. Variables aleatorias Gaussianas	119
B.3.1. Transformaciones lineales	121
B.4. Procesos aleatorios	122
B.4.1. Clasificación de los procesos aleatorios	123
B.4.2. Procesos Aleatorios Determinísticos y No Determinísticos	126
B.4.3. Caracterización de los procesos aleatorios	127
B.4.4. Procesos aleatorios estacionarios	128
B.5. Funciones de correlación	131
B.5.1. Autocorrelación	131
B.5.2. Correlación	136
B.5.3. Autocovarianza	137
B.5.4. Covarianza	138
B.6. Procesos Aleatorios discretos	138
B.6.1. Caracterización parcial de una secuencia aleatoria discreta	139
B.6.2. Secuencias de autocorrelación y correlación cruzada	141
B.6.3. Matriz de autocorrelación	142
C. Variables de Estado	147
C.1. Modelado en variables de estado	147
C.1.1. Definiciones	147
C.2. Representaciones en el espacio de estados de sistemas en tiempo discreto	149
C.2.1. Forma canónica controlable	150

ÍNDICE GENERAL

C.2.2. Forma canónica observable	151
C.2.3. Forma canónica diagonal	151
C.2.4. Forma canónica de Jordan	152

Índice de figuras

2.1. Sistema lineal causal, con entrada $v[n]$ y salida $\hat{s}[n]$, encargado de modelar la señal $s[n]$	9
2.2. Estructura de un filtro FIR en su forma directa	15
2.3. Estructura de un filtro FIR en su forma de cascada	15
2.4. Estructura de un filtro FIR en celosía	16
2.5. Estructura de un filtro IIR en su forma directa	17
3.1. Diagrama a bloques que representa el problema del filtrado de Wiener	23
3.2. Interpretación geométrica del principio de ortogonalidad	25
3.3. Filtro de Wiener FIR con estructura directa, caracterizado por los coeficientes w_0, w_1, \dots, w_p	28
3.4. Superficie de desempeño de 2 dimensiones	30
3.5. Diagrama a bloques de aplicación del filtro de Wiener	33
3.6. Filtro de Wiener de orden 0	34
3.7. Filtro de Wiener de orden 1	35
3.8. Filtro de Wiener de orden 2	36
3.9. Filtro de Wiener de orden 3	37
3.10. Filtro de Wiener de orden 4	38
3.11. Filtro de Wiener de orden 5	39
3.12. Gráfica del error cuadrático medio mínimo J_{min} para distintos órdenes	40
4.1. Representación gráfica del espacio $Y[k]$ y los subespacios $\mathcal{Y}[k-1]$ y $\mathcal{Z}[k]$	51
4.2. El ciclo del filtrado de Kalman se dividen dos grupos y se repite	55
4.3. Ciclo completo del filtrado de Kalman	57
4.4. Diagrama a bloques de aplicación del filtro de Kalman	61
4.5. Simulación del filtro de Kalman aplicado al problema en la sección 4.5	62
4.6. Convergencia de la varianza del error	63
4.7. Simulación del filtro de Wiener de orden 10	63
4.8. Comparación de las señales obtenidas por Kalman y Wiener junto con la señal deseada	64

ÍNDICE DE FIGURAS

5.1. Ilustración de un robot diferencial con tres puntos de apoyo, dos de los cuales son las ruedas móviles	66
5.2. Trayectoria ideal descrita por un robot móvil desde un punto inicial hasta el punto objetivo	71
5.3. Función de densidad de las mediciones obtenidas de un GPS con media $\mu = 0$ y una matriz de covarianza \mathbf{R}	73
5.4. Diagrama de flujo de la implementación del filtro de Kalman	74
5.5. Gráfica de las mediciones de las señales de entrada ΔR y ΔL del modelo	77
5.6. Gráfica de los estados óptimos estimados del modelo	78
5.7. Trayectoria óptima estimada a partir de las mediciones	79
5.8. Acercamiento de la trayectoria estimada	80
A.1. Representación gráfica de una señal en tiempo discreto	85
A.2. Señal analógica y señal discreta	86
A.3. Secuencias elementales presentes en los sistemas en tiempo discreto	87
A.4. Secuencias exponenciales reales	90
A.5. Señales exponenciales complejas	91
A.6. Representación gráfica de un sistema discreto en el tiempo	95
B.1. Correspondencia del espacio S hacia el eje x	104
B.2. Representación gráfica de la relación entre una función de distribución y su función de densidad de probabilidad	107
B.3. Interpretación del valor medio	112
B.4. Representación gráfica de la correlación y la independencia	120
B.5. Correspondencia de un espacio de eventos S hacia distintas funciones $X_i(t)$	123
B.6. Proceso aleatorio continuo	124
B.7. Proceso aleatorio discreto	125
B.8. Secuencia aleatoria continua	125
B.9. Secuencia aleatoria discreta	126
B.10. Proceso aleatorio lento $X(t)$ y proceso aleatorio rápido $Y(t)$	134
B.11. Autocorrelación de procesos rápido ($X(t)$) y lento ($Y(t)$)	135

Introducción

1.1. Objetivo

- Estudio y revisión de los principios básicos, imprescindibles, para adentrarse en el área teoría de la estimación lineal basada en el sentido del criterio de minimización del error cuadrático medio (MSE), particularmente el trabajo relativo al problema de filtrado de señales aleatorias.
- Estudio, revisión y comparación de los diferentes enfoques teóricos para el caso de procesos estacionarios y no estacionarios tanto para la predicción a un paso y la cancelación de ruido, considerando modelos entrada-salida (externos), como para el filtrado utilizando modelos de estado (internos).
- Análisis, programación, validación y comparación de los algoritmos de estimación de parámetros del modelo y de las señales en lenguaje de alto nivel, como bloques o módulos independientemente de la aplicación.
- Establecer puntos y criterios de comparación entre los diferentes enfoques y algoritmos para la resolución del problema de filtrado óptimo.
- Simulación, implementación, validación y evaluación del desempeño de los algoritmos bajo el contexto de la predicción y el filtrado para el caso de la cancelación de ruido e interferencias.
- Desarrollo modular de las funciones de los algoritmos, y de la simulación bajo el contexto de la predicción y filtrado. La biblioteca de funciones resultante, podrá ser utilizada para otras simulaciones o aplicaciones reales en distintas plataformas de implementación.

1.2. Planteamiento del problema

En muchas aplicaciones teóricas y prácticas, se dispone de información que ha sido corrompida, lo cual limita el funcionamiento además, de dificultar el análisis de los sistemas. Esto se debe a que ningún modelo matemático es perfecto, los sistemas son susceptibles a perturbaciones y los sensores no son ideales y añaden ruido no deseado, que tiene una naturaleza aleatoria.

Por lo anterior, es útil diseñar un filtro que suprima o minimice las interferencias no deseadas de manera óptima. En consecuencia, la tarea consiste en encontrar los parámetros del filtro que realicen la discriminación de la información de manera óptima de acuerdo con algún criterio de evaluación preestablecido.

1.3. Metodología

Para establecer los puntos de comparación entre los distintos enfoques de algoritmos utilizados en la resolución del problema, y particularmente los algoritmos correspondientes considerados como módulos entrada-salida independientes, es necesario desarrollar las bases teóricas de acuerdo con cada aproximación, de esta manera comprender los alcances y requerimientos de cada algoritmo en particular.

Posteriormente, se realiza la implementación de los mismos en algún dispositivo digital, de esta manera se comparan los recursos necesarios para su funcionamiento óptimo, además de obtener datos sintéticos que permitan evaluar el desempeño en distintas condiciones de trabajo.

Finalmente, la comparación de los resultados permite determinar la viabilidad o facilidad de implementación, así como ventajas y desventajas, de un algoritmo en específico para ciertas condiciones preestablecidas.

La metodología se generaliza para la integración de los algoritmos bajo el entorno de la predicción, del filtrado y de la estimación de trayectorias de objetos móviles.

1.4. Motivación

El análisis de señales se ha convertido en los últimos años en una herramienta esencial para el procesamiento y manipulación de datos a través de sistemas digitales que cada vez son más comunes debido a la disminución de su tamaño y precio, además de su considerable incremento en la capacidad de procesamiento. Esta idea básica nos lleva a pensar en la implantación de dispositivos que consuman menos recursos energéticos y que al mismo tiempo puedan realizar cada vez más tareas.

Es por ello que se busca siempre implementar aplicaciones en pequeñas micro-computadoras, cuidando el desempeño para que a su vez sean integradas a sistemas más complejos, y de esta manera contribuir a obtener un mayor nivel de integración en los dispositivos existentes en el mercado.

En específico el enfoque probabilístico que se le da al filtrado puede ser una herramienta de gran ayuda para el modelado, análisis e implementación de los sistemas dinámicos, tales como lo puede ser un robot móvil en el cuál se centra este trabajo.

1.5. Contexto histórico

Durante los últimos años se ha dado una fusión entre la teoría de probabilidad y la teoría de sistemas, en especial la teoría de control, esto debido a que se encuentran altamente relacionadas de la ciencia y en conjunto estas dos ramas han permitido grandes avances en la tecnología, principalmente durante el siglo XX.

Los primeros indicios del desarrollo de la teoría de estimación, una rama de la probabilidad la cual se dedica a la estimación de valores o parámetros basados en las mediciones o datos obtenidos de manera empírica que tienen una componente aleatoria, tiene sus inicios en estudios astronómicos realizados por Galileo Galilei en 1632 al tratar de minimizar ciertos errores. Sin embargo, de acuerdo con [Haykin, 1991, p. 65], el inicio se le atribuye a Carl Fredrik Gauss, quien a la edad de 18 años en 1795 realizó una generalización del método de los mínimos cuadrados (*method of least squares*) para estudiar el movimiento de los cuerpos celestes.

A partir de entonces y hasta finales del siglo XIX se presentaron pequeños detalles en ambas ramas de la ciencia, tanto en el campo de la teoría de estimación como en la teoría de control; salvo con algunas publicaciones como las de Routh y

1. INTRODUCCIÓN

Hurwitz, que sentaron las bases y adquirieron relevancia en una época posterior en especial la teoría desarrollada por Alexandr Lyapunov en 1892.

El periodo siguiente, que comprende desde finales del siglo *XIX* hasta la segunda guerra mundial, se caracterizó por un amplio desarrollo en el campo de la electricidad y el magnetismo. A este periodo se le conoce como la Edad de la Electrónica, según [Hendricks et al., 2008, p. 4], en ella se llegó a grandes descubrimientos que revolucionaron la industria de los circuitos y las comunicaciones, contribuciones como: el amplificador termoiónico, que conllevaron a la aplicación de conceptos como la *retroalimentación* en sistemas prácticos que desencadenaron en amplificadores de radio u osciladores además de sentar las bases teóricas para el análisis de los circuitos de corriente alterna.

Estos avances permitieron trabajos posteriores muy importantes en diferentes áreas. Por un lado, los primeros estudios de la estimación con el enfoque del valor medio mínimo en procesos aleatorios fueron desarrollados por Kolmogorov, Krein y Wiener, siendo Wiener quien formuló el problema de la predicción lineal y el filtrado al estimar un proceso corrompido por ruido, teniendo como resultado una fórmula explícita para el estimado óptimo [Wiener-Hopf, 1931] [Haykin, 1991].

Mientras tanto, en la teoría de control, Harry Nyquist (1932) y Hendrik Bode (1940) realizaban trabajos relacionados con el análisis de estabilidad y la respuesta en magnitud y fase para sistemas en lazo abierto y lazo cerrado, ambos sobre sistemas descritos mediante funciones de transferencia.

Llegando la Segunda Guerra Mundial, se utilizaron todos estos conceptos desarrollados hasta entonces además de ser estudiados con mayor profundidad y aplicados en diversos dispositivos de propósito bélico. Al finalizar este periodo, continuó el intenso desarrollo de aplicaciones prácticas utilizando herramientas teóricas como: métodos en el dominio de la frecuencia, mediante la transformada de Laplace, análisis mediante función de transferencia y el plano de s . Este periodo está caracterizado además por el uso de sistemas SISO, por sus siglas en inglés (*Single Input, Single Output*).

Finalmente, a finales de la década de los 50's, con el inicio de la Era Espacial, distintos autores trabajaron en esta área, realizando trabajos que van desde controlar objetos balísticos hasta la estimación de procesos no estacionarios. Este nuevo periodo tuvo como requerimiento analizar y controlar sistemas cada vez más complejos, sistemas MIMO (*Multiple Input, Multiple Output*).

Finalmente, durante este periodo, en específico entre 1960 y 1961, Rudolf Kal-

man y sus colaboradores publicaron cuatro artículos, entre los cuales se habla de los detalles del algoritmo de estimación óptima, discreta y continua.

1.6. Estructura de la tesis

Este trabajo está dividido en cinco capítulos, con los cuales se desarrolla una solución para el seguimiento de la trayectoria de un robot móvil a partir de la estimación de los estados mediante un filtro de Kalman discreto.

En el primer capítulo, la introducción, se plantean los alcances del trabajo mediante los objetivos así como el problema a resolver, describiendo la metodología utilizada para la resolución de la problemática.

Además, se da una breve explicación de los motivos por los cuales se considera el enfoque propuesto y, por supuesto, se brinda una breve reseña histórica mediante la cual se puede apreciar mejor el desarrollo de las dos áreas en las cuales se basa el trabajo, la probabilidad y la teoría de control.

En el segundo capítulo, se considera una breve descripción acerca de los modelos entrada-salida, los cuales son esenciales para la comprensión de la teoría de sistemas así como del filtrado lineal. Siendo una parte importante los modelos paramétricos que son ampliamente utilizados en la teoría de estimación.

Posteriormente, se brinda también una descripción de las estructuras más utilizadas para la implementación de dichos modelos paramétricos, mediante filtros lineales discretos.

En el tercer capítulo, se puede encontrar el desarrollo del filtro de Wiener de una manera detallada, describiendo el planteamiento del problema así como el criterio utilizado para alcanzar el funcionamiento óptimo, para, finalmente, llegar a una expresión en concreto que permite la implementación del filtro.

En especial, se trata este filtro ya que sirve como introducción para abordar de manera más fluida el filtrado de Kalman.

A lo largo del cuarto capítulo, se abordan las ventajas de utilizar un filtro adaptable, sobre un filtro de Wiener, para tratar diferentes tipos de señales. En esta sección también se trata la importancia del porqué se decidió utilizar un filtro de Kalman, en donde se describen las ventajas más destacadas, y se describen las principales

1. INTRODUCCIÓN

fases con las que cuenta el algoritmo.

Finalmente, en el último capítulo, se utilizan todos los conceptos descritos, principalmente en el capítulo tres y cuatro, para desarrollar la problemática propuesta al intentar dar un estimado de la trayectoria de un robot móvil.

Modelos y estimación paramétrica en filtrado lineal

2.1. Modelos entrada-salida

Los modelos de entrada salida fueron desarrollados en una etapa de gran desarrollo teórico, en el periodo previo a la segunda guerra mundial, periodo en el cual se desarrollaron conceptos o dispositivos tan importantes como el amplificador, la retroalimentación (*feedback*), positiva y negativa, que hicieron posibles los sistemas de comunicación. Hasta entonces la gran parte de la teoría disponible fue desarrollada gracias a la industria de la electrónica y las comunicaciones[Hendricks et al., 2008, p. 2].

Durante la segunda guerra mundial, numerosos desarrollos se dieron en el ámbito de la tecnología que, una vez finalizada, permitieron un mejor entendimiento de las teorías y conceptos utilizados en los sistemas. Normalmente estos sistemas eran diseñados con un enfoque entrada-salida, o *SISO*, que eran aplicados a sistemas invariantes en el tiempo *LTI*.

Las bases teóricas para el análisis de estos sistemas fueron desarrolladas por Nyquist, Bode, Evans y Nichols mucho antes de que hubiese una conexión entre estos métodos. El desempeño de estos sistemas se podía entonces describir mediante las propiedades del sistema en lazo abierto, además de parámetros, en el dominio de la frecuencia, como: ancho de banda, ganancia, fase, tiempo de asentamiento, porcentaje de sobrepaso, error en estado estacionario, resonancia y amortiguamiento[Hendricks et al., 2008, p. 1-8].

Normalmente estas representaciones de los sistemas lineales en el tiempo discreto se dan mediante:

- Respuesta al impulso
- Función de transferencia
- Ecuaciones en diferencias

Para una gran parte de la teoría de los sistemas discretos en el tiempo, se utiliza una representación mediante ecuaciones en diferencias. La representación mediante ecuación en diferencias es ampliamente utilizada pues permite una implementación fácil mediante las operaciones que se pueden realizar en algún dispositivo o computadora digital, tales como suma, multiplicación y almacenamiento de muestras pasadas.

Otro enfoque para los sistemas es el planteado por los modelos en variables de estado, el cual tuvo un gran auge en la década de 1960, durante la era aeroespacial.

2.2. Modelado paramétrico

El término *modelo* es usado como una representación abstracta para explicar el comportamiento de las leyes que se supone describen o rigen en general algún fenómeno de interés, de tal manera que puede ser una réplica exacta de éste o ser una abstracción de las propiedades determinantes del fenómeno.

El modelado paramétrico es un procedimiento que fue originalmente ideado por Yule en el año de 1927 [Yule, 1927]. De acuerdo con el procedimiento, el objetivo es encontrar un modelo matemático, con una estructura predefinida y un número predefinido de parámetros, para representar una señal. En otras palabras, una señal en específico $s[n]$ se intenta representar mediante $\hat{s}[n]$ que se supone sea lo más cercana posible en algún sentido a la señal dada.

Dicho modelo matemático se representa mediante un sistema lineal, en este caso lineal y discreto en el tiempo, descrito mediante una función de transferencia $H[z]$, representada por la expresión 2.1. Al momento de agregar restricciones se hace posible la obtención de los parámetros de $H[z]$ para que dada una señal de entrada estadísticamente independiente $v[n]$ se obtenga una salida $\hat{s}[n]$ [Haykin, 1991, p. 8].

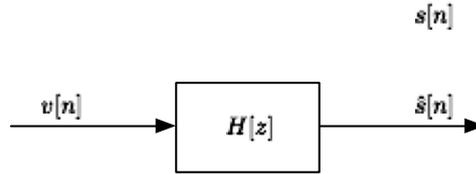


Figura 2.1: Sistema lineal causal, con entrada $v[n]$ y salida $\hat{s}[n]$, encargado de modelar la señal $s[n]$

$$H[z] = \frac{B[z]}{A[z]} = \frac{\sum_{k=0}^Q b_k z^{-k}}{1 - \sum_{k=1}^P a_k z^{-k}} \quad (2.1)$$

Entonces la señal queda modelada por los valores de los coeficientes de los polinomios $A[z]$ y $B[z]$, junto con la señal de entrada $v[n]$. La señal de entrada se elige dependiendo si la señal que se desea modelar es determinística o aleatoria, para las primeras suele utilizarse un impulso unitario, mientras que para las señales aleatorias generalmente se escoge una señal aleatoria Gaussiana con media cero y varianza constante, también conocida como ruido gaussiano [Proakis and Manolakis, 2007, p. 744-746].

Por lo anterior se establece que un proceso aleatorio estacionario en sentido amplio puede representarse como la salida de un sistema lineal excitado por ruido blanco, la tarea restante es encontrar los parámetros a_k y b_k para una estructura predefinida que aproxime la señal de salida $\hat{s}[n]$ lo mejor posible a $s[n]$. Al ser $H[z]$ un sistema lineal, causal e invariante en el tiempo, la salida se puede relacionar con la entrada mediante la ecuación en diferencias:

$$\hat{s}[n] = \sum_{k=0}^Q b_k v[n-k] - \sum_{k=1}^P a_k \hat{s}[n-k] \quad (2.2)$$

La manera en la que la combinación lineal de los valores pasados de la entrada $v[n-k]$ y los valores pasados de la salida $\hat{s}[n-k]$ son establecidos define la estructura del filtro lineal. Por lo general se pueden identificar tres principales tipos de modelos con diferentes ventajas y desventajas.

2.2.1. Modelo autoregresivo (AR)

El modelo autoregresivo (AR), por nombre en inglés *Autoregressive Model* [Haykin, 1991, p. 90], es un filtro lineal $H[z]$ descrito solamente por polos y ningún cero.

$$H[z] = \frac{1}{A[z]} = \frac{1}{1 - \sum_{k=1}^P a_k z^{-k}} \quad (2.3)$$

En una representación mediante una ecuación en diferencias

$$\hat{s}[n] = v[n] - \sum_{k=1}^P a_k \hat{s}[n - k] \quad (2.4)$$

De esta manera se puede observar que el modelo depende únicamente del valor actual de la entrada y los valores pasados de la salida. Lo que significa que al hacer $Q = 0$ y $b_0 = 1$ en la ecuación 2.2 se modela un proceso autoregresivo.

Este modelo se recomienda cuando la señal a modelar presenta muchos picos y se requiere utilizar el modelo más sencillo, pues para obtener sus parámetros, a_k para $k = 1, 2, \dots, P$, se utiliza un sistema de ecuaciones lineales.

2.2.2. Modelo de promedio móvil (MA)

El modelo de promedio móvil (MA), por sus siglas en inglés *Moving Average* [Haykin, 1991, p. 93], es un filtro lineal $H[z]$ descrito por sólo ceros y ningún polo, de acuerdo con la función de transferencia

$$H[z] = \frac{B[z]}{1} = \frac{\sum_{k=0}^Q b_k z^{-k}}{1} \quad (2.5)$$

La cual, representada mediante una ecuación en diferencias, tiene la forma siguiente

$$\hat{s}[n] = \sum_{k=0}^Q b_k v[n - k] \quad (2.6)$$

En donde a partir de 2.2 y al hacer $P = 0$ y $a_0 = 1$ se puede observar que el modelo depende de los valores pasados y actual de la entrada.

Este modelo es más complicado que el modelo anterior ya que al obtener los parámetros b_k para $k = 0, 1, \dots, Q$ se involucra la resolución de un conjunto de ecuaciones no lineales. Se recomienda el uso de este modelo para señales que se desean modelar con abundancia de valles.

2.2.3. Modelo Autoregresivo de promedio móvil (ARMA)

Finalmente, el modelo autoregresivo de promedio móvil (ARMA) es una combinación de los dos modelos anteriores, y se puede decir que es el modelo más general pues se representa mediante una función de transferencia $H[z]$ con polos y ceros, como la descrita en la ecuación 2.1.

En ella se puede observar que su representación está dada por la ecuación en diferencias 2.2, en donde el modelo depende tanto de los valores pasados de la entrada como los valores pasados de la salida. Este modelo al ser una combinación de los modelos *AR* y *MA* presenta la ventaja de ambos, pero también se tiene la desventaja de la resolución de sistemas no lineales para obtener los parámetros.

2.3. Filtros lineales

2.3.1. Definición de filtro

El término filtro usualmente se utiliza para describir un dispositivo que es utilizado para discriminar objetos aplicados en su entrada con algún atributo. Un filtro digital normalmente en la forma de una pieza física, hardware ó software, es aplicado en una serie de datos ruidosos con el fin de extraer una cierta información de interés preestablecida. [Haykin, 1991, p. 1] [Proakis and Manolakis, 2007, p. 291]

2.3.2. Operaciones básicas

Una vez definido el concepto de filtro es útil saber que es lo que se puede realizar con dicho dispositivo, para ello se pueden distinguir tres operaciones básicas que se pueden realizar con él.

- **Filtrado:** Es la acción que realiza un filtro de extraer información de una serie de datos dados $d[n] = u[n]$, en donde $d[n]$ es la señal deseada y $u[n]$ es la entrada actual.

- **Suavizado:** Lo que lo difiere del filtrado, es que la información de interés no se encuentra disponible en el momento t , y la información recabada en un tiempo después del tiempo t puede ser usada para obtener la información pasada. Lo cual significa que en el caso del *suavizado* existe un retraso al producir el resultado de interés. Es por ello que se espera que el suavizado sea una operación mas precisa de lo que es el filtrado. $d[n] = u[n - D]$, para $D > 0$.
- **Predicción:** Es la acción mediante la cual se puede dar un pronóstico de la información que se procesa. $d[n] = u[n + D]$, para $D > 0$.

2.3.3. Definición del filtro lineal

Un filtro se dice que es lineal si las operaciones básicas de filtrado, suavizado y predicción a la salida del dispositivo son una función lineal de las observaciones aplicadas a la entrada del filtro. [Haykin \[1991\]](#)

2.3.4. Planteamiento y Solución al problema del filtrado lineal

En muchas aplicaciones prácticas se dispone de una señal de entrada $u[n]$ que ha sido corrompida por la suma de un ruido no deseado $v[n]$, y es necesario diseñar un filtro que suprima las interferencias no deseadas a la vez que conserve las características de la señal deseada $d[n]$.

Para resolver éste problema, el requerimiento es diseñar un filtro lineal, que tenga como entrada una señal ruidosa, encargado de suprimir las interferencias no deseadas de la señal deseada, es decir, minimizar los efectos del ruido a la salida del filtro.

Existen diferentes enfoques para revolver el problema planteado anteriormente, pero uno de los más utilizados es un enfoque estadístico, para el cual se asume la disponibilidad de los siguientes puntos:

- Ciertos parámetros estadísticos (media, correlación, etc.)
- Una señal útil
- Ruido aditivo no deseado

El enfoque estadístico utilizado normalmente es minimizar el valor medio de la señal de error, que se define como la diferencia entre una respuesta deseada y la salida actual del filtro lineal. Cuando la señal de entrada es una señal estacionaria la

solución resultante es comúnmente conocida como filtro de Wiener. El nombre viene porque fue Wiener quien primero desarrolló la solución a dicho problema para señales en tiempo continuo [Douglas, 1999, p.14], en cuyo caso se dice que la solución es óptima en el sentido del error cuadrático medio.

Algunas de las desventajas en el uso del filtro de Wiener es que es inadecuado para el caso de señales no estacionarias a la entrada, en tales situaciones el filtro asume un comportamiento que varía con el tiempo. Para este caso es posible recomendar un filtro de Kalman. Otro de los requerimientos para poder utilizar un filtro de Wiener como solución al problema del filtrado lineal es el hecho de que es necesario conocer la información estadística de los datos a ser procesados, ya que el filtro sólo será óptimo cuando las características estadísticas de la entrada concuerden con las características estadísticas usadas para diseñar el filtro, de otra manera no será posible diseñar el filtro o puede no comportarse en una manera óptima.

2.4. Estructuras de filtros

Como ya se sabe, un sistema lineal invariante en el tiempo con función de transferencia racional tiene la propiedad de que puede ser representado mediante una ecuación en diferencias lineal con coeficientes constantes, representada en la ecuación A.41.

$$\sum_{k=0}^P a_k y[n-k] = \sum_{j=0}^Q b_j x[n-j], \quad \forall n \quad (2.7)$$

Para la implementación de los filtros digitales lineales, que de acuerdo con lo anterior pueden ser descritos por un sistema lineal invariante en el tiempo, se utiliza hardware digital o algún programa de computadora. Para ello es necesario convertir la función de transferencia o la ecuación en diferencias en algún algoritmo o estructura para que así pueda ser utilizada la tecnología deseada.

La implementación concreta de esas estructuras depende de la tecnología que se desee utilizar. Normalmente un conjunto de operaciones puede ordenarse para construirse diagramas a bloques que constan de elementos básicos como retardo, multiplicador y sumador. Dependiendo de la forma de organizar dichos elementos básicos se pueden definir diferentes estructuras para su implementación en un hardware digital como: microcomputadoras o DSP (Digital Signal Processor), por sus siglas en inglés.

Si el sistema se desea implementar por software, el mismo conjunto de ecuaciones expresado por A.41, puede convertirse en un programa para ser ejecutado por una computadora.

Un punto que se debe de considerar en la elección de alguna estructura en específico, es la complejidad de cálculo. Este factor se refiere al número de operaciones aritméticas (multiplicaciones, divisiones y sumas) que son necesarias para calcular la salida del sistema. Además, también se debe de tener en cuenta la cantidad de memoria necesaria, que corresponde al número de posiciones que se necesitan para almacenar los datos a utilizar en la implementación de cierto algoritmo.

2.4.1. Estructuras de Filtros FIR

Como ya se sabe, un sistema invariante en el tiempo está completamente caracterizado en el dominio del tiempo por la respuesta al impulso $h[n]$. Sin embargo, una subclase de los sistemas invariantes en el tiempo son los sistemas con una respuesta al impulso de duración finita (FIR, *finite-duration impulse response*).

Un sistema FIR tienen una respuesta al impulso que es cero fuera de un determinado intervalo de tiempo finito. Por ello siempre es posible realizarlos mediante un sistema no recursivo con una relación de entrada-salida

$$y[n] = \sum_{k=0}^Q b_k x[n-k] \quad (2.8)$$

Comparando la ecuación anterior con la ecuación 2.6 se puede decir que un modelo de promedio móvil está representado por un sistema FIR.

Forma directa

La realización de la forma directa se obtiene inmediatamente de la ecuación en diferencias no recursiva 2.8, que es una suma ponderada de Q valores pasados de la entrada. Esta estructura necesita Q elementos de memoria para poder almacenar las muestras pasadas de la entrada, además de $Q + 1$ elementos multiplicadores y Q sumadores.

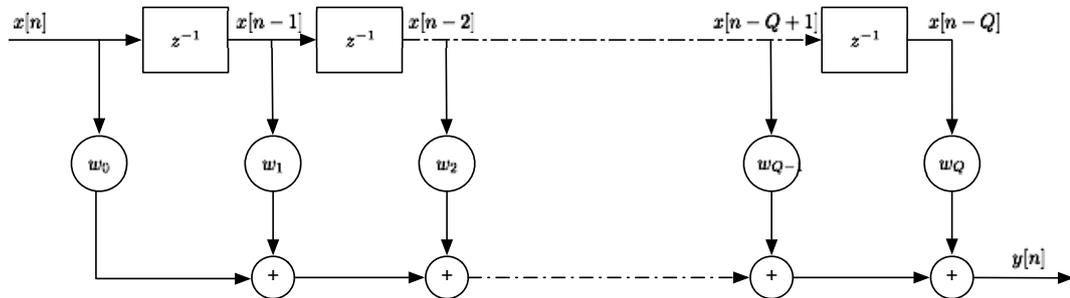


Figura 2.2: Estructura de un filtro FIR en su forma directa

Cascada

La estructura en cascada se obtiene al factorizar el polinomio de la función de transferencia $B[z]$ en los sistemas FIR, de modo que

$$H[z] = \prod_{k=1}^Q H_k[z] \quad (2.9)$$

donde cada elemento $H_k[z]$ son los ceros de la función de transferencia $H[z]$.

$$H_k[z] = (z - z_k), \quad k = 1, 2, \dots, Q \quad (2.10)$$

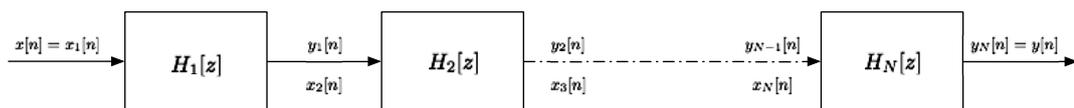


Figura 2.3: Estructura de un filtro FIR en su forma de cascada

Celosía

Los filtros en celosía son un tipo de estructura muy extendido en la implementación de filtros adaptables. Se puede demostrar que una estructura directa se puede implementar mediante una estructura en celosía de la forma representada en la figura 2.4.

En donde

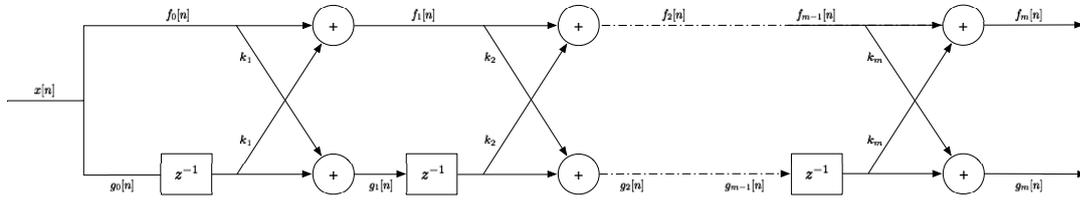


Figura 2.4: Estructura de un filtro FIR en celosía

$$f_0[n] = g_0[n] = x[n] \quad (2.11)$$

$$f_m[n] = f_{m-1}[n] + K_m g_{m-1}[n-1], \quad m = 1, 2, \dots, M-1 \quad (2.12)$$

$$g_m[n] = K_m f_{m-1}[n] + g_{m-1}[n-1], \quad m = 1, 2, \dots, M-1 \quad (2.13)$$

A cada uno de los parámetros K_m se le denomina coeficiente de reflexión.

2.4.2. Estructuras de Filtros IIR

Otra clase de los sistemas lineales invariantes en el tiempo son los sistemas con una respuesta al impulso de duración infinita (IIR, *infinite-duration impulse response*).

Este tipo de sistemas presentan una respuesta al impulso que al ser infinita no puede ser representada mediante un sistema no recursivo, por lo tanto es necesario utilizar las muestras pasadas de la salida para implementarlo. Por ello la ecuación en diferencias que los caracteriza es la ecuación general A.41. Comparando ahora con la ecuación 2.2 se puede concluir que un modelo AR y uno ARMA puede ser representado por un sistema IIR.

Forma directa

La función de transferencia dada por 2.1 que caracteriza a un sistema IIR puede interpretarse como dos sistemas conectados en cascada, es decir

$$H[z] = H_1[z]H_2[z] \quad (2.14)$$

en donde

$$H_1[z] = \sum_{k=0}^Q b_k z^{-k} \quad (2.15)$$

y

$$H_2[z] = \frac{1}{1 + \sum_{k=1}^P a_k z^{-k}} \quad (2.16)$$

Dado que H_1 es un sistema FIR se puede representar por una estructura en forma directa y entonces al conectar en cascada el $H_2[z]$, se obtiene la forma directa mostrada en la imagen 2.5.

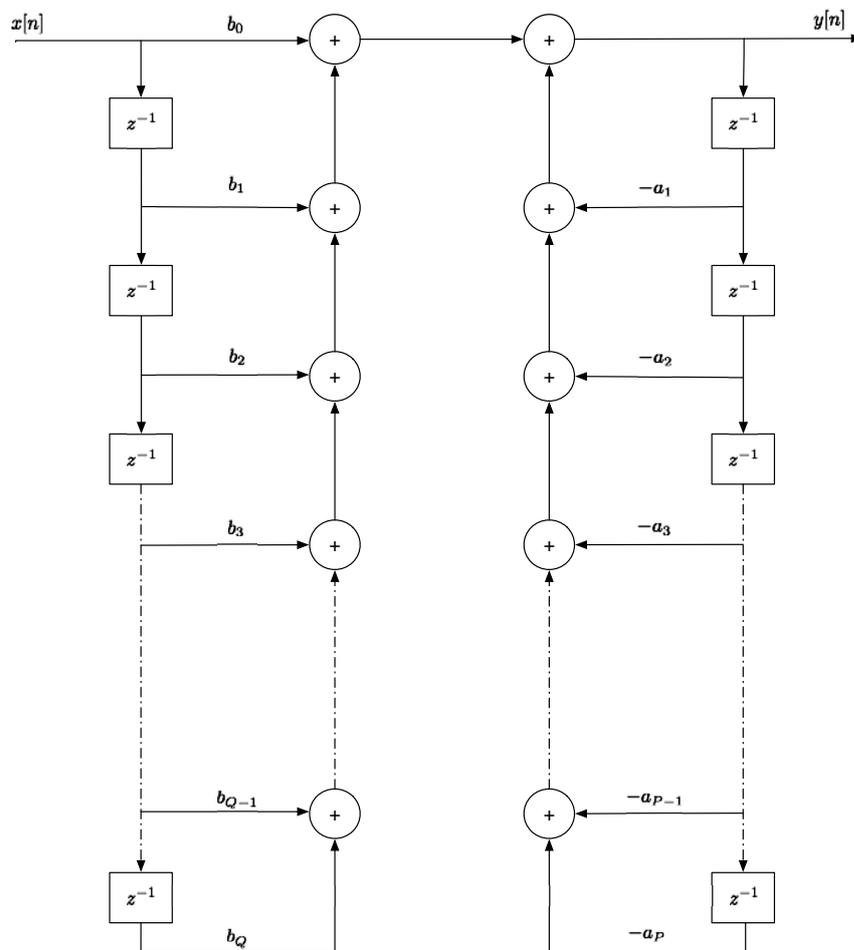


Figura 2.5: Estructura de un filtro IIR en su forma directa

2. MODELOS Y ESTIMACIÓN PARAMÉTRICA EN FILTRADO LINEAL

En la imagen [2.5](#) se puede apreciar que los requerimientos son $Q + P + 1$ multiplicadores, $Q + P$ sumas y $Q + P + 1$ posiciones de memoria.

Filtro de Wiener

Como ya se mencionó anteriormente un enfoque para la solución del problema del filtrado lineal es el enfoque estadístico. Asumiendo que se dispone de los parámetros estadísticos de una cierta señal (valor medio, correlación, etc.), una señal de interés y ruido aditivo no deseado, entonces se requiere diseñar un filtro lineal, que al aplicarle a la entrada una señal, entregue a la salida la señal de interés con un menor efecto del ruido aditivo sobre la señal de interés de acuerdo con algún criterio estadístico.

El punto clave para el diseño del filtro óptimo es encontrar los parámetros óptimos que realicen dicha tarea, un criterio para encontrar estos parámetros es minimizar el *valor cuadrático medio* de la señal de *error*. En donde el error se puede interpretar como una medición de la efectividad del filtro, de tal forma que si el error es muy grande entonces no se realiza de manera correcta la tarea, pero si el error es nulo significa que se presenta el caso ideal en el que el filtro realiza la tarea para la que fue diseñado, eliminar el ruido de la señal de interés, a la perfección.

Cuando una señal estacionaria está involucrada, el filtro lineal que resuelve el problema antes planteado se conoce como *Filtro de Wiener*, el cual se dice que es *óptimo en el sentido del valor cuadrático medio* [Haykin, 1991, p. 2].

3.1. Consideraciones y restricciones

Para poder realizar el diseño de un filtro es necesario conocer de manera precisa los parámetros estadísticos de la señal que se desea procesar. Es sólo en este caso cuando se puede proceder a diseñar el filtro con los parámetros conocidos. En caso de que la información estadística del proceso en cuestión no se conozca por com-

pleto, el filtro no podrá ser diseñado o el filtrado que realice no será el óptimo.

Otra consideración y requerimiento es que la señal que se desea filtrar debe de poder ser modelada mediante un proceso aleatorio estacionario. Esto debido a que el filtro de Wiener no es adecuado para lidiar con señales *no estacionarias*. En tales situaciones la señal cambia sus parámetros estadísticos con el tiempo, por lo cual los parámetros utilizados para el diseño no serán los mismos que los de la señal de entrada y por ende el filtro no será óptimo.

3.2. Filtro lineal óptimo de Wiener

Considérese un filtro lineal discreto en el tiempo descrito por la función de transferencia $H[z]$.

$$H[z] = w_k z^{-k}, \quad k = 0, 1, \dots, P \quad (3.1)$$

Con una entrada $\mathbf{x}[n]$ que representa una señal estacionaria y una salida $\mathbf{y}[n]$. La representación mediante una ecuación en diferencias es:

$$y[n] = \sum_{i=0}^P w_i x[n-i] \quad (3.2)$$

O de la manera matricial la salida en el instante n ($y[n]$) se define como el producto interno de los vectores $\mathbf{w}^*[n]$ y $\mathbf{x}[n]$, ambos de longitud $P + 1$, en donde $*$ denota complejo conjugado, de la siguiente manera:

$$y[n] = \mathbf{w}^*[n] \mathbf{x}[n] \quad (3.3)$$

$$y[n] = [w_0 \quad w_1 \quad \dots \quad w_P]^* \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-P] \end{bmatrix} \quad (3.4)$$

La salida del filtro es utilizada para proveer un estimado de una respuesta deseada $d[n]$. El desempeño del estimador puede ser evaluado mediante el error de estimación $e[n]$ que es definido como la diferencia entre la respuesta deseada y la salida del filtro. Como requerimiento se tiene que lograr que el error sea lo más pequeño posible para así asegurar un óptimo desempeño del filtro en estimar la respuesta deseada.

Los detalles finales para la especificación del filtro dependen de dos elecciones que tienen que ser hechas:

1. La respuesta al impulso del filtro tiene una duración finita o infinita (FIR o IIR).
2. El criterio estadístico que se usará para la optimización.

Para ello se comenzará con la elección de la estructura del filtro y posteriormente se procederá con el criterio estadístico para la optimización. Las consideraciones para la elección de un filtro FIR y uno IIR dependen de la aplicación que se le vaya a dar al sistema, aunque es más sencilla la implementación y diseño del filtro de Wiener mediante un filtro FIR, ya que para un sistema IIR involucra la resolución de un sistema de ecuaciones no lineales. Una vez establecida la respuesta que tendrá el filtro se procede a la elección del criterio estadístico, como ya se mencionó existen diversos métodos o criterios para la optimización del diseño del filtro, en específico se logrará mediante la minimización de una función costo o desempeño [Haykin, 1991, p.16].

3.2.1. Criterio de optimización

En el ámbito de la teoría de optimización, un concepto importante es el de la *función costo*. Una función costo define una transformación de un espacio vectorial hacia un espacio escalar, o de manera más estricta se puede decir que es una función escalar de variable vectorial. Esta función establece una regla de correspondencia del espacio de vectores, que es la variable independiente, hacia un espacio escalar, que es la variable dependiente [Haykin, 1991].

Esta función provee una medición cuantitativa para evaluar la calidad o desempeño. Es por eso que este concepto es básico en la optimización de los filtros digitales. Una vez definidas las características de una función costo es pertinente seleccionar un criterio, entre las cuales una pequeña lista de posibilidades se muestra a continuación :

- Valor medio del error de estimación
- Valor esperado del valor absoluto del error de estimación
- Valor esperado de la tercera o mayor potencia del valor absoluto del error de estimación

Una función costo muy popular para realizar el proceso de optimización es el establecido por el criterio del error cuadrático medio, que se define como el valor cuadrático medio de un error de estimación. El error de estimación se define de la siguiente manera:

$$e[n] = d[n] - y[n] \quad (3.5)$$

$$= d[n] - \mathbf{w}^*[n]\mathbf{x}[n] \quad (3.6)$$

Y la función costo queda definida como:

$$\begin{aligned} J &= E\{|e[n]|^2\} \\ J &= E\{|d[n] - y[n]|^2\} \end{aligned} \quad (3.7)$$

Considerando que la señal $y[n]$ es la señal provista por el filtro lineal óptimo, entonces está en función de la entrada aplicada y los parámetros que describen al filtro. Por lo tanto el proceso de optimización consiste en encontrar los parámetros que minimicen la función costo J . Esta idea establece la base para el diseño de un filtro de Wiener.

En particular y de acuerdo con [Haykin, 1991, p. 159] el criterio del error cuadrático medio resulta en una dependencia de segundo orden para la función costo respecto a los coeficientes desconocidos que caracterizan la respuesta al impulso del filtro. Además, al ser una función de segundo orden cuenta con un único mínimo que define el diseño estadístico óptimo para el filtro.

3.2.2. Planteamiento del problema del filtrado de Wiener

El problema planteado para el filtrado de Wiener se puede enunciar de la siguiente manera: diseñar un filtro lineal discreto cuya salida $y[n]$ provea un estimado de una respuesta deseada, dado un conjunto de muestras a la entrada $\mathbf{x}[n] = \{x[0], x[1], x[2], \dots\}$, tal que el valor medio del error de predicción $e[n]$, definido como la diferencia entre la respuesta deseada y la respuesta actual $y[n]$, sea minimizado. [Haykin, 1991, p. 160]

3.3. Principio de Ortogonalidad

Considerando el enfoque estadístico dado por el filtro de Wiener para la resolución del problema del filtrado lineal, observado en la imagen 3.1.

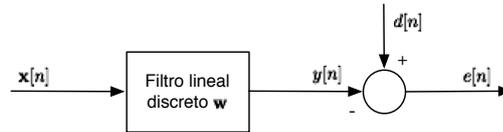


Figura 3.1: Diagrama a bloques que representa el problema del filtrado de Wiener

El propósito del filtro es producir un estimado de la respuesta deseada $d[n]$. Asumiendo que la entrada del filtro $x[n]$ y la respuesta deseada $d[n]$ son secuencias aleatorias discretas estacionarias en sentido amplio, ambas con valor medio cero.

Para optimizar el diseño del filtro, y de acuerdo con el criterio de optimización, la tarea consiste en minimizar el valor cuadrático medio del error de estimación $e[n]$, definido en la ecuación 3.6. Definiendo la función costo del error cuadrático medio:

$$\begin{aligned} J &= E \{ |e[n]|^2 \} \\ J &= E \{ e[n] e^*[n] \} \end{aligned} \quad (3.8)$$

Definiendo el operador ∇_k , como la derivada parcial con respecto del coeficiente w_k del filtro.

$$\nabla_k = \frac{\partial}{\partial w_k} \quad (3.9)$$

El operador gradiente siempre es utilizado en el contexto de encontrar puntos estacionarios de una función¹. Para la función costo J , el obtener su punto mínimo, implica que todos los elementos del vector gradiente $\nabla(J)$ deben de ser cero.

¹En matemáticas, en particular en cálculo, un punto estacionario o punto crítico es un punto en el dominio de una función diferenciable, donde la derivada es igual a cero (equivalentemente, el lugar en donde la pendiente de la gráfica es cero): es un punto en donde la función deja de crecer o decrecer. Para una función diferenciable de muchas variables, un punto estacionario o punto crítico es una entrada (un valor para cada variable) donde todas las derivadas parciales son cero (el gradiente es cero).

Los puntos estacionarios son fáciles de visualizar en una gráfica de la función, corresponden a los puntos donde la tangente es paralela al eje x . Para una función de dos variables corresponde a los puntos en la gráfica donde el plano tangente es paralelo al plano xy

$$\nabla_k(J) = \begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \frac{\partial J}{\partial w_2} \\ \vdots \\ \frac{\partial J}{\partial w_k} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (3.10)$$

o de manera equivalente

$$\nabla_k(J) = 0, \quad k = 0, 1, \dots \quad (3.11)$$

Dado este conjunto de condiciones, el filtro se dice que opera de manera óptima en el sentido del error cuadrático medio. De esta manera, al substituir 3.6 y derivar respecto de \mathbf{w} , se obtiene:

$$\nabla_k(J) = \frac{\partial J}{\partial \mathbf{w}} = E \left\{ \frac{\partial (e[n]e^*[n])}{\partial \mathbf{w}} \right\} \quad (3.12)$$

$$= E \left\{ e[n] \frac{\partial e^*[n]}{\partial \mathbf{w}} + \frac{\partial e[n]}{\partial \mathbf{w}} e^*[n] \right\} \quad (3.13)$$

y dado que

$$\frac{\partial e[n]}{\partial \mathbf{w}} = -\mathbf{x} \quad (3.14)$$

$$\nabla_k(J) = \frac{\partial J}{\partial \mathbf{w}} = E \{ -e[n]\mathbf{x}^*[n] - \mathbf{x}[n]e^*[n] \} \quad (3.15)$$

$$= -2E \{ \mathbf{x}[n]e^*[n] \} \quad (3.16)$$

Entonces es posible determinar las condiciones bajo las cuales el filtro minimiza la función costo J . Supongase e_0 como el valor del error de estimación resultante cuando el filtro opera en su estado óptimo. Las condiciones de operación para el estado óptimo están dadas por 3.10, o lo que es equivalente a

$$E \{ \mathbf{x}[n]e_0^*[n] \} = 0 \quad (3.17)$$

En otras palabras:

La condición necesaria y suficiente para que la función costo J alcance su valor mínimo es que el correspondiente valor del error de predicción $e_0[n]$ sea ortogonal a cada valor de la señal de entrada que se utiliza en la estimación de la respuesta deseada en el tiempo n . [Haykin, 1991, p. 163]

Esta afirmación constituye el *principio de ortogonalidad*, y representa uno de los teoremas más elegantes en el campo del filtrado óptimo, de acuerdo con [Haykin, 1991, p. 163].

Si al principio de ortogonalidad establecido en la expresión 3.17 se premultiplica por el vector de coeficientes del filtro \mathbf{w} se tiene

$$\mathbf{w}E\{\mathbf{x}[n]e_0^*[n]\} = 0 \quad (3.18)$$

$$E\{\mathbf{w}\mathbf{x}[n]e_0^*[n]\} = 0 \quad (3.19)$$

$$E\{y_0[n]e_0^*[n]\} = 0 \quad (3.20)$$

en donde $y_0[n]$ denota la salida del filtro en el instante n cuando está en un funcionamiento óptimo en el sentido del error cuadrático medio. Sea $\hat{d}(n|\mathcal{X}_n)$ el estimado de la respuesta deseada $d[n]$, generado por el espacio \mathcal{X}_n que consiste en una combinación lineal de las variables aleatorias x_1, x_2, \dots, x_n , entonces el estimado óptimo será

$$\hat{d}(n|\mathcal{X}_n) = y_0[n] \quad (3.21)$$

Esta otra forma de expresar el principio de ortogonalidad ofrece una interpretación geométrica, figura 3.2. En donde se puede observar que la diferencia de los vectores d y y_0 dan como resultado el error en estado óptimo e_0 , que representa la parte de d que no puede ser representada por la información disponible, el espacio generado por las muestras pasadas \mathcal{X}_n , y que es ortogonal a y_0 . Además, también se puede visualizar que entre más pequeño sea el error de estimación más tiende a parecerse d y y_0 o en otras palabras cada vez tiene más y_0 de d .

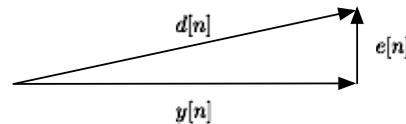


Figura 3.2: Interpretación geométrica del principio de ortogonalidad

Error cuadrático medio mínimo

Cuando el filtro discreto trabaja en su estado óptimo, produce una salida óptima que a su vez da como resultado un error de predicción mínimo.

$$e_0[n] = d[n] - y_0[n] \quad (3.22)$$

$$e_0[n] = d[n] - \hat{d}(n|\mathcal{X}_n) \quad (3.23)$$

reacomodando se tiene

$$d[n] = \hat{d}(n|\mathcal{X}_n) + e_0[n] \quad (3.24)$$

Si se denota el error cuadrático medio mínimo, producido por el filtro en un estado óptimo.

$$J_{min} = E \{ |e_0[n]|^2 \} \quad (3.25)$$

Al aplicar el operador valor esperado en ambos lados de la ecuación 3.24, y utilizar el principio de ortogonalidad, lo que se obtiene es lo siguiente:

$$E \{ |d[n]|^2 \} = E \{ (\hat{d}(n|\mathcal{X}_n) + e_0[n])^2 \} \quad (3.26)$$

$$E \{ |d[n]|^2 \} = E \{ |\hat{d}(n|\mathcal{X}_n)|^2 + 2\hat{d}(n|\mathcal{X}_n)e_0[n] + |e_0[n]|^2 \} \quad (3.27)$$

$$E \{ |d[n]|^2 \} = E \{ |\hat{d}(n|\mathcal{X}_n)|^2 \} + 2E \{ \hat{d}(n|\mathcal{X}_n)e_0[n] \} + E \{ |e_0[n]|^2 \} \quad (3.28)$$

$$(3.29)$$

Aplicando el principio de ortogonalidad, se puede observar que el segundo término del lado derecho es cero ya que la salida óptima estimada $\hat{d}(n|\mathcal{X}_n)$ es ortogonal con $e_0[n]$. Entonces se tiene

$$E \{ |d[n]|^2 \} = E \{ |\hat{d}(n|\mathcal{X}_n)|^2 \} + E \{ |e_0[n]|^2 \} \quad (3.30)$$

$$\sigma_d^2 = \sigma_{\hat{d}}^2 + J_{min} \quad (3.31)$$

Esto al recordar que el valor medio de la señal deseada es cero $\bar{d} = 0$, por lo que el valor cuadrático medio es igual a la variación de la señal σ_d^2 . Lo mismo sucede para el caso de la señal $\hat{d}(n|\mathcal{X}_n)$ y resolviendo para J_{min} , se tiene

$$J_{min} = \sigma_d^2 - \sigma_{\hat{d}}^2 \quad (3.32)$$

Es conveniente normalizar el valor del error cuadrático medio, con el fin de tener un valor que permita medir el desempeño del filtro, esto se consigue al hacer que el valor J_{min} varíe sólo entre cero y uno.

$$\frac{J_{min}}{\sigma_d^2} = 1 - \frac{\sigma_{\hat{d}}^2}{\sigma_d^2} \quad (3.33)$$

Al definir el error cuadrático medio normalizado como:

$$\epsilon = \frac{J_{min}}{\sigma_d^2} \quad (3.34)$$

Se puede observar que el rango de ϵ es $0 \leq \epsilon \leq 1$, en donde si ϵ es igual a cero se dice que el filtro funciona de manera óptima al ser completamente igual la señal estimada $\hat{d}(n|\mathcal{X}_n)$ y la señal deseada $d[n]$. En el caso de que ϵ sea uno entonces se presenta el peor de los casos en el que no hay ninguna concordancia entre la señal estimada y la señal deseada [Haykin, 1991].

3.4. Ecuaciones de Wiener-Hopf

Una vez establecidas las condiciones necesarias y suficientes de la operación óptima del filtro digital, a partir del principio de ortogonalidad, la siguiente tarea es obtener de alguna manera los coeficientes que permitan obtener el error cuadrático medio mínimo.

Al substituir 3.6 en 3.17 se tiene

$$E \{ \mathbf{x}[n] e_0^*[n] \} = 0 \quad (3.35)$$

$$E \{ \mathbf{x}[n] (d^*[n] - \mathbf{x}^*[n] \mathbf{w}_0) \} = 0 \quad (3.36)$$

$$E \{ \mathbf{x}[n] d^*[n] - \mathbf{x}[n] \mathbf{x}^*[n] \mathbf{w}_0 \} = 0 \quad (3.37)$$

$$E \{ \mathbf{x}[n] d^*[n] \} = E \{ \mathbf{x}[n] \mathbf{x}^*[n] \mathbf{w}_0 \} \quad (3.38)$$

$$E \{ \mathbf{x}[n] \mathbf{x}^*[n] \} \mathbf{w}_0 = E \{ \mathbf{x}[n] d^*[n] \} \quad (3.39)$$

En donde los términos del lado derecho e izquierdo se pueden interpretar como sigue.

1. El término $E \{ \mathbf{x}[n] d^*[n] \}$ es la definición de la función de correlación cruzada (B.46) entre la señal de entrada $\mathbf{x}[n]$ y el valor actual de la señal deseada $d^*[n]$, pero que para el caso discreto se define en la ecuación B.134.

$$E \{ \mathbf{x}[n] d^*[n] \} = C_{Xd}[-k] = \mathbf{p} = \begin{bmatrix} p[0] \\ p[-1] \\ \vdots \\ p[1-P] \end{bmatrix} \quad (3.40)$$

3. FILTRO DE WIENER

2. El término $E \{ \mathbf{x}[n] \mathbf{x}^*[n] \}$ es la definición de la matriz de autocorrelación B.136 de la señal $\mathbf{x}[n]$

$$E \{ \mathbf{x}[n] \mathbf{x}^*[n] \} = \mathbf{R} = \begin{bmatrix} r[0] & r[1] & \cdots & r[P-1] \\ r^*[1] & r[0] & \cdots & r[P-2] \\ \vdots & \vdots & \ddots & \vdots \\ r^*[P-1] & r^*[P-2] & \cdots & r[0] \end{bmatrix} \quad (3.41)$$

Por lo tanto se pueden reescribir las condiciones necesarias, al hacer uso de los conceptos de la matriz de autocorrelación y la correlación cruzada, de la siguiente manera:

$$\mathbf{R} \mathbf{w}_0 = \mathbf{p} \quad (3.42)$$

En donde \mathbf{w}_0 representa al vector de coeficientes de longitud $P + 1$, que caracteriza al filtro como se muestra en la figura 3.3. De esta manera, el conjunto de ecuaciones lineales 3.42 define los coeficientes óptimos del filtro en términos de las funciones de autocorrelación de la señal de entrada $\mathbf{x}[n]$ y la correlación con la señal deseada $d[n]$. A este conjunto de ecuaciones se les conoce como *ecuaciones de Wiener-Hopf*.

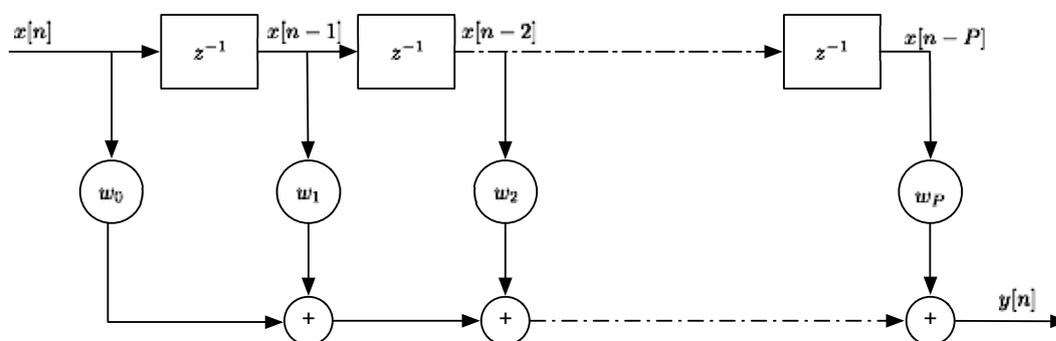


Figura 3.3: Filtro de Wiener FIR con estructura directa, caracterizado por los coeficientes w_0, w_1, \dots, w_P

3.4.1. Solución de ecuaciones de Wiener-Hopf

La solución de las ecuaciones de Wiener-Hopf para sistemas lineales son fáciles de resolver, ya que no involucran ecuaciones no lineales. Teniendo en cuenta que

la matriz de autocorrelación \mathbf{R} , de acuerdo con sus propiedades, es una matriz no singular, es decir existe su inversa \mathbf{R}^{-1} . El sistema de ecuaciones se puede resolver para \mathbf{w}_0 simplemente obteniendo la inversa de la matriz de autocorrelación y pre-multiplicando de ambos lados de la ecuación matricial 3.42.

$$\mathbf{w}_0 = \mathbf{R}^{-1}\mathbf{p} \quad (3.43)$$

De esta forma se puede observar que el calculo de los coeficientes \mathbf{w}_0 , que definen el comportamiento óptimo del filtro FIR de Wiener, requieren conocer sólo dos cantidades. Que son la matriz \mathbf{R} y el vector \mathbf{p} .

3.5. Superficie de desempeño

Otra opción de desarrollo para obtener las ecuaciones de Wiener-Hopf, es revisar la dependencia que existe entre la función costo 3.8 y los coeficientes del filtro. Al expresar la función costo en función de los coeficientes mediante la sustitución de 3.6 y 3.4 se obtiene

$$\begin{aligned} J &= E \{e[n]e^*[n]\} \\ &= E \{(d[n] - y[n])(d[n] - y[n])^*\} \\ &= E \{(d[n] - \mathbf{w}^* \mathbf{x}[n])(d[n] - \mathbf{w}^* \mathbf{x}[n])^*\} \\ &= E \{(d[n] - \mathbf{w}^* \mathbf{x}[n])(d^*[n] - \mathbf{x}^*[n]\mathbf{w})\} \\ &= E \{|d[n]|^2 - \mathbf{w}^* \mathbf{x}[n]d^*[n] - d[n]\mathbf{x}^*[n]\mathbf{w} + \mathbf{w}^* \mathbf{x}[n]\mathbf{x}^*[n]\mathbf{w}\} \\ J &= E \{|d[n]|^2\} - \mathbf{w}^* E \{\mathbf{x}[n]d^*[n]\} - E \{d[n]\mathbf{x}^*[n]\} \mathbf{w} + \mathbf{w}^* E \{\mathbf{x}[n]\mathbf{x}^*[n]\} \mathbf{w} \end{aligned}$$

En donde, se puede observar que se encuentran presentes: la definición de la matriz de autocorrelación B.136 de la señal $\mathbf{x}[n]$ definida mediante el vector B.135 además de la definición de la correlación cruzada B.134 de la señal $\mathbf{x}[n]$ con $d[n]$. Al sustituir estas definiciones en la ecuación anterior y recordar que la varianza de una señal con media cero es igual a su valor cuadrático medio, se obtiene:

$$J = \sigma_d^2 - \mathbf{w}^* \mathbf{p} - \mathbf{p}^* \mathbf{w} + \mathbf{w}^* \mathbf{R} \mathbf{w} \quad (3.44)$$

La ecuación 3.44 establece para el caso de que tanto la entrada del filtro como la señal deseada sean señales estacionarias, la función costo J , o el error cuadrático medio es una función de segundo orden o cuadrática de los coeficientes \mathbf{w} del filtro.

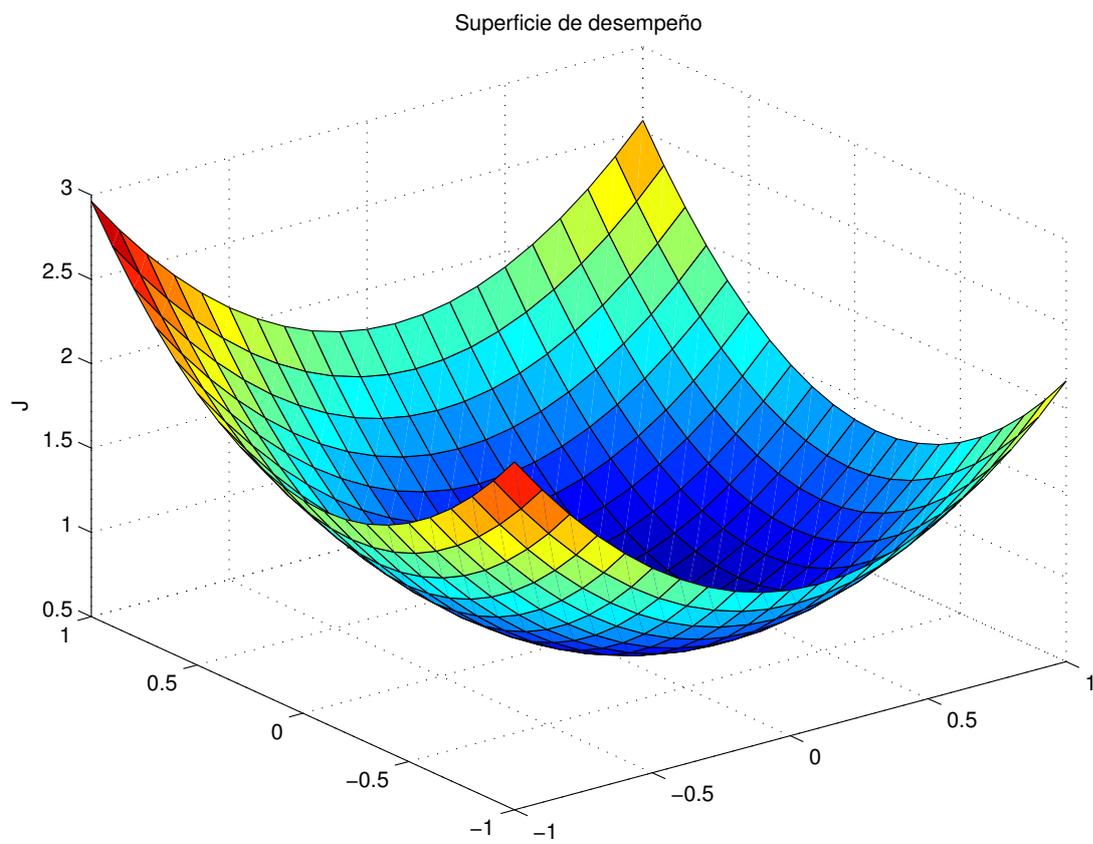


Figura 3.4: Superficie de desempeño de 2 dimensiones

En consecuencia se puede visualizar esta dependencia como una superficie de dimensión $P + 1$ que forma un paraboloide elíptico. La cual se caracteriza por poseer un único mínimo. Esta superficie es lo que se le conoce como superficie de desempeño. En el fondo de la superficie de desempeño, es donde la función costo J alcanza su valor mínimo denotado por J_{min} . Y es donde el gradiente $\nabla(J)$ es idéntico a cero y el plano tangente es paralelo al plano conformado por los coeficientes \mathbf{w} . Al aplicar entonces el operador *nabla* a 3.44 e igualar con cero, se obtiene:

$$\nabla(J) = -2\mathbf{p} + 2\mathbf{R}\mathbf{w} = 0 \quad (3.45)$$

$$\mathbf{R}\mathbf{w} = \mathbf{p} \quad (3.46)$$

Lo cual nos lleva a las ecuaciones de Wiener-Hopf, y al resolver para \mathbf{w} , se encuentran las coordenadas del punto mínimo en la superficie de desempeño.

3.6. Problema planteado

Considérese un proceso autoregresivo $\mathbf{d}[n]$, que es transmitido a través de una línea de transmisión. La salida $\mathbf{u}[n]$ es corrompida tanto por el proceso de transmisión como por ruido gaussiano aditivo que se presenta en la medición a la salida del canal.

3.6.1. Caracterización del problema

El proceso autoregresivo es generado mediante un sistema autoregresivo, que tiene a la entrada un ruido gaussiano \mathbf{v}_1 de media cero y varianza $\sigma_1^2 = 0.27$, el cual está descrito mediante la siguiente función de transferencia:

$$H_1[z] = \frac{1}{1 + 0.8458z^{-1}} \quad (3.47)$$

El canal de transmisión se modela mediante la siguiente función de transferencia:

$$H_2[z] = \frac{1}{1 - 0.9458z^{-1}} \quad (3.48)$$

Finalmente la medición a la salida del canal de transmisión se ve contaminada por un ruido aditivo $\mathbf{v}_2[n]$, en específico un ruido gaussiano con varianza $\sigma_2^2 = 3$.

Para caracterizar y modelar la problemática en un sistema digital, se procede a representar las funciones de transferencia mediante ecuaciones en diferencias de la siguiente manera:

1. La ecuación de transferencia que representa al modelo autoregresivo (3.47), puede representarse mediante un sistema que tiene como entrada a la señal v_1 y a la salida es la representación del proceso autoregresivo que es la señal deseada $\mathbf{d}[n]$. Su representación mediante una ecuación en diferencias es la siguiente:

$$d[n] = v_1[n] - 0.8458d[n - 1] \quad (3.49)$$

2. Mientras tanto el canal se modela mediante el sistema descrito por la función de transferencia 3.48, que tiene a la entrada la señal deseada $\mathbf{d}[n]$, que es transmitida, y a la salida la señal $\mathbf{x}[n]$. Entonces la ecuación que relaciona la salida en el instante $x[n]$ con la entrada $d[n]$ está dada por la siguiente ecuación en diferencias:

$$x[n] = d[n] + 0.9458x[n - 1] \quad (3.50)$$

3. Finalmente el ruido aditivo $\mathbf{v}_2[n]$ se suma a la señal de salida $\mathbf{x}[n]$, lo que da como resultado la señal medida a la salida del canal de transmisión, denotada por $\mathbf{u}[n]$.

$$\mathbf{u}[n] = \mathbf{x}[n] + \mathbf{v}_2[n] \quad (3.51)$$

3.6.2. Resolución del problema

El objetivo es diseñar un filtro lineal óptimo, Filtro de Wiener, que se colocará a la salida de la línea de transmisión con el fin de minimizar la distorsión debida al canal de transmisión y la interferencia del ruido aditivo en la medición, para obtener un estimado de la señal deseada $\hat{\mathbf{d}}_1[n]$ lo más próxima a la señal deseada $\mathbf{d}[n]$.

Filtro de Wiener de distintos órdenes

Para resolver el problema del filtro óptimo, en el sentido del error cuadrático medio, se deben de encontrar los parámetros que minimicen el valor cuadrático

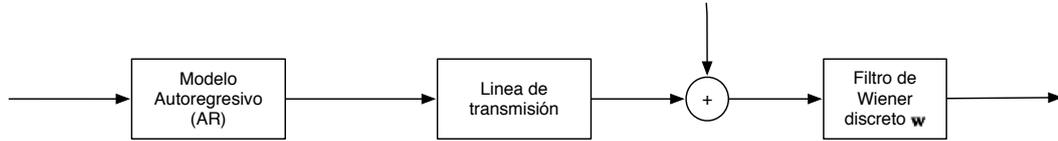


Figura 3.5: Diagrama a bloques de aplicación del filtro de Wiener

medio del error entre la señal estimada $\hat{\mathbf{d}}[n]$ y la señal deseada $\mathbf{d}[n]$.

Suponiendo que el filtro tiene una estructura directa como la mostrada en la imagen 3.3. Entonces la tarea consiste en encontrar los valores de los coeficientes, w_0, w_1, \dots, w_P del filtro de orden P , mediante la resolución de las ecuaciones de *Wiener-Hopf*.

Definiendo a \mathbf{R} como la matriz de autocorrelación de la entrada $\mathbf{u}[n]$ de orden P como:

$$\mathbf{R} = \begin{bmatrix} r[0] & r[1] & \cdots & r[P] \\ r^*[1] & r[0] & \cdots & r[P-1] \\ \vdots & \vdots & \ddots & \vdots \\ r^*[P] & r^*[P-1] & \cdots & r[0] \end{bmatrix} \quad (3.52)$$

y al vector de correlación cruzada entre la entrada $\mathbf{u}[n]$ y la señal deseada $\mathbf{d}[n]$ como:

$$\mathbf{p} = \begin{bmatrix} p[0] \\ p[-1] \\ \vdots \\ p[-P] \end{bmatrix} \quad (3.53)$$

Finalmente el vector de los coeficientes óptimos está dado por la ecuación

$$\mathbf{w}_0 = \mathbf{R}^{-1} \mathbf{p} \quad (3.54)$$

Simulaciones

Como resultado de obtener los coeficientes \mathbf{w}_0 para distintos ordenes del filtro de Wiener, se comprobó el resultado del filtro mediante las siguientes simulaciones:

3. FILTRO DE WIENER

Filtro de Wiener de orden $P = 0$. Para el filtro de Wiener de orden $P = 0$ sólo es necesario contar con un sólo coeficiente además de los valores instantáneos de la señal $\mathbf{u}[n]$. El desempeño del filtro se muestra en la gráfica 3.6.

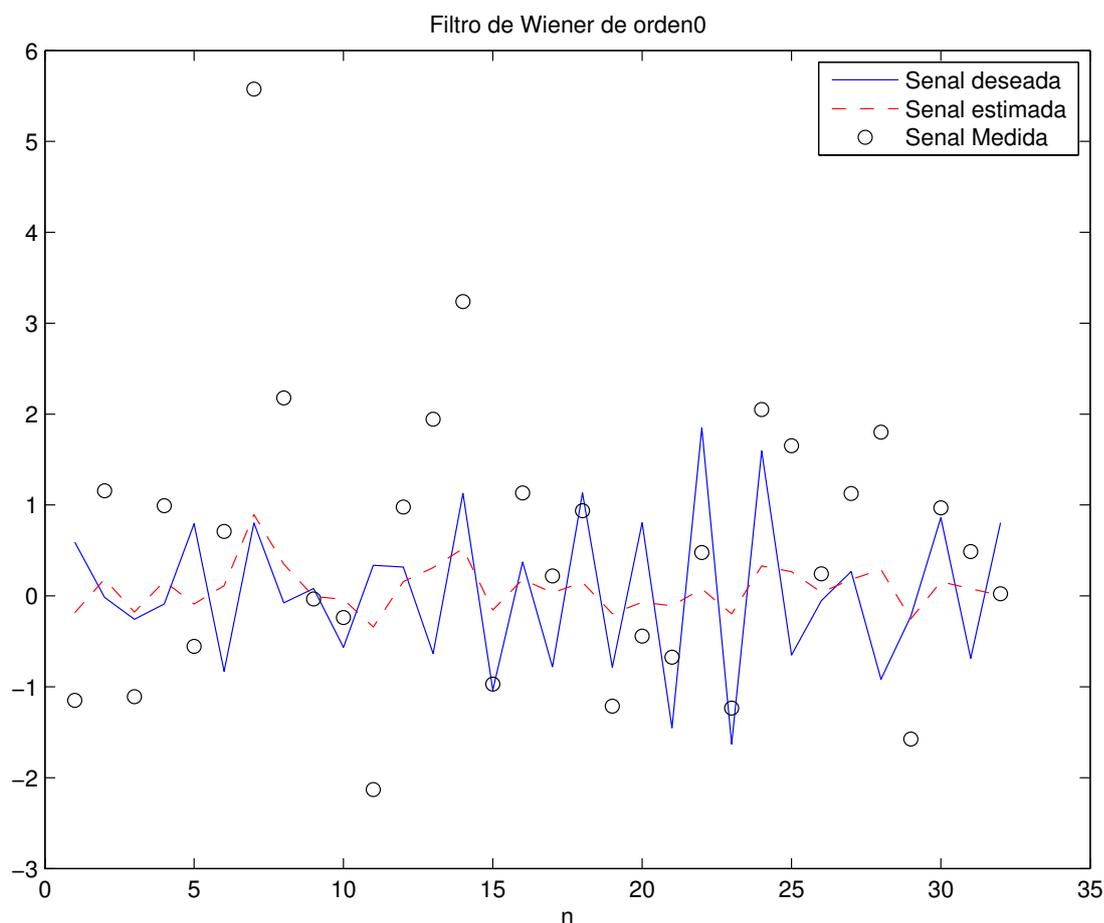
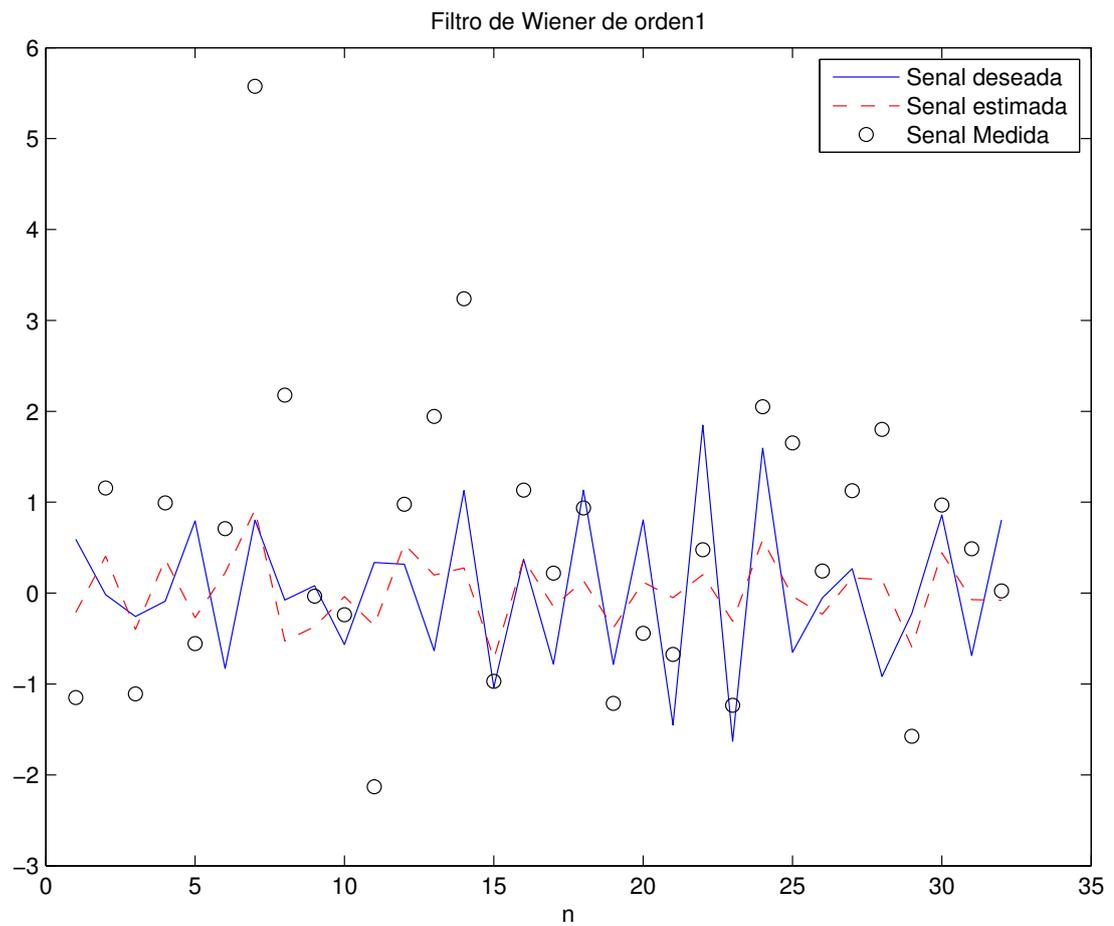


Figura 3.6: Filtro de Wiener de orden 0

Filtro de Wiener de orden $P = 1$. Para el filtro de Wiener de orden $P = 1$ sólo es necesario calcular dos coeficientes (w_0 y w_1) además de los valores instantáneos y anterior de la señal $\mathbf{u}[n]$. El desempeño del filtro se muestra en la gráfica 3.7.

Filtro de Wiener de orden $P = 2$. Para el filtro de Wiener de orden $P = 2$ se calcularán tres coeficientes y el filtro necesitará almacenar dos valores pasados de la secuencia de entrada $\mathbf{u}[n]$, es decir $\mathbf{u}[n - 1]$ y $\mathbf{u}[n - 2]$. El desempeño del filtro se muestra en la gráfica 3.8.

**Figura 3.7:** Filtro de Wiener de orden 1

3. FILTRO DE WIENER

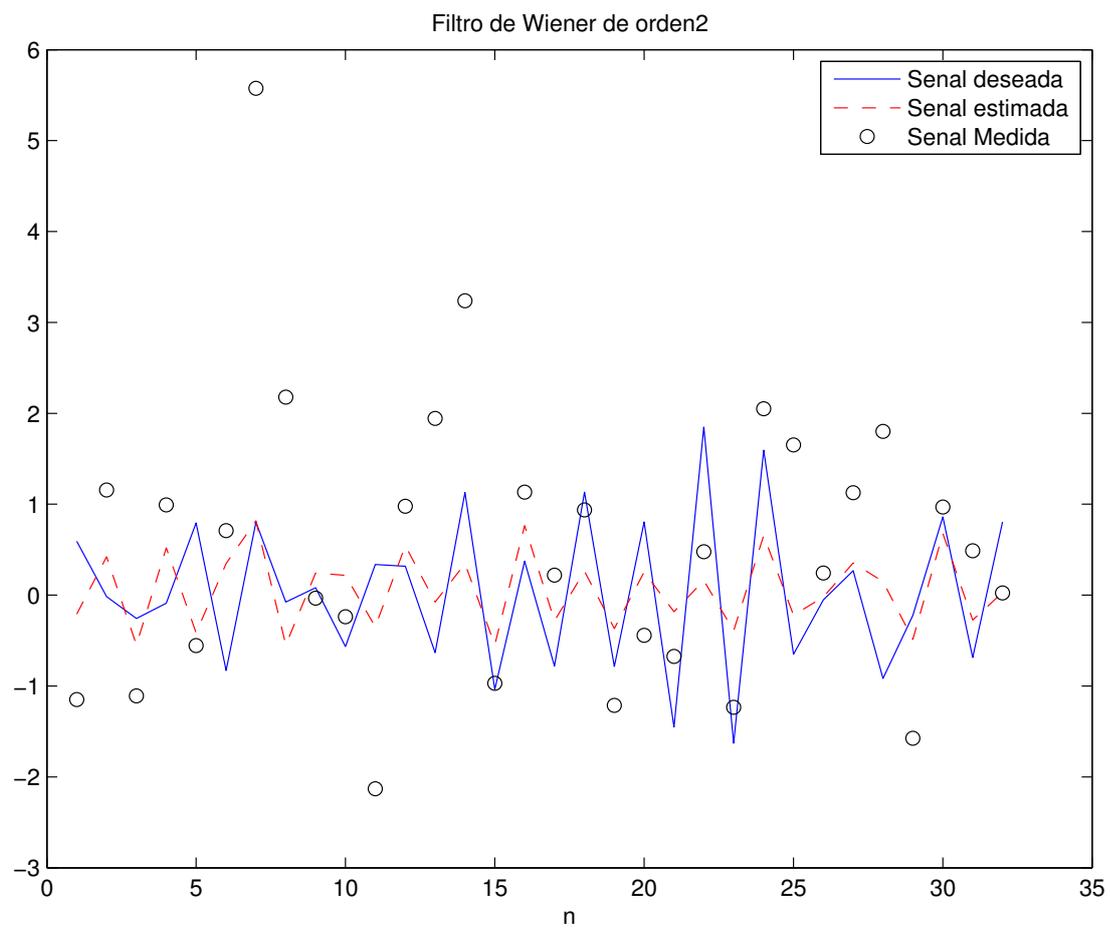
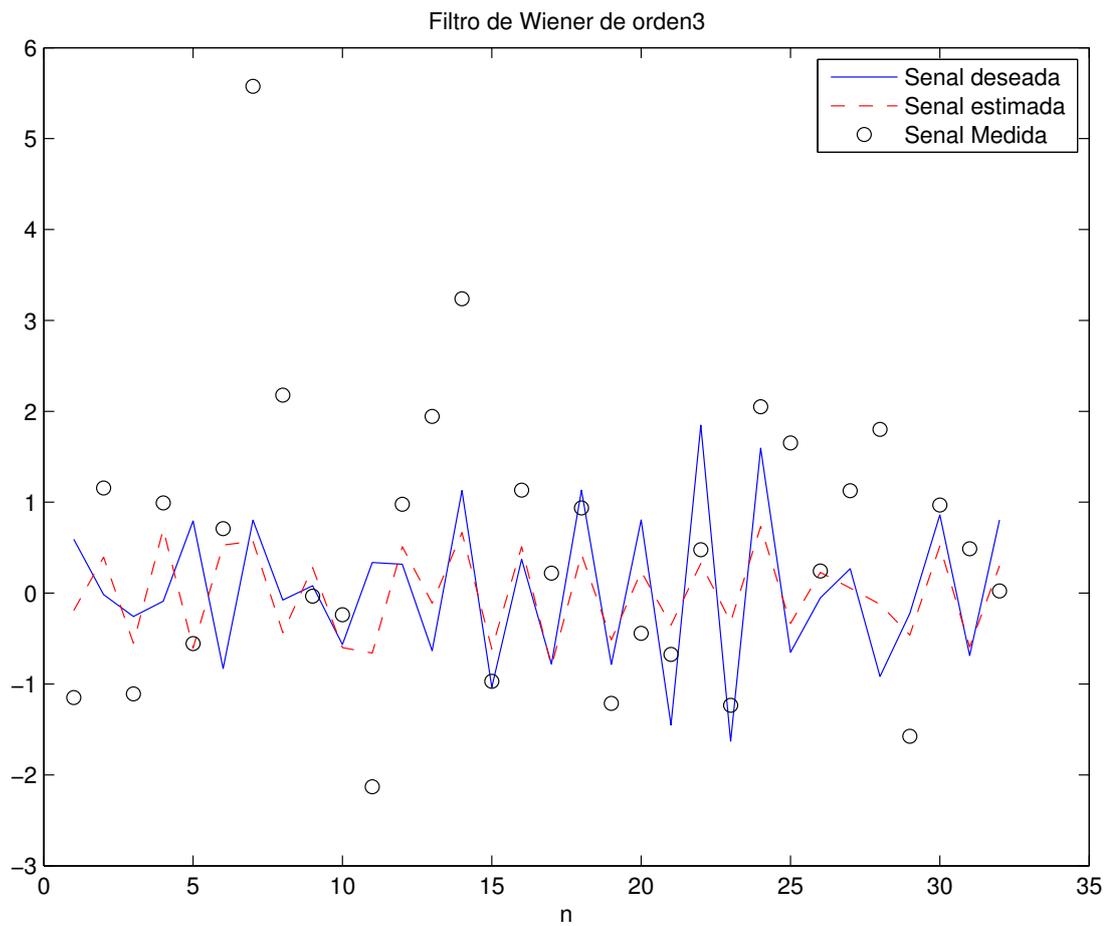


Figura 3.8: Filtro de Wiener de orden 2

**Figura 3.9:** Filtro de Wiener de orden 3

3. FILTRO DE WIENER

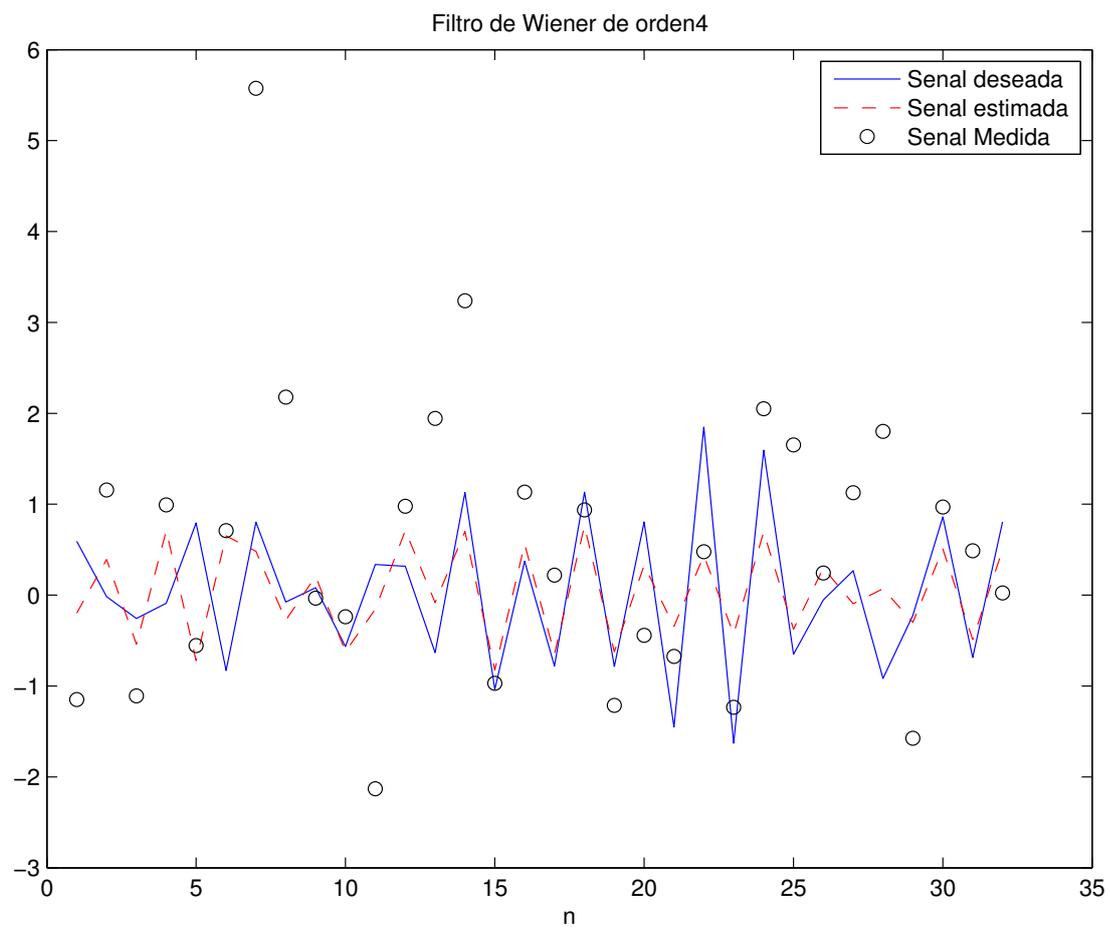
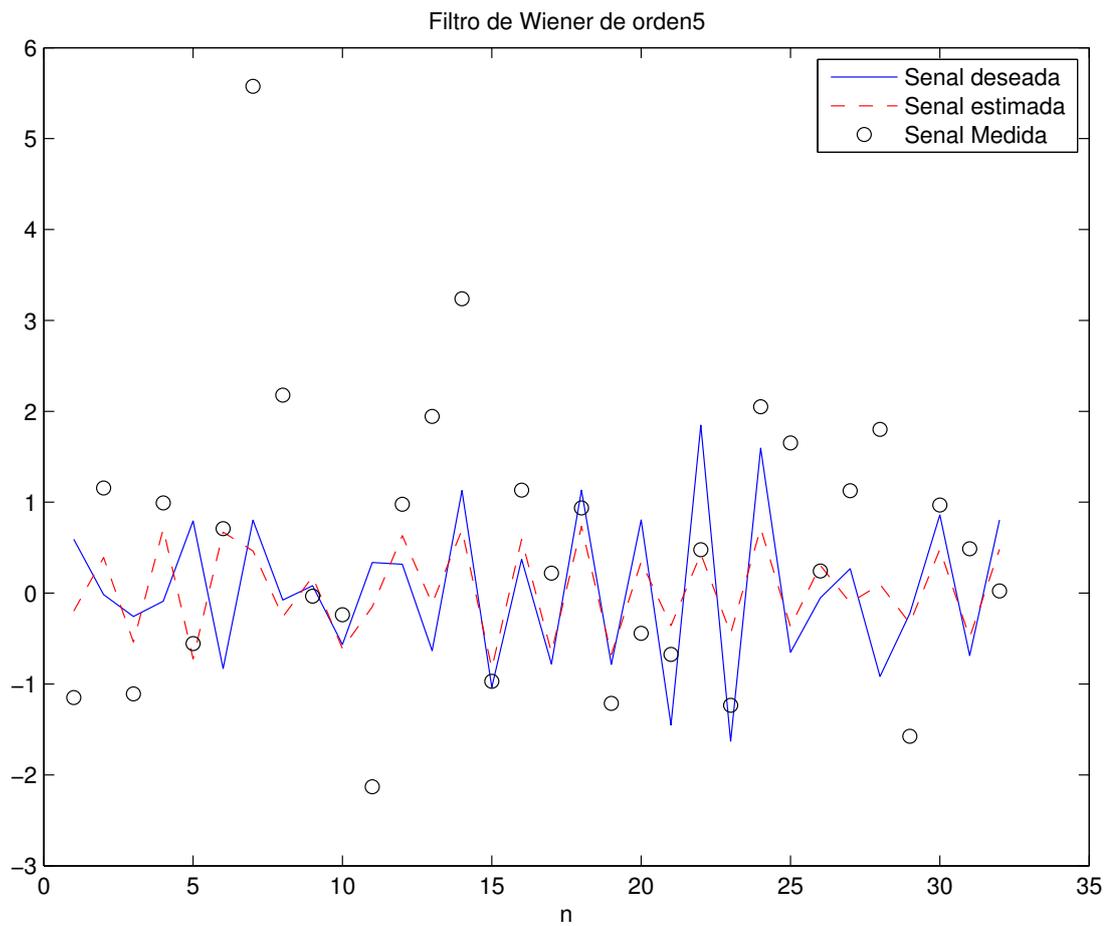


Figura 3.10: Filtro de Wiener de orden 4

**Figura 3.11:** Filtro de Wiener de orden 5

Comparación del desempeño. Una manera de comparar el desempeño de cada uno de los filtros de Wiener obtenidos, es mediante la comparación del error cuadrático medio mínimo J_{min} . En la gráfica 3.12 se puede observar que, el error cuadrático medio mínimo disminuye a medida que incrementa el orden del filtro.

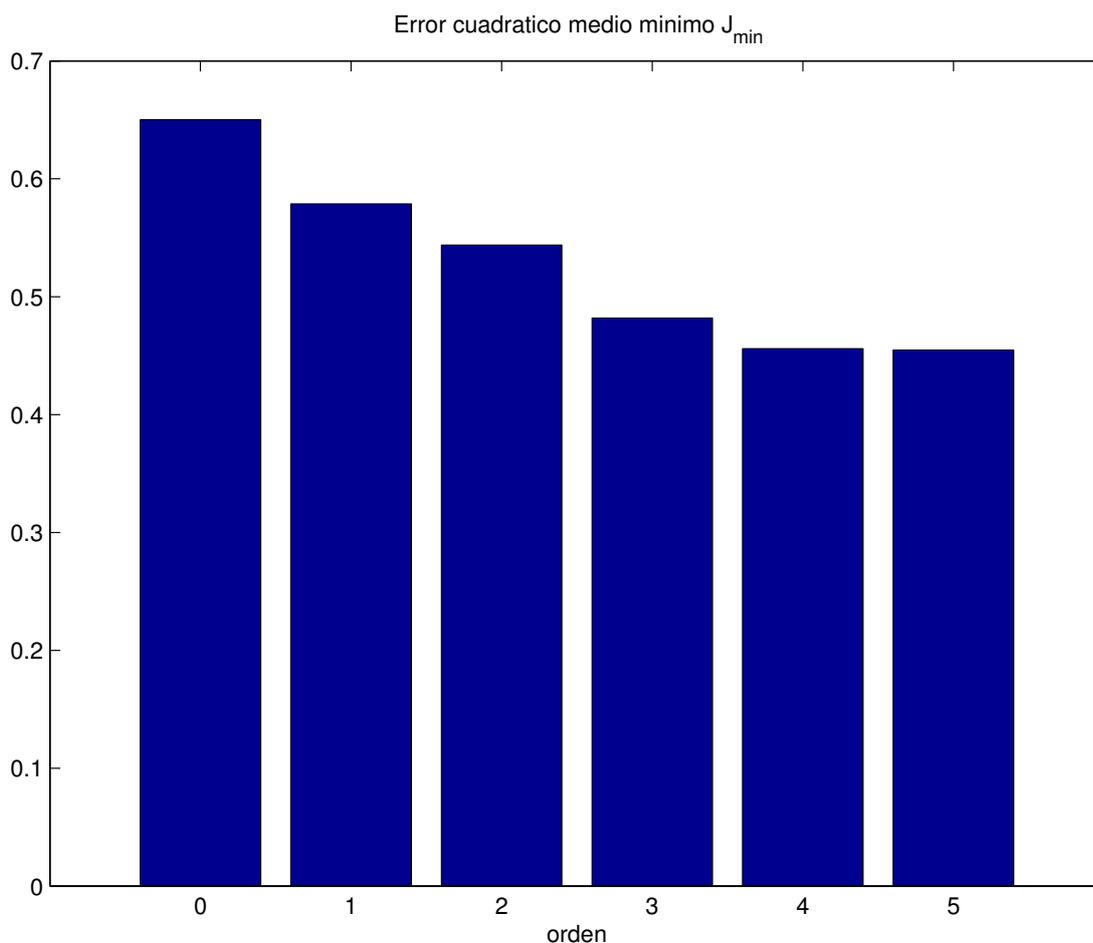


Figura 3.12: Gráfica del error cuadrático medio mínimo J_{min} para distintos órdenes

3.7. Conclusiones

El filtro de Wiener es un filtro discreto que debido al criterio de diseño, minimización del error entre una señal deseada y la señal estimada de esta señal dada por

el filtro, se considera un filtro óptimo en el sentido del valor cuadrático medio.

Una característica que en lo personal me parece importante es que al al considerarse un filtro con una estructura FIR, sólo es necesario calcular los coeficientes w_0 para caracterizar por completo al filtro, esto permite que la implementación del mismo sea muy sencilla en comparación con algún otro algoritmo, en el sentido de que la complejidad de cálculo depende sólo de operaciones básicas: multiplicación, suma y elementos de memoria.

Pero este filtro a pesar de ser óptimo tiene diversas limitaciones ó dificultades que pueden ver afectado el desempeño de la estimación.

Una de las dificultades que se presentan es que se deben de conocer de manera precisa los parámetros estadísticos de la señal que se desea estimar, los cuales además deben de permanecer constantes.

En otras palabras, la señal que se desea estimar debe de ser un proceso aleatorio estacionario con parámetros conocidos. Esto lleva a la limitante de que en el caso de que no se conozcan los parámetros del proceso o que los dichos parámetros no se mantengan constantes, provocaría que el filtro no fuese óptimo y en dado caso tendía que volverse a calcular los coeficientes que caracterizan al filtro.

Por lo anterior, se puede considerar que el filtro de Wiener funciona bien y es ampliamente recomendable para entornos estáticos, es decir que no cambian con el tiempo, como lo pueden ser señales que se tienen almacenadas para un procesamiento posterior. De esta manera se asegura el funcionamiento óptimo del filtro, el cual no podría alcanzarse al encontrarse en un entorno dinámico.

Filtro de Kalman

4.1. Filtros Adaptables

El filtro de Wiener es inadecuado para lidiar con señales no estacionarias, señales en las que sus parámetros estadísticos cambian con el tiempo. En tales situaciones, el filtro óptimo debe de tener una forma variante en el tiempo también. Como se mencionó anteriormente el diseño de un filtro de Wiener requiere el conocimiento de las características estadísticas de la señal de entrada. Cuando esta información estadística no se conoce un método eficiente es utilizar un filtro adaptable.

Un filtro adaptable es un dispositivo que se autoconfigura ó autodiseña basando su operación en un algoritmo recursivo, el cual comienza a partir de condiciones iniciales que representan un desconocimiento total del entorno.

Un filtro adaptable es útil tanto para señales estacionarias como para señales no estacionarias. En un ambiente estacionario, después de varias iteraciones, el algoritmo converge hacia la solución óptima de Wiener. Al tratarse de señales no estacionarias, el algoritmo tiene la habilidad conocida como *seguimiento* (tracking), gracias a la cual puede seguir las variaciones que se pudieran presentar en los parámetros estadísticos de la entrada [Haykin, 1991, p. 3].

Como consecuencia directa de la aplicación de un algoritmo recursivo, en donde los parámetros de un filtro adaptable se actualizan con cada iteración, los parámetros se vuelven independientes de los datos a procesar. Esto por lo tanto significa que, un filtro adaptable es un sistema no lineal en el sentido de que no cumple con el principio de superposición.

En otro contexto, existe también el término *filtro lineal adaptable* que normalmente se refiere a la linealidad en el sentido de que el estimado de la señal que se obtiene mediante un algoritmo de filtrado adaptable, que utiliza una combinación lineal del conjunto de observaciones aplicadas a la entrada del filtro para generar la salida deseada.

Funcionamiento y enfoque de un filtro adaptable

El funcionamiento de un filtro adaptable involucra un algoritmo en el que se pueden identificar dos grandes procesos:

1. **Proceso de filtrado:** Diseñado para producir una salida en respuesta a una secuencia de entrada determinada.
2. **Proceso adaptable:** Este proceso tiene como tarea proveer un mecanismo para el ajuste adaptable del conjunto de parámetros utilizados en el proceso de filtrado.

Sabiendo los dos procesos que se pueden identificar a grandes rasgos, es posible identificar y resolver fácilmente el primer proceso mediante alguna estructura de un filtro digital. Para el caso de proceso adaptable no hay una solución única. El reto que se debe de superar es entender las ventajas y desventajas que puede presentar cada algoritmo, y seleccionar alguno que sea apropiado para la aplicación requerida.

Básicamente, se pueden identificar tres distintos métodos o enfoques que derivan en algoritmos recursivos, necesarios para la operación de un filtro adaptable.

1. Enfoque basado en la teoría del filtro de Wiener.
2. Enfoque basado en la teoría del filtro de Kalman.
3. Enfoque basado en el método de los Mínimos cuadrados.

4.1.1. Enfoque de operación basado en la teoría del filtro de Kalman

El problema del filtrado de Kalman para un sistema lineal dinámico se formula en términos de dos ecuaciones básicas:

Ecuación del proceso: Es la ecuación que describe la dinámica del sistema en términos de un vector de estados.

Ecuación de medición: Es la ecuación que describe las mediciones de los errores que pudiese presentar el sistema.

La solución al problema se expresa mediante un conjunto de expresiones matriciales recursivas. En este caso la implementación no se basa en una estructura directa, si no en una implementación basada en el espacio de estados. Para este caso los parámetros que caracterizan al filtro pueden ser identificados como el vector de estado.

Cuando una señal estacionaria se utiliza, el modelo se mantiene fijo, lo que significa que los parámetros del filtro se mantienen constantes, para el caso de señales no estacionarias, se recurre al uso de un modelo ruidoso en el cual los parámetros del sistema se mueven al rededor de un valor medio que caracteriza al sistema.

Las ventajas de este tipo de algoritmos que se basan en la teoría del filtrado de Kalman, es que tienen una velocidad de convergencia mucho mayor, comparados con aquellos basados en el método de los mínimos cuadrados, además de ser robustos y del hecho que pueden trabajar con señales estacionarias y no estacionarias. Pero presentan la desventaja que son complejos en un sentido computacional, es decir requieren de cálculos complejos[Haykin, 1991, p.2-13].

4.2. Filtro de Kalman

El filtro de kalman es un conjunto de ecuaciones matemáticas que proveen un cálculo eficiente para estimar el estado de un proceso, en el sentido de que minimiza el error cuadrático medio. El filtro es muy poderoso en distintos aspectos entre ellos tiene la capacidad de realizar una estimación aún cuando la el modelo preciso de un sistema no se conoce[Welch and Bishop, 2006].

El filtro de Kalman es un algoritmo de procesamiento óptimo recursivo, o estimador cuadrático lineal. Pero existen muchos criterios para determinar que un algoritmo es óptimo, y dependen del elegido para evaluar su desempeño. Se dice que es óptimo ya que utiliza toda la información disponible a lo largo del tiempo, con ruido inclusive, y produce un estimado del valor de la variable actual que tiende a ser más precisa que aquellos valores basados en una sola medición[Maybeck, 1979].

4. FILTRO DE KALMAN

Para poder dar el estimado de las variables de interés, es necesario tener a la mano una serie de requisitos, los cuales son:

1. Conocimiento de la dinámica del sistema y la dinámica de medición.
2. Una descripción estadística del ruido del sistema.
3. Cualquier información disponible acerca de las condiciones iniciales del sistema.

Una de las ventajas que presenta el filtro de Kalman es el uso de toda la información disponible a lo largo del tiempo, sin necesidad de almacenar todas las mediciones anteriores, esto se logra gracias a la naturaleza recursiva del algoritmo. La palabra recursiva significa que el filtro solamente necesita almacenar el valor anterior en el cual tiene incluida la información recabada hasta ese momento.

Este filtro está basado en la idea del estimador de varianza mínima (LVM) de sistemas discretos. Se dice que un estimador es sin sesgo si el valor medio de la salida del estimador es igual al valor medio de la cantidad estimada.

$$E\{\hat{\mathbf{x}}\} = E\{\mathbf{x}\} \quad (4.1)$$

De acuerdo con [Reid, 2001] un estimador de varianza mínima sin sesgo (MVUE) que minimiza el error cuadrático medio está dado por:

$$\hat{\mathbf{x}} = \underset{\hat{\mathbf{x}}}{\operatorname{arg\,min}} E\{|\hat{\mathbf{x}} - \mathbf{x}|^2 | z\} = E\{\mathbf{x} | z\} \quad (4.2)$$

en donde el término $E\{|\hat{x} - x|^2\}$ es la varianza del error, el cual se encuentra estrechamente relacionado con la matriz de covarianza del error

$$E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^*\} \quad (4.3)$$

Específicamente la varianza del error de un estimador es igual a la traza¹, denotada por $tr(\cdot)$, de la matriz de covarianza del error.

$$E\{|\hat{x} - x|^2\} = tr(E\{(\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^*\}) \quad (4.4)$$

Por lo tanto el filtro de Kalman es un filtro de varianza mínima del error, el cual es el mejor de todas las clases de filtros. En el caso de que el vector de estado y el vector de medición, sean variables Gaussianas el MVUE es una función lineal de las mediciones pasadas y por lo tanto MVUE es también un estimador LMV[Reid, 2001].

¹En álgebra lineal, la traza de una matriz cuadrada A de $n \times n$ está definida como la suma de los elementos de la diagonal principal de A .

4.2.1. ¿Por qué el filtro de Kalman?

De acuerdo con [Kalman, 1960] existe una clase importante de problemas teóricos y prácticos en el área de las comunicaciones y el control que tienen una naturaleza aleatoria que motivaron a Rudolf Emil Kálmán al diseño de un filtro. Tales problemas son:

1. Predicción de señales aleatorias.
2. Separación de señales aleatorias de ruido aleatorio.
3. Detección de señales con una forma conocida en la presencia de ruido aleatorio.

De acuerdo con él, Wiener demostró que los problemas de predicción y separación se pueden realizar mediante la teoría que conlleva a las ecuaciones de *Wiener-Hopf*, además de que establece un método para la resolución de dichas ecuaciones en presencia del caso especial de señales estacionarias.

Desde entonces muchos métodos han sido presentados para resolver el filtrado de Wiener, pero todos ellos cuentan con numerosas limitaciones que reducen ampliamente su utilidad práctica.

Estas limitaciones, según [Maybeck, 1979], del porqué sistemas determinísticos y las teorías de control no proveen los suficientes medios para analizar (observar el comportamiento actual de un sistema, y proveer entradas que puedan generar una respuesta deseada) y diseñar, son las siguientes:

1. Ningún modelo matemático es perfecto.
2. Los sistemas, son controlados no solo por las señales de entrada, también por las perturbaciones que no se pueden controlar o modelar de manera determinística.
3. Los sensores no proveen datos completamente perfectos acerca de un sistema, estos dispositivos siempre presentan ruido.

4.3. Problema del filtrado de Kalman

4.3.1. Modelo en espacio de estados

El filtro de Kalman plantea el problema general de intentar estimar el vector de estados $\mathbf{x} \in \mathbb{R}^n$ de un proceso en tiempo discreto que está descrito por la ecuación en diferencias

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k] + \mathbf{B}\mathbf{u}[k] + \mathbf{w}[k] \quad (4.5)$$

con una medición $\mathbf{z} \in \mathbb{R}^n$

$$\mathbf{y}[k] = \mathbf{C}\mathbf{x}[k] + \mathbf{v}[k] \quad (4.6)$$

En donde las variables \mathbf{w} y \mathbf{v} representan el ruido del proceso y de medición respectivamente, las cuales se asume que son independientes gaussianas, con media cero y varianza \mathbf{Q} y \mathbf{R} respectivamente.

$$\begin{aligned} p(\mathbf{w}) & N(0, \mathbf{Q}) \\ p(\mathbf{v}) & N(0, \mathbf{R}) \end{aligned} \quad (4.7)$$

En la práctica las matrices de covarianza \mathbf{Q} y \mathbf{R} de los procesos pueden cambiar con el tiempo, pero en esta ocasión se asumirá que se mantienen constantes.

4.3.2. Proyecciones ortogonales

De acuerdo con [Kalman, 1960] el cálculo explícito de un estimado óptimo como una función de los valores observados de las variables es, en general, imposible. Pero existe una excepción, si dos procesos $X_1[t]$ y $X_2[t]$ son gaussianos.

Considerese las variables aleatorias $y[t_0], \dots, y[t]$. El conjunto de todas las combinaciones lineales de estas variables aleatorias con coeficientes reales

$$\sum_{i=t_0}^t a_i y[i] \quad (4.8)$$

forman un espacio vectorial denotado por $\mathcal{Y}[t]$. Se confía entonces que cualquier expresión con la forma 4.8 es un punto o un vector en el espacio $\mathcal{Y}[t]$. Utilizando el proceso de ortogonalización descrito por Schmidt en [Schmidt, 1907] es posible encontrar una base ortonormal en $\mathcal{Y}[t]$, esto significa que un conjunto de

vectores denotados por e_{t_0}, \dots, e_t pertenecientes a $\mathcal{Y}[t]$ pueden representar cualquier otro vector en $\mathcal{Y}[t]$ como una combinación lineal única de e_{t_0}, \dots, e_t . Si además

$$E\{e_i e_j\} = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (4.9)$$

Entonces cualquier vector $\mathbf{x} \in \mathcal{Y}[t]$, está dado por:

$$\mathbf{x} = \sum_{i=t_0}^t a_i e_i \quad (4.10)$$

y los coeficientes a_i pueden calcularse con ayuda de 4.9 de la siguiente manera:

$$E\{\mathbf{x} e_j\} = E\left\{\left(\sum_{i=t_0}^t a_i e_i\right) e_j\right\} = \sum_{i=t_0}^t a_i E\{e_i e_j\} = \sum_{i=t_0}^t a_i \delta_{ij} = a_j \quad (4.11)$$

Una vez establecido lo anterior se establece que cualquier variable aleatoria x , perteneciente ó no, a $\mathcal{Y}[t]$ puede ser descompuesta de manera única en dos partes:

1. La parte de x perteneciente a $\mathcal{Y}[t]$, $\bar{x} \in \mathcal{Y}[t]$.
2. La parte de x no perteneciente a $\mathcal{Y}[t]$ ó dicho de otra manera la parte x ortogonal a $\mathcal{Y}[t]$, $\tilde{x} \notin \mathcal{Y}[t]$.

Esto se puede expresar matemáticamente de la siguiente manera:

$$\mathbf{x} = \bar{\mathbf{x}} + \tilde{\mathbf{x}} = (E\{\mathbf{x} e_i\}) e_i + \tilde{\mathbf{x}} \quad (4.12)$$

Para comprobar que $\tilde{\mathbf{x}}$ es ortogonal a $\mathcal{Y}[t]$, se obtienen las coordenadas de $\tilde{\mathbf{x}}$ respecto a la base e_{t_0}, \dots, e_t de la misma manera que con $\bar{\mathbf{x}}$.

$$E\{\tilde{\mathbf{x}} e_i\} = E\{(\mathbf{x} - \bar{\mathbf{x}}) e_i\} = E\{\mathbf{x} e_i\} - E\{\bar{\mathbf{x}} e_i\} \quad (4.13)$$

Dado que $E\{\bar{\mathbf{x}} e_i\} = 0$, ya que es ortogonal al espacio $\mathcal{Y}[t]$ y por lo tanto a cada base e_i , las coordenadas de $\tilde{\mathbf{x}}$ respecto a la base e_{t_0}, \dots, e_t son nulas. Por lo tanto se dice que $\tilde{\mathbf{x}}$ es la *proyección ortogonal* de \mathbf{x} sobre $\mathcal{Y}[t]$.

Utilizando el *Teorema 2* establecido por Kalman, de igual manera en [Kalman, 1960], el cual dice lo siguiente:

Teorema 2. Sea $\{x[t]\}, \{y[t]\}$ procesos aleatorio con media cero ($E\{x[t]\} = E\{y[t]\} = 0$ para todo t). Observando $y[t_0], \dots, y[t]$. Si cualquiera de las siguientes:

- Los procesos aleatorios $\{x[t]\}, \{y[t]\}$ son gaussianos

- El estimado óptimo es restringido a ser una función lineal de las variables aleatorias observadas

Entonces

$$\hat{\mathbf{x}}[t_1|t] = \text{estimado óptimo de } \mathbf{x}[t_1] \text{ dado } \mathbf{y}[t_0], \dots, \mathbf{y}[t] \quad (4.14)$$

$$= \text{proyección ortogonal } \bar{\mathbf{x}}[t_1|t] \text{ de } \mathbf{x}[t_1] \text{ en } \mathcal{Y}[t] \quad (4.15)$$

Otra notación utilizada es:

$$\bar{\mathbf{x}}[t_1|t] = \hat{\mathbf{x}}[t_1|t] = \hat{E}\{\mathbf{x}[t_1]|\mathcal{Y}[t]\} \quad (4.16)$$

Este teorema establece que el estimado óptimo bajo las dos condiciones establecidas son una combinación lineal de todas las observaciones pasadas. Como consecuencia se puede ver que el estimado puede ser la salida de un filtro lineal que tiene a la entrada los valores observables de las variables aleatorias.

4.3.3. Solución al problema de filtrado

Considerando el modelo dinámico de un sistema discreto en variables de estado dado por las ecuaciones 4.5 y 4.6 y modificando un poco la notación para simplificar la escritura, se tiene:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \end{aligned} \quad (4.17)$$

Donde \mathbf{u} es un vector de longitud p , \mathbf{x} es un vector de longitud n , \mathbf{y} es un vector de longitud q , \mathbf{A} es la matriz de transición de estados de $n \times n$, \mathbf{C} es una matriz de $q \times n$ y \mathbf{B} es una matriz de $n \times p$. \mathbf{w} y \mathbf{v} son procesos gaussianos independientes con media cero.

El problema del filtrado de Kalman se puede enunciar de la siguiente manera: *Dados los valores observados $\mathbf{y}_0, \dots, \mathbf{y}_k$, encontrar un estimado $\hat{\mathbf{x}}_{k+1|k}$ de \mathbf{x}_{k+1} el cual minimice la pérdida esperada.*

Del Teorema 2 se sabe que la solución a dicho problema consiste en encontrar la proyección ortogonal de \mathbf{x}_k en el espacio $\mathcal{Y}[k]$ generado por las variables aleatorias observadas.

Para este caso se separará el espacio $\mathcal{Y}[k]$ en dos subespacios. El primer subespacio será $\mathcal{Y}[k-1]$ el cual es el espacio generado por todos aquellos valores observados ya medidos, es decir, $\mathbf{y}_0, \dots, \mathbf{y}_{k-1}$ en el cual se presenta una componente de \mathbf{y}_k denotada por $\bar{\mathbf{y}}_{k|k-1}$.

Ahora sea $\tilde{\mathbf{y}}_{k|k-1}$ la componente de \mathbf{y}_k ortogonal a $\mathcal{Y}[k-1]$, la cual genera un subespacio vectorial perteneciente $\mathcal{Y}[k]$ denotado por $\mathcal{Z}[k]$. Ocurre de manera similar que cuando se trató las proyecciones ortogonales teniendo la variable aleatoria \mathbf{y}_k dos componentes, una componente es su proyección ortogonal sobre $\mathcal{Y}[k-1]$ y la otra es ortogonal a $\mathcal{Y}[k-1]$:

$$\mathbf{y}_k = \tilde{\mathbf{y}}_{k|k-1} + \bar{\mathbf{y}}_{k|k-1} \tag{4.18}$$

Si la componente ortogonal al espacio $\mathcal{Y}[k-1]$ es cero, $\tilde{\mathbf{y}}_{k|k-1}=0$, significa que entonces el espacio $\mathcal{Y}[k]$ es el mismo espacio que $\mathcal{Y}[k-1]$ y por lo tanto la medición de \mathbf{y}_k no conlleva ninguna información adicional. Por lo tanto el espacio $\mathcal{Y}[k-1]$ junto con el espacio $\mathcal{Z}[k]$ conforman al espacio $\mathcal{Y}[k]$.

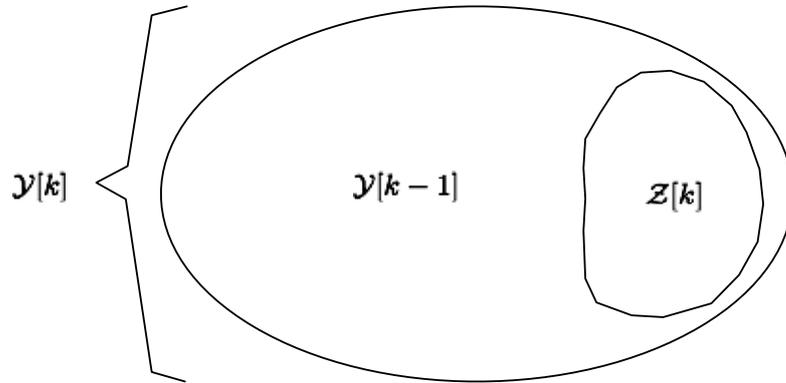


Figura 4.1: Representación gráfica del espacio $\mathcal{Y}[k]$ y los subespacios $\mathcal{Y}[k-1]$ y $\mathcal{Z}[k]$

Se sabe entonces que:

$$\hat{\mathbf{x}}_{k+1|k} = \hat{E}\{\mathbf{x}_{k+1}|\mathcal{Y}[k]\} \tag{4.19}$$

$$= \hat{E}\{\mathbf{x}_k|\mathcal{Y}[k-1]\} + \hat{E}\{\mathbf{x}_k|\mathcal{Z}[k]\} \tag{4.20}$$

Utilizando la ecuación 4.5 entonces se puede reescribir la ecuación anterior de la siguiente manera:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A}\hat{\mathbf{x}}_{k|k-1} + \hat{E}\{\mathbf{B}\mathbf{u}_{k-1}|\mathcal{Y}[k-1]\} + \hat{E}\{\mathbf{w}_k|\mathcal{Y}[k-1]\} + \hat{E}\{\mathbf{x}_k|\mathcal{Z}[k]\} \quad (4.21)$$

Sabiendo que el ruido del proceso tiene media cero y no se encuentra correlacionado con $\mathcal{Y}[k-1]$ además que \mathbf{u}_k se conoce precisamente y dado que el tercer término debe de ser generado por un operación lineal, debido a las restricciones, se debe de definir un operación lineal para la componente que genera al espacio $\mathcal{Z}[k]$ de la siguiente manera:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A}\hat{\mathbf{x}}_{k|k-1} + \mathbf{B}\mathbf{u}_k + \hat{E}\{\mathbf{x}_k|\mathcal{Z}[k]\} \quad (4.22)$$

$$= \mathbf{A}\hat{\mathbf{x}}_{k|k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{K}_k\tilde{\mathbf{y}}_{k|k-1} \quad (4.23)$$

En donde \mathbf{K}_k es una matriz de $n \times q$, esta matriz se conoce comúnmente como la *ganancia de Kalman*. A partir de 4.18 y sabiendo que $\tilde{\mathbf{y}}_{k|k-1}$ se encuentra en el espacio $\mathcal{Y}[k-1]$ y por lo tanto se puede representar mediante la ecuación 4.6 se obtiene:

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A}\hat{\mathbf{x}}_{k|k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{K}_k(\mathbf{y}_k - \bar{\mathbf{y}}_{k|k-1}) \quad (4.24)$$

$$= \mathbf{A}\hat{\mathbf{x}}_{k|k-1} + \mathbf{B}\mathbf{u}_k + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}) \quad (4.25)$$

Observando la ecuación 4.25 es fácil determinar que se trata de un algoritmo recursivo, pues para calcular el estimado óptimo $\hat{\mathbf{x}}_{k+1|k}$ es necesario el valor estimado pasado $\hat{\mathbf{x}}_{k|k-1}$. Para conocer el desempeño del filtro es necesario definir el error de estimación como la diferencia entre x_{k+1} y $\hat{\mathbf{x}}_{k+1|k}$, al observar la ecuación 4.12 entonces se observa que el error de estimación es la componente ortogonal de x_{k+1} al espacio $\mathcal{Y}[k]$.

$$\tilde{\mathbf{x}}_{k+1|k} = \mathbf{x}_{k+1} - \hat{\mathbf{x}}_{k+1|k} \quad (4.26)$$

Haciendo uso nuevamente de la ecuación 4.5 y de la ecuación 4.25 se puede escribir el error de estimación de la siguiente manera:

$$\tilde{\mathbf{x}}_{k+1|k} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k + \mathbf{w}_k - \mathbf{A}\hat{\mathbf{x}}_{k|k-1} - \mathbf{B}\mathbf{u}_k - \mathbf{K}_k\tilde{\mathbf{y}}_{k|k-1} \quad (4.27)$$

$$= \mathbf{A}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) - \mathbf{K}_k\tilde{\mathbf{y}}_{k|k-1} + \mathbf{w}_k \quad (4.28)$$

$$= \mathbf{A}\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k\tilde{\mathbf{y}}_{k|k-1} + \mathbf{w}_k \quad (4.29)$$

De esta manera se observa que el error de estimación calculado mediante la ecuación 4.29 es también una operación recursiva al necesitar del error de estimación pasado. Al definir la matriz de covarianza del error de estimación como:

$$\mathbf{P}_{k+1} = E \left\{ \tilde{\mathbf{x}}_{k+1|k} \tilde{\mathbf{x}}_{k+1|k}^* \right\} \quad (4.30)$$

Se puede entonces a partir de la ecuación 4.29 y de la definición anterior (4.30) encontrar una expresión recursiva para el calculo de la matriz de covarianza como sigue:

$$\mathbf{P}_{k+1} = E \left\{ \tilde{\mathbf{x}}_{k+1|k} \tilde{\mathbf{x}}_{k+1|k}^* \right\} \quad (4.31)$$

$$= E \left\{ (\mathbf{A}\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1} + \mathbf{w}_k) (\mathbf{A}\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1} + \mathbf{w}_k)^* \right\} \quad (4.32)$$

$$= E \left\{ (\mathbf{A}\tilde{\mathbf{x}}_{k|k-1} - \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1} + \mathbf{w}_k) (\tilde{\mathbf{x}}_{k|k-1}^* \mathbf{A}^* - \tilde{\mathbf{y}}_{k|k-1}^* \mathbf{K}_k^* + \mathbf{w}_k^*) \right\} \quad (4.33)$$

$$\begin{aligned} &= E \left\{ \mathbf{A}\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \mathbf{A}^* \right\} - E \left\{ \mathbf{A}\tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{y}}_{k|k-1}^* \mathbf{K}_k^* \right\} + E \left\{ \mathbf{A}\tilde{\mathbf{x}}_{k|k-1} \mathbf{w}_k^* \right\} \\ &\quad - E \left\{ \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \mathbf{A}^* \right\} + E \left\{ \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1} \tilde{\mathbf{y}}_{k|k-1}^* \mathbf{K}_k^* \right\} - E \left\{ \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1} \mathbf{w}_k^* \right\} \\ &\quad + E \left\{ \mathbf{w}_k \tilde{\mathbf{x}}_{k|k-1}^* \mathbf{A}^* \right\} - E \left\{ \mathbf{w}_k \tilde{\mathbf{y}}_{k|k-1}^* \mathbf{K}_k^* \right\} + E \left\{ \mathbf{w}_k \mathbf{w}_k^* \right\} \end{aligned} \quad (4.34)$$

Dado que el ruido \mathbf{w} no está correlacionada ni con los \mathbf{x} ni con $\tilde{\mathbf{y}}$ y al utilizar la expresión 4.6, la expresión anterior se reduce a la siguiente manera:

$$\begin{aligned} \mathbf{P}_{k+1} &= \mathbf{A}E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{A}^* + E \left\{ \mathbf{w}_k \mathbf{w}_k^* \right\} - \mathbf{A}E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} \mathbf{K}_k^* \\ &\quad - \mathbf{K}_k E \left\{ \tilde{\mathbf{y}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{A}^* + \mathbf{K}_k E \left\{ \tilde{\mathbf{y}}_{k|k-1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} \mathbf{K}_k^* \end{aligned} \quad (4.35)$$

$$\begin{aligned} &= \mathbf{A} \mathbf{P}_k \mathbf{A}^* + \mathbf{Q} - \mathbf{A}E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} \mathbf{K}_k^* - \mathbf{K}_k E \left\{ \tilde{\mathbf{y}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{A}^* \\ &\quad + \mathbf{K}_k \mathbf{C} E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{C}^* \mathbf{K}_k^* + \mathbf{K}_k E \left\{ \mathbf{v}_k \mathbf{v}_k^* \right\} \mathbf{K}_k^* \end{aligned} \quad (4.36)$$

$$\begin{aligned} &= \mathbf{A} \mathbf{P}_k \mathbf{A}^* + \mathbf{Q} - \mathbf{A}E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{C} \mathbf{K}_k^* - \mathbf{K}_k \mathbf{C} E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{A}^* \\ &\quad + \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* \mathbf{K}_k^* + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^* \end{aligned} \quad (4.37)$$

Para finalmente reducirse a

$$\mathbf{P}_{k+1} = \mathbf{A} \mathbf{P}_k \mathbf{A}^* + \mathbf{Q} - \mathbf{A} \mathbf{P}_k \mathbf{C} \mathbf{K}_k^* - \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{A}^* + \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* \mathbf{K}_k^* + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^* \quad (4.38)$$

Finalmente sólo resta definir la última \mathbf{K}_k que interviene en el cálculo del estimado óptimo (4.25) y en la covarianza del error (4.38). Dado que el error debido al espacio $Z[k]$ es la diferencia entre el valor \mathbf{x}_{k+1} y el estimado óptimo debido a $\mathcal{Z}[k]$, con ayuda de la expresión 4.20, se tiene:

$$\tilde{\mathbf{x}}_{k+1|\mathcal{Z}[k]} = \mathbf{x}_{k+1} - \hat{E} \{ \mathbf{x}_{k+1} | \mathcal{Z}[k] \} \quad (4.39)$$

Entonces el estimado óptimo debido a $\mathcal{Z}[k]$ es ortogonal la componente de \mathbf{y}_k ortogonal a $\mathcal{Y}[k-1]$, es decir $\tilde{\mathbf{y}}_{k|k-1}$. Por lo tanto:

$$E \left\{ (\mathbf{x}_{k+1} - \mathbf{K}_k \tilde{\mathbf{y}}_{k|k-1}) \tilde{\mathbf{y}}_{k|k-1}^* \right\} = 0 \quad (4.40)$$

$$E \left\{ \mathbf{x}_{k+1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} = \mathbf{K}_k E \left\{ \tilde{\mathbf{y}}_{k|k-1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} \quad (4.41)$$

$$E \left\{ \mathbf{x}_{k+1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.42)$$

Al utilizar la expresión 4.5 en la ecuación anterior, descomponer al valor actual \mathbf{x}_k en sus componentes y sabiendo que $\hat{\mathbf{x}}_{k|k-1}$ es ortogonal a $\mathcal{Z}[k]$, se tiene:

$$E \left\{ \mathbf{x}_{k+1} \tilde{\mathbf{y}}_{k|k-1}^* \right\} = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.43)$$

$$E \left\{ (\mathbf{A} \mathbf{x}_k + \mathbf{B} \mathbf{u}_k) \tilde{\mathbf{y}}_{k|k-1}^* \right\} = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.44)$$

$$E \left\{ (\mathbf{A}(\hat{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{x}}_{k|k-1}) \tilde{\mathbf{y}}_{k|k-1}^* + \mathbf{B} \mathbf{u}_k \tilde{\mathbf{y}}_{k|k-1}^* \right\} = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.45)$$

$$\mathbf{A} E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{C}^* + \mathbf{A} E \left\{ \tilde{\mathbf{x}}_{k|k-1} \mathbf{v}_k^* \right\} + \mathbf{B} E \left\{ \mathbf{u}_k \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{C}^* + \mathbf{B} E \left\{ \mathbf{u}_k \mathbf{v}_k \right\} = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.46)$$

Puesto que el ruido es independiente y la entrada \mathbf{u} no está correlacionada con $\tilde{\mathbf{x}}$, se reduce a

$$\mathbf{A} E \left\{ \tilde{\mathbf{x}}_{k|k-1} \tilde{\mathbf{x}}_{k|k-1}^* \right\} \mathbf{C}^* = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.47)$$

$$\mathbf{A} \mathbf{P}_k \mathbf{C}^* = \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{K}_k \mathbf{R} \quad (4.48)$$

Finalmente al factorizar y resolver para \mathbf{K}_k se llega a la siguiente ecuación:

$$\mathbf{K}_k = \mathbf{A} \mathbf{P}_k \mathbf{C}^* [\mathbf{C} \mathbf{P}_k \mathbf{C}^* + \mathbf{R}]^{-1} \quad (4.49)$$

Esta matriz tiene la tarea de minimizar el error, mediante la corrección del estimador [Welch and Bishop, 2006, p. 3].

4.4. Algoritmo para el filtrado de Kalman

El filtro de Kalman estima un proceso mediante el uso de retroalimentación: el filtro estima los estados del proceso en algún tiempo dado y después obtiene retro-

alimentación en la forma de mediciones ruidosas. De esta manera se pueden dividir las ecuaciones del filtro de Kalman en dos grupos que se alternan indefinidamente. Para ello se utiliza información obtenida antes de comenzar el ciclo que se le conoce como información *a priori* y otra que se obtiene después de iniciar el ciclo, a esta información se le conoce como información *a posteriori* [Welch and Bishop, 2006]. Estos dos grupos son:

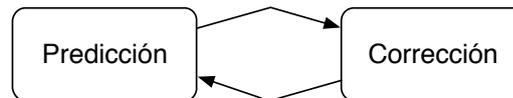


Figura 4.2: El ciclo del filtrado de Kalman se dividen dos grupos y se repite

Predicción: Las ecuaciones del grupo de predicción son responsables de hacer una proyección en el tiempo del estado actual y del estimado de matriz de covarianza del error para obtener los estimados a priori del siguiente ciclo. Estas ecuaciones pueden ser vistas como las ecuaciones de un predictor.

Corrección: Las ecuaciones de corrección son responsables de la retroalimentación, al incorporar una nueva medición al estimado a priori calculado por las ecuaciones de predicción para obtener una versión mejorada que es el estimado *a posteriori*. Las ecuaciones pueden ser vista como ecuaciones para corregir y actualizar los estimados.

Para obtener las ecuaciones correspondientes de cada grupo, es necesario suponer que no se tiene disponibilidad de la medición actual, es decir \mathbf{y}_k . Sabiendo entonces que el grupo de predicción se encarga de dar una estimación a priori del estado y de la matriz de covarianza del error. Se utilizan las ecuaciones 4.25 y 4.38 pero es necesario adecuarlas recordando que la medición actual no se encuentra disponible.

4.4.1. Predicción

Las ecuaciones para la predicción son las siguientes:

- Para la estimación del estado a priori, denotado por $\hat{\mathbf{x}}_{k+1|k}^-$, se utiliza la ecuación 4.25 pero se elimina el término que involucra la variable \mathbf{K}_k junto con la medición del estado actual \mathbf{y}_k , la cual se supone no está disponible en este momento.

$$\hat{\mathbf{x}}_{k+1|k}^- = \mathbf{A}\hat{\mathbf{x}}_{k|k-1} + \mathbf{B}\mathbf{u}_k \quad (4.50)$$

- El estimado a priori de la matriz de covarianza del error se calcula mediante la expresión 4.38, pero eliminando todos los términos que involucran a la matriz \mathbf{K}_k .

$$\mathbf{P}_{k+1}^- = \mathbf{A}\mathbf{P}_k\mathbf{A}^* + \mathbf{Q} \quad (4.51)$$

Hasta este momento se considera la disponibilidad del estado estimado pasado $\hat{\mathbf{x}}_{k|k-1}$, la entrada actual \mathbf{u}_k y la matriz de covarianza del error pasada \mathbf{P}_k . Con estos datos además de las matrices que describen al sistema y la covarianza del ruido de los estados \mathbf{Q} .

4.4.2. Corrección

Una vez que se han calculado los estimados, es decir los estados a priori y la matriz de covarianza del error a priori, se continua con la tarea de corregir las posibles imprecisiones, asumiendo que se encuentra disponible a partir de ahora el valor actual de medición \mathbf{y}_k . De igual manera se utilizan las expresiones 4.25 y 4.38, pero se observa que es necesario calcular antes la matriz \mathbf{K}_k .

Las ecuaciones para la corrección son las siguientes:

- El primer paso es calcular la matriz \mathbf{K}_k mediante la ecuación 4.49, al observar esta expresión se nota que es necesario contar con la matriz \mathbf{P}_k , pero ya que no se ha calculado aún y solo se cuenta con el estimado a priori \mathbf{P}_{k+1}^- , calculado en 4.51, se utiliza este dato.

$$\mathbf{K}_k = \mathbf{A}\mathbf{P}_k^- \mathbf{C}^* [\mathbf{C}\mathbf{P}_k^- \mathbf{C}^* + \mathbf{R}]^{-1} \quad (4.52)$$

- Ya que se cuenta con la matriz \mathbf{K}_k es posible entonces hacer la corrección del estimado a priori al utilizar el valor medido actual \mathbf{y}_k en la ecuación 4.25. Se puede observar que los primeros dos términos son el estimado a priori $\hat{\mathbf{x}}_{k+1|k}^-$, por lo que la ecuación se expresa de la siguiente manera:

$$\hat{\mathbf{x}}_{k+1|k} = \hat{\mathbf{x}}_{k+1|k}^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}) \quad (4.53)$$

- Finalmente sólo resta corregir el estimado a priori de la matriz de covarianza del error al calcular los términos que no se consideraron previamente.

$$\mathbf{P}_{k+1} = \mathbf{P}_{k+1}^- - \mathbf{A}\mathbf{P}_k\mathbf{C}\mathbf{K}_k^* - \mathbf{K}_k\mathbf{C}\mathbf{P}_k\mathbf{A}^* + \mathbf{K}_k\mathbf{C}\mathbf{P}_k\mathbf{C}^*\mathbf{K}_k^* + \mathbf{K}_k\mathbf{R}\mathbf{K}_k^* \quad (4.54)$$

Teniendo completamente descritos los dos grupos de ecuaciones se puede observar claramente la naturaleza recursiva del filtro de Kalman. Esto significa que sólo necesita el estimado pasado $\hat{\mathbf{x}}_{k|k-1}$, la medición actual \mathbf{y}_k y la matriz de covarianza del error pasado \mathbf{P}_k para funcionar, en contraste con otras técnicas de estimación que necesitan un historial de observaciones o estimados para funcionar. Para el caso de la primera iteración será necesario proporcionar las condiciones iniciales, que son los parámetros necesarios para comenzar con el cálculo recursivo que realiza el filtro.

El filtro de Kalman podría implementarse directamente mediante las ecuaciones 4.25, 4.38 y 4.49, pero se divide para distinguir las dos fases de predicción y corrección.

La fase predicción se conoce como estimado a priori debido a la carencia del último dato medido, que matemáticamente se representa mediante el espacio de información $\mathcal{Y}[k-1]$. Mientras que la fase de corrección se conoce como estimado a posteriori ya que la medición del último dato marca un punto de quiebre al disponer de nueva información, la cual se representa matemáticamente mediante el espacio $\mathcal{Z}[k]$. Algunos otros autores como [Haykin, 1991] llaman a esta nueva información *innovación*.

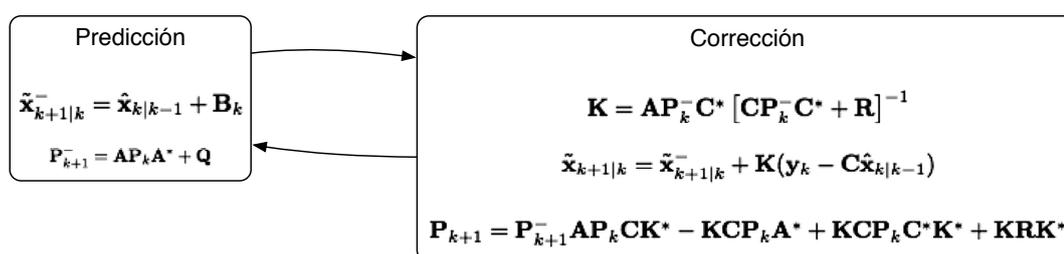


Figura 4.3: Ciclo completo del filtrado de Kalman

Normalmente estas dos fases alternan su funcionamiento produciendo un estimado del siguiente dato y corrigiéndolo. Pero esto no siempre es necesario, si por alguna razón no se encuentra disponible en algún instante de tiempo la medición del

sistema, el proceso de corrección puede ser ignorado o saltado y producirse múltiples fases de predicción.

Por otra parte si se presentan múltiples mediciones en un mismo instante, se puede realizar varias veces la fase de corrección para actualizar los datos con toda la información disponible de estas mediciones, normalmente utilizando diferentes matrices de observación \mathbf{C} [Wikipedia, 2014b].

4.4.3. Parámetros del Filtro de Kalman

La matriz de covarianza del ruido de medición \mathbf{R} normalmente es medida antes de la operación del filtro. Mientras que la determinación de la matriz de covarianza del ruido de los estados \mathbf{Q} generalmente es más difícil de obtener debido a que no se tiene acceso a la observación directa de los estados que se están estimando.

En cualquier caso, se tenga o no disponibilidad de estos parámetros necesarios para la operación del filtro, usualmente un desempeño superior del filtro, en un enfoque estadístico, se puede obtener al variar los parámetros \mathbf{Q} y \mathbf{R} . Esta variación a veces suele llamarse sintonización, este proceso normalmente se realiza antes de que funcione el filtro, a veces ayudado de otro filtro de Kalman distinto que realiza una tarea conocida como *identificación de sistema* [Welch and Bishop, 2006].

Se puede observar que bajo ciertas condiciones \mathbf{Q} y \mathbf{R} son constantes, por lo tanto la matriz de covarianza del error \mathbf{P}_k y la matriz \mathbf{K}_k también conocida como *ganancia de Kalman* se estabilizarán y permanecerán constantes también. En este caso se pueden calcular de manera previa estos parámetros modificar la velocidad de convergencia del filtro.

Si al analizar la expresión 4.52, no tomando en cuenta momentáneamente a la matriz \mathbf{A} , se observa que a medida que la matriz de covarianza del ruido de medición \mathbf{R} se aproxima a cero, la ganancia de Kalman tiende al inverso de la matriz de medición.

$$\lim_{\mathbf{R} \rightarrow 0} \mathbf{K}_k = \mathbf{C}^{-1} \quad (4.55)$$

Si ahora con este resultado se analiza la ecuación 4.53 se observa que la ganancia \mathbf{K}_k contrarresta de cierta manera el efecto del estimado pasado $\hat{\mathbf{x}}_{k|k-1}$ sobre la corrección que se realiza. Esto se puede pensar como que la medición actual \mathbf{y}_k es más confiable a medida que \mathbf{R} se aproxima a cero, mientras que el dato arrojado en

la predicción se vuelve menos confiable.

En otro análisis, si la matriz de covarianza del error a priori \mathbf{P}_k^- se aproxima a cero, la ganancia \mathbf{K}_k tiende a ser cero también.

$$\lim_{\mathbf{P} \rightarrow 0} \mathbf{K}_k = \mathbf{0} \quad (4.56)$$

Si ahora se analiza la ecuación 4.53 se observa que la ganancia \mathbf{K}_k contrarresta tanto el efecto de la medición actual \mathbf{y}_k como del estimado pasado, para prácticamente sólo tomarse en cuenta el valor predicho $\tilde{\mathbf{x}}_{k+1|k}^-$, sin corrección alguna. Entendiendo de esta manera que la predicción que da el filtro es muy buena ya que el error es prácticamente nulo, y no son necesarias las mediciones observadas [Welch and Bishop, 2006].

4.5. Problema planteado

Considérese el mismo problema planteado en la sección 3.6. Un proceso autoregresivo $\mathbf{d}[n]$, que es transmitido a través de una línea de transmisión. La salida $\mathbf{u}[n]$ es corrompida tanto por el proceso de transmisión como por ruido gaussiano aditivo que se presenta en la medición a la salida del canal.

4.5.1. Caracterización del problema

El proceso autoregresivo es generado mediante un sistema autoregresivo, que tiene a la entrada un ruido gaussiano \mathbf{v}_1 de media cero y varianza $\sigma_1^2 = 0.27$, el cual está descrito mediante la misma función de transferencia dada en 3.47, la cual es la siguiente:

$$H_1[z] = \frac{1}{1 + 0.8458z^{-1}} \quad (4.57)$$

El canal de transmisión se modela mediante la siguiente función de transferencia:

$$H_2[z] = \frac{1}{1 - 0.9458z^{-1}} \quad (4.58)$$

Finalmente la medición a la salida del canal de transmisión se ve contaminada por un ruido aditivo $\mathbf{v}_2[n]$, en específico un ruido gaussiano con varianza $\sigma_2^2 = 3$.

Para caracterizar y modelar la problemática en un sistema digital, se procede a representar las funciones de transferencia mediante un modelo en variables de estado equivalente. Existen muchas formas para llevar a cabo representaciones en el espacio de estados para sistemas en tiempo discreto descritos por funciones de transferencia, el método utilizado será expresar la función de transferencia en la forma canónica controlable según se describe en C.2, quedando de la siguiente manera:

1. La ecuación de transferencia que representa al modelo autoregresivo (4.57), puede representarse mediante un modelo en variables de estado que tiene como entrada a la señal v_1 y a la salida es la representación del proceso autoregresivo que es la señal deseada $\mathbf{d}[n]$. Su representación mediante un modelo en variables de estado es la siguiente:

$$x[k+1] = \mathbf{A}x[k] + \mathbf{B}v_1[k] + w[k] \quad (4.59)$$

$$d[k] = \mathbf{C}x[k] + \mathbf{D}v_1[k] + v[k] \quad (4.60)$$

en donde:

$$\mathbf{A} = [-0.8458] \quad (4.61)$$

$$\mathbf{B} = [1] \quad (4.62)$$

$$\mathbf{C} = [-0.8458] \quad (4.63)$$

$$\mathbf{D} = [1] \quad (4.64)$$

2. Mientras tanto el canal se modela mediante el sistema descrito por la función de transferencia 4.58, que tiene a la entrada la señal deseada $\mathbf{d}[n]$, que es transmitida, y a la salida la señal $\mathbf{x}[n]$. El modelo del canal no es necesario expresarlo en variables de estado, ya que su efecto se considerará como ruido a ser eliminado. Entonces la ecuación que relaciona la salida en el instante $x[n]$ con la entrada $d[n]$ está dada por la siguiente ecuación en diferencias:

$$x[n] = d[n] + 0.9458x[n-1] \quad (4.65)$$

3. Finalmente el ruido aditivo $\mathbf{v}_2[n]$ se suma a la señal de salida $\mathbf{x}[n]$, lo que da como resultado la señal medida a la salida del canal de transmisión, denotada por $\mathbf{u}[n]$.

$$\mathbf{u}[n] = \mathbf{x}[n] + \mathbf{v}_2[n] \quad (4.66)$$

4.5.2. Resolución del problema

El objetivo es diseñar un filtro lineal óptimo, que se colocará a la salida de la línea de transmisión con el fin de minimizar la distorsión debida al canal de transmisión y la interferencia del ruido aditivo en la medición, para obtener un estimado de la señal deseada $\hat{\mathbf{d}}[n]$ lo más próxima a la señal deseada $\mathbf{d}[n]$.

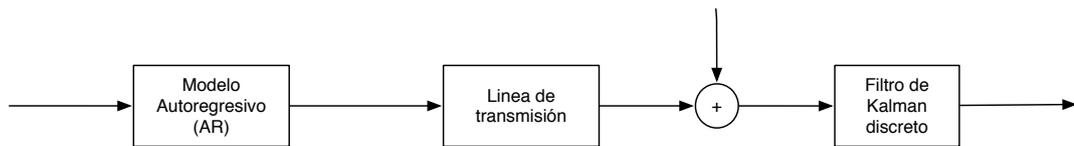


Figura 4.4: Diagrama a bloques de aplicación del filtro de Kalman

Parámetros del Filtro

Para aplicar el filtro de Kalman los efectos de ruido causados por la distorsión del canal y el ruido de medición se consideran como el ruido $v[k]$ en la ecuación 4.60. Teóricamente el ruido tiene una varianza igual a la de $v_2[n]$, $\sigma^2 = 3$, pero para considerar la matriz \mathbf{R} se puede dar un mayor grado considerando tal vez al ruido $v[k]$ con una varianza $\sigma_v^2 = 5$.

$$\mathbf{R} = [5] \quad (4.67)$$

El término $w[k]$ podría seleccionarse observando la dependencia de los estados internos, al ser un modelo autoregresivo se sabe que depende sólo de la entrada por lo cual se podría considerar que no tienen ruido los estados, esto ayudaría a la convergencia del filtro. Por lo tanto:

$$\mathbf{Q} = [0] \quad (4.68)$$

Condiciones iniciales

Finalmente al saber que es un proceso aleatorio estacionario con media cero, se puede elegir este valor como condición inicial para los estados $x_0[0]$. Para la matriz

de covarianza del error se puede escoger algún otro valor mayor que cero, como por ejemplo $P = 1$.

4.5.3. Simulaciones

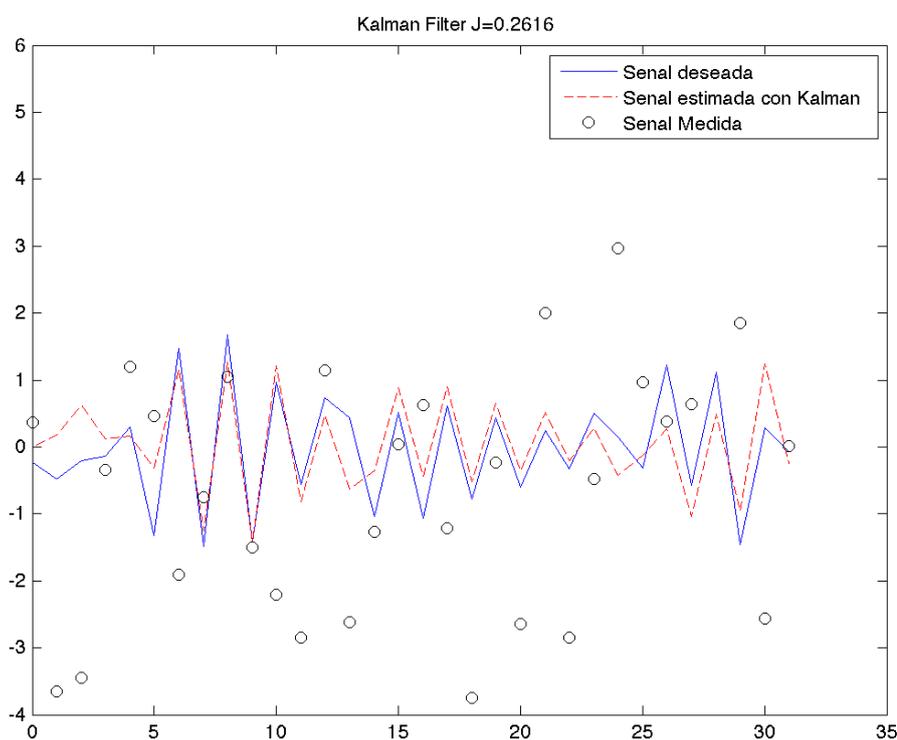


Figura 4.5: Simulación del filtro de Kalman aplicado al problema en la sección 4.5

En la imagen 4.5 se muestran los resultados obtenidos al implementar el filtro de Kalman para el problema anterior. Además, en la gráfica 4.6 se puede observar como se minimiza la varianza del error, alcanzando un valor de prácticamente cero después de 10 iteraciones, la cual se obtuvo de acuerdo con la ecuación 4.4, conforme pasan las iteraciones.

Al comparar el desempeño del filtro de Kalman con un filtro de Wiener de orden 10 (figura 4.7), operando bajo las mismas condiciones, se puede observar que el

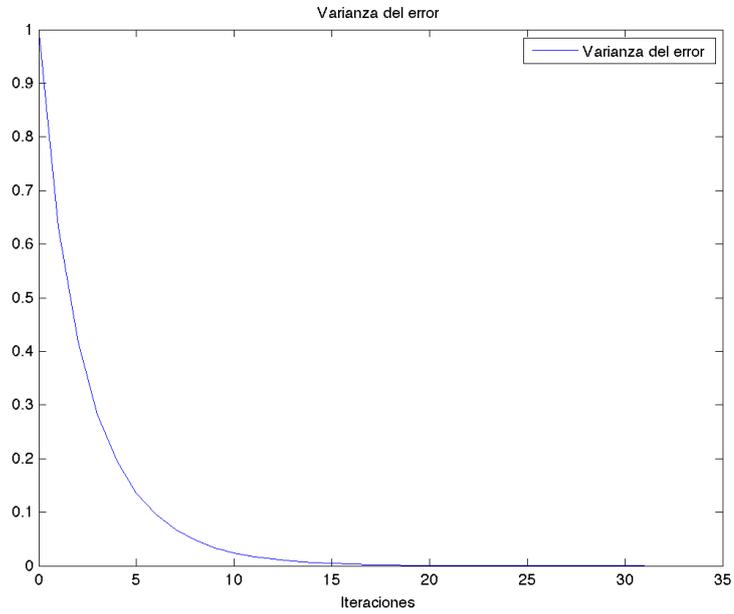


Figura 4.6: Convergencia de la varianza del error

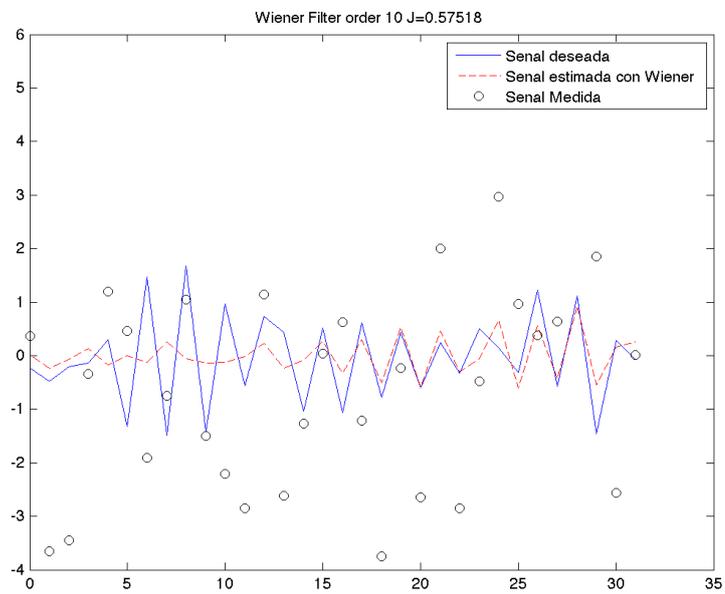


Figura 4.7: Simulación del filtro de Wiener de orden 10

4. FILTRO DE KALMAN

desempeño a simple vista parece ser muy similar. Pero al comparar el error cuadrático medio que se obtiene con cada uno de los filtros, se observa que el desempeño fue mejor con el filtro de Kalman en un sentido probabilístico, al tener un error cuadrático medio menor que el obtenido a partir del filtro de Wiener.

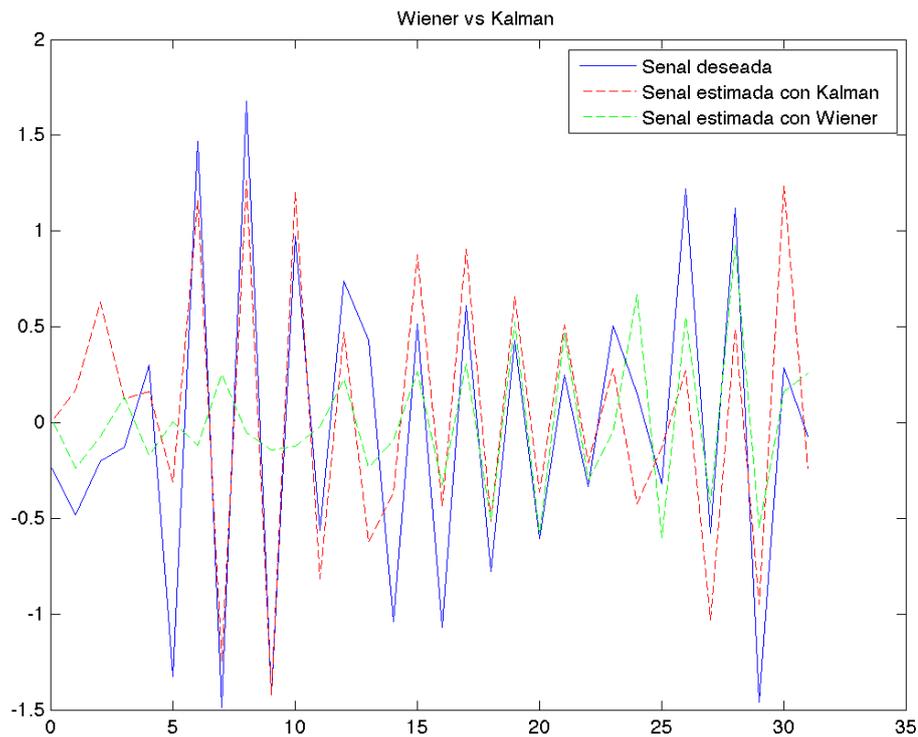


Figura 4.8: Comparación de las señales obtenidas por Kalman y Wiener junto con la señal deseada

Seguimiento de trayectorias

5.1. Planteamiento del problema

Normalmente los robots están diseñados para desempeñar distintas tareas programadas por los seres humanos, en específico un robot móvil debe de realizar distintas instrucciones en lugares diferentes, por lo tanto es necesario que se desplace a través del medio en el que se encuentra. Estos desplazamientos frecuentemente presentan obstáculos que dificultan el movimiento del robot, los cuales por sentido común deben de ser evitados, por lo que es necesario realizar la planificación de una ruta para evadir los obstáculos conocidos o encontrar una manera precisa de localizarlos para así esquivarlos.

Para poder realizar todas estas tareas y demás, es indispensable contar con la posición del robot de manera precisa, en un marco de referencia. Esto se logra mediante el uso de sensores que brindan información acerca de la posición o de la dinámica del robot, tal como orientación, velocidad o aceleración.

Pero recordando lo que menciona [Maybeck, 1979], las mediciones obtenidas por los sensores con los que cuentan los robots no son completamente certeras, es decir se encuentran corrompidas por ruido, el cual de acuerdo con [Hendricks et al., 2008, p. 392] puede propagarse a través del sistema, haciendo que se presenten imprecisiones en el desempeño del robot.

Para resolver entonces el problema de la posición de un robot en un marco de referencia, usualmente en un espacio de dos dimensiones ($2D$), se utilizan una gran variedad de sensores que brindan información de distintos tipos, con distintas características. De esta manera se puede utilizar toda la información disponible obtenida por los sensores para intentar mejorar de manera conjunta la precisión de

las mediciones. Frecuentemente esto se logra mediante la fusión de todos los datos mediante algún algoritmo de procesamiento.

Comúnmente, el uso de un filtro de Kalman es utilizado para realizar esta tarea, como se puede comprobar en los siguientes artículos [Mirats Tur, 2007], [Nepali et al., 2014], [Marín et al., 2013], [Mishra et al., 2014], [Nitulescu, 2005], ya que se utiliza toda la información disponible para mejorar la precisión de las mediciones y de esta manera dar un estimado más confiable de la posición del móvil. Siendo posible utilizar distintos sensores tales como GPS (*Global Positioning System*), cámaras, LRF (*Laser Range Finder*), sensores ultrasónicos, etc.

5.2. Robot diferencial

En este trabajo se propone utilizar un filtro de Kalman para dar un estimado de la posición de un robot diferencial. Un robot diferencial, de acuerdo con [Nitulescu, 2005], es aquel robot el cual cuenta con al menos tres puntos de apoyo, es decir tres ruedas, de las cuales dos de ellas pueden controlarse para hacer que el robot se mueva hacia adelante, atrás ó para cambiar de orientación.

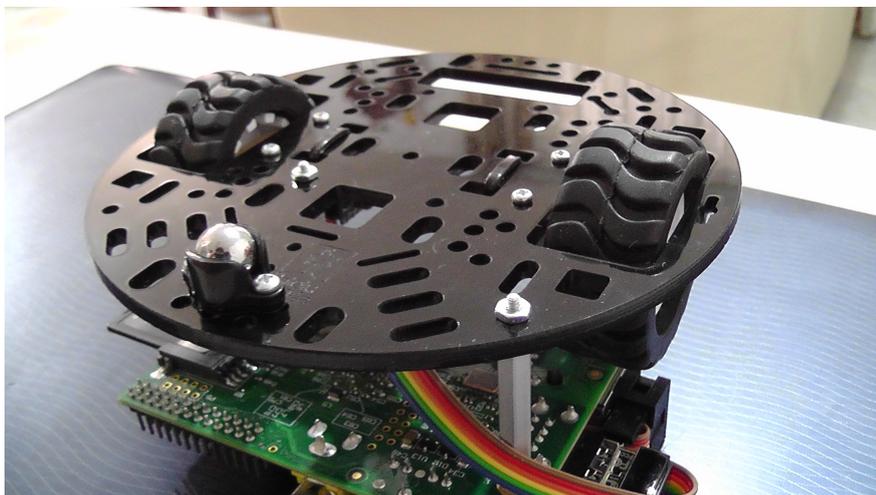


Figura 5.1: Ilustración de un robot diferencial con tres puntos de apoyo, dos de los cuales son las ruedas móviles

Normalmente, y dependiendo de la construcción, la tercera rueda se encuentra

ubicada al frente ó en la parte posterior de robot. De esta manera brinda el apoyo para que las dos ruedas móviles puedan hacer que se realicen sólo dos tipos de desplazamientos, lineales o rotacionales, por lo que se dice que este tipo de robots cuentan con grados de libertad.

Para este tipo de robots en particular, es muy importante tener en cuenta que los movimientos realizados sean suaves y no muy bruscos, esto para evitar que se presente deslizamiento en las ruedas ó sacudidas en la estructura.

5.2.1. Modelo dinámico

El modelo dinámico se obtiene a partir del conocimiento de que el robot tiene dos grados de libertad, es decir, puede realizar dos tipos de desplazamientos: rotacional y lineal. Este tipo de desplazamientos están en función de la velocidad de giro de las llantas, siendo respectivamente las siguientes ecuaciones:

$$\omega = \frac{r(\omega_R - \omega_L)}{D} \quad (5.1)$$

$$V = \frac{r(\omega_R + \omega_L)}{2} \quad (5.2)$$

En donde r representa el radio de las ruedas, D la distancia entre las ruedas y ω_R , ω_L representan las velocidades angulares de la rueda derecha e izquierda respectivamente. De esta manera, contando con las ecuaciones 5.1 y 5.2, se puede describir la dinámica del robot mediante el siguiente conjunto de ecuaciones:

$$\dot{x} = \frac{r(\omega_R + \omega_L)}{2} \cos(\theta) \quad (5.3)$$

$$\dot{y} = \frac{r(\omega_R + \omega_L)}{2} \sin(\theta) \quad (5.4)$$

$$\dot{\theta} = \frac{r(\omega_R - \omega_L)}{D} \quad (5.5)$$

5.2.2. Modelo discreto

Siendo x , y y θ las coordenadas en el eje x , en el eje y y la orientación del robot respectivamente. Estas tres variables son suficientes para determinar la dinámica del robot. Ahora suponiendo que se desea implementar el algoritmo en un dispositivo digital, como puede ser una computadora, un microcontrolador, o inclusive un DSP, es necesario contar con el modelo del sistema en tiempo discreto, es decir, que

5. SEGUIMIENTO DE TRAYECTORIAS

se encuentre expresado mediante ecuaciones en diferencias en lugar de ecuaciones diferenciales. Para ello se procede a hacer un proceso llamado *discretización aproximada*, mediante el cual se aproxima la derivada de las ecuaciones de la siguiente manera:

Sabiendo que por definición la derivada de una función es la siguiente:

$$\dot{x} = \lim_{\Delta t \rightarrow 0} \frac{f(x + \Delta t) - f(x)}{\Delta t} = \frac{r(\omega_R + \omega_L)}{2} \cos(\theta) \quad (5.6)$$

Suponiendo que se cuenta con un muestreo de la señal a periodos igual a T_s entonces se puede decir que, el periodo de muestreo $T_s = \Delta t$ por lo tanto, si el periodo de muestreo es lo suficientemente pequeño, es decir $T_s \approx \Delta t$, entonces:

$$\dot{x} \approx \frac{x[t + T_s] - x[t]}{T_s} = \quad (5.7)$$

En donde, $x[k + 1]$ representa el valor en el siguiente instante de muestreo y $x[k]$ es el valor en el instante actual. Igualando la ecuación anterior con la definición de \dot{x} se tiene

$$\frac{x[k + 1] - x[k]}{T_s} \approx \frac{r(\omega_R + \omega_L)}{2} \cos(\theta[k]) \quad (5.8)$$

$$x[k + 1] - x[k] \approx T_s \frac{r(\omega_R + \omega_L)}{2} \cos(\theta[k]) \quad (5.9)$$

$$x[k + 1] \approx x[k] + T_s \frac{r(\omega_R + \omega_L)}{2} \cos(\theta[k]) \quad (5.10)$$

Siguiendo el mismo procedimiento, para las ecuaciones representadas en 5.5, se obtiene el modelo discreto del robot diferencial.

$$x[k + 1] \approx x[k] + T_s \frac{r(\omega_R + \omega_L)}{2} \cos(\theta[k]) \quad (5.11)$$

$$y[k + 1] \approx y[k] + T_s \frac{r(\omega_R + \omega_L)}{2} \sin(\theta[k]) \quad (5.12)$$

$$\theta[k + 1] \approx \theta[k] + T_s \frac{r(\omega_R - \omega_L)}{D} \quad (5.13)$$

Ahora, sabiendo que la velocidad angular de las ruedas derecha e izquierda son ω_R y ω_L , se puede utilizar la expresión 5.14 para dejar en términos de la velocidad tangencial de las ruedas.

$$v = r \cdot \omega \quad (5.14)$$

Si además, se utiliza la expresión 5.16 y se sustituye 5.14.

$$v = \frac{d}{t} \quad (5.15)$$

$$d = vt \quad (5.16)$$

Se conocerá la distancia recorrida por las ruedas durante un periodo de muestreo T_s , como lo muestran las expresiones 5.17, en donde ΔR y ΔL son las distancias recorridas por la rueda derecha e izquierda respectivamente.

$$\begin{aligned} \Delta R &= T_s r \cdot \omega_R \\ \Delta L &= T_s r \cdot \omega_L \end{aligned} \quad (5.17)$$

Por lo tanto, al utilizar la expresión anterior se pueden simplificar las ecuaciones del modelo discreto del robot diferencial de la siguiente manera:

$$x[k+1] \approx x[k] + \frac{\cos(\theta[k])}{2} (\Delta R + \Delta L) \quad (5.18)$$

$$y[k+1] \approx y[k] + \frac{\sin(\theta[k])}{2} (\Delta R + \Delta L) \quad (5.19)$$

$$\theta[k+1] \approx \theta[k] + \frac{1}{D} (\Delta R - \Delta L) \quad (5.20)$$

Dichas ecuaciones simplificadas, se pueden expresar en un modelo en variables de estado, como se expresa a continuación:

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \theta_{k+1} \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} \frac{\cos(\theta_k)}{2} & \frac{\cos(\theta_k)}{2} \\ \frac{\sin(\theta_k)}{2} & \frac{\sin(\theta_k)}{2} \\ \frac{1}{D} & -\frac{1}{D} \end{bmatrix} \begin{bmatrix} \Delta R \\ \Delta L \end{bmatrix} \quad (5.21)$$

Al observar el modelo en variables de estado discreto (5.21), se constata que la ecuación del proceso cuenta con operaciones lineales, pero que es variable respecto del tiempo. Para el caso de la ecuación de medición, la cual supone se elija de tal manera que la salida sean cantidades que se pueden medir físicamente, se define de la siguiente manera:

$$\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} + \mathbf{v}_2 \quad (5.22)$$

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} + \mathbf{v}_2 \quad (5.23)$$

La cual será medida a través de un GPS, en donde v_2 es el ruido aditivo debido a las mediciones y el cual se considera Gaussiano. De esta manera se pueden agregar distintos sensores, por ejemplo una brújula para medir la orientación del robot, los cuales se utilizarían mediante otra matriz de medición.

5.3. Planificación de trayectoria

Como ya se ha mencionado, algunas de las tareas relacionadas con los robots móviles es el llegar a un punto en el espacio siguiendo una ruta planificada. En este caso se establece un punto en el plano XY al cual se desea que llegue el robot, desde un punto inicial, describiendo una trayectoria que se supone ideal, la cual de acuerdo con [Hendricks et al., 2008, p. 13] es la ruta que describe el vector de estados que describe el comportamiento del sistema.

5.3.1. Trayectoria ideal

Esta trayectoria se supone ideal pues es la trayectoria que seguiría el móvil al tener implementado un controlador de la forma 5.24 y no presentarse ruido en las mediciones, es decir que las mediciones de la posición fuesen totalmente precisas. En donde K_θ es el error angular que existe entre la orientación actual del robot y el punto deseado y r representa la distancia que hay entre la posición actual y el punto deseado, es decir el módulo del error de posición del móvil.

$$\begin{aligned}v_L &= -K_\theta e_\theta + K_p r \left(1 - \frac{|\text{sat}(e_\theta)|}{K_s}\right) \\v_R &= K_\theta e_\theta + K_p r \left(1 - \frac{|\text{sat}(e_\theta)|}{K_s}\right)\end{aligned}\tag{5.24}$$

Los coeficientes K_θ , K_p y K_s determinan el comportamiento del controlador. En específico K_θ es la constante de proporcionalidad del error angular, K_p es la constante de proporcionalidad del error de posición y K_s es la constante de saturación del error angular. Los valores utilizados para estas constantes se pueden observar en la siguiente tabla.

$$\begin{aligned}K_\theta &= 10 \\K_p &= 20 \\K_s &= \frac{3\pi}{4}\end{aligned}$$

Finalmente, la trayectoria ideal, descrita por el móvil e implementando el controlador 5.24 y las constantes antes mencionadas, se puede observar en la figura 5.2.

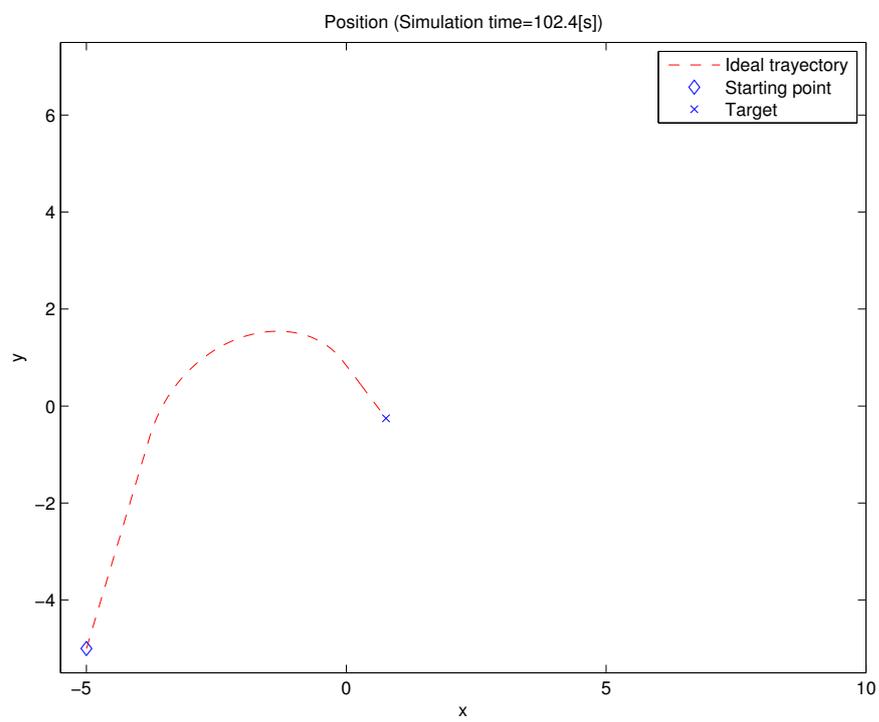


Figura 5.2: Trayectoria ideal descrita por un robot móvil desde un punto inicial hasta el punto objetivo

5.3.2. Trayectoria con ruido

La trayectoria ruidosa, es decir la trayectoria que sigue el robot según los sensores los cuales, como ya se ha hablado antes, introducen ruido a las mediciones. En el caso de las las mediciones obtenidas por el GPS y de acuerdo con [Mishra et al., 2014] la posición horizontal y vertical tienen una precisión de 3 y 5 metros respectivamente en el 95 % de los casos.

Asumiendo que las mediciones tienen una distribución Gaussiana, y recordando que 3 veces la desviación estándar de este tipo de distribuciones equivalen al 99.9% de la probabilidad. Se puede deducir la varianza de las mediciones a partir de la afirmación de [Mishra et al., 2014].

$$\sigma_{GPSx}^2 = \left(\frac{3}{3}\right)^2 = 1[\text{m}^2] \quad (5.25)$$

$$\sigma_{GPSy}^2 = \left(\frac{5}{3}\right)^2 = \frac{25}{9}[\text{m}^2] \quad (5.26)$$

Asumiendo además de que las mediciones en ambos sentidos no están correlacionadas, se obtiene la matriz de covarianza de las mediciones (5.27), y la función de densidad se puede observar en la gráfica de la figura 5.3, la cual muestra que la probabilidad es mayor cerca de la posición en la que se encuentra el móvil.

$$\mathbf{R} = \begin{bmatrix} \sigma_{GPSx}^2 & 0 \\ 0 & \sigma_{GPSy}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{25}{9} \end{bmatrix} \quad (5.27)$$

5.4. Estimación óptima de trayectoria

Para realizar la estimación de la trayectoria, se utiliza un filtro de Kalman mediante el cual se puede *fusionar* la información de dos sensores o más, para mejorar la precisión de las mediciones y con ello poder dar un estimado con menos ruido de la posición del móvil.

Recordando el modelo del robot diferencial (5.28), el cual cuenta con una matriz que es variante con el tiempo, se debe de tener una consideración práctica para la implementación del filtro de Kalman.

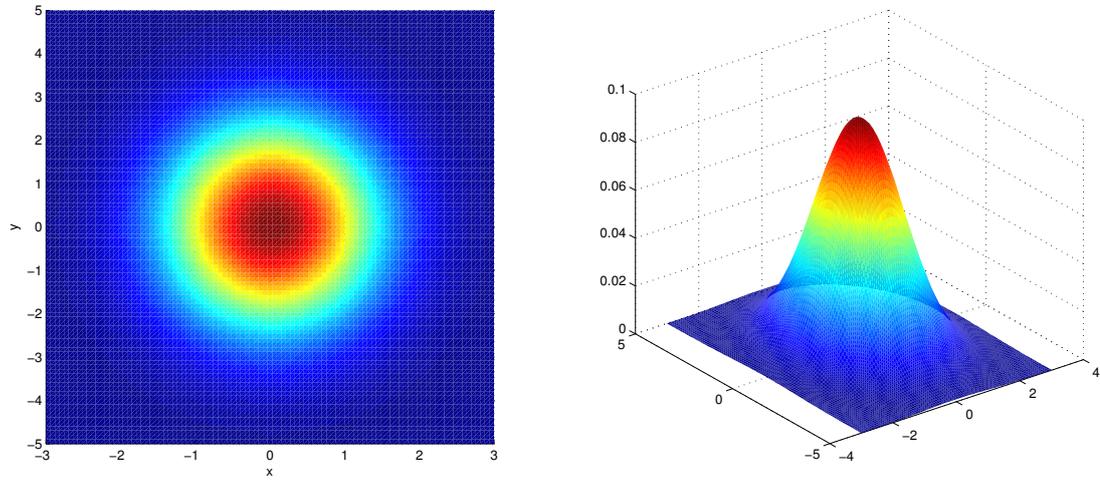


Figura 5.3: Función de densidad de las mediciones obtenidas de un GPS con media $\mu = 0$ y una matriz de covarianza \mathbf{R} .

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} \frac{\cos(\theta_k)}{2} & \frac{\cos(\theta_k)}{2} \\ \frac{\sin(\theta_k)}{2} & \frac{\sin(\theta_k)}{2} \\ \frac{1}{D} & -\frac{1}{D} \end{bmatrix} \begin{bmatrix} \Delta R \\ \Delta L \end{bmatrix} + \mathbf{v}_1 \quad (5.28)$$

Esta matriz variable (5.29), como se puede observar en 5.28, es la matriz \mathbf{B} . Por lo tanto, para realizar la implementación es necesario actualizar los elementos de la matriz \mathbf{B} en cada iteración del algoritmo antes de utilizarla, como se muestra en el diagrama de la figura 5.4.

$$\mathbf{B} = \begin{bmatrix} \frac{\cos(\theta_k)}{2} & \frac{\cos(\theta_k)}{2} \\ \frac{\sin(\theta_k)}{2} & \frac{\sin(\theta_k)}{2} \\ \frac{1}{D} & -\frac{1}{D} \end{bmatrix} \quad (5.29)$$

Para obtener la trayectoria estimada, el procedimiento consta de dos etapas principales, la etapa de predicción y la etapa de corrección. Estas etapas se dividen a su vez en una serie de tareas más simples, que se observan de mejor manera en el diagrama de la figura 5.4.

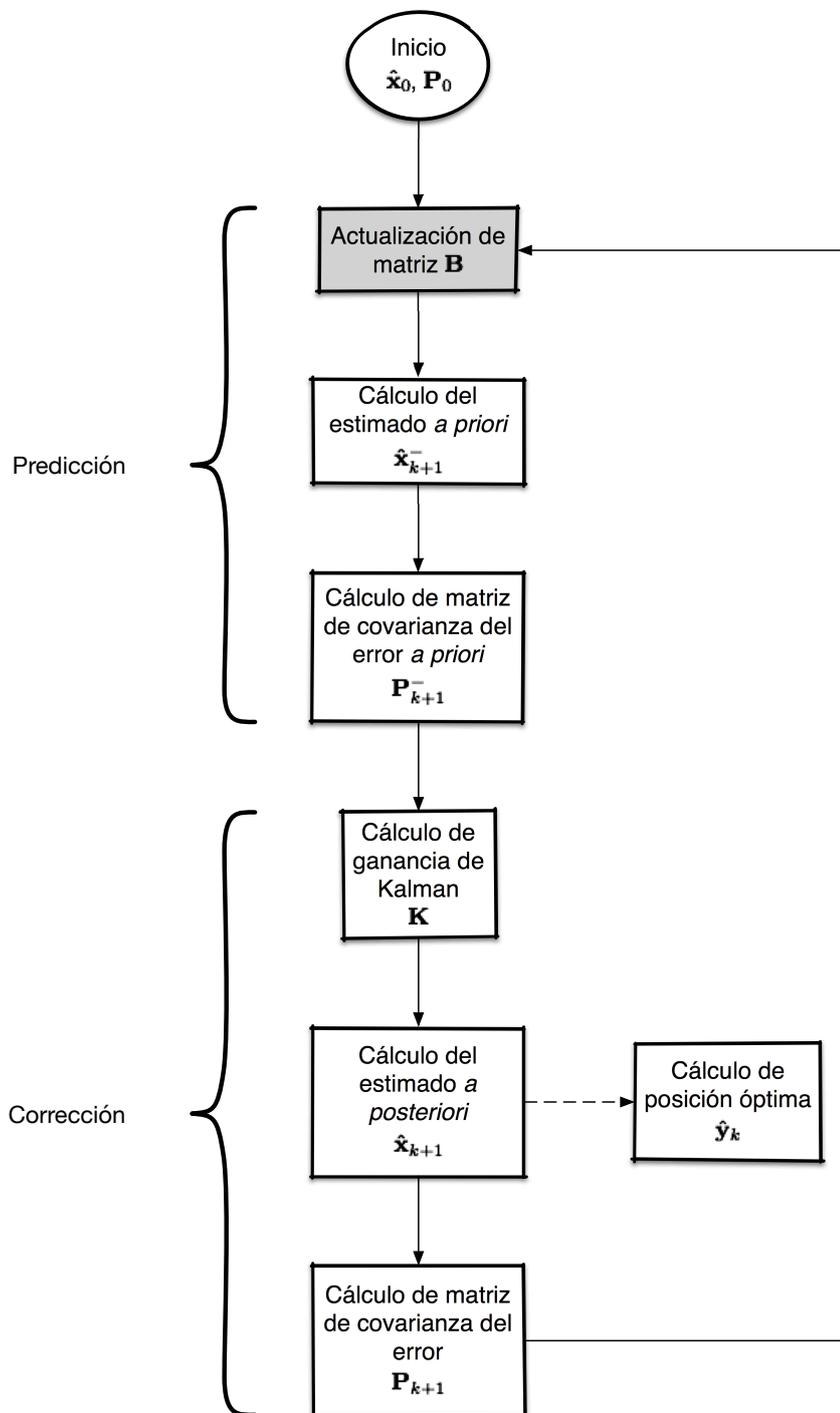


Figura 5.4: Diagrama de flujo de la implementación del filtro de Kalman

5.4.1. Predicción

Actualización de matriz B: Para la actualización de esta matriz es necesario contar con el valor actual de θ_k , que se obtiene a partir del estimado del estado siguiente que se da en la iteración anterior. Para el caso especial de la primera iteración, se utiliza el valor inicial θ_0 contenido en el vector de estado inicial \mathbf{x}_0 .

$$\mathbf{B}[k] = \begin{bmatrix} \frac{\cos(\theta_k)}{2} & \frac{\cos(\theta_k)}{2} \\ \frac{\sin(\theta_k)}{2} & \frac{\sin(\theta_k)}{2} \\ \frac{1}{D} & -\frac{1}{D} \end{bmatrix} \quad (5.30)$$

Estimado a priori: El estimado a priori está dado por la ecuación 4.50, en donde se utiliza el valor actualizado de la matriz \mathbf{B} e identificando que la matriz \mathbf{A} es la matriz identidad, queda entonces:

$$\hat{\mathbf{x}}_{k+1|k}^- = \hat{\mathbf{x}}_{k|k-1} + \mathbf{B}[k]\mathbf{u}_k \quad (5.31)$$

Matriz del error de covarianza a priori: La matriz del error de covarianza a priori se calcula a partir de 4.51, en donde \mathbf{Q} es la matriz de covarianza del ruido en los estados. Este ruido está indicado en 5.28, el cual se puede considerar que se debe a las mediciones de las entradas ΔR y ΔL , por lo tanto es necesario caracterizarlo. Para ello, y recordando que las entradas representan la distancia recorrida por las ruedas en una iteración, el error máximo se supondrá como la distancia de una vuelta de las ruedas.

Realizando un procedimiento similar que el hecho para obtener la matriz de covarianza \mathbf{R} y haciendo uso de 5.17 se puede conocer entonces el error máximo.

$$\Delta R_{max} = T_s r \cdot \omega_{Rmax} \quad (5.32)$$

$$\Delta L_{max} = T_s r \cdot \omega_{Lmax} \quad (5.33)$$

Ahora, si además se asume que este ruido es Gaussiano e independiente y que es necesario multiplicar por el máximo valor de la matriz \mathbf{B} , la varianza del error quedaría de la siguiente manera:

$$\begin{bmatrix} \sigma_x^2 \\ \sigma_y^2 \\ \sigma_\theta^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{D} & -\frac{1}{D} \end{bmatrix} \begin{bmatrix} \left(\frac{\Delta R_{max}}{3}\right)^2 \\ \left(\frac{\Delta L_{max}}{3}\right)^2 \end{bmatrix} \quad (5.34)$$

Y la matriz de covarianza a partir de la ecuación anterior:

$$\mathbf{Q} = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \sigma_\theta^2 \end{bmatrix} \quad (5.35)$$

Finalmente, el cálculo de la matriz de covarianza del error a priori se calcula de la siguiente manera:

$$\mathbf{P}_{k+1}^- = \mathbf{P}_k^* + \mathbf{Q} \quad (5.36)$$

5.4.2. Corrección

Esta etapa se puede realizar de manera múltiple en caso de haber información redundante, como múltiples sensores que arrojen la misma medición o distintas mediciones, en tal caso se utilizaría una matriz de medición \mathbf{C} distinta.

Ganancia de Kalman: La ganancia de Kalman está dada por la ecuación 4.52, que se simplifica un poco gracias a la matriz \mathbf{A} , igual a la matriz identidad.

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}^* [\mathbf{C} \mathbf{P}_k^- \mathbf{C}^* + \mathbf{R}]^{-1} \quad (5.37)$$

Estimado a posteriori: El estimado óptimo se puede calcular una vez que se tiene la ganancia de Kalman, mediante la ecuación 4.53. En este punto se puede entonces obtener la posición estimada óptima al utilizar la ecuación de medición del proceso (5.23).

Matriz de covarianza del error: Por último, el cálculo de la matriz de covarianza del error es el último punto de cada iteración, ésta se calcula mediante la ecuación 4.54, simplificándose un poco.

$$\mathbf{P}_{k+1} = \mathbf{P}_{k+1}^- - \mathbf{P}_k \mathbf{C} \mathbf{K}_k^* - \mathbf{K}_k \mathbf{C} \mathbf{P}_k^* + \mathbf{K}_k \mathbf{C} \mathbf{P}_k \mathbf{C}^* \mathbf{K}_k^* + \mathbf{K}_k \mathbf{R} \mathbf{K}_k^* \quad (5.38)$$

5.5. Resultados

Implementando el algoritmo antes descrito mediante una simulación, se obtuvieron los siguientes resultados que se pueden observar claramente en las siguientes imágenes.

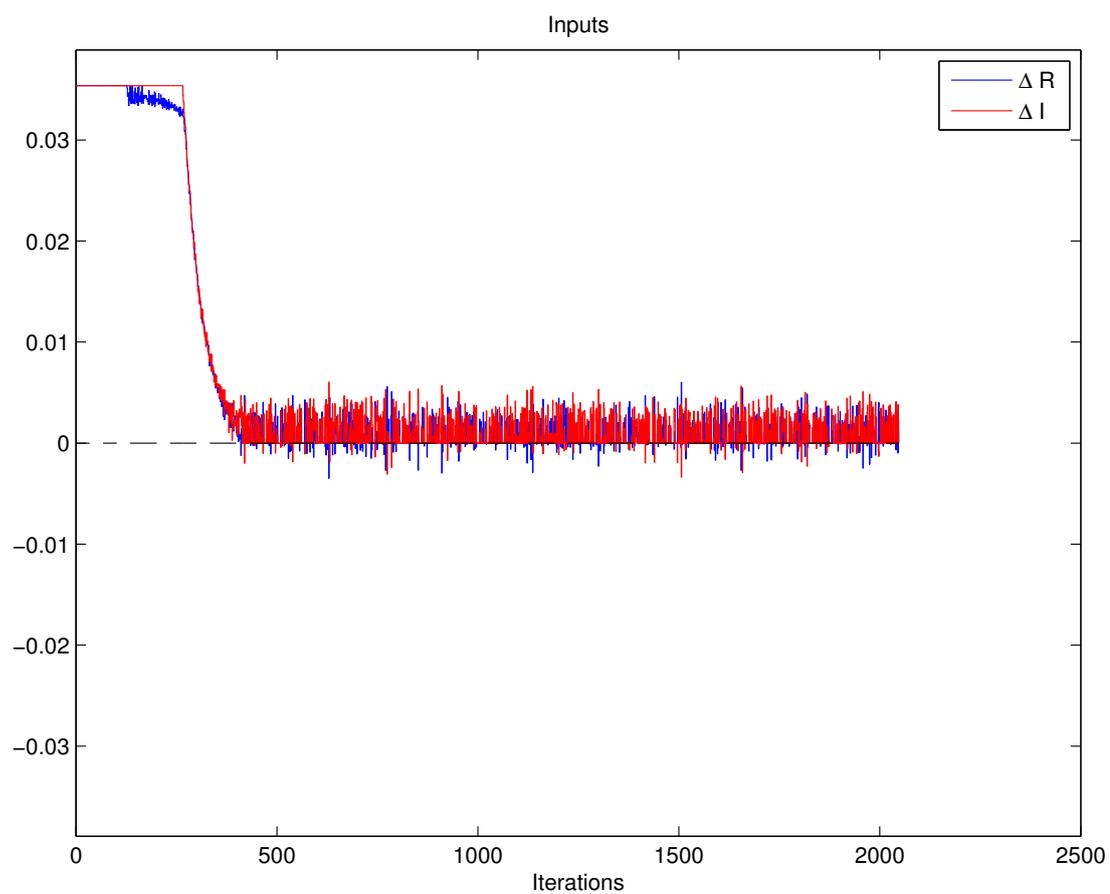


Figura 5.5: Gráfica de las mediciones de las señales de entrada ΔR y ΔL del modelo

5. SEGUIMIENTO DE TRAYECTORIAS

Como ya se menciono antes, la tarea del filtro de Kalman es dar un estimado de la trayectoria, es decir la ruta que describe el vector de estados que describe al sistema a lo largo de periodo de tiempo conocido, con una mayor precisión debido a la naturaleza del filtro el cual minimiza la varianza del error.

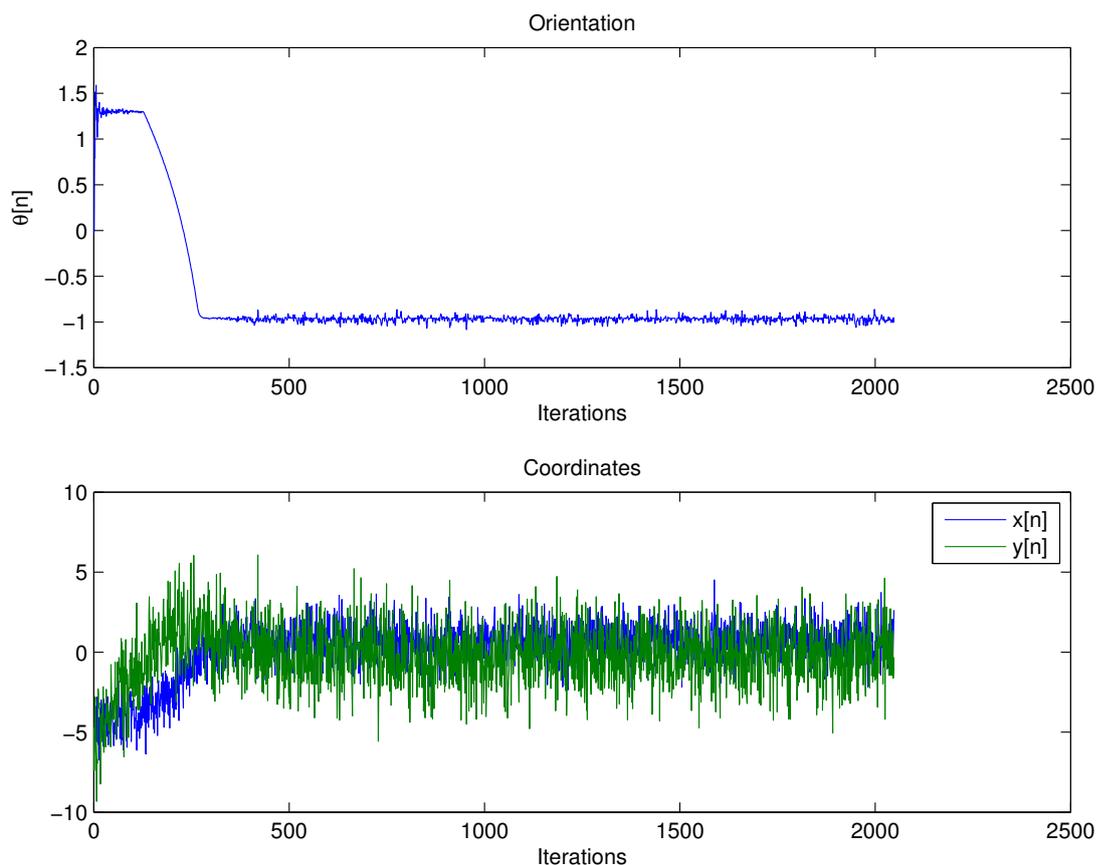


Figura 5.6: Gráfica de los estados óptimos estimados del modelo

Las trayectorias estimadas descritas a lo largo del experimento se ilustran en la figura 5.6, que se obtuvieron a partir de las señales de control, las señales de entrada del sistema, que se muestran en la imagen 5.5.

Una vez que se tiene la trayectoria de los estados internos del sistema es fácil obtener la tarea principal, que de una manera más práctica se puede describir como, obtener la ruta seguida por el móvil a lo largo de un periodo de tiempo. En concreto la ruta descrita se obtiene a partir de las trayectorias en las coordenadas X y Y , a

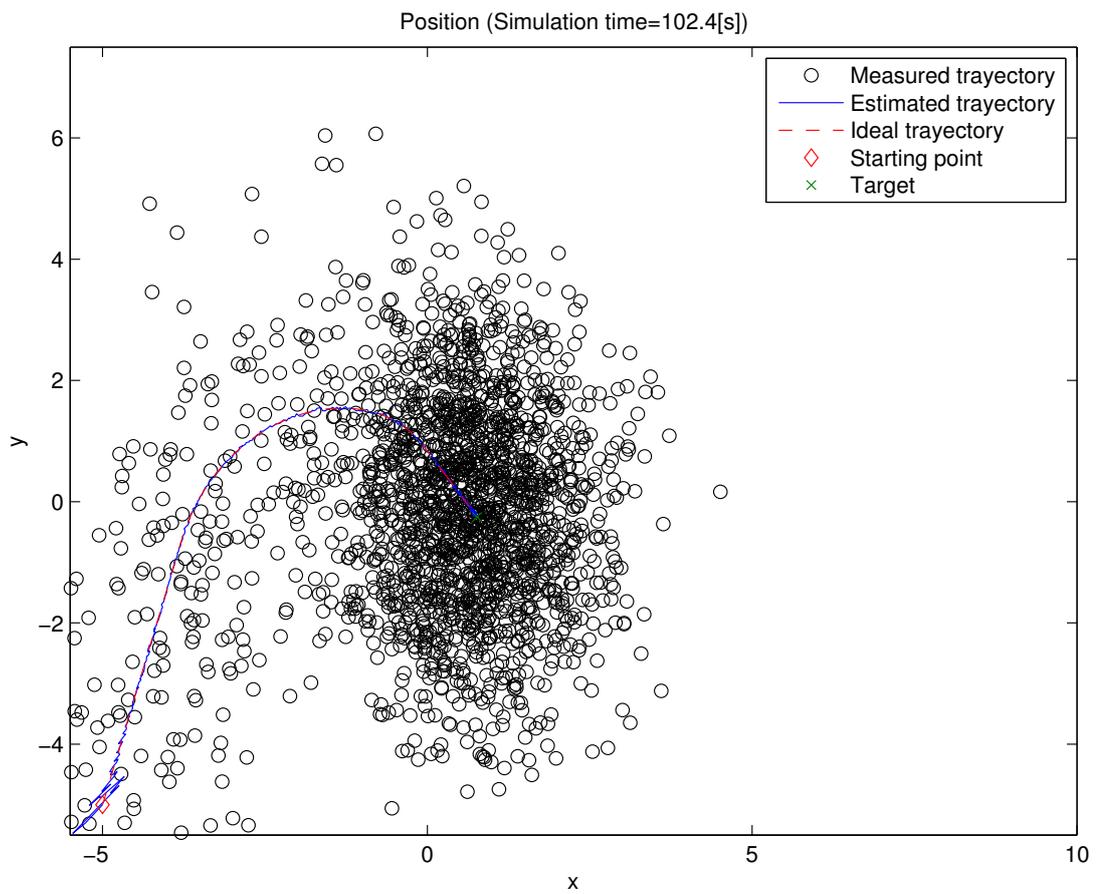


Figura 5.7: Trayectoria óptima estimada a partir de las mediciones

5. SEGUIMIENTO DE TRAYECTORIAS

partir de punto inicial \mathbf{x}_0 .

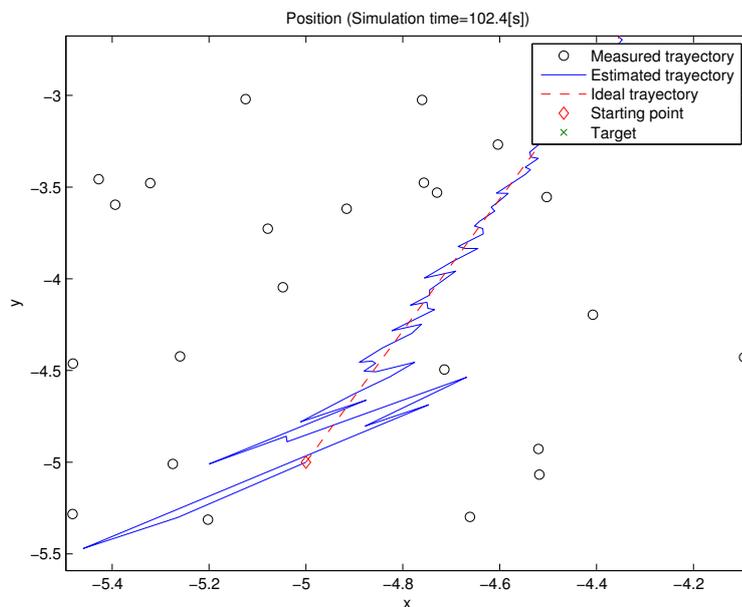


Figura 5.8: Acercamiento de la trayectoria estimada

Esta ruta se puede observar en la figura 5.7, además de comparar con la ruta ideal que supuestamente seguiría el móvil sin la presencia de ruido, es decir si las mediciones fuesen completamente precisas.

Finalmente, los resultados muestran la viabilidad y funcionalidad de la implementación de un filtro de Kalman aplicado a un modelo dinámico, en este caso un robot móvil, para la estimación de la trayectoria del vector de estados el cual cumple la tarea de eliminar los efectos aleatorios del ruido que se pudiese presentar, además de contar con la característica de utilizar toda la información disponible, la fusión de la información provista por diversos sensores, para mejorar aún más la precisión de la estimación.

Conclusiones

Como resultado de este trabajo, se estudiaron los principios básicos necesarios para adentrarse en la teoría de los sistemas, en el modelado y la clasificación. Además de presentarse la bases probabilísticas, conceptos e ideas, que se utilizaron para caracterizar las posibles perturbaciones de naturaleza aleatoria que se llegan a presentar en los modelos matemáticos de los sistemas.

Bajo las premisas anteriores, fue posible estudiar los conceptos básicos de la estimación de una señal, basándose en la minimización del error de la señal estimada. Para ello se presentaron distintos enfoques, como el criterio del error cuadrático medio que es ampliamente utilizado en la teoría de *Wiener* así como el propuesto en la teoría de *Kalman*, los cuales presentan diferentes ventajas dependiendo del problema que se desee resolver, siendo ambos enfoques muy efectivos además de ampliamente utilizados.

Una vez comprendidas las bases teóricas de ambas vertientes, es posible hacer un análisis de las restricciones o limitaciones, dando como resultado natural una comparación implícita de las desventajas presentes en cada uno de los filtros al momento de implementarse.

En este punto vale la pena mencionar las amplias ventajas que se observaron en el uso de un filtro de Kalman en comparación con un filtro *estático* como el de Wiener, que no cuenta con la característica de variar sus parámetros para compensar los cambios de las características de la señal a filtrar. En este sentido el filtro de Kalman se puede llegar a clasificar como un filtro adaptable dado que puede variar para ajustar y mejorar la estimación.

Gracias a esta característica con la que cuenta el filtro de Kalman, es posible encontrar una amplia variedad de aplicaciones prácticas en las que se puede im-

6. CONCLUSIONES

plementar de entre las cuales, el seguimiento de la trayectoria de un móvil, como se presentó en este trabajo, pero además cuenta con un valor agregado que es la posibilidad de mejorar la precisión de la estimación mediante la fusión de información recabada por distintos sensores, o fuentes de información.

De esta manera, junto con la naturaleza recursiva del algoritmo, el filtro utiliza toda la información disponible, en un instante de tiempo debido a múltiples sensores ó la información a lo largo de todo el tiempo de filtrado, para realizar una estimación de los estados internos que describen el comportamiento del sistema. Siendo una consecuencia de esta precisión la eliminación de los efectos del ruido en las mediciones realizadas.

Todo esto a costa de una mayor complejidad matemática y computacional, en comparación con la complejidad requerida para la implementación de un filtro de Wiener, la cual se puede justificar mediante la premisa establecida en la motivación de este trabajo (1.4), en la cual se menciona el considerable incremento que han tenido las microcomputadoras en su capacidad de procesamiento y la disminución del tamaño y costo, lo cual permite la implementación de algoritmos más complejos como filtro de Kalman.

Sistemas y Señales en tiempo discreto

En este apartado se describen los conceptos básicos de la teoría relacionada con los sistemas y señales, en específico en su enfoque en el tiempo discreto.

Con esto se busca introducir las bases, ó en su caso refrescar al memoria, que pueden ser importantes para la comprensión ideas más complejas basadas en estos conceptos elementales, como lo puede ser los filtros lineales el cual es un término utilizado frecuentemente en este trabajo.

A.1. Señal

El término señal es uno de los conceptos más importantes y utilizados dentro de la teoría de sistemas, y a veces no es fácil definir y comprenderlo. De acuerdo con [Oppenheim and Schaffer, 2011, p. 9] la palabra señal se aplica generalmente a *algo* que lleva información. Se dice que las señales llevan generalmente información sobre el estado o el comportamiento de un sistema físico y a menudo se sintetizan señales con el propósito de comunicar información entre seres humanos, o entre seres humanos y máquinas.

Las señales se pueden representar de diferentes formas, comúnmente la representación matemática de las funciones es mediante funciones de una o más variables independientes, pero lo que siempre tienen en común las distintas representaciones, es el hecho de que, la información se encuentra contenida en algún patrón de variaciones.

A.1.1. Señales continuas y discretas

En la teoría de sistemas, ampliamente utilizada en el campo de la ingeniería, una señal se representa normalmente de manera matemática en función de una o más variables independientes. Estas variables independientes que describen a la señal pueden ser continuas o discretas, lo cual nos brinda un punto de comparación entre los tipos de señales representadas por variables continuas y las señales representadas por variables discretas.

Las señales en tiempo continuo, o simplemente señales continuas, se definen en un continuo temporal y se representan por lo tanto con una variable independiente continua. Estas señales se denominan frecuentemente como señales analógicas, esto dado que la variable independiente puede representar cualquier cantidad física [Ingle and Proakis, 2010, p. 22]. Por otra parte, las señales en tiempo discreto se definen en instantes discretos del tiempo, y por lo tanto, la variable independiente toma valores discretos. Otro parámetro para la clasificación de las señales puede ser la amplitud de la misma, la cual puede ser de igual manera continua o discreta, cuando una señal es discreta tanto en la amplitud como en la variable independiente se le conoce entonces como una señal digital [Oppenheim and Schaffer, 2011, p. 9].

A.2. Señales en tiempo discreto

De una manera más amplia se pueden describir a las señales discretas en el tiempo en el ámbito de la teoría de sistemas de manera matemática como secuencias de números. Una secuencia de números x , en los que el n -ésimo número se indica como $x[n]$ y se escribe formalmente usando la siguiente notación [Ingle and Proakis, 2010, Oppenheim and Schaffer, 2011, p. 22, p. 10]

$$x = \{x[n]\}, \quad -\infty < n < \infty \quad (\text{A.1})$$

siendo n un número entero. En casos prácticos estas secuencias surgen de muestrear una señal analógica $x(t)$, en dado caso el n -ésimo número de la secuencia es igual al valor de la señal analógica, $x(t)$, en el instante temporal nT , es decir

$$x[n] = x(nT), \quad -\infty < n < \infty \quad (\text{A.2})$$

en donde la cantidad T se denomina periodo de muestreo, es decir el intervalo entre muestra y muestra, y su inversa es la frecuencia de muestreo f_s [Proakis and Manolakis, 2007][Ingle and Proakis, 2010][Oppenheim and Schaffer, 2011].

Otra manera de representar las señales en tiempo discreto, es decir las secuencias, es mediante una representación gráfica la cual se puede apreciar en la figura A.1.

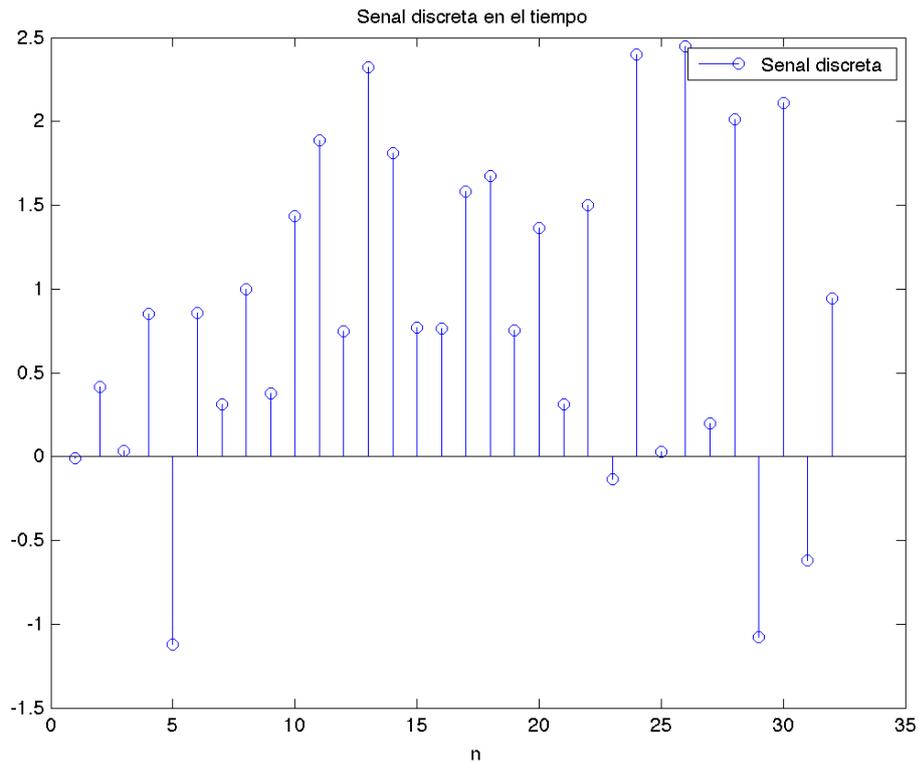


Figura A.1: Representación gráfica de una señal en tiempo discreto

Es importante resaltar que, aunque el eje de la variable independiente se representa mediante una línea continua, la señal $x[n]$ está definida sólo para valores enteros de n , y no se encuentra definida en los instantes entre dos muestras sucesivas. Además de que es incorrecto pensar que $x[n]$ es igual a cero si n no es un entero, simplemente no está definida para valores no enteros de n [Proakis and Manolakis, 2007, pp. 37].

Otro ejemplo de una señal discreta puede ser la que se muestra en la figura A.2, la cual representa una señal analógica muestreada para obtener una representación discreta de la señal.

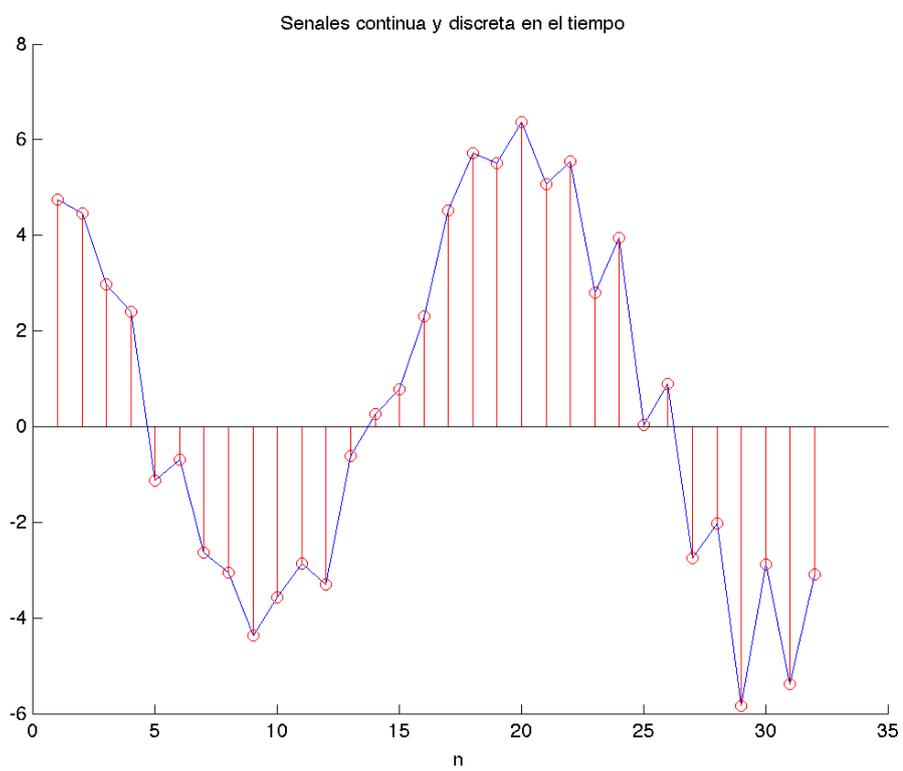


Figura A.2: Señal analógica y señal discreta

A.2.1. Señales elementales

Al presentar la teoría de las señales y los sistemas en tiempo discreto, aparecen varias secuencias básicas de particular importancia. Estas secuencias se muestran en la imagen A.3 [Proakis and Manolakis, 2007][Oppenheim and Schaffer, 2011].

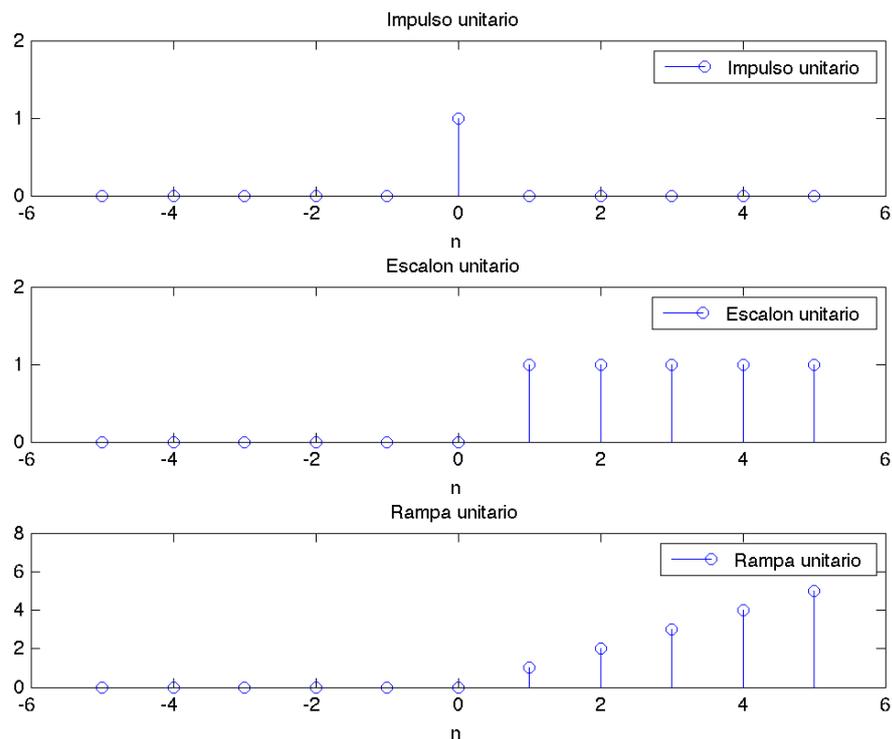


Figura A.3: Secuencias elementales presentes en los sistemas en tiempo discreto

Impulso unitario

La secuencia impulso unitario o muestra unitaria se define como

$$\delta[n] = \begin{cases} 0, & n \neq 0 \\ 1, & n = 0 \end{cases} \quad (\text{A.3})$$

La secuencia muestra unitaria juega el mismo papel en señales y sistemas en tiempo discreto que la función impulso para señales y sistemas en tiempo continuo

[Oppenheim and Schaffer, 2011, p 12-13]. Uno de los aspectos más importantes de la secuencia impulso es que permite expresar una secuencia arbitraria como una suma de impulsos desplazados y escalados, o muestras unitarias ponderadas. En términos más generales, cualquier secuencia se puede expresar de esta forma [Proakis and Manolakis, 2007][Ingle and Proakis, 2010]:

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n-k] \quad (\text{A.4})$$

Esta forma de representar una señal discreta completamente arbitraria se usa específicamente al hablar de la representación de los sistemas lineales en tiempo discreto.

Escalón unitario

La secuencia escalón unitario se define como [Oppenheim and Schaffer, 2011, p. 13]

$$u[n] = \begin{cases} 1, & n \geq 0 \\ 0, & n < 0 \end{cases} \quad (\text{A.5})$$

La secuencia escalón unitario se define como

$$u[n] = \sum_{k=-\infty}^n \delta[k] \quad (\text{A.6})$$

es decir, el valor de la secuencia escalón unitario en el índice temporal n es igual a la suma acumulada hasta el índice n de todos los valores anteriores de la secuencia impulso. O bien se puede representar como la suma de impulsos retardados

$$u[n] = \sum_{k=0}^{\infty} \delta[n-k] \quad (\text{A.7})$$

Rampa unitaria

La señal rampa unitaria se denota como $r[n]$ y se define como [Proakis and Manolakis, 2007, p. 39]

$$r[n] = \begin{cases} n, & n \geq 0 \\ 0, & n < 0 \end{cases} \quad (\text{A.8})$$

Secuencia exponencial

Las secuencias exponenciales son otra clase importante de las señales básicas. La forma general de una secuencia exponencial es [Ingle and Proakis, 2010, p. 24]

$$x[n] = z^n, \quad \forall n \quad (\text{A.9})$$

Exponencial real. En dado caso que la base z de la secuencia exponencial se trate de un número real se pueden presentar los siguientes casos [Oppenheim and Schaffer, 2011, p. 13-15]:

- Si $|z| > 1$, exponencial creciente.
- Si $0 \leq |z| \leq 1$, exponencial decreciente.
- Si $-1 \leq |z| \leq 0$, exponencial decreciente con signo alternante.
- Si $|z| < -1$, entonces se trata de una exponencial creciente, con signo alternante.

Estos casos se ejemplifican en la imagen A.4.

Exponencial compleja. Si en la secuencia exponencial z es un número complejo entonces tiene parte real e imaginaria y se puede expresar mediante la fórmula de Euler [Proakis and Manolakis, 2007, p. 40]

$$z = |z|e^{i\theta_z} = |z|[\cos(\theta_z) + i \sin(\theta_z)] \quad (\text{A.10})$$

De esta manera la ecuación A.9 se puede expresar de la siguiente forma

$$x[n] = z^n \quad (\text{A.11})$$

$$= |z|^n e^{i\theta_z n} \quad (\text{A.12})$$

$$= |z|^n [\cos(\theta_z n) + i \sin(\theta_z n)] \quad (\text{A.13})$$

La ecuación A.13 nos denota que la señal se puede expresar mediante sinusoides ponderadas exponencialmente en donde $|z|$ es el módulo del número complejo z , y representa una atenuación si $|z| < 1$ o una amplificación si $|z| > 1$, y θ_z es la frecuencia de oscilación en radianes de la señal sinusoidal. Se pueden apreciar en la imagen A.5

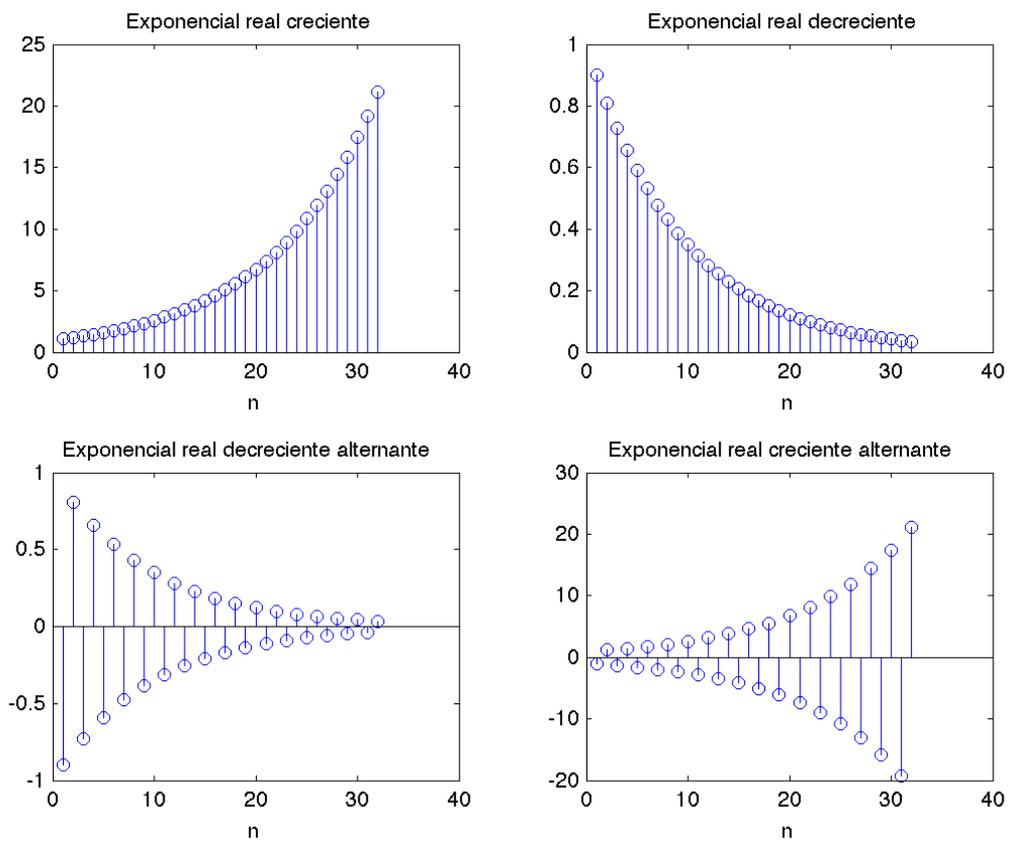


Figura A.4: Secuencias exponenciales reales

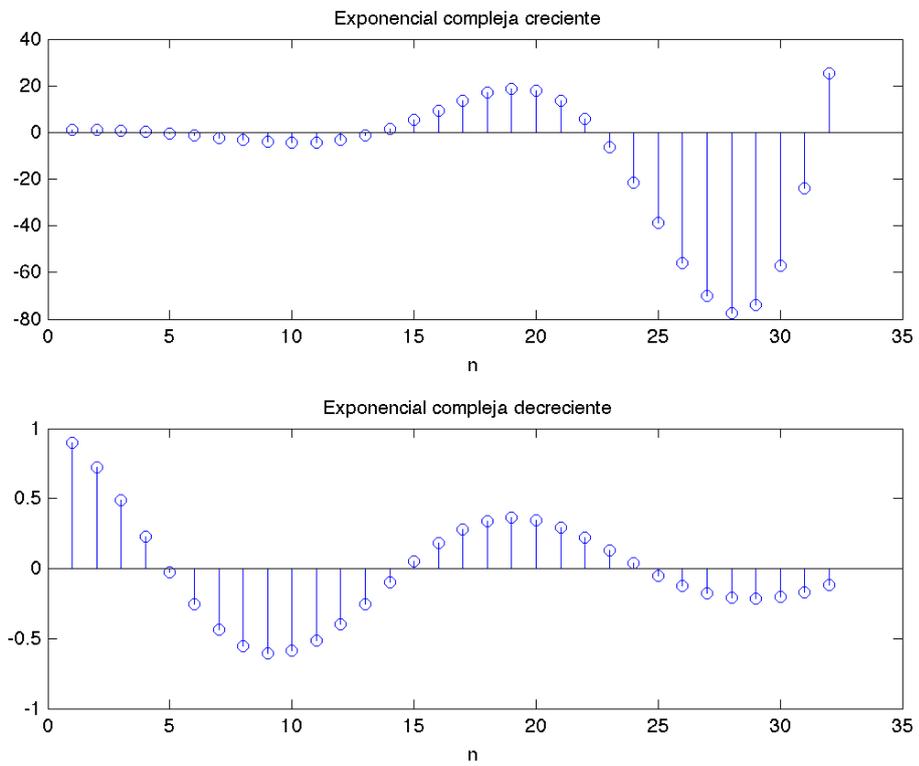


Figura A.5: Señales exponenciales complejas

Secuencias aleatorias

Otro tipo muy importante de señales discretas en el tiempo son las secuencias aleatorias, esto se debe a que muchas situaciones prácticas no pueden ser descritas mediante expresiones matemáticas, ya que los procesos que generan las señales son tan complejos que realizar una descripción precisa de una señal se hace extremadamente difícil e incluso imposible. En estos casos resulta útil modelar analíticamente la señal como un proceso estocástico.

Una señal aleatoria se considera como miembro de un conjunto de señales en tiempo discreto caracterizada por un conjunto de funciones de densidad de probabilidad. Más concretamente, dada una señal particular en un instante concreto, su amplitud en ese instante se supone determinada por un esquema de probabilidades subyacente. Es decir, cada muestra individual $x[n]$ de una señal particular se supone que es una realización de una variable aleatoria subyacente x_n , asociada a funciones de densidad de probabilidad que la describen [Proakis and Manolakis, 2007, p. 64-66][Ingle and Proakis, 2010, p. 25].

A.2.2. Clasificación de las señales discretas en el tiempo

Una vez que se conoce el concepto de *señal* y más específicamente el concepto de *señal discreta en el tiempo*, es importante conocer las características de las señales que se pueden presentar, para que de esta manera se puedan analizar. Esto se debe a que los métodos matemáticos empleados en el análisis de señales y sistemas discretos en el tiempo dependen de las características que presenten las señales.

Para clasificar las señales discretas en el tiempo se pueden utilizar diversos criterios para agrupar las semejanzas que se puedan presentar en las señales.

Señales de energía y potencia

Un criterio de clasificación de las señales discretas en el tiempo es la energía o la potencia promedio de la señal, dado que las señales que tienen una energía media finita presentan características similares entre si, de igual manera lo hacen las señales con potencia media finita.

La energía E de una señal discreta en el tiempo $x[n]$ se define como

$$E = \sum_{n=-\infty}^{\infty} |x[n]|^2 \quad (\text{A.14})$$

Si E es *finita* entonces se dice que $x[n]$ es una *señal de energía* [Proakis and Manolakis, 2007, p. 42]

Dado que muchas señales $x[n]$ poseen energía infinita no se pueden clasificar como señales de energía, pero cuentan con potencia media finita la cual se define de la siguiente manera

$$P = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N |x[n]|^2 \quad (\text{A.15})$$

Si se define la energía de la señal $x[n]$ en el intervalo finito $-N \leq n \leq N$ como

$$E_N \equiv \sum_{n=-N}^N |x[n]|^2 \quad (\text{A.16})$$

De esta manera se puede expresar la energía de la señal

$$E \equiv \lim_{N \rightarrow \infty} E_N \quad (\text{A.17})$$

y la potencia media de la señal $x[n]$ como

$$P \equiv \lim_{N \rightarrow \infty} \frac{1}{2N+1} E_N \quad (\text{A.18})$$

De esta manera si la energía E de la señal es finita, la potencia media será igual a cero, $P = 0$. En caso contrario si la energía es infinita, la potencia media P puede ser finita o infinita. De esta manera se puede decir que si P es finita y distinta de cero, entonces se dice que se trata de una *señal de potencia* [Proakis and Manolakis, 2007, p. 43].

Señales periódicas y aperiódicas

De nuevo otro criterio de clasificación de las señales discretas en el tiempo es la periodicidad de la señal, saber si una señal discreta es periódica o no permite realizar posteriormente un análisis adecuado.

Una señal $x[n]$ es periódica de periodo N si y sólo si

$$x[n+N] = x[n], \quad \forall n \quad (\text{A.19})$$

El valor más pequeño de N para el que la ecuación anterior se cumple se denomina periodo fundamental. En dado caso de que no exista ningún valor de N que satisfaga la expresión anterior, se dice que es *no periódica* o *aperiódica* [Proakis and Manolakis, 2007, p. 43].

Para comprobar esto y sabiendo que la señal sinusoidal de la forma

$$x[n] = A \sin(2\pi f_0 n) \quad (\text{A.20})$$

es periódica cuando f_0 es un número racional, es decir, si f_0 puede expresarse como

$$f_0 = \frac{k}{N} \quad (\text{A.21})$$

donde k y N son enteros.

Un caso especial de análisis es revisar la energía de una señal periódica $x[n]$ en un sólo periodo, es decir, en el intervalo $0 \leq n \leq N - 1$. Sin embargo, la energía de la señal periódica para $-\infty < n < \infty$ es infinita. Por el contrario, la potencia media de la señal periódica es finita e igual a la potencia media en un solo periodo. Por tanto, si $x[n]$ es una señal periódica de periodo fundamental N y toma valores finitos, su potencia viene dada por

$$P = \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2 \quad (\text{A.22})$$

En consecuencia, las señales periódicas son siempre señales de potencia.

A.3. Sistemas en tiempo discreto

Una vez descrito el concepto de señal discreta en el tiempo se puede proceder a describir el concepto, igual de abstracto, de *sistema*. De acuerdo con [Wikipedia, 2014a] un sistema es un objeto complejo cuyos componentes se relacionan con al menos algún otro elemento; puede ser material o conceptual. Otra definición puede ser, "*Un sistema es un conjunto o arreglo de elementos y procesos que están relacionados y cuyo comportamiento satisface las necesidades operacionales y provee ó contempla el sostenimiento para un ciclo vital de los productos*" [Doran, 2006].

Las definiciones anteriores son definiciones muy generales que nos brindan una visión general de lo complejo que puede ser un sistema, pero hablando en especial de los sistemas discretos se podría definir mejor de la siguiente manera.

Sistema discreto Un sistema discreto en el tiempo es un dispositivo o algoritmo que opera sobre una señal discreta en el tiempo, que es la entrada o excitación, de acuerdo con una determinada regla bien definida, para producir una señal discreta en el tiempo, que es la salida o respuesta del sistema. En general, decimos que un sistema es una operación o conjunto de operaciones que se realizan sobre la señal de entrada $x[n]$ para generar la salida $y[n]$ [Proakis and Manolakis, 2007, p. 48].

Matemáticamente un sistema discreto se define como una transformación u operador que transforma una secuencia de entrada con valores $x[n]$ en una secuencia de salida con valores $y[n]$ [Oppenheim and Schaffer, 2011, p. 17]. Esto se puede expresar de la siguiente manera

$$y[n] = T\{x[n]\} \quad (\text{A.23})$$

Esta relación matemática se puede representar gráficamente como se muestra en la figura A.6.

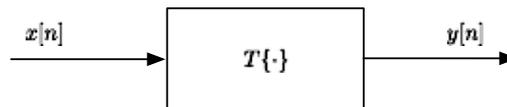


Figura A.6: Representación gráfica de un sistema discreto en el tiempo

A.3.1. Clasificación

Una vez que se definió el concepto de *sistema*, al igual que con las señales, es necesario clasificar a los diferentes sistemas puesto que resulta cómodo de acuerdo con las propiedades generales que satisfacen para posteriormente poder analizarlos, para ello se pueden establecer restricciones en la transformación $T\{\cdot\}$ [Proakis and Manolakis, 2007][Ingle and Proakis, 2010][Oppenheim and Schaffer, 2011].

Se debe destacar que para que un sistema posea una determinada propiedad, es preciso que dicha propiedad se satisfaga para cualquier posible señal de entrada que se aplique al sistema. Si la propiedad sólo se cumple para algunas señales de entrada pero no para otras, quiere decir que el sistema no posee dicha propiedad. Por lo tanto, un único contraejemplo es suficiente para demostrar que un sistema no posee una propiedad dada.

Sistemas sin memoria

Una de las propiedades más básicas de los sistemas es la presencia de memoria, se dice que un sistema es sin memoria si la salida $y[n]$ para cualquier valor de n depende sólo de la entrada $x[n]$ en el mismo valor de n [Oppenheim and Schaffer, 2011, p. 18].

Sistemas estáticos

Se dice que un sistema discreto en el tiempo es estático si su salida en cualquier instante n depende a lo sumo de la muestra de entrada en dicho instante, pero no de muestras pasadas o futuras de la entrada [Proakis and Manolakis, 2007, p. 54]. En cualquier otro caso, se dice que en el sistema es dinámico o que tiene memoria. Nótese que, en general, los sistemas estáticos o sin memoria se describen mediante ecuaciones de entrada-salida de la forma

$$y[n] = T\{x[n], n\} \quad (\text{A.24})$$

y no incluyen elementos de retardo (memoria).

Sistemas lineales

La clase general de sistemas puede subdividirse en sistemas lineales y sistemas no lineales. Para poder determinar si un sistema es lineal o no, se debe de comprobar si la transformación que describe al sistema discreto en el tiempo satisface el principio de superposición, el cual se define de la siguiente manera

$$T\{x_1[n] + x_2[n]\} = T\{x_1[n]\} + T\{x_2[n]\} = y_1[n] + y_2[n] \quad (\text{A.25})$$

y

$$T\{ax[n]\} = aT\{x[n]\} = ay[n] \quad (\text{A.26})$$

La expresión A.25 se denomina *propiedad de aditividad* mientras que a la expresión A.26 se le conoce como *propiedad de homogeneidad* o *escalado*, estas propiedades de los operadores $T\{\cdot\}$ se suelen combinar en el *principio de superposición*, que se formula así [Oppenheim and Schaffer, 2011, p. 18-19]

$$T\{ax_1[n] + bx_2[n]\} = aT\{x_1[n]\} + bT\{x_2[n]\} = ay_1[n] + by_2[n] \quad (\text{A.27})$$

Entonces para que un sistema sea considerado lineal tiene que satisfacer el principio de superposición para cualquier secuencia de entrada $x_1[n]$, $x_2[n]$, y cualquier constante arbitraria a_1 , a_2 .

Extrapolando, la condición de linealidad expresada anteriormente puede ampliarse arbitrariamente a cualquier combinación lineal ponderada de señales [Proakis and Manolakis, 2007, p. 56-57]. En general, se tiene que

$$x[n] \xrightarrow{T} y[n] \quad (\text{A.28})$$

$$\sum_{k=1}^{Q-1} a_k x_k[n] \xrightarrow{T} \sum_{k=1}^{Q-1} a_k y_k[n] \quad (\text{A.29})$$

Suponiendo que cualquier secuencia arbitraria $x[n]$ puede ser expresada mediante una suma ponderada de impulsos retrasados, como se mencionó antes en la expresión A.4, entonces la respuesta $y[n]$ de un sistema lineal a una secuencia de entrada arbitraria $x[n]$ está dada por

$$y[n] = T\{x[n]\} \quad (\text{A.30})$$

$$= T \left\{ \sum_{k=-\infty}^{\infty} x[k] \delta[n-k] \right\} \quad (\text{A.31})$$

$$= \sum_{k=-\infty}^{\infty} x[k] T\{\delta[n-k]\} \quad (\text{A.32})$$

Entonces la respuesta $T\{\delta[n-k]\}$ puede ser interpretada como la respuesta de un sistema lineal en el tiempo n debido a una muestra unitaria (secuencia impulso unitario) en el tiempo k . A esto se le conoce como *respuesta al impulso* [Ingle and Proakis, 2010, p. 37] y se denota por $h[n, k]$, quedando así

$$y[n] = \sum_{k=-\infty}^{\infty} x[k] h[n, k] \quad (\text{A.33})$$

Sistemas causales

El concepto de causalidad es importante y a veces necesario para asegurar que los sistemas pueden ser construidos. Se dice que un sistema es *causal* si la salida del mismo en cualquier instante n (es decir, $y[n]$) sólo depende de las entradas actuales y pasadas (es decir, $x[n]$, $x[n-1]$, $x[n-2]$, \dots) pero no depende de las entradas futuras ($x[n+1]$, $x[n+2]$, $x[n+3]$, \dots) [Proakis and Manolakis, 2007, p. 59-60]. En términos matemáticos, la salida de un sistema causal satisface una ecuación de la forma

$$y[n] = F\{x[n], x[n-1], x[n-2], \dots\} \quad (\text{A.34})$$

donde $F\{\cdot\}$ es alguna función arbitraria. Si un sistema no satisface dicha definición se dice que es *no causal*. Un sistema así tiene una salida que depende no sólo de las entradas actual y pasada, sino también de las entradas futuras.

Sistemas estables

La estabilidad de los sistemas discretos es otra de las grandes propiedades que brindan un punto de comparación para poder clasificar a los sistemas, esta propiedad es además de vital importancia ya que debe tenerse en cuenta al momento de diseñar un sistema para cualquier aplicación práctica. Los sistemas inestables normalmente presentan un comportamiento errático y extremo, y producen desbordamiento en cualquier implementación práctica [Proakis and Manolakis, 2007, p. 60].

Se dice que un *sistema* es *estable* en el sentido de entrada acotada, salida acotada (*BIBO, Bounded Input-Bounded Output*) si y sólo si cualquier secuencia acotada a su entrada produce una secuencia de salida acotada. La condición de que la secuencia de entrada $x[n]$ y la secuencia de salida $y[n]$ sean acotadas se expresa matemáticamente estableciendo que existen determinados números finitos positivos [Oppenheim and Schaffer, 2011, Proakis and Manolakis, 2007, p. 60, p. 22], como por ejemplo, M_x y M_y , tales que

$$|x[n]| \leq M_x < \infty, \quad \forall x, n \quad (\text{A.35})$$

$$|y[n]| \leq M_y < \infty, \quad \forall y, n \quad (\text{A.36})$$

Si, para determinada secuencia de entrada acotada $x[n]$, la salida no está acotada (es infinita), el *sistema* se clasifica como *no estable*.

Invariantes en el tiempo

Un sistema invariante con el tiempo (a menudo denominado también sistema invariante al desplazamiento) es un sistema para el que un desplazamiento temporal o retardo de la secuencia de entrada provoca el mismo desplazamiento o retardo en la secuencia de salida. Supongamos que se dispone de un sistema $T\{\cdot\}$ en estado de reposo que, cuando se excita con una señal de entrada $x[n]$, genera una señal de salida $y[n]$, por lo que podemos escribir

$$y[n] = T\{x[n]\} \quad (\text{A.37})$$

Supongamos que la misma señal de entrada se retarda k unidades de tiempo para proporcionar $x[n-k]$, y de nuevo se aplica al mismo sistema. Si las características del sistema no cambian con el tiempo, la salida del sistema en reposo será $y[n-k]$. Es decir, la salida será la misma que la respuesta a $x[n]$, excepto en que estará retardada k unidades en el tiempo al igual que la entrada. Esto lleva a la siguiente definición de sistema invariante en el tiempo o invariante a desplazamientos [Ingle and Proakis, 2010, Proakis and Manolakis, 2007, p. 54-55, p. 38]:

$$y[n] = T\{x[n]\} \implies T\{x[n-k]\} = y[n-k] \quad (\text{A.38})$$

Supongamos que ahora se tiene un sistema lineal invariante con el tiempo (*LTI*) con una secuencia de entrada $x[n]$ y una respuesta $y[n]$ para dicha entrada. Entonces la respuesta al impulso variante con el tiempo $h[n, k]$ se convierte en una función invariante con el tiempo $h[n-k]$, y la salida de la respuesta al impulso esta dada por

$$y[n] = LTI\{x[n]\} = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \quad (\text{A.39})$$

Entonces se dice que la respuesta al impulso de un sistema invariante en el tiempo (*LTI*) está dado por $h[n]$, una función que no se modifica con el tiempo. La operación matemática de la expresión A.39 se conoce como una suma de convolución lineal y se denota por

$$y[n] = x[n] * h[n] \quad (\text{A.40})$$

Por lo tanto un sistema invariante en el tiempo está completamente caracterizado en el dominio del tiempo por la respuesta al impulso $h[n]$ [Ingle and Proakis, 2010, p. 38-40].

A.4. Ecuaciones en diferencias

Una clase importante de sistemas lineales e invariantes con el tiempo, recursivos y no recursivos, está formada por aquellos sistemas en los que la entrada y la salida satisfacen una ecuación en diferencias lineal con coeficientes constantes de orden P [Ingle and Proakis, 2010, Oppenheim and Schaffer, 2011, Proakis and Manolakis, 2007], para $0 \leq P < \infty$. Las ecuaciones en diferencias son muy importantes en el ámbito del procesamiento digital y juegan un papel determinante en dicho contexto, un análogo podría ser el papel que desempeñan las ecuaciones diferenciales

de coeficientes constantes para describir a los sistemas continuos en el tiempo. La forma general de sistemas descritos mediante ecuaciones en diferencias lineales y coeficientes constantes es

$$\sum_{i=0}^P a_i y[n-i] = \sum_{j=0}^Q b_j x[n-j], \quad \forall n \quad (\text{A.41})$$

En donde el número entero P define el orden de la ecuación en diferencias del sistema. Otra manera de expresar la misma ecuación, considerando que $a_0 = 1$, es la siguiente

$$y[n] = \sum_{j=0}^Q b_j x[n-j] - \sum_{i=1}^P a_i y[n-i] \quad (\text{A.42})$$

La expresión A.42 representa la forma general de los sistemas *recursivos* descritos mediante ecuaciones en diferencias [Proakis and Manolakis, 2007, p. 83]. Esta ecuación expresa la salida del sistema en el instante n directamente como una suma ponderada de las salidas $y[n-1], y[n-2], \dots, y[n-P]$ así como las muestras de las señales de entrada pasadas y presente ($x[n], x[n-1], x[n-2], \dots, x[n-Q]$).

Solución de las ecuaciones en diferencias lineales

Como en el caso de las ecuaciones diferenciales lineales con coeficientes constantes para sistemas en tiempo continuo sin restricciones adicionales, una ecuación en diferencias lineal con coeficientes constantes para sistemas en tiempo discreto no proporciona una especificación única de la salida para una entrada determinada. Para ello el objetivo es determinar una expresión explícita para la salida $y[n]$, el método que se utiliza se conoce como *método directo* [Proakis and Manolakis, 2007, p. 87] y dicta que la solución total es la suma de dos partes

$$y[n] = y_h[n] + y_p[n] \quad (\text{A.43})$$

siendo $y_h[n]$ la parte que se conoce como solución homogénea o complementaria, mientras que $y_p[n]$ es la solución particular.

La solución homogénea es la solución a la ecuación en diferencias homogénea la cual es la ecuación A.41 para una secuencia de entrada $x[n] = 0, \forall n$

$$\sum_{i=0}^P a_i y[n-i] = 0 \quad (\text{A.44})$$

El procedimiento para resolver la ecuación en diferencias homogénea es muy similar al procedimiento utilizado para resolver una ecuación diferencial lineal de coeficientes constantes. Básicamente se supone que la solución pertenece a una familia de soluciones de la forma [Oppenheim and Schafer, 2011]:

$$y[n] = \sum_{i=1}^P a_i z_i^n \quad (\text{A.45})$$

Dicha ecuación forma el *polinomio característico* del sistema. En general tiene P raíces, que designamos como z_1, z_2, \dots, z_P . Las raíces pueden ser reales o complejas. En la práctica, normalmente, los coeficientes a_1, a_2, \dots, a_P son reales. Las raíces complejas aparecen como pares de complejos conjugados. Algunas de las P raíces pueden ser iguales, en cuyo caso se tienen raíces de orden múltiple. El polinomio característico es importante al determinar la estabilidad del sistema. Si las raíces z_i satisfacen la condición

$$|z_i| < 1, \quad i = 1, 2, \dots, P \quad (\text{A.46})$$

entonces el sistema causal descrito por la expresión A.41 es estable [Ingle and Proakis, 2010].

Si se especifica una entrada $x[n]$ entonces es necesario calcular la solución particular del sistema, la cual se determina a partir del lado derecho de la ecuación general A.41, para ello será necesario un conjunto de datos auxiliares que se conocen como *condiciones iniciales*. Estas condiciones auxiliares pueden consistir en unos valores especificados de $y[n]$ para valores concretos de n , como $y[-1], y[-2], \dots, y[-P]$.

Si se especifica la entrada $x[n], \forall n$, junto con un conjunto de valores auxiliares, por ejemplo $y[-1], y[-2], \dots, y[-P]$, utilizando la ecuación en diferencias se puede determinar $y[0]$ y después $y[1]$ y así sucesivamente.

Si un sistema está caracterizado por una ecuación en diferencias lineal con coeficientes constantes y se especifica adicionalmente que debe de ser lineal, invariante con el tiempo y causal, la solución es única. En este caso, las condiciones auxiliares se denominan a menudo condiciones de reposo inicial, o condiciones iniciales nulas. En otras palabras, la información auxiliar es que si la entrada $x[n]$ es cero para valores de n menores de algún instante n_0 ($n < n_0$), la salida $y[n]$ debe valer cero para n menor que n_0 [Oppenheim and Schafer, 2011, p. 40].

Procesos Aleatorios

Este apéndice tiene como objetivo introducir las bases teóricas necesarias para la comprensión del filtrado basado en un enfoque probabilístico, al desarrollar los conceptos que a mi parecer son los más importantes, sin los cuales muy difícil entender la importancia que juega la probabilidad en el diseño, implementación y análisis de los sistemas que presentan perturbaciones de naturaleza aleatoria.

De esta manera, se puede añadir un enfoque diferente a la teoría de control y los sistemas, al considerar elementos que normalmente se consideran ideales, y tomar en cuenta los efectos del ruido que tiene naturaleza aleatoria para contar con una herramienta distinta que permite describir el comportamiento de los sistemas y en dado caso mitigar los efectos no deseados que pudiese tener.

B.1. Variables aleatorias

El concepto de una variable aleatoria es uno de los más importantes en la teoría de la probabilidad, pero también es un poco difícil de entender tal concepto de manera sencilla así como darle una justificación a su uso.

Para hacer esta idea abstracta un poco más fácil de explicar, será necesario hacer uso de los dos experimentos aleatorios más simples y utilizados: lanzar un dado y un volado.

Cuando se realiza el experimento de lanzar un dado, el resultado obtenido son el número de puntos en una de sus caras (1 - 6). Mientras tanto en un volado solo hay dos posibles resultados "águila" ó "sol". En cada uno de los casos los resultados se dice que son: *resultados de palabra* [Dolecek, 2013].

Generalmente, es mas conveniente *asignar valores numéricos* a los resultados de los experimentos aleatorios [Dolecek, 2013, p. 28] para que de esta manera se pueda obtener una *expresión matemática* que puede ser utilizada para analizar dichos resultados aleatorios, esta idea nos da la pauta para entender el concepto de una variable aleatoria.

Definición de una variable aleatoria

Una variable aleatoria X es una *función* la cual asigna un número $X(s_i)$, real o complejo, a cada resultado aleatorio s_i , de acuerdo con ciertas reglas.

Cada punto del espacio S , de todos los resultados, debe corresponder con uno y solamente un punto en el eje x . Esto significa que no es posible que haya una correspondencia entre un resultado en particular s_i hacia dos o más puntos en el eje x . Pero si es posible que un punto en específico del eje x sea correspondido por distintos resultados. Entonces cada uno de los valores posibles que se pueden tomar en el eje x conforman lo que se conoce como el rango de la variable aleatoria [Dolecek, 2013].

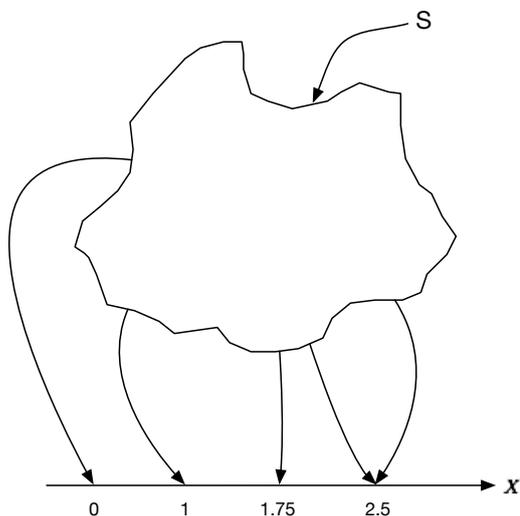


Figura B.1: Correspondencia del espacio S hacia el eje x

Sabiendo lo que es el rango de una variable aleatoria, se puede determinar que

si dicho rango es continuo entonces el espacio de muestras también debe de serlo, En caso contrario, algunos valores del rango continuo de x no tendrían una correspondencia con el espacio S .

Un espacio continuo no significa que siempre resultará en una variable continua. Generalmente, de un espacio continuo pueden obtenerse variables aleatorias continuas y discretas e inclusive pueden ser mixtas[Dolecek, 2013, p. 29-30].

B.1.1. Caracterización

Con el fin de describir completamente una variable aleatoria, es necesario saber no solamente todos los posibles valores que puede tomar si no también que tan frecuentemente se presentan dichos valores. Por ello es necesario establecer las probabilidades correspondientes para cada posible resultado [Peyton Z. Peebles, 2001].

Función de distribución

Para conseguir caracterizar una variable aleatoria es necesario definir una función matemática que exprese dichas probabilidades, a esta función se le conoce como función de distribución de una dimensión o simplemente función de distribución. Se dice que la probabilidad de que una variable aleatoria X tenga un valor menor o igual que x se expresa de la siguiente manera [Dolecek, 2013, p. 35]

$$F_X(x) = P\{X \leq x\}, \quad -\infty < x < \infty \quad (\text{B.1})$$

Esta expresión es válida tanto para variables continuas como para variables aleatorias.

Una manera de calcular dicha función es dividir todo el eje x en subintervalos de acuerdo a los valores correspondientes de la variable aleatoria. Posteriormente es necesario encontrar la probabilidad de cada intervalo, es decir de cada valor conocido.

Propiedades de una función de distribución

Para que una función sea considerada como una función de distribución debe de cumplir con las siguientes propiedades [Dolecek, 2013, p. 40-41]:

Propiedad 1.

$$0 \leq F_X(x) \leq 1 \quad (\text{B.2})$$

Esta propiedad es fácil de justificar sabiendo que de acuerdo con B.1 la distribución es la probabilidad y esta debe tener valores entre 0 y 1.

Propiedad 2.

$$F_X(-\infty) = 0 \quad (\text{B.3})$$

$$F_X(\infty) = 1 \quad (\text{B.4})$$

La primera ecuación viene del hecho de que $P\{X \leq -\infty\}$ es la probabilidad de que un conjunto vacío, por lo tanto es igual a 0. La segunda ecuación denota la probabilidad de un evento completamente certero lo cual es igual a 1 [Dolecek, 2013].

Propiedad 3.

$$F_X(x_1) \leq F_X(x_2), \quad \forall x_1 < x_2 \quad (\text{B.5})$$

Esta propiedad establece que una función de distribución es una función no decreciente.

Propiedad 4.

$$F_X(x^+) = F_X(x) \quad (\text{B.6})$$

Esta propiedad dicta que una función de distribución es una función continua por la derecha. Esto significa que, mientras nos acercamos a x por la derecha, el límite del valor de la función de distribución debe de ser el valor de la función de distribución en dicho punto [Dolecek, 2013, p. 40-41].

Función de densidad de probabilidad

La función de distribución introducida en B.1 es una herramienta matemática usada para describir a una variable aleatoria. Sin embargo, a veces es un tanto complejo y laborioso trabajar con ella. Una alternativa más simple de trabajar es usar la función de densidad de probabilidad. Una función de densidad de probabilidad, o simplemente función de densidad, es la derivada de la función de distribución $F_X(x)$ y se define de la siguiente manera

$$f_X(x) = \frac{dF_X(x)}{dx} \quad (\text{B.7})$$

o de manera alternativa

$$F_X(x) = \int_{-\infty}^x f_X(v) dv \quad (\text{B.8})$$

Entonces, de acuerdo con B.8, la función de distribución en un punto x corresponde al área bajo la curva de la función de densidad desde $-\infty$ hasta x

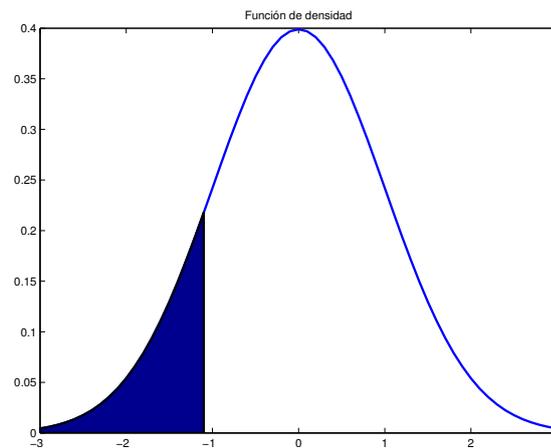


Figura B.2: Representación gráfica de la relación entre una función de distribución y su función de densidad de probabilidad

Propiedades de una función de densidad de probabilidad

Para que una función sea considerada como una función de densidad debe de cumplir con las siguientes propiedades [Dolecek, 2013, p. 49-50]:

Propiedad 1.

$$f_X(x) \geq 0, \quad \forall x \quad (\text{B.9})$$

Esta propiedad se puede interpretar al afirmar que una función de densidad es una función que nunca es negativa. Esto se basa en la propiedad 3 (B.5) de una función de distribución, en consecuencia la derivada de una función no decreciente no puede ser una función negativa.

Propiedad 2.

$$\int_{-\infty}^{\infty} f_X(x) dx = 1 \quad (\text{B.10})$$

Esta propiedad viene de la definición alternativa de la misma función (B.8) la cual al reemplazar x en el límite superior por ∞ se tiene

$$\begin{aligned} F_X(\infty) &= 1 \\ &= P\{X \leq \infty\} &= \int_{-\infty}^{\infty} f_X(x) dx \end{aligned}$$

Propiedad 3.

$$P\{x_1 \leq X \leq x_2\} = \int_{x_1}^{x_2} f_X(x) dx \quad (\text{B.11})$$

En este caso usando la definición alternativa nuevamente (B.8) y la definición de la función de distribución (B.1) se tiene que

$$\begin{aligned} P\{x_1 \leq X \leq x_2\} &= P\{X \leq x_2\} - P\{X \leq x_1\} \\ &= F_X(x_2) - F_X(x_1) \\ &= \int_{-\infty}^{x_2} f_X(x) dx - \int_{-\infty}^{x_1} f_X(x) dx \\ &= \int_{x_1}^{x_2} f_X(x) dx \end{aligned}$$

Lo que nos dicta que la probabilidad de que una variable aleatoria se encuentre dentro del intervalo $[x_1, x_2]$ es igual al área bajo la curva de la función de densidad en dicho intervalo.

Una vez conocidas las funciones de distribución y densidad de probabilidad es importante saber también que otra posible manera de caracterizar a las variables aleatorias es mediante las transformadas de *Fourier* y *Laplace*, las cuales son cruciales en la descripción de propiedades de señales determinísticas y sistemas. De manera similar, es posible aplicar estas transformadas a las funciones de densidad y distribución, esto es posible debido a que las transformadas de *Fourier* y *Laplace* son funciones determinísticas [Dolecek, 2013, p. 110].

B.1.2. Momentos de una variable aleatoria

Una manera de caracterizar de manera parcial a las variables aleatorias es mediante el uso de momentos, los cuales nos brindan información útil acerca de la variable aleatoria que se desea caracterizar y proveen una visión general de la misma.

Las funciones de densidad y distribución proveen toda la información acerca de una variable aleatoria, pero existen muchos casos en los que no es necesaria tanta información o solo se cuenta con los momentos que nos dan el conocimiento necesario de una variable en particular sin necesidad de conocer sus funciones características.

Para poder continuar con la definición de los momentos es importante introducir el concepto del operador *valor esperado* o *esperanza matemática* [Dolecek, 2013, p. 75]. Este se define como

$$E\{g(X)\} = \int_{-\infty}^{\infty} g(X)f_X(x)dx \quad (\text{B.12})$$

dicho operador cumple con una serie de propiedades que son las siguientes:

Propiedad 1. El valor esperado de una constante es la constante misma

$$E\{b\} = bx1 = b, \quad b = \text{constante} \quad (\text{B.13})$$

Propiedad 2. El valor esperado de un valor esperado es el valor esperado mismo

$$E\{E\{x\}\} = E\{x\} \quad (\text{B.14})$$

Propiedad 3. El valor esperado de una variable aleatoria X multiplicada por una constante a es igual al producto de la constante a por el valor esperado de la variable x

$$E\{aX\} = aE\{X\} \quad (\text{B.15})$$

Propiedad 4. Una combinación de las propiedades anteriores sería

$$E\{aX + b\} = aE\{X\} + b \quad (\text{B.16})$$

Momentos respecto al origen

Los momentos respecto al origen, se denotan como m_n donde n es el momento n -ésimo y son una clase importante. Se definen de la siguiente manera

$$m_n = E\{X^n\} = \int_{-\infty}^{\infty} x^n f_X(x) dx \quad (\text{B.17})$$

En donde el operador $E\{\cdot\}$ es el operador lineal *valor esperado* e integra el producto de la función de densidad por el operando, en este caso x^n .

Momento central cero. El momento central cero se define como

$$m_0 = E\{X^0\} = E\{1\} = 1 \quad (\text{B.18})$$

Los momentos más importantes de una variable aleatoria son los primeros y segundos momentos los cuales se les conoce comúnmente como *valor medio* y *valor cuadrático medio* respectivamente.

Valor medio

El valor medio de una variable aleatoria es uno de los más importantes parámetros asociados al concepto y teoría de las variables aleatorias y juega un papel importante cuando la caracterización parcial de la variable es necesaria o es la única información.

Normalmente, todo mundo se encuentra familiarizado con el término *promedio* cuando se refiere a un conjunto finito conocido de valores. Por ejemplo, considerando un conjunto de temperaturas de todos los días del año de una cierta ubicación se puede saber toda la información necesaria de la temperatura del lugar, pero si se calcula la temperatura promedio, por ejemplo 16.5°C, puede ser que ese valor nunca se haya presentado o se presente. Sin embargo, hablando de manera general se conoce suficiente información acerca del clima de ese lugar, se puede aseverar entonces que no va nevar y que además no hará un calor excesivo.

Para comprender mas profundamente dicho concepto se definirá mediante un ejemplo. Si se tiene conocimiento de un conjunto de valores x_i , $i = 1, 2, \dots, N$, entonces se sabe que el valor medio o promedio (m_{av}) es la suma de los valores del conjunto S dividido entre el número total de elementos.

$$m_{av} = \frac{S}{N} = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (\text{B.19})$$

Entonces el valor medio generalizado para un conjunto de valores x_i se puede interpretar como el punto y para el cual la distancia al cuadrado de cada uno de los puntos del conjunto S es mínima [Dolecek, 2013, p. 76-77]. Esto se expresa de la siguiente manera

$$D = \sum_{i=1}^N (y - x_i)^2 \quad (\text{B.20})$$

Para encontrar el punto crítico¹ en donde la distancia es mínima es necesario derivar respecto a y e igualar a cero

$$\frac{dD}{dy} = 0 \quad (\text{B.21})$$

$$2 \sum_{i=1}^N (y - x_i) = \quad (\text{B.22})$$

$$2 \sum_{i=1}^N y - 2 \sum_{i=1}^N x_i = \quad (\text{B.23})$$

$$2Ny - 2 \sum_{i=1}^N x_i = 0 \quad (\text{B.24})$$

reacomodando la expresión anterior

$$2Ny = 2 \sum_{i=1}^N x_i$$

$$Ny = \sum_{i=1}^N x_i$$

$$y = \frac{\sum_{i=1}^N x_i}{N}$$

Al expresar la sumatoria de la forma extendida entonces se llega a la siguiente expresión, la cual se puede observar que es la misma que la misma ecuación necesaria para calcular el promedio de un conjunto finito de datos (B.19).

¹En matemáticas un punto crítico es un punto del dominio de una función diferenciable, donde la derivada es cero

$$y = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (\text{B.25})$$

Entonces, el promedio m_{av} de un conjunto de números $x_i, i = 1, 2, \dots, N$, puede ser visto como el número el cual se encuentra a una misma distancia de todos los números en el conjunto [Dolecek, 2013, p. 75-77].

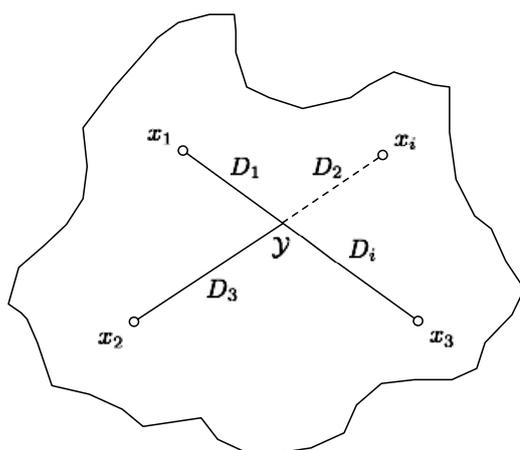


Figura B.3: Interpretación del valor medio

Finalmente, una vez que se comprende el concepto del valor medio y su importancia debido a la información que brinda, el primer momento cuenta con una notación especial \bar{X} , el cual indica el valor medio de X y se define de la siguiente manera

$$\bar{X} = m_1 = E\{X^1\} = E\{X\} \quad (\text{B.26})$$

Segundo momento. El segundo momento se llama valor cuadrático medio y es una medición de la fuerza o *potencia* de la variable aleatoria [Dolecek, 2013, p. 97], y se define de la siguiente manera

$$m_2 = E\{X^2\} \quad (\text{B.27})$$

La raíz cuadrada del valor cuadrático medio es conocida como el valor *rms* (*root mean squared*) de X .

$$X_{rms} = rms(X) = \sqrt{E\{X^2\}} \quad (\text{B.28})$$

Otra interpretación acerca del primer y segundo momento es pensar en $f_X(X)$ como una masa distribuida a lo largo del eje x [Dolecek, 2013, p. 97], entonces

- El primer momento calcula el centro de gravedad de la masa.
- El segundo momento es el momento central de inercia de la distribución de la masa al rededor del centro de gravedad.

Momentos centrales

Otra clase de momentos son los momentos centrales, los cuales a diferencia de los momentos respecto al origen, nos brindan una medición acerca de la aleatoriedad de la variable aleatoria. Considerando el caso en el que el valor medio de una variable aleatoria es muy grande, entonces se puede decir que la componente determinística de la variable es dominante, por lo cual casi se pierde la aleatoriedad [Dolecek, 2013, p. 98-99].

Los momentos centrales están definidos como el valor esperado de la función

$$g(X) = (X - E\{X\})^n \quad (\text{B.29})$$

Los momentos centrales se denotan por μ_n y quedan definidos entonces como

$$\mu_n = E\{(X - E\{X\})^n\} \quad (\text{B.30})$$

$$= E\{(X - \bar{X})^n\} \quad (\text{B.31})$$

De la ecuación B.31 se puede decir que el momento central cero es igual a 1, mientras que el primer momento central es igual a cero

$$\mu_0 = 1 \quad (\text{B.32})$$

$$\mu_1 = 0 \quad (\text{B.33})$$

Varianza

El momento central más importante es el segundo, μ_2

Es tan importante que inclusive cuenta con su propia denominación (varianza) y su propia notación, σ^2 . Cuando se quiere especificar que σ^2 es la varianza de una variable aleatoria X , entonces se escribe σ_X^2 . Entonces la varianza queda definida como

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \bar{X})^2 f_X(x) dx \quad (\text{B.34})$$

Otra expresión equivalente puede ser

$$\sigma_X^2 = E\{X^2\} - \bar{X}^2 \quad (\text{B.35})$$

Lo cual nos denota que $\sigma_X^2 \geq 0$. En dado caso de que la varianza sea cero, esto significa que no hay variación alguna de los resultados al rededor del valor medio de x y en consecuencia la variable X se convierte en una constante y por lo tanto la variable no es más una variable aleatoria. En este sentido es por eso que la varianza puede ofrecer una medición de la aleatoriedad de una variable aleatoria [Dolecek, 2013, p. 99-100].

La raíz cuadrada positiva de la varianza es conocida como desviación estandar

$$\sigma_X = \sqrt{\sigma_X^2} \quad (\text{B.36})$$

Y de igual manera provee una medición acerca de la dispersión de la variable aleatoria al rededor del valor medio \bar{X} .

B.2. Variables aleatorias multidimensionales

Una vez definido el concepto de una variable aleatoria, es necesario entender que existen muchos problemas prácticos en los cuales es necesario crear una correspondencia de no sólo una sola dimensión, si no una correspondencia de dos ó más dimensiones. Por ejemplo en el caso contar con un espacio bidimensional se obtiene una correspondencia hacia una variable aleatoria de dos dimensiones (X_1, X_2) .

Dado que X_1 y X_2 son resultado del mismo experimento, será necesario hacer una caracterización conjunta para estas dos variables. En el caso general, la correspondencia del espacio S es hacia un conjunto de N dimensiones y esto conlleva a una caracterización conjunta para una variable aleatoria de N dimensiones [Dolecek, 2013, p. 155].

B.2.1. Caracterización

Función de distribución

La función de densidad conjunta, o simplemente función de distribución para un par de variables aleatorias X_1 y X_2 está definida como la probabilidad conjunta de los eventos

$$\{X_1 \leq x_1; X_2 \leq x_2\} \quad (\text{B.37})$$

se denota como $F_{X_1 X_2}(x_1, x_2)$ y está definida por

$$F_{X_1 X_2}(x_1, x_2) = P\{X_1 \leq x_1; X_2 \leq x_2\}, \quad \forall x_1, x_2 \quad (\text{B.38})$$

donde x_1 y x_2 son los valores que pueden tomar las variables aleatorias de un espacio bidimensional. La función de distribución denotada por B.38 es llamada también función de distribución bidimensional. La versión más general es una función de distribución de N dimensiones que describe la probabilidad conjunta para N variables aleatorias

$$(X_1, X_2, \dots, X_N) \quad (\text{B.39})$$

esta función se denota por $F_{X_1 \dots X_N}(x_1, \dots, x_N)$ y está definida por [Dolecek, 2013, p. 160-162]

$$F_{X_1 \dots X_N}(x_1, \dots, x_N) = P\{X_1 \leq x_1; \dots; X_N \leq x_N\} \quad (\text{B.40})$$

Las propiedades de la función de distribución unidimensional se siguen aplicando a las funciones de densidad de N variables aleatorias, sólo es necesario considerar que se cuenta con un espacio de dimensión N .

Función de densidad

La función de densidad de probabilidad conjunta o simplemente función de densidad para dos variables aleatorias X_1 y X_2 , se denota como $f_{X_1 X_2}(x_1, x_2)$, y está definida de la siguiente manera

$$f_{X_1 X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1 X_2}(x_1, x_2)}{\partial x_1 \partial x_2} \quad (\text{B.41})$$

De manera equivalente a la función de densidad de una dimensión, en donde la probabilidad representaba el área bajo la función de densidad, para una función de un par de variables aleatorias, la probabilidad es el volumen debajo de la superficie

formada por la función de densidad conjunta.

De igual manera que sucede con la función de distribución, las propiedades de la función de densidad se conservan para un espacio dimensional de tamaño N . En el cual la función de densidad se define así [Dolecek, 2013, p. 162-165]

$$f_{X_1 \dots X_N}(x_1, \dots, x_N) = \frac{\partial^N F_{X_1 \dots X_N}(x_1, \dots, x_N)}{\partial x_1 \dots \partial x_N} \quad (\text{B.42})$$

B.2.2. Momentos

Después de establecer las formas de caracterizar a una variable aleatoria multidimensional mediante las funciones de distribución y de densidad conjunta de dimensión N . Es importante extender el concepto de los momentos a un ámbito más general aplicable a N variables aleatorias conjuntas.

Momentos respecto al origen

Para el caso especial de dos variables aleatorias X y Y , los momentos respecto al se definen como

$$m_{nk} = E\{X^n Y^k\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy \quad (\text{B.43})$$

en donde se dice que el momento es de orden $r = n + k$.

Esta ecuación puede ser generalizada para un conjunto de N variables aleatorias (X_1, \dots, X_N) [Dolecek, 2013, p. 174] de la siguiente manera

$$E\{X_1^{n_1}, \dots, X_N^{n_N}\} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{n_1}, \dots, x_N^{n_N} f_{X_1 \dots X_N}(x_1, \dots, x_N) dx_1 \dots dx_N \quad (\text{B.44})$$

En donde el orden del momento está dado por

$$r = \sum_{i=1}^N n_i \quad (\text{B.45})$$

Los momentos de primer orden $m_{01} = E\{Y\}$ y $m_{10} = E\{X\}$ son los valores esperados de X (\bar{X}) y Y (\bar{Y}), y son las coordenadas del centro de gravedad de la función $f_{XY}(x, y)$ [Peyton Z. Peebles, 2001, p. 119].

Para el momento de segundo orden en donde $n = k = 1$ (m_{11}) cobra especial importancia a tal grado que tiene su propia notación (R_{XY}), se le llama *correlación* y se define de la siguiente manera

$$R_{XY} = E\{XY\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{XY}(x, y) dx dy \quad (\text{B.46})$$

Si la correlación puede ser escrita en la forma

$$R_{XY} = E\{X\}E\{Y\} \quad (\text{B.47})$$

entonces se dice que las variables aleatorias X y Y son *independientes* o *no correlacionadas*. Si la correlación es cero se dice que las variables son *ortogonales* [Dolecek, 2013, Peyton Z. Peebles, 2001].

$$R_{XY} = 0 \quad (\text{B.48})$$

Momentos centrales

En el caso de contar con un par de variables aleatorias X y Y , los momentos centrales se definen como [Dolecek, 2013, p. 175]

$$\mu_{nk} = E\{(X - E\{X\})^n (Y - E\{Y\})^k\} \quad (\text{B.49})$$

$$= E\{(X - \bar{X})^n (y - \bar{Y})^k\} \quad (\text{B.50})$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})^n (y - \bar{Y})^k f_{XY}(x, y) dx dy \quad (\text{B.51})$$

en donde se dice que el momento es de orden $r = n + k$.

Esta expresión puede ser generalizada para un conjunto de N variables aleatorias (X_1, \dots, X_N) [Dolecek, 2013, p. 176] de la siguiente manera

$$\mu_{n_1 \dots n_N} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (x_1 - \bar{X}_1)^{n_1} \dots (x_N - \bar{X}_N)^{n_N} f_{X_1 \dots X_N}(x_1, \dots, x_N) dx_1 \dots dx_N \quad (\text{B.52})$$

Los momentos de segundo orden μ_{20} y μ_{02} son las varianzas de X y Y respectivamente [Peyton Z. Peebles, 2001, p. 120].

$$\mu_{20} = E\{(X - \bar{X})^2\} = \sigma_X^2 \quad (\text{B.53})$$

$$\mu_{02} = E\{(Y - \bar{Y})^2\} = \sigma_Y^2 \quad (\text{B.54})$$

El momento de segundo orden en donde $n = k = 1$ (μ_{11}) es muy importante. Se llama covarianza de X y Y y está dado por el símbolo C_{XY} y definido como

$$C_{XY} = \mu_{11} = E\{(X - \bar{X})(Y - \bar{Y})\} \quad (\text{B.55})$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{X})(y - \bar{Y}) f_{XY}(x, y) dx dy \quad (\text{B.56})$$

$$C_{XY} = R_{XY} - E\{X\}E\{Y\} \quad (\text{B.57})$$

En dado caso de que X y Y sean variables independientes (B.47) la ecuación B.57 se simplifica y la covarianza es cero [Dolecek, 2013, Peyton Z. Peebles, 2001].

$$C_{XY} = R_{XY} - E\{X\}E\{Y\} \quad (\text{B.58})$$

$$= E\{X\}E\{Y\} - E\{X\}E\{Y\} \quad (\text{B.59})$$

$$C_{XY} = 0 \quad (\text{B.60})$$

El otro caso especial de que X y Y sean variables ortogonales (B.48) la ecuación B.57 se reduce de la siguiente manera [Dolecek, 2013, Peyton Z. Peebles, 2001]

$$C_{XY} = R_{XY} - E\{X\}E\{Y\} \quad (\text{B.61})$$

$$= -E\{X\}E\{Y\} \quad (\text{B.62})$$

$$(\text{B.63})$$

B.2.3. Correlación e Independencia

Se dice que las variables aleatorias son independientes si una variable no tiene ninguna influencia sobre los valores de la otra variable, y viceversa. Para variables no aleatorias, el término dependencia significa que si se conoce una variable se pueden encontrar los valores exactos de la otra variable.

Por otra parte, la dependencia en las variables aleatorias es diferente, es por eso que se dice que existen diferentes grados de dependencia entre las variables aleatorias.

Si la relación entre dos variables aleatorias X y Y se puede expresar mediante una relación matemática exacta, como por ejemplo $Y = 2X + 2$, entonces la dependencia se llama *funcional*. En caso opuesto la manera de expresar la dependencia entre variables se complica un poco.

Estas dependencias que parecen no tener una expresión matemática exacta entre variables aleatorias se pueden expresar mediante la covarianza. La covarianza se usa ya que contiene información no solo de la dependencia entre las variables aleatorias (correlación), si no que además contiene información acerca de la dispersión de las variables al rededor de sus valores medios.

Pero, en dado caso de que la dispersión al rededor de sus valores promedio fuese muy pequeña, entonces la covarianza C_{XY} también lo sería para cualquier grado de dependencia, no importando si es mucho o poco. Esto hace difícil medir el grado de dependencia entre cada una de las variables. Para ello se hace uso de la *covarianza normalizada* [Peyton Z. Peebles, 2001, p. 121] o también conocido como *coeficiente de correlación* [Dolecek, 2013, p. 182], el cual se denota por ρ_{XY} y se define como

$$\rho_{XY} = \frac{\mu_{11}}{\sqrt{\mu_{20}\mu_{02}}} = \frac{C_{XY}}{\sqrt{\sigma_X^2\sigma_Y^2}} \quad (\text{B.64})$$

De esta manera el coeficiente de correlación puede tomar valores $-1 \leq \rho \leq 1$. Además de que se puede concluir de que la dependencia es una condición con mayor peso que la correlación.

Esto significa que si unas variables son independientes es condición suficiente para asegurar que que también son no correlacionadas. Pero si dos variables no están correlacionadas, no necesariamente significa que son independientes. Un caso especial sucede con las variables aleatorias *Gaussianas* las cuales si no son correlacionadas, entonces se sabe que también son independientes.

B.3. Variables aleatorias Gaussianas

Las variables aleatorias Gaussianas son muy importantes porque aparecen casi en todas las áreas de la ciencia y la ingeniería.

Para dos variables aleatorias X y Y se dice que son variables aleatorias Gaussianas si su función de densidad conjunta es de la forma [Peyton Z. Peebles, 2001, p. 124-126]:

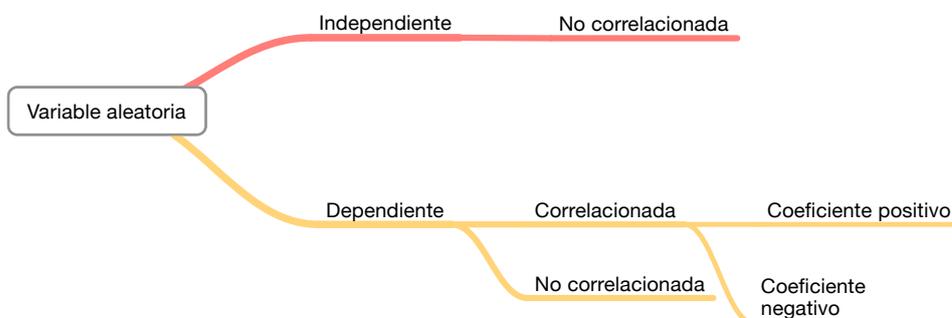


Figura B.4: Representación gráfica de la correlación y la independencia

$$f_{XY}(x, y) = \frac{\sqrt{1 - \rho_{XY}^2}}{2\pi\sigma_X\sigma_Y} \exp \left\{ \frac{-1}{2(1 - \rho_{XY}^2)} \left[\frac{(x - \bar{X})^2}{\sigma_X^2} - \frac{2\rho_{XY}(x - \bar{X})(y - \bar{Y})}{\sigma_X\sigma_Y} + \frac{(y - \bar{Y})^2}{\sigma_Y^2} \right] \right\} \quad (\text{B.65})$$

La cual a veces es conocida también como densidad Gaussiana bivalente. En donde

$$\begin{aligned} \bar{X} &= E\{X\} \\ \bar{Y} &= E\{Y\} \\ \sigma_X^2 &= E\{(X - \bar{X})^2\} \\ \sigma_Y^2 &= E\{(Y - \bar{Y})^2\} \\ \rho_{XY} &= \frac{C_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} \end{aligned}$$

Inspeccionando B.65 se puede observar que la función de densidad se puede caracterizar prácticamente con los valores medios y varianzas de las variables, además el valor máximo está dado por el coeficiente $\frac{\sqrt{1 - \rho_{XY}^2}}{2\pi\sigma_X\sigma_Y}$. Si $\rho = 0$, lo que corresponde a variables no correlacionadas, por lo tanto B.65 puede ser escrita de la siguiente manera

$$f_{XY}(x, y) = f_X(x) f_Y(y) \quad (\text{B.66})$$

Lo cual es suficiente para garantizar que X y Y son estadísticamente independientes. Entonces se puede concluir que cualquier variable Gaussiana *no correlacionada* es también *independiente* [Peyton Z. Peebles, 2001, p. 124-126]. Y finalmente quedar completamente caracterizado el fenómeno aleatorio por los valores medios y varianzas de cada una de las variables aleatorias.

B.3.1. Transformaciones lineales

Al igual que con cualquier tipo de variable, sobre las variables Gaussianas se pueden aplicar una serie de transformaciones, lineales o no lineales, para obtener una nueva variable o conjunto de variables. Pero las transformaciones lineales sobre las variables Gaussianas son algo un poco más especial.

Tranformando linealmente un conjunto de variables aleatorias Gaussianas X_1, X_2, \dots, X_N , se pueden obtener un conjunto de nuevas variables Y_1, Y_2, \dots, Y_N de la siguiente manera

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1N}X_N \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2N}X_N \\ &\vdots \\ Y_N &= a_{N1}X_1 + a_{N2}X_2 + \dots + a_{NN}X_N \end{aligned} \tag{B.67}$$

donde los coeficientes a_{ij} para $i = j = 1, 2, \dots, N$ son números reales. Ahora se puede definir la relación mediante una matriz que agrupa a todos los coeficientes

$$\mathbf{T} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{bmatrix} \tag{B.68}$$

A dicha matriz se le conoce como *matriz de transformación* y establece las relaciones entre el conjunto de variables aleatorias $\mathbf{X}^T = [X_1, X_2, \dots, X_N]$ con $\mathbf{Y}^T = [Y_1, Y_2, \dots, Y_N]$

$$\begin{aligned} \mathbf{Y} &= \mathbf{TX} \\ \mathbf{X} &= \mathbf{T}^{-1}\mathbf{Y} \\ [\mathbf{Y} - \tilde{\mathbf{Y}}] &= \mathbf{T}[\mathbf{X} - \tilde{\mathbf{X}}] \\ [\mathbf{X} - \tilde{\mathbf{X}}] &= \mathbf{T}^{-1}[\mathbf{Y} - \tilde{\mathbf{Y}}] \\ \mathbf{C}_Y &= \mathbf{TC}_X\mathbf{T}^{-1} \end{aligned}$$

Finalmente se puede establecer que las nuevas variables aleatorias $\mathbf{Y}^T = [Y_1, Y_2, \dots, Y_N]$ son variables aleatorias Gaussianas debido a que su función de densidad tienen la forma de una función de densidad Gaussiana de N variables aleatorias.

En conclusión, esto muestra que una *transformación lineal* de una variable aleatoria Gaussiana produce otra variable aleatoria Gaussiana [Peyton Z. Peebles, 2001, p. 130-132].

B.4. Procesos aleatorios

En el mundo real en las ciencias y en la ingeniería, es necesario trabajar frecuentemente con formas de onda, de las cuales en muchos casos prácticos se tratan de ondas aleatorias que inclusive llegan a cambiar con el tiempo.

Por otra parte, una señal normalmente se encuentra acompañada por una señal aleatoria que no deseamos que se presente, ruido. El ruido interfiere con la información que tiene la señal e inclusive llega a limitar el desempeño de los sistemas involucrados. Por eso es importante poder describir de alguna manera aquellas señales aleatorias para así poder analizarlas y en dado caso disminuir la interferencia con la información e inclusive tratar de eliminarlas.

Para ello es necesario introducir el concepto de *proceso aleatorio*. La idea de un proceso aleatorio consiste en generalizar aun más el significado de una variable aleatoria al incluir el tiempo. Dado que una variable aleatoria es, por definición, una función de los posibles resultados s_i de un experimento, ahora se convierte en una función de el tiempo y los posibles resultados s_i [Peyton Z. Peebles, 2001, p. 142].

Recordando que una variable aleatoria es el resultado de asignar una serie de ejes-x a cada uno de los resultados posibles s_i de un espacio de resultados S . Ahora si en lugar de asignarse a un sólo eje se asigna una correspondencia a funciones del tiempo $X_i(t)$, de esta manera se obtiene un proceso aleatorio. Ahora si consideramos que cada resultado s_i , es un instante de tiempo se tiene una serie de variables aleatorias $X(s_i)$ que a su vez se puede expresar como $X_i(t)$ y en conjunto conforman el proceso aleatorio $X(t, s)$, siendo $X_i(t)$ una realización en específico en un instante de tiempo t [Dolecek, 2013, p. 371].

Entonces claramente un proceso aleatorio $X(t, s)$ representa una familia o conjunto de funciones del tiempo. Entonces a cada una de las funciones se le llama realización del proceso y representa una variable aleatoria cuando el tiempo se man-

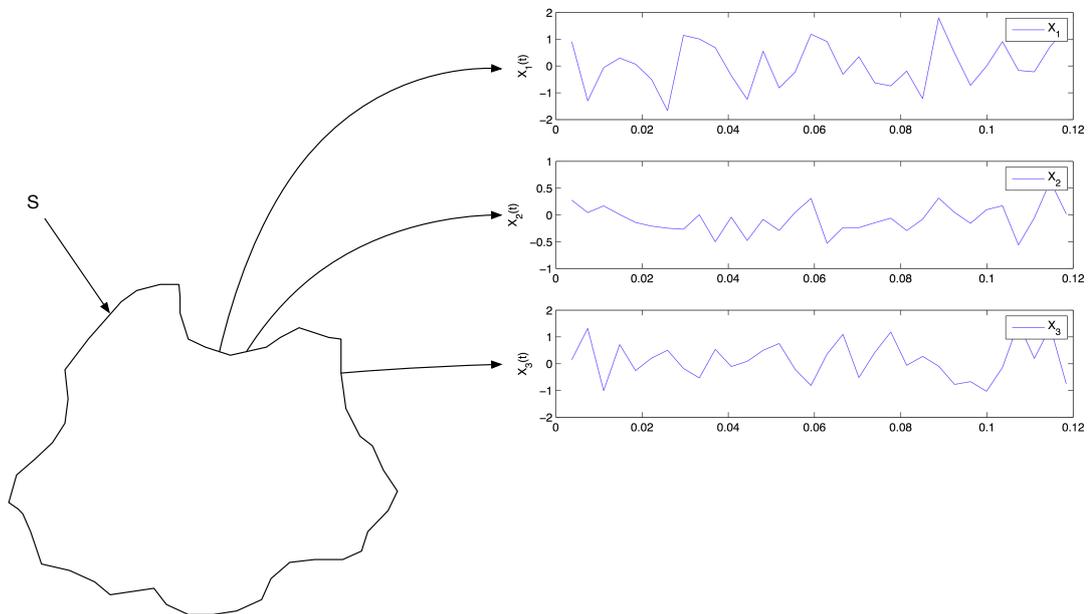


Figura B.5: Correspondencia de un espacio de eventos S hacia distintas funciones $X_i(t)$

tiene fijo, pero cuando el resultado s_i se mantiene fijo entonces solo se tiene una variable función del tiempo [Peyton Z. Peebles, 2001, p. 143].

Entendiendo que un proceso aleatorio depende tanto de s como de t , se puede simplificar la notación a $X(t)$ asumiendo que se sabe que se trata de un proceso aleatorio que depende de s .

B.4.1. Clasificación de los procesos aleatorios

Al igual que los sistemas, es conveniente clasificar a los procesos aleatorios para poder realizar un correcto análisis posteriormente. Existen distintos criterios para la clasificación de los procesos aleatorios de acuerdo con las características de la variable aleatoria $X_i(t)$ en el tiempo t . Uno de los más utilizados es a partir de la amplitud del proceso y del tiempo en el que se presenta.

Proceso aleatorio continuo. Si x es continua y t puede tener cualquier valor continuo, entonces $X(t)$ es llamado un *proceso aleatorio continuo* [Peyton Z. Peebles, 2001, p. 144-145]. (Figura B.6)

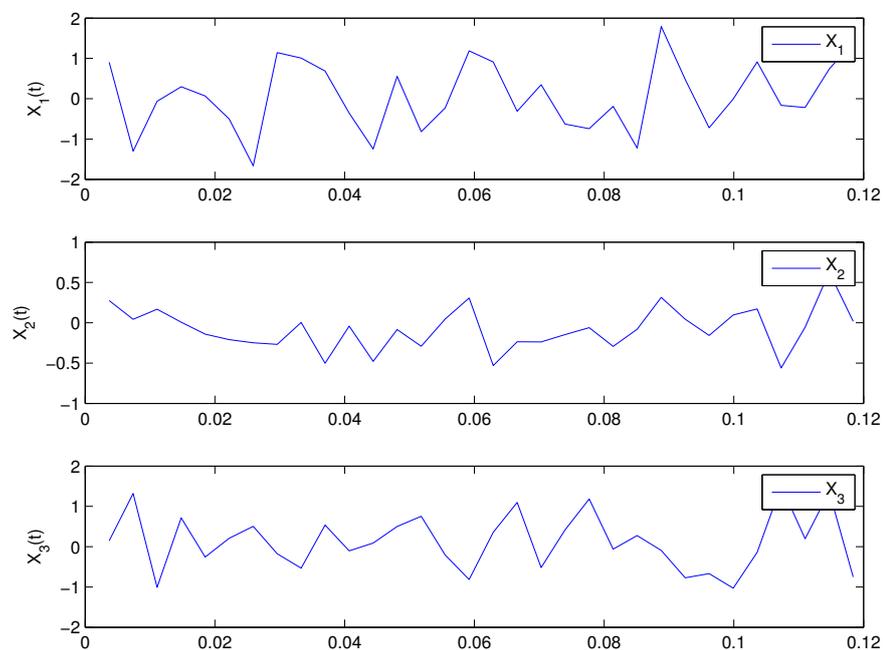


Figura B.6: Proceso aleatorio continuo

Proceso aleatorio discreto. La clase llamada *proceso aleatorio discreto* corresponde a la variable x que tiene solo valores discretos mientras t asume valores continuos [Peyton Z. Peebles, 2001, p. 144-145]. (Figura B.7)

Secuencia aleatoria continua. Un proceso aleatorio para el cual x es continua pero el tiempo asume solo valores discretos se llama *secuencia aleatoria continua* [Peyton Z. Peebles, 2001, p. 145-146].

Este tipo de secuencia puede ser obtenida mientras el muestreo periódico de un conjunto de funciones miembro o realizaciones de un proceso aleatorio continuo. (Figura B.8)

Secuencia aleatoria discreta. Otra clase de proceso aleatorio, llamado secuencia aleatoria discreta, corresponde a las variables x y t ambas son discretas [Peyton Z. Peebles, 2001, p. 146]. (Figura B.9)

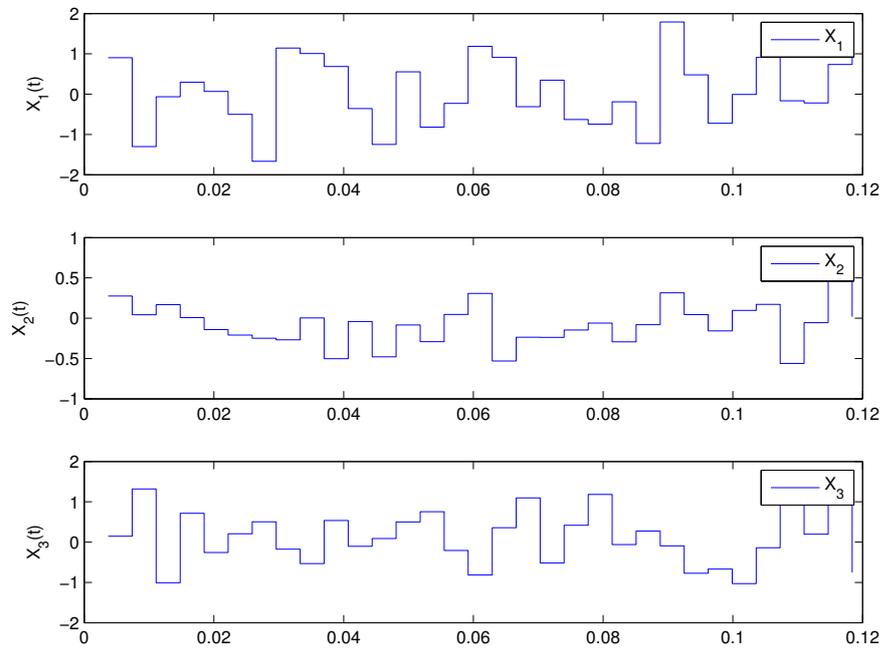


Figura B.7: Proceso aleatorio discreto

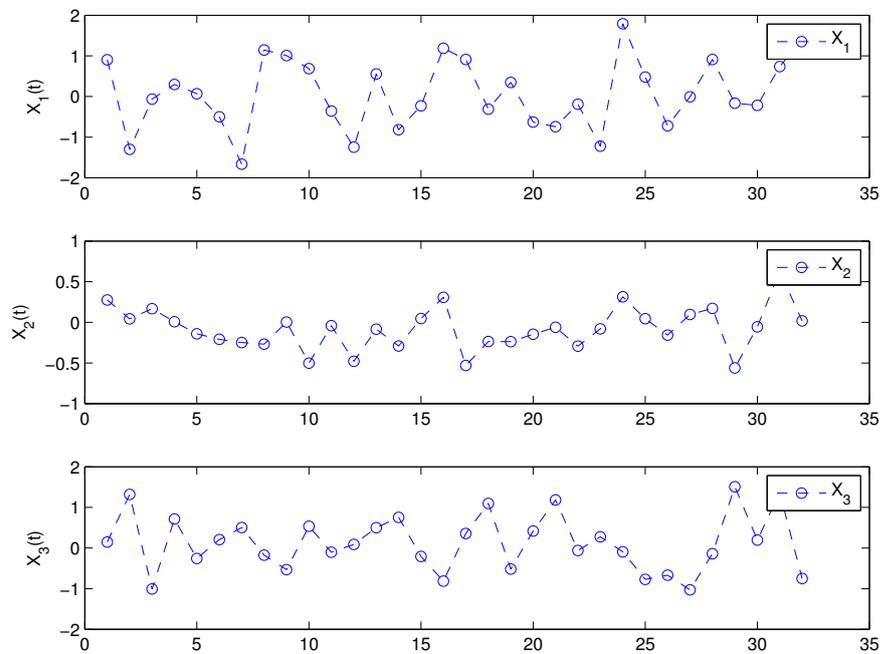


Figura B.8: Secuencia aleatoria continua

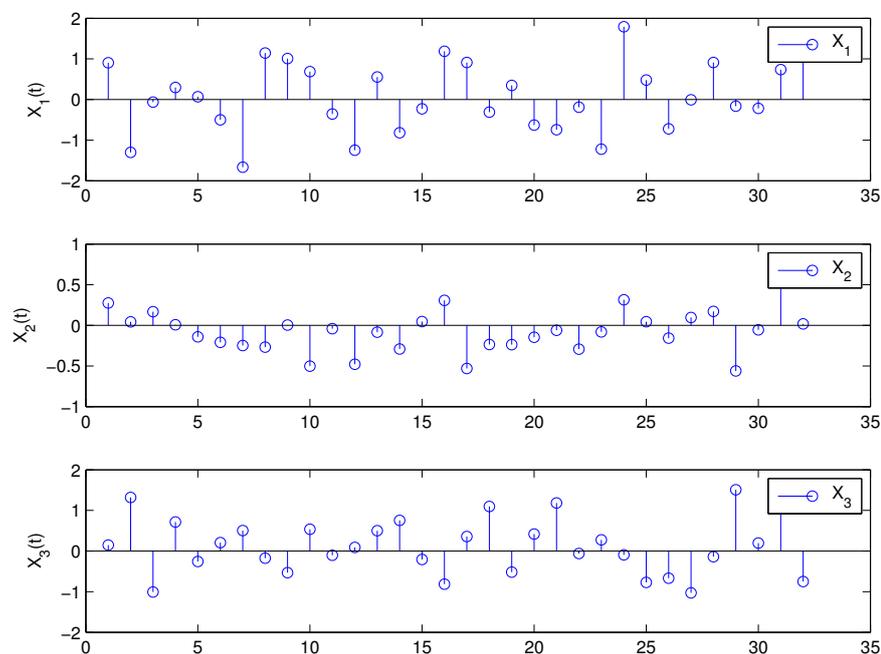


Figura B.9: Secuencia aleatoria discreta

B.4.2. Procesos Aleatorios Determinísticos y No Determinísticos

Un proceso aleatorio se dice ser determinístico siempre y cuando todos los valores de cualquier realización futura pueden ser determinados a través de sus valores pasados. Un ejemplo de ello puede ser una forma de onda senoidal, la cual posee ciertas propiedades como una amplitud conocida y una frecuencia de oscilación, y sabiendo dichos parámetros y conociendo algún valor de una realización pasada es posible determinar los valores futuros de la señal.

En caso contrario, se puede decir que para aquel proceso aleatorio no determinístico no se puede determinar con precisión sus valores futuros, tal es el caso de una señal de voltaje con un ruido aditivo, el cual gracias a la naturaleza aleatoria del ruido hace imposible determinar el valor preciso que tendrá la señal en alguna realización posterior.

B.4.3. Caracterización de los procesos aleatorios

Como ya se mencionó antes, un proceso aleatorio especificado en un instante de tiempo t , describe a una sola variable aleatoria. Partiendo de esta idea, se pueden caracterizar a los procesos aleatorios en distintos instantes de tiempo. Esto se muestra a continuación.

Descripción de un proceso aleatorio en un punto

Considerando el instante de tiempo t_1 . El proceso aleatorio descrito en dicho instante de tiempo es una variable aleatoria, digamos X_1 . Entonces la manera de caracterizar a dicha variable aleatoria es mediante una función de distribución denotada como $F_{X_1}(x_1, t_1)$ y se define como [Dolecek, 2013, Peyton Z. Peebles, 2001]:

$$F_{X_1}(x_1, t_1) = P\{X_1 \leq x_1; t_1\} \quad (\text{B.69})$$

Notese que la definición B.69 es la misma que la definición de una función de distribución de una dimensión (B.1), la única diferencia es que ahora la notación ha sido alterada para reflejar la dependencia del tiempo t_1 .

La correspondiente función de densidad se obtiene mediante la derivada de la función de distribución,

$$f_{X_1}(x_1, t_1) = \frac{\partial F_{X_1}(x_1, t_1)}{\partial x_1} \quad (\text{B.70})$$

Descripción de un proceso aleatorio en dos puntos

Ahora considerando que se desea describir al proceso aleatorio $X(t)$ en dos puntos, es decir en dos instantes de tiempo t_1 y t_2 , significa que ahora se cuentan con dos variables aleatorias, X_1 y X_2 . La forma de caracterizar se realiza ahora mediante una función de distribución conjunta de dos dimensiones, definida como [Dolecek, 2013, Peyton Z. Peebles, 2001]:

$$F_{X_1 X_2}(x_1, x_2; t_1 t_2) = P\{X_1(t_1) \leq x_1; X_2(t_2) \leq x_2\}, \quad \forall x_1, x_2 \quad (\text{B.71})$$

Nótese de nuevo que esta es la misma definición que la de una función de distribución conjunta de dos variables aleatorias (B.38) con la excepción que de se modificó la notación para añadir la dependencia de los instantes de tiempo t_1, t_2 .

Su correspondiente función de densidad está dada por

$$f_{X_1 X_2}(x_1, x_2; t_1, t_2) = \frac{\partial^2 F_{X_1 X_2}(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2} \quad (\text{B.72})$$

Descripción de un proceso aleatorio en N puntos

Finamente al observar al proceso aleatorio en N instantes de tiempo, se consiguen N variables aleatorias (X_1, X_2, \dots, X_N) . Entonces se defina la función de distribución conjunta de N dimensiones [Dolecek, 2013, Peyton Z. Peebles, 2001]

$$F_{X_1 \dots X_N}(x_1, \dots, x_N; t_1, \dots, t_N) = P\{X_1(t_1) \leq x_1; \dots; X_N(t_N) \leq x_N\} \quad (\text{B.73})$$

Y su función de densidad conjunta como:

$$f_{X_1 \dots X_N}(x_1, \dots, x_N; t_1, \dots, t_N) = \frac{\partial^N F_{X_1 \dots X_N}(x_1, \dots, x_N; t_1, \dots, t_N)}{\partial x_1 \dots \partial x_N} \quad (\text{B.74})$$

B.4.4. Procesos aleatorios estacionarios

Se sabe que una variable aleatoria tiene propiedades estadísticas, tales como valor medio, momentos, varianza, etc., las cuales están estrechamente relacionados como la función de densidad.

Considerando que un proceso aleatorio es descrito en dos instantes de tiempo, se entiende que entonces se tendrán dos variables aleatorias que lo describan, cada una con propiedades estadísticas relacionadas con la función de densidad conjunta. De manera más general si se tiene un proceso aleatorio descrito por N variables aleatorias entonces se contará con N variables cada una con sus propiedades estadísticas relacionadas a la función de densidad conjunta.

Al presentarse esta situación se puede intuir que las propiedades estadísticas de cada una de las variables pueden ser iguales o no. En el caso de dichas propiedades sean diferentes significa que las propiedades del proceso cambian con el tiempo, lo cual complica su análisis. El caso más simple es considerar que las propiedades se mantienen.

Hablando más específicamente, un proceso aleatorio se dice *estacionario* si todas sus propiedades estadísticas no cambian con el tiempo, de otra manera se le llama *no estacionario* [Peyton Z. Peebles, 2001, p. 147]. Aun así, se pueden considerar diferentes niveles o tipos de *estacionaridad*.

Procesos estacionarios de primer orden

Un proceso se dice que es estacionario de primer orden si su función de densidad de primer orden, o de una dimensión, es independiente a un desplazamiento en el tiempo Δ [Dolecek, 2013, p. 377-378], esto es:

$$f_X(x, t) = f_X(x, t + \Delta) \quad \forall t, \Delta \quad (\text{B.75})$$

Por lo tanto

$$f_X(x, t) = f_X(x) \quad (\text{B.76})$$

En general, el valor medio de un proceso depende del tiempo, esto gracias a la definición de la función de densidad, en el caso de un proceso descrito en un sólo punto su valor medio es:

$$\overline{X(t)} = E\{X(t)\} = \int_{-\infty}^{\infty} x f_X(x, t) dx \quad (\text{B.77})$$

Si el proceso descrito en un sólo punto es estacionario de primer orden, de acuerdo con B.76, el valor medio en consecuencia es una constante [Dolecek, 2013, p. 380-381].

$$\overline{X(t)} = E\{X(t)\} = \int_{-\infty}^{\infty} x f_X(x) dx \quad (\text{B.78})$$

$$= E\{X\} \quad (\text{B.79})$$

$$\overline{X(t)} = \bar{X} \quad (\text{B.80})$$

Procesos estacionarios de segundo orden

De manera similar, un proceso se dice que es estacionario de segundo orden si su función de densidad conjunta, de orden dos, tiene la propiedad de ser independiente a desplazamientos en el tiempo.

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1' + \Delta, t_2 + \Delta) \quad (\text{B.81})$$

Para todo t_1, t_2 y Δ . Nótese que en este caso, la función de densidad conjunta depende sólo de una diferencia entre los instantes de tiempo t_1 y t_2 ,

$$\tau = t_2 - t_1 \quad (\text{B.82})$$

Entonces la función de densidad puede expresarse de la siguiente manera

$$f_X(x_1, x_2; t_1, t_2) = f_X(x_1, x_2; t_1' + \Delta, t_2 + \Delta) = f_X(x_1, x_2; \tau) \quad (\text{B.83})$$

Como consecuencia de la ecuación B.83, ahora la correlación $E\{X_1 X_2\} = E\{X_1\} E\{X_2\}$ de un proceso, que en general es una función que involucra a t_1 y a t_2 , ahora será una función de τ . Más aun, considerando que las dos variables X_1 y X_2 son del mismo proceso entonces se le conoce como *función de autocorrelación* del proceso aleatorio. Y se denota como [Dolecek, 2013, Peyton Z. Peebles, 2001]:

$$R_{XX}(t_1, t_2) = E\{X_1\} E\{X(t_1 + \tau)\} = R_{XX}(\tau) \quad (\text{B.84})$$

Proceso estacionario en sentido amplio (*Wide-Sense Stationary Process*) Un tipo de estacionaridad importante es cuando un proceso se dice ser estacionario en sentido amplio. Se dice esto cuando un proceso cumple las siguientes dos condiciones [Dolecek, 2013, Peyton Z. Peebles, 2001]:

Condición 1. Valor medio contante

$$E\{X(t)\} = \bar{X} = \text{const.} \quad (\text{B.85})$$

Condición 2. La función de autocorrelación depende sólo de la diferencia de los instantes de tiempo t_1 y t_2 , para $\tau = t_2 - t_1$

$$E\{X(t_1)\} E\{X(t_1 + \tau)\} = R_{XX}(\tau) \quad (\text{B.86})$$

Entonces un proceso estacionario de orden dos es claramente estacionario en sentido amplio. Pero la afirmación inversa no necesariamente es verdad.

Si se está tratando con dos procesos aleatorios $X(t)$ y $Y(t)$, podemos decir que son estacionarios en sentido amplio, si cada uno de manera independiente satisfacen las condiciones necesarias para ser estacionarios en sentido amplio y además la *correlación cruzada* es una función que depende solamente de la diferencia del tiempo τ [Dolecek, 2013, Peyton Z. Peebles, 2001].

Condición 1. Procesos estacionarios en sentido amplio de manera individual

$$\begin{aligned} E\{X(t)\} = \bar{X} = \text{const.} & \quad E\{Y(t)\} = \bar{Y} = \text{const.} \\ E\{X(t)\} E\{X(t + \tau)\} = R_{XX}(\tau) & \quad E\{Y(t)\} E\{Y(t + \tau)\} = R_{YY}(\tau) \end{aligned} \quad (\text{B.87})$$

Condición 2. La función de correlación cruzada depende sólo de la diferencia de los instantes de tiempo t_1 y t_2 , para $\tau = t_2 - t_1$

$$E\{X(t) Y(t + \tau)\} = R_{XY}(\tau) \quad (\text{B.88})$$

Procesos estacionarios de orden N

Extendiendo la estacionaridad a N variables aleatorias X_1, X_2, \dots, X_N , se dice que es un proceso estacionario de orden N si su función de densidad conjunta de orden N es invariante a desplazamientos en el tiempo.

$$f_X(x_1, \dots, x_N; t_1, \dots, t_N) = f_X(x_1, \dots, x_N; t_1 + \Delta, \dots, t_N + \Delta) \quad (\text{B.89})$$

Para todo t_1, \dots, t_N y Δ . Estacionaridad de orden N implica estacionaridad para todos los ordenes $k \leq N$. Un proceso estacionario para todos los ordenes $N = 1, 2, \dots$, es llamado *estacionario en sentido estricto* [Peyton Z. Peebles, 2001, p. 151].

B.5. Funciones de correlación

Hablando de varias variables aleatorias se dicen que estas son independientes entre si, si una variable no tiene ninguna influencia de otra, y viceversa. En dicho caso de que llegase a existir una cierta dependencia de una variable sobre alguna otra variable distinta es necesario expresar dicho grado de relación existente mediante una expresión matemática. De esta manera se puede obtener información entre la similitud de dos señales para poder compararlas y medir un retardo entre ellas o simplemente buscar una señal conocida dentro de una señal con ruido aditivo.

Con los procesos aleatorios puede suceder lo mismo, puede llegar a existir dependencia entre uno y otro proceso, y por ende al igual que con las variables aleatorias se puede expresar dicha relación mediante alguna expresión matemática. Para conseguir lo anterior, es necesario definir algunas las funciones dedicadas a expresar esta dependencia para variables aleatorias, pero ahora para procesos aleatorios.

B.5.1. Autocorrelación

El análisis estadístico de un proceso aleatorio puede ser realizado mediante la función de *autocorrelación*, lo cual la convierte en una de las funciones más importantes dentro del ámbito de la teoría de los procesos estocásticos.

Una función de autocorrelación se define como el valor esperado de las variables aleatorias de un proceso, definidas en los instantes de tiempo t_1 y t_2 [Peyton Z. Peebles, 2001, p. 152] [Dolecek, 2013, p. 381].

$$R_{XX}(t_1, t_2) = E\{X(t_1)X(t_2)\} = \int_{-\infty}^{\infty} x_1 x_2 f_{X_1 X_2}(x_1, x_2; t_1, t_2) dx_1 dx_2 \quad (\text{B.90})$$

Como se puede observar normalmente la función de autocorrelación depende de los instantes de tiempo t_1 y t_2 . Pero si se considera por simplicidad que se trata de un proceso aleatorio estacionario en sentido amplio (*wide-sense stationary*), en consecuencia y debido a la ecuación B.83, entonces ahora la función de autocorrelación dependerá solamente de la diferencia en los instantes de tiempo.

$$\tau = t_2 - t_1 \quad (\text{B.91})$$

Esto permite reescribir la función de autocorrelación, para el caso de procesos aleatorios estacionarios en sentido amplio [Peyton Z. Peebles, 2001, p. 153] [Dolecek, 2013, p. 382], de la siguiente manera:

$$R_{XX} = E\{X(t)X(t + \tau)\} \quad (\text{B.92})$$

Propiedades de la función de autocorrelación para procesos estacionarios en sentido amplio

Una función de autocorrelación de un proceso estacionario en sentido amplio cuenta con las siguientes propiedades [Peyton Z. Peebles, 2001, p. 153] [Dolecek, 2013, p. 385-389]:

Propiedad 1. El máximo valor de la función de autocorrelación se presenta en $\tau = 0$.

$$|R_{XY}(\tau)| \leq R_{XX}(0) \quad (\text{B.93})$$

Propiedad 2. La función de autocorrelación es una función par, lo que indica que tiene simetría respecto al origen.

$$R_{XX}(\tau) = R_{XX}(-\tau) \quad (\text{B.94})$$

Propiedad 3. El máximo de la función de autocorrelación es igual al valor cuadrático medio, que es la potencia del proceso.

$$|R_{XX}(0)| = E\{X^2(t)\} \quad (\text{B.95})$$

Propiedad 4. Si $X(t)$ tiene una componente periódica, entonces $R_{XX}(\tau)$ también tendrá componente periódica con el mismo periodo.

Propiedad 5. Si $E\{X(t)\} \neq 0$ y $X(t)$ no tiene componente periódica, entonces

$$\lim_{|\tau| \rightarrow \infty} R_{XX}(\tau) = E\{X^2(t)\} \quad (\text{B.96})$$

Propiedad 6. Si $X(t)$ es ergódico, con valor medio cero y no tiene componente periódica, entonces

$$\lim_{|\tau| \rightarrow \infty} R_{XX}(\tau) = 0 \quad (\text{B.97})$$

Propiedad 7. $R_{XX}(\tau)$ no puede tener una forma arbitraria. Dado que $R_{XX}(\tau)$ está relacionada con el espectro de densidad de potencia a través de la transformada de Fourier, la forma no puede ser arbitraria.

¿Por qué es importante la función de autocorrelación?

De la definición de una función de autocorrelación, se sabe que es una caracterización estadística de un proceso aleatorio y que se obtiene al observar el proceso en dos instantes de tiempo, pero esta misma caracterización en dos puntos se puede hacer con una función de densidad conjunta de dos dimensiones en los mismos dos instantes de tiempo. La pregunta radica en saber porque es necesaria la función de autocorrelación si ya se cuenta con la función de densidad.

Considerese un proceso aleatorio $X(t)$, asumiendo que es estacionario en sentido amplio, se sabe que tiene una función de densidad de primer orden de la forma:

$$f_X(x, t) = f_X(x) \quad (\text{B.98})$$

Ahora considérese al proceso $Y(t)$ que es obtenido del proceso $X(t)$ al acelerarlo, es decir a al tener una mayor rapidez en el cambio de los valores del proceso. Entonces, es obvio que ambos procesos tendrán las mismas características de amplitud, por lo tanto ambos procesos tendrán la misma función de densidad [Dolecek, 2013, p. 383-385].

$$f_X(x) = f_Y(y) \quad (\text{B.99})$$

Ahora, a pesar de que ambos procesos cuentan con las mismas características estadísticas, descritas por la misma función de densidad, son obviamente diferentes. El proceso $X(t)$ cambia su amplitud de manera lenta según cambia el tiempo,

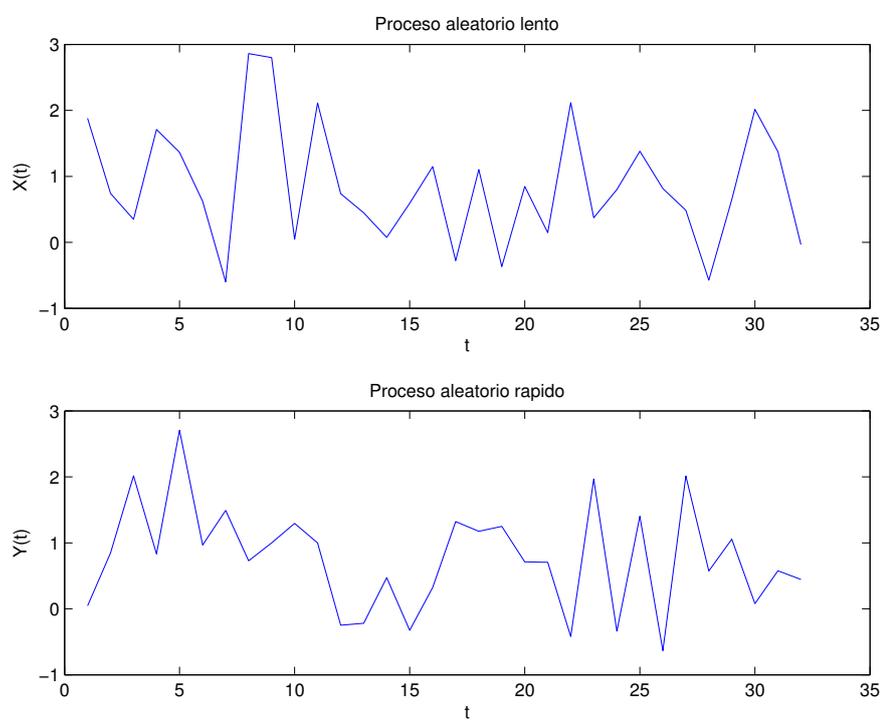


Figura B.10: Proceso aleatorio lento $X(t)$ y proceso aleatorio rápido $Y(t)$

mientras el proceso $Y(t)$ lo hace de manera más rápida en comparación con el anterior. Como resultado las amplitudes del proceso $X(t)$ serán más similares en los tiempos t_1 y t_2 que las amplitudes de $Y(t)$ en los mismos instantes. Con el fin de medir la similitud que tienen las amplitudes de ambos procesos cuando el tiempo cambia, se utiliza la función de autocorrelación.

Suponiendo que las amplitudes pueden ser positivas y negativas, se puede observar que al realizar la correlación el producto $x_1 x_2$ será positivo la mayor de las veces, esto debido a que no cambia mucho entre el instante t_1 y t_2 y no da tiempo de que x_2 alcance un valor negativo. En caso contrario habrán más valores negativos para el proceso $Y(t)$, en donde si pueden presentarse en el mismo periodo valores negativos. Por ello el valor medio de $X(t_1)X(t_2)$ será mayor que el de $Y(t_1)Y(t_2)$, debido a la falta de valores negativos.

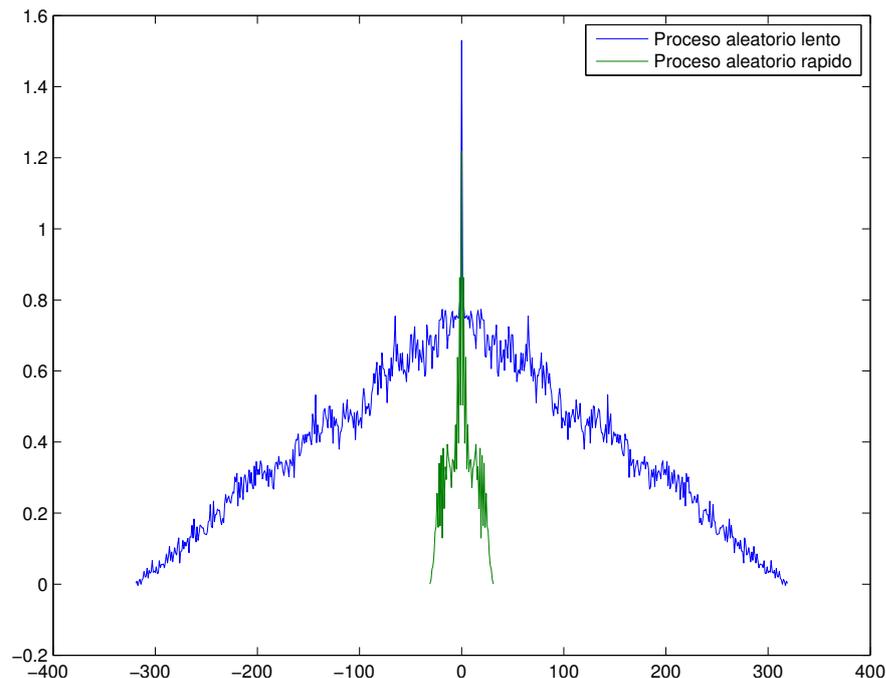


Figura B.11: Autocorrelación de procesos rápido ($X(t)$) y lento ($Y(t)$)

En este sentido, la autocorrelación brinda información acerca de si un proceso cambia rápida o lentamente, en otras palabras puede brindar información de la velocidad de cambio del proceso [Dolecek, 2013, p. 383-385].

B.5.2. Correlación

Otra función que también es de suma importancia es la *correlación* o *correlación cruzada*. Esta se utiliza cuando se trata de dos procesos aleatorios $X(t)$ y $Y(t)$ y se define de la siguiente manera

$$R_{XY}(t_1, t_2) = E\{X(t_1)Y(t_2)\} \quad (\text{B.100})$$

En donde $X(t_1)$ y $Y(t_2)$ son variables aleatorias de los procesos $X(t)$ y $Y(t)$ en los instantes de tiempo t_1 y t_2 respectivamente. Considerando entonces que la correlación depende sólo de la diferencia de los tiempos $\tau = t_1 - t_2$.

$$R_{XY}(\tau) = E\{X(t)Y(t + \tau)\} \quad (\text{B.101})$$

En dado caso de que la función de correlación sea cero, entonces se dice que los procesos son *ortogonales*

$$R_{XY}(\tau) = 0 \quad (\text{B.102})$$

Si los procesos son independientes, entonces la correlación puede escribirse de la siguiente forma

$$R_{XY}(\tau) = E\{X(t)\}E\{Y(t + \tau)\} \quad (\text{B.103})$$

Y adicionalmente si los procesos son estacionarios en sentido amplio, sus valores medios no dependen del tiempo y son constantes. Lo que facilita aun más la correlación

$$R_{XY}(\tau) = \bar{X}\bar{Y} \quad (\text{B.104})$$

En conclusión, la correlación es una constante si se trata de procesos aleatorios estacionarios en sentido amplio e independientes [Peyton Z. Peebles, 2001] [Dolecek, 2013].

Algunas de las propiedades de la función de correlación para procesos aleatorios estacionario de al menos orden dos son:

Propiedad 1. En general, una función de correlación cruzada, no es par ni impar. Se rige a partir de la siguiente relación

$$R_{XY}(-\tau) = R_{YX}(\tau) \quad (\text{B.105})$$

Propiedad 2. El límite superior de una función de correlación es la media geométrica de las correspondientes funciones de autocorrelación valuadas en el origen.

$$|R_{XY}(\tau)| \leq \sqrt{R_{XX}(0)R_{YY}(0)} \quad (\text{B.106})$$

Propiedad 3.

$$|R_{XY}(\tau)| \leq \frac{1}{2}[R_{XX}(0) + R_{YY}(0)] \quad (\text{B.107})$$

Propiedad 4. A diferencia de lo sucedido con la función de autocorrelación no hay una manera de interpretar físicamente $R_{XY}(0)$ o $R_{YX}(0)$, y puede además no ser el valor máximo de la función de correlación [Dolecek, 2013, p. 397].

B.5.3. Autocovarianza

La autocovarianza de un proceso, denotada como $C_{XX}(t_1, t_2)$, está definida como la covarianza de las variables aleatorias $X(t_1)$ y $X(t_2)$ [Peyton Z. Peebles, 2001, p. 155][Dolecek, 2013, p. 392].

$$C_{XX}(t_1, t_2) = E\{(X(t_1) - E\{X(t_1)\})(X(t_2) - E\{X(t_2)\})\} \quad (\text{B.108})$$

$$= R_{XX}(t_1, t_2) - E\{X(t_1)\}E\{X(t_2)\} \quad (\text{B.109})$$

Para $t_1 = t_2 = t$, la covarianza es igual a la varianza del proceso

$$C_{XX}(t, t) = E\{(X(t) - E\{X(t)\})^2\} = \sigma_{XX}^2(t) \quad (\text{B.110})$$

Si el proceso $X(t)$ es estacionario en sentido amplio, entonces el valor medio es constante, y la autocorrelación depende sólo de la diferencia de los instantes de tiempo τ .

$$C_{XX}(\tau) = R_{XX}(\tau) - \bar{X}^2 \quad (\text{B.111})$$

En caso de que la diferencia de tiempos $\tau = 0$ entonces la expresión se reduce a

$$C_{XX}(0) = R_{XX}(0) - \bar{X}^2 \quad (\text{B.112})$$

$$= E\{X^2\} - \bar{X}^2 \quad (\text{B.113})$$

$$= \sigma_{XX}^2 \quad (\text{B.114})$$

Lo cual es la varianza del sistema, y claramente es constante.

B.5.4. Covarianza

La covarianza, o covarianza cruzada, de dos procesos aleatorios se define como

$$C_{XY}(t_1, t_2) = E\{(X(t_1) - E\{X(t_1)\})(Y(t_2) - E\{Y(t_2)\})\} \quad (\text{B.115})$$

$$= R_{XY}(t_1, t_2) - E\{X(t_1)\}E\{Y(t_2)\} \quad (\text{B.116})$$

Si los procesos son estacionarios en sentido amplio, es decir si cumplen las condiciones establecidas en B.87 y B.88, entonces la expresión anterior puede escribirse de la siguiente manera

$$C_{XY}(\tau) = R_{XY}(\tau) - \bar{X}\bar{Y} \quad (\text{B.117})$$

En caso de que la covarianza sea igual a cero, entonces los procesos $X(t)$ y $Y(t)$ se dice que son *no correlacionados*. Entonces utilizando la expresión B.117, se puede expresar la siguiente relación

$$R_{XY}(\tau) = \bar{X}\bar{Y} \quad (\text{B.118})$$

A partir de esta expresión se puede determinar que procesos independientes son también procesos no correlacionados. Pero la afirmación inversa no siempre se puede sostener, a menos de que se trate de procesos Gaussianos [Peyton Z. Peebles, 2001, p. 156][Dolecek, 2013, p. 399].

B.6. Procesos Aleatorios discretos

Como ya se mencionó anteriormente, hay diferentes clases de procesos aleatorios, en este caso será necesario enfocarse en los procesos aleatorios discretos, más específicamente en las *secuencias aleatorias discretas*. Las cuales por su naturaleza digital pueden ser analizadas o procesadas en dispositivos digitales, tal como lo es un microprocesador o inclusive un DSP (Digital Signal Processor), esto muestreando un proceso aleatorio continuo de manera uniforme antes de realizar la etapa de procesamiento.

Al ser una secuencia aleatoria un subconjunto de la clase general de los procesos aleatorios, la teoría descrita anteriormente para caracterizar y analizar a los procesos es igualmente aplicable, para ello sólo es necesario hacer unas pequeñas consideraciones, dado que el análisis se realizará a partir de ahora con secuencias de números.

Sea el proceso aleatorio continuo $V(t)$, al ser muestreado de manera uniforme se obtiene una secuencia aleatoria discreta

$$V[n] = \{v[n], v[n-1], v[n-2], \dots, v[n-k]\}, \quad k = 0, 1, 2, \dots \quad (\text{B.119})$$

La cual expresada en una forma vectorial, se tiene el vector columna de longitud N

$$\mathbf{v} = \begin{bmatrix} v[n] \\ v[n-1] \\ \vdots \\ v[n-N] \end{bmatrix} \quad (\text{B.120})$$

B.6.1. Caracterización parcial de una secuencia aleatoria discreta

Comunmente no es posible determinar la función de densidad de probabilidad conjunta para un conjunto arbitrario de observaciones que conforman una secuencia aleatoria discreta. De acuerdo con esto, es necesario dar una caracterización parcial del proceso al especificar sus momentos de primer y segundo orden.

Considerando una secuencia aleatoria discreta, representada por $X[n]$, la cual puede ser compleja. El valor medio del proceso se define como [Haykin, 1991, p. 79]:

$$\mu[n] = E\{X[n]\} \quad (\text{B.121})$$

donde $E\{\cdot\}$ denota al operador valor esperado. Posteriormente, la definición de la función de autocorrelación de un proceso está dada por

$$r_{XX}[n, n-k] = E\{x[n]x^*[n-k]\}, \quad k = 0, \pm 1, \pm 2, \dots \quad (\text{B.122})$$

en donde $*$ denota que es el valor $x[n-k]$ conjugando. Continuando la autocovarianza del proceso se define como:

$$c_{XX}[n, n-k] = E\{(x[n] - \mu[k])(x[n-k] - \mu[n-k])^*\}, \quad k = 0, \pm 1, \pm 2, \dots \quad (\text{B.123})$$

Además de acuerdo a la expresión B.109 el valor medio, la autocorrelación y la autocovarianza están relacionados por

$$c_{XX}[n, n-k] = r_{XX}[n, n-k] - \mu[n]\mu^*[n-k] \quad (\text{B.124})$$

De esta manera, para una caracterización parcial del proceso, es necesario especificar:

1. El valor medio del proceso $\mu[n]$.
2. La función de autocorrelación $r[n, n-k]$ o la función de autocovarianza $c_{XX}[n, n-k]$ para varios valores de n y k , dependiendo del interés.

Esta caracterización ofrece dos principales ventajas, las cuales son:

1. Permite realizar mediciones prácticas del proceso.
2. La caracterización es muy adecuada para realizar operaciones lineales sobre la secuencia aleatoria discreta.

Para un proceso aleatorio que es estrictamente estacionario [Peyton Z. Peebles, 2001, p. 151], las tres cantidades anteriores asumen una forma más simple. El primer momento, el valor medio, es una constante.

$$\mu[n] = \mu, \quad \forall n \quad (\text{B.125})$$

Posteriormente, se sabe que la correlación y la autocovarianza dependen sólo de la diferencia entre los tiempos de observación n y $n-k$.

$$r_{XX}[n, n-k] = r[k] \quad (\text{B.126})$$

$$c_{XX}[n, n-k] = c[k] \quad (\text{B.127})$$

Nótese además que cuando $k=0$, correspondiente a una diferencia o desplazamiento de cero, $r_{XX}[0]$ es igual al valor cuadrático medio de $X[n]$

$$r_{XX}[0] = E\{x[n]x^*[n]\} = E\{|x[n]|^2\} \quad (\text{B.128})$$

y $c_{XX}[0]$ es igual a la varianza de $X[n]$

$$c_{XX}[0] = \sigma_{XX}^2 \quad (\text{B.129})$$

Estas condiciones anteriores no son suficientes para garantizar que la secuencia aleatoria discreta sea estrictamente estacionaria. Pero se dice que si es estacionaria en sentido amplio, o estacionaria de segundo orden. Un proceso estrictamente estacionario $X[n]$ es estacionario en sentido amplio si y sólo si

$$E\{|x[n]|^2\} \leq \infty, \quad \forall n \quad (\text{B.130})$$

Esta condición normalmente se satisface por los procesos aleatorios que se encuentran envueltos en las ciencias y la ingeniería. También es interesante hacer notar que si se trata de un proceso aleatorio Gaussiano estacionario en sentido amplio, el proceso es estrictamente estacionario también, además de quedar completamente caracterizado por las tres cantidades dadas anteriormente (μ , $r_{XX}[k]$ y $c_{XX}[k]$) [Haykin, 1991, p. 79-80].

B.6.2. Secuencias de autocorrelación y correlación cruzada

Como ya se vió antes, para caracterizar parcialmente a una secuencia aleatoria discreta es necesario conocer la función de autocorrelación. Dicha función, por definición, hace uso de la función de densidad, la cual ya se mencionó igualmente, a veces es imposible de determinar. Para ello es necesario estimar la función de autocorrelación de la siguiente manera.

Supongamos que se tiene un registro finito de una señal aleatoria $X[n]$ estacionaria en sentido amplio y de tamaño N , en donde

$$V[n] = \begin{cases} X[n], & 0 \leq n \leq N-1 \\ 0, & n < 0, n \geq N \end{cases} \quad (\text{B.131})$$

Entonces, se considera un estimador de la secuencia de autocorrelación como

$$\hat{r}_{XX}[k] = \frac{1}{N} \sum_{n=0}^{N-1} v[n]v[n-k], \quad k = 0, \pm 1, \pm 2, \dots \quad (\text{B.132})$$

Para determinar las propiedades de este estimador, y de acuerdo con la definición [B.122](#)

$$\begin{aligned} E\{\hat{r}_{XX}[k]\} &= E\left\{\frac{1}{N} \sum_{n=0}^{N-1} v[n]v[n-k]\right\} \\ &= \frac{1}{N} \sum_{n=0}^{N-1} E\{v[n]v[n-k]\} \\ &= \frac{1}{N} \sum_{n=0}^{Q-1} r_{XX}[k] \end{aligned}$$

Dado que se trata de una señal aleatoria discreta estacionaria en sentido amplio, $r_{XX}[k]$ no depende en absoluto de la variable n lo que significa que se trata de una suma de N veces el valor de $r_{XX}[k]$. Pero dado que se trata de una secuencia finita con un desplazamiento de k muestras sólo se toman en cuenta $N - k$ productos.

$$E\{\hat{r}_{XX}[k]\} = \frac{N-k}{N} r_{XX}[k] \quad (\text{B.133})$$

Observando la ecuación anterior se puede identificar a $\hat{r}_{xx}[k]$ como un estimador sesgado de $r_{xx}[k]$, ya que la media del estimador no es igual a la función de autocorrelación, pero el sesgo es pequeño si $k \ll N$. [[Oppenheim and Schafer, 2011](#), p.825-826] [[Proakis and Manolakis, 2007](#), p. 104-106]

De manera similar la secuencia de correlación cruzada queda definida de la siguiente manera [Proakis and Manolakis, 2007, p. 104-106]:

$$\hat{c}_{XY}[k] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]y[n-k], \quad k = 0, \pm 1, \pm 2, \dots \quad (\text{B.134})$$

B.6.3. Matriz de autocorrelación

Sea el vector de observación $\mathbf{x}[n]$ de longitud P .

$$\mathbf{x}[n] = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-P+1] \end{bmatrix} \quad (\text{B.135})$$

Definimos la matriz de autocorrelación de una secuencia aleatoria discreta estacionaria como el valor esperado del producto del vector de observación con el mismo [Haykin, 1991, p. 82]. Se denotará a la matriz de correlación como \mathbf{R}_{XX} de la siguiente manera

$$\mathbf{R}_{XX} = E \{ \mathbf{x}[n] \mathbf{x}^* [n] \} \quad (\text{B.136})$$

Donde $*$ denota una transposición *Hermitiana* (conjugado transpuesto). Expresada además en la forma expandida

$$\mathbf{R}_{XX} = \begin{bmatrix} r[0] & r[1] & \dots & r[P-1] \\ r[-1] & r[0] & \dots & r[P-2] \\ \vdots & \vdots & \ddots & \vdots \\ r[1-P] & r[2-P] & \dots & r[0] \end{bmatrix} \quad (\text{B.137})$$

El elemento $r[0]$ en la diagonal principal es siempre un valor real. Para datos complejos, los demás elementos pueden ser complejos.

Propiedades de la matriz de autocorrelación

La matriz de autocorrelación juega un rol importantísimo en el análisis estadístico, además en el diseño de filtros discretos en el tiempo.

La matriz de autocorrelación tiene las siguientes propiedades [Haykin, 1991, p. 82-87]:

Propiedad 1. La matriz de autocorrelación de un proceso estacionario discreto es *Hermitiana*. Se dice que una matriz, con elementos complejos, es Hermitiana si es igual a su conjugado transpuesto.

$$\mathbf{R}^* = \mathbf{R} \quad (\text{B.138})$$

Donde \mathbf{R}^* es el conjugado transpuesto de \mathbf{R} . Se puede proponer otra manera de exponer la propiedad Hermitiana, de la matriz de autocorrelación, al escribir

$$r[-k] = r^*[k] \quad (\text{B.139})$$

donde $r[k]$ es el valor de la función de autocorrelación con un desplazamiento de k .

De acuerdo con ello, para una secuencia aleatoria discreta estacionaria en sentido amplio, sólo se necesitan P valores de la función de autocorrelación $r[k]$ para $k = 0, 1, \dots, P - 1$ para definir completamente la matriz de autocorrelación \mathbf{R} , entonces se dice que la matriz es de orden P . Se puede expresar a la matriz \mathbf{R} de la forma expandida como sigue

$$\mathbf{R} = \begin{bmatrix} r[0] & r[1] & \cdots & r[P-1] \\ r^*[1] & r[0] & \cdots & r[P-2] \\ \vdots & \vdots & \ddots & \vdots \\ r^*[P-1] & r^*[P-2] & \cdots & r[0] \end{bmatrix} \quad (\text{B.140})$$

Nótese que para el caso especial de una secuencia real, la función de autocorrelación $r[k]$ es real para todo k , y la matriz de autocorrelación \mathbf{R} es simétrica.

Propiedad 2. La matriz de autocorrelación de un proceso aleatorio estacionario es *Teopltiz*.

Se dice que una matriz cuadrada es Teopltiz si todos los elementos en su diagonal principal son iguales, y si los elementos en cualquier otra diagonal paralela a la diagonal principal son de la misma manera iguales.

Es importante reconocer que si una secuencia aleatoria discreta es estacionaria en sentido amplio, en consecuencia su matriz de autocorrelación \mathbf{R} debe de ser Teopltiz, y viceversa, si la matriz de autocorrelación \mathbf{R} es Teopltiz,

entonces la secuencia aleatoria discreta debe de ser estacionaria en sentido amplio.

Propiedad 3. La matriz de autocorrelación de una secuencia aleatoria discreta es siempre positiva semidefinida y casi siempre positiva definida.

Si la forma cuadrática satisface la siguiente condición

$$\mathbf{x}^* \mathbf{R} \mathbf{x}, \quad \mathbf{x} \neq \mathbf{0} \quad (\text{B.141})$$

Se dice entonces que la matriz de autocorrelación \mathbf{R} es positiva definida. Esta condición se satisface para una proceso estacionario en sentido amplio, a menos de que hayan dependencias entre las variables aleatorias que constituyen los P elementos de el vector de observación $\mathbf{x}[n]$.

El hecho de que la matriz de autocorrelación sea positiva definida implica que el determinante y todos los menores principales son mayores que cero. Estas condiciones implican que la matriz de autocorrelación no es singular. Se dice que una matriz no es singular si su inversa existe. De acuerdo con esto, se puede afirmar que la matriz de autocorrelación siempre es no singular, es decir siempre existe su inversa \mathbf{R}^{-1} .

Propiedad 4. Cuando los elementos que constituyen al vector de observación $\mathbf{x}[n]$ son reacomodados de manera inversa, los efectos son equivalentes a transponer la matriz de autocorrelación.

Propiedad 5. Las matrices de autocorrelación de orden P (R_P) y de orden $P + 1$ (R_{P+1}) de una secuencia aleatoria discreta estacionaria, que pertenecen a P y $P + 1$ observaciones respectivamente. Están relacionadas por

$$\mathbf{R}_{P+1} = \left[\begin{array}{c|c} r[0] & r^* \\ \hline r & \mathbf{R}_M \end{array} \right] \quad (\text{B.142})$$

o de manera equivalente

$$\mathbf{R}_{P+1} = \left[\begin{array}{c|c} \mathbf{R}_M & r_p \\ \hline r_p^* & r[0] \end{array} \right] \quad (\text{B.143})$$

en donde

$$r^T = [r^*[1] \quad r^*[2] \quad \cdots \quad r^*[P]] \quad (\text{B.144})$$

y

$$r_p^T = [r[P] \quad r[P-1] \quad \cdots \quad r[1]] \quad (\text{B.145})$$

Para probar la relación se expresa la matriz de autocorrelación de orden $P+1$ en su forma expandida

$$\mathbf{R}_{P+1} = \left[\begin{array}{cccc|c} r[0] & r[1] & \cdots & r[P-1] & r[P] \\ r^*[1] & r[0] & \cdots & r[P-2] & r[P-1] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r^*[P-1] & r^*[P-2] & \cdots & r[0] & r[1] \\ \hline r^*[P] & r^*[P-1] & \cdots & r^*[1] & r[0] \end{array} \right] \quad (\text{B.146})$$

Variables de Estado

Un sistema de control moderno puede tener muchas entradas y muchas salidas, y pueden estar relacionadas de manera complicada. Los métodos en el espacio de estados para el análisis y modelado de sistemas de control son más adecuados para tratar sistemas con varias entradas y varias salidas que requiere que sean óptimos en algún sentido [Ogata, 2013, p. 293][Hendricks et al., 2008, p. 9].

C.1. Modelado en variables de estado

Este método se basa en la descripción del sistema en términos de n ecuaciones en diferencias o diferenciales de primer orden, que pueden combinarse en una ecuación matricial en diferencias o diferencial. Esta notación matricial tiene como fin simplificar la representación y cálculos matemático que se puedan realizar sobre las ecuaciones.

C.1.1. Definiciones

Algunas definiciones importantes de los conceptos utilizados para este método se describen a continuación:

Estado. El estado de un sistema dinámico es el conjunto más pequeño de variables (variables de estado) tales que el conocimiento de dichas variables en $t = t_0$, junto con el conocimiento de la entrada para $t \geq 0$, determinan por completo el comportamiento del sistema para cualquier tiempo $t \geq 0$. El concepto de

estado de ninguna manera está limitado a sistemas físicos, también se aplica en sistemas biológicos, sistemas económicos, sistemas sociales y otros.

VARIABLES DE ESTADO. Las variables de estado de un sistema dinámico son las que conforman el conjunto más pequeño de variables que determinan el estado del sistema dinámico. Si para describir en su totalidad el comportamiento de un sistema dinámico se requiere de por lo menos n variables x_1, x_2, \dots, x_n (de tal forma que una vez dada la entrada para $t \geq t_0$ y el estado inicial $t = t_0$, el estado futuro del sistema queda completamente determinado), entonces dichas n variables se consideran un conjunto de variables de estado.

VECTOR DE ESTADO. Si se necesitan n variables de estado para describir completamente el comportamiento de un sistema dado, entonces estas n variables de estado se pueden considerar como las n componentes de un vector \mathbf{x} . Dicho vector se conoce como vector de estado. Un vector de estado es un vector que determina en forma única el estado $\mathbf{x}(t)$ del sistema para cualquier tiempo $t \geq t_0$, una vez dado el estado en $t = t_0$ y especificada la entrada $\mathbf{u}(t)$ para $t \geq t_0$.

ESPACIO DE ESTADOS. El espacio de n dimensiones cuyos ejes coordenados están formados por el eje x_1, x_2, \dots, x_n se conoce como *espacio de estado*. Cualquier estado puede representarse por un punto dentro del espacio de estado.

ECUACIONES EN EL ESPACIO DE ESTADOS. En el análisis en el espacio de estados se trata con tres tipos de variables que están involucradas en el modelado de sistemas dinámicos:

- Variables de entrada
- Variables de salida
- Variables de estado

Se sabe también que la representación en el espacio de estados para un sistema dado, no es única.

Para sistemas, ya sea lineales o no lineales en tiempo discreto variantes en el tiempo, la ecuación de estado se puede escribir como:

$$\mathbf{x}[k+1] = \mathbf{f}[\mathbf{x}[k], \mathbf{u}[k], k] \quad (\text{C.1})$$

y la ecuación de salida como:

$$\mathbf{y}[k] = \mathbf{g}[\mathbf{x}[k], \mathbf{u}[k], k] \quad (\text{C.2})$$

Para el caso especial de los sistemas lineales en tiempo discreto variantes en el tiempo, la ecuación de estado y la ecuación de salida se pueden simplificar a

$$\mathbf{x}[k+1] = \mathbf{A}[k]\mathbf{x}[k] + \mathbf{B}[k]\mathbf{u}[k] \quad (\text{C.3})$$

$$\mathbf{y}[k] = \mathbf{C}[k]\mathbf{x}[k] + \mathbf{D}[k]\mathbf{u}[k] \quad (\text{C.4})$$

en donde:

$\mathbf{x}[k]$ Vector de estado de tamaño n .

$\mathbf{y}[k]$ Vector de salida de tamaño m .

$\mathbf{u}[k]$ Vector de entrada de tamaño r .

$\mathbf{A}[k]$ Matriz de estado de tamaño $n \times n$.

$\mathbf{B}[k]$ Matriz de entrada de tamaño $n \times r$.

$\mathbf{C}[k]$ Matriz de salida de tamaño $m \times n$.

$\mathbf{D}[k]$ Matriz de transmisión directa de tamaño $m \times r$.

La presencia de la variable k en las matrices $\mathbf{A}[k]$, $\mathbf{B}[k]$, $\mathbf{C}[k]$ y $\mathbf{D}[k]$ implica que estas matrices varían con el tiempo. Para el caso de que se trate con un sistema invariante en el tiempo, estas matrices permanecen constante por lo que las dos últimas ecuaciones se pueden simplificar como:

$$\mathbf{x}[k+1] = \mathbf{A}\mathbf{x}[k] + \mathbf{B}\mathbf{u}[k] \quad (\text{C.5})$$

$$\mathbf{y}[k] = \mathbf{C}\mathbf{x}[k] + \mathbf{D}\mathbf{u}[k] \quad (\text{C.6})$$

C.2. Representaciones en el espacio de estados de sistemas en tiempo discreto

Existen muchas técnicas para obtener representaciones en el espacio de estados correspondientes a sistemas en tiempo discreto. Considérese el sistema en tiempo discreto descrito por la ecuación en diferencias:

$$y[k] + a_1 y[k-1] + \dots + a_n y[k-n] = b_0 u[k] + b_1 u[k-1] + \dots + b_n u[k-n] \quad (\text{C.7})$$

donde $u[k]$ es la entrada y $y[k]$ es la salida del sistema en el instante de muestreo k . La ecuación en diferencias puede escribirse de la siguiente manera:

$$\frac{Y[z]}{U[z]} = \frac{b_0 + b_1 z^{-1} + \dots + b_n z^{-n}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}} \quad (\text{C.8})$$

Existen muchas formas de llevar a cabo representaciones en el espacio de estados para el sistema en tiempo discreto descrito por C.7 o por C.8. Algunas de ellas son [Ogata, 2013, p. 293-300][Hendricks et al., 2008, p. 21]:

1. Forma canónica controlable
2. Forma canónica observable
3. Forma canónica diagonal
4. Forma canónica de Jordan

C.2.1. Forma canónica controlable

La representación en el espacio de estados del sistema en tiempo discreto obtenido por C.7 o por C.8 se puede expresar en la forma dada por las ecuaciones siguientes:

$$\begin{bmatrix} x_1[k+1] \\ x_2[k+1] \\ \vdots \\ x_{n-1}[k+1] \\ x_n[k+1] \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_{n-1}[k] \\ x_n[k] \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u[k] \quad (\text{C.9})$$

$$y[k] = [b_n - a_n b_0 \quad b_{n-1} - a_{n-1} b_0 \quad \cdots \quad b_1 - a_1 b_0] \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_n[k] \end{bmatrix} + b_0 u[k] \quad (\text{C.10})$$

Las ecuaciones C.9 y C.10 son las ecuaciones de estado y salida, respectivamente. A esta representación se le conoce comúnmente como *forma canónica controlable*.

C.2.2. Forma canónica observable

La representación en el espacio de estados del sistema en tiempo discreto obtenido por C.7 o por C.8 se puede expresar en la forma dada por las ecuaciones siguientes:

$$\begin{bmatrix} x_1[k+1] \\ x_2[k+1] \\ \vdots \\ x_{n-1}[k+1] \\ x_n[k+1] \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & -a_n \\ 1 & 0 & \cdots & 0 & 0 & -a_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & -a_2 \\ 0 & 0 & \cdots & 0 & 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_{n-2}[k] \\ x_{n-1}[k] \\ x_n[k] \end{bmatrix} + \begin{bmatrix} b_n - a_n b_0 \\ b_{n-1} - a_{n-1} b_0 \\ \vdots \\ b_3 - a_3 b_0 \\ b_2 - a_2 b_0 \\ b_1 - a_1 b_0 \end{bmatrix} u[k] \quad (\text{C.11})$$

$$y[k] = [0 \quad 0 \quad \cdots \quad 0 \quad 1] \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_{n-1}[k] \\ x_n[k] \end{bmatrix} + b_0 u[k] \quad (\text{C.12})$$

Las ecuaciones C.11 y C.14 son las ecuaciones de estado y salida, respectivamente. A esta representación se le conoce comúnmente como *forma canónica observable*. Nótese que la matriz de estado de $n \times n$ de la ecuación C.9 es la transpuesta de la matriz C.11.

C.2.3. Forma canónica diagonal

Si los polos de la función de transferencia, p_i $i = 1, \dots, n$, dados por la ecuación C.8 son distintos, entonces la representación en el espacio de estados se puede expresar de la forma canónica diagonal como sigue:

$$\begin{bmatrix} x_1[k+1] \\ x_2[k+1] \\ \vdots \\ x_n[k+1] \end{bmatrix} = \begin{bmatrix} p_1 & 0 & \cdots & 0 \\ 0 & p_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_n \end{bmatrix} \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_n[k] \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} u[k] \quad (\text{C.13})$$

$$y[k] = [c_1 \quad c_2 \quad \cdots \quad c_n] \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_n[k] \end{bmatrix} + b_0 u[k] \quad (\text{C.14})$$

C.2.4. Forma canónica de Jordan

Si la función de transferencia dada por la ecuación C.8 incluye un polo múltiple del orden m en $z = p_1$ y todos los demás polos son distintos, entonces la ecuación de estado y la ecuación de salida se pueden expresar como sigue:

$$\begin{bmatrix} x_1[k+1] \\ x_2[k+1] \\ \vdots \\ x_m[k+1] \\ x_{m+1}[k+1] \\ \vdots \\ x_n[k+1] \end{bmatrix} = \begin{bmatrix} p_1 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & p_1 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & p_1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & p_{m+1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & p_n \end{bmatrix} \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_m[k] \\ x_{m+1}[k] \\ \vdots \\ x_n[k] \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} u[k] \quad (\text{C.15})$$

$$y[k] = [c_1 \quad c_2 \quad \cdots \quad c_n] \begin{bmatrix} x_1[k] \\ x_2[k] \\ \vdots \\ x_n[k] \end{bmatrix} + b_0 u[k] \quad (\text{C.16})$$

Bibliografía

- Gordana Javanovic Dolecek. *Random Signals and Processes Primer with MATLAB*. Springer, 2013.
- T. Doran. IEEE 1220: for practical systems engineering. *IEEE Computer Society*, pages 92 – 94, May 2006.
- Scott C. Douglas. *Introduction to Adaptive Filters*, 1999. URL <http://www.ece.mcmaster.ca/faculty/reilly/coe4tl4/adaptive%20filters%20Scot%20Douglas.PDF>.
- Simon Haykin. *Adaptive Filter Theory*. Prentice Hall, 2^o edition, 1991.
- Elbert Hendricks, Ole Jannerup, and Paul Haase Sørensen. *Linear Systems Control*. Springer, 2008.
- Vinay K. Ingle and John G. Proakis. *Digital Signal Processing using MATLAB*. CENGAGE Learning, 3^o edition, 2010.
- Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.
- Leonardo Marín, Marina Vallés, Ángel Soriano, Ángel Valera, and Pedro Albertos. Multi sensor fusion framework for indoor-outdoor localization of limited resource mobile robots. *Sensors*, 13(10):14133–14160, 2013.
- Peter S. Maybeck. *Stochastic models, estimation, and control*. Academic Press, 1979.
- J.M. Mirats Tur. Onto computing the uncertainty for the odometry pose estimate of a mobile robot. In *Emerging Technologies and Factory Automation, 2007. EFTA. IEEE Conference on*, pages 1340–1345, Sept 2007. doi: 10.1109/EFTA.2007.4416936.

BIBLIOGRAFÍA

- Chinmaya Kumar Mishra, BK Mishra, and J Sam Jebakumar. EKF based GPS/odometer data fusion for precise robot localization in outdoor environment. *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)*, 11:70 – 75, March - April 2014.
- Min Raj Nepali, Dubey Aditya Housila Prasad Ashutosh, Susheel Balasubramaniam, and EN Venkatesh. A simple integrative solution for simultaneous localization and mapping. *International Journal of Robotics and Automation (IJRA)*, 5(2):24, 2014.
- Mircea Nitulescu. Controlling a mobile robot along planned trajectories. *Journal of Control Engineering and Applied Informatics*, 7(2):18–24, 2005.
- Katsuhiko Ogata. *Sistemas de Control en Tiempo Discreto*. Pearson, March 2013.
- Alan V. Oppenheim and Ronald W. Schaffer. *Tratamiento de Señales en Tiempo Discreto*. Pearson, 3^o edition, 2011.
- Jr. Peyton Z. Peebles. *Probability, Random Variables and Random Signal Principles*. McGraw-Hill International Edition, 4^o edition, 2001.
- John G. Proakis and Dimitris G. Manolakis. *Tratamiento Digital de Señales*. Prentice Hall, 4^o edition, 2007.
- Ian Reid. Estimation II. Última visita: 2015-01-01, 2001. URL <http://www.robots.ox.ac.uk/~ian/Teaching/Estimation/LectureNotes2.pdf>.
- Erhard Schmidt. Zur theorie der linearen und nichtlinearen integralgleichungen. i. teil: Entwicklung willkürlicher funktionen nach systemen vorgeschriebener. *Mathematische Annalen*, 63:433–476, 1907. URL <http://eudml.org/doc/158296>.
- Greg Welch and Gary Bishop. An introduction to the kalman filter. *UNC-Chapel Hill*, TR 95-041, July 2006.
- Wikipedia. Sistema, sep 2014a. URL <http://es.wikipedia.org/wiki/Sistema>.
- Wikipedia. Kalman filter, 2014b. URL http://en.wikipedia.org/wiki/Kalman_filter#Details.
- G. Udny Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226(636-646):267–298, 1927. doi: 10.1098/rsta.1927.0007. URL <http://rsta.royalsocietypublishing.org/content/226/636-646/267.short>.