

Depto 2

(Abril 1989)

MODULACION SIGMA-DELTA
SOBREMUESTREADA

*Tesis Doctoral
por*

Ning He

Universidad Nacional Autónoma de México
Facultad de Ingeniería
División de Estudios de Posgrado
Apartado Postal 70-256
04510, México D.F.
MEXICO

TESIS CON
FALLA DE ORIGEN

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

CONTENIDO

PREFACIO	iv
CAPITULO 1 INTRODUCCION	1
1.1 Antecedentes y Motivación	1
1.2 Un Breve Resumen de Resultados Tradicionales y Recientes	5
1.3 Un Bosquejo de la Tesis	8
CAPITULO 2 MODELADO	10
2.1 Modulador Sigma-Delta de Lazo Simple	10
2.2 Moduladores Sigma-Delta Multilazo	13
2.3 Moduladores Sigma-Delta Multietapa	14
2.4 Modelos de Señal de Entrada	16
2.5 Configuración y Filtrado de Ruido de Cuantización	18
CAPITULO 3 ANALISIS EN EL DOMINIO DE TIEMPO	23
3.1 Cuantización Uniforme	23
3.2 Modulador Sigma-Delta de Lazo Simple	26
3.3 Moduladores Sigma-Delta Multietapa	30
3.4 Moduladores Sigma-Delta Multilazo	32
CAPITULO 4 PROPIEDADES ASINTOTICAS COMO PROMEDIOS TEMPORALES	38
4.1 Definiciones y el Método de Transformada	38
4.2 Algunos Resultados Relevantes de la Teoría Ergódica	46
4.3 El Resultado Fundamental: Entrada Constante	54
4.4 El Resultado Fundamental: Entrada Senoidal	58
4.5 Característica Espectral de la Secuencia Modulada	62

CAPITULO 5 DESEMPEÑO DE SISTEMA	64
5.1 Modulación Sigma-Delta y Cuantización Uniforme	64
5.2 Desempeño Asintótico de Sistema	67
5.3 Evaluación de Desempeño para Varios Filtros de Decodificación	70
5.4 Comparaciones	76
CAPITULO 6 DESEMPEÑO BAJO IMPERFECCIONES DE CIRCUITO	78
6.1 Motivación	78
6.2 Modelos de Imperfecciones de Circuito	80
6.3 Derivación de Desempeño	84
6.4 Una Estimación	88
CAPITULO 7 COMENTARIOS SOBRE INVESTIGACION FUTURA	91
REFERENCIAS	95

PREFACIO

Esta disertación doctoral presenta un análisis teórico de modulación sigma-delta sobremuestreada, la cual ha sido recientemente promovida como una alternativa promisorio a las técnicas convencionales para implementar convertidores analógico/digital de alta resolución con tecnología VLSI.

Aquellos tópicos de interés general son discutidos, incluyendo el principio de operación, arquitecturas, propiedades básicas tales como el comportamiento estadístico y compromisos de su desempeño; consideraciones detalladas de circuitos y de tecnología en el diseño de cuantizadores sigma-delta prácticos para aplicaciones específicas no están cubiertas aquí, excepto un tratamiento general de algunas cuestiones de robustez.

El desarrollo sigue un enfoque de "análisis exacto", y todos los resultados son rigurosamente derivados en el sentido de que ninguna de las aproximaciones ni por linealización ni el modelo de ruido aditivo blanco son usados. Consecuentemente, no existe la necesidad de recurrir a simulaciones por computadoras.

Esta selección particular de temas y de metodología refleja el interés de investigación del autor. Aunque el tratamiento es altamente matemático, las conclusiones principales son generalmente penetrantes y fáciles de interpretar. La teoría puede proporcionar información útil para la comprensión y el diseño de sistemas prácticos de cuantización sigma-delta.

Durante el desarrollo de esta tesis, el autor ha recibido asistencia de un buen número de sus amigos y colegas, y es un placer expresar aquí su aprecio por la ayuda. El autor agradece profundamente a su asesor doctoral, el profesor Federico Kuhlmann, cuyo ánimo y dirección han tenido una influencia estimulante sobre el desarrollo de esta investigación; muchos de sus comentarios penetrantes y sugerencias expertas han contribuido de manera importante al mejoramiento de la calidad y el estilo de presentación.

Reconocimiento también se dirige al doctor Chao-Ming Zeng, colega del

autor, por su colaboración productiva en el "análisis de robustez" en el capítulo 6 de esta tesis; al Profesor Andrés Buzo, por haber introducido al autor al interesante campo de cuantización, y por su entusiasta enseñanza durante su estudio de posgrado en México.

Un crédito especial se debe al profesor Robert M. Gray, quien gentilmente ha compartido con el autor muchos de sus resultados recientes de investigación sobre la materia (todos están en forma manuscrita).

Finalmente el autor desea expresar su sincera gratitud a la Universidad Nacional Autónoma de México (UNAM), particularmente la División de Estudios de Posgrado de la Facultad de Ingeniería (DEPFI), a la Comisión Federal de Electricidad (CFE) y el Instituto de Investigaciones Eléctricas (IIE), por los apoyos generosamente brindados durante este largo proyecto.

N. H.

Abril de 1989, Ciudad de México

CAPITULO 1 INTRODUCCION

1.1 Antecedentes y Motivación

Conversión analógico/digital juega un papel fundamental en sistemas de comunicaciones digitales y ha sido un tópico de extenso estudio teórico por más de 40 años. Entre las diversas técnicas para conversión analógico/digital de alta resolución, modulación sigma-delta (o delta-sigma) sobremuestreada ha sido recientemente promovida como un candidato promisorio para implementaciones con tecnologías LSI y VLSI, debido a su simplicidad y su robustez contra imperfecciones del circuito. El ejemplo más sencillo de moduladores sigma-delta es el modulador sigma-delta de lazo simple que consiste en un integrador ideal y un cuantizador binario rodeados por un lazo de retroalimentación, mostrado en las Figuras 2.1 y 2.3 en el Capítulo 2. Claramente este es un dispositivo *no lineal* ya que uno de sus componentes es un cuantizador.

Convertidores analógico/digital sobremuestreados intentan lograr cuantización de alta resolución por medio de cuantizadores uniformes de tasas bajas (usualmente binarios) dentro de los lazos de retroalimentación junto con filtros lineales. Muestreando la forma de onda analógica en la entrada muchas veces más rápidamente que la tasa de Nyquist, varios bits sucesivos del cuantizador crudo pueden producir una sola aproximación de tasa alta de las muestras de la señal original a la tasa de Nyquist. Entonces la idea básica detrás de estos sistemas es el compromiso entre la resolución del cuantizador y la precisión y rapidez de temporización: uno usa un cuantizador de tasa baja muchas veces en lugar de un cuantizador de tasa alta una vez para producir cada muestra digitalizada en la salida. Como la tolerancia de precisión del circuito requerida para implementar cuantizadores de tasas bajas se relaja significativamente y como la precisión de sobremuestreo es frecuentemente fácil de controlar, convertidores analógico/digital sobremuestreados son adecuados para tecnologías VLSI y por lo tanto, esta

técnica actualmente recibe atención creciente como una alternativa atractiva a las técnicas convencionales.

Una variedad de arquitecturas específicas para cumplir este objetivo ha sido propuesta en la literatura. Convertidores analógico/digitales generales pueden ser complejos porque pueden tener varios cuantizadores dentro de más de un lazo de retroalimentación. Cuando sólo un lazo de retroalimentación alrededor del cuantizador es usado, un sistema general de cuantización con retroalimentación puede tener filtros lineales en ambas trayectorias, directa y/o de retroalimentación. Cuando el filtrado lineal se encuentra totalmente en la trayectoria de retroalimentación, el sistema se llama codificador predictivo; si el filtrado está únicamente en la trayectoria directa, el sistema es denominado codificador interpolativo o de configuración de ruido. El ejemplo clásico de codificadores predictivos es el modulador delta, mientras los moduladores sigma-delta constituyen la clase más importante de codificadores interpolativos. La característica común de moduladores sigma-delta es que sólo se usan integradores ideales como filtros lineales. El modulador sigma-delta más sencillo tiene sólo un lazo de retroalimentación y el cuantizador tiene únicamente dos niveles de salida. Moduladores más sofisticados pueden ser construidos por medio de un cuantizador multibit dentro de más de un lazo de retroalimentación, llamado modulador sigma-delta multilazo, o bien por medio de una cascada de varios moduladores de lazo simple llamado modulador sigma-delta multietapa. En general se cree que entre más complejo sea un sistema, mejor será el desempeño global de cuantización.

Un problema persistente en las aplicaciones de los sistemas de cuantización retroalimentados ha sido la dificultad en el análisis y predicción de su comportamiento, porque la complejidad debida a la operación no lineal es empeorada por su presencia dentro de un lazo de retroalimentación. Para evitar el enfrentamiento a esta dificultad, se ha desarrollado una variedad de aproximaciones. La validez rara vez prueba analíticamente mientras muchas de tales técnicas, por experiencia y simulación, se sabe que son capaces de llegar a resultados buenos en algunos casos y pobres en otros. Por mucho la aproximación más común consiste en usar técnicas de linealización, en las cuales el cuantizador crudo es reemplazado por una fuente de ruido blanco aditivo uniforme e independiente de la señal de entrada. Como este sistema nuevo es lineal, algunas de sus propiedades pueden obtenerse fácilmente usando técnicas convencionales de sistemas

lineales. Esta aproximación conduce a, por ejemplo, desempeños globales y razón señal a ruido de cuantización, que concuerdan razonablemente bien con resultados de simulación para sistemas de modulación sigma-delta de órdenes altos, y se cree popularmente que mientras el orden (el número de lazos o etapas) sea mayor, las aproximaciones son mejores. Este tipo de análisis, sin embargo, no puede hacerse riguroso para el modulador sigma-delta de lazo simple, porque el espectro del proceso de error del cuantizador binario consiste en líneas espectrales discretas cuyas amplitudes y frecuencias dependen de la entrada, como fue recientemente reportado por Gray. Además, para algunas aplicaciones tales como audio digital, la razón señal a ruido de cuantización per se no es tan significativa para señales con espectros discretos, y, para moduladores multilazo, simulación y codificadores actuales exhiben frecuentemente comportamientos de inestabilidad, que no son predichos por el método de aproximación.

Existen buenas razones que motivan el uso de sistemas más sofisticados tales como moduladores multilazo o multietapa. Como ya hemos mencionado, el modulador de lazo simple tiene por lo menos dos deficiencias a pesar de que su estructura es bastante simple. Para implementar convertidores de alta resolución, se requiere de velocidades de operación muy altas (más precisamente, de razones de sobremuestreo grandes). Sin embargo, generalmente son deseables pequeñas razones de sobremuestreo porque para el máximo ancho de banda de una tecnología dada, menores razones de sobremuestreo permiten manejar señales de espectros más amplios con lo cual se incrementa el número de posibles aplicaciones. Otra desventaja es que el ruido en el cuantizador y la salida tienen espectros discretos, y la frecuencia y magnitud de cada una de las líneas espectrales dependen de manera complicada de la entrada del modulador. Este hecho sugiere que el modulador de lazo simple es desfavorable porque la dependencia de entrada de los espectros complica el diseño de filtros de decodificación. También la naturaleza discreta de los espectros convierte la razón señal a ruido de cuantización en una medida de desempeño subjetivamente pobre para ciertas aplicaciones tales como audio digital, donde la calidad del sonido puede ser mucho peor que aquella sugerida por la razón señal a ruido de cuantización.

La superioridad en desempeño de sistemas multilazo ha sido observada como un resultado de numerosas simulaciones y experimentos. Modulación sigma-delta multietapa es relativamente nueva, y la experimentación ha

indicado que es también una técnica muy promisoría. Contrario al modulador sigma-delta de lazo simple, estos moduladores sofisticados sí exhiben el comportamiento predicho por el análisis con fuentes de ruido blanco. Desafortunadamente, existen muy pocos estudios teóricos en la literatura, y en particular la conjetura de que el ruido en el cuantizador es blanco y uniforme no ha sido demostrada matemáticamente. Por tanto, un reto de investigación debe ser una descripción comprensiva del comportamiento complejo de tales sistemas altamente no lineales, porque las propiedades básicas pueden comprenderse más fácilmente, y la predicción del desempeño del sistema (que es una asistencia valiosa para diseñar sistemas prácticos) puede derivarse con facilidad a través del estudio cuantitativo de un modelo ideal. Más específicamente, un análisis teórico idealmente debe proporcionar respuestas rigurosas a las siguientes cuestiones fundamentales:

1. ¿Cuál es el comportamiento estadístico del ruido en el cuantizador interno, es decir, los promedios temporales tales como la media muestral, la potencia, la autocorrelación, el espectro de potencia y la distribución muestral?

2. ¿Cómo varía el desempeño, e.g., el ruido de cuantización medio cuadrático, como una función de la razón de sobremuestreo y del número de lazos o etapas en el modulador sigma-delta?

3. ¿Bajo qué condiciones proporcionan los diferentes métodos de análisis predicciones correctas sobre el comportamiento de un modulador sigma-delta? También debe ser capaz de proporcionar una visión profunda y valiosa a otros tópicos importantes tales como problemas de inestabilidad que aparecen en moduladores sigma-delta multilazo y la degradación del desempeño debida a las imperfecciones de circuito.

El objetivo de esta disertación es presentar una colección de resultados exactos sobre la clase de modulación sigma-delta multilazo que proporcionen una descripción teórica del comportamiento de estos sistemas. "Exacto" o "riguroso" significa que las ecuaciones en diferencia no-lineales que describen estos sistemas de tiempo discreto son resueltas exactamente sin linealización y sin suposiciones sobre el ruido de cuantización. En particular, estamos interesados en evaluar el desempeño global cuando se usan algunos filtros digitales simples para demodular los datos modulados y comparar estos desempeños y sus implicaciones con aquellos derivados usando aproximaciones comunes.

La base de las técnicas del análisis exacto está contenida en los artículos fundamentales de Gray sobre los moduladores sigma-delta de lazo simple. Extendemos su enfoque al caso de multilazo, el cual incluye el de lazo simple como un ejemplo especial. El análisis en el dominio de tiempo no requiere matemáticas muy sofisticadas, aunque el algebra se complica ocasionalmente con fórmulas. Muchos de los trabajos originales para desarrollar estos resultados consisten en adivinar la solución a las ecuaciones en diferencias no lineales. Una vez logrado esto, la solución puede verificarse fácilmente usando inducción matemática. Las propiedades básicas de estabilidad de estos sistemas también siguen la misma línea de razonamiento.

En el análisis asintótico, restringimos nuestra atención a formas de onda de entrada constante y senoidal. Se enfatiza la entrada constante porque es la forma de onda más simple y es una buena aproximación para señales generales que varían lentamente cuando la razón de sobremuestreo es grande. Las senoides son importantes para aplicaciones porque son comúnmente usadas en las pruebas de desempeño de sistema, pero los detalles del análisis se complican. El método de transformada o de función característica, así como algunos resultados clásicos en la teoría ergódica han demostrado su utilidad.

Nuestros resultados se enfocan principalmente a moduladores sigma-delta multilazo, aunque son inmediatamente aplicables al caso de multietapa. Los resultados para este último caso serán discutidos cuando sea conveniente, pero la presentación será algunas veces breve o simplemente omitida.

1.2 Un Breve Resumen de los Resultados Tradicionales y Recientes

Modulación sigma-delta fue propuesta por primera vez por Inose, Yasuda y Murakami (1962), e Inose y Yasuda (1963), como un medio eficiente para alcanzar cuantización de alta resolución usando sólo circuitos con componentes de baja precisión.

Muchos de los resultados originales y el entendimiento penetrante del comportamiento de moduladores sigma-delta así como la popularidad moderna de estos sistemas se deben a Candy y sus colegas. Los resultados de Candy (1974, 1985, 1986) y Candy, Ching, y Alexander (1976) y de muchos otros autores, por

ejemplo, Agrawal y Shenoi (1983), Ritchie (1977), están basados en simulación y experimentación o la aproximación de que el error del cuantizador es una sucesión independiente y uniformemente distribuida. La esperanza fue que esta suposición comúnmente usada pudiera ser justificada por el resultado clásico de que el ruido granular de cuantización puede ser modelado razonablemente bien como ruido aditivo blanco uniforme, desarrollado por Bennett (1948), y Widrow (1956). (Ver Gersho (1978) para un resumen tutorial de estos resultados clásicos.) Candy y Benjamin (1981) también usaron un modelo simple que aproxima un modulador de tiempo continuo para obtener un análisis espectral del ruido de cuantización. Su método es una extensión de Iversen (1969) para analizar sistemas de modulación delta. Candy (1986) y Candy, Ching, y Alexander (1976) también consideraron el problema de decimación y propusieron algunos filtros decodificadores simples para aquel propósito.

Una variación de la aproximación con ruido aditivo blanco es un enfoque basado en funciones descriptivas usadas en el campo de control no lineal. La función descriptiva representa la componente fundamental de la salida del cuantizador y los componentes restantes son entonces resumidos colectivamente como una fuente aditiva de ruido blanco con una ganancia que se determina en el sentido de mínimos cuadrados, así como la función descriptiva. Este método fue primero aplicado a moduladores sigma-delta por Ardalan y Paolos (1987) y Ardalan (1988).

Una variedad de arquitecturas específicas de moduladores sigma-delta se ha propuesto en la literatura. Los moduladores sigma-delta multilazo son introducidos primero por Candy (1974), y se ha observado los problemas de estabilidad cuando sólo se usan cuantizadores binarios en estos sistemas multilazo, y este comportamiento no se puede describir adecuadamente por el análisis aproximado. Moduladores sigma-delta multietapa son propuestos primero por Uchimura, Hayashi, Kimura, y Iwata (1986) y Matsuya, Uchimura, Kobayashi, y Ishikawa (1987). Sus resultados fueron principalmente experimentales y no contemplaron análisis teórico. Topologías más generales para codificadores interpolativos y predictivos fueron recientemente propuestas por Tewksbury y Hallock (1978), Lee y Sodini (1987), Giancarlo y Sodini (1986), y Scott, Lee, Giancarlo, y Sodini (1988).

Modulación sigma-delta es también aplicable a la conversión digital/analógica, como fue propuesta por Ritchie, Candy y Ninke (1974). Candy y Huynh (1986) usaron moduladores sigma-delta multilazo para obtener

mejor desempeño. Como en este caso la entrada es digital y por lo tanto sólo toma valores racionales, el comportamiento del sistema es caracterizado por ciclos límite. Algunos resultados teóricos interesantes sobre este comportamiento periódico fueron recientemente reportados por Friedman (1988).

Aunque modulación sigma-delta sobremuestreada fue propuesta hace casi 30 años, las propiedades básicas y el análisis de desempeño de estos sistemas nunca han sido derivados previamente de manera exacta hasta los trabajos recientes de Gray. En su artículo fundamental, Gray (1987), se efectuó un análisis exacto en el dominio de tiempo para un modelo de tiempo discreto (Figura 2.3) del modulador sigma-delta de lazo simple excitado por una entrada constante, y varias propiedades se derivan rigurosamente sin las suposiciones sobre el espectro o la distribución del proceso de error. Su conclusión es que si la entrada es sobremuestreada a, digamos, N veces la tasa de Nyquist, y la salida es filtrada (paso-baja) y decimada tomando una de cada N muestras, entonces la razón señal a ruido de cuantización es una buena aproximación al efecto de un cuantizador uniforme de N niveles a la tasa original de muestreo. Gray (1988) también reportó un análisis espectral usando series de Bohr-Fourier y una aplicación novedosa de la teoría ergódica para determinar el espectro del ruido de cuantizador. Gray (1988) concluyó entonces que el modelo de ruido aditivo blanco, independiente de la señal, es totalmente inconsistente, porque el verdadero proceso de error de cuantización tiene un espectro discreto que consiste en líneas espectrales cuyas amplitudes y frecuencias dependen en una manera complicada del nivel de entrada. Este análisis fue también extendido recientemente al caso de entradas senoidales por Gray, Chou, y Wong (1988). Debemos mencionar que muchos de los resultados de Gray se asemejan a los trabajos previos por Clavier, Panter, y Grieg (1947.a, b).

Este enfoque de análisis riguroso ha sido recientemente extendido por He, Kuhlmann, y Buzo (1988.a,b), y He, Buzo, y Kuhlmann (1988) a moduladores sigma-delta multilazo. Cuantizadores multinivel son usados en su análisis para garantizar la llamada "estabilidad de no-sobrecarga", una propiedad que es crucial para que el método funcione. Se establecieron condiciones necesarias y suficientes para estabilidad. Quizá sorprendentemente, la secuencia del error de cuantizador es en efecto un ruido blanco con distribución uniforme para ambas entradas, constante o senoidal (en el último caso se requiere de al menos tres lazos de retroalimentación).

Independientemente, Wong y Gray obtuvieron resultados similares a los resultados de He et.al., para moduladores sigma-delta de dos etapas, mientras que Chou, Wong, y Gray (1988) estudiaron moduladores sigma-delta multietapa generales.

Finalmente, Zeng, He, y Kuhlmann (1988) han formulado y estudiado el problema de robustez de modulación sigma-delta en la presencia de imperfecciones de los circuitos.

Todos estos resultados exactos sobre moduladores sigma-delta se han reportado muy recientemente y la mayoría de ellos aún no han sido publicados. Para una discusión tutorial completa y actualizada sobre el tema, el lector es referido a Gray (1989).

1.3 Un Bosquejo de la Tesis

Una colección de resultados teóricos exactos sobre modulación sigma-delta sobremuestreada obtenidos recientemente por el autor, se presenta aquí de manera unificada. El material se estructura en 5 capítulos (Capítulos 2-6). En el Capítulo 2 proporcionamos una descripción detallada de las diferentes arquitecturas de moduladores sigma-delta así como su efecto en la configuración del ruido de cuantización. Modelos de tiempo discreto para sistemas multilazo son enfatizados.

El análisis en el dominio de tiempo se lleva a cabo para entrada general en el Capítulo 3, y se establecen y resuelven las ecuaciones en diferencias básicas que describen los moduladores sigma-delta. Debido a la estructura recursiva del modulador sigma-delta de orden alto, muchas propiedades pueden derivarse usando argumentos simples de inducción. Se han usado cuantizadores multinivel en moduladores multilazo para evitar sobrecarga en el cuantizador, hecho que podría invalidar nuestro análisis.

Análisis asintótico es el tópico principal del Capítulo 4. Además de los espectros de potencia de varios procesos, también estudiamos otros promedios en tiempo tales como la media, la potencia, la distribución muestral, etc. La técnica básica parte de un resultado clásico de teoría ergódica sobre la distribución asintótica módulo uno de polinomios con coeficientes irracionales, obtenido por Weyl (1916), y Furstenberg (1981). Una variación

del método clásico de funciones características de Rice (1954) o el método de transformada de Davenport y Root (1958) también ha probado su utilidad. Un entendimiento de la representación por series de Bohr-Fourier para sucesiones cuasi periódicas [cf. Bohr (1947)] podría ayudar, pero esta herramienta no se usa explícitamente en nuestro desarrollo.

El Capítulo 5 se dedica al análisis del desempeño. El desempeño de un sistema sigma-delta sobremuestreado (modulador sigma-delta más filtro decodificador) se mide por el compromiso entre el error de cuantización y la razón de sobremuestreo N y el número de lazos o etapas L . Varios filtros de decodificación simples son propuestos para demodular o decodificar los datos modulados, y su error medio cuadrático asintótico es derivado rigurosamente. En algunos casos encontramos una conexión interesante entre un cuantizador sigma-delta y un cuantizador uniforme: ellos son equivalentes en el sentido de entrada-salida si el estado del modulador sigma-delta es inicializado adecuadamente en cada período de conversión de N muestras.

La robustez se estudia en el Capítulo 6. El análisis consiste en modelar algunas imperfecciones de circuito, y comparar los desempeños de los modelos ideal e imperfecto. Se demuestra entonces que pequeños cambios de ciertos parámetros de sistema resultan en pequeños cambios en el desempeño global del sistema. Aquellas imperfecciones que pueden producir mayores degradaciones de desempeño son también identificadas.

Concluimos la tesis dando algunas sugerencias para investigación futura en el Capítulo 7. Aunque toda la tesis tiene contribuciones originales, la mayor parte de ellas se concentra en los Capítulos 3-6.

CAPITULO 2 MODELADO

A través de esta tesis se estudian sólo modelos ideales de modulación sigma-delta, es decir, en nuestro análisis exacto, no consideraremos efectos tales como fuga de corriente en los integradores, desajuste de las ganancias, imprecisiones del cuantizador, imprecisiones de temporización, entradas ruidosas, ciclos límite, etc., porque estos efectos no ideales sólo oscurecen el entendimiento de los problemas fundamentales. En el Capítulo 6, sin embargo, proporcionaremos un estudio sobre el importante tópico de robustez, basado en un modulador sigma-delta de lazo simple con varios tipos de imperfecciones de circuito.

2.1 Modulador Sigma-Delta de Lazo Simple

Un modulador sigma-delta de lazo simple consiste de un integrador ideal y un cuantizador binario dentro de un lazo de retroalimentación. En la Figura 2.1 se muestra un modelo de tiempo continuo de tal sistema.

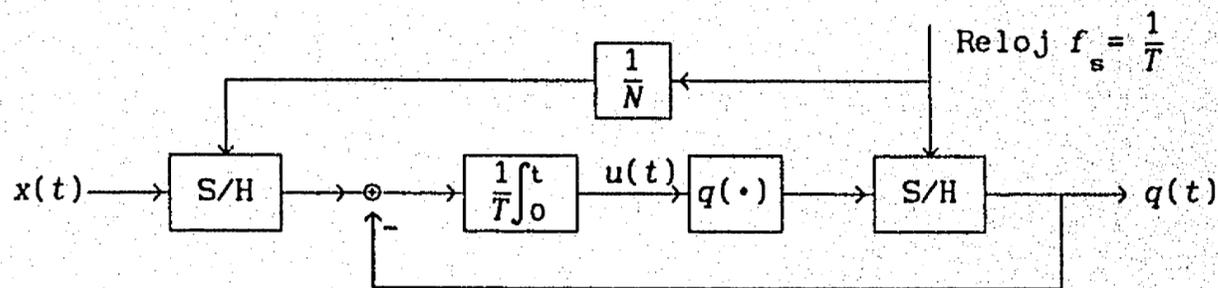


Figura 2.1 Modelo de Tiempo Continuo

Aquí S/H representa un circuito de muestreador/retenedor, y $q(\cdot)$ denota un

cuantizador binario, definido por

$$q(u) = \begin{cases} +b, & \text{si } u \geq 0, \\ -b, & \text{si } u < 0, \end{cases}$$

y esta característica de entrada y salida se ilustra en la Figura 2.2.

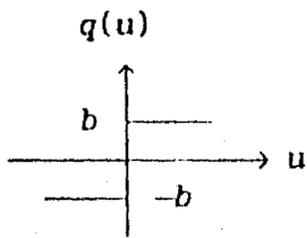


Figura 2.2 Cuantizador Binario

La salida del modulador sigma-delta $q(t)$, una secuencia de datos binarios, está sincronizada por un reloj de frecuencia $f_s = \frac{1}{T}$, la cual es N veces mayor que la de Nyquist de la señal de entrada $x(t)$. El parámetro N se denomina *razón de sobremuestreo*, cuyos valores típicos se encuentran en el rango de 100 a 1000. Más precisamente, supóngase que el conjunto de todas las entradas al sistema consiste en las señales analógicas $x(t)$ que no tienen componentes espectrales con frecuencias mayores que algún valor finito, digamos W hz; si el modulador opera a una tasa de f_s hz, entonces la razón de sobremuestreo se define por

$$N = f_s / 2W.$$

La ecuación básica que describe el estado del integrador $u(t)$ es

$$u(t) = u(0) + \frac{1}{T} \int_0^t [x(t) - q(t)] dt.$$

Un modelo más conveniente para nuestro propósito se obtiene muestreando el modelo de tiempo continuo en $t = nT$, $n = 0, 1, \dots$, y la versión de tiempo discreto o muestreada se muestra en la Figura 2.3.

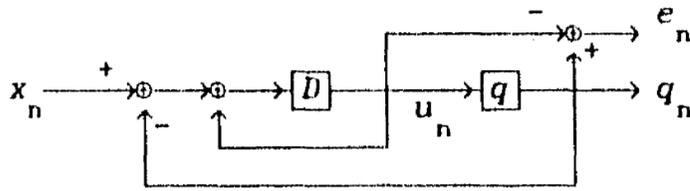


Figura 2.3 Modulador Sigma-Delta de Lazo Simple

Aquí y en las figuras subsecuentes, D denota el elemento de *retraso unitario* (su función de transferencia es z^{-1}). Las señales analógicas y de tiempo discreto están relacionadas vía las siguientes ecuaciones:

$$x_n = x(nT), \quad u_n = u(nT), \quad q_n = q(nT), \quad n = 0, 1, 2, \dots,$$

donde $T = 1/f_s$ es el periodo de muestreo. Explícitamente indicado y enfatizado también en la Figura 2.3 está la secuencia de error del cuantizador binario e_n , definida como

$$e_n = q_n - u_n.$$

El estado del integrador u_n satisface la siguiente relación recursiva

$$u_{n+1} = u_n + x_n - q(u_n), \quad n = 0, 1, 2, \dots,$$

la cual es obviamente una ecuación en diferencias no lineal.

Se muestra en la Figura 2.4 un modelo equivalente del modulador sigma-delta de lazo simple como un integrador de tiempo discreto seguido de un modulador delta, hecho que origina el nombre de "modulador *sigma-delta*".

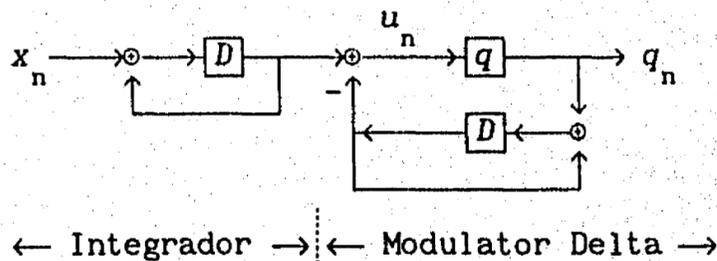


Figura 2.4 Circuito Equivalente de la Figura 2.3

Otro modelo equivalente es mostrado en la Figura 2.5, el cual podría llamarse sistema con perturbación determinística porque la entrada al cuantizador binario es la entrada al modulador más un término de error que se comporta como una señal de perturbación.

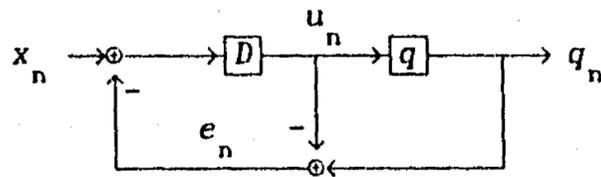


Figura 2.5 Perturbación Determinística

2.2 Moduladores Sigma-Delta Multilazo

Los moduladores sigma-delta multilazos se forman con un cuantizador uniforme (generalmente multibit) dentro de varios lazos de retroalimentación. La Figura 2.6 muestra un modulador de doble lazo con un cuantizador de 4 niveles.

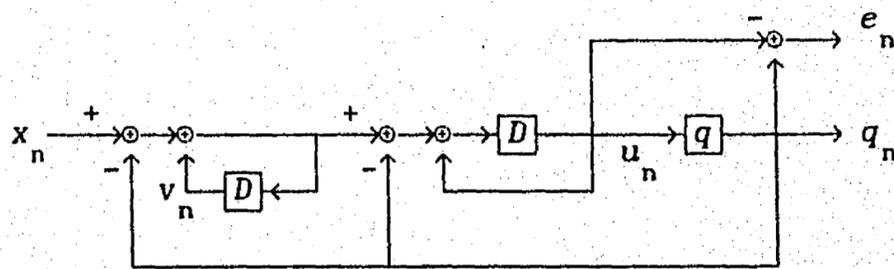


Figura 2.6 Modulador Sigma-Delta de Doble Lazos

En esta figura, $q(\cdot)$ denota un cuantizador de 4 niveles, cuya característica está dada gráficamente en la Figura 2.7.

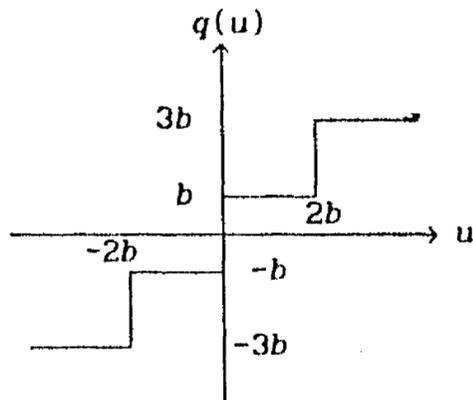


Figura 2.7 Cuantizador de 2 Bits

De la Figura 2.6 se puede escribir, por inspección, las siguientes ecuaciones en diferencias para el estado del integrador:

$$v_{n+1} = x_n + v_n - q_n, \quad n = 0, 1, 2, \dots$$

$$u_{n+1} = v_{n+1} + u_n - q_n, \quad n = 0, 1, 2, \dots,$$

donde $q_n = q(u_n)$ es la salida del cuantizador de 2 bits.

Un modulador general de L lazos puede definirse recursivamente por medio de un modulador de orden menor dentro del lazo exterior (L -ésimo) como se ilustra en la Figura 2.8. Un análisis detallado de este sistema en el dominio del tiempo se llevará a cabo en el Capítulo 3.

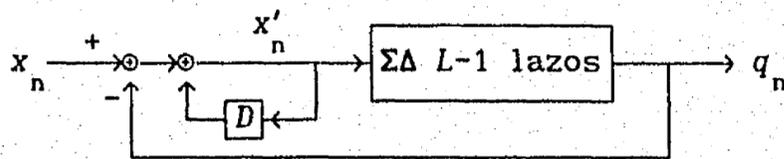


Figura 2.8 Modulador Sigma-Delta de L Lazos

2.3 Moduladores Sigma-Delta Multietapa

Modulador sigma-delta multietapa es otra generalización importante del modulador sigma-delta de lazo simple. Se puede formar como una conexión en

cascada de varios moduladores sigma-delta de lazo simple junto con una red digital lineal. En la Figura 2.9, se muestra el modelo de tiempo discreto de un modulador sigma-delta de dos etapas.

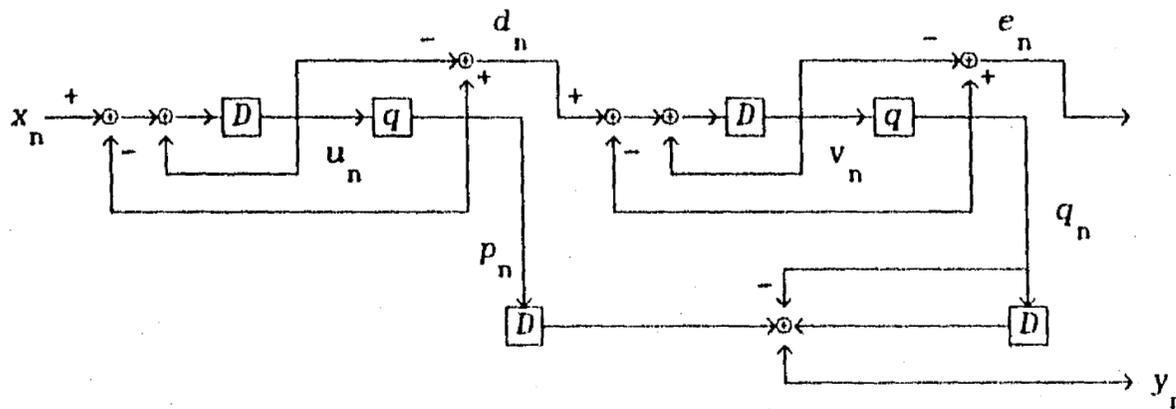


Figura 2.9 Modulador Sigma-Delta de Dos Etapas

Los estados del sistema, u_n y v_n , se describen mediante las siguientes ecuaciones:

$$u_n = u_{n-1} - q(u_{n-1}) + x_{n-1}, \quad n = 1, 2, \dots,$$

$$v_n = v_{n-1} - q(v_{n-1}) + d_{n-1}, \quad n = 1, 2, \dots,$$

donde

$$d_n = p_n - u_n = q(u_n) - u_n, \quad n = 0, 1, 2, \dots,$$

y

$$e_n = q_n - v_n = q(v_n) - v_n, \quad n = 0, 1, 2, \dots$$

son los errores de los cuantizadores binarios, en la primera y la segunda etapa, respectivamente, y $x_n \in [-b, b)$, $n = 0, 1, 2, \dots$ es la sucesión de entrada. La red lineal combinatorial está definida por

$$y_n = p_{n-1} - q_n + q_{n-1}, \quad n = 1, 2, \dots,$$

donde y_n es la sucesión de salida del modulador sigma-delta de dos etapas.

Un modulador general de L etapas puede ser construido similarmente. La sucesión de error del l -ésima etapa se conecta a la entrada de la etapa $(l+1)$ -ésima, mientras que las salidas de todas las etapas se combinan vía un circuito lineal digital para formar la salida global, como se ilustra en la

Figura 2.10, en la cual ∇^l representa un operador de diferencia hacia atrás de orden l (cuya función de transferencia es $(1-z^{-1})^l$). La salida de la red lineal de combinación está dada por

$$y_n = \sum_{l=1}^L (-1)^{l-1} \nabla^{l-1} (q_{n-L+1}^{(l)}).$$

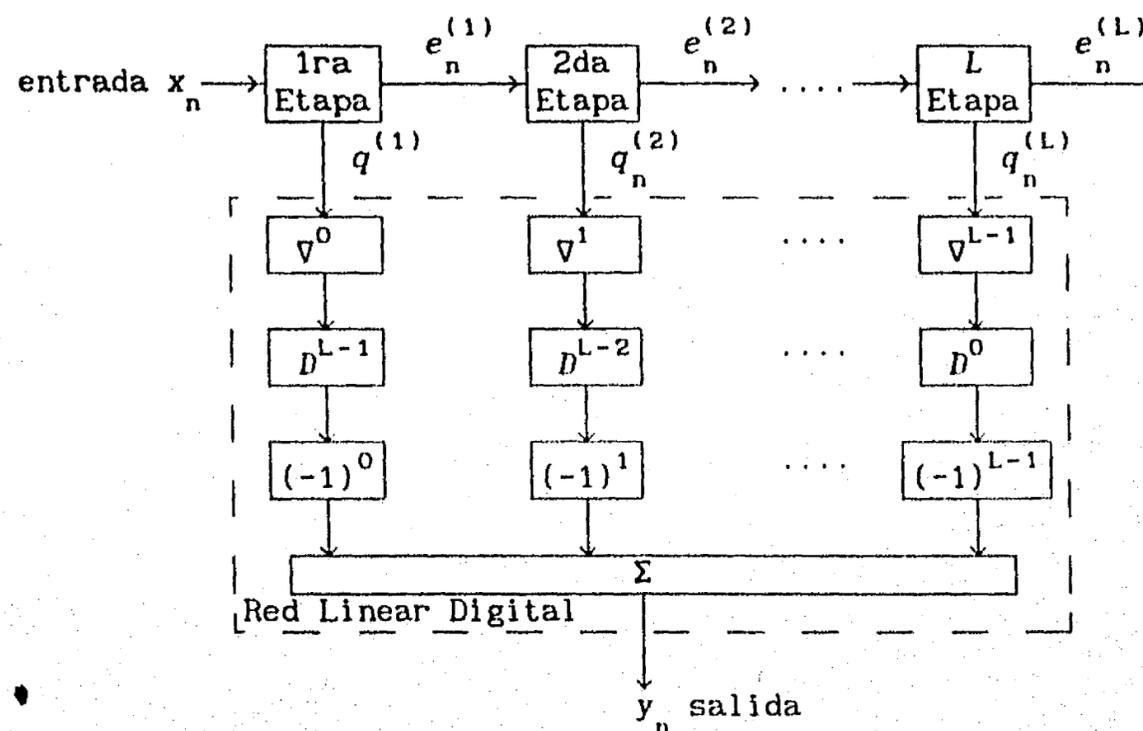


Figura 2.10 Modulador Sigma-Delta de L Etapas

2.4 Modelos de Señal de Entrada

Supondremos siempre que todas las sucesiones de entrada x_n tienen amplitud acotada, i.e.,

$$x_n \in [-b, b], \quad n = 0, 1, 2, \dots,$$

donde el intervalo $[-b, b]$ es el rango dinámico de la señal de entrada; cualquiera que tenga mayor rango dinámico debe ser escalada apropiadamente.

Para simplificar nuestro análisis asintótico, confinaremos la atención a dos modelos simples de señales determinísticas, que son sucesiones constantes

(entrada constante) y senoides puras de tiempo discreto (entrada senoidal). Mientras que las formas de onda senoidales se usan comúnmente en las pruebas de desempeño del sistema, la entrada constante representa una idealización razonable y útil de señales más generales de variación lenta. Si se usa un circuito de muestreo/retención para muestrear la señal de entrada analógica como en la Figura 1, entonces la entrada discreta se mantiene constante durante cada período de conversión. Más específicamente, hacemos la siguiente suposición.

Entrada Constante:

$$x_n = x, \quad \text{para todo } n = 0, 1, \dots,$$

donde $x \in [-b, b]$ es un múltiplo irracional de b , i.e., x/b es un número irracional. Esta suposición es razonable si x se selecciona en el intervalo $[-b, b]$ de acuerdo a una función de densidad de probabilidad, entonces la suposición se satisface con probabilidad uno.

Entrada Senoidal:

$$x_n = a \cos(2\pi n f / f_s + \theta) = a \cos(\omega n + \theta), \quad n = 0, 1, \dots,$$

donde $a \in [0, b]$, $f \in [0, W]$ y $\theta \in [0, 2\pi)$ son respectivamente la amplitud, la frecuencia y la fase de la senoide, y f_s es la frecuencia de muestreo (cf. Figura 2.1) la cual es considerablemente mayor que f (ya que la razón de sobremuestreo $N = f_s / f$ es usualmente grande). Esto implica que $\omega = 2\pi f / f_s$ es pequeña y consecuentemente x_n es una secuencia senoidal de variación lenta. Suponemos que los parámetros a , f y θ satisfacen algún tipo de relación de irracionalidad. Nuevamente, tal suposición se satisface con probabilidad uno si la terna (a, f, θ) se escoge del conjunto $[0, b] \times [0, W] \times [0, 2\pi)$ de acuerdo a una función de densidad de probabilidad.

Las suposiciones anteriores de irracionalidad permiten simplificar nuestro análisis asintótico en el Capítulo 4 donde se evalúan algunas de las propiedades estadísticas como promedios en el tiempo de varios procesos. Las entradas racionales también son de interés práctico por dos razones. Primero, cuando se simula un sistema mediante una computadora digital, todas las muestras de señales son necesariamente números racionales. Segundo, en

aplicaciones de conversión digital/analógica la señal de entrada es digital y por tanto racional. En estos casos, sin embargo, la suposición de irracionalidad aún proporciona una buena aproximación de entradas racionales porque cualquier número racional q/p (p y q son enteros relativamente primos) asemeja un número irracional si el denominador p es suficientemente grande (el cual representa el período de los ciclos límite).

2.5 Configuración y Filtrado de Ruido de Cuantización

El principio básico de operación de modulación sigma-delta puede resumirse como sigue [6]: "El circuito intenta minimizar el ruido mediante los valores cuantizados oscilando entre los niveles de manera tal, que el promedio de $q(t)$ se aproxima al promedio de $x(t)$."

Para hacer más plausible este razonamiento intuitivo, considérese el siguiente modelo de un modulador sigma-delta de lazo simple donde el cuantizador binario es reemplazado por una fuente de ruido aditivo $e_n = q_n - u_n$:

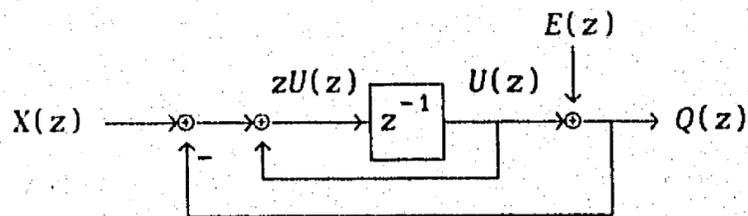


Figura 2.11 Representación en el Dominio de Transformada

donde $X(z)$, $U(z)$, $E(z)$, y $Q(z)$ denotan las transformadas z de las sucesiones correspondientes. Por inspección de la Figura 2.11, escribamos

$$zU(z) = U(z) + X(z) - Q(z),$$

y

$$Q(z) = U(z) + E(z).$$

Resolviendo $Q(z)$ en términos de $X(z)$ y $E(z)$, se sigue que

$$Q(z) = z^{-1} X(z) + (1-z^{-1}) E(z),$$

y la expresión equivalente en el dominio de tiempo es

$$q_n = x_{n-1} + \nabla(e_n),$$

donde el operador de diferencia de primer orden

$$\nabla(e_n) = e_n - e_{n-1}$$

es un filtro lineal con función de transferencia $(1-z^{-1})$.

El resultado similar para moduladores multilazo es,

$$q_n = x_{n-1} + \nabla^L(e_n),$$

donde $L = 1, 2, \dots$ es el número de lazos de retroalimentación, y $\nabla^L(\cdot)$ denota el operador de diferencia de orden L con función de transferencia $(1-z^{-1})^L$. Para demostrar esto, redibuje la Figura 2.8 como se muestra en la Figura 2.12,

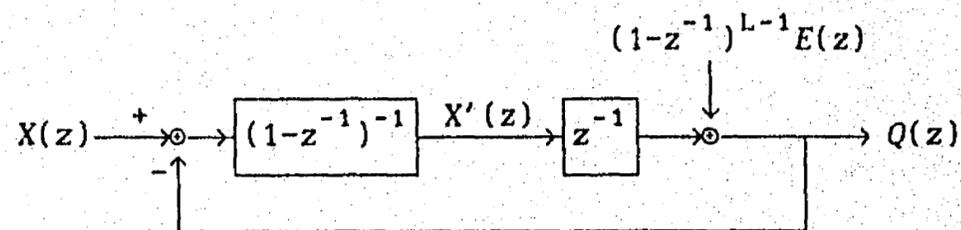


Figura 2.12 Representación en el Dominio de Transformada de Fig.2.8

donde la entrada al modulador de $L-1$ lazos situado en la trayectoria directa es

$$X'(z) = (1-z^{-1})^{-1} [X(z) - Q(z)],$$

y por una hipótesis de inducción, la relación de entrada y salida del modulador de $L-1$ lazos es

$$Q(z) = z^{-1} X'(z) + (1-z^{-1})^{L-1} E(z).$$

Resolviendo para obtener $Q(z)$ tenemos

$$Q(z) = z^{-1} X(z) + (1-z^{-1})^L E(z),$$

la cual es exactamente la equivalente a la expresión deseada en el dominio de la transformación.

Las ecuaciones anteriores tienen una interpretación intuitiva: la salida del modulador o del cuantizador q_n se expresa como la entrada retrasada x_{n-1} más un término de error que es la derivada de tiempo discreto de orden L del ruido del cuantizador e_n .

Para moduladores sigma-delta multietapas, si la red de combinación lineal se define como en la Figura 2.10, entonces se demuestra por inducción sobre L que

$$y_n = x_{n-L} + (-1)^{L-1} \nabla^L(e_n),$$

donde y_n es la salida del modulador de L etapas, $q_n^{(l)}$ denota la salida de la etapa l -ésima, $l = 1, 2, 3, \dots, L$, y e_n denota el error de cuantización de la última (L -ésima) etapa.

Para ambos moduladores sigma-delta de L lazos y L etapas, definamos el ruido global de cuantización como

$$N_n = \nabla^L(e_n).$$

Si conocemos el espectro de potencia $S_e(f)$ del error e_n , de la teoría estándar de sistemas lineales se sigue que el espectro de potencia $S_N(f)$ de N_n está dado por

$$S_N(f) = [2\sin(\pi f)]^{2L} S_e(f),$$

ya que la función de transferencia del sistema lineal $\nabla^L(\cdot)$ es

$$(1-z^{-1})^L \Big|_{z=e^{j2\pi f}} = [j2\sin(\pi f)]^L e^{-j\pi L f},$$

la cual tiene una gran atenuación en la banda de frecuencia de la entrada del modulador. Nótese que la frecuencia está normalizada de modo tal que $f \in [0, 1)$, y la banda de entrada es $[0, 1/2N) \cup [1-1/2N, 1)$, donde N es la razón de sobremuestreo. Se espera que N_n sea una señal de tipo paso-alta, i.e., la mayoría de su potencia se ubique fuera de la banda de la señal de entrada, y consecuentemente pueda ser removida fácilmente por un filtro paso-baja para obtener la señal de entrada original (ver la Figura 2.13). Este efecto de configurar el espectro del ruido de cuantización incrementa exponencialmente con el orden (i.e., el número de lazos o etapas) de los moduladores, por lo cual los moduladores de orden mayor pueden proporcionar mejores desempeños.

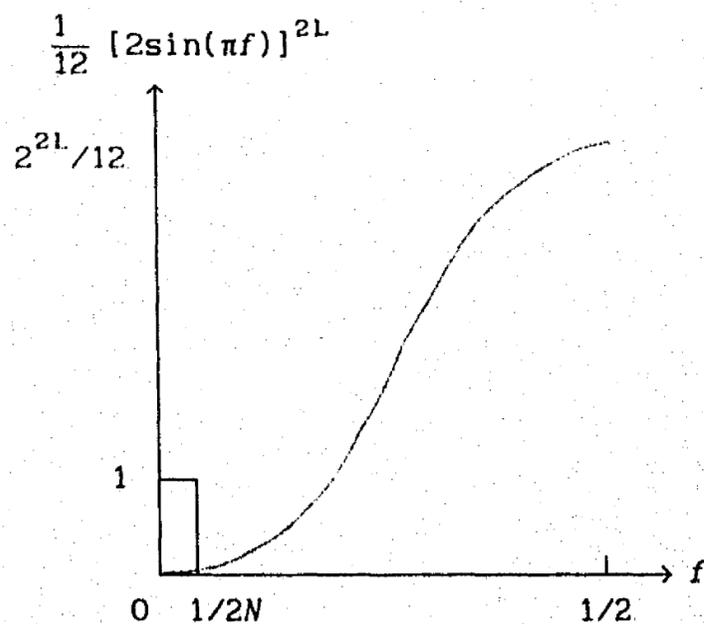


Figura 2.13 Efecto de Configurar el Ruido

Nótese que, para completar este análisis espectral, todo lo que se necesita conocer es el espectro de potencia del ruido de cuantización e_n . Sin embargo, la teoría clásica de sistemas lineales es incapaz de derivar $S_e(f)$ porque este proceso es producido por un cuantizador, dispositivo

inherentemente no lineal. En el método tradicional por aproximación, e_n es por suposición un ruido aditivo blanco, independiente de la entrada y uniformemente distribuido, i.e., $S_e(f) = b^2/3$. Esta suposición es definitivamente falsa para moduladores de un solo lazo con entradas constantes o senoidales, pero sí es correcta para moduladores multilazo o multietapa con ambos tipos de entrada, como se demostrará rigurosamente en los Capítulos 3 y 4.

La función para eliminar el ruido global de cuantización N_n de la salida del modulador puede cumplirse mediante filtros digitales paso-baja llamados filtros decodificadores. La sucesión digital producida por un modulador sigma-delta se alimenta al filtro de decodificación. Este último procesa los datos modulados para minimizar la potencia del ruido de cuantización en la reproducción, es decir, remover lo más posible la componente ruidosa de la banda de la señal original. Decimación o submuestreo es también incorporado en esta etapa de decodificación, tomando una de cada N muestras de la salida sobremuestreada para obtener la señal digitalizada a la tasa original de Nyquist. El objeto es obtener una representación compacta de la señal digitalizada, la cual es mucho más conveniente para su transmisión, almacenamiento o procesamiento posteriores. Por simplicidad, sólo los filtros lineales e invariantes en el tiempo (ambos FIR y IIR) son considerados en la literatura. Idealmente los coeficientes deben ser optimizados para proporcionar la razón señal a ruido de cuantización lo más grande posible. Este tópico de diseñar buenos decodificadores y determinar sus desempeños será tratado en el Capítulo 5.

CAPITULO 3 ANALISIS EN EL DOMINIO DE TIEMPO

Este capítulo tiene dos propósitos: desarrollar una representación útil de los errores de cuantización en los moduladores sigma-delta, usando una transformación afín de la entrada de sistema seguida de una no-linealidad sin memoria, en ausencia de sobrecarga del cuantizador, y encontrar condiciones sobre los parámetros de sistema y el rango de entrada tales que el cuantizador no se sobrecargue.

3.1 Cuantización Uniforme

El componente básico común a todos los moduladores sigma-delta es el cuantizador uniforme (binario o multinivel). Es uniforme por su sencillez; como no se dispone del conocimiento a priori sobre el comportamiento estadístico de la entrada del cuantizador o de la entrada original del sistema, no hay manera de optimizar los niveles de salida para la entrada.

Considérese un cuantizador uniforme con un número par, digamos M , de niveles de salida y un paso de cuantización Δ (la distancia entre niveles adyacentes). La salida es el nivel más cercano a la entrada; este mapeo de error mínimo puede ser expresado como

$$q(u) = \begin{cases} (M-1)\Delta/2, & (M/2-1)\Delta \leq u; \\ (k-1/2)\Delta, & (k-1)\Delta \leq u < k\Delta, \quad k = (-M/2+2), \dots, (M/2-1); \\ (-M+1)\Delta/2, & u < (-M/2+1)\Delta, \end{cases} \quad (3.1.1)$$

donde la colección de todos los niveles de salida

$$\{ (k-1/2)\Delta, \quad k = -M/2+1, \dots, M/2 \}$$

es llamada el alfabeto de reproducción del cuantizador.

Defínase el error como $e = q(u) - u$. Nótese que si la entrada u del cuantizador está limitada en el rango $[-M\Delta/2, M\Delta/2]$, entonces la magnitud del error $|e|$ está acotada por $\Delta/2$ y este error de cuantización se llama ruido granular. Si la entrada u se encuentra fuera de este rango, la magnitud será mayor que $\Delta/2$, y decimos que el cuantizador está sobrecargado. El rango $[-M\Delta/2, M\Delta/2]$ se denomina *rango de no-sobrecarga* y tiene tamaño $M\Delta$. En el análisis de sistemas de cuantización retroalimentados, los resultados dependen de manera crucial de no sobrecargar el cuantizador, y por lo tanto es importante determinar las condiciones que debe cumplir el sistema para que el cuantizador en un lazo de retroalimentación no se sobrecargue.

Desarrollemos ahora una expresión extremadamente útil para el ruido granular normalizado. Se define el error normalizado como

$$\varepsilon = e/\Delta,$$

así que $|\varepsilon| \leq 1/2$. Demostraremos que si el cuantizador no se sobrecarga entonces $\varepsilon = 1/2 - \langle u/\Delta \rangle$, donde $\langle \cdot \rangle$ es el operador fraccional definido para todos los números reales como

$$\langle r \rangle = r \bmod 1,$$

i.e., $\langle r \rangle = r - [r]$, donde $[r]$ denota el mayor entero menor o igual que r . Para demostrar este postulado, observemos de (3.1.1) que

$$-\varepsilon = (u - q(u))/\Delta$$

$$= u/\Delta - (k-1) - 1/2, \quad (k-1) \leq u/\Delta < k, \quad k = (-M/2+1), \dots, M/2, \quad (3.1.2)$$

siempre y cuando $u \in [-M\Delta/2, M\Delta/2]$. También notando que $[u/\Delta] = k-1$, y $u/\Delta - (k-1) = \langle u/\Delta \rangle$, se obtiene la expresión deseada

$$\varepsilon = 1/2 - \langle u/\Delta \rangle. \quad (3.1.3)$$

La gráfica del error normalizado — función de la entrada u del

cuantizador

$$\epsilon(\bar{u}) = e(u)/\Delta = (q(u)-u)/\Delta,$$

es mostrada en la Figura 3.1. Nótese que ésta corresponde a una función periódica con periodo unitario si la entrada u está limitada en la región de no-sobrecarga $[-M\Delta/2, M\Delta/2]$.

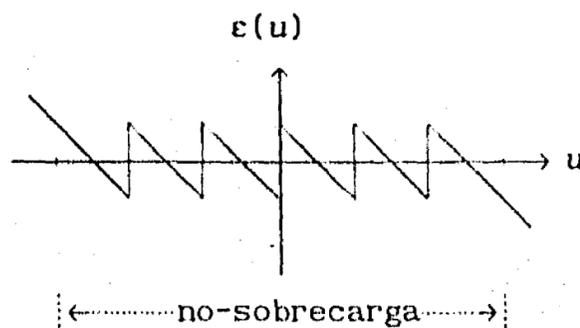


Figura 3.1 Error de Cuantizador

La representación por serie de Fourier para ϵ está dada por

$$\epsilon(u) = \sum_{l \neq 0} \frac{1}{j2\pi l} e^{j2\pi l u / \Delta} = \sum_{l=1}^{\infty} \frac{1}{\pi l} \sin(2\pi l u / \Delta), \quad (3.1.4)$$

la cual es válida para todos los valores de u excepto donde la función sea discontinua.

Finalmente, mediante un argumento similar, se puede mostrar que la expresión para el error granular de un cuantizador uniforme de un número impar de niveles de salida es

$$e = \Delta (1/2 - \langle 1/2 + u/\Delta \rangle).$$

Combinando los dos casos, encontramos la expresión general para cualquier número de niveles M , par o impar, presentada formalmente en el siguiente lema para referencia posterior:

Lema 3.1: El ruido granular de un cuantizador uniforme con paso de

cuantización Δ y M niveles de salida es

$$\frac{e}{\Delta} = \left\lfloor \frac{1}{2} - \left\langle \frac{M+u}{2\Delta} \right\rangle \right\rfloor, \text{ para } u \in [-M\Delta/2, M\Delta/2].$$

3.2 Modulador Sigma-Delta de Lazo Simple

El sistema de ecuaciones no-lineales en diferencias que gobierna este modulador (Figura 2.3) es

$$u_n = u_{n-1} + x_{n-1} - q(u_{n-1}), \quad n = 1, 2, \dots, \quad (3.2.1)$$

donde u_n , $n = 1, 2, \dots$ es la salida del integrador o la entrada al cuantizador binario, generalmente referida como el estado del sistema, y u_0 es el estado inicial; x_n , $n = 0, 1, 2, \dots$ es la entrada del modulador, una señal de tiempo discreto que se obtiene usualmente mediante muestreo de una señal analógica; $q(\cdot)$ es un cuantizador binario con paso de cuantización Δ , definido por

$$q(u) = \begin{cases} +\Delta/2, & \text{si } u \geq 0; \\ -\Delta/2, & \text{si } u < 0, \end{cases}$$

y la sucesión binaria $q_n = q(u_n)$ es la salida del modulador, la cual será posteriormente procesada por un filtro digital para recuperar la señal de entrada x_n .

Como $u_n = q(u_n) - e_n$, se sigue de (3.2.1) que

$$q_n = x_{n-1} + e_n - e_{n-1}, \quad (3.2.2)$$

y

$$u_n = x_{n-1} - e_{n-1}. \quad (3.2.3)$$

Nótese que la sucesión de error e_n es un proceso clave para nuestro análisis, ya que el estado u_n y la salida del modulador q_n se pueden expresar como funciones lineales de la sucesión de error e_n y la entrada x_n , y por lo tanto sus propiedades pueden derivarse inmediatamente usando teoría de sistemas

lineales, una vez obtenida la descripción del comportamiento de e_n .

Ahora dirigimos nuestra atención al problema central de este capítulo, i.e., la derivación de las ecuaciones en diferencias no-lineales para la sucesión de error y su solución exacta. Nos enfocaremos al error normalizado ϵ_n definido por $\epsilon_n = e_n/\Delta$. Supóngase primero que el cuantizador no se sobrecarga, los resultados obtenidos en la última sección permiten escribir

$$\epsilon_n = 1/2 - \langle u_n/\Delta \rangle, \quad n = 1, 2, \dots$$

Como $u_n/\Delta = (x_{n-1} - e_{n-1})/\Delta = x_{n-1}/\Delta - \epsilon_{n-1}$, una sustitución en la ecuación anterior conduce a los resultados deseados:

$$\epsilon_n = 1/2 - \langle x_{n-1}/\Delta - \epsilon_{n-1} \rangle, \quad n = 1, 2, \dots \quad (3.2.4)$$

Este sistema de ecuaciones en diferencia no-lineales, aparentemente difícil, puede resolverse fácilmente, y la solución exacta expresada en términos de la entrada x_n y el estado inicial u_0 del sistema, está dada por

Teorema 3.1: La sucesión de error de un modulador sigma-delta libre de sobrecarga está dada por

$$\epsilon_n = 1/2 - \langle u_0/\Delta + \sigma_n \rangle, \quad n = 0, 1, 2, \dots \quad (3.2.5)$$

donde

$$\sigma_n = \begin{cases} \sum_{k=0}^{n-1} \left(\frac{1}{2} + \frac{x_k}{\Delta} \right), & \text{si } n \geq 1; \\ 0, & \text{si } n=0. \end{cases} \quad (3.2.6)$$

Prueba: Demostramos (3.2.5) mediante inducción. Para $n=0$, (3.2.5) se reduce a $\epsilon_0 = 1/2 - \langle u_0/\Delta \rangle$, un hecho probado ya en la última sección. Suponga ahora que (3.2.5) es cierto para n , entonces de (3.2.4) se sigue que

$$\epsilon_{n+1} = 1/2 - \langle x_n/\Delta - \epsilon_n \rangle = 1/2 - \langle x_n/\Delta - 1/2 + \langle u_0/\Delta + \sigma_n \rangle \rangle.$$

Para simplificar esta expresión, usamos las siguientes propiedades de la

función $\langle \cdot \rangle$: para cualesquier números reales r y s , y todo entero l ,

$$\langle r \pm \langle s \rangle \rangle = \langle r \pm s \rangle, \quad \text{y} \quad \langle r \rangle = \langle r + l \rangle.$$

Entonces

$$\begin{aligned} \varepsilon_{n+1} &= 1/2 - \langle 1 + x_n/\Delta - 1/2 + u_0/\Delta + \sigma_n \rangle \\ &= 1/2 - \langle u_0/\Delta + \sigma_{n+1} \rangle, \end{aligned}$$

ya que $\sigma_{n+1} = \sigma_n + (1/2 + x_n/\Delta)$, y por lo tanto (3.2.5) es también cierto para $n+1$.

Ecuación (3.2.5) expresa el proceso de error ε_n como una transformación afín σ_n de la entrada x_n seguida de una no-linealidad sin memoria $\langle \cdot \rangle$. Esta es una expresión notablemente útil, puesto que las propiedades asintóticas pueden estudiarse analíticamente, como presentaremos en el siguiente capítulo.

Debemos enfatizar que el análisis previo no funciona, sin embargo, si el cuantizador binario se sobrecarga, y tenemos que encontrar las condiciones para que el cuantizador no se sobrecargue. Sea $[-b, b]$ el rango máximo permisible de la entrada, i.e., $-b \leq x_n \leq b$, $n=0, 1, \dots$. Intuitivamente deseamos escoger valores grandes para b con el objeto de tener variaciones grandes de la entrada, pero por otro lado b no debe ser demasiado grande o el cuantizador podría sobrecargarse. La pregunta que deseamos contestar es la siguiente: dado el rango de no-sobrecarga $[-\Delta, \Delta]$ del cuantizador binario, ¿cuál es el valor máximo de b tal que el cuantizador no se sobrecargue nunca, si la entrada x_n se mantiene siempre en $[-b, b]$?

La respuesta es $b = \Delta/2$, i.e., la amplitud de la entrada debe satisfacer

$$|x_n| \leq \Delta/2, \quad \text{para todo } n. \quad (3.2.7)$$

Esta es una condición suficiente y necesaria para garantizar la propiedad de no-sobrecarga en el cuantizador binario. Para probar la suficiencia, suponemos que el estado inicial u_0 se encuentra en el rango de no-sobrecarga $[-\Delta, \Delta]$, así que $e_0 \in [-\Delta/2, \Delta/2]$ por la definición de esta propiedad. Para asegurar que el cuantizador no se sobrecargue, necesitamos $u_1 = x_0 - e_0 \in [-\Delta, \Delta]$, o $|x_0 - e_0| \leq \Delta$; pero como e_0 está acotado por $\Delta/2$, la condición

$$|x_0| \leq b = \Delta/2$$

garantiza que el cuantizador no se sobrecarga para $n = 1$:

$$|u_1| = |x_0 - e_0| \leq |x_0| + |e_0|$$

$$\leq \Delta/2 + \Delta/2 = \Delta.$$

Procedemos ahora por inducción. Suponga que hasta n el cuantizador no ha sido sobrecargado, así que $|e_n| \leq b$. Pero tenemos entonces que

$$|u_{n+1}| = |x_n - e_n| \leq |x_n| + |e_n| \leq b + b = \Delta,$$

i.e., u_{n+1} está dentro del rango de no-sobrecarga y por lo tanto el cuantizador tampoco se sobrecarga para $n+1$.

La condición $|x_n| \leq b = \Delta/2$, $n = 0, 1, \dots$, es también necesaria. Esto debe de interpretarse con precaución ya que la entrada al sistema x_n no está conectada directamente al cuantizador, y esta puede salir de aquel rango ocasionalmente sin causar sobrecarga. Por necesidad se entiende que existe por lo menos una entrada que no satisface (3.2.7), para la cual el cuantizador sí sufre sobrecarga. Tal entrada obviamente existe (por ejemplo, la entrada constante).

Existe una manera más intuitiva de ver la transición del estado del sistema u_n cuando el modulador está excitado por una entrada constante $x_n = x$. Recuerde que

$$u_{n+1} = u_n + x - q(u_n) = \phi(u_n, x), \quad n = 0, 1, 2, \dots$$

La transición del estado del sistema, de u_n a u_{n+1} depende de manera crucial de si la magnitud de la constante x es mayor o no que la mitad del paso de cuantización. Dos muestras de trayectoria de transición de estado se muestran en la Figura 3.2.

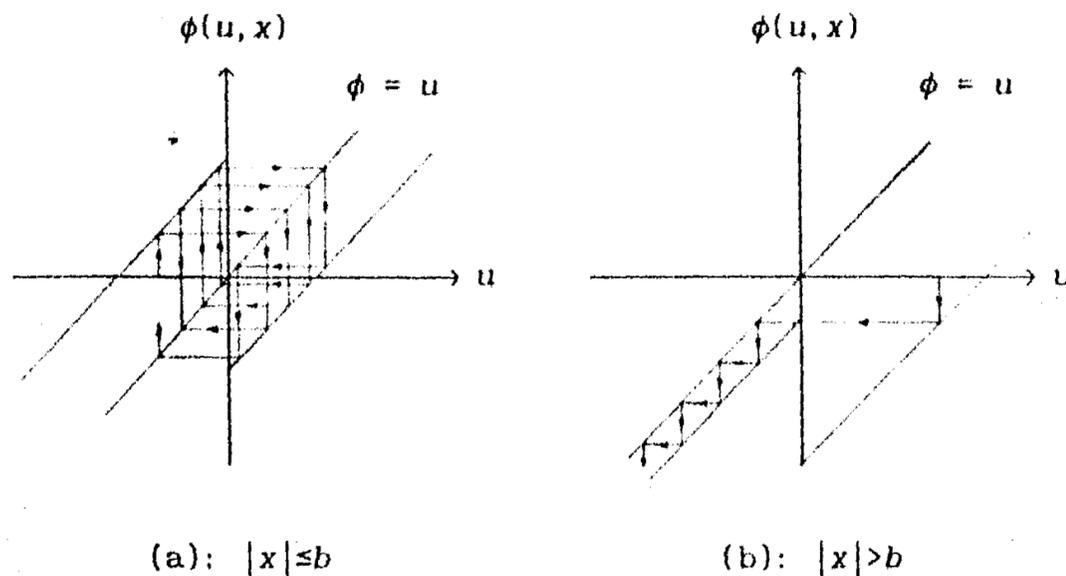


Figura 3.2 Trayectorias de Transición de Estado

En el caso donde $|x| \leq b$, u_n permanece en el rango de no-sobrecarga, esto es, $|u_n| < \Delta$, una vez que el estado inicial u_0 esté en ese rango. Por otro lado, si $|x| > b$, entonces el estado u_n se encontrará eventualmente fuera de $[-\Delta, \Delta]$, independientemente de su valor inicial, y consecuentemente el cuantizador permanecerá sobrecargado.

3.3 Moduladores Sigma-Delta Multietapa

Un modulador sigma-delta multietapa general se forma como una conexión en cascada de L moduladores idénticos de lazo simple, donde el error $e_n^{(l)}$ resultante del cuantizador binario en la etapa l -ésima está conectado a la terminal de entrada de la etapa $(l+1)$ -ésima, $l=1, 2, \dots, L-1$. Sea $u_0^{(l)}$ el estado inicial dado, tal que $|u_0^{(l)}| \leq \Delta$, para toda l , y suponga que la entrada satisface $|x_n| \leq \Delta/2$, para todo n . Concluimos por el análisis de la última sección que el cuantizador binario en la primera etapa no se sobrecarga, y el proceso de error $e_n^{(1)}$ está garantizado de estar siempre en $[-b, b]$, — el rango de entrada que no causa sobrecarga en el cuantizador de la segunda etapa. Por lo tanto, ninguno de los cuantizadores se sobrecarga ya que este argumento puede extenderse a todas las etapas. Aplicando a cada uno de los moduladores de lazo simple en el sistema multietapa la expresión para el error desarrollada en la última sección, encontramos que

$$\frac{1}{\Delta} e_n^{(l)} = \begin{cases} \frac{1}{2} - \left\langle u_0^{(l)} / \Delta + \sum_{k=0}^{n-1} \left(\frac{1}{2} + \frac{x_k}{\Delta} \right) \right\rangle, & l = 1; \\ \frac{1}{2} - \left\langle u_0^{(l)} / \Delta + \sum_{k=0}^{n-1} \left(\frac{1}{2} + \frac{1}{\Delta} e_0^{(l-1)} \right) \right\rangle, & l = 2, \dots, L. \end{cases} \quad (3.3.1)$$

En lo que sigue, por simplicidad, fijamos a cero el estado inicial del sistema, i.e., $u_0^{(l)} = 0$, $l=1, 2, \dots, L$. La forma general del error normalizado de los cuantizadores $\varepsilon_n^{(l)}$, $l=1, 2, \dots, L$, expresada sólo en términos de la entrada del sistema x_n , está dada en el siguiente teorema.

Teorema 3.2:

$$\varepsilon_n^{(l)} = (-1)^{l-1} \left[\frac{1}{2} - \langle \sigma_n^{(l)} \rangle \right], \quad n \geq l \quad (3.3.2)$$

donde

$$\sigma_n^{(l)} = \begin{cases} \sum_{i=0}^{n-1} \dots \sum_{i_1=0}^{i_2-1} \left(\frac{1}{2} + \frac{x_{i_1}}{\Delta} \right), & \text{si } n \geq l; \\ 0, & \text{si } 0 \leq n \leq l-1. \end{cases} \quad (3.3.3)$$

Nótese que la relación recursiva de $\sigma_n^{(l)}$ es,

$$\sigma_n^{(l)} = \sum_{i=0}^{n-1} \sigma_i^{(l-1)}.$$

Prueba: (Por inducción sobre l , el número de etapas.) Cuando $l=1$, (3.3.2) coincide con (3.2.5). Suponga ahora que (3.3.2) se mantiene para $l \geq 1$, entonces,

$$\begin{aligned} \varepsilon_n^{(l+1)} &= \frac{1}{2} - \left\langle \sum_{i_1=0}^{n-1} \left(\frac{1}{2} + \varepsilon_{i_1}^{(l)} \right) \right\rangle \\ &= \frac{1}{2} - \left\langle \sum_{i_1=0}^{n-1} \left[\frac{1}{2} + (-1)^{l-1} \frac{1}{2} + (-1)^l \langle \sigma_{i_1}^{(l)} \rangle \right] \right\rangle \\ &= \frac{1}{2} - \left\langle (-1)^l \sum_{i_1=0}^{n-1} \sigma_{i_1}^{(l)} \right\rangle = \frac{1}{2} - \langle (-1)^l \sigma_n^{(l+1)} \rangle, \end{aligned} \quad (3.3.4)$$

puesto que el término $1/2 + (-1)^{l-1}/2$ no contribuye a la parte fraccionaria y como $\langle\langle x \rangle \pm \langle y \rangle\rangle = \langle x \pm y \rangle$ para todo x , y reales. Para l par, tenemos que $(-1)^l = 1$, por lo que (3.3.4) se convierte en

$$\epsilon_n^{(l+1)} = \frac{1}{2} - \langle \sigma_n^{(l+1)} \rangle,$$

lo cual concuerda con (3.3.2). Similarmente, $(-1)^l = -1$ para l non y (3.3.4) se convierte en

$$\begin{aligned} \epsilon_n^{(l+1)} &= \frac{1}{2} - \langle -\sigma_n^{(l+1)} \rangle = \langle \sigma_n^{(l+1)} \rangle - \frac{1}{2} \\ &= (-1)^l \left[\frac{1}{2} - \langle \sigma_n^{(l+1)} \rangle \right], \end{aligned}$$

puesto que $\langle -x \rangle = 1 - \langle x \rangle$. Esto completa la prueba.

3.4 Moduladores Sigma-Delta Multilazo

Efectuamos ahora un análisis similar para moduladores multilazo. Para este fin adoptamos una estrategia diferente: en lugar de analizar un modulador sigma-delta multilazo general, ejemplificamos primero nuestros resultados y técnicas por medio de un caso especial — el modulador sigma-delta de doble lazo, porque este es el modulador multilazo más simple, y consecuentemente los detalles de derivación se mantienen al mínimo, y este sistema es a la vez suficientemente general para ilustrar aquellas propiedades básicas comunes a sistemas más generales. Estas técnicas pueden extenderse inmediatamente a casos generales de multilazo y el resultado relevante se presenta sin mayor detalle.

De la Figura 6 del Capítulo 2, se pueden escribir por inspección las siguientes ecuaciones en diferencia que describen el modulador sigma-delta de doble lazo:

$$u_{n+1} = v_{n+1} + u_n - q_n, \quad (3.4.1)$$

$$v_{n+1} = x_n + v_n - q_n, \quad (3.4.2)$$

donde, como antes, x_n es la entrada de tiempo discreto, u_n y v_n son el estado del sistema y $q_n = q(u_n)$ es la salida del cuantizador con entrada u_n . El proceso clave en el análisis que sigue es de nuevo el error del cuantizador, $e_n = q_n - u_n$, donde el signo es tal que la salida del cuantizador puede expresarse como una suma de la entrada al cuantizador u_n y un término de ruido e_n . También es conveniente definir

$$d_n = q_n - v_n, \quad n = 0, 1, 2, \dots$$

Demostramos que varios procesos pueden expresarse sólo en términos de la entrada x_n y el proceso de error e_n . Restando q_{n+1} de ambos (3.4.1) y (3.4.2) y aplicando la definición de e_n y de d_n , encontramos que

$$q_{n+1} = x_n + d_{n+1} - d_n, \quad (3.4.3)$$

y

$$d_{n+1} = e_{n+1} - e_n. \quad (3.4.4)$$

Substituyendo (3.4.4) en (3.4.3) se sigue que

$$q_{n+1} = x_n + e_{n+1} - 2e_n + e_{n-1}, \quad (3.4.5)$$

y entonces

$$u_{n+1} = q_{n+1} - e_{n+1} = x_n - 2e_n + e_{n-1}. \quad (3.4.6)$$

Finalmente,

$$v_{n+1} = x_n - e_n + e_{n-1} \quad (3.4.7)$$

resulta de (3.4.1) cuando $q_n - v_n$ es reemplazado por $e_n - e_{n-1} = d_n$.

En particular la sucesión de error e_n satisface la siguiente ecuación en diferencias no-lineal, la cual se obtiene por substitución de (3.4.6) en

$$e_{n+1} = e(u_{n+1}):$$

$$e_{n+1} = e(x_n - 2e_n + e_{n-1}), \quad n = 1, 2, \dots, \quad (3.4.8)$$

con las siguientes condiciones iniciales

$$e_0 = e(u_0), \quad e_1 = e(x_0 - e_0 + d_0), \quad \text{y} \quad d_0 = q(u_0) - v_0, \quad (3.4.9)$$

donde la función $e(\cdot)$ está definida como antes por $e(u) = q(u) - u$.

Antes de proceder al desarrollo de una solución exacta al sistema anterior de ecuaciones no-lineales en diferencia, necesitamos determinar el número mínimo de niveles de salida en el cuantizador, necesario para garantizar la estabilidad de no-sobrecarga. Considérese un cuantizador uniforme de M niveles con paso de cuantización Δ . Como siempre, se supone que la entrada x_n está acotada en magnitud por b . Si este cuantizador no se sobrecarga, su entrada $u_n = x_n - 2e_{n-1} + e_{n-2}$ debe estar contenida en el rango de no-sobrecarga $[-M\Delta/2, M\Delta/2]$. Pero como e_n puede asumir en principio cualquier valor en $[-\Delta/2, \Delta/2]$, y similarmente x_n puede estar en cualquier parte de $[-b, b]$, se tiene que cumplir $M\Delta/2 \geq b + 3\Delta/2$, o $b \leq (M-3)\Delta/2$. Concluimos pues que el rango máximo permisible para la entrada está dado por el intervalo $[-(M-3)\Delta/2, (M-3)\Delta/2]$. Un cuantizador binario es obviamente inapropiado para un modulador sigma-delta de doble lazo y el número mínimo de niveles de cuantización es 4 y en este caso $b = \Delta/2$.

Bajo estas suposiciones, i.e., el modulador sigma-delta de doble lazo usa un cuantizador de 4 niveles de salida con paso de cuantización Δ y el rango de entrada es $[-\Delta/2, \Delta/2]$, el cuantizador no se sobrecarga, y el sistema de ecuaciones (3.4.8) puede reescribirse como sigue, tomando en cuenta la expresión exacta para ruido granular desarrollada previamente:

$$\varepsilon_{n+1} = 1/2 - \langle x_n/\Delta - 2\varepsilon_n + \varepsilon_{n-1} \rangle, \quad n = 1, 2, \dots \quad (3.4.10)$$

Por la similitud, uno puede conjeturar que la forma general de la solución a (3.4.10) debe ser algo similar a la de modulador multietapa. De hecho,

$$\varepsilon_n = \frac{1}{2} - \left\langle \frac{1}{2} - \varepsilon_0 - n\delta_0 + s_n^{(2)} \right\rangle, \quad (3.4.11)$$

donde

$$s_n^{(2)} = \sum_{k=0}^{n-1} \sum_{i=0}^k \left(\frac{1}{2} + \frac{x_i}{\Delta} \right). \quad (3.4.12)$$

Observe que la condición inicial $\epsilon_0 = e_0/\Delta$ y $\delta_0 = d_0/\Delta$ está presente en la expresión anterior. La siguiente prueba usa inducción. Nótese primero que (3.4.11) es trivial para $n = 0$, y $n = 1$. Supóngase que (3.4.11) es cierto para $n = 0, 1, \dots, m$, entonces para $n = m+1$ tenemos

$$\begin{aligned}\epsilon_{m+1} &= 1/2 - \langle x_m/\Delta + \epsilon_{m-1} - 2\epsilon_m \rangle \\ &= 1/2 - \langle x_m/\Delta + \langle \epsilon_{m-1} - \epsilon_m \rangle - \epsilon_m \rangle,\end{aligned}$$

donde

$$\begin{aligned}\langle \epsilon_{m-1} - \epsilon_m \rangle &= \langle \epsilon_0 + \delta_0(m-1) - s_{m-1}^{(2)} - \epsilon_0 - \delta_0 m + s_m^{(2)} \rangle \\ &= \langle \sum_{k=0}^{m-1} \left(\frac{1}{2} + \frac{x_k}{\Delta} \right) - \delta_0 \rangle,\end{aligned}$$

y

$$\begin{aligned}\langle x_m/\Delta + \langle \epsilon_{m-1} - \epsilon_m \rangle - \epsilon_m \rangle \\ &= \langle \frac{1}{2} + \left[\frac{x_m}{\Delta} + \frac{1}{2} \right] + \sum_{k=0}^{m-1} \left[\frac{1}{2} + \frac{x_k}{\Delta} \right] - \delta_0 - \epsilon_0 - \delta_0 m + s_m^{(2)} \rangle \\ &= \langle \frac{1}{2} - \epsilon_0 - \delta_0(m+1) + s_{m+1}^{(2)} \rangle.\end{aligned}$$

Pues (3.4.11) es también correcto para $n = m+1$, y la prueba está completa.

Si el estado del modulador es inicialmente igual a algún nivel de cuantización, digamos $u_0 = v_0 = \Delta/2$, entonces $e_0 = d_0 = 0$, o $\epsilon_0 = \delta_0 = 0$, y consecuentemente,

$$\epsilon_n = \frac{1}{2} - \langle s_n^{(2)} \rangle, \quad n = 0, 1, \dots$$

Ahora bosquejamos el resultado para moduladores multilazo generales. Sea $L \geq 2$ el número de lazos de retroalimentación. Denota por $u_n^{(l)}$ la salida del integrador en el lazo l -ésimo, $l = 1, 2, \dots, L$. Aquí usamos la convención de que el primer lazo es el interior mientras el L -ésimo es el exterior. Por consiguiente, la entrada al cuantizador es $u_n = u_n^{(1)}$, y el error del cuantizador está dado por $e_n = e_n^{(1)} = q_n - u_n^{(1)}$, donde $q_n = q(u_n^{(1)})$ es la salida.

Usando inducción sobre el número de lazos, se puede demostrar la

siguiente relación entre q_n , u_n y e_n :

$$q_n = x_{n-1} + \nabla^L(e_n), \quad n \geq L, \quad (3.4.13)$$

y por consiguiente

$$u_n = q_n - e_n = x_{n-1} + \nabla^L(e_n) - e_n, \quad (3.4.14)$$

donde

$$\nabla^L(e_n) = \sum_{l=0}^L (-1)^l \binom{L}{l} e_{n-l}, \quad \text{para } n \geq L \quad (3.4.15)$$

es la diferencia (hacia atrás) de orden L de e_n , y $\binom{L}{l}$ denota el coeficiente binomial estándar.

Sean Δ , M y b como antes. Una aplicación de la siguiente identidad para los coeficientes binomiales

$$\sum_{l=0}^L \binom{L}{l} = 2^L$$

a (3.4.14), resulta en

$$|u_n| \leq |x_{n-1}| + |\nabla^L(e_n) - e_n| \leq b + \frac{\Delta}{2} (2^L - 1).$$

Entonces, si se requiere que el cuantizador esté libre de sobrecarga, M tiene que satisfacer

$$b + \frac{\Delta}{2} (2^L - 1) \leq M \frac{\Delta}{2} \quad \text{o} \quad b \leq \frac{\Delta}{2} (M + 1 - 2^L),$$

y concluimos por lo tanto

$$M = 2^L \quad \text{y} \quad b = \Delta/2. \quad (3.4.16)$$

Supóngase que se han escogido estos parámetros, tales que el sistema es estable en el sentido de no-sobrecarga; el error normalizado $\epsilon_n = e_n/\Delta$ satisface la siguiente relación:

$$\epsilon_n = 1/2 - \langle u_n \rangle = 1/2 - \langle x_{n-1}/\Delta + \nabla^L(\epsilon_n) - \epsilon_n \rangle,$$

y por inducción se puede demostrar que la solución exacta está dada por el siguiente teorema, cuya prueba se encuentra en [32].

Teorema 3.3:

$$\epsilon_n = \frac{1}{2} - \left\langle \frac{1}{2} + s_n^{(L)} - \sum_{l=1}^L \binom{n+l-2}{l-1} \epsilon_0^{(l)} \right\rangle, \quad (3.4.17)$$

donde

$$s_n^{(L)} = \sum_{l_1=0}^{n-1} \dots \sum_{l_1=0}^{l_2} \left(\frac{1}{2} + \frac{x_{l_1}}{\Delta} \right),$$

y

$$\epsilon_0^{(l)} = \frac{1}{\Delta} (q_0 - u_0^{(l)}), \quad l = 1, 2, \dots, L.$$

Nótese la diferencia entre $\sigma_n^{(L)}$ y $s_n^{(L)}$ para $L > 1$.

Especializando la fórmula (3.4.17) a los casos de uno y dos lazos ($L = 1, 2$), ésta coincide con los resultados previos, i.e., (3.2.5) y (3.4.11), respectivamente.

CAPITULO 4 PROPIEDADES ASINTOTICAS COMO PROMEDIOS TEMPORALES

En este capítulo estamos interesados principalmente en el comportamiento estadístico de los moduladores sigma-delta multilazo y multietapa con entrada constante o senoidal. Derivamos rigurosamente las propiedades como promedios en el tiempo, tales como momentos, espectros de potencia, así como distribuciones muestrales para varios de los procesos involucrados. Todos los resultados obtenidos en este capítulo son de naturaleza asintótica, y proporcionan descripciones interesantes del comportamiento estadístico del sistema. Nuestros resultados son consistentes con el hecho de que las razones de sobremuestreo son usualmente grandes; también son comparables con las suposiciones usadas en el análisis aproximado.

4.1 Definiciones y el Método de Transformada

En esta sección se introduce la terminología y el método de la transformada. La siguiente notación se usa a través del desarrollo de este capítulo: dado un proceso aleatorio de tiempo discreto x_n , $n = 1, 2, \dots$, denotamos

$$M\{x_n\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E(x_n), \quad (4.1.1)$$

siempre y cuando exista el límite. Si x_n es estacionario, entonces $M\{x_n\}$ se reduce a la media probabilística $E(x_n)$, y si x_n es una sucesión determinística, $M\{x_n\}$ es simplemente el promedio temporal o la media muestral. El segundo momento o la potencia promedio, y la función de autocorrelación se definen similarmente por

$$M(x_n^2) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E(|x_n|^2) \quad (4.1.2)$$

y

$$r_x(k) = M(x_n x_{n+k}^*) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N E(x_n x_{n+k}^*) \quad (4.1.3)$$

respectivamente.

Para asegurar la existencia de dichos límites, restringimos nuestra atención a una clase de procesos suficientemente general denominados *cuasi estacionarios*: un proceso de tiempo discreto x_n es llamado *cuasi-estacionario* si

1) existe una constante finita C tal que

$$|E(x_n)| \leq C, \quad \text{para todo } n,$$

y

$$|E(x_n x_m)| \leq C, \quad \text{para todo } n \text{ y } m;$$

2) y los límites

$$m_x = M\{x_n\}, \text{ y } r_x(k) = M\{x_n x_{n+k}\} \text{ para todo } k$$

existen. Procesos *cuasi estacionarios* incluyen todas las sucesiones determinísticas con media muestral y autocorrelación acotadas, todos los procesos aleatorio estacionarios de segundo orden (en el sentido amplio), así como sus combinaciones lineales.

El espectro de potencia o la densidad espectral de potencia $S_x(f)$ está definido como la transformada de Fourier de tiempo discreto de la autocorrelación $r_x(k)$, $k = 0, \pm 1, \pm 2, \dots$, i.e.,

$$S_x(f) = \sum_k r_x(k) e^{-j2\pi f k}, \quad f \in [0, 1). \quad (4.1.4)$$

Un proceso *cuasi estacionario* x_n con media cero, $M\{x_n\} = 0$, se llama *ruido blanco*, si su espectro es una constante $S_x(f) = r_x(0)$, o equivalentemente, $r_x(k) = 0$ para todo $k \neq 0$.

Sea $\chi(x; I)$ la *función indicadora* de cualquier intervalo $I = [a, b]$, i.e.,

$$\chi(x; I) = \begin{cases} 1, & \text{si } x \in I; \\ 0, & \text{caso contrario.} \end{cases}$$

Llamemos la siguiente función (definida sobre los intervalos) la *distribución marginal* o medida de una sucesión x_n :

$$\mu_x(I) = M\{\chi(x_n; I)\}. \quad (4.1.5)$$

Esta definición se interpreta como la frecuencia relativa de que los puntos de la secuencia x_n caigan en el intervalo I . Más precisamente, fija N , hay exactamente

$$v_I(N) = \sum_{n=1}^N \chi(x_n; I)$$

de los N elementos x_1, \dots, x_N que caen dentro de I . Entoces la medida es justamente el límite de las razones $v_I(N)/N$, cuando N tiende a infinito.

Un proceso x_n en $[a, b]$ es *uniformemente distribuido* en $[a, b]$ si para cada subintervalo $I = [c, d] \subseteq [a, b]$,

$$\mu_x(I) = |I|/(b-a). \quad (4.1.6)$$

donde $|I| = d-c$ es la longitud del intervalo I .

La *función de distribución conjunta* se define similarmente para varios procesos. Por ejemplo,

$$\mu_{xy}(I \times J) = M\{\chi[(x_n, y_n); I \times J]\}$$

es la distribución conjunta de los procesos x_n y y_n , definida para cada producto cartesiano de los intervalos I, J .

Dos procesos x_n y y_n son *independientes* si se cumple que

$$\mu_{xy}(I \times J) = \mu_x(I) \mu_y(J) \quad (4.1.7)$$

para cada I y J . Además, si son uniformes en $[a, b]$ y $[c, d]$ respectivamente, entonces x_n y y_n son *conjuntamente uniformes* en $[a, b] \times [c, d]$, y en este caso

$$\mu_{xy}(I \times J) = |I||J|/(d-c)(b-a), \quad (4.1.8)$$

con tales que $I \subseteq [a,b]$ y $J \subseteq [c,d]$.

Definamos la función característica marginal de x_n por

$$\Phi_x(t) = M\left\{e^{j2\pi t x_n}\right\}, \quad (4.1.9)$$

y la función característica conjunta de x_n y y_n por

$$\Phi_{xy}(t,s) = M\left\{e^{j2\pi(t x_n + s y_n)}\right\}.$$

Una técnica general para evaluar los límites en el tiempo de las secuencias involucradas en sistemas no-lineales es un análogo de tiempo discreto al llamado método de función característica [42] o de transformada [15]. La idea básica de este método se resume como sigue. Considere un sistema no-lineal que consiste en un filtro lineal seguido por una no-linealidad sin memoria $g(\cdot)$, i.e., la relación de entrada y salida puede expresarse como

$$y_n = g(s_n), \quad n = 1, 2, \dots,$$

donde s_n denota la transformación lineal o afín de la entrada x_n . Suponga además, que la no-linealidad $g(\cdot)$ es una función periódica con periodo 1, que posee representación por serie de Fourier, i.e.,

$$g(s) = \sum_1 c_1 e^{j2\pi_1 s},$$

con los coeficientes c_1 dados por

$$c_1 = \int_0^1 g(s) e^{-j2\pi_1 s} ds.$$

El problema es encontrar los promedios temporales del proceso de salida para un proceso de entrada dado. Como $y_n = \sum_1 c_1 e^{j2\pi_1 s_n}$ y como el operador $M\{\cdot\}$ es

lineal, i.e.,

$$M\{ax_n + by_n\} = aM\{x_n\} + bM\{y_n\},$$

lo cual implica que

$$m_y = M\{y_n\} = \sum_l c_l M\{e^{j2\pi l s_n}\} = \sum_l c_l \Phi_s(l),$$

y

$$\begin{aligned} r_y(k) &= M\{y_n y_{n+k}^*\} = \sum_l \sum_l c_l c_l^* M\{e^{j2\pi(l s_n - l s_{n+k})}\} \\ &= \sum_l \sum_l c_l c_l^* \Phi_s^k(l, -l), \end{aligned}$$

donde se supone que las funciones características $\Phi_s(k)$ y $\Phi_s^k(l, -l)$ están dadas. Indicamos que en la derivación de estas fórmulas los límites $M\{\cdot\}$ y la serie de Fourier se han intercambiado, y una condición suficiente (pero no necesaria) para que este intercambio de límites sea válido es que las funciones características se anulen cuando su argumento tiende a infinito.

Para ilustrar las definiciones y el método de transformada, consideremos un ejemplo sencillo pero bastante general. Sean α y β números reales cualesquiera, y defina

$$\epsilon_n = 1/2 - \langle \alpha n + \beta \rangle, \quad n = 1, 2, \dots, \quad (4.1.10)$$

donde $\langle \cdot \rangle$ denota el operador fraccional introducido en el último capítulo. Claramente, esta es una secuencia determinística con rango $[-1/2, 1/2]$. El problema es determinar su media muestral, potencia, autocorrelación y espectro, así como su distribución muestral. Recuérdese que la función $g(u) = 1/2 - \langle u \rangle$ tiene una serie de Fourier dada por

$$g(u) = \sum_{l \neq 0} \frac{1}{j2\pi l} e^{j2\pi l u},$$

entonces

$$\varepsilon_n = g(\alpha n + \beta) = \sum_{l \neq 0} \frac{1}{j2\pi l} e^{j2\pi l(\alpha n + \beta)}. \quad (4.1.11)$$

De la linealidad del operador $M\{\cdot\}$ se sigue que

$$M\{\varepsilon_n\} = \sum_{l \neq 0} \frac{1}{j2\pi l} e^{j2\pi l\beta} M\{e^{j2\pi l\alpha n}\}.$$

Obviamente, $M\{e^{j2\pi l\alpha n}\}$ es igual a 1 para $l = 0$. Para otros valores de l , este depende de si el número α es irracional or racional. Si α es irracional entonces

$$M\{e^{j2\pi l\alpha n}\} = \lim_{N \rightarrow \infty} \frac{1}{N} \frac{1 - e^{j2\pi l\alpha N}}{1 - e^{j2\pi l\alpha}} = 0.$$

Si α es racional, digamos, $\alpha = q/p$, siendo p y q enteros relativamente primos, entonces por la periodicidad de la función exponencial compleja,

$$\begin{aligned} M\{e^{j2\pi l\alpha n}\} &= \lim_{k \rightarrow \infty} \frac{1}{kp} \sum_{n=1}^{kp} e^{j2\pi l(q/p)n} = \frac{1}{p} \sum_{n=1}^p e^{j2\pi l(q/p)n} \\ &= \delta(l \bmod p) = \begin{cases} 1, & \text{si } l = mp \text{ (} m \text{ entero);} \\ 0, & \text{caso contrario.} \end{cases} \end{aligned}$$

En resumen, encontramos

$$M\{e^{j2\pi l\alpha n}\} = \begin{cases} \delta(l), & \text{si } \alpha \text{ es irracional;} \\ \delta(l \bmod p), & \text{si } \alpha \text{ es racional (} q/p \text{),} \end{cases}$$

y por lo tanto la media de ε_n es

$$M\{\varepsilon_n\} = \begin{cases} 0, & \text{si } \alpha \text{ es irracional;} \\ g(p\beta)/p, & \text{si } \alpha \text{ es racional con denominador } p. \end{cases} \quad (4.1.12)$$

Obsérvese que este resultado es asintóticamente consistente en el sentido de que si tomamos una sucesión de números racionales q/p que

convergen a un número irracional, entonces $g(p\beta)/p$ converge a 0 cuando el período p crece indefinidamente.

El mismo método se usa para calcular la autocorrelación:

$$\begin{aligned} r_{\varepsilon}(k) &= M\{\varepsilon_n \varepsilon_{n+k}^*\} = M\{g(\alpha n + \beta) g^*(\alpha(n+k) + \beta)\} \\ &= \sum_{l \neq 0} \sum_{i \neq 0} \frac{1}{4\pi^2 |l|} M\{e^{j2\pi l(\alpha n + \beta)} e^{-j2\pi l(\alpha(n+k) + \beta)}\} \\ &= \sum_{l \neq 0} \sum_{i \neq 0} \frac{1}{4\pi^2 |l|} e^{j2\pi(l\beta - l\alpha k - l\beta)} M\{e^{j2\pi(l-1)\alpha n}\}. \end{aligned}$$

Si α es irracional, entonces

$$M\{e^{j2\pi(l-1)\alpha n}\} = \delta(l-1) = \begin{cases} 1, & \text{si } l=1; \\ 0, & \text{si } l \neq 1. \end{cases}$$

Substituyendo se llega a

$$\begin{aligned} r_{\varepsilon}(k) &= \sum_{l \neq 0} \frac{1}{4\pi^2 |l|^2} e^{j2\pi \alpha k l} \\ &= \frac{1}{2\pi^2} \sum_{l=1}^{\infty} \frac{\cos 2\pi l \langle k\alpha \rangle}{l^2} \\ &= \frac{1}{12} - \frac{1}{2} \langle k\alpha \rangle (1 - \langle k\alpha \rangle), \end{aligned} \tag{4.1.13}$$

donde el último paso sigue de la fórmula 1.443.3 en [22], p.39. Obsérvese que ε_n no es un ruido blanco, con potencia promedio $r_{\varepsilon}(0) = 1/12$.

Como en este caso la autocorrelación está ya expresada como una expansión de Fourier en la forma

$$r_{\varepsilon}(k) = \sum_l s_{\varepsilon}(l) e^{j2\pi \lambda_l k}, \tag{4.1.14}$$

el espectro $S_{\varepsilon}(f)$ es discreto y consiste en un conjunto de líneas espectrales con pesos

$$s_{\varepsilon}(l) = \begin{cases} 0, & \text{si } l = 0; \\ (2\pi l)^{-2}, & \text{si } l \neq 0, \end{cases} \quad (4.1.15)$$

en las frecuencias $\lambda_l = \langle \alpha l \rangle$.

Si $\alpha = q/p$ es racional, se demuestra similarmente que

$$r_{\varepsilon}(k) = \frac{1}{12} - \frac{1}{2} \langle k\alpha \rangle (1 - \langle k\alpha \rangle) + \frac{1}{6p^2},$$

y

$$r_{\varepsilon}(0) = 1/12 + 1/6p^2.$$

como hemos notado anteriormente, la diferencia entre resultados racional e irracional desaparece a medida que el período p crece a infinito.

Un argumento similar prueba que si α es irracional entonces el proceso ε_n es uniformemente distribuido en el intervalo unitario $[-1/2, 1/2]$. La función característica de ε_n , para argumentos enteros i , es como antes

$$\begin{aligned} \Phi_{\varepsilon}(i) &= M\left\{e^{j2\pi i \varepsilon_n}\right\} = e^{j\pi i(1-2\beta)} M\left\{e^{j2\pi i n\alpha}\right\} \\ &= e^{j\pi i(1-2\beta)} \delta(i) = \begin{cases} 1, & \text{si } i=0 \\ 0, & \text{si } i=\pm 1, \pm 2, \dots \end{cases} \end{aligned} \quad (4.1.16)$$

Sea I cualquier subintervalo de $[-1/2, 1/2]$ y sea la serie de Fourier para la función indicadora de I dada por

$$\chi(\varepsilon; I) = \sum_1 c_1 e^{j2\pi 1 \varepsilon}.$$

Entonces

$$\begin{aligned} M\{\chi(\varepsilon_n; I)\} &= M\left\{\sum_1 c_1 e^{j2\pi 1 \varepsilon_n}\right\} = \sum_1 c_1 M\left\{e^{j2\pi 1 \varepsilon_n}\right\} \\ &= \sum_1 c_1 \Phi_{\varepsilon}(1) = \sum_1 c_1 \delta(1) = c_0. \end{aligned}$$

Pero como

$$c_0 = \int \chi(\epsilon; I) d\epsilon = \int_I d\epsilon = |I|,$$

por lo tanto $\mu_\epsilon(I) = |I|$, y ϵ_n es en efecto uniforme en $[-1/2, 1/2]$.

Antes de cerrar esta sección, es conveniente indicar que el proceso $\epsilon_n = 1/2 - \langle \alpha n + \beta \rangle$ estudiado en el ejemplo corresponde al proceso de error en un modulador sigma-delta de lazo simple excitado por una entrada constante x , y con un estado inicial u_0 , tales que $\alpha = (1/2 + x/\Delta)$ y $\beta = u_0/\Delta$. Consecuentemente la conclusión de este ejemplo es que el proceso de error en un modulador sigma-delta de lazo simple con entrada constante es uniformemente distribuido en $[-1/2, 1/2]$, y su espectro de potencia consiste de un número infinito de líneas espectrales, cuyo peso y localización en la frecuencia dependen de la entrada (a través de α irracional).

4.2 Algunos Resultados Relevantes de la Teoría Ergódica

Como hemos visto, el paso más crucial para calcular las varias propiedades estadísticas es la evaluación del límite $M\{\cdot\}$. Para simplificar esta tarea generalmente difícil, varios resultados de la teoría ergódica demuestran su utilidad. Los siguientes teoremas acerca de la distribución de las partes fraccionarias de polinomios se deben originalmente a Weyl (1916).

Teorema 4.1:

1) Sea $P(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_L t^L$ un polinomio de grado L ($L \geq 1$) con coeficientes reales. Entre $\alpha_1, \alpha_2, \dots, \alpha_L$, si existe por lo menos uno irracional, entonces el proceso $x_n = \langle P(n) \rangle$, $n = 1, 2, \dots$, i.e., la parte fraccionaria del polinomio $P(n)$, es uniformemente distribuido en el intervalo $[0, 1]$.

2) Sean $P_l(t)$, $l = 1, 2, \dots, L$, L polinomios de grado por lo menos uno con coeficientes reales. Entre los coeficientes (excepto el término constante) del polinomio $k_1 P_1(n) + k_2 P_2(n) + \dots + k_L P_L(n)$ si existe por lo menos un número irracional para cada L -ada de enteros $(k_1, k_2, \dots, k_L) \neq (0, 0, \dots, 0)$, entonces las sucesiones $\langle P_l(n) \rangle$, $l = 1, 2, \dots, L$ son conjuntamente uniforme módulo 1.

Vea [27, 41, 49] para las pruebas. Consulte también [40, 18] para una

discusión general y referencias adicionales.

El teorema siguiente proporciona un método efectivo para la determinación de las propiedades de procesos conjunta uniformemente distribuidos vía integrales simples, y un criterio de verificación de la uniformidad conjunta.

Teorema 4.2: Sean $x^{(l)}$ $l = 1, 2, \dots, L$, sucesiones determinísticas, y denotan por $(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(L)})$, $n = 1, 2, \dots$, la secuencia vectorial definida sobre el hipercubo unitario L -dimensional $[0, 1]^L$, el producto cartesiano de $[0, 1]$ L veces. Entonces las siguientes proposiciones son equivalentes:

(1) $(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(L)})$ es uniformemente distribuido en $[0, 1]^L$, en otras palabras, los procesos componentes $x_n^{(l)}$ son conjuntamente uniformes en $[0, 1]^L$.

(2) La igualdad

$$M\{F(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(L)})\} = \int_0^1 \int_0^1 \dots \int_0^1 F(r_1, r_2, \dots, r_L) dr_1 dr_2 \dots dr_L \quad (4.2.1)$$

se mantiene para toda función Riemann-integrable F en $[0, 1]^L$.

(3) La función característica de $(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(L)})$, i.e., la función característica conjunta de los procesos componentes,

$$\Phi_X(t_1, t_2, \dots, t_L) = M\{\exp[j2\pi(t_1 x_n^{(1)} + t_2 x_n^{(2)} + \dots + t_L x_n^{(L)})]\},$$

satisface

$$\Phi_X(k_1, k_2, \dots, k_L) = 0, \quad (4.2.2)$$

para toda L -ada de enteros $(k_1, k_2, \dots, k_L) \neq (0, 0, \dots, 0)$.

Prueba:

(1) \Rightarrow (2): Si $F(r_1, \dots, r_L)$ es Riemann-integrable en $[0, 1]^L$, entonces dado $\epsilon > 0$, existe una partición de $[0, 1]^L$ y dos funciones simples (i.e., combinaciones lineales finitas de funciones indicadoras), $s(r_1, \dots, r_L)$ y $S(r_1, \dots, r_L)$ definidas sobre esa partición, tales que

$$s(r_1, \dots, r_L) \leq F(r_1, \dots, r_L) \leq S(r_1, \dots, r_L),$$

para todo $(r_1, \dots, r_L) \in [0, 1]^L$, y

$$\int \dots \int S dr_1 \dots dr_L - \int \dots \int s dr_1 \dots dr_L \leq \epsilon.$$

tomando $M\{\cdot\}$ e integrando, encontramos respectivamente

$$M\{s(x_n^{(1)}, \dots, x_n^{(L)})\} \leq M\{F(x_n^{(1)}, \dots, x_n^{(L)})\} \leq M\{S(x_n^{(1)}, \dots, x_n^{(L)})\},$$

$$\int \dots \int s dr_1 \dots dr_L \leq \int \dots \int F dr_1 \dots dr_L \leq \int \dots \int S dr_1 \dots dr_L.$$

Pero (1) obviamente implica que

$$M\{s(x_n^{(1)}, \dots, x_n^{(L)})\} = \int \dots \int s dr_1 \dots dr_L,$$

$$M\{S(x_n^{(1)}, \dots, x_n^{(L)})\} = \int \dots \int S dr_1 \dots dr_L,$$

puesto que s y S son funciones simples, y por lo tanto

$$\begin{aligned} & |M\{F(x_n^{(1)}, \dots, x_n^{(L)})\} - \int \dots \int F dr_1 \dots dr_L| \\ & \leq \int \dots \int S dr_1 \dots dr_L - \int \dots \int s dr_1 \dots dr_L \leq \epsilon. \end{aligned}$$

La igualdad en (2) sigue del hecho de que $\epsilon > 0$ es arbitrario.

(2) \Rightarrow (3): Reemplace el integrando $F(\cdot)$ en (3) por

$$\exp[j2\pi(k_1 x_n^{(1)} + \dots + k_L x_n^{(L)})],$$

entonces (3) sigue de una evaluación directa de aquella integral.

(3) \Rightarrow (1): Escriba, para cualquier $I_1 \times \dots \times I_L \subseteq [0, 1]^L$,

$$\chi(x_n^{(1)}, \dots, x_n^{(L)}; I_1 \times \dots \times I_L) = \sum_{i_1} \dots \sum_{i_L} c(i_1, \dots, i_L) \exp[j2\pi(x_n^{(1)} i_1 + \dots + x_n^{(L)} i_L)],$$

entonces (3) implica que

$$\begin{aligned} \mu_x(I_1 \times \dots \times I_L) &= \sum_{i_1} \dots \sum_{i_L} c(i_1, \dots, i_L) \Phi_x(i_1, \dots, i_L) \\ &= c(0, \dots, 0) = \int_0^1 \dots \int_0^1 \chi(r_1, \dots, r_L; I_1 \times \dots \times I_L) dr_1 \dots dr_L \\ &= \int_{I_1} dr_1 \dots \int_{I_L} dr_L = |I_1| \dots |I_L|, \end{aligned}$$

y (1) resulta. Note que el intercambio de límites es legítimo. Esto completa la prueba.

Para ilustrar la utilidad de los teoremas anteriores, recalculamos la función de autocorrelación del ejemplo en la Sección 4.1, donde tenemos $\epsilon_n = 1/2 - \langle \alpha n + \beta \rangle$ con α irracional. Entonces

$$\begin{aligned} r_\epsilon(k) &= M\{\epsilon_n \epsilon_{n+k}\} = M\{(1/2 - \langle \alpha n + \beta \rangle) (1/2 - \langle \alpha n + \beta + k\alpha \rangle)\} \\ &= M\left\{\frac{1}{4} - \frac{1}{2}\langle \alpha n + \beta \rangle - \frac{1}{2}\langle \alpha n + \beta + k\alpha \rangle + \langle \alpha n + \beta \rangle \langle \alpha n + \beta + k\alpha \rangle\right\}. \end{aligned}$$

Como $\langle \alpha n + \beta \rangle$ y $\langle \alpha n + \beta + k\alpha \rangle$ son uniformemente distribuidos en $[0, 1]$ por el Teorema 4.1, se sigue del Teorema 4.2 que

$$\begin{aligned} r_\epsilon(k) &= \frac{1}{4} - \frac{1}{2} \int_0^1 s ds - \frac{1}{2} \int_0^1 s ds + M\{\langle \alpha n + \beta \rangle \langle \alpha n + \beta + k\alpha \rangle\} \\ &= -1/4 + M\{\langle \alpha n + \beta \rangle \langle \alpha n + \beta + k\alpha \rangle\}. \end{aligned}$$

Una aplicación simple de la siguiente igualdad

$$\langle x+y \rangle = \langle x \rangle + \langle y \rangle - \chi(\langle x \rangle + \langle y \rangle \geq 1)$$

donde

$$\chi(\langle x \rangle + \langle y \rangle \geq 1) = \begin{cases} 1, & \text{si } \langle x \rangle + \langle y \rangle \geq 1; \\ 0, & \text{si } \langle x \rangle + \langle y \rangle < 1 \end{cases}$$

al último término resulta

$$\begin{aligned} & M\{\langle \alpha n + \beta \rangle (\langle \alpha n + \beta \rangle + \langle k\alpha \rangle - \chi(\langle \alpha n + \beta \rangle + \langle k\alpha \rangle \geq 1))\} \\ &= M\{\langle \alpha n + \beta \rangle^2\} + \langle k\alpha \rangle M\{\langle \alpha n + \beta \rangle\} - M\{\langle \alpha n + \beta \rangle \chi(\langle \alpha n + \beta \rangle \geq 1 - \langle k\alpha \rangle)\}. \end{aligned}$$

Aplicando el Teorema 4.2 otra vez se obtiene

$$\begin{aligned} & \int_0^1 s^2 ds + \langle k\alpha \rangle \int_0^1 s ds - \int_0^1 s \chi(s \geq 1 - \langle k\alpha \rangle) ds \\ &= 1/3 + \langle k\alpha \rangle / 2 - \int_{1 - \langle k\alpha \rangle}^1 s ds, \end{aligned}$$

y de donde

$$r_{\epsilon}(k) = \frac{1}{12} - \frac{1}{2} \langle k\alpha \rangle (1 - \langle k\alpha \rangle),$$

el cual concuerda con el resultado previo (4.1.13).

Como la distribución de las partes fraccionarias de polinomios depende sólo de si existe entre sus coeficientes al menos un número irracional, usaremos una notación de polinomios genéricos para evitar la determinación exacta de todos los coeficientes de polinomios involucrados en el desarrollo. Represente por $\mathcal{R}^k(t)$ cualquier polinomio de grado k con únicamente coeficientes racionales; $\mathcal{J}^k(t)$ aquellos con por lo menos un coeficiente irracional (excepto el término constante). Cuando un polinomio de grado k cuyo tipo de coeficientes es irrelevante para el cálculo, lo denotemos por $\mathcal{P}^k(t)$. Usando esta notación podemos escribir, por ejemplo,

$$\mathcal{R}^k(t) + \mathcal{R}^l(t) = \mathcal{R}^k(t), \text{ si } l < k,$$

$$\mathcal{R}^l(t) \mathcal{J}^k(t) = \mathcal{J}^{k+1}(t),$$

las cuales significan que la suma de dos polinomios con coeficientes racionales es otro polinomio con coeficientes racionales, y el producto de un polinomio con coeficientes racionales por otro con coeficientes irracionales

resulta en uno con coeficientes irracionales.

Lema 4.1: Sea $P^L(x) = \alpha_L x^L + \dots + \alpha_0$ un polinomio de grado $L \geq 2$ cuyo primer coeficiente α_L es irracional. Entonces para todos los enteros $n, i, k,$ y $l,$ se mantiene la siguiente igualdad:

$$iP^L(n) + lP^L(n+k) = \begin{cases} g^L(n); & \text{si } i+l \neq 0 \\ g^{L-1}(n); & \text{si } i+l = 0 \text{ y } kl \neq 0, \\ g^{L-2}(n); & \text{si } i+l = 0 \text{ y } kl = 0. \end{cases} \quad (4.2.3)$$

Prueba: En efecto,

$$iP_L^L(n) + lP_L^L(n+k) = (i+l)\alpha_L n^L + [(i+l)\alpha_{L-1} + lkl\alpha_L]n^{L-1} + g^{L-2}(n).$$

Lema 4.2: Sea $P^L(x)$ un polinomio de grado $L \geq 1$ cuyo primer coeficiente es irracional. Entonces existen $L-1$ polinomios $P^l(x), l = 1, \dots, L-1,$ de grado l y con primer coeficiente irracional, con la propiedad de que para cada $l = 1, \dots, L-1,$ existen enteros positivos $c(l, i), i = 0, 1, \dots, l,$ tales que

$$P^L(x+1) = \sum_{i=0}^l c(l, i)P^{l-1}(x)$$

para cada número real $x.$

Este lema significa que $P^L(x+1)$ puede expresarse como una combinación lineal de $P^l(x)$ (polinomio de grado l con primer coeficiente irracional), donde los pesos son enteros positivos.

Prueba: Dado un polinomio $P^L(x) = \alpha_L x^L + \alpha_{L-1} x^{L-1} + \dots + \alpha_0,$ formemos recursivamente los polinomios $P^l(x), l = L-1, \dots, 1$ de acuerdo con la siguiente regla:

$$P^l(x) = P^{l+1}(x+1) - P^{l+1}(x).$$

Obviamente $P^l(x)$ es de grado $l.$

Si el primer coeficiente α_L de $P^L(x)$ es un número irracional, entonces el polinomio

$$\begin{aligned}
P^{l-1}(x) &= P^l(x+1) - P^l(x) \\
&= \alpha_L^*(x+1)^L + \alpha_{L-1}^*(x+1)^{L-1} + \dots + \alpha_0^* - (\alpha_L x^L + \alpha_{L-1} x^{L-1} + \dots + \alpha_0) \\
&= L\alpha_L x^{L-1} + P_{L-2}^l(x)
\end{aligned}$$

también tiene primer coeficiente $L\alpha_L$ irracional. Repitiendo el mismo argumento para $l = L-2, \dots, 2, 1$, concluimos que todos los $P^l(x)$ tienen primer coeficiente irracional.

El resto sigue por inducción sobre L . El lema es trivial para $L = 1$, i.e., $P^1(x) = P^1(x)$, y procedemos asumiendo que el lema es verdadero para $1, \dots, L$, y probándolo para $L+1$. Como $P^L(x+1) = P^L(x) + P^{L-1}(x)$, entonces

$$\begin{aligned}
P^L(x+1) &= P^L(x+1-1) + P^{L-1}(x+1-1) = \dots \\
&= P^L(x) + \sum_{k=1}^1 P^{L-1}(x+1-k).
\end{aligned}$$

Aplicando la hipótesis de inducción a la última suma, se sigue que

$$\begin{aligned}
P^L(x+1) &= P^L(x) + \sum_{k=1}^1 \left[\sum_{i=0}^{1-k} c(l-k, i) P^{L-1}(x) \right] \\
&= P^L(x) + \sum_{i=0}^{1-1} \left[\sum_{k=1}^{1-1} c(l-k, i) \right] P^{L-1}(x) \\
&= P^L(x) + \sum_{i=1}^1 \left[\sum_{k=1}^{1+1-i} c(l-k, i-1) \right] P^{L+1-i}(x).
\end{aligned}$$

Esta es la combinación lineal deseada donde todos los coeficientes de ponderación (las sumas entre paréntesis) son obviamente enteros positivos, y esto termina la prueba.

El teorema que sigue es una consecuencia del Lema 4.2 y el Teorema 4.1.

Teorema 4.3: Si $P^L(x)$ es un polinomio de grado $L \geq 1$ con su primer coeficiente irracional, entonces los procesos $\langle P^L(n) \rangle$, $\langle P^L(n+1) \rangle$, ..., $\langle P^L(n+L-1) \rangle$ son conjuntamente uniforme módulo 1.

Prueba: El teorema sigue de la segunda parte del Teorema 4.1, si podemos demostrar que el polinomio $\sum_{l=0}^{L-1} k_l P^l(x+1)$ tiene por lo menos un coeficiente irracional (excepto el término constante) para cada L -ada de enteros $(k_0, k_1, \dots, k_{L-1}) \neq (0, 0, \dots, 0)$. Aplicando el Lema 4.2, encontramos

$$\begin{aligned} \sum_{l=0}^{L-1} k_l P^l(x+1) &= \sum_{l=0}^{L-1} k_l \left[\sum_{i=0}^l c(l, i) P^{l-i}(x) \right] \\ &= \sum_{i=0}^{L-1} P^i(x) \left[\sum_{l=i}^{L-1} k_l c(l, i) \right] = \sum_{i=0}^{L-1} k'_i P^i(x), \end{aligned}$$

donde $P^i(x)$ es de grado i , $i = 0, 1, \dots, L$, y tiene primer coeficiente irracional, y

$$k'_i = \sum_{l=i}^{L-1} c(l, i) k_l, \quad i = 0, 1, \dots, L-1,$$

son enteros porque $c(l, i)$ lo son. Además,

$$(k'_0, k'_1, \dots, k'_{L-1}) = (0, 0, \dots, 0),$$

si y sólo si

$$(k_0, k_1, \dots, k_{L-1}) = (0, 0, \dots, 0),$$

porque estos vectores están relacionados por un sistema de L ecuaciones lineales cuya matriz de coeficientes es triangular con entradas positivas $c(l, l)$ sobre la diagonal principal y por tanto es invertible. Usando el hecho de que todos los $P^i(x)$ tienen diferentes grados y el primer coeficiente irracional, concluimos que el polinomio obtenido por combinación entera

$$\sum_{i=0}^{L-1} k'_i P^i(x),$$

tiene por lo menos un coeficiente irracional, a menos que

$$(k'_0, k'_1, \dots, k'_{L-1}) = (0, 0, \dots, 0),$$

y esto implica a su vez que $\sum_{l=0}^{L-1} k_l P^L(x+l)$ debe tener por lo menos un coeficiente irracional para todo $(k_0, k_1, \dots, k_{L-1}) \neq (0, 0, \dots, 0)$. Esto termina la prueba.

4.3 El Resultado Fundamental: Entrada Constante

Ahora enunciamos uno de los resultados más importantes de esta investigación, acerca del comportamiento estadístico asintótico de moduladores sigma-delta con entradas constantes, resumido en el siguiente teorema.

Teorema 4.4: El proceso de error del cuantizador e_n en cualquier modulador sigma-delta con entrada constante es, con probabilidad uno, uniformemente distribuido en $[-\Delta/2, \Delta/2]$. Además, e_n es también un ruido blanco independiente de la entrada, con la única excepción del modulador sigma-delta de lazo simple.

Prueba: No tratamos el caso de multietapas, porque la prueba es similar a la prueba para moduladores multilazos.

Sea $L=1, 2, \dots$ el número de lazos y considere un modulador sigma-delta de L lazos con entrada constante $x \in [-b, b]$. Asíumase que $\alpha = (1/2 + x/\Delta)$ es un número irracional. Esta suposición se cumple con probabilidad uno si x es escogida de $[-b, b]$ de acuerdo a una función de densidad de probabilidad. Especializando la ecuación (3.4.17) para una entrada constante, se convierte en

$$\begin{aligned} \epsilon_n &= \frac{1}{2} - \left\langle \frac{1}{2} - \sum_{l=1}^L \binom{n+l-2}{l-1} \epsilon_0^{(l)} + \sum_{l=L}^{n-1} \dots \sum_{l=1}^2 \alpha \right\rangle \\ &= 1/2 - \langle P^L(n) \rangle, \end{aligned} \tag{4.3.1}$$

donde $P^L(n)$ tiene un primer coeficiente irracional. Esto se puede ver usando

la notación de polinomios genéricos y el hecho de que

$$\sum_{l_1=0}^{n-1} \dots \sum_{l_1=0}^{l_2} 1 = \binom{n+L-1}{L} = \mathcal{R}^L(n),$$

y se sigue que

$$P^L(n) = \mathcal{P}^{L-1}(n) + \alpha \mathcal{R}^L(n) = \mathcal{J}^L(n),$$

ya que α es irracional. Por conveniencia, definimos un proceso en $[0,1]$ por

$$w_n = 1/2 - \varepsilon_n = \langle P^L(n) \rangle. \quad (4.3.2)$$

Como w_n es la parte fraccionaria de un polinomio irracional de grado $L \geq 1$, es entonces uniformemente distribuido en $[0,1]$ por el Teorema 4.1. Esto implica a su vez que $\varepsilon_n = 1/2 - w_n$ es uniformemente distribuido en el intervalo $[-1/2, 1/2]$, y por lo tanto $e_n = \Delta \varepsilon_n$ es uniforme en $[-\Delta/2, \Delta/2]$. Note que e_n tiene media cero y potencia $\Delta^2/12$.

Para demostrar que el proceso ε_n es también blanco para $L \geq 2$, o equivalentemente $r_\varepsilon(k) = 0$, para todo $k \neq 0$, es suficiente probar que

$$\Phi_\varepsilon^k(i, l) = 0, \text{ para todo } (i, l) \neq (0, 0), \text{ y } k \neq 0, \quad (4.3.3)$$

donde

$$\begin{aligned} \Phi_\varepsilon^k(i, l) &= \mathbb{M}\{\exp[j2\pi(i\varepsilon_n + l\varepsilon_{n+k})]\} \\ &= \exp[j\pi(i+l)] \mathbb{M}\{\exp[-j2\pi(iP^L(n) + lP^L(n+k))]\}, \end{aligned}$$

usando el método de transformada, esto implica

$$r_\varepsilon(k) = \sum_{l \neq 0} \sum_{i \neq 0} \frac{1}{4\pi^2 il} \Phi_\varepsilon^k(i, -l) = 0.$$

Sin embargo, por el Lema 4.1

$$iP^L(n) + lP^L(n+k) = \mathcal{J}^L(n) \text{ o } \mathcal{J}^{L-1}(n), \text{ a menos que } i+l = 0 \text{ y } lk = 0,$$

en otras palabras, es un polinomio irracional de grado por lo menos $L-1$, a menos que $l=i=0$, puesto que k es por suposición diferente de 0. Como $L \geq 2$, los Teoremas 4.1 y 4.2 conducen a (4.3.3)

$$\Phi_c^k(i, l) = e^{j2\pi(l+1)} \int_0^1 e^{-j2\pi r} dr = 0.$$

Remarcamos que este argumento no es aplicable al modulador de lazo simple, ya que en tal caso ($L = 1$) el proceso de error de cuantizador no es blanco, como ya hemos demostrado en la Sección 4.1.

Para terminar la prueba, demostramos que e_n es independiente de la entrada constante x , en el sentido definido en la Sección 4.1, Ec.(4.1.6). Para $I \times J \subseteq [-b, b]^2$, tenemos

$$\chi(x, e_n; I \times J) = \begin{cases} 0, & \text{si } x \notin I; \\ \chi(e_n; J), & \text{si } x \in I, \end{cases}$$

entonces

$$\begin{aligned} \mu_{x_e}(I \times J) &= M\{\chi(x, e_n; I \times J)\} = \begin{cases} 0, & \text{si } x \notin I; \\ \mu_e(J), & \text{si } x \in I \end{cases} \\ &= \mu_x(I) \mu_e(J), \end{aligned}$$

porque $\mu_x(I) = M\{\chi(x; I)\} = 1$, si $x \in I$; $= 0$, de otra manera. Esto termina la prueba.

Intuitivamente, los moduladores sigma-delta de orden superior deben producir ruido de cuantización más "blanco", pero esta idea no está reflejada en el teorema anterior. El resultado siguiente proporciona una exposición precisa de esta intuición.

Una secuencia x_n con rango $[a, b]$ es llamada L -equidistribuida si [37]

1) x_n y sus $L-1$ versiones recorridas

$$x_n^{(l)} = x_{n+l}, \quad l = 1, \dots, L-1, \quad (4.3.4)$$

son conjuntamente uniforme en $[a, b]^L$, y

2) cualesquiera $l > L$ secuencias recorridas no son conjuntamente uniformes.

Claramente, un proceso L -equidistribuido es también l -equidistribuido para $l \leq L$, pero no viceversa. Note también que la noción de "ruido blanco" es aproximadamente equivalente a 2-equidistribución. Por ejemplo, el proceso de error en un modulador de lazo simple tiene una distribución uniforme pero no es ruido blanco, por lo tanto es 1-equidistribuido en $[-1/2, 1/2]$. Una generalización a moduladores sigma-delta multilazo o multietapa se presenta en el teorema siguiente.

Teorema 4.5: El proceso de error normalizado de cuantizador ε_n en un modulador sigma-delta de L -lazos o L -etapas con una entrada constante es L -equidistribuido en el intervalo $[-1/2, 1/2]$.

Prueba: Con probabilidad uno, $\alpha = 1/2 + x/\Delta$ es un número irracional, y el resultado que derivamos bajo esta suposición se cumple casi seguramente. Primero, defina como antes $w_n = 1/2 - \varepsilon_n$, entonces tenemos que probar que $w_n^{(0)}, w_n^{(1)}, \dots, w_n^{(L-1)}$ son conjuntamente uniformes en $[0, 1]^L$. Pero esto es justamente una consecuencia inmediata del Teorema 4.3, ya que $w_n = \langle P^L(n) \rangle$, donde $P^L(\cdot)$ es un polinomio de grado L cuyo primer coeficiente es un número irracional.

Demostramos ahora que los $L+1$ procesos $e_n, e_{n+1}, \dots, e_{n+L}$ no son conjuntamente uniformes. Sea A el alfabeto de reproducción del cuantizador interno del modulador sigma-delta de L lazos, entonces A es un conjunto finito que contiene 2^L niveles de salida. Considere la función

$$F(e_n, e_{n+1}, \dots, e_{n+L}) = \chi \left(x + \sum_{l=1}^L (-1)^l \binom{L}{l} e_{n-l}; A \right) \quad (4.3.5)$$

donde $\chi(\cdot; \cdot)$ denota la función indicadora. Como A es finito, F es Riemann integrable y su integral sobre $[-1/2, 1/2]^{L+1}$ es cero:

$$\iiint \dots \int F(r_0, r_1, \dots, r_L) dr_0 dr_1 \dots dr_L = 0.$$

Por otro lado, recuerde del Capítulo 3, Ec. (3.4.13) que

$$x + \sum_{l=1}^L (-1)^l \binom{L}{l} e_{n-l} = q_n \in A,$$

es el proceso de salida del cuantizador, por lo tanto

$$M\{F(e_n, e_{n+1}, \dots, e_{n+L})\} = M\{\chi(q_n; A)\} = 1.$$

Se sigue del Teorema 4.1 que $e_n, e_{n+1}, \dots, e_{n+L}$ no son conjuntamente uniformes, y el teorema está demostrado.

4.4 El Resultado Fundamental: Entrada Senoidal

En el desarrollo siguiente, confinamos atención a moduladores sigma-delta multietapa, pero las técnicas y resultados son aplicables a sistemas multilazo.

Sea la entrada x_n a un modulador sigma-delta una senoide pura, i.e.,

$$x_n = a \cos(\omega n + \theta), \quad n = 0, 1, \dots,$$

donde $a \in [0, b]$, $\omega = 2\pi f/f_s$, y $\theta \in [0, 2\pi)$, como definidas en la Sección 2.4.

Lema 4.3:

$$\sum_{l=0}^{n-1} \dots \sum_{l_1=0}^{l_2-1} \cos(n\omega + \theta) = P(n) + \alpha^L \cos 2\pi Q(n), \quad (4.4.1)$$

donde

$$\alpha = [2\sin(\omega/2)]^{-1}, \quad (4.4.2)$$

$$P(n) = \alpha \sin(\omega/2 - \theta) \frac{n^{L-1}}{(L-1)!} + \mathcal{P}^{L-2}(n), \quad (4.4.3)$$

$$Q(n) = \frac{\omega}{2\pi} n - \frac{L}{4\pi}(\omega + \pi) + \frac{\theta}{2\pi}. \quad (4.4.4)$$

El lema se puede demostrar usando inducción sobre L . Omitimos, sin embargo, los detalles de la prueba, y sólo indicamos que las siguientes

fórmulas son útiles (cf. página 39 de [22]):

$$\sum_{l=0}^{n-1} \cos(l\omega) = \frac{\sin(\omega/2) + \sin(n-1/2)\omega}{2\sin(\omega/2)} = \alpha \left[\sin \frac{\omega}{2} + \sin(n\omega - \frac{\omega}{2}) \right],$$

$$\sum_{l=0}^{n-1} \sin(l\omega) = \frac{\cos(\omega/2) - \cos(n-1/2)\omega}{2\sin(\omega/2)} = \alpha \left[\cos \frac{\omega}{2} + \cos(n\omega - \frac{\omega}{2}) \right].$$

El siguiente teorema es una analogía al Teorema 4.4.

Teorema 4.6: Si la entrada es una senoide pura, entonces con probabilidad uno el proceso de error del cuantizador en todos los moduladores multilazo o multietapa con al menos tres lazos o etapas es un ruido blanco uniforme e independiente de la entrada.

Prueba: Nuestra prueba se basa en los moduladores multietapa; la prueba para moduladores multilazo es similar. El error de cuantizador en la última etapa de un modulador sigma-delta de L etapas con entrada senoidal es [cf. (3.3.2)]

$$\begin{aligned} \epsilon_n &= (-1)^{L-1} \left[\frac{1}{2} - \left\langle \sum_{l_L=0}^{n-1} \dots \sum_{l_1=0}^{l_2-1} \frac{a}{\Delta} \cos(\omega n + \theta) \right\rangle \right] \\ &= (-1)^{L-1} \left[\frac{1}{2} - \left\langle \frac{a}{\Delta} P(n) + \frac{a}{\Delta} \alpha^L \cos 2\pi Q(n) \right\rangle \right], \end{aligned}$$

donde hemos usado el Lema 4.3 en el último paso. Definase por conveniencia

$$w_n = \left\langle \frac{a}{\Delta} P(n) + \frac{a}{\Delta} \alpha^L \cos 2\pi Q(n) \right\rangle, \quad n = 1, 2, \dots, \quad (4.4.5)$$

y deseamos demostrar que w_n es uniformemente distribuido en $[0, 1]$.

Note que para $L \geq 2$, los polinomios

$$P'(n) = \frac{a}{\Delta} P(n) = \frac{a}{\Delta} \alpha \sin(\omega/2 - \theta) \frac{n^{L-1}}{(L-1)!} + P^{L-2}(n),$$

$$= \begin{cases} \beta = \Delta^{-1} \alpha \sin(\omega/2 - \theta), & \text{si } L = 1; \\ \beta n + \mathcal{P}_0, & \text{si } L = 2; \\ \beta n^{L-1} / (L-1)! + \mathcal{P}^{L-2}(n), & \text{si } L \geq 3 \end{cases} \quad (4.4.6)$$

y $Q(n)$ en (4.4.4) satisfacen las condiciones del Teorema 4.1, parte (2), con probabilidad uno, i.e., entre los coeficientes del polinomio (de grado $L-1 \geq 1$) $IP'(n) + kQ(n)$ existe por lo menos un número irracional, para todos los pares de enteros $(l, k) \neq (0, 0)$, y por lo tanto concluimos que $\langle P'(n) \rangle$ y $\langle Q(n) \rangle$ son conjuntamente uniformes.

Usando el Teorema 4.2, la función característica de w_n se calcula para argumentos enteros, como sigue:

$$\begin{aligned} \Phi_w(k) &= M\{\exp(j2\pi k w_n)\} \\ &= M\left\{\exp\left[j2\pi k \left[\langle P'(n) \rangle + \gamma \cos 2\pi \langle Q(n) \rangle\right]\right]\right\} \end{aligned}$$

$$= \int_0^1 e^{j2\pi k s} ds \int_0^1 e^{j2\pi \gamma k \cos 2\pi t} dt = \begin{cases} 0, & \text{si } k \neq \pm 1, \pm 2, \dots; \\ 1, & \text{si } k = 0, \end{cases}$$

donde $\gamma = \frac{a}{\Delta} \alpha^L$. Esto prueba que w_n es uniformemente distribuido en $[0, 1]$, en vista del Teorema 4.2.

Si $L = 1$, $P'(n)$ se reduce en una constante β , y por lo tanto

$$\Phi_w(k) = e^{j2\pi k \beta} \int_0^1 e^{j2\pi \gamma k \cos 2\pi t} dt = e^{j2\pi k \beta} J_0(2\pi \gamma k),$$

donde $J_m(\cdot)$ denota la función de Bessel de orden m . En este caso, $k \neq 0$ no implica necesariamente que $\Phi_w(k) = 0$, así que w_n ya no es uniforme.

Mostramos que ε_n es un ruido blanco cuando $L \geq 3$. Para ello, sólo necesitamos probar que para $k \neq 0$,

$$\Phi_w^k(i, l) = M\left\{e^{j2\pi(lw_n + lw_{n+k})}\right\} = \begin{cases} 1, & \text{si } (i, l) = (0, 0); \\ 0, & \text{de otro modo,} \end{cases} \quad (4.4.7)$$

ya que esto implica que w_n y w_{n+k} son conjuntamente uniformes en $[0, 1]^2$, o equivalentemente, ε_n y ε_{n+k} conjuntamente uniforme en $[-1/2, 1/2]$, y por lo

tanto ε_n es blanco. Considere para $k \neq 0$ y el par de enteros $(i, l) \neq (0, 0)$,

$$I\{P'(n) + \gamma \cos 2\pi Q(n)\} + I\{P'(n+k) + \gamma \cos 2\pi Q(n+k)\} = \\ \{IP'(n) + IP'(n+k)\} + \gamma\{i \cos 2\pi Q(n) + l \cos[2\pi Q(n) + k\omega]\}.$$

Aplicando el Lema 4.1 al primer término, tenemos que

$$IP'(n) + IP'(n+k) = \mathcal{J}^{L-1}(n) \text{ or } \mathcal{J}^{L-2}(n),$$

pero como $L \geq 3$ o $L-2 \geq 1$, este es conjuntamente uniforme con $Q(n)$, y entonces una aplicación del Teorema 4.2 obtiene (4.4.7):

$$\Phi_W^k(i, l) = \int_0^1 e^{j2\pi s} ds \int_0^1 e^{j2\pi \gamma \{i \cos 2\pi t + l \cos(2\pi t + k\omega)\}} dt = 0.$$

Los casos $L = 1, 2$ necesitan un tratamiento especial y el resultado es que en ambos moduladores de una y dos etapas, el proceso de error ε_n tiene un espectro discreto, el cual se calcula de $\Phi_W^k(i, l)$ usando el método de la transformada. Omitimos los detalles.

Finalmente necesitamos demostrar que el proceso de error es independiente de la entrada al modulador. Observe primero que la función característica conjunta o cruzada $\Phi_{xw}(i, l) = 0$, o equivalentemente, $\Phi_{x\varepsilon}(i, l) = 0$, para $l \neq 0$. Esto sigue del hecho de que $x_n = a \cos[2\pi \langle Q(n) \rangle + \mathcal{P}_0]$ y el mismo argumento usado para probar (4.4.7). Ahora con el método de transformada en mente, esto implica el resultado deseado:

$$\mu_{x\varepsilon}(I \times J) = \sum_{i, l} c(i, l) \Phi_{x\varepsilon}(i, l) = \sum_i c(i, 0) \Phi_{x\varepsilon}(i, 0) = \sum_i c(i, 0) \Phi_x(i) \\ = \mu_\varepsilon(J) \sum_i c(i) \Phi_x(i) = \mu_\varepsilon(J) \mu_x(I),$$

donde hemos usado el hecho de que

$$c(i, 0) = \int_I \int_J e^{j2\pi(i x + 0 \varepsilon)} dx d\varepsilon = c(i) |J| = c(i) \mu_\varepsilon(J).$$

Esto completa la prueba del teorema.

4.5 Característica Espectral de la Secuencia Modulada

El comportamiento estadístico de segundo orden de la secuencia de salida q_n de un modulador sigma-delta se obtiene de la secuencia de error ϵ_n como sigue. Recuerde que

$$q_n = x'_n + \nabla^L(\epsilon_n) = x'_n + \Delta \sum_{l=0}^L (-1)^l \binom{L}{l} \epsilon_{n-l}, \quad \text{si } n \geq L,$$

donde x'_n es una versión retrasada de la entrada. Note que los valores de q_n para $0 \leq n < L$ no influye el resultado final del cálculo de los promedios en el tiempo, — propiedades asintóticas de la sucesión entera y cualquier efecto de términos finitos eventualmente desaparece.

Considere primero entrada constante, $x'_n = x$. La media muestral de q_n es

$$M\{q_n\} = x + \nabla^L(M\{\epsilon_n\}) = x, \quad (4.5.1)$$

como se espera. Como los procesos del error y de la entrada no están correlacionados, la autocorrelación de la salida está dada por

$$\begin{aligned} R_q(k) &= x^2 + \Delta^2 \sum_{l=0}^L \sum_{m=0}^L (-1)^{l+m} \binom{L}{l} \binom{L}{m} M\{\epsilon_{n-l} \epsilon_{n+k-m}^*\} \\ &= x^2 + \frac{\Delta^2}{12} (-1)^k \sum_{l=0}^{L-k} \binom{L}{l} \binom{L}{k+l} = x^2 + \frac{\Delta^2}{12} (-1)^k \binom{2L}{L+k}, \end{aligned} \quad (4.5.2)$$

donde hemos usado el hecho de que ϵ_n es ruido blanco y la siguiente identidad [16, p.58]

$$\sum_{r=0}^{n-k} \binom{n}{k+r} \binom{m}{r} = \binom{m+n}{n-k}.$$

Tomando transformada de Fourier, el espectro de q_n es

$$S_q(f) = x^2 \delta(f) + \frac{\Delta^2}{12} \sum_{l=-\infty}^{\infty} (-1)^l \binom{2L}{L+l} e^{-j2\pi l f}$$

$$= x^2 \delta(f) + \frac{\Delta^2}{12} |2 \sin(\pi f)|^{2L}. \quad (4.5.3)$$

En el caso de entrada senoidal, $x_n = a \cos(\omega n + \theta)$, se puede demostrar similarmente que

$$S_q(f) = \frac{a^2}{4} \left[\delta\left(f - \frac{\omega}{2\pi}\right) + \delta\left(f - 1 + \frac{\omega}{2\pi}\right) \right] + \frac{\Delta^2}{12} |2 \sin(\pi f)|^{2L}, \quad f \in [0, 1).$$

En ambos casos (constante y senoidal), el primer término del espectro de salida es exactamente el espectro de la señal de entrada, y el término restante es por lo tanto considerado como el término de ruido. Este ruido es común en ambos casos y es independiente de la entrada (contrario a los moduladores de lazo simple).

Para concluir, indicamos que nuestro resultado es consistente con la aproximación por ruido blanco usado en el método usual de linealización. Este método aproximado produce predicción correcta sobre el comportamiento estadístico de los moduladores sigma-delta con entrada constante o senoidal.

CAPITULO 5 DESEMPEÑO DEL SISTEMA

El propósito de este capítulo es la derivación del desempeño global del sistema cuando se recupera la señal de entrada analógica a partir de la salida producida por un modulador sigma-delta, usando un filtro decodificador. Consideramos solamente filtros lineales e invariantes en el tiempo por su simplicidad en la implementación y por su buen desempeño.

5.1 Modulación Sigma-Delta y Cuantización Uniforme

La propiedad fundamental de la modulación sigma-delta es que la señal de entrada, analógica puede recuperarse a la tasa original de muestreo de Nyquist, a partir de la salida modulada, y la precisión de conversión depende principalmente de la razón de sobremuestreo. En esta sección, consideramos solamente entradas constantes y la medida de desempeño del sistema de cuantización sigma-delta es la magnitud de error. Resultados más generales sobre el error cuadrático medio de cuantización se presentan en el resto del capítulo.

Considérese primero un sistema de cuantización sigma-delta que consiste en un modulador sigma-delta de lazo simple seguido por un filtro "acumula y vacía" o filtro "peine" definido por

$$\hat{x} = \frac{1}{N} \sum_{n=1}^N q_n, \quad (5.1.1)$$

donde N es la razón de sobremuestreo. Asuma que la entrada analógica es una constante x en el rango $[-b, b]$, y el sistema empieza en un estado inicial arbitrario $u_0 \in [-\Delta, \Delta]$, donde $\Delta = 2b$. Estamos interesados en determinar qué tan cercana es la salida del sistema \hat{x} a la entrada x , y comparar su desempeño con aquel de un cuantizador uniforme escalar que tiene un número

apropiado de niveles de salida.

Observe que la salida \hat{x} puede tomar cualquiera de los $N+1$ diferentes valores

$$A = \left\{ \frac{2b}{N} k - b; k = 0, 1, 2, \dots, N \right\}$$

porque q_n es $+b$ o bien $-b$, y el número de ocurrencias de $+b$ en la secuencia de $N \pm b$'s puede ser $k = 0, 1, \dots, N$. Obviamente este alfabeto de reproducción A coincide con aquel producido por un cuantizador uniforme para el intervalo $[-b-b/N, b+b/N]$, con paso de cuantización $2b/N$ y $N+1$ niveles. Denote por $Q(\cdot)$ este cuantizador uniforme, y recuerde que este es un mapeo de distorsión mínima, i.e.,

$$|x - Q(x)| = \min\{|x - y|, y \in A\}. \quad (5.1.2)$$

Como una consecuencia, el error debido a la cuantización sigma-delta está siempre acotado inferiormente (en magnitud), por el error del cuantizador uniforme:

$$|\hat{x} - x| \geq |Q(x) - x|.$$

Una cota superior a $|\hat{x} - x|$ se encuentra como sigue. Recuerde que la salida del modulador q_n puede expresarse como una función del error de cuantizador binario, i.e.,

$$q_n = x + e_n - e_{n-1}, \quad (5.1.3)$$

substitución de (5.1.3) en (5.1.1) resulta en

$$\frac{1}{N} \sum_{n=1}^N q_n = \frac{1}{N} \sum_{n=1}^N (x + e_n - e_{n-1}) = x + (e_N - e_0)/N,$$

donde la simplificación se debe a las cancelaciones en la suma telescópica. Reescribiendo la ecuación anterior en términos de ε_n , la versión normalizada de e_n , y recordando que $\varepsilon_n \in (-1/2, 1/2]$, encontramos que

$$|\hat{x} - x| = \frac{\Delta}{N} |\varepsilon_N - \varepsilon_0| \leq \frac{\Delta}{N}. \quad (5.1.4)$$

Como se espera, el cuantizador sigma-delta no puede tener mejor desempeño que el correspondiente cuantizador de distorsión mínima $Q(\cdot)$, porque en el último caso tenemos una cota superior más estrecha [cf. (5.1.2)]

$$|Q(x) - x| \leq b/N.$$

Ambas cotas, sin embargo, son proporcionales inversamente a la razón de sobremuestreo N .

Obviamente, las cota anteriores implican que $|Q(x) - \hat{x}| \leq 3b/N$, así que el desempeño del sistema de cuantización sigma-delta es bastante cercano al del cuantizador uniforme $Q(\cdot)$. Sin embargo, se puede encontrar cotas más estrechas y a veces muy remarcables. Como $Q(\cdot)$ tiene $N+1$ niveles de salida y paso de cuantización Δ/N , usando el Lema 3.1, el error puede ser expresado como

$$Q(x) - x = \frac{\Delta}{N} \left[\frac{1}{2} - \left\langle \frac{Nx + N + 1}{\Delta} \right\rangle \right].$$

Recuerde también que [cf. (3.2.5)]

$$\epsilon_N = \frac{1}{2} - \left\langle \frac{1}{2} + N \left(\frac{x}{\Delta} + \frac{1}{2} \right) - \epsilon_0 \right\rangle = \frac{1}{2} - \left\langle \frac{1}{2} + N\alpha - \epsilon_0 \right\rangle,$$

donde $\alpha = \frac{x}{\Delta} + \frac{1}{2}$. Entonces,

$$\begin{aligned} |Q(x) - \hat{x}| &\leq \frac{\Delta}{N} \left| \frac{1}{2} - \left\langle \frac{1}{2} + N\alpha - \epsilon_0 \right\rangle - \epsilon_0 - \frac{1}{2} + \left\langle \frac{1}{2} + N\alpha \right\rangle \right| \\ &= \frac{\Delta}{N} \left| \left\lfloor \frac{1}{2} + N\alpha - \epsilon_0 \right\rfloor - \left\lfloor \frac{1}{2} + N\alpha \right\rfloor \right| \leq \frac{\Delta}{N}, \end{aligned} \quad (5.1.5)$$

ya que, por suposición, $\epsilon_0 \in (-1/2, 1/2]$, y entonces el valor absoluto puede ser solamente 0 o 1. Indicamos que si el estado inicial del modulador sigma-delta de lazo simple es $u_0 = \pm b$, tal que $u_0 \in [x-b, x+b]$ o equivalentemente $\epsilon_0 = 0$, entonces $\hat{x} = Q(x)$, y significa que estos sistemas de cuantización son equivalentes en el sentido de entrada y salida.

Para generalizar este método en el dominio del tiempo a moduladores de órdenes superiores, digamos un modulador de L lazos, consideramos la

siguiente extensión directa del filtro peine, llamado filtro sinc^L:

$$\hat{x} = N^{-L} \sum_{i_1=0}^{N-1} \dots \sum_{i_L=0}^{N-1} q_{LN-i_1-\dots-i_L}$$

Usando el hecho de que $q_n = x + \nabla^L(e_n)$, y simplificando la suma telescópica resultante, encontramos que

$$\hat{x} - x = N^{-L} \sum_{i=0}^L (-1)^i \binom{L}{i} e_{N(L-i)},$$

y de donde

$$|\hat{x} - x| < N^{-L} b 2^L = b \left(\frac{2}{N}\right)^L \quad (5.1.6)$$

se sigue con la ayuda de la igualdad $\sum_1 \binom{L}{i} = 2^L$. Note que (5.1.6) es en efecto una generalización de (5.1.4), la cota derivada para el sistema de lazo simple.

Desde luego, cotas superiores para error de cuantización medio cuadrático se obtiene fácilmente elevando al cuadrado las cotas correspondientes para la magnitud de error de cuantización, y son proporcionales a N^{-2L} . Con el método en el dominio de tiempo, esta es la mejor tasa a la cual el error de cuantización decrece a cero con la razón de sobremuestreo N . Sin embargo, es posible una evaluación más precisa del desempeño global de sistema, como veremos más tarde.

5.2 Desempeño Asintótico de Sistema

Las medidas de desempeño más populares para convertidores analógico/digitales son el error cuadrático medio y la razón señal a ruido de cuantización. Estas son, contrario a las cotas obtenidas en la última sección, descripciones estadísticas del desempeño del sistema, y por lo tanto se deben definir y entender como propiedades asintóticas en la misma terminología usada en el Capítulo 4. La ventaja definitiva de este enfoque asintótico es que permite

el uso de argumentos espectrales para evaluar la potencia del ruido de cuantización sigma-delta global.

Suponga que un modulador sigma-delta multilazo tiene una secuencia de entrada x_n , $n = 0, 1, \dots$, y su secuencia de salida q_n , $n = 1, 2, \dots$, se alimenta a un filtro decodificador, el cual es, por simplicidad, lineal e invariante en el tiempo, con una respuesta a impulso unitario dada por h_n , $n = 0, 1, \dots$. Denotando por \hat{x}_n la salida del filtro, entonces

$$\hat{x}_n = \sum_{k=0}^{\infty} h_k q_{n-k}.$$

Como la salida del modulador (i.e., la entrada al filtro decodificador), es la suma de la entrada retrasada más el ruido de modulación, i.e.,

$$q_n = x_{n-1} + N_n,$$

la señal a la salida del filtro decodificador es

$$\hat{x}_n = \sum_{k=0}^{\infty} h_k x_{n-1-k} + \sum_{k=0}^{\infty} h_k N_{n-k},$$

donde la primera suma es una versión filtrada de la señal y el segundo término es el ruido filtrado. Llamemos la componente ruidosa del proceso \hat{x}_n el ruido de cuantización sigma-delta, y su potencia promedio, el error cuadrático medio de cuantización sigma-delta, denotado por σ^2 .

Recuerde del Capítulo 4 que el espectro de potencia de N_n es

$$S_N(f) = [2\sin(\pi f)]^{2L} S_e(f) = [2\sin(\pi f)]^{2L} \frac{\Delta^2}{12} \quad (5.2.1)$$

para sistemas multilazo y multietapa con entradas constante y senoidal (por lo menos tres lazos o etapas en el último caso), porque en estos casos el espectro de error de cuantizador $S_e(f)$ es simplemente $\frac{\Delta^2}{12}$. Sea $H(f)$ la función de transferencia del filtro decodificador, i.e.,

$$H(f) = \sum_{k=0}^{\infty} h_k e^{-j2\pi kf}. \quad (5.2.2)$$

Usando la teoría de sistemas lineales, el espectro del ruido filtrado es igual a

$$|H(f)|^2 S_N(f),$$

y su potencia se calcula como la integral del espectro de potencia. Este resultado es resumido en el teorema que sigue.

Teorema 5.1: Si la salida de moduladores sigma-delta multilazo o multietapa con entrada constante o senoidal (en el último caso se requiere de por lo menos tres lazos o etapas) es decodificada por un filtro lineal e invariante en el tiempo $H(f)$, entonces la potencia promedio del ruido de cuantización sigma-delta se calcula vía la siguiente fórmula:

$$\sigma^2 = \frac{\Delta^2}{12} \int_0^1 |H(f)|^2 [2\sin(\pi f)]^{2L} df. \quad (5.2.3)$$

Una vez encontrada σ^2 , La razón señal a ruido de cuantización (SQNR) se determina por

$$\text{SQNR} = \text{potencia de señal} / \sigma^2 = \begin{cases} x^2 / \sigma^2, & \text{entrada constante;} \\ a^2 / 2\sigma^2, & \text{entrada senoidal.} \end{cases}$$

En la próxima sección calculamos la potencia normalizada de cuantización sigma-delta

$$\Psi = \int_0^1 |H(f)|^2 [2\sin(\pi f)]^{2L} df \quad (5.2.4)$$

para varios filtros decodificadores específicos.

Antes de cerrar esta sección, consideramos cómo determinar el desempeño de un cuantizador sigma-delta de lazo simple excitado por una entrada constante. No es difícil mostrar que el espectro de la salida es discreto, dado por

$$r_q(k) = \sum_i s_q(i) e^{j2\pi k i \alpha},$$

donde

$$s_q(k) = \begin{cases} x, & \text{si } k=0; \\ \frac{\sin^2(\pi k \alpha) \Delta^2}{(\pi k)^2} & \text{si } k \neq 0. \end{cases}$$

De donde la potencia de cuantización sigma-delta se puede calcular como una suma infinita

$$\sigma^2 = \sum_{k \neq 0} s_q(k) |H(\langle k \alpha \rangle)|^2 = \Delta^2 \sum_{k \neq 0} \frac{\sin^2 \pi \langle k \alpha \rangle}{(\pi k)^2} |H(\langle k \alpha \rangle)|^2. \quad (5.2.5)$$

Para simplificar la expresión necesitamos una suposición adicional de que $\alpha = 1/2 + x/\Delta$ es una variable aleatoria uniformemente distribuida en $[0, 1]$, y considere la siguiente media probabilística

$$E \left\{ \Delta^2 \sum_{k \neq 0} \frac{\sin^2 \pi \langle k \alpha \rangle}{(\pi k)^2} |H(\langle k \alpha \rangle)|^2 \right\} = \Delta^2 \sum_{k \neq 0} \frac{1}{(\pi k)^2} \int_0^1 \sin^2 \pi \langle k r \rangle |H(\langle k r \rangle)|^2 dr,$$

observando que todas las integrales son iguales debido a la periodicidad de los integrandos, la anterior se convierte en

$$\sigma^2 = \frac{\Delta^2}{3} \int_0^1 \sin^2 \pi f |H(f)|^2 df, \quad (5.2.6)$$

donde hemos usado la identidad

$$\sum_{k=1}^{\infty} k^{-2} = \frac{\pi^2}{6}.$$

Note que si especializamos la Ec.(5.2.3) a $L=1$, tenemos (5.2.6), y esto significa que la potencia normalizada de cuantización sigma-delta de un sistema de lazo simple puede calcularse en la misma manera para los sistemas multilazo o multietapa.

5.3 Evaluación de Desempeño para Varios Filtros de Decodificación

Se considera los siguientes decodificadores: filtro paso-baja ideal, filtro sinc y filtro FIR óptimo. N denota la razón de sobremuestreo en el cálculo.

El filtro paso-baja ideal está dado por

$$H(f) = \begin{cases} 1, & \text{si } |f| \leq 1/2N; \\ 0, & \text{de otro modo.} \end{cases} \quad (5.3.1)$$

Observe que la banda de paso de este filtro paso-baja es la misma que la banda de la señal de entrada original. Como la función de transferencia es periódica con período unitario, la potencia normalizada del ruido de cuantización es

$$\left(\int_0^{1/2N} + \int_{1-1/2N}^1 \right) [2\sin(\pi f)]^{2L} df = 2^{2L+1} \int_0^{1/2N} (\sin \pi f)^{2L} df.$$

Para N grande, necesitamos considerar sólo el integrando para valores pequeños de f , y podemos usar la serie de Taylor $\sin(x) = x + O(x^3)$. Entonces

$$\begin{aligned} \Psi &= 2^{2L+1} \pi^{2L} \int_0^{1/2N} f^{2L} df + \int_0^{1/2N} O(f^{6L}) df \\ &\approx \frac{\pi^{2L}}{(2L+1)N^{2L+1}}. \end{aligned} \quad (5.3.2)$$

Observe que la potencia promedio del ruido de cuantización a la salida del decodificador tiende a cero a una tasa de $1/N^{2L+1}$, la cual es mejor que la tasa obtenida en la sección anterior usando sólo técnicas en el dominio del tiempo. Es de interés encontrar filtros más simples que puedan proporcionar similares compromisos de desempeño, puesto que el filtro paso-baja ideal sólo puede aproximarse en la práctica.

Un filtro sinc^k de N coeficientes puede construirse mediante una conexión en cascada de k filtros peine de N/k coeficientes, donde la razón de sobremuestreo N es por suposición un múltiplo de k . Los filtros sinc^1 y sinc^2 se llaman algunas veces peine y triangular respectivamente, por la forma de su respuesta a impulso unitario. El nombre "sinc" sigue del hecho de que la magnitud de su función de transferencia está dada por

$$|H(f)| = \left| \frac{k}{N} \frac{\sin(\pi f N/k)}{\sin(\pi f)} \right|^k. \quad (5.3.3)$$

La evaluación de la integral

$$\Psi = \int_0^1 \left| \frac{k}{N} \frac{\sin(\pi f N/k)}{\sin(\pi f)} \right|^{2k} [2\sin(\pi f)]^{2L} df$$

es una tarea generalmente difícil, pero para los casos más interesantes donde $k = L$ y $k = L+1$, se pueden obtener expresiones en forma cerrada como sigue.

$k=L$: sea $m=N/L$, y haciendo cambio de variable $\varphi = \pi f m$, se tiene

$$\begin{aligned} \int_0^1 |\sin(m\pi f)|^{2L} df &= \frac{1}{m\pi} \int_0^{m\pi} |\sin\varphi|^{2L} d\varphi = \frac{2}{\pi} \int_0^{\pi/2} (\sin\varphi)^{2L} d\varphi \\ &= \frac{(2L-1)!!}{(2L)!!} = \frac{(2L)!}{L! L!} 2^{-2L}, \end{aligned}$$

donde el último paso es una identidad bien conocida, entonces

$$\Psi = \left(\frac{2L}{N}\right)^{2L} \frac{(2L)!}{L! L!} 2^{-2L} = \left(\frac{2L}{L}\right) \left(\frac{L}{N}\right)^{2L} \quad (5.3.4)$$

$k=L+1$: sea $m=N/(L+1)$ y suponga que $M = \frac{m-1}{2}$ es un entero. La integral se convierte en

$$\Psi = m^{-(2L+2)} \int_0^1 \left| \frac{\sin(\pi f m)}{\sin(\pi f)} \right|^2 [2\sin(\pi f m)]^{2L} df.$$

Como

$$\frac{\sin(m\pi f)}{\sin(\pi f)} = \sum_{n=-M}^M e^{j2\pi f n}, \quad \text{y} \quad [2\sin(m\pi f)]^L = (-1)^L e^{-j\pi L m f} \sum_{l=0}^L \binom{L}{l} e^{j2\pi m f l},$$

entonces el integrando se convierte en el cuadrado de

$$\frac{\sin(m\pi f)}{\sin(\pi f)} [2\sin(m\pi f)]^L = (-1)^L e^{-j\pi L m f} \sum_{n=-M}^M \sum_{l=0}^L \binom{L}{l} e^{j2\pi f (ml+n)}.$$

Aplicando la igualdad de Parseval, su integral resulta en

$$\int_0^1 \left| \frac{\sin(m\pi f)}{\sin(\pi f)} \right|^2 [2\sin(m\pi f)]^{2L} df = m \sum_{l=0}^L \binom{L}{l}^2 = m \binom{2L}{L},$$

y de donde la integral original resulta en

$$\Psi = \binom{2L}{L} \left(\frac{L+1}{N}\right)^{2L+1}. \quad (5.3.5)$$

Es natural intentar el filtro de mayor orden. Sin embargo, el siguiente teorema indica que el compromiso de desempeño entre potencia de ruido de cuantización sigma-selta y la razón de sobremuestreo no puede mejorarse substancialmente aumentando el orden de los filtros sinc^k .

Teorema 5.2: La solución del siguiente problema de optimización con restricciones:

$$\text{Minimizar } \Psi = \int_0^1 \left| \sum_{k=0}^N h_k e^{-j2\pi k f} \right|^2 [2\sin(\pi f)]^{2L} df, \quad (5.3.6)$$

sobre todos los filtros lineales con $N+1$ parámetros h_n , $n=0,1,\dots,N$, sujeto a la restricción de *ganancia unitaria en la componente directa*

$$\sum_{k=0}^N h_k = H(0) = 1, \quad (5.3.7)$$

está dada por

$$h_k^* = \frac{\binom{L+k}{L} \binom{L+N-k}{L}}{\binom{N+2L+1}{2L+1}}, \quad k = 0, 1, \dots, N, \quad (5.3.8)$$

y

$$\Psi^* = \binom{2L}{L} \binom{N+2L+1}{2L+1}^{-1}. \quad (5.3.9)$$

Para N grande las siguientes aproximaciones son útiles:

$$h_k \cong \frac{(2L+1)!}{L! L!} \frac{1}{N} \left(\frac{k}{N}\right)^L \left(\frac{N-k}{N}\right)^L, \quad k=0,1,\dots,N,$$

y

$$\Psi^* \cong \binom{2L}{L} \frac{(2L+1)!}{N^{2L+1}}.$$

Note también que la restricción (5.3.7) permite que entradas constantes o la componente directa de las entradas pasen sin ser escaladas.

Prueba: Expandiendo la suma en (5.3.6) encontramos que

$$\Psi = \sum_{k=0}^N \sum_{l=0}^N h_k h_l r_{k-l} = r_0 \sum_{k=0}^N h_k^2 + 2 \sum_{l=1}^L r_l \sum_{k=0}^{N-l} h_k h_{k+l},$$

donde r_l es la autocorrelación de q_n , dada por (cf. Sec. 4.5)

$$r_l = \int_0^1 |2 \sin(\pi f)|^2 e^{j2\pi l f} df = \begin{cases} (-1)^l \binom{2L}{L+l}, & \text{si } |l| \leq L, \\ 0, & \text{si } |l| > L. \end{cases}$$

Introduciendo un multiplicador de Lagrange -2λ , diferenciando

$$\Psi - 2\lambda \sum_{k=0}^N h_k$$

con respecto a h_k e igualando las derivadas parciales a 0, se llega las ecuaciones

$$\begin{aligned} \sum_{l=-k}^L r_l h_{k+l} &= \lambda, & \text{si } 0 \leq k < L, \\ \sum_{l=-L}^L r_l h_{k+l} &= \lambda, & \text{si } L \leq k \leq N-L, \\ \sum_{l=-L}^{N-k} r_l h_{k+l} &= \lambda, & \text{si } N-L < k \leq N, \end{aligned}$$

donde r_l están dadas arriba. Estas $N+1$ ecuaciones más la restricción de ganancia unitaria (5.3.7) proporcionan $N+2$ ecuaciones para las $N+2$ incógnitas h_k y λ ; la solución está dada por (5.3.8) como se puede verificar en lo que sigue.

$\sum_k h_k^* = 1$ es una consecuencia inmediata de la siguiente identidad [16, p. 58]:

$$\sum_{k=0}^a \binom{a-k}{m} \binom{b+k}{n} = \binom{a+b+1}{m+n+1}.$$

Ahora denote

$$\lambda' = \sum_1 (-1)^1 \binom{2L}{L+1} \binom{k+L+1}{L} \binom{N-k+L-1}{L}.$$

Encontramos después de una manipulación simple (pero tediosa) que

$$\lambda' = \binom{2L}{L},$$

y de donde

$$\lambda = \binom{2L}{L} \binom{N+2L+1}{2L+1}^{-1}.$$

Finalmente,

$$\begin{aligned} \Psi^* &= r_0 \sum_k (h_k^*)^2 + 2 \sum_1 r_1 \sum_k h_k^* h_{k+1}^* \\ &= \sum_k h_k^* \left(r_0 h_k^* + 2 \sum_1 r_1 h_{k+1}^* \right) = \sum_k h_k^* \lambda = \lambda, \end{aligned}$$

entonces se sigue el teorema.

Comparando el filtro FIR óptimo y el filtro paso-baja ideal, concluimos que filtros FIR e IIR pueden proporcionar desempeños del mismo orden de magnitud, i.e., $\Psi = O(N^{-(2L+1)})$, pero el primero es generalmente menos complejo y consecuentemente más favorable.

En la derivación de estas expresiones exactas del desempeño asintótico del sistema, la suposición de razones de sobremuestreo grandes es usada implícitamente, así que es natural preguntar ¿qué tan grande debe ser la razón de sobremuestreo, para que estas predicciones teóricas sean razonablemente precisas? Desafortunadamente, esta cuestión se puede contestar sólo por simulación, porque los varios teoremas de la teoría ergódica no son lo suficientemente fuertes como para proporcionar información sobre la velocidad de convergencia. Podemos, sin embargo, usar todavía las cotas superiores desarrolladas en la Sección 5.1, las cuales son razonablemente cercanas a los resultados asintóticos correspondientes para valores pequeños de la razón de sobremuestreo, digamos $N \leq 50$.

5.4 Comparaciones

Como la modulación sigma-delta intenta asistir en la implementación de convertidores analógico/digitales de alta resolución, es interesante comparar su desempeño con el de un cuantizador escalar uniforme estándar. La pregunta es: dado un sistema de cuantización sigma-delta, ¿qué tan grande debe ser la razón de sobremuestreo, con el objeto de alcanzar el mismo error medio cuadrático de un cuantizador uniforme con un (gran) número dado de niveles de salida?

Para contestar esta pregunta, recordemos el siguiente resultado clásico de la teoría de cuantización asintótica [19,21]: si una variable aleatoria x con un rango dinámico apropiado, digamos $[-1/2, 1/2]$ y una función de densidad de probabilidad suave, es la entrada a un cuantizador uniforme de paso de cuantización Δ , entonces el error de cuantización cuadrático medio es aproximadamente $\Delta^2/12$, cuando el número de niveles de salida $M = 1/\Delta$ es suficientemente grande. En otras palabras,

$$E|x-q(x)|^2 \cong 2^{-2B}/12,$$

Si $B = \log_2 M$, la tasa (en bits) del cuantizador $q(\cdot)$ es grande, digamos $B \geq 7$.

Por otro lado, las potencias de ruido de cuantización sigma-delta se derivan como función de la razón de sobremuestreo N para varios sistemas representativos, resumidas como sigue ($\Delta=1$):

1. Lazo simple:

Filtro FIR óptimo:	$\sigma^2 \cong 1/N^3$
Filtro triangular:	$\sigma^2 = 4/3N^3$
Filtro peine:	$\sigma^2 = 1/6N^2$

2. Doble lazo o dos etapas:

Filtro FIR óptimo:	$\sigma^2 \cong 60/N^5$
Filtro sinc ³ :	$\sigma^2 = 243/2N^5$
Filtro triangular:	$\sigma^2 = 8/N^4$

3. Triple lazo o tres etapas:

Filtro FIR óptimo:	$\sigma^2 \cong 8400/N^7$
Filtro sinc ⁴ :	$\sigma^2 = 81920/3N^7$

Filtro sinc³:

$$\sigma^2 = 1215/N^6$$

Igualando las potencias $\sigma^2 = E|x-q(x)|^2$, y expresando N en términos de B , tenemos la siguiente tabla (la Tabla 5.1).

PCM		lazo simple			doble lazos			triple lazos		
Bits	SQNR dB	opt.	sinc ²	sinc ¹	opt.	sinc ³	sinc ²	opt.	sinc ⁴	sinc ³
12	72.2	587	646	5793	104	120	202			
13	78.2	930	1024		138	159	284			
14	84.3	1477	1628		181	210	402	83	100	126
15	90.3	2345	2582		238	276	568	102	120	159
16	96.3				315	363	802	124	148	201
17	102.3				416	480	1134	151	180	252
18	108.4				549	633	1604	184	220	318

Tabla 5.1 Razón de Sobremuestreo contra Tasa de Bit

Nota: SQNR (dB) $\approx 6.02 B$ (bits)

Se puede concluir que los sistemas multilazo o multietapa son mucho más eficientes que los sistemas de lazo simple. Observe que el sistema de doble lazo o dos etapas con filtro sinc³ y una razón de sobremuestreo moderado puede proporcionar un desempeño suficientemente bueno para la mayoría de los propósitos prácticos.

CAPITULO 6 DESEMPEÑO BAJO IMPERFECCIONES DE CIRCUITO

Desde luego, diversas imperfecciones o imprecisiones de circuito se pueden aparecer en las implementaciones prácticas de modulación sigma-delta, como en cualquier implementación de sistemas desarrollados teóricamente. Todos los efectos no ideales sólo pueden degradar el desempeño del sistema y hacen nuestras predicciones teóricas inadecuadas sobre el comportamiento del sistema, ya que todos los resultados previos fueron derivados basados en modelos idealizados. En este capítulo presentamos un modelo comprensivo para el modulador sigma-delta de lazo simple bajo imperfecciones de circuito. Estimamos su desempeño cuando la entrada es constante y el filtro de decodificación es el filtro "acumula y vacía". Comparación con el desempeño ideal revela que el sistema de modulación sigma-delta es robusto contra ciertas imperfecciones de circuito. Aquellas imperfecciones que pueden producir degradaciones grandes también están identificadas.

6.1 Motivación

Se cree popularmente que la insensibilidad a imperfecciones de los circuitos es una de las características deseables de los moduladores sigma-delta, porque (1) se usan sólo componentes de baja precisión tales como cuantizadores binarios, y (2) precisión de temporización o sincronización es fácil de obtener con VLSI. Esto se basa principalmente en observaciones de las implementaciones experimentales y parece intuitivamente razonable. Pero mientras no exista un análisis cuidadoso para hacerlo riguroso, quedará siempre como un resultado empírico.

Considere, por ejemplo, el sistema mostrado en la Figura 6.1, el cual representa una variante de un convertidor analógico/digital por aproximaciones sucesivas.

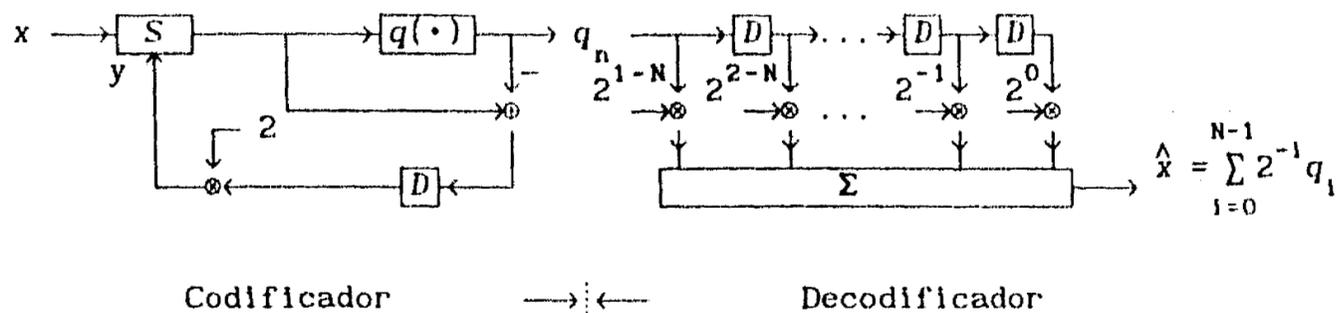


Figura 6.1 Convertidor A/D por Aproximaciones Sucesivas

En esta figura, "S" denota un interruptor, cuya salida es igual a x si $n=0$, y es igual a y , si $n=1, \dots, N$; $q(\cdot)$ es un cuantizador binario cuya salida es $\pm b$. No es difícil ver que el error de cuantización está acotado por

$$|x - \hat{x}| \leq \Delta/2^N, \quad (\Delta = 2b)$$

así que es mucho más eficiente que el cuantizador sigma-delta formado por un modulador sigma-delta de lazo simple y filtro peine, porque en este último caso tenemos

$$|x - \hat{x}| \leq \Delta/N.$$

Observe también que ambos sistemas tienen una complejidad estructural comparable. Entonces, ¿por qué el cuantizador sigma-delta es atractivo?

Hablando *grosso modo*, en la presencia de imperfecciones de circuito, la modulación sigma-delta alcanza todavía un desempeño razonablemente bueno, i.e., se puede demostrar que el error de cuantización sigma-delta decrece a cero conforme la razón de sobremuestreo se incrementa a infinito, aunque la velocidad es más lenta. Esto se debe al efecto de promedio del decodificador, de modo que algunos efectos no ideales se cancelan parcialmente a su salida. Por otro lado el sistema en la Figura 6.1 usa un decodificador diferente que carece de tal efecto de promedio, por eso es extremadamente sensible a, por ejemplo, los errores de umbral del cuantizador: una vez que un bit en la salida del codificador esté erróneamente codificado, no puede ser corregido por el decodificador más tarde. Entonces, una desviación ligera del sistema

ideal puede resultar en una degradación drástica de desempeño, y por lo tanto el sistema en la Figura 6.1 no es robusto en este sentido.

El resto del capítulo proporciona un análisis del desempeño del cuantizador sigma-delta en la presencia de imperfecciones de circuito.

6.2 Modelo de Imperfecciones de Circuito

Existen muchas formas de imperfecciones de circuito, tales como fuga de corriente y desajuste de ganancia en el integrador, error de umbral y desajustes de los niveles de salida en el cuantizador binario, efectos no lineales en los sumadores y el integrador, retrasos parásitos en el lazo de retroalimentación e imprecisiones de temporización, etc. Por simplicidad, consideramos solamente cinco tipos de imperfecciones que pueden aparecer en las componentes de circuito de un modulador sigma-delta de lazo simple. Más precisamente, consideramos

- 1) *Error de umbral* en el cuantizador binario, cuyo efecto es particularmente severo cuando la operación es rápida. La causa es la variación del voltaje de umbral debida al ruido térmico y el efecto de histéresis.
- 2) *Desajuste de niveles de salida* en el cuantizador binario. En el caso ideal se supone $\pm b = \pm \Delta/2$.
- 3) *Desajuste de ganancia* en el integrador. Mientras que el integrador ideal tiene ganancia unitaria, la ganancia del integrador actual $G_I > 0$ puede ser diferente de 1.
- 4) *Fuga de corriente* en el integrador. Como un resultado, el integrador ideal debe modificarse de manera complicada, y el sistema resultante es bastante difícil de analizar teóricamente.
- 5) *Imprecisiones de temporización* significa que el intervalo de tiempo entre los instantes adyacentes de muestreo pueden diferir del período de muestreo ideal. Estas diferencias son usualmente pequeñas y se llaman error de temporización.

Un cuantizador binario con las imperfecciones (1) y (2) puede modelarse razonablemente bien por uno mostrado en la Figura 6.2.

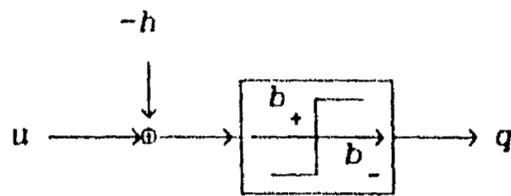


Figura 6.2 Cuantizador Binario Imperfecto

La relación de entrada y salida está dada por

$$q = \begin{cases} b_+, & \text{si } u \geq h; \\ b_-, & \text{si } u < h, \end{cases} \quad (6.2.1)$$

donde h denota el error de umbral, el cual se supone que es un ruido con magnitud acotada, i.e.,

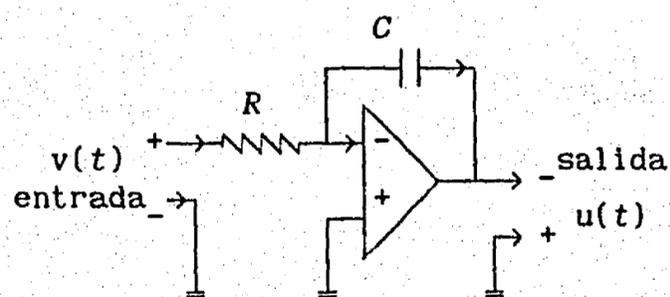
$$|h| \leq H. \quad (6.2.2)$$

Los parámetros b_+ y b_- representan los niveles de salida reales, y pueden expresarse, sin pérdida de generalidad, como

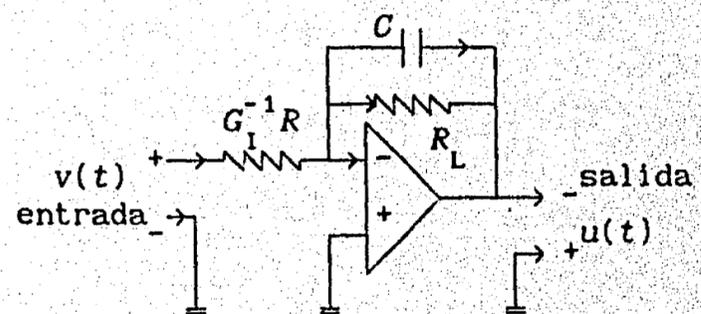
$$b_+ = G_q b + \delta, \quad \text{y} \quad b_- = -G_q b + \delta, \quad (6.2.3)$$

donde $b = \Delta/2$ es el valor ideal; $G_q > 0$ y δ modelan el desajuste de los niveles de salida.

Un modelo apropiado para un integrador con las imperfecciones (3) y (4) se obtiene modificando uno ideal, como se muestra en la Figura 6.3.



(a) Integrador Ideal



(b) Integrador Imperfecto

Figura 6.3 Modelos para Integradores Ideal y No Ideal

En la Figura 6.3(a), si la resistencia R y la capacitancia C se escogen de manera tal que $RC=T_s$, — el periodo de muestreo del modulador, la salida $u(t)$ y la entrada $v(t)$ están relacionadas por

$$u(t) = u(0) + \frac{1}{T_s} \int_0^t v(\tau) d\tau, \quad (6.2.4)$$

donde $u(0)$ es el estado inicial del integrador en $t_0=0$. En la Figura 6.3(b), los efectos del desajuste de ganancia y de fuga de corriente están representados, respectivamente, por un factor $G_I > 0$ y una resistencia adicional grande R_L conectada en paralelo al capacitor. Denotando por $T_L = R_L C$, la ecuación correspondiente es

$$u(t) = u(0) \exp(-t/T_L) + \frac{G_I}{T_s} \int_0^t \exp(\tau/T_L) v(\tau) d\tau. \quad (6.2.5)$$

Comparando las Ecs. (6.2.5) y (6.2.4), se puede ver que la relación de entrada y salida actual es mucho más compleja, y el integrador ideal es un caso especial del no ideal con $T_L = \infty$ y $G_I = 1$.

Si se usa un reloj perfecto para sincronizar el modulador sigma-delta, los instantes de muestreo t_n , $n=0,1,\dots$ están separados por el periodo de muestreo fijo T_s . Como se presentan los errores de temporización, sin embargo, tenemos

$$t_{n+1} - t_n = T_s + \tau_n, \quad n=0,1,\dots, \quad (6.2.6)$$

donde los errores de temporización τ_n son usualmente pequeños (en magnitud) relativo a T_s y pueden considerarse como una secuencia aleatoria independiente e idénticamente distribuida con una distribución marginal sobre un rango apropiado, digamos $[-\tau, \tau]$, con $\tau < T_s/2$.

Ahora tomando en cuenta todas las cinco imperfecciones, y muestreando el modulador de tiempo continuo en la Figura 2.1 en t_n , $n=0,1,\dots$, ($t_0=0$), i.e., definiendo $u_n = u(t_n)$, $q_n = q(t_n)$, $h_n = h(t_n)$, etc., obtenemos el modelo de tiempo discreto mostrado en la Figura 6.2. En este modulador nuevo tenemos una fuente de ruido aditivo $-h_n$ delante del cuantizador binario ideal, la cual causa el error de umbral; dos niveles de salida desajustados $+b_+$ y b_- en

lugar de $\pm b$; una ganancia extra del integrador G_I ; y dos parámetros adicionales α_n y β_n debidos a la fuga de corriente en el integrador, los cuales varían con el tiempo cuando también están presentes imprecisiones de temporización.

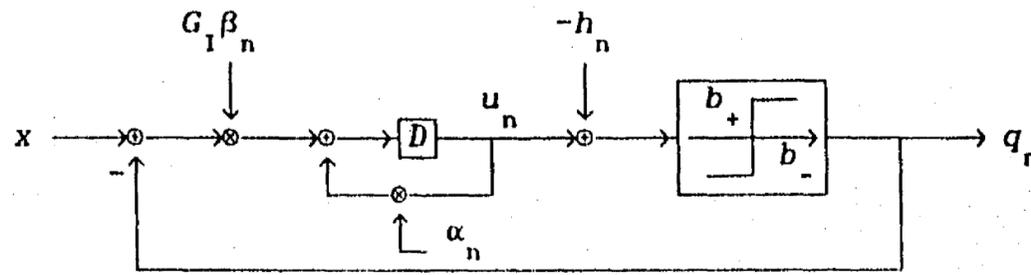


Figura 6.4 Modulador Sigma-Delat Imperfecto

La ecuación en diferencias que gobierna este modulador imperfecto es

$$u_{n+1} = \alpha_n u_n + G_I \beta_n (x - q_n), \quad (6.2.7)$$

donde

$$q_n = G q_n^* + \delta, \quad \text{con} \quad q_n^* = b \operatorname{sgn}(u_n - h_n), \quad (6.2.8)$$

$$\alpha_n = \exp\left(-\frac{T_s + \tau_n}{T_L}\right), \quad \text{y} \quad \beta_n = T_L / T_s (1 - \alpha_n). \quad (6.2.9)$$

Asumiendo $|\tau_n| \ll T_s \ll T_L$ y usando una expansión de Taylor, estos parámetros pueden aproximarse por

$$\alpha_n \approx \exp(-T_s / T_L) = \alpha, \quad \text{y} \quad \beta_n \approx (1 - \alpha) T_L / T_s + \tau_n / T_s = \beta + \tau'_n. \quad (6.2.10)$$

A través de todo el capítulo consideramos únicamente entrada constante, $x \in [b_-, b_+]$. Se alimenta la salida del modulador a un filtro peine, y denota su salida por \hat{x} . Como el modulador y el decodificador permiten implementarse separadamente, sin embargo, \hat{x} debe definirse más precisamente por

$$\hat{x} = \frac{1}{N} \sum_{n=1}^N q_n^* = \frac{b}{N} \sum_{n=1}^N \operatorname{sgn}(q_n). \quad (6.2.11)$$

donde $\text{sgn}(\cdot)$ denota la función signo y N es la razón de sobremuestreo.

Estamos interesados en determinar el desempeño de este cuantizador sigma-delta imperfecto, y comparar su desempeño con el sistema ideal. Si un "pequeño" cambio en los parámetros del sistema resulta en degradaciones del desempeño de sistema global también "pequeñas", entonces el sistema de cuantización sigma-delta debe considerarse robusto contra las correspondientes imperfecciones de circuito.

6.3 Derivación del Desempeño

Para simplificar nuestro análisis, supongamos que no existe ni fuga de corriente en el integrador ni imprecisión de temporización, i.e., $T_L = \infty$, y $\tau_n = 0$, o equivalentemente, $\alpha_n = \beta_n = 1$. Se considera sólo un periodo de conversión, i.e., $n=0,1,\dots,N$, y el estado inicial del modulador puede considerarse como el estado final del periodo previo de conversión. Con estas consideraciones en mente, podemos describir el sistema de la Figura 6.1 por las ecuaciones en diferencia

$$u_{n+1} = u_n + G_I(x - q_n), \quad n = 0, 1, \dots, N, \quad (6.3.1)$$

donde

$$q_n = \begin{cases} b_+, & \text{si } u_n \geq h_n; \\ b_-, & \text{si } u_n < h_n, \end{cases}$$

y u_0 es la condición inicial. El problema de desajuste de los niveles de salida se puede tratar substituyendo (6.2.8) en (6.3.1) para encontrar

$$u_{n+1} = u_n + G_I G_q \left[G_q^{-1}(x - \delta) - q_n^* \right],$$

la cual puede expresarse también como

$$u_{n+1} = u_n + G(x' - q_n^*), \quad (6.3.2)$$

con el obvio cambio de variables:

$$x' = G_q^{-1}(x - \delta), \quad y \quad G = G_1 G_q. \quad (6.3.3)$$

La Ecuación (6.3.2) representa un modelo equivalente pero más conveniente y tiene una interpretación intuitiva: el cuantizador binario con niveles de salida malajustados puede reemplazarse por uno sin el desajuste, pero la ganancia y la entrada deben modificarse en el sentido de la Ecuación (6.3.3) para mantener la misma relación del sistema.

Sumando (6.3.2) sobre $n=0,1,\dots,N-1$, y simplificando la suma telescópica, y dividiendo entonces entre N , obtenemos

$$\frac{u_N}{N} = \frac{u_0}{N} + Gx' + \frac{G}{N} \sum_{n=0}^{N-1} q_n^*.$$

la cual, después de resolver para \hat{x} , resulta en

$$\hat{x} = x' + \frac{1}{N} \left[\left[q_N^* - \frac{u_N}{G} \right] - \left[q_0^* - \frac{u_0}{G} \right] \right]. \quad (6.3.4)$$

Ahora estamos listos de derivar una cota superior general a la magnitud del "error de cuantización" $|\hat{x} - x'|$. El teorema siguiente afirma que el máximo de error de cuantización está acotado en magnitud por algo ligeramente mayor que Δ/N , la cota para el cuantizador sigma-delta ideal encontrado en el Capítulo 5.

Teorema 6.1: Dado un modulador sigma-delta imperfecto (imperfecciones 1,2,y 3) con entrada constante $x \in [b_-, b_+]$, and un estado inicial $u_0 \in [G_1(x - b_+) - H, G_1(x - b_-) + H]$. Si el decodificador es un filtro peine con razón de sobremuestreo N , entonces

$$|x - \hat{x}| \leq \frac{\Delta}{N}(1 + \mu). \quad (6.3.5)$$

donde $\Delta = 2b$, y

$$\mu = H/Gb. \quad (6.3.6)$$

Prueba: Observe primero que las suposiciones

$$x \in [b_-, b_+] \quad \text{y} \quad u_0 \in [G_1(x-b_+)-H, G_1(x-b_-)+H]$$

son equivalentes a [cf. (6.3.3) y (6.2.3)]

$$x' \in [-b, b] \quad \text{y} \quad u_0 \in [G(x'-b)-H, G(x'+b)+H],$$

respectivamente. Como (6.3.2) puede escribirse también como

$$u_n = \begin{cases} u_{n-1} + G(x'-b), & \text{si } u_{n-1} \geq h_{n-1}; \\ u_{n-1} + G(x'+b), & \text{si } u_{n-1} < h_{n-1}, \end{cases}$$

entonces de esto y $|h_n| < H$, no es difícil ver que si el estado inicial u_0 está en $[G(x'-b)-H, G(x'+b)+H]$, entonces todos los estados futuros u_n , $n=1, 2, \dots, N$ estarán dentro del mismo intervalo para todo $x' \in [-b, b]$, y entonces u_n/G está en $[x'-b(1+\mu), x'+b(1+\mu)]$.

Si $u_n \geq h_n$, esto implica que $q_n^* = b$, entonces

$$q_n^* - u_n/G = b - u_n/G \geq b - [x'+b(1+\mu)] = (b-x') - b(1+\mu) \geq -b(1+\mu),$$

y usando $|h_n| < H$,

$$q_n^* - u_n/G = b - u_n/G \leq b - h_n/G \leq b + H/G = b(1+\mu).$$

Por otro lado, si $u_n < h_n$, implica $q_n^* = -b$, tenemos

$$q_n^* - u_n/G = -b - u_n/G \leq -b - [x-b(1+\mu)] = -(b+x) + b(1+\mu) \leq b(1+\mu),$$

y

$$q_n^* - u_n/G = -b - u_n/G > -b - h_n/G \geq -b - H/G = -b(1+\mu).$$

En resumen, hemos demostrado que

$$|q_n^* - u_n/G| \leq b(1+\mu), \quad \text{para todo } n = 0, 1, \dots, N,$$

entonces la cota (6.3.5) sigue inmediatamente de (6.3.4) y la prueba termina.

Note que la salida \hat{x} del cuantizador sigma-delta se aproxima a la entrada x' del modelo equivalente y no la entrada actual x del sistema original. Como x' es sólo una amplificación (o atenuación) por G_q^{-1} de la entrada verdadera x recorrida por una pequeña componente de corriente directa $-\delta$, no existe seria degradación de desempeño de sistema, y este efecto puede corregirse (al menos en principio) sumando δ a la entrada x y luego multiplicando por G_q , antes de usar el cuantizador sigma-delta.

Indicamos también que la cantidad $\mu = H/Gb$ puede interpretarse como una medida apropiada del grado de imperfecciones del sistema. Como H es usualmente pequeño comparado con b , y como G es usualmente cercano a la unidad, entonces $\mu \ll 1$, y por lo tanto la cota $(1+\mu)\Delta/N$ es bastante similar a la cota anterior Δ/N .

Para determinar el error de cuantización sigma-delta cuadrático medio, necesitamos una expresión exacta para $x' - \hat{x}$, la cual sea válida para un rango de entradas ligeramente menor, dado por el lema siguiente.

Lema 6.1: Considere el mismo modulador sigma-delta que el del teorema previo, pero ahora la entrada x es limitada a $[b_-(1-2\mu), b_+(1-2\mu)]$, entonces

$$x' - \hat{x} = \frac{\Delta}{N} \left[\langle N\alpha + \frac{u_0 - h_N}{G\Delta} \rangle - \langle \frac{u_0 - h_0}{G\Delta} \rangle + \frac{h_N - h_0}{G\Delta} \right]$$

donde $\alpha = 1/2 + x'/\Delta$.

Una prueba está contenida en [53]. El siguiente teorema es una consecuencia del Lema 6.1, el cual contiene una cota superior no-trivial al error de cuantización sigma-delta medio cuadrático.

Teorema 6.2: Bajo las mismas condiciones del Lema 6.1, asumiendo que la entrada x es una variable aleatoria con una función de densidad de probabilidad suave, y las variables aleatorias h_n , $n=0,1,\dots,N$ son independientes de x , entonces para N grande,

$$E|x' - \hat{x}|^2 \leq \left[\frac{1}{3} + \mu(1+\mu) \right] \left(\frac{\Delta}{N} \right)^2. \quad (6.3.7)$$

Prueba: (Bosquejo.) Reescribase, por simplicidad, la expresión en el Lema 6.1 como

$$x' - \hat{x} = \frac{\Delta}{N} \left[\langle N\alpha + z_0 \rangle - 1/2 + 1/2 - \langle z_1 \rangle + z_2 \right].$$

Como $\langle N\alpha + z_0 \rangle - 1/2$ tiende a ser uniformemente distribuido en $[-1/2, 1/2]$ e independiente de las demás variables, y como $|1/2 - \langle z_1 \rangle| \leq 1/2$, tenemos

$$E|x' - \hat{x}|^2 \cong \left(\frac{\Delta}{N}\right)^2 [1/12 + (1/2 + z_2)^2],$$

y el resultado deseado (6.3.7) sigue del hecho de que $|z_2| \leq 2H/G\Delta = \mu$.

En resumen, hemos derivado cotas superiores a la magnitud del error y al error cuadrático medio, las cuales son proporcionales inversamente a la razón de sobremuestreo y su cuadrado, respectivamente. Estas son las mismas cotas derivadas en el Capítulo 5, aunque las constantes de proporcionalidad son ahora ligeramente mayores. Consecuentemente concluimos que el cuantizador sigma-delta es en efecto robusto contra las imperfecciones correspondientes del circuito, es decir, los errores de umbral, el desajuste de los niveles de salida en el cuantizador binario y el desajuste de ganancia en el integrador.

6.4 Una Estimación

Ahora dirigimos nuestra atención a las imperfecciones restantes, i.e., fuga de corriente en el integrador e imprecisión de temporización. Como un análisis teórico de un modulador sigma-delta con estos efectos no-ideales es sumamente difícil, sólo presentamos una estimación cruda al error de cuantización sigma-delta cuadrático medio, mientras que un resultado más general y convincente podría obtenerse vía técnicas de simulación.

En la derivación siguiente, supongamos que $|\tau_n| \ll T_s \ll T_L$ y usemos las aproximaciones para α_n y β_n dadas por la Ecuación (6.2.10), y consecuentemente la Ecuación (6.2.7) es ahora

$$u_{n+1} = \alpha u_n + (\beta + \tau'_n)(x - q_n), \quad (6.4.1)$$

donde $q_n = b \operatorname{sgn}(u_n)$. Después de una manipulación algebraica encontramos

$$\hat{x} - \alpha x = \frac{e_N - e_0}{N} + (1-\alpha) \frac{1}{N} \sum_{n=0}^{N-1} e_n + \frac{\alpha}{\beta} \frac{1}{N} \sum_{n=1}^N \tau'_n (x - q_n), \quad (6.4.2)$$

donde $e_n = q_n - (\alpha/\beta)u_n$ es granular.

En el miembro derecho de (6.4.2), el primer término es el mismo para un sistema perfecto, mientras el segundo y el tercero representan la degradación del desempeño del sistema causado por la fuga de corriente y los errores de sincronización, respectivamente. El factor de escalamiento adicional α en el lado izquierdo se debe también a la fuga de corriente en el integrador, pero no contribuye a la degradación del desempeño, porque su efecto puede cancelarse multiplicando x previamente por α^{-1} .

Para acotar el error cuadrático medio $E|\alpha x - \hat{x}|^2$, hacemos uso de las siguientes suposiciones de aproximación:

1) el error de cuantizador e_n , $n=1,2,\dots,N$, tiene una autocorrelación empírica que satisface

$$\frac{1}{N} \sum_{n=0}^{N-1-k} e_n e_{n+k} = 0, \text{ para todo } k > p; \quad (6.4.3)$$

2) los errores de temporización τ'_n , $n=1,2,\dots,N$, son variables aleatorias independientes e idénticamente distribuidas con media cero, y existe una constante τ , tal que $|\tau'_n| \leq \tau$. En particular supóngase que τ'_n y q_n son estadísticamente independientes.

Las suposiciones anteriores necesitan una explicación. Hemos demostrado en el Capítulo 4 que el error de cuantizador en un modulador sigma-delta ideal con entrada constante no es ruido blanco. Sin embargo, en el presente caso es razonable suponer que las muestras separadas a una distancia mayor que cierta p tengan una correlación despreciable. Por otro lado la dependencia estadística entre q_n y τ'_n es también débil, ya que $q_{n+1} = \pm b$, y su signo depende principalmente de $\alpha u_n + \beta(x - q_n)$. La suposición (1) implica directamente

$$\begin{aligned} \left| (1-\alpha) \frac{1}{N} \sum_{n=0}^{N-1} e_n \right|^2 &= (1-\alpha)^2 \frac{1}{N} \left[\frac{1}{N} \sum_{n=0}^{N-1} e_n^2 + \sum_{k=1}^{p-1} \frac{2}{N} \sum_{n=0}^{N-1-k} e_n e_{n+k} \right] \\ &\leq (1-\alpha)^2 (2p-1) b^2 / N. \end{aligned} \quad (6.4.4)$$

La suposición (2) implica que

$$E_{\tau} \left| \frac{1}{N} \sum_{n=1}^N \tau'_n (x - q_n) \right|^2 = \frac{1}{N} \left\{ \frac{1}{N} \sum_{n=1}^N E_{\tau} (\tau'_n)^2 (x - q_n)^2 \right\} \leq \frac{1}{N} \tau^2 \left[(x-b)^2 + (x+b)^2 \right] / 2$$

$$\leq 2\tau^2 b^2 / N. \quad (6.4.5)$$

La cota para el cuadrado del primer término en el miembro derecho de (6.4.2) es simplemente $4b^2/N^2$. Despreciando los demás términos en la expansión de $E|\alpha x - \hat{x}|^2$ (puesto que son insignificantes comparado con (6.2.4) y (6.2.5) cuando N es suficientemente grande), el error medio cuadrático de cuantización sigma-delta es entonces

$$E|\alpha x - \hat{x}|^2 \lesssim \frac{b^2}{N} \left[\frac{4}{N} + (1-\alpha)^2 (2p-1) + 2\tau^2 \right], \quad (6.4.6)$$

y esta es nuestra estimación deseada.

Comparando esta cota con el desempeño ideal encontrado en el Capítulo 5, que es de orden $O(N^{-2})$, se puede ver que un cambio "pequeño" de parámetros del sistema debido a la fuga de corriente y errores de temporización puede producir un efecto "grande" sobre el compromiso de tasa-distorsión entre error de cuantización y la razón de sobremuestreo. Por eso no podemos concluir que el modulador sigma-delta sea robusto con respecto a estos efectos no-ideales. Sin embargo, el error cuadrático medio decrece de todas maneras a cero cuando la razón de sobremuestreo tiende a infinito con una velocidad de $O(N^{-1})$, y por tanto si alcanza cualquier error medio cuadrático a un costo de requerir razones de sobremuestreo mucho más (desesperadamente a veces) grandes.

Finalmente, conjeturamos que el desempeño puede degradar aún más si las cinco imperfecciones descritas en la Sección 6.2 se toman en cuenta al mismo tiempo. Por ejemplo, el error de umbral en el cuantizador puede no cancelarse en la presencia de un integrador con fuga. No trataremos el problema, sin embargo, por ser los detalles extremadamente confusos, sino que sugerimos el uso de técnicas de simulación.

CAPITULO 7 COMENTARIOS SOBRE INVESTIGACION FUTURA

En esta disertación, hemos presentado una descripción teórica comprehensiva de algunas propiedades básicas de modulación sigma-delta sobremuestreada. Todos los tópicos considerados fueron desarrollados en una forma unificada. Es justo decir que las cuestiones en la introducción (Capítulo 1) fueron contestadas satisfactoriamente, se ganó una visión profunda del comportamiento de los moduladores sigma-delta. Aunque nos hemos enfocado principalmente a los aspectos de análisis, nuestro resultado es seguramente informativo y útil también para el diseño y la implementación, puesto que la interrelación entre estas áreas es bastante fuerte. Esta teoría proporciona también un ejemplo interesante de aplicaciones novedosas de las técnicas matemáticas avanzadas a problemas específicos de ingeniería.

La metodología y las técnicas usadas y/o desarrolladas en esta tesis pueden extenderse para analizar una clase amplia de sistemas de cuantización, siempre y cuando se puede encontrar una expresión, tal que las funciones características de dimensión uno y dos puedan ser evaluadas. En efecto, Iwersen [36] llevó a cabo un análisis del ruido de cuantización en sistemas de modulación delta usando esencialmente el mismo enfoque.

En esta investigación, hemos limitado el análisis a moduladores sigma-delta solamente, un caso especial de los codificadores interpolativos. Todos los moduladores sigma-delta tienen una característica común: el filtrado lineal en el lazo de retroalimentación consiste en integradores ideales. Una dirección promisoría de investigación futura es desarrollar un análisis exacto similar para los codificadores de interés particular como el que se muestra en la Figura 7.1, donde el integrador ideal $(z-1)^{-1}$ se reemplaza por un filtro general $H(z)$ con por lo menos una unidad de retraso.

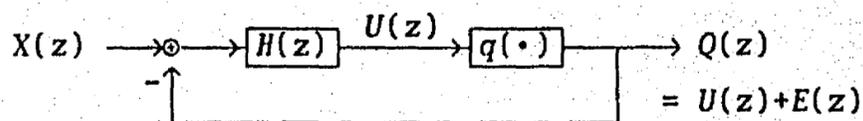


Figura 7.1 Codificador Interpolativo de Lazo Simple

En el dominio de transformada z , la salida del sistema $Q(z)$ está relacionada a la entrada $X(z)$ y el error de cuantizador $E(z)$ por

$$Q(z) = F(z) X(z) + G(z) E(z),$$

donde

$$F(z) = \frac{H(z)}{1+H(z)}, \quad \text{y} \quad G(z) = \frac{1}{1+H(z)},$$

son las funciones de transferencia para señal y ruido, respectivamente. Esta relación sencilla proporciona información sobre el principio general de configurar el ruido de cuantización: El filtro $H(z)$ se debe escoger de modo que (1) la función de transferencia para señal $F(z)$ tenga un espectro plano en las frecuencias bajas, pasando la señal de entrada $X(z)$ o las componentes deseadas de ella, y (2) la función de transferencia para ruido $G(z)$ tenga una forma de filtro paso-alta, atenuando efectivamente el ruido de cuantización $E(z)$ en las frecuencias bajas.

Como $H(z)$ no está restringido de ser un integrador, conjeturamos que un codificador interpolativo con un filtro lineal adecuado puede proporcionar mejor desempeño de sistema que el modulador sigma-delta de lazo simple. Por lo que, un problema interesante es determinar cuál es el compromiso de desempeño que este codificador puede ofrecerse y la mejor selección de $H(z)$ en alguna clase de filtros, por ejemplo, de la forma de

$$H(z) = \frac{G}{(1-z)} \frac{z-b}{z-a}.$$

En el desarrollo del Capítulo 3 hemos usado extensamente la suposición de que el cuantizador interno no se sobrecarga y encontramos las condiciones para prevenir la sobrecarga en los cuantizadores. En la práctica es común usar sólo cuantizadores binarios para implementar moduladores sigma-delta multilazo, y estos sistemas no tienen la propiedad de no-sobrecarga. El resultante desempeño global del sistema aún puede ser aceptable si el cuantizador no se sobrecarga demasiado seguido y si el efecto no es catastrófico. Sin embargo, nuestra predicción teórica recomienda fuertemente el uso de cuantizadores multinivel en los moduladores multilazos (e.g., 4 niveles en 2 lazos y 8 niveles en 3 lazos) para prevenir la sobrecarga, particularmente cuando estos convertidores analógico/digital se usan como

componentes de sistemas de comunicaciones de alto desempeño, ya que este costo extra es muy moderado. No está claro si el enfoque de análisis exacto puede modificarse para tratar también ruido de sobrecarga. (Un remedio es usar un limitador dentro del lazo para estabilizar el sistema, ver [45].)

Quedan todavía dos cuestiones abiertas de considerable interés, las cuales están relacionadas con el modelo de entrada. Primero, en la derivación de las propiedades estadísticas asintóticas hemos limitado el estudio a entradas constante y senoidal solamente, aunque el análisis en el dominio del tiempo no usó estas suposiciones. Una generalización posible es extender el análisis asintótico a aquellas señales que pueden expresarse como una suma finita de senoides

$$x_n = \sum_1 a_1 \cos(\omega_1 n + \theta_1),$$

tales que sus amplitudes estén restringidas aún en el rango apropiado de entrada, i.e.,

$$|x_n| \leq \sum_1 a_1 \leq b.$$

La esperanza es que los resultados desarrollados en el Capítulo 4 para una senoide pura se mantengan también para senoides múltiples, y entonces para una clase bastante general de entradas que pueden aproximarse razonablemente por sumas finitas de senoides.

Segundo, excepto en pocos casos, no hemos considerado una amplia clase de entradas que consiste en procesos estocásticos, así que una dirección importante para investigación futura es examinar si puede desarrollarse una teoría paralela. Ejemplos interesantes incluyen procesos independientes e idénticamente distribuidos (iid) y gaussianos. Otro ejemplo de interés particular es de la forma siguiente:

$$x'_n = x_n + v_n,$$

donde x_n es la señal de entrada original y v_n es un proceso iid llamado proceso de perturbación, que es independiente de x_n y que la amplitud de la suma x'_n satisface todavía una restricción apropiada. Aunque este ruido de entrada es generalmente indeseable, se ha usado en la práctica como un medio para mejorar la calidad subjetiva de voz e imagen cuantizada. Además, se cree

que un proceso de perturbación bien escogido puede forzar a que el ruido de cuantizador sea blanco y por lo tanto se simplifica el análisis.

Como la teoría desarrollada aquí está principalmente enfocada a las aplicaciones en conversión analógico/digital, el aspecto de ciclos límite no ha tratado con detalle, ya que este fenómeno no es típico. Es importante, sin embargo, cuando modulación sigma-delta se aplica a conversión digital/analógico, y entonces debe comprenderse cabalmente. Sería interesante extender los resultados reportados por Friedman [17] y por Gray [24] para moduladores más complejos.

En la evaluación del desempeño global de sistemas de cuantización sigma-delta (Capítulo 5), hemos estudiado con detalle varios filtros decodificadores *lineales e invariantes en el tiempo*, y encontramos sus compromisos de desempeño entre el error cuadrático medio y la razón de sobremuestreo. Una pregunta natural es entonces: ¿puede mejorar el desempeño de sistema, si la restricción de linealidad e invariancia de los filtros de decodificación es removida?

El problema difícil pero importante de robustez contra las imperfecciones de circuito fue tratado en el Capítulo 6. Esto representa un primer paso valioso hacia la solución final, aunque estos resultados en el Capítulo 6 son restrictivos porque consideramos sólo el cuantizador sigma-delta más simple y sólo entrada constante. Pensamos que este tópico es suficientemente importante para merecer un esfuerzo adicional. Mientras que es obvio que los moduladores más complejos son inherentemente más susceptibles a imperfecciones de circuito, no se dispone de ningún resultado cuantitativo para describir el balance entre el mejoramiento y la degradación del desempeño de sistema debidos al incremento de la complejidad de sistema. Investigaciones futuras deben enfocarse sobre moduladores multilazo o multietapa, pero si nuestro método puede extenderse para ese propósito es un problema interesante por sí mismo.

Como una observación final, indicamos que los conceptos y las técnicas en la caracterización del comportamiento estadístico de modulación sigma-delta (Capítulo 4) puede aplicarse para analizar y diseñar generadores de números (pseudo) aleatorios con una medida específica de aleatoriedad, los cuales son ampliamente usados en las técnicas de simulación.

REFERENCIAS

- [1] B.P.Agrawal and K.Shenoi (1983), "Design methodology for $\Sigma\Delta$," *IEEE Transactions on Communications*, Vol.COM-31, pp.360-369.
- [2] S.H.Ardalan and J.J.Paolos (1987), "An analysis of nonlinear behavior in delta-sigma modulators," *IEEE Transactions on Circuits and Systems*, Vol.CAS-34, pp.593-603.
- [3] S.H.Ardalan (1988), "Analysis of delta-sigma modulators with bandlimited gaussian inputs," en *Proceedings of 1988 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.1866-1869.
- [4] W.R.Bennett (1948), "Spectra of quantized signals," *Bell Systems Technical Journal*, Vol.27, pp.446-472.
- [5] H.Bohr (1947), *Almost Periodic Functions*, Chelsea, New York. (Traducido por H.Cohn.)
- [6] J.C.Candy (1974), "A use of limit cycle oscillations to obtain robust analog-to-digital converters," *IEEE Transactions on Communications*, Vol.COM-22, pp.298-305.
- [7] J.C.Candy (1985), "A use of double integration in sigma delta modulation," *IEEE Transactions on Communications*, Vol.COM-33, pp.249-258.
- [8] J.C.Candy (1986), "Decimation for sigma delta modulation," *IEEE Transactions on Communications*, Vol.COM-34, pp.72-76.
- [9] J.C.Candy, Y.C.Ching, and D.S.Alexander (1974), "Using triangular weighted interpolation to get 13-bit PCM from $\Sigma\Delta$ modulator," *IEEE*

Transactions on Communications, Vol.COM-24, pp. 1268-1275.

- [10] J.C.Candy and O.J.Benjamin (1981), "The structure of quantization noise from sigma-delta modulation," *IEEE Transactions on Communications*, Vol.COM-29, pp.1316-1323.
- [11] J.C.Candy and A-N.Huynh (1986), "Double interpolation for digital-to-analog conversion," *IEEE Transactions on Communications*, Vol.COM-34, pp.77-81.
- [12] W.Chou, P.W.Wong, and R.M.Gray (1988), "Multi-stage sigma-delta quantization," *IEEE Transactions on Information Theory*, por aparecer.
- [13] A.G.Clavier, P.F.Panter, and D.D.Grieg (1947.a), "Distortion in a pulse count modulation system," *AIEE Transactions*, Vol.66, pp.989-1005.
- [14] A.G.Clavier, P.F.Panter, and D.D.Grieg (1947.b), "PCM distortion analysis," *Electrical Engineering*, pp.1110-1122.
- [15] W.B.Davenport, Jr. and W.L.Root (1958), *Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York.
- [16] W.Feller (1968), *An Introduction to Probability and Its Applications*, Vol.1, Wiley, New York. 3ra ed.
- [17] V.Friedman (1988), "The structure of the limit cycles in sigma delta modulation," *IEEE Transactions on Communications*, Vol.COM-36, pp.972-979.
- [18] H.Furstenberg (1981), *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, Princeton.
- [19] A.Gersho (1978), "Principles of quantization," *IEEE Transactions on Circuits and Systems*, Vol.CAS-25, pp.427-436.

- [20] C.H.Giancarlo and C.G.Sodini (1986), "A slop adaptive delta modulator for VLSI signal processing systems," *IEEE Transactions on Circuits and Systems*, Vol.CAS-33, pp.51-58.
- [21] H.Gish and J.N.Pierce (1968), "Asymptotically efficient quantizing," *IEEE Transactions on Information Theory*, Vol.IT-14, pp.676-683.
- [22] I.S.Gradshteyn and I.M.Ryzhik (1980), *Table of Integrals, Series and Products*, Academic Press, New York, edición corregida y aumentada.
- [23] R.M.Gray (1987), "Oversampled sigma-delta modulation," *IEEE Transactions on Communications*, Vol.COM-35, pp.481-489.
- [24] R.M.Gray (1988), "Spectral analysis of quantization noise in a single-loop sigma-delta modulator with dc input," *IEEE Transactions on Communications*, por aparecer.
- [25] R.M.Gray, W.Chou, and P.W.Wong (1988), "Quantization noise in single-loop sigma-delta modulation with sinusoidal inputs," *IEEE Transactions on Communications*, por aparecer.
- [26] R.M.Gray (1989), "Quantization noise spectra," sometido a *IEEE Transactions on Information Theory*.
- [27] F.J.Hahn (1964), "On affine transformations of compact Abelian groups," *American Journal of Mathematics*, Vol.58, pp.428-446. Errata en Vol.59, pp.463-464.
- [28] M.W.Hauser and R.W.Brodersen (1986), "Circuit and technology considerations for MOS delta-sigma A/D converters," en *Proceedings of 1986 IEEE International Symposium on Circuits and Systems*.
- [29] T.Hayashi, Y.Inabe, K.Uchimura, and T.Kimura (1986), "A multistage delta-sigma modulator without double integration loop," en *ISSCC Digest of Technical Papers*, pp.182-183.

- [30] N.He, F.Kuhlmann, and A.Buzo (1988.a), "Double-loop sigma-delta modulation with dc input," *IEEE Transactions on Communications*, por aparecer.
- [31] N.He, F.Kuhlmann, and A.Buzo (1988.b), "Oversampled high-order sigma-delta modulation," en *Proceedings of 31st Midwest Symposium on Circuits and Systems*.
- [32] N.He, A.Buzo, and F.Kuhlmann (1988a), "Multiloop sigma-delta quantization: time-domain analysis," manuscrito.
- [33] N.He, A Buzo, and F. Kuhlmann (1988b), "Multi-loop sigma-delta quantization: spectral analysis," en *Proceedings of 1988 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.1870-1873.
- [34] H.Inose, Y.Yasuda, and J.Murakami (1962), "A telemetering system by code modulation — Σ - Δ modulation," *IRE Transactions on Space Electronics and Telemetry*, Vol.SET-8, pp.204-209.
- [35] H.Inose and Y.Yasuda (1963), "A unit bit coding method by negative feedback," en *Proceedings of the IEEE*, Vol.51, pp.1524-1535.
- [36] J.E.Iwersen (1969), "Calculated quantizing noise of single integration delta-modulation coders," *Bell Systems Technical Journal*, Vol.48, pp.2359-2389.
- [37] D.E.Knuth (1968), *The Art of Computer Programming*, Vol.2: *Semi-Arithmetical Algorithms*, Addison-wesley.
- [38] W.L.Lee and C.G.Sodini (1987), "A topology for high order interpolative coders," en *Proceedings of 1986 IEEE International Symposium on Circuits and Systems*.

- [39] Y. Matsuya, K. Uchimura, A. Iwata, T. Kobayashi, and M. Ishikawa (1988), "A 16b oversampling conversion technology using triple integration noise shaping," en *Proceedings of 1987 IEEE International Solid-State Circuits Conference*, pp. 48-49.
- [40] K. Petersen (1983), *Ergodic Theory*, Cambridge University Press, Cambridge.
- [41] A.G Portnikov (1966), "Ergodic Problems in the theory of congruences and of diophantine approximations," en *Proceedings of the Steklov Institute of Mathematics*, No.82, pp.1-128. Traducción por American Mathematics Society.
- [42] S.O. Rice (1954), "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, editado por N. Wax, pp.133-294, Dover, New York. (Reimpresión de *Bell Systems Technical Journal*, Vol.23 and 24.)
- [43] G.R. Ritchie (1977), "High order interpolative analog to digital converters," *Ph.D. Dissertation*, University of Pennsylvania.
- [44] G.R. Ritchie, J.C. Candy, and W.H. Ninke (1974), "Interpolative digital-to-analog converters," *IEEE Transactions on Communications*, Vol.COM-22, pp.1797-1806.
- [45] E.F. Stikvoort (1988), "Some remarks on the stability and performance of the noise shaper or sigma-delta modulator," *IEEE Transactions on Communications*, Vol.COM-36, pp.1157-1162.
- [46] S.K. Tewsbury and R.W. Hallock (1978), "Oversampled, linear predictive and noise-shaping coders of order $N > 1$," *IEEE Transactions on Circuits and Systems*, Vol.CAS-25, pp.436-447.
- [47] K. Uchimura, T. Hayashi, T. Kimura, and A. Iwata (1986), "VLSI A-to-D and D-to-A converters with multistage noise shaping modulators," en

Proceedings of 1986 IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.1545-1548.

- [48] K.Uchimura, T.Hayashi, T.Kimura, and A.Iwata (1988), "Oversampling A-to-D and D-to-A converters with multistage noise shaping modulators," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol.36, pp.1899-1905.
- [49] H.Weyl (1916), "Über die gleichverteilung von zahlen mod eins," *Mathematische Annalen*, Vol.77, pp.313-352.
- [50] B.Widrow (1956), "A study of rough amplitude quantization by means of Nyquist sampling theory," *IRE Transactions on Circuit Theory*, Vol.CT-3, pp.266-276.
- [51] P.W.Wong and R.M.Gray (1988), "Two stage sigma-delta modulation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, por aparecer.
- [52] C-M.Zeng, N.He, and F.Kuhlmann (1988a), "On the robustness of sigma-delta quantizers against circuit imperfections," en *Proceedings of 31st Midwest Symposium on Circuits and Systems*.
- [53] C-M.Zeng, N.He, and F.Kuhlmann (1988b), "Performance analysis of sigma-delta quantizers in the presence of circuit imperfections," sometido a *IEEE Transactions on Communications*.