

**ALGORITMOS DE DITHERING CON TIEMPO DE
PROCESAMIENTO REDUCIDO**

Por

Samira E Ortiz-Rodríguez

Tesis presentada en cumplimiento parcial de los requisitos para el grado de

MAESTRÍA EN CIENCIAS

en

INGENIERÍA ELÉCTRICA

UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGÜEZ
2016

Aprobado por:

Shawn Hunt, PhD.
Presidente, Comité Graduado

Fecha

Domingo Rodríguez, PhD.
Miembro, Comité Graduado

Fecha

Dámaris Santana Morant, PhD.
Miembro, Comité Graduado

Fecha

Isabel Rios, MBA.
Representante de Estudios Graduados

Fecha

José Colom Ustáriz, PhD.
Director del Departamento
Ingeniería Eléctrica y Computadoras

Fecha

Resumen de Disertación Presentado a Escuela Graduada
de la Universidad de Puerto Rico como requisito parcial de los
Requerimientos para el grado de Maestría en Ciencias

**ALGORITMOS DE DITHERING CON TIEMPO DE PROCESAMIENTO
REDUCIDO**

Por

Samira E Ortiz-Rodríguez

2016

Consejero: Shawn D. Hunt
Departamento: Ingeniería Eléctrica y Computadoras

En esta investigación se trabajó con el algoritmo Pendiente Máxima (SD o *Steepest Descent*, por sus siglas en inglés) para hallar una señal de dither (señal aleatoria) que hace que el ruido de cuantización (diferencia entre la señal de entrada y de salida del cuantizador) sea blanco y no esté correlacionado con la señal original. El tiempo de procesamiento resultó ser menor al registrado por la investigación de punta. El algoritmo propuesto resultó ser menos complejo, empleó 99.2% menos multiplicaciones y 17.9% menos sumas que el algoritmo de Benitez-Quiroz [2]. El algoritmo de Pendiente Máxima tuvo un tiempo de procesamiento un 83.02% menor que el Benitez-Quiroz.

Se usa además la autocorrelación circular del error de cuantización para verificar que sea un ruido blanco. Se obtienen valores de SNR mayor al que se obtiene con el dither triangular.

Abstract of Dissertation Presented to the Graduate School
of the University of Puerto Rico in Partial Fulfillment of the
Requirements for the Degree of Master of Science

DITHERING ALGORITHMS WITH REDUCED PROCESSING TIME

By

Samira E Ortiz-Rodríguez

2016

Chair: Shawn D. Hunt

Major Department: Electrical and Computer Engineering

This research worked with the Steepest Descent (SD) algorithm to find a dither signal that ensures that the quantization noise (difference between the input signal and output quantizer) is white and is uncorrelated with the original signal. The processing time was less than state of art algorithms. The proposed algorithm proved to be less complex, used 99.2% fewer multiplications and 17.9% fewer sums than the Benitez-Quiroz[2]. SD algorithm processing time was 83.02% less than the Benitez-Quiroz algorithm [2]

Circular autocorrelation was used to verify that the quantization error is white noise. Higher SNR values than triangular dither were obtained.

TABLA DE CONTENIDO

ALGORITMOS DE DITHERING CON TIEMPO DE PROCESAMIENTO REDUCIDO.....	ii
DITHERING ALGORITHMS WITH REDUCED PROCESSING TIME	iii
TABLA DE CONTENIDO	iv
LISTA DE FIGURAS	vi
INDICE DE SIMBOLOS	viii
CAPÍTULO 1 INTRODUCCIÓN.....	1
CAPÍTULO 2 REVISIÓN DE LITERATURA	9
2.1. Cuantización.....	9
2.2. Dithering	11
2.3. Resumen	15
CAPÍTULO 3 METODOLOGIA.....	16
3.1. Cuantización.....	16
3.1.1. Función de densidad de potencia de la señal cuantizada	19
3.2. Ruido de cuantización.....	21
3.2.1. Estadísticas del ruido de cuantización.....	23
3.3. Dither	29
3.3.1 Error total	29
3.3.2. FDP del error total	31
3.3.3. Dithering Adaptativo	32
3.4. Algoritmo de pendiente máxima (SD o Steepest descent, por sus siglas en inglés).....	36
3.4.1 Generalidades	36
3.4.2 Planteamiento del problema	37
3.5. Prueba de Ljung-Box	39

3.6. Operaciones de punto flotante.....	41
3.7. Resumen	43
CAPÍTULO 4 EXPERIMENTOS Y RESULTADOS	44
4.1. Niveles de cuantización	44
4.1.1. Señales de audio	44
4.1.2. Comparación de tiempo.....	53
4.2. Varianza.....	54
4.3. Imágenes	56
4.4. Resumen	57
CAPÍTULO 5 CONCLUSIONES Y TRABAJOS FUTUROS.....	59
REFERENCIAS	62

LISTA DE FIGURAS

Figura 1-1 (a) Señal original (b) Espectro de la señal original (c) Señal cuantizada (d) Espectro de la señal cuantizada.....	4
Figura 1-2 (a) Señal cuantizada y (b)Espectro de la señal cuantizada después de agregar dither.....	5
Figura 1-3 Espectro de las señales de error de cuantización con (a) dither y (b) sin dither.	5
Figura 1-4 (a)FDP del ruido uniforme 1 (b) FDP del ruido uniforme 2 (c)FDP del ruido triangular	7
Figura 2-1	10
Figura 2-2	14
Figura 3-1 Señal analógica y señal cuantizada.....	17
Figura 3-2 Diagrama de bloques del cuantizador.	17
Figura 3-3 Funciones de transferencia de cuantización (a) Función piso (b) Redondeo	18
Figura 3-4 FDP de la señal cuantizada.....	20
Figura 3-5 Formación de FDP de la señal cuantizada (a) Ventana de entrada (b) Tren de funciones delta de Dirac.	20
Figura 3-6 Diagrama de bloques: error de cuantización	22
Figura 3-7 Error de cuantización.....	22
Figura 3-8 Construcción de la FDP del ruido de cuantización.	24
Figura 3-9 Autocorrelación del error de cuantización (a) sin dither (b) con dither triangular.....	27
Figura 3-10 Esquemas de cuantización (a) Sin dither (b) Con dither sustractivo (c) Con dither no sustractivo	30
Figura 3-11 Pendiente máxima.	36
Figura 4-1 Espectro de la (a) señal de audio sintética y la (b) señal real cuantizada x_q para $N=8$	46

Figura 4-2 Autocorrelación para distintos niveles de cuantización de la señal sintética sin ruido de dither (a) $N = 10$ y (b) $N = 4$, con ruido triangular para (c) $N = 10$ y (d) $N = 4$, y con dither de SD (e) $N = 10$ y (f) $N = 4$	47
Figura 4-3 Autocorrelación para distintos niveles de cuantización de la señal de audio real para $N=4$ para (a) Señal sin dither (b) Señal con dither triangular	48
Figura 4-4 Varianza del error de cuantización ϵ para distintos niveles de cuantización (a) Señal sintética sin dither (b) Señal sintética con dither (c) Señal real sin dither (d) Señal real con dither	49
Figura 4-5 Valor de p para distintos niveles de cuantización (a) Señal sintética sin dither (b) Señal sintética con dither (c) Señal real sin dither (d) Señal real con dither.....	50
Figura 4-6 SNR para distintos niveles de cuantización para (a) Señal sintética (b) Señal real.....	51
Figura 4-7 Error de cuantización para (a) señal sintética cuando $N=10$ y (b) señal real cuando $N=8$	52
Figura 4-8 Comparación de varianzas obtenidas con los distintos dithers para señal de audio sintética	54
Figura 4-9 Comparación de varianzas obtenidas con los distintos dithers para señal de audio real.....	55
Figura 4-10 (a) Imagen original (b) Imagen cuantizada a 4 bits (c) Imagen con dither triangular (d) Imagen con dither SD	57

INDICE DE SIMBOLOS

x_q	señal cuantizada
$x[n]$	señal original
$Q(\cdot)$	operación de cuantización
$\varepsilon[n]$	error de cuantización
$E\{\cdot\}$	valor esperado
SNR	relación señal a ruido
Δ	tamaño del escalón
$\rho(\cdot)$	función de probabilidad de densidad
σ^2	varianza
$R[\cdot]$	autocorrelacion
τ	diferencia en tiempo de dos variables aleatorias
ϑ	error total
d	señal de dither
p	valor p (<i>p-value</i>)
N	numero de bits después de cuantizar la señal

Capítulo 1.

INTRODUCCIÓN

Las señales digitales son discretas en tiempo y en amplitud, y se usan ampliamente en numerosas aplicaciones. Si una señal es discreta en tiempo y en amplitud, se denomina señal digital. El proceso de conversión de una señal analógica a digital se lleva a cabo a través de dos procedimientos, el primer procedimiento se efectúa en tiempo y se denomina muestreo. Para definir el segundo procedimiento que consiste en convertir la señal analógica en amplitud se usará el término cuantizar. Típicamente este término se traduce cuantificar en el diccionario de la Real Academia Española (RAE), debido a que hay dos términos en inglés, “quantize” y “quantify”, con la misma traducción al español. Estas dos termologías se refieren a dos procedimientos completamente distintos. Mientras que “quantify” o en español cuantificar, se refiere según RAE a expresar numéricamente una magnitud de algo,

CAPÍTULO 1. INTRODUCCIÓN

“quantize” consiste en restringir una cantidad a un conjunto de valores finitos que son caracterizados por números cuánticos.

Dentro de los métodos más básicos para llevar a cabo el proceso de cuantización se encuentran el truncamiento, que consiste en descartar los bits menos significativos, también encontramos el redondeo (Figura 3-3), que consiste en llevar una señal a la representación digital más cercana y la función de piso con la que la señal se representa con el entero más cercano que se sea menor o igual al valor de la señal analógica. La señal cuantizada se denomina:

$$x_q = Q(x). \quad (1.1)$$

Donde $Q(x)$ es la operación de cuantización y x es la señal que se desea cuantizar.

Hay aplicaciones que tienen ciertas especificaciones para su almacenamiento o transmisión requiriendo que se disminuya la precisión en amplitud de las señales discretas. En estas aplicaciones se reduce el número de bits y por lo tanto, la resolución en amplitud de una señal digital.

Al cuantizar una señal se puede degradar la señal original e introducir un error de cuantización. El error de cuantización $\varepsilon[n]$ es la diferencia entre la señal original y la señal cuantizada de salida:

$$\varepsilon[n] \triangleq Q(x[n]) - x[n], \quad (1.2)$$

el símbolo \triangleq para indicar igualdad por definición.

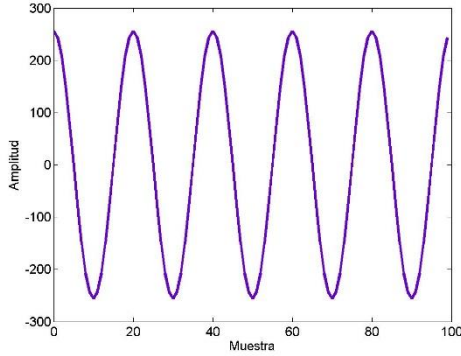
Cuando la señal presenta un cambio de amplitud grande entre una muestra y la próxima, se asume que el ruido de cuantización es blanco y no está correlacionado

CAPÍTULO 1. INTRODUCCIÓN

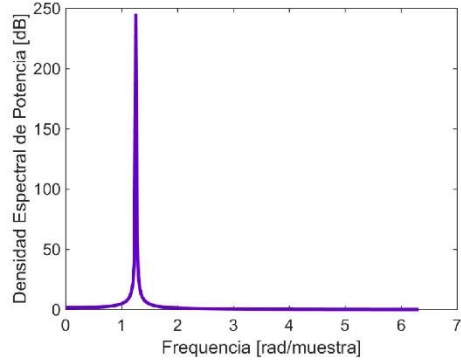
con la señal de entrada. El ruido blanco se caracteriza por tener una densidad espectral de potencia (DEP) que contiene todas las frecuencias a una misma potencia. Sin embargo, cuando el cambio de amplitud entre una muestra y la próxima es pequeño, la cuantización puede tener armónicos indeseados si se observa su espectro de frecuencia a través de la transformada de Fourier.

CAPÍTULO 1. INTRODUCCIÓN

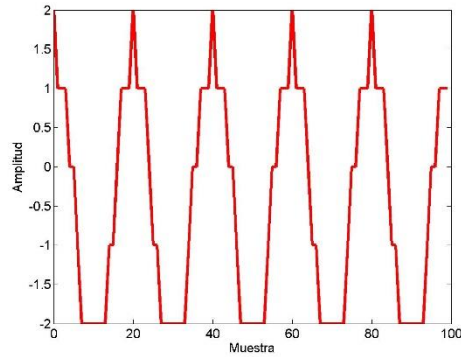
(a)



(b)



(c)



(d)

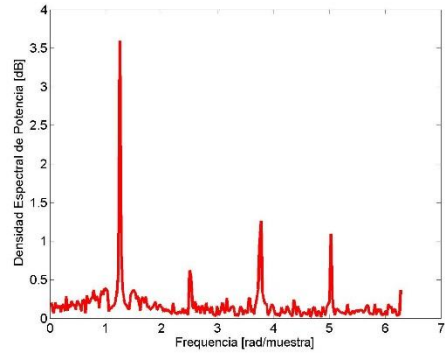


Figura 1-1 (a) Señal original (b) Espectro de la señal original (c) Señal cuantizada (d) Espectro de la señal cuantizada.

El espectro de la señal original no muestra ningún armónico, sólo tiene una frecuencia principal, que es la frecuencia de la señal coseno, sin embargo, en la (d) se observan armónicos.

Para evitar estos armónicos se le puede sumar a la señal de entrada una señal de ruido d denominada dither (Figura 1-2):

$$x_d = x + d, \quad (1.3)$$

x_d representa la señal de entrada con dither agregado.

(a)

(b)

CAPÍTULO 1. INTRODUCCIÓN

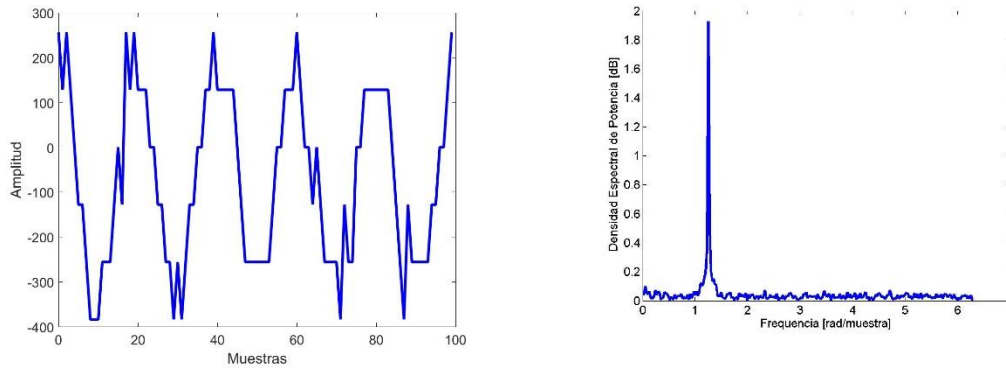


Figura 1-2 (a) Señal cuantizada y (b) Espectro de la señal cuantizada después de agregar dither. No se presentan armónicos indeseados, solo se ve un impulso en la frecuencia equivalente a la frecuencia del coseno.

El propósito de dither es asegurar que el error de cuantización sea un ruido blanco.

En la Figura 1-3 los espectros de los errores de cuantización se presentan.

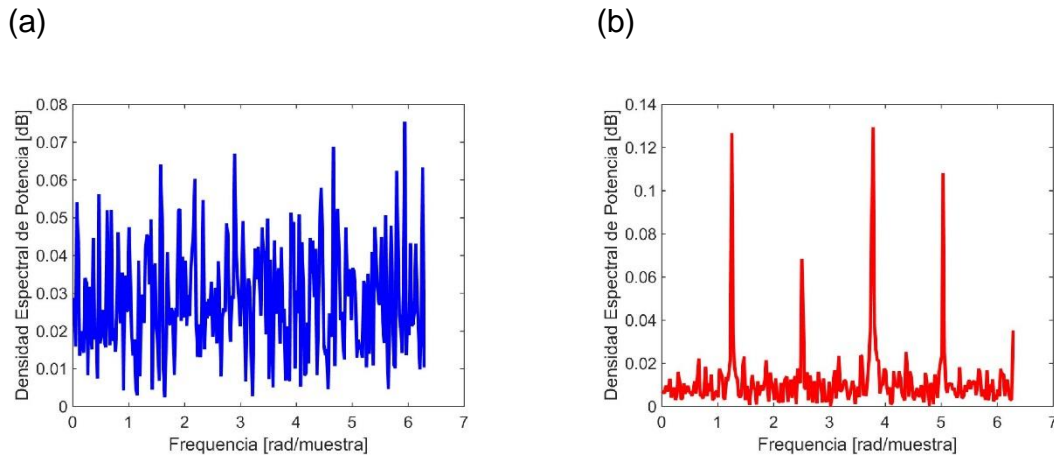


Figura 1-3 Espectro de las señales de error de cuantización con (a) dither y (b) sin dither. Cuando se agrega dither el error de cuantización oscila entre valores pequeños entre 0 y 0.07, mostrando una potencia casi constante mientras que el espectro del ruido de cuantización cuando no se agrega dither presenta una densidad espectral de potencia variable.

La potencia P de la señal $x[n]$ es:

$$P = \lim_{T \rightarrow \infty} \frac{1}{2K + 1} \sum_{n=-K}^K |x(n)|^2. \quad (1.4)$$

Asumiendo que el proceso $x[n]$ es ergódico entonces:

CAPÍTULO 1. INTRODUCCIÓN

$$E\{x^2[n]\} = \lim_{T \rightarrow \infty} \frac{1}{2K+1} \sum_{n=-K}^K |x(n)|^2. \quad (1.5)$$

Donde $E\{\cdot\}$ representa el valor esperado y $E\{x^2[n]\}$ es la potencia total en $x[n]$.

Ahora, la varianza de la señal $x[n]$ es:

$$\sigma_x^2 = E\{|x[n] - \mu_x|\} = E\{x^2[n]\} - E^2\{x[n]\}. \quad (1.6)$$

Entonces, se tiene que:

$$E\{x^2[n]\} = \sigma_x^2 + \mu_x^2, \quad (1.7)$$

donde σ_x^2 es la potencia AC y μ_x^2 es la potencia DC.

La relación entre la potencia de la señal original y la potencia del ruido, se denomina relación señal/ruido (SNR o *Signal to noise ratio*, por sus siglas en inglés). La Ecuación (1.4) y (1.5) describen el SNR y el SNR en decibeles respectivamente. Agregar una señal de dither implica la suma de un ruido adicional, de manera que la el SNR disminuye, representado así, una de las desventajas principales de la estrategia.

El dither adaptativo es un dither que se determina basado en la señal de entrada. En esta investigación se diseña un dither adaptativo que depende de la señal de entrada y tiene varianza constante.

$$SNR = \frac{P_{señal}}{P_{ruido}}, \quad (1.8)$$

$$SNR_{dB} = 10 \log\left(\frac{P_{señal}}{P_{ruido}}\right). \quad (1.9)$$

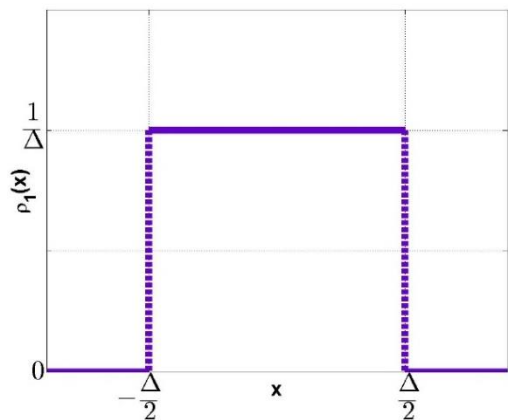
Donde $P_{señal}$ es la potencia de la señal y P_{ruido} es la potencia del ruido. En el modelo clásico de dithering uno de los ruidos más comúnmente usado es la función de

CAPÍTULO 1. INTRODUCCIÓN

densidad de probabilidad (FDP) uniforme. Una distribución uniforme se caracteriza porque todos los intervalos de una misma longitud en la distribución de su rango $(-\frac{\Delta}{2}, \frac{\Delta}{2})$ tienen la misma probabilidad. La Ecuación (1.10) y las Figura 1-4(a) describen la función de densidad para una distribución uniforme.

$$\rho_x(x) = \begin{cases} \frac{1}{\Delta} & \text{para } x \in (-\frac{\Delta}{2}, \frac{\Delta}{2}) \\ 0 & \text{para } x \notin (-\frac{\Delta}{2}, \frac{\Delta}{2}) \end{cases} \quad (1.10)$$

(a)



(b)

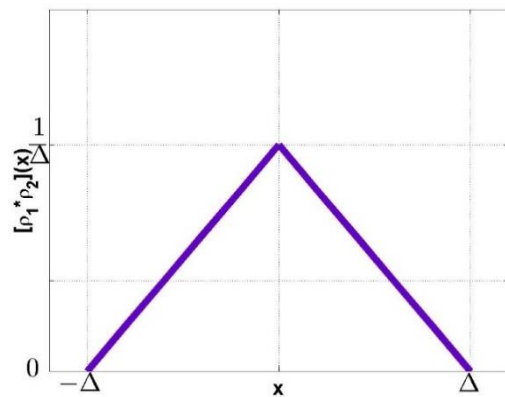


Figura 1-4 (a)FDP del ruido uniforme 1 (b) FDP del ruido uniforme 2 (c)FDP del ruido triangular

Para señales de dither con funciones de densidad como la descrita en la Ecuación (1.6) la señal de error puede causar la dependencia del primer y segundo momento con la señal de entrada [2]. En [3] Wannamaker, Lipshitz y Vanderkooy recomiendan un dither de FDP triangular. El FDP triangular se genera a través de la convolución de dos señales con FDP uniformes. En [4], Lipshitz y Vanderkooy probaron que no era posible hacer un dither independiente de la señal de entrada del sistema, que generara un error total de cuantización con una varianza constante e inferior al que

CAPÍTULO 1. INTRODUCCIÓN

se obtuvo con FDP triangular. Adicionalmente con una FDP triangular no se encuentran niveles de ruido que pudiesen ser detectados por el oído humano.

En el trabajo de Benítez [1], se propuso un método que encuentra un dither adaptativo con una varianza inferior al de dither triangular tradicional, pero con alta demanda de tiempo de procesamiento. Por lo tanto, a través del desarrollo del presente trabajo se investigará un algoritmo con un tiempo de procesamiento menor para desarrollar dithering adaptativo con el que se mantenga o incremente la relación señal/ruido que se encuentra en la investigación de punta.

La tesis está organizada de la siguiente forma: en el Capítulo 2 se incluye una revisión de literatura de cuantización y dithering. En el Capítulo 3, se trata la metodología abarcando el contenido teórico que permite el desarrollo de la tesis, se tratan temas como la cuantización, los estadísticos, dithering, el algoritmo de pendiente máximo y se introduce la prueba de Ljung-Box. Finalmente, en los Capítulos 4 y 5 se explica los experimentos llevados a cabo y sus resultados y se discuten las conclusiones con base en los resultados obtenidos de diferentes algoritmos y los trabajos futuros.

Capítulo 2.

REVISIÓN DE LITERATURA

Se realizó una revisión de la investigación de punta resumiendo los avances más importantes y recientes en dithering. En el capítulo se tratan aspectos históricos relevantes en cuantización y dithering, así como ejemplos de aplicaciones que usan dithering.

2.1. CUANTIZACIÓN

Se han realizados muchas investigaciones en cuantización y dithering. Existe la cuantización uniforme y no uniforme en la conversión analógico/digital. Se hace cuantización uniforme cuando la diferencia entre los niveles de cuantización permanece constante. Se estará trabajando con la cuantización uniforme en esta investigación.

El estudio de las propiedades de la cuantización y sus efectos ha incrementado después de la mitad del siglo XX. En 1948, en [5], Oliver, Pierce y. Shannon

CAPÍTULO 2. REVISIÓN DE LITERATURA

reconocieron el papel que tenía la cuantización en modulación y en la conversión analógica-digital de señales mientras se desarrollaban sistemas de modulación por impulsos codificados (MIC). MIC fue la primera técnica digital para transmitir una señal analógica, típicamente la voz, a través de un canal analógico.

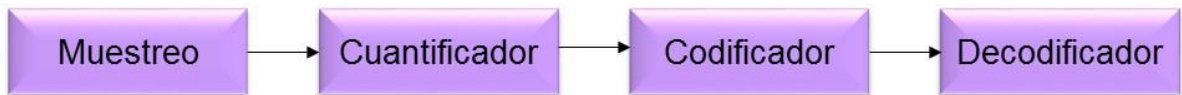


Figura 2-1

Los elementos de un sistema MIC son el muestreo, la cuantización, la codificación y la decodificación (Ver Figura 2-1). En la etapa de muestreo una señal continua $x(t)$ se convierte en una secuencia de muestras:

$$x[n] = x(n/f_s), \quad (2.1)$$

donde f_s es la frecuencia de muestreo. Puesto que el objetivo principal de un sistema de transmisión es representar en la salida como una función de tiempo la señal que se tiene en la entrada del sistema, con la cuantización se representa la señal a través de ciertos niveles permitidos para luego proceder a su codificación.

En 1948, Bennet [6] publicó un análisis exacto de ruido de cuantización para procesos aleatorios normales y demostró que bajo ciertas asunciones de densidades “suaves”, el error de cuantización se comporta como un ruido aleatorio, es decir, tiene una correlación poco significativa con la señal de entrada y tiene un espectro aproximadamente plano, que lo haría un ruido blanco. Esto le permitió desarrollar un modelo aditivo del error de cuantización, ya que de esta manera se representaría la salida del cuantizador como la suma de la señal y ruido blanco.

CAPÍTULO 2. REVISIÓN DE LITERATURA

Widrow en [7] [8] desarrolló muchas de las bases matemáticas de la cuantización, analizando los sistemas de cuantización sin dithering demostró la utilidad de las funciones características y explicó cómo éstas se pueden recuperar a través de la distribución de la salida del cuantizador y a través de la relación entre la entrada y la salida, resultando ser similar al teorema de muestreo de Shannon [9]. Widrow explicó cómo bajo ciertas condiciones el error de cuantización se puede modelar como una secuencia independiente e idénticamente distribuida (i.i.d o independent and identically distributed, por sus siglas en inglés), definiendo momentos estadísticos de orden superior para el error y el estudio de los momentos estadísticos en la salida del cuantizador.

2.2. DITHERING

El término “dither” se publicó en libros de computación analógica y armas controladas hidráulicamente poco después de la Segunda Guerra Mundial [10]. Una de las primeras aplicaciones de dither vino durante de la Segunda Guerra Mundial [11]. Para realizar cálculos de navegación y de la trayectoria de bombas, los bombarderos aéreos utilizaban equipos mecánicos, los cuales funcionaban con mayor exactitud cuándo se encontraban a bordo del avión en vuelo que cuando se encontraban en tierra. Se encontró que este fenómeno se debía a que la vibración del avión reducía el error de sus partes móviles, ya que no se movían en pasos cortos sino continuamente. Se construyeron entonces motores con pequeñas vibraciones y a estas vibraciones se les denominó dither [11].

CAPÍTULO 2. REVISIÓN DE LITERATURA

Robert en [12] [13] introdujo dithering en la cuantización, como un medio para volver aleatorios los efectos de la cuantización uniforme. Esto consistió en agregar una señal de dither para minimizar la distorsión o ruido introducido por la cuantización, haciendo que las imágenes se vieran mejor al disminuir los efectos de este ruido. Robert experimentó con dithering sustractivo en aplicaciones de video [12]. En dither sustractivo la señal de dither se resta después de la cuantización (Sección 3.2). Robert trabajó en la transmisión de imágenes de televisión sobre un canal digital, al agregar una señal de dither para sustraerla posteriormente en el receptor, pudo reducir la tasa de fuente a tres bits por muestra. Para una buena calidad de imagen, la MIC necesita que la longitud de cada muestra sea al menos de seis bits. En esta investigación, Robert agregó cantidades pequeñas de ruido antes de la cuantización para causar el mismo efecto en el ojo con menos bits. El ojo humano puede distinguir cambios pequeños en la intensidad. La Figura 2-2 muestra algunos ejemplos de [13] y se puede apreciar cómo a pesar del ruido que se agrega en 2-2(b) y 2-2(d), las imágenes se ven mejoradas con respecto a la 2-2(c) y 2-2(e).

El dithering sustractivo fue posteriormente adaptado para la codificación de voz [14]. La propiedad de dither de volver el error de cuantización blanco mejora notablemente la calidad de la voz muestreada a 6 KHz.

Luego de esto, Schuchman en [15], exploró las estadísticas unidimensionales del dither sustractivo y obtuvieron las condiciones de la señal de dither para que sea considerada independiente de la señal de entrada. Después, Sherwood en [16]

CAPÍTULO 2. REVISIÓN DE LITERATURA

presentó un tratamiento completo de los sistemas de dither sustractivo y probó teoremas relacionados con el diseño de la secuencia de dither.

En 1993, Gray y Stockham en [17] presentaron una derivación unificada del cuantizador con dither, dither sustractivo y dither no sustractivo, explicando además sus propiedades estadísticas. Sherwood [18] desarrolló una investigación detallada en dither no sustractivo, mostrando que al agregar dither con una función de densidad de probabilidad apropiada, los momentos del error total pueden ser independientes de la señal de entrada.



(a) Original de seis bits

CAPÍTULO 2. REVISIÓN DE LITERATURA

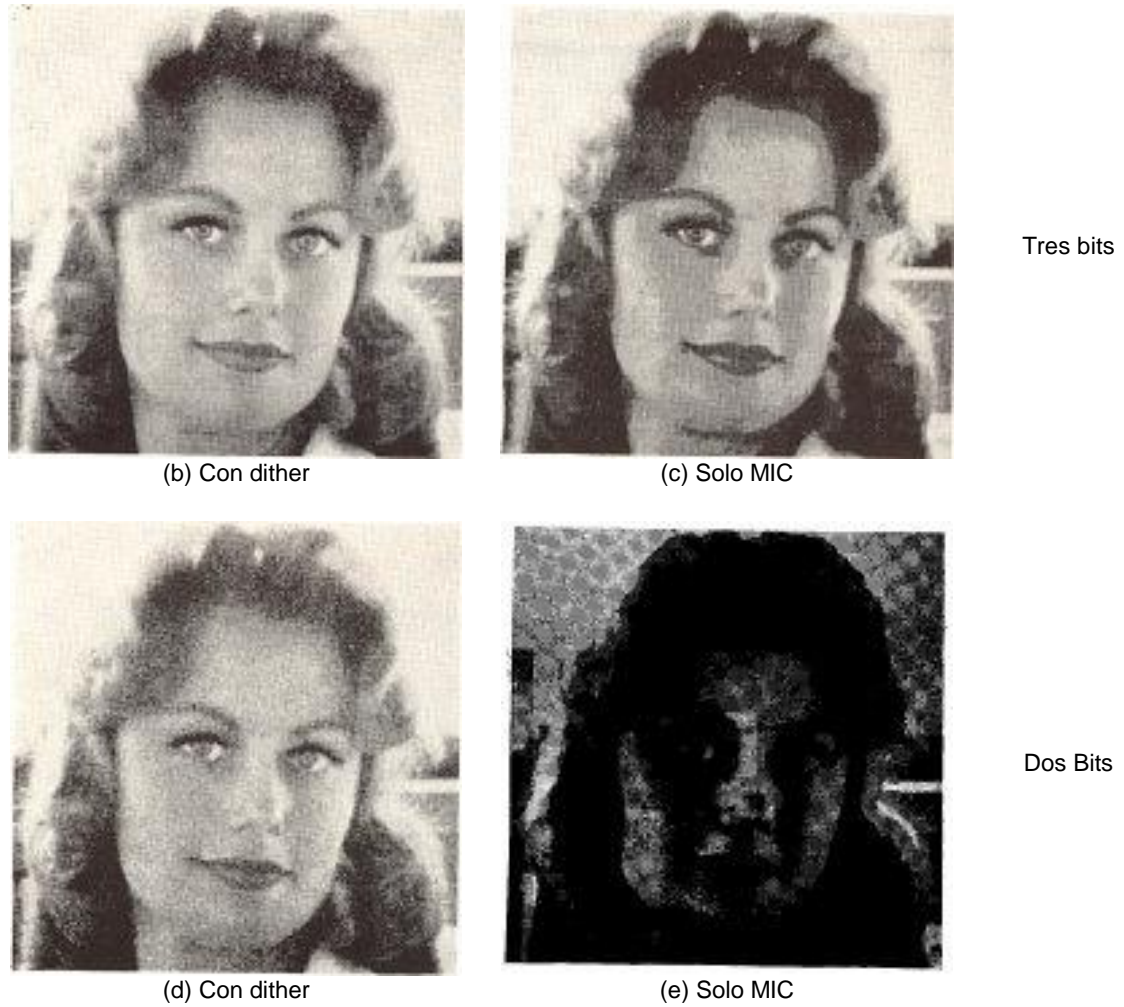


Figura 2-2

A mediados de los años 80, el procesamiento digital se extendió cada vez más y se hizo necesario incluir dither en los sistemas digitales [2]. En [4] Lipshitz y Vanderkooy, utilizan dither con FDP triangular e introducen el dither digital junto con la cuantización digital. El uso del dither con FDP triangular se usa en [19] para aplicaciones digitales de audio y en general es la técnica más usada para señales de una dimensión [1]

CAPÍTULO 2. REVISIÓN DE LITERATURA

2.3. RESUMEN

En este capítulo se hizo una revisión del estado del arte de cuantización y dithering incluyendo aplicaciones. Se comentó la función de la cuantización dentro del proceso de conversión analógico/digital y la distorsión que produce. A lo largo del capítulo se muestran distintas aplicaciones donde se usa dither para disminuir los efectos del ruido que introducen el proceso de cuantización..

Capítulo 3

METODOLOGIA

En este capítulo se explica la teoría concerniente al proceso de cuantización, dithering, el algoritmo de pendiente máxima y la prueba de Ljung-Box. La primera y segunda sección abarcan las características generales de la re-cuantización y el ruido de cuantización, explicando los estadísticos que son útiles para desarrollar esta investigación. En la tercera sección se deriva la FDP del error total utilizando el teorema de convolución y el *area sampling*. En la cuarta sección se explica el algoritmo de optimización de pendiente máxima que permite encontrar una señal de dither adecuada. Finalmente, en la quinta sección se presentan pruebas de hipótesis para evaluar si el error total es un ruido blanco.

3.1. CUANTIZACIÓN

El proceso de llevar una señal analógica a la representación digital más cercana en la conversión A/D se denomina cuantización [20] (Ver Figura 3-1). La resolución de la señal digital está determinada por el esquema de cuantización utilizado.

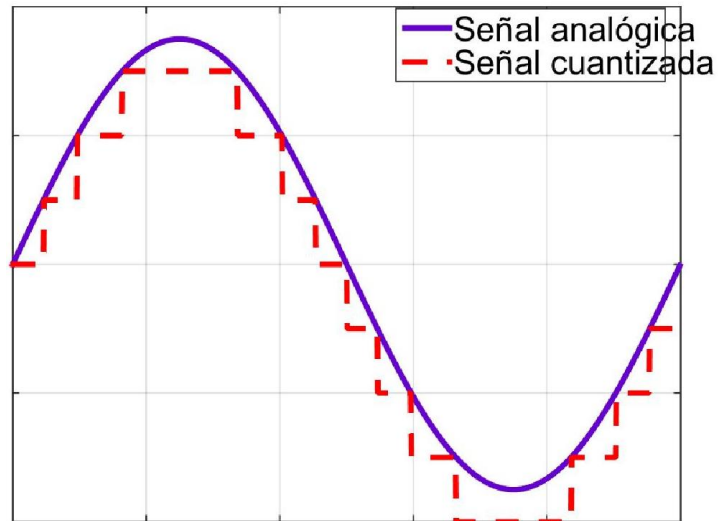


Figura 3-1 Señal analógica y señal cuantizada.
La señal cuantizada tiene tres bits de resolución (8 niveles)

El proceso de cuantización también consiste en reducir la resolución de la señal digital. La operación de cuantización, descrita a través del diagrama de bloques en la Figura 3-2, hace que el sistema no sea lineal. La relación existente entre la señal analógica (x) y la señal cuantizada ($Q(x)$) está representada por una función escalonada. La Figura 3-2 muestra el diagrama de bloques del cuantizador y la Figura 3-3 muestra las funciones de transferencia para distintos esquemas de cuantización.

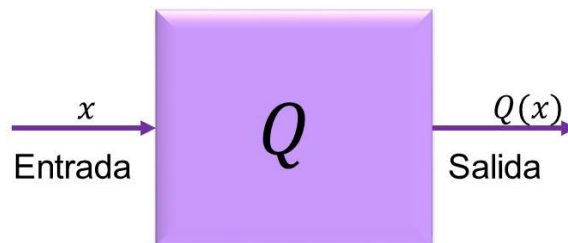


Figura 3-2 Diagrama de bloques del cuantizador.
La entrada del sistema es (x) y la señal cuantizada $Q(x)$ es la salida del sistema

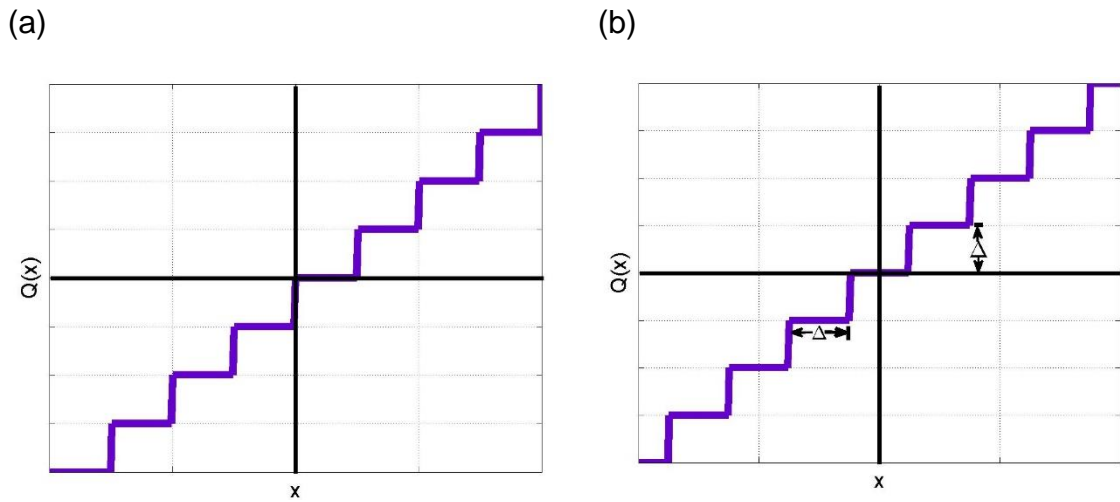


Figura 3-3 Funciones de transferencia de cuantización (a) Función piso (b) Redondeo

La mayoría de las representaciones discretas en amplitud usan el sistema de codificación binaria. El número de niveles N corresponde a la potencia de base dos y a la potencia N se denomina resolución y define el número de niveles en los cuales se puede cuantizar una señal [20]. Una representación en el sistema binario sería el que se muestra en la Ecuación (3.1). El coeficiente de 2^{N-1} se denomina el bit más significativo (MSB ó *Most Significant Bit*, por sus siglas en inglés) y a el coeficiente de 2^0 , bit menos significativo (LSB ó *Least Significant Bit*, por sus siglas en inglés)

$$A_s = \sum_{i=0}^{N-1} a_i 2^i = a_0 2^0 + a_1 2^1 + \dots + a_{n-1} + \dots + a_{N-1} 2^{N-1}. \quad (3.1)$$

Al tamaño del escalón (Δ), en la Ecuación (3.1) a_0 , comúnmente se le llama LSB, debido a que un cambio de un escalón en la señal de entrada corresponde al cambio en el LSB de la señal de salida en el sistema binario, es decir, el LSB representa la

CAPÍTULO 3. METODOLOGIA

diferencia entre la codificación digital binaria para dos niveles adyacentes en la salida.

En este trabajo solo se considerarán cuantizadores que sean uniformes, es decir, que los niveles o escalones de la función escalonada estén equidistantes. La Figura 3-3 muestra las características de un cuantizador uniforme.

3.1.1. Función de densidad de potencia de la señal cuantizada

Las funciones de densidad de potencia (FDP) de las señales de entrada (x) y de salida ($Q(x)$) del sistema de cuantización del diagrama de bloques de la Figura 3-2, están relacionadas a través de un tipo especial de muestreo [21]. La FDP de salida es un tren de funciones delta de Dirac, cuyas áreas corresponden a las áreas bajo la FDP de entrada dentro de los límites de cada escalón. Se realiza un *area sampling*, donde la FDP de entrada ρ_x se corta en bandas, entonces el área de cada banda se comprime en un impulso en el centro de la banda para formar la FDP de salida ρ_{x_q} de la Figura 3-4 FDP de la señal cuantizada. La FDP de la señal cuantizada ρ_{x_q} es el resultado de multiplicar la FDP de la señal de entrada ρ_x de la **¡Error! No se encuentra el origen de la referencia.**(a) con el tren de pulsos de la **¡Error! No se encuentra el origen de la referencia.**(b).

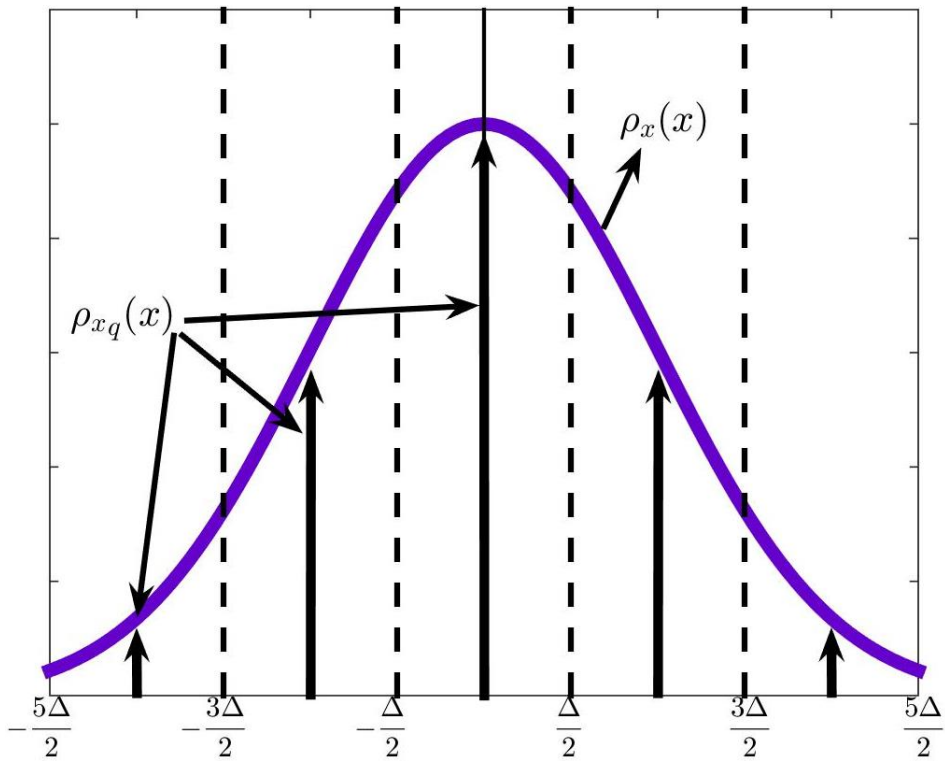


Figura 3-4 FDP de la señal cuantizada.
 La figura muestra la FDP de la señal de entrada ρ_x con línea violeta y el tren de pulsos de color negro es la FDP de la señal cuantizada ρ_{x_q}

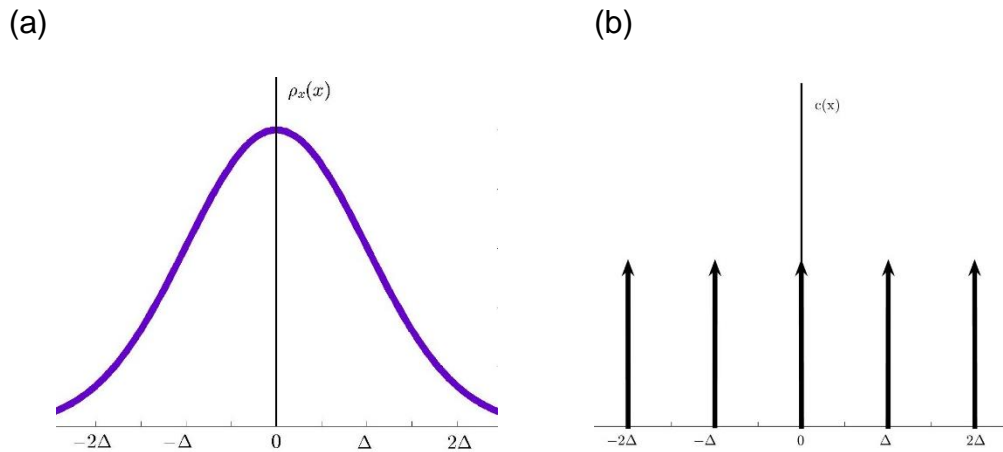


Figura 3-5 Formación de FDP de la señal cuantizada (a) Ventana de entrada (b) Tren de funciones delta de Dirac.
 Cuando se multiplican (a) y (b) se tiene como resultado la FDP de la señal cuantizada ρ_{x_q}

3.2. RUIDO DE CUANTIZACIÓN

El proceso de cuantización cambia la señal de entrada, generando un error de cuantización (ε), que se define como la diferencia entre la entrada del cuantizador $x[n]$ y la salida $x_q[n]$ (Ver Figura 3-6).

$$\varepsilon[n] \triangleq Q(x[n]) - x[n], \quad (3.2)$$

donde $Q(\cdot)$ es la operación de cuantización y consiste en llevar una señal a la representación digital más cercana. $x_q[n] = Q(x[n])$ es la señal cuantizada y x es la señal original. Se estará trabajando con la función de piso, en la cual se redondea al entero más cercano de menor valor y se observa que:

$$|\varepsilon[n]| < \Delta. \quad (3.3)$$

Los valores del ruido de cuantización están en el rango de 0 a $+\Delta$. La Figura 3-7 muestra el error de cuantización para la señal de la Figura 3-1 con el esquema de cuantización de la Figura 3-3(b)

La potencia que se asocia con este error de cuantización define la calidad del proceso realizado en la conversión A/D.

Esta investigación estará trabajando con señales discretas representadas a través del complemento a dos del sistema binario, por lo tanto, la cuantización con función de piso estaría representada por la Ecuación (3.4).

$$x_q[n] = 2^M \left\lfloor x[n] * \frac{2^R}{2^M} \right\rfloor \quad (3.4)$$

Donde la señal es cuantizada de R a $R - M$ bits y se usa el símbolo $\lfloor \cdot \rfloor$ para simbolizar la función piso. En el caso de $R = 2$ y $M = 1$, cuando se disminuye la resolución de la señal por un bit y la señal cuantizada tendrá la mitad de la resolución.

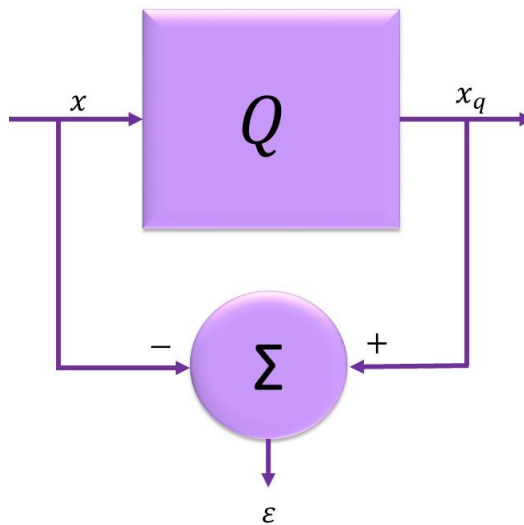


Figura 3-6 Diagrama de bloques: error de cuantización

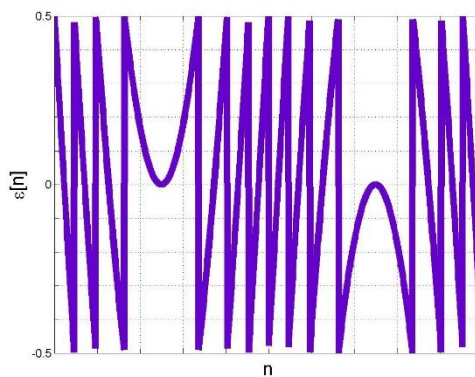


Figura 3-7 Error de cuantización.
La diferencia entre $x_q[n]$ y $[n]$ da como resultado esta señal de error.

3.2.1. Estadísticas del ruido de cuantización

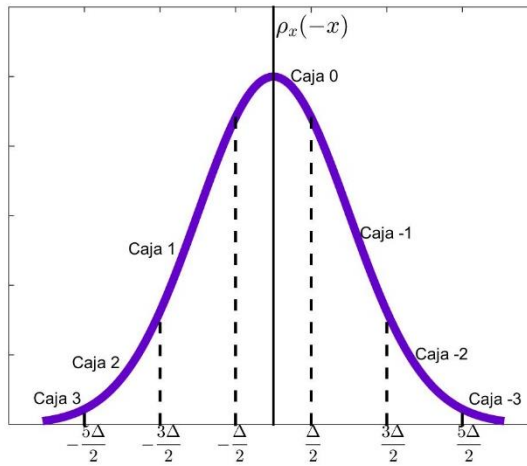
La FDP del ruido de cuantización $\rho_\varepsilon(\varepsilon)$ se puede obtener cortando la FDP de x , ρ_x , en bandas (Ver Figura 3-4) para luego sumarlas [21], la Figura 3-8 (a) tiene la división de ρ_x en siete cajas. Los valores del error de cuantización ε son el resultado de la cuantización de x dentro del escalón o la caja de cuantización correspondiente de la Figura 3-8 (b). La probabilidad de obtener un valor determinado para ε es la suma de todas las probabilidades de las cajas de cuantización de la Figura 3-8(c).

En [22] se enuncia que el ruido de cuantización $\varepsilon[n]$ se puede modelar como un proceso aleatorio de carácter aditivo que no está correlacionado con la entrada, es independiente y esta idénticamente distribuido. La función de densidad de probabilidad

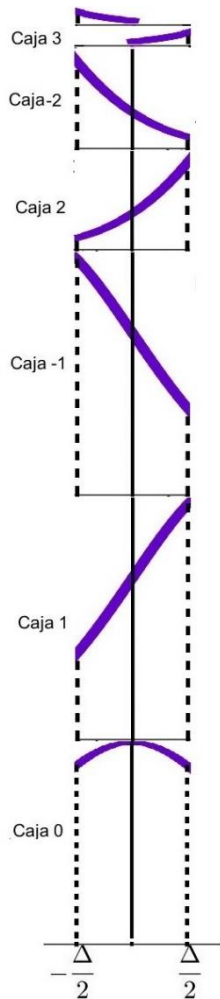
$$\rho_\varepsilon(\varepsilon) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} < \varepsilon < \frac{\Delta}{2} \\ 0, & \varepsilon < -\frac{\Delta}{2} \text{ o } \varepsilon > \frac{\Delta}{2} \end{cases} \quad (3.5)$$

CAPÍTULO 3. METODOLOGIA

(a)



(b)



(c)

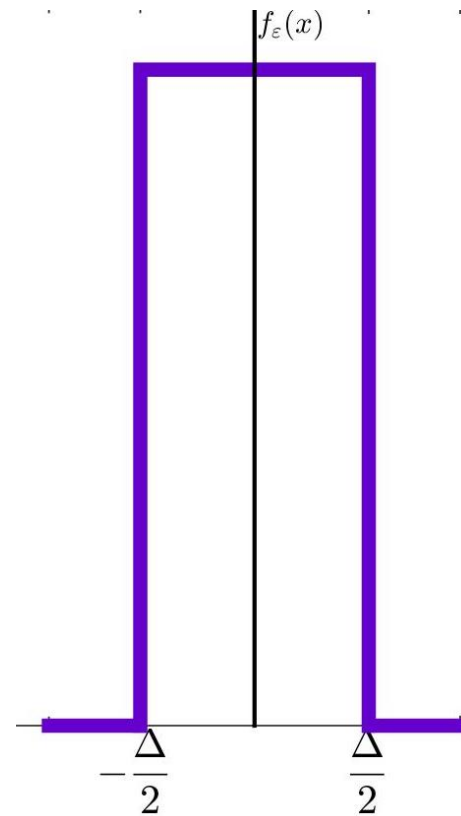


Figura 3-8 Construcción de la FDP del ruido de cuantización.

En la (a) se encuentra la FDP de x , $\rho_x(-x)$ dividida en siete cajas, en la (b) se encuentran los valores de la cuantización dentro de su respectiva caja y la (c) tiene el resultado su suma.

CAPÍTULO 3. METODOLOGIA

La Ecuación (3.5) describe un error con una función de probabilidad uniforme y bajo estas condiciones se tiene que la media o el primer momento es el que describe la Ecuación (3.6), el segundo momento o la varianza es el que describe la Ecuación (3.7) y representa la potencia del ruido, ésta se distribuye uniformemente entre cero y la frecuencia de muestreo. La función que generaliza el momento k es la que describe la Ecuación (3.8)

$$E[\varepsilon] = 0, \quad (3.6)$$

$$E[\varepsilon^2] = \frac{\Delta^2}{12}, \quad (3.7)$$

$$E[\varepsilon^k] = \begin{cases} \frac{1}{k+1} \left(\frac{\Delta^2}{12}\right)^k, & \text{para } k \text{ par.} \\ 0, & \text{para } k \text{ impar} \end{cases} \quad (3.8)$$

Cuando se cumplen las condiciones anteriores, el ruido de cuantización es blanco. Este caso se presenta cuando el cambio de amplitud es grande de muestra a muestra, pero para señales muy pequeñas, el error de cuantización es una distorsión que depende de la entrada y no puede modelarse como una fuente independiente de ruido. Para mitigar las propiedades estadísticas del ruido de cuantización se le suma una señal de ruido, denominada dither, a la señal de entrada antes de ser cuantizada, la Sección 3.3 describe las características de esta señal.

CAPÍTULO 3. METODOLOGIA

3.2.1.1. Autocorrelación

La autocorrelación $R_{\varepsilon\varepsilon}[n_1, n_2]$ de $\varepsilon[n]$ es el valor esperado del producto $\varepsilon[n_1]\varepsilon[n_2]$ y representa la relación lineal entre dos puntos de la señal del ruido de cuantización. n_1 y n_2 son las variables de tiempo y representan dos muestras cualesquiera del vector de autocorrelación $R_{\varepsilon\varepsilon}$.

$$R_{\varepsilon\varepsilon}[n_1, n_2] = E[\varepsilon[n_1]\varepsilon[n_2]]. \quad (3.9)$$

El valor de $R_{\varepsilon\varepsilon}[n_1, n_2]$ en la diagonal $n_1 = n_2 = n$ es la varianza de la señal y representa la potencia promedio de la misma.

Si $\varepsilon[n]$ es un proceso estacionario en sentido amplio (WSS o *wide sense stationary*, por sus siglas en inglés), el valor esperado del proceso es una constante y por lo tanto independiente de la variable de tiempo y su autocorrelación se puede escribir como una función de la diferencia de muestras $\tau = n_1 - n_2$, entonces $R_{\varepsilon\varepsilon}[n_1, n_2] = R_{\varepsilon\varepsilon}[n_1 - n_2] = R_{\varepsilon\varepsilon}[\tau]$ y su autocorrelación se define como muestra la Ecuación (3.11)

$$R_{\varepsilon\varepsilon}[\tau] = E[\varepsilon[n + \tau]\varepsilon[n]] = \begin{cases} \sum_{n=0}^{J-\tau-1} \varepsilon[n + \tau]\varepsilon[n], \tau \geq 0 \\ R_{\varepsilon\varepsilon}[-\tau], \tau < 0, \end{cases} \quad (3.10)$$

donde τ es la diferencia en tiempo de dos variables aleatorias, K representa el número de elementos del vector ε y el vector $R_{\varepsilon\varepsilon}[\tau]$ tiene elementos dados por: $n = 1, 2, \dots, 2K - 1$.

CAPÍTULO 3. METODOLOGIA

Para el ruido blanco, cuando τ es diferente de cero, la autocorrelación sería cero y cuando τ es cero, entonces la autocorrelación, como se había mencionado, es la varianza o la potencia de la señal.

$$R_{\varepsilon\varepsilon}[0] = E[|\varepsilon[n]|^2]. \quad (3.11)$$

Resumiendo, la autocorrelación de un ruido blanco se expresa matemáticamente así:

$$R_{\varepsilon\varepsilon}[\tau] = \begin{cases} \sigma^2, & \text{para } \tau = 0 \\ 0, & \text{para } \tau \neq 0. \end{cases} \quad (3.1)$$

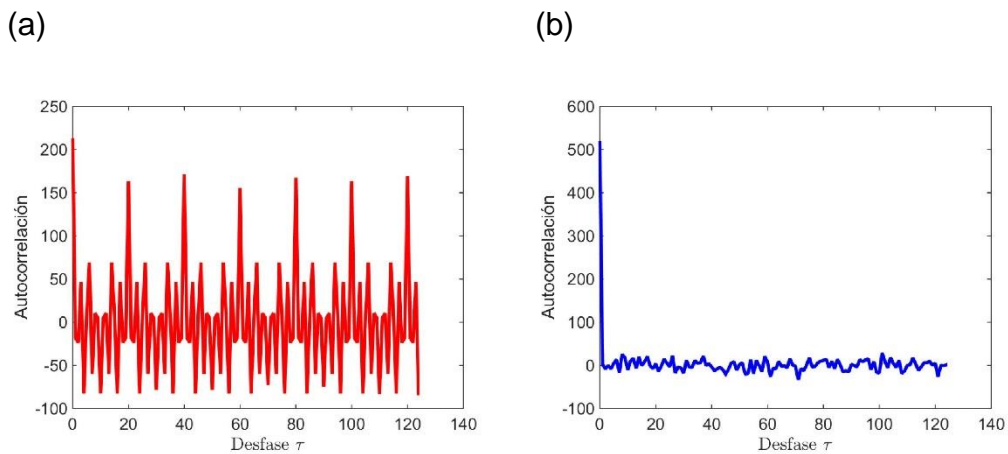


Figura 3-9 Autocorrelación del error de cuantización (a) sin dither (b) con dither triangular.

Cuando no se agrega dither (a) se presentan valores notables en desfases distintos de 0 y cuando se agrega un dither triangular antes de se cuantizada (b) los valores de ka autocorrelación en desfases distintos de 0 son cercanos a 0.

3.2.1.2. Coeficiente de autocorrelación

La media y la varianza del error de cuantización están dada por:

$$\mu_{\varepsilon} = \frac{1}{J} \sum_{n=0}^{J-1} \varepsilon[n], \quad (3.13)$$

$$\sigma_{\varepsilon}^2 = E\{\varepsilon^2[n]\} - E^2\{\varepsilon[n]\}. \quad (3.14)$$

CAPÍTULO 3. METODOLOGIA

Si $\varepsilon[n]$ tiene media cero, el coeficiente de autocorrelación es:

$$r_{\varepsilon}[n] = \frac{R_{\varepsilon\varepsilon}}{\sigma_{\varepsilon}^2}. \quad (3.15)$$

El coeficiente de autocorrelación mide la dependencia lineal existente entre $\varepsilon[n]$ y $\varepsilon[n + \tau]$, si son independientes entonces la correlación es cero para todo $\tau \neq 0$. y su coeficiente de autocorrelación también es cero.

3.2.1.3. Densidad Espectral de Potencia

La densidad espectral de potencia (DEP) de un proceso aleatorio WSS es, por definición, la transformada de Fourier en tiempo discreto (TFTD) de su función de autocorrelación y mide la potencia de la señal en el dominio de la frecuencia:

$$F_{DT}[\varepsilon](w) \triangleq \sum_{n=-\infty}^{\infty} R_{\varepsilon\varepsilon}[n] e^{-jwn}, \quad (3.16)$$

Donde $F_{DT}[\cdot]$ es la transformada de Fourier en tiempo discreto, w es la frecuencia angular. $S_{\varepsilon\varepsilon}(f)$ es la DEP, entonces, se tiene que:

$$R_{\varepsilon\varepsilon}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{\varepsilon\varepsilon}(w) e^{jwn} dw. \quad (3.17)$$

En el caso donde el error de cuantización es un ruido blanco con media cero, entonces la sumatoria de la ecuación 3.16 solo tendría un valor que equivale a la varianza cuando $n = 0$, de tal forma que $S_{\varepsilon\varepsilon}(w) = \sigma^2$. Esta es la razón por la cual el ruido blanco tiene componentes en todas las frecuencias del espectro con la misma magnitud.

3.3. DITHER

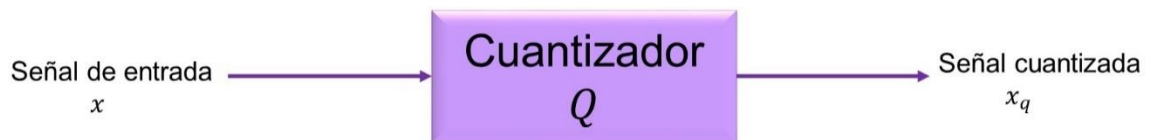
Una señal de dither es una señal de ruido que se le suma a la señal de entrada x para evitar armónicos indeseados. Dentro de los esquemas de cuantización se encuentra: sin dither (UN o *undithered*, por sus siglas en inglés), con dither sustractivo (SD o *Subtractively Dithered*, por sus siglas en inglés) y con dither no sustractivo (NSD o *Non-Subtractively Dithered*). En esta investigación se trabaja con este último. La Figura 3-10 muestra los esquemas de cuantización.

3.3.1 Error total

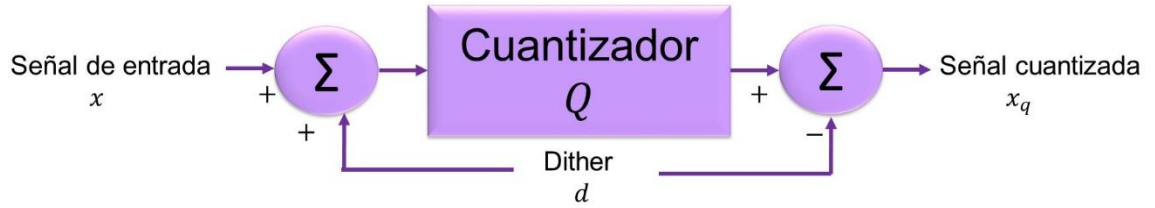
El error total ϑ está definido como la diferencia entre la entrada y la salida del sistema de cuantización, donde se tiene en cuenta la señal de dither d sumada a la señal original antes de ser cuantizada. El error total es diferente del error de cuantización, el cual es la diferencia entre la entrada y salida del cuantizador. Para el sistema sin dither, la señal cuantizada $x_q = Q(x)$, por tanto, el error total sería equivalente al error de cuantización:

$$\vartheta = Q(x) - x \quad (3.18)$$

(a)



(b)



(c)

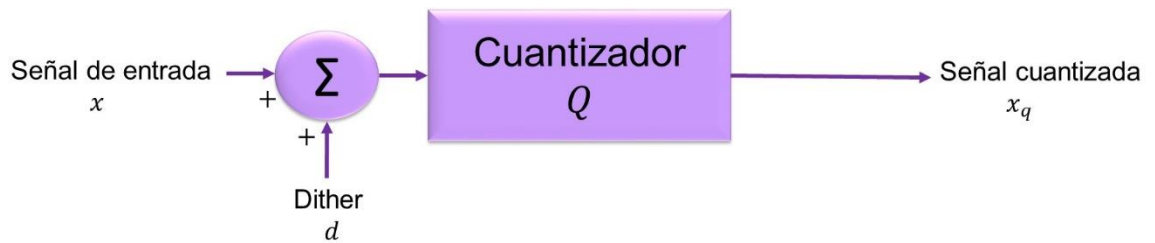


Figura 3-10 Esquemas de cuantización (a) Sin dither (b) Con dither sustractivo (c) Con dither no sustractivo

Para el esquema con dither sustractivo, la señal de entrada al cuantizador será el resultado de la adición de la señal de entrada x y el dither d y la señal cuantizada $x_q = Q(x + d)$, por tanto, el error total es el que se muestra:

$$\vartheta = Q(x + d) - (x + d). \quad (3.19)$$

El error total para el esquema de dither no sustractivo, se muestra en la Ecuación (3.21) y se puede apreciar que se omite la operación de sustracción de la señal de dither en la salida del cuantizador.

$$\vartheta = Q(x + d) - x. \quad (3.20)$$

3.3.2. FDP del error total

Tal como se mostró la Sección 3.2.1, la FDP del ruido de cuantización es la FDP de la señal de entrada cortada en intervalos de longitud Δ , donde la FDP de la señal de entrada se multiplica con el tren de pulsos (Ver Figura 3-5 (b)). La FDP del error de cuantización esta en función de delta de Dirac $\delta(\cdot)$ y se representa como se muestra en la Ecuación (3.21).

$$\rho_q(q) = \sum_{k=-\infty}^{\infty} \delta(q - k\Delta) \int_{-\frac{\Delta}{2}+k\Delta}^{\frac{\Delta}{2}+k\Delta} \rho_x(x) dx, \quad (3.21)$$

donde q es la señal de salida y $\rho_x(x)$ es la FDP de la señal de entrada. Se puede, entonces, deducir la FDP del error total usando esta ecuación. Sea $\rho_{\vartheta|x}(y, x)$ la FDP condicional del error total en un sistema de dither no sustractivo. Donde $y = x + d$ es la entrada del cuantizador, la FDP condicional de la entrada del cuantizador dado x (señal de entrada) es la que se describe en la Ecuación (3.22) y usando *area sampling*, la FDP del error condicional es el que muestra la Ecuación (3.23). Multiplicando por $\rho_x(x)$ e integrando con respecto a x , la FDP marginal de ϑ es el que muestra la Ecuación (3.24). Mostrando la dependencia existente entre el error total y la señal de entrada en un sistema de dither no sustractivo [23].

$$\rho_{y|x}(y, x) = \rho_d(y - x), \quad (3.22)$$

$$\rho_{\vartheta|x}(\vartheta, x) = \sum_{k=-\infty}^{\infty} \delta(\vartheta + x - k\Delta) \int_{-\frac{\Delta}{2}+k\Delta}^{\frac{\Delta}{2}+k\Delta} \rho_d(y - x) dy, \quad (3.23)$$

$$\rho_{\vartheta}(\vartheta) = \int_{-\infty}^{\infty} \rho_{\vartheta|x}(\vartheta, x) \rho_x(x) dx. \quad (3.24)$$

3.3.3. Dithering Adaptativo

La señal de dither debe tener un error total con varianza constante, esta varianza deber ser igual o menor que la que se registra con dither triangular, de tal manera que tenga menor ruido. Se usarán métodos numéricos de optimización para determinar la señal de dither.

La señal cuantizada, x_q está definida como enuncia la Ecuación (3.25):

$$x_q[n] = Q\left(\frac{x[n]}{2^m} + d[n]\right)2^m. \quad (3.25)$$

El objetivo del trabajo consiste en encontrar una señal de dither, tal que, el error total tenga un espectro blanco y una varianza constante.

El error total, descrito en la Ecuación (3.20), es la diferencia entre la señal cuantizada dividida por 2^m y la señal de entrada (con resolución disminuida):

$$\vartheta[n] = \frac{x[n] - x_q[n]}{2^m} \quad (3.26)$$

Como se describió en la Sección 3.2.1.1 con la autocorrelación del error total se verifica si éste es un ruido blanco, puesto a pesar que el error total depende de la señal que se cuantiza, tal como enuncia la Ecuación (3.13), el ruido blanco será un impulso, equivalente a la magnitud de su varianza cuando $\tau = 0$.

El vector error $\vec{\vartheta}[n]$ de tamaño S es:

$$\vec{\vartheta}[n] = \begin{bmatrix} \vartheta[n] \\ \vartheta[n-1] \\ \vdots \\ \vartheta[n-S-1] \end{bmatrix}. \quad (3.27)$$

La autocorrelación, que viene dada por la Ecuación (3.11) se puede escribir en forma vectorial:

$$R_{\vartheta\vartheta}[\tau] = \begin{bmatrix} R_{\vartheta\vartheta}[0] \\ R_{\vartheta\vartheta}[1] \\ \vdots \\ R_{\vartheta\vartheta}[S-1] \end{bmatrix}. \quad (3.28)$$

$$R_{\vartheta\vartheta}[\tau] = \begin{bmatrix} \vartheta[0]\vartheta[0] & +\vartheta[1]\vartheta[1] & +\dots & +\vartheta[S-1]\vartheta[S-1] \\ 0 & +\vartheta[0]\vartheta[1] & +\dots & +\vartheta[S-2]\vartheta[S-1] \\ \vdots & +\ddots & +\dots & \vdots \\ 0 & +\dots & 0 & +\vartheta[0]\vartheta[S-1] \end{bmatrix}. \quad (3.29)$$

De la Ecuación (3.29) se puede apreciar que a medida que aumenta τ , el número de elementos de la suma para hallar cada elemento del vector $R_{\vartheta\vartheta}$ disminuye, es así que para hallar, por ejemplo, $R_{\vartheta\vartheta}[S-1]$ solo se tiene en cuenta $\vartheta[0]\vartheta[S-1]$. En esta investigación se estará usando la autocorrelación circular, la cual es simétrica cuando $\tau = \frac{S}{2}$ y tiene la misma cantidad de elementos para hallar cada uno de los valores de $R_{\vartheta\vartheta}$. El nuevo sistema de ecuaciones está dado por:

CAPÍTULO 3. METODOLOGIA

$$R_{\vartheta\vartheta}[\tau] = \begin{bmatrix} \vartheta[0]\vartheta[0] & +\vartheta[1]\vartheta[1] & + \dots & +\vartheta[S-1]\vartheta[S-1] \\ \vartheta[0]\vartheta[S-1] & +\vartheta[0]\vartheta[1] & + \dots & +\vartheta[S-2]\vartheta[S-1] \\ \vartheta[0]\vartheta[S-2] & +\vartheta[0]\vartheta[S-1] & + \dots & +\vartheta[S-3]\vartheta[S-1] \\ \vdots & \ddots & & \vdots \\ \vartheta[0]\vartheta\left[\frac{S}{2}\right] & \dots & & +\vartheta\left[\frac{S}{2}\right]\vartheta[S-1] \end{bmatrix} \quad (3.30)$$

Para que sea ruido blanco la Ecuación (3.28) sería equivalente a:

$$R_{\vartheta\vartheta}[\tau] = \begin{bmatrix} R_{\vartheta\vartheta}[0] \\ R_{\vartheta\vartheta}[1] \\ \vdots \\ R_{\vartheta\vartheta}[S-1] \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \quad (3.31)$$

Se puede apreciar que el sistema de ecuaciones depende de ϑ , si se reemplaza (3.25) en (3.26):

$$\vartheta[n] = \frac{x[n]}{2^m} - Q_r \left(\frac{x[n]}{2^m} + d[n] \right), \quad (3.32)$$

$x[n]$ tiene una parte entera $x_e[n]$ y una parte decimal $x_d[n]$, de manera que:

$$x[n] = x_e[n] + x_d[n]. \quad (3.33)$$

Si la señal de entrada $x[n]$ tiene parte entera, ésta no es afectada por el proceso de cuantificación, así que la Ecuación (3.32) se puede reescribir así:

$$\vartheta[n] = \frac{x_e[n]}{2^m} - Q_r \left(\frac{x_d[n]}{2^m} + d[n] \right). \quad (3.34)$$

CAPÍTULO 3. METODOLOGIA

De la Ecuación (3.34) se observa que el error total depende solamente de la parte decimal $x_d[n]$, que es una constante conocida, y depende de la señal de dither $d[n]$. Además, depende de la operación de redondeo, que es una operación no lineal.

Este sistema de ecuaciones de la Ecuación (3.30) no es lineal debido a que la operación de función de piso no es diferenciable y se resuelve utilizando el algoritmo de optimización de pendiente máxima, tal como se describe en la Sección 3.3.3. Para linealizar el sistema en el punto de operación, normalmente se usan derivadas o aproximaciones finitas. El estimador más simple de $Q_r(x)$ es $\hat{Q}_r(x) = x$, dado $\frac{d\hat{Q}_r(x)}{dx} = 1$. Este estimador se utilizó en [2] arrojando resultados satisfactorios.

3.4. ALGORITMO DE PENDIENTE MÁXIMA (SD O STEEPEST DESCENT, POR SUS SIGLAS EN INGLÉS)

3.4.1 Generalidades

De manera general, en el algoritmo de pendiente máxima hay una función de pesos denominada $L(\vec{w})$. Si los pesos cambian en cada iteración, entonces $L(\vec{w}(n))$, donde n indica el número de iteración.

Con este algoritmo se busca que $L(\vec{w}(n)) \geq L(\vec{w}(n + 1))$, la cual es una secuencia decreciente acotada abajo, con la que el algoritmo cambia de pesos para disminuir el error en cada iteración, minimizando J .

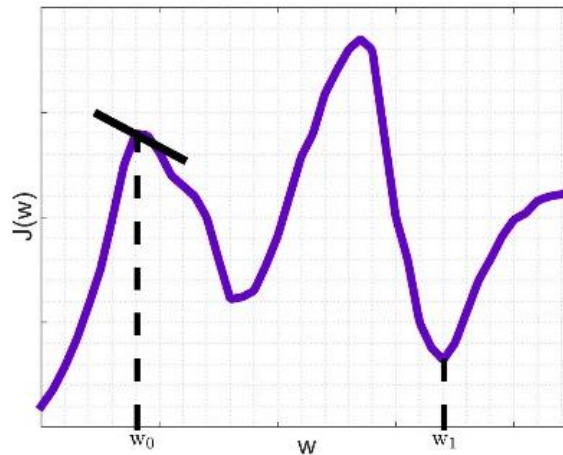


Figura 3-11 Pendiente máxima.

Algoritmo iterativo que busca mínimos locales, si se empieza en w_0 , se debe incrementar el peso puesto que no llega ningún mínimo.

Si la función que se desea optimizar es lineal se puede hacer en un paso, en caso contrario, se le asigna un valor inicial y se calcula el error. En la Figura 3-11, si se empieza en w_0 , se desea incrementar el peso, puesto que no va a converger en el mínimo local, si empieza en w_1 , la pendiente sería cero y no se necesitan más

CAPÍTULO 3. METODOLOGIA

iteraciones. Se usa la derivada para saber la dirección, tal como muestra la Ecuación (3.35)

$$\nabla L(\vec{w}) = \frac{\partial L(\vec{w})}{\partial \vec{w}}. \quad (3.35)$$

Para actualizar esta ecuación se utiliza la Ecuación (3.36)

$$\vec{w}(n+1) = \vec{w}(n) - \alpha \nabla J L(\vec{w}), \quad (3.36)$$

donde el coeficiente α es la razón de cambio de J

Expandiendo en series de Taylor a $L(\vec{w})$ con $\vec{w}(n)$:

$$L(\vec{w}(n+1)) = L(\vec{w}(n)) + \frac{\partial L(\vec{w}(n))}{\partial \vec{w}(n)} (\vec{w}(n+1) - \vec{w}(n)) + \dots \quad (3.37)$$

$$L(\vec{w}(n+1)) \cong L(\vec{w}(n)) + \frac{\partial L(\vec{w}(n))}{\partial \vec{w}(n)} (\vec{w}(n+1) - \vec{w}(n)). \quad (3.38)$$

Usando los pesos se actualiza la ecuación:

$$L(\vec{w}(n+1)) = L(\vec{w}(n)) + \nabla L(\vec{w}(n)) [-\alpha \nabla L(\vec{w}(n))], \quad (3.39)$$

$$L(\vec{w}(n+1)) = L(\vec{w}(n)) - \alpha \|\nabla L(\vec{w}(n))\|^2. \quad (3.40)$$

En general, se tiene que $L(\vec{w}(n+1)) < L(\vec{w}(n))$.

3.4.2 Planteamiento del problema

La función de costo que permitirá evaluar el dither que se encuentra a través del algoritmo de pendiente máxima es la suma de cuadrados de los residuales o errores (SSE o Sum of Squared Errors, por sus siglas en inglés). El objetivo es minimizar el SSE y está definido como enuncia la Ecuación (3.34):

CAPÍTULO 3. METODOLOGIA

$$SSE = \sum_{k=0}^{S-1} (R_{\vartheta\vartheta}[k] - \overline{R_{\vartheta\vartheta}})^2, \quad (3.41)$$

donde M representa el número de elementos del vector de autocorrelación $R_{\vartheta\vartheta}$.

Para el algoritmo de pendiente máxima, el dither se va modificando en cada iteración, puesto que el SSE se obtiene a partir del vector autocorrelación $R_{\vartheta\vartheta}$ y este a su vez, tal como se mencionó en la Sección 3.3.3, depende de $d[n]$. Así, la Ecuación (3.35) se reescribe:

$$\nabla SSE(\vec{d}) = \frac{\partial SSE(\vec{d})}{\partial \vec{d}}. \quad (3.42)$$

Para actualizar esta ecuación, se modifica la Ecuación (3.36):

$$\vec{d}[n+1] = \vec{d}[n] - \alpha \nabla SSE[\vec{d}] \quad (3.43)$$

Usando el vector de dither se actualiza la Ecuación (3.40):

$$SSE(\vec{d}[n+1]) = SSE(\vec{d}(n)) - \alpha \|\nabla SSE(\vec{d}(n))\|^2 \quad (3.44)$$

3.5. PRUEBA DE LJUNG-BOX

Si el ruido total, que es la diferencia entre la entrada y la salida del sistema (Figura 3-10(c)), es un ruido blanco, no se agregan armónicos indeseados a la señal de entrada. Para determinar si el ruido total es blanco, se verifica si es una secuencia que se distribuye independientemente, puesto que una secuencia independiente tiene un espectro blanco. Idealmente el error de cuantización es independiente de la señal a cuantizar $x[n]$, pero de manera general se acepta que el primer y el segundo momento no estén correlacionados [1][4].

En la Ecuación (3.12) se planteó las características que debía cumplir la autocorrelación del ruido de cuantización para ser considerado blanco, sin embargo, esto se cumple cuando la señal es de longitud infinita. Para señales de longitud finita la autocorrelación y el coeficiente de autocorrelación nunca serán exactamente cero, por lo tanto, para determinar si una secuencia finita de datos discreta puede ser considerado blanco se usará la prueba de Ljung-Box [24]. Esta prueba no se hace en cada desfase τ para evaluar su aleatoriedad, sino que se evalúa con base en el número de desfases y de esta manera se determinan su aleatoriedad de una forma generalizada.

Debido a que la autocorrelación de un ruido blanco, tal como describe la Ecuación (3.12) será cero en los desfases cuyo valor sean distintos de cero, la prueba de Ljung-Box se usa para verificar la hipótesis nula H_0 del ruido blanco, es decir, permitirá verificar si los datos de ruido total se distribuyen de manera independiente. Si se distribuyen de manera independiente, las autocorrelaciones del error total son cero demostrando la aleatoriedad del ruido. Si se rechaza la hipótesis nula,

CAPÍTULO 3. METODOLOGIA

entonces se acepta la hipótesis alterna: los datos no se distribuyen de manera independiente.

La prueba de Ljung-Box utiliza los primeros τ desfases del coeficiente de autocorrelación (Sección 3.2.1.2) para calcular el siguiente estadístico:

$$Q_{JB} = S(S + 2) \sum_{n=1}^h \frac{r_e^2[n]}{S - n}, \quad (3.45)$$

r_e es el coeficiente de autocorrelación y tal como se enunció en la sección 3.2.1.2 si éste es cero, entonces $\varepsilon[n]$ y $\varepsilon[n + \tau]$ son independientes y por lo tanto su autocorrelación es cero para todos $\tau \neq 0$, siendo esta una de las características principales del ruido blanco.

Q_{JB} se acerca asintóticamente a la distribución chi-cuadrado (X^2) con h grados de libertad para incrementar M bajo la hipótesis nula de ruido blanco. Para que Q_{JB} se acerque a su distribución asintótica, el número de desfases τ debe ser menor que el número de muestras S .

La región para el rechazo a la hipótesis nula por el nivel de significancia α es $Q_{JB} > X_{1-\alpha, h}^2$ donde $X_{1-\alpha, h}^2$ es el α -cuantil de la distribución chi-cuadrado con h grados de libertad [24].

Cuando la hipótesis nula es cierta el valor p se define como la probabilidad de obtener un valor para el estadístico Q_{JB} igual o más extremo que el obtenido. Por lo tanto, un valor p pequeño ($p < \alpha$) la hipótesis nula debe ser rechazada.

CAPÍTULO 3. METODOLOGIA

3.6. OPERACIONES DE PUNTO FLOTANTE

Para medir la complejidad del algoritmo propuesto y su tiempo de procesamiento se comparan los resultados con los propuestos en [1], donde se usa el algoritmo de Levenberg-Marquardt (LM) en lugar de SD.

En [1], a partir de una señal original, se toma una ventana de 512 muestras, con las que se obtiene una señal de dither adaptativo que pasa la prueba de Ljung-Box. Una vez se tiene una señal de dither inicial, se procede a calcular una señal de dither para el resto de la señal usando una ventana de 512 valores que se desplaza 256 muestras, de tal forma que se tienen 256 incógnitas.

Se cuentan, entonces, el número de operaciones de coma flotante, donde se incluye el número de sumas y multiplicaciones por iteración, y se calcula un promedio de iteraciones para verificar la rapidez de los algoritmos con las señales de dither triangular, el dither calculado con SD y con LM.

Con el algoritmo de LM hay un promedio de 780 iteraciones para calcular las 256 incógnitas. Para solucionar la Ecuación (3.30), para el segmento equivalente a 256, se necesitarían 131,072 multiplicaciones, resultado de multiplicar 512 por 256, y 256 sumas en una iteración. Al multiplicar por el número de iteraciones promedio, se tendrá un total de 102,236,160 multiplicaciones y 200,448 sumas para una señal de 256 muestras.

Para SD, la ventana se está moviendo una a la vez, de tal manera que se tiene una sola incógnita. Como se empieza con una señal de 512, se necesitarán 256 valores adicionales y el promedio de iteraciones para encontrar cada uno de los 256 valores

CAPÍTULO 3. METODOLOGIA

es 5, así que serán 513 multiplicaciones y 1 suma en la primera iteración. Luego el número de multiplicaciones y sumas aumentará en 1 en cada muestra, de tal manera que para la segunda iteración serán 514 multiplicaciones y 2 sumas. Debido a que el promedio es de 5 iteraciones, el total de multiplicaciones para encontrar el segmento equivalente a 256 será 163,968, que se obtiene del resultado de la sumatoria $\sum_{k=513}^{768} k$ y 32,896 sumas, resultado de $\sum_{k=1}^{256} k$, y si se obtiene el valor del número total de multiplicaciones para el promedio de 5 iteraciones será 819,840 y el de las sumas será 164,480.

3.7. RESUMEN

En este capítulo se abarcaron conceptos básicos de re-cuantización, dithering, el algoritmo de pendiente máxima y la prueba de Ljung-Box. En las primeras dos secciones se desarrolló el análisis matemático del proceso de re-cuantización y el ruido de cuantización. De manera general el ruido se asume blanco cuando el cambio de amplitud es grande de muestra a muestra, de tal manera que la autocorrelación será un impulso en el desfase equivalente a cero. En la tercera sección, se explica el dithering y su efecto en el ruido total, se concluye que la autocorrelación depende de la señal de dither, para luego, en la cuarta sección, encontrar esta señal de dither que minimice el SSE y mantenga una varianza constante con un ruido mínimo. En la quinta sección se introduce la prueba de Ljung-Box para evaluar si el error total que se obtiene con la suma del dither encontrada a través del algoritmo de pendiente máxima, es un ruido blanco.

Capítulo 4

EXPERIMENTOS Y RESULTADOS

En este capítulo se muestran experimentos con señales de audio sintéticas, estos experimentos están diseñados para realizar evaluaciones que permitan determinar si el ruido total es un ruido blanco. Además, se verifica si el dithering adaptativo alcanza la varianza deseada con cantidades distintas de muestras, en diferentes niveles de cuantización y con valores distintos de varianza deseada. Se calcula el SNR y se compara con los resultados obtenidos en [2].

Empezaremos diciendo que la hipótesis nula es que el error total es blanco y se usará un nivel de significancia del 10%.

4.1. NIVELES DE CUANTIZACIÓN

4.1.1. Señales de audio

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

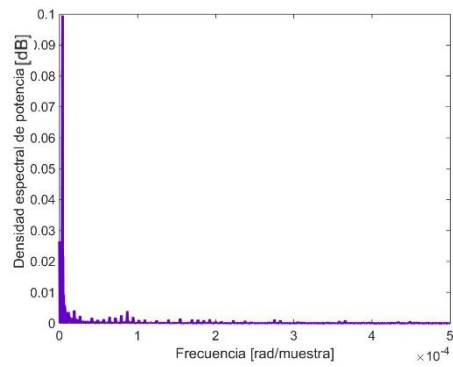
El propósito de los primeros experimentos es determinar la relación existente entre los niveles de cuantización y la naturaleza del ruido total obtenido, de manera que se pueda decir si se acepta o se rechaza la hipótesis nula (Ecuación 3.39).

Para esta primera parte se usan dos señales, una sintética generada a partir de una señal coseno y una señal de audio real, ambas de longitud total de 3,873,448, con una frecuencia de 628 rad/s para la sintética y de 44.1 kHz para la real y una precisión de $N = 24$ bits para ambas señales. Se cuantizan a distintos niveles de cuantización: (Ecuación 3.4). En esta parte no se le agrega ningún ruido adicional y se compara cuando se le agregan ruidos, que son el ruido triangular, el ruido de obtenido a partir del algoritmo de Levenberg-Marquardt propuesto por Benítez en [2] y un ruido que se calcula a través del algoritmo de pendiente máxima. Estos ruidos se agregan a las señales por separado, antes de ser cuantizadas para mitigar los efectos del proceso de cuantización. El propósito de unir los experimentos con los dos ruidos es comparar el algoritmo propuesto con el que se encuentra en la investigación de punta.

Una vez cuantizadas las señales, se obtiene el ruido de cuantización, se verifica sus varianzas y su autocorrelaciones para evaluar si tienen las propiedades estadísticas de un ruido blanco (Ecuación 3.13). Se calculan además los valores para ambas señales obtenidos a partir de las pruebas de hipótesis.

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

(a)



(b)

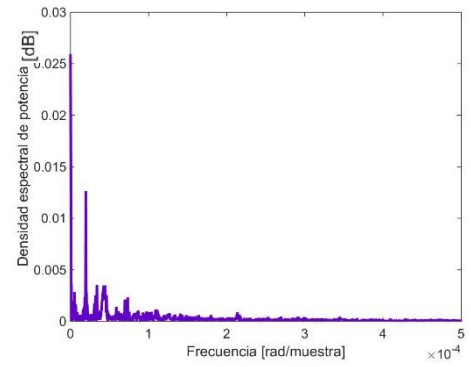
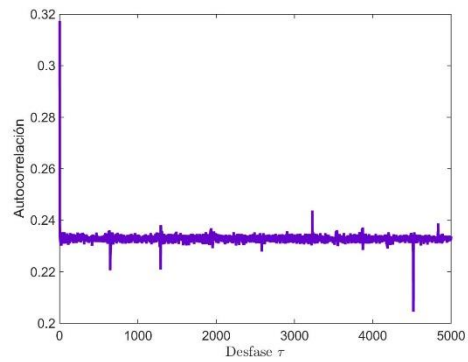
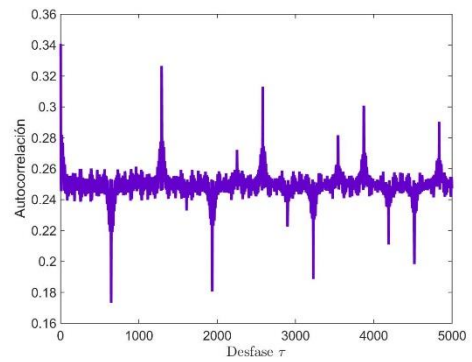


Figura 0-1 Espectro de la (a) señal de audio sintética y la (b) señal real cuantizada x_q para $N=8$

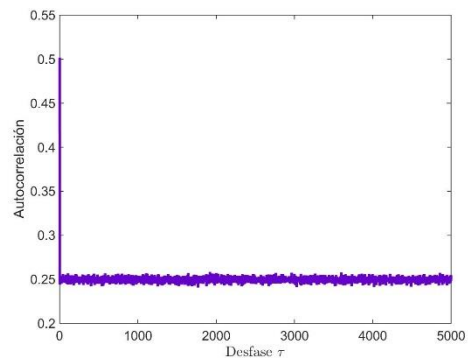
(a)



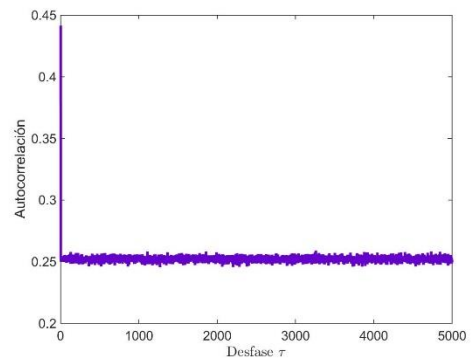
(b)



(c)

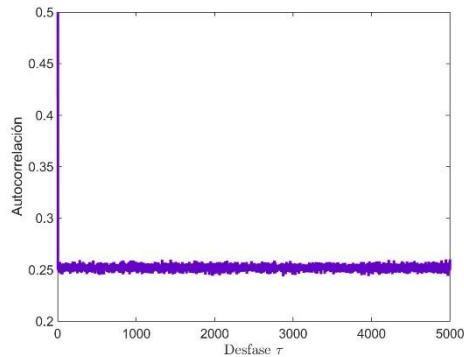


(d)



CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

(e)



(f)

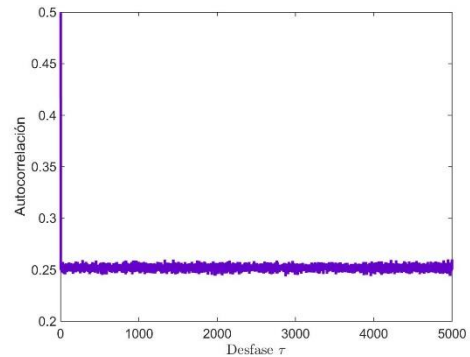


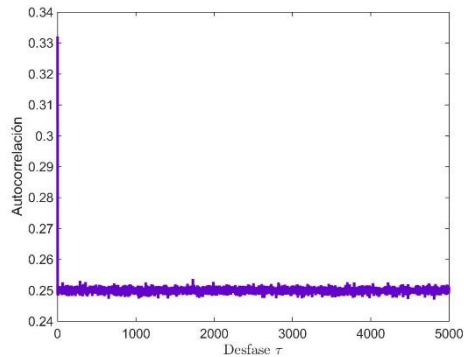
Figura 0-2 Autocorrelación para distintos niveles de cuantización de la señal sintética sin ruido de dither (a) $N = 10$ y (b) $N = 4$, con ruido triangular para (c) $N = 10$ y (d) $N = 4$, y con dither de SD (e) $N = 10$ y (f) $N = 4$

En el espectro de las señales de la Figura 0-1 se muestran armónicos indeseados y se puede apreciar en la Figura 0-2 (a) y (b) que la autocorrelación del ruido de cuantización presenta varios picos de valores considerables en desfases distintos de 0, mostrando propiedades que no son propias del ruido blanco. Para los casos donde N tiene valores correspondientes a $\{16,17,18,19,20,21,23\}$, el valor de autocorrelación para todos los desfases es despreciable debido a que el error de cuantización es significativamente pequeño para todas las muestras, ya que el cambio en la precisión tiene un valor que no afecta la resolución de señal.

Por su parte, en la señal real para los valores de $N = \{8,9,10,11,12,15\}$ se presenta una gráfica de autocorrelación como la de la Figura 0-3(b), cuando $N = 4$, la gráfica de autocorrelación del ruido de cuantización presenta varias oscilaciones, mostrando que en la cuantización a valores bajos, se afecta la resolución de la señal.

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

(a)



(b)

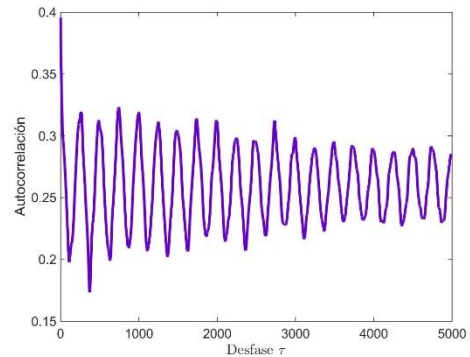
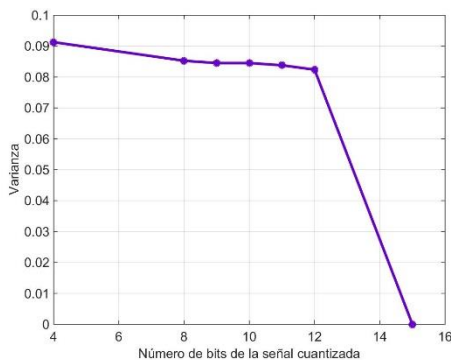


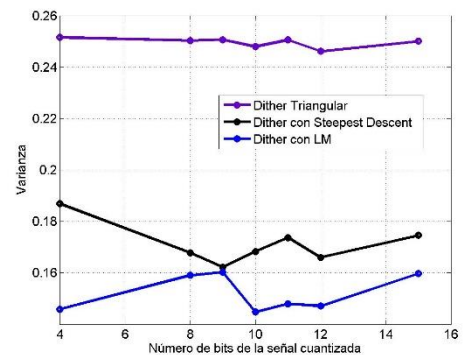
Figura 0-3 Autocorrelación para distintos niveles de cuantización de la señal de audio real para $N=4$ para (a) Señal sin dither (b) Señal con dither triangular

En la Figura 0-4 se puede apreciar, que la varianza de los ruidos de cuantización aumenta a medida que aumentan los niveles de cuantización, por lo general se presentó un cambio considerable en los valores de varianza para números de bits superiores a 9 para la señal sin dither. En cuanto a las señales con dither, se presenta en (b) y (d), que para el dither triangular no se puede escoger una potencia específica para la señal de dither, manteniéndose en aproximadamente 0.25. En cuanto a LM y SD, el dither cumple con las especificaciones de varianza, que para este caso fue de 0.15.

(a)

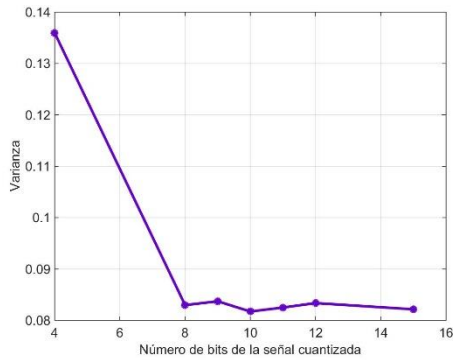


(b)



CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

(c)



(d)

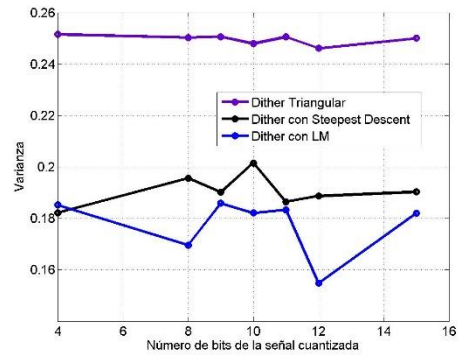
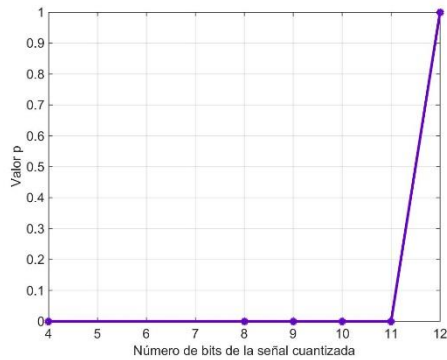


Figura 0-4 Varianza del error de cuantización ϵ para distintos niveles de cuantización (a) Señal sintética sin dither (b) Señal sintética con dither (c) Señal real sin dither (d) Señal real con dither

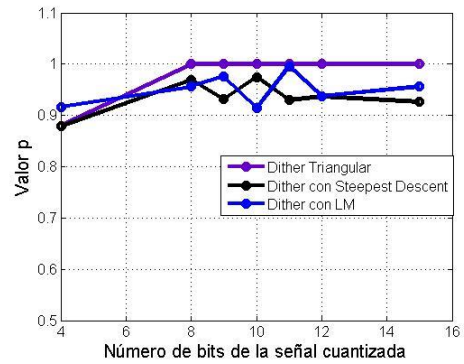
Se observa además en la Figura 0-5 como se rechaza la hipótesis nula para niveles de cuantización alrededor de $N = 12$. Para verificar si se acepta la hipótesis nula o la alterna, se obtienen nuevamente los valores p a través de la prueba de Ljung-Box para las señales con dither, observándose como el valor de p empieza a aumentar y para varios niveles de cuantización en los que se rechazó la hipótesis nula, ahora se acepta, y en Figura 0-5 (b) y (d) se puede apreciar como en algunos niveles de cuantización con el dither de SD el valor de p es 1 mientras con el ruido de FDP triangular los valores se mantienen por debajo de 0.5, de tal manera que para algunos niveles de cuantización, el ruido total es un ruido que cumple con las propiedades estadísticas de un ruido blanco. Sin embargo, se observar que el algoritmo de [2] cuando la señal es real, presenta un mejor desempeño, puesto que para niveles de cuantización superiores a 8 bits, el ruido resulta ser blanco.

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

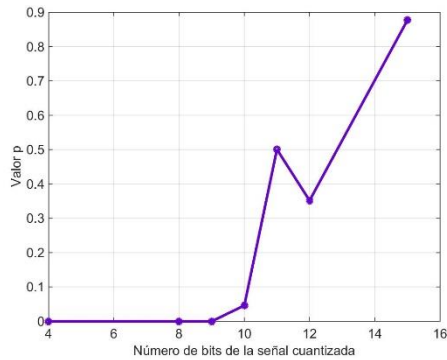
(a)



(b)



(c)



(d)

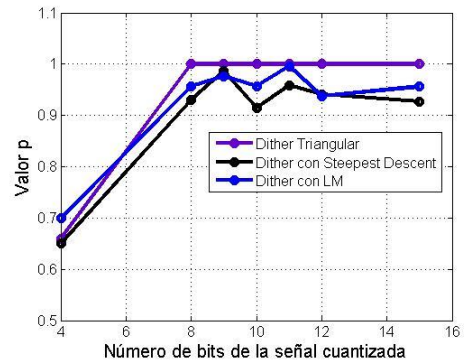
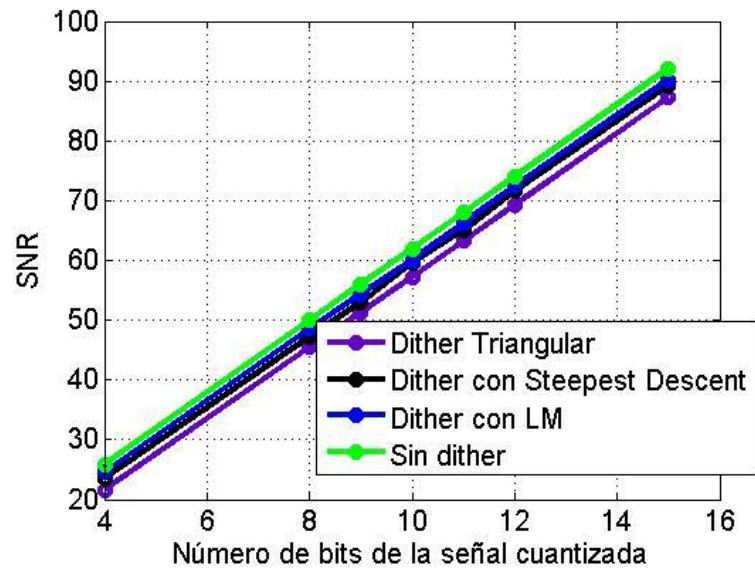


Figura 0-5 Valor de p para distintos niveles de cuantización (a) Señal sintética sin dither (b) Señal sintética con dither (c) Señal real sin dither (d) Señal real con dither

De la Figura 0-6 podemos apreciar como el valor de SNR para la autocorrelación que se obtuvo cuando se sumó dither de SD, tiene valores superiores a los que se obtuvieron con el ruido triangular. Se puede aseverar entonces, que el proceso de sumar una señal de dither hace que las muestras no estén correlacionadas con el error total de cuantización.

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

(a)



(b)

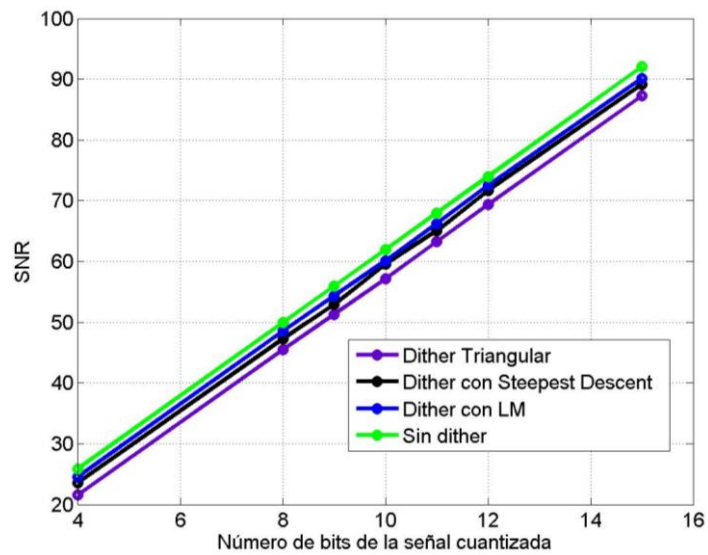


Figura 0-6 SNR para distintos niveles de cuantización para (a) Señal sintética (b) Señal real

La Figura 0-7 muestra el error de cuantización para las primeras 250 muestras para $N = 10$.

(a)

(b)

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

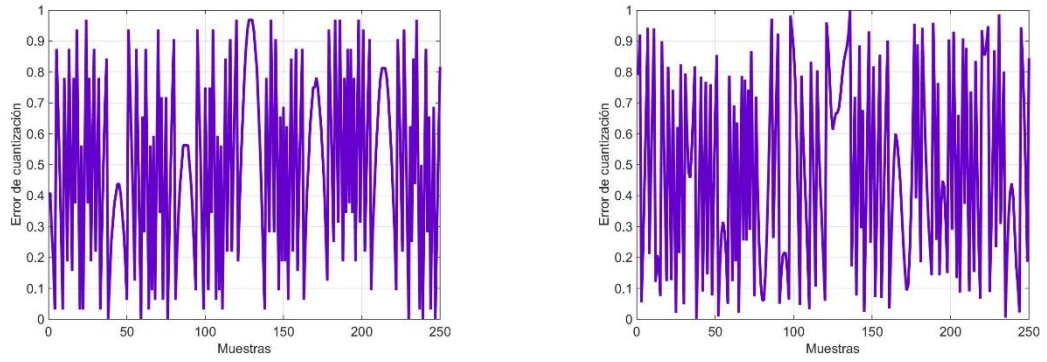


Figura 0-7 Error de cuantización para (a) señal sintética cuando $N=10$ y (b) señal real cuando $N=8$

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

4.1.2. Comparación de tiempo

Para comparar el tiempo que tomó cada algoritmo en correr, se evalúan los respectivos tiempos en segundos para los experimentos llevados a cabo en la sección anterior.

Tabla 1Tiempo en segundos de algoritmos

<i>q</i>	<i>SD</i>	<i>LM</i>
20	420.87	3092.46
16	435.04	3066.00
15	431.82	3198.22
14	431.29	3319.64
13	447.04	3341.69
12	425.64	5340.36
9	1050.54	3036.13

Con base en las operaciones de punto flotante, el número de multiplicaciones de LM por iteración en promedio son aproximadamente cuatro veces la de SD, en cuanto a los tiempos, en promedio se obtuvo que el algoritmo de SD es aproximadamente cinco veces más rápido.

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

4.2. VARIANZA

El propósito de este experimento es verificar con las señales de audio sintético y real si se alcanza la varianza deseada del error total cuando se configura a distintos valores. En la Figura 4-8 y Figura 0-9 en el eje horizontal están los valores de varianza deseado, debido a que para los algoritmos de SD y LM [2] se puede ajustar el nivel de ruido. En el eje vertical se encuentran los valores obtenidos de las varianzas del error total usando los distintos dithers.

Las varianzas se fijaron a valores de 0.11, 0.12, 0.13, 0.15, 0.2, 0.25 y 0.3, entonces se aprecia, por ejemplo, como en la Figura 4-8 cuando se desea que la varianza del error total sea 0.11, el algoritmo de LM alcanza una varianza de aproximadamente 0.11 y el de SD una varianza de aproximadamente 0.13.

Para el dither triangular en la Figura 4-8 y Figura 0-9, la varianza del error total queda fija en aproximadamente 0.25, puesto que a diferencia de SD y LM no se puede escoger un nivel de ruido específico.

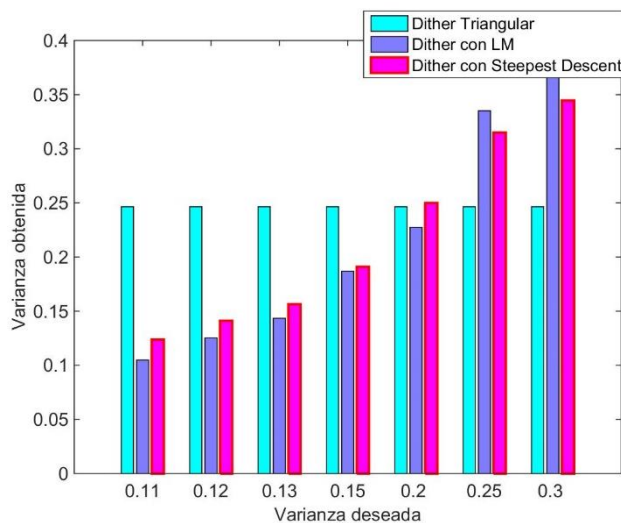


Figura 0-8 Comparación de varianzas obtenidas con los distintos dithers para señal de audio sintética

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

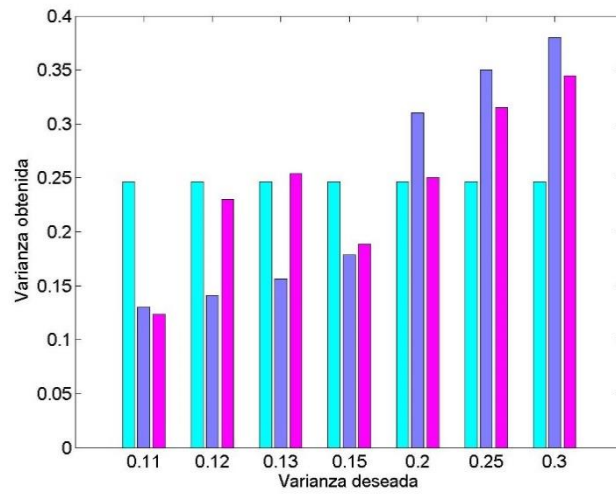


Figura 0-9 Comparación de varianzas obtenidas con los distintos dithers para señal de audio real

4.3. IMÁGENES

Para mostrar los efectos de la cuantización y de la suma de dither, se muestran las siguientes imágenes. De igual manera que con audio se toma una imagen, se cuantiza sin dither a 4 bits, se compara con la misma imagen cuantizada a 4 bits con dither triangular y el dither obtenido a partir de del algoritmo de SD antes de ser cuantizada.

(a)



(b)



(c)



(d)



Figura 0-10 (a) Imagen original (b) Imagen cuantizada a 4 bits (c) Imagen con dither triangular (d) Imagen con dither SD

Se observa cómo se pierde información de la imagen cuando se cuantiza sin ningún dither, mientras que en (c) y (d) los ruidos agregados hacen que no se presenten cambios abruptos entre las diferentes texturas y tonos de la imagen, por lo tanto, la calidad de la imagen es mejor. Con (d) se pueden apreciar cambios bastantes tenues entre las texturas presentando un mejor efecto visual.

4.4. RESUMEN

En este capítulo se realizaron una serie de experimentos para determinar la validez del algoritmo propuesto, se realizó la cuantización a distintos niveles de cuantización sin dither y con ruido y se determinó si el ruido obtenido era blanco para cada uno de los casos. Se analizaron las densidades espectrales de potencia y se determinó

CAPÍTULO 4. EXPERIMENTOS Y RESULTADOS

la relación señal a ruido. De igual manera, se verificó si con distintas varianzas fijas para el error total, el algoritmo cumplía con las especificaciones.

Se compararon los tiempos con el algoritmo de LM propuesto en [2] para determinar si concordaban con el análisis de las operaciones de punto flotante mencionadas en la Sección 3.6.

Capítulo 5

CONCLUSIONES Y TRABAJOS FUTUROS

Con esta investigación se realizaron pruebas de hipótesis que permitieron demostrar si el ruido calculado generaba un error total con espectro blanco. Con las pruebas realizadas a la señal de audio sintética, se permitió evaluar si los valores de p para rechazar o aceptar la hipótesis nula, usando un nivel de significancia del 10%, el ruido resultó ser blanco.

Cada vez que se agregaba una nueva muestra a la señal para calcular el valor del ruido correspondiente a la posición de esta señal se realizaban las pruebas de hipótesis para determinar si el error total seguía siendo blanco. Además, se verificaban los valores de varianza y de SNR.

CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS

Para la mayoría de los experimentos se cambia la resolución de la señal original de 24 a 16 bits. Para solucionar el sistema de ecuaciones no lineal se usó el algoritmo de Steepest Descent, que usaba una sola derivada. Debido a que al cambiar una sola muestra no se alteraban considerablemente las propiedades estadísticas de la señal, el algoritmo resultó ser considerablemente rápido. Se usó una aproximación lineal para resolver el sistema no lineal de ecuaciones.

Adicionalmente se usó la autocorrelación circular para resolver el sistema de ecuaciones, de esta manera se realizaban la mitad de las operaciones, se contaba con todas las muestras en cada operación y se eliminaba la correlación en el algoritmo de SD. La hipótesis nula no se rechazó con este estimador de autocorrelación circular.

Se realizaron distintos experimentos para evaluar los efectos la cantidad de bits para realizar la cuantización y además se escogieron distintos valores de varianza para determinar en qué escenarios los resultados eran más satisfactorios. Se compararon los resultados con sistemas donde no se le sumaba ninguna señal de dither y donde se le sumaba dither con FDP triangular. Los experimentos se realizaron utilizando una señal de audio sintética, una señal de audio real y una imagen.

CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS

El SNR fue ligeramente mayor al que registra la revisión de literatura, por tanto, la varianza del error fue inferior. Adicionalmente el número de operaciones empleado fue menor que el de la investigación de punta. Al evaluar los tiempos a través de la herramienta MATLAB® también se registraron tiempos inferiores.

REFERENCIAS

- [1]. C. Fabian Benítez-Quiroz and S. Hunt, "Adaptive dithering of one dimensional signals," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 3757–3760.
- [2]. C. Fabian Benítez-Quiroz. "Adaptive dithering of one dimensional
- [3]. R. A. Wannamaker, S. P. Lipshitz, J. Vanderkooy, and J. N. Wright, "A theory of nonsubtractive dither," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 499–516, 2000.
- [4]. S. P. Lipshitz and J. Vanderkooy, "Digital Dither," in *Audio Engineering Society Convention 81*, 1986.
- [5]. B. M. Oliver, J. R. Pierce, and C. E. Shannon, "The Philosophy of PCM," *Proc. IRE*, vol. 36, no. 11, pp. 1324–1331, Nov. 1948.
- [6]. W. R. Bennett, "Spectra of Quantized Signals," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 446–472, Jul. 1948.
- [7]. Widrow, B., "A Study of Rough Amplitude Quantization by Means of Nyquist Sampling Theory," *Circuit Theory, IRE Transactions on* , vol.3, no.4, pp. 266-276, Dec. 1956

- [8]. Widrow, B., "Statistical analysis of amplitude-quantized sampled-data systems," American Institute of Electrical Engineers, Part II: Applications and Industry, Transactions of the, vol.79, no.6, pp. 555-568, Jan. 1961.
- [9]. C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
- [10]. G. Zames and N. Shneydor, "Dither in nonlinear systems," *IEEE Trans. Automat. Contr.*, vol. 21, no. 5, pp. 660–667, Oct. 1976.
- [11]. F. A. Malekzadeh, R. Mahmoudi, and A. van Roermund, *Analog Dithering Techniques for Wireless Transmitters*. Springer Science & Business Media, 2012.
- [12]. Roberts, L., "Picture coding using pseudo-random noise," Information Theory, IRE Transactions on, vol.8, no.2, pp. 145-154, February 1962.
- [13]. L. G. Roberts. Picture Coding Using Pseudo-Random Noise. MIT, S.M. thesis, 1961.
- [14]. Jayant, N.S.; Rabiner, L.R., "The Application of dither to the quantization of speech signals," Bell System Technical Journal, The, vol.51, no. 6, pp. 1293-1304, July-Aug. 1972.
- [15]. Schuchman, L., "Dither Signals and Their Effect on Quantization Noise," Communication Technology, IEEE Transactions on, vol.12, no. 4, pp. 162-165, December 1964.

- [16]. Sherwood, D., "Some Theorems On Quantization And An Example Using Dither," *Circuits, Systems and Computers*, 1985. Nineteenth Asilomar Conference on, vol., no., pp. 207-212, 6-8 Nov. 1985.
- [17]. Gray, R.M.; Stockham, T.G., Jr., "Dithered quantizers," *Information Theory, IEEE Transactions on*, vol.39, no. 3, pp. 805-812, May 1993.
- [18]. Wannamaker, R.A.; Lipshitz, S.; Vanderkooy, J.; Wright, J.N., "A theory of nonsubtractive dither," *Signal Processing, IEEE Transactions on*, vol. 48, no. 2, pp. 499-516, Feb 2000.
- [19]. J. Vanderkooy and S. P. Lipshitz, "Dither in Digital Audio," *J. Audio Eng. Soc.*, vol. 35, no. 12, pp. 966–975, Dec. 1987.
- [20]. M. J. M. Pelgrom, *Analog-to-Digital Conversion*. Dordrecht: Springer Netherlands, 2010.
- [21]. B. Widrow and I. Kollar, "Statistical theory of quantization," *IEEE Trans. Instrum. Meas.*, vol. 45, no. 2, pp. 353–361, Apr. 1996.
- [22]. Sanjit Mitra, "Digital Signal Processing – A Computer Based Approach 2nd ed.," McGraw-Hill Irwin, 2001.
- [23]. R. A. Wannamaker. *The Theory of Dithered Quantization*,. University of Waterloo, 1997.
- [24]. G. M. LJUNG and G. E. P. BOX, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, Aug. 1978.