

Introducción al reconocimiento de patrones: Preprocesamiento de señales

M. Sc. Saúl Calderón Ramírez
Instituto Tecnológico de Costa Rica,

Escuela de Computación, bachillerato en Ingeniería en Computación,
PAtern Recognition and MArine Learning Group (PARMA-Group)

April 22, 2019

Abstract

Este material está basado en el libro *Análisis de señales*, de Pablo Irarrazabal[2],
y el libro de Procesamiento digital de imágenes, de Rafael González, [1].

1 Introducción

Una señal puede definirse como una serie de valores, con origen usualmente de fenómenos físicos: sonidos, correspondientes a alteraciones en la presión del aire, intensidades lumínicas en distintas longitudes de onda, de origen biológico como la presión arterial, etc. Existen tres grandes grupos, que se resumen en el siguiente cuadro:

Señales	Espacio	Amplitud	Ejemplos
Analógicas	Continua	Continua	Señales sísmicas, electrocardiogramas, ondas cerebrales, sonido, etc.
Discretas	Discreta	Continua	Señales muestreadas en dispositivos analógicos
Digitales	Discreta	Discreta	Imágenes, señales y sonidos en un computador digital

Table 1: Tipos de señales.

Las señales naturalmente son analógicas, y dado que gran parte de los computadores actuales son digitales, es necesario realizar el proceso de muestreo, el cual consiste en dos etapas: discretizar primero el dominio del tiempo, tomando una cantidad de muestras finitas y almacenandolas en un dispositivo analógico como un condensador, para posteriormente realizar la etapa de cuantificación

(codificación en binario) de las muestras, de modo que la amplitud sea también discreta. La Figura 1 muestra el diagrama de un circuito de muestreo en el tiempo y en amplitud básico.

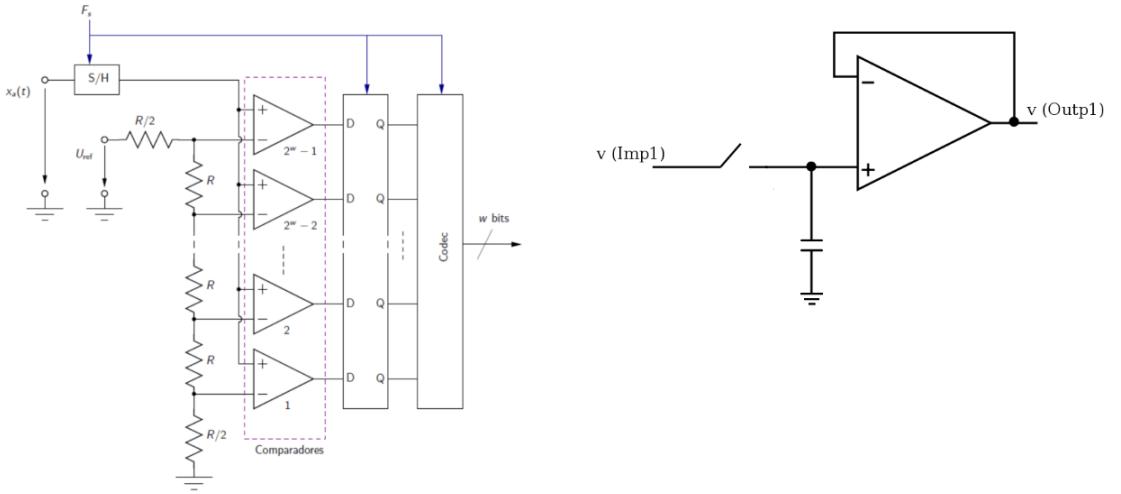


Figure 1: Circuito de muestreo.

La Figura 2 muestra el proceso de digitalización de una señal mencionado. En el primer gráfico, únicamente el dominio del tiempo es discreto, en el segundo, tanto el dominio del tiempo como el de la intensidad o amplitud de la señal son discretos, lo que corresponde entonces a una señal digital.

2 Números complejos

Un número complejo $z \in \mathbb{C}$ se define como aquel que tiene una parte real y otra imaginaria, por lo que está dado de la forma:

$$z = x + yi,$$

donde $\operatorname{Re}(z) = x$ corresponde a la parte real y $\operatorname{Im}(z) = y$ a la imaginaria. La constante i es la unidad imaginaria que conceptualmente cumple con:

$$i^2 = -1.$$

Todo número complejo $z \in \mathbb{C}$ con $\operatorname{Re}(z) = 0$ se refiere como un número imaginario.

Un número complejo se puede representar como un par ordenado $z = (x, y)$ o un vector $z = [x \ y]$, por lo que entonces se puede dibujar en el plano cartesiano, donde el eje de las abscisas representa la parte real un par de números complejos, como se muestra en la Figura 3

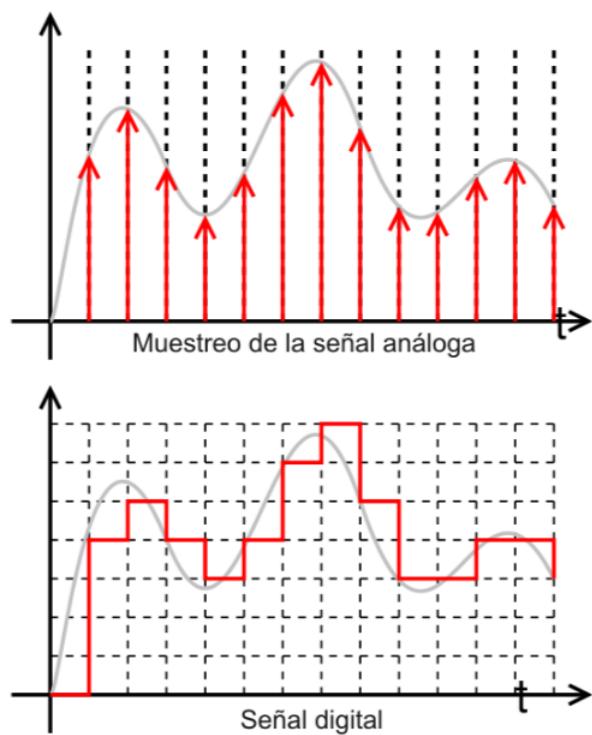


Figure 2: Señal digital. Tomado de <https://ccvrede1.wordpress.com/2012/10/06/senales-y-codificacion-parte-iv/>

Esto significa que todo número complejo tiene una magnitud y un ángulo, como también lo ilustra la Figura 3, definidos respectivamente como:

$$r = \|z\| = \sqrt{x^2 + y^2}$$

$$\varphi = \tan^{-1} \left(\frac{y}{x} \right).$$

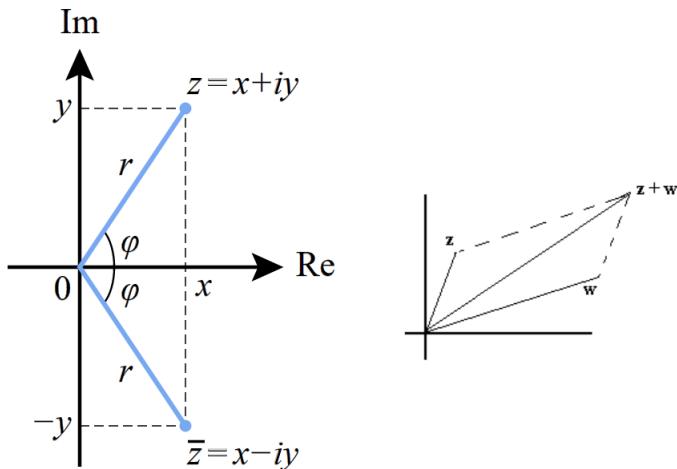


Figure 3: Número complejo en coordenadas cartesianas. Tomado de https://en.wikipedia.org/wiki/Complex_number

La suma, resta, producto y división por componente $z, w \in \mathbb{C}$ se definen tal cual se hace para vectores en \mathbb{R}^2 , como se muestra también en la Figura 3.

El conjugado de un número complejo $z = x + iy$, se define como $\bar{z} = x - iy$ y se muestra también en la Figura 3.

Representación en coordenadas polares para números complejos

De la Figura 3, se puede extraer fácilmente por trigonometría básica que la parte real e imaginaria de un número complejo $z = x + iy$ se puede expresar en términos únicamente del ángulo φ :

$$\cos(\varphi) = \frac{x}{r} \Rightarrow x = r \cos(\varphi)$$

$$\sin(\varphi) = \frac{y}{r i} \Rightarrow y = r i \sin(\varphi),$$

lo cual implica que el número puede representarse en términos únicamente de su magnitud y ángulo, como lo sugiere la Figura 3, con lo cual:

$$z = r (\cos(\varphi) + i \sin(\varphi)).$$

Por el teorema de Euler que establece que:

$$e^{i\varphi} = \cos(\varphi) + i \sin(\varphi).$$

El teorema de Euler se deduce al utilizar las series de McLaurin para desarrollar la parte real e imaginaria de z en coordenadas polares:

$$\begin{aligned} i \sin(\varphi) &= i \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} \varphi^{2n+1} = i\varphi - i \frac{\varphi^3}{3!} + i \frac{\varphi^5}{5!} - \dots \\ \cos(\varphi) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} \varphi^{2n} = 1 - \frac{\varphi^2}{2!} + \frac{\varphi^4}{4!} - \dots \\ e^z &= \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \dots \end{aligned}$$

Observe que si se sustituye $z = i\varphi$ en la serie de McLaurin de $e^z = e^{i\varphi}$ se tiene que:

$$e^{i\varphi} = \sum_{n=0}^{\infty} \frac{(i\varphi)^n}{n!} = 1 + (i\varphi) + \frac{(i\varphi)^2}{2!} + \frac{(i\varphi)^3}{3!} + \dots$$

y dado que $i^n = -1$ con n par y $i^m = -i$ con m impar, se tiene que:

$$e^{i\varphi} = \sum_{n=0}^{\infty} \frac{(i\varphi)^n}{n!} = 1 + (i\varphi) - \frac{\varphi^2}{2!} - \frac{i\varphi^3}{3!} + \dots$$

que equivale a hacer la suma de las series de $i \sin(\varphi)$ y $\cos(\varphi)$ se obtiene la equivalencia:

$$e^{i\varphi} = \cos(\varphi) + i \sin(\varphi) = 1 + (i\varphi) - \frac{\varphi^2}{2!} - \frac{i\varphi^3}{3!} + \dots.$$

Con ello se concluye que cualquier número complejo puede representarse en términos de su magnitud y ángulo (coordenadas polares), de forma compacta con una exponencial compleja, según el teorema de Euler:

$$z = r e^{i\varphi}.$$

2.1 Funciones complejas

Pueden caracterizarse tres tipos de funciones complejas:

1. **Funciones con argumento complejo y codominio real:** $y = f(x)$ con $x \in \mathbb{C}$ y $y \in \mathbb{R}$. Su gráfico corresponde a la graficación de una superficie o *manto* sobre el dominio de dimensión dos, como se muestra en la Figura 4. Muchas veces es innecesario graficar todos los puntos en la superficie, y es suficiente mostrar donde se encuentran los **polos** (puntos donde la

función $f(x)$ se indefine o diverge) y **ceros** (puntos donde la función $f(x)$ se hace cero), como también se muestra en la Figura 4, graficación de la función:

$$f_1(x) = \frac{x - (5 + 5i)}{(x - (-3 + 3i))(x - (-3 - 3i))}$$

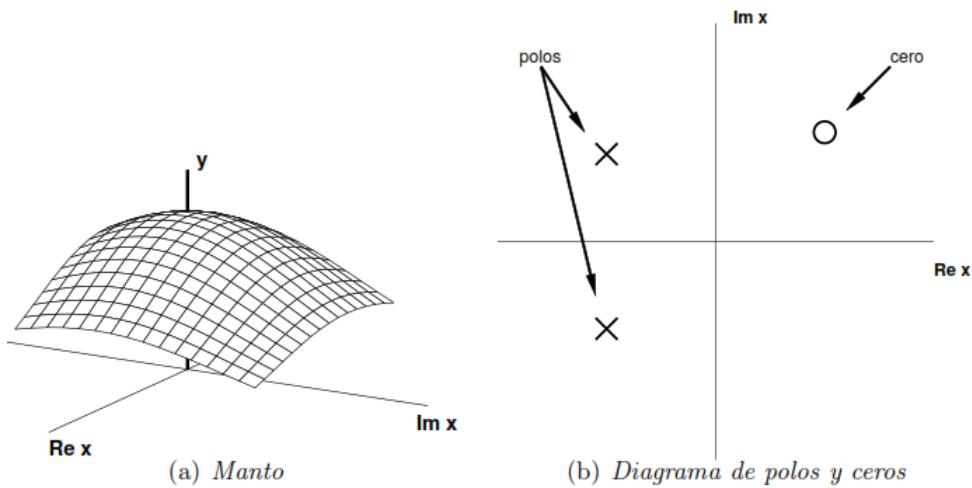


Figure 4: Diagramación de una función con argumento complejo y salida real.
Tomado de [2]

2. **Funciones con argumento real y codominio complejo:** $y = f(x)$ con $y \in \mathbb{C}$ y $x \in \mathbb{R}$. Estas funciones tienen definida su entrada en una dimensión, y su salida en dos dimensiones. Su graficación puede no ser sencilla, como muestra el apartado (a) de la Figura 5, por lo que es usual graficar la parte real e imaginaria por separado, ya sea en coordenadas rectangulares o polares (en términos de la magnitud y la fase). Un ejemplo de una función de este tipo es $y = f(x) = x + i$.

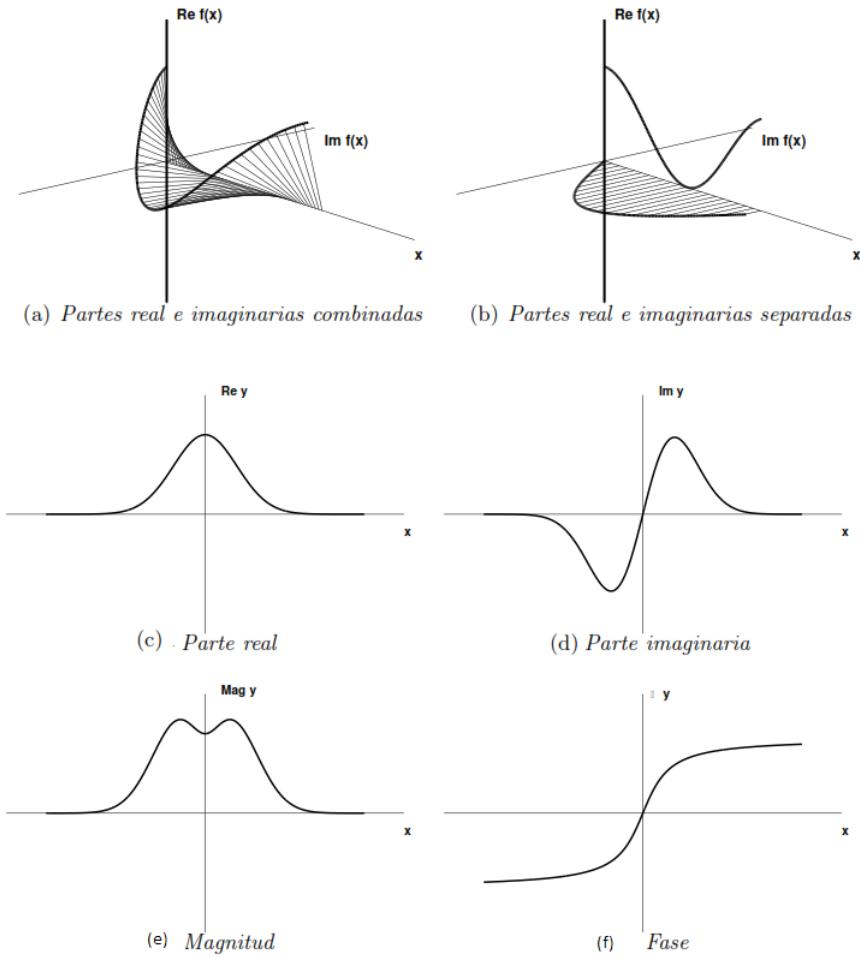


Figure 5: Graficación de funciones con argumento real y codominio complejo.
Tomado de [2].

3. **Funciones con argumento y codominio complejo:** $y = f(x)$ con $y, x \in \mathbb{C}$. Graficarlas requiere de dibujar una superficie por la parte real e imaginaria del codominio y , o un diagrama de polos y ceros por cada parte del codominio, como muestra la Figura 6.

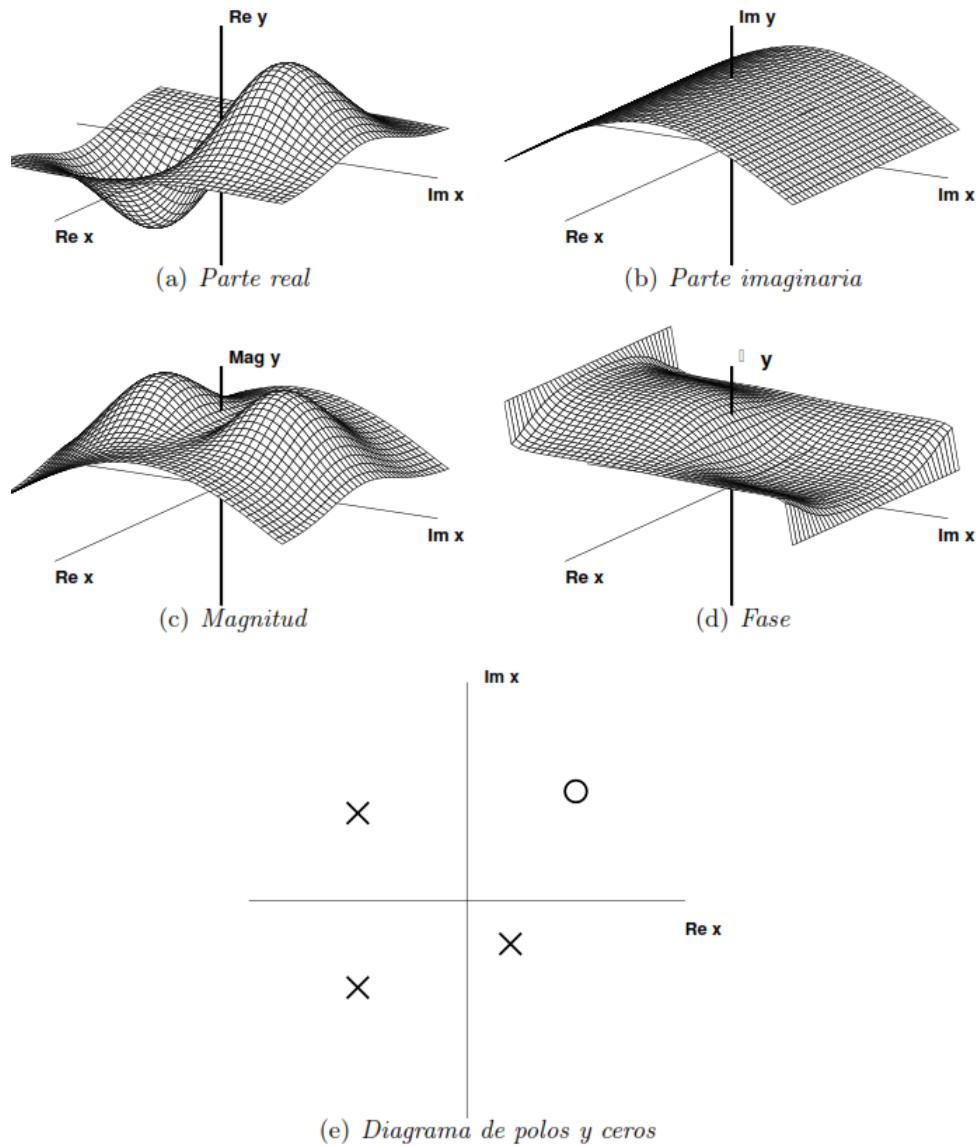


Figure 6: Diagramas de ejemplo para funciones de dominio y codominio complejo. Tomado de [2] .

3 Funciones importantes

A continuación se presentan algunas funciones importantes en el análisis de señales y sus propiedades.

3.1 Funciones simétricas

La simetría es una propiedad en las funciones que facilita el análisis de señales. Se define una función par aquella que cumple con:

$$f(x) = f(-x) \forall x,$$

lo que gráficamente quiere decir que las funciones pares son simétricas con respecto al eje de las ordenadas.

Una función es impar si cumple con:

$$f(-x) = -f(x) \forall x,$$

lo cual corresponde con una anti-simetría respecto al eje y , como se ilustra en la Figura 7.

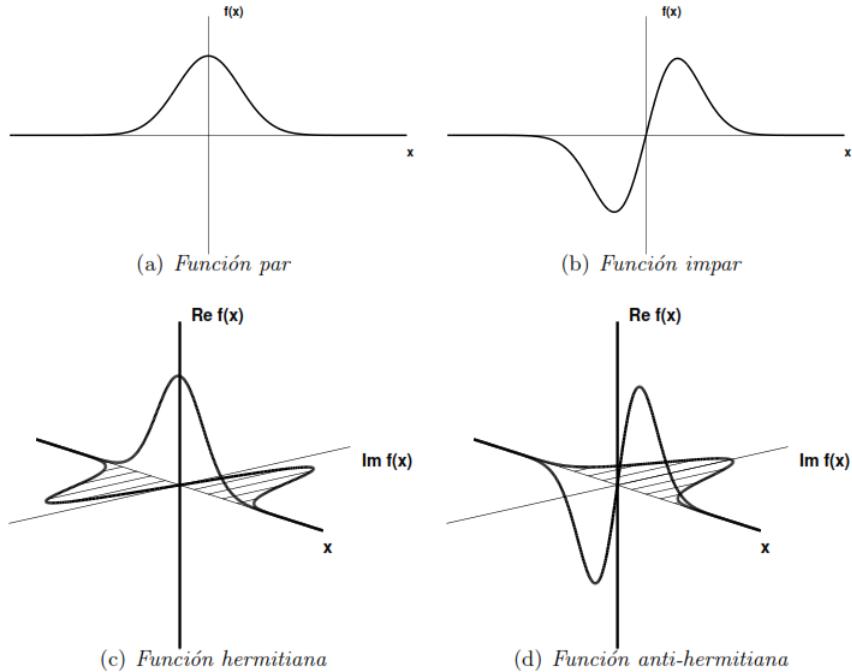


Figure 7: Funciones pares, impares, hermitiana y anti-hermitiana. Tomado de [2]

En el caso de las funciones complejas $y = f(x)$ con $x \in \mathbb{R}$ y $y \in \mathbb{C}$ con dominio real y codominio complejo por ejemplo, es necesario considerar la parte real e imaginaria de la salida de la función. Se dice entonces que una función compleja es **hermitiana** cuando su parte real es par y su parte imaginaria es

impar (apartado (c) de la Figura 7), y anti-hermitiana cuando su parte real es impar y su imaginaria par (apartado (d) de la Figura 7).

Cuando una función no cumple ninguna de las características anteriores, se dice que es una **función asimétrica**. Muchas veces el análisis de una función asimétrica se facilita si se descompone en partes par e impar, reescribiéndola de la siguiente forma:

$$f(x) = P(x) + I(x),$$

con $P(x)$ una función par y $I(x)$ una función impar. Para cualquier función $f(x)$ se puede escribir tal función par e impar, respectivamente, de la siguiente forma:

$$P(x) = \frac{1}{2}(f(x) + f(-x))$$

$$I(x) = \frac{1}{2}(f(x) - f(-x))$$

Tome por ejemplo una función real $f(x) = x^2 + x^3$, la cual se grafica en azul en la Figura 8. Observe que la función $f(-x) = x^2 - x^3$ equivale a *reflejar* la función $f(x)$ respecto al eje de las ordenadas, como muestra la gráfica roja de la Figura 8. Es por ello que en este caso, la función par está dada por:

$$P(x) = \frac{1}{2}(f(x) + f(-x)) = \frac{1}{2}(x^2 + x^3 + x^2 - x^3) = x^2,$$

lo cual corresponde a una parábola cóncava hacia arriba, una función par. Para el caso de la función impar, se tiene que:

$$I(x) = \frac{1}{2}(f(x) - f(-x)) = \frac{1}{2}(x^2 + x^3 - x^2 + x^3) = x^3,$$

por lo que entonces evidentemente $f(x) = P(x) + I(x) = x^2 + x^3$.

3.2 Seno y coseno

La función sinusoidal o seno es una función impar con período 2π (π corresponde a la razón entre la circunferencia y el diámetro de un círculo) la cual oscila suavemente hacia el infinito, y está dada en su forma más sencilla como:

$$f(x) = \sin(x),$$

y su gráfica se muestra en la Figura 9. Observe que la función se encuentra entre 1 y -1, lo que se refiere a la **amplitud del seno**, y efectivamente tiene un período de $2\pi \approx 6.2832$. Otras formas frecuentes de la función sinusoidal son $f_1(x) = \sin(\pi x)$ y $f_2(x) = \sin(2\pi x)$, las cuales se grafican en la Figura 10. Como se puede observar, el aumentar el coeficiente de la variable libre x acorta el período de repetición de la función sinusoidal T , el cual para la función f_1 corresponde a $T = 2$ y para la función f_2 a $T = 1$. Por esta razón es usual encontrar la forma de f_2 para la función seno, puesto que tanto su amplitud $a = 1$ y su período es $T = 1$.

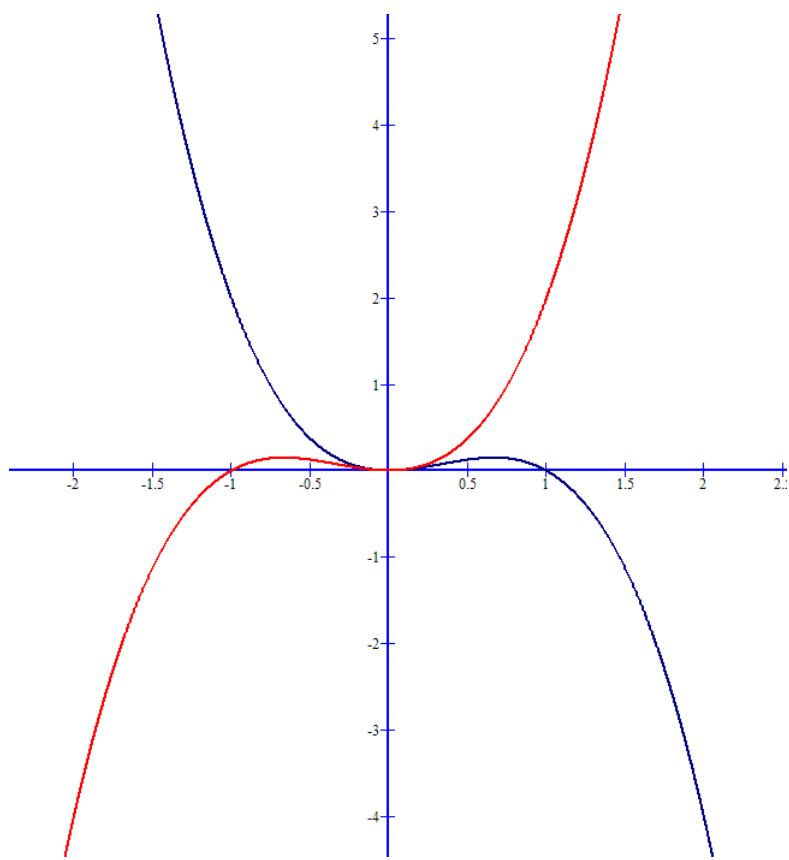


Figure 8: Funciones $f(x) = x^2 + x^3$ y $f(-x)$.

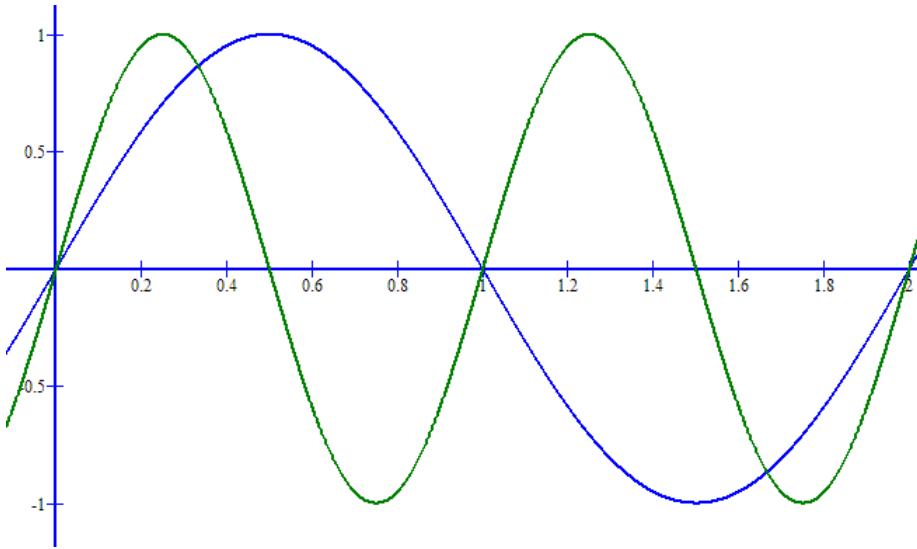


Figure 10: Funciones $f(x) = \sin(\pi x)$ (azul) y $f(x) = \sin(2\pi x)$ (verde).

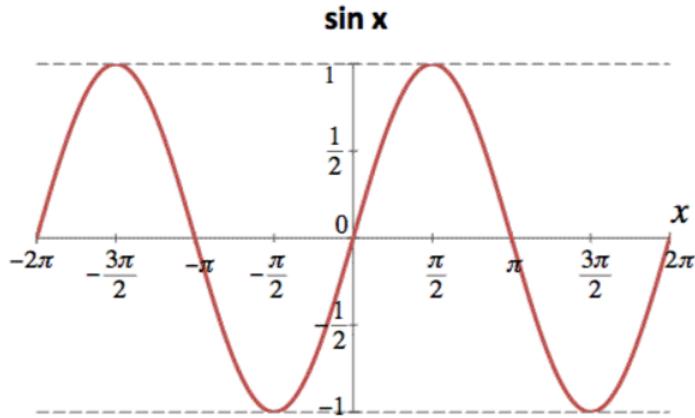


Figure 9: Función seno.

La forma general de la función seno viene dada por:

$$f(t) = a \sin(\omega f t + \phi),$$

con $\omega = 2\pi$ donde f define la frecuencia de oscilación de la función seno. Suponiendo que el codominio de la función está dado en segundos (s) observe a la función $f_3(t) = 2 \sin(\omega \frac{1}{2}t + \frac{\pi}{2})$, la cual presenta un período $T = 2s$, una amplitud $a = 2$ y un cambio de fase o desplazamiento $\phi = \frac{\pi}{2}$, el cual en la función equivale a un retraso de la onda por $\frac{\phi}{\omega f}$ (lo que en este caso

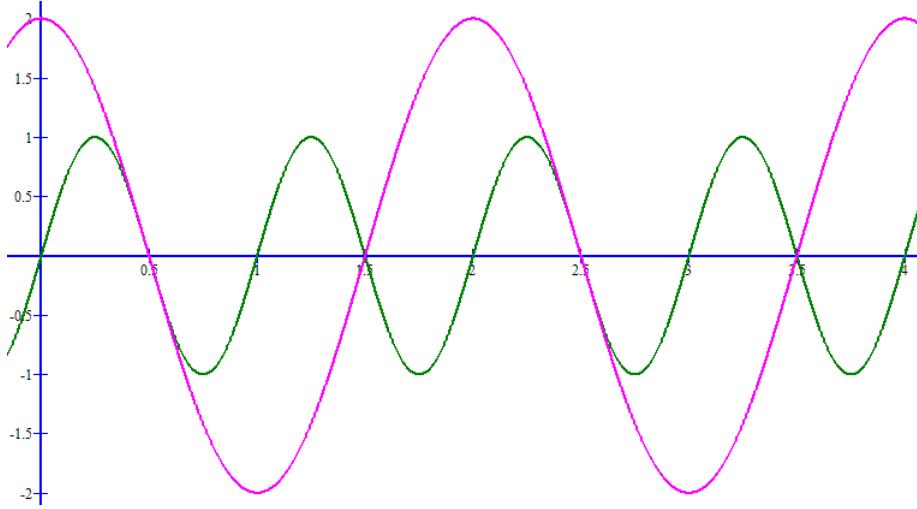


Figure 11: Graficación de las funciones f_2 (verde) y f_3 (magenta) con distinta amplitud, fase y frecuencia.

corresponde a un desplazamiento por $-\frac{1}{2}$), como se observa en la Figura 11. Se deduce además la relación inversa entre el período y la frecuencia:

$$f = \frac{1}{T}$$

En general, el **coseno** (función par) corresponde a una función sinusoidal con un desplazamiento por $-\frac{1}{2}$, por lo que entonces:

$$\sin\left(2\pi x + \frac{\pi}{2}\right) = \cos(2\pi x)$$

Las derivadas e integrales del seno y coseno están dadas como sigue:

$$\frac{d}{dx} \sin(\pi x) = \pi \cos(\pi x) \quad \int \sin(\pi x) dx = -\frac{1}{\pi} \cos(\pi x)$$

$$\frac{d}{dx} \cos(\pi x) = -\pi \sin(\pi x) \quad \int \cos(\pi x) dx = \frac{1}{\pi} \sin(\pi x)$$

Para el caso de la integral definida

$$\int_{-\infty}^{\infty} \sin(\pi x) dx$$

formalmente diverge. Sin embargo, si consideramos que físicamente la onda sinusoidal es finita (matemáticamente podemos pensar que la función puede

acotarse por alguna función par $P(x)$ de modo que $g(x) = P(x)$), y examinando que en un período la integral de la función es negativa y positiva en igual cantidad de intervalos, se puede considerar que:

$$\int_{-\infty}^{\infty} \sin(\pi x) dx \cong \int_{-\infty}^{\infty} \cos(\pi x) dx \cong 0,$$

un resultado útil como se verá más adelante.

3.3 Función exponencial compleja

La función exponencial compleja es una función con argumento real y salida compleja, definida como:

$$f(x) = e^{2\pi i x},$$

la cual se grafica en la Figura 12. Por la igualdad de Euler demostrada anteriormente, se tiene que la función exponencial compleja equivale a:

$$f(x) = e^{2\pi i x} = \cos(2\pi x) + i \sin(2\pi x).$$

La función exponencial compleja forma el espacio base para la transformada de Fourier, incluyendo en una sola función una función impar (senos) y otra par (cosenos). Observe en la Figura 12 que la magnitud se mantiene constante en 1 y la fase varía linealmente con x , en sentido contra las agujas del reloj.

3.4 Función impulso

La función impulso o también conocida como función *delta dirac* o *delta Kronecker* $\delta(x)$ no es estrictamente una función matemática, sino más bien una abstracción, que fija las siguientes propiedades para la función:

$$\delta(x) = \begin{cases} 0 & \text{si } x \neq 0 \\ \infty & x = 0 \end{cases}$$

y

$$\int_{-\infty}^{\infty} \delta(x) dx = 1.$$

La representación gráfica corresponde a una flecha en $x = 0$ representando el hecho de que en el eje y la función tiende al infinito, y el eje x la función tiende a cero, pero a pesar de ello, manteniendo un área bajo la curva de 1.

La necesidad de la abstracción de la función impulso deriva con lo usual de encontrar magnitudes físicas que concentran mucha energía en un instante de tiempo muy corto o espacio muy pequeño. La función impulso se puede aproximar a partir por ejemplo, de una función rectángulo (la cual se trata en la siguiente subsección):

$$\int_{-\infty}^{\infty} \delta(x) dx \cong \lim_{\tau \rightarrow 0} \frac{1}{\tau} \int_{-\infty}^{\infty} \Pi\left(\frac{x}{\tau}\right) dx.$$

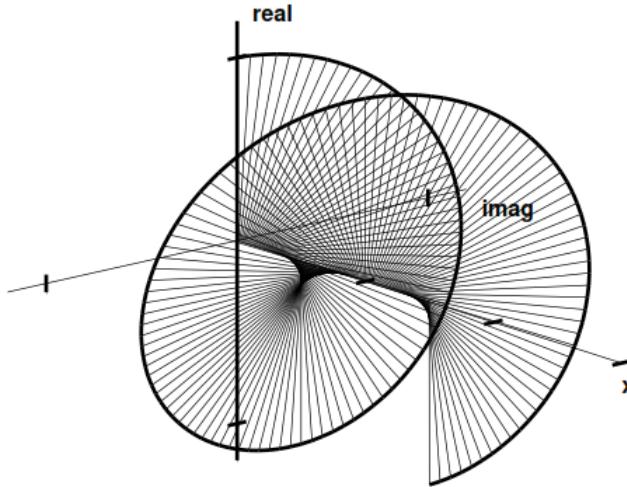


Figure 12: Función exponencial compleja. Tomado de [2].

Observe que el multiplicar la función rectangular por $\frac{1}{\tau}$ hace que la función rectangular, originalmente con amplitud unitaria, tenga una altura de $\frac{1}{\tau}$ y el dividir el argumento por τ hace que la base del rectángulo sea de magnitud τ . La Figura 16 muestra como a partir de la función rectangular se approxima al impulso, conforme $\tau \rightarrow \infty$. En realidad, la función impulso se puede aproximar con cualquier función simétrica con valor mayor que cero alrededor del origen, y decreciente conforme se aleja, como por ejemplo una función Gaussiana, ilustrado en la animación de la Figura 14.

3.4.1 Función impulso con dominio de dos dimensiones

La función impulso $\delta(x, y)$ definida en un dominio en \mathbb{R}^2 se define como:

$$\delta(x, y) = \begin{cases} 0 & \text{si } x \neq 0 \text{ o } y \neq 0 \\ \infty & x = y = 0 \end{cases}$$

$$y \quad \int_{-\infty}^{\infty} \delta(x, y) dx dy = 1.$$

y puede graficarse como se muestra en la Figura 15.

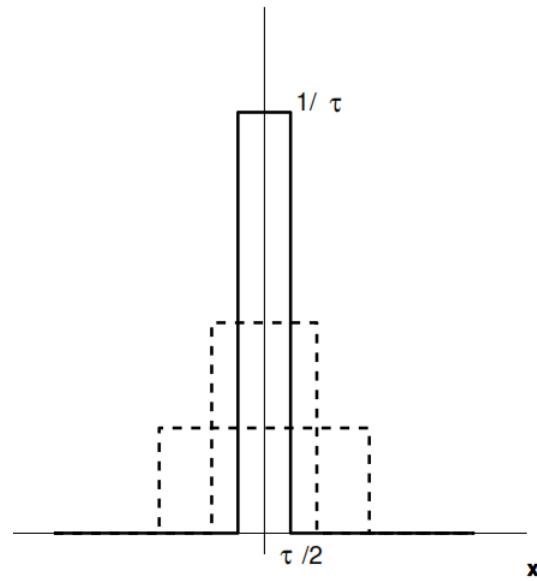


Figure 13: Aproximación con la función *rect* del impulso. Tomado de [2]

(Loading...)

Figure 14: Aproximación de la función impulso con una función Gaussiana.

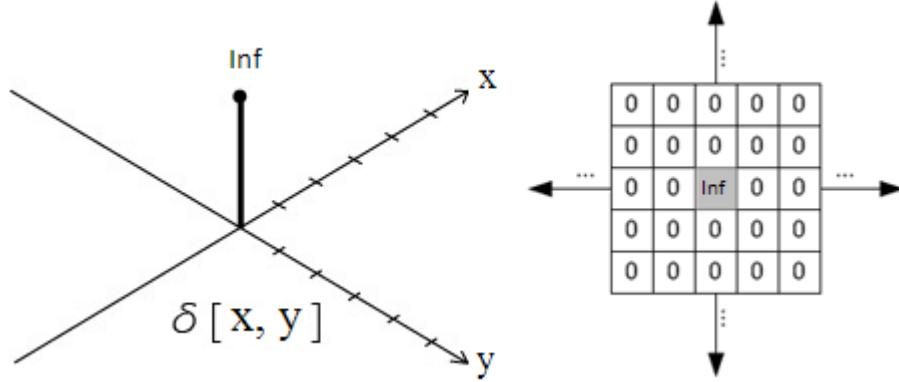


Figure 15: Graficación de la función impulso.

3.4.2 La propiedad del cedazo

Observe la siguiente integral:

$$\int_{-\infty}^{\infty} \delta(x) f(x) dx,$$

la cual, reemplazando la función *delta dirac* por su equivalente en términos de la función *rect* vendría dada por:

$$\frac{1}{\tau} \int_{-\infty}^{\infty} \prod \left(\frac{x}{\tau} \right) f(x) dx = \int_{-\frac{\tau}{2}}^{\frac{\tau}{2}} \frac{1}{\tau} f(x) dx,$$

y recordemos que $\tau \rightarrow 0$, por lo que entonces la integral es distinta de cero en el intervalo $-\frac{\tau}{2}$ a $\frac{\tau}{2}$ (según la definición de la función *rect*), con lo que entonces se tiene que:

$$\int_{-\infty}^{\infty} \delta(x) f(x) dx = f(0)$$

y en general, para un impulso centrado en a :

$$\int_{-\infty}^{\infty} \delta(x - a) f(x) dx = f(a)$$

y cuando la función f está desplazada por a :

$$\int_{-\infty}^{\infty} \delta(x) f(x - a) dx = f(-a).$$

A esta propiedad se le llama la propiedad del cedazo, pues extrae un valor puntual de la función f mediante la función impulso.

3.5 Función rectangulo

La función conocida como rectangular o rect, la cual se muestra en la Figura 16 :

$$\Pi(x) = \begin{cases} 1 & \text{si } |x| \leq \frac{1}{2} \\ 0 & \text{sino} \end{cases}$$

Ello significa entonces que

$$\int_{-\infty}^{\infty} \Pi(x) dx = 1$$

En los puntos $x = -\frac{1}{2}$ y $x = \frac{1}{2}$ se puede afirmar que existe una discontinuidad, por lo que la derivada no está definida. Sin embargo, se puede suponer una transición más suave en tales puntos, como la línea punteada dibujada en la Figura 16, función que etiquetaremos como $\Pi_\alpha(x)$, de modo que la derivada en la práctica tienda a ∞ . La constante α define la *suavidad* para la aproximación de la función rectangular.

De este modo la derivada de la función *rect* en los puntos $x = -\frac{1}{2}$ y $x = \frac{1}{2}$ se puede aproximar como:

$$\frac{d}{dx} \Pi(x) = \delta\left(x + \frac{1}{2}\right) - \delta\left(x - \frac{1}{2}\right).$$

Una modificación usual que conserva el valor de la integral unitario es la siguiente:

$$\frac{1}{\tau} \Pi\left(\frac{x}{\tau}\right) = \begin{cases} \frac{1}{\tau} & \text{si } |x| \leq \frac{\tau}{2} \\ 0 & \text{sino} \end{cases},$$

de esta forma, como se vió en la Figura 13, conforme $\tau \rightarrow 0$, se tiene que $\frac{1}{\tau} \Pi\left(\frac{x}{\tau}\right) \rightarrow \delta(x)$.

3.5.1 Función rectangulo con dominio de dos dimensiones

La función *rect* $\Pi(x, y)$ en dos dimensiones se define como:

$$\Pi(x, y) = \begin{cases} 1 & \text{si } |x| \leq \frac{1}{2} \text{ y } |y| \leq \frac{1}{2} \\ 0 & \text{sino} \end{cases}.$$

La Figura 17 muestra la graficación de la función *rect* desde dos perspectivas (lateral y desde arriba, en una imagen).

3.6 Función escalón o de Heaviside

La función escalón o también conocida como de Heaviside $\Gamma(x)$ se define como:

$$\Gamma(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x < 0 \end{cases}$$

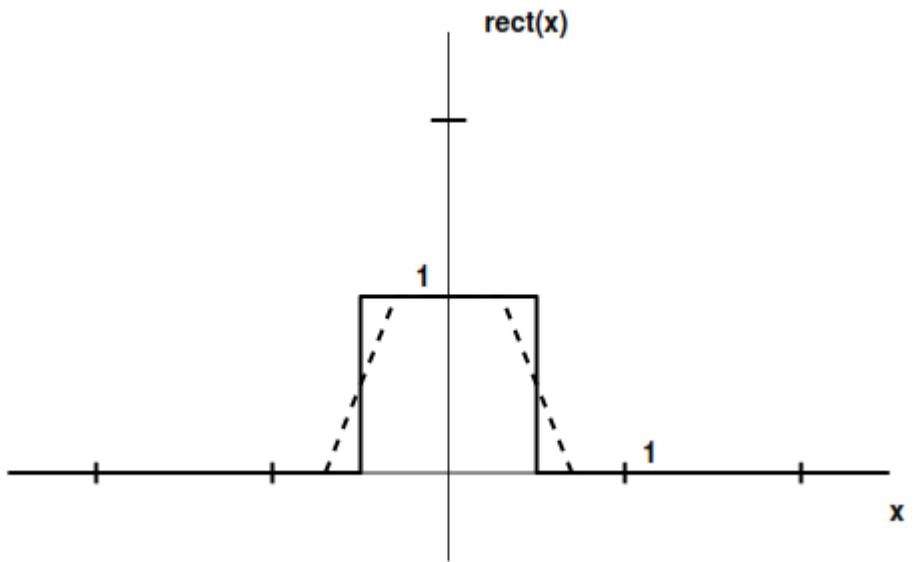


Figure 16: Función rect $\Pi(x)$ y su aproximación $\Pi_\alpha(x)$ en línea punteada.
Tomado de [2]

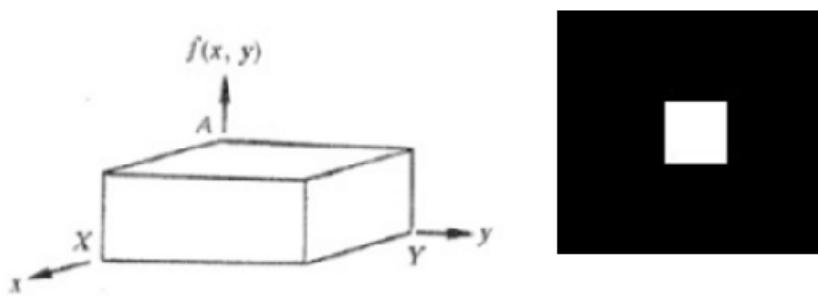


Figure 17: Función *rect* con dominio en dos dimensiones.

Recordando la definición de la derivada para la función *rect*, para el caso de la función de Heaviside, se define su derivada como:

$$\frac{d}{dx} \Gamma(x) = \delta(x)$$

Para hacer que la función esté centrada en a , como es usual, se hace $\Gamma(x - a)$, con lo que:

$$\Gamma(x) = \begin{cases} 1 & \text{si } x > a \\ 0 & \text{si } x < a \end{cases}$$

3.7 Función signo

La función signo, representada como $\text{sgn}(x)$, está definida como sigue:

$$\text{sgn}(x) = \begin{cases} 1 & \text{si } x > 0 \\ -1 & \text{si } x < 0 \end{cases}$$

Para la cual la derivada está dada por:

$$\frac{d}{dx} \text{sgn}(x) = \delta(x)$$

Y para hacer que la función esté centrada en a , como es usual, se hace $\Gamma(x - a)$, con lo que:

$$\Gamma(x - a) = \begin{cases} 1 & \text{si } x > a \\ 0 & \text{si } x < a \end{cases}$$

La Figura 18 muestra las gráficas de las funciones signo y Heaviside.

3.8 Función sinc o seno cardenal

La función *sinc* o conocida también como seno cardenal se define como sigue:

$$\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x},$$

y cruza por cero cuando $|x| = 1, 2, 3, \dots$, su valor cuando $x = 0$ es uno (lo cual se demuestra más abajo, por L'Hopital), y el área total bajo la curva es uno también:

$$\text{sinc}(0) = \lim_{x \rightarrow 0} \frac{\sin(\pi x)}{\pi x} = \lim_{x \rightarrow 0} \frac{\cos(\pi x)}{\pi} = 1,$$

$$\int_{-\infty}^{\infty} \text{sinc}(x) dx = 1.$$

Todo ello se ilustra en la Figura 19. Las ondas con cruces por cero se les llama lóbulos, y tienen magnitudes decrecientes (el lóbulo central tiene magnitud 1, el segundo 0,21, y el tercero 0,13 aproximadamente).

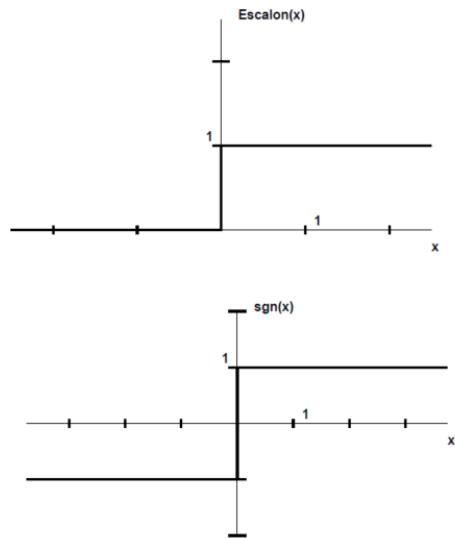


Figure 18: Funciones signo y de Heaviside. Tomado de [2]

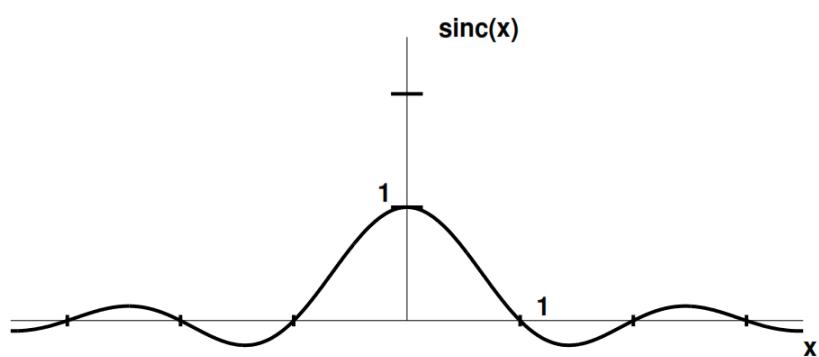


Figure 19: Función de seno cardenal.

4 Convolución

La convolución es una operación muy común en el procesado de señales. Implementa el concepto de “ventana deslizante” en el que básicamente una función se desliza sobre otra, y se calcula la respuesta de multiplicarlas en cada punto del “deslizamiento”. Formalmente, la integral de convolución $h(x)$ de dos funciones $f(x)$ y $g(x)$ se define como:

$$h(x) = f(x) * g(x) = \{f * g\}(x) = \int_{-\infty}^{\infty} f(\tau) g(x - \tau) d\tau$$

Observe que el término $g(x - \tau)$ corresponde a una traslación de la función $g(x)$ por τ unidades, con τ asociado a la variable de integración que va de $-\infty$ a ∞ .

Como ejemplo, considere la función $\text{rect}[\cdot](x)$, convolucionada consigo misma:

$$h(x) = \text{rect}(x) * \text{rect}(x) = \int_{-\infty}^{\infty} \text{rect}(\tau) \text{rect}(x - \tau) d\tau.$$

Observemos el proceso de convolución ilustrado gráficamente en la Figura 20. Tome el ejemplo de evaluar la convolución en $x = -500$ (un valor muy negativo, tendiendo a $-\infty$)

$$h(-500) = \int_{-\infty}^{\infty} \text{rect}(\tau) \text{rect}(-500 - \tau) d\tau = \int_{-\infty}^{\infty} \text{rect}(\tau) \text{rect}(-(500 + \tau)) d\tau,$$

con la equivalencia anterior es fácil ver que la función $\text{rect}(-(500 + \tau))$ tiene su origen en $\tau = -500$, por lo que está centrada en tal punto, y además, como su argumento está negado, la función está reflejada respecto a su origen, sin embargo al ser una función simétrica, ello no tiene ningún efecto.

La integral de menos infinito a infinito, suma la multiplicación de todos los puntos, con lo que en este caso, dado que no existe intersección entre ambas funciones, la integral resulta en cero. Lo anterior se ilustra en la primer sección de arriba a abajo de la Figura 20. Posteriormente, conforme x se acerca a cero, la integral de la intersección entre ambos rectángulos comienza a aumentar, lo que hace que la función $h(x)$ comience a crecer y alcanza su máximo cuando $h(0)$ (correspondiente a $x = 0$), pues en tal caso se tiene la siguiente integral:

$$h(0) = \int_{-\infty}^{\infty} \text{rect}(\tau) \text{rect}(-\tau) d\tau = \int_{-\infty}^{\infty} \text{rect}(\tau) \text{rect}(\tau) d\tau = \int_{-\frac{1}{2}}^{\frac{1}{2}} 1 d\tau = 1,$$

lo cual implica que la función de convolución tiene su máximo en $x = 0$, puesto que conforme $x \rightarrow \infty$, la integral decrece hasta ser nula cuando no existe intersección entre ambos rectángulos, lo cual lo ilustra también la Figura 20.

La Figura 21 muestra una animación del proceso de convolucionar dos rectángulos pero de distinta base, lo que resulta en un trapecio.

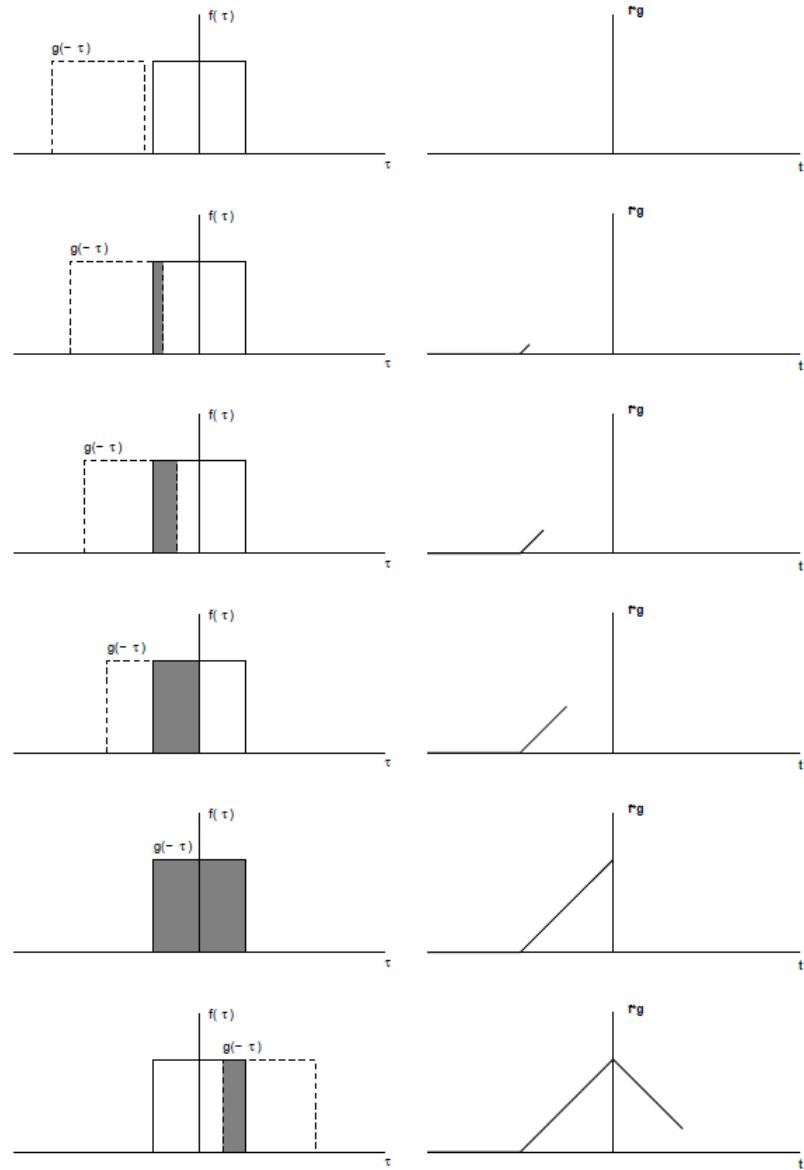


Figure 20: Distintos puntos de evaluación para la convolución. Tomado de [2]

(Loading...)

Figure 21: Convolución de dos rectángulos de distinta base.

Otro ejemplo de la convolución es la convolución de una función cualquiera f con la función impulso:

$$h(x) = f(x) * \delta(x) = \int_{-\infty}^{\infty} f(\tau) \delta(x - \tau) d\tau.$$

Recuerde que para un punto cualquiera $x = a$, se tiene que $f(\tau) \delta(a - \tau) = f(a)$ por la propiedad del cedazo, por lo que entonces, la convolución anterior corresponde a aplicar la propiedad del cedazo para todos los puntos en x , por lo que se tiene que:

$$h(x) = f(x) * \delta(x) = f(x).$$

Observe además que tomando la convolución:

$$h(x) = \delta(x) * f(x) = \int_{-\infty}^{\infty} \delta(\tau) f(x - \tau) d\tau = f(x),$$

lo cual sugiere la comutatividad de la convolución.

La convolución discreta para dos funciones discretas $f[n]$ y $g[n]$ se define como:

$$h[n] = f[n] * g[n] = \sum_{m=-\infty}^{\infty} f[m] g[n - m]$$

Ejemplo 1

Sean las funciones $f[n] = (10 \quad 10 \quad 10)$ y $g[n] = (\frac{1}{2} \quad \frac{1}{2})$, su convolución viene dada entonces por:

$$h[n] = (10 \quad 10 \quad 10) * (\frac{1}{2} \quad \frac{1}{2}).$$

Tomemos en cuenta que $h[n]$ se encuentra definida para $n \geq 0$ y lo mismo sucede para las funciones $f[n]$ y $g[n]$. Si no lo está, se asume que su valor es

cero. Esto implica que, por ejemplo, para obtener el valor de $h[0]$ (usando a cero como el índice inicial) se tiene que:

$$h[0] = \sum_{m=0}^0 f[m]g[0-m] = f[0] \cdot g[0] = 10 \cdot \frac{1}{2} = 5$$

y para otros casos, de forma similar se tiene que:

$$\begin{aligned} h[1] &= \sum_{m=0}^1 f[m]g[1-m] = f[0] \cdot g[1] + f[1] \cdot g[0] = 10 \cdot \frac{1}{2} + 10 \cdot \frac{1}{2} = 10 \\ h[2] &= \sum_{m=0}^2 f[m]g[2-m] = f[1] \cdot g[1] + f[2] \cdot g[0] = 10 \cdot \frac{1}{2} + 10 \cdot \frac{1}{2} = 10 \\ h[3] &= \sum_{m=0}^3 f[m]g[3-m] = f[2] \cdot g[1] = 10 \cdot \frac{1}{2} = 5 \end{aligned}$$

Lo cual si se grafica, tiene forma similar a lo mostrado en la Figura 20.

Ejemplo 2

Sean las funciones $f[n] = (1 \ 2 \ 4 \ 3 \ 2)$ y $g[n] = (1 \ 2 \ 2)$, su convolución viene dada entonces por:

$$h[n] = (1 \ 2 \ 4 \ 3 \ 2) * (1 \ 2 \ 2) = (1 \ 4 \ 10 \ 15 \ 16 \ 10 \ 4).$$

Para entender como se obtuvo tal resultado, tomemos en cuenta que $h[n]$ se encuentra definida para $n \geq 0$ y lo mismo sucede para las funciones $f[n]$ y $g[n]$. Esto implica que, por ejemplo, para obtener el valor de $h[0]$ (usando a cero como el índice inicial) se tiene que:

$$h[0] = \sum_{m=0}^0 f[m]g[0-m] = f[0] \cdot g[0] = 1 \cdot 1 = 1$$

y para otros casos, de forma similar se tiene que:

$$\begin{aligned} h[1] &= \sum_{m=0}^1 f[m]g[1-m] = f[0] \cdot g[1] + f[1] \cdot g[0] = 1 \cdot 2 + 2 \cdot 1 = 4 \\ h[2] &= \sum_{m=0}^2 f[m]g[2-m] = f[0] \cdot g[2] + f[1] \cdot g[1] + f[2] \cdot g[0] = 1 \cdot 2 + 2 \cdot 2 + 4 \cdot 1 = 10 \end{aligned}$$

4.1 Correlación cruzada

La correlación cruzada es una **medida de similitud** entre las funciones reales f y g y viene dada por:

$$c(x) = f(x) \star g(x) = \{f \star g\}(x) = \int_{-\infty}^{\infty} f(\tau)g(x+\tau)d\tau$$

y en su forma discreta viene dada por:

$$c[n] = f[n] \star g[n] = \sum_{m=-\infty}^{\infty} f[m]g[n+m]$$

y equivale a realizar la convolución sin reflejar o rotar la función g . Observe que si la función g es simétrica la correlación y la convolución son equivalentes:

$$f(x) \star g(x) = f(x) * g(x)$$

4.2 Autocorrelación

La autocorrelación de una función f equivale a la correlación de la función f consigo misma:

$$c(x) = f(x) \star f(x) = \int_{-\infty}^{\infty} f(\tau) f(x + \tau) d\tau$$

y en forma discreta:

$$c[n] = f[n] \star f[n] = \sum_{m=-\infty}^{\infty} f[m] f[n+m]$$

la autocorrelación es utilizada en el procesamiento de señales para averiguar si una función f es periódica o no, pues si c tiene dos o más máximos de igual magnitud, la señal fue equivalente dos o más veces a sí misma, con un desplazamiento.

4.3 Convolución con dominio \mathbb{R}^2

La convolución en dos ejes se define para dos funciones continuas $g(x, y)$ y $f(x, y)$ como:

$$(g * f)(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x-u, y-v) f(u, v) du dv$$

y para dos funciones discretas $g[x, y]$ y $f[x, y]$ como:

$$(g * f)[x, y] = \sum_{u=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} g[x-u, y-v] f[u, v] \quad (1)$$

En el presente curso, será usual utilizar matrices digitales (muestras finitas de valores discretos), por lo que los funcionales con dominio definido en \mathbb{R}^2 y codominio en \mathbb{R} serán representados como tales matrices digitales: $U \in \mathbb{R}^{N \times M}$ y $F \in \mathbb{R}^{K \times L}$, con $F, G : \mathbb{R}^2 \rightarrow \mathbb{R}$ y respecto a las dimensiones $K < N$ y $L < M$. Para usar matrices con origen en 0,0, se utiliza la siguiente definición y una notación que prescinda de rotaciones del núcleo F (lo cuál lo hace similar a la correlación cruzada):

$$(U * F)[x, y] = \sum_{u=0}^{K-1} \sum_{v=0}^{L-1} U \left[x - \frac{(K-1)}{2} + u, y - \frac{(L-1)}{2} + v \right] F[u, v]$$

Ejemplo 1:

Sean las matrices

$$U = \begin{bmatrix} 2 & 3 & 4 & 1 & 5 \\ 2 & 9 & 20 & 1 & 5 \\ 1 & 5 & 3 & 7 & 8 \\ 1 & 2 & 3 & 7 & 5 \\ 5 & 7 & 8 & 9 & 10 \end{bmatrix} \quad F = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

La respuesta de la convolución

$$(U * F) [2, 2] = \sum_{u=0}^2 \sum_{v=0}^2 U [2 - 1 + u, y - 1 + v] F [u, v]$$

$$(U * F) [2, 2] = \begin{matrix} U [1, 1] F [0, 0] + U [1, 2] F [0, 1] + U [1, 3] F [0, 2] \\ U [2, 1] F [1, 0] + U [2, 2] F [1, 1] + U [2, 3] F [1, 2] \\ U [3, 1] F [2, 0] + U [3, 2] F [2, 1] + U [3, 3] F [2, 2] \end{matrix}$$

$$(U * F) [2, 2] = \frac{9 \cdot -1 + 20 \cdot 0 + 1 \cdot 1}{2 \cdot -1 + 3 \cdot 0 + 7 \cdot 1} = -1$$

El resultado es realizar el producto punto de la sub-matriz $U_{1:3,1:3}$ con el filtro F . Observe que por ejemplo, la respuesta de la convolución en $(U * F) [0, 0]$ no es posible de calcular, lo cual genera un efecto indeseable de cambio de la dimensionalidad en la matriz resultante de la convolución. Es por esto que es frecuente concatenar filas y columnas llenas de ceros (0's) en los bordes de la matriz con mayores dimensiones U , lo cual se refiere como *padding*. La cantidad de filas con ceros a concatenar se representa con Δf y la cantidad de columnas con Δc , por lo que entonces la concatenación de ceros en filas y columnas está dada por la operación P:

$$p(U) = U' \in \mathbb{R}^{(N+2z_f) \times (M+2z_c)}$$

, por lo que para preservar la dimensionalidad de la matriz resultante en la convolución, se hace que: $z_c = \frac{(K-1)}{2}$ y $z_f = \frac{(L-1)}{2}$.

La definición de la convolución según la Ecuación 3 se ilustra en la Figura 22, donde se observa como la función con menor dimensionalidad básicamente se *desliza* sobre la función con mayor dimensionalidad (para ello, por conveniencia, se fija como el segundo operando). Por cada *deslizamiento* de la función G sobre F , se calcula el **producto punto entre ambas matrices**, centrado en el elemento actual (x, y) de tal deslizamiento:

$$(U * F) [x, y] = \alpha U_{x - \frac{(K-1)}{2} : x + \frac{(K-1)}{2}, y - \frac{(L-1)}{2} : y + \frac{(L-1)}{2}} \cdot F, \quad (2)$$

donde el **producto punto se puede definir para dos matrices** $A, B \in \mathbb{R}^{N \times M}$ como:

$$A \cdot B = \sum_{i=0}^N \sum_{j=0}^M A_{i,j} B_{i,j}$$

con $s \in \mathbb{R}$.

- **Ejemplo:** Sean las matrices o funciones F y $U, F, U : \mathbb{R}^2 \rightarrow \mathbb{R}$ dadas por:

$$U = \begin{bmatrix} 2 & 3 & 2 \\ 3 & 25 & 3 \\ 2 & 3 & 2 \end{bmatrix} \quad F = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

calcule la **convolución normalizada**. Realice el *padding* necesario para conservar la dimensionalidad de la señal. En este caso se tiene que $K = L = 3$, $\alpha = \frac{1}{9}$. A la señal U se le realiza una concatenación de una fila y columna de ceros, de forma que las columnas negativas en una unidad sean cero, lo que resulta en:

$$U' = p(U) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 3 & 2 & 0 \\ 0 & 3 & 25 & 3 & 0 \\ 0 & 2 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

por lo que entonces se tiene que:

$$\begin{aligned} (U * F)[0,0] &= (U' * F)[1,1] = \frac{1}{9} \sum_{u=0}^2 \sum_{v=0}^2 U'[1-1+u, 1-1+v] F[u, v] = U'_{0:2,0:2} \cdot F \\ &= \frac{1}{9} U'[0,0] F[0,0] + U'[0,1] F[0,1] + U[0,2] F[0,2] \\ &\quad U'[1,0] F[1,0] + U'[1,1] F[1,1] + U'[1,2] F[1,2] \\ &\quad U'[2,0] F[2,0] + U'[2,1] F[2,1] + U'[2,2] F[2,2] = 3.6667 \end{aligned}$$

La notación de producto punto es más clara, veamos como se calcula la respuesta en $[0,1]$:

$$(U * F)[0,1] = \alpha U_{0-1:0+1,1-1:1+1} \cdot F = \frac{1}{9} U_{-1:1,0:2} \cdot F = \frac{1}{9} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 3 & 2 \\ 3 & 25 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = 4.2222$$

Realizando todos los cálculos para todas las entradas, se tiene que:

$$(U * F) = \begin{bmatrix} 3.66 & 4.22 & 3.66 \\ 4.22 & 5 & 4.22 \\ 3.66 & 4.22 & 3.66 \end{bmatrix}$$

En MATLAB se puede usar la función *conv2*, la cual, por defecto, hace el padding para las dos señales, conservando la dimensionalidad a la salida.

```
F = [1/9 1/9 1/9; 1/9 1/9 1/9; 1/9 1/9 1/9];
U = [2 3 2; 3 25 3; 2 3 2]
C = conv2(U, F);
```

Muchas veces, con el afán de ahorrar recursos computacionales, el deslizamiento de la función U sobre F se hace con un *paso* mayor (en inglés conocido como *stride*), de forma que el deslizamiento se hace a *saltos*. La cantidad de filas y columnas que se salta en el deslizamiento se define por Δ , la convolución se puede reescribir como:

$$(U * F)[x,y] = \sum_{u=0}^{K-1} \sum_{v=0}^{L-1} U \left[x\Delta - u + \frac{(K-1)}{2}, y\Delta - v + \frac{(L-1)}{2} \right] F[u,v]. \quad (3)$$

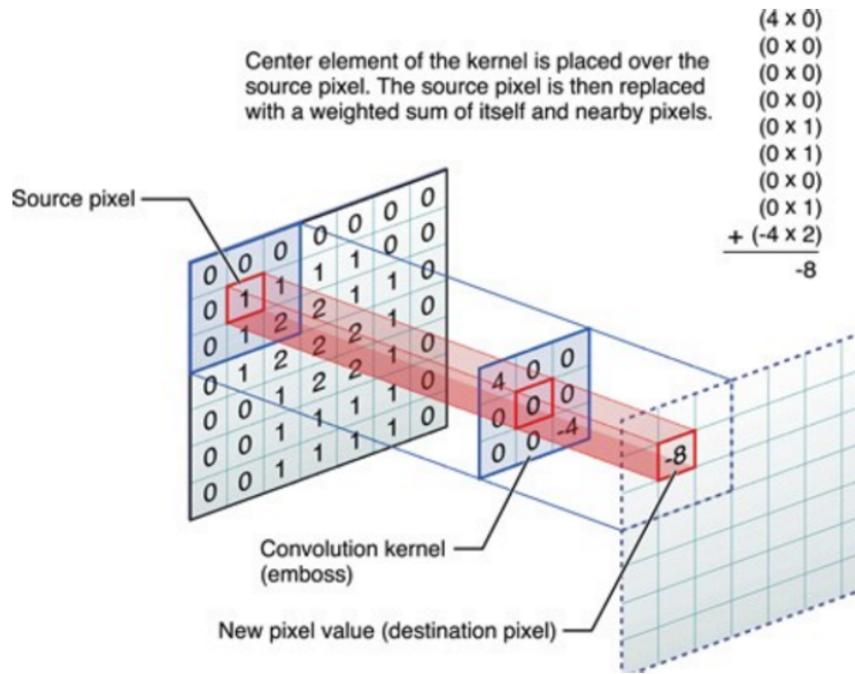


Figure 22: Deslizamiento en la convolución. Tomado de *Deep evolution of image representations for handwritten digit recognition*

lo cual reduce la dimensionalidad de la matriz resultante de la convolución, submuestreando el resultado original.

En las páginas https://github.com/vdumoulin/conv_arithmetic/blob/master/README.md y <http://cs231n.github.io/convolutional-networks/> se muestra la operación de la convolución animada.

En general, para determinar la dimensionalidad de la imagen convolucionada o la salida $Y \in \mathbb{R}^{A \times B}$, con una matriz de entrada $U \in \mathbb{R}^{N \times M}$ y un filtro $F \in \mathbb{R}^{K \times L}$, con un paso Δ y un padding z , se puede utilizar la siguiente formulación:

$$A = \frac{N - K + 2z}{\Delta} + 1, \quad B = \frac{M - L + 2z}{\Delta} + 1$$

4.4 Propiedades de la convolución

Las siguientes son propiedades generales de la convolución y convoluciones conocidas:

- Conmutatividad $f(x) * g(x) = g(x) * h(x)$.
- Homogeneidad $f(x) * a_1 g(x) = a_1 (h(x) * g(x))$ con $a_1 \in \mathbb{R}$.
- Distributividad $g(x) * (h_1(x) + h_2(x)) = g(x) * h_1(x) + g(x) * h_2(x)$.
- Asociatividad $g(x) * (h_1(x) * h_2(x)) = (g(x) * h_1(x)) * h_2(x)$.
- Amplificación $g(x) * a_1 \delta(x) = a_1 g(x)$ con $a_1 \in \mathbb{R}$.
- Desplazamiento $g(x) * \delta(x+s) = g(x+s)$
- Eco: $g(x) * (\delta(x) + k\delta(n-s)) = g(x) + k g(x-s)$.

4.5 Algunas convoluciones importantes: Aproximación discreta de la primer derivada

Tome la siguiente convolución discreta:

$$h[n] = g[n] * f[n],$$

con $g[n] = (1 \ -1)$ y $f[n]$ cualquier función, examinemos por ejemplo los valores $h[2]$ y $h[3]$:

$$h[2] = \sum_{m=1}^2 f[m] g[2-m] = f[1]g[1] + f[2]g[0] = f[2] - f[1]$$

$$h[3] = \sum_{m=2}^3 f[m] g[3-m] = f[2]g[1] + f[3]g[0] = f[3] - f[2]$$

Generalizando lo anterior, se concluye que:

$$d_-[n] = f[n] - f[n-1] \quad (4)$$

y haciendo el reemplazo de $n = u + 1$ se obtiene lo que llamaremos la aproximación por diferencias finitas $d[u]$:

$$d_+[u] = f[u+1] - f[u] \quad (5)$$

Veamos ahora la definición formal de la derivada de una función continua $f(x)$, la cual mide la tasa de cambio del funcional $f(x)$ en un punto x (tomando en cuenta la dirección del mismo):

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}.$$

Aproximación de la derivada por diferencias finitas: basados en la derivada según la definición continua, para el dominio discreto, hacemos que h tenga un

valor discreto menor posible, el cual corresponde a $h = 1$, por lo que entonces se tiene que, a partir de la definición formal de la primera derivada:

$$f' [n] \cong d_+ [n] = \frac{f [n + h] - f [n]}{h} = f [n + 1] - f [n].$$

La fórmula anterior se denomina **aproximación por diferencias hacia adelante**, pues involucra un desplazamiento hacia adelante de la función y coincide con lo arribado en la ecuación 5. La aproximación por **diferencias hacia atrás** se define entonces como:

$$f' [n] \cong d_- [n] = \frac{f [n] - f [n - h]}{h} = f [n] - f [n - 1],$$

lo cual coindice con lo planteado en la ecuación 4. Una mejor aproximación es la de las **diferencias centrales**, con un error menor y menor sensibilidad al ruido, la cual promedia las aproximaciones hacia adelante y atrás:

$$f' [n] \cong \frac{1}{2} \left(\frac{f [n + h] - f [n]}{h} + \frac{f [n] - f [n - h]}{h} \right) = \frac{f [n + h] - f [n - h]}{2h} = \frac{f [n + 1] - f [n - 1]}{2} \quad (6)$$

La última igualdad, considerando a $h = 1$. Observe entonces que la convolución de una función $f [n]$ con una **máscara o núcleo** (*kernel* en inglés) $d_+ [n] = (1 \ -1)$ resulta en una aproximación de la primer derivada por diferencias hacia adelante. Para realizar una aproximación por diferencias centrales, según lo fijado en la ecuación 6 se usa la máscara:

$$d_c [n] = \left(\frac{1}{2} \ 0 \ -\frac{1}{2} \right).$$

Es usual la necesidad de amplificar la magnitud de la aproximación de la primer derivada en muchos casos, por lo que por ejemplo, por la propiedad de la homogeneidad de la convolución podemos crear una máscara nueva $d_c^* [n] = \alpha d_c [n]$, por lo que con $\alpha = 2$, se obtiene la máscara:

$$d_c^* [n] = (1 \ 0 \ -1).$$

Por ello entonces, la primera derivada $f' [n]$ en términos de la convolución se puede aproximar como:

$$f' [n] \cong f [n] * d_c^* [n]$$

4.6 Aproximación discreta de la segunda derivada

La segunda derivada de una función f corresponde a la derivada de la primer derivada $f' (x)$, lo que conceptualmente se refiere entonces a la *aceleración* con que cambian los valores del funcional f respecto a x . Es por ello que entonces la segunda derivada corresponde a:

$$f'' (x) = \lim_{h \rightarrow 0} \frac{f' (x + h) - f' (x)}{h} = \lim_{h \rightarrow 0} \frac{\lim_{h \rightarrow 0} \frac{f(x+2h)-f(x+h)}{h} - \lim_{h \rightarrow 0} \frac{f(x+h)-f(x)}{h}}{h}$$

$$\Rightarrow f''(x) = \lim_{h \rightarrow 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2},$$

con lo cual tomando $h = 1$ se tiene que, en términos discretos, se puede aproximar la segunda derivada como sigue:

$$\Rightarrow f''(x) \cong f(x+2) - 2f(x+1) + f(x),$$

y su máscara correspondiente se define entonces como:

$$s[n] = (1 \quad -2 \quad 1)$$

y de manera similar se tiene entonces que, en términos de una convolución, la segunda derivada se define como:

$$f''[n] \cong f[n] * s[n] \quad (7)$$

5 Filtros lineales

5.1 Filtros paso bajos o de eliminación de ruido

Como se observó anteriormente para el caso de la función rect $\Pi(x, y)$, su convolución con cualquier otra función $f(x, y)$ tiende a resultar en un *suavizado* de $f(x, y)$, y en tal caso visto anteriormente, la convolución $\Pi(x, y) * \Pi(x, y) = \Delta(x)$ resultó en la función triángulo, la cual puede conceptualizarse como una aproximación más suave de la función rect . La función rect con dominio en dos dimensiones, puede representarse una matriz de 1's $\Pi \in \mathbb{R}^{N \times M}$ en todas sus entradas:

$$\Pi = \alpha \begin{bmatrix} 1 & \dots & 1_{1,M} \\ \vdots & \ddots & 1 \\ 1_{N,1} & \dots & 1_{N,M} \end{bmatrix}$$

con $\alpha = \frac{1}{NM}$ el coeficiente de normalización que se asegura de preservar la magnitud de la señal original, por lo que también se llama el filtro de promediado.

Visto en el dominio de Fourier (transformación $\mathcal{F}(U)$) que descompone la señal U en una sumatoria de senos y cosenos complejos), los cambios abruptos o fuertes en una señal necesitan de componentes de alta frecuencia con alta amplitud, para aproximar la función original, por lo que entonces se puede afirmar que los bordes o cambios de alta magnitud y abruptos se asocian con la energía en las altas frecuencias. La energía entonces en las bajas frecuencias, se asocian a cambios suaves en la señal. Por ejemplo, para una imagen, los bordes y el ruido en general tienden a generar energía en las altas frecuencias, e información como la textura genera la energía en intervalos de frecuencia más bajos. Es por esto que convolucionar una señal U con un Π , $U * \Pi$, resulta

en una atenuación fuerte de la energía en las altas frecuencias, y menor en las bajas frecuencias, por lo que el filtro de promediado es un tipo de filtro que deja pasar la energía en las bajas frecuencias, o *paso bajos*.

Un filtro más utilizado aún para los mismos efectos, es el filtro Gaussiano $\mathcal{N}(x, y)$, el cual tiene la agradable propiedad de no tener discontinuidades y ser una función *suave*, distinto al filtro basado en la función *rect*, cuyas discontinuidades hacen que sea necesario una cantidad infinita de senos y cosenos para representarlo, puesto que:

$$\mathcal{F}(\Pi(x)) = \text{sinc}(x)$$

lo cual introduce elementos de frecuencia no deseados. En cambio, para el caso de la función Gaussiana, su transformada es también una función Gaussiana $\mathcal{F}(\mathcal{N}(x, y)) = \mathcal{N}(x, y)$, por lo que truncarla para que tenga un dominio finito, provoca menos inexactitud, además de incluir menos energía de componentes no deseados.

Usualmente la función Gaussiana implementada para el filtrado es *isotrópica*, es decir, es equivalente sin importar la dirección desde la que se le examine. Es por esto que entonces la función viene dada entonces por:

$$F(x, y) = \mathcal{N}(x, y) = \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right)$$

Por ejemplo, utilizando la función simétrica en ambas dimensiones, con desviación estándar $\sigma_x = \sigma_y = 1$ y dimensiones $\mathcal{N} \in \mathbb{R}^{3 \times 3}$

`fspecial('gaussian', 3, 1)`

se obtiene la máscara (con el coeficiente de normalizado incluido):

$$\mathcal{N}[x, y] = \begin{bmatrix} 0.0751 & 0.1238 & 0.0751 \\ 0.1238 & 0.2042 & 0.1238 \\ 0.0751 & 0.1238 & 0.0751 \end{bmatrix}$$

la cual ya está normalizada de forma que la sumatoria de todos los elementos resulta en 1, e **implementa una covarianza nula entre las dos dimensiones**.

La Figura 23 ilustra la superficie generada por una función Gaussiana. Además, la Figura 23 ilustra el efecto de convolucionar una imagen con un núcleo Gaussiano.

5.2 Filtros paso altos: detección y mejora de bordes

5.2.1 Filtros de aproximación al gradiente de primer grado (primera derivada)

Los filtros de Prewitt y Sobel para la detección de bordes, aproximan al gradiente para una función (matriz) $U \in \mathbb{R}^{N \times M}$ con $z = U[x, y]$ como:

$$\nabla U = \frac{d}{dx} U[x, y] \hat{i} + \frac{d}{dy} U[x, y] \hat{j} = \nabla_x U \hat{i} + \nabla_y U \hat{j}$$

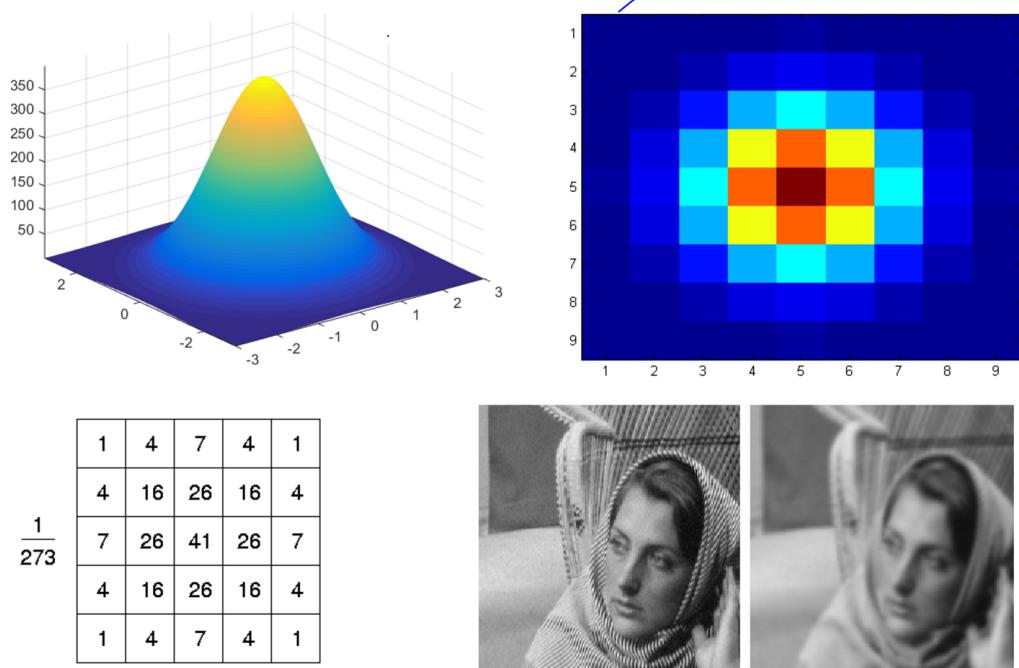


Figure 23: Ilustraciones de la función de densidad Gaussiana en distintas perspectivas, y el resultado de convolucionar una imagen con un núcleo Gaussiano.

Las máscaras en general utilizadas para realizar la aproximación de las derivadas parciales $\frac{d}{dx}U[x, y]$ y $\frac{d}{dy}F[x, y]$ están dadas, respectivamente como:

$$d_x = \alpha \begin{bmatrix} 1 & 0 & -1 \\ k & 0 & -k \\ 1 & 0 & -1 \end{bmatrix} \quad d_y = \alpha \begin{bmatrix} 1 & k & 1 \\ 0 & 0 & 0 \\ -1 & -k & -1 \end{bmatrix}$$

donde para el filtro de Prewitt $k = 1$ y $\alpha = \frac{1}{6}$, y para el filtro de Sobel $k = 2$ y $\alpha = \frac{1}{8}$. El coeficiente α se asegura de preservar la amplitud original de la señal para las derivadas parciales, y el coeficiente k hace más sensible el operador a los cambios en la señal. De esta manera, las derivadas parciales se aproximan según Prewitt y Sobel como:

$$\frac{d}{dx}U \approx U * d_x \quad \frac{d}{dy}U \approx U * d_y.$$

Para la detección de bordes, es suficiente conocer la magnitud del gradiente en cada punto, por lo que entonces se puede aproximar el mismo como:

$$\|\nabla U(x, y)\| = \sqrt{\left(\frac{d}{dx}U[x, y]\right)^2 + \left(\frac{d}{dy}U[x, y]\right)^2}$$

Una umbralización automática, usando por ejemplo el algoritmo de Otsu, permite detectar los bordes a partir de la matriz de magnitudes de gradientes $\|\nabla f(x, y)\|$. Un problema notable de la aproximación de la magnitud del gradiente es su sensibilidad al ruido, por lo que para el caso de la detección de bordes, se utilizan los cruces por cero de la segunda derivada. La Figura 24 muestra como la aproximación de la segunda derivada facilita la detección de bordes, usando el concepto de cruces por cero.

5.2.2 Filtros de aproximación al gradiente de segundo grado (segunda derivada), el Laplaciano del Gaussiano (LoG)

Ya anteriormente se demostró que la convolución de una función $u(t)$, $u : \mathbb{R} \rightarrow \mathbb{R}$ con una máscara $s[n] = (1 \ -2 \ 1)$ aproxima a la segunda derivada. La extensión de esta máscara como una función $L(x, y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ para aproximar la segunda derivada se puede definir como:

$$L = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -k & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

con $k = 4$ o $k = 8$ por ejemplo, la máscara aproxima primero la segunda derivada respecto a x , y luego respecto a y , en una sola máscara:

$$(U * L)[x, y] = \nabla_{x,y}U[x, y].$$

La Figura 24 muestra la aproximación a la segunda derivada. Para mejorar la robustez ante el ruido, es común convolucionar la imagen de entrada U con un

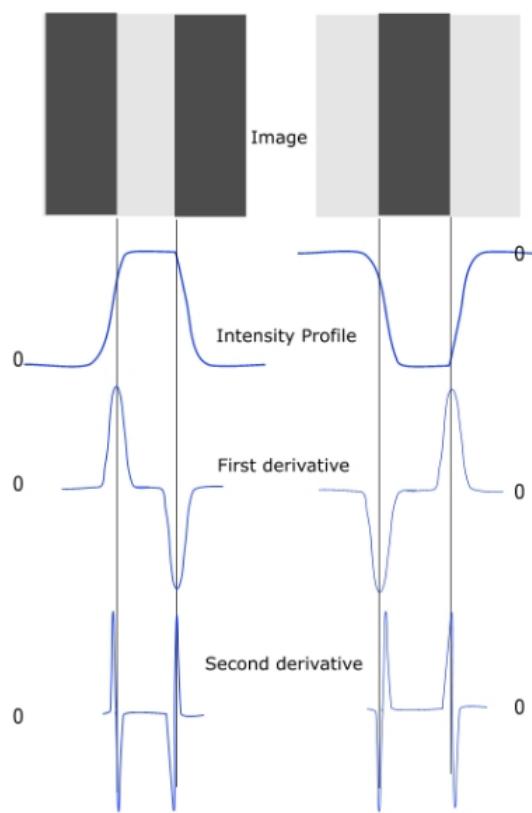


Figure 24: Aproximación de la primera y segunda derivada.

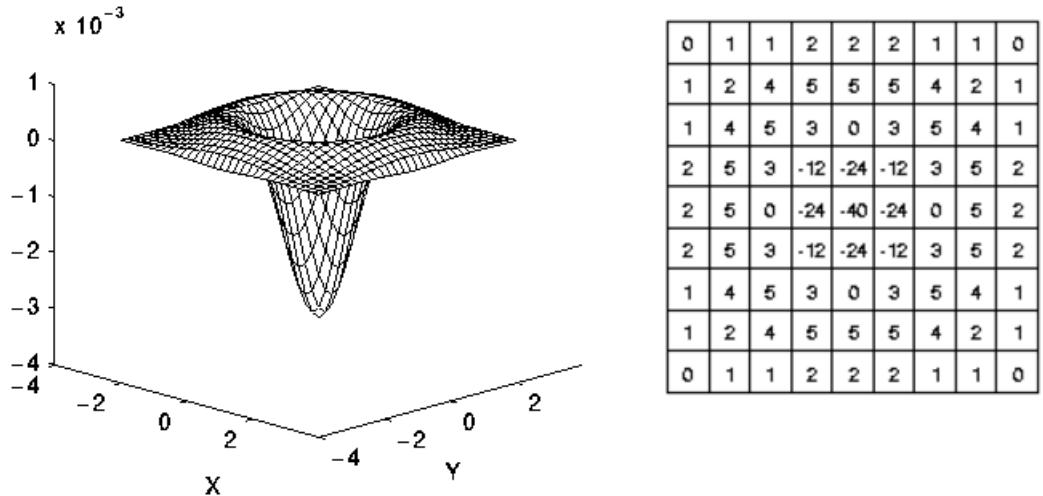


Figure 25: Laplaciano del Gaussiano, ejemplo. Tomado de <https://www.quora.com/What-happens-to-the-image-when-Laplacian-of-Gaussian-is-applied>

núcleo Gaussiano \mathcal{N} , para posteriormente convolucionar la matriz resultante por la máscara Gaussiana L , de modo que

$$((U * \mathcal{N}) * L)[x, y]$$

, lo que supone dos convoluciones o deslizamientos sobre la imagen $U \in \mathbb{R}^{N \times M}$. Para ahorrar costo computacional, utilizamos las propiedades de la conmutatividad y la asociatividad, de modo que

$$((U * \mathcal{N}) * L)[x, y] = (U * (L * \mathcal{N}))[x, y],$$

lo que resulta en una nueva máscara conocida como el Laplaciano del Gaussiano, o LoG:

$$\text{LoG}[x, y] = (L * \mathcal{N})[x, y],$$

observe que los coeficientes de la máscara LoG dependen de las desviaciones estándar usadas en la máscara Gaussiana σ_x y σ_y . La Figura 25 muestra gráficamente una máscara LoG de ejemplo.

La convolución de la imagen de entrada U con el núcleo LoG resulta en una imagen

$$E = U * \text{LoG}$$

con cruces por cero donde existan bordes, como ilustra la Figura 26, la cual fue normalizada en el rango de 0-255 para evitar la representación de los valores negativos.

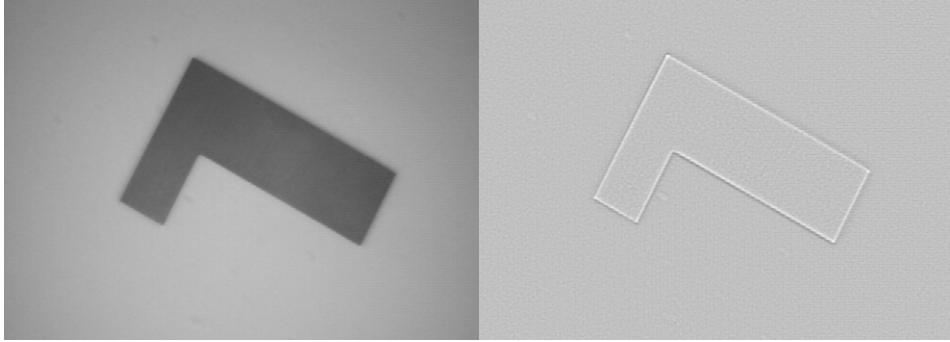


Figure 26: Visualización de la imagen $E = U * \text{LoG}$.

Una forma alternativa de calcular la respuesta al filtro LoG E es utilizando lo que se conoce como la diferencia entre Gaussianas (DoG), la cual consiste en tomar la diferencia entre la imagen original U y el resultado de convolucionar tal imagen con un núcleo Gaussiano \mathcal{N} , por lo que entonces se tiene que:

$$E \approx U - (U * \mathcal{N})$$

a la imagen $B = -E$ se le conoce como la máscara deforme o *unsharp mask*. La Figura 27 ilustra el proceso de aproximación de la segunda derivada usando DoG, y el proceso que sucede al observar la gráfica de un borde.

5.2.3 Mejora de bordes: *Unsharp Masking*

Para la mejora de bordes, se puede usar la máscara deforme o unsharp mask definida anteriormente como $B = -E$ y super-ponerla para producir la imagen con bordes mejorados G :

$$G = U + \lambda B$$

donde $\lambda \in \mathbb{R}$ es el coeficiente que controla la mejora de los bordes. La Figura 28 ilustra el proceso de mejora de bordes definido anteriormente, y conocido también como *unsharped masking*.

6 Filtros no lineales

Los filtros no lineales permiten generar comportamientos más sofisticados, y evitar problemas como por ejemplo la degradación de bordes en el caso de los filtros paso bajos, o la magnificación del ruido en los filtros paso altos. A continuación se presentan algunos filtros no lineales para distintos propósitos.

6.1 Filtro de medianas

El filtro de medianas es un filtro no lineal, el cual toma la mediana de cada ventana. La mediana se define como el valor a la *mitad* de un conjunto de

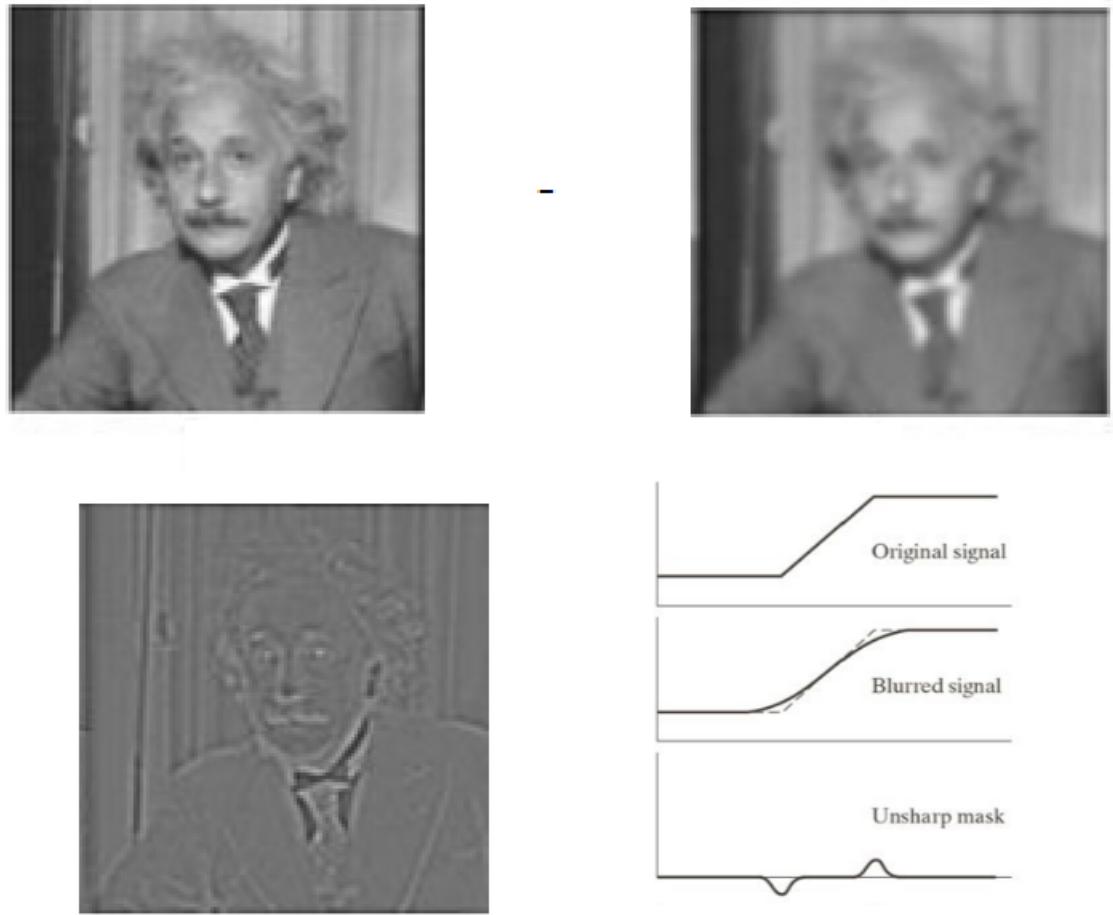


Figure 27: Ilustración de DoG. Tomado de http://ict.udlap.mx/people/oleg/docencia/imagenes/chapter3/image_32_IS548.html



Figure 28: Unsharped Masking con distintos valores de λ . Tomado de http://ict.udlap.mx/people/oleg/docencia/imagenes/chapter3/image_32_IS548.html



Figure 29: Imagen contaminada con ruido de tipo sal y pimienta y el resultado de filtrarla con un filtro de medianas.

datos ordenados. Por ejemplo, para la siguientes matriz $A \in \mathbb{R}^{N \times N}$, con en este caso $N = 3$:

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 5 & 7 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

al calcular un vector \vec{a} con todos los elementos ordenados de mayor a menor se tiene que:

$$\vec{a} = [0 \ 1 \ 1 \ 2 \ \mathbf{2} \ 3 \ 4 \ 5 \ 7]$$

el elemento a la mitad o en la mediana del arreglo ordenado, en la posición $\frac{N \times N - 1}{2}$ (indizado desde 0) de los elementos de A viene dado por:

$$\text{mediana}(A) = 2.$$

El filtro de medianas define con una ventana de $K \times L$ dimensiones, realiza el cálculo de la mediana para esa ventana deslizante en una imagen de entrada $U \in \mathbb{R}^{N \times M}$:

$$G[x, y] = \text{mediana}\left(U_{x - \frac{(K-1)}{2}:x + \frac{(K-1)}{2}, y - \frac{(L-1)}{2}:y + \frac{(L-1)}{2}}\right), \quad (8)$$

El filtro de medianas es un filtro de la familia de filtros de *orden de rango* o *rank order filters*, los cuales obtienen resultados satisfactorios en señales contaminadas con ruido de tipo *sal y pimienta* como el ilustrado en la Figura 29.

6.2 Filtro bilateral

El filtro bilateral, además de pesar el dominio espacial con una función Gaussiana, pesa el dominio de la intensidad con otra función Gaussiana, de modo que los pixeles muy distintos al pixel en el que está centrado la ventana deslizante, obtienen un peso más bajo, evitando la degradación de los bordes.

El filtro bilateral [?] implementa una función dependiente de la entrada U :

$$S_{\text{BF}}[x, y] = \frac{1}{\alpha} \sum_{u=0}^{K-1} \sum_{v=0}^{L-1} U\left[x + u - \left(\frac{K-1}{2}\right), y + v - \left(\frac{L-1}{2}\right)\right] \psi_{\text{BF}}[[u, v], [x, y]]. \quad (9)$$

$$\psi_{\text{BF}}[[u, v], [x, y]] = \exp\left(-\frac{(u-x)^2 + (v-y)^2}{2\sigma_s^2}\right) \exp\left(-\frac{(U[x, y] - U[u, v])^2}{2\sigma_r^2}\right).$$

Como se puede observar, se definen dos parámetros, σ_r y σ_s , los cuales controlan el ancho de la campana Gaussiana en el dominio de la intensidad y del espacio, respectivamente. La Figura 30 muestra el efecto de cada parámetro.

Analizando el comportamiento del Filtro bilateral matemáticamente, y según la Figura 30, se observa que si $\sigma_s \rightarrow \infty$:

$$F[x, y] = \exp\left(-\frac{(u-x)^2 + (v-y)^2}{2\sigma_s^2}\right) \overset{1}{\exp}\left(-\frac{(U[x, y] - U[u, v])^2}{2\sigma_r^2}\right) = \exp\left(-\frac{(U[x, y] - U[u, v])^2}{2\sigma_r^2}\right),$$

el filtro pasa a usar una ventana rectangular ideal, pues no existe un pesado en el dominio espacial. Lo anterior simplifica el cómputo, pero conlleva el efecto de *aliasing* y difuminación sensible a la dirección (borroneado cuadrado).

Respecto al efecto de hacer $\sigma_r \rightarrow \infty$ se tiene que:

$$F[x, y] = \exp\left(-\frac{(u-x)^2 + (v-y)^2}{2\sigma_s^2}\right) \overset{1}{\exp}\left(-\frac{(U[x, y] - U[u, v])^2}{2\sigma_r^2}\right) = \exp\left(-\frac{(u-x)^2 + (v-y)^2}{2\sigma_s^2}\right),$$

por lo que el filtro se comporta como un filtro Gaussiano, el cual no toma en cuenta la intensidad de la imagen, por lo que los bordes se degradan.

6.3 Filtro de promediado no local

Ante situaciones de mucho ruido (una relación señal a ruido muy baja) el filtro bilateral no presenta un buen desempeño. El Filtro de promediado no local (Non Local Means o NLM) utiliza la premisa de pesar no solo la similitud de intensidad entre dos pixeles $p_1 = (x_1, y_1)$ y $p_2 = (x_2, y_2)$, sino también la similitud de los vecindarios de ambos pixeles.

La Figura 31 muestra a los pixeles q_1 , q_2 y q_3 , además de los vecindarios de cada uno en el recuadro blanco de $M \times M$ pixeles, donde en tal caso los pixeles q_1 y q_2 son bastante más similares entre si que los pixeles q_3 y q_1 . De esta forma se minimiza la incorrecta interpretación que hace el filtro Bilateral de comparar las intensidades de dos pixeles p_1 y p_2 que sean similares en cuanto

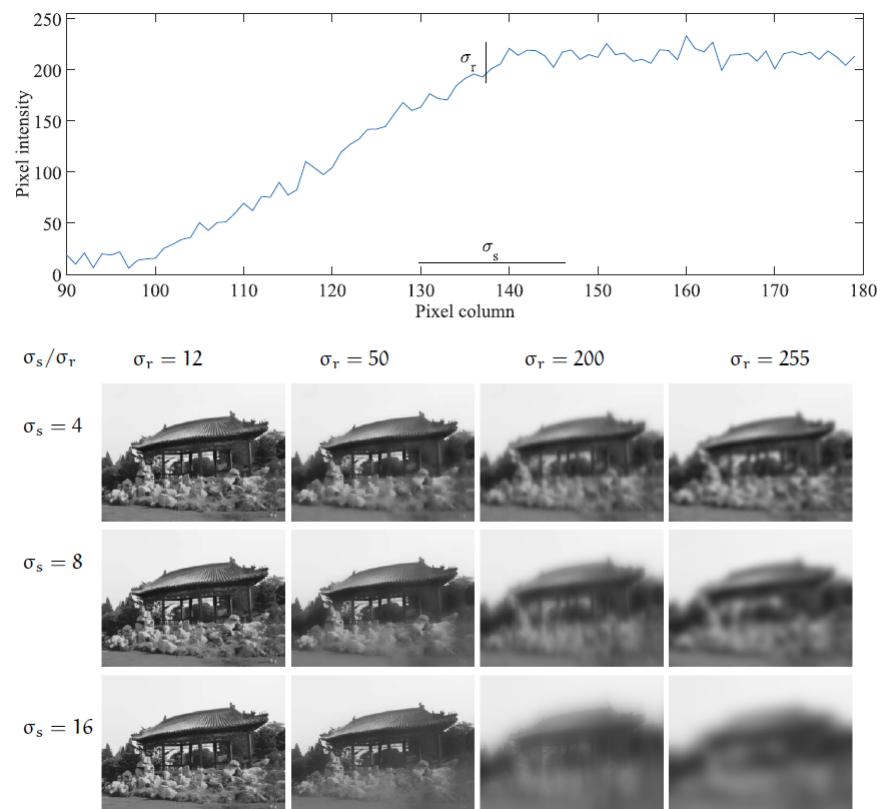


Figure 30: Efectos de los dos parámetros del filtro bilateral, tomado de P. Kornprobst and J. Tumblin, *Bilateral filtering: Theory and applications*. Now Publishers Inc, 2009..

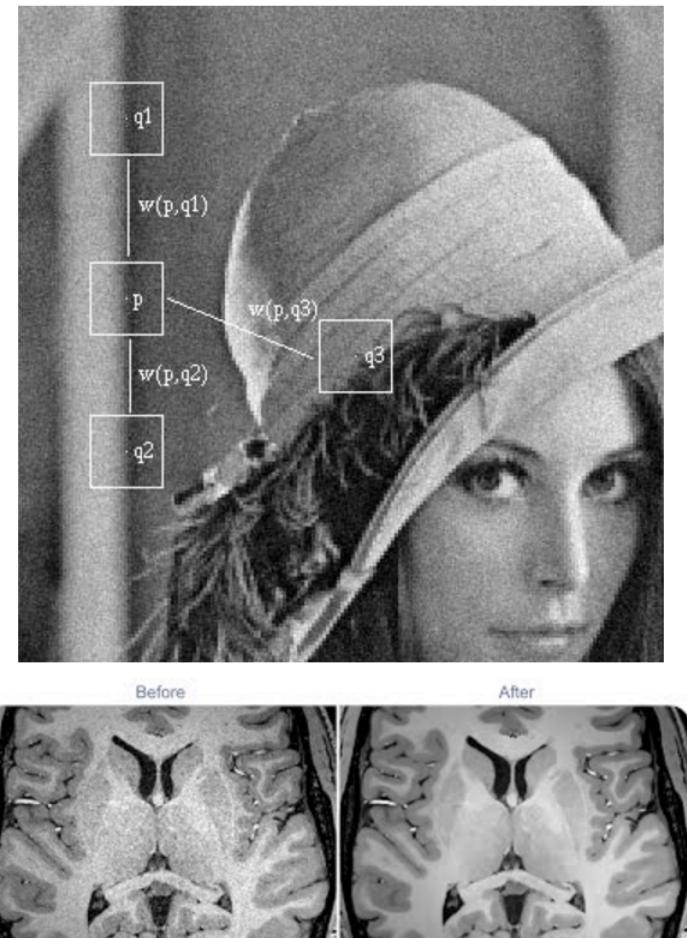


Figure 31: Filtro de promediado no local, tomado de *A non-local algorithm for image denoising*.

a intensidad por causa del ruido (en cuyo caso usualmente no existe una similitud estructural de sus vecindarios). En tal caso, el filtro NLM resultaría en un peso bajo al comparar los vecindarios de los pixeles p_1 y p_2 .

Para asignar la similitud de los vecindarios entre pixeles, el NLM vectoriza los vecindarios de ambos pixeles en $\vec{\eta}(p_1)$ y $\vec{\eta}(p_2)$, respectivamente, y realiza la comparación de tales vecindarios usando por ejemplo la norma euclidiana $\|\vec{\eta}(p_1) - \vec{\eta}(p_2)\|_2$. De esta forma, la ecuación del filtro se expresa como:

$$F[x, y] = \exp\left(-\frac{(x^2 + y^2)}{2\sigma_s^2}\right) \exp\left(-\frac{\|\vec{\eta}(p_1) - \vec{\eta}(p_2)\|_2}{2\sigma_r^2}\right).$$

$$\psi_{\text{NLM}}[[u, v], [x, y]] = \exp\left(-\frac{((u-x)^2 + (v-y)^2)}{2\sigma_s^2}\right) \exp\left(-\frac{\|\vec{\eta}[x, y] - \vec{\eta}[u, v]\|}{2\sigma_r^2}\right).$$

Una implementación a *fuerza bruta* del NLM es considerablemente más costosa que por ejemplo una implementación del mismo tipo del filtro bilateral. Es por eso que la implementación de este tipo usual del NLM elimina el término de pesado espacial, usando una ventana rectangular:

$$F[x, y] = \exp\left(-\frac{\|\vec{\eta}[x, y] - \vec{\eta}[u, v]\|}{2\sigma_r^2}\right).$$

Se pueden además considerar otras distancias más rápidas de calcular, como una del tipo ℓ_∞ .

6.4 Filtros de detección de bordes: Detección de bordes Laplaciana

Un aspecto posiblemente deseable para la posterior etapa de extracción de características es la detección de bordes en una imagen. Ya se introdujeron los filtros de Sobel, Prewitt y Laplacianos para la aproximación de la primera y segunda derivadas. Basados en tales aproximaciones a la primer y segunda derivada, puede ya sea umbralizarse el resultado según la magnitud del gradiente de primer grado, o buscar los cruces por cero para el gradiente de segundo grado. La umbralización consiste en escoger un valor T el cual permita crear una nueva imagen umbralizada W la cual esté dada por:

$$W(i, j) = 1 \quad \text{Si} \quad U(i, j) \geq T$$

$$W(i, j) = 0 \quad \text{Si} \quad U(i, j) < T$$

La imagen umbralizada W es entonces una imagen binaria. El determinar el valor óptimo T es de capital importancia para una detección de bordes satisfactoria. Un valor incorrecto puede tomar pixeles correspondientes a ruido. Otro problema es la aparición de bordes *dobles*, dado que usualmente es necesario detectar bordes de un pixel de ancho, que además no se separan o *rompan*.

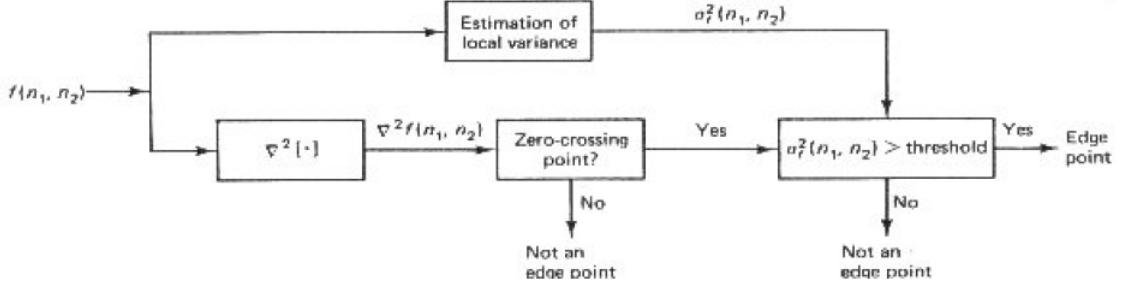


Figure 32: Detección de bordes según el detector Laplaciano. Tomado de https://nptel.ac.in/courses/117104069/chapter_8/8_23.html

El siguiente algoritmo baja la tasa de falsos positivos al ser más robusto ante condiciones ruidosas, usando la aproximación de la segunda derivada resultante al convolucionar la imagen de entrada U con la máscara Laplaciana del Gaussiano $\nabla^2 U \approx U * \text{LoG}$. Un cruce por cero puede ser provocado por pixeles ruidosos, por lo que el algoritmo de *detección de bordes Laplaciana* calcula la imagen de varianza local, la cual se define como:

$$V[x, y] = \text{var} \left(U_{x - \frac{(K-1)}{2} : x + \frac{(K-1)}{2}, y - \frac{(L-1)}{2} : y + \frac{(L-1)}{2}} \right), \quad (10)$$

donde la varianza para una matriz $A \in \mathbb{R}^{K \times L}$ con una ventana de $P \times P$ (con $P = 3$ usualmente) se define como:

$$\sigma^2(A) = \text{var}(A) = \frac{1}{P^2 - 1} \sum_{i=1}^P \sum_{j=1}^P (A[i, j] - \mu_A)^2$$

con:

$$\mu_A = \frac{1}{P^2} \sum_{i=1}^P \sum_{j=1}^P A[i, j]$$

Basado en el cálculo de la varianza local, el algoritmo de detección de bordes Laplaciana implementa el flujo mostrado en la Figura 32.

De esta forma, si un punto es un cruce por cero, y además tiene una alta varianza, se clasifica como un punto de borde. Existen algoritmos de detección de bordes más sofisticados como el algoritmo de Canny, que además incluyen etapas de postprocesado para reconstruir bordes discontinuos. Para detectar líneas en una imagen, se usan algoritmos como la transformada de Hough.

7 Métricas para la eliminación del ruido y exactitud de bordes

7.1 Métricas para la eliminación del ruido: Relación señal a ruido

Existen distintas métricas para medir la cantidad de ruido presente en una señal. La relación pico señal-ruido (*peak signal to noise ratio*) es muy utilizada en Ingeniería Eléctrica, y la medición de las ondas sonoras. Se basa en el logaritmo de base 10 para representar la relación entre la potencia eléctrica y la potencia de la señal (para bajas potencias eléctricas, un cambio tiene mayor efecto en la potencia de la señal que en altas potencias eléctricas). Su fórmula viene dada por:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) = 20 \log_{10} \left(\frac{\text{MAX}}{\sqrt{\text{MSE}}} \right) = 20 \log_{10} (\text{MAX}) - 10 \log_{10} (\text{MSE}),$$

donde MAX se refiere al máximo valor posible de la señal, por ejemplo, para imágenes a escala de grises $\text{MAX} = 255$ y el error medio cuadrático (*mean squared error*) de la diferencia entre la señal original, sin contaminar por ruido $U \in \mathbb{R}^{m \times n}$ (la imagen de entrada), y la imagen Y contaminada por ruido:

$$\text{MSE} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (U(i, j) - Y(i, j))^2.$$

La relación $\frac{\text{MAX}^2}{\text{MSE}}$ asegura que el argumento del logaritmo sea siempre mayor que 1, pues el máximo error promedio puede llegar a ser MAX^2 . El logaritmo multiplicado por 10 (de allí el prefijo *decí*) se comporta como lo mostrado en la Figura 33, donde se observa que para bajos valores, un cambio supone una mayor ganancia, y en altos valores, un cambio supone una ganancia menor.

Observe que en caso de existir un alto error medio cuadrático, relación señal a ruido será más baja, de lo contrario, en presencia de poco ruido, la relación señal a ruido es bastante más alta, como lo ilustra la Figura 34.

7.2 Métricas para la comparación entre señales binarias

Para evaluar la precisión de la segmentación, las métricas usuales evalúan la diferencia de señales binarias (donde por ejemplo los pixeles etiquetados con 1 se refieren a los objetos de interés y los etiquetados con 0 al fondo). Una métrica muy utilizada para tales efectos es el coeficiente de *dice* o de Sørensen, el cual se puede usar para evaluar la similitud estructural entre dos imágenes binarias $U \in \mathbb{R}^{m \times n}$ y $V \in \mathbb{R}^{m \times n}$, y viene dado por:

$$S = \frac{2 \left(\sum_j^m \sum_i^n U[i, j] V[i, j] \right)}{\left(\sum_j^m \sum_i^n U[i, j] \right) + \left(\sum_j^m \sum_i^n V[i, j] \right)}$$

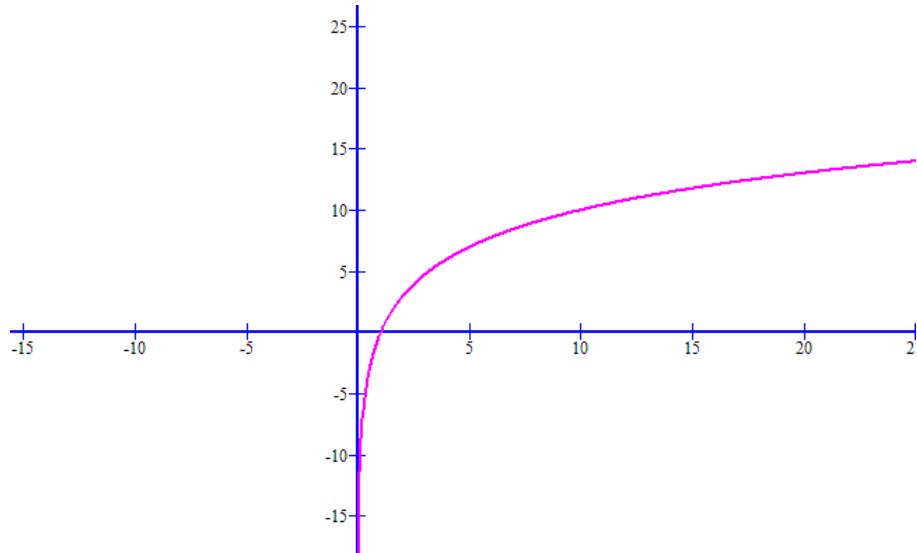


Figure 33: Graficación de la función \log_{10} .

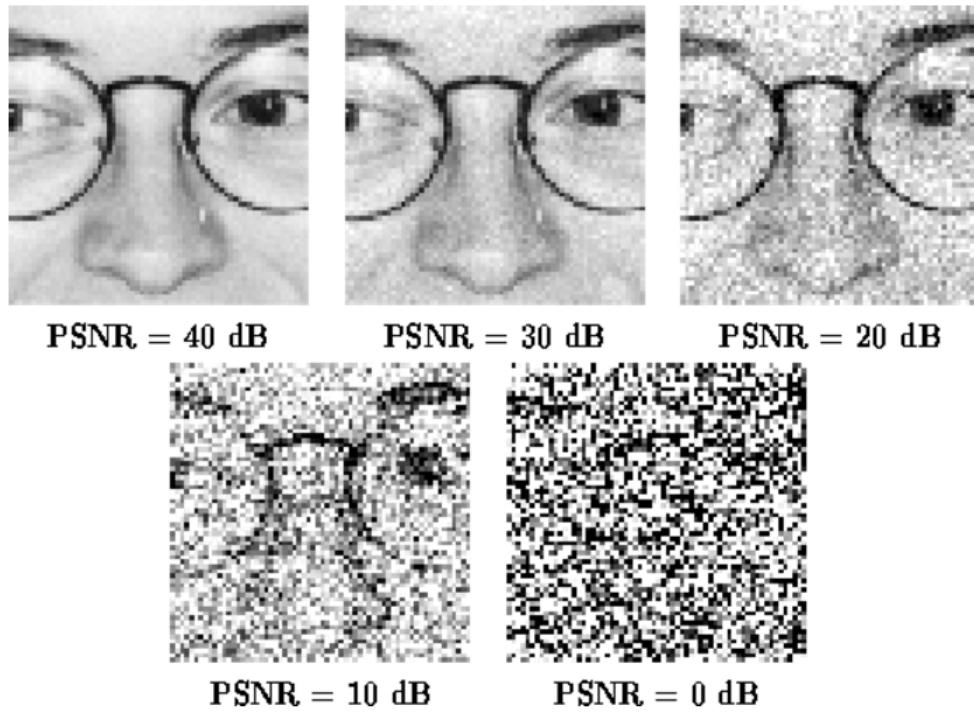


Figure 34: PSNR de distintos valores. Tomado de http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/VELDHUIZEN/node18.html



Figure 35: *Groundtruth* a la izquierda, e imagen automáticamente segmentada a la derecha. Tomado de <https://blog.altoros.com/experimenting-with-deep-neural-networks-for-x-ray-image-segmentation.html>.

donde el numerador equivale conceptualmente a la cantidad de pixeles del objeto de interés intersecados en ambas imágenes U y V : $2|U \cap V|$ (donde las barras denotan la cantidad de pixeles con 1's en la matriz) y el denominador a la suma de los pixeles del objeto de interés en ambas imágenes: $|U| + |V|$, lo que entonces significa que se puede reescribir S como:

$$S = \frac{2|U \cap V|}{|U| + |V|}$$

La imagen U puede asociarse a la imagen segmentada por el algoritmo, y V la imagen segmentada manualmente, también conocida como imagen de *groundtruth*. De este modo, una segmentación perfecta según el *groundtruth* resulta en un coeficiente de *dice* $S = 1$, y una segmentación completamente errada un coeficiente $S = 0$. La Figura 35 ilustra la segmentación automática de una radiografía pulmonar, y el *groundtruth* de la misma.

References

- [1] Rafael C Gonzalez, Richard Eugene Woods, and Steven L Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.
- [2] Pablo Irarrázaval. *Análisis de señales*. McGraw-Hill Interamericana, 1999.