Departamento de Computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires

Reducción de ruido con Transformada Discreta del Coseno

Trabajo Práctico 1,
Métodos Numéricos,
Primer Cuatrimestre de 2013

Apellido y Nombre	LU	E-mail
María Candela Capra Coarasa	234/11	canduh_27@hotmail.com
Leandro Lovisolo	645/11	leandro@leandro.me
Lautaro José Petaccio	443/11	lausuper@gmail.com

Resumen:
Se aplican distintos métodos para agregar ruido a señales sonoras e imágenes para luego, utilizando la Transformada Discreta del Coseno, plantear y sacar conclusiones sobre la eficiencia de varias técnicas de reducción del ruido agregado.

Palabras claves: Reducción, ruido, DST, sonido, imágenes, frecuencia.

Índice

1. Introducción Teórica

En este trabajo exploramos un conjunto de métodos basados en la Transformada Discreta del Coseno para eliminar ruidos aperiódicos sobre muestras de señales de una y dos dimensiones (ejemplo: audio e imágenes, respectivamente).

La Transformada Discreta del Coseno, de ahora en más DCT, permite expresar un vector en $\mathbb{R}^n$ como combinación lineal de vectores de la forma $\{(cos(0*t_0), \ldots cos(0*t_{n-1})), \ldots (cos(n-1*t_0), \ldots cos(n-1*t_{n-1}))\}$, donde n es el tamaño de la muestra y $t_i = (i+\frac{1}{2})\frac{\pi}{n}$. Gráficamente, se puede interpretar este cambio de base como la escritura de la muestra como una suma de cosenos de distintas frecuencias, donde las coordenadas del vector transformado son coeficientes que determinan la amplitud de cada coseno, y se presentan ordenados de menor a mayor frecuencia.

En el caso de señales bidimensionales representadas con matrices $\mathbb{R}^{n\times n}$, la matriz transformada obtenida es análoga al caso unidimensional, en la que se pueden obtener los coeficientes ordenados de menor a mayor frecuencia recorriendo las coordenadas diagonalmente de derecha a izquierda y arriba a abajo, partiendo de la coordenada superior izquierda.

Los tipos de ruido aplicados a las muestras son aditivo e impulsivo, descritos a continuación:

Ruido aditivo: Suma a cada elemento de la muestra una variable aleatoria con distribución normal, con media cero y distintas varianzas de acuerdo al experimento.

Ruido impulsivo: Reemplaza cada elemento de la muestra por máx o mín con probabilidad p, donde máx y mín representan el valor máximo y mínimo que adquieren los elementos de la muestra, y p variable de acuerdo al experimento.

2. Desarrollo

2.1. Atenuar

2.1.1. Ruido aditivo en audio

Muestra ramp1234.txt, varianza 10.
PSNR de la muestra con ruido agregado: 22.55 dB
PSNR de la muestra luego de aplicar el método: 25.03 dB

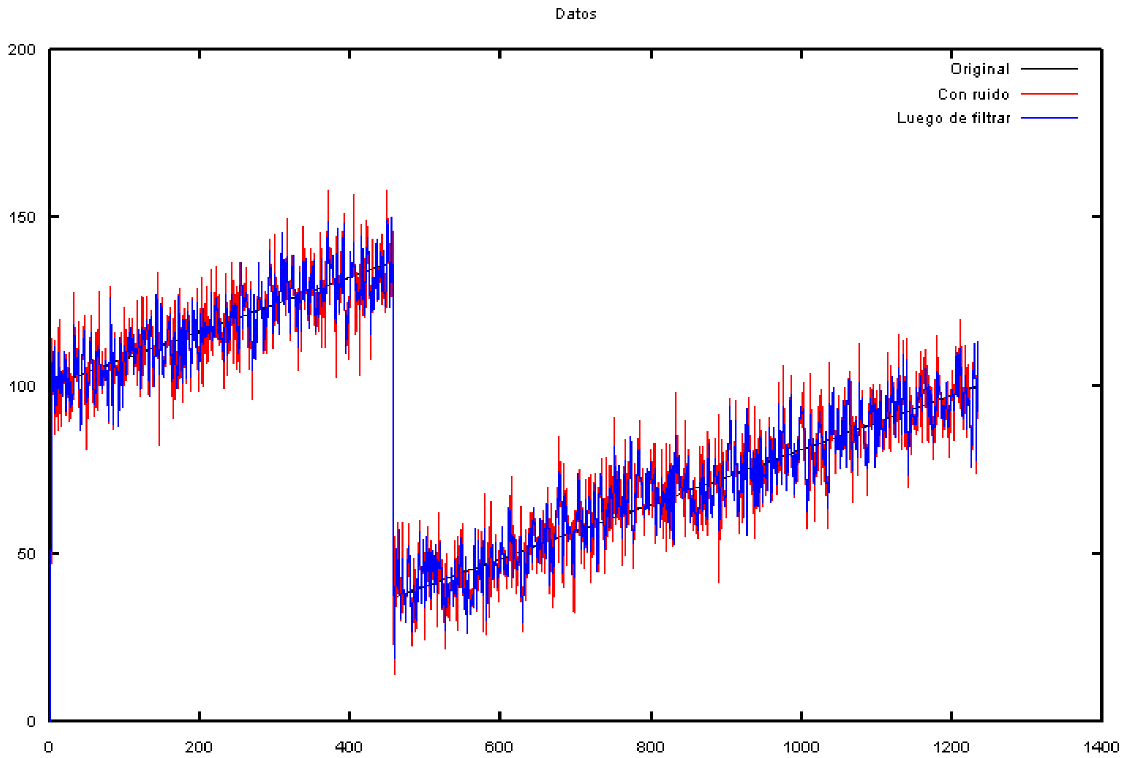


Figura 1: Muestra en base canónica.

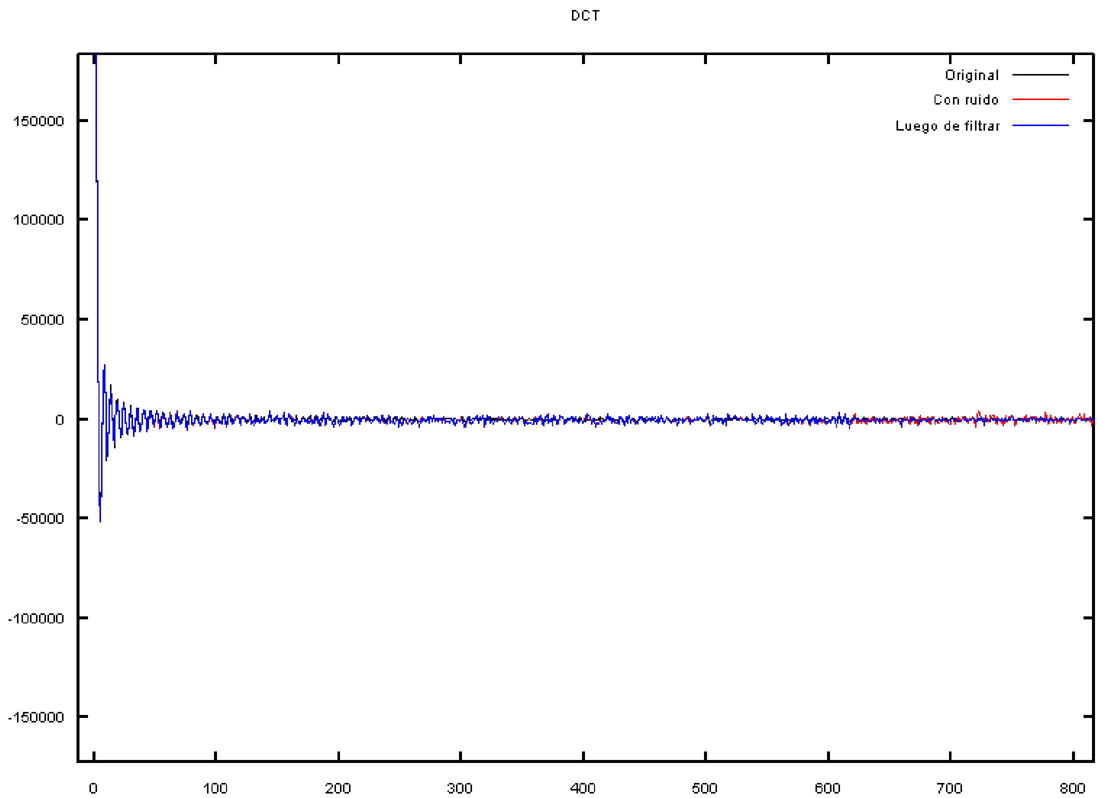


Figura 2: Muestra en base DCT.

Muestra dopp512.txt, varianza 10.
PSNR de la muestra con ruido agregado: 14.02 dB

PSNR de la muestra luego de aplicar el método: 16.28 dB

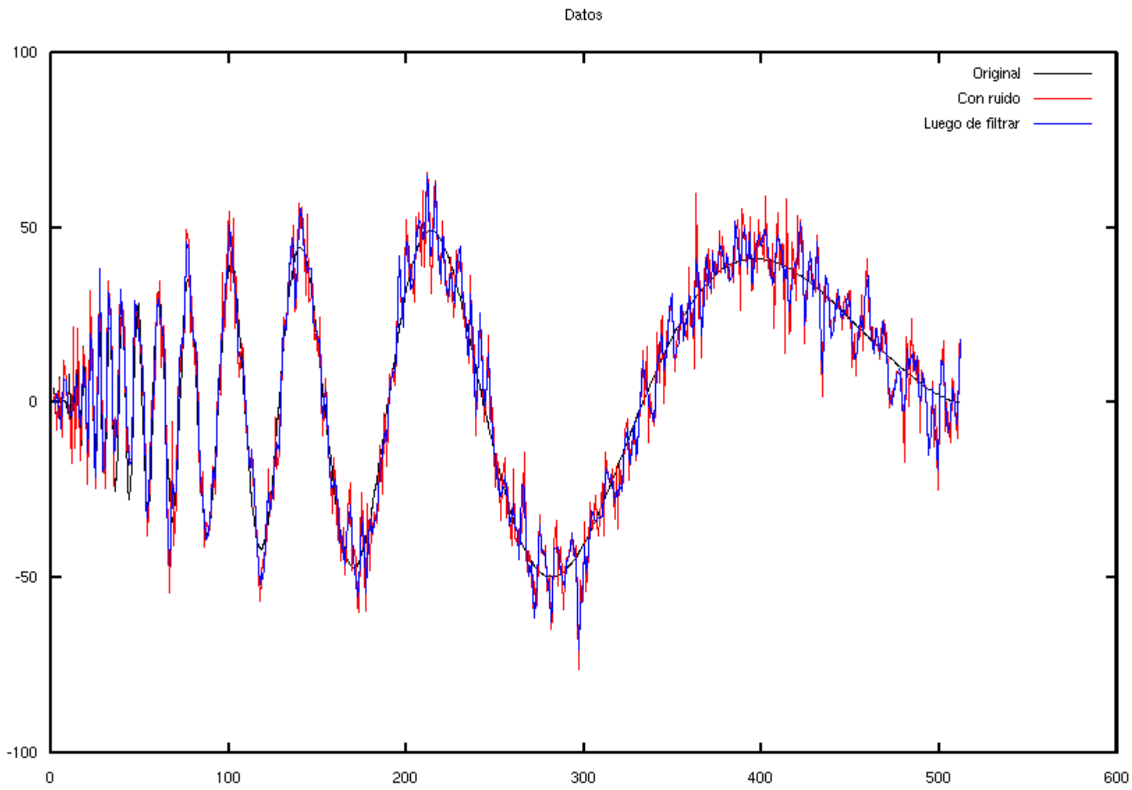


Figura 3: Muestra en base canónica.

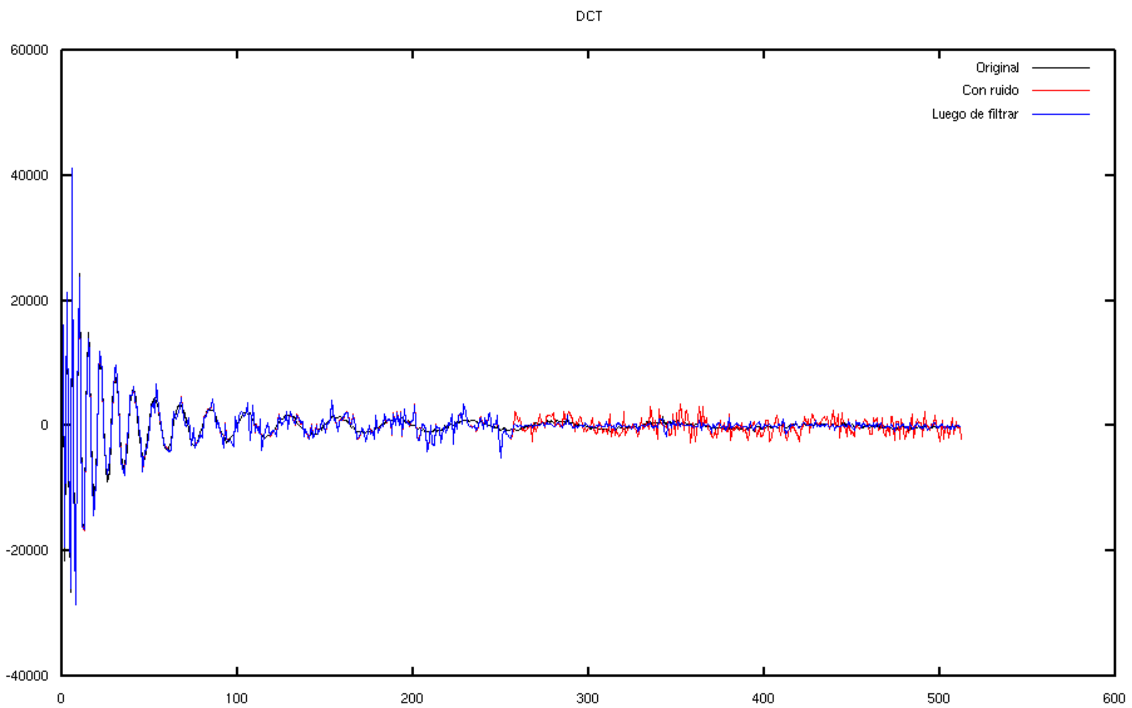


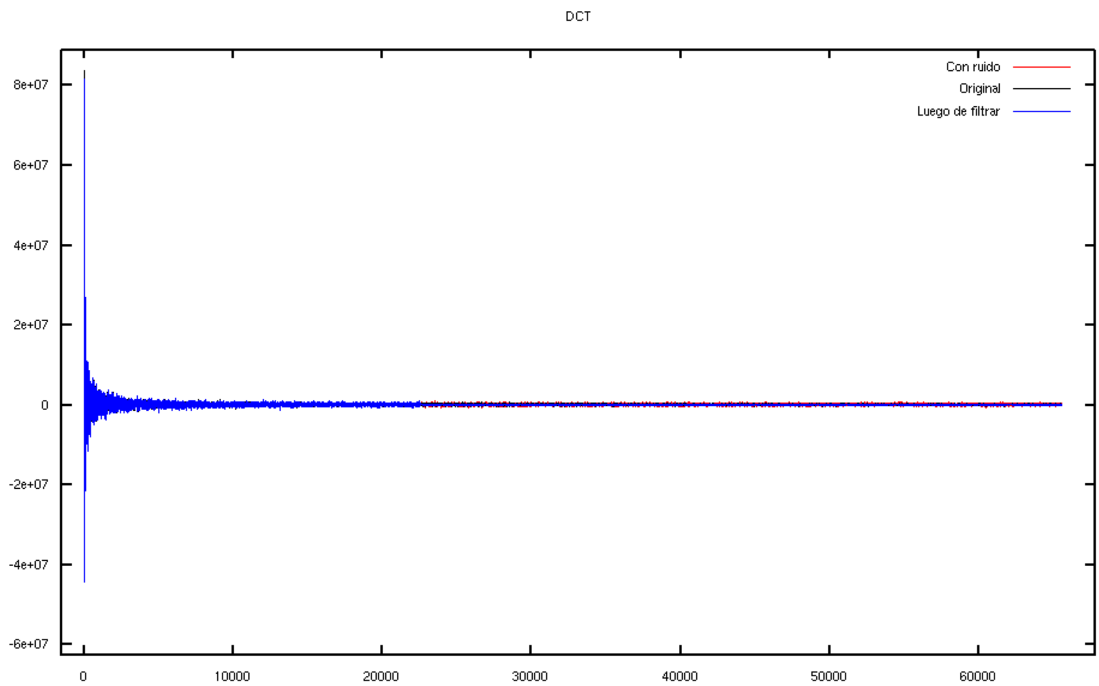
Figura 4: Muestra en base DCT.

2.1.2. Ruido aditivo en imágenes

Muestra lena.pgm, varianza 10.
PSNR de la muestra con ruido agregado: 27.23 dB
PSNR de la muestra luego de aplicar el método: 28.62 dB

(a) Antes de atenuar. (b) Después de atenuar.

Figura 5: Ruido aditivo en imágenes.



2.1.3. Ruido impulsivo en audio

Muestra dopp512.txt, varianza 10.
PSNR de la muestra con ruido agregado: 16.2445
PSNR de la muestra luego de aplicar el método: 18.2166

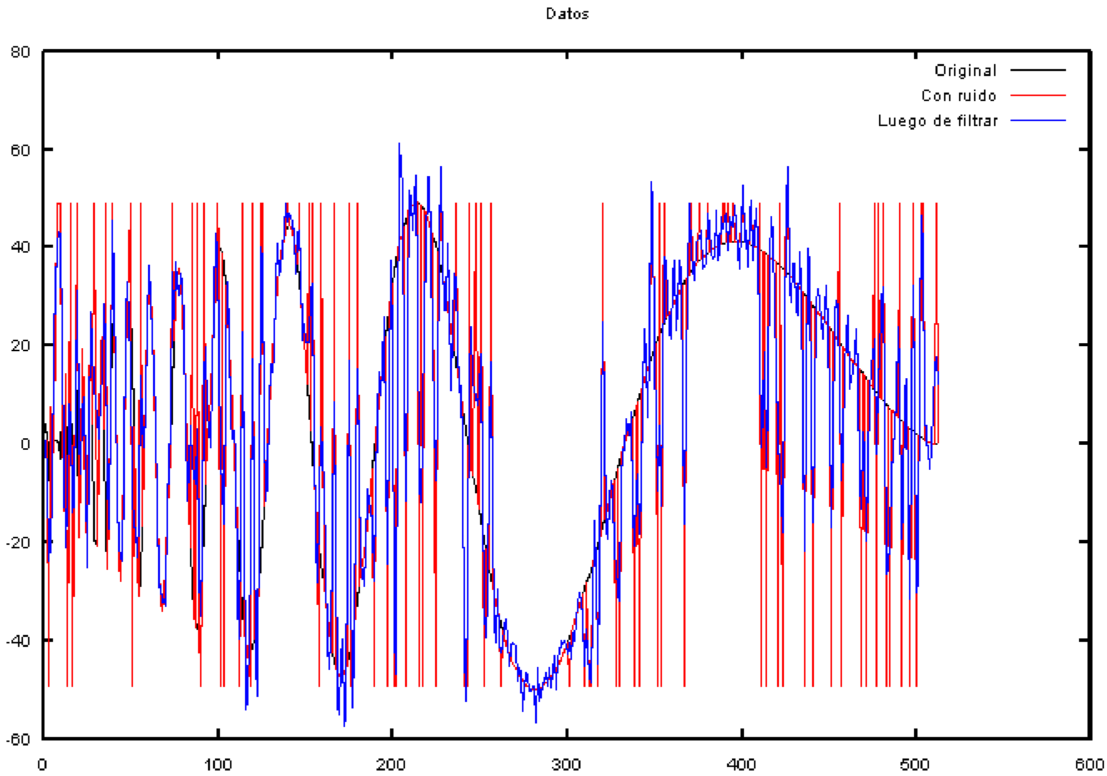


Figura 6: Muestra en base canónica.

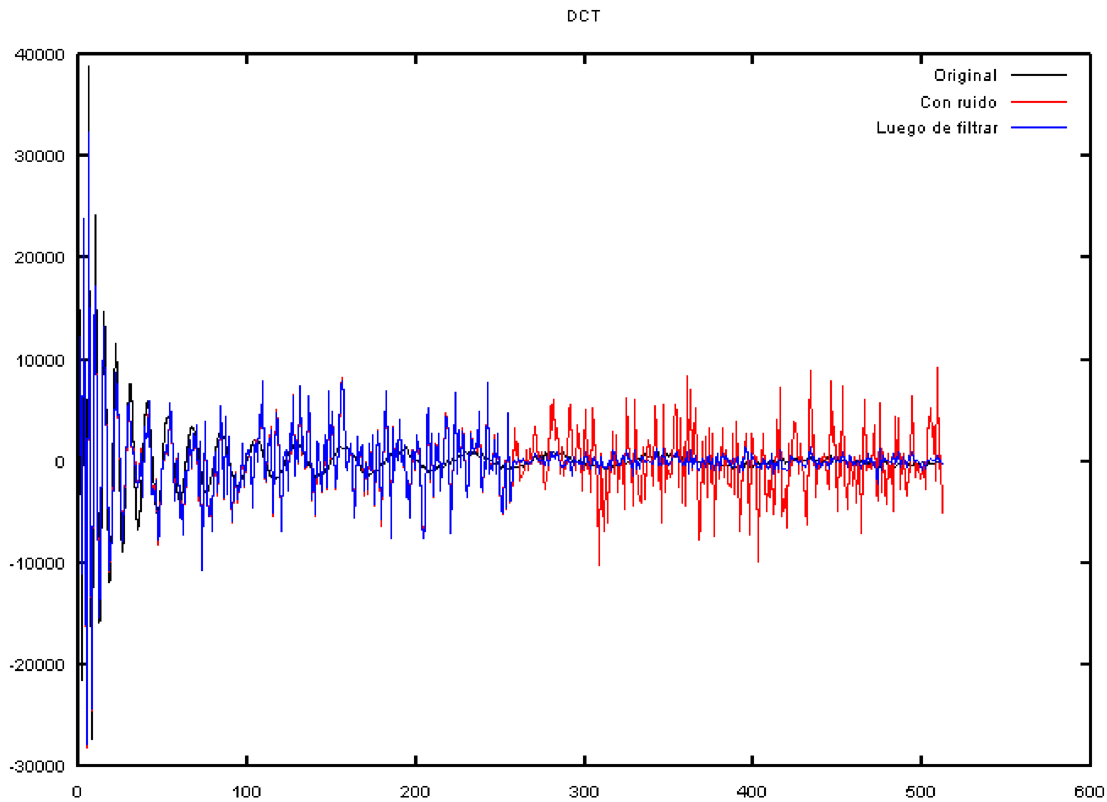


Figura 7: Muestra en base DCT.

3. Resultados

Pendiente.

4. Discusión

Pendiente.

5. Conclusiones

Pendiente.

6. Apéndice A: Enunciado del Trabajo Práctico

Introducción

La Transformada Discreta del Coseno (DCT, por sus siglas en inglés) es una herramienta que nos permite representar cualquier señal en el plano de las frecuencias. Dado que es utilizada por el estándar de compresión de imágenes JPEG y formato de video MPEG, se encuentra implementada en más lugares de lo que pensamos: en cada cámara digital o teléfono móvil. La DCT no solo tiene aplicaciones al mundo de la compresión (donde los valores transformados pueden ser codificados de forma eficiente), sino también al procesamiento: el análisis de qué frecuencias están presentes en las señales es esencial en ciertos contextos de aplicación.

La idea intuitiva de esta transformada, en el plano continuo, consiste en representar una función $f : \mathbb{R} \to \mathbb{R}$ en la base de funciones $\mathcal{B} = \{1, \cos(x), \cos(2x), ...\}$. En el plano discreto, la DCT se corresponde a un cambio de base: cada una de las funciones de la base $\mathcal{B}$ se discretiza en ciertos puntos pasando a ser una base de vectores en $\mathbb{R}^n$, (donde n es la dimensión del vector o señal a transformar). Es decir, dado un vector o señal $x \in \mathbb{R}^n$ existe una matriz $M \in \mathbb{R}^{n\times n}$ de cambio de base que define la transformada DCT, donde $y = Mx$ es el vector o señal transformado al espacio de frecuencias por la DCT (ver apéndice A). Esta operación es fácilmente extensible a señales de dos dimensiones (ver apéndice A.1).

Enunciado

El objetivo del trabajo es eliminar ruido sobre una señal ruidosa $x \in \mathbb{R}^n$. Para ello se realiza el siguiente proceso:

1. $y := Mx$ [Transformar usando Ec. (1) de Ap. A]
2. $\tilde{y} := f(y)$ [Modificar]
3. Resolver $M\tilde{x} = \tilde{y}$ [Reconstruir]

Una forma de medir la calidad visual de la señal reconstruida $\tilde{x}$, es a través del PSNR (*Peak Signal-to-Noise Ratio*). EL PSNR es una métrica 'perceptual' (acorde a lo que perciben los humanos) y nos da una forma de medir la calidad de una imagen perturbada, siempre y cuando se cuente con la señal original. Cuanto mayor es el PSNR, mayor es la calidad de la imagen. La unidad de medida es el decibel (db) y se considera que una diferencia de 0.5 db ya es notada por la vista humana. El PSNR se define como:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_x^2}{ECM}\right)$$

donde MAX_x define el rango máximo de la señal (en caso de entradas de 8 bits sin signo, sería 255) y ECM es el *error cuadrático medio*, definido como: $\frac{1}{n}\sum_i (x_i - \tilde{x}_i)^2$, donde n es la cantidad de elementos de la señal, x es la señal original y $\tilde{x}$ es la señal recuperada.

En la implementación realizada deben llevar a cabo los siguientes experimentos:

- Para varias señales con distintos niveles de ruido, se deberán experimentar con al menos 2 estrategias (definiendo f de forma conveniente) para modificar la señal transformada y (paso 2) con el objetivo de que la señal recuperada $\tilde{x}$ contenga menos ruido; se deberán extraer conclusiones en cuanto a la calidad de la señal recuperada, en función de la estrategia utilizada.
- Se deberán repetir los anteriores experimentos también sobre imágenes adaptando el método para aplicar la transformada DCT en dos dimensiones según se explica en la apéndice A.1.
- **(Opcional)** Se deberá analizar la aplicación de la DCT 'por bloques' sobre imágenes. Por ejemplo, si tenemos una imagen de 64×64 píxeles podemos subdividirla en: 4 bloques de 32×32, o 16 bloques de 16×16, o 64 bloques de 8×8, y aplicar la DCT en 2D sobre cada uno de los bloques (considerar un tamaño mínimo de 8×8 para cada bloque).
  Elegir una estrategia utilizada para señales unidimiensionales y sacar conclusiones respondiendo a

los siguientes interrogantes (realizando experimentos que justifiquen la respuesta): ¿Es lo mismo eliminar ruido sobre la imagen entera que de a bloques? ¿Qué forma es más conveniente en cuanto a la calidad visual? ¿Qué forma es más rápida?

Formatos de archivos de entrada

Las señales serán leídas de un archivo de texto en cuya primer línea figuran la cantidad de datos y en la línea siguiente se encuentran los datos en ASCII separados por espacios. Para leer y escribir imágenes sugerimos utilizar el formato *raw* binario `.pgm`[1]. El mismo es muy sencillo de implementar y compatible con muchos gestores de fotos[2] y Matlab.

Fecha de entrega:

- *Formato electrónico:* jueves 16 de mayo de 2013, hasta las 23:59 hs., enviando el trabajo (informe+código) a `metnum.lab@gmail.com`. El subject del email debe comenzar con el texto `[TP2]` seguido de la lista de apellidos de los integrantes del grupo.
- *Formato físico:* viernes 17 de abril de 2013, de 18 a 20hs (en la clase del labo).

[1] `http://netpbm.sourceforge.net/doc/pgm.html`
[2] XnView `http://www.xnview.com/`

A. Transformada Coseno Discreta

Para generar la matriz $M \in \mathbb{R}^{n\times n}$ que define la transformada de Coseno Discreta definimos:

- Vector de frecuencias: $g = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ n-1 \end{pmatrix}$
- Vector de muestreo: $s = \dfrac{\pi}{n} \begin{pmatrix} \frac{1}{2} \\ 1 + \frac{1}{2} \\ \vdots \\ (n-1) + \frac{1}{2} \end{pmatrix}$
- Constante de normalización: $C(k) = \begin{cases} \sqrt{\frac{1}{n}} & k = 1 \\ \sqrt{\frac{2}{n}} & k > 1 \end{cases}$

Siendo $T = \cos(g \cdot s^t)$ la matriz resultante de aplicarle el coseno a cada elemento de la matriz $g \cdot s^t$, finalmente definimos, $\widehat{M}_{i,j} = C(i) \cdot T_{i,j}$

Para obtener una versión entera de la transformación que define la matriz M, la cual será aplicada a señales (o vectores) enteras en el rango $[0, q]$, definimos:

$$M = \left\lfloor \frac{q\widehat{M} + 1}{2} \right\rfloor \tag{1}$$

donde $\lfloor \cdot \rfloor$ indica la parte entera inferior[3]. (Es decir, escalamos los elementos de la matriz M por $q/2$ y luego redondeamos los valores.)

A.1. Extensión a 2D

Dada una matriz $B \in \mathbb{R}^{n\times n}$, podemos extender fácilmente la transformada DCT a señales de dos dimensiones. Para ello, aplicamos la transformación por filas y por columnas: $M\,B\,M^t$

[3] El redondeo de un número m pude definirse como $\lfloor m + 1/2 \rfloor$. Luego, M se define como el redondeo de $\widehat{M} \cdot (q/2)$.