

Representación Robusta de Señales con Coeficientes Poco Densos en Dicionarios Ortogonales y Redundantes

Juan Marcos Ramírez Rondón, M.S.
Tutor: José Luis Paredes Quintero, Ph.D.



Programa de Doctorado en Ciencia Aplicadas
Facultad de Ingeniería
Universidad de Los Andes



Mayo, 2015

Tabla de Contenidos

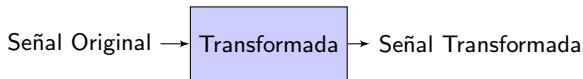
- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones
- 7 Trabajo a Futuro

Tabla de Contenidos

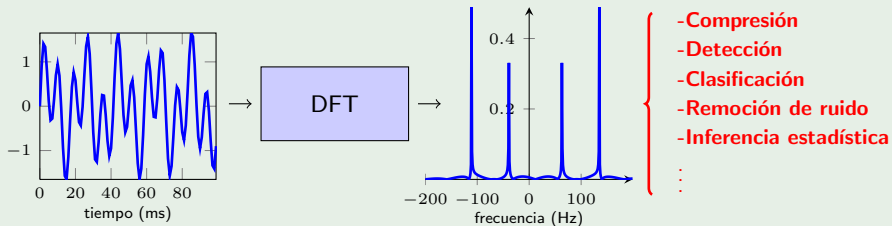
- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones
- 7 Trabajo a Futuro

Transformada

- Es un operador que representa la señal en un dominio alternativo, donde ciertas características implícitas en la señal son realzadas para un posterior análisis y procesamiento.



Transformada Discreta de Fourier



Transformada Unitaria

Desde la perspectiva matricial la transformada unitaria se puede especificar como:

$$\mathbf{X} = \Phi \mathbf{z} \quad (1)$$

- $\mathbf{z} = [z_1, \dots, z_N]^T$, es la señal N -dimensional en su dominio original
- $\mathbf{X} = [X_1, \dots, X_N]^T$, es la señal transformada
- $\Phi \in \mathbb{C}^{N \times N}$, es la matriz de transformación directa
- $\phi_\ell|_{\ell=1}^N$, son los vectores columnas de la matriz de transformación
- En el contexto de las transformadas unitarias los vectores $\phi_\ell|_{\ell=1}^N$
 - Son ortonormales y una base que expande todo el espacio \mathbb{C}^N
 - Existe una amplia gama de transformadas unitarias discretas (DFT, DCT, Wavelet Ortogonal, Hadamard, etc.)

Transformada Discreta de Fourier

- Bases de la DFT

$$\phi_\ell[n] = \frac{1}{\sqrt{N}} e^{-\frac{2i\pi}{N}(n-1)(\ell-1)}, \quad \text{para } n = 1, \dots, N$$

Transformada Inversa

La señal en su dominio convencional puede recuperarse a partir de su transformada inversa mediante:

$$z = \Psi \mathbf{X} = \sum_{\ell=1}^N X_{\ell} \psi_{\ell} \quad (2)$$

- $\Psi \in \mathbb{C}^{N \times N}$, es la matriz de transformación inversa
- $\Psi = \Phi^{-1}$, pero en el contexto de las transformadas unitarias

$$\Psi = \Phi^H \quad \text{y} \quad \Psi \Phi = \mathbf{I}$$

- $\psi_{\ell}|_{\ell=1}^N$, son los vectores columnas de la matriz de transformación inversa
- La señal original puede recuperarse como una expansión lineal de $\psi_{\ell}|_{\ell=1}^N$

Transformada Discreta de Fourier

- Bases de la IDFT

$$\psi_{\ell}[n] = \frac{1}{\sqrt{N}} e^{\frac{2i\pi}{N}(n-1)(\ell-1)}, \quad \text{para } n = 1, \dots, N$$

- Bajo la perspectiva de representación de señales en diccionarios redundantes, los dominios tiempo-frecuencia exhiben la propiedad de **incoherencia máxima**

Modelo de Adquisición de la Señal

La representación de una señal, en cualquier dominio de transformación, debe considerar un modelo apropiado del proceso de adquisición

$$\mathbf{y} = \mathbf{z} + \boldsymbol{\eta} \quad (3)$$

- $\mathbf{y} = [y_1, \dots, y_N]^T$, es la señal observada
- $\mathbf{z} = [z_1, \dots, z_N]^T$, es la señal de interés (señal sin ruido)
- $\boldsymbol{\eta}$, es el vector de **ruido aditivo** que reúne los errores debido a:
 - Acoplamiento de ruido del entorno
 - Errores en el proceso de muestreo y adquisición
- A partir del modelo de adquisición (3), se han desarrollado un importante número de métodos de estimación de los coeficientes de la señal de interés
- En este trabajo, el problema de estimación de la transformada de la señal de interés se analizará bajo la perspectiva de la estimación estadística
 - **Estimación de verosimilitud máxima (maximum likelihood estimation, MLE)**
 - **Estimación de máxima probabilidad (maximum a posteriori, MAP)**

La Transformada Bajo el Enfoque MLE

A partir de las restricciones impuestas por el modelo de adquisición

$$\mathbf{y} = \mathbf{z} + \boldsymbol{\eta},$$

el enfoque MLE se desarrolla en los siguientes términos:

- Asume que las componentes del vector de ruido aditivo $\boldsymbol{\eta}$ obedece a un modelo estadístico particular
- Estima los coeficientes que maximice la probabilidad conjunta de las componentes del ruido
- En otras palabras, determina los coeficientes resolviendo el siguiente problema de optimización:

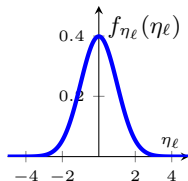
$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X}} \{p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})\} \quad (4)$$

- $p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$ es la función de verosimilitud
 - Describe la relación entre las observaciones \mathbf{y} y la señal deseada $\mathbf{z} = \boldsymbol{\Psi}\mathbf{X}$
 - Asociado al modelo estadístico que mejor describe el vector de ruido $\boldsymbol{\eta}$

La Transformada Bajo el Enfoque MLE

- Asuma que las componentes del ruido aditivo $\eta_\ell |_{\ell=1}^N$ son modeladas como muestras independientes e idénticamente distribuidas (i.i.d.) $\sim \mathcal{N}(0, \sigma_\eta^2)$

$$f_{\eta_\ell}(\eta_\ell) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{\eta_\ell^2}{2\sigma_\eta^2}}$$



- Bajo el enfoque de verosimilitud máxima, la estimación de los coeficientes de transformación se reduce a:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \frac{1}{2} \|\mathbf{y} - \Psi \mathbf{X}\|_2^2 \right\} \quad (5)$$

- La solución del problema de minimización (5) conduce a la transformada estándar del vector de observaciones ruidosas como estimador óptimo¹, esto es

$$\hat{\mathbf{X}} = \Phi \mathbf{y} \quad (6)$$

¹Peter Huber. *Robust statistics*. Springer, 2011.

Transformada Bajo la Perspectiva de Máxima Probabilidad

Considerando el modelo de adquisición

$$\mathbf{y} = \mathbf{z} + \boldsymbol{\eta},$$

el enfoque MAP se desarrolla en los siguientes términos:

- Estima los coeficientes que maximice la probabilidad del ruido, tomando en cuenta información de los coeficientes previamente conocida
- Busca la solución al siguiente problema de optimización:

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X}} \left\{ p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X}) p_{\mathbf{X}}(\mathbf{X}) \right\} \quad (7)$$

- $p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$, es la función de verosimilitud
 - Describe la relación entre las observaciones \mathbf{y} y la señal deseada \mathbf{z}
 - Asociado al modelo estadístico que mejor describe el vector de ruido $\boldsymbol{\eta}$
- $p_{\mathbf{X}}(\mathbf{X})$, es la función de distribución a priori (*prior distribution*)
 - Contiene el modelo estadístico de una característica previamente conocida del vector de coeficientes (conocimiento a priori)

Algoritmos de Contracción (*Shrinkage Algorithms*)

- Bajo el enfoque MAP, se han propuesto un conjunto de algoritmos para la estimación de los coeficientes de transformación de la señal sin ruido en el contexto de remoción de ruido en señales^{2,3,4}
 - **Signal Denoising**
- Características:
 - Asumen que las componentes de ruido aditivo son i.i.d $\sim \mathcal{N}(0, \sigma_\eta^2)$
 - La función de distribución a priori es seleccionada de modo que el vector de coeficientes tenga la menor cantidad de componentes distintos de cero \rightarrow **vector de coeficientes poco denso (sparse signal)**
 - Las condiciones anteriores conducen a la solución del siguiente problema de optimización:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \frac{1}{2} \|\mathbf{y} - \Psi \mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_p \right\} \quad \text{con } 0 \leq p \leq 1$$

²D. Donoho and J. Johnstone. "Ideal spatial adaptation by wavelet shrinkage". In: *Biometrika* 81.3 (1994).

³D. Donoho. "De-noising by soft-thresholding". In: *IEEE Trans. Inf. Theory* 41.3 (1995).

⁴D. Donoho and J. Johnstone. "Adapting to unknown smoothness via wavelet shrinkage". In: *Journal of the american statistical association* 90.432 (1995).

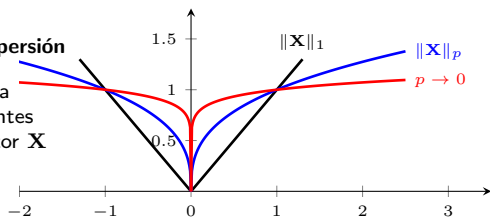
Algoritmos de Contracción (*Shrinkage Algorithms*)

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \frac{1}{2} \|\mathbf{y} - \Psi \mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_p \right\} \quad \text{con } 0 \leq p \leq 1 \quad (8)$$

- $\frac{1}{2} \|\mathbf{y} - \Psi \mathbf{X}\|_2^2$, es el término de ajuste de datos (*data fitting term*)
- $\|\mathbf{X}\|_p$, es el término que induce poca densidad (*sparsity*) en el vector de coeficientes, también llamado término de regularización

Términos que inducen dispersión

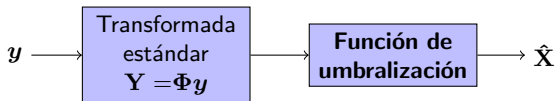
$p \rightarrow 0$, se obtiene una cuenta de los componentes distintos de cero del vector \mathbf{X}



- $\lambda \geq 0$, es el parámetro de regularización que controla el compromiso entre el término de ajuste de datos y el término que induce poca densidad

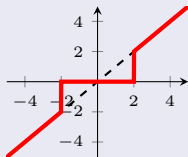
Algoritmos de Contracción (*Shrinkage Algorithms*)

- Obtiene el estimado de los coeficientes de la señal sin ruido aplicando una **función de umbralización** sobre la transformada de la señal ruidosa



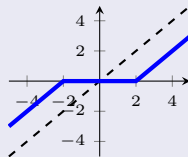
- El estimado de la señal sin ruido en su dominio original se describe mediante la transformada inversa $\hat{z} = \Psi \hat{X}$

Umbralización dura ($p = 0$)



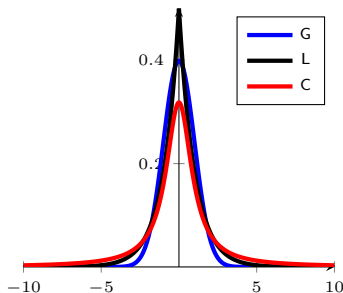
- Función no continua
- Muy sensible al ruido

Umbralización suave ($p = 1$)



- Función continua
- Estimaciones con sesgo (bias)

- Los métodos de transformación, basados en la suposición de que el ruido es gaussiano, degradan severamente su desempeño ante ruido impulsivo
- **Ruido impulsivo:** señales con picos elevados en instantes aleatorios
- Esta clase de métodos también es afectada notablemente cuando el ruido que contamina la señal está modelado por distribuciones de colas más pesadas que las exhibidas por el modelo gaussiano



- Es necesario el desarrollo de estrategias robustas de estimación de los coeficientes de transformación que mejoren su desempeño ante ruido de naturaleza impulsiva

Tabla de Contenidos

- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta**
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones
- 7 Trabajo a Futuro

La Transformada Robusta Bajo el Enfoque MAP

- Se fundamenta en resolver el siguiente problema de optimización

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \rho_{\mathbf{y}|\mathbf{X}} + \lambda \rho_{\mathbf{X}} \right\} \quad (9)$$

- donde

$$\rho_{\mathbf{y}|\mathbf{X}} = -\log \left[p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X}) \right] \qquad \rho_{\mathbf{X}} = -\log [p_{\mathbf{X}}(\mathbf{X})]$$

- $p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$, es el modelo estadístico que describe el ruido que contamina la señal de interés
 - En nuestro caso, con el objeto de minimizar el efecto del ruido impulsivo en la señal original, se modelará $p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X})$ usando distribuciones de colas más pesadas que las exhibidas por la distribución gaussiana (enfoques tradicionales)
- $p_{\mathbf{X}}(\mathbf{X})$, es el término que tiene la información *a priori* del vector de coeficientes
 - Se seleccionará $p_{\mathbf{X}}(\mathbf{X})$ de modo que la transformada tenga la menor cantidad de coeficientes distintos de cero (*sparse vector*)

La Transformada Robusta Bajo el Enfoque MAP

- Asume que el ruido aditivo es un conjunto de muestras i.i.d. que obedecen a una distribución laplaciana de media cero y factor de escala σ_η

$$p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X}) = \prod_{i=1}^N \frac{1}{2\sigma_\eta} e^{-\frac{|(\mathbf{y}-\Psi\mathbf{X})_i|}{\sigma_\eta}} \quad (10)$$

- Los coeficientes también son modelados como muestras i.i.d. de media cero y factor de escala $\sigma_{\mathbf{x}}$ (con el objeto de inducir dispersión), esto es

$$p_{\mathbf{x}}(\mathbf{X}) = \prod_{j=1}^N \frac{1}{2\sigma_{\mathbf{x}}} e^{-\frac{|x_j|}{\sigma_{\mathbf{x}}}} \quad (11)$$

- La estimación de los coeficientes de transformación, bajo el enfoque de estimación MAP, se reduce a solucionar el siguiente problema de optimización

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X}} \left\{ \prod_{i=1}^N \frac{1}{2\sigma_\eta} e^{-\frac{|(\mathbf{y}-\Psi\mathbf{X})_i|}{\sigma_\eta}} \prod_{j=1}^N \frac{1}{2\sigma_{\mathbf{x}}} e^{-\frac{|x_j|}{\sigma_{\mathbf{x}}}} \right\} \quad (12)$$

Dado que el logaritmo es una función estrictamente no decreciente

$$\hat{\mathbf{X}} = \arg \max_{\mathbf{X}} \left\{ \log \left[\prod_{i=1}^N \frac{1}{2\sigma_\eta} e^{-\frac{|(\mathbf{y} - \Psi \mathbf{X})_i|}{\sigma_\eta}} \prod_{j=1}^N \frac{1}{2\sigma_{\mathbf{X}}} e^{-\frac{|X_j|}{\sigma_{\mathbf{X}}}} \right] \right\}$$

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \left\{ \frac{1}{\sigma_\eta} \sum_{i=1}^N |(\mathbf{y} - \Psi \mathbf{X})_i| + \frac{1}{\sigma_{\mathbf{X}}} \sum_{j=1}^N |X_j| \right\}$$

Estimación robusta de los coeficientes de transformación

Se reduce a resolver la siguiente minimización

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \Psi \mathbf{X}\|_1 + \lambda \|\mathbf{X}\|_1 \} \quad (13)$$

donde

$$\lambda = \frac{\sigma_\eta}{\sigma_{\mathbf{X}}} \quad (14)$$

Solución Usando Descenso de Coordenadas

- La solución del problema previamente definido se abordará usando la estrategia de **descenso de coordenadas**
- Cada coeficiente de transformación X_k se actualiza de forma iterativa
 - Los coeficientes restantes del vector de transformación $\mathbf{X}^{(m)}$ son considerados como valores fijos y de algún modo conocidos
 - m , es el índice de iteración
- El problema de la estimación del vector de transformación reduce a determinar cada coeficiente de forma independiente, esto es

$$\hat{X}_k^{(m+1)} = \arg \min_{X_k} \left\{ \|\mathbf{y} - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)} - \psi_k X_k\|_1 + \lambda \|\mathbf{X}\|_1 \right\} \quad (15)$$

- ψ_k , es el vector columna de la matriz de transformación inversa asociado con el coeficiente X_k
- $\Psi \mathbf{X}^{(m)} - \psi_k X_k^{(m)}$, descarta la actualización de X_k —obtenida en la iteración previa— en la estimación actual

Después de aplicar algunas manipulaciones algebraicas sobre la expresión (15), tenemos

$$\hat{X}_k^{(m+1)} = \arg \min_{X_k} \left\{ \underbrace{\sum_{i=1}^N |\psi_{k,i}| \left| \frac{(\mathbf{y} - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}} - X_k \right|}_{\text{Suma de desviaciones absolutas ponderadas}} + \underbrace{\lambda |X_k|}_{\text{Regularización}} \right\}$$

Suma de desviaciones absolutas ponderadas

- $\frac{(\mathbf{y} - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}}$ para $i = 1, \dots, N$, las muestras
- $|\psi_{k,i}|$ para $i = 1, \dots, N$, las ponderaciones
- X_k , el parámetro de localización bajo el enfoque de verosimilitud máxima

Término de regularización

- **Se puede incorporar en la suma de desviaciones absolutas ponderadas**

$$\hat{X}_k^{(m+1)} = \arg \min_{X_k} \left\{ \sum_{i=1}^{N+1} |\omega_i| |Y_i - X_k| \right\} \quad (16)$$

- $Y_i = \begin{cases} \frac{(y - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}} & \text{para } i = 1, \dots, N \\ 0 & i = N + 1 \end{cases}$
- $w_i = \begin{cases} |\psi_{k,i}| & \text{para } i = 1, \dots, N \\ \lambda & i = N + 1 \end{cases}$

La solución de la minimización (16) se reduce a aplicar el operador de **mediana ponderada**, más precisamente:

$$\hat{X}_k^{(m+1)} = \text{MEDIANA} \left(w_i \diamond Y_i \Big|_{i=1}^{N+1} \right) \quad (17)$$

- Y_i para $i = 1, \dots, N + 1$, son las muestras
- ω_i para $i = 1, \dots, N + 1$, son las ponderaciones
- \diamond , es el operador de ponderación

La regularización es efectuada agregando la muestra cero ponderada por λ en el operador de mediana

- $\lambda \rightarrow \infty$, la muestra cero tendrá mayor peso, induciendo poca densidad

Algoritmo para el Cálculo de la Mediana Ponderada

Ponderaciones reales positivas

La salida del operador de mediana ponderada se obtiene mediante el siguiente algoritmo⁵:

- 1 Ordenar las muestras en orden decreciente
- 2 Sumar los pesos respectivos, hasta que la suma parcial exceda al umbral $T_h = \frac{1}{2} \sum_{i=1}^M W_i$
- 3 La mediana ponderada es la muestra asociada con el último peso incorporado en la suma parcial

Ejemplo

Muestras: $X = [1, 5, 8, 11, 2]^T$; Ponderaciones: $W = [0.1, 0.2, 0.3, 0.2, 0.1]^T$.

$$Z = \text{MEDIANA}(W \diamond X); \quad T_h = \frac{1}{2} \sum_{i=1}^5 W_i = 0.45$$

MEDIANA PODERADA					
Muestras ordenadas	11	8	5	2	1
Pesos correspondientes	0.2	0.3	0.2	0.1	0.1
Suma parcial	0.2	0.5			

Diagrama de flujo: Una flecha apunta desde el encabezado "MEDIANA PODERADA" hacia el valor 8 en la fila de "Muestras ordenadas". Otra flecha apunta desde el valor 0.5 en la fila de "Suma parcial" hacia un recuadro que contiene la expresión $0.5 > Th = 0.45$.

⁵L. Yin, R. Yang, M. Gabbouj, and Y. Neuvo. "Weighted median filters: a tutorial". In: *IEEE Trans. Circuits Syst. II* 43.3 (1996).

$$\hat{X}_k^{(m+1)} = \text{MEDIANA} \left(w_i \diamond Y_i \Big|_{i=1}^{N+1} \right)$$

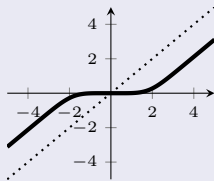
- $$Y_i = \begin{cases} \frac{(y - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}} & \text{para } i = 1, \dots, N \\ 0 & i = N + 1 \end{cases}$$
- $$w_i = \begin{cases} |\psi_{k,i}| & \text{para } i = 1, \dots, N \\ \lambda & i = N + 1 \end{cases}$$

Las componentes del vector de muestras $Y_i \Big|_{i=1}^{N+1}$ se determinan tomando en cuenta los coeficientes estimados en las iteraciones previas

Parámetro de Regularización Fijo

- Usar el mismo λ impone un efecto de atracción hacia cero en los coeficientes estimados (*shrinking*)
- Similar a la función de umbralización suave
 - Los coeficientes con valores grandes son acortados (usando relativamente el mismo sesgo)
 - Los coeficientes con valores pequeños son forzados a cero

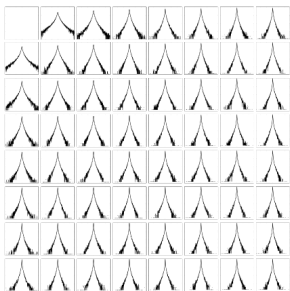
Función de umbralización promediada usando λ fijo



- Estimaciones con sesgo (bias)
- Comportamiento continuo

Bajo el enfoque MAP : $\lambda = \frac{\sigma_\eta}{\sigma_{\mathbf{x}}}$

- La función de distribución a priori $p_{\mathbf{X}}(\mathbf{X})$ modela los coeficientes como muestras i.i.d. que obedecen a una distribución laplaciana
 - Las distribuciones que modelan los coeficientes tienen un factor de escala común $\sigma_{\mathbf{x}}$
- La variabilidad en el factor de escala de los coeficientes⁶



- Ejemplos numéricos de la distribución de los coeficientes DCT en una amplia base de datos de imágenes
 - **Comportamiento de tipo laplaciano**
 - **Alta variabilidad en el factor de escala**

⁶I. Ramírez and G. Sapiro. "Universal Regularizers for Robust Sparse Coding and Modeling". In: *IEEE Trans. Image Process.* 21.9 (2012).

Parámetro de Regularización Sintonizable

- Modela los coeficientes como muestras independientes (**mas no idénticamente distribuidas**) que siguen una distribución laplaciana, esto conlleva a:
 - Parámetro de escala particular σ_{X_k} para cada coeficiente
 - Parámetro de regularización $\lambda_k = \frac{\sigma_\eta}{\sigma_{X_k}}$ para cada coeficiente
- Bajo el enfoque MAP, el σ_{X_k} óptimo se obtiene resolviendo:

$$\begin{aligned} \hat{\sigma}_{X_k} &= \arg \min_{\sigma_{X_k}} \left\{ \frac{1}{\sigma_\eta} \sum_{i=1}^N |(\mathbf{y} - \Psi \mathbf{X})_i| - N \log \left(\frac{1}{2\sigma_\eta} \right) \right. \\ &\quad \left. \dots + \sum_{j=1}^N \frac{|X_j|}{\sigma_{X_j}} - \sum_{j=1}^N \log \left(\frac{1}{2\sigma_{X_j}} \right) \right\} \end{aligned} \quad (18)$$

- Lo que conduce a

$$\sigma_{X_k} = |X_k| \quad (19)$$

$$\sigma_{X_k} = |X_k|$$

- σ_{X_k} está sujeto al conocimiento previo de X_k
- X_k es el principal objetivo en la determinación de la transformada

- Algoritmo de doble lazo anidado
 - **Lazo interno:** se actualiza λ_k y X_k
 - **Lazo externo:** se refina todo el vector de transformación
- Parámetro de regularización sintonizable

$$\lambda_k^{(m+1)} = \frac{\tau}{|X_k^{(m)}|} \quad (20)$$

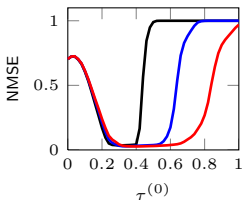
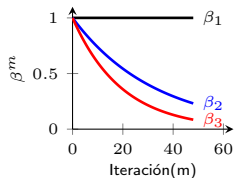
- $\tau \in \mathbb{R}^+$, está relacionado con el factor de escala del ruido ($\lambda_k = \frac{\sigma_\eta}{\sigma_{X_k}}$)

$$\hat{X}_k^{(m+1)} = \text{MEDIANA} \left(w_i \diamond Y_i \Big|_{i=1}^{N+1} \right)$$

$$Y_i = \begin{cases} \frac{(y - \Psi X^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}} & \text{para } i = 1, \dots, N \\ 0 & i = N + 1 \end{cases} \quad w_i = \begin{cases} |\psi_{k,i}| & \text{para } i = 1, \dots, N \\ \lambda_k^{(m+1)} & i = N + 1 \end{cases}$$

$$\lambda_k^{(m+1)} = \frac{\tau}{|X_k^{(m)}| + \varepsilon}; \text{ con } 0 < \varepsilon \leq 1$$

- La selección del valor óptimo de τ es una tarea compleja (ruido desconocido)
- Este problema es abordado usando un **enfoque de continuación**⁷
- El algoritmo inicia con un $\tau = \tau^{(0)}$ elevado
 - Estima los coeficientes de mayor valor
 - Favorece la dispersión
- A medida que algoritmo progresa, disminuye τ
 - Se identifican nuevos coeficientes
 - Las estimaciones de iteraciones previas son refinadas
- β es la tasa de decrecimiento de τ por iteración
 - $\beta_1 = 1$ [sin continuación], $\beta_2 = 0.97$, $\beta_3 = 0.95$



⁷E. Hale, W. Yin, and Y. Zhang. "Fixed-point continuation for ℓ_1 -minimization: Methodology and convergence". In: *SIAM Journal on Optimization* 19.3 (2008).

Selección del τ Inicial

- El algoritmo de transformada robusta inicia con un valor elevado de τ con el objeto de inducir poca densidad en el vector de coeficientes
- Bajo un examen detallado del algoritmo de transformación, es posible **determinar el mínimo valor de τ que genere el vector de coeficientes cero, $\hat{\mathbf{X}} = 0$**
- Seleccionar ese valor como $\tau = \tau^{(0)}$

$$Y_i = \begin{cases} \frac{(y - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}} & \text{para } i = 1, \dots, N \\ 0 & \text{para } i = N + 1 \end{cases} \quad w_i = \begin{cases} |\psi_{k,i}| & \text{para } i = 1, \dots, N \\ \lambda_k^{(m+1)} & \text{para } i = N + 1 \end{cases}$$

$$\hat{X}_k^{(m+1)} = \text{MEDIANA} \left(w_i \diamond Y_i \Big|_{i=1}^{N+1} \right)$$

Algoritmo de mediana ponderada

- Ordenar las muestras
- Sumar los pesos correspondientes en orden descendente
- La mediana ponderada es la muestra correspondiente al ponderación cuya suma parcial excede el umbral $T_h = \frac{1}{2} \sum_{i=1}^{N+1} w_i$ (suma parcial estrictamente mayor a T_h)

Selección del τ Inicial

$$T_h = \frac{1}{2} \left(\sum_{i=1}^N |\psi_{k,i}| + \lambda_k \right) = \frac{1}{2} (\|\boldsymbol{\psi}_k\|_1 + \lambda_k) \quad (22)$$

Propiedad: la salida del vector de coeficientes es el vector cero si y solo si

$$\lambda_k > \max \left\{ \|\boldsymbol{\psi}_\ell\|_1 \mid \ell=1 \right\} \quad (23)$$

Prueba: asuma que $\lambda_k > \max \left\{ \|\boldsymbol{\psi}_\ell\|_1 \mid \ell=1 \right\}$

$$T_h = \frac{1}{2} (\|\boldsymbol{\psi}_k\|_1 + \lambda_k) \geq \|\boldsymbol{\psi}_k\|_1 \quad (24)$$

- Si se suman todos los pesos de Y_i distintas a la muestra cero, tenemos $\|\boldsymbol{\psi}_k\|_1 \leq T_h$
- Si se incorpora el peso de la muestra cero (ubicada en cualquier posición), entonces $\lambda_k + \gamma > T_h$, para cualquier $\gamma \geq 0$
- Por tanto, la suma parcial de los pesos será estrictamente mayor que el umbral sólo cuando se adiciona el peso asociado a la muestra cero, generando $\hat{\mathbf{X}} = \mathbf{0}$ \square

El rango efectivo de λ_k para la estimación de los coeficientes distintos de cero

$$0 \leq \lambda_k \leq \max \left\{ \|\boldsymbol{\psi}_\ell\|_1 \mid \ell=1 \right\}$$

Finalmente, un valor apropiado para el τ inicial es $\tau^{(0)} = \varepsilon \max \left\{ \|\boldsymbol{\psi}_\ell\|_1 \mid \ell=1 \right\}$

Algoritmo de Transformada Robusta

```

Input      :  $\mathbf{y}, \Psi, \tau^{(0)}, itmax, tolerance, \beta$ 
 $\hat{\mathbf{X}}^{(0)} \leftarrow \mathbf{0}_N$ ;
 $m = 0$ ;
 $error = \frac{\|\mathbf{y} - \Psi \hat{\mathbf{X}}^{(0)}\|_2^2}{\|\mathbf{y}\|_2^2}$ ;
while  $m < itmax$  and  $error > tolerance$  do
     $\tau^{(m)} = \tau^{(0)} \beta^m$ ;
    for  $k = 1$  to  $N$  do
         $\lambda_k^{(m)} = \frac{\tau^{(m)}}{\varepsilon + |\hat{X}_k^{(m)}|}$ ;
        
$$Y_i = \begin{cases} \frac{(\mathbf{y} - \Psi \hat{\mathbf{X}} + \psi_k \hat{X}_k^{(m)})_i}{\psi_{k,i}} & \text{for } i = 1, \dots, N \\ 0 & i = N + 1 \end{cases}$$
;
        
$$w_i = \begin{cases} |\psi_{k,i}| & \text{for } i = 1, \dots, N \\ \lambda_k^{(m)} & i = N + 1 \end{cases}$$
;
         $\hat{X}_k^{(m+1)} = \text{MEDIAN} \left( w_i \diamond Y_i \Big|_{i=1}^{N+1} \right)$ ;
    end
     $error = \frac{\|\mathbf{y} - \Psi \hat{\mathbf{X}}^{(m+1)}\|_2^2}{\|\mathbf{y}\|_2^2}$ ;
     $m = m + 1$ ;
end
output    :  $\hat{\mathbf{X}} = [\hat{X}_1, \hat{X}_2, \dots, \hat{X}_N]^T$ 

```

Señales y matrices de transformación complejas

- Es común que las matrices de transformación contengan componentes de valor complejo (por ejemplo, en la transformada discreta de Fourier)
- Bajo la perspectiva de la transformada robusta, la estimación de los coeficientes se reduce a la mediana ponderada de muestras complejas
- Sin embargo, la mediana ponderada sobre muestras complejas no está bien definida
- Para aplicar el algoritmo propuesto se recurre a un enfoque que define la mediana ponderada sobre muestras complejas con pesos reales⁸

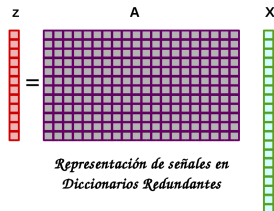
$$\hat{X}_k = \text{MEDIAN}(w_i \diamond \Re(Y_i)|_{i=1}^{N+1}) + j\text{MEDIAN}(w_i \diamond \Im(Y_i)|_{i=1}^{N+1}), \quad (25)$$

⁸S. Hoyos, Y. Li, J. Bacca, and G. Arce. "Weighted median filters admitting complex-valued weights and their optimization". In: *IEEE Trans. Signal Process.* 52.10 (2004).

Representación de Señales en Diccionarios Redundantes

$$z = \mathbf{A}\mathbf{X} \quad (26)$$

- $z \in \mathbb{R}^M$ es la señal proyectada
- $\mathbf{X} \in \mathbb{R}^N$ es el vector de coeficientes de interés
- $\mathbf{A} \in \mathbb{R}^{M \times N}$ es la matriz de proyección o **diccionario** con $N \geq M$



El Diccionario

- Es una colección de formas de onda parametrizadas, generalmente ajustada a la aplicación de interés
 - Diccionarios Ortogonales o Matrices de Transformación Lineal ($N = M$)
 - Diccionarios de Frecuencia Redundantes (DFT, DCT)
 - Diccionarios holográficos (muestreo comprimido)
 - Megadiccionarios (concatenación de diccionarios ortogonales)
- $\mathbf{a}_\ell \big|_{\ell=1}^N$ son los vectores columnas de \mathbf{A}
 - Señales discretas de dimensión M llamadas también **átomos**

Diccionarios Redundantes $N > M$

- No son invertibles, y el sistema de ecuaciones $z = \mathbf{A}\mathbf{X}$ tiene múltiples soluciones
- Dependiendo de la restricción impuesta se obtendrá una solución particular

Solución de Magnitud Mínima

Selecciona aquella solución que minimice la magnitud en el vector de coeficientes, esto es

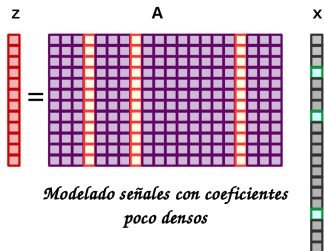
$$\min_{\mathbf{X}} \|\mathbf{X}\|_2 \quad \text{sujeto a } z = \mathbf{A}\mathbf{X}$$

La solución es obtenida usando la inversa generalizada de la matriz \mathbf{A} , i.e.,

$$\hat{\mathbf{X}} = \mathbf{A}^T (\mathbf{A}\mathbf{A}^T)^{-1} z$$

Limitación: no preserva la poca densidad en el vector de coeficientes

Representación de Señales con Coeficientes Poco Densos en Dicciones Redundantes



- El vector de coeficientes \mathbf{X} contiene pocos elementos distintos de cero (ubicados en posiciones aleatorias)
- La señal proyectada z se construye como **la combinación lineal de pocos átomos del diccionario**

En esta situación particular se busca resolver el siguiente problema de optimización

$$\min_{\mathbf{X}} \|\mathbf{X}\|_0 \quad \text{sujeto a} \quad z = \mathbf{A}\mathbf{X} \quad (27)$$

En el caso de que la señal proyectada esté contaminada de ruido aditivo, $\mathbf{y} = \mathbf{A}\mathbf{X} + \boldsymbol{\eta}$, la optimización se ajusta de la siguiente forma

$$\min_{\mathbf{X}} \|\mathbf{X}\|_0 \quad \text{sujeto a} \quad \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_p < \epsilon \quad (28)$$

donde la norma $\|\cdot\|_p$ depende del modelo estadístico asumido para el ruido

Algoritmos de Aproximación

Asumiendo que las componentes $\eta_i|_{i=1}^M$ son muestras i.i.d. que siguen una distribución gaussiana, el problema de estimación se reduce a:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_0 \} \quad (29)$$

- La minimización (29) **es un problema combinatorial NP-completo**⁹ cuyo costo computacional es demasiado elevado aún para señales de dimensiones modestas
- Para resolver el problema de estimación del vector de coeficientes \mathbf{X} se recurre a aproximaciones

Métodos de Relajación Convexa

- Suaviza la pseudonorma $\|\cdot\|_0$
- Implementa técnicas de optimización convexa

Métodos de Búsqueda Voraz

- Matching pursuit algorithms
- Determina un coeficiente en cada iteración

⁹A. Bruckstein, D. Donoho, and M. Elad. "From sparse solutions of systems of equations to sparse modeling of signals and images". In: *SIAM review* (2009).

Algoritmos de Relajación Convexa

En vez de resolver

$$\arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_0 \}$$



Resuelve

$$\arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2 + \lambda \|\mathbf{X}\|_1 \}$$

- Reemplaza la norma $\|\cdot\|_0$ por la norma $\|\cdot\|_1$
- El problema de optimización definido es convexo
- Este problema también es conocido como *Basis Pursuit Denoising (BPDN)*
- Puede resolverse usando diversas técnicas eficientes de optimización
 - Método de punto interior¹⁰
 - Algoritmos de contracción iterativos^{11,12}
- $\|\mathbf{y} - \mathbf{A}\mathbf{X}\|_2^2$ se basa en la suposición de que el ruido sigue un modelo gaussiano

¹⁰S. Chen, D. Donoho, and M. Saunders. "Atomic decomposition by basis pursuit". In: *SIAM journal on scientific computing* (1998).

¹¹M. Elad. "Why simple shrinkage is still relevant for redundant representations". In: *IEEE Trans. Inf. Theory* (2006).

¹²M. Zibulevsky and M. Elad. "L1-L2 Optimization in Signal and Image Processing". In: *IEEE Signal Process. Mag.* (2010).

Algoritmos de Búsqueda Voraz

- Los métodos basados en el algoritmo de **matching pursuit**¹³ (MP) estiman un coeficiente a la vez de forma iterativa
 - Selecciona el átomo que tiene la mayor correlación con un vector de residuo (vector de observaciones en la primera iteración)
 - La magnitud del coeficiente correspondiente se determina como la correlación entre el átomo seleccionado y el vector de residuo
 - El algoritmo se detiene cuando el error $\|y - AX\|_2^2$ tenga un valor menor que un umbral previamente establecido
- El algoritmo de **orthogonal MP**¹⁴ (OMP) es una versión mejorada que evalúa los coeficientes en cada iteración usando mínimos cuadrados

Otros algoritmos basados en MP

- *Regularized OMP*
- *Stagewise OMP*
- *Subspace MP*
- *Compressive Samplig MP (CoSaMP)*

El uso de la correlación lineal se deriva de la suposición de que el ruido que contamina las proyecciones sigue un modelo gaussiano

¹³S. G Mallat and Z. Zhang. "Matching pursuits with time-frequency dictionaries". In: *Signal Processing, IEEE Transactions on* (1993).

¹⁴Y. Pati, R. Rezaiifar, and PS Krishnaprasad. "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition". In: *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. 1993.

- Es ampliamente conocido que los métodos basados en la suposición de que el ruido es de tipo gaussiano degradan notablemente su desempeño ante ruido impulsivo
- Es necesario el desarrollo de técnicas robustas de representación de señales que recupere una versión confiable de la señal en presencia de ruido impulsivo

Algoritmo Robusto de Representación

- Reemplaza $\|\mathbf{y} - \mathbf{AX}\|_2^2$ por $\|\mathbf{y} - \mathbf{AX}\|_1$, esto conduce a

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{AX}\|_1 + \lambda \|\mathbf{X}\|_0 \} \quad (30)$$

- Emerge cuando el ruido aditivo es modelado como muestras i.i.d. que obedecen a una distribución laplaciana de media cero
- (30) es un problema de optimización combinatorial NP-completo^a
- Suaviza el término de regularización usando $\|\mathbf{x}\|_1$, obteniéndose

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{AX}\|_1 + \lambda \|\mathbf{X}\|_1 \} \quad (31)$$

^aX. Huo and J. Chen. "Hardness on Numerical Realization of Some Penalized Likelihood Estimators". In: *Industrial and Technology Statistics Group, Georgia Inst. Technol., Atlanta, GA, Tech. Paper 8* (2006).

Algoritmo de Robusto de Representación de Señales Poco Densas

Busca resolver:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_1 + \lambda \|\mathbf{X}\|_1 \}$$

- Con el objeto de obtener la representación robusta de una señal con coeficientes poco densos en diccionarios redundantes, se implementa un algoritmo iterativo basado en el operador de mediana ponderada
- El algoritmo a implementar sigue un desarrollo idéntico al algoritmo de transformada robusta basado en la mediana ponderada
 - Enfoque de descenso de coordenadas
 - Parámetro de regularización sintonizable
 - Estrategia de Continuación
- Considera las dimensiones tanto de los vectores como de los diccionarios en la especificación del algoritmo

Tabla de Contenidos

- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares**
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones
- 7 Trabajo a Futuro

Simulaciones

Transformada Discreta de Fourier en Ruido Impulsivo

Los resultados del desempeño del algoritmo de transformada robusta propuesto se reportaron en¹⁵

Señal discreta

- longitud $N = 256$
- Se construye como la suma de M bases de la matriz de transformación inversa

$$z = \sum_{i=1}^M \psi_{n_i}$$

- $n_i \in \{1, \dots, 256\}$, son índices generados aleatoriamente

Ruido Aditivo

- Se uso el siguiente modelo de ruido impulsivo¹⁶

$$\eta = \sigma_1 \mathcal{N}(0, 1) + (\sigma_2 \mathcal{N}(0, 1))^3 + j \left(\sigma_1 \mathcal{N}(0, 1) + (\sigma_2 \mathcal{N}(0, 1))^3 \right)$$

Parámetros de inicialización del algoritmo

- $\varepsilon = 0.01$, $\text{itmax} = 32$, $\beta = 0.98$, $\text{tolerance} = 1 \times 10^{-6}$
- $\tau^{(0)} = \varepsilon \max(\|\psi_k\|_1 |_{k=1}^N)$

¹⁵J. Ramirez and J. Paredes. "Robust Transforms Based on the Weighted Median Operator". In: *IEEE Signal Process. Lett.* (2015).

¹⁶I. Djurovic, L. Stankovic, and J. Bohme. "Robust L-estimation based forms of signal transforms and time-frequency representations". In: *IEEE Trans. Signal Process.* 51.7 (2003).

Simulaciones

Transformada Discreta de Fourier en Ruido Impulsivo

- Cada punto en las curvas de resultados es el promedio de 1000 realizaciones del experimento respectivo

Métrica de desempeño

- Error medio cuadrático normalizado (NMSE, *normalized mean square error*)

$$\text{NMSE}(dB) = 10 \log_{10} \left[\frac{\|\mathbf{X} - \hat{\mathbf{X}}\|_2^2}{\|\mathbf{X}\|_2^2} \right] \quad (32)$$

- \mathbf{X} , es la transformada de la señal sin ruido
- $\hat{\mathbf{X}}$, es la transformada estimada mediante diversos métodos de transformación

Métodos implementados

- **DFT**: La transformada discreta de Fourier estándar
- **L-based(1)**: transformada basada en estimación- L con $\alpha = 0.5^{17}$
- **L-based(2)**: transformada basada en estimación- L y muestreo comprimido¹⁸
- **WM-based**: transformada basada en mediana ponderada (propuesta)

¹⁷I. Djurovic, L. Stankovic, and J. Bohme. "Robust L-estimation based forms of signal transforms and time-frequency representations". In: *IEEE Trans. Signal Process.* 51.7 (2003).

¹⁸L. Stankovic, S. Stankovic, I. Orovic, and M. Amin. "Robust Time-Frequency Analysis Based on the L-Estimation and Compressive Sensing". In: *IEEE Signal Process. Lett.* 20.5 (2013).

Simulaciones

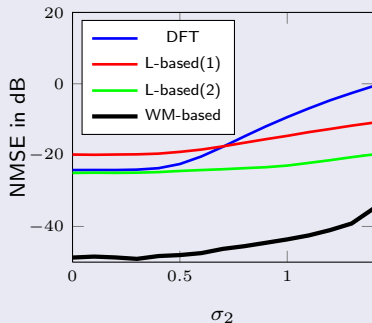
Transformada Discreta de Fourier en Ruido Impulsivo

$$z = \sum_{i=1}^M \psi_{n_i}$$

$$\eta = \sigma_1 \mathcal{N}(0, 1) + (\sigma_2 \mathcal{N}(0, 1))^3 + j (\sigma_1 \mathcal{N}(0, 1) + (\sigma_2 \mathcal{N}(0, 1))^3)$$

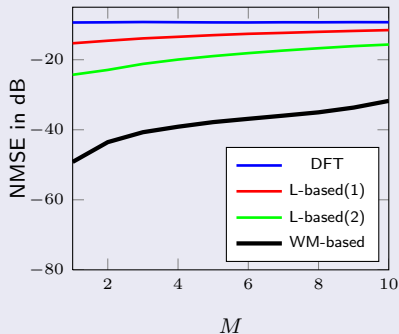
NMSE versus σ_2

● $M = 2, \sigma_1 = 0.7$



NMSE versus M

● $\sigma_1 = 0.7, \sigma_2 = 1.0$



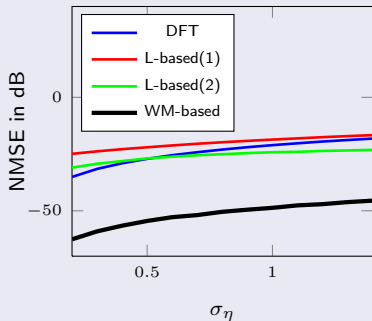
Simulaciones

Transformada Discreta de Fourier en Ruido Laplaciano

$$\eta_i \sim \frac{1}{2\sigma_\eta} e^{-\frac{|\eta_i|}{\sigma_\eta}}$$

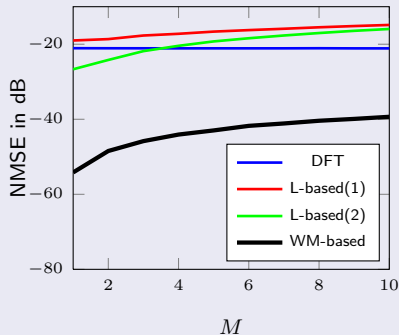
NMSE versus σ_η

• $M = 2$



NMSE versus M

• $\sigma_\eta = 1.0$



Aplicación

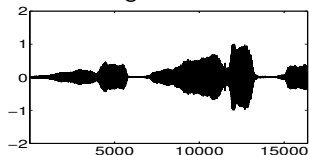
Remoción de ruido en audio (*audio denoising*)

- Las señales de audio son afectadas con frecuencia por distorsiones locales y ruido de tipo impulsivo (clicks)
- Las técnicas tradicionales intentan atenuar el ruido, preservando con la mayor fidelidad posible la información de interés
 - 1 Calculan el espectrograma de la señal de audio, donde las transformadas locales tienen, en general, una representación poco densa
 - 2 Aplican una operación sobre los coeficientes del espectrograma para atenuar el ruido
 - 3 Mediante un proceso de transformación inversa recuperan un estimado de la señal de audio
- **No obstante, están fundamentadas en la suposición de que el ruido es Gaussiano**
- En nuestro caso, se implementan los algoritmos de transformación robustos con el fin de obtener las transformadas locales del espectrograma de la señal ruidosa
 - Remoción de ruido implícito
 - El proceso de transformación inversa se calcula sobre el espectrograma obtenido, con el objeto de evaluar la calidad la señal recuperada

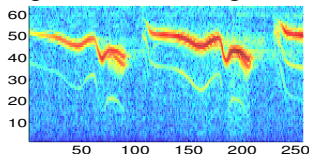
Aplicación

Remoción de ruido en audio (*audio denoising*)

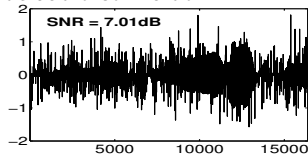
- Señal de audio original



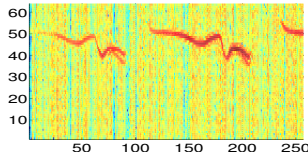
- Espectrograma del audio original



- Señal de audio con ruido



- Espectrograma del audio con ruido



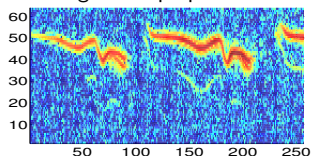
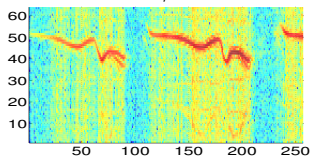
Características del ruido

- Ruido impulsivo e-contaminated o p-Gaussian
 - $f_{\eta}(\eta) = (1 - \epsilon)\mathcal{N}(0, \sigma_{\epsilon}) + \epsilon\mathcal{N}(0, 50\sigma_{\epsilon})$
 - $\epsilon = 0.05$ (nivel de impulsividad 5%), $\sigma_{\epsilon} = 0.01$

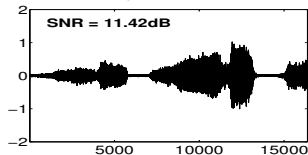
Aplicación

Remoción de ruido en audio (*audio denoising*)

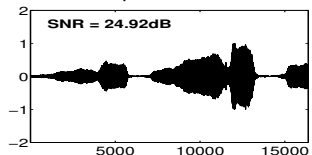
- El espectrograma se construyó usando la transformada basada en estimación-L con un parámetro de descarte $\alpha = 0.20$ ¹⁹ y la transformada robusta basada en la mediana ponderada (propuesta)
- Espectrograma de la señal ruidosa basado en estimación-L, $\alpha = 0.20$
- Espectrograma de la señal ruidosa basado en el algoritmo propuesto



- Señal de audio recuperada



- Señal de audio recuperada



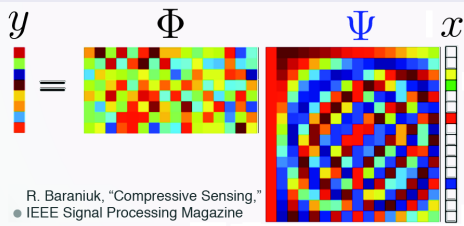
¹⁹I. Djurovic, L. Stankovic, and J. Bohme. "Robust L-estimation based forms of signal transforms and time-frequency representations". In: *IEEE Trans. Signal Process.* 51.7 (2003).

Aplicación

Muestreo Comprimido (*Compressive Sensing*)

- Es un paradigma alternativo de adquisición de señales que captura los datos en formato comprimida

$$\mathbf{y} = \Phi\Psi\mathbf{x} + \boldsymbol{\eta} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta} \quad (33)$$



- $\mathbf{y} \in \mathbb{R}^M$ es el vector de mediciones
- $\mathbf{x} \in \mathbb{R}^N$ es el vector de coeficientes de interés con $N > M$
 - Es un vector poco denso cuya cardinalidad $S \ll N$
- $\boldsymbol{\eta}$ es el vector de ruido aditivo
- $\mathbf{A} = \Phi\Psi \in \mathbb{R}^{M \times N}$ es la matriz de mediciones (matriz holográfica)
 - Debe satisfacer la propiedad de isometría restringida (*Restricted Isometry Property, RIP*)

- El vector de coeficientes se obtiene resolviendo:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_p + \lambda \|\mathbf{x}\|_0 \}$$

Simulaciones

Muestreo Comprimido (*Compressive Sensing*)

Los resultados generados por algoritmo de representación robusta de señales con coeficientes poco densos en diccionarios redundantes se reportaron en²⁰

Vector de coeficientes sin ruido

- Longitud $N = 256$, cardinalidad $S = 25$
- Los coeficientes distintos de cero $\sim \mathcal{N}(0, 1)$ en posiciones aleatorias

Matrices de proyección

- Se estimarán señales poco densas en el dominio canónico $\Psi = \mathbf{I}$
- Φ es un matriz de dimensiones 256×512 cuyas componentes $\sim \mathcal{N}(0, 1)$
- Las columnas de Φ se normalizan de modo que $\|\phi_k\|_2^2 = 1$ para $k = 1, \dots, N$

Parámetros de inicialización del algoritmo

- $\varepsilon = 0.01$, $\text{itmax} = 100$, $\beta = 0.98$, $\text{tolerance} = 1 \times 10^{-6}$
- $\tau^{(0)} = \varepsilon \max(\|\psi_k\|_1 |_{k=1}^N)$

²⁰J. Ramirez and J. Paredes. "Robust Sparse Signal Recovery Based on Weighted Median Operator". In: *Acoustics, Speech and Signal Processing, 2014. ICASSP 2014 Proceedings. 2014 IEEE International Conference on* IEEE, 2014. 

Simulaciones

Muestreo Comprimido (*Compressive Sensing*)

- Cada punto en las curvas de resultados es el promedio de 1000 realizaciones del experimento correspondiente

Métrica de desempeño

- Error medio cuadrático normalizado (NMSE, *normalized mean square error*)

$$\text{NMSE}(dB) = 10 \log_{10} \left[\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}{\|\mathbf{x}\|_2^2} \right] \quad (34)$$

Métodos implementados

- **ROMP**: OMP Regularizado (*Regularized OMP*)²¹
- **l1-ls**: Mínimos cuadrados regularizado por la norma ℓ_1 ²²
- **LAD-lasso**: Regresión robusta a través de LAD-LASSO²³
- **WM-HT**: Mediana ponderada y umbralización dura²⁴
- **Proposed**: Algoritmo propuesto

²¹D. Needell and R. Vershynin. "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit". In: *Foundations of computational mathematics* (2009).

²²S.-J Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. "An Interior-Point Method for Large-Scale l1-Regularized Least Squares". In: *IEEE J. Sel. Topics Signal Process.* (2007).

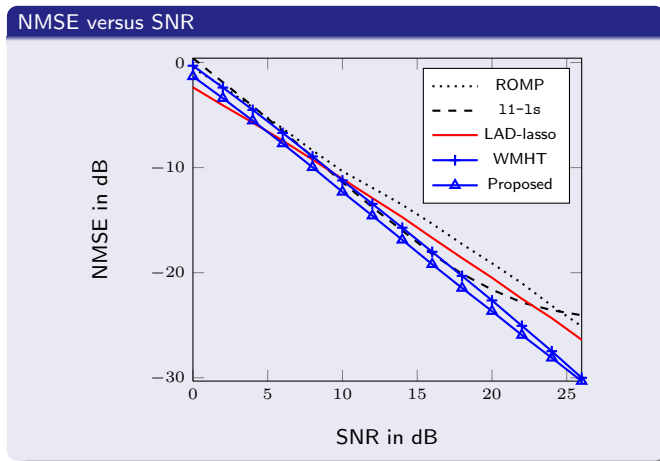
²³H. Wang, G. Li, and G. Jiang. "Robust regression shrinkage and consistent variable selection through the LAD-Lasso". In: *Journal of Business & Economic Statistics* (2007).

²⁴J. L. Paredes and G. Arce. "Compressive sensing signal reconstruction by weighted median regression estimates". In: *IEEE Trans. Signal Process.* (2011).

Simulaciones

Recuperación de la señal contaminada de ruido gaussiano

$$\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_{\eta}^2); \quad \text{SNR} = \frac{(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x})}{M\sigma_{\eta}}$$

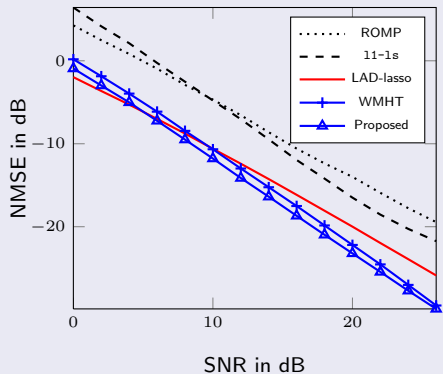


Simulaciones

Recuperación de la señal contaminada de ruido impulsivo

$$\boldsymbol{\eta} \sim (1 - \epsilon)\mathcal{N}(0, \sigma_1) + \epsilon\mathcal{N}(0, 100\sigma_1); \quad \text{RSNR}[\text{dB}] = -\text{NMSE}[\text{dB}]$$

Nivel de impulsividad fija $\epsilon = 3\%$



Varianza fija SNR = 12dB

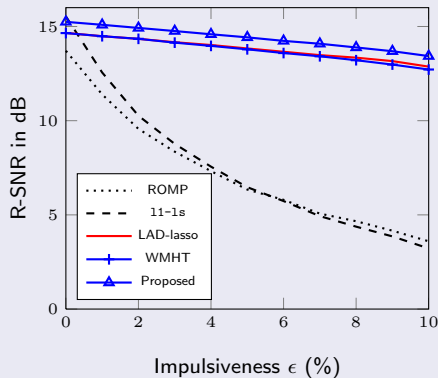


Tabla de Contenidos

- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales**
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones
- 7 Trabajo a Futuro

Modelado del Ruido Mediante la Distribución de Cauchy

- En general, se trabaja con el siguiente modelo de adquisición

$$\mathbf{y} = \mathbf{A}\mathbf{X} + \boldsymbol{\eta} \quad (35)$$

- $\mathbf{X} \in \mathbb{R}^N$, $\mathbf{y} \in \mathbb{R}^M$, $\mathbf{A} \in \mathbb{R}^{N \times M}$, con $N \geq M$
- Bajo el enfoque MAP, el ruido aditivo se puede modelar como muestras i.i.d. que siguen una distribución de Cauchy con parámetro de dispersión K

$$p_{\mathbf{y}|\mathbf{X}}(\mathbf{y}|\mathbf{X}) = \prod_{i=1}^M \frac{K}{\pi} \left[\frac{1}{K^2 + ((\mathbf{y} - \mathbf{A}\mathbf{X})_i)^2} \right] \quad (36)$$

- Los coeficientes de representación se modelan como muestras i.i.d. que siguen una distribución laplaciana de media cero y factor de escala $\sigma_{\mathbf{x}}$, esto es

$$p_{\mathbf{x}}(\mathbf{X}) = \prod_{j=1}^N \frac{1}{2\sigma_{\mathbf{x}}} e^{-\frac{|x_j|}{\sigma_{\mathbf{x}}}} \quad (37)$$

Modelado del Ruido Mediante la Distribución de Cauchy

- La estimación de los coeficientes de representación se reduce a solucionar el siguiente problema de optimización

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \{ \|\mathbf{y} - \mathbf{A}\mathbf{X}\|_{LL} + \lambda \|\mathbf{X}\|_1 \} \quad (38)$$

- donde

$$\|\mathbf{y} - \mathbf{A}\mathbf{X}\|_{LL} = \sum_{i=1}^M \log [K^2 + ((\mathbf{y} - \mathbf{A}\mathbf{X})_i)^2] \quad (39)$$

- Se han reportado algoritmos de representación basados en el operador **MYRIAD**²⁵
 - Estos algoritmos están diseñados para operar sobre proyecciones obtenidas con una matriz con componentes aleatorios que siguen una distribución Cauchy
 - No consideran matrices unitarias ni matrices aleatorias con componentes distintos a los entregados por la distribución de Cauchy
- El objetivo es desarrollar un algoritmo de representación que minimize (38), y que considere matrices de transformación y una gama amplia de diccionarios**

²⁵A. Ramirez, G. Arce, D. Otero, J. Paredes, and B. Sadler. "Reconstruction of Sparse Signals From Dimensionality-Reduced Cauchy Random Projections". In: *IEEE Trans. Signal Process.* (2012).

- Bajo un examen detallado de la estrategia de descenso de coordenadas, se observó que la estimación de cada coeficiente conlleva a la solución de la siguiente minimización:

$$\hat{X}_k = \arg \min_{X_k} \left\{ \sum_{i=1}^N \rho \left[\psi_{k,i} \left(\frac{(\mathbf{y} - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}} - X_k \right) \right] + \lambda |X_k| \right\},$$

- $\frac{(\mathbf{y} - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})_i}{\psi_{k,i}}$ son las muestras $(Y_i |_{i=1}^M)$
 - $\psi_{k,i} |_{i=1}^M$ son las ponderaciones $(\omega_i |_{i=1}^M)$
 - X_k es el parámetro de localización
 - $\rho(\cdot)$ depende del modelo estadístico asumido para el ruido aditivo
- Cuando las componentes del ruido aditivo se modelan como muestras independientes que obedecen a una distribución de Cauchy con parámetro de dispersión K , tenemos

$$\rho(X_k, Y_i |_{i=1}^M, \omega_i |_{i=1}^M, K) = \sum_{i=1}^M \log [K^2 + \omega_i (Y_i - X_k)^2] \quad (40)$$

donde el estimado de verosimilitud máxima del parámetro de localización X_k se reduce al operador de **myriad ponderado** (fundamento de los filtros myriad ponderados)

- La estimación de cada coeficiente considera la última actualización del vector de transformación

$$\hat{X}_k = \arg \min_{X_k} \left\{ \sum_{i=1}^N \rho \left[\psi_{k,i} \left(\frac{(\mathbf{y} - \Psi \mathbf{X}^{(m)} + \psi_k X_k^{(m)})}{\psi_{k,i}} - X_k \right) \right] + \lambda |X_k| \right\},$$

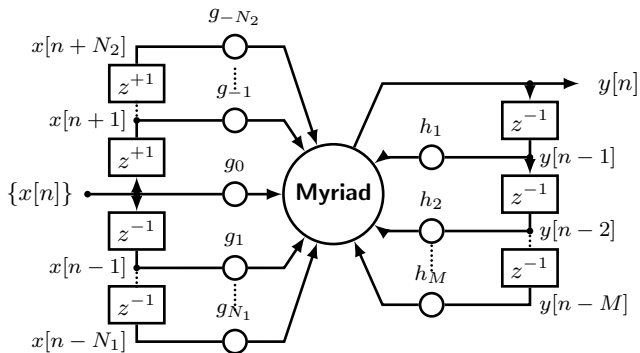
Son los únicos términos que cambian en la actualización de las muestras

Las muestras afectadas por este conjunto pueden considerarse las salidas previas de un sistema (p.e., un filtro)

Las muestras modificadas por este grupo pueden considerarse las entradas actuales del mismo sistema

- El proceso de estimación de los coeficientes, bajo la estrategia de descenso de coordenadas, tiene estructura recursiva implícita
- En este marco, se observó que estructuras recursivas de filtrado basadas en el operador de myriad ponderado no estaban definidas

Durante la etapa de investigación, se desarrolló la propuesta de filtros recursivos de frecuencia selectiva basados en el operador myriad ponderado



● Filtro Recursivo Myriad Ponderado (*Recursive Weighted Myriad Filter*)

- Considera las entradas y las salidas previas para estimar la salida actual del filtro
- Asume que las entradas y las salidas previas son afectadas por ruido de tipo Cauchy
- Se propone un filtro normalizado y un filtro escalado (amplifica la salida)
- Este filtro incluye como casos especiales el filtro lineal IIR, el filtro no recursivo de myriad ponderado, y el filtro de moda muestral

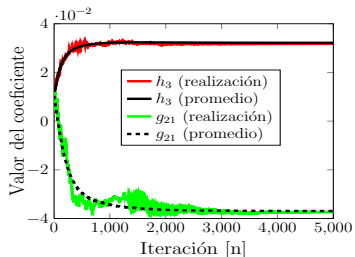
● Filtro Recursivo Myriad Híbrido (*Recursive Hybrid Myriad Filter*)

- Asume que las entradas están contaminadas de ruido tipo Cauchy, y que las salidas previas están afectadas por ruido gaussiano
- Fundamentado en el hecho que el estimado myriad se aproxima asintóticamente a una distribución normal
- A partir de un análisis detallado de la función de coste objetivo, se determinaron algunas propiedades del filtro híbrido
- Se desarrolló un algoritmo para el cálculo rápido de la salida filtro bajo el enfoque de búsqueda de puntos fijos
- Se propone un filtro normalizado y un filtro escalado (amplifica la salida)

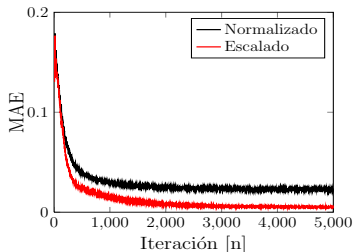
● Algoritmos Adaptativos de Diseño

- Se desarrollaron algoritmos adaptativos para el diseño de los filtros propuestos
- Intentan minimizar el error absoluto medio (*mean absolute error*, MAE) entre la salida del filtro y una señal deseada
- Los coeficientes se determinan mediante la aplicación de un algoritmo iterativo de descenso más pronunciado (*steepest descent*)
- El gradiente es determinado usando el enfoque de formulación de la ecuación de error (*equation error formulation approach*)

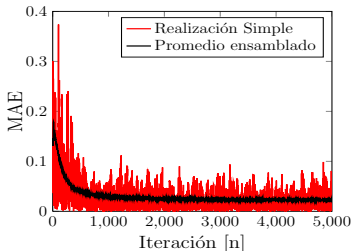
Desempeño de los Algoritmos Adaptativos de Diseño



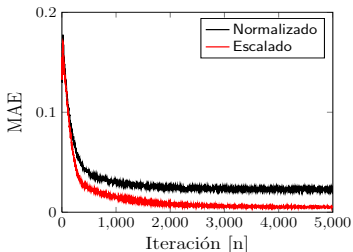
Comportamiento de los coeficientes



Filtro Recursivo Myriad Ponderado



Filtro Recursivo Myriad Ponderado



Filtro Recursivo Myriad Híbrido

Desempeño de los Filtros de Tipo Paso Banda

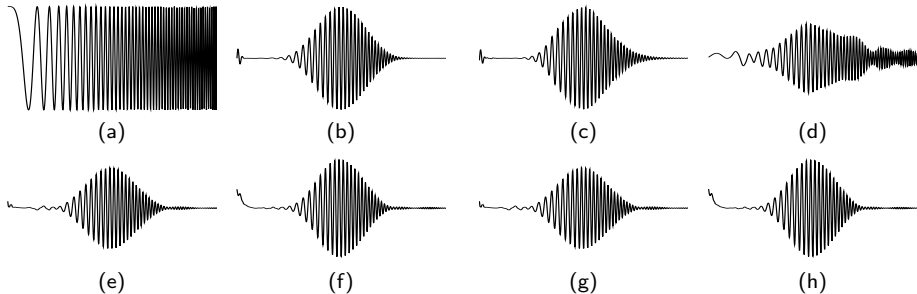


Figura: (a) señal de prueba [chirp], (b) filtro FIR [señal deseada], (c) filtro IIR, (d) filtro myriad ponderado no recursivo, MAE = 0.1593, (e) filtro recursivo de myriad ponderado, MAE = 0.0395, (f) filtro recursivo de myriad ponderado escalado, MAE = 0.0204, (g) filtro recursivo myriad híbrido, MAE = 0.0395, (h) filtro recursivo myriad híbrido escalado, MAE = 0.0206.

Desempeño de los Filtros Ante Ruido Impulsivo

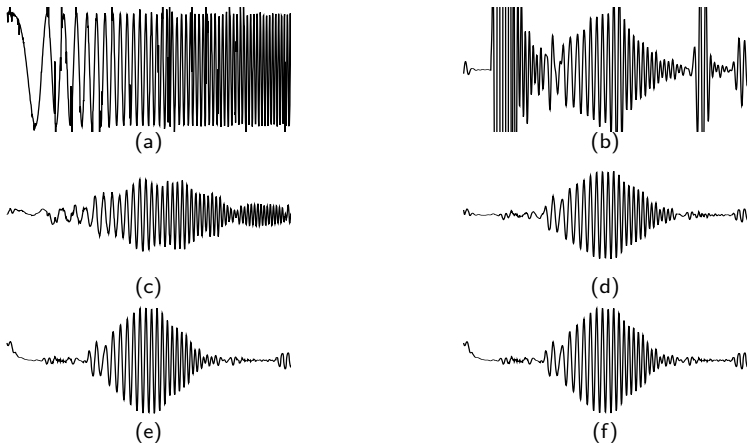


Figura: (a) señal ruidosa, (b) filtro IIR, MAE = 0.5108, (c) filtro myriad ponderado no recursivo, MAE = 0.1658, (d) filtro recursivo myriad ponderado, MAE = 0.0576, (e) filtro recursivo myriad ponderado escalado, MAE = 0.0529, (f) filtro recursivo myriad híbrido escalado, MAE = 0.0529.

Desempeño del Filtrado Paso Bajo y del Escalamiento

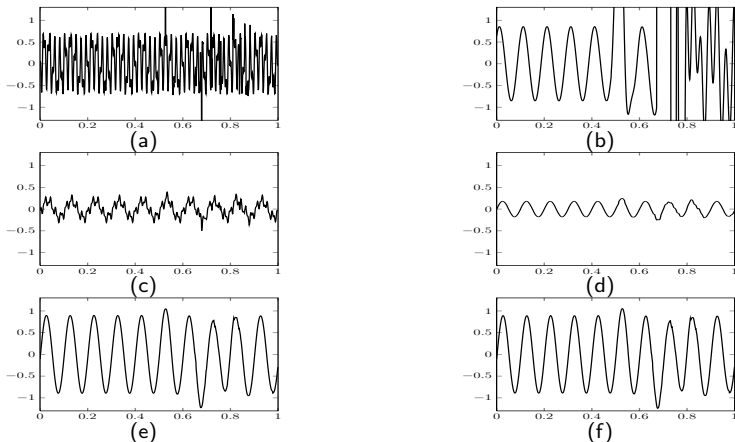


Figura: (a) señal ruidosa, (b) filtro IIR, (c) filtro myriad ponderado no recursivo, (d) filtro recursivo myriad ponderado, (e) filtro recursivo myriad ponderado escalado, (f) filtro recursivo myriad híbrido escalado.

Tabla de Contenidos

- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación**
- 6 Conclusiones
- 7 Trabajo a Futuro

Trabajos Publicados y Trabajos en Proceso de Evaluación

● Trabajos publicados

- Juan Ramírez and José Paredes. “Robust Sparse Signal Recovery Based on Weighted Median Operator”. In: *Acoustics, Speech and Signal Processing, 2014. ICASSP 2014 Proceedings. 2014 IEEE International Conference on*. IEEE. 2014, pp. 1050–1054
- Juan Ramírez and José Paredes. “Robust Transforms Based on the Weighted Median Operator”. In: *Signal Processing Letters, IEEE 22.1 (2015)*, pp. 120–124

● Trabajos en proceso de evaluación

- Juan Ramírez and José Paredes. “Recursive Weighted Myriad Filters and Their Optimizations”. In: *IEEE Trans. Signal Process.* (submitted for publication)

● Otras publicaciones

- Daniel García, José Paredes, Juan Ramírez, Luis Vega, and Ronal Celaya. “Síntesis en FPGA de filtros de mediana ponderada reconfigurables de frecuencia selectiva”. In: *Memorias del XI CIMENICS 2011*. 2011, PS25–PS30
- Ronal Celaya, Luis Vega, José Paredes, and Juan Ramírez. “Implementación de un Arquitectura Basada en Regresión de Mediana Ponderada para la Reconstrucción de Señales Poco Densas Usando Computación Reconfigurable”. In: *Memorias del XI CIMENICS 2011*. 2011, PS1–PS6
- Yanet Moreno, Juan Ramírez, Ícaro Alzúru, César Fuguet, and Gerard Páez. “Diseño de la Arquitectura de un microprocesador RISC modular de 8 bits”. In: *Memorias del XI CIMENICS 2011*. 2011, E7–E12

Tabla de Contenidos

- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones**
- 7 Trabajo a Futuro

Conclusiones

- El presente trabajo se ha concentrado en el estudio de la representación robusta de señales poco densas en diccionarios redundantes, bajo la perspectiva de estimación de máxima probabilidad (MAP)
- En este contexto, se desarrolló un método de transformada robusta basado en el operador de mediana ponderada
- El problema de estimación robusta de la transformada se desarrolló asumiendo:
 - Las componentes del vector de ruido pueden describirse como muestras i.i.d. que obedecen a una distribución laplaciana
 - Los coeficientes a determinar se pueden modelar como muestras independientes que siguen una distribución laplaciana
- El problema de estimación de los coeficientes se aborda usando:
 - Enfoque de descenso de coordenadas (*coordinate descent*)
 - Definición de un parámetro de regularización adaptativo cuyo valor es ajustado mediante una estrategia de continuación
- El algoritmo propuesto se evaluó mediante simulaciones y en la solución de un problema en aplicaciones prácticas (remoción de ruido en audio y muestreo comprimido)

Conclusiones

- En el marco de las técnicas robustas de procesamiento de señales, se propone una nueva clase de filtros recursivos basados en el operador de myriad
- Filtro Recursivo Myriad Ponderado
 - Incluye como casos especiales el filtro lineal IIR, el filtro de moda muestral, y el filtro (no recursivo) myriad ponderado
 - Se desarrolló una versión normalizada y una versión escalada
- Filtro Recursivo Myriad Híbrido
 - Se basa en el hecho que la salida de un filtro myriad tiende asintóticamente a una distribución normal
 - Se diseñó un algoritmo de cálculo rápido de la salida de filtro híbrido basado en el enfoque de búsqueda de puntos fijos
 - Se desarrolló una versión normalizada y una versión escalada
- Se desarrollaron los algoritmos adaptativos para la obtención de los coeficientes óptimos de los filtros recursivos propuestos basados en el operador de myriad

Tabla de Contenidos

- 1 Transformadas Unitarias
- 2 El Algoritmo de Transformada Robusta
- 3 Resultados Preliminares
- 4 Trabajo Adicional en Procesamiento Robusto de Señales
- 5 Trabajos Publicados y Trabajos en Proceso de Evaluación
- 6 Conclusiones
- 7 Trabajo a Futuro**

Trabajo a Futuro

- Desarrollar técnicas que permita la aceleración de los algoritmos de representación propuestos
- Considerar otras distribuciones de colas pesadas en el modelado del ruido aditivo, con el objeto de desarrollar nuevas técnicas robustas de representación de señales con coeficientes poco densos en diccionarios redundantes
- Incorporar los algoritmos de representación propuestos en diversas aplicaciones de procesamiento de señales e imágenes
 - Estimación del canal de comunicaciones
 - Remoción de ruido impulsivo en imágenes
 - Restauración de imágenes
 - Pre-procesamiento de imágenes médicas