

## ANÁLISIS DE LA EFICIENCIA DEL ALGORITMO DE RECONSTRUCCIÓN DE SEÑALES BASADO EN LA MEDIANA PONDERADA A TRAVÉS DE LA REGRESIÓN DE COX

Jabbour George<sup>1</sup> y Paredes Jose Luis<sup>1</sup>  
(Recibido febrero 2012, Aceptado mayo 2012)  
<sup>1</sup>Universidad de los Andes  
jabbour@ula.ve

**Resumen:** En el presente artículo se analiza, a través del modelo de regresión de Cox, la eficiencia del algoritmo de reconstrucción de señales basado en la regresión de Mediana Ponderada, en el contexto del Sensado Comprimido. Se realizaron 1620 reconstrucciones de señales de diferente tamaño, con diferente tamaño de soporte (K), y variando además, el número de medidas (M) y el parámetro de regularización ( $\alpha$ ), que forma parte esencial del algoritmo bajo estudio. Entre los resultados más importantes obtenidos se encuentra que la eficiencia del algoritmo tiene un comportamiento parabólico invertido con respecto a  $\alpha$ , alcanzando su máximo en  $\alpha=0.8$ . Además, se encontró que este algoritmo tiende a ser sensible a variaciones de  $\alpha$  y M, de manera que con ligeros cambios en estos parámetros el algoritmo puede cambiar radicalmente su velocidad de convergencia. Así, ajustando adecuadamente el valor de M se logra controlar el comportamiento inestable antes descrito, pues si la relación M/K pasa de 7 a 9, la probabilidad de que el algoritmo presente un buen desempeño aumenta del 40% al 70%. De igual manera, si el valor de  $\alpha$  cambia de 0.5 a 0.9, la probabilidad de que el algoritmo presente un buen desempeño aumenta de 0.14 a 0.96.

**Palabras clave:** Reconstrucción de señales/ Sensado comprimido/ Mediana ponderada/ Regresión de Cox/ Análisis de supervivencia.

## ANALYSIS OF THE EFFICIENCY OF WEIGHTED MEDIAN-BASED SIGNAL RECONSTRUCTION ALGORITHM THROUGH COX REGRESSION

**Abstract:** In this paper, the efficiency of the algorithm for compressive sensing (CS) signal reconstruction based on weighted median regression (WMR) is analyzed through a Cox-regression model. We perform 1620 reconstructions for signals with different dimension (N), sparsity (K), number of measurements (M) and regularization parameter ( $\alpha$ ) that induces sparsity in the solution. Among the most relevant results, we find that the algorithm efficiency, as a function of the regularization parameter, follows an inverted parabolic function reaching its maximum at  $\alpha = 0.8$ . Furthermore, we show that the reconstruction algorithm is quite sensible to  $\alpha$  and M. Thus, a slight change on those parameters leads to a notable variation on the algorithm's convergence speed. Therefore, by suitably tuning the number of measurements, we can control the volatile described above. Thus, if the ratio N/M goes from 7 to 9, the probability of having a good performance increases from 0.4 to 0.7. Furthermore, if  $\alpha$  changes from 0.5 to 0.9 this probability increases from 0.14 to 0.96.

**Keywords:** Signal Reconstruction/ Compressive Sensing/ Weighted Median/ Cox Regression/ Survival Analysis.

### I. INTRODUCCIÓN

En el campo del muestreo/reconstrucción de señales, el teorema de Shannon-Nyquist establece que no habrá pérdida de información si la señal original es muestreada con una frecuencia de al menos el doble del ancho de banda de la misma [1]. Dependiendo de las características de la señal original, esta condición puede representar una seria limitación, especialmente, cuando la señal es de banda ultra-ancha, demandando el uso de sensores o

dispositivos de adquisición de datos extremadamente rápidos; o cuando la señal es de alta resolución (p.e. una fotografía de alta resolución o un video de alta calidad), requiriendo de espacio de almacenamiento, canal de comunicación y procesamiento (hardware y suministro de energía) de gran escala [2][3][7]. Adicionalmente, en el área de telecomunicaciones siempre ha existido un marcado interés por reducir el tamaño de las señales a transmitir, y con ello, disminuir los costos de transmisión [4].

En definitiva, es indiscutible el gran beneficio que representa el disponer de una herramienta que sea capaz de generar representaciones de señales de baja dimensionalidad (por debajo de la tasa de Nyquist) con poca o ninguna pérdida de información; y es esto precisamente lo que ha servido de punto de partida para la génesis del método llamado Sensado Comprimido (SC) [5][6], cuya eficacia no depende del ancho de banda de la señal bajo análisis sino de la cantidad de información de la misma [9]. En este sentido, el SC ha surgido como una técnica vanguardista en el campo del muestreo/reconstrucción de señales, permitiendo representar señales mediante un número reducido de medidas, y a su vez, ofreciendo herramientas capaces de reconstruir las señales originales con muy poca pérdida de información. Así, el SC simplifica el proceso de adquisición a un costo de tener una mayor complejidad en el proceso de reconstrucción.

Aún así, existe la necesidad de mejorar los resultados de este novedoso enfoque de adquisición de señales, apuntando a simplificar su aplicación, y a su vez, hacerlo cada vez más robusto, ya que entre sus principales desventajas se encuentra la complejidad y la gran cantidad de cómputos asociados a sus algoritmos en la fase de reconstrucción [2][3][4][7][8]. Por esta razón, el presente artículo pretende generar resultados que permitan a los investigadores y usuarios de este tipo de métodos conocer aspectos técnicos relacionados con la eficiencia de los algoritmos que forman parte del mismo, y así, facilitar su uso. Específicamente, se realiza un análisis estadístico de la influencia de las características relativas a la señal a recuperar/reconstruir y del parámetro de regularización sobre la eficiencia del algoritmo de reconstrucción de señales basado en la regresión de mediana ponderada. Se escoge este algoritmo de reconstrucción en particular dado que según fue mostrado en [8], es robusto a una gran variedad de tipos de ruido que pudiera contaminar las mediciones, además de la facilidad para seleccionar el parámetro de regularización que define el grado de densidad de la señal a reconstruir.

La organización del artículo es como sigue. Primero, en las secciones 2 y 3, se presentan los fundamentos básicos que forman parte del método SC, el método de reconstrucción de señales basado en la mediana ponderada, así como el modelo de regresión de Cox; luego, en la sección 4 se describen los experimentos realizados, y en la sección 5 se presentan los resultados obtenidos. Finalmente, en la sección 6, se presentan las conclusiones.

## II. DESARROLLO

### 1. Sensado Comprimido

#### 1.1 Muestreo/compresión de señales

Dada una señal real en tiempo discreto, finita y comprimible,  $x[n]$  ( $n=1,2,\dots,N$ ), el objetivo del método SC radica en encontrar una secuencia de medidas,  $y[m]$  ( $m=1,2,\dots,M$ ), de manera que sea posible reconstruir la

señal original  $x[n]$  a partir de  $y[m]$  con ninguna o poca pérdida de información, siendo  $M$  un valor mucho menor que  $N$  (un  $M$  muy pequeño implica una alta tasa de compresión pero con relativa baja probabilidad de reconstrucción, y viceversa) [8][9]. La condición de que la señal de interés debe ser discreta en tiempo obedece al único propósito de simplificar la presentación del método, sin embargo, el concepto de SC puede ser aplicado a señales analógicas donde el proceso de proyección ocurre en el dominio analógico.

Matemáticamente, la secuencia de medidas,  $y[m]$ , se obtiene al proyectar el vector  $\mathbf{x}=[x[1],\dots,x[N]]^T$  en un nuevo espacio  $M$ -dimensional, representado por una base obtenida aleatoriamente. Esto es:

$$\mathbf{y} = \Phi \mathbf{x};$$

$$\Phi = \begin{bmatrix} \phi_{11} & \dots & \phi_{1N} \\ \vdots & \ddots & \vdots \\ \phi_{M1} & \dots & \phi_{MN} \end{bmatrix} = [\Phi_1 \dots \Phi_N];$$

$$\Phi_n = [\phi_{1n} \dots \phi_{Mn}]^T; \quad (n=1,\dots,N) \quad (1)$$

en donde  $\Phi$  es conocida como la matriz de proyecciones aleatorias o matriz de medidas, cuyas componentes son realizaciones de una variable aleatoria tipo Gaussiana o Bernoulli. Para que la proyección resultante,  $y[m]$ , contenga la información relevante que permita reconstruir a  $x[n]$  con una pérdida de información despreciable, ésta última debe ser una señal poco densa en un dominio  $\Psi$  denominado diccionario de representación y la matriz de proyección debe ser incoherente con este diccionario [9]. Estrictamente hablando,  $x[n]$  debe ser una señal con  $K$  componentes no nulos en el dominio del tiempo o en cualquier otro dominio creado por funciones parametrizadas que conformen un diccionario, tal como se define a continuación:

**Señal K-sparse.** Sea  $\Psi=[\Psi_1 \dots \Psi_N]$ ;  $\Psi_n=[\psi_{1n},\dots,\psi_{Nn}]^T$  ( $n=1,\dots,N$ ); una base ortonormal que define los elementos de un diccionario, y  $\mathbf{s}=[s_1,\dots,s_N]$  un vector de coeficientes tales que:

$$\mathbf{x} = \Psi \mathbf{s} = \sum_{n=1}^N s_n \Psi_n \quad (2)$$

Así,  $\mathbf{x}$  es una señal  $K$ -sparse si es posible aproximarla como una combinación lineal de sólo  $K$  vectores de  $\Psi$ , es decir, si  $K$  coeficientes  $s_n$  poseen valores significativos [10]. Obviamente, sólo son de interés aquellos casos en donde  $K$  es mucho menor que  $N$  ( $K \ll N$ ), pues esto significa que  $\mathbf{x}$  es poco densa, y por lo tanto, es comprimible en un vector de medidas,  $\mathbf{y}$ , de menor dimensión, justificándose así el uso de esta técnica.

#### 1.2 Selección del diccionario

Es necesario encontrar una matriz  $\Psi$  (que es un espacio  $N$ -dimensional) tal que exista una secuencia  $\{s\}$  con sólo  $K$  elementos no nulos, y cuya proyección sobre  $\Psi$  genere

a la señal  $x$ . Si  $x$  es K-sparse en el dominio del tiempo no es necesario proyectarla en un espacio alternativo, y en su lugar, simplemente se asume  $s=x$ , lo cual implica  $\Psi=I$ ; es decir, si  $x$  es K-sparse en el dominio del tiempo, entonces, el diccionario se puede tomar como la matriz identidad de rango N. En este caso, se dice que la señal es poco densa en el dominio canónico.

En caso de que  $x$  no sea poco densa en el dominio canónico, el vector  $s$  no presenta las características deseadas, de manera que se hace necesario seleccionar un diccionario diferente, como por ejemplo, un diccionario conformado por las bases de la transformada de Fourier, de la transformada del coseno, o de la transformada wavelet, o incluso de la combinación de estas bases ortogonales [9]. De esta forma, en alguno de estos diccionarios la señal  $x$  tendrá una representación con sólo K coeficientes significativos

Otro aspecto a tomar en cuenta en la selección del diccionario trata sobre su relación con la matriz de proyecciones. En concreto, la matriz de proyecciones se debe seleccionar de manera que sea incoherente con el diccionario donde la señal tiene una representación poco densa, según se describe a continuación.

**1.3 Incoherencia entre el diccionario  $\Psi$  y la matriz de proyecciones  $\Phi$**

La coherencia entre las matrices  $\Phi$  y  $\Psi$  es una medida que refleja el grado de correlación que existe entre estas dos matrices, lo cual incide sobre la probabilidad de reconstruir la señal original con la menor pérdida de información posible: a menor coherencia entre  $\Phi$  y  $\Psi$ , mayor probabilidad de éxito en la reconstrucción se tendría. Esta medida se define como sigue:

$$\mu(\Phi, \Psi) = \sqrt{n} * \max_{1 \leq m \leq M; 1 \leq n \leq N} \left| \langle \varphi^m, \psi_n \rangle \right| \tag{3}$$

En la Ec. (3),  $\varphi^m$  es la m-ésima fila de  $\Phi$ ,  $\psi_n$  es la n-ésima columna de  $\Psi$  y  $\langle \cdot, \cdot \rangle$  denota el producto escalar. Mientras más grande sea la correlación entre estos vectores (en valor absoluto), mayor será la coherencia entre las matrices  $\Phi$  y  $\Psi$  [9]. Se ha probado que si los elementos  $\phi_{ij}$  siguen una distribución gaussiana o de Bernoulli, la matriz resultante es incoherente con una gran variedad de diccionarios, entre ellos los creados a partir de las bases de Fourier, DCT y wavelet.

**1.4 Reconstrucción de la señal original: enfoque basado en estimaciones obtenidas mediante la regresión de mediana ponderada**

La reconstrucción de una señal consiste en realizar el proceso inverso al de su compresión, es decir: a partir del vector de medidas,  $y$ , y la matriz de proyecciones,  $\Phi$ , el objetivo es obtener la secuencia  $x$  original, aquella cuya compresión condujo precisamente a  $y$  [8][9]. Dado que  $x \in \mathfrak{R}^N$ ,  $y \in \mathfrak{R}^M$  y  $M < N$ , existen múltiples soluciones al problema inverso, por lo que se plantea la minimización

de la diferencia entre la señal original y la señal reconstruida en el espacio de las proyecciones, y esto implica que esta última señal posea el mismo grado de densidad presente en la señal original, lo cual añade una considerable dificultad al proceso de reconstrucción. Matemáticamente, la señal reconstruida se obtiene al resolver el siguiente problema de optimización [8]:

$$\hat{x} = \arg \min_x \|y - \Phi X\|_{l_{p_1}} + \tau \|X\|_{l_{p_2}} \tag{4}$$

donde  $\|\cdot\|_{l_p}$  denota la norma  $l_p$ , de órdenes  $p_1$  y  $p_2$ , cuyos valores inciden notablemente sobre el error presente en la señal reconstruida, así como el grado de densidad (tamaño del soporte) de la misma. El primer término de (4), asociado a la norma  $l_{p_1}$ , tiene el propósito de reducir el error entre la señal medida y la señal proyectada, por lo cual, generalmente se establece  $p_1=2$  inducido por la suposición de Gaussianidad asociada al ruido que contamina las proyecciones. Esto conduce a una optimización por mínimos cuadrados. También es común encontrar el uso de  $p_1=1$ , norma que es óptima bajo el criterio de máxima verosimilitud cuando el ruido es Laplaciano. Esta norma conduce a una minimización de desviación absoluta. Sin embargo, este primer término por sí solo no resuelve por completo el problema bajo consideración, ya que además de la necesidad de un error mínimo (al utilizar  $p_1=1$  ó  $p_1=2$ ) existe el requerimiento de que la señal reconstruida sea poco densa al igual que la señal original. La probabilidad de que esta última condición sea satisfecha por una optimización basada sólo en el primer término (mínimos cuadrados o mínima desviación absoluta) es sumamente baja; razón por la cual se incorpora el término asociado a la norma  $l_{p_2}$ , cuyo rol en el proceso de optimización se centra en controlar que el tamaño del soporte de la señal reconstruida no sea mayor que K. Esto se logra estableciendo  $p_2=0$ , es decir, haciendo que el segundo término de (4) se corresponda con el tamaño de soporte (seudo-norma  $l_0$ ) del vector X, buscando así, que la optimización genere una señal que no solo presente un error mínimo sino que también tenga un tamaño de soporte adecuado. Adicionalmente, se tiene el parámetro de regularización,  $\tau$ , cuya función es establecer un compromiso entre la magnitud del error y el tamaño de soporte de la señal reconstruida, pues  $\tau=0$  equivale a ignorar el requerimiento relacionado con el tamaño del soporte, mientras que un  $\tau$  muy grande generaría una señal con un tamaño de soporte pequeño, pero con una diferencia grande con respecto a la señal original, de manera que es importante ajustar este parámetro adecuadamente, para así obtener una solución que sea satisfactoria desde ambos puntos de vista.

La solución del problema de optimización planteado en (4) se puede obtener a través de múltiples métodos, cada uno de los cuales se especializa en determinadas situaciones, dependiendo principalmente del grado de densidad y de las características de la contaminación de la señal original [8][9][10].

En el caso particular de este trabajo, se hace uso de un método que consiste en un enfoque regresivo basado en estimaciones obtenidas mediante el operador de Mediana Ponderada [9]. Este método es de especial interés cuando la señal original está contaminada con ruido cuya distribución estadística presenta colas con una densidad mayor que la de una distribución normal, lo que lo hace robusto ante la presencia de ruido de naturaleza impulsiva. Este enfoque consiste en dividir el problema de optimización N-dimensional (definido en (4)) en N problemas unidimensionales, de la forma:

$$\hat{x}_n = \arg \min_{X_n} \sum_{i=1}^M |y[i] - \phi_n X_n| + \tau_0 \|X_n\|_{l_0}; n=1, \dots, N \quad (5)$$

cuya solución,  $\hat{x}_n$ , se obtiene como la mediana ponderada de la señal residuo, seguido de un proceso de umbralización, esto es [8]:

$$\hat{x}_n = \begin{cases} \text{Mediana} \left( \left\{ \phi_n \frac{|y[i] - \sum_{j=1; j \neq n}^N \phi_j \hat{x}_j}{\phi_n} \right\}_{i=1}^M \right) & \text{si } \|r_n\|_{l_1} - \|r_n - \hat{x}_n \Phi_n\|_{l_1} > \tau_0 \\ 0 & \text{en otro caso} \end{cases} \quad (6)$$

para  $n=1, \dots, N$ .

En donde  $r_n = y - \sum_{j=1; j \neq n}^N \Phi_j \hat{x}_j$ ,  $\Phi_n$  es la n-ésima columna de  $\Phi$  y  $\|\cdot\|_{l_p}$  denota la norma  $l_p$ . Esta expresión se utiliza iterativamente para estimar los N elementos de la secuencia original, aunque, dado que en cada optimización individual se toma en cuenta una sola variable a la vez, con N iteraciones (una iteración para cada variable) muy difícilmente se logre alcanzar el óptimo global N-dimensional. Esto conlleva a la necesidad de repetir múltiples ciclos de estimación para alcanzar la solución óptima global, en donde se define como ciclo al proceso de estimar a cada  $x_i$  una vez, es decir, un ciclo corresponde a N iteraciones. Uno de los criterios más utilizados para detener al algoritmo se basa en la energía residual normalizada, en donde, el algoritmo se detiene cuando se cumple que  $\|y - \Phi \hat{x}\|_2^2 / \|y\|_2^2 \leq \varepsilon$ ; y como segundo criterio está el número máximo de ciclos, del cual no se puede prescindir puesto que no existe una garantía de que el criterio anterior se vaya a satisfacer en algún momento, ya que  $\varepsilon$  es fijado en función del grado de exactitud que se desea obtener en la señal reconstruida y en función del tiempo de ejecución del algoritmo. Para finalizar, vale la pena señalar que el umbral  $\tau$ , que en el primer ciclo toma el valor  $\tau_0$ , va disminuyendo progresivamente de un ciclo a otro con el fin de permitir la incorporación de nuevos elementos no nulos a la secuencia X. El algoritmo inicia el proceso de cálculo asumiendo que  $x[i]=0$  ( $i=1, \dots, N$ ), y dependiendo de la importancia de cada elemento se van asignando valores no nulos a los mismos, y a medida que se disminuye el valor del umbral, más elementos se van incorporando al soporte de la secuencia a recuperar. Es por esto que, en el k-ésimo ciclo, el umbral se define como  $\tau_k = \tau_0 \alpha^k$ , en donde  $\alpha$  es una

constante comprendida en el intervalo (0,1); haciendo que el umbral vaya reduciéndose gradualmente. El valor de  $\tau_0$  fue seleccionado siguiendo los criterios establecidos en [6].

## 2. Análisis de Supervivencia a través de la Regresión de Cox

En el análisis de supervivencia, uno de los principales objetivos consiste en determinar la influencia que poseen ciertos factores sobre el tiempo que transcurre desde que se inicia algún proceso en particular hasta que ocurre cierto evento de interés [10].

Entre los conceptos asociados al análisis de supervivencia, son de especial interés el tiempo de supervivencia, la función de supervivencia, la función de riesgo y la esperanza de vida [12]. En detalle, el tiempo de supervivencia, T, es una variable aleatoria no negativa que representa el tiempo que transcurre desde que se comienza a observar el individuo hasta que ocurre cierto evento de interés llamado muerte. Mientras no ocurra el evento de interés, o muerte, se asume que el individuo se encuentra vivo, o en estado de supervivencia. En general, T es una variable aleatoria continua con función de densidad de probabilidad  $f_T(t)$ , y función de densidad acumulada  $F_T(t)=P(T \leq t)$ . La función de supervivencia se define como  $S(t)=1-F(t)$ , y representa la probabilidad de que cierto individuo esté vivo en el instante t. Por otro lado, la función de riesgo es una medida alternativa para caracterizar a la supervivencia, y se define como  $H(t)=f(t)/S(t)$ , de manera que la función riesgo es inversamente proporcional a la función de supervivencia, y por lo tanto, a mayor supervivencia menor riesgo, y viceversa. Por último, se tiene el concepto de esperanza de vida, que no es más que la esperanza estadística de T. Si la presencia de cierto factor aumenta la esperanza de vida de un individuo, significa que contribuye a incrementar el tiempo esperado que transcurre hasta que ocurre el evento muerte, es decir, aumenta la función de supervivencia, o equivalentemente, disminuye la función de riesgo [12][13].

Dentro del conjunto de métodos que forman parte del análisis de supervivencia se encuentra la regresión de Cox, la cual se basa en un modelo que permite determinar la influencia que poseen ciertos factores sobre la función de riesgo, y por ende, sobre el tiempo esperado que transcurre hasta la ocurrencia del evento muerte en un contexto dado. Por lo tanto, el modelo de regresión de Cox es una función cuya variable dependiente es el riesgo en el que se encuentra un individuo en el instante t, y cuyas variables independientes son factores que se cree que pueden influir sobre tal riesgo [13]; específicamente:

$$H(t) = H_0(t) e^{\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p} \quad (7)$$

en donde,  $H_0(t)$  es el riesgo en el que se encontraría el individuo en el instante t si  $X_1=X_2=\dots=X_p=0$ . Es un riesgo referencial, que equivale a ignorar el efecto de las variables independientes. La interpretación de  $H_0(t)$  es

irrelevante, ya que no está relacionado con ninguna de las variables independientes del modelo, que representan el objetivo del estudio. Por otro lado, el interés se centra en las variables independientes, cuyo efecto se mide directamente a través de sus coeficientes  $\beta_i$  ( $i=1, \dots, p$ ). Si  $\beta_i > 0$ , entonces la presencia del  $i$ -ésimo factor aumenta el riesgo, mientras que, si  $\beta_i < 0$ , entonces el  $i$ -ésimo factor reduce el riesgo [13].

### 3. Descripción de los Experimentos

En el contexto de esta investigación se desea analizar la influencia de los factores S (señal madre), N (tamaño de la señal), K (soporte de la señal), M (número de mediciones) y  $\alpha$  (parámetro de regularización) sobre el número de ciclos necesarios hasta que el algoritmo de la mediana ponderada regresiva alcance la condición de energía residual normalizada menor a 0.1, es decir:

$$\|y - \Phi \hat{x}\|_2^2 / \|y\|_2^2 \leq 10^{-1},$$

haciendo uso de la regresión de Cox. Para tal fin se realizaron pruebas de reconstrucción para 12 señales, obtenidas a partir de 3 señales madre, S1, S2 y S3, generadas aleatoriamente y con una longitud de 1024 cada una. A partir de cada una de estas 3 señales se generaron 4 señales: desde el componente 1 hasta el componente  $N=256 \cdot L$ , para  $L=1, 2, 3, 4$ ; de manera que en total se tienen 3 señales de longitud 256, 3 de 512, 3 de 768 y 3 de 1024. Para  $N=256$ , se utilizaron valores de K (en el sentido K-sparse) de 10, 14 y 18; para  $N=512$ , se utilizaron valores de K de 20, 28 y 36; para  $N=768$ , se utilizaron valores de K de 30, 42 y 54; y  $N=1024$ , se utilizaron valores de K de 40, 56 y 72. Además, Para cada valor de N (de cada una de las señales madre) y K, se utilizaron 3 valores de M:  $7 \cdot K$ ,  $8 \cdot K$  y  $9 \cdot K$ . Por otro lado, para cada combinación de parámetros descritos anteriormente, se aplicó el algoritmo de la mediana ponderada para valores de  $\alpha$  comprendidos entre 0.2 y 0.9, con incrementos de 0.05:  $\alpha = \{0.20, 0.25, \dots, 0.90\}$ . En resumen, se realizaron 1620 ( $3S \times 4N \times 3K \times 3M \times 15\alpha$ ) pruebas de reconstrucción, para las cuales se ha registrado el número de ciclos (denotado por C) para el cual se alcanzó la condición  $\|y - \Phi \hat{x}\|_2^2 / \|y\|_2^2 \leq 10^{-1}$ .

Adicionalmente, se utilizó la misma matriz de proyecciones para todas pruebas para garantizar que estas se realizaran en las mismas condiciones, pues no es conveniente comparar los resultados de dos pruebas si estas se han basado en diferentes matrices de proyección.

En cuanto a la aplicación del modelo de regresión de Cox, se utiliza a la variable C como el tiempo de supervivencia, y se dice que ocurre el evento muerte si  $C \leq 100$ , y por lo tanto, a partir de los valores de C se estima el riesgo para cada caso. Además como variables independientes de este modelo se toman las variables S, N, K, M y  $\alpha$ ; con lo cual, el modelo de regresión a evaluar tendría la siguiente forma:

$$H(t) = H_0(t) e^{\beta_1 S + \beta_2 N + \beta_3 K + \beta_4 M + \beta_5 \alpha} \quad (8)$$

Los resultados de este modelo aportarán evidencia

estadística sobre la forma en la que influye cada uno de los parámetros bajo estudio.

### 4. Resultados

A continuación se presentan resultados estadísticos tanto descriptivos como analíticos de acuerdo a lo planteado anteriormente, haciendo énfasis en los parámetros  $\alpha$  y M puesto que estos valores son establecidos por el investigador, mientras que los demás valores (S, N y K) son propios de la señal, y por lo tanto, no forman parte del proceso de toma de decisiones al momento de aplicar el algoritmo bajo estudio.

#### 4.1 Análisis descriptivo

De las 1620 reconstrucciones de señal llevadas a cabo, en 552 casos ocurrió el evento de interés (en una prueba ocurre el evento de interés si  $C \leq 100$ ): 22 casos (3.4%) para  $\alpha \leq 0.45$  y 530 casos (54.53%) para  $\alpha \geq 0.5$ . Dado que es necesario contar con resultados en donde exista diversidad de casos para poder analizar los factores que influyen sobre el desempeño del algoritmo bajo estudio, se descartan los registros asociados a  $\alpha \leq 0.45$ , dado que los mismos no ofrecen tal diversidad, pues en el 96.6% de estas pruebas no ocurrió el evento de interés, y esto a su vez conlleva a un primer resultado que indica que no es recomendable utilizar valores de  $\alpha$  inferiores a 0.5, ya que indiferentemente de los valores de los parámetros restantes, el algoritmo va a ser ineficiente con muy alta probabilidad, lo cual concuerda con lo expuesto por los autores de este método, quienes recomienda utiliza un  $0.75 \leq \alpha \leq 0.95$  [7]. En consecuencia, el análisis que se realiza en lo sucesivo se basa sólo en las 972 reconstrucciones de señal asociadas a  $\alpha \geq 0.5$ , en donde, en 530 casos ocurrió el evento de interés, distribuidos de la siguiente manera:

Tabla I. Resultados según el valor de  $\alpha$

$\alpha$	$\bar{C}$	No. de Casos	% Casos
0.50	7	15	13.89
0.55	9	21	19.44
0.60	9.08	39	36.11
0.65	8.92	36	33.33
0.70	8.56	54	50.00
0.75	9.88	78	72.22
0.80	9.86	87	80.56
0.85	10.53	96	88.89
0.90	11.91	104	96.30

Tabla II. Resultados según la relación M/K

M/K	$\bar{C}$	No. de Casos	% Casos
7	10.69	128	39.51
8	10.07	177	54.63
9	9.6	225	69.44

En las Tablas I y II,  $\bar{C}$  denota el promedio de ciclos registrados en las pruebas asociadas al respectivo caso, pero sólo en donde ocurrió el evento de interés; "No. de

Casos” se refiere al número de casos en donde ocurrió el evento de interés, mientras que “% Casos” indica el porcentaje de casos en los cuales ocurrió el evento de interés, que a su vez puede ser interpretado como una estimación de la probabilidad de que ocurra el evento de interés.

En los 530 casos en los que ocurrió el evento de interés, los valores de C se mantuvieron dentro del rango [6,16], con un promedio de 10.02, es decir, que el error establecido ( $\epsilon=10^{-1}$ ) o era alcanzado en  $6 \leq C \leq 16$  (en 530 casos) o era alcanzando en  $C > 100$  (en los 442 casos restantes), es decir, ninguna de las 972 corridas duró entre 17 y 100 ciclos. Esta brecha puede ser interpretada como un indicador de la inestabilidad del algoritmo, o alternativamente, que la velocidad de convergencia del algoritmo es altamente no lineal con respecto a las características de la señal y con respecto al parámetro de regularización  $\alpha$ , de manera que en algunos casos el algoritmo es muy rápido y en los demás casos es lento, sin posibilidad de términos medios. Lo antes afirmado puede significar simultáneamente una ventaja y una desventaja, ya que si el usuario posee el nivel de conocimiento adecuado sobre este algoritmo, encontrará en el mismo una herramienta sumamente eficiente, y viceversa.

Así mismo, en la Tabla I se puede observar que a mayor valor de  $\alpha$ , más tarda el algoritmo para alcanzar el error establecido ( $\epsilon=10^{-1}$ ), aunque logra esta meta con mayor probabilidad. Por ejemplo, con  $\alpha=0.5$  hicieron falta sólo 7 ciclos en promedio ( $\bar{C}=7$ ), pero apenas en el 13.89% de los casos ocurrió el evento de interés, mientras con  $\alpha=0.9$  el evento de interés se alcanzó en el 96.30% de los casos. Lo antes afirmado refleja una fuerte sensibilidad del desempeño del algoritmo de reconstrucción con respecto a  $\alpha$ . Así, si el parámetro de regularización decae muy rápidamente, el algoritmo de reconstrucción no puede seleccionar adecuadamente el soporte de la señal de interés, induciendo errores de reconstrucción; y por el contrario, si el parámetro de regularización decae lentamente, el algoritmo detecta correctamente el soporte de la señal y estima sus valores apropiadamente a un costo de requerir un mayor número de ciclos.

Por otro lado, se tiene la relación M/K, que indica cuántas medidas se utilizan por cada elemento no nulo de la señal original. En la Tabla II se puede apreciar que mientras más grande sea el vector de medidas con respecto al grado de densidad de la señal original, más rápido será el algoritmo, y además, existe una mayor probabilidad de ocurrencia del evento de interés, siendo esto último la más significativo del resultado, ya que si en vez de utilizar 7 medidas se utilizan 9 por cada elemento no nulo de la señal original, se logra aumentar de 39.51% a 69.44% tal probabilidad.

**4.2 Resultados del modelo de regresión de Cox**

En las Tablas III y IV se resumen los resultados correspondientes a los parámetros del modelo de regresión de Cox (correspondiente a la ec. 8), en donde se

asume que todas las variables son cuantitativas, excepto S (indicadora de la señal madre) que es una variable categórica:

**Tabla III. Resultados del modelo de Cox asumiendo a  $\alpha$  como variable cuantitativa.**

Var	$\beta$	Wald	Sig.
N	-.001	10.690	.001
K	-.078	34.735	.000
M	.011	58.498	.000
$\alpha$	4.571	168.663	.000
S		.003	.998
S(1)	.003	.001	.978
S(2)	-.003	.001	.977

**Var:** nombre de la variable;  **$\beta$ :** valor de su coeficiente en el modelo de Cox ( $\beta_i$ ); **Wald:** valor del estadístico de Wald, que es un estadístico de prueba de la hipótesis nula que establece que la respectiva variable no es significativa (su coeficiente es nulo). **Sig:** nivel de significancia del estadístico de Wald. Si  $Sig < p$ ,  $0 \leq p \leq 1$ , entonces la respectiva variable es significativa al  $100*(1-p)\%$ .

**Tabla IV. Prueba del Omnibus (bondad de ajuste).**

D = -2 Log (Verosimilitud)	Chi-cuadrado	GL	Sig.
6731.064	254.624	6	.000

**D:** es un estadístico de prueba de la bondad de ajuste del modelo, cuya hipótesis nula establece que todos los coeficientes del modelo son nulos, y por lo tanto, el modelo no se ajusta bien a los datos. Este estadístico se distribuye chi-cuadrado con **GL** grados de libertad. **Sig:** nivel de significancia del estadístico antes mencionado. Si  $Sig < p$ ,  $0 \leq p \leq 1$ , entonces se rechaza la hipótesis nula, y en tal caso, se concluye que el modelo se ajusta bien a los datos con un nivel de confianza del  $100*(1-p)\%$ .

En primer lugar, la Tabla III indica que, según la prueba del Omnibus [13], no se rechaza la hipótesis de que el modelo se ajusta bien a los datos, lo cual indica que el modelo es válido. Por otro lado, según la Tabla IV, se puede concluir con un 99.9% de confianza lo siguiente: a mayor valor de N y K (mientras más grande sea la secuencia, y mientras mayor sea su soporte), menor riesgo, o equivalentemente, mayor supervivencia, y por lo tanto, mayor será el número de ciclos necesarios hasta que los residuos alcancen el error establecido. Por el contrario, a mayor valor de M y  $\alpha$  en el intervalo [0.5 - 0.9] menor será el número de ciclos necesarios. En cuanto a las señales madre, estas no han influido en lo absoluto sobre los resultados, es decir, los resultados han sido independientes de las señales madre que se utilizaron.

Entre los resultado antes presentados, causa especial inquietud el relacionado con el coeficiente  $\alpha$ , para el cual se afirmó que a mayor valor del mismo, mayor será la velocidad de convergencia del algoritmo, sin embargo, vale la pena investigar si esta afirmación es absolutamente cierta, o si este comportamiento simplemente refleja una tendencia general, pero que analizada en detalle, se pudiese descubrir algún otro comportamiento un tanto más complejo, diferente a la influencia monótona obtenida en los resultados anteriores. Para investigar este aspecto, se decidió asumir al coeficiente  $\alpha$  como una

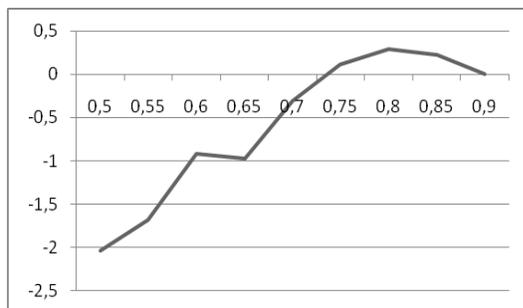
variable cualitativa, con el fin de analizar por separado el efecto de cada uno de sus posibles valores. Al realizar este cambio, y reestimar los parámetros del modelo de Cox, se obtuvieron los siguientes resultados:

**Tabla V. Resultados del modelo de Cox asumiendo a  $\alpha$  como variable cualitativa.**

Var	$\beta$	Wald	Sig.
N	-.001	10.947	.001
K	-.085	40.800	.000
M	.012	69.034	.000
$\alpha$		178.999	.000
$\alpha(1)$	-2,033	53.716	.000
$\alpha(2)$	-1,684	49.000	.000
$\alpha(3)$	-.915	23.381	.000
$\alpha(4)$	-.973	24.956	.000
$\alpha(5)$	-.306	3.269	.071
$\alpha(6)$	.117	.595	.440
$\alpha(7)$	.292	3.926	.048
$\alpha(8)$	.227	2.540	.111
S		.009	.996
S(1)	.008	.006	.937
S(2)	.000	.000	.999

Los significados de las columnas de esta tabla son los mismos que los de la Tabla IV.

En este caso, se tomó  $\alpha=0.90$  como la categoría de referencia, y se codificaron los valores de 0.50, 0.55, ... 0.85 como  $\alpha(1)$ ,  $\alpha(2)$ , ...,  $\alpha(8)$ . En la Tabla V se puede observar que a medida que se incrementa a  $\alpha$ , el algoritmo se acelera hasta llegar a  $\alpha=0.80$ ,  $\alpha(7)$ , en donde el algoritmo alcanza su máxima eficiencia, para comenzar a desmejorar en los siguientes valores de  $\alpha$ . El siguiente gráfico permite visualizar de una forma más adecuada el efecto de los valores de  $\alpha$  sobre la eficiencia del algoritmo:



**Figura 1.  $\alpha$  vs.  $\beta_i$**

En este gráfico, el eje horizontal representa los diferentes valores de  $\alpha$  evaluados (de 0.50 hasta 0.90), mientras que el eje vertical corresponde a los coeficientes del modelo de Cox asociados a cada valor de  $\alpha$ .

En la Figura 1, la curva obtenida demuestra que la relación entre la eficiencia del algoritmo con respecto a  $\alpha$  no es monótona creciente sino que sigue, aproximadamente, una parábola invertida que alcanza su máximo en  $\alpha=0.80$ , de manera que entre 0 y 0.8 la eficiencia del algoritmo mejora a medida que  $\alpha$  aumenta, y a partir de ahí, comienza a desmejorar. Es importante

señalar que estos resultados no contradicen lo presentado en la Tabla I, ya que en ese caso,  $\bar{C}$  fue calculado tomando en cuenta sólo las pruebas en donde el algoritmo convergió ( $C \leq 100$ ), a diferencia de los resultados del modelo de Cox, en donde se consideran todas las pruebas realizadas.

### III. CONCLUSIONES

1. Los resultados obtenidos en esta investigación permiten llegar a las siguientes conclusiones sobre el desempeño del algoritmo de reconstrucción de señales basado en la regresión de mediana ponderada.
2. A pesar de la gran variedad de condiciones, cuando el algoritmo alcanzó el error establecido ( $\epsilon=10^{-1}$ ), lo hizo recorriendo entre 10 y 16 ciclos, lo cual es considerado como un buen desempeño, mientras que en los demás casos se realizaron 100 ciclos sin que esto ocurriera, de manera que se evidencian dos tipos de comportamientos sumamente homogéneos y diferentes entre sí: uno lento y otro rápido; sin la posibilidad de resultados intermedios.
3. Sin embargo, ajustando adecuadamente los valores de  $\alpha$  y M se puede reducir drásticamente la probabilidad de que el algoritmo se comporte en forma lenta: con  $\alpha=0.5$  hicieron falta sólo 7 ciclos en promedio ( $\bar{C}=7$ ), pero apenas en el 13.89% de los casos el algoritmo requirió menos de 100 ciclos para converger, mientras con  $\alpha=0.9$  hicieron falta en promedio 11.91 ciclos, pero la convergencia se logró en el 96.30% de los casos. En ningún caso es recomendable utilizar  $\alpha \leq 0.45$ , pues con estos valores se obtuvo que en el 96.6% de los casos el algoritmo fue ineficiente, indiferentemente de los valores de los parámetros restantes.
4. Por otro lado, existe la posibilidad de ajustar el valor de M con el fin de aumentar la probabilidad de que el algoritmo se comporte en forma rápida, por ejemplo, si en vez de utilizar  $M=7K$  se utiliza  $M=9K$ , se logra aumentar de 39.51% a 69.44% tal probabilidad. En este caso, es importante tener presente que es de interés que M sea lo más pequeño posible, o lo que es equivalente, que  $N/M$  sea lo más grande posible, pues uno de los objetivos del SC es comprimir las señales, por lo tanto, es necesario tener cuidado al aumentar el valor de M.
5. En cuanto al resto de las características de la señal, según el modelo de regresión de Cox, se obtuvo que a mayor valor de N y K, menor riesgo (mayor supervivencia), y por lo tanto, mayor será el número de ciclos necesarios hasta que los residuos alcancen el error establecido. Además, con este modelo se confirmaron los resultados asociados a  $\alpha$  y M, expuestos anteriormente; aunque en el caso de  $\alpha$ , se obtuvo una información adicional que indica que la relación entre la eficiencia del algoritmo con respecto a  $\alpha$  no es monótona creciente sino que sigue,

aproximadamente, una parábola invertida que alcanza su máximo en  $\alpha=0.80$ , valor que representa un punto de equilibrio entre un número de ciclos bajo y la probabilidad de que el algoritmo no se desempeñe en forma lenta.

## V. REFERENCIAS

1. Mitra, S., Digital Signal Processing, A Computer Based Approach, 3era Edición, pp. 117-232, McGraw Hill, 2006.
2. Majumdar, A., y Ward, R., Compressed Sensing of Color Images. Signal Processing, Elsevier, Vol. 90, No. 12, pp. 3122-3127, 2010.
3. Han, B., Wu, F., y Wu, D., Image representation by compressive sensing for visual sensor networks. Journal of Visual Communication and Image Representation, Vol. 21, No. 4, pp. 325-333, 2010.
4. Bajwa, W., Sayeed, A., y Nowak, R., Compressed Sensing of Wireless Channels in Time, Frequency, and Space. Proc. 42nd Asilomar Conf. Signals, Systems, and Computers, Pacific Grove, CA, 2008.
5. Donoho, D., Compressed Sensing. IEEE Trans. on Information Theory, Vol. 52, No. 4, pp. 1289-1306, 2006.
6. Candes, E., Compressive Sampling, Proc. Int. Cong. Mathematicians, Madrid, Spain, Vol. 3, pp. 1433-1452, 2006.
7. Paredes, J., Arce, G., y Wang, Z., Ultra-Wideband Compressed Sensing: Channel Estimation. IEEE Journal of Selected Topics in Signal Processing, Vol. 1, No. 3, pp 383-395, 2007.
8. Paredes, J., y Arce, G., Compressive Sensing Signal Reconstruction by Weighted Median Regression Estimates. IEEE Transactions on Signal Processing, Vol. 59, No. 6, pp. 2585-2601, 2011.
9. Baraniuk, R., Compressive Sensing. IEEE Signal Processing Magazine, Vol. 24, pp. 118-124, 2007.
10. Tropp, J., y Gilbert, A., Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit. IEEE Transactions on Information Theory, Vol. 53, No. 12, pp 4655-4666, 2007.
11. Chen, Q., Oppenheim, A., y Wang, D., Encyclopedia of Data Warehousing and Mining, pp. 1077-1082, Idea Group Inc., 2006.
12. Berry, M., y Linoff, G., Data Mining Techniques, 2da Edición, pp. 383-420, Wiley, 2004.
13. Harrell, F., Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis, 1era Edición, pp. 465 - 507, Springer, 2001.