Información Tecnológica
Vol. 29(4), 281-290 (2018)
http://dx.doi.org/10.4067/S0718-07642018000400281

Una Revisión de los Estimadores de Matrices de Bajo Rango y Matrices Dispersas

Juan P. Hoyos[(1)] y Pablo E. Jojoa[(2)]
(1) Corporación Universitaria Comfacauca Unicomfacauca, Facultad de Ingeniería, Santander de Quilichao-Colombia (e-mail: jhoyos@unicomfacauca.edu.co)
(2) Universidad del Cauca, Facultad de Ingeniería Electrónica y Telecomunicaciones, Popayán-Colombia (email: pjojoa@unicauca.edu.co)

Recibido Oct. 26, 2017; Aceptado Ene. 11, 2018; Versión final Feb. 14, 2018, Publicado Ago. 2018

Resumen

Se presenta una revisión selectiva de los desarrollos más recientes en la estimación de la matriz de covarianza con un número de muestras menores a la dimensión ambiente. En particular se consideran estimadores para las estructuras bajo rango y dispersa, dando especial atención a los algoritmos más utilizados en la práctica. Se presentan estimadores que recurren al uso de técnicas clásicas como umbrales y descomposición de valor único (SVD), así como estimadores más novedosos basados en matrices aleatorias y optimización convexa. Una de las principales conclusiones del estudio es que a pesar de los grandes avances en el desarrollo de nuevos estimadores, temas relacionados al tiempo de cómputo de los programas convexos que utilizan los estimadores han sido poco explorados. Además, se observa la falta de trabajos sobre el desarrollo de estimadores para matrices con estructura conjuntamente dispersa y bajo rango.

Palabras clave: matriz de covarianza; bajo rango; dispersa; sketch; estimadores; matrices aleatorias

A Review of Low Rank and Sparse Matrix Estimators

Abstract

A selective review of the most recent developments in the estimation of the covariance matrix with a number of samples smaller than the ambient dimension is presented. In particular, estimators are considered for low rank and sparse structures, paying special attention to the algorithms most used in practice. Estimators that resort to the use of classical techniques such as thresholds and singular-value decomposition (SVD), as well as newest estimators based on random matrices and convex optimization are presented. One of the main conclusions of the study is that, in spite of the great advances in the development of new estimators, issues related to the computation time of the convex programs that use the estimators have been little explored. Additionally, lack of works on the development of estimators for matrices with jointly sparse and low rank structure is observed

Keywords: covariance matrix; low rank; sparse; sketch; estimators; random matrices

INTRODUCCIÓN

La teoría de señales ha sido un área de constante investigación desde sus primeros usos en la teoría de la información (Shannon, 2001). Sus avances han permitido el desarrollo de novedosos sistemas que han impactado notablemente la manera de vivir, llevando a la sociedad a la era de las Tecnologías de la Información y las Comunicaciones (TIC). El rápido crecimiento en la generación de datos e información por los diferentes sistemas de comunicaciones, y las actuales limitantes han ocasionado que los problemas de inferencia estadística y procesamiento de datos de altas dimensiones se hayan convertido en un reto de alta complejidad, ya que buscan cada vez métodos más eficientes de estimación de las estadísticas de segundo orden, manteniendo niveles óptimos de precisión (Chen et al., 2015; Leus y Tian, 2011). Entre las limitantes actuales se encuentra la capacidad de procesamiento y la memoria disponible, por lo cual técnicas como censado compresivo o *compressed sensing* (Bryan y Leise, 2013) y *sketches* (Chen et al., 2015) recurren al uso de optimizaciones y matrices aleatorias para realizar una recuperación a partir de mediciones y procesamiento de los datos sin requerir de su almacenamiento.

La estimación de matrices juega un papel preponderante en muchas aplicaciones, en especial la estimación de la matriz de covarianza es fundamental en aplicaciones como procesamiento de señales (Krim y Viberg, 1996), reconocimiento de patrones, geometría convexa computacional, estadística de altas dimensiones, aprendizaje de máquina,..., ya que captura la estructura de la covarianza de las variables aleatorias (Vershynin, 2012). Sea x un vector aleatorio en R^n distribuido acorde a alguna función de distribución μ. La matriz de covarianza viene definida como

$$\Sigma = \mathrm{E}\{(x-\mathrm{E}x)(x-\mathrm{E}x)^T\} = \mathrm{E}\{(x-\mathrm{E}x)\otimes(x-\mathrm{E}x)\} \tag{1}$$

y tiene las propiedades de ser simétrica y semi-definida positiva. Los autovectores de la matriz de covarianza son llamados los componentes principales, he aquí la importancia en Análisis de Componentes Principales o PCA (por sus siglas en ingles). La estimación de la matriz de covarianza se puede realizar a partir de muestras mediante

$$\Sigma_k = \frac{1}{k}\sum_{i=1}^{k}(x-\mathrm{E}x)(x-\mathrm{E}x)^T \tag{2}$$

y de acuerdo a la ley de grandes números se garantiza que el estimador se vuelva ajustado cuando el número de muestras k va al infinito, es decir $\Sigma_k \to \Sigma$ cuando $k \to \infty$. La dificultad radica en cuán grande debe ser el número de muestras para asegurar que la muestra de covarianza converja a la covarianza verdadera?, o equivalentemente cuantas muestras debo tener para lograr cierta precisión (ε=0.001) en el operador norma

$$\|\Sigma_k - \Sigma\| \le \varepsilon\|\Sigma\|, \tag{3}$$

donde $\|\cdot\|$ es la norma espectral. En (Tropp, 2015) se demuestra que si el número de muestras es mayor a $c\varepsilon^{-2}n\log n$ el estimador muestral provee una relativa precisión en la estimación de la verdadera matriz de covarianza Σ. Pero con la gran cantidad de datos, generalmente las matrices son extremadamente grandes y con las facilidades en el censado y adquisición de datos, el tamaño tiende cada vez a aumentar.

Por limitaciones de volumen y costos es común no poder muestrear completamente los datos de salida para obtener la matriz completa, o en campos como la biología o medicina el número de muestras disponibles es reducido y generalmente menor que el tamaño de variables (Kwan, 2011). Por tanto se hace necesario desarrollar técnicas de estimación de matrices de covarianza capaces de lograr un bajo error sujetos a que el número de muestras k sea moderado comparado a n.

El estimador bien condicionado desarrollado por Ledoit y Wolf (2003, 2004) conocido como estimación *shrinkage* que al igual que el estimador muestral tiene la ventaja de no necesitar conocer la estructura de la matriz de covarianza y logra un mejor desempeño que el estimador muestral (Ledoit y Wolf, 2012). Otros métodos basados en teoría de matrices aleatorias, optimización convexa y muestreo compresivo han demostrado obtener una mejor estimación que los anteriores métodos, incluso cuando se tiene ruido (Bai y Shi, 2011; Walden y Schneider-Luftman, 2015). Además, sus resultados son mejores cuando se tiene en cuenta en la estimación la estructura que podría presentar la matriz de covarianza.

ESTIMACIÓN DE MATRICES BAJO RANGO

Las matrices de bajo rango son muy comunes debido a su aparición en diferentes campos: sistemas de identificación (Liu y Vandenberghe, 2009), restauración de películas (Yu et al., 2016), recuperación del movimiento en 3-D (Wang et al., 2016), ensamble de señales (Davies y Eldar, 2012), aprendizaje de maquina o machine learning (Argyriou et al., 2008; Guan et al., 2016).

Definición:(Matriz de bajo rango) Una matriz $X \in R^{mxn}$ se le llama de bajo rango si su rango r << min(m,n). Entonces la matriz puede ser expresada exactamente o aproximadamente mediante:

$$X \approx \sum_{i=1}^{r} \alpha_i \, u_i \, v_i^T, \tag{4}$$

donde $\alpha_i \geq 0$, i=1,…,r , $u_1,\ldots, u_r \in R^m$ y $v_1,\ldots, v_r$ R^n son vectores ortonormales. Una interpretación es que los escalares son los valores singulares de X y los vectores ortonormales son los correspondientes vectores singulares de la Descomposición en Valores Singulares o SVD (por sus siglas en ingles).

Los diferentes métodos desarrollados hasta el momento se basan en variaciones en el diseño del operador A, el cual se encarga de la toma de mediciones o muestras, por tanto encontramos (Davenport y Romberg, 2016): operador aleatorio, operador que retorna un subconjunto de las entradas y operador de rango uno. En lo que sigue, se listan los estimadores más representativos en la recuperación de matrices bajo rango.

Aproximación de bajo rango: La idea principal es encontrar la mejor aproximación de bajo rango de una matriz dada $X \in R^{mxn}$. Cuando se habla de la mejor aproximación significa que se desea determinar la matriz más cercana en norma de Frobenius, es decir:

$$\underset{X}{\text{minimize}} \ \|X_l - X\|_F^2 \quad \text{sujeto a rango}(X_l) \tag{5}$$

donde $\|\cdot\|_F^2$ es el cuadrado de la norma de Frobenius, y r es el rango deseado de la aproximación. A pesar que el problema es no convexo, una forma para solucionarlo de manera fácil es utilizar la SVD

$$X = U\Sigma V^T \tag{6}$$

donde U,V son matrices ortonormales y Σ es una matriz diagonal que contiene los valores singulares ordenados. La solución se encuentra truncando la expansión, es decir aplicando un umbral o *thresholding* a los valores singulares para mantener los más grandes y a partir de ellos se reconstruya la matriz (Horn y Johnson, 1994)

$$\hat{X} = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{7}$$

También se puede usar Lagrange en (5) para obtener:

$$\underset{X}{\text{minimize}} \ \|X_l - X\|_F^2 + \lambda \cdot \text{rango}(X) \tag{8}$$

donde el conjunto de soluciones obtenidas para $0<\lambda<\infty$ es igual al conjunto de soluciones obtenidas en (5) para $1 \leq r \leq k$ con k=min(m,n). Para un λ especifico, (7) se puede solucionar calculando la SVD y sometiendo los valores singulares a un umbral duro. Cuando el umbral es igual a $\lambda/2$ el resultado es igual a la solución de:

$$\underset{X}{\text{minimize}} \ \|X_l - X\|_F^2 + \lambda \cdot \|X\|_*, \tag{9}$$

donde $\|\cdot\|_*$ es la norma nuclear o norma traza (suma de valores singulares) que es una relajación convexa del rango, muy utilizada en problemas de optimización (Boyd y Vandenberghe, 2004).

Recuperación bajo rango y minimización por traza: Generalmente, en cambio de observar directamente la matriz X, lo que se hace es obtener mediciones de la forma y= $\mathcal{A}$(X)+η, donde η es un término de error acotado y $\mathcal{A}$: $R^{mxn} \to R^k$ es un operador lineal de medición que actúa sobre la matriz X tomando el producto interno estándar con matrices previamente establecidas $A_1,\ldots, A_k$. Así, cada componente del vector $y \in R^k$ viene dada por:

$$y_i = \langle A_i, X \rangle + n_i. \tag{10}$$

El objetivo es recuperar una buena aproximación de X con pocas medidas tanto como sean posibles. Una aproximación de bajo rango puede venir dada por

$$\underset{T}{\text{minimize}} \ \|y - \mathcal{A}(T)\|_F^2 \quad \text{sujeto a rango}(T) = r \tag{11}$$

pero el problema de minimización de rango es en general NP-difícil (Boyd y Vandenberghe, 2004). En cambio si se usa el programa de optimización convexa (9) se tiene

$$\underset{T}{\text{minimize}} \ \|y - \mathcal{A}(T)\|_2^2 + \lambda \|T\|_*, \tag{12}$$

que puede ser solucionado sistemáticamente mediante algoritmos como el proximal o FISTA (Davenport y Romberg, 2016). La soluciones a (12) para cualquier β>0 obedecerán la condición de punto fijo

$$T = \text{prox}_\beta(T - \beta \mathcal{A}^*(\mathcal{A}(T) - y)), \tag{13}$$

donde el operador proximal viene dado por

$$\text{prox}(M) = \underset{T}{\text{argmin}} \|T - M\|_F^2 + \beta\lambda \|T\|_*, \tag{14}$$

λ define el margen tolerable de error entre la solución y las mediciones y, así que si hay un nivel alto de fidelidad entonces se podría dejar λ =0 y replantear (12) como

$$\underset{T}{\text{minimize}} \ \|T\|_* \quad \text{sujeto a } \mathcal{A}(T) = y, \tag{15}$$

presentando la ventaja de ser una optimización convexa, ya que la matriz objetivo X pertenece a dos conjuntos convexos: el primero es la bola norma nuclear de radio $\gamma_x = \|X\|_*$, dada por

$$X \in \mathcal{C} = \{T : \|T\|_* \le \gamma_X\}, \tag{16}$$

y el segundo es el espacio afín de todas las matrices $m \times n$ que tienen las mismas mediciones

$$X \in \mathcal{E} := \{T : \mathcal{A}(T) = y\}, \tag{17}$$

entonces por definición X será la única solución si y sólo si es la única matriz en la intersección C∩Ɛ.

Recuperación de fase: Un problema que ha recibido gran atención ha sido la recuperación de la señal $x \in R^n$ a partir de *m* mediciones cuadráticas (intensidad) de la forma $x_i = |\langle x, a_i \rangle|$,i=1,…,m. Para recuperar la señal además de la magnitud se necesita determinar la fase, y este problema es NP-difícil (Sahinoglou y Cabrera, 1991). En un muy interesante trabajo conocido como *PhaseLift* desarrollado por Candès et al. (2013) demostraron que si los vectores Gaussianos a_i son muestreados independiente y uniformemente sobre la esfera unitaria entonces la señal puede ser recuperada exactamente a partir de las mediciones de magnitud. El problema original puede ser reescrito como

$$|\langle x, a_i \rangle|^2 = x^T a_i a_i^T x = \langle X, A_i \rangle, \tag{18}$$

donde $X=xx^T$ y $A_i=a_ia_i^T$ son matrices simétricas de rango 1. Entonces el problema puede ser convertido a un problema de recuperación de matrices

$$\underset{T}{\text{minimize}}\ \|T\|_* \quad \text{sujeto a } \langle X, A_i \rangle = y_i^2 \qquad T \pm 0 \tag{19}$$

y la estimación de la señal original x es exitosa cuando el número de muestras (mediciones) depende logarítmicamente de la dimensión, es decir m ≥ n log(n). Además, este concepto ha tenido gran impacto en la recuperación de imágenes, matrices y tensores.

Mediciones cuadráticas o sketch: En los anteriores estimadores los vectores de censado eran variables Gaussianas Independientes e Idénticamente Distribuidas (i.i.d.), en (Chen et al., 2015) se extendió el análisis a vectores de censado sub-Gaussianos, permitiendo utilizar una clase más amplia de mecanismos de muestreo. Además, utilizando los vectores de medición o *sketch* se obtiene mediciones cuadráticas a partir de las cuales demostraron que es posible la estimación de una matriz de covarianza. Cada medición cuadrática viene dada por:

$$y_i = a_i^T C_x a_i + \eta_i \tag{20}$$

y C_x es la matriz de covarianza. Cuando las mediciones han sido contaminadas por ruido acotado, en especial ruido gaussiano el problema de optimización puede ser rescrito como:

$$\hat{T} = \underset{T}{\text{argmin}}\ \text{rango}(T) \ \text{ sujeto a } T \pm 0 \qquad \|y - \mathcal{A}(T)\|_F^2 \le \varepsilon_1 \tag{21}$$

donde $\|\cdot\|$ es a norma espectral y ε_1 es la cota superior de $\|\eta\|_1$. Pero como hemos visto el problema de la minimización del rango es un problema NP-difícil, una relajación convexa que permite volver el problema tratable viene dada por:

$$\hat{T} = \underset{T}{\text{argmin}}\ \text{Tr}(T) \ \text{ sujeto a } T \pm 0 \qquad \|y - \mathcal{A}(T)\|_F^2 \le \varepsilon_1 \tag{22}$$

donde Tr(·) es la traza. Chen et al. (2015) demostraron que en ausencia de ruido la recuperación es exacta y el número de mediciones necesarias es del orden de nr. Cuando hay ruido demostraron que la estimación de la matriz es estable y el error está acotado por:

$$\|T - C_x\| = O\left(\frac{\varepsilon}{m}\right) \tag{23}$$

Sus resultados coinciden con los encontrados con *PhaseLift* y al recurrir al uso de los sketches permite recuperar matrices de covarianza sin la necesidad de almacenar todas las realizaciones del vector de variables aleatorias.

ESTIMACIÓN DE MATRICES DISPERSAS O SPARSE

Las matrices dispersas o *sparse* se caracterizan por ser matrices con muchas de sus entradas iguales cero; es decir que contiene muy pocos elementos diferentes de cero por columna. La relación entre el número de entradas cero y el total de entradas es conocida como el nivel de dispersión o *sparsity*. La importancia de estas matrices radica en el impacto que tiene en la reducción en la complejidad de los algoritmos, la computación y el almacenamiento de las matrices (Arora et al., 2006; Drineas y Zouzias, 2011; Astaiza-Hoyos et al., 2017), por lo cual se ha convertido en una área muy activa de investigación que ha dado origen a nuevos métodos de muestreo, estimación y aprendizaje (Chen et al., 2015; Dasarathy et al., 2015; Gilbert y Indyk, 2010). De forma general podemos clasificar los esquemas de estimación de matrices dispersas de como umbral o aleatorios.

Umbral o Thresholding

Uno de los métodos más fáciles para estimar matrices de covarianza dispersas ha sido el uso de un umbral, el cual establece elementos pequeños (menores al umbral) a cero (Bickel y Levina, 2008). Cada componente fuera de la diagonal principal de la matriz de covarianza Σ_k es sometido a un umbral especifico w_T, entonces la matriz dispersa viene dada por:

$$\hat{\Sigma}_k = (\hat{\sigma}_{ij})_{n\times n} = \begin{cases} \sigma_{ij} & \text{si } i = j \\ \sigma_{ij}\mathbf{1}(|\sigma_{ij}| > w_T) & \text{si } i \neq j \end{cases} \tag{24}$$

donde **1** es la función indicador que realiza el proceso de umbral

$$\mathbf{1}(|\sigma_{ij}| > w_T) = \begin{cases} x & \text{si } |x| > w_T \\ 0 & \text{si } |x| \leq w_T \end{cases} \tag{25}$$

a este umbral se le conoce como umbral duro. Cuando los datos son Gaussianos o sub-Gaussianos, el umbral puede ser escogido como:

$$w_T = c\sqrt{\frac{\log k}{n}},\ c > 0 \tag{26}$$

y una ventaja del umbral es que evita estimar elementos pequeños ayudando a reducir el nivel de ruido presente en los datos. Por otra parte, Rothman et al. (2009) combinaron el método de umbral y *shrinkage* para crear reglas de umbral más flexibles. Cai y Liu (2011) propusieron un umbral el cual es adaptativo a la variabilidad de las entradas y Cai y Yuan (2012) consideraron la estimación adaptativa de la matriz de covarianza utilizando un umbral en bloque. Pero la gran desventaja de este tipo de estimadores es que no garantiza que la matriz de covarianza estimada vaya a ser definida positiva.

Métodos Aleatorios

Dentro de las clasificaciones de métodos aleatorios se encuentran:

Producto de Tensores: Cuando se desea recuperar la matriz $T \in R^{nxn}$ de la matriz de mediciones

$$Y = ATB^T \in R^{m\times m} \tag{27}$$

donde A y B son matrices de censado o *Sketching* de mxn con m << n, formando un ensamble bipartita-δ, es decir consiste del conjunto de todas las matrices 0-1 que tienen al menos δ unos por columna. Si la matriz tiene filas y columnas con pocas entradas diferentes de cero, a esta estructura se le conoce como dispersión distribuida. (Dasarathy et al., 2015) mostraron que la estimación es posible resolviendo la optimización convexa

$$\underset{T}{\text{minimize}}\ \|T\| \quad \text{sujeto a } ATB^T = Y, \tag{28}$$

donde $\|\cdot\|_1$ es la norma de vectorizar (la vectorización es una transformación lineal que toma las columnas de una matriz y las apila una tras otra para obtener un vector columna) a T (vect(T)). El concepto de tensores aparece cuando reescribimos el producto matricial como:

$$y = (B \otimes A)x, \tag{29}$$

con x=vect(X) y y=vect(Y). Este tipo de aproximaciones presenta dos grandes desventajas ya que al hacer uso del producto Kronecher se incrementa la dimensión y se pierde algo de la estructura del problema, en especial es difícil hacer cumplir la propiedad de semi definida positiva en la forma vectorial. Entonces, la aproximación generalmente ignora este requerimiento y la matriz de covarianza estimada puede no ser una matriz de covarianza valida.

Minimización Alternante: Para solucionar (28) Bioucas-Dias et al. (2014) propusieron explotar la información a priori sobre la matriz de covarianza para recuperar la matriz a partir de mediciones compresivas. Desarrollaron un algoritmo llamado CoVALSA, el cual produce formulaciones semi definidas para matrices dispersas, Toeplitz, bajo rango y bajo rango permutadas, las cuales son solucionadas eficientemente por el algoritmo SALSA (Afonso et al., 2011). Sea $x \in R^n$ un vector aleatorio de media cero con matriz de covarianza X, la idea consiste en tomar mediciones compresivas dadas por:

$$Y = AXA^T \tag{30}$$

donde A ϵ R^{mxn} es la matriz de muestreo con m < n. Como el número de mediciones es menor que la dimensión del vector, entonces para poder recuperar la matriz de covarianza se debe tener en cuenta la estructura de la matriz, lo que les permitió plantear el siguiente problema de optimización

$$\underset{T}{\text{minimize}} \ \frac{1}{2}\left\|Y - ATA^T\right\|_F^2 + \lambda\phi(T) \quad \text{sujeto a } T \pm \mathbf{0} \tag{31}$$

donde ϕ es una función cerrada, propia y convexa que captura la estructura deseada de la matriz (bajo rango, sparse, Toeplitz). λ es un parámetro de regularización. La ventaja que esa aproximación presenta es que en el problema de optimización el espacio de matrices está limitado a semi-definidas positivas, lo cual nos asegurara que la matriz recuperada es una matriz de covarianza valida, y aunque Y es una matriz de menor dimensión se hace necesario un almacenamiento importante, por ejemplo muestran que para una matriz de dimensión n=100 se logra una buena recuperación si m=50, que en ultimas se pasa de tener una matriz de 100 x 100 a una matriz de 50 x 50, aunque una buena relación estamos interesados en buscar algoritmos que permitan una mejor reducción, y además de ser posible tratar que Y sea un vector y no una matriz para una computación más rápida.

Mediciones cuadráticas o sketch: El modelo de estimación desarrollado por Chen et al. (2015) también abarca el caso cuando la matriz de covarianza a estimar es dispersa. Si X es dispersa, es conocido que la norma ℓ_0 promueve la dispersión, entonces el problema de estimación puede ser planteado como

$$\begin{aligned} \hat{X} = \underset{T}{\text{argmin}} \|T\|_0 \quad & \text{sujeto a } T \pm 0 \\ \|y - \mathcal{A}(T)\|_1 \leq \varepsilon, & \end{aligned} \tag{32}$$

donde ε es una cota superior de $\|\eta\|_1$. Pero al utilizar la norma ℓ_0 el problema de minimización es NP-difícil (Foucart y Rauhut, 2013), es decir es intratable. Entonces en (Chen et al., 2015) se planteó utilizar una relajación convexa dada por

$$\begin{aligned} \hat{X} = \underset{T}{\text{argmin}} \|T\|_1 \quad & \text{sujeto a } T \pm 0 \\ \|y - \mathcal{A}(T)\|_1 \leq \varepsilon, & \end{aligned} \tag{33}$$

donde la norma ℓ_1 es la relajación convexa del tamaño de soporte y que ha demostrado ser exitosa en muchos algoritmos de censado compresivo promoviendo la dispersión (Candès, 2008; Foucart y Rauhut, 2013). Además, demostraron que cuando la matriz es exactamente p-dispersa y no hay presencia de ruido, la solución a (33) es exactamente igual a la verdadera siempre que el número de mediciones satisfaga k > p log (n^2/p). Inclusive, para matrices aproximadamente dispersas la estimación es robusta frente al ruido indicando que las mediciones cuadráticas son al menos tan buenas como las mediciones lineales.
.
La idea de sketch también ha sido utilizada en (Pedarsani et al., 2015) para la estimación de matrices dispersas mediante códigos de grafos dispersos, en (Bahmani y Romberg, 2015) para la estimación de matrices de bajo rango dispersas, o en el producto de tensores expuesto anteriormente. Otra forma de aproximarse a *sketching* es encontrar un pequeño subconjunto de columnas o filas de la matriz que aproximen la matriz entera, esto es conocido como Problema de Selección de Subconjunto Columna (Boutsidis et al., 2009, 2014; Drineas y Kannan, 2003). Un último enfoque ha sido Direcciones Frecuentes, el cual se realiza de forma determinística (Ghashami et al., 2016).

CONCLUSIONES

La estimación de la matriz de covarianza con un número de muestras menores que la dimensión ambiente u original se ha convertido en un campo de rápido desarrollo, ya que el estimador clásico presenta serias desventajas ante esta condición.

Estimadores basado en métodos conocidos como SVD logran una aceptable recuperación de la matriz al incluir información de la estructura de bajo rango que presenta la matriz de covarianza, mientras que nuevos métodos basado muestreo aleatorio y optimización convexa superan los desempeños de los métodos comunes, aún en presencia de alteraciones en las muestras. Además para reducir el tiempo de convergencia de los algoritmos se utilizan técnicas como métodos proximales.

Por otra parte, la estimación de matrices dispersas ha tomado un gran auge debido en parte al desarrollo de representaciones dispersas de las señales en bases o *frames*, y su fácil computación e almacenamiento. Algoritmos basados en el método de umbral logran satisfacer la condición de pocas muestras para su estimación, aunque su desempeño es superado por los nuevos algoritmos basados en matrices aleatorias y optimización convexa que al incorporar la condición de dispersión de la matriz logran reducir notablemente el número de muestras manteniendo un alto grado de precisión en la estimación, incluso en presencia de ruido.

A pesar de los grandes avances en el desarrollo de nuevos estimadores, temas relacionados al tiempo de computo de los programas convexos que utilizan los estimadores quedan por explorar, como también la inclusión de nueva información en la recuperación y el desarrollo de estimadores para matrices con una estructura conjuntamente dispersa y bajo rango.

AGRADECIMIENTOS

Los autores agradecen el apoyo de Colciencias a través de la convocatoria 528 y de la Universidad del Cauca.

REFERENCIAS

Afonso, M.V., J.M. Bioucas-Dias y M.A.T. Figueiredo, An Augmented Lagrangian Approach to the Constrained Optimization Formulation of Imaging Inverse Problems, IEEE Trans. Image Process, 20, 681-695 (2011)

Argyriou, A., T. Evgeniou y M. Pontil, Convex multi-task feature learning, Mach. Learn, 73, 243-272 (2008)

Arora, S., E. Hazan y S. Kale, A Fast Random Sampling Algorithm for Sparsifying Matrices. In Approximation, Randomization, and Combinatorial Optimization, Algorithms and Techniques, Springer, Berlin, Heidelberg, 272-279 (2006)

Astaiza-Hoyos, E., H. Bermúdez-Orozco y F. Muñoz, Sensado de Espectro Local de Banda Ancha para Radios Cognitivos Multi-antena basado en Compleción de Matrices y Muestreo Sub-Nyquist Uniforme en el Dominio Disperso, Información Tecnológica, 28(3), 185-196 (2017)

Bahmani, S. y J.K. Romberg, Sketching for simultaneously sparse and low-rank covariance matrices. In 6th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015, Cancun, Mexico, pp. 357–360, 13-16 de December (2015)

Bai, J. y S. Shi, Estimating High Dimensional Covariance Matrices and its Applications, Annals of Economics and Finance, 12, 199-215 (2011)

Bickel, P.J. y E. Levina, Covariance regularization by thresholding, The Annals of Statistics, 36, 2577-2604 (2008)

Bioucas-Dias, J.M., D. Cohen e Y.C. Eldar, Covalsa: Covariance estimation from compressive measurements using alternating minimization. In 2014 22nd European Signal Processing Conference (EUSIPCO), 999-1003 (2014)

Boutsidis, C., M. Mahoney y P. Drineas, An Improved Approximation Algorithm for the Column Subset Selection Problem. In Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, (Society for Industrial and Applied Mathematics), 968-977 (2009)

Boutsidis, C., P. Drineas y M. Magdon-Ismail, Near-Optimal Column-Based Matrix Reconstruction, SIAM J. Comput., 43, 687-717 (2014)

Boyd, S. y L. Vandenberghe, Convex Optimization, New York, NY, USA, Cambridge University Press (2004)

Bryan, K. y T. Leise, Making Do with Less: An Introduction to Compressed Sensing, SIAM Rev., 55, 547-566 (2013)

Cai, T. y W. Liu, Adaptive Thresholding for Sparse Covariance Matrix Estimation, ArXiv11022237 Stat (2011)

Cai, T.T. y M. Yuan, Adaptive covariance matrix estimation through block thresholding, Ann. Stat., 40, 2014-2042 (2012)

Candès, E.J., The restricted isometry property and its implications for compressed sensing, Comptes Rendus Math., 346, 589-592 (2008)

Candès, E.J., T. Strohmer y V. Voroninski, PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming. Commun. Pure Appl. Math., 66, 1241-1274 (2013)

Chen, Y., Y. Chi y A.J. Goldsmith, Exact and Stable Covariance Estimation from Quadratic Sampling via Convex Programming. Inf. Theory IEEE Trans. Appear (2015)

Dasarathy, G., P. Shah, B.N. Bhaskar y R.D. Nowak, Sketching Sparse Matrices, Covariances, and Graphs via Tensor Products. Inf. Theory IEEE Trans., On 61, 1373-1388 (2015)

Davenport, M.A. y J. Romberg, An Overview of Low-Rank Matrix Recovery From Incomplete Observations, IEEE J. Sel. Top. Signal Process, 10, 608-622 (2016)

Davies, M.E. e Y.C. Eldar, Rank Awareness in Joint Sparse Recovery, IEEE Trans. Inf. Theory, 58, 1135-1146 (2012)

Drineas, P. y R. Kannan, Pass Efficient Algorithms for Approximating Large Matrices. In Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, PA, USA, Society for Industrial and Applied Mathematics, pp. 223–232 (2003)

Drineas, P. y A. Zouzias, A note on element-wise matrix sparsification via a matrix-valued Bernstein inequality. Inf. Process. Lett., 111, 385-389 (2011)

Foucart, S. y H. Rauhut, A Mathematical Introduction to Compressive Sensing, New York, Birkhäuser (2013)

Ghashami, M., E. Liberty, J. Phillips y D. Woodruff, Frequent Directions: Simple and Deterministic Matrix Sketching, SIAM J. Comput., 45, 1762-1792 (2016)

Gilbert, A. y P. Indyk, Sparse Recovery Using Sparse Matrices. Proc. IEEE, 98, 937-947 (2010)

Guan, Y., T. Lu y otros cuatro autores, Efficient low-rank supported extreme learning machine for robust face recognition, In 2016 Visual Communications and Image Processing (VCIP), 1-4 (2016)

Horn, R.A. y C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press (1994)

Krim, H. y M. Viberg, Two decades of array signal processing research: the parametric approach. IEEE Signal Process. Mag., 13, 67-94 (1996)

Kwan, C., An Introduction to Shrinkage Estimation of the Covariance Matrix: A Pedagogic Illustration, Spreadsheets Educ. EJSiE, 4 (2011)

Ledoit, O. y M. Wolf, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, J. Empir. Finance, 10, 603-621 (2003)

Ledoit, O. y M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, J. Multivar. Anal., 88, 365-411 (2004)

Ledoit, O. y M. Wolf, Nonlinear shrinkage estimation of large-dimensional covariance matrices, The Annals of Statistics, 40, 1024-1060 (2012)

Leus, G. y Z. Tian, Recovering second-order statistics from compressive measurements. In 2011 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), pp. 337-340 (2011)

Liu, Z. y L. Vandenberghe, Interior-Point Method for Nuclear Norm Approximation with Application to System Identification, SIAM J. Matrix Anal. Appl., 31, 1235-1256 (2009)

Pedarsani, R., K. Lee y K. Ramchandran, Sparse covariance estimation based on sparse-graph codes. In 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), 612-619 (2015)

Rothman, A.J., E. Levina y J. Zhu, Generalized Thresholding of Large Covariance Matrices, J. Am. Stat. Assoc., 104, 177-186 (2009)

Sahinoglou, H. y S.D. Cabrera, On phase retrieval of finite-length sequences using the initial time sample. IEEE Trans. Circuits Syst., 38, 954-958 (1991)

Shannon, C.E., A Mathematical Theory of Communication. SIGMOBILE Mob Comput. Commun. Rev., 5, 3–55 (2001)

Tropp, J.A., An Introduction to Matrix Concentration Inequalities, Found. Trends® Mach, Learn, 8, 1-230 (2015)

Vershynin, R., How Close is the Sample Covariance Matrix to the Actual Covariance Matrix? J. Theor. Probab., 25, 655-686 (2012)

Walden, A.T. y D. Schneider-Luftman, Random Matrix Derived Shrinkage of Spectral Precision Matrices. IEEE Trans. Signal Process, 63, 4689-4699 (2015)

Wang, M., K. Li y otros tres autores, 3-D motion recovery via low rank matrix analysis. In 2016 Visual Communications and Image Processing (VCIP), 1-4 (2016)

Yu, B., Y. Ding, X. Huang y B. Wu, B., Corrupted old film sequences restoration using improved PatchMatch and low-rank matrix recovery. In 2016 International Conference on Audio, Language and Image Processing (ICALIP), 295-300 (2016)