

# Una breve introducción a la adquisición comprimida

por

Carlos Fernandez-Granda

RESUMEN. La teoría de la adquisición comprimida supuso una revolución en el procesamiento de señales, demostrando que las medidas aleatorias y la optimización convexa permiten adquirir y reconstruir señales compresibles de manera eficiente. En este artículo describimos algunos de los resultados fundamentales de esta teoría y los demostramos rigurosamente.

En el procesamiento de señales hay dos problemas fundamentales: (1) cómo muestrear las señales de manera que se preserve la información de interés; (2) cómo comprimir esta información para almacenarla eficientemente. El paradigma tradicional consiste en resolver estos dos problemas por separado. El objetivo que motiva la adquisición comprimida (*compressed sensing* o *compressive sensing* en inglés<sup>1</sup>) es combinar estos dos problemas: si una señal es compresible, deberíamos poder muestrearla más eficientemente. Hace unos quince años, Emmanuel Candès, David Donoho, Justin Romberg y Terence Tao demostraron que esto es posible en una serie de artículos de enorme impacto que revolucionaron el procesamiento de señales moderno [8, 9, 10, 11, 13]. Nuestro objetivo aquí es describir los principios básicos de la adquisición comprimida de una manera sencilla pero matemáticamente rigurosa. Para más información referimos al lector a los numerosos tutoriales existentes en la literatura (por ejemplo, [12]). Este artículo, escrito con ocasión de los Premios Princesa de Asturias,<sup>2</sup> está dedicado a Emmanuel Candès, mi director de tesis, a quien estoy enormemente agradecido por su apoyo todos estos años.

## 1. LOS DOS PILARES DE LA ADQUISICIÓN COMPRIMIDA

Nuestro objetivo es estudiar la adquisición de señales *compresibles*, que se pueden representar con un número reducido de parámetros. En la teoría básica de la adquisición comprimida modelamos las señales de interés como vectores *dispersos* (*sparse* en inglés) en  $\mathbb{R}^d$ , que sólo tienen  $s$  entradas no nulas ( $s < d$ ). Dichas señales dependen tan solo de  $2s$  parámetros: la posición de las entradas no nulas, y

<sup>1</sup>En español otros autores la llaman *percepción comprimida*.

<sup>2</sup>Como se recoge en el artículo *Huellas del Análisis de Fourier en el Premio Princesa de Asturias de Investigación Científica y Técnica 2020*, de Luz Roncal, que aparece en este mismo número de LA GACETA, la adquisición comprimida es una de las técnicas matemáticas que el jurado ha destacado en su motivación de la concesión de dicho premio a Emmanuel Candès, Ingrid Daubechies, Yves Meyer y Terence Tao.

sus valores. Las señales son muestreadas de manera lineal; modelamos las muestras correspondientes a un vector  $x^* \in \mathbb{R}^d$  como  $y := Ax^*$ , donde  $A$  es una matriz de dimensiones  $m \times d$  y  $m$  es el número de muestras. En este artículo consideramos dos preguntas fundamentales:

- ¿Es posible diseñar una matriz  $A$  que permita reconstruir *cualquier* vector con  $s$  entradas no nulas a partir de un número de muestras  $m$  *proporcional a  $s$* ?
- ¿Es posible realizar la reconstrucción de manera *eficiente*?

Una condición necesaria para que la reconstrucción sea posible es que ninguna de nuestras señales de interés esté en el núcleo de la matriz  $A$ . De hecho, para asegurarnos de que la reconstrucción sea robusta necesitamos que las medidas preserven la energía de cualquier señal dispersa. Esto motiva la definición de la propiedad de isometría restringida.

**DEFINICIÓN 1.1.** Una matriz  $A$  de dimensiones  $m \times d$ , con  $m < d$ , satisface la *propiedad de isometría restringida* con constante  $\varepsilon$  si, para todo vector  $x$  de dimensión  $d$  con  $s$  entradas no nulas, se cumple

$$(1 - \varepsilon) \|x\|_2 \leq \|Ax\|_2 \leq (1 + \varepsilon) \|x\|_2.$$

Como su nombre indica, una matriz con esta propiedad representa una aplicación que es aproximadamente isométrica si restringimos su acción al conjunto de vectores dispersos. Uno de los descubrimientos clave de la adquisición comprimida es que las matrices con entradas aleatorias satisfacen la propiedad de isometría restringida con alta probabilidad. Aquí consideramos un modelo canónico en la literatura, matrices con entradas gaussianas, pero se han obtenido resultados teóricos similares para otros modelos, como las matrices de Fourier aleatorias [11, 14].

**TEOREMA 1.2** (Propiedad de isometría restringida para matrices gaussianas). *Sea  $\mathbf{A} \in \mathbb{R}^{m \times d}$  una matriz con entradas independientes e idénticamente distribuidas que siguen una distribución normal estándar. La matriz  $\frac{1}{\sqrt{m}}\mathbf{A}$  satisface la propiedad de isometría restringida para una constante  $\varepsilon$  con probabilidad  $1 - \frac{C_2}{d}$  si el número de filas  $m$  cumple*

$$m \geq \frac{C_1 s}{\varepsilon^2} \log \left( \frac{d}{s} \right), \quad (1)$$

donde  $C_1, C_2 > 0$  son dos constantes.

Ignorando factores logarítmicos, el teorema establece que la matriz aleatoria es una isometría restringida si el número de medidas es proporcional a  $s$ , el número de entradas no nulas, y no depende directamente de la dimensión ambiente  $d$ .

El teorema 1.2 implica que la reconstrucción de un vector disperso a partir de medidas aleatorias es posible, porque no puede haber otro vector disperso que corresponda a los mismos datos. Sin embargo, esto no quiere decir que la reconstrucción sea computacionalmente eficiente. Otro descubrimiento clave en la adquisición comprimida es que la minimización en norma  $\ell_1$  produce una reconstrucción exacta. Este problema de optimización es convexo y puede resolverse de manera eficiente.

**TEOREMA 1.3** (Reconstrucción exacta por minimización en norma  $\ell_1$ ). Sean  $\mathbf{A} \in \mathbb{R}^{m \times d}$  una matriz con entradas independientes e idénticamente distribuidas que sigan una distribución normal estándar, y  $x^* \in \mathbb{R}^d$  un vector con  $s$  entradas no nulas. Si  $y := \mathbf{A}x^*$ , la probabilidad de que  $x^*$  sea la única solución al problema de minimización en norma  $\ell_1$

$$\min_{x \in \mathbb{R}^d} \|x\|_1 \quad \text{sujeto a} \quad \mathbf{A}x = y$$

es al menos  $1 - \frac{1}{d}$  siempre que el número de medidas satisfaga

$$m \geq Cs \log d,$$

donde  $C > 0$  es una constante.

El teorema 1.3 es esencialmente óptimo. Si conociésemos de antemano la posición de las  $s$  entradas no nulas del vector disperso necesitaríamos al menos  $s$  medidas para realizar la reconstrucción (invirtiendo la correspondiente submatriz de  $A$ ). Sorprendentemente, la minimización en norma  $\ell_1$  consigue reconstruir la señal a partir de un número de medidas que es proporcional a  $s$ , y no tiene dependencia directa con la dimensión ambiente  $d$  (ignorando factores logarítmicos).

El resto de este artículo contiene las demostraciones de los teoremas 1.2 (sección 2) y 1.3 (sección 4). En ellas introducimos algunas de las principales técnicas matemáticas desarrolladas en la teoría de la adquisición comprimida.

## 2. DEMOSTRACIÓN DE LA PROPIEDAD DE ISOMETRÍA RESTRINGIDA (TEOREMA 1.2)

La propiedad de isometría restringida se puede expresar en términos de los valores singulares de ciertas submatrices de  $\mathbf{A}$ . Si fijamos el conjunto  $\mathcal{S} \subset \{1, \dots, d\}$  de las  $s$  entradas no nulas del vector disperso, el mayor ( $\sigma_1$ ) y menor ( $\sigma_s$ ) valor singular de la submatriz  $\mathbf{A}_{\mathcal{S}}$  formada por las columnas correspondientes a los elementos de  $\mathcal{S}$  satisfacen

$$\sigma_s \leq \|\mathbf{A}_{\mathcal{S}}x_{\mathcal{S}}\|_2 = \|\mathbf{A}x\|_2 \leq \sigma_1,$$

para cualquier vector  $x$  cuyas entradas no nulas pertenezcan a  $\mathcal{S}$  ( $x_{\mathcal{S}} \in \mathbb{R}^s$  contiene las entradas no nulas de  $x$ ). Para demostrar el teorema 1.2 seguimos la estrategia propuesta en [1]: primero acotamos los valores singulares de una submatriz arbitraria pero fija, y luego extendemos las cotas a todas las posibles submatrices usando la desigualdad de Boole.<sup>3</sup>

Necesitamos controlar los valores singulares de matrices de dimensiones  $m \times s$  con entradas iid<sup>4</sup> gaussianas estándar. Numéricamente se observa que, si fijamos  $s$  y aumentamos  $m$ , los  $s$  valores singulares convergen a  $\sqrt{m}$ . Esto implica que

<sup>3</sup>Es la desigualdad que nos dice que para una familia finita o numerable de sucesos  $A_1, A_2, A_3, \dots$  se cumple  $P\left(\bigcup_n A_n\right) \leq \sum_n P(A_n)$ .

<sup>4</sup>A partir de ahora utilizaremos iid para abreviar «independientes e idénticamente distribuidas».

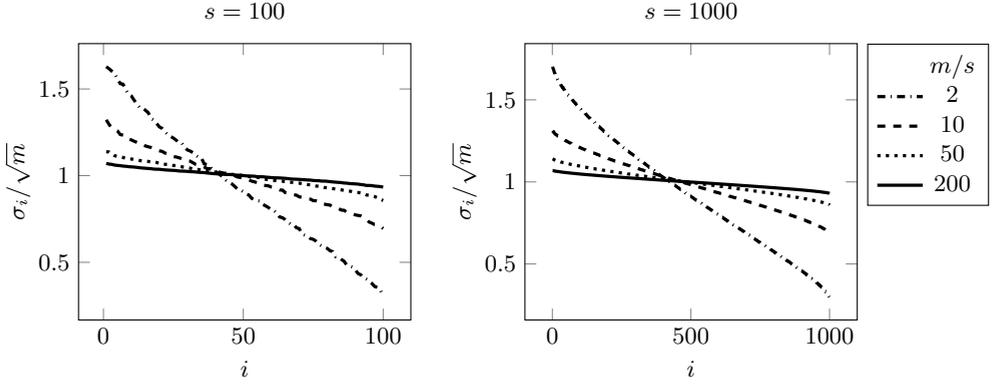


Figura 1: Valores singulares de matrices de dimensiones  $m \times s$  con entradas iid gaussianas estándar para diferentes valores de  $m$  y  $s$ .

la matriz es aproximadamente igual a  $\sqrt{m}UV^T$ , donde  $U$  y  $V$  son dos matrices ortogonales. Por consiguiente  $\mathbf{A}$  tiende a ser aproximadamente ortogonal (una vez la multiplicamos por  $1/\sqrt{m}$ ). Geométricamente, si generamos un número fijo de vectores gaussianos estándar en espacios de dimensión cada vez mayor, los vectores tenderán a ser casi ortogonales con alta probabilidad. El siguiente teorema establece cotas no asintóticas sobre los valores singulares usando un argumento basado en los números de recubrimiento propuesto en [1] (recomendamos al lector interesado en este tipo de técnicas el excelente libro [15]). La sección 3 contiene la demostración del siguiente resultado.

**TEOREMA 2.1** (Valores singulares de una matriz gaussiana). *Sea  $\mathbf{M}$  una matriz  $m \times s$  con entradas iid gaussianas estándar y  $m > s$ . Para cualquier constante  $\varepsilon > 0$  fija, los valores singulares de  $\mathbf{M}$  satisfacen*

$$\sqrt{m}(1 - \varepsilon) \leq \sigma_s \leq \sigma_1 \leq \sqrt{m}(1 + \varepsilon)$$

con probabilidad al menos  $1 - 2\left(\frac{12}{\varepsilon}\right)^s \exp\left(-\frac{m\varepsilon^2}{32}\right)$ .

Para cualquier submatriz  $\mathbf{A}_{\mathcal{S}}$  de  $\mathbf{A}$  con  $s$  columnas, el teorema implica que el mayor ( $\sigma_1$ ) y menor ( $\sigma_s$ ) valor singular de  $\mathbf{A}_{\mathcal{S}}$  satisfacen

$$\sqrt{m}(1 - \varepsilon) \leq \sigma_s \leq \sigma_1 \leq \sqrt{m}(1 + \varepsilon).$$

Por tanto, para cualquier vector  $x$  cuyas entradas no nulas están restringidas a  $\mathcal{S}$ ,

$$\sqrt{1 - \varepsilon} \|x\|_2 \leq \frac{1}{\sqrt{m}} \|\mathbf{A}x\|_2 \leq \sqrt{1 + \varepsilon} \|x\|_2, \quad (2)$$

pero esto no es suficiente para demostrar el teorema 1.2. Necesitamos establecer las desigualdades para *cualquier* conjunto de  $s$  entradas, es decir, para todas las combinaciones posibles de  $s$  columnas seleccionadas de entre las  $d$  columnas de  $\mathbf{A}$ .

Para ello aplicamos la desigualdad de Boole, usando la siguiente cota sobre el número de combinaciones:

$$\binom{d}{s} \leq \left(\frac{ed}{s}\right)^s.$$

Concluimos que todas las posibles submatrices satisfacen (2) para cualquier vector con  $s$  entradas no nulas con probabilidad, al menos,

$$\begin{aligned} & 1 - 2 \left(\frac{ed}{s}\right)^s \left(\frac{12}{\varepsilon}\right)^s \exp\left(-\frac{m\varepsilon^2}{32}\right) \\ &= 1 - \exp\left(\log 2 + s + s \log\left(\frac{d}{s}\right) + s \log\left(\frac{12}{\varepsilon}\right) - \frac{m\varepsilon^2}{2}\right) \leq 1 - \frac{C_2}{d} \end{aligned}$$

para una cierta constante  $C_2 > 0$ , siempre que  $m$  satisfaga (1).

### 3. VALORES SINGULARES DE MATRICES GAUSSIANAS (TEOREMA 2.1)

Para establecer las cotas sobre los valores singulares de la matriz  $\mathbf{M}$ , tenemos que demostrar que, para todo vector  $v$  con norma  $\ell_2$  igual a uno,

$$\sqrt{m}(1 - \varepsilon) < \|\mathbf{M}v\|_2 < \sqrt{m}(1 + \varepsilon).$$

Empecemos controlando  $\|\mathbf{M}v\|_2$  para un vector  $v$  fijo, usando la siguiente cota de Chernoff. Incluimos la demostración, aunque es bastante estándar, en el apartado 3.3.

**TEOREMA 3.1** (Cota de Chernoff sobre la norma  $\ell_2$  de un vector gaussiano). *Sea  $\mathbf{x}$  un vector aleatorio iid gaussiano estándar de dimensión  $m$ . Para todo  $\varepsilon \in (0, 1)$ ,*

$$P\left(m(1 - \varepsilon) < \|\mathbf{x}\|_2^2 < m(1 + \varepsilon)\right) \geq 1 - 2 \exp\left(-\frac{m\varepsilon^2}{8}\right).$$

Para un vector  $v \in \mathbb{R}^s$  fijo,  $\frac{1}{\sqrt{m}}\mathbf{M}v$  es un vector iid gaussiano estándar (el resultado de combinar linealmente vectores gaussianos es también gaussiano), lo cual nos permite acotar  $\|\mathbf{M}v\|_2$ .

**COROLARIO 3.2.** *Sea  $\mathbf{M}$  una matriz  $m \times s$  con entradas iid gaussianas estándar,  $m > s$ . Para cualquier vector  $v \in \mathbb{R}^s$  con norma  $\ell_2$  igual a uno y cualquier  $\varepsilon \in (0, 1)$ ,*

$$\sqrt{1 - \varepsilon} \leq \left\| \frac{1}{\sqrt{m}}\mathbf{M}v \right\|_2 \leq \sqrt{1 + \varepsilon}$$

con probabilidad al menos  $1 - 2 \exp(-m\varepsilon^2/8)$ .

El corolario no es suficiente para nuestros objetivos porque las cotas sólo se cumplen con la probabilidad indicada para un vector  $v$  fijo (aunque arbitrario), no para todos los vectores de la esfera  $\mathbb{S}^{s-1} := \{v \mid \|v\|_2 = 1\}$  a la vez. Una posible estrategia para usar el resultado es combinarlo con la desigualdad de Boole para extender las cotas a toda la esfera. ¡El problema es que la esfera es un conjunto de cardinalidad infinita, con lo cual la desigualdad de Boole es vacua! Por suerte, una sutil modificación permite sortear este obstáculo.

- Primero, establecemos las cotas en un subconjunto finito de la esfera  $\mathbb{S}^{s-1}$ , llamado  $\varepsilon$ -red ( $\varepsilon$ -net en inglés). La  $\varepsilon$ -red cubre  $\mathbb{S}^{s-1}$ ; todos los elementos de  $\mathbb{S}^{s-1}$  están a una distancia  $\varepsilon$  de algún elemento de la  $\varepsilon$ -red.
- Segundo, demostramos que las cotas pueden extenderse a todo punto que está suficientemente cerca de un elemento de la  $\varepsilon$ -red.

El resto de la sección implementa esta estrategia.

### 3.1. COTAS EN LA $\varepsilon$ -RED

Empezamos definiendo  $\varepsilon$ -red formalmente.

DEFINICIÓN 3.3. Un subconjunto  $\mathcal{N}_\varepsilon \subseteq \mathcal{X}$  es una  $\varepsilon$ -red de un conjunto  $\mathcal{X} \subseteq \mathbb{R}^s$  si para todo  $x \in \mathcal{X}$  existe al menos un vector  $y \in \mathcal{N}_\varepsilon$  que está a menos de  $\varepsilon$  de  $x$ ,

$$\|x - y\|_2 \leq \varepsilon.$$

El menor número posible de puntos en una  $\varepsilon$ -red de un conjunto se llama su *número de recubrimiento*.

DEFINICIÓN 3.4. El *número de recubrimiento*  $\mathcal{N}(\mathcal{X}, \varepsilon)$  de un conjunto  $\mathcal{X}$  a escala  $\varepsilon$  es la cardinalidad mínima de las  $\varepsilon$ -redes de  $\mathcal{X}$ , o, equivalentemente, el menor número de bolas de radio  $\varepsilon$  que necesitamos para cubrir  $\mathcal{X}$ .

El siguiente teorema acota el número de recubrimiento de la esfera.

TEOREMA 3.5 (Número de recubrimiento de la esfera). *El número de recubrimiento de la esfera  $s$ -dimensional  $\mathbb{S}^{s-1}$  a escala  $\varepsilon$  satisface*

$$\mathcal{N}(\mathbb{S}^{s-1}, \varepsilon) \leq \left(\frac{2 + \varepsilon}{\varepsilon}\right)^s \leq \left(\frac{3}{\varepsilon}\right)^s.$$

DEMOSTRACIÓN. Construimos una  $\varepsilon$ -red  $\mathcal{N}_\varepsilon \subseteq \mathbb{S}^{s-1}$  recursivamente. Inicializamos  $\mathcal{N}_\varepsilon$  como el conjunto vacío. Después elegimos un punto  $x \in \mathbb{S}^{s-1}$  tal que  $\|x - y\|_2 > \varepsilon$  para cualquier  $y \in \mathcal{N}_\varepsilon$  y añadimos  $x$  a  $\mathcal{N}_\varepsilon$ . Repetimos este proceso hasta que no quedan puntos en  $\mathbb{S}^{s-1}$  que estén a una distancia mayor que  $\varepsilon$  de todos los puntos de  $\mathcal{N}_\varepsilon$ . Este proceso termina necesariamente después de un número finito de repeticiones porque la esfera es un conjunto compacto (si no, habríamos generado una secuencia infinita sin subsecuencias convergentes).

Consideremos ahora las bolas de radio  $\varepsilon/2$  centradas en cada elemento de  $\mathcal{N}_\varepsilon$ . Estas bolas no se solapan, ya que sus centros están separados por una distancia de al menos  $\varepsilon$ . Además, todas se encuentran en una bola de radio  $1 + \varepsilon/2$  centrada en el origen porque sus centros pertenecen a  $\mathbb{S}^{s-1}$ . Por lo tanto,

$$\text{Vol}\left(\mathbb{B}_{1+\varepsilon/2}^s(\vec{0})\right) \geq \text{Vol}\left(\bigcup_{x \in \mathcal{N}_\varepsilon} \mathbb{B}_{\varepsilon/2}^s(x)\right) = |\mathcal{N}_\varepsilon| \text{Vol}\left(\mathbb{B}_{\varepsilon/2}^s(\vec{0})\right) \quad (3)$$

donde  $\mathbb{B}_r^s(x)$  es la bola de radio  $r$  centrada en  $x$ , y  $\vec{0}$  es el origen. Por cálculo multivariable básico tenemos

$$\text{Vol}\left(\mathbb{B}_r^s(\vec{0})\right) = r^s \text{Vol}\left(\mathbb{B}_1^s(\vec{0})\right),$$

con lo cual (3) implica

$$(1 + \varepsilon/2)^s \geq |\mathcal{N}_\varepsilon| (\varepsilon/2)^s. \quad \square$$

Ahora combinamos el corolario 3.2 con la desigualdad de Boole. Sean  $\varepsilon_1 := \varepsilon/4$  y  $\varepsilon_2 := \varepsilon/2$ . Consideramos una  $\varepsilon_1$ -red  $\mathcal{N}_{\varepsilon_1}$  de  $\mathbb{S}^{s-1}$ . Definimos el evento aleatorio

$$\mathcal{E}_{v,\varepsilon_2} := \left\{ m(1 - \varepsilon_2) \|v\|_2^2 \leq \|\mathbf{M}v\|_2^2 \leq m(1 + \varepsilon_2) \|v\|_2^2 \right\}.$$

Por el corolario 3.2, para todo  $v \in \mathbb{R}^s$  fijo,  $P(\mathcal{E}_{v,\varepsilon_2}^c) \leq 2 \exp(-m\varepsilon^2/32)$ . Aplicando la desigualdad de Boole, obtenemos

$$\begin{aligned} P\left(\bigcup_{v \in \mathcal{N}_{\varepsilon_1}} \mathcal{E}_{v,\varepsilon_2}^c\right) &\leq \sum_{v \in \mathcal{N}_{\varepsilon_1}} P(\mathcal{E}_{v,\varepsilon_2}^c) \\ &\leq |\mathcal{N}_{\varepsilon_1}| \max_{v \in \mathcal{N}_{\varepsilon_1}} P(\mathcal{E}_{v,\varepsilon_2}^c) \leq 2 \left(\frac{12}{\varepsilon}\right)^s \exp\left(-\frac{m\varepsilon^2}{32}\right). \end{aligned}$$

Por tanto

$$P\left(\bigcap_{v \in \mathcal{N}_{\varepsilon_1}} \mathcal{E}_{v,\varepsilon_2}\right) \geq 1 - 2 \left(\frac{12}{\varepsilon}\right)^s \exp\left(-\frac{m\varepsilon^2}{32}\right), \quad (4)$$

que es la cota para la probabilidad que buscamos en el teorema 2.1.

### 3.2. EXTENSIÓN DE LAS COTAS AL RESTO DE LA ESFERA

Para completar la demostración del teorema 2.1 basta con demostrar que en el evento  $\bigcap_{v \in \mathcal{N}_{\varepsilon_1}} \mathcal{E}_{v,\varepsilon_2}$  se cumplen las cotas para los valores singulares que indica el teorema. Para ello usaremos la desigualdad triangular y el hecho de que para el mayor ( $\sigma_1$ ) y menor valor singular ( $\sigma_s$ ) de  $\mathbf{M}$  se tiene

$$\sigma_1 = \max_{\|y\|_2=1} \|\mathbf{M}y\|_2, \quad (5)$$

$$\sigma_s = \min_{\|y\|_2=1} \|\mathbf{M}y\|_2. \quad (6)$$

Fijemos un vector arbitrario  $x \in \mathbb{S}^{s-1}$  sobre la esfera. Por definición de  $\varepsilon$ -red, existe un vector  $v$  en  $\mathcal{N}(\mathcal{X}, \varepsilon_1)$  que satisface  $\|x - v\|_2 \leq \varepsilon/4$ . La desigualdad triangular y (5) implican que, en el evento  $\mathcal{E}_{v,\varepsilon_2}$ ,

$$\begin{aligned} \|\mathbf{M}x\|_2 &\leq \|\mathbf{M}v\|_2 + \|\mathbf{M}(x - v)\|_2 \leq \sqrt{m} \sqrt{1 + \frac{\varepsilon}{2}} + \|\mathbf{M}(x - v)\|_2 \\ &\leq \sqrt{m} \left(1 + \frac{\varepsilon}{2}\right) + \|\mathbf{M}(x - v)\|_2 \leq \sqrt{m} \left(1 + \frac{\varepsilon}{2}\right) + \sigma_1 \|x - v\|_2 \\ &\leq \sqrt{m} \left(1 + \frac{\varepsilon}{2}\right) + \frac{\sigma_1 \varepsilon}{4}. \end{aligned} \quad (7)$$

Por (5),  $\sigma_1$  es la menor cota superior sobre  $\|\mathbf{M}x\|_2$  para todo  $x$  en la esfera, así que cuando estemos en  $\bigcap_{v \in \mathcal{N}_{\varepsilon_1}} \mathcal{E}_{v,\varepsilon_2}$  la cota (7) tiene que ser mayor que  $\sigma_1$ . Esto implica

$$\sigma_1 \leq \sqrt{m} \left(1 + \frac{\varepsilon}{2}\right) + \frac{\sigma_1 \varepsilon}{4},$$

lo cual nos permite concluir

$$\sigma_1 \leq \sqrt{m} \left( \frac{1 + \varepsilon/2}{1 - \varepsilon/4} \right) = \sqrt{m} \left( 1 + \varepsilon - \frac{\varepsilon(1 - \varepsilon)}{4 - \varepsilon} \right) \leq \sqrt{m} (1 + \varepsilon). \quad (8)$$

Obtenemos la cota inferior sobre  $\sigma_s$  combinando una idea muy similar con (8). Por la desigualdad triangular, (5) y (8), en el evento  $\mathcal{E}_{v, \varepsilon_2}$  se tiene

$$\begin{aligned} \|\mathbf{M}x\|_2 &\geq \|\mathbf{M}v\|_2 - \|\mathbf{M}(x - v)\|_2 \geq \sqrt{m} \sqrt{1 - \frac{\varepsilon}{2}} - \|\mathbf{M}(x - v)\|_2 \\ &\geq \sqrt{m} \left( 1 - \frac{\varepsilon}{2} \right) - \|\mathbf{M}(x - v)\|_2 \geq \sqrt{m} \left( 1 - \frac{\varepsilon}{2} \right) - \sigma_1 \|x - v\|_2 \\ &\geq \sqrt{m} \left( 1 - \frac{\varepsilon}{2} \right) - \frac{\varepsilon}{4} \sqrt{m} (1 + \varepsilon) = \sqrt{m} (1 - \varepsilon). \end{aligned} \quad (9)$$

A partir de (6) se llega, por un razonamiento similar al anterior usando ahora la desigualdad (9), a que  $\sigma_s \geq \sqrt{m} (1 - \varepsilon)$  siempre que estemos en  $\cap_{v \in \mathcal{N}_{\varepsilon_1}} \mathcal{E}_{v, \varepsilon_2}$ . Por la propiedad (4), esto concluye la demostración del teorema 2.1.

### 3.3. COTA DE CHERNOFF SOBRE LA NORMA $\ell_2$ DE UN VECTOR GAUSSIANO (TEOREMA 3.1)

Sea  $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$  e  $\mathbf{y} := \|\mathbf{x}\|_2^2$ . Las dos siguientes desigualdades implican el resultado deseado:

$$P(\mathbf{y} > m(1 + \varepsilon)) \leq \exp\left(-\frac{m\varepsilon^2}{8}\right), \quad (10)$$

$$P(\mathbf{y} < m(1 - \varepsilon)) \leq \exp\left(-\frac{m\varepsilon^2}{8}\right). \quad (11)$$

Demostramos (10); la demostración de (11) es muy similar. Sea  $t > 0$  una constante positiva. Por la desigualdad de Markov y la independencia de  $\mathbf{x}_1, \dots, \mathbf{x}_m$ ,

$$\begin{aligned} P(\mathbf{y} > a) &= P(\exp(t\mathbf{y}) > \exp(at)) \leq \exp(-at) E(\exp(t\mathbf{y})) \\ &= \exp(-at) E\left(\exp\left(\sum_{i=1}^m t\mathbf{x}_i^2\right)\right) = \exp(-at) \prod_{i=1}^m E(\exp(t\mathbf{x}_i^2)) \\ &= \frac{\exp(-at)}{(1 - 2t)^{\frac{m}{2}}}, \end{aligned} \quad (12)$$

donde el último paso es consecuencia de que, si  $\mathbf{z}$  es una variable gaussiana estándar y  $t < 1/2$ ,

$$E(\exp(t\mathbf{z}^2)) = \frac{1}{\sqrt{1 - 2t}}.$$

Fijamos  $a := m(1 + \varepsilon)$  y elegimos el valor de  $t \in (0, 1/2)$  que minimiza (12),

$$t := \frac{1}{2} - \frac{1}{2(1 + \varepsilon)},$$

para obtener

$$\begin{aligned} P(\mathbf{y} > m(1 + \varepsilon)) &\leq (1 + \varepsilon)^{\frac{m}{2}} \exp\left(-\frac{m\varepsilon}{2}\right) \\ &= \exp\left(-\frac{m}{2}(\varepsilon - \log(1 + \varepsilon))\right) \leq \exp\left(-\frac{m\varepsilon^2}{8}\right), \end{aligned}$$

donde el último paso es consecuencia de que la función  $g(x) := x - \frac{x^2}{4} - \log(1 + x)$  es no negativa entre 0 y 1.

#### 4. SUBGRADIENTES Y CERTIFICADOS DUALES

En esta sección presentamos un concepto clave en la adquisición comprimida: los certificados duales. En realidad, para medidas gaussianas esta técnica no es necesaria; es posible usar la propiedad de isometría restringida establecida en la sección 2 para demostrar el teorema 1.3 [2]. Sin embargo, hemos decidido describir la técnica de certificados duales, introducida inicialmente en [10], porque proporciona una mejor intuición geométrica sobre la minimización en norma  $\ell_1$ . Además, esta técnica se ha aplicado con éxito a varios problemas inversos donde las medidas no satisfacen la propiedad de isometría restringida como, por ejemplo, reconstrucción de matrices de rango pequeño con datos incompletos [6] y corruptos [5], o estimación de vectores a partir de medidas sin fase [3] o de baja resolución [4].

Como todas las normas, la norma  $\ell_1$  es una función convexa. Una estrategia razonable para caracterizar el mínimo de una función convexa diferenciable es usar su gradiente. Sin embargo, la norma  $\ell_1$  no es diferenciable en los puntos que tienen entradas iguales a cero (¡precisamente nuestro objeto de interés!). Por suerte, podemos recurrir a una generalización del gradiente: el subgradiente.

**DEFINICIÓN 4.1.** Un *subgradiente* de una función  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  en  $x \in \mathbb{R}^d$  es un vector  $g \in \mathbb{R}^d$  que satisface

$$f(x + h) \geq f(x) + g^T(h), \quad \text{para todo } h \in \mathbb{R}^d.$$

Cada subgradiente en un punto  $x$  está asociado a un hiperplano en el que se apoya la superficie definida por el grafo de la función, que es convexa en  $x$ . El siguiente teorema caracteriza los subgradientes de la norma  $\ell_1$ . La figura 2 muestra un ejemplo.

**TEOREMA 4.2** (Subgradientes de la norma  $\ell_1$ ). *Un vector  $g \in \mathbb{R}^d$  es un subgradiente de la norma  $\ell_1$  en  $x \in \mathbb{R}^d$  si y sólo si*

$$g_i = \text{signo}(x_i) \quad \text{si } x_i \neq 0, \tag{13}$$

$$|g_i| \leq 1 \quad \text{si } x_i = 0. \tag{14}$$

**DEMOSTRACIÓN.** Para no extendernos demasiado, sólo demostramos que un vector con esta estructura es un subgradiente, que es lo único que necesitamos.

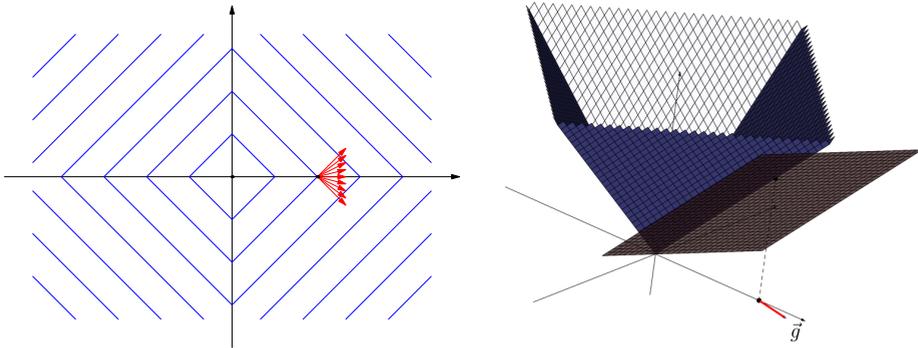


Figura 2: La imagen de la izquierda muestra líneas de contorno de la norma  $\ell_1$  en  $\mathbb{R}^2$ . Las flechas son subgradietes en un punto en el cual la función no es diferenciable. La imagen de la derecha muestra cómo la superficie correspondiente a la función se apoya en ese punto en el hiperplano (plano en este caso) correspondiente a uno de los subgradietes, denotado por  $\vec{g}$ .

La desigualdad de Hölder implica  $\|x + h\|_1 \geq \langle g, x + h \rangle$  para cualquier  $h$  (porque  $\|g\|_\infty \leq 1$ ). Por tanto,

$$\|x + h\|_1 \geq \langle g, x + h \rangle = \langle g, x \rangle + \langle g, h \rangle = \|x\|_1 + \langle g, h \rangle. \quad \square$$

Aquí olvidamos por un momento que nos interesan la matrices aleatorias y los vectores dispersos. Nos centramos en una situación genérica donde  $y := Ax^*$ , para una matriz  $A \in \mathbb{R}^{m \times d}$  y un vector  $x^* \in \mathbb{R}^d$  fijos arbitrarios. Nuestro objetivo es demostrar que la existencia de un cierto subgradiente implica que  $x^*$  es el único mínimo de la norma  $\ell_1$  en el conjunto  $\{x \in \mathbb{R}^d \mid Ax = y\}$ . La clave es considerar subgradietes en  $x^*$  que sean combinaciones lineales de las filas de  $A$ , es decir,  $g := A^T a$  para algún vector  $a \in \mathbb{R}^m$ . Si dicho  $g$  existe, para cualquier  $x$  que satisfaga  $Ax = y$  se tiene

$$\|x\|_1 \geq \|x^*\|_1 + \langle g, x - x^* \rangle = \|x^*\|_1 + \langle g, A(x - x^*) \rangle = \|x^*\|_1 + \langle g, y - y \rangle = \|x^*\|_1.$$

Esto es casi lo que necesitamos: la existencia del subgradiente implica que  $x^*$  es un mínimo. El siguiente teorema establece que una pequeña condición adicional sobre el subgradiente (que la desigualdad (14) sea estricta) nos asegura que el mínimo sea único.

**TEOREMA 4.3** (Certificado dual para la minimización en norma  $\ell_1$ ). *Sea  $A \in \mathbb{R}^{m \times d}$  y sea  $x^* \in \mathbb{R}^d$  con entradas no nulas en el conjunto de índices  $\mathcal{S}$  tal que  $Ax^* = y$ . Supongamos que la submatriz  $A_{\mathcal{S}}$  de columnas de  $A$  con índices en  $\mathcal{S}$  tenga rango completo. Si existe un vector  $a_{\text{cert}} \in \mathbb{R}^m$  tal que  $g_{\text{cert}} := A^T a_{\text{cert}}$  satisfaga*

$$(g_{\text{cert}})_i = \text{signo}(x_i^*) \quad \text{si } x_i^* \neq 0, \quad (15)$$

$$|(g_{\text{cert}})_i| < 1 \quad \text{si } x_i^* = 0, \quad (16)$$

entonces  $x^*$  es la única solución al problema de optimización

$$\min_{x \in \mathbb{R}^d} \|x\|_1 \quad \text{sujeto a} \quad Ax = y. \quad (17)$$

El vector  $a_{\text{cert}} \in \mathbb{R}^m$  se suele denominar *certificado dual* porque es una solución al problema dual de (17).

DEMOSTRACIÓN. Para cualquier  $x \in \mathbb{R}^d$  tal que  $Ax = y$ , sea  $h = x - x^*$ . Como  $A_{\mathcal{S}}$  tiene rango completo, el subvector  $h_{\mathcal{S}^c}$  de entradas de  $h$  con índice en  $\mathcal{S}^c$  cumple  $h_{\mathcal{S}^c} \neq 0$  a menos que  $h = 0$ , porque si no el subvector  $h_{\mathcal{S}}$  de entradas con índice en  $\mathcal{S}$  sería un vector no nulo en el núcleo de  $A_{\mathcal{S}}$ . La condición (16) implica

$$\|h_{\mathcal{S}^c}\|_1 > g_{\text{cert}}^T h_{\mathcal{S}^c}. \quad (18)$$

Sea  $\mathcal{P}_{\mathcal{S}}(\cdot)$  una proyección que convierte en cero todas las entradas de un vector excepto las que tienen índice en  $\mathcal{S}$ . Tenemos, por (18),

$$\begin{aligned} \|x\|_1 &= \|x^* + \mathcal{P}_{\mathcal{S}}(h)\|_1 + \|h_{\mathcal{S}^c}\|_1 \\ &> \|x^*\|_1 + g_{\text{cert}}^T \mathcal{P}_{\mathcal{S}}(h) + (A^T a_{\text{cert}})^T \mathcal{P}_{\mathcal{S}^c}(h) = \|x^*\|_1 + a_{\text{cert}}^T Ah = \|x^*\|_1. \quad \square \end{aligned}$$

## 5. CONSTRUCCIÓN DEL CERTIFICADO DUAL

El teorema 4.3 nos proporciona un plan para demostrar el teorema 1.3: construir un certificado dual adaptado a un vector  $x^*$  arbitrario con  $s$  entradas no nulas, y demostrar que satisface las condiciones con alta probabilidad si  $m$  es suficientemente grande. En esta sección describimos una construcción cuando la matriz de medidas  $\mathbf{A}$  tiene entradas iid gaussianas estándar. Concretamente, nuestro objetivo es usar las filas de  $\mathbf{A}$  para interpolar el signo de  $x^*$  en  $\mathcal{S}$  (el conjunto de índices donde  $x^*$  tiene entradas no nulas). Para ello seguimos la estrategia propuesta en [7].

Nuestra construcción se basa en una propiedad fundamental de las matrices aleatorias: sus columnas son prácticamente ortogonales si su dimensión es suficientemente grande, como establece el teorema 2.1. Consideremos el vector  $\mathbf{c}(i)$  que contiene el producto interior de la columna  $i$  de  $\mathbf{A}$ , que denominamos  $\mathbf{A}_i \in \mathbb{R}^m$ , con las demás columnas,

$$\mathbf{c}(i) := \mathbf{A}^T \mathbf{A}_i, \quad 1 \leq i \leq m.$$

El vector  $\mathbf{c}(i)$  es una combinación lineal de las filas de  $\mathbf{A}$ . Según el teorema 3.1, su entrada  $i$ -ésima cumple  $\mathbf{c}(i)_i = \|\mathbf{A}_i\|_2^2 \approx d$ . Las demás entradas son muy pequeñas con alta probabilidad por la ortogonalidad de las columnas. ¡Esto es perfecto para nuestro objetivo! Utilizamos las correlaciones  $\mathbf{c}(i)$ , correspondientes a  $i \in \mathcal{S}$ , para construir el certificado

$$\mathbf{g}_{\text{cert}} := \sum_{i \in \mathcal{S}} \mathbf{w}(i) \mathbf{c}(i).$$

Los coeficientes  $\mathbf{w}(i)$ ,  $i \in \mathcal{S}$ , deben ajustarse para que el certificado interpole el signo de  $x^*$  en  $\mathcal{S}$ , es decir, que las coordenadas de  $\mathbf{g}_{\text{cert}}$  satisfagan

$$(\mathbf{g}_{\text{cert}})_j = \text{signo}(x^*)_j \quad \text{para todo } j \in \mathcal{S}.$$

Si  $v_{\mathcal{S}}$  denota el subvector de entradas con índice en  $\mathcal{S}$  de cualquier vector  $v \in \mathbb{R}^d$ , las condiciones sobre los coeficientes son equivalentes al sistema de ecuaciones

$$\text{signo}(x^*)_{\mathcal{S}} = \left( \sum_{i \in \mathcal{S}} \mathbf{w}(i) \mathbf{c}(i) \right)_{\mathcal{S}} = \sum_{i \in \mathcal{S}} \mathbf{w}(i) \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_i = \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \mathbf{w}_{\text{cert}},$$

donde  $\mathbf{w}_{\text{cert}}$  contiene los  $s$  coeficientes  $\mathbf{w}(i)$ ,  $i \in \mathcal{S}$ . Resolviendo el sistema, tenemos

$$\mathbf{w}_{\text{cert}} := \left( \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \text{signo}(x^*)_{\mathcal{S}}.$$

El certificado (o mejor dicho, candidato a ser un certificado) es por tanto igual a

$$\mathbf{g}_{\text{cert}} = \sum_{i \in \mathcal{S}} \mathbf{w}(i) \mathbf{c}(i) = \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \mathbf{w}_{\text{cert}} = \mathbf{A}^T \mathbf{A}_{\mathcal{S}} \left( \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \text{signo}(x^*)_{\mathcal{S}}.$$

Para completar la demostración del teorema 1.3 necesitamos verificar que este vector existe y satisface las condiciones del teorema 4.3 con suficiente probabilidad. Sea  $\sigma_s$  el menor valor singular de  $\mathbf{A}_{\mathcal{S}}$ . Fijando  $\varepsilon := 0.5$  en el teorema 2.1, sea  $\mathcal{E}$  el evento aleatorio

$$0.5\sqrt{m} \leq \sigma_s \leq \sigma_1 \leq 1.5\sqrt{m}.$$

Entonces

$$\mathbb{P}(\mathcal{E}) \geq 1 - \exp\left(-C' \frac{m}{s}\right)$$

para una constante  $C' > 0$ . Condicionando sobre  $\mathcal{E}$ ,  $\mathbf{A}_{\mathcal{S}}$  es de rango completo y  $\mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}}$  es invertible, con lo cual  $\mathbf{g}_{\text{cert}}$  existe y satisface la condición (15).

Sólo nos queda comprobar que el certificado satisface la condición (16). Sea

$$\boldsymbol{\alpha}_{\text{cert}} := \mathbf{A}_{\mathcal{S}} \mathbf{w}_{\text{cert}} = \mathbf{A}_{\mathcal{S}} \left( \mathbf{A}_{\mathcal{S}}^T \mathbf{A}_{\mathcal{S}} \right)^{-1} \text{signo}(x^*)_{\mathcal{S}},$$

tal que  $\mathbf{g}_{\text{cert}} = \mathbf{A}^T \boldsymbol{\alpha}_{\text{cert}}$ , y sea  $\mathbf{U} \mathbf{S} \mathbf{V}^T$  la descomposición en valores singulares reducida de  $\mathbf{A}_{\mathcal{S}}$ . Condicionando en  $\mathcal{E}$ , tenemos

$$\|\boldsymbol{\alpha}_{\text{cert}}\|_2 = \|\mathbf{U} \mathbf{S}^{-1} \mathbf{V}^T \text{signo}(x^*)_{\mathcal{S}}\|_2 \leq \frac{\|\text{signo}(x^*)_{\mathcal{S}}\|_2}{\sigma_s} \leq 2 \sqrt{\frac{s}{m}}. \quad (19)$$

Para un índice fijo  $i \in \mathcal{S}^c$  y un vector fijo  $v \in \mathbb{R}^m$ ,  $\mathbf{A}_i^T v / \|v\|_2$  es una variable aleatoria gaussiana estándar, lo cual implica

$$\mathbb{P}(|\mathbf{A}_i^T v| \geq 1) = \mathbb{P}\left(\frac{|\mathbf{A}_i^T v|}{\|v\|_2} \geq \frac{1}{\|v\|_2}\right) \leq 2 \exp\left(-\frac{1}{2\|v\|_2^2}\right)$$

ya que, para una variable gaussiana estándar  $\mathbf{u}$ ,

$$\begin{aligned} \mathbb{P}(\mathbf{u} \geq t) &= \mathbb{P}(\exp(\mathbf{u}t) \geq \exp(t^2)) \leq \mathbb{E}(\exp(\mathbf{u}t - t^2)) \\ &= \exp\left(-\frac{t^2}{2}\right) \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-t)^2}{2}\right) dx = \exp\left(-\frac{t^2}{2}\right). \end{aligned}$$

Si  $i \notin \mathcal{S}$  entonces  $\mathbf{A}_i$  y  $\alpha_{\text{cert}}$  son independientes (dependen de diferentes columnas de  $\mathbf{A}$ ). Por consiguiente, (19) implica

$$P(|\mathbf{A}_i^T \alpha_{\text{cert}}| \geq 1 \mid \mathcal{E}) = P\left(|\mathbf{A}_i^T v| \geq 1 \mid \|v\|_2 \leq 2\sqrt{\frac{s}{m}}\right) \leq 2 \exp\left(-\frac{m}{8s}\right),$$

y, a su vez,

$$\begin{aligned} P(|\mathbf{A}_i^T \alpha_{\text{cert}}| \geq 1) &\leq P(|\mathbf{A}_i^T \alpha_{\text{cert}}| \geq 1 \mid \mathcal{E}) + P(\mathcal{E}^c) \\ &\leq 2 \exp\left(-\frac{m}{8s}\right) + \exp\left(-C' \frac{m}{s}\right). \end{aligned}$$

Para completar la demostración aplicamos la desigualdad de Boole para obtener una cota válida para todo  $i \in \mathcal{S}^c$ . Pero  $\mathcal{S}^c$  tiene como mucho  $d$  elementos, así que

$$P\left(\bigcup_{i \in \mathcal{S}^c} \{|\mathbf{A}_i^T \alpha_{\text{cert}}| \geq 1\}\right) \leq d \left(2 \exp\left(-\frac{m}{8s}\right) + \exp\left(-C' \frac{m}{s}\right)\right).$$

Podemos elegir por tanto una constante  $C$  tal que, si el número de medidas satisface

$$m \geq Cs \log d,$$

la minimización en norma  $\ell_1$  reconstruirá  $x^*$  exactamente con probabilidad  $1 - \frac{1}{d}$ .

AGRADECIMIENTOS. El autor agradece el apoyo de la National Science Foundation, a través del proyecto DMS 2009752.

## REFERENCIAS

- [1] R. BARANIUK, M. DAVENPORT, R. DEVORE Y M. WAKIN, A simple proof of the restricted isometry property for random matrices, *Constr. Approx.* **28** (2008), no. 3, 253–263.
- [2] E. J. CANDÈS, The restricted isometry property and its implications for compressed sensing, *C. R. Math. Acad. Sci. Paris* **346** (2008), no. 9-10, 589–592.
- [3] E. J. CANDÈS, Y. C. ELДАР, T. STROHMER Y V. VORONINSKI, Phase retrieval via matrix completion, *SIAM Rev.* **57** (2015), no. 2, 225–251.
- [4] E. J. CANDÈS Y C. FERNANDEZ-GRANDA, Towards a mathematical theory of super-resolution, *Comm. Pure Appl. Math.* **67** (2014), no. 6, 906–956.
- [5] E. J. CANDÈS, X. LI, Y. MA Y J. WRIGHT, Robust principal component analysis? *J. ACM* **58** (2011), no. 3, Art. 11, 37 pp.
- [6] E. J. CANDÈS Y B. RECHT, Exact matrix completion via convex optimization, *Found. Comput. Math.* **9** (2009), no. 6, 717–772.
- [7] E. CANDÈS Y B. RECHT, Simple bounds for recovering low-complexity models, *Math. Program.* **141** (2013), no. 1-2, Ser. A, 577–589.

- [8] E. J. CANDÈS, J. K. ROMBERG Y T. TAO, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inform. Theory* **52** (2006), no. 2, 489–509.
- [9] E. J. CANDÈS, J. K. ROMBERG Y T. TAO, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.* **59** (2006), no. 8, 1207–1223.
- [10] E. J. CANDÈS Y T. TAO, Decoding by linear programming, *IEEE Trans. Inform. Theory* **51** (2005), no. 12, 4203–4215.
- [11] E. J. CANDÈS Y T. TAO, Near-optimal signal recovery from random projections: universal encoding strategies?, *IEEE Trans. Inform. Theory* **52** (2006), no. 12, 5406–5425.
- [12] E. J. CANDÈS Y M. B. WAKIN, An introduction to compressive sampling, *IEEE Signal Process. Mag.* **25** (2008), no. 2, 21–30.
- [13] D. L. DONOHO, Compressed sensing, *IEEE Trans. Inform. Theory* **52** (2006), no. 4, 1289–1306.
- [14] M. RUDELSON Y R. VERSHYNIN, On sparse reconstruction from Fourier and Gaussian measurements, *Comm. Pure Appl. Math.* **61** (2008), no. 8, 1025–1045.
- [15] R. VERSHYNIN, Introduction to the non-asymptotic analysis of random matrices, *Compressed Sensing: Theory and Applications* (Y. Eldar y G. Kutyniok, eds.), 210–268, Cambridge University Press, 2012.

CARLOS FERNANDEZ-GRANDA, COURANT INSTITUTE OF MATHEMATICAL SCIENCES Y CENTER FOR DATA SCIENCE, NEW YORK UNIVERSITY

Correo electrónico: [cfgranda@cims.nyu.edu](mailto:cfgranda@cims.nyu.edu)

Página web: <https://cims.nyu.edu/~cfgranda/>