# Apéndice 1: Demostración del *Teorema 5[11]*.

*Teorema 5([11]):* (OMP con matriz de mediciones admisible): fijamos $\delta \in (0, 0.36)$ y elegimos $m \geq Ck \ln(n/\delta)$. Suponemos que x es una señal k-sparse en $\mathbb{R}^n$ y tomamos una matriz de medicones admisible Φ de dimensión $mxn$ independiente de la señal. Dado el dato y = Φx, OMP puede reconstruir la señal con una probabilidad mayor que $1-\delta$. La constante cumple $C \leq 20$. Para valores elevados de k, se puede reducir ese valor a $C \leq 4+\eta$ siendo η cualquier número positivo. [12].

*Prueba del Teorema 5([11]):* sin perder la generalidad, asumimos que las k primeras entradas de la señal original x son distintas de cero y las siguientes n-k entradas son cero. Luego, el vector y es una combinación lineal de las k primeras columnas de la matriz Φ. Representamos la matriz como $\Phi = [\Phi_{opt}|\Psi]$, por lo que $\Phi_{opt}$ tiene k columnas y Ψ tiene n-k columnas. El vector y = Φx es estadísticamente independiente de la matriz Ψ.

Considerar el evento $E_{succ}$ en el cual el algoritmo identifica correctamente la señal x tras k iteraciones. Sólo decrementamos la probabilidad de éxito si imponemos el requerimiento adicional de que el valor singular más pequeño de $\Phi_{opt}$ encuentre una cota inferior. Con este fin se define el evento

(65) $$\Sigma \overset{def}{=} \left\{\sigma_k\left(\Phi_{opt}\right) \geq 0.5\right\}$$

Aplicando la definición de probabilidad condicional,

(66) $$P(E_{succ}) \geq P(E_{succ} \cap \Sigma) = P(E_{succ} \mid \Sigma) \cdot P(\Sigma)$$

Para comprobar que $E_{succ}$ ocurre condicionadamente a ∑, es suficiente comprobar que el algoritmo identifica correctamente las columnas de $\Phi_{opt}$. Dichas columnas determinan qué entradas de la señal son distintas de cero.

A continuación mostramos cómo el algoritmo localiza dichas columnas. Para un vector $r \in \mathbb{R}^n$, el criterio empleado para la selección

(67) $$\rho(r) \overset{def}{=} \frac{\|\Psi^T r\|_\infty}{\|\Phi_{opt}^T r\|_\infty} = \frac{\max_\Psi |\langle \psi, r\rangle|}{\|\Phi_{opt}^T r\|_\infty}$$

donde la maximización se realiza sobre las columnas de Ψ. Si r es el vector residuo que aparece en el paso II del algoritmo, el algoritmo cogerá una columna de $\Phi_{opt}$ siempre que ρ(r) < 1. En el caso de que ρ(r) = 1, una columna óptima y otra no óptima consiguen el producto interior máximo, y el algoritmo no tiene forma de elegir una u otra por lo que la elección puede no ser correcta.

Imaginemos que podemos ejecutar k iteraciones del algoritmo OMP con la señal de entrada x y la matriz de mediciones reducida $\Phi_{opt}$ para obtener una secuencia de residuos $q_0,q_1,\ldots,q_{k-1}$ y una secuencia de índices de columnas $w_1,w_2,\ldots,w_k$. El algoritmo es determinista, por lo que estas secuencias son función de x y de $\Phi_{opt}$. En particular, los residuos son estadísticamente independientes de $\Psi$.

Ejecutamos OMP con la señal de entrada x y la matriz completa $\Phi$, obteniendo la secuencia real de residuos $r_0,r_1,\ldots,r_{k-1}$ y la secuencia real de índices de columna $\lambda_1,\lambda_2,\ldots,\lambda_k$. OMP reconstruye con éxito la señal x después de k iteraciones si y sólo si el algoritmo selecciona las k columnas de $\Phi_{opt}$ , en cualquier orden.

El algoritmo asegura que el residuo inicial satisface $q_0 = r_0$, por lo que la condición $\rho(q_0) < 1$ asegura que $\rho(r_0) < 1$. De forma que la iteración actual elige la columna $\lambda_1$ de $\Phi_{opt}$ cuyo producto interno con $r_0$ tiene el valor máximo. Por lo tanto, la iteración imaginaria elige la columna $w_1$ de $\Phi_{opt}$ cuyo producto interno con $q_0$ tiene el máximo valor. Evidentemente $\lambda_1 = w_1$.

Suponemos que durante las t primeras iteraciones, la ejecución real de OMP elige las mismas columnas que la ejecución imaginaria., esto es $\lambda_j = w_j$ para $j = 1,2\ldots t$. Los residuos actual e imaginario deben ser iguales al principio de la iteración t, es decir, $q_k = r_k$. Una consecuencia obvia es que $\rho(q_t) < 1$ implica que $\rho(r_t) < 1$. Repitiendo este argumento se establece que $\lambda_{t+1} = w_{t+1}$.

Concluimos que la probabilidad condicional satisface

(68) $$P\left(E_{succ} \mid \Sigma\right) \geq P\left(\max_t \rho\left(q_t\right) < 1 \mid \Sigma\right)$$

donde $\{q_t\}$ es una secuencia de k vectores aleatorios que pertenecen a las columnas de $\Phi_{opt}$ y que son estadísticamente independientes de $\Psi$.

Asumimos que $\sum$ ocurre. Para cada índice $t = 0,1,\ldots,k-1$, tenemos

(69) $$\rho\left(q_t\right) = \frac{\max_\Psi |\langle \psi, q_t \rangle|}{\| \Phi_{opt}^T q_t \|_\infty}$$

Como $\Phi_{opt}^T q_t$ es un vector de dimensión k

(70) $$\rho\left(q_t\right) \leq \frac{\sqrt{k}\, \max_\Psi |\langle \psi, q_t \rangle|}{\| \Phi_{opt}^T q_t \|_2}$$

Para simplificar esta expresión, definimos el vector

(71) $$u_t \overset{def}{=} \frac{0.5 q_t}{\| \Phi_{opt}^T q_t \|_2}$$

De la propiedad del valor singular más pequeño obtenemos la desigualdad

(72) $$\frac{\| \Phi_{opt}^T \|_2}{\| q \|_2} \geq \sigma_k\left(\Phi_{opt}\right) \geq 0.5$$

Para cualquier vector q en el rango de $\Phi_{opt}$. Como $u_t$ pertenece a este rango tendremos $\| u_t \|_2 \leq 1$. Resumiendo

(73) $$\rho\left(q_t\right) \leq 2\sqrt{k}\ \max_\Psi |\langle \psi, u_t \rangle|$$

para cada índice t. A cuenta de este hecho

(74) $$P\left\{\max_t \rho\left(q_t\right) < 1 | \Sigma\right\} \geq P\left\{\max_t \max_\Psi |\langle \psi, u_t \rangle| < \frac{1}{2\sqrt{k}} | \Sigma\right\}$$

Intercambiamos las dos maximizaciones y usando la independencia de las columnas de $\Psi$ obtenemos

(75) $$P\left\{\max_t \rho\left(q_t\right) < 1 | \Sigma\right\} \geq \prod_\Psi P\left\{\max_t |\langle \psi, u_t \rangle| < \frac{1}{2\sqrt{k}} | \Sigma\right\}$$

Como cada columna de la matriz $\Psi$ es independiente de {ut} y de $\sum$, la Propiedad 3 del apartado 2.4.2 establece una cota inferior para cada uno de los n-k términos que aparecen en el producto

(76) $$P\left\{\max_t \rho\left(q_t\right) < 1 | \Sigma\right\} \geq \left[1 - 2ke^{-cm/4k}\right]^{n-k}$$

De la Propiedad 4 y la ecuación (41) obtenemos

(77) $$P\left(\Sigma\right) = P\left\{\sigma_m\left(\Phi_{opt}\right) \geq 0.5\right\} \geq 1 - e^{-Cm}$$

Introduciendo (44) y sustituyendo el resultado obtenido en (42) llegamos a

(78) $$P\left(E_{succ}\right) \geq \left[1 - 2ke^{-Cm/4k}\right]^{n-k} \left[1 - e^{-Cm}\right]$$

Aplicando la desigualdad $(1-x)^k \geq 1 - kx$ valido para $k \geq 1$ y $x \leq 1$, obtenemos

(79) $$P\left(E_{succ}\right) \geq 1 - 2k(n-k)e^{-Cm/4k} - e^{-Cm}$$

Observamos que $k(n-k) \leq n^2/4$, luego

(80) $$P\left(E_{succ}\right) \geq 1 - n^2 e^{-Cm/k}$$

Como conclusión, la elección de $m \geq C \cdot k \cdot \log(n/\delta)$ es suficiente para reducir la probabilidad de fallo por debajo de $\delta$.