Mathematical Methods for Machine Learning and Signal Processing SS 2019

Lecture 11: Matching Pursuit and Orthogonal Matching Pursuit

Céline Aubel (corrections by V. Morgenshtern)

Let $\mathbf{D} \in \mathbb{C}^{K\times N}$ be a dictionary of $N > K$ vectors having unit ℓ_2-norm, i.e., $\|\mathbf{d}_\ell\|_2 = 1$ for all $\ell \in [\![1, N]\!]$. This dictionary is supposed to be complete, which means that it includes K linearly independent vectors that define a basis of the signal space $\mathbb{C}^K$.

1 Matching Pursuit (MP)

The Matching Pursuit (MP) algorithm was introduced by S. Mallat and Z. Zhang in 1993. It computes an s-term approximation of a signal $\mathbf{x} \in \mathbb{C}^K$ in $\mathbf{D}$, by iteratively selecting a single dictionary vector at a time.

Principle of the MP algorithm. The algorithm starts by projecting $\mathbf{x}$ orthogonally onto a vector $\mathbf{d}_{\ell_0}$, $\ell_0 \in [\![1, N]\!]$,

$$\mathbf{x} = \langle \mathbf{x}, \mathbf{d}_{\ell_0}\rangle \mathbf{d}_{\ell_0} + \mathbf{r}^{(1)},$$

$\mathbf{r}^{(1)}$ being the residual. Since $\mathbf{r}^{(1)}$ is orthogonal to $\mathbf{d}_{\ell_0}$, it follows from Pythagoras theorem that

$$\|\mathbf{x}\|_2^2 = |\langle \mathbf{x}, \mathbf{d}_{\ell_0}\rangle|^2 + \|\mathbf{r}^{(1)}\|_2^2.$$

In order to minimize $\|\mathbf{r}^{(1)}\|_2$, we must choose $\mathbf{d}_{\ell_0}$, $\ell_0 \in [\![1, N]\!]$, such that $|\langle \mathbf{x}, \mathbf{d}_{\ell_0}\rangle|$ is maximum[1], that is,

$$\ell_0 \in \arg\max_{\ell\in[\![1,N]\!]} |\langle \mathbf{x}, \mathbf{d}_\ell\rangle| .$$

The MP algorithm iterates this procedure by subdecomposing the residual. Assume that the mth residual $\mathbf{r}^{(m)}$ has already been computed for $m \geqslant 0$ (initially, we take $\mathbf{r}^{(0)} = \mathbf{x}$). Then, the next iteration chooses $\mathbf{d}_{\ell_m}$, $\ell_m \in [\![1, N]\!]$ such that[2]

$$\ell_m \in \arg\max_{\ell\in[\![1,N]\!]} \left|\left\langle \mathbf{r}^{(m)}, \mathbf{d}_\ell\right\rangle\right| ,$$

and projects $\mathbf{r}^{(m)}$ onto $\mathbf{d}_{\ell_m}$, thus defining the $(m + 1)$th residual $\mathbf{r}^{(m+1)}$,

$$\mathbf{r}^{(m)} = \left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m}\right\rangle \mathbf{d}_{\ell_m} + \mathbf{r}^{(m+1)}. \tag{1}$$

Again, the orthogonality of $\mathbf{r}^{(m+1)}$ and $\mathbf{d}_{\ell_m}$ implies that

$$\left\|\mathbf{r}^{(m)}\right\|_2^2 = \left|\left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m}\right\rangle\right|^2 + \left\|\mathbf{r}^{(m+1)}\right\|_2^2. \tag{2}$$

[1] In practice, it is sometimes computationally more efficient to find a vector $\mathbf{d}_{\ell_0}$ that is almost optimal, that is, $|\langle \mathbf{x}, \mathbf{d}_{\ell_0}\rangle| \geqslant \alpha \max_{\ell\in[\![1,N]\!]} |\langle \mathbf{x}, \mathbf{d}_\ell\rangle|$, where $\alpha \in (0, 1]$ is a relaxation factor. In this case, the algorithm is referred to as *weak MP*.

[2] This step becomes $\left|\left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m}\right\rangle\right| \geqslant \alpha \max_{\ell\in[\![1,N]\!]} \left|\left\langle \mathbf{r}^{(m)}, \mathbf{d}_\ell\right\rangle\right|$ for weak MP with relaxation parameter $\alpha \in (0, 1]$.

Summing (1) and (2) for m between 0 and $M-1$ gives

$$\mathbf{x} = \sum_{m=0}^{M-1} \left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m} \right\rangle \mathbf{d}_{\ell_m} + \mathbf{r}^{(M)} \tag{3}$$

and

$$\|\mathbf{x}\|_2^2 = \sum_{m=0}^{M-1} \left| \left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m} \right\rangle \right|^2 + \left\| \mathbf{r}^{(M)} \right\|_2^2. \tag{4}$$

Convergence of the MP algorithm. We can observe from (3) and (4) that the convergence of the MP algorithm depends on the rate of decay of the residual $\left\|\mathbf{r}^{(m)}\right\|_2^2$. The following theorem shows that it has an exponential decay.

Theorem 1 (Convergence of the MP algorithm)**.** The residual $\mathbf{r}^{(m)}$ computed by the MP algorithm satisfies[3]

$$\left\| \mathbf{r}^{(m)} \right\|_2^2 \leqslant \left(1 - \mu_{\min}^2(\mathbf{D})\right)^m \|\mathbf{x}\|_2^2,$$

where

$$\mu_{\min}(\mathbf{D}) = \inf_{\substack{\mathbf{r}\in\mathbb{C}^K \\ \mathbf{r}\neq\mathbf{0}}} \mu(\mathbf{r}, \mathbf{D}) > 0$$

and

$$\mu(\mathbf{r}, \mathbf{D}) = \max_{\ell\in[\![1,N]\!]} \left| \left\langle \frac{\mathbf{r}}{\|\mathbf{r}\|_2}, \mathbf{d}_\ell \right\rangle \right| \leqslant 1$$

is the coherence of the vector $\mathbf{r}$ relative to the dictionary. As a consequence,

$$\mathbf{x} = \sum_{m=0}^{\infty} \left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m} \right\rangle \mathbf{d}_{\ell_m} \quad \text{and} \quad \|\mathbf{x}\|_2^2 = \sum_{m=0}^{\infty} \left| \left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m} \right\rangle \right|^2.$$

Proof. For $k \in [\![0, m-1]\!]$, Pythagoras theorem stated in (2) implies that

$$\frac{\left\|\mathbf{r}^{(k+1)}\right\|_2^2}{\left\|\mathbf{r}^{(k)}\right\|_2^2} = 1 - \left| \left\langle \frac{\mathbf{r}^{(k)}}{\left\|\mathbf{r}^{(k)}\right\|_2}, \mathbf{d}_{\ell_k} \right\rangle \right|^2.$$

By definition of $\mathbf{d}_{\ell_k}$, it holds that

$$\frac{\left\|\mathbf{r}^{(k+1)}\right\|_2^2}{\left\|\mathbf{r}^{(k)}\right\|_2^2} = 1 - \max_{\ell\in[\![1,N]\!]} \left| \left\langle \frac{\mathbf{r}^{(k)}}{\left\|\mathbf{r}^{(k)}\right\|_2}, \mathbf{d}_\ell \right\rangle \right|^2 = 1 - \mu^2\left(\mathbf{r}^{(k)}, \mathbf{D}\right).$$

We have then

$$\frac{\left\|\mathbf{r}^{(k+1)}\right\|_2^2}{\left\|\mathbf{r}^{(k)}\right\|_2^2} \leqslant \sup_{\substack{\mathbf{r}\in\mathbb{C}^K \\ \mathbf{r}\neq\mathbf{0}}} \left(1 - \mu^2(\mathbf{r}, \mathbf{D})\right) = 1 - \inf_{\substack{\mathbf{r}\in\mathbb{C}^K \\ \mathbf{r}\neq\mathbf{0}}} \mu^2(\mathbf{r}, \mathbf{D}) = 1 - \mu_{\min}^2(\mathbf{D}).$$

Multiplying this last inequality for k between 0 and $m-1$ and using the fact that $\mathbf{r}^{(0)} = \mathbf{x}$ yields

$$\left\| \mathbf{r}^{(m)} \right\|_2^2 \leqslant \left(1 - \alpha^2\mu_{\min}^2(\mathbf{D})\right)^m \|\mathbf{x}\|_2^2.$$

[3] The residual satisfies $\left\| \mathbf{r}^{(m)} \right\|_2^2 \leqslant \left(1 - \alpha^2\mu_{\min}^2(\mathbf{D})\right)^m \|\mathbf{x}\|_2^2$ for weak MP with relaxation parameter $\alpha \in (0, 1]$.

We now have to verify that $\mu_{\min}(\mathbf{D}) > 0$. By way of contradiction, assume that $\mu_{\min}(\mathbf{D}) = 0$. Then, one can find a sequence $\{\mathbf{x}_n\}_{n\in\mathbb{N}}$ of $\mathbb{C}^K$ with $\|\mathbf{x}_n\|_2 = 1$ such that

$$\lim_{n\to\infty} \max_{\ell\in[\![1,N]\!]} |\langle \mathbf{x}_n, \mathbf{d}_\ell\rangle| = 0.$$

Since the unit sphere of $\mathbb{C}^N$ is compact, there exists a subsequence $\{\mathbf{z}_{n_k}\}_{k\in\mathbb{N}}$ of $\{\mathbf{z}_n\}_{n\in\mathbb{N}}$ that converges to a unit vector $\mathbf{z}\in\mathbb{C}^N$, $\|\mathbf{z}\|_2 = 1$. It follows that

$$\max_{\ell\in[\![1,N]\!]} |\langle \mathbf{z}, \mathbf{d}_\ell\rangle| = \lim_{k\to\infty} \max_{\ell\in[\![1,N]\!]} |\langle \mathbf{z}_{n_k}, \mathbf{d}_\ell\rangle| = 0.$$

Hence, $\langle \mathbf{z}, \mathbf{d}_\ell\rangle = 0$ for all $\ell \in [\![1,N]\!]$. Since $\{\mathbf{d}_\ell\}_{\ell\in[\![1,N]\!]}$ contains a basis for $\mathbb{C}^K$, necessarily, $\mathbf{z} = \mathbf{0}$, which contradicts the fact that $\|\mathbf{z}\|_2 = 1$. Therefore, $\mu_{\min}(\mathbf{D}) > 0$. As a result, $1-\alpha^2\mu_{\min}(\mathbf{D}) < 1$, implying that $\lim_{m\to\infty}\left\|\mathbf{r}^{(m)}\right\|_2 = 0$. We can conclude by noting that (3) and (4) becomes

$$\mathbf{x} = \sum_{m=0}^{\infty} \left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m}\right\rangle \mathbf{d}_{m_\ell} \quad \text{and} \quad \|\mathbf{x}\|_2^2 = \sum_{m=0}^{\infty} \left|\left\langle \mathbf{r}^{(m)}, \mathbf{d}_{\ell_m}\right\rangle\right|^2$$

as $M\to\infty$. □

2 Orthogonal Matching Pursuit (OMP)

At each iteration m, the vector $\mathbf{d}_{\ell_m}$ selected by the Matching Pursuit algorithm is a priori not orthogonal to the previously selected vectors $\mathbf{d}_{\ell_k}$, $k\in[\![0,m-1]\!]$. The Orthogonal Matching Pursuit (OMP) algorithm improves the MP algorithm by orthogonalizing the directions of projections.

The Gram-Schmidt algorithm orthogonalizes $\mathbf{d}_{\ell_m}$ with respect to $\{\mathbf{d}_{\ell_k}\}_{k\in[\![0,m-1]\!]}$,

$$\begin{aligned} \mathbf{u}_0 &= \mathbf{d}_{\ell_0} \\ \mathbf{u}_m &= \mathbf{d}_{\ell_m} - \sum_{k=0}^{m-1} \frac{\langle \mathbf{d}_{\ell_m}, \mathbf{u}_k\rangle}{\|\mathbf{u}_k\|_2^2}\mathbf{u}_k. \end{aligned} \tag{5}$$

The residual $\mathbf{r}^{(m)}$ is projected onto $\mathbf{u}_m$ instead of $\mathbf{d}_{\ell_m}$,

$$\mathbf{r}^{(m)} = \frac{\left\langle \mathbf{r}^{(m)}, \mathbf{u}_m\right\rangle}{\|\mathbf{u}_m\|_2^2}\mathbf{u}_m + \mathbf{r}^{(m+1)} = \frac{\mathbf{u}_m^H\mathbf{r}^{(m)}}{\|\mathbf{u}_m\|_2^2}\mathbf{u}_m + \mathbf{r}^{(m+1)} = \frac{\mathbf{u}_m\mathbf{u}_m^H}{\|\mathbf{u}_m\|_2^2}\mathbf{r}^{(m)} + \mathbf{r}^{(m+1)},$$

which yields the relation

$$\left(\mathbf{I}_K - \frac{\mathbf{u}_m\mathbf{u}_m^H}{\|\mathbf{u}_m\|_2^2}\right)\mathbf{r}^{(m)} = \mathbf{r}^{(m+1)}.$$

Since $\mathbf{r}^{(0)} = \mathbf{x}$ and the vectors $\mathbf{u}_m$ are orthogonal, it results that

$$\underbrace{\left(\mathbf{I}_K - \frac{\mathbf{u}_0\mathbf{u}_0^H}{\|\mathbf{u}_0\|_2^2}\right)\cdots\left(\mathbf{I}_K - \frac{\mathbf{u}_{M-1}\mathbf{u}_{M-1}^H}{\|\mathbf{u}_{M-1}\|_2^2}\right)\mathbf{x}}_{=(\mathbf{I}_K-\mathbf{P}_{\mathcal{V}_M})\mathbf{x}} = \mathbf{r}^{(M)},$$

where

$$\mathbf{P}_{\mathcal{V}_M} = \sum_{m=0}^{M-1} \frac{\mathbf{u}_m\left(\mathbf{u}_m\right)^H}{\|\mathbf{u}_m\|_2^2}$$

is the orthogonal projector onto the space $\mathcal{V}_M = \text{span}\{\mathbf{u}_m\}_{m\in[\![0,M-1]\!]}$. It follows that

$$\mathbf{x} = \mathbf{P}_{\mathcal{V}_M}\mathbf{x} + \mathbf{r}^{(M)}.$$

The Gram-Schmidt algorithm ensures that $\{\mathbf{d}_m\}_{m\in[\![0,M-1]\!]}$ is also a basis for $\mathcal{V}_M$. The residual $\mathbf{r}^{(M)}$ is the component of $\mathbf{x}$ that is orthogonal to $\mathcal{V}_M$. For $m = M$, (5) implies that

$$\left\langle \mathbf{r}^{(M)}, \mathbf{u}_M \right\rangle = \left\langle \mathbf{r}^{(M)}, \mathbf{d}_{l_M} \right\rangle.$$

Since $\mathcal{V}_M$ has dimension M, there exists $P \leqslant K$ such that $\mathbf{x} \in \mathcal{V}_P$. We have then $\mathbf{r}^{(P)} = 0$ and

$$\mathbf{x} = \sum_{m=0}^{P-1} \frac{\left\langle \mathbf{r}^{(m)}, \mathbf{d}_{l_m} \right\rangle}{\|\mathbf{u}_m\|_2^2}\mathbf{u}_m.$$

The algorithm stops after $P \leqslant K$ iterations. The energy conservation resulting from the decomposition is

$$\|\mathbf{x}\|_2^2 = \sum_{m=0}^{M-1} \frac{\left|\left\langle \mathbf{r}^{(m)}, \mathbf{d}_{l_m} \right\rangle\right|^2}{\|\mathbf{u}_m\|_2^2}.$$

To expand $\mathbf{x}$ over the original dictionary vectors $\mathbf{d}_m$, we can perform a change of basis. The OMP algorithm is summarized as Algorithm 1.

Algorithm 1: Orthogonal Matching Pursuit (OMP)

Input : $\mathbf{D}$, a dictionary in $\mathbb{C}^{K\times N}$,
$\mathbf{x}$, a signal in $\mathbb{C}^K$,
s, the sparsity level of the ideal signal.

Output: $\widehat{\boldsymbol{\alpha}}$, a sparse representation of the signal in $\mathbb{C}^N$,
$\mathcal{I}$, the support of the estimated signal, i.e., the set containing the position of the nonzero elements of $\hat{\boldsymbol{\alpha}}$.

Initialize the sparse representation $\boldsymbol{\alpha}^{(0)} = 0$, the index set $\mathcal{I}^{(0)} = \emptyset$, the matrix of chosen atoms $\mathbf{D}^{(0)} = [\,]$, and the iteration counter $t = 1$.
while $t < s$ **do**
  Calculate the residual:
  $\mathbf{r}^{(t)} = \mathbf{x} - \mathbf{D}^{(t-1)}\boldsymbol{\alpha}^{(t-1)}$.
  Find the index of the column of $\mathbf{D}$ that is most correlated with $\mathbf{r}^{(t)}$:
  $i^{(t)} = \underset{j=1,\dots,N}{\text{argmax}} \left|\left\langle \mathbf{r}^{(t)}, \mathbf{D}_j \right\rangle\right|$.
  If the maximum occurs for multiple indices, choose one arbitrarily.
  Augment the index set $\mathcal{I}^{(t)} = \mathcal{I}^{(t-1)} \cup \{i^{(t)}\}$
  and the matrix of chosen atoms $\mathbf{D}^{(t)} = \text{concat}(\mathbf{D}^{(t-1)}, \mathbf{D}_{i^{(t)}})$.
  Update the signal estimate by solving the least square problem:
  $\boldsymbol{\alpha}^{(t)} = \underset{\boldsymbol{\alpha}\in\mathbb{C}^N}{\text{argmin}} \|\mathbf{x} - \mathbf{D}^{(t)}\boldsymbol{\alpha}\|_2$, i.e., $\boldsymbol{\alpha}^{(t)} = {\mathbf{D}^{(t)}}^{\dagger}\mathbf{x}$.
  $t = t + 1$.
end while
$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^{(t)}$ and $\mathcal{I} = \mathcal{I}^{(t)}$.
return $\widehat{\boldsymbol{\alpha}}, \mathcal{I}$

3 Exact Recovery Conditions

Theorem 2. For $\Lambda \subseteq [\![1,N]\!]$, let $\{\tilde{\mathbf{d}}_\ell\}_{\ell\in\Lambda}$ be the canonical dual basis of $\{\mathbf{d}_\ell\}_{\ell\in\Lambda}$ in $\mathcal{V}_\Lambda = \operatorname{span}\{\mathbf{d}_\ell\}_\Lambda$ and define

$$\mathrm{ERC}(\Lambda) = \max_{m\in\Lambda^c} \sum_{\ell\in\Lambda} \left| \left\langle \tilde{\mathbf{d}}_\ell, \mathbf{d}_m \right\rangle \right|,$$

where $\Lambda^c = [\![1,N]\!] \setminus \Lambda$ denotes the complement of Λ. Let $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} \in \mathbb{C}^K$, where $\boldsymbol{\alpha} \in \mathbb{C}^N$, and denote $\mathcal{S} = \operatorname{supp}\alpha = \{\ell \in [\![1,N]\!] \colon |\alpha_\ell| \neq 0\}$. If the exact recovery condition (ERC),

$$\mathrm{ERC}(\mathcal{S}) < 1,$$

is satisfied, then the matching pursuit algorithm selects only vectors in $\{\mathbf{d}_\ell\}_{\ell\in\mathcal{S}}$ and the orthogonal matching pursuit algorithm recovers $\mathbf{x}$ with at most $|\mathcal{S}|$ iterations.

Proof. At each iteration m, the MP and OMP algorithms selects a vector $\mathbf{d}_\ell$ with $\ell \in \Lambda$ if and only if the correlation of the residual $\mathbf{r}^{(m)}$ with vectors indexed by the complement of Λ is smaller than the correlation with vectors indexed by Λ: $C(\mathbf{r}^{(m)}, \Lambda^c) < 1$, where we define the correlation of a vector $\mathbf{h}$ with vectors in Λ^c relative to Λ,

$$C(\mathbf{h}, \Lambda) = \frac{\max_{m\in\Lambda^c} |\langle \mathbf{h}, \mathbf{d}_m \rangle|}{\max_{\ell\in\Lambda} |\langle \mathbf{h}, \mathbf{d}_\ell \rangle|}.$$

Let us first prove that for all $\Lambda \subseteq [\![1,N]\!]$,

$$\sup_{\mathbf{h}\in\mathcal{V}_\Lambda} C(\mathbf{h}, \Lambda) \leqslant \mathrm{ERC}(\Lambda). \tag{6}$$

Let $\mathbf{D}_\Lambda^\dagger = (\mathbf{D}_\Lambda^H \mathbf{D}_\Lambda)^{-1} \mathbf{D}_\Lambda^H$ be the Moore-Penrose pseudo-inverse of $\mathbf{D}_\Lambda$. We know that $\mathbf{D}_\Lambda \mathbf{D}_\Lambda^\dagger = (\mathbf{D}_\Lambda^\dagger)^H \mathbf{D}_\Lambda^H$ is the orthogonal projector onto $\mathcal{V}_\Lambda$. Thus, if $\mathbf{h} \in \mathcal{V}_\Lambda$ and $m \in \Lambda^c$, it holds that

$$\begin{aligned}
|\langle \mathbf{h}, \mathbf{d}_m \rangle| = \left| \left\langle (\mathbf{D}_\Lambda^\dagger)^H \mathbf{D}_\Lambda^H \mathbf{h}, \mathbf{d}_m \right\rangle \right| &= \left| \left\langle \mathbf{D}_\Lambda^H \mathbf{h}, \mathbf{D}_\Lambda^\dagger \mathbf{d}_m \right\rangle \right| \leqslant \left\| \mathbf{D}_\Lambda^H \mathbf{h} \right\|_\infty \left\| \mathbf{D}_\Lambda^\dagger \mathbf{d}_m \right\|_1 \\
&\leqslant \left\| \mathbf{D}_\Lambda^H \mathbf{h} \right\|_\infty \max_{m'\in\Lambda^c} \left\| \mathbf{D}_\Lambda^\dagger \mathbf{d}_{m'} \right\|_1 .
\end{aligned}$$

Since $\widetilde{\mathbf{D}}$ is the dual basis of $\mathbf{D}$ in the space $\mathcal{V} = \operatorname{span}\{\mathbf{d}_\ell\}_{\ell\in\Lambda}$, we know that $\widetilde{\mathbf{D}}_\Lambda^H = \mathbf{D}_\Lambda^\dagger$. Therefore, it holds that

$$\mathrm{ERC}(\Lambda) = \max_{m'\in\Lambda^c} \sum_{\ell\in\Lambda} \left| \left\langle \tilde{\mathbf{d}}_\ell, \mathbf{d}_{m'} \right\rangle \right| = \max_{m'\in\Lambda^c} \left\| \widetilde{\mathbf{D}}_\Lambda^H \mathbf{d}_{m'} \right\|_1 = \max_{m'\in\Lambda^c} \left\| \mathbf{D}_\Lambda^\dagger \mathbf{d}_{m'} \right\|_1 .$$

Moreover, we have

$$\left\| \mathbf{D}_\Lambda^H \mathbf{h} \right\|_\infty = \max_{\ell\in\Lambda} |\langle \mathbf{h}, \mathbf{d}_\ell \rangle| .$$

As a consequence, it holds for all $\mathbf{h} \in \mathcal{V}_\Lambda$ that

$$\max_{m\in\Lambda^c} |\langle \mathbf{h}, \mathbf{d}_m \rangle| \leqslant \mathrm{ERC}(\Lambda) \max_{\ell\in\Lambda} |\langle \mathbf{h}, \mathbf{d}_\ell \rangle| ,$$

which proves (6).

We now prove the reverse inequality. Let $m_0 \in \Lambda^c$ be such that

$$m_0 \in \arg\max_{m\in\Lambda^c} \sum_{\ell\in\Lambda} \left| \left\langle \tilde{\mathbf{d}}_\ell, \mathbf{d}_m \right\rangle \right| .$$

Introducing

$$\mathbf{h} = \sum_{\ell\in\Lambda} \text{sign}\left(\left\langle \tilde{\mathbf{d}}_\ell, \mathbf{d}_{m_0}\right\rangle\right) \tilde{\mathbf{d}}_\ell$$

leads to

$$\text{ERC}(\Lambda) = \sum_{\ell\in\Lambda} \left|\left\langle \tilde{\mathbf{d}}_\ell, \mathbf{d}_{m_0}\right\rangle\right| = |\langle \mathbf{h}, \mathbf{d}_{m_0}\rangle| \leqslant \max_{m\in\Lambda^c} |\langle \mathbf{h}, \mathbf{d}_m\rangle| \leqslant C(\mathbf{h}, \Lambda^c) \max_{\ell\in\Lambda} |\langle \mathbf{h}, \mathbf{d}_\ell\rangle| .$$

Since $|\langle \mathbf{h}, \mathbf{d}_\ell\rangle| = \left|\text{sign}\left(\left\langle \tilde{\mathbf{d}}_\ell, \mathbf{d}_{m_0}\right\rangle\right)\right| = 1$, it results that $\text{ERC}(\Lambda) \leqslant C(\mathbf{h}, \Lambda^c)$ and therefore,

$$\text{ERC}(\Lambda) \leqslant \sup_{\mathbf{h}\in\mathcal{V}_\Lambda} C(\mathbf{h}, \Lambda^c),$$

which, combined with (6), shows that

$$\text{ERC}(\Lambda) = \sup_{\mathbf{h}\in\mathcal{V}_\Lambda} C(\mathbf{h}, \Lambda^c).$$

Now, to prove the claim of the theorem, suppose that $\mathbf{x} = \mathbf{r}^{(0)} \in \mathcal{V}_\mathcal{S}$ and $\text{ERC}(\mathcal{S}) < 1$. We prove by induction that the MP algorithm selects only vectors in $\{\mathbf{d}_\ell\}_{\ell\in\mathcal{S}}$. Suppose that the first $m < M$ vectors selected by the MP algorithm are in $\{\mathbf{d}_\ell\}_{\ell\in\mathcal{S}}$, and thus, that $\mathbf{r}^{(m)} \in \mathcal{V}_\mathcal{S}$. If $\mathbf{r}^{(m)} \neq \mathbf{0}$, then the condition $\text{ERC}(\mathcal{S}) < 1$ implies that $C(\mathbf{r}^{(m)}, \mathcal{S}^c) < 1$ and thus the next vector is selected in $\mathcal{S}$. Since $\dim(\mathcal{V}_\mathcal{S}) \leqslant |\mathcal{S}|$, the OMP algorithm converges in less that $|\mathcal{S}|$ iterations. In the $|\mathcal{S}|$th step, we are left with $\mathbf{r}^{(|\mathcal{S}|)} = (\mathbf{I} - \mathbf{P}_{\mathcal{V}_{|\mathcal{S}|-1}})\mathbf{x} = \mathbf{0}$ and hence, the algorithm stops.
□

Proposition 3.1. For any $\Lambda \subseteq [\![1, N]\!]$, we have that

$$\text{ERC}(\Lambda) \leqslant \frac{|\Lambda|\mu(\mathbf{D})}{1 - (|\Lambda| - 1)\mu(\mathbf{D})}.$$

Proof. We have shown in the proof of Theorem 2 that

$$\text{ERC}(\Lambda) = \max_{m\in\Lambda^c} \left\|\mathbf{D}_\Lambda^\dagger \mathbf{d}_m\right\|_1 .$$

Since $\mathbf{D}_\Lambda^\dagger = (\mathbf{D}_\Lambda^H \mathbf{D}_\Lambda)^{-1}\mathbf{D}_\Lambda^H$, we have

$$\text{ERC}(\Lambda) = \max_{m\in\Lambda^c} \left\|(\mathbf{D}_\Lambda^H \mathbf{D}_\Lambda)^{-1}\mathbf{D}_\Lambda^H \mathbf{d}_m\right\|_1 \leqslant \left\|(\mathbf{D}_\Lambda^H \mathbf{D}_\Lambda)^{-1}\right\|_{1\to1} \max_{m\in\Lambda^c} \left\|\mathbf{D}_\Lambda^H \mathbf{d}_m\right\|_1 , \tag{7}$$

where we used the matrix norm $\|\cdot\|_{1\to1}$ defined as

$$\|\mathbf{A}\|_{1\to1} = \max_{\substack{\mathbf{u}\in\mathbb{C}^N\\ \mathbf{u}\neq\mathbf{0}}} \frac{\|\mathbf{A}\mathbf{u}\|_1}{\|\mathbf{u}\|_1} = \max_{\ell\in[\![1,N]\!]} \|\mathbf{a}_\ell\|_1 \tag{8}$$

for a matrix $\mathbf{A} = \{\mathbf{a}_\ell\}_{\ell\in[\![1,N]\!]}$. The second term of the upper bound in (7) equals

$$\max_{m\in\Lambda^c} \left\|\mathbf{D}_\Lambda^H \mathbf{d}_m\right\|_1 = \max_{m\in\Lambda^c} \sum_{\ell\in\Lambda} |\langle \mathbf{d}_m, \mathbf{d}_\ell\rangle| .$$

By definition of the coherence $\mu(\mathbf{D})$ of the dictionary $\mathbf{D}$, each term $|\langle \mathbf{d}_m, \mathbf{d}_\ell \rangle|$ is smaller than $\mu(\mathbf{D})$. Therefore,

$$\max_{m \in \Lambda^c} \left\| \mathbf{D}_\Lambda^H \mathbf{d}_m \right\|_1 \leqslant \max_{m \in \Lambda^c} \sum_{\ell \in \Lambda} \mu(\mathbf{D}) = |\Lambda| \mu(\mathbf{D}). \tag{9}$$

For the first term of the upper bound in (7), we can use the Neumann theorem to write that

$$\left\| (\mathbf{D}_\Lambda^H \mathbf{D}_\Lambda)^{-1} \right\|_{1\to 1} \leqslant \sum_{k=0}^{\infty} \left\| \mathbf{I}_{|\Lambda|} - \mathbf{D}_\Lambda^H \mathbf{D}_\Lambda \right\|_{1\to 1}^k = \frac{1}{1 - \left\| \mathbf{I}_{|\Lambda|} - \mathbf{D}_\Lambda^H \mathbf{D}_\Lambda \right\|_{1\to 1}}.$$

Given that $\|\mathbf{d}_\ell\|_2 = 1$ for all $\ell \in [\![1, N]\!]$, we have

$$\left\| \mathbf{I}_{|\Lambda|} - \mathbf{D}_\Lambda^H \mathbf{D}_\Lambda \right\|_{1\to 1} = \max_{\ell' \in \Lambda} \sum_{\substack{\ell \in \Lambda \\ \ell \neq \ell'}} |\langle \mathbf{d}_\ell, \mathbf{d}_{\ell'} \rangle| \leqslant \max_{\ell' \in \Lambda} \sum_{\substack{\ell \in \Lambda \\ \ell \neq \ell'}} \mu(\mathbf{D}) = \mu(\mathbf{D})(|\Lambda| - 1). \tag{10}$$

Combining (7), (9), and (10) gives

$$\mathrm{ERC}(\Lambda) \leqslant \frac{|\Lambda| \mu(\mathbf{D})}{1 - (|\Lambda| - 1)\mu(\mathbf{D})}.$$

□

Corollary 3.1. Let $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} \in \mathbb{C}^K$, where $\boldsymbol{\alpha} \in \mathbb{C}^N$ and denote $\mathcal{S} = \operatorname{supp} \alpha$. If the condition

$$|\mathcal{S}| < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right)$$

is satisfied, then the orthogonal matching pursuit algorithm recovers $\mathbf{x}$ in less than $|\mathcal{S}|$ iterations.

Proof. If

$$|\mathcal{S}| < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right),$$

then using Proposition 3.1, we have that

$$\mathrm{ERC}(\mathcal{S}) \leqslant \frac{|\mathcal{S}| \mu(\mathbf{D})}{1 - (|\mathcal{S}| - 1)\mu(\mathbf{D})} < 1,$$

and thus, we can invoke Theorem 2 to complete the proof. □