Dictionary-Sparse Recovery via Thresholding-Based Algorithms

Simon Foucart*— University of Georgia

Abstract

It is shown that the iterative hard thresholding and hard thresholding pursuit algorithms provide the same theoretical guarantees as ℓ_1-minimization for the recovery from imperfect compressive measurements of signals that have almost sparse analysis expansions in a fixed dictionary. Unlike other signal space algorithms targeting the recovery of signals with sparse synthesis expansions, the ability to compute (near) best approximations by synthesis-sparse signals is not necessary. The results are first established for tight frame dictionaries, before being extended to arbitrary dictionaries modulo an adjustment of the measurement process.

Key words and phrases: compressive sensing, sparse recovery, iterative hard thresholding, hard thresholding pursuit, restricted isometry property adapted to a dictionary, tight frames.

AMS classification: 42C15, 65F10, 94A12.

1 Introduction

This note deals with the recovery of signals with sparse expansions in a fixed dictionary when these signals are acquired through compressive linear measurements. The dictionary is represented by a matrix $\mathbf{D} \in \mathbb{C}^{n \times N}$, where n is the dimension of the signal space identified to $\mathbb{C}^n$ and $N > n$ is the dimension of the coefficient space $\mathbb{C}^N$. On first instance, we suppose that the dictionary is a tight frame, i.e., that

$$\mathbf{D}\mathbf{D}^* = \mathbf{I}_n.$$

This assumption will be removed in Section 4. The signals $\mathbf{f} \in \mathbb{C}^n$ of interest are assumed to be (nearly) s-sparse with respect to $\mathbf{D}$, which may have one of the two meanings:

- it is synthesis-sparse, i.e., $\mathbf{f} = \mathbf{D}\mathbf{x}$ for some s-sparse $\mathbf{x} \in \mathbb{C}^N$,
- it is analysis-sparse, i.e., $\mathbf{D}^*\mathbf{f} \in \mathbb{C}^N$ is s-sparse.

*S. F. partially supported by NSF grant number DMS-1120622

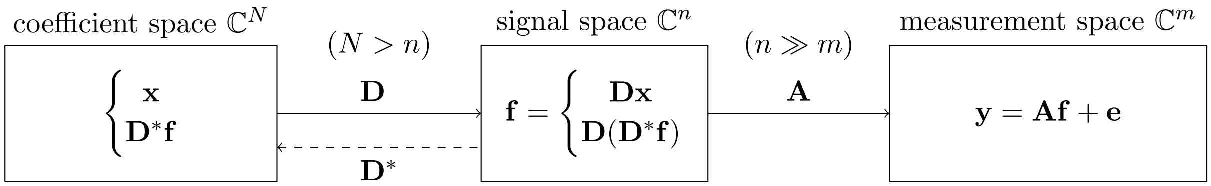

Figure 1: The coefficient, signal, and measurement spaces

The set of analysis-sparse vectors is contained in the set of synthesis-sparse vectors by virtue of $\mathbf{f} = \mathbf{D}(\mathbf{D}^*\mathbf{f})$. We target the stable and robust recovery of vectors in (the vicinity of) the smaller set of analysis-sparse signals. The recovery method uses some linear information about $\mathbf{f}$ arranged as a measurement vector $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e} \in \mathbb{C}^m$ with $m \ll n$. Here, $\mathbf{A} \in \mathbb{C}^{m\times n}$ is called the measurement matrix and $\mathbf{e} \in \mathbb{C}^m$ represents a measurement error which equals the zero vector only in an idealized setting but for which a bound $\|\mathbf{e}\|_2 \le \eta$ is typically available. The situation is summarized in Figure 1.

In [2], the presumably first theoretical result on the recovery of dictionary-sparse vectors involved the ℓ_1-minimizer

$$\mathbf{f}^\sharp = \operatorname*{argmin}_{\mathbf{g}} \{\|\mathbf{D}^*\mathbf{g}\|_1 \quad \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{g}\|_2 \le \eta\}\,.$$

For a tight frame $\mathbf{D}$, with $\sigma_s(\mathbf{D}^*\mathbf{f})_1$ representing the error of best s-term approximation to $\mathbf{D}^*\mathbf{f}$ in ℓ_1^N, it was established that

$$\|\mathbf{f} - \mathbf{f}^\sharp\|_2 \le c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}} + d\,\eta \tag{1}$$

is valid for all $\mathbf{f}$ and all $\mathbf{e}$ with $\|\mathbf{e}\|_2 \le \eta$ as soon as a restricted isometry condition for $\mathbf{A}$ adapted to $\mathbf{D}$ holds. Precisely, the condition reads $\delta_{2s}^{\mathbf{D}} < \delta_* := 0.08$, where $\delta_t^{\mathbf{D}}$ is the smallest constant $\delta \ge 0$ such that

$$\text{(D-RIP)} \quad (1-\delta)\|\mathbf{g}\|_2^2 \le \|\mathbf{A}\mathbf{g}\|_2^2 \le (1+\delta)\|\mathbf{g}\|_2^2 \qquad \text{whenever } \mathbf{g} = \mathbf{D}\mathbf{z} \text{ for some } s\text{-sparse } \mathbf{z} \in \mathbb{C}^N.$$

This condition is fulfilled with overwhelming probability by $m \times n$ matrices populated by iid subgaussian entries with variance $1/m$ provided $m \ge c(\delta_*)s\ln(eN/s)$. It is also fulfilled with high probability for random partial Fourier matrices after sign randomization of their columns, see [10]. The result (1) was extended in [6] to reconstruction errors measured at the coefficient level in ℓ_p for any $p \in [1,2]$[1], to loose (i.e., non-tight) frames provided the Moore–Penrose pseudoinverse $\mathbf{D}^\dagger$ replaces $\mathbf{D}^*$, and to pregaussian (aka subexponential) random matrices thanks to a modification of (D-RIP) where the inner norm is taken to be the ℓ_1-norm and the outer norm depends on the pregaussian distribution (this modified restricted isometry property appeared first in [7] for the traditional case $\mathbf{D} = \mathbf{I}$).

[1] meaning that a bound on $\|\mathbf{D}^*\mathbf{f} - \mathbf{D}^*\mathbf{f}^\sharp\|_p$ was established — for $p = 2$ and for tight frames, it reduces to the bound (1) on $\|\mathbf{f} - \mathbf{f}^\sharp\|_2$.

While the ℓ_1-analysis approach mentioned above concerns recovery of analysis-sparse vectors, other popular compressive sensing algorithms adapted to the dictionary case target the recovery of synthesis-sparse vectors. The theoretical guarantees, though, require the ability to best-approximate by synthesis-sparse vectors [1, 3] or at least to near-best-approximate [9]. These tasks may be achievable in specific situations, but they are hard in general. The purpose of this note is to show that such a strong requirement is not necessary to obtain weaker reconstruction bounds similar to (1). To illustrate this point, we state the following representative result.

Theorem 1. Suppose that $\mathbf{D} \in \mathbb{C}^{n \times N}$ satisfies $\mathbf{D}\mathbf{D}^* = \mathbf{I}_n$ and that

$$m \geq C\, s \ln\left(\frac{eN}{s}\right).$$

If $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a random matrix populated by iid subgaussian entries with variance $1/m$, then with overwhelming probability

$$\|\mathbf{f} - \mathbf{f}^{\mathrm{HTP}}\|_2 \leq c\, \frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}} + d\, \|\mathbf{e}\|_2,$$

holds for all $\mathbf{f} \in \mathbb{C}^n$ and all $\mathbf{e} \in \mathbb{C}^m$. Here, $\mathbf{f}^{\mathrm{HTP}}$ is the output of the hard thresholding pursuit algorithm of order $2s$ adapted to $\mathbf{D}$ and applied to $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e}$. It is obtained after a finite number of iterations.

In this statement, the constants C, c, and d are universal. Throughout the note, they will represent constants with values changing from line to line and possibly dependent on parameters δ and μ to be fixed in due course. Theorem 1 is proven in Section 3. It relies on arguments put forward in Section 2 to derive a similar statement for the simpler iterative hard thresholding algorithm. We conclude in Section 4 by highlighting the adjustment needed to handle loose frames.

2 Iterative Hard Thresholding

We start by presenting the core of the argument on the illustrative case of an analysis-sparse signal $\mathbf{f} \in \mathbb{C}^n$ — it is assumed that $\mathbf{D}^*\mathbf{f}$ is exactly s-sparse, see a future comment — which is measured with perfect precision, so that $\mathbf{y} = \mathbf{A}\mathbf{f}$. We consider the sequence $(\mathbf{f}^k)$ of synthesis-sparse *[sic]* signals defined by $\mathbf{f}^0 = \mathbf{0}$ and, for $k \geq 0$, by

$$\mathbf{f}^{k+1} = \mathbf{D}\left\{H_s(\mathbf{D}^*\mathbf{g}^k)\right\}, \qquad \text{where} \quad \mathbf{g}^k := \mathbf{f}^k + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k). \tag{2}$$

Here, $H_s(\mathbf{z})$ is the hard thresholding operator applied to a vector $\mathbf{z} \in \mathbb{C}^N$: it keeps the s largest absolute entries of $\mathbf{z}$ and annihilates the other ones. The argument exploits the restricted isometry property adapted to $\mathbf{D}$, or rather one of its consequences established below.

Lemma 2. For all signals $\mathbf{g}, \mathbf{h} \in \mathbb{C}^n$ written as $\mathbf{g} := \mathbf{D}\mathbf{u}$ and $\mathbf{h} = \mathbf{D}\mathbf{v}$ with $\mathbf{u}, \mathbf{v} \in \mathbb{C}^N$,

$$|\langle \mathbf{g}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{h}\rangle| \le \delta_t^{\mathbf{D}} \|\mathbf{g}\|_2 \|\mathbf{h}\|_2, \qquad \text{where } t := \operatorname{card}\{\operatorname{supp}(\mathbf{u}) \cup \operatorname{supp}(\mathbf{v})\}.$$

Proof. Let $\mathbf{g}' := \mathbf{g}/\|\mathbf{g}\|_2$ and $\mathbf{h}' = e^{i\theta}\mathbf{h}/\|\mathbf{h}\|_2$ for $\theta \in [-\pi, \pi]$ chosen as the argument of the complex number $\langle \mathbf{g}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{h}\rangle$. We have

$$\begin{aligned}
\frac{|\langle \mathbf{g}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{h}\rangle|}{\|\mathbf{g}\|_2\|\mathbf{h}\|_2} &= \operatorname{Re}\left(\frac{e^{-i\theta}\langle \mathbf{g}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{h}\rangle}{\|\mathbf{g}\|_2\|\mathbf{h}\|_2}\right) \\
&= \operatorname{Re}\left(\langle \mathbf{g}', (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{h}'\rangle\right) = \operatorname{Re}\left(\langle \mathbf{A}\mathbf{g}', \mathbf{A}\mathbf{h}'\rangle\right) - \operatorname{Re}\left(\langle \mathbf{g}', \mathbf{h}'\rangle\right) \\
&= \frac{1}{4}\left(\|\mathbf{A}(\mathbf{g}' + \mathbf{h}')\|_2^2 - \|\mathbf{A}(\mathbf{g}' - \mathbf{h}')\|_2^2\right) - \frac{1}{4}\left(\|\mathbf{g}' + \mathbf{h}'\|_2^2 - \|\mathbf{g}' - \mathbf{h}'\|_2^2\right) \\
&= \frac{1}{4}\left(\|\mathbf{A}(\mathbf{g}' + \mathbf{h}')\|_2^2 - \|\mathbf{g}' + \mathbf{h}'\|_2^2\right) - \frac{1}{4}\left(\|\mathbf{A}(\mathbf{g}' - \mathbf{h}')\|_2^2 - \|\mathbf{g}' - \mathbf{h}'\|_2^2\right) \\
&\le \frac{1}{4}\delta_t^{\mathbf{D}}\|\mathbf{g}' + \mathbf{h}'\|_2^2 + \frac{1}{4}\delta_t^{\mathbf{D}}\|\mathbf{g}' - \mathbf{h}'\|_2^2 = \frac{1}{4}\delta_t^{\mathbf{D}}\left(2\|\mathbf{g}'\|_2^2 + 2\|\mathbf{h}'\|_2^2\right) = \delta_t^{\mathbf{D}}.
\end{aligned}$$

The required result follows immediately after multiplying throughout by $\|\mathbf{g}\|_2\|\mathbf{h}\|_2$. □

With Lemma 2 established, we can prove that the sequence $(\mathbf{f}^k)$ converges to $\mathbf{f}$ following the argument proposed in [4] for the usual iterative hard thresholding algorithm. Thus, we start from the observation that $H_s(\mathbf{D}^*\mathbf{g}^k)$ is a better s-sparse approximation to $\mathbf{D}^*\mathbf{g}^k$ than $\mathbf{D}^*\mathbf{f}$ is to write

$$\|\mathbf{D}^*\mathbf{g}^k - H_s(\mathbf{D}^*\mathbf{g}^k)\|_2^2 \le \|\mathbf{D}^*\mathbf{g}^k - \mathbf{D}^*\mathbf{f}\|_2^2, \qquad \text{i.e.,} \quad \|\mathbf{D}^*(\mathbf{g}^k - \mathbf{f}) + \mathbf{D}^*\mathbf{f} - H_s(\mathbf{D}^*\mathbf{g}^k)\|_2^2 \le \|\mathbf{D}^*(\mathbf{g}^k - \mathbf{f})\|_2^2.$$

After expanding the square on the left-hand side and rearranging, we arrive at

$$\begin{aligned}
\|\mathbf{D}^*\mathbf{f} - H_s(\mathbf{D}^*\mathbf{g}^k)\|_2^2 &\le 2\operatorname{Re}\langle \mathbf{D}^*\mathbf{f} - H_s(\mathbf{D}^*\mathbf{g}^k), \mathbf{D}^*(\mathbf{f} - \mathbf{g}^k)\rangle = 2\operatorname{Re}\langle \mathbf{D}(\mathbf{D}^*\mathbf{f} - H_s(\mathbf{D}^*\mathbf{g}^k)), \mathbf{f} - \mathbf{g}^k\rangle \\
&= 2\operatorname{Re}\langle \mathbf{f} - \mathbf{f}^{k+1}, \mathbf{f} - \mathbf{g}^k\rangle = 2\operatorname{Re}\langle \mathbf{f} - \mathbf{f}^{k+1}, (\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{f} - \mathbf{f}^k)\rangle.
\end{aligned}$$

We now invoke Lemma 2, as well as the fact that $\|\mathbf{D}\|_{2\to 2} = 1$ (since $\mathbf{D}$ has all its nonzero singular values equal to one by virtue of $\mathbf{D}\mathbf{D}^* = \mathbf{I}$), to derive

$$\|\mathbf{f} - \mathbf{f}^{k+1}\|_2^2 = \|\mathbf{D}(\mathbf{D}^*\mathbf{f} - H_s(\mathbf{D}^*\mathbf{g}^k))\|_2^2 \le \|\mathbf{D}^*\mathbf{f} - H_s(\mathbf{D}^*\mathbf{g}^k)\|_2^2 \le 2\,\delta_{3s}^{\mathbf{D}}\,\|\mathbf{f} - \mathbf{f}^{k+1}\|_2\|\mathbf{f} - \mathbf{f}^k\|_2,$$

or after simplification

$$\|\mathbf{f} - \mathbf{f}^{k+1}\|_2 \le \rho\|\mathbf{f} - \mathbf{f}^k\|_2, \qquad \rho := 2\delta_{3s}^{\mathbf{D}}.$$

This ensures recovery of $\mathbf{f}$ as the limit of $\mathbf{f}^k$ when $k \to \infty$ as soon as $\rho < 1$, i.e., $\delta_{3s}^{\mathbf{D}} < 1/2$.

The unabridged argument is a variation on the simple case just presented. We do take note, though, that the simple case is quite unrealistic, because requiring $\mathbf{D}^*\mathbf{f}$ to be exactly s-sparse for $\mathbf{f} \neq \mathbf{0}$ is very restrictive. Indeed, if the columns $\mathbf{d}_1, \ldots, \mathbf{d}_N \in \mathbb{C}^n$ of $\mathbf{D}$ are in general position, then $(\mathbf{D}^*\mathbf{f})_j = 0$ for all $j \notin S := \operatorname{supp}(\mathbf{D}^*\mathbf{f})$, i.e., $\langle \mathbf{f}, \mathbf{d}_j\rangle = 0$ for all $j \in \overline{S}$, means that $\mathbf{f}$ lies in the orthogonal complement of the space $\mathrm{span}\{\mathbf{d}_j, j \in \overline{S}\}$ which has full dimension when $N - s \geq n$, so that exact sparsity is only possible for $s > N - n$. Let us point out in passing that this observation has sparked some recent investigations into the cosparse analysis model, see e.g. [8, 11]. Back to our setting, we therefore emphasize that the realistic situation occurs when $\sigma_s(\mathbf{D}^* f)_1$ decays quickly, rather than vanishes quickly, as s increases. Practical examples of this situation are given in [2]. As a consequence, we do not make any sparsity assumption on $\mathbf{D}^*\mathbf{f}$ from now on. Instead, we group the indices in sets S_0, S_1, S_2, ... of size s arranged by decreasing absolute entries of $\mathbf{D}^*\mathbf{f}$. We then write $\mathbf{f} = \mathbf{D}(\mathbf{D}^*\mathbf{f})$ as the sum of a $2s$-synthesis-sparse signal $\widehat{\mathbf{f}}$ and a remainder signal $\overline{\mathbf{f}}$ as follows:

$$\mathbf{f} = \widehat{\mathbf{f}} + \overline{\mathbf{f}}, \qquad \text{where} \quad \begin{cases} \widehat{\mathbf{f}} & := \mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{S_0 \cup S_1}\big), \\ \overline{\mathbf{f}} & := \mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{\overline{S_0 \cup S_1}}\big). \end{cases}$$

The remainder signal $\overline{\mathbf{f}}$ is then absorbed in the measurement error by writing

$$\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e} = \mathbf{A}\widehat{\mathbf{f}} + \overline{\mathbf{e}}, \qquad \text{where} \quad \overline{\mathbf{e}} := \mathbf{A}\overline{\mathbf{f}} + \mathbf{e}.$$

It is useful to isolate the following estimates.

Lemma 3. One has

$$\|\overline{\mathbf{f}}\|_2 \leq \Sigma \qquad \text{and} \qquad \|\overline{\mathbf{e}}\|_2 \leq \sqrt{1 + \delta_s^{\mathbf{D}}}\, \Sigma + \|\mathbf{e}\|_2,$$

where the quantity Σ satisfies

$$\Sigma := \sum_{k \geq 2} \|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_2 \leq \frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}}.$$

Proof. The bound on Σ is folklore: it is based on the inequalities

$$\|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_2/\sqrt{s} \leq \max_{j \in S_k} |(\mathbf{D}^*\mathbf{f})_j| \leq \min_{j \in S_{k-1}} |(\mathbf{D}^*\mathbf{f})_j| \leq \|[\mathbf{D}^*\mathbf{f}]_{S_{k-1}}\|_1/s$$

written as $\|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_2 \leq \|[\mathbf{D}^*\mathbf{f}]_{S_{k-1}}\|_1/\sqrt{s}$ and then summed for $k \geq 2$. For the bound on $\|\overline{\mathbf{f}}\|_2$, we simply invoke the triangle inequality $\|\overline{\mathbf{f}}\|_2 = \|\sum_{k\geq 2} \mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{S_k}\big)\|_2 \leq \sum_{k \geq 2} \|\mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{S_k}\big)\|_2$ together with the fact that $\|\mathbf{D}\|_{2\to 2} \leq 1$. For the bound on $\|\overline{\mathbf{e}}\|_2$, we start from $\|\overline{\mathbf{e}}\|_2 \leq \|\mathbf{A}\overline{\mathbf{f}}\|_2 + \|\mathbf{e}\|_2$ and bound $\|\mathbf{A}\overline{\mathbf{f}}\|_2$ with the help of (D-RIP) as

$$\begin{aligned} \|\mathbf{A}\overline{\mathbf{f}}\|_2 &= \|\sum_{k\geq 2} \mathbf{A}\mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{S_k}\big)\|_2 \leq \sum_{k \geq 2} \|\mathbf{A}\mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{S_k}\big)\|_2 \leq \sum_{k\geq 2} \sqrt{1+\delta_s^{\mathbf{D}}} \|\mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{S_k}\big)\|_2 \\ &\leq \sqrt{1+\delta_s^{\mathbf{D}}} \sum_{k\geq 2} \|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_2 = \sqrt{1+\delta_s^{\mathbf{D}}}\, \Sigma. \end{aligned}$$

The announced bound on $\|\overline{\mathbf{e}}\|_2$ is now apparent. $\square$

In order to tackle the realistic situation discussed above, we replace the parameter s by $2s$ in (2). For extra generality, we attach a free parameter μ in front of $\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k)$ in the definition of $\mathbf{g}^k$. All in all, the iterative hard thresholding algorithm of order $2s$ with parameter μ adapted to $\mathbf{D}$ — $\mathbf{D}$-IHT$^\mu_{2s}$ for short — applied to $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e} = \mathbf{A}\widehat{\mathbf{f}} + \overline{\mathbf{e}}$ consists in constructing a sequence $(\mathbf{f}^k)$ of synthesis-sparse signals defined by $\mathbf{f}^0 = \mathbf{0}$ and, for $k \geq 0$, by

$$(\mathbf{D}\text{-IHT}^\mu_{2s}) \qquad \mathbf{f}^{k+1} = \mathbf{D}\left\{H_{2s}(\mathbf{D}^*\mathbf{g}^k)\right\}, \qquad \mathbf{g}^k := \mathbf{f}^k + \mu\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k).$$

With the notation introduced above, the main result of this section reads as follows.

Theorem 4. Given a tight frame $\mathbf{D} \in \mathbb{C}^{n\times N}$ and given $\mu \in (1/2, 3/2)$, suppose that the matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ has a restricted isometry constant adapted to $\mathbf{D}$ of order $6s$ satisfying

$$\delta^{\mathbf{D}}_{6s} < \frac{1/2 - |1-\mu|}{\mu}. \tag{3}$$

Then, for all $\mathbf{f} \in \mathbb{C}^n$ and all $\mathbf{e} \in \mathbb{C}^m$, the sequence $(\mathbf{f}^k)$ produced by $(\mathbf{D}\text{-IHT}^\mu_{2s})$ applied to $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e}$ satisfies, for any $k \geq 0$,

$$\|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 \leq \rho^k \|\mathbf{f}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2, \tag{4}$$

where $\rho := 2(|1-\mu| + \mu\delta^{\mathbf{D}}_{6s}) < 1$. The constants c and d also depend only on μ and $\delta^{\mathbf{D}}_{6s}$.

Proof. Let us use the shorthand notation δ for $\delta^{\mathbf{D}}_{6s}$ throughout the proof. In order to obtain the bound (4), we could work as in the illustrative case at the signal level and show that, for any $k \geq 0$,

$$\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \leq \rho\|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2. \tag{5}$$

We shall instead work at the coefficient level: with $\mathbf{x}^0 = \mathbf{0}$ and $\mathbf{x}^{k+1} := H_{2s}(\mathbf{D}^*\mathbf{g}^k)$ designating the $2s$-sparse vector such that $\mathbf{f}^{k+1} = \mathbf{D}\mathbf{x}^{x+1}$, we are going to establish that, for any $k \geq 0$,

$$\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2 \leq \rho\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2. \tag{6}$$

This immediately implies

$$\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2 \leq \rho^k\|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2, \tag{7}$$

and the estimate announced in (4) follows from $\|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 = \|\mathbf{D}(\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k)\|_2 \leq \|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2$ and from $\|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 = \|\widehat{\mathbf{f}}\|_2 \leq \|\mathbf{f}\|_2 + \|\overline{\mathbf{f}}\|_2 \leq \|\mathbf{f}\|_2 + \Sigma$. Now, in order to establish (6), we start with the observation that $\mathbf{x}^{k+1}$ is a better $2s$-term approximation to $\mathbf{D}^*\mathbf{g}^k$ than $[\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1}$ is to derive

$$\|\mathbf{D}^*\mathbf{g}^k - \mathbf{x}^{k+1}\|_2^2 \leq \|\mathbf{D}^*\mathbf{g}^k - [\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1}\|_2^2.$$

Exploiting the fact that $[\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1} = \mathbf{D}^*\mathbf{f} - [\mathbf{D}^*\mathbf{f}]_{\overline{S_0\cup S_1}} = \mathbf{D}^*\widehat{\mathbf{f}} + \mathbf{D}^*\overline{\mathbf{f}} - [\mathbf{D}^*\mathbf{f}]_{\overline{S_0\cup S_1}}$, this reads

$$\|\mathbf{D}^*(\mathbf{g}^k - \widehat{\mathbf{f}}) + \mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2^2 \leq \|\mathbf{D}^*(\mathbf{g}^k - \widehat{\mathbf{f}}) - \mathbf{D}^*\overline{\mathbf{f}} + [\mathbf{D}^*\mathbf{f}]_{\overline{S_0\cup S_1}}\|_2^2.$$

Expanding the squares and rearranging, we arrive at

$$\begin{aligned}\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2^2 \le\; & 2\,\mathrm{Re}\langle \mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}, \mathbf{D}^*(\widehat{\mathbf{f}} - \mathbf{g}^k)\rangle \\ & + 2\,\mathrm{Re}\langle \mathbf{D}^*(\widehat{\mathbf{f}} - \mathbf{g}^k), \mathbf{D}^*\overline{\mathbf{f}} - [\mathbf{D}^*\mathbf{f}]_{\overline{S_0 \cup S_1}}\rangle \\ & + \|\mathbf{D}^*\overline{\mathbf{f}} - [\mathbf{D}^*\mathbf{f}]_{\overline{S_0 \cup S_1}}\|_2^2. \end{aligned} \tag{8}$$

The square root of the last term in the right-hand side of (8) is bounded as

$$\|\mathbf{D}^*\overline{\mathbf{f}} - [\mathbf{D}^*\mathbf{f}]_{\overline{S_0 \cup S_1}}\|_2 \le \|\mathbf{D}^*\overline{\mathbf{f}}\|_2 + \sum_{k \ge 2} \|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_2 = \|\overline{\mathbf{f}}\|_2 + \Sigma \le 2\Sigma. \tag{9}$$

The middle term in the right-hand side of (8) is simply zero, as it reduces to

$$2\,\mathrm{Re}\langle \widehat{\mathbf{f}} - \mathbf{g}^k, \mathbf{D}\mathbf{D}^*\overline{\mathbf{f}} - \mathbf{D}\big([\mathbf{D}^*\mathbf{f}]_{\overline{S_0 \cup S_1}}\big)\rangle = 2\,\mathrm{Re}\langle \widehat{\mathbf{f}} - \mathbf{g}^k, \overline{\mathbf{f}} - \overline{\mathbf{f}}\rangle = 0. \tag{10}$$

As for the first term in the right-hand side of (8), it takes the form

$$\begin{aligned}(11)\quad & 2\,\mathrm{Re}\langle \mathbf{D}\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{D}\mathbf{x}^{k+1}, \widehat{\mathbf{f}} - \mathbf{g}^k\rangle = 2\,\mathrm{Re}\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, \widehat{\mathbf{f}} - \mathbf{f}^k - \mu\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k)\rangle \\ & = 2(1-\mu)\,\mathrm{Re}\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, \widehat{\mathbf{f}} - \mathbf{f}^k\rangle + 2\mu\,\mathrm{Re}\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, (\mathbf{I} - \mathbf{A}^*\mathbf{A})(\widehat{\mathbf{f}} - \mathbf{f}^k)\rangle - 2\mu\,\mathrm{Re}\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, \mathbf{A}^*\overline{\mathbf{e}}\rangle. \end{aligned}$$

In view of Cauchy–Schwarz inequality, of Lemma 2, and of (D-RIP), we notice that

$$\begin{aligned} |\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, \widehat{\mathbf{f}} - \mathbf{f}^k\rangle| & \le \|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2, \\ |\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, (\mathbf{I} - \mathbf{A}^*\mathbf{A})(\widehat{\mathbf{f}} - \mathbf{f}^k)\rangle| & \le \delta\,\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2, \\ |\langle \widehat{\mathbf{f}} - \mathbf{f}^{k+1}, \mathbf{A}^*\overline{\mathbf{e}}\rangle| & = |\langle \mathbf{A}(\widehat{\mathbf{f}} - \mathbf{f}^{k+1}), \overline{\mathbf{e}}\rangle| \le \|\mathbf{A}(\widehat{\mathbf{f}} - \mathbf{f}^{k+1})\|_2 \|\overline{\mathbf{e}}\|_2 \\ & \le \sqrt{1+\delta}\,\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \|\overline{\mathbf{e}}\|_2. \end{aligned}$$

Using these inequalities in (11) allows us to bound the first term in the right-hand side of (8) by

$$2\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \left\{ |1-\mu| \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + \mu\,\delta\,\|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + \mu\,\sqrt{1+\delta}\,\|\overline{\mathbf{e}}\|_2 \right\}. \tag{12}$$

Substituting (9), (10), and (12) into (8) yields

$$\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2^2 \le 2\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \left\{ (|1-\mu| + \mu\,\delta) \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + \mu\,\sqrt{1+\delta}\,\|\overline{\mathbf{e}}\|_2 \right\} + 4\Sigma^2. \tag{13}$$

Taking $\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 = \|\mathbf{D}(\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1})\|_2 \le \|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2$ into account, we could deduce from (13) that

$$\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2^2 \le 2\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \left\{ (|1-\mu| + \mu\,\delta) \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + \mu\,\sqrt{1+\delta}\,\|\overline{\mathbf{e}}\|_2 \right\} + 4\Sigma^2$$

and proceed as we do below to obtain the bound (5). We shall instead deduce from (13), still using $\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \le \|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2$ but also $\|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 \le \|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2$, that

$$\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2^2 \le 2\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2 \left\{ (|1-\mu| + \mu\,\delta) \|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2 + \mu\,\sqrt{1+\delta}\,\|\overline{\mathbf{e}}\|_2 \right\} + 4\Sigma^2.$$

With {} denoting the above quantity in curly brackets, this inequality reads

$$\Big(\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2 - \{\}\Big)^2 \le \{\}^2 + 4\Sigma^2 \le (\{\} + 2\Sigma)^2 .$$

It then follows that

$$\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^{k+1}\|_2 \le \{\} + (\{\} + 2\Sigma) = 2\,(\{\} + \Sigma)\,,$$

which does take the form of (7) after invoking Lemma 3 to bound $\|\overline{\mathbf{e}}\|_2$. It is easily verified that $\rho := 2(|1-\mu| + \mu\,\delta)$ satisfies $\rho < 1$ when (3) holds. This concludes the proof. □

The argument presented above does not guarantee convergence of the sequence $(\mathbf{f}^k)$. However, since (4) implies boundedness of this sequence, at least it admits a convergent subsequence. We can give error bounds for the recovery of $\mathbf{f}$ as the limit of such a convergent subsequence. In fact, we can also give error bounds for the recovery of $\mathbf{f}$ as the iterate $\mathbf{f}^k$ when the index k is large enough. Here are the precise statements.

Corollary 5. Under the same hypotheses as in Theorem 4, if $\mathbf{f}^\flat$ denotes a cluster point of the sequence $(\mathbf{f}^k)$, then

$$\|\mathbf{f} - \mathbf{f}^\flat\|_2 \le c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}} + d\,\|\mathbf{e}\|_2. \tag{14}$$

Moreover, if a bound $\|\mathbf{e}\|_2 \le \eta$ is available, then

$$\|\mathbf{f} - \mathbf{f}^k\|_2 \le c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}} + d\,\eta \qquad \text{as soon as} \quad k \ge \frac{\ln(\|\mathbf{y}\|_2/\eta)}{\ln(1/\rho)}. \tag{15}$$

The constants c and d — different on both occurrences — depend only on μ and $\delta_{6s}^{\mathbf{D}}$.

Proof. From the bound (4), we derive that, for any $k \ge 0$,

$$\|\mathbf{f}-\mathbf{f}^k\|_2 \le \|\widehat{\mathbf{f}}-\mathbf{f}^k\|_2 + \|\overline{\mathbf{f}}\|_2 \le \rho^k\|\mathbf{f}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2 + \|\overline{\mathbf{f}}\|_2 \le \rho^k\|\widehat{\mathbf{f}}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2 + (1+\rho^k)\|\overline{\mathbf{f}}\|_2,$$

and invoking Lemma 3 to bound $\|\overline{\mathbf{f}}\|_2$ gives

$$\|\mathbf{f} - \mathbf{f}^k\|_2 \le \rho^k\|\widehat{\mathbf{f}}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2. \tag{16}$$

For a convergent subsequence $(\mathbf{f}^{k_j})$, taking the limit as $j \to \infty$ yields $\|\mathbf{f}-\mathbf{f}^\flat\|_2 \le c\,\Sigma + d\,\|\mathbf{e}\|_2$, which implies (14) in view of the bound on Σ from Lemma 3. Next, when $k \ge \ln(\|\mathbf{y}\|_2/\eta)/\ln(1/\rho)$, we have $\rho^k \le \eta/\|\mathbf{y}\|_2$. Moreover, since $\sqrt{1-\delta_{2s}^{\mathbf{D}}}\,\|\widehat{\mathbf{f}}\|_2 \le \|\mathbf{A}\widehat{\mathbf{f}}\|_2 = \|\mathbf{y}-\overline{\mathbf{e}}\|_2 \le \|\mathbf{y}\|_2 + \|\overline{\mathbf{e}}\|_2$, we obtain $\rho^k\|\widehat{\mathbf{f}}\|_2 \le [(\eta/\|\mathbf{y}\|_2)\|\mathbf{y}\|_2 + \rho^k\|\overline{\mathbf{e}}\|_2]/\sqrt{1-\delta_{2s}^{\mathbf{D}}} \le [\eta + \|\overline{\mathbf{e}}\|_2]/\sqrt{1/2}$. Substituting this inequality into (16) and using the bounds on $\|\overline{\mathbf{e}}\|_2$ and Σ from Lemma 3 implies (15). □

Remark. It is also possible to derive reconstruction error bounds in ℓ_p for any $p \in [1,2]$. To see this, we start from the estimate (7) in the proof of Theorem 4, precisely from

$$\|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2 \le \rho^k \|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2.$$

Noticing that $\mathbf{D}^*\widehat{\mathbf{f}} = \mathbf{D}^*\mathbf{f} - \mathbf{D}^*\overline{\mathbf{f}} = [\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1} + [\mathbf{D}^*\mathbf{f}]_{\overline{S_0\cup S_1}} - \mathbf{D}^*\overline{\mathbf{f}}$, this gives

$$\begin{aligned}\|[\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1} - \mathbf{x}^k\|_2 &\le \|\mathbf{D}^*\widehat{\mathbf{f}} - \mathbf{x}^k\|_2 + \|[\mathbf{D}^*\mathbf{f}]_{\overline{S_0\cup S_1}}\|_2 + \|\mathbf{D}^*\overline{\mathbf{f}}\|_2 \\ &\le \rho^k\|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2 + \sum_{k\ge 2}\|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_2 + \|\overline{\mathbf{f}}\|_2 \\ &\le \rho^k\|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2 \le \rho^k\|\mathbf{D}\widehat{\mathbf{f}}\|_2 + c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}} + d\,\|\mathbf{e}\|_2,\end{aligned}$$

where Lemma 3 was used in the last two steps. Since $[\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1} - \mathbf{x}^k$ is $4s$-sparse, it follows that

$$\begin{aligned}\text{(17)}\qquad \|[\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1} - \mathbf{x}^k\|_p &\le (4s)^{1/p-1/2}\|[\mathbf{D}^*\mathbf{f}]_{S_0\cup S_1} - \mathbf{x}^k\|_2 \\ &\le \rho^k(4s)^{1/p-1/2}\|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 + c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{s^{1-1/p}} + d\,s^{1/p-1/2}\|\mathbf{e}\|_2.\end{aligned}$$

Moreover, it is easy to extend the argument of Lemma 3 to obtain

$$\text{(18)}\qquad \|[\mathbf{D}^*\mathbf{f}]_{\overline{S_0\cup S_1}}\|_p \le \sum_{k\ge 2}\|[\mathbf{D}^*\mathbf{f}]_{S_k}\|_p \le \frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{s^{1-1/p}}.$$

Summing (17) and (18) yields

$$\|\mathbf{D}^*\mathbf{f} - \mathbf{x}^k\|_p \le \rho^k(4s)^{1/p-1/2}\|\mathbf{D}^*\widehat{\mathbf{f}}\|_2 + c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{s^{1-1/p}} + d\,s^{1/p-1/2}\|\mathbf{e}\|_2.$$

We deduce a bound at the coefficient level for the recovery of $\mathbf{D}^*\mathbf{f}$ as a cluster point of the sequence $(\mathbf{x}^k)$, namely

$$\|\mathbf{D}^*\mathbf{f} - \mathbf{x}^\flat\|_p \le c\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{s^{1-1/p}} + d\,s^{1/p-1/2}\|\mathbf{e}\|_2, \qquad p\in[1,2].$$

At the signal level, in view of $\|\mathbf{f} - \mathbf{f}^k\|_p = \|\mathbf{D}\mathbf{D}^*\mathbf{f} - \mathbf{D}\mathbf{x}^k\|_p \le \|\mathbf{D}\|_{p\to p}\|\mathbf{D}^*\mathbf{f} - \mathbf{x}^k\|_p$, a bound

$$\|\mathbf{f} - \mathbf{f}^\flat\|_p \le c'\,\frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{s^{1-1/p}} + d'\,s^{1/p-1/2}\|\mathbf{e}\|_2, \qquad p\in[1,2],$$

also holds for the recovery of $\mathbf{f}$ as a cluster point of the sequence $(\mathbf{f}^k)$. The constants c' and d' involving $\|\mathbf{D}\|_{p\to p}$ can be made independent of p thanks to Riesz–Thorin theorem, but they a priori depend on $\mathbf{D}$, hence possibly on n and N.

3 Hard Thresholding Pursuit

The results established in the previous section for the iterative hard thresholding algorithm are somewhat unsatisfying for a couple of reasons. First, even though the error bound (15) compares to (1), an estimation s for the sparsity level is required as an input of the algorithm, whereas the ℓ_1-analysis minimization only requires an estimation η for the magnitude of the measurement error as an input (besides $\mathbf{y}$, $\mathbf{A}$, and $\mathbf{D}$, obviously). Second, the stopping criterion for the number of iterations suggested by (15) — namely, $k = \lceil \ln(\|\mathbf{y}\|_2/\eta)/\ln(1/\rho) \rceil$ — does not apply when no measurement error is present ($\eta = 0$). To obtain exact recovery in this case, one really needs an infinite number of iterations, which is of course impractical. This second drawback is resolved by considering the hard thresholding pursuit algorithm of order $2s$ with parameter μ adapted to $\mathbf{D}$ — $\mathbf{D}$-HTP$^{\mu}_{2s}$ for short. This algorithm consists in constructing a sequence $(\mathbf{f}^k)$ of synthesis-sparse signals defined by $\mathbf{f}^0 = \mathbf{0}$ and, for $k \geq 0$, by

$$(\mathbf{D}\text{-HTP}^{\mu}_{2s}) \qquad \begin{cases} S^{k+1} = \text{index set of } 2s \text{ largest absolute entries of } \mathbf{D}^* \left(\mathbf{f}^k + \mu \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k)\right), \\ \mathbf{f}^{k+1} = \mathbf{D} \operatorname{argmin}\{\|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{z}\|_2, \ \operatorname{supp}(\mathbf{z}) \subseteq S^{k+1}\}. \end{cases}$$

A natural stopping criterion here is $S^{k+1} = S^k$, since the sequence becomes stationary afterwards. We also point out that the sequence (S^k) is eventually periodic, hence so is $(\mathbf{f}^k)$, because there are only finitely many index sets of size $2s$. Therefore, if the sequences (S^k) and $(\mathbf{f}^k)$ converge, then they are eventually stationary, and $S^{k+1} = S^k$ does occur for some $k \geq 0$. Note that, in the classical setting and under some restricted isometry condition, convergence is guaranteed unconditionally on the input $\mathbf{y}$ provided $\mu < 1$ is suitably chosen, see [5], but the validity of this phenomenon in the dictionary case is put in doubt by some modest numerical experiments. Regardless of convergence issues, theoretical bounds for the recovery error exist and resemble the ones presented in the previous section for iterative hard thresholding. They are stated below.

Theorem 6. Given a tight frame $\mathbf{D} \in \mathbb{C}^{n \times N}$ and given $\mu \in (1/2, 3/2)$, suppose that the matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ has a restricted isometry constant adapted to $\mathbf{D}$ of order $6s$ small enough to ensure that

$$\delta^{\mathbf{D}}_{6s} < \frac{\sqrt{1 - \delta^{\mathbf{D}}_{6s}} / \left(2\sqrt{1 + \delta^{\mathbf{D}}_{6s}}\right) - |1 - \mu|}{\mu}. \tag{19}$$

Then, for all $\mathbf{f} \in \mathbb{C}^n$ and all $\mathbf{e} \in \mathbb{C}^m$, the sequence $(\mathbf{f}^k)$ produced by $(\mathbf{D}\text{-HTP}^{\mu}_{2s})$ with $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e}$ satisfies, for any $k \geq 0$,

$$\|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 \leq \varrho^k \|\mathbf{f}\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2, \tag{20}$$

where $\varrho := 2(|1 - \mu| + \mu\,\delta^{\mathbf{D}}_{6s})\sqrt{1 + \delta^{\mathbf{D}}_{6s}}/\sqrt{1 - \delta^{\mathbf{D}}_{6s}} < 1$. The constants c and d also depend only on μ and $\delta^{\mathbf{D}}_{6s}$.

Proof. As in the proof of Theorem 4, we use the shorthand notation δ for $\delta^{\mathbf{D}}_{6s}$. For any $k \geq 0$, we target the inequality

$$\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \leq \varrho \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2 \tag{21}$$

for some $\varrho < 1$. In fact, we rely on the estimate (5) from the proof of Theorem 4, namely

$$\|\widehat{\mathbf{f}} - \check{\mathbf{f}}^{k+1}\|_2 \leq \rho \|\widehat{\mathbf{f}} - \mathbf{f}^k\|_2 + c\,\Sigma + d\,\|\mathbf{e}\|_2, \qquad \rho = 2(|1 - \mu| + \mu\delta), \tag{22}$$

where $\check{\mathbf{f}}^{k+1} := \mathbf{D}\{H_{2s}(\mathbf{D}^*(\mathbf{f}^k + \mu \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k)))\}$. By definition of $\mathbf{f}^{k+1}$, we have

$$\|\mathbf{y} - \mathbf{A}\mathbf{f}^{k+1}\|_2 \le \|\mathbf{y} - \mathbf{A}\check{\mathbf{f}}^{k+1}\|_2, \qquad \text{i.e.,} \quad \|\mathbf{A}(\widehat{\mathbf{f}} - \mathbf{f}^{k+1}) + \overline{\mathbf{e}}\|_2 \le \|\mathbf{A}(\widehat{\mathbf{f}} - \check{\mathbf{f}}^{k+1}) + \overline{\mathbf{e}}\|_2.$$

Since the triangle inequality and the restricted isometry property adapted to $\mathbf{D}$ yield

$$\|\mathbf{A}(\widehat{\mathbf{f}} - \mathbf{f}^{k+1}) + \overline{\mathbf{e}}\|_2 \ge \sqrt{1-\delta}\, \|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 - \|\overline{\mathbf{e}}\|_2,$$
$$\|\mathbf{A}(\widehat{\mathbf{f}} - \check{\mathbf{f}}^{k+1}) + \overline{\mathbf{e}}\|_2 \le \sqrt{1+\delta}\, \|\widehat{\mathbf{f}} - \check{\mathbf{f}}^{k+1}\|_2 + \|\overline{\mathbf{e}}\|_2,$$

we deduce the estimate

$$\|\widehat{\mathbf{f}} - \mathbf{f}^{k+1}\|_2 \le \sqrt{\frac{1+\delta}{1-\delta}}\, \|\widehat{\mathbf{f}} - \check{\mathbf{f}}^{k+1}\|_2 + \frac{2}{\sqrt{1-\delta}} \|\overline{\mathbf{e}}\|_2.$$

After substituting (22) into the latter inequality and invoking Lemma 3 to bound $\|\overline{\mathbf{e}}\|_2$, we see that the desired inequality (21) holds with ϱ given above. The fact that $\varrho < 1$ is ensured by (19). □

We can now deduce error bounds for the recovery of $\mathbf{f}$ as one of the signals produced by the hard thresholding pursuit algorithm after a finite number of iterations. Note that an estimation for the magnitude of the measurement error is not necessary here.

Corollary 7. With the same assumptions and notation as in Theorem 6, if $\mathbf{f}^\flat$ denotes the output of ($\mathbf{D}$-HTP$^\mu_{2s}$) after the stopping criterion is met, then

$$\|\mathbf{f} - \mathbf{f}^\flat\|_2 \le c\, \frac{\sigma_s(\mathbf{D}^*\mathbf{f})_1}{\sqrt{s}} + d\, \|\mathbf{e}\|_2.$$

Even if the stopping criterion is not met, the previous bound is valid for any element $\mathbf{f}^\flat$ of the eventual cycle of the sequence $(\mathbf{f}^k)$. The constants c and d depend only on μ and $\delta^{\mathbf{D}}_{6s}$.

Proof. One essentially reproduce the proof of Corollary 5, coupled with the fact that the sequence $(\mathbf{f}^k)$ is at least eventually periodic if not eventually stationary. □

4 Loose frame dictionaries

In this section, we consider the case of an arbitrary dictionary, i.e., we remove the requirement that $\mathbf{D} \in \mathbb{C}^{n \times N}$ is a tight frame. It so happens that the previous recovery guarantees are still valid, provided the measurement process is slightly modified. Precisely, the measurement matrix now takes the form

$$\mathbf{A} = \mathbf{G}(\mathbf{D}\mathbf{D}^*)^{-1/2},$$

where $\mathbf{G} \in \mathbb{C}^{m \times n}$ is populated by iid subgaussian entries with variance $1/m$. The iterative hard thresholding and hard thresholding pursuit algorithms are simply adjusted by replacing $\mathbf{D}^*$ with the Moore–Penrose pseudo-inverse

$$\mathbf{D}^\dagger = \mathbf{D}^*(\mathbf{D}\mathbf{D}^*)^{-1}.$$

More explicitly, the adjusted algorithms are respectively

$$(\mathbf{D}\text{-IHT}_{2s}^{\mu}) \qquad \mathbf{f}^{k+1} = \mathbf{D}\left\{H_{2s}\left(\mathbf{D}^\dagger\left(\mathbf{f}^k + \mu\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k)\right)\right)\right\},$$

$$(\mathbf{D}\text{-HTP}_{2s}^{\mu}) \qquad \begin{cases} S^{k+1} = \text{index set of } 2s \text{ largest absolute entries of } \mathbf{D}^\dagger\left(\mathbf{f}^k + \mu\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{f}^k)\right), \\ \mathbf{f}^{k+1} = \mathbf{D}\ \mathrm{argmin}\{\|\mathbf{y} - \mathbf{A}\mathbf{D}\mathbf{z}\|_2,\ \mathrm{supp}(\mathbf{z}) \subseteq S^{k+1}\}. \end{cases}$$

We summarize the corresponding theoretical results in the following theorem.

Theorem 8. Given a full-rank matrix $\mathbf{D} \in \mathbb{C}^{n \times N}$ and given $\mu \in (1/2, 3/2)$, suppose that

$$m \geq C\, s\, \ln\left(\frac{eN}{s}\right).$$

If $\mathbf{A} = \mathbf{G}(\mathbf{D}\mathbf{D}^*)^{-1/2}$ where $\mathbf{G} \in \mathbb{C}^{m \times n}$ is a random matrix populated by iid subgaussian entries with variance $1/m$, then with overwhelming probability

$$\|\mathbf{f} - \mathbf{f}^\flat\|_2 \leq c\, \frac{\sigma_s(\mathbf{D}^\dagger\mathbf{f})_1}{\sqrt{s}} + d\, \|\mathbf{e}\|_2$$

holds for all $\mathbf{f} \in \mathbb{C}^n$ and all $\mathbf{e} \in \mathbb{C}^m$. Here, $\mathbf{f}^\flat$ is any cluster point of the sequence produced by the iterative hard thresholding or hard thresholding pursuit algorithm of order $2s$ with parameter μ adapted to $\mathbf{D}$ and applied to $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e}$. The constant C is universal, while the constants c and d depend only on μ and $\mathbf{D}$.

Proof. The key to the argument is the observation that

$$\widetilde{\mathbf{D}} := (\mathbf{D}\mathbf{D}^*)^{-1/2}\mathbf{D}$$

is a tight frame. In order to capitalize on the results established so far, instead of working with the true signals that are generically denoted $\mathbf{g}$, we work with transposed signals that are generically denoted $\widetilde{\mathbf{g}}$. The transposition rule is

$$\widetilde{\mathbf{g}} = (\mathbf{D}\mathbf{D}^*)^{-1/2}\mathbf{g}.$$

Thus, measuring the signal $\mathbf{f}$ as $\mathbf{y} = \mathbf{A}\mathbf{f} + \mathbf{e}$ through the matrix $\mathbf{A}$ is the same as measuring the signal $\widetilde{\mathbf{f}}$ as $\mathbf{y} = \mathbf{G}\widetilde{\mathbf{f}} + \mathbf{e}$ through the matrix $\mathbf{G}$. The assumptions of the theorem ensure that this matrix $\mathbf{G}$ has the restricted isometry property adapted to $\widetilde{\mathbf{D}}$. Its restricted isometry constant $\delta_{6s}^{\widetilde{\mathbf{D}}}$ can be chosen small enough to make the results of Sections 2 and 3 applicable to the sequences produced by the thresholding-based algorithms with inputs $\mathbf{y}$, $\mathbf{G}$, and $\widetilde{\mathbf{D}}$. These sequences are directly related to the sequences produced by the thresholding-based algorithms with inputs $\mathbf{y}$, $\mathbf{A}$, and $\mathbf{D}$, since the transposed versions of ($\mathbf{D}$-IHT$_{2s}^{\mu}$) and ($\mathbf{D}$-HTP$_{2s}^{\mu}$) readily amount to

$$\widetilde{\mathbf{f}}^{k+1} = \widetilde{\mathbf{D}} \left\{ H_{2s} \left(\widetilde{\mathbf{D}}^* \left(\widetilde{\mathbf{f}}^k + \mu \mathbf{G}^* (\mathbf{y} - \mathbf{G} \widetilde{\mathbf{f}}^k) \right) \right) \right\},$$

$$\begin{cases} S^{k+1} = \text{index set of } 2s \text{ largest absolute entries of } \widetilde{\mathbf{D}}^* \left(\widetilde{\mathbf{f}}^k + \mu \mathbf{G}^* (\mathbf{y} - \mathbf{G} \widetilde{\mathbf{f}}^k) \right), \\ \widetilde{\mathbf{f}}^{k+1} = \widetilde{\mathbf{D}} \operatorname{argmin} \{ \|\mathbf{y} - \mathbf{G} \widetilde{\mathbf{D}} \mathbf{z}\|_2, \ \operatorname{supp}(\mathbf{z}) \subseteq S^{k+1} \}. \end{cases}$$

Therefore, as a consequence of the results from the previous sections, we obtain

$$\|\widetilde{\mathbf{f}} - \widetilde{\mathbf{f}}^\flat\|_2 \le c \, \frac{\sigma_s(\widetilde{\mathbf{D}}^* \widetilde{\mathbf{f}})_1}{\sqrt{s}} + d \, \|\mathbf{e}\|_2.$$

It remains to take note of $\widetilde{\mathbf{D}}^* \widetilde{\mathbf{f}} = \mathbf{D}^\dagger \mathbf{f}$ and of $\|\mathbf{f} - \mathbf{f}^\flat\|_2 = \|(\mathbf{D}\mathbf{D}^*)^{1/2} (\widetilde{\mathbf{f}} - \widetilde{\mathbf{f}}^\flat)\|_2 \le \|\mathbf{D}\|_{2 \to 2} \|\widetilde{\mathbf{f}} - \widetilde{\mathbf{f}}^\flat\|_2$. □

References

[1] T. Blumensath. *Sampling and reconstructing signals from a union of linear subspaces*. IEEE Transactions on Information Theory, 57/7, 4660–4671, 2011.

[2] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. *Compressed sensing with coherent and redundant dictionaries*. Applied and Computational Harmonic Analysis, 31/1, 59–73, 2011.

[3] M. Davenport, D. Needell, and M. Wakin. *Signal space CoSaMP for sparse recovery with redundant dictionaries*. IEEE Transactions on Information Theory, 59/10, 6820–6829, 2013.

[4] S. Foucart. *Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants*. In: Approximation Theory XIII: San Antonio 2010 (Springer), 65–77, 2012.

[5] S. Foucart. *Hard thresholding pursuit: an algorithm for compressive sensing*. SIAM Journal on Numerical Analysis, 49/6, 2543–2563, 2011.

[6] S. Foucart. *Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries*. Linear Algebra and its Applications, 441, 4–21, 2014.

[7] S. Foucart and M.-J. Lai. *Sparse recovery with pre-Gaussian random matrices*. Studia Mathematica, 200/1, 91–102, 2010.

[8] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies. *Greedy-like algorithms for the cosparse analysis model*. Linear Algebra and its Applications, 441, 22–60, 2014.

[9] R. Giryes and D. Needell. *Greedy signal space methods for incoherence and beyond*. Applied and Computational Harmonic Analysis, To appear.

[10] F. Krahmer and R. Ward. *New and improved Johnson–Lindenstrauss embeddings via the restricted isometry property*. SIAM Journal on Mathematical Analysis, 43/3, 1269–1281, 2011.

[11] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. *The cosparse analysis model and algorithms*. Applied and Computational Harmonic Analysis, 34/1, 30–56, 2013.