

Hard Thresholding Pursuit Algorithms: Number of Iterations

Jean-Luc Bouchot, Simon Foucart, Pawel Hitzenko — Drexel University *

Abstract

The Hard Thresholding Pursuit algorithm for sparse recovery is revisited using a new theoretical analysis. The main result states that all sparse vectors can be exactly recovered from compressive linear measurements in a number of iterations at most proportional to the sparsity level as soon as the measurement matrix obeys a certain restricted isometry condition. The recovery is also robust to measurement error. The same conclusions are derived for a variation of Hard Thresholding Pursuit, called Graded Hard Thresholding Pursuit, which is a natural companion to Orthogonal Matching Pursuit and runs without a prior estimation of the sparsity level. In addition, for two extreme cases of the vector shape, it is shown that, with high probability on the draw of random measurements, a fixed sparse vector is robustly recovered in a number of iterations precisely equal to the sparsity level. These theoretical findings are experimentally validated, too.

Key words and phrases: compressive sensing, uniform sparse recovery, nonuniform sparse recovery, random measurements, iterative algorithms, hard thresholding.

AMS classification: 65F10, 65J20, 15A29, 94A12.

1 Introduction

This paper deals with the standard compressive sensing problem, i.e., the reconstruction of vectors $\mathbf{x} \in \mathbb{C}^N$ from an incomplete set of $m \ll N$ linear measurements organized in the form $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^m$ for some matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$. It is now well known that if \mathbf{x} is s -sparse (i.e., has only s nonzero entries) and if \mathbf{A} is a random matrix whose number m of rows scales like s times some logarithmic factors, then the reconstruction of \mathbf{x} is achievable via a variety of methods. The ℓ_1 -minimization is probably the most popular one, but simple iterative algorithms do provide alternative methods. We consider here the hard thresholding pursuit (HTP) algorithm [6] as well as a novel variation and we focus on the number of iterations needed for the reconstruction. This reconstruction is addressed in two settings:

*S. F. and J.-L. B. partially supported by NSF (DMS-1120622), P. H. by Simons Foundation (grant 208766)

- An idealized situation, where the vectors $\mathbf{x} \in \mathbb{C}^N$ are exactly sparse and where the measurements $\mathbf{y} \in \mathbb{C}^m$ are exactly equal to $\mathbf{A}\mathbf{x}$. In this case, the exact reconstruction of $\mathbf{x} \in \mathbb{C}^N$ is targeted.
- A realistic situation, where the vectors $\mathbf{x} \in \mathbb{C}^N$ are not exactly sparse and where the measurements $\mathbf{y} \in \mathbb{C}^m$ contain errors $\mathbf{e} \in \mathbb{C}^m$, i.e., $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$. In this case, only an approximate reconstruction of $\mathbf{x} \in \mathbb{C}^N$ is targeted. Precisely, the reconstruction error should be controlled by the sparsity defect and by the measurement error. The sparsity defect can be incorporated in the measurement error if $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ is rewritten as $\mathbf{y} = \mathbf{A}\mathbf{x}_S + \mathbf{e}'$ where S is an index set of s largest absolute entries of \mathbf{x} and $\mathbf{e}' := \mathbf{A}\mathbf{x}_{\bar{S}} + \mathbf{e}$.

We shall mainly state and prove our results in the realistic situation. They specialize to the idealized situation simply by setting $\mathbf{e}' = \mathbf{0}$. In fact, setting $\mathbf{e}' = \mathbf{0}$ inside the proofs would simplify them considerably.

Let us now recall that (HTP) consists in constructing a sequence (\mathbf{x}^n) of s -sparse vectors, starting with an initial s -sparse $\mathbf{x}^0 \in \mathbb{C}^N$ — we take $\mathbf{x}^0 = \mathbf{0}$ — and iterating the scheme¹

$$\begin{aligned} (\text{HTP}_1) \quad & S^n := \text{index set of } s \text{ largest absolute entries of } \mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}), \\ (\text{HTP}_2) \quad & \mathbf{x}^n := \operatorname{argmin}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^n\}, \end{aligned}$$

until a stopping criterion is met. It had been shown [10] that, in the idealized situation, exact reconstruction of every s -sparse $\mathbf{x} \in \mathbb{C}^N$ is achieved in s iterations of (HTP) with $\mathbf{y} = \mathbf{A}\mathbf{x}$ provided the coherence of the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies $\mu < 1/(3s)$ (note that this condition can be fulfilled when $m \asymp s^2$). Exact and approximate reconstructions were treated in [6], where it was in particular shown that every s -sparse $\mathbf{x} \in \mathbb{C}^N$ is the limit of the sequence (\mathbf{x}^n) produced by (HTP) with $\mathbf{y} = \mathbf{A}\mathbf{x}$ provided the 3st restricted isometry constant of the measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ obeys $\delta_{3s} < 1/\sqrt{3} \approx 0.577$ (note that this condition can be fulfilled when $m \asymp s \ln(N/s)$). As a reminder, the k th restricted isometry constant δ_k of \mathbf{A} is defined as the smallest constant $\delta \geq 0$ such that

$$(1 - \delta)\|\mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{z}\|_2^2 \leq (1 + \delta)\|\mathbf{z}\|_2^2 \quad \text{for all } k\text{-sparse } \mathbf{z} \in \mathbb{C}^N.$$

In fact, it was shown in [6] that the convergence is achieved in a finite number \bar{n} of iterations. In the idealized situation, it can be estimated as

$$(1) \quad \bar{n} \leq \left\lceil \frac{\ln(\sqrt{2/3}\|\mathbf{x}\|_2/\xi)}{\ln(1/\rho_{3s})} \right\rceil, \quad \rho_{3s} := \sqrt{\frac{2\delta_{3s}^2}{1 - \delta_{3s}^2}} < 1, \quad \xi := \min_{j \in \operatorname{supp}(\mathbf{x})} |x_j|.$$

This paper establishes that the number of iterations can be estimated independently of the shape of \mathbf{x} : under a restricted isometry condition, it is at most proportional to the sparsity s ,

¹exact arithmetic is assumed and, among several candidates for the index set of largest absolute entries, the smallest one in lexicographic order is always chosen

see Theorem 5 and a robust version in Theorem 6. This is reminiscent of the work of T. Zhang [15] on orthogonal matching pursuit (OMP), see also [7, Theorem 6.25] where it is proved that $\bar{n} \leq 12s$ provided that $\delta_{13s} < 1/6$.

However, (HTP) presents a significant drawback in that a prior estimation of the sparsity s is required to run the algorithm, while (OMP) does not (although stopping (OMP) at iteration $12s$ does require this estimation). We therefore consider a variation of (HTP) avoiding the prior estimation of s . We call it graded hard thresholding pursuit (GHTP) algorithm, because the index set has a size that increases with the iteration. Precisely, starting with $\mathbf{x}^0 = \mathbf{0}$, a sequence (\mathbf{x}^n) of n -sparse vectors is constructed according to

$$\begin{aligned} \text{(GHTP}_1\text{)} \quad S^n &:= \text{index set of } n \text{ largest absolute entries of } \mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}), \\ \text{(GHTP}_2\text{)} \quad \mathbf{x}^n &:= \operatorname{argmin}\{\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2, \operatorname{supp}(\mathbf{z}) \subseteq S^n\}, \end{aligned}$$

until a stopping criterion is met. For (GHTP), too, it is established that a restricted isometry condition implies that the number of iterations needed for robust reconstruction is at most proportional to the sparsity s , see Theorem 8.

One expects the number of (GHTP) iterations needed for s -sparse recovery to be exactly equal to s , though. Such a statement is indeed proved, but in the nonuniform setting rather than in the uniform setting (i.e., when successful recovery is sought with high probability on the draw of random measurements for a fixed $\mathbf{x} \in \mathbb{C}^N$ rather than for all $\mathbf{x} \in \mathbb{C}^N$ simultaneously) and in two extreme cases on the shape of the s -sparse vector to be recovered. The first case deals with vectors that are ‘flat’ on their support: robust nonuniform recovery is guaranteed with $m \asymp s \ln(N)$ Gaussian measurements, see Proposition 9. The second case deals with vectors that are decaying exponentially fast on their support: robust nonuniform recovery is guaranteed with $m \asymp \max\{s, \ln(N)\}$ Gaussian measurements or with $m \asymp s \ln(N)$ Fourier measurements, see Proposition 10. This is an improvement (attenuated by the extra shape condition) on uniform recovery results based on the restricted isometry property that require $m \asymp s \ln(N/s)$ Gaussian measurements or (currently) $m \asymp s \ln^4(N)$ Fourier measurements.

To conclude, the (HTP) and (GHTP) algorithms are tested numerically and benchmarked against the (OMP) algorithm. Their performance is roughly similar in terms of success rate, with different behaviors mostly explained by the shape of the vectors to be recovered. In terms of number of iterations, however, (HTP) does perform best, owing to the fact that an a priori knowledge of the sparsity s is used to run the algorithm (which also provides a convenient stopping criterion). It also seems that (GHTP) always requires less iterations than (OMP). The empirical observations that ‘flat’ vectors are least favorable, coupled with the theoretical findings of Sections 3 and 5, support the beliefs that uniform recovery via (HTP) is achievable in much less iterations than the upper bound c_s of Theorem 5 and that nonuniform recovery via (GHTP) is achievable in exactly s iterations whatever the shape of the vector. Finally, the numerical part ends with an experimental study of the influence of the measurement error.

Throughout the paper, we make use of the following notation:

- $\mathbf{x}^* \in \mathbb{R}_+^N$ is the nonincreasing rearrangement of a vector $\mathbf{x} \in \mathbb{C}^N$, i.e.,

$$x_1^* \geq x_2^* \geq \dots \geq x_N^* \geq 0$$

and there exists a permutation π of $\{1, \dots, N\}$ such that $x_j^* = |x_{\pi(j)}|$ for all $j \in \{1, \dots, N\}$;

- S is the support or an index set of s largest absolute entries of a vector $\mathbf{x} \in \mathbb{C}^N$;
- \mathbf{x}_T is the restriction of a vector $\mathbf{x} \in \mathbb{C}^N$ to an index set T — depending on the context, it is interpreted either as a vector in \mathbb{C}^T or as a vector in \mathbb{C}^N ;
- \bar{T} is the complementary set of an index set T , i.e., $\bar{T} = \{1, \dots, N\} \setminus T$;
- $T \Delta T'$ is the symmetric difference of sets T and T' , i.e., $T \Delta T' = (T \setminus T') \cup (T' \setminus T)$;
- ρ_s, τ_s are quantities associated to the s th restricted isometry constant $\delta_s < 1$ via

$$(2) \quad \rho_s = \sqrt{\frac{2\delta_s^2}{1-\delta_s^2}}, \quad \tau_s = \sqrt{\frac{2}{1-\delta_s}} + \frac{\sqrt{1+\delta_s}}{1-\delta_s}.$$

2 Preliminary Results

This section collects the key facts needed in the subsequent analyses of (HTP) and (GHTP). It also includes a discussion on appropriate stopping criteria.

2.1 Extensions of previous arguments

The arguments used in [6] to prove the convergence of the sequence (\mathbf{x}^n) produced by (HTP) in the idealized situation, say, relied on the exponential decay of $\|\mathbf{x} - \mathbf{x}^n\|_2$, precisely

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \rho_{3s} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2, \quad n \geq 1,$$

where $\rho_{3s} < 1$ provided $\delta_{3s} < 1/\sqrt{3}$. This followed from two inequalities derived as results of (HTP₁) and (HTP₂), namely

$$(3) \quad \|\mathbf{x}_{\bar{S}^n}\|_2 \leq \sqrt{2\delta_{3s}^2} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2,$$

$$(4) \quad \|\mathbf{x} - \mathbf{x}^n\|_2 \leq \sqrt{\frac{1}{1-\delta_{2s}^2}} \|\mathbf{x}_{\bar{S}^n}\|_2.$$

For the analyses of (HTP) and (GHTP), similar arguments are needed in the more general form below.

1. *Consequence of (HTP₁)-(GHTP₁).*

If $\mathbf{x} \in \mathbb{C}^N$ is s -sparse, if $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{C}^m$, if $\mathbf{x}' \in \mathbb{C}^N$ is s' -sparse, and if T is an index set of $t \geq s$ largest absolute entries of $\mathbf{x}' + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}')$, then

$$(5) \quad \|\mathbf{x}_{\overline{T}}\|_2 \leq \sqrt{2\delta_{s+s'+t}^2} \|\mathbf{x} - \mathbf{x}'\|_2 + \sqrt{2} \|(\mathbf{A}^*\mathbf{e})_{S\Delta T}\|_2.$$

Indeed, with $S := \text{supp}(\mathbf{x})$, we first notice that

$$\|(\mathbf{x}' + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_T\|_2^2 \geq \|(\mathbf{x}' + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_S\|_2^2.$$

Eliminating the contribution on $S \cap T$ gives

$$\|(\mathbf{x}' + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{T \setminus S}\|_2 \geq \|(\mathbf{x}' + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{S \setminus T}\|_2.$$

The left-hand side satisfies

$$\|(\mathbf{x}' + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{T \setminus S}\|_2 = \|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{T \setminus S}\|_2,$$

while the right-hand side satisfies

$$\|(\mathbf{x}' + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{S \setminus T}\|_2 \geq \|\mathbf{x}_{S \setminus T}\|_2 - \|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{S \setminus T}\|_2.$$

Therefore, we obtain

$$(6) \quad \begin{aligned} \|\mathbf{x}_{S \setminus T}\|_2 &\leq \|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{T \setminus S}\|_2 + \|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{S \setminus T}\|_2 \\ &\leq \sqrt{2} \|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}') + \mathbf{A}^*\mathbf{e})_{T\Delta S}\|_2 \\ &\leq \sqrt{2} \|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}'))_{T\Delta S}\|_2 + \sqrt{2} \|(\mathbf{A}^*\mathbf{e})_{T\Delta S}\|_2 \\ &\leq \sqrt{2} \delta_{s+s'+t} \|\mathbf{x} - \mathbf{x}'\|_2 + \sqrt{2} \|(\mathbf{A}^*\mathbf{e})_{T\Delta S}\|_2, \end{aligned}$$

where the last step used (3.8) of [6]. The resulting inequality is the one announced in (5).

2. *Consequence of (HTP₂)-(GHTP₂).*

If $\mathbf{x} \in \mathbb{C}^N$ is s -sparse, if $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{C}^m$, if T is an index set of size t , and if \mathbf{x}' is a minimizer of $\|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2$ subject to $\text{supp}(\mathbf{z}) \subseteq T$, then

$$(7) \quad \|\mathbf{x} - \mathbf{x}'\|_2 \leq \sqrt{\frac{1}{1 - \delta_{s+t}^2}} \|\mathbf{x}_{\overline{T}}\|_2 + \frac{1}{1 - \delta_{s+t}} \|(\mathbf{A}^*\mathbf{e})_T\|_2.$$

Indeed, the vector \mathbf{x}' is characterized by the orthogonality condition $(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}'))_T = \mathbf{0}$ (in other words, by the normal equations), therefore

$$(8) \quad \begin{aligned} \|(\mathbf{x} - \mathbf{x}')_T\|_2^2 &= \langle (\mathbf{x} - \mathbf{x}')_T, (\mathbf{x} - \mathbf{x}')_T \rangle = \langle (\mathbf{x} - \mathbf{x}' - \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}'))_T, (\mathbf{x} - \mathbf{x}')_T \rangle \\ &\leq |\langle ((\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{x} - \mathbf{x}'))_T, (\mathbf{x} - \mathbf{x}')_T \rangle| + |\langle (\mathbf{A}^*\mathbf{e})_T, (\mathbf{x} - \mathbf{x}')_T \rangle| \\ &\leq \|((\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{x} - \mathbf{x}'))_T\|_2 \|(\mathbf{x} - \mathbf{x}')_T\|_2 + \|(\mathbf{A}^*\mathbf{e})_T\|_2 \|(\mathbf{x} - \mathbf{x}')_T\|_2 \\ &\leq \delta_{s+t} \|\mathbf{x} - \mathbf{x}'\|_2 \|(\mathbf{x} - \mathbf{x}')_T\|_2 + \|(\mathbf{A}^*\mathbf{e})_T\|_2 \|(\mathbf{x} - \mathbf{x}')_T\|_2, \end{aligned}$$

where the last step used (3.8) of [6]. After dividing through by $\|(\mathbf{x} - \mathbf{x}')_T\|_2$, we obtain

$$\|(\mathbf{x} - \mathbf{x}')_T\|_2 \leq \delta_{s+t} \|\mathbf{x} - \mathbf{x}'\|_2 + \|(\mathbf{A}^* \mathbf{e})_T\|_2.$$

It then follows that

$$\|\mathbf{x} - \mathbf{x}'\|_2^2 = \|(\mathbf{x} - \mathbf{x}')_T\|_2^2 + \|(\mathbf{x} - \mathbf{x}')_{\bar{T}}\|_2^2 \leq (\delta_{s+t} \|\mathbf{x} - \mathbf{x}'\|_2 + \|(\mathbf{A}^* \mathbf{e})_T\|_2)^2 + \|\mathbf{x}_{\bar{T}}\|_2^2,$$

in other words $P(\|\mathbf{x} - \mathbf{x}'\|_2) \leq 0$, where the quadratic polynomial P is defined by

$$P(z) = (1 - \delta_{s+t}^2)z^2 - 2\delta_{s+t}\|(\mathbf{A}^* \mathbf{e})_T\|_2 z - \|(\mathbf{A}^* \mathbf{e})_T\|_2^2 - \|\mathbf{x}_{\bar{T}}\|_2^2.$$

This implies that $\|\mathbf{x} - \mathbf{x}'\|_2$ is smaller than the largest root of P , i.e.,

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}'\|_2 &\leq \frac{\delta_{s+t}\|(\mathbf{A}^* \mathbf{e})_T\|_2 + \sqrt{\|(\mathbf{A}^* \mathbf{e})_T\|_2^2 + (1 - \delta_{s+t}^2)\|\mathbf{x}_{\bar{T}}\|_2^2}}{1 - \delta_{s+t}^2} \\ &\leq \frac{(1 + \delta_{s+t})\|(\mathbf{A}^* \mathbf{e})_T\|_2 + \sqrt{(1 - \delta_{s+t}^2)\|\mathbf{x}_{\bar{T}}\|_2}}{1 - \delta_{s+t}^2}. \end{aligned}$$

This is the inequality announced in (7).

Remark 1. The full restricted isometry property is not needed to derive (5) and (7). The restricted isometry constants only appear in (6) and (8), where $\|((\mathbf{I} - \mathbf{A}^* \mathbf{A})(\mathbf{x} - \mathbf{x}'))_U\|_2$ is bounded for $U = T\Delta S$ and $U = T$. This is easier to fulfill than the restricted isometry property if $\text{supp}(\mathbf{x})$, $\text{supp}(\mathbf{x}')$, and T are all included in a fixed index set, as will be the case in Section 5.

From inequalities (5) and (7), we deduce the decay of the sequences $(\|\mathbf{x} - \mathbf{x}^n\|_2)_{n \geq 0}$ for (HTP) and $(\|\mathbf{x} - \mathbf{x}^n\|_2)_{n \geq s}$ for (GHTP).

Corollary 2. Let $\mathbf{x} \in \mathbb{C}^N$ be s -sparse and let $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{C}^m$. If (\mathbf{x}^n) is the sequence produced by (HTP), then

$$(9) \quad \|\mathbf{x} - \mathbf{x}^n\|_2 \leq \rho_{3s} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \tau_{2s} \|\mathbf{e}\|_2 \quad \text{for any } n \geq 1.$$

If (\mathbf{x}^n) is the sequence produced by (GHTP), then

$$(10) \quad \|\mathbf{x} - \mathbf{x}^n\|_2 \leq \rho_{s+2n-1} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \tau_{s+n} \|\mathbf{e}\|_2 \quad \text{for any } n \geq s.$$

Proof. We consider (HTP) first. Applying (5) to $\mathbf{x}' = \mathbf{x}^{n-1}$ and $T = S^n$ gives

$$\|\mathbf{x}_{S^n}\|_2 \leq \sqrt{2\delta_{3s}^2} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \sqrt{2} \|(\mathbf{A}^* \mathbf{e})_{S\Delta S^n}\|_2$$

and applying (7) to $T = S^n$ and $\mathbf{x}' = \mathbf{x}^n$ gives

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \sqrt{\frac{1}{1 - \delta_{2s}^2}} \|\mathbf{x}_{S^n}\|_2 + \frac{1}{1 - \delta_{2s}} \|(\mathbf{A}^* \mathbf{e})_{S^n}\|_2.$$

Combining these two inequalities, while using $\delta_{2s} \leq \delta_{3s}$, yields

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \sqrt{\frac{2\delta_{3s}^2}{1 - \delta_{3s}^2}} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \left(\sqrt{\frac{2}{1 - \delta_{2s}^2}} + \frac{1}{1 - \delta_{2s}} \right) \|(\mathbf{A}^* \mathbf{e})_{S \cup S^n}\|_2.$$

In view of $\|(\mathbf{A}^* \mathbf{e})_{S \cup S^n}\|_2 \leq \sqrt{1 + \delta_{2s}} \|\mathbf{e}\|_2$ (see e.g. (3.19) in [6]), we recognize (9) with the values of ρ_{3s} and τ_{2s} given in (2).

Let us now consider (GHTP). Applying (5) to $\mathbf{x}' = \mathbf{x}^{n-1}$ and $T = S^n$ for $n \geq s$ gives

$$\|\mathbf{x}_{S^n}\|_2 \leq \sqrt{2\delta_{s+2n-1}^2} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \sqrt{2} \|(\mathbf{A}^* \mathbf{e})_{S \Delta S^n}\|_2$$

and applying (7) to $T = S^n$ and $\mathbf{x}' = \mathbf{x}^n$ gives

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \sqrt{\frac{1}{1 - \delta_{s+n}^2}} \|\mathbf{x}_{S^n}\|_2 + \frac{1}{1 - \delta_{s+n}} \|(\mathbf{A}^* \mathbf{e})_{S^n}\|_2.$$

Combining these two inequalities, while using $\delta_{s+n} \leq \delta_{s+2n-1}$, yields

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \sqrt{\frac{2\delta_{s+2n-1}^2}{1 - \delta_{s+2n-1}^2}} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \left(\sqrt{\frac{2}{1 - \delta_{s+n}^2}} + \frac{1}{1 - \delta_{s+n}} \right) \|(\mathbf{A}^* \mathbf{e})_{S \cup S^n}\|_2.$$

In view of $\|(\mathbf{A}^* \mathbf{e})_{S \cup S^n}\|_2 \leq \sqrt{1 + \delta_{s+n}} \|\mathbf{e}\|_2$, we recognize (10) with the values of ρ_{s+2n-1} and τ_{s+n} given in (2). \square

2.2 Additional arguments

While the above argument probed only the closeness of \mathbf{x}^n to \mathbf{x} , the following lemmas examine the indices of nonzero entries of \mathbf{x} that are correctly captured in the support sets produced by (HTP) and (GHTP). They show how many iterations are necessary to increase the number of correct indices by a specified amount. The first lemma concerns (HTP).

Lemma 3. Let $\mathbf{x} \in \mathbb{C}^N$ be s -sparse and let (S^n) be the sequence of index sets produced by (HTP) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{C}^m$. For integers $n, p \geq 0$, suppose that S^n contains the indices of p largest absolute entries of \mathbf{x} . Then, for integers $k, q \geq 1$, S^{n+k} contains the indices of $p + q$ largest absolute entries of \mathbf{x} , provided

$$(11) \quad x_{p+q}^* > \rho_{3s}^k \|\mathbf{x}_{\{p+1, \dots, s\}}^*\|_2 + \kappa_{3s} \|\mathbf{e}\|_2,$$

where the constant κ_{3s} depends only on δ_{3s} . For instance, $\delta_{3s} < 1/3$ yields $\kappa_{3s} \leq 5.72$.

Proof. Let π be the permutation of $\{1, 2, \dots, N\}$ for which $|x_{\pi(j)}| = x_j^*$ for all $j \in \{1, 2, \dots, N\}$. Our hypothesis is $\pi(\{1, \dots, p\}) \subseteq S^n$ and our aim is to prove that $\pi(\{1, \dots, p + q\}) \subseteq S^{n+k}$.

In other words, we aim at proving that the $|(\mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\pi(j)}|$ for $j \in \{1, \dots, p+q\}$ are among the s largest values of $|(\mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_i|$ for $i \in \{1, \dots, N\}$. With $\text{supp}(\mathbf{x}) \subseteq S$, $\text{card}(S) = s$, it is enough to show that

$$(12) \quad \min_{j \in \{1, \dots, p+q\}} |(\mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\pi(j)}| > \max_{\ell \in \bar{S}} |(\mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\ell}|.$$

We notice that, for $j \in \{1, \dots, p+q\}$ and $\ell \in \bar{S}$,

$$\begin{aligned} |(\mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\pi(j)}| &\geq |x_{\pi(j)}| - |(-\mathbf{x} + \mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\pi(j)}| \\ &\geq x_{p+q}^* - |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n+k-1}) + \mathbf{A}^*\mathbf{e})_{\pi(j)}|, \\ |(\mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\ell}| &= |(-\mathbf{x} + \mathbf{x}^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\ell}| \\ &= |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n+k-1}) + \mathbf{A}^*\mathbf{e})_{\ell}|. \end{aligned}$$

Therefore, (12) will be shown as soon as it is proved that, for all $j \in \{1, \dots, p+q\}$ and $\ell \in \bar{S}$,

$$(13) \quad x_{p+q}^* > |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n+k-1}) + \mathbf{A}^*\mathbf{e})_{\pi(j)}| + |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n+k-1}) + \mathbf{A}^*\mathbf{e})_{\ell}|.$$

The right-hand side can be bounded by

$$\begin{aligned} &\sqrt{2} \left\| ((\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{x} - \mathbf{x}^{n+k-1}) + \mathbf{A}^*\mathbf{e})_{\{\pi(j), \ell\}} \right\|_2 \\ &\leq \sqrt{2} \left\| ((\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{x} - \mathbf{x}^{n+k-1}))_{\{\pi(j), \ell\}} \right\|_2 + \sqrt{2} \left\| (\mathbf{A}^*\mathbf{e})_{\{\pi(j), \ell\}} \right\|_2 \\ &\leq \sqrt{2} \delta_{2s+2} \|\mathbf{x} - \mathbf{x}^{n+k-1}\|_2 + \sqrt{2} \sqrt{1 + \delta_2} \|\mathbf{e}\|_2 \\ &\leq \sqrt{2} \delta_{3s} \left(\rho_{3s}^{k-1} \|\mathbf{x} - \mathbf{x}^n\|_2 + \frac{\tau_{2s}}{1 - \rho_{3s}} \|\mathbf{e}\|_2 \right) + \sqrt{2(1 + \delta_{2s})} \|\mathbf{e}\|_2, \end{aligned}$$

where (9) was used $k-1$ times in the last step. Using (7) and the assumption that $\pi(\{1, \dots, p\}) \subseteq S^n$, we derive

$$\begin{aligned} &\sqrt{2} \left\| ((\mathbf{I} - \mathbf{A}^*\mathbf{A})(\mathbf{x} - \mathbf{x}^{n+k-1}) + \mathbf{A}^*\mathbf{e})_{\{\pi(j), \ell\}} \right\|_2 \\ &\leq \sqrt{2} \delta_{3s} \rho_{3s}^{k-1} \left(\frac{1}{\sqrt{1 - \delta_{2s}^2}} \|\mathbf{x}_{S^n}\|_2 + \frac{1}{1 - \delta_{2s}} \|(\mathbf{A}^*\mathbf{e})_{S^n}\|_2 \right) + \left(\frac{\sqrt{2} \delta_{3s} \tau_{2s}}{1 - \rho_{3s}} + \sqrt{2(1 + \delta_{2s})} \right) \|\mathbf{e}\|_2 \\ &\leq \rho_{3s}^k \|\mathbf{x}_{\pi(\{1, \dots, p\})}\|_2 + \left(\frac{\sqrt{2} \delta_{3s} \rho_{3s}^{k-1} \sqrt{1 + \delta_s}}{1 - \delta_{2s}} + \frac{\sqrt{2} \delta_{3s} \tau_{2s}}{1 - \rho_{3s}} + \sqrt{2(1 + \delta_{2s})} \right) \|\mathbf{e}\|_2 \\ &\leq \rho_{3s}^k \|\mathbf{x}_{\{p+1, \dots, s\}}^*\|_2 + \kappa_{3s} \|\mathbf{e}\|_2, \end{aligned}$$

where the constant κ_{3s} is determined by

$$\kappa_{3s} = \frac{\sqrt{2(1 + \delta_{3s})}}{1 - \delta_{3s}} + \frac{\sqrt{2} \delta_{3s} \tau_{3s}}{1 - \rho_{3s}}.$$

Therefore, (13) is proved as soon as condition (11) holds. The proof is completed by noticing that $\delta_{3s} \leq 1/3$ yields $\rho_{3s} \leq 1/2$, $\tau_{3s} \leq 2\sqrt{3}$, and finally $\kappa_{3s} \leq 14/\sqrt{6}$. \square

For (GHTP), an analog of Lemma 3 is needed in the form below.

Lemma 4. Let $\mathbf{x} \in \mathbb{C}^N$ be s -sparse and let (S^n) be the sequence of index sets produced by (GHTP) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some $\mathbf{e} \in \mathbb{C}^m$. For integers $n \geq s$ and $p \geq 0$, suppose that S^n contains the indices of p largest absolute entries of \mathbf{x} . Then, for integers $k, q \geq 1$, S^{n+k} contains the indices of $p+q$ largest absolute entries of \mathbf{x} , provided

$$(14) \quad x_{p+q}^* > \rho_{s+2n+2k}^k \|\mathbf{x}_{\{p+1, \dots, s\}}^*\|_2 + \kappa_{s+2n+2k} \|\mathbf{e}\|_2,$$

where the constant $\kappa_{s+2n+2k}$ depends only on $\delta_{s+2n+2k}$.

Proof. Given the permutation π of $\{1, 2, \dots, N\}$ for which $|x_{\pi(j)}| = x_j^*$ for all $j \in \{1, \dots, N\}$, our hypothesis is $\pi(\{1, \dots, p\}) \subseteq S^n$ and our aim is to prove that $\pi(\{1, \dots, p+q\}) \subseteq S^{n+k}$. For this purpose, we aim at proving that the $|(x^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_{\pi(j)}|$ for $j \in \{1, \dots, p+q\}$ are among the $n+k$ largest values of $|(x^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_i|$, $i \in \{1, \dots, N\}$. Since $n+k \geq s$, it is enough to show that they are among the s largest values of $|(x^{n+k-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n+k-1}))_i|$, $i \in \{1, \dots, N\}$. The rest of the proof duplicates the proof of Lemma 3 modulo the change imposed by (10) for the indices of δ and κ . \square

2.3 Stopping the (HTP) and (GHTP) algorithms

The stopping criterion, which is an integral part of the algorithm, has not been addressed yet. We take a few moments to discuss this issue here. For (HTP), the natural stopping criterion² is $S^n = S^{n-1}$, since the algorithm always produce the same vector afterwards. Moreover, under the condition $\delta_{3s} < 1/\sqrt{3}$,

1. if $S^n = S^{n-1}$ occurs, then robustness is guaranteed, since $\mathbf{x}^n = \mathbf{x}^{n-1}$ substituted in (9) yields $\|\mathbf{x} - \mathbf{x}^{n-1}\|_2 \leq \rho_{3s} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 + \tau_{2s} \|\mathbf{e}\|_2$, i.e.,

$$\|\mathbf{x} - \mathbf{x}^{n-1}\|_2 \leq \frac{\tau_{2s}}{1 - \rho_{3s}} \|\mathbf{e}\|_2;$$

2. and $S^n = S^{n-1}$ does occur provided $\|\mathbf{e}\|_2 < 4[\tau_{2s}/(1 - \rho_{3s})] x_s^*$. Indeed, for $\|\mathbf{e}\|_2 > 0^3$ and for n large enough to have $\rho_{3s}^{n-1} \|\mathbf{x}\|_2 \leq [\tau_{2s}/(1 - \rho_{3s})] \|\mathbf{e}\|_2$, the inequality (9) applied $n-1$ times implies that $\|\mathbf{x} - \mathbf{x}^{n-1}\|_2 \leq \rho_{3s}^{n-1} \|\mathbf{x}\|_2 + [\tau_{2s}/(1 - \rho_{3s})] \|\mathbf{e}\|_2 \leq 2[\tau_{2s}/(1 - \rho_{3s})] \|\mathbf{e}\|_2$. In turn, for $j \in \text{supp}(\mathbf{x})$ and $\ell \notin \text{supp}(\mathbf{x})$, the inequalities

$$\begin{aligned} |x_j^{n-1}| &\geq |x_j| - |(\mathbf{x} - \mathbf{x}^{n-1})_j| \geq x_s^* - 2[\tau_{2s}/(1 - \rho_{3s})] \|\mathbf{e}\|_2 > 2[\tau_{2s}/(1 - \rho_{3s})] \|\mathbf{e}\|_2, \\ |x_\ell^{n-1}| &= |(\mathbf{x} - \mathbf{x}^{n-1})_\ell| \leq 2[\tau_{2s}/(1 - \rho_{3s})] \|\mathbf{e}\|_2 \end{aligned}$$

show that $S^{n-1} = \text{supp}(\mathbf{x})$. The same reasoning gives $S^n = \text{supp}(\mathbf{x})$, hence $S^n = S^{n-1}$.

²again, exact arithmetic is assumed and the lexicographic rule to break ties applies

³in the case $\mathbf{e} = \mathbf{0}$, we have $\mathbf{x}^{n-1} = \mathbf{x}$ for some $n \geq 1$ (see (1)), hence $\mathbf{x}^n = \mathbf{x}$, and in particular $S^{n-1} = S^n$

For (GHTP), a stopping criterion not requiring any estimation of s or $\|\mathbf{e}\|_2$ seems to be lost. It is worth recalling that iterative algorithms such as CoSaMP [12], IHT [2], or HTP [6] typically involve an estimation of s and that optimization methods such as quadratically constrained ℓ_1 -minimization typically involve an estimation of $\|\mathbf{e}\|_2$ — unless the measurement matrix satisfies the ℓ_1 -quotient property, see [14]. Here, we settle for a stopping criterion requiring an estimation of either one of s or $\|\mathbf{e}\|_2$, precisely (GHTP) is stopped either when $n = 4s$ or when $\|\mathbf{y} - \mathbf{A}\mathbf{x}^n\|_2 \leq d'\|\mathbf{e}\|_2$ for some constant $d' \geq 2.83$, say. Indeed, under the condition $\delta_{9s} < 1/3$, robustness is guaranteed from Theorem 8 by the fact that $\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq d\|\mathbf{e}\|_2$ for some $\bar{n} \leq 4s$ and some $d \leq 2.45$. Thus, in the alternative $n = 4s$,

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \rho_{9s}^{4s-\bar{n}} \|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 + \frac{\tau_{9s}}{1-\rho_{9s}} \|\mathbf{e}\|_2 \leq \left(d + \frac{\tau_{9s}}{1-\rho_{9s}} \right) \|\mathbf{e}\|_2 \leq d'' \|\mathbf{e}\|_2$$

and in the alternative $\|\mathbf{y} - \mathbf{A}\mathbf{x}^n\|_2 \leq d'\|\mathbf{e}\|_2$ (which is guaranteed to occur for some $n \leq \bar{n}$ by virtue of $\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\bar{n}}\|_2 \leq \sqrt{1+\delta_{9s}} \|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq \sqrt{1+\delta_{9s}} d \|\mathbf{e}\|_2$),

$$\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \frac{1}{\sqrt{1-\delta_{9s}}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^n)\|_2 \leq \frac{1}{\sqrt{1-\delta_{9s}}} (\|\mathbf{y} - \mathbf{A}\mathbf{x}^n\|_2 + \|\mathbf{e}\|_2) \leq \frac{d'+1}{\sqrt{1-\delta_{9s}}} \|\mathbf{e}\|_2 \leq d'' \|\mathbf{e}\|_2,$$

where $d'' := \max\{d + \tau_{9s}/(1-\rho_{9s}), (d'+1)/\sqrt{1-\delta_{9s}}\} \leq 9.38$. We point out that, according to the result of [15], a similar stopping criterion involving either one of s or $\|\mathbf{e}\|_2$ can be used for (OMP). This makes the resemblance between (GHTP) and (OMP) even more compelling. Their empirical performances will be compared in Section 6.

3 Uniform Recovery via Hard Thresholding Pursuit

This section is dedicated to the main results about (HTP), namely that sparse recovery is achieved in a number of iterations at most proportional to the sparsity level, independently of the shape of the vector to be recovered. The results are uniform in the sense that the same measurement matrix allows for the recovery of all s -sparse vectors simultaneously. For clarity of exposition, we first state and prove a result in the idealized situation — only the realistic situation is considered thereafter.

Theorem 5. If the restricted isometry constant of the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ obeys

$$\delta_{3s} \leq \frac{1}{\sqrt{5}},$$

then every s -sparse vector $\mathbf{x} \in \mathbb{C}^N$ is recovered from the measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^m$ via a number \bar{n} of iterations of (HTP) satisfying

$$\bar{n} \leq cs.$$

The constant $c \leq 3$ depends only on δ_{3s} . For instance, $\delta_{3s} \leq 1/3$ yields $c \leq 2$.

Proof. Let π be the permutation of $\{1, 2, \dots, N\}$ for which $|x_{\pi(j)}| = x_j^*$ for all $j \in \{1, 2, \dots, N\}$. Our goal is to prove that $\text{supp}(\mathbf{x}) \subseteq S^{\bar{n}}$ for $\bar{n} \leq 3s$, hence (HTP₂) ensures that $\mathbf{A}\mathbf{x}^{\bar{n}} = \mathbf{A}\mathbf{x}$ and $\mathbf{x}^{\bar{n}} = \mathbf{x}$ follows from $\delta_{3s} < 1$. For this purpose, we shall consider a partition $Q_1 \cup Q_2 \cup \dots \cup Q_r$, $r \leq s$, of $\text{supp}(\mathbf{x}) = \pi(\{1, \dots, s\})$. Lemma 3 enables to argue repeatedly that, if $Q_1 \cup \dots \cup Q_{i-1}$ has been correctly captured, then a suitable number k_i of additional (HTP) iterations captures $Q_1 \cup \dots \cup Q_{i-1} \cup Q_i$. The remaining step is just the estimation of $k_1 + \dots + k_r$. Let us set $q_0 = 0$ and let us define the sets Q_1, \dots, Q_r inductively by

$$(15) \quad Q_i = \pi(\{q_{i-1} + 1, \dots, q_i\}), \quad q_i := \text{maximum index} \geq q_{i-1} + 1 \text{ such that } x_{q_i}^* > \frac{1}{\sqrt{2}}x_{q_{i-1}+1}^*.$$

We notice that $x_{q_{i+1}}^* \leq x_{q_{i-1}+1}^*/\sqrt{2}$ for all $i \in \{1, \dots, r-1\}$. With the artificial introduction of $Q_0 = \emptyset$ and $k_0 = 0$, we now prove by induction on $i \in \{0, \dots, r\}$ that

$$(16) \quad Q_0 \cup Q_1 \cup \dots \cup Q_i \subseteq S^{k_0+k_1+\dots+k_i},$$

where k_1, \dots, k_r are defined, with $\rho := \rho_{3s} < 1$, as

$$(17) \quad k_i := \left\lceil \frac{\ln(2(\text{card}(Q_i) + \text{card}(Q_{i+1})/2 + \dots + \text{card}(Q_r)/2^{r-i}))}{\ln(1/\rho^2)} \right\rceil.$$

For $i = 0$, the result holds trivially. Next, if (16) holds for $i-1$, $i \in \{1, \dots, r\}$, Lemma 3 (with $\mathbf{e} = \mathbf{0}$) guarantees that (16) holds for i , provided

$$(18) \quad (x_{q_i}^*)^2 > \rho^{2k_i} (\|\mathbf{x}_{Q_i}\|_2^2 + \|\mathbf{x}_{Q_{i+1}}\|_2^2 + \dots + \|\mathbf{x}_{Q_r}\|_2^2).$$

In view of $(x_{q_i}^*)^2 > (x_{q_{i-1}+1}^*)^2/2$ and of

$$\begin{aligned} & \|\mathbf{x}_{Q_i}\|_2^2 + \|\mathbf{x}_{Q_{i+1}}\|_2^2 + \dots + \|\mathbf{x}_{Q_r}\|_2^2 \\ & \leq (x_{q_{i-1}+1}^*)^2 \text{card}(Q_i) + (x_{q_i+1}^*)^2 \text{card}(Q_{i+1}) + \dots + (x_{q_{r-1}+1}^*)^2 \text{card}(Q_r) \\ & \leq (x_{q_{i-1}+1}^*)^2 \left(\text{card}(Q_i) + \frac{1}{2} \text{card}(Q_{i+1}) + \dots + \frac{1}{2^{r-i}} \text{card}(Q_r) \right), \end{aligned}$$

we verify that condition (18) is fulfilled thanks to the definition (17) of k_i . This concludes the inductive proof. We derive that the support $\pi(\{1, \dots, s\})$ of \mathbf{x} is recovered in a number of iterations at most

$$\begin{aligned} \bar{n} &= \sum_{i=1}^r k_i \leq \sum_{i=1}^r \left(1 + \frac{\ln(2(\text{card}(Q_i) + \text{card}(Q_{i+1})/2 + \dots + \text{card}(Q_r)/2^{r-i}))}{\ln(1/\rho^2)} \right) \\ &= r + \frac{r}{\ln(1/\rho^2)} \sum_{i=1}^r \frac{1}{r} \ln(2(\text{card}(Q_i) + \text{card}(Q_{i+1})/2 + \dots + \text{card}(Q_r)/2^{r-i})). \end{aligned}$$

Using the concavity of the logarithm, we obtain

$$\begin{aligned} \ln(1/\rho^2) \frac{\bar{n} - r}{r} &\leq \ln \left(\sum_{i=1}^r \frac{2}{r} (\text{card}(Q_i) + \text{card}(Q_{i+1})/2 + \dots + \text{card}(Q_r)/2^{r-i}) \right) \\ &= \ln \left(\frac{2}{r} (\text{card}(Q_1) + (1 + 1/2)\text{card}(Q_2) + \dots + (1 + 1/2 + \dots + 1/2^{r-1})\text{card}(Q_r)) \right) \\ &\leq \ln \left(\frac{4}{r} (\text{card}(Q_1) + \text{card}(Q_2) + \dots + \text{card}(Q_r)) \right) = \ln \left(\frac{4s}{r} \right). \end{aligned}$$

Since $\ln(4x)/x \leq \ln(4)$ for $x \geq 1$, we deduce that

$$\ln(1/\rho^2) \frac{\bar{n} - r}{s} \leq \frac{\ln(4s/r)}{s/r} \leq \ln(4),$$

and we subsequently obtain

$$\bar{n} \leq r + \frac{\ln(4)}{\ln(1/\rho^2)} s \leq \left(1 + \frac{\ln(4)}{\ln(1/\rho^2)}\right) s = \frac{\ln(4/\rho^2)}{\ln(1/\rho^2)} s.$$

This is the desired result with $c = \ln(2(1 - \delta_{3s}^2)/\delta_{3s}^2)/\ln((1 - \delta_{3s}^2)/(2\delta_{3s}^2))$. We observe that $\delta_{3s} \leq 1/\sqrt{5}$ yields $c \leq \ln(8)/\ln(2) = 3$ and that $\delta_{3s} \leq 1/3$ yields $c \leq \ln(16)/\ln(4) = 2$. The proof is complete. \square

The approach above does not allow for the reduction of the number of iterations below $c s$. Indeed, if the s -sparse vector $\mathbf{x} \in \mathbb{C}^N$ has nonzero entries $x_j = \mu^j$ for $\mu \in (0, 1/2)$, then each set Q_i consists of one element, so there are s of these sets, and $\bar{n} = \sum_{i=1}^s k_i \geq s$. However, such a vector whose nonincreasing rearrangement decays exponentially fast is highly compressible — applying Theorem 6 turns out to be more appropriate (see Remark 7). For other types of sparse vectors, the number of iterations can be significantly lowered. We give two examples below. In each of them, the partition involves a single set Q_1 . This corresponds to invoking (1).

Sparse vectors whose nonincreasing rearrangement decays slowly. For s -sparse vectors $\mathbf{x} \in \mathbb{C}^N$ satisfying $x_j^* = 1/j^\alpha$ for all $j \in \{1, \dots, s\}$ and for some $\alpha \geq 0$ (‘flat’ vectors are included with the choice $\alpha = 0$), the number of iterations is at most proportional to $\ln(s)$. Indeed, we have

$$\frac{\|\mathbf{x}\|_2^2}{(x_s^*)^2} = \frac{1 + \dots + 1/s^{2\alpha}}{1/s^{2\alpha}} \leq \frac{s}{1/s^{2\alpha}} = s^{2\alpha+1},$$

and it follows from (1) that

$$\bar{n} \leq \left\lceil \frac{\ln(\sqrt{2/3} s^{\alpha+1/2})}{\ln(1/\rho_{3s})} \right\rceil \leq c_{\delta_{3s}, \alpha} \ln(s).$$

Gaussian vectors. In numerical experiments, the nonzero entries of an s -sparse vector $\mathbf{g} \in \mathbb{C}^N$ are often taken as independent standard Gaussian random variables g_j , $j \in S$. In this case, with probability at least $1 - \varepsilon$, the vector \mathbf{g} is recovered from $\mathbf{y} = \mathbf{A}\mathbf{g}$ with a number of (HTP) iterations at most proportional to $\ln(s/\varepsilon)$. Indeed, we establish below that

$$(19) \quad \mathbb{P}\left(g_s^* < \sqrt{\frac{\pi}{8}} \frac{\varepsilon}{s}\right) \leq \frac{\varepsilon}{2},$$

$$(20) \quad \mathbb{P}\left(\|\mathbf{g}\|_2^2 > \frac{3\pi}{2} \frac{s}{\varepsilon}\right) \leq \frac{\varepsilon}{2}.$$

Thus, with failure probability at most ε , we have $g_s^* \geq \sqrt{\pi/8} \varepsilon/s$ and $\|\mathbf{g}\|_2 \leq \sqrt{3\pi/2} \sqrt{s/\varepsilon}$. It follows from (1) that the number of iterations necessary to recover \mathbf{g} satisfies

$$\bar{n} \leq \left\lceil \frac{\ln((2s/\varepsilon)^{3/2})}{\ln(1/\rho_{3s})} \right\rceil = \left\lceil \frac{3 \ln(2s/\varepsilon)}{2 \ln(1/\rho_{3s})} \right\rceil \leq c_{\delta_{3s}} \ln(s/\varepsilon).$$

To obtain (19), with g denoting a standard Gaussian random variable and with $t := \sqrt{\pi/8} \varepsilon/s$, we write

$$\mathbb{P}(g_s^* < t) = \mathbb{P}(|g_j| < t \text{ for some } j \in S) \leq s \mathbb{P}(|g| < t) = s \int_{-t}^t \frac{\exp(-u^2/2)}{\sqrt{2\pi}} du \leq s \sqrt{\frac{2}{\pi}} t = \varepsilon/2.$$

To obtain (20), with $u \in (0, 1/4)$ to be chosen later and with $t := (3\pi/2)s/\varepsilon$, we write

$$\mathbb{P}(\|\mathbf{g}\|_2^2 > t) = \mathbb{P}\left(\sum_{j \in S} g_j^2 > t\right) \leq \frac{\mathbb{E}\left(\exp\left(u \sum_{j \in S} g_j^2\right)\right)}{\exp(ut)} = \frac{\prod_{j \in S} \mathbb{E}\left(\exp\left(ug_j^2\right)\right)}{\exp(ut)} = \frac{((1-2u)^{-1/2})^s}{\exp(ut)}.$$

Using $(1-2u)^{-1/2} = (1+2u/(1-2u))^{1/2} \leq (1+4u)^{1/2} \leq \exp(2u)$ for $u \in (0, 1/4)$, we derive

$$\mathbb{P}(\|\mathbf{g}\|_2^2 > t) \leq \exp(2us - ut) = \exp\left(-us \left(\frac{3\pi}{2\varepsilon} - 2\right)\right) = \frac{\varepsilon}{2},$$

where we have chosen $us = \ln(2/\varepsilon)/(3\pi/(2\varepsilon) - 2)$, so that $u \leq us < 1/4$ when $\varepsilon < 1/2$.

We now state the extension of Theorem 5 to the realistic situation. We need to assume that the measurement error is not too large compared with the smallest nonzero absolute entry of the sparse vector — such a condition (quite common in the literature, see e.g. [11, Theorem 3.2] or [9, Theorem 4.1]) is further discussed in Subsection 6.3.

Theorem 6. If the restricted isometry constant of the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ obeys

$$\delta_{3s} \leq \frac{1}{3},$$

then the sequence (\mathbf{x}^n) produced by (HTP) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{C}^m$ for some s -sparse $\mathbf{x} \in \mathbb{C}^N$ and some $\mathbf{e} \in \mathbb{C}^m$ with $\|\mathbf{e}\|_2 \leq \gamma x_s^*$ satisfies

$$\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq d\|\mathbf{e}\|_2, \quad \bar{n} \leq cs.$$

The constants $c \leq 3$, $d \leq 2.45$, and $\gamma \geq 0.079$ depend only on δ_{3s} .

Proof. Let π be the permutation of $\{1, 2, \dots, N\}$ for which $|x_{\pi(j)}| = x_j^*$ for all $j \in \{1, 2, \dots, N\}$. As before, we partition $\text{supp}(\mathbf{x}) = \pi(\{1, \dots, s\})$ as $Q_1 \cup Q_2 \cup \dots \cup Q_r$, $r \leq s$, where the Q_i are defined as in (15). With $Q_0 = \emptyset$ and $k_0 = 0$, we prove by induction on $i \in \{0, \dots, r\}$ that $Q_0 \cup Q_1 \cup \dots \cup Q_i \subseteq S^{k_0+k_1+\dots+k_i}$, where k_1, \dots, k_r are defined, with $\rho := \rho_{3s} < 1$, by

$$(21) \quad k_i := \left\lceil \frac{\ln(16(\text{card}(Q_i) + \text{card}(Q_{i+1})/2 + \dots + \text{card}(Q_r)/2^{r-i}))}{\ln(1/\rho^2)} \right\rceil.$$

The induction hypothesis holds trivially for $i = 0$. Next, if it holds for $i-1$, $i \in \{1, \dots, r\}$, then Lemma 3 guarantees that it holds for i , provided

$$(22) \quad x_{q_i}^* > \rho^{k_i} \sqrt{\|\mathbf{x}_{Q_i}\|_2^2 + \|\mathbf{x}_{Q_{i+1}}\|_2^2 + \dots + \|\mathbf{x}_{Q_r}\|_2^2 + \kappa_{3s}\|\mathbf{e}\|_2}.$$

With the choice $\gamma := (2\sqrt{2} - 1)/(4\kappa_{3s})$, we have $\kappa_{3s}\|\mathbf{e}\|_2 \leq \kappa_{3s}\gamma x_s^* \leq (1/\sqrt{2} - 1/4)x_{q_{i-1}+1}^*$. Moreover, in view of $x_{q_i}^* > x_{q_{i-1}+1}^*/\sqrt{2}$ and of

$$\sqrt{\|\mathbf{x}_{Q_i}\|_2^2 + \|\mathbf{x}_{Q_{i+1}}\|_2^2 + \cdots + \|\mathbf{x}_{Q_r}\|_2^2} \leq x_{q_{i-1}+1}^* \sqrt{\text{card}(Q_i) + \frac{1}{2}\text{card}(Q_{i+1}) + \cdots + \frac{1}{2^{r-i}}\text{card}(Q_r)},$$

we verify that (22) is fulfilled thanks to the definition (21) of k_i . This concludes the induction. In particular, the support $\pi(\{1, \dots, s\})$ of \mathbf{x} is included in $S^{\bar{n}}$, where \bar{n} can be estimated (using similar arguments as before) by

$$\bar{n} = \sum_{i=1}^r k_i \leq c s, \quad c := \frac{\ln(16/\rho^2)}{\ln(1/\rho^2)}.$$

It follows from (HTP₂) that $\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\bar{n}}\|_2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 = \|\mathbf{e}\|_2$. In turn, we derive that

$$\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq \frac{1}{\sqrt{1 - \delta_{\bar{n}}}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^{\bar{n}})\|_2 \leq \frac{1}{\sqrt{1 - \delta_{cs}}} (\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\bar{n}}\|_2 + \|\mathbf{e}\|_2) \leq \frac{2}{\sqrt{1 - \delta_{cs}}} \|\mathbf{e}\|_2 =: d\|\mathbf{e}\|_2.$$

We note that $\delta_{3s} \leq 1/3$ yields $c \leq 3$, then $d \leq \sqrt{6}$, and finally $\gamma \geq (4 - \sqrt{2})\sqrt{3}/56$. \square

Remark 7. When the vector \mathbf{x} is not exactly s -sparse, we denote by S an index set of its s largest absolute entries. Writing $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ as $\mathbf{y} = \mathbf{A}\mathbf{x}_S + \mathbf{e}'$ with $\mathbf{e}' := \mathbf{A}\mathbf{x}_{\bar{S}} + \mathbf{e}$, Theorem 6 guarantees that, under the restricted isometry condition $\delta_{3s} \leq 1/3$,

$$\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq \|\mathbf{x}_{\bar{S}}\|_2 + \|\mathbf{x}_S - \mathbf{x}^{\bar{n}}\|_2 \leq \|\mathbf{x}_{\bar{S}}\|_2 + d\|\mathbf{e}'\|_2, \quad \bar{n} \leq c s,$$

provided $\|\mathbf{e}'\|_2 \leq \gamma x_s^*$. Let us now assume that $\mathbf{e} = \mathbf{0}$ for simplicity. With S_1, S_2, \dots denoting index sets of s largest absolute entries of $\mathbf{x}_{\bar{S}}$, of next s largest absolute entries of $\mathbf{x}_{\bar{S}}$, etc., we notice that

$$\|\mathbf{e}'\|_2 = \|\mathbf{A}\mathbf{x}_{\bar{S}}\|_2 \leq \sum_{k \geq 1} \|\mathbf{A}\mathbf{x}_{S_k}\|_2 \leq \sqrt{1 + \delta_s} \sum_{k \geq 1} \|\mathbf{x}_{S_k}\|_2.$$

In particular, vectors whose nonincreasing rearrangement decays as $x_j^* = \mu^j$ for $\mu \in (0, 1/2)$ satisfy $\|\mathbf{A}\mathbf{x}_{\bar{S}}\|_2 \leq c_{\delta_s} \mu^{s+1}$, so that $\|\mathbf{e}'\|_2 \leq \gamma x_s^*$ holds for small enough μ . Taking $\|\mathbf{x}_{\bar{S}}\|_2 \leq c' \mu^{s+1}$ into account, too, the reconstruction bound becomes $\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq c''_{\delta_s} \mu^{s+1}$. As a matter of fact, the same bound is valid if s is replaced by a smaller value t (even $t = 1$), so the reconstruction error becomes quite satisfactory after a relatively small number ct of iterations.

4 Uniform Recovery via Graded Hard Thresholding Pursuit

This section contains the main result about (GHTP) in the uniform setting, namely the fact that the reconstruction of all s -sparse vectors is achieved under a restricted isometry condition in a number of iterations at most proportional to s . This is comparable to the main result about (HTP), but we recall the added benefits that (GHTP) does not need an estimation of

the sparsity to run and that the stopping criterion involves an estimation of either one of the sparsity or the measurement error — the only other such algorithm we are aware of is (OMP). In passing, we note that (in both uniform and nonuniform settings) the number of (GHTP) iterations for s -sparse recovery cannot be less than s , since $t < s$ iterations only produce t -sparse vectors. We prove the result directly in the realistic situation and we invite the reader to weaken the restricted isometry condition in the idealized situation.

Theorem 8. If the restricted isometry constant of the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ obeys

$$\delta_{9s} \leq \frac{1}{3},$$

then the sequence (\mathbf{x}^n) produced by (GHTP) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{C}^m$ for some s -sparse $\mathbf{x} \in \mathbb{C}^N$ and some $\mathbf{e} \in \mathbb{C}^m$ with $\|\mathbf{e}\|_2 \leq \gamma x_s^*$ satisfies

$$\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq d\|\mathbf{e}\|_2, \quad \bar{n} \leq cs.$$

The constants $c \leq 4$, $d \leq 2.45$, and $\gamma \geq 0.079$ depend only on δ_{9s} .

Proof. Let π be the permutation of $\{1, 2, \dots, N\}$ for which $|x_{\pi(j)}| = x_j^*$ for all $j \in \{1, 2, \dots, N\}$. Our goal is to show that $\text{supp}(\mathbf{x}) = \pi(\{1, \dots, s\}) \subseteq S^{\bar{n}}$ for some $\bar{n} \leq 4s$. The arguments are very similar to the ones used in the proof of Theorem 5, except for the notable difference that the first s iterations are ignored. We still partition $\pi(\{1, \dots, s\})$ as $Q_1 \cup Q_2 \cup \dots \cup Q_r$, $r \leq s$, where the Q_i are defined as in (15), and we prove by induction on $i \in \{0, \dots, r\}$ that

$$(23) \quad Q_0 \cup Q_1 \cup \dots \cup Q_i \subseteq S^{s+k_0+k_1+\dots+k_i},$$

where $Q_0 = \emptyset$, $k_0 = 0$, and k_1, \dots, k_r are defined, with $\rho := \rho_{9s} \leq 1/2$, by

$$(24) \quad k_i := \left\lceil \frac{\ln(16(\text{card}(Q_i) + \text{card}(Q_{i+1})/2 + \dots + \text{card}(Q_r)/2^{r-i}))}{\ln(1/\rho^2)} \right\rceil.$$

Based on arguments already used, we observe that

$$\sum_{i=1}^r k_i \leq \frac{\ln(4/\rho^2)}{\ln(1/\rho^2)} s \leq 3s.$$

We now note that (23) holds trivially for $i = 0$. Next, if (23) holds for $i - 1$, $i \in \{1, \dots, r\}$, then Lemma 4 guarantees that (23) holds for i , provided

$$(25) \quad x_{q_i}^* > \rho^{k_i} \sqrt{\|\mathbf{x}_{Q_i}\|_2^2 + \|\mathbf{x}_{Q_{i+1}}\|_2^2 + \dots + \|\mathbf{x}_{Q_r}\|_2^2 + \kappa_{9s}\|\mathbf{e}\|_2},$$

since $\rho_{s+2(s+k_1+\dots+k_{i-1})+2k_i} \leq \rho_{9s}$. With $\gamma := (2\sqrt{2} - 1)/(4\kappa_{9s}) \geq (4 - \sqrt{2})\sqrt{3}/56$, we verify that (25) is fulfilled thanks to the definition (24) of k_i . This concludes the induction. In particular, the support $\pi(\{1, \dots, s\})$ of \mathbf{x} is included in $S^{\bar{n}}$ with $\bar{n} = s + k_1 + \dots + k_r \leq 4s$. We derive that $\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\bar{n}}\|_2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 = \|\mathbf{e}\|_2$ from (HTP₂), and then

$$\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq \frac{1}{\sqrt{1 - \delta_{\bar{n}}}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^{\bar{n}})\|_2 \leq \frac{1}{\sqrt{1 - \delta_{4s}}} (\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\bar{n}}\|_2 + \|\mathbf{e}\|_2) \leq \frac{2}{\sqrt{1 - \delta_{9s}}} \|\mathbf{e}\|_2,$$

which is the desired result with $d = 2/\sqrt{1 - \delta_{9s}} \leq \sqrt{6}$. \square

5 Nonuniform Recovery via Graded Hard Thresholding Pursuit

In this section, we demonstrate that s -sparse recovery via exactly s iterations of (GHTP) can be expected in the nonuniform setting, i.e., given an s -sparse vector $\mathbf{x} \in \mathbb{C}^N$, a random measurement matrix allows for the recovery of this specific \mathbf{x} with high probability. Precisely, we obtain results in two extreme cases: when the nonincreasing rearrangement of the sparse vector does not decay much and when it decays exponentially fast. Proposition 9 covers the first case, suspected to be the least favorable one — see Subsection 6.2. The result and its proof have similarities with the work of [13], which was extended to incorporate measurement error in [9]. The result is stated directly in this realistic situation, and it is worth pointing out its validity for all measurement errors $\mathbf{e} \in \mathbb{C}^m$ simultaneously.

Proposition 9. Let $\lambda \geq 1$ and let $\mathbf{x} \in \mathbb{C}^N$ be an s -sparse vector such that $x_1^* \leq \lambda x_s^*$. If $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a (normalized) Gaussian random matrix with

$$m \geq Cs \ln(N),$$

then it occurs with probability at least $1 - 2N^{-c}$ that, for all $\mathbf{e} \in \mathbb{C}^m$ with $\|\mathbf{e}\|_2 \leq \gamma x_s^*$, the sequences (S^n) and (\mathbf{x}^n) produced by (GHTP) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfy, at iteration s ,

$$S^s = \text{supp}(\mathbf{x}) \quad \text{and} \quad \|\mathbf{x} - \mathbf{x}^s\|_2 \leq d \|\mathbf{e}\|_2.$$

The constants γ and d depend only on λ , while the constant C depends on γ and c .

Proof. The result is in fact valid if (normalized) Gaussian random matrices are replaced by matrices with independent subgaussian entries with mean zero and variance $1/m$. One part of the argument relies on the fact that, for a fixed index set S of size s ,

$$(26) \quad \mathbb{P}(\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}\|_{2 \rightarrow 2} > \delta) \leq 2 \exp(-c' \delta^2 m)$$

provided $m \geq C's/\delta^2$, with c', C' depending only on the subgaussian distribution (see e.g. [7, Theorem 9.9], which leads to the improvement [7, Theorem 9.11] of [1, Theorem 5.2] in terms of dependence of constants on δ). Another part of the argument relies on the fact that, for a vector $\mathbf{v} \in \mathbb{C}^N$ and an index $\ell \in \{1, \dots, N\}$, with \mathbf{a}_ℓ denoting the ℓ th column of the matrix \mathbf{A} ,

$$(27) \quad \mathbb{P}(|\langle \mathbf{a}_\ell, \mathbf{v} \rangle| > t \|\mathbf{v}\|_2) \leq 4 \exp(-c'' t^2 m)$$

for a constant c'' depending only on the subgaussian distribution (see e.g. [7, Theorem 7.27] in the case $\mathbf{v} \in \mathbb{R}^N$). With $S := \text{supp}(\mathbf{x})$, let us now define, for $n \in \{1, \dots, s\}$, two random variables χ_n and ζ_n as

$$\begin{aligned} \chi_n &:= [(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_S]_n^*, \\ \zeta_n &:= [(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_{\bar{S}}]_1^*, \end{aligned}$$

i.e., χ_n is the n th largest value of $|(\mathbf{x}^{n-1} + \mathbf{A}^* \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}))_j|$ when j runs through S and ζ_n is the largest value of $|(\mathbf{x}^{n-1} + \mathbf{A}^* \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}))_\ell|$ when ℓ runs through \bar{S} . We are going to prove that, with high probability, $S^n \subseteq S$ for all $n \in \{1, \dots, s\}$, which follows from $\chi_n > \zeta_n$ for all $n \in \{1, \dots, s\}$. The failure probability of this event is

$$P := \mathbb{P}(\exists n \in \{1, \dots, s\} : \zeta_n \geq \chi_n \text{ and } (\chi_{n-1} > \zeta_{n-1}, \dots, \chi_1 > \zeta_1))$$

$$(28) \leq \mathbb{P}\left(\|\mathbf{A}_{S \cup \{\ell\}}^* \mathbf{A}_{S \cup \{\ell\}} - \mathbf{I}\|_{2 \rightarrow 2} > \delta \text{ for some } \ell \in \bar{S}\right)$$

$$(29) + \sum_{n=1}^s \mathbb{P}\left(\zeta_n \geq \chi_n, (\chi_{n-1} > \zeta_{n-1}, \dots, \chi_1 > \zeta_1), (\|\mathbf{A}_{S \cup \{\ell\}}^* \mathbf{A}_{S \cup \{\ell\}} - \mathbf{I}\|_{2 \rightarrow 2} \leq \delta \text{ for all } \ell \in \bar{S})\right),$$

where the constant $\delta = \delta_\lambda$ is to be chosen later. According to (26), the term in (28) is bounded by $2(N - s) \exp(-c'\delta^2 m)$ since a proper choice of C (depending on λ through δ) guarantees that $m \geq C'(s + 1)/\delta^2$. Let us turn to the term in (29) and let us use the shorthand $\mathbb{P}'(E)$ to denote the probability of an event E intersected with the events $(\chi_{n-1} > \zeta_{n-1}, \dots, \chi_1 > \zeta_1)$ and $(\|\mathbf{A}_{S \cup \{\ell\}}^* \mathbf{A}_{S \cup \{\ell\}} - \mathbf{I}\|_{2 \rightarrow 2} \leq \delta \text{ for all } \ell \in \bar{S})$, which are assumed to hold below. On the one hand, with $T^{s-n+1} \subseteq S$ representing an index set corresponding to $s - n + 1$ smallest values of $|(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_j|$ when j runs through S , we notice that

$$\begin{aligned} \chi_n &\geq \frac{1}{\sqrt{s-n+1}} \|(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_{T^{s-n+1}}\|_2 \\ &\geq \frac{1}{\sqrt{s-n+1}} (\|\mathbf{x}_{T^{s-n+1}}\|_2 - \|((\mathbf{A}^* \mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_{T^{s-n+1}}\|_2 - \|(\mathbf{A}^* \mathbf{e})_{T^{s-n+1}}\|_2). \end{aligned}$$

Then, since the condition $\chi_{n-1} > \zeta_{n-1}$ yields $S^{n-1} \subseteq S$, we obtain

$$\|((\mathbf{A}^* \mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_{T^{s-n+1}}\|_2 \leq \|(\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1})\|_2 \leq \delta \|\mathbf{x} - \mathbf{x}^{n-1}\|_2.$$

We also use the inequality $\|(\mathbf{A}^* \mathbf{e})_{T^{s-n+1}}\|_2 \leq \sqrt{1 + \delta} \|\mathbf{e}\|_2$ to derive

$$\chi_n \geq \frac{1}{\sqrt{s-n+1}} \left(\|\mathbf{x}_{T^{s-n+1}}\|_2 - \delta \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 - \sqrt{1 + \delta} \|\mathbf{e}\|_2 \right).$$

On the other hand, since $S^{n-1} \subseteq S$, we have

$$\begin{aligned} \zeta_n &= \max_{\ell \in \bar{S}} |(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_\ell| \leq \max_{\ell \in \bar{S}} |(\mathbf{A}^* \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}))_\ell| + \max_{\ell \in \bar{S}} |(\mathbf{A}^* \mathbf{e})_\ell| \\ &\leq \max_{\ell \in \bar{S}} |\langle \mathbf{a}_\ell, \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}) \rangle| + \sqrt{1 + \delta} \|\mathbf{e}\|_2. \end{aligned}$$

It follows that

$$\begin{aligned} &\mathbb{P}'(\zeta_n \geq \chi_n) \\ &\leq \mathbb{P}'\left(\max_{\ell \in \bar{S}} |\langle \mathbf{a}_\ell, \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}) \rangle| \geq \frac{1}{\sqrt{s-n+1}} (\|\mathbf{x}_{T^{s-n+1}}\|_2 - \delta \|\mathbf{x} - \mathbf{x}^{n-1}\|_2) - 2\sqrt{1 + \delta} \|\mathbf{e}\|_2\right) \\ &\leq \mathbb{P}'\left(\max_{\ell \in \bar{S}} |\langle \mathbf{a}_\ell, \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}) \rangle| \geq \frac{\delta}{\sqrt{s}} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2\right), \end{aligned}$$

where the last step will be justified by the inequality

$$\frac{1}{\sqrt{s-n+1}} \|\mathbf{x}_{T^{s-n+1}}\|_2 - 2\sqrt{1+\delta} \|\mathbf{e}\|_2 \geq \frac{2\delta}{\sqrt{s-n+1}} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2.$$

In order to fulfill this inequality, we remark that $\|\mathbf{x}_{T^{s-n+1}}\|_2 \geq \sqrt{s-n+1} x_s^*$, that $\|\mathbf{e}\|_2 \leq \gamma x_s^*$, and, thanks to the extension of (7) mentioned in Remark 1, that

$$\begin{aligned} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 &\leq \frac{1}{\sqrt{1-\delta^2}} \|\mathbf{x}_{S^{n-1}}\|_2 + \frac{1}{1-\delta} \|(\mathbf{A}^* \mathbf{e})_{S^{n-1}}\|_2 \leq \frac{\sqrt{s-n+1}}{\sqrt{1-\delta^2}} x_1^* + \frac{\sqrt{1+\delta}}{1-\delta} \|\mathbf{e}\|_2 \\ &\leq \sqrt{s-n+1} \left(\frac{\lambda}{\sqrt{1-\delta^2}} x_s^* + \frac{\sqrt{1+\delta} \gamma}{1-\delta} x_s^* \right), \end{aligned}$$

so that it is enough to choose δ then γ (depending on λ) small enough to have

$$1 - 2\sqrt{1+\delta} \gamma \geq 2\delta \left(\frac{\lambda}{\sqrt{1-\delta^2}} + \frac{\sqrt{1+\delta} \gamma}{1-\delta} \right).$$

Taking $\|\mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1})\|_2 \leq \sqrt{1+\delta} \|\mathbf{x} - \mathbf{x}^{n-1}\|_2$ into account, we obtain

$$(30) \quad \mathbb{P}'(\zeta_n \geq \chi_n) \leq \mathbb{P}' \left(\max_{\ell \in \bar{S}} |\langle \mathbf{a}_\ell, \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}) \rangle| \geq \frac{\delta}{\sqrt{1+\delta}\sqrt{s}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1})\|_2 \right).$$

Let us now remark that the condition $\chi_1 = [(\mathbf{A}^* \mathbf{A}_S \mathbf{x} + \mathbf{A}^* \mathbf{e})_{S^1}]_1^* > \zeta_1 = [(\mathbf{A}^* \mathbf{A}_S \mathbf{x} + \mathbf{A}^* \mathbf{e})_{\bar{S}^1}]_{\bar{S}^1}^*$ implies that the random set S^1 (corresponding to the largest value of $\mathbf{A}_S^* \mathbf{A}_S \mathbf{x} + \mathbf{A}_S \mathbf{e}$) depends only on the random submatrix \mathbf{A}_S , so that the random vector $\mathbf{x}^1 = \mathbf{A}_{S^1}^\dagger \mathbf{y}$ depends only on \mathbf{A}_S , too; then, since \mathbf{x}^1 is supported on $S^1 \subseteq S$, the condition $\chi_2 = [(\mathbf{x}^1 + \mathbf{A}^* \mathbf{A}_S(\mathbf{x} - \mathbf{x}^1) + \mathbf{A}^* \mathbf{e})_{S^2}]_2^* > \zeta_2 = [(\mathbf{x}^1 + \mathbf{A}^* \mathbf{A}_S(\mathbf{x} - \mathbf{x}^1) + \mathbf{A}^* \mathbf{e})_{\bar{S}^2}]_{\bar{S}^2}^*$ implies that the random set S^2 (corresponding to the two largest values of $\mathbf{x}_S^1 + \mathbf{A}_S^* \mathbf{A}_S(\mathbf{x} - \mathbf{x}^1) + \mathbf{A}_S^* \mathbf{e}$) depends only on the random submatrix \mathbf{A}_S , so that the random vector $\mathbf{x}^2 = \mathbf{A}_{S^2}^\dagger \mathbf{y}$ depends only on \mathbf{A}_S , too; and so on until we infer that the random vector \mathbf{x}^{n-1} , supported on $S^{n-1} \subseteq S$, depends only on \mathbf{A}_S , and in turn so does $\mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1})$. Exploiting the independence of $\mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1})$ and of the columns \mathbf{a}_ℓ for $\ell \in \bar{S}$, the estimate (27) can be applied in (30) to yield

$$\begin{aligned} \mathbb{P}'(\zeta_n \geq \chi_n) &\leq \sum_{\ell \in \bar{S}} \mathbb{P} \left(|\langle \mathbf{a}_\ell, \mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1}) \rangle| \geq \frac{\delta}{\sqrt{1+\delta}\sqrt{s}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^{n-1})\|_2 \right) \\ &\leq 4(N-s) \exp \left(-\frac{c'' \delta^2 m}{(1+\delta)s} \right). \end{aligned}$$

Altogether, the failure probability P satisfies

$$\begin{aligned} P &\leq 2(N-s) \exp(-c' \delta^2 m) + 4s(N-s) \exp \left(-\frac{c'' \delta^2 m}{(1+\delta)s} \right) \\ &\leq 2(N-1) \exp(-c' \delta^2 m) + N^2 \exp \left(-\frac{c'' \delta^2 m}{(1+\delta)s} \right). \end{aligned}$$

In particular, we have $\chi_s > \zeta_s$, and in turn $S^s = S$, with failure probability at most P . Then, since $\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}\|_{2 \rightarrow 2} \leq \delta$ with failure probability given in (26), we also have

$$\|\mathbf{x} - \mathbf{x}^s\|_2 \leq \frac{1}{\sqrt{1-\delta}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^s)\|_2 \leq \frac{1}{\sqrt{1-\delta}} (\|\mathbf{y} - \mathbf{A}\mathbf{x}^s\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2) \leq \frac{2}{\sqrt{1-\delta}} \|\mathbf{e}\|_2,$$

which is the desired result with $d = 2/\sqrt{1-\delta}$. The resulting failure probability is bounded by

$$2N \exp(-c' \delta^2 m) + N^2 \exp\left(-\frac{c'' \delta^2 m}{(1+\delta)s}\right) \leq 2N^2 \exp\left(-\frac{c''' m}{s}\right) \leq 2N^{-c},$$

where the last inequality holds with a proper choice of the constant C in $m \geq Cs \ln(N)$. \square

The previous result is limited by its nonapplicability to random partial Fourier matrices. For such matrices, the uniform results of the two previous sections — or any recovery result based on the restricted isometry property — do apply, but the restricted isometry property is only known to be fulfilled with a number of measurements satisfying $m \geq cs \ln^4(N)$. This threshold can be lowered in the nonuniform setting by reducing the power of the logarithm. This is typically achieved via ℓ_1 -minimization (see [4, Theorem 1.3], [3, Theorem 1.1], or [8, V³]), for which the successful recovery of a vector does not depend on its shape but only on its sign pattern. We establish below a comparable result for (GHTP): fewer measurements guarantee the high probability of nonuniform s -sparse recovery via exactly s iterations of (GHTP). There is an additional proviso on the vector shape, namely its nonincreasing rearrangement should decay exponentially fast. This is admittedly a strong assumption, but possible improvements in the proof strategy (perhaps incorporating ideas from Proposition 9) suggest that the result could be extended beyond this class of sparse vectors for Fourier measurements. Incidentally, for this class of vectors, it is striking that the number of Gaussian measurements can be reduced even below the classical reduction (see [5, Theorems 1.2 and 1.3]) of the constant c in $m \approx cs \ln(N/s)$ when passing from the uniform to the nonuniform setting.

Proposition 10. Let $\mu \in (0, 1)$ and let $\mathbf{x} \in \mathbb{C}^N$ be an s -sparse vector such that $x_{j+1}^* \leq \mu x_j^*$ for all $j \in \{1, \dots, s-1\}$. If $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a (normalized) Gaussian random matrix with

$$m \geq \max\{Cs, C' \ln(N)\},$$

or if $\mathbf{A} \in \mathbb{C}^{m \times N}$ is a (normalized) random partial Fourier matrix with

$$m \geq C'' s \ln(N),$$

then it occurs with probability at least $1 - 2N^{-c}$ that, for all $\mathbf{e} \in \mathbb{C}^m$ with $\|\mathbf{e}\|_2 \leq \gamma x_s^*$, the sequences (S^n) and (\mathbf{x}^n) produced by (GHTP) with $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfy, at iteration s ,

$$S^s = \text{supp}(\mathbf{x}) \quad \text{and} \quad \|\mathbf{x} - \mathbf{x}^s\|_2 \leq d \|\mathbf{e}\|_2.$$

The constants C , γ , and d depend only on μ , while the constants C' and C'' depend on μ and c .

Proof. The result would also be valid if (normalized) Gaussian random matrices were replaced by (normalized) subgaussian random matrices and if (normalized) random partial Fourier matrices were replaced by (normalized) random sampling matrices associated to bounded orthonormal systems. We recall (see [7, chapter 12] for details) that a bounded orthonormal system is a system (ϕ_1, \dots, ϕ_N) of functions, defined on a set \mathcal{D} endowed with a probability measure ν , that satisfy, for some constant $K \geq 1$,

$$(31) \quad \sup_{t \in \mathcal{D}} |\phi_j(t)| \leq K \quad \text{and} \quad \int_{\mathcal{D}} \phi_k(t) \overline{\phi_\ell(t)} d\nu(t) = \delta_{k,\ell}$$

for all $j, k, \ell \in \{1, \dots, N\}$. The normalized random sampling matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ associated to (ϕ_1, \dots, ϕ_N) has entries

$$A_{k,\ell} = \frac{1}{\sqrt{m}} \phi_\ell(t_k), \quad k \in \{1, \dots, m\}, \quad \ell \in \{1, \dots, N\},$$

where the sampling points $t_1, \dots, t_m \in \mathcal{D}$ are selected independently at random according to the probability measure ν . For instance, the random partial Fourier matrix corresponds to the choices $\mathcal{D} = \{1, \dots, N\}$, $\nu(T) = \text{card}(T)/N$ for all $T \subseteq \{1, \dots, N\}$, and $\phi_k(t) = \exp(i2\pi kt)$, in which case $K = 1$. The argument here relies on the fact (see [7, Theorem 12.12]) that, for a fixed index set S of size s ,

$$(32) \quad \mathbb{P}(\|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}\|_{2 \rightarrow 2} > \delta) \leq 2s \exp\left(-\frac{3\delta^2 m}{8K^2 s}\right).$$

Let π be the permutation of $\{1, 2, \dots, N\}$ for which $|x_{\pi(j)}| = x_j^*$ for all $j \in \{1, 2, \dots, N\}$ and let Π_n denote the set $\pi(\{1, 2, \dots, n\})$ for each $n \in \{1, \dots, N\}$. Let $\delta = \delta_\mu > 0$ be chosen small enough to have $2\delta/\sqrt{1-\delta^2} \leq (1-\mu)\sqrt{1-\mu^2}/3$ — for reason that will become apparent later. Invoking (26) and (32) for the sets $\Pi_s \cup \{k\}$, $k \notin \Pi_s$, ensures that

$$(33) \quad \|\mathbf{A}_{\Pi_s \cup \{k\}}^* \mathbf{A}_{\Pi_s \cup \{k\}} - \mathbf{I}\|_{2 \rightarrow 2} \leq \delta.$$

In the Gaussian case, provided $m \geq C_\mu s$, this holds with failure probability at most

$$2(N-s) \exp(-c_\mu m) \leq 2N \exp(-c_\mu C'_{\mu,c} \ln(N)) \leq 2N^{-c}$$

where $C'_{\mu,c}$ is chosen large enough. In the Fourier case, this holds with failure probability at most

$$2(N-s)s \exp\left(-\frac{c_\mu m}{s}\right) \leq 2N^2 \exp(-c_\mu C''_{\mu,c} \ln(N)) \leq 2N^{-c}$$

where $C''_{\mu,c}$ is chosen large enough. We are going to prove by induction that $S^n = \Pi_n$ for all $n \in \{0, 1, \dots, s\}$. The result holds trivially for $n = 0$. To increment the induction from $n-1$ to n , $n \in \{1, \dots, s\}$, we need to show that

$$\min_{j=1, \dots, n} |(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_{\pi(j)}| > \max_{\ell=n+1, \dots, N} |(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_{\pi(\ell)}|.$$

We observe that for $j \in \{1, \dots, n\}$ and $\ell \in \{n+1, \dots, N\}$,

$$\begin{aligned} |(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_{\pi(j)}| &\geq |x_{\pi(j)}| - |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}) + \mathbf{A}^*\mathbf{e})_{\pi(j)}| \\ &\geq x_n^* - |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_{\pi(j)}| - |(\mathbf{A}^*\mathbf{e})_{\pi(j)}|, \\ |(\mathbf{x}^{n-1} + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^{n-1}))_{\pi(\ell)}| &\leq |x_{\pi(\ell)}| + |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}) + \mathbf{A}^*\mathbf{e})_{\pi(\ell)}| \\ &\leq x_{n+1}^* + |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_{\pi(\ell)}| + |(\mathbf{A}^*\mathbf{e})_{\pi(\ell)}|. \end{aligned}$$

Hence, it is enough to prove that, for all $j \in \{1, \dots, n\}$ and $\ell \in \{n+1, \dots, N\}$,

$$(34) \quad x_n^* > x_{n+1}^* + |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_{\pi(j)}| + |(\mathbf{A}^*\mathbf{e})_{\pi(j)}| + |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_{\pi(\ell)}| + |(\mathbf{A}^*\mathbf{e})_{\pi(\ell)}|.$$

For any $k \in \{1, \dots, N\}$, using the the induction hypothesis and (33), we obtain

$$(35) \quad |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_k| \leq \|(\mathbf{A}_{\Pi_s \cup \{k}}^* \mathbf{A}_{\Pi_s \cup \{k}} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1})\|_2 \leq \delta \|\mathbf{x} - \mathbf{x}^{n-1}\|_2.$$

Moreover, in view of the extension of (7) mentioned in Remark 1, we have

$$(36) \quad \|\mathbf{x} - \mathbf{x}^{n-1}\|_2 \leq \frac{1}{\sqrt{1-\delta^2}} \|\mathbf{x}_{\Pi_{n-1}}\|_2 + \frac{1}{1-\delta} \|(\mathbf{A}^*\mathbf{e})_{\Pi_{n-1}}\|_2.$$

Combining (35) and (36) gives, for any $k \in \{1, \dots, N\}$,

$$\begin{aligned} |((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_k| + |(\mathbf{A}^*\mathbf{e})_k| &\leq \frac{\delta}{\sqrt{1-\delta^2}} \|\mathbf{x}_{\Pi_{n-1}}\|_2 + \frac{\delta}{1-\delta} \|(\mathbf{A}^*\mathbf{e})_{\Pi_{n-1}}\|_2 + |(\mathbf{A}^*\mathbf{e})_k| \\ &\leq \frac{\delta}{\sqrt{1-\delta^2}\sqrt{1-\mu^2}} x_n^* + \frac{1}{1-\delta} \|(\mathbf{A}^*\mathbf{e})_{\Pi_s \cup \{k}\}\|_2, \end{aligned}$$

where we have used $\|\mathbf{x}_{\Pi_{n-1}}\|_2^2 \leq (x_n^*)^2(1 + \mu^2 + \dots + (\mu^2)^{s-n}) \leq (x_n^*)^2/(1 - \mu^2)$ in the last step. Using also the estimate imposed by (33) for the maximal singular value of $\mathbf{A}_{\Pi_s \cup \{k}\}$ in

$$\|(\mathbf{A}^*\mathbf{e})_{\Pi_s \cup \{k}\}\|_2 \leq \sqrt{1+\delta} \|\mathbf{e}\|_2 \leq \sqrt{1+\delta} \gamma x_s^* \leq \sqrt{1+\delta} \gamma x_n^*,$$

we arrive at

$$|((\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^{n-1}))_k| + |(\mathbf{A}^*\mathbf{e})_k| \leq \left(\frac{\delta}{\sqrt{1-\delta^2}\sqrt{1-\mu^2}} + \frac{\sqrt{1+\delta}\gamma}{1-\delta} \right) x_n^*.$$

Taking the fact that $x_{n+1}^* \leq \mu x_n^*$ into account, we observe that (34) is fulfilled as soon as

$$1 > \mu + \frac{2\delta}{\sqrt{1-\delta^2}\sqrt{1-\mu^2}} + \frac{2\sqrt{1+\delta}\gamma}{1-\delta}.$$

The latter holds with the choice of δ made at the beginning and then with an appropriate choice of γ (depending only on μ). At this point, we have justified inductively that $S^n = \Pi_n$ for all $n \in \{1, \dots, s\}$, hence in particular that $S^s = \Pi_s = \text{supp}(\mathbf{x})$. To conclude the proof, we remark that (GHTP₂) yields $\|\mathbf{y} - \mathbf{A}\mathbf{x}^s\|_2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 = \|\mathbf{e}\|_2$ and we also use the estimate imposed by (33) for the minimal singular value of \mathbf{A}_{Π_s} to derive

$$\|\mathbf{x} - \mathbf{x}^s\|_2 \leq \frac{1}{\sqrt{1-\delta}} \|\mathbf{A}(\mathbf{x} - \mathbf{x}^s)\|_2 \leq \frac{1}{\sqrt{1-\delta}} (\|\mathbf{y} - \mathbf{A}\mathbf{x}^s\|_2 + \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2) \leq \frac{2}{\sqrt{1-\delta}} \|\mathbf{e}\|_2.$$

The constant $d := 2/\sqrt{1-\delta}$ depends only on μ . □

6 Numerical experiments

This section provides some empirical validation of our theoretical results. We focus on the performances of (HTP), (GHTP), and its natural companion (OMP). All the tests were carried out for Gaussian random matrices $\mathbf{A} \in \mathbb{R}^{m \times N}$ with $m = 200$ and $N = 1000$. We generated 100 such random matrices. To assess the influence of the decay in the nonincreasing rearrangement of s -sparse vectors $\mathbf{x} \in \mathbb{R}^N$ to be recovered, we tested ‘flat’ vectors with $x_j^* = 1$ for $j \in \{1, \dots, s\}$, ‘linear’ vectors with $x_j^* = (s + 1 - j)/s$ for $j \in \{1, \dots, s\}$, and Gaussian vectors whose s nonzero entries are independent standard normal random variables. We generated 10 such vectors per matrix. All these experiments can be reproduced by downloading the associated MATLAB files from the authors’ webpages.

6.1 Frequency of successful recoveries

We present here a comparison between (HTP), (GHTP), and (OMP) in terms of successful recovery — this complements the experiments of [6] where (HTP) was compared to basis pursuit, compressive sampling matching pursuit, iterative hard thresholding, and normalized iterative hard thresholding. The natural stopping criterion $S^n = S^{n-1}$ was used for (HTP), and we used the criterion⁴ [$\text{supp}(\mathbf{x}) \subseteq S^n$ or $\|\mathbf{x} - \mathbf{x}^n\|_2 / \|\mathbf{x}\|_2 < 10^{-4}$] for (GHTP) and (OMP).

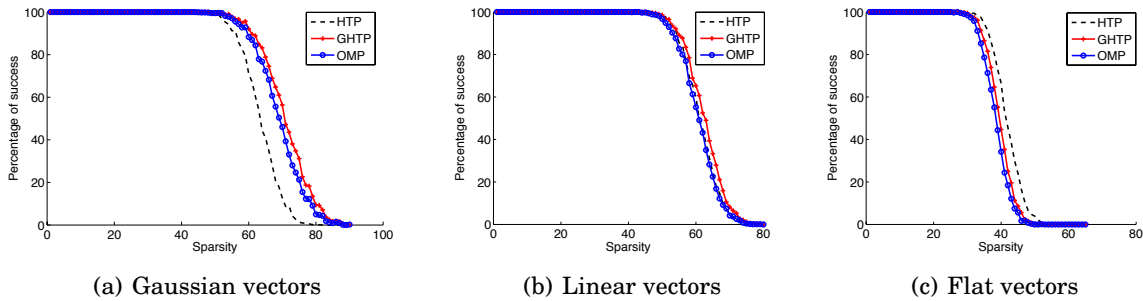


Figure 1: Frequency of success for (HTP), (GHTP), and (OMP) using Gaussian measurements

Figure 1 suggests that no algorithm consistently outperforms the other ones, although (GHTP) is always marginally better than (OMP). The shape of the vector to be recovered has a strong influence on the success rate, in contrast with basis pursuit whose success depend only on the sign pattern.

⁴the true sparse solution is available in this experiment, unlike in a practical scenario

6.2 Number of iterations and indices correctly captured

We now focus on the information provided by our experiments in terms of number of iterations needed for the recovery of a sparse vector and in terms of number of indices from the true support correctly captured in the active sets as a function of the iteration count. First, Figure 2 displays the maximum over all our trials for the number $\bar{n}(s)$ of iterations needed for s -sparse recovery as a function of s in the sparsity regions where recovery is certain (see Figure 1). With this indicator, (HTP) is the best algorithm — thanks to the fact that a prior knowledge of s is exploited. For this algorithm, no notable difference seems to result from the vector shape (but remember that it did influence the success rate), so the upper bound $\bar{n}(s) \leq cs$ from Theorem 5 is likely to be an overestimation to be replaced by the bound $\bar{n}(s) \leq c \ln(s)$, which is known to be valid for ‘flat’ vectors. As for (GHTP) and (OMP), we observe that the vector shape influences the regions (predicted by Propositions 9 and 10) of s -values for which $\bar{n}(s) = s$ and that (GHTP) generally requires fewer iterations than (OMP), probably thanks to the fact that incorrectly selected indices can be rejected at the next (GHTP) iteration (however, this complicates speed-ups based on QR -updates).

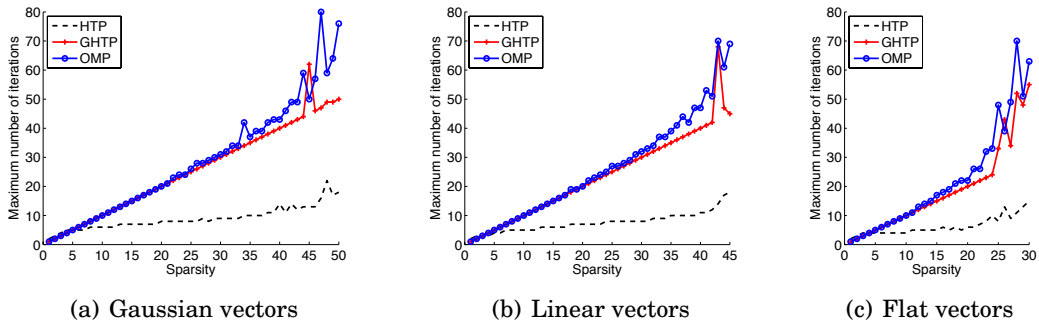
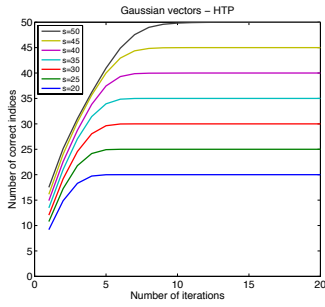
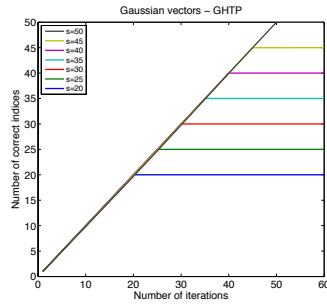


Figure 2: Number of iterations for (HTP), (GHTP), and (OMP) using Gaussian measurements

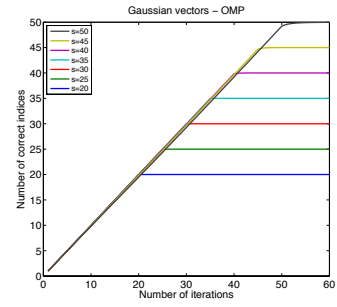
Next, Figures 3, 4, and 5 display the number $i(n)$ of indices from the true support that are correctly captured in the active set S^n as a function of the iteration count n . We point out that these figures do not reflect the worst case anymore, but instead they present an average over all successful recoveries. Once again, (HTP) is the superior algorithm because several correct indices can be captured at once. For (GHTP) and (OMP), the ideal situation corresponds to the selection of correct indices only, i.e., $i(n) = n$ for all $n \leq s$. This seems to occur for small sparsities, with larger regions of validity for vectors whose nonincreasing rearrangements decay quickly. This is an encouraging fact since $i(n) = n$ has been justified for all $n \leq s$ in the proof of Proposition 9 for the least favorable case of ‘flat’ vectors. For these vectors, Figure 5 also reveals that, even in the region where s -sparse recovery via (GHTP) and (OMP) is certain, the recovery is sometimes achieved in more than s iterations.



(a) Recovery via (HTP)

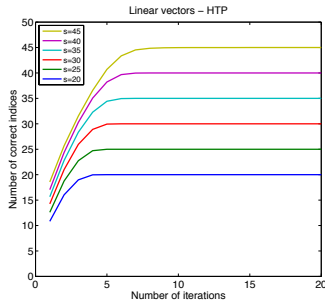


(b) Recovery via (GHTP)

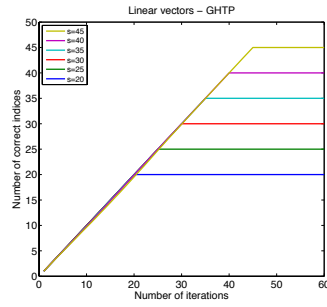


(c) Recovery via (OMP)

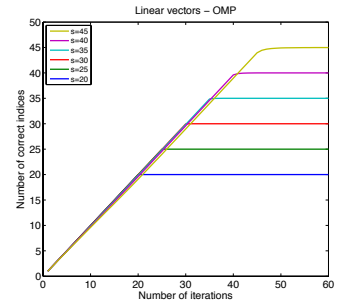
Figure 3: Number of support indices of Gaussian vectors correctly captured by the active set as a function of the iteration count for (HTP), (GHTP), and (OMP) using Gaussian measurements



(a) Recovery via (HTP)

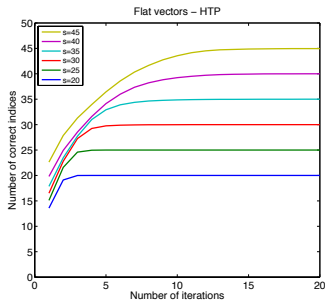


(b) Recovery via (GHTP)

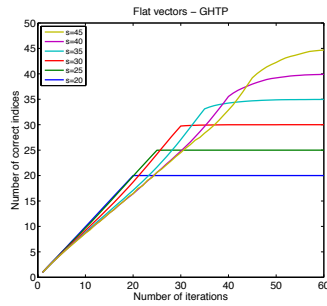


(c) Recovery via (OMP)

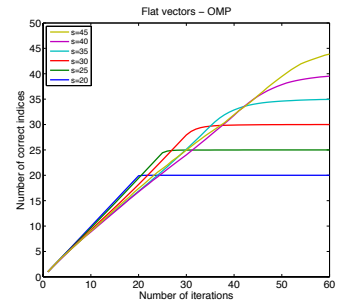
Figure 4: Number of support indices of linear vectors correctly captured by the active set as a function of the iteration count for (HTP), (GHTP), and (OMP) using Gaussian measurements



(a) Recovery via (HTP)



(b) Recovery via (GHTP)



(c) Recovery via (OMP)

Figure 5: Number of support indices of flat vectors correctly captured by the active set as a function of the iteration count for (HTP), (GHTP), and (OMP) using Gaussian measurements

6.3 Robustness to measurement error

Our final experiment highlights the effect of the measurement error $\mathbf{e} \in \mathbb{C}^m$ on the recovery via (HTP), (GHTP), and (OMP) of sparse vectors $\mathbf{x} \in \mathbb{C}^N$ acquired through $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{C}^m$. For a sparsity set to 25, with other parameters still fixed at $m = 200$ and $N = 1000$, the test was carried out on ‘linear’ and ‘flat’ s -sparse vectors. For an error level η varying from 0 to 5 by increment of 0.01, we generated 100 Gaussian matrices \mathbf{A} , then 10 instances of a random support S and a Gaussian noise vector \mathbf{e} normalized by $\|\mathbf{e}\|_2 = \eta x_s^*$, and we run the three algorithms on the resulting 1000 measurement vectors per error level. After the algorithms exited, we averaged the number of indices from the true support correctly caught, as well as the reconstruction error measured in ℓ_2 -norm. Under restricted isometry conditions, the proofs of Theorems 6 and 8 guarantee that all indices are correctly caught provided η is below some threshold η_* . This phenomenon is apparent in Figure 6 (abscissa: $\eta = \|\mathbf{e}\|_2/x_s^*$, ordinate: $\text{card}(S \cap S^\sharp)$, where S^\sharp is the support of the output \mathbf{x}^\sharp of the algorithm). The figure also shows that the threshold is approximately the same for the three algorithms and that it is higher for ‘flat’ vectors. For ‘linear’ vectors, the value $\eta_* \approx 0.1$ does not differ too much from the value $\gamma \approx 0.079$ obtained in Theorems 6 and 8. Once the threshold is passed, we also notice that (HTP) catches fewer correct indices than (GHTP) and (OMP), probably due to the fact that it produces index sets of smaller size.

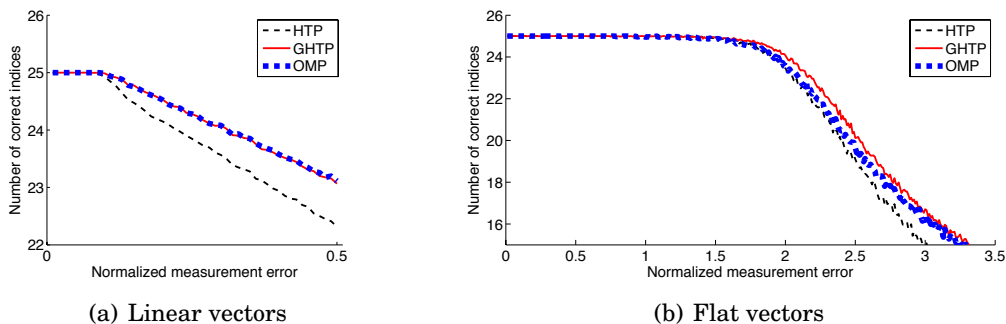


Figure 6: Number of indices correctly captured as a function of the measurement error

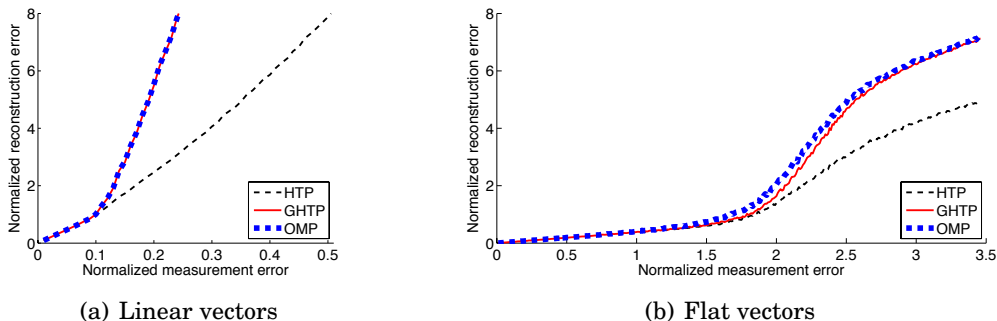


Figure 7: Reconstruction error as a function of the measurement error

In terms of reconstruction error, Figure 7 (abscissa: $\eta = \|\mathbf{e}\|_2/x_s^*$, ordinate: $\|\mathbf{x} - \mathbf{x}^\sharp\|_2/x_s^*$) displays transitions around the same values η_* as Figure 6. Although the ratio $\|\mathbf{x} - \mathbf{x}^\sharp\|_2/\|\mathbf{e}\|_2$ of reconstruction error by measurement error increases after the threshold, it does not rule out a robustness estimate of the type $\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq d\|\mathbf{e}\|_2$ valid for all $\mathbf{e} \in \mathbb{C}^m$ with an absolute constant d (this is established in [6] for (HTP) under a restricted isometry constant), since the prospective estimate is intuitively easier to satisfy when $\|\mathbf{e}\|_2$ becomes larger. We finally notice that (HTP) delivers lower reconstruction errors than (GHTP) and (OMP), probably due to the smaller index set involved in the final orthogonal projection.

References

- [1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3), 253–263, 2008.
- [2] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.
- [3] E. J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *IEEE Transactions on Information Theory*, 57(11):7235–7254, 2011.
- [4] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2): 489–509, 2006.
- [5] D. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of the American Mathematical Society*, 22(1):1–53, 2009.
- [6] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM Journal on Numerical Analysis*, 49(6):2543–2563, 2011.
- [7] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser. In press.
- [8] J.-P. Kahane. Variantes sur un théorème de Candès, Romberg et Tao. *Annales de l’Institut Fourier*. To appear.
- [9] J. Lin and S. Li. Nonuniform support recovery from noisy random measurements by orthogonal matching pursuit. *Journal of Approximation Theory*, 165:20–40, 2013.
- [10] A. Maleki. Coherence analysis of iterative thresholding algorithms. In *Proc. of 47th Annual Allerton Conference on Communication, Control, and Computing*, 236–243, 2009.

- [11] D. Needell. Noisy signal recovery via iterative reweighted L1-minimization. In Proc. Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA Nov. 2009.
- [12] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.
- [13] J. A. Tropp and A. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inform. Theory*, 53:4655–4666, 2007.
- [14] P. Wojtaszczyk. Stability and instance optimality for Gaussian measurements in compressed sensing. *Foundations of Computational Mathematics*, 10(1): 1–13, 2010.
- [15] T. Zhang. Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Trans. Inform. Theory*, 57(9):6215–6221, 2011.