Sparse disjointed recovery from noninflating measurements

Simon Foucart[g,*], Michael F. Minner[d], Tom Needham[g]
[g]Department of Mathematics, University of Georgia
[d]Department of Mathematics, Drexel University

Abstract

We investigate the minimal number of linear measurements needed to recover sparse disjointed vectors robustly in the presence of measurement error. First, we analyze an iterative hard thresholding algorithm relying on a dynamic program computing sparse disjointed projections to upper-bound the order of the minimal number of measurements. Next, we show that this order cannot be reduced by any robust algorithm handling noninflating measurements. As a consequence, we conclude that there is no benefit in knowing the simultaneity of sparsity and disjointedness over knowing only one of these structures.

Key words and phrases: sparse vectors, disjointed vectors, simultaneously structured models, dynamic programming, compressive sensing, subgaussian matrices, restricted isometry property, iterative hard thresholding.

1 Introduction and Main Result

In this note, we examine the recovery of sparse disjointed vectors $\mathbf{x} \in \mathbb{C}^N$ from linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^m$ with $m \ll N$. We recall that a vector $\mathbf{x} \in \mathbb{C}^N$ is said to be s-sparse if it has no more than s nonzero entries, i.e., if $\mathrm{card}(\mathrm{supp}(\mathbf{x})) \leq s$, where $\mathrm{supp}(\mathbf{x}) := \{i \in [\![1 : N]\!] : x_i \neq 0\}$. It is said to be d-disjointed if there are always at least d zero entries between two nonzero entries, i.e., if $|j - i| > d$ for all distinct $i, j \in \mathrm{supp}(\mathbf{x})$. We investigate here vectors that are simultaneously s-sparse and d-disjointed. This investigation was prompted by grid discretizations in MIMO radar problems [9]: the nonzero entries represent the positions of airplanes in an observation frame, so it is natural to assume that their number is low and that they are not too close to one another. Sparse disjointed vectors also serve as a pertinent model for neural spike trains, see [8] which already established recovery results akin to those presented in Section 3. In this note, however, we emphasize the question of the *minimal* number of measurements needed for robust uniform recovery of sparse disjointed vectors. We provide a complete answer with regard to noninflating measurements relative to this model (see Section 4 for the explanation of this terminology). As a reminder, the uniform recovery of s-sparse vectors is achievable from

$$(1) \qquad m \asymp m_{\mathsf{spa}} := s \ln\left(e\frac{N}{s}\right)$$

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

random linear measurements. It can be carried out efficiently using convex optimization or iterative greedy algorithms. The recovery is robust with respect to measurement error and stable with respect to sparsity defect. The number of measurements in (1) is optimal when stability is required. As for the uniform recovery of d-disjointed vectors, it is achievable from

$$m \asymp m_{\mathsf{dis}} := \frac{N}{d} \tag{2}$$

deterministic Fourier measurements and it can be carried out efficiently using convex optimization (see [4, Corollary 1.4]). The number of measurements (2) is easily seen to be optimal, even without requiring stability. Concerning simultaneously sparse and disjointed vectors, our main result is informally stated below.

Theorem 1. The minimal number of noninflating measurements needed to achieve robust uniform recovery of s-sparse d-disjointed vectors is of the order of

$$m_{\mathsf{spa\&dis}} := s \ln \left(e \frac{N - d(s-1)}{s} \right). \tag{3}$$

The significance of this result lies in its interpretation: for $m_{\mathsf{spa\&dis}}$ to be of smaller order than m_{spa}, we need $t := (N - d(s-1))/s \le N/(2s)$; but then $d = (N - st)/(s-1) \ge (N - N/2)/(s-1) \ge N/(2s)$, i.e., $N/d \le 2s$, which implies that m_{dis} is of smaller order than $m_{\mathsf{spa\&dis}}$. In short, we arrive at

$$m_{\mathsf{spa\&dis}} \asymp \min \{ m_{\mathsf{spa}}, m_{\mathsf{dis}} \}. \tag{4}$$

Expressed differently, there is no benefit in knowing the simultaneity of sparsity and disjointness as far as the number of noninflating measurements is concerned. This echoes the message of [10], which showed that vectors possessing certain structures simultaneously require at least as many Gaussian random measurements for their recovery via combined convex relaxations as what could have been achieved via the convex relaxation associated to one of the structures. Our result is narrower since it focuses on a particular simultaneity of structures, but no limitation is placed on the nature of the recovery algorithm and the measurements are only assumed to be noninflating instead of Gaussian. Note that restricting to ℓ_1-minimization and Gaussian measurements would have been irrelevant here, because even nonuniform recovery, i.e., the recovery of a single sparse vector—a fortiori of a disjointed one—already requires a number of measurements of order at least m_{spa}, as inferred from known results on phase transition (see [5] for the original arguments and [1] for recent arguments).

The rest of this note is organized as follows. In Section 2, we discuss basic facts about sparse disjointed vectors. In particular, we reveal how projections onto the set of sparse disjointed vectors can be computed by dynamic programming. The ability to compute these projections would allow for the modification of virtually all sparse recovery iterative greedy algorithms to fit the sparse disjointed framework, but we focus only on iterative hard thresholding (IHT)—arguably the simplest of these algorithms—in Section 3. There, we give a short justification that robust uniform recovery can be carried out efficiently based on random measurements (which are noninflating) provided their number has order at least $m_{\mathsf{spa\&dis}}$. Finally, Section 4 contains the crucial result that robust uniform recovery schemes for sparse disjointed vectors cannot exist if the number of noninflating measurements has order less than $m_{\mathsf{spa\&dis}}$.

2 Preliminary Considerations on Sparse Disjointed Vectors

We observe first that sparsity and disjointness are not totally independent structures, since a highly disjointed vector is automatically quite sparse when its length N is fixed. Viewed differently, an exactly s-sparse vector that is d-disjointed cannot have too small a length. Precisely, there holds $N \geq s + d(s-1)$, because there must be s nonzero entries and at least d zero entries in each of the $s-1$ spaces between them. In connection with (3), we rephrase this inequality as $N - d(s-1) \geq s$ to highlight that the logarithmic factor is at least equal to 1. Figure 1 depicts an alternative way to think of an s-sparse d-disjointed vector that is sometimes useful. Namely, we artificially insert d zero entries at the right end, hence forming a vector of length $N + d$ containing s blocks of size $d + 1$. Each block consists of one nonzero entry followed by d zero entries. Then, identifying every block with a collapsed object of length 1 reveals a one-to-one correspondence between s-sparse d-disjointed vectors of length N and s-sparse vectors of length $N - d(s-1)$. This correspondence will be used again in Section 4, but for now it immediately explains the following fact already exploited in [8].

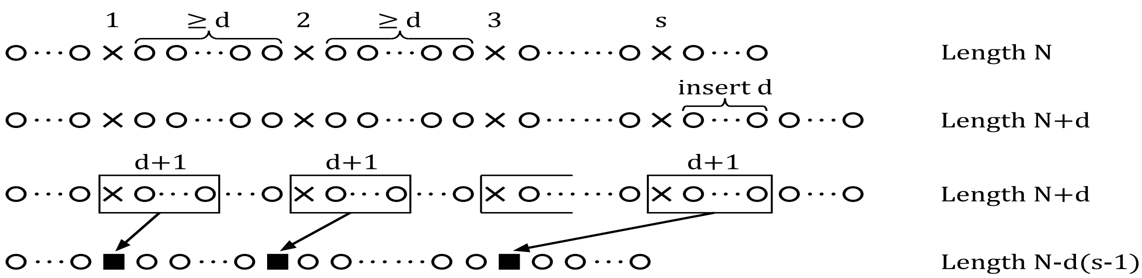

Figure 1: Pictorial representation of sparse disjointed vectors (hollow circles represent zero values).

Fact 2. The number of d-disjointed subsets of $[\![1:N]\!]$ with size s is $\binom{N-d(s-1)}{s}$.

This formula could instead be justified by the inductive process at the basis of Fact 3 below, which concerns the computation of best approximations by s-sparse d-disjointed vectors in ℓ_p for $p \in (0, \infty)$ (one talks about projections if $p = 2$). This task is not immediate, unlike the computation of the best approximations by s-sparse vectors (called hard thresholding), which simply consists of keeping the s largest absolute entries and setting the other entries to zero. Perhaps counterintuitively, the largest absolute entry need not be part of the support of the best approximation in the sparse disjointed case[1], as illustrated on Figure 2 by the example $\mathbf{x} = (1, 0, 1, 2^{1/4}, 1, 0, 2^{-1/2})$ whose best 3-sparse 1-separated approximation is $(1, 0, 1, 0, 1, 0, 0)$ for $p = 2$. Note that the best approximation is $(1, 0, 0, 2^{1/4}, 0, 0, 2^{-1/2})$ for $p = 4$, highlighting another difference with the sparse case, namely the possible dependence on p of the best approximation. The strategy adopted in [8] for computing best sparse disjointed approximations consisted in recasting the problem as an integer program and

[1] in particular, the nested approximation property of [2] does not hold in this case

relaxing it to a linear program proved to yield the same solution. We propose a different approach here. The corresponding MATLAB implementation is accessible on the first author's webpage as part of the reproducible accompanying this note.

Fact 3. The best ℓ_p-approximation of a vector $\mathbf{x} \in \mathbb{C}^N$ from the set of s-sparse d-disjointed vectors can be efficiently computed by dynamic programming for any $p \in (0, \infty)$.

The program determines $F(N,s)$, where $F(n,r)$ is defined for $n \in [\![1:N]\!]$ and $r \in [\![0:s]\!]$ by

$$F(n,r) := \min \left\{ \sum_{j=1}^{n} |x_j - z_j|^p : \mathbf{z} \in \mathbb{C}^n \text{ is } r\text{-sparse } d\text{-disjointed} \right\}. \tag{5}$$

We claim that, for $n \in [\![d+2, N]\!]$ and $r \in [\![1, s]\!]$,

$$F(n,r) = \min \begin{cases} F(n-1,r) + |x_n|^p, \\ F(n-d-1, r-1) + \sum\limits_{j=n-d}^{n-1} |x_j|^p. \end{cases} \tag{6}$$

This relation may be considered straightforward, as it simply distinguishes between a zero and a nonzero value at the last entry of the minimizer for $F(n,r)$. To be more rigorous, we establish first the lower estimate on $F(n,r)$ by considering a minimizer $\widehat{\mathbf{x}} \in \mathbb{C}^n$ for $F(n,r)$: if $\widehat{x}_n = 0$, then $F(n,r) = \sum_{j=1}^{n-1} |x_j - \widehat{x}_j|^p + |x_n|^p \geq F(n-1,r) + |x_n|^p$ since $\widehat{\mathbf{x}}_{[\![1:n-1]\!]} \in \mathbb{C}^{n-1}$ is r-sparse d-disjointed; if $\widehat{x}_n \neq 0$, so that $\widehat{x}_{n-d} = \cdots = \widehat{x}_{n-1} = 0$ by d-disjointedness, then $F(n,r) = \sum_{j=1}^{n-d-1} |x_j - \widehat{x}_j|^p + \sum_{j=n-d}^{n-1} |x_j|^p + |x_n - \widehat{x}_n|^p \geq F(n-d-1, r-1) + \sum_{j=n-d}^{n-1} |x_j|^p$ since $\widehat{\mathbf{x}}_{[\![1:n-d-1]\!]} \in \mathbb{C}^{n-d-1}$ is $(r-1)$-sparse d-disjointed. Second, we establish the upper estimate on $F(n,r)$ by separating cases for the minimum in (6): if $F(n-1,r) + |x_n|^p$ is the smallest value, selecting a minimizer $\widetilde{\mathbf{x}} \in \mathbb{C}^{n-1}$ for $F(n-1,r)$ and considering the r-sparse d-disjointed vector $\widehat{\mathbf{x}} := (\widetilde{\mathbf{x}}, 0) \in \mathbb{C}^n$ yields $F(n,r) \leq \sum_{j=1}^{n} |x_j - \widehat{x}_j|^p = \sum_{j=1}^{n-1} |x_j - \widetilde{x}_j|^p + |x_n|^p = F(n-1,r) + |x_n|^p$; if $F(n-d-1, r-1) + \sum_{j=n-d}^{n-1} |x_j|^p$ is the smallest value, selecting a minimizer $\widetilde{\mathbf{x}} \in \mathbb{C}^{n-d-1}$ for $F(n-d-1, r-1)$ and considering the r-sparse d-separated vector $\widehat{\mathbf{x}} := (\widetilde{\mathbf{x}}, 0, \ldots, 0, x_n) \in \mathbb{C}^n$ yields $F(n,r) \leq \sum_{j=1}^{n} |x_j - \widehat{x}_j|^p = \sum_{j=1}^{n-d-1} |x_j - \widetilde{x}_j|^p + \sum_{j=n-d}^{n-1} |x_j|^p = F(n-d-1, r-1) + \sum_{j=n-d}^{n-1} |x_j|^p$. With the relation (6) now fully justified, we can fill in a table of values for $F(n,r)$ from the initial values

$$F(n,0) = \|\mathbf{x}_{[\![1:n]\!]}\|_p^p, \quad n \in [\![1:N]\!], \qquad F(n,r) = \|\mathbf{x}_{[\![1:n]\!]}\|_p^p - \max_{j \in [\![1:n]\!]} |x_j|^p, \quad n \in [\![1:d+1]\!], r \in [\![1:s]\!].$$

The latter relation reflects the absence of exactly r-sparse d-disjointed vectors in $\mathbb{C}^n$ when $r \geq 2$ and $n \leq d+1$. Let us observe that, according to (6), determining one entry of the table requires $\mathcal{O}(d)$ arithmetic operations and that there are $\mathcal{O}(sN)$ entries, so computing the error of best approximation by s-sparse d-disjointed vectors requires a total of $\mathcal{O}(dsN) = \mathcal{O}(N^2)$ arithmetic operations (to compare with $\mathcal{O}(N^{3.5})$ for the linear programming strategy of [8]). As for the best approximation itself, we need to keep track of the cases producing the minima in (6)[2]. As illustrated in Figure 2, we follow the path of bold arrows starting from the (N,s)th box until the italicized section: if an arrow points northwest from the (n,r)th box, then the index n is selected for the support of the best approximation, and in the italicized section (first column excluded), the underlined index is selected. Once the support is determined in this way, the best approximation is the vector equal to $\mathbf{x}$ on the support.

[2] we break possible ties arbitrarily by choosing preferentially the second case

$\mathbf{x}$
1
0
1
1.1892
1
0
0.7071

$F(n,r)$	$r=0$	$r=1$	$r=2$	$r=3$
$n=1$	1	0	0	0
$n=2$	1	0	0	0
$n=3$	2	1	0	0
$n=4$	3.4142	2	1	1
$n=5$	4.4142	3	2	1.4142
$n=6$	4.4142	3	2	1.4142
$n=7$	4.9142	3.5	2.5	1.9142

$F(n,r)$	$r=0$	$r=1$	$r=2$	$r=3$
$n=1$	1	0	0	0
$n=2$	1	0	0	0
$n=3$	2	1	0	0
$n=4$	4	2	1	1
$n=5$	5	3	2	2
$n=6$	5	3	2	2
$n=7$	5.25	3.25	2.25	2

Figure 2: Sketch of the dynamic program computing the best s-sparse d-disjointed approximations to $\mathbf{x} = (1, 0, 1, 2^{1/4}, 1, 0, 2^{-1/2})$ with $s = 3$ and $d = 1$ for $p = 2$ (left) and $p = 4$ (right).

3 Sufficient Number of Measurements

The purpose of this section is to show that a number of measurements proportional to $m_{\mathsf{spa\&dis}}$ is enough to ensure robust recovery of sparse disjointed vectors. Such a result was already stated in [8, Theorem 3], see [2] for the proof. The algorithm considered in these articles is an adaptation of CoSaMP, whereas we analyze an adaptation of IHT. We include all the details here mostly for completeness, but also because we believe that they simplify existing arguments. The main tool is a restricted-isometry-like property valid in the general context of union of subspaces, see [3, Theorem 3.3]. The slight difference with the theorem stated below is that the scaling in δ has been reduced from $\ln(1/\delta)/\delta^2$ to the optimal $1/\delta^2$.

Theorem 4. Let $\delta \in (0,1)$ and let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be populated by independent identically distributed subgaussian random variables with variance $1/m$. Then, with probability at least $1 - 2\exp(-c\delta^2 m)$,

$$(1-\delta)\|\mathbf{z}+\mathbf{z}'+\mathbf{z}''\|_2^2 \le \|\mathbf{A}(\mathbf{z}+\mathbf{z}'+\mathbf{z}'')\|_2^2 \le (1+\delta)\|\mathbf{z}+\mathbf{z}'+\mathbf{z}''\|_2^2 \tag{7}$$

for all s-sparse d-disjointed $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{C}^N$, provided

$$m \ge \frac{C}{\delta^2} s \ln\left(e\frac{N - d(s-1)}{s}\right).$$

The constants $c, C > 0$ depend only on the subgaussian distribution.

Proof. We prove the equivalent statement that, with probability at least $1 - 2\exp(-c\delta^2 m)$,

$$\|\mathbf{A}_T^*\mathbf{A}_T - \mathbf{I}\|_{2\to2} \le \delta$$

for all sets $T \subseteq [\![1:N]\!]$ of the form $S \cup S' \cup S''$ where S, S', S'' are d-disjointed subsets of $[\![1:N]\!]$ with size s. Such sets are of size at most $3s$ and their number is upper-bounded by the cube of the number of d-disjointed subsets with size s, i.e., by

$$\binom{N-d(s-1)}{s}^3 \le \left(e\frac{N-d(s-1)}{s}\right)^{3s}.$$

Now, if T is fixed, it is known (see e.g. [7, Theorem 9.9 and Equation (9.12)]) that

$$\mathbb{P}\left(\|\mathbf{A}_T^*\mathbf{A}_T - \mathbf{I}\|_{2\to 2} > \delta\right) \le 2\exp\left(-c'\delta^2 m + c'' s\right)$$

with constants $c', c'' > 0$ depending only on the subgaussian distribution. Taking a union bound over all possible T, we see that the desired result holds with failure probability at most

$$\begin{aligned}\left(e\frac{N-d(s-1)}{s}\right)^{3s} 2\exp\left(-c'\delta^2 m + c'' s\right) &\le 2\exp\left(-c'\delta^2 m + (c''+3)s\ln\left(e\frac{N-d(s-1)}{s}\right)\right)\\ &\le 2\exp(-c'\delta^2 m/2),\end{aligned}$$

where the last inequality holds if one imposes $m \ge [2(c''+3)/c']\delta^{-2} s\ln(e(N-d(s-1))/s)$. □

The restricted-isometry-like property will not be used directly in the form (7), but rather through its two consequences below.

Proposition 5. Suppose that $\mathbf{A}$ satisfies (7). Then, for all s-sparse d-disjointed $\mathbf{x}, \mathbf{x}', \mathbf{x}'' \in \mathbb{C}^N$,

$$|\langle \mathbf{x}-\mathbf{x}', (\mathbf{A}^*\mathbf{A}-\mathbf{I})(\mathbf{x}-\mathbf{x}'')\rangle| \le \delta\|\mathbf{x}-\mathbf{x}'\|_2\|\mathbf{x}-\mathbf{x}''\|_2, \tag{8}$$

and, for all $\mathbf{e} \in \mathbb{C}^m$ and all d-disjointed subsets S, S' of $[\![1:N]\!]$ with size s,

$$\|(\mathbf{A}^*\mathbf{e})_{S\cup S'}\|_2 \le \sqrt{1+\delta}\,\|\mathbf{e}\|_2. \tag{9}$$

Proof. Setting $\mathbf{u} = e^{i\theta}(\mathbf{x}-\mathbf{x}')/\|\mathbf{x}-\mathbf{x}'\|_2$ and $\mathbf{v} = e^{i\upsilon}(\mathbf{x}-\mathbf{x}'')/\|\mathbf{x}-\mathbf{x}''\|_2^2$ for properly chosen $\theta, \upsilon \in [-\pi, \pi]$, we have

$$\begin{aligned}\frac{|\langle \mathbf{x}-\mathbf{x}', (\mathbf{A}^*\mathbf{A}-\mathbf{I})(\mathbf{x}-\mathbf{x}'')\rangle|}{\|\mathbf{x}-\mathbf{x}'\|_2\|\mathbf{x}-\mathbf{x}''\|_2} &= \mathrm{Re}\langle \mathbf{u}, (\mathbf{A}^*\mathbf{A}-\mathbf{I})\mathbf{v}\rangle = \mathrm{Re}\langle \mathbf{A}\mathbf{u}, \mathbf{A}\mathbf{v}\rangle - \mathrm{Re}\langle \mathbf{u}, \mathbf{v}\rangle\\ &= \frac{1}{4}\left(\|\mathbf{A}(\mathbf{u}+\mathbf{v})\|_2^2 - \|\mathbf{A}(\mathbf{u}-\mathbf{v})\|_2^2\right) - \frac{1}{4}\left(\|\mathbf{u}+\mathbf{v}\|_2^2 - \|\mathbf{u}-\mathbf{v}\|_2^2\right)\\ &\le \frac{1}{4}\left|\|\mathbf{A}(\mathbf{u}+\mathbf{v})\|_2^2 - \|\mathbf{u}+\mathbf{v}\|_2^2\right| + \frac{1}{4}\left|\|\mathbf{A}(\mathbf{u}-\mathbf{v})\|_2^2 - \|\mathbf{u}-\mathbf{v}\|_2^2\right|.\end{aligned}$$

Noticing that both $\mathbf{u}+\mathbf{v}$ and $\mathbf{u}-\mathbf{v}$ take the form $\mathbf{z}+\mathbf{z}'+\mathbf{z}''$ for some s-sparse d-disjointed $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{C}^N$, we apply (7) to deduce

$$\frac{|\langle \mathbf{x}-\mathbf{x}', (\mathbf{A}^*\mathbf{A}-\mathbf{I})(\mathbf{x}-\mathbf{x}'')\rangle|}{\|\mathbf{x}-\mathbf{x}'\|_2\|\mathbf{x}-\mathbf{x}''\|_2} \le \frac{\delta}{4}\left(\|\mathbf{u}+\mathbf{v}\|_2^2 + \|\mathbf{u}-\mathbf{v}\|_2^2\right) = \frac{\delta}{2}\left(\|\mathbf{u}\|_2^2 + \|\mathbf{v}\|_2^2\right) = \delta.$$

This shows the desired inequality (8). To prove inequality (9), we write

$$\|(\mathbf{A}^*\mathbf{e})_{S\cup S'}\|_2^2 = \langle (\mathbf{A}^*\mathbf{e})_{S\cup S'}, \mathbf{A}^*\mathbf{e}\rangle = \langle \mathbf{A}((\mathbf{A}^*\mathbf{e})_{S\cup S'}), \mathbf{e}\rangle \le \|\mathbf{A}((\mathbf{A}^*\mathbf{e})_{S\cup S'})\|_2\|\mathbf{e}\|_2.$$

It now suffices to notice that $(\mathbf{A}^*\mathbf{e})_{S\cup S'}$ takes the form $\mathbf{z}+\mathbf{z}'+\mathbf{z}''$ for some s-sparse d-disjointed $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \in \mathbb{C}^N$ (with $\mathbf{z}'' = \mathbf{0}$) to derive $\|\mathbf{A}((\mathbf{A}^*\mathbf{e})_{S\cup S'})\|_2 \le \sqrt{1+\delta}\,\|(\mathbf{A}^*\mathbf{e})_{S\cup S'}\|_2$ from (7). The inequality (9) follows after simplifying by $\|(\mathbf{A}^*\mathbf{e})_{S\cup S'}\|_2$. □

With Proposition 5 at hand, robust uniform recovery is quickly established for the iterative hard thresholding algorithm adapted to the framework of s-sparse d-disjointed vectors. In the description of this algorithm below, $\mathbf{P}_{s,d}$ represents the projection onto (i.e., best ℓ_2-approximation by) s-sparse d-disjointed vectors discussed in Section 2.

Iterative hard thresholding (IHT)

Input: measurement matrix $\mathbf{A}$, measurement vector $\mathbf{y}$, sparsity level s, disjointedness level d.
Initialization: s-sparse d-disjointed vector $\mathbf{x}^0$, typically $\mathbf{x}^0 = \mathbf{0}$.
Iteration: repeat until a stopping criterion is met at $n = \bar{n}$:

$$\mathbf{x}^{n+1} = \mathbf{P}_{s,d}(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)). \tag{IHT}$$

Output: the s-sparse d-disjointed vector $\mathbf{x}^\sharp = \mathbf{x}^{\bar{n}}$.

Theorem 6. Suppose that $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies (7) with $\delta < 1/2$. Then for every s-sparse d-disjointed vector $\mathbf{x} \in \mathbb{C}^N$ acquired via $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \in \mathbb{C}^m$ with an adversarial error $\mathbf{e} \in \mathbb{C}^m$, the output $\mathbf{x}^\sharp := \lim_{n \to \infty} \mathbf{x}^n$ of IHT approximates $\mathbf{x}$ with ℓ_2-error

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq D \|\mathbf{e}\|_2, \tag{10}$$

where $D > 0$ is a constant depending only on δ. In particular, this conclusion is valid with high probability on the draw of a matrix populated by independent zero-mean Gaussian entries with variance $1/m$ provided

$$m \geq C\, m_{\mathsf{spa\&dis}}$$

for some absolute constant $C > 0$.

Proof. We observe that $\mathbf{x}^{n+1}$ is a better s-sparse d-disjointed ℓ_2-approximation to the vector $\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n) = \mathbf{x}^n + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}$ than $\mathbf{x}$ is to derive

$$\|(\mathbf{x}^n + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}) - \mathbf{x}^{n+1}\|_2^2 \leq \|(\mathbf{x}^n + \mathbf{A}^*\mathbf{A}(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}) - \mathbf{x}\|_2^2,$$

i.e.,

$$\|\mathbf{x} - \mathbf{x}^{n+1} + (\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}\|_2^2 \leq \|(\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}\|_2^2.$$

After expanding the squares and rearranging, we deduce

$$\begin{aligned}\|\mathbf{x} - \mathbf{x}^{n+1}\|_2^2 &\leq -2\langle \mathbf{x} - \mathbf{x}^{n+1}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}\rangle \\ &\leq 2|\langle \mathbf{x} - \mathbf{x}^{n+1}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^n)\rangle| + 2\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \|(\mathbf{A}^*\mathbf{e})_{S \cup S^{n+1}}\|_2,\end{aligned}$$

where S and S^{n+1} denote the supports of $\mathbf{x}$ and $\mathbf{x}^{n+1}$, respectively. Applying (8) and (9) and simplifying by $\|\mathbf{x} - \mathbf{x}^{n+1}\|_2$ gives

$$\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \leq 2\delta \|\mathbf{x} - \mathbf{x}^n\|_2 + 2\sqrt{1+\delta}\, \|\mathbf{e}\|_2.$$

With $\delta < 1/2$, this inequality readily implies the desired result (10) with $D := 2\sqrt{1+\delta}/(1-2\delta)$. Combining it with Theorem 4 yields the rest of the statement. $\square$

4 Necessary Number of Measurements

The purpose of this section is to show that a number of noninflating measurements at least proportional to $m_{\mathsf{spa\&dis}}$ is necessary to ensure robust recovery of sparse disjointed vectors. By noninflating measurements relative to the s-sparse d-disjointed model, we mean that the matrix $\mathbf{A}$ associated with the measurement process satisfies, for some absolute constant $c > 0$,

$$\|\mathbf{A}\mathbf{z}\|_2 \le c\|\mathbf{z}\|_2 \qquad \text{whenever } \mathbf{z} \in \mathbb{C}^N \text{ is } s\text{-sparse } d\text{-disjointed}.$$

Figuratively, the energy of a signal with the targeted structure is not inflated by the measurement process. According to Theorems 4 and 6, a random measurement process is likely to be noninflating (with constant $c \le 1+\delta$) and it enables robust uniform recovery of s-sparse d-disjointed vectors when the number of measurements obeys $m \ge C\, m_{\mathsf{spa\&dis}}$. We show below that this is optimal. The key to the argument is a generalization of a lemma used in [6] (see also [7, Lemma 10.12]).

Lemma 7. There exist

$$n \ge \left(\frac{N-d(s-1)}{c_1 s}\right)^{c_2 s} \tag{11}$$

d-disjointed subsets $S_1,\ldots,S_n$ of $[\![1:N]\!]$ such that

$$\operatorname{card}(S_i) = s \quad \text{for all } i \qquad \text{and} \qquad \operatorname{card}(S_i \cap S_j) < \frac{s}{2} \quad \text{for all } i \ne j.$$

The constants $c_1, c_2 > 0$ are universal—precisely, one can take $c_1 = 12e$ and $c_2 = 1/2$.

Proof. Let $\mathcal{A}$ be the collection of all d-disjointed subsets of $[\![1:N]\!]$ with size s. Let us fix an arbitrary $S_1 \in \mathcal{A}$. We then consider the collection $\mathcal{A}_1$ of sets in $\mathcal{A}$ whose intersection with S_1 has size $s/2$ or more, i.e.,

$$\mathcal{A}_1 = \bigcup_{j=\lceil s/2\rceil}^{s} \mathcal{A}_1^j, \qquad \text{where} \quad \mathcal{A}_1^j := \{S \in \mathcal{A} : \operatorname{card}(S_1 \cap S) = j\}.$$

We claim that, for any $j \in [\![\lceil s/2\rceil : s]\!]$,

$$\operatorname{card}(\mathcal{A}_1^j) \le \binom{s}{j}\binom{N-d(s-1)}{s-j} \le \binom{s}{j}\binom{N-d(s-1)}{\lfloor s/2\rfloor}. \tag{12}$$

The first factor upper-bounds the possible choices of the intersection $J := S_1 \cap S$. The second factor upper-bounds the number of d-disjointed subsets of $[\![1:N]\!]$ with size s whose intersection with S_1 is a fixed set J of size j: indeed, by thinking of s-sparse d-disjointed vectors of length N as s blocks with size $d+1$ inside a set with size $N+d$, as we did in Figure 1, we observe in Figure 3 that the d-disjointed subsets of $[\![1:N]\!]$ with size s whose intersection with S_1 equals J inject into the d-disjointed subsets of $[\![1:N-(d+1)j]\!]$ with size $s-j$ by the process of removing the blocks attached to J, so the desired number is at most

$$\binom{N-(d+1)j-d(s-j-1)}{s-j} = \binom{N-d(s-1)-j}{s-j} \le \binom{N-d(s-1)}{s-j}.$$

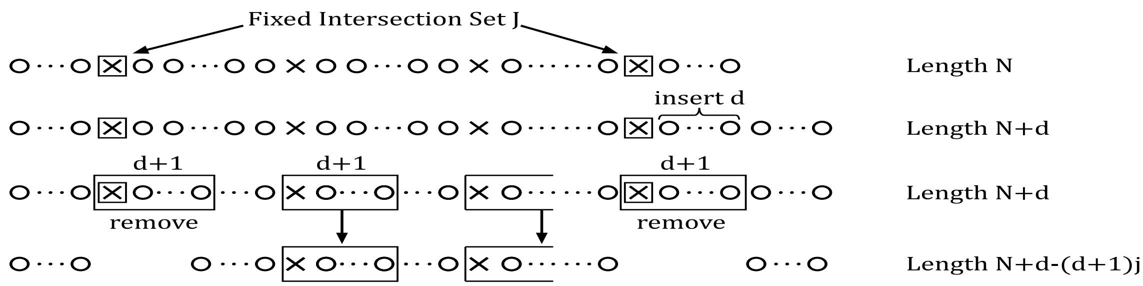

Figure 3: Illustration of the counting argument in the proof of Lemma 7.

This finishes the justification of the first inequality in (12). The second inequality holds because $s - j \le \lfloor s/2 \rfloor \le \lceil (N - d(s-1))/2 \rceil$. It then follows from (12) that

$$\operatorname{card}(\mathcal{A}_1) \le \sum_{j=\lceil s/2 \rceil}^{s} \binom{s}{j} \binom{N - d(s-1)}{\lfloor s/2 \rfloor} \le 2^s \binom{N - d(s-1)}{\lfloor s/2 \rfloor}.$$

Let us now fix an arbitrary set $S_2 \in \mathcal{A} \setminus \mathcal{A}_1$, provided the latter is nonempty. We consider the collection $\mathcal{A}_2$ of sets in $\mathcal{A} \setminus \mathcal{A}_1$ whose intersection with S_2 has size $s/2$ or more, i.e.,

$$\mathcal{A}_2 = \bigcup_{j=\lceil s/2 \rceil}^{s} \mathcal{A}_2^j, \qquad \text{where} \quad \mathcal{A}_2^j := \{ S \in \mathcal{A} \setminus \mathcal{A}_1 : \operatorname{card}(S_2 \cap S) = j \}.$$

The same reasoning as before yields

$$\operatorname{card}(\mathcal{A}_2) \le 2^s \binom{N - d(s-1)}{\lfloor s/2 \rfloor}.$$

We repeat the procedure of selecting sets $S_1, \ldots, S_n$ until $\mathcal{A} \setminus (\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_n)$ becomes empty. In this way, for any $i < j$, the condition $\operatorname{card}(S_i \cap S_j) < s/2$ is automatically fulfilled by virtue of $S_j \in \mathcal{A} \setminus (\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_{j-1}) \subseteq \mathcal{A} \setminus (\mathcal{A}_1 \cup \cdots \cup \mathcal{A}_i)$. Finally, the number n of subsets satisfies

$$n \ge \frac{\operatorname{card}(\mathcal{A})}{\max\limits_{i \in [\![1:n]\!]} \operatorname{card}(\mathcal{A}_i)} \ge \frac{\binom{N - d(s-1)}{s}}{2^s \binom{N - d(s-1)}{\lfloor s/2 \rfloor}} \ge \frac{\left(\frac{N - d(s-1)}{s} \right)^s}{2^s \left(e \frac{N - d(s-1)}{\lfloor s/2 \rfloor} \right)^{\lfloor s/2 \rfloor}} \ge \frac{\left(\frac{N - d(s-1)}{s} \right)^s}{2^s \left(e \frac{N - d(s-1)}{s/3} \right)^{s/2}}$$
$$= \left(\frac{N - d(s-1)}{12es} \right)^{s/2}. \qquad \square$$

We can now turn to the main result of this section.

Theorem 8. Let $\mathbf{A} \in \mathbb{C}^{m \times N}$ be the matrix of a noninflating measurement process with constant c relative to the s-sparse d-disjointed model and let $\Delta : \mathbb{C}^m \to \mathbb{C}^N$ be a reconstruction map providing a robustness estimate

$$\|\mathbf{x} - \Delta(\mathbf{A}\mathbf{x} + \mathbf{e})\|_2 \le D \|\mathbf{e}\|_2 \tag{13}$$

which is valid for all s-sparse d-disjointed vectors $\mathbf{x} \in \mathbb{C}^N$ and all measurement error $\mathbf{e} \in \mathbb{C}^m$. Then the number m of measurements is lower-bounded as

$$m \geq C s \ln\left(e \frac{N - d(s-1)}{s}\right).$$

The constant $C > 0$ depends only on c and D.

Proof. With each S_i of Lemma 7, we associate an s-sparse d-disjointed vector $\mathbf{x}^i \in \mathbb{C}^N$ defined by

$$x^i_\ell = \begin{cases} 1/\sqrt{s}, & \text{if } \ell \in S_i, \\ 0, & \text{if } \ell \notin S_i. \end{cases}$$

We notice that each $\mathbf{x}^i$ satisfies $\|\mathbf{x}^i\|_2 = 1$. We also notice that each $\mathbf{x}^i - \mathbf{x}^j$, $i \neq j$, is supported on the symmetric difference $(S_i \cup S_j) \setminus (S_i \cap S_j)$, which has size larger than s, and since its nonzero entries are $\pm 1/\sqrt{s}$, we have

$$\|\mathbf{x}^i - \mathbf{x}^j\|_2 > 1, \quad i \neq j.$$

Setting $\rho := 1/(2D)$, let us consider the balls in $\mathbb{C}^m \equiv \mathbb{R}^{2m}$ with radius ρ and centered at the $\mathbf{A}\mathbf{x}^i$, i.e.,

$$\mathcal{B}_i := \mathbf{A}\mathbf{x}^i + \rho B_2^m.$$

Using the noninflating property, we easily see that each ball $\mathcal{B}_i$ is contained in $(c+\rho)B_2^m$. This implies that

$$\mathrm{Vol}\left(\bigcup_{i=1}^n \mathcal{B}_i\right) \leq \mathrm{Vol}\left((c+\rho)B_2^m\right) = (c+\rho)^{2m}\mathrm{Vol}(B_2^m). \tag{14}$$

Moreover, we claim that the balls $\mathcal{B}_i$ are disjoint. Indeed, if $\mathcal{B}_i \cap \mathcal{B}_j \neq \emptyset$ for some $i \neq j$, then there would exist $\mathbf{e}^i, \mathbf{e}^j \in B_2^m$ such that $\mathbf{A}\mathbf{x}^i + \rho\mathbf{e}^i = \mathbf{A}\mathbf{x}^j + \rho\mathbf{e}^j =: \mathbf{y}$. Exploiting (13), we could then write

$$1 < \|\mathbf{x}^i - \mathbf{x}^j\|_2 \leq \|\mathbf{x}^i - \Delta(\mathbf{y})\|_2 + \|\mathbf{x}^j - \Delta(\mathbf{y})\|_2 \leq D\|\rho\mathbf{e}^i\|_2 + D\|\rho\mathbf{e}^j\|_2 \leq 2D\rho = 1,$$

which is absurd. It follows that

$$\mathrm{Vol}\left(\bigcup_{i=1}^n \mathcal{B}_i\right) = \sum_{i=1}^n \mathrm{Vol}(\mathcal{B}_i) = n\mathrm{Vol}(\rho B_2^m) = n\rho^{2m}\mathrm{Vol}(B_2^m). \tag{15}$$

Putting (14) and (15) together before making use of (11), we obtain

$$\left(1 + \frac{c}{\rho}\right)^{2m} \geq n, \qquad \text{hence} \quad 2m \ln\left(1 + \frac{c}{\rho}\right) \geq \ln(n) \geq c_2 s \ln\left(\frac{N - d(s-1)}{c_1 s}\right).$$

In view of $\rho = 1/(2D)$, we have arrived at

$$\frac{m}{s} \geq c_3 \ln\left(\frac{N - d(s-1)}{c_1 s}\right), \qquad c_3 := \frac{c_2}{2\ln(1 + 2cD)}.$$

To conclude, we take the obvious lower bound $m/s \geq 1$ into account to derive

$$\left(\frac{1}{c_3} + \ln(ec_1)\right)\frac{m}{s} \geq \ln\left(\frac{N - d(s-1)}{c_1 s}\right) + \ln(ec_1) = \ln\left(e\frac{N - d(s-1)}{s}\right).$$

This is the desired result with $C := c_3/(1 + c_3 \ln(ec_1))$. $\square$

References

[1] D. Amelunxen, M. Lotz, M. McCoy, and J. Tropp, Living on the edge: A geometric theory of phase transitions in convex optimization. Information and Inference, 3(3), 224–294, 2014.

[2] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, Model-based compressive sensing. IEEE Transactions on Information Theory, 56(4), 1982–2001, 2010.

[3] T. Blumensath and M. Davies, Sampling theorems for signals from the union of finite-dimensional linear subspaces. IEEE Transactions on Information Theory, 55(4), 1872–1882, 2009.

[4] E. Candès and C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Communications on Pure and Applied Mathematics, 67(6), 906–956, 2014.

[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, A. Pajor, H. Rauhut, and T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. Journal of Complexity, 26(6), 629–640, 2010.

[7] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing. Birkhäuser, 2013.

[8] C. Hegde, M. Duarte, and V. Cevher, Compressive sensing recovery of spike trains using a structured sparsity model. In: Signal Processing with Adaptive Sparse Structured Representations (SPARS), Saint-Malo, 2009.

[9] M. Minner, On-Grid MIMO Radar via Compressive Sensing. In: 2nd International Workshop on Compressed Sensing applied to Radar (CoSeRa), Bonn, 2013.

[10] S. Oymak, A. Jalali, M. Fazel, Y. Eldar, and B. Hassibi, Simultaneously structured models with application to sparse and low-rank matrices. IEEE Transactions on Information Theory, 61(5), 2886–2908, 2015.