Flavors of Compressive Sensing

Simon Foucart

Abstract About a decade ago, a couple of groundbreaking articles revealed the possibility of faithfully recovering high-dimensional signals from some seemingly incomplete information about them. Perhaps more importantly, practical procedures to perform the recovery were also provided. These realizations had a tremendous impact in science and engineering. They gave rise to a field called "compressive sensing", which is now in a mature state and whose foundations rely on an elegant mathematical theory. This survey presents an overview of the field, accentuating elements from approximation theory, but also highlighting connections with other disciplines that have enriched the theory, e.g. statistics, sampling theory, probability, optimization, metagenomics, graph theory, frame theory, and Banach space geometry.

Compressive sensing was born circa 2005 after influential works of Candès–Romberg–Tao [20] and of Donoho [27]. These works triggered rapid developments benefiting from fruitful interactions between Mathematics, Engineering, Computer Science, and Statistics. The one-sentence summary is:

> Stable and robust recovery of s-sparse vectors in $\mathbb{C}^N$ from an optimal number $m \asymp s ln(eN/s)$ of measurements is achievable via various efficient algorithms.

This survey will clarify the precise meaning of this sentence and its hidden deep theoretical points, all of which can be found in the book [39]. For complete results, I will mostly refer to [39] instead of original sources, which can be found in the Notes sections of [39]. As a rule, I shall not repeat the full underlying arguments here. Several exceptions will be made when proofs are provided either because some assertions are now stronger or because their explanations have been simplified since the publication of [39]. This survey is not intended to be comprehensive, as my first goal is to showcase a view of compressive sensing through an approximation theory lens. In particular, I will not discuss motivating applications such as magnetic resonance imaging, error correction, single-pixel camera, radar, or machine learning, which are all touched upon in [39]. This survey is biased, too, as my second goal is to summarize my own contribution to the field.

Simon Foucart
Department of Mathematics, Texas A&M University, e-mail: foucart@tamu.edu

1 Stability and Robustness in Compressive Sensing

The scenario considered throughout this survey involves a signal space of very large dimension N. One can only acquire a small number $m \ll N$ of linear measurements about the signals from this space. To enable the recovery of signals from the few available measurements, a structural assumption on the signals must be made. A realistic assumption consists in stipulating that the signals of interest have sparse representations in some basis, and we identify from the onset signals of interest with their coefficient vectors $\mathbf{x} \in \mathbb{C}^N$ relative to this basis. We say that $\mathbf{x} \in \mathbb{C}^N$ is s-sparse if

$$\|\mathbf{x}\|_0 := \mathrm{card}\{j \in [\![1:N]\!] : x_j \neq 0\} \leq s.$$

The standard compressive sensing problem reads, in idealized form:[1]

Find measurement matrices $\mathbf{A} \in \mathbb{C}^{m \times N}$ and recovery maps $\Delta : \mathbb{C}^m \to \mathbb{C}^N$ such that

$$\Delta(\mathbf{A}\mathbf{x}) = \mathbf{x} \qquad \text{for all } s\text{-sparse vectors } \mathbf{x} \in \mathbb{C}^N. \tag{1}$$

To make the problem solvable, the number of measurements cannot be too small compared to the sparsity level, precisely $m \geq 2s$ is required (see e.g. [39, p.49]). It turns out that the idealized problem is in fact solvable with $m = 2s$, with the added perks that the matrix $\mathbf{A}$ amounts to discrete Fourier measurements and that the map Δ executes an efficient recovery algorithm related to the age-old Prony's method. Its main lines are (see [39, Theorem 2.15] for more details):

- identify $\mathbf{x} \in \mathbb{C}^N$ with a function on $[\![0:N-1]\!]$ and get the $2s$ Fourier coefficients

$$\hat{x}(j) = \sum_{k=0}^{N-1} x(k) \exp(-\imath 2\pi jk/N), \qquad j \in [\![0:2s-1]\!];$$

- with $S := \mathrm{supp}(x)$, with consider the trigonometric polynomial vanishing on S defined by

$$p(t) := \prod_{k \in S} (1 - \exp(-\imath 2\pi k/N)\exp(\imath 2\pi t/N));$$

- observe that $p \times x = 0$ and deduce by discrete convolution that

$$0 = (\hat{p} * \hat{x})(j) = \sum_{k=0}^{N-1} \hat{p}(k)\hat{x}(j-k), \qquad j \in [\![0:N-1]\!]; \tag{2}$$

- use the facts that $\hat{p}(0) = 1$ and that $\hat{p}(k) = 0$ when $k > s$ to transform the equations (2) for $j \in [\![s:2s-1]\!]$ into a (Toeplitz) system with unknowns $\hat{p}(1), \ldots, \hat{p}(s)$;
- solve the system to determine $\hat{p}$, and in turn p, S, and finally x.

---

[1] Although the problem is stated in the complex setting, our account will often be presented in the real setting. There are almost no differences in the theory, but this sidestep avoids discrepancy with existing litterature concerning e.g. Gelfand widths.

However, this algorithm will go astray[2] if the original vector $\mathbf{x}$ is not exactly sparse or if the measurements do not exactly equal $\mathbf{Ax}$. These two issues are fundamental in compressive sensing: one demands recovery maps that are stable — i.e., they cope well with sparsity defect — and that are robust — i.e., they cope well with measurement error. Let us point out informally that robustness implies stability. Indeed, sparsity defect can be incorporated into the measurement error: if $\mathbf{x}$ is not exactly sparse, consider an index set S of s largest absolute entries of $\mathbf{x}$, so that the inexact measurement vector $\mathbf{y} := \mathbf{Ax} + \mathbf{e}$ can be written as $\mathbf{y} = \mathbf{Ax}_S + \mathbf{e}'$ with $\mathbf{e}' := \mathbf{Ax}_{\overline{S}} + \mathbf{e}$. We shall put a special emphasis on stability in this survey. Therefore, with $\sigma_s(\mathbf{x})_1$ denoting the error of best s-term approximation to a vector $\mathbf{x} \in \mathbb{C}^N$, i.e.,

$$\sigma_s(\mathbf{x})_1 := \min\{\|\mathbf{x} - \mathbf{z}\|_1,\ \mathbf{z} \in \mathbb{C}^N \text{ is } s\text{-sparse}\},$$

we replace the idealized problem (1) by the following refined version of the standard compressive sensing problem:[3]

Find measurement matrices $\mathbf{A} \in \mathbb{C}^{m \times N}$ and recovery maps $\Delta : \mathbb{C}^m \to \mathbb{C}^N$ such that

$$\|\mathbf{x} - \Delta(\mathbf{Ax})\|_2 \le \frac{C}{\sqrt{s}} \sigma_s(\mathbf{x})_1 \qquad \text{for all vectors } \mathbf{x} \in \mathbb{C}^N. \tag{3}$$

In the above formulation, $C > 0$ denotes an absolute constant, i.e., a constant that does not depend on any other parameter of the problem. Throughout this survey, the notation $c, c', C, C', d, d', D, D', \ldots$ always stands for positive absolute constants. They are introduced without notice and they may not represent the same value from line to line. Other pieces of notation used in this survey are quite standard, except possibly $[\![k : \ell]\!]$, which stands for the set $\{k, k+1, \ldots, \ell\}$.

Here is a brief overview of the exposition that follows. Section 2 highlights a connection with Gelfand widths which enables to establish that, due to the stability requirement, the minimal number of measurements increases from $2s$ to linear in s times a logarithmic factor. The choice of suitable measurement matrices is then addressed in Section 3, where the popular restricted isometry property (RIP) is introduced and proved for Gaussian random matrices. In Sections 4 and 5, two recovery procedures, orthogonal matching pursuit (OMP) and basis pursuit (BP), are presented and analyzed under the RIP of the measurement matrix. In Section 6, we exploit the notions introduced in Sections 3, 4, and 5 to shed a different light on the connection between compressive sensing and Gelfand widths — namely, instead of viewing Gelfand widths as a cornerstone of the compressive sensing theory as in Section 2, we derive Gelfand width estimates from pure compressive sensing techniques. In Section 7, nongaussian random measurement matrices are examined. In Section 8, we discuss extensions of the standard compressive sensing problem. In particular, we give new and simplified arguments for the one-bit compressive sensing problem. We close this survey by listing in Section 9 a selection of questions left open in the compressive sensing theory.

[2] This is illustrated in the reproducible MATLAB file found on the author's webpage.

[3] A 'weaker' formulation asks for the estimate $\|\mathbf{x} - \Delta(\mathbf{Ax})\|_1 \le C\sigma_s(\mathbf{x})_1$ for all vectors $\mathbf{x} \in \mathbb{C}^N$.

2 Compressive Sensing: Byproduct of Gelfand Widths Estimates

This section establishes a connection between compressive sensing and the study of widths. Only one direction of this connection is described for now, namely the impact of Gelfand widths on compressive sensing. We work here in the real setting.

2.1 Gelfand widths of the ℓ_1-ball

Widths are a well-studied concept in approximation theory, as illustrated e.g. by the monograph [58]. For a subset K of a real normed space X, the Kolmogorov m-width and Gelfand m-width are defined, respectively, by

$$d_m(K,X) := \inf \left\{ \sup_{\mathbf{x} \in K} \inf_{\mathbf{z} \in X_m} \|\mathbf{x} - \mathbf{z}\|,\ X_m \text{ subspace of } X \text{ with } \dim(X_m) \le m \right\},$$
$$d^m(K,X) := \inf \left\{ \sup_{\mathbf{x} \in K \cap L^m} \|\mathbf{x}\|,\ L^m \text{ subspace of } X \text{ with } \operatorname{codim}(L^m) \le m \right\}.$$

We concentrate on the case of $X = \ell_q^N$ and $K = B_p^N$, i.e., of $\mathbb{R}^N$ equipped with the ℓ_q-norm and the unit ball of ℓ_p^N. In this situation, there is a duality between Kolmogorov and Gelfand widths given by the relation

$$d_m(B_p^N, \ell_q^N) = d^m(B_{q'}^N, \ell_{p'}^N),$$

where $p', q' \ge 1$ are the conjugate exponents of $p, q \ge 1$, i.e., $1/p + 1/p' = 1$ and $1/q + 1/q' = 1$ (see e.g. [39, Theorem 10.14] for a proof). The behaviors of the widths of B_p^N in ℓ_q^N were almost all known in the mid-1970's, but the Gelfand width of B_1^N in ℓ_2^N was notably missing from the picture. Then Kashin [46] and Garnaev–Gluskin [40] proved that

$$d^m(B_1^N, \ell_2^N) \asymp \min\left\{1, \sqrt{\frac{\ln(eN/m)}{m}}\right\}.$$

Precisely, Kashin proved the upper estimate $d^m(B_1^N, \ell_2^N) \le C \min\{1, \sqrt{\ln(eN/m)/m}\}$ (strictly speaking, his estimate featured $\ln(eN/m)^{3/2}$ instead of $\ln(eN/m)^{1/2}$) and Garnaev–Gluskin proved the lower estimate $d^m(B_1^N, \ell_2^N) \ge c \min\{1, \sqrt{\ln(eN/m)/m}\}$. We highlight below the implication that each estimate has for compressive sensing. Note that we shall informally say that a pair $(\mathbf{A}, \Delta)$ of measurement matrix and recovery map is stable of order s with constant C if (3) holds. This corresponds to the notion of mixed (ℓ_2, ℓ_1)-instance optimality in the terminology of [23].

2.2 Consequence of the lower estimate

Briefly stated, *the existence of a stable pair* $(\mathbf{A},\Delta)$ *of order s forces* $m \geq cs\ln(eN/s)$. Indeed, let us consider $\mathbf{v} \in \ker \mathbf{A}$ and let us apply (3) to $-\mathbf{v}_S$ and to $\mathbf{v}_{\overline{S}}$ for some index set S of size s to obtain

$$\|-\mathbf{v}_S - \Delta(\mathbf{A}(-\mathbf{v}_S))\|_2 \leq 0, \qquad \text{i.e.,} \quad -\mathbf{v}_S = \Delta(\mathbf{A}(-\mathbf{v}_S)) = \Delta(\mathbf{A}\mathbf{v}_{\overline{S}}),$$
$$\|\mathbf{v}_{\overline{S}} - \Delta(\mathbf{A}\mathbf{v}_{\overline{S}})\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{v}_{\overline{S}})_1, \quad \text{i.e.,} \quad \|\mathbf{v}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{v}_{\overline{S}})_1 \leq \frac{C}{\sqrt{s}}\|\mathbf{v}\|_1.$$

Given that the latter inequality holds for all $\mathbf{v} \in \ker \mathbf{A}$, which is a subspace of ℓ_2^N of codimension at most m, we deduce that $d^m(B_1^N, \ell_2^N) \leq C/\sqrt{s}$. But since $d^m(B_1^N, \ell_2^N) \geq c\min\{1, \sqrt{\ln(eN/m)/m}\}$, we conclude that $m \geq cs\ln(eN/m)$ (the case $s \leq C$ being set aside). This looks like, but is not exactly the same as, the desired condition $m \geq cs\ln(eN/s)$. It is not hard to see that these two conditions are actually equivalent up to changing the constant c, as justified in [39, Lemma C.6].

2.3 Consequence of the upper estimate

Briefly stated, *the upper estimate provides a pair* $(\mathbf{A},\Delta)$ *which is stable of order* $s \asymp m/ln(eN/m)$. Indeed, in view of $d^m(B_1^N, \ell_2^N) \leq C\sqrt{\ln(eN/m)/m}$, if we set $s \approx m/(8C^2\ln(eN/m))$, then there is a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ (whose null space is an optimal subspace L^m in the definition of the Gelfand width) such that

$$\|\mathbf{v}\|_2 \leq \frac{1}{\sqrt{8s}}\|\mathbf{v}\|_1 \qquad \text{for all } \mathbf{v} \in \ker \mathbf{A}.$$

The recovery map is defined, for any $\mathbf{y} \in \mathbb{R}^m$, by

$$\Delta(\mathbf{y}) = \underset{\mathbf{z} \in \mathbb{R}^N}{\operatorname{argmin}}\, \sigma_s(\mathbf{z})_1 \qquad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}. \tag{4}$$

For a vector $\mathbf{x} \in \mathbb{R}^N$, let us write $\widetilde{\mathbf{x}} := \Delta(\mathbf{A}\mathbf{x})$ for short and let us consider index sets S and $\widetilde{S}$ of s largest absolute entries of $\mathbf{x}$ and $\widetilde{\mathbf{x}}$, respectively. By the definition of $\widetilde{\mathbf{x}}$, we have

$$\|\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_{\widetilde{S}}\|_1 \leq \|\mathbf{x} - \mathbf{x}_S\|_1 = \sigma_s(\mathbf{x})_1.$$

In particular, there holds

$$\|\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_{S \cup \widetilde{S}}\|_1 \leq \|\widetilde{\mathbf{x}} - \widetilde{\mathbf{x}}_{\widetilde{S}}\|_1 \leq \sigma_s(\mathbf{x})_1, \quad \text{as well as} \quad \|\mathbf{x} - \mathbf{x}_{S \cup \widetilde{S}}\|_1 \leq \|\mathbf{x} - \mathbf{x}_S\|_1 = \sigma_s(\mathbf{x})_1.$$

Moreover, the fact that $(\mathbf{x} - \widetilde{\mathbf{x}})_{S \cup \widetilde{S}}$ is $2s$-sparse implies that

$$\|(\mathbf{x} - \widetilde{\mathbf{x}})_{S \cup \widetilde{S}}\|_1 \leq \sqrt{2s}\|(\mathbf{x} - \widetilde{\mathbf{x}})_{S \cup \widetilde{S}}\|_2 \leq \sqrt{2s}\|\mathbf{x} - \widetilde{\mathbf{x}}\|_2.$$

Therefore, observing that $\mathbf{x}-\widetilde{\mathbf{x}} \in \ker \mathbf{A}$, we derive

$$\begin{aligned}\|\mathbf{x}-\widetilde{\mathbf{x}}\|_2 \leq & \frac{1}{\sqrt{8s}}\|\mathbf{x}-\widetilde{\mathbf{x}}\|_1 \leq \frac{1}{\sqrt{8s}}\left(\|(\mathbf{x}-\widetilde{\mathbf{x}})_{\overline{S\cup\widetilde{S}}}\|_1 + \|\mathbf{x}-\mathbf{x}_{S\cup\widetilde{S}}\|_1 + \|\widetilde{\mathbf{x}}-\widetilde{\mathbf{x}}_{S\cup\widetilde{S}}\|_1\right) \\ \leq & \frac{1}{\sqrt{8s}}\left(\sqrt{2s}\|\mathbf{x}-\widetilde{\mathbf{x}}\|_2 + 2\sigma_s(\mathbf{x})_1\right) = \frac{1}{2}\|\mathbf{x}-\widetilde{\mathbf{x}}\|_2 + \frac{1}{\sqrt{2s}}\sigma_s(\mathbf{x})_1.\end{aligned} \tag{5}$$

After rearrangement, we obtain the desired inequality (3) with constant $C=\sqrt{2}$.

2.4 Discussion

Arguably, the story of compressive sensing could very well end here... After all, the standard compressive sensing problem has been solved in its stable version (3). Of course, the optimization program (4) is impractical, but it can be replaced by the computationally-friendly recovery map $\Delta_{\rm BP}$ arising from the ℓ_1-minimization (16). Indeed, it is known (though not really well known) that, if a pair $(\mathbf{A},\Delta)$ is stable of order s, then the pair $(\mathbf{A},\Delta_{\rm BP})$ is stable of order s/C, see [39, Exercise 11.5]. In fact, we will present at the beginning of Section 6 a short argument showing that the upper estimate implies that the pair $(\mathbf{A},\Delta_{\rm BP})$ is stable of order $s \asymp m/\ln(eN/m)$. So why not end the story here, really? Well, for starter, we have to take these difficult Gelfand width estimates for granted. In contrast, the compressive sensing theory can now dissociate itself from Gelfand widths entirely, as we will see in Subsection 6.3. Besides, the theory created some tools that justify the Gelfand width estimates differently, as exposed in Subsections 6.1 and 6.2. The breakthrough in Kashin's original argument for the upper bound was the use of probablisitic techniques. The compressive sensing justification of the upper estimate also relies on such techniques, but it has the appealing feature of decoupling the probabilistic ingredients from the deterministic ingredients. This feature is explained in the next section, where we focus on a property of the matrix $\mathbf{A}$ that implies stability of many pairs $(\mathbf{A},\Delta)$ and we prove that this property is satisfied with high probability by Gaussian random matrices. Whether one counts the story of compressive sensing by beginning from Gelfand widths or by building an independent theory yielding Gelfand width estimates as corollaries is probably a matter of personal taste. It is the latter alternative that has my preference, as it incidentally provided me with some intuition not only on widths but also on other aspects of Banach space geometry such as Kashin's decomposition theorem (see [39, Section 10.3]).

3 The Restricted Isometry Property

The deterministic property mentioned a few lines above is the so-called restricted isometry property (RIP) introduced in [21].

Definition 1. A matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies the restricted isometry property of order s with constant $\delta \in (0,1)$ if

$$(1-\delta)\|\mathbf{z}\|_2^2 \le \|\mathbf{A}\mathbf{z}\|_2^2 \le (1+\delta)\|\mathbf{z}\|_2^2 \qquad \text{for all } s\text{-sparse vectors } \mathbf{z} \in \mathbb{C}^N. \tag{6}$$

The smallest such constant δ is called the restricted isometry constant of order s. It is denoted by δ_s and is also expressed as

$$\delta_s = \max_{\text{card}(S)=s} \|\mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}\|_{2\to 2},$$

where $\mathbf{A}_S$ represents the column-submatrix of $\mathbf{A}$ indexed by S. Our first task is to verify that, in the optimal regime of parameters s, m, and N, RIP is really a nonvoid property. This is achieved via probabilistic techniques. Next, we present a simple algorithm which, coupled to a matrix with RIP, solves the standard compressive sensing problem in an expeditious fashion.

3.1 RIP matrices do exist

For a matrix $\mathbf{A}$ with real entries, it is straightforward to see that the complex RIP (6) holds if and only if its real counterpart involving only real-valued vectors $\mathbf{z}$ holds. It is this real version that is established for Gaussian matrices in the theorem below. Its proof is included because the argument for the concentration inequality is simpler in the Gaussian case than in the more general case treated in [39, Section 9.1].

Theorem 1. *Let $\mathbf{A} \in \mathbb{R}^{m \times N}$ be a random matrix populated by independent Gaussian variables with mean zero and variance $1/m$. If $m \ge C\delta^{-2} s \ln(eN/s)$, then, with probability at least $1 - 2\exp(-c\delta^2 m)$, the matrix $\mathbf{A}$ satisfies the restricted isometry property of order s with constant δ.*

Proof. The classical proof presented here consists of three steps.

Concentration inequality: For a fixed vector $\mathbf{x} \in \mathbb{R}^N$, and for any $i \in [\![1:m]\!]$, since a linear combination of Gaussian random variables is still Gaussian, we have

$$(\mathbf{A}\mathbf{x})_i = \sum_{j=1}^{N} A_{i,j} x_j = \frac{\|\mathbf{x}\|_2}{\sqrt{m}} g_i,$$

where g_i denotes a standard Gaussian random variable (i.e., it has mean zero and variance one). Then, given $t \in (0,1)$, since $\|\mathbf{A}\mathbf{x}\|_2^2 = \sum_{i=1}^m (\mathbf{A}\mathbf{x})_i^2$, we have

$$\begin{aligned}\mathbb{P}\left(\|\mathbf{A}\mathbf{x}\|_2^2 > (1+t)\|\mathbf{x}\|_2^2\right) &= \mathbb{P}\left(\sum_{i=1}^m g_i^2 \ge m(1+t)\right)\\ &= \mathbb{P}\left(\exp\left(u\sum_{i=1}^m g_i^2\right) > \exp(um(1+t))\right),\end{aligned}$$

where the specific value of $u > 0$ will be chosen later. Using Markov's inequality (see e.g. [39, Theorem 7.3]), the independence of the random variables $g_1,\dots,g_m$, as well as the expression of moment generating function $\theta \mapsto \mathbb{E}(\exp(\theta g^2))$ of a squared standard Gaussian random variable (see e.g. [39, Lemma 7.6]), we derive

$$\begin{aligned}\mathbb{P}\left(\|\mathbf{A}\mathbf{x}\|_2^2 > (1+t)\|\mathbf{x}\|_2^2\right) &\le \frac{\mathbb{E}\left(\exp\left(u\sum_{i=1}^m g_i^2\right)\right)}{\exp(um(1+t))} = \frac{\mathbb{E}\left(\prod_{i=1}^m \exp\left(ug_i^2\right)\right)}{\prod_{i=1}^m \exp(u(1+t))}\\ &= \prod_{i=1}^m \frac{\mathbb{E}\left(\exp(ug_i^2)\right)}{\exp(u(1+t))} = \left(\frac{1/\sqrt{1-2u}}{\exp(u(1+t))}\right)^m.\end{aligned}$$

We note that imposing $u < 1/4$ allows us to write

$$\begin{aligned}\frac{1}{\sqrt{1-2u}} = \left(1+\frac{2u}{1-2u}\right)^{1/2} &\le (1+2u(1+4u))^{1/2} \le \exp(2u(1+4u))^{1/2}\\ &= \exp(u+4u^2).\end{aligned}$$

It follows that

$$\mathbb{P}\left(\|\mathbf{A}\mathbf{x}\|_2^2 > (1+t)\|\mathbf{x}\|_2^2\right) \le \left(\exp(4u^2-ut)\right)^m = \exp\left(\frac{t^2}{16}-\frac{t^2}{8}\right)^m = \exp\left(-\frac{mt^2}{16}\right),$$

where we have made the optimal choice $u = t/8 < 1/4$. Likewise, we can show that

$$\mathbb{P}\left(\|\mathbf{A}\mathbf{x}\|_2^2 < (1-t)\|\mathbf{x}\|_2^2\right) \le \exp\left(-\frac{mt^2}{16}\right).$$

All in all, for a fixed $\mathbf{x} \in \mathbb{R}^N$, we have obtained the concentration inequality

$$\mathbb{P}\left(\left|\|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\right| > t\|\mathbf{x}\|_2^2\right) \le 2\exp\left(-\frac{mt^2}{16}\right).$$

Covering argument: We rely on the fact that the unit sphere of $\mathbb{R}^s$ can be covered with $n \le (1+2/\rho)^s$ balls of radius ρ (see e.g. [39, Proposition C.3]). Fixing an index set S of size s and identifying the space $\Sigma_S := \{\mathbf{z} \in \mathbb{R}^N : \mathrm{supp}(\mathbf{z}) \subseteq S\}$ with $\mathbb{R}^s$, this means that there are ℓ_2-normalized vectors $\mathbf{u}_1,\dots,\mathbf{u}_n \in \Sigma_S$, $n \le (1+2/\rho)^s$, such that any ℓ_2-normalized $\mathbf{x} \in \Sigma_S$ satisfies $\|\mathbf{x}-\mathbf{u}_k\|_2 \le \rho$ for some $k \in [\![1:n]\!]$. Given $t \in (0,1)$, the concentration inequality reads, for any $k \in [\![1:n]\!]$,

$$\mathbb{P}\left(\left|\langle(\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I})\mathbf{u}_k,\mathbf{u}_k\rangle\right| > t\right) = \mathbb{P}\left(\left|\|\mathbf{A}\mathbf{u}_k\|_2^2 - \|\mathbf{u}_k\|_2^2\right| > t\right) \le 2\exp\left(-\frac{mt^2}{16}\right).$$

Therefore, writing $\mathbf{B} := \mathbf{A}_S^* \mathbf{A}_S - \mathbf{I}$ for short, by the union bound

$$\begin{aligned}\mathbb{P}\left(|\langle \mathbf{B}\mathbf{u}_k, \mathbf{u}_k\rangle| > t \text{ for some } k \in [\![1:n]\!]\right) &\leq n \times 2\exp\left(-\frac{mt^2}{16}\right) \\ &\leq 2\left(1+\frac{2}{\rho}\right)^s \exp\left(-\frac{mt^2}{16}\right). \end{aligned} \tag{7}$$

Let us place ourselves in the likely situation where $|\langle \mathbf{B}\mathbf{u}_k, \mathbf{u}_k\rangle| \leq t$ for all $k \in [\![1:n]\!]$. Then, for $\mathbf{x} \in \Sigma_S$ with $\|\mathbf{x}\|_2 = 1$, we consider $k \in [\![1:n]\!]$ such that $\|\mathbf{x} - \mathbf{u}_k\|_2 \leq \rho$. We obtain, thanks to the self-adjointness of $\mathbf{B}$,

$$\begin{aligned}|\langle \mathbf{B}\mathbf{x}, \mathbf{x}\rangle| &= |\langle \mathbf{B}\mathbf{u}_k, \mathbf{u}_k\rangle + \langle \mathbf{B}(\mathbf{x}+\mathbf{u}_k), \mathbf{x}-\mathbf{u}_k\rangle| \leq |\langle \mathbf{B}\mathbf{u}_k, \mathbf{u}_k\rangle| + |\langle \mathbf{B}(\mathbf{x}+\mathbf{u}_k), \mathbf{x}-\mathbf{u}_k\rangle| \\ &\leq t + \|\mathbf{B}\|_{2\to 2}\|\mathbf{x}+\mathbf{u}_k\|_2\|\mathbf{x}-\mathbf{u}_k\|_2 \leq t + 2\rho\|\mathbf{B}\|_{2\to 2}.\end{aligned}$$

Taking the supremum over $\mathbf{x}$ yields

$$\|\mathbf{B}\|_{2\to 2} \leq t + 2\rho\|\mathbf{B}\|_{2\to 2}.$$

With the choice $\rho = 1/4$ and $t = \delta/2$, a rearrangement gives

$$\|\mathbf{B}\|_{2\to 2} \leq \frac{t}{1-2\rho} = \delta.$$

This occurs with failure probability at most (7), i.e.,

$$\begin{aligned}\mathbb{P}\left(\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}\|_{2\to 2} > \delta\right) &\leq 2\left(1+\frac{2}{1/4}\right)^s \exp\left(-\frac{m(\delta/2)^2}{16}\right) \\ &= 2\exp\left(\ln(9)s - \frac{m\delta^2}{64}\right).\end{aligned}$$

Union bound: It remains to unfix the index set S. This leads to

$$\begin{aligned}\mathbb{P}(\delta_s > \delta) &= \mathbb{P}\left(\|\mathbf{A}_S^*\mathbf{A}_S - \mathbf{I}\|_{2\to 2} > \delta \text{ for some } S \subseteq [\![1:N]\!] \text{ with } \mathrm{card}(S) = s\right) \\ &\leq \binom{N}{s} \times 2\exp\left(\ln(9)s - \frac{m\delta^2}{64}\right) \leq 2\exp\left(s\ln\left(\frac{eN}{s}\right) + \ln(9)s - \frac{m\delta^2}{64}\right) \\ &\leq 2\exp\left(\ln(9e)s\ln\left(\frac{eN}{s}\right) - \frac{m\delta^2}{64}\right),\end{aligned}$$

where we have used the inequality $\binom{N}{s} \leq (eN/s)^s$ (see e.g. [39, Lemma C.5]). We conclude that

$$\mathbb{P}(\delta_s > \delta) \leq 2\exp\left(-\frac{m\delta^2}{128}\right),$$

provided that $\ln(9e)s\ln(eN/s) \leq m\delta^2/128$, i.e., that $m \geq C\delta^{-2}s\ln(eN/s)$ with $C = 128\ln(9e)$, which is exactly our starting assumption. □

3.2 Recovery for the impatient: iterative hard thresholding

The RIP guarantees that various sparse recovery procedures are stable and robust. Typically, a condition of the type $\delta_{\kappa s} < \delta_*$ for some small integer $\kappa = 1, 2, 3, \ldots$ and some threshold $\delta_* \in (0,1)$ allows one to prove a robustness estimate of the type

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

valid for all s-sparse $\mathbf{x} \in \mathbb{C}^N$ and all vectors $\mathbf{e} \in \mathbb{C}^m$. We will study recovery maps Δ associated with orthogonal matching pursuit and with ℓ_1-minimizations in the next two sections. But the sparse recovery procedure with the easiest justification turns out to be iterative hard thresholding (IHT), so we present the argument here. The algorithm consists in mimicking classical iterative methods for solving the square system $\mathbf{A}^*\mathbf{A}\mathbf{z} = \mathbf{A}^*\mathbf{y}$ and 'sparsifying' each iterate. Precisely, starting from $\mathbf{x}^0 = \mathbf{0}$, we construct a sequence $(\mathbf{x}^n)_{n\ge 0}$ according to

$$\mathbf{x}^{n+1} = H_s(\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)), \qquad n \ge 0.$$

The operator H_s is the hard thresholding operator that keeps s largest absolute entries of a vector and sends the other ones to zero. The condition $\delta_{3s} < 1/2$ below is not optimal, as a little more effort reveals that $\delta_{3s} < 1/\sqrt{3}$ is also sufficient, see [39, Theorem 6.18].

Theorem 2. *Suppose that a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ has a restricted isometry constant $\delta_{3s} < 1/2$. Then, for all s-sparse $\mathbf{x} \in \mathbb{C}^N$ and all $\mathbf{e} \in \mathbb{C}^m$, the output $\mathbf{x}^\infty := \lim_{n\to\infty} \mathbf{x}^n$ of IHT run on $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfies*

$$\|\mathbf{x} - \mathbf{x}^\infty\|_2 \le \frac{d}{1 - 2\delta_{3s}} \|\mathbf{e}\|_2.$$

Proof. It is enough to prove that there are constants $0 < \rho < 1$ and $d > 0$ such that, for any $n \ge 0$,

$$\|\mathbf{x} - \mathbf{x}^{n+1}\|_2 \le \rho\|\mathbf{x} - \mathbf{x}^n\|_2 + d\|\mathbf{e}\|_2.$$

To derive this inequality, we first remark that $\mathbf{x}^{n+1}$ is a better s-term approximation to $\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n)$ than $\mathbf{x}$ is, so that

$$\|\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n) - \mathbf{x}^{n+1}\|_2^2 \le \|\mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n) - \mathbf{x}\|_2^2.$$

Next, introducing $\mathbf{x}$ in the left-hand side and expanding the square gives

$$\begin{aligned}
\|\mathbf{x} - \mathbf{x}^{n+1}\|_2^2 &\le -2\langle \mathbf{x} - \mathbf{x}^{n+1}, \mathbf{x}^n + \mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n) - \mathbf{x}\rangle \\
&= -2\langle \mathbf{x} - \mathbf{x}^{n+1}, (\mathbf{A}^*\mathbf{A} - \mathbf{I})(\mathbf{x} - \mathbf{x}^n) + \mathbf{A}^*\mathbf{e}\rangle \\
&= -2\left(\langle \mathbf{x} - \mathbf{x}^{n+1}, (\mathbf{A}_T^*\mathbf{A}_T - \mathbf{I})(\mathbf{x} - \mathbf{x}^n)\rangle + \langle \mathbf{A}(\mathbf{x} - \mathbf{x}^{n+1}), \mathbf{e}\rangle\right),
\end{aligned}$$

where the index set $T := \mathrm{supp}(\mathbf{x}) \cup \mathrm{supp}(\mathbf{x}^n) \cup \mathrm{supp}(\mathbf{x}^{n+1})$ has size at most $3s$. It then follows that

$$\|\mathbf{x}-\mathbf{x}^{n+1}\|_2^2 \le 2\|\mathbf{x}-\mathbf{x}^{n+1}\|_2\|\mathbf{A}_T^*\mathbf{A}_T-\mathbf{I}\|_{2\to 2}\|\mathbf{x}-\mathbf{x}^n\|_2+2\|\mathbf{A}(\mathbf{x}-\mathbf{x}^{n+1})\|_2\|\mathbf{e}\|_2$$
$$\le 2\delta_{3s}\|\mathbf{x}-\mathbf{x}^{n+1}\|_2\|\mathbf{x}-\mathbf{x}^n\|_2+2\sqrt{1+\delta_{3s}}\|\mathbf{x}-\mathbf{x}^{n+1}\|_2\|\mathbf{e}\|_2.$$

After simplifying by $\|\mathbf{x}-\mathbf{x}^{n+1}\|_2$, we obtain the desired result with $\rho := 2\delta_{3s} < 1$ and $d := 2\sqrt{1+\delta_{3s}} < \sqrt{6}$. □

4 Orthogonal Matching Pursuit

The IHT procedure presented above has a critical deficiency, in that an estimation of the sparsity level is needed for the algorithm to run. This is not the case for the recovery procedures put forward in this section and the next[4]. We start by reviewing the orthogonal matching pursuit (OMP) procedure.

4.1 The algorithm

Orthogonal matching pursuit prevails as a popular algorithm in approximation theory, specifically in the subfield of sparse approximation, where it is better known as orthogonal greedy algorithm (see the book [64] for a comprehensive treatment). There is a subtle difference in perspective, though, in that compressive sensing aims at approximating a vector $\mathbf{x}\in\mathbb{C}^N$ by some $\widetilde{\mathbf{x}}\in\mathbb{C}^N$ computed from $\mathbf{y}=\mathbf{A}\mathbf{x}\in\mathbb{C}^m$ only, while the goal in sparse approximation is to approximate a vector $\mathbf{y}\in\mathbb{C}^m$, not necessarily of the form $\mathbf{y}=\mathbf{A}\mathbf{x}$, by some $\mathbf{A}\widetilde{\mathbf{x}}\in\mathbb{C}^m$ where $\widetilde{\mathbf{x}}\in\mathbb{C}^N$ is sparse. Under RIP, these tasks are somewhat equivalent since $\|\mathbf{A}(\mathbf{x}-\widetilde{\mathbf{x}})\|_2\approx\|\mathbf{x}-\widetilde{\mathbf{x}}\|_2$ when $\mathbf{x}$ and $\widetilde{\mathbf{x}}$ are sparse. Regardless of the intended usage, the OMP algorithm does not change — it is an iterative algorithm that consists in building a candidate support by adding new indices one by one and producing candidate vectors with the canditate support that best fit the measurements. More formally, starting from the index set $S^0=\emptyset$ and the vector $\mathbf{x}^0=\mathbf{0}$, a sequence $(S^n,\mathbf{x}^n)_{n\ge 0}$ is constructed according to

$$S^{n+1} := S^n \cup \{j^{n+1}\}, \qquad \text{where } j^{n+1} := \underset{j\in[\![1:N]\!]}{\operatorname{argmin}}\, |(\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_j|, \qquad (\mathrm{OMP}_1)$$

$$\mathbf{x}^{n+1} := \underset{\mathbf{z}\in\mathbb{C}^N}{\operatorname{argmin}}\,\|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2 \ \text{ subject to } \operatorname{supp}(\mathbf{z})\subseteq S^{n+1}, \qquad (\mathrm{OMP}_2)$$

until a stopping criterion is met. The rule for choosing the new index j^{n+1} is, in retrospect, justified by (10) when the columns $\mathbf{a}_1,\ldots,\mathbf{a}_N$ of $\mathbf{A}$ are ℓ_2-normalized, which we assume for the rest of this subsection. As a matter of fact, if one wishes to be as greedy as one can be and decrease the norm of the residual as much as possible at every iteration, the new index j^{n+1} should maximize the quantity $|(\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_j|/\operatorname{dist}(\mathbf{a}_j,\operatorname{span}\{\mathbf{a}_{j^1},\ldots,\mathbf{a}_{j^n}\})$ instead of $|(\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_j|$. Under RIP, the former is approximately equal to the latter, though. See [30] for details.

---

[4] Arguably, orthogonal matching pursuit could require an estimation of the sparsity level if an estimation of the magnitude of the measurement error is not available.

4.2 RIP analysis

If one wants to recover s-sparse vectors via OMP, it seems quite natural to stop the algorithm after s iterations (although this would require a prior estimation of s). In fact, when $m \asymp s\ln(N)$, it was established in [67] that, given a fixed s-sparse vector $\mathbf{x} \in \mathbb{C}^N$, one has $\mathbb{P}(\mathbf{x}^s = \mathbf{x}) \geq 1 - 1/N$. But a stronger result of the type $\mathbb{P}(\mathbf{x}^s = \mathbf{x} \text{ for all } s\text{-sparse } \mathbf{x} \in \mathbb{C}^N) \geq 1 - o(N)$ is not possible, see [27, Section 7]. However, such a result becomes possible if the number of iterations is allowed to exceed s. Of course, this number of iterations should remain small, ideally at most proportional to s, as is the case for e.g. Hard Thresholding Pursuit [9]. This objective is achievable under RIP. Here is a precise statement, taking robustness into account. In this statement, one could take $\kappa = 12$ and $\delta_* = 1/6$, but other choices could be made, too.

Theorem 3. *Suppose that a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ with ℓ_2-normalized columns has restricted isometry constant $\delta_{(1+\kappa)s} < \delta_*$. Then, for all s-sparse $\mathbf{x} \in \mathbb{C}^N$ and all $\mathbf{e} \in \mathbb{C}^m$, the output of the OMP algorithm run on $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfies*

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}^{\kappa s}\|_2 \leq D\|\mathbf{e}\|_2. \tag{8}$$

The original proof of this result is rather convoluted, see [69]. It was slightly simplified in [30] (see also [39, Theorem 6.25]), and later much simplified in [24]. Even so, the arguments feel artificial and a natural proof is still coveted. There is an easy explanation, presented below, for the recovery of sparse vectors that are ‘flat’ on their supports. Numerical evidence suggest that this is in fact the worst case, see Figure 4.2. The best case seems to be the one of sparse vectors whose nonincreasing rearrangements decay geometrically, because each new index j^{n+1} can be shown to belong to the true support. Intuitively, one feels that there should be a simple reasoning bridging these two extreme cases.

Proposition 1. *Given $\gamma \geq 1$, let $\mathbf{x} \in \mathbb{C}^N$ be an s-sparse vector which satisfies $\max\{|x_j|, j \in \mathrm{supp}(\mathbf{x})\} \leq \gamma \min\{|x_j|, j \in \mathrm{supp}(\mathbf{x})\}$. With $t := \lceil 12\gamma^4 s \rceil$, if a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ with ℓ_2-normalized columns has restricted isometry constant $\delta_{s+t} \leq 1/5$, then for all $\mathbf{e} \in \mathbb{C}^m$, the output of the OMP algorithm run on $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ satisfies*

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}^t\|_2 \leq 4\|\mathbf{e}\|_2.$$

Proof. We recall the following observations (see [39, Lemmas 3.4 and 3.3]):

$$(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_{S^n} = \mathbf{0}, \tag{9}$$

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}^{n+1}\|_2^2 \leq \|\mathbf{y} - \mathbf{A}\mathbf{x}^n\|_2^2 - |(\mathbf{A}^*(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_{j^{n+1}}|^2. \tag{10}$$

With S denoting the support of $\mathbf{x}$, we now remark that

$$\begin{aligned}
\|\mathbf{x}_{S\setminus S^n}\|_1 |(\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_{j^{n+1}}| &\geq \sum_{j\in S\setminus S^n} x_j (\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_j \\
&= \sum_{j\in S\cup S^n} (\mathbf{x}-\mathbf{x}^n)_j (\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_j = \langle \mathbf{x}-\mathbf{x}^n, \mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n)\rangle \\
&= \langle \mathbf{A}(\mathbf{x}-\mathbf{x}^n), \mathbf{y}-\mathbf{A}\mathbf{x}^n\rangle = \|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2^2 + \langle \mathbf{A}(\mathbf{x}-\mathbf{x}^n), \mathbf{e}\rangle \\
&\geq \|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2^2 - \|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2 \|\mathbf{e}\|_2.
\end{aligned}$$

For $n < t$, we may assume that $\|\mathbf{e}\|_2 \leq (1-c)\|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2$, where $c := \sqrt{5}/4$, otherwise $\|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2 \leq \|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2 + \|\mathbf{e}\|_2 \leq (1/(1-c)+1)\|\mathbf{e}\|_2 \leq 4\|\mathbf{e}\|_2$, so the result is acquired from $\|\mathbf{y}-\mathbf{A}\mathbf{x}^t\|_2 \leq \|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2$. We may also assume that $\|\mathbf{x}_{S\setminus S^n}\|_1 > 0$, otherwise $S \subseteq S^n$ and $\|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2 \leq \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{e}\|_2$, so the result is again acquired from $\|\mathbf{y}-\mathbf{A}\mathbf{x}^t\|_2 \leq \|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2$. Therefore, with $\delta := \delta_{s+t}$, we obtain $|(\mathbf{A}^*(\mathbf{y}-\mathbf{A}\mathbf{x}^n))_{j^{n+1}}| \geq c\|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2^2/\|\mathbf{x}_{S\setminus S^n}\|_1$, and in turn

$$\begin{aligned}
\|\mathbf{y}-\mathbf{A}\mathbf{x}^{n+1}\|_2^2 &\leq \|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2^2 - \frac{c^2\|\mathbf{A}(\mathbf{x}-\mathbf{x}^n)\|_2^4}{\|\mathbf{x}_{S\setminus S^n}\|_1^2} \\
&\leq \|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2^2 - \frac{c^2(1-\delta)^2\|\mathbf{x}-\mathbf{x}^n\|_2^4}{\|\mathbf{x}_{S\setminus S^n}\|_1^2}.
\end{aligned}$$

With $\alpha := \min\{|x_j|, j\in S\}$ and $\beta := \max\{|x_j|, j \in S\}$, we now notice that

$$\begin{aligned}
&\|\mathbf{x}-\mathbf{x}^n\|_2^2 \geq \|\mathbf{x}_{S\setminus S^n}\|_2^2 \geq \alpha^2 \mathrm{card}(S\setminus S^n) = \frac{\beta^2}{\gamma^2}\mathrm{card}(S\setminus S^n), \\
&\|\mathbf{x}_{S\setminus S^n}\|_1 \leq \beta\,\mathrm{card}(S\setminus S^n).
\end{aligned}$$

These inequalities allows us to derive that

$$\|\mathbf{y}-\mathbf{A}\mathbf{x}^{n+1}\|_2^2 \leq \|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2^2 - \frac{c^2(1-\delta^2)}{\gamma^4}\beta^2 \leq \|\mathbf{y}-\mathbf{A}\mathbf{x}^n\|_2^2 - \frac{c^2(1-\delta^2)}{\gamma^4}\frac{\|\mathbf{x}\|_2^2}{s}.$$

By immediate induction, we arrive at

$$\begin{aligned}
\|\mathbf{y}-\mathbf{A}\mathbf{x}^t\|_2^2 &\leq \|\mathbf{y}\|_2^2 - t\frac{c^2(1-\delta^2)}{\gamma^4}\frac{\|\mathbf{x}\|_2^2}{s} \leq 2\|\mathbf{A}\mathbf{x}\|_2^2 + 2\|\mathbf{e}\|_2^2 - \frac{t}{s}\frac{c^2(1-\delta^2)}{\gamma^4}\|\mathbf{x}\|_2^2 \\
&\leq \left(2(1+\delta) - \frac{t}{s}\frac{c^2(1-\delta)^2}{\gamma^4}\right)\|\mathbf{x}\|_2^2 + 2\|\mathbf{e}\|_2^2 \\
&= \left(\frac{12}{5} - \frac{t}{s}\frac{1}{5\gamma^4}\right)\|\mathbf{x}\|_2^2 + 2\|\mathbf{e}\|_2^2 \leq 2\|\mathbf{e}\|_2^2,
\end{aligned}$$

where the last two steps followed from the choices $\delta = 1/5$ and $t = \lceil 12\gamma^4 s\rceil$. $\square$

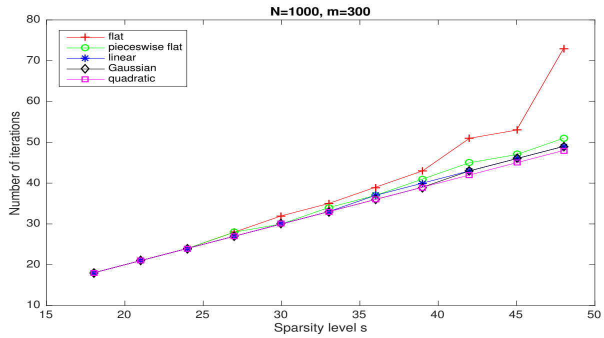

Fig. 1 Number of iterations OMP needs to recovery s-sparse vectors, maximized over 200 tries, when the sparse vectors are: flat ($x_j = 1$), piecewise flat ($x_j \in \{1,2\}$), linear ($x_j = j$), Gaussian ($x_j \sim \mathcal{N}(0,1)$), and quadratic ($x_j = j^2$).

4.3 Stability via the sort-and-split technique

The bound (8) takes care of measurement error, but what about sparsity defect? As mentioned in Section 1, this can be incorporated into the measurement error. Below is an argument that leads to stability and robustness estimates. For $\mathbf{x} \in \mathbb{C}^N$, not necessarily s-sparse, we consider an index set T of $t = 2s$ largest absolute entries of $\mathbf{x}$ and we write $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ as $\mathbf{y} = \mathbf{A}\mathbf{x}_T + \mathbf{e}'$ with $\mathbf{e}' := \mathbf{A}\mathbf{x}_{\overline{T}} + \mathbf{e}$. Then, provided $\delta_{2(1+\kappa)s} < \delta_*$, we can apply Theorem 3 to obtain

$$\|\mathbf{y} - \mathbf{A}\mathbf{x}^{2\kappa s}\|_2 \le D\|\mathbf{e}'\|_2 \le D\|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 + D\|\mathbf{e}\|_2.$$

To get a bound for $\|\mathbf{x} - \mathbf{x}^{2\kappa s}\|_2$ instead — a 'compressive sensing' estimate rather than a 'sparse approximation' estimate, so to speak — we start by noticing that

$$\|\mathbf{x} - \mathbf{x}^{2\kappa s}\|_2 \le \|\mathbf{x}_T - \mathbf{x}^{2\kappa s}\|_2 + \|\mathbf{x}_{\overline{T}}\|_2 \le \|\mathbf{x}_T - \mathbf{x}^{2\kappa s}\|_2 + \frac{1}{\sqrt{s}}\sigma_s(\mathbf{x})_1, \tag{11}$$

where we have applied Stechkin bound for the ℓ_q-norm of the tail of a vector in terms of its ℓ_p-norm, $p < q$ (see [39, Proposition 2.3]). Then, using the RIP, we have

$$\begin{aligned}\|\mathbf{x}_T - \mathbf{x}^{2\kappa s}\|_2 &\le \frac{1}{\sqrt{1-\delta_*}}\|\mathbf{A}(\mathbf{x}_T - \mathbf{x}^{2\kappa s})\|_2 = \frac{1}{\sqrt{1-\delta_*}}\|\mathbf{y} - \mathbf{A}\mathbf{x}^{2\kappa s} - (\mathbf{A}\mathbf{x}_{\overline{T}} + \mathbf{e})\|_2 \\ &\le \frac{1}{\sqrt{1-\delta_*}}\left(\|\mathbf{y} - \mathbf{A}\mathbf{x}^{2\kappa s}\|_2 + \|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 + \|\mathbf{e}\|_2\right) \\ &\le C\|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 + D\|\mathbf{e}\|_2. \end{aligned} \tag{12}$$

It would be nice to try and bound $\|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2$ by an absolute constant times $\|\mathbf{x}_{\overline{T}}\|_2$, but this is doomed (in the terminology of [23], it would imply ℓ_2-instance optimality, which is only possible for $m \asymp N$). Instead, the key is to use a technique that has now become routine in compressive sensing — I call it the sort-and-split technique in this survey. We decompose $[\![1:N]\!]$ into index sets $S_0, S_1, S_2, \dots$ of size s in such a way that S_0 contains s largest absolute entries of $\mathbf{x}$, S_1 contains s next largest absolute entries of $\mathbf{x}$, etc. We invoke the RIP again to write

$$\|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 = \left\|\mathbf{A}\left(\sum_{k\geq 2}\mathbf{x}_{S_k}\right)\right\|_2 \leq \sum_{k\geq 2}\|\mathbf{A}\mathbf{x}_{S_k}\|_2 \leq \sqrt{1+\delta_*}\sum_{k\geq 2}\|\mathbf{x}_{S_k}\|_2.$$

Now, since any absolute entry of $\mathbf{x}_{S_k}$ is bounded by any absolute entry of $\mathbf{x}_{S_{k-1}}$, we can easily check that, for each $k \geq 2$,

$$\|\mathbf{x}_{S_k}\|_2 \leq \frac{1}{\sqrt{s}}\|\mathbf{x}_{S_{k-1}}\|_1.$$

This yields

$$\|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 \leq \frac{\sqrt{1+\delta_*}}{\sqrt{s}}\sum_{k\geq 2}\|\mathbf{x}_{S_{k-1}}\|_1 \leq \frac{\sqrt{1+\delta_*}}{\sqrt{s}}\|\mathbf{x}_{\overline{S_0}}\|_1 = \frac{\sqrt{1+\delta_*}}{\sqrt{s}}\sigma_s(\mathbf{x})_1. \tag{13}$$

Putting (11), (12), and (13) together gives

$$\|\mathbf{x}-\mathbf{x}^{2\kappa s}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\|\mathbf{e}\|_2. \tag{14}$$

This is the stability and robustness estimate incorporating (3) that we were after. Note that the argument remains valid for other algorithms, e.g. for IHT by starting from the result of Theorem 2.

We finally point out that a stability and robustness estimate in the 'weaker' sense of footnote 3 can also be obtained via the sort-and-split technique. Indeed, instead of (11), we could also have written

$$\|\mathbf{x}-\mathbf{x}^{2\kappa s}\|_1 \leq \|\mathbf{x}_T-\mathbf{x}^{2\kappa s}\|_1 + \|\mathbf{x}_{\overline{T}}\|_1 \leq \sqrt{2(1+\kappa)s}\|\mathbf{x}_T-\mathbf{x}^{2\kappa s}\|_2 + \sigma_s(\mathbf{x})_1.$$

The latter term is at most the right-hand side of (11) multiplied by $\sqrt{2(1+\kappa)s}$. Therefore, calling upon (12) and (13) again, we arrive at the following recovery estimate measuring error in ℓ_1 rather than ℓ_2:

$$\|\mathbf{x}-\mathbf{x}^{2\kappa s}\|_1 \leq C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\|\mathbf{e}\|_2. \tag{15}$$

5 Basis Pursuits

In this section, we consider both the equality-constrained ℓ_1-minimization

$$\underset{\mathbf{z} \in \mathbb{K}^N}{\text{minimize}} \, \|\mathbf{z}\|_1 \qquad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}, \tag{16}$$

as well as the quadratically-constrained ℓ_1-minimization

$$\underset{\mathbf{z} \in \mathbb{K}^N}{\text{minimize}} \, \|\mathbf{z}\|_1 \qquad \text{subject to } \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \le \eta. \tag{17}$$

The first program is relevant when the measurements are perfectly accurate, while the second program seems more relevant when there is some measurement error in $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$, but a bound $\|\mathbf{e}\|_2 \le \eta$ on the magnitude of the measurement error is available. There are several intuitive explanations for why ℓ_1-minimization does promote sparsity. One of them is the fact that a minimizer of (16) or (17) is automatically m-sparse, provided it is unique (see [39, Theorem 3.1 and Exercise 3.3]). Beware here of a subtle difference between the real and complex settings — this statement, true when $\mathbb{K} = \mathbb{R}$, becomes false when $\mathbb{K} = \mathbb{C}$ (see [39, Exercise 3.2]). As another subtle difference between the two settings, the optimization problem (16) can be transformed into a linear program when $\mathbb{K} = \mathbb{R}$, but is only into a second-order-cone program when $\mathbb{K} = \mathbb{C}$ (see [39, pp. 63–64]). These considerations set aside, for exact sparse recovery from uncorrupted measurements, the equivalence between the following two assertions is easy to verify (approximation theorists can view it as resulting from characterizations of best approximation in ℓ_1-norm):

(a) every s-sparse $\mathbf{x} \in \mathbb{K}^N$ is the unique minimizer of (16) with $\mathbf{y} = \mathbf{A}\mathbf{x}$,
(b) for all $\mathbf{v} \in \ker \mathbf{A} \setminus \{\mathbf{0}\}$, for all index sets S of size s,

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\overline{S}}\|_1.$$

The latter property is called the null space property (NSP). The name emphasizes that the success of sparse recovery via ℓ_1-minimization depends only on the null space of the measurement matrix, not on the matrix itself, so e.g. one can safely rescale, reshuffle, or add measurements (see [39, Remark 4.6]). Note that, given a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, (b) could be interpreted in two different ways, whether $\ker \mathbf{A}$ is considered a real vector space or a complex vector space. Remarkably, these two versions of the NSP are equivalent (see [34] or [39, Theorem 4.7]).

5.1 *Robust null space property*

The recovery procedure considered in this subsection corresponds to (17) and is called quadratically-constrained basis pursuit — as the reader has guessed already, basis pursuit is synonymous to ℓ_1-minimization. We will use a specific notation for the associated recovery map, namely

$$\Delta_{\mathrm{BP}}^{\eta}(\mathbf{y}) := \operatorname{argmin} \|\mathbf{z}\|_1 \qquad \text{subject to } \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_2 \le \eta.$$

We target stability and robustness estimates akin to (14), with η replacing $\|\mathbf{e}\|_2$, i.e.,

$$\|\mathbf{x}-\Delta_{\mathrm{BP}}^{\eta}(\mathbf{A}\mathbf{x}+\mathbf{e})\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\eta \tag{18}$$

valid for all $\mathbf{x} \in \mathbb{C}^N$ and all $\mathbf{e} \in \mathbb{C}^m$ with $\|\mathbf{e}\|_2 \le \eta$. Assuming that (18) holds, if we make the particular choice $\mathbf{x} = \mathbf{v} \in \mathbb{C}^N$, $\mathbf{e} = -\mathbf{A}\mathbf{v}$, and $\eta = \|\mathbf{A}\mathbf{v}\|_2$, we have $\Delta_{\mathrm{BP}}^{\eta}(\mathbf{A}\mathbf{x}+\mathbf{e}) = \mathbf{0}$, hence

$$\|\mathbf{v}\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(\mathbf{v})_1 + D\|\mathbf{A}\mathbf{v}\|_2.$$

It follows in particular that, for any $\mathbf{v} \in \mathbb{C}^N$ and any index set S of size s,

$$\|\mathbf{v}_S\|_2 \le \frac{C}{\sqrt{s}}\|\mathbf{v}_{\overline{S}}\|_1 + D\|\mathbf{A}\mathbf{v}\|_2.$$

This is called the robust null space property (RNSP) of order s with constants $C, D > 0$. We have just pointed out that it is a necessary condition for the stability and robustness estimate (18) to hold. As it turns out, the RNSP is also a sufficient condition, with the proviso that $C < 1$ (this proviso makes sense in view of the NSP mentioned in (b)). Here is a more general result (see [39, Theorem 4.25]). To see why (19) below is more general than (18), just apply (19) to $\mathbf{z} := \Delta_{\mathrm{BP}}^{\eta}(\mathbf{A}\mathbf{x}+\mathbf{e})$, so that $\|\mathbf{z}\|_1 \le \|\mathbf{x}\|_1$ and $\|\mathbf{A}(\mathbf{z}-\mathbf{x})\|_2 \le \|\mathbf{y}-\mathbf{A}\mathbf{z}\|_2 + \|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2 \le 2\eta$.

Theorem 4. *If a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ satisfies the RNSP of order s with constants $0 < \rho < 1$ and $\tau > 0$, then for any $\mathbf{x}, \mathbf{z} \in \mathbb{C}^N$,*

$$\|\mathbf{x}-\mathbf{z}\|_2 \le \frac{C}{\sqrt{s}}\left(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\sigma_s(\mathbf{x})_1\right) + D\|\mathbf{A}(\mathbf{z}-\mathbf{x})\|_2, \tag{19}$$

with constants $C, D > 0$ depending only on ρ and τ.

We now point out that the RNSP can be deduced from the RIP. The easiest way to do so would be to reproduce the argument proposed in [39, Theorem 6.9 and Exercise 6.12], which invokes the condition $\delta_{2s} < 1/3$. We provide a less stringent condition below. For several years, researchers improved sufficient conditions of the type $\delta_{2s} < \delta_*$ by raising the threshold δ_*. As mentioned in [26], it is impossible to take $\delta_* > 1/\sqrt{2}$. It has recently been proved in [14] that one can take $\delta_* = 1/\sqrt{2}$. We paraphrase the argument to fit the RNSP framework.

Theorem 5. *If a matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ has restricted isometry constant*

$$\delta_{2s} < \frac{1}{\sqrt{2}},$$

then it satisfies the RNSP with constants $0 < \rho < 1$ and $\tau > 0$ depending only on δ_{2s}.

Proof. Given a vector $\mathbf{v} \in \mathbb{C}^N$ and an index set S of size s, our goal is to find constants $0 < \rho < 1$ and $\tau > 0$ such that

$$\|\mathbf{v}_S\|_2 \le \frac{\rho}{\sqrt{s}} \|\mathbf{v}_{\overline{S}}\|_1 + \tau \|\mathbf{A}\mathbf{v}\|_2.$$

Note that we may assume that $\mathbf{v} \in \mathbb{R}^N$. Indeed, we can multiply by a diagonal matrix $\mathbf{D}$ whose nonzero entries have modulus one and replace $\mathbf{v} \in \mathbb{C}^N$ by $\mathbf{D}\mathbf{v} \in \mathbb{R}^N$. At the same time, we would replace $\mathbf{A}$ by $\mathbf{A}\mathbf{D}^{-1}$, which does not change the restricted isometry constants. We now partition the complement $\overline{S}$ of S as $\overline{S} = S' \cup S''$, where

$$S' := \left\{ j \in \overline{S} : |v_j| > \frac{\|\mathbf{v}_{\overline{S}}\|_1}{s} \right\}, \qquad S'' := \left\{ j \in \overline{S} : |v_j| \le \frac{\|\mathbf{v}_{\overline{S}}\|_1}{s} \right\}.$$

With $s' := \mathrm{card}(S')$, we derive that $s' \le s$ from $\|\mathbf{v}_{S'}\|_1 \ge s' \|\mathbf{v}_{\overline{S}}\|_1 / s \ge s' \|\mathbf{v}_{S'}\|_1 / s$. We further remark that

$$\|\mathbf{v}_{S''}\|_1 = \|\mathbf{v}_{\overline{S}}\|_1 - \|\mathbf{v}_{S'}\|_1 \le (s - s') \frac{\|\mathbf{v}_{\overline{S}}\|_1}{s},$$

while we also have

$$\|\mathbf{v}_{S''}\|_\infty \le \frac{\|\mathbf{v}_{\overline{S}}\|_1}{s}.$$

Thus, the vector $\mathbf{v}_{S''}$ belong to a scaled version of the polytope $(s - s') B_1^{s''} \cap B_\infty^{s''}$, $s'' := \mathrm{card}(S'')$. A key to the argument of [14] was to recognize and prove that this polytope can be represented as the convex hull of $(s-s')$-sparse vectors. In fact, this result is similar to [54, Lemma 5.2]. It allows us to write $\mathbf{v}_{S''}$ as a convex combination

$$\mathbf{v}_{S''} = \sum_{i=1}^n \lambda_i \mathbf{u}^i,$$

of $(s - s')$-sparse vectors $\mathbf{u}^1, \ldots, \mathbf{u}^n$ supported on S'' that satisfy $\|\mathbf{u}^i\|_\infty \le \|\mathbf{v}_{\overline{S}}\|_1 / s$ for all $i \in [\![1 : n]\!]$. We now observe that

$$\begin{aligned}
&\sum_{i=1}^n \lambda_i \left[\|\mathbf{A}((1+\delta)(\mathbf{v}_S + \mathbf{v}_{S'}) + \delta \mathbf{u}^i)\|_2^2 - \|\mathbf{A}((1-\delta)(\mathbf{v}_S + \mathbf{v}_{S'}) - \delta \mathbf{u}^i)\|_2^2 \right] \\
&\qquad = \sum_{i=1}^n \lambda_i \left[\langle \mathbf{A}(2(\mathbf{v}_S + \mathbf{v}_{S'})), \mathbf{A}(2\delta(\mathbf{v}_S + \mathbf{v}_{S'}) + 2\delta \mathbf{u}^i) \rangle \right] \\
&\qquad = 4\delta \langle \mathbf{A}(\mathbf{v}_s + \mathbf{v}_{S'}), \mathbf{A}(\mathbf{v}_S + \mathbf{v}_{S'} + \sum_{i=1}^n \lambda_i \mathbf{u}^i) \rangle \\
&\qquad = 4\delta \langle \mathbf{A}(\mathbf{v}_S + \mathbf{v}_{S'}), \mathbf{A}\mathbf{v} \rangle. \qquad (20)
\end{aligned}$$

Because the sparsity of the vectors involved in the left-hand side of (20) is at most $s + s' + (s - s') = 2s$, with $\delta := \delta_{2s}$, each summand in this left-hand side is bounded from below by

$$
\begin{aligned}
&(1-\delta)\|(1+\delta)(\mathbf{v}_S+\mathbf{v}_{S'})+\delta\mathbf{u}^i\|_2^2-(1+\delta)\|(1-\delta)(\mathbf{v}_S+\mathbf{v}_{S'})-\delta\mathbf{u}^i\|_2^2\\
&=(1-\delta)\left((1+\delta)^2\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2^2+\delta^2\|\mathbf{u}^i\|_2^2\right)-(1+\delta)\left((1-\delta)^2\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2^2+\delta^2\|\mathbf{u}^i\|_2^2\right)\\
&=(1-\delta^2)2\delta\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2^2-2\delta^3\|\mathbf{u}^i\|_2^2.
\end{aligned}
$$

Using the fact that $\|\mathbf{u}^i\|_2^2 \le (s-s')\|\mathbf{u}^i\|_\infty^2 \le (s-s')\|\mathbf{v}_{\overline{S}}\|_1^2/s^2 \le \|\mathbf{v}_{\overline{S}}\|_1^2/s$, multiplying by λ_i and summing over i shows that the left-hand side of (20) is bounded from below by

$$(1-\delta^2)2\delta\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2^2-2\delta^3\frac{\|\mathbf{v}_{\overline{S}}\|_1^2}{s}.$$

On the other hand, the right-hand side of (20) is bounded from above by

$$4\delta\|\mathbf{A}(\mathbf{v}_S+\mathbf{v}_{S'})\|_2\|\mathbf{A}\mathbf{v}\|_2 \le 4\delta\sqrt{1+\delta}\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2\|\mathbf{A}\mathbf{v}\|_2.$$

Combining the two bounds gives

$$(1-\delta^2)\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2^2-\delta^2\frac{\|\mathbf{v}_{\overline{S}}\|_1^2}{s} \le 2\sqrt{1+\delta}\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2\|\mathbf{A}\mathbf{v}\|_2,$$

which is equivalent to

$$\left(\sqrt{1-\delta^2}\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2-\frac{1}{\sqrt{1-\delta}}\|\mathbf{A}\mathbf{v}\|_2\right)^2 \le \delta^2\frac{\|\mathbf{v}_{\overline{S}}\|_1^2}{s}+\frac{1}{1-\delta}\|\mathbf{A}\mathbf{v}\|_2^2.$$

As a result, we obtain

$$
\begin{aligned}
\sqrt{1-\delta^2}\|\mathbf{v}_S+\mathbf{v}_{S'}\|_2 &\le \left(\delta^2\frac{\|\mathbf{v}_{\overline{S}}\|_1^2}{s}+\frac{1}{1-\delta}\|\mathbf{A}\mathbf{v}\|_2^2\right)^{1/2}+\frac{1}{\sqrt{1-\delta}}\|\mathbf{A}\mathbf{v}\|_2\\
&\le \delta\frac{\|\mathbf{v}_{\overline{S}}\|_1}{\sqrt{s}}+\frac{2}{\sqrt{1-\delta}}\|\mathbf{A}\mathbf{v}\|_2.
\end{aligned}
$$

It remains to use the fact that $\|\mathbf{v}_S\|_2 \le \|\mathbf{v}_S+\mathbf{v}_{S'}\|_2$ to arrive at

$$\|\mathbf{v}_S\|_2 \le \frac{\delta}{\sqrt{1-\delta^2}}\frac{\|\mathbf{v}_{\overline{S}}\|_1}{\sqrt{s}}+\frac{2}{(1-\delta)\sqrt{1+\delta}}\|\mathbf{A}\mathbf{v}\|_2,$$

which is the desired inequality with $\rho := \delta/\sqrt{1-\delta^2} < 1$ when $\delta < 1/\sqrt{2}$ and with $\tau := 2/((1-\delta)\sqrt{1+\delta}) < 5.25$. □

From Theorems 4 and 5, we now know that, under the condition $\delta_{2s} < 1/\sqrt{2}$, the stability and robustness estimate (18) holds. Similarly to OMP, we can also give a stability and robustness estimate with recovery error measured in ℓ_1 instead of ℓ_2. One way to achieve this is based on the following lemma.

Lemma 1. *If $\mathbf{x} \in \mathbb{C}^N$ is s-sparse and if $\mathbf{x}^\sharp = \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x})$ or $\mathbf{x}^\sharp = \Delta_{\mathrm{BP}}^{\eta}(\mathbf{A}\mathbf{x}+\mathbf{e})$ with $\|\mathbf{e}\|_2 \le \eta$, then $\mathbf{x}-\mathbf{x}^\sharp$ is effectively $4s$-sparse, meaning that*

$$\|\mathbf{x}-\mathbf{x}^\sharp\|_1 \le \sqrt{4s}\|\mathbf{x}-\mathbf{x}^\sharp\|_2.$$

More generally, for any vector $\mathbf{x} \in \mathbb{C}^N$ *and any index set* $S \subseteq [\![1:N]\!]$*, if* $\mathbf{x}^\sharp$ *is a minimizer of* $\|\mathbf{z}\|_1$ *subject to some constraint met by* $\mathbf{x}$*, then*

$$\|(\mathbf{x}-\mathbf{x}^\sharp)_{\overline{S}}\|_1 \le \|(\mathbf{x}-\mathbf{x}^\sharp)_S\|_1 + 2\|\mathbf{x}_{\overline{S}}\|_1.$$

Proof. To establish the second statement, we combine the fact that $\|\mathbf{x}^\sharp\|_1 \le \|\mathbf{x}\|_1$ with the observations that

$$\|\mathbf{x}^\sharp\|_1 = \|\mathbf{x}^\sharp_S\|_1 + \|\mathbf{x}^\sharp_{\overline{S}}\|_1 \ge \|\mathbf{x}^\sharp_S\|_1 + \|(\mathbf{x}-\mathbf{x}^\sharp)_{\overline{S}}\|_1 - \|\mathbf{x}_{\overline{S}}\|_1,$$
$$\|\mathbf{x}\|_1 = \|\mathbf{x}_S\|_1 + \|\mathbf{x}_{\overline{S}}\|_1 \le \|\mathbf{x}^\sharp_S\|_1 + \|(\mathbf{x}-\mathbf{x}^\sharp)_S\|_1 + \|\mathbf{x}_{\overline{S}}\|_1.$$

To prove the first statement, we write the second one as $\|(\mathbf{x}-\mathbf{x}^\sharp)_{\overline{S}}\|_1 \le \|(\mathbf{x}-\mathbf{x}^\sharp)_S\|_1$ when $\mathbf{x}$ is s-sparse and S is the support of $\mathbf{x}$. We then derive that

$$\|\mathbf{x}-\mathbf{x}^\sharp\|_1 \le 2\|(\mathbf{x}-\mathbf{x}^\sharp)_S\|_1 \le 2\sqrt{s}\|(\mathbf{x}-\mathbf{x}^\sharp)_S\|_2 \le \sqrt{4s}\|\mathbf{x}-\mathbf{x}^\sharp\|_2,$$

which is the desired result. □

We sum up the considerations about quadratically-constrained basis pursuit presented so far with the following statement about the recovery error in ℓ_p for $p=1$ and $p=2$. It is not hard to extend it to ℓ_p for any $p \in [1,2]$.

Theorem 6. *If a matrix* $\mathbf{A} \in \mathbb{C}^{m\times N}$ *has restricted isometry constant* $\delta_{2s} < 1/\sqrt{2}$*, then, for any* $\mathbf{x} \in \mathbb{C}^N$ *and any* $\mathbf{e} \in \mathbb{C}^m$ *with* $\|\mathbf{e}\|_2 \le \eta$*, a minimizer of the quadratically-constrained* ℓ_1*-minimization* (17) *approximates* $\mathbf{x}$ *with error*

$$\|\mathbf{x} - \Delta^\eta_{\mathrm{BP}}(\mathbf{A}\mathbf{x}+\mathbf{e})\|_1 \le C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\eta, \tag{21}$$
$$\|\mathbf{x} - \Delta^\eta_{\mathrm{BP}}(\mathbf{A}\mathbf{x}+\mathbf{e})\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\eta. \tag{22}$$

5.2 Quotient property

Although the result of Theorem 6 is quite satisfying, there is a minor issue one wishes to overcome, namely an estimation of the magnitude η of the measurement error is needed for (17) to be executed (in parallel, recall that typically iterative algorithms require an estimation of the sparsity level s to run). Could it happen that recovery via equality-constrained ℓ_1-minimization provides stability and robustness even in the presence of measurement error? The answer is affirmative, as we shall see. Precisely, using the notation

$$\Delta_{\mathrm{BP}}(\mathbf{y}) := \operatorname{argmin}\ \|\mathbf{z}\|_1 \qquad \text{subject to } \mathbf{A}\mathbf{z} = \mathbf{y}, \tag{23}$$

we target an estimate of the type (22), i.e.,

$$\|\mathbf{x} - \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x} + \mathbf{e})\|_2 \leq \frac{C}{\sqrt{s}} \sigma_s(\mathbf{x})_1 + D\|\mathbf{e}\|_2 \tag{24}$$

valid for all $\mathbf{x} \in \mathbb{C}^N$ and all $\mathbf{e} \in \mathbb{C}^m$ in the optimal regime of parameters $s \asymp s_*$, where

$$s_* := \frac{m}{\ln(eN/m)}.$$

Note that (24) holds for all $s \lesssim s_*$ as soon as it holds for $s \asymp s_*$. We also target, for $s \asymp s_*$, an estimate similar to (21) which measures recovery error in ℓ_1-norm, precisely

$$\|\mathbf{x} - \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x} + \mathbf{e})\|_1 \leq C\sigma_{s_*}(\mathbf{x})_1 + D\sqrt{s_*}\|\mathbf{e}\|_2. \tag{25}$$

We can see that the RNSP is still a necessary condition to guareantee (24) — as before, we make the choice $\mathbf{x} = \mathbf{v} \in \mathbb{C}^N$ and $\mathbf{e} = -\mathbf{A}\mathbf{v}$. Moreover, if we make the choice $\mathbf{x} = \mathbf{0}$ in (25), we obtain that, for any $\mathbf{e} \in \mathbb{C}^m$,

$$\|\Delta_{\mathrm{BP}}(\mathbf{e})\|_1 \leq D\sqrt{s_*}\|\mathbf{e}\|_2.$$

From the definition of $\Delta_{\mathrm{BP}}(\mathbf{e})$, this is equivalent to the following statement:

$$\text{for every } \mathbf{e} \in \mathbb{C}^m, \text{ there exists } \mathbf{u} \in \mathbb{C}^N \text{ with } \mathbf{A}\mathbf{u} = \mathbf{e} \text{ and } \|\mathbf{u}\|_1 \leq D\sqrt{s_*}\|\mathbf{e}\|_2.$$

This is called the quotient property (QP). We finish this subsection by proving that the RNSP and the QP are not only necessary conditions for the estimates (24)-(25) to hold, but they are also sufficient conditions. The argument given below simplifies the ones from [39, Section 11.2]. Of course, we should establish that matrices with the RNSP and QP actually do exist. Propitiously, Gaussian matrices with $N \geq 2m$ do exhibit the QP with high probability, as proved in [39, Theorem 11.19].

Theorem 7. *If a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies the QP and the RNSP of order cs_*, $s_* := m/\ln(eN/m)$, then for any $\mathbf{x} \in \mathbb{C}^N$ and any $\mathbf{e} \in \mathbb{C}^m$, a minimizer of the equality-constrained ℓ_1-minimization* (16) *applied to $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ approximates $\mathbf{x}$ with error*

$$\|\mathbf{x} - \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x} + \mathbf{e})\|_1 \leq C\sigma_{cs_*}(\mathbf{x})_1 + D\sqrt{s_*}\|\mathbf{e}\|_2,$$
$$\|\mathbf{x} - \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x} + \mathbf{e})\|_2 \leq \frac{C}{\sqrt{s_*}}\sigma_{cs_*}(\mathbf{x})_1 + D\|\mathbf{e}\|_2.$$

Proof. By the quotient property, there exists $\mathbf{u} \in \mathbb{C}^N$ such that

$$\mathbf{A}\mathbf{u} = \mathbf{e} \qquad \text{and} \qquad \|\mathbf{u}\|_1 \leq D\sqrt{s_*}\|\mathbf{e}\|_2.$$

We notice that $\mathbf{x} + \mathbf{u}$ satisfies $\mathbf{A}(\mathbf{x} + \mathbf{u}) = \mathbf{A}\mathbf{x} + \mathbf{e} = \mathbf{y}$. Therefore, by definition of $\mathbf{x}^\sharp := \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x} + \mathbf{e})$, we have

$$\|\mathbf{x}^\sharp\|_1 \leq \|\mathbf{x} + \mathbf{u}\|_1 \leq \|\mathbf{x}\|_1 + \|\mathbf{u}\|_1 \leq \|\mathbf{x}\|_1 + D\sqrt{s_*}\|\mathbf{e}\|_2.$$

We also notice that $\mathbf{A}(\mathbf{x}^\sharp - \mathbf{x}) = \mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{e}$. By Theorem 4, the RNSP implies

$$\|\mathbf{x}-\mathbf{x}^\sharp\|_2 \le \frac{C}{\sqrt{cs_*}}\Big(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}\|_1 + 2\sigma_{cs_*}(\mathbf{x})_1\Big) + D\|\mathbf{A}(\mathbf{x}^\sharp-\mathbf{x})\|_2$$
$$\le \frac{C}{\sqrt{s_*}}(D\sqrt{s_*}\|\mathbf{e}\|_2 + 2\sigma_{cs_*}(\mathbf{x})_1) + D\|\mathbf{e}\|_2 \le \frac{C}{\sqrt{s_*}}\sigma_{cs_*}(\mathbf{x})_1 + D\|\mathbf{e}\|_2.$$

This is the result for $p=2$. To obtain the result for $p=1$, we make use of Lemma 1 applied to $\mathbf{x}+\mathbf{u}$ and to an index set S of cs_* largest absolute entries of $\mathbf{x}$. We obtain

$$\|\mathbf{x}-\mathbf{x}^\sharp\|_1 \le \|\mathbf{x}+\mathbf{u}-\mathbf{x}^\sharp\|_1 + \|\mathbf{u}\|_1 \le 2\|(\mathbf{x}+\mathbf{u}-\mathbf{x}^\sharp)_S\|_1 + 2\|\mathbf{x}_{\overline{S}}\|_1 + \|\mathbf{u}\|_1$$
$$\le 2\|(\mathbf{x}-\mathbf{x}^\sharp)_S\|_1 + 2\sigma_{cs_*}(\mathbf{x})_1 + 3\|\mathbf{u}\|_1$$
$$\le 2\sqrt{cs_*}\|\mathbf{x}-\mathbf{x}^\sharp\|_2 + 2\sigma_{cs_*}(\mathbf{x})_1 + 3D\sqrt{s_*}\|\mathbf{e}\|_2,$$

so that the bound for $p=1$ follows from the bound for $p=2$ derived above. □

5.3 Regularizations

We have not discussed algorithms to perform the ℓ_1-minimizations (16) and (17). There are efficient algorithms[5] to perform this task exactly, but it is also popular, especially in Statistics, to solve the related unconstrained optimization problem

$$\underset{\mathbf{z}}{\text{minimize}}\ \|\mathbf{z}\|_1 + \lambda\|\mathbf{A}\mathbf{z}-\mathbf{y}\|_2^2. \tag{26}$$

Intuitively, choosing a large parameter λ imposes $\|\mathbf{A}\mathbf{z}-\mathbf{y}\|_2$ to be small, so a solution of (26) is potentially also a solution of (17). To be a little more precise, (17) and (26) are equivalent in the sense that a minimizer of (17) is also a minimizer of (26) for some choice of parameter λ, and conversely a minimizer of (26) is also a minimizer of (17) for some choice of parameter η (see e.g [39, Proposition 3.2] for details). The catch is that the parameter in one program cannot be chosen beforehand as it depends on the minimizer of the other program. Problem (17) is also equivalent (with the same meaning of equivalence) to the optimization problem

$$\underset{\mathbf{z}}{\text{minimize}}\ \|\mathbf{A}\mathbf{z}-\mathbf{y}\|_2 \qquad \text{subject to } \|\mathbf{z}\|_1 \le \tau. \tag{27}$$

Problem (27) is called LASSO and Problem (26) is called basis pursuit denoising, but there is often some confusion in the terminology between (17), (26), and (27).

Instead of (26), I personally prefer a regularized problem where the ℓ_1-norm is squared, namely

$$\underset{\mathbf{z}}{\text{minimize}}\ \|\mathbf{z}\|_1^2 + \lambda\|\mathbf{A}\mathbf{z}-\mathbf{y}\|_2^2. \tag{28}$$

Apart from the seemingly more natural homogeneity in $\mathbf{z}$, the reason for my preference comes from the fact that (28) can be transformed into a nonnegative

[5] For instance, ℓ_1-MAGIC, NESTA, and YALL1 are freely available online.

least squares problem (see [33] for a theoretical analysis[6]). For a quick explanation, consider the situation where the sparse vector $\mathbf{x} \in \mathbb{R}^N$ to recover is nonnegative (it is a problem in metagenomics, where $\mathbf{x}$ is a vector of bacterial concentrations, that motivated this scenario — see [47, 48] for the end-results). In such a case, we readily see that the optimization problem

$$\underset{\mathbf{z}}{\text{minimize}} \; \|\mathbf{z}\|_1^2 + \lambda \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2^2 \qquad \text{subject to } \mathbf{z} \geq \mathbf{0}$$

can be recast as the nonnegative least squares problem

$$\underset{\mathbf{z}}{\text{minimize}} \; \left\| \left[\frac{\mathbf{1}}{\sqrt{\lambda}\mathbf{A}} \right] \mathbf{z} - \left[\frac{0}{\sqrt{\lambda}\mathbf{y}} \right] \right\|_2^2 \qquad \text{subject to } \mathbf{z} \geq \mathbf{0},$$

where $\mathbf{1}$ is the row-vector of all ones. A classical algorithm for nonnegative least squares, due to Lawson and Hanson, dates back to 1973 with the first edition of [50], see Chapter 23. It turns out to be very well suited for sparse recovery because it is an 'active set' method. In order to give some details, let us change notation and consider the problem

$$\underset{\mathbf{z}}{\text{minimize}} \, \|\widetilde{\mathbf{A}}\mathbf{z} - \widetilde{\mathbf{y}}\|_2^2 \qquad \text{subject to } \mathbf{z} \geq \mathbf{0}.$$

Lawson–Hanson algorithm proceeds iteratively in a manner reminiscent of OMP. In fact, the similarity is striking by writing the algorithm as

$$S^{n+1} := S^n \cup \{j^{n+1}\}, \qquad \text{where } j^{n+1} := \underset{j \in [\![1:N]\!]}{\text{argmin}} \left(\widetilde{\mathbf{A}}^*(\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{x}^n)\right)_j \tag{LH$_1$}$$

$$\mathbf{x}^{n+1} := \underset{\mathbf{z} \in \mathbb{R}^N}{\text{argmin}} \, \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{z}\|_2 \;\; \text{subject to } \text{supp}(\mathbf{z}) \subseteq S^{n+1} \tag{LH$_2$}$$

& enter an inner loop to adjust $\mathbf{x}^{n+1}$ until all its entries are nonnegative.

We see that (LH$_2$) is the same as (OMP$_2$), save for the presence of the inner loop (which in practice is rarely entered), and that (LH$_1$) is the same as (OMP$_1$), save for the absence of the absolute value (which guarantees the creation of one nonnegative entry, in view of $x^{n+1}_{j^{n+1}} = (\widetilde{\mathbf{A}}^*(\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{x}^n))_{j^{n+1}} / \text{dist}(\widetilde{\mathbf{a}}_{j^{n+1}}, \text{span}\{\widetilde{\mathbf{a}}_{j^1}, \ldots, \widetilde{\mathbf{a}}_{j^n}\})$, see [30] for details).

6 Gelfand Widths Estimates: Byproduct of Compressive Sensing

It is time to come back to the connection between Gelfand widths and compressive sensing, now that we are more informed about the compressive sensing theory from Sections 3, 4, and 5. The intent is to turn the story around by showing that the compressive sensing theory is self-sufficient and that it provides Gelfand width estimates as side results. Before doind so, I promised in Subsection 2.4 to replace

[6] See also the MATLAB reproducible for a numerical illustration.

the impractical recovery map (4) by the recovery map (23) while maintaining the message of Subsection 2.3, namely that the upper estimate for the Gelfand width provides a measurement matrix $\mathbf{A}$ such that the pair $(\mathbf{A}, \Delta_{\mathrm{BP}})$ is stable of order $s \asymp m/\ln(eN/m)$. This can be easily seen by replacing $\widetilde{\mathbf{x}}$ by $\mathbf{x}^\sharp = \Delta_{\mathrm{BP}}(\mathbf{A}\mathbf{x})$ until (5), where we would call upon Lemma 1 to derive

$$\begin{aligned}\|\mathbf{x} - \mathbf{x}^\sharp\|_2 &\le \frac{1}{\sqrt{8s}} \|\mathbf{x} - \mathbf{x}^\sharp\|_1 = \frac{1}{\sqrt{8s}} \left(\|(\mathbf{x} - \mathbf{x}^\sharp)_{\overline{S}}\|_1 + \|(\mathbf{x} - \mathbf{x}^\sharp)_S\|_1 \right) \\ &\le \frac{1}{\sqrt{8s}} \left(2\|(\mathbf{x} - \mathbf{x}^\sharp)_S\|_1 + 2\sigma_s(\mathbf{x})_1 \right) \le \frac{1}{\sqrt{2}} \|\mathbf{x} - \mathbf{x}^\sharp\|_2 + \frac{1}{\sqrt{2s}} \sigma_s(\mathbf{x})_1,\end{aligned}$$

so a rearrangement provides the desired inequality (3) with constant $C = \sqrt{2} + 1$.

After this digression, let us come back to our main goal of proving the upper and lower estimates for the Gelfand width of the ℓ_1-ball B_1^N relative to ℓ_2^N using compressive sensing tools only and highlighting that Gelfand widths could be left out of the compressive sensing theory if so desired. We return to the real setting for the rest of this section.

6.1 Derivation of the upper estimate

Bounding $d^m(B_1^N, \ell_2^N)$ from above requires the RIP and the sort-and-split technique. Theorem 1 guarantees that we can find a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ with restricted isometry constant $\delta_{2s} < 4/5$, say, for $s \asymp m/ln(eN/m)$. Since the space $L^m := \ker \mathbf{A}$ has codimension at most m, the definition of the Gelfand width yields

$$d^m(B_1^N, \ell_2^N) \le \sup_{\mathbf{v} \in \ker \mathbf{A}} \frac{\|\mathbf{v}\|_2}{\|\mathbf{v}\|_1}.$$

Then, given $\mathbf{v} \in \ker \mathbf{A}$, we notice that $\mathbf{A}(\mathbf{v}_S) = -\mathbf{A}(\mathbf{v}_{\overline{S}})$ with S denoting an index set of s largest absolute entries of $\mathbf{v}$. Thanks to the RIP and the sort-and-split technique, we obtain (as in (13))

$$\|\mathbf{v}_S\|_2 \le \frac{1}{\sqrt{1-\delta}} \|\mathbf{A}\mathbf{v}_S\|_2 = \frac{1}{\sqrt{1-\delta}} \|\mathbf{A}\mathbf{v}_{\overline{S}}\|_2 \le \frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} \frac{\|\mathbf{v}\|_1}{\sqrt{s}}.$$

Moreover, Stechkin bound gives $\|\mathbf{v}_{\overline{S}}\|_2 \le \|\mathbf{v}\|_1/\sqrt{s}$. It follows that

$$\|\mathbf{v}\|_2 \le \|\mathbf{v}_S\|_2 + \|\mathbf{v}_{\overline{S}}\|_2 \le \left(\frac{\sqrt{1+\delta}}{\sqrt{1-\delta}} + 1 \right) \frac{\|\mathbf{v}\|_1}{\sqrt{s}} \le 4 \frac{\|\mathbf{v}\|_1}{\sqrt{s}}.$$

Taking $s \asymp m/ln(eN/m)$ into account, we arrive at

$$d^m(B_1^N, \ell_2^N) \le C \sqrt{\frac{\ln(eN/m)}{m}}.$$

Since $d^m(B_1^N, \ell_2^N) \le 1$ (the ℓ_2-ball contains the ℓ_1-ball), the upper bound is proved.

6.2 Derivation of the lower estimate

Subsection 2.2 can give the impression that the lower estimate for $d^m(B_1^N, \ell_2^N)$ is essential in the compressive sensing theory, as it enables to prove that the number of measurements $m \asymp s\ln(eN/s)$ is optimal. The next subsection shows that this dependence on the lower estimate can be lifted, but for now we want to stress that the compressive sensing theory provides tools, and perhaps more importantly intuition, to retrieve this lower estimate. Basically, the two insights:

(i) small $d^m(B_1^N, \ell_2^N)$ yields exact s-sparse recovery via ℓ_1-minimization for 'large' s,
(ii) exact s-sparse recovery via ℓ_1-minimization is only possible for 'moderate' s,

inform us that the Gelfand width cannot be too small, i.e., they establish a lower bound. A formal argument can be found in [38], where the ℓ_1-ball B_1^N is replaced by the ℓ_p-balls B_p^N for $0 < p \le 1$. Here, we settle for an informal justification of the two insights (i) and (ii).

For (i), we pick an optimal subspace L^m in the definition of the Gelfand width and a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ with $\ker \mathbf{A} = L^m$. For any $\mathbf{v} \in \ker \mathbf{A}$, we have $\|\mathbf{v}\|_2 \le \omega \|\mathbf{v}\|_1$, where we wrote $\omega := d^m(B_1^N, \ell_2^N)$ for short. Then, for any index set S of size s,

$$\|\mathbf{v}_S\|_1 \le \sqrt{s}\|\mathbf{v}_S\|_2 \le \sqrt{s}\|\mathbf{v}\|_2 \le \sqrt{s}\omega\|\mathbf{v}\|_1 = \sqrt{s}\omega\|\mathbf{v}_S\|_1 + \sqrt{s}\omega\|\mathbf{v}_{\overline{S}}\|_1.$$

Choosing s such that $\sqrt{s}\omega = 1/2$, this reduces to the NSP (b), which is equivalent to exact s-sparse recovery as stated in (a).

For (ii), the key ingredient is a combinatorial lemma that appears independently in many places (we have seen it in [13, 41, 53, 55, 56]) and we refer to [39, Lemma 10.12] for a proof.

Lemma 2. *One can find a large number*

$$n \ge \left(\frac{N}{4s}\right)^{s/2}$$

of subsets $S_1, \ldots, S_n$ of $[\![1 : N]\!]$ with size s that are almost disjoint in the sense that

$$\mathrm{card}(S_i \cap S_j) < \frac{s}{2} \qquad \textit{whenever } i \ne j.$$

Now, consider the vectors $\mathbf{x}^1, \ldots, \mathbf{x}^n \in \mathbb{R}^N$ supported on $S_1, \ldots, S_n$, respectively, that equal $1/s$ on their supports. We have $\|\mathbf{x}^i\|_1 = 1$ for all $i \in [\![1 : n]\!]$ and, thanks to Lemma 2, $\|\mathbf{x}^i - \mathbf{x}^j\|_1 \ge 1$ for all distinct $i, j \in [\![1 : n]\!]$. Since the $\mathbf{x}^i$ and $\mathbf{x}^i - \mathbf{x}^j$ are sparse vectors, they are ℓ_1-minimizers, so that $\|[\mathbf{x}^i]\| = 1$ and $\|[\mathbf{x}^i] - [\mathbf{x}^j]\| \ge 1$ in the quotient space $\ell_1^N / \ker \mathbf{A}$ equipped with the norm $\|[\mathbf{z}]\| := \inf\{\|\mathbf{z}'\|_1 : \mathbf{A}\mathbf{z}' = \mathbf{A}\mathbf{z}\}$. This supplies a set of $n \ge (N/4s)^{s/2}$ points on the unit sphere of the m-dimensional space $\ell_1^N / \ker \mathbf{A}$ which are all separated by a distance at least 1. Since such a set cannot be too large (see e.g. [39, Proposition C.3]), this imposes the desired restriction on the size of s.

6.3 Distantiation from the Gelfand width

Sections 2 and 6 described a close connection between the Gelfand width of ℓ_1-balls and compressive sensing and how one topic can be seen as a ramification of the other and vice versa. As an epilogue, I want to highlight that the detour through Gelfand widths is no longer necessary to build the compressive sensing theory (the shortcut presented below was discovered after studying Gelfand widths, though). Specifically, given $p \in [1,2]$, let us show that a pair $(\mathbf{A},\Delta)$ providing the stability estimate

$$\|\mathbf{x}-\Delta(\mathbf{A}\mathbf{x})\|_p \leq \frac{C}{s^{1-1/p}}\sigma_s(\mathbf{x})_1$$

valid for all $\mathbf{x} \in \mathbb{C}^N$ can only exist if $m \geq Cs\ln(eN/s)$. Note that, for $p=1$, the result could not be established using Gelfand widths. We still consider the s-sparse vectors $\mathbf{x}^1,\ldots,\mathbf{x}^n$ built from the index sets $S_1,\ldots,S_n$ of Lemma 2. With $\rho := 1/(4(C+1))$, we consider the subsets of $\mathbb{R}^m$ defined by

$$E_i := \mathbf{A}\left(\mathbf{x}^i + \rho\left(B_1^N \cap \frac{1}{s^{1-1/p}}B_p^N\right)\right), \qquad i \in [\![1:n]\!].$$

We claim that $E_1,\ldots,E_n$ are all disjoint. Indeed, if there was an element in $E_i \cap E_j$, $i \neq j$, say $\mathbf{y} = \mathbf{A}(\mathbf{x}^i+\rho\mathbf{z}) = \mathbf{A}(\mathbf{x}^j+\rho\mathbf{z}')$ with $\mathbf{z},\mathbf{z}' \in B_1^N \cap (1/s^{1-1/p})B_p^N$, then

$$\begin{aligned}
\|\mathbf{x}^i-\mathbf{x}^j\|_p &\leq \|\mathbf{x}^i+\rho\mathbf{z}-\Delta(\mathbf{y})\|_p + \|\mathbf{x}^j+\rho\mathbf{z}'-\Delta(\mathbf{y})\|_p + \|\rho\mathbf{z}-\rho\mathbf{z}'\|_p\\
&\leq \frac{C}{s^{1-1/p}}\sigma_s(\mathbf{x}^i+\rho\mathbf{z})_1 + \frac{C}{s^{1-1/p}}\sigma_s(\mathbf{x}^j+\rho\mathbf{z}')_1 + \rho\|\mathbf{z}-\mathbf{z}'\|_p\\
&\leq \frac{C}{s^{1-1/p}}\rho\left(\|\mathbf{z}\|_1+\|\mathbf{z}'\|_1\right) + \rho\left(\|\mathbf{z}\|_p+\|\mathbf{z}'\|_p\right) \leq \frac{C}{s^{1-1/p}}2\rho + \rho\frac{2}{s^{1-1/p}}\\
&\leq \frac{2(C+1)\rho}{s^{1-1/p}} = \frac{1}{2s^{1-1/p}}.
\end{aligned}$$

This contradicts the fact that $\|\mathbf{x}^i-\mathbf{x}^j\|_p \geq (1/s)\mathrm{card}(S_i\Delta S_j)^{1/p} \geq 1/s^{1-1/p}$. We now also claim that, because all the $\mathbf{x}^i$ belong to $B_1^N \cap (1/s^{1-1/p})B_p^N$, all the sets E_i are contained in the ball $(1+\rho)\mathbf{A}\left(B_1^N \cap (1/s^{1-1/p})B_p^N\right)$. From these two claims, we derive that $\sum_{i=1}^n \mathrm{Vol(E_i)} = \mathrm{Vol}\left(\cup_{i=1}^n E_i\right) \leq \mathrm{Vol}\left((1+\rho)\mathbf{A}\left(B_1^N \cap (1/s^{1-1/p})B_p^N\right)\right)$. With $\mathscr{V} := \mathrm{Vol}\left(\mathbf{A}\left(B_1^N \cap (1/s^{1-1/p})B_p^N\right)\right)$ and $r := \mathrm{rank}(\mathbf{A}) \leq m$, this inequality reads $n\rho^r\mathscr{V} \leq (1+\rho)^r\mathscr{V}$. This imples that

$$\left(\frac{N}{4s}\right)^{s/2} \leq n \leq \left(1+\frac{1}{\rho}\right)^r \leq \left(1+\frac{1}{\rho}\right)^m, \quad \text{hence} \quad \frac{s}{2}\ln\left(\frac{N}{4s}\right) \leq m\ln\left(1+\frac{1}{\rho}\right).$$

We have obtained $m \geq cs\ln(N/4s)$, which is equivalent, up to changing the constant, to the desired inequality $m \geq cs\ln(eN/s)$, see [39, Lemma C.6].

7 Other Matrices Suitable for Compressive Sensing

The measurement matrices put forward so far have been realizations of Gaussian matrices. We now discuss other types of matrices that can be used in compressive sensing. Again, they will be realizations of some random matrices.

7.1 Subgaussian matrices

A mean-zero random variable ξ is called subgaussian if its tails are dominated by the tails of a Gaussian random variable, which is equivalent to the existence of constants $\alpha, \beta > 0$ such that

$$\mathbb{P}(|\xi| > t) \leq \alpha \exp(-\beta t^2) \qquad \text{for all } t > 0.$$

Another equivalent condition involving the moment generating function of ξ reads

$$\mathbb{E}\,(\exp(\theta\xi)) \leq \exp(\gamma\theta^2) \qquad \text{for all } \theta \in \mathbb{R}.$$

Gaussian random variables are subgaussian, of course. Other important examples of subgaussian random variables include random variables uniformly distributed on an interval $[-c, c]$ and Rademacher random variables taking values -1 and $+1$ with probability $1/2$ each. By a subgaussian random matrix, we mean a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ populated by independent (but not necessarily identically distributed) subgaussian random variables with a common parameter γ and normalized to have variance $1/m$. Such matrices are suitable compressive sensing matrices because they possess, with overwhelming probability, the RIP in the optimal regime $m \asymp s \ln(eN/s)$. The argument is almost identical to the proof of Theorem 1: more work is required to establish the concentration inequality, but the the covering argument and the union bound are similar. See [39, Section 9.1] for details.

7.2 Subexponential matrices

A mean-zero random variable is called subexponential if there exist constants $\alpha, \beta > 0$ such that

$$\mathbb{P}(|\xi| > t) \leq \alpha \exp(-\beta t) \qquad \text{for all } t > 0.$$

For instance, Laplace random variables are subexponential. By a subexponential random matrix, we mean a matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ populated by independent (but not necessarily identically distributed) subexponential random variables with common parameters α, β and normalized to have variance $1/m$. Such matrices do not have the RIP in the optimal regime, since $m \geq C s \ln(eN/s)^2$ is required, as shown in [1].

But these matrices still satisfiy the RNSP when $m \asymp s\ln(eN/s)$, as shown in [36], hence they allow for stable and robust sparse recovery via quadratically-constrained basis pursuit. In fact, they also allow for stable and robust sparse recovery via other algorithms, as shown in [35]. This can be seen as a consequence of the modified RIP

$$(1-\delta)\, /\!\!/ \mathbf{z} /\!\!/ \leq \|\mathbf{A}\mathbf{z}\|_1 \leq (1+\delta)\, /\!\!/ \mathbf{z} /\!\!/ \qquad \text{for all } s\text{-sparse vectors } \mathbf{z} \in \mathbb{R}^N,$$

which features the ℓ_1-norm as the inner norm. The outer norm, whose precise expression is somewhat irrelevant, depends a priori on the probability distributions of the entries, but it satisfies, for some absolute constants $c, C > 0$,

$$c\sqrt{m}\|\mathbf{z}\|_2 \leq /\!\!/ \mathbf{z} /\!\!/ \leq C\sqrt{m}\|\mathbf{z}\|_2 \qquad \text{for all vectors } \mathbf{z} \in \mathbb{C}^N.$$

In particular, the result applies to Weibull matrices with parameters $r \in [1,\infty]$, whose entries are independent symmetric random variables obeying

$$\mathbb{P}(|\xi| > t) = \exp\left(-\left(\sqrt{\Gamma(1+2/r)m}\,t\right)^r\right).$$

If we further make the restriction $r \in [1,2]$, then, with overwhelming probability, these matrices also satisfy the QP, as shown in [31]. Therefore, they will allow for stable and robust sparse recovery via equality-constrained basis pursuit even in the presence of measurement error.

We close this subsection by mentioning that the NSP can also be obtained, with high probability, for a wider random matrix class than subexponential matrices. Indeed, [51] establishes the NSP under rather weak conditions on the moments of the entries of the random matrices.

7.3 Fourier submatrices and affiliates

Consider the matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ with entries

$$A_{k,\ell} = \exp(\imath 2\pi \ell t_k), \qquad k \in [\![1:m]\!], \quad \ell \in [\![1:N]\!],$$

where the points $t_1,\ldots,t_m \in [0,1]$ are independently chosen uniformly at random from the grid $\{0, 1/N, \ldots, (N-1)/N\}$ or from the interval $[0,1]$. These two cases are of practical importance: the first case corresponds to randomly selecting rows of the discrete $N \times N$ Fourier matrix, so it provides fast matrix-vector multiplications exploited in the recovery algorithms; in the second case, samples of a trigonometric polynomial $f(t) = \sum_{\ell=0}^{N-1} x_\ell \exp(\imath 2\pi \ell t)$ at random points $t_1,\ldots,t_m \in [0,1]$ produce a vector $\mathbf{A}\mathbf{x} \in \mathbb{C}^m$, from where f can be recovered provided the coefficient vector $\mathbf{x} \in \mathbb{C}^N$ is s-sparse and $m \asymp s\ln^4(N)$ — this drastically improves on $m \asymp N$ predicted from the Shannon–Nyquist theorem traditionally invoked in signal processing. It can indeed be proved that both types of random matrices possess the RIP of order s when $m \asymp s\ln^4(N)$. This remains true for the random selection of rows from

Hadamard matrices, and more generally for random sampling matrices associated with bounded orthonormal systems (BOS). These matrices have entries $\Phi_\ell(t_k)$, $k \in [\![1:m]\!]$, $\ell \in [\![1:N]\!]$, where the functions $\Phi_1, \ldots, \Phi_N$ are uniformly bounded on a domain $\mathscr{D}$ by some absolute constant K and form an orthonormal system for the inner product associated to a probability measure ν on $\mathscr{D}$, and where the points $t_1, \ldots, t_m \in \mathscr{D}$ are independently chosen at random according to ν. We refer to [39, Chapter 12] for details. The power of the logarithm factor in the number of required measurements is improvable, with $m \asymp s \ln^2(s) \ln(N)$ seemingly the best result so far (see [22]). It is conceivable that the logarithmic factors may be further reduced, but not below $\ln(N)$ (see [39, Section 12.2]), thus not down to $\ln(eN/s)$ as is the case for subgaussian matrices.

7.4 Adjacency matrices of lossless expanders

Expander graphs, and their bipartite counterpart lossless expanders, are very popular objects in Computer Science. They have made their appearance in compressive sensing via [6] and constitute now an important part of the theory, too. In short, an (s,d,θ)-lossless expander is a bipartite graph with left vertices indexed by $[\![1:N]\!]$ and right vertices indexed by $[\![1:m]\!]$ such that each left vertex is connected to d right vertices and that

$$\operatorname{card}(R(S)) \geq (1-\theta)\, d \operatorname{card}(S)$$

for every set S of left vertices of size at most s, with $R(S)$ denoting the set of right vertices connected to S. Using probabilistic arguments, it can be shown that (s,d,θ)-lossless expanders exist with $d \asymp \theta^{-1} \ln(eN/s)$ and $m \asymp \theta^{-2} s \ln(eN/s)$. The adjacency matrix of such a graph is the matrix $\mathbf{A} \in \{0,1\}^{m \times N}$ defined for $i \in [\![1:m]\!]$ and $j \in [\![1:N]\!]$ by

$$A_{i,j} = \begin{cases} 1 & \text{if there is an edge from } j \text{ to } i, \\ 0 & \text{otherwise.} \end{cases}$$

Although these matrices (properly normalized) do not satisfy the RIP, they still allow for stable and robust sparse recovery (in the weak sense of footnote 3) via basis pursuit or via an adpated iterative thresholding algorithm. All the details can be found in [39, Chapter 13].

8 Nonstandard Compressive Sensing

The name compressive sensing originates from the possibility of simultaneously acquiring (sensing) and compressing high-dimensional vectors $\mathbf{x} \in \mathbb{C}^N$ by taking only few linear measurements $y_1 = \langle \mathbf{a}_1, \mathbf{x} \rangle, \ldots, y_m = \langle \mathbf{a}_m, \mathbf{x} \rangle$ (with the understanding, of course, that efficient decompression procedures are also provided). The sparsity of the vectors of interest is what made this specific task feasible. But any task that exploits some structure of high-dimensional objects of interest to simultaneously acquire and compress them by means of few observations (while providing efficient decompression procedures) are also to be put under the umbrella of compressive sensing. We present below several extensions of the standard compressive sensing problem that fall in this category.

8.1 Low-rank recovery

The standard compressive sensing problem can sometimes be referred to as sparse recovery. The low-rank recovery problem is very similar: s-sparse vectors $\mathbf{x} \in \mathbb{C}^N$ are replaced by rank-r matrices $\mathbf{X} \in \mathbb{C}^{n_1 \times n_2}$ — for simplicity of exposition, we assume here that $n_1 = n_2 =: n$. The matrices are still acquired by means of m linear measurements organized in the vector $\mathbf{y} = \mathscr{A}(\mathbf{X}) \in \mathbb{C}^m$, where $\mathscr{A} : \mathbb{C}^{n \times n} \to \mathbb{C}^m$ is a linear map. The minimal number of measurements for stable and robust rank-r recovery is $m \asymp rn$ — there is no logarithmic factor. The theoretical explanation follows a route almost identical to the sparse recovery route, as outlined below. A popular optimization-based recovery procedure consists in solving

$$\underset{\mathbf{Z} \in \mathbb{C}^{n \times n}}{\text{minimize}} \ \|\mathbf{Z}\|_* \qquad \text{subject to } \|\mathscr{A}(\mathbf{Z}) - \mathbf{y}\|_2 \le \eta,$$

where the nuclear norm is the sum of the singular values $\sigma_1(\mathbf{Z}) \ge \cdots \ge \sigma_n(\mathbf{Z}) \ge 0$, i.e.,

$$\|\mathbf{Z}\|_* = \sum_{k=1}^{n} \sigma_k(\mathbf{Z}). \tag{29}$$

An analogy with basis pursuit is offered by the facts that a matrix has rank at most r if and only if its vector of singular values is r-sparse and that the nuclear norm is just the ℓ_1-norm of the vector of singular values. The problem (29) can be solved by semidefinite programming, thanks to the observation that

$$\|\mathbf{Z}\|_* = \inf \left\{ \operatorname{tr}(\mathbf{M}) + \operatorname{tr}(\mathbf{N}), \ \begin{bmatrix} \mathbf{M} & \mathbf{Z} \\ \mathbf{Z}^* & \mathbf{N} \end{bmatrix} \succeq \mathbf{0} \right\}. \tag{30}$$

There is a null space property which is equivalent to the success of exact rank-r recovery via nuclear norm minimization (see [39, Theorem 4.40]) and a robust version that guarantees stability and robustness of the recovery (see [39, Exercise 4.20]). With constants $0 < \rho < 1$ and $\tau > 0$, it reads:

$$\text{for all } \mathbf{M} \in \mathbb{C}^{n \times n}, \qquad \left[\sum_{k=1}^{r} \sigma_k(\mathbf{M})^2 \right]^{1/2} \le \frac{\rho}{\sqrt{r}} \sum_{k=r+1}^{n} \sigma_k(\mathbf{M}) + \tau \|\mathscr{A}(\mathbf{M})\|_2. \tag{31}$$

This property is implied by a matrix version of the restricted isometry property, which reads

$$(1-\delta)\|\mathbf{Z}\|_F^2 \le \|\mathscr{A}(\mathbf{Z})\|_2^2 \le (1+\delta)\|\mathbf{Z}\|_F^2 \qquad \text{for all } \mathbf{Z} \in \mathbb{C}^{n\times n} \text{ with } \operatorname{rank}(\mathbf{Z}) \le r.$$

As in the vector case, we write δ_r for the smallest such constant δ. Given a linear map $\mathscr{A} : \mathbf{M} \in \mathbb{C}^{n\times n} \mapsto \left(\sum_{k,\ell=1}^n \mathscr{A}_{i,k,\ell} M_{k,\ell}\right)_{i=1}^m \in \mathbb{C}^m$ where the $\mathscr{A}_{i,k,\ell}$ are independent Gaussian random variables with mean zero and variance $1/m$, it can be shown that $\delta_r < \delta_*$ with overwhelming probability provided that $m \ge C\delta_*^{-2} rn$ (see [39, Exercise 9.12]). The logarithmic factor disappears because a union bound is unnecessary to estimate the covering number of the set $\{\mathbf{Z} \in \mathbb{C}^{n\times n} : \|\mathbf{Z}\|_F = 1, \operatorname{rank}(\mathbf{Z}) \le r\}$. Finally, seeing why the restricted isometry condition $\delta_{2r} < 1/\sqrt{2}$ (or any sufficient condition $\delta_{\kappa r} < \delta_*$ from the vector case) implies the matrix version of the RNSP requires almost no work. Indeed, given $\mathbf{M} \in \mathbb{C}^{n\times n}$, we consider the singular value decomposition $\mathbf{M} = \mathbf{U}\operatorname{diag}[\sigma(\mathbf{M})]\mathbf{V}^*$, and we define a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ via its action

$$\mathbf{z} \in \mathbb{C}^n \mapsto \mathscr{A}(\mathbf{U}\operatorname{diag}[\mathbf{z}]\mathbf{V}^*) \in \mathbb{C}^m.$$

This matrix obeys the restricted isometry condition, because, for every $2r$-sparse $\mathbf{z} \in \mathbb{C}^n$, the matrix $\mathbf{U}\operatorname{diag}[\mathbf{z}]\mathbf{V}^*$ has rank at most $2r$, hence

$$\begin{aligned}\big|\|\mathbf{A}(\mathbf{z})\|_2^2 - \|\mathbf{z}\|_2^2\big| &= \big|\|\mathscr{A}(\mathbf{U}\operatorname{diag}(\mathbf{z})\mathbf{V}^*)\|_2^2 - \|\mathbf{U}\operatorname{diag}[\mathbf{z}]\mathbf{V}^*\|_F^2\big| \le \delta_{2r}\|\mathbf{U}\operatorname{diag}[\mathbf{z}]\mathbf{V}^*\|_F^2 \\ &< \delta_*\|\mathbf{z}\|_2^2.\end{aligned}$$

Therefore, the matrix $\mathbf{A}$ possesses the RNSP of order r. In particular, writing it for the vector $\sigma(\mathbf{M})$ and the index set $S = \{1,\dots,r\}$ reduces to (31).

8.2 Low-rank matrix completion

Low-rank matrix completion is a particular instance of low-rank recovery where the linear measurements are restricted in nature. Precisely, what is observed consists of few entries of the matrix. It can be proved that $m \asymp rn$ entries observed at random suffice to efficiently recover, with high probability, ‘most’ rank-r matrices. Note that the recovery of all rank-r matrices is impossible, as the zero matrix and the rank-1 matrix with only one nonzero entry cannot be distinguished from a typical set of observations. For details, see [18] where theoretical results were first derived, [42, 61] where such results were improved and simplified, [17] where robustness to observation errors was taken into account, and [7] where uniform recovery results were obtained.

8.3 Sparse phaseless recovery

The sparse phaseless recovery scenario is very close to the plain sparse recovery scenario, except that the linear measurements $\langle \mathbf{a}_i, \mathbf{x} \rangle$ lose their phase, so only their magnitude is available. In other words, sparse vectors $\mathbf{x} \in \mathbb{C}^N$ are acquired through the nonlinear measurements $y_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2$ and one wishes to recover them (up to a phase factor). The nonsparse version of the problem has been around for a long time. The article [19] was the first one to propose a recovery strategy based on convex optimization. The following 'lifting trick' is a central part of the argument:

$$y_i = |\langle \mathbf{a}_i, \mathbf{x} \rangle|^2 = \langle \mathbf{x}, \mathbf{a}_i \rangle \langle \mathbf{a}_i, \mathbf{x} \rangle = \mathbf{a}_i^* \mathbf{x}\mathbf{x}^* \mathbf{a}_i = \mathrm{tr}(\mathbf{a}_i^* \mathbf{x}\mathbf{x}^* \mathbf{a}_i) = \mathrm{tr}(\mathbf{a}_i \mathbf{a}_i^* \mathbf{x}\mathbf{x}^*) = \mathrm{tr}(\mathbf{a}_i \mathbf{a}_i^* \mathbf{X}),$$

where $\mathbf{X}$ is the rank-one positive semidefinite matrix defined by $\mathbf{X} := \mathbf{x}\mathbf{x}^*$. Thus, the recovery problem reduces to

$$\underset{\mathbf{Z} \in \mathbb{C}^{n \times n}}{\text{minimize}} \ \mathrm{rank}(\mathbf{Z}) \qquad \text{subject to} \quad \mathrm{tr}(\mathbf{a}_i \mathbf{a}_i^* \mathbf{Z}) = y_i, \quad i \in [\![1:m]\!], \qquad \text{and to} \quad \mathbf{Z} \succeq \mathbf{0}.$$

Since this optimization problem is not computationally tractable, it is relaxed to a convex optimization problem by replacing the rank by the nuclear norm. The expression (30) allows one to treat the resulting optimization problem, namely

$$\underset{\mathbf{Z} \in \mathbb{C}^{n \times n}}{\text{minimize}} \ \|\mathbf{Z}\|_* \qquad \text{subject to} \quad \mathrm{tr}(\mathbf{a}_i \mathbf{a}_i^* \mathbf{Z}) = y_i, \quad i \in [\![1:m]\!], \qquad \text{and to} \quad \mathbf{Z} \succeq \mathbf{0} \tag{32}$$

as a semidefinite program. If a solution to (32) happen to be of rank one, call it $\mathbf{X} \in \mathbb{C}^{n \times n}$, then a solution to the original phaseless recovery problem is derived as $\mathbf{x} = (\|\mathbf{X}\|_F^{1/2} / \|\mathbf{X}\mathbf{u}\|_2)\mathbf{X}\mathbf{u}$ for an arbitrary $\mathbf{u} \in \mathbb{C}^n$. It was shown in [19] that taking $m \asymp N \ln(N)$ Gaussian measurements guarantees that the solution to (32) is indeed of rank one, and the estimate for the number of measurements was improved to $m \asymp N$ in [16]. The number of measurements can be further reduced if the vectors to be recovered are s-sparse: [52] originally showed (in the real setting) that the order $m \asymp s^2 \ln(N)$ is achievable via convex programming, but that $m \asymp s \ln(N)$ is not achievable for a class of naive semidefinite relaxations, and later [45] showed that there is a choice of $m \asymp s \ln(eN/s)$ phaseless measurements allowing for s-sparse recovery.

8.4 One-bit Compressive Sensing

The one-bit compressive sensing problem is in some sense complementary to the sparse phaseless recovery problem: instead of keeping only the magnitudes of linear measurements, these magnitudes are lost and only the sign information is kept —

we are now working in the real setting. The problem, originally proposed in [10], was born out of the practical consideration that measurements really have to be quantized. Although sophisticated quantization scheme such as $\Sigma\Delta$ quantization should be used in practice (see [43] for a $\Sigma\Delta$ analysis of compressive sensing), the one-bit compressive sensing problem concentrates on the extreme scenario where only some binary information about the measurements $\langle \mathbf{a}_i, \mathbf{x}\rangle$ made on sparse vectors $\mathbf{x} \in \mathbb{R}^N$ is retained, i.e., only $y_1 = \mathrm{sgn}\langle \mathbf{a}_1, \mathbf{x}\rangle, \ldots, y_m = \mathrm{sgn}\langle \mathbf{a}_m, \mathbf{x}\rangle$ are now available. Of course, exact s-sparse recovery is now impossible. Instead, the question is phrased as: how few measurements allow us to recover the direction of s-sparse vectors up to accuracy ε? Note that we cannot recover more than the direction. The theory was mostly developed in [59, 60]. We propose below some swift shortcuts through the theory (to keep things simple, stability and robustness issues are left out), relying solely on a modified version of the restricted isometry property with ℓ_1-norm as the inner norm. Precisely, we consider a matrix $\mathbf{A} \in \mathbb{R}^{m\times N}$ such that

$$(1-\delta)\|\mathbf{z}\|_2 \le \|\mathbf{A}\mathbf{z}\|_1 \le (1+\delta)\|\mathbf{z}\|_2 \qquad \text{for all } s\text{-sparse vectors } \mathbf{z} \in \mathbb{R}^N, \tag{33}$$

in which case we write that $\mathbf{A}$ satisfies $\mathrm{RIP}_1(s,\delta)$. For Gaussian matrices populated by independent random variables with mean zero and standard deviation $\sqrt{\pi/2}/m$, $\mathrm{RIP}_1(s,\delta)$ occurs with high probability provided $m \asymp \delta^{-\mu} s \ln(eN/s)$. This can be shown in the same way as Theorem 1, with the caveat that the power of δ^{-1} will have a nonoptimal value $\mu > 2$. The value $\mu = 2$ can be achieved by other means, which in fact allow to prove this modified RIP not only for genuinely s-sparse vectors but also for effectively s-sparse vectors (see [8, 62]). We write $\mathrm{RIP}_1^{\mathrm{eff}}(s,\delta)$ for this extension of the modified RIP. Everything is now in place to present our simplified argument on sparse recovery from one-bit measurements via a thresholding-based strategy and via an optimization-based strategy.

Recovery via hard thresholding: The following observation is the crucial part in the proof of the next theorem.

Lemma 3. *If* $\mathbf{A}$ *satisfies* $RIP_1(s,\delta)$*, then for any* ℓ_2*-normalized vector* $\mathbf{x} \in \mathbb{R}^N$ *supported on an index set* S *of size* s*,*

$$\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S - \mathbf{x}\|_2^2 \le 5\delta.$$

Proof. Expanding the square gives

$$\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S - \mathbf{x}\|_2^2 = \|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\|_2^2 - 2\langle (\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S, \mathbf{x}\rangle + \|\mathbf{x}\|_2^2. \tag{34}$$

The first term on the right-hand side of (34) satisfies

$$\begin{aligned}\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\|_2^2 &= \langle \mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}), (\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\rangle = \langle \mathrm{sgn}(\mathbf{A}\mathbf{x}), \mathbf{A}\left((\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\right)\rangle \\ &\le \|\mathbf{A}\left((\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\right)\|_1 \le (1+\delta)\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\|_2,\end{aligned}$$

so that, after simplification,

$$\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S\|_2 \leq (1+\delta).$$

For the second term on the right-hand side of (34), we notice that

$$\begin{aligned}\langle(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}))_S, \mathbf{x}\rangle &= \langle\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x}), \mathbf{x}\rangle = \langle\mathrm{sgn}(\mathbf{A}\mathbf{x}), \mathbf{A}\mathbf{x}\rangle = \|\mathbf{A}\mathbf{x}\|_1 \\ &\geq (1-\delta)\|\mathbf{x}\|_2 = (1-\delta).\end{aligned}$$

The third term on the right-hand side of (34) is simply $\|\mathbf{x}\|_2^2 = 1$. Altogether, we arrive at

$$\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x})_S - \mathbf{x}\|_2^2 \leq (1+\delta)^2 - 2(1-\delta) + 1 = 4\delta + \delta^2 \leq 5\delta,$$

which is the announced result. ☐

We now consider the simple (noniterative) hard thresholding procedure given by

$$\Delta_{\mathrm{1B-HT}}(\mathbf{y}) := H_s(\mathbf{A}^*\mathbf{y}). \tag{35}$$

Theorem 8. *If* $\mathbf{A}$ *satisfies* $RIP_1(2s,\delta)$*, then every* ℓ_2*-normalized and* s*-sparse vector* $\mathbf{x} \in \mathbb{R}^N$ *observed via* $\mathbf{y} = \mathrm{sgn}(\mathbf{A}\mathbf{x}) \in \{\pm 1\}^m$ *is approximated by the output of the hard thresholding procedure* (35) *with error*

$$\|\mathbf{x} - \Delta_{\mathrm{1B-HT}}(\mathbf{y})\|_2 \leq C\sqrt{\delta}.$$

Proof. Let $S := \mathrm{supp}(\mathbf{x})$ and $T := \mathrm{supp}(\mathbf{x}^{\mathrm{HT}})$. We notice that $\mathbf{x}^{\mathrm{HT}}$ is also the best s-term approximation to $(\mathbf{A}^*\mathbf{y})_{S\cup T}$, so that

$$\begin{aligned}\|\mathbf{x} - \mathbf{x}^{\mathrm{HT}}\|_2 &\leq \|(\mathbf{A}^*\mathbf{y})_{S\cup T} - \mathbf{x}^{\mathrm{HT}}\|_2 + \|(\mathbf{A}^*\mathbf{y})_{S\cup T} - \mathbf{x}\|_2 \leq 2\|(\mathbf{A}^*\mathbf{y})_{S\cup T} - \mathbf{x}\|_2 \\ &\leq 2\sqrt{5\delta},\end{aligned}$$

where we made use of Lemma 3. This establishes our result. ☐

Recovery via linear programming: We consider here the optimization problem

$$\Delta_{\mathrm{1B-LP}}(\mathbf{y}) = \mathrm{argmin}\ \|\mathbf{z}\|_1 \qquad \text{subject to } \mathrm{sgn}(\mathbf{A}\mathbf{z}) = \mathbf{y} \quad \text{and} \quad \|\mathbf{A}\mathbf{z}\|_1 = 1. \tag{36}$$

By writing the constraints as $y_i\langle\mathbf{a}_i, \mathbf{z}\rangle \geq 0$ for $i \in [\![1:m]\!]$ and $\sum_{i=1}^m y_i\langle\mathbf{a}_i, \mathbf{z}\rangle = 1$, we see that this optimization problem can be recast as a linear program. We need the following intermediate result before justifying the recovery guarantee.

Lemma 4. *If* $\mathbf{A}$ *satisfies* $RIP_1(9s, 1/5)$ *and if* $\mathbf{x} \in \mathbb{R}^N$ *is* s*-sparse, then any convex combination of* $\mathbf{x}$ *and* $\Delta_{\mathrm{1B-LP}}(\mathrm{sgn}(\mathbf{A}\mathbf{x}))$ *is effectively* $9s$*-sparse. The same conclusion holds under* $RIP_1^{\mathrm{eff}}(9s, 1/5)$ *if* $\mathbf{x}$ *is effectively* s*-sparse.*

Proof. Let us use the notation $t = 9s$, $\delta = 1/5$, and $\mathbf{x}^\sharp = \Delta_{\mathrm{1B-LP}}(\mathrm{sgn}(\mathbf{A}\mathbf{x}))$. By the defining property of $\mathbf{x}^\sharp$, we have

$$\|\mathbf{x}^\sharp\|_1 \leq \left\|\frac{\mathbf{x}}{\|\mathbf{A}\mathbf{x}\|_1}\right\|_1 \leq \frac{\sqrt{s}\|\mathbf{x}\|_2}{\|\mathbf{A}\mathbf{x}\|_1} \leq \frac{\sqrt{s}}{1-\delta} = \frac{5\sqrt{s}}{4}.$$

If $\widehat{\mathbf{x}} = (1-\lambda)\mathbf{x} + \lambda \mathbf{x}^\sharp$ denotes a convex combination of $\mathbf{x}$ and $\mathbf{x}^\sharp$, then we obtain

$$\|\widehat{\mathbf{x}}\|_1 \le (1-\lambda)\|\mathbf{x}\|_1 + \lambda \|\mathbf{x}^\sharp\|_1 \le (1-\lambda)\sqrt{s} + \lambda \frac{5\sqrt{s}}{4} = \left(1 + \frac{\lambda}{4}\right)\sqrt{s}.$$

Notice also that $\operatorname{sgn}(\mathbf{A}\widehat{\mathbf{x}}) = \operatorname{sgn}(\mathbf{A}\mathbf{x}^\sharp) = \operatorname{sgn}(\mathbf{A}\mathbf{x})$, so that

$$\|\mathbf{A}\widehat{\mathbf{x}}\|_1 = (1-\lambda)\|\mathbf{A}\mathbf{x}\|_1 + \lambda \|\mathbf{A}\mathbf{x}^\sharp\|_1 \ge (1-\lambda)(1-\delta) + \lambda = \frac{4}{5}\left(1 + \frac{\lambda}{4}\right).$$

Next, we use the sort-and-split technique and consider an index set T_0 of t largest absolute entries of $\widehat{\mathbf{x}}$, an index set T_1 of t next largest absolute entries of $\widehat{\mathbf{x}}$, an index set T_2 of t next largest absolute entries of $\widehat{\mathbf{x}}$, etc. Thus,

$$\begin{aligned}
\frac{4}{5}\left(1+\frac{\lambda}{4}\right) &\le \|\mathbf{A}\widehat{\mathbf{x}}\|_1 \le \sum_{k\ge 0} \|\mathbf{A}\widehat{\mathbf{x}}_{T_k}\|_1 \le (1+\delta)\left(\|\widehat{\mathbf{x}}_{T_0}\|_2 + \sum_{k\ge 1}\|\widehat{\mathbf{x}}_{T_k}\|_2\right)\\
&\le (1+\delta)\left(\|\widehat{\mathbf{x}}\|_2 + \frac{1}{\sqrt{t}}\|\widehat{\mathbf{x}}\|_1\right) \le (1+\delta)\left(\|\widehat{\mathbf{x}}\|_2 + \sqrt{\frac{s}{t}}\left(1+\frac{\lambda}{4}\right)\right)\\
&= \frac{6}{5}\left(\|\widehat{\mathbf{x}}\|_2 + \frac{1}{3}\left(1+\frac{\lambda}{4}\right)\right).
\end{aligned}$$

This implies that $\|\widehat{\mathbf{x}}\|_2 \ge (1+\lambda/4)/3$. In turn, we derive

$$\frac{\|\widehat{\mathbf{x}}\|_1}{\|\widehat{\mathbf{x}}\|_2} \le \frac{(1+\lambda/4)\sqrt{s}}{(1+\lambda/4)/3} = \sqrt{9s}.$$

This means that $\widehat{\mathbf{x}}$ is effectively $9s$-sparse, as announced. □

We can now state and prove the recovery result for the linear program.

Theorem 9. *If* $\mathbf{A}$ *satisfies* $RIP_1^{\mathrm{eff}}(9s,\delta)$ *with* $\delta \le 1/5$*, then every* ℓ_2*-normalized and effectively* s*-sparse vector* $\mathbf{x} \in \mathbb{R}^N$ *observed via* $\mathbf{y} = \operatorname{sgn}(\mathbf{A}\mathbf{x}) \in \{\pm 1\}^m$ *is approximated by the output of the linear programming procedure* (36) *with error*

$$\|\mathbf{x} - \Delta_{\mathrm{1B-LP}}(\mathbf{y})\|_2 \le C\sqrt{\delta}.$$

Proof. We still use the notation $\mathbf{x}^\sharp$ for $\Delta_{\mathrm{1B-LP}}(\mathbf{y})$. By Lemma 4, the vector $(\mathbf{x}+\mathbf{x}^\sharp)/2$ is effectively $9s$-sparse, and using the fact that $\operatorname{sgn}(\mathbf{A}\mathbf{x}^\sharp) = \operatorname{sgn}(\mathbf{A}\mathbf{x})$, we have

$$\begin{aligned}
\left\|\frac{\mathbf{x}+\mathbf{x}^\sharp}{2}\right\|_2 &\ge \frac{1}{1+\delta}\left\|\mathbf{A}\left(\frac{\mathbf{x}+\mathbf{x}^\sharp}{2}\right)\right\|_1 = \frac{1}{1+\delta}\frac{\|\mathbf{A}\mathbf{x}\|_1 + \|\mathbf{A}\mathbf{x}^\sharp\|_1}{2} \ge \frac{1}{1+\delta}\frac{(1-\delta)+1}{2}\\
&= \frac{1-\delta/2}{1+\delta}.
\end{aligned}$$

Lemma 4 also implies that $\mathbf{x}^\sharp$ is effectively $9s$-sparse, so we can write

$$\left\| \frac{\mathbf{x}-\mathbf{x}^\sharp}{2} \right\|_2^2 = \frac{1+\|\mathbf{x}^\sharp\|_2^2}{2} - \left\| \frac{\mathbf{x}+\mathbf{x}^\sharp}{2} \right\|_2^2 \quad \leq \frac{1+1/(1-\delta)^2}{2} - \frac{(1-\delta/2)^2}{(1+\delta)^2}$$
$$\leq C\delta.$$

This immediately implies the announced result. □

Recovery of more than the direction: To sum up the previous results, it is possible to recover the direction of s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ acquired via $\mathbf{y} = \mathrm{sgn}(\mathbf{A}\mathbf{x}) \in \{\pm 1\}^m$ with accuracy ε provided $m \geq C\varepsilon^4 s \ln(eN/s)$. Stated differently, there are recovery maps $\Delta : \{\pm 1\}^m \to \mathbb{R}^N$ such that

$$\|\mathbf{x} - \Delta(\mathrm{sgn}(\mathbf{A}\mathbf{x}))\|_2 \leq C \left[\frac{m}{s \ln(eN/s)} \right]^{1/4} \qquad \text{whenever } \|\mathbf{x}\|_0 \leq s, \|\mathbf{x}\|_2 = 1.$$

It is also possible to estimate the magnitude of $\mathbf{x}$ if one allows thresholds $\tau_1, \ldots, \tau_m$ in the binary measurements $y_1 = \mathrm{sgn}(\langle \mathbf{a}_i, \mathbf{x} \rangle - \tau_i)$, $i \in [\![1:m]\!]$, see [4, 5, 49]. In fact, if these thresholds can be chosen adaptively, we proved in [4] that one can even achieve the bound

$$\|\mathbf{x} - \Delta(\mathrm{sgn}(\mathbf{A}\mathbf{x} - \tau))\|_2 \leq C \exp\left(-c \left[\frac{m}{s \ln(eN/s)} \right] \right) \quad \text{whenever } \|\mathbf{x}\|_0 \leq s, \|\mathbf{x}\|_2 \leq 1.$$

9 Open Questions

We close this survey by describing some compressive sensing problems that remain currently unresolved. The first one, concerning Gelfand widths, sits at the boundary of the topic, but it is included in view of the central part played by Gelfand widths in this survey. The other ones are bona fide compressive sensing problems.

Gelfand width of the ℓ_1-ball in ℓ_∞: We have seen how two-sided estimates for the Gelfand width $d^m(B_1^N, \ell_2^N)$ can be obtained from methods of compressive sensing. The same methods also yield two-sided estimates for the Gelfand width $d^m(B_1^N, \ell_q^N)$ for any $q \in [1,2]$. But for $q > 2$, known lower and upper estimates do not quite match. In particular, for $q = \infty$, it is only known that

$$c \max\left\{ \sqrt{\frac{1}{m}}, \min\left\{1, \frac{\ln(eN/m)}{m}\right\} \right\} \leq d^m(B_1^N, \ell_\infty^N) \leq C \min\left\{1, \sqrt{\frac{\ln(eN/m)}{m}}\right\},$$

see [38, 68] for details. Finding the correct order for $d^m(B_1^N, \ell_\infty^N)$ seems challenging.

The quest for deterministic compressive sensing matrices: We have seen that measurement matrices $\mathbf{A} \in \mathbb{C}^{m \times N}$ allowing for stable and robust s-sparse recovery in the optimal regime $m \asymp s \ln(eN/s)$ do exist. In fact, many such matrices should exist, as the argument consists in proving that random matrices satisfy, with high probability, some property that guarantee sparse recovery, e.g. the RIP. To date, however, the derandomization problem of exhibiting concrete examples of matrices allowing for stable and robust sparse recovery remains open — arguably, it is the most significant open problem in the field. Note that selecting a realization of a random matrix and verifying that it has the RIP, say, is not a viable strategy, since this task is NP-hard (see [3, 66]). Using the concept of coherence, it is easy (see e.g. [39, Chapter 5]) to deterministically produce matrices that allow for stable and robust s-sparse recovery with a number m of rows of the order of s^2. Reducing this number of rows to $m \asymp s^{\beta}$ with $\beta < 2$ would already be a breakthrough. In fact, such a breakthrough was achieved in [12]: an explicit choice of $m \asymp s^{2-\varepsilon}$ rows from the discrete Fourier matrix was shown to yield a matrix with the RIP of order s, but $\varepsilon > 0$ was ever so small that this theoretical feat is of no practical interest. Finally, it has to be noted that, should one settle for a weaker notion of stability and robustness (in the spirit of footnote 3), deterministic measurement matrices in the near-optimal regime $m \asymp s^{\alpha}$ are supposedly available for any $\alpha > 1$, since [44] claims an explicit construction of (s,d,θ)-lossless expanders with this number of right vertices.

Universality of phase transition for random matrices: In [28], the authors look at the standard compressive sensing problem from a different perspective. Defining $\delta := m/N \in [0,1]$ (not the be confused with a restricted isometry constant) and $\rho := s/m \in [0,1]$, they investigate the phase diagram for the function

$$P(\delta,\rho) := \lim_{N \to \infty} \mathbb{P}(\mathbf{x} \text{ is exactly recovered from } \mathbf{y} = \mathbf{A}\mathbf{x} \text{ by } \ell_1\text{-minimization}),$$

where $\mathbf{x} \in \mathbb{R}^N$ is an arbitrary (but fixed) vector and where $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a Gaussian random matrix. Exploiting sophisticated results about random polytopes, they could show that $P^{-1}(\{0\})$ and $P^{-1}(\{1\})$ partition the square $[0,1]^2$. In other words, when N is large, ℓ_1-recovery from random measurements either fails or succeeds with high probability depending on the values of the ratios s/m and m/N. They could also give an implicit expression of the curve separating the regions $P^{-1}(\{0\})$ and $P^{-1}(\{1\})$. Their original arguments have been simplified and extended in [2] by relying on tools from integral geometry. The authors also carried out extensive numerical experiments for non-Gaussian measurements. These experiments, reported in [29], suggest that the phase transition phenomenon is ubiquitous across a wide variety of random matrices, as the observed phase diagrams were in many cases identical to the one derived theoretically in [28]. An explanation for this universality is currently lacking.

Logarithmic factors in the RIP for BOS: We have seen in Subsection 7.3 that random sampling matrices associated with bounded orthonormal systems possess the RIP of order s provided $m \asymp s\ln^3(N)$. Reducing the power of the logarithm factor further seem to require considerable efforts. It constitutes a noteworthy open problem, not so much because of practical implications (as the hidden constants could dwarf the logarithm factor) but because of theoretical implications. Indeed, achieving $m \asymp s\ln(N)$ would solve a standing conjecture about the Λ_1-problem studied by Bourgain [11] and Talagrand [63]. See [39, Section 12.7] for details about this connection.

Dictionary sparsity: We have started this journey by identifying from the on-set signals with their coefficient vectors in a given basis, and these vectors were assumed to be sparse. But it is often more realistic to assume that the signals have a sparse representation not in a basis but rather in a dictionary (i.e., an overcomplete system). In order to give a few details, suppose to simplify that $\mathbf{D} \in \mathbb{C}^{n \times N}$ is a tight frame, in the sense that $\mathbf{D}\mathbf{D}^* = \mathbf{I}$. Saying that a signal $\mathbf{f} \in \mathbb{C}^n$ is sparse with respect to the dictionary $\mathbf{D}$ may have two meanings:

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

Synthesis sparsity is a more realistic assumption to make, and there are theoretical algorithms that allow for stable and robust s-synthesis-sparse recovery from a number $m \asymp s\ln(eN/s)$ of random measurements, see e.g. [25]. The term'theoretical' algorithms was used because these algorithms involve finding best or near-best approximations from the set of s-synthesis-sparse vectors, which is sometimes a difficult task in itself (see [65]). In contrast, the algorithms proposed for analysis-sparse recovery are really practical. The optimization-based strategy of [15] and the thresholding-based iterative strategy of [32] both yields the same stability and robustness estimate featuring the error $\sigma_s(\mathbf{D}^*\mathbf{f})_1$ of best s-term approximation to $\mathbf{D}^*\mathbf{f}$ in ℓ_1-norm. However, it is not clear in all situations that this is a small quantity. Certainly, if the columns of $\mathbf{D}$ are in general position, then $\sigma_s(\mathbf{D}^*\mathbf{f})_1$ cannot equal zero for $s \le N - n$ unless $\mathbf{f} = \{\mathbf{0}\}$, otherwise $(\mathbf{D}^*\mathbf{f})_j = 0$ for $N - s$ indices j means that $\mathbf{f} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ is orthogonal to $N - s \ge n$ linearly independent vectors, which is impossible. To sum up, the two notions of dictionary sparsity are both undermined by specific drawbacks for certain dictionaries. It is desirable to determine which dictionaries can circumvent the pitfalls of synthesis-sparsity recovery and of analysis-sparsity recovery simultaneously.

Structure other than sparsity, combination of structures: We have argued that compressive sensing does not reduce to the standard problem (3) — it concerns more generally the recovery of structured high-dimensional objects acquired from fewer measurements than conventionally thought necessary, as exemplified in Subsection 8.1, say. There are other structured objects to be considered besides sparse vectors and low-rank matrices, e.g. the recovery of low-rank tensors is currently an active research niche. The recovery of objects possessing several structures simultaneously leads to further interesting inquiries. For fairly general structures, the authors of [57] showed that, roughly speaking, as far as the asymptotic number of Gaussian measurements is concerned and when optimization-based recovery methods are used, there is no gain in exploiting the combination of structures instead of exploiting each one of them separately. Such an observation was corroborated in [37], where a similar conclusion was reached for a quite restrictive combination of structures — sparsity and disjointedness — but for arbitrary measurement and recovery schemes. However, this does not seem to be a general phenomenon. As a case in point, it can be shown that recovering matrices that are both low-rank and sparse is possible from fewer measurements than the numbers needed for the recovery of low-rank matrices alone and for the recovery of sparse matrices alone. The recovery map is not optimization-based — this would contradict [57] — but it is a thresholding-based iterative algorithm, with the caveat that it involves the impractical task of finding best or near-best approximations from the set of simultaneously low-rank and sparse matrices. But besides solving this minor issue, a challenging program consists in understanding exactly when the combination of structures turns out to be beneficial. Such a program may need to be carried out on a case-to-case basis.

Acknowledgements I thank the organizers of the International Conference on Approximation Theory for running this important series of triennial meetings. It was a plenary address by Ron DeVore at the 2007 meeting that drove me into the subject of compressive sensing. His talk was entitled "A Taste of Compressed sensing" and my title is clearly a reference to his. Furthermore, I acknowledge support from the NSF under the grant DMS-1622134. Finally, I am also indebted to the AIM SQuaRE program for funding and hosting a collaboration on one-bit compressive sensing.

References

1. R. Adamczak, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann: Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. Constructive Approximation. **34**, 61–88 (2011).
2. D. Amelunxen, M. Lotz, M. McCoy, J. Tropp: Living on the edge: Phase transitions in convex programs with random data. Information and Inference. iau005 (2014).
3. A. Bandeira, E. Dobriban, D. Mixon, and W. Sawin: Certifying the restricted isometry property is hard. IEEE Trans. Inform. Theory. **59**, 3448–3450 (2013).
4. R. Baraniuk, S. Foucart, D. Needell, Y. Plan, M. Wootters: Exponential decay of reconstruction error from binary measurements of sparse signals. IEEE Trans. Inform. Theory. To appear.
5. R. Baraniuk, S. Foucart, D. Needell, Y. Plan, M. Wootters: One-bit compressive sensing of dictionary-sparse signals. Preprint.
6. R. Berinde, A. Gilbert, P. Indyk, H. Karloff, M. Strauss: Combining geometry and combinatorics: a unified approach to sparse signal recovery. In: Proc. of 46th Annual Allerton Conference on Communication, Control, and Computing, pp. 798–805 (2008).
7. S. Bhojanapalli, P. Jain: Universal matrix completion. In: Proceedings of the 31st International Conference on Machine Learning (ICML). MIT Press (2014).
8. D. Bilyk, M. T. Lacey: Random tessellations, restricted isometric embeddings, and one bit sensing. arXiv preprint arXiv:1512.06697 (2015).
9. J.-L. Bouchot, S. Foucart, P. Hitczenko: Hard thresholding pursuit algorithms: number of iterations. Applied and Computational Harmonic Analysis. **41**, 412–435 (2016).
10. P. Boufounos, R. Baraniuk: 1-bit compressive sensing. In: Proceedings of the 42nd Annual Conference on Information Sciences and Systems (CISS), pp. 16–21. IEEE (2008).
11. J. Bourgain: Bounded orthogonal systems and the $\Lambda(p)$-set problem. Acta Math. **162**, 227–245 (1989).
12. J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, D. Kutzarova: Explicit constructions of RIP matrices and related problems. Duke Mathematical Journal. **159**, 145–185 (2011).
13. H. Buhrman, P. Miltersen, J. Radhakrishnan, S. Venkatesh: Are bitvectors optimal? In: Proceedings of the 32nd Annual ACM Symposium on Theory of Computing (STOC), pp. 449–458. ACM (2000).
14. T. Cai, A. Zhang: Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. IEEE Trans. Inform. Theory. **60**, 122–132 (2014).

15. E. Candès, Y. Eldar, D. Needell, P. Randall: Compressed sensing with coherent and redundant dictionaries. Applied and Computational Harmonic Analysis. **31**, 59–73 (2011).
16. E. Candès, X. Li: Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. Foundations of Computational Mathematics. **14**, 1017–1026 (2014).
17. E. Candès, Y. Plan: Matrix completion with noise. Proceedings of the IEEE. **98**, 925–936 (2010).
18. E. Candès, B. Recht: Exact matrix completion via convex optimization. Found. of Comput. Math. **9**, 717–772 (2009).
19. E. Candès, T. Strohmer, V. Voroninski: Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. Commun. Pure Appl. Math. **66**, 1241–1274 (2013).
20. E. Candès, J. Romberg, T. Tao: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory. **52**, 489–509 (2006).
21. E. Candès, T. Tao: Decoding by linear programming. IEEE Trans. Inform. Theory. **51**, 4203–4215 (2005).
22. A. Chkifa, N. Dexter, H. Tran, C. Webster: Polynomial approximation via compressed sensing of high-dimensional functions on lower sets. Preprint.
23. A. Cohen, W. Dahmen, R. DeVore: Compressed sensing and best k-term approximation. J. Amer. Math. Soc. **22**, 211–231 (2009).
24. A. Cohen, W. Dahmen, R. DeVore: Orthogonal Matching Pursuit under the Restricted Isometry Property. Constructive Approximation. **45**, 113–127 (2017).
25. M. Davenport, D. Needell, M. Wakin: Signal space CoSaMP for sparse recovery with redundant dictionaries. IEEE Trans. Inform. Theory. **59**, 6820–6829 (2013).
26. M. Davies, R. Gribonval: Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$. IEEE Trans. Inform. Theory. **55**, 2203–2214 (2009).
27. D. Donoho: For most large underdetermined systems of linear equations the minimal ℓ^1 solution is also the sparsest solution. Commun. Pure Appl. Math. **59**, 797–829 (2006).
28. D. Donoho, J. Tanner: Counting faces of randomly projected polytopes when the projection radically lowers dimension. Journal of the American Mathematical Society. **22**, 1–53 (2009).
29. D. Donoho, J. Tanner: Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. **367**, 4273–4293 (2009).
30. S. Foucart: Stability and robustness of weak orthogonal matching pursuits. In: Bilyk, D., De Carli, L., Petukhov, A., Stokolos, A.M., Wick, B.D. (eds.), Recent advances in harmonic analysis and applications, pp. 395–405. Springer, New York (2012).
31. S. Foucart: Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries. Linear Algebra and its Applications. **441**, 4–21 (2014).
32. S. Foucart: Dictionary-sparse recovery via thresholding-based algorithms. Journal of Fourier Analysis and Applications, **22**, 6–19 (2016).
33. S. Foucart, D. Koslicki: Sparse recovery by means of nonnegative least squares. IEEE Signal Processing Letters. **21**, 498–502 (2014).
34. S. Foucart, R. Gribonval: Real vs. complex null space properties for sparse vector recovery. C. R. Math. Acad. Sci. Paris. **348**, 863–865 (2010).
35. S. Foucart, G. Lecué: An IHT algorithm for sparse recovery from subexponential measurements. Preprint.
36. S. Foucart, M.-J. Lai: Sparse recovery with pre-Gaussian random matrices. Studia Math. **200**, 91–102 (2010).
37. S. Foucart, M. Minner, T. Needham: Sparse disjointed recovery from noninflating measurements. Applied and Computational Harmonic Analysis. **39**, 558–567 (2015).
38. S. Foucart, A. Pajor, H. Rauhut, T. Ullrich: The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. J. Complexity. **26**, 629–640 (2010).
39. S. Foucart, H. Rauhut: A mathematical introduction to compressive sensing. Birkhäuser, Boston (2013).

40. A. Garnaev, E. Gluskin: On widths of the Euclidean ball. Sov. Math., Dokl. **30**, 200–204 (1984).
41. R. Graham, N. Sloane: Lower bounds for constant weight codes. IEEE Trans. Inform. Theory. **26**, 37–43 (1980).
42. D. Gross: Recovering low-rank matrices from few coefficients in any basis. IEEE Trans. Inform. Theory. **57**, 1548–1566 (2011).
43. C. Güntürk, M. Lammers, A. Powell, R. Saab, Ö. Yılmaz: Sigma-Delta quantization for compressed sensing. In: Proceedings of the 44th Annual Conference on Information Sciences and Systems (CISS). IEEE, (2010).
44. V. Guruswani, C. Umans, S. Vadhan: Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. In: IEEE Conference on Computational Complexity, pp. 237–246 (2007).
45. M. Iwen, A. Viswanathan, Y. Wang: Robust sparse phase retrieval made easy. Applied and Computational Harmonic Analysis. **42**, 135–142 (2017).
46. B. Kashin: Diameters of some finite-dimensional sets and classes of smooth functions. Math. USSR, Izv. **11**, 317–333 (1977).
47. D. Koslicki, S. Foucart, and G. Rosen: Quikr: a method for rapid reconstruction of bacterial communities via compressive sensing. Bioinformatics. btt336 (2013).
48. D. Koslicki, S. Foucart, and G. Rosen: WGSQuikr: fast whole-genome shotgun metagenomic classification. PloS one. **9** e91784 (2014).
49. K. Knudson, R. Saab, R. Ward: One-bit compressive sensing with norm estimation. IEEE Trans. Inform. Theory. **62**, 2748–2758 (2016).
50. C. Lawson, R. Hanson: Solving least squares problems. SIAM, Philadelphia (1995).
51. G. Lecué, S. Mendelson: Sparse recovery under weak moment assumptions. Journal of the European Mathematical Society. **19**, 881–904 (2017).
52. X. Li, V. Voroninski: Sparse signal recovery from quadratic measurements via convex programming. SIAM Journal on Mathematical Analysis. **45**, 3019–3033 (2013).
53. N. Linial, I. Novik: How neighborly can a centrally symmetric polytope be? Discr. Comput. Geom. **36**, 273–281 (2006).
54. G. Lorentz, M. von Golitschek, and Y. Makovoz. Constructive approximation: advanced problems. Springer, Berlin (1996).
55. S. Mendelson, A. Pajor, M. Rudelson: The geometry of random $\{-1,1\}$-polytopes. Discr. Comput. Geom. **34**, 365–379 (2005).
56. N. Noam, W. Avi: Hardness vs randomness. Journal of Computer and System Sciences. **49**, 149–167 (1994).
57. S. Oymak, A. Jalali, M. Fazel, Y. Eldar, B. Hassibi: Simultaneously structured models with application to sparse and low-rank matrices. IEEE Trans. Inform. Theory. **61**, 2886–2908 (2015).
58. A. Pinkus: *n*-Widths in Approximation Theory. Springer-Verlag, Berlin (1985).
59. Y. Plan, R. Vershynin: One-bit compressed sensing by linear programming. Commun. Pure Appl. Math. **66**, 1275–1297 (2013).
60. Y. Plan, R. Vershynin: Robust 1-bit compressed sensing and sparse logistic regression: a convex programming approach. IEEE Trans. Inform. Theory. **59**, 482–494 (2013).
61. B. Recht: A simpler approach to matrix completion. Journal of Machine Learning Research. **12**, 3413–3430 (2011).
62. G. Schechtman: Two observations regarding embedding subsets of Euclidean spaces in normed spaces. Advances in Mathematics. **200**, 125–135 (2006).
63. M. Talagrand: Selecting a proportion of characters. Israel J. Math. **108**, 173–191 (1998).
64. V. Temlyakov: Greedy approximation. Cambridge University Press, Cambridge (2011).
65. A. Tillmann, R. Gribonval, M. Pfetsch: Projection onto the cosparse set is NP-hard. In: Proceedings of the 2014 Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, (2014).
66. A. Tillmann, M. Pfetsch: The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. IEEE Trans. Inform. Theory. **60**, 1248–1259 (2014).

67. J. Tropp, A. Gilbert: Signal recovery from random measurements via orthogonal matching pursuit. IEEE Trans. Inform. Theory. **53**, 4655–4666 (2007).
68. J. Vybíral: Widths of embeddings in function spaces. J. Complexity. **24**, 545–570 (2008).
69. T. Zhang: Sparse recovery with orthogonal matching pursuit under RIP. IEEE Trans. Inform. Theory. **57**, 6215–6221 (2011).