Sparse Recovery from Inaccurate Saturated Measurements[*]

Simon Foucart[†] and Jiangyuan Li — Texas A&M University

Abstract

This article studies a variation of the standard compressive sensing problem, in which sparse vectors $\mathbf{x} \in \mathbb{R}^N$ are acquired through inaccurate saturated measurements $\mathbf{y} = \mathcal{S}(\mathbf{A}\mathbf{x} + \mathbf{e}) \in \mathbb{R}^m$, $m \ll N$. The saturation function $\mathcal{S}$ acts componentwise by sending entries that are large in absolute value to plus-or-minus a threshold while keeping the other entries unchanged. The present study focuses on the effect of the presaturation error $\mathbf{e} \in \mathbb{R}^m$. The existing theory for accurate saturated measurements, i.e., the case $\mathbf{e} = \mathbf{0}$, which exhibits two regimes depending on the magnitude of $\mathbf{x} \in \mathbb{R}^N$, is extended here. A recovery procedure based on convex optimization is proposed and shown to be robust to presaturation error in both regimes. Another procedure ignoring the presaturation error is also analyzed and shown to be robust in the small magnitude regime.

Key words and phrases: Compressive sensing, saturation, convex programming.

AMS classification: 94A12, 90C05, 60D05.

1 Introduction

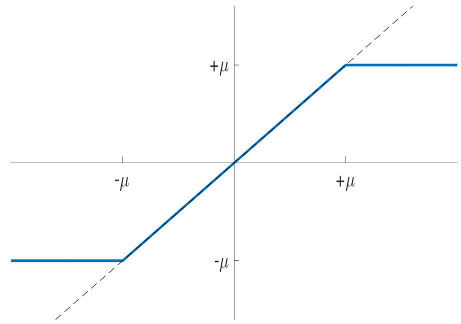
Suppose that vectors $\mathbf{x} \in \mathbb{R}^N$ are acquired through m linear measurements $y_i = \langle \mathbf{a}_i, \mathbf{x} \rangle = \mathbf{a}_i^\top \mathbf{x}$, $i = 1, \ldots, m$, for some $\mathbf{a}_1, \ldots, \mathbf{a}_m \in \mathbb{R}^N$. In condensed form, this reads $\mathbf{y} = \mathbf{A}\mathbf{x}$, where the matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ has rows $\mathbf{a}_1^\top, \ldots, \mathbf{a}_m^\top$. When m is smaller than N, it is in general inconceivable to recover $\mathbf{x} \in \mathbb{R}^N$ from the mere knowledge of $\mathbf{y} \in \mathbb{R}^m$, but the theory of compressive sensing [4, 5, 8] made it clear that such a recovery task is in fact possible when some prior information about the structure of $\mathbf{x}$ is available. Typically, it is assumed that the vectors of interest are s-sparse, i.e., that the $\mathbf{x} \in \mathbb{R}^N$ possess at most s nonzero entries. In this case, it is now well known that only $m \asymp s \ln(eN/s)$ random measurements enable the exact recovery of all s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ via various efficient algorithms taking the compressed version $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ as an input (together with the matrix $\mathbf{A}$ itself, of course).

[*]This work is the result of Jiangyuan Li's capstone project, carried out at Texas A&M University as part of an exchange with Beihang University.

[†]S. F. is partially supported by the NSF under the grant DMS-1622134.

Now suppose that sparse vectors $\mathbf{x} \in \mathbb{R}^N$ are not acquired through $\mathbf{y} = \mathbf{A}\mathbf{x}$ but rather through $\mathcal{S}(\mathbf{y}) = \mathcal{S}(\mathbf{A}\mathbf{x})$, where $\mathcal{S} = \mathcal{S}_\mu$ is the saturation function depicted below.

$$\mathcal{S}(t) = \begin{cases} -\mu, & t \in (-\infty, -\mu], \\ t, & t \in (-\mu, +\mu), \\ +\mu, & t \in [+\mu, +\infty). \end{cases}$$



Intuitively, at one extreme, when the threshold μ is very large (compared to the magnitude of $\mathbf{x}$), the saturation function $\mathcal{S}$ has no effect and the setting of standard compressive sensing prevails, so one expects exact recovery of $\mathbf{x}$ to be possible. At the other extreme, when the threshold μ is very small (compared to the magnitude of $\mathbf{x}$), knowing $\mathcal{S}(\mathbf{y})$ reduces to knowing the binary information $\mathrm{sgn}(\mathbf{A}\mathbf{x})$, so one expects approximate recovery of the direction of $\mathbf{x}$ to be possible (and not more), similarly to the setting of one-bit compressive sensing (see [3] where the scenario was put forward, [11] where the first theoretical results were established, and [6, Section 8.4] where they were simplified). Between these two extremes, one expects more than direction recovery to be possible, since more than just the sign information is available. This intuition was formalized in [7] for the recovery map given by

$$\Lambda_1(\mathbf{y}) := \underset{\mathbf{z} \in \mathbb{R}^N}{\mathrm{argmin}} \; \|\mathbf{z}\|_1 \qquad \text{subject to} \quad \mathcal{S}(\mathbf{A}\mathbf{z}) = \mathcal{S}(\mathbf{y}). \tag{1}$$

In short, the article [7] showed that, with high probability on the draw of a random matrix $\mathbf{A}$ populated by independent $\mathcal{N}(0, \sigma^2)$ entries,

(i) there is a constant $\alpha > 0$ such that, for every s-sparse $\mathbf{x} \in \mathbb{R}^N$ satisfying $\sigma\|\mathbf{x}\|_2 \le \alpha\mu$, one has

$$\Lambda_1(\mathcal{S}(\mathbf{A}\mathbf{x})) = \mathbf{x} \tag{2}$$

provided $m \ge \kappa s \ln(eN/s)$ for some absolute constant $\kappa > 0$;

(ii) given a fixed constant $\beta > \alpha$, and given any $\delta \in (0,1)$, for every s-sparse $\mathbf{x} \in \mathbb{R}^N$ satisfying $\alpha\mu < \sigma\|\mathbf{x}\|_2 \le \beta\mu$, one has

$$\|\mathbf{x} - \Lambda_1(\mathcal{S}(\mathbf{A}\mathbf{x}))\|_2 \le \delta \|\mathbf{x}\|_2 \tag{3}$$

provided $m \ge \kappa_{\beta,\delta} s \ln(eN/s)$ for some constant $\kappa_{\beta,\delta} > 0$ depending on β and δ.

The article [7] did not touch upon potential complications that might occur when the measurements are corrupted. This is the focus of this paper, which will establish robustness of optimization-based recovery methods in this situation. So from now on, the inaccurate compressed version of a vector $\mathbf{x} \in \mathbb{R}^N$ takes the form $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ for some measurement error $\mathbf{e} \in \mathbb{R}^m$. In standard compressive sensing, when a bound $\|\mathbf{e}\|_2 \le \eta$ is known a priori, it is classical to approximate $\mathbf{x}$ by

$$(4) \qquad \Delta_{1,\eta}(\mathbf{y}) := \underset{\mathbf{z} \in \mathbb{R}^N}{\operatorname{argmin}} \ \|\mathbf{z}\|_1 \qquad \text{subject to} \quad \|\mathbf{A}\mathbf{z} - \mathbf{y}\|_2 \le \eta.$$

Under some condition on the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$, there exist absolute constants $\kappa, D > 0$ such that the recovery bound

$$(5) \qquad \|\mathbf{x} - \Delta_{1,\eta}(\mathbf{A}\mathbf{x} + \mathbf{e})\|_2 \le D\eta$$

is valid for all s-sparse $\mathbf{x} \in \mathbb{R}^N$ and all $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\|_2 \le \eta$, provided $m \ge \kappa s \ln(eN/s)$. The required condition on the matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is fulfilled with high probability when its entries are independent Gaussian random variables with mean zero and standard deviation $\sigma = 1/\sqrt{m}$, which is what we assume in the rest of the paper.

We now place ourselves in the situation where the measurement vector $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ is only available through its saturated version $\mathcal{S}(\mathbf{y})$, with the knowledge of a bound $\|\mathbf{e}\|_2 \le \eta$ on the presaturation error. Our goal is to establish recovery bounds bridging (2)-(3) and (5). Namely, we aim at uncovering a recovery procedure $\Lambda_{1,\eta} : [-\mu, +\mu]^m \to \mathbb{R}^N$ such that

- in a regime of small magnitude, i.e., when the s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ satisfy the inequality $\|\mathbf{x}\|_2 \le \alpha\mu\sqrt{m}$, one has

$$(6) \qquad \|\mathbf{x} - \Lambda_{1,\eta}(\mathcal{S}(\mathbf{A}\mathbf{x} + \mathbf{e}))\|_2 \le D\eta;$$

- in a regime of intermediate magnitude, i.e., when the s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ satisfy the inequalities $\alpha\mu\sqrt{m} < \|\mathbf{x}\|_2 \le \beta\mu\sqrt{m}$, one has

$$(7) \qquad \|\mathbf{x} - \Lambda_{1,\eta}(\mathcal{S}(\mathbf{A}\mathbf{x} + \mathbf{e}))\|_2 \le \delta\|\mathbf{x}\|_2 + D\eta.$$

The recovery procedure we propose, which we call noise-cognizant ℓ_1-minimization (to distinguish it from the noise-ignoring ℓ_1-minimization discussed in Section 5) consists in producing from $\mathcal{S}(\mathbf{y})$ the vector $\Lambda_{1,\eta}(\mathcal{S}(\mathbf{y})) = \mathbf{x}^\sharp$, where

$$(8) \qquad (\mathbf{x}^\sharp, \mathbf{e}^\sharp) := \underset{(\mathbf{z},\mathbf{w}) \in \mathbb{R}^N \times \mathbb{R}^m}{\operatorname{argmin}} \ \|\mathbf{z}\|_1 \qquad \text{subject to} \quad \begin{cases} \mathcal{S}(\mathbf{A}\mathbf{z} + \mathbf{w}) = \mathcal{S}(\mathbf{y}), \\ \|\mathbf{w}\|_2 \le \gamma\,\eta, \\ \|\mathbf{z}\|_2 \le \gamma'\mu\sqrt{m}. \end{cases}$$

We will explain later how the parameters $\gamma, \gamma' > 0$ are chosen. It is not clear if the last constraint, namely $\|\mathbf{z}\|_2 \le \gamma'\mu\sqrt{m}$, is absolutely necessary, but it certainly eases the argument. We emphasize that our recovery procedure is (recast as) a second-order cone program. Indeed, the first constraint, namely $\mathcal{S}(\mathbf{Az}+\mathbf{w}) = \mathcal{S}(\mathbf{y})$, can be written as a set of linear constraints. To see this, we introduce the index sets of nonsaturated and of saturated measurements as defined by

$$(9) \qquad \mathcal{I}_{\rm nonsat} := \{i \in [\![1:m]\!] : |y_i| < \mu\} \qquad \text{and} \qquad \mathcal{I}_{\rm sat} := \{i \in [\![1:m]\!] : |y_i| \geq \mu\},$$

where the latter is further decomposed into index sets of negatively and of positively saturated measurements, i.e.,

$$(10) \qquad \mathcal{I}_{\rm negsat} := \{i \in [\![1:m]\!] : y_i \leq -\mu\} \qquad \text{and} \qquad \mathcal{I}_{\rm possat} := \{i \in [\![1:m]\!] : y_i \geq +\mu\},$$

and we can observe that

$$(11) \qquad [\mathcal{S}(\mathbf{Az}+\mathbf{w}) = \mathcal{S}(\mathbf{y})] \iff \begin{cases} \langle \mathbf{a}_i, \mathbf{z}\rangle + w_i \;=\; \quad y_i, & i \in \mathcal{I}_{\rm nonsat}, \\ \langle \mathbf{a}_i, \mathbf{z}\rangle + w_i \leq -\mu, & i \in \mathcal{I}_{\rm negsat}, \\ \langle \mathbf{a}_i, \mathbf{z}\rangle + w_i \geq +\mu, & i \in \mathcal{I}_{\rm possat}. \end{cases}$$

Moreover, it is usual to deal with the minimization of the ℓ_1-norm $\|\mathbf{z}\|_1$ by introducing a slack vector $\mathbf{c} \in \mathbb{R}^N$ and minimizing $\sum_{j=1}^N c_j$ subject to $|z_j| \leq c_j$, i.e., $-c_j \leq z_j \leq c_j$, for all $j \in [\![1:N]\!]$.

In order to establish the results outlined above, we need several properties of Gaussian matrices. These properties are formulated in Section 2 below. Then, in Section 3, we show that estimate (6) is valid in the small magnitude regime for $\|\mathbf{x}\|_2$. In Section 4, we consider the intermediate magnitude regime for $\|\mathbf{x}\|_2$ and show that estimate (7) is valid, too. In Section 5, we discuss a procedure different from (8), for which we establish a recovery result in the small magnitude regime only. Section 6 reports on some modest numerical experiments. Finally, an appendix collects justifications of some unproven statements made in Section 2.

2 Technical Tools

The arguments presented in Sections 3, 4, and 5 below rely on several properties of a random matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ populated by independent $\mathcal{N}(0, 1/m)$ entries. These properties are explicitly stated in this section. They involve some constants whose values are fixed for the remainder of the paper.

2.1 Standard restricted isometry property and consequences

A matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ is said to satisfy the restricted isometry property of order s with constant $\delta \in (0,1)$ if

$$(12) \qquad (1-\delta)\|\mathbf{z}\|_2^2 \leq \|\mathbf{Az}\|_2^2 \leq (1+\delta)\|\mathbf{z}\|_2^2 \qquad \text{for all } s\text{-sparse } \mathbf{z} \in \mathbb{R}^N.$$

It is known (see e.g. [8, Theorems 9.2 and 9.27]) that there exist absolute constants $\kappa_1, \upsilon_1, c_1 > 0$ such that, if $m \geq \kappa_1 \delta^{-2} s \ln(eN/s)$, then a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ (meaning here that its entries are independent Gaussian random variables with mean zero and variance $1/m$) satisfies the restricted isometry property of order s with constant $\delta \in (0,1)$ with failure probability at most $\upsilon_1 \exp(-c_1 \delta^2 m)$. We will also invoke the two consequences of this property formulated below.

1. Restricted isometry property for effectively s-sparse vectors (i.e., vectors $\mathbf{z} \in \mathbb{R}^N$ such that $\|\mathbf{z}\|_1 \leq \sqrt{s}\|\mathbf{z}\|_2$): there exist absolute constants $\kappa_2, \upsilon_2, c_2 > 0$ such that, if $m \geq \kappa_2 \delta^{-2} s \ln(eN/s)$, then a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ satisfies

$$(13) \qquad (1-\delta)\|\mathbf{z}\|_2^2 \leq \|\mathbf{A}\mathbf{z}\|_2^2 \leq (1+\delta)\|\mathbf{z}\|_2^2 \qquad \text{for all effectively } s\text{-sparse } \mathbf{z} \in \mathbb{R}^N$$

with failure probability at most $\upsilon_2 \exp(-c_2 \delta^2 m)$. Property (13) does follow from Property (12) using the sort-and-split technique. A proof of this somewhat folklore implication is included in the appendix.

2. Democratic robust null space property: there exist absolute constants $\kappa_3, \upsilon_3, c_3, C_3, D_3, \rho_3 > 0$ such that, if $m \geq \kappa_3 s \ln(eN/s)$, then a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ allows for the property

$$(14) \qquad \|\mathbf{x}-\mathbf{z}\|_2 \leq \frac{C_3}{\sqrt{s}}\left(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\sigma_s(\mathbf{x})_1\right) + D_3\|\mathbf{A}_{\mathcal{I}}(\mathbf{x}-\mathbf{z})\|_2, \qquad \mathbf{x}, \mathbf{z} \in \mathbb{R}^N,$$

to hold simultaneously for all sets $\mathcal{I} \subseteq [\![1:m]\!]$ of size $m' \geq (1-\rho_3)m$ with failure probability at most $\upsilon_3 \exp(-c_3 m)$. Here, the notation $\sigma_s(\mathbf{x})_1$ stands for the error of best s-term approximation to $\mathbf{x}$ in ℓ_1-norm, i.e.,

$$(15) \qquad \sigma_s(\mathbf{x})_1 := \min\{\|\mathbf{x}-\mathbf{z}\|_1, \mathbf{z} \in \mathbb{R}^N \text{ is } s\text{-sparse}\},$$

and the matrix $\mathbf{A}_{\mathcal{I}}$ represents the row-submatrix of $\mathbf{A}$ indexed by $\mathcal{I}$. The above property appeared implicitly in [7] and the name 'democratic' is inspired by [10]. Its proof involves a simple union bound, as outlined in the appendix.

2.2 Saturated restricted isometry property

An essential tool put forward in [7] was a version of the restricted isometry property for saturated measurements. Precisely, [7, Lemma 8] established that, given $\delta \in (0,1)$ and $\beta_0 > \alpha_0 > 0$, there are positive constants $\kappa' = \kappa'(\alpha_0, \beta_0)$, $c' = c'(\beta_0)$, and υ' such that, if $m \geq \kappa' \delta^{-4} s \ln(eN/s)$ and if $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a random matrix populated by independent $\mathcal{N}(0, 1/m)$ entries, then with probability at most $1 - \upsilon' \exp(-c'\delta^2 m)$, one has

$$(16) \qquad (1-\delta)\, m\, \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{u}\|_2}{\sqrt{m}}\right) \leq \|\mathcal{S}(\mathbf{A}\mathbf{u})\|_1 \leq (1+\delta)\, m\, \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{u}\|_2}{\sqrt{m}}\right)$$

for all effectively s-sparse vectors $\mathbf{u} \in \mathbb{R}^N$ satisfying $\alpha_0 \mu < \sigma\|\mathbf{u}\|_2 \leq \beta_0 \mu$. The explicit expression of the function $\widetilde{\mathcal{S}}$ is not really important — it is given in [7], along with a couple of useful properties.

The one property we shall exploit here is the fact that, for any $\beta' > 0$, there is a constant $\nu_{\beta'} > 0$ such that

$$\frac{d\widetilde{\mathcal{S}}(t)}{dt} \geq \nu_{\beta'} \qquad \text{whenever } t \in [0, \beta'\mu]. \tag{17}$$

2.3 Tessellation of the 'effectively sparse sphere' under perturbation

Early theoretical works such as [9, 11] made it apparent that the one-bit compressive sensing problem is intimately connected to a tessellation property for the sets of ℓ_2-normalized genuinely and effectively sparse vectors. A stronger tessellation property for arbitrary sets was studied in [12], but it does not seem to incorporate the result needed here, except in case there is no prequantization error ($\mathbf{e} = \mathbf{e}' = \mathbf{0}$). Precisely, we need the fact that there are absolute constants $\kappa'', \upsilon'', c'', d'', \omega'' > 0$ such that, if $m \geq \kappa''\delta^{-9} s \ln(eN/s)$ and if $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a random matrix populated by independent $\mathcal{N}(0, 1/m)$ entries, then with probability at least $1 - \upsilon'' \exp(-c''\delta^2 m)$, for all effectively s-sparse $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$ and all $\mathbf{e}, \mathbf{e}' \in \mathbb{R}^m$ satisfying $\|\mathbf{e}\|_2 \leq \omega''\|\mathbf{x}\|_2$ and $\|\mathbf{e}'\|_2 \leq \omega''\|\mathbf{x}'\|_2$, one has

$$\Big[\mathrm{sgn}(\mathbf{A}\mathbf{x} + \mathbf{e}) = \mathrm{sgn}(\mathbf{A}\mathbf{x}' + \mathbf{e}')\Big] \implies \left[\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}'}{\|\mathbf{x}'\|_2} \right\|_2 \leq \delta + d'' \max\left\{ \frac{\|\mathbf{e}\|_2}{\|\mathbf{x}\|_2}, \frac{\|\mathbf{e}'\|_2}{\|\mathbf{x}'\|_2} \right\}\right]. \tag{18}$$

The proof of this statement is included in the appendix.

2.4 Simultaneous (ℓ_2, ℓ_1)-quotient property

In standard compressive sensing, the so-called quotient property introduced in [13] is crucial to establish that ℓ_1-minimization is robust in the absence of an a priori bound on the measurement error. A slightly stronger statement is valid, namely there are absolute constants $c''', d''' > 0$ such that, if $N \geq 2m$, then a Gaussian matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ satisfies, with failure probability at most $\exp(-c'''m)$, the property that

$$\text{for all } \mathbf{e} \in \mathbb{R}^m, \text{ there exists } \mathbf{u} \in \mathbb{R}^N \text{ with } \mathbf{A}\mathbf{u} = \mathbf{e} \text{ and } \begin{cases} \|\mathbf{u}\|_2 \leq d'''\|\mathbf{e}\|_2, \\ \|\mathbf{u}\|_1 \leq d'''\sqrt{s_*}\|\mathbf{e}\|_2, \end{cases} \tag{19}$$

where the notation $s_* := m/\ln(eN/m)$ has been used. This statement is obtained by combining Theorem 6.13, Lemma 11.16, and Theorem 11.19 of [8].

3 Small Magnitude Regime

In this section, we place ourselves in the situation where

$$\|\mathbf{x}\|_2 \leq \alpha\mu\sqrt{m}, \tag{20}$$

with constant $\alpha > 0$ selected as (recall the introduction of the constant ρ_3 in Subsection 2.1)

$$\alpha := \frac{\sqrt{\rho_3}}{2}, \tag{21}$$

and with parameters $\gamma, \gamma' > 0$ chosen to satisfy

$$\gamma \geq \max\{1, \theta\}, \quad \theta := \frac{8}{3\sqrt{\rho_3}}, \qquad \text{and} \qquad \gamma' \geq \alpha. \tag{22}$$

The performance of the recovery procedure introduced in (8) is attested by the following result, which generalizes the one announced in (6) from genuine sparsity to effective sparsity.

Proposition 1. Suppose that $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a random matrix populated by independent Gaussian entries with mean zero and standard deviation $1/\sqrt{m}$. There are absolute constants $\kappa, \upsilon, c, C, D > 0$ such that, if $m \geq \kappa s \ln(eN/s)$, then with probability at least $1 - \upsilon \exp(-cm)$, one has

$$\|\mathbf{x} - \Lambda_{1,\eta}(\mathcal{S}(\mathbf{A}\mathbf{x} + \mathbf{e}))\|_2 \leq \frac{C}{\sqrt{s}} \sigma_s(\mathbf{x})_1 + D\eta \tag{23}$$

for all effectively s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ satisfying $\|\mathbf{x}\|_2 \leq \alpha\mu\sqrt{m}$ and all $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\|_2 \leq \eta$.

Proof. We separate two cases based on the size of the presaturation error.

Case 1: $\eta \geq \dfrac{\mu\sqrt{m}}{\theta}$.

Let $\mathbf{v} \in \mathbb{R}^m$ be defined by $v_i = y_i$ for $i \in \mathcal{I}_{\rm nonsat}$, $v_i = -\mu$ for $i \in \mathcal{I}_{\rm negsat}$, and $v_i = +\mu$ for $i \in \mathcal{I}_{\rm possat}$. We claim that the couple $(\mathbf{0}, \mathbf{v}) \in \mathbb{R}^N \times \mathbb{R}^m$ is feasible for the optimization program of (8). Indeed, if $(\mathbf{z}, \mathbf{w}) = (\mathbf{0}, \mathbf{v})$, then the constraints are satisfied by virtue of

$$\langle \mathbf{a}_i, \mathbf{z} \rangle + w_i = v_i \begin{cases} = \quad y_i, & i \in \mathcal{I}_{\rm nonsat}, \\ \leq -\mu, & i \in \mathcal{I}_{\rm negsat}, \\ \geq +\mu, & i \in \mathcal{I}_{\rm possat}, \end{cases} \tag{24}$$

as well as $\|\mathbf{w}\|_2 \leq \mu\sqrt{m} \leq \theta\eta \leq \gamma\eta$ and of course $\|\mathbf{z}\|_2 = 0 \leq \gamma'\mu\sqrt{m}$. This implies that $\mathbf{x}^\sharp = \mathbf{0}$, and hence that

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 = \|\mathbf{x}\|_2 \leq \alpha\mu\sqrt{m} \leq \alpha\theta\eta, \tag{25}$$

which implies the required estimate (23) with $D = \alpha\theta = 4/3$.

Case 2: $\eta < \dfrac{\mu\sqrt{m}}{\theta}$.

With κ chosen as $\kappa := \max\{\kappa_2, \kappa_3\}$, the assumption $m \geq \kappa s \ln(eN/s)$ guarantees that both the restricted isometry property (13) of order s with constant $\delta = 9/16$ (so that $\sqrt{1+\delta} = 5/4$)

and the democratic robust null space property (14) of order s hold, which occurs with failure probability at most $\upsilon_2 \exp(-c_2(9/16)^2 m) + \upsilon_3 \exp(-c_3 m) \le \upsilon \exp(-cm)$, where $\upsilon := \upsilon_2 + \upsilon_3$ and $c := \min\{c_1(9/16)^2, c_3\}$. In view of (13), we have on the one hand

$$\|\mathbf{A}\mathbf{x}\|_2 \le \sqrt{1+\delta}\|\mathbf{x}\|_2 \le \frac{5}{4}\alpha\mu\sqrt{m}. \tag{26}$$

On the other hand, with m_{sat} denoting the cardinality of $\mathcal{I}_{\text{sat}}$, we also have

$$\|\mathbf{A}\mathbf{x}\|_2 = \|\mathbf{y}-\mathbf{e}\|_2 \ge \|\mathbf{y}\|_2 - \|\mathbf{e}\|_2 \ge \|\mathbf{y}_{\mathcal{I}_{\text{sat}}}\|_2 - \eta \ge \mu\sqrt{m_{\text{sat}}} - \frac{\mu\sqrt{m}}{\theta}. \tag{27}$$

By combining (26) and (27), we derive

$$\sqrt{m_{\text{sat}}} \le \left(\frac{5}{4}\alpha + \frac{1}{\theta}\right)\sqrt{m} = \sqrt{\rho_3}\sqrt{m}, \qquad \text{i.e.,} \quad m_{\text{nonsat}} = m - m_{\text{sat}} \ge (1-\rho_3)m. \tag{28}$$

Thus, we can apply (14) to obtain

$$\|\mathbf{x}-\mathbf{x}^\sharp\|_2 \le \frac{C_3}{\sqrt{s}}(\|\mathbf{x}^\sharp\|_1 - \|\mathbf{x}\|_1 + 2\sigma_s(\mathbf{x})_1) + D_3\|\mathbf{A}_{\mathcal{I}_{\text{nonsat}}}(\mathbf{x}-\mathbf{x}^\sharp)\|_2. \tag{29}$$

We notice that $\|\mathbf{x}^\sharp\|_1 \le \|\mathbf{x}\|_1$ because the couple $(\mathbf{x},\mathbf{e}) \in \mathbb{R}^N \times \mathbb{R}^m$ is feasible for the optimization program (8). We also notice that

$$\begin{aligned}\|\mathbf{A}_{\mathcal{I}_{\text{nonsat}}}(\mathbf{x}-\mathbf{x}^\sharp)\|_2 &\le \|\mathbf{y}_{\mathcal{I}_{\text{nonsat}}} - \mathbf{A}_{\mathcal{I}_{\text{nonsat}}}\mathbf{x}\|_2 + \|\mathbf{y}_{\mathcal{I}_{\text{nonsat}}} - \mathbf{A}_{\mathcal{I}_{\text{nonsat}}}\mathbf{x}^\sharp\|_2 \\ &= \|\mathbf{e}_{\mathcal{I}_{\text{nonsat}}}\|_2 + \|\mathbf{e}^\sharp_{\mathcal{I}_{\text{nonsat}}}\|_2 \le \|\mathbf{e}\|_2 + \|\mathbf{e}^\sharp\|_2 \le \eta + \gamma\eta = (1+\gamma)\eta.\end{aligned} \tag{30}$$

All in all, we arrive at

$$\|\mathbf{x}-\mathbf{x}^\sharp\|_2 \le \frac{2C_3}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D_3(1+\gamma)\eta, \tag{31}$$

which is the required estimate (23) with $C = 2C_3$ and $D = D_3(1+\gamma)$. □

4 Intermediate Magnitude Regime

In this section, we place ourselves in the situation where

$$\alpha\mu\sqrt{m} < \|\mathbf{x}\|_2 \le \beta\mu\sqrt{m}. \tag{32}$$

The constant $\alpha > 0$ is the same as in Section 3, and the constant $\beta > \alpha$ is arbitrary, although we think of it as a fixed multiple of α. The parameters $\gamma, \gamma' > 0$ are chosen to satisfy

$$\gamma \ge 1 \qquad \text{and} \qquad \gamma' \ge \beta. \tag{33}$$

The following result quantifies the performance of our recovery procedure, again generalizing the result announced in (7) from genuine sparsity to effective sparsity.

Proposition 2. Let $\delta \in (0, \delta_*)$, $\delta_* := \min\{1/2, 3\alpha\nu_{\gamma'}/64\}$. Suppose that $\mathbf{A} \in \mathbb{R}^{m \times N}$ is a random matrix populated by independent Gaussian entries with mean zero and standard deviation $1/\sqrt{m}$. There are constants $\upsilon > 0$, $\kappa, c, C > 0$ depending on γ', and $D > 0$ depending on γ and γ' such that, if $m \geq \kappa \delta^{-9} s \ln(eN/s)$, then with probability at least $1 - \upsilon \exp(-c\delta^2 m)$, one has

$$\|\mathbf{x} - \Lambda_{1,\eta}(\mathcal{S}(\mathbf{A}\mathbf{x} + \mathbf{e}))\|_2 \leq C\delta\|\mathbf{x}\|_2 + D\eta \tag{34}$$

for all effectively s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ satisfying $\alpha\mu\sqrt{m} \leq \|\mathbf{x}\|_2 \leq \beta\mu\sqrt{m}$ and all $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\|_2 \leq \eta$.

Proof. We again separate two cases based on the size of the presaturation error. In what follows, the constant $\theta > 0$ is chosen large enough to ensure that

$$\frac{\gamma + 1}{\nu_{\gamma'}\alpha\theta} \leq \frac{1}{8} \qquad \text{and} \qquad \frac{4\gamma}{\alpha\theta} \leq \omega''. \tag{35}$$

Case 1: $\eta \geq \dfrac{\mu\sqrt{m}}{\theta}$.

We simply take into account the constraint $\|\mathbf{x}^\sharp\|_2 \leq \gamma'\mu\sqrt{m}$, while also keeping in mind that $\|\mathbf{x}\|_2 \leq \beta\mu\sqrt{m}$, to arrive at

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{x}^\sharp\|_2 \leq (\beta + \gamma')\mu\sqrt{m} \leq 2\gamma'\theta\eta, \tag{36}$$

which implies the required estimate (34) with $D = 2\gamma'\theta$.

Case 2: $\eta < \dfrac{\mu\sqrt{m}}{\theta}$.

With $\kappa := 16\max\{\kappa'(\alpha/4, \gamma'), \kappa''\}$, the assumption $m \geq \kappa\delta^{-9} s \ln(eN/s)$ guarantees that the saturated restricted isometry property (16) holds for effectively $16s$-sparse vectors in the magnitude range $[\alpha/4\mu\sqrt{m}, \gamma'\mu\sqrt{m}]$ and that the tessellation property (18) holds for effectively $16s$-sparse vectors. The failure probability is at most $\upsilon' \exp(-c'\delta^2 m) + \upsilon'' \exp(-c''\delta^2 m) \leq \upsilon \exp(-c\delta^2 m)$, with $\upsilon = \upsilon' + \upsilon''$ and $c = \min\{c', c''\}$. We now shall proceed in three steps, namely:

(i) proving that $\mathbf{x}^\sharp$ is effectively $16s$-sparse and satisfies

$$\alpha_0\mu\sqrt{m} < \|\mathbf{x}^\sharp\|_2 \leq \beta_0\mu\sqrt{m}, \tag{37}$$

where $\alpha_0 := \alpha/4$ and $\beta_0 := \gamma'$;

(ii) proving that the magnitudes of $\mathbf{x}$ and of $\mathbf{x}^\sharp$ are close, in the sense that

$$\Big| \|\mathbf{x}\|_2 - \|\mathbf{x}^\sharp\|_2 \Big| \leq C_1\delta\|\mathbf{x}\|_2 + D_1\eta \tag{38}$$

for some constants $C_1, D_1 > 0$;

(iii) proving that the directions of $\mathbf{x}$ and of $\mathbf{x}^\sharp$ are close, in the sense that

$$\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}^\sharp}{\|\mathbf{x}^\sharp\|_2} \right\|_2 \leq C_2 \delta + D_2 \frac{\eta}{\|\mathbf{x}\|_2} \tag{39}$$

for some constants $C_2, D_2 > 0$.

The required estimate (34) will follow from

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq \left| \|\mathbf{x}\|_2 - \|\mathbf{x}^\sharp\|_2 \right| + \|\mathbf{x}\|_2 \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}^\sharp}{\|\mathbf{x}^\sharp\|_2} \right\|_2 \leq C\delta \|\mathbf{x}\|_2 + D\eta \tag{40}$$

with $C = C_1 + C_2$ and $D = D_1 + D_2$. Thus, it remains to establish (i), (ii), and (iii). The proofs of (i) and (ii) are intertwined in order to assemble a shorter argument centered around the vector

$$\mathbf{x}^\flat := \begin{cases} \mathbf{x}^\sharp & \text{if } \|\mathbf{x}^\sharp\|_2 \geq \dfrac{\|\mathbf{x}\|_2}{2}, \\ \dfrac{\mathbf{x} + \mathbf{x}^\sharp}{2} & \text{if } \|\mathbf{x}^\sharp\|_2 < \dfrac{\|\mathbf{x}\|_2}{2}. \end{cases} \tag{41}$$

We note, on the one hand, that $\|\mathbf{x}^\sharp\|_1 \leq \|\mathbf{x}\|_1$ because $(\mathbf{x}, \mathbf{e})$ is feasible for the optimization program in (8), and consequently $\|\mathbf{x}^\flat\|_1 \leq \|\mathbf{x}\|_1$, too. On the other hand, we observe that

$$\|\mathbf{x}^\flat\|_2 \geq \begin{cases} \|\mathbf{x}^\sharp\|_2 \geq \dfrac{\|\mathbf{x}\|_2}{2} & \text{if } \|\mathbf{x}^\sharp\|_2 \geq \dfrac{\|\mathbf{x}\|_2}{2}, \\ \dfrac{\|\mathbf{x}\|_2 - \|\mathbf{x}^\sharp\|_2}{2} > \dfrac{\|\mathbf{x}\|_2}{4} & \text{if } \|\mathbf{x}^\sharp\|_2 < \dfrac{\|\mathbf{x}\|_2}{2}. \end{cases} \tag{42}$$

Therefore, it is always the case that $\mathbf{x}^\flat$ is effectively $16s$-sparse, since

$$\frac{\|\mathbf{x}^\flat\|_1}{\|\mathbf{x}^\flat\|_2} \leq \frac{\|\mathbf{x}\|_1}{\|\mathbf{x}\|_2/4} \leq 4\sqrt{s}. \tag{43}$$

Moreover, we have $\|\mathbf{x}^\flat\|_2 \geq \|\mathbf{x}\|_2/4 > (\alpha/4)\mu\sqrt{m}$, as well as $\|\mathbf{x}^\flat\|_2 \leq \max\{\|\mathbf{x}\|_2, \|\mathbf{x}^\sharp\|_2\} \leq \gamma' \mu \sqrt{m}$. At this point, we have almost established (i), except that it concerns $\mathbf{x}^\flat$ instead of $\mathbf{x}^\sharp$. Still, the saturated restricted isometry property (16) allows us to write

$$(1-\delta)\, m\, \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}^\flat\|_2}{\sqrt{m}} \right) \leq \|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 \leq (1+\delta)\, m\, \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}^\flat\|_2}{\sqrt{m}} \right). \tag{44}$$

Similarly, because $\mathbf{x}$ itself is effectively $16s$-sparse and in the magnitude range $[\alpha_0 \mu \sqrt{m}, \beta_0 \mu \sqrt{m}]$, the saturated restricted isometry property (16) also allows us to write

$$(1-\delta)\, m\, \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}\|_2}{\sqrt{m}} \right) \leq \|\mathcal{S}(\mathbf{A}\mathbf{x})\|_1 \leq (1+\delta)\, m\, \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}\|_2}{\sqrt{m}} \right). \tag{45}$$

With $\mathbf{e}^\flat \in \mathbb{R}^N$ defined as $\mathbf{e}^\flat = \mathbf{e}^\sharp$ if $\|\mathbf{x}^\sharp\|_2 \geq \|\mathbf{x}\|_2/2$ and $\mathbf{e}^\flat = (\mathbf{e}+\mathbf{e}^\sharp)/2$ if $\|\mathbf{x}^\sharp\|_2 < \|\mathbf{x}\|_2/2$, we have

$$\begin{aligned}(46)\qquad \Big|\|\mathcal{S}(\mathbf{y})\|_1 - \|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1\Big| &\leq \Big\|\mathcal{S}(\mathbf{y}) - \mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\Big\|_1 = \Big\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat + \mathbf{e}^\flat) - \mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\Big\|_1 \\ &= \sum_{i=1}^m \Big|\mathcal{S}(\langle \mathbf{a}_i, \mathbf{x}^\flat\rangle + e_i^\flat) - \mathcal{S}(\langle \mathbf{a}_i, \mathbf{x}^\flat\rangle)\Big| \leq \sum_{i=1}^m \mathrm{Lip}(\mathcal{S})|e_i^\flat| = \sum_{i=1}^m |e_i^\flat| \leq \sqrt{m}\|\mathbf{e}^\flat\|_2 \\ &\leq \sqrt{m}\max\{\|\mathbf{e}\|_2, \|\mathbf{e}^\sharp\|_2\} \leq \sqrt{m}\gamma\eta.\end{aligned}$$

In a similar fashion, we would derive

$$(47)\qquad \Big|\|\mathcal{S}(\mathbf{y})\|_1 - \|\mathcal{S}(\mathbf{A}\mathbf{x})\|_1\Big| \leq \sqrt{m}\eta.$$

Therefore, we obtain

$$(48)\qquad \Big|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 - \|\mathcal{S}(\mathbf{A}\mathbf{x})\|_1\Big| \leq (\gamma+1)\sqrt{m}\eta.$$

In view of (44) and (45), it follows that

$$\begin{aligned}(49)\qquad \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}^\flat\|_2}{\sqrt{m}}\right) - \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}\|_2}{\sqrt{m}}\right) &\leq \frac{1}{1-\delta}\frac{1}{m}\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 - \frac{1}{1+\delta}\frac{1}{m}\|\mathcal{S}(\mathbf{A}\mathbf{x})\|_1 \\ &= \frac{2\delta}{1-\delta^2}\frac{1}{m}\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 + \frac{1}{1+\delta}\frac{1}{m}\Big(\|\mathcal{S}(\mathbf{A}\mathbf{x}^\flat)\|_1 - \|\mathcal{S}(\mathbf{A}\mathbf{x})\|_1\Big) \\ &\leq \frac{2\delta}{1-\delta^2}\mu + \frac{1}{1+\delta}\frac{(\gamma+1)\eta}{\sqrt{m}}.\end{aligned}$$

This bound is in fact valid not only for the left-hand side itself but also for its absolute value, because we can exchange the roles of $\mathbf{x}^\flat$ and $\mathbf{x}$ in the argument. Note that (17) also implies

$$(50)\qquad \left|\widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}^\flat\|_2}{\sqrt{m}}\right) - \widetilde{\mathcal{S}}\left(\frac{\|\mathbf{x}\|_2}{\sqrt{m}}\right)\right| \geq \nu_{\gamma'}\left|\frac{\|\mathbf{x}^\flat\|_2}{\sqrt{m}} - \frac{\|\mathbf{x}\|_2}{\sqrt{m}}\right|,$$

hence we can deduce that

$$(51)\qquad \Big|\|\mathbf{x}^\flat\|_2 - \|\mathbf{x}\|_2\Big| \leq \frac{2\delta\mu\sqrt{m}}{\nu_{\gamma'}(1-\delta^2)} + \frac{(\gamma+1)\eta}{\nu_{\gamma'}(1+\delta)} \leq C_1\delta\|\mathbf{x}\|_2 + D_1\eta,$$

where $C_1 := 8/(3\alpha\nu_{\gamma'})$ and $D_1 := (\gamma+1)/\nu_{\gamma'}$. We have almost established (ii), except that again it concerns $\mathbf{x}^\flat$ instead of $\mathbf{x}^\sharp$. But in reality $\mathbf{x}^\flat = \mathbf{x}^\sharp$, since $\|\mathbf{x}^\sharp\|_2 < \|\mathbf{x}\|_2/2$ cannot occur, otherwise

$$(52)\qquad \|\mathbf{x}^\flat\|_2 = \left\|\frac{\mathbf{x}+\mathbf{x}^\sharp}{2}\right\|_2 \leq \frac{\|\mathbf{x}\|_2 + \|\mathbf{x}^\sharp\|_2}{2} < \frac{3}{4}\|\mathbf{x}\|_2,$$

which would contradict

$$(53)\quad \|\mathbf{x}^\flat\|_2 \geq \|\mathbf{x}\|_2 - \Big|\|\mathbf{x}^\flat\|_2 - \|\mathbf{x}\|_2\Big| \geq \|\mathbf{x}\|_2 - (C_1\delta\|\mathbf{x}\|_2 + D_1\eta) \geq \|\mathbf{x}\|_2 - \left(C_1\delta\|\mathbf{x}\|_2 + \frac{D_1}{\alpha\theta}\|\mathbf{x}\|_2\right)$$

when $\delta \leq \delta_*$ and (35) are enforced. Thus, (i) and (ii) have now been truly established. To conclude, we have to verify that (iii) is valid. To do so, we remark that $\mathbf{x}, \mathbf{x}^\sharp \in \mathbb{R}^N$ are effectively $16s$-sparse vectors that satisfy $\mathrm{sgn}(\mathbf{Ax}+\mathbf{e}) = \mathrm{sgn}(\mathbf{Ax}^\sharp + \mathbf{e}^\sharp)$, as a consequence of $\mathcal{S}(\mathbf{Ax}+\mathbf{e}) = \mathcal{S}(\mathbf{Ax}^\sharp + \mathbf{e}^\sharp)$, and that $\|\mathbf{e}\|_2 \le \omega''\|\mathbf{x}\|_2$ and $\|\mathbf{e}^\sharp\|_2 \le \omega''\|\mathbf{x}^\sharp\|_2$, as a consequence of

$$(54) \qquad \frac{\|\mathbf{e}\|_2}{\|\mathbf{x}\|_2} \le \frac{\eta}{\alpha\mu\sqrt{m}} \le \frac{1}{\alpha\theta} \le \omega'' \qquad \text{and} \qquad \frac{\|\mathbf{e}^\sharp\|_2}{\|\mathbf{x}^\sharp\|_2} \le \frac{\gamma\eta}{\alpha_0\mu\sqrt{m}} \le \frac{4\gamma}{\alpha\theta} \le \omega''.$$

Therefore, the tessellation property (18) yields

$$(55) \qquad \left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}^\sharp}{\|\mathbf{x}^\sharp\|_2} \right\|_2 \le \delta + d'' \frac{\gamma\eta}{\alpha_0\mu\sqrt{m}} \le \delta + \frac{d''\gamma}{\alpha_0/\beta}\frac{\eta}{\|\mathbf{x}\|_2},$$

which implies (39) with $C_2 = 1$ and $D_2 = 4d''\gamma\gamma'/\alpha$. The proof is now complete. □

5 Noise-ignoring ℓ_1-minimization

The noise-cognizant procedure intoduced in (8) requires the knowledge of an upper bound η on the magnitude $\|\mathbf{e}\|_2$ of the presaturation error. Can we derive recovery estimates similar to (6) and (7) even without this knowledge? We only give a partial affirmative answer to this question by showing that (6) can still be guaranteed in the small magnitude regime. The recovery procedure consists in simply ignoring the presaturation error, i.e., pretending that $\eta = 0$, and in outputting

$$(56) \qquad \mathbf{x}^\star = \Lambda_{1,0}(\mathcal{S}(\mathbf{y})) = \underset{\mathbf{z}\in\mathbb{R}^N}{\mathrm{argmin}}\ \|\mathbf{z}\|_1 \qquad \text{subject to} \quad \begin{cases} \mathcal{S}(\mathbf{Az}) = \mathcal{S}(\mathbf{y}), \\ \|\mathbf{z}\|_2 \le \gamma''\mu\sqrt{m}, \end{cases}$$

even if the sparse vector $\mathbf{x} \in \mathbb{R}^N$ to be recovered from $\mathcal{S}(\mathbf{y})$, $\mathbf{y} = \mathbf{Ax}+\mathbf{e}$, is not feasible for this optimization program. With θ still denoting the constant appearing in (22), the parameter γ'' is chosen to satisfy

$$(57) \qquad \gamma'' \ge \max\left\{ d''', \alpha + \frac{d'''}{\theta} \right\}.$$

Proposition 3. Suppose that $\mathbf{A} \in \mathbb{R}^{m\times N}$ is a random matrix populated by independent Gaussian entries with mean zero and standard deviation $1/\sqrt{m}$. There are absolute constants $\kappa', \upsilon, c, C, D > 0$ such that, if $N \ge 2m$ and $m \ge \kappa' s \ln(eN/s)$, then with probability at least $1 - \upsilon\exp(-cm)$, one has

$$(58) \qquad \|\mathbf{x} - \Lambda_{1,0}(\mathcal{S}(\mathbf{Ax}+\mathbf{e}))\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\|\mathbf{e}\|_2$$

for all effectively s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ satisfying $\|\mathbf{x}\|_2 \le \alpha\mu\sqrt{m}$ and all $\mathbf{e} \in \mathbb{R}^m$.

Proof. First of all, we remark that $\mathbf{x}^\star$ is properly defined, i.e., that the optimization program in (56) is feasible. To see this, we observe that the hypotheses of Proposition 3 guarantee the validity of the simultaneous (ℓ_2, ℓ_1)-quotient property (19), which we invoke to justify the existence of $\mathbf{z} \in \mathbb{R}^N$ such that $\mathbf{Az} = \mathcal{S}(\mathbf{y})$, hence $\mathcal{S}(\mathbf{Az}) = \mathcal{S}(\mathbf{y})$, and

$$\|\mathbf{z}\|_2 \le d''' \|\mathcal{S}(\mathbf{y})\|_2 \le d''' \mu\sqrt{m} \le \gamma'' \mu \sqrt{m}. \tag{59}$$

At this point, we still separate two cases based on the size of the presaturation error.

Case 1: $\|\mathbf{e}\|_2 \ge \dfrac{\mu\sqrt{m}}{\theta}$.

We simply that into account the constraint $\|\mathbf{x}^\star\|_2 \le \gamma''\mu\sqrt{m}$, while also keeping in mind that $\|\mathbf{x}\|_2 \le \alpha\mu\sqrt{m}$, to obtain

$$\|\mathbf{x} - \mathbf{x}^\star\|_2 \le \|\mathbf{x}\|_2 + \|\mathbf{x}^\star\|_2 \le (\alpha + \gamma'')\mu\sqrt{m} \le (\alpha+\gamma'')\theta\|\mathbf{e}\|_2, \tag{60}$$

which implies the required estimate (58) with $D := (\alpha + \gamma'')\theta$.

Case 2: $\|\mathbf{e}\|_2 < \dfrac{\mu\sqrt{m}}{\theta}$.

We observe that $m \ge \kappa' s \ln(eN/s)$ implies $m \ge \kappa' s \ln(eN/m)$, i.e., that $s \le \lfloor s_*/\kappa' \rfloor =: s'$. Thus, by monotonicity of the right-hand side of (58), it is enough to prove the result for $s = s'$. By choosing the constant $\kappa' > 0$ large enough to have $\kappa'/(1+\ln(\kappa')) \ge \kappa$, where $\kappa > 0$ is the constant appearing in Proposition 1, we claim that $m \ge \kappa s' \ln(eN/s')$. Indeed, from $s' \le s_*/\kappa' = m/(\kappa' \ln(eN/m))$, we derive that $s'/m \le 1/\kappa'$, and in turn

$$\begin{aligned} m &\ge \kappa' s' \ln\left(\frac{eN}{m}\right) = \kappa' s' \ln\left(\frac{eN}{s'}\right) + \kappa' m \left(\frac{s'}{m} \ln\left(\frac{s'}{m}\right)\right) \\ &\ge \kappa' s' \ln\left(\frac{eN}{s'}\right) + \kappa' m \left(\frac{1}{\kappa'} \ln\left(\frac{1}{\kappa'}\right)\right) = \kappa' s' \ln\left(\frac{eN}{s'}\right) - m\ln(\kappa'), \end{aligned} \tag{61}$$

so a rearrangement gives

$$m \ge \frac{\kappa'}{1+\ln(\kappa')} s' \ln\left(\frac{eN}{s'}\right) \ge \kappa s' \ln\left(\frac{eN}{s'}\right), \tag{62}$$

as claimed. We can now replicate the argument in the proof of Proposition 1, Case 2 (based on the restricted isometry property (13) and the democratic robust null space property (14), both of order s') until (29) to arrive at

$$\|\mathbf{x} - \mathbf{x}^\star\|_2 \le \frac{C_3}{\sqrt{s'}} (\|\mathbf{x}^\star\|_1 - \|\mathbf{x}\|_1 + 2\sigma_{s'}(\mathbf{x})_1) + D_3 \|\mathbf{A}_{\mathcal{I}_{\rm nonsat}}(\mathbf{x} - \mathbf{x}^\star)\|_2. \tag{63}$$

We again invoke the simultaneous (ℓ_2, ℓ_1)-quotient property (19) to guarantee the existence of $\mathbf{u} \in \mathbb{R}^N$ such that

$$\mathbf{Au} = \mathbf{e} \qquad \text{with} \qquad \begin{cases} \|\mathbf{u}\|_2 \le d''' \|\mathbf{e}\|_2, \\ \|\mathbf{u}\|_1 \le d''' \sqrt{s_*} \|\mathbf{e}\|_2. \end{cases} \tag{64}$$

We remark that the vector $\mathbf{x}+\mathbf{u}$ is feasible, because the first constraint is satisfied by virtue of $\mathbf{A}(\mathbf{x}+\mathbf{u}) = \mathbf{A}\mathbf{x}+\mathbf{e} = \mathbf{y}$ and the second constraint is satisfied by virtue of

$$\|\mathbf{x}+\mathbf{u}\|_2 \leq \|\mathbf{x}\|_2 + \|\mathbf{u}\|_2 \leq \alpha\mu\sqrt{m} + d'''\|\mathbf{e}\|_2 \leq \alpha\mu\sqrt{m} + d'''\frac{\mu\sqrt{m}}{\theta} \leq \gamma''\mu\sqrt{m}. \tag{65}$$

Then, the minimizing property of $\mathbf{x}^\star$ yields

$$\|\mathbf{x}^\star\|_1 \leq \|\mathbf{x}+\mathbf{u}\|_1 \leq \|\mathbf{x}\|_1 + \|\mathbf{u}\|_1 \leq \|\mathbf{x}\|_1 + d'''\sqrt{s_*}\|\mathbf{e}\|_2. \tag{66}$$

Furthermore, since the constraint $\mathcal{S}(\mathbf{A}\mathbf{x}^\star) = \mathcal{S}(\mathbf{y})$ imposes $\mathbf{A}_{\mathcal{I}_{\text{nonsat}}}\mathbf{x}^\star = \mathbf{y}_{\mathcal{I}_{\text{nonsat}}}$, we have

$$\|\mathbf{A}_{\mathcal{I}_{\text{nonsat}}}(\mathbf{x}-\mathbf{x}^\star)\|_2 = \|\mathbf{A}_{\mathcal{I}_{\text{nonsat}}}\mathbf{x} - \mathbf{y}_{\mathcal{I}_{\text{nonsat}}}\|_2 = \|\mathbf{e}_{\mathcal{I}_{\text{nonsat}}}\|_2 \leq \|\mathbf{e}\|_2. \tag{67}$$

Substituting (66) and (67) into (63) gives

$$\|\mathbf{x}-\mathbf{x}^\star\|_2 \leq \frac{C_3}{\sqrt{s'}}(d'''\sqrt{s_*}\|\mathbf{e}\|_2 + 2\sigma_{s'}(\mathbf{x})_1) + D_3\|\mathbf{e}\|_2 \leq \frac{2C_3}{\sqrt{s'}}\sigma_{s'}(\mathbf{x})_1 + \left(\frac{C_3 d'''}{\sqrt{1/(2\kappa')}} + D_3\right)\|\mathbf{e}\|_2, \tag{68}$$

which implies the required estimate (58) with $C = 2C_3$ and $D = \sqrt{2\kappa'}C_3 d''' + D_3$. □

6 Numerical Validation

We have implemented the noise-cognizant recovery procedure described in (8) using CVX [1], a package for specifying and solving convex programs. The noise-ignoring procedure (56) does not require a separate implementation, as it is just the procedure (8) with parameter γ set to zero. Our implementation can be downloaded from the first author's webpage, along with the MATLAB file containing the code to reproduce the experiments presented below. By no means do these experiments constitute an exhaustive investigation. They are included here for illustrative purposes only and one should refrain from drawing any strong conclusions out of them.

6.1 Influence of the magnitude η of the presaturation error

In a first experiment, see Figure 1, we run the noise-cognizant ℓ_1-minimization (8) on several random s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ with varying magnitude $\xi = \|\mathbf{x}\|_2$ acquired through inaccurate saturated measurements $\mathcal{S}(\mathbf{A}\mathbf{x}+\mathbf{e}) \in [-\mu, +\mu]^m$. We can discern two regimes depending on the magnitude of $\mathbf{x}$: when ξ is small, the recovery error is roughly constant and when ξ passes a certain breakpoint, the recovery error depends somewhat linearly on ξ. Our theory suggests a breakpoint independent of the bound η on the ℓ_2-norm of the presaturation error $\mathbf{e}$, which is empirically compelling when η is small. Note that the recovery error grows with η in the small magnitude regime, but that this intuitive phenomenon is dampened in the intermediate magnitude regime.

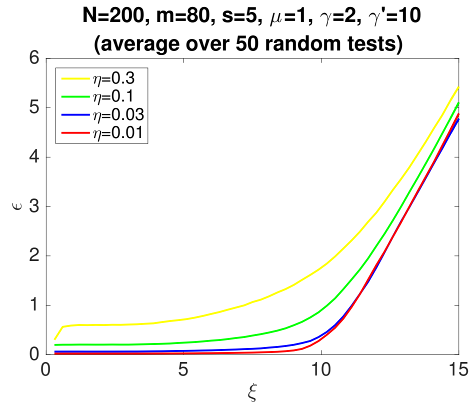
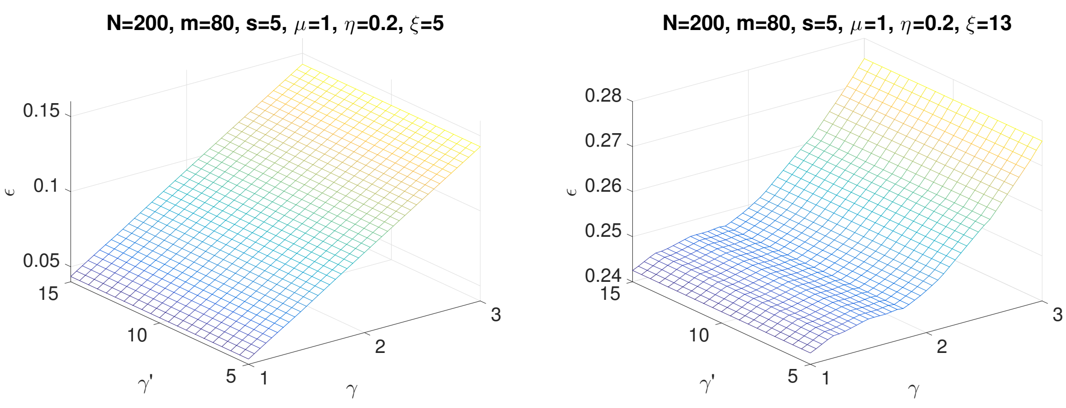


Figure 1: Recovery error $\varepsilon = \|\mathbf{x} - \mathbf{x}^\sharp\|_2$ vs magnitude $\xi = \|\mathbf{x}\|_2$ of the sparse vector to be recovered

6.2 Influence of the parameters γ and γ' of the ℓ_1-minimization

In a second experiment, we inquire into the best empirical choices of parameters γ and γ' in the noise-cognizant ℓ_1-minimization (8). With the same problem dimensions as in the previous experiment, we consider one sparse vector in the small magnitude regime and another sparse vector in the intermediate magnitude regime. For each of these two vectors, we run the noise-cognizant procedures for several different values of γ and γ' and we record the corresponding recovery errors. The results are displayed in Figure 2. In both cases, the experiments confirm the intuition that better recovery occurs when the parameter $\gamma \geq 1$ in the constraint $\|\mathbf{w}\|_2 \leq \gamma\eta$ is close to one. As for the parameter γ' in the constraint $\|\mathbf{z}\|_2 \leq \gamma'\mu\sqrt{m}$, it seems to have no influence on the accuracy of the recovery. This suggests that one could just remove this constraint from the optimization program. Unfortunately, we were unable to establish theoretical guarantees in this situation.



(a) Small magnitude regime

(b) Intermediate magnitude regime

Figure 2: Relative recovery error $\varepsilon = \|\mathbf{x} - \mathbf{x}^\sharp\|_2 / \|\mathbf{x}\|_2$ as a function of the parameters γ and γ'

6.3 Comparison of the noise-cognizant and noise-ignoring minimizations

In a third and final experiment, we test the noise-ignoring ℓ_1-minimization (56) and compare its performance with the noise-cognizant ℓ_1-minimization (8), see Figure 3. Besides the obvious advantages that (56) does not require any (over)estimation η of $\|\mathbf{e}\|_2$ to be run and that it is faster than (8) (since the optimization program features less variables and less constraints), it also appears to be slightly more accurate, at least for certain parameter values. Unfortunately, we were unable to establish full theoretical guarantees for the noise-ignoring procedure.

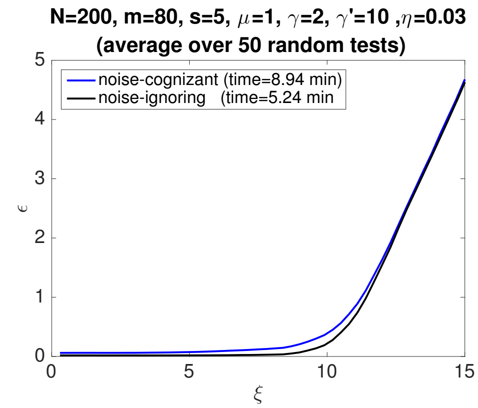


Figure 3: Recovery error $\varepsilon = \|\mathbf{x} - \hat{\mathbf{x}}\|_2$, $\hat{\mathbf{x}} \in \{\mathbf{x}^\sharp, \mathbf{x}^\star\}$, vs magnitude $\xi = \|\mathbf{x}\|_2$ of the sparse vector

References

[1] CVX Research, Inc. CVX: MATLAB software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[2] R. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters. *Exponential decay of reconstruction error from binary measurements of sparse signals.* IEEE Transactions on Information Theory 63.6 (2017): 3368–3385.

[3] P. Boufounos and R. Baraniuk. *1-bit compressive sensing.* In: Proceedings of the 42nd Annual Conference on Information Sciences and Systems (2008): 16–21.

[4] E. Candès. *Compressive sampling.* In: Proceedings of the International Congress of Mathematicians Vol. 3 (2006).

[5] D. Donoho. *Compressed sensing.* IEEE Transactions on Information Theory 52.4 (2006): 1289–1306.

[6] S. Foucart. *Flavors of compressive sensing.* In: Approximation Theory XV: San Antonio 2016, Springer Proceedings in Mathematics & Statistics 201 (2017): 61–104.

[7] S. Foucart and T. Needham. *Sparse recovery from saturated measurements.* Information and Inference 6.2 (2017): 196–212.

[8] S. Foucart and H. Rauhut. A Mathematical Introduction to Compressive Sensing. Birkhäuser (2013).

[9] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk. *Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors.* IEEE Transactions on Information Theory 59.4 (2013): 2082–2102.

[10] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk. *Democracy in action: quantization, saturation, and compressive sensing.* Applied and Computational Harmonic Analysis 31.3 (2011): 429–443.

[11] Y. Plan and R. Vershynin. *One-bit compressed sensing by linear programming.* Communications on Pure and Applied Mathematics 66.8 (2013): 1275–1297.

[12] Y. Plan and R. Vershynin. *Dimension reduction by random hyperplane tessellations.* Discrete & Computational Geometry 51.2 (2014): 438–461.

[13] P. Wojtaszczyk. *Stability and instance optimality for Gaussian measurements in compressed sensing.* Foundations of Computational Mathematics 10.1 (2010): 1–13.

Appendix

Proof of the restricted isometry property for effectively sparse vectors

We justify here that the restricted isometry property (13), valid for effectively s-sparse vectors, is a consequence of the restricted isometry property (12), valid for genuinely $2s$-sparse vectors, in which δ is replaced by $\delta/4$. To do so, given an effectively s-sparse vector $\mathbf{z} \in \mathbb{R}^N$, we use the sort-and-split technique to decompose $[\![1:N]\!]$ into index sets $S_0, S_1, S_2, \ldots$ of size s in such a way that S_0 corresponds to s largest absolute entries of $\mathbf{x}$, S_1 corresponds to s next largest absolute entries of $\mathbf{x}$, etc. We then write

$$\begin{aligned}(69)\quad \big|\|\mathbf{A}\mathbf{z}\|_2^2 - \|\mathbf{z}\|_2^2\big| &= \big|\langle(\mathbf{A}^*\mathbf{A}-\mathbf{I})\mathbf{z},\mathbf{z}\rangle\big| = \Big|\Big\langle \sum_{k\geq 0}(\mathbf{A}^*\mathbf{A}-\mathbf{I})\mathbf{z}_{S_k}, \sum_{k\geq 0}\mathbf{z}_{S_k}\Big\rangle\Big| \\ &\leq |\langle(\mathbf{A}^*\mathbf{A}-\mathbf{I})\mathbf{z}_{S_0},\mathbf{z}_{S_0}\rangle| + 2\Big|\sum_{k\geq 1}\langle(\mathbf{A}^*\mathbf{A}-\mathbf{I})\mathbf{z}_{S_0},\mathbf{z}_{S_k}\rangle\Big| + \Big|\sum_{k,\ell\geq 1}\langle(\mathbf{A}^*\mathbf{A}-\mathbf{I})\mathbf{z}_{S_k},\mathbf{z}_{S_\ell}\rangle\Big|.\end{aligned}$$

Thanks to the restricted isometry for genuinely sparse vectors, the first term in the right-hand side of (69) is bounded as

$$|\langle (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{z}_{S_0}, \mathbf{z}_{S_0} \rangle| \le \frac{\delta}{4} \|\mathbf{z}_{S_0}\|_2^2 \le \frac{\delta}{4} \|\mathbf{z}\|_2^2. \tag{70}$$

As for the second term in the right-hand side of (69), again thanks to the restricted isometry for genuinely sparse vectors, it is bounded (see [8, Proposition 6.3]) via

$$\begin{aligned} \left| \sum_{k\ge1} \langle (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{z}_{S_k}, \mathbf{z}_{S_0} \rangle \right| &\le \sum_{k\ge1} |\langle (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{z}_{S_0}, \mathbf{z}_{S_k} \rangle| \le \sum_{k\ge1} \frac{\delta}{4} \|\mathbf{z}_{S_0}\|_2 \|\mathbf{z}_{S_k}\|_2 \\ &\le \frac{\delta}{4} \|\mathbf{z}_{S_0}\|_2 \sum_{k\ge1} \frac{\|\mathbf{z}_{S_{k-1}}\|_1}{\sqrt{s}} \le \frac{\delta}{4} \|\mathbf{z}_{S_0}\|_2 \frac{\|\mathbf{z}\|_1}{\sqrt{s}} \le \frac{\delta}{4} \|\mathbf{z}_{S_0}\|_2 \|\mathbf{z}\|_2 \le \frac{\delta}{4} \|\mathbf{z}\|_2^2. \end{aligned} \tag{71}$$

Finally, the last term in the right-hand side of (69) satisfies

$$\begin{aligned} \left| \sum_{k,\ell\ge1} \langle (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{z}_{S_k}, \mathbf{z}_{S_\ell} \rangle \right| &\le \sum_{k,\ell\ge1} |\langle (\mathbf{A}^*\mathbf{A} - \mathbf{I})\mathbf{z}_{S_k}, \mathbf{z}_{S_\ell} \rangle| \le \sum_{k,\ell\ge1} \frac{\delta}{4} \|\mathbf{z}_{S_k}\|_2 \|\mathbf{z}_{S_\ell}\|_2 \\ &= \frac{\delta}{4} \left(\sum_{k\ge1} \|\mathbf{z}_{S_k}\|_2 \right)^2 \le \frac{\delta}{4} \left(\frac{\|\mathbf{z}\|_1}{\sqrt{s}} \right)^2 \le \frac{\delta}{4} \|\mathbf{z}\|_2^2. \end{aligned} \tag{72}$$

All in all, we arrive at

$$\big| \|\mathbf{A}\mathbf{z}\|_2^2 - \|\mathbf{z}\|_2^2 \big| \le \frac{\delta}{4} \|\mathbf{z}\|_2^2 + 2\frac{\delta}{4} \|\mathbf{z}\|_2^2 + \frac{\delta}{4} \|\mathbf{z}\|_2^2 = \delta \|\mathbf{z}\|_2^2, \tag{73}$$

which directly implies the required result.

Proof of the democratic robust null space property

We justify here a second statement made in Subsection 2.1, namely that the democratic robust null space property is a consequence of the restricted isometry property. So let us assume that $\mathbf{A} \in \mathbb{R}^{m\times N}$ is a Gaussian matrix with $m \ge \kappa_3 s \ln(eN/s)$, where $\kappa_3 = (81/128)\kappa_1/(1-\rho_3)$ and $\rho_3 \in (0,1)$ is to be chosen later. Setting $\delta = 9/16$, say, and fixing a set $\mathcal{I} \subseteq [\![1:m]\!]$ of size m' at least $(1-\rho_3)m$, we have $m' \ge \kappa_1 \delta^{-2}(2s)\ln(eN/2s)$, so that the matrix $\sqrt{m/m'}\mathbf{A}_\mathcal{I}$ satisfies the standard restricted isometry property (12) of order $2s$ with constant δ with failure probability at most $\upsilon_1 \exp(-c_1\delta^2 m')$. Then, as $\mathcal{I}$ is allowed to vary, all the matrices $\sqrt{m/m'}\mathbf{A}_\mathcal{I}$ satisfy the standard restricted isometry property (12) of order $2s$ with constant δ with failure probability at most

$$\begin{aligned} \sum_{m'\ge(1-\rho_3)m} \binom{m}{m'} \times \upsilon_1 \exp\left(-c_1\delta^2 m'\right) &\le \left(\frac{e}{\rho_3}\right)^{\rho_3 m} \times \upsilon_1 \exp\left(-c_1\delta^2(1-\rho_3)m\right) \\ &= \upsilon_1 \exp\left(\rho_3 m \ln\left(\frac{e}{\rho_3}\right) - c_1\delta^2(1-\rho_3)m\right) \le \upsilon_1 \exp\left(-\frac{c_1}{4}\delta^2 m\right), \end{aligned} \tag{74}$$

provided ρ_3 is chosen small enough to guarantee that $\rho_3 \ln(e/\rho_3) \le c_1\delta^2/4 \le c_1\delta^2(1-\rho_3)/2$. Then, Theorems 4.25 and 6.13 of [8] ensures that, for all $\mathcal{I} \subseteq [\![1:m]\!]$ of size $m' \ge (1-\rho_\delta)m$ and for all $\mathbf{x}, \mathbf{z} \in \mathbb{R}^N$,

$$\begin{aligned}
\|\mathbf{x}-\mathbf{z}\|_2 &\le \frac{C}{\sqrt{s}}\left(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\sigma_s(\mathbf{x})_1\right) + D\|\sqrt{m/m'}\mathbf{A}_{\mathcal{I}}(\mathbf{x}-\mathbf{z})\|_2 \\
&\le \frac{C}{\sqrt{s}}\left(\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\sigma_s(\mathbf{x})_1\right) + \frac{D}{\sqrt{1-\rho_3}}\|\mathbf{A}_{\mathcal{I}}(\mathbf{x}-\mathbf{z})\|_2,
\end{aligned} \tag{75}$$

which is Property (14) with $C_3 = C$ and $D_3 = D/\sqrt{1-\rho_3}$.

Proof of the tessellation property under perturbation

We finally justify the tessellation property (18) for the 'effectively sparse sphere' in the presence of prequantization error. To do so, we recall from [2, Theorem 3] that there are absolute constants $\kappa'', \upsilon'', c'', d'', \omega'' > 0$ such that, if $m \ge \kappa''\delta^{-7}s\ln(eN/s)$ and if $\mathbf{A} \in \mathbb{R}^{m\times N}$ is a random matrix populated by independent $\mathcal{N}(0,1/m)$ entries, then with probability at least $1-\upsilon''\exp(-c''\delta^2 m)$, for any s-sparse $\mathbf{x} \in \mathbb{R}^N$ and any $\mathbf{e} \in \mathbb{R}^m$ satisfying $\|\mathbf{e}\|_2 \le \omega''\|\mathbf{x}\|_2$, one has

$$\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - H_s'(\mathbf{A}^\top \mathrm{sgn}(\mathbf{A}\mathbf{x}+\mathbf{e}))\right\|_2 \le \delta + d''\frac{\|\mathbf{e}\|_2}{\|\mathbf{x}\|_2}, \tag{76}$$

where $H_s'(\mathbf{u})$ is the ℓ_2-normalized output of a hard thresholding procedure with input $\mathbf{u} \in \mathbb{R}^N$. We are now going to prove that this remains true (up to some changes in the constants and in the power of δ^{-1}) when $\mathbf{x} \in \mathbb{R}^N$ is effectively, rather than genuinely, s-sparse. Therefore, if $\mathbf{x}' \in \mathbb{R}^N$ is another effectively s-sparse vector and if $\mathbf{e}' \in \mathbb{R}^m$ satisfies $\|\mathbf{e}'\|_2 \le \omega''\|\mathbf{x}\|_2$, since (76) also holds for $\mathbf{x}'$ in place of $\mathbf{x}$, a triangle inequality gives (18) in the form

$$\Big[\mathrm{sgn}(\mathbf{A}\mathbf{x}+\mathbf{e}) = \mathrm{sgn}(\mathbf{A}\mathbf{x}'+\mathbf{e}')\Big] \Longrightarrow \left[\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}'}{\|\mathbf{x}'\|_2}\right\|_2 \le 2\delta + 2d''\max\left\{\frac{\|\mathbf{e}\|_2}{\|\mathbf{x}\|_2}, \frac{\|\mathbf{e}'\|_2}{\|\mathbf{x}'\|_2}\right\}\right]. \tag{77}$$

So, given an effectively s-sparse $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{e} \in \mathbb{R}^m$ with $\|\mathbf{e}\|_2 \le \omega'''\|\mathbf{x}\|_2$, ω''' to be chosen later, we use the sort-and-split technique to decompose $[\![1:N]\!]$ into index sets $T = T_0, T_1, T_2, \ldots$ of size $t = \lceil \delta^{-2}s\rceil$ in such a way that T_0 corresponds to t largest absolute entries of $\mathbf{x}$, T_1 corresponds to t next largest absolute entries of $\mathbf{x}$, etc. Interpreting $\mathrm{sgn}(\mathbf{A}\mathbf{x}+\mathbf{e})$ as $\mathrm{sgn}(\mathbf{A}\mathbf{x}_T+\widetilde{\mathbf{e}})$ with $\widetilde{\mathbf{e}} := \mathbf{A}\mathbf{x}_{\overline{T}}+\mathbf{e}$, (76) gives

$$\left\| \frac{\mathbf{x}_T}{\|\mathbf{x}_T\|_2} - H_t'(\mathbf{A}^\top \mathrm{sgn}(\mathbf{A}\mathbf{x}+\mathbf{e}))\right\|_2 \le \delta + d''\frac{\|\widetilde{\mathbf{e}}\|_2}{\|\mathbf{x}_T\|_2}, \tag{78}$$

so long as $\|\widetilde{\mathbf{e}}\|_2 \le \omega''\|\mathbf{x}_T\|_2$. But this can be fulfilled if δ and ω''' are small enough to guarantee that $5\delta/4 + \omega''' \le (1-\delta/2)\omega''$, in view of the estimates, to be established below,

$$\|\widetilde{\mathbf{e}}\|_2 \le \left(\frac{5\delta}{4} + \omega'''\right)\|\mathbf{x}\|_2, \tag{79}$$

$$\|\mathbf{x}_T\|_2 \ge \left(1 - \frac{\delta}{2}\right)\|\mathbf{x}\|_2. \tag{80}$$

To establish (80), we invoke [8, Theorem 2.5] and the effective sparsity of $\mathbf{x}$ to write

$$\|\mathbf{x}_{\overline{T}}\|_2 \le \frac{1}{2\sqrt{t}}\|\mathbf{x}\|_1 \le \frac{1}{2}\sqrt{\frac{s}{t}}\|\mathbf{x}\|_2 \le \frac{\delta}{2}\|\mathbf{x}\|_2, \tag{81}$$

so that $\|\mathbf{x}_T\|_2 \ge \|\mathbf{x}\|_2 - \|\mathbf{x}_{\overline{T}}\|_2 \ge (1-\delta/2)\,\|\mathbf{x}\|_2$, as announced. To establish (79), since we can assume that $\mathbf{A}$ has the restricted isometry property of order t with constant $9/16$, say, we have

$$\begin{aligned}\|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 &= \Big\|\sum_{k\ge1}\mathbf{A}\mathbf{x}_{T_k}\Big\|_2 \le \sum_{k\ge1}\|\mathbf{A}\mathbf{x}_{T_k}\|_2 \le \sum_{k\ge1}\sqrt{1+9/16}\|\mathbf{x}_{T_k}\|_2\\ &\le \frac{5}{4}\sum_{k\ge1}\frac{\|\mathbf{x}_{T_{k-1}}\|_1}{\sqrt{t}} \le \frac{5}{4}\frac{\|\mathbf{x}\|_1}{\sqrt{t}} \le \frac{5}{4}\sqrt{\frac{s}{t}}\|\mathbf{x}\|_2 \le \frac{5\delta}{4}\|\mathbf{x}\|_2.\end{aligned} \tag{82}$$

The estimate (79) then follows from $\|\mathbf{e}\|_2 \le \omega'''\|\mathbf{x}\|_2$ and

$$\|\widetilde{\mathbf{e}}\|_2 \le \|\mathbf{A}\mathbf{x}_{\overline{T}}\|_2 + \|\mathbf{e}\|_2 \le \frac{5\delta}{4}\|\mathbf{x}\|_2 + \|\mathbf{e}\|_2. \tag{83}$$

We have now validated (78). Next, we take into account that

$$\begin{aligned}\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{x}_T}{\|\mathbf{x}_T\|_2}\right\|_2^2 &= 2 - 2\frac{\langle \mathbf{x}, \mathbf{x}_T\rangle}{\|\mathbf{x}\|_2\|\mathbf{x}_T\|_2} = 2\frac{\|\mathbf{x}\|_2 - \|\mathbf{x}_T\|_2}{\|\mathbf{x}\|_2} = 2\frac{\|\mathbf{x}\|_2^2 - \|\mathbf{x}_T\|_2^2}{\|\mathbf{x}\|_2(\|\mathbf{x}\|_2 + \|\mathbf{x}_T\|_2)}\\ &\le \frac{2\|\mathbf{x}_{\overline{T}}\|_2^2}{2\|\mathbf{x}\|_2^2} \le \frac{s}{4t} = \frac{\delta^2}{4}.\end{aligned} \tag{84}$$

In view of (78) and of (84), a triangle inequality gives

$$\left\|\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - H_t'(\mathbf{A}^\top \mathrm{sgn}(\mathbf{A}\mathbf{x}+\mathbf{e}))\right\|_2 \le \frac{3\delta}{2} + d''\frac{\|\widetilde{\mathbf{e}}\|_2}{\|\mathbf{x}_T\|_2} \le \frac{3\delta}{2} + d''\frac{(5\delta/4)\|\mathbf{x}\|_2 + \|\mathbf{e}\|_2}{(1-\delta/2)\|\mathbf{x}\|_2} \le c\delta + d\frac{\|\mathbf{e}\|_2}{\|\mathbf{x}\|_2}, \tag{85}$$

with $c = 3/2 + 5d''/2$ and $d = 2d''$. This concludes the proof.