An IHT algorithm for sparse recovery from subexponential measurements

Simon Foucart[*], Texas A&M University
Guillaume Lecué, CREST, CNRS, Université Paris Saclay

Abstract

A matrix whose entries are independent subexponential random variables is not likely to satisfy the classical restricted isometry property in the optimal regime of parameters. However, it is known that uniform sparse recovery is still possible with high probability in the optimal regime if ones uses ℓ_1-minimization as a recovery algorithm. We show in this note that such a statement remains valid if one uses a new variation of iterative hard thresholding as a recovery algorithm. The argument is based on a modified restricted isometry property featuring the ℓ_1-norm as the inner norm.

Key words and phrases: compressive sensing, sparse recovery, iterative hard thresholding, restricted isometry property, subexponential random variable.

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

We consider the standard compressive sensing problem where one aims at recovering s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ acquired as $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^m$ with $m \ll N$. It is by now well known that if the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ satisfies the so-called restricted isometry property of order s, i.e., if

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

with some small $\delta \in (0,1)$, then this task can be carried out in a stable and robust way by a variety of recovery algorithms, e.g. basis pursuit (i.e., ℓ_1-minimization), orthogonal matching pursuit, compressive sampling matching pursuit, iterative hard thresholding, or hard thresholding pursuit, to name just a few. We refer to the textbook [5] which gathers background information in one place. It is also well known that the restricted isometry property of order s holds with high probability for random matrices $\mathbf{A} = \mathbf{B}/\sqrt{m} \in \mathbb{R}^{m \times N}$ in the optimal regime of parameters $m \asymp s\ln(eN/s)$ when $\mathbf{B} \in \mathbb{R}^{m \times N}$ is populated with independent identically distributed mean-zero subgaussian random variables normalized so that $\mathbb{E}(b_{i,j}^2) = 1$, see e.g. [5, Section 9.1]. If subgaussian random variables are replaced by subexponential (aka Ψ_1 or pregaussian) random variables, then the resulting matrix $\mathbf{A} \in \mathbb{R}^{m \times N}$ satisfies the restricted isometry property of order s with high

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

probability provided $m \asymp s\ln^2(eN/s)$ and this number of measurements cannot be reduced to match the subgaussian case, as established in [1]. Nonetheless, with $m \asymp s\ln(eN/s)$, it was shown in [4] that random matrices $\mathbf{A} = \mathbf{B}/m \in \mathbb{R}^{m\times N}$, where $\mathbf{B} \in \mathbb{R}^{m\times N}$ is populated with independent identically distributed mean-zero subexponential random variables normalized so that $\mathbb{E}(|b_{i,j}|) = 1$, do satisfy a version of the restricted isometry property modified to feature the ℓ_1-norm as the inner norm. Precisely, with failure probability at most $C\exp(-c\delta^3 m)$, one has

$$(1-\delta) /\!\!/ \mathbf{z} /\!\!/ \le \|\mathbf{A}\mathbf{z}\|_1 \le (1+\delta) /\!\!/ \mathbf{z} /\!\!/ \qquad \text{for all } s\text{-sparse vectors } \mathbf{z} \in \mathbb{R}^N, \tag{2}$$

and we shall denote by δ_s the smallest such constant $\delta \in (0,1)$. Contrary to (1), the outer norm $/\!\!/ \cdot /\!\!/$ is not the ℓ_2-norm anymore. It depends a priori on the probability measure μ associated to the subexponential distribution via

$$/\!\!/ \mathbf{z} /\!\!/ := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left| \sum_{j=1}^{N} t_j z_j \right| d\mu(t_1)\cdots d\mu(t_N). \tag{3}$$

However, it is comparable to the ℓ_2-norm in the sense that there are constants $\beta \ge \alpha > 0$ such that

$$\alpha\|\mathbf{z}\|_2 \le /\!\!/ \mathbf{z} /\!\!/ \le \beta\|\mathbf{z}\|_2 \qquad \text{for all vectors } \mathbf{z} \in \mathbb{R}^N. \tag{4}$$

Explicit values involving the variance $\sigma = \mathbb{E}(b_{i,j}^2) \ge 1$ were derived in [4] for the constants α, β, leading to

$$\gamma := \frac{\beta}{\alpha} \in \left[1, \sqrt{8}\,\sigma\right]. \tag{5}$$

For instance, for symmetric Weibull random variables with exponent $r \ge 1$, we have $\gamma \in [1,4]$, since $\sigma = \sqrt{\Gamma(1+2/r)}/\Gamma(1+1/r) \le \sqrt{\Gamma(3)}/\Gamma(2) = \sqrt{2}$. The importance of the modified restricted isometry property lies in the fact that (2) is enough to guarantee stable and robust sparse recovery by ℓ_1-minimization, so that s-sparse recovery from subexponetial measurements is still highly likely in the regime $m \asymp s\ln(eN/s)$. This was the main message of [4]. However, this article focused only on ℓ_1-minimization and left open the possibility of sparse recovery via other algorithms. This is the problem addressed in this note. Since it seems rather unlikely that, say, the unaltered iterative hard thresolding algorithm as proposed in [2] would allow for sparse recovery in the optimal regime of parameters, we put forward a fresh variation on the underlying principle. Precisely, our novel algorithm consists in producing a sequence $(\mathbf{x}^n)_{n\ge 0}$ of sparse vectors constructed from $\mathbf{y} \in \mathbb{R}^m$ by iterating the scheme initiated at $\mathbf{x}^0 = \mathbf{0}$ and given as

$$\mathbf{x}^{n+1} = H_{\kappa s}\big(\mathbf{x}^n + \nu_n\, \mathbf{A}^*\mathrm{sgn}(\mathbf{y} - \mathbf{A}\mathbf{x}^n)\big). \tag{6}$$

Here, H_t denotes the usual hard thresholding operator that keeps t largest absolute entries of vector and sends the other ones to zero, $\mathbf{A}^*$ denotes the transpose of $\mathbf{A}$, and sgn denotes the sign function applied componentwise. Intuitively, (6) can be interpreted as the sparsification of a (sub)gradient descent step for the function $\mathbf{z} \mapsto \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_1$, instead of $\mathbf{z} \mapsto \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2^2$ in classical iterative hard thresholding. The exact values of the parameters $\kappa, \nu_n \in \mathbb{R}$ are revealed in the following theorem, which is the main result of this note. The reason for our specific choice is revealed by the proof.

Theorem 1. Let $\mathbf{A} \in \mathbb{R}^{m \times N}$ be a matrix satisfying (2) with $\delta_{(2\kappa+1)s} < 1/8$. For all $\mathbf{x} \in \mathbb{R}^N$ and all $\mathbf{e} \in \mathbb{R}^m$, if $\mathbf{x}^\sharp$ denotes any cluster point of the sequence $(\mathbf{x}^n)_{n \geq 0}$ produced by (6) applied to $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ and with parameters depending on the constants from (4)-(5) as

$$\kappa := \lceil 16\gamma^4 \rceil, \tag{7}$$

$$\nu_n := \frac{\|\mathbf{y} - \mathbf{A}\mathbf{x}^n\|_1}{\beta^2}, \tag{8}$$

then, for any index set $S \subseteq \{1, \ldots, N\}$ of size s,

$$\|\mathbf{x}_S - \mathbf{x}^\sharp\|_2 \leq d \|\mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}\|_1, \tag{9}$$

where the constant d depends only on α, β, and $\delta_{(2\kappa+1)s}$.

Remark. Before turning to the proof, a few comments are in order.

(i) In the ideal situation where $\mathbf{x}$ is exactly s-sparse (i.e., $\mathbf{x}_S = \mathbf{x}$ and $\mathbf{x}_{\overline{S}} = \mathbf{0}$ for some index set S of size s) and where there is no measurement error (i.e., $\mathbf{e} = \mathbf{0}$), the result guarantees that $\mathbf{x}^\sharp = \mathbf{x}$, meaning that $\mathbf{x}$ is exactly recovered as the limit of the sequence $(\mathbf{x}^n)_{n \geq 0}$. Moreover, the proof informs us that the convergence is geometric, i.e., there is a $\rho \in (0,1)$ such that $\|\mathbf{x} - \mathbf{x}^n\|_2 \leq \rho^n \|\mathbf{x}\|_2$ for all $n \geq 0$. Note in particular that after a finite number $\bar{n} := \lceil \ln(\|\mathbf{x}\|_2/\varepsilon)/\ln(1/\rho) \rceil$ of iterations, we have $\|\mathbf{x} - \mathbf{x}^{\bar{n}}\|_2 \leq \varepsilon$.

(ii) In the more realistic situation incorporating sparsity defect and measurement error, the convergence of the sequence $(\mathbf{x}^n)_{n \geq 0}$ is not necessarily guaranteed but the existence of cluster points is, and any cluster point $\mathbf{x}^\sharp$ satisfies (9). We also point out that the error estimate (9) can be replaced by

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq \frac{C}{\sqrt{s}} \min_{s-\text{sparse } \mathbf{z} \in \mathbb{R}^N} \|\mathbf{x} - \mathbf{z}\|_1 + D\|\mathbf{e}\|_1, \tag{10}$$

provided the algorithm is run with parameter $2s$ instead of s, so that $H_{2\kappa s}$ appears in lieu of $H_{\kappa s}$ in (6) and the condition $\delta_{(4\kappa+2)s} < 1/8$ prevails instead of $\delta_{(2\kappa+1)s} < 1/8$. It is somewhat folklore to derive (10) from (9) using the sort-and-split technique, see e.g. [3, Section 4.3]. However, since the argument is typically based on the standard restricted isometry property (1) rather than the modified version (2), we give a full justification for completeness. Given $\mathbf{x} \in \mathbb{R}^N$, we consider an index set S_0 corresponding to s largest absolute entries of $\mathbf{x}$, an index set S_1 corresponding to s next largest absolute entries of $\mathbf{x}$, an index set S_2 corresponding to s next largest absolute entries of $\mathbf{x}$, etc. Then we write

$$\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \leq \|\mathbf{x}_{\overline{S_0 \cup S_1}}\|_2 + \|\mathbf{x}_{S_0 \cup S_1} - \mathbf{x}^\sharp\|_2. \tag{11}$$

The first term in the right-hand side of (11) is bounded as

$$\|\mathbf{x}_{\overline{S_0 \cup S_1}}\|_2 \leq \frac{\|\mathbf{x}_{\overline{S_0}}\|_1}{\sqrt{s}}, \tag{12}$$

see e.g. [5, Proposition 2.3], while the second term is bounded with the help of Theorem 1 (applied with $2s$ instead of s) as

$$(13)\qquad \|\mathbf{x}_{S_0\cup S_1} - \mathbf{x}^\sharp\|_2 \le d\|\mathbf{A}\mathbf{x}_{\overline{S_0\cup S_1}} + \mathbf{e}\|_1 = d\left\|\mathbf{A}\left(\sum_{k\ge 2}\mathbf{x}_{S_k}\right) + \mathbf{e}\right\|_1 \le d\sum_{k\ge 2}\|\mathbf{A}\mathbf{x}_{S_k}\|_1 + d\|\mathbf{e}\|_1$$
$$\le d\sum_{k\ge 2}(1+\delta_s)\beta\|\mathbf{x}_{S_k}\|_2 + d\|\mathbf{e}\|_1 \le d(1+\delta_s)\beta\sum_{k\ge 2}\frac{\|\mathbf{x}_{S_{k-1}}\|_1}{\sqrt{s}} + d\|\mathbf{e}\|_1$$
$$\le d(1+\delta_s)\beta\frac{\|\mathbf{x}_{\overline{S_0}}\|_1}{\sqrt{s}} + d\|\mathbf{e}\|_1.$$

Substituting (12) and (13) into (11) gives the desired result.

(iii) The error estimate (10) is the most classical, except that it usually features $\|\mathbf{e}\|_2$ instead of $\|\mathbf{e}\|_1$ for subgaussian measurements, see e.g. [5, Theorems 6.12, 6.21, 6.25, 6.28]. The discrepancy is in fact due to the normalization of the measurement matrix $\mathbf{A}$, and (10) is in reality slightly better than the usual version with $\|\mathbf{e}\|_2$. Indeed, suppose that an s-sparse vector $\mathbf{x}\in\mathbb{R}^N$ is observed via $\mathbf{v} = \mathbf{B}\mathbf{x} + \mathbf{w}$ for an unnormalized matrix $\mathbf{B}\in\mathbb{R}^{m\times N}$. In the subgaussian case where we take $\mathbf{A} = \mathbf{B}/\sqrt{m}$, we deal with the measurement vector $\mathbf{y} = \mathbf{v}/\sqrt{m}$ and measurement error $\mathbf{e} = \mathbf{w}/\sqrt{m}$, so the usual error estimate would read $\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \le D\|\mathbf{e}\|_2 = D\|\mathbf{w}\|_2/\sqrt{m}$. In the subexponetial case where we take $\mathbf{A} = \mathbf{B}/m$, we deal with the measurement vector $\mathbf{y} = \mathbf{v}/m$ and measurement error $\mathbf{e} = \mathbf{w}/m$, so the error estimate (10) reads $\|\mathbf{x} - \mathbf{x}^\sharp\|_2 \le D\|\mathbf{e}\|_1 = D\|\mathbf{w}\|_1/m$, which is slightly better by virtue of $\|\mathbf{w}\|_1 \le \sqrt{m}\|\mathbf{w}\|_2$.

We now turn to the proof of Theorem 1, which relies on two key lemmas stated below, the first one being borrowed from [6].

Lemma 2. For any $\mathbf{u},\mathbf{v}\in\mathbb{R}^N$, if $\mathbf{u}$ is s-sparse, then (provided $N \ge (\kappa+1)s$)

$$(14)\qquad \|\mathbf{u} - H_{\kappa s}(\mathbf{v})\|_2 \le \eta\|\mathbf{u} - \mathbf{v}\|_2, \qquad \eta := \frac{1+\sqrt{4\kappa+1}}{\sqrt{4\kappa}}.$$

Lemma 3. For any $\mathbf{u}\in\mathbb{R}^N$ and any $\mathbf{e}\in\mathbb{R}^m$, if $\mathbf{u}$ is supported on an index set T of size t and if $\nu := \|\mathbf{A}\mathbf{u} + \mathbf{e}\|_1/\beta^2$, then

$$(15)\qquad \|\mathbf{u} - \nu(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u} + \mathbf{e}))_T\|_2 \le \left(1 - \frac{1-4\delta_t}{2\gamma^2}\right)\|\mathbf{u}\|_2 + d\|\mathbf{e}\|_1,$$

where d is a constant depending only on α, β, and $\delta_t \le \sqrt{2}-1$.

Proof. Let us first prove a variation of the estimate (15) where ν is replaced by $\widetilde{\nu} := \|\mathbf{A}\mathbf{u}\|_1/\beta^2$. By expanding squares, we have

$$(16)\quad \|\mathbf{u} - \tilde{\nu}(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u} + \mathbf{e}))_T\|_2^2 = \|\mathbf{u}\|_2^2 - 2\widetilde{\nu}\langle\mathbf{u}, (\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u} + \mathbf{e}))_T\rangle + \widetilde{\nu}^2\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{x} + \mathbf{e}))_T\|_2^2.$$

Since $\mathbf{u}$ is supported on T, we observe that

$$
\begin{aligned}
(17)\qquad \langle \mathbf{u}, (\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T \rangle &= \langle \mathbf{u}, \mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}) \rangle = \langle \mathbf{A}\mathbf{u}, \mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}) \rangle \\
&= \langle \mathbf{A}\mathbf{u}+\mathbf{e}, \mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}) \rangle - \langle \mathbf{e}, \mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}) \rangle \\
&= \|\mathbf{A}\mathbf{u}+\mathbf{e}\|_1 - \langle \mathbf{e}, \mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}) \rangle \\
&\geq \|\mathbf{A}\mathbf{u}\|_1 - 2\|\mathbf{e}\|_1.
\end{aligned}
$$

We also deduce from (2) and (4) that

$$
\begin{aligned}
(18)\qquad \|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2^2 &= \langle \mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}), (\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T \rangle \\
&= \langle \mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}), \mathbf{A}((\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T) \rangle \\
&\leq \|\mathbf{A}((\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T)\|_1 \\
&\leq (1+\delta_t) \,/\!\!/\, (\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T \,/\!\!/ \\
&\leq (1+\delta_t)\beta \|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2,
\end{aligned}
$$

which, after simplification, yields the estimate

$$
(19)\qquad \|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2 \leq (1+\delta_t)\beta.
$$

Substituting (17) and (19) into (16), while taking $\widetilde{\nu} = \|\mathbf{A}\mathbf{u}\|_1/\beta^2$ into account, we arrive at

$$
(20)\qquad \|\mathbf{u} - \widetilde{\nu}(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2^2 \leq \|\mathbf{u}\|_2^2 - \frac{2-(1+\delta_t)^2}{\beta^2}\|\mathbf{A}\mathbf{u}\|_1^2 + \frac{4}{\beta^2}\|\mathbf{A}\mathbf{u}\|_1\|\mathbf{e}\|_1.
$$

Invoking (2) and (4) once more, we also notice that

$$
\begin{aligned}
(21)\qquad \|\mathbf{A}\mathbf{u}\|_1 &\geq (1-\delta_t) \,/\!\!/\, \mathbf{u} \,/\!\!/\, \geq (1-\delta_t)\alpha\|\mathbf{u}\|_2, \\
(22)\qquad \|\mathbf{A}\mathbf{u}\|_1 &\leq (1+\delta_t) \,/\!\!/\, \mathbf{u} \,/\!\!/\, \leq (1+\delta_t)\beta\|\mathbf{u}\|_2.
\end{aligned}
$$

Next, using (21) and (22) in (20), while making sure that $2-(1+\delta_t)^2 \geq 0$, gives

$$
\|\mathbf{u} - \widetilde{\nu}(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2^2 \leq (1-c')\|\mathbf{u}\|_2^2 + d'\|\mathbf{u}\|_2\|\mathbf{e}\|_1,
$$

with $c' := (2-(1+\delta_t)^2)(1-\delta_t)^2/\gamma^2 = ((1-2\delta_t)^2 - \delta_t^4)/\gamma^2 \geq (1-4\delta_t)/\gamma^2 =: c''$ and $d' := 4(1+\delta_t)/\beta$. It follows that

$$
\begin{aligned}
(23)\qquad \|\mathbf{u} - \widetilde{\nu}(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2 &\leq \left[(1-c'')\left(\|\mathbf{u}\|_2^2 + \frac{d'}{1-c''}\|\mathbf{u}\|_2\|\mathbf{e}\|_1\right)\right]^{1/2} \\
&\leq \left(1-\frac{c''}{2}\right)\left(\|\mathbf{u}\|_2 + \frac{d'}{2(1-c'')}\|\mathbf{e}\|_1\right).
\end{aligned}
$$

This is (15) with ν replaced by $\widetilde{\nu}$. To obtain the genuine estimate (15), we simply write

$$
(24)\quad \|\mathbf{u} - \nu(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2 \leq \|\mathbf{u} - \widetilde{\nu}(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2 + |\nu - \widetilde{\nu}|\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2.
$$

We bound the first term on the right-hand side as in (23), and we bound the second term by $(1+\delta_t)\|\mathbf{e}\|_1/\beta$ by using $|\nu - \widetilde{\nu}| = |\|\mathbf{A}\mathbf{u}+\mathbf{e}\|_1 - \|\mathbf{A}\mathbf{u}\|_1|/\beta^2 \leq \|\mathbf{e}\|_1/\beta^2$ while keeping in mind the fact established in (19) that $\|(\mathbf{A}^*\mathrm{sgn}(\mathbf{A}\mathbf{u}+\mathbf{e}))_T\|_2 \leq (1+\delta_t)\beta$. This completes the proof. □

With Lemmas 2 and 3 now at our disposal, we can conclude this note swiftly by supplying the awaited proof of our main result.

Proof of Theorem 1. It is enough to show that, for all $n \geq 0$,

$$\|\mathbf{x}_S - \mathbf{x}^n\|_2 \leq \rho^n \|\mathbf{x}_S - \mathbf{x}^0\|_2 + d\|\mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}\|_1 \tag{25}$$

for some $\rho \in (0,1)$, which will follow from the validity, for all $n \geq 0$, of the prospective inequality

$$\|\mathbf{x}_S - \mathbf{x}^{n+1}\|_2 \leq \rho \|\mathbf{x}_S - \mathbf{x}^n\|_2 + d'\|\mathbf{e}'\|_1, \tag{26}$$

where $\mathbf{e}' := \mathbf{A}\mathbf{x}_{\overline{S}} + \mathbf{e}$. To prove the latter, let us consider the index sets $S^n := \mathrm{supp}(\mathbf{x}^n)$ and $S^{n+1} := \mathrm{supp}(\mathbf{x}^{n+1})$ of size κs, and let us remark that $\mathbf{x}^{n+1}$ defined in (6) is also obtained by applying the hard thresholding operator $H_{\kappa s}$ to $(\mathbf{x}^n + \nu_n \mathbf{A}^* \mathrm{sgn}(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_{S \cup S^n \cup S^{n+1}}$. Therefore, by Lemma 2 applied with $\mathbf{u} = \mathbf{x}_S$ and $\mathbf{v} = (\mathbf{x}^n + \nu_n \mathbf{A}^* \mathrm{sgn}(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_{S \cup S^n \cup S^{n+1}}$ yields

$$\begin{aligned} \|\mathbf{x}_S - \mathbf{x}^{n+1}\|_2 &\leq \eta \|\mathbf{x}_S - (\mathbf{x}^n + \nu_n (\mathbf{A}^* \mathrm{sgn}(\mathbf{y} - \mathbf{A}\mathbf{x}^n))_{S \cup S^n \cup S^{n+1}}\|_2 \\ &= \eta \|\mathbf{x}_S - \mathbf{x}^n - \nu_n (\mathbf{A}^* \mathrm{sgn}(\mathbf{A}(\mathbf{x}_S - \mathbf{x}^n) + \mathbf{e}'))_{S \cup S^n \cup S^{n+1}}\|_2, \end{aligned} \tag{27}$$

where, thanks to our choice of $\kappa = \lceil 16\gamma^4 \rceil$,

$$\eta = \frac{1 + \sqrt{1 + 4\kappa}}{\sqrt{4\kappa}} \leq \frac{2 + 2\sqrt{\kappa}}{2\sqrt{\kappa}} = 1 + \frac{1}{\sqrt{\kappa}} \leq 1 + \frac{1}{4\gamma^2}. \tag{28}$$

Now, calling upon Lemma 3 applied with $\mathbf{u} = \mathbf{x}_S - \mathbf{x}^n$, $T = S \cup S^n \cup S^{n+1}$, and $\mathbf{e}$ replaced by $\mathbf{e}'$, while using the fact that $\delta_{(2\kappa+1)s} \leq 1/8$, we obtain

$$\|\mathbf{x}_S - \mathbf{x}^n - \nu_n (\mathbf{A}^* \mathrm{sgn}(\mathbf{A}(\mathbf{x}_S - \mathbf{x}^n) + \mathbf{e}'))_{S \cup S^n \cup S^{n+1}}\|_2 \leq \left(1 - \frac{1}{4\gamma^2}\right) \|\mathbf{x}_S - \mathbf{x}^n\|_2 + d\|\mathbf{e}'\|_1. \tag{29}$$

Combining (27), (28), and (29) leads to

$$\|\mathbf{x}_S - \mathbf{x}^{n+1}\|_2 \leq \left(1 - \frac{1}{16\gamma^4}\right) \|\mathbf{x}_S - \mathbf{x}^n\|_2 + d'\|\mathbf{e}'\|_1. \tag{30}$$

This is the prospective inequality (26) with $\rho := 1 - 1/(16\gamma^4) < 1$, and our proof is now complete. □

Remark. Similar theoretical guarantees still hold if the stepsize ν_n in (6) is replaced by $\tau_n \nu_n$ with $\tau_n \in (\tau_-, \tau_+)$, where $\tau_- := 1/(1 + 4\gamma^2)$ and $\tau_+ := (2 + 1/4\gamma^2)/((1 + 1/4\gamma^2)(2 - 1/4\gamma^2))$. Indeed, for each $n \geq 1$, (27) remains valid in the form

$$\|\mathbf{x}_S - H_{\kappa s}(\mathbf{x}^n + \tau_n \nu_n \mathbf{A}^* \mathrm{sgn}(\mathbf{y} - \mathbf{A}\mathbf{x}^n))\|_2 \leq \eta \|\mathbf{x}_S - \mathbf{x}^n - \tau_n \nu_n (\mathbf{A}^* \mathrm{sgn}(\mathbf{A}(\mathbf{x}_S - \mathbf{x}^n) + \mathbf{e}'))_{S \cup S^n \cup S^{n+1}}\|_2,$$

and (29) can be replaced by

$$\begin{aligned} &\|\mathbf{x}_S - \mathbf{x}^n - \tau_n \nu_n (\mathbf{A}^* \mathrm{sgn}(\mathbf{A}(\mathbf{x}_S - \mathbf{x}^n) + \mathbf{e}'))_{S \cup S^n \cup S^{n+1}}\|_2 \\ &\qquad \leq \|(1 - \tau_n)(\mathbf{x}_S - \mathbf{x}^n)\|_2 + \tau_n \|\mathbf{x}_S - \mathbf{x}^n - \nu_n (\mathbf{A}^* \mathrm{sgn}(\mathbf{A}(\mathbf{x}_S - \mathbf{x}^n) + \mathbf{e}'))_{S \cup S^n \cup S^{n+1}}\|_2 \\ &\qquad \leq |1 - \tau_n| \|\mathbf{x}_S - \mathbf{x}^n\|_2 + \tau_n \left[(1 - 1/(4\gamma^2))\|\mathbf{x}_S - \mathbf{x}^n\|_2 + d\|\mathbf{e}'\|_1\right]. \end{aligned}$$

Therefore, we can derive that

$$\|\mathbf{x}_S - H_{\kappa s}(\mathbf{x}^n + \tau_n \nu_n \mathbf{A}^* \mathrm{sgn}(\mathbf{y} - \mathbf{A}\mathbf{x}^n))\|_2 \le \theta \|\mathbf{x}_S - \mathbf{x}^n\|_2 + d' \|\mathbf{e}'\|_1,$$

where $\theta := (1 + 1/4\gamma^2)(|1 - \tau_n| + \tau_n(1 - 1/4\gamma^2))$ satisfies $\theta < 1$. If $\tau_n \in [\tau_-, 1)$, this follows from $\theta = (1 + 1/4\gamma^2)(1 - \tau_n/4\gamma^2) < (1 + 1/4\gamma^2)(1 - \tau_-/4\gamma^2) = 1$, and if $\tau_n \in (1, \tau_+]$, this follows from $\theta = (1 + 1/4\gamma^2)(\tau_n(2 - 1/4\gamma^2) - 1) < (1 + 1/4\gamma^2)(\tau_+(2 - 1/4\gamma^2) - 1) = 1$.

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] T. Blumensath, M. Davies: Iterative hard thresholding for compressed sensing. Applied and Computational Harmonic Analysis. **27**, 265–274 (2009).

[3] S. Foucart: Flavors of Compressive Sensing. Approximation Theory XV: San Antonio 2016, Springer Proceedings in Mathematics & Statistics. To appear.

[4] S. Foucart, M.-J. Lai: Sparse recovery with pre-Gaussian random matrices. Studia Math. **200**, 91–102 (2010).

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

[6] J. Shen, P. Li: A tight bound of hard thresholding. arXiv preprint arXiv:1605.01656 (2016).