# Sparse Recovery Algorithms: Sufficient Conditions in terms of Restricted Isometry Constants

Simon Foucart

**Abstract**  We review three recovery algorithms used in Compressive Sensing for the reconstruction $s$-sparse vectors $\mathbf{x} \in \mathbb{C}^N$ from the mere knowledge of linear measurements $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^m$, $m < N$. For each of the algorithms, we derive improved conditions on the restricted isometry constants of the measurement matrix $A$ that guarantee the success of the reconstruction. These conditions are $\delta_{2s} < 0.4652$ for basis pursuit, $\delta_{3s} < 0.5$ and $\delta_{2s} < 0.25$ for iterative hard thresholding, and $\delta_{4s} < 0.3843$ for compressive sampling matching pursuit. The arguments also applies to almost sparse vectors and corrupted measurements. The analysis of iterative hard thresholding is surprisingly simple. The analysis of basis pursuit features a new inequality that encompasses several inequalities encountered in Compressive Sensing.

## 1 Introduction

In this paper, we address the Compressive Sensing problem that consists in reconstructing an $s$-sparse vector $\mathbf{x} \in \mathbb{C}^N$ from the mere knowledge of the measurement vector $\mathbf{y} = A\mathbf{x} \in \mathbb{C}^m$ when $m \ll N$. We do not focus on the design of suitable measurement matrices $A \in \mathbb{C}^{m \times N}$, since we take for granted the existence of matrices having small restricted isometry constants (see Section 2 for the definition of these constants). Instead, we focus on three popular reconstruction algorithms that allow sparse recovery in a stable and robust fashion. For each algorithm, we present some sufficient conditions in terms of restricted isometry constants that improve on the ones currently found in the literature. The algorithms under consideration are:

Simon Foucart

Université Pierre et Marie Curie, Paris, France, e-mail: foucart@ann.jussieu.fr

1

**Basis Pursuit:** solve the convex optimization problem

$$\underset{\mathbf{z}\in\mathbb{C}^N}{\text{minimize}} \ \|\mathbf{z}\|_1 \quad \text{subject to } A\mathbf{z}=\mathbf{y}. \tag{BP}$$

**Iterative Hard Thresholding:** from $\mathbf{x}^0$ $s$-sparse, iterate

$$\mathbf{x}^{n+1} = H_s(\mathbf{x}^n + A^*(\mathbf{y}-A\mathbf{x}^n)), \tag{IHT}$$

where the nonlinear operator $H_s$ keeps $s$ largest (in modulus) entries of a vector and sets the other ones to zero, so that $H_s(\mathbf{z})$ is a — not necessarily unique — best $s$-term approximation to $\mathbf{z}\in\mathbb{C}^N$.

**Compressive Sampling Matching Pursuit:** from $\mathbf{x}^0$ $s$-sparse, iterate

$$T^n := \big\{\text{indices of } 2s \text{ largest (in modulus) entries of } A^*(\mathbf{y}-A\mathbf{x}^n)\big\}, \tag{CSMP$_1$}$$
$$U^n := T^n \cup S^n, \qquad \text{where } S^n := \text{supp}(\mathbf{x}^n), \tag{CSMP$_2$}$$
$$\mathbf{u}^n := \text{argmin}\big\{\|\mathbf{y}-A\mathbf{z}\|_2, \text{supp}(\mathbf{z})\subseteq U^n\big\}, \tag{CSMP$_3$}$$
$$\mathbf{x}^{n+1} := H_s(\mathbf{u}^n). \tag{CSMP$_4$}$$

## 2 Restricted Isometry Constants

We recall the definition of restricted isometry constants introduced in [3]. We also draw attention to a less common, though sometimes preferable, characterization.

**Definition 1.** The $s$-th order restricted isometry constant $\delta_s = \delta_s(A)$ of a matrix $A\in\mathbb{C}^{m\times N}$ is the smallest $\delta\geq 0$ such that

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

An alternative characterization reads

$$\delta_s = \max_{S\subseteq[N],|S|\leq s} \|A_S^*A_S - \text{Id}\|_{2\to 2}. \tag{2}$$

*Proof.* To justify the equivalence between the two definitions, we start by noticing that (1) is equivalent to

$$\big|\|A_S\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\big| \leq \delta\|\mathbf{x}\|_2^2 \quad \text{for all } S\subseteq[N], |S|\leq s, \text{ and all } \mathbf{x}\in\mathbb{C}^{|S|}.$$

We then observe that

$$\|A_S\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 = \langle A_S\mathbf{x}, A_S\mathbf{x}\rangle - \langle \mathbf{x}, \mathbf{x}\rangle = \langle (A_S^*A_S - \text{Id})\mathbf{x}, \mathbf{x}\rangle.$$

Now, since the matrix $(A_S^*A_S - \text{Id})$ is hermitian, we have

$$\max_{\mathbf{x}\in\mathbb{C}^{|S|}\backslash\{0\}} \frac{\langle (A_S^*A_S - \mathrm{Id})\mathbf{x}, \mathbf{x}\rangle}{\|\mathbf{x}\|_2} = \|A_S^*A_S - \mathrm{Id}\|_{2\to 2},$$

so that (1) is equivalent to

$$\max_{S\subseteq[N],|S|\leq s} \|A_S^*A_S - \mathrm{Id}\|_{2\to 2} \leq \delta.$$

This establishes the identity (2), because $\delta_s$ is the smallest such $\delta$.   □

The expression (2) gives, for instance, an easy explanation of

$$|\langle A\mathbf{u}, A\mathbf{v}\rangle| \leq \delta_{\mathrm{supp}(\mathbf{u})\cup\mathrm{supp}(\mathbf{v})}\|\mathbf{u}\|_2\|\mathbf{v}\|_2 \qquad \text{if } \mathbf{u} \text{ and } \mathbf{v} \text{ are disjointly supported,}$$

a statement that can be derived in the real setting using a polarization formula, see e.g. [2]. Indeed, with $S := \mathrm{supp}(\mathbf{u}) \cup \mathrm{supp}(\mathbf{v})$, we just have to write (with slightly abusive notations)

$$|\langle A\mathbf{u}, A\mathbf{v}\rangle| = |\langle A_S\mathbf{u}, A_S\mathbf{v}\rangle - \langle \mathbf{u}, \mathbf{v}\rangle| = |\langle (A_S^*A_S - \mathrm{Id})\mathbf{u}, \mathbf{v}\rangle| \leq \|(A_S^*A_S - \mathrm{Id})\mathbf{u}\|_2\|\mathbf{v}\|_2$$
$$\leq \|A_S^*A_S - \mathrm{Id}\|_{2\to 2}\|\mathbf{u}\|_2\|\mathbf{v}\|_2 \leq \delta_{\mathrm{supp}(\mathbf{u})\cup\mathrm{supp}(\mathbf{v})}\|\mathbf{u}\|_2\|\mathbf{v}\|_2.$$

The concept of restricted isometry constant offers an elegant way to formulate sufficient conditions for the success of all the algorithms under consideration. Informally, if the restricted isometry constants are small, then all three algorithms are guaranteed to succeed in reconstructing sparse vectors. Slightly more precisely, if $\delta_t$ is small enough for some $t$ related to $s$, then any $s$-sparse vector $\mathbf{x} \in \mathbb{C}^N$ is recovered as the output of the algorithms. The object of what follows is to quantify this statement. We note that a sufficient condition in terms of some $\delta_t$ can always be imposed by a sufficient condition in terms of some other $\delta_{t'}$, according to the comparison result given in Proposition 1 below. For instance, in view of $\delta_{3s} \leq 3\delta_{2s}$, the sufficient condition $\delta_{3s} < 1/2$ obtained in Section 4 for iterative hard thresholding can be imposed by the condition $\delta_{2s} < 1/6$ — which will actually be improved to $\delta_{2s} < 1/4$. A heuristic way to compare such sufficient conditions is to recall that, given a prescribed $\delta > 0$, it is typical to have ($c$ denoting an absolute constant)

$$\delta_t \leq \delta \quad \text{provided} \quad m \geq c\,\frac{t}{\delta^2}\,\ln(eN/t)$$

for random measurement matrices. Therefore, it is desirable to make the ratio $t/\delta^2$ as small as possible in order to minimize the necessary number of measurements. As such, the sufficient condition $\delta_{3s} < 1/2$ is better than the condition $\delta_{2s} < 1/4$, since $3s/(1/2)^2 < 2s/(1/4)^2$. Let us now state the aforementioned comparison result, which is just an extension of [7, Corollary 3.4] to the case where $t$ is not a multiple of $s$.

**Proposition 1.** *For integers $t \geq s \geq 1$,*

$$\delta_s \leq \delta_t \leq \frac{t-d}{s} \delta_{2s} + \frac{d}{s} \delta_s, \qquad where \; d := \gcd(s,t).$$

*Proof.* The first inequality is clear. As for the second one, if $d$ denotes a common divisor of $s$ and $t$, we introduce the integers $k, n$ such that

$$s = kd, \qquad t = nd.$$

Given a $t$-sparse vector $\mathbf{u} \in \mathbb{C}^N$, we need to show that

$$\left| \|A\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2 \right| \leq \left( \frac{t-d}{s} \delta_{2s} + \frac{d}{s} \delta_s \right) \|\mathbf{u}\|_2^2. \tag{3}$$

Let $T =: \{j_1, j_2, \ldots, j_t\}$ denote the support of $\mathbf{u}$. We define $n$ subsets $S_1, S_2, \ldots, S_n$ of $T$, each of size $s$, by (the indices are meant modulo $t$)

$$S_i = \{j_{(i-1)d+1}, j_{(i-1)d+2}, \ldots, j_{(i-1)d+s}\}.$$

In this way, each $j \in T$ belongs to exactly $s/d = k$ sets $S_i$, so that

$$\mathbf{u} = \frac{1}{k} \sum_{1 \leq i \leq n} \mathbf{u}_{S_i} \qquad \text{and} \qquad \|\mathbf{u}\|_2^2 = \frac{1}{k} \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2^2.$$

Inequality (3) then follows from

$$\left| \|A\mathbf{u}\|_2^2 - \|\mathbf{u}\|_2^2 \right| = \left| \langle (A^*A - \mathrm{Id})\mathbf{u}, \mathbf{u} \rangle \right| \leq \frac{1}{k^2} \sum_{1 \leq i \leq n} \sum_{1 \leq j \leq n} \left| \langle (A^*A - \mathrm{Id})\mathbf{u}_{S_i}, \mathbf{u}_{S_j} \rangle \right|$$

$$= \frac{1}{k^2} \left( \sum_{1 \leq i \neq j \leq n} \left| \langle (A_{S_i \cup S_j}^* A_{S_i \cup S_j} - \mathrm{Id})\mathbf{u}_{S_i}, \mathbf{u}_{S_j} \rangle \right| + \sum_{1 \leq i \leq n} \left| \langle (A_{S_i}^* A_{S_i} - \mathrm{Id})\mathbf{u}_{S_i}, \mathbf{u}_{S_i} \rangle \right| \right)$$

$$\leq \frac{1}{k^2} \left( \sum_{1 \leq i \neq j \leq n} \delta_{2s} \|\mathbf{u}_{S_i}\|_2 \|\mathbf{u}_{S_j}\|_2 + \sum_{1 \leq i \leq n} \delta_s \|\mathbf{u}_{S_i}\|_2^2 \right)$$

$$= \frac{\delta_{2s}}{k^2} \left( \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2 \right)^2 - \frac{\delta_{2s}}{k^2} \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2^2 + \frac{\delta_s}{k^2} \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2^2$$

$$\leq \frac{\delta_{2s} n}{k^2} \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2^2 - \frac{\delta_{2s} - \delta_s}{k^2} \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2^2 = \left( \frac{\delta_{2s} n}{k^2} - \frac{\delta_{2s} - \delta_s}{k^2} \right) \sum_{1 \leq i \leq n} \|\mathbf{u}_{S_i}\|_2^2$$

$$= \left( \frac{n}{k} \delta_{2s} - \frac{1}{k}(\delta_{2s} - \delta_s) \right) \|\mathbf{u}\|_2^2 = \left( \frac{t}{s} \delta_{2s} - \frac{1}{k}(\delta_{2s} - \delta_s) \right) \|\mathbf{u}\|_2^2.$$

In order to make the latter as small as possible, we need to take $k$ as small as possible, i.e., to take $d$ as large as possible, hence the choice $d := \gcd(s,t)$. This finishes the proof. $\square$

## 3 Basis Pursuit

In this section, we recall that $s$-sparse recovery via basis pursuit succeeds as soon as $\delta_{2s} < 0.46515$. Contrary to the other sections, we do not give a full proof of this statement, as this was done in [5]. Instead, we outline the main steps of a classical argument before highlighting a particular inequality which is to be adjusted for the proof of the refined result. We justify the adjustment with a general inequality that encompasses two further inequalities often used in Compressive Sensing.

**Theorem 1.** *Suppose that the $2s$-th order restricted isometry constant of the matrix $A \in \mathbb{C}^{m \times N}$ satisfies*

$$\delta_{2s} < \frac{3}{4 + \sqrt{6}} \approx 0.46515.$$

*If $\mathbf{x} \in \mathbb{C}^N$ is an $s$-sparse vector, then it is recovered as a solution of* (BP) *with $\mathbf{y} = A\mathbf{x}$.*
*More generally, if $S$ denotes an index set of $s$ largest (in modulus) entries of a vector $\mathbf{x} \in \mathbb{C}^N$ and if $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ for some error term $\mathbf{e} \in \mathbb{C}^m$ satisfying $\|\mathbf{e}\|_2 \leq \eta$, then a minimizer $\mathbf{x}^\star$ of $\|\mathbf{z}\|_1$ subject to $\|A\mathbf{z} - \mathbf{y}\|_2 \leq \eta$ approximates the vector $\mathbf{x}$ with error*

$$\|\mathbf{x} - \mathbf{x}^\star\|_p \leq \frac{C}{s^{1-1/p}} \|\mathbf{x}_{\overline{S}}\|_1 + D\, s^{1/p-1/2} \eta, \qquad all \ p \in [1,2],$$

*where the constants $C$ and $D$ depend only on $\delta_{2s}$.*

*Proof (Sketch for $\delta_{2s} < 1/3$).* For exactly sparse vectors measured without error, it is necessary and sufficient to prove the *null space property* in the form

$$\|\mathbf{v}_S\|_1 \leq \frac{1}{2} \|\mathbf{v}\|_1 \qquad \text{for all } \mathbf{v} \in \ker A \setminus \{0\} \text{ and all } S \subseteq [N] \text{ with } |S| = s. \quad (4)$$

Take $\mathbf{v} \in \ker A \setminus \{0\}$. It is enough to prove (4) when $S$ is a set $S_0$ of indices corresponding to $s$ largest (in modulus) entries of $\mathbf{v}$. Then consider sets $S_1, S_2, \dots$ of $s$ indices ordered by decreasing moduli of entries of $\mathbf{v}$. We have

$$\|\mathbf{v}_{S_0}\|_2^2 \leq \frac{1}{1 - \delta_s} \|A(\mathbf{v}_{S_0})\|_2^2 = \frac{1}{1 - \delta_s} \sum_{k \geq 1} \langle A(\mathbf{v}_{S_0}), A(-\mathbf{v}_{S_k}) \rangle$$

$$\leq \frac{1}{1 - \delta_s} \sum_{k \geq 1} \delta_{2s} \|\mathbf{v}_{S_0}\|_2 \|\mathbf{v}_{S_k}\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \|\mathbf{v}_{S_0}\|_2 \sum_{k \geq 1} \|\mathbf{v}_{S_k}\|_2.$$

Simplify by $\|\mathbf{v}_{S_0}\|_2$ and observe that

$$\|\mathbf{v}_{S_k}\|_2 \leq \frac{1}{\sqrt{s}} \|\mathbf{v}_{S_{k-1}}\|_1, \qquad k \geq 1, \quad (5)$$

to obtain

$$\|\mathbf{v}_{S_0}\|_2 \leq \frac{\delta_{2s}}{1 - \delta_{2s}} \frac{1}{\sqrt{s}} \|\mathbf{v}\|_1.$$

The result follows from $\|\mathbf{v}_{S_0}\|_1 \leq \sqrt{s} \|\mathbf{v}_{S_0}\|_2$ and $\delta_{2s}/(1-\delta_{2s}) < 1/2$ for $\delta_{2s} < 1/3$. In the more general case of arbitrary signals measured with some errors, it is sufficient to prove a stronger form of (4), namely

$$\|\mathbf{v}_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|\mathbf{v}_{\overline{S}}\|_1 + \tau \|A\mathbf{v}\|_2 \qquad \text{for all } \mathbf{v} \in \mathbb{C}^N \text{ and all } S \subseteq [N] \text{ with } |S| = s,$$

for some positive constants $\rho < 1$ and $\tau$.   $\square$

For a proof of Theorem 1, the main point to adjust from the previous arguments is (5). This was first done in [1] by Cai, Wang, and Xu, who introduced what they called a *shifting inequality*. Instead of comparing the $\ell_2$-norm of the subvector $\mathbf{v}_{S_k}$ with the $\ell_1$-norm of the shifted subvector $\mathbf{v}_{S_{k-1}}$, they suggested to reduce the size of the shift from $s$ to roughly $s/4$. Precisely, they showed that

$$\text{for } k \geq s/4, \text{ if } \underbrace{a_1 \geq \cdots \geq \overbrace{a_{k+1} \geq \cdots \geq a_s}^{\mathbf{v}} \geq \cdots \geq a_{k+s}}_{\mathbf{u}} \geq 0, \text{ then}$$

$$\|\mathbf{v}\|_2 \leq \frac{1}{\sqrt{s}} \|\mathbf{u}\|_1.$$

This is the particular case $p = 1$, $q = 2$, $t = s$, of the following result, which generalizes the *shifting inequality* to other norms and to other vector sizes. A proof is included in the Appendix.

**Theorem 2.** *If* $0 < p < q$ *and* $\underbrace{a_1 \geq \cdots \geq \overbrace{a_{k+1} \geq \cdots \geq a_s}^{\mathbf{v}} \geq \cdots \geq a_{k+t}}_{\mathbf{u}} \geq 0$, *then*

$$\|\mathbf{v}\|_q \leq C_{p,q}(k,s,t) \|\mathbf{u}\|_p,$$

*where*

$$C_{p,q}(k,s,t) = \max\left\{ \frac{t^{p/q}}{s}, \left(\frac{p}{q}\right)^{p/q} \left(1 - \frac{p}{q}\right)^{1-p/q} \frac{1}{k^{1-p/q}} \right\}^{1/p}. \qquad (6)$$

*When* $\mathbf{u}$ *and* $\mathbf{v}$ *do not overlap much, the second term can be discarded, i.e.,*

$$C_{p,q}(k,s,t) = \frac{t^{1/q}}{s^{1/p}} \qquad \text{provided } s - k \leq \left(\frac{p}{q}\right)s. \qquad (7)$$

It is interesting to point out that two classical inequalities — the first one due to Stechkin — from Approximation Theory and often used in Compressive Sensing are contained in this result. These inequalities are,

for $0 < p < q$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\sigma_k(\mathbf{x})_q \leq \begin{cases} \dfrac{1}{k^{1/p-1/q}} \|\mathbf{x}\|_p, \\ \dfrac{D_{p,q}}{k^{1/p-1/q}} \|\mathbf{x}\|_p, \quad D_{p,q} := \dfrac{1}{(q/p-1)^{1/q}} = \left( \dfrac{(p/q)^{p/q}}{(1-p/q)^{p/q}} \right)^{1/p}. \end{cases}$$

This corresponds to the case $s = n$, $t = n - k$, for which we indeed have

$$k^{1-p/q} C_{p,q}(k,n,n-k)^p \leq \min\{1, D_{p,q}^p\},$$

since the left-hand side reduces to

$$\max \left\{ \left(1 - \frac{k}{n}\right)^{p/q} \left(\frac{k}{n}\right)^{1-p/q}, \left(\frac{p}{q}\right)^{p/q} \left(1 - \frac{p}{q}\right)^{1-p/q} \right\} = \left(\frac{p}{q}\right)^{p/q} \left(1 - \frac{p}{q}\right)^{1-p/q},$$

and the latter is readily seen to be bounded by $\min\{1, D_{p,q}^p\}$.

## 4 Iterative Hard Thresholding

In this section, we give an elegant and surprisingly simple justification of the success of $s$-sparse recovery via iterative hard thresholding as soon as $\delta_{3s} < 1/2$. This improves the result of [4], where the sufficient condition $\delta_{3s} < 1/\sqrt{8}$ was obtained — although the main theorem was stated for $\delta_{3s} < 1/\sqrt{32}$ in order to achieve a rate of convergence equal to $\rho = 1/2$.

**Theorem 3.** *Suppose that the $3s$-th order restricted isometry constant of the matrix $A \in \mathbb{C}^{m \times N}$ satisfies*

$$\delta_{3s} < \frac{1}{2}.$$

*If $\mathbf{x} \in \mathbb{C}^N$ is an $s$-sparse vector, then the sequence $(\mathbf{x}^n)$ defined by* (IHT) *with $\mathbf{y} = A\mathbf{x}$ converges to the vector $\mathbf{x}$.*
*More generally, if $S$ denotes an index set of $s$ largest (in modulus) entries of a vector $\mathbf{x} \in \mathbb{C}^N$ and if $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ for some error term $\mathbf{e} \in \mathbb{C}^m$, then*

$$\|\mathbf{x}^n - \mathbf{x}_S\|_2 \leq \rho^n \|\mathbf{x}^0 - \mathbf{x}_S\|_2 + \tau \|A\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2, \qquad all \ n \geq 0, \tag{8}$$

*where*

$$\rho := 2\delta_{3s} < 1 \qquad and \qquad \tau := \frac{2\sqrt{1+\delta_{2s}}}{1-2\delta_{3s}}.$$

*Remark 1.* The value $\tau = 6$ was obtained in [4] for $\rho \leq 1/2$, which was ensured by $\delta_{3s} \leq 1/\sqrt{32}$. Theorem 3 gives the value $\tau \approx 4.4721$ for $\rho \leq 1/2$, i.e., for $\delta_{3s} \leq 1/4$, and the value $\tau \approx 3.3562$ for $\delta_{3s} \leq 1/\sqrt{32}$.

*Proof.* We simply use the fact that the $s$-sparse vector $\mathbf{x}^{n+1}$ is a better $s$-term approximation to

$$\mathbf{v}^n := \mathbf{x}^n + A^*(\mathbf{y} - A\mathbf{x}^n) = \mathbf{x}^n + A^*A(\mathbf{x}_S - \mathbf{x}^n) + A^*(A\mathbf{x}_{\overline{S}} + \mathbf{e})$$

than the $s$-sparse vector $\mathbf{x}_S$ to write

$$\|(\mathbf{v}^n - \mathbf{x}_S) + (\mathbf{x}_S - \mathbf{x}^{n+1})\|_2^2 \leq \|\mathbf{v}^n - \mathbf{x}_S\|_2^2.$$

Expanding the left-hand side and eliminating $\|\mathbf{v}^n - \mathbf{x}_S\|_2^2$ leads to, with $\mathbf{e}' := A\mathbf{x}_{\overline{S}} + \mathbf{e}$ and $V := \operatorname{supp}(\mathbf{x}) \cup \operatorname{supp}(\mathbf{x}^n) \cup \operatorname{supp}(\mathbf{x}^{n+1})$,

$$\begin{aligned}
\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2^2 &\leq 2\Re\langle \mathbf{v}^n - \mathbf{x}_S, \mathbf{x}^{n+1} - \mathbf{x}_S \rangle \\
&= 2\Re\langle (\mathrm{Id} - A^*A)(\mathbf{x}^n - \mathbf{x}_S) + A^*\mathbf{e}', \mathbf{x}^{n+1} - \mathbf{x}_S \rangle \\
&\leq 2\Re\langle (\mathrm{Id} - A_V^*A_V)(\mathbf{x}^n - \mathbf{x}_S), \mathbf{x}^{n+1} - \mathbf{x}_S \rangle + 2\Re\langle \mathbf{e}', A(\mathbf{x}^{n+1} - \mathbf{x}_S) \rangle \\
&\leq 2\|\mathrm{Id} - A_V^*A_V\|_{2\to 2}\|\mathbf{x}^n - \mathbf{x}_S\|_2\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2 + 2\|\mathbf{e}'\|_2\|A(\mathbf{x}^{n+1} - \mathbf{x}_S)\|_2 \\
&\leq 2\delta_{3s}\|\mathbf{x}^n - \mathbf{x}_S\|_2\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2 + 2\|\mathbf{e}'\|_2\sqrt{1 + \delta_{2s}}\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2.
\end{aligned}$$

Simplifying by $\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2$, we derive

$$\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2 \leq 2\delta_{3s}\|\mathbf{x}^n - \mathbf{x}_S\|_2 + 2\sqrt{1 + \delta_{2s}}\|\mathbf{e}'\|_2.$$

This easily implies the estimate (8). In particular that, if $\mathbf{x}$ is an $s$-sparse vector ($\mathbf{x}_{\overline{S}} = 0$) and if the measurements are accurate ($\mathbf{e} = 0$), then

$$\|\mathbf{x}^n - \mathbf{x}\|_2 \leq \rho^n\|\mathbf{x}^0 - \mathbf{x}\|,$$

so the sequence $(\mathbf{x}^n)$ converges to $\mathbf{x}$ as soon as $\rho < 1$, i.e., $\delta_{3s} < 1/2$.  $\square$

Using the same technique, it is also possible to formulate a sufficient condition in terms of $\delta_{2s}$ for the success of $s$-sparse recovery via iterative hard thresholding. For simplicity, we only state the result in the case of exactly sparse vectors measured with perfect accuracy.

**Theorem 4.** *Suppose that the $2s$-th order restricted isometry constant of the matrix $A \in \mathbb{C}^{m\times N}$ satisfies*

$$\delta_{2s} < \frac{1}{4}.$$

*If $\mathbf{x} \in \mathbb{C}^N$ is an $s$-sparse vector, then the sequence $(\mathbf{x}^n)$ defined by* (IHT) *with $\mathbf{y} = A\mathbf{x}$ converges to the vector $\mathbf{x}$.*

*Proof.* We use what has been done in the proof of Theorem 3, specified to the case $\mathbf{e}' = 0$, to write

$$\|\mathbf{x}^{n+1} - \mathbf{x}\|_2^2 \leq 2\Re\langle (\mathrm{Id} - A^*A)(\mathbf{x}^n - \mathbf{x}), \mathbf{x}^{n+1} - \mathbf{x} \rangle.$$

Let us decompose $\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^n) \cup \text{supp}(\mathbf{x}^{n+1})$ into the three disjoint sets

$$
\begin{aligned}
V_1 &:= \quad (\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^n)) \quad \cap \quad (\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^{n+1})), \\
V_2 &:= \quad (\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^n)) \quad \setminus \quad (\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^{n+1})), \\
V_3 &:= \quad (\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^{n+1})) \quad \setminus \quad (\text{supp}(\mathbf{x}) \cup \text{supp}(\mathbf{x}^n)).
\end{aligned}
$$

Since $V_1 \cup V_2$, $V_2 \cup V_3$, and $V_2 \cup V_3$ all have size at most $2s$, we have

$$
\begin{aligned}
\|\mathbf{x}^{n+1} - \mathbf{x}\|_2^2 &= 2\Re\langle (\text{Id} - A^*A)((\mathbf{x}^n - \mathbf{x})_{V_1} + (\mathbf{x}^n - \mathbf{x})_{V_2}), (\mathbf{x}^{n+1} - \mathbf{x})_{V_1} + (\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\rangle \\
&= 2\Re\langle (\text{Id} - A^*A)((\mathbf{x}^n - \mathbf{x})_{V_1}), (\mathbf{x}^{n+1} - \mathbf{x})_{V_1}\rangle \\
&\quad + 2\Re\langle (\text{Id} - A^*A)((\mathbf{x}^n - \mathbf{x})_{V_1}), (\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\rangle \\
&\quad + 2\Re\langle (\text{Id} - A^*A)((\mathbf{x}^n - \mathbf{x})_{V_2}), (\mathbf{x}^{n+1} - \mathbf{x})_{V_1}\rangle \\
&\quad + 2\Re\langle (\text{Id} - A^*A)((\mathbf{x}^n - \mathbf{x})_{V_2}), (\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\rangle \\
&\leq 2\delta_{2s}\big(\|(\mathbf{x}^n - \mathbf{x})_{V_1}\|_2\|(\mathbf{x}^{n+1} - \mathbf{x})_{V_1}\|_2 + \|(\mathbf{x}^n - \mathbf{x})_{V_1}\|_2\|(\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\|_2 \\
&\qquad + \|(\mathbf{x}^n - \mathbf{x})_{V_2}\|_2\|(\mathbf{x}^{n+1} - \mathbf{x})_{V_1}\|_2 + \|(\mathbf{x}^n - \mathbf{x})_{V_2}\|_2\|(\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\|_2\big) \\
&\leq 2\delta_{2s}\big(\|(\mathbf{x}^n - \mathbf{x})_{V_1}\|_2^2 + \|(\mathbf{x}^n - \mathbf{x})_{V_1}\|_2^2 + \|(\mathbf{x}^n - \mathbf{x})_{V_2}\|_2^2 + \|(\mathbf{x}^n - \mathbf{x})_{V_2}\|_2^2\big)^{1/2} \\
&\qquad \times \big(\|(\mathbf{x}^{n+1} - \mathbf{x})_{V_1}\|_2^2 + \|(\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\|_2^2 + \|(\mathbf{x}^{n+1} - \mathbf{x})_{V_1}\|_2^2 + \|(\mathbf{x}^{n+1} - \mathbf{x})_{V_3}\|_2^2\big)^{1/2} \\
&= 2\delta_{2s}\big(2\|\mathbf{x}^n - \mathbf{x}\|_2^2\big)^{1/2}\big(2\|\mathbf{x}^{n+1} - \mathbf{x}\|_2^2\big)^{1/2} = 4\delta_{2s}\|\mathbf{x}^n - \mathbf{x}\|_2\|\mathbf{x}^{n+1} - \mathbf{x}\|_2.
\end{aligned}
$$

This yields, after simplification by $\|\mathbf{x}^{n+1} - \mathbf{x}\|_2$,

$$
\|\mathbf{x}^{n+1} - \mathbf{x}\|_2 \leq \rho\,\|\mathbf{x}^n - \mathbf{x}\|_2, \qquad \rho := 4\delta_{2s}.
$$

Convergence of the sequence $(\mathbf{x}_n)$ towards $\mathbf{x}$ is therefore guaranteed as soon as $\rho < 1$, i.e., $\delta_{2s} < 1/4$. $\quad\square$

*Remark 2.* The better sufficient condition $\delta_{2s} < 1/3$ was obtained in [6] with a slight modification of the iterative hard thresholding algorithm, namely the iteration

$$
\mathbf{x}^{n+1} = H_s\left(\mathbf{x}^n + \frac{3}{4}A^*(\mathbf{y} - A\mathbf{x}^n)\right). \tag{IHT$_{3/4}$}
$$

Note, however, that the condition $\delta_{2s} < 1/3$ is not heuristically better than the condition $\delta_{3s} < 1/2$, since $(2s)/(1/3)^2 > (3s)/(1/2)^2$.

## 5 Compressive Sampling Matching Pursuit

In this section, we present a proof of the success of $s$-sparse recovery via compressive sampling matching pursuit as soon as $\delta_{4s} < 0.38427$. This improves the original condition of [7]. There, the authors targeted a rate

of convergence equal to $\rho = 1/2$, so that they gave the sufficient condition $\delta_{4s} \leq 0.1$, but their arguments actually yield $\rho < 1$ as soon as $\delta_{4s} < 0.17157$.

**Theorem 5.** *Suppose that the $4s$-th order restricted isometry constant of the matrix $A \in \mathbb{C}^{m \times N}$ satisfies*

$$\delta_{4s} < \sqrt{\frac{2}{5 + \sqrt{73}}} \approx 0.38427.$$

*If $\mathbf{x} \in \mathbb{C}^N$ is an $s$-sparse vector, then the sequence $(\mathbf{x}^n)$ defined by $(\mathrm{CSMP}_{1-4})$ with $\mathbf{y} = A\mathbf{x}$ converges to the vector $\mathbf{x}$.*
*More generally, if $S$ denotes an index set of $s$ largest (in modulus) entries of a vector $\mathbf{x} \in \mathbb{C}^N$ and if $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ for some error term $\mathbf{e} \in \mathbb{C}^m$, then*

$$\|\mathbf{x}^n - \mathbf{x}_S\|_2 \leq \rho^n \|\mathbf{x}^0 - \mathbf{x}_S\|_2 + \tau \|A\mathbf{x}_{\overline{S}} + \mathbf{e}\|_2, \qquad all \ n \geq 0, \tag{9}$$

*where the positive constants $\rho < 1$ and $\tau$ depend only on $\delta_{4s}$.*

*Remark 3.* The explicit expressions for $\rho$ and $\tau$ are given at the end of the proof (the constant $\tau$ is made dependent only on $\delta_{4s}$ by using $\delta_{3s} \leq \delta_{4s}$). Note that the value $\tau = 15$ was obtained in [7] for $\rho \leq 1/2$, which was ensured by $\delta_{4s} \leq 0.1$. Theorem 5 gives the value $\tau \approx 10.369$ for $\rho \leq 1/2$, i.e., for $\delta_{4s} \leq 0.22665$, and the value $\tau \approx 5.6686$ for $\delta_{4s} \leq 0.1$.

*Proof.* Step $(\mathrm{CSMP}_3)$ says that $A\mathbf{u}^n$ is the best $\ell_2$-approximation to $\mathbf{y}$ from the space $\{A\mathbf{z}, \mathrm{supp}(\mathbf{z}) \subseteq U^n\}$, hence it is characterized by

$$\langle A\mathbf{u}^n - \mathbf{y}, A\mathbf{z} \rangle = 0 \qquad \text{whenever } \mathrm{supp}(\mathbf{z}) \subseteq U^n. \tag{10}$$

Setting $\mathbf{e}' := A\mathbf{x}_{\overline{S}} + \mathbf{e}$ to have $\mathbf{y} = A\mathbf{x}_S + \mathbf{e}'$, this can be rewritten as

$$\langle \mathbf{u}^n - \mathbf{x}_S, A^*A\mathbf{z} \rangle = \langle \mathbf{e}', A\mathbf{z} \rangle \qquad \text{whenever } \mathrm{supp}(\mathbf{z}) \subseteq U^n. \tag{11}$$

This yields in particular

$$\|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2^2 = \langle \mathbf{u}^n - \mathbf{x}_S, (\mathbf{u}^n - \mathbf{x}_S)_{U^n} \rangle$$
$$= \langle \mathbf{u}^n - \mathbf{x}_S, (\mathrm{Id} - A^*A)\big((\mathbf{u}^n - \mathbf{x}_S)_{U^n}\big) \rangle + \langle \mathbf{e}', A\big((\mathbf{u}^n - \mathbf{x}_S)_{U^n}\big) \rangle$$
$$\leq \delta_{4s} \|\mathbf{u}^n - \mathbf{x}_S\|_2 \|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2 + \|\mathbf{e}'\|_2 \sqrt{1 + \delta_{3s}} \|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2,$$

which gives, after simplification by $\|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2$,

$$\|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2 \leq \delta_{4s} \|\mathbf{u}^n - \mathbf{x}_S\|_2 + \sqrt{1 + \delta_{3s}} \|\mathbf{e}'\|_2. \tag{12}$$

It follows that

$$\|\mathbf{u}^n - \mathbf{x}_S\|_2^2 = \|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2^2 + \|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2^2$$
$$\leq \|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2^2 + \big(\delta_{4s} \|\mathbf{u}^n - \mathbf{x}_S\|_2 + \sqrt{1 + \delta_{3s}} \|\mathbf{e}'\|_2\big)^2.$$

This reads $p(\|\mathbf{u}^n - \mathbf{x}_S\|_2) \leq 0$ for the quadratic polynomial defined by

$$p(t) := (1 - \delta_{4s}^2)\, t^2 - (2\delta_{4s}\sqrt{1 + \delta_{3s}}\, \|\mathbf{e}'\|_2)\, t - (\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2^2 + (1 + \delta_{3s})\, \|\mathbf{e}'\|_2^2).$$

This proves that $\|\mathbf{u}^n - \mathbf{x}_S\|_2$ is bounded by the largest root of $p$, i.e.,

$$\|\mathbf{u}^n - \mathbf{x}_S\|_2 \leq \frac{\delta_{4s}\sqrt{1 + \delta_{3s}}\|\mathbf{e}'\|_2 + \sqrt{(1 - \delta_{4s}^2)\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2^2 + (1 + \delta_{3s})\,\|\mathbf{e}'\|_2^2}}{1 - \delta_{4s}^2}$$

$$\leq \frac{1}{\sqrt{1 - \delta_{4s}^2}}\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2 + \frac{\sqrt{1 + \delta_{3s}}}{1 - \delta_{4s}}\|\mathbf{e}'\|_2. \tag{13}$$

We now turn to the estimate for $\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2$. We start by writing

$$\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2^2 = \|(\mathbf{u}^n - \mathbf{x}_S) - (\mathbf{u}^n - \mathbf{x}^{n+1})\|_2^2$$
$$= \|\mathbf{u}^n - \mathbf{x}_S\|_2^2 + \|\mathbf{u}^n - \mathbf{x}^{n+1}\|_2^2 - 2\Re\langle\mathbf{u}^n - \mathbf{x}_S, \mathbf{u}^n - \mathbf{x}^{n+1}\rangle. \tag{14}$$

Step (CSMP$_4$) implies that $\mathbf{x}^{n+1}$ is a better $s$-term approximation to $\mathbf{u}^n$ than $\mathbf{x}_{S\cap U^n}$, so that

$$\|\mathbf{u}^n - \mathbf{x}^{n+1}\|_2 \leq \|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2. \tag{15}$$

We also note, in view of (11) and of $\operatorname{supp}(\mathbf{u}^n - \mathbf{x}^{n+1}) \subseteq U^n$, that

$$|\langle\mathbf{u}^n - \mathbf{x}_S, \mathbf{u}^n - \mathbf{x}^{n+1}\rangle| = |\langle\mathbf{u}^n - \mathbf{x}_S, (\operatorname{Id} - A^*A)(\mathbf{u}^n - \mathbf{x}^{n+1})\rangle + \langle\mathbf{e}', A(\mathbf{u}^n - \mathbf{x}^{n+1})\rangle|$$
$$\leq \delta_{4s}\|\mathbf{u}^n - \mathbf{x}_S\|_2\|\mathbf{u}^n - \mathbf{x}^{n+1}\|_2 + \|\mathbf{e}'\|_2\sqrt{1 + \delta_{3s}}\|\mathbf{u}^n - \mathbf{x}^{n+1}\|_2. \tag{16}$$

Substituting (15) and (16) into (14), then using (12), we obtain

$$\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2^2 \leq \|\mathbf{u}^n - \mathbf{x}_S\|_2^2 + \|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2^2 + 2\delta_{4s}\|\mathbf{u}^n - \mathbf{x}_S\|_2\|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2$$
$$+ 2\sqrt{1 + \delta_{3s}}\|\mathbf{e}'\|_2\|(\mathbf{u}^n - \mathbf{x}_S)_{U^n}\|_2$$
$$= (1 + 3\delta_{4s}^2)\|\mathbf{u}^n - \mathbf{x}_S\|_2^2 + 6\delta_{4s}\sqrt{1 + \delta_{3s}}\|\mathbf{u}^n - \mathbf{x}_S\|_2\|\mathbf{e}'\|_2 + 3(1 + \delta_{3s})\|\mathbf{e}'\|_2^2$$
$$\leq (1 + 3\delta_{4s}^2)\left(\|\mathbf{u}^n - \mathbf{x}_S\|_2 + \sqrt{\frac{3(1 + \delta_{3s})}{1 + 3\delta_{4s}^2}}\|\mathbf{e}'\|_2\right)^2.$$

Combining the latter with (13), we deduce

$$\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2 \leq \sqrt{\frac{1 + 3\delta_{4s}^2}{1 - \delta_{4s}^2}}\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2 + \left(\frac{\sqrt{1 + 3\delta_{4s}^2}}{1 - \delta_{4s}} + \sqrt{3}\right)\sqrt{1 + \delta_{3s}}\|\mathbf{e}'\|_2. \tag{17}$$

It remains to bound $\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2$ in terms of $\|\mathbf{x}^n - \mathbf{x}_S\|_2$. For this, we notice that $\mathbf{u}^n_{\overline{U^n}} = 0 = \mathbf{x}^n_{\overline{U^n}}$, so that

$$\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2 = \|(\mathbf{x}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2 \le \|(\mathbf{x}^n - \mathbf{x}_S)_{\overline{T^n}}\|_2 = \|(\mathbf{x}^n - \mathbf{x}_S)_{(S \cup S^n) \setminus T^n}\|_2. \quad (18)$$

Step $(\mathrm{CSMP}_1)$ means that $A^*(\mathbf{y} - A\mathbf{x}^n)_{T^n}$ is a best $2s$-term approximation to $A^*(\mathbf{y} - A\mathbf{x}^n)_{S \cup S^n \cup T^n}$ among all vectors supported on $S \cup S^n \cup T^n$. In particular,

$$\|A^*(\mathbf{y} - A\mathbf{x}^n)_{(S \cup S^n) \setminus T^n}\|_2 \le \|A^*(\mathbf{y} - A\mathbf{x}^n)_{T^n \setminus (S \cup S^n)}\|_2$$
$$\le \|A^* A(\mathbf{x}_S - \mathbf{x}^n)_{T^n \setminus (S \cup S^n)}\|_2 + \|(A^* \mathbf{e}')_{T^n \setminus (S \cup S^n)}\|_2$$
$$= \|((A^* A - \mathrm{Id})(\mathbf{x}_S - \mathbf{x}^n))_{T^n \setminus (S \cup S^n)}\|_2 + \|(A^* \mathbf{e}')_{T^n \setminus (S \cup S^n)}\|_2$$
$$\le \delta_{4s}\|\mathbf{x}_S - \mathbf{x}^n\|_2 + \|(A^* \mathbf{e}')_{T^n \setminus (S \cup S^n)}\|_2. \quad (19)$$

On the other hand, we have

$$\|A^*(\mathbf{y} - A\mathbf{x}^n)_{(S \cup S^n) \setminus T^n}\|_2 \ge \|A^* A(\mathbf{x}_S - \mathbf{x}^n)_{(S \cup S^n) \setminus T^n}\|_2 - \|(A^* \mathbf{e}')_{(S \cup S^n) \setminus T^n}\|_2$$
$$\ge \|(\mathbf{x}_S - \mathbf{x}^n)_{(S \cup S^n) \setminus T^n}\|_2 - \|((A^* A - \mathrm{Id})(\mathbf{x}_S - \mathbf{x}^n))_{(S \cup S^n) \setminus T^n}\|_2 - \|(A^* \mathbf{e}')_{(S \cup S^n) \setminus T^n}\|_2$$
$$\ge \|(\mathbf{x}_S - \mathbf{x}^n)_{(S \cup S^n) \setminus T^n}\|_2 - \delta_{2s}\|\mathbf{x}_S - \mathbf{x}^n\|_2 - \|(A^* \mathbf{e}')_{(S \cup S^n) \setminus T^n}\|_2. \quad (20)$$

From (19), (20), and (18), we derive that

$$\|(\mathbf{u}^n - \mathbf{x}_S)_{\overline{U^n}}\|_2 \le (\delta_{2s} + \delta_{4s})\|\mathbf{x}_S - \mathbf{x}^n\|_2 + \|(A^* \mathbf{e}')_{T^n \setminus (S \cup S^n)}\|_2 + \|(A^* \mathbf{e}')_{(S \cup S^n) \setminus T^n}\|_2$$
$$\le 2\delta_{4s}\|\mathbf{x}_S - \mathbf{x}^n\|_2 + \sqrt{2}\,\|(A^* \mathbf{e}')_{(S \cup S^n) \Delta T^n}\|_2$$
$$\le 2\delta_{4s}\|\mathbf{x}_S - \mathbf{x}^n\|_2 + \sqrt{2(1 + \delta_{4s})}\,\|\mathbf{e}'\|_2. \quad (21)$$

Putting (17) and (21) together, we finally conclude

$$\|\mathbf{x}^{n+1} - \mathbf{x}_S\|_2 \le \rho\,\|\mathbf{x}^n - \mathbf{x}_S\|_2 + (1 - \rho)\tau\,\|\mathbf{e}'\|_2. \quad (22)$$

where

$$\rho := \sqrt{\frac{4\delta_{4s}^2(1 + 3\delta_{4s}^2)}{1 - \delta_{4s}^2}},$$

$$(1 - \rho)\tau := \sqrt{\frac{2(1 + 3\delta_{4s}^2)}{1 - \delta_{4s}}} + \frac{\sqrt{(1 + 3\delta_{4s}^2)(1 + \delta_{3s})}}{1 - \delta_{4s}} + \sqrt{3(1 + \delta_{3s})}.$$

To finish, we point out that the constant $\rho$ is less than one when

$$12\delta_{4s}^4 + 5\delta_{4s}^2 - 1 < 0, \qquad \text{i.e.,} \qquad \delta_{4s} < \sqrt{\frac{2}{5 + \sqrt{73}}} \approx 0.38427,$$

and that (22) readily implies (9). $\quad\square$

## Appendix: Proof of Theorem 2

The constant $C_{p,q}(k,s,t)^q$ is the solution of the maximization problem

maximize $a_{k+1}^q + \cdots + a_{k+t}^q$     subject to $a_1^p + \cdots + a_s^p \leq 1$, $a_1 \geq \cdots \geq a_{k+t} \geq 0$.

Thus, setting $r := q/p > 1$, we aim at maximizing the convex function

$$f(x_1, \ldots, x_{k+t}) := x_{k+1}^r + \cdots + x_{k+t}^r$$

over the convex polytope

$$\mathscr{C} := \left\{ \mathbf{x} \in \mathbb{R}^{k+t} : x_1 + \cdots + x_s \leq 1, x_1 \geq \cdots \geq x_{k+t} \geq 0 \right\}.$$

The maximum is attained at one of the vertices of the convex polytope $\mathscr{C}$. These vertices are obtained by turning into equalities $k+t$ of the $k+t+1$ inequalities defining $\mathscr{C}$. We have to separate several cases:

- if $x_1 = \cdots = x_{k+t} = 0$, then

$$f(x_1, \ldots, x_{k+t}) = 0;$$

- if $x_1 + \cdots + x_s = 1$ and $x_1 = \cdots = x_h > x_{h+1} = x_{k+t} = 0$ for some $1 \leq h \leq k$, then

$$f(x_1, \ldots, x_{k+t}) = 0;$$

- if $x_1 + \cdots + x_s = 1$ and $x_1 = \cdots = x_h > x_{h+1} = x_{k+t} = 0$ for some $k \leq h \leq s$, then $x_1 = \cdots = x_h = 1/h$, and

$$f(x_1, \ldots, x_{k+t}) = \frac{h-k}{h^r};$$

- if $x_1 + \cdots + x_s = 1$ and $x_1 = \cdots = x_h > x_{h+1} = x_{k+t} = 0$ for some $s \leq h \leq k+t$, then $x_1 = \cdots = x_h = 1/s$, and

$$f(x_1, \ldots, x_{k+t}) = \frac{h-k}{s^r}.$$

It follows that the desired constant is

$$C_{p,q}(k,s,t)^q = \max \left\{ \max_{k \leq h \leq s} \frac{h-k}{h^r}, \max_{s \leq h \leq k+t} \frac{h-k}{s^r} \right\}.$$

Considering $h$ as a continuous variable, we observe that the function $g(h) := (h-k)/h^r$ is increasing until the critical point $h^* := \left( r/(r-1) \right) k$

and decreasing thereafter, so that the first maximum is no larger than $g(h^*) = \big((r-1)^{r-1}/r^r\big)/k^{r-1}$, or than $g(s) = (s-k)/s^r$ if $h^* \geq s$. Now taking into account that $(h-k)/s^r$ increases with $h$ on $[s, k+t]$, we deduce

$$C_{p,q}(k,s,t)^q \begin{cases} \leq \max\left\{\dfrac{(r-1)^{r-1}}{r^r}\dfrac{1}{k^{r-1}}, \dfrac{t}{s^r}\right\}, \\ = \dfrac{t}{s^r} \end{cases} \quad \text{if} \quad \dfrac{r}{r-1}k \geq s.$$

We simply obtain (6) and (7) by rearranging the latter. It is worth noting that the constants appearing in (6) and (7) cannot be improved. $\square$

# References

1. Cai, T.T., Wang, L., Xu, G.: Shifting inequality and recovery of sparse signals. IEEE Transactions on Signal Processing, To appear.
2. Candès, E.J.: The restricted isometry property and its implications for compressed sensing. Comptes Rendus de l'Académie des Sciences, Série I, **346**, 589–592 (2008).
3. Candès, E., Tao. T.: Decoding by linear programing. IEEE Trans. Inf. Theory **51**, 4203–4215 (2005).
4. Davis, M.E., Blumensath, T.: Iterative hard thresholding for compressed sensing. Appl. Comput. Harmon. Anal. **27**, 265274 (2009).
5. Foucart, S.: A note on guaranteed sparse recovery via $\ell_1$-minimization. Applied and Comput. Harmonic Analysis, To appear.
6. Garg, R., Khandekar, R.: Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property. In: Bottou, L., Littman, M. (eds.) Proceedings of the 26 th International Confer- ence on Machine Learning, pp. 337-344.
7. Needell, D., Tropp, J.A.: CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26** 301–321 (2009).