Sparsest Solutions of Underdetermined Linear Systems via ℓ_q-minimization for $0 < q \leq 1$

Simon Foucart*
Department of Mathematics
Vanderbilt University
Nashville, TN 37240

Ming-Jun Lai†
Department of Mathematics
University of Georgia
Athens, GA 30602

September 9, 2008

Abstract

We present a condition on the matrix of an underdetermined linear system which guarantees that the solution of the system with minimal ℓ_q-quasinorm is also the sparsest one. This generalizes, and slightly improves, a similar result for the ℓ_1-norm. We then introduce a simple numerical scheme to compute solutions with minimal ℓ_q-quasinorm, and we study its convergence. Finally, we display the results of some experiments which indicate that the ℓ_q-method performs better than other available methods.

1 Introduction

Our objective in this paper is to find the sparsest solutions of a linear system $A\mathbf{z} = \mathbf{y}$. Here we think of the fixed vector $\mathbf{y}$ as an incomplete set of m linear measurements taken of a signal $\mathbf{x} \in \mathbb{R}^N$, thus it is represented as $\mathbf{y} = A\mathbf{x}$ for some $m \times N$ matrix A. Since the number m of measurements is smaller than the dimension N of the signal space – typically, much smaller – the linear system $A\mathbf{z} = \mathbf{y}$ has many solutions, among which we wish to single out the sparsest ones, i.e. the solutions of $A\mathbf{z} = \mathbf{y}$ with a minimal number of nonzero components. Following the tradition, we write $\|\mathbf{z}\|_0$ for the number of nonzero components of a vector $\mathbf{z}$, and we rephrase the problem as

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_0 \qquad \text{subject to} \quad A\mathbf{z} = \mathbf{y}. \tag{P$_0$}$$

*simon.foucart@vanderbilt.edu
†mjlai@math.uga.edu. This author is supported by the National Science Foundation under grant #DMS 0713807.

One can easily observe that a solution $\mathbf{z}$ of (P_0) is guaranteed to be unique as soon as $2\|\mathbf{z}\|_0 < \mathrm{spark(A)}$, where $\mathrm{spark(A)} \leq \mathrm{rank}(A) + 1$ is the smallest integer σ for which σ columns of A are linearly dependent, see [8]. Uniqueness can also be characterized in terms of the Restricted Isometry Constants δ_k of the matrix A. We recall that these are the smallest constants $0 < \delta_k \leq 1$ for which the matrix A satisfies the Restricted Isometry Property of order k, that is

$$(1-\delta_k)\|\mathbf{z}\|_2^2 \leq \|A\mathbf{z}\|_2^2 \leq (1+\delta_k)\|\mathbf{z}\|_2^2 \qquad \text{whenever } \|\mathbf{z}\|_0 \leq k. \tag{1}$$

It is then easy to observe that any s-sparse vector $\mathbf{x}$ is recovered via the minimization (P_0) in which $\mathbf{y} = A\mathbf{x}$ if and only if the strict inequality $\delta_{2s} < 1$ holds, see e.g. [4].

However appealing (P_0) might seem, it remains an NP-problem [11] that cannot be solved in practice. Nonetheless, assuming certain conditions on the matrix A, alternative strategies to find sparsest solutions have been put forward, such as orthogonal greedy algorithms or basis pursuit. The latter replaces the problem (P_0) by the ℓ_1-minimization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_1 \qquad \text{subject to} \quad A\mathbf{z} = \mathbf{y}. \tag{P_1}$$

Candès and Tao [5] showed for instance that any s-sparse vector is exactly recovered via the minimization (P_1) as soon as $\delta_{3s} + 3\delta_{4s} < 2$. Note that a condition involving only δ_{2s} would seem more natural, in view of the previous considerations. Candès provided just that in [2] when he established exact recovery of s-sparse vectors via ℓ_1-minimization under the condition

$$\delta_{2s} < \sqrt{2} - 1 \approx 0.4142. \tag{2}$$

We shall now adopt a strategy that lies between the minimizations (P_0) and (P_1). Namely, we consider, for some $0 < q \leq 1$, the minimization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_q \qquad \text{subject to} \quad A\mathbf{z} = \mathbf{y}. \tag{P_q}$$

This is by no means a brand new approach. Gribonval and Nielsen, see e.g. [10], studied the ℓ_q-minimization in terms of Coherence. Chartrand [7] studied it in terms of Restricted Isometry Constants. He stated that s-sparse vectors can be exactly recovered by solving (P_q) under the assumption that $\delta_{as} + b\delta_{(a+1)s} < b - 1$ holds for some $b > 1$ and $a := b^{q/(2-q)}$. He then claimed that exact recovery of s-sparse vectors can be obtained from the solution of (P_q) for some $q > 0$ small enough, provided that $\delta_{2s+1} < 1$. There was a minor imprecision in his arguments, as he neglected the fact that as must be an integer when he chose the number a under the requirement $1 < a < 1 + 1/s$. A correct justification would be to define $a := 1 + 1/s$, so that the sufficient condition $\delta_{as} + b\delta_{(a+1)s} < b - 1$, where $b := a^{(2-q)/q} > 1$, becomes feasible for $q > 0$ small enough as long as $\delta_{2s+1} < 1$.

Let us describe our contribution to the question while explaining the organization of the paper. In Section 2, we discuss exact recovery from perfect data via ℓ_q-minimization. In particular, we derive from Theorem 2.1 a sufficient condition slightly weaker than (2), as

well as another version of Chartrand's result. Theorem 2.1 actually follows from the more general Theorem 3.1, which is stated and proved in Section 3. This theorem deals with the more realistic situation of a measurement $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ containing a perturbation vector $\mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \vartheta$ for some fixed amount $\vartheta \geq 0$. This framework is exactly the one introduced by Candès, Romberg, and Tao in [3] for the case $q = 1$. Next, in Section 4, we propose a numerical algorithm to approximate the minimization (P_q). We then discuss convergence issues and prove that the output of the algorithm is not merely an approximation, but is in fact exact. Finally, we compare in Section 5 our ℓ_q-algorithm with four existing methods: the orthogonal greedy algorithm, see e.g. [13], the regularized orthogonal matching pursuit, see [12], the ℓ_1-minimization, and the reweighted ℓ_1-minimization, see [6]. The last two, as well as our ℓ_q-algorithm, use the ℓ_1-magic software available on Candès' web page. It comes as a small surprise that the ℓ_q-method performs best.

2 Exact recovery via ℓ_q-minimization

Our main theorem is similar in flavor to many previous ones – in fact, its proof is inspired by theirs – except that we avoid Restricted Isometry Constants, as we felt that the non-homogeneity of the Restricted Isometry Property (1) contradicted the consistency of the problem with respect to measurement amplification, or in other words, that it was in conflict with the equivalence of all the linear systems $(c\,A)\mathbf{z} = c\,\mathbf{y}$, $c \in \mathbb{R}$. Instead, we introduce $\alpha_k, \beta_k \geq 0$ to be the best constants in the inequalities

$$\alpha_k \|\mathbf{z}\|_2 \leq \|A\mathbf{z}\|_2 \leq \beta_k \|\mathbf{z}\|_2, \qquad \|\mathbf{z}\|_0 \leq k.$$

Our results are to be stated in terms of a quantity invariant under the change $A \leftarrow c\,A$, namely

$$\gamma_{2s} := \frac{{\beta_{2s}}^2}{{\alpha_{2s}}^2} \geq 1.$$

The quantity γ_{2s} can be made arbitrarily close to 1 by taking the entries of A to be e.g. independent realizations of Gaussian random variables of mean zero and identical variance, provided that $m \geq c \cdot s \log(N/s)$, where c is a constant depending on $\gamma_{2s} - 1$. We refer the reader to [1] for a precise statement and a simple proof based on concentration of measure inequalities.

We start by illustrating Theorem 3.1 in the special case of s-sparse vectors that are measured with infinite precision, which means that both the error $\sigma_s(\mathbf{x})_q$ of best s-term approximation to $\mathbf{x}$ with respect to the ℓ_q-quasinorm and the relative measurement error θ are equal to zero.

Theorem 2.1 *Given* $0 < q \leq 1$, *if*

$$\gamma_{2t} - 1 \;<\; 4(\sqrt{2} - 1)\left(\frac{t}{s}\right)^{1/q - 1/2} \qquad \textit{for some integer } t \geq s, \tag{3}$$

then every s-sparse vector is exactly recovered by solving (P_q).

Let us remark that, in practice, we do not solve (P_q) but an approximated problem, which still yields exact solutions. There are two special instances of the above result that are worth pointing out. The first one corresponds to the choices $t = s$ and $q = 1$.

Corollary 2.1 *Under the assumption that*

$$\gamma_{2s} < 4\sqrt{2} - 3 \approx 2.6569, \tag{4}$$

every s-sparse vector is exactly recovered by solving (P_1).

This slightly improves Candès' condition (2), since the constant γ_{2s} is expressed in terms of the Restricted Isometry Constant δ_{2s} as

$$\gamma_{2s} = \frac{1 + \delta_{2s}}{1 - \delta_{2s}},$$

hence the condition (4) becomes $\delta_{2s} < 2(3 - \sqrt{2})/7 \approx 0.4531$.

The second special instance we are pointing out corresponds to the choice $t = s + 1$. In this case, Condition (3) reads

$$\gamma_{2s+2} < 1 + 4(\sqrt{2} - 1)\left(1 + \frac{1}{s}\right)^{1/q-1/2}.$$

The right-hand side of this inequality tends to infinity as q approaches zero. The following result is then straightforward.

Corollary 2.2 *Under the assumption that*

$$\gamma_{2s+2} < +\infty,$$

every s-sparse vector is exactly recovered by solving (P_q) for some $q > 0$ small enough.

Let us note that the condition $\gamma_{2s+2} < +\infty$ is equivalent to the condition $\delta_{2s+2} < 1$. This result is almost optimal, since it says that if one could recover every $(s + 1)$-sparse vector via the theoretical program (P_0), then one can actually recover every s-sparse vector via the program (P_q) for some $q > 0$.

Theorem 2.1 is an immediate consequence of Theorem 3.1 to be given in the next section, hence we do not provide a separate proof. Let us nonetheless comment briefly on this potential proof, as it helps to elucidate the structure of the main proof. Let us consider first a vector $\mathbf{v}$ in the null-space of A and an index set S with $|S| \leq s$. The vector $\mathbf{v}_S$, i.e. the vector which equals $\mathbf{v}$ on S and vanishes on the complement $\overline{S}$ of S in $\{1, \ldots, N\}$, has the same image under A as the vector $-\mathbf{v}_{\overline{S}}$. Since $\mathbf{v}_S$ is s-sparse, the anticipated result implies $\|\mathbf{v}_S\|_q < \|\mathbf{v}_{\overline{S}}\|_q$, unless $\mathbf{v}_S = \mathbf{v}_{\overline{S}}$, i.e. $\mathbf{v} = 0$. This necessary condition turns out to be sufficient, too. It is established in Step 1 of the main proof using the assumption on γ_{2t}. Then, using the ℓ_q-minimization in Step 2, we establish a reverse inequality $\|\mathbf{v}_{\overline{S}}\|_q \leq \|\mathbf{v}_S\|_q$ for the support S of the vector $\mathbf{x}$ and for $\mathbf{v} := \mathbf{x} - \mathbf{x}^*$, where $\mathbf{x}^*$ is a solution of (P_q). Clearly, the two inequalities imply that $\mathbf{v} = 0$, or equivalently that any solution $\mathbf{x}^*$ of (P_q) equals the original vector $\mathbf{x}$, as expected.

3 Approximate recovery from imperfect data

We now consider the situation where the measurements $\mathbf{y}$ are moderately flawed, i.e. we suppose

$$\|A\mathbf{x} - \mathbf{y}\|_2 \leq \beta_{2s} \cdot \theta.$$

Note that θ represents a *relative* error between accurate and inaccurate measurements, so that it makes the previous bound invariant under the change $A \leftarrow c\,A$, $y \leftarrow c\,y$. This differs slightly from the formulation of [3] for $q = 1$, where the *absolute* error was considered. Of course, the choice of the homogeneous constant β_{2s} is somewhat arbitrary, it is merely dictated by the nice estimates (5) and (6). In order to approximately recover the original vector $\mathbf{x} \in \mathbb{R}^N$ from the knowledge of $\mathbf{y}$, we shall solve the minimization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_q \qquad \text{subject to} \quad \|A\mathbf{z} - \mathbf{y}\|_2 \leq \beta_{2s} \cdot \theta. \tag{$\mathrm{P}_{q,\theta}$}$$

Before anything else, let us make sure that this minimization is solvable.

Lemma 3.1 *A solution of ($P_{q,\theta}$) exists for any $0 < q \leq 1$ and any $\theta \geq 0$.*

Proof. Let κ be the value of the minimum in ($\mathrm{P}_{q,\theta}$). It is straightforward to see that ($\mathrm{P}_{q,\theta}$) is equivalent to, say, the minimization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}}\ \|\mathbf{z}\|_q \qquad \text{subject to} \quad \|\mathbf{z}\|_q \leq 2\kappa \quad \text{and} \quad \|A\mathbf{z} - \mathbf{y}\|_2 \leq \beta_{2s} \cdot \theta.$$

Because the set $\left\{\mathbf{z} \in \mathbb{R}^N : \|\mathbf{z}\|_q \leq 2\kappa, \|A\mathbf{z} - \mathbf{y}\|_2 \leq \beta_{2s} \cdot \theta\right\}$ is compact and because the ℓ_q-quasinorm is a continuous function, we can conclude that a minimizer exists. ■

We are now in a position to state the main theoretical result of the paper. In what follows, the quantity $\sigma_s(\mathbf{x})_q$ denotes the error of best s-term approximation to $\mathbf{x}$ with respect to the ℓ_q-quasinorm, that is

$$\sigma_s(\mathbf{x})_q := \inf_{\|\mathbf{z}\|_0 \leq s} \|\mathbf{x} - \mathbf{z}\|_q.$$

Theorem 3.1 *Given $0 < q \leq 1$, if Condition (3) holds, i.e. if*

$$\gamma_{2t} - 1 \ < \ 4(\sqrt{2} - 1)\left(\frac{t}{s}\right)^{1/q-1/2} \qquad \textit{for some integer } t \geq s,$$

then a solution $\mathbf{x}^*$ of ($P_{q,\theta}$) approximate the original vector $\mathbf{x}$ with errors

$$\|\mathbf{x} - \mathbf{x}^*\|_q \ \leq \ C_1 \cdot \sigma_s(\mathbf{x})_q \ + D_1 \cdot s^{1/q-1/2} \cdot \theta, \tag{5}$$

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \ \leq \ C_2 \cdot \frac{\sigma_s(\mathbf{x})_q}{t^{1/q-1/2}} + D_2 \cdot \theta. \tag{6}$$

The constants C_1, C_2, D_1, and D_2 depend only on q, γ_{2t}, and the ratio s/t.

Proof. The proof involves some properties of the ℓ_q-quasinorm which must be recalled. Namely, for any vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, one has

$$\|\mathbf{u}\|_1 \leq \|\mathbf{u}\|_q, \qquad \|\mathbf{u}\|_q \leq n^{1/q-1/2}\,\|\mathbf{u}\|_2, \qquad \|\mathbf{u}+\mathbf{v}\|_q^q \leq \|\mathbf{u}\|_q^q + \|\mathbf{v}\|_q^q. \tag{7}$$

Step 1: Consequence of the assumption on γ_{2t}.
We consider an arbitrary index set $S =: S_0$ with $|S| \leq s$. Let $\mathbf{v}$ be a vector in $\mathbb{R}^N$, which will often — but not always — be an element of the null-space of A. For instance, we will take $\mathbf{v} := \mathbf{x} - \mathbf{x}^*$ in Step 2. We partition the complement of S in $\{1,\ldots,N\}$ as $\overline{S} = S_1 \cup S_2 \cup \ldots$, where

$$\begin{aligned} S_1 &:= \{\text{indices of the } t \text{ largest absolute-value components of } \mathbf{v} \text{ in } \overline{S}\},\\ S_2 &:= \{\text{indices of the next } t \text{ largest absolute-value components of } \mathbf{v} \text{ in } \overline{S}\},\\ &\vdots \end{aligned}$$

We first observe that

$$\|\mathbf{v}_{S_0}\|_2^2 + \|\mathbf{v}_{S_1}\|_2^2 = \|\mathbf{v}_{S_0} + \mathbf{v}_{S_1}\|_2^2 \;\leq\; \frac{1}{\alpha_{2t}^2}\,\|A(\mathbf{v}_{S_0} + \mathbf{v}_{S_1})\|_2^2 \tag{8}$$

$$= \frac{1}{\alpha_{2t}^2}\,\langle A(\mathbf{v} - \mathbf{v}_{S_2} - \mathbf{v}_{S_3} - \cdots), A(\mathbf{v}_{S_0} + \mathbf{v}_{S_1})\rangle \tag{9}$$

$$= \frac{1}{\alpha_{2t}^2}\,\langle A\mathbf{v}, A(\mathbf{v}_{S_0} + \mathbf{v}_{S_1})\rangle + \frac{1}{\alpha_{2t}^2}\sum_{k\geq 2}\big[\langle A(-\mathbf{v}_{S_k}), A\mathbf{v}_{S_0}\rangle + \langle A(-\mathbf{v}_{S_k}), A\mathbf{v}_{S_1}\rangle\big]. \tag{10}$$

Let us renormalize the vectors $-\mathbf{v}_{S_k}$ and $\mathbf{v}_{S_0}$ so that their ℓ_2-norms equal one by setting $\mathbf{u}_k := -\mathbf{v}_{S_k}/\|\mathbf{v}_{S_k}\|_2$ and $\mathbf{u}_0 := \mathbf{v}_{S_0}/\|\mathbf{v}_{S_0}\|_2$. We then obtain

$$\begin{aligned} \frac{\langle A(-\mathbf{v}_{S_k}), A\mathbf{v}_{S_0}\rangle}{\|\mathbf{v}_{S_k}\|_2\,\|\mathbf{v}_{S_0}\|_2} &= \langle A\mathbf{u}_k, A\mathbf{u}_0\rangle = \frac{1}{4}\,\big[\|A(\mathbf{u}_k+\mathbf{u}_0)\|_2^2 - \|A(\mathbf{u}_k-\mathbf{u}_0)\|_2^2\big]\\ &\leq \frac{1}{4}\,\big[\beta_{2t}^2\,\|\mathbf{u}_k+\mathbf{u}_0\|_2^2 - \alpha_{2t}^2\,\|\mathbf{u}_k-\mathbf{u}_0\|_2^2\big] = \frac{1}{2}\,[\beta_{2t}^2 - \alpha_{2t}^2]. \end{aligned}$$

With a similar argument with S_1 in place of S_0, we can derive

$$\langle A(-\mathbf{v}_{S_k}), A\mathbf{v}_{S_0}\rangle + \langle A(-\mathbf{v}_{S_k}), A\mathbf{v}_{S_1}\rangle \leq \frac{\beta_{2t}^2 - \alpha_{2t}^2}{2}\,\|\mathbf{v}_{S_k}\|_2\big[\|\mathbf{v}_{S_0}\|_2 + \|\mathbf{v}_{S_1}\|_2\big]. \tag{11}$$

Besides, we have

$$\langle A\mathbf{v}, A(\mathbf{v}_{S_0} + \mathbf{v}_{S_1})\rangle \leq \|A\mathbf{v}\|_2 \cdot \|A(\mathbf{v}_{S_0} + \mathbf{v}_{S_1})\|_2 \leq \|A\mathbf{v}\|_2 \cdot \beta_{2t}\,\big[\|\mathbf{v}_{S_0}\|_2 + \|\mathbf{v}_{S_1}\|_2\big]. \tag{12}$$

Substituting the inequalities (11) and (12) into (10), we have

$$\|\mathbf{v}_{S_0}\|_2^2 + \|\mathbf{v}_{S_1}\|_2^2 \leq \Big(\frac{\gamma_{2t}}{\beta_{2t}}\|A\mathbf{v}\|_2 + \frac{\gamma_{2t}-1}{2}\sum_{k\geq 2}\|\mathbf{v}_{S_k}\|_2\Big)\big[\|\mathbf{v}_{S_0}\|_2 + \|\mathbf{v}_{S_1}\|_2\big].$$

With $c := \gamma_{2t}/\beta_{2t} \cdot \|A\mathbf{v}\|_2$, $d := (\gamma_{2t}-1)/2$, and $\Sigma := \sum_{k\geq 2} \|\mathbf{v}_{S_k}\|_2$, this reads

$$\Big[\|\mathbf{v}_{S_0}\|_2 - \frac{c+d\Sigma}{2}\Big]^2 + \Big[\|\mathbf{v}_{S_1}\|_2 - \frac{c+d\Sigma}{2}\Big]^2 \leq \frac{(c+d\Sigma)^2}{2}.$$

The above inequality easily implies

$$\|\mathbf{v}_{S_0}\|_2 \leq \frac{c+d\Sigma}{2} + \frac{c+d\Sigma}{\sqrt{2}} = \frac{(1+\sqrt{2})}{2} \cdot (c+d\Sigma). \tag{13}$$

By Hölder's inequality mentioned in (7), we get

$$\|\mathbf{v}_{S_0}\|_q \leq s^{1/q-1/2}\, \|\mathbf{v}_{S_0}\|_2 \leq \frac{1+\sqrt{2}}{2} \cdot (c+d\Sigma) \cdot s^{1/q-1/2}. \tag{14}$$

It now remains to bound Σ. Given an integer $k \geq 2$, let us consider $i \in S_k$ and $j \in S_{k-1}$. From the inequality $|v_i| \leq |v_j|$ raised to the power q, we derive that $|v_i|^q \leq t^{-1}\|\mathbf{v}_{S_{k-1}}\|_q^q$ by averaging over j. In turn, this yields the inequality $\|\mathbf{v}_{S_k}\|_2^2 \leq t^{1-2/q}\|\mathbf{v}_{S_{k-1}}\|_q^2$ by raising to the power $2/q$ and summing over i. It follows that

$$\Sigma = \sum_{k\geq 2} \|\mathbf{v}_{S_k}\|_2 \leq t^{1/2-1/q} \sum_{k\geq 1} \|\mathbf{v}_{S_k}\|_q \leq t^{1/2-1/q} \Big[\sum_{k\geq 1} \|\mathbf{v}_{S_k}\|_q^q\Big]^{1/q} = t^{1/2-1/q}\|\mathbf{v}_{\overline{S}}\|_q.$$

Combining the above inequality with (14), we obtain the partial conclusion:

$$\|\mathbf{v}_S\|_q \leq \frac{\lambda}{2\beta_{2t}} \cdot \|A\mathbf{v}\|_2 \cdot s^{1/q-1/2} + \mu \cdot \|\mathbf{v}_{\overline{S}}\|_q, \qquad \mathbf{v} \in \mathbb{R}^N,\ |S| \leq s, \tag{15}$$

where the constants λ and μ are given by

$$\lambda := (1+\sqrt{2})\gamma_{2t} \quad \text{and} \quad \mu := \frac{1}{4}(1+\sqrt{2})(\gamma_{2t}-1)\Big(\frac{s}{t}\Big)^{1/q-1/2}. \tag{16}$$

Note that the assumption on γ_{2t} translates into the inequality $\mu < 1$.

Step 2: Consequence of the ℓ_q-minimization.

Now let S be specified as the set of indices of the s largest absolute-value components of $\mathbf{x}$, and let $\mathbf{v}$ be specified as $\mathbf{v} := \mathbf{x} - \mathbf{x}^*$. Because $\mathbf{x}^*$ is a minimizer of $(\mathrm{P}_{q,\theta})$, we have

$$\|\mathbf{x}\|_q^q \ \geq\ \|\mathbf{x}^*\|_q^q, \qquad \text{i.e.} \qquad \|\mathbf{x}_S\|_q^q + \|\mathbf{x}_{\overline{S}}\|_q^q \ \geq\ \|\mathbf{x}_S^*\|_q^q + \|\mathbf{x}_{\overline{S}}^*\|_q^q.$$

By the triangular inequality mentioned in (7), we obtain

$$\|\mathbf{x}_S\|_q^q + \|\mathbf{x}_{\overline{S}}\|_q^q \ \geq\ \|\mathbf{x}_S\|_q^q - \|\mathbf{v}_S\|_q^q \ +\ \|\mathbf{v}_{\overline{S}}\|_q^q - \|\mathbf{x}_{\overline{S}}\|_q^q.$$

Rearranging the latter yields the inequality

$$\|\mathbf{v}_{\overline{S}}\|_q^q \ \leq\ 2\,\|\mathbf{x}_{\overline{S}}\|_q^q + \|\mathbf{v}_S\|_q^q = 2\,\sigma_s(\mathbf{x})_q^q + \|\mathbf{v}_S\|_q^q. \tag{17}$$

Step 3: Error estimates.
We now take into account the bound

$$\|A\mathbf{v}\|_2 = \|A\mathbf{x} - A\mathbf{x}^*\|_2 \le \|A\mathbf{x} - \mathbf{y}\|_2 + \|A\mathbf{x}^* - \mathbf{y}\|_2 \le \beta_{2s} \cdot \theta + \beta_{2s} \cdot \theta \le 2\beta_{2t} \cdot \theta.$$

For the ℓ_q-error, we combine the estimates (15) and (17) to get

$$\|\mathbf{v}_{\overline{S}}\|_q^q \le 2\,\sigma_s(\mathbf{x})_q^q + \lambda^q \cdot s^{1-q/2} \cdot \theta^q + \mu^q \cdot \|\mathbf{v}_{\overline{S}}\|_q^q.$$

As a consequence of $\mu < 1$, we now obtain

$$\|\mathbf{v}_{\overline{S}}\|_q^q \le \frac{2}{1-\mu^q} \cdot \sigma_s(\mathbf{x})_q^q + \frac{\lambda^q}{1-\mu^q} \cdot s^{1-q/2} \cdot \theta^q.$$

Using the estimate (15) once more, we can derive that

$$\begin{aligned}
\|\mathbf{v}\|_q &= [\,\|\mathbf{v}_S\|_q^q + \|\mathbf{v}_{\overline{S}}\|_q^q\,]^{1/q} \le \big[\,(1+\mu^q) \cdot \|\mathbf{v}_{\overline{S}}\|_q^q + \lambda^q \cdot s^{1-q/2} \cdot \theta^q\,\big]^{1/q} \\
&\le \left[\frac{2\,(1+\mu^q)}{1-\mu^q} \cdot \sigma_s(\mathbf{x})_q^q + \frac{2\,\lambda^q}{1-\mu^q} \cdot s^{1-q/2} \cdot \theta^q\right]^{1/q} \\
&\le 2^{1/q-1}\left[\frac{2^{1/q}\,(1+\mu^q)^{1/q}}{(1-\mu^q)^{1/q}} \cdot \sigma_s(\mathbf{x})_q + \frac{2^{1/q}\,\lambda}{(1-\mu^q)^{1/q}} \cdot s^{1/q-1/2} \cdot \theta\right],
\end{aligned}$$

where we have made use of the inequality $[a^q + b^q]^{1/q} \le 2^{1/q-1}[a+b]$ for $a, b \ge 0$. The estimate (5) follows with

$$C_1 := \frac{2^{2/q-1}\,(1+\mu^q)^{1/q}}{(1-\mu^q)^{1/q}} \qquad \text{and} \qquad D_1 := \frac{2^{2/q-1}\,\lambda}{(1-\mu^q)^{1/q}}.$$

As for the ℓ_2-error, we remark that the bound (13) also holds for $\|\mathbf{v}_{S_1}\|_2$ in place of $\|\mathbf{v}_{S_0}\|_2$ to obtain

$$\|\mathbf{v}\|_2 = \Big[\sum_{k\ge 0} \|\mathbf{v}_{S_k}\|_2^2\Big]^{1/2} \le \sum_{k\ge 0} \|\mathbf{v}_{S_k}\|_2 \le (1+\sqrt{2}) \cdot (c + d\Sigma) + \Sigma \le \nu \cdot \Sigma + 2\lambda \cdot \theta,$$

where $\nu := (\lambda + 1 - \sqrt{2})/2$. Then, in view of the bound

$$\Sigma \le t^{1/2-1/q}\,\|\mathbf{v}_{\overline{S}}\|_q \le t^{1/2-1/q} \cdot 2^{1/q-1} \cdot \Big[\frac{2^{1/q}}{(1-\mu^q)^{1/q}} \cdot \sigma_s(\mathbf{x})_q + \frac{\lambda}{(1-\mu^q)^{1/q}} \cdot s^{1/q-1/2} \cdot \theta\Big],$$

we may finally conclude that

$$\|\mathbf{v}\|_2 \le \frac{2^{2/q-1}\,\nu}{(1-\mu^q)^{1/q}} \cdot \frac{\sigma_s(\mathbf{x})_q}{t^{1/q-1/2}} + \Big[\frac{2^{1/q-1}\,\lambda\,\nu}{(1-\mu^q)^{1/q}} \cdot \Big(\frac{s}{t}\Big)^{1/q-1/2} + 2\lambda\Big]\theta.$$

This leads to the estimate (6) with

$$C_2 := \frac{2^{2/q-2}\,(\lambda + 1 - \sqrt{2})}{(1-\mu^q)^{1/q}} \qquad \text{and} \qquad D_2 := \frac{2^{1/q-2}\,\lambda\,(\lambda + 1 - \sqrt{2})}{(1-\mu^q)^{1/q}} + 2\lambda.$$

The reader is invited to verify that the constants C_1, D_1, C_2, and D_2 depend only on q, γ_{2t}, and the ratio s/t. However, they grow exponentially fast when q tends to zero. ■

4 Description of the algorithm

We assume from now on that $\mathbf{x}$ is an s-sparse vector. The minimization problem (P_q) suggested to recover $\mathbf{x}$ is nonconvex, so it needs to be approximated. We propose in this section an algorithm to compute a minimizer of the approximated problem, for which we give an informal but detailed justification.

We shall proceed iteratively, starting from a vector $\mathbf{z_0}$ satisfying $A\mathbf{z_0} = \mathbf{y}$, which is a reasonable guess for $\mathbf{x}$, and constructing a sequence $(\mathbf{z}_n)$ recursively by defining $\mathbf{z}_{n+1}$ as a solution of the minimization problem

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \sum_{i=1}^{N} \frac{|z_i|}{(|z_{n,i}| + \epsilon_n)^{1-q}} \qquad \text{subject to} \quad A\mathbf{z} = \mathbf{y}. \tag{18}$$

Here, the sequence (ϵ_n) is a nonincreasing sequence of positive numbers. It might be prescribed from the start or defined during the iterative process. In practice, we will take $\lim_{n\to\infty} \epsilon_n = 0$ to facilitate the use of Proposition 4.2. However, we also allow the case $\lim_{n\to\infty} \epsilon_n > 0$ in order not to exclude constant sequences (ϵ_n) from the theory. We point out that the scheme is easy to implement, since each step reduces to an ℓ_1-minimization problem (P_1) relatively to the renormalized matrix

$$A_n := A \times \mathrm{Diag}[(|z_{n,i}| + \epsilon_n)^{1-q}].$$

We shall now concentrate on convergence issues. We start with the following lemma.

Proposition 4.1 *For any nonincreasing sequence (ϵ_n) of positive numbers and for any initial vector $\mathbf{z}_0$ satisfying $A\mathbf{z}_0 = \mathbf{y}$, the sequence $(\mathbf{z}_n)$ defined by (18) admits a convergent subsequence.*

Proof. Using the monotonicity of the sequence (ϵ_n), Hölder's inequality, and the minimality property of $\mathbf{z}_{n+1}$, we may write

$$\begin{aligned}
\sum_{i=1}^{N} (|z_{n+1,i}| + \epsilon_{n+1})^q &\leq \sum_{i=1}^{N} \frac{(|z_{n+1,i}| + \epsilon_n)^q}{(|z_{n,i}| + \epsilon_n)^{q(1-q)}} \cdot (|z_{n,i}| + \epsilon_n)^{q(1-q)} \\
&\leq \Big[\sum_{i=1}^{N} \frac{|z_{n+1,i}| + \epsilon_n}{(|z_{n,i}| + \epsilon_n)^{1-q}}\Big]^q \Big[\sum_{i=1}^{N} (|z_{n,i}| + \epsilon_n)^q\Big]^{1-q} \\
&\leq \Big[\sum_{i=1}^{N} \frac{|z_{n,i}| + \epsilon_n}{(|z_{n,i}| + \epsilon_n)^{1-q}}\Big]^q \Big[\sum_{i=1}^{N} (|z_{n,i}| + \epsilon_n)^q\Big]^{1-q},
\end{aligned}$$

that is to say

$$\sum_{i=1}^{N} (|z_{n+1,i}| + \epsilon_{n+1})^q \leq \sum_{i=1}^{N} (|z_{n,i}| + \epsilon_n)^q. \tag{19}$$

In particular, we obtain

$$\|\mathbf{z}_n\|_\infty \le \|\mathbf{z}_n\|_q \le \Big[\sum_{i=1}^N (|z_{n,i}| + \epsilon_n)^q\Big]^{1/q} \le \Big[\sum_{i=1}^N (|z_{0,i}| + \epsilon_0)^q\Big]^{1/q}.$$

The boundedness of $(\mathbf{z}_n)$ implies the existence of a convergent subsequence. ■

Unfortunately, the convergence of the whole sequence $(\mathbf{z}_n)$ could not be established rigorously. However, several points beside the numerical experiments of Section 5 hint at its convergence to the original s-sparse vector $\mathbf{x}$. First, with $\epsilon := \lim_{n\to\infty} \epsilon_n$ and in view of the inequality (19), it is reasonable to expect that any cluster point of the sequence $(\mathbf{z}_n)$ is a minimizer of

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \ \sum_{i=1}^N (|z_i| + \epsilon)^q \qquad \text{subject to} \quad A\mathbf{z} = \mathbf{y}, \tag{$\mathcal{P}_{q,\epsilon}$}$$

at least under appropriate conditions on $\mathbf{z}_0$ — this question is further discussed in the remark at the end of this section. In the case of a constant sequence (ϵ_n), for instance, if $\mathbf{z}_0$ is chosen as a minimizer of $(\mathcal{P}_{q,\epsilon})$, then $\mathbf{z}_n$ is also a minimizer of $(\mathcal{P}_{q,\epsilon})$ for every $n \ge 0$. Then, as we shall see in Proposition 4.3, when $\epsilon > 0$ is small enough, any minimizer of $(\mathcal{P}_{q,\epsilon})$ turns out to be the vector $\mathbf{x}$ itself, provided that Condition (3) is fulfilled. Thus, under Condition (3) and the appropriate conditions on $\mathbf{z}_0$, we can expect $\mathbf{x}$ to be a cluster point of the sequence $(\mathbf{z}_n)$. This implies, by virtue of the forthcoming Proposition 4.2, that $\mathbf{z}_n$ is actually equal to $\mathbf{x}$ for n large enough. Proposition 4.2 is noteworthy: it states that the algorithm (18) recovers the vector $\mathbf{x}$ exactly, not just approximately. The proof is based on the following lemma, of independent interest.

Lemma 4.1 *Given an s-sparse vector $\mathbf{x}$ supported on a set S, and given a weight vector $\mathbf{w} \in \mathbb{R}^N_+$, if*

$$\gamma_{2t} - 1 \ < \ 4(\sqrt{2}-1)\left(\frac{t}{s}\right)^{1/2} \cdot \frac{\min_{i\in S} w_i}{\max_{i\in \overline{S}} w_i} \qquad \textit{for some integer } t \ge s, \tag{20}$$

then the vector $\mathbf{x}$ is exactly recovered by the minimization

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\textit{minimize}} \ \sum_{i=1}^N \frac{|z_i|}{w_i} \qquad \textit{subject to} \quad A\mathbf{z} = \mathbf{y}. \tag{P_w}$$

Proof. Consider the weighted ℓ_1-norm defined by

$$\|\mathbf{z}\|_w := \sum_{i=1}^N \frac{|z_i|}{w_i}, \qquad \mathbf{z} \in \mathbb{R}^N.$$

Let $\bar{\mathbf{z}}$ be a minimizer of (P_w). Our objective is to show that $\mathbf{v} := \mathbf{x} - \bar{\mathbf{z}}$ equals zero. We shall follow the proof of Theorem 3.1. First, if S denotes the support of $\mathbf{x}$, we can reproduce Step 2 to get

$$\|\mathbf{v}_{\overline{S}}\|_w \;\leq\; \|\mathbf{v}_S\|_w. \tag{21}$$

On the other hand, since $\mathbf{v} \in \ker A$, an estimate analogous to (13) reads

$$\|\mathbf{v}_S\|_2 \leq \frac{1+\sqrt{2}}{4}(\gamma_{2t}-1) \cdot \sum_{k\geq 2} \|\mathbf{v}_{T_k}\|_2, \tag{22}$$

where we have partitioned $\overline{S}$ as $T_1 \cup T_2 \cup \ldots$ with

$$\begin{aligned} T_1 &:= \{\text{indices of the } t \text{ largest values of } |v_i|/w_i \text{ in } \overline{S}\}, \\ T_2 &:= \{\text{indices of the next } t \text{ largest values of } |v_i|/w_i \text{ in } \overline{S}\}, \\ &\;\vdots \end{aligned}$$

Let us observe that

$$\|\mathbf{v}_S\|_w = \sum_{i\in S} \frac{|v_i|}{w_i} \leq \Big[\sum_{i\in S} \frac{1}{w_i^2}\Big]^{1/2} \Big[\sum_{i\in S} v_i^2\Big]^{1/2} \leq \frac{1}{\min_{i\in S} w_i} \cdot s^{1/2} \cdot \|\mathbf{v}_S\|_2. \tag{23}$$

The inequalities (23) and (22) therefore imply

$$\|\mathbf{v}_S\|_w \leq \frac{1}{\min_{i\in S} w_i} \cdot \frac{1+\sqrt{2}}{4}(\gamma_{2t}-1) \cdot s^{1/2} \sum_{k\geq 2} \|\mathbf{v}_{T_k}\|_2. \tag{24}$$

Now, given an integer $k \geq 2$, let us consider $i \in T_k$ and $j \in T_{k-1}$. From the inequality $|v_i|/w_i \leq |v_j|/w_j$, we derive that $|v_i| \leq w_i \cdot t^{-1}\|\mathbf{v}_{T_{k-1}}\|_w$ by averaging over j. In turn, this yields the inequality $\|\mathbf{v}_{T_k}\|_2^2 \leq \big[\sum_{i\in T_k} w_i^2\big] \cdot t^{-2}\|\mathbf{v}_{T_{k-1}}\|_w^2 \leq \max_{i\in\overline{S}} w_i^2 \cdot t^{-1}\|\mathbf{v}_{T_{k-1}}\|_w^2$ by squaring and summing over i. It follows that

$$\sum_{k\geq 2} \|\mathbf{v}_{T_k}\|_2 \leq \max_{i\in\overline{S}} w_i \cdot t^{-1/2} \sum_{k\geq 2} \|\mathbf{v}_{T_{k-1}}\|_w \leq \max_{i\in\overline{S}} w_i \cdot t^{-1/2} \|\mathbf{v}_{\overline{S}}\|_w. \tag{25}$$

In view of (24) and (25), we obtain

$$\|\mathbf{v}_S\|_w \leq \frac{\max_{i\in\overline{S}} w_i}{\min_{i\in S} w_i} \cdot \frac{1+\sqrt{2}}{4}(\gamma_{2t}-1) \left(\frac{s}{t}\right)^{1/2} \cdot \|\mathbf{v}_{\overline{S}}\|_w =: \bar{\mu} \cdot \|\mathbf{v}_{\overline{S}}\|_w. \tag{26}$$

The estimates (21) and (26) together imply the conclusion $\mathbf{v} = 0$, provided that $\bar{\mu} < 1$, which simply reduces to Condition (20). ■

Proposition 4.2 *Given $0 < q < 1$ and the original s-sparse vector $\mathbf{x}$, there exists $\eta > 0$ such that, if*

$$\epsilon_n < \eta \qquad \text{and} \qquad \|\mathbf{z}_n - \mathbf{x}\|_\infty < \eta \qquad \text{for some } n, \tag{27}$$

then one has

$$\mathbf{z}_k = \mathbf{x} \qquad \text{for all } k > n.$$

The constant η depends only on q, $\mathbf{x}$, and γ_{2s}.

Proof. Let us denote by S the support of the vector $\mathbf{x}$ and by ξ the positive number defined by $\xi := \min_{i\in S} |x_i|$. We take η small enough so that

$$\gamma_{2s} - 1 < 4(\sqrt{2}-1)\Big(\frac{\xi-\eta}{2\eta}\Big)^{1-q}.$$

The vector $\mathbf{z}_{n+1}$ is obtained from the minimization (P_w) where $w_i := (|z_{n,i}| + \epsilon_n)^{1-q}$. We observe that

$$|z_{n,i}| + \epsilon_n \begin{cases} \geq & |x_i| - |z_{n,i} - x_i| + \epsilon_n \;\geq\; \xi - \eta, \qquad i \in S, \\ \leq & |x_i| + |z_{n,i} - x_i| + \epsilon_n \;\leq\; 2\eta, \qquad i \in \overline{S}. \end{cases}$$

We deduce that

$$\frac{\min_{i\in S} w_i}{\max_{i\in \overline{S}} w_i} \geq \Big(\frac{\xi-\eta}{2\eta}\Big)^{1-q}.$$

Therefore, Condition (20) is fulfilled with $t = s$, and Lemma 4.1 implies that $\mathbf{z}_{n+1} = \mathbf{x}$. The conditions of (27) are now obviously satisfied for $n+1$ instead of n, which implies that $\mathbf{z}_{n+2} = \mathbf{x}$. It follows by immediate induction that $\mathbf{z}_k = \mathbf{x}$ for all $k > n$. ■

To continue this section, we now justify that the minimization $(\mathcal{P}_{q,\epsilon})$ also guarantees exact recovery when $\epsilon > 0$ is small enough. First, we isolate the following lemma.

Lemma 4.2 *Given $0 < q \leq 1$ and an s-sparse vector $\mathbf{x}$, if Condition (3) holds, i.e. if*

$$\gamma_{2t} - 1 \;<\; 4(\sqrt{2}-1)\left(\frac{t}{s}\right)^{1/q-1/2} \qquad \text{for some integer } t \geq s,$$

then for any vector $\mathbf{z}$ satisfying $A\mathbf{z} = \mathbf{y}$, one has

$$\|\mathbf{z} - \mathbf{x}\|_q^q \leq C\left[\|\mathbf{z}\|_q^q - \|\mathbf{x}\|_q^q\right],$$

for some constant C depending only on q, γ_{2t}, and the ratio s/t.

Proof. Let us set $\mathbf{v} := \mathbf{z} - \mathbf{x}$, and let S denote the support of $\mathbf{x}$. We recall that, since $\mathbf{v} \in \ker A$, the estimate (15) yields

$$\|\mathbf{v}_S\|_q \leq \mu \cdot \|\mathbf{v}_{\overline{S}}\|_q, \qquad \mu := \frac{1}{4}(1+\sqrt{2})(\gamma_{2t}-1)\Big(\frac{s}{t}\Big)^{1/q-1/2}. \tag{28}$$

According to Condition (3), we have $\mu < 1$. Let us now observe that

$$\|\mathbf{v}_{\overline{S}}\|_q^q = \|\mathbf{z}_{\overline{S}}\|_q^q = \|\mathbf{z}\|_q^q - \|\mathbf{z}_S\|_q^q \le \|\mathbf{z}\|_q^q - \big(\|\mathbf{x}_S\|_q^q - \|\mathbf{v}_S\|_q^q\big) = \|\mathbf{v}_S\|_q^q + \big[\|\mathbf{z}\|_q^q - \|\mathbf{x}\|_q^q\big].$$

Then, using (28), we derive

$$\|\mathbf{v}_{\overline{S}}\|_q^q \le \mu^q \, \|\mathbf{v}_{\overline{S}}\|_q^q + \big[\|\mathbf{z}\|_q^q - \|\mathbf{x}\|_q^q\big], \quad \text{i.e.} \quad \|\mathbf{v}_{\overline{S}}\|_q^q \le \frac{1}{1-\mu^q}\,\big[\|\mathbf{z}\|_q^q - \|\mathbf{x}\|_q^q\big].$$

Using (28) once more, we obtain

$$\|\mathbf{v}\|_q^q = \|\mathbf{v}_{\overline{S}}\|_q^q + \|\mathbf{v}_S\|_q^q \le (1+\mu^q)\|\mathbf{v}_{\overline{S}}\|_q^q \le \frac{1+\mu^q}{1-\mu^q}\,\big[\|\mathbf{z}\|_q^q - \|\mathbf{x}\|_q^q\big].$$

This is the expected inequality, with $C := (1+\mu^q)/(1-\mu^q)$. ■

Proposition 4.3 *Given $0 < q < 1$ and the original s-sparse vector $\mathbf{x}$, if Condition (3) holds, i.e. if*

$$\gamma_{2t} - 1 \;<\; 4(\sqrt{2}-1)\left(\frac{t}{s}\right)^{1/q-1/2} \qquad \textit{for some integer } t \ge s,$$

then there exists $\zeta > 0$ such that, for any nonnegative ϵ less than ζ, the vector $\mathbf{x}$ is exactly recovered by solving $(\mathcal{P}_{q,\epsilon})$. The constant ζ depends only on N, q, $\mathbf{x}$, γ_{2t}, and the ratio s/t.

Proof. Let $\mathbf{z}_\epsilon$ be a minimizer of $(\mathcal{P}_{q,\epsilon})$. We have

$$\|\mathbf{z}_\epsilon\|_q^q - \|\mathbf{x}\|_q^q = \sum_{i=1}^N |z_{\epsilon,i}|^q - \sum_{i=1}^N |x_i|^q \le \sum_{i=1}^N (|z_{\epsilon,i}|+\epsilon)^q - \Big(\sum_{i=1}^N (|x_i|+\epsilon)^q - \sum_{i=1}^N \epsilon^q\Big) \le N\epsilon^q. \quad (29)$$

We define $\zeta := (CN)^{-1/q}\,\eta$, where η is the constant of Proposition 4.2. Given $\epsilon < \zeta$, we have $\epsilon < \eta$, and, in view of (29), we also have

$$\|\mathbf{z}_\epsilon - \mathbf{x}\|_\infty \le \|\mathbf{z}_\epsilon - \mathbf{x}\|_q \le C^{1/q}\big[\|\mathbf{z}_\epsilon\|_q^q - \|\mathbf{x}\|_q^q\big]^{1/q} \le (CN)^{1/q}\epsilon < \eta.$$

Therefore, according to Proposition 4.2, we infer that $\mathbf{z}_k = \mathbf{x}$ for all $k \ge 1$, where the sequence $(\mathbf{z}_n)$ is defined by the iteration (18) with $\mathbf{z}_0 = \mathbf{z}_\epsilon$ and $\epsilon_n = \epsilon$ for all n. On the other hand, for a vector $\mathbf{z}$ satisfying $A\mathbf{z} = \mathbf{y}$, the inequalities

$$\sum_{i=1}^N (|z_{\epsilon,i}|+\epsilon)^q \le \sum_{i=1}^N (|z_i|+\epsilon)^q \le \Big[\sum_{i=1}^N \frac{|z_i|+\epsilon}{(|z_{\epsilon,i}|+\epsilon)^{1-q}}\Big]^q \Big[\sum_{i=1}^N (|z_{\epsilon,i}|+\epsilon)^q\Big]^{1-q}$$

yield the lower bound

$$\sum_{i=1}^N \frac{|z_i|+\epsilon}{(|z_{\epsilon,i}|+\epsilon)^{1-q}} \ge \sum_{i=1}^N (|z_{\epsilon,i}|+\epsilon)^q.$$

This means that we can chose $\mathbf{z}_1 = \mathbf{z}_\epsilon$ as a minimizer in (18) when $n = 0$. Since we have also proved that $\mathbf{z}_1 = \mathbf{x}$, we conclude that $\mathbf{z}_\epsilon = \mathbf{x}$, as desired. ■

We conclude this section with an auxiliary discussion on convergence issues.

Remark. In our rough explanation of the convergence of the sequence $(\mathbf{z}_n)$ towards the original s-sparse vector $\mathbf{x}$, we insisted on certain appropriate conditions on the initial vector $\mathbf{z}_0$. We highlight here that the convergence towards $\mathbf{x}$ cannot be achieved without such conditions. Indeed, let us consider the 1-sparse vector $\mathbf{x}$ and the 3×4 matrix A defined by

$$\mathbf{x} := [1,\, 0,\, 0,\, 0]^\top, \qquad A := \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

Note that the null-space of A is spanned by $[1,\, 1,\, 1,\, 1]^\top$. Thus, any vector $\mathbf{z}$ satisfying $A\mathbf{z} = A\mathbf{x}$ is of the form

$$\mathbf{z} = [1 + t,\, t,\, t,\, t]^\top, \qquad t \in \mathbb{R}.$$

In this case, the minimization (P_q) reads

$$\underset{t \in \mathbb{R}}{\text{minimize}} \;\; |1 + t|^q + 3|t|^q.$$

It is easy to check that, for $0 < q \leq 1$, the minimum is achieved for $t = 0$, i.e. for $\mathbf{z} = \mathbf{x}$, so that the vector $\mathbf{x}$ is recovered by ℓ_q-minimization — for $q > 0$ small enough, this was guaranteed by the fact that $\delta_3 < 1$. However, if we start the iterative scheme at $\mathbf{z}_0 = [0,\, -1,\, -1,\, -1]^\top$, the minimization (18) reads

$$\underset{t \in \mathbb{R}}{\text{minimize}} \;\; \frac{|1 + t|}{\epsilon^{1-q}} + 3\frac{|t|}{(1 + \epsilon)^{1-q}}.$$

It is easy to check that, if $\epsilon > 1/(3^{1/(1-q)} - 1)$, the minimum is achieved for $t = 0$, so that $\mathbf{z}_1 = \mathbf{x}$. But if $\epsilon < 1/(3^{1/(1-q)} - 1)$, the minimum is achieved for $t = -1$, so that $\mathbf{z}_1 = \mathbf{z}_0$. In this case, we obtain $\mathbf{z}_n = \mathbf{z}_0$ for all n, by immediate induction. Therefore, the sequence $(\mathbf{z}_n)$ does not converge to $\mathbf{x}$ independently of the choice of $\mathbf{z}_0$. It should also be noticed that, even though the vector $[0,\, -1,\, -1,\, -1]^\top$ is the limit of a sequence $(\mathbf{z}_n)$, it is not a local minimizer of $(\mathcal{P}_{q,\epsilon})$ when q is close to 1 and ϵ close to 0.

5 Numerical experiments

We compare in this section the algorithm described in Section 4 with four other existing algorithms, namely the orthogonal greedy algorithm (OGA, see [13]), the regularized orthogonal matching pursuit (ROMP, see [12]), the ℓ_1-minimization (L1), and the reweighted ℓ_1-minimization (RWL1, see [6]).

There are many greedy algorithms available in the literature, see e.g. [14], [15], [16], and [9], but that we find the orthogonal greedy algorithm of [13] more efficient due to two

of its features: one is to select multiple columns from A during each greedy iteration and the other is to use an iterative computational algorithm to find the best approximation in each greedy computation. We thank Alex Petukhov for providing us with his MATLAB code. The MATLAB codes for the regularized orthogonal matching pursuit and for the ℓ_1-minimization can be found online. As for the code associated to the ℓ_q-method of this paper, it is available on the authors' web pages.

We point out that the reweighted ℓ_1-minimization discussed in [6], which came to our attention while we were testing this scheme, is the special instance of the algorithm (18) with

$$q = 0, \qquad \epsilon_n = \epsilon, \qquad \mathbf{z}_0 = \text{minimizer of } (\mathrm{P}_1).$$

Thus, as the approximation of the original problem (P_0), one would intuitively expect that, among the approximations of the problems (P_q), the reweighted ℓ_1-minimization is the best option to recover sparse vectors. This is not the case, though, and there appears to be some advantages in letting the parameter q vary, as demonstrated by the numerical experiments below.

In our first experiment, we justify the values attributed by default to the number of iterations n, the exponent q, and the sequence (ϵ_k) in our ℓ_q-algorithm. The choice is based on the computations summarized in Figures 1, 2, and 3. Here, we have selected $N = 512$ and $m = 128$. For each sparsity level s between 40 and 64, we have picked 150 pairs of s-sparse vector $\mathbf{x}$ and matrix A at random. This means that the support of the vector $\mathbf{x}$ is chosen as the first s values of a random permutation of $\{1, \ldots, N\}$, and that the entries of $\mathbf{x}$ on its support, as well as the entries of A, are independent identically distributed Gaussian random variables with zero mean and unit variance. Then, for each pair $(\mathbf{x}, A)$, using only the partial information $\mathbf{y} = A\mathbf{x}$, we have tried to recover $\mathbf{x}$ as the vector $\mathbf{z}_n$ produced by the iterative scheme (18) started at the minimizer $\mathbf{z}_0$ of (P_1). This was done for several values of the parameters n, q, and (ϵ_k). The recovery was considered a success if $\|\mathbf{x} - \mathbf{z}_n\|_2 < 10^{-3}$.

In Figure 1, the frequency of success appears stationary after a relatively small number of iterations. Figures 2 and 3 suggests that a sequence (ϵ_k) decreasing too fast does not improve the efficiency of the computations. Let us mention that the constant sequence defined by $\epsilon_k = 1/2$ will not be used, as the value of the constant should depend on the expected amplitude of the vector $\mathbf{x}$. Figure 2 also shows that the choice $q = 0$ is not unequivocally the best choice for a single q. Based on these considerations, our preferred values for the parameters are

- number of iteration: $n = 10$,
- ϵ-sequence: $\epsilon_k = \dfrac{1}{k+2}$,
- exponents: $q \in \{0, 0.05, 0.1, 0.2\}$.

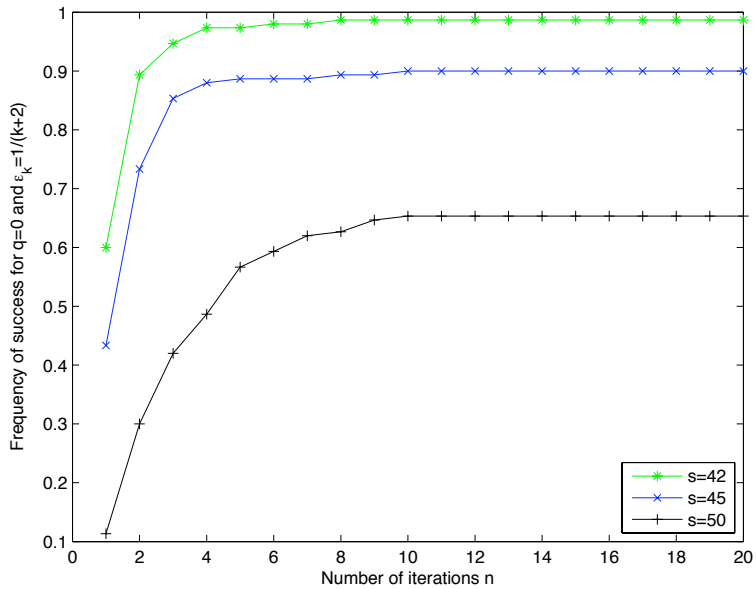

Figure 1: Frequency of success vs. number of iterations n

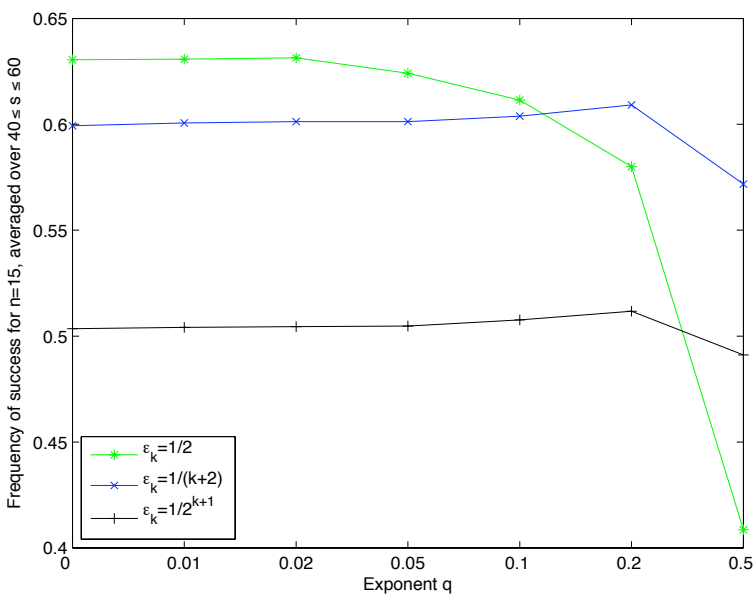

Figure 2: Frequency of success vs. exponent q

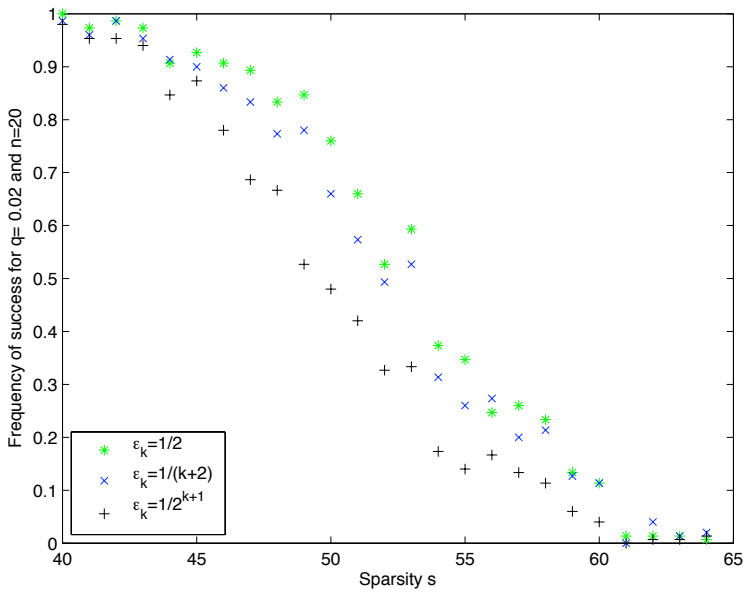

Figure 3: Frequency of success vs. sparsity s

Let us point out that our ℓ_q-algorithm allows several choices for q, including $q = 0$, and that the sparsest output produced from these choices is eventually retained. In this way, it is no surprise that the ℓ_q-method performs at least as well as the reweighted ℓ_1-minimization. It is surprising, however, that it does perform better, even by a small margin. This improvement is obtained at a default cost of 4 times a 10-iteration reweighted ℓ_1-minimization, i.e. at a cost of 40 times an ℓ_1-minimization.

In our last experiment, we present an extensive comparison of the algorithms previously mentioned. We used 100 random pairs $(\mathbf{x}, A)$ for this test, with $N = 512$ and $m = 128$. For each pair, we run each of the five algorithms to obtain a vector $\widetilde{\mathbf{x}}$, and we have strengthened the condition for successful recovery to $\|\mathbf{x} - \widetilde{\mathbf{x}}\|_\infty < 10^{-5}$. In the reweighted ℓ_1-minimization, we have taken $n = 20$ and $\epsilon_k = 1/10$, while in the ℓ_q-algorithm, we have taken $n = 20$, $\epsilon_k = 1/2^k$, and $q = 0$, $q = 0.1$, $q = 0.2$, $\ldots$, $q = 0.9$ successively. In fact, not all these values of q were necessarily used. The knowledge of the original vector $\mathbf{x}$ was indeed exploited in this particular experiment to stop the program at the first occurrence of a successful recovery. All the same, even with the non-optimal parameters we chose, Figure 4 reveals that an ℓ_q-method with varying q gives somewhat better success rates than the other four algorithms considered.

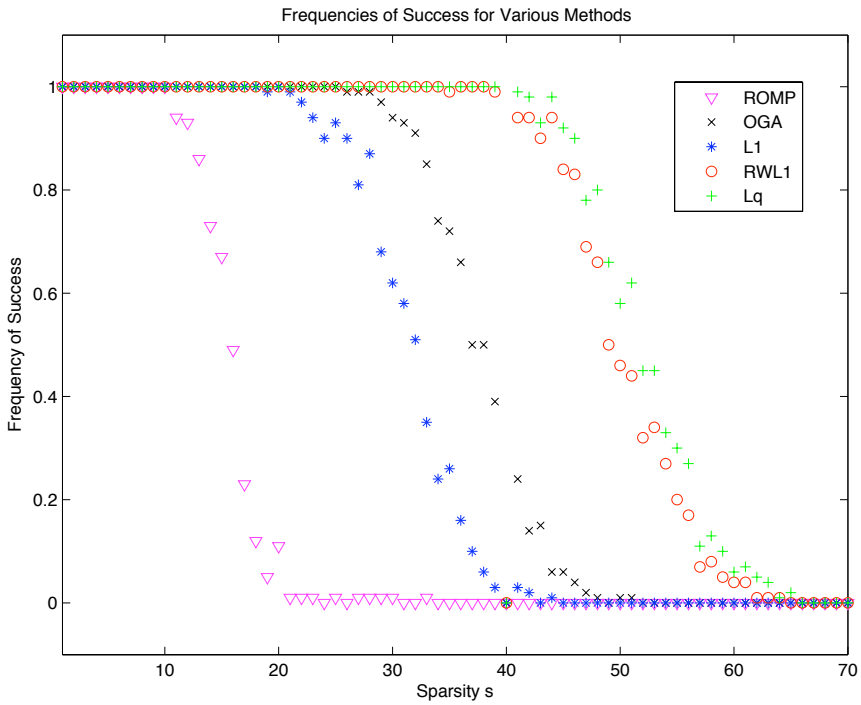

Figure 4: Comparison of the five algorithms for sparse vectors with arbitrary entries

References

[1] Baraniuk, R., M. Davenport, R. DeVore, and M. Wakin, A simple proof of the restricted isometry property for random matrices, Constructive Approximation, 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 (2008), 589–592.

[3] Candès, E. J., J. K. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. 59 (2006), 1207–1223.

[4] Candès, E. J. and T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (2005), no. 12, 4203–4215.

[5] Candès, E. J. and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies, IEEE Trans. Inform. Theory 52 (2006), no. 12, 5406–5425.

[6] Candès, E. J., M. Watkin, and S. Boyd, Enhancing Sparsity by Reweighted l_1 Minimization, To appear in J. Fourier Anal. Appl.

[7] Chartrand, R., Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett., 14 (2007), 707–710.

[8] Donoho, D. L. and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization, Proc. Natl. Acad. Sci. USA 100 (2003), no. 5, 2197–2202.

[9] Donoho, D. L., M. Elad, and V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inform. Theory, 52 (2006), 6–18.

[10] Gribonval, R. and M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure. Appl. Comp. Harm. Anal. 22 (2007), 335–355.

[11] Natarajan, B. K., Sparse approximate solutions to linear systems, SIAM J. Comput., vol. 24, pp. 227–234, 1995.

[12] Needell, D. and R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, To appear in Foundations of Comp. Math.

[13] Petukhov, A., Fast implementation of orthogonal greedy algorithm for tight wavelet frames, Signal Processing, 86 (2006), 471–479.

[14] Temlyakov, V. N., Weak greedy algorithms, Adv. Comput. Math. 12 (2000), 213–227.

[15] Temlyakov, V. N., Nonlinear methods of approximation, Foundations of Comp. Math., 3 (2003), 33–107.

[16] Tropp, J. A., Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory, 50 (2004), 2231–2242.