HAL
open science

Accuracy guarantees for L1-recovery

Anatoli B. Juditsky, Arkadii S. Nemirovski

▶ To cite this version:

Anatoli B. Juditsky, Arkadii S. Nemirovski. Accuracy guarantees for L1-recovery. IEEE Transactions on Information Theory, 2011, 57 (12), pp.7818-7839. 10.1109/TIT.2011.2162569 . hal-00510689v3

HAL Id: hal-00510689
https://hal.science/hal-00510689v3

Submitted on 25 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accuracy guarantees for ℓ_1-recovery *

Anatoli Juditsky
LJK, Université J. Fourier, B.P. 53, 38041 Grenoble Cedex 9, France
Anatoli.Juditsky@imag.fr

Arkadi Nemirovski
Georgia Institute of Technology, Atlanta, Georgia 30332, USA
nemirovs@isye.gatech.edu

May 25, 2011

Abstract

We discuss two new methods of recovery of sparse signals from noisy observation based on ℓ_1-minimization. While they are closely related to the well-known techniques such as Lasso and Dantzig Selector, these estimators come with *efficiently verifiable guaranties of performance*. By optimizing these bounds with respect to the method parameters we are able to construct the estimators which possess better statistical properties than the commonly used ones.

We link our performance estimations to the well known results of Compressive Sensing and justify our proposed approach with an oracle inequality which links the properties of the recovery algorithms and the best estimation performance when the signal support is known. We also show how the estimates can be computed using the Non-Euclidean Basis Pursuit algorithm.

Key words : sparse recovery, linear estimation, oracle inequalities, nonparametric estimation by convex optimization

AMS Subject Classification : 62G08, 90C25

1 Introduction

Recently several methods of estimation and selection which refer to the ℓ_1-minimization received much attention in the statistical literature. For instance, *Lasso estimator*, which is the ℓ_1-penalized least-squares method is probably the most studied (a theoretical analysis of the Lasso estimator is provided in, e.g., [2, 3, 4, 19, 20, 21, 17, 18], see also the references cited therein). Another, closely related to the Lasso, statistical estimator is the *Dantzig Selector* [7, 2, 16, 17]. To be more precise, let us consider the estimation problem as follows. Assume that an observation

$$y = Ax + \sigma\xi \in \mathbf{R}^m \tag{1}$$

is available, where $x \in \mathbf{R}^n$ is an unknown signal and $A \in \mathbf{R}^{m\times n}$ is a known *sensing matrix*. We suppose that $\sigma\xi$ is a Gaussian disturbance with $\xi \sim N(0, I_m)$ (i.e., $\xi = (\xi_1, ..., \xi_n)^T$, where ξ_i are independent normal r.v.'s with zero mean and unit variance), and $\sigma > 0$ is a known deterministic noise level. Our focus is on the recovery of unknown signal x.

*Research of the second author was supported by the Office of Naval Research grant # N000140811104 and the NSF grant DMS-0914785.

The Dantzig Selector estimator $\widehat{x}_{\rm DS}$ of the signal x is defined as follows [7]:

$$\widehat{x}_{\rm DS}(y) \in \underset{v\in\mathbf{R}^n}{\rm Argmin}\{\|v\|_1 \mid \|A^T(Av-y)\|_\infty \le \rho\}$$

where $\rho = O\left(\sigma\sqrt{\ln n}\right)$ is the algorithm's parameter. Since $\widehat{x}_{\rm DS}$ is obtained as a solution of an linear program, it is very attractive by its low computational cost. Accuracy bounds for this estimator are readily available. For instance, a well known result about this estimator (cf. [7, Theorem 1.1]) is that if $\rho = O\left(\sigma\sqrt{\ln(n/\epsilon)}\right)$ then

$$\|\widehat{x}_{\rm DS}(y) - x\|_2 \le K\sigma\sqrt{s\log(n\epsilon^{-1})}$$

with probability $1-\epsilon$ if a) the signal x is s-sparse, i.e. has at most s non-vanishing components, and b) the sensing matrix A with unit columns possesses the *Restricted Isometry Property* RIP(δ,k) with parameters $0<\delta<\frac{1}{1+\sqrt{2}}$ and $k\ge 3s$. [1] Further, in this case one has $K = C(1-\delta)^{-1}$, where C is a moderate absolute constant. This result is quite impressive, in part due to the fact (see, e.g. [5, 6]) that there exist $m\times n$ random matrices, with $m<n$, which possess the RIP with probability close to 1, δ close to zero and the value of k as large as $O\left(m\ln^{-1}(n/m)\right)$. Similar performance guarantees are known for *Lasso recovery*

$$\widehat{x}_{\rm lasso}(y) \in \underset{v\in\mathbf{R}^n}{\rm Argmin}\left\{\|v\|_1 + \varkappa\|Av-y\|_2^2\right\},$$

with properly chosen penalty parameter $\varkappa$. The available accuracy bounds for Lasso and Dantzig Selector rely upon the Restricted Isometry Property or less restrictive assumptions about the sensing matrix, such as Restricted Eigenvalue [2] or Compatibility [3] conditions (a complete overview of those and several other assumptions with description of how they relate to each other is provided in [19]). However, these assumptions cannot be verified efficiently. The latter implies that there is currently no way to provide any guaranties (e.g., confidence sets) of the performance of the proposed procedures. A notable exception from this rule is the *Mutual Incoherence* assumption (see, e.g. [10, 11, 12] and [21] for the case of, respectively, deterministic and random observation noise) which can be used to compute the accuracy bounds for recovery algorithms: a matrix A with columns of unit ℓ_2-norm and *mutual incoherence* $\mu(A)$ possesses RIP(δ,k) with $\delta = (m-1)\mu(A)$.[2] Unfortunately, the latter relation implies that $\mu(A)$ should be very small to certify the possibility of accurate ℓ_1-recovery of non-trivial sparse signals, so that performance guarantees based on mutual incoherence are very conservative. This "theoretical observation" is supported by numerical experiments – the practical guarantees which may be obtained using the mutual incoherence are generally quite poor even for the problems with nice theoretical properties (cf. [14, 15]).

Recently the authors have proposed a new approach for efficient computing of upper and lower bounds on the "level of goodness" of a sensing matrix A, i.e. the maximal s such that the ℓ_1-recovery of *all* signals with no more than s non-vanishing components is accurate in the case where the measurement noise vanishes (see [14]). In the present paper we aim to use the related *verifiable sufficient conditions* of "goodness" of a sensing matrix A to provide efficiently computable bounds for the error of ℓ_1 recovery procedures in the case when the observations are affected by random noise.

The main body of the paper is organized as follows:

[1] Recall that RIP(δ,k), called also *uniform uncertainty principle*, means that for any $v\in\mathbf{R}^n$ with at most k nonzero entries,

$$(1-\delta)\|v\|_2 \le \|Av\|_2^2 \le (1+\delta)\|v\|_2$$

This property essentially requires that every set of columns of A with cardinality less than k approximately behaves like an orthonormal system.

[2] The *mutual incoherence* $\mu(A)$ of a sensing matrix $A=[A_1,...,A_n]$ is computed according to

$$\mu(A) = \max_{i\ne j}\frac{|A_i^TA_j|}{A_i^TA_i}.$$

Obviously, the mutual incoherence can be easily computed even for large matrices.

1. We start with Section 2.1 where we formulate the sparse recovery problem and introduce our core assumption – a verifiable condition $\mathbf{H}_{s,\infty}(\kappa)$ linking matrix $A \in \mathbf{R}^{m\times n}$ and a *contrast matrix* $H \in \mathbf{R}^{m\times n}$. In Sections 2.2, 2.3 we present two recovery routines with contrast matrices:

   - *regular recovery*:
   $$\widehat{x}_{\rm reg}(y) \in \mathop{\rm Argmin}_{v\in\mathbf{R}^n}\{\|v\|_1 : \|H^T(Av-y)\|_\infty \le \rho\},$$
   - *penalized recovery*:
   $$\widehat{x}_{\rm pen}(y) \in \mathop{\rm Argmin}_{v\in\mathbf{R}^n}\{\|v\|_1 + \theta s\|H^T(Av-y)\|_\infty\},$$
   (s is our guess for the number of nonzero entries in the true signal, $\theta > 0$ is the penalty parameter)

   along with their performance guarantees under condition $\mathbf{H}_{s,\infty}(\kappa)$ with $\kappa < 1/2$, that is, explicit upper bounds on the confidence levels of the recovery errors $\|\widehat{x}-x\|_p$. The novelty here is that our bounds are of the form
   $$\text{Prob}\left\{\|\widehat{x}-x\|_p \le O\left(s^{1/p}\sigma\sqrt{\ln(n/\epsilon)}\right) \text{ for every } s\text{-sparse signal } x \text{ and all } 1\le p\le\infty\right\} \ge 1-\epsilon \quad (2)$$
   (with hidden factors in $O(\cdot)$ independent of ϵ, σ), and *are valid in the entire range* $1 \le p \le \infty$ *of values of* p. Note that similar error bounds for Dantzig Selector and Lasso are only known for $1 \le p \le 2$, whatever be the assumptions on "essentially nonsquare" matrix A.

2. Our interest in condition $\mathbf{H}_{s,\infty}(\kappa)$ stems from the fact that this condition, in contrast to the majority of the known sufficient conditions for the validity of ℓ_1-based sparse recovery (e.g., Restricted Isometry/Eigenvalue/Compatibility), is efficiently verifiable. Moreover, it turns out that one can efficiently optimize the error bounds of the associated with this verifiable condition regular/penalized recovery routines over the contrast matrix H. The related issues are considered in Section 3. In Section 4 we provide some additional justification of the condition $\mathbf{H}$, in particular, by linking it with the Mutual Incoherence and Restricted Isometry properties. This, in particular, implies that the condition $\mathbf{H}_{s,\infty}(\kappa)$ with, say, $\kappa = \frac{1}{3}$ associated with randomly selected $m \times n$ matrices A is feasible, with probability approaching 1 as m, n grow, for s as large as $O(\sqrt{m/\ln(n)})$. We also establish *limits of performance* of the condition, specifically, show that unless A is nearly square, $\mathbf{H}_{s,\infty}(\kappa)$ with $\kappa < 1/2$ can be feasible only when $s \le O(1)\sqrt{m}$, meaning that the tractability of the condition has a heavy price: when designing and validating ℓ_1 minimization based sparse recovery routines, this condition can be useful only in a severely restricted range of the sparsity parameter s.

3. In Section 5 we show that the condition $\mathbf{H}_{s,\infty}(\kappa)$ is the strongest (and seemingly the only verifiable one) in a natural family of conditions $\mathbf{H}_{s,q}(\kappa)$ linking a sensing and a contrast matrix; here s is the number of nonzeros in the sparse signal to be recovered $q \in [1,\infty]$. We demonstrate that when a contrast matrix H satisfies $\mathbf{H}_{s,q}(\kappa)$ with $\kappa < 1/2$, the associated regular and penalized ℓ_1 recoveries admit error bounds similar to (2), but now in the restricted range $1 \le p \le q$ of values of p. We demonstrate also that feasibility of $\mathbf{H}_{s,q}(\kappa)$ with $\kappa < 1/2$ implies instructive (although slightly worse than those in (2)) error bounds for the Dantzig Selector and Lasso recovering routines.

4. In Section 6, we present numerical results on comparison of regular/penalized ℓ_1 recovery with the Dantzig Selector and Lasso algorithms. The conclusion suggested by these preliminary numerical results is that *when the former procedures are applicable* (i.e., when the techniques of Section 3 allow to build a "not too large" contrast matrix satisfying the condition $\mathbf{H}_{s,\infty}(\kappa)$ with, say, $\kappa = 1/3$), *our procedures outperform significantly the Dantzig Selector and work exactly as well as the Lasso algorithm with "ideal" (unrealistic in actual applications) choice of the regularization parameter*[3].

[3] With "theoretically optimal," rather than "ideal," choice of the regularization parameter in Lasso, this algorithm is essentially worse than our algorithms utilizing the contrast matrix.

5. In the concluding Section 7 we present a "Non-Euclidean Matching Pursuit algorithm" (similar to the one presented in [15]) with the same performance characteristics as those of regular/penalized ℓ_1 recoveries; this algorithm, however, does not require optimization and can be considered as a computationally cheap alternative to ℓ_1 recoveries, especially in the case when one needs to process a series of recovery problems with common sensing matrix.

All proofs are placed in the Appendix.

2 Accuracy bounds for ℓ_1-Recovery Routines

2.1 Problem statement

Notation. For a vector $x \in \mathbf{R}^n$ and $1 \leq s \leq n$ we denote x^s the vector obtained from x by setting to 0 all but the s largest in magnitude entries of x. Ties, if any, could be resolved arbitrarily; for the sake of definiteness assume that among entries of equal magnitudes, those with smaller indexes have priority (e.g., with $x = [1; 2; 2; 3]$ one has $x^2 = [0; 2; 0; 3]$). $\|x\|_{s,p}$ stands for the usual ℓ_p-norm of x^s (so that $\|x\|_{s,\infty} = \|x\|_\infty$). We say that a vector z is *s-sparse* if it has at most s nonzero entries. Finally, for a set $I \subset \{1, ..., n\}$ we denote by J its complement $\{1, ..., n\} \backslash I$; given $x \in \mathbf{R}^n$, we denote by x_I the vector obtained from x by zeroing the entries with indices outside of I, so that $x = x_I + x_J$.

Given a norm $\nu(\cdot)$ on $\mathbf{R}^m$ and a matrix $H = [h_1, ..., h_N] \in \mathbf{R}^{m \times N}$, we set $\nu(H) = \max\limits_{i \leq N} \nu(h_i)$.

The problem. We consider an observation $y \in \mathbf{R}^m$

$$y = Ax + u + \sigma\xi, \tag{3}$$

where $x \in \mathbf{R}^n$ is an unknown signal and $A \in \mathbf{R}^{m \times n}$ is *the sensing matrix*. We suppose that $\sigma\xi$ is a Gaussian disturbance, where $\xi \sim N(0, I_m)$ (i.e., $\xi = (\xi_1, ..., \xi_n)^T$ with independent normal random variables ξ_i with zero mean and unit variance), $\sigma > 0$ being known, and u is a nuisance parameter known to belong to a given *uncertainty set* $\mathcal{U} \subset \mathbf{R}^m$ which we will suppose to be convex, compact and symmetric w.r.t. the origin. Our goal is to recover x from y, provided that x is "nearly s-sparse." Specifically, we consider the sets

$$X(s, \upsilon) = \{x \in \mathbf{R}^n : \|x - x^s\|_1 \leq \upsilon\}$$

of signals which admit s-sparse approximation of $\|\cdot\|_1$-accuracy υ. Given p, $1 \leq p \leq \infty$, and a confidence level $1 - \epsilon$, $\epsilon \in (0, 1)$, we quantify a recovery routine — a Borel function $\mathbf{R}^m \ni y \mapsto \widehat{x}(y) \in \mathbf{R}^n$ — by its worst-case, over $x \in X(s, \upsilon)$, confidence interval, taken w.r.t. $\|\cdot\|_p$-norm of the error. Specifically, we define the *risks* of a recovery routine as

$$\mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon, \sigma, s, \upsilon) = \inf \left\{\delta : \mathrm{Prob}\{\xi : \exists x \in X(s, \upsilon), u \in \mathcal{U} : \|\widehat{x}(Ax + \sigma\xi + u) - x\|_p > \delta\} \leq \epsilon\right\}.$$

Equivalently: $\mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon, \sigma, s, \upsilon) \leq \delta$ if and only if there exists a set Ξ of "good" realizations of ξ with $\mathrm{Prob}\{\xi \in \Xi\} \geq 1 - \epsilon$ such that whenever $\xi \in \Xi$, one has $\|\widehat{x}(Ax + \sigma\xi + u) - x\|_p \leq \delta$ *for all* $x \in X(s, \upsilon)$ *and all* $u \in \mathcal{U}$.

Norm $\nu(\cdot)$. Given ϵ and $\sigma > 0$ let us denote

$$\nu(v) = \nu_{\epsilon,\sigma,\mathcal{U}}(v) = \sup_{u \in \mathcal{U}} u^T v + \sigma\sqrt{2\ln(n/\epsilon)}\|v\|_2. \tag{4}$$

Since $\mathcal{U}$ is convex, closed and symmetric with respect to the origin, $\nu(\cdot)$ is a norm. Let ν_* be the norm on $\mathbf{R}^n$ conjugate to ν:

$$\nu_*(u) = \max_v \{v^T u : \nu(v) \leq 1\}.$$

Conditions $\mathbf{H}(\gamma)$ and $\mathbf{H}_{s,\infty}(\kappa)$. Let $\gamma = (\gamma_1, ..., \gamma_n) \in \mathbf{R}^n_+$. Given $A \in \mathbf{R}^{m\times n}$, consider the following condition on a matrix $H = [h_1, ..., h_n] \in \mathbf{R}^{m\times n}$:

$\mathbf{H}(\gamma)$: *for all $x \in \mathbf{R}^n$ and $1 \le i \le n$ one has*

$$|x_i| \le |h_i^T Ax| + \gamma_i \|x\|_1. \tag{5}$$

Now let s be a positive integer and $\kappa > 0$. Given $A \in \mathbf{R}^{m\times n}$, we say that a matrix $H = [h_1, ..., h_n] \in \mathbf{R}^{m\times n}$ satisfies condition $\mathbf{H}_{s.\infty}(\kappa)$ [4], if

$$\forall x \in \mathbf{R}^n : \|x\|_\infty \le \|H^T Ax\|_\infty + s^{-1}\kappa\|x\|_1. \tag{6}$$

The conditions we have introduced are closely related to each other:

Lemma 1 *If H satisfies $\mathbf{H}(\gamma)$, then H satisfies $H_{s,\infty}(s\|\gamma\|_\infty)$, and "nearly vice versa:" given $H \in \mathbf{R}^{m\times n}$ satisfying $\mathbf{H}_{s,\infty}(\kappa)$, one can build efficiently a matrix $H' \in \mathbf{R}^{m\times n}$ satisfying $\mathbf{H}(\gamma)$ with $\gamma = \frac{\kappa}{s}[1; ...; 1]$ (i.e., $\kappa = s\|\gamma\|_\infty$) and such that the columns of H' are convex combinations of the columns of H and $-H$, so that $\nu(H') \le \nu(H)$ for every norm $\nu(\cdot)$ on $\mathbf{R}^m$.*

2.2 Regular ℓ_1 Recovery

In this section we discuss the properties of the *regular ℓ_1-recovery* $\widehat{x}_{\rm reg}$ given by:

$$\widehat{x}_{\rm reg} = \widehat{x}_{\rm reg}(y) \in \operatorname*{Argmin}_{v\in\mathbf{R}^n}\{\|v\|_1 : |h_i^T(Av - y)| \le \rho_i, \ \ i = 1, ..., n\}, \tag{7}$$

where y is as in (3), h_i, $i = 1, ..., n$, are some vectors in $\mathbf{R}^m$ and $\rho_i > 0$, $i = 1, ..., n$. We refer to the matrix $H = [h_1, ..., h_n]$ as to the *contrast matrix* underlying the recovering procedure.

The starting point of our developments is the following

Proposition 1 *Given an $m \times n$ sensing matrix A, noise intensity σ, uncertainty set $\mathcal{U}$ and a tolerance $\epsilon \in (0, 1)$, let the matrix $H = [h_1, ..., h_n]$ from (7) satisfy the condition $\mathbf{H}(\gamma)$ for some $\gamma \in \mathbf{R}^n_+$, and let ρ_i in (7) satisfy the relation*

$$\rho_i \ge \nu_i := \nu(h_i), \ i = 1, ..., n \tag{8}$$

where $\nu(\cdot)$ is given by (4). Then there exists a set $\Xi \subset \mathbf{R}^m$, $\mathrm{Prob}\{\xi \in \Xi\} \ge 1 - \epsilon$, of "good" realizations of ξ such that

(i) *Whenever $\xi \in \Xi$, for every $x \in \mathbf{R}^n$, every $u \in \mathcal{U}$ and every subset $I \subset \{1, ..., n\}$ such that*

$$\gamma_I := \sum_{i\in I} \gamma_i < \tfrac{1}{2}, \tag{9}$$

the regular ℓ_1-recovery $\widehat{x}_{\rm reg}$ given by (7) satisfies:

$$\begin{array}{lll} (a) & \|\widehat{x}_{\rm reg}(Ax + \sigma\xi + u) - x\|_1 & \le \dfrac{2\|x_J\|_1 + 2\rho_I + 2\nu_I}{1 - 2\gamma_I}; \\ (b) & |\,[\widehat{x}_{\rm reg}(Ax + \sigma\xi + u) - x]_i\,| & \le \rho_i + \nu_i + \gamma_i\|\widehat{x}_{\rm reg}(y) - x\|_1 \\ & & \le \rho_i + \nu_i + \gamma_i \dfrac{2\|x_J\|_1 + 2\rho_I + 2\nu_I}{1 - 2\gamma_I}, \ i = 1, ..., n, \end{array} \tag{10}$$

where $\rho_I = \sum_{i\in I} \rho_i$ and $\nu_I = \sum_{i\in I} \nu_i$.

(ii) *In particular, when setting*

$$\begin{array}{l} \widehat{\rho}_s = \|[\rho_1; ...; \rho_n]\|_{s,1}, \ \ \widehat{\nu}_s = \|[\nu(h_1); ...; \nu(h_n)]\|_{s,1}, \ \ \widehat{\gamma}_s = \|[\gamma_1; ...; \gamma_n]\|_{s,1}, \\ \widehat{\rho} = \widehat{\rho}_1 = \max_i \rho_i, \ \ \nu(H) = \widehat{\nu}_1 = \max_i \nu(h_i), \ \ \widehat{\gamma} = \widehat{\gamma}_1 = \max_i \gamma_i, \end{array} \tag{11}$$

[4] The reason for this cumbersome, at the first glance, notation will become clear later, in Section 5.

and assuming $\widehat{\gamma}_s < \frac{1}{2}$, *for every* $x \in \mathbf{R}^n$, $\xi \in \Xi$ *and* $u \in \mathcal{U}$ *it holds*

$$\begin{array}{rcl} \|\widehat{x}_{\rm reg}(Ax+\sigma\xi+u)-x\|_1 & \leq & 2\dfrac{\|x-x^s\|_1+\widehat{\rho}_s+\widehat{\nu}_s}{1-2\widehat{\gamma}_s} \leq 2\dfrac{\|x-x^s\|_1}{1-2\widehat{\gamma}_s}+2s\dfrac{\widehat{\rho}+\nu(H)}{1-2\widehat{\gamma}_s}; \\ \|\widehat{x}_{\rm reg}(Ax+\sigma\xi+u)-x\|_\infty & \leq & 2\widehat{\gamma}\dfrac{\|x-x^s\|_1}{1-2\widehat{\gamma}_s}+\dfrac{[1+2s\widehat{\gamma}-2\widehat{\gamma}_s][\widehat{\rho}+\nu(H)]}{1-2\widehat{\gamma}_s} \end{array}$$

(iii) Finally, assuming $s\widehat{\gamma} < 1/2$, *for every* $\xi \in \Xi$, $x \in \mathbf{R}^n$ *and* $u \in \mathcal{U}$ *one has*

$$\begin{array}{rcl} \|\widehat{x}_{\rm reg}(Ax+\sigma\xi+u)-x\|_1 & \leq & 2\frac{\|x-x^s\|_1}{1-2s\widehat{\gamma}}+2s\frac{\widehat{\rho}+\nu(H)}{1-2s\widehat{\gamma}}; \\ \|\widehat{x}_{\rm reg}(Ax+\sigma\xi+u)-x\|_\infty & \leq & s^{-1}\frac{\|x-x^s\|_1}{1-2s\widehat{\gamma}}+\frac{\widehat{\rho}+\nu(H)}{1-2s\widehat{\gamma}}. \end{array} \tag{12}$$

The following result is an immediate corollary of Proposition 1:

Lemma 2 *Under the premise of Proposition 1, assume that* $\widehat{\gamma}_s < \frac{1}{2}$. *Then for all* $1 \leq p \leq \infty$ *and* $\upsilon \geq 0$*:*

$$\mathrm{Risk}_p(\widehat{x}_{\rm reg}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq \frac{2}{1-2\widehat{\gamma}_s}\left[\upsilon+\widehat{\rho}_s+\widehat{\nu}_s\right]^{\frac{1}{p}}\left[\widehat{\gamma}\upsilon+[\tfrac{1}{2}-\widehat{\gamma}_s][\widehat{\rho}+\nu(H)]+\widehat{\gamma}[\widehat{\nu}_s+\widehat{\rho}_s]\right]^{\frac{p-1}{p}} \tag{13}$$

(for notation, see (11)). Further, if $s\widehat{\gamma} < 1/2$, *we have also*

$$1 \leq p \leq \infty \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\rm reg}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq \frac{(2s)^{\frac{1}{p}}}{1-2s\widehat{\gamma}}(s^{-1}\upsilon+\widehat{\rho}+\nu(H)). \tag{14}$$

The next statement is similar to the cases of $\kappa := s\widehat{\gamma} < 1/2$ in Proposition 1 and Lemma 2; the difference is that now we assume that H satisfies $\mathbf{H}_{s,\infty}(\kappa)$, which, by Lemma 1, is a weaker requirement on H than to satisfy $\mathbf{H}(\gamma)$ with $s\widehat{\gamma} \equiv s\|\gamma\|_\infty = \kappa$.

Proposition 2 *Given an* $m \times n$ *sensing matrix* A, *noise intensity* σ, *uncertainty set* $\mathcal{U}$ *and a tolerance* $\epsilon \in (0,1)$, *let the matrix* $H = [h_1,...,h_n]$ *from (7) satisfy the condition* $\mathbf{H}_{s,\infty}(\kappa)$ *for some* $\kappa < 1/2$, *and let* ρ_i *in (7) satisfy the relation (8). Then there exists a set* $\Xi \subset \mathbf{R}^m$, $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$, *of "good" realizations of* ξ *such that whenever* $\xi \in \Xi$, *for every* $x \in \mathbf{R}^n$ *and every* $u \in \mathcal{U}$ *one has*

$$\begin{array}{rcl} \|\widehat{x}_{\rm reg}(Ax+\sigma\xi+u)-x\|_1 & \leq & 2\frac{\|x-x^s\|_1}{1-2\kappa}+2s\frac{\widehat{\rho}+\nu(H)}{1-2\kappa}; \\ \|\widehat{x}_{\rm reg}(Ax+\sigma\xi+u)-x\|_\infty & \leq & s^{-1}\frac{\|x-x^s\|_1}{1-2\kappa}+\frac{\widehat{\rho}+\nu(H)}{1-2\kappa}. \end{array} \tag{15}$$

In particular,

$$1 \leq p \leq \infty \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\rm reg}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq \frac{(2s)^{\frac{1}{p}}}{1-2\kappa}(s^{-1}\upsilon+\widehat{\rho}+\nu(H)). \tag{16}$$

2.3 Penalized ℓ_1 Recovery

Now consider the *penalized* ℓ_1*-recovery* $\widehat{x}_{\rm pen}$ as follows:

$$\widehat{x}_{\rm pen}(y) \in \mathop{\mathrm{Argmin}}_{v\in\mathbf{R}^n}\{\|v\|_1+\theta s\|H^T(Av-y)\|_\infty\}, \tag{17}$$

where y is as in (3), and an integer $s \leq n$, a positive θ, and a matrix H are parameters of the construction.

Proposition 3 *Given an* $m \times n$ *sensing matrix* A, *an integer* $s \leq n$, *a matrix* $H = [h_1,...,h_n] \in \mathbf{R}^{m\times n}$ *and positive reals* γ_i, $1 \leq i \leq n$, *satisfying the condition* $\mathbf{H}(\gamma)$, *and a* $\theta > 0$, *assume that*

$$\widehat{\gamma}_s := \|\gamma\|_{s,1} < \tfrac{1}{2} \tag{18}$$

and

$$(1-\widehat{\gamma}_s)^{-1} < \theta < (\widehat{\gamma}_s)^{-1} \tag{19}$$

Further, let $\sigma \geq 0$, $\epsilon \in (0,1)$, *and let*

$$\nu_i = \nu_{\epsilon,\sigma,\mathcal{U}}(h_i),\ i=1,...,n,\ \nu(H) = \max_i \nu_i. \tag{20}$$

Consider the penalized recovery $\widehat{x}_{\rm pen}(\cdot)$ *associated with* H, s, θ. *There exists a set* $\Xi \subset \mathbf{R}^m$, $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$, *of "good" realizations of* ξ *such that*

(i) *Whenever* $\xi \in \Xi$, *for every signal* $x \in \mathbf{R}^n$ *and every* $u \in \mathcal{U}$ *one has*

$$\begin{array}{rrcl} (a) & \|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u)-x\|_1 & \leq & \frac{2\|x-x^s\|_1+2s\theta\nu(H)}{\min[\theta(1-\widehat{\gamma}_s)-1,1-\theta\widehat{\gamma}_s]} \\ (b) & \|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u))-x\|_\infty & \leq & \left(\frac{1}{s\theta}+\widehat{\gamma}\right)\|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u)-x\|_1+2\nu(H) \\ & & \leq & 2\left(\frac{1}{s\theta}+\widehat{\gamma}\right)\frac{\|x-x^s\|_1}{\min[\theta(1-\widehat{\gamma}_s)-1,1-\theta\widehat{\gamma}_s]}+2\nu(H)\left[\frac{1+s\theta\widehat{\gamma}}{\min[\theta(1-\widehat{\gamma}_s)-1,1-\theta\widehat{\gamma}_s]}+1\right], \end{array} \tag{21}$$

where, as in Lemma 2, $\widehat{\gamma} = \max_i \gamma_i$.

(ii) *When* $\theta = 2$ *and* $\widehat{\gamma} < \frac{1}{2s}$, *one has for every* $x \in \mathbf{R}^n$, $u \in \mathcal{U}$ *and* $\xi \in \Xi$:

$$\begin{array}{rrcl} (a) & \|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u)-x\|_1 & \leq & 2\frac{\|x-x^s\|_1}{1-2s\widehat{\gamma}}+4s\frac{\nu(H)}{1-2s\widehat{\gamma}} \\ (b) & \|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u))-x\|_\infty & \leq & 2s^{-1}\frac{\|x-x^s\|_1}{1-2s\widehat{\gamma}}+4\frac{\nu(H)}{1-2s\widehat{\gamma}}, \end{array} \tag{22}$$

whence for every $\upsilon \geq 0$ *and* $1 \leq p \leq \infty$:

$$\mathrm{Risk}_p(\widehat{x}_{\rm pen}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq \frac{2s^{\frac{1}{p}}}{1-2s\widehat{\gamma}}(s^{-1}\upsilon+2\nu(H)). \tag{23}$$

The next statement is in the same relation to Proposition 3 as Proposition 2 is to Proposition 1 and Lemma 2.

Proposition 4 *Given an* $m \times n$ *sensing matrix* A, *noise intensity* σ, *uncertainty set* $\mathcal{U}$ *and a tolerance* $\epsilon \in (0,1)$, *let the matrix* $H = [h_1,...,h_n]$ *from (17) satisfy the condition* $\mathbf{H}_{s,\infty}(\kappa)$ *for some* $\kappa < 1/2$, *and let* $\theta = 2$. *Then there exists a set* $\Xi \subset \mathbf{R}^m$, $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$, *of "good" realizations of* ξ *such that whenever* $\xi \in \Xi$, *for every* $x \in \mathbf{R}^n$ *and every* $u \in \mathcal{U}$ *one has*

$$\begin{array}{rcl} \|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u)-x\|_1 & \leq & 2\frac{\|x-x^s\|_1}{1-2\kappa}+4s\frac{\nu(H)}{1-2\kappa}; \\ \|\widehat{x}_{\rm pen}(Ax+\sigma\xi+u)-x\|_\infty & \leq & 2s^{-1}\frac{\|x-x^s\|_1}{1-2\kappa}+4\frac{\nu(H)}{1-2\kappa}. \end{array} \tag{24}$$

In particular,

$$1 \leq p \leq \infty \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\rm pen}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq \frac{2s^{\frac{1}{p}}}{1-2\kappa}(s^{-1}\upsilon+2\nu(H)). \tag{25}$$

Note that under the premise of Proposition 2, the smallest possible values of ρ_i are the quantities ν_i, which results in $\widehat{\rho} = \nu(H)$; with this choice of ρ_i, the risk bound for the regular recovery, as given by the right hand side in (16), coincides within factor 2 with the risk bound for the penalized recovery with $\theta = 2$ as given by (25); both bounds assume that H satisfies $\mathbf{H}_{s,\infty}(\kappa)$ with $\kappa < 1/2$ and imply that

$$1 \leq p \leq \infty \Rightarrow \mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq 2\frac{s^{\frac{1}{p}}}{1-2\kappa}(s^{-1}\upsilon+2\nu(H)). \tag{26}$$

When $\upsilon = 0$, the latter bound admits a quite transparent interpretation: everything is *as if* we were observing the sum of an unknown *s-dimensional* signal and an observation error of the uniform norm $O(1)\nu(H)$.

3 Efficient construction of the contrast matrix H

In what follows, we fix A, the "environment parameters" ϵ, $\sigma, \mathcal{U}$ and the "level of sparsity" s of signals x we intend to recover, and are interested in building the contrast matrix $H = [h_1, ..., h_n]$ resulting in as small as possible error bound (26). All we need to this end is to answer the following question (where we should specify the norm $\varphi(\cdot)$ as $\nu_{\epsilon,\sigma,\mathcal{U}}(\cdot)$):

> (?) *Let $\varphi(\cdot)$ be a norm on $\mathbf{R}^m$, and s be a positive integer. What is the domain G_s of pairs $(\omega, \kappa) \in \mathbf{R}^2_+$ such that $\kappa < 1/2$ and there exists matrix $H = [h_1, ..., h_n] \in \mathbf{R}^{m\times n}$ satisfying the condition $\mathbf{H}_{s,\infty}(\kappa)$ and the relation $\varphi(H) := \max_i \varphi(h_i) \leq \omega$? How to find such an H, provided it exists?*

Invoking Lemma 1, we can reformulate this question as follows:

> (??) *Let $\varphi(\cdot)$ and s be as in (?). Given $(\omega, \kappa) \in \mathbf{R}^2_+$, how to find vectors $h_i \in \mathbf{R}^m$, $1 \leq i \leq n$, satisfying*
>
> $$(a): \ \varphi(h_i) \leq \omega; \ \& \ (b): \ |x_i| \leq |h_i^T Ax| + s^{-1}\kappa\|x\|_1 \, \forall x \in \mathbf{R}^n \qquad (\mathcal{P}_i)$$
>
> *for every i, or to detect correctly that no such collection of vectors exists?*

Indeed, by Lemma 1, if H' satisfies $\mathbf{H}_{s,\infty}(\kappa)$ and $\varphi(H') \leq \omega$, then there exists $H = [h_1, ..., h_n]$ such that h_i satisfy $(\mathcal{P}_i.b)$ for all i and $\varphi(H) \leq \varphi(H') \leq \omega$, so that h_i satisfy $(\mathcal{P}_i.a)$ for all i as well. Vice versa, if h_i satisfy $(\mathcal{P}_i)$, $1 \leq i \leq n$, then the matrix $H = [h_1, ..., h_n]$ clearly satisfies $\mathbf{H}_{s,\infty}(\kappa)$, and $\varphi(H) \leq \omega$.

The answer to (??) is given by the following

Lemma 3 *Given $\kappa > 0$, $\omega \geq 0$, and a positive integer s, let $\gamma = \kappa/s$. For every $i \leq n$, the following three properties are equivalent to each other:*

(i) *There exists $h = h_i$ satisfying $(\mathcal{P}_i)$;*

(ii) *The optimal value in the optimization problem*

$$\mathrm{Opt}_i(\gamma) = \min_h \left\{\varphi(h) : \|A^T h - e_i\|_\infty \leq \gamma\right\} \qquad (P_i^\gamma)$$

where e_i is i-th standard basic orth in $\mathbf{R}^n$, is $\leq \omega$;

(iii) *One has*

$$\forall x \in \mathbf{R}^n : |x_i| \leq \omega\varphi_*(Ax) + \gamma\|x\|_1, \qquad (27)$$

where $\varphi_*(u) = \max_{\varphi(v)\leq 1} u^T v$ is the norm on $\mathbf{R}^m$ conjugate to $\varphi(\cdot)$.

Whenever one (and then – all) of these properties take place, problem (P_i^γ) is solvable, and its optimal solution h_i satisfies $(\mathcal{P}_i)$.

3.1 Optimal contrasts for regular and penalized recoveries

As an immediate consequence of Lemma 3, we get the following description of the domain G_s associated with the norm $\varphi(\cdot) = \nu_{\epsilon,\sigma,\mathcal{U}}(\cdot)$:

$$\begin{array}{lrcl} (a) & G_s & = & \left\{(\omega,\kappa) \geq 0 : s^{-1}\kappa \geq \gamma_*, \, \omega \geq \omega_*(s^{-1}\kappa)\right\}, \\ \text{where} & & & \\ (b) & \gamma_* & = & \max\limits_{1\leq i\leq n} \min\limits_h \|A^T h - e_i\|_\infty = \max\limits_i \max\limits_x \left\{x_i : \|x\|_1 \leq 1, Ax = 0\right\}, \\ (c) & \omega_*(\gamma) & = & \max\limits_{1\leq i\leq n} \mathrm{Opt}_i(\gamma) \end{array} \qquad (28)$$

where $\phi(\cdot)$ in (P_i^γ) is specified as $\nu_{\epsilon,\sigma,\mathcal{U}}(\cdot)$. Note that the second equality in (b) is given by Linear Programming duality. Indeed, by (b), γ_* is the smallest γ for which all problems (P_i^γ), $i = 1, ..., n$, are feasible, and thus, by Lemma 3, $(\gamma, \kappa) \in G_s$ if and only if $\kappa/s \geq \gamma_*$ and $\omega \geq \omega_*(\kappa/s)$.

Note that the quantity γ_* depends solely of A, while $\omega_*(\cdot)$ depends on $\epsilon, \sigma, \mathcal{U}$, as on parameters, but is independent of s.

The outlined results suggest the following scheme of building the contrast matrix H:

- we compute γ_* by solving n Linear Programming problems in (28.b); if $s\gamma_* \geq \frac{1}{2}$, then G_s does not contain points (ω, κ) with $\kappa < 1/2$, so that our recovery routines are not applicable (or, at least, we cannot justify them theoretically);

- when $s\gamma_* < \frac{1}{2}$, the set G_s is nonempty, and its Pareto frontier (the set of pairs $(\omega, \kappa) \in \mathbf{R}^2_+$ such that $(\omega, \kappa) \geq (\omega', \kappa) \in G_s$ is possible if and only if $\omega' = \omega$) is the curve $(\omega_*(\gamma), s\gamma)$, $\gamma_* \leq \gamma < \frac{1}{2s}$. We choose a "working point" on this curve, that is, a point $\bar{\gamma} \in [\gamma_*, \frac{1}{2s}]$ and compute $\omega_*(\bar{\gamma})$ by solving the convex optimization programs $(P_i^{\bar{\gamma}})$, $i = 1, ..., n$, with $\phi(\cdot)$ specified as $\nu_{\epsilon,\sigma,\mathcal{U}}(\cdot)$. $\omega_*(\bar{\gamma})$ is nothing but the maximum, over i, of the optimal values of these problems, and the optimal solutions h_i to the problems induce the matrix $H = H(\bar{\gamma}) = [h_1, ..., h_n]$ which satisfies $\mathbf{H}_{s,\infty}(s\bar{\gamma})$ and has $\nu(H) \leq \omega_*(\bar{\gamma})$. By reasoning which led us to (??),

$$\nu(H(\bar{\gamma})) = \omega_*(\bar{\gamma}) = \min_{H'} \left\{ \nu(H') : H' \text{ satisfies } \mathbf{H}_{s,\infty}(s\bar{\gamma}) \right\},$$

  that is, $H = H(\bar{\gamma})$ is the best for our purposes contrast matrices satisfying $\mathbf{H}_{s,\infty}(s\bar{\gamma})$. With this contrast matrix, the error bound (26) for regular/penalized ℓ_1 recoveries (in the former, $\rho_i = \nu_i(h_i)$, in the latter, $\theta = 2$) read

$$1 \leq p \leq \infty \Rightarrow \mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon, \sigma, s, \upsilon) \leq 2\frac{s^{\frac{1}{p}}}{1 - 2s\bar{\gamma}}(s^{-1}\upsilon + 2\omega_*(\bar{\gamma})). \tag{29}$$

The outlined strategy does not explain how to choose $\bar{\gamma}$. This issue could be resolved, e.g., as follows. We choose an upper bound on the sensitivity of the risk (29) to υ, i.e., to the $\|\cdot\|_1$-deviation of a signal to be recovered from the set of s-sparse signals. This sensitivity is proportional to $\frac{1}{1-2s\bar{\gamma}}$, so that an upper bound on the sensitivity translates into an upper bound $\gamma^+ < \frac{1}{2s}$ on $\bar{\gamma}$. We can now choose $\bar{\gamma}$ by minimizing the remaining term in the risk bound over $\gamma \in [\gamma_*, \gamma^+]$, which amounts to solving the optimization problem

$$\max \left\{ \tau : \ \tau\omega_*(\gamma) \leq 1 - 2s\gamma, \ \gamma_* \leq \gamma \leq \gamma^+ \right\}.$$

Observing that $\omega_*(\cdot)$ is, by its origin, a convex function, we can solve the resulting problem efficiently by bisection in τ. A step of this bisection requires solving a *univariate* convex feasibility problem with efficiently computable constraint and thus is easy, at least for moderate values of n.

4 Range of feasibility of condition $\mathbf{H}_{s,\infty}(\kappa)$

We address the crucial question of what can be said about the magnitude of the quantity $\omega_*(\cdot)$, see (28) and the risk bound (29). One way to answer it is just to compute the (efficiently computable!) quantity $\omega_*(\gamma)$ for a desired value of γ. Yet it is natural to know theoretical upper bounds on ω_* in some "reference" situations. Below, we provide three results of this type.

At this point, it makes sense to express in the notation that $\omega_*(\gamma)$ depends, as on parameters, on the sensing matrix A and the "environment parameters" $\epsilon, \sigma, \mathcal{U}$, so that in this section we write $\omega_*(\gamma|A, \epsilon, \sigma, \mathcal{U})$ instead of $\omega_*(\gamma)$.

4.1 Bounding $\omega_*(\cdot)$ via mutual incoherence

Recall that for an $m \times n$ sensing matrix $A = [A_1, ..., A_n]$ with no zero columns, its *mutual incoherence* is defined as

$$\mu(A) = \max_{1 \le i \ne j \le n} \frac{|A_i^T A_j|}{A_i^T A_i}.$$

Compressed Sensing literature contains numerous mutual-incoherence-related results (see, e.g., [10, 11, 12] and references therein). To the best of our knowledge, all these results state that if s is a positive integer and A is a sensing matrix such that $\frac{s\mu(A)}{\mu(A)+1} < \frac{1}{2}$, then ℓ_1-based sparse recovery is well suited for recovering s-sparse signals (e.g., recovers them exactly when there is no observation noise, admit explicit error bounds when there is noise and/or the signal is only nearly s-sparse, etc.). To the best of our knowledge, all these results, up to the values of absolute constant factors in error bounds, are covered by the risk bounds (29) combined with the following immediate observation:

Observation 1 *Whenever $A = [A_1, ..., A_n]$ is an $m \times n$ matrix with no zero columns and s is a positive integer, the matrix $H(A) = \frac{1}{\mu(A)+1}[A_1/A_1^T A_1, A_2/A_2^T A_2, ..., A_n/A_n^T A_n]$ satisfies the condition $\mathbf{H}_{s,\infty}\left(\frac{s\mu(A)}{\mu(A)+1}\right)$.*

Verification is immediate: the diagonal entries in the matrix $Z = I - H^T A$ are equal to $\gamma := 1 - \frac{1}{\mu(A)+1} = \frac{\mu(A)}{\mu(A)+1}$, while the magnitudes of the off-diagonal entries in Z do not exceed γ. Therefore

$$\begin{aligned} & x \in \mathbf{R}^n \Rightarrow \gamma\|x\|_1 \ge \|Zx\|_\infty = \|x - H^T Ax\|_\infty \ge \|x\|_\infty - \|H^T Ax\|_\infty \Leftrightarrow \|x\|_\infty \le \|H^T Ax\|_\infty + \gamma\|x\|_1 \\ & \Leftrightarrow \ H \text{ satisfies } \mathbf{H}_{s,\infty}(s\gamma). \end{aligned}$$

Observe that the Euclidean norms of the columns in $H(A)$ do not exceed $\left[\min_i \|A_i\|_2\right]^{-1}$, whence $\nu(H(A)) \le r(\mathcal{U}) + \frac{\sigma\sqrt{2\ln(n/\epsilon)}}{\min_i \|A_i\|_2}$, where $r(\mathcal{U}) = \max_{u \in \mathcal{U}} \|u\|_2$. In the notation from Section 3, our observations can be summarized as follows:

Corollary 1 *For every $m \times n$ matrix A with no zero columns, one has $\gamma_* \le \gamma := \frac{\mu(A)}{\mu(A)+1}$ and $\omega_*(\gamma|A, \epsilon, \sigma, \mathcal{U}) \le \nu(H(A)) \le r(\mathcal{U}) + \frac{\sigma\sqrt{2\ln(n/\epsilon)}}{\min_i \|A_i\|_2}$. In particular,*

$$s \le \frac{\mu(A)+1}{3\mu} \Rightarrow \omega_*(\frac{1}{3s}|A, \epsilon, \sigma, \mathcal{U}) \le r(\mathcal{U}) + \frac{\sigma\sqrt{2\ln(n/\epsilon)}}{\min_i \|A_i\|_2}.$$

It should be added that as m, n grow in such a way that $\ln(n) \le O(1)\ln m$, realizations A of "typical" random $m \times n$ matrices (e.g., those with independent $\mathcal{N}(0, 1/m)$ entries or with independent entries taking values $\pm 1/\sqrt{m}$) with overwhelming probability satisfy $\mu(A) \le O(1)\sqrt{\ln(n)/m}$ and $\|A_i\|_2 \le 0.9$ for all i. By Corollary 1, it follows that for these A the condition $\mathbf{H}_{s,\infty}(\kappa)$ with, say, $\kappa = 1/3$ can be satisfied for s as large as $O(1)\sqrt{m/\ln(n)}$ merely by the choice $H = H(A)$, which ensures that $\nu(H) \le O(1)[r(\mathcal{U}) + \sigma\sqrt{2\ln(n/\epsilon)}]$; in particular, in the indicated range of values of s one has $\omega_*(\frac{1}{3s}) \le O(1)[r(\mathcal{U}) + \sigma\sqrt{2\ln(n/\epsilon)}]$.

4.2 The case of A satisfying the Restricted Isometry Property

Proposition 5 *Let A satisfy $\mathrm{RIP}(\delta, k)$ with some $\delta \in (0,1)$ and with $k > 1$. Then there exists matrix $H(A)$ which, for every positive integer s, satisfies the condition $\mathbf{H}_{s,\infty}(s\gamma(\delta, k))$, with*

$$\gamma(\delta, k) = \frac{\sqrt{2}\delta}{(1-\delta)\sqrt{k-1}}, \tag{30}$$

and is such that $\nu(H(A)) \le \left[r(\mathcal{U}) + \sigma\sqrt{2\ln(n/\epsilon)}\right]/\sqrt{1-\delta}$. In particular,

$$s \le \frac{1-\delta}{3\sqrt{2}}\sqrt{k-1} \Rightarrow \omega_*(\frac{1}{3s}|A, \epsilon, \sigma, \mathcal{U}) \le \frac{1}{\sqrt{1-\delta}}\left[r(\mathcal{U}) + \sigma\sqrt{2\ln(n/\epsilon)}\right]. \tag{31}$$

4.3 Oracle inequality

Here we assume that $A \in \mathbf{R}^{m\times n}$ possesses the following property (where S is a positive integer and $\varphi > 0$):

> $\mathbf{O}(S,\omega)$: For every $i \in \{1,...,n\}$ and every S-element subset $I \ni i$ of $\{1,...,n\}$ there exists a routine $\mathcal{R}_{i,I}$ for recovering x_i from a noisy observation
>
> $$y = Ax + u + \sigma e, \qquad [e \sim \mathcal{N}(0, I_m), u \in \mathcal{U}]$$
>
> of *unknown* signal $x \in \mathbf{R}^n$, *known* to be supported on I such that for every such signal and every $u \in \mathcal{U}$ one has
>
> $$\mathrm{Prob}\{|\mathcal{R}_{i,I}(Ax + u + \sigma e) - x_i| \geq \omega\} \leq \epsilon.$$

We intend to demonstrate that in this situation for all s in certain range (which extends as S grows and ω decreases) the uniform error of the regular and the penalized recoveries associated with properly selected contrast matrix is, with probability $\geq 1-\epsilon$, "close" to ω. The precise statement is as follows:

Proposition 6 *Given A and the "environment parameters" $\epsilon < 1/16$, σ, $\mathcal{U}$, assume that A satisfies the condition $\mathbf{O}(S,\gamma)$ with certain S,γ. Then for every integer s from the range*

$$1 \leq s \leq \frac{\sigma\sqrt{2S\ln(1/\epsilon)}}{4\omega\|A\|} \tag{32}$$

(here $\|\cdot\|$ is the standard matrix norm, the largest singular value) there exists a contrast matrix H satisfying the condition $\mathbf{H}_{s,\infty}(\frac{1}{4})$ and such that $\nu(H) \leq 2\sqrt{1+\ln(n)/\ln(1/\epsilon)}\omega$, so that in the outlined range of values of H one has $\omega_*(\frac{1}{4s}) \leq 2\sqrt{1+\ln(n)/\ln(1/\epsilon)}\omega$, and the associated with H error bound (29) for regular/penalized ℓ_1 recovery is

$$\mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon,\sigma,s,\upsilon) \leq 16 s^{\frac{1}{p}}\left[\omega\sqrt{1+\frac{\ln n}{\ln(1/\epsilon)}} + \frac{\upsilon}{4s}\right]. \tag{33}$$

Proposition 6 justifies to some extent, our approach; it says that if there exists a routine which recovers S-sparse signals *with a priori known sparsity pattern* within certain accuracy (measured component-wise), then our recovering routines exhibit "close" performance without any knowledge of the sparsity pattern, albeit in a smaller range of values of the sparsity parameter.

4.4 Condition $\mathbf{H}_{s,\infty}(\kappa)$: limits of performance

Recall that when recovering s-sparse signals, the condition $\mathbf{H}_{s,\infty}(\kappa)$ helps *only when* $\kappa < 1/2$. Unfortunately, with these κ, the condition is feasible in a severely restricted range of values of s. Specifically, from [15, Proposition 5.1] and Lemma 1 it immediately follows that

> (*) *If $A \in \mathbf{R}^{m\times n}$ is not "nearly square," that is, if $n > 2(2\sqrt{m}+1)^2$, then the condition $\mathbf{H}_{s,\infty}(\kappa)$ with $\kappa < 1/2$ cannot be satisfied when s is "large", namely, when $s > 2\sqrt{2m}+1$.*

Note that from the discussion at the end of section 4.1 we know that the "$O(\sqrt{m})$ limit of performance" of the condition $\mathbf{H}_{s,\infty}(\cdot)$ stated in (*) is "nearly sharp:" – when $s \leq O(1)\sqrt{m}$, the condition $\mathbf{H}_{s,\infty}(\frac{1}{3})$ associated with a typical randomly generated $m \times n$ sensing matrix A is feasible and can be satisfies with a contrast matrix H with quite moderate $\nu(H)$.

(*) says that unless A is nearly square, the condition $\mathbf{H}_{s,\infty}(\cdot)$ can validate ℓ_1 sparse recovery only in a severely restricted range $s \leq O(\sqrt{m})$ of values of the sparsity parameter. This is in sharp contrast with *unverifiable* sufficient conditions for "goodness" of ℓ_1 recovery, like RIP: it is well known that when m,n grow, realizations of "typical" random $m \times n$ matrices, like those mentioned at the end of Section 4.1, with overwhelming probability possess RIP$(0.1, 2s)$ with s as large as $O(m/\ln(2n/m))$. As a result, "unverifiable" sufficient conditions, like RIP, can justify the validity of ℓ_1 recovery routines in a much wider (and in fact –

the widest possible) range of values of the sparsity parameter s than the "fully computationally tractable" condition $\mathbf{H}_{s,\infty}(\cdot)$. This being said, note that this comparison is not completely fair. Indeed, aside of its tractability, the condition $\mathbf{H}_{s,\infty}(\kappa)$ with $\kappa < 1/2$ ensures the error bounds (29) *in the entire range* $1 \le p \le \infty$ *of values of* p, which perhaps is not the case with conditions like RIP. Specifically, consider the "no nuisance" case $\mathcal{U} = \{0\}$, and let A satisfy RIP$(0.1, 2S)$ for certain S. It is well known (see, e.g., the next section) that in this case the Dantzig Selector recovery ensures for every $s \le S$ and every s-sparse signal x that

$$\|\widehat{x}_{\mathrm{DS}} - x\|_p \le O(1)\sigma\sqrt{\ln(n/\epsilon)}s^{1/p},\ 1 \le p \le 2,$$

with probability $\ge 1 - \epsilon$. However, we are not aware of similar bounds (under whatever conditions) for "large" s and $p > 2$. For comparison: in the case in question, for "small" s, namely, $s \le O(1)\sqrt{S}$, we have $\omega_*(\frac{1}{3s}) \le O(1)\sigma\sqrt{\ln(n/\epsilon)}$ (by Proposition 5), whence for regular and penalized ℓ_1 recoveries with appropriately chosen contrast matrix (which can be built efficiently!) one has for all s-sparse x

$$\|\widehat{x} - x\|_p \le O(1)\sigma\sqrt{\ln(n/\epsilon)}s^{\frac{1}{p}} \quad \forall p \in [1, \infty]$$

with probability $\ge 1 - \epsilon$ (see (29)). We wonder whether a similar (perhaps, with extra logarithmic factors) bound can be obtained *for large* s (e.g., $s \ge m^{\frac{1}{2}+\delta}$) for a *whatever* ℓ_1 recovery routine and a *whatever* essentially nonsquare (say, $m < n/2$) $m \times n$ sensing matrix A with columns of Euclidean length $\le O(1)$.

5 Extensions

We are about to demonstrate that the pivot element of the preceding sections — the condition $\mathbf{H}_{s,\infty}(\kappa)$ — is the strongest (and seemingly the only verifiable one) in a natural parametric series of conditions on a contrast matrix H; every one of these conditions validates the regular and the penalized ℓ_1 recoveries associated with H in certain *restricted* range of values of p in the error bounds (29).

5.1 Conditions $\mathbf{H}_{s,q}(\kappa)$

Let us fix an $m \times n$ sensing matrix A. Given a positive integer $s \le m$, a $q \in [1, \infty]$ and a real $\kappa > 0$, let us say that an $m \times n$ contrast matrix H satisfies condition $\mathbf{H}_{s,q}(\kappa)$, if

$$\forall x \in \mathbf{R}^n : \|x\|_{s,q} \le s^{\frac{1}{q}}\|H^T Ax\|_\infty + \kappa s^{\frac{1}{q}-1}\|x\|_1, \tag{34}$$

where $\|x\|_{s,q} = \|x^s\|_q$ and x^s, as always, is the vector obtained from x by zeroing all but the s largest in magnitude entries. Observe that

- What used to be denoted $\mathbf{H}_{s,\infty}(\kappa)$ before, is exactly what is called $\mathbf{H}_{s,\infty}(\kappa)$ now;
- If H satisfies $\mathbf{H}_{s,q}(\kappa)$, H satisfies $\mathbf{H}_{s,q'}(\kappa)$ for all $q' \in [1, q]$ (since for s-sparse vector x^s we have $\|x^s\|_{q'} \le s^{\frac{1}{q'}-\frac{1}{q}}\|x^s\|_{s,q}$).

Less immediate observations are as follows:

- Let A be an $m \times n$ matrix and let $s \le n$ be a positive integer. We say that A is *s-good* if for all s-sparse $x \in \mathbf{R}^n$ the ℓ_1-recovery
$$\widehat{x} \in \operatorname*{Argmin}_v\{\|v\|_1 : Av = y\}$$
is exact in the case of noiseless observation $y = Ax$. It turns out that feasibility of $\mathbf{H}_{s,1}(\kappa)$ with $\kappa < \frac{1}{2}$ is intimately related to s-goodness of A:

  Lemma 4 *A is s-good if and only if there exist $\kappa < \frac{1}{2}$ and $H \in \mathbf{R}^{m\times n}$ satisfying $\mathbf{H}_{s,1}(\kappa)$.*

- The Restricted Isometry Property implies feasibility of $\mathbf{H}_{s,2}(\kappa)$ with small κ:

  Lemma 5 *Let A satisfy* RIP$(\delta, 2s)$ *with $\delta < \frac{1}{3}$. Then the matrix $H = \frac{1}{1-\delta}A$ satisfies the condition $\mathbf{H}_{s,2}(\kappa)$ with $\kappa = \frac{\delta}{1-\delta} < \frac{1}{2}$.*

5.2 Regular and penalized ℓ_1 recoveries with contrast matrices satisfying $\mathbf{H}_{s,q}(\kappa)$

Our immediate goal is to obtain the following extension of the main results of Section 2, specifically, Propositions 2, 4:

Proposition 7 *Assume we are given an $m \times n$ sensing matrix $A = [a_1, ..., a_n]$, an integer $s \leq m$, $\kappa < 1/2$, a contrast matrix $H = [h_1, ..., h_n] \in \mathbf{R}^{m \times n}$, and $q \in [1, \infty]$ such that H satisfies the condition $\mathbf{H}_{s,q}(\kappa)$. Denote $\nu_i = \nu_{\epsilon,\sigma,\mathcal{U}}(h_i)$, where the norm $\nu_{\epsilon,\sigma,\mathcal{U}}(\cdot)$ is defined in (4), and $\nu(H) = \max_i \nu_i$. Let also noise intensity σ, uncertainty set $\mathcal{U}$ and tolerance $\epsilon \in (0, 1)$ be given.*

(i) *Consider the regular recovery (7) with the contrast matrix H and the parameters ρ_i satisfying the relations*

$$\rho_i \geq \nu_i, \, 1 \leq i \leq n,$$

and let $\widehat{\rho} = \max_i \rho_i$. Then

$$1 \leq p \leq q \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\mathrm{reg}}(\cdot)|\epsilon, \sigma, s, \upsilon) \leq (3s)^{\frac{1}{p}} \frac{\widehat{\rho} + \nu(H) + s^{-1}\upsilon}{1 - 2\kappa}. \tag{35}$$

(ii) *Consider the penalized recovery (17) with the contrast matrix H and $\theta = 2$. Then*

$$1 \leq p \leq q \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\mathrm{pen}}(\cdot)|\epsilon, \sigma, s, \upsilon) \leq 3s^{\frac{1}{p}} \frac{2\nu(H) + s^{-1}\upsilon}{1 - 2\kappa}. \tag{36}$$

5.3 Error bounds for Lasso and Dantzig Selector under condition $\mathbf{H}_{s,q}(\kappa)$

We are about to demonstrate that the feasibility of condition $\mathbf{H}_{s,q}(\kappa)$ with $\kappa < \frac{1}{2}$ implies some consequences for the performance of Lasso and Dantzig Selector when recovering s-sparse signals in $\|\cdot\|_p$ norms, $1 \leq p \leq q$. This might look strange at the first glance, since neither Lasso nor Dantzig Selector use contrast matrices. The surprise, however, is eliminated by the following observation:

(!) *Let H satisfy $\mathbf{H}_{s,q}(\kappa)$ and let $\widehat{\lambda}$ be the maximum of the Euclidean norms of columns in H. Then*

$$\forall x \in \mathbf{R}^n : \|x\|_{s,q} \leq \widehat{\lambda} s^{\frac{1}{q}} \|Ax\|_2 + \kappa s^{\frac{1}{q}-1} \|x\|_1. \tag{37}$$

The fact that a condition like (37) with $\kappa < 1/2$ plays a crucial role in the performance analysis of Lasso and Dantzig Selector is neither surprising nor too novel. For example, the standard error bounds for the latter algorithms under the RIP assumption are in fact based on the validity of (37) with $\widehat{\lambda} = O(1)$ for $q = 2$ (see Lemma 5). Another example is given by the Restricted Eigenvalue [2] and the Compatibility conditions [3, 19]. Specifically, the Restricted Eigenvalue condition $\mathrm{RE}(s, \rho, \varkappa)$ (s is positive integer, $\rho > 1$, $\varkappa > 0$ states that

$$\|x^s\|_2 \leq \frac{1}{\varkappa} \|Ax\|_2 \text{ whenever } \rho \|x^s\|_1 \geq \|x - x^s\|_2,$$

whence $\|x^s\|_1 \leq \frac{\sqrt{s}}{\varkappa} \|Ax\|_2$ whenever $(\rho + 1)\|x^s\|_1 \geq \|x\|_1$, so that

$$\forall x \in \mathbf{R}^n : \|x\|_{s,1} \leq \frac{s^{1/2}}{\varkappa} \|Ax\|_2 + \frac{1}{1 + \rho} \|x\|_1;. \tag{38}$$

Further, the Compatibility condition of [19] is nothing but (38) with $\rho = 3$. We see that both Restricted Eigenvalue and Compatibility conditions imply (37) with $q = 1$, $\widehat{\lambda} = (\varkappa\sqrt{s})^{-1}$ and certain $\kappa < 1/2$.

We are about to present a simple result on the performance of Lasso and Dantzig Selector algorithms in the case when A satisfies the condition (37). The result is as follows:

Proposition 8 *Let $m \times n$ matrix $A = [a_1, ..., a_n]$ satisfy (37) with $\kappa < \frac{1}{2}$ and some $q \in [1, \infty]$, and let $\beta = \max_i \|a_i\|_2$. Let also the "environment parameters" $\sigma > 0$, $\epsilon \in (0,1)$ be given, and let there be no nuisance: $\mathcal{U} = \{0\}$.*

(i) *Consider the Dantzig Selector recovery*

$$\widehat{x}_{\mathrm{DS}}(y) \in \operatorname*{Argmin}_v \left\{ \|v\|_1 : \|A^T(Av - y)\|_\infty \le \rho \right\},$$

where

$$\rho \ge \varrho := \sigma\beta\sqrt{2\ln(n/\epsilon)}. \tag{39}$$

Then

$$1 \le p \le q \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\mathrm{DS}}(\cdot)|\epsilon, \sigma, s, \upsilon) \le \frac{2(3s)^{\frac{1}{p}}}{1-2\kappa}\left[\frac{2s\widehat{\lambda}^2(\rho+\varrho)}{1-2\kappa} + s^{-1}\upsilon\right]. \tag{40}$$

(ii) *Consider the Lasso recovery*

$$\widehat{x}_{\mathrm{lasso}}(y) \in \operatorname*{Argmin}_v \left\{ \|v\|_1 + \varkappa\|Av - y\|_2^2 \right\},$$

and let $\varkappa$ satisfy the relation

$$2\kappa + 2\varrho\varkappa < 1,$$

where ϱ is given by (39). Then

$$1 \le p \le q \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\mathrm{lasso}}(\cdot)|\epsilon, \sigma, s, \upsilon) \le \frac{4s^{\frac{1}{p}}}{1-2\kappa-2\varrho\varkappa}\left[\frac{2s\widehat{\lambda}^2}{\varkappa} + s^{-1}\upsilon\right]. \tag{41}$$

In particular, with

$$\varkappa = \frac{1-2\kappa}{4\varrho}, \tag{42}$$

one has

$$1 \le p \le q \Rightarrow \mathrm{Risk}_p(\widehat{x}_{\mathrm{lasso}}(\cdot)|\epsilon, \sigma, s, \upsilon) \le \frac{8s^{\frac{1}{p}}}{1-2\kappa}\left[\frac{8s\varrho\widehat{\lambda}^2}{1-2\kappa} + s^{-1}\upsilon\right]. \tag{43}$$

Discussion. Let us compare the error bounds given by Propositions 7, 8. Assume that there is no nuisance ($\mathcal{U} = \{0\}$) and A is such that the condition $H_{s,q}(\frac{1}{4})$ is satisfied by certain matrix H, the maximum of Euclidean norms of the columns of H being $\widehat{\lambda}$. Assuming that the penalized recovery uses $\theta = 2$, and the regular recovery uses $\widehat{\rho} = \nu(H) = \widehat{\lambda}\sigma\sqrt{2\ln(n/\epsilon)})$, the associated risk bounds as given by Proposition 7 become

$$\mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon, \sigma, s, \upsilon) \le O(1)s^{\frac{1}{p}}\left[\widehat{\lambda}\sigma\sqrt{2\ln(n/\epsilon)} + s^{-1}\upsilon\right] \;\; 1 \le p \le q. \tag{44}$$

Note that these bounds admit a transparent interpretation: in the range $1 \le p \le q$ an s-sparse signal is recovered *as if* we were identifying correctly its support and estimating the entries with the uniform error $O(1)\widehat{\lambda}\sigma\sqrt{2\ln(n/\epsilon)}$.

Now, as we have already explained, the existence of a matrix H satisfying $\mathbf{H}_{s,q}(\frac{1}{4})$ with columns in H being of Euclidean lengths $\le \widehat{\lambda}$ implies validity of (37) with $\kappa = \frac{1}{4}$. Assuming that in Dantzig Selector one uses $\rho = \varrho$, and that $\varkappa$ in Lasso is chosen according to (42), the error bounds for Dantzig Selector and Lasso as given by Proposition 8 become

$$\mathrm{Risk}_p(\widehat{x}(\cdot)|\epsilon, \sigma, s, \upsilon) \le O(1)s^{\frac{1}{p}}\left[[\beta\widehat{\lambda}]s\widehat{\lambda}\sigma\sqrt{2\ln(n/\epsilon)} + s^{-1}\upsilon\right] \;\; 1 \le p \le q. \tag{45}$$

Observe that $\beta\widehat{\lambda} \ge O(1)$ (look what happens with (37) when x is the i-th basic orth). We see that the bounds (45) are worse than the bounds (44), primarily due to the presence of the factor s in the first bracketed

term in (45). At this point it is unclear whether this drawback is an artifact caused by poor analysis of the Dantzig Selector and Lasso algorithms or it indeed "reflects reality." Some related numerical results presented in Section 6.1 suggest that the latter option could be the actual one.

Moreover, consider an example of the recovery problem with a 2×2 matrix A with unit columns and singular values 1 and ε. It can be easily seen that if x is aligned with the second right singular vector of A (corresponding to the singular value ε) the error of the Dantzig Selector may be as large as $O(\varepsilon^{-2}\sigma)$, while the error of "$H$-conscious" recovery will be $O(\varepsilon^{-1}\sigma)$ up to the logarithmic factor in ϵ (indeed, choosing $H = A^{-1}$ results in $\lambda = \varepsilon^{-1}$). This toy example suggests that the extra λ factor in the bound (45), at least for Dantzig Selector, is not only due to our clumsy analysis.

This being said, it should be stressed that the comparison of regularized/penalised ℓ_1 recoveries with Dantzig Selector and Lasso based solely on above the error bounds is somehow biased against Dantzig Selector and Lasso. Indeed, in order for regular/penalized ℓ_1 recoveries to enjoy their "good" error bounds, we should specify the required contrast matrix, which is not the case for Lasso and Dantzig Selector: the bounds (45) require only existence of such a matrix[5]. Besides this, there is at least one case where error bounds for Dantzig Selector are as good as (44), specifically, the case when A possesses, say, RIP$(0.1, 2s)$. Indeed, in this case, by Lemma 5, the matrix $H = O(1)A$ satisfies $\mathbf{H}_{s,2}(\frac{1}{4})$, meaning that Dantzig Selector with properly chosen ρ is nothing but the regular recovery with contrast matrix H and as such obeys the bounds (44) with $q = 2$.

It is time to point out that the above discussion is somehow scholastic: when $q < \infty$ and s is nontrivial, we do not know how to verify efficiently the fact that the condition $\mathbf{H}_{s,q}(\kappa)$ is satisfied by a given H, not speaking about efficient synthesis of H satisfying this condition. One should not think that these tractability issues concern only our algorithms which need a good contrast matrix. In fact, *all* conditions which allow to validate Dantzig Selector and Lasso beyond the scope of the "fully tractable" condition $H_{s,\infty}(\kappa)$ are, to the best of our knowledge, unverifiable – they cannot be checked efficiently, and thus we never can be sure that Lasso and Dantzig Selector (or any other known computationally efficient technique for sparse recovery) indeed work well for a *given sensing matrix*. As we have seen in Section 3, the situation improves dramatically when passing from unverifiable conditions $\mathbf{H}_{s,q}(\kappa)$, $q < \infty$, to the efficiently verifiable condition $\mathbf{H}_{s,\infty}(\kappa)$, although in a severely restricted range of values of s.

6 Numerical examples

We present here a small simulation study.

6.1 Regular/penalized recovery vs. Lasso: no-nuisance case

To illustrate the discussion in Section 5.3, we compare numerical performance of Lasso and penalized recovery in the observation model (1) without nuisance:

$$y = Ax + \sigma\xi, \quad \xi \sim N(0, I_m),$$

where $\sigma > 0$ is known. The sensing matrix A is specified by selecting at random $m = 120$ rows of the 128×128 Hadamard matrix[6], and "suppressing" the first of the selected rows by multiplying it by 1.e-3. The resulting 120×128 sensing matrix has orthogonal rows; 119 of its 120 singular values are equal to $8\sqrt{2}$, and the remaining singular value is $0.008\sqrt{2}$.

We have processed A as explained in Section 3 (a reader is referred to this section for the description of

[5] And even less than that, since feasibility of $H_{s,q}(\kappa)$ is just a *sufficient* condition for the validity of (37), the condition which indeed underlies Proposition 8.

[6] The k-th Hadamard matrix H^k is given by the recurrence $H^0 = 1$, $H^{p+1} = [H^p; H^p; H^p, -H^p]$. It is a $2^k \times 2^k$ matrix with orthogonal rows and all entries equal to ± 1.

			$\|\widehat{x}-x\|_p$		
Recovery	σ	$\varkappa$	$p=1$	$p=2$	$p=\infty$
Penalized		N/A	2.1e-4	6.5e-5	3.8e-5
Lasso	1.e-4	3.74e-3*	2.1e-4	5.2e-4	3.9e-5
Lasso		4.01e-2†	1.6e-3	5.2e-4	2.0e-4
Penalized		N/A	2.2e-5	6.0e-6	2.7e-6
Lasso	1.e-5	4.78e-4*	3.1e-5	8.1e-6	3.4e-6
Lasso		4.01e-3†	1.8e-4	5.8e-5	2.1e-5
Penalized		N/A	2.1e-6	6.2e-7	2.5e-7
Lasso	1.e-6	1.10e-4*	8.8e-6	1.6e-6	5.9e-7
Lasso		4.01e-4†	1.8e-5	5.4e-6	1.9e-6

Table 1: . Lasso vs. penalized ℓ_1 recovery. Choice of $\varkappa$: * – "ideal" choice; † – theoretical choice.

entities involved).[7] We started with computing γ_*, which turned out to be 0.0287, meaning that the level of s-goodness of A is at least 17. In our experiment, we aimed at recovering signals with at most $s = 10$ nonzero entries and with no nuisance ($\mathcal{U} = \{0\}$). The synthesis of the corresponding "optimal" contrast matrix $H = H_*$ as outlined in Section 3 results in $\bar{\gamma} = 0.294$, $\omega_*(\bar{\gamma}) = 0.0899\sqrt{2\ln(n/\epsilon)}$. Note that we are in the case of $\mathcal{U} = \{0\}$, and in this case the optimal H is independent of the values of σ and ϵ.

We compare the penalized ℓ_1-recovery with the contrast matrix H_* and $\theta = 2$ with the Lasso recovery on randomly generated signals x with 10 nonzero entries. We consider two choices of the penalty $\varkappa$ in Lasso: the "theoretically optimal" choice (42) and the "ideal" choice, where we scanned the "fine grid" $(1.05)^k$, $k = 0, \pm1, \pm2, ...$ of values of $\varkappa$ and selected the value for which the Lasso recovery was at the smallest $\|\cdot\|_1$-distance from the true signal. The confidence parameter ϵ in (42) was set to 0.01.

The results of a typical experiment are presented in Table 1. We see that as compared to the penalized ℓ_1 recovery, the accuracy of Lasso with the theoretically optimal choice of the penalty is nearly 10 times worse. With the "ideal" (unrealistic!) choice of penalty, Lasso is never better than the penalized ℓ_1 recovery, and for the smallest value of σ is nearly 4 times worse than the latter routine.

6.2 The nuisance case

In the second experiment we study the behavior of recovery procedures in the situation when an "input nuisance" is present:

$$y = A(x + v) + \sigma\xi,$$

where $x \in \mathbf{R}^n$ is an unknown sparse signal, $v \in \mathcal{V}$ with known $\mathcal{V} \subset \mathbf{R}^n$, σ is known and $\xi \in \mathbf{R}^m$ is standard normal $\xi \sim N(0, I_m)$; in terms of (3), $u = Av$ and $\mathcal{U} = A\mathcal{V}$. We compare the performance of the regular and penalized recoveries to that of the Lasso and Dantzig Selector algorithms. To handle the nuisance, the latter methods were modified as follows: instead of the standard Lasso estimator we use the estimator

$$\widehat{x}_{\rm lasso}(y) \in \operatorname*{Argmin}_{x\in\mathbf{R}^n} \min_{v\in\mathcal{V}} \left\{\|x\|_1 + \varkappa\|A(x+v) - y\|_2^2\right\},$$

where the penalization coefficient $\varkappa$ is chosen according to [2, Theorem 4.1]; in turn, the Dantzig Selector is substituted by

$$\widehat{x}_{\rm DS}(y) \in \operatorname*{Argmin}_{x\in\mathbf{R}^m} \min_{v\in\mathcal{V}} \left\{\|x\|_1 \ : |[A^T(A(x+v) - y)]_i| \le \varrho_i, \ \ i = 1, ..., m\right\}$$

[7] It is worth to mention that when A is comprised of (perhaps, scaled) rows of an Hadamard matrix (and in fact, of scaled rows of any other Fourier transform matrix associated with a finite Abelian group) the synthesis described in Section 3 simplifies dramatically due to the fact that all problems (P_i^γ) turn out to be equivalent to each other, and their optimal solutions are obtained from each other by simple linear transformations. As a result, we can work with a *single* problem (P_1^γ) instead of working with n of them.

with $\varrho_i = \sigma\sqrt{2\ln(n/\epsilon)}\|A_i\|_2$, where A_i are the columns of A and ϵ is given (in what follows $\epsilon = 0.01$).

We present below the simulation results for two setups with $n = 256$:

1. *Gaussian setup*: a 161×256 sensing matrix A_{Gauss} with independent $N(0,1)$ entries is generated, then its columns are normalized. The nuisance set $\mathcal{V} = \mathcal{V}(L) \subset \mathbf{R}^{256}$ is as follows:

$$\mathcal{V}(L) = \{v \in \mathbf{R}^{256},\ |v_{i+1} - 2v_i + v_{i-1}| \le L,\ \text{for } i = 2, ..., 255,\ v_2 = v_1 = 0\},$$

   where L is a known parameter; in other words, we observe the sum of a sparse signal and "smooth background."

2. *Convolution setup*: a 240×256 sensing matrix A_{conv} is constructed as follows: consider a signal x "living" on $\mathbf{Z}^2$ and supported on the 16×16 grid $\Gamma = \{(i,j) \in \mathbf{Z}^2 :\ 0 \le i, j \le 15\}$. We subject such a signal to discrete time convolution with a kernel supported on the set $\{(i,j) \in \mathbf{Z}^2 : -7 \le i, j \le 7\}$, and then restrict the result on the 16×15 grid $\Gamma_+ = \{(i,j) \in \Gamma : 1 \le j \le 15\}$. This way we obtain a linear mapping $x \mapsto A_{\text{conv}}x :\ \mathbb{R}^{256} \to \mathbb{R}^{240}$. The nuisance set $\mathcal{V} = \mathcal{V}(L) \subset \mathbf{R}^{256}$ is composed of zero-mean signals u on Γ which satisfy

$$|[D^2u]_{i,j}| \le L,$$

   where D is the discrete (periodic) homogeneous Laplace operator:

$$[Du]_{i,j} = \frac{1}{4}\left(u_{i,\overline{j-1}} + u_{\overline{i-1},j} + u_{i,\overline{j+1}} + u_{\overline{i+1},j} - 4u_{i,j}\right),\quad i, j = 1, ..., 16,$$

   with $\overline{i} = i \mod 16$, $\overline{j} = j \mod 16$.

In the simulations we acted as follows: given the sensing matrix A, the nuisance set $\mathcal{U} = A\mathcal{V}$ and the values of s and σ, we compute the contrast matrix H by choosing a "reasonable" value $\bar{\gamma} > \gamma_*$ of γ and specifying H as the matrix satisfying $\mathbf{H}_{s,\infty}(s\bar{\gamma})$ and such that $\nu(H) = \omega_*(\bar{\gamma})$, see Section 3. Then N samples of *random* signal x, random nuisance $v \in \mathcal{V}$ and random perturbation ξ were generated, and the corresponding observations were processed by every one of the algorithms we are comparing[8]. The plots below present the average, over these $N = 100$ experiments, ℓ_∞ and ℓ_1 recovery errors. All recovery procedures were using `Mosek` optimization software [1].

We start with Gaussian setup in which the signal x has $s = 2$ non-vanishing components, randomly drawn, with $\|x\|_1 = 10$. For the penalized and the regular recovery algorithms the contrast matrix H was computed using $\bar{\gamma} = 0.1$. On Figure 1 we plot the average recovery error as a function of the value of the parameter L of the nuisance set $\mathcal{V}$, for fixed $\sigma = 0.1$, and on Figure 2 — as a function of σ for fixed $L = 0.01$.

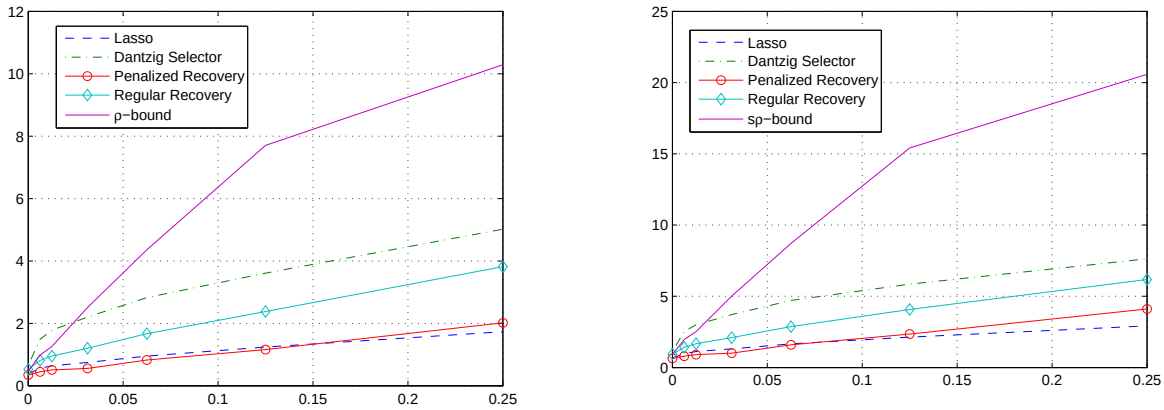
In the next experiment we fix the "environmental parameters" σ, L and vary the number s of nonzero entries in the signal x (of norm $\|x\|_1 = 5s$). On Figure 3 we present the recovery error as a function of s.

We run the same simulations in the convolution setup. The contrast matrix H for the penalized and the regular recoveries is computed using $\bar{\gamma} = 0.2$. On Figure 4 we plot the average recovery error as a function of the "size" L of the nuisance set $\mathcal{V}$ for fixed $\sigma = 0.1$, on Figure 5 — as a function of σ for fixed $L = 0.01$, and on Figure 6 — as a function of s.

We observe quite different behavior of the recovery procedures in our two setups. In the Gaussian setup the nuisance signal $v \in \mathcal{V}$ does not mask the true signal x, and the performance of the Lasso and Dantzig Selector is quite good in this case. The situation changes dramatically in the convolution setup, where the performance of the Lasso and Dantzig Selector degrades rapidly when the parameter L of the nuisance set increases.[9] The conclusion suggested by the outlined numerical results is that the penalized ℓ_1 recovery, while sometimes losing slightly to Lasso, in some of the experiments outperforms significantly all other algorithms we are comparing.

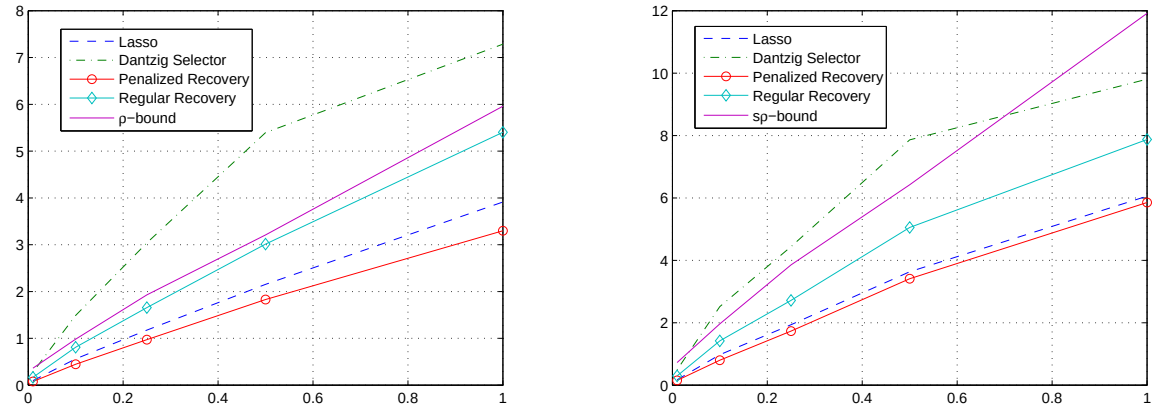
[8] Randomness of the sparse signal x is important. Using the techniques of [14], one can verify that in the convolution setup there are signals with only 3 non-vanishing components which cannot be recovered by ℓ_1 minimization even in the noiseless case $\mathcal{V} = \{0\}$, $\sigma = 0$. In other words, the s-goodness characteristic of the corresponding matrix A is equal to 2.

[9] The error plot for these estimators on Figure 4 flatters for higher values of L simply because they always underestimate the signal, and the error of recovery is always less than the corresponding norm of the signal.



ℓ_∞-error ℓ_1-error

Figure 1: Mean recovery error as a function of the nuisance magnitude L. Gaussian setup parameters: $\sigma = 0.1$, $s = 2$, $\mu = 0.1$, $\|x\|_1 = 10$.



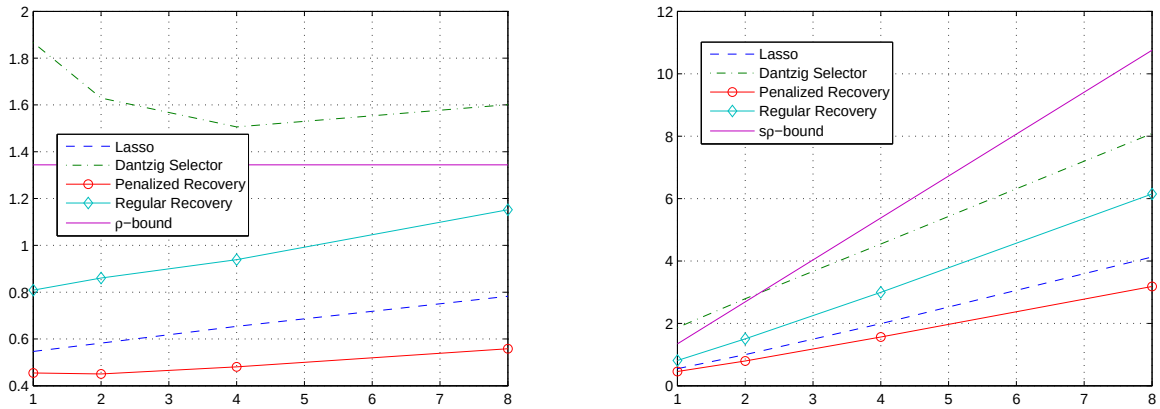
ℓ_∞-error ℓ_1-error

Figure 2: Mean recovery error as a function of the noise StD σ. Gaussian setup parameters: $L = 0.01$, $s = 2$, $\mu = 0.1$, $\|x\|_1 = 10$.

7 Non-Euclidean matching pursuit algorithm

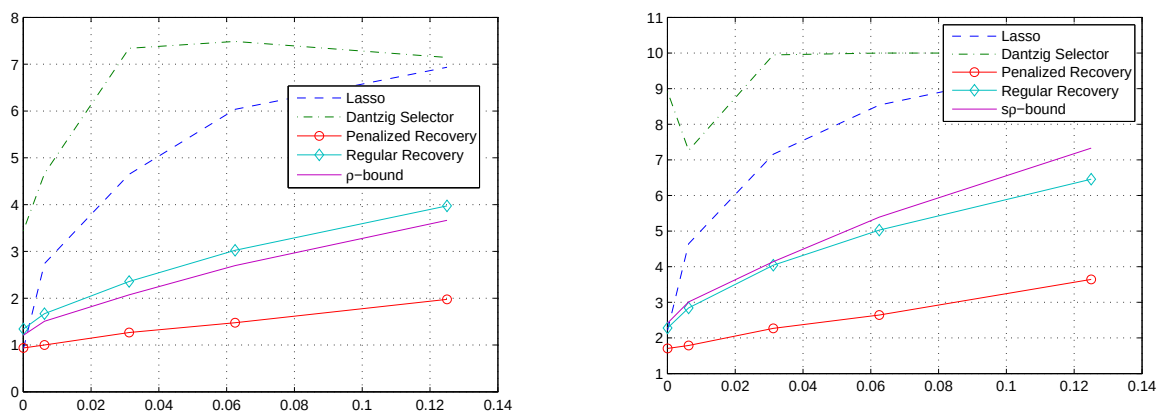
The Matching Pursuit algorithm for sparse recovery is motivated by the desire to provide a reduced complexity alternative to the algorithms using ℓ_1-minimization. Several implementations of Matching Pursuit has been proposed in the Compressive Sensing literature (see, e.g., [11, 10, 12]). They are based on successive Euclidean projections of the signal and the corresponding performance results rely upon the bounds on mutual incoherence parameter $\mu(A)$ of the sensing matrix. We are about to show how the construction of Section 3 can be used to design a specific version of the Matching Pursuit algorithm which we refer to as *Non-Euclidean Matching Pursuit (NEMP) algorithm*. The NEMP algorithm can be an interesting option if the ℓ_1-recovery is to be used repeatedly on the observations obtained with the same sensing matrix A; the numerical complexity of the pursuit algorithm for a given matrix A may only be a fraction of that of the recovery, especially when used on high-dimensional data.

Suppose that we have in our disposal $\bar{\gamma} \geq 0$ such that the condition $\mathbf{H}(\bar{\gamma}[1; ..., 1])$ is feasible; invoking



ℓ_∞-error $\qquad$ ℓ_1-error

Figure 3: Mean recovery error as a function of the number s of nonzero entries in the signal. Gaussian setup parameters: $L = 0.01$, $\sigma = 0.1$, $\bar{\gamma} = 0.1$, $\|x\|_1 = 5s$.



ℓ_∞-error $\qquad$ ℓ_1-error

Figure 4: Mean recovery error as a function of the nuisance magnitude L. Convolution setup parameters: $\sigma = 0.1$, $s = 2$, $\bar{\gamma} = 0.2$, $\|x\|_1 = 10$.

Lemma 3, in this case we can find efficiently a contrast matrix $H = [h_1, ..., h_n]$ such that

$$|[I - H^T A]_{ij}| \leq \bar{\gamma}, \;\; \nu(H) = \omega_*(\bar{\gamma}), \tag{46}$$

where, as always, $\nu(H) = \max_i \nu(h_i)$ with $\nu(h) = \nu_{\epsilon,\sigma,\mathcal{U}}(h)$ given by (4).

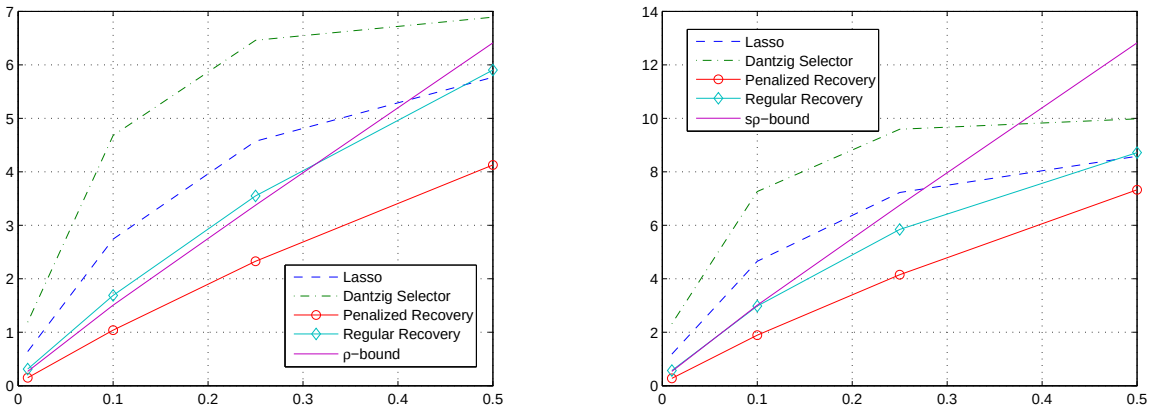
Consider a signal $x \in \mathbf{R}^n$ such that $\|x - x^s\|_1 \leq \upsilon$, where, as usual, x^s is the vector obtained from x by replacing all but the s largest in magnitude entries in x with zeros, and let y be an observation as in (3).

Suppose that $s\bar{\gamma} < 1$, and let $\upsilon \geq 0$ be given. Consider the following iterative procedure:

Algorithm 1

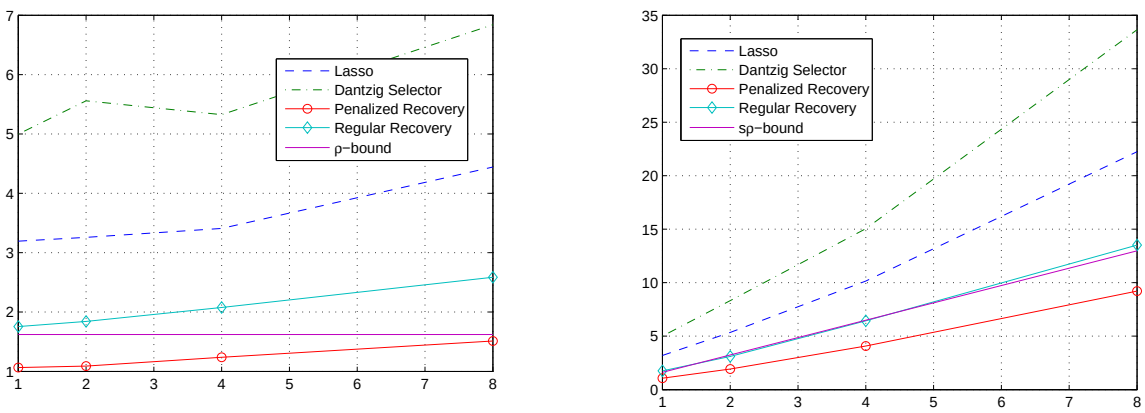
1. *Initialization:* *Set* $v^{(0)} = 0$,
$$\alpha_0 = \frac{\|H^T y\|_{s,1} + s\nu(H) + \upsilon}{1 - s\bar{\gamma}}.$$

2. *Step* k, $k = 1, 2, ...$: *Given* $v^{(k-1)} \in \mathbf{R}^n$ *and* $\alpha_{k-1} \geq 0$, *compute*



ℓ_∞-error $\qquad$ ℓ_1-error

Figure 5: Mean recovery error as a function of the noise StD σ. Convolution setup parameters: $L = 0.01$, $s = 2$, $\bar{\gamma} = 0.2$, $\|x\|_1 = 10$.



ℓ_∞-error $\qquad$ ℓ_1-error

Figure 6: Mean recovery error as a function of the number s of nonzero entries in the signal. Convolution setup parameters: $L = 0.01$, $\sigma = 0.1$, $\bar{\gamma} = 0.2$, $\|x\|_1 = 5s$.

(a) $u = H^T(y - Av^{(k-1)})$ *and vector* $\Delta \in \mathbf{R}^n$ *with the entries*

$$\Delta_i = \mathrm{sign}(u_i)[|u_i| - \bar{\gamma}\alpha_{k-1} - \nu(H)]_+, \quad 1 \le i \le n$$

(here $[a]_+ = \max[0, a]$*).*

(b) Set $v^{(k)} = v^{(k-1)} + \Delta$ *and*

$$\alpha_k = 2s\bar{\gamma}\alpha_{k-1} + 2s\nu(H) + \upsilon. \tag{47}$$

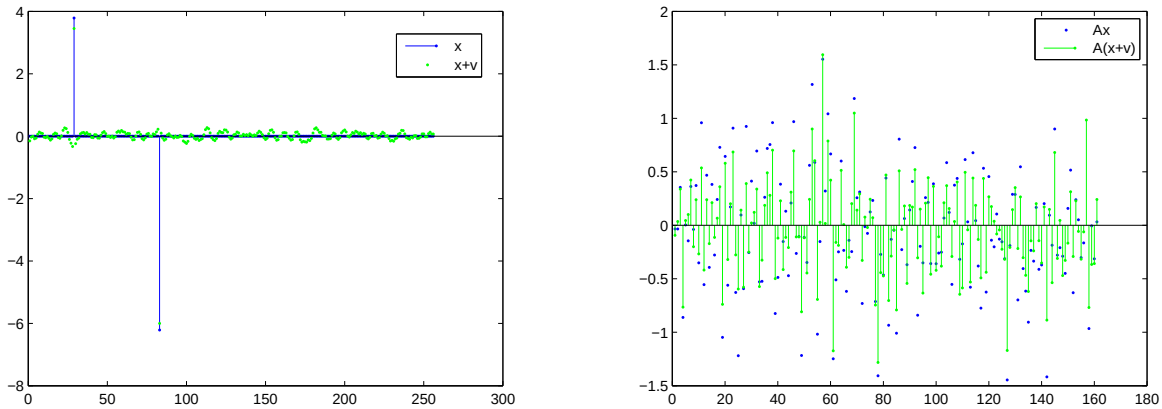
and loop to step $k + 1$*.*

3. The approximate solution found after k *iterations is* $v^{(k)}$*.*

Proposition 9 *Assume that* $s\bar{\gamma} < 1$ *and an* $\upsilon \ge 0$ *is given. Then there exists a set* $\Xi \subset \mathbf{R}^m$, $\mathrm{Prob}\{\xi \in \Xi\} \ge 1 - \epsilon$, *of "good" realizations of* ξ *such that whenever* $\xi \in \Xi$, *for every* $x \in \mathbf{R}^n$ *satisfying* $\|x - x^s\|_1 \le \upsilon$ *and every* $u \in \mathcal{U}$, *the approximate solution* $v^{(k)}$ *and the value* α_k *after the* k*-th step of Algorithm 1 satisfy*

(a_k) *for all* i $\;v_i^{(k)} \in \mathrm{Conv}\{0; x_i\}$

(b_k) $\;\|x - v^{(k)}\|_1 \le \alpha_k$ *and* $\|x - v^{(k+1)}\|_\infty \le 2\bar{\gamma}\alpha_k + 2\omega_*(\bar{\gamma})$.



signal x and nuisance v image Ax and $A(x+v)$

Figure 7: A typical signal/worst Lasso nuisance. Gaussian setup with parameters: $L = 0.05$, $\sigma = 0.1$, $s = 2$, $\|x\|_1 = 10$, $\bar{\gamma} = 0.1$.

Note that if $2s\bar{\gamma} < 1$ then also $s\bar{\gamma} < 1$ and Proposition 9 holds true. Furthermore, by (47) the sequence α_k converges exponentially fast to the limit $\alpha_\infty := \frac{2s\omega_*(\bar{\gamma})+\upsilon}{1-2s\bar{\gamma}}$:

$$\|v^{(k)} - x\|_1 \leq \alpha_k = (2s\bar{\gamma})^k[\alpha_0 - \alpha_\infty] + \alpha_\infty.$$

Along with the second inequality of (b_k) this implies the bounds:

$$\|v^{(k)} - x\|_\infty \leq 2\bar{\gamma}\alpha_{k-1} + 2\omega_*(\bar{\gamma}) \leq \frac{\alpha_k}{s},$$

and, since $\|x\|_p \leq \|x\|_1^{\frac{1}{p}}\|x\|_\infty^{\frac{p-1}{p}}$ for $1 \leq p \leq \infty$,

$$\|v^{(k)} - x\|_p \leq s^{\frac{1-p}{p}}\left((2s\bar{\gamma})^k[\alpha_0 - \alpha_\infty] + \alpha_\infty\right).$$

The bottom line here is as follows:

Corollary 2 *Let $\bar{\gamma} < 1/(2s)$ be such that the condition $\mathbf{H}(\bar{\gamma}[1;...;1])$ is feasible, so that we can find efficiently a contrast matrix H satisfying (46). With Algorithm 1 associated with H and some $\upsilon \geq 0$, one ensures that for every $t = 1, 2, ...$, the approximate solution $v^{(t)}$ found after t iterations satisfies*

$$\mathrm{Risk}_p(v^{(t)}|\epsilon, \sigma, s, \upsilon) \leq s^{\frac{1}{p}}\left(\frac{2\omega_*(\bar{\gamma})+s^{-1}\upsilon}{1-2s\bar{\gamma}} + (2s\bar{\gamma})^t\left[\frac{\omega_*(\bar{\gamma})+s^{-1}(\|H^Ty\|_{s,1}+\upsilon)}{1-s\bar{\gamma}} - \frac{2\omega_*(\bar{\gamma})+s^{-1}\upsilon}{1-2s\bar{\gamma}}\right]\right), \ 1 \leq p \leq \infty.$$

(cf. (29)).

To put this result in proper perspective, note that the mutual incoherence based condition

$$\frac{\mu(A)}{1+\mu(A)} < \frac{1}{2s}$$

underlying typical convergence results for the Matching Pursuit algorithms as applied to recovery of s-sparse signals (see, e.g. [11, 10, 12]) definitely is sufficient for convergence of the NEMP algorithm with $\bar{\gamma} = \frac{\mu(H)}{\mu(H)+1}$, see Section 4.1. It follows that the scope of NEMP is at least as wide as that of "theoretically valid" Matching Pursuit algorithms known from the literature; in the situation in question Corollary 2 recovers some results from [10, 11, 12].

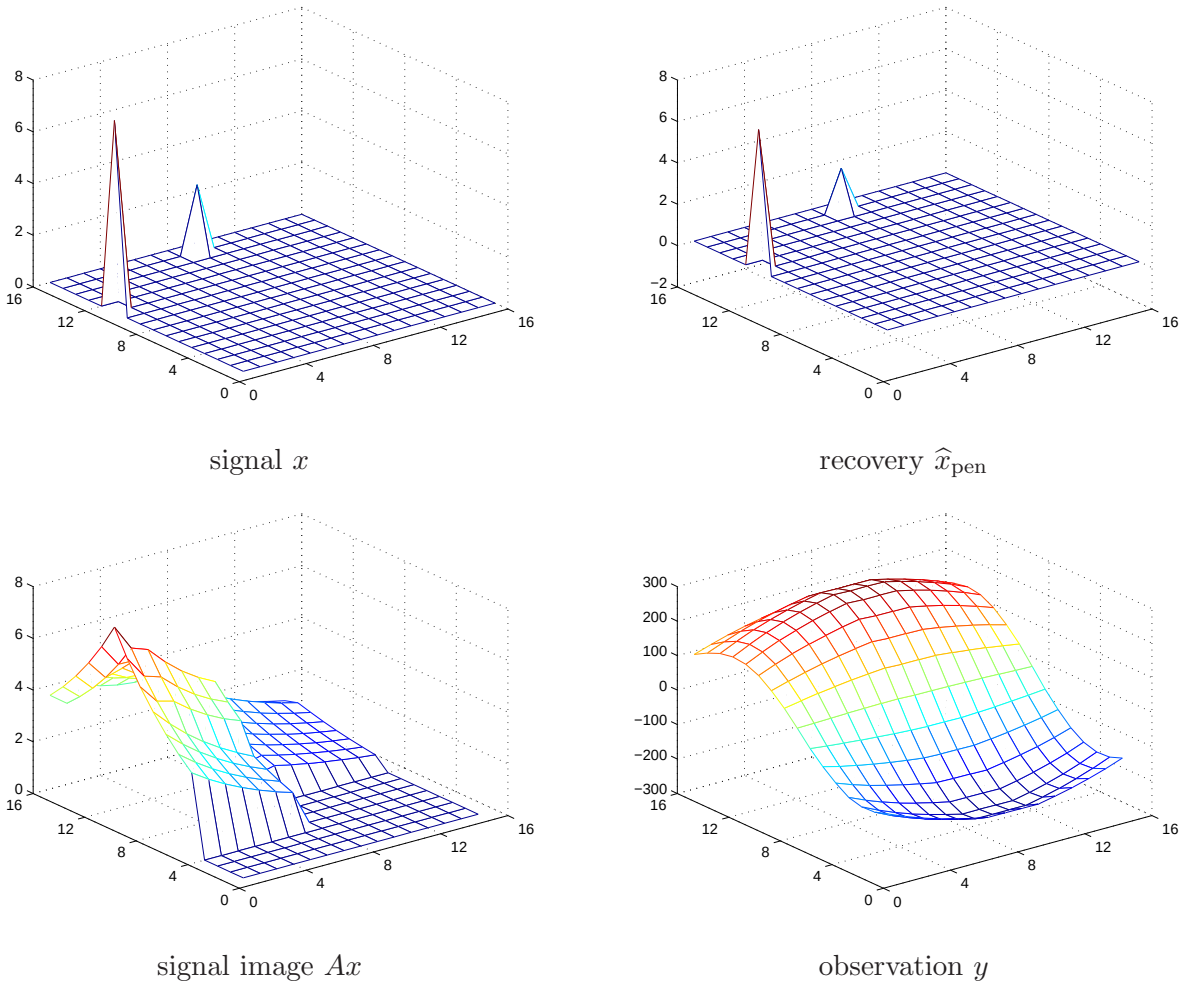


Figure 8: A typical signal/recovery in Convolution setup. Parameters: $L = 0.025$, $\sigma = 0.1$, $s = 2$, $\|x\|_1 = 10$, $\bar{\gamma} = 0.2$.

References

[1] Andersen, E. D., Andersen, K. D. *The MOSEK optimization tools manual. Version 5.0* http://www.mosek.com/fileadmin/products/5_0/tools/doc/html/tools/index.html

[2] Bickel, P., Ritov, Y. and Tsybakov, A. B. Simultaneous analysis of Lasso and Dantzig selector. *Ann. Statist.*, **37**, 1705-1732 (2009).

[3] Bühlmann, P., van de Geer, S. On the conditions used to prove oracle results for the Lasso *Electron. J. Statist.*, **3**, 1360-1392 (2009).

[4] Bunea, F., Tsybakov, A.B. and Wegkamp, M.H. Sparsity oracle inequalities for the Lasso. *Electron. J. Stat.*, **1**, 169-194 (2007).

[5] Candès, E., Romberg, J., Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory,* **52** 489-509 (2006).

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

[7] Candès, E. and Tao, T. The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.*, **35**, 2313-2351 (2007).

[8] Candès, E. J. The restricted isometry property and its implications for compressed sensing. *Comptes Rendus de l'Acad. des Sci.*, Serie I, 346, 589592 (2008).

[9] Donoho, D., Statistical estimation and optimal recovery. *The Annals of Statistics* **22**, 1, 238-270 (1995).

[10] Donoho, D., Elad, M. Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization. *Proc. Natl. Acad. Sci. USA*, **100**, 2197-2202 (2003).

[11] Elad, E., Bruckstein, A.M. A generalized uncertainty principle and sparse representation in pairs of $\mathbf{R}^n$ bases. *IEEE Trans. Inform. Theory*, **48**, 2558-2567 (2002).

[12] Gribonval, R., Nielsen, R. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory*, **49**, 3320-3325 (2003).

[13] Juditsky, A. B., Nemirovski A.S. Nonparametric estimation by convex programming. *Ann. Statist.*, **37**, 5a, 2278-2300 (2009).

[14] Juditsky, A., Nemirovski, A. On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization. *Mathematical Programming*, **127:1**, 57-88 (2011).

[15] Juditsky, A., Kilinç Karzan, F., Nemirovski, A. Verifiable conditions of ℓ_1-recovery of sparse signals with sign restrictions. *Mathematical Programming*, **127:1**, 89-122 (2011).

[16] Koltchinskii, V. The Dantzig selector and sparsity oracle inequalities. *Bernoulli*, **15** 799-828 (2009).

[17] Lounici, K. Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. *Electronic Journal of Statistics*, **2**, 90-102 (2008).

[18] Meinshausen, N. and Yu, B. Lasso-type recovery of sparse representations for high-dimensional data. *Annals of Statistics*, **37**, 246-270 (2009).

[19] van de Geer, S. The deterministic Lasso. In JSM proceedings, American Statistical Association (2007) (see also http://stat.ethz.ch/research/research_reports/2007/140).

[20] Zhang, C.-H., Huang, J. The sparsity and bias of the Lasso selection in highdimensional linear regression. *Ann. Statist.*, **36**, 1567-1594 (2008).

[21] Zhang, T. Some sharp performance bounds for least squares regression with L_1 regularization. *Ann. Statist.*, **37**, 2109-2144 (2009).

A Proofs

A.1 Proofs for section 2

A.1.1 Proof or Lemma 1

The first claim is evident. Now let $H = [h_1, ..., h_n]$ satisfy $\mathbf{H}_{s,\infty}(\kappa)$, and let $h_{n+i} = -h_i$, $1 \le i \le n$. Then for every $i \le n$ and every $x \in \mathbf{R}^n$ with $\|x\|_1 \le 1$ we have

$$\frac{\kappa}{s}\|H^T Ax\|_\infty = \frac{\kappa}{s} + \max_{\lambda \in \mathbf{R}^{2n}_+ : \sum_i \lambda_i = 1} \left[\sum\nolimits_{i=1}^{2n} \lambda_i h_i\right]^T Ax \ge x_i,$$

or, which is the same,

$$\min_{x:\|x\|_1\leq 1}\max_{\lambda\in\mathbf{R}^{2n}_+:\sum_i\lambda_i=1}\left[\frac{\kappa}{s}+\left[\sum_{i=1}^{2n}\lambda_i h_i\right]^T Ax-x_i\right]\geq 0.$$

By von Neumann lemma, this is the same as

$$\max_{\lambda\in\mathbf{R}^{2n}_+:\sum_i\lambda_i=1}\min_{x:\|x\|_1\leq 1}\left[\frac{\kappa}{s}+\left[\sum_{i=1}^{2n}\lambda_i h_i\right]^T Ax-x_i\right]\geq 0,$$

and the outer max clearly is achieved, meaning that there exists $\lambda^i\geq 0$, $\sum_{j=1}^{2n}\lambda^i_j=1$, such that with $h'_i=\sum_{j=1}^{2n}\lambda^i_j h_j$ one has

$$\frac{\kappa}{s}+[h'_i]^T Ax-x_i\geq 0\ \forall x:\|x\|_1\leq 1,$$

so that for every x with $\|x\|_1\leq 1$ one has $\frac{\kappa}{s}+|[h'_i]^T Ax|-x_i\geq 0$; applying the latter inequality to $-x$ in the role of x, we get $\frac{\kappa}{s}+|[h'_i]^T Ax|\geq |x_i|$ whenever $\|x\|_1\leq 1$, whence, of course, $\frac{\kappa}{s}\|x\|_1+|[h'_i]^T Ax|\geq |x_i|$ for all x. We conclude that the matrix $H'=[h'_1,...,h'_n]$ satisfies $\mathbf{H}(\frac{\kappa}{s}[1;...;1])$. It remains to note that by construction the columns of H' are convex combinations of the columns of H and $-H$, and that building H' reduces to solving n matrix games and thus can be carried out efficiently. ∎

A.1.2 Proof of Proposition 1

Let

$$\Xi=\{\xi:\ |h_i^T y|\leq\sqrt{2\ln(n/\epsilon)}\|h_i\|_2\ 1\leq i\leq n\},$$

so that $\mathrm{Prob}\{\xi\in\Xi\}\geq 1-\epsilon$. Let us fix $\xi\in\Xi$, a set $I=\{1,...,n\}\backslash J\subset\{1,...,n\}$ satisfying (9), a signal $x\in\mathbf{R}^n$ and a realization $u\in\mathcal{U}$ of the nuisance, and let $\widehat{x}$ be the value of the estimate (7) at the observation $y=Ax+u+\sigma\xi$. We are about to verify that $\widehat{x}$ satisfies (10), which, of course, will complete the proof.

Observe that because of $\xi\in\Xi$ we have

$$|h_i^T(Ax-y)|=|h_i^T(u+\sigma\xi)|\leq\max_{u'\in\mathcal{U}}|h_i^T u'|+\sigma\sqrt{2\ln(n/\epsilon)}\|h_i\|_2=\nu(h_i)=\nu_i,\ \ 1\leq i\leq n.$$

Now, $\rho_i\geq\nu_i$ by (8), whence $|h_i^T(y-Ax)|\leq\rho_i$ for all i, and thus x is a feasible solution to the optimization problem in (7) and thus $\|x\|_1\geq\|\widehat{x}\|_1$. Setting $z=\widehat{x}-x$, we now have $\|\widehat{x}_J\|_1=\|\widehat{x}\|_1-\|\widehat{x}_I\|_1\leq\|x\|_1-\|\widehat{x}_I\|_1\leq\|x\|_1-\|x_I\|_1+\|z_I\|_1=\|x_J\|_1+\|z_I\|_1$, whence $\|z_J\|_1\leq\|\widehat{x}_J\|_1+\|x_J\|_1\leq 2\|x_J\|_1+\|z_I\|_1$. It follows that

$$\|z\|_1\leq 2\|z_I\|_1+2\|x_J\|. \tag{48}$$

Further, $|h_i^T A(\widehat{x}-x)|\leq|h_i^T(A\widehat{x}-y)|+|h_i^T(Ax-y)|$. Since $\widehat{x}$ is feasible for the optimization problem in (7), we have $|h_i^T(A\widehat{x}-y)|\leq\rho_i$, and we have already seen that $|h_i^T(Ax-y)|\leq\nu_i$, hence

$$|h_i^T Az|\leq\rho_i+\nu_i \tag{49}$$

for all $1\leq i\leq n$. Applying (5) we now get

$$\begin{aligned}\|z_I\|_1 &= \sum_{i\in I}|z_i|\leq\sum_{i\in I}[|h_i^T Az|+\gamma_i\|z\|_1]\leq\sum_{i\in I}(\rho_i+\nu_i)+\left[\sum_{i\in I}\gamma_i\right][\|z_I\|_1+\|z_J\|_1]\\ &= \rho_I+\nu_I+\gamma_I\left[\|z_I\|_1+\|z_J\|_1\right]\leq\rho_I+\nu_I+2\gamma_I\left[\|z_I\|_1+\|x_J\|_1\right],\end{aligned}$$

where the concluding $\leq$ is given by (48). Taking into account that $\gamma_I<\frac{1}{2}$, we get

$$\|z_I\|_1\leq\frac{\rho_I+\nu_I+2\gamma_I\|x_J\|_1}{1-2\gamma_I}.$$

Invoking (48) once again, we finally get

$$\|z\|_1 = \|z_I\|_1 + \|z_J\|_1 \leq 2\|z_I\|_1 + 2\|x_J\|_1 \leq 2\frac{\rho_I + \nu_I + 2\gamma_I\|x_J\|_1}{1 - 2\gamma_I} + 2\|x_J\|_1,$$

and we arrive at (10.a).

To prove (10.b), we apply (5) to $z = \widehat{x} - x$, thus getting

$$|z_i| \leq |h_i^T Az| + \gamma_i\|z\|_1.$$

As we have already seen, $|h_i^T Az| \leq \rho_i + \nu_i$, and the first "$\leq$" in (10.$b$) follows; the second "$\leq$" in (10.$b$) is then readily given by (10.a). Now (ii) and (iii) are immediate consequences of (10) and the fact that $\widehat{\gamma}_s \leq s\widehat{\gamma}$. ∎

A.1.3 Proof of Lemma 2

In what follows, we use the notation from Proposition 1. For $x \in X(s, \upsilon)$, denoting by I the support of x^s, we have

$$\|x_J\|_1 \leq \upsilon,\ \rho_I \leq \widehat{\rho}_s \leq s\widehat{\rho},\ \nu_I \leq \widehat{\nu}_s \leq s\nu(H),\ \gamma_I \leq \widehat{\nu}_s \leq s\widehat{\gamma}.$$

Assuming $\widehat{\gamma}_s < \frac{1}{2}$, for $\xi \in \Xi$ (which happens with probability $\geq 1 - \epsilon$), (12) implies that for all $u \in \mathcal{U}$ it holds

$$\|\widehat{x}_{\rm reg}(y) - x\|_1 \leq \underbrace{\frac{2}{1 - 2\widehat{\gamma}_s}[\upsilon + \widehat{\rho}_s + \widehat{\nu}_s]}_{P}, \quad \text{and} \quad \|\widehat{x}_{\rm reg}(y) - x\|_\infty \leq \underbrace{\widehat{\rho} + \nu(H) + 2\widehat{\gamma}\frac{\upsilon + \widehat{\rho}_s + \widehat{\nu}_s}{1 - 2\widehat{\gamma}_s}}_{Q},$$

which combines with the standard bound $\|z\|_p \leq \|z\|_1^{\frac{1}{p}}\|z\|_\infty^{\frac{p-1}{p}}$ to imply (13). When $s\widehat{\gamma} < \frac{1}{2}$, we clearly have

$$P \leq \frac{2}{1 - 2s\widehat{\gamma}}[\upsilon + s(\widehat{\rho} + \nu(H))],\ Q \leq \widehat{\rho} + \nu(H) + \frac{2\widehat{\gamma}}{1 - 2s\widehat{\gamma}}[\upsilon + s(\widehat{\rho} + \nu(H))] = \frac{2}{1 - 2s\widehat{\gamma}}[\widehat{\gamma}\upsilon + \tfrac{1}{2}[\widehat{\rho} + \nu(H)]],$$

and (14) follows due to $\|\widehat{x}_{\rm reg}(y) - x\|_p = P^{\frac{1}{p}}Q^{\frac{p-1}{p}}$. ∎

A.1.4 Proof of Proposition 2

The proof is obtained by minor modifications from the one of Proposition 1. Same as in the latter proof, let $\Xi = \{\xi :\ |h_i^T y| \leq \sqrt{2\ln(n/\epsilon)}\|h_i\|_2\ 1 \leq i \leq n\}$, where h_i are the columns of H, so that Prob$\{\xi \in \Xi\} \geq 1 - \epsilon$. Let us fix $\xi \in \Xi$, $x \in \mathbf{R}^n$, $u \in \mathcal{U}$, let $\eta = \sigma\xi + u$, $y = Ax + \eta$, $\widehat{x} = \widehat{x}_{\rm reg}(y)$, $z = \widehat{x} - x$. Finally, let I be the support of x^s.

Due to $\xi \in \Xi$, we have

$$|h_i^T(Ax - y)| = |h_i^T(u + \sigma\xi)| \leq \max_{u' \in \mathcal{U}}|h_i^T u'| + \sigma\sqrt{2\ln(n/\epsilon)}\|h_i\|_2 = \nu(h_i) = \nu_i,\ \ 1 \leq i \leq n,$$

whence, by (8), x is a feasible solution to the optimization problem in (7) and thus $\|x\|_1 \geq \|\widehat{x}\|_1$. The latter, exactly as in the proof of Proposition (1), implies the validity of (48):

$$\|z\|_1 \leq 2\|z_I\|1 + 2\|x_J\|_1. \tag{50}$$

Besides this, the same reasoning as in the proof of Proposition 1 results in (49), whence

$$\|H^T Az\|_\infty \leq \widehat{\rho} + \nu(H). \tag{51}$$

Applying (6) to z, we get

$$\|z_I\|_1 \leq s\|z\|_\infty \leq s\|H^T Az\|_\infty + \kappa\|z\|_1 \leq s(\widehat{\rho} + \nu(H)) + \kappa\|z\|_1,$$

which combines with (50) to imply that

$$\|z\|_1 \leq \frac{1}{1-2\kappa}\left[2s(\widehat{\rho} + \nu(H)) + 2\|x_J\|_1\right], \tag{52}$$

which is nothing but the first relation in (15). Applying to z (6) once again, we get

$$\|z\|_\infty \leq \|H^T Az\|_\infty + s^{-1}\kappa\|z\|_1,$$

which combines with (52) to imply the second relation in (15). Relation (15) combines with the Moment inequality to imply (16). ∎

A.1.5 Proof of Proposition 3

(i): Given σ, ϵ, let, same as in the proof of Proposition 1, $\Xi = \{\xi : |h_i^T\xi| \leq \sqrt{2\ln(n/\epsilon)}\|h_i\|_2,\ 1 \leq i \leq n\}$, so that $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$. Let us fix $\xi \in \Xi$, $u \in \mathcal{U}$ and a signal $x \in \mathbf{R}^n$, and let us prove that for these data (21) takes place; this clearly will prove (i). Let us set $y = Ax + \sigma\xi + u$, $\widehat{x} = \widehat{x}_{\mathrm{pen}}(y)$, $z = \widehat{x} - x$, $\eta = u + \sigma\xi$. Let also I be the support of x^s.

Observe that by the origin of $\widehat{x}$, we have

$$\|\widehat{x}\|_1 + s\theta\|H^T(A\widehat{x} - y)\|_\infty \leq \|x\|_1 + s\theta\|H^T(Ax - y)\|_\infty = \|x\|_1 + s\theta\|H^T\eta\|_\infty, \tag{53}$$

and

$$\|H^T(A\widehat{x} - y)\|_\infty = \|H^T(Az + Ax - y)\|_\infty \geq \|H^T Az\|_\infty - \|H^T(Ax - y)\|_\infty = \|H^T Az\|_\infty - \|H^T\eta\|_\infty.$$

Combining the resulting inequality with (53), we get

$$\|\widehat{x}\|_1 + s\theta\|H^T Az\|_\infty \leq \|x\|_1 + 2s\theta\|H^T\eta\|_\infty \leq \|x\|_1 + 2s\theta\nu(H), \tag{54}$$

where the concluding $\leq$ is due to $\xi \in \Xi$ combined with (20). Further,

$$\|\widehat{x}\|_1 = \|x + z\|_1 = \|x_I + z_I\|_1 + \|x_J + z_J\|_1 \geq \|x_I\|_1 - \|z_I\|_1 + \|z_J\|_1 - \|x_J\|_1,$$

which combines with (54) to imply that

$$\|x_I\|_1 - \|z_I\|_1 + \|z_J\|_1 - \|x_J\|_1 + s\theta\|H^T Az\|_\infty \leq \|x\|_1 + 2s\theta\nu(H),$$

or, which is the same,

$$\|z_J\|_1 - \|z_I\|_1 + s\theta\|H^T Az\|_\infty \leq 2\|x_J\|_1 + 2s\theta\nu(H). \tag{55}$$

By (5), we have

$$\forall i : |z_i| \leq \|H^T Az\|_\infty + \gamma_i\|z\|_1, \tag{56}$$

whence $\|z_I\|_1 \leq s\|H^T Az\|_\infty + \widehat{\gamma}_s\|z\|_1$ and therefore

$$(1 - \widehat{\gamma}_s)\|z_I\|_1 - \widehat{\gamma}_s\|z_J\|_1 - s\|H^T Az\|_\infty \leq 0.$$

Multiplying the latter inequality by θ and summing up with (55), we get

$$[\theta(1 - \widehat{\gamma}_s) - 1]\|z_I\|_1 + (1 - \theta\widehat{\gamma}_s)\|z_J\|_1 \leq 2\|x_J\|_1 + 2s\theta\nu(H).$$

In view of condition (19), the coefficients in the left hand side are positive, and (21.a) follows.

To prove (21.b), note that from (54) it follows that

$$\|H^T Az\|_\infty \leq \frac{1}{s\theta}[\|x\|_1 - \|\widehat{x}\|_1] + 2\nu(H) \leq \frac{1}{s\theta}\|z\|_1 + 2\nu(H),$$

which combines with (56) to imply that

$$\|z\|_\infty \leq \frac{1}{s\theta}\|z\|_1 + 2\nu(H) + \widehat{\gamma}\|z\|_1.$$

Recalling that $z = \widehat{x} - x$ and invoking (21.a), (21.b) follows.

(ii)–(iii): (22) is an immediate consequence of (21) due to $\widehat{\gamma}_s \leq s\widehat{\gamma}$. Assuming that $x \in X(s, \upsilon)$ and $\theta = 2$ and taking into account that $\widehat{\gamma}_s \leq s\widehat{\gamma}$, we obtain from (22) that uniformly on $\xi \in \Xi$ and $u \in \mathcal{U}$

$$\|\bar{x}_{\rm reg}(y) - x\|_1 \leq \underbrace{\left[\frac{2\upsilon + 4s\nu(H)}{1 - 2s\widehat{\gamma}}\right]}_{P}, \quad \|\bar{x}_{\rm reg}(y) - x\|_\infty \leq \underbrace{\left[\frac{(s^{-1} + 2\widehat{\gamma})\upsilon + 4\nu(H)}{1 - 2s\widehat{\gamma}}\right]}_{Q}.$$

Using, as in the proof of Lemma 2, the standard bound

$$\|z\|_p \leq \|z\|_1^{\frac{1}{p}} \|z\|_\infty^{\frac{p-1}{p}} \leq P^{\frac{1}{p}} Q^{\frac{p-1}{p}}$$

we come to (23). ∎

A.1.6 Proof of Proposition 4

The proof is obtained by minor modifications from the one of Proposition 1. Same as in the latter proof, let $\Xi = \{\xi : |h_i^T \xi| \leq \sqrt{2\ln(n/\epsilon)}\|h_i\|_2,\ 1 \leq i \leq n\}$, so that $\mathrm{Prob}\{\xi \in \Xi\} \geq 1 - \epsilon$.

Let us fix $\xi \in \Xi$, $u \in \mathcal{U}$ and a signal $x \in \mathbf{R}^n$. Let us set $y = Ax + \sigma\xi + u$, $\widehat{x} = \widehat{x}_{\rm pen}(y)$, $z = \widehat{x} - x$, $\eta = u + \sigma\xi$. Let also I be the support of x^s.

Observe that by the origin of $\widehat{x}$ and due to $\theta = 2$ we have

$$\|\widehat{x}\|_1 + s\theta\|H^T(A\widehat{x} - y)\|_\infty \leq \|x\|_1 + s\theta\|H^T(Ax - y)\|_\infty = \|x\|_1 + 2s\|H^T\eta\|_\infty, \tag{57}$$

and

$$\|H^T(A\widehat{x} - y)\|_\infty = \|H^T(Az + Ax - y)\|_\infty \geq \|H^T Az\|_\infty - \|H^T(Ax - y)\|_\infty = \|H^T Az\|_\infty - \|H^T\eta\|_\infty.$$

Combining the resulting inequality with (57), we get

$$\|\widehat{x}\|_1 + 2s\|H^T Az\|_\infty \leq \|x\|_1 + 4s\|H^T\eta\|_\infty \leq \|x\|_1 + 4s\nu(H), \tag{58}$$

where the concluding $\leq$ is due to $\xi \in \Xi$ combined with the definition of $\nu(H)$. Further,

$$\|\widehat{x}\|_1 = \|x + z\|_1 = \|x_I + z_I\|_1 + \|x_J + z_J\|_1 \geq \|x_I\|_1 - \|z_I\|_1 + \|z_J\|_1 - \|x_J\|_1,$$

which combines with (58) to imply that

$$\|x_I\|_1 - \|z_I\|_1 + \|z_J\|_1 - \|x_J\|_1 + 2s\|H^T Az\|_\infty \leq \|x\|_1 + 4s\nu(H),$$

or, which is the same,

$$\|z_J\|_1 - \|z_I\|_1 + 2s\|H^T Az\|_\infty \leq 2\|x_J\|_1 + 4s\nu(H). \tag{59}$$

By (6) we have

$$\|z\|_\infty \leq \|H^T Az\|_\infty + \frac{\kappa}{s}\|z\|_1, \tag{60}$$

whence $\|z_I\|_1 \leq s\|H^T Az\|_\infty + \kappa\|z\|_1$ and therefore

$$(1-\kappa)\|z_I\|_1 - \kappa\|z_J\|_1 - s\|H^T Az\|_\infty \leq 0.$$

Multiplying the latter inequality by 2 and summing up with (59), we get

$$(1-2\kappa)\|z\|_1 \leq 2\|x_J\|_1 + 4s\nu(H),$$

and the first relation in (24). The second relation in (24) is readily given by the first one combined with (6). We have proved that (24) holds true whenever $\xi \in \Xi$; since $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$, (25) follows. ∎

A.2 Proofs for sections 3, 4

A.2.1 Proof of Lemma 3

(i)⇒(iii): If h_i satisfies $(\mathcal{P}_i)$, then for every x we have

$$|x_i| \leq |h_i^T Ax| + \gamma\|x\|_1 \leq \omega\varphi_*(Ax) + \gamma\|x\|_1,$$

where the first and the second inequalities are given by $(\mathcal{P}_i.b)$ and $(\mathcal{P}_i.a)$, respectively. □

(iii)⇒(ii): Assume that (iii) takes place; then, by homogeneity, $\omega\varphi_*(Ax)+\gamma \geq x_i$ for every x with $\|x\|_1 \leq 1$, or, which is the same, the optimal value in the conic problem

$$\min_x \{\omega\varphi_*(Ax) - x_i : \|x\|_1 \leq 1\}$$

is $\geq -\gamma$. The problem clearly is strictly feasible and bounded, so that by Conic Duality Theorem the dual problem is solvable with the same optimal value. Now, the dual problem reads

$$\max_{g,h,s}\{-s : \ \varphi(h) \leq \omega, \ \|g\|_\infty \leq s, \ A^T h + g = e_i\},$$

and the fact that it is solvable with the optimal value $\geq -\gamma$ means that there exist h, g such that $\varphi(h) \leq \omega$, $\|g\|_\infty \leq \gamma$ and $A^T h + g = e_i$, whence h is a feasible solution to (P_i^γ) with the value of the objective $\leq \omega$. □

(ii)⇒(i): If (P_i^γ) is feasible, it clearly is solvable; thus, in the case of (ii) there exists h with $\varphi(h) \leq \omega$ and $\|A^T h - e_i\|_\infty \leq \gamma$. From the latter inequality it follows that $|e_i^T x - h^T Ax| \leq \gamma\|x\|_1$ for every x, so that $|x_i| - |h^T Ax| \leq \gamma\|x\|_1$ for all x. We see that h satisfies $(\mathcal{P}_i)$, and thus (i) takes place. This reasoning shows also that whenever (P_i^γ) is feasible with optimal value $\leq \omega$, it is solvable, and its optimal solution satisfies $(\mathcal{P}_i)$. ∎

A.2.2 Proof of Proposition 5

Let $\gamma = \gamma(\delta, k)$, $\lambda = r(\mathcal{U}) + \sigma\sqrt{2\ln(n/\epsilon)}$, so that what we need to prove is that there exists a matrix H satisfying $\mathbf{H}_{s,\infty}(\gamma)$ and such that $\nu(H) \leq \lambda$. Invoking Lemma 3, all we need to this end is to show that

$$\forall x \in \mathbf{R}^n : \|x\|_\infty \leq \frac{1}{\sqrt{1-\delta}}\lambda\nu_*(Ax) + \gamma\|x\|_1 \tag{61}$$

Now, we clearly have $\nu(h) \leq \max_{u\in\mathcal{U}} u^T h + \sigma\sqrt{2\ln(n/\epsilon)}\|h\|_2 \leq \lambda\|h\|_2$ for all h, whence $\phi_*(\eta) \geq \lambda^{-1}\|\eta\|_2$ for all η. Therefore all we need in order to justify (61) is to prove that

$$\forall s \in \mathbf{R}^n : \|x\|_\infty \leq \frac{1}{\sqrt{1-\delta}}\|Ax\|_2 + \gamma\|x\|_1. \tag{62}$$

Let $x \in \mathbf{R}^n$. Setting $s = \text{floor}(k/2)$, let vectors $x^1, ..., x^q$ be obtained from x by the procedure as follows: x^1 is obtained by zeroing all but the s largest in magnitude entries of x; x^2 is obtained by the same procedure from $x - x^1$, x^3 is obtained by the same procedure from $x - x^1 - x^2$, and so on, until the step q where we get $x = x^1 + ... + x^q$. We clearly have $\|x^j\|_\infty \leq s^{-1}\|x^{j-1}\|_1$, $2 \leq j \leq q$, whence also $\|x^j\|_2 \leq s^{-1/2}\|x^{j-1}\|_1$, $2 \leq j \leq q$, since the vectors x^j are s-sparse. Setting $\|Ax\|_2 = \alpha$ and $\|Ax^1\|_2 = \beta$, we have

$$\alpha\beta = \|Ax\|_2\|Ax^1\|_2 \geq (Ax)^T Ax^1 = [x^1]^T A^T Ax^1 + \sum_{j=2}^{q} [x^1]^T A^T Ax^j \geq \beta^2 - \sum_{j=2}^{q} \delta\|x^1\|_2\|x^j\|_2,$$

where the last $\geq$ is given by the following well-known fact: [8]:

> (!) *If A is* RIP(δ, k) *and u, v are supported on a common set of indices I of cardinality k and are orthogonal, then $|u^T A^T Av| \leq \delta\|u\|_2\|v\|_2$.*

It follows that

$$\alpha\beta \geq \beta^2 - \delta\|x^1\|_2 \sum_{j=2}^{q} \|x^j\|_2 \geq \beta^2 - \delta s^{-1/2}\|x^1\|_2 \sum_{j=2}^{q} \|x^{j-1}\|_1 \geq \beta^2 - \delta s^{-1/2}\|x^1\|_2\|x\|_1.$$

Hence

$$\beta \leq \alpha + \frac{\delta\|x\|_1\|x^1\|_2}{\sqrt{s}\beta} \leq \alpha + \frac{\delta\|x\|_1}{\sqrt{s(1-\delta)}},$$

where the second inequality is due to the fact that $\|x^1\|_2/\beta \leq 1/\sqrt{1-\delta}$ by RIP. Thus,

$$\|x\|_\infty \leq \|x^1\|_2 \leq \frac{\beta}{\sqrt{1-\delta}} \leq \frac{\alpha}{\sqrt{1-\delta}} + \frac{\delta\|x\|_1}{(1-\delta)\sqrt{s}} \leq \frac{\alpha}{\sqrt{1-\delta}} + \gamma\|x\|_1,$$

where the concluding inequality is due to $s \geq (k-1)/2$ and $\gamma = \gamma(\delta, k)$. Recalling that $\alpha = \|Ax\|_2$, (62) follows. ∎

A.2.3 Proof of Proposition 6

Proof. We start with analysis of $\mathbf{O}(S, \omega)$. Let $i \in \{1, ..., n\}$, and let $I \ni i$ be a subset of $\{1, ..., n\}$ of cardinality S. Let $\mathbf{R}^{(S)}$ be the linear space of all vectors from $\mathbf{R}^n$ supported on I, and let $X_R = \{x \in \mathbf{R}^{(S)} : \|x\|_2 \leq R\}$. Assume that we are given a noisy observation $y = Ax + u + \sigma\xi$ of a signal $z = (x, u) \in (X_R, \mathcal{U})$, and that we want to recover from this observation the linear form x_i of the signal. From $\mathbf{O}(S, \omega)$ it follows that there exists a recovering routine such that for every $x \in X_R$ and $u \in \mathcal{U}$ the probability of recovering error to be $\geq \omega$ is $\leq \epsilon$. Assuming $\epsilon \leq 1/16$ and applying the celebrated result of Donoho [9], there exists a *linear* estimate $\phi_R^T y$ such that for every $x \in X_R$ and $u \in \mathcal{U}$ the probability for the error of this estimate to be $\geq 1.22\omega$ is $\leq \epsilon$. Moreover (cf. Proposition 4.2 of [13]), one can pick ϕ_R such that

$$\forall x \in X_R,\ u \in \mathcal{U}, \quad \left\{ \begin{array}{lcll} \text{Prob}\{\phi_R^T[u + \sigma\xi + A_I x] - x_i > 1.22\omega\} & \leq & \epsilon/2, & (a) \\ \text{Prob}\{\phi_R^T[u + \sigma\xi + A_I x] - x_i < -1.22\omega\} & \leq & \epsilon/2, & (b) \end{array} \right.$$

where A_I is the matrix obtained from A by zeroing columns with indexes not belonging to I. Let $p(R) = \max_{u \in \mathcal{U}} |\phi_R^T u|$ and $r(R) = \|A_I^T \phi_R - e_i\|_2$, where e_i is the i-th basic orth (so that $x_i = e_i^T x$). Specifying $\bar{x}$ as the vector from X_R such that $\bar{x}^T(A_I^T \phi_R - e_i) = Rr(R)$, and $\bar{u}$ as the vector from $\mathcal{U}$ such that $\phi_R^T \bar{u} = p(R)$ (the required $\bar{x}$, $\bar{u}$ clearly exist) and applying (a) to the pair $(x, u) = (\bar{x}, \bar{u})$, and (b) to the pair $(x, u) = (-\bar{x}), -\bar{u})$, we get

$$\begin{array}{lcl} \text{Prob}\{\sigma\phi_R^T\xi > 1.22\omega - Rr(R) - p(R)\} & \leq & \epsilon/2, \\ \text{Prob}\{\sigma\phi_R^T\xi < -1.22\omega + Rr(R) + p(R)\} & \leq & \epsilon/2. \end{array}$$

Hence, denoting by $\mathrm{erfinv}(\epsilon)$ the value of the inverse error function at ϵ, we obtain

$$\mathrm{erfinv}\left(\frac{\epsilon}{2}\right)\sigma\|\phi_R\|_2 \leq 1.22\omega - Rr(R) - p(R).$$

It follows that as $R \to \infty$, ϕ_R remains bounded and $r(R) = \|e_i - A_I^T\phi_R\|_2 \to 0$. Thus, there exists a sequence $R_k \to +\infty$, of values of R such that ϕ_{R_k} goes to a limit ϕ as $k \to \infty$, and this limit satisfies the relations

$$\mathrm{erfinv}\left(\frac{\epsilon}{2}\right)\sigma\|\phi\|_2 \leq 1.22\omega, \quad \text{and } A_I^T\phi = e_i.$$

Taking into account that $\mathrm{erfinv}\left(\frac{\epsilon}{2}\right) \geq 0.92\sqrt{\ln(1/\epsilon)}$ when $\epsilon \leq 1/16$, we arrive at the following result:

Lemma 6 *Under assumption* $\mathbf{O}(S,\omega)$*, for every* $i \leq n$ *and every* S*-element subset* $I \ni i$ *of* $\{1,...,n\}$ *there exists* $\phi \in \mathbf{R}^m$ *such that* $\phi^T a_i = 1$*,* $\phi^T a_j = 0$ *for all* $j \in I$*,* $j \neq i$ *(here* $a_1,...,a_n$ *are the columns of* A*), and*

$$\max_{u\in\mathcal{U}} |u^T\phi| + \sigma\sqrt{\ln(1/\epsilon)}\|\phi\|_2 \leq \sqrt{2}\omega. \tag{63}$$

We claim that in this case for all $x \in \mathbf{R}^n$ it holds:

$$\|x\|_\infty \leq \bar{\omega}\nu_*(Ax) + \overbrace{\frac{\omega\|A\|}{\sigma\sqrt{2S\ln(1/\epsilon)}}}^{\widehat{\gamma}}\|x\|_1. \tag{64}$$

Taking this claim for granted, and invoking Lemma 3, we immediately arrive at the desired conclusion. Indeed, given s satisfying (32), we have $\frac{1}{4s} \geq \widehat{\gamma}$, so that (64) implies that

$$\forall x \in \mathbf{R}^n : \|x\|_\infty \leq \bar{\omega}\nu_*(Ax) + \frac{1}{4s}\|x\|_1,$$

whence, by Lemma 3, there exists H satisfying the condition $\mathbf{H}_{s,\infty}(\frac{1}{4})$ and such that $\nu(H) \leq \bar{\omega}$, which is exactly what Proposition 6 states.

It remains to prove (64). Let us fix $x \in \mathbf{R}^n$, and let I be set of indices of the S largest in magnitude entries in x. Denoting by J the complement of I in $\{1,...,n\}$, we have $\|x_J\|_\infty \leq S^{-1}\|x_I\|_1$, whence

$$\|x_J\|_2 \leq \|x_J\|_\infty^{1/2}\|x_J\|_1^{1/2} \leq S^{-1/2}\|x_I\|_1^{1/2}\|x_J\|_1^{1/2} \leq \frac{1}{2}S^{-1/2}\|x\|_1. \tag{65}$$

Let $i_* \in I$ be the index of the largest in magnitude entry of x. By Lemma 6 there exists $\phi \in \mathbf{R}^m$ satisfying (63) and such that $\phi^T a_{i_*} = \mathrm{sign}(x_{i_*})$, $\phi^T a_i = 0$ for $i \in I\backslash\{i_*\}$. We have

$$\begin{array}{l} \nu(\phi) \equiv \nu_{\epsilon,\sigma,\mathcal{U}}(\phi) = \max\limits_{u\in\mathcal{U}} u^T\phi + \sigma\sqrt{2\ln(n/\epsilon)}\|\phi\|_2 \\ \leq \sqrt{2[1+\ln(n)/\ln(1/\epsilon)]}\left[\max\limits_{u\in\mathcal{U}} u^T\phi + \sigma\sqrt{\ln(1/\epsilon)}\|\phi\|_2\right] \leq 2\sqrt{1+\ln(n)/\ln(1/\epsilon)}\omega, \end{array} \tag{66}$$

where the concluding $\leq$ is given by (63). Now,

$$\begin{array}{rcl} \nu(\phi)\nu_*(Ax) & \geq & \phi^T Ax = \phi^T Ax_I + \phi^T Ax_J = |x_{i_*}| + \phi^T Ax_J = \|x\|_\infty + \phi^T Ax_J \\ & \geq & \|x\|_\infty - \|\phi\|_2\|Ax_J\|_2 \geq \|x\|_\infty - \|\phi\|_2\|A\|\|x_J\|_2 \geq \|x\|_\infty - \frac{1}{2}\|A\|S^{-1/2}\|\phi\|_2\|x\|_1, \end{array}$$

with the concluding $\leq$ given by (65). The resulting inequality, in view of (66) and the bound $\|\phi\|_2 \leq \frac{\sqrt{2}\omega}{\sigma\sqrt{\ln(1/\epsilon)}}$ given by (63) implies (64). ∎

A.3 Proofs for section 5

A.3.1 Proof of Lemma 4

Recall (cf., e.g., Theorem 2.1 in [14]) that a necessary and sufficient condition for an $m \times n$ matrix A to be s-good is the *nullspace property* as follows: there exists $\kappa < 1/2$ such that

$$\|x\|_{s,1} \leq \kappa \|x\|_1 \ \forall (x \in \mathbf{R}^n, Ax = 0). \tag{67}$$

Assume that this condition is satisfied, and let $u \in \mathbf{R}^n$ be a vector with s nonzero coordinates, equal to ± 1. (67) says that the optimal value in the Linear Programming problem

$$\max_x \left\{ u^T x : Ax = 0, \|x\|_1 \leq 1 \right\}$$

is at most κ; passing to the dual problem, we conclude that there exist h_u and g_u such that $A^T h_u + g_u = u$ and $\|g_u\|_\infty \leq \kappa$, whence for every $x \in \mathbf{R}^n$ it holds

$$u^T x = x^T A^T h_u + g_u^T x \leq \|h_u\|_1 \|Ax\|_\infty + \kappa \|x\|_1.$$

Since the set U of the outlined vectors u is finite, the quantity $L = \max_{u \in U} \|h_u\|_1$ is finite, and

$$\|x\|_{s,1} = \max_{u \in U} u^T x \leq L \|Ax\|_\infty + \kappa \|x\|_1 \ \forall x,$$

meaning that the condition $\mathbf{H}_{s,1}(\kappa)$ holds true for $H = [LI_m, 0_{m \times n-m}]$. Vice versa, the existence of $\kappa < 1/2$ and H satisfying $\mathbf{H}_{s,q}(\kappa)$ clearly implies the validity of (67) with the same κ and this implies the s-goodness of A. ∎

A.3.2 Proof of Lemma 5

Let A satisfy RIP$(\delta, 2s)$ with $\delta < 1$; we want to prove that then the matrix $\frac{1}{1-\delta} A$ satisfies the condition $\mathbf{H}_{s,2}(\frac{\delta}{1-\delta})$. Indeed, let $x \in \mathbf{R}^n$. Let vectors x^1, x^2, ..., x^q be obtained from x as follows: x^1 is obtained by zeroing all but the s largest in magnitude entries of x and keeping the latter entries intact, then x^2 is obtained by applying the same procedure to $x - x^1$, and so on. We stop at step q where we get $x = x^1 + ... + x^q$. Observe that $\|x^j\|_\infty \leq \frac{1}{s} \|x^{j-1}\|_1$, whence also $\|x^j\|_2 \leq \|x^{j-1}\|_1 s^{-1/2}$ (since x^j is s-sparse). We now have

$$\begin{aligned}
\sqrt{s} \|x^1\|_2 \|A^T A x\|_\infty &\geq \|x^1\|_1 \|A^T A x\|_\infty \geq [x^1]^T A^T A x = [x^1]^T A^T A x^1 + \sum_{j=2}^q [x^1]^T A^T A x^j \\
&\geq (1-\delta) \|x^1\|_2^2 - \delta \sum_{j=2}^q \|x^1\|_2 \|x^j\|_2 \qquad (*) \\
&\geq (1-\delta) \|x^2\|_2^2 - \delta s^{-1/2} \sum_{j=2}^q \|x^1\|_2 \|x^{j-1}\|_1 \geq (1-\delta) \|x^2\|_2^2 - \delta s^{-1/2} \|x^1\|_2 \|x\|_1 \\
\Rightarrow \quad \|x\|_{s,2} &= \|x^1\|_2 \leq \frac{1}{1-\delta} \|A^T A x\|_\infty + \frac{\delta}{1-\delta} \|x\|_1
\end{aligned}$$

(in the above chain, step $(*)$ is valid due to $[x^1]^T A^T A x^1 \geq (1-\delta) \|x^1\|_2$ (since A is RIP$(\delta, 2s)$) and the statement (!), see the proof of Proposition 5). The concluding inequality in the chain says that $\frac{1}{1-\delta} A$ satisfies $\mathbf{H}_{s,2}(\frac{\delta}{1-\delta})$. ∎

A.3.3 Proof of Proposition 7

We present here the proof of (i), which is a straightforward modification of the proof of Proposition 2. The proof of (ii) can be obtained by equally straightforward modification of the proof of Proposition 4.

Thus, suppose we are under the premise of (i), and let Ξ be defined exactly as in the proof of Proposition 1, so that $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$ and $|(\sigma\xi+u)^T h_i| \leq \nu_{\epsilon,\sigma,\mathcal{U}}(h_i) \leq \nu(H)$ for all $\xi \in \Xi$, $u \in \mathcal{U}$ and all i. Let us fix $x \in \mathbf{R}^n$, $\xi \in \Xi$ and $u \in \mathcal{U}$, let I be the set of indices of the s largest in magnitude entries in x, and let $\eta = \sigma\xi + u$, $y = Ax + \eta$, $\widehat{x} = \widehat{x}_{\mathrm{reg}}(y)$, and $z = \widehat{x} - x$.

Since $\xi \in \Xi$ and $u \in \mathcal{U}$, we have $|h_i^T(Ax-y)| \leq \nu_i \leq \rho_i$ for all i, whence x is a feasible solution to the optimization problem defining $\widehat{x}$, whence, exactly as in the proof of Proposition 1,

$$\begin{array}{lll} (a) & \|z\|_1 & \leq 2\|z_I\|_1 + 2\|x_J\|_1, \\ (b) & |h_i^T Az| & \leq \rho_i + \nu_i,\ 1 \leq i \leq n \\ \Rightarrow & \|H^T Az\|_\infty & \leq \widehat{\rho} + \widehat{\omega}. \end{array} \tag{68}$$

Now, H satisfies the condition $\mathbf{H}_{s,q}(\kappa)$ and thus satisfies the condition $\mathbf{H}_{s,1}(\kappa)$. Applying the latter condition, we get

$$\|z_I\|_1 \leq s\|H^T Az\|_\infty + \kappa\|z\|_1.$$

Invoking (68), we conclude that

$$\|z\|_1 \leq 2s\|H^T Az\|_\infty + 2\kappa\|z\|_1 + 2\|x_J\|_1 \leq 2s[\widehat{\rho} + \widehat{\omega} + s^{-1}\|x_J\|_1] + 2\kappa\|z\|_1, \tag{69}$$

thus

$$\|z\|_1 \leq \frac{2s}{1-2\kappa}[\widehat{\rho} + \widehat{\omega} + s^{-1}\|x_J\|_1]. \tag{70}$$

Next, H satisfies $\mathbf{H}_{s,q}(\kappa)$, whence $\|z\|_{s,q} \leq \|H^T Az\|_\infty + \kappa\|z\|_1$. Therefore, we get from (68):

$$\|z\|_{s,q} \leq s^{\frac{1}{q}}\|H^T Az\|_\infty + \kappa s^{\frac{1}{q}-1}\|z\|_1 \leq s^{\frac{1}{q}}[\widehat{\rho} + \widehat{\omega}] + 2\kappa s^{\frac{1}{q}}\frac{\widehat{\rho} + \widehat{\omega} + s^{-1}\|x_J\|_1}{1-2\kappa} \tag{71}$$

$$\leq s^{\frac{1}{q}}\frac{\widehat{\rho} + \widehat{\omega} + 2\kappa s^{-1}\|x_J\|_1}{1-2\kappa} \leq s^{\frac{1}{q}}\frac{\widehat{\rho} + \widehat{\omega} + s^{-1}\|x_J\|_1}{1-2\kappa}. \tag{72}$$

All we need in order to extract (i) from (69) and (72) is to verify that

$$1 \leq p \leq q \ \Rightarrow\ \|z\|_p \leq (3s)^{1/p}\theta, \quad \theta = \frac{\widehat{\rho} + \widehat{\omega} + s^{-1}\|x_J\|_1}{1-2\kappa}.$$

The desired inequality holds true when $p = 1$ (see (69)), thus, invoking the Hölder inequality, all we need is to verify that

$$\|z\|_q \leq (3s)^{1/q}\theta. \tag{73}$$

When $q = \infty$, (73) is implied by (72), so let us assume that $q < \infty$. Let λ be the $(s+1)$-st largest of the magnitudes of entries in z. By (72) we have $\lambda^q s \leq \|z\|_{s,q}^q \leq s\theta^q$, and $\lambda \leq \theta$. Hence, setting $z' = z - z^s$, we get

$$\|z'\|_q^q \leq \lambda^{q-1}\|z'\|_1 \leq \theta^{q-1}\|z\|_1 \leq \theta^{q-1}2s\theta,$$

where the concluding inequality is given by (69). Thus, $\|z'\|_q^q \leq 2s\theta^q$, while $\|z^s\|_q^q \leq s\theta^q$ by (72). We see that $\|z\|_q^q \leq \|z^s\|_q^q + \|z'\|_q^q \leq 3s\theta^q$, as required in (73). ∎

A.3.4 Proof of Proposition 8

(i): Let $\Xi = \{\xi \in \mathbf{R}^m : |\xi^T a_i| \le \varrho, 1 \le i \le n\}$, so that $\mathrm{Prob}\{\xi \in \Xi\} \ge 1-\epsilon$. Let us fix $\xi \in \Xi$ and $x \in \mathbf{R}^n$, and let $y = Ax + \sigma\xi$, $\widehat{x} = \widehat{x}_{\mathrm{DS}}(y)$. We have $\|A^T(Ax-y)\|_\infty = \|A^T\sigma\xi\|_\infty \le \varrho \le \rho$, so that x is a feasible solution to the optimization problem specifying $\widehat{x}_{\mathrm{DS}}(y)$ and therefore $\|\widehat{x}\|_1 \le \|x\|_1$. Denoting by I the support of x^s, setting $z = \widehat{x} - z$ and acting exactly as when deriving (48), we arrive at

$$\|z\|_1 \le 2\|z_I\|_1 + 2\|x_J\|_1. \tag{74}$$

Further,

$$\|A^TAz\|_\infty = \|A^T(A\widehat{x} - y + \sigma\xi)\|_\infty \le \|A^T\sigma\xi\|_\infty + \|A^T(A\widehat{x}-y)\|_\infty \le \rho + \varrho,$$

and therefore

$$\|Az\|_2^2 = z^TA^TAz \le \|z\|_1\|A^TAz\|_\infty \le (\rho+\varrho)\|z\|_1. \tag{75}$$

On the other hand, by (37) we have

$$\|z_I\|_1 \le \|z\|_{s,1} \le s^{1-\frac{1}{q}}\|z\|_{s,q} \le s\widehat{\lambda}\|Az\|_2 + \kappa\|z\|_1 \le s\widehat{\lambda}(\rho+\varrho)^{1/2}\|z\|_1^{1/2} + \kappa\|z\|_1.$$

Substituting the above bound into (74), we get

$$\|z\|_1 \le 2\kappa\|z\|_1 + 2s\widehat{\lambda}(\rho+\varrho)^{1/2}\|z\|_1^{1/2} + 2\|x_J\|_1,$$

whence by elementary calculations

$$\tau := \|z\|_1^{1/2} \le a + b^{1/2}, \quad \text{where } a = \frac{2s\widehat{\lambda}\sqrt{\rho+\varrho}}{1-2\kappa}, \ b = \frac{2\|x_J\|_1}{1-2\kappa}. \tag{76}$$

Invoking (37) and (75), we have

$$\begin{aligned}
\|z\|_{s,q} &\le s^{\frac{1}{q}}\widehat{\lambda}\|Az\|_2 + \kappa s^{\frac{1}{q}-1}\|z\|_1 \le s^{\frac{1}{q}-1}\left[s\widehat{\lambda}\sqrt{\rho+\varrho}\|z\|_1^{1/2} + \kappa\|z\|_1\right] = s^{\frac{1}{q}-1}\left[\frac{(1-2\kappa)}{2}a\tau + \kappa\tau^2\right] \\
&\le s^{\frac{1}{q}-1}\left[\frac{1-2\kappa}{2}a^2 + \frac{1-2\kappa}{2}ab^{1/2} + \kappa[a+b^{1/2}]^2\right] \le s^{\frac{1}{q}-1}\left[\frac{a^2}{2} + \frac{1+2\kappa}{2}ab^{1/2} + \kappa b\right] \\
&\le \frac{s^{\frac{1}{q}-1}}{2}[a+b^{1/2}]^2,
\end{aligned} \tag{77}$$

where the last inequality of the chain is due to $\kappa < \frac{1}{2}$. Assuming for a moment that $1 < q < \infty$ and denoting by μ the $(s+1)$-st largest magnitude of entries in z, we conclude from the latter inequality that $\mu \le \frac{1}{2s}[a+b^{1/2}]^2$. Hence, when setting $z' = z - z^s$ we obtain (cf. the verification of (73)) $\|z'\|_q^q \le \mu^{q-1}\|z'\|_1 \le (2s)^{1-q}[a+b^{1/2}]^{2q}$. Invoking (77) one more time we get

$$\|z\|_q^q \le \|z^s\|_q^q + \|z'\|_q^q \le \frac{3}{2}(2s)^{1-q}[a+b^{1/2}]^{2q}.$$

The resulting inequality combines with (76) and the Hölder inequality to imply that

$$\|z\|_p \le \frac{3^{\frac{1}{p}}s^{\frac{1}{p}-1}}{2}[a+b^{1/2}]^2 \le 3^{\frac{1}{p}}s^{\frac{1}{p}-1}[a^2+b], \quad 1 \le p \le q. \tag{78}$$

Note that the derivation of (78) was carried out under the additional assumption that $1 < q < \infty$. This assumption can now be removed: when $q = 1$, (78) is readily given by (76). When $q = \infty$, A satisfies (37) for $q = \infty$ and thus – for every value of q from $[1, \infty]$, meaning that (78) holds true for every $q < \infty$, whence (78) holds true for $q = \infty$ as well.

Recalling that relation (78) is valid whenever $\xi \in \Xi$ and $x \in \mathbf{R}^n$ and plugging in the values of a and b, we arrive at (40). (i) is proved.

(ii): Same as above, let $\Xi = \{\xi : |a_i^T\xi| \le \sqrt{2\ln(n/\epsilon)}\|a_i\|_2,\ 1 \le i \le n\}$, so that $\mathrm{Prob}\{\xi \notin \Xi\} \le \epsilon$. Let us fix $x \in \mathbf{R}^n$, $\xi \in \Xi$, and let I be the support of x^s. Let also $y = Ax + \sigma\xi$, $\widehat{x} = \widehat{x}_{\mathrm{lasso}}(y)$, $z = \widehat{x} - x$. We have

$$\|\widehat{x}\|_1 + \varkappa\|A\widehat{x} - y\|_2^2 \le \|x\|_1 + \varkappa\|Ax - y\|_2^2 \le \|x\|_1 + \sigma^2\varkappa\xi^T\xi,$$

or, which is the same due to $A\widehat{x} - y = Az - \sigma\xi$,

$$\|\widehat{x}\|_1 + \varkappa\|Az\|_2^2 - 2\sigma\varkappa\xi^T Az \le \|x\|_1.$$

It follows that

$$\begin{array}{rcl} 0 & \ge & \|\widehat{x}\|_1 - \|x\|_1 + \varkappa(\|Az\|_2^2 - 2\sigma\xi^T Az) \\ & = & (\|x_I + z_I\|_1 - \|x_I\|_1) + (\|x_J + z_J\|_1 - \|x_J\|_1) + \varkappa(\|Az\|_2^2 - 2\sigma\xi^T Az) \\ & \ge & -\|z_I\|_1 + \|z_J\|_1 - 2\|x_J\|_1 + \varkappa(\|Az\|_2^2 - 2\sigma\xi^T Az) \\ & \ge & -\|z_I\|_1 + \|z_J\|_1 - 2\|x_J\|_1 - 2\varkappa\varrho\|z\|_1 + \varkappa\|Az\|_2^2, \end{array}$$

where the last $\ge$ is readily given by the fact that $\|A^T\sigma\xi\|_\infty \le \varrho$ for $\xi \in \Xi$. We conclude that

$$\|z_J\|_1 \le \|z_I\|_1 + 2\varkappa\varrho\|z\|_1 - \varkappa\|Az\|_2^2 + 2\|x_J\|_1,$$

and therefore

$$\|z\|_1 \le 2\|z_I\|_1 + 2\varkappa\varrho\|z\|_1 - \varkappa\|Az\|_2^2 + 2\|x_J\|_1. \tag{79}$$

Now, we have

$$\|z_I\|_1 \le s^{1-\frac{1}{q}}\|z\|_{s,q} \le s\widehat{\lambda}\|Az\|_2 + \kappa\|z\|_1,$$

where the concluding inequality is given by (37). Combining the resulting inequality with (79), we get

$$\|z\|_1(1 - 2\kappa - 2\varkappa\varrho) \le 2s\widehat{\lambda}\|Az\|_2 - \varkappa\|Az\|_2^2 + 2\|x_J\|_1.$$

Combining this inequality with (79), we get the first inequality in the following chain:

$$\begin{array}{rcl} \|z\|_1 & \le & 2(\kappa + \varkappa\varrho)\|z\|_1 + 2\|x_J\|_1 + \varkappa\left(2\|Az\|_2\dfrac{s\widehat{\lambda}}{\varkappa} - \|Az\|_2^2\right) \\ & \le & 2(\kappa + \varkappa\varrho)\|z\|_1 + 2\|x_J\|_1 + \dfrac{s^2\widehat{\lambda}^2}{\varkappa} \end{array} \tag{80}$$

and since $2\kappa + 2\varkappa\varrho < 1$, we arrive at

$$\|z\|_1 \le a := \frac{1}{1 - 2(\kappa + \varkappa\varrho)}\left[\frac{s^2\widehat{\lambda}^2}{\varkappa} + 2\|x_J\|_1\right]. \tag{81}$$

Since $2(\kappa + \varrho\varkappa) < 1$, the first inequality in (80) is possible only if

$$\|Az\|_2^2 - \frac{2s\widehat{\lambda}}{\varkappa}\|Az\|_2 - \frac{2\|x_J\|_1}{\varkappa} \le 0,$$

whence

$$\|Az\|_2 \le \frac{2s\widehat{\lambda}}{\varkappa} + \frac{\|x_J\|_1}{s\widehat{\lambda}}. \tag{82}$$

Invoking (37), we get $\|z\|_{s,q} \le s^{\frac{1}{q}-1}\left[s\widehat{\lambda}\|Az\|_2 + \kappa\|z\|_1\right]$, which combines with (82) and (81) to imply that

$$\|z\|_{s,q} \le 2s^{\frac{1}{q}-1}a. \tag{83}$$

Denoting by μ the $(s+1)$-st largest of the magnitudes of entries in z, we conclude from (83) that $\mu \leq 2s^{-1}a$, whence, denoting $z' = z - z^s$,

$$\|z'\|_q \leq \mu^{\frac{q-1}{q}} \|z'\|_1^{\frac{1}{q}} \leq (2s^{-1}a)^{\frac{q-1}{q}} \|z\|_1^{\frac{1}{q}} \leq 2^{\frac{q-1}{q}} s^{\frac{1}{q}-1} a,$$

(we have used (81)), which combines with (83) to imply that

$$\|z\|_q \leq 4s^{\frac{1}{q}-1} a. \tag{84}$$

Combining (84), (81) and the Hölder inequality, we get

$$1 \leq p \leq q \Rightarrow \|z\|_p \leq 4s^{\frac{1}{p}-1} a. \tag{85}$$

Plugging in the value of a (see (81)) and recalling that (85) takes place whenever $\xi \in \Xi$ with $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$, we arrive at (41). ∎

A.4 Proof of Proposition 9

The proof below follows the lines of the proof of Proposition 7 of [15]. Given $\epsilon \in (0,1)$, let $\Xi = \{\xi : |h_i^T \xi| \leq \sqrt{2\ln(n/\epsilon)} \|h_i\|_2,\ 1 \leq i \leq n\}$, so that $\mathrm{Prob}\{\xi \in \Xi\} \geq 1-\epsilon$. Let us fix $\xi \in \Xi$, $x \in \mathbf{R}^n$ such that $\|x - x^s\|_1 \leq \upsilon$, and $u \in \mathcal{U}$. For $\eta = y - Ax = u + \sigma\xi$, by the definition (4) of the norm ν and because of $\nu(h_i) \leq \nu(H)$, we have $\|H^T\eta\|_\infty \leq \nu(H) \leq \omega_*(\bar\gamma)$.

We intend to prove the relations (a_k), (b_k) by induction in k. First, let us show that (a_{k-1}, b_{k-1}) implies (a_k, b_k). Thus, assume that (a_{k-1}, b_{k-1}) holds true. Let $z^{(k-1)} = x - v^{(k-1)}$. By (a_{k-1}), $z^{(k-1)}$ is supported on the support of x. Note that

$$\begin{array}{rcl} z^{(k-1)} - u & = & x - v^{(k-1)} - H^T(y - Av^{(k-1)}) = (I - H^TA)(x - v^{(k-1)}) - H^T\eta \\ & = & (I - H^TA)z^{(k-1)} - H^T\eta, \end{array}$$

Then by (46) for any $1 \leq i \leq n$,

$$-\bar\gamma \sum_j |z_j^{(k-1)}| - \omega_*(\bar\gamma) \leq z_i^{(k-1)} - u_i \leq \bar\gamma \sum_j z_j^{(k-1)} + \omega_*(\bar\gamma),$$

consequently,

$$-\gamma := -\bar\gamma\alpha_{k-1} - \omega_*(\bar\gamma) \leq z_i^{(k-1)} - u_i \leq \bar\gamma\alpha_{k-1} + \omega_*(\bar\gamma) := \gamma, \tag{86}$$

so that the segment $S_i = [u_i - \gamma,\ u_i + \gamma]$ of the width $\ell = 2\bar\gamma\alpha_{k-1} + 2\omega_*(\bar\gamma)$, covers $z_i^{(k-1)}$, and the closest to zero point of this interval is

$$\widetilde{\Delta}_i = \begin{cases} [u_i - \gamma]_+, & u_i \geq 0, \\ -[|u_i| - \gamma]_+, & u_i < 0, \end{cases}$$

that is, $\widetilde{\Delta}_i = \Delta_i$ for all i. Since the segment S_i covers $z_i^{(k-1)}$ and Δ_i is the closest to 0 point in S_i, while the width of S_i is at most ℓ, we clearly have

$$(a) \quad \Delta_i \in \mathrm{Conv}\left\{0, z_i^{(k-1)}\right\}, \quad (b) \quad |z_i^{(k-1)} - \Delta_i| \leq \ell. \tag{87}$$

Since (a_{k-1}) is valid, (87.a) implies that

$$v_i^{(k)} = v_i^{(k-1)} + \Delta_i \in \left[v_i^{(k-1)} + \mathrm{Conv}\left\{0, x_i - v_i^{(k-1)}\right\}\right] \subseteq \mathrm{Conv}\{0, x_i\},$$

and (a_k) holds. Further, let I be the support of x^s. Relation (a_k) clearly implies that $|z_i^{(k)}| \leq |x_i|$, and we can write due to (87.b):

$$\|x - v^{(k)}\|_1 = \sum_{i\in I} |x_i - [v_i^{(k-1)} + \Delta_i]| + \sum_{i\notin I} |z_i^{(k)}| \leq \sum_{i\in I} |z_i^{(k-1)} - \Delta_i| + \sum_{i\notin I} |x_i| \leq s\ell + \upsilon = \alpha_k.$$

Since by (87.b)

$$\|x - v^{(k)}\|_\infty \quad = \quad \|x - v^{(k-1)} - \Delta\|_\infty \leq \ell = 2\bar{\gamma}\alpha_{k-1} + 2\omega_*(\bar{\gamma}),$$

we conclude that (b_k) holds true. The induction step is justified.

It remains to show that (a_0, b_0) holds true. Since (a_0) is evident, all we need is to justify (b_0). Let

$$\alpha_* = \|x\|_1,$$

and let $u = H^T y$. Same as above (cf. (86)), we have for all i:

$$|x_i - u_i| \leq \bar{\gamma}\alpha_* + \omega_*(\bar{\gamma}).$$

Then

$$\alpha_* = \sum_{i\in I} |x_i| + \sum_{i\notin I} |x_i| \leq \sum_{i\in I} [|u_i| + \bar{\gamma}\alpha_* + \omega_*(\bar{\gamma})] + \upsilon \leq \|u\|_{s,1} + s\bar{\gamma}\alpha_* + s\omega_*(\bar{\gamma}) + \upsilon.$$

Hence

$$\alpha_* \leq \alpha_0 = \frac{\|u\|_{s,1} + s\omega_*(\bar{\gamma}) + \upsilon}{1 - s\bar{\gamma}},$$

which implies (b_0). ∎