The Annals of Statistics
2012, Vol. 40, No. 6, 3077–3107
DOI: 10.1214/12-AOS1057

ACCURACY GUARANTIES FOR ℓ_1 RECOVERY OF BLOCK-SPARSE SIGNALS

BY ANATOLI JUDITSKY, FATMA KILINÇ KARZAN[1], ARKADI NEMIROVSKI[1,2] AND BORIS POLYAK

Université J. Fourier de Grenoble, Carnegie Mellon University, Georgia Institute of Technology and Institute of Control Sciences

We introduce a general framework to handle structured models (sparse and block-sparse with possibly overlapping blocks). We discuss new methods for their recovery from incomplete observation, corrupted with deterministic and stochastic noise, using block-ℓ_1 regularization. While the current theory provides promising bounds for the recovery errors under a number of different, yet mostly hard to verify conditions, our emphasis is on verifiable conditions on the problem parameters (sensing matrix and the block structure) which guarantee accurate recovery. Verifiability of our conditions not only leads to efficiently computable bounds for the recovery error but also allows us to optimize these error bounds with respect to the method parameters, and therefore construct estimators with improved statistical properties. To justify our approach, we also provide an oracle inequality, which links the properties of the proposed recovery algorithms and the best estimation performance. Furthermore, utilizing these verifiable conditions, we develop a computationally cheap alternative to block-ℓ_1 minimization, the non-Euclidean Block Matching Pursuit algorithm. We close by presenting a numerical study to investigate the effect of different block regularizations and demonstrate the performance of the proposed recoveries.

1. Introduction.

The problem. Our goal in this paper is to estimate a linear transform $Bx \in \mathbb{R}^N$ of a vector $x \in \mathbb{R}^n$ from the observations

$$y = Ax + u + \xi. \tag{1.1}$$

Here A is a given $m \times n$ sensing matrix, B is a given $N \times n$ matrix, and $u + \xi$ is the observation error; in this error, u is an unknown *nuisance* known to belong to a given compact convex set $\mathcal{U} \subset \mathbb{R}^m$ symmetric w.r.t. the origin, and ξ is random noise with known distribution P.

Received November 2011; revised September 2012.
[1]Supported by the Office of Naval Research Grant N000140811104.
[2]Supported by the NSF Grant DMS-09-14785.
MSC2010 subject classifications. Primary 62G08, 62H12; secondary 90C90.
Key words and phrases. Sparse recovery, nonparametric estimation by convex optimization, oracle inequalities.

We assume that the space $\mathbb{R}^N$ where Bx lives is represented as $\mathbb{R}^N = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_K}$, so that a vector $w \in \mathbb{R}^N$ is a block vector: $w = [w[1]; \dots; w[K]]$ with blocks $w[k] \in \mathbb{R}^{n_k}$, $1 \le k \le K$.[3] In particular, $Bx = [B[1]x; \dots; B[K]x]$ with $n_k \times n$ matrices $B[k]$, $1 \le k \le K$. While we do not assume that the vector x is sparse in the usual sense, we do assume that the linear transform Bx to be estimated is *s-block sparse*, meaning that at most a given number, s, of the blocks $B[k]x$, $1 \le k \le K$, are nonzero.

The recovery routines we intend to consider are based on *block-ℓ_1 minimization*, that is, the estimate $\widehat{w}(y)$ of $w = Bx$ is $B\widehat{z}(y)$, where $\widehat{z}(y)$ is obtained by minimizing the norm $\sum_{k=1}^K \|B[k]z\|_{(k)}$ over signals $z \in \mathbb{R}^n$ with Az "fitting," in a certain precise sense, the observations y. Above, $\|\cdot\|_{(k)}$ are given in advance norms on the spaces $\mathbb{R}^{n_k}$ where the blocks of Bx take their values.

In the sequel we refer to the given in advance collection $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$ as the *representation structure* (r.s.). Given such a representation structure $\mathcal{S}$ and a sensing matrix A, our ultimate goal is to understand how well one can recover the s-block-sparse transform Bx by appropriately implementing block-ℓ_1 minimization.

Related Compressed Sensing research. Our situation and goal form a straightforward extension of the usual sparsity/block sparsity framework of Compressed Sensing. Indeed, the *standard representation structure* with $B = I_n$, $n_k = 1$, and $\|\cdot\|_{(k)} = |\cdot|$, $1 \le k \le K = n$, leads to the standard Compressed Sensing setting—recovering a sparse signal $x \in \mathbb{R}^n$ from its noisy observations (1.1) via ℓ_1 minimization. The case of nontrivial block structure $\{n_k, \|\cdot\|_{(k)}\}_{k=1}^K$ and $B = I_n$ is generally referred to as *block-sparse*, and has been considered in numerous recent papers. Block-sparsity (with $B = I_n$) arises naturally (see, e.g., [13] and references therein) in a number of applications such as multi-band signals, measurements of gene expression levels or estimation of multiple measurement vectors sharing a joint sparsity pattern. Several methods of estimation and selection extending the "plain" ℓ_1-minimization to block sparsity were proposed and investigated recently. Most of the related research focused so far on *block regularization schemes*—group Lasso recovery of the form

$$\widehat{x}(y) \in \operatorname*{Arg\,min}_{z=[z^1;\dots;z^K]\in\mathbb{R}^n=\mathbb{R}^{n_1}\times\cdots\times\mathbb{R}^{n_K}} \left\{ \|Az - y\|_2^2 + \lambda \sum_{k=1}^K \|z[k]\|_2 \right\}$$

(here $\|\cdot\|_2$ is the Euclidean norm of the block). In particular, the literature on "plain Lasso" (the case of $n_k = 1$, $1 \le k \le K = n$) has an important counterpart

[3] We use MATLAB notation: $[u, v, \dots, z]$ is the horizontal concatenation of matrices $u, v, \dots, z$ of common height, while $[u; v; \dots; z]$ is the vertical concatenation of matrices $u, v, \dots, z$ of common width. All vectors are column vectors.

on group Lasso; see, for example, [2, 4, 9, 11–15, 22, 25–28, 30, 32] and the references therein. Another celebrated technique of sparse recovery, the Dantzig selector, originating from [6], has also been extended to handle block-sparse structures [16, 23]. Most of the cited papers focus on bounding recovery errors in terms of the magnitude of the observation noise and "s-concentration" of the true signal x (the distance from the space of signals with at most s nonzero blocks—the sum of magnitudes $\|x[k]\|_2$ of all but the s largest in magnitude blocks in x). Typically, these results rely on natural block analogy ("Block RIP;" see, e.g., [13]) of the celebrated Restricted Isometry Property introduced by Candés and Tao [7, 8] or on block analogies [24] of the Restricted Eigenvalue Property introduced in [5]. In addition to the usual (block)-sparse recovery, our framework also allows to handle group sparse recovery with overlapping groups by properly defining the corresponding B matrix.

Contributions of this paper. The first (by itself, minor) novelty in our problem setting is the presence of the linear mapping B. We are not aware of any preceding work handling the case of a "nontrivial" (i.e., different from the identity) B. We qualify this novelty as minor, since in fact the case of a nontrivial B can be reduced to the one of $B = I_n$.[4] However, "can be reduced" is not the same as "should be reduced," since problems with nontrivial B mappings arise in many applications. This is the case, for example, when x is the solution of a linear finite-difference equation with a sparse right-hand side ("evolution of a linear plant corrected from time to time by impulse control"), where B is the matrix of the corresponding finite-difference operator. Therefore, introducing B adds some useful flexibility (and as a matter of fact costs nothing, as far as the theoretical analysis is concerned).

We believe, however, that the major novelty in what follows is the emphasis on *verifiable* conditions on matrix A and the r.s. $\mathcal{S}$ which *guarantee* good recovery of the transform Bx from noisy observations of Ax, provided that the transform in question is nearly s-block sparse, and the observation noise is low. Note that such efficiently verifiable guarantees cannot be obtained from the "classical" conditions[5] used when studying theoretical properties of block-sparse recovery (with a notable exception of the Mutual Block-Incoherence condition of [12]). For example, given A and $\mathcal{S}$, one cannot answer in any reasonable time if the (Block-) Restricted Isometry or Restricted Eigenvalue property holds with given parameters. While the efficient verifiability is by no means necessary for a condition to be

[4] Assuming, for example, that $x \mapsto Bx$ is an "onto" mapping, we can treat Bx as our signal, the observations being Py, where P is the projector onto the orthogonal complement to the linear subspace $A \cdot \operatorname{Ker} B$ in $\mathbb{R}^m$; with $y = Ax + u + \xi$, we have $Py = GBx + P(u + \xi)$ with an explicitly given matrix G.

[5] Note that it has been recently proved in [29] that computing the parameters involved in verification of Nullspace condition as well as RIP for sparse recovery is NP-hard.

meaningful and useful, we believe that verifiability has its value and is worthy of being investigated. In particular, it allows us to design new recovery routines with explicit confidence bounds for the recovery error and then optimize these bounds with respect to the method parameters. In this respect, the current work extends the results of [19–21], where ℓ_1 recovery of the "usual" sparse vectors was considered (in the first two papers—in the case of uncertain-but-bounded observation errors, and in the third—in the case of Gaussian observation noise). Specifically, we propose here new routines of block-sparse recovery which explicitly utilize a *contrast matrix*, a kind of "validity certificate," and show how these routines may be tuned to attain the best performance bounds. In addition to this, verifiable conditions pave the way of efficiently designing sensing matrices which possess certifiably good recovery properties for block-sparse recovery (see [17] for implementation of such an approach in the usual sparsity setting).

The main body of the paper is organized as follows: in Section 2 we formulate the block-sparse recovery problem and introduce our core assumption—a family of conditions $\mathbf{Q}_{s,q}$, $1 \le q \le \infty$, which links the representation structure $\mathcal{S}$ and sensing matrix $A \in \mathbb{R}^{m\times n}$ with a contrast matrix $H \in \mathbb{R}^{m\times M}$. Specifically, given s and $q \in [1,\infty]$ and a norm $\|\cdot\|$, the condition $\mathbf{Q}_{s,q}$ on an $m \times M$ *contrast matrix* H requires $\exists\kappa \in [0, 1/2)$ such that

$$\forall\big(x \in \mathbb{R}^n\big) \qquad L_{s,q}(Bx) \le s^{1/q}\big\|H^T Ax\big\| + \kappa s^{1/q-1} L_1(Bx)$$

holds, where for $w = [w[1]; \dots; w[K]] \in \mathbb{R}^N$ and $p \in [1,\infty]$,

$$L_p(w) = \big\|\big[\|w[1]\|_{(1)}; \dots; \|w[K]\|_{(K)}\big]\big\|_p$$

and

$$L_{s,p}(w) = \big\|\big[\|w[1]\|_{(1)}; \dots; \|w[K]\|_{(K)}\big]\big\|_{s,p},$$

where $\|u\|_{s,p}$ is the norm on $\mathbb{R}^K$ defined as follows: we zero out all but the s largest in magnitude entries in vector u, and take the $\|\cdot\|_p$-norm of the resulting s-sparse vector. Then, by restricting our attention to the standard representation structures, we study the relation between condition $\mathbf{Q}_{s,q}$ and the usual assumptions used to validate block-sparse recovery, for example, Restricted Isometry/Eigenvalue Properties and their block versions.

In Section 3 we introduce two recovery routines based on the $L_1(\cdot)$ norm:

- *regular ℓ_1 recovery* [cf. (block-) Dantzig selector]

$$\widehat{x}_{\text{reg}}(y) \in \operatorname*{Arg\,min}_{z\in\mathbb{R}^n}\big\{L_1(Bz) : \big\|H^T(y - Az)\big\|_\infty \le \rho\big\},$$

  where with probability $1-\varepsilon$, $\rho[=\rho(H,\varepsilon)]$ is an upper bound on the $\|\cdot\|$-norm of the observation error;

- *penalized ℓ_1 recovery* [*cf.* (*block-*) *Lasso*]

$$\widehat{x}_{\mathrm{pen}}(y) \in \operatorname*{Arg\,min}_{z\in\mathbb{R}^n}\big[L_1(Bz) + 2s\|H^T(y - Az)\|_\infty\big],$$

where s is our guess for the number of nonvanishing blocks in the true signal Bx.

Under condition $\mathbf{Q}_{s,q}$, we establish performance guarantees of these recoveries, that is, explicit upper bounds on the size of confidence sets for the recovery error $L_p(B(\widehat{x} - x))$, $1 \le p \le q$. Our performance guarantees have the usual natural interpretation—as far as recovery of transforms Bx with small s-block concentration[6] is concerned, everything is as if we were given the direct observations of Bx contaminated by noise of small L_∞ magnitude.

Similar to the usual assumptions from the literature, conditions $\mathbf{Q}_{s,q}$ are generally computationally intractable, nonetheless, we point out a notable exception in Section 4. When all block norms are $\|\cdot\|_{(k)} = \|\cdot\|_\infty$, the condition $\mathbf{Q}_{s,\infty}$, the strongest among our family of conditions, is efficiently verifiable. Besides, in this situation, the latter condition is "fully computationally tractable," meaning that one can optimize efficiently the bounds for the recovery error over the contrast matrices H satisfying $\mathbf{Q}_{s,\infty}$ to design optimal recovery routines. In addition to this, in Section 4.2, we establish an oracle inequality which shows that existence of the contrast matrix H satisfying condition $\mathbf{Q}_{s,\infty}$ is not only sufficient but also necessary for "good recovery" of block-sparse signals in the L_∞-norm when $\|\cdot\|_{(k)} = \|\cdot\|_\infty$.

In Section 5 we provide a verifiable sufficient condition for the validity of $\mathbf{Q}_{s,q}$ for general q, assuming that $\mathcal{S}$ is ℓ_r-r.s. [i.e., $\|\cdot\|_{(k)} = \|\cdot\|_r$, $1 \le k \le K$], and, in addition, $r \in \{1, 2, \infty\}$. This sufficient condition can be used to build a "quasi-optimal" contrast matrix H. We also relate this condition to the Mutual Block-Incoherence condition of [12] developed for the case of ℓ_2-r.s. with $B = I_n$. In particular, we show in Section 5.4 that the Mutual Block-Incoherence is more conservative than our verifiable condition, and thus is "covered" by the latter. "Limits of performance" of our verifiable sufficient conditions are investigated in Section 5.3.

In Section 6 we describe a computationally cheap alternative to block-ℓ_1 recoveries—a non-Euclidean Block Matching Pursuit (NEBMP) algorithm. Assuming that $\mathcal{S}$ is either ℓ_2-, or ℓ_∞-r.s. and that the verifiable sufficient condition $\mathbf{Q}_{s,\infty}$ is satisfied, we show that this algorithm (which does not require optimization) provides performance guarantees similar to those of regular/penalized ℓ_1 recoveries.

We close by presenting a small simulation study in Section 7.

Proofs of all results are given in the supplementary article [18].

---

[6] s-block concentration of a block vector w is defined as $L_1(w) - L_{s,1}(w)$.

2. Problem statement.

Notation. In the sequel, we deal with:

- *signals*—vectors $x = [x_1; \dots; x_n] \in \mathbb{R}^n$, and an $m \times n$ *sensing matrix* A;
- *representations of signals*—block vectors $w = [w[1]; \dots; w[K]] \in \mathcal{W} := \mathbb{R}^{n_1}_{w[1]} \times \cdots \times \mathbb{R}^{n_K}_{w[K]}$, and the *representation matrix* $B = [B[1]; \dots; B[K]]$, $B[k] \in \mathbb{R}^{n_k \times n}$; the representation of a signal $x \in \mathbb{R}^n$ is the block vector $w = Bx$ with the blocks $B[1]x, \dots, B[K]x$.

From now on, the dimension of $\mathcal{W}$ is denoted by N:

$$N = n_1 + \cdots + n_K.$$

The factors $\mathbb{R}^{n_k}$ of the representation space $\mathcal{W}$ are equipped with norms $\|\cdot\|_{(k)}$; the conjugate norms are denoted by $\|\cdot\|_{(k,*)}$. A vector $w = [w[1]; \dots; w[K]]$ from $\mathcal{W}$ is called *s-block-sparse*, if the number of nonzero blocks $w[k] \in \mathbb{R}^{n_k}$ in w is at most s. A vector $x \in \mathbb{R}^n$ will be called s-block-sparse, if its representation Bx is so. We refer to the collection $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$ as the *representation structure* (r.s. for short). The *standard* r.s. is given by $B = I_n$, $K = N$, $n_1 = \cdots = n_n = 1$ and $\|\cdot\|_{(k)} = |\cdot|$, $1 \le k \le N$, and *an ℓ_r-r.s.* is the r.s. with $\|\cdot\|_{(k)} = \|\cdot\|_r$, $1 \le k \le K$.

For $w \in \mathcal{W}$, we call the number $\|w[k]\|_{(k)}$ the *magnitude* of the kth block in w and denote by w^s the representation vector obtained from w by zeroing out all but the s largest in magnitude blocks in w (with the ties resolved arbitrarily). For $I \subset \{1, \dots, K\}$ and a representation vector w, w_I denotes the vector obtained from w by keeping intact the blocks $w[k]$ with $k \in I$ and zeroing out all remaining blocks. For $w \in \mathcal{W}$ and $1 \le p \le \infty$, we denote by $L_p(w)$ the $\|\cdot\|_p$-norm of the vector $[\|w[1]\|_{(1)}; \dots; \|w[K]\|_{(K)}]$, so that $L_p(\cdot)$ is a norm on $\mathcal{W}$ with the conjugate norm $L_p^*(w) = \|[\|w[1]\|_{(1,*)}; \dots; \|w[K]\|_{(K,*)}]\|_{p_*}$ where $p_* = \frac{p}{p-1}$. Given a positive integer $s \le K$, we set $L_{s,p}(w) = L_p(w^s)$. Note that $L_{s,p}(\cdot)$ is a norm on $\mathcal{W}$. We define the *s-block concentration* of a vector w as $\upsilon_s(w) = L_1(w - w^s)$.

Problem of interest. Given an observation

$$y = Ax + u + \xi, \tag{2.1}$$

of unknown signal $x \in \mathbb{R}^n$, we want to recover the representation Bx of x, knowing in advance that this representation is "nearly s-block-sparse," that is, the representation can be approximated by an s-block-sparse one; the L_1-error of this approximation, that is, the s-block concentration, $\upsilon_s(Bx)$, will be present in our error bounds.

In (2.1) the term $u + \xi$ is the observation error; in this error, u is an unknown *nuisance* known to belong to a given compact convex set $\mathcal{U} \subset \mathbb{R}^m$ symmetric w.r.t. the origin, and ξ is random noise with known distribution P.

Condition $\mathbf{Q}_{s,q}(\kappa)$. We start with introducing the condition which will be instrumental in all subsequent constructions and results. Let a sensing matrix A and an r.s. $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$ be given, and let $s \le K$ be a positive integer, $q \in [1, \infty]$ and $\kappa \ge 0$. We say that a pair $(H, \|\cdot\|)$, where $H \in \mathbb{R}^{m\times M}$ and $\|\cdot\|$ is a norm on $\mathbb{R}^M$, satisfies the condition $\mathbf{Q}_{s,q}(\kappa)$ associated with the matrix A and the r.s. $\mathcal{S}$, if

$$(2.2)\qquad \forall x \in \mathbb{R}^n \qquad L_{s,q}(Bx) \le s^{1/q}\|H^T Ax\| + \kappa s^{1/q-1} L_1(Bx).$$

The following observation is evident:

OBSERVATION 2.1. *Given A and an r.s. $\mathcal{S}$, let $(H, \|\cdot\|)$ satisfy $\mathbf{Q}_{s,q}(\kappa)$. Then $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,q'}(\kappa')$ for all $q' \in (1, q)$ and $\kappa' \ge \kappa$. Besides this, if $s' \le s$ is a positive integer, $((s/s')^{1/q} H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s',q}((s'/s)^{1-1/q}\kappa)$. Furthermore, if $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,q}(\kappa)$, and $q' \ge q$, a positive integer $s' \le s$, and κ' are such that $\kappa'(s')^{1/q'-1} \ge \kappa s^{1/q-1}$, then $(s^{1/q}(s')^{-1/q'} H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s',q'}(\kappa')$. In particular, when $s' \le s^{1-1/q}$, the fact that $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,q}(\kappa)$ implies that $(s^{1/q} H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s',\infty}(\kappa)$.*

Relation to known conditions for the validity of sparse ℓ_1 recovery. Note that whenever

$$\mathcal{S} = \big(B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)}\big)$$

is the standard r.s., the condition $\mathbf{Q}_{s,q}(\kappa)$ reduces to the condition $\mathbf{H}_{s,q}(\kappa)$ introduced in [19]. On the other hand, condition $\mathbf{Q}_{s,p}(\kappa)$ is closely related to other known conditions, introduced to study the properties of recovery routines in the context of block-sparsity. Specifically, consider an r.s. with $B = I_n$, and let us make the following observation:

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

$$(2.3)\qquad \forall x \in \mathbb{R}^n \qquad L_{s,q}(x) \le \widehat{\lambda} s^{1/q}\|Ax\|_2 + \kappa s^{1/q-1} L_1(x).$$

Let us fix the r.s. $\mathcal{S}_2 = (I_n, n_1, \dots, n_K, \|\cdot\|_2, \dots, \|\cdot\|_2)$. Condition (2.3) with $\kappa < 1/2$ plays a crucial role in the performance analysis of the group-Lasso and Dantzig Selector. For example, the error bounds for Lasso recovery obtained in [24] rely upon the Restricted Eigenvalue assumption RE$(s, \varkappa)$ as follows: there exists $\varkappa > 0$ such that

$$L_2(x^s) \le \frac{1}{\varkappa}\|Ax\|_2 \qquad \text{whenever } 3L_1(x^s) \ge L_1(x - x^s).$$

In this case $L_{s,1}(x) \le \sqrt{s} L_{s,2}(x) \le \frac{\sqrt{s}}{\varkappa}\|Ax\|_2$ whenever $4L_{s,1}(x) \ge L_1(x)$, so that

$$(2.4)\qquad \forall x \in \mathbb{R}^n \qquad L_{s,1}(x) \le \frac{s^{1/2}}{\varkappa}\|Ax\|_2 + \frac{1}{4}L_1(x),$$

which is exactly (2.3) with $q = 1$, $\kappa = 1/4$ and $\widehat{\lambda} = (\varkappa\sqrt{s})^{-1}$ (observe that (2.4) is nothing but the "block version" of the Compatibility condition from [31]).

Recall that a sensing matrix $A \in \mathbb{R}^{m\times n}$ satisfies the Block Restricted Isometry Property BRIP(δ, k) (see, e.g., [13]) with $\delta \geq 0$ and a positive integer k if for every $x \in \mathbb{R}^n$ with at most k nonvanishing blocks one has

$$(1-\delta)\|x\|_2^2 \leq x^T A^T A x \leq (1+\delta)\|x\|_2^2. \tag{2.5}$$

PROPOSITION 2.1. *Let $A \in \mathbb{R}^{m\times n}$ satisfy* BRIP($\delta, 2s$) *for some $\delta < 1$ and positive integer s. Then*:

(i) *The pair* ($H = \frac{s^{-1/2}}{\sqrt{1-\delta}} I_m, \|\cdot\|_2$) *satisfies the condition* $\mathbf{Q}_{s,2}(\frac{\delta}{1-\delta})$ *associated with A and the r.s.* $\mathcal{S}_2$.

(ii) *The pair* ($H = \frac{1}{1-\delta} A, L_\infty(\cdot)$) *satisfies the condition* $\mathbf{Q}_{s,2}(\frac{\delta}{1-\delta})$ *associated with A and the r.s.* $\mathcal{S}_2$.

Our last observation here is as follows: let $(H, \|\cdot\|)$ satisfy $\mathbf{Q}_{s,q}(\kappa)$ for the r.s. given by $(B, n_1, \dots, n_K, \|\cdot\|_2, \dots, \|\cdot\|_2)$, and let $d = \max_k n_k$. Then $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,q}(\sqrt{d}\kappa)$ for the r.s. given by $(B, n_1, \dots, n_K, \|\cdot\|_\infty, \dots, \|\cdot\|_\infty)$.

3. Accuracy bounds for ℓ_1 block recovery routines. Throughout this section we fix an r.s. $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$ and a sensing matrix A.

3.1. *Regular ℓ_1 recovery.* We define the *regular ℓ_1 recovery* as

$$\widehat{x}_{\mathrm{reg}}(y) \in \operatorname*{Arg\,min}_u \{L_1(Bu) : \|H^T(Au - y)\| \leq \rho\}, \tag{3.1}$$

where the *contrast matrix* $H \in \mathbb{R}^{m\times M}$, the norm $\|\cdot\|$ and $\rho > 0$ are parameters of the construction.

THEOREM 3.1. *Let s be a positive integer, $q \in [1, \infty]$, $\kappa \in (0, 1/2)$. Assume that the pair $(H, \|\cdot\|)$ satisfies the condition $\mathbf{Q}_{s,q}(\kappa)$ associated with A and r.s. $\mathcal{S}$, and let*

$$\Xi = \Xi_{\rho,\mathcal{U}} = \{\xi : \|H^T(u + \xi)\| \leq \rho \ \forall u \in \mathcal{U}\}. \tag{3.2}$$

Then for all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$ and $\xi \in \Xi$ one has

$$\begin{aligned} &L_p\big(B[\widehat{x}_{\mathrm{reg}}(Ax + u + \xi) - x]\big) \\ &\quad \leq \frac{4(2s)^{1/p}}{1-2\kappa}\left[\rho + \frac{1}{2s} L_1\big(Bx - [Bx]^s\big)\right], \qquad 1 \leq p \leq q. \end{aligned} \tag{3.3}$$

The above result can be slightly strengthened by replacing the assumption that $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,q}(\kappa)$, $\kappa < 1/2$, with a weaker, by Observation 2.1, assumption that $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,1}(\varkappa)$ with $\varkappa < 1/2$ and satisfies $\mathbf{Q}_{s,q}(\kappa)$ with some (perhaps large) κ:

THEOREM 3.2. *Given A, r.s. $\mathcal{S}$, integer $s > 0$, $q \in [1, \infty]$ and $\varepsilon \in (0, 1)$, assume that $(H, \|\cdot\|)$ satisfies the condition $\mathbf{Q}_{s,1}(\varkappa)$ with $\varkappa < 1/2$ and the condition $\mathbf{Q}_{s,q}(\kappa)$ with some $\kappa \geq \varkappa$, and let Ξ be given by* (3.2). *Then for all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$, $\xi \in \Xi$ and p, $1 \leq p \leq q$, it holds*

$$L_p\big(B\big[\widehat{x}_{\text{reg}}(Ax + u + \xi) - x\big]\big) \leq \frac{4(2s)^{1/p}[1 + \kappa - \varkappa]^{q(p-1)/(p(q-1))}}{1 - 2\varkappa}\bigg[\rho + \frac{L_1(Bx - [Bx]^s)}{2s}\bigg]. \tag{3.4}$$

3.2. *Penalized ℓ_1 recovery.* The penalized ℓ_1 recovery is

$$\widehat{x}_{\text{pen}}(y) \in \operatorname*{Arg\,min}_u \big\{L_1(Bu) + \lambda \big\|H^T(Ax - y)\big\|\big\}, \tag{3.5}$$

where $H \in \mathbb{R}^{m \times M}$, $\|\cdot\|$ and a positive real λ are parameters of the construction.

THEOREM 3.3. *Given A, r.s. $\mathcal{S}$, integer s, $q \in [1, \infty]$ and $\varepsilon \in (0, 1)$, assume that $(H, \|\cdot\|)$ satisfies the conditions $\mathbf{Q}_{s,q}(\kappa)$ and $\mathbf{Q}_{s,1}(\varkappa)$ with $\varkappa < 1/2$ and $\kappa \geq \varkappa$.*

(i) *Let $\lambda \geq 2s$. Then for all $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ it holds for $1 \leq p \leq q$*

$$\begin{aligned} &L_p\big(B\big[\widehat{x}_{\text{pen}}(y) - x\big]\big) \\ &\quad \leq \frac{4\lambda^{1/p}}{1 - 2\varkappa}\bigg[1 + \frac{\kappa\lambda}{2s} - \varkappa\bigg]^{q(p-1)/(p(q-1))} \\ &\quad\quad \times \bigg[\big\|H^T(Ax - y)\big\| + \frac{1}{2s}L_1\big(Bx - [Bx]^s\big)\bigg]. \end{aligned} \tag{3.6}$$

In particular, with $\lambda = 2s$ we have for $1 \leq p \leq q$

$$\begin{aligned} &L_p\big(B\big[\widehat{x}_{\text{pen}}(y) - x\big]\big) \\ &\quad \leq \frac{4(2s)^{1/p}}{1 - 2\varkappa}[1 + \kappa - \varkappa]^{q(p-1)/(p(q-1))} \\ &\quad\quad \times \bigg[\big\|H^T(Ax - y)\big\| + \frac{1}{2s}L_1\big(Bx - [Bx]^s\big)\bigg]. \end{aligned} \tag{3.7}$$

(ii) *Let $\rho \geq 0$ and Ξ be given by* (3.2). *Then for all $x \in \mathbb{R}^n$, $u \in \mathcal{U}$ and all $\xi \in \Xi$ one has for $1 \leq p \leq q$*

$$\begin{aligned} \lambda \geq 2s \quad \Rightarrow \quad &L_p\big(B\big[\widehat{x}_{\text{pen}}(Ax + u + \xi) - x\big]\big) \\ &\quad \leq \frac{4\lambda^{1/p}}{1 - 2\varkappa}\bigg[1 + \frac{\kappa\lambda}{2s} - \varkappa\bigg]^{q(p-1)/(p(q-1))} \\ &\quad\quad \times \bigg[\rho + \frac{1}{2s}L_1\big(Bx - [Bx]^s\big)\bigg], \end{aligned} \tag{3.8}$$

$$\lambda = 2s \quad \Rightarrow \quad L_p\big(B[\widehat{x}_{\text{pen}}(Ax+u+\xi) - x]\big) \\ \le \frac{4(2s)^{1/p}}{1-2\varkappa}[1+\kappa-\varkappa]^{q(p-1)/(p(q-1))} \\ \times \left[\rho + \frac{1}{2s}L_1\big(Bx - [Bx]^s\big)\right].$$

Discussion. Let us compare the error bounds of the regular and the penalized ℓ_1 recoveries associated with the same pair $(H, \|\cdot\|)$ satisfying the condition $\mathbf{Q}_{s,q}(\kappa)$ with $\kappa < 1/2$. Given $\varepsilon \in (0,1)$, let

$$\rho_\varepsilon\big[H, \|\cdot\|\big] = \min\big\{\rho : \mathrm{Prob}\big\{\xi : \big\|H^T(u+\xi)\big\| \le \rho \ \forall u \in \mathcal{U}\big\} \ge 1-\varepsilon\big\}; \tag{3.9}$$

this is nothing but the smallest ρ such that

$$\mathrm{Prob}(\xi \in \Xi_{\rho,\varepsilon}) \ge 1-\varepsilon \tag{3.10}$$

[see (3.2)] and, thus, the smallest ρ for which the error bound (3.3) for the regular ℓ_1 recovery holds true with probability $1-\varepsilon$ (or at least the smallest ρ for which the latter claim is supported by Theorem 3.1). With $\rho = \rho_\varepsilon[H, \|\cdot\|]$, the regular ℓ_1 recovery guarantees (and that is the best guarantee one can extract from Theorem 3.1) that

> (!) For some set Ξ, $\mathrm{Prob}\{\xi \in \Xi\} \ge 1-\varepsilon$, of "good" realizations of the random component ξ of the observation error, one has
>
> $$L_p\big(B[\widehat{x}(Ax+u+\xi) - x]\big) \\ \le \frac{4(2s)^{1/p}}{1-2\kappa}\left[\rho_\varepsilon\big[H, \|\cdot\|\big] + \frac{L_1(Bx-[Bx]^s)}{2s}\right], \qquad 1 \le p \le q, \tag{3.11}$$
>
> whenever $x \in \mathbb{R}^n$, $u \in \mathcal{U}$, and $\xi \in \Xi$.

The error bound (3.7) [where we can safely set $\varkappa = \kappa$, since $\mathbf{Q}_{s,q}(\kappa)$ implies $\mathbf{Q}_{s,1}(\kappa)$] says that *(!) holds true for the penalized ℓ_1 recovery with $\lambda = 2s$*. The latter observation suggests that the penalized ℓ_1 recovery associated with $(H, \|\cdot\|)$ and $\lambda = 2s$ is better than its regular counterpart, the reason being twofold. First, in order to ensure (!) with the regular recovery, the "built in" parameter ρ of this recovery should be set to $\rho_\varepsilon[H, \|\cdot\|]$, and the latter quantity is not always easy to identify. In contrast to this, the construction of the penalized ℓ_1 recovery is completely independent of a priori assumptions on the structure of observation errors, while automatically ensuring (!) for the error model we use. Second, and more importantly, for the penalized recovery the bound (3.11) is no more than the "worst, with confidence $1-\varepsilon$, case," and the typical values of the quantity $\|H^T(u+\xi)\|$ which indeed participates in the error bound (3.6) are essentially smaller than $\rho_\varepsilon[H, \|\cdot\|]$. Our numerical experience fully supports the above suggestion: the difference in observed performance of the two routines in question, although not dramatic, is definitely in favor of the penalized recovery. The only potential disadvantage of the latter routine is that the penalty parameter λ should be tuned to the level s of sparsity we aim at, while the regular recovery is free of any guess of this type. Of course, the "tuning" is rather loose—all we need (and experiments show that we indeed need this) is the relation $\lambda \geq 2s$, so that a rough upper bound on s will do; note, however, that the bound (3.6) deteriorates as λ grows.

4. Tractability of condition $\mathbf{Q}_{s,\infty}(\kappa)$, ℓ_∞-norm of the blocks. We have seen in Section 3 that given a sensing matrix A and an r.s. $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$ such that the associated conditions $\mathbf{Q}_{s,q}(\kappa)$ are satisfiable, we can validate the ℓ_1 recovery of nearly s-block-sparse signals, specifically, we can point out ℓ_1-type recoveries with controlled (and small, provided so are the observation error and the deviation of the signal from an s-block-sparse one). The bad news here is that, in general, condition $\mathbf{Q}_{s,q}(\kappa)$, as well as other conditions for the validity of ℓ_1 recovery, like Block RE/RIP, cannot be verified efficiently. The latter means that given a sensing matrix A and a r.s. $\mathcal{S}$, it is difficult to verify that a given candidate pair $(H, \|\cdot\|)$ satisfies condition $\mathbf{Q}_{s,q}(\kappa)$ associated with A and $\mathcal{S}$. Fortunately, one can construct "tractable approximations" of condition $\mathbf{Q}_{s,q}(\kappa)$, that is, verifiable sufficient conditions for the validity of $\mathbf{Q}_{s,q}(\kappa)$. The first good news is that when all $\|\cdot\|_{(k)}$ are the uniform norms $\|\cdot\|_\infty$ and, in addition, $q = \infty$ [which, by Observation 2.1, corresponds to the strongest among the conditions $\mathbf{Q}_{s,q}(\kappa)$ and ensures the validity of (3.3) and (3.6) in the largest possible range $1 \leq p \leq \infty$ of values of p], the condition $\mathbf{Q}_{s,q}(\kappa)$ becomes "fully computationally tractable." We intend to demonstrate also that the condition $\mathbf{Q}_{s,\infty}(\kappa)$ is in fact necessary for the risk bounds of the form (3.3)–(3.8) to be valid when $p = \infty$.

4.1. *Condition $\mathbf{Q}_{s,\infty}(\kappa)$: Tractability and the optimal choice of the contrast matrix H.*

Notation. In the sequel, given $r, \theta \in [1, \infty]$ and a matrix M, we denote by $\|M\|_{r,\theta}$ the norm of the linear operator $u \mapsto Mu$ induced by the norms $\|\cdot\|_r$ and $\|\cdot\|_\theta$ on the argument and the image spaces:

$$\|M\|_{r,\theta} = \max_{u:\,\|u\|_r \leq 1} \|Mu\|_\theta.$$

We denote by $\|M\|_{(\ell,k)}$ the norm of the linear mapping $u \mapsto Mu : \mathbb{R}^{n_\ell} \to \mathbb{R}^{n_k}$ induced by the norms $\|\cdot\|_{(\ell)}$, $\|\cdot\|_{(k)}$ on the argument and on the image spaces. Further, $\mathrm{Row}_k[M]$ stands for the transpose of the kth row of M and $\mathrm{Col}_k[M]$ stands for kth column of M. Finally, $\|u\|_{s,q}$ is the ℓ_q-norm of the vector obtained from a vector $u \in \mathbb{R}^k$ by zeroing all but the s largest in magnitude entries in u.

Main result. Consider r.s. $\mathcal{S}_\infty = (B, n_1, \dots, n_K, \|\cdot\|_\infty, \dots, \|\cdot\|_\infty)$. We claim that in this case the condition $\mathbf{Q}_{s,\infty}(\kappa)$ becomes fully tractable. Specifically, we have the following.

PROPOSITION 4.1. *Let a matrix $A \in \mathbb{R}^{m\times n}$, the r.s. $\mathcal{S}_\infty$, a positive integer s and reals $\kappa > 0$, $\varepsilon \in (0,1)$ be given.*

(i) *Assume that a triple $(H, \|\cdot\|, \rho)$, where $H \in \mathbb{R}^{m\times M}$, $\|\cdot\|$ is a norm on $\mathbb{R}^M$, and $\rho \geq 0$, is such that*

(!) $(H, \|\cdot\|)$ *satisfies* $\mathbf{Q}_{s,\infty}(\kappa)$, *and the set* $\Xi = \{\xi : \|H^T[u+\xi]\| \leq \rho\ \forall u \in \mathcal{U}\}$ *satisfies* $\mathrm{Prob}(\xi \in \Xi) \geq 1-\varepsilon$.

Given H, $\|\cdot\|$, ρ, one can find efficiently $N = n_1 + \cdots + n_K$ vectors $h^1, \dots, h^N$ in $\mathbb{R}^m$ and $N \times N$ block matrix $V = [V^{k\ell}]_{k,\ell=1}^K$ (the blocks $V^{k\ell}$ of V are $n_k \times n_\ell$ matrices) such that

$$
\begin{aligned}
&\text{(a)} \quad B = VB + \big[h^1, \dots, h^N\big]^T A,\\
(4.1)\quad &\text{(b)} \quad \big\|V^{k\ell}\big\|_{\infty,\infty} \leq s^{-1}\kappa \qquad \forall k, \ell \leq K,\\
&\text{(c)} \quad \mathrm{Prob}_\xi\Big(\Xi^+ := \Big\{\xi : \max_{u\in\mathcal{U}} u^T h^i + |\xi^T h^i| \leq \rho, 1 \leq i \leq N\Big\}\Big) \geq 1-\varepsilon
\end{aligned}
$$

(*note that the matrix norm $\|A\|_{\infty,\infty} = \max_j \|\mathrm{Row}_j[A]\|_1$ is simply the maximum ℓ_1-norm of the rows of A*).

(ii) *Whenever vectors $h^1, \dots, h^N \in \mathbb{R}^m$ and a matrix $V = [V^{k\ell}]_{k,\ell=1}^K$ with $n_k \times n_\ell$ blocks $V^{k\ell}$ satisfy* (4.1), *the $m \times N$ matrix $\widehat{H} = [h^1, \dots, h^N]$, the norm $\|\cdot\|_\infty$ on $\mathbb{R}^N$ and ρ form a triple satisfying* (!).

Discussion. Let a sensing matrix $A \in \mathbb{R}^{m\times n}$ and a r.s. $\mathcal{S}_\infty$ be given, along with a positive integer s, an uncertainty set $\mathcal{U}$, a distribution P of ξ and $\varepsilon \in (0,1)$. Theorems 3.1 and 3.3 say that if a triple $(H, \|\cdot\|, \rho)$ is such that $(H, \|\cdot\|)$ satisfies $\mathbf{Q}_{s,\infty}(\kappa)$ with $\kappa < 1/2$ and H, ρ are such that the set Ξ given by (3.2) satisfies (3.10), then for the regular ℓ_1 recovery associated with $(H, \|\cdot\|, \rho)$ and for the penalized ℓ_1 recovery associated with $(H, \|\cdot\|)$ and $\lambda = 2s$, the following holds:

$$
\begin{aligned}
&\forall \big(x \in \mathbb{R}^n, u \in \mathcal{U}, \xi \in \Xi\big)\\
(4.2)\qquad &L_p\big(B\big[\widehat{x}(Ax+u+\xi) - x\big]\big) \leq \frac{4(2s)^{1/p}}{1-2\kappa}\Big[\rho + \frac{1}{2s} L_1\big(Bx - [Bx]^s\big)\Big],\\
&\qquad\qquad 1 \leq p \leq \infty.
\end{aligned}
$$

Proposition 4.1 states that when applying this result, we lose nothing by restricting ourselves with triples $H = [h^1, \dots, h^N] \in \mathbb{R}^{m\times N}$, $N = n_1 + \cdots + n_K$, $\|\cdot\| = L_\infty(\cdot)$, $\rho \geq 0$ which can be augmented by an appropriately chosen $N \times N$ matrix V to satisfy relations (4.1). In the rest of this discussion, it is assumed that we are speaking about triples $(H, \|\cdot\|, \rho)$ satisfying the just defined restrictions.

The bound (4.2) is completely determined by two parameters—κ (which should be $< 1/2$) and ρ; the smaller are these parameters, the better are the bounds. In what follows we address the issue of efficient synthesis of matrices H with "as good as possible" values of κ and ρ.

Observe first that $H = [h^1, \dots, h^N]$ and κ should admit an extension by a matrix V to a solution of the system of convex constraints (4.1)(a), (4.1)(b). In the case of $\xi \equiv 0$ the best choice of ρ, given H, is

$$\rho = \max_i \mu_{\mathcal{U}}(h^i) \qquad \text{where } \mu_{\mathcal{U}}(h) = \max_{u \in \mathcal{U}} u^T h.$$

Consequently, in this case the "achievable pairs" ρ, κ form a computationally tractable convex set

$$\begin{aligned} G_s = \Bigg\{ (\kappa, \rho) : \exists H = [h^1, \dots, h^N] \in \mathbb{R}^{m \times N}, \\ V = [V^{k\ell} \in \mathbb{R}^{n_k \times n_\ell}]_{k,\ell=1}^K : B = VB + H^T A, \\ \|V^{k\ell}\|_{\infty,\infty} \le \frac{\kappa}{s}, \mu_{\mathcal{U}}(h^i) \le \rho, 1 \le i \le N \Bigg\}. \end{aligned}$$

When ξ does not vanish, the situation is complicated by the necessity to maintain the validity of the restriction

$$\begin{aligned} \operatorname{Prob}(\xi \in \Xi^+) := \operatorname{Prob}\{\xi : \mu_{\mathcal{U}}(h^i) + |\xi^T h^i| \le \rho, 1 \le i \le N\} \\ \ge 1 - \varepsilon, \end{aligned} \tag{4.3}$$

which is a chance constraint in variables $h^1, \dots, h^N$, ρ and as such can be "computationally intractable." Let us consider the "standard" case of Gaussian zero mean noise ξ, that is, assume that $\xi = D\eta$ with $\eta \sim \mathcal{N}(0, I_m)$ and known $D \in \mathbb{R}^{m \times m}$. Then (4.3) implies that

$$\rho \ge \max_i \left[\mu_{\mathcal{U}}(h^i) + \operatorname{Erfinv}\left(\frac{\varepsilon}{2}\right) \|D^T h^i\|_2 \right].$$

On the other hand, (4.3) is clearly implied by

$$\rho \ge \max_i \left[\mu_{\mathcal{U}}(h^i) + \operatorname{Erfinv}\left(\frac{\varepsilon}{2N}\right) \|D^T h^i\|_2 \right].$$

Ignoring the "gap" between $\operatorname{Erfinv}(\frac{\varepsilon}{2})$ and $\operatorname{Erfinv}(\frac{\varepsilon}{2N})$, we can safely model the restriction (4.3) by the system of convex constraints

$$\mu_{\mathcal{U}}(h^i) + \operatorname{Erfinv}\left(\frac{\varepsilon}{2N}\right) \|D^T h^i\|_2 \le \rho, \qquad 1 \le i \le N. \tag{4.4}$$

Thus, the set G_s of admissible κ, ρ can be safely approximated by the computationally tractable convex set

$$G_s^* = \Big\{(\kappa,\rho): \exists\big[H=[h^1,\dots,h^N]\in\mathbb{R}^{m\times N},\ V=\big[V^{k\ell}\in\mathbb{R}^{n_k\times n_\ell}\big]_{k,\ell=1}^K\big]: \Big\{B = BV + H^TA,\ \|V^{k\ell}\|_{\infty,\infty}\le\frac{\kappa}{s},\ 1\le k,\ell\le K,\ \max_{u\in\mathcal{U}} u^Th^i + \mathrm{Erfinv}\Big(\frac{\varepsilon}{2N}\Big)\|D^Th^i\|_2\le\rho,\ 1\le i\le N\Big\}\Big\}. \tag{4.5}$$

4.2. *Condition* $\mathbf{Q}_{s,\infty}(\kappa)$: *Necessity.* In this section, as above, we assume that all norms $\|\cdot\|_{(k)}$ in the r.s. $\mathcal{S}_\infty$ are ℓ_∞-norms; we assume, in addition, that ξ is a zero mean Gaussian noise: $\xi = D\eta$ with $\eta\sim\mathcal{N}(0,I_m)$ and known $D\in\mathbb{R}^{m\times m}$. From the above discussion we know that if, for some $\kappa<1/2$ and $\rho>0$, there exist $H=[h^1,\dots,h^N]\in\mathbb{R}^{m\times N}$ and $V=[V^{k\ell}\in\mathbb{R}^{n_k\times n_\ell}]_{k,\ell=1}^K$ satisfying (4.1), then regular and penalized ℓ_1 recoveries with appropriate choice of parameters ensure that

$$\forall(x\in\mathbb{R}^n, u\in\mathcal{U})\quad \mathrm{Prob}_\xi\Big(\big\|B\big[x-\widehat{x}(Ax+u+\xi)\big]\big\|_\infty\le\frac{4}{1-2\kappa}\Big[\rho+\frac{L_1(Bx-[Bx]^s)}{2s}\Big]\Big)\ge 1-\varepsilon. \tag{4.6}$$

We are about to demonstrate that this implication can be "nearly inverted":

PROPOSITION 4.2. *Let a sensing matrix A, an r.s. $\mathcal{S}_\infty$ with $\|\cdot\|_{(k)}=\|\cdot\|_\infty$, $1\le k\le K$, an uncertainty set $\mathcal{U}$, and reals $\kappa>0$, $\varepsilon\in(0,1/2)$ be given. Suppose that the observation error "is present," specifically, that for every $r>0$, the set $\{u+De: u\in\mathcal{U}, \|e\|_2\le r\}$ contains a neighborhood of the origin.*

Given a positive integer S, assume that there exists a recovering routine $\widehat{x}$ satisfying an error bound of the form (4.6), *specifically, such that for all $x\in\mathbb{R}^n$, $u\in\mathcal{U}$,*

$$\mathrm{Prob}_\xi\big(\big\|B\big[x-\widehat{x}(Ax+u+\xi)\big]\big\|_\infty\le\alpha+S^{-1}L_1\big(Bx-[Bx]^S\big)\big)\ge 1-\varepsilon \tag{4.7}$$

for some $\alpha>0$. Then there exist $H=[h^1,\dots,h^N]\in\mathbb{R}^{m\times N}$ and $V=[V^{k\ell}\in\mathbb{R}^{n_k\times n_\ell}]_{k,\ell=1}^K$ satisfying

$$\begin{aligned}&\text{(a)}\quad B = VB + H^TA,\\ &\text{(b)}\quad \|V^{k\ell}\|_{\infty,\infty}\le 2S^{-1}\qquad \forall k,\ell\le K,\\ &\text{(c)}\quad \textit{with } \rho := \max_{1\le i\le N}\Big[\max_{u\in\mathcal{U}} u^Th^i + \mathrm{Erfinv}\Big(\frac{\varepsilon}{2N}\Big)\|D^Th^i\|_2\Big],\end{aligned} \tag{4.8}$$

one has $\rho \le 2\alpha$ *when* $D = 0$, $\rho \le 2\alpha \frac{\mathrm{Erfinv}(\varepsilon/(2N))}{\mathrm{Erfinv}(\varepsilon)}$ *when* $D \neq 0$, *and for* $\xi = D\eta, \eta \sim \mathcal{N}(0, I_m)$ *one has*

$$\mathrm{Prob}_\xi\Big(\Xi^+ := \Big\{\xi : \max_{u\in\mathcal{U}} u^T h^i + |\xi^T h^i| \le \rho, 1 \le i \le N\Big\}\Big) \ge 1 - \varepsilon.$$

In other words (see Proposition 4.1), $(H, L_\infty(\cdot))$ satisfies $\mathbf{Q}_{s,\infty}(\kappa)$ for s "nearly as large as S," namely, $s \le \frac{\kappa}{2}S$, and $H = [h^1, \dots, h^k]$, ρ satisfy conditions (4.4) with ρ being "nearly α," namely, $\rho \le 2\alpha$ in the case of $D = 0$ and $\rho \le 2\frac{\mathrm{Erfinv}(\varepsilon/(2N))}{\mathrm{Erfinv}(\varepsilon)}$ when $D \neq 0$. In particular, under the premise of Proposition 4.2, the contrast optimization procedure of Section 4.1 supplies the matrix H such that the corresponding regular or penalized recovery $\widehat{x}(\cdot)$ for all $s \le \frac{S}{8}$ satisfies

$$\mathrm{Prob}_\xi\left\{\|B[x-\widehat{x}(y)]\|_\infty \le 4\left[4\frac{\mathrm{Erfinv}(\varepsilon/(2N))}{\mathrm{Erfinv}(\varepsilon)}\alpha + s^{-1}L_1\big(Bx-[Bx]^s\big)\right]\right\} \ge 1-\varepsilon.$$

5. Tractable approximations of $\mathbf{Q}_{s,q}(\kappa)$. Aside from the important case of $q = \infty$, $\|\cdot\|_{(k)} = \|\cdot\|_\infty$ considered in Sections 4.1 and 4.2, condition $\mathbf{Q}_{s,q}(\kappa)$ "as it is" seems to be computationally intractable: unless $s = O(1)$, it is unknown how to check efficiently that a given pair $(H, \|\cdot\|)$ satisfies this condition, not speaking about synthesis of a pair satisfying this condition and resulting in the best possible error bound (3.3), (3.6) for regular and penalized ℓ_1-recoveries. We are about to present *verifiable sufficient conditions* for the validity of $\mathbf{Q}_{s,q}(\kappa)$ which may become an interesting alternative for condition $\mathbf{Q}_{s,q}(\kappa)$ for that purposes.

5.1. *Sufficient condition for* $\mathbf{Q}_{s,q}(\kappa)$.

PROPOSITION 5.1. *Suppose that a sensing matrix* A, *an r.s.* $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$, *and* $\kappa \ge 0$ *are given.*

Let $N = n_1 + \cdots + n_K$, *and let* $N \times N$ *matrix* $V = [V^{k\ell}]_{k,\ell=1}^K$ ($V^{k\ell}$ *are* $n_k \times n_\ell$) *and* $m \times N$ *matrix* H *satisfy the relation*

$$B = VB + H^T A. \tag{5.1}$$

Let us denote

$$\nu^*_{s,q}(V) = \max_{1\le\ell\le K}\ \max_{w^\ell\in\mathbb{R}^{n_\ell}:\,\|w^\ell\|_{(\ell)}\le 1} L_{s,q}\big([V^{1\ell}w^\ell; \dots; V^{K\ell}w^\ell]\big).$$

Then for all $s \le K$ *and all* $q \in [1, \infty]$, *we have*

$$L_{s,q}(Bx) \le s^{1/q}L_\infty\big(H^T Ax\big) + \nu^*_{s,q}(V)L_1(Bx) \qquad \forall x \in \mathbf{R}^n. \tag{5.2}$$

The result of Proposition 5.1 is a step toward developing a verifiable sufficient condition for the validity of $\mathbf{Q}_{s,q}$. To get such a condition, we need an efficiently computable upper bound of the quantity $\nu^*_{s,q}$. In particular, if for a given positive integer $s \leq K$ and a real $q \in [1, \infty]$ there exist an upper bounding function $\nu_{s,q}(V)$ such that

$$\nu_{s,q}(\cdot) \text{ is convex} \quad \text{and} \quad \nu_{s,q}(V) \geq \nu^*_{s,q}(V) \qquad \forall V \tag{5.3}$$

and a matrix V such that

$$\nu_{s,q}(V) \leq s^{1/q-1}\kappa, \tag{5.4}$$

then the pair $(H, L_\infty(\cdot))$ satisfies $\mathbf{Q}_{s,q}(\kappa)$. An important example of the upper bound for $\nu^*_{s,q}(V)$ which satisfies (5.4) is provided in the following statement.

PROPOSITION 5.2. *Let Ω be a $K \times K$ matrix with entries $[\Omega]_{k,\ell} = \|V^{k\ell}\|_{(\ell,k)}$, $1 \leq k, \ell \leq K$. Then*

$$\widehat{\nu}_{s,q}(V) := \max_{1\leq k\leq K} \|\mathrm{Col}_k[\Omega]\|_{s,q} \geq \nu^*_{s,q}(V) \qquad \forall V \tag{5.5}$$

[*note that the inequality in* (5.5) *becomes equality when either $q = \infty$ or $s = 1$*], *so that the condition*

$$\widehat{\nu}_{s,q}(V) \leq s^{1/q-1}\kappa \tag{5.6}$$

taken along with (5.1) *is sufficient for $(H, L_\infty(\cdot))$ to satisfy $\mathbf{Q}_{s,q}(\kappa)$.*

When all $\|\cdot\|_{(k)}$ are the ℓ_∞-norms and $q = \infty$, the results of Propositions 5.1 and 5.2 recover Proposition 4.1. In the general case, they suggest a way to synthesize matrices $H \in \mathbb{R}^{m\times N}$ which, augmented by the norm $\|\cdot\| = L_\infty(\cdot)$, provably satisfies the condition $\mathbf{Q}_{s,q}(\kappa)$, along with a certificate V for this fact. Namely, H and V should satisfy the system of linear equations (5.1) and, in addition, (5.4) should hold for V with $\nu_{s,q}(\cdot)$ satisfying (5.3). Further, for such a $\nu_{s,q}(\cdot)$, (5.4) is a system of convex constraints on V. Whenever these constraints are efficiently computable, we get a computationally tractable sufficient condition for $(H, L_\infty(\cdot))$ to satisfy $\mathbf{Q}_{s,q}(\kappa)$—a condition which is expressed by an explicit system of efficiently computable convex constraints (5.1), (5.4) on H and additional matrix variable V.

5.2. *Tractable sufficient conditions and contrast optimization.* The quantity $\widehat{\nu}_{s,q}(\cdot)$ is the simplest choice of $\nu_{s,q}(\cdot)$ satisfying (5.3). In this case, efficient computability of the constraints (5.4) is the same as efficient computability of norms $\|\cdot\|_{(k,\ell)}$. Assuming that $\|\cdot\|_{(k)} = \|\cdot\|_{r_k}$ for every k in the r.s. $\mathcal{S}$, the computability issue becomes the one of efficient computation of the norms $\|\cdot\|_{r_\ell,r_k}$. The norm $\|\cdot\|_{r,\theta}$ is known to be generically efficiently computable in only three cases:

(1) $\theta = \infty$, where $\|M\|_{r,\infty} = \|M^T\|_{1,r/(r-1)} = \max_i \|\mathrm{Row}_i^T(M)\|_{r/(r-1)}$;
(2) $r = 1$, where $\|M\|_{1,\theta} = \max_j \|\mathrm{Col}_j[M]\|_\theta$;
(3) $r = \theta = 2$, where $\|M\|_{2,2} = \sigma_{\max}(M)$ is the spectral norm of M.

Assuming for the sake of simplicity that in our r.s. $\|\cdot\|_{(k)}$ are r-norms with common value of r, let us look at three "tractable cases" as specified by the above discussion—those of $r=\infty$, $r=1$ and $r=2$. In these cases, candidate contrast matrices H are $m\times N$, the associated norm $\|\cdot\|$ is $L_\infty(\cdot)$, and our sufficient condition for H to be *good* [i.e., for $(H, L_\infty(\cdot))$ to satisfy $\mathbf{Q}_{s,q}(\kappa)$ with given $\kappa<1/2$ and q] becomes a system $\mathbf{S}=\mathbf{S}_{\kappa,q}$ of explicit efficiently computable convex constraints on H and additional matrix variable $V\in\mathbb{R}^{N\times N}$, implying that the set $\mathcal{H}$ of good H is convex and computationally tractable, so that we can minimize efficiently over $\mathcal{H}$ any convex and efficiently computable function. In our context, a natural way to use $\mathbf{S}$ is to optimize over $H\in\mathcal{H}$ the error bound (3.11) or, which is the same, to minimize over $\mathcal{H}$ the function $\rho(H)=\rho_\varepsilon[H,L_\infty(\cdot)]$; see (3.9), where $\varepsilon<1$ is a given tolerance. Taken literally, this problem still can be difficult, since the function $\rho(H)$ is not necessarily convex and can be difficult to compute even in the convex case. To overcome this difficulty, we again can use a *verifiable sufficient condition* for the relation $\rho(H)\le\rho$, that is, a system $\mathbf{T}=\mathbf{T}_\varepsilon$ of explicit efficiently computable convex constraints on variables H and ρ (and, perhaps some slack variables ζ) such that $\rho(H)\le\rho$ for the (H,ρ)-component of every feasible solution of $\mathbf{T}$. With this approach, the design of the best, as allowed by $\mathbf{S}$ and $\mathbf{T}$, contrast matrix H reduces to solving a convex optimization problem with efficiently computable constraints in variables H, V, ρ, specifically, the problem

$$\min_{\rho,H,V,\zeta}\{\rho: H, V \text{ satisfy } \mathbf{S};\ H,\rho,\zeta \text{ satisfy } \mathbf{T}\}. \tag{5.7}$$

In the rest of this section we present explicitly the systems $\mathbf{S}$ and $\mathbf{T}$ for the three tractable cases we are interested in, assuming the following model of observation errors:

$$\mathcal{U}=\{u=Ev:\|v\|_2\le 1\};\qquad \xi=D\eta,\ \eta\sim\mathcal{N}(0,I_m),$$

where $E, D\in\mathbb{R}^{m\times m}$ are given.

We use the following notation: the $m\times N$ matrix H is partitioned into $m\times n_k$ blocks $H[k]$, $1\le k\le K$, according to the block structure of the representation vectors; the tth column in $H[k]$ is denoted $h^{kt}\in\mathbb{R}^m$, $1\le t\le n_k$.

For derivations of the results to follow, see Section A.7 of the supplementary article [18].

The case of $r=\infty$. The case of $q=\infty$ was considered in full details in Section 4.1. When $q\le\infty$, one has

$$\mathbf{S}_{\kappa,q}:\begin{cases} B=VB+H^TA, & \\ \Omega_{k\ell}:=\|V^{k\ell}\|_{\infty,\infty}=\max\limits_{1\le t\le n_k}\|\mathrm{Row}_t[V^{k\ell}]\|_1, & 1\le k,\ell\le K,\\ \|\mathrm{Col}_\ell[\Omega]\|_{s,q}\le s^{1/q-1}\kappa, & 1\le\ell\le K,\end{cases} \tag{5.8}$$

$$\mathbf{T}_\varepsilon : \mathrm{Erfinv}\left(\frac{\varepsilon}{2N}\right)\|D^T h^{kt}\|_2 + \|E^T h^{kt}\|_2 \le \rho, \qquad 1 \le t \le n_k, 1 \le k \le K.$$

The case of $r = 2$. Here

$$\mathbf{S}_{\kappa,q} : \begin{cases} B = VB + H^T A, & \\ \Omega_{k\ell} := \|V^{k\ell}\|_{2,2} = \sigma_{\max}(V^{k\ell}), & 1 \le k, \ell \le K, \\ \|\mathrm{Col}_\ell[\Omega]\|_{s,q} \le s^{1/q-1}\kappa, & 1 \le \ell \le K, \end{cases}$$

(5.9)
$$\mathbf{T}_\varepsilon : \exists \{W_k \in \mathbb{S}^m, \alpha_k, \beta_k, \gamma_k \in \mathbb{R}\}_{k=1}^K :$$
$$\begin{cases} \sigma_{\max}(E^T H[k]) + \alpha_k \le \rho, \\ \begin{bmatrix} W_k & D^T H[k] \\ H^T[k] D & \alpha_k I_{n_k} \end{bmatrix} \succeq 0, \\ \|\lambda(W_k)\|_\infty \le \beta_k, \qquad \|\lambda(W_k)\|_2 \le \gamma_k, \\ \mathrm{Tr}(W_k) + 2[\delta\beta_k + \sqrt{\delta^2\beta_k^2 + 2\delta\gamma_k^2}] \le \alpha_k, \end{cases} \qquad 1 \le k \le K,$$
$$\delta := \ln(K/\varepsilon),$$

where $\mathbb{S}^m$ is the space of $m \times m$ symmetric matrices, and $\lambda(W)$ is the vector of eigenvalues of $W \in \mathbb{S}^m$.

The case of $r = 1$. Here

$$\mathbf{S}_{\kappa,q} : \begin{cases} B = VB + H^T A, & \\ \Omega_{k\ell} := \|V^{k\ell}\|_{1,1} = \max\limits_{1 \le t \le n_\ell} \|\mathrm{Col}_t[V^{k\ell}]\|_1, & 1 \le k, \ell \le K, \\ \|\mathrm{Col}_\ell[\Omega]\|_{s,q} \le s^{1/q-1}\kappa, & 1 \le \ell \le K, \end{cases}$$

(5.10)
$$\mathbf{T}_\varepsilon : \exists \{\lambda^k \in \mathbb{R}^m_+, \mu^k \ge 0\}_{k=1}^K \qquad \forall (k \le K, t \le n_k)$$
$$\begin{cases} \mathrm{Erfinv}\left(\frac{\varepsilon}{2Kn_k}\right) \sum_{t=1}^{n_k} \|D^T h^{kt}\|_2 + \frac{1}{2}\sum_i \lambda_i^k + \frac{1}{2}\mu^k \le \rho, \\ \begin{bmatrix} \mathrm{Diag}\{\lambda^k\} & H^T[K]E \\ E^T H[k] & \mu^k I_{n_k} \end{bmatrix} \succeq 0. \end{cases}$$

5.3. *Tractable sufficient conditions*: *Limits of performance*. Consider the situation where all the norms $\|\cdot\|_{(k)}$ are $\|\cdot\|_r$, with $r \in \{1, 2, \infty\}$. A natural question about verifiable sufficient conditions for a pair $(H, L_\infty(\cdot))$ to satisfy $\mathbf{Q}_{s,q}(\kappa)$ is, what are the "limits of performance" of these sufficient conditions? Specifically, how large could be the range of s for which the condition can be satisfied by at least one contrast matrix? Here is a partial answer to this question:

PROPOSITION 5.3. *Let A be an $m \times n$ sensing matrix which is "essentially nonsquare," specifically, such that $2m \le n$, let the r.s. $\mathcal{S}$ be such that $B = I_n$, and let $n_k = d$, $\|\cdot\|_{(k)} = \|\cdot\|_r$, $1 \le k \le K$, with $r \in \{1, 2, \infty\}$. Whenever an $m \times n$ matrix H and $n \times n$ matrix V satisfy the conditions*

$$
\begin{aligned}
&I = V + H^T A \quad \textit{and} \\
&\max_{1\le \ell\le K} \|[\|V^{1\ell}\|_{r,r}; \|V^{2\ell}\|_{r,r}; \dots; \|V^{K\ell}\|_{r,r}]\|_{s,q} \le \frac{1}{2} s^{1/q-1}
\end{aligned}
\tag{5.11}
$$

[*cf.* (5.1), (5.5) *and* (5.4)] *with $q \ge 1$, one has*

$$
s \le \frac{3\sqrt{m}}{2\sqrt{d}}. \tag{5.12}
$$

Discussion. Let the r.s. $\mathcal{S}$ in question be the same as in Proposition 5.3, and let $m \times n$ sensing matrix A have $2m \le n$. Proposition 5.3 says that in this case, the verifiable sufficient condition, stated by Proposition 5.1, for satisfiability of $\mathbf{Q}_{s,q}(\kappa)$ with $\kappa < 1/2$ has rather restricted scope—it cannot certify the satisfiability of $\mathbf{Q}_{s,q}(\kappa)$, $\kappa \le 1/2$, when $s \ge \frac{3\sqrt{m}}{2\sqrt{d}}$. Yet, the condition $\mathbf{Q}_{s,q}(\kappa)$ *may be satisfiable* in a much larger range of values of s. For instance, when the r.s. in question is the standard one and A is a random Gaussian $m \times n$ matrix, the matrix A satisfies, with overwhelming probability as m, n grow, the RIP$(\frac{1}{5}, s)$ condition for s as large as $O(1)m/\sqrt{\ln(n/m)}$ (cf. [8]). By Proposition 2.1, this implies that $(\frac{5}{4}A, \|\cdot\|_\infty)$ satisfies the condition $\mathbf{Q}_{s,2}(\frac{1}{4})$ in essentially the same large range of s. There is, however, an important case where the "limits of performance" of our *verifiable sufficient condition* for the satisfiability of $\mathbf{Q}_{s,q}(\kappa)$ implies severe restrictions on the range of values of s in which the "true" condition $\mathbf{Q}_{s,q}(\kappa)$ is satisfiable—this is the case when $q = \infty$ and $r = \infty$. Combining Propositions 4.1 and 5.3, we conclude that *in the case of r.s. from Proposition* 5.3 *with $r = \infty$ and "sufficiently nonsquare"* ($2m \le n$) *$m \times n$ sensing matrix A, the associated condition $\mathbf{Q}_{s,\infty}(\frac{1}{2})$ cannot be satisfied when $s > \frac{3\sqrt{m}}{2\sqrt{d}}$.*

5.4. *Tractable sufficient conditions and Mutual Block-Incoherence.* We have mentioned in the Introduction that, to the best of our knowledge, the only previously proposed verifiable sufficient condition for the validity of ℓ_1 block recovery is the "mutual block incoherence condition" of [12]. Our immediate goal is to show that this condition is covered by Proposition 5.1.

Consider an r.s. with $B = I_n$ and with ℓ_2-norms in the role of $\|\cdot\|_{(k)}$, $1 \le k \le K$, and let the sensing matrix A in question be partitioned as $A = [A[1], \dots, A[K]]$, where $A[k]$ has n_k columns. Let us define the *mutual block-incoherence μ* of A w.r.t. the r.s. in question as follows:

$$
\mu = \max_{\substack{1\le k,\ell\le K,\\ k\ne\ell}} \sigma_{\max}\big(C_k^{-1} A^T[k] A[\ell]\big) \qquad \big[\text{where } C_k := A^T[k]A[k]\big], \tag{5.13}
$$

provided that all matrices C_k, $1 \le k \le K$, are nonsingular, otherwise $\mu = \infty$. Note that in the case of the standard r.s., the just defined quantity is nothing but the standard mutual incoherence known from the Compressed Sensing literature (see, e.g., [10]).

In [12], the authors consider the same r.s. and assume that $n_k = d$, $1 \le k \le K$, and that the columns of A are of unit $\|\cdot\|_2$-norm. They introduce the quantities

$$\begin{aligned} \nu &= \max_{1\le k\le K} \max_{1\le j\ne j'\le K} \left|\mathrm{Col}_j^T[A[k]]\mathrm{Col}_{j'}[A[k]]\right|, \\ \mu_B &= \frac{1}{d} \max_{\substack{1\le k,\ell\le K,\\ k\ne\ell}} \sigma_{\max}\left(A^T[k]A[\ell]\right) \end{aligned} \tag{5.14}$$

and prove that an appropriate version of block-ℓ_1 recovery allows to recover exactly every s-block-sparse signal x from the noiseless observations $y = Ax$, provided that

$$1 - (d-1)\nu > 0 \quad \text{and} \quad s < \chi := \frac{1-(d-1)\nu + d\mu_B}{2d\mu_B}. \tag{5.15}$$

The following observation is almost immediate:

PROPOSITION 5.4. *Given an $m \times n$ sensing matrix A and an r.s. $\mathcal{S}$ with $B = I_n$, $\|\cdot\|_{(k)} = \|\cdot\|_2$, $1 \le k \le K$, let $A = [A[1], \dots, A[K]]$ be the corresponding partition of A.*

(i) *Let μ be the mutual block-incoherence of A w.r.t. $\mathcal{S}$. Assuming $\mu < \infty$, we set*

$$H = \frac{1}{1+\mu}\left[A[1]C_1^{-1}, A[2]C_2^{-1}, \dots, A[K]C_K^{-1}\right] \quad \text{where } C_k = A^T[k]A[k]. \tag{5.16}$$

Then the contrast matrix H along with the matrix $I_n - H^T A$ satisfies condition (5.1) (where $B = I_n$) and condition (5.6) with $q = \infty$ and

$$\kappa = \frac{\mu s}{1+\mu}.$$

As a result, applying Proposition 5.1, we conclude that whenever

$$s < \frac{1+\mu}{2\mu}, \tag{5.17}$$

the pair $(H, L_\infty(\cdot))$ satisfies $\mathbf{Q}_{s,\infty}(\kappa)$ with $\kappa = \frac{\mu s}{1+\mu} < 1/2$.

(ii) *Suppose that $n_k = d$, $k = 1, \dots, K$, and let the quantities ν and μ_B defined in (5.14) satisfy the relations (5.15). Then the mutual block-incoherence of A w.r.t. the r.s. in question does not exceed $\bar\mu = \frac{d\mu_B}{1-(d-1)\nu}$. Furthermore, we have $\frac{1+\bar\mu}{2\bar\mu} = \chi$, and (5.17) holds, and thus ensures that the contrast H, as defined in (5.16), and $L_\infty(\cdot)$ satisfy $\mathbf{Q}_{s,\infty}(\kappa)$ with some $\kappa < \frac{1}{2}$.*

Let $A = [A_{ij}] \in \mathbb{R}^{m\times n}$ be a random matrix with i.i.d. entries $A_{ij} \sim \mathcal{N}(0, m^{-1})$. We have the following simple result.

PROPOSITION 5.5. *Assume that $B = I_n$, $n_k = d$ and $\|\cdot\|_{(k)} = \|\cdot\|_2$ for all k. There are absolute constants $C_1, C_2 < \infty$ (the corresponding bounds are provided in Section* A.10 *of the supplementary article* [18]) *such that if $m \geq C_1(d + \ln(n))$, then the mutual block-incoherence μ of A satisfies*

$$\mu \leq C_2\sqrt{\frac{d+\ln(n)}{m}} \tag{5.18}$$

with probability at least $1 - \frac{1}{n}$.

The bound (5.18), along with Proposition 5.4(i), implies that when A is a Gaussian matrix, all block-norms are the ℓ_2-norms and all $n_k = d$ with d "large enough" [such that $d^{-1}\ln n = O(1)$], the verifiable sufficient condition for $\mathbf{Q}_{s,\infty}(\frac{1}{3})$ holds with overwhelming probability for $s = O(\sqrt{\frac{m}{d}})$. In other words, in this case the (verifiable!) condition $Q_{s,\infty}(\kappa)$ attains (up to an absolute factor) the limit of performance stated in Proposition 5.3.

6. Matching pursuit algorithm for block recovery. The Matching Pursuit algorithm for block-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 for block-sparse recovery have been proposed in the Compressed Sensing literature [3, 4, 12, 13]. In this section we aim to show that a pair H, V satisfying (5.1) and (5.4) where $\kappa < 1/2$ [and thus, by Proposition 5.1, such that $(H, L_\infty(\cdot))$ satisfies $\mathbf{Q}_{s,\infty}(\kappa)$] can be used to design a specific version of the Matching Pursuit algorithm which we refer to as the *non-Euclidean Block Matching Pursuit* (*NEBMP*) *algorithm* for block-sparse recovery.

We fix an r.s. $\mathcal{S} = (B, n_1, \dots, n_K, \|\cdot\|_{(1)}, \dots, \|\cdot\|_{(K)})$ and assume that the block norms $\|\cdot\|_{(k)}$, $k = 1, \dots, K$, are either $\|\cdot\|_\infty$- or $\|\cdot\|_2$-norms. Furthermore, we suppose that the matrix B is of full row rank, so that, given $z \in \mathbb{R}^N$, one can compute x such that $z = Bx$ [e.g., $x = B^+z$ where $B^+ = B^T(BB^T)^{-1}$ is the pseudo-inverse of B]. Let the noise ξ in the observation $y = Ax + u + \xi$ be Gaussian, $\xi \sim \mathcal{N}(0, D)$, $D \in \mathbb{R}^{m\times m}$ is known. Finally, we assume that we are in the situation of Section 5.2, that is, we have at our disposal an $m \times N$, $N = n_1 + \cdots + n_K$, matrix H, an $N \times N$ block matrix $V = [V^{k\ell} \in \mathbb{R}^{n_k\times n_\ell}]_{k,\ell=1}^K$, a $\bar{\gamma} > 0$ and $\rho \geq 0$ such that

$$\begin{aligned}
&\text{(a)}\quad B = VB + H^TA,\\
&\text{(b)}\quad \|V^{k\ell}\|_{(\ell,k)} = [\Omega]_{k,\ell} \leq \bar{\gamma} \qquad \forall k, \ell \leq K,\\
&\text{(c)}\quad \mathrm{Prob}_\xi\{\Xi^+ := \{\xi : L_\infty(H^T[u+\xi]) \leq \rho\ \forall u \in \mathcal{U}\}\} \geq 1 - \varepsilon.
\end{aligned} \tag{6.1}$$

Algorithm 1 Non-Euclidean Block Matching Pursuit

1. *Initialization*: Set $\upsilon^{(0)} = 0$, $\alpha_0 = \frac{L_{s,1}(H^T y)+s\rho+\upsilon}{1-s\bar{\gamma}}$.
2. *Step k*, $k = 1, 2, \dots$: Given $\upsilon^{(k-1)} \in \mathbb{R}^n$ and $\alpha_{k-1} \geq 0$, compute

   2.1. $g = H^T(y - A\upsilon^{(k-1)})$ and vector $\Delta = [\Delta[1], \dots, \Delta[K]] \in \mathbb{R}^N$ by setting for $j = 1, \dots, K$:

$$\Delta[j] = \frac{g[j]}{\|g[j]\|_2}\big[\|g[j]\|_2 - \bar{\gamma}\alpha_{k-1} - \rho\big]_+ \qquad \text{if } \|\cdot\|_{(j)} = \|\cdot\|_2;$$

(6.2)
$$\Delta_{ji} = \operatorname{sign}(g_{ji})\big[|g_{ji}| - \bar{\gamma}\alpha_{k-1} - \rho\big]_+, \qquad 1 \leq i \leq n_j, \qquad \text{if } \|\cdot\|_{(j)} = \|\cdot\|_\infty,$$

   where w_{ji} is ith entry in jth block of a representation vector w and $[a]_+ = \max\{a, 0\}$.

   2.2. Choose $\upsilon^{(k)}$ such that $B(\upsilon^{(k)} - \upsilon^{(k-1)}) = \Delta$, set

(6.3)
$$\alpha_k = 2s\bar{\gamma}\alpha_{k-1} + 2s\rho + \upsilon.$$

   and loop to step $k + 1$.

3. *Output*: The approximate solution found after k iterations is $\upsilon^{(k)}$.

Given observation y, a positive integer s and a real $\upsilon \geq 0$ [υ is our guess for an upper bound on $L_1(Bx - [Bx]^s)$], consider Algorithm 1. Its convergence analysis is based upon the following:

LEMMA 6.1. *In the situation of* (6.1), *let $s\bar{\gamma} < 1$. Then whenever $\xi \in \Xi^+$, for every $x \in \mathbb{R}^n$ with $L_1(Bx - [Bx]^s) \leq \upsilon$ and every $u \in \mathcal{U}$, the following holds true. When applying Algorithm 1 to $y = Ax + u + \xi$, the resulting approximations $B\upsilon^{(k)}$ to Bx and the quantities α_k for all k satisfy the relations*

$$(\mathrm{a}_k) \quad \textit{for all } 1 \leq j \leq K \qquad \big\|(B\upsilon^{(k)} - Bx)[j]\big\|_{(j)} \leq \big\|(Bx)[j]\big\|_{(j)},$$

$$(\mathrm{b}_k) \quad L_1\big(Bx - B\upsilon^{(k)}\big) \leq \alpha_k \quad \textit{and} \quad L_\infty\big(Bx - B\upsilon^{(k+1)}\big) \leq 2\bar{\gamma}\alpha_k + 2\rho.$$

Note that if $2s\bar{\gamma} < 1$, then also $s\bar{\gamma} < 1$, so that Lemma 6.1 is applicable. Furthermore, in this case, by (6.3), the sequence α_k converges exponentially fast to the limit $\alpha_\infty := \frac{2s\rho+\upsilon}{1-2s\bar{\gamma}}$:

$$L_1\big(B\upsilon^{(k)} - Bx\big) \leq \alpha_k = (2s\bar{\gamma})^k[\alpha_0 - \alpha_\infty] + \alpha_\infty.$$

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

$$L_\infty\big(B\upsilon^{(k)} - Bx\big) \leq 2\bar{\gamma}\alpha_{k-1} + 2\rho \leq \frac{\alpha_k}{s},$$

and since $L_p(w) \leq L_1(w)^{1/p} L_\infty(w)^{(p-1)/p}$ for $1 \leq p \leq \infty$, we have

$$L_p\big(B\upsilon^{(k)} - Bx\big) \leq s^{(1-p)/p}\big[(2s\bar{\gamma})^k[\alpha_0 - \alpha_\infty] + \alpha_\infty\big].$$

The bottom line here is as follows.

PROPOSITION 6.1. *Suppose that a collection* $(H, L_\infty(\cdot), \rho, \bar{\gamma}, \varepsilon)$ *satisfies* (6.1), *and let the parameter* s *of Algorithm* 1 *satisfy* $2\kappa := 2s\bar{\gamma} < 1$. *Then for all* $\xi \in \Xi^+$, $u \in \mathcal{U}$, $x \in \mathbb{R}^n$ *such that* $L_1(Bx - [Bx]^s) \le \upsilon$, *Algorithm* 1 *as applied to* $y = Ax + u + \xi$ *ensures that for every* $t = 1, 2, \ldots$ *one has*

$$L_p(Bv^{(t)} - Bx) \le s^{1/p}\left[\frac{2\rho + s^{-1}\upsilon}{1 - 2\kappa} + (2\kappa)^t\left(\frac{s^{-1}(L_{s,1}(H^T y) + \upsilon) + \rho}{1 - \kappa} - \frac{2\rho + s^{-1}\upsilon}{1 - 2\kappa}\right)\right]$$

for all $1 \le p \le \infty$*'s* [cf. (4.2)].

Note that Proposition 6.1 combined with Proposition 5.4 essentially covers the results of [12] on the properties of the Matching Pursuit algorithm for the block-sparse recovery proposed in this reference.

7. Numerical illustration. In the theoretical part of this paper we considered the situation where the sensing matrix A and the r.s. $\mathcal{S} = (B, n_1, \ldots, n_K, \|\cdot\|_{(1)}, \ldots, \|\cdot\|_{(K)})$ were given, and we were interested in understanding:

(A) whether ℓ_1 recovery allows to recover the representations Bx of all s-block-sparse signals with a given s in the absence of observation noise, and

(B) how to choose the best (resulting in the smallest possible error bounds) pair $(H, \|\cdot\|)$.[7]

Note that our problem setup involves a number of components. While in typical applications sensing matrix A, representation matrix B and the dimensions $n_1, \ldots, n_K$ of the block vectors may be thought as given by the "problem's physics," it is not the case for the block norms $\|\cdot\|_{(k)}$. Their choice (which does affect the ℓ_1 recovery routines) appears to be unrelated to the model of the data.

The first goal of our experiments is to understand how to choose the block norms in order to validate ℓ_1 recovery for the largest possible value of the sparsity parameter s; here "to validate" means to provide guarantees of small recovery error for all s-block-sparse signals when the observation error is small (which implies, of course, the exactness of the recovery in the case of noiseless observation). Here we restrict ourselves to the case of ℓ_r-r.s. with $r \in \{1, 2, \infty\}$. By reasons explained in the discussion in Section 3, we consider here only the case of the penalized ℓ_1 recovery with $m \times N$ contrast matrix H (where, as always, $N = n_1 + \cdots + n_K$),

[7]Needless to say, the results presented so far do not pretend to provide full answers to these questions. Our verifiable sufficient conditions for the validity of ℓ_1 block recovery supply only *lower bounds* on the largest $s = s_*$ for which the answer to (A) is positive. Similarly, aside of the case $q = \infty$, $\|\cdot\|_{(k)} = \|\cdot\|_\infty$, $1 \le k \le K$, our conditions for the validity of block-ℓ_1 recovery are only sufficient, meaning that optimizing the error bound over $(H, \|\cdot\|)$ allowed by these conditions may only yield *sub*optimal recovery routines.

$\|\cdot\| = L_\infty(\cdot)$,[8] and with $\lambda = 2s$ [see (3.5)]. Besides this, we assume, mainly for the sake of notational convenience, that $B = I_n$.

Let us fix $A \in \mathbb{R}^{m\times n}$, $B = I_n, K, n_1, \ldots, n_K$ ($n_1 + \cdots + n_K = n =: N$). By Proposition 5.1, for every matrix $H \in \mathbb{R}^{m\times n}$ setting

$$
\begin{aligned}
V &\equiv \left[V^{k\ell} \in \mathbb{R}^{n_k\times n_\ell}\right]_{k\ell=1}^K = I - H^T A,\\
\Omega^r(H) &= \left[\|V^{k\ell}\|_{r,r}\right]_{k,\ell=1}^K,\\
\kappa_1^{r,s}(H) &= \max_{1\le \ell\le K} \|\mathrm{Col}_\ell[\Omega^r(H)]\|_{s,1},\\
\kappa_\infty^{r,s}(H) &= s \max_{1\le k,\ell\le K} \left[\Omega^r(H)\right]_{k,\ell},
\end{aligned}
\tag{7.1}
$$

the pair $(H, L_\infty(\cdot))$ satisfies the conditions $\mathbf{Q}_{s,q}(\kappa_q^{r,s}(H))$, $q = 1$ and $q = \infty$, provided that the block norms are the ℓ_r-ones. In particular, when $\kappa_1^{r,s}(H) < 1/2$, the penalized ℓ_1/ℓ_r recovery [i.e., the recovery (3.1) with all block norms being the ℓ_r-ones] "is valid" on s-block-sparse signals, meaning exactly that this recovery ensures the validity of the error bounds (3.8) with $q = \infty$, $\varkappa = \kappa_1^{r,s}$, $\kappa = \kappa_\infty^{r,s}$ (and, in particular, recovers exactly all s-block-sparse signals when there is no observation noise).

Our strategy is as follows. For each value of $r \in \{1, 2, \infty\}$, we consider the convex optimization problem

$$
\min_{H\in\mathbb{R}^{m\times n}} \left\{\kappa_1^{r,s}(H) := \max_{\ell\le K} \|\mathrm{Col}_\ell[\Omega^r(H)]\|_{s,1}\right\},
$$

find the largest $s = s(r)$ for which the optimal value in this problem is $< 1/2$, and denote by $H^{(r)}$, $r \in \{1, 2, \infty\}$ the corresponding optimal solution. In addition to these "marked" contrast matrices, we consider two more contrasts, $H^{(\mathrm{MI})}$ and $H^{(\mathrm{MBI})}$, based on the mutual block-incoherence condition and given by the calculation (5.13) for the cases of the "standard" (1-element blocks in $x = Bx$) and the actual block structures, respectively.

Now, given the set $\mathcal{H} = \{H^{(\mathrm{MI})}, H^{(\mathrm{MBI})}, H^{(1)}, H^{(2)}, H^{(\infty)}\}$ of $m \times n$ candidate contrast matrices, we can choose the "most powerful" penalized ℓ_1/ℓ_r recovery suggested by $\mathcal{H}$ as follows: for every $H \in \mathcal{H}$ and for every $p \in \{1, 2, \infty\}$, we find the largest $s = s(H, p)$ for which $\kappa_1^{r,p}(H) < 1/2$, and then define the quantity $s_* = s_*(\mathcal{H}) = \max\{s(H, p) : H \in \mathcal{H}, p \in \{1, 2, \infty\}\}$ along with $H_* \in \mathcal{H}$ and $p_* \in \{1, 2, \infty\}$ such that $s_* = s(H_*, p_*)$. The penalized ℓ_1/ℓ_{p_*} recovery utilizing the contrast matrix H_* and the norm $L_\infty(\cdot)$ associated with block norms $\|\cdot\|_{p_*}$ of the blocks is definitely valid for $s = s_*(\mathcal{H})$, and this is the largest sparsity range, as certified by our sufficient conditions, for the validity of ℓ_1/ℓ_r recovery, which we can get with contrast matrices from $\mathcal{H}$. Note that $s_* \ge \max[s(1), s(2), s(\infty)]$, that

[8]These are exactly the pairs $(H, \|\cdot\|)$ covered by the sufficient conditions for the validity of ℓ_1 recovery; see Proposition 5.1.

is, the resulting range of values of s is also the largest we can certify using our sufficient conditions, with *no* restriction on the contrast matrices.

Implementation. We have tested the outlined strategy in the following problem setup:

- the sensing matrices A are of *size* $(m = 96) \times (n = 128)$, $B = I$ with $K = 32$ four-element blocks in $Bx = x$;
- the 96×128 *sensing matrices* A are built as follows: we first draw a matrix at random from one of the following distributions:
  - *type* `H`: randomly selected 96×128 submatrix of the 128×128 Hadamard matrix,[9]
  - *type* `G`: 96×128 matrix with independent $\mathcal{N}(0, 1)$ entries,
  - *type* `R`: 96×128 matrix with independent entries taking values ± 1 with equal probabilities,
  - *type* `T`: random 96×128 matrix of the structure arising in *Multi-Task Learning* (see, e.g., [1] and references therein): the consecutive 4-column parts of the matrix are block-diagonal with four 24×1 diagonal blocks with independent $\mathcal{N}(0, 1)$ entries,

  and then scale the columns of the selected matrix to have their $\|\cdot\|_2$-norms equal to 1.

The results we report describe 4 experiments differing from each other by the type of the (randomly selected) matrix A.[10]

In Table 1, we display the certified sparsity levels of penalized ℓ_1/ℓ_r recoveries for the candidate contrast matrices. In addition, we present valid *upper bounds* $\bar{s}(r)$ on the "r-goodness" $s^*(A, r)$ of A, defined as the largest s such that the ℓ_1/ℓ_r recovery in the noiseless case recovers *exactly* the representations of *all* s-block-sparse vectors, that is,

$$s^*(A, r) = \max\left\{s : x = \operatorname*{Arg\,min}_{z \in \mathbb{R}^n}\left\{\sum_{k=1}^{K} \|[z]_k\|_r : Az = Ax\right\}\right.$$
$$\left.\text{for all } s\text{-block-sparse } x.\right\}$$

[9] The Hadamard matrices H_k of order $2^k \times 2^k$, $k = 0, 1, \dots$, are given by the recurrence $H_0 = 1$, $H_{k+1} = [H_k, H_k; H_k, -H_k]$. They are symmetric matrices with ± 1 entries and rows orthogonal to each other.

[10] As far as our experience shows, the results remain nearly the same across instances of A drawn from the same distribution, so that only one experiment for each type of distribution in question appears to be representative enough.

TABLE 1

Certified sparsity levels for penalized ℓ_1/ℓ_r-recoveries for candidate contrast matrices. For each candidate and each value of r we present in the corresponding cells the triple $s(H,r)|\kappa_1^{r,s(H,r)}(H)|\kappa_\infty^{r,s(H,r)}(H)$. $\bar{s}(r)$: a computed upper bound on r-goodness $s^*(A,r)$ of A. Italic: the best sparsity $s_*(\mathcal{H})$ certified by our sufficient conditions for the validity of penalized recovery

A	**r**		$\boldsymbol{H^{(\mathrm{MI})}}$			$\boldsymbol{H^{(\mathrm{MBI})}}$			$\boldsymbol{H^{(1)}}$	
H	1	2	0.4727	0.509	2	0.444	0.460	*3*	0.429	0.429
	2	2	0.436	0.436	2	0.429	0.429	*3*	0.429	0.429
	∞	2	0.473	0.509	2	0.444	0.460	*3*	0429	0.429
G	1	0	0.000	0.000	0	0.000	0.000	*3*	0.467	0.900
	2	0	0.000	0.000	1	0.368	0.368	1	0.300	0.300
	∞	0	0.000	0.000	0	0.000	0.000	0	0.000	0.000
R	1	0	0.0000	0.000	0	0.000	0.000	*3*	0.477	0.853
	2	0	0.000	0.000	1	0.354	0.354	1	0.284	0.284
	∞	0	0.000	0.000	0	0.000	0.000	1	0.482	0.482
T	1	1	0.384	0.384	1	0.399	0.399	*2*	0.383	0.383
	2	1	0.384	0.384	1	0.399	0.399	*2*	0.383	0.383
	∞	1	0.384	0.384	1	0.399	0.399	*2*	0.383	0.383

A	**r**		$\boldsymbol{H^{(2)}}$			$\boldsymbol{H^{(\infty)}}$		$\boldsymbol{\bar{s}(r)}$
H	1	2	0.487	0.519	*3*	0.429	0.429	4
	2	*3*	0.429	0.429	*3*	0.429	0.429	3
	∞	2	0.487	0.519	*3*	0.429	0.429	3
G	1	1	0.301	0.301	1	0.489	0.489	5
	2	*3*	0.447	0.458	2	0.479	0.549	5
	∞	1	0.305	0.305	*3*	0.483	0.823	4
R	1	1	0.291	0.291	1	0.498	0.498	5
	2	*3*	0.438	0.440	1	0.264	0.264	5
	∞	1	0.286	0.286	*3*	0.489	0.739	5
T	1	*2*	0.383	0.383	*2*	0.383	0.383	3
	2	*2*	0.383	0.383	*2*	0.383	0.383	3
	∞	*2*	0.383	0.383	*2*	0.383	0.383	3

We present on Figure 1 examples of "bad" signals [i.e., $(\bar{s}(r)+1)$-block-sparse signals which are *not* recovered correctly by the latter procedure].[11]

On the basis of this experiment we can make two tentative conclusions:

[11] It is immediately seen that whenever B is of full row rank, the *nullspace property* "$L_{s,1}(Bx) < \frac{1}{2}L_1(Bx)$ for all $x \in \operatorname{Ker} A$ with $Bx \neq 0$" is necessary for s to be $\leq s^*(A,\cdot)$. As a result, for B's of full row rank, $s^*(A,r)$ can be upper-bounded in a manner completely similar to the case of the standard r.s.; see [21], Section 4.1.

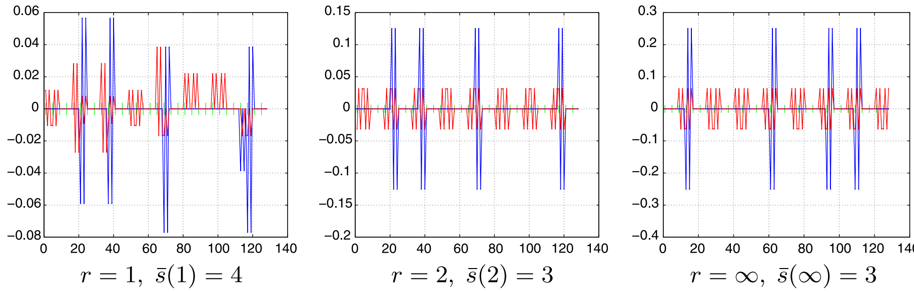

FIG. 1. *"Bad" $(\bar{s}(r)+1)$-block-sparse signals (blue) and their ℓ_1/ℓ_r recoveries (red) from noiseless observations*, H-*matrix* A.

- the ℓ_1/ℓ_2 recovery with the contrast matrix $H^{(2)}$ and the ℓ_1/ℓ_∞ recovery with the contrast matrix $H^{(\infty)}$ were able to certify the best levels of allowed sparsity (when compared to other candidate matrices from $\mathcal{H}$);
- in our experiments, the upper bounds $\bar{s}(r)$ on the r-goodness $s^*(A, r)$ of A are close to the corresponding certified lower bounds $s_*(\mathcal{H}, r) = \max_{H\in\mathcal{H}} s(H, r)$.

Numerical evaluation of recovery errors. The objective of the next experiment is to evaluate the accuracy of penalized ℓ_1/ℓ_r recoveries in the noisy setting. As above, we consider the contrast matrices from $\mathcal{H} = \{H^{(\mathrm{MI})}, H^{(\mathrm{MBI})}, H^{(1)}, H^{(2)}, H^{(\infty)}\}$. Note that it is possible to improve the error bound by optimizing it over H as it was done in Section 5.2. In the experiments to be reported this additional optimization, however, did not yield a significant improvement (which perhaps reflects the "nice conditioning" of the sensing matrices we dealt with), and we do not present the simulation results for optimized contrasts here:

- We ran four series of simulations corresponding to the four instances of the sensing matrix A we used. The series associated with a particular A was as follows:
- Given A, we associate with it the five aforementioned candidate contrast matrices from $\mathcal{H}$. Combining these matrices with 3 values of r ($r = 1, 2, \infty$), we get 15 recovery routines. In addition to these 15 routines, we also included the block Lasso recovery as described in [24]. In our notation, this recovery is (cf. [24], (2.2))

$$\widehat{x}_{\mathrm{Lasso}}(y) \in \operatorname*{Arg\,min}_z \left\{ \frac{1}{m}\|Az - y\|_2^2 + 2\sum_{k=1}^{K} \lambda_k \|z[k]\|_2 \right\}$$

  ($z[k]$, $1 \le k \le K$, are the blocks in $z = Bz$), with the penalty coefficients λ_k chosen according to the equality version of the relations in [24], Theorem 3.1, used with $q = 2$.

Each of the 16 resulting recovery routines was tested on two samples, each containing 100 randomly generated recovery problem instances. In each problem instance the true signal was randomly generated with s nonzero blocks, and the observations were corrupted by pure Gaussian white noise: $y = Ax + \sigma\xi$, $\xi \sim \mathcal{N}(0, I)$. In the first sample, s was set to the best value $s_*(\mathcal{H})$ of block sparsity we were able to certify; in the second, $s = 2s_*(\mathcal{H})$ was used. The parameter λ of the penalized recoveries was set to $2s$ (and thus was tuned to the actual sparsity of test signals). In both samples, we used $\sigma = 0.001$.

We compare the recovery routines on the basis of their *ratings* computed as follows: given a recovery problem instance from the sample, we applied to it every one of our 16 recovery routines and measured the 16 resulting $\|\cdot\|_\infty$-errors. Dividing the smallest of these errors by the error of a given routine, we obtain "the rating" of the routine in this particular simulation. Thus, all ratings are ≤ 1; and the routine which attains the best $\|\cdot\|_\infty$ recovery error for the current data is rated "1.0." For the remaining routines, the closer to 1 is the rating of the routine, the closer is the routine to the "winner" of the current simulation. The final rating of a given recovery routine is its average rating over all $800 = 4 \times 2 \times 100$ recovery problem instances processed in the experiment.

The resulting ratings are presented in Table 2. The "winner" is the routine associated with $r = 2$ and $H = H^{(2)}$. Surprisingly, the second best routine is associated with the same $r = 2$ and the simplest contrast $H^{(\mathrm{MI})}$, an outsider in terms of the data presented in Table 1. This inconsistency may be explained by the fact that the data in Table 1 describe the guaranteed worst-case behavior of our recovery routines, which may be quite different from their "average behavior," reflected by Table 2. Our tentative conclusion on the basis of the data from Tables 1 and 2 is that the penalized ℓ_1/ℓ_2 recovery associated with the contrast matrix $H^{(2)}$ may be favorable when recovery guarantees are to be associated with good numerical performance.

The above comparison was carried out for σ set to 0.001. The conducted experiments show that for the routines in question and our purely Gaussian model of observation errors, the recovery errors are, typically, proportional to σ. This is illustrated by the plots on Figure 2 where we traced the average (over 40 experiments for every grid value of σ) *signal-to-noise ratio* (the ratio of the $\|\cdot\|_\infty$-error of the recovery to σ) of our favorable recovery ($r = 2$, $H = H^{(2)}$) and the corresponding performance figure for block Lasso.

TABLE 2
Ratings of recovery routines

r	$H^{(\mathrm{MI})}$	$H^{(\mathrm{MBI})}$	$H^{(1)}$	$H^{(2)}$	$H^{(\infty)}$	**Lasso**
1	0.30	0.20	0.53	0.60	0.54	N/A
2	0.76	0.51	0.75	0.79	0.75	0.19
∞	0.25	0.18	0.44	0.48	0.44	N/A

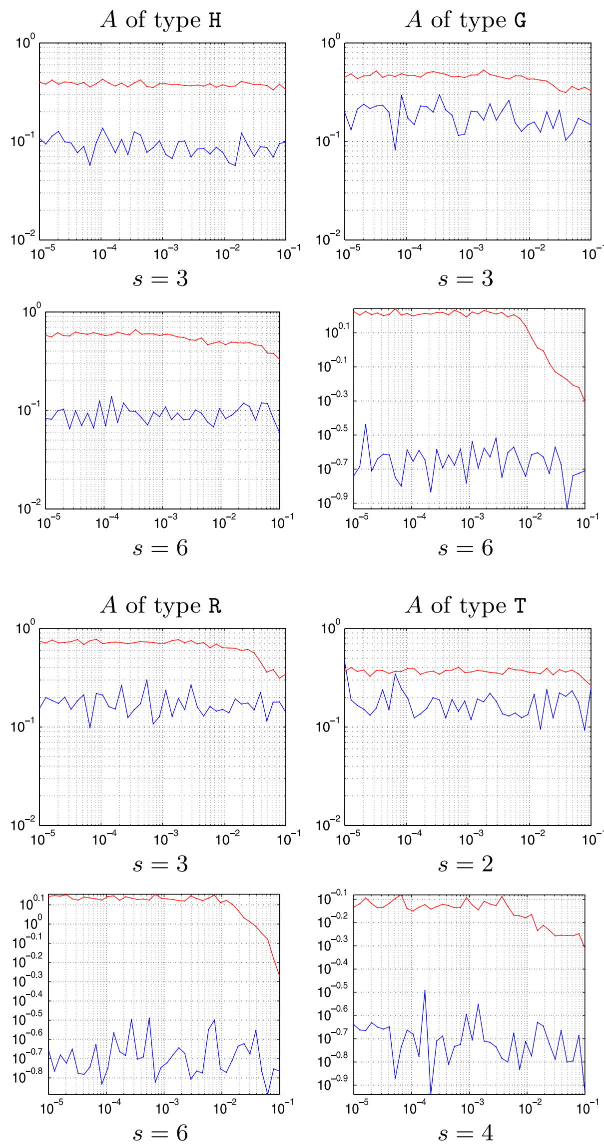

FIG. 2. *Average over* 40 *experiments ratio of* $\|\cdot\|_\infty$ *recovery error to* σ *vs.* σ. *In blue*: ℓ_1/ℓ_2 *recovery with* $H = H^{(2)}$; *in red*: *Lasso recovery.*

SUPPLEMENTARY MATERIAL

Supplement to "Accuracy guaranties for ℓ_1 recovery of block-sparse signals" (DOI: 10.1214/12-AOS1057SUPP; .pdf). The proofs of the results stated in the paper and the derivations for Section 5.2 are provided in the supplementary article [18].

REFERENCES

[1] ARGYRIOU, A., EVGENIOU, T. and PONTIL, M. (2008). Convex multi-task feature learning. *Machine Learning* **73** 243–272.

[2] BACH, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.* **9** 1179–1225. MR2417268

[3] BARANIUK, R. G., CEVHER, V., DUARTE, M. F. and HEGDE, C. (2010). Model-based compressive sensing. *IEEE Trans. Inform. Theory* **56** 1982–2001. MR2654489

[4] BEN-HAIM, Z. and ELDAR, Y. (2011). Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements. *IEEE J. Selected Topics in Signal Processing* **5** 1032–1047.

[5] BICKEL, P. J., RITOV, Y. and TSYBAKOV, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann. Statist.* **37** 1705–1732. MR2533469

[6] CANDES, E. and TAO, T. (2007). The Dantzig selector: Statistical estimation when p is much larger than n. *Ann. Statist.* **35** 2313–2351. MR2382644

[7] CANDÈS, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris* **346** 589–592. MR2412803

[8] CANDES, E. J. and TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Inform. Theory* **51** 4203–4215. MR2243152

[9] CHESNEAU, C. and HEBIRI, M. (2008). Some theoretical results on the grouped variables Lasso. *Math. Methods Statist.* **17** 317–326. MR2483460

[10] DONOHO, D. L., ELAD, M. and TEMLYAKOV, V. N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory* **52** 6–18. MR2237332

[11] DUARTE, M., BAJWA, W. and CALDERBANK, R. (2011). The performance of group Lasso for linear regression of grouped variables. Technical report 2010-10, Dept. Computer Science, Duke Univ., Durham, NC. Available at http://www.rci.rutgers.edu/~wub1/pubs/sampta11_tr.pdf.

[12] ELDAR, Y. C., KUPPINGER, P. and BÖLCSKEI, H. (2010). Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.* **58** 3042–3054. MR2730246

[13] ELDAR, Y. C. and MISHALI, M. (2009). Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inform. Theory* **55** 5302–5316. MR2596977

[14] GRIBONVAL, R. and NIELSEN, M. (2003). Sparse representations in unions of bases. *IEEE Trans. Inform. Theory* **49** 3320–3325. MR2045813

[15] HUANG, J. and ZHANG, T. (2010). The benefit of group sparsity. *Ann. Statist.* **38** 1978–2004. MR2676881

[16] JAMES, G. M., RADCHENKO, P. and LV, J. (2009). DASSO: Connections between the Dantzig selector and lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 127–142. MR2655526

[17] JUDITSKY, A., KILINÇ KARZAN, F. and NEMIROVSKI, A. (2011). On low rank matrix approximations with applications to synthesis problem in compressed sensing. *SIAM J. Matrix Anal. Appl.* **32** 1019–1029. MR2837590

[18] JUDITSKY, A., KILINÇ KARZAN, F., NEMIROVSKI, A. and POLYAK, V. (2013). Supplement to "Accuracy guaranties for ℓ_1 recovery of block-sparse signals." DOI:10.1214/12-AOS1057SUPP.

[19] JUDITSKY, A. and NEMIROVSKI, A. (2011). Accuracy guarantees for ℓ_1-recovery. *IEEE Trans. Inform. Theory* **57** 7818–7839. MR2895363

[20] JUDITSKY, A. B., KILINÇ-KARZAN, F. and NEMIROVSKI, A. S. (2011). Verifiable conditions of ℓ_1 recovery for sparse signals with sign restrictions. *Math. Program.* **127** 89–122. MR2776711

[21] JUDITSKY, A. B. and NEMIROVSKI, A. S. (2011). On verifiable sufficient conditions for sparse signal recovery via ℓ_1 minimization. *Math. Program.* **127** 57–88. MR2776710

[22] LIU, H. and ZHANG, J. (2009). Estimation consistency of the group Lasso and its applications. *J. Mach. Learn. Res. Proceedings Track* **5** 376–383.

[23] LIU, H., ZHANG, J., JIANG, X. and LIU, J. (2010). The group Dantzig selector. *J. Mach. Learn. Res. Proceedings Track* **9** 461–468.

[24] LOUNICI, K., PONTIL, M., VAN DE GEER, S. and TSYBAKOV, A. B. (2011). Oracle inequalities and optimal inference under group sparsity. *Ann. Statist.* **39** 2164–2204. MR2893865

[25] MEIER, L., VAN DE GEER, S. and BÜHLMANN, P. (2008). The group Lasso for logistic regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70** 53–71. MR2412631

[26] NARDI, Y. and RINALDO, A. (2008). On the asymptotic properties of the group lasso estimator for linear models. *Electron. J. Stat.* **2** 605–633. MR2426104

[27] OBOZINSKI, G., WAINWRIGHT, M. J. and JORDAN, M. I. (2011). Support union recovery in high-dimensional multivariate regression. *Ann. Statist.* **39** 1–47. MR2797839

[28] PARVARESH, F., VIKALO, H., MISRA, S. and HASSIBI, B. (2008). Recovering sparse signals using sparse measurement matrices in compressed DNA microarrays. *IEEE J. Selected Topics in Signal Processing* **2** 275–285.

[29] PFETSCH, M. E. and TILLMANN, A. M. (2012). The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. Technical report. Available at http://arxiv.org/abs/1205.2081.

[30] STOJNIC, M., PARVARESH, F. and HASSIBI, B. (2009). On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Signal Process.* **57** 3075–3085. MR2723043

[31] VAN DE GEER, S. A. and BÜHLMANN, P. (2009). On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.* **3** 1360–1392. MR2576316

[32] YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 49–67. MR2212574

A. JUDITSKY
LJK
UNIVERSITÉ J. FOURIER
B.P. 53
38041 GRENOBLE CEDEX 9
FRANCE
E-MAIL: juditsky@imag.com

F. KILINÇ KARZAN
TEPPER SCHOOL OF BUSINESS
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PENNSYLVANIA 15213
USA
E-MAIL: fkilinc@andrew.cmu.edu

A. NEMIROVSKI
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332
USA
E-MAIL: nemirovs@isye.gatech.edu

D. POLYAK
INSTITUTE OF CONTROL SCIENCES
OF RUSSIAN ACADEMY OF SCIENCES
MOSCOW 117997
RUSSIA
E-MAIL: boris@ipu.rssi.ru