# Verifiable conditions of $\ell_1$-recovery for sparse signals with sign restrictions*

Anatoli Juditsky Fatma Kılınç Karzan Arkadi Nemirovski

September 17, 2010

**Abstract**

We propose necessary and sufficient conditions for a sensing matrix to be "$s$-semigood" – to allow for exact $\ell_1$-recovery of sparse signals with at most $s$ nonzero entries under sign restrictions on part of the entries. We express error bounds for imperfect $\ell_1$-recovery in terms of the characteristics underlying these conditions. These characteristics, although difficult to evaluate, lead to verifiable sufficient conditions for exact sparse $\ell_1$-recovery and thus efficiently computable upper bounds on those $s$ for which a given sensing matrix is $s$-semigood. We examine the properties of proposed verifiable sufficient conditions, describe their limits of performance and provide numerical examples comparing them with other verifiable conditions from the literature.

## 1 Introduction

Assessing a sparse signal from an observation has been one of the main research areas in Compressed Sensing and sparse signal recovery. In practice, *a priori* information about the signal to be recovered often exists and will be beneficial if taken into account in the recovery procedure. In this paper, we suppose that the *a priori* information about a *sparse* signal $w \in \mathbf{R}^n$ amounts to the *sign restrictions*, and is given as the subsets $P_+$ and $P_-$ of $\{1, ..., n\}$, $P_+ \cap P_- = \emptyset$, such that $w_i \geq 0$ for $i \in P_+$ and $w_i \leq 0$ for $i \in P_-$. Therefore we address the following recovery problem: given an observation $y \in \mathbf{R}^m$,

$$y = Aw + e, \tag{1}$$

where $A \in \mathbf{R}^{m \times n}$ (in this context $m < n$) is a given matrix, $e \in \mathbf{R}^m$ is the observation error, assess a *sparse* signal $w \in \mathbf{R}^n$ satisfying *sign restrictions.*

A celebrated solution to the problem is given by the $\ell_1$-recovery, which amounts to taking, as an estimate of $w$, an optimal solution $\widehat{w}$ to the optimization problem

$$\widehat{w} \in \mathrm{Argmin}_x \left\{ \|x\|_1 : \ \|Ax - y\| \leq \varepsilon, \ x_i \geq 0 \ \forall i \in P_+, \ x_i \leq 0 \ \forall i \in P_- \right\} \tag{2}$$

(here $\varepsilon$ is an *a priori* bound on the norm $\|e\|$ of the observation error, $\|\cdot\|$ being some norm on $\mathbf{R}^m$). When there are no sign restrictions (i.e. $P_+ = P_- = \emptyset$), we arrive at the estimator playing the central role in the Compressive Sensing theory. The central result here is that when signal $w$

*Research of the second and the third authors was supported by the Office of Naval Research grant # N000140811104.

is $s$-sparse (i.e., with at most $s$ nonzero entries) and the matrix $A$ possesses a certain well-defined (although difficult to verify) property, then the $\ell_1$-recovery $\widehat{w}$ is close to $w$, provided the error bound $\varepsilon$ is small (for a comprehensive survey see [4] and references therein). Our goal here is to propose efficiently verifiable sufficient conditions on $A$ which allow for similar 'consistency" results, with emphasis on the case where sign restrictions are present.

To outline our results and to position them with respect to what is already known, let us start with noiseless recovery (i.e., $\varepsilon = 0$ and $y = Aw$). Here we are interested to answer the question:

> *Whether $A$ is such that whenever the true signal $w$ in (1) is $s$-sparse and satisfies the sign constraints $w_i \geq 0$, $i \in P_+$, $w_i \leq 0$, $i \in P_-$, the $\ell_1$-recovery*
>
> $$\widehat{w} \in \mathrm{Argmin}_x \{\|x\|_1 : \quad Ax = y,\ x_i \geq 0\ \forall i \in P_+,\ x_i \leq 0\ \forall i \in P_-\} \tag{3}$$
>
> *recovers $w$ exactly.*

If the answer is positive, we say that $A$ is *$s$-semigood*[1].

The theory of Compressive Sensing provides several sufficient/necessary and sufficient conditions for the $\ell_1$-recovery to be exact. For example, when no sign constraints are imposed on $w$, Donoho and Huo [9] prove that $A$ is $s$-good if for any set $I \subset \{1, ..., n\}$ of cardinality $\leq s$ it holds

$$\sum_{i \in I} |z_i| < \sum_{i \notin I} |z_i| \ \text{ for any } \ z \in \mathrm{Ker} A. \tag{4}$$

This condition has been extensively investigated. Its necessity has been established in [8]; it has been discussed in [16, 18] (under the name of *strict $s$-balancedness*), where its link to the geometric necessary and sufficient condition of $s$-goodness from [11] has been discussed. In [6], this condition has was also related to the sufficient condition ("*Null Space Property*") for successful combinatorial recovery.

The first characterization of $s$-semigoodness for the case when $w$ is nonnegative (i.e. $P_+ = \{1, ..., n\}$) was proposed in the founding paper of Donoho and Tanner [10] in terms of neighboring properties of the polytope $AS$, $S$ being the standard simplex $S = \{x \in \mathbf{R}^n : \ x \geq 0,\ \sum_i x_i \leq 1\}$. This paper contains also several important examples of $m \times n$ matrices which are $\lfloor \frac{m}{2} \rfloor$-semigood (here $\lfloor a \rfloor$ stands for the integer part of $a$) and demonstrates that various types of randomly generated matrices possess this property with overwhelming probability. Extending the results from Donoho and Huo [9], an equivalent characterization of $s$-semigoodness has been provided in the nonnegative case by Zhang in [17, 18], where it is shown that $A$ is $s$-semigood if and only if the kernel of $A$, $\mathrm{Ker}A$, is *strictly half $s$-balanced*, meaning that for any set $I \subset \{1, ..., n\}$ of cardinality $\leq s$ it holds

$$\sum_{i \in I} z_i < \sum_{i \notin I} |z_i| \ \text{ for any } \ z \in \mathrm{Ker} A \text{ such that } z_i \leq 0, \text{ for all } i \notin I. \tag{5}$$

It should be mentioned that the necessary and sufficient conditions for $s$-semigoodness from (4), (5) and [10, 11] share a common drawback – they seemingly cannot be verified in a computationally efficient way. To the best of our knowledge, the only *efficiently verifiable* conditions

[1] We use the term "$s$-semigoodness" to comply with the terminology of the companion paper [14], where we used the name *$s$-goodness* to indicate that $\ell_1$-recovery as in (3) *without* the sign restrictions is exact.

for $s$-semigoodness offered by the existing Compressive Sensing theory are the *sufficient* conditions based on the *mutual incoherence*

$$\mu(A) = \max_{i \neq j} \frac{|A_i^T A_j|}{A_i^T A_i} \tag{6}$$

where $A_i$ are columns of $A$ (assumed to be nonzero). Clearly, the mutual incoherence can be easily computed even for large matrices. Unfortunately, it turns out that that the estimates of "level of (semi)goodness" of a sensing matrix based on mutual incoherence usually are too conservative, in particular, they are provably dominated by the verifiable Linear Programming (LP) based sufficient conditions for $s$-goodness proposed in the companion paper [14] and based on characterization of $s$-goodness given in (4). Another verifiable sufficient condition for $s$-goodness, which uses the Semidefinite Programming (SDP) relaxation, has been recently proposed in [7].

The contributions of this paper, which follow the approach developed in [14], are as follows.

1. Taking existing characterizations of (semi)goodness (4), (5) as a starting point, we develop in Section 2, several equivalent necessary and sufficient conditions for $s$-semigoodness of a matrix $A$ in the case of general-type sign restrictions. Then in Section 3, we establish error bounds for inexact $\ell_1$-recovery (noisy observation (1), imprecise optimization in (2), nearly-sparse true signals); these bounds are expressed in the same terms as the necessary and sufficient conditions for $s$-semigoodness from Section 2. These bounds can be seen as an extension to the sign restricted case of bounds of Section 3 in [14] and as a special case of the bounds provided in Theorem 4.1 of [18]. To the best of our knowledge, these bounds that incorporate sign information of the signal are new.

2. The major goal of this paper is to use the LP relaxation techniques from [14] to derive novel *efficiently verifiable* sufficient conditions for $s$-semigoodness. These conditions allow one to build, in a computationally efficient fashion, lower bounds on the "level of $s$-semigoodness" of a given matrix $A$, that is, on the largest $s = s_*(A)$ for which $A$ is $s$-semigood with respect to given $P_\pm$. Some properties of these verifiable conditions, same as limits of their performance, are studied in Sections 4, 5, where we provide also a computationally efficient scheme for upper bounding of $s_*(A)$. In Section 6, we develop another efficiently computable lower bound for $s_*(A)$ by applying the SDP relaxation, similar to the approach developed in [7] for the "unsigned" case $P_\pm = \emptyset$. In Section 7 we report on numerical experiments aimed at comparing the "power" of our LP-based sufficient conditions for $s$-semigoodness, their "unsigned" prototypes from [14], and conditions based on mutual incoherence. We show that incorporating the sign information can improve the bounds on the level of $s$-semigoodness, and that the bounds based on LP relaxations clearly outperform the bounds based on mutual incoherence.

3. It turns out that our verifiable sufficient conditions for $s$-semigoodness can be expressed in terms of specific properties of the *linear* recovery $\widehat{w}^{\text{lin}} = Y^T y$ associated with an appropriate $m \times n$ matrix $Y$. In Section 8, we propose and justify a new non-Euclidean *Matching Pursuit* algorithm associated with this linear recovery.

## 2 Necessary and sufficient conditions for $s$-semigoodness

Let $A$ be an $m \times n$ matrix, let $s$, $1 \leq s \leq m$, be an integer, and let $P_+$, $P_-$ and $P_n$ be a partition of $\{1, \ldots, n\}$ into three non-overlapping subsets. We say that $A$ is *$s$-semigood*, if for every vector $w$

with at most $s$ nonzero entries satisfying $w_i \geq 0$ for $i \in P_+$, and $w_i \leq 0$ for $i \in P_-$, $w$ is the unique optimal solution to the problem

$$\mathrm{Opt} = \min_z \{\|z\|_1 : Az = Aw,\ z_i \geq 0\ \forall i \in P_+,\ z_i \leq 0\ \forall i \in P_-\}\,. \tag{7}$$

Our primary goals are to find necessary and sufficient and *verifiable* sufficient conditions for $A$ to be $s$-semigood.

Note that without loss of generality we may assume $P_- = \emptyset$. Indeed, by replacing the partition $P_+, P_-, P_n$ with the partition $\overline{P}_+ = P_+ \cup P_-, \overline{P}_- = \emptyset, \overline{P}_n = P_n$ and matrix $A$ – with the matrix $\overline{A}$ obtained from $A$ by multiplying the columns with indices $i \in P_-$ by $-1$, $s$-semigoodness of $A$ with respect to the original sign restrictions given by $P_\pm, P_n$ is equivalent to the $s$-semigoodness of the new matrix $\overline{A}$ with respect to the new sign restrictions. By this reason, *we assume from now on that* $P_- = \emptyset$. Besides this, we assume without loss of generality that $P_+ = \{1, ..., p\}$ and $P_n = \{p+1, ..., n\}$ for some $p$. From now on, we denote by $\mathcal{P}_n$ the set of all signals satisfying the sign restrictions:

$$\mathcal{P}_n = \{w \in \mathbf{R}^n : w_i \geq 0\ \ \forall i \in P_+\}.$$

Note that since $P_- = \emptyset$, (7) simplifies to

$$\mathrm{Opt} = \min_z \{\|z\|_1 : Az = Aw,\ z_i \geq 0\ \forall i \in P_+\}\,. \tag{8}$$

Let us fix a norm $\|\cdot\|$ on $\mathbf{R}^n$, and let $\|\cdot\|_*$ be the conjugate norm.

**Proposition 2.1** *Let $m, n, s$ and $P_+$ be given. The following six conditions on an $m \times n$ matrix $A$ are equivalent to each other:*

(i) *$A$ is $s$-semigood;*

(ii) *For every subset $J$ of $\{1, ..., n\}$ with $\mathrm{Card}(J) \leq s$, and any $x \in \mathrm{Ker}A \backslash \{0\}$ such that $x_i \leq 0$ for all $i \in P_+ \setminus J$ one has*

$$\sum_{i \in J \cap P_+} x_i + \sum_{i \in J \cap P_n} |x_i| < \sum_{i \notin J} |x_i|.$$

(iii) *There exists $\xi \in (0, 1)$ such that for every subset $J$ of $\{1, ..., n\}$ with $\mathrm{Card}(J) \leq s$ and any $x \in \mathrm{Ker}A$ such that $x_i \leq 0$ for all $i \in P_+ \setminus J$ one has*

$$\sum_{i \in J \cap P_+} x_i + \sum_{i \in J \cap P_n} |x_i| \leq \xi \sum_{i \notin J} |x_i|.$$

(iv) *There exist $\xi \in (0, 1)$ and $\theta \in [1, \infty)$ such that $A$ satisfies the condition $\mathbf{SG}_s(\xi, \theta)$ as follows: for every $x \in \mathrm{Ker}A$ and every subset $J$ of $\{1, ..., n\}$ with $\mathrm{Card}(J) \leq s$, one has*

$$\sum_{i \in J \cap P_+} x_i + \sum_{i \in J \cap P_n} |x_i| \leq \xi \left( \sum_{i \in P_n \backslash J} |x_i| + \sum_{i \in P_+ \backslash J} \psi(x_i) \right),\ \ \psi(t) = \max[-t, \theta t],$$

*or, equivalently: for all $x \in \mathrm{Ker}A$, $\Theta(x) \leq \xi \Psi(x)$ where*

$$\begin{array}{l} \Theta(x) := \max\limits_{\substack{J \subset \{1,...,n\}, \\ \mathrm{Card}(J) \leq s}} \left[ \sum_{i \in J \cap P_+} \max[(1-\xi)x_i, (1+\theta\xi)x_i] + \sum_{i \in J \cap P_n} (1+\xi)|x_i| \right] \\ \Psi(x) := \sum_{i \in P_+} \max[-x_i, \theta x_i] + \sum_{i \in P_n} |x_i| \end{array} \tag{9}$$

(v) *There exist* $\xi \in (0,1)$, $\theta \in [1,\infty)$ *and* $\beta \in [0,\infty)$ *such that* $A$ *satisfies the condition* $\mathbf{SG}_{s,\beta}(\xi,\theta)$ *as follows:*
*for every* $x \in \mathbf{R}^n$ *and every subset* $J$ *of* $\{1,...,n\}$ *with* $\mathrm{Card}(J) \leq s$, *one has*

$$\sum_{i\in J\cap P_+} x_i + \sum_{i\in J\cap P_n} |x_i| \leq \beta\|Ax\| + \xi\left(\sum_{i\in P_n\setminus J} |x_i| + \sum_{i\in P_+\setminus J} \psi(x_i)\right), \quad \psi(t) = \max[-t, \theta t].$$

(vi) *There exist* $\xi \in (0,1)$ *and* $\beta \in [0,\infty)$ *such that* $A$ *satisfies the condition* $\mathbf{SG}_{s,\beta}(\xi)$ *as follows: for every* $J \subset \{1,...,n\}$ *with* $\mathrm{Card}(J) \leq s$ *and any* $x \in \mathbf{R}^n$ *such that* $x_i \leq 0$ *for all* $i \in P_+ \setminus J$, *one has*

$$\sum_{i\in J\cap P_+} x_i + \sum_{i\in J\cap P_n} |x_i| \leq \beta\|Ax\| + \xi\sum_{i\notin J} |x_i|.$$

We provide the proof of Proposition 2.1 in Appendix A.

As we have already mentioned in Introduction, when $P_n = \emptyset$ or $P_+ = \emptyset$, the characterizations (i)–(iv) of $s$-semigoodness are not completely new. For instance, when $P_n = \emptyset$, a necessary and sufficient condition for $s$-semigoodness of $A$ in the form (ii) has been established in [17] (compare (ii) to the definition (5) of half $s$-balancedness of $\mathrm{Ker}A$). On the other hand, the equivalent formulation of this characterization in terms of conditions $\mathbf{SG}_{s,\beta}(\xi,\theta)$ and $\mathbf{SG}_{s,\beta}(\xi)$ seems to be new. We are about to demonstrate that the latter two conditions allow to control the error of $\ell_1$-recovery in the case when the vector $w \in \mathbf{R}^n$ is not $s$-sparse and the problem (8) is not solved to exact optimality.

# 3 Error bounds for imperfect $\ell_1$-recovery

We have seen that the conditions provided in Proposition 2.1 are responsible for $s$-semigoodness of a sensing matrix $A$, that is, for the exactness of $\ell_1$-recovery in the "ideal case" when the true signal $w$ is $s$-sparse, there is no observation error, and the optimization problem (8) is solved to exact optimality. Below we demonstrate that these conditions control also the error of $\ell_1$-recovery in the case when the signal $w \in \mathcal{P}_n$ is not exactly $s$-sparse, there is observation noise and problem (8) is not solved to exact optimality. The corresponding error bound (cf [14, Proposition 3.1, Theorem 3.1]) is as follows:

**Proposition 3.1** *Let* $w \in \mathcal{P}_n$ *be such that* $\|w-w^s\|_1 \leq \mu$, *where* $w^s$ *is the vector obtained from* $w$ *by replacing all but the* $s$ *largest in magnitude entries in* $w$ *with zeros, let* $y$ *be such that* $\|Aw-y\| \leq \varepsilon$, *and let, finally,* $x$ *be an approximate solution to the optimization problem*

$$\mathrm{Opt} = \min_z \left\{\|z\|_1 : \|Az-y\| \leq \varepsilon,\ z_i \geq 0\ \forall i \in P_+\right\}. \tag{10}$$

*such that* $\|x\|_1 \leq \mathrm{Opt} + \nu$ *and* $\|Ax-y\| \leq \delta$.

1. *If* $A$ *satisfies the condition* $\mathbf{SG}_{s,\beta}(\xi,\theta)$ *with some* $\xi \in (0,1)$, $\beta \in [0,\infty)$ *and* $\theta \in [1,\infty)$, *then*

$$\|x-w\|_1 \leq \frac{1+\xi}{1-\xi}\nu + \frac{2(1+\xi\theta)}{1-\xi}\mu + \frac{2\beta}{1-\xi}(\varepsilon+\delta). \tag{11}$$

2. *If $A$ satisfies the condition $\mathbf{SG}_{s,\beta}(\xi)$ with some $\xi \in (0,1)$, $\beta \in [0,\infty)$, then*

$$\|x-w\|_1 \leq \frac{1+\xi}{1-\xi}\nu + \frac{2(1+\beta\alpha)}{1-\xi}\mu + \frac{2\beta}{1-\xi}(\varepsilon+\delta). \tag{12}$$

*where $\alpha$ stands for the maximum of $\|\cdot\|$-norms of the columns in $A$.*

For proof, see Appendix B.

# 4 Verifiable conditions for $s$-semigoodness

We are about to demonstrate that condition $\mathbf{SG}_{s,\beta}(\xi,\theta)$ from Proposition 2.1 leads to efficiently computable lower and upper bounds on the level of $s$-semigoodness.

## 4.1 Verifiable sufficient conditions for $s$-semigoodness by Linear Programming

Let

$$\mathcal{U}_s = \{u \in \mathbf{R}^n : \ \|u\|_1 \leq s, \ \|u\|_\infty \leq 1\},$$

so that $\mathcal{U}_s$ is the convex hull of all $\{-1,0,1\}$ vectors with at most $s$ nonzero entries, and for $x \in \mathbf{R}^n$, let $\|x\|_{s,1}$ be the sum of the $s$ largest magnitudes of entries in $x$, or, equivalently,

$$\|x\|_{s,1} = \max_{u\in\mathcal{U}_s} u^T x.$$

Let

$$(D_\theta[x])_i = \left\{\begin{array}{ll} [1+\theta\xi]\max[x_i,0], & i \in P_+ \\ (1+\xi)|x_i|, & i \notin P_+ \end{array}\right., \qquad \Phi(x) = \|D_\theta[x]\|_{s,1}.$$

Suppose $\xi \in [0,1)$, $\theta \in [1,\infty)$ and $\rho, \sigma \in [0,\infty)$ are given. Consider the following condition on an $m \times n$ matrix $A$:

$\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$: *There exist $m \times n$ matrix $Y = [y_1,...,y_n]$ and a vector $v \in \mathbf{R}^m$ such that*

$$\begin{array}{rcll} \Phi_s(-C_i[Y,A]) + (A^T v)_i & \leq & \xi, \ 1 \leq i \leq n & (a) \\ \Phi_s(C_i[Y,A]) - (A^T v)_i & \leq & \xi, \ i \notin P_+ & (b) \\ \Phi_s(C_i[Y,A]) - (A^T v)_i & \leq & \theta\xi, \ i \in P_+ & (c) \\ \|y_i\|_* & \leq & \sigma, \ 1 \leq i \leq n & (d) \\ \|v\|_* & \leq & \rho & (e) \end{array} \tag{13}$$

*where $C_i[Y,A]$ is the $i$-th column of the matrix $I - Y^T A$.*

Observe that this condition is verifiable, since (13) is a system of explicit convex constraints on $Y$ and $v$.

**Proposition 4.1** *Let $A$ satisfy $\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$ with some $\xi \in [0,1)$, $\theta \in [1,\infty)$, and $\rho,\sigma \in [0,\infty)$. Then $A$ satisfies $\mathbf{SG}_{s,\beta}(\xi,\theta)$ with*

$$\beta = \rho + \sigma \max_{k_+,k_n}\left\{ k_+(1+\theta\xi) + k_n(1+\xi) : \begin{array}{l} 0 \leq k_+ \leq \mathrm{Card}(P_+) \\ 0 \leq k_n \leq \mathrm{Card}(P_n) \\ k_+ + k_n \leq s \end{array}\right\} \leq \rho + \sigma s(1+\theta\xi). \tag{14}$$

*In particular, $A$ is $s$-semigood.*

For proof, see Appendix C.

Some comments are in order.

**Origin of the condition $\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$.** The condition $\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$ is yielded by a simple and general construction, and we believe it makes sense to present this construction in its general form. The essence of the matter is in building a verifiable sufficient condition for the validity of (9), see Proposition 2.1.iv. By positive homogeneity of degree 1 of the convex functions $\Theta,\Psi$ participating in (9), the latter condition is equivalent to

$$\mathrm{Opt} := \max_x \{\Theta(x) : Ax = 0, x \in X\} \le \xi, \quad X = \{x : \Psi(x) \le 1\}. \tag{15}$$

A verifiable sufficient condition for (15) is basically the same as an efficiently computable upper bound for Opt; the sufficient condition for the validity of (15) associated with such a bound merely states that the bound is $\le \xi$. Now observe that from the origin of $\Psi$ (see (9)) it is clear that $X$ has a moderate number, $N$, of readily available extreme points $x^1,...,x^N$ (in the case of (9), $N = 2n$), so that the only difficulty in computing Opt exactly comes from linear constraints $Ax = 0$. The standard way to circumvent this difficulty and to efficiently bound Opt from above is to use the Lagrange relaxation: for any $v \in \mathbf{R}^m$,

$$\begin{array}{rcl}\mathrm{Opt} &=& \max\limits_{x\in X}\left\{\Theta(x) + v^TAx : Ax = 0, x \in X\right\}\\ &\le& \max\limits_{x}\left\{\Theta(x) + v^TAx : x \in X\right\} = \max\limits_{1\le i\le N}[\Theta(x^i) + v^TAx^i],\end{array}$$

and hence the efficiently computable *Lagrange relaxation bound* $\inf_v \max_{1\le i\le N}[\Theta(x^i) + v^TAx^i]$ is an upper bound on Opt. Unfortunately, in our situation this bound can be very poor; e.g., when $X$ is symmetric with respect to the origin and $\Theta$ is even (as it happens in (9) when $P_+ = \emptyset$), it is immediately seen that the bound becomes the trivial bound $\mathrm{Opt} \le \max_{x\in X}\Theta(x) = \max_i \Theta(x^i)$. In order to strengthen the relaxation, we pass to the Fenchel-type representation of $\Theta$

$$\Theta(x) = \sup_u \left[[Pu + q]^T x - \Theta_*(u)\right]$$

with a proper convex function $\Theta_*$; such a representation, even with $Pu + p \equiv u$, exists whenever $\Theta$ is a proper convex function (and can be easily found for $\Theta$ we are interested in). We now have for any $Y \in \mathbf{R}^{m\times n}$, $v \in \mathbf{R}^m$,

$$\begin{array}{rcl}\mathrm{Opt} &=& \max\limits_x \{\Theta(x) : Ax = 0, x \in X\} = \sup\limits_{x,u}\left\{[Pu + p]^T x - \Theta_*(u) : Ax = 0, x \in X\right\}\\ &=& \sup\limits_{x,u}\left\{[Pu + p]^T[x - Y^TAx] + v^TAx - \Theta_*(u) : Ax = 0, x \in X\right\}\\ &\le& \sup\limits_{x,u}\left\{[Pu + p]^T[x - Y^TAx] + v^TAx - \Theta_*(u) : x \in X\right\}\\ &=& \max\limits_{1\le i\le N}\underbrace{\sup\limits_u\left\{[Pu + p]^T[x^i - Y^TAx^i] + v^TAx^i - \Theta_*(u)\right\}}_{:=\Theta_i(Y,v)},\end{array}$$

so that the condition

$$\exists (Y \in \mathbf{R}^{m\times n}, v \in \mathbf{R}^m) : \Theta_i(Y, v) \le \xi, \ 1 \le i \le N, \tag{16}$$

is sufficient for the validity of (15). Note that the functions $\Theta_i$, by their origin, are convex, so that the condition (16) is efficiently verifiable, provided that $\Theta_i(\cdot)$ are efficiently computable.

In the case we are interested in, the extreme points of $X$ are the $2n$ vectors $-e_i$ for $1 \leq i \leq n$, $e_i$ for $i \in P_n$, and $\theta^{-1}e_i$ for $i \in P_+$, where $e_i$ is the $i$-th basic orth. Implementing the outlined bounding scheme and adding additional restrictions (13.$d$,$e$) to get a control over $\beta$, we arrive at (13). It should be stressed that the outlined scheme can be applied to bounding from above the optimal value of a whatever problem of the form (15) with a convex polytope $X$ and a proper convex objective $\Theta$; all what matters is that $X$ is given as $\mathrm{Conv}\{x^1, ..., x^N\}$ and $\Theta$ is efficiently computable. Note also that when $X$ is a polytope given by list of $M$ linear inequalities, we can efficiently represent it as the intersection of $M$-dimensional standard simplex and an affine plane, so that the outlined scheme is applicable to a whatever problem of maximizing an efficiently computable proper convex function under a (finite) system of linear inequality and equality constraints.

**Effect of increasing $\beta, \theta, \xi$.** The condition $\mathbf{SG}_{s,\beta}(\xi, \theta)$ appearing in Proposition 2.1.v clearly is "monotone" in the parameters $\beta, \theta, \xi$: whenever $A$ satisfies this condition and $\beta' \geq \beta$, $\theta' \geq \theta$ and $\xi' \geq \xi$, $A$ satisfies the condition $\mathbf{SG}_{s,\beta'}(\xi', \theta')$ as well. Proposition 4.1 offers a verifiable sufficient condition for the validity of $\mathbf{SG}_{s,\beta}(\xi, \theta)$, specifically,

$\mathbf{VSG}^*_{s,\beta}(\xi, \theta)$: $\exists Y, v\ \rho,\ \sigma$ satisfying (13) and the relation $\rho + \sigma s(1 + \theta\xi) \leq \beta$.

A natural question is, whether this verifiable condition possesses the same monotonicity properties as the "target" condition $\mathbf{SG}_{s,\beta}(\xi, \theta)$. In the case of the affirmative answer, in order to conclude that $A$ is $s$-semigood, we could check the validity of $\mathbf{VSG}^*_{s,\beta}(\xi, \theta)$ for appropriately large values of $\beta, \theta$ and a close to one value of $\xi < 1$; if the condition is satisfied, $A$ is $s$-semigood, and error bounds from Proposition 3.1 take place. Were the condition $\mathbf{VSG}^*_{s,\beta}(\xi, \theta)$ "not monotone," to justify the $s$-semigoodness of $A$ via this condition would require a problematic and time-consuming search in the space of parameters $\beta, \theta, \xi$. Fortunately, the condition $\mathbf{VSG}^*_{s,\beta}(\xi, \theta)$ indeed is monotone:

**Proposition 4.2** *Let $A$ satisfy $\mathbf{VSG}^*_{s,\beta}(\xi, \theta)$, and let $Y, v, \sigma, \rho$ be the corresponding certificate, that is, $\rho + \sigma s(1 + \theta\xi) \leq \beta$ and $Y, v, \sigma, \rho$ satisfy (13). Then $A$ satisfies $\mathbf{VSG}^*_{s,\beta'}(\xi', \theta')$ whenever $\beta' \geq \beta$, $\theta' \geq \theta$ and $\xi' \in (\xi, 1)$, the certificate being $(Y', v, \sigma, \rho)$, where the columns $Y'_i$ of $Y'$ are multiples of the columns $Y_i$ of $Y$, namely,*

$$Y'_i = a_i Y_i;\ \ [0,1] \ni a_i = \begin{cases} (1+\xi\theta)/(1+\xi'\theta'), & i \in P_+ \\ (1+\xi)/(1+\xi'), & i \in P_n \end{cases}$$

For proof, see Online Supplement F.1.

**Relation to the sufficient condition for $s$-goodness from [14] and the Restricted Isometry Property.** The verifiable sufficient condition for $s$-goodness from [14] requires from an $m \times n$ matrix $A$ the existence of $\gamma < 1/2$ and $Y = [y_1, ..., y_n] \in \mathbf{R}^{m \times n}$ such that

$$\|C_i[Y, A]\|_{s,1} \leq \gamma, \ \text{ for all } 1 \leq i \leq n,$$

Setting $\theta = 1$ and $\xi = \frac{\gamma}{1-\gamma}$ (so that $\xi < 1$ and $\gamma = \frac{\xi}{1+\xi}$) and taking into account that in the case of $\theta = 1$ we have $\Phi_s(z) \leq (1 + \xi)\|z\|_{s,1}$, the latter condition implies that

$$\Phi_s(\pm C_i[Y, A]) \leq (1 + \xi)\gamma = \xi, \quad \forall i,$$

that is, it implies the validity of $\mathbf{VSG}_s(\xi, 1, 0, \sigma)$, provided that $\sigma$ is large enough, specifically, $\sigma \geq \|y_i\|_*$ for all $i$.

As it was shown in the companion paper [14], when $A$ satisfies the *Restricted Isometry Property* RIP$(\delta, k)$ with parameters $\delta \in (0,1)$, $k > 1$, the above sufficient condition for $s$-goodness is satisfied with $\gamma = 1/3$ for $s$ as large as $O(1)(1-\delta)\sqrt{k}$; as a result, a RIP$(\delta, k)$-matrix satisfies $\mathbf{VSG}_s(\frac{1}{2}, 1, 0, \sigma)$ provided that $\sigma$ is large enough and $s \leq O(1)(1-\delta)\sqrt{k}$. Since for large $m, n$, $m < n$, typical random matrices possess, with overwhelming probability, property RIP$(\frac{1}{2}, k)$ with $k$ as large as $O(1)m/\ln(n/m)$, we see that our verifiable sufficient condition for $s$-semigoodness can certify the latter property for $s$ as large as $O(1)\sqrt{m/\ln(n/m)}$, provided that the matrix in question is "good enough".

### 4.2 Upper bounding the level of $s$-semigoodness

Here we address the issue of bounding from above the maximal $s = s_*(A)$ for which $A$ is $s$-semigood. The construction to follow is motivated by item (iv) of Proposition 2.1. A necessary and sufficient condition for the $s$-semigoodness of $A$ is the existence of $\xi < 1$ and $\theta \geq 1$ such that for all $x \in \mathrm{Ker}A$ and any set $I$ of indices with $\mathrm{Card}(I) \leq s$

$$\sum_{i \in I \cap P_+} \max[(1-\xi)x_i, (1+\theta\xi)x_i] + \sum_{i \in I \cap P_n} (1+\xi)|x_i| \leq \xi \Psi(x)$$

where

$$\Psi(x) = \sum_{i \in P_+} \max[-x_i, \theta x_i] + \sum_{i \in P_n} |x_i|, \tag{17}$$

or, equivalently,

> (!) for every $x \in \mathrm{Ker}A$ and every vector $v$ with at most $s$ nonzero entries and nonzero entries $v_i$ belonging to $[1-\xi, 1+\xi\theta]$ if $i \in P_+$ and belonging to $[-1-\xi, 1+\xi]$ if $i \in P_n$, one has
> $$v^T x \leq \xi \Psi(x).$$

Observe that the convex hull of the vectors $v$ in question is exactly the set

$$\mathcal{U}^{\xi,\theta} = \left\{ v \in \mathbf{R}^n : \begin{array}{c} 0 \leq v_i \leq 1+\theta\xi,\ i \in P_+,\ |v_i| \leq 1+\xi,\ i \in P_n, \\ \sum_{i \in P_+} \frac{v_i}{1+\theta\xi} + \sum_{i \in P_n} \frac{|v_i|}{1+\xi} \leq s \end{array} \right\}.$$

Recalling that $P_+ = \{1, ..., p\}$, setting $q = n - p = \mathrm{Card}(P_n)$ and

$$\mathcal{U} = \{u \in \mathbf{R}^n : \ \|u\|_1 \leq s,\ \|u\|_\infty \leq 1,\ u_i \geq 0 \text{ for } i \in P_+\} \tag{18}$$

we see that

$$\mathcal{U}^{\xi,\theta} = C^{\xi,\theta}\mathcal{U}, \quad \text{where} \quad C^{\xi,\theta} = \left[\begin{array}{c|c} (1+\xi\theta)I_p & 0 \\ \hline 0 & (1+\xi)I_q \end{array}\right]. \tag{19}$$

The condition (!) now reads

$$\max_{v \in \mathcal{U}^{\xi,\theta}} v^T x \leq \xi \Psi(x) \ \text{ for all } x \in \mathrm{Ker}A.$$

Setting $\mathcal{X} = \{x \in \mathrm{Ker}A : \ \Psi(x) \leq 1\}$ the latter condition, by homogeneity reason, is the same as

$$\mathrm{Opt} = \mathrm{Opt}(\xi, \theta) := \max_{v,x} \left\{ v^T x : \ v \in \mathcal{U}^{\xi,\theta}, \ x \in \mathcal{X} \right\} \leq \xi; \tag{20}$$

recall that $A$ is $s$-semigood if and only if there exist $\theta \geq 1$ and $\xi < 1$ such that (20) takes place.

We can use (20) in order to bound $s_*(A)$ from above, as follows. In order to certify that $s_*(A) < s$ for a given $s$ ($s$ is the input to our algorithm), we fix a large $\theta$ and a close to one $\xi < 1$ (these are the parameters of the algorithm) and run the iterations

$$u_0 \in \mathcal{U}^{\xi,\theta} \mapsto x_1 \in \mathrm{Argmax}_{x \in \mathcal{X}} u_0^T x \mapsto u_1 \in \mathrm{Argmax}_{u \in \mathcal{U}^{\xi,\theta}} u^T x_1 \mapsto ...$$

initiating them by a picked at random vertex $u_0$ of $\mathcal{U}^{\xi,\theta}$. Note that the quantities $u_i^T x_i$, $i = 1, 2, ...$ clearly form a nondecreasing sequence of lower bounds on Opt. We terminate the outlined iterations when the progress in the bounds – the difference $u_i^T x_i - u_{i-1}^T x_{i-1}$ – falls below a given small threshold, and we run this process a predetermined number of times from different randomly chosen starting points. As a result, we get a set of lower bounds on Opt of the form $u^T x$, where $u$ is a vertex of $\mathcal{U}^{\xi,\theta}$ and $x \in \mathcal{X}$. If our goal were merely to certify that (23) is not valid for given $s, \theta, \xi$, we could terminate this process at the first step, if any, when the current lower bound $u^T x$ becomes $> \xi$ (cf. [14, Section 4.1]). We, however, want to certify that $s > s_*(A)$, or, which is the same by Proposition 2.1.iv, that (23) fails to be true for *all* $\theta$ and all $\xi < 1$, and not only for those $\theta, \xi$ we have selected for our test. To overcome this difficulty, we accompany every step $u \mapsto x \in \mathrm{Argmax}_{x \in \mathcal{X}} u^T x$ by an additional computation as follows. In our process, $u$ is an extreme point of $\mathcal{U}^{\xi,\theta}$, that is, a point with $s_u \leq s$ nonzero entries, let the set of indices of these entries be $I$. Setting $\epsilon_i = \mathrm{sign}(u_i)$, we solve the following LP problem

$$\max_x \left\{ \sum_{i \in I \cap P_+} x_i + \sum_{i \in I \cap P_n} \epsilon_i x_i : \left\{ \begin{array}{l} x_i \leq 0, \ i \in P_+ \backslash I \\ Ax = 0 \\ \sum_{i \notin I} |x_i| \leq 1 \end{array} \right\} \right\}.$$

If the optimal value in this problem is $\geq 1$, we terminate our test and claim that $A$ is not $s$-good; by Proposition 2.1.ii, this indeed is the case.

As applied to a given input $s$, the outlined test either terminates with a valid claim "$s > s_*(A)$", or terminates with no conclusion at all, in which case we could pass to testing a larger value of $s$.

# 5 Limits of performance of LP-based sufficient conditions for $s$-semigoodness

Unfortunately, the condition in question, same as its predecessor from [14], can*not* certify $s$-semigoodness of an $m \times n$ matrix in the case of $s > O(1)\sqrt{m}$, unless the matrix is "nearly square". The precise statement is as follows (cf. [14, Proposition 4.2]):

**Proposition 5.1** *Let*

$$n > 2(2\sqrt{2m} + 1)^2 \tag{21}$$

*and let $\xi < 1, \theta \geq 1, \sigma \geq 0, \rho \geq 0$, an integer $s$ and an $m \times n$ matrix $A$ be such that $A$ satisfies* $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$. *Then*

$$s \leq 2\sqrt{2m} + 1. \tag{22}$$

For proof, see Appendix D.

The results from Proposition 5.1 show that our verifiable sufficient conditions can only certify $s$-semigoodness of an $m \times n$ matrix at a suboptimal rate of $s \leq O(1)\sqrt{m}$, unless the matrix is "nearly square". In fact this verifiable bound can still give a very poor impression on the true largest $s = s_*(A)$ for which $A$ is $s$-semigood. An instructive example in this direction is as follows. Consider the case of $P_+ = \{1, ..., n\}$, let $m = 2d + 1$ be odd, and let the rows of $A$ be comprised of the values of basic trigonometric polynomials

$$p_0(\phi) \equiv 1, \;\; p_{2i-1}(\phi) = \cos(i\phi), \;\; p_{2i}(\phi) = \sin(i\phi), \;\; 1 \leq i \leq d,$$

taken along the regular grid $\phi_j = 2\pi j/n$, $0 \leq j < n$, so that $A_{ij} = p_i(\phi_j)$, $0 \leq i < m$, $0 \leq j < n$ (we enumerate rows and columns starting with 0 rather than with 1). It is well known [5, 10] that in this case $A$ is $s$-semigood for $s = d$. In contrast to this, when $A$ is not "nearly square", specifically, when $n > 4\pi d$, $A$ can satisfy the condition $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$ only for $s \leq 2$, no matter how large $\theta, \sigma, \rho$ are and how close to 1 $\xi < 1$ is, see Online Supplement F.2.

# 6 Verifiable sufficient conditions for $s$-semigoodness by Semidefinite Relaxation

Following d'Aspremont and El Ghaoui [7], we are about to derive another verifiable sufficient condition for $s$-semigoodness, now - via semidefinite relaxation. The construction to follow is motivated by the development in the beginning of Section 4.2, according to which $s$-semigoodness of $A$ is implied by the validity of (20) for $\theta > 1$ and $\xi < 1$.

Let, as before,

$$\mathcal{X} = \{x \in \mathrm{Ker} A : \;\Psi(x) \leq 1\} \;\text{ and }\; \mathcal{U}^{\xi,\theta} = \{C^{\xi,\theta} u : \; u \in \mathcal{U}\},$$

where $\Psi$, $\mathcal{U}$ and $C^{\xi,\theta}$ are defined in, respectively, (17), (18) and (19). The condition (20) is equivalent to

$$\max_{u,x} \left\{ (C^{\xi,\theta} u)^T x : \; u \in \mathcal{U}, \; x \in \mathcal{X} \right\} \leq \xi. \tag{23}$$

Observe that for $x\in\mathcal{X}$, $u\in\mathcal{U}$ the matrices $Z=xx^T$, $V=xu^T$ and $Q=uu^T$ satisfy the relations

$$\begin{array}{ll}
\multicolumn{2}{l}{\exists t\in\mathbf{R}^n, R\in\mathbf{S}^{2n}, \Lambda\in\mathbf{S}^{2n}:}\\
(a) & G=\left[\begin{array}{c|c|c}1 & x^T & u^T\\ \hline x & Z & V\\ \hline u & V^T & Q\end{array}\right]\succeq 0;\\
(b) & \left\{\begin{array}{l} Q=\underbrace{\left[\begin{array}{c|c} I_n & -I_n\end{array}\right]}_{:=L}\underbrace{\left[\begin{array}{c|c}R^{11} & R^{12}\\ \hline R^{12} & R^{11}\end{array}\right]}_{:=R}L^T,\\ 0\leq R_{ij}\leq \frac{1}{2},\quad R\succeq 0,\quad R^{12}=[R^{12}]^T,\ \mathrm{Tr}(R)\leq s,\\ \sum_{i,j}R_{ij}\leq s^2,\quad R^{12}_{ij}=0\ \forall i,j\in P_+;\end{array}\right.\\
(c) & Z=\underbrace{\left[\begin{array}{c|c|c|c}-I_p & 0 & \frac{1}{\theta}I_p & 0\\ \hline 0 & -I_q & 0 & I_q\end{array}\right]}_{:=F}\Lambda F^T,\quad 0\leq\Lambda_{ij},\quad \Lambda\succeq 0,\quad \sum_{i,j}\Lambda_{ij}\leq 1;\\
(d_1) & \left\{\begin{array}{ll}-t_i\leq V_{ij}\leq\frac{t_i}{\theta}, & \forall i,j\in P_+\\ |V_{ij}|\leq t_i, & \text{otherwise};\end{array}\right.\\
(d_2) & \sum_{j\in P_+}\max[-V_{ij},\theta V_{ij}]+\sum_{j\in P_n}|V_{ij}|\leq st_i, \forall i\in P_+;\\
(d_3) & \sum_j |V_{ij}|\leq st_i, \forall i\in P_n;\\
(d_4) & \sum_i t_i\leq 1;\\
(e) & AZA^T=0\\
(f) & x\in\mathcal{X},\quad u\in\mathcal{U}.
\end{array}\tag{24}$$

Besides this,

$$u^T(C^{\xi,\theta})^Tx=\mathrm{Tr}(C^{\xi,\theta}V).$$

Indeed, the latter relation, same as (24.$a$), (24.$e$) and (24.$f$), is evident. To verify (24.$b$), let $u_+=\max[u,0]$, $u_-=\max[-u,0]$, where max is acting coordinate-wise. Then

$$\begin{array}{rcl}
Q &=& L\left[\begin{array}{c|c}u_+u_+^T & u_+u_-^T\\ \hline u_-u_+^T & u_-u_-^T\end{array}\right]L^T = L\left[\begin{array}{c|c}u_-u_-^T & u_-u_+^T\\ \hline u_+u_-^T & u_+u_+^T\end{array}\right]L^T\\
&=& L\underbrace{\left[\begin{array}{c|c}\frac{1}{2}[u_+u_+^T+u_-u_-^T] & \frac{1}{2}[u_+u_-^T+u_-u_+^T]\\ \hline \frac{1}{2}[u_-u_+^T+u_+u_-^T] & \frac{1}{2}[u_-u_-^T+u_+u_+^T]\end{array}\right]}_{R}L^T,
\end{array}$$

and the matrix $R$ we have just defined clearly satisfies all requirements from (24.$b$). To verify (24.$c$), observe that the extreme points of the set $\mathcal{X}^+=\{x:\Psi(x)\leq 1\}\supset\mathcal{X}$ are the vectors $\pm e_i$, $i>p$, and $-e_i,\theta^{-1}e_i$, $i\leq p$, so that $x=F\lambda$ with $\lambda\in\mathbf{R}^{2n}_+$, $\sum_i\lambda_i\leq 1$; setting $\Lambda=\lambda\lambda^T$, we satisfy (24.$c$). To satisfy (24.$d$), it suffices to set $t_i=|x_i|$ for all $i>p$ and $t_i=\max[-x_i,\theta x_i]$ for $i\leq p$ and to take into account that $\max[-V_{ij},\theta V_{ij}]\geq|V_{ij}|$ for all $i,j$ due to $\theta\geq 1$, and that $u_i\geq 0$ for $i\in P_+$.

It follows that a sufficient condition for (23) is

$$\mathrm{Opt}^{\xi,\theta}:=\max_{\substack{Z,Q\in\mathbf{S}^n,\ R,\Lambda\in\mathbf{S}^{2n},\\ V\in\mathbf{R}^{n\times n},\ t\in\mathbf{R}^n}}\left\{\mathrm{Tr}(C^{\xi,\theta}V):\ (24)\text{ is satisfied}\right\}\leq\xi.\tag{25}$$

The optimization problem in (25) clearly reduces to a semidefinite maximization program $\mathcal{S}$; by weak duality, the optimal value in the semidefinite dual $\mathcal{D}$ to $\mathcal{S}$ is $\geq \mathrm{Opt}^{\xi,\theta}$. It follows that the efficiently verifiable condition

$$\mathrm{Opt}(\mathcal{D}) \leq \xi$$

is a sufficient condition for $s$-semigoodness of $A$. Note that the above construction depends on $\theta \geq 1$ and $\xi < 1$ as parameters.

**Remark.** Consider the case of $P_+ = \emptyset$, where $\mathcal{X} = \{x \in \mathbf{R}^n : \|x\|_1 \leq 1, Ax = 0\} \supset \mathcal{Z} = \{x \in \mathbf{R}^n : \|x\|_1 \leq 1\}$. In this case, the standard semidefinite relaxation of the set $\mathcal{C}_* = \mathrm{Conv}\{xx^T : x \in \mathcal{Z}\}$ is

$$\mathcal{C} = \left\{ Z : \ Z \succeq 0, \sum_{i,j} |Z_{ij}| \leq 1 \right\}$$

(cf. [7]). Note that (24.$c$) uses another semidefinite relaxation of $\mathcal{C}_*$, namely,

$$\mathcal{C}' = \left\{ Z : \exists \Lambda \in \mathbf{S}^{2n} : \begin{array}{l} \Lambda \succeq 0, \Lambda_{i,j} \geq 0 \ \forall i,j, \ \sum_{i,j} \Lambda_{ij} \leq 1 \\ Z = [I_n, -I_n]\Lambda[I_n, -I_n]^T \end{array} \right\}.$$

It is immediately seen that $\mathcal{C}_* \subset \mathcal{C}' \subset \mathcal{C}$; a surprising fact is that the second of these inclusions is strict. Thus, the relaxation of $\mathcal{C}_*$ given by $\mathcal{C}'$ is less conservative than the standard relaxation given by $\mathcal{C}$. As observed by A. d'Aspremont (private communication), the relaxation $\mathcal{C}'$ can be further improved, namely, by replacing $\mathcal{C}'$ with

$$\mathcal{C}^+ = \left\{ Z : \exists \Lambda = \left[ \begin{array}{cc} \Lambda^{11} & \Lambda^{12} \\ \Lambda^{21} & \Lambda^{22} \end{array} \right] \in \mathbf{S}^{2n} : \begin{array}{l} \Lambda^{\mu\nu} \in \mathbf{R}^{n\times n}, \ \Lambda \succeq 0, \ \Lambda_{i,j} \geq 0 \ \forall i,j \\ \sum_{i,j} \Lambda_{ij} \leq 1, \ \Lambda^{12}_{ii} = 0, \ 1 \leq i \leq n \\ Z = [I_n, -I_n]\Lambda[I_n, -I_n]^T \end{array} \right\}.$$

Note that this idea can be used to improve the semidefinite relaxation given by $\mathcal{C}$ as well. Specifically, the matrix $R$ as built in the justification of (24) clearly satisfies $(R^{12})_{ii} = 0$, $1 \leq i \leq n$, and we can add these linear constraints on $R$ to (24.$b$). Similarly, when representing a vector $x \in \mathcal{X}^+$ as $F\lambda$ with $\lambda \in \mathbf{R}^{2n}_+$, $\sum_i \lambda_i \leq 1$, see the justification of (24), we clearly can ensure that $\lambda_i \lambda_{n+i} = 0$, $1 \leq i \leq n$, that is, the matrix $\Lambda$ we have built in fact satisfies $\Lambda_{i,n+i} = \Lambda_{n+i,i} = 0$, $1 \leq i \leq n$, and we can add these linear constraints on $\Lambda$ to (24.$c$).

**Proposition 6.1** *If* $\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$ *with* $\rho = \sigma = \infty$ *holds, then* $\mathrm{Opt}^{\xi,\theta} \leq \xi$.

For proof, see Online Supplement F.3.

Although Proposition 6.1 states that the verifiable sufficient conditions based on semidefinite programming are at least as good as the ones based on linear programming, i.e. $\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$, in terms of their computational cost, conditions based on linear programming are far more advantageous.

## 7 Numerical results

In order to compare the performance of the proposed bounds on the maximal $s = s_*(A)$ for which a given matrix, $A$, is $s$-semigood, with the bounds known from the literature, we present some preliminary numerical results for relatively small sensing matrices. Our goal is to see if the sign information on a signal allows to improve the bounds for $s_*(A)$ as compared to the bounds on the largest $s = s_0(A)$ for which $A$ is $s$-good.

We generate four sets of random matrices, which are normalizations (all columns scaled to be of $\|\cdot\|_2$-norm 1) of (a) Rademacher matrices (i.i.d. entries taking values $\pm 1$ with probabilities 0.5), (b) Gaussian matrices (iid $\mathcal{N}(0,1)$ entries), (c) Fourier matrices — $m \times n$ submatrices of the matrix of $n \times n$ Discrete Fourier Transform, and (d) Hadamard matrices — $m \times n$ submatrices of the $n \times n$ Hadamard matrix[2]; in the cases (c,d), the $m$ rows comprising the submatrix were drawn at random from the $n$ rows of the "parent" matrix. For each type, we set the number of columns to $n = 256$ and vary the number of rows, $m = 0.5n, \dots, 0.95n$.

We bound from below the value $s_0(A)$ using the bound $s[\mu]$ by mutual incoherence and the bounds $s[\alpha_1]$ and $s[\alpha_s]$, computed through the LP-based verifiable sufficient conditions for $s$-goodness (see [14, Section 6]).

*The lower bound on* $s_*(A)$ is computed by invoking condition $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$, where $\rho = \sigma = \infty$ and $\theta$ is set to once for ever fixed "large enough" value, and $\xi$ is set to 0.9999, see section 4.1 and Propositions 4.1, 4.2. Note that given a matrix $Y$, and setting $v = 0$, one can compute the largest $s$ satisfying (13) and thus ensuring the validity of $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$. We first compute the best lower bound $\underline{s}$ on $s_*(A)$ given by the $Y$-matrices generated when bounding $s_0(A)$. Then we compute the "improved" lower bound for $s_*(A)$ as follows: we check whether the condition $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$ holds true for $s = \underline{s} + 1$, if it is the case, check whether this condition holds true for $s = \underline{s} + 2$, and so on.

While the outlined lower bounds on $s_*(A)$ and $s_0(A)$ are efficiently computable via LP (when $\sigma = \rho = \infty$, the sufficient condition is easily checked by solving a Linear Programming program), the sizes of the resulting LPs are rather large. For instance, when $A$ is $m \times n$, the LP associated with (13) has a $(2n^2+2n+1) \times ((m+2n)(n+1)+2)$ constraint matrix (compared to $(2n^2+n) \times (n(m+n+1)+1)$ constraint matrices arising when computing lower bounds for $s_0(A)$). For instance, for $m = 230$ and $n = 256$, bounding $s_*(A)$ results in an LP program of the size $131,585 \times 190,696$, while computing a lower bound on $s_0(A)$ requires solving an LP problem of size $131,328 \times 124,673$. In all the computations, we used the state-of-the-art commercial LP solver `mosekopt` [1].

*The upper bounds* on $s_*(A)$ and on $s_0(A)$ are computed by the techniques from Section 4.2 and [14, Section 4.1].

The results of our experiments and related CPU times are presented in Table 1. The computations were carried out on a single core of an 8-core Intel Xeon E5520@2.27GHz CPU Linux workstation.

The results in Table 1 merit some comments. We observe that our LP-based efficiently computable lower bounds on $s_0(A)$ and $s_*(A)$ clearly outperform the bounds based on mutual incoherence. We notice that for Fourier and Hadamard matrices, the lower bounds on $s_*(A)$ and $s_0(A)$ are nearly always the same, except for two Hadamard instances with $m = 230$ and $m = 242$. On the other hand, for Gaussian and Rademacher matrices, as the number of rows $m$ approaches the number of columns $n$, the difference between the best certified lower bounds on $s_*(A)$ and on $s_0(A)$ increases (for the sizes we have considered, this difference attains 5 for the Gaussian matrix with $m = 242$). While for Gaussian, Rademacher and Fourier matrices, the upper bounds on $s_*(A)$ become loose (they are twice or three times higher than the upper bounds on $s_0(A)$), these bounds

[2]The Hadamard matrix $H_d$, $d = 0, 1, 2, \dots$, has order $2^d \times 2^d$ and is given by the recurrence $H_0 = 1$, $H_{d+1} = [H_d, H_d; H_d, -H_d]$.

Table 1: Comparison of efficiently computable bounds on $s_*(A)$, $n = 256$

**Fourier matrices**

| | **Unsigned** | | | | **Nonnegative** | | **CPU time (s)** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | LBs on $s_0(A)$ | | | UB | LB | UB | Unsigned | | | Nonnegative | |
| $m$ | $s[\mu]$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ |
| 128 | 3 | 5 | 5 | 12 | 5 | 47 | 0.8 | 1054.0 | 146.0 | 3114.4 | 172.9 |
| 128 | 3 | 5 | 5 | 11 | 5 | 32 | 0.9 | 986.0 | 169.4 | 2891.5 | 311.5 |
| 152 | 2 | 6 | 6 | 11 | 6 | 49 | 1.1 | 898.5 | 252.5 | 3680.2 | 179.6 |
| 152 | 3 | 6 | 6 | 11 | 6 | 53 | 1.3 | 899.3 | 161.7 | 3836.7 | 183.5 |
| 178 | 2 | 6 | 6 | 12 | 6 | 47 | 1.1 | 866.5 | 228.6 | 3976.0 | 294.0 |
| 178 | 3 | 7 | 7 | 16 | 7 | 42 | 0.7 | 484.8 | 365.2 | 3216.8 | 416.9 |
| 204 | 4 | 8 | 8 | 17 | 8 | 67 | 1.0 | 828.5 | 235.4 | 3829.7 | 209.2 |
| 204 | 3 | 7 | 7 | 15 | 7 | 65 | 1.1 | 906.8 | 220.2 | 3914.4 | 197.4 |
| 230 | 4 | 10 | 10 | 21 | 10 | 70 | 1.1 | 1879.9 | 300.5 | 4287.6 | 384.6 |
| 230 | 4 | 9 | 9 | 20 | 9 | 65 | 1.0 | 856.6 | 286.5 | 4040.2 | 362.0 |
| 242 | 5 | 11 | 11 | 26 | 11 | 89 | 1.7 | 1425.1 | 290.5 | 6444.1 | 513.0 |
| 242 | 4 | 10 | 10 | 19 | 10 | 75 | 1.2 | 1920.6 | 265.3 | 4069.1 | 232.8 |

**Hadamard matrices**

| | **Unsigned** | | | | **Nonnegative** | | **CPU time (s)** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | LBs on $s_0(A)$ | | | UB | LB | UB | Unsigned | | | Nonnegative | |
| $m$ | $s[\mu]$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ |
| 128 | 3 | 5 | 5 | 7 | 5 | 8 | 0.2 | 1148.1 | 77.8 | 3007.0 | 68.5 |
| 128 | 2 | 5 | 5 | 7 | 5 | 7 | 0.3 | 1297.1 | 73.4 | 2894.4 | 116.8 |
| 152 | 3 | 7 | 7 | 7 | 7 | 58 | 0.3 | 1224.4 | 47.9 | 3997.0 | 186.8 |
| 152 | 4 | 7 | 7 | 13 | 7 | 58 | 0.2 | 1205.8 | 245.0 | 3962.6 | 310.4 |
| 178 | 4 | 9 | 9 | 15 | 9 | 70 | 0.2 | 1269.8 | 238.9 | 4828.2 | 212.0 |
| 178 | 4 | 9 | 9 | 15 | 9 | 19 | 0.3 | 1340.7 | 271.1 | 4923.3 | 342.8 |
| 204 | 4 | 12 | 12 | 15 | 12 | 16 | 0.5 | 2908.1 | 131.2 | 6409.9 | 385.4 |
| 204 | 5 | 12 | 12 | 15 | 12 | 16 | 0.4 | 2996.7 | 148.9 | 5507.9 | 253.9 |
| 230 | 8 | 18 | 18 | 31 | 19 | 31 | 0.3 | 1860.1 | 250.8 | 9046.7 | 331.1 |
| 230 | 8 | 18 | 18 | 31 | 18 | 39 | 0.4 | 2100.2 | 282.8 | 4081.3 | 396.8 |
| 242 | 12 | 26 | 26 | 31 | 27 | 31 | 0.3 | 2015.1 | 92.7 | 7478.2 | 176.2 |
| 242 | 12 | 26 | 26 | 31 | 26 | 31 | 0.3 | 1976.7 | 116.8 | 3597.9 | 412.0 |

**Rademacher matrices**

| | **Unsigned** | | | | **Nonnegative** | | **CPU time (s)** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | LBs on $s_0(A)$ | | | UB | LB | UB | Unsigned | | | Nonnegative | |
| $m$ | $s[\mu]$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ |
| 128 | 1 | 5 | 5 | 14 | 5 | 53 | 27.8 | 1253.1 | 171.6 | 3388.7 | 124.8 |
| 128 | 1 | 5 | 5 | 15 | 5 | 48 | 27.8 | 1361.5 | 191.1 | 3291.6 | 123.4 |
| 152 | 2 | 6 | 6 | 18 | 7 | 65 | 38.4 | 1426.3 | 322.7 | 9592.1 | 136.3 |
| 152 | 1 | 6 | 6 | 19 | 7 | 66 | 38.3 | 1183.0 | 218.9 | 9146.3 | 139.0 |
| 178 | 2 | 7 | 8 | 25 | 9 | 78 | 44.2 | 2819.1 | 258.9 | 8032.1 | 225.8 |
| 178 | 2 | 7 | 8 | 24 | 9 | 78 | 41.8 | 2481.7 | 256.0 | 8306.3 | 168.2 |
| 204 | 2 | 10 | 11 | 32 | 12 | 92 | 51.1 | 1434.2 | 291.8 | 9738.5 | 209.3 |
| 204 | 2 | 10 | 11 | 30 | 12 | 90 | 50.8 | 1316.6 | 448.3 | 9146.8 | 345.4 |
| 230 | 2 | 14 | 16 | 41 | 19 | 107 | 61.8 | 2422.9 | 302.7 | 15235.2 | 162.2 |
| 230 | 2 | 14 | 16 | 39 | 19 | 107 | 61.7 | 2466.2 | 624.0 | 15578.4 | 161.9 |
| 242 | 2 | 20 | 23 | 47 | 27 | 116 | 64.8 | 3929.4 | 269.2 | 19828.7 | 178.1 |
| 242 | 2 | 19 | 23 | 47 | 27 | 111 | 68.0 | 4242.4 | 277.8 | 20506.7 | 270.5 |

**Gaussian matrices**

| | **Unsigned** | | | | **Nonnegative** | | **CPU time (s)** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | LBs on $s_0(A)$ | | | UB | LB | UB | Unsigned | | | Nonnegative | |
| $m$ | $s[\mu]$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_1]$ | $s[\alpha_s]$ | $\bar{s}$ | $s[\alpha_s]$ | $\bar{s}$ |
| 128 | 1 | 5 | 5 | 14 | 5 | 44 | 28.2 | 852.1 | 172.4 | 3283.2 | 114.7 |
| 128 | 1 | 4 | 5 | 15 | 5 | 52 | 27.7 | 1913.9 | 177.7 | 3712.0 | 124.6 |
| 152 | 2 | 6 | 6 | 19 | 7 | 58 | 35.4 | 981.0 | 214.1 | 8433.5 | 392.8 |
| 152 | 1 | 6 | 6 | 19 | 7 | 58 | 38.9 | 1004.0 | 242.6 | 8231.7 | 373.3 |
| 178 | 2 | 7 | 8 | 24 | 9 | 79 | 43.0 | 2164.4 | 393.9 | 10294.7 | 368.2 |
| 178 | 2 | 7 | 8 | 25 | 9 | 77 | 47.6 | 2390.3 | 263.1 | 9548.8 | 374.0 |
| 204 | 2 | 10 | 11 | 32 | 12 | 88 | 58.0 | 1363.6 | 293.3 | 11496.7 | 274.1 |
| 204 | 2 | 10 | 11 | 32 | 12 | 91 | 51.7 | 1218.4 | 293.4 | 12497.2 | 529.5 |
| 230 | 2 | 14 | 17 | 41 | 19 | 102 | 70.4 | 3200.9 | 339.7 | 18771.3 | 431.6 |
| 230 | 2 | 14 | 16 | 39 | 19 | 106 | 61.5 | 2118.4 | 485.4 | 18959.5 | 435.0 |
| 242 | 2 | 19 | 22 | 46 | 27 | 113 | 73.6 | 2212.8 | 277.4 | 26874.6 | 269.2 |
| 242 | 2 | 20 | 23 | 47 | 27 | 112 | 65.3 | 2995.2 | 426.7 | 21308.7 | 191.7 |

become tighter in the case of Hadamard matrices. Further, for some matrices the lower and the upper bound on $s_0(A)$ match (e.g., the Hadamard matrix with $m = 152$), what allows to identify the exact value of $s_0(A)$ . Moreover, we have observed samples of smaller random Hadamard matrices (with $n = 128$) for which the lower bounds and upper bounds on both $s_*(A)$ and $s_0(A)$ coincide, which implies $s_*(A) = s_0(A)$ in these cases.

# 8 Matching pursuit algorithm

The Matching Pursuit algorithm for signal recovery has been first introduced in [15] and is motivated by the desire to provide a reduced complexity alternative to the $\ell_1$-recovery problem. Several implementations of Matching Pursuit has been proposed in the Compressive Sensing literature (see, e.g., the review [2]). All of them are based on successive Euclidean projections of the signal and the corresponding performance results rely upon the bounds on mutual incoherence $\mu(A)$ of the sensing matrix. We are about to show that the LP-based verifiable sufficient conditions from the previous section can be used to construct a specific version of the Matching Pursuit algorithm which we refer to as *Non-Euclidean Matching Pursuit (NEMP) algorithm.*

Suppose that we have in our disposal $\tau, \tau_\pm \geq 0$ and a matrix $Y = [y_1, ..., y_n]$, such that

$$\begin{array}{lll} (a) & -\tau_- \leq [I - Y^T A]_{ij} \leq \tau_+, & \forall i \in P_+,\ \forall j, \\ (b) & -\tau \leq [I - Y^T A]_{ij} \leq \tau, & \forall i \in P_n,\ \forall j, \\ (c) & \|y_j\|_* \leq \sigma, & \forall j. \end{array} \tag{26}$$

Consider a signal $w \in \mathcal{P}_n$ such that $\|w - w^s\|_1 \leq \mu$, where $w^s$ is the vector obtained from $w$ by replacing all but $s$ largest magnitudes of entries in $w$ with zeros, and let $y$ and $\delta$ be such that $\|Aw - y\| \leq \delta$.

Suppose that

$$\rho = s \max\{\tau_+, \tau_-, \tau\} < 1. \tag{27}$$

To simplify notation, we denote $\max[a, b]$ by $a \vee b$. Consider the following iterative procedure:

**Algorithm 1**

1. *Initialization:* *Set* $v^{(0)} = 0$, $\alpha_0 = \frac{\|Y^T y\|_{s,1} + s\sigma\delta + \mu}{1-\rho}$.

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

   (a) $u = Y^T(y - Av^{(k-1)})$ *and* $n$ *segments*

$$S_i = \left\{ \begin{array}{ll} [u_i - \tau_- \alpha_{k-1} - \sigma\delta,\ u_i + \tau_+ \alpha_{k-1} + \sigma\delta], & i \in P_+, \\ [u_i - \tau\alpha_{k-1} - \sigma\delta,\ u_i + \tau\alpha_{k-1} + \sigma\delta], & i \in P_n. \end{array} \right.$$

   *Define* $\Delta \in \mathbf{R}^n$ *by setting*

$$\Delta_i = \left\{ \begin{array}{ll} [u_i - \tau_- \alpha_{k-1} - \sigma\delta]_+, & i \in P_+, \\ [u_i - \tau\alpha_{k-1} - \sigma\delta]_+, & i \in P_n,\ \ u_i \geq 0, \\ -[|u_i| - \tau\alpha_{k-1} - \sigma\delta]_+, & i \in P_n,\ \ u_i < 0 \end{array} \right.$$

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

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

$$\alpha_k = s[2\tau \vee (\tau_- + \tau_+)]\alpha_{k-1} + 2s\sigma\delta + \mu. \tag{28}$$

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

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

**Proposition 8.1** *Assume that* $w_i \geq 0$ *for* $i \in P_+$, (27) *takes place, and that* $\|w - w^s\|_1 \leq \mu$ *with a known in advance value of* $\mu$. *Then the approximate solution* $v^{(k)}$ *and the value* $\alpha_k$ *after the* $k$*-th step of Algorithm 1 satisfy*

$$(a_k) \quad \textit{for all } i \;\; v_i^{(k)} \in \mathrm{Conv}\{0; w_i\}, \qquad (b_k) \quad \|w - v^{(k)}\|_1 \leq \alpha_k.$$

For proof, see Appendix E.

Let

$$\lambda = s[2\tau \vee (\tau_- + \tau_+)];$$

if $\lambda < 1$, then also $\rho < 1$, so that Proposition 8.1 holds true. Furthermore, by (28) the sequence $\alpha_k$ converges exponentially fast to the limit $\alpha_\infty := \frac{2s\sigma\delta+\mu}{1-\lambda}$:

$$\alpha_k = \lambda^k[\alpha_0 - \alpha_\infty] + \alpha_\infty.$$

Note that when $P_+ = \emptyset$, we can set $\tau_- = \tau_+ = 0$ to obtain $\lambda = 2s\tau$; in the case of $P_n = \emptyset$, by setting $\tau = 0$, we have $\lambda = s(\tau_- + \tau_+)$.

The bottom line is: if the optimal value in the convex program

$$\mathrm{Opt} = \min_{\tau,\tau_\pm,Y} \left\{ s[2\tau \vee (\tau_- + \tau_+)] : \begin{array}{cc} -\tau_- \leq [I - Y^T A]_{ij} \leq \tau_+, & \forall i \in P_+,\ \forall j \\ -\tau \leq [I - Y^T A]_{ij} \leq \tau, & \forall i \in P_n,\ \forall j \\ \tau, \tau_\pm \geq 0 & \end{array} \right\}$$

is $< 1$, the above procedure, as yielded by an optimal solution to the latter problem, possesses the following properties:

1. All approximations $v^{(k)}$, $k = 0, 1, ...$ of $w$ are supported on the support of $w$;
2. For $i \in P_+$, $v_i^{(k)} \geq 0$ are nondecreasing in $k$ and are $\leq w_i$ for all $k$;
3. For $i \in P_n$,
   - if $w_i > 0$, then $0 \leq v_i^{(k)} \leq w_i$ and $v_i^{(k)}$ are nondecreasing in $k$;
   - if $w_i < 0$, then $w_i \leq v_i^{(k)} \leq 0$ and $v_i^{(k)}$ are nonincreasing in $k$;
4. As $k$ grows, the upper bound $\alpha_k$ on the $\ell_1$-error of approximating $w$ by $v^{(k)}$ goes exponentially fast to

$$\alpha_\infty = \frac{2s\sigma\delta + \mu}{1 - \mathrm{Opt}}.$$

Let now $\xi \in [0, 1)$, $\sigma \geq 0$ and $\theta \geq 1$ and suppose that an $m \times n$ matrix $A$ satisfies the following condition:

$\overline{\mathbf{VSG}}_s(\xi,\sigma,\theta)$: *There exists $m\times n$ matrix $Y=[y_1,...,y_n]$ such that $\|y_i\|_*\leq\sigma$ for all $i$ and*

$$\begin{array}{ll} -\frac{\xi}{(1+\xi)s}\leq [I-Y^TA]_{ij}\leq \frac{\xi}{(1+\xi)s} & \forall i\notin P_+,\ \forall j,\\ -\frac{\xi}{(1+\xi\theta)s}\leq [I-Y^TA]_{ij}\leq \frac{\xi}{(1+\xi\theta)s} & \forall i\in P_+,\ \forall j\notin P_+,\\ -\frac{\xi}{(1+\xi\theta)s}\leq [I-Y^TA]_{ij}\leq \frac{\xi\theta}{(1+\xi\theta)s} & \forall i,j\in P_+. \end{array} \tag{29}$$

Observe that (29) is a system of convex inequalities in $Y$. Further, $\overline{\mathbf{VSG}}_s(\xi,\sigma,\theta)$ certainly implies $\mathbf{VSG}_s(\xi,\theta,0,\sigma)$, and is therefore sufficient condition for $s$-semigoodness of the matrix $A$.

When $\overline{\mathbf{VSG}}_s(\xi,\sigma,\theta)$ is satisfied with $\xi\in(0,1)$ and $\theta>1$, by taking

$$\tau_-=\frac{\xi}{(1+\xi\theta)s},\quad \tau_+=\frac{\xi\theta}{(1+\xi\theta)s}\quad\text{and}\quad \tau=\frac{\xi}{(1+\xi)s},$$

we obtain

$$\lambda=\max\left(\frac{\xi+\xi\theta}{1+\xi\theta},\ \frac{2\xi}{1+\xi}\right)<1. \tag{30}$$

Combining this condition with Proposition 8.1 gives:

**Corollary 8.1** *Suppose that $A$ satisfies the condition $\overline{\mathbf{VSG}}_s(\xi,\sigma,\theta)$ with certain $\xi\in(0,1)$, $\sigma\geq 0$ and $\theta\geq 1$. Let $w\in\mathcal{P}_n$ be a vector with $\|w-w^s\|_1\leq\mu$ where $w^s$ is the vector obtained from $w$ by replacing all but $s$ largest in magnitude entries in $w$ with zeros, and let $y$ be such that $\|Aw-y\|\leq\delta$. Then the approximate solution $v^{(t)}$ found by Algorithm 1 after $t$ iterations satisfies $v_i^{(t)}\geq 0$ for all $i\in P_+$ and*

$$\|w-v^{(t)}\|_1\leq\frac{2s\sigma\delta+\mu}{1-\lambda}+\lambda^t\left[\frac{\|Y^Ty\|_{s,1}+s\sigma\delta+\mu}{1-\rho}-\frac{2s\sigma\delta+\mu}{1-\lambda}\right],$$

*where $\lambda$ is given by (30) and $\rho=\frac{\xi\theta}{1+\xi\theta}$.*

It should be noted the NEMP algorithm has several drawbacks as compared with the $\ell_1$-recovery. First, the pursuit algorithm requires a priori knowledge of several parameters ($\sigma$, $Y$, $\tau$, $\tau_-$, $\tau_+$, $s$ and $\mu$). Second, the value $(1-\lambda)^{-1}(2s\sigma\delta+\mu)$ is a conservative upper bound on the error of the $\ell_1$-recovery, but the error bound in Corollary 8.1 is exact. On the other hand, the NEMP algorithm can be an interesting option if the $\ell_1$-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 $\ell_1$-recovery, especially when used on high-dimensional data.

Our concluding remark is on the condition

$$\frac{\mu(A)}{1+\mu(A)}<\frac{1}{2s}, \tag{31}$$

where $\mu(A)$ is the mutual incoherence of $A$ (see (6)). This condition is usually used in order to establish convergence results for the Matching Pursuit algorithms (see, e.g. [12, 13, 3]). As it is immediately seen, when $\mu(A)$ is well defined (i.e., all columns in $A$ are nonzero), the matrix $Y=[y_1,...,y_n]$ with the columns

$$y_i=\frac{A_i}{(1+\mu(A))A_i^TA_i}$$

satisfies for all $i = 1, ..., m$ and $j = 1, ..., n$ the relations

$$|[I - Y^T A]_{ij}| \leq \frac{\mu(A)}{1 + \mu(A)}.$$

In the case of (31), setting $\theta = 1$ and specifying $\xi$ from the relation $\frac{\xi}{1+\xi} = \frac{s\mu(A)}{1+\mu(A)}$, we get $0 < \xi < 1$ and meet all inequalities in (29). It follows that $Y$ certifies the validity of the condition $\overline{\mathbf{VSG}}_s(\xi, \sigma, 1)$ with the outlined $\xi$ and with all $\sigma \geq \max_i \frac{\|A_i\|_*}{(1+\mu(A))\|A_i\|_2^2}$, and thus the above $Y$ can be readily used in Matching Pursuit. Note that in the situation in question Corollary 8.1 recovers some results from [12, 13, 3].

## References


[1] Andersen, E. D., Andersen, K. D., *The MOSEK optimization tools manual.* http://www.mosek.com/fileadmin/products/6_0/tools/doc/pdf/tools.pdf.

[2] Bruckstein, A., Donoho, D., Elad, M., From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images, *SIAM Review*, **51**(1), 34-81 (2009).

[3] Bruckstein, A., Elad, M., Zibulevsky, M., A non-negative and sparse enough solution of an underdetermined linear system of equations is unique, *IEEE Transactions on Information Theory*, **54**(11), 4813-4820 (2008).

[4] Candes, E.J., Compressive sampling, Marta Sanz-Solé, Javier Soria, Juan Luis Varona, Joan Verdera, Eds. *International Congress of Mathematicians*, Madrid 2006, Vol. III, 1437-1452. European Mathematical Society Publishing House, (2006).

[5] Carathéodory, C., Ueber den variabilitaetsbereich der fourierschen konstanten von positiven harmonischen funktionen, *Rend. Circ. Mat. Palermo*, **32**, 193-217 (1911).

[6] Cohen, A., Dahmen, W., DeVore, R., Compressed sensing and best k-term approximation, `http://www.math.sc.edu/devore/publications/CDDSensing 6.pdf`, submitted for publication (2006).

[7] d'Aspremont, A., El Ghaoui, L., Testing the Nullspace Property using Semidefinite Programming, `http://arxiv.org/abs/0807.3520`, to appear in *Math. Programm.* (2008).

[8] Donoho, D., Elad, M., Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell_1$ minimization *Proc. of the National Academy of Sciences*, **100**(5), 2197-2202, (2003).

[9] Donoho, D., Huo, X., Uncertainty principles and ideal atomic decomposition *IEEE Transactions on Information Theory* **47**(7), 2845-2862 (2001).

[10] Donoho, D., Tanner, J., Sparse Nonnegative Solutions of Underdetermined Linear Equations by Linear Programming *Proc. of the National Academy of Sciences*, **102**(27), 9446-9451, (2005).

[11] Donoho, D., Tanner, J., Neighborliness of randomly-projected simplices in high dimensions *Proc. of the National Academy of Sciences*, **102**(27), 9452-9457, (2005)

[12] Donoho, D., Elad, M., Temlyakov, V., On Lebesgue-Type Inequalities for Greedy Approximation *Journal of Approximation Theory*, **147**(2), 185-195 (2007).

[13] Elad, M., Optimized projections for compressed sensing *IEEE Trans. on Signal Processing*, **55**(12), 5695-5702 (2007).

[14] Juditsky, A., Nemirovski, A., On Verifiable Sufficient Conditions for Sparse Signal Recovery via $\ell_1$ Minimization `http://hal.archives-ouvertes.fr/hal-00321775/` , submitted to *Math. Programm.* (2008).

[15] Mallat, S., Zhang, A., Matching pursuits with time-frequency dictionaries *IEEE Transactions on Signal Processing*, **41**(12), 3397-3415 (1993).

[16] Zhang, Y., A simple proof for recoverability of ell-1-minimization: go over or under? Technical Report TR05-09, Department of Computational and Applied Mathematics, Rice University, Houston, TX (2005).

[17] Zhang, Y., A simple proof for recoverability of ell-1-minimization (II): the nonnegative case. Technical report TR05-10, Department of Computational and Applied Mathematics, Rice University, Houston, TX (2005).

[18] Zhang, Y., Theory of Compressive Sensing via l1 minimization: a non-rip analysis and extensions. Technical report TR08-11, Department of Computational and Applied Mathematics, Rice University, Houston, TX (2008).

# A Proof of Proposition 2.1

**(i)⇒(ii):** Let $A$ be $s$-semigood, and let, in contrast to what is stated by (ii), $J$ be a subset of $\{1,...,n\}$ with $\mathrm{Card}(J) \leq s$ and $x \in \mathrm{Ker}A \backslash \{0\}$ be such that $x_i \leq 0$ for all $i \in P_+ \setminus J$ and

$$\sum_{i\in J\cap P_+} x_i + \sum_{i\in J\cap P_n} |x_i| \geq \sum_{i\notin J} |x_i|.$$

Let $I = (J \cap P_n) \cup \{i \in J \cap P_+ : x_i \geq 0\}$ so that $I \subseteq J$. From the construction of $I$, we have $x_i \leq 0$ for $i \in J \setminus I$ implying that $x_i \leq 0$ for $i \in P_+ \setminus I$. Further,

$$\begin{aligned}\sum_{i\in I\cap P_+} x_i + \sum_{i\in I\cap P_n} |x_i| &= \sum_{i\in J\cap P_+} x_i - \sum_{i\in J\setminus I} x_i + \sum_{i\in J\cap P_n} |x_i| \\ &\geq \sum_{i\notin J} |x_i| - \sum_{i\in J\setminus I} x_i = \sum_{i\notin J} |x_i| + \sum_{i\in J\setminus I} |x_i| = \sum_{i\notin I} |x_i|.\end{aligned}$$

Hence $I$ also violates the condition in (ii). Setting $u_i = x_i$ when $i \in I$ and $u_i = 0$ otherwise and setting $v = u - x$, we have $u_i \geq 0$ for any $i \in I \cap P_+$, $u_i = 0$ for any $i \in P_+ \setminus I$, and $v_i \geq 0$ for $i \in P_+ \setminus I$, $v_i = 0$ for $i \in I \cap P_+$ and $\sum_i |u_i| \geq \sum_i |v_i|$. In addition, $Au = Av$ due to $Ax = 0$, and $u$ is $s$-sparse; finally, $u \neq v$ due to $x \neq 0$. We see that the $s$-sparse vector $u \in \mathcal{P}_n$ is not the unique solution to

$$\min_z \left\{ \sum_i |z_i| : \; Az = Au, \quad z_i \geq 0 \; \forall i \in P_+ \right\},$$

which is a desired contradiction.

**(ii)⇒(iii):** Let $A$ satisfy (ii). Let $\mathcal{J}$ be the family of all subsets $J$ of $\{1,...,n\}$ of cardinality $\leq s$. For $J \in \mathcal{J}$, let

$$X_J = \{x \in \mathrm{Ker} A : \|x\|_1 = 1,\ x_i \leq 0\ \ \forall i \in P_+ \setminus J\}.$$

Assuming that $X_J \neq \emptyset$, let $x \in X_J$. By (ii), we have

$$\sum_{i \in J \cap P_+} x_i + \sum_{i \in J \cap P_n} |x_i| < \sum_{i \notin J} |x_i|.$$

We claim that $\sum_{i \notin J} |x_i| > 0$.

> Indeed, otherwise $x_i \neq 0$ implies that $i \in J$. Let $I_+$ and $I_-$ be the subsets of $J$ such that $x_i > 0$ for $i \in I_-$ and $x_i < 0$ for $i \in I_+$. At least one of these sets is nonempty due to $x \neq 0$. W.l.o.g. we can assume that $\sum_{i \in I_+} x_i \geq \sum_{i \in I_-} |x_i|$ (otherwise we could replace $x$ with $-x$ and swap $I_+$ and $I_-$). Applying (ii) to $x$ and to $I_+$ in the role of $J$, we should have
>
> $$\sum_{i \in I_+ \cap P_+} x_i + \sum_{i \in I_+ \cap P_n} |x_i| = \sum_{i \in I_+} x_i < \sum_{i \notin I_+} |x_i| = \sum_{i \in I_-} |x_i|,$$
>
> which is not the case. This contradiction shows that $\sum_{i \notin J} |x_i| > 0$ whenever $x \in X_J$.

From our claim it follows that the function

$$\frac{\sum_{i \in J \cap P_+} x_i + \sum_{i \in J \cap P_n} |x_i|}{\sum_{i \notin J} |x_i|}$$

is continuous on $X_J$ and is $< 1$ at every point of this set. Since $X_J$ is compact, we conclude that when $J \in \mathcal{J}$ is such that $X_J \neq \emptyset$, there exists $\xi_J < 1$ such that

$$\sum_{i \in J \cap P_+} x_i + \sum_{i \in J \cap P_n} |x_i| \leq \xi_J \sum_{i \notin J} |x_i| \text{ for any } x \in X_J.$$

Setting $\xi = \max\limits_{J \in \mathcal{J} : X_J \neq \emptyset} \xi_J$, we clearly ensure the validity of (iii). The implication (ii)⇒(iii) is proved.

**(iii)⇒(i):** Let (iii) take place; let us prove that $A$ is $s$-semigood. Thus, let $u$ with $u_i \geq 0$ for all $i \in P_+$ be $s$-sparse; we should prove that $u$ is the unique optimal solution to the problem

$$\min_z \left\{ \sum_i |z_i| : Az = Au,\ z_i \geq 0\ \forall i \in P_+ \right\}.$$

Assume, on the contrary to what should be proved, that the latter problem has an optimal solution $v$ different from $u$, and let $x = u - v$, so that $x \in \mathrm{Ker} A$ and $x \neq 0$. Setting $I = \{i : u_i \neq 0\}$, we have $\mathrm{Card}(I) \leq s$ and $x_i \leq 0$ when $i \in P_+ \setminus I$, whence by (iii)

$$\sum_{i \in I \cap P_+} x_i + \sum_{i \in I \cap P_n} |x_i| \leq \xi \sum_{i \notin I} |x_i| = \xi \sum_{i \notin I} |v_i|,$$

whence also

$$\underbrace{\sum_{i\in I\cap P_+} u_i + \sum_{i\in I\cap P_n} |u_i|}_{=\sum_{i\in I}|u_i|} \leq \underbrace{\sum_{i\in I\cap P_+} v_i + \sum_{i\in I\cap P_n} |v_i|}_{=\sum_{i\in I}|v_i|} + \xi\sum_{i\notin I}|v_i|. \tag{32}$$

Since $\sum_i |v_i| \leq \sum_i |u_i| = \sum_{i\in I}|u_i|$ due to the origin of $v$, (32) implies that $\sum_{i\notin I}|v_i| = 0$, that is, both $u$ and $v$ are supported on $I$, so that $x$ is supported on $I$ as well. Now let $I_+ = \{i \in I\cap P_+ : x_i \geq 0\}$, $I_- = \{i \in I \cap P_+ : x_i < 0\}$ and $I_n = I \cap P_n$. Replacing, if necessary, $x$ with $-x$ and swapping $I_+$ and $I_-$, we can assume that $\sum_{i\in I_+} x_i = \sum_{i\in I_+}|x_i| \geq \sum_{i\in I_-}|x_i|$. Applying (iii) to $x$ and to $I_+ \cup I_n$ in the role of $J$, we get

$$\sum_{i\in I_+} x_i + \sum_{i\in I_n}|x_i| \leq \xi\sum_{i\in I_-}|x_i|,$$

thereby $\sum_{i\in I_+} x_i = \sum_{i\in I_n}|x_i| = \sum_{i\in I_-}|x_i| = 0$ due to $\sum_{i\in I_+} x_i \geq \sum_{i\in I_-}|x_i|$. Thus, $x = 0$, which is a desired contradiction.

We have proved that the properties (i) – (iii) of $A$ are equivalent to each other.

**(iii)⇔(iv):** The implication (iv)⇒(iii) is evident. Let us prove the inverse implication. Thus, let $A$ satisfy (iii) (and thus – (i) – (ii) as well), and let $\xi' \in (\xi, 1)$. Let, as above, $\mathcal{J}$ be the family of all subsets $J$ of $\{1,...,n\}$ of cardinality $\leq s$. Let $X = \{x \in \mathrm{Ker}A : \|x\|_1 = 1\}$, and let $J \in \mathcal{J}$. Let $x \in X$. We claim that there exists a neighborhood $U_x$ of $x$ in $X$ and $\theta_{J,x} \in [1,\infty)$ such that for any $u \in U_x$ and $\theta \geq \theta_{J,x}$ it holds

$$\sum_{i\in J\cap P_+} u_i + \sum_{i\in J\cap P_n}|u_i| \leq \xi'\left(\sum_{i\in P_n\setminus J}|u_i| + \sum_{i\in P_+\setminus J}\max[-u_i, \theta u_i]\right). \tag{33}$$

The claim is clearly true when there exists $i \in P_+ \setminus J$ such that $x_i > 0$. Now assume that $x_i \leq 0$ for $i \in P_+ \setminus J$. Then $\sum_{i\notin J}|x_i| > 0$. Indeed, otherwise $x_i = 0$ for all $i \notin J$, which combines with $s$-semigoodness of $A$ and the relation $Ax = 0$ to imply that $x = 0$ (since assuming $x \neq 0$, we have $x = u - v$ with $s$-sparse $u \geq 0, v \geq 0$ with non-overlapping supports, and $Au = Av$ due to $Ax = 0$, which of course contradicts the $s$-semigoodness of $A$), while $x$ definitely is nonzero (since $\|x\|_1 = 1$ due to $x \in X$). Now, since $x \in \mathrm{Ker}A$ and $x_i \leq 0$, $i \in P_+ \setminus J$, we have

$$\sum_{i\in J\cap P_+} x_i + \sum_{i\in J\cap P_n}|x_i| \leq \xi\sum_{i\notin J}|x_i| < \xi'\sum_{i\notin J}|x_i|$$

where the first inequality is due to (iii), and the second – due to $\sum_{i\notin J}|x_i| > 0$. The concluding strict inequality clearly implies the validity of (33) with $\theta = 1$, provided that $U_x$ is a small enough neighborhood of $x$. Thus, our claim is true.

From the validity of our claim, extracting from the covering $\{U_x\}_{x\in X}$ of the compact set $X$ a finite subcovering, we conclude that there exists $\theta_J \in [1,\infty)$ such that

$$\forall (x \in X,\ \theta \geq \theta_J): \sum_{i\in J\cap P_+} x_i + \sum_{i\in J\cap P_n}|x_i| \leq \xi'\left(\sum_{i\in P_n\setminus J}|x_i| + \sum_{i\in P_+\setminus J}\max[-x_i, \theta x_i]\right).$$

Setting $\theta = \max_{J\in\mathcal{J}} \theta_J$, we see that $A$ satisfies $\mathbf{SG}_s(\xi', \theta)$.

**(iv)⇒(v):** Let $A$ satisfy $\mathbf{SG}_s(\xi, \theta)$ for certain $\xi \in (0, 1)$, $\theta \in [1, \infty)$ and let $\|\cdot\|$ be a norm on $\mathbf{R}^m$. Let, further, $P$ be the orthogonal projector of $\mathbf{R}^n$ on $\mathrm{Ker}A$. Then clearly with a properly chosen $C$ one has

$$\|Px - x\|_1 \le C\|Ax\|$$

for any $x \in \mathbf{R}^n$. Now let $J$ be a subset of $\{1, ..., n\}$ of cardinality $\le s$, $x \in \mathbf{R}^n$ and $u = Px$. We have

$$\begin{aligned}
\sum_{i\in J\cap P_+} x_i + \sum_{i\in J\cap P_n} |x_i| &\le \sum_{i\in J\cap P_+} u_i + \sum_{i\in J\cap P_n} |u_i| + \sum_{i\in J} |u_i - x_i| \\
&\le \xi \left[\sum_{i\in P_n\setminus J} |u_i| + \sum_{i\in P_+\setminus J} \max[-u_i, \theta u_i]\right] + \sum_{i\in J} |u_i - x_i| \\
&\le \xi \left[\sum_{i\in P_n\setminus J} [|x_i| + |u_i - x_i|] + \sum_{i\in P_+\setminus J} [\max[-x_i, \theta x_i] + \theta|x_i - u_i|]\right] + \sum_{i\in J} |u_i - x_i| \\
&\le \xi \left[\sum_{i\in P_n\setminus J} |x_i| + \sum_{i\in P_+\setminus J} \max[-x_i, \theta x_i]\right] + \max[1, \theta\xi]\|x - u\|_1 \\
&\le \xi \left[\sum_{i\in P_n\setminus J} |x_i| + \sum_{i\in P_+\setminus J} \max[-x_i, \theta x_i]\right] + \max[1, \theta\xi]C\|Ax\|,
\end{aligned}$$

so that $A$ satisfies $\mathbf{SG}_{s,\beta}(\xi, \theta)$ with $\beta = \max(1, \theta\xi)C$. The implication (iv)⇒(v) is proved.

**(v)⇒(vi)⇒(iii):** These implications are evident. □

# B Proof of Proposition 3.1

Let $I$ be the support of $w^s$, $\bar{I}$ be the complement of $I$ in $\{1, ..., n\}$, and let $z = w - x$. We denote $I_+ = \{i \in I : z_i \ge 0\}$, $\bar{I}_+ = \{i \in \bar{I} : z_i \ge 0\}$, and $I_- = I \setminus I_+$, $\bar{I}_- = \bar{I} \setminus \bar{I}_+$. Observe that $w$ is a feasible solution to (10), so that

$$\|x\|_1 \le \|w\|_1 + \nu. \tag{34}$$

Obviously, $|x_i| - |w_i| \geq -|z_i|$ and $|x_i| - |w_i| \geq |z_i| - 2|w_i|$. Now using $x_i, w_i \geq 0\ \forall i \in P_+$, and $z_i \geq 0\ \forall i \in I_+$, we get

$$\begin{array}{rcl}
\nu & \geq & \sum_i [|x_i| - |w_i|] \qquad \text{[by (34)]} \\
& \geq & \sum_{i\in I_+\cap P_+} \underbrace{(x_i - w_i)}_{=-z_i} + \sum_{i\in I_-\cap P_+} \underbrace{(x_i - w_i)}_{=-z_i=|z_i|} + \sum_{i\in \bar{I}_-\cap P_+} \underbrace{(x_i - w_i)}_{=-z_i=|z_i|} + \sum_{i\in \bar{I}_+\cap P_+} \underbrace{(x_i - w_i)}_{=-z_i\geq -w_i} \\
& & + \sum_{i\in P_n} (|x_i| - |w_i|) \\
& \geq & - \sum_{i\in I_+\cap P_+} z_i + \sum_{i\in I_-\cap P_+} |z_i| + \sum_{i\in \bar{I}_-\cap P_+} |z_i| - \sum_{i\in \bar{I}_+\cap P_+} w_i \\
& & - \sum_{i\in I\cap P_n} |z_i| + \sum_{i\in \bar{I}\cap P_n} (|z_i| - 2|w_i|),
\end{array}$$

or, equivalently,

$$\begin{array}{l}
\sum_{i\in I_-\cap P_+} |z_i| + \sum_{i\in \bar{I}_-\cap P_+} |z_i| + \sum_{i\in \bar{I}\cap P_n} |z_i| \\
\leq \nu + \sum_{i\in I_+\cap P_+} z_i + \sum_{i\in I\cap P_n} |z_i| + \sum_{i\in \bar{I}_+\cap P_+} w_i + 2\sum_{i\in \bar{I}\cap P_n} |w_i|.
\end{array} \tag{35}$$

On the other hand, we have

$$\|Az\| = \|Aw - Ax\| \leq \|Aw - y\| + \|Ax - y\| \leq \varepsilon + \delta. \tag{36}$$

Then by condition $\mathbf{SG}_{s,\beta}(\xi,\theta)$ with $(I_+ \cap P_+) \cup (I \cap P_n)$ in the role of $J$, we get

$$\begin{array}{l}
\underbrace{\sum_{i\in I_+\cap P_+} z_i + \sum_{i\in I\cap P_n} |z_i|}_{:=\kappa} \leq \beta\|Az\| + \xi \left[\sum_{i\in \bar{I}\cap P_n} |z_i| + \sum_{i\in (\bar{I}\cap P_+)\cup(I_-\cap P_+)} \psi(z_i)\right] \\
\kappa \leq \beta\|Az\| + \xi \Big[ \underbrace{\sum_{i\in \bar{I}\cap P_n} |z_i| + \sum_{i\in I_-\cap P_+} |z_i| + \sum_{i\in \bar{I}_-\cap P_+} |z_i| + \theta \sum_{i\in \bar{I}_+\cap P_+} z_i}_{:=\tau(\theta)} \Big]
\end{array} \tag{37}$$

Let us derive a bound on $\tau(\theta)$. Now (35) implies, independently of whether $\mathbf{SG}_{s,\beta}(\xi,\theta)$ is or is not true, the first inequality in the following chain:

$$\begin{array}{rcl}
\tau(\theta) & \leq & \nu + \sum_{i\in I_+\cap P_+} z_i + \sum_{i\in I\cap P_n} |z_i| + \sum_{i\in \bar{I}_+\cap P_+} w_i + 2\sum_{i\in \bar{I}\cap P_n} |w_i| + \theta \sum_{i\in \bar{I}_+\cap P_+} z_i \\
& \leq & \nu + \kappa + (1+\theta) \sum_{i\in \bar{I}_+\cap P_+} w_i + 2\sum_{i\in \bar{I}\cap P_n} |w_i| \qquad \text{[since } w_i \geq z_i \text{ for } i \in P_+] \\
& \leq & \nu + \kappa + (1+\theta)\mu, \qquad \text{[since } \theta \geq 1 \text{ and } \sum_{i\in \bar{I}} |w_i| \leq \mu],
\end{array} \tag{38}$$

and, in particular,

$$\tau(1) = \sum_{i\in I_-\cap P_+} |z_i| + \sum_{i\in \bar{I}} |z_i| \leq \nu + \kappa + 2\mu. \tag{39}$$

Combining (36), (37) and (38), we obtain

$$\kappa \le \beta(\varepsilon+\delta) + \xi\left[\nu+\kappa+(1+\theta)\mu\right],$$

and thereby,

$$\kappa = \sum_{i\in I_+\cap P_+} z_i + \sum_{i\in I\cap P_n} |z_i| \le \frac{\beta(\varepsilon+\delta)+\xi(\nu+(\theta+1)\mu)}{1-\xi}.$$

Summing up the latter inequality and (39), we obtain

$$\begin{array}{rcl}\|z\|_1 &=& \sum_{i\in I\cap P_n} |z_i| + \sum_{i\in I_+\cap P_+} z_i + \left[\sum_{i\in I_-\cap P_+} |z_i| + \sum_{i\in \bar I} |z_i|\right] \le \nu+2\mu+2\kappa \\ &\le& \nu+2\mu+\frac{2\beta(\varepsilon+\delta)+2\xi(\nu+(\theta+1)\mu)}{1-\xi} = \frac{1+\xi}{1-\xi}\nu + \frac{2(1+\xi\theta)}{1-\xi}\mu + \frac{2\beta}{1-\xi}(\varepsilon+\delta),\end{array}$$

which is (11).

To show (12) observe that increasing $\varepsilon$ to $\varepsilon' = \varepsilon + \alpha\mu$, we can think that the true signal underlying the observation $y$ is $w^s$ rather than $w$; note that (34) implies that

$$\|x\|_1 \le \|w^s\|_1 + \nu',\ \ \nu' = \nu + \mu. \tag{40}$$

We can now repeat the reasoning which follows (34), with (40) in the role of (34), $w^s$ in the role of $w$, $\varepsilon'$ in the role of $\varepsilon$ and 0 in the role of $\mu$, thus arriving at the following analogy of the bound (11):

$$\|x-w^s\|_1 \le \frac{1+\xi}{1-\xi}\nu' + \frac{2\beta}{1-\xi}(\varepsilon'+\delta),$$

whence

$$\|x-w\|_1 \le \frac{1+\xi}{1-\xi}\nu' + \frac{2\beta}{1-\xi}(\varepsilon'+\delta)+\mu,$$

which is nothing but (12). □

# C Proof of Proposition 4.1

Let $A$ satisfy $\mathbf{VSG}_s(\xi,\theta,\rho,\sigma)$, and let $Y=[y_1,...,y_n]$ and $v$ satisfy (13). Let, further, $I\subset\{1,...,n\}$ be such that $\mathrm{Card}(I)\le s$, and let $x\in\mathbf{R}^n$. Let $u\in\mathbf{R}^n$ be given by

$$u_i = \left\{\begin{array}{ll} 1+\theta\xi, & i\in P_+\cap I,\ x_i\ge 0\\ 1-\xi, & i\in P_+\cap I,\ x_i<0\\ (1+\xi)\,\mathrm{sign}(x_i), & i\in P_n\cap I\\ 0, & i\notin I\end{array}\right..$$

Note that $u$ has at most $s$ nonzero entries, the entries of $u$ with indices from $P_+$ belong to $[0,1+\theta\xi]$, and the modulae of entries in $u$ with indices from $P_n$ are $\le 1+\xi$, so that $u^Tz\le\Phi_s(z)$ for all $z$.

We have

$$\begin{array}{rcl}
u^T[I-Y^TA]x &=& \sum_i u^TC_i[Y,A]x_i = \sum_{i:x_i\geq 0} u^TC_i[Y,A]x_i + \sum_{i:x_i<0} u^T[-C_i[Y,A]]|x_i| \\
&\leq& \sum_{i:x_i\geq 0}\Phi_s(C_i[Y,A])x_i + \sum_{i:x_i<0}\Phi_s(-C_i[Y,A])|x_i| \qquad [\text{since } u^Tz\leq \Phi_s(z)] \\
&\leq& \sum_{i:x_i\geq 0, i\not\in P_+}[\xi+(A^Tv)_i]x_i + \sum_{i:x_i\geq 0, i\in P_+}[\theta\xi+(A^Tv)_i]x_i + \sum_{i:x_i<0}[\xi-(A^Tv)_i]|x_i| \quad [\text{by (13)}] \\
&=& \xi\left[\sum_{i:x_i\geq 0, i\not\in P_+} x_i + \theta\sum_{i:x_i\geq 0, i\in P_+} x_i + \sum_{i:x_i<0}|x_i|\right] + x^TA^Tv \\
&=& \xi\left[\sum_{i\in P_+}\max[-x_i,\theta x_i] + \sum_{i\in P_n}|x_i|\right] + x^TA^Tv,
\end{array}$$

whence

$$u^T[I-Y^TA]x \leq \xi\left[\sum_{i\in P_+}\max[-x_i,\theta x_i] + \sum_{i\in P_n}|x_i|\right] + \rho\|Ax\| \tag{41}$$

(recall that $\|v\|_*\leq\rho$). On the other hand, recalling the definition of $u$ and that $\|y_i\|_*\leq\sigma$, we have

$$\begin{array}{rcl}
u^T[I-Y^TA]x &=& u^Tx - \sum_{i\in I} u_iy_i^TAx \\
&=& \sum_{i\in I\cap P_+}\max[(1-\xi)x_i,(1+\theta\xi)x_i] + (1+\xi)\sum_{i\in I\cap P_n}|x_i| - \sum_{i\in I}u_iy_i^TAx \\
&\geq& \sum_{i\in I\cap P_+}\max[(1-\xi)x_i,(1+\theta\xi)x_i] + (1+\xi)\sum_{i\in I\cap P_n}|x_i| \\
&& \underbrace{-\,\sigma\left[\sum_{i\in I\cap P_+}(1+\theta\xi) + \sum_{i\in I\cap P_n}(1+\xi)\right]}_{\leq\beta-\rho}\|Ax\|.
\end{array}$$

Combining the resulting inequality with (41), we get

$$\sum_{i\in I\cap P_+}[x_i+\xi\max[-x_i,\theta x_i]] + (1+\xi)\sum_{i\in I\cap P_n}|x_i| \leq \beta\|Ax\| + \xi\left[\sum_{i\in P_+}\max[-x_i,\theta x_i] + \sum_{i\in P_n}|x_i|\right]$$

with $\beta$ given by (14), or, equivalently,

$$\sum_{i\in I\cap P_+} x_i + \sum_{i\in I\cap P_n}|x_i| \leq \beta\|Ax\| + \xi\left[\sum_{i\in P_+\backslash I}\max[-x_i,\theta x_i] + \sum_{i\in P_n\backslash I}|x_i|\right].$$

The latter relation holds true for every $x\in\mathbf{R}^n$ and for every set $I\subset\{1,...,n\}$ of cardinality $\leq s$, so that $A$ satisfies $\mathbf{SG}_{s,\beta}(\xi,\theta)$. $\square$

## D Proof of Proposition 5.1

Proof is based on the following

**Lemma D.1** *Let $Z$ be a $\nu \times \nu$ matrix of rank $m$, $s > 1$ be a positive integer, and $\delta_i \in (0,1]$, $1 \le i \le \nu$, be such that for the columns $C_i$ of the matrix $I_\nu - Z$ it holds $\|C_i\|_{s,1} \le 1 - \delta_i$. Assume that*

$$\nu > (2\sqrt{2m}+1)^2. \tag{42}$$

*Then*

$$s \le 2\sqrt{2m}+1. \tag{43}$$

**Proof of the lemma.** Let $\sigma_i = Z_{ii}$, and let $\gamma_i$ be the sum of $s-1$ largest magnitudes of the entries in $C_i$ with indices different from $i$. We have

$$1 - \sigma_i + \gamma_i \le \|C_i\|_{s,1} \le 1 - \delta_i,$$

consequently $\sigma_i \ge \delta_i + \gamma_i > 0$. Let us set $\lambda_i = \frac{1}{\sigma_i}$, and let $\bar Z$ be the matrix with the columns $\bar Z_i = \lambda_i Z_i$, where $Z_i$ is the $i$-th column in $Z$. Note that $\bar Z$ is of the same rank $m$ as $Z$, and that $\bar Z_{ii} = 1$ for all $i$. Recalling that $\gamma_i < \sigma_i$, we have also

$$\|\bar Z_i\|_{s-1,1} = \lambda_i \|Z_i\|_{s-1,1} \le \lambda_i[\gamma_i + \sigma_i] \le 2\lambda_i\sigma_i = 2.$$

Now let $\bar s = \min[s-1, \lfloor \nu^{1/2} \rfloor]$, so that $\bar s \ge 1$ due to $s > 1$. We have $\|\bar Z_i\|_{\bar s,1} \le \|\bar Z_i\|_{s-1,1} \le 2$ and $\bar s^2 \le \nu$. From the latter inequality and due to $\|\bar Z_i\|_2^2 \le \max\{1, \nu \bar s^{-2}\}\|\bar Z_i\|_{\bar s,1}^2$ (cf. the proof of [14, Proposition 4.2]), it follows that $\|\bar Z_i\|_2^2 \le 4\nu\bar s^{-2}$. We conclude that $\|\bar Z\|_2^2 \le 4\nu^2 \bar s^{-2}$, where for a matrix $B$, $\|B\|_2$ is the Frobenius norm of $B$. Setting $H = \frac{1}{2}[\bar Z + \bar Z^T]$, we have therefore $\|H\|_2^2 \le 4\nu^2\bar s^{-2}$. On the other hand, $\mathrm{Tr}(H) = \sum_{i=1}^{\nu} \bar Z_{ii} = \nu$, while $\mathrm{rank}(H) \le 2m$, whence, denoting by $\mu_i$, $1 \le i \le p \le 2m$, the nonzero eigenvalues of $H$, we have

$$\|H\|_2^2 = \sum_{i=1}^{p} \mu_i^2 \ge (\sum_{i=1}^{p} \mu_i)^2/p = (\mathrm{Tr}(H))^2/p \ge \nu^2/(2m).$$

We arrive at the inequality $4\nu^2\bar s^{-2} \ge \|H\|_2^2 \ge \nu^2/(2m)$, thereby

$$\bar s^2 \le 8m. \tag{44}$$

Assuming that $\bar s = \lfloor \nu^{1/2} \rfloor$, (44) says that $\nu \le (2\sqrt{2m}+1)^2$, which is impossible. The only other option is that $\bar s = s-1$, and we arrive at (43). □

**Lemma D.1 ⇒ Proposition 5.1:** Let $Y, v$ satisfy (13). Consider first the case when $\nu := \mathrm{Card}(P_n) \ge n/2$. Denoting by $\widehat{C}_i$ the $\nu$-dimensional vector comprised of the last $\nu$ entries in $C_i = C_i[Y, A]$ (i.e., entries with indices from $P_n$). By (13), for every $i \in P_n$ and for every set $I \subset P_n$ with $\mathrm{Card}(I) \le s$ we have

$$\textstyle\sum_{j\in I}(1+\xi)|[C_i]_j| \le \Phi_s(-C_i) \le \xi - (A^Tv)_i, \quad \sum_{j\in I}(1+\xi)|[C_i]_j| \le \Phi_s(C_i) \le \xi + (A^Tv)_i,$$

thus for any $i \in P_n$,

$$2(1+\xi)\|\widehat{C}_i\|_{s,1} \le \Phi_s(-C_i) + \Phi_s(C_i) \le 2\xi,$$

so that $\|\widehat{C}_i\|_{s,1} < 1/2$. We see that the South-Eastern $\nu \times \nu$ submatrix $Z$ of $Y^TA$ satisfies the premise of Lemma D.1, while the size $\nu$ of $Z$ satisfies (42) due to (21) and $\nu \geq n/2$. Applying the lemma, we arrive at (22).

Now consider the case when $\mathrm{Card}(P_n) < n/2$, that is, $\nu := \mathrm{Card}(P_+) \geq n/2$. By (13), setting $C_i = C_i[Y, A]$, for every set $I \subset P_+$ with $\mathrm{Card}(I) \leq s$ and every $i \in P_+$ we have

$$\begin{array}{l}\sum_{j\in I}(1+\theta\xi)\max[-[C_i]_j, 0] \leq \Phi_s(-C_i) \leq \xi - (A^Tv)_i,\\ \sum_{j\in I}(1+\theta\xi)\max[[C_i]_j, 0] \leq \Phi_s(C_i) \leq \theta\xi + (A^Tv)_i,\end{array}$$

whence

$$\sum_{j\in I}|[C_i]_j| \leq \frac{\xi(1+\theta)}{1+\theta\xi} < 1.$$

Since the latter inequality holds true for every subset $I$ of $P_+$ with $\mathrm{Card}(I) \leq s$, when denoting by $\bar{C}_i$ the part of $C_i$ comprised of the first $\nu$ entries (those with indexes from $P_+$), we have for all $i \in P_+$:

$$\|\bar{C}_i\|_{s,1} < 1.$$

Now the proof can be completed exactly as in the previous case, with the North-Western $\nu \times \nu$ submatrix of $Y^TA$ in the role of $Z$. $\square$

# E Proof of Proposition 8.1

Let us proceed by induction. 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 $z^{(k-1)} = w - v^{(k-1)}$. By $(a_{k-1})$, $z^{(k-1)}$ is supported on the support of $w$ and is such that $z_i^{(k-1)} \geq 0$ for $i \in P_+$. Note that

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

where $e = y - Aw$ with $\|Y^Te\|_\infty \leq \sigma\delta$ due to (26.c). Then by (26.a,b) for any $i \in P_+$,

$$-\tau_-\left[\sum_{j\in P_+} z_j^{(k-1)} + \sum_{j\in P_n}|z_j^{(k-1)}|\right] - \sigma\delta \leq z_i^{(k-1)} - u_i \leq \tau_+\left[\sum_{j\in P_+} z_j^{(k-1)} + \sum_{j\in P_n}|z_j^{(k-1)}|\right] + \sigma\delta,$$

consequently,

$$-\gamma_- := -\tau_-\alpha_{k-1} - \sigma\delta \leq z_i^{(k-1)} - u_i \leq \gamma_+ := \tau_+\alpha_{k-1} + \sigma\delta. \tag{45}$$

We conclude that for any $i \in P_+$ the interval $S_i = [u_i - \gamma_-,\, u_i + \gamma_+]$ of the width

$$\ell_+ = [\tau_- + \tau_+]\alpha_{k-1} + 2\sigma\delta,$$

covers $z_i^{(k-1)}$. In the same way for any $i \in P_n$

$$-\gamma := -\tau\alpha_{k-1} - \sigma\delta \leq z_i^{(k-1)} - u_i \leq \tau\alpha_{k-1} + \sigma\delta = \gamma,$$

so that the interval $S_i = [u_i - \gamma, u_i + \gamma]$ of the width

$$\ell = 2\tau\alpha_{k-1} + 2\sigma\delta,$$

covers $z_i^{(k-1)}$ when $i \in P_n$.

Recalling that $z_i^{(k-1)} \geq 0$ for $i \in P_+$, the closest to 0 point of $S_i$ is

$$\begin{array}{rclcrcl} \widetilde{\Delta}_i &=& [u_i - \gamma_-]_+ \quad \text{for } i \in P_+, & & \widetilde{\Delta}_i &=& [u_i - \gamma]_+ \quad \text{for } i \in P_n,\ \ u_i \geq 0, \\ \widetilde{\Delta}_i &=& -[|u_i| - \gamma]_+ \quad \text{for } i \in P_n,\ \ u_i < 0, & & & & \end{array}$$

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

$$(a) \quad \Delta_i \in \mathrm{Conv}\left\{0, z_i^{(k-1)}\right\}, \qquad (b) \quad |z_i^{(k-1)} - \Delta_i| \leq \ell \vee \ell_+. \tag{46}$$

Since $(a_{k-1})$ is valid, (46.a) implies that

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

and $(a_k)$ holds. Further, let $I$ be the support of $w^s$. Relation $(a_k)$ clearly implies that $|z_i^{(k)}| \leq |w_i|$, and we can write due to (46.b):

$$\begin{aligned} \|w - v^{(k)}\|_1 &= \sum_{i \in I} |w - [v_i^{(k-1)} + \Delta_i]| + \sum_{i \notin I} |z_i^{(k)}| \\ &\leq \sum_{i \in I} |z_i^{(k-1)} - \Delta_i| + \sum_{i \notin I} |w_i| \leq s[\ell \vee \ell_+] + \mu = \alpha_k, \end{aligned}$$

which is $(b_k)$. 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

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

and let $u = Y^T y$. Same as above (cf. (45)), we have for all $i$:

$$|w_i - u_i| \leq \max\{\tau_-, \tau_+, \tau\}\alpha_* + \sigma\delta = \frac{\rho}{s}\alpha_* + \sigma\delta.$$

Then

$$\alpha_* = \sum_{i \in I} |w_i| + \sum_{i \notin I} |w_i| \leq \sum_{i \in I} [|u_i| + \frac{\rho}{s}\alpha_* + \sigma\delta] + \mu \leq \|u\|_{s,1} + \rho\alpha_* + s\sigma\delta + \mu.$$

Hence

$$\alpha_* \leq \alpha_0 = \frac{\|u\|_{s,1} + s\sigma\delta + \mu}{1 - \rho},$$

which implies $(b_0)$. $\square$

# F ONLINE SUPPLEMENT

## F.1 Proof of Proposition 4.2

Let $Y=[Y_1,...,Y_n],v,\sigma,\rho$ certify the validity of $\mathbf{VSG}^*_{s,\beta}(\xi,\theta)$, and let $\beta'\geq\beta$, $\theta'\geq\theta$ and $\xi'\in[\xi,1)$. Let us set

$$\lambda=\frac{1+\theta\xi}{1+\theta'\xi'},\ \mu=\frac{1+\xi}{1+\xi'}.$$

so that $\lambda,\mu\in[0,1]$, and let $Y'$ be as in the assertion to be proved, that is, the columns of $Y'$ are multiples of those of $Y$: $Y'_i=\lambda Y_i$ when $i\in P_+$ and $Y'_i=\mu Y_i$ otherwise. All we need to prove is that $(Y',v,\sigma,\rho)$ certify the validity of $\mathbf{VSG}^*_{s,\beta'}(\xi',\theta')$, and this immediately reduces to verification of the following fact:

**Lemma F.1** *Let $i$, $1\leq i\leq n$, be fixed, and let $z\in\mathbf{R}^n$ for any $I\subset\{1,...,n\}$ of cardinality $s$ satisfy the relations*

$$\begin{array}{ll}(a) & (1+\theta\xi)\sum\limits_{j\in P_+\cap I}\max[z_j-\delta_{ij},0]+(1+\xi)\sum\limits_{j\in P_n\cap I}|z_j-\delta_{ij}|+(Av)_i\leq\xi,\\ (b) & (1+\theta\xi)\sum\limits_{j\in P_+\cap I}\max[\delta_{ij}-z_j,0]+(1+\xi)\sum\limits_{j\in P_n\cap I}|z_j-\delta_{ij}|-(Av)_i\\ & \leq\eta=\left\{\begin{array}{ll}\theta\xi, & i\in P_+,\\ \xi, & i\in P_n,\end{array}\right.\end{array}\tag{47}$$

*where* $\delta_{ij}=\left\{\begin{array}{ll}0, & j\neq i,\\ 1, & i=j.\end{array}\right.$ *Then for every set $I\subset\{1,...,n\}$ of cardinality $s$ we have*

$$\begin{array}{ll}(a) & (1+\theta'\xi')\sum\limits_{j\in P_+\cap I}\max[\lambda z_j-\delta_{ij},0]+(1+\xi')\sum\limits_{j\in P_n\cap I}|\mu z_j-\delta_{ij}|+(Av)_i\leq\xi',\\ (b) & (1+\theta'\xi')\sum\limits_{j\in P_+\cap I}\max[\delta_{ij}-\lambda z_j,0]+(1+\xi')\sum\limits_{j\in P_n\cap I}|\mu z_j-\delta_{ij}|-(Av)_i\\ & \leq\eta_+=\left\{\begin{array}{ll}\theta'\xi', & i\in P_+,\\ \xi', & i\in P_n.\end{array}\right.\end{array}\tag{48}$$

**Proof.** Taking into account the definition of $\lambda,\mu$, in the case of $i\notin I$ the relations (48) are readily given by (47), hence we can assume $i\in I$. Consider two possible cases: $i\in P_+\cap I$ and $i\in P_n\cap I$.

**The case of $i\in P_+\cap I$.** In this case (47) reads:

$$\begin{array}{ll}(a) & (1+\theta\xi)\max[z_i-1,0]+(1+\theta\xi)\sum\limits_{j\in P_+\cap I,j\neq i}\max[z_j,0]\\ & +(1+\xi)\sum\limits_{j\in P_n\cap I}|z_j|+(Av)_i\leq\xi,\\ (b) & (1+\theta\xi)\max[1-z_i,0]+(1+\theta\xi)\sum\limits_{j\in P_+\cap I,j\neq i}\max[-z_j,0]\\ & +(1+\xi)\sum\limits_{j\in P_n\cap I}|z_j|-(Av)_i\leq\theta\xi,\end{array}\tag{49}$$

and our goal is to verify that then

$$\begin{array}{ll}(a) & (1+\theta'\xi')\max[\lambda z_i-1,0]\\ & +\overbrace{(1+\theta'\xi')\lambda}^{=1+\theta\xi}\sum\limits_{j\in P_+\cap I,j\neq i}\max[z_j,0]+\overbrace{(1+\xi')\mu}^{=1+\xi}\sum\limits_{j\in P_n\cap I}|z_j|+(Av)_i\leq \xi',\\ (b) & (1+\theta'\xi')\max[1-\lambda z_i,0]\\ & +\underbrace{(1+\theta\xi)\sum\limits_{j\in P_+\cap I,j\neq i}\max[-z_j,0]+(1+\xi)\sum\limits_{j\in P_n\cap I}|z_j|-(Av)_i}_{:=R}\leq\theta'\xi'.\end{array}\tag{50}$$

We have $\lambda z_i - 1 \leq \lambda(z_i-1)$ due to $\lambda\leq 1$, consequently

$$\max[\lambda z_i-1,0]\leq\max[\lambda(z_i-1),0]=\lambda\max[z_i-1,0],$$

and therefore (50.$a$) follows from (49.$a$) due to $(1+\theta'\xi')\lambda=1+\theta\xi$ and $\xi'\geq\xi$. It remains to verify (50.$b$). Assume, first, that $\lambda z_i\leq 1$. From (49.$b$) it follows that

$$(1+\theta\xi)[1-z_i]+R\leq(1+\theta\xi)\max[1-z_i,0]+R\leq\theta\xi,$$

implying $z_i\geq\frac{1+R}{1+\theta\xi}$ and therefore

$$1-\lambda z_i\leq 1-\frac{1+R}{1+\theta'\xi'}=\frac{\theta'\xi'-R}{1+\theta'\xi'}.$$

Since we are in the case $1-\lambda z_i\geq 0$, we arrive at

$$(1+\theta'\xi')\max[1-\lambda z_i,0]+R=(1+\theta'\xi')[1-\lambda z_i]+R\leq(1+\theta'\xi')\frac{\theta'\xi'-R}{1+\theta'\xi'}+R=\theta'\xi',$$

as required in (50.$b$). The case of $1-\lambda z_i\leq 0$ is trivial, since here the left hand side in (50.$b$) clearly is $\leq$ the left hand side in (49.$b$), while $\theta'\xi'\geq\theta\xi$, so that (50.$b$) is readily given by (49.$b$). Thus, when $i\in P_+\cap I$, (50) follows from (49).

**The case of $i\in P_n\cap I$.** In this case (47) means that

$$\begin{array}{ll}(a) & (1+\theta\xi)\sum\limits_{j\in P_+\cap I,j\neq i}\max[z_j,0]+(1+\xi)|1-z_i|+(1+\xi)\sum\limits_{j\in P_n\cap I,j\neq i}|z_j|+(Av)_i\leq\xi,\\ (b) & (1+\theta\xi)\sum\limits_{j\in P_+\cap I}\max[-z_j,0]+(1+\xi)|1-z_i|+(1+\xi)\sum\limits_{j\in P_n\cap I,j\neq i}|z_j|-(Av)_i\leq\xi,\end{array}\tag{51}$$

and our goal is to verify that then

$$\begin{array}{ll}(a) & (1+\theta'\xi')\sum\limits_{j\in P_+\cap I,j\neq i}\max[\lambda z_j,0]\\ & +(1+\xi')|1-\mu z_i|+(1+\xi')\mu\sum\limits_{j\in P_n\cap I,j\neq i}|z_j|+(Av)_i\leq\xi',\\ (b) & (1+\theta'\xi')\sum\limits_{j\in P_+\cap I}\max[-\lambda z_j,0]\\ & +(1+\xi')|1-\mu z_i|+(1+\xi')\sum\limits_{j\in P_n\cap I,j\neq i}|\mu z_j|-(Av)_i\leq\xi'.\end{array}\tag{52}$$

Comparing (51.$a$) with (52.$a$), and (51.$b$) with (52.$b$), we see that all we need in order to derive (52) from (51) is to verify the following statement: if $(1+\xi)|1-z| \leq \xi + a$, then $(1+\xi')|1-\mu z| \leq \xi' + a$. This is immediate: assuming $(1+\xi)|1-z| \leq \xi + a$, the premises in the following two implication chains hold true:

$$\begin{array}{l}(1+\xi)[1-z] \leq \xi + a \Rightarrow z \geq \frac{1-a}{1+\xi} \Rightarrow \mu z \geq \frac{1-a}{1+\xi'} \Rightarrow 1-\mu z \leq 1 - \frac{1-a}{1+\xi'} = \frac{\xi'+a}{1+\xi'}\\ \qquad\qquad \Rightarrow (1+\xi')[1-\mu z] \leq \xi' + a,\\ (1+\xi)[z-1] \leq \xi + a \Rightarrow z \leq 1 + \frac{\xi+a}{1+\xi} \Rightarrow \mu z \leq \frac{1+2\xi+a}{1+\xi'} \Rightarrow \mu z - 1 \leq \frac{2\xi-\xi'+a}{1+\xi'}\\ \qquad\qquad \Rightarrow (1+\xi')[\mu z - 1] \leq 2\xi - \xi' + a \Rightarrow (1+\xi')[\mu z - 1] \leq \xi' + a,\end{array}$$

while the resulting inequalities in these chains lead to the desired conclusion $(1+\xi')|1-\mu z| \leq \xi' + a$. □

## F.2 "Trigonometric polynomials" example

The validity of the claim concluding Section 5 is readily given by the following

**Lemma F.2** *For any positive integer $d$, let $n \geq 4\pi d$, and $A$ be the matrix obtained from the basic trigonometric polynomials as described in Section 5, then the condition $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$ can hold true for $s \leq 2$ only.*

**Proof.** Let $L$ be the $n \times n$ permutation matrix corresponding to the cyclic shift $e_j \mapsto e_{j_+}$, $j_+ = (j+1) \bmod n$, of the standard basic orths $e_0, ..., e_{n-1}$ in $\mathbf{R}^n$, and $R$ be the $m \times m$ orthogonal block-diagonal matrix with the North-Western block 1 and $d$ additional $2 \times 2$ diagonal blocks $\left[\begin{array}{cc}\cos(2\pi i/n) & -\sin(2\pi i/n)\\ \sin(2\pi i/n) & \cos(2\pi i/n)\end{array}\right]$, $1 \leq i \leq d$. Denoting by $A_j$ the $j$-th column of $A$, $0 \leq j \leq n-1$, we clearly have $RA_j = A_{j_+}$, hence $A = RAL^{-1}$ and therefore also $A = R^i A L^{-i}$ for $1 \leq i \leq n$. Now assume that $Y, v$ satisfy (13) for certain $\xi < 1$, $\theta \geq 1$, $\rho$, $\sigma$. Then

$$\max_i \left[\Phi_s(-C_i[Y,A]) + \Phi_s(C_i[Y,A])\right] \leq \xi(1+\theta),$$

in this way, it is immediately seen, $\max_i \|C_i[Y,A]\|_{s,1} \leq \kappa := \frac{\xi(1+\theta)}{1+\theta\xi} < 1$, or, which is the same,

$$\Gamma(I - Y^T A) \leq \kappa < 1,$$

where $\Gamma(Z)$ is the maximum of the $\|\cdot\|_{s,1}$-norms of columns of $Z \in \mathbf{R}^{n\times n}$. Observe that $\Gamma$ is a convex function which is symmetric in the sense that $\Gamma(PZP^T) = \Gamma(Z)$ whenever $P$ is a permutation matrix. Now let $\bar{Y} = \frac{1}{n}\sum_{i=1}^n R^{-i} Y L^i$. Since $L^n = I_n$, $R^{-n} = I_m$, we have $R^{-1}\bar{Y}L = \bar{Y}$. We claim that

$$\Gamma(I - \bar{Y}^T A) \leq \kappa.$$

Indeed, we have

$$
\begin{aligned}
\Gamma(I-\bar{Y}^TA) &= \Gamma(\frac{1}{n}\sum_{i=1}^n[I-L^{-i}Y^TR^iA])\\
&\leq \frac{1}{n}\sum_{i=1}^n\Gamma(I-L^{-i}Y^TR^iA) \quad \text{[since } \Gamma \text{ is convex]}\\
&= \frac{1}{n}\sum_{i=1}^n\Gamma(L^{-i}\left[I-Y^T[R^iAL^{-i}]\right]L^i)\\
&= \frac{1}{n}\sum_{i=1}^n\Gamma(I-Y^TA) \quad \text{[since } \Gamma \text{ is symmetric and } R^iAL^{-i}=A]\\
&= \Gamma(I-Y^TA)
\end{aligned}
$$

Now let

$$
y_j(\phi)=\bar{Y}_{0j}+\sum_{i=1}^d[\bar{Y}_{2i-1,j}\cos(i\phi)+\bar{Y}_{2i,j}\sin(i\phi)].
$$

We have $R^{-1}\bar{Y}L=\bar{Y}$, that is, $R^{-1}\bar{Y}=\bar{Y}L^{-1}$. In other words, the columns $\bar{Y}_j$ of $\bar{Y}$ satisfy the relation $\bar{Y}_j=R\bar{Y}_{j_-}$, where $j_-=(j-1)\bmod n$. This is nothing but $y_j(\phi)\equiv y_{j_-}(\phi-\delta)$, $\delta=2\pi/n$, whence $y_j(\phi)=y_0(\phi-j\delta)$. Observe that the $j$-th column in $\bar{Y}^TA$ has the entries

$$
\bar{Y}_i^TA_j=y_i(j\delta)=y_0((j-i)\delta),\ \ 0\leq i\leq n-1,
$$

meaning that the columns in the matrix $I-\bar{Y}^TA$ are cyclic shifts of each other (so that the $\|\cdot\|_{s,1}$-norms of all columns are the same), and the zero column is comprised of the values of the trigonometric polynomial $1-y_0(\phi)$ on the grid $G=\{\phi_j=\frac{2\pi j}{n}:0\leq j<n\}$. Assuming $s>1$, when denoting by $\gamma$ the sum of $s-1$ largest magnitudes of entries in the $(n-1)$-dimensional vector $\{y_0(\phi_i)\}_{i=1}^{n-1}$, we have

$$
1-y_0(0)+\gamma\leq\|C_0[\bar{Y},A]\|_{s,1}\leq\kappa<1,
$$

thereby $\mu:=y_0(0)>\gamma$. Now let $M=\max\limits_{0\leq\phi\leq2\pi}|y_0(\phi)|$, and let $\bar{\phi}\in\mathrm{Argmax}_\phi|y_0(\phi)|$, so that $y_0'(\bar{\phi})=0$. By Bernstein theorem, we have $|y_0''(\phi)|\leq d^2M$ for all $\phi$, whence $|y_0(\phi)|\geq M/2$ when $|\phi-\bar{\phi}|\leq 1/d$, so that

$$
\mathrm{Card}\{j:|y_0(\phi_j)|\geq M/2\}>\frac{n}{\pi d}-1.
$$

It follows that $\gamma\geq\min\left[s-1,\frac{n}{\pi d}-2\right]M/2$, while $\mu=y_0(0)\leq M$. Thus, the relation $\mu>\gamma$ implies that

$$
\min[s-1,\frac{n}{\pi d}-2]<2,
$$

that is, $s\leq 2$ provided that $n\geq 4\pi d$. $\square$

### F.3 Proof of Proposition 6.1

We will prove the following more general result which implies Proposition 6.1. Consider the problem of bounding from above the quantity

$$
\mathrm{Opt}=\max_{x,u}\left\{x^T[Pu+p]:x\in X,Ax=0,u\in U\right\},\quad X=\mathrm{Conv}\{x^1,...,x^N\},\tag{53}
$$

where $x^i \in \mathbf{R}^n$, the set $\{x \in X : Ax = 0\}$ is nonempty, and $U \subset \mathbf{R}^n$ is a computationally tractable compact convex set which contains the origin in its interior.

In this setting the linear programming based relaxation scheme corresponds to

$$\mathrm{Opt}^+ = \inf_{Y,v} \max_{1\le i\le N} \left[ \max_{u\in U} [(I - Y^T A)x^i]^T [Pu + p] + v^T A x^i \right], \tag{54}$$

and in section 4, while demonstrating the origin of $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$, linear programming based verifiable sufficient condition, we have shown that $\mathrm{Opt}^+ \ge \mathrm{Opt}$. Since the only role of $\rho$ and $\sigma$ in $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$ is to get a control over $\beta$, and there is no such control in the SDP based condition (25), we will not include them in the analysis here.

Let $\phi(u)$ be the Minkowski function of $U$, that is, a positively homogeneous, of order 1, function on $\mathbf{R}^n$ such that $U = \{u : \phi(u) \le 1\}$, let $\mathcal{U}$ be the cone $\{(u,t) : \phi(u) \le t\}$. Note that $\mathcal{U}$ is a closed pointed convex cone with a nonempty interior, and its dual cone is

$$\mathcal{U}_* = \{(\omega, \gamma) : \phi_*(-\omega) \le \gamma\}, \ \phi_*(\omega) = \max\left\{\omega^T u : u \in U\right\}.$$

Now,

$$\begin{aligned}
\mathrm{Opt}^+ &= \inf_{Y,v} \max_{1\le i\le N} \left[ \max_{u\in U} [(I - Y^T A)x^i]^T [Pu + p] + v^T A x^i \right] \\
&= \inf_{Y,v,\tau} \left\{ \tau : \ \max_{u\in U} u^T P^T [I - Y^T A] x^i + [p^T (I - Y^T A) + v^T A] x^i \le \tau, \ \ \text{for } 1 \le i \le N \right\} \\
&= \inf_{Y,v,\tau} \left\{ \tau : \ \phi_*(-P^T [I - Y^T A] x^i) + [p^T (I - Y^T A) + v^T A] x^i \le \tau, \ \ \text{for } 1 \le i \le N \right\} \\
&= \inf_{Y,v,\tau,\gamma} \left\{ \tau : \ \begin{array}{l} \phi_*(-P^T [I - Y^T A] x^i) \le \gamma_i, \ \ \text{for } 1 \le i \le N \\ \gamma_i + [p^T (I - Y^T A) + v^T A] x^i \le \tau, \ \ \text{for } 1 \le i \le N \end{array} \right\}
\end{aligned}$$

Since the constraints $\phi_*(-P^T [I - Y^T A] x^i) \le \gamma_i$ imply exactly that $(P^T [I - Y^T A] x^i, \gamma_i) \in \mathcal{U}_*$, $\mathrm{Opt}^+$ is the optimal value of a conic minimization problem. Moreover it is immediately seen that this problem is strictly feasible and bounded, so that the dual problem is solvable with the optimal value $\mathrm{Opt}^+$, which amounts to

$$\begin{aligned}
\mathrm{Opt}^+ &= \max_{w^i, t_i} \left\{ \sum_i \mathrm{Tr}(x^i [P w^i + t_i p]^T) : \ \begin{array}{l} A \sum_{i=1}^N t_i x^i = 0 \\ A \left[ \sum_{i=1}^N x^i [P w^i + t_i p] \right] = 0 \\ \phi(w^i) \le t_i, \ \ \text{for } 1 \le i \le N \\ \sum_{i=1}^N t_i = 1 \end{array} \right\} \\
&= \max_{V \in \mathcal{V}} \{\mathrm{Tr}(V)\},
\end{aligned}$$

where

$$\mathcal{V} = \left\{ V = \sum_{i=1}^N x^i [P w^i + t_i p]^T : \ \ \phi(w^i) \le t_i, \ \sum_{i=1}^N t_i = 1, \ A \sum_{i=1}^N t_i x^i = 0, \ AV = 0 \right\}.$$

Note that $\mathcal{V}$ is a computationally tractable convex compact set. Moreover the set $\mathcal{V}$ admits a simple interpretation. Specifically, setting

$$\mathcal{W} = \mathrm{Conv}\left\{ [x, u, x[Pu + p]^T] : \ x \in X, \ u \in U \right\},$$

we have

$$\mathcal{V} = \{V : \exists \bar{x}, \bar{u} : [\bar{x}, \bar{u}, V] \in \mathcal{W},\ AV = 0,\ A\bar{x} = 0\}.$$

Indeed, if $V \in \mathcal{V}$, that is, $V = \sum_{i=1}^N x^i[Pw^i + t_i p]^T$ with $\phi(w^i) \le t_i$, $\sum_{i=1}^N t_i = 1$ and $AV = 0$, $A\sum_{i=1}^N t_i x^i = 0$, then $w^i = t_i u^i$ with $u^i \in U$, so that, setting $\bar{x} = \sum_{i=1}^N t_i x^i$ and $\bar{u} = \sum_{i=1}^N t_i u^i$, we have

$$[\bar{x}, \bar{u}, V] = \left[\sum_{i=1}^N t_i x^i, \sum_{i=1}^N t_i u^i, \sum_{i=1}^N t_i x^i [Pu^i + p]^T\right] \in \mathcal{W}.$$

Vice versa, if $[\bar{x}, \bar{u}, V] \in \mathcal{W}$ and $A\bar{x} = 0, AV = 0$, then $[\bar{x}, \bar{u}, V] = \sum_{k=1}^K \lambda_k [\hat{x}^k, \hat{u}^k, \hat{x}^k[P\hat{u}^k + p]^T]$ with $\hat{u}^k \in U$, $\hat{x}^k \in X$ and nonnegative $\lambda_k$ summing up to 1. Representing $\hat{x}^k = \sum_{i=1}^N \mu_{ki} x^i$ with nonnegative $\mu_{ki}$, $\sum_{i=1}^N \mu_{ki} = 1$, we have

$$\begin{aligned}&[\bar{x}, V] = \textstyle\sum_{k=1}^K \lambda_k [\hat{x}^k, \hat{x}^k[P\hat{u}^k + p]^T] = \sum_{k=1}^K \sum_{i=1}^N \lambda_k \mu_{ki} [x^i, x^i[P\hat{u}^k + p]^T] \\ &= \left[\textstyle\sum_{i=1}^N t_i x^i, \sum_{i=1}^N x^i [Pw^i + t_i p]^T\right],\end{aligned}$$

where $w^i = \sum_{k=1}^K \lambda_k \mu_{ki} \hat{u}^k$, $t_i = \sum_{k=1}^K \lambda_k \mu_{ki}$. Clearly $\sum_{i=1}^N t_i = 1$ and since $\phi(\hat{u}^k) \le 1$ and $\phi(\cdot)$ is a convex function, we have $\phi(w^i) \le t_i$. Thus, $V = \sum_{i=1}^N x^i[Pw^i + t_i p]^T$ with $\phi(w^i) \le t_i$ and $t_i$ summing up to 1 and such that $A\sum_{i=1}^N t_i x^i = 0$, that is, $V \in \mathcal{V}$.

We will next examine the connection with SDP based condition. Let $\mathcal{X} = \{x \in X : Ax = 0\}$. Given $x \in \mathcal{X}$ and $u \in U$, consider the positive semidefinite matrix

$$\Delta(x, u) := [1; x; [Pu + p]][1; x; [Pu + p]]^T = \left[\begin{array}{c|c|c} 1 & x^T & [Pu + p]^T \\ \hline x & xx^T & x^T[Pu + p]^T \\ \hline [Pu + p] & [Pu + p]x^T & [Pu + p][Pu + p]^T \end{array}\right]$$

The convex hull $\mathcal{K}_*$ of these matrices is contained in every set of the form

$$\mathcal{K} = \left\{\Delta = \left[\begin{array}{c|c|c} 1 & x^T & [Pu + p]^T \\ \hline x & Z & V \\ \hline [Pu + p] & V^T & Q \end{array}\right] : \Delta \succeq 0, AZ = 0, [x, u, V] \in \mathcal{W}, (*)\right\},$$

where $(*)$ is a set of efficiently computable convex constraints on $\Delta$ which are valid for matrices $\Delta(x, u)$ given by $x \in \mathcal{X}$, $u \in U$. When $\Delta \succeq 0$ and $AZ = 0$, we automatically have $Ax = 0, AV = 0$, that is,

$$\mathcal{K}_* \subset \left\{\Delta = \left[\begin{array}{c|c|c} 1 & x^T & [Pu + p]^T \\ \hline x & Z & V \\ \hline [Pu + p] & V^T & Q \end{array}\right] : \Delta \succeq 0, AZ = 0, V \in \mathcal{V}, (*)\right\}.$$

Let us denote

$$\mathrm{Opt}^* := \max_{Z,Q,V,x,u} \left\{\mathrm{Tr}(V) : \Delta = \left[\begin{array}{c|c|c} 1 & x^T & [Pu + p]^T \\ \hline x & Z & V \\ \hline [Pu + p] & V^T & Q \end{array}\right] \succeq 0, AZ = 0, V \in \mathcal{V}, (*)\right\},$$

Then $\mathrm{Opt}^*$ is efficiently computable and it follows that $\mathrm{Opt} \le \mathrm{Opt}^* \le \mathrm{Opt}^+$.

In our particular case, for the derivation of verifiable sufficient conditions, $p = 0$ and $P = C^{\xi,\theta}$ which is defined in (19). The extreme points of $X$ are the $2n$ vectors $-e_i$ for $1 \le i \le n$, $e_i$ for

$i \in P_n$, and $\theta^{-1}e_i$ for $i \in P_+$, where $e_i$ is the $i$-th basic orth. Moreover $U = \mathcal{U}$ as defined in (18). Our LP based verifiable sufficient condition $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$ with $\rho = \sigma = \infty$ is exactly $\mathrm{Opt}^+ \leq \xi$. In addition to this, our SDP bound given in (25) is at least as good as $\mathrm{Opt}^*$ without any inequalities included in $(*)$. Note that in our case $\Delta = HGH^T$ where $H = \left[\begin{array}{c|c} I_{n+1} & \\ \hline & C^{\xi,\theta} \end{array}\right]$, under this connection it is clear that the objective functions in two SDPs are the same. Furthermore, (24.$a$), i.e. $G \succeq 0$, holds if and only if $\Delta \succeq 0$, with the same transformation, (24.$b$) and (24.$c$) correspond to constraints in $(*)$, (24.$d$) together with (24.$f$) characterize the set $\mathcal{W}$ and (24.$e$) is equivalent to $AZ = 0$ in $\mathrm{Opt}^*$ .

Hence $\mathrm{Opt}^{\xi,\theta} \leq \mathrm{Opt}^* \leq \mathrm{Opt}^+ \leq \xi$ where the last inequality holds whenever $\mathbf{VSG}_s(\xi, \theta, \rho, \sigma)$ with $\rho = \sigma = \infty$ holds. □