Available online at www.sciencedirect.com
ScienceDirect
Journal of Approximation Theory 163 (2011) 388–412
JOURNAL OF Approximation Theory
www.elsevier.com/locate/jat

Sparse representations and approximation theory

Allan Pinkus

Department of Mathematics, Technion, 32000 Haifa, Israel

Received 28 March 2010; accepted 28 October 2010
Available online 5 November 2010

Communicated by Paul Nevai

Abstract

This paper is an attempt to both expound and expand upon, from an approximation theorist's point of view, some of the theoretical results that have been obtained in the sparse representation (compressed sensing) literature. In particular, we consider in detail ℓ_1^m-approximation, which is fundamental in the theory of sparse representations, and the connection between the theory of sparse representations and certain n-width concepts. We try to illustrate how the theory of sparse representation leads to new and interesting problems in approximation theory, while the results and techniques of approximation theory can further add to the theory of sparse representations.

Keywords: Sparse representations; ℓ_1^m-approximation

1. Introduction

Recent years have witnessed an abundance of research devoted to the subject of sparse and redundant representations; see Bruckstein, Donoho, Elad [1], Elad [7]. The fundamental problem therein is the following. Assume that we are given an $n \times m$ matrix A of rank n over $\mathbb{R}$ (or $\mathbb{C}$). Given any vector $\mathbf{b} \in \mathbb{R}^n$ (or $\mathbb{C}^n$), the solution set to $A\mathbf{x} = \mathbf{b}$ is an affine space of dimension $m - n$; i.e., we have for $m > n$ an underdetermined system of linear equations. The aim is then to characterize and identify a solution with minimal support (the sparsest representation). A successful solution to this problem has applications in signal and image processing.

The main idea behind characterizing sparse representations is to look for a solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ of minimal ℓ_1^m-norm. Finding a minimal ℓ_1^m solution is equivalent to a linear programming

E-mail address: pinkus@tx.technion.ac.il.

0021-9045/$ - see front matter
doi:10.1016/j.jat.2010.10.007

problem, and thus is considered, to some extent, as tractable. On the other hand, searching for the solution $\mathbf{x}$ to $A\mathbf{x} = \mathbf{b}$ of minimal support is not, in general, computationally efficient. The theory then asks for conditions on when the minimal ℓ_1^m solution is also necessarily a solution of minimal support. Minimal norm solutions are generally not solutions of minimal support. However, the ℓ_1^m-norm can be, and often is, the exception to this rule.

In this paper, we consider this problem from an approximation theory perspective. For any vector $\mathbf{b}$, the solution set $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}\}$ is given by $\mathbf{x} + Y$, where $\mathbf{x}$ is any fixed element thereof and Y is the right kernel of A, and hence is an $(m - n)$-dimensional subspace, independent of $\mathbf{b}$. Thus, finding a minimal norm solution to $A\mathbf{x} = \mathbf{b}$ is equivalent to looking for $\mathbf{x} - \mathbf{y}^*$, where $\mathbf{y}^*$ is a best approximation to $\mathbf{x}$ from Y. In Section 2, we ask for conditions on when, if $\mathbf{x}$ has its support $\{i : x_i \neq 0\}$ in some fixed index set I, its best ℓ_1^m-approximation from Y is necessarily the $\mathbf{0}$ vector. That is, what are conditions on A and I such that if the support of $\mathbf{x}$ lies in I then $\mathbf{x}$ is a minimal ℓ_1^m-norm solution to $A\mathbf{x} = \mathbf{b}$. In Section 3, we consider this same problem when $\mathbf{x}$ is any vector whose support is of a certain size (cardinality), i.e., when $\mathbf{x}$ is sufficiently sparse. This is a central problem considered in the sparse representation literature. We take a somewhat different approach from that found in the literature by introducing an ancillary norm on $\mathbb{R}^m$ and considering its dual norm and approximation therein. We show, in what we believe is a rather intuitive manner, some of the central results of the theory for matrices with given mutual incoherence. In Section 4, we look, in somewhat more detail, at mutual incoherence and what it implies. Finally, in Section 5, we look at the best possible estimates available on the size of the support of the vector $\mathbf{x}$ guaranteeing that it is a minimal ℓ_1^m-norm solution to $A\mathbf{x} = \mathbf{b}$, as we vary over all $n \times m$ matrices. That is, assuming we want to maximize the possible size of the support while maintaining for every vector $\mathbf{x}$, with support of that size, the minimal ℓ_1^m-norm property, what is the best we can hope for and how can we construct the associated $n \times m$ matrix A? This leads us to a consideration of n-widths, and we review known results and their relevance. We state all results in this paper for $\mathbb{R}$ rather than $\mathbb{C}$. Many of these results are also valid over $\mathbb{C}$.

This paper is an attempt to both expound and expand upon, from an approximation theorist's point of view, some of the theoretical results that have been obtained in the sparse representation literature; see also DeVore [2, Section 5], for a somewhat different approach. We do not consider many of the other aspects of the theory, neither the more practical and computational parts of this theory nor the probabilistic approach to these problems. It is gratifying to see approximation theory applied to such problems. It is also a pleasure to acknowledge the insights and the new results in approximation theory that have resulted from this theory. This paper is written for the approximation theorist. Hopefully, however, the ideas, results, and techniques of approximation theory can further add to the theory of sparse representations.

2. Fixed minimal support

Set

$$\|\mathbf{x}\|_1 := \sum_{i=1}^{m} |x_i|,$$

and let Y be any finite-dimensional subspace of ℓ_1^m in $\mathbb{R}^m$. To each $\mathbf{x} \in \mathbb{R}^m$ there exists a best approximation to $\mathbf{x}$ from Y. We start by summarizing well-known results concerning characterization, uniqueness, and strong uniqueness of the best ℓ_1^m-approximations from Y to elements of $\mathbb{R}^m$.

Theorem 2.1. *Let Y be any finite-dimensional subspace of ℓ_1^m in $\mathbb{R}^m$. The vector $\mathbf{y}^* \in Y$ is a best approximation to $\mathbf{x}$ from Y if and only if*

$$\left| \sum_{\{i: x_i - y_i^* \neq 0\}} [\operatorname{sgn}(x_i - y_i^*)] y_i \right| \leq \sum_{\{i: x_i - y_i^* = 0\}} |y_i|$$

for all $\mathbf{y} \in Y$. The vector $\mathbf{y}^* \in Y$ is the unique best approximation to $\mathbf{x}$ from Y if and only if

$$\left| \sum_{\{i: x_i - y_i^* \neq 0\}} [\operatorname{sgn}(x_i - y_i^*)] y_i \right| < \sum_{\{i: x_i - y_i^* = 0\}} |y_i|$$

for all $\mathbf{y} \in Y \setminus \{\mathbf{0}\}$. Assuming that $\mathbf{y}^* \in Y$ is the unique best approximation to $\mathbf{x}$ from Y, then we also always have strong uniqueness, namely

$$\|\mathbf{x} - \mathbf{y}\|_1 - \|\mathbf{x} - \mathbf{y}^*\|_1 \geq \gamma \|\mathbf{y} - \mathbf{y}^*\|_1$$

for all $\mathbf{y} \in Y$, where the optimal $\gamma > 0$ in this inequality is given by

$$\gamma := \min_{\mathbf{y} \in Y, \|\mathbf{y}\|_1 = 1} \left(\sum_{\{i: x_i - y_i^* = 0\}} |y_i| - \sum_{\{i: x_i - y_i^* \neq 0\}} [\operatorname{sgn}(x_i - y_i^*)] y_i \right).$$

This result may be found in Pinkus [28], see Chapters 2 and 6. See also the more exact references therein, especially James [16] and Kripke, Rivlin [18].

Consider now the equation

$$A\mathbf{x} = \mathbf{b},$$

where A is an $n \times m$ matrix of rank n. For each $\mathbf{b}$, the solution set is an affine subspace (flat) of dimension $m - n$. That is, setting

$$Y := \{\mathbf{y} : A\mathbf{y} = \mathbf{0}\},$$

the right kernel of A, then Y is a subspace of dimension $m - n$, and for any $\mathbf{x} \in \mathbb{R}^m$ satisfying $A\mathbf{x} = \mathbf{b}$ we have $A(\mathbf{x} - \mathbf{y}) = \mathbf{b}$ if and only if $\mathbf{y} \in Y$.

Let $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$ denote an index set, I^c the complimentary set to I in $\{1, \ldots, m\}$, and $|I| = k$ the cardinality (size) of I. For each $\mathbf{x} = (x_1, \ldots, x_m)$, let

$$\operatorname{supp} \mathbf{x} := \{i : x_i \neq 0\}.$$

The first question we ask is as follows. What are conditions on I implying that if $\operatorname{supp} \mathbf{x} \subseteq I$ then necessarily

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z} = A\mathbf{x}} \|\mathbf{z}\|_1?$$

This is not the central problem considered within the context of sparse representations. However, it is of interest both in and of itself and because the results herein will also be used in Section 3.

Proposition 2.2. *Let A, Y, and I be as above. Then we have the equivalence of the following. For all $\mathbf{x} \in \mathbb{R}^m$ satisfying $\operatorname{supp} \mathbf{x} \subseteq I$,*

(a) $\|\mathbf{x}\|_1 = \min_{A\mathbf{z} = A\mathbf{x}} \|\mathbf{z}\|_1$.
(b) $\min_{\mathbf{y} \in Y} \|\mathbf{x} - \mathbf{y}\|_1 = \|\mathbf{x}\|_1$.
(c) $\sum_{i \in I} |y_i| \leq \sum_{i \in I^c} |y_i|$ *for all* $\mathbf{y} \in Y$.

Proof. (a) ⇔ (b). The set $\{\mathbf{x} - \mathbf{y} : \mathbf{y} \in Y\}$ represents all $\mathbf{z}$ satisfying $A\mathbf{z} = \mathbf{b}$.

(b) ⇔ (c). From Theorem 2.1, the vector $\mathbf{0}$ is a best approximation to $\mathbf{x}$ from Y if and only if

$$\left|\sum_{\{i: x_i \neq 0\}} [\operatorname{sgn} x_i] y_i\right| \leq \sum_{\{i: x_i = 0\}} |y_i|$$

for all $\mathbf{y} \in Y$. If (c) holds, then for each $\mathbf{x} \in \mathbb{R}^m$ with $\operatorname{supp} \mathbf{x} \subseteq I$ we have, for all $\mathbf{y} \in Y$,

$$\left|\sum_{\{i: x_i \neq 0\}} [\operatorname{sgn} x_i] y_i\right| \leq \sum_{\{i: x_i \neq 0\}} |y_i| \leq \sum_{i \in I} |y_i| \leq \sum_{i \in I^c} |y_i| \leq \sum_{\{i: x_i = 0\}} |y_i|,$$

and thus (b) holds. Now assume that (b) holds for all $\mathbf{x} \in \mathbb{R}^m$ with $\operatorname{supp} \mathbf{x} \subseteq I$. Then, for all such $\mathbf{x}$,

$$\left|\sum_{\{i: x_i \neq 0\}} [\operatorname{sgn} x_i] y_i\right| \leq \sum_{\{i: x_i = 0\}} |y_i|$$

for all $\mathbf{y} \in Y$. Given $\mathbf{y} \in Y$, choose any $\mathbf{x}$ such that $\operatorname{sgn} x_i = y_i$ for $i \in I$, and $x_i = 0$ for $i \in I^c$. Thus we obtain (c). □

There is one additional characterization of when every $\mathbf{x} \in \mathbb{R}^m$ with $\operatorname{supp} \mathbf{x} \subseteq I$ satisfies Proposition 2.2; see also Fuchs [11, Theorem 4], for an approach using duality in linear programming. In what follows, we let $\mathbf{c}^\ell$, $\ell = 1, \ldots, m$, denote the ℓth column vector of A.

Proposition 2.3. *Let A, Y, and I be as above. Then, for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $\operatorname{supp} \mathbf{x} \subseteq I$, we have*

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z} = A\mathbf{x}} \|\mathbf{z}\|_1$$

if and only if, for every given $\varepsilon \in \{-1, 1\}^I$, there exists a $\mathbf{d}^\varepsilon \in \mathbb{R}^n$ satisfying $\langle \mathbf{d}^\varepsilon, \mathbf{c}^i \rangle = \varepsilon_i$, $i \in I$, and $|\langle \mathbf{d}^\varepsilon, \mathbf{c}^\ell \rangle| \leq 1$, $\ell = 1, \ldots, m$.

Proof. (⇐). Assume that for every given $\varepsilon \in \{-1, 1\}^I$ there exists a $\mathbf{d}^\varepsilon \in \mathbb{R}^n$ satisfying $\langle \mathbf{d}^\varepsilon, \mathbf{c}^i \rangle = \varepsilon_i$, $i \in I$, and $|\langle \mathbf{d}^\varepsilon, \mathbf{c}^\ell \rangle| \leq 1$, $\ell = 1, \ldots, m$. Given $\mathbf{y} \in Y$ and $i \in I$, let $\varepsilon_i = \operatorname{sgn} y_i$ if $y_i \neq 0$ and $\varepsilon_i \in \{-1, 1\}$ be arbitrarily chosen if $y_i = 0$. Let $\mathbf{d}^\varepsilon \in \mathbb{R}^n$ be as guaranteed above. Since $\mathbf{y} \in Y$, we have

$$\sum_{\ell=1}^{m} y_\ell \langle \mathbf{d}^\varepsilon, \mathbf{c}^\ell \rangle = \langle \mathbf{d}^\varepsilon, A\mathbf{y} \rangle = 0.$$

Thus,

$$\begin{aligned}\sum_{i \in I} |y_i| &= \sum_{i \in I} y_i [\operatorname{sgn} y_i] = \sum_{i \in I} y_i \langle \mathbf{d}^\varepsilon, \mathbf{c}^i \rangle \\ &= -\sum_{i \in I^c} y_i \langle \mathbf{d}^\varepsilon, \mathbf{c}^i \rangle \leq \sum_{i \in I^c} |y_i| |\langle \mathbf{d}^\varepsilon, \mathbf{c}^i \rangle| \leq \sum_{i \in I^c} |y_i|,\end{aligned}$$

and therefore (c) of Proposition 2.2 holds.

(⇒). Assume that for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $\operatorname{supp} \mathbf{x} \subseteq I$ we have

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z} = A\mathbf{x}} \|\mathbf{z}\|_1.$$

From Proposition 2.2 this implies that

$$\min_{\mathbf{y}\in Y} \|\mathbf{x}-\mathbf{y}\|_1 = \|\mathbf{x}\|_1.$$

We recall from duality that

$$\min_{\mathbf{y}\in Y} \|\mathbf{x}-\mathbf{y}\|_1 = \max_{\mathbf{g}\in Y^\perp} \frac{|\langle \mathbf{x}, \mathbf{g}\rangle|}{\|\mathbf{g}\|_\infty},$$

where $Y^\perp$ is the orthogonal complement to Y. Now, since $Y = \{\mathbf{y} : A\mathbf{y} = \mathbf{0}\}$, it follows that $Y^\perp$ is the n-dimensional space given by the span of the rows of the matrix A. Thus $\mathbf{g} \in Y^\perp$ if and only if $\mathbf{g} = (\langle \mathbf{d}, \mathbf{c}^1\rangle, \ldots, \langle \mathbf{d}, \mathbf{c}^m\rangle)$ for some $\mathbf{d} \in \mathbb{R}^n$. For $\mathbf{x}$ with supp $\mathbf{x} \subseteq I$, we therefore have

$$\|\mathbf{x}\|_1 = \min_{\mathbf{y}\in Y} \|\mathbf{x}-\mathbf{y}\|_1 = \max_{\mathbf{d}} \frac{\left|\sum_{i=1}^m x_i \langle \mathbf{d}, \mathbf{c}^i\rangle\right|}{\max_{j=1,\ldots,m} |\langle \mathbf{d}, \mathbf{c}^j\rangle|}.$$

Given $\mathbf{x}$, with supp $\mathbf{x} \subseteq I$, there therefore exists a $\mathbf{d} \in \mathbb{R}^n$ for which

$$\max_{j=1,\ldots,m} |\langle \mathbf{d}, \mathbf{c}^j\rangle| = 1$$

and

$$\left|\sum_{i=1}^m x_i \langle \mathbf{d}, \mathbf{c}^i\rangle\right| = \sum_{i=1}^m |x_i|;$$

i.e.,

$$\left|\sum_{j=1}^k x_{i_j} \langle \mathbf{d}, \mathbf{c}^{i_j}\rangle\right| = \sum_{j=1}^k |x_{i_j}|.$$

This implies that

$$\langle \mathbf{d}, \mathbf{c}^{i_j}\rangle = \operatorname{sgn} x_{i_j}$$

if $x_{i_j} \neq 0$. Given ε, as in the statement of the proposition, take any $\mathbf{x}$ with supp $\mathbf{x} \subseteq I$ satisfying $\operatorname{sgn} x_{i_j} = \varepsilon_{i_j}$. Thus there exists a $\mathbf{d} \in \mathbb{R}^n$ satisfying the desired conditions. □

The conditions of Proposition 2.3 are difficult to verify. There is, however, a somewhat simpler sufficient condition given by this next result. In the following, $\mathbf{e}^j$ denotes the jth coordinate direction in $\mathbb{R}^k$.

Proposition 2.4. *Let A and $I = \{i_1, \ldots, i_k\}$ be as above. Assume that there exists a $k \times n$ matrix D satisfying*

$$D\mathbf{c}^{i_j} = \mathbf{e}^j, \quad j = 1, \ldots, k,$$

and

$$\|D\mathbf{c}^\ell\|_1 \leq 1, \quad \ell = 1, \ldots, m.$$

Then the conditions of Proposition 2.3 hold.

Proof. For any given ε, as in Proposition 2.3, set

$$\mathbf{d}^{\varepsilon} = \sum_{j=1}^{k} \varepsilon_{i_j} \mathbf{d}^j,$$

where the $\mathbf{d}^j$, $j = 1, \ldots, k$, are the rows of D. Then the result holds trivially, since

$$|\langle \mathbf{d}^{\varepsilon}, \mathbf{c}^{\ell} \rangle| \le \sum_{j=1}^{k} |\langle \mathbf{d}^j, \mathbf{c}^{\ell} \rangle| = \|D\mathbf{c}^{\ell}\|_1 \le 1. \quad \square$$

Remark. A particular case of Proposition 2.4 may be found in Tropp [31, Theorem 3.3]. (Note that, if $\|D\mathbf{c}^{\ell}\|_1 < 1$, $\ell \notin \{i_1, \ldots, i_k\}$, then, since $\|D\mathbf{c}^{\ell}\|_2 < 1$, it follows, as in the terminology of Tropp [31], that both BP and OMP converge.)

The condition in Proposition 2.4 is probably not equivalent to that in Proposition 2.3. However, it is equivalent in certain cases.

Proposition 2.5. *Let A, Y, and I be as above. For $k = 1$, $k = 2$, and $k = n$, the conditions of Propositions 2.3 and 2.4 are equivalent.*

Proof. If $k = 1$, then the result is immediate.

Let us consider the case $k = n$. Assume, without loss of generality, that $I = \{1, \ldots, n\}$. Let $A = [B, C]$, where B is an $n \times n$ matrix and C an $n \times (m-n)$ matrix. From (c) of Proposition 2.2 it easily follows that the first n columns of A, i.e., the columns of B, must be linearly independent. Set $D = B^{-1}$. Let $\mathbf{d}^i$ denote the ith row of $D = B^{-1}$. Since the $\mathbf{d}^{\varepsilon}$ of Proposition 2.3 satisfy

$$\langle \mathbf{d}^{\varepsilon}, \mathbf{c}^j \rangle = \varepsilon_j, \quad j = 1, \ldots, n,$$

it follows that the $\mathbf{d}^{\varepsilon}$ must have the explicit form

$$\mathbf{d}^{\varepsilon} = \sum_{i=1}^{n} \varepsilon_i \mathbf{d}^i.$$

In addition, from Proposition 2.3, we have

$$1 \ge |\langle \mathbf{d}^{\varepsilon}, \mathbf{c}^j \rangle| = \left| \left\langle \sum_{i=1}^{n} \varepsilon_i \mathbf{d}^i, \mathbf{c}^j \right\rangle \right|$$

for each $j = n+1, \ldots, m$, and for every choice of $\varepsilon_i \in \{-1, 1\}$. Choosing $\varepsilon_i = \operatorname{sgn} \langle \mathbf{d}^i, \mathbf{c}^j \rangle$, we obtain

$$\sum_{i=1}^{n} |\langle \mathbf{d}^i, \mathbf{c}^j \rangle| \le 1.$$

This proves the result for $k = n$.

If it can be shown that the vectors $\mathbf{d}^{\varepsilon}$ of Proposition 2.3, as ε varies over all possible choices as in Proposition 2.3, span a k-dimensional space, then the conditions in Propositions 2.3 and 2.4 are equivalent. This follows easily using the above method of proof. When $k = 2$, this is valid as it suffices to consider only the two vectors $\mathbf{d}^{\varepsilon}$ associated with $\varepsilon = (1, 1)$ and $\varepsilon = (1, -1)$. $\square$

As a consequence of Theorem 2.1 and the above analysis, we also have this next result. Note that strong uniqueness is said to provide for numerical stability.

Proposition 2.6. *Each* $\mathbf{x} \in \mathbb{R}^m$ *satisfying* $\operatorname{supp} \mathbf{x} \subseteq I$ *is the unique solution to*

$$\min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1 = \|\mathbf{x}\|_1$$

if and only if

$$\sum_{i\in I} |y_i| < \sum_{i\in I^c} |y_i|$$

for all $\mathbf{y} \in Y \setminus \{\mathbf{0}\}$. *Furthermore, in this case, there exists a* $\gamma^* > 0$ *such that, for all* $\mathbf{z}$ *satisfying* $A\mathbf{z} = A\mathbf{x}$, *we have*

$$\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 \geq \gamma^* \|\mathbf{z} - \mathbf{x}\|_1.$$

The largest γ^*, *dependent on* A *and* I, *but independent of* $\mathbf{x}$, *is given by*

$$\gamma^* := \min_{\mathbf{y}\in Y, \|\mathbf{y}\|_1=1} \left(\sum_{i\in I^c} |y_i| - \sum_{i\in I} |y_i| \right).$$

How large can $|I|$ be, assuming that it satisfies Proposition 2.2? In fact, it can be n. As Y is of dimension $m-n$, this is the largest value of $|I|$ for which (c) of Proposition 2.2 can possibly hold.

Proposition 2.7. *Given any* $m > n$, *there exists an* $n \times m$ *matrix* A *with every* n *columns of* A *linearly independent such that* $I = \{1, \ldots, n\}$ *satisfies the conditions of Proposition 2.2.*

Proof. Let $Y = \operatorname{span}\{\mathbf{e}^{n+1}, \ldots, \mathbf{e}^m\}$, where $\mathbf{e}^j$ is the jth unit vector in $\mathbb{R}^m$. Set $I = \{1, \ldots, n\}$. Then obviously

$$0 = \sum_{i=1}^{n} |y_i| < \sum_{i=n+1}^{m} |y_i|$$

for all $\mathbf{y} \in Y$, $\mathbf{y} \neq \mathbf{0}$.

Any matrix A associated with Y has columns $n+1, \ldots, m$ all zero. No one likes matrices with $m-n$ zero columns. But a small perturbation of A implies a small perturbation of Y which, in turn, implies that the above property of I is maintained under small perturbations. Thus we can get rid of the zero columns and in fact ensure that for the new matrix A all choices of n columns are linearly independent. □

A more interesting question is that of the largest possible value of k such that for *every* $n \times m$ matrix A of rank n there exists an $I \subseteq \{1, \ldots, m\}$ satisfying $|I| \geq k$ and such that Proposition 2.2 holds. A simple lower bound is the following.

Proposition 2.8. *Let* A *and* Y *be as above. Then there always exists an* $I \subseteq \{1, \ldots, m\}$ *satisfying the conditions of Proposition 2.2 with* $|I| \geq [m/(m-n+1)]$.

Proof. Let $r = [m/(m-n+1)]$. We divide the m indices $\{1, \ldots, m\}$ into r groups of $m-n+1$ distinct indices and discard whatever might be left over. Since Y is a subspace of dimension $m-n$, for each given group of indices $\{j_1, \ldots, j_{m-n+1}\}$ there exists an $\mathbf{a} \in \mathbb{R}^m \setminus \{\mathbf{0}\}$ such that $a_i = 0$, $i \neq j_k$, and $\langle \mathbf{a}, \mathbf{y} \rangle = 0$ for all $\mathbf{y} \in Y$. Let j_p be such that $1 = |a_{j_p}| \geq |a_{j_k}|$, $k = 1, \ldots, m-n+1$. Thus we have

$$|y_{j_p}| = \left| \sum_{k=1,k\neq p}^{m-n+1} a_{j_k} y_{j_k} \right| \leq \sum_{k=1,k\neq p}^{m-n+1} |y_{j_k}|$$

for all $\mathbf{y} \in Y$. Summing over the r groups of indices, the result now follows. □

A less elementary result is the following asymptotic result in m (or n) for $m - n$ fixed. We do not present its proof here. It may be found in Pinchasi, Pinkus [25].

Theorem 2.9. *Let A be any $n \times m$ matrix of rank n. For $m-n$ fixed, there exists a subset I of the indices $\{1, 2, \ldots, m\}$ of cardinality $(1/2 - o(1))m$ satisfying the conditions of Proposition 2.2.*

Note that this is not really a result about A, but is a result about the $(m - n)$-dimensional subspace Y of $\mathbb{R}^m$. It says that asymptotically, for m and n large but $m - n$ fixed, there is a fixed subset of indices of cardinality almost $m/2$ such that, for every $\mathbf{x} \in \mathbb{R}^m$ with support in this subset, the $\mathbf{0}$-vector is a best approximation to $\mathbf{x}$ from Y.

3. k-sparsity

We consider the same setting as in Section 2, but now ask the following question. What are conditions on A implying that, if supp $\mathbf{x} \subseteq I$ and I is *any* subset of $\{1, \ldots, m\}$ with $|I| \leq k$, then for all $\mathbf{x}$ with supp $\mathbf{x} \subseteq I$ we have

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1?$$

For ease of expression, we will say that $\mathbf{x}$ is k-**sparse** if it has at most k nonzero coordinates, i.e., $|\text{supp}\, \mathbf{x}| \leq k$. Thus we ask when we can find and identify a positive integer k such that for all k-sparse $\mathbf{x}$ we have

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1.$$

From Proposition 2.2(c), we see that this is equivalent to demanding that

$$\sum_{i\in I} |y_i| \leq \sum_{i\in I^c} |y_i|$$

for all $\mathbf{y} \in Y$ and for all $I \subseteq \{1, \ldots, m\}$ with $|I| \leq k$. We rewrite the above inequality as

$$2\sum_{i\in I} |y_i| \leq \|\mathbf{y}\|_1.$$

While $\sum_{i\in I} |y_i|$ is not a norm on $\mathbb{R}^m$ for a fixed I and $k < m$, the following is a norm:

$$|||\mathbf{y}|||_k := \max_{|I|=k} \sum_{i\in I} |y_i|.$$

For convenience, set

$$R_k := \max_{\mathbf{y}\in Y} \frac{|||\mathbf{y}|||_k}{\|\mathbf{y}\|_1}.$$

Thus we have the following.

Proposition 3.1. *Let A and Y be as previously. For all k-sparse $\mathbf{x} \in \mathbb{R}^m$ we have*

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1$$

if and only if

$$R_k \le \frac{1}{2}.$$

Furthermore, if $R_k < 1/2$, then, for all $\mathbf{x}$, as above, and for every $\mathbf{z}$ satisfying $A\mathbf{z} = A\mathbf{x}$, we have

$$\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 \ge (1 - 2R_k)\|\mathbf{z} - \mathbf{x}\|_1.$$

The latter statement in the above proposition is a direct consequence of Proposition 2.6. The constant $1 - 2R_k$ therein is best possible.

Thus we wish to determine conditions on A (or Y) and k such that

$$R_k \le \frac{1}{2}.$$

We start with a simple bound.

Theorem 3.2. *Let A and R_k be as above. If $R_k \le 1/2$ then necessarily*

$$k \le \frac{n+1}{2}.$$

Proof. We will show that for every Y of dimension $m - n$ we have

$$R_k \ge \frac{k}{n+1},$$

from which the above result immediately follows.

Let X be any subspace of dimension $n + 1$. As Y is a subspace of dimension $m - n$, the subspace $X \cap Y$ is of dimension at least 1. Thus,

$$\max_{\mathbf{y}\in Y} \frac{\|\!|\mathbf{y}|\!\|_k}{\|\mathbf{y}\|_1} \ge \min_{\mathbf{x}\in X} \frac{\|\!|\mathbf{x}|\!\|_k}{\|\mathbf{x}\|_1}$$

for every such X. Thus,

$$\max_{\mathbf{y}\in Y} \frac{\|\!|\mathbf{y}|\!\|_k}{\|\mathbf{y}\|_1} \ge \max_{\dim X=n+1} \min_{\mathbf{x}\in X} \frac{\|\!|\mathbf{x}|\!\|_k}{\|\mathbf{x}\|_1}.$$

(This latter quantity is called the **Bernstein n-width** of Y.)

Let $X = \mathrm{span}\{\mathbf{e}^1, \ldots, \mathbf{e}^{n+1}\}$. For $\mathbf{x} \in X$, we have $\mathbf{x} = (x_1, \ldots, x_{n+1}, 0, \ldots, 0)$. Assume without loss of generality that $\|\!|\mathbf{x}|\!\|_k = 1$ and $|x_1| \ge |x_2| \ge \cdots \ge |x_{n+1}|$. Thus,

$$1 = \|\!|\mathbf{x}|\!\|_k = \sum_{i=1}^{k} |x_i|$$

and

$$|x_i| \le |x_k| \le \frac{1}{k}$$

for $i = k+1, \ldots, n+1$. Therefore,

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n+1} |x_i| = \sum_{i=1}^{k} |x_i| + \sum_{i=k+1}^{n+1} |x_i| \le 1 + (n+1-k)\frac{1}{k} = \frac{n+1}{k},$$

implying that

$$\min_{\mathbf{x}\in X} \frac{|||\mathbf{x}|||_k}{\|\mathbf{x}\|_1} \ge \frac{k}{n+1}.$$

Here is a simpler proof of this same fact. As Y is of dimension $m-n$, there exists a nontrivial $\mathbf{y} \in Y \setminus \{\mathbf{0}\}$ that vanishes on any given $m-n-1$ indices; i.e., there exists a $\mathbf{y} \in Y \setminus \{\mathbf{0}\}$ with

$$|\text{supp}\, \mathbf{y}| \le n+1.$$

If $k > (n+1)/2$, then for this $\mathbf{y}$ we necessarily have

$$\frac{|||\mathbf{y}|||_k}{\|\mathbf{y}\|_1} > \frac{1}{2}. \quad \square$$

What we have shown above is that the appropriate Bernstein n-width is bounded below by $k/(n+1)$. It is an open question as to whether this is the true value of this n-width.

We now consider in somewhat more detail the norm $|||\cdot|||_k$. We defer to the Appendix a proof of the following facts.

Proposition 3.3. *The dual norm of* $|||\cdot|||_k$ *is given by*

$$|||\mathbf{x}|||_k^* = \max\left\{\|\mathbf{x}\|_\infty, \frac{\|\mathbf{x}\|_1}{k}\right\}.$$

In addition, the extreme points of the unit ball of $|||\cdot|||_k^*$ *are the vectors of the form*

$$\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$$

for all choices of $\varepsilon_j \in \{-1, 1\}$, *and for all choices of* $1 \le i_1 < \cdots < i_k \le m$.

Let $\mathbf{a}^1, \ldots, \mathbf{a}^n$ denote the (linearly independent) rows of A. Thus,

$$Y = \{\mathbf{y} : \langle \mathbf{a}^k, \mathbf{y}\rangle = 0,\ k = 1, \ldots, n\}.$$

Let $\mathcal{A} := \text{span}\{\mathbf{a}^1, \ldots, \mathbf{a}^n\}$. The subspace $\mathcal{A} = Y^\perp$ is always uniquely defined, although, for any given $(m-n)$-dimensional subspace Y, as above, the matrix A is not uniquely defined.

We will use the dual norm and the following well-known duality result.

Proposition 3.4. *For* $\mathcal{A}$ *and* Y *as above,*

$$R_k = \max_{\mathbf{y}\in Y} \frac{|||\mathbf{y}|||_k}{\|\mathbf{y}\|_1} = \max_{|||\mathbf{x}|||_k^* \le 1} \min_{\mathbf{a}\in\mathcal{A}} \|\mathbf{x} - \mathbf{a}\|_\infty.$$

When maximizing on the right-hand side it suffices to maximize over all extreme points. Thus, from Propositions 3.3 and 3.4, we have

$$R_k = \max_{\substack{\varepsilon_j \in\{-1,1\}\\ i_1<\cdots<i_k}} \min_{\mathbf{a}\in\mathcal{A}} \left\| \sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j} - \mathbf{a} \right\|_\infty .$$

This is the problem we now consider in detail in the remainder of this section. We recall that we want conditions on A and k such that $R_k \le 1/2$. From the above characterization of R_k we realize that our problem is one of bounding the error in approximating, in the uniform norm, certain specific linear combinations of k unit vectors. Given our $n \times m$ matrix A, then, for every matrix B with n columns, the rows of BA are in $\mathcal{A}$. In what follows, we look for suitable choices of B so as to facilitate this approximation problem. We choose $m \times n$ matrices B so that BA has the desirable property that its m rows are reasonably good approximations to the m unit vectors. This may be far from optimal, but with a lack of any further structure it is a reasonable strategy.

In the remainder of this section we assume that the column vectors $\mathbf{c}^\ell$, $\ell = 1, \dots, m$, of our $n \times m$ matrix A are all nonzero. Set

$$M := \max_{i \neq j} \frac{|\langle \mathbf{c}^i, \mathbf{c}^j \rangle|}{\langle \mathbf{c}^i, \mathbf{c}^i \rangle}.$$

We also assume, in what follows, that $M < 1$. In the literature of sparse representations, one considers matrices A with column vectors of unit Euclidean norm, and then the above M is called the **mutual incoherence** of A. This is a minor generalization thereof. (We will discuss bounds on M in the next section.)

We start with a now classical result due to Donoho, Huo [6].

Proposition 3.5. *Let A and M be as above. Then*

$$R_k \le \frac{kM}{1+M}.$$

Thus, if

$$k \le \frac{1}{2}\left(1 + \frac{1}{M}\right),$$

then, for all k-sparse $\mathbf{x} \in \mathbb{R}^m$, we have

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1.$$

Proof. We present two proofs of this result.

For every choice of $\varepsilon_j \in \{-1, 1\}$, and $1 \le i_1 < \cdots < i_k \le m$, we have

$$\min_{\mathbf{a}\in\mathcal{A}} \left\| \sum_{j=1}^k \varepsilon_j \mathbf{e}^{i_j} - \mathbf{a} \right\|_\infty \le \sum_{j=1}^k \min_{\mathbf{a}\in\mathcal{A}} \|\mathbf{e}^{i_j} - \mathbf{a}\|_\infty.$$

Let D denote the $m \times m$ diagonal matrix with diagonal entries $d_{ii} = 1/\langle \mathbf{c}^i, \mathbf{c}^i \rangle$, $i = 1, \dots, m$. The matrix $G = DA^T A$ is an $m \times m$ matrix whose rows all lie in $\mathcal{A}$, i.e., are linear combinations of the rows of A. The ith row, which we denote by $\mathbf{g}^i$, has jth entry equal to $\langle \mathbf{c}^i, \mathbf{c}^j \rangle / \langle \mathbf{c}^i, \mathbf{c}^i \rangle$, $i = 1, \dots, m$. Thus its ith entry equals 1, and its jth entry, $j \neq i$, is bounded in absolute value by M.

Consider $\|\mathbf{e}^{i_j} - \alpha_j \mathbf{g}^{i_j}\|_\infty$. This equals

$$\max\{|1-\alpha_j|, |\alpha_j||\langle \mathbf{c}^{i_j}, \mathbf{c}^k \rangle|/\langle \mathbf{c}^{i_j}, \mathbf{c}^{i_j} \rangle : k \neq i_j,\ k = 1, \dots, m\}.$$

We wish to choose α_j which minimizes this quantity. It is easily checked that a good universal choice of α_j is

$$\alpha_j = \frac{1}{1+M},$$

where the M is defined as above. Thus, for each j, we have

$$\|\mathbf{e}^{i_j} - \alpha_j \mathbf{g}^{i_j}\|_\infty \le \frac{M}{1+M},$$

and therefore

$$R_k \le \sum_{j=1}^{k} \frac{M}{1+M} = \frac{kM}{1+M}.$$

Here is an even simpler proof due to Gribonal, Nielsen [14]. Let $\mathbf{y} \in Y$, i.e., $A\mathbf{y} = \mathbf{0}$. Thus,

$$\sum_{j=1}^{m} y_j \mathbf{c}^j = \mathbf{0},$$

whence

$$y_i \mathbf{c}^i = -\sum_{j \ne i} y_j \mathbf{c}^j.$$

Therefore,

$$y_i \langle \mathbf{c}^i, \mathbf{c}^i \rangle = \left\langle \mathbf{c}^i, -\sum_{j \ne i} y_j \mathbf{c}^j \right\rangle,$$

implying that

$$|y_i| \le \sum_{j \ne i} |y_j| M.$$

The above can be written as

$$(1+M)|y_i| \le \|\mathbf{y}\|_1 M.$$

Thus, for each $i \in \{1, \ldots, m\}$, we have

$$\frac{|y_i|}{\|\mathbf{y}\|_1} \le \frac{M}{1+M},$$

and therefore

$$\frac{|||\mathbf{y}|||_k}{\|\mathbf{y}\|_1} \le \frac{kM}{1+M}. \quad \square$$

Another approach is the following. Set

$$\mu(s) := \max_{|\Gamma|=s} \max_{j \notin \Gamma} \sum_{i \in \Gamma} \frac{|\langle \mathbf{c}^i, \mathbf{c}^j \rangle|}{\langle \mathbf{c}^i, \mathbf{c}^i \rangle},$$

where Γ is any subset of $\{1, \ldots, m\}$. This result is due to Tropp [31].

Proposition 3.6. *Let A and $\mu(s)$ be as above. Then*

$$R_k \le \frac{\mu(k)}{\mu(k) - \mu(k-1) + 1}.$$

Thus, $R_k \le 1/2$ if

$$\mu(k) + \mu(k-1) \le 1.$$

Proof. We recall that $G = DA^T A$, as in the proof of Proposition 3.5, is an $m \times m$ matrix whose rows all lie in $\mathcal{A}$, i.e., are linear combinations of the rows of A. We approximate $\sum_{j=1}^k \varepsilon_j \mathbf{e}^{i_j}$ by $\alpha(\sum_{j=1}^k \varepsilon_j \mathbf{g}^{i_j})$ for some appropriate α. As $(\mathbf{g}^i)_i = 1$, a simple calculation shows that, for each $i \in \{i_1, \ldots, i_k\}$, we have

$$\left|\left(\sum_{j=1}^k \varepsilon_j \mathbf{e}^{i_j} - \alpha\left(\sum_{j=1}^k \varepsilon_j \mathbf{g}^{i_j}\right)\right)_i\right| \le |1-\alpha| + |\alpha|\mu(k-1),$$

while, for $i \in \{1, \ldots, m\} \setminus \{i_1, \ldots, i_k\}$, we have

$$\left|\left(\sum_{j=1}^k \varepsilon_j \mathbf{e}^{i_j} - \alpha\left(\sum_{j=1}^k \varepsilon_j \mathbf{g}^{i_j}\right)\right)_i\right| \le |\alpha|\mu(k).$$

Thus, the uniform norm of $\sum_{j=1}^k \varepsilon_j \mathbf{e}^{i_j} - \alpha(\sum_{j=1}^k \varepsilon_j \mathbf{g}^{i_j})$ is bounded above by

$$\max\{|1-\alpha| + |\alpha|\mu(k-1), |\alpha|\mu(k)\}$$

for each α. As $\mu(k) \ge \mu(k-1)$, we take

$$\alpha = \frac{1}{\mu(k) - \mu(k-1) + 1},$$

from which the result follows. □

Note that $M = \mu(1) \le \mu(2) \le \cdots \le \mu(m-1)$, and $\mu(s) \le sM$ for each s. Thus the bottom line of Proposition 3.5 also follows from this result.

In the sparse representation literature there are considered various estimates in the special case where A is an $n \times nL$ matrix of the form

$$A = [\Phi_1, \ldots, \Phi_L],$$

and the $\Phi_1, \ldots, \Phi_L$ are L $n \times n$ unitary matrices, i.e., $\Phi_i^T \Phi_i = I$. Note that, for convenience, we have assumed that the columns vectors of A are all of unit Euclidean norm.

The main result here is the following due to Donoho, Elad [5]; see p. 530.

Theorem 3.7. *Let $\Phi_1, \ldots, \Phi_L$ be L $n \times n$ unitary matrices. Let A denote the $n \times nL$ matrix*

$$A := [\Phi_1, \ldots, \Phi_L].$$

Set $M = \max_{i \ne j} |\langle \mathbf{c}^i, \mathbf{c}^j \rangle|$, where the $\mathbf{c}^i$ are the columns of A. Then, for all $L \ge 2$, we have $R_k \le 1/2$ if

$$k \le \left[\sqrt{2} - 1 + \frac{1}{2(L-1)}\right]\frac{1}{M}.$$

For $L \ge 3$, we also have that $R_k \le 1/2$ if

$$k \le \left[\frac{1}{2 - \frac{1}{L-1}}\right]\frac{1}{M}.$$

This latter bound is better when $L \ge 4$.

Proof. From the previous analysis, there exist $1 \le i_1 < \cdots < i_k \le nL$ and $\varepsilon_j \in \{-1, 1\}$ such that

$$R_k = \min_{\mathbf{a}\in\mathcal{A}} \left\| \sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j} - \mathbf{a} \right\|_\infty .$$

Define

$$C_p := \{j : (p-1)n + 1 \le i_j \le pn\},$$

and set $r_p := |C_p|$. Thus $r_1 + \cdots + r_L = k$. Consider

$$A^T A = \begin{bmatrix} I & \Phi_1^T \Phi_2 & \cdots & \Phi_1^T \Phi_L \\ \Phi_2^T \Phi_1 & I & \cdots & \Phi_2^T \Phi_L \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_L^T \Phi_1 & \Phi_L^T \Phi_2 & \cdots & I \end{bmatrix}.$$

We will approximate $\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$ using rows $i_1, \ldots, i_k$ from $G = A^T A$. We use the approximant

$$\sum_{p=1}^{L} \alpha_p \left(\sum_{i_j \in C_p} \varepsilon_j \mathbf{g}^{i_j} \right),$$

where $\mathbf{g}^i$ is the ith row of G. It follows that the error is bounded above by

$$\max_{p=1,\ldots,L} \left[|1-\alpha_p| + \sum_{j=1, j\ne p}^{L} |\alpha_j| r_j M \right],$$

since the diagonal entries of C are identically 1, while the off-diagonal entries are either 0 or bounded above by M, as may be seen from the form of C. Assume, without loss of generality, that $r_1 \ge r_2 \ge \cdots \ge r_L$. Set

$$\alpha_p := \frac{1 + r_L M}{1 + r_p M}, \quad p = 1, \ldots, L.$$

Note that we therefore have $0 < \alpha_1 \le \alpha_2 \le \cdots \le \alpha_L = 1$, and, for each $p = 1, \ldots, L$,

$$|1-\alpha_p| + \sum_{j=1, j\ne p}^{L} |\alpha_j| r_j M = (1 + r_L M) \left(\sum_{j=1}^{L-1} \frac{r_j M}{1 + r_j M} \right).$$

Set $r := r_L$, and thus $r_1 + \cdots + r_{L-1} = k - r$. As $r = r_L \le r_j$ for all j and $r_1 + \cdots + r_L = k$, we have $r \le k/L$. It is easily verified that

$$\max_{r_1+\cdots+r_{L-1}=k-r} \sum_{j=1}^{L-1} \frac{r_j M}{1 + r_j M} = \frac{(k-r)M}{1 + \frac{(k-r)M}{L-1}},$$

i.e., the maximum is attained by choosing $r_i = (k-r)/(L-1)$, $i = 1, \ldots, L-1$. Thus our error is bounded above by

$$\max_{0 \le r \le k/L} \frac{(1 + rM)(k-r)M}{1 + \frac{(k-r)M}{L-1}}. \tag{1}$$

We now determine the maximum of the above quantity as a function of r. Differentiating with respect to rM, we obtain that the derivative vanishes at

$$rM = kM + L - 1 - \sqrt{(kM + L)(L - 1)}. \tag{2}$$

We recall that $kM < 1$ and L is an integer, of value at least 2.

Thus the maximum of the expression (1) is attained at one of the three values $r = 0$, $r = k/L$, and the above r as in (2) if it lies in $[0, k/L]$. It may easily be verified that the value of the above maximum at $r = 0$ is

$$\frac{kM}{1 + \frac{kM}{L-1}},$$

while its value at $r = k/L$ is

$$\frac{kM(L - 1)}{L}.$$

Thus, assuming that $kM \le 1$, the value at $r = 0$ is larger than that at $r = k/L$.

Whether the maximum of (1) over $r \in [0, k/L]$ is attained at (2) depends on whether the r satisfying (2) is contained in $[0, k/L]$.

Substituting r satisfying (2) in (1), we have that

$$\frac{(1 + rM)(k - r)M}{1 + \frac{(k-r)M}{L-1}} = (L - 1)\left(\sqrt{kM + L} - \sqrt{L - 1}\right)^2.$$

This value is an upper bound on R_k. Thus this value is at most 1/2 if and only if

$$(L - 1)(\sqrt{kM + L} - \sqrt{L - 1})^2 \le \frac{1}{2}.$$

A calculation shows that this holds if and only if

$$kM \le \sqrt{2} - 1 + \frac{1}{2(L - 1)}. \tag{3}$$

Now $r \ge 0$ if and only if

$$kM \ge \sqrt{L - 1}\left(\frac{\sqrt{L + 3} - \sqrt{L - 1}}{2}\right). \tag{4}$$

If (3) does not hold, then the quantity in (1), as a function of r in $[0, k/L]$, is decreasing, and thus its maximum is at $r = 0$. Note that for (3) and (4) to hold we must have $L = 2$ and $L = 3$. For $L \ge 4$ they cannot simultaneously hold.

Thus, if (4) does not hold, i.e.,

$$kM \le \sqrt{L - 1}\left(\frac{\sqrt{L + 3} - \sqrt{L - 1}}{2}\right),$$

then the maximum equals

$$\frac{kM}{1 + \frac{kM}{L-1}}.$$

In this case, $R_k \le 1/2$ if

$$kM \le \frac{1}{2-\frac{1}{L-1}} = \frac{L-1}{2L-3}.$$

Thus, if

$$kM \le \min\left\{\sqrt{L-1}\left(\frac{\sqrt{L+3}-\sqrt{L-1}}{2}\right), \frac{1}{2-\frac{1}{L-1}}\right\},$$

then $R_k \le 1/2$. For $L \ge 3$, it follows that

$$\min\left\{\sqrt{L-1}\left(\frac{\sqrt{L+3}-\sqrt{L-1}}{2}\right), \frac{1}{2-\frac{1}{L-1}}\right\} = \frac{1}{2-\frac{1}{L-1}}.$$

Therefore, if $L \ge 3$ and

$$kM \le \frac{1}{2-\frac{1}{L-1}},$$

then $R_k \le 1/2$.

For $L = 2, 3$, the former bound is better. For $L \ge 4$, this bound is better. □

Remark. The above Theorem 3.7 is the last (so far) in a series of estimates. Elad, Bruckstein [8] first proved the bound $R_k \le 1/2$ if

$$k \le \frac{\sqrt{2}-1/2}{M},$$

when $L = 2$. In addition, it has been shown that this constant of $\sqrt{2}-1/2$ is best possible in the above form of the inequality when considering all pairs of unitary matrices; see Feuer, Nemirovski [9]. The inequality $R_k \le 1/2$ if

$$k \le \left[\sqrt{2}-1+\frac{1}{2(L-1)}\right]\frac{1}{M}$$

is contained in Gribonval, Nielsen [14]. An easier proof also gives the bound of $R_k \le 1/2$ if

$$k \le \frac{1}{(2-\frac{1}{L})M}.$$

4. Grassmannian

The results discussed in this section may be found in Pinkus [26, Chap. VI], where proofs are also provided. These results are also used in the theory of n-widths, which is the reason they appear in Pinkus [26]. We will utilize this connection in the next section, when we consider the problem of best possible $n \times m$ matrices A.

We recall that, for an $n \times m$ matrix A, with normalized column vectors $\mathbf{c}^i$, its mutual incoherence is defined by

$$M := \max_{i \ne j} |\langle \mathbf{c}^i, \mathbf{c}^j \rangle|.$$

As the bounds of Proposition 3.5 and Theorem 3.7 depend upon M (in an inverse manner), it is of interest to try to determine the minimal possible value of M (depending on n and m), when it can be attained, and what attaining it implies regarding the structure of A. We review this theory in the section.

We start with the following result, which is essentially due to van Lint, Seidel [32]. It also follows from n-width results as found in Pinkus [27].

Proposition 4.1. *For any unit vectors* $\{\mathbf{c}^1, \ldots, \mathbf{c}^m\}$ *in* $\mathbb{R}^n$, $m > n$, *we always have*

$$M = \max_{i \neq j} |\langle \mathbf{c}^i, \mathbf{c}^j \rangle| \geq \left(\frac{m-n}{n(m-1)} \right)^{\frac{1}{2}}.$$

Furthermore, we have equality if and only if

$$|\langle \mathbf{c}^i, \mathbf{c}^j \rangle| = \left(\frac{m-n}{n(m-1)} \right)^{\frac{1}{2}}$$

for all $i \neq j$.

This next result considers the case of equality in more detail. It is proven in Melkman [22]. It also uses results from the theory of n-widths.

Theorem 4.2. *The following are equivalent.*

(a) *There exists an* $m \times m$ *rank* n *matrix* C *satisfying* $c_{ii} = 1$, *for all* i, *and* $|c_{ij}| = M$, *for all* $i \neq j$, *where*

$$M = \left(\frac{m-n}{n(m-1)} \right)^{\frac{1}{2}}.$$

(b) *There exists an* $n \times m$ *matrix* A *such that* $C = A^T A$ *is as in (a) and* $AA^T = \frac{m}{n} I$.

(c) *There exists an* $m \times m$ *rank* $m - n$ *matrix* D *satisfying* $d_{ii} = 1$, *for all* i, *and* $|d_{ij}| = L$, *for all* $i \neq j$, *where*

$$L = \left(\frac{n}{(m-n)(m-1)} \right)^{\frac{1}{2}}.$$

(d) *There exists an* $(m - n) \times m$ *matrix* B *such that* $D = B^T B$ *is as in* (c) *and* $BB^T = \frac{m}{(m-n)} I$.

Unfortunately, it is known that such matrices do not exist for all values of m and n. In fact, we have the following limiting values on m and n. This result is contained in [21]. Its proof, however, is from [19].

Theorem 4.3. *Assume that matrices* C, A, D, *or* B *exist satisfying Theorem* 4.2. *If* $1 < n < m - 1$, *then we then necessarily have*

$$m \leq \min\{n(n+1)/2, (m-n)(m-n+1)/2\}.$$

In certain cases, we can attain the desired bound. For example, as pointed out by Melkman [21], see also Lemmens, Seidel [29], equality in Proposition 4.1 holds when $m = 2n$ if there exists a symmetric conference matrix. Sufficient conditions for the existence of such matrices are $m = p^k + 1$, if p is a prime and $m = 2 \pmod 4$. There are other cases where Theorem 4.2 holds; see Melkman [22] and Strohmer, Heath [30]. In Strohmer, Heath [30], some of these results were reported upon. They also discuss other contexts where these concepts arise. They call such matrices **Grassmannian**. As may be seen (it follows easily from the proofs of the above result), Grassmannianism is more concerned with approximating the unit vectors. That is, it is more applicable to approximating the unit ball of ℓ_1^m in the ℓ_∞^m-norm than to approximating the unit ball of the $\|\cdot\|_k^*$-norm. We return to this fact in the next section.

Remark. The results in this section are also valid when we replace $\mathbb{R}$ by $\mathbb{C}$. There are slight differences. For example, in Theorem 4.3, we have, over $\mathbb{C}$ and if $1 < n < m - 1$, the bounds

$$m \le \min\{n^2, (m-n)^2\}.$$

There are additional constructions over $\mathbb{C}$ of matrices satisfying equality in Theorem 4.2. For example, when $m = 2n$, if there exists an $m \times m$ skew Hadamard matrix S (S has entries 1 or -1, its rows are mutually orthogonal, and $S + S^T = 2I$), then we can construct C as in Theorem 4.2, over $\mathbb{C}^m$. C is given by

$$C = \left(1 - \frac{i}{\sqrt{2n-1}}\right) I + \frac{i}{\sqrt{2n-1}} S.$$

This result is from Melkman [21].

5. *n*-widths and optimal sparse representations

As we vary over all $n \times m$ matrices A, what is the largest possible value of k such that, if $\mathbf{x} \in \mathbb{R}^m$ is k-sparse, then $\mathbf{x}$ satisfies

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1?$$

That is, what is the largest possible value of k such that $R_k \le 1/2$ when considering all $n \times m$ matrices A? And what are good matrices A? These questions are essentially questions regarding n-widths. Let us explain. We recall from Proposition 3.4 that

$$R_k = \max_{\mathbf{y}\in Y} \frac{\|\mathbf{y}\|_k}{\|\mathbf{y}\|_1} = \max_{\|\mathbf{x}\|_k^* \le 1} \min_{\mathbf{a}\in\mathcal{A}} \|\mathbf{x} - \mathbf{a}\|_\infty.$$

Here, Y is a given $(m-n)$-dimensional subspace and $\mathcal{A} = Y^\perp$ is an n-dimensional subspace. If we wish to find an $n \times m$ matrix A with largest k for which $R_k \le 1/2$, then we are led to the consideration, for fixed n, m, and k, of the minimum of R_k over all possible $(m-n)$-dimensional subspaces Y, or equivalently all possible n-dimensional subspaces $\mathcal{A}$. Thus, we should consider

$$\min_Y \max_{\mathbf{y}\in Y} \frac{\|\mathbf{y}\|_k}{\|\mathbf{y}\|_1} = \min_{\mathcal{A}} \max_{\|\mathbf{x}\|_k^* \le 1} \min_{\mathbf{a}\in\mathcal{A}} \|\mathbf{x} - \mathbf{a}\|_\infty,$$

where Y and $\mathcal{A}$ vary, as above. These are general n-width quantities that have been considered in other contexts.

To explain, we start with some definitions. Assume that X is a normed linear space with norm $\|\cdot\|_X$, and that C is a closed, convex, and centrally symmetric (i.e., $x \in C$ implies that $-x \in C$) subset of X. The quantity

$$d_n(C; X) := \inf_{X_n} \sup_{x\in C} \inf_{y\in X_n} \|x - y\|_X$$

is called the **Kolmogorov** ***n*-width** of C in X. Here, X_n varies over all possible n-dimensional subspaces of X. The quantity

$$d^n(C; X) := \inf_{L^n} \sup_{x\in C\cap L^n} \|x\|_X$$

is called the **Gel'fand** n**-width** of C in X. Here, L^n varies over all possible subspaces of codimension n in X. We also define the quantity

$$\delta_n(C; X) := \inf_{P_n} \sup_{x \in C} \|x - P_n x\|_X.$$

This is called the **linear** n**-width** of C in X. Here, P_n varies over all possible linear operators of rank at most n taking X into X. For more on n-widths (including the Bernstein n-width defined earlier), see Pinkus [26] and Lorentz, v. Golitschek, Makovoz [20].

Setting $X = \ell_\infty^m$ in $\mathbb{R}^m$, and $C_k := \{\mathbf{x} : |||\mathbf{x}|||_k^* \le 1\}$, we see that

$$d_n(C_k; \ell_\infty^m) = \min_{\mathcal{A}} \max_{|||\mathbf{x}|||_k^* \le 1} \min_{\mathbf{a} \in \mathcal{A}} \|\mathbf{x} - \mathbf{a}\|_\infty,$$

where $\mathcal{A}$ is as above. Let $B(X)$ denote the unit ball in the normed linear space X, and Z_k be the normed linear space $\mathbb{R}^m$ with norm $|||\cdot|||_k$. Then

$$d^n(B(\ell_1^m); Z_k) = \min_Y \max_{\mathbf{y} \in Y} \frac{|||\mathbf{y}|||_k}{\|\mathbf{y}\|_1},$$

since an $(m - n)$-dimensional subspace in $\mathbb{R}^m$ is exactly the same as a subspace of codimension n in $\mathbb{R}^m$. The fact that for these choices of C_k and Z_k we have

$$d_n(C_k; \ell_\infty^m) = d^n(B(\ell_1^m); Z_k)$$

is a special case of a general duality result; see Proposition 3.4.

Note that we are considering these n-width quantities from a somewhat nonstandard point of view. One usually keeps the sets fixed and looks at d_n or d^n as a function of n. Here, we are considering d_n and d^n, for fixed n, and letting C_k and Z_k vary with k. Obviously these n-widths are nondecreasing functions of k, for fixed n. We therefore seek the largest value of k such that the associated n-width is less than or equal to $1/2$. Unfortunately, we know of no precise values for these n-widths. We only have estimates.

We already have one estimate, and that is the lower bound contained in Theorem 3.2. Namely, $d_n(C_k; \ell_\infty^m) = d^n(B(\ell_1^m); Z_k) \ge k/(n + 1)$ (from which follows the bound $k \le (n + 1)/2$ on possible "admissible" k). Note that there is no m in this bound. This is not optimal.

Let us start by considering upper bounds. These upper bounds are obtained using known estimates of other, more classical, n-widths, and come from a simple application of Hölder's inequality. Let $\|\cdot\|_p$ denote the ℓ_p^m norm on $\mathbb{R}^m$. For each Y, as previously defined, set

$$S_p := \max_{\mathbf{y} \in Y} \frac{\|\mathbf{y}\|_p}{\|\mathbf{y}\|_1}.$$

Proposition 5.1. *Let A, Y, R_k, and S_p be as above. Then, for each $p \in (1, \infty]$,*

$$R_k \le k^{1/p'} S_p,$$

where $1/p + 1/p' = 1$. Therefore, if k satisfies

$$k \le \frac{1}{(2S_p)^{p'}},$$

then, for all k-sparse $\mathbf{x} \in \mathbb{R}^m$, we have

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z} = A\mathbf{x}} \|\mathbf{z}\|_1.$$

Furthermore,

$$d_n(C_k;\ell_\infty^m) = d^n(B(\ell_1^m);Z_k) \le k^{1/p'} d_n(B(\ell_{p'}^m);\ell_\infty^m) = k^{1/p'} d^n(B(\ell_1^m);\ell_p^m).$$

Proof. The proposition is an immediate consequence of Hölder's inequality, namely,

$$|||\mathbf{y}|||_k = \sum_{j=1}^k |y_{i_j}| \le k^{1/p'} \left(\sum_{j=1}^k |y_{i_j}|^p\right)^{1/p} \le k^{1/p'} \|\mathbf{y}\|_p. \quad \square$$

We will consider and apply the above result in the two cases: $p=\infty$ and $p=2$.

For $p=\infty$, since the extreme points of the unit ball in ℓ_1^m are the coordinate directions, we have that

$$\begin{aligned} d_n(B(\ell_1^m);\ell_\infty^m) &= d^n(B(\ell_1^m);\ell_\infty^m) = \delta_n(B(\ell_1^m);\ell_\infty^m) \\ &= \min_{P_n} \max_{i=1,\dots,m} \|\mathbf{e}^i - P_n(\mathbf{e}^i)\|_\infty = \min_{Q_n} \max_{i,j=1,\dots,m} |\delta_{ij} - q_{ij}|, \end{aligned}$$

where P_n varies over all linear operators from $\mathbb{R}^m$ to $\mathbb{R}^m$ of rank at most n, and Q_n varies over all $m \times m$ matrices of rank at most n.

The exact value of this n-width is not known for all m and n. We do have both upper and lower bounds. From Proposition 5.1, it would seem that only the upper bounds are relevant here. However, the known lower bounds are sometimes the exact value of these n-widths. The following lower bound is from Pinkus [27].

Theorem 5.2. *For each $n < m$, we have*

$$d_n(B(\ell_1^m);\ell_\infty^m) = d^n(B(\ell_1^m);\ell_\infty^m) = \delta_n(B(\ell_1^m);\ell_\infty^m) \ge \frac{1}{1+\left(\frac{n(m-1)}{m-n}\right)^{\frac{1}{2}}}.$$

Furthermore, a necessary and sufficient condition for equality is that the conditions of Theorem 4.2 hold.

As there is a symmetry in Theorem 4.2 between n and $m-n$, equality in the above is equivalent to

$$d_{m-n}(B(\ell_1^m);\ell_\infty^m) = d^{m-n}(B(\ell_1^m);\ell_\infty^m) = \delta_{m-n}(B(\ell_1^m);\ell_\infty^m) = \frac{1}{1+\left(\frac{(m-n)(m-1)}{n}\right)^{\frac{1}{2}}}.$$

We may have equality in the above, but, as evidenced in Theorem 4.3 and the remarks thereafter, it may only hold for limited values of n and m. When it does hold, we have the following.

Corollary 5.3. *If equality holds in Theorem 5.2, i.e.,*

$$d^n(B(\ell_1^m);\ell_\infty^m) = \frac{1}{1+\left(\frac{n(m-1)}{m-n}\right)^{\frac{1}{2}}},$$

then there exists an $n \times m$ matrix A such that, if k satisfies

$$k \le \frac{1+\left(\frac{n(m-1)}{m-n}\right)^{\frac{1}{2}}}{2},$$

then, for all k-sparse $\mathbf{x} \in \mathbb{R}^m$*, we have*

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1.$$

The upper bounds for $d^n(B(\ell_1^m); \ell_\infty^m)$ are varied. The best of which we are aware is that of Höllig [15]. It is

$$d^n(B(\ell_1^m); \ell_\infty^m) \le \frac{8 \ln m}{\sqrt{n} \ln n}.$$

(A proof may be found in Pinkus [26].) From this we obtain the following.

Corollary 5.4. *For each* $m > n$*, there exists a real* $n \times m$ *matrix* A *such that, if*

$$k \le \frac{\sqrt{n} \ln n}{16 \ln m},$$

then, for all k-sparse $\mathbf{x} \in \mathbb{R}^m$*, we have*

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1.$$

Furthermore, such matrices can be explicitly constructed. A similar result, together with the construction of A, is contained in DeVore [3]. The existence of such a construction is undoubtedly aided by the fact that $d^n(B(\ell_1^m); \ell_\infty^m) = \delta_n(B(\ell_1^m); \ell_\infty^m)$, i.e., the problem has a linear element.

Remark. When $m = 2n$, we have from Elad, Bruckstein [8], see Theorem 3.7 or the remark near the end of Section 3, when $A = [\Phi_1, \Phi_2]$, where Φ_1, Φ_2 are $n \times n$ unitary matrices, that if

$$k \le \frac{\sqrt{2} - 1/2}{M}$$

then for all k-sparse $\mathbf{x} \in \mathbb{R}^m$ we have

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1.$$

There exist A, as above, with $M = 1/\sqrt{n}$ for many choices of n. The resulting bound $(\sqrt{2} - 1/2)\sqrt{n}$ is better than the bounds of Corollaries 5.3 and 5.4.

We now consider the case $p = p' = 2$. In this case, we get significantly better bounds. It is known from work of Kashin [17], Gluskin [13] and Garnaev, Gluskin [12] that the n-widths $d_n(B(\ell_2^m); \ell_\infty^m) = d^n(B(\ell_1^m); \ell_2^m)$ are of the order

$$C \min\left\{1, \sqrt{\frac{\ln(m/n) + 1}{n}}\right\}.$$

As an immediate consequence we have the following stronger corollary.

Corollary 5.5. *There exists a constant* C' *(independent of* m *and* n*) such that for each* $m > n$ *there exists an* $n \times m$ *matrix* A *for which we have that, if*

$$k \le \frac{C' n}{\ln(m/n) + 1},$$

then, for all k-sparse $\mathbf{x} \in \mathbb{R}^m$*, we have*

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1.$$

When $m = rn$, for any fixed r, we have here a bound of the order n, while the previous methods when $m = 2n$ only gave us bounds on the order of $\sqrt{n}$. Unfortunately, I am unaware of any explicit method of constructing such matrices. Any deterministic construction giving bounds better than of the order of $\sqrt{n}$ would be of significant interest. It should be noted, in this regard, that $\delta_n(B(\ell_1^m); \ell_2^m) = \sqrt{(m-n)/m}$ is of a totally different order. Another probabilistic approach giving bounds of this same order for $d^n(B(\ell_1^m); \ell_2^m)$ may be found in Donoho [4]. The relevance of these n-width estimates to this problem can be found in Mendelson, Pajor, Tomczak-Jaegermann [24].

The following was recently proven in Foucart, Pajor, Rauhut, Ullrich [10], based on a combinatorial result of Mendelson, Pajor, Rudelson [23] and volume estimates.

Theorem 5.6. *If for all k-sparse* $\mathbf{x} \in \mathbb{R}^m$ *we have*

$$\|\mathbf{x}\|_1 = \min_{A\mathbf{z}=A\mathbf{x}} \|\mathbf{z}\|_1,$$

for some $n \times m$ *matrix A, then k necessarily satisfies*

$$\left(\frac{m}{2k}\right)^{\frac{k}{4}} \le 3^n.$$

Thus, since $k \le (n+1)/2 \le m$*, we have*

$$k \le \frac{Cn}{\ln(m/(2k))} \le \frac{Cn}{\ln(m/(n+1))},$$

where $C = 4\ln 3$.

Note, however, that this can give less information than one might suppose. The maximum of

$$\left(\frac{m}{2k}\right)^{\frac{k}{4}}$$

as a function of k on $(0, \infty)$ is attained at $m/2e$. Thus, if $m \le cn$, where $c = 8e \ln 3 \approx 23.89\ldots$, then

$$\left(\frac{m}{2k}\right)^{\frac{k}{4}} \le e^{\frac{m}{8e}} \le 3^n,$$

and this bound on k is always satisfied. That is, getting reasonable and useful estimates for the constants in these various inequalities is also important.

Acknowledgments

My (rather tardy) introduction to sparse representations was based on a course given by Michael (Miki) Elad which I audited last semester. I want to thank him for his help. I also thank Carl de Boor, Albert Cohen, Shahar Mendelson, and the referees for their assistance.

Appendix

As promised in Section 3, we present here a proof of Proposition 3.3 regarding the dual norm $|||\cdot|||_k^*$ to $|||\cdot|||_k$. We first recall the statement of Proposition 3.3.

Proposition 3.3. *The dual norm of* $|||\cdot|||_k$ *is given by*

$$|||\mathbf{x}|||_k^* := \max\left\{\|\mathbf{x}\|_\infty, \frac{\|\mathbf{x}\|_1}{k}\right\}.$$

In addition, the extreme points of the unit ball of $|||\cdot|||_k^*$ *are the vectors of the form*

$$\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$$

for all choices of $\varepsilon_j \in \{-1, 1\}$, *and for all choices of* $1 \le i_1 < \cdots < i_k \le m$.

Proof. To verify the duality we need to prove either

$$\max_{|||\mathbf{y}|||_k \le 1} \langle \mathbf{x}, \mathbf{y} \rangle = |||\mathbf{x}|||_k^*$$

or

$$\max_{|||\mathbf{x}|||_k^* \le 1} \langle \mathbf{x}, \mathbf{y} \rangle = |||\mathbf{y}|||_k.$$

We prove the former.

Consider

$$\max_{|||\mathbf{y}|||_k \le 1} \langle \mathbf{x}, \mathbf{y} \rangle = \max_{|||\mathbf{y}|||_k \le 1} \sum_{i=1}^{m} |x_i||y_i|.$$

Assume that the above maximum is attained by some $\mathbf{y}$ with

$$1 = |||\mathbf{y}|||_k = \sum_{j=1}^{k} |y_{i_j}|.$$

Thus we also assume that $|y_r| \le \min_{j=1,\ldots,k} |y_{i_j}|$ for each $r \notin \{i_1, \ldots, i_k\}$. For every such $\mathbf{y}$, we do not alter the value $|||\mathbf{y}|||_k$ if we set $|y_r| = \min_{j=1,\ldots,k} |y_{i_j}|$ for each $r \notin \{i_1, \ldots, i_k\}$, and this choice does not decrease the value of

$$\max_{|||\mathbf{y}|||_k \le 1} \sum_{i=1}^{m} |x_i||y_i|.$$

Let $\min\{|y_{i_1}|, \ldots, |y_{i_k}|\} = \varepsilon$. As $\sum_{j=1}^{k} |y_{i_j}| = 1$, we have $0 \le \varepsilon \le 1/k$. Thus, we have

$$\begin{aligned}
\sum_{i=1}^{m} |x_i||y_i| &\le |x_{i_1}||y_{i_1}| + \cdots + |x_{i_k}||y_{i_k}| + \varepsilon \sum_{r \notin \{i_1,\ldots,i_k\}} |x_r| \\
&= |x_{i_1}|(|y_{i_1}| - \varepsilon) + \cdots + |x_{i_k}|(|y_{i_k}| - \varepsilon) + \varepsilon\|\mathbf{x}\|_1 \\
&\le [(|y_{i_1}| - \varepsilon) + \cdots + (|y_{i_k}| - \varepsilon)]\|\mathbf{x}\|_\infty + \varepsilon\|\mathbf{x}\|_1 \\
&= (1 - k\varepsilon)\|\mathbf{x}\|_\infty + \varepsilon\|\mathbf{x}\|_1 \le \max\left\{\|\mathbf{x}\|_\infty, \frac{\|\mathbf{x}\|_1}{k}\right\}.
\end{aligned}$$

This proves the upper bound.

It is easily verified that

$$\max_{|||\mathbf{y}|||_k \le 1} \sum_{i=1}^{m} |x_i||y_i|$$

is at least as large as both $\|\mathbf{x}\|_\infty$ and $\|\mathbf{x}\|_1/k$. We get the former by taking $|y_i| = 1$ for some i, where $|x_i| = \|\mathbf{x}\|_\infty$ and $y_j = 0$ for $j \neq i$. Taking $|y_i| = 1/k$ for all i gives a bound of $\|\mathbf{x}\|_1/k$. This proves the duality.

It remains to prove that the extreme points of the unit ball of $|||\cdot|||_k^*$ are the vectors of the form

$$\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$$

for all choices of $\varepsilon_j \in \{-1, 1\}$, and for all choices of $1 \le i_1 < \cdots < i_k \le m$.

For each such vector, consider

$$\max_{|||\mathbf{x}|||_k^* \le 1} \left\langle \mathbf{x}, \sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j} \right\rangle.$$

As

$$\left\langle \mathbf{x}, \sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j} \right\rangle = \sum_{j=1}^{k} \varepsilon_j x_{i_j}$$

and $|||\mathbf{x}|||_k^* \le 1$, it readily follows that the maximum is uniquely attained by the vector $\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$ itself. Thus, $\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$ is an extreme point of the unit ball of $|||\cdot|||_k^*$.

Now, for any $\mathbf{y} \neq \mathbf{0}$, consider

$$\max_{|||\mathbf{x}|||_k^* \le 1} \langle \mathbf{x}, \mathbf{y} \rangle.$$

Assume, without loss of generality, that

$$|y_1| \ge |y_2| \ge \cdots \ge |y_k| \ge \cdots.$$

If $|y_k| > 0$, set $\mathbf{x} = \sum_{j=1}^{k} (\text{sgn } y_j)\mathbf{e}^j$. If $|y_\ell| > y_{\ell+1} = \cdots = y_k = 0$, let $\mathbf{x} = \sum_{j=1}^{\ell} (\text{sgn } y_j)\mathbf{e}^j + \sum_{j=\ell+1}^{k} \varepsilon_j \mathbf{e}^j$ for arbitrary $\varepsilon_j \in \{-1, 1\}$. Thus, $|||\mathbf{x}|||_k^* = 1$ and

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{k} |y_i| = |||\mathbf{y}|||_k.$$

We see that the above maximum is always attained by a vector of the form

$$\sum_{j=1}^{k} \varepsilon_j \mathbf{e}^{i_j}$$

for some choice of $\varepsilon_j \in \{-1, 1\}$, and for some choice of $1 \le i_1 < \cdots < i_k \le m$. This implies that the above set of vectors is all the extreme points of the unit ball of $|||\cdot|||_k^*$. □

References

[1] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (2009) 34–81.
[2] R.A. DeVore, Optimal computation, in: Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006, vol. I, EMS, Zürich, 2007, pp. 187–215.
[3] R. DeVore, Deterministic constructions of compressed sensing matrices, J. Complexity 23 (2007) 918–925.

[4] D.L. Donoho, For most large underdetermined systems of equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution, Comm. Pure Appl. Math. 59 (2006) 907–934.
[5] D.L. Donoho, M. Elad, On the stability of the basis pursuit in the presence of noise, Signal Process. 86 (2006) 511–532.
[6] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Information Theory 47 (2001) 2845–2862.
[7] M. Elad, Sparse and Redundant Representations — From Theory to Applications in Signal and Image Processing, Springer, 2010.
[8] M. Elad, A.M. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. Information Theory 48 (2002) 2558–2567.
[9] A. Feuer, A. Nemirovski, On sparse representation in pairs of bases, IEEE Trans. Information Theory 49 (2003) 1579–1581.
[10] S. Foucart, A. Pajor, H. Rauhut, T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$, J. Complexity 26 (2010) 629–640.
[11] J.J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Information Theory 50 (2004) 1341–1344.
[12] A.Yu. Garnaev, E.D. Gluskin, The widths of a Euclidean ball, Dokl. Akad. Nauk SSSR 277 (1984) 1048–1052.
[13] E.D. Gluskin, Norms of random matrices and widths of finite-dimensional sets, Mat. Sb. (N.S.) 120 (1983) 180–189; transl. in Math. USSR Sb. 48 (1984) 173–182.
[14] R. Gribonval, M. Nielsen, Sparse decompositions in unions of bases, IEEE Trans. Information Theory 49 (2003) 3320–3325.
[15] K. Höllig, Approximationszahlen von Sobolev–Einbettungen, Math. Ann. 242 (1979) 273–281.
[16] R.C. James, Orthogonality and linear functionals in normed linear spaces, Trans. Amer. Math. Soc. 61 (1947) 265–292.
[17] B. Kashin, The widths of certain finite dimensional sets and classes of smooth functions, Izv. Akad. Nauk SSSR 41 (1977) 334–351; transl. in Math. USSR Izv. 11 (1977) 317–333.
[18] B.R. Kripke, T.J. Rivlin, Approximation in the metric of $L^1(X, \mu)$, Trans. Amer. Math. Soc. 119 (1965) 101–122.
[19] P.W.H. Lemmens, J.J. Seidel, Equiangular lines, J. Algebra 24 (1973) 494–512.
[20] G.G. Lorentz, M.v. Golitschek, Y. Makovoz, Constructive Approximation: Advanced Problems, in: Grundlehren, vol. 304, Springer-Verlag, Berlin, 1996.
[21] A.A. Melkman, n-widths of octahedra, in: Quantitative Approximation (Proc. Internat. Sympos., Bonn, 1979), Academic Press, New York, London, 1980, pp. 209–216.
[22] A.A. Melkman, The distance of a subspace of $\mathbb{R}^m$ from its axes and n-widths of octahedra, J. Approx. Theory 42 (1984) 245–256.
[23] S. Mendelson, A. Pajor, M. Rudelson, The geometry of random $\{-1, 1\}$-polytopes, Disc. Comput. Geom. 34 (2005) 365–379.
[24] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Reconstruction and subgaussian operators in asymptotic geometric analysis, Geom. Funct. Anal. 17 (2007) 1248–1282.
[25] R. Pinchasi, A. Pinkus, Dominating subsets under projections, SIAM J. Discrete Math. 24 (2010) 910–920.
[26] A. Pinkus, n-Widths in Approximation Theory, in: Ergebnisse, vol. 7, Springer-Verlag, Berlin, 1985.
[27] A. Pinkus, Matrices and n-widths, Linear Algebra Appl. 27 (1979) 245–278.
[28] A. Pinkus, On L^1-Approximation, in: Cambridge Tracts in Mathematics, vol. 93, Cambridge University Press, 1989.
[29] J.J. Seidel, A survey in 2-graphs, in: Proceedings Int. Colloq. Teorie Combinatorie, Accad. Naz. Lincei, Rome, 1976, pp. 481–511.
[30] T. Strohmer, R.W. Heath, Grassmannian frames with applications to coding and communication, Applied and Computational Harmonic Analysis 14 (2003) 257–275.
[31] J.A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Information Theory 50 (2004) 2231–2242.
[32] J.H. van Lint, J.J. Seidel, Equilateral points sets in elliptic geometry, Indag. Math. 28 (1966) 335–348.