Geometric Approach to Error-Correcting Codes and Reconstruction of Signals

Mark Rudelson and Roman Vershynin

1 Error-correcting codes and transform coding

The results of this paper can be stated in three equivalent ways—in terms of the sparse recovery problem, the error-correction problem, and the problem of existence of certain extremal (neighborly) polytopes.

Error-correcting codes are used in modern technology to protect information from errors. Information is formed by finite words over some alphabet $\mathbb{F}$. The encoder transforms an n-letter word x into an m-letter word y with $m > n$. The decoder must be able to recover x correctly when up to r letters of y are corrupted in any way. Such an encoder-decoder pair is called an (n, m, r)-*error-correcting code*.

Development of algorithmically efficient error correcting codes has attracted attention of engineers, computer scientists, and applied mathematicians for the past five decades. Known constructions involve deep algebraic and combinatorial methods, see [34, 35, 36]. This paper develops an approach to error-correcting codes from the viewpoint of geometric functional analysis (asymptotic convex geometry). It thus belongs to a common ground of coding theory, signal processing, combinatorial geometry, and geometric functional analysis. Our argument, outlined in Section 3, may be of independent interest in geometric functional analysis.

Our main focus will be on words over the alphabet $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. In applications, these words may be formed of the coefficients of some signal (such as image or audio)

with respect to some basis or overcomplete system (Fourier, wavelet, etc.). Finite alphabets will be discussed in Section 5.

The simplest and most natural way to encode a vector $x \in \mathbb{R}^n$ into a vector $y \in \mathbb{R}^m$ is a linear transform

$$y = Qx, \tag{1.1}$$

where Q is given by an $m \times n$ matrix. A linear algebra argument gives that if $m \geq n + 2r$ and the range of Q is generic[1] then x can be recovered from y even if r coordinates of y are corrupted. This gives an (n, m, r)-error-correcting code. However, the decoder for this code has a huge computational complexity, as it involves a search through all r-element subsets of the components of y. Then the problem is: *how to reconstruct a vector y in an n-dimensional subspace Y of $\mathbb{R}^m$ from a vector $y' \in \mathbb{R}^m$ that differs from y in at most r coordinates?* An important feature inherent to this error-correction problem over the reals or complex numbers is that the errors may be of arbitrary magnitude. This is not the case in the more classical error-correction problem over finite alphabets (such as $\mathbb{F} = \{0, 1\}$). This accounts for a need of completely different method to deal with such errors, as well as for new challenges (such as stability of reconstruction).

A traditional and simple approach to denoising y', used in applications such as signal processing, is the mean least square (MLS) minimization. One hopes that y is well approximated by a solution to the minimization problem

$$\min_{u \in Y} \|u - y'\|_2, \tag{MLS}$$

where $\|x\|_2^2 = \sum_i |x_i|^2$. The solution to (MLS) is simply the orthogonal projection of y' onto Y. This cannot recover y exactly, and even the approximation is typically poor since we have no control of the magnitude of the errors in the corrupted coordinates. A promising alternative approach is the *basis pursuit* (BP). We simply replace the 1-norm by the 2-norm and expect y to be the *exact* and unique solution to the minimization problem

$$\min_{u \in Y} \|u - y'\|_1, \tag{BP}$$

where $\|x\|_1 = \sum_i |x_i|$. Thus a solution to (BP) is the metric projection of y' onto Y with

---

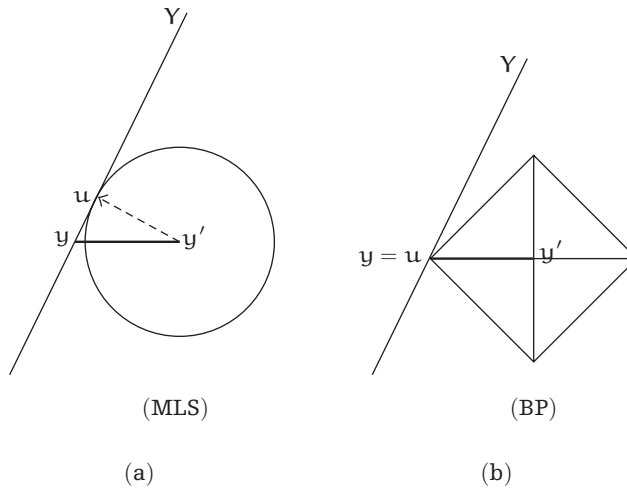[1] That is, in general position with respect to all subspaces $\mathbb{R}^I$, $|I| = r$.

Figure 1.1

respect to the 1-norm. (BP) is cast as a linear programming problem, and can be attacked with a variety of methods, such as the classical simplex method or more recent interior point methods that yield polynomial time algorithms [9].

The potential of basis pursuit for exact reconstruction is illustrated by the following heuristics, essentially due to [15]. The solution u to (MLS) is the contact point where the smallest Euclidean ball centered at y' meets the subspace Y, see Figure 1.1a. That contact point is in general different from y. The situation is much better in (BP): typically the solution coincides with y. The solution u to (BP) is the contact point where the smallest octahedron centered at y' (the ball with respect to the 1-norm) meets Y, see Figure 1.1b . Because the vector $y - y'$ lies in a low-dimensional coordinate subspace, the octahedron has a corner at y (or a wedge in high dimension). Thus, many subspaces Y through y will miss the octahedron of radius $y - y'$ (as opposed to the Euclidean ball). This forces the solution u to (BP), which is the contact point of the octahedron, to coincide with y.

The idea of using the 1-norm instead of the 2-norm for better data recovery has been explored since mid-seventies in various applied areas, in particular geophysics, and statistics (early history can be found in [38]). With the subsequent development of fast interior point methods in linear programming, (BP) turned into an effectively solvable problem, and was put forward more recently by Donoho and his collaborators, triggering massive experimental and theoretical work [4, 5, 7, 9, 12, 13, 14, 15, 16, 18, 19, 20, 21, 27, 37, 38, 39].

The main result of this paper validates the basis pursuit method for most subspaces Y under an asymptotically sharp condition on m, n, r. We thus prove that *the basis pursuit yields exact reconstruction for most subspaces* Y in the Grassmanian $G_{m,n}$ of n-dimensional subspaces of $\mathbb{R}^m$, equipped with the normalized Haar measure. Positive absolute constants will be denoted throughout the paper by $C, c, C_1, \ldots$.

Theorem 1.1. Let $m, n,$ and $r < cm$ be positive integers such that

$$m = n + R, \quad \text{where } R \geq Cr \log\left(\frac{m}{r}\right). \tag{1.2}$$

Then a random n-dimensional subspace Y in $\mathbb{R}^m$ in the Grassmanian satisfies the following with probability at least $1 - e^{-cR}$. Let $y \in Y$ be an unknown vector, and let $y' \in \mathbb{R}^m$ be a vector which differs from y on at most r coordinates. Then y can be exactly reconstructed from y' as the solution to the minimization problem (BP). □

In an equivalent form, this theorem is an improvement of recent results of Donoho [13] and of Candes and Tao [7], see Theorem 2.1.

1.1   Error-correcting codes

Theorem 1.1 implies a natural (n, m, r)-error-correcting code over $\mathbb{R}$. The encoder (1.1) is given by an $m \times n$ random orthogonal matrix[2] Q. Its range Y is a random n-dimensional subspace in $\mathbb{R}^m$. The decoder takes a corrupted vector y', solves (BP), and outputs $Q^\top u = Q^{-1}u$. Theorem 1.1 states that this encoder-decoder pair is an (n, m, r)-error-correcting code with exponentially good probability $\geq 1 - e^{-cR}$, provided the assumption (1.2) holds.

Assumption (1.2) meets, up to an absolute constant, the *Gilbert-Varshamov bound* which is fundamental in coding theory (see [34]): $n/m \geq 1 - H(Cr/n)$, where $H(x)$ is the entropy function. The encoder runs in quadratic time in the size n of the input, the decoder runs in polynomial time.

1.2   Sharpness

The sufficient condition (1.2) is sharp up to an absolute constant C (see Section 5) and is only slightly stronger than the necessary condition $m \geq n + 2r$. The ratio $\varepsilon = r/m$ in

---

[2]One can view it as the first n columns of a random matrix from the orthogonal group $O(m)$ equipped with the normalized Haar measure.

(1.2) is the number of errors per letter in the noisy communication channel that maps y to y'. Thus ε should be considered as a quality of the channel, which is independent of the message. Thus (1.2) is equivalent to

$$m \geq \left(1 + C\varepsilon \log \frac{1}{\varepsilon}\right)n. \tag{1.3}$$

1.3 Transform coding

In the signal processing, the linear codes (1.1) are known as *transform codes*. The general paradigm about transform codes is that the redundancies in the coefficients of y that come from the excess of the dimension $m > n$ should guarantee a stability of the signal with respect to noise, quantization, erasures, and so forth. This is confirmed by an extensive experimental and some theoretical work, see, for example, [3, 8, 10, 23, 24, 25, 26, 29] and the bibliography contained therein. Theorem 1.1 states that *most orthogonal transform codes are good error-correcting codes*.

2 Reconstruction of signals from linear measurements

A heuristic idea that motivates the reconstruction problem is that *a function f from a small class should be determined by few linear measurements*. Linear measurements are generally given by some linear functionals X_k in the dual space, which are fixed (in particular are independent of f). Most common measurements are point evaluation functionals; the problem there is to interpolate f between known values while keeping f in the known (small) class. When the evaluation points are chosen at random, this becomes the "proper learning" problem of the statistical learning theory (see [33]).

　　We will, however, be interested in general linear measurements. The proposal to learn f from general linear measurements (*sensing*) has been originated recently from a criticism of the current methodology of signal compression. Most of real-life signals, such as images and sounds, seem to belong to small classes. This is because they carry much of unwanted information that can be discarded with almost no perceptual loss, which makes such signals easily compressible. Donoho [11] then questions the conventional scheme of signal processing, where the whole signal must be first acquired (together with lots of unwanted information) and only then be compressed (throwing away the unwanted part). Instead, can one *directly acquire* (sense) the essential part of the signal, via few linear measurements? Similar issues are raised in [7]. We will operate under the assumption that some technology allows us to take linear measurements in certain fixed "directions" X_k.

We will assume that our signal f is discrete, so we view it as a vector in $\mathbb{R}^m$. Suppose we can take linear measurements $\langle f, X_k \rangle$ with some fixed vectors $X_1, X_2, \ldots, X_R$ in $\mathbb{R}^m$. Assuming that f belongs to a small class, how many measurements R are needed to reconstruct f? And even when we prove that R measurements do determine f (uniquely or approximately), the algorithmic issue remains unsettled: how can one reconstruct f from these measurements?

The previous section suggests to reconstruct f as a solution to the basis pursuit minimization problem

$$\min \|g\|_1 \quad \text{subject to } \langle g, X_k \rangle = \langle f, X_k \rangle, \quad k = 1, \ldots, R. \tag{BP$'$}$$

For the basis pursuit to work, the vectors X_k must be in a good position with respect to all coordinate subspaces $\mathbb{R}^I$, $|I| \le r$. A typical choice for such vectors would be the independent standard Gaussian vectors[3] X_k.

2.1   Functions with small support

In the class of functions with small support, one can hope for exact reconstruction. Candes and Tao [7] have indeed proved that every *fixed* function f with support $|\operatorname{supp} f| \le r$ can indeed be recovered by (BP$'$), correctly with the polynomial probability $1 - m^{-\text{const}}$, from the $R = Cr \log m$ Gaussian measurements. However, the polynomial probability is clearly not sufficient to deduce that there is *one* set vectors X_k that can be used to reconstruct all functions f of small support.

The following equivalent form of **Theorem 1.1** does yield a uniform exact reconstruction. It provides us with *one set* of linear measurements from which we can effectively reconstruct *every* signal of small support.

Theorem 2.1 (uniform exact reconstruction). Let $m, r < cm$ and R be positive integers satisfying $R \ge Cr \log(m/r)$. The independent standard Gaussian vectors X_k in $\mathbb{R}^m$ satisfy the following with probability at least $1 - e^{-cR}$. Let $f \in \mathbb{R}^m$ be an unknown function of small support, $|\operatorname{supp} f| \le r$. Then f can be exactly reconstructed from R measurements $\langle f, X_k \rangle$ as a solution to the basis pursuit problem (BP$'$).   □

This theorem gives uniformity in Candes-Tao result [7], improves the polynomial probability to an exponential probability, and improves upon the number R of measurements (which was $R \ge Cr \log m$ in [7]). Donoho [11] proved a weaker form of **Theorem 2.1** with R/r bounded below by some function of m/r.

---

[3]All the components of X_k are independent standard Gaussian random variables.

Proof. Write $g = f - u$ for some $u \in \mathbb{R}^m$. Then (BP$'$) reads as

$$\min \|u - f\|_1 \quad \text{subject to} \quad \langle u, X_k \rangle = 0, \quad k = 1, \ldots, R. \tag{2.1}$$

∎

The constraints here define a random ($n = m - R$)-dimensional subspace Y of $\mathbb{R}^m$. Now apply Theorem 1.1 with $y = 0$ and $y' = f$. It states that the unique solution to (2.1) is $u = 0$. Therefore, the unique solution to (BP$'$) is f.

2.2 Compressible functions

In a larger class of compressible functions [11], we can only hope for an approximate reconstruction. This is a class of functions f that are well compressible by a known orthogonal transform, such as Fourier or wavelet. This means that the coefficients of f with respect to a certain known orthogonal basis have a power decay. By applying an appropriate rotation, we can assume that this basis is the canonical basis of $\mathbb{R}^m$, thus f satisfies

$$f^*(s) \leq s^{-1/p}, \quad s = 1, \ldots, m, \tag{2.2}$$

where f^* denotes a nonincreasing rearrangement of $|f|$. Many natural signals are compressible for some $0 < p < 1$, such as smooth signals and signals with bounded variations (see [7]), in particular most photographic images. Theorem 2.1 implies, by the argument of [7], that functions compressible in some basis can be approximately reconstructed from few fixed linear measurements. This is an improvement of a result of Donoho [11].

Corollary 2.2 (uniform approximate reconstruction). Let m and r be positive integers. The independent standard Gaussian vectors X_k in $\mathbb{R}^m$ satisfy the following with probability at least $1 - e^{-cR}$. Assume that an unknown function $f \in \mathbb{R}^m$ satisfies either (2.2) for some $0 < p < 1$ or $\|f\|_1 \leq 1$ for $p = 1$. Then f can be approximately reconstructed from R measurements $\langle f, X_k \rangle$: a unique solution g to the basis pursuit problem (BP$'$) satisfies

$$\|f - g\|_2 \leq C_p \left(\frac{\log\left(\frac{m}{R}\right)}{R} \right)^{(1/p)-(1/2)}, \tag{2.3}$$

where C_p depends on p only. □

Corollary 2.2 was proved by Donoho [11] under an additional assumption that $m \sim CR^\alpha$ for some $\alpha > 1$. Notice that in this case $\log(m/R) \sim \log m$. Now this assumption

is removed. Candes and Tao [7] proved **Corollary 2.2** without the uniformity in f due to a weaker (polynomial) probability. Finally, **Corollary 2.2** also improves upon the approximation error (there is now the ratio m/r instead of m in the logarithm).

In a recent work independent of ours, Candes and Tao [6] have sharpened their previous work [7]. They obtained results similar to those in the present paper, but their approach is different.

3   Counting low-dimensional facets of polytopes

Theorem 1.1 turns out to be equivalent to a problem of counting lower-dimensional facets of polytopes. Let B_1^m denote the unit ball with respect to the 1-norm; it is sometimes called the unit octahedron. The polar body is the unit cube $B_\infty^m = [-1, 1]^m$. The conclusion of **Theorem 1.1** is then equivalent to the following statement: the affine subspace $z + Y$ is tangent to the unit octahedron at point z, where $z = y' - y$. This should happen for all z from the coordinate subspaces $\mathbb{R}^I$ with $|I| = r$. By duality, this means that the subspace $Y^\perp$ intersects all $(m - r)$-dimensional facets of the unit cube. The section of the cube by the subspace $Y^\perp$ forms an origin-symmetric polytope of dimension R and with $2m$ facets.

Our problem can thus be stated as a problem of counting lower-dimensional facets of polytopes. *Consider an R-dimensional origin-symmetric polytope with $2m$ facets. How many $(R - r)$-dimensional facets can it have?* Clearly,[4] no more than $2^r \binom{m}{r}$. Does there exist a polytope with that many facets? Our ability to construct such a polytope is equivalent to the existence of the efficient error-correcting code. Indeed, looking at the canonical realization of such a polytope as a section of the unit cube by a subspace $Y^\perp$, we see that $Y^\perp$ intersects all the $(m - r)$-dimensional facets of the cube. Thus Y satisfies the conclusion of **Theorem 1.1**. We can thus state **Theorem 1.1** in the following form.

Theorem 3.1. There exists an R-dimensional symmetric polytope with m facets and with the maximal number of $(R - r)$-dimensional facets (which is $2^r \binom{m}{r}$), provided $R \geq Cr \log(m/r)$. A random section of the cube forms such a polytope with probability $1 - e^{-cR}$.
$$\square$$

Note also that **Theorem 3.1** provides a construction of the so-called neighborly polytopes (see [28, Chapter 7] or [40]). An n-dimensional convex polytope is called

---

[4]Any such facet is the intersection of some r facets of the polytope of full dimension $R - 1$; there are m facets to choose from, each coming with its opposite by the symmetry.

m-neighborly if any set of m vertices forms an $(m-1)$-dimensional face. For symmetric polytopes this definition has to be modified, since a face cannot contain a pair of opposite vertices. A symmetric polytope is called m-neighborly if any set of m of its vertices, which does not contain opposite pairs, forms an $(m-1)$-dimensional face. The neighborly polytopes have been extensively studied recently in connection with computer science problems (see, e.g., [17]). In this terminology Theorem 3.1 can be reformulated as follows.

Theorem 3.2. Let m, r, R be positive integers satisfying

$$R \geq Cr \log \left(\frac{m}{r} \right). \tag{3.1}$$

Then a random R-dimensional projection of the m-dimensional octahedron B_1^m is an r-neighborly symmetric polytope with 2m vertices with probability greater than $1 - e^{-cR}$.

□

The rest of the paper is organized as follows. In Section 4 we prove Theorem 1.1. In Section 5 we discuss some optimality and robustness of the basis pursuit with applications to error-correcting codes over finite alphabets.

4   Proof of **Theorem 1.1**

4.1   Outline of the proof

We will use duality, which gives the following equivalent form of Theorem 1.1: a random $(m-n)$-dimensional subspace E intersects all $(m-r)$-dimensional facets of a cube in $\mathbb{R}^m$. It will then suffice to show that the probability of intersection with a fixed facet is suitably sufficiently large. This will be established by a proper use of the concentration of measure technique.

The main difficulty is that the ℓ_∞-norm defined by the unit cube (more precisely, by its facet) has a bad Lipschitz constant, which impedes the use of concentration inequalities. To improve the Lipschitzness, we first project the facet onto a random subspace (within its affine span); the kernel of this projection will form a part of the random subspace E. This will create a big Euclidean ball inside the projected facet. To prove this, we will use the full strength of the estimate of Garnaev and Gluskin [22] on Euclidean projections of a cube. The existence of the Euclidean ball inside a body creates the needed Lipschitzness, so we can now use the concentration of measure technique.

4.2   Notation

We will use the following standard notations throughout the proof. The p-norm ($1 \leq p < \infty$) on $\mathbb{R}^m$ is defined by $\|x\|_p^p = \sum_i |x_i|^p$, and for $p = \infty$ it is $\|x\|_\infty = \max_i |x_i|$. The unit ball with respect to the p-norm on $\mathbb{R}^n$ is denoted by B_p^m. When the p-norm is considered on a coordinate subspace $\mathbb{R}^I$, $I \subset \{1, \ldots, m\}$, the corresponding unit ball is denoted by B_p^I.

The unit Euclidean sphere in a subspace E is denoted by $S(E)$. The normalized rotational invariant Lebesgue measure on $S(E)$ is denoted by σ_E. The orthogonal projection onto a subspace E is denoted by P_E. The standard Gaussian measure on E (with the identity covariance matrix) is denoted by γ_H. When $E = \mathbb{R}^d$, we write σ_{d-1} for σ_E and γ_d for γ_E.

4.3   Duality

We begin the proof of **Theorem 1.1** with a typical duality argument, leading to the same reformulation of the problem as in [7]. We claim that the conclusion of **Theorem 1.1** follows from (and is actually equivalent to) the following separation condition:

$$(z + Y) \cap \text{ interior } (B_1^m) = \varnothing \quad \forall z \in \bigcup_{|I|=r} B_1^I. \tag{4.1}$$

Indeed, suppose (4.1) holds. We apply it for

$$z := \frac{y - y'}{\|y - y'\|_1} \tag{4.2}$$

noting that $z \in \bigcup_{|I|=r} B_1^I$ holds, because y and y' differ in at most r coordinates. By (4.1),

$$(z + v) \cap \text{ interior } (B_1^m) = \varnothing \quad \forall v \in Y \tag{4.3}$$

which implies

$$\|z + v\|_1 \geq 1 \quad \forall v \in Y. \tag{4.4}$$

Let $u \in Y$ be arbitrary. Using the inequality above for $v := (u - y)/\|u - y\|_1$, we conclude that

$$\|u - y\|_1 \geq \|y - y'\|_1 \quad \forall u \in Y. \tag{4.5}$$

This proves that y is indeed a solution to (**BP**). The solution to (**BP**) is unique with probability 1 in the Grassmanian. This follows from a direct dimension argument, see, for example, [7].

By Hahn-Banach theorem, the separation condition (4.1) is equivalent to the following: for every r-element set $I \subset \{1, \ldots, m\}$ and for every point z on the boundary of B_1^I there exists $w = w(z) \in Y^\perp$ such that

$$\langle w, z \rangle = \sup_{x \in B_1^m} \langle w, x \rangle = \|w\|_\infty. \tag{4.6}$$

This holds if and only if the components of w satisfy

$$\begin{aligned} w_j &= \text{sign}(z_j) \quad \text{for } j \in I, \\ |w_j| &\leq 1 \quad \text{for } j \in I^c. \end{aligned} \tag{4.7}$$

The set of vectors w in $\mathbb{R}^m$ that satisfy (4.7) form a $(m - r)$-dimensional facet of the unit cube B_∞^m. Then with $E := Y^\perp$, we can say that the conclusion of **Theorem 1.1** is equivalent to the following. *A random R-dimensional subspace E in $\mathbb{R}^m$ intersects all the $(m - r)$-dimensional facets of the unit cube with probability at least $1 - e^{-cR}$.*

It will be enough to show that E intersects *one fixed* facet with the probability $1 - e^{-cR}$. Indeed, the probability that E misses some facet would be at most Ne^{-cR}, where

$$N = 2^r \binom{m}{r} \leq \exp\left(Cr \log \frac{m}{r}\right) \tag{4.8}$$

is the total number of facets. An appropriate choice of the constant C in (1.2) yields $Ne^{-cR} \leq e^{-c'R}$.

4.4   Realizing a random subspace

We are to show that a random R-dimensional subspace E intersects one fixed $(m - r)$-dimensional facet of the unit cube B_∞^m with high probability. Such a facet $F = F_{I,w}$ is determined by a subset I of $\{1, \ldots, n\}$ of cardinality r and by a vector $w \in \{-1, 1\}^I$. Then $F_{I,w}$ consists of all points in the cube B_∞^n whose coordinates in I coincide with w. Without loss of generality, we can assume that our facet is

$$F = \left\{(w_1, \ldots, w_{m-r}, 1, \ldots, 1), \text{ all } |w_j| \leq 1\right\}, \tag{4.9}$$
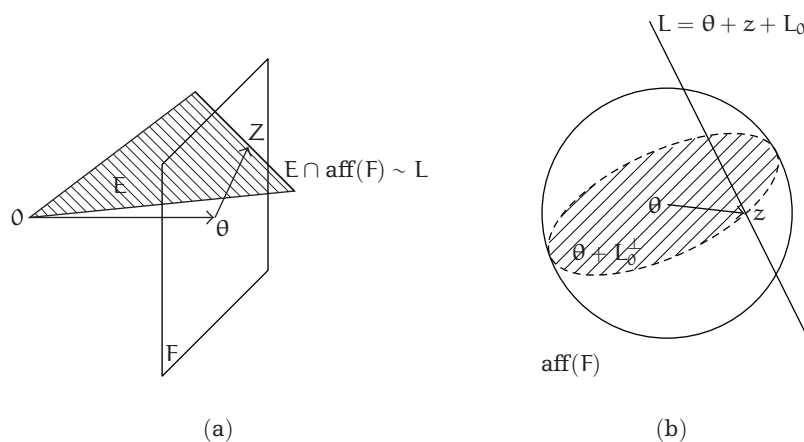
Figure 4.1

whose center is

$$\theta = \left(\underbrace{0, \ldots, 0}_{m-r}, 1, \ldots, 1 \right). \tag{4.10}$$

The probability we are interested in is

$$P := \text{Prob}\left\{ E \cap F \neq \varnothing \right\}. \tag{4.11}$$

We will restrict our attention to the linear span of F,

$$\text{lin}(F) = \left\{ (w_1, \ldots, w_{m-r}, t, \ldots, t), \text{ all } w_j \in \mathbb{R}, \, t \in \mathbb{R} \right\}, \tag{4.12}$$

and even to its affine span,

$$\text{aff}(F) = \left\{ (w_1, \ldots, w_{m-r}, 1, \ldots, 1), \text{ all } w_j \in \mathbb{R} \right\}. \tag{4.13}$$

Only the random affine subspace $E \cap \text{aff}(F)$ matters for us, because

$$P = \text{Prob}\left\{ \left(E \cap \text{aff}(F) \right) \cap F \neq \varnothing \right\}. \tag{4.14}$$

The dimension of that affine subspace is almost surely

$$l := \dim\left(E \cap \mathrm{aff}(F)\right) = R - r. \tag{4.15}$$

We can realize the random affine subspace $E \cap \mathrm{aff}(F)$ (or rather a random subspace with the same law) by the following algorithm which is illustrated in Figure 4.1.

(1) Select a random variable D with the same law as $\mathrm{dist}(\theta, E \cap \mathrm{aff}(F))$.

(2) Select a random subspace L_0 in the Grassmanian $G_{m-r,l}$. It will realize the "direction" of $E \cap \mathrm{aff}(F)$ in $\mathrm{aff}(F)$.

(3) Select a random point z on the Euclidean sphere $D \cdot S(L_0^\perp)$ of radius D, according to the uniform distribution on the sphere. Here $L_0^\perp$ is the orthogonal complement of L_0 in $\mathbb{R}^{m-r}$. The vector z will realize the distance from the affine subspace $E \cap \mathrm{aff}(F)$ to the center θ of F.

(4) Set $L = \theta + z + L_0$. Thus the random affine subspace L has the same law as $E \cap \mathrm{aff}(F)$.

Hence

$$P = \mathrm{Prob}\left\{L \cap F \neq \varnothing\right\} = \mathrm{Prob}\left\{(z + L_0 v) \cap B_\infty^{m-r} \neq \varnothing\right\} = \mathrm{Prob}\left\{z \in P_{L_0^\perp} B_\infty^{m-r}\right\}. \tag{4.16}$$

$H := L_0^\perp$ is a random subspace in $G_{m-r,m-r-l} = G_{m-r,m-R}$. By the rotational invariance of $z \in D \cdot S(H)$,

$$P = \int_{\mathbb{R}^+} \int_{G_{m-r,m-R}} \sigma_H\left(D^{-1} P_H B_\infty^{m-r}\right) d\nu(H) d\mu(D), \tag{4.17}$$

where ν is the normalized Haar measure on $G_{m-r,m-R}$ and μ is the law of D. We will bound P in two steps:

(1) prove that the distance D is small with high probability;

(2) prove that a suitable multiple of the random projection $P_H B_\infty^{m-r}$ has an almost full Gaussian (thus also spherical) measure.

4.5   The distance D from the center of the facet to a random subspace

We will first relate D, the distance to the affine subspace $E \cap \mathrm{aff}(F)$, to the distance to the linear subspace $E \cap \mathrm{lin}(F)$. Equivalently, we compute the length of the projection onto $E \cap \mathrm{lin}(F)$.
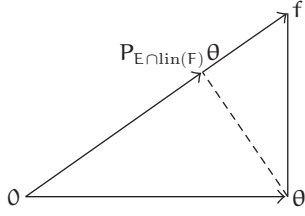
Figure 4.2

Lemma 4.1.

$$\left\|P_{E\cap\mathrm{lin}(F)}\theta\right\|_2 = \sqrt{\frac{r}{r+D^2}}\|\theta\|_2. \tag{4.18}$$

$\square$

Proof. Let f be the orthogonal projection of θ to the affine space $L = E \cap \mathrm{aff}(F)$. Then $f - \theta \perp \theta$ and $D = \|f - \theta\|_2$. Note that the vector f is orthogonal to L. Indeed, $\theta \perp \mathrm{aff}(F)$ and $f - \theta \perp L$. Also, for any $x \in L$,

$$\left\langle P_{E\cap\mathrm{lin}(F)}\theta, x\right\rangle = \left\langle\theta, P_{E\cap\mathrm{lin}(F)}x\right\rangle = \langle\theta, x\rangle = 0, \tag{4.19}$$

so $P_{E\cap\mathrm{lin}(F)}\theta \perp L$ as well. Therefore, both f and $P_{E\cap\mathrm{lin}(F)}\theta$ belong to the space $E\cap\mathrm{lin}(F)\cap L^{\perp}$, and since $\dim(E \cap \mathrm{lin}(F) \cap L^{\perp}) = 1$, f is a multiple of $P_{E\cap\mathrm{lin}(F)}\theta$.

By the similarity of the triangles with the vertices $(0, \theta, P_{E\cap\mathrm{lin}(F)}\theta)$ and $(0, f, \theta)$, we conclude that

$$\left\|P_{E\cap\mathrm{lin}(F)}\theta\right\|_2 = \frac{r}{\sqrt{r+D^2}} = \sqrt{\frac{r}{r+D^2}}\|\theta\|_2 \tag{4.20}$$

because $\|\theta\|_2 = \sqrt{r}$. This completes the proof. ■

The length of the projection of a fixed vector onto a random subspace in **Lemma 4.1** is well known. The asymptotically sharp estimate was computed by Artstein [2], but we will be satisfied with a much weaker elementary estimate, see, for example, [32, Lemma 15.2.2].

Lemma 4.2. Let $\theta \in \mathbb{R}^{d-1}$ and let G be a random subspace in $G_{d,k}$. Then

$$\mathrm{Prob}\left\{c\sqrt{\frac{k}{d}}\|\theta\|_2 \leq \|P_G\theta\|_2 \leq C\sqrt{\frac{k}{d}}\|\theta\|_2\right\} \geq 1 - 2e^{-ck}. \tag{4.21}$$

$\square$

We apply this lemma for $G = E \cap \operatorname{lin}(F)$, which is a random subspace in the Grassmanian of $(l+1)$-dimensional subspaces of $\operatorname{lin}(F)$. Since $\dim \operatorname{lin}(F) = m - r + 1$, we have

$$\operatorname{Prob}\left\{\|P_{E \cap \operatorname{lin}(F)}\theta\|_2 \geq c\sqrt{\frac{l+1}{m-r+1}}\|\theta\|_2\right\} \geq 1 - 2e^{-cl}. \tag{4.22}$$

Together with Lemma 4.1 this gives

$$\operatorname{Prob}\left\{D \leq c\sqrt{m-r}\sqrt{\frac{r}{l}}\right\} \geq 1 - 2e^{-cl}. \tag{4.23}$$

Note that $\sqrt{m-r}$ is the radius of the Euclidean ball circumscribed on the facet F. The statement $D \leq \sqrt{m-r}$ would only tell us that the random subspace E intersects the circumscribed ball, not yet the facet itself. The ratio r/l in (4.23) will be chosen logarithmically small, which will force E intersect also the facet F.

4.6 Gaussian measure of random projections of the cube

By (4.17) and (4.23),

$$P \geq \int_{G_{m-r,m-R}} \sigma_H\left(\frac{c}{\sqrt{m-r}}\sqrt{\frac{l}{r}}P_H B_\infty^{m-r}\right) d\nu(H) - 2e^{-cl}. \tag{4.24}$$

We can replace the spherical measure σ_H by the Gaussian measure γ_H via a simple and known lemma, whose proof we include for reader's convenience. A set K in a linear vector space is called star shaped if $x \in K$ implies $tx \in K$ for all $t \in [0,1]$.

Lemma 4.3. Let K be a star-shaped set in $\mathbb{R}^d$. Then

$$\gamma_d(c\sqrt{d} \cdot K) - e^{-d} \leq \sigma_{d-1}(K) \leq \gamma_d(C\sqrt{d} \cdot K) \cdot (1 + e^{-d}). \tag{4.25}$$

$\square$

Proof. Passing to polar coordinates, by the rotational invariance of the Gaussian measure we see that there exists a probability measure μ on $\mathbb{R}^+$ so that the Gaussian measure of every set A can be computed as $\int_{\mathbb{R}^+} \sigma^t(A)d\mu(t)$, where σ^t denotes the normalized Lebesgue measure on the Euclidean sphere of radius t in $\mathbb{R}^d$. Since K is star shaped, $\sigma^t(K)$ is a nonincreasing function of t. Hence

$$\gamma_d(K) \geq \int_0^{C\sqrt{d}} \sigma^t(K)d\mu(t) \geq \sigma^{C\sqrt{d}}(K) \cdot \gamma_d(C\sqrt{d}B_2^d),$$

$$\gamma_d(K) \leq \int_0^{c\sqrt{d}} d\mu(t) + \sigma^{c\sqrt{d}}(K)\int_{c\sqrt{d}}^\infty d\mu(t) \leq \gamma_d(c\sqrt{d} \cdot B_2^d) + \sigma^{c\sqrt{d}}(K). \tag{4.26}$$

The classical large deviation inequalities imply $\gamma_d(c\sqrt{d}\cdot B_2^d) \leq e^{-d}$ and $\gamma_d(C\sqrt{d}B_2^d) \geq 1 - e^{-d}/2$. Using the above argument for $c\sqrt{d}\cdot K$, we conclude that $\gamma_d(c\sqrt{d}\cdot K) \leq e^{-d} + \sigma_{d-1}(K)$ and $\gamma_d(C\sqrt{d}\cdot K) \geq \sigma_{d-1}(K)\cdot(1 - e^{-d}/2)$. $\blacksquare$

Using **Lemma 4.3** in the space H of dimension $d = m - R$, we obtain

$$P \geq \int_{G_{m-r,\,m-R}} \gamma_H\left(c\sqrt{\frac{m-R}{m-r}}\sqrt{\frac{l}{r}}P_H B_\infty^{m-r}\right) d\nu(H) - 2e^{-cl} - e^{m-R}. \tag{4.27}$$

By choosing the absolute constant c in the assumption $r < cm$ appropriately small, we can assume that $2r < R < m/2$. Thus

$$P \geq \int_{G_{m-r,\,m-R}} \gamma_H\left(c\sqrt{\frac{R}{r}}P_H B_\infty^{m-r}\right) d\nu(H) - 2e^{-cR}. \tag{4.28}$$

We now compute the Gaussian measure of random projections of the cube.

Proposition 4.4. Let H be a random subspace in $G_{n,n-k}$, $k < n/2$. Then the inequality

$$\gamma_H\left(C\sqrt{\log\frac{n}{k}}P_H B_\infty^n\right) \geq 1 - e^{-ck} \tag{4.29}$$

holds with probability at least $1 - e^{-ck}$ in the Grassmanian. $\square$

The proof of this estimate will follow from the concentration of Gaussian measure, combined with the existence of a big Euclidean ball inside a random projection of the cube.

Lemma 4.5 (concentration of Gaussian measure)**.** Let A be a measurable set in $\mathbb{R}^n$. Then for $\varepsilon > 0$,

$$\gamma_n(A) \geq e^{-\varepsilon^2 n} \quad \text{implies } \gamma_n\left(A + C\varepsilon\sqrt{n}B_2^n\right) \geq 1 - e^{-\varepsilon^2 n}. \tag{4.30}$$
$\square$

With the stronger assumption $\gamma(A) \geq 1/2$, this lemma is the classical concentration inequality, see [30, inequality (1.4)]. The fact that the concentration holds also for exponentially small sets follows formally by a simple extension argument that was first noticed by Amir and Milman in [1], see [30, Lemma 1.1].

The optimal result on random projections of the cube is due to Garnaev and Gluskin [22].

Theorem 4.6 (Euclidean projections of the cube [22]). Let H be a random subspace in $G_{n,n-k}$, where $k = \alpha n < n/2$. Then with probability at least $1 - e^{-ck}$ in the Grassmanian,

$$c(\alpha)P_H\left(\sqrt{n}B_2^n\right) \subseteq P_H\left(B_\infty^n\right) \subseteq P_H\left(\sqrt{n}B_2^n\right), \tag{4.31}$$

where

$$c(\alpha) = c\sqrt{\frac{\alpha}{\log\frac{1}{\alpha}}}. \tag{4.32}$$

$\square$

Proof of Proposition 4.4. Let $g_1, g_2, \ldots$ be independent standard Gaussian random variables. Then for a suitable positive absolute constant c and for every $0 < \varepsilon < 1/2$,

$$\gamma_n\left(C\sqrt{\log\frac{1}{\varepsilon}}B_\infty^n\right) = \text{Prob}\left\{\max_{1\leq j\leq n}|g_i| \leq C\sqrt{\log\frac{1}{\varepsilon}}\right\} \geq \left(\frac{1-\varepsilon^2}{10}\right)^n \geq e^{-\varepsilon^2 n}. \tag{4.33}$$

Since for every measurable set A and every subspace H one has $\gamma_H(P_H A) \geq \gamma(A)$, we conclude that

$$\gamma_H\left(C\sqrt{\log\frac{1}{\varepsilon}}P_H B_\infty^n\right) \geq e^{-\varepsilon^2 n} \quad \text{for } 0 < \varepsilon < \frac{1}{2}. \tag{4.34}$$

Then by Lemma 4.5,

$$\gamma_H\left(C\sqrt{\log\frac{1}{\varepsilon}}P_H B_\infty^n + C\varepsilon\sqrt{n}P_H B_2^n\right) \geq 1 - e^{-\varepsilon^2 n} \quad \text{for } 0 < \varepsilon < \frac{1}{2}. \tag{4.35}$$

Theorem 4.6 tells us that for a random subspace H if $\varepsilon = c\sqrt{\alpha} = c\sqrt{k/n}$, then Euclidean ball is absorbed by the projection of the cube in (4.35):

$$\varepsilon\sqrt{n}P_H B_2^n \subset C\sqrt{\log\frac{1}{\varepsilon}}P_H B_\infty^n. \tag{4.36}$$

Hence for a random subspace H and for ε as above we have

$$\gamma_H\left(C\sqrt{\log\frac{1}{\varepsilon}}P_H B_\infty^n\right) \geq 1 - e^{-\varepsilon^2 n}, \tag{4.37}$$

which completes the proof. ∎

Coming back to (4.28), we will use **Proposition 4.4** for a random subspace H in the Grassmanian $G_{m-r,m-R}$. We conclude that if

$$c\sqrt{\frac{R}{r}} \geq C\sqrt{\log \frac{m-r}{R-r}},\tag{4.38}$$

then with probability at least $1 - e^{-cR}$ in the Grassmanian,

$$\gamma_H\left(c\sqrt{\frac{R}{r}}P_H B_\infty^{m-r}\right) \geq 1 - e^{-cR}.\tag{4.39}$$

Since $(m-r)/(R-r) \leq m/r$, the choice of R in (1.2) satisfies condition (4.38). Thus (4.28) implies

$$P \geq 1 - 3e^{-cR}.\tag{4.40}$$

This completes the proof of **Theorem 1.1**.

5   Optimality, robustness, and finite alphabets

5.1   Optimality

The logarithmic term in Theorems **1.1** and **2.1** is necessary, at least in the case of small r. Indeed, combining formula (4.17) and Lemmas **4.1**, **4.2**, **4.3**, we obtain

$$P \leq \int_{G_{m-r,m-R}} \gamma_H\left(c\sqrt{\frac{R}{r}}P_H B_\infty^{m-r}\right) d\nu(H) + 2e^{-cR}.\tag{5.1}$$

To estimate the Gaussian measure we need the following.

Lemma 5.1. Let $x_1, \ldots, x_s$ be vectors in $\mathbb{R}^s$. Then

$$\gamma_s\left(\sum_{j=1}^s \left[-x_j, x_j\right]\right) \leq \gamma_s\left(M \cdot B_\infty^s\right),\tag{5.2}$$

where $M = \max_{j=1,\ldots,s} \|x_j\|_2$. $\square$

The sum in the lemma is understood as the Minkowski sum of sets of vectors, $A + B = \{a + b \mid a \in A, \ b \in B\}$.

Proof. Let $F = \mathrm{span}(x_1, \ldots, x_{s-1})$ and let $V = F^\perp$. Let $v \in V$ be a unit vector. Set $Z = \sum_{j=1}^{s-1} [-x_j, x_j]$. Then

$$\gamma_s \left(\sum_{j=1}^{s} \left[-x_j, x_j \right] \right) = \int_V \gamma_F \left(\left(\sum_{j=1}^{s} \left[-x_j, x_j \right] - tv \right) \cap F \right) d\gamma_V(t)$$

$$= \int_{[-P_V x_s, P_V x_s]} \gamma_F \left(Z + t P_F x_s \right) d\gamma_V(t). \tag{5.3}$$

By Anderson's lemma (see [31]), $\gamma_F(Z + t P_F x_s) \leq \gamma_F(Z)$. Thus,

$$\gamma_s \left(\sum_{j=1}^{s} \left[-x_j, x_j \right] \right) \leq \gamma_V \left(\left[- P_V x_s, P_V x_s \right] \right) \cdot \gamma_F(Z) \leq \gamma_1 \left([-M, M] \right) \cdot \gamma_F(Z). \tag{5.4}$$

The proof of the lemma is completed by induction. ∎

The Gaussian measure of a projection of the cube can be estimated as follows.

Proposition 5.2. Let H be any subspace in $G_{n,n-k}, k < n/2$. Then

$$\gamma_H \left(\frac{c}{\sqrt{k}} \sqrt{\log \frac{n}{k}} P_H B_\infty^n \right) \leq e^{-cn/k}. \tag{5.5}$$

$\square$

Proof. Decompose I into the disjoint union of the sets $J_1, \ldots, J_{s+1}$, so that each of the sets $J_1, \ldots, J_s$ contains $k + 1$ elements and $(k + 1)s < n \leq (k + 1)(s + 1)$. Let $1 \leq j \leq s$. Let $U_j = H \cap (P_H e_i, \ i \in \{1, \ldots, n\} \setminus J_j)^\perp$, where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$. Then U_j is a one-dimensional subspace of H. Set

$$x_j = \sum_{i \in J_j} \varepsilon_i P_H e_i, \tag{5.6}$$

where the signs $\varepsilon_i \in \{-1, 1\}$ are chosen to maximize $\|P_{U_j} x_j\|_2$. Let $E = \mathrm{span}(x_1, \ldots, x_{s-1})$. Since $P_{U_j} B_\infty^n = [-x_j, x_j]$, we get

$$P_H B_\infty^n \cap E = \sum_{j=1}^{s} \left[-x_j, x_j \right], \tag{5.7}$$

where the sum is understood in the sense of Minkowski addition. Since $\|P_{U_j}\| = 1$, $\|x_j\|_2 \leq C\sqrt{k}$ and by Lemma 5.1,

$$\gamma_E \left(\frac{\bar{c}\sqrt{\log s}}{\sqrt{k}} \sum_{j=1}^{s} \left[-x_j, x_j \right] \right) \leq \gamma_E \left(c'\sqrt{\log s} \cdot B_\infty^E \right) \leq e^{-cs} \tag{5.8}$$

for some appropriately chosen constant $\bar{c}$. Finally, log-concavity of the Gaussian measure implies that for any convex symmetric body $K \subset H$,

$$\gamma_H(K) \leq \gamma_E(K \cap E). \tag{5.9}$$

∎

Combining (5.1) and (5.5) we obtain $P \leq 2e^{-cR}$, whenever $R \leq c \log(m/r)$.

5.2   Robustness and codes for finite alphabets

Robustness is a natural feature of the basis pursuit method. The solution to (BP) is stable with respect to the 1-norm in the same way as the solution to (MLS) is stable with respect to the 2-norm. Indeed, once Theorem 1.1 holds, the unknown vector y in Theorem 1.1 can be approximately recovered from $y'' = y' + h$, where $h \in \mathbb{R}^m$ is any additional error vector of small 1-norm. Namely, the solution u to the basis pursuit problem

$$\min_{u \in Y} \|u - y''\|_1 \tag{5.10}$$

satisfies

$$\|u - y\|_1 \leq C'\|h\|_1 \tag{5.11}$$

(see, e.g., [7]). This implies a possibility of quantization of the coefficients in the process of encoding and yields *robust error-correcting codes over alphabets of polynomial size*, *with a Gilbert-Varshamov-type bound*, *and with quadratic time encoders and polynomial time decoders*.

The following is the (m, n, r)-error-correcting code under the Gilbert-Varshamov-type assumption (1.2), with input words x over the alphabet $\{1, \ldots, p\}$ and the encoded words y over the alphabet $\{1, \ldots, Cpn^{3/2}\}$.

The construction is the same as in (1.1); we just have to introduce quantization. The encoder takes $x \in \{1, \ldots, p\}^n$, computes $y = Qx$, and outputs the $\widehat{y}$ whose coefficients are the coefficients of y quantized with the uniform quantizer with step $1/4C'm$.

Then $\widehat{y} \in (1/4C'm)\mathbb{Z}^m \cap [-p\sqrt{n}, p\sqrt{n}]^m$, which by rescaling can be identified with $\{1, \ldots, Cpn^{3/2}\}^m$ because we can assume that $m \leq 2n$. The decoder takes $y' \in (1/4C'm)\mathbb{Z}^m$, finds solution u to (BP) with $Y = \text{range}(Q)$, inverts to $x' = Q^T u$, and outputs $\widehat{x'}$ whose coefficients are coefficients of x' quantized with the uniform quantizer with step 1.

This is indeed an (m, n, r)-error-correcting code. If y' differs from $\widehat{y}$ on at most r coordinates, this and the condition $\|\widehat{y} - y\|_1 \leq 1/4C'$ imply by the robustness that $\|u - y\|_1 \leq 1/4$. Hence $\|x' - x\|_2 = \|Q^T(u - y)\|_2 = \|u - y\|_2 \leq \|u - y\|_1 \leq 1/4$. Thus $\widehat{x'} = x$, so the decoder recovers x from y' correctly.

The robustness also implies a "continuity" of our error-correcting codes. If the number of corrupted coordinates in the received message y' is bigger than r but is still a small fraction, then the (m, n, r)-error-correcting code above can still recover y up to some small fraction of the coordinates.

Acknowledgments

References

[1]   D. Amir and V. D. Milman, *Unconditional and symmetric sets in n-dimensional normed spaces*, Israel J. Math. **37** (1980), no. 1-2, 3–20.

[2]   S. Artstein, *Proportional concentration phenomena on the sphere*, Israel J. Math. **132** (2002), 337–358.

[3]   B. Beferull-Lozano and A. Ortega, *Efficient quantization for overcomplete expansions in $\mathbb{R}^N$*, IEEE Trans. Inform. Theory **49** (2003), no. 1, 129–150.

[4]   E. Candes and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, to appear in Found. Comput. Math.

[5]   E. Candes, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, preprint, 2004.

[6]   E. Candes and T. Tao, *Decoding by linear programming*, to appear in IEEE Trans. Inform. Theory.

[7]   ———, *Near optimal signal recovery from random projections: universal encoding strategies?*, preprint, 2004.

[8]   P. G. Casazza and J. Kovačević, *Equal-norm tight frames with erasures. Frames*, Adv. Comput. Math. **18** (2003), no. 2–4, 387–430.

[9]   S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. **20** (1998), no. 1, 33–61, reprinted in SIAM Rev. 43 (2001), no. 1, 129–159.

[10]   I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 61, SIAM, Pennsylvania, 1992.

[11]   D. L. Donoho, *Compressed sensing*, preprint, 2004.

[12]   ———, *For most large underdetermined systems of linear equations*, *the minimal ℓ^1-norm near-solution approximates the sparsest near-solution*, preprint, 2004.

[13]   ———, *For most large underdetermined systems of linear equations*, *the minimal ℓ^1-norm solution is also the sparsest solution*, preprint, 2004.

[14]   D. L. Donoho and M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization*, Proc. Nat. Acad. Sci. U.S.A. **100** (2003), no. 5, 2197–2202.

[15]   D. L. Donoho, M. Elad, and V. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, to appear in IEEE Trans. Inform. Theory.

[16]   D. L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory **47** (2001), no. 7, 2845–2862.

[17]   D. L. Donoho and J. Tanner, *Sparse nonnegative solution of underdetermined linear equations by linear programming*, Proc. Nat. Acad. Sci. U.S.A. **102** (2005), no. 27, 9446–9451.

[18]   D. L. Donoho and Y. Tsaig, *Breakdown of equivalence between the minimal ℓ^1-norm solution and the sparsest solution*, to appear in EURASIP Signal Processing Journal.

[19]   ———, *Extensions of compresed sensing*, preprint, 2004.

[20]   M. Elad and A. M. Bruckstein, *A generalized uncertainty principle and sparse representation in pairs of bases*, IEEE Trans. Inform. Theory **48** (2002), no. 9, 2558–2567.

[21]   A. Feuer and A. Nemirovski, *On sparse representation in pairs of bases*, IEEE Trans. Inform. Theory **49** (2003), no. 6, 1579–1581.

[22]   A. Yu. Garnaev and E. D. Gluskin, *The widths of a Euclidean ball*, Dokl. Akad. Nauk SSSR **277** (1984), no. 5, 1048–1052 (Russian), English translation: Soviet Math. Dokl. 30 (1984), 200–204.

[23]   V. K. Goyal, *Multiple description coding: compression meets the network*, IEEE Signal Processing Mag. **18** (2001), no. 5, 74–93.

[24]   ———, *Theoretical foundations of transform coding*, IEEE Signal Processing Mag. **18** (2001), no. 5, 9–21.

[25]   V. K. Goyal, J. Kovačević, and J. A. Kelner, *Quantized frame expansions with erasures*, Appl. Comput. Harmon. Anal. **10** (2001), no. 3, 203–233.

[26]   V. K. Goyal, M. Vetterli, and N. T. Thao, *Quantized overcomplete expansions in $\mathbf{R}^N$: analysis*, *synthesis*, *and algorithms*, IEEE Trans. Inform. Theory **44** (1998), no. 1, 16–31.

[27]   R. Gribonval and M. Nielsen, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory **49** (2003), no. 12, 3320–3325.

[28]   B. Grünbaum, *Convex Polytopes*, 2nd ed., Graduate Texts in Mathematics, vol. 221, Springer, New York, 2003.

[29] J. Kovačević, P. L. Dragotti, and V. K. Goyal, *Filter bank frame expansions with erasures*, IEEE Trans. Inform. Theory **48** (2002), no. 6, 1439–1450, Special issue on Shannon theory: perspective, trends, and applications.

[30] M. Ledoux, *The Concentration of Measure Phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Rhode Island, 2001.

[31] M. A. Lifshits, *Gaussian Random Functions*, Mathematics and Its Applications, vol. 322, Kluwer Academic, Dordrecht, 1995.

[32] J. Matoušek, *Lectures on Discrete Geometry*, Graduate Texts in Mathematics, vol. 212, Springer, New York, 2002.

[33] S. Mendelson, *Geometric parameters in learning theory*, Geometric Aspects of Functional Analysis, Lecture Notes in Math., vol. 1850, Springer, Berlin, 2004, pp. 193–235.

[34] V. S. Pless, W. C. Huffman, and R. A. Brualdi (eds.), *Handbook of Coding Theory. Vol. I*, *II*, North-Holland, Amsterdam, 1998.

[35] D. A. Spielman, *The complexity of error-correcting codes*, Fundamentals of Computation Theory (Kraków, 1997), Lecture Notes in Comput. Sci., vol. 1279, Springer, Berlin, 1997, pp. 67–84.

[36] ———, *Constructing error-correcting codes from expander graphs*, Emerging Applications of Number Theory (Minneapolis, Minn, 1996), IMA Vol. Math. Appl., vol. 109, Springer, New York, 1999, pp. 591–600.

[37] J. A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory **50** (2004), no. 10, 2231–2242.

[38] ———, *Just relax: convex programming methods for subset selection and sparse approximation*, ICES Report 04-04, University of Texas at Austin, Texas, 2004.

[39] ———, *Recovery of short*, *complex linear combinations via ℓ_1 minimization*, IEEE Trans. Inform. Theory **51** (2005), no. 4, 1568–1570.

[40] G. M. Ziegler, *Lectures on Polytopes*, Graduate Texts in Mathematics, vol. 152, Springer, New York, 1995.

Mark Rudelson: Department of Mathematics, University of Missouri, Columbia, MO 65211, USA
E-mail address: rudelson@math.missouri.edu

Roman Vershynin: Department of Mathematics, University of California, Davis, CA 95616, USA
E-mail address: vershynin@math.ucdavis.edu