Error Correction via Linear Programming

Emmanuel Candes
Applied and Computational Mathematics, Caltech,
Pasadena, CA 91125, USA
emmanuel@acm.caltech.edu

Mark Rudelson
Department of Mathematics
University of Missouri,
Columbia, MO 65203, USA
rudelson@math.missouri.edu

Terence Tao
Department of Mathematics
University of California,
Los Angeles, CA 90095, USA
tao@math.ucla.edu

Roman Vershynin
Department of Mathematics
University of California,
Davis, CA 9595616, USA
vershynin@math.ucdavis.edu

Abstract

Suppose we wish to transmit a vector $f \in \mathbb{R}^n$ reliably. A frequently discussed approach consists in encoding f with an m by n coding matrix A. Assume now that a fraction of the entries of Af are corrupted in a completely arbitrary fashion by an error e. We do not know which entries are affected nor do we know how they are affected. Is it possible to recover f exactly from the corrupted m-dimensional vector $y' = Af + e$?

This paper proves that under suitable conditions on the coding matrix A, the input f is the unique solution to the ℓ_1-minimization problem ($\|x\|_{\ell_1} := \sum_i |x_i|$)

$$\min_{\tilde{f} \in \mathbb{R}^n} \|y' - A\tilde{f}\|_{\ell_1}$$

provided that the fraction of corrupted entries is not too large, i.e. does not exceed some strictly positive constant ρ^* (numerical values for ρ^* are actually given). In other words, f can be recovered exactly by solving a simple convex optimization problem; in fact, a linear program. We report on numerical experiments suggesting that ℓ_1-minimization is amazingly effective; f is recovered exactly even in situations where a very significant fraction of the output is corrupted.

In the case when the measurement matrix A is Gaussian, the problem is equivalent to that of counting low-dimensional facets of a convex polytope, and in particular of a random section of the unit cube. In this case we can strengthen the results somewhat by using a geometric functional analysis approach.

Keywords. Linear codes, decoding of (random) linear codes, sparse solutions to underdetermined systems, ℓ_1-minimization, linear programming, restricted orthonormality, Gaussian random matrices.

1 Introduction

1.1 The error correction problem

This paper considers the model problem of recovering an input vector $f \in \mathbb{R}^n$ from corrupted measurements $y' = Af + e$. Here, A is an m by n matrix (we will assume throughout the paper that $m > n$), and $e \in \mathbb{R}^m$ is an unknown vector of errors. We will assume that at most r entries are corrupted, thus at most r entries of e are non-zero, but apart from this restriction e will be arbitrary. The problem we consider is whether it is possible to recover f exactly from the data y. And if so, how?

In its abstract form, our problem is of course equivalent to the classical error correcting problem which arises in coding theory as we may think of A as a *linear code*; a linear code is a given collection of codewords which are vectors $a_1, \ldots, a_n \in \mathbb{R}^m$—the columns of the matrix A. Given a vector $f \in \mathbb{R}^n$ (the "plaintext") we can then generate a vector Af in $\mathbb{R}^m$ (the "ciphertext"); if A has full rank, then one can clearly recover the plaintext f from the ciphertext Af. But now we suppose that the ciphertext Af is corrupted by an arbitrary vector $e \in \mathbb{R}^m$ with at most r non-zero entries so that the corrupted ciphertext is of the form $Af + e$. The question is then: given the coding matrix A and $Af + e$, can one recover f exactly?

Let us say that the linear code A is a (m, n, r)-*error correcting code* if one can recover f from $Af + e$ whenever

e has at most r non-zero coefficients. As is well-known, if the number r of corrupted entries is too large, then of course we have no hope of having a (m, n, r)-error correcting code. For instance, if $n + 2r > m$ then elementary linear algebra shows that there exist plaintexts $f, f' \in \mathbb{R}^n$ and errors $e, e' \in \mathbb{R}^m$ with at most r non-zero coefficients each such that $Af + e = Af' + e'$, and so one cannot distinguish f from f' in this case. In particular, if the fraction $\rho := \frac{r}{m}$ of corrupted entries exceeds $1/2$ then an (m, n, r)-error correcting code is impossible regardless of how large one makes m with respect to n.

This situation raises an important question: for which fraction ρ of the corrupted entries is accurate decoding possible with practical algorithms? That is, with algorithms whose complexity is at most polynomial in the length m of the codewords?

Setting $y := Af$, and letting $Y \subset \mathbb{R}^m$ be the image of A, we can rephrase the problem more geometrically as follows: *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?*

If the matrix A is chosen randomly, for instance by the Gaussian ensemble, then it is easy to show that with probability one that all the plaintexts can be distinguished in an information-theoretic sense as soon as $n + 2r \leq m$. However, this result provides no algorithm for recovering the plaintext f from the corrupted ciphertext $Af + e$, other than brute force search, which has exponential complexity in m. Based on analogy with discrete (e.g. finite field) analogues of this problem, to obtain a polynomial-time recovery algorithm it is more reasonable to expect as a necessary condition the *Gilbert-Varshamov bound*

$$n/m \geq 1 - H(Cr/n) \tag{1.1}$$

which is fundamental in coding theory (see [34]); here $H(x)$ is the entropy function, and C, c, c_1, etc. will be used to denote various positive absolute constants. This heuristic can be made rigorous if one requires a certain stability property for the recovery algorithm; see Section 6.

One can instead consider a mean square approach, based on the minimization problem

$$(P_2) \qquad \min_{\tilde{f} \in \mathbb{R}^n} \|y' - A\tilde{f}\|_{\ell_2}$$

or equivalently

$$(P_2') \qquad \min_{\tilde{y} \in Y} \|y' - \tilde{y}\|_{\ell_2}$$

but the minimizer $f^\star$ (resp. $y^\star$) may be arbitrarily far away from the plaintext f (resp. y) since we have no size control on the error e.

1.2 Solution via ℓ_1-minimization

To recover f accurately from corrupted data $y' = Af + e$, we consider solving the following ℓ_1-minimization (or *Basis Pursuit*) problem

$$(P_1) \qquad \min_{\tilde{f} \in \mathbb{R}^n} \|y' - A\tilde{f}\|_{\ell_1} \tag{1.2}$$

or equivalently

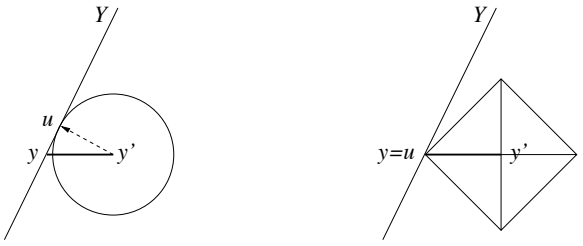
$$(P_1') \qquad \min_{\tilde{y} \in \mathbb{R}^n} \|y' - y\|_{\ell_1}$$

Thus the minimizer $y^\star$ to (P_1') is the metric projection of y' onto the vector space Y with respect to the ℓ_1 norm. This is a convex program which can be classically reformulated as a linear program. Indeed, the ℓ_1-minimization problem (P_1) is equivalent to

$$\min \sum_{i=1}^{m} t_i, \qquad -t \leq y' - A\tilde{f} \leq t,$$

where the optimization variables are $t \in R^m$ and $\tilde{f} \in \mathbb{R}^n$ (as is standard, the generalized vector inequality $x \leq y$ means that $x_i \leq y_i$ for every coordinate i). Hence, (P_1) is an LP with inequality constraints and can be solved efficiently using standard or specialized optimization algorithms, see [4], [5].

The main claim of this paper is that for suitable coding matrices A, the solution $f^\star$ to our linear program is actually exact; $f^\star = f$! (Equivalently, $y^\star = y$.)



(MLS) (BP)

The potential of Basis Pursuit for exact reconstruction is illustrated by the following heuristics, essentially due to [18]. The minimizer u to (P_2') is the contact point where the smallest Euclidean ball centered at y' meets the subspace Y. That contact point is in general different from y. The situation is much better in (P_1'): typically the solution coincides with y. The minimizer u to (P_1') is the contact point where the smallest octahedron centered at y' (the ball with respect to the 1-norm) meets Y. Because the vector $y - y'$ lies in a low-dimensional coordinate subspace, the octahedron has a wedge at y. 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 (P_1'), 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 [47]). With the subsequent development of fast interior point methods in Linear Programming, (P_1) turned into an effectively solvable problem, and was put forward more recently by Donoho and his collaborators, triggering massive experimental and theoretical work [5, 7, 8, 10, 14, 15, 17–21, 23, 27, 33, 45–47].

We shall rigorously validate the above heuristics in two slightly different ways. First we present a deterministic and axiomatic approach, in which we assume a certain *restricted isometry condition* on the measurement matrix A, and deduce that the minimizer $f^\star$ to (P_1) equals f exactly. Then we present a more geometric functional analysis approach, in which A is a random Gaussian matrix, and establish the claim as a consequence of geometric facts about the facets of random sections of the unit cube.

1.3 Restricted isometry matrices

We begin by introducing the restricted isometry condition. Consider a fixed p by m matrix B and let B_T, $T \subset \{1, \dots, m\}$ be the $p \times |T|$ submatrix obtained by extracting the columns of B corresponding to the indices in T. Then [11] defines the S-restricted isometry constant δ_S of B which is the smallest quantity such that

$$(1 - \delta_S) \|c\|_{\ell_2}^2 \leq \|B_T c\|_{\ell_2}^2 \leq (1 + \delta_S) \|c\|_{\ell_2}^2 \qquad (1.3)$$

for all subsets T with $|T| \leq S$ and coefficient sequences $(c_j)_{j \in T}$. This property essentially requires that every set of columns with cardinality less than S approximately behaves like an orthonormal system.

Let us return to our error correction problem, and consider a matrix B whose kernel equals the range of A, so in particular $BA = 0$ (B is any $(m - n) \times n$ matrix whose kernel is the range of A in $\mathbb{R}^m$). Apply B on both sides of the equation $y' = Af + e$, and obtain

$$By' = B(Af + e) = Be \qquad (1.4)$$

since $BA = 0$. Therefore, the decoding problem is reduced to that of recovering the error vector e from the known vector $Be = By'$. Once e is known, $Af = y' - e$ is known and, therefore, f is also known since A has full rank.

To solve the underdetermined system of linear equations $Be = By'$, we search among all vector $\tilde{e} \in \mathbb{R}^m$ obeying $B\tilde{e} = By'$ for that with minimum ℓ_1-norm

$$(P_1'') \qquad \min_{\tilde{e} \in \mathbb{R}^m} \|\tilde{e}\|_{\ell_1}, \qquad B\tilde{e} = By', \qquad (1.5)$$

This convex program (P_1'') is easily seen to be equivalent to (P_1) or (P_1'), and may be recast as an LP.

We now state the first main result of this paper, which we prove in Section 2.

Theorem 1.1. *Let $f \in \mathbb{R}^n$, let A be an $m \times n$ matrix, let e have at most r non-zero entries, let $y' = Af + e$, and let B be a matrix whose kernel equals the range of A. Suppose we also have the condition*

$$\delta_{3r} + 3\delta_{4r} < 2, \qquad (1.6)$$

and let $e \in \mathbb{R}^m$ have at most r entries non-zero. Then the solution to (P_1'') (resp. (P_1), (P_1')) is unique and equal to e (resp. f, y). In particular, A is a (m, n, r) error-correcting code, with (P_1) as exact recovery algorithm.

This last theorem claims, perhaps, a rather surprising result. In effect, it says that minimizing ℓ_1 recovers *all* input signals $f \in \mathbb{R}^n$ regardless of the corruption patterns, provided of course that the support of the error vector is not too large. In particular, one can introduce errors of arbitrary large sizes and still recover the input vector f exactly, by solving a convenient linear program; in other words, as long as the fraction of corrupted entries is not too large, there is nothing a malevolent adversary can do to corrupt Af as to fool the simple decoding strategy (1.2).

1.4 The Gaussian ensemble

For Theorem 1.1 to be of real interest, one should use matrices B with good restricted isometry constants δ_S; that is, such that the condition of Theorem 1.1 holds with large values of S. How to design such matrices is a delicate question, and we do not know of any matrix which provably obeys (1.6) for interesting values of S. However, if we simply sample a matrix B with i.i.d. entries, it will obey (1.6) for large values of S with overwhelming probability. For instance, by using concentration of measure inequalities and the Marchenko-Pastur law [39] as in [11, Theorem 1.6] (see also [22, 35, 44] for some relevant results) one can establish

Theorem 1.2. *Assume $n < m$, let $p := m - n$, and let B be a p by m matrix whose entries are i.i.d. Gaussian with mean zero and variance $1/p$. Then the condition of Theorem 1.1 holds with probability at least $1 - O(e^{-\alpha m})$ for some fixed constant $\alpha > 0$, provided that $r \leq \rho^* m$, where ρ^* depends only on the ratio n/m. For large values of n and m, one can show that $\rho^* \geq 1/3,000$ for $m = 2n$, and $\rho^* \geq 1/2,000$ for $m = 4n$.*

In particular, we see that a n by m Gaussian matrix will be a (m, n, r)-error correcting code with high probability, as long as the fraction $\rho = \frac{r}{m}$ of corrupted entries is less than a constant ρ^* depending only on n/m. This is because

the annihilator B of a random Gaussian matrix can be chosen to be another random Gaussian matrix.

Similar statements with different constants hold for other types of ensembles, e.g. for binary matrices with i.i.d. entries taking values $\pm 1/\sqrt{p}$ with probability 1/2. It is interesting that our methods actually give numerical values, instead of the traditional "for some positive constant ρ." However, the numerical bounds we derived in this paper are overly pessimistic. We are confident that finer arguments and perhaps new ideas will allow to derive versions of Theorem 1.2 with better bounds. Numerical experiments actually suggests that the threshold is indeed much higher, see Section 5.

Returning to the Gaussian case, it turns out that when r is somewhat small then we can come close to the theoretical limit $n + 2r \leq m$. More precisely, in Section 3 we will prove

Theorem 1.3. *Let m, n and $r < cm$ be positive integers such that*

$$m = n + R, \quad \textit{where } R \geq Cr \log(m/r). \tag{1.7}$$

Let G be an $m \times n$ matrix whose entries are independent $N(0,1)$ normal random variables. Then, with probability at least $1 - e^{-cR}$, the matrix G is an (m, n, r) error-correcting code with exact recovery algorithm (P_1').

The assymption (1.7) meets, up to a constant, the Gilbert-Varshamov bound (1.1), and can be rephrased in terms of the corruption rate $\rho = r/m$ as $m \geq (1 + C\rho \log \frac{1}{\rho})n$. Theorem 1.3 then asserts that $m \times n$ Gaussian matrices will be an (m, n, r)-error correcting code with high probability once this condition is attained.

In the signal processing, linear codes 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, etc. This is confirmed by an extensive experimental and some theoretical work, see e.g. [3, 6, 13, 29–32, 36] and the bibliography contained therein. Theorem 1.3 thus states that *most orthogonal transform codes are good error-correcting codes*.

Acknowledgements.

E. C. is partially supported in part by a National Science Foundation grant DMS 01-40698 (FRG) and by an Alfred P. Sloan Fellowship. M. R. is partially supported by the NSF grant DMS 0245380. T. T. is supported by a grant from the Packard Foundation. R. V. is an Alfred P. Sloan Research Fellow He was also partially supported by the NSF grant DMS 0401032 and by the Miller Scholarship from the University of Missouri-Columbia. R. V. is grateful to University of Missouri for their hospitality during this period, when part of this research was started. E. C. and T. T. would like to thank Rafail Ostrovsky for pointing out possible connections between their earlier work and the error correction problem. Parts of this paper are an abridged version of [11] and [43].

2 Proof of Theorem 1.1

The proof of the theorem makes use of two geometrical special facts about the solution $d^\star$ to (P_1'). First, $Bd^\star = By'$ which geometrically says that $d^\star$ belongs to a known plane of co-dimension p where p. Second, because e is feasible, we must have $\|d^\star\|_{\ell_1} \leq \|e\|_{\ell_1}$. Decompose $d^\star$ as $d^\star = e + h$, thus $Bh = 0$. As observed in [19]

$$\|e\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} \leq \|e + h\|_{\ell_1} \leq \|e\|_{\ell_1},$$

where T_0 is the support of e, and $h_{T_0}(t) = h(t)$ for $t \in T_0$ and zero elsewhere (similarly for $h_{T_0^c}$). Hence, h obeys the cone constraint

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} \tag{2.1}$$

which expresses the geometric idea that h must lie in the cone of descent of the ℓ_1-norm at e. Exact recovery occurs provided that the null vector is the only point in the intersection between $\{h : Bh = 0\}$ and the set of h obeying (2.1).

We begin by dividing T_0^c into subsets of size M (we will choose M later) and enumerate T_0^c as

$$n_1, n_2, \ldots, n_{m-|T_0|}$$

in decreasing order of magnitude of $h_{T_0^c}$. Set $T_j = \{n_\ell, (j-1)M + 1 \leq \ell \leq jM\}$. That is, T_1 contains the indices of the M largest coefficients of $h_{T_0^c}$, T_2 contains the indices of the next M largest coefficients, and so on.

With this decomposition, the ℓ_2-norm of h is concentrated on $T_{01} = T_0 \cup T_1$. Indeed, the kth largest value of $h_{T_0^c}$ obeys

$$|h_{T_0^c}|_{(k)} \ \leq \ \|h_{T_0^c}\|_{\ell_1}/k$$

and, therefore,

$$\|h_{T_{01}^c}\|_{\ell_2}^2 \ \leq \ \|h_{T_0^c}\|_{\ell_1}^2 \sum_{k=M+1}^{m} 1/k^2 \ \leq \ \|h_{T_0^c}\|_{\ell_1}^2/M.$$

Further, the ℓ_1-cone constraint gives

$$\|h_{T_{01}^c}\|_{\ell_2}^2 \ \leq \ \|h_{T_0}\|_{\ell_1}^2/M \ \leq \ \|h_{T_0}\|_{\ell_2}^2 \cdot |T_0|/M$$

and thus

$$\begin{aligned} \|h\|_{\ell_2}^2 &= \|h_{T_{01}}\|_{\ell_2}^2 + \|h_{T_{01}^c}\|_{\ell_2}^2 \\ &\leq (1 + |T_0|/M) \cdot \|h_{T_{01}}\|_{\ell_2}^2. \end{aligned} \tag{2.2}$$

Observe now that

$$\begin{aligned}\|Bh\|_{\ell_2} &= \|B_{T_{01}} h_{T_{01}} + \sum_{j\geq 2} B_{T_j} h_{T_j}\|_{\ell_2} \\ &\geq \|B_{T_{01}} h_{T_{01}}\|_{\ell_2} - \|\sum_{j\geq 2} B_{T_j} h_{T_j}\|_{\ell_2} \\ &\geq \|B_{T_{01}} h_{T_{01}}\|_{\ell_2} - \sum_{j\geq 2} \|B_{T_j} h_{T_j}\|_{\ell_2} \\ &\geq \sqrt{1-\delta_{M+|T_0|}}\,\|h_{T_{01}}\|_{\ell_2} - \\ &\quad \sqrt{1+\delta_M}\sum_{j\geq 2}\|h_{T_j}\|_{\ell_2}.\end{aligned}$$

Set $\rho_M = |T_0|/M$. As we shall see later,

$$\sum_{j\geq 2}\|h_{T_j}\|_{\ell_2} \leq \sqrt{\rho_M}\cdot\|h_{T_0}\|_{\ell_2}, \tag{2.3}$$

and since $Bh = 0$, this gives

$$[\sqrt{1-\delta_{M+|T_0|}} - \sqrt{\rho_M}\,\sqrt{1+\delta_M}]\cdot\|h_{T_{01}}\|_{\ell_2} \leq 0. \tag{2.4}$$

It then follows from (2.2) that $h = 0$ provided that the quantity $\sqrt{1-\delta_{M+|T_0|}} - \sqrt{\rho_M}\,\sqrt{1+\delta_M}$ is positive. Take $M = 3|T_0|$ for example. Then this quantity is positive if $\delta_{3|T_0|} + 3\delta_{4|T_0|} < 2$. Since $|T_0| \leq r$, this follows from (1.6).

It remains to argue about (2.3). Observe that by construction, the magnitude of each coefficient in T_{j+1} is less than the average of the magnitudes in T_j:

$$|h_{T_{j+1}}(t)| \leq \|h_{T_j}\|_{\ell_1}/M.$$

Then

$$\|h_{T_{j+1}}\|_{\ell_2}^2 \leq \|h_{T_j}\|_{\ell_1}^2/M$$

and (2.3) follows from

$$\begin{aligned}\sum_{j\geq 2}\|h_{T_j}\|_{\ell_2} &\leq \sum_{j\geq 1}\|h_{T_j}\|_{\ell_1}/\sqrt{M} \\ &\leq \|h_{T_0}\|_{\ell_1}/\sqrt{M} \leq \sqrt{|T_0|/M}\cdot\|h_{T_0}\|_{\ell_2}.\end{aligned}$$

3 Proof of Theorem 1.3

3.1 Low-dimensional facets of polytopes.

Theorem 1.3 turns out to be equivaent 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 $B_\infty^m := [-1,1]^m$. Note that the range of the matrix G is an n-dimensional subspace uniformly distributed over the Grassmannian. Thus the conclusion of Theorem 1.3 can be reformulated as follows. Let $y \in Y$ be an unknown vector, and we are given a vector y' in $\mathbb{R}^m$ that differs from y on at most r coordinates. Then y can be exactly reconstructed from y' as the solution to the minimization problem (P_1'). This means that 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 the 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[1], 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 1.3. We can thus state Theorem 1.3 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 $1 - e^{-cR}$.*

3.2 Notation.

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 in 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.

3.3 Duality.

The proof of Theorem 1.3 begins with a typical duality argument, leading to the same reformulation of the prob-

[1] 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.

lem as in [10]. The conclusion of Theorem 1.3 is actually equivalent to the fact that Y forms a tangent space to the unit octahedron at all points whose support size is r (see the picture on p.2):

$$(z+Y)\cap \text{ interior}\,(B_1^m)=\emptyset \quad \text{for all } z\in \bigcup_{|I|=r} B_1^I.$$

By Hahn-Banach theorem, this separation is equivalent to the following (denoting $E=Y^\perp$):

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, since the total number of the facets is $N=2^r\binom{m}{r}$, the probability that E misses some facet would be at most $Ne^{-cR}\le e^{-c_1R}$ with an appropriate choice of the absolute constant in (1.7).

3.4 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. Without loss of generality, we can assume that our facet is

$$F=\{(w_1,\dots,w_{m-r},1,\dots,1),\ \text{ all } |w_j|\le 1\},$$

whose center is $\theta=(0,\dots,0,1,\dots,1)$ (with $m-r$ zeroes). We are interested in is

$$P:=\mathbb{P}\{E\cap F\ne\emptyset\}.$$

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

$$\text{lin}(F):=\{(w_1,\dots,w_{m-r},t,\dots,t): t,w_1,\dots,w_{m-r}\in\mathbb{R}\},$$

and even to its the affine span

$$\text{aff}(F):=\{(w_1,\dots,w_{m-r},1,\dots,1): w_1,\dots,w_{m-r}\in\mathbb{R}\}.$$

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

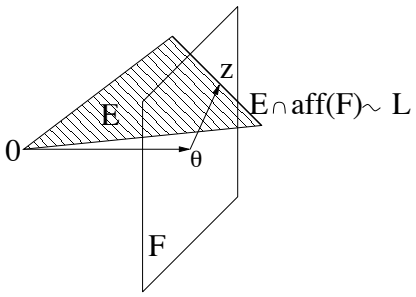
$$P=\mathbb{P}\Big\{(E\cap\text{aff}(F))\cap F\ne\emptyset\Big\}.$$

The dimension of that affine subspace is almost surely

$$l:=\dim(E\cap\text{aff}(F))=R-r.$$

We can realize the random affine subspace $E\cap\text{aff}(F)$ (or rather a random subspace with the same law) by the following algorithm:

1. Select a random variable D with the same law as $\text{dist}(\theta, E\cap\text{aff}(F))$.
2. Select a random subspace L_0 in the Grassmanian $G_{m-r,l}$. It will realize the "direction" of $E\cap\text{aff}(F)$ in $\text{aff}(F)$.
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 $E\cap\text{aff}(F)$ to the center θ of F.
4. Set $L=\theta+z+L_0$. Thus the random affine subspace L has the same law as $E\cap\text{aff}(F)$.



Hence

$$\begin{aligned}P&=\mathbb{P}\{L\cap F\ne\emptyset\}=\mathbb{P}\{(z+L_0)\cap B_\infty^{m-r}\ne\emptyset\}\\&=\mathbb{P}\{z\in P_{L_0^\perp}B_\infty^{m-r}\}.\end{aligned}$$

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

$$P=\int_{\mathbb{R}^+}\int_{G_{m-r,m-R}}\sigma_H(D^{-1}P_HB_\infty^{m-r})\,d\nu(H)\,d\mu(D) \tag{3.1}$$

where ν is the normalized Haar measure on $G_{m-r,m-R}$ and μ is the law of D. We shall 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 $P_HB_\infty^{m-r}$ has an almost full Gaussian (thus also spherical) measure.

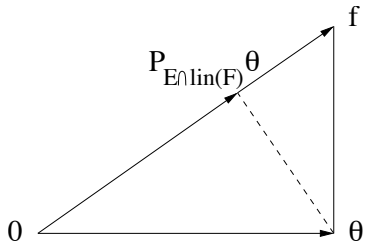
3.5 The distance D from the center of the facet to a random subspace

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

Lemma 3.2.

$$\|P_{E\cap\text{lin}(F)}\theta\|_2=\sqrt{\frac{r}{r+D^2}}\,\|\theta\|_2.$$

Proof. Let f be the multiple of the vector $P_{E\cap\mathrm{lin}(F)}\theta$ such that $f-\theta$ is orthogonal to θ. Such a multiple exists and is unique, as this is a two-dimensional problem.



Then $f \in E \cap \mathrm{aff}(F)$. Notice that $D = \|f-\theta\|_2$. 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

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

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

The length of the projection of a fixed vector onto a random subspace in Lemma 3.2 is well known. The asymptotically sharp estimate was computed by S. Artstein [1], but we will be satisfied with a much weaker elementary estimate, see e.g. [40, Theorem 15.2.2].

Lemma 3.3. *Let $\theta\in\mathbb{R}^{d-1}$ and let G be a random subspace in $G_{d,k}$. Then*

$$\mathbb{P}\Big\{c\sqrt{\frac{k}{d}}\,\|\theta\|_2 \le \|P_G\theta\|_2 \le C\sqrt{\frac{k}{d}}\,\|\theta\|_2\Big\} \ge 1-2e^{-ck}.$$

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

$$\mathbb{P}\Big\{\|P_{E\cap\mathrm{lin}(F)}\theta\|_2 \ge c\sqrt{\frac{l+1}{m-r+1}}\,\|\theta\|_2\Big\} \ge 1-2e^{-cl}.$$

Together with Lemma 3.2 this gives

$$\mathbb{P}\Big\{D \le c\sqrt{m-r}\sqrt{\frac{r}{l}}\Big\} \ge 1-2e^{-cl}. \qquad (3.2)$$

Note that $\sqrt{m-r}$ is the radius of the Euclidean ball circumscribed on the facet F. The statement $D\le\sqrt{m-r}$ would only tell us that the random subspace E intersects the circumscribed ball, not yet the facet itself. The ratio r/l in (3.2) will be chosen logarithmically small, which will force E intersect also the facet F.

3.6 Gaussian measure of random projections of the cube

By (3.1) and (3.2),

$$P \ge \int\limits_{G_{m-r,m-R}} \sigma_H\Big(\frac{c}{\sqrt{m-r}}\sqrt{\frac{l}{r}}\,P_H B_\infty^{m-r}\Big)\,d\nu(H) - 2e^{-cl}.$$

We can replace the spherical measure σ_H by the Gaussian measure γ_H via a simple lemma:

Lemma 3.4. *Let K be a star-shaped set in $\mathbb{R}^d$. Then*

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

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 non-increasing function of t. Hence

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

and

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

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

Using Lemma 3.4 in the space H of dimension $d=m-R$, we obtain

$$P \ge \int\limits_{G_{m-r,m-R}} \gamma_H\Big(c\sqrt{\frac{m-R}{m-r}}\sqrt{\frac{l}{r}}\,P_H B_\infty^{m-r}\Big)\,d\nu(H) - 2e^{-cl} - e^{m-R}.$$

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

$$P \ge \int\limits_{G_{m-r,m-R}} \gamma_H\Big(c\sqrt{\frac{R}{r}}\,P_H B_\infty^{m-r}\Big)\,d\nu(H) - 2e^{-cR}. \qquad (3.3)$$

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

Proposition 3.5. *Let H be a random subspace in $G_{n,n-k}$, $k < n/2$. Then the inequality*

$$\gamma_H\Big(C\sqrt{\log\frac{n}{k}}\,P_H B_\infty^n\Big) \geq 1 - e^{-ck}$$

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

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 3.6 (Concentration of Gaussian measure)**.** *Let $\varepsilon > 0$ and let $A \subset \mathbb{R}^n$ be a measurable set such that $\gamma_n(A) \geq e^{-\varepsilon^2 n}$. Then*

$$\gamma_n(A + C\varepsilon\sqrt{n}B_2^n) \geq 1 - e^{-\varepsilon^2 n}.$$

With the stronger assumption $\gamma(A) \geq 1/2$, this lemma is the classical concentration inequality, see [37] 1.1. The fact that the concentration holds also for exponentially small sets follows formally by a simple extension argument that was first noticed by D. Amir and V. Milman in [2], see [37] Lemma 1.1.

The optimal result on random projections of the cube is due to Garnaev and Gluskin [28].

Theorem 3.7 (Euclidean projections of the cube [28])**.** *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, we have*

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

where

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

Proof of Proposition 3.5. Let $g_1, g_2, \ldots$ be independent standard Gaussian random variables. Then for a suitable positive absolute constant c and for every $0 < \varepsilon < 1/2$,

$$\begin{aligned}\gamma_n\Big(C\sqrt{\log\frac{1}{\varepsilon}}\,B_\infty^n\Big) &= \mathbb{P}\Big\{\max_{1\leq j\leq n}|g_i| \leq C\sqrt{\log\frac{1}{\varepsilon}}\Big\}\\ &\geq (1-\varepsilon^2/10)^n \geq e^{-\varepsilon^2 n}.\end{aligned}$$

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

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

Then by Lemma 3.6,

$$\gamma_H\Big(C\sqrt{\log\frac{1}{\varepsilon}}\,P_H B_\infty^n + C\varepsilon\sqrt{n}\,P_H B_2^n\Big) \geq 1 - e^{-\varepsilon^2 n} \tag{3.4}$$

for $0 < \varepsilon < 1/2$. Theorem 3.7 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 (3.4):

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

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

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

which completes the proof. □

Coming back to (3.3), we shall use Lemma 3.5 for a random subspace H in the Grassmanian $G_{m-r,m-R}$. We conclude that if

$$c\sqrt{\frac{R}{r}} \geq C\sqrt{\log\frac{m-r}{R-r}}, \tag{3.5}$$

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

$$\gamma_H\Big(c\sqrt{\frac{R}{r}}\,P_H B_\infty^{m-r}\Big) \geq 1 - e^{-cR}.$$

Since $\frac{m-r}{R-r} \leq \frac{m}{r}$, the choice of R in (1.7) satisfies condition (3.5). Thus (3.3) implies

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

This completes the proof. ■

3.7 Optimality

The logarithmic term in Theorems 1.3 and 4.1 is necessary, at least in the case of small r. Indeed, combining formula (3.1) and Lemmas 3.2, 3.3, 3.4, we obtain

$$P \leq \int\limits_{G_{m-r,m-R}} \gamma_H\Big(c\sqrt{\frac{R}{r}}\,P_H B_\infty^{m-r}\Big)\,d\nu(H) + 2e^{-cR}. \tag{3.6}$$

To estimate the Gaussian measure we need the following

Lemma 3.8. *Let $x_1, \ldots x_s$ be vectors in $\mathbb{R}^s$. Then*

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

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

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 $F = \mathrm{span}(x_1, \dots x_{s-1})$ 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

$$\begin{aligned}
&\gamma_s\Big(\sum_{j=1}^{s}[-x_j, x_j]\Big) \\
&= \int_V \gamma_F\Big(\Big(\sum_{j=1}^{s}[-x_j, x_j] - tv\Big) \cap F\Big)\, d\gamma_V(t) \\
&= \int_{[-P_V x_s, P_V x_s]} \gamma_F(Z + tP_F x_s) d\gamma_V(t).
\end{aligned}$$

By Anderson's Lemma (see [38]), $\gamma_F(Z + tP_F x_s) \le \gamma_F(Z)$. Thus,

$$\begin{aligned}
\gamma_s\Big(\sum_{j=1}^{s}[-x_j, x_j]\Big) &\le \gamma_V([-P_V x_s, P_V x_s]) \cdot \gamma_F(Z) \\
&\le \gamma_1([-M, M]) \cdot \gamma_F(Z).
\end{aligned}$$

The proof of the Lemma is completed by induction. □

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

Proposition 3.9. *Let H be any subspace in $G_{n,n-k}$, $k < n/2$. Then*

$$\gamma_H\Big(\frac{c}{\sqrt{k}}\sqrt{\log\frac{n}{k}}\, P_H B_\infty^n\Big) \le e^{-cn/k}. \qquad (3.7)$$

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

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

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

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

where the sum is understood in the sense of Minkowski addition. Since $\|P_{U_j}\| = 1$, $\|x_j\|_2 \le C\sqrt{k}$ and by Lemma 3.8,

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

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

$$\gamma_H(K) \le \gamma_E(K \cap E). \qquad \square$$

Combining (3.6) and (3.7) we obtain $P \le 2e^{-cR}$, whenever $R \le c\log(m/r)$.

4 Reconstruction of signals from linear measurements.

The heuristic idea that guides Statistical Learning Theory 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 [41]).

We shall 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 seem to belong to small classes, as they carry much of unwanted information that can be discarded. Donoho [16] then questions the conventional scheme of signal processing, where the whole signal must first be acquired and only then compressed. Instead, can one *directly acquire* ('sense') the essential part of the signal, via few linear measurements? Similar issues are raised in [10]. We shall 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 $X_1, X_2, \dots, X_R$ in $\mathbb{R}^m$. The discussion in the introduction suggests to reconstruct f as a solution to the Basis Pursuit minimization problem

$$(BP) \quad \min \|g\|_1 \text{ subject to } \langle g, X_k\rangle = \langle f, X_k\rangle \ \ \forall k.$$

4.1 Functions with small support

In the class of functions with small support, one can hope for exact reconstruction. In previous work [10], two of the authors showed that every *fixed* function f with support $|\mathrm{supp} f| \le r$ can indeed be recovered by (BP), correctly with the polynomial probability $1 - m^{-\mathrm{const}}$, 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 1.3 does yield a uniform exact reconstruction. It provides us with *one set* of linear measurements from from which we can effectively reconstruct *every* signal of small support.

Theorem 4.1 (Uniform Exact Reconstruction)**.** *Let m, $r < cm$ and R be positive integers satisfying $R \geq Cr \log(m/r)$. 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, $|\mathrm{supp} f| \leq r$, and we are given R measurements $\langle f, X_k \rangle$. Then f can be exactly reconstructed from these measurements as a solution to the Basis Pursuit problem (BP).*

This theorem gives uniformity in [10], improves the polynomial probability to an exponential probability, and improves upon the number R of measurements (which was $R \geq Cr \log m$ in [10]). Donoho [16] proved a weaker form of Theorem 4.1 with R/r bounded below by some function of m/r.

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

$$\min \|u - f\|_1 \ \text{ subject to } \ \langle u, X_k \rangle = 0, \ \ \forall k.$$

The constraints here define a random $(n = m - R)$-dimensional subspace Y of $\mathbb{R}^m$. Now apply Theorem 1.3 with $y = 0$ and $y' = f$. It states that the unique solution to the minimization problem above is $u = 0$. Therefore, the unique solution to (BP′) is f. □

4.2 Compressible functions

In a larger class of compressible functions [16], 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:

$$f^*(s) \leq s^{-1/p}, \quad s = 1, \ldots, m \tag{4.1}$$

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 [10]), Theorem 4.1 implies, by the argument of [10], that functions compressible in some basis can be approximately reconstructed from few fixed linear measurements. This is an improvement of a result of Donoho [16].

Corollary 4.2 (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* (4.1) *for some $0 < p < 1$ or $\|f\|_1 \leq 1$ for $p = 1$. Suppose that we are given R measurements $\langle f, X_k \rangle$. Then f can be approximately reconstructed from these measurements: a unique solution g to the Basis Pursuit problem (BP) satisfies*

$$\|f - g\|_2 \leq C_p \Big(\frac{\log(m/R)}{R} \Big)^{\frac{1}{p} - \frac{1}{2}}$$

where C_p depends on p only.

Corollary 4.2 was proved by Donoho [16] 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. In [10] Corollary 4.2 was proven without the uniformity in f due to a weaker (polynomial) probability. Finally, Corollary 4.2 also improves upon the approximation error (there is now the ratio m/r instead of m in the logarithm).

5 Numerical Experiments

In this section, we empirically investigates the performance of our decoding strategy. Of special interest is the location of the breakpoint beyond which ℓ_1 fails to decode accurately. To study this issue, we performed a first series of experiments as follows:

1. select n (the size of the input signal) and m so that with the same notations as before, A is an m by n matrix; sample A with independent Gaussian entries and select the plaintext f at random;

2. select S as a percentage of m;

3. select a support set T of size $|T| = S$ uniformly at random, and sample a vector e on T with independent and identically distributed Gaussian entries, and with standard deviation about that of the coordinates of the output (Af) (the errors are then quite large compared to the "clean" coordinates of Af)[2];

4. make $\tilde{y} = Af + e$, solve (P_1) and obtain $f^\star$; compare f to $f^\star$;

5. repeat 100 times for each S, and for various sizes of n and m.

[2]The results presented here do not seem to depend on the actual distribution used to sample the errors.

The results are presented in Figure 1. In these experiments, we choose $n = 128$, and set $m = 2n$ (Figure 1(a)) or $m = 4n$ (Figure 1(b)). Our experiments show that the linear program recovers the input vector *all the time* as long as the fraction of the corrupted entries is less or equal to 15% in the case where $m = 2n$ and less or equal to 35% in the case where $m = 4n$. We repeated these experiments for different values of n, e.g. $n = 256$ and obtained very similar recovery curves.

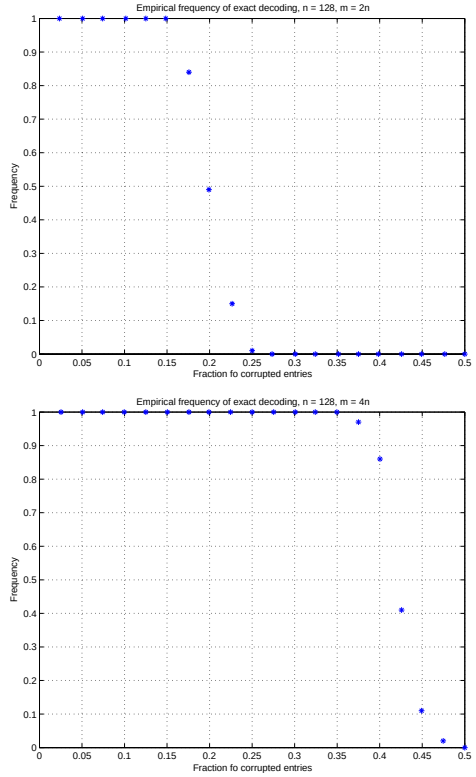


Figure 1. ℓ_1-recovery of an input signal from $y' = Af + e$ with A an m by n matrix with independent Gaussian entries. In these experiments, we set $n = 128$. (Top) Success rate of (P_1) for $m = 2n$. (Bottom) Success rate of (P_1) for $m = 4n$. On top, exact recovery occurs as long as the corruption rate does not exceed 15%. The bottom breakdown is near 35%.

It is clear that versions of Theorem 1.2 exist for other type of random matrices, e.g. binary matrices. In the next experiment, we take the plaintext f as a binary sequence of zeros and ones (which is generated at random), and sample A with i.i.d entries taking on values in $\{\pm 1\}$, each with probability 1/2. To recover f, we solve the linear program

$$\min_{g \in \mathbb{R}^n} \|y - Ag\|_{\ell_1} \quad \text{subject to} \quad 0 \le g \le 1, \qquad (5.1)$$

and round up the coordinates of the solution to the nearest integer. We follow the same procedure as before except that now, we select S locations of Af at random (the corruption rate is again S/m) and flip the sign of the selected coordinates. We are again interested in the location of the breakpoint.

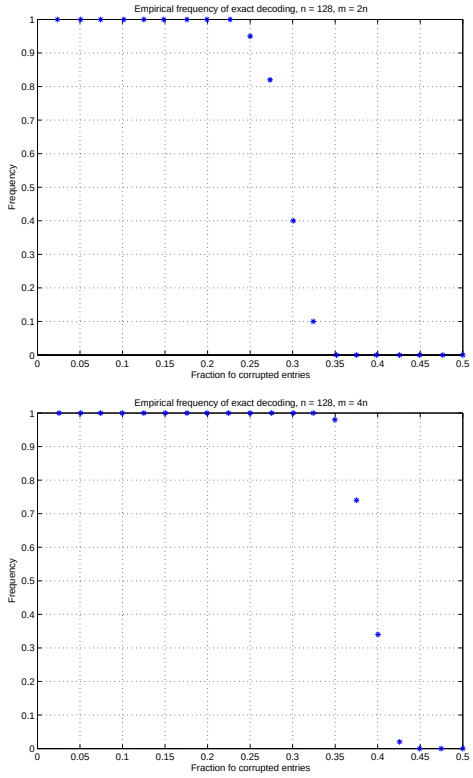


Figure 2. ℓ_1-recovery of a binary sequence from corrupted data y'; A an m by n matrix with independent binary entries and the vector of errors is obtained by randomly selecting coordinates of Af and flipping their sign. In these experiments, we set $n = 128$. (Top) Success rate for $m = 2n$. (Bottom) Success rate for $m = 4n$. On top, exact recovery occurs as long as the corruption rate does not exceed 22.5%. The bottom breakdown is near 35%.

The results are presented in Figure 2. In these experiments, we choose $n = 128$ as before, and set $m = 2n$ (Figure 2(a)) or $m = 4n$ (Figure 2(b)). Our experiments show that the linear program recovers the input vector *all the time* as long as the fraction of the corrupted entries is less or equal to 22.5% in the case where $m = 2n$ and less than about 35% in the case where $m = 4n$. We repeated these experiments for different values of n, e.g. $n = 256$

and obtained similar recovery curves.

In conclusion, our error correcting strategy seems to enjoy a wide range of effectiveness.

6 Discussion

A first impulse to find the sparsest solution to an underdetermined system of linear equations might be to solve the combinatorial problem

$$(P_0'') \qquad \min_{d\in\mathbb{R}^m} \|d\|_{\ell_0} \quad \text{subject to} \quad Bd = Be.$$

To the best of our knowledge, solving this problem essentially require exhaustive searches over all subsets of columns of B and is NP-hard [42]. Our results, however, establish a formal equivalence between (P_0'') and (P_1'') provided that the unknown vector e is sufficiently sparse. In this direction, we would like to mention a series of papers [17, 19, 33, 46] showing the exact equivalence between the two programs (P_0'') and (P_1'') for special matrices obtained by concatenation of two orthonormal bases. In this literature, equivalence holds if e has fewer than $\rho \cdot \sqrt{m}$ entries; compare with Theorem 1.2 which tolerates a fraction of nonzero entries proportional to m.

For Gaussian random matrices, however, [14] proved that the equivalence holds when the number of nonzero entries may be as large as $\rho \cdot m$, where $\rho > 0$ is some very small and unspecified positive constant independent of m. This finding is of course similar to ours but the ideas in this paper go much further. First, the paper establishes *deterministic* results showing that exact decoding occurs provided the coding matrix A obeys the conditions of Theorem 1.1. It is of interest because our own work [8, 10] shows that the condition of Theorem 1.1 with large values of r for many other types of matrices, and especially matrices obtained by sampling rows or columns of larger Fourier matrices. These alternatives might be of great practical interest because they would come with fast algorithms for applying A or A^* to an arbitrary vector g and, hence, speed up the computations to find the ℓ_1-minimizer. And second, the paper of course links solutions to sparse underdetermined systems to a linear programming problem for error correction, which we believe is new.

An natural feature of our error correction code is its *robustness*. Simple linear algebra yields that the solution to (P_1') is stable with respect to the 1-norm – in the same way as the solution to (P_2') is stable with respect to the 2-norm, see [10]. Indeed, it is not hard to show that, once Theorem 1.3 holds, the unknown vector y in Theorem 1.3 can be approximately recovered from $y'' = y' + h$, where $h \in \mathbb{R}^m$ is any additional error vector of small 1-norm (see [10]). Namely, the solution u to the Basis Pursuit problem

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

satisfies $\|u - y\|_1 \le 4\|h\|_1$.

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*. Indeed, we can now describe an (m, n, r)-error correcting code under the Gilbert-Varshamov type assumption (1.7), with input words x of length n over the alphabet $\{1, \dots, p\}$ and the encoded words y of length m over the alphabet $\{1, \dots, Cpn^{3/2}\}$. The encoder takes $x \in \{1, \dots, p\}^n$, computes $y = Qx$ where Q is the orthogonal projection onto $Y = \mathrm{range}(A)$, and outputs the $\hat{y}$ whose coefficients are the quantized coefficients of y with step $\frac{1}{10m}$. Then $\hat{y} \in \frac{1}{10m}\mathbb{Z}^m \cap [-p\sqrt{m}, p\sqrt{m}]^m$, which by rescaling can be identified with $\{1, \dots, Cpn^{3/2}\}$, because we can assume that $m \le 2n$. The decoder takes $y' \in \frac{1}{10m}\mathbb{Z}^m$, finds the minimizer u to (P_1'), inverts to $x' = Q^T u$ and outputs $\hat{x'}$ whose coefficients are the quantized coefficients of x' with step 1.

This is indeed an (m, n, r)-error correcting code. If y' differs from $\hat{y}$ on at most r coordinates, this and the condition $\|\hat{y} - y\|_1 \le \frac{1}{10}$ implies by the robustness that $\|u - y\|_1 \le 0.4$. Hence $\|x' - x\|_2 = \|Q^T(u - y)\|_2 = \|u - y\|_2 \le \|u - y\|_1 \le 0.4$. Thus $\hat{x'} = x$, 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. See [9] for some further discussion of stability of basis pursuit methods.

In our linear programming model, the plaintext and ciphertext had real-valued components. Another intensively studied model occurs when the plaintext and ciphertext take values in the finite field $F_2 := \{0, 1\}$, and the transformation $x \mapsto Ax$ is linear with respect to F_2 rather than $\mathbb{R}$ (note that these are not the same as the quantized linear transformations discussed previously). In recent work of Feldman et al. [24], [25], [26], linear programming methods (based on relaxing the space of codewords to a convex polytope) were developed to establish a polynomial-time decoder which can correct a constant fraction of errors, and also achieve the information-theoretic capacity of the code. There is thus some intriguing parallels between those works and the results in this paper, however there appears to be no direct overlap as our methods are restricted to real-valued texts, and the work cited above requires texts in F_2. Also, our error analysis is deterministic (assuming the isometry condition) and is thus guaranteed to correct arbitrary errors provided that they are sufficiently sparse.

We would like to close this paper by pointing out that for Gaussian matrices, say, there is a critical point ρ_c (depend-

ing on n and m) such that accurate decoding occurs for all plaintexts and corrupted patterns (in the sense of Theorem 1.1) as long as the fraction of corrupted entries does not exceed ρ_c. It would be of theoretical interest to identify this critical threshold, at least in the limit of large m and n, with perhaps n/m converging to a fixed ratio. From a different viewpoint, this is asking about how far does the equivalence between a combinatorial and a related convex problem hold. We pose this as an interesting challenge.

References

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

[2] D. Amir, and V. D. Milman. Unconditional and symmetric sets in n-dimensional normed spaces. *Israel J. Math.* 37: 3–20, 1980.

[3] B. Beferull-Lozano, and A. Ortega. Efficient quantization for overcomplete expansions in $\mathbb{R}^n$. *IEEE Trans. Inform. Theory* 49: 129–150, 2003.

[4] S. Boyd, and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[5] P. G. Casazza, and J. Kovacević. Equal-norm tight frames with erasures. *Adv. Comput. Math.* 18: 387–430, 2003.

[6] E. J. Candès, and J. Romberg, Quantitative robust uncertainty principles and optimally sparse decompositions. To appear *Foundations of Computational Mathematics*, November 2004.

[7] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. To appear *IEEE Transactions on Information Theory*, June 2004. Available on the ArXiV preprint server: `math.NA/0409186`.

[8] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. To appear *Comm. Pure Appl. Math.* Available on the ArXiV preprint server: `math.NA/0409186`.

[9] E. J. Candès, and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? Submitted to *IEEE Transactions on Information Theory*, October 2004. Available on the ArXiV preprint server: `math.CA/0410542`.

[10] E. J. Candès, and T. Tao. Decoding by linear programming. Submitted, December 2004. Available on the ArXiV preprint server: `math.MG/0502327`.

[11] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Scientific Computing* 20: 33–61, 1998.

[12] I.Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, 1992.

[13] D. L. Donoho. For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution. Manuscript, September 2004.

[14] D. L. Donoho. For most large undetermined systems of linear equations the minimal ℓ_1-norm near-solution is also the sparsest near-solution. Manuscript September 2004.

[15] D. Donoho. Compressed sensing. Manuscript, September 2004.

[16] D. L. Donoho, and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. *Proc. Natl. Acad. Sci. USA* 100: 2197–2202 (2003).

[17] D. Donoho, M. Elad, and V. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. Manuscript, 2004.

[18] D. L. Donoho, and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47:2845–2862, 2001.

[19] D. L. Donoho, and Y. Tsaig. Extensions of compresed sensing. Preprint, 2004.

[20] D. Donoho, and Y. Tsaig. Breakdown of equivalence between the minimal ℓ_1-norm solution and the sparsest solution. Preprint, 2004.

[21] N. El Karoui. New Results about Random Covariance Matrices and Statistical Applications. Stanford Ph. .D. Thesis, August 2004.

[22] M. Elad, and A. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Inform. Theory* 48: 2558–2567, 2002.

[23] J. Feldman. Decoding Error-Correcting Codes via Linear Programming. Ph.D. Thesis 2003, Massachussets Institute of Technology.

[24] J. Feldman, LP decoding achieves capacity, 2005 ACM-SIAM Symposium on Discrete Algorithms (SODA), preprint (2005).

[25] J. Feldman, T. Malkin, C. Stein, R. A. Servedio, and M. J. Wainwright, LP decoding corrects a constant fraction of errors. Proc. IEEE International Symposium on Information Theory (ISIT), June 2004.

[26] A. Feuer, and A. Nemirovski. On sparse representation in pairs of bases. *IEEE Trans. Inform. Theory* 49: 1579–1581, 2003.

[27] A. Yu. Garnaev, E. D. Gluskin, The widths of a Euclidean ball (Russian), *Dokl. Akad. Nauk SSSR* 277: 1048–1052, 1984. English translation: *Soviet Math. Dokl.* 30: 200–204, 1984.

[28] V. K. Goyal. Theoretical foundations of transform coding. *IEEE Signal Processing Magazine* 18(5): 9–21, 2001.

[29] V. K. Goyal. Multiple description coding: compression meets the network. *IEEE Signal Processing Magazine* 18(5): 74–93, 2001.

[30] V. K. Goyal, J. Kovacevic, and J. A. Kelner. Quantized frame expansions with erasures. *Applied and Computational Harmonic Analysis* 10: 203–233, 2001.

[31] V. K. Goyal, M. Vetterli, and N. T. Thao. Quantized overcomplete expansions in $\mathbb{R}^N$: analysis, synthesis and algorithms, *IEEE Trans. on Information Theory* 44: 16–31, 1998.

[32] R. Gribonval, and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. Inform. Theory* 49: 3320–3325, 2003.

[33] *Handbook of coding theory. Vol. I, II.* Edited by V. S. Pless, W. C. Huffman and R. A. Brualdi. North-Holland, Amsterdam, 1998.

[34] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* 29: 295–327, 2001.

[35] J. Kovacevic, P. Dragotti, and V. Goyal. Filter bank frame expansions with erasures. *IEEE Trans. on Information Theory*, 48: 1439–1450, 2002.

[36] M. Ledoux. *The concentration of measure phenomenon*. Mathematical Surveys and Monographs 89, American Mathematical Society, Providence, RI, 2001.

[37] M. A. Lifshits, *Gaussian random functions*. Mathematics and its Applications, 322. Kluwer Academic Publishers, Dordrecht, 1995.

[38] V. A. Marchenko, and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* 72: 407–535, 1967 (in Russian).

[39] J. Matousek. *Lectures on discrete geometry*. Graduate Texts in Mathematics, 212. Springer-Verlag, New York, 2002.

[40] S. Mendelson. Geometric parameters in learning theory. *Geometric aspects of functional analysis*. Lecture Notes in Mathematics 1850: 193–235, Springer, Berlin, 2004.

[41] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.* 24: 227–234, 1995.

[42] M. Rudelson, and R. Vershynin. Geometric approach to error correcting codes and reconstruction of signals. Submitted, 2005. Available on the ArXiV preprint server: `math.FA/0502299`.

[43] S. J. Szarek. Condition numbers of random matrices. *J. Complexity* 7:131–149, 1991.

[44] J. Tropp. Recovery of short, complex linear combinations via ℓ_1 minimization. To appear *IEEE Trans. Inform. Theory*.

[45] J. Tropp, Greed is good: Algorithmic results for sparse approximation, *IEEE Trans. Inform. Theory,* 50(10): 2231-2242, October 2004.

[46] J. Tropp. Just relax: Convex programming methods for subset selection and sparse approximation. *ICES Report* 04-04, UT-Austin, 2004.