Highly Robust Error Correction by Convex Programming

Emmanuel J. Candès and Paige A. Randall

***Abstract*—This paper discusses a stylized communications problem where one wishes to transmit a real-valued signal $x \in \mathbb{R}^n$ (a block of n pieces of information) to a remote receiver. We ask whether it is possible to transmit this information reliably when a fraction of the transmitted codeword is corrupted by arbitrary gross errors, and when in addition, all the entries of the codeword are contaminated by smaller errors (e.g., quantization errors).**

We show that if one encodes the information as Ax where $A \in \mathbb{R}^{m \times n}$ $(m \geq n)$ is a suitable coding matrix, there are two decoding schemes that allow the recovery of the block of n pieces of information x with nearly the same accuracy as if no gross errors occurred upon transmission (or equivalently as if one had an oracle supplying perfect information about the sites and amplitudes of the gross errors). Moreover, both decoding strategies are very concrete and only involve solving simple convex optimization programs, either a linear program or a second-order cone program. We complement our study with numerical simulations showing that the encoder/decoder pair performs remarkably well.

***Index Terms*—Decoding of (random) linear codes, Gaussian random matrices and random projections, linear codes, ℓ_1 minimization, linear programming, restricted orthonormality, second-order cone programming, sparse solutions to underdetermined systems, the Dantzig selector.**

I. Introduction

THIS paper discusses a coding problem over the reals. We wish to transmit a block of n real values—a vector $x \in \mathbb{R}^n$—to a remote receiver. A possible way to address this problem is to communicate the codeword Ax where A is an m by n coding matrix with $m \geq n$. Now a recurrent problem with real communication or storage devices is that some portions of the transmitted codeword may become corrupted; when this occurs, parts of the received codeword are unreliable and may have nothing to do with their original values. We represent this as receiving a distorted codeword $y = Ax + z_0$. The question is whether one can recover the signal x from the received data y.

It has recently been shown [1], [2] that one could recover the information x exactly—under suitable conditions on the coding matrix A—provided that the fraction of corrupted entries of Ax is not too large. In greater details, [1] proved that if the corruption z_0 contains at most a fixed fraction of nonzero entries, then the signal $x \in \mathbb{R}^n$ is the unique solution of the minimum-ℓ_{x1} approximation problem

$$\min_{\tilde{x} \in \mathbb{R}^n} \|y - A\tilde{x}\|_{\ell_1}. \tag{I.1}$$

What may appear as a surprise is the fact that this requires no assumption whatsoever about the corruption pattern z_0 except that it must be sparse. In particular, the decoding algorithm is provably exact even though the entries of z_0—and thus of y as well—may be arbitrary large, for example.

While this is interesting, it may not be realistic to assume that except for some gross errors, one is able to receive the values of Ax with infinite precision. A better model would assume instead that the receiver gets

$$y = Ax + z_0, \quad z_0 = e + z \tag{I.2}$$

where e is a possibly sparse vector of gross errors and z is a vector of small errors affecting all the entries. In other words, one is willing to assume that there are malicious errors affecting a fraction of the entries of the transmitted codeword *and* in addition, smaller errors affecting all the entries. For instance, one could think of z as some sort of quantization error which limits the precision/resolution of the transmitted information. In this more practical scenario, we ask whether it is still possible to recover the signal x accurately. The subject of this paper is to show that it is in fact possible to recover the original signal with nearly the same accuracy as if one had a perfect communication system in which no gross errors occurred upon transmission. Further, the recovery algorithms are very concrete and practical; they involve solving very convenient convex optimization problems.

Before expanding on our results, we would like to comment on the practical relevance of our model. Coding theory generally assumes that data take on values in a finite field, but there are a number of applications where encoding over the reals is of direct interest. We give two examples. The first example concerns orthogonal frequency-division multiplexing for wireless and wideband digital communication. Here, one can experience deep fades at certain frequencies (because of mulipathing for instance) and/or frequency jamming because of strong interferers so that large parts of the data are unreliable. The second example is in the area of digital computations. Here, researchers are currently interested in error correction over the reals to protect real-valued results of onboard computations which are executed by circuits that are subject to faults due, for example, to radiation. As we will see, our work introduces an encoding strategy which is robust to such errors, which runs in polynomial time, and which provably obeys optimal bounds.

To understand the claims of this paper in a more quantitative fashion, suppose that we had a perfect channel in which no gross

Manuscript received December 17, 2006; revised December 4, 2007. This work was supported in part by the National Science Foundation (NSF) under Grants ITR ACI-0204932 and CCF515362 and by the 2006 Waterman Award (NSF). The material in this paper was presented at WavE 2006, Lausanne, Switzerland, July 2006.

The authors are with the Department of Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA 91125 USA (e-mail: emmanuel@acm.caltech.edu; paige@acm.caltech.edu).

Communicated by A. J. Goldsmith, Associate Editor for Communications.

Color versions of Figures 1 and 2 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2008.924688

errors ever occurred; that is, we assume $e = 0$ in (I.2). Then we would receive $y = Ax + z$ and would reconstruct x by the method of least squares which, assuming that A has full rank, takes the form

$$x^{\mathrm{Ideal}} = (A^*A)^{-1}A^*y. \tag{I.3}$$

In this ideal situation, the reconstruction error would then obey

$$\|x^{\mathrm{Ideal}} - x\|_{\ell_2} = \|(A^*A)^{-1}A^*z\|_{\ell_2}. \tag{I.4}$$

Suppose we design the coding matrix A with orthonormal columns so that $A^*A = I$. Then we would obtain a reconstruction error whose maximum size is just about that of z. If the smaller errors z_i are independent and identically distributed (i.i.d.) $N(0, \sigma^2)$, then the mean-squared error (MSE) would obey

$$\boldsymbol{E}\|x^{\mathrm{Ideal}} - x\|_{\ell_2}^2 = \sigma^2 \mathrm{Tr}((A^*A)^{-1}).$$

If $A^*A = I$, then the MSE is equal to $n\sigma^2$.

The question then is, can one hope to do almost as well as this optimal MSE without knowing e or even the support of e in advance? This paper shows that one can in fact do almost as well by solving very simple convex programs. This holds for *all* signals $x \in \mathbb{R}^n$ and *all* sparse gross errors no matter how adversary.

Two concrete decoding strategies are introduced: one based on second-order cone programming (SOCP) in Section II, and another based on linear programming (LP) in Section III. We introduce two different decoding strategies because in certain situations it may be preferable to solve an LP over an SOCP or *vice versa*. Also, we show theoretically that the two methods scale differently, so in a particular setup one method could outperform the other. For instance, it is an open question whether or not the SOCP decoder can achieve the adaptive bounds of the LP decoder. In Section IV, we compare the empirical performances of the two decoders in a series of numerical experiments before proving our results in Section V, followed by a discussion in Section VI.

We conclude the introduction by noting that this paper is part of a larger body of work. In particular, besides the obvious connections with [1], [2], it draws on recent results [3], [4] showing that the theory and practice of compressed sensing (also known as compressive sampling) is robust vis a vis noise. The connection with this work should become clear in our proofs.

II. Decoding by Second-Order Cone Programming

To recover the signal x from the corrupted vector y (I.2) we propose solving the following optimization program:

$$\begin{array}{lll} (P_2) & \min & \|y - A\tilde{x} - \tilde{z}\|_{\ell_1} \\ & \text{subject to} & \|\tilde{z}\|_{\ell_2} \le \varepsilon \\ & & A^*\tilde{z} = 0 \end{array} \tag{II.1}$$

with variables $\tilde{x} \in \mathbb{R}^n$ and $\tilde{z} \in \mathbb{R}^m$. The parameter ε above depends on the magnitude of the small errors and shall be specified later. The program (P_2) is equivalent to

$$\begin{array}{lll} \min \ \mathbf{1}^*\tilde{u}, & \text{subject to} & -\tilde{u} \le y - A\tilde{x} - \tilde{z} \le \tilde{u} \\ & & \|\tilde{z}\|_{\ell_2} \le \varepsilon \\ & & A^*\tilde{z} = 0 \end{array} \tag{II.2}$$

where we added the slack optimization variable $\tilde{u} \in \mathbb{R}^m$. In the above formulation, $\mathbf{1}$ is a vector of ones and the vector inequality $u \le v$ means componentwise, i.e., $u_i \le v_i$ for all i. The program (II.2) is a second-order cone program and as a result, (P_2) can be solved efficiently using standard optimization algorithms, see [5].

The first key point of this paper is that the SOCP decoder is highly robust against imperfections in communication channels. Here and below, V denotes the subspace spanned by the columns of A, and $Q \in \mathbb{R}^{m\times(m-n)}$ is a matrix whose columns form an orthobasis of $V^\perp$, the orthogonal complement to V. Such a matrix Q is a kind of parity-check matrix since $Q^*A = 0$. Applying Q^* on both sides of (I.2) gives

$$Q^*y = Q^*e + Q^*z. \tag{II.3}$$

Now if we could somehow get an accurate estimate $\hat{e}$ of e from Q^*y, we could reconstruct x by applying the method of least squares to the vector y corrected for the gross errors

$$\hat{x} = (A^*A)^{-1}A^*(y - \hat{e}). \tag{II.4}$$

If $\hat{e}$ were very accurate, we would probably do very well.

The point is that under suitable conditions, (P_2) provides such accurate estimates. Introduce $\tilde{e} = y - A\tilde{x} - \tilde{z}$, and observe the following equivalence:

$$\begin{array}{lllll} (P_2) & \Leftrightarrow & \min & \|\tilde{e}\|_{\ell_1} & \\ & & \text{subject to} & \tilde{e} = y - A\tilde{x} - \tilde{z}, & \\ & & & A^*\tilde{z} = 0, \|\tilde{z}\|_{\ell_2} \le \varepsilon & \\ \Leftrightarrow & (P_2') & \min & \|\tilde{e}\|_{\ell_1} & \\ & & \text{subject to} & \|Q^*(y - \tilde{e})\|_{\ell_2} \le \varepsilon. & \end{array} \tag{II.5}$$

We only need to argue about the second equivalence since the first is immediate. Observe that the condition $A^*\tilde{z} = 0$ decomposes $y - \tilde{e}$ as the superposition of an arbitrary element in V (the vector $A\tilde{x}$) and of an element in $V^\perp$ (the vector $\tilde{z}$) whose Euclidean length is less than ε. In other words, $\tilde{z} = P_{V^\perp}(y - \tilde{e})$ where $P_{V^\perp} = QQ^*$ is the orthonormal projector onto $V^\perp$ so that the problem is that of minimizing the ℓ_1 norm of $\tilde{e}$ under the constraint $\|P_{V^\perp}(y - \tilde{e})\|_{\ell_2} \le \varepsilon$. The claim follows from the identity $\|P_{V^\perp}v\|_{\ell_2} = \|Q^*v\|_{\ell_2}$ which holds for all $v \in \mathbb{R}^m$.

The equivalence between (P_2) and (P_2') asserts that if $(\hat{x}, \hat{z})$ is solution to (P_2), then $\hat{e} = y - A\hat{x} - \hat{z}$ is solution to (P_2') and *vice versa*; if $\hat{e}$ is solution to (P_2'), then there is a unique way to write $y - \hat{e}$ as the sum $A\hat{x} + \hat{z}$ with $z \in V^\perp$, and the pair $(\hat{x}, \hat{z})$ is solution to (P_2). We note, and this is important, that the solution $\hat{x}$ to (P_2) is also given by the corrected least squares formula (II.4). Equally important is to note that even though we use the matrix Q to explain the rationale behind the methodology, one should keep in mind that Q does not play any special role in (P_2).

The issue here is that if $\|P_{V^\perp}v\|_{\ell_2}$ is approximately proportional to $\|v\|_{\ell_2}$ for all sparse vectors $v \in \mathbb{R}^m$, then the solution $\hat{e}$ to (P_2') is close to e, provided that e is sufficiently sparse [3]. Quantitatively speaking, if ε is chosen so that $\|P_{V^\perp}z\|_{\ell_2} \le \varepsilon$, then $\|e - \hat{e}\|$ is less than a numerical constant times ε; that is, the reconstruction error is within the noise level. The key concept underlying this theory is the so-called restricted isometry property.

Definition 2.1: Define the isometry constant δ_k of a matrix Φ as the smallest number such that

$$(1-\delta_k)\|x\|_{\ell_2}^2 \le \|\Phi x\|_{\ell_2}^2 \le (1+\delta_k)\|x\|_{\ell_2}^2 \tag{II.6}$$

holds for all k-sparse vectors x (a k-sparse vector has at most k nonzero entries).

In the sequel, we shall be concerned with the isometry constants of A^* times a scalar. Since AA^* is the orthogonal projection P_V onto V, we will be thus interested in subspaces V such that P_V nearly acts as an isometry on sparse vectors. Our first result states that the SOCP decoder is provably accurate.

Theorem 2.2: Choose a coding matrix $A \in \mathbb{R}^{m\times n}$ with orthonormal columns spanning V, and let (δ_k) be the isometry constants of the rescaled matrix $\sqrt{\frac{m}{n}}\,A^*$. Suppose $\|P_{V^\perp} z\|_{\ell_2} \le \varepsilon$. Then the solution $\hat{x}$ to (P_2) obeys

$$\|\hat{x}-x\|_{\ell_2} \le C_2 \cdot \frac{\varepsilon}{\sqrt{1-\frac{n}{m}}} + \|x^{\mathrm{Ideal}}-x\|_{\ell_2} \tag{II.7}$$

for some $C_2 = C_2(c)$ provided that the number k of gross errors obeys $\delta_{3k} + \frac{1}{2}\delta_{2k} < \frac{c}{2}\left(\frac{m}{n}-1\right)$ for some $c<1$; x^{Ideal} is the ideal solution (I.3) one would get if no gross errors ever occurred $(e=0)$.

If the (orthonormal) columns of A are selected uniformly at random, then with probability at least $1-O(e^{-\gamma(m-n)})$ for some positive constant γ, the estimate (II.7) holds for $k \asymp \rho\cdot m$, provided $\rho \le \rho^*(n/m)$, which is a constant depending only n/m.[1]

This theorem is of significant appeal because it says that the reconstruction error is in some sense within a constant factor of the ideal solution. Indeed, suppose all we know about z is that $\|z\|_{\ell_2} \le \varepsilon$. Then $\|x^{\mathrm{Ideal}}-x\|_{\ell_2} = \|A^*z\|_{\ell_2}$ may be as large as ε. Thus, for $m=2n$, say, (II.7) asserts that the reconstruction error is bounded by a constant times the ideal reconstruction error. In addition, if one selects a coding matrix with random orthonormal columns (one way of doing so is to sample $X \in \mathbb{R}^{m\times n}$ with i.i.d. $N(0,1)$ entries and orthonormalize the columns by means of the QR factorization), then one can correct a positive fraction of arbitrarily corrupted entries, in a near-ideal fashion.

Note that in the case where there are no small errors $(z=0)$, the decoding is exact since $\varepsilon = 0$ and $x^{\mathrm{Ideal}} = x$. Hence, this generalizes earlier results [1]. We would like to emphasize that there is nothing special about the fact that the columns of A are taken to be orthonormal in Theorem 2.2. In fact, one could just as well obtain equivalent statements for general matrices. Our assumption only allows us to formulate simple and useful results.

While the previous result discussed arbitrary small errors, the next is about stochastic errors.

Corollary 2.3: Suppose the small errors are i.i.d. $N(0,\sigma^2)$ and set $\varepsilon := \sqrt{(m-n)(1+t)}\cdot\sigma$ for some fixed $t>0$. Then under the same hypotheses about the restricted isometry constants of A and the number of gross errors as in Theorem 2.2, the solution to (P_2) obeys

$$\|\hat{x}-x\|_{\ell_2}^2 \le C_2' \cdot m \cdot \sigma^2 \tag{II.8}$$

for some numerical constant C_2' with probability exceeding $1-e^{-\gamma^2(m-n)/2}-e^{-m/2}$ where $\gamma = \frac{\sqrt{1+2t}-1}{\sqrt{2}}$. In particular, this last statement holds with overwhelming probability if A is chosen at random as in Theorem 2.2.

Suppose, for instance, that $m=2n$ to make things concrete so that the MSE of the ideal estimate is equal to $m/2\cdot\sigma^2$. Then the SOCP reconstruction is within a multiplicative factor $2C$ of the ideal MSE. Our experiments show that in practice the constant is small: e.g., when $m=2n$, one can correct 15% of arbitrary errors, and in the overwhelming majority of cases obtain a decoded vector whose MSE is less than three times larger than the ideal MSE.

III. Decoding by Linear Programming

Another way to recover the signal x from the corrupted vector y (I.2) is by linear programming

$$\begin{aligned}(P_\infty)\quad \min\quad & \|y-A\tilde{x}-\tilde{z}\|_{\ell_1}\\ \text{subject to}\quad & \|\tilde{z}\|_{\ell_\infty} \le \lambda\\ & A^*\tilde{z}=0\end{aligned} \tag{III.1}$$

with variables $\tilde{x}\in\mathbb{R}^n$ and $\tilde{z}\in\mathbb{R}^m$. As is well known, the program (P_∞) may also be re-expressed as a linear program by introducing slack variables just as in (P_2); we omit the standard details. As with (P_2), the parameter λ here is related to the size of the small errors and will be discussed shortly. In the sequel, we shall also be interested in the more general formulation of (P_∞)

$$\begin{aligned}\min\quad & \|y-A\tilde{x}-z\|_{\ell_1}\\ \text{subject to}\quad & |\tilde{z}|_i \le \lambda_i,\quad 1\le i\le m\\ & A^*\tilde{z}=0\end{aligned} \tag{III.2}$$

which gives additional flexibility for adjusting the thresholds $\lambda_1,\lambda_2,\ldots,\lambda_m$ to the noise level.

The same arguments as before prove that (P_∞) is equivalent to

$$\begin{aligned}(P_\infty')\qquad \min\quad & \|\tilde{e}\|_{\ell_1}\\ \text{subject to}\quad & \|QQ^*(y-\tilde{e})\|_{\ell_\infty} \le \lambda\end{aligned} \tag{III.3}$$

where we recall that $P_{V^\perp} = QQ^*$ is the orthonormal projector onto $V^\perp$ (V is the column space of A); that is, if $\hat{e}$ is solution to (P_∞'), then there is a unique decomposition $y-\hat{e} = A\hat{x}+\hat{z}$ where $A^*\hat{z}=0$ and $(\hat{x},\hat{z})$ is solution to (P_∞). The converse is also true. Similarly, the more general program (III.2) is equivalent to minimizing the ℓ_1 norm of $\tilde{e}$ under the constraint $|P_{V^\perp}(y-\tilde{e})|_i \le \lambda_i$, $1\le i\le m$.

In statistics, the estimator $\hat{e}$ solution to (P_∞') is known as the *Dantzig selector* [4]. It was originally introduced to estimate the

[1]Analysis shows ρ^* to be of the form $\rho^* = O\left(\frac{n/m-1}{\log(1-n/m)}\right)$ but this is not informative because the constant is unknown. Determining the constant is extremely challenging; for an analysis with sparse errors see [6], [7].

vector e from the data y' and the model

$$y' = Q^*e + z' \tag{III.4}$$

where z' is a vector of stochastic errors, e.g., independent mean-zero Gaussian random variables. The connection with our problem is clear since applying the parity-check matrix Q^* on both sides of (I.2) gives

$$Q^*y = Q^*e + Q^*z$$

as before. If z is stochastic noise, we can use the Dantzig selector to recover e from Q^*y. Moreover, available statistical theory asserts that if Q^* obeys nice restricted isometry properties and e is sufficiently sparse just as before, then this estimation procedure is extremely accurate and in some sense optimal.

It remains to discuss how one should specify the parameter λ in (III.1)–(III.3) which is easy. Suppose the small errors are stochastic. Then we fix λ so that the true vector e is feasible for (P'_∞) with very high probability; i.e., we adjust λ so that

$$\|P_{V^\perp}(y-e)\|_{\ell_\infty} = \|P_{V^\perp}z\|_{\ell_\infty} \le \lambda$$

with high probability. In the more general formulation, the thresholds are adjusted so that $\sup_{1\le i\le m} |P_{V^\perp}z|_i/\lambda_i \le 1$ with high probability.

The main result of this section is that the LP decoder is also provably accurate.

Theorem 3.1: Choose a coding matrix $A \in \mathbb{R}^{m\times n}$ with orthonormal columns spanning V, and let (δ_k) be the isometry constants of the rescaled matrix $\sqrt{\frac{m}{n}}\,A^*$. Suppose $\|P_{V^\perp}z\|_{\ell_\infty} \le \lambda$. Then the solution $\hat{x}$ to (P_∞) obeys

$$\|\hat{x}-x\|_{\ell_2} \le C_1\sqrt{k}\cdot\frac{\lambda}{1-\frac{n}{m}} + \|x^{\mathrm{Ideal}}-x\|_{\ell_2} \tag{III.5}$$

for some $C_1 = C_1(c)$ provided that the number k of gross errors obeys $\delta_{3k}+\delta_{2k} < c\left(\frac{m}{n}-1\right)$ for some $c<1$; x^{Ideal} is the ideal solution (I.3) one would get if no gross errors ever occurred.

If the (orthonormal) columns of A are selected uniformly at random, then with probability at least $1-O(e^{-\gamma(m-n)})$ for some positive constant γ, the estimate (III.5) holds for $k \asymp \rho\cdot m$, provided $\rho \le \rho^*(n/m)$.

In effect, the LP decoder efficiently corrects a positive fraction of arbitrarily corrupted entries. Again, when there are no small errors ($z=0$), the decoding is exact. (Also and just as before, there is nothing special about the fact that the columns of A are taken to be orthonormal.) We now consider the interesting case in which the small errors are stochastic. Below, we conveniently adjust the thresholds λ_j so that the true vector e is feasible with high probability, see Section V-D for details.

Corollary 3.2: Choose a coding matrix A with (orthonormal) columns selected uniformly at random and suppose the small errors are i.i.d. $N(0,\sigma^2)$. Fix

$$\lambda_i = \sqrt{2\log m}\cdot\sqrt{1-\|A_{i,.}\|^2_{\ell_2}}\cdot\sigma$$

in (III.2), where $\|A_{i,.}\|_{\ell_2} = \left(\sum_{1\le j\le n} A^2_{i,j}\right)^{1/2}$ is the ℓ_2 norm of the ith row. Then if the number k of gross errors is no more than a fraction of m as in Theorem 3.1, the solution $\hat{x}$ obeys

$$\|\hat{x}-x\|^2_{\ell_2} \le \left[1+C'_1 s\right]^2\cdot\|x^{\mathrm{Ideal}}-x\|^2_{\ell_2} \tag{III.6}$$

with very large probability, where C'_1 is some numerical constant and

$$s^2 = \frac{k}{m}\cdot\frac{\log m}{\frac{n}{m}\left(1-\frac{n}{m}\right)}.$$

In effect, $\|\hat{x}-x\|^2_{\ell_2}$ is bounded by just about $[1+C'_1 s]^2\cdot n\sigma^2$ since $\|x^{\mathrm{Ideal}}-x\|^2_{\ell_2}$ is distributed as σ^2 times a chi-square with n degrees of freedom, and is tightly concentrated around $n\sigma^2$.

Recall that the MSE is equal to $n\sigma^2$ when there are no gross errors and, therefore, this last result asserts that the reconstruction error is bounded by a constant times the ideal reconstruction error. Suppose for instance that $m = 2n$. Then $s^2 = 4k(\log m)/m$ and we see that s is small when there are few gross errors. In this case, the recovery error is very close to that attained by the ideal procedure. Our experiments show that in practice, the constant C'_1 is quite small: for instance, when $m = 2n$, one can correct 15% of arbitrary errors, and in the overwhelming majority of cases obtain a decoded vector whose MSE is less than three times larger than the ideal MSE. Finally, this last result is in some way more subtle than the corresponding result for the SOCP decoder. Indeed, note the explicit dependence on k of the scaling factor in (III.6) that is not present in the corresponding expression for the SOCP decoder (II.8). This says that in some sense the accuracy of the LP decoder automatically adapts to the number k of gross errors which were introduced. The smaller this number, the smaller the recovery error. For small values of k, the bound in (III.6) may in fact be considerably smaller than its analog (II.8).

IV. Numerical Experiments

As mentioned earlier, numerical studies show that the empirical performance of the proposed decoding strategies is noticeable. To confirm these findings, this section discusses an experimental setup and presents numerical results. The reader wanting to reproduce our results may find the matlab file available at `http://www.acm.caltech.edu/~emmanuel/ConvexDecode.m` useful. Here are the steps we used.

1) Choose a pair (n,m) and sample an m by n matrix A with independent standard normal entries; the coding matrix is fixed throughout.
2) Choose a fraction ρ of grossly corrupted entries and define the number of corrupted entries as $k = \mathtt{round}(\rho\cdot m)$; e.g., if $m = 512$ and 10% of the entries are corrupted, $k = 51$.
3) Sample a block of information $x \in \mathbb{R}^n$ with i.i.d. Gaussian entries. Compute Ax.
4) Select k locations uniformly at random and flip the signs of Ax at these locations.
5) Sample the vector $z = (z_1,\ldots,z_m)$ of smaller errors with z_i i.i.d. $N(0,\sigma^2)$, and add z to the outcome of the previous step. Obtain y.
6) Obtain $\hat{x}$ by solving both (P_2) and (P_∞) followed by a reprojection step discussed below [4].
7) Repeat steps 3)–6) 500 times.

We briefly discuss the reprojection step. As observed in [4], both programs (P'_2) and (P'_∞) have a tendency to underestimate

the vector e (they tend to be akin to soft-thresholding procedures). One can easily correct for this bias as follows.

1) Solve (P_2') or (P_∞') and obtain $\hat{e}$.
2) Estimate the support of the gross errors e via $I := \{i : |\hat{e}_i| > \sigma\}$, where σ is the standard deviation of the smaller errors; recall that $y' := Q^*y = Q^*e + Q^*z$ and update the estimate by regressing y' onto the selected columns of Q^* via the method of least squares

$$\hat{e} = \operatorname{argmin} \|y' - Q^*\tilde{e}\|_{\ell_2}^2 \text{ subject to } \tilde{e}_i = 0, i \in I^c.$$

3) Finally, obtain $\hat{x}$ via $(A^*A)^{-1}A^*(y - \hat{e})$ where $\hat{e}$ is the reprojected estimate calculated in the previous step.

In our series of experiments, we used $m = 2n = 512$ and a corruption rate of 10%. The standard deviation σ is selected in such a way that just about the first three binary digits of each entry of the codeword Ax are reliable. Formally, $\sigma = \text{median}\,|Ax|/16$. Finally and to be complete, we set the threshold ε in (P_2) so that $\|Q^*z\|_{\ell_2} \le \varepsilon$ with probability 0.95; in other words, $\varepsilon^2 = \chi^2_{m-n}(0.95) \cdot \sigma^2$, where $\chi^2_{m-n}(0.95)$ is the 95th percentile of a chi-squared distribution with $m - n$ degrees of freedom. We also set the thresholds in the general formulation (III.2) of (P_∞) in a similar fashion. The distribution of $(QQ^*z)_i$ is normal with mean 0 and variance $s_i^2 = (QQ^*)_{i,i} \cdot \sigma^2$ so that the variable $z_i' = (QQ^*z)_i/s_i$ is standard normal. We choose $\lambda_i = \lambda \cdot s_i$ where λ obeys

$$\sup_{1 \le i \le m} |z_i'| \le \lambda$$

with probability at least 0.95. In both cases, our selection makes the true vector e of gross errors feasible with probability at least 0.95. In our simulations, the thresholds for the SOCP and LP decoders (the parameters $\chi^2_{m-n}(0.95)$ and λ) were computed by Monte Carlo simulations.

To evaluate the accuracy of the decoders, we report two statistics

$$\rho^{\text{Ideal}} = \frac{\|\hat{x} - x\|}{\|x^{\text{Ideal}} - x\|} \quad \text{and} \quad \rho^{\text{Oracle}} = \frac{\|\hat{x} - x\|}{\|x^{\text{Oracle}} - x\|} \tag{IV.1}$$

which compare the performance of our decoders with that of ideal strategies which assume either exact knowledge of the gross errors or exact knowledge of their locations. As discussed earlier, x^{Ideal} is the reconstructed vector one would obtain if the gross errors were known to the receiver *exactly* (which is of course equivalent to having no gross errors at all). The reconstruction x^{Oracle} is that one would obtain if, instead, one had available an oracle supplying perfect information about the location of the gross errors (but not their value). Then one could simply delete the corrupted entries of the received codeword y and reconstruct x by the method of least squares, i.e., find the solution to $\|y^{\text{Oracle}} - A^{\text{Oracle}}\tilde{x}\|_{\ell_2}$, where A^{Oracle} (resp., y^{Oracle}) is obtained from A (resp., y) by deleting the corrupted rows.

The results are presented in Fig. 1 and summarized in Table I. These results show that both our approaches work extremely well. As one can see, our methods give reconstruction errors which are nearly as sharp as if no gross errors had occurred or as if one knew the locations of these large errors exactly. Put in a different way, the constants appearing in our quantitative bounds are in practice very small. Finally, the SOCP and LP decoders have about the same performance although upon closer inspection, one could argue that the LP decoder is perhaps a tiny bit more accurate.

We also repeated the same experiment but with a coding matrix A consisting of $n = 256$ randomly sampled columns of the 512×512 discrete Fourier transform, and obtained very similar results. The results are presented in Fig. 2 and summarized in Table II. The numbers are remarkably close to our earlier findings and again both our methods work extremely well (again the LP decoder is a tiny bit more accurate). This experiment is of special interest since it suggests that one can apply our decoding algorithms to very large data vectors, e.g., with sizes ranging in the hundreds of thousands. The reason is that one can use off-the-shelf interior point algorithms which only need to be able to apply A or A^* to arbitrary vectors (and never need to manipulate the entries of A or even store them). When A is a partial Fourier transform, one can evaluate Ax and A^*y by means of the fast Fourier transform (FFT) and, hence, this is well suited for very large problems. See [8] for very large scale experiments of a similar flavor.

V. Proofs

In this section, we prove all of our results. We begin with some preliminaries which will be used throughout, then prove the claims about the SOCP decoder, and end this section with the LP decoder. Our work builds on [3] and [4].

A. Preliminaries

We shall make extensive use of two simple lemmas that we now record.

Lemma 5.1: Let $Y_d \sim \chi^2_d$ be distributed as a chi-squared random variable with d degrees of freedom. Then for each $t > 0$

$$\begin{aligned} \boldsymbol{P}(Y_d - d \ge t\sqrt{2d} + t^2) &\le e^{-t^2/2} \quad \text{and} \\ \boldsymbol{P}(Y_d - d \le -t\sqrt{2d}) &\le e^{-t^2/2}. \end{aligned} \tag{V.1}$$

This is fairly standard [9], see also [10] for very slightly refined estimates. We will use (V.1) as follows: for each $\epsilon \in (0, 1)$ we have

$$\begin{aligned} \boldsymbol{P}(Y_d \ge d(1-\epsilon)^{-1}) &\le e^{-\epsilon^2 d/4} \quad \text{and} \\ \boldsymbol{P}(Y_d \le d(1-\epsilon)) &\le e^{-\epsilon^2 d/4}. \end{aligned} \tag{V.2}$$

A consequence of these large deviation bounds is the following estimate.

Lemma 5.2: Let $(u_1, u_2, \ldots, u_m)$ be a vector uniformly distributed on the unit sphere in m dimensions and $Z_n = u_1^2 + \cdots + u_n^2$ be the squared length of its first n components. Then for each $t < 1$

$$\boldsymbol{P}\left(Z_n \le \frac{n}{m}(1-t)\right) \le e^{-nt^2/16} + e^{-mt^2/16} \tag{V.3}$$

and

$$\boldsymbol{P}\left(Z_n \ge \frac{n}{m}\frac{1}{1-t}\right) \le e^{-nt^2/16} + e^{-mt^2/16}. \tag{V.4}$$

Proof: A result of this kind would essentially follow from the measure concentration on the sphere [11], but we

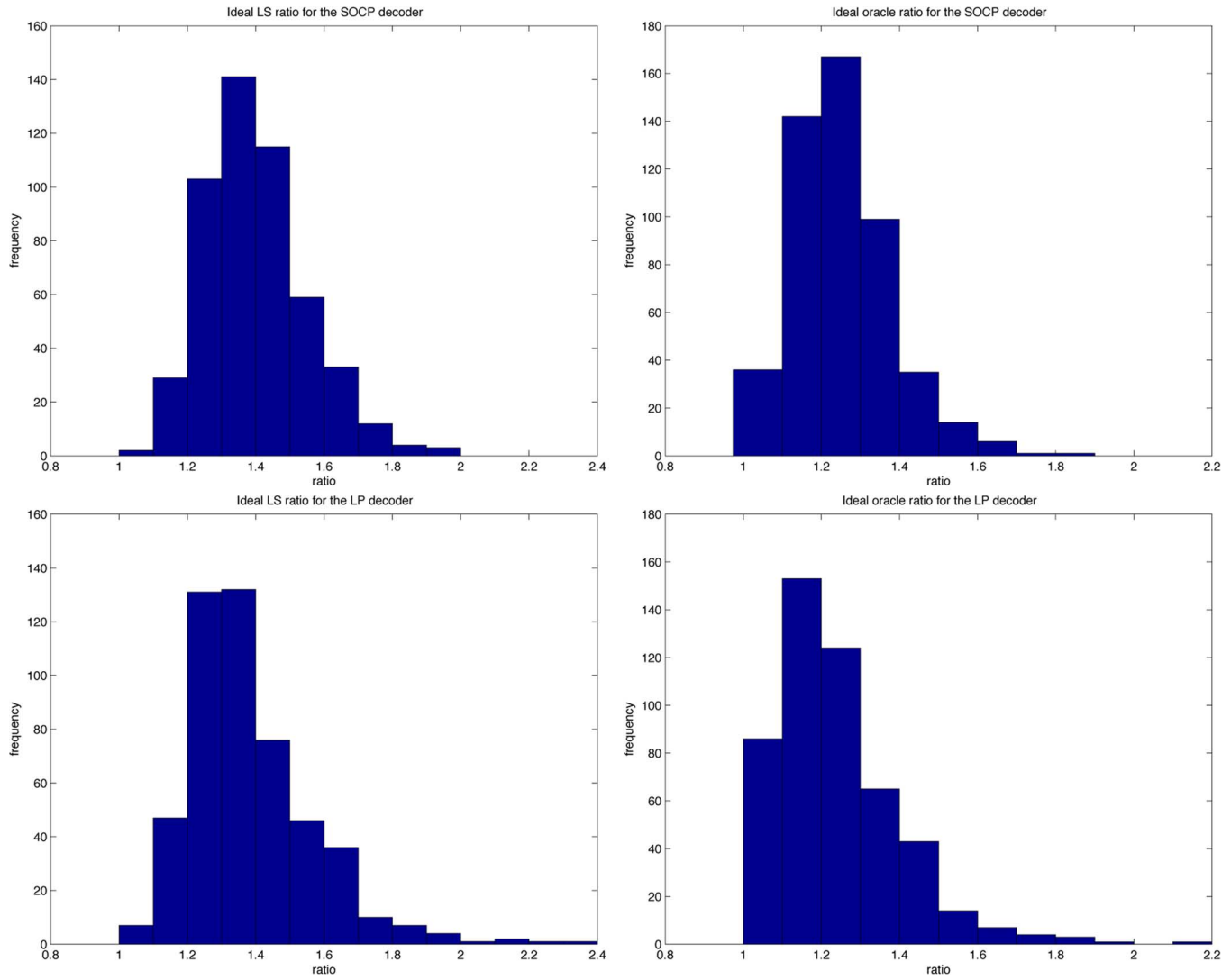

Fig. 1. Statistics of the ratios (IV.1) ρ^{Ideal} (first column) and ρ^{Oracle} (second column) which compare the performance of the proposed decoders with that of ideal strategies which assume either exact knowledge of the gross errors or exact knowledge of their locations. The first row shows the performance of the SOCP decoder, the second that of the LP decoder.

prefer giving a short and elementary argument. Suppose $X_1, X_2, \ldots, X_m$ are i.i.d. $N(0,1)$. Then the distribution of $(u_1, u_2, \ldots, u_m)$ is that of the vector $X/\|X\|_{\ell_2}$ and, therefore, the law of Z_n is that of Y_n/Y_m, where $Y_k = \sum_{j\le k} X_j^2$. For a fixed $t \in (0,1)$, define the events $A = \{Y_m \ge m/(1-t/2)\}$ and $B = \{Y_n/Y_m \le n/m(1-t)\}$. We have

$$\begin{aligned}\boldsymbol{P}(B) &= \boldsymbol{P}(B|A^c)\boldsymbol{P}(A^c) + \boldsymbol{P}(B|A)\boldsymbol{P}(A)\\ &\le \boldsymbol{P}(Y_n \le n(1-t)/(1-t/2))\\ &\quad + \boldsymbol{P}(Y_m \ge m/(1-t/2)).\end{aligned}$$

For $0 \le t \le 1$, we have $(1-t)/(1-t/2) \le 1 - t/2$ and thus

$$\begin{aligned}&\boldsymbol{P}(Z_n \le n/m(1-t))\\ &\quad \le \boldsymbol{P}(Y_n \le n(1-t/2)) + \boldsymbol{P}(Y_m \ge m/(1-t/2))\\ &\quad \le e^{-nt^2/16} + e^{-mt^2/16}\end{aligned}$$

which follows from (V.2).

For the second inequality, we employ a similar strategy with $A = \{Y_m \le m(1-t/2)\}$ and $B = \{Y_n/Y_m \ge n/m\,(1-t)^{-1}\}$, which leads to

$$\begin{aligned}&\boldsymbol{P}(Z_n \ge n/m(1-t)^{-1})\\ &\quad \le \boldsymbol{P}(Y_n \ge n/(1-t/2)) + \boldsymbol{P}(Y_m \le m(1-t/2))\\ &\quad \le e^{-nt^2/16} + e^{-mt^2/16},\end{aligned}$$

as claimed. □

B. Restricted Isometries

For a matrix Φ, define the sequences (a_k) and (b_k) as respectively the largest and smallest numbers obeying

$$a_k\|x\|_{\ell_2} \le \|\Phi x\|_{\ell_2} \le b_k\|x\|_{\ell_2} \tag{V.5}$$

for all k-sparse vectors. In other words, if we list all the singular values of all the submatrices of Φ with k columns, a_k is the smallest element from that list and b_k the largest. Note of course the resemblance to (II.6)—only this is slightly more general.

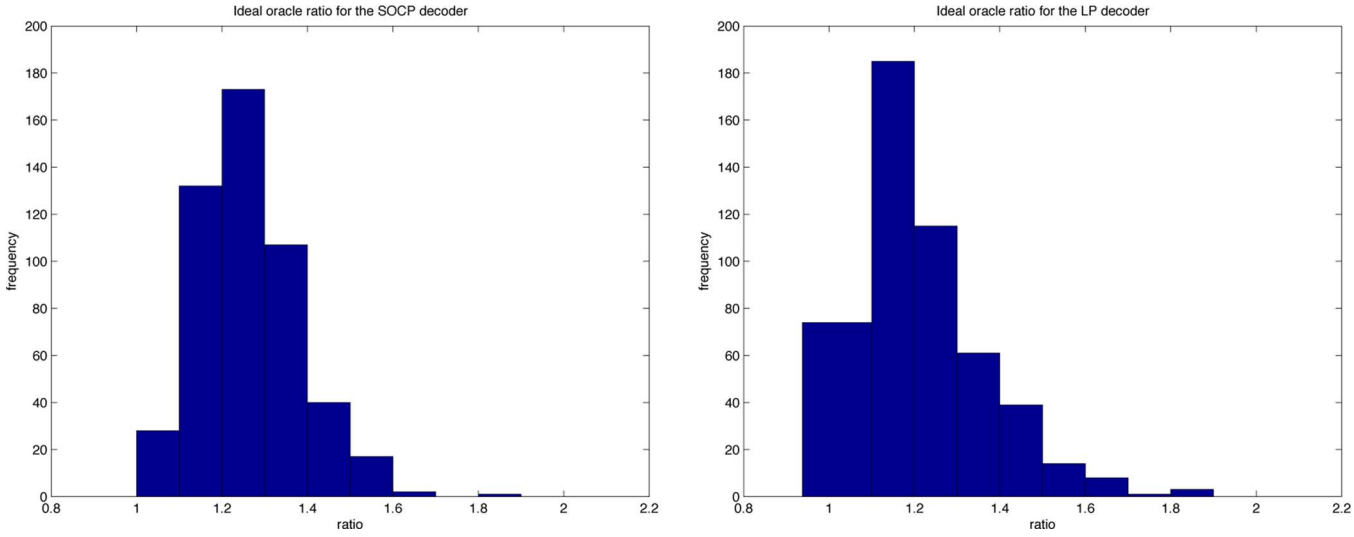

Fig. 2. Statistics of the ratios ρ^{Oracle} for the SOCP decoder (first column) and the LP decoder (second column) in the case where the coding matrix is a partial Fourier transform.

TABLE I
SUMMARY STATISTICS OF THE RATIOS ρ^{Ideal} AND ρ^{Oracle} (IV.1) FOR THE GAUSSIAN CODING MATRIX

	median of ρ^{Ideal}	mean of ρ^{Ideal}	median of ρ^{Oracle}	mean of ρ^{Oracle}
SOCP decoder	1.386	1.401	1.241	1.253
LP decoder	1.346	1.386	1.212	1.239

TABLE II
SUMMARY STATISTICS OF THE RATIOS ρ^{Ideal} AND ρ^{Oracle} (IV.1) FOR THE FOURIER CODING MATRIX

	median of ρ^{Ideal}	mean of ρ^{Ideal}	median of ρ^{Oracle}	mean of ρ^{Oracle}
SOCP decoder	1.390	1.401	1.244	1.262
LP decoder	1.337	1.375	1.195	1.230

Restricted extremal singular values of random orthonormal projections will play an important role in the sequel. The following lemma states that for an $r \times m$ random orthogonal projection, the numbers a_k and b_k are about $\sqrt{r/m}$.

Lemma 5.3: Let Φ be the first r rows of a random orthogonal matrix (sampled from the Haar measure). Then the restricted extremal singular values of Φ obey

$$\boldsymbol{P}\left(a_k(\Phi) \le \frac{7}{8}\sqrt{\frac{r}{m}}\right) \le c_0 e^{-\gamma_0 r} \tag{V.6}$$

and

$$\boldsymbol{P}\left(b_k(\Phi) \ge \frac{9}{8}\sqrt{\frac{r}{m}}\right) \le c_0' e^{-\gamma_0' r} \tag{V.7}$$

for some universal positive constants $c_0, c_0', \gamma_0, \gamma_0'$ provided that $k \le c_1 r/\log(m/r)$ for some $c_1 > 0$.

Proof: Put Σ_k for the set of all unit-normed k-sparse vectors. By definition

$$a_k(\Phi) = \inf_{x \in \Sigma_k} \|\Phi x\|_{\ell_2}.$$

Take a fixed vector x in Σ_k. Then $\|\Phi x\|_{\ell_2}^2$ is distributed as Z_r in Lemma 5.2. To see why this is true, note that Φx are the first r components of Ux where U is an $m \times m$ random orthogonal matrix. The claim follows from the fact that Ux is uniformly distributed on the $(m-1)$-dimensional unit sphere. This is useful because Lemma 5.2 can be employed to show that for a fixed $x \in \Sigma_k$, $\|\Phi x\|_{\ell_2}$ cannot deviate much from $\sqrt{r/m}$. To develop an inequality concerning all sparse vectors, we now employ a covering number argument. Consider an ϵ-net $\mathcal{N}(\epsilon)$ of Σ_k. An ϵ-net is a subset $\mathcal{N}(\epsilon)$ of Σ_k such that for all $x \in \Sigma_k$, there is an $x_0 \in \mathcal{N}(\epsilon)$ such that $\|x - x_0\|_{\ell_2} \le \epsilon$. In other words, $\mathcal{N}(\epsilon)$ approximates Σ_k to within distance ϵ. For each $x \in \Sigma_k$

$$\|\Phi x\| \ge \|\Phi x_0\|_{\ell_2} - \|\Phi(x - x_0)\|_{\ell_2} \ge \|\Phi x_0\| - \epsilon$$

for some $x_0 \in \mathcal{N}(\epsilon)$ obeying $\|x - x_0\|_{\ell_2} \le \epsilon$, where the last inequality follows from the fact that the operator norm of Φ is bounded by 1. Hence

$$a_k(\Phi) \ge \inf_{x_0 \in \mathcal{N}(\epsilon)} \|\Phi x_0\| - \epsilon.$$

Now set $\epsilon = 1/16 \cdot \sqrt{r/m}$. Then

$$\begin{aligned}
&\boldsymbol{P}\left(a_k(\Phi) < \frac{7}{8}\sqrt{\frac{r}{m}}\right) \\
&\quad \le \boldsymbol{P}\left(\inf_{x_0 \in \mathcal{N}(\epsilon)} \|\Phi x_0\|_{\ell_2} \le \sqrt{\frac{r}{m}}\left(1 - \frac{1}{16}\right)\right) \\
&\quad \le |\mathcal{N}(\epsilon)| \cdot \boldsymbol{P}\left(Z_r < \frac{r}{m}\left(1 - \frac{1}{16}\right)^2\right)
\end{aligned}$$

which comes from the union bound together with $\|\Phi x_0\|_{\ell_2}^2 \sim Z_r$ for each x_0. Further, one can find $\mathcal{N}(\epsilon)$ obeying

$$|\mathcal{N}(\epsilon)| \le (3/\epsilon)^k \binom{m}{k}.$$

The reason is simple. First, one can find an ϵ-net of the $k-1$-dimensional sphere whose cardinality does not exceed $(3/\epsilon)^k$, see [12, Lemma 4.16]. And second, Σ_k is a union of $\binom{m}{k}$ $k-1$-dimensional spheres. We then apply this fact together with Lemma 5.2, and obtain

$$\boldsymbol{P}\left(a_k(\Phi) < \frac{7}{8}\sqrt{\frac{r}{m}}\right) \le 48^k (m/r)^{k/2}\binom{m}{k} e^{-r/133}.$$

Next, there is a bound on binomial coefficients of the form $\log\binom{m}{k} \le k(\log(m/k)+1)$ so that

$$48^k (m/r)^{k/2}\binom{m}{k} \le \exp\left(k(5+0.5\log(m/r)+\log(m/k))\right).$$

One can check that if $k \le c_0 \cdot r/(\log m/r)$ for c_0 sufficiently small, the right-hand side of the last inequality is bounded by $e^{\beta_0 r}$ for some $\beta_0 < 1/133$. This establishes the first part of the theorem, namely, (V.6).

The second part is nearly identical and is only sketched. We have that

$$b_k(\Phi) = \sup_{x\in\Sigma_k} \|\Phi x\|_{\ell_2} \le \sup_{x_0\in\mathcal{N}(\epsilon)} \|\Phi x_0\|_{\ell_2} + \epsilon.$$

The proof now proceeds as before noting that (V.4) gives a bound on the probability that for each x_0, $\|\Phi x_0\|_{\ell_2}^2$ exceeds r/m times a small multiplicative factor. □

Note that in this proof, we have not tried to derive the optimal constants, and a more refined analysis would surely yield far better numerical constants.

C. The SOCP Decoder

We begin by adapting an important result from [3].

Lemma 5.4 (Adapted From [3]): Set $\Phi \in \mathbb{R}^{r\times m}$ and let (a_k) and (b_k) be the restricted extremal singular values of Φ as in (V.5). Any point $\tilde{x} \in \mathbb{R}^m$ obeying

$$\|\tilde{x}\|_{\ell_1} \le \|x\|_{\ell_1}, \text{ and } \|\Phi\tilde{x} - \Phi x\|_{\ell_2} \le 2\varepsilon \tag{V.8}$$

also obeys

$$\|\tilde{x} - x\|_{\ell_2} \le \frac{\sqrt{6}\varepsilon}{a_{3k}(\Phi) - \frac{1}{\sqrt{2}} b_{2k}(\Phi)}, \tag{V.9}$$

provided that x is k-sparse with k such that $a_{3k}(\Phi) - \frac{1}{\sqrt{2}} b_{2k}(\Phi) > 0$.

The proof follows the same steps as that of Theorem 1.1 in [3], and is omitted. In particular, it follows from [3, eq. (2.6)] with $M = 2|T_0|$ and $a_{M+|T_0|}$ (resp., b_M) in place of $\sqrt{1-\delta_{|T_0|+M}}$ (resp., $\sqrt{1+\delta_M}$) in the definition of $C_{|T_0|,M}$.

1) Proof of Theorem 2.2: Recall that the solution $(\hat{x}, \hat{z})$ to (P_2) obeys (II.4) where $\hat{e}$ is the solution to (P_2'). Replacing y in (II.4) with $Ax + e + z$ gives

$$\begin{aligned}\hat{x} - x &= (A^*A)^{-1}A^*(e-\hat{e}) + (A^*A)^{-1}A^*z \\ &= (A^*A)^{-1}A^*(e-\hat{e}) + x^{\text{Ideal}} - x \end{aligned} \tag{V.10}$$

and since $A^*A = I$

$$\|\hat{x} - x\|_{\ell_2} \le \|A^*(e-\hat{e})\|_{\ell_2} + \|x^{\text{Ideal}} - x\|_{\ell_2}.$$

To prove (II.7), it then suffices to show that $\|e-\hat{e}\|_{\ell_2} \le \frac{C\,\varepsilon}{\sqrt{1-\frac{n}{m}}}$ since the 2-norm of A^* is at most 1.

By assumption, $\|Q^*(y-e)\|_{\ell_2} = \|Q^*z\|_{\ell_2} \le \varepsilon$ and thus, e is feasible for (P_2') which implies $\|\hat{e}\|_{\ell_1} \le \|e\|_{\ell_1}$. Moreover

$$\|Q^*e - Q^*\hat{e}\|_{\ell_2} \le \|Q^*(y-e)\|_{\ell_2} + \|Q^*(y-\hat{e})\|_{\ell_2} \le 2\varepsilon.$$

We then apply Lemma 5.4 (with $\Phi = Q^*$) and obtain

$$\|e-\hat{e}\|_{\ell_2} \le \frac{\sqrt{6}\varepsilon}{a_{3k}(Q^*) - \frac{1}{\sqrt{2}} b_{2k}(Q^*)}. \tag{V.11}$$

Now since the $m \times m$ matrix obtained by concatenating the columns of A and Q is an isometry, we have

$$\|A^*x\|_{\ell_2}^2 + \|Q^*x\|_{\ell_2}^2 = \|x\|_{\ell_2}^2, \quad \forall\, x \in \mathbb{R}^m$$

whence

$$\begin{aligned} a_k^2(Q^*) &= 1 - b_k^2(A^*) \\ b_k^2(Q^*) &= 1 - a_k^2(A^*). \end{aligned}$$

Assuming that $a_{3k}(Q^*) \ge \frac{1}{\sqrt{2}} b_{2k}(Q^*)$, we deduce from (V.11) that

$$\begin{aligned}\|e-\hat{e}\|_{\ell_2} &\le \sqrt{6}\varepsilon \cdot \frac{a_{3k}(Q^*) + \frac{1}{\sqrt{2}} b_{2k}(Q^*)}{1 - b_{3k}^2(A^*) - \frac{1}{2}(1 - a_{2k}^2(A^*))} \\ &\le 2\sqrt{6}\varepsilon \cdot \frac{a_{3k}(Q^*)}{\frac{1}{2} + \frac{1}{2}a_{2k}^2(A^*) - b_{3k}^2(A^*)}. \end{aligned} \tag{V.12}$$

Recall that (δ_k) are the restricted isometry constants of $\sqrt{\frac{m}{n}}\,A^*$, and observe that by definition for each $k = 1, 2, \ldots$

$$a_k^2(A^*) \ge \frac{n}{m}(1-\delta_k), \quad b_k^2(A^*) \le \frac{n}{m}(1+\delta_k).$$

It follows that the denominator on the right-hand side of (V.12) is greater or equal to

$$\begin{aligned}\frac{1}{2} + \frac{n}{2m}(1-\delta_{2k}) - \frac{n}{m}(1+\delta_{3k}) \\ = \frac{1}{2}\left(1 - \frac{n}{m}\right) - \frac{n}{m}\left(\delta_{3k} + \frac{1}{2}\delta_{2k}\right).\end{aligned}$$

Now suppose that for some $0 < c < 1$

$$\delta_{3k} + \frac{1}{2}\delta_{2k} \le \frac{c}{2} \cdot \left(\frac{m}{n} - 1\right).$$

This automatically implies $a_{3k}(Q^*) \ge \frac{1}{\sqrt{2}} b_{2k}(Q^*)$, and the denominator on the right-hand side of (V.12) is greater or equal to $\frac{1}{2}(1-c)\left(1-\frac{n}{m}\right)$. The numerator obeys

$$a_{3k}^2(Q^*) = 1 - b_{3k}^2(A^*) \le 1 - a_{3k}^2(A^*) \le 1 - (1-\delta_{3k})\frac{n}{m}.$$

Since $\frac{n}{m}\,\delta_{3k} \le \frac{c}{2}\left(1-\frac{n}{m}\right)$, we also have $a_{3k}^2(Q^*) \le \left(1+\frac{c}{2}\right)\left(1-\frac{n}{m}\right)$. In summary, (V.12) gives

$$\|e-\hat{e}\|_{\ell_2} \le C_2 \cdot \frac{\varepsilon}{\sqrt{1-\frac{n}{m}}}$$

where one can take C_2 as $4\sqrt{6(1+c/2)}/(1-c)$. This establishes the first part of the claim.

We now turn to the second part of the theorem and argue that if the orthonormal columns of A are chosen uniformly at random, the error bound (II.7) is valid as long as we have a constant fraction of gross errors. Put $r = m - n$ and let X be an m by r matrix with independent Gaussian entries with mean 0 and variance $1/m$. Consider now the reduced singular value decomposition of X

$$X = U\Sigma V^*, \quad U \in \mathbb{R}^{m\times r}, \quad \text{and } \Sigma, V \in \mathbb{R}^{r\times r}.$$

Then the columns of U are r orthonormal vectors selected uniformly at random and thus U and Q have the same distribution. Thus, we can think of Q as being the left singular vectors of a Gaussian matrix X with independent entries. From now on, we identify U with Q. Observe now that

$$\begin{aligned}\|X^*(\hat{e}-e)\|_{\ell_2} &= \|V\Sigma Q^*(\hat{e}-e)\|_{\ell_2} = \|\Sigma Q^*(\hat{e}-e)\|_{\ell_2}\\ &\le \sigma_1(X)\|Q^*(\hat{e}-e)\|_{\ell_2}\end{aligned}$$

where $\sigma_1(X)$ is the largest singular value of X. The singular values of Gaussian matrices are well concentrated and a classical result [13, Theorem 2.13] shows that

$$\boldsymbol{P}\left(\sigma_1(X) > 1 + \sqrt{\frac{r}{m}} + t\right) \le e^{-mt^2/2}. \tag{V.13}$$

By choosing $t = 1$ in the above formula, we have

$$\|X^*(\hat{e}-e)\|_{\ell_2} \le 3\|Q^*(\hat{e}-e)\|_{\ell_2} \le 6\varepsilon$$

with probability at least $1 - e^{-m/2}$ since $\|Q^*(\hat{e}-e)\|_{\ell_2} \le 2\epsilon$. We now apply Lemma 5.4 with $\Phi = X^*$, which gives

$$\begin{aligned}\|e-\hat{e}\|_{\ell_2} &\le \frac{3\sqrt{6}\varepsilon}{a_{3k}(X^*) - \frac{1}{\sqrt{2}}b_{2k}(X^*)}\\ &= \sqrt{\frac{m}{r}}\cdot\frac{3\sqrt{6}\varepsilon}{a_{3k}(Y^*) - \frac{1}{\sqrt{2}}b_{2k}(Y^*)}\end{aligned} \tag{V.14}$$

where $Y = \sqrt{\frac{m}{r}}X$. The theorem is proved since it is well known that if $k \le c_0 \cdot r/\log(m/r)$ for some constant c_0, we have $a_{3k}(Y^*) - \frac{1}{\sqrt{2}}\, b_{2k}(Y^*) \ge c_1$ with probability at least $1 - O(e^{-\gamma' r})$ for some universal constants c_1 and γ; this follows from available bounds on the restricted isometry constants of Gaussian matrices [1], [14]–[16].

2) Proof of Corollary 2.3: First, we can just assume that $\sigma = 1$ as the general case is treated by a simple rescaling. Put $r = m - n$. Since the random vector z follows a multivariate normal distribution with mean zero and covariance matrix I_m (I_m is the identity matrix in m dimensions), Q^*z is also multivariate normal with mean zero and covariance matrix $Q^*Q = I_r$. Consequently, $\|Q^*z\|_{\ell_2}^2$ is distributed as a chi-squared variable with r degrees of freedom. Pick $\lambda = \gamma\sqrt{r}$ in (V.1), and obtain

$$\boldsymbol{P}\left(\|Q^*z\|_{\ell_2}^2 \ge (1 + \gamma\sqrt{2} + \gamma^2)r\right) \le e^{-\gamma^2 r/2}.$$

With $t = \gamma\sqrt{2} + \gamma^2$ so that $\gamma = \left(\sqrt{1+2t} - 1\right)/\sqrt{2}$, we have $\|Q^*z\|_{\ell_2} \le \sqrt{r(1+t)}$ with probability at least $1 - e^{-\gamma^2(m-n)/2}$. On this event, Theorem 2.2 asserts that

$$\|\hat{x} - x\|_{\ell_2} \le C\sqrt{m(1+t)} + \|x - x^{\mathrm{Ideal}}\|_{\ell_2}.$$

This essentially concludes the proof of the corollary since the size of $\|x - x^{\mathrm{Ideal}}\|_{\ell_2}$ is about $\sqrt{n}$. Indeed, $\|x - x^{\mathrm{Ideal}}\|_{\ell_2}^2 = \|A^*z\|_{\ell_2}^2 \sim \chi_n^2$ as observed earlier. As a consequence, for each $t_0 > 0$, we have $\|x - x^{\mathrm{Ideal}}\|_{\ell_2} \le \sqrt{n(1+t_0)}\cdot\sigma$ with probability at least $1 - e^{-\gamma_0^2 n/2}$, where γ_0 is the same function of t_0 as before. Selecting t_0 as $t_0 = m/n$, say, gives the result.

D. The LP Decoder

Before we begin, we introduce the number $\theta_{k,k'}$ of a matrix $\Phi \in \mathbb{R}^{r\times m}$ for $k + k' \le m$ called the k, k'-restricted orthogonality constants. This is the smallest quantity such that

$$|\langle \Phi v, \Phi v'\rangle| \le \theta_{k,k'} \cdot \|v\|_{\ell_2}\|v'\|_{\ell_2} \tag{V.15}$$

holds for all k and k'-sparse vectors supported on disjoint sets. Small values of restricted orthogonality constants indicate that disjoint subsets of columns span nearly orthogonal subspaces. The following lemma which relates the number $\theta_{k,k'}$ to the extremal singular values will prove useful.

Lemma 5.5: For any matrix $\Phi \in \mathbb{R}^{r\times m}$, we have

$$\theta_{k,k'}(\Phi) \le \frac{1}{2}\left(b_{k+k'}^2(\Phi) - a_{k+k'}^2(\Phi)\right).$$

Proof: Consider two vectors v and v' which are, respectively, k and k'-sparse. By definition, we have

$$\begin{aligned}2a_{k+k'}^2(\Phi) \le \|\Phi v + \Phi v'\|_{\ell_2}^2 \le 2b_{k+k'}^2(\Phi)\\ 2a_{k+k'}^2(\Phi) \le \|\Phi v - \Phi v'\|_{\ell_2}^2 \le 2b_{k+k'}^2(\Phi)\end{aligned}$$

and the conclusion follows from the parallelogram identity

$$\begin{aligned}|\langle \Phi v, \Phi v'\rangle| &= \frac{1}{4}\left|\|\Phi v + \Phi v'\|_{\ell_2}^2 - \|\Phi v - \Phi v'\|_{\ell_2}^2\right|\\ &\le \frac{1}{2}\left(b_{k+k'}^2(\Phi) - a_{k+k'}^2(\Phi)\right).\end{aligned}$$

□

The argument underlying Theorem 3.1 uses an intermediate result whose proof may be found in the Appendix. Here and in the remainder of this paper, x_I is the restriction of the vector x to an index set I, and for a matrix X, X_I is the submatrix formed by selecting the columns of X with indices in I.

Lemma 5.6: Let Φ be an $r \times m$-dimensional matrix and suppose T_0 is a set of cardinality k. For a vector $h \in \mathbb{R}^m$, we let T_1 be the k' largest positions of h outside of T_0. Put $T_{01} = T_0 \cup T_1$ and let $\Phi_{T_{01}}^*$ and $h_{T_{01}}$ be the coordinate restrictions of Φ^* and h to T_{01}, respectively. Then

$$\|h_{T_{01}}\|_{\ell_2} \le \frac{1}{a_{k+k'}^2(\Phi)}\|\Phi_{T_{01}}^*\Phi h\|_{\ell_2} + \frac{\theta_{k',k+k'}(\Phi)}{a_{k+k'}^2(\Phi)\sqrt{k'}}\|h_{T_0^c}\|_{\ell_1} \tag{V.16}$$

and

$$\|h\|_{\ell_2}^2 \le \|h_{T_{01}}\|_{\ell_2}^2 + \frac{1}{k'}\|h_{T_0^c}\|_{\ell_1}^2. \tag{V.17}$$

1) Proof of Theorem 3.1: Just as before, it suffices to show that $\|e-\hat{e}\|_{\ell_2} \le C\sqrt{k}\cdot\lambda\cdot(1-n/m)^{-1}$. Set $h=\hat{e}-e$ and let T_0 be the support of e (which has size k). Because e is feasible for (P'_∞), we have on the one hand $\|\hat{e}\|_{\ell_1}\le\|e\|_{\ell_1}$, which gives

$$\|e_{T_0}\|_{\ell_1}-\|h_{T_0}\|_{\ell_1}+\|h_{T_0^c}\|_{\ell_1}\le\|e+h\|_{\ell_1}\le\|e\|_{\ell_1} \Rightarrow \|h_{T_0^c}\|_{\ell_1}\le\|h_{T_0}\|_{\ell_1}.$$

Note that this has an interesting consequence since

$$\|h_{T_0^c}\|_{\ell_1}\le\|h_{T_0}\|_{\ell_1}\le\sqrt{k}\cdot\|h_{T_0}\|_{\ell_2} \tag{V.18}$$

by Cauchy–Schwarz. On the other hand

$$\|QQ^*h\|_{\ell_\infty}\le\|QQ^*(\hat{e}-y)\|_{\ell_\infty}+\|QQ^*(y-e)\|_{\ell_\infty}\le 2\lambda. \tag{V.19}$$

The ingredients are now in place to establish the claim. We set $k'=k$, apply Lemma 5.6 with $\Phi=Q^*$ to the vector $h=\hat{e}-e$, and obtain

$$\|h\|_{\ell_2}\le\sqrt{2}\|h_{T_{01}}\|_{\ell_2},\ \text{and}$$
$$\|h_{T_{01}}\|_{\ell_2}\le\frac{1}{a^2_{2k}(Q^*)-\theta_{k,2k}(Q^*)}\|Q_{T_{01}}Q^*h\|_{\ell_2}. \tag{V.20}$$

Since each component of $Q_{T_{01}}Q^*h$ is at most equal to 2λ, see (V.19), we have $\|Q_{T_{01}}Q^*h\|_{\ell_2}\le\sqrt{2k}\cdot 2\lambda$. We then conclude from Lemma 5.5 that

$$\|h\|_{\ell_2}\le 2\sqrt{k}\cdot\frac{2\lambda}{a^2_{2k}(Q^*)+\frac{1}{2}a^2_{3k}(Q^*)-\frac{1}{2}b^2_{3k}(Q^*)}. \tag{V.21}$$

For each k, recall the relations $a_k^2(Q^*)=1-b_k^2(A^*)$ and $b_k^2(Q^*)=1-a_k^2(A^*)$ which give

$$\|h\|_{\ell_2}\le 4\sqrt{k}\cdot\frac{\lambda}{D},$$
$$D:=1-b^2_{2k}(A^*)-\frac{1}{2}b^2_{3k}(A^*)+\frac{1}{2}a^2_{3k}(A^*).$$

Now just as before, it follows from our definitions that for each k, $b_k^2(A^*)\le\frac{n}{m}(1+\delta_k)$ and $a_k^2(A^*)\ge\frac{n}{m}(1-\delta_k)$. These inequalities imply

$$D\ge 1-\frac{n}{m}(1+\delta_{2k}+\delta_{3k}).$$

Therefore, if one assumes that

$$\delta_{2k}+\delta_{3k}\le c\left(\frac{m}{n}-1\right),$$

for some fixed constant $0<c<1$, then

$$\|e-\hat{e}\|_{\ell_2}=\|h\|_{\ell_2}\le\frac{4\sqrt{k}}{1-c}\cdot\frac{\lambda}{1-\frac{n}{m}}.$$

This establishes the first part of the theorem.

We turn to the second part of the claim; if the orthonormal columns of A are chosen uniformly at random, we show that the error bound (III.5) is valid with large probability as long as we have a constant fraction of gross errors. To do this, it suffices to show that the denominator D in (V.21) obeys

$$D\ge\frac{3}{2}a^2_{3k}(Q^*)-\frac{1}{2}b^2_{3k}(Q^*)\ge\frac{r}{2m}.$$

This follows from Lemma 5.3. If k is sufficiently small, we have that $a^2_{3k}(Q^*)\ge(7/8)^2r/m$ and $b^2_{3k}(Q^*)\le(9/8)^2r/m$ except on a set of exponentially small probability, which gives

$$D\ge\frac{r}{m}\left(\frac{3}{2}\left(\frac{7}{8}\right)^2-\frac{1}{2}\left(\frac{9}{8}\right)^2\right)\ge\frac{r}{2m}.$$

2) Proof of Corollary 3.2: First, we can just assume that $\sigma=1$ as the general case is treated by a simple rescaling. The random vector QQ^*z follows a multivariate normal distribution with mean zero and covariance matrix QQ^*. In particular, $(QQ^*z)_i\sim N(0,s_i^2)$, where $s_i^2=(QQ^*)_{i,i}$. This implies that $z'_i=(QQ^*z)_i/s_i$ is standard normal with density $\phi(t)=(2\pi)^{-1/2}e^{-t^2/2}$. For each i, $\boldsymbol{P}(|z'_i|>t)\le\phi(t)/t$ and thus

$$\boldsymbol{P}\left(\sup_{1\le i\le m}|z'_i|\ge t\right)\le 2m\cdot\phi(t)/t.$$

With $t=\sqrt{2\log m}$, this gives

$$\boldsymbol{P}\left(\sup_{1\le i\le m}|z'_i|\ge\sqrt{2\log m}\right)\le 1/\sqrt{\pi\log m}.$$

Better bounds are possible but we will not pursue these refinements here. Observe now that $s_i^2=\|Q_{i,\cdot}\|_{\ell_2}^2=1-\|A_{i,\cdot}\|_{\ell_2}^2$, and since $\lambda_i=\sqrt{2\log m}\,\|Q_{i,\cdot}\|_{\ell_2}$, we have that

$$|QQ^*z_i|\le\lambda_i,\quad\forall\, i \tag{V.22}$$

with probability at least $1-1/\sqrt{\pi\log m}$. On the event(V.22), Theorem 3.1 then shows that

$$\|\hat{x}-x\|_{\ell_2}\le C\sqrt{k}\cdot(m/r)\cdot\max_i|\lambda_i|+\|x-x^{\mathrm{Ideal}}\|_{\ell_2}. \tag{V.23}$$

We claim that

$$\frac{\max_i|\lambda_i|}{\sqrt{2\log m}}=\max_i\|Q_{i,\cdot}\|_{\ell_2}\le\sqrt{\frac{3r}{m}} \tag{V.24}$$

with probability at least $1-2e^{-\gamma m}$ for some positive constant γ. Combining (V.23) and(V.24) yields

$$\|\hat{x}-x\|_{\ell_2}\le 2C\cdot\sqrt{\frac{m\log m}{m-n}}\cdot\sqrt{k}+\|x-x^{\mathrm{Ideal}}\|_{\ell_2}.$$

This would essentially conclude the proof of the corollary since the size of $\|x-x^{\mathrm{Ideal}}\|_{\ell_2}$ is about $\sqrt{n}$. Exact bounds for $\|x-x^{\mathrm{Ideal}}\|_{\ell_2}$ are found in the proof of Corollary 2.3 and we do not repeat the argument.

It remains to check why (V.24) is true. For $r\ge m/3$ and since $\|Q_{i,\cdot}\|_{\ell_2}\le 1$, the claim holds with probability 1 because $3r/m\ge 1$! For $r\le m/3$, it follows from $\|Q_{i,\cdot}\|_{\ell_2}^2+\|A_{i,\cdot}\|_{\ell_2}^2=1$ that

$$\boldsymbol{P}\left(\max_i\|Q_{i,\cdot}\|_{\ell_2}^2\ge\frac{2r}{m}\right)=\boldsymbol{P}\left(\min_i\|A_{i,\cdot}\|_{\ell_2}^2\le\frac{n}{m}\left(1-\frac{r}{n}\right)\right)$$
$$\le m\boldsymbol{P}\left(\|A_{1,\cdot}\|_{\ell_2}^2\le\frac{n}{m}\left(1-\frac{r}{n}\right)\right).$$

The claim follows by applying Lemma 5.2 since $r/n\le 1/2$.

VI. Discussion

We have introduced two decoding strategies for recovering a block $x\in\mathbb{R}^n$ of n pieces of information from a codeword Ax which has been corrupted both by adversary and small errors. Our methods are concrete, efficient, and guaranteed to perform well. Because we are working with real-valued inputs, we emphasize that this work has nothing to do with the use

of linear programming methods proposed by Feldman and his colleagues to decode binary codes such as turbo codes or low-density parity-check codes [17]–[19]. Instead, it has much to do with the recent literature on compressive sampling or compressed sensing [14], [20]–[24], see also [25], [26] for related work.

On the practical end, we truly recommend using the two-step refinement discussed in Section IV—the reprojection step—as this really tends to enhance the performance. We anticipate that other tweaks of this kind might also work and provide additional enhancement. On the theoretical end, we have not tried to obtain the best possible constants and there is little doubt that a more careful analysis will provide sharper constants. Also, we presented some results for coding matrices with orthonormal columns for ease of exposition but this is unessential. In fact, our results can be extended to nonorthogonal matrices. For instance, one could just as well obtain similar results for $m \times n$ coding matrices A with independent Gaussian entries. There are also variations on how one might want to decode. We focused on constraints of the form $\|P_{V^\perp}\tilde{z}\|$ where $\|\cdot\|$ is either the ℓ_2 norm or the ℓ_∞ norm, and $P_{V^\perp}$ is the orthoprojector onto $V^\perp$, the orthogonal subspace to the column space of A. But one could also imagine choosing other types of constraints, e.g., of the form $\|X^*\tilde{z}\|_{\ell_2} \le \varepsilon$ for (P_2) or $\|XX^*\tilde{z}\|_{\ell_\infty} \le \lambda$ for (P_∞) (or constraints about the individual magnitudes of the coordinates $(XX^*\tilde{z})_i$ in the more general formulation), where the columns of X span $V^\perp$. In fact, one could choose the decoding matrix X *first*, and then A so that the ranges of A and X are orthogonal. Choosing $X \in R^{m\times r}$ with i.i.d. mean-zero Gaussian entries and applying the LP decoder with a constraint on $\|XX^*\tilde{z}\|_{\ell_\infty}$ instead of $\|\tilde{z}\|_{\ell_\infty}$ would simplify the argument since restricted isometry constants for Gaussian matrices are already readily available [1], [14]–[16]!

Finally, we discussed the use of coding matrices which have fast algorithms, thus enabling large scale problems. Exploring further opportunities in this area seems a worthy pursuit.

APPENDIX
PROOF OF LEMMA 5.6

The proof is a variation on that of Lemma 3.1 in [4]. In the sequel, $T_0 \subset \{1,\dots,m\}$ is a set of size k, T_1 is the k' largest positions of h outside of T_0, $T_{01} = T_0 \cup T_1$. Next, divide T_0^c into subsets of size k' and enumerate T_0^c as $n_1, n_2, \dots, n_{m-|T_0|}$ in decreasing order of magnitude of $h_{T_0^c}$. Set $T_j = \{n_\ell, (j-1)k'+1 \le \ell \le jk'\}$. That is, T_1 is as before and contains the indices of the k' largest coefficients of $h_{T_0^c}$, T_2 contains the indices of the next k' largest coefficients, and so on.

Observe that $\Phi h_{T_{01}} = \Phi h - \sum_{j\ge 2}\Phi h_{T_j}$ so that

$$\|\Phi h_{T_{01}}\|_{\ell_2}^2 = \langle \Phi h_{T_{01}}, \Phi h\rangle - \sum_{j\ge 2}\langle \Phi h_{T_{01}}, \Phi h_{T_j}\rangle.$$

On the one hand, we have

$$|\langle \Phi h_{T_{01}}, \Phi h\rangle| = \langle h_{T_{01}}, \Phi^*_{T_{01}}\Phi h\rangle \le \|h_{T_{01}}\|_{\ell_2}\|\Phi^*_{T_{01}}\Phi h\|_{\ell_2}$$

and on the other

$$|\langle \Phi h_{T_{01}}, \Phi h_{T_j}\rangle| \le \theta_{k+k',k'}\|h_{T_{01}}\|_{\ell_2}\|h_{T_j}\|_{\ell_2}.$$

This gives

$$\begin{aligned} a_{k+k'}^2\|h_{T_{01}}\|_{\ell_2}^2 &\le \|\Phi h_{T_{01}}\|_{\ell_2}^2 \\ &= \|h_{T_{01}}\|_{\ell_2}\left(\|\Phi^*_{T_{01}}\Phi h\|_{\ell_2} + \theta_{k+k',k'}\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}\right) \end{aligned} \quad \text{(A1)}$$

where for simplicity, we have omitted the dependence on Φ in the constants $a_k(\Phi)$ and $\theta_{k,k'}(\Phi)$. We then develop an upper bound on $\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2}$ as in [3]. By construction, the magnitude of each coefficient in T_{j+1} is less than the average of the magnitudes in T_j

$$\|h_{T_{j+1}}\|_{\ell_\infty} \le \|h_{T_j}\|_{\ell_1}/k' \Rightarrow \|h_{T_{j+1}}\|_{\ell_2}^2 \le \|h_{T_j}\|_{\ell_1}^2/k'.$$

Therefore

$$\sum_{j\ge 2}\|h_{T_j}\|_{\ell_2} \le \sum_{j\ge 1}\|h_{T_j}\|_{\ell_1}/\sqrt{k'} = \|h\|_{\ell_1(T_0^c)}/\sqrt{k'}. \quad \text{(A2)}$$

Hence, we deduce from (A1) that

$$\|h_{T_{01}}\|_{\ell_2} \le \frac{\|\Phi^*_{T_{01}}\Phi h\|_{\ell_2}}{a_{k+k'}^2} + \frac{\theta_{k+k',k'}\|h\|_{\ell_1(T_0^c)}}{a_{k+k'}^2\sqrt{k'}},$$

which proves the first part of the lemma.

For the second part, it follows from (A2) that

$$\|h_{T_{01}^c}\|_{\ell_2} = \left\|\sum_{j\ge 2}h_{T_j}\right\|_{\ell_2} \le \sum_{j\ge 2}\|h_{T_j}\|_{\ell_2} \le \|h_{T_0^c}\|_{\ell_1}/\sqrt{k'}.$$

ACKNOWLEDGMENT

E. J. Candès would like to thank the Centre Interfacultaire Bernoulli of the Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, for hospitality during June and July 2006. The authors would like to thank Mike Wakin for his careful reading of the manuscript and the anonymous referees and the Associate Editor for their useful comments.

REFERENCES

[1] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[2] E. Candès, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *Proc. 46th Annu. IEEE Symp. Foundations of Computer Science (FOCS)*, Pittsburgh, PA, Oct. 2005, pp. 295–308.
[3] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[4] E. Candès and T. Tao, "The Dantzig selector: Statistical estimation when p is much larger than n," *Ann. Statist.*, vol. 35, no. 6, pp. 2313–2351, 2007.
[5] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[6] D. Donoho, "Neighborly polytopes and sparse solutions of underdetermined linear equations," Stanford Univ., Stanford, CA, 2005, Tech. Rep.

[7] C. Dwork, F. McSherry, and K. Talwar, "The price of privacy and the limits of LP decoding," in *STOC '07: Proc. 39th Annu. ACM Symp. Theory of Computing*, New York, NY, 2007, pp. 85–94.
[8] E. Candès and J. Romberg, "Practical signal recovery from random projections," in *Proc. SPIE Int. Symp. Electoric Imaging: Computational Imaging III*, 2005.
[9] B. Laurent and P. Massart, "Adaptive estimation of a quadratic functional by model selection," *Ann. Statist.*, vol. 28, no. 5, pp. 1302–1338, 2000.
[10] I. M. Johnstone, "Chi-square oracle inequalities," in *State of the Art in Probability and Statistics (Leiden, 1999)*, ser. IMS Lecture Notes Monographs Ser.. Beachwood, OH: Inst. Math. Statist., 2001, vol. 36, pp. 399–418.
[11] A. Barvinok, *A Course in Convexity*, ser. Graduate Studies in Mathematics. Providence, RI: AMS, 2002, vol. 54.
[12] G. Pisier, *The Volume of Convex Bodies and Banach Space Geometry*, ser. Cambridge Tracts in Mathematics. Cambridge, U.K.: Cambridge Univ. Press, 1989, vol. 94.
[13] K. R. Davidson and S. J. Szarek, "Addenda and corrigenda to: Local operator theory, random matrices and banach spaces," in *Handbook of the Geometry of Banach Spaces*. Amsterdam, The Netherlands: North-Holland, 2001, vol. I, pp. 317–366.
[14] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[15] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Commun. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, 2006.
[16] M. Rudelson and R. Vershynin, "Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements," in *Proc. 2006 40th. Annu. Conf. Information Sciences and Systems,* , Mar. 2006, pp. 207–212.
[17] J. Feldman and D. R. Karger, "Decoding turbo-like codes via linear programming," *J. Comput. Syst. Sci.*, vol. 68, no. 4, pp. 733–752, 2004.
[18] J. Feldman and C. Stein, "LP decoding achieves capacity," in *Proc. 16th Annu. ACM-SIAM Symp. Discrete Algorithms*, New York, 2005, pp. 460–469.
[19] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright, "LP decoding corrects a constant fraction of errors," *IEEE Trans. Inf. Theory*, vol. 53, no. 1, pp. 82–89, Jan. 2007.
[20] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[21] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[22] J. Tropp and A. Gilbert, "Signal Recovery From Partial Information via Orthogonal Matching Pursuit," 2005, unpublished manuscript.
[23] A. Cohen, W. Dahmen, and R. DeVore, "Compressed Sensing and the Best k-Term Approximation 2006, unpublished manuscript.
[24] M. Rudelson and R. Vershynin, "Geometric approach to error-correcting codes and reconstruction of signals," *Int. Math. Res. Not.*, no. 64, pp. 4019–4041, 2005.
[25] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.
[26] P. Marziliano, M. Vetterli, and T. Blu, "Sampling and exact reconstruction of bandlimited signals with additive shot noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2230–2233, May 2006.