# The Fundamental Limits of Stable Recovery in Compressed Sensing

Galen Reeves
Departments of Electrical & Computer Engineering and Statistical Science
Duke University

***Abstract*—Compressed sensing has shown that a wide variety of structured signals can be recovered from a limited number of noisy linear measurements. This paper considers the extent to which such recovery is robust to signal and measurement uncertainty. The main result is a non-asymptotic upper bound on the reconstruction error in terms of two key quantities: the best approximation error of the signal (with respect to a user-defined approximation set) and the measurement error. We assume a random Gaussian sensing matrix but place no restrictions on the signal or the noise. This result provides a simple and yet powerful framework for analyzing the fundamental limits of stable recovery, allowing us to sharpen existing results as well as derive new ones.**

## I. Introduction

Consider the problem of recovering an unknown signal $\mathbf{x}^\star \in \mathbb{R}^n$ from measurements $\mathbf{y} \in \mathbb{R}^m$ generated by the linear observation model

$$\mathbf{y} = A\mathbf{x}^\star + \mathbf{w} \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$ is a known sensing matrix which compresses the signal from $n$ to $m < n$ dimensions and $\mathbf{w}$ is unknown noise. Given the observations $(\mathbf{y}, A)$ we seek a reconstruction $\hat{\mathbf{x}}$ that is close to $\mathbf{x}^\star$ in mean-squared error.

This problem has been studied extensively in the compressed sensing literature. A major focus of previous work has been the tradeoff between the number of measurements and the reconstruction error under a variety of signal models — e.g. sparse signals [1]–[3], compressible signals [4], [5], unions of subspaces [6]–[10], random discrete-continuous mixtures [11], [12], and block-sparse models [13].

One approach (e.g. [1], [6]–[10]) has been to derive sufficient conditions in terms of certain properties of the measurement matrix, such as the restricted isometry property, the mutual incoherence, or generalizations thereof. Conditional on these properties holding, it is shown that the number of measurements required for stability is proportional to the number of "degrees of freedom" in the signal class. One limitation of this approach, however, is that the resulting conditions are loose — stability is guaranteed only if the number of measurements exceeds a critical cutoff point, and this cutoff point does not match the number of measurements needed for exact recovery in the absence of noise.[1]

[1]For example, although it is well known that a $k$-sparse signal can be recovered from $m = k + 1$ randomly generated *noiseless* measurements, results based on the restricted isometry property require $m = \Omega(k \log(n/k))$ for stable recovery in the noisy setting (see e.g. [1], [2] and subsequent work).

TABLE I
Examples of the stability constant given in (3) .

| measurements ($m$) | max. dim. ($d$) | no. subspaces ($N$) | stability constant ($C$) |
|---|---|---|---|
| 12 | 2 | 5 | 39.42 |
| 15 | 10 | 100 | $2.54 \times 10^3$ |
| 150 | 100 | 1000 | 25.92 |
| 1500 | 1000 | $10^7$ | 6.3 |

Another approach (e.g. [12], [14]–[18]) has been to study the fundamental behavior for random signals generated from a known distribution (or class of distributions). In these cases, careful analysis has shown the exact location of certain phase transitions in the large system setting. These results, however, are asymptotic and require prior knowledge of a distribution.

The main result of this paper is a non-asymptotic bound that overcomes the limitations outlined above; this bound provides a sharp characterization of recovery that makes no assumptions about the signal $\mathbf{x}$ or the noise $\mathbf{w}$.

To state these bounds, we focus on approximation sets which can be expressed as the union of a finite number of affine linear subspaces:

$$\mathcal{M} = \bigcup_{i=1}^{N} \mathcal{M}_i, \quad \mathcal{M}_i = \{\mathbf{b}_i + \operatorname{span}(B_i)\}, \quad \dim(\mathcal{M}_i) \le d.$$

The complexity is measured by the maximum dimension $d$ and number of subspaces $N$. The best approximation error of signal $\mathbf{x} \in \mathbb{R}^n$ is defined by the squared distance:

$$\rho^2_{\mathcal{M}}(\mathbf{x}) = \inf_{\mathbf{u} \in \mathcal{M}} \|\mathbf{x} - \mathbf{u}\|^2.$$

We study the performance of the constrained least squares estimator defined by

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x} \in \mathcal{M}} \|\mathbf{y} - A\mathbf{x}\|^2. \tag{2}$$

We show that if $A$ is an i.i.d. Gaussian matrix with $m > d + 2$, then the reconstruction error of (2) obeys

$$\mathbf{E}\big[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2\big] \le (1 + C) \cdot \rho^2_{\mathcal{M}}(\mathbf{x}^\star) + C \cdot \|\mathbf{w}\|^2. \tag{3}$$

Here, expectation is with respect to the sensing matrix and the constant $C$ is given explicitly in terms of the number of measurements $m$, the maximum dimension $d$, and the number of subspaces $N$; specific examples are given in Table 1.

## II. Main Result

This section gives a precise statement of our main result and provides some examples of how it can be used to derive fundamental phase transitions.

**Theorem 1.** *Let $A$ be a random $m \times n$ matrix whose entries are i.i.d. $N(0, 1/m)$, let $\mathcal{M}$ be a union of $N$ subspaces with maximum dimension $d < m$, and let $\hat{\mathbf{x}}$ be given by (2). For any $t \geq 1$, the inequality*

$$\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2 \leq s_{m,d,N} \cdot \left(\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2\right) \cdot t - \|\mathbf{w}\|^2$$

*holds with probability at least $1 - 2\exp(-\frac{m-k}{2}\Delta(t))$ where probability is with respect to the sensing matrix and*

$$s_{m,d,N} = \tfrac{m+1}{m-d+1}\left[\mathcal{L}^{-1}_{[1,\infty)}\left(\tfrac{\log N}{m-d}\right)\right]^2$$
$$\Delta(t) = 2\mathcal{L}_{[1,\infty)}\left(\frac{t}{1+\sqrt{t\mathcal{L}_{[1,\infty)}(t)}}\right)$$

*with $\mathcal{L}_{[1,\infty)}(t) = [\log(t) + 1/t - 1]\mathbf{1}(t \in [1,\infty))$. The function $\Delta(t)$ is positive and strictly increasing on the interval $(1,\infty)$.*

We now make several remarks:

- One immediate consequence of Theorem 1 is that the necessary and sufficient conditions for *stable* recovery match the necessary and sufficient conditions for *exact* recovery in the absence of approximation and measurement errors. Specifically:

  $$\text{recovery is stable in probability} \quad \Longleftrightarrow \quad m > d.$$

  Remarkably, this condition does not depend on the signal length $n$ or the number of subspaces $N$.
- This bound is universal in the sense that it holds for any signal $\mathbf{x}^\star$ and error term $\mathbf{w}$. However, the independence of the sensing matrix is crucial — the signal and noise cannot depend on the realization of the sensing matrix.
- The constant $s_{m,d,N}$ characterizes the sensitivity to the approximation error and measurement error. Its scaling behavior is given by

  $$s_{m,d,N} \sim \frac{m+1}{m-d+1}(N/e)^{\frac{2}{m-d}}, \qquad N \to \infty.$$

- The exponent $\Delta(t)$ obeys $\Delta(t) \sim \log(t)$ as $t \to \infty$. Its behavior is illustrated in Figure 1.

**Theorem 2.** *Consider the assumptions of Theorem 1. If $m > d+2$, then the expected reconstruction error is finite and obeys*

$$\mathbf{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right] \leq (1 + C_{m,d,N}) \cdot \rho^2_{\mathcal{U}}(\mathbf{x}) + C_{m,d,N} \cdot \|\mathbf{w}\|^2$$

*where the expectation is with respect to the sensing matrix and*

$$C_{m,d,N} = s_{m,d,N} \cdot \left(1 + \int_1^\infty 2\exp(-\tfrac{m-k}{2}\Delta(t))dt\right) - 1.$$

*Proof.* Since the reconstruction error is nonnegative, we have

$$\mathbf{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2\right] = \int_0^\infty \mathbf{P}\left[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2 \geq u\right]du.$$

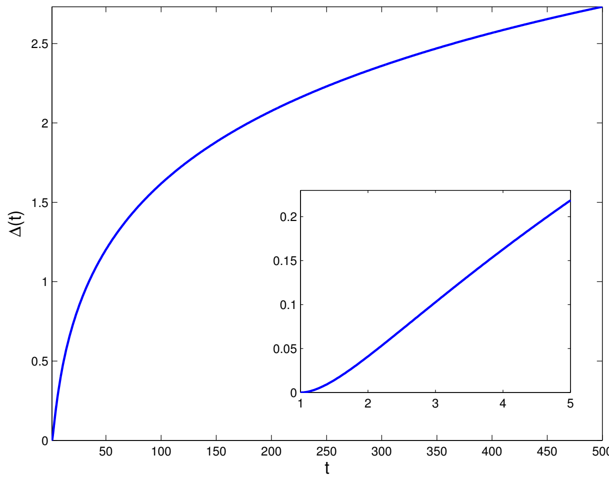


Fig. 1. Illustration of $\Delta(t)$ as a function of $t$. Note that $\Delta(t)$ is positive and strictly increasing for all $t > 1$.

Making the substitution $t = (u + \|\mathbf{w}\|^2)/s_{m,d,N} \cdot (\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2)$, and combining Theorem 1 with the fact that probabilities are upper bounded by one completes the proof. □

The following sections summarize some applications of Theorems 2. To simplify notation we write $C = C_{m,d,N}$.

### A. Improved and generalized bounds

Suppose that a subset $\mathcal{A} \subset \mathbb{R}^n$ can be approximated by a union of subspaces $\mathcal{M}$ such that $\sup_{\mathbf{x}\in\mathcal{A}} \rho_{\mathcal{M}}(\mathbf{x}) \leq \kappa$. By Theorem 2 and the triangle inequality, we then have

$$\mathbf{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2\right] \leq (1 + C) \cdot [\kappa + \rho_{\mathcal{A}}(\mathbf{x}^\star)]^2 + C \cdot \|\mathbf{w}\|^2.$$

This relationship allows us to extend our results to signal models, such as manifolds, which can be well approximated using a finite number of subspaces.

This relationship can also be used to strengthen our results by providing a tradeoff between the maximum dimension $d$ and the number of subspaces $N$. For example, suppose that $\mathcal{M} \subset \tilde{\mathcal{M}}$ where $\tilde{\mathcal{M}}$ is a union of $\tilde{N} < N$ subspaces with maximum dimension $\tilde{d} > d$. Then the reconstruction error of the estimator (3) constrained to the set $\tilde{\mathcal{M}}$ obeys

$$\mathbf{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2\right] \leq (1 + C_{m,\tilde{d},\tilde{N}}) \cdot \rho^2_{\mathcal{M}}(\mathbf{x}^\star) + C_{m,\tilde{d},\tilde{N}} \cdot \|\mathbf{w}\|^2,$$

where we have used the fact that $\rho_{\tilde{\mathcal{M}}}(\mathbf{x}) \leq \rho_{\mathcal{M}}(\mathbf{x})$. In some cases, $C_{m,\tilde{d},\tilde{N}}$ can be much smaller than $C_{m,d,N}$, thus leading to an improved bound.

### B. Optimal phase transitions for random signals

For any distribution on $\mathbf{x}^\star$, the expected reconstruction error is bounded by the expected approximation error:

$$\mathbf{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2\right] \leq (1 + C) \cdot \mathbf{E}\left[\rho^2_{\mathcal{M}}(\mathbf{x}^\star)\right] + C \cdot \|\mathbf{w}\|^2.$$

Now suppose that the signal entries are i.i.d. according to the discrete continuous mixture:

$$P_X = (1 - \varepsilon)P_d + \varepsilon P_c$$

where $P_c$ is absolutely continuous and $P_d$ is discrete with finite support of cardinality $b$. Wu and Verdu [12] showed that if $P_X$ is known, then the necessary and sufficient condition for stable recovery in the large system setting is given by:

$$\liminf_{n\to\infty} m_n/n \geq \varepsilon. \tag{4}$$

Using Theorem 2, we can show that this same result holds even if the only information we have about $P_X$ is the size of the support of $P_d$. The basic idea is that any typical realization of $\mathbf{x}^\star$ can be well approximated by the set of all vectors containing at most $\varepsilon n + b$ distinct values (see [19]). Since this set can also be represented as a finite union of subspaces, the sufficiency of (4) follows from Theorem 2 and the fact that $b$ is constant.

### C. Universal scalar quantization

Lastly, let $\mathcal{M}$ be the set of all signals of length $n$ with at most $d$ different values. Since this set can be represented as a union of $N = d^N/d!$ subspaces of maximum dimension $d$, it follows that

$$\mathbf{E}\big[\|\hat{\mathbf{x}} - \mathbf{x}\|^2\big] \leq (1+C)\cdot D_d(\mathbf{x}) + C \cdot \|\mathbf{w}\|^2$$

where $C = O\Big(\frac{m}{m-d} d^{\frac{2n}{m-d}}\Big)$ and

$$D_d(\mathbf{x}) = \min_{\mathbf{z}\in\mathbb{R}^d} \sum_{i=1}^n \min_{j\in[d]} (x_i - z_j)^2,$$

is the squared distortion of the optimal scalar quantizer using only $d$ values. To the best of our knowledge, this result is new.

## III. Distribution of least squares

This section considers recovery with respect to a single subspace and characterizes the bivariate distribution on two key quantities: the reconstruction error and the residual error.

Let $\mathcal{M} = \{\mathbf{b} + \mathrm{span}(B)\}$ be an affine linear subspace with dimension $d = \mathrm{rank}(B)$. If the number of measurements $m$ is greater than the dimension $d$, the least squares solution is unique almost surely (with respect to the distribution on $A$) and is given in closed form as:

$$\hat{\mathbf{x}} = \mathbf{b} + P(AP)^\dagger(\mathbf{y} - A\mathbf{b}) \tag{5}$$

where $P \in \mathbb{R}^{m\times d}$ an orthonormal basis for the range space of $B$ and $(\cdot)^\dagger$ denotes the Moore-Penrose pseudo inverse.

We are interested in the behavior of two random variables: the reconstruction error, defined by the sum of squared errors

$$\mathsf{SSE} = \|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2, \tag{6}$$

and the residue sum of squares, defined by

$$\mathsf{RSS} = \|\mathbf{y} - A\hat{\mathbf{x}}\|^2. \tag{7}$$

**Theorem 3.** *Let $A$ be a random $m \times n$ matrix whose entries are i.i.d. $N(0, 1/m)$. If $m > d$, then the distribution of the random pair $(\mathsf{SSE}, \mathsf{RSS})$ can be characterized as follows:*

*(a) If $\rho_{\mathcal{M}}(\mathbf{x}^\star) = 0$ then*

$$\big(\mathsf{SSE}, \mathsf{RSS}\big) \overset{D}{=} \left(\frac{\|\mathbf{w}\|^2 \cdot (1-\Theta)}{T}, \|\mathbf{w}\|^2 \cdot \Theta\right)$$

*where $T \sim \chi^2_{m-d+1}$ and $\Theta \sim \mathrm{Beta}(\frac{m-d}{2}, \frac{d}{2})$ are independent.*

*(b) If $\rho_{\mathcal{M}}(\mathbf{x}^\star) > 0$ then*

$$\big(\mathsf{SSE}, \mathsf{RSS}\big) \overset{D}{=} \left(\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \frac{\rho^2_{\mathcal{V}}(\mathbf{x}^\star)\cdot V_1}{T}, \frac{\rho^2_{\mathcal{V}}(\mathbf{x}^\star)\cdot V_2}{m}\right)$$

*where $T \sim \chi^2_{m-d+1}$ is independent of $(V_1, V_2)$ and*

$$V_1 \sim \chi^2_d\bigg(m \cdot \frac{\|\mathbf{w}\|^2}{\rho^2_{\mathcal{V}}(\mathbf{x}^\star)} \cdot (1-\Theta)\bigg)$$

$$V_2 \sim \chi^2_{m-d}\bigg(m \cdot \frac{\|\mathbf{w}\|^2}{\rho^2_{\mathcal{V}}(\mathbf{x}^\star)} \cdot \Theta\bigg)$$

$$B \sim \mathrm{Beta}\bigg(\frac{m-d}{2}, \frac{d}{2}\bigg)$$

*where $V_1$ and $V_2$ conditionally independent given $B$.*

We note that several special cases of Theorem 3 have been shown previously in the context of sparse signal models, i.e. the setting where the subspace $\mathcal{M}$ is characterized by a subset of the columns of the identity matrix. For example:

- Wainwright [20, pg. 5740] derived the marginal distribution of $\mathsf{RSS}$ in the Gaussian noise model.
- Cartis and Thomson [21, Lemma 4.3] characterized the marginal distributions of $\mathsf{SSE}$ and $\mathsf{RSS}$ in the case of noiseless measurements and also in the case of Gaussian noise and zero approximation error (i.e. $\rho_{\mathcal{V}}(\mathbf{x}) = 0$).
- Reeves and Donoho [3, Theorem 1] derived the bivariate distribution on $(\mathsf{SSE}, \mathsf{RSS})$ for the case of Gaussian noise and zero approximation error.

In addition to the work mentioned above, it is possible that other parts of Theorem 3 have also been shown previously, in some form or another. However, to the best of our knowledge, the full statement of Theorem 3 is new. The proof is omitted due to space constraints.

The next result provides closed form bounds on the cumulative distribution function of the distribution given in Theorem 3. The proof is omitted due to space constraints.

**Lemma 1.** *Under the assumptions of Theorem 3, the following bounds hold for all $r, s \in \mathbb{R}$:*

$$\mathbf{P}[\mathsf{SSE} \geq s, \mathsf{RSS} \leq r] \leq e^{-\frac{m-d+1}{2}E_1(s) - \frac{m-d}{2}E_2(r)} \tag{8}$$

$$\mathbf{P}[\mathsf{RSS} \geq r] \leq e^{-\frac{m-d}{2}E_3(r)} \tag{9}$$

*where:*

$$E_1(s) = \mathcal{L}_{[1,\infty)}\bigg(\frac{m-d+1}{m+1}\cdot\frac{s + \|\mathbf{w}\|^2}{\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2}\bigg)$$

$$E_2(r) = \mathcal{L}_{[1,\infty)}\bigg(\frac{m-d}{m}\cdot\frac{\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2}{r}\bigg)$$

$$E_3(r) = \mathcal{L}_{(0,1]}\bigg(\frac{m-d}{m}\cdot\frac{\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2}{r}\bigg).$$

*with $\mathcal{L}(x) = \log(x) + 1/x - 1$ and $\mathcal{L}_{\mathcal{A}}(x) = \mathcal{L}(x)\mathbf{1}(x \in \mathcal{A})$.*

## IV. Recovery in a union of subspaces

This section considers recovery with respect to a finite union of subspaces. Building upon the results of Section III, we derive a sequence of upper bounds on the reconstruction error of the restricted least squares estimator.

Let $\mathcal{M}$ be a union of $N$ affine subspaces with maximum dimension $d < m$. Observe that the minimization problem in (2) can be expressed in two stages as

$$\min_i \left\{ \min_{\mathbf{x} \in \mathcal{M}_i} \|\mathbf{y} - A\mathbf{x}\|^2 \right\} \tag{10}$$

where the outer minimization is over all subspaces and the inner minimization corresponds to the setting of Section III.

With probability one, each of the inner minimization problems has a unique solution $\hat{\mathbf{x}}(i)$ given in closed form by (5). The corresponding sum of squared errors and residual sum of squares are defined by

$$\mathsf{SSE}(i) = \|\hat{\mathbf{x}}(i) - \mathbf{x}^\star\|^2 \quad \text{and} \quad \mathsf{RSS}(i) = \|\mathbf{y} - A\hat{\mathbf{x}}(i)\|^2$$

Using this notation, the least squares estimate can be expressed as $\hat{\mathbf{x}}(i^\star)$ where $i^\star$ achieves the outer minimization in (10), i.e.

$$\mathsf{RSS}(i^\star) = \min_i \mathsf{RSS}(i). \tag{11}$$

Note that $i^\star$ is unique with probability one if the subspaces are distinct. The corresponding reconstruction error is $\mathsf{SSE}(i^\star)$.

The analysis given in the remainder of this section is based on the fact that we need not identify the *best* subspace in $\mathcal{M}$ in order to bound the reconstruction error in terms of $\rho^2_{\mathcal{M}}(\mathbf{x}^\star)$. This fact is supported by the following intuition:

- At low SNR, reconstruction with respect to any subspace (even one that is a bad match for the signal) will result in an error that is small relative to the noise.
- At high SNR, reconstruction with respect to a bad subspace will result in an error that is large relative to the noise. However, since the SNR is high, it is possible to identify subspaces which are a good match for the signal (though not necessarily optimal). Reconstruction with respect to these subspaces will result in an error that is small relative to the noise.

### A. An initial bound

To begin, observe that by (11), the following series of implications holds for any index $j \in [N]$ and $r, s \in \mathbb{R}$:

$$\begin{aligned} &\big\{ \mathsf{SSE}(i^\star) \geq s, \mathsf{RSS}(j) \leq r \big\} \\ \implies &\big\{ \mathsf{SSE}(i^\star) \geq s, \mathsf{RSS}(i^\star) \leq r \big\} \\ \implies &\bigcup_i \big\{ \mathsf{SSE}(i) \geq s, \mathsf{RSS}(i) \leq r \big\}. \end{aligned}$$

Thus, by the union bound, we can write

$$\begin{aligned} &\mathbf{P}[\mathsf{SSE}(i^\star) \geq s, \mathsf{RSS}(j) \leq r] \\ &\leq \sum_i \mathbf{P}[\mathsf{SSE}(i) \geq s, \mathsf{RSS}(i) \leq r], \end{aligned}$$

and by the law of total probability,

$$\begin{aligned} \mathbf{P}[\mathsf{SSE}(i^\star) \geq s] &= \mathbf{P}[\mathsf{SSE}(i^\star) \geq s, \mathsf{RSS}(j) \leq r] \\ &\quad + \mathbf{P}[\mathsf{SSE}(i^\star) \geq s, \mathsf{RSS}(j) > r] \\ &\leq \sum_i \mathbf{P}[\mathsf{SSE}(i) \geq s, \mathsf{RSS}(i) \leq r] \\ &\quad + \mathbf{P}[\mathsf{RSS}(j) > r]. \end{aligned} \tag{12}$$

Combining (12) with Theorem 1 and taking the minimum over the index $j$ gives the following result.

**Theorem 4.** *Under the assumptions of Theorem 1, the following bound holds for any $s, r \in \mathbb{R}$:*

$$\begin{aligned} \mathbf{P}\big[\|\hat{\mathbf{x}} - \mathbf{x}^\star\|^2 \geq s\big] &\leq \sum_{i=1}^N e^{-\frac{m-d_i+1}{2} E_1^{(i)}(s) - \frac{m-d_i}{2} E_2^{(i)}(r)} \\ &\quad + \min_{i \in [N]} e^{-\frac{m-d_i}{2} E_3^{(i)}(r)} \end{aligned} \tag{13}$$

*where $E_1^{(i)}(s)$, $E_2^{(i)}(r)$, and $E_3^{(i)}(r)$ correspond to the functions given Theorem 1 evaluated with $d_i$ and $\rho_{\mathcal{M}_i}(\mathbf{x}^\star)$.*

### B. Proof of Theorem 1

We now prove Theorem 1 by showing that, after an appropriate change of variables, the right hand side of (13) can be bounded uniformly for all $\mathbf{x}^\star$ and $\mathbf{w}$.

For each subspace $\mathcal{M}_i$, we define the ratio

$$\gamma_i = \frac{\rho^2_{\mathcal{M}_i}(\mathbf{x}^\star) + \|\mathbf{w}\|^2}{\rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2}. \tag{14}$$

Note that $1 \leq \gamma_i < \infty$ where equality on the left is attained for at least one subspace in $\mathcal{M}$. Also, we define

$$\begin{aligned} \tilde{s} &= \tfrac{m+1}{m-d+1} \big[ \big( \rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2 \big) \cdot s - \|\mathbf{w}\|^2 \big] \\ \tilde{r} &= \tfrac{m-d}{m} \big( \rho^2_{\mathcal{M}}(\mathbf{x}^\star) + \|\mathbf{w}\|^2 \big) \cdot r. \end{aligned}$$

Using this notation, the exponent in the first term in (13) can be bounded as follows:

$$\begin{aligned} &\tfrac{m-d_i+1}{2} E_1^{(i)}(\tilde{s}) + \tfrac{m-d_i}{2} E_2^{(i)}(\tilde{r}) \\ &= \tfrac{m-d_i+1}{2} \mathcal{L}_{[1,\infty)}\Big( \tfrac{m-d_i+1}{m-d+1} \frac{s}{\gamma_i} \Big) + \tfrac{m-d_i}{2} \mathcal{L}_{[1,\infty)}\Big( \tfrac{m-d_i}{m-d} \frac{\gamma_i}{r} \Big) \\ &\geq \tfrac{m-d}{2} \Big[ \mathcal{L}_{[1,\infty)}\Big( \frac{s}{\gamma_i} \Big) + \mathcal{L}_{[1,\infty)}\Big( \frac{\gamma_i}{r} \Big) \Big] \end{aligned}$$

where we have used the fact that $\mathcal{L}_{[1,\infty)}(x)$ is nondecreasing and the fact that $d_i \leq d$. Similarly, the exponent in the second term in (13) obeys:

$$\begin{aligned} \max_i \tfrac{m-d_i}{2} E_3^{(i)}(\tilde{r}) &= \max_i \tfrac{m-d_i}{2} \mathcal{L}_{(0,1]}\Big( \tfrac{m-d_i}{m-d} \frac{\gamma_i}{r} \Big) \\ &\geq \max_i \tfrac{m-d}{2} \mathcal{L}_{(0,1]}\Big( \tfrac{m-d}{m-d} \frac{\gamma_i}{r} \Big) \\ &= \tfrac{m-d}{2} \mathcal{L}_{(0,1]}(1/r) \end{aligned}$$

where the inequality follows from the fact that, for any $x \in \mathbb{R}$, the function $g(\alpha) = \alpha\, \mathcal{L}_{(0,1]}(\alpha\, x)$ is nonincreasing in $\alpha$.

Plugging these inequalities back into (13) leads to

$$\begin{aligned}&\mathbf{P}\Big[\|\hat{\mathbf{x}}-\mathbf{x}^\star\|^2 \ge \tfrac{m+1}{m-d+1}\big(\rho^2_{\mathcal{M}}(\mathbf{x}^\star)+\|\mathbf{w}\|^2\big)\cdot s-\|\mathbf{w}\|^2\Big]\\ &\le \sum_{i=1}^N e^{-\frac{m-d}{2}\left[\mathcal{L}_{[1,\infty)}\left(\frac{s}{\gamma_i}\right)+\mathcal{L}_{[1,\infty)}\left(\frac{\gamma_i}{r}\right)\right]}+e^{-\frac{m-d}{2}\mathcal{L}_{(0,1]}\left(\frac{1}{r}\right)}.\end{aligned}$$

Next, we upper bound each term in the summation. By differentiation, it can be shown that the worst case value of $\gamma_i$ is given by $\gamma^* = \sqrt{rs}$, and thus

$$\inf_{\gamma\ge 1}\left[\mathcal{L}_{[1,\infty)}\left(\frac{s}{\gamma}\right)+\mathcal{L}_{[1,\infty)}\left(\frac{\gamma}{r}\right)\right]=2\mathcal{L}_{[1,\infty)}\left(\sqrt{\frac{s}{r}}\right). \quad (15)$$

Applying this inequality leads to

$$\begin{aligned}&\mathbf{P}\Big[\|\hat{\mathbf{x}}-\mathbf{x}^\star\|^2 \ge \tfrac{m+1}{m-d+1}\big(\rho^2_{\mathcal{M}}(\mathbf{x}^\star)+\|\mathbf{w}\|^2\big)\cdot s-\|\mathbf{w}\|^2\Big]\\ &\le N\cdot 2e^{-\frac{m-d}{2}\left[2\mathcal{L}_{[1,\infty)}\left(\sqrt{s/r}\right)\right]}+3e^{-\frac{m-d}{2}\mathcal{L}_{(0,1]}\left(\frac{1}{r}\right)}. \quad (16)\end{aligned}$$

Note that this bound is universal in the sense that the right hand side does not depend on the signal $\mathbf{x}^\star$ or the noise $\mathbf{w}$.

To further simplify the bound, let $s$ be parametrized as:

$$s=\left[\mathcal{L}^{-1}_{[1,\infty)}\left(\tfrac{\log N}{m-d}\right)\right]^2\cdot t.$$

It is then straightforward to verify that, for all $0<r\le t$,

$$2\mathcal{L}_{[1,\infty)}\left(\sqrt{s/r}\right)\ge\frac{2\log N}{m-d}+2\mathcal{L}_{[1,\infty)}\left(\sqrt{t/r}\right). \quad (17)$$

Returning to (16), we can then write:

$$\begin{aligned}&\mathbf{P}\big[\|\hat{\mathbf{x}}-\mathbf{x}^\star\|^2 \ge s_{m,d,N}\cdot\big(\rho^2_{\mathcal{M}}(\mathbf{x}^\star)+\|\mathbf{w}\|^2\big)\cdot t-\|\mathbf{w}\|^2\big]\\ &\quad\le e^{-\frac{m-d}{2}\left[2\mathcal{L}_{[1,\infty)}\left(\sqrt{t/r}\right)\right]}+e^{-\frac{m-d}{2}\mathcal{L}_{(0,1]}\left(\frac{1}{r}\right)}\\ &\quad\le 2e^{-\frac{m-d}{2}\min\left\{2\mathcal{L}_{[1,\infty)}\left(\sqrt{t/r}\right),\mathcal{L}_{(0,1]}\left(\frac{1}{r}\right)\right\}} \quad (18)\end{aligned}$$

where the first inequality is due to (17). Evaluating (18) with $r=\big(1/\sqrt{t}+\sqrt{\mathcal{L}_{[1,\infty)}(t)}\big)^2$ leads to Theorem 1.

## V. Discussion

The paper studies the fundamental limits of stable recovery with respect to a finite union of subspaces. Our main results are upper bounds on the reconstruction error of an arbitary signal in terms of the approximation error and measurement error. These bounds show that the error is bounded in probability if $m > d$ and bounded in expectation if $m > d + 2$ where $d$ is the maximum dimension in the subspace.

These bounds provide a framework for analyzing the precise limits of stable recovery from the compressed sensing measurements. We show how this framework can be used to characterize certain phase transitions as well as provide new insights, such as the fact that the reconstruction error can be bounded in terms of the distortion of the optimal scalar quantizer using only $m-3$ values.

An important question to address in future work is the the extent to which the fundamental limits derived in this paper can attained, either exactly or approximately, using computationally efficient methods.

## References


[1] E. J. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, pp. 1207–1223, February 2006.

[2] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, January 2006.

[3] G. Reeves and D. L. Donoho, "The minimax noise sensitivity in compressed sensing," in *Proceedings of the IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013.

[4] D. L. Donoho, I. M. Johnstone, A. Maleki, and A. Montanari, "Compressed sensing over $\ell_p$ -balls: Minimax mean square error," in *Proceedings of the IEEE International Symposium on Information Theory*, St. Petersburg, Russia, August 2011.

[5] S. Jalali, A. Maleki, and R. G. Baraniuk, "Minimum complexity pursuit for universal compressed sensing," 2012, see http://arxiv.org/abs/1208.5814.

[6] U. M. Lu and M. N. Do, "A theory for sampling signals from a union of subspaces," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2334–2345, June 2008.

[7] M. F. Duarte, C. Hegde, V. Cevher, and R. G. Baraniuk, "Recovery of compressible signals in unions of subspaces," in *Proceedings of the Conference on Inofrmation Sciences and Systems (CISS)*, Baltimore, MD, March 2009.

[8] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5302–5316, November 2009.

[9] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, April 2010.

[10] T. Blumensath, "Sampling and reconstructing signals from a union of linear subspaces," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4660 – 4671, June 2011.

[11] D. L. Donoho and J. Tanner, "Precise undersampling theorems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.

[12] Y. Wu and S. Verdu, "Optimal phase transitions in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6241 – 6263, October 2012.

[13] V. Cevher, P. Indyk, C. Hegde, and R. G. Baraniuk, "Recovery of clustered sparse signals from compressive measurements," in *Proceedings of the 8-th International conference on Sampling Theory and Applications*, Marseille, France, May 2009.

[14] D. L. Donoho, I. M. Johnstone, and A. Montanari, "Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising," 2011, http://arxiv.org/abs/1111.1041.

[15] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of map estimation via the replica method and applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1902 – 1923, March 2012.

[16] D. L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2012)*, Cambridge, MA, July 2012.

[17] F. Krzakala, M. Mezard, F. Sausset, Y. F. Sun, and L. Zdeborová, "Statistical-physics-based reconstruction in compressed sensing," *Physical Review X*, vol. 2, no. 2, May 2012.

[18] D. L. Donoho and G. Reeves, "The sensitivity of compressed sensing performance to relaxation of sparsity," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2012)*, Cambridge, MA, July 2012.

[19] G. Reeves, "Beyond sparsity: Universally stable compressed sensing when the number of 'free' values is less than the number of observations," in *CAMSAP*, Saint Martin, December 2013.

[20] M. J. Wainwright, "Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting," *IEEE Transactions on Information Theory*, vol. 55, no. 12, pp. 5728–5741, December 2009.

[21] C. Cartis and A. Thompson, "A new and improved quantitative recovery analysis for iterative hard thresholding algorithms in compressed sensing," 2013, see http://arxiv.org/abs/1309.5406.