Beyond Sparsity: Universally Stable Compressed Sensing when the number of 'free' values is less than the number of observations

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

***Abstract*—Recent results in compressed sensing have shown that a wide variety of structured signals can be recovered from undersampled and noisy linear observations. In this paper, we show that many of these signal structures can be modeled using an union of affine subspaces, and that the fundamental number of observations needed for stable recovery is given by the number of "free" values, i.e. the dimension of the largest subspace in the union. One surprising consequence of our results is that the fundamental phase transition for random discrete–continuous signal models can be attained by a universal estimator that does not depend on the distribution.**

I. Introduction

Suppose that we want to recover a fixed, but unknown, signal $\mathbf{x} \in \mathbb{R}^n$ from a set of underdetermined and noisy linear equations of the form

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

where $A \in \mathbb{R}^{m \times n}$ is a known measurement matrix with $m < n$ and $\mathbf{w} \sim \mathcal{N}(0, \sigma^2 I)$ is Gaussian white noise. Given the observations $(\mathbf{y}, A)$ we seek a reconstruction $\hat{\mathbf{x}}$ that is close to $\mathbf{x}$ in mean-squared error.

A great deal of work in the compressed sensing literature has addressed this problem by selecting an estimate $\hat{\mathbf{x}}$ that is both consistent with the observations and also close to an assumed signal model. Examples of signal models include sparse signals [1]–[3], compressible signals [4]–[6], unions of subspaces [7]–[11], random discrete-continuous mixtures [12], [13], and block-sparse models [14], [15]. Reconstruction is said to be *stable*, with respect to a given signal model, if there exist universal constants C_1 and C_2 such that the event

$$\|\hat{\mathbf{x}} - \mathbf{x}\|^2 \le C_1 \|\mathbf{w}\|^2 + C_2 \|\mathbf{x}^* - \mathbf{x}\|^2 \tag{2}$$

occurs with high probability where $\mathbf{x}^*$ denotes the best approximation of $\mathbf{x}$ in the signal model (e.g. under the assumption of sparsity, $\mathbf{x}^*$ is the best k-term approximation of $\mathbf{x}$).

The tradeoff between the number of measurements and the reconstruction error has been studied extensively. One approach (e.g. [1], [7]–[11]) has been to derive sufficient conditions in terms of certain properties of the measurement matrix and the signal class, 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]

A different approach (e.g. [13], [16]–[20]) 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.

This paper makes the following contributions:

- **Union of affine subspaces model:** We show that many of the signal models studied in compressed sensing can be modeled using a finite union of affine subspaces (UAS):
$$\mathcal{U} = \bigcup_{i=1}^{N} \mathcal{V}_i, \quad \mathcal{V}_i = \{\tilde{\mathbf{v}}_i + \mathrm{span}(V_i)\},\ V_i \in \mathbb{R}^{n \times d_i}. \tag{3}$$
While non-affine subspace models have been studied previously [7]–[11], the addition of an affine component allows us to include important structural information about the unknown signal. For example, in Section III we draw a connection between the UAS model and random signals whose entries are drawn i.i.d. according to a discrete-continuous mixture. Using this connection, we show that universally stable recovery is possible even if the underlying distribution is unknown.
- **Fundamental limits of stability:** We derive an explicit and non-asymptotic result (Theorem 1) which shows that the minimum number of measurements m needed for stable recovery in the UAS model is given by the dimension d of the largest subspace, i.e.
$$m > d = \max_i \mathrm{rank}(V_i) \implies \text{stable recovery.}$$
With respect to previous work on subspace models, this result closes several existing gaps between the sufficient conditions for exact recovery in the absence of noise and the sufficient conditions for stable recovery in noise.

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

II. Main Result

We focus on recovery with respect to the finite UAS model given in (3). The dimension of the largest subspace

$$d = \max_{i \in \{1,2,\cdots,N\}} \operatorname{rank}(V_i) \tag{4}$$

is referred to as the number of free values.

For any vector $\mathbf{x} \in \mathbb{R}^n$ and subspace $\mathcal{U} \subset \mathbb{R}^n$, we define the dimension-normalized distance

$$\rho_{\mathcal{U}}(\mathbf{x}) = \min_{\mathbf{u} \in \mathcal{U}} \tfrac{1}{n} \|\mathbf{x} - \mathbf{u}\|.$$

Given observations $(\mathbf{y}, A)$, we consider the least squares estimator restricted to the set $\mathcal{U}$:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{u} \in \mathcal{U}} \|\mathbf{y} - A\mathbf{u}\|^2. \tag{5}$$

It is important to note that, depending the choice of $\mathcal{U}$, this optimization problem may be computationally infeasible in practice. The reason we study its performance is to illustrate the fundamental limits of recovery.

Finally, we study the behavior when the measurement matrix A is a random $m \times n$ matrix, generated independently of $\mathbf{x}$ and $\mathbf{w}$, with i.i.d. Gaussian entries.

We now state our main result which is a non-asymptotic upper bound on the probability that the reconstruction error is large. In this result, the parameter t provides a explicit tradeoff between the size of the constants C_1 and C_2 given in (2) and the probability that the inequality holds.

Theorem 1 (Stability). *Fix any vector $\mathbf{x} \in \mathbb{R}^n$ and finite union of affine subspaces $\mathcal{U} \subset \mathbb{R}^n$. Consider the observation model (1) where A has entries i.i.d. $\mathcal{N}(0, 1/m)$, independent of $\mathbf{w}$. If the number of measurements m exceeds the number of free values d, then recovery using the constrained least-squares estimator given in (5) is stable in the sense that, for all $t > 1$,*

$$\mathbb{P}\left[\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|^2}{n} \geq \frac{m}{n}\left[\left(\sigma^2 + \rho^2_{\mathcal{U}}(\mathbf{x})\right) C_{m,d,N} \cdot t - \sigma^2\right]\right] \leq 2e^{-\frac{m-d}{2}\Delta(t)} \tag{6}$$

where probability is taken with respect to A and $\mathbf{w}$, and

$$C_{m,d,N} = \frac{(m+1)}{(m-d+1)}\left[\mathcal{L}^{-1}\left(\frac{\log(N)}{m-d}\right)\right]^2$$

$$\Delta(t) = 2\mathcal{L}\left(\frac{t}{1+\sqrt{t\mathcal{L}(t)}}\right)$$

with

$$\mathcal{L}(x) = \begin{cases} \log(x) + 1/x - 1, & x \geq 1 \\ 0, & x < 1 \end{cases}.$$

Proof Sketch: A proof is given in [3, Section IV] for the special case where $\mathcal{U}$ models the set of k-sparse signals and it is assumed that $\mathbf{x} \in \mathcal{U}$ (and hence $\rho_{\mathcal{U}}(\mathbf{x}) = 0$). The extension to general subspaces relies heavily on the rotational invariance of the Gaussian distribution of A and $\mathbf{w}$, and the properties of affine projections. ■

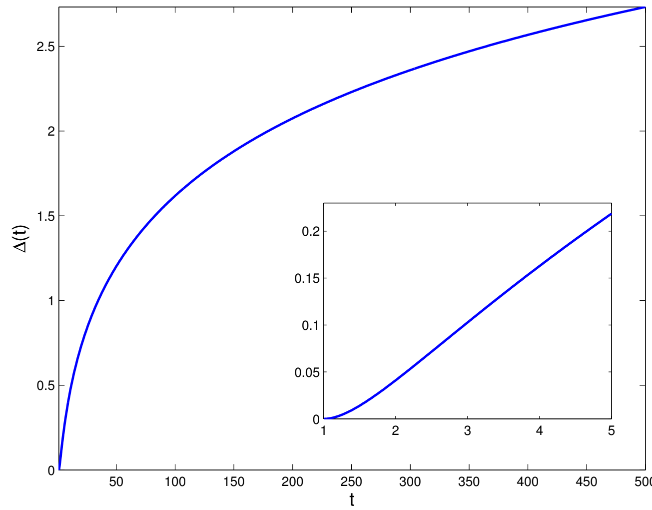

Fig. 1. Illustration of $\Delta(t)$ as a function of t. Note that $\Delta(t)$ is positive and strictly increasing for all $t > 1$.

We remark that the condition for stability in Theorem 1 depends only on the number of free values d, but not the number of subspaces, N, nor the relative positions of the subspaces — it is possible that some of the subspaces are very close to each other while others are far apart.

Moreover, Theorem 1 does not require us to verify whether certain properties of A (e.g. mutual incoherence or restricted isometry) hold. In fact, it shows that recovery is possible in settings where it is known that these properties cannot hold (e.g. when $m \approx d$). As a consequence, Theorem 1 improves over previous results which guarantee stability only when the ratio m/d exceeds a critical cutoff that is greater than one.

Finally, we remark that Theorem 1 makes no assumptions a priori about the relationship between $\mathbf{x}$ and $\mathcal{U}$. If it happens that $\mathbf{x}$ is an element of $\mathcal{U}$ then $\rho^*_{\mathcal{U}}(\mathbf{x}) = 0$ and the reconstruction error is proportional to the noise power σ^2. The worst case ratio between the expected mean-squared error and the noise power is referred to as the *noise sensitivity* [3], [21] and is given by

$$\sup_{\mathbf{x} \in \mathcal{U}} \sup_{\sigma^2 \geq 0} \frac{\mathbb{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right]}{n\,\sigma^2} \tag{7}$$

where the expectation is taken with respect to A and $\mathbf{w}$.

The following result shows that boundedness of the noise sensitivity corresponds directly to the number of free values.

Theorem 2 (Noise Sensitivity). *Consider the setting of Theorem 1. The noise sensitivity is finite if $m > d+2$ and is infinite if $m < d+2$.*

Proof Sketch: To prove finiteness, we start with the fact

$$\mathbb{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right] = \int_0^\infty \mathbb{P}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|^2 > u\right] du,$$

and then apply Theorem 1. Since $\Delta(t)$ obeys $\Delta(t)/\log(t) \to 1$ as $t \to \infty$, the integral if finite for $m > d+2$. The converse follows similarly to the proof of [3, Theorem 4]. ■

III. Examples of the UAS Model

This section shows how the UAS model and Theorem 1 can be used to derive the fundamental limits for several different signal models considered in compressed sensing. The number of affine subspaces, N, and the number of free values, d, for these models is summarized in Table I below.

A. Strict Sparsity

Consider the set of all vectors with at most k nonzero entries. It is well known that this set can be expressed as a union of $N = \binom{n}{k}$ linear subspaces, i.e.

$$\{x \in \mathbb{R}^n \,:\, \|\mathbf{x}\|_0 \leq k\} = \bigcup_{i=1}^{N} \operatorname{span}(V_i)$$

where each $V_i \in \mathbb{R}^{n\times k}$ is a concatenation of k distinct columns from the $n \times n$ identity matrix. For this special case, the asymptotic behavior of Theorem 1 is studied in [3]. Interestingly, it is shown by comparison with lower bounds on the mean-squared error, that the behavior of the constant $C_{m,d,N}$ is relatively tight.

B. Small number of distinct values

A different notion of structure is to consider the set of all vectors with at most k distinct values, i.e.

$$\{x \in \mathbb{R}^n \,:\, \operatorname{val}(\mathbf{x}) \leq k\}$$

where $\operatorname{val}(\mathbf{x}) = |\{x \,:\, x = x_i \text{ for some } i \in [n]\}|$. This set models all signals which can be represented exactly by at most k distinct quantization points. No assumptions are made about the nature of these values.

The class of k-valued vectors has been studied in various forms for representing signals with *clustered* entries (see e.g. [14], [22]) It is straightforward to show that this class can be expressed as a union of $N = k^n/k!$ linear subspaces. Here, the numerator comes from all possible allocations of n entries to k values and the denominator comes from the fact that the ordering of the values does not matter.

C. Discrete–Continuous Mixtures

A more general signal model is to assume that all but k entries belong to some finite set $B = \{b_1, b_2, \cdots, b_\ell\}$, i.e.

$$\{\mathbf{x} \in \mathbb{R}^n \,:\, |\{i \,:\, x_i \notin B\}| \leq k\}.$$

The special case $B = \{0\}$ corresponds to the set of k-sparse vectors. For the special case $B = \{0, 1\}$, this set, combined with the constraint $0 \leq x_i \leq 1$ for all i, corresponds to the class of "simple" signals studied by Donoho and Tanner [12].

This class can also be used to model a typical realization of a random vector $\mathbf{x}$ whose entries are i.i.d. P_X where

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

for some mixing weight $\varepsilon \in (0,1)$, absolutely continuous distribution P_c, and discrete distribution P_d with finite support B. By the law of large numbers,

$$\frac{1}{n}|\{i \,:\, x_i \notin B\}| \to \varepsilon, \quad \text{as} \quad n \to \infty,$$

and thus we can choose $k \approx \varepsilon n$ so that $\mathbf{x}$ belongs to the UAS model with high probability.

Let us first consider the case where the set B is known. The discrete-continuous mixture class can be modeled using a bipartite graph as shown in Figure 2. Each entry in $\mathbf{x}$ is connected to either a "free" variable (whose value is unconstrained) or a "fixed" variable, belonging to the set B. Each edge structure E defined on the this graph can be mapped to an affine subspace $\mathcal{V} = \{\tilde{\mathbf{v}} + \operatorname{span}(V)\}$ where

$$\tilde{v}_\alpha = \begin{cases} b_\beta, & \text{if there is an edge from } x_\alpha \text{ to } b_\beta \\ 0, & \text{otherwise} \end{cases}$$

and V a diagonal matrix with

$$V_{\alpha,\alpha} = \begin{cases} 1, & \text{if there is an edge from } x_\alpha \text{ to a `free' variable} \\ 0, & \text{otherwise} \end{cases}$$

The total number of affine subspaces is given by the total number of edge structures defined on the graph (up to permutations of the free variables), and thus

$$N = \ell^{n-k} \times \binom{n}{k}.$$

where the first term comes from number of sequences drawn from B and the second term corresponds to the allocation of the free variables.

Alternatively, we may consider the case where we know the size ℓ of the set B, but the values $\{b_i\}$ are *unknown*. This means that for a typical realization of $\mathbf{x}$, the entries can be categorized into one of two groups: one group has about εn distinct entries and the other group has about $(1-\varepsilon)n$ entries with no more than ℓ values. Thus, we consider the set

$$\{\mathbf{x} \in \mathbb{R}^n \,:\, |\{i \,:\, x_i \notin \tilde{B}\}| \leq k, \text{for some } \tilde{B} \text{ with } |\tilde{B}| \leq \ell\}$$

In this case there are a total of $d = k + \ell$ free values. Each edge structure E can be mapped to a linear subspace that corresponds to an allocation of these values to the entries in $\mathbf{x}$. There are

$$N = \frac{\ell^{n-k}}{\ell!} \times \binom{n}{k}$$

subspaces where the division by $\ell!$ is due to the fact that the order of the variables assigned to the set $\tilde{B}$ does not matter.

This example shows that when the support is unknown, the number of measurements needed for stability increases from k to $k + \ell$. In the asymptotic setting, when ℓ is fixed and $k/n \to \varepsilon$, this means that the undersampling ratio m/n is dictated by the weight of the "continuous" part of the signal, regardless of whether or not the support of the discrete part is known.

D. Finite Set

Finally, it is clear that any finite set of signals

$$\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$$

is a union of N affine subspaces, each with dimension 0. Thus, our results cover cases where $\mathbf{x}$ can be quantized to a known set of points.

TABLE I
EXAMPLES OF SIGNAL CLASSES COVERED BY THE UNION OF AFFINE SUBSPACES MODEL. THE NUMBER OF FREE VARIABLES IS THE DIMENSION OF THE LARGEST SUBSPACE IN THE UNION.

signal class	# of subspaces N	# of free variables d
k-sparse	$\binom{n}{k}$	k
k-valued	$k^n/k!$	k
discrete–continuous (known support)	$\ell^{n-k} \times \binom{n}{k}$	k
discrete–continuous (unknown support)	$\frac{\ell^{n-k}}{\ell!} \times \binom{n}{k}$	$k + \ell$
finite set	N	0

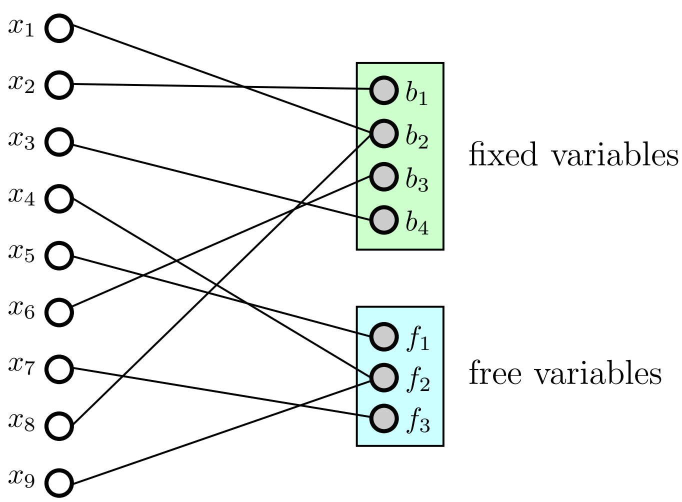

Fig. 2. Illustration of bipartite graph G. An edge structure E is an assignment of edges from variables in $\mathbf{x}$ to either the fixed or free variables. The set of k-sparse signals corresponds to a graph with a single fixed variable at 0, k free variables, and a least $n-k$ entries in $\mathbf{x}$ connected to 0. The set of k-valued signals corresponds to a graph with no fixed variables and k free variables.

IV. DISCUSSION

This paper studies the fundamental limits of stable recovery with respect to the UAS model. We show that the critical number of measurements depends only on the number of free values which is given by the dimension of the largest subspace.

We also demonstrate how the UAS model applies to a number of interesting signal classes considered in compressed sensing. Unlike models based on compressible signals, which can be well approximated by a discrete set of quantization points, the UAS model places no restrictions on the number of signals in the model nor their magnitudes. Also, unlike the non-affine subspace models, the UAS model allows us to incorporate additional structural information, such as the knowledge that some fraction of the entries in $\mathbf{x}$ belong to a known finite set while the remaining entries are unconstrained.

A key question for future work is the extent to which the fundamental stability results studied herein can be realized using computationally efficient recovery algorithms.

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 ℓ_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] S. Jalali and A. Maleki, “Minimum complexity pursuit: Stability analysis,” in *Proceedings of the IEEE International Symposium on Information Theory*, Cambridge, MA, July 2012.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] D. L. Donoho and J. Tanner, “Precise undersampling theorems,” *Proceedings of the IEEE*, vol. 98, no. 6, pp. 913–924, 2010.
[13] 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.
[14] 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.
[15] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, “Bayesian compressive sensing for cluster structured sparse signals,” *Signal Processing*, vol. 92, no. 1, pp. 256–269, January 2012.
[16] 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.
[17] 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.
[18] 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.
[19] 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.
[20] 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.
[21] D. Donoho, A. Maleki, and A. Montanari, “The noise-sensitivity phase transition in compressed sensing,” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6920–6941, October 2011.
[22] A. Jalali and M. Fazel, “Convex methods for learning *d*-valued vectors,” Presented at the Duke Workshop on Sensing and Analysis of High-Dimensional Data, July 2013.