Recovering the Sparsest Element in Subspace

Laurent Demanet and Paul Hand

Massachusetts Institute of Technology, Department of Mathematics,
77 Massachusetts Avenue, Cambridge, MA 02139

September 2013

Abstract

We address the problem of recovering a single sparse n-vector from a basis of a subspace spanned by the vector and k random vectors. We show that the sparse vector will be the output to at least one of n linear programs with high probability, provided that its support size s satisfies $s \lesssim n/\sqrt{k \log n}$. Except for the log factor, this scaling can not be improved under our subspace model. The scaling law still holds when the desired vector is approximately sparse. We also present results of computer simulations that empirically reveal when recovery is successful.

Acknowledgements. The authors acknowledge generous funding from the NSF, the Alfred P. Sloan Foundation, TOTAL S.A., and the AFOSR. The authors would also like to thank Vladislav Voroninski for helpful discussions.

1 Introduction

In this paper, we are interested in finding the sparsest nonzero element in a given subspace of $\mathbb{R}^n$. Such a task is interesting because it can be used to construct a basis of sparse vectors. Finding such a basis is an important part of many problems in dictionary learning [5, 1], blind source separation [3], and optimization theory [4]. Finding a sparse nonzero vector in a nullspace also has applications in spectral estimation, such as with Prony's method. **[citation?]**.

Minimizing $\|\cdot\|_0$ over nonzero elements in the subspace is NP-hard [4]. It is natural to attempt to minimize $\|\cdot\|_1$ instead. Spielman et al [1] provide an algorithm for finding sparse vectors in a given subspace. They provide scaling laws for high probability recovery in the case that there is a basis consisting entirely of sparse vectors.

In the present paper, we will study the procedure of Spielman et al [1] in a different context. Instead of considering a subspace with a sparse basis, we will prove scaling laws for a subspace with one sparse element together with random vectors. Additionally, we will show that the method is robust when the special vector is close to being sparse.

Note that the task of finding the sparsest element in a subspace is different from the standard problem of compressed sensing because the sparsest element in any subspace is zero. To get around this trivial solution, we add an inhomogeneous linear constraint by setting one of the coefficients to one. The resulting measurement matrix would have rows that are i.i.d. random vectors orthogonal to a fixed sparse vector. It is not clear that such a matrix behaves like a restricted isometry, so we find an alternative proof approach.

1.1 Exact Recovery

Consider the task of finding a sparse nonzero vector v in $\mathbb{R}^n$ given a subspace spanned by it and k random vectors. Precisely, let $W = \text{span}\{v, \tilde{v}_{(1)}, \dots, \tilde{v}_{(k)}\}$ where each $\tilde{v}_{(j)}$ is i.i.d. $\mathcal{N}(0, I_n)$. Such a subspace could be specified by a basis of columns of the matrix $[v, \tilde{v}_{(1)}, \dots, \tilde{v}_{(k)}]B$ for some unknown invertible matrix B. We can not simply read off v because a generic basis will not contain it in isolation.

As in [1], we attempt to recover v by solving the collection of n linear programs

$$\min \|z\|_1 \text{ such that } z \in W, z(i) = 1 \tag{1}$$

for each $1 \leq i \leq n$. From the n outputs, we select the sparsest as our estimate of v. Note that we can only hope to recover v up to a multiplicative factor that may be negative. Given an arbitrary basis of W, the subspace constraint can be implemented by computing a basis of the orthogonal complement $W^\perp$ and enforcing orthogonality of z against those vectors.

If v is sparse enough and if the subspace W is of low enough dimension, we anticipate that a multiple of v will be the solution to (1) for at least one value of i. We expect that recovery will be most likely when i is the index of one of the largest components of v. In practice, no information is known about which indices are large. Thus, we will only find v with high probability if we solve all n programs given by (1). Precisely, we have the following theorem.

Theorem 1. *Fix a nonzero $v \in \mathbb{R}^n$, let $i^* \in \text{argmax}_i |v(i)|$, and let $k \leq n/32$. There exists a universal constant c such that for sufficiently large n,*

$$\|v\|_0 \leq c\frac{n/\sqrt{\log n}}{\sqrt{k}} \Rightarrow \frac{v}{v(i^*)} \text{ is the unique solution to (1) for } i = i^*, \tag{2}$$

with probability at least $1 - 2e^{-n/64} - \gamma_1 e^{-\gamma_2 n/2} - ke^{-\lfloor c\sqrt{n/\log n}\rfloor} - \frac{k}{n^2}$. Here, γ_1 and γ_2 are universal constants.

From the scaling law, we observe the following scaling limits on the permissible sparsity in terms of the dimensionality of the search space:

$$k \sim 1 \Longrightarrow \|v\|_0 \lesssim n/\sqrt{\log n} \tag{3}$$

$$k \sim n \Longrightarrow \|v\|_0 \lesssim \sqrt{n}/\sqrt{\log n} \tag{4}$$

That is, a search space of constant size permits the discovery of a vector whose support size is almost a constant fraction of n. Similarly, a search space of fixed and sufficiently small fraction of the ambient dimension allows recovery of a vector whose support size is almost on the the order of the square root of that dimension.

Except for the logarithmic factor, the scaling law between n, k, and $\|v\|_0$ in Theorem 1 can not be improved. To see this, note that if $i^* \in \text{argmax}_i |v(i)|$, then $v/v(i^*)$ is feasible for (1) with $i = i^*$. A necessary condition for successful recovery is that the sparse vector gives a lower value of $\|\cdot\|_1$ than the minimum value attainable by the span of the random vectors:

$$\frac{\|v\|_1}{\|v\|_\infty} \leq \min \|z\|_1 \text{ such that } z \in \text{span}\{\tilde{v}_{(1)}, \dots, \tilde{v}_{(k)}\}, z(i^*) = 1, \tag{5}$$

We will show in Section 2.1 that the right hand side of (5) is on the order of $n/\sqrt{k}$ when k is at most some constant fraction of n. As $\|v\|_1/\|v\|_\infty \leq \|v\|_0$ for all v, and the equality is attained for some v, we conclude that high probability recovery of arbitrary v is possible only if $\|v\|_0 \lesssim n/\sqrt{k}$.

1.2 Stable Recovery

We now consider the corresponding task of recovering an approximately sparse v. As before, we solve the n linear programs (1) for each $1 \leq i \leq n$. Of the n results, we will select the one that is the closest to being sparse. Such a selection method could involve choosing the output with the smallest value of $\|\cdot\|_0$ after thresholding small entries. Alternatively, it could select the output with the smallest values of $\|\cdot\|_1/\|\cdot\|_2$ or $\|\cdot\|_1/\|\cdot\|_\infty$. We expect the solution to (1) to be a sparse approximation of v when i is the index of one of the largest components of v. Let v_s be the best s-sparse approximation of v. This recovery method is robust in the sense of the following theorem.

Theorem 2. *Fix a nonzero $v \in \mathbb{R}^n$, let $i^* \in \operatorname{argmax}_i |v(i)|$, and let $k \leq n/32$. There exists universal constants c, C such that for sufficiently large n, for $s = \lfloor c \frac{n/\sqrt{\log n}}{\sqrt{k}} \rfloor$, and for $i = i^*$, any minimizer $z^\#$ of (1) satisfies*

$$\left\| z^\# - \frac{v}{v(i^*)} \right\|_2 \leq C \frac{\sqrt{k \log n}}{\sqrt{n}} \frac{\|v - v_s\|_1}{\|v\|_\infty} \tag{6}$$

with probability at least $1 - 2e^{-n/64} - 2e^{-n/32} - \gamma_1 e^{-\gamma_2 n/2} - k e^{-\lfloor c\sqrt{n/\log n} \rfloor} - k/n^2$.

This theorem has a favorable constant in the error bound provided that $k \lesssim n/\log n$. In the case that $k \sim n$, the error bound has a mildly unfavorable constant, growing like $\sqrt{\log n}$. The $\sqrt{n/k}$ behavior of the error constant plays the roll of the $1/\sqrt{s}$ term that arises in the noisy compressed sensing problem [16]. The estimate (6) is slightly worse, as $\sqrt{k/n} \sim k^{1/4}/\sqrt{s}$, ignoring logarithmic factors. We believe that this behavior of the error bound could be improved.

1.3 Organization of the paper

In Section 2, we prove both theorems. In section 2.1, we derive the scaling law and proves its optimality in the context of Theorem 1. In Section 3, we present numerical simulations.

2 Proofs

To prove the theorems, we note that (1), (2), (6), and the value of $i^* \in \operatorname{argmax}_i |v(i)|$ are all invariant to any rescaling of v. Without loss of generality, it suffices to take $\|v\|_\infty = 1$ and $v(i^*) = 1$.

We begin with some notation. Let $V = [v, \tilde{V}]$, where $\tilde{V} = [\tilde{v}_{(1)}, \tilde{v}_{(2)}, \ldots, \tilde{v}_{(k)}]$. Let $V_{i^*,:}$ be the i^*-th row of V. Write $V_{i^*,:} = [1, \tilde{a}^t]$, where $\tilde{a} \in \mathbb{R}^k$. For a set S, write $\tilde{V}_S$ and $\tilde{V}_{S^c}$ as the restrictions of $\tilde{V}$ to the rows given by S and S^c, respectively. Let 1_S be the vector that is 1 on S and 0 on S^c.

Our aim is to prove that v is or is near the solution to (1) when $i = i^*$. We begin by noting that $W = \operatorname{range}(V)$. Hence, changing variables by $z = Vx$, (1) is equivalent to

$$\min \|Vx\|_1 \text{ such that } V_{i^*,:}x = 1 \tag{7}$$

We will show that $x = e_1$ is the solution to (7) in the exact case and is near the solution in the noisy case. Write $x = [x(1), \tilde{x}]$ in order to separately study the behavior of x on and away from the first coefficient. Our overall proof approach is to show that if n is larger than the given scaling, a nonzero $\tilde{x}$ gives rise to a large contribution to the ℓ^1 norm of Vx from coefficients off the support of v.

2.1 Derivation of Scaling Law

In this section, we derive the scaling law in the exact case of Theorem 1. We also prove its optimality up to the log factor. Recalling that $|v(i^*)| = \|v\|_\infty$, a necessary condition for recovery is that the normalized v is smaller in ℓ^1 than any linear combination of the random vectors:

$$\frac{\|v\|_1}{\|v\|_\infty} \leq \min \|\tilde{V}\tilde{x}\|_1 \text{ such that } \tilde{V}_{i^*,:}\tilde{x} = 1 \tag{8}$$

Because range($\tilde{V}$) is a k-dimensional random subspace, we can appeal to the uniform equivalence of the ℓ^1 and ℓ^2 norms, as given by the following lemma.

Lemma 3. *Fix $\eta < 1$. For every y in a randomly chosen (with respect to the natural Grassmannian measure) ηn-dimensional subspace of $\mathbb{R}^n$,*

$$c_\eta \sqrt{n}\|y\|_2 \leq \|y\|_1 \leq \sqrt{n}\|y\|_2$$

with probability $1 - \gamma_1 e^{-\gamma_2 n}$ for universal constants γ_1, γ_2.

This result is well known [10, 11]. Related results with different types of random subspaces can be found at [8, 7, 9, 6]. Thus, with high probability,

$$\|\tilde{V}\tilde{x}\|_1 \approx \sqrt{n}\|\tilde{V}\tilde{x}\|_2 \text{ for all } \tilde{x} \tag{9}$$

We now appeal to nonasymptotic estimates of the singular values of $\tilde{V}$. Corollary 5.35 in [12] gives that for a matrix $A \in \mathbb{R}^{n\times k}$ with $k \leq n/16$ and i.i.d. $\mathcal{N}(0, 1)$ entries,

$$\mathbb{P}\Big(\frac{\sqrt{n}}{2} \leq \sigma_{\min}(A) \leq \sigma_{\max}(A) \leq \frac{3\sqrt{n}}{2}\Big) \geq 1 - 2e^{-n/32}. \tag{10}$$

Thus, with high probability,

$$\|\tilde{V}\tilde{x}\|_2 \approx \sqrt{n}\|\tilde{x}\|_2 \tag{11}$$

Combining (9) and (11), we get $\|\tilde{V}\tilde{x}\|_1 \approx n\|\tilde{x}\|_2$ with high probability. Hence, the minimum values of the following two programs are within fixed constant multiples of each other:

$$\min \|\tilde{V}\tilde{x}\|_1 \text{ such that } \tilde{V}_{i^*,:}\tilde{x} = 1 \quad \approx \quad \min n\|\tilde{x}\|_2 \text{ such that } \tilde{V}_{i^*,:}\tilde{x} = 1 \tag{12}$$

By the Cauchy-Schwarz inequality and concentration estimates of the length of a Gaussian vector, any feasible point in the programs (12) satisfies

$$\|\tilde{x}\|_2 \geq \frac{1}{\|\tilde{V}_{i^*,:}\|_2} \approx \frac{1}{\sqrt{k}}, \tag{13}$$

with failure probability that decays exponentially in k. We have thus shown that

$$\frac{n}{\sqrt{k}} \approx \min \|\tilde{V}\tilde{x}\|_1 \text{ such that } \tilde{V}_{i^*,:}\tilde{x} = 1 \tag{14}$$

with high probability. Hence, linear combinations of columns of $\tilde{V}$ can reach ℓ^1 values as low as $n/\sqrt{k}$ under the provided normalization. For (1) to succeed at finding v with high probability, it is necessary that $\|v\|_1/\|v\|_\infty$ stay below the $n/\sqrt{k}$ level. To get successful recovery of any s-sparse v with high probability, this necessary condition becomes $s \lesssim n/\sqrt{k}$.

2.2 Proof of Theorem 1

The proof of Theorem 1 hinges on the following lemma. Let S be a superset of the support of v. Relative to the candidate $x = e_1$, any nonzero $\tilde{x}$ gives components on S^c that can only increase $\|Vx\|_1$. Nonzero $\tilde{x}$ can give components on S that decrease $\|Vx\|_1$. If the ℓ^1 norm of $\tilde{V}\tilde{x}$ on S^c is large enough and the ℓ^1 norm of $\tilde{V}\tilde{x}$ on S is small enough, then the minimizer must be e_1.

Lemma 4. *Let $V = [v, \tilde{V}]$ with $\|v\|_\infty = 1$, $V_{i^*,:} = [1, \tilde{a}^t]$, supp$(v) \subseteq S$, and card$(S) = s$. Suppose that $\|\tilde{V}_S\tilde{x}\|_1 \leq 2s\|\tilde{x}\|_1$ and $\|\tilde{V}_{S^c}\tilde{x}\|_1 \geq (2\|\tilde{a}\|_\infty + 2)s\|\tilde{x}\|_1$ for all $\tilde{x}$. Then, e_1 is the unique solution to* (7).

Proof. For any x, observe that

$$\|Vx\|_1 = \|vx(1) + \tilde{V}_S\tilde{x}\|_1 + \|\tilde{V}_{S^c}\tilde{x}\|_1 \tag{15}$$
$$\geq \|v\|_1|x(1)| - 2s\|\tilde{x}\|_1 + \|\tilde{V}_{S^c}\tilde{x}\|_1 \tag{16}$$
$$\geq \|v\|_1|x(1)| + 2\|\tilde{a}\|_\infty s\|\tilde{x}\|_1 \tag{17}$$

where the first inequality is from the upper bound on $\|\tilde{V}_S\tilde{x}\|_1$ and the second inequality is from the lower bound on $\|\tilde{V}_{S^c}\tilde{x}\|_1$. Note that $x = e_1$ is feasible and has value $\|Ve_1\|_1 = \|v\|_1$. Hence, at a minimizer $\tilde{x}^\#$,

$$\|v\|_1|x^\#(1)| + 2\|\tilde{a}\|_\infty s\|\tilde{x}^\#\|_1 \leq \|v\|_1. \tag{18}$$

Using the constraint $x^\#(1) + \tilde{a}^t\tilde{x}^\# = 1$, a minimizer must satisfy

$$\|v\|_1(1 - \|\tilde{a}\|_\infty\|\tilde{x}^\#\|_1) + 2\|\tilde{a}\|_\infty s\|\tilde{x}^\#\|_1 \leq \|v\|_1. \tag{19}$$

Noting that $\|v\|_1 \leq s$, a minimizer must satisfy

$$2\|\tilde{a}\|_\infty s\|\tilde{x}^\#\|_1 \leq \|\tilde{a}\|_\infty s\|\tilde{x}^\#\|_1. \tag{20}$$

Hence, $\tilde{x}^\# = 0$. The constraint provides $x^\#(1) = 1$, which proves that e_1 is the unique solution to (7). □

To prove Theorem 1 by applying Lemma 4, we need to study the minimum value of $\|\tilde{V}_{S^c}\tilde{x}\|_1/\|\tilde{x}\|_1$ for matrices $\tilde{V}_{S^c}$ with i.i.d. $\mathcal{N}(0, 1)$ entries. Precisely, we will show the following lemma.

Lemma 5. *Let A be a $n \times k$ matrix with i.i.d. $\mathcal{N}(0, 1)$ entries, with $k \leq n/16$. There is a universal constant $\tilde{c}$, such that with high probability, $\|Ax\|_1/\|x\|_1 \geq \tilde{c}n/\sqrt{k}$ for all $x \neq 0$. This probability is at least $1 - 2e^{-n/32} - \gamma_1 e^{-\gamma_2 n}$.*

Proof of Lemma 5. We are to study the problem

$$\min \|Ax\|_1 \text{ such that } \|x\|_1 = 1, \tag{21}$$

which is equivalent to

$$\min \|Ax\|_1 \text{ such that } \|x\|_1 \geq 1. \tag{22}$$

The minimum value of (22) can be bounded from below by that of

$$\min \|Ax\|_1 \text{ such that } \|x\|_2 \geq 1/\sqrt{k} \tag{23}$$

because the feasible set of (22) is included in the feasible set of (23). We now write both the objective and constraint in terms of Ax. To that end, we apply the lower bound in (10) to get

$$\mathbb{P}(\|x\|_2 \leq 2\frac{\|Ax\|_2}{\sqrt{n}} \text{ for all } x) \geq 1 - 2e^{-n/32} \tag{24}$$

The feasible set of (23) is contained by the set $\{x \mid \|Ax\|_2 \geq \frac{1}{2}\sqrt{\frac{n}{k}}\}$ with high probability. Hence, a lower bound to (23) is with high probability given by

$$\min \|Ax\|_1 \text{ such that } \|Ax\|_2 \geq \frac{1}{2}\sqrt{\frac{n}{k}} \tag{25}$$

In order to find a lower bound on (25), we apply Lemma 3 to the range of A, which is a k-dimensional random subspace of $\mathbb{R}^n$ with $k \leq n/16$. Taking $\eta = 1/16$, we see that with high probability, the minimal value of (25) is bounded from below by $\frac{c_\eta}{2}\frac{n}{\sqrt{k}}$. The minimal value of (25), and hence of (21), is bounded from below by $\tilde{c}n/\sqrt{k}$ for some universal constant $\tilde{c}$ with probability at least $1 - 2e^{-n/32} - \gamma_1 e^{-\gamma_2 n}$. □

To prove the theorem by applying Lemma 4, we also need to study the maximum value of $\|\tilde{V}_S\tilde{x}\|_1/\|\tilde{x}\|_1$ for matrices $\tilde{V}_S$ with i.i.d. $\mathcal{N}(0,1)$ entries.

Lemma 6. *Let A be a $s \times k$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Then $\sup_{x\neq 0} \|Ax\|_1/\|x\|_1 \leq 2s$ with probability at least $1 - ke^{-s}$.*

Proof. Note that elementary matrix theory gives that the $\ell^1 \to \ell^1$ operator norm of A is

$$\max_{x\neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1\leq i\leq k} \|Ae_i\|_1 \tag{26}$$

As Ae_i is an $s \times 1$ vector of i.i.d. standard normals, we have

$$\mathbb{P}(\|Ae_i\|_1 > t) \leq 2^s e^{-t^2/2s} \tag{27}$$

Hence,

$$\mathbb{P}(\max_i \|Ae_i\|_1 > t) \leq k2^s e^{-t^2/2s} \tag{28}$$

Taking $t = 2s$, we conclude

$$\mathbb{P}(\max_i \|Ae_i\|_1 > 2s) \leq k2^s e^{-2s} \leq ke^{-s} \tag{29}$$

□

We can now combine Lemmas 4, 5, and 6 to prove Theorem 1.

Proof of Theorem 1. Let $\tilde{c}$ be the universal constant given by Lemma 5 and let $c = \tilde{c}/5$. We will show that for $\|v\|_0 \leq c\frac{n/\sqrt{\log n}}{\sqrt{k}}$, the minimizer to (1) is v with at least the stated probability.

Let S be any superset of supp(v) with cardinality $s = \lfloor c\frac{n/\sqrt{\log n}}{\sqrt{k}} \rfloor$. As per Lemma 4, e_1 is the solution to (7), and hence v is the unique solution to (1), if the following events occur simultaneously:

$$\|\tilde{V}_S\tilde{x}\|_1 \leq 2s\|\tilde{x}\|_1 \text{ for all } \tilde{x} \tag{30}$$

$$\|\tilde{a}\|_\infty \leq 2\sqrt{\log n} \tag{31}$$

$$\|\tilde{V}_{S^c}\tilde{x}\|_1 \geq 5\sqrt{\log n}s\|\tilde{x}\|_1 \text{ for all } \tilde{x} \tag{32}$$

Applying Lemma 6 to the $s \times k$ matrix $\tilde{V}_S$, we get that (30) holds with probability at least $1-ke^{-s} = 1-ke^{-\lfloor c\sqrt{n/\log n}\rfloor}$. Classical results on the maximum of a gaussian vector establishes that (31) holds with probability at least $1 - k/n^2$. Because $s \leq n/2$ and $k \leq n/32$, we have that $\tilde{V}_{S^c}$ has height at least $n/2$ and width at most $n/32$. Hence, Lemma 5 gives that $\|\tilde{V}_{S^c}\tilde{x}\|_1/\|\tilde{x}\|_1 \geq \tilde{c}n/\sqrt{k}$ for all $\tilde{x} \neq 0$ with probability at least $1 - 2e^{-n/64} - \gamma_1 e^{-\gamma_2 n/2}$. Because $s \leq \frac{\tilde{c}}{5}\frac{n/\sqrt{\log n}}{\sqrt{k}}$, we conclude (32), allowing us to apply Lemma 4. Hence, successful recovery occurs with probability at least $1 - 2e^{-n/64} - \gamma_1 e^{-\gamma_2 n/2} - ke^{-\lfloor c\sqrt{n/\log n}\rfloor} - k/n^2$.

$\square$

2.3 Proof of Theorem 2

We will prove the following lemma, of which Theorem 2 is a special case.

Lemma 7. *Fix a nonzero $v \in \mathbb{R}^n$, let $i^* \in \operatorname{argmax}_i |v(i)|$, and let $k \leq n/32$. There exists universal constants c, C such that for sufficiently large n, for all $s \leq c\frac{n/\sqrt{\log n}}{\sqrt{k}}$, and for $i = i^*$, any minimizer $z^\#$ of (1) satisfies*

$$\left\| z^\# - \frac{v}{v(i^*)} \right\|_2 \leq C\frac{\sqrt{n}}{s}\frac{\|v - v_s\|_1}{\|v\|_\infty} \tag{33}$$

with probability at least $1 - 2e^{-n/64} - 2e^{-n/32} - \gamma_1 e^{-\gamma_2 n/2} - ke^{-s} - k/n^2$.

At first glance, this lemma appears to have poor error bounds for large n and poor probabilistic guarantees for small s. On further inspection, the bounds can be improved by simply considering a larger s, possibly even larger than the size of the support of v. Larger values of s simultaneously increase the denominator and decrease the s-term approximation error in the numerator of (33). Taking the largest permissible value $s = \lfloor c\frac{n/\sqrt{\log n}}{\sqrt{k}} \rfloor$, we arrive at Theorem 2.

Lemma 7 hinges on the following analog of Lemma 4.

Lemma 8. *Fix $1 \leq s < n$ and $\alpha > 0$. Let $V = [v, \tilde{V}]$ with $\|v\|_\infty = 1$, $V_{i^*,:} = [1, \tilde{a}^t]$, $\delta = \|v - v_s\|_1$, supp$(v) \subseteq S$, and card$(S) = s$. If $\|\tilde{V}_S\tilde{x}\|_1 \leq 2s\|\tilde{x}\|_1$ and $\|\tilde{V}_{S^c}\tilde{x}\|_1 \geq (2\|\tilde{a}\|_\infty + 2 + \alpha)s\|\tilde{x}\|_1$ for all $\tilde{x} \in \mathbb{R}^k$, then any $x^\#$ minimizing (7) satisfies*

$$|x_1^\# - 1| \leq \frac{2\delta}{s}, \qquad \text{and} \qquad \|\tilde{x}^\#\|_1 \leq \frac{2\delta}{s(\|\tilde{a}\|_\infty + \alpha)}. \tag{34}$$

Proof. For any x, observe that

$$\|Vx\|_1 = \|v \cdot x(1) + \tilde{V}_S \tilde{x}\|_1 + \|\tilde{V}_{S^c}\tilde{x}\|_1 \tag{35}$$
$$\geq \|v\|_1 |x(1)| - 2s\|\tilde{x}\|_1 + \|\tilde{V}_{S^c}\tilde{x}\|_1 \tag{36}$$
$$\geq \|v\|_1 |x(1)| + (2\|\tilde{a}\|_\infty + \alpha)s\|\tilde{x}\|_1 \tag{37}$$
$$\geq (\|v_s\|_1 - \delta)|x(1)| + (2\|\tilde{a}\|_\infty + \alpha)s\|\tilde{x}\|_1 \tag{38}$$

where the first inequality is from the upper bound on $\|\tilde{V}_S\tilde{x}\|_1$ and the second inequality is from the lower bound on $\|\tilde{V}_{S^c}\tilde{x}\|_1$. Note that $x = e_1$ is feasible and has value $\|Ve_1\|_1 = \|v\|_1 \leq \|v_s\|_1 + \delta$. Hence, at a minimizer $\tilde{x}^\#$,

$$(\|v_s\|_1 - \delta)|x^\#(1)| + (2\|\tilde{a}\|_\infty + \alpha)s\|\tilde{x}^\#\|_1 \leq \|v_s\|_1 + \delta. \tag{39}$$

Using the constraint $x^\#(1) + \tilde{a}\tilde{x}^\# = 1$, a minimizer must satisfy

$$(\|v_s\|_1 - \delta)(1 - \|\tilde{a}\|_\infty \|\tilde{x}^\#\|_1) + (2\|\tilde{a}\|_\infty + \alpha)s\|\tilde{x}^\#\|_1 \leq \|v_s\|_1 + \delta. \tag{40}$$

Noting that $\|v_s\|_1 \leq s$, a minimizer must satisfy

$$\|\tilde{x}^\#\|_1 \leq \frac{2\delta}{(\|\tilde{a}\|_\infty + \alpha)s}. \tag{41}$$

Applying the constraint again, we get

$$|x^\#(1) - 1| \leq \frac{2\delta}{s}. \tag{42}$$

□

We now complete the proof of Theorem 2 by proving Lemma 7.

Proof of Lemma 7. Let $\tilde{c}$ be the universal constant given by Lemma 5 and let $c = \tilde{c}/6$. We will show that for any $s \leq c\frac{n/\sqrt{\log n}}{\sqrt{k}}$, the minimizer to (1) is near v with at least the stated probability.

Let S be any superset of $\text{supp}(v_s)$ with cardinality s. Applying Lemma 8 with $\alpha = \sqrt{\log n}$, we observe that a minimizer $x^\#$ to (7) satisfies $|x^\#(1) - 1| \leq 2\delta/s$ and $\|\tilde{x}^\#\|_1 \leq 2\delta/(s\sqrt{\log n})$ if the following events occur simultaneously:

$$\|\tilde{V}_S\tilde{x}\|_1 \leq 2s\|\tilde{x}\|_1 \text{ for all } \tilde{x} \tag{43}$$
$$\|\tilde{a}\|_\infty \leq 2\sqrt{\log n} \tag{44}$$
$$\|\tilde{V}_{S^c}\tilde{x}\|_1 \geq 6\sqrt{\log n}s\|\tilde{x}\|_1 \text{ for all } \tilde{x} \tag{45}$$

Applying Lemma 6 to the $s \times k$ matrix $\tilde{V}_S$, we get that (43) holds with probability at least $1 - ke^{-s}$. Classical results on the maximum of a gaussian vector establishes that (44) holds with probability at least $1 - k/n^2$. Because $s \leq n/2$ and $k \leq n/32$, we have that $\tilde{V}_{S^c}$ has height at least $n/2$ and width at most $n/32$. Hence, Lemma 5 gives that $\|\tilde{V}_{S^c}\tilde{x}\|_1/\|\tilde{x}\|_1 \geq \tilde{c}n/\sqrt{k}$ for all $\tilde{x} \neq 0$ with probability at least $1 - 2e^{-n/64} - \gamma_1 e^{-\gamma_2 n/2}$. If $s \leq \frac{\tilde{c}}{6}\frac{n/\sqrt{\log n}}{\sqrt{k}}$, we conclude (45), allowing us to apply Lemma 4.

It remains to show that $Vx^\#$ is near v. Observe that

$$\begin{aligned}
\|Vx^\# - v\|_2 &= \|Vx^\# - Ve_1\|_2 && (46)\\
&\leq \|v\|_2|x^\#(1) - 1| + \|\tilde{V}\tilde{x}^\#\|_2 && (47)\\
&\leq \|v\|_2|x^\#(1) - 1| + \sigma_{\max}(\tilde{V})\|\tilde{x}^\#\|_2 && (48)\\
&\leq \sqrt{n}|x^\#(1) - 1| + \frac{3}{2}\sqrt{n}\|\tilde{x}^\#\|_1 && (49)\\
&\leq \sqrt{n}\frac{2\delta}{s} + \frac{3}{2}\sqrt{n}\frac{2\delta}{s\sqrt{\log n}} && (50)\\
&\leq C\frac{\sqrt{n}}{s}\delta && (51)
\end{aligned}$$

The the third inequality uses the fact that $\|v\|_\infty = 1$ and $\sigma_{\max}(\tilde{V}) \leq \frac{3}{2}\sqrt{n}$, which occurs with probability at least $1 - 2e^{-n/32}$ due to the upper bound in (10). Hence, approximate recovery occurs with probability at least $1 - 2e^{-n/64} - 2e^{-n/32} - \gamma_1 e^{-\gamma_2 n/2} - ke^{-s} - k/n^2$.

□

3 Simulations

In this section, we present computer simulations that demonstrate when solving n linear programs of the form (1) can find an approximately sparse vector $v \in \mathbb{R}^n$ from a subspace spanned by it and k random vectors.

Let $n = 100$, $S = \{1, \ldots, s\}$, and $v = 1_S + \epsilon u$, where $\epsilon = 0.01$ and u is i.i.d. Gaussian and normalized such that $\|u\|_1 = 1$. As before, let $i^* = \operatorname{argmax}_i |v(i)|$. We solve (1) for $1 \leq i \leq n$ using YALMIP [13] with the SDPT3 solver [14, 15]. Among these n outputs, we let $z^\#$ be either the one corresponding to $i = i^*$ or the one that has the smallest value of $\|\cdot\|_1/\|\cdot\|_2$. We call a recovery successful if $\left\|z^\# - v/v(i^*)\right\|_2 \leq \epsilon$. Figure 1 shows the probability of successful recovery, as computed over 10 independent trials, for many values of k and the approximate sparsity s. Near and below the visible curve, simulations were performed for all even values of k and s. In the large region to the top-right of the curve, simulations were performed only for values of k and s that are multiples of 5. In this region, the probably of recovery was always zero.

We observe that the recoverable sparsity decays rapidly in k for small values of k. For large k, outside the scope of the theorems of this paper, the maximal recoverable sparsity decays slowly.

Unsurprisingly, when we select the best of all n programs (1), we can outperform the result from a single program, even if an oracle tells us the index of the largest coefficient, i^*. This effect is more pronounced for small values of k. To see why, note that successful recovery is expected when $\|\tilde{V}_{i,:}\|_\infty$ is small. If k is small a large deviation of $\|\tilde{V}_{i,:}\|_\infty$ is fairly likely. If there are many i where v is large, it is likely that (1) will recover v for one of these i. If k is large, a large deviation of $\|\tilde{V}_{i,:}\|_\infty$ is extremely unlikely.

When selecting a signal among the solutions to (1) for each $1 \leq i \leq n$, we see that we do not recover a signal with support size much greater than 50 out of 100, even if that signal is the minimizer for $i = i^*$. In this case, the selector $\|\cdot\|_1/\|\cdot\|_2$ may preferentially select random vectors because v is neither exactly or approximately sparse.

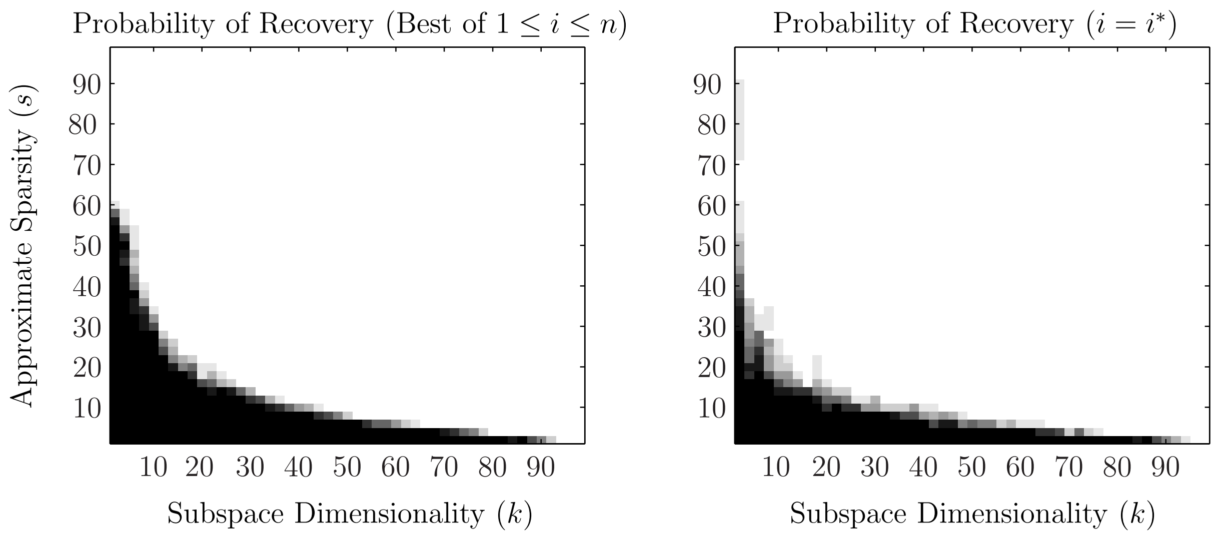

Figure 1: Empirical probability of recovery versus approximate sparsity and subspace dimensionality k. In the left panel, the n programs (1) were solved, and the output with smallest value of $\| \cdot \|_1 / \| \cdot \|_2$ was selected. In the right panel, the one corresponding to $i^* = \max_i |v(i)|$ was selected. Each set of parameters was simulated with 10 independent trials. White represents a recovery with probability zero. Black represents recovery with probability 1.

References

[1] D. Spielman, H. Wang, J. Wright. Exact Recovery of Sparsely-Used Dictionaries. *J. Machine Learning Research - Proceedings Track 23* 37.1-37.18, 2012

[2] L-A. Gottlieb, T. Neylon. Matrix sparsification and the sparse null space problem. *APPROX and RANDOM*, 6302:205-218, 2010.

[3] M. Zibulevsky and B. Pearlmutter. Blind source separation by sparse decomposition. *Neural Computation*, 13(4), 2001.

[4] T.F. Coleman, A. Pothen. The null space problem I. Complexity. *SIAM J. Algebraic and Discrete Methods*, 7(4):527-537, 1986.

[5] F. Bach, J. Mairal, J. Ponce. Convex Sparse Matrix Factorizations *arXiv preprint 0812.1869*, 2008.

[6] V. Guruswami, J. Lee, A. Razborov. Almost Euclidean subspaces of l_1^n via expander codes. *Combinatorica* 30(1): 47-68, 2010.

[7] S. Artstein-Avidan, V. Milman. Logarithmic reduction of the level of randomness in some probabilistic geometric constructions. *J. Functional Analysis* 235, 297-329, 2006.

[8] S. Lovett, S. Sodin. Almost Euclidean sections of the N-dimensional cross-polytope using O(N) random bits. *Electronic Colloquium on Computational Complexity, Report* 7-12. 2007.

[9] J. Lee. Kernels of Random Sign Matrices. Tcs math blog post. Available: `tcsmath.wordpress.com/2008/05/08/kernels-of-random-sign-matrices/`. 2008.

[10] T. Figiel, J. Lindenstrauss, V. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53-94, 1977.

[11] B. Kashin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334-351, 478, 1977.

[12] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. C. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2010.

[13] J. Löfberg. YALMIP : A Toolbox for Modeling and Optimization in MATLAB. *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[14] K.C. Toh, M.J. Todd, R.H. Tutuncu. SDPT3 — a Matlab software package for semidefinite programming, *Optimization Methods and Software*, 11, 545-581, 1999

[15] R.H Tutuncu, K.C. Toh, M.J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3, *Mathematical Programming Ser. B*, 95, 189-217, 2003.

[16] E. J. Candès, J. Romberg, T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59 1207-1223, 2005.