

COMPRESSED SENSING WITH SEQUENTIAL OBSERVATIONS

D. M. Malioutov, S. Sanghavi, A. S. Willsky

Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge MA

ABSTRACT

Compressed sensing allows perfect recovery of a signal that is known to be sparse in some basis using only a small number of measurements. The results in the literature have focused on the asymptotics of how many samples are required and the probability of making an error for a fixed batch of samples. We investigate an alternative scenario where observations are available in sequence and can be stopped as soon as there is reasonable certainty of correct reconstruction. For the random Gaussian ensemble we show that a simple stopping rule gives the absolute minimum number of observations required for exact recovery, with probability one. However, for other ensembles like Bernoulli or Fourier, this is no longer true, and the rule is modified to trade off delay in stopping and probability of error. We also describe a stopping rule for the near-sparse case which tells when enough observations are made to reach a desired tolerance in reconstruction. Sequential approach to compressed sensing involves the solution of a sequence of linear programs, and we outline how this sequence can be solved efficiently.

Index Terms— Sequential compressed sensing

1. INTRODUCTION

In compressed sensing (CS) a few random measurements of a signal are taken, and the signal is recovered using sparse representation algorithms such as the ℓ_1 -based *basis pursuit* [1]. This is most useful when the cost of taking measurements is much larger than the computational overhead of recovering the signal, hence minimizing the number of required measurements is a primary concern. Existing analytical results provide guidelines on how many measurements are needed to ensure exact recovery with high probability, but these are often seen to be pessimistic [1, 2] and rely on information about the sparsity of the unknown signal. We consider an alternative scenario where one is able to access observations in sequence and perform computations in between observations. Exact recovery may now be possible from the smallest

required number of observations without any *a-priori* knowledge about how sparse the underlying signal is. This, however, requires a computationally efficient decoder which can detect exactly when enough samples have been received.

We first consider the case when noiseless measurements are taken using the random Gaussian ensemble, and we show that performing a simple check yields such a decoder. In particular, at time M the decoder solves for

$$\hat{\mathbf{x}}_M = \arg \min \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{a}'_i \mathbf{x} = y_i, \quad i = 1, \dots, M \quad (1)$$

In case of one-step agreement, i.e. $\hat{\mathbf{x}}_M = \hat{\mathbf{x}}_{M-1}$, the decoder declares $\hat{\mathbf{x}}_M$ to be the reconstruction and stops requesting new measurements. In Section 2 we show that this decoder gives exact reconstruction with probability one.

The sequential stopping rule can also be used with other measurement ensembles, such as random Bernoulli and the ensemble of random rows from a Fourier basis. However, in this case the stopping rule no longer has probability zero of making an error. We modify the rule to take this into account: now we wait until T subsequent solutions $\hat{\mathbf{x}}^M, \dots, \hat{\mathbf{x}}^{M+T}$ all agree. In Section 3 we show that in the Bernoulli case the probability of making an error using this stopping rule decays exponentially with T allowing trade-off of error probability and delay. In Section 4 we consider the near-sparse case where the signal mostly has small but non-zero components. We describe when to stop requesting measurements to have the error in the reconstruction below a desired tolerance.

We propose an efficient way to solve the sequential problem in Section 5. Rather than re-solving the linear program from scratch after an additional measurement is received, we use an augmented linear program that uses the solution at step M to guide its search for the new solution. We show that this significantly reduces computational complexity.

2. STOPPING RULE IN THE GAUSSIAN CASE

We now analyze the sequential CS approach with the Gaussian measurement ensemble. Suppose that the underlying sparse signal $\mathbf{x}^* \in \mathbb{R}^N$, has K non-zero components (we denote the number of non-zero entries in \mathbf{x} by $\|\mathbf{x}\|_0$). We sequentially receive random measurements $y_i = \mathbf{a}'_i \mathbf{x}^*$, where $\mathbf{a}_i \sim \mathcal{N}(0, I)$ is a N -vector of i.i.d. Gaussian samples. At

This research was supported by the Air Force Office of Scientific Research under Grant FA9550-04-1-0351, the Army Research Office under MURI Grant W911NF-06-1-0076.

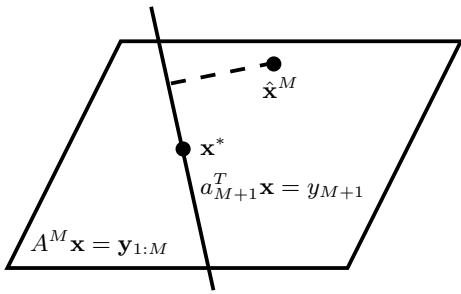


Fig. 1. A new constraint is added: $\mathbf{a}'_{M+1} \mathbf{x} = y_{M+1}$. Probability that this hyperplane passing through \mathbf{x}^* also passes through $\hat{\mathbf{x}}^M$ is zero.

step M we solve the basis-pursuit problem in (1) using all the received data. Results in compressed sensing [1, 2] indicate that after receiving around $M \propto K \log(N)$ measurements, solving (1) recovers the signal \mathbf{x}^* with high probability. This requires the knowledge of K , which may not be available, and only rough bounds on the scaling constants are known. Our approach is different - we compare the solutions at step M and $M + 1$, and if they agree, we declare correct recovery.

Proposition 1 *If $\hat{\mathbf{x}}^{M+1} = \hat{\mathbf{x}}^M$ in the Gaussian measurement ensemble, then $\hat{\mathbf{x}}^M = \mathbf{x}^*$, with probability 1.*

Proof. Let $\mathbf{y}_{1:M} \triangleq [y_1, \dots, y_M]'$, and $A^M \triangleq [\mathbf{a}'_1, \dots, \mathbf{a}'_M]'$. Suppose that $\hat{\mathbf{x}}^M \neq \mathbf{x}^*$ and $\mathbf{y}_{1:M} = A^M \hat{\mathbf{x}}^M$. We also have $\mathbf{y}_{1:M} = A^M \mathbf{x}^*$, so both \mathbf{x}^* and $\hat{\mathbf{x}}^M$ belong to the $(N - M)$ -dimensional affine space $\{\mathbf{x} \mid \mathbf{y}_{1:M} = A^M \mathbf{x}\}$. The next measurement passes a random hyperplane $y_{M+1} = \mathbf{a}'_{M+1} \mathbf{x}^*$ through \mathbf{x}^* and reduces the dimension of the affine subspace of feasible solutions by 1. The nullspace $\text{Null}(A^{M+1}) = \text{Null}(A^M) \cap \text{Null}(\mathbf{a}_{M+1})$. Since \mathbf{a}_{M+1} is random, and independent of previous samples $\mathbf{a}_1, \dots, \mathbf{a}_M$, the probability that $\mathbf{x}^* - \hat{\mathbf{x}}^M$ falls in the lower-dimensional subspace $\text{Null}(A^{M+1})$ is 0. See Figure 1 for illustration. \square

Clearly, if we obtain $\hat{\mathbf{x}}^M = \mathbf{x}^*$, then the solution will not change with additional samples: \mathbf{x}^* is always in the feasible set, and the feasible set is shrinking with each new sample. In the Gaussian case the stopping rule can be simplified further: if $\hat{\mathbf{x}}^M$ has fewer than M non-zero entries, then $\hat{\mathbf{x}}^M = \mathbf{x}^*$.

Proposition 2 *If $\|\hat{\mathbf{x}}^M\|_0 < M$, then $\hat{\mathbf{x}}^M = \mathbf{x}^*$ with prob. 1.*

Proof. Let $A = A^M$ to simplify notation. Then A is $M \times N$, with $M < N$. The key fact about random Gaussian matrices is that any $M \times M$ submatrix of A is non-singular with probability 1¹. Let \mathcal{I} be the support of \mathbf{x}^* , i.e. $\mathcal{I} = \{i \mid x_i^* \neq 0\}$. Suppose that there is another sparse feasible

¹This is easy to see: fix $T \subset \{1, \dots, N\}$ with $|T| = M$. Then probability that $A_{T_M} \in \text{span}(A_{T_1}, \dots, A_{T_{M-1}})$ is zero, as A_{T_M} is a random vector in \mathbb{R}^M and the remaining columns span a lower-dimensional subspace.

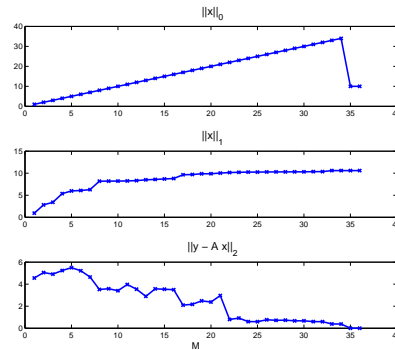


Fig. 2. An example with the Gaussian measurement ensemble. $N = 100$, and $K = 10$. Top plot: $\|\hat{\mathbf{x}}^M\|_0$. Medium plot: $\|\hat{\mathbf{x}}^M\|_1$. Bottom plot: $\|\mathbf{x}^* - \hat{\mathbf{x}}^M\|_2$.

vector $\hat{\mathbf{x}} \neq \mathbf{x}^*$ with support \mathcal{J} , such that $|\mathcal{J}| < M$. There are two possibilities: $\mathcal{I} \subset \mathcal{J}$ and $\mathcal{I} \not\subset \mathcal{I} \cap \mathcal{J}$. We show that in both cases $\hat{\mathbf{x}} \neq \mathbf{x}^*$ can occur only with probability zero.

First suppose $\mathcal{I} \subset \mathcal{J}$. Then $\hat{\mathbf{x}} - \mathbf{x}^* \in \text{Null}(A)$, and support of $\hat{\mathbf{x}} - \mathbf{x}^*$ is a subset of \mathcal{J} , hence it is smaller than M . But that means that fewer than M columns of A are linearly dependent, which only happens with probability zero.

Now if $\mathcal{I} \not\subset \mathcal{I} \cap \mathcal{J}$, then the probability that $\mathbf{y} = A\mathbf{x}^*$ falls into $\text{span}(A_{\mathcal{I} \cap \mathcal{J}})$ is zero, as it is a smaller dimensional subspace of $\text{span}(A_{\mathcal{I}})$. Thus with probability 1 there is only one solution with $\|\mathbf{x}\|_0 < M$, namely \mathbf{x}^* . \square

Consider an example in Figure 2 with $N = 100$, and $K = 10$. We keep solving (1) until agreement, $\hat{\mathbf{x}}^M = \hat{\mathbf{x}}^{M+1}$. The top plot shows that $\|\hat{\mathbf{x}}^M\|_0$ increases linearly with M until $M = 35$, at which point it drops to $K = 10$ and we have $\hat{\mathbf{x}}^M = \mathbf{x}^*$. The middle plot shows the monotonic increase in $\|\hat{\mathbf{x}}^M\|_1$ (as the feasible set is shrinking with M). The bottom plot shows the error-norm of the solution, $\|\hat{\mathbf{x}}^M - \mathbf{x}^*\|_2$. On average it tends to go down with more observations, but non-monotonically. After $M = 35$ the error becomes zero.

3. STOPPING RULE IN THE BERNOULLI CASE

Now suppose that the measurement vectors \mathbf{a}_i have equiprobable i.i.d. Bernoulli entries ± 1 . A difference emerges from the Gaussian case: the probability that all $M \times M$ submatrices of A^M are non-singular is no longer 0. This makes it possible (with non-zero probability) for $\hat{\mathbf{x}}^{M+1}$ to agree with $\hat{\mathbf{x}}^M$ when $\hat{\mathbf{x}}^M \neq \mathbf{x}^*$, and for erroneous solutions $\hat{\mathbf{x}}^M$ to have cardinality less than M . We modify the stopping rule to require agreement for several steps - success is declared only when last T solutions all agree. We show in proposition 3 that the probability of error decays exponentially with T . We use the following Lemma from [3]:

Lemma 1 *Let \mathbf{a} be an i.i.d. Bernoulli vector with $\mathbf{a} \in \{-1, 1\}^N$. Let W be a deterministic d -dimensional subspace of \mathbb{R}^N , $0 \leq$*

$d < N$. Then $P(\mathbf{a} \in W) \leq 2^{d-N}$.

We are now ready to establish the following claim:

Proposition 3 *Consider the Bernoulli measurement case. Suppose \mathbf{x}^* is in general position (e.g. has i.i.d. Gaussian entries). If $\hat{\mathbf{x}}^M = \hat{\mathbf{x}}^{M+1} = \dots = \hat{\mathbf{x}}^{M+T}$, then $\hat{\mathbf{x}}^M = \mathbf{x}^*$ with probability greater than $1 - 2^{-T}$.*

Proof. Suppose that $\hat{\mathbf{x}}^M \neq \mathbf{x}^*$. Denote the support of \mathbf{x}^* and $\hat{\mathbf{x}}^M$ by \mathcal{I} and \mathcal{J} respectively, and denote i -th column of A^M as A_i^M . We have $\mathbf{y}_{1:M} = A^M \mathbf{x}^* = A^M \hat{\mathbf{x}}^M$, i.e.

$$\sum_{i \in \mathcal{I}} A_i^M x_i^* = \sum_{i \in \mathcal{J}} A_i^M \hat{x}_i^M, \quad (2)$$

Since \mathbf{x}^* is in general position, then $\sum_{i \in \mathcal{I}} A_i^M x_i^*$ is a general point in the span of $\{A_i^M, i \in \mathcal{I}\}$, and hence, it can be shown that $\text{span}(A_i^M, i \in \mathcal{I}) \subset \text{span}(A_i^M, i \in \mathcal{J})$ with probability 1. Let $L = \|\hat{\mathbf{x}}^M\|_0$, and let A^L contain L linearly independent rows of $A_{\mathcal{J}}^M$. Now $A_{\mathcal{J}}^L$ is a square $L \times L$ invertible matrix. The matrix $A_{\mathcal{I} \cup \mathcal{J}}^L$ has rank L , as the span of $A_{\mathcal{I}}^L$ belongs to span of $A_{\mathcal{J}}^L$.

Now suppose we receive an additional measurement, and it still holds that $\mathbf{y}_{1:M+1} = A^{M+1} \hat{\mathbf{x}}^M$. This can happen in two cases: either the part of the new row $[\mathbf{a}_{M+1}]_{\mathcal{I} \cup \mathcal{J}}$ is linearly dependent on rows of $A_{\mathcal{I} \cup \mathcal{J}}^L$, or it is linearly independent, but $\hat{\mathbf{x}}^M - \mathbf{x}^*$ still falls into $\text{Null}(A^{M+1})$, which is a smaller dimensional subspace of $\text{Null}(A^M)$. The second case happens with probability 0: suppose that $A_{\mathcal{I} \cup \mathcal{J}}^{L+1}$ has rank $L+1$ (here A^{L+1} is formed by appending row a_{M+1} to A^L). Then some column of $A_{\mathcal{I}}^{L+1}$ does not belong to $\text{span}(A_{\mathcal{J}}^{L+1})$. Since \mathbf{x}^* is in general position, then $A^{L+1} \mathbf{x}^* \neq A^{L+1} \hat{\mathbf{x}}^M$.

So it remains to analyze the possibility that \mathbf{a}_{M+1} is linearly dependent on the previous M measurements in locations $\mathcal{I} \cup \mathcal{J}$. Using Lemma 1 this happens with probability at most $1/2$: since $\mathcal{I} \neq \mathcal{J}$, $|\mathcal{I} \cup \mathcal{J}| \geq |\mathcal{J}| + 1$. Hence, after waiting for T steps the probability that all T new measurements are linearly dependent on the first M is 2^{-T} . \square

We now pursue an alternative line of analysis, similar to that of Proposition 2. For the Bernoulli case, $\|\hat{\mathbf{x}}^M\|_0 < M$ does not imply $\hat{\mathbf{x}}^M = \mathbf{x}^*$. However, we believe that if $N^2 2^{1-M} \ll 1$, then $\hat{\mathbf{x}}^M = \mathbf{x}^*$ with high probability. Since the elements of \mathbf{a}_j belong to finite set $\{-1, 1\}$, an $M \times M$ submatrix of A^M can be singular with non-zero probability. Surprisingly, characterizing this probability is a very hard question. It is conjectured [3] that the dominant source of singularity is the event that two columns or two rows are equal or opposite in sign. This leads to the following estimate (here X_M is $M \times M$):²

$$P(\det X_M = 0) = (1 + o(1))M^2 2^{1-M} \quad (3)$$

However the very recent best provable bound on this probability is still rather far: $P(\det X_M = 0) = ((\frac{3}{4} + o(1))^M)$ [3]. If

²Probability that two columns are equal or opposite in sign is 2^{1-M} , and there are $O(M^2)$ pairs of columns.

we assume that the simple estimate based on pairs of columns is accurate, similar analysis shows that the probability that a random ± 1 $M \times N$ matrix with $M \ll N$ having all $M \times M$ submatrices non-singular is $(1 + o(1))N^2 2^{1-M}$.

4. NEAR-SPARSE SIGNALS

In practical settings, e.g. when taking Fourier and wavelet transforms of smooth signals, we may only have approximate sparseness: a few values are large, and most are very small. Results in the CS literature assume power-law decay of entries of \mathbf{x}^* (upon sorting) and show that with roughly $O(K \log N)$ samples, $\hat{\mathbf{x}}^M$ in (1) will have similar error to that of keeping the K largest entries in \mathbf{x}^* [1]. We consider a different line of analysis that does not assume a model of decay, and that compares subsequent solutions (in our sequential setting) to bound the reconstruction error.

We consider Gaussian A^M for this section. The stopping rule from Section 2 is vacuous for near-sparse signals, as $\|\mathbf{x}^*\|_0 = N$ and all samples are needed for perfect recovery. We modify the rule to stop when $d(\hat{\mathbf{x}}^{M+T}, \hat{\mathbf{x}}^M) = \|\hat{\mathbf{x}}^{M+T} - \hat{\mathbf{x}}^M\|_2$ is sufficiently small, and show that this guarantees that the error in the approximation $d(\hat{\mathbf{x}}^M, \mathbf{x}^*)$ is at a desired tolerance.

Consider Figure 1 again. Solution $\hat{\mathbf{x}}^{M+T}$ lies on the hyperplane $H_{M+T} \triangleq \{\mathbf{x} \mid y_i = \mathbf{a}_i^T \mathbf{x}, i = 1, \dots, M+T\}$. Let θ_T be the angle between the line connecting \mathbf{x}^* with $\hat{\mathbf{x}}^M$, and H_{M+T} . We have

$$d(\mathbf{x}^*, \hat{\mathbf{x}}^M) = \frac{d(\hat{\mathbf{x}}^M, H_{M+T})}{\sin(\theta_T)} \leq \frac{d(\hat{\mathbf{x}}^{M+T}, \hat{\mathbf{x}}^M)}{\sin(\theta_T)} \quad (4)$$

θ_T is a random variable - the angle between a fixed vector in \mathbb{R}^{N-M} and a random $N - (M+T)$ dimensional hyperplane. We next analyze the distribution of θ_T , which allows to infer when the reconstruction error $d(\mathbf{x}^*, \hat{\mathbf{x}}^M)$ is small enough.

Let $L = N - M$. In the Gaussian case (due to invariance to orthogonal transformations) it is equivalent to consider the angle θ between a fixed $(L - T)$ -dimensional subspace H and a random vector \mathbf{h} in \mathbb{R}^L . Let H be the span of the last $L - T$ coordinate vectors, and \mathbf{h} be i.i.d. Gaussian. Then: $\frac{1}{\sin(\theta)} = \sqrt{\sum_{i=1}^L x_i^2} / \sqrt{\sum_{i=1}^T x_i^2}$.

Using the properties of χ_L , χ_L^2 , and inverse- χ_L^2 distributions [4] and Jensen's inequality, we have $E[\frac{1}{\sin(\theta)}] \geq \sqrt{\frac{L}{T}}$ (we also have $E[\frac{1}{\sin(\theta)}] \approx \sqrt{\frac{L}{T}}$ as shown in Figure 3), and an upper bound on the variance:³

$$\text{Var} \left[\frac{1}{\sin(\theta)} \right] \leq \frac{L-2}{T-2} - \frac{L}{T} \quad (5)$$

³Consider $E[\sin(\theta)^2] = \left(\sum_{i=1}^T x_i^2 \right) / \|\mathbf{x}\|_2^2$. We have $E[\frac{x_i^2}{\|\mathbf{x}\|_2^2}] = \frac{1}{L}$ (Dirichlet dist.), so $E[\sin(\theta)^2] = \frac{T}{L}$. Using Jensen's ineq. with $\sqrt{1/x}$, $E[1/\sin(\theta)] \geq \sqrt{\frac{L}{T}}$. Finally, $E[\frac{1}{\sin(\theta)^2}] = \frac{L-2}{T-2}$ (for $T > 2$).

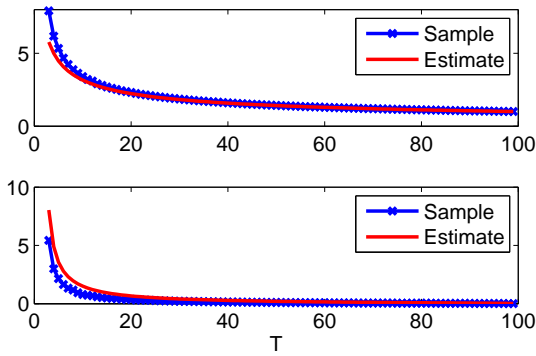


Fig. 3. Mean estimate (top) and standard deviation bound (bottom) of $\frac{1}{\sin(\theta)}$ vs. averages over 2500 samples. $L = 100$.

In Figure 3 we plot the mean estimate and the standard deviation bound for $L = 100$ and a range of T . We compare them to sample mean and standard deviation of $\frac{1}{\sin(\theta)}$ based on 2500 samples. They give very good approximation for most of the range of $T > 2$. Standard deviations quickly fall off with T , giving tight confidence intervals (by Chebyshev ineq. $p(|a - E[a]| \geq k\sigma_a) \leq \frac{1}{k^2}$). We can use this to gauge the reconstruction error: if $d(\hat{\mathbf{x}}^M, \hat{\mathbf{x}}^{M+T})$ is small, then we have a confidence interval on how small is $d(\hat{\mathbf{x}}^M, \mathbf{x}^*)$.

5. EFFICIENT SEQUENTIAL SOLUTION

The main motivation for the sequential approach is to reduce the number of measurements to as few as possible. Yet, we would also like to keep the computational complexity of the sequential approach low. Instead of re-solving the linear program (1) after each new sample, we would like to use the solution to the previous problem to guide the current problem. We now investigate a linear programming approach to accomplish this. In related work, [5] proposed to use Row-action methods for compressed sensing, which rely on a quadratic programming formulation equivalent to (1) and can take advantage of sequential measurements.

We can not use the solution $\hat{\mathbf{x}}^M$ directly as a starting point for the new problem at step $M + 1$, because it in general will not be feasible. In the Gaussian measurement case, unless $\hat{\mathbf{x}}^M = \mathbf{x}^*$, the new constraint $\mathbf{a}'_{M+1}\hat{\mathbf{x}}^M = y(M + 1)$ will be violated. One way to handle this is through a dual formulation, but we instead use an augmented primal formulation [6].

First, to model (1) as a linear program we use the standard trick: define $x_i^+ = \max(x_i, 0)$, $x_i^- = \max(-x_i, 0)$, and $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$. This gives a linear program in standard form:

$$\begin{aligned} \min \mathbf{1}'\mathbf{x}^+ + \mathbf{1}'\mathbf{x}^- \\ \mathbf{y}_{1:M} = [A^M - A^M] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{bmatrix}, \text{ and } \mathbf{x}^+, \mathbf{x}^- \geq 0 \end{aligned} \quad (6)$$

Next we need to add an extra constraint $y_{M+1} = \mathbf{a}'_{M+1}\mathbf{x}^+ - \mathbf{a}'_{M+1}\mathbf{x}^-$. Suppose that $\mathbf{a}'_{M+1}\hat{\mathbf{x}}^M > y(M + 1)$. We add an

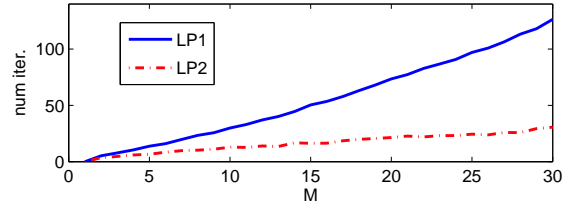


Fig. 4. A comparison of the number of simplex iterations when solving (1) from scratch (LP1) and using the solution at step $M - 1$ (LP2). Plot shows # iter. vs. M , over 100 trials.

extra slack variable z to the linear program, and a high positive cost Q on z . This gives the following linear program:

$$\min \mathbf{1}'\mathbf{x}^+ + \mathbf{1}'\mathbf{x}^- + Qz \quad (7)$$

$$\mathbf{y}_{1:M} = [A^M - A^M] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \end{bmatrix}, \text{ and } \mathbf{x}^+, \mathbf{x}^- \geq 0$$

$$y_{M+1} = \mathbf{a}'_{M+1}\mathbf{x}^+ - \mathbf{a}'_{M+1}\mathbf{x}^- - z, \text{ and } z \geq 0$$

Now using $\hat{\mathbf{x}}^M$ and $z = \mathbf{a}'_{M+1}(\hat{\mathbf{x}}^M)^+ - \mathbf{a}'_{M+1}(\hat{\mathbf{x}}^M)^- - y_{M+1}$ yields a basic feasible solution to this augmented problem. By selecting Q large enough, z will be removed from the optimal basis (i.e. z is set to 0), and the solutions to this problem and the $(M + 1)$ -th sequential problem are the same.

We test the approach on an example with $N = 200$, $K = 10$, and 100 trials. In Figure 4 we plot the number of iterations of the simplex method required to solve the problem (1) at step M from scratch (LP1) and using the formulation in (7) (LP2). To solve (6) we first have to find a basic feasible solution (phase 1) and then move from it to the optimal BFS. An important advantage of (7) is that we start with a basic feasible solution, so phase 1 is not required. The figure illustrates that for large M the approach LP2 is significantly faster.

6. REFERENCES

- [1] E. J. Candes, “Compressive sampling,” in *Proc. Int. Congress of Math.*, 2006, Madrid, Spain.
- [2] M. Rudelson and R. Vershynin, “Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements,” in *CISS 2006*, 2006.
- [3] T. Tao and V. Vu, “On the singularity probability of random Bernoulli matrices,” *Journal Amer. Math. Soc.*, vol. 20, pp. 603–628, 2007.
- [4] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*, Wiley and Sons, 2000.
- [5] S. Sra and J. A. Tropp, “Row-action methods for compressed sensing,” in *IEEE Intern. Conf. on Acoustics, Speech, and Signal Proc.*, 2006, vol. 3, pp. 868–871.
- [6] D. Bertsimas and J. N. Tsitsiklis, *Introduction to linear optimization*, Athena Scientific, 1997.