ROW-ACTION METHODS FOR COMPRESSED SENSING

Suvrit Sra[t] *and Joel A. Tropp*[m]

[t] Department of Computer Sciences, The University of Texas at Austin
[m] Department of Mathematics, The University of Michigan at Ann Arbor

ABSTRACT

Compressed Sensing uses a small number of random, linear measurements to acquire a sparse signal. Nonlinear algorithms, such as ℓ_1 minimization, are used to reconstruct the signal from the measured data. This paper proposes *row-action* methods as a computational approach to solving the ℓ_1 optimization problem. This paper presents a specific row-action method and provides extensive empirical evidence that it is an effective technique for signal reconstruction. This approach offers several advantages over interior-point methods, including minimal storage and computational requirements, scalability, and robustness.

1. COMPRESSED SENSING

The usual paradigm for source coding places the computational burden on the encoder so that the decoder can operate efficiently. This approach is appropriate for applications such as video compression, where real-time decoding is essential. One can imagine other types of applications—in sensor networks, for example—where the encoder is severely constrained but the decoder can afford to do some work. In this setting, it is natural to ask how to design an encoding process that performs a limited amount of nonadaptive computation.

An approach called *Compressed Sensing* has recently been promoted by the computational harmonic analysis community [1, 2, 3, 4]. The basic idea is that certain classes of compressible signals are determined almost completely by a low-dimensional random projection. To implement this compression scheme, the encoder performs a linear operation on the signal. In compensation, decoding requires more costly nonlinear algorithms.

To be more concrete, we consider the set of d-dimensional real-valued signals with no more than m nonzero entries. We refer to signals of this form as *m-sparse*. It is evident that no coding scheme can represent an m-sparse signal using fewer than m real numbers. In fact, a comparable amount of *nonadaptive* information is sufficient to compress these signals.

Theorem 1 (Rudelson–Vershynin [3]) *Fix a number $N \geq Cm\log(d/m)$, and draw a random $N \times d$ matrix $\mathbf{\Phi}$ whose entries are iid* $\textsc{Normal}(0,1)$. *With overwhelming probability, all m-sparse signals in $\mathbb{R}^d$ can be recovered from their images under $\mathbf{\Phi}$. The number C is an absolute constant.*

In other words, an m-sparse signal in $\mathbb{R}^d$ can be compressed *losslessly* using about $m\log(d/m)$ linear measurements. Similar results hold for other signal classes, such as the ℓ_1 unit ball, and for other types of random matrices [1, 2, 5].

To recover a signal from its compressed version, at least three different algorithmic approaches are available.

- Orthogonal Matching Pursuit [4]
- Nonlinear programming [6]
- Convex relaxation [1, 2, 3]

This work concentrates on the latter approach. Suppose that $\boldsymbol{s}$ is a signal that compresses to $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{s}$. To reconstruct $\boldsymbol{s}$, convex relaxation elicits the signal $\widehat{\boldsymbol{s}}$ of minimal ℓ_1 norm that compresses to $\boldsymbol{y}$. That is,

$$\widehat{\boldsymbol{s}} = \arg\min_{\boldsymbol{x}} \;\|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{x} = \boldsymbol{y}. \qquad \text{(P1)}$$

Under the hypotheses of Theorem 1, the solution to (P1) is identical with the original signal: $\widehat{\boldsymbol{s}} = \boldsymbol{s}$.

The convex relaxation method provides a powerful approach to signal reconstruction. Unfortunately, the literature on Compressed Sensing does not discuss some of the important algorithmic issues that arise in engineering contexts. Most work assumes that an interior-point method is appropriate for solving (P1). In this article, we describe another approach, called a *row-action method*, which has several qualities that make it appealing for Compressed Sensing:

- **Linear algebra.** Row-action methods require only vector arithmetic, such as inner products and *saxpy* operations.
- **Scalability.** Row-action methods can be used for extremely large-scale problems.
- **Streaming data.** Row-action methods can take advantage of a measurements that are presented sequentially.
- **Robustness.** Row-action methods perform well even if the compressed data or measurement matrix are corrupted.

JAT was supported by NSF DMS Grant No. 0503299.
E-mail: suvrit@cs.utexas.edu, jtropp@umich.edu

1-4244-0469-X/06/$20.00 ©2006 IEEE

The major shortcoming of row-action methods is that they typically converge linearly, in contrast with the quadratic convergence of interior-point methods. This issue is not critical unless it is necessary to reconstruct signals with very high accuracy.

2. ROW-ACTION METHODS

The basic idea behind a row-action method is to enforce the constraints in an optimization problem sequentially. This approach can be traced back to Hildreth's Method for quadratic programming [7] and Bregman's Method for convex programming with linear constraints. These methods all require that the objective function be strictly convex. To develop a row-action method for solving (P1), therefore, we must first replace it with an equivalent quadratic program.

To that end, we rewrite (P1) as a linear program

$$\min_{\boldsymbol{x},\boldsymbol{f}} \ \langle \mathbf{0},\ \boldsymbol{x}\rangle + \langle \mathbf{1},\ \boldsymbol{f}\rangle \quad \text{subject to}$$

$$\boldsymbol{\Phi x} = \boldsymbol{y}, \ \ -\boldsymbol{x} - \boldsymbol{f} \le \mathbf{0}, \ \ \boldsymbol{x} - \boldsymbol{f} \le \mathbf{0} \quad \text{(LP)}$$

where the auxiliary variable $\boldsymbol{f} = |\boldsymbol{x}|$. Define the concatenation $\boldsymbol{z} = [\boldsymbol{x}; \boldsymbol{f}]$ and the conformal vector $\boldsymbol{c} = [\mathbf{0}; \mathbf{1}]$. Now, we replace the linear program (LP) with the quadratic program

$$\min_{\boldsymbol{z} \in Z} \ \|\boldsymbol{z} + \lambda \boldsymbol{c}\|_2 \qquad \text{(QP)}$$

where Z is the feasible set for (LP) and λ is a positive parameter. Theorem 2.1 of [8] states that the unique minimizer of (QP) is identical with the minimum ℓ_2-norm solution of the linear program (LP), provided λ exceeds a constant $\lambda_0 > 0$.

The quadratic program (QP) can be solved using a row-action method. At each iteration, this approach enforces all the equality constraints. Then it enforces the inequality constraints, applying corrections to ensure convergence to the global optimum. Algorithm 1 implements this procedure.

Theorem 2 *Algorithm 1 converges to the solution of* (QP).

Proof. The problem (QP) has a strictly convex objective function and a convex constraint set. Standard convergence results for row-action methods in [9, Chapter 6] ensure that Algorithm 1 converges to the optimal solution of (QP). ■

Algorithm 1 requires three inputs: the measurement matrix $\boldsymbol{\Phi}$, the compressed signal $\boldsymbol{y}$, and a parameter λ. It can be shown that λ_0, the minimum value for this parameter, is obtained as the solution to a linear-fractional program. Unfortunately, this approach does not seem to yield theoretical predictions for λ_0. Experiments with different values of λ reveal that the algorithm is quite stable with respect to a range of λ values.

The empirical convergence of Algorithm 1 is linear. This fact yields error estimates that can be used to halt the algorithm automatically. We omit this discussion.

ALGORITHM 1: Row-action method for (QP).

```
ROW_ACTION($\boldsymbol{\Phi}$, $\lambda$, $\boldsymbol{y}$)
Output: $\boldsymbol{z} = \arg\min_{\boldsymbol{z} \in \boldsymbol{Z}} \|\boldsymbol{z} + \lambda \boldsymbol{c}\|_2$
$\boldsymbol{x} \leftarrow \mathbf{0}$; $\boldsymbol{f} \leftarrow -\lambda \mathbf{1}$; $\boldsymbol{\mu} \leftarrow \boldsymbol{\nu} \leftarrow \mathbf{0}$.
while not converged
    {Enforce the hyperplane constraints}
    foreach $i : \langle \boldsymbol{\phi}_i,\ \boldsymbol{x}\rangle = y_i$
        $\boldsymbol{x} \leftarrow \boldsymbol{x} - \|\boldsymbol{\phi}_i\|^{-2} [\langle \boldsymbol{\phi}_i,\ \boldsymbol{x}\rangle - y_i]\, \boldsymbol{\phi}_i$
    end
    {Enforce the inequalities}
    $\boldsymbol{\theta} \leftarrow \min\{\boldsymbol{\mu}, \frac{1}{2}(\boldsymbol{f} + \boldsymbol{x})\}$
    $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu} - \boldsymbol{\theta}$
    $\boldsymbol{x} \leftarrow \boldsymbol{x} - \boldsymbol{\theta}$;  $\boldsymbol{f} \leftarrow \boldsymbol{f} - \boldsymbol{\theta}$
    $\boldsymbol{\theta} \leftarrow \min\{\boldsymbol{\nu}, \frac{1}{2}(\boldsymbol{f} - \boldsymbol{x})\}$
    $\boldsymbol{\nu} \leftarrow \boldsymbol{\nu} - \boldsymbol{\theta}$
    $\boldsymbol{x} \leftarrow \boldsymbol{x} + \boldsymbol{\theta}$;  $\boldsymbol{f} \leftarrow \boldsymbol{f} - \boldsymbol{\theta}$
end.
return $\{\boldsymbol{z} = [\boldsymbol{x}; \boldsymbol{f}]\}$.
```

2.1. Nonnegative signals

We also mention the problem of compressing a m-sparse nonnegative signal $\boldsymbol{s}$. One approach to reconstructing the signal from data $\boldsymbol{y} = \boldsymbol{\Phi s}$ uses the following convex relaxation [10].

$$\widehat{\boldsymbol{s}} = \arg\min_{\boldsymbol{x}} \ \langle \boldsymbol{x},\ \mathbf{1}\rangle \quad \text{subject to}$$

$$\boldsymbol{\Phi x} = \boldsymbol{y}, \quad \boldsymbol{x} \ge \mathbf{0} \quad \text{(P1+)}$$

The row-action method for this problem is very similar, so we omit the derivation.

3. NUMERICAL EXPERIMENTS

We have carried out numerous experiments to demonstrate that row-action methods provide a competitive approach to reconstructing signals from random measurements. We implemented the row-action methods in C++, and we performed a series of experiments on a Linux-based P4 3.0GHz machine equipped with 1GB RAM. In the sequel, a random measurement matrix is a matrix with iid entries drawn from $\{0, \pm 1\}$ with probabilities 0.5, 0.25 and 0.25. For Compressed Sensing, these matrices behave like Gaussian matrices [5], but they permit faster arithmetic and they can be stored efficiently. A random m-sparse signal has its nonzero entries at uniformly random locations, and each nonzero is iid $\textsc{Normal}(0, 1)$. Unless otherwise stated, all experiments use Algorithm 1 to solve (P1), halting after 5000 iterations.

Our first experiment is designed to provide some insight on selecting the parameter λ. For three different choices of the sparsity level m, the number of measurements N, and the signal length d, we draw a random measurement matrix. For each choice of λ, we execute the algorithm with 20 random signals, and we calculate the median relative error. The results appear in Figure 1. It is clear that, when λ is sufficiently large, at least half the random experiments lead to reconstruction

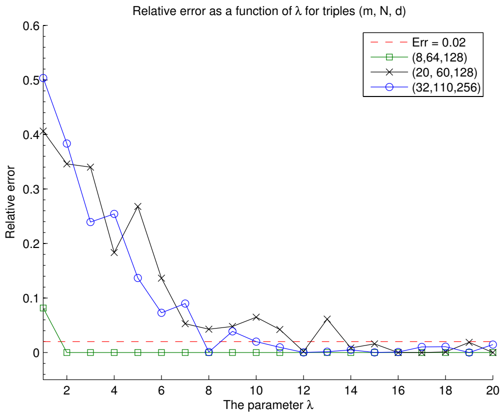

Fig. 1. *The median relative error in reconstruction as a function of λ for three different pairs (m, N, d). The dashed line marks a relative error of* 0.02.

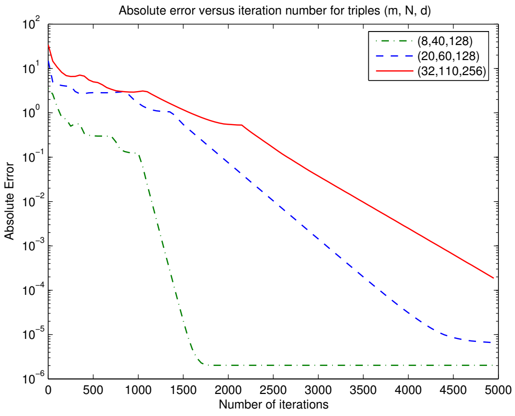

Fig. 2. *Absolute reconstruction error as a function of iteration number for three choices of* (m, N, d).

with a relative error $\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_2 / \|\boldsymbol{s}\|_2 \leq 0.02$. The empirical evidence also suggests that the choice $\lambda = km$ for a small constant k leads to accurate reconstruction in fewer iterations. The remaining experiments select λ according to this rule.

The aim of the next experiment is to identify the order of convergence of Algorithm 1 for m-sparse signals. Figure 2 displays the absolute error $\|\widehat{\boldsymbol{s}} - \boldsymbol{s}\|_2$ as a function of the iteration number for three different choices of (m, N, d). The semilog plot clearly indicates a linear order of convergence in all three sequences. It appears that arithmetic errors prevent the row-action method from attaining an arbitrarily small error. Note that the algorithm yields comparable relative errors in all three experiments, which may not be apparent because the input signals have different norms.

Although the empirical convergence of Algorithm 1 is only linear, the method is still quite fast because of its simplicity. The next set of experiments attempts to determine

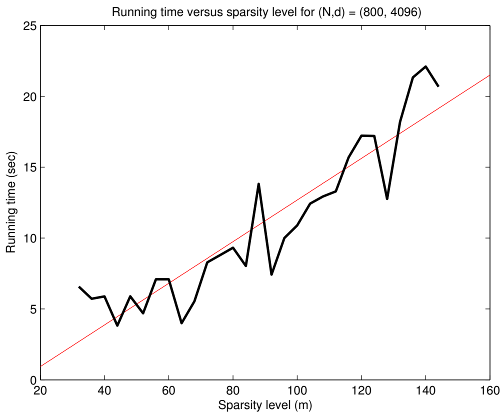

Fig. 3. *Running time to achieve a relative error of 0.02 as a function of the sparsity level m. Here, $N = 800$ and $d = 4096$. The regression line has equation* $t = 0.147m - 2.01$.

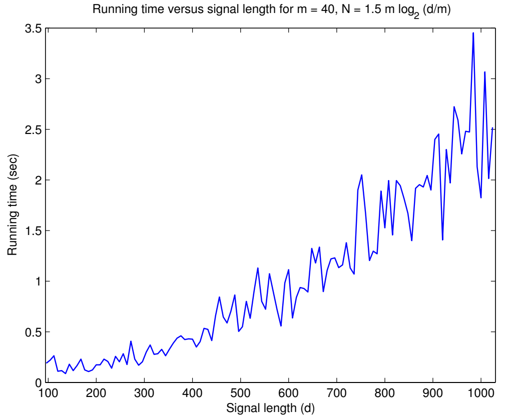

Fig. 4. *Running time to achieve a relative error of 0.02 as a function of the signal length d. Here, $m = 40$ and $N =$* $1.5\, m \log_2(d/m)$.

how the running time of the algorithm depends on (m, N, d). In these cases, we halt the algorithm as soon as the relative error declines below 0.02. Figure 3 demonstrates that the running time is approximately a linear function of the sparsity level m, when the number of measurements N and the signal length d are fixed. As a specific example, the row-action method achieves a relative error of 0.02 in less than 20 seconds for a 156-sparse signal of length $d = 4096$ using $N = 800$ measurements. Figure 4 reports the dependence of the running time on the signal length d. In this experiment, $m = 40$ and $N = 1.5\, m \log_2(d/m)$. For these choices, it appears that the running time is approximately $O(d \log d)$.

Our final two experiments are designed to show that row-action methods succeed across the entire regime where N measurements characterize m-sparse signals from $\mathbb{R}^d$, according to theory [11, 10]. Fix the signal length $d = 200$. We

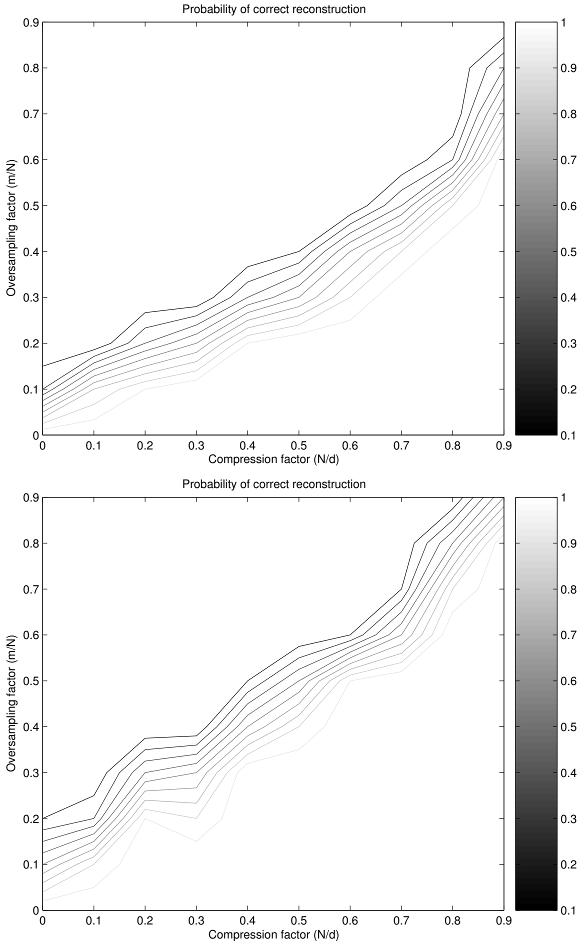

Fig. 5. [Top] *Probability of correctly reconstructing an m-sparse signal using Algorithm 1 to solve* (P1). [Bottom] *Probability of reconstructing a nonnegative m-sparse signal using a row-action method to solve* (P1+). *The horizontal axis is the compression factor N/d, and the vertical axis is the oversampling factor m/N. The signal length $d = 200$.*

let the number of measurements N range from 0 to d in 10 equal steps, and for each value of N, we let the sparsity level m range from 0 to N in 10 equal steps. For each choice of (m, N), we draw a random matrix. Then we use Algorithm 1 to solve (P1) for each of 20 random vectors. If the relative error in reconstruction is less than 0.02, we deem the experiment a success. Figure 5 charts the probability of success. Figure 5 also exhibits the results of a similar experiment using a row-action method to solve (P1+). For the second experiment, the nonzero entries of the random signals have UNIFORM(0,1) entries. Observe the sharp phase transition between success and failure, which matches the theoretical predictions in Donoho's work [11, 10].

Table 1 reports comparisons against CPLEX$^{\text{TM}}$, a state-of-the-art optimization package. Row-action methods achieve relative errors 10^{-2} smaller in the same amount of time.

m	C1 (Time/Err)	C2 (Time/Err)	Row (Time/Err)
32	279s / 3.7e-7	7.3s / 2.1e-1	6.7s / 9.1e-4
140	595s / 5.0e-7	83s / 9.1e-1	79.8s / 5.0e-3

Table 1. *Relative errors: CPLEX vs. row-action. Headings C1, C2 refer to running CPLEX until convergence or halting it early. Signal parameters are $(N, d) = (800, 4096)$*

4. CONCLUSIONS AND FUTURE WORK

These experiments show that row-action methods offer a compelling approach to signal reconstruction in Compressed Sensing problems. We are pursuing several other directions:

- Experiments for other signal classes, e.g., weak ℓ_p balls.
- Study potential gains arising from structured measurement matrices.
- Extension to large-scale problems, via (parallel) SOR row-action schemes.
- Develop row-action methods for related problems, such as mixed ℓ_1/ℓ_2 objective functions.

5. REFERENCES

[1] D. L. Donoho, "Compressed sensing," unpublished manuscript, Oct. 2004.

[2] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," submitted for publication, Nov. 2004.

[3] M. Rudelson and R. Vershynin, "Geometric approach to error correcting codes and reconstruction of signals," submitted for publication, Feb. 2005.

[4] J. A. Tropp and A. C. Gilbert, "Signal recovery from partial information via Orthogonal Matching Pursuit," submitted to *IEEE Trans. Inform. Theory*, April 2005.

[5] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian processes," *Comptes Rendus Acad. Sci.*, , no. 340, pp. 885–888, 2005.

[6] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," submitted to *IEEE Trans. Info. Theory*, 2005.

[7] C. Hildreth, "A quadratic programming procedure," *Naval Res. Logistics Quart.*, vol. 4, pp. 79–85, 1957.

[8] O. L. Mangasarian, "Normal solutions of linear programs," *Mathematical Programming Study*, vol. 22, pp. 206–216, 1984.

[9] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Num. Math. and Sci. Comp.. Oxford Univ. Press, 1997.

[10] D. L. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," Dept. of Statistics TR 2005-6, Stanford Univ., Apr. 2005.

[11] D. L. Donoho, "Neighborly polytopes and sparse solution of underdetermined linear equations," Dept. of Statistics TR 2005-4, Stanford Univ., Feb. 2005.