AN EFFICIENT METHOD FOR COMPRESSED SENSING

Seung-Jean Kim, Kwangmoo Koh, Michael Lustig, and Stephen Boyd

Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9510 USA
{sjkim,deneb1,mlustig,boyd}@stanford.edu

ABSTRACT

Compressed sensing or compressive sampling (CS) has been receiving a lot of interest as a promising method for signal recovery and sampling. CS problems can be cast as convex problems, and then solved by several standard methods such as interior-point methods, at least for small and medium size problems. In this paper we describe a specialized interior-point method for solving CS problems that uses a preconditioned conjugate gradient method to compute the search step. The method can efficiently solve large CS problems, by exploiting fast algorithms for the signal transforms used. The method is demonstrated with a medical resonance imaging (MRI) example.

Index Terms— compressed sensing, compressive sampling, ℓ_1 regularization, interior-point methods, preconditioned conjugate gradients.

1. INTRODUCTION

1.1. Compressed sensing

Let z be an unknown vector in $\mathbf{R}^n$. Suppose that we have m noisy linear measurements of z of the form

$$y_i = \langle \phi_i, z \rangle + v_i, \quad i = 1, \ldots, m,$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product, $v \in \mathbf{R}^m$ is the noise, and $\phi_i \in \mathbf{R}^n$ are known signals. Standard reconstruction methods require at least n samples. Suppose we know a priori that z is compressible or has a sparse representation in a transform domain, described by $W \in \mathbf{R}^{n \times n}$ (after expanding the real and imaginary parts if necessary). In this case, if the measurement vectors are well chosen, then the number of measurements m can be dramatically smaller than n usually considered necessary.

Compressed sensing [1] or compressive sampling [2] exploits the sparsity or compressibility in the transform domain by solving a problem of the form

$$\text{minimize} \quad \|\Phi z - y\|_2^2 + \lambda \|Wz\|_1 \tag{1}$$

where the variable is $z \in \mathbf{R}^n$ and $\|x\|_1 = \sum_i |x_i|$ denotes the ℓ_1 norm. Here, $\Phi = [\phi_1 \cdots \phi_m]^T \in \mathbf{R}^{m \times n}$ is called the compressed sensing matrix, $\lambda > 0$ is the regularization parameter, and W is called the sparsifying transform.

1.2. Solution methods

When W is invertible, the CS problem (1) can be reformulated as the ℓ_1-regularized least squares problem (LSP)

$$\text{minimize} \quad \|Ax - y\|_2^2 + \lambda \|x\|_1 \tag{2}$$

where the variable is $x \in \mathbf{R}^n$ and the problem data or parameters are $A = \Phi W^{-1} \in \mathbf{R}^{m \times n}$ and $y \in \mathbf{R}^m$. The ℓ_1-regularized problem (2) can be transformed to a convex quadratic program (QP), with linear inequality constraints,

$$
\begin{aligned}
\text{minimize} \quad & \|Ax - y\|^2 + \sum_{i=1}^n \lambda u_i \\
\text{subject to} \quad & -u_i \leq x_i \leq u_i, \quad i = 1, \ldots, n,
\end{aligned} \tag{3}
$$

where the variables are $x \in \mathbf{R}^n$ and $u \in \mathbf{R}^n$.

The data matrix A is typically fully dense, and so small and medium sized problems can be solved by standard convex optimization methods such as interior-point methods. The QP (3) that arises in compressed sensing applications has an important difference from general dense QPs: there are a fast method for multiplying a vector by A and a fast method for multiplying a vector by A^T, based on fast algorithms for the sparsifying transform and its inverse transform. A specialized interior-point method that exploits such algorithms may scale to large problems [3]. An example is `l1-magic` [4], which uses the conjugate gradient (CG) method to compute the search step.

Specialized computational methods for problems of the form (2) include path-following methods and variants [5, 6, 7]. When the optimal solution of (2) is extremely sparse, path-following methods can be very fast. Path-following methods tend to be slow, as the number of nonzeros at the optimal solution increases. Other recently developed computational methods for ℓ_1-regularized LSPs include coordinate-wise descent methods [8], bound optimization methods [9], sequential subspace optimization methods [10]), iterated shrinkage methods [11, 12], and gradient projection algorithms [13]. Some of these methods can handle very large problems with modest accuracy.

The main goal of this paper is to describe a specialized interior-point method for solving the QP (3). The method uses a preconditioned conjugate gradient (PCG) method to compute the search step and therefore can exploit fast algorithms for the sparsifying transform and its inverse transform.

The specialized method is far more efficient than (primal-dual) interior-point methods that use direct or CG methods to compute the search step. Compared with first-order methods such as coordinate descent methods, the specialized method is comparable in solving large problems with modest accuracy, but is able to solve them with high accuracy with relatively small additional computational cost. The method is demonstrated with an MRI example.

2. PRELIMINARIES

We describe some basic ingredients necessary for the interior-point method described in Section 3.

2.1. Dual problem

To derive a Lagrange dual of (2), we first write it in the equivalent form

$$\text{minimize} \quad z^T z + \sum_{i=1}^n \lambda |x_i|$$
$$\text{subject to} \quad z = Ax - y,$$

where the variables are $x \in \mathbf{R}^n$ and $z \in \mathbf{R}^m$. We associate dual variables $\nu_i \in \mathbf{R}$, $i = 1, \ldots, m$ with the equality constraints $z_i = (Ax - y)_i$. The Lagrange dual of (2) can be written as

$$\text{maximize} \quad G(\nu) = -(1/4)\nu^T \nu - \nu^T y$$
$$\text{subject to} \quad \|A^T \nu\|_\infty \leq \lambda. \tag{4}$$

The dual problem (4) is a convex optimization problem with a variable $\nu \in \mathbf{R}^m$. We say that $\nu \in \mathbf{R}^m$ is *dual feasible* if it satisfies the constraints in the dual problem (4). (See [14, §4] for more on Lagrange duality.)

Any dual feasible point ν gives a lower bound on the optimal value $p^\star$ of the primal problem (2), *i.e.*, $G(\nu) \leq p^\star$, which is called weak duality. Furthermore, the optimal value of the primal and dual are equal since the primal problem (3) satisfies Slater's condition, which is called strong duality [14].

2.2. Suboptimality bound

We are able to derive an easily computed bound on the suboptimality of x, by constructing a dual feasible point ν from an arbitrary x. The dual point

$$\nu = 2s(Ax - y), \tag{5}$$

with the scaling constant $s = \min\{1, \lambda/\|A^T \nu\|_\infty\}$, is dual feasible. Therefore $G(\nu)$ is a lower bound on the optimal value of (2). The difference between the primal objective value and the associated lower bound $G(\nu)$ is called the *duality gap* and denoted η:

$$\eta = \|Ax - y\|_2^2 + \sum_{i=1}^n \lambda |x_i| - G(\nu). \tag{6}$$

The duality gap is always nonnegative by weak duality, and is zero at the optimal point.

3. AN INTERIOR-POINT METHOD

We start by defining the logarithmic barrier for the bound constraints $-u_i \leq x_i \leq u_i$ in (3),

$$\Phi(x, u) = -\sum_{i=1}^n \log(u_i + x_i) - \sum_{i=1}^n \log(u_i - x_i)$$

with domain $\mathbf{dom}\, \Phi = \{(x, u) \in \mathbf{R}^n \times \mathbf{R}^n \mid |x_i| < u_i,\ i = 1, \ldots, n\}$. The central path consists of the unique minimizer of the convex function

$$\phi_t(x, u) = t\|Ax - y\|_2^2 + t\sum_{i=1}^n \lambda u_i + \Phi(x, u),$$

as the parameter t varies from 0 to ∞.

In the primal interior-point method, we compute a sequence of points on the central path, for an increasing sequence of values of t, starting from the previously computed central point. In the primal barrier method, Newton's method is used to minimize $\phi_t(x, u)$, *i.e.*, the search direction is computed as the exact solution to the Newton system

$$\nabla^2 \phi_t(x, u) \begin{bmatrix} \Delta x \\ \Delta u \end{bmatrix} = -\nabla \phi_t(x, u). \tag{7}$$

(The reader is referred to [14, Chap.11] for more on the primal barrier method.)

For a large ℓ_1-regularized LSP, solving the Newton system exactly is not computationally practical. In the method described below, the search direction is computed as an approximate solution to the Newton system, using a truncated Newton method.

In the primal barrier method, the parameter t is held constant until ϕ_t is (approximately) minimized, *i.e.*, $\|\nabla \phi_t\|_2$ is small; when this occurs, t is increased by a factor typically between 2 and 50. The method described below attempts to update the parameter t at each iteration, using the observation made above that we can cheaply compute a dual feasible point and associated duality gap for *any* x.

TRUNCATED NEWTON INTERIOR-POINT METHOD.

given relative tolerance $\epsilon_{\text{rel}} > 0$, $\alpha \in (0, 1/2)$, $\beta \in (0, 1)$

Initialize. $t := 1/\lambda$, $x := 0$, $u := \mathbf{1} = (1, \ldots, 1) \in \mathbf{R}^n$.

repeat
1. Compute the search direction $(\Delta x, \Delta u)$ as an approximate solution to the Newton system (7).
2. *Backtracking line search.*
   Find the smallest integer $k \geq 0$ that satisfies
   $$\phi_t(x + \beta^k \Delta x, u + \beta^k \Delta u)$$
   $$\leq \phi_t(x, u) + \alpha \beta^k \nabla \phi_t(x, u)^T \begin{bmatrix} \Delta x \\ \Delta u \end{bmatrix}.$$
3. *Update.* $(x, u) := (x, u) + \beta^k (\Delta x, \Delta u)$.
4. Construct dual feasible point ν from (5).

5. Evaluate duality gap η from (6).
6. **quit** if $\eta/G(\nu) \leq \epsilon_{\text{rel}}$.
7. *Update t.*

As a stopping criterion, the method uses the duality gap divided by the dual objective value. By weak duality, the ratio is an upper bound on the relative suboptimality.

The update rule we propose is

$$t := \begin{cases} \max\left\{\mu \min\{\hat{t}, t\}, t\right\}, & s \geq s_{\min} \\ t, & s < s_{\min} \end{cases}$$

where $\hat{t} = 2n/\eta$, and $s = \beta^k$ is the step length chosen in the line search. Here $\mu > 1$ and $s_{\min} \in (0, 1]$ are algorithm parameters. The choice of $\mu = 2$ and $s_{\min} = 0.5$ appears to give good performance for a wide range of problems. This rule has been used in solving l_1-regularized logistic regression problems in [15]. See [15] for an informal justification of convergence of the interior-point method based on this update rule (with exact search directions).

We compute the search direction approximately, applying the PCG method [16, §6.6] to the Newton system (7). It uses a (symmetric positive definite) preconditioner $P \in \mathbf{R}^{2n \times 2n}$ that approximates the Hessian of $\phi_t(x, u)$,

$$\nabla^2 \phi_t(x, u) = t\nabla^2 \|Ax - y\|^2 + \nabla^2 \Phi(x, u) \in \mathbf{R}^{2n \times 2n}.$$

The preconditioner approximates the first term with its diagonal entries, while retaining the second term:

$$P = \text{diag}(t\nabla^2\|Ax - y\|^2) + \nabla^2\Phi(x, u) \in \mathbf{R}^{2n \times 2n}.$$

(Here $\text{diag}(S)$ is the diagonal matrix obtained by setting the off-diagonal entries of the matrix S to zero.) The cost of computing the diagonal entries can be amortized over all interior-point iterations since we need to compute them only once.

The computational effort of each iteration of the PCG algorithm is dominated by one matrix-vector product of the form Hp with the Hessian $H = \nabla^2\phi_t(x, u)$ and one solve step of the form $P^{-1}r$ with the preconditioner P. The cost of computing Hp is cheap, since we can use fast algorithms for the transforms Φ and W (*e.g.*, fast discrete wavelet and Fourier transforms). The cost of computing $P^{-1}r_k$ is $O(n)$ flops.

The PCG algorithm has two parameters: the initial point x_0 and relative tolerance ϵ_{pcg}. For the initial point in the PCG algorithm, we use the previous search direction. The PCG relative tolerance parameter ϵ_{pcg} has to be carefully chosen to obtain good efficiency in the interior-point method. We change the relative tolerance adaptively as $\epsilon_{\text{pcg}} = \min\{0.1, \xi\eta/\|g\|_2\}$, where η is the duality gap at the current iterate and ξ is an algorithm parameter. (The choice of $\xi = 0.01$ appears to work well for a wide range of problems.) Thus, we solve the Newton system with low accuracy at early iterations, and solve it more accurately as the duality gap decreases.

4. APPLICATION TO SPARSE MRI

In this section we demonstrate the interior-point method described in Section 3 with real Magnetic Resonance Imaging (MRI) data, using algorithm parameters $\alpha = 0.01$, $\beta = 0.5$, $s_{\min} = 0.5$, $\mu = 2$, $\xi = 0.01$, and $\epsilon_{\text{rel}} = 0.05$. The regularization parameter is taken as $\lambda = 0.01$. The method was implemented in Matlab, and run on a 3.2GHz Pentium IV under Linux. (The Matlab implementation is available from `http://www.stanford.edu/~boyd/l1_ls/`.)

In MRI, samples are collected directly in the spatial frequency domain of the object of interest. The scan time in MRI is often proportional to the number of Fourier coefficients required for reconstruction. Using compressed sensing, one can significantly reduce the number of acquisition samples and hence the scan time. This approach is referred to as sparse or compressed sensing MRI [17].

We scanned the brain of a healthy volunteer. We obtained 205 out of 512 possible parallel lines in the spatial frequency of the image. The lines were chosen randomly with higher density sampling at low frequency achieving a 2.5 scan-time reduction factor, as illustrated in the left panel of Fig. 1. The compressed sensing matrix Φ in (1) is therefore a matrix obtained by randomly removing many rows of the discrete two-dimensional Fourier transform (DFT) matrix, called a random Fourier ensemble. Brain images have a sparse representation in the wavelet domain. In the example shown, we use the Daubechies 4 wavelet transform as the sparsifying transform W in (1).

We compared the sparse MRI method with a linear reconstruction method, which sets unobserved Fourier coefficients to zero and then performs the inverse Fourier transform. Fig. 1 shows the two reconstruction results. The linear reconstruction suffers from incoherent noise-like streaking artifacts (pointed by the arrow) due to undersampling, whereas the artifacts are much less noticeable in the compressed sensing reconstruction.

The QP for compressed sensing reconstruction has around $4 \times 512^2 \approx 10^6$ variables. (Here one half are the real and imaginary wavelet coefficients and the other half are new variables added in transforming the CS problem into a QP.) The run time of the Matlab implementation of our interior-point method was around 3 minutes, and the total number of PCG steps required over all interior-point iterations was 137. MOSEK [18] could not handle the QP, since forming the Hessian H (let alone computing the search direction) is prohibitively expensive for direct methods.

5. EXTENSIONS

Although not described here in detail, the interior-point method described in Section 3 can be readily extended to other problems that have a similar form. For instance, it can be readily extended to CS problems where the ℓ_1 norm of a complex
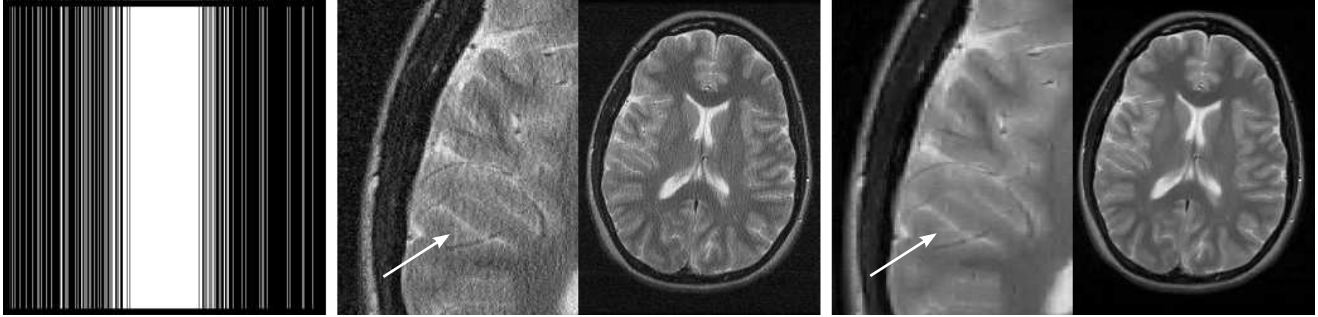
Figure 1: Brain image reconstruction results. *Left.* Collected partial Fourier coefficients (in white). *Middle.* Linear reconstruction. *Right.* Compressed sensing reconstruction.

vector is the sum of the absolute values of the complex elements.

6. REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. Candès, "Compressive sampling," in *Proceedings of International Congress of Mathematics*, 2006.

[3] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[4] E. Candès and J. Romberg, l_1-*magic: A collection of MATLAB routines for solving the convex optimization programs central to compressive sampling*, 2006, Available from www.acm.caltech.edu/l1magic/.

[5] D. Donoho and Y. Tsaig, "Fast solution of l_1-norm minimization problems when the solution may be sparse," 2006, Manuscript. Available from http://www.stanford.edu/%7Etsaig/Papers/FastL1.pdf.

[6] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.

[7] M. Osborne, B. Presnell, and B. Turlach, "A new approach to variable selection in least squares problems," *IMA Journal of Numerical Analysis*, vol. 20, no. 3, pp. 389–403, 2000.

[8] J. Friedman, T. Hastie T, and R. Tibshirani, "Pathwise coordinate optimization," 2007, Manuscript available from www-stat.stanford.edu/~hastie/pub.htm.

[9] M. Figueiredo and R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, 2005, pp. 782–5.

[10] G. Narkiss and M. Zibulevsky, "Sequential subspace optimization method for large-scale unconstrained problems," 2005, Tech. Report CCIT No 559, EE Dept., Technion.

[11] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, pp. 1413–1541, 2004.

[12] M. Elad, B. Matalon, and M. Zibulevsky, "Image denoising with shrinkage and redundant representations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recogn (CVPR)*, 2006, vol. 2, pp. 1924–1931.

[13] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," 2007, Manuscript. Available from http://www.lx.it.pt/~mtf/GPSR/.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[15] K. Koh, S.-J. Kim, and S. Boyd, "An interior-point method for ℓ_1-regularized logistic regression," 2007, To appear in *Journal of Machine Learning Research*. Available from www.stanford.edu/~boyd/l1_logistic_reg.html.

[16] J. Demmel, *Applied Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, 1997.

[17] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," 2007, Revised for publication in *Magnetic Resonance in Medicine*.

[18] MOSEK ApS, *The MOSEK Optimization Tools Version 2.5. User's Manual and Reference*, 2002, Available from www.mosek.com.