# Convex optimization-based methods*

Wotao Yin

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

**Abstract**

This module provides an overview of convex optimization approaches to sparse signal recovery.

An important class of sparse recovery algorithms[1] fall under the purview of *convex optimization*. Algorithms in this category seek to optimize a convex function $f(\cdot)$ of the unknown variable $x$ over a (possibly unbounded) convex subset of $\mathbb{R}^N$.

## 1 Setup

Let $J(x)$ be a convex *sparsity-promoting* cost function (i.e., $J(x)$ is small for sparse $x$.) To recover a sparse signal representation $\hat{x}$ from measurements $y = \Phi x, \Phi \in \mathbb{R}^{M \times N}$, we may either solve

$$\min_x \{J(x) : y = \Phi x\}, \tag{1}$$

when there is no noise, or solve

$$\min_x \{J(x) : H(\Phi x, y) \leq \varepsilon\} \tag{2}$$

when there is noise in the measurements. Here, $H$ is a cost function that penalizes the distance between the vectors $\Phi x$ and $y$. For an appropriate penalty parameter $\mu$, (2) is equivalent to the *unconstrained* formulation:

$$\min_x J(x) + \mu H(\Phi x, y) \tag{3}$$

for some $\mu > 0$. The parameter $\mu$ may be chosen by trial-and-error, or by statistical techniques such as cross-validation [2].

For convex programming algorithms, the most common choices of $J$ and $H$ are usually chosen as follows: $J(x) = \| x \|_1$, the $\ell_1$-norm of $x$, and $H(\Phi x, y) = \frac{1}{2} \| \Phi x - y \|_2^2$, the $\ell_2$-norm of the error between the observed measurements and the linear projections of the target vector $x$. In statistics, minimizing this $H$ subject to $\| x \|_1 \leq \delta$ is known as the *Lasso* problem. More generally, $J(\cdot)$ acts as a regularization term and can be replaced by other, more complex, functions; for example, the desired signal may be piecewise constant, and simultaneously have a sparse representation under a known basis transform $\Psi$. In this case, we may use a mixed regularization term:

$$J(x) = TV(x) + \lambda \| x \|_1 \tag{4}$$

*Version 1.5: Sep 23, 2011 7:46 pm -0500

†http://creativecommons.org/licenses/by/3.0/

[1] "Sparse recovery algorithms" <http://cnx.org/content/m37292/latest/>

It might be tempting to use conventional convex optimization packages for the above formulations ((1), (2), and (3)). Nevertheless, the above problems pose two key challenges which are specific to practical problems encountered in CS[2]: (i) real-world applications are invariably large-scale (an image of a resolution of $1024 \times 1024$ pixels leads to optimization over a million variables, well beyond the reach of any standard optimization software package); (ii) the objective function is nonsmooth, and standard smoothing techniques do not yield very good results. Hence, for these problems, conventional algorithms (typically involving matrix factorizations) are not effective or even applicable. These unique challenges encountered in the context of CS have led to considerable interest in developing improved sparse recovery algorithms in the optimization community.

## 2 Linear programming

In the noiseless[3] case, the $\ell_1$-minimization[4] problem (obtained by substituting $J(x) = \| x \|_1$ in (1)) can be recast as a linear program (LP) with equality constraints. These can be solved in polynomial time ($O\left(N^3\right)$) using standard interior-point methods [3]. This was the first feasible reconstruction algorithm used for CS recovery and has strong theoretical guarantees, as shown earlier in this course[5]. In the noisy case, the problem can be recast as a second-order cone program (SOCP) with quadratic constraints. Solving LPs and SOCPs is a principal thrust in optimization research; nevertheless, their application in practical CS problems is limited due to the fact that both the signal dimension $N$, and the number of constraints $M$, can be very large in many scenarios. Note that both LPs and SOCPs correspond to the constrained formulations in (1) and (2) and are solved using *first order* interior-point methods.

A newer algorithm called "l1_ls" [12] is based on an interior-point algorithm that uses a preconditioned conjugate gradient (PCG) method to approximately solve linear systems in a truncated-Newton framework. The algorithm exploits the structure of the Hessian to construct their preconditioner; thus, this is a second order method. Computational results show that about a hundred PCG steps are sufficient for obtaining accurate reconstruction. This method has been typically shown to be slower than first-order methods, but could be faster in cases where the true target signal is highly sparse.

## 3 Fixed-point continuation

As opposed to solving the constrained formulation, an alternate approach is to solve the unconstrained formulation in (3). A widely used method for solving $\ell_1$-minimization problems of the form

$$\underset{x}{min}\ \mu \| x \|_1 + H(x), \tag{5}$$

for a convex and differentiable $H$, is an iterative procedure based on *shrinkage* (also called soft thresholding; see (6) below). In the context of solving (5) with a quadratic $H$, this method was independently proposed and analyzed in [1], [9], [13], [14], and then further studied or extended in [4], [5], [6], [7], [11], [16]. Shrinkage is a classic method used in wavelet-based image denoising. The shrinkage operator on any scalar component can be defined as follows:

$$\text{shrink}(t, \alpha) = \{ \begin{array}{cc} t - \alpha & \text{if} t > \alpha, \\ 0 & \text{if} -\alpha \leq t \leq \alpha, \text{and} \\ t + \alpha & \text{if} t < -\alpha. \end{array} \tag{6}$$

This concept can be used effectively to solve (5). In particular, the basic algorithm can be written as following the fixed-point iteration: for $i = 1, ..., N$, the $i^{\text{th}}$ coefficient of $x$ at the $(k+1)^{\text{th}}$ time step is given

[2] "Introduction to compressive sensing" <http://cnx.org/content/m37172/latest/>
[3] "Noise-free signal recovery" <http://cnx.org/content/m37181/latest/>
[4] "Signal recovery via $\ell$_1 minimization" <http://cnx.org/content/m37179/latest/>
[5] "Signal recovery via $\ell$_1 minimization" <http://cnx.org/content/m37179/latest/>

by

$$x_i^{k+1} = \text{shrink}\left(\left(x^k - \tau\,\texttt{[U+25BD]}\,H\left(x^k\right)\right)_i, \mu\tau\right) \tag{7}$$

where $\tau > 0$ serves as a step-length for gradient descent (which may vary with $k$) and $\mu$ is as specified by the user. It is easy to see that the larger $\mu$ is, the larger the allowable distance between $x^{k+1}$ and $x^k$. For a quadratic penalty term $H(\cdot)$, the gradient [U+25BD]$H$ can be easily computed as a linear function of $x^k$; thus each iteration of (7) essentially boils down to a small number of matrix-vector multiplications.

The simplicity of the iterative approach is quite appealing, both from a computational, as well as a code-design standpoint. Various modifications, enhancements, and generalizations to this approach have been proposed, both to improve the efficiency of the basic iteration in (7), and to extend its applicability to various kinds of $J$[8], [10], [16]. In principle, the basic iteration in (7) would not be practically effective without a continuation (or path-following) strategy [11], [16] in which we choose a gradually decreasing sequence of values for the parameter $\mu$ to guide the intermediate iterates towards the final optimal solution.

This procedure is known as *continuation*; in [11], the performance of an algorithm known as Fixed-Point Continuation (FPC) has been compared favorably with another similar method known as Gradient Projection for Sparse Reconstruction (GPSR) [10] and "l1_ls" [12]. A key aspect to solving the unconstrained optimization problem is the choice of the parameter $\mu$. As discussed above, for CS recovery, $\mu$ may be chosen by trial and error; for the noiseless constrained formulation, we may solve the corresponding unconstrained minimization by choosing a large value for $\mu$.

In the case of recovery from noisy compressive measurements, a commonly used choice for the convex cost function $H(x)$ is the square of the norm of the *residual*. Thus we have:

$$\begin{aligned} H(x) &= \| y - \Phi x \|_2^2 \\ \texttt{[U+25BD]}\,H(x) &= 2\Phi^\top (y - \Phi x). \end{aligned} \tag{8}$$

For this particular choice of penalty function, (7) reduces to the following iteration:

$$x_i^{k+1} = \text{shrink}\left(\left(x^k - \tau\,\texttt{[U+25BD]}\,H\left(y - \Phi x^k\right)_i, \mu\tau\right) \tag{9}$$

which is run until convergence to a fixed point. The algorithm is detailed in pseudocode form below.

```
    Inputs: CS matrix Φ , signal measurements y, parameter sequence μ _n
Outputs: Signal estimate x̂
initialize: x̂_0 = 0, r = y, k = 0.
while alting criterion false do
    1. k ← k + 1
    2. x ←x̂ −τ Φ ˆTr {take a gradient step}
    3. x̂← shrink(x, μ _kτ ) {perform soft thresholding}
    4. r ← y − Φ x̂ {update measurement residual}
end while
return x̂←x̂
```

## 4 Bregman iteration methods

It turns out that an efficient method to obtain the solution to the constrained optimization problem in (1) can be devised by solving a small number of the unconstrained problems in the form of (3). These subproblems

are commonly referred to as *Bregman iterations*. A simple version can be written as follows:

$$\begin{aligned} y^{k+1} &= y^k + y - \Phi x^k \\ x^{k+1} &= argmin\ J(x) + \frac{\mu}{2} \| \Phi x - y^{k+1} \|^2. \end{aligned} \tag{10}$$

The problem in the second step can be solved by the algorithms reviewed above. Bregman iterations were introduced in [15] for constrained total variation minimization problems, and was proved to converge for closed, convex functions $J(x)$. In [17], it is applied to (1) for $J(x) = \| x \|_1$ and shown to converge in a finite number of steps for any $\mu > 0$. For moderate $\mu$, the number of iterations needed is typically lesser than 5. Compared to the alternate approach that solves (1) through directly solving the unconstrained problem in (3) with a very large $\mu$, Bregman iterations are often more stable and sometimes much faster.

## 5 Discussion

All the methods discussed in this section optimize a convex function (usually the $\ell_1$-norm) over a convex (possibly unbounded) set. This implies *guaranteed* convergence to the global optimum. In other words, given that the sampling matrix $\Phi$ satisfies the conditions specified in "Signal recovery via $\ell_1$ minimization"[6], convex optimization methods will recover the underlying signal $x$. In addition, convex relaxation methods also guarantee *stable* recovery by reformulating the recovery problem as the SOCP, or the unconstrained formulation.

# References


[1] J. Bect, L. Blanc-Feraud, G. Aubert, and A. Chambolle. A -unified variational framework for image restoration. In *Proc. European Conf. Comp. Vision (ECCV)*, Prague, Czech Republic, May 2004.

[2] P. Boufounos, M. Duarte, and R. Baraniuk. Sparse signal reconstruction from noisy compressive measurements using cross validation. In *Proc. IEEE Work. Stat. Signal Processing*, Madison, WI, Aug. 2007.

[3] S. Boyd and L. Vanderberghe. *Convex Optimization*. Cambridge Univ. Press, Cambridge, England, 2004.

[4] P. Combettes and J. Pesquet. Proximal thresholding algorithm for minimization over orthonormal bases. *SIAM J. Optimization*, 18(4):13518211;1376, 2007.

[5] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):14138211;1457, 2004.

[6] M. Elad. Why simple shrinkage is still relevant for redundant representations? *IEEE Trans. Inform. Theory*, 52(12):55598211;5569, 2006.

[7] M. Elad, B. Matalon, J. Shtok, and M. Zibulevsky. A wide-angle view at iterated shrinkage algorithms. In *Proc. SPIE Optics Photonics: Wavelets*, San Diego, CA, Apr. 2007.

[8] M. Elad, B. Matalon, and M. Zibulevsky. Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *Appl. Comput. Harmon. Anal.*, 23(3):3468211;367, 2007.

[9] M. Figueiredo and R. Nowak. An em algorithm for wavelet-based image restoration. *IEEE Trans. Image Processing*, 12(8):9068211;916, 2003.


[6] "Signal recovery via $\ell$_1 minimization" <http://cnx.org/content/m37179/latest/>

[10] M. Figueiredo, R. Nowak, and S. Wright. Gradient projections for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Select. Top. Signal Processing*, 1(4):5868211;597, 2007.

[11] E. Hale, W. Yin, and Y. Zhang. A fixed-point continuation method for -regularized minimization with applications to compressed sensing. Technical report TR07-07, Rice Univ., CAAM Dept., 2007.

[12] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior point method for large-scale -regularized least squares. *IEEE J. Select. Top. Signal Processing*, 1(4):6068211;617, 2007.

[13] C. De Mol and M. Defrise. A note on wavelet-based inversion algorithms. *Contemporary Math.*, 313:858211;96, 2002.

[14] R. Nowak and M. Figueiredo. Fast wavelet-based image deconvolution using the em algorithm. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 2001.

[15] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. *SIAM J. Multiscale Modeling and Simulation*, 4(2):4608211;489, 2005.

[16] S. Wright, R. Nowak, and M. Figueiredo. Sparse reconstruction by separable approximation. *IEEE Trans. Signal Processing*, 57(7):24798211;2493, 2009.

[17] W. Yin, S. Osher, D. Goldfarb, and J. Darbon. Bregman iterative algorithms for -minimization with applications to compressed sensing. *SIAM J. Imag. Sci.*, 1(1):1438211;168, 2008.