A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization and continuation [Wen,Yin,Goldfarb,Zhang 2009]

Yongjia Song

University of Wisconsin-Madison

April 22, 2010

Introduction

Fundamental principal of Compressive Sensing

K-sparse signal $\bar{x} \in R^n$ can be recovered from relatively few incomplete measurements, $b = A\bar{x}$ by solving a l_0-minimization problem:

$$\min_{x \in R^n} \|x\|_0 \text{ s.t. } Ax = b \tag{1}$$

Basis Pursuit

$$\min_{x \in R^n} \|x\|_1 \text{ s.t. } Ax = b \tag{2}$$

which is proposed by Candes,Romberg,Tao, is more tractable. Under some reasonable conditions on $\bar{x}$ and A, the sparsest solution $\bar{x}$ of problem (1) can be found by solving (2).

Introduction

Greedy algorithm for (1): Orthogonal Matching Pursuit

Solve a sequence of subspace optimization problems of the form:

$$\min_x \|A_T x_T - b\|_2^2 \text{ s.t. } x_i = 0, \forall i \notin T \tag{3}$$

Start from $T = \varnothing, x = 0$, and in each iteration,add to T the index of the largest component of the current gradient of $\|Ax - b\|_2^2$

Regularization for (2)

$$\min_{x \in R^n} \psi_\mu(x) = \mu\|x\|_1 + \|Ax - b\|_2^2 \tag{4}$$

Theory of penalty functions implies that the solution to (4) goes to solution to (2) when μ goes to 0.

Introduction

Challenges for standard LP and QP

- Large-scale real-world applications
- A is usually dense
- real-time processing is required

Special structure

- Measurement matrix A often corresponds to partial transform matrices(eg, discrete Fourier), so fast matrix-vector multiplication are available.
- Sparsity feature of the solutions.

Outline of the algorithm

This paper proposed a two-stage algorithm for the regularized l_1 minimization problem, it combines the good features of both greedy algorithms and convex optimization approach:

- step1 convex optimization
  - Do not require prior information.
  - A first-order method based on shrinkage is applied.
  - obtain an approximation solution and identify a working set T.
- step2 greedy algorithm
  - takes advantage of the sparsity structure.
  - A second-order method is applied
  - solve a smooth subspace problem.
- Embed the two-stage algorithm into a continuation approach by assigning a decreasing sequence of values to μ.

Shrinkage

Iterative Shrinkage Algorithm

For the problem

$$\min_{x \in R^n} \phi(x) := \frac{1}{2}\|Ax - y\|_2^2 + \tau c(x) \tag{5}$$

Iterative Shrinkage algorithm:

$$x^{k+1} = \psi_{\tau/\alpha}(x^k - \frac{1}{\alpha}A^T(Ax^k - y)) \tag{6}$$

where

$$\psi_\lambda(u) := \arg\min_z \frac{1}{2}\|z - u\|_2^2 + \lambda c(z) \tag{7}$$

Features

- Products by A and A^T are efficiently computable.
- Ensure convergence under mild conditions.

Shrinkage

Derivation of Iterative Shrinkage

- Expectation-Maximization [Figueiredo, Nowak, 2001].
- Majorization-Minimization [Daubechies,Defrise,DeMol, 2004].
- Forward-Backward Splitting [Hale, Yin, Zhang, 2007].
- Separable Approximation [Wright,Figueiredo,Nowak,2009].

Benefit of Iterative Shrinkage

Iterative shrinkage yields the support and the signs of the optimal solution x^* of the problem (4) in a finite number of iterations, that is, $\exists \bar{k}$ such that $sgn(x^k) \equiv sgn(x^*)$ for all $k > \bar{k}$.

Use shrinkage in step 1

Notation

- $f(x) = \|Ax - b\|_2^2$
- $g(x) = \nabla f(x)$
- $x \odot y$: component-wise product of x and y

Special shrinkage for $c(x) = \|x\|_1$

$$S(y, \gamma) \triangleq sgn(y) \odot \max\{|y| - \gamma, 0\} \quad (8)$$

which is the unique minimizer of the function

$$\gamma \|x\|_1 + \frac{1}{2}\|x - y\|_2^2 \quad (9)$$

and we use the iteration:

$$x^{k+1} \triangleq S(x^k - \gamma g^k, \mu\lambda), \lambda > 0 \quad (10)$$

Shrinkage

Shrinkage phase of the algorithm

- Select a parameter λ^k and compute a direction $S(x^k - \lambda^k g^k, \mu_k \lambda^k) - x^k$
- Do a line search on the direction d^k,and set the new iteration point $x^{k+1} = x^k + \alpha_k d^k$
- The set of indices corresponding to 0 and nearly 0 components of x^k(say, $x^k \leq \xi^k$) is selected as a working set,which during the step 2, the subspace problem is formed by fixing these components to be 0.

Note

We may not do a subspace optimization problem in every iteration, we only do that if certain conditions hold for the iteration point in shrinkage phase.

Subspace optimization phase

motivation

- The iterative shrinkage scheme essentially reduces to a gradient projection method for solving a subspace minimization problem after sufficiently many iterations.
- A second-order method might be faster than the iterative shrinkage to solve the subspace problem.

subspace problem, definition

- The active set $A(x) := \{i \in \{1, ..., n\} | |x_i| = 0\}$
- The inactive set $I(x) := \{i \in \{1, ..., n\} | |x_i| > 0\}$

When $A(x^k)$ is a good estimate of the true active set, we approximate $sgn(x_i^*)$ by $sgn(x_i^k)$ and replace the original $\phi_\mu(x)$ by the smooth function(suppose the approximation set are A_k, I_k)

$$\varphi_\mu(x) := \mu sgn(x_{I_k}^k)^T x_{I_k} + f(x) \tag{11}$$

Subspace optimization phase

Smooth subspace problem

Now we come up with a nice subspace problem, we can use efficient and fast algorithm in NLP to solve this simply-constrained problem:

$$\min \varphi_\mu(x)$$

$$\text{s.t. } x \in \Omega(x^k)$$

$$\text{where } \Omega(x^k) := \{x \in R^n | sgn(x_i^k)x_i \geq 0, i \in I_k; x_i = 0, i \in A_k\}$$

In the paper, they use a limited-memory quasi-Newton method with simple bound constraints.

Alternating strategy of the two phases

Idea of stopping criteria for shrinkage phase

- We want to start subspace optimization as soon as possible
- we want the active set that defines the subspace optimization problem to be as accurate as possible
- In all, we want to make both phases work efficiently, if either of them does not, switch to the other

Idea of identification of the active set

- The efficiency of the algorithm depends on how fast and how well the active set is identified.
- One approach is to replace the active set $A(x^k)$ and the support $I(x^k)$ by the sets:

$$A(x^k, \xi_k) := \{i \in \{1, ..., n\} \| x_i^k| \leq \xi_k\},$$
$$I(x^k, \xi_k) := \{i \in \{1, ..., n\} \| x_i^k| > \xi_k\}$$

Continuation strategy

Idea

- objective: create a path of solutions that converge to the solution of the original problem
- A fixed fractional reduction of μ_k should be enforced

Summary

- objective: solve a l_1-regularized minimization problem
- approach: two-phases
  - shrinkage phase: do not use prior information, partially use sparsity, take some iterations to obtain an estimate subset of the active set and the support.
  - subspace phase: take advantage of the prior information, reconstruct the problem to be a smooth problem, use fast algorithms.
  - Reasonable strategy to switch between the two phases.