Fast Sparse Recovery via Non-Convex Optimization

Laming Chen and Yuantau Gu

Department of Electronic Engineering
Tsinghua University
Email: chen-lm06@mails.tsinghua.edu.cn, gyt@tsinghua.edu.cn
Homepage: http://gu.ee.tsinghua.edu.cn/

Dec. 16, 2015

Contents

1. Preliminary
2. Algorithm
3. Convergence Results
4. Simulation Results
5. Summary

Contents

1. Preliminary
2. Algorithm
3. Convergence Results
4. Simulation Results
5. Summary

Preliminary

Sparse Recovery Problem

- One tries to find a sparse solution to an underdetermined linear system.
- Using ℓ_1 norm to induce sparsity is a standard technique.
- Some certain non-convex functions tend to outperform ℓ_1 norm empirically in sparse recovery.

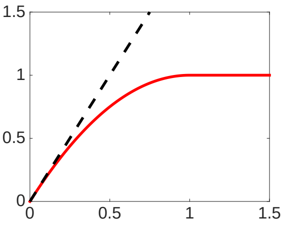
Motivation

- The convergence results of these non-convex algorithms are still very limited.
- We aim to devise a fast algorithm and to provide its convergence results.

Preliminary

Weak Convexity

- The non-convex $F(\cdot)$ becomes convex by adding a quadratic term.
- Let $\rho > 0$ be the smallest quantity such that $F(x) + \rho x^2$ is convex.
- These exists $\alpha > 0$ such that $F(x)/x \to \alpha$ as $x \to 0^+$.



$\frac{\rho}{\alpha}$: non-convexity of $F(\cdot)$

Preliminary

Problem Setup

- Consider the optimization problem

$$\arg\min_{\mathbf{x}} J(\mathbf{x}) + \frac{\tau}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$$

where $J(\mathbf{x}) = \sum_{i=1}^{N} F(x_i)$ is fully separable, and non-convex scalar function $F : \mathbb{R} \to \mathbb{R}^+$ satisfies:

(a) $F(0) = 0$, $F(\cdot)$ is even and not identically zero;
(b) $F(\cdot)$ is non-decreasing on $[0, +\infty)$;
(c) The function $x \mapsto F(x)/x$ is non-increasing on $(0, +\infty)$;
(d) $F(\cdot)$ is weakly convex on $[0, +\infty)$.

- Such $J(\mathbf{x})$ is common in sparse recovery literatures.

Concrete Examples of $F(\cdot)$

Requirements: $0 \leq p < 1$ and $\sigma > 0$

No.	$F(x)$	ρ	α
1.	$\frac{\lvert x\rvert}{(\lvert x\rvert+\sigma)^{1-p}}$	$(1-p)\sigma^{p-2}$	σ^{p-1}
2.	$1-\mathrm{e}^{-\sigma\lvert x\rvert}$	$\sigma^2/2$	σ
3.	$\ln(1+\sigma\lvert x\rvert)$	$\sigma^2/2$	σ
4.	$\mathrm{atan}(\sigma\lvert x\rvert)$	$3\sqrt{3}\sigma^2/16$	σ
5.	$(2\sigma\lvert x\rvert-\sigma^2x^2)\mathbf{1}_{\lvert x\rvert\leq\frac{1}{\sigma}}+\mathbf{1}_{\lvert x\rvert>\frac{1}{\sigma}}$	σ^2	2σ

Contents

1. Preliminary
2. Algorithm
3. Convergence Results
4. Simulation Results
5. Summary

Algorithm

For the optimization problem

$$\arg\min_{\mathbf{x}} J(\mathbf{x}) + \frac{\tau}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$$

Algorithm 1 Proposed Algorithm

Require: $\mathbf{y}$, $\mathbf{A}$, $\tau > 0$, $\delta > 0$

Initialize: $l = 0$, $\mathbf{x}^0 = \mathbf{0}$;
while not converge **do**
  $\nabla^l = \mathbf{x}^l - \delta\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x}^l - \mathbf{y})$;
  $\mathbf{x}^{l+1} = \mathrm{prox}_J(\nabla^l, \delta/\tau)$ [1];
  $l = l + 1$;
end while

[1] $\mathrm{prox}_J(\mathbf{v}, \lambda) = \arg\min_{\mathbf{x}} J(\mathbf{x}) + \frac{1}{2\lambda}\|\mathbf{x} - \mathbf{v}\|_2^2$

Contents

1. Preliminary
2. Algorithm
3. Convergence Results
4. Simulation Results
5. Summary

Convergence Results

Define the objective function

$$G(\mathbf{x}) = J(\mathbf{x}) + \frac{\tau}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$$

Theorem 1. (Convergence to stationary point)

Assume sequence $\{\mathbf{x}^l\}$ is generated by Algorithm 1. If δ satisfies $0 < \delta < \min\{1/\|\mathbf{A}^T\mathbf{A}\|_2, \tau/(2\rho)\}$, then

(a) The sequence $\{G(\mathbf{x}^l)\}$ is non-increasing and convergent;

(b) Any limit point of $\{\mathbf{x}^l\}$ is a stationary point of the problem.

Convergence Results

Theorem 2. (Convergence to sparse signal)

(a) Assume $\|\mathbf{x}^*\|_0 \leq K$, $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$, and $M_0 = \|\mathbf{x}^0 - \mathbf{x}^*\|_2$;

(b) Suppose $\gamma(J, \mathbf{A}, K) < 1$ and the non-convexity $\rho/\alpha \leq c_1/M_0$;

(c) Suppose $0 < \delta < \min\{1/(\|\mathbf{A}^T\mathbf{A}\|_2 + \|\mathbf{y}\|_2^2/\|\mathbf{x}^*\|_2^2), \tau/(2\rho)\}$, and the regularization parameter $\tau = c_2/\|\mathbf{e}\|_2$;

(d) Assume the sequence $\{\mathbf{x}^l\}$ is generated by Algorithm 1, and $\|\mathbf{x}^l - \mathbf{x}^*\|_2 \geq c_3\|\mathbf{e}\|_2$ holds for any $1 \leq l \leq k$.

Then for any $1 \leq l \leq k$,

$$\|\mathbf{x}^l - \mathbf{x}^*\|_2^2 + \frac{2\delta c_2}{\tau c_3}\|\mathbf{x}^l - \mathbf{x}^*\|_2 \leq \|\mathbf{x}^{l-1} - \mathbf{x}^*\|_2^2.$$

Contents

1. Preliminary
2. Algorithm
3. Convergence Results
4. Simulation Results
5. Summary

Simulation Results

General settings

- We choose

$$F(x) = (|x| - \rho x^2)\mathbf{1}_{|x| \leq 1/(2\rho)} + 1/(4\rho)\mathbf{1}_{|x| > 1/(2\rho)}.$$

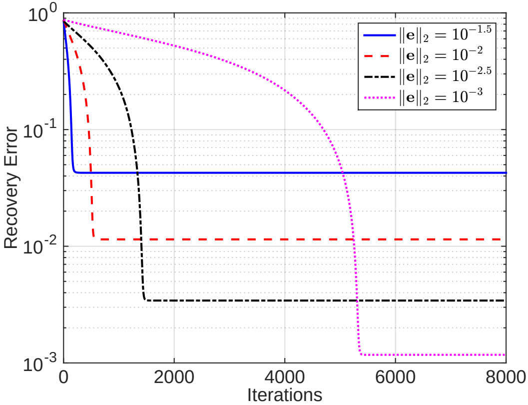
Then it can be calculated that when $\lambda < 1/(2\rho)$,

$$\mathrm{prox}_F(v_i, \lambda) = \frac{v_i - \lambda \mathrm{sign}(v_i)}{1 - 2\lambda\rho}\mathbf{1}_{\lambda < |v_i| \leq 1/(2\rho)} + v_i \mathbf{1}_{|v_i| > 1/(2\rho)}.$$

- We take random partial DCT measurements of a normalized sparse signal.

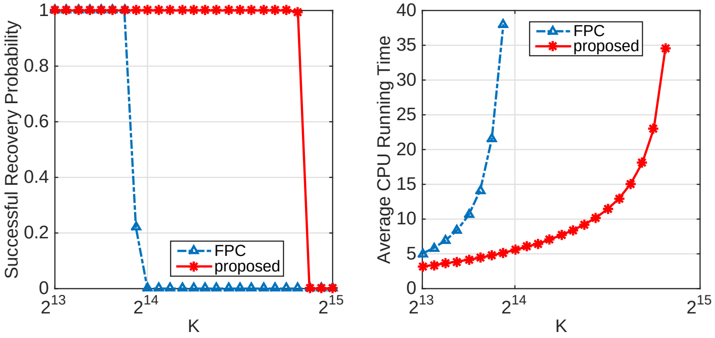
Experiment 1: Rate of Convergence

$$N = 2^{10},\ M = 2^8,\ K = 2^5,\ \rho = 5,\ \delta = 1, \text{ and } \tau = \sqrt{N}/\|\mathbf{e}\|_2$$



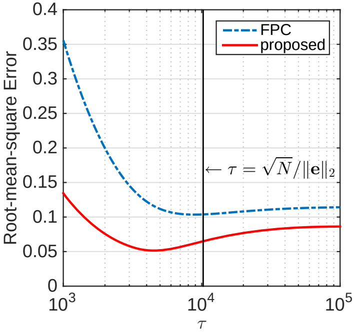
Experiment 2: Comparison in the Noiseless Case

$N = 2^{18}$, $M = 2^{16}$, K varies from 2^{13} to 2^{15}, $\rho = 120$, and $\tau = 10^6$



Experiment 3: Comparison in the Noisy Case

$N = 2^{18}$, $M = 2^{16}$, $K = 2^{13}$, $\rho = 120$, and (left) $\|\mathbf{e}\|_2 = 0.05$; (right) $\tau = \sqrt{N}/\|\mathbf{e}\|_2$



Contents

1 Preliminary

2 Algorithm

3 Convergence Results

4 Simulation Results

5 Summary

Summary

- We propose a fast algorithm for the non-convex function regularized least squares problem.
- We prove that under some conditions, the iterates converge to a neighborhood of the sparse signal with superlinear convergence rate.
- Simulation results verify the theoretical results and show the superiority of the proposed algorithm compared with its convex counterpart.