Robust Sparse Recovery via Non-Convex Optimization

Laming Chen and Yuantao Gu

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

August 23, 2014

Contents

Contents

Introduction

Sparse Recovery Problem

- Sparse recovery problem is expressed as

$$\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_0, \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}$$

  where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a “fat” sensing matrix.
- Computationally intractable

Sparse Recovery Algorithms

- Greedy pursuits (OMP, CoSaMP, SP, etc)
- Convex optimization based algorithms
- Non-convex optimization based algorithms
  - Better performance tends to be derived
  - Still lacking complete convergence analysis

Introduction

Current Convergence Analysis

- For p-IRLS $(0 < p < 1)$, the convergence is guaranteed in a sufficiently small neighborhood of the sparse signal.
- For MM subspace algorithm, it is shown that the generated sequence will converge to a critical point.
- For SL0, the complete convergence analysis is done due to the "local convexity" of the penalties, and the algorithm needs to solve a sequence of optimization problems rather than a single one to guarantee convergence.

Objective

- To provide theoretical convergence guarantees of a non-convex approach for sparse recovery from the initial solution to the global optimal.

Contents

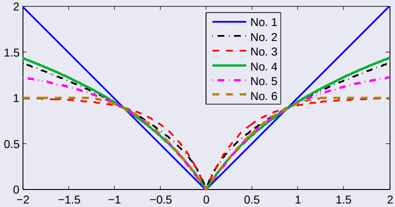
Preliminary

Problem Setup

- Non-convex optimization problem is formulated as

$$\underset{\mathbf{x}}{\text{argmin}} J(\mathbf{x}) = \sum_{i=1}^{N} F(x_i), \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{x}$$

- $F(\cdot)$ belongs to a class of sparseness measures (proposed by Rémi Gribonval et al.)



Preliminary

Null Space Property

- Define $\gamma(J, \mathbf{A}, K)$ as the smallest quantity such that

  $$J(\mathbf{z}_S) \leq \gamma(J, \mathbf{A}, K) J(\mathbf{z}_{S^c})$$

  holds for any set $S \subset \{1, 2, \ldots, N\}$ with $\#S \leq K$ and for any $\mathbf{z} \in \mathcal{N}(\mathbf{A})$.
- Performance analysis with $\gamma(J, \mathbf{A}, K)$ (provided by Rémi Gribonval et al.)
  - If $\gamma(J, \mathbf{A}, K) < 1$, for any K-sparse $\mathbf{x}^*$ and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$, the optimization problem returns $\mathbf{x}^*$;
  - If $\gamma(J, \mathbf{A}, K) > 1$, there exists a K-sparse $\mathbf{x}^*$ and $\mathbf{y} = \mathbf{A}\mathbf{x}^*$ such that the problem returns a signal differs from $\mathbf{x}^*$;
  - $\gamma(\ell_0, \mathbf{A}, K) \leq \gamma(J, \mathbf{A}, K) \leq \gamma(\ell_1, \mathbf{A}, K)$.

Preliminary

Weak Convexity

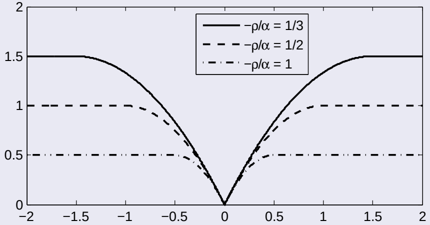
- $F(\cdot)$ is ρ-convex if and only if ρ is the largest quantity such that there exists convex $H(\cdot)$ and $F(t) = H(t) + \rho t^2$.
- $F(\cdot)$ is
  - strongly convex if $\rho > 0$;
  - convex if $\rho = 0$;
  - weakly convex (also known as semi-convex) if $\rho < 0$.
- Generalized gradient (a generalization of subgradient for convex functions) for weakly convex functions can be seen as any element in the set $\partial F(t) := \partial H(t) + 2\rho t$.

Contents

Main Contribution

Sparsity-inducing Penalty

- $F(\cdot)$ is a weakly convex sparseness measure;
- Vast majority of non-convex sparsity-inducing penalties are formed by weakly convex sparseness measures;
- There exists α such that $F(t) \leq \alpha|t|$. Define $-\rho/\alpha$ as the non-convexity of the penalty (a measure of how quickly the generalized gradient of $F(\cdot)$ decreases).



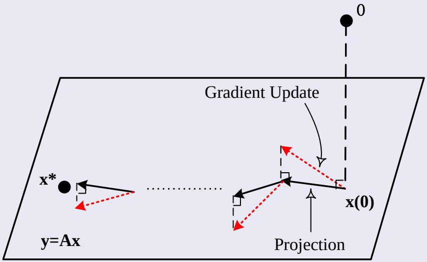
Main Contribution

Projected Generalized Gradient (PGG) Method

- Initialization: Calculate $\mathbf{A}^\dagger$, $\mathbf{x}(0) = \mathbf{A}^\dagger \mathbf{y}$, $n = 0$;
- Repeat:

$$\tilde{\mathbf{x}}(n+1) = \mathbf{x}(n) - \kappa \nabla J(\mathbf{x}(n));$$
$$\mathbf{x}(n+1) = \tilde{\mathbf{x}}(n+1) + \mathbf{A}^\dagger(\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}(n+1));$$
$$n = n+1;$$

- Until: Stopping criterion satisfied;



Main Contribution

Performance of PGG

Theorem 1. Assume $\mathbf{y} = \mathbf{A}\mathbf{x}^* + \mathbf{e}$ and $\|\mathbf{x}^*\|_0 \leq K$. For any tuple $(J, \mathbf{A}, K)$ with $J(\cdot)$ formed by weakly convex sparseness measure $F(\cdot)$ and $\gamma(J, \mathbf{A}, K) < 1$, and for any positive constant M_0, if the non-convexity of $J(\cdot)$ satisfies

$$\frac{-\rho}{\alpha} \leq \frac{1}{M_0} \frac{1 - \gamma(J, \mathbf{A}, K)}{5 + 3\gamma(J, \mathbf{A}, K)},$$

the recovered solution $\hat{\mathbf{x}}$ by PGG satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 \leq \frac{4\alpha^2 N}{C_1}\kappa + 8C_2\|\mathbf{e}\|_2$$

provided that $\|\mathbf{x}(0) - \mathbf{x}^*\|_2 \leq M_0$.

Main Contribution

Approximate PGG (APGG) Method

- To reduce the computational burden of calculating $\mathbf{A}^{\dagger} = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{A}^{\mathrm{T}})^{-1}$, assume we adopt $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ to approximate $\mathbf{A}^{\dagger}$, and let $\zeta = \|\mathbf{I} - \mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{B}\|_2$.
- **Theorem 2.** Under the same assumptions as Theorem 1, if the non-convexity of $J(\cdot)$ satisfies

$$\frac{-\rho}{\alpha} \leq \frac{1}{M_0} \frac{1-\gamma(J,\mathbf{A},K)}{5+3\gamma(J,\mathbf{A},K)},$$

and $\zeta < 1$, the recovered solution $\hat{\mathbf{x}}$ by APGG satisfies

$$\|\hat{\mathbf{x}} - \mathbf{x}^*\|_2 \leq 2C_3(\zeta)\kappa + 2C_4(\zeta)\|\mathbf{e}\|_2$$

provided that $\|\mathbf{x}(0) - \mathbf{x}^*\|_2 \leq M_0$.

Contents

Simulation

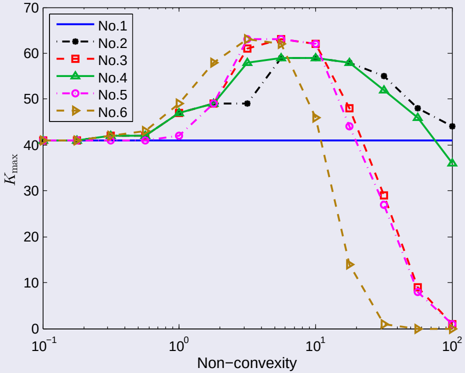
Parameter Setting

- Matrix $\mathbf{A}$: 200×1000; independent and identically distributed Gaussian entries;
- Vector $\mathbf{x}^*$: nonzero entries independently satisfy Gaussian distribution; normalized to have unit ℓ_2 norm;
- The approximate $\mathbf{A}^\dagger$ is calculated using an iterative method (introduced by Ben-Israel et al.);

Simulation

First Experiment

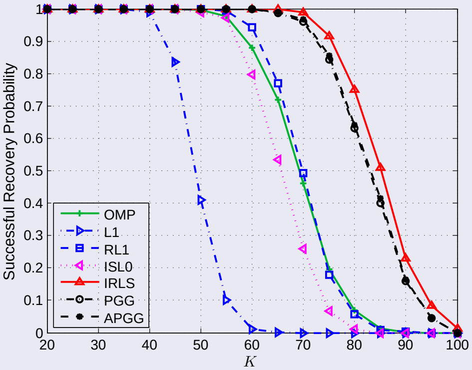
- Recovery performance of the PGG method is tested in the noiseless scenario with different sparsity-inducing penalties and different choices of non-convexity.



Simulation

Second Experiment

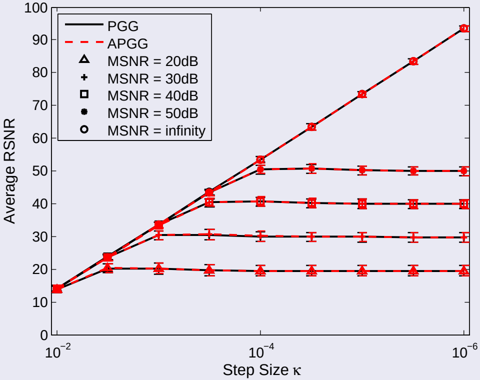
- The recovery performance of the (A)PGG method is compared in the noiseless scenario with some typical sparse recovery algorithms.



Simulation

Third Experiment

- The recovery precisions of the (A)PGG method are simulated under different settings of step size and measurement noise.



Contents

Summary

Conclusion

- A class of weakly convex sparseness measures is adopted to constitute the sparsity-inducing penalties;
- The convergence analysis of the (A)PGG method reveals that when the non-convexity is below a threshold, the recovery error is linear in both the step size and the noise term;
- As for the APGG method, the influence of the approximate projection is reflected in the coefficients instead of an additional error term.

Related Works

[1] Laming Chen and Yuantao Gu, The Convergence Guarantees of a Non-convex Approach for Sparse Recovery, *IEEE Transactions on Signal Processing*, 62(15):3754-3767, 2014.
[2] Laming Chen and Yuantao Gu, The Convergence Guarantees of a Non-convex Approach for Sparse Recovery Using Regularized Least Squares, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 3374-3378, May 4-9, 2014, Florence, Italy.
[3] Laming Chen and Yuantao Gu, Robust Recovery of Low-Rank Matrices via Non-Convex Optimization, *International Conference on Digital Signal Processing (DSP)*, 355-360, Aug. 20-23, 2014, Hong Kong, China.

http://gu.ee.tsinghua.edu.cn/

Thanks!

Supplementary

Weakly Convex Sparseness Measures with Parameter ρ

No.	$F(t)$	ρ
1.	$\|t\|$	0
2.	$\frac{\|t\|}{(\|t\|+\sigma)^{1-p}}$	$(p-1)\sigma^{p-2}$
3.	$1-\mathrm{e}^{-\sigma\|t\|}$	$-\sigma^2/2$
4.	$\ln(1+\sigma\|t\|)$	$-\sigma^2/2$
5.	$\mathrm{atan}(\sigma\|t\|)$	$-3\sqrt{3}\sigma^2/16$
6.	$(2\sigma\|t\|-\sigma^2t^2)\mathcal{X}_{\|t\|\leq\frac{1}{\sigma}}+\mathcal{X}_{\|t\|>\frac{1}{\sigma}}$	$-\sigma^2$