Robust Recovery of Low-Rank Matrices 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 22, 2014

Contents

1. Introduction
2. Preliminary
3. Main Contribution
4. Simulation
5. Summary

Contents

1 Introduction

2 Preliminary

3 Main Contribution

4 Simulation

5 Summary

Introduction

Linearly Constrained Rank Minimization Problem

- The problem is expressed as

$$\underset{\mathbf{X}}{\operatorname{argmin}} \ \operatorname{rank}(\mathbf{X}), \ \text{subject to } \mathbf{y} = \mathcal{A}(\mathbf{X})$$

  where $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^m$ is a given linear operator.
- Computationally intractable

Linearly Constrained Rank Minimization Algorithms

- Rank constrained LS problem (AM, ADMiRA, SVP, etc)
- Linearly constrained convex optimization
- Linearly constrained non-convex optimization
  - Better performance tends to be derived
  - Still lacking complete convergence analysis

Introduction

Current Convergence Analysis

- For ℓ_p-Proximal-Gradient algorithm, it is proved that any cluster point of the solution sequence is a generalized fixed point of a majorizer of the penalty, and it may not be a global minimizer.
- The Smooth Rank Function algorithm considers solving a sequence of non-convex optimization problems, and it is only shown that their global optimums converge to the low-rank matrix.

Objective

- To provide theoretical convergence guarantees of a non-convex approach for linearly constrained non-convex optimization from the initialization to the low-rank matrix.

Contents

1 Introduction

2 Preliminary

3 Main Contribution

4 Simulation

5 Summary

Preliminary

Notation

- For $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$, let $\sigma_1(\mathbf{X}) \geq \cdots \geq \sigma_n(\mathbf{X})$ be its singular values where $n = \min\{n_1, n_2\}$.

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

1. Introduction
2. Preliminary
3. Main Contribution
4. Simulation
5. Summary

Main Contribution

Low-Rank-Inducing Penalty

- The low-rank-inducing penalty is defined as

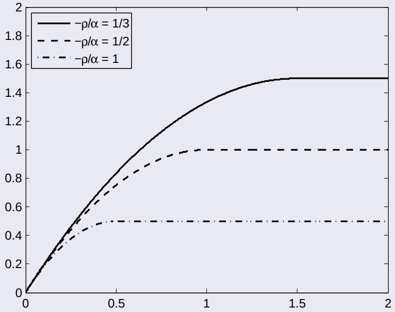
$$J(\mathbf{X}) = \sum_{i=1}^{n} F(\sigma_i(\mathbf{X}))$$

- The function $F : [0, +\infty) \to \mathbb{R}$ satisfies
  - $F(0) = 0$ and $F(\cdot)$ is not identically zero;
  - $F(\cdot)$ is non-decreasing;
  - $F(\cdot)$ is concave;
  - $F(\cdot)$ is weakly convex.
- Vast majority of low-rank-inducing penalties satisfy these conditions.

Main Contribution

Non-Convexity of the Penalty

- 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

Generalized Gradient of the Penalty

- If the reduced SVD of $\mathbf{X}$ is $\mathbf{X} = \mathbf{U}\mathrm{diag}(\sigma_1, \cdots, \sigma_n)\mathbf{V}^{\mathrm{T}}$, the generalized gradients of $J(\mathbf{X})$ with respect to $\mathbf{X}$ belong to

$$\partial J(\mathbf{X}) = \{\mathbf{U}\mathrm{diag}(f(\sigma_1), \cdots, f(\sigma_n))\mathbf{V}^{\mathrm{T}}\},$$

where $f(\sigma_i) \in \partial F(\sigma_i)$ for $i = 1, \cdots, n$.

Projected Generalized Gradient for Low Rank Recovery

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

$$\mathbf{X}(\ell+1) = \mathbf{X}(0) + \mathcal{P}_{\mathcal{N}(\mathcal{A})}(\mathbf{X}(\ell) - \kappa \nabla J(\mathbf{X}(\ell)));$$
$$\ell = \ell + 1;$$

- Until: Stopping criterion satisfied;

Main Contribution

Null Space Constant

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

$$\sum_{i=1}^{k} F(\sigma_i(\mathbf{Z})) \leq \gamma(J, \mathcal{A}, k) \sum_{i=k+1}^{n} F(\sigma_i(\mathbf{Z}))$$

  holds for all $\mathbf{Z} \in \mathcal{N}(\mathcal{A})$.
- The definition is a generalization of the null space constant for the Schatten-p quasi norm with $0 < p < 1$.
- The null space constant characterizes how spreading out the singular values of elements in the null space are.

Main Contribution

Performance of PGG-LRR

Theorem 1. Assume $\mathbf{y} = \mathcal{A}(\mathbf{X}_0) + \mathbf{e}$ and $\text{rank}(\mathbf{X}_0) \leq k$. For any tuple $(J, \mathcal{A}, k)$ with $J(\cdot)$ formed by $F(\cdot)$ satisfying the aforementioned definition and with $\gamma(J, \mathcal{A}, 2k) < 1$, and for any positive constant M_0, if the non-convexity of $J(\cdot)$ satisfies

$$\frac{-\rho}{\alpha} \leq \frac{1}{M_0} \frac{1 - \gamma(J, \mathcal{A}, 2k)}{5 + 3\gamma(J, \mathcal{A}, 2k)},$$

the recovered solution $\hat{\mathbf{X}}$ by PGG-LRR satisfies

$$\|\hat{\mathbf{X}} - \mathbf{X}_0\|_F \leq \frac{2\alpha^2 n}{C_1}\kappa + 2C_2\|\mathbf{e}\|_2$$

provided that $\|\mathbf{X}(0) - \mathbf{X}_0\|_F \leq M_0$.

Contents

1 Introduction

2 Preliminary

3 Main Contribution

4 Simulation

5 Summary

Simulation

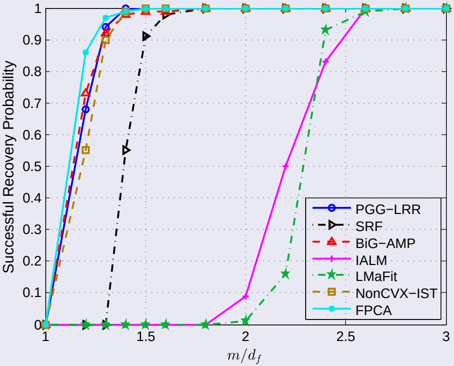
Parameter Setting

- Matrix $\mathbf{X}_0$: 100×100; of the form $\mathbf{X}_0 = \mathbf{L}\mathbf{R}^{\mathrm{T}}$, where the entries of $\mathbf{L} \in \mathbb{R}^{100\times k}$ and $\mathbf{R} \in \mathbb{R}^{100\times k}$ are i.i.d. standard normal; normalized to have unit Frobenius norm;
- $d_f = k(n_1 + n_2 - k)$: the degrees of freedom of $\mathbf{X}_0$;
- Modes of measurements
  - For random sampling, the locations of the samples are uniformly chosen among all possible choices;
  - For random projection, each measurement is the inner product of $\mathbf{X}_0$ and a matrix of the same size whose entries are i.i.d. Gaussian with zero mean and variance $1/m$;

Simulation

First Experiment

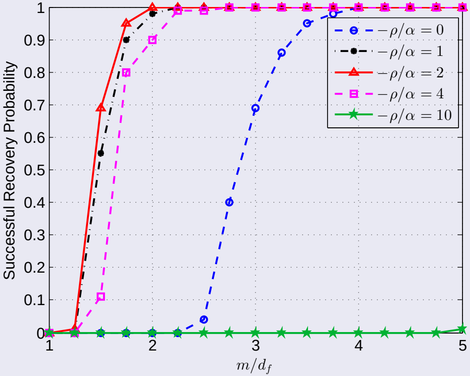
- The recovery performance of PGG-LRR is compared in the noiseless scenario of random sampling with some matrix completion algorithms.



Simulation

Second Experiment

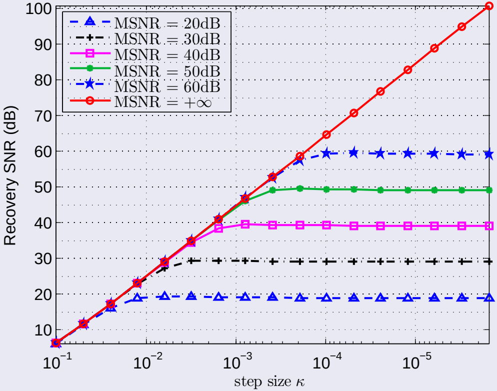
- The recovery performance of PGG-LRR is tested in the noiseless scenario of random sampling with different choices of non-convexity.



Simulation

Third Experiment

- The recovery precisions of PGG-LRR are demonstrated in the noisy scenario of random projection with different step size κ and measurement SNR.



Contents

1 Introduction

2 Preliminary

3 Main Contribution

4 Simulation

5 Summary

Summary

Conclusion

- A class of low-rank-inducing penalties with characterization of their non-convexity is adopted as the relaxation of the rank function;
- The convergence guarantees of PGG-LRR are provided in the noisy scenario from the initialization to the low-rank matrix;
- Numerical experiments verify the theoretical results in this paper, and the performance of the proposed non-convex approach is among the best.

Thanks!

Supplementary

Examples of $F(\cdot)$ 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.	$(2\sigma t-\sigma^2 t^2)\mathcal{X}_{t\leq\frac{1}{\sigma}}+\mathcal{X}_{t>\frac{1}{\sigma}}$	$-\sigma^2$