Generalized Singular Value Thresholding

Canyi Lu[1], Changbo Zhu[1], Chunyan Xu[2], Shuicheng Yan[1], Zhouchen Lin[3]

[1]National University of Singapore, [2]Huazhong University of Science and Technology, [3]Peking University

Nonconvex Nonsmooth Low-rank Minimization Problem

This paper aims to solve the following nonconvex nonsmooth problem

$$\min_{\mathbf{X}\in\mathbb{R}^{m\times n}} F(\mathbf{X}) = \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + h(\mathbf{X}), \tag{1}$$

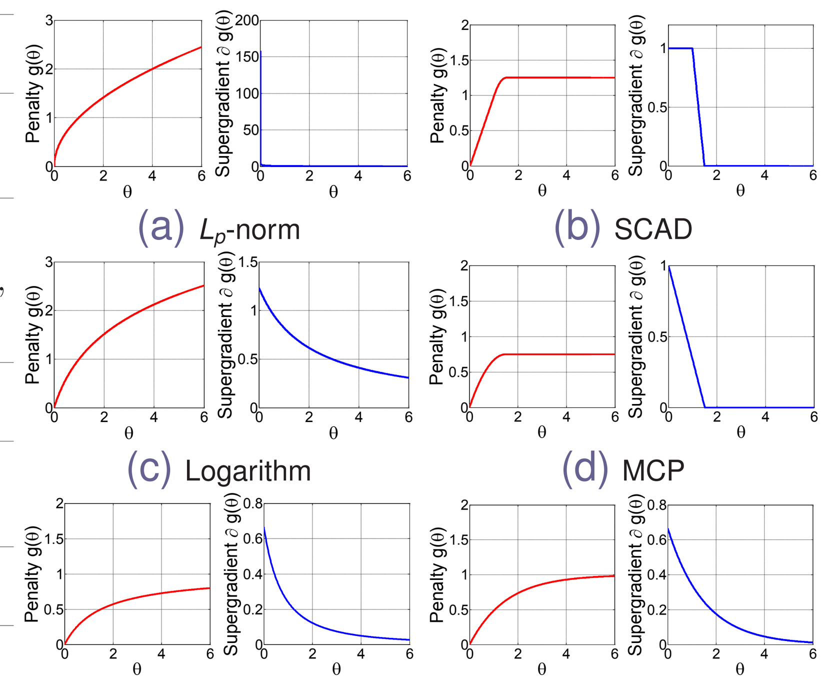
where $\sigma_i(\mathbf{X})$'s denote the singular values of $\mathbf{X}\in\mathbb{R}^{m\times n}$ (assume $m\leq n$), and

- h: $\mathbb{R}^{m\times n}\to\mathbb{R}^{+}$ is a smooth function with Lipschitz continuous gradient, i.e.,

  $$\|\nabla h(\mathbf{X})-\nabla h(\mathbf{Y})\|_F \leq L(h)\|\mathbf{X}-\mathbf{Y}\|_F,\ \forall\, \mathbf{X},\mathbf{Y}\in\mathbb{R}^{m\times n},$$

  where $L(h)>0$ is called the Lipschitz constant of ∇h. h is possibly nonconvex.
  Examples: squared loss $\frac{1}{2}\|\mathbf{A}\mathbf{X}-\mathbf{b}\|^2$ and logistic loss.
- $g:\mathbb{R}^{+}\to\mathbb{R}^{+}$ is continuous, concave and monotonically increasing on $[0,\infty)$. g is possibly nonsmooth.
  Examples: the nonconvex surrogate functions of L_0-norm.

Penalty	Formula $g_\lambda(\theta)$, $\theta\geq 0$, $\lambda>0$	Supergradient $\partial g_\lambda(\theta)$
L_p-norm	$\lambda\theta^p$	$\begin{cases}\infty, & \text{if } \theta=0,\\ \lambda p\theta^{p-1}, & \text{if } \theta>0.\end{cases}$
SCAD	$\begin{cases}\lambda\theta, & \text{if } \theta\leq\lambda,\\ \frac{-\theta^2+2\gamma\lambda\theta-\lambda^2}{2(\gamma-1)}, & \text{if } \lambda<\theta\leq\gamma\lambda,\\ \frac{\lambda^2(\gamma+1)}{2}, & \text{if } \theta>\gamma\lambda.\end{cases}$	$\begin{cases}\lambda, & \text{if } \theta\leq\lambda,\\ \frac{\gamma\lambda-\theta}{\gamma-1}, & \text{if } \lambda<\theta\leq\gamma\lambda,\\ 0, & \text{if } \theta>\gamma\lambda.\end{cases}$
Logarithm	$\frac{\lambda}{\log(\gamma+1)}\log(\gamma\theta+1)$	$\frac{\gamma\lambda}{(\gamma\theta+1)\log(\gamma+1)}$
MCP	$\begin{cases}\lambda\theta-\frac{\theta^2}{2\gamma}, & \text{if } \theta<\gamma\lambda,\\ \frac{1}{2}\gamma\lambda^2, & \text{if } \theta\geq\gamma\lambda.\end{cases}$	$\begin{cases}\lambda-\frac{\theta}{\gamma}, & \text{if } \theta<\gamma\lambda,\\ 0, & \text{if } \theta\geq\gamma\lambda.\end{cases}$
Geman	$\frac{\lambda\theta}{\theta+\gamma}$	$\frac{\lambda\gamma}{(\theta+\gamma)^2}$
Laplace	$\lambda(1-\exp(-\frac{\theta}{\gamma}))$	$\frac{\lambda}{\gamma}\exp(-\frac{\theta}{\gamma})$



(a) L_p-norm (b) SCAD (c) Logarithm (d) MCP (e) Geman (f) Laplace

Generalized Proximal Gradient (GPG) Algorithm for (1)

- Since $\nabla h(\mathbf{X})$ is Lipschitz continuous, we have

$$h(\mathbf{X}) \leq h(\mathbf{X}^k) + \langle\nabla h(\mathbf{X}^k), \mathbf{X}-\mathbf{X}^k\rangle + \frac{\mu}{2}\|\mathbf{X}-\mathbf{X}^k\|_F^2,\quad \forall \mu\geq L(h). \tag{2}$$

- Update $\mathbf{X}$ by

$$\begin{aligned}\mathbf{X}^{k+1} &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + \langle\nabla h(\mathbf{X}^k), \mathbf{X}-\mathbf{X}^k\rangle + \frac{\mu}{2}\|\mathbf{X}-\mathbf{X}^k\|_F^2\\ &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + \frac{\mu}{2}\left\|\mathbf{X}-\left(\mathbf{X}^k-\frac{1}{\mu}\nabla h(\mathbf{X}^k)\right)\right\|_F^2,\\ &= \arg\min_{\mathbf{X}} \mathbf{Prox}^{\sigma}_{\frac{1}{\mu}g}\left(\mathbf{X}^k-\frac{1}{\mu}\nabla h(\mathbf{X}^k)\right)\end{aligned} \tag{3}$$

Theorem 1 *If $\mu>L(h)$, the sequence $\{\mathbf{X}^k\}$ generated by (3) satisfies the following properties:*

(1) *$F(\mathbf{X}^k)$ is monotonically decreasing.*

(2) $\lim_{k\to+\infty}(\mathbf{X}^k-\mathbf{X}^{k+1})=\mathbf{0}$;

(3) *If $F(\mathbf{X})\to+\infty$ when $\|\mathbf{X}\|_F\to+\infty$, then any limit point of $\{\mathbf{X}^k\}$ is a stationary point.*

Generalized Singular Value Thresholding (GSVT)

Solving (3) requires computing the GSVT operator associated with g, i.e.,

$$\mathbf{Prox}^{\sigma}_{g}(\mathbf{B}) = \arg\min_{\mathbf{X}} \sum_{i=1}^{m} g(\sigma_i(\mathbf{X})) + \frac{1}{2}\|\mathbf{X}-\mathbf{B}\|_F^2. \tag{4}$$

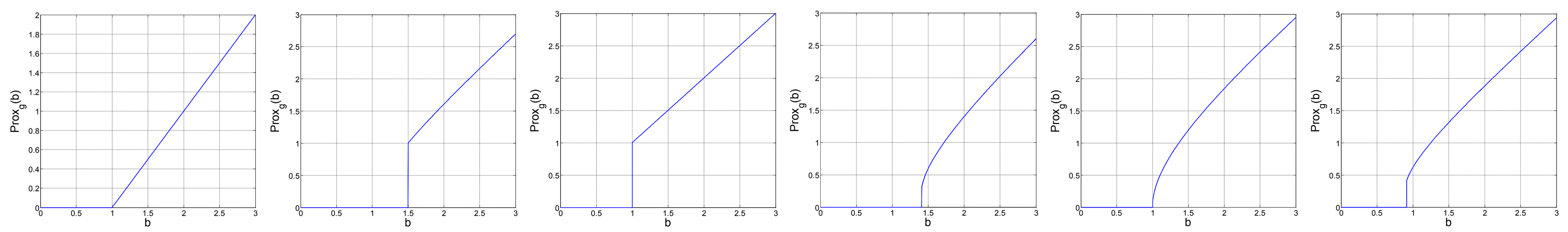
Theorem 2 *Let $g:\mathbb{R}^{+}\to\mathbb{R}^{+}$ be a function such that $\mathbf{Prox}_g(\cdot)$ is monotone. Let $\mathbf{B}=\mathbf{U}\,\mathrm{Diag}(\boldsymbol{\sigma}(\mathbf{B}))\mathbf{V}^T$ be the SVD of $\mathbf{B}\in\mathbb{R}^{m\times n}$. Then an optimal solution to (4) is*

$$\mathbf{X}^* = \mathbf{U}\,\mathrm{Diag}(\boldsymbol{\varrho}^*)\mathbf{V}^T, \tag{5}$$

where $\boldsymbol{\varrho}^*$ satisfies $\varrho_1^*\geq\varrho_2^*\geq\cdots\geq\varrho_m^*$, $i=1,\cdots,m$, and

$$\varrho_i^* \in \mathbf{Prox}_g(\sigma_i(\mathbf{B})) = \operatorname*{argmin}_{\varrho_i\geq 0} g(\varrho_i) + \frac{1}{2}(\varrho_i-\sigma_i(\mathbf{B}))^2. \tag{6}$$

Theorem 3 *For any lower bounded function g, its proximal operator $\mathbf{Prox}_g(\cdot)$ is monotone, i.e., for any $p_i^*\in\mathbf{Prox}_g(x_i)$, $i=1,2$, $p_1^*\geq p_2^*$, when $x_1>x_2$.*



(a) L_1-norm (b) L_p-norm (c) MCP (d) Logarithm (e) Laplace (f) Geman

Figure: Plots of b v.s. $\mathbf{Prox}_g(b)$ for different choices of g. $\mathbf{Prox}_g(b)$ is monotone.

Computing $\mathbf{Prox}^{\sigma}_{g}(\cdot)$ in (4) is equivalent to computing $\mathbf{Prox}_g(\cdot)$ in (6).

Proximal Operator of Nonconvex Function

If g satisfies some assumptions, then it is easy to compute its proximal operator, i.e.,

$$\mathbf{Prox}_g(b) = \arg\min_{x\geq 0} f_b(x) = g(x) + \frac{1}{2}(x-b)^2. \tag{7}$$

Assumption 1 $g:\mathbb{R}^{+}\to\mathbb{R}^{+}$, $g(0)=0$. g is concave, nondecreasing and differentiable. The gradient ∇g is convex.

Theorem 4 *Given g satisfying **Assumption** 1. Denote $\hat{x}^b=\max\{x|\nabla f_b(x)=0, 0\leq x\leq b\}$ and $x^*=\arg\min_{x\in\{0,\hat{x}^b\}} f_b(x)$. Then x^* is optimal to (7).*

Algorithm 1: A general solver to (7) in which g satisfying **Assumption 1**

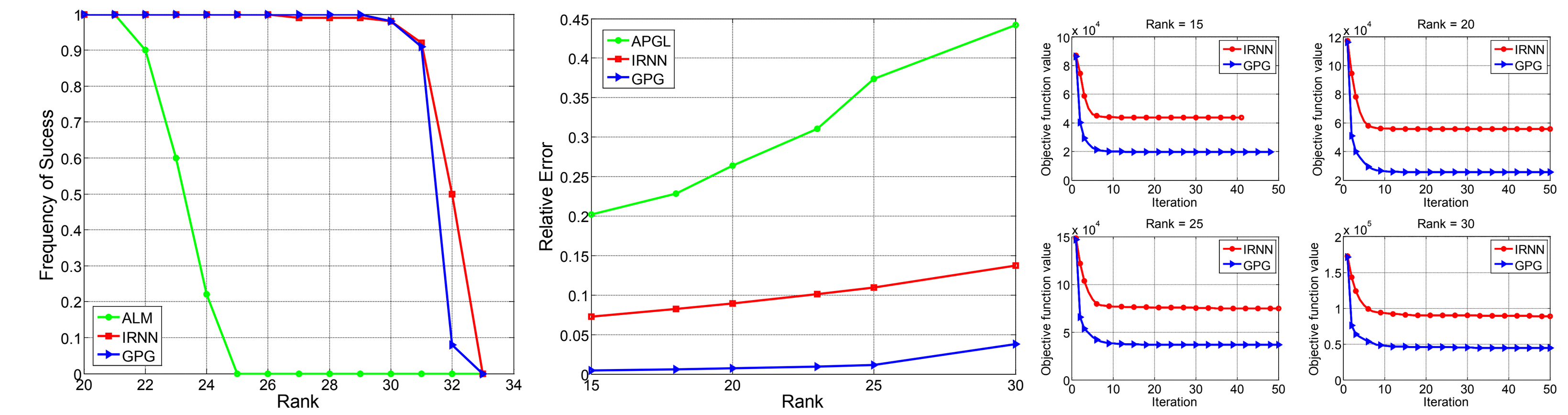
Input: $b\geq 0$.
Output: Identify an optimal solution, 0 or $\hat{x}^b=\max\{x|\nabla f_b(x)=0, 0\leq x\leq b\}$.
if $\nabla g(b)=0$ **then**
  return $\hat{x}^b=b$;
else
  // find $\hat{x}^b$ by fixed point iteration.
  $x_0=b$. // Initialization.
  while not converge **do**
    $x_{k+1}=b-\nabla g(x_k)$;
    if $x_{k+1}<0$ **then**
      return $\hat{x}^b=0$;
      break;
    end
  end
end
Compare $f_b(0)$ and $f_b(\hat{x}^b)$ to identify the optimal one.

Experiment: Low-rank Matrix Completion on Random Data

Test on the following problem with different nonconvex surrogate functions

$$\min_{\mathbf{X}} \sum_{i=1}^{m} g_\lambda(\sigma_i(\mathbf{X})) + \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{X}-\mathbf{M})\|_F^2, \tag{8}$$

where Ω is the index set, and $\mathcal{P}_\Omega:\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}$ is a linear operator that keeps the entries in Ω unchanged and those outside Ω zeros.
Compared methods: Iteratively Reweighted Nuclear Norm (IRNN) [1].

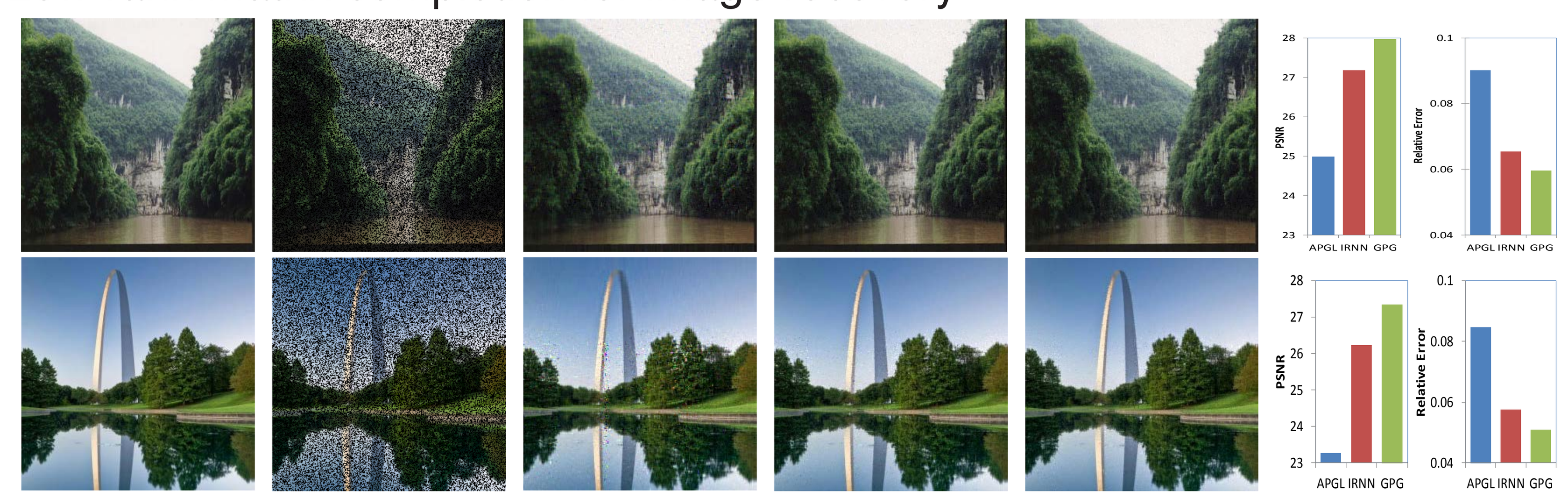


(a) (b) (c)

Figure: Results of low rank matrix recovery on random data. (a) Frequency of Success for a noise free case. (b) Relative error for a noisy case. (c) Convergence curves of algorithms.

Experiment: Low-rank Matrix Completion on Real Data

- Low-rank matrix completion for image recovery



(a) Original (b) Noisy (c) APGL (d) IRNN (e) GPG (f) PSNR & error

Figure: Image inpainting by APGL, IRNN, and GPG.

- Low-rank matrix completion for collaborative filtering
  - To predict the unknown preference of a user on a set of unrated items.
  - Test on the MovieLens data set which includes three problems: moive-100K, moive-1M and moive-10M.
  - Normalized Mean Absolute Error (NMAE), i.e., $\|\mathcal{P}_\Omega(\mathbf{X}^*)-\mathcal{P}_\Omega(\mathbf{M})\|_1/|\Omega|$.

Table: Comparison of NMAE of APGL, IRNN and GPG for collaborative filtering.

Problem	size of $\mathbf{M}$: (m,n)	APGL	IRNN	GPG
moive-100K	(943, 1682)	2.76e-3	2.60e-3	**2.53e-3**
moive-1M	(6040, 3706)	2.66e-1	2.52e-1	**2.47e-1**
moive-10M	(71567, 10677)	3.13e-1	3.01e-1	**2.89e-1**

[1] Canyi Lu, et al. Generalized Nonconvex Nonsmooth Low-Rank Minimization, CVPR, 2014.

Department of Electrical and Computer Engineering, National University of Singapore. Email: canyilu@nus.edu.sg

Matlab code available at our homepage: https://sites.google.com/site/canyilu/