Generalized Nonconvex Nonsmooth Low-Rank Minimization

Canyi Lu[1], Jinhui Tang[2], Shuicheng Yan[1], Zhouchen Lin[3]

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

Motivation

- The ℓ_1-norm is a convex surrogate of ℓ_0-norm, but it may be too loose.
- A better choice is to use nonconvex surrogate functions, e.g., ℓ_p-norm ($0 < p < 1$).
- The nuclear norm is a convex surrogate of the rank function, but it may be too loose.
- A better choice is to apply the nonconvex surrogate functions of ℓ_0-norm on the singular values.
- However, the nonconvex low-rank minimization is much more challenging than the nonconvex sparse minimization.
- We propose a general solver for nonconvex nonsmooth low-rank minimization.

Many Nonconvex Surrogate Functions of the ℓ_0-norm

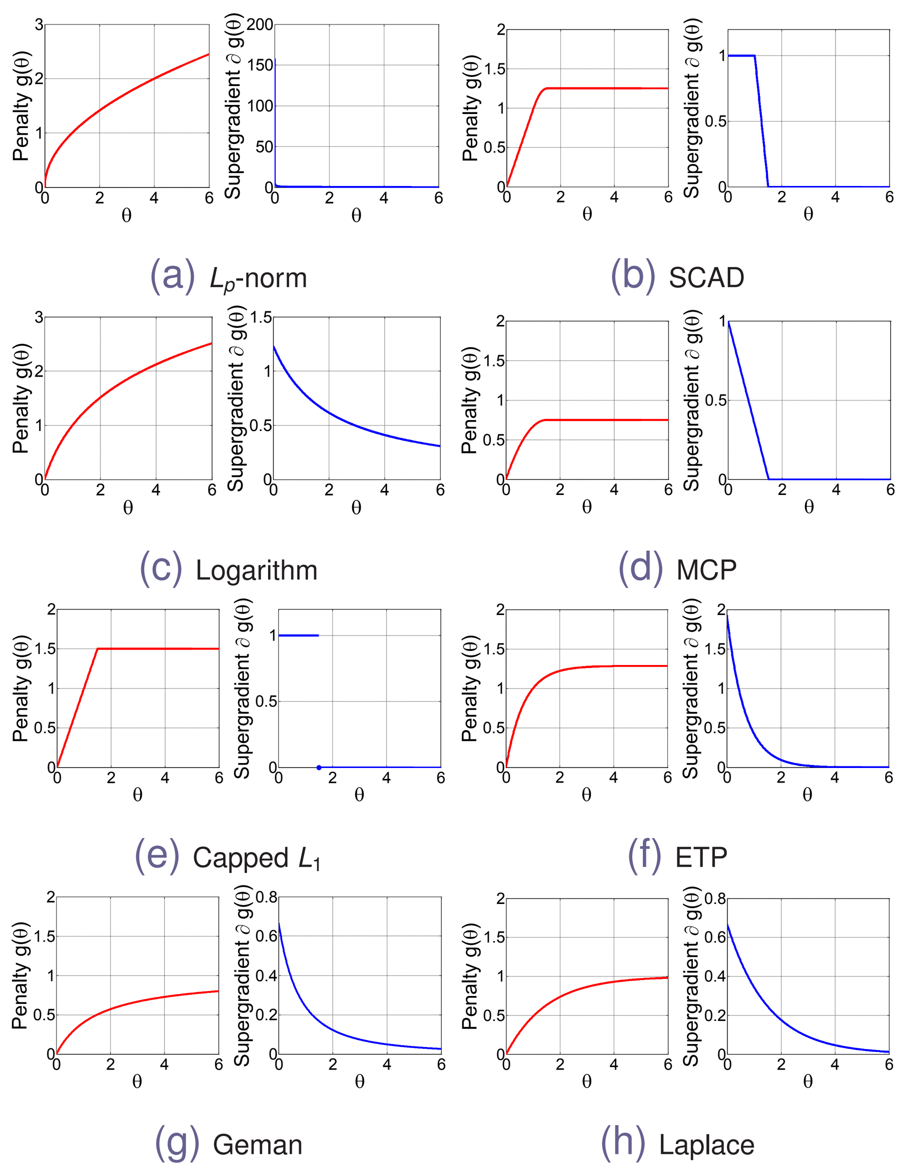
Table: Popular nonconvex surrogate functions of $||\theta||_0$ and their supergradients.

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	$\begin{cases} \lambda\theta - \frac{\theta^2}{2\gamma}, & \text{if } \theta < \gamma\lambda, \\ \frac{1}{2}\gamma\lambda^2, & \text{if } \theta \geq \gamma\lambda. \end{cases}$	$\begin{cases} \lambda - \frac{\theta}{\gamma}, & \text{if } \theta < \gamma\lambda, \\ 0, & \text{if } \theta \geq \gamma\lambda. \end{cases}$
Capped L_1	$\begin{cases} \lambda\theta, & \text{if } \theta < \gamma, \\ \lambda\gamma, & \text{if } \theta \geq \gamma. \end{cases}$	$\begin{cases} \lambda, & \text{if } \theta < \gamma, \\ [0,\lambda], & \text{if } \theta = \gamma, \\ 0, & \text{if } \theta > \gamma. \end{cases}$
ETP	$\frac{\lambda}{1-\exp(-\gamma)}(1-\exp(-\gamma\theta))$	$\frac{\lambda\gamma}{1-\exp(-\gamma)}\exp(-\gamma\theta)$
Geman	$\frac{\lambda\theta}{\theta+\gamma}$	$\frac{\lambda\gamma}{(\theta+\gamma)^2}$
Laplace	$\lambda(1-\exp(-\frac{\theta}{\gamma}))$	$\frac{\lambda}{\gamma}\exp(-\frac{\theta}{\gamma})$

The same properties:

- All the above nonconvex surrogate functions are concave and monotonically increasing on $[0, \infty)$.
- Their supergradients are nonnegative and monotonically decreasing.

Figure: Illustration of the popular nonconvex surrogate functions of $||\theta||_0$ (left), and their supergradients (right).



Supergradient of a Concave Function g

Definition *Let $g : \mathbb{R}^n \to \mathbb{R}$ be concave. A vector $\mathbf{v}$ is a supergradient of g at the point $\mathbf{x} \in \mathbb{R}^n$ if for every $\mathbf{y} \in \mathbb{R}^n$, the following inequality holds*

$$g(\mathbf{x}) + \langle \mathbf{v}, \mathbf{y} - \mathbf{x} \rangle \geq g(\mathbf{y}).$$

All supergradients of g at $\mathbf{x}$ are called superdifferential, denoted as $\partial g(\mathbf{x})$.

Key Lemma *The superdifferential of a concave function g is an antimonotone operator, i.e.,*

$$\langle \mathbf{u} - \mathbf{v}, \mathbf{x} - \mathbf{y} \rangle \leq 0, \qquad (1)$$

for any $\mathbf{u} \in \partial g(\mathbf{x})$, $\mathbf{v} \in \partial g(\mathbf{y})$.

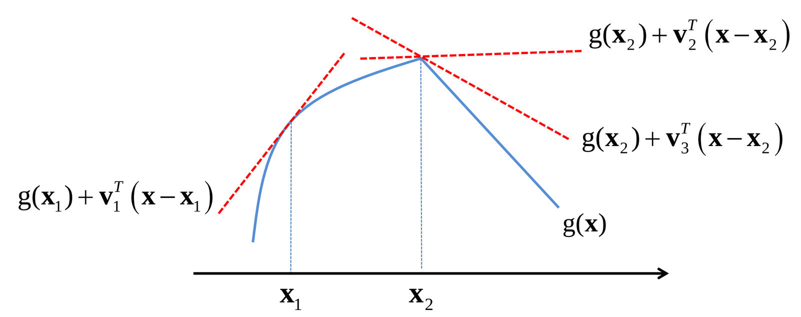


Figure: Supergraidients of a concave function. $\mathbf{v}_1$ is a supergradient at $\mathbf{x}_1$, and $\mathbf{v}_2$ and $\mathbf{v}_3$ are supergradients at $\mathbf{x}_2$.

Nonconvex Nonsmooth Low-Rank Minimization

This work aims to solve the following nonconvex nonsmooth low-rank minimization problem

$$\min_{\mathbf{X} \in \mathbb{R}^{m \times n}} F(\mathbf{X}) = \sum_{i=1}^{m} g_\lambda(\sigma_i(\mathbf{X})) + f(\mathbf{X}), \qquad (2)$$

where $\sigma_i(\mathbf{X})$ denotes the i-th singular value of $\mathbf{X} \in \mathbb{R}^{m \times n}$ (assume $m \leq n$).
The penalty function g_λ and loss function f satisfy the following assumptions:

- $g_\lambda : \mathbb{R} \to \mathbb{R}^+$ is continuous, concave and monotonically increasing on $[0, \infty)$. g_λ is possibly nonsmooth.
  Examples: all the known nonconvex surrogate functions of ℓ_0-norm.
- f: $\mathbb{R}^{m \times n} \to \mathbb{R}^+$ is a smooth function with Lipschitz continuous gradient, i.e.,
  $$||\nabla f(\mathbf{X}) - \nabla f(\mathbf{Y})||_F \leq L(f)||\mathbf{X} - \mathbf{Y}||_F, \ \forall\, \mathbf{X}, \mathbf{Y} \in \mathbb{R}^{m \times n},$$
  where $L(f) > 0$ is called the Lipschitz constant of ∇f. f is possibly nonconvex.
  Examples: squared loss $||\mathcal{A}(\mathbf{X}) - \mathbf{b}||^2$ and logistic loss.
- $F(\mathbf{X}) \to \infty$ iff $||\mathbf{X}||_F \to \infty$.

Iteratively Reweighted Nuclear Norm (IRNN): A General Solver to (2)

Method: when updating $\mathbf{X}^{k+1}$, linearize g_λ and f at $\mathbf{X}^k$, simultaneously.
Motivation: two key inequalities: (3) and (5).

- Since g_λ is concave on $[0, \infty)$, by the definition of the supergradient, we have

$$g_\lambda(\sigma_i(\mathbf{X})) \leq g_\lambda(\sigma_i(\mathbf{X}^k)) + w_i^k(\sigma_i(\mathbf{X}) - \sigma_i(\mathbf{X}^k)), \qquad (3)$$

where

$$w_i^k \in \partial g_\lambda(\sigma_i(\mathbf{X}^k)). \qquad (4)$$

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

$$f(\mathbf{X}) \leq f(\mathbf{X}^k) + \langle \nabla f(\mathbf{X}^k), \mathbf{X} - \mathbf{X}^k \rangle + \frac{\mu}{2}||\mathbf{X} - \mathbf{X}^k||_F^2, \quad \forall \mu \geq L(f). \qquad (5)$$

Combining (3) and (5), we update $\mathbf{X}$ by

$$\begin{aligned} \mathbf{X}^{k+1} &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} w_i^k \sigma_i(\mathbf{X}) + \langle \nabla f(\mathbf{X}^k), \mathbf{X} - \mathbf{X}^k \rangle + \frac{\mu}{2}||\mathbf{X} - \mathbf{X}^k||_F^2 \\ &= \arg\min_{\mathbf{X}} \sum_{i=1}^{m} w_i^k \sigma_i(\mathbf{X}) + \frac{\mu}{2}\left\|\mathbf{X} - \left(\mathbf{X}^k - \frac{1}{\mu}\nabla f(\mathbf{X}^k)\right)\right\|_F^2. \end{aligned} \qquad (6)$$

Since $\sigma_1(\mathbf{X}^k) \geq \sigma_2(\mathbf{X}^k) \geq \cdots \geq \sigma_m(\mathbf{X}^k) \geq 0$, by the antimonotone property of supergradient (1), we have

$$0 \leq w_1^k \leq w_2^k \leq \cdots \leq w_m^k.$$

This key property guarantees that problem (6) has a closed form solution by Weighted Singular Value Thresholding.
Iteratively Reweighted Nuclear Norm (IRNN) algorithm: Alternately updating w by (4) and $\mathbf{X}$ by (6) to solve (2).

Convergence Analysis of IRNN

Theorem *The sequence $\{\mathbf{X}^k\}$ generated by IRNN satisfies the following properties:*

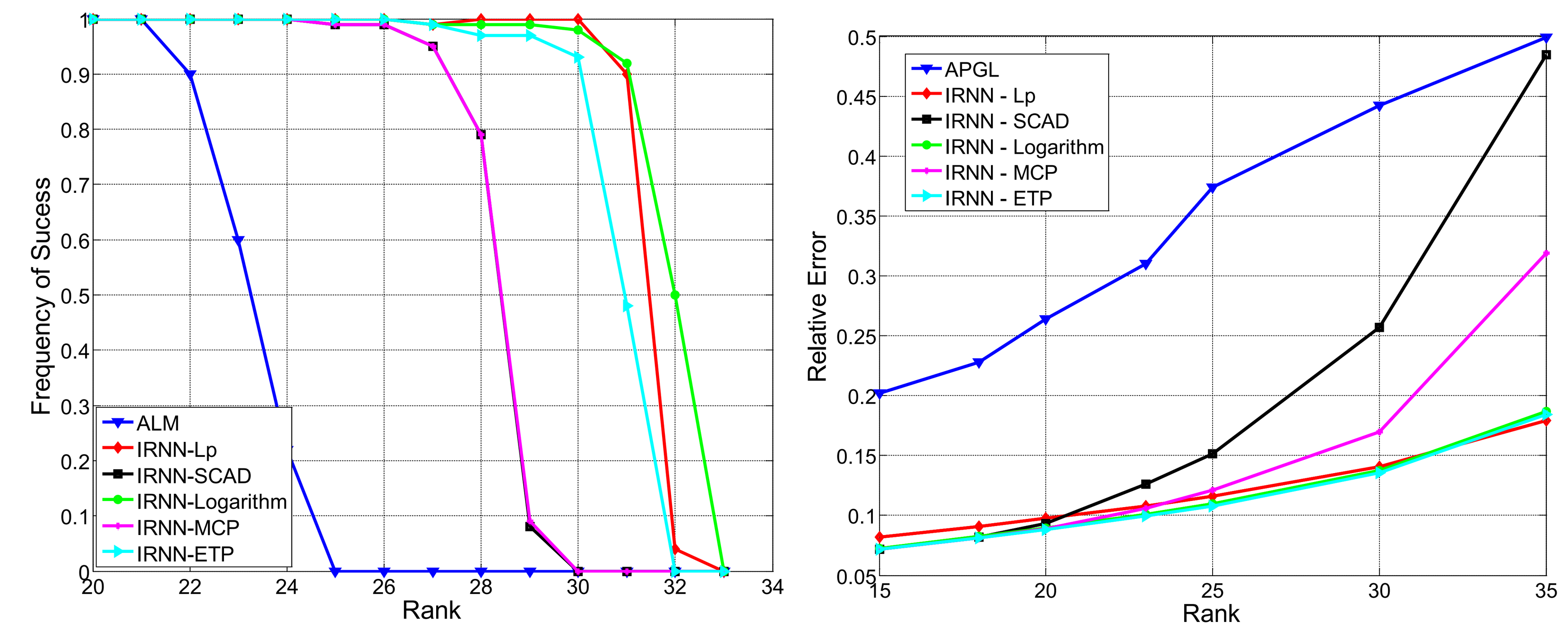
- *$F(\mathbf{X}^k)$ is monotonically decreasing. Indeed, $F(\mathbf{X}^k) - F(\mathbf{X}^{k+1}) \geq \frac{\mu - L(f)}{2}||\mathbf{X}^k - \mathbf{X}^{k+1}||_F^2 \geq 0$;*
- $\lim_{k \to \infty} (\mathbf{X}^k - \mathbf{X}^{k+1}) = \mathbf{0}$;
- *The sequence $\{\mathbf{X}^k\}$ is bounded;*
- *Any accumulation point of $\{\mathbf{X}^k\}$ is a stationary point to problem (2).*

Experiment: Low-rank Matrix Completion on Random Data

Test on the following problem with different nonconvex surrogate functions

$$\min_{\mathbf{X}} \sum_{i=1}^{m} g_\lambda(\sigma_i(\mathbf{X})) + \frac{1}{2}||\mathcal{P}_\Omega(\mathbf{X} - \mathbf{M})||_F^2, \qquad (7)$$

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.



(a) random data without noise (b) random data with noise

Figure: Low-rank matrix completion on random data.

Experiment: Low-rank Matrix Completion for Image Recovery

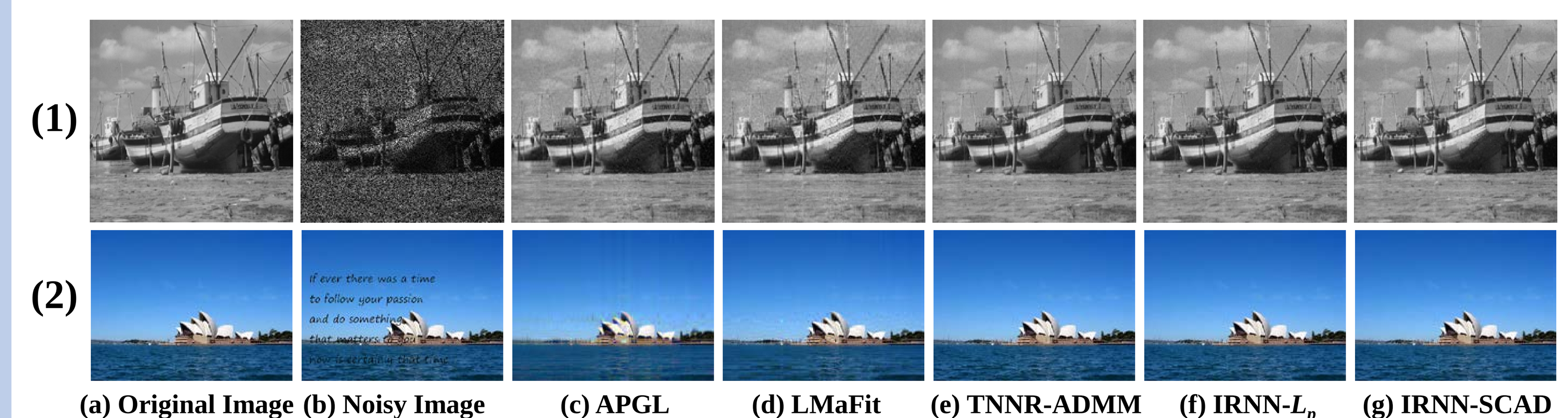


Figure: (a) Original image. (b) Noisy image. (c)-(g) Recovered images by APGL, LMaFit, TNNR-ADMM, IRNN-L_p, and IRNN-SCAD, respectively.

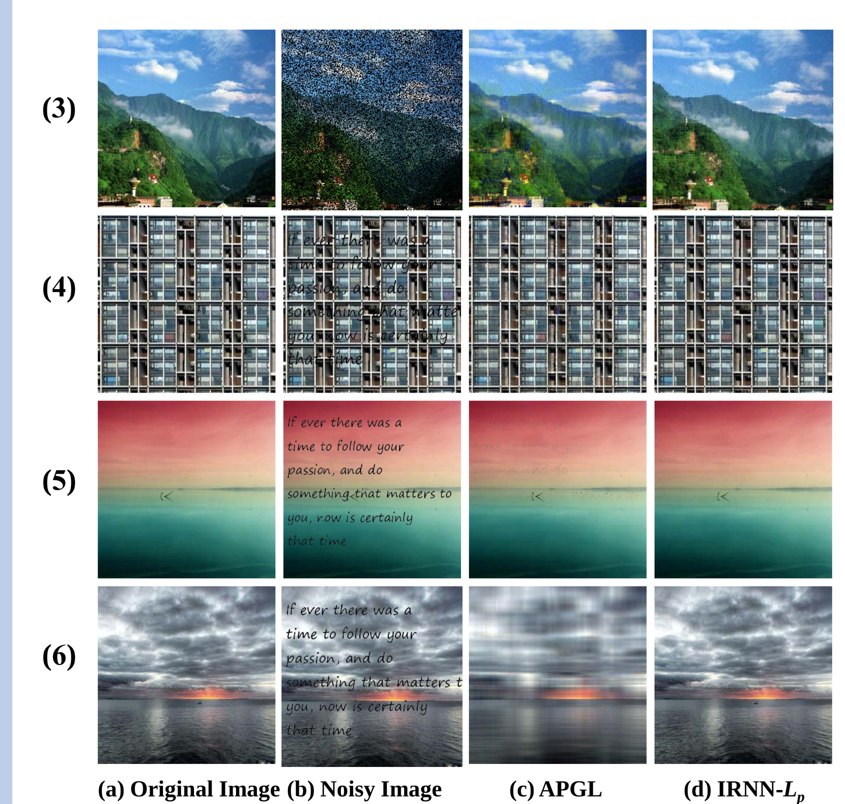


Figure: Image recovery on more images.

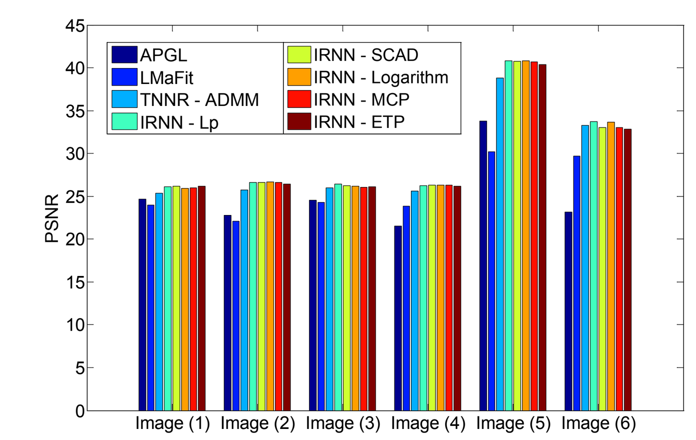


Figure: Comparison of the PSNR values by different matrix completion algorithms.

Take Home Message

- Use nonconvex surrogate rank function instead of nuclear norm;
- Iteratively Reweighted Nuclear Norm (IRNN), the first general solver to (2), was proposed.

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/