

A Proximal Splitting Algorithm for Linearly Involved Generalized Minimax Concave Penalty Models

**Isao Yamada
Tokyo Institute of Technology**

Based on a joint work with Jiro Abe & Masao Yamagishi

**2nd Workshop on Operator Splitting Methods in Data Analysis
@ Center for Computational Mathematics,
Flatiron Institute, NY**

March 20, 2019, New York

Thanks for invitation to this dreaming place !



One of my most favorite buildings
on the globe !

It was 30 years ago when I came
here for the 1st time.

I was so impressed with this
beautiful shape that I took
many photos from
the very low angle lying
on the street.

Thanks for invitation to this dreaming place !



I believe you also like this !

Why ?

Thanks for invitation to this dreaming place !



I believe you also like this !

Why ?

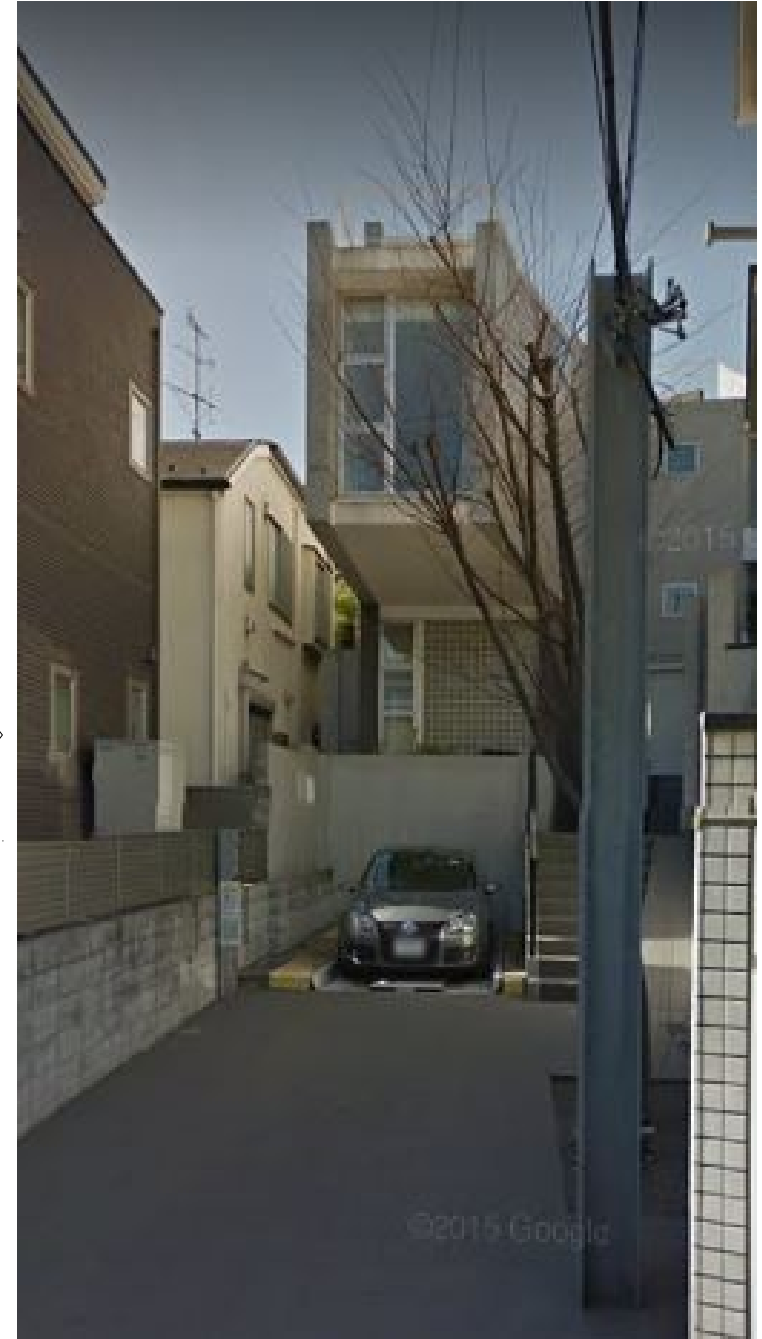
Because we are very strange
persons who like
optimization
subject to severe constraints !

Very severe constraints found in



New York

Tokyo



Many problems in sparsity-aware signal processing and data analytics have been translated into

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \ J_{\Psi \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi \circ L(\mathbf{x}), \ \mu > 0,$$

where $\mathbf{y} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{l \times n}$ and

$\Psi : \mathbb{R}^l \rightarrow \mathbb{R}_+$ is a certain approximation of $\|\cdot\|_0$.

Popular Examples of Ψ :

$$\|(z_1, z_2, \dots, z_l)\|_p := \left(\sum_{k=1}^l |z_k|^p \right)^{1/p}, \quad 0 < p \leq 1.$$

ℓ_0 pseudo-norm

➡ $\|\cdot\|_1$ has been a standard choice due to convexity

..., LASSO [Tibishirani '96], TV [Rudin, Osher, Fatemi'92], ...

Convexity is certainly a key for global optimization

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi \circ L(\mathbf{x}), \quad \mu > 0,$$

But

Convex relaxation of $\|\cdot\|_0$ would not be an only realistic compromise for convexity of $J_{\Psi \circ L}$

For better sparsity,

How about designing nonconvex Ψ
to ensure the overall convexity of $J_{\Psi \circ L}$?

... [Blake, Zisserman'87], [Nikolova'98], [Nikolova'11] ...

[Möllenhoff, Strekalovskiy, Moeller, Cremers'15]...

[Bayram'16], [Lanza, Morigi, Sgallari'16] ... [Selesnick'17] ...

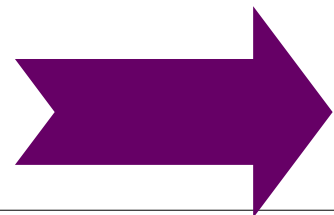
Generalized Minimax Concave (GMC) Penalty [Selesnick '17]

$$(B \in \mathbb{R}^{m \times l}) \Psi_B : \mathbb{R}^l \rightarrow \mathbb{R}_+ : \mathbf{z} \mapsto \|\mathbf{z}\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\mathbf{z} - \mathbf{v})\|^2 \right]$$

is a parametric penalty function satisfying

Nonconvex but includes $\|\cdot\|_1$ as special case

$$\Psi_B(\mathbf{z}) = \|\mathbf{z}\|_1 - \frac{1}{2} \|B\mathbf{z}\|^2 \Leftrightarrow \|B^\top B\mathbf{z}\|_\infty \leq 1,$$



$$\Psi_B = \begin{cases} \|\cdot\|_1 & \text{if } B = O_{m \times l}, \\ \text{nonconvex} & \text{otherwise.} \end{cases}$$

+

Overall convexity for $L = \text{Id}$

For $(A, B, \mu, \mathbf{y}) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \times \mathbb{R}_{++} \times \mathbb{R}^m$,

$$A^\top A - \mu B^\top B \succeq O_n \Rightarrow \frac{1}{2} \|\mathbf{y} - A(\cdot)\|^2 + \mu \Psi_B(\cdot) \in \Gamma_0(\mathbb{R}^n)$$

Generalized Minimax-Concave (GMC) Penalty

I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol.65, 2017.

$$(B \in \mathbb{R}^{m \times l}) \quad \Psi_B(\mathbf{z}) := \|\mathbf{z}\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\mathbf{z} - \mathbf{v})\|^2 \right]$$



Parametrized Multidimensional Extension



Minimax-Concave (MC) Penalty [Zhang '10]

$$\psi : \mathbb{R} \rightarrow \mathbb{R} : z \mapsto \psi(z) := \begin{cases} |z| - \frac{1}{2} z^2, & \text{if } |z| \leq 1 \\ \frac{1}{2}, & \text{otherwise} \end{cases}$$

C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol.38, 2010.

Generalized Minimax-Concave (GMC) Penalty

I. Selesnick, "Sparse regularization via convex analysis,"
IEEE Trans. Signal Process., vol.65, 2017.

$$(B \in \mathbb{R}^{m \times l}) \quad \Psi_B(\mathbf{z}) := \|\mathbf{z}\|_1 - \underbrace{\min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\mathbf{z} - \mathbf{v})\|^2 \right]}_{\text{Generalized Huber function}}$$

Generalized Huber function

$$(\forall \mathbf{z} \in \mathbb{R}^l) \quad \Psi_0(\mathbf{z}) = \|\mathbf{z}\|_1$$

Convex Penalty

I like this beautiful idea of GMC penalty very much because it has a great potential for dealing with many nice nonconvex penalties under single umbrella of the modern convex analysis.



We are interested in global optimization of

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \quad \mu > 0, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{l \times n}$, $B \in \mathbb{R}^{m \times l}$, and

$$\Psi_B(\cdot) := \|\cdot\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\cdot - \mathbf{v})\|^2 \right],$$

under as much general overall-convexity cond for B as possible !

Q1. What is general cond for B to ensure overall-convexity of (1) ?

[Selesnick '17]

$$A^\top A - \mu B^\top B \succeq O_n \Rightarrow \frac{1}{2} \|\mathbf{y} - A(\cdot)\|^2 + \mu \Psi_B \circ \text{Id}(\cdot) \in \Gamma_0(\mathbb{R}^n)$$

But

$$? \longleftrightarrow \frac{1}{2} \|\mathbf{y} - A(\cdot)\|^2 + \mu \Psi_B \circ L(\cdot) \in \Gamma_0(\mathbb{R}^n)$$

does not seem to have been clarified yet.

We are interested in global optimization of

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \ J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \ \mu > 0, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $L \in \mathbb{R}^{l \times n}$, $B \in \mathbb{R}^{m \times l}$, and

$$\Psi_B(\cdot) := \|\cdot\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\cdot - \mathbf{v})\|^2 \right],$$

under as much general overall-convexity cond for B as possible !

Q2. Can we establish any iterative algorithm of guaranteed convergence to globally optimal solution of (1) under general overall-convexity condition ?

[Selesnick '17]

A forward-backward splitting type algorithm is found but only for a special case: $L = \text{Id}$ and $B^T B = (\theta/\mu) A^T A$, $0 \leq \theta \leq 1$



Any algorithm applicable to fully general cases does not seem to have been reported yet.

We are interested in global optimization of

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \quad \mu > 0, \quad (1)$$

$$\text{where} \quad \Psi_B(\cdot) := \|\cdot\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\cdot - \mathbf{v})\|^2 \right],$$

Q1. What is general cond for B to ensure overall-convexity of (1) ?

Q2. Can we establish any iterative algorithm of guaranteed convergence to globally optimal solution of (1) under general overall-convexity condition ?

Q3. Can we choose B flexibly to ensure overall-convexity of (1) ?

These questions would be keys for **further evolutions** as well as **broader applications** of the GMC penalties !

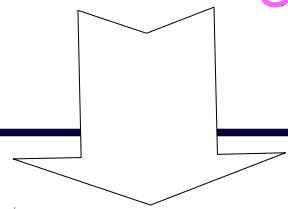
Straightforward limited applications of certain linearly involved GMC penalties have been reported very recently in [Du, Liu'18], [Zhong, Yi, Xiao, Zhang, Wu '18] without any theoretical consideration on these key questions.

We are interested in global optimization of

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \quad \mu > 0, \quad (1)$$

$$\text{where} \quad \Psi_B(\cdot) := \|\cdot\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\cdot - \mathbf{v})\|^2 \right],$$

Q1. What is general cond for B to ensure overall-convexity of (1) ?



[Ayy-ICASSP'19, Proposition 1]

$$A^\top A - \mu L^\top B^\top B L \succeq O_n$$



$$\frac{1}{2} \|\mathbf{y} - A(\cdot)\|^2 + \mu \Psi_B \circ L(\cdot) \in \Gamma_0(\mathbb{R}^n)$$

[Ayy-ICASSP'19] J.Abe, M.Yamagishi, I. Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," *IEEE ICASSP 2019, May, 2019*.

We are interested in global optimization of

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \quad \mu > 0, \quad (1)$$

$$\text{where} \quad \Psi_B(\cdot) := \|\cdot\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\cdot - \mathbf{v})\|^2 \right],$$

Q2. Can we establish any iterative algorithm of guaranteed convergence to globally optimal solution of (1) under general overall-convexity condition ?

[Ayy-ICASSP'19, Theorem 1]

Although Ψ_B is nonsmooth and nonconvex,

we can express the set of all globally optimal solutions in terms of the fixed-point set of computable nonexpansive operator in a certain Hilbert space and therefore can solve (1).

[Ayy-ICASSP'19] J.Abe, M.Yamagishi, I. Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," *IEEE ICASSP 2019, May, 2019*.

Theorem 1(ICASSP'19) Assume $1/2\|\mathbf{y} - A(\cdot)\|^2 + \mu\Psi_B \circ L(\cdot) \in \Gamma_0(\mathbb{R}^n)$ and define $T_{\text{LCP}} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \rightarrow \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l : (\mathbf{x}, \mathbf{v}, \mathbf{w}) \mapsto (\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\eta})$ by

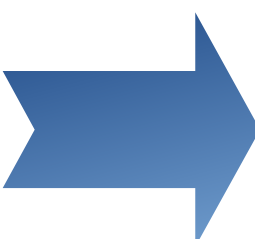
$$\boldsymbol{\xi} := \left[\text{Id} - \frac{1}{\sigma} (A^\top A - \mu L^\top B^\top B L) \right] \mathbf{x} - \frac{\mu}{\sigma} L^\top B^\top B \mathbf{v} - \frac{\mu}{\sigma} L^\top \mathbf{w} + \frac{1}{\sigma} A^\top \mathbf{y}$$

$$\boldsymbol{\zeta} := \text{Soft}_{\frac{\mu}{\tau}} \left[\frac{2\mu}{\tau} B^\top B L \boldsymbol{\xi} - \frac{\mu}{\tau} B^\top B L \mathbf{x} + \left(\text{Id} - \frac{\mu}{\tau} B^\top B \right) \mathbf{v} \right]$$

$$\boldsymbol{\eta} := P_{[-1,1]^l} (2L \boldsymbol{\xi} - L \mathbf{x} + \mathbf{w}),$$

where $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ is chosen to satisfy

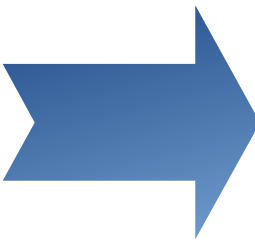
$$\begin{cases} \sigma \text{Id} - \frac{\mu^2}{\tau} L^\top (B^\top B)^2 L - \mu L^\top L \succ O_n \\ \sigma \text{Id} - \frac{\kappa}{2} A^\top A - \mu L^\top L \succ O_n \\ \tau \geq \left(\frac{\kappa}{2} + \frac{2}{\kappa} \right) \mu \|B^\top B\|_2. \end{cases}$$



$$\operatorname{argmin} J_{\Psi_B \circ L} = \mathcal{Q}(\text{Fix}(T_{\text{LCP}})),$$

where $\mathcal{Q} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \rightarrow \mathbb{R}^n : (\mathbf{x}, \mathbf{v}, \mathbf{w}) \mapsto \mathbf{x}$

and $\text{Fix}(T_{\text{LCP}}) := \{(\mathbf{x}, \mathbf{v}, \mathbf{w}) \mid T_{\text{LCP}}(\mathbf{x}, \mathbf{v}, \mathbf{w}) = (\mathbf{x}, \mathbf{v}, \mathbf{w})\}$



$$\mathcal{P} := \begin{bmatrix} \sigma \text{Id} & -\mu L^\top B^\top B & -\mu L^\top \\ -\mu^\top B^\top B L & \tau \text{Id} & O_l \\ -\mu L & O_l & \mu \text{Id} \end{bmatrix} \succ O_{n+2l} \text{ and}$$

$T_{\text{LCP}} : \mathcal{H} (:= \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l) \rightarrow \mathcal{H}$ is $\frac{\kappa}{2\kappa-1}$ -averaged nonexpansive in the Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{P}}, \|\cdot\|_{\mathcal{P}})$, i.e.,

$$(\forall \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{H})$$

$$\|T_{\text{LCP}}(\mathbf{z}_1) - T_{\text{LCP}}(\mathbf{z}_2)\|_{\mathcal{P}} \leq \|\mathbf{z}_1 - \mathbf{z}_2\|_{\mathcal{P}} - \frac{\kappa - 1}{\kappa} \|(\text{Id} - T_{\text{LCP}})(\mathbf{z}_1) - (\text{Id} - T_{\text{LCP}})(\mathbf{z}_2)\|_{\mathcal{P}}$$



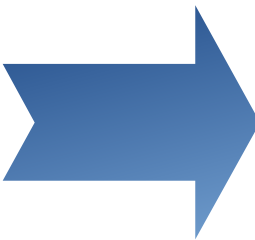
For any initial point $(\mathbf{x}_0, \mathbf{v}_0, \mathbf{w}_0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$, the sequence $(\mathbf{x}_n, \mathbf{v}_n, \mathbf{w}_n)_{n \in \mathbb{N}} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$ generated by

[Krasnoselskii-Mann]

$$(\mathbf{x}_{n+1}, \mathbf{v}_{n+1}, \mathbf{w}_{n+1}) := T_{\text{LCP}}(\mathbf{x}_n, \mathbf{v}_n, \mathbf{w}_n)$$

converges to a point $(\mathbf{x}^*, \mathbf{v}^*, \mathbf{w}^*) \in \text{Fix}(T_{\text{LCP}})$ and

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}^* \in \text{argmin } J_{\Psi_{B \circ L}}$$



$$\mathcal{P} := \begin{bmatrix} \sigma \text{Id} & -\mu L^\top B^\top B & -\mu L^\top \\ -\mu^\top B^\top B L & \tau \text{Id} & O_l \\ -\mu L & O_l & \mu \text{Id} \end{bmatrix} \succ O_{n+2l} \text{ and}$$

$T_{\text{LCP}} : \mathcal{H} \left(:= \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \right) \rightarrow \mathcal{H}$ is $\frac{\kappa}{2\kappa-1}$ -averaged nonexpansive in the Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{P}}, \| \cdot \|_{\mathcal{P}})$, i.e.,

$$(\forall \mathbf{z}_1, \mathbf{z}_2 \in \mathcal{H})$$

$$\|T_{\text{LCP}}(\mathbf{z}_1) - T_{\text{LCP}}(\mathbf{z}_2)\|_{\mathcal{P}} \leq \|\mathbf{z}_1 - \mathbf{z}_2\|_{\mathcal{P}} - \frac{\kappa - 1}{\kappa} \|(\text{Id} - T_{\text{LCP}})(\mathbf{z}_1) - (\text{Id} - T_{\text{LCP}})(\mathbf{z}_2)\|_{\mathcal{P}}$$



For any initial point $(\mathbf{x}_0, \mathbf{v}_0, \mathbf{w}_0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$, the sequence $(\mathbf{x}_n, \mathbf{v}_n, \mathbf{w}_n)_{n \in \mathbb{N}} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$ generated by

[Krasnoselskii-Mann]

$$(\mathbf{x}_{n+1}, \mathbf{v}_{n+1}, \mathbf{w}_{n+1}) := T_{\text{LCP}}(\mathbf{x}_n, \mathbf{v}_n, \mathbf{w}_n)$$

Further optimization is possible if you use **HSDM!**

converges to a point $(\mathbf{x}^*, \mathbf{v}^*, \mathbf{w}^*) \in \text{Fix}(T_{\text{LCP}})$ and

$$\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}^* \in \argmin J_{\Psi_B \circ L}$$



The proof sketch of [AYY-ICASSP'19, Theorem 1]

STEP1

Translate

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \ J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \ \mu > 0,$$

into **a monotone inclusion** by using

Fact (Regular/Fréchet Subdifferential [e.g., Rockafellar-Wets'09])

Let $f \in \Gamma_0(\mathbb{R}^n)$. Assume $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is continuously differentiable in a neighborhood of $\bar{\mathbf{x}} \in \text{dom}(f)$ and $f + g \in \Gamma_0(\mathbb{R}^n)$.

$$\Rightarrow \partial(f + g)(\bar{\mathbf{x}}) = \hat{\partial}(f + g)(\bar{\mathbf{x}}) = \hat{\partial}f(\bar{\mathbf{x}}) + \nabla g(\bar{\mathbf{x}}) = \partial f(\bar{\mathbf{x}}) + \nabla g(\bar{\mathbf{x}}),$$

$$\text{where } \hat{\partial}\varphi(\bar{\mathbf{x}}) := \left\{ \mathbf{u} \in \mathbb{R}^n \mid \liminf_{\mathbf{x} \rightarrow \bar{\mathbf{x}}} \frac{\varphi(\mathbf{x}) - \varphi(\bar{\mathbf{x}}) - \langle \mathbf{u}, \mathbf{x} - \bar{\mathbf{x}} \rangle}{\|\mathbf{x} - \bar{\mathbf{x}}\|} \geq 0 \right\}.$$

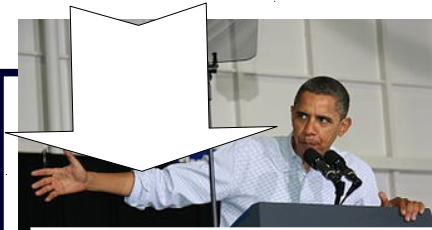
STEP2 Follow [Condat'13/Vũ'13]+[Ogura-IY'02/Combettes-IY'15] to design an averaged operator whose fixed-point set can express the set of all solutions of the monotone inclusion.

How can we apply this optimization model ?

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_B \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_B \circ L(\mathbf{x}), \quad \mu > 0, \quad (1)$$

$$\text{where } \Psi_B(\cdot) := \|\cdot\|_1 - \min_{\mathbf{v} \in \mathbb{R}^l} \left[\|\mathbf{v}\|_1 + \frac{1}{2} \|B(\cdot - \mathbf{v})\|^2 \right].$$

Q3. Can we choose B flexibly to ensure overall-convexity of (1) ?



[Ayy-ICASSP'19, Proposition 2]

Suppose $L \in \mathbb{R}^{l \times n}$ satisfies $\text{rank}(L) = l$.

Choose a nonsingular $\tilde{L} \in \mathbb{R}^{n \times n}$, s.t., $\begin{bmatrix} O_{l \times (n-l)} & \mathbf{I}_l \end{bmatrix} \tilde{L} = L$.

$B_\theta := \sqrt{\theta/\mu} \Lambda^{1/2} U^T$, $\theta \in [0, 1]$, ensures the convexity of $J_{\Psi_{B_\theta} \circ L}$,

where $U \Lambda U^T := \tilde{A}_2^\top \tilde{A}_2 - \tilde{A}_2^\top \tilde{A}_1 \left(\tilde{A}_1^\top \tilde{A}_1 \right)^\dagger \tilde{A}_1^\top \tilde{A}_2 \in \mathbb{R}^{l \times l}$
is the EVD with $A(\tilde{L})^{-1} = \begin{bmatrix} \tilde{A}_1 & \tilde{A}_2 \end{bmatrix}$.

[Ayy-ICASSP'19] J.Abe, M.Yamagishi, I. Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," *IEEE ICASSP 2019, May, 2019*.

Numerical Experiments (Convexity-Edge-Preserving smoother)

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_{B_\theta} \circ L}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_{B_\theta} \circ L(\mathbf{x}), \quad \mu > 0.$$

We applied the proposed algorithm to

$$\mathbf{y} = A\mathbf{x}_\star + \varepsilon \quad \begin{cases} \mathbf{x}_\star \in \mathbb{R}^{128} : \text{a piecewise constant signal} \\ A \in \mathbb{R}^{100 \times 128} : \text{drawn from } \mathcal{N}(0, 1) \text{ and fixed} \\ \varepsilon \in \mathbb{R}^{100} : \text{AWG to achieve SNR (15dB)} \end{cases}$$

To achieve convexity-edge-preserving smoother, we used

$$L = D := \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1) \times n}, \quad \tilde{L} = \tilde{D} := \begin{bmatrix} 1 & 0 & \cdots & 0 \\ & D & & \end{bmatrix} \in \mathbb{R}^{n \times n}$$

and $B_\theta := \sqrt{\theta/\mu} \Lambda^{1/2} U^T$, $\theta \in [0, 1]$.

$\theta := 0$ (Standard TV)

$\theta := 0.9$ (Nonconvex)

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \quad J_{\Psi_{B_\theta} \circ D}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_{B_\theta} \circ D(\mathbf{x}), \quad \mu > 0.$$

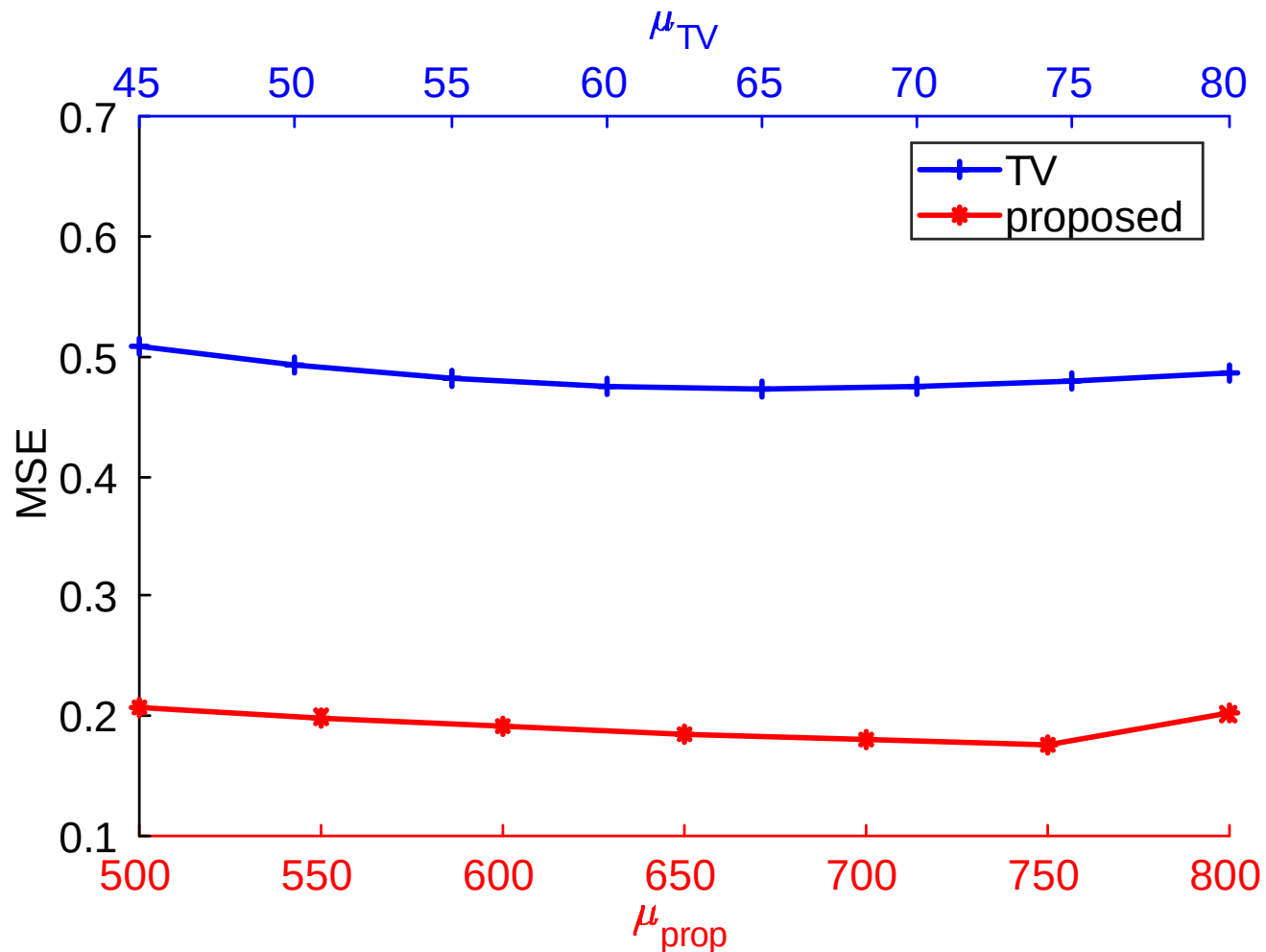


Fig. 1 Dependency of recovering performance on μ .
MSE (mean squared error): Average of $\|\mathbf{x}_{20000} - \mathbf{x}_\star\|^2$
over 100 indepent realizations of ε .

$$\underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimize}} \ J_{\Psi_{B_\theta} \circ D}(\mathbf{x}) := \frac{1}{2} \|\mathbf{y} - A\mathbf{x}\|^2 + \mu \Psi_{B_\theta} \circ D(\mathbf{x}), \ \mu > 0.$$

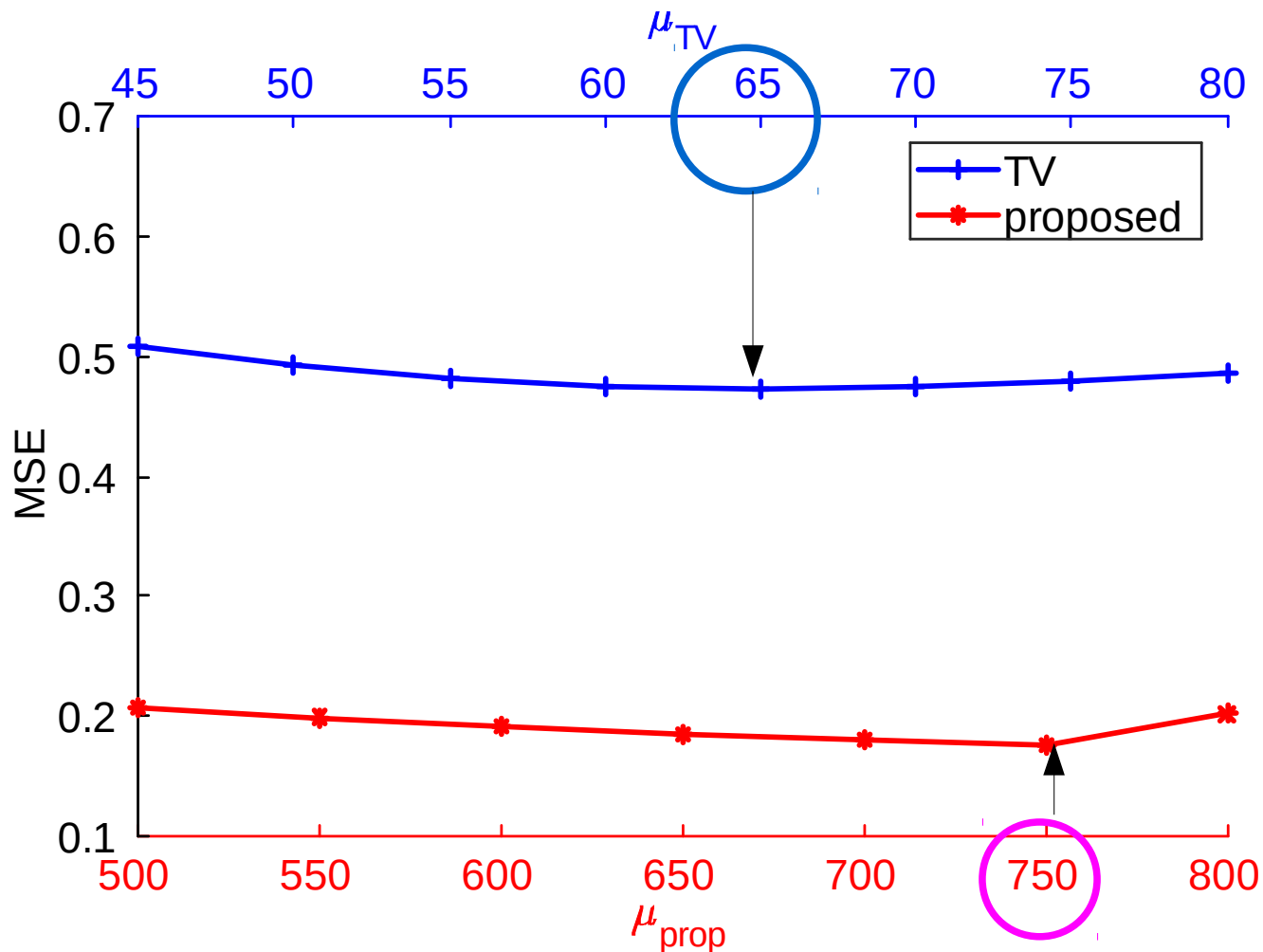


Fig. 1 Dependency of recovering performance on μ .
MSE (mean squared error): Average of $\|\mathbf{x}_{20000} - \mathbf{x}_\star\|^2$
over 100 indepent realizations of ε .

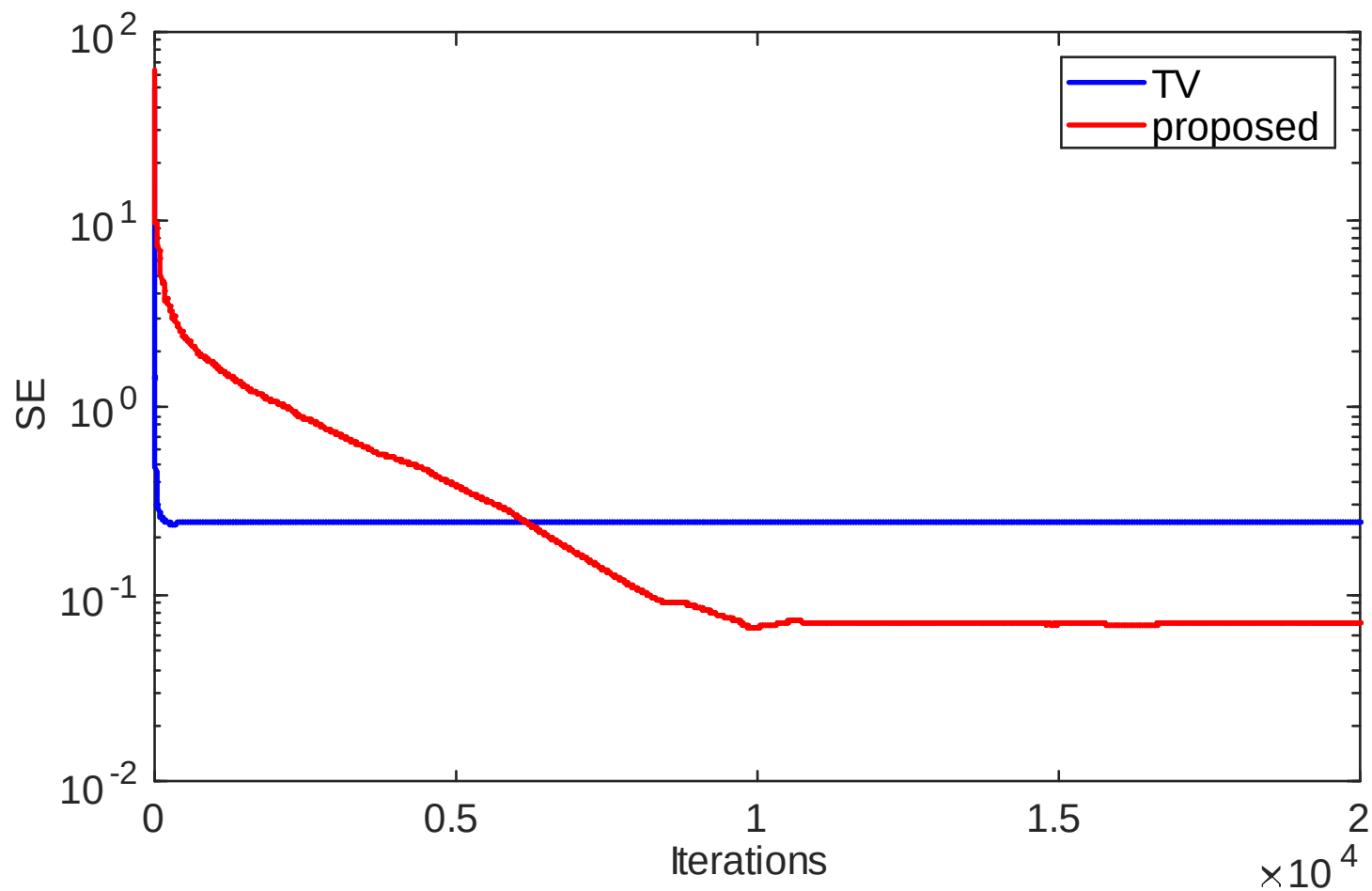


Fig. 2 $\|\mathbf{x}_k - \mathbf{x}_\star\|^2$ vs the iteration number k

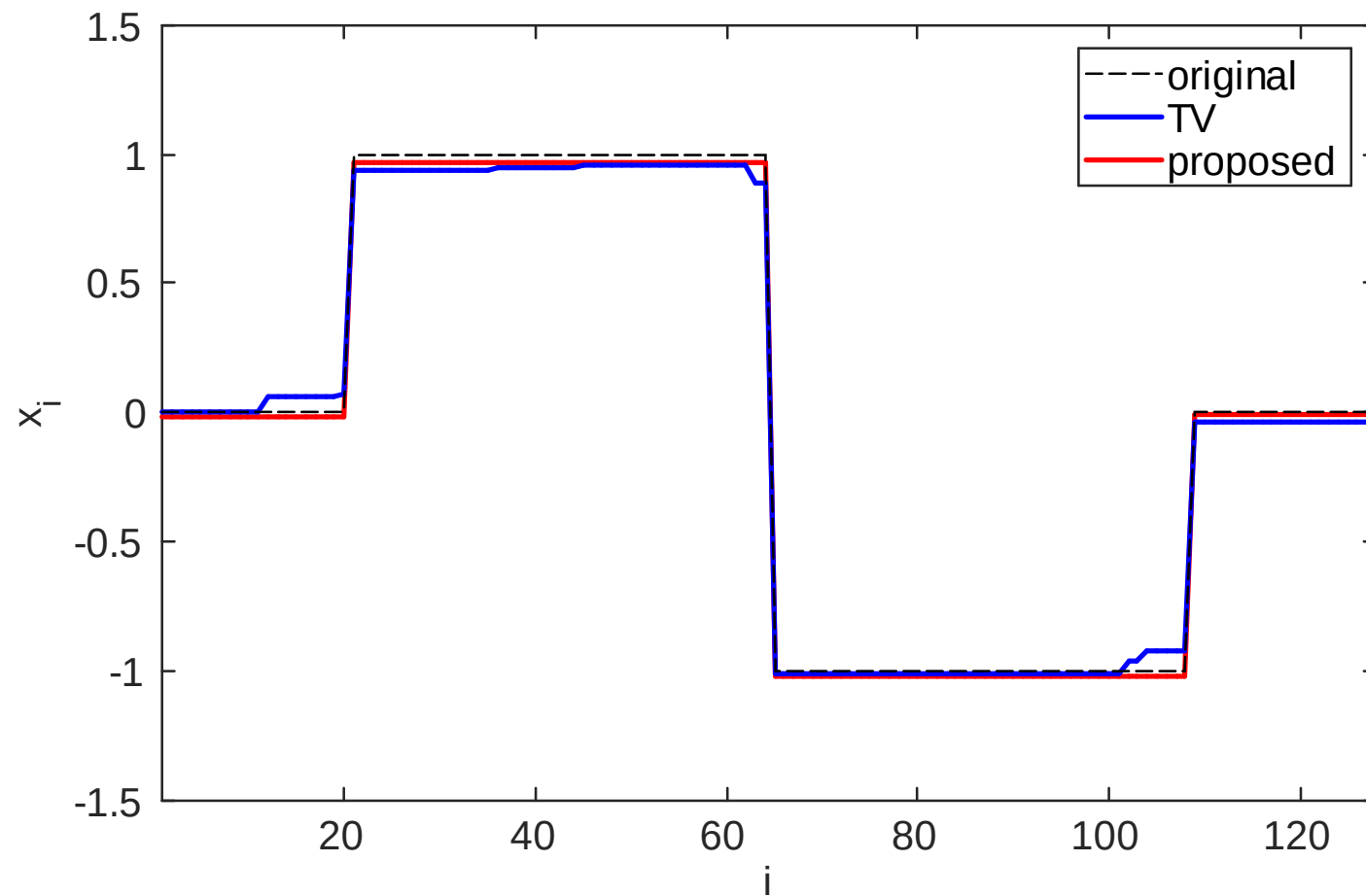


Fig.3 Recovered signals by the proposed algorithm after 20000 iterations

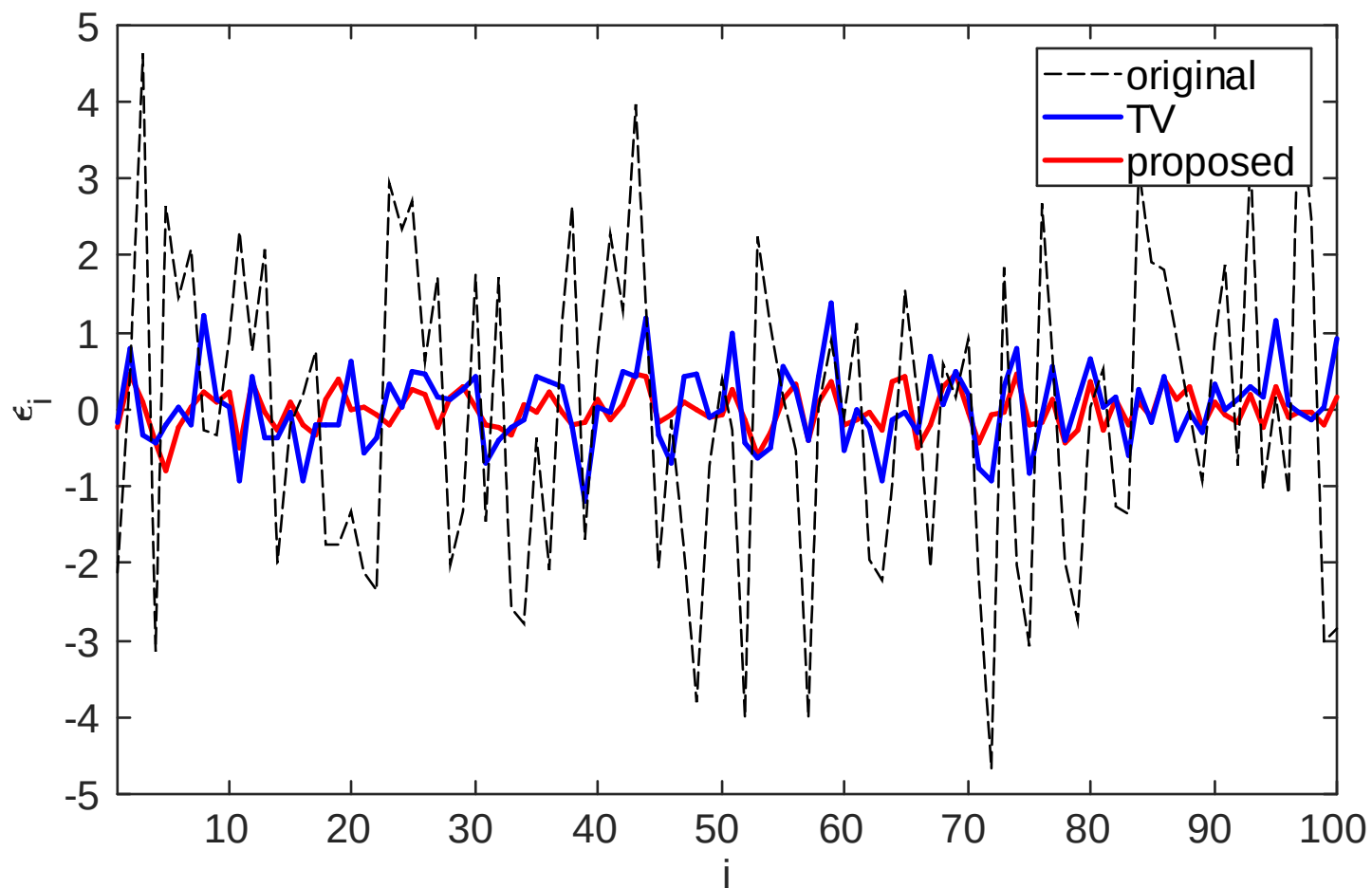


Fig.4 Noise suppression by Standard TV model and Proposed model :
Entries in $y - Ax_*$, $Ax_{TV} - Ax_*$, and $Ax_{prop} - Ax_*$

Conclusion and Open Questions

We have introduced a **proximal splitting algorithm** which has theoretical guarantee of convergence to a globally optimal solution of the **linearly involved generalized minimax concave (GMC) penalty models**.

The proposed algorithm can be applied to fully general cases of these beautiful nonconvex penalty models under overall-convexity condition.

J.Abe, M.Yamagishi and I.Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," IEEE ICASSP 2019, May 2019.

Further extension of the proposed algorithm will be presented at **ICCOPT'2019**, Aug., 2019.

Conclusion and Open Questions

We have introduced a **proximal splitting algorithm** which has theoretical guarantee of convergence to a globally optimal solution of the **linearly involved generalized minimax concave (GMC) penalty models**.

The proposed algorithm can be applied to fully general cases of these beautiful nonconvex penalty models under overall-convexity condition.

J.Abe, M.Yamagishi and I.Yamada, "Convexity-edge-preserving signal recovery with linearly involved generalized minimax concave penalty function," IEEE ICASSP 2019, May 2019.

Further extension of the proposed algorithm will be presented at **ICCOPT'2019**, Aug., 2019.

Is there any clever statistical strategy for tuning good parameters $B \in \mathbb{R}^{m \times l}$ and $\mu > 0$?

How about extension of TREX to GMC penalty models ?

