# CONVEXITY-EDGE-PRESERVING SIGNAL RECOVERY WITH LINEARLY INVOLVED GENERALIZED MINIMAX CONCAVE PENALTY FUNCTION

*Jiro Abe, Masao Yamagishi, and Isao Yamada*

Dept. of Information and Communications Engineering, Tokyo Institute of Technology
Email:{abe,myamagi,isao}@sp.ce.titech.ac.jp

## ABSTRACT

In this paper, we propose a new linearly involved convexity-preserving model for signal recovery by extending the idea in the generalized minimax concave (GMC) penalty [Selesnick'17]. The proposed model can use nonconvex penalties but maintain the overall convexity and is applicable to much more general scenarios of signal recovery than the original GMC model. We also propose a new iterative algorithm which has theoretical guarantee of convergence to a global minimizer of the proposed model. A numerical experiment for noise suppression shows excellent edge-preserving performance of the proposed smoother in comparison with the standard convex TV smoother.

***Index Terms***— Signal recovery, nonconvex penalty, generalized minimax concave penalty function, linearly involved convexity-preserving model, nonexpansive operator

## 1. INTRODUCTION

Many methods for sparsity aware signal recovery rely on the following type of optimization problems:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\ J(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \mu\psi(\boldsymbol{x}),\ \ \mu > 0, \quad (1)$$

where $\boldsymbol{y} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$, and $\psi : \mathbb{R}^n \to \mathbb{R}$ is a certain sparsity promoting function. The $\ell_1$ norm, $\|\boldsymbol{x}\|_1 := \sum_{i=1}^n |x_i|$, has been used extensively as $\psi$, e.g., in LASSO [1] because it is the convex envelope of $\ell_0$ pseudo-norm in the vicinity of $\mathbf{0} \in \mathbb{R}^n$. Nevertheless, it has been often reported that certain nonconvex penalty functions approximate $\ell_0$ pseudo-norm better than $\ell_1$ norm does and that the solutions of the optimization problem (1) with such nonconvex penalties $\psi$ result in better signal recovery [2–5]. However, the computation of a global minimizer of (1) has been challenging because the overall convexity of $J$ in (1) is not guaranteed in general. To overcome the computational difficulty, in problem (1), caused by such nonconvex penalties, convexity-preserving nonconvex penalties were introduced originally by Blake, Zisserman, and Nikolova [6–8]. Remarkably, convexity-preserving nonconvex penalties can maintain the overall convexity of $J$ in (1) under certain conditions. For recent developments of the convexity-preserving nonconvex penalties, see [9–12] and references therein. Most of these works rely on the presence of a strongly convex term, which corresponds to the assumption for nonsingularity of $A^\mathsf{T}A$ in the scenario of (1). An exceptional example which is free from such an assumption is found in the *generalized minimax concave (GMC) penalty function*[1] $\Psi_B$ in [15] with $B \in \mathbb{R}^{m\times n}$ (see Fact 1). Indeed, for any $A \in \mathbb{R}^{m\times n}$, $\Psi_B$ can maintain the overall convexity of $J$ in (1) with appropriate choice of $B$ (see Fact 1(b)). In [15], to demonstrate the effectiveness of the GMC penalty, an iterative algorithm was proposed for problem (1) with $\psi = \Psi_B$ but it is applicable only to a particular case satisfying $B^\mathsf{T}B = (\theta/\mu)A^\mathsf{T}A$, $0 \le \theta \le 1$.

In this paper, we consider linearly involved model:

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

where the dimension of the domain of $\psi$ is properly adjusted to $\mathbb{R}^l$ for $L \in \mathbb{R}^{l\times n}$. Note that, even in the case of convex penalty $\psi$, algorithmic strategies for problem (2) require much more elaborated idea than those for problem (1) (see, e.g., the art of *proximal-splitting techniques* in [16–18]). A typical example of this model is found in an edge-preserving smoothing model:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\ J_D(\boldsymbol{x}) := \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \mu\psi \circ D(\boldsymbol{x}),\ \ \mu > 0, \ (3)$$

where $D \in \mathbb{R}^{(n-1)\times n}$ is the first-order finite difference operator defined as

$$D := \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}. \quad (4)$$

If we employ $\psi(\cdot) = \|\cdot\|_1$, the $\psi \circ D$ is nothing but the total variation (TV) [19] and (3) is the so-called TV smoother used widely as a convex edge-preserving smoother. A question arises: *can we establish a more effective edge-preserving smoother than the TV smoother by employing* $\Psi_B$ *in place of* $\|\cdot\|_1$*?*

[1] The GMC penalty function $\Psi_B$ is a generalization of the so-called *minimax concave (MC) penalty function* [4] (see also [13, 14]). In fact, $\Psi_B$ with $B^\mathsf{T}B = \beta\mathrm{I}_n$ ($\beta \in \mathbb{R}_+$) reproduces the MC penalty function.

To answer this question[2], we first present a condition for the overall convexity of the linearly involved model:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\ \frac{1}{2}\|\boldsymbol{y}-A\boldsymbol{x}\|^2+\mu\Psi_B\circ L(\boldsymbol{x}),\ \ \mu>0. \quad (5)$$

Under this condition, i.e., *linearly involved convexity-preserving condition* (see Proposition 1), we also propose a new iterative algorithm (Algorithm 1) for problem (5) for general $B$. The proposed algorithm is designed in a way similar to an idea behind the primal-dual splitting method [22–24] and has theoretical guarantee of convergence to a global minimizer of (5). The proposed algorithm is applicable to much wider nonconvex penalty than the algorithm introduced in [15].

To demonstrate the effectiveness of the proposed linearly involved convexity-preserving model (5) and Algorithm 1, we present a numerical experiment in a scenario of edge-preserving smoother, i.e., $L=D$ in (4). The numerical example shows that the proposed smoother has excellent edge-preserving performance in comparison with the convex TV smoother.

## 2. PRELIMINARIES

### 2.1. Notation

Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}_{++}$ be the sets of natural numbers, real numbers, nonnegative real numbers, and positive real numbers, respectively. Bold face small letters express vectors. The superscript $(\cdot)^\mathsf{T}$ denotes transpose. For a vector $\boldsymbol{x}:=(x_1,x_2,\ldots,x_n)\in\mathbb{R}^n$, we use $\|\boldsymbol{x}\|_p:=(\sum_{i=1}^n|x_i|^p)^{1/p}$ $(0<p<\infty)$, $\|\boldsymbol{x}\|_\infty:=\max\{|x_1|,\ldots,|x_n|\}$, and $\|\boldsymbol{x}\|_0:=\#\{i\in\mathbb{N}\cap[1,n]\mid x_i\neq 0\}$. For a matrix $A\in\mathbb{R}^{m\times n}$, $A^\dagger\in\mathbb{R}^{n\times m}$ stands for the Moore-Penrose pseudo inverse of $A$, and $\|A\|_2:=\max_{\|\boldsymbol{x}\|_2\le 1}\|A\boldsymbol{x}\|_2$ is given by the maximum singular value of $A$. $\mathrm{O}_{m\times n}\in\mathbb{R}^{m\times n}$ and $\mathrm{O}_n\in\mathbb{R}^{n\times n}$ stand for the zero matrices, and $\mathrm{I}_n\in\mathbb{R}^{n\times n}$ the identity matrix. The positive definiteness and positive semidefiniteness of a symmetric matrix $A\in\mathbb{R}^{n\times n}$ are expressed respectively as $A\succ\mathrm{O}_n$ and $A\succeq\mathrm{O}_n$. For any $A\succ\mathrm{O}_n$, by defining an inner product $\langle\cdot,\cdot\rangle_A:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}\colon(\boldsymbol{x},\boldsymbol{y})\mapsto\boldsymbol{x}^\mathsf{T}A\boldsymbol{y}$ and its induced norm $\|\boldsymbol{x}\|_A:=\sqrt{\langle\boldsymbol{x},\boldsymbol{x}\rangle_A}$, $(\mathbb{R}^n,\langle\cdot,\cdot\rangle_A,\|\boldsymbol{x}\|_A)$ becomes a real Hilbert space whose identity operator is denoted by $\mathrm{Id}$. The set of all proper lower semicontinuous convex functions is denoted by $\Gamma_0(\mathbb{R}^n)$ [16].

### 2.2. Generalized Minimax Concave (GMC) Penalty

The *generalized minimax concave (GMC) penalty function* [15] was introduced as a sparsity-promoting non-convex penalty function.

[2]In a recent paper [20], we have found a positive but a partial answer, to this question, where the authors reported that the model (3) with $\psi=\Psi_B$ in a special case $(A,B^\mathsf{T}B)=(\mathrm{I}_n,\beta\mathrm{I}_{n-1})$ $(\beta\in\mathbb{R}_+)$ shows superior denoising performance not only to (i) the model (3) with $\psi=\|\cdot\|_1$ but also to (ii) the Moreau-enhanced TV denoising in [21].

**Fact 1** (Properties of GMC penalty function [15])**.** *For a given $B\in\mathbb{R}^{m\times n}$, the GMC penalty function $\Psi_B\colon\mathbb{R}^n\to\mathbb{R}_+$ defined as*

$$\Psi_B(\boldsymbol{x}):=\|\boldsymbol{x}\|_1-\min_{\boldsymbol{v}\in\mathbb{R}^n}\left[\|\boldsymbol{v}\|_1+\frac{1}{2}\|B(\boldsymbol{x}-\boldsymbol{v})\|^2\right]$$

*satisfies the following properties:*

*(a) (Nonconvexity [15, Corollary 2]) $\Psi_B(\boldsymbol{x})=\|\boldsymbol{x}\|_1-\frac{1}{2}\|B\boldsymbol{x}\|^2$ if and only if $\|B^\mathsf{T}B\boldsymbol{x}\|_\infty\le 1$. This implies that $\Psi_B=\|\cdot\|_1$ if $B=\mathrm{O}_{m\times n}$, and $\Psi_B$ is nonconvex otherwise.*

*(b) (Convexity-preserving property [15, Theorem 1]) For $(A,B,\mu,\boldsymbol{y})\in\mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}\times\mathbb{R}_{++}\times\mathbb{R}^m$ satisfying $A^\mathsf{T}A-\mu B^\mathsf{T}B\succeq\mathrm{O}_n$, $\frac{1}{2}\|\boldsymbol{y}-A(\cdot)\|^2+\mu\Psi_B(\cdot)\in\Gamma_0(\mathbb{R}^n)$.*

## 3. CONVEXITY-EDGE-PRESERVING SMOOTHER WITH GMC PENALTY

### 3.1. Linearly involved convexity-preserving model for signal recovery

To broaden the applicability of the GMC penalty, we propose

$$\Psi_B\circ L\colon\mathbb{R}^n\to\mathbb{R}_+ \quad (6)$$

as a penalty with $B\in\mathbb{R}^{m\times l}$ and $L\in\mathbb{R}^{l\times n}$ of $\mathrm{rank}(L)=l$. For example, if we choose $L=D$ in (4) and $B=\mathrm{O}_{m\times(n-1)}$, $\Psi_B\circ D$ reproduces the total variation which has been used widely for convex edge-preserving smoother. This means that $\Psi_B\circ D$ with $B\neq\mathrm{O}_{m\times(n-1)}$ can serve as a nonconvex penalty for convexity-edge-preserving smoother.

**Definition 1.** $\Psi_B\circ L$ in (6) is said to satisfy the *linearly involved convexity-preserving condition* for $(A,\mu)\in\mathbb{R}^{m\times n}\times\mathbb{R}_{++}$ if

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

**Proposition 1.** *For $\Psi_B\circ L$ in (6), the following are equivalent:*

*(a) $\Psi_B\circ L$ satisfies the linearly involved convexity-preserving condition for $(A,\mu)\in\mathbb{R}^{m\times n}\times\mathbb{R}_{++}$.*

*(b) $A^\mathsf{T}A-\mu L^\mathsf{T}B^\mathsf{T}BL\succeq\mathrm{O}_n$.*

*(c) $A^\mathsf{T}A-\mu\tilde{L}^\mathsf{T}[\mathrm{O}_{l\times(n-l)}\ \mathrm{I}_l]^\mathsf{T}B^\mathsf{T}B[\mathrm{O}_{l\times(n-l)}\ \mathrm{I}_l]\tilde{L}\succeq\mathrm{O}_n$, where $\tilde{L}\in\mathbb{R}^{n\times n}$ is any nonsingular matrix satisfying*

$$[\mathrm{O}_{l\times(n-l)}\ \mathrm{I}_l]\tilde{L}:=L. \quad (7)$$

Remark that Proposition 1 is not only a generalization but also a refinement of Fact 1(b) because *(a)* $\Leftrightarrow$ *(b)* in Proposition 1 for $L=\mathrm{I}_n$ reproduces $A^\mathsf{T}A-\mu B^\mathsf{T}B\succeq\mathrm{O}_n\Leftrightarrow\frac{1}{2}\|\boldsymbol{y}-A(\cdot)\|^2+\mu\Psi_B(\cdot)\in\Gamma_0(\mathbb{R}^n)$.

**Proposition 2.** *For $(A,L,\mu)\in\mathbb{R}^{m\times n}\times\mathbb{R}^{l\times n}\times\mathbb{R}_{++}$, let*

$$B=\sqrt{\theta/\mu}\Lambda^{1/2}U^\mathsf{T},\quad \theta\in[0,1],$$

*where $U\Lambda U^\mathsf{T}:=\tilde{A}_2^\mathsf{T}\tilde{A}_2-\tilde{A}_2^\mathsf{T}\tilde{A}_1(\tilde{A}_1^\mathsf{T}\tilde{A}_1)^\dagger\tilde{A}_1^\mathsf{T}\tilde{A}_2\in\mathbb{R}^{l\times l}$ is an eigendecomposition with $[\tilde{A}_1\ \tilde{A}_2]:=A(\tilde{L})^{-1}\in\mathbb{R}^{m\times n}$ for $\tilde{L}$ in (7). Then $\Psi_B\circ L$ satisfies the linearly involved convexity-preserving condition for $(A,\mu)$.*

By applying Proposition 2 to $(A, L, \mu) = (A, D, \mu)$, we obtain a novel nonconvex penalty for convexity-edge-preserving smoother. For example, a simplest choice of $\tilde{L}$ in (7) is given by

$$\tilde{L} = \tilde{D} := \left[\, \boldsymbol{e}_1 \,|\, D^\mathsf{T} \,\right]^\mathsf{T} \in \mathbb{R}^{n \times n}, \tag{8}$$

where $\boldsymbol{e}_1 \in \mathbb{R}^n$ is the vector whose 1st component is 1 and whose remaining components are zero.

Now our target is the following convex optimization problem:

**Problem 1** (Linearly involved convexity-preserving model)**.** Suppose that $\Psi_B \circ L$ in (6) satisfies the linearly involved convexity-preserving condition in Definition 1 for $(A, \mu) \in \mathbb{R}^{m \times n} \times \mathbb{R}_{++}$. Then for a given $\boldsymbol{y} \in \mathbb{R}^m$,

$$\text{find } \boldsymbol{x}^\star \in \mathcal{S} := \underset{\boldsymbol{x} \in \mathbb{R}^n}{\arg\min} \left[ \frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|^2 + \mu \Psi_B \circ L(\boldsymbol{x}) \right].$$

### 3.2. How to solve the linearly involved convexity-preserving model

Remark that we cannot apply standard proximal-splitting techniques in [16] to Problem 1 because of the nonconvexity of $\Psi_B$. To establish an iterative algorithm for Problem 1, we first characterize the solution set $\mathcal{S}$ in terms of the fixed point set of a nonexpansive operator in a way similar to an idea behind the primal-dual splitting method [22–24].

**Theorem 1** (Nonexpansive operator $T_{\mathrm{LCP}}$)**.** *In Problem 1, define* $T_{\mathrm{LCP}} : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \to \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l : (\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w}) \mapsto (\boldsymbol{\xi}, \boldsymbol{\zeta}, \boldsymbol{\eta})$ *by*

$$\boldsymbol{\xi} := \left[ \mathrm{I}_n - \frac{1}{\sigma}(A^\mathsf{T} A - \mu L^\mathsf{T} B^\mathsf{T} B L) \right] \boldsymbol{x} - \frac{\mu}{\sigma} L^\mathsf{T} B^\mathsf{T} B \boldsymbol{v} - \frac{\mu}{\sigma} L^\mathsf{T} \boldsymbol{w} + \frac{1}{\sigma} A^\mathsf{T} \boldsymbol{y},$$

$$\boldsymbol{\zeta} := \mathrm{Soft}_{\frac{\mu}{\tau}} \left[ \frac{2\mu}{\tau} B^\mathsf{T} B L \boldsymbol{\xi} - \frac{\mu}{\tau} B^\mathsf{T} B L \boldsymbol{x} + \left( \mathrm{I}_l - \frac{\mu}{\tau} B^\mathsf{T} B \right) \boldsymbol{v} \right],$$

$$\boldsymbol{\eta} := \mathrm{P}_{[-1,1]^l} \left( 2L\boldsymbol{\xi} - L\boldsymbol{x} + \boldsymbol{w} \right),$$

*where* $(\sigma, \tau) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$*, and*

$$\mathrm{Soft}_{\frac{\mu}{\tau}}(\boldsymbol{x})_i := \begin{cases} 0, & \text{if } |x_i| \le \frac{\mu}{\tau}, \\ (|x_i| - \frac{\mu}{\tau}) \frac{x_i}{|x_i|}, & \text{otherwise}, \end{cases} \tag{9}$$

$$\mathrm{P}_{[-1,1]^l}(\boldsymbol{x})_i := \begin{cases} x_i, & \text{if } |x_i| \le 1, \\ \frac{x_i}{|x_i|}, & \text{otherwise}. \end{cases} \tag{10}$$

*Then we have:*

*(a)* $\mathcal{S} = \mathcal{Q}(\mathrm{Fix}(T_{\mathrm{LCP}}))$*, where* $\mathcal{Q} : \mathcal{H} := \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \to \mathbb{R}^n : (\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w}) \mapsto \boldsymbol{x}$*, and* $\mathrm{Fix}(T_{\mathrm{LCP}}) := \{(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w}) \in \mathcal{H} \mid T_{\mathrm{LCP}}(\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w}) = (\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{w})\}$.

*(b) Choose* $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ *satisfying*[3]

$$\left[ \begin{array}{l} \sigma \mathrm{I}_n - \frac{\mu^2}{\tau} L^\mathsf{T} (B^\mathsf{T} B)^2 L - \mu L^\mathsf{T} L \succ \mathrm{O}_n, \\ \sigma \mathrm{I}_n - \frac{\kappa}{2} A^\mathsf{T} A - \mu L^\mathsf{T} L \succeq \mathrm{O}_n, \\ \tau \ge \left( \frac{\kappa}{2} + \frac{2}{\kappa} \right) \mu \|B^\mathsf{T} B\|_2^2. \end{array} \right. \tag{12}$$

*Then*

$$\mathcal{P} := \begin{bmatrix} \sigma \mathrm{I}_n & -\mu L^\mathsf{T} B^\mathsf{T} B & -\mu L^\mathsf{T} \\ -\mu B^\mathsf{T} B L & \tau \mathrm{I}_l & \mathrm{O}_l \\ -\mu L & \mathrm{O}_l & \mu \mathrm{I}_l \end{bmatrix} \succ \mathrm{O}_{n+2l} \tag{13}$$

*and* $T_{\mathrm{LCP}}$ *is* $\frac{\kappa}{2\kappa - 1}$*-averaged nonexpansive [17] in the Hilbert space* $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{P}}, \| \cdot \|_{\mathcal{P}})$*, i.e., for any* $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathcal{H}$*,*

$$\frac{\kappa - 1}{\kappa} \|(\mathrm{Id} - T_{\mathrm{LCP}})(\boldsymbol{z}_1) - (\mathrm{Id} - T_{\mathrm{LCP}})(\boldsymbol{z}_2)\|_{\mathcal{P}}^2 \le \|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_{\mathcal{P}}^2 - \|T_{\mathrm{LCP}}(\boldsymbol{z}_1) - T_{\mathrm{LCP}}(\boldsymbol{z}_2)\|_{\mathcal{P}}^2.$$

[3] For example, any $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ chosen as

$$\left[ \begin{array}{l} \sigma = \| \frac{\kappa}{2} A^\mathsf{T} A + \mu L^\mathsf{T} L \|_2^2 + (\kappa - 1) > 0, \\ \tau = (\frac{\kappa}{2} + \frac{2}{\kappa}) \mu \|B^\mathsf{T} B\|_2^2 + (\kappa - 1) > 0, \\ \kappa > 1, \end{array} \right. \tag{11}$$

satisfies (12).

**Algorithm 1** for Problem 1.

Choose $(\boldsymbol{x}_0, \boldsymbol{v}_0, \boldsymbol{w}_0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$.
Let $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ satisfying (12).
$\triangleright$ See also (11) for selections of $(\sigma, \tau, \kappa)$.
Define $\mathcal{P}$ as in (13).
$k \leftarrow 0$.
**Do**
$\boldsymbol{x}_{k+1} \leftarrow \left[ \mathrm{I}_n - \frac{1}{\sigma}(A^\mathsf{T} A - \mu L^\mathsf{T} B^\mathsf{T} B L) \right] \boldsymbol{x}_k - \frac{\mu}{\sigma} L^\mathsf{T} B^\mathsf{T} B \boldsymbol{v}_k - \frac{\mu}{\sigma} L^\mathsf{T} \boldsymbol{w}_k + \frac{1}{\sigma} A^\mathsf{T} \boldsymbol{y}$
$\boldsymbol{v}_{k+1} \leftarrow \mathrm{Soft}_{\frac{\mu}{\tau}} \left[ \frac{2\mu}{\tau} B^\mathsf{T} B L \boldsymbol{x}_{k+1} - \frac{\mu}{\tau} B^\mathsf{T} B L \boldsymbol{x}_k + \left( \mathrm{I}_l - \frac{\mu}{\tau} B^\mathsf{T} B \right) \boldsymbol{v}_k \right]$
$\triangleright$ $\mathrm{Soft}_{\frac{\mu}{\tau}}$ is defined in (9).
$\boldsymbol{w}_{k+1} \leftarrow \mathrm{P}_{[-1,1]^l} \left( 2L\boldsymbol{x}_{k+1} - L\boldsymbol{x}_k + \boldsymbol{w}_k \right)$
$\triangleright$ $\mathrm{P}_{[-1,1]^l}$ is defined in (10).
$k \leftarrow k + 1$
**while** $\|(\boldsymbol{x}_k, \boldsymbol{v}_k, \boldsymbol{w}_k) - (\boldsymbol{x}_{k-1}, \boldsymbol{v}_{k-1}, \boldsymbol{w}_{k-1})\|_{\mathcal{P}}$ is not sufficiently small
**return** $\boldsymbol{x}_k$

Now, by applying Fact 2 in Appendix to $T_{\mathrm{LCP}}$ in Theorem 1, we obtain an iterative algorithm for Problem 1.

**Theorem 2** (Proposed algorithm and its convergence)**.** *Assume the conditions in Problem 1 and Theorem 1. Then for any initial point* $(\boldsymbol{x}_0, \boldsymbol{v}_0, \boldsymbol{w}_0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$*, the sequence* $(\boldsymbol{x}_k, \boldsymbol{v}_k, \boldsymbol{w}_k)_{k \in \mathbb{N}} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$ *generated by*

$$(\boldsymbol{x}_{k+1}, \boldsymbol{v}_{k+1}, \boldsymbol{w}_{k+1}) = T_{\mathrm{LCP}}(\boldsymbol{x}_k, \boldsymbol{v}_k, \boldsymbol{w}_k)$$

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

$$\lim_{k \to \infty} \boldsymbol{x}_k = \boldsymbol{x}^\star \in \mathcal{S}.$$

## 4. NUMERICAL EXPERIMENT

To demonstrate the effectiveness of the proposed linearly involved convexity-preserving model, we present a numerical experiment in a scenario of convexity-edge-preserving smoother by considering Problem 1 with $L = D$ under the following settings. Entries of $A \in \mathbb{R}^{100 \times 128}$ are drawn from i.i.d. zero-mean white Gaussian noise with unit variance. The observation $\boldsymbol{y} \in \mathbb{R}^{100}$ is generated by $\boldsymbol{y} = A\boldsymbol{x}_\star + \boldsymbol{\varepsilon}$, where $\boldsymbol{x}_\star \in \mathbb{R}^{128}$ is the piecewise constant signal (Fig. 3: dotted) and $\boldsymbol{\varepsilon} \in \mathbb{R}^{100}$ is additive white Gaussian noise. The signal-to-noise ratio (SNR) is 15dB. We compared two penalties: one is $\Psi_{\mathrm{O}_{m \times l}} \circ D$ which reproduces the standard convex TV penalty,

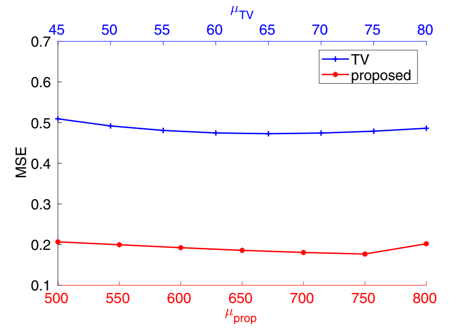


**Fig. 1**: MSE for the proposed model (red) and the standard convex TV (blue) after $k = 20,000$ iterations. $\mu_{\rm prop}$ and $\mu_{\rm TV}$ denote the parameter $\mu$ in Problem 1 for the proposed model and the standard convex TV, respectively.

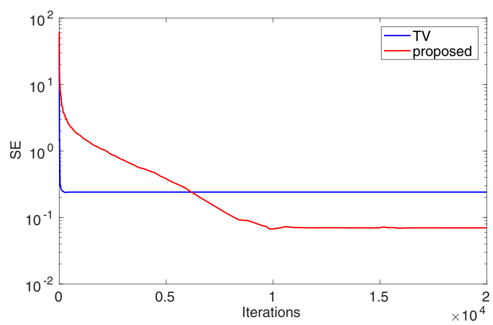


**Fig. 2**: SE versus iterations for the proposed model (red) and the standard convex TV (blue).

the other is the proposed linearly involved convexity-edge-preserving penalty $\Psi_B \circ D$ obtained by applying Proposition 2 with $\tilde{L}$ in (8) and $\theta = 0.9$ to $(A, D, \mu)$. Algorithm 1 with (11) of $\kappa = 1.001$ is applied to the optimization problems.

Fig. 1 shows dependency of recovering performance on the parameter $\mu$ in Problem 1. The performance is measured by mean squared error (MSE) defined as the average of

$$\text{squared error (SE):}\ \ \|\boldsymbol{x}_k - \boldsymbol{x}_\star\|^2$$

over 100 independent realizations of the additive noise. From Fig. 1, we can see that (i) both methods have fair robustness against choice of $\mu$, (ii) the proposed convexity-edge-preserving smoother using $B \neq \mathrm{O}_{m\times l}$ outperforms the standard convex TV smoother in the quality of the recovered signal, and (iii) the best weights of the penalties of both methods are respectively $\mu_{\rm prop} := 750$ for the proposed convexity-edge-preserving smoother and $\mu_{\rm TV} := 65$ for the standard convex TV smoother.

Fig. 2 shows dependency of the SE on the number of iterations under weights $(\mu_{\rm prop}, \mu_{\rm TV})$. Although the recovering is slow possibly due to the nonconvexity of $\Psi_B \circ D$, overwhelming accuracy of the approximation is archived by the proposed method in the end.

Fig. 3 shows the recovered signals by these methods after $20,000$ iterations. The proposed convexity-edge-preserving smoother can maintain much more successfully the sharp edges than the standard convex TV smoother, which also results in excellent noise suppression depicted in Fig.4.

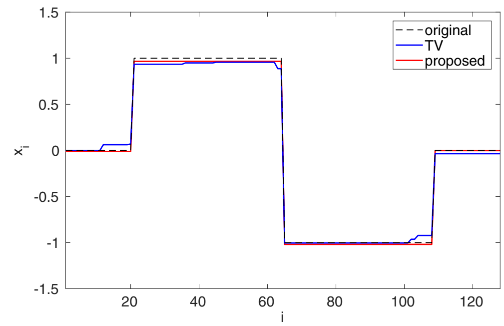


**Fig. 3**: Entries in piecewise constant signal ($\boldsymbol{x}_\star$: dotted black), recovered by the proposed model ($\boldsymbol{x}_{\rm prop}$: red), and by the standard convex TV ($\boldsymbol{x}_{\rm TV}$: blue)

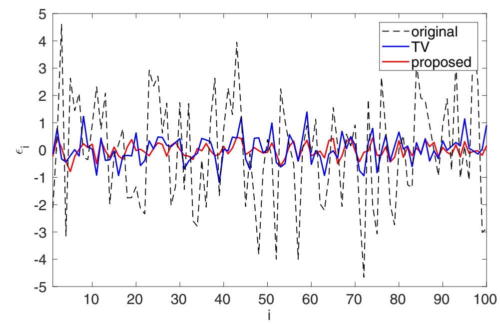


**Fig. 4**: Entries in $\boldsymbol{y} - A\boldsymbol{x}_\star$ (dotted black), $A\boldsymbol{x}_{\rm prop} - A\boldsymbol{x}_\star$ (red), and $A\boldsymbol{x}_{\rm TV} - A\boldsymbol{x}_\star$ (blue), for $\boldsymbol{x}_\star$, $\boldsymbol{x}_{\rm prop}$, and $\boldsymbol{x}_{\rm TV}$ in Fig.3.

## 5. CONCLUDING REMARK

We have shown a necessary and sufficient condition for linearly involved convexity-preserving in Problem 1 (see Proposition 1). Under this general condition, we have also presented a new iterative algorithm (see Algorithm 1). The proposed algorithm has theoretical guarantee of convergence to a global minimizer of Problem 1 (see Theorem 2). A numerical example, in a scenario of edge-preserving smoother, shows that the proposed smoother has excellent edge-preserving performance in comparison with the standard convex TV smoother.

### Appendix

**Fact 2** (Krasnosel'skiĭ-Mann Iteration [25] [16, Section 5.2] for Finding a Fixed Point of Nonexpansive Operator )**.** *Let* $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ *be a real Hilbert space. Let* $T\colon \mathcal{H} \to \mathcal{H}$ *be a nonexpansive operator, i.e.,*

$$\|T(\boldsymbol{z}_1) - T(\boldsymbol{z}_2)\|_{\mathcal{H}} \leq \|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_{\mathcal{H}} \quad (\forall \boldsymbol{z}_1,\ \boldsymbol{z}_2 \in \mathcal{H}),$$

*and* $\mathrm{Fix}(T) := \{\boldsymbol{z} \in \mathcal{H} \mid T(\boldsymbol{z}) = \boldsymbol{z}\} \neq \emptyset$. *Then for any initial point* $\boldsymbol{z}_0 \in \mathcal{H}$*, the sequence* $(\boldsymbol{z}_k)_{k\in\mathbb{N}}$ *generated by*

$$\boldsymbol{z}_{k+1} = [(1-\alpha_k)\mathrm{Id} + \alpha_k T](\boldsymbol{z}_k)$$

*converges weakly to a point in* $\mathrm{Fix}(T)$ *if* $(\alpha_k)_{k\in\mathbb{N}} \subset [0,1]$ *satisfies* $\sum_{k\in\mathbb{N}} \alpha_k(1-\alpha_k) = \infty$. *In particular, if* $T$ *is* $\alpha$*-averaged for some* $\alpha \in (0,1)$*, the sequence* $(\boldsymbol{z}_k)_{k\in\mathbb{N}}$ *generated by a simple iteration*

$$\boldsymbol{z}_{k+1} = T(\boldsymbol{z}_k)$$

*converges weakly to a point in* $\mathrm{Fix}(T)$.

## 6. REFERENCES


[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B*, vol. 58, pp. 267–288, 1996.

[2] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 96, pp. 1348–1360, 2001.

[3] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.

[4] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, pp. 894–942, 2010.

[5] J. Woodworth and R. Chartrand, "Compressed sensing recovery via nonconvex shrinkage penalties," *Inverse Problems*, vol. 32, pp. 75004–75028, 2016.

[6] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, 1987.

[7] M. Nikolova, "Estimation of binary images by minimizing convex criteria," *Proc. IEEE Int. Conf. Image Process.*, vol. 2, pp. 108–112, 1998.

[8] M. Nikolova, "Energy minimization methods," in *Handbook of Mathematical Methods in Imaging*, O. Scherzer, Ed., pp. 138–186. Springer, 2011.

[9] T. Möllenhoff, E. Strekalovskiy, M. Moeller, and D. Cremers, "The primal-dual hybrid gradient method for semiconvex splittings," *SIAM J. Imaging Sci.*, vol. 8, pp. 827–857, 2015.

[10] İ. Bayram, "On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty," *IEEE Trans. Signal Process.*, vol. 64, pp. 1597–1608, 2016.

[11] A. Lanza, S. morigi, and F. Sgallari, "Convex image denoising via non-convex regularization with parameter selection," *J. Math. Imaging and Vision*, pp. 1–26, 2016.

[12] M. Malek-Mohammadi, C. R. Rojas, and B. Wahlberg, "A class of nonconvex penalties preserving overall convexity in optimization-based mean filtering," *IEEE Trans. Signal Process.*, vol. 64, pp. 6650–6664, 2016.

[13] İ. Bayram, "Penalty functions derived from monotone mappings," *IEEE Signal Process. Lett.*, vol. 22, pp. 265–269, 2015.

[14] M. Fornasier and H. Rauhut, "Iterative thresholding algorithms," *J. Appl. Comput. Harmon. Anal.*, vol. 25, pp. 187–208, 2008.

[15] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, pp. 4481–4494, 2017.

[16] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., Springer, 2017.

[17] P. L. Combettes and I. Yamada, "Compositions and convex combinations of averaged nonexpansive operators," *J. Math. Anal. Appl.*, vol. 425, pp. 55–70, 2015.

[18] I. Yamada and M. Yamagishi, "Hierarchical convex optimization by the hybrid steepest descent method with proximal splitting operators - enhancements of SVM and Lasso," in *Splitting Algorithm, Modern Operator Theory and Applications*, H. H. Bauschke, D. R. Luke, and R. Burachik, Ed., 74pp., Springer, 2019 (in press).

[19] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[20] H. Du and Y. Liu, "Minmax-concave total variation denoising," *Signal, Image and Video Processing*, vol. 12, pp. 1027–1034, 2018.

[21] I. Selesnick, "Total variation denoising via the Moreau envelope," *IEEE Signal Process. Lett.*, vol. 24, pp. 216–220, 2017.

[22] B. C. Vũ, "A splitting algorithm for dual monotone inclusions involving cocoercive operators," *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.

[23] L. Condat, "A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms," *J. Optim. Theory Appl.*, vol. 158, pp. 460–479, 2013.

[24] S. Ono and I. Yamada, "Hierarchical convex optimization with primal-dual splitting," *IEEE Trans. Signal Process.*, vol. 63, pp. 373–388, 2015.

[25] C. W. Groetsch, "A note on segmenting Mann iterates," *J. Math. Anal. Appl.*, vol. 40, pp. 369–372, 1972.