A Convexity Preserving Nonconvex Regularization for Inverse Problems under Non-Gaussian Noise

Wataru Yata, Keita Kume, Isao Yamada
Dept. of Information and Communications Engineering
Institute of Science Tokyo, Tokyo, Japan
Email: {yata, kume, isao}@sp.ict.e.titech.ac.jp

***Abstract*—We propose a nonconvexly regularized convex model for linear regression problems under non-Gaussian noise. The cost function of the proposed model is designed with a possibly non-quadratic data fidelity term and a nonconvex regularizer via the generalized Moreau enhancement of a seed convex regularizer. We present sufficient conditions (i) for the cost function of the proposed model to be convex over the entire space, and (ii) for the existence of a minimizer of the proposed model. Under such conditions, we propose a proximal splitting type algorithm with guaranteed convergence to a global minimizer of the proposed model. As an application, we enhance nonconvexly a convex sparsity-promoting regularizer in a scenario of simultaneous declipping and denoising.**

***Index Terms*—generalized Moreau enhancement, convex optimization, non-quadratic data fidelity, proximal splitting.**

I. Introduction

Sparsity-aware estimation of a target signal $x^\star \in \mathscr{X}$ from its noisy observation:

$$y = Ax^\star + \varepsilon \in \mathscr{Y} \tag{1}$$

is a central goal in inverse problems and signal processing [1], [2], where $\mathscr{X}$ and $\mathscr{Y}$ are finite-dimensional real Hilbert spaces, $A \in \mathscr{B}(\mathscr{X},\mathscr{Y})$ is a known linear operator and $\varepsilon \in \mathscr{Y}$ is noise (see Section II-A for mathematical notation).

In this paper, we consider the following nonconvexly regularized convex (NRC) model[1] for estimation of $x^\star$ in (1) by using its prior knowledge on the sparsity of $\mathfrak{L}x^\star$ with a certain known linear operator $\mathfrak{L}$.

Problem 1 (Target NRC model)**.** Let $\mathscr{X},\mathscr{Y},\mathscr{Z},\widetilde{\mathscr{Z}}$ and $\mathfrak{Z}$ be finite-dimensional real Hilbert spaces. Under the following settings (see Section II-A for definitions of *italicized technical terms*):

(a) $\mathbf{C}(\subset \mathfrak{Z})$ is a *simple* closed convex set and $\mathfrak{C} \in \mathscr{B}(\mathscr{X},\mathfrak{Z})$;
(b) $A \in \mathscr{B}(\mathscr{X},\mathscr{Y})$ and $f \in \Gamma_0(\mathscr{Y})$ is continuously differentiable over $\mathscr{Y}$;
(c) We use $\mu \in \mathbb{R}_{++}$, $\mathfrak{L} \in \mathscr{B}(\mathscr{X},\mathscr{Z})$ and $B \in \mathscr{B}\left(\mathscr{Z},\widetilde{\mathscr{Z}}\right)$. $\Psi \in \Gamma_0(\mathscr{Z})$ is *coercive* and *prox-friendly*;
(d) $\mathrm{dom}(\Psi \circ \mathfrak{L}) \cap \mathfrak{C}^{-1}(\mathbf{C}) \neq \emptyset$,

consider a nonconvexly regularized convex model:

$$\underset{\mathfrak{C}x \in \mathbf{C}}{\text{minimize}}\ J(x) := f \circ A(x) + \mu \Psi_B \circ \mathfrak{L}(x), \tag{2}$$

where $\Psi_B : \mathscr{Z} \to (-\infty,\infty]$ is the *generalized Moreau enhancement (GME)* [3] of a seed convex function Ψ:

$$\Psi_B(\cdot) := \Psi(\cdot) - \min_{v \in \mathscr{Z}} \left[\Psi(v) + \frac{1}{2}\|B(\cdot - v)\|^2_{\widetilde{\mathscr{Z}}}\right]. \tag{3}$$

(Note: The GME function Ψ_B in (2) with B being a zero matrix reproduces the seed convex function Ψ.)

We formulate the proposed model (2) motivated by recent studies [3]–[8] on the *convexity-preserving nonconvex regularizers* (see "Related works" in the end of this section). Indeed, the model (2) is an extension of the *(constrained) Linearly involved Generalized Moreau Enhanced (LiGME) model* [3], [7] in order to use a more flexible convex data fidelity term than the quadratic data fidelity case $f := \frac{1}{2}\|y - \cdot\|^2_{\mathscr{Y}}$ where (i) the cost function J in (2) achieves the convexity by a strategical choice [9] of a tuning matrix B and (ii) inner-loop free algorithms were given with convergence guarantees to a global minimizer of the model (2). For broader applications such as computed tomography [10], an extension of f in the model in [7] has also been studied in [8] together with (i) the overall convexity condition of the extended model therein and (ii) an applicable difference-of-convex (DC) type algorithm, along the standard strategy in the DC programming [11], [12], which requires inner loops.

The goal of this paper is to establish an inner-loop free algorithm for the model (2) by extending proximal splitting type algorithms in [3], [7]. More precisely, we address the following research questions according to the model (2) and its algorithm:

(Q1) Under what conditions can we guarantee the existence of a minimizer of the model (2) ?
(Q2) How can we extend the existing proximal splitting type algorithms, developed specially for the quadratic data fidelity case [3], [7], to be applicable to the extended model (2) ?

The remainder of this paper is organized as follows. In Section II, as preliminaries, we introduce notation and selected tools in convex analysis and fixed point theory. In Section III, we first introduce a sufficient condition, called *the overall convexity condition*, for the convexity of J in (2) over $\mathscr{X}$ (see Proposition 1). Under the overall convexity condition, we present in Theorem 1 sufficient conditions for the existence of a minimizer using a classical theorem by Auslender [13]. Under the overall convexity condition and the existence of a minimizer, we also propose a proximal splitting type algorithm with guaranteed convergence to a global minimizer of (2)

This work was supported by JSPS Grants-in-Aid (19H04134, 23KJ0945, 24K23885).

[1]The cost function J in (2) enjoys the convexity by a technical condition (see Proposition 1).

(see Theorem 2). The proposed algorithm requires Lipschitz continuity of ∇f over $\mathscr{Y}$. However, for some applications (e.g., see Section IV), Lipschitz continuity of ∇f can be assumed only over the constraint set $\mathfrak{C}^{-1}(\mathbf{C})$. As a remedy of this issue, under a certain assumption, we present a way to reformulate such an optimization model into a model to which the proposed algorithm can be applied (see Proposition 2). In Section IV, as an application of the proposed methods, we enhance nonconvexly a convex sparsity-promoting regularizer in a convex model [14] for simultaneous declipping and denoising.

Related works

The *convexity-preserving nonconvex regularizers* was pioneered by Blake and Zisserman [15] and by Nikolova [16], [17], and has been developed further. The most of convexity-preserving nonconvex regularizers depend on the strong convexity of the least squares term $\frac{1}{2}\|y-A\cdot\|_{\mathscr{Y}}^2$, i.e., the nonsingularity of A^*A. As an exceptional example which is free from the nonsingularity of A^*A, Selesnick [5] proposed the *generalized minimax concave (GMC) penalty* as a nonconvex enhancement of the ℓ_1 norm[2] $\|\cdot\|_1$ ($\Psi_B\circ\mathfrak{L}$ reproduces the GMC penalty with $\Psi=\|\cdot\|_1$ and $\mathfrak{L}=\mathrm{Id}$). The GMC penalty is a nonseparable multidimensional extension of the minimax concave penalty [4]. To extend the idea of the GMC penalty for a general seed convex function Ψ and a linear operator $\mathfrak{L}$, the LiGME regularizer $\Psi_B\circ\mathfrak{L}$ in (3) has been proposed [3]. See, e.g., [6], [19], [20] for further advancements of the GMC model and [7], [8], [21]–[24] for those[3] of the LiGME model.

II. Preliminaries

A. Notation

Symbols $\mathbb{N}_0$, $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{R}_{++}$ denote respectively all nonnegative integers, all real numbers, all nonnegative real numbers and all positive real numbers. Let $\mathscr{H}$ and $\mathscr{K}$ be finite dimensional real Hilbert spaces. A Hilbert space $\mathscr{H}$ is equipped with an inner product $\langle\cdot,\cdot\rangle_{\mathscr{H}}$ and its induced norm $\|\cdot\|_{\mathscr{H}}$. $\mathscr{B}(\mathscr{H},\mathscr{K})$ denotes the set of all linear operators from $\mathscr{H}$ to $\mathscr{K}$. For $L\in\mathscr{B}(\mathscr{H},\mathscr{K})$, $\|L\|_{\mathrm{op}}$ denotes the operator norm of L (i.e., $\|L\|_{\mathrm{op}}:=\sup_{x\in\mathscr{H},\|x\|_{\mathscr{H}}\le 1}\|Lx\|_{\mathscr{K}}$) and $L^*\in\mathscr{B}(\mathscr{K},\mathscr{H})$ the adjoint operator of L (i.e., $(\forall x\in\mathscr{H})(\forall y\in\mathscr{K})\langle Lx,y\rangle_{\mathscr{K}}=\langle x,L^*y\rangle_{\mathscr{H}}$). The identity operator is denoted by Id and the zero operator from $\mathscr{H}$ to $\mathscr{K}$ by $\mathrm{O}_{\mathscr{H},\mathscr{K}}$. In particular, we use the simplified notation $\mathrm{O}_{\mathscr{H}}$ for the zero operator from $\mathscr{H}$ to $\mathscr{H}$. We express the positive definiteness and the positive semidefiniteness of a self-adjoint operator $L\in\mathscr{B}(\mathscr{H},\mathscr{H})$ as $L\succ\mathrm{O}_{\mathscr{H}}$ and $L\succeq\mathrm{O}_{\mathscr{H}}$, respectively. Any $L\succ\mathrm{O}_{\mathscr{H}}$ defines a new real Hilbert space $(\mathscr{H},\langle\cdot,\cdot\rangle_L,\|\cdot\|_L)$ where $\langle x_1,x_2\rangle_L:=\langle x_1,Lx_2\rangle_{\mathscr{H}}$ and $\|x\|_L:=\sqrt{\langle x,Lx\rangle_{\mathscr{H}}}$. For a linear operator $L\in\mathscr{B}(\mathscr{H},\mathscr{K})$ and sets $S_{\mathscr{H}}\subset\mathscr{H}$ and $S_{\mathscr{K}}\subset\mathscr{K}$, $L(S_{\mathscr{H}}):=\{Lx\in\mathscr{K}\,|\,x\in S_{\mathscr{H}}\}$ is the image of $S_{\mathscr{H}}$ under L, and $L^{-1}(S_{\mathscr{K}}):=\{x\in\mathscr{H}\,|\,Lx\in S_{\mathscr{K}}\}$ is the preimage of $S_{\mathscr{K}}$ under L. A set K is a cone if $(\forall x\in K)(\forall\alpha\in\mathbb{R}_{++})$ $\alpha x\in K$. A set $C\subset\mathscr{H}$ is convex if $(\forall x\in\mathscr{H})(\forall u\in\mathscr{H})(\forall\alpha\in[0,1])$ $\alpha x+(1-\alpha)u\in C$. A closed convex set $C\subset\mathscr{H}$ is said to be *simple* if its metric projection $P_C:\mathscr{H}\to\mathscr{H}:x\to\operatorname{argmin}_{u\in C}\|x-u\|_{\mathscr{H}}$ is available as a computable operator. A function $f:\mathscr{H}\to(-\infty,\infty]$ is said to be (i) proper if $\operatorname{dom}f:=\{x\in\mathscr{H}\,|\,f(x)<\infty\}\neq\emptyset$, (ii) lower semicontinuous if $\operatorname{lev}_{\le\alpha}f:=\{x\in\mathscr{H}\,|\,f(x)\le\alpha\}$ is closed for every $\alpha\in\mathbb{R}$, (iii) convex if $f(\alpha x+(1-\alpha)y)\le\alpha f(x)+(1-\alpha)f(y)$ for every $x,y\in\mathscr{H},0<\alpha<1$. The set of all proper lower semicontinuous convex functions defined on $\mathscr{H}$ is denoted by $\Gamma_0(\mathscr{H})$. $f\in\Gamma_0(\mathscr{H})$ is said to be coercive if $\lim_{\|x\|_{\mathscr{H}}\to\infty}f(x)=\infty$. $f\in\Gamma_0(\mathscr{H})$ is said to be prox-friendly if $\operatorname{Prox}_{\gamma f}:\mathscr{H}\to\mathscr{H}:x\mapsto\operatorname{argmin}_{v\in\mathscr{H}}\left[f(v)+\frac{1}{2\gamma}\|v-x\|_{\mathscr{H}}^2\right]$ is available as a computable operator for every $\gamma>0$. For a nonempty closed convex set $C\subset\mathscr{H}$, the proximity operator of the indicator function $\iota_C\in\Gamma_0(\mathscr{H})$ defined by $\iota_C(x):=\begin{cases}0 & x\in C\\ \infty & x\notin C\end{cases}$ coincides with the metric projection operator onto C, i.e., $\operatorname{Prox}_{\gamma\iota_C}=P_C$ $(\forall\gamma>0)$.

B. Selected tools in fixed point theory

An operator $T:\mathscr{H}\to\mathscr{H}$ is said to be *nonexpansive* if

$$(\forall x\in\mathscr{H})(\forall y\in\mathscr{H})\ \ \|Tx-Ty\|_{\mathscr{H}}\le\|x-y\|_{\mathscr{H}}.$$

In particular, T is α*-averaged nonexpansive* with $\alpha\in(0,1)$ if there exists a nonexpansive operator $\widehat{T}:\mathscr{H}\to\mathscr{H}$ such that $T=(1-\alpha)\mathrm{Id}+\alpha\widehat{T}$.

A fixed point of a nonexpansive operator can be approximated successively via the Krasnosel'skiĭ-Mann iteration.

Fact 1 (Krasnosel'skiĭ-Mann iteration (See, e.g., [25]))**.** Let $T:\mathscr{H}\to\mathscr{H}$ be a nonexpansive operator such that $\operatorname{Fix}(T):=\{x\in\mathscr{H}\mid x=T(x)\}\neq\emptyset$. For any initial point $x_0\in\mathscr{H}$, set $(x_k)_{k\in\mathbb{N}_0}$ by

$$(\forall k\in\mathbb{N}_0)\ x_{k+1}=[(1-\lambda_k)\mathrm{Id}+\lambda_k T](x_k)$$

with a sequence $(\lambda_k)_{k\in\mathbb{N}_0}$ satisfying $(\forall k\in\mathbb{N}_0)$ $\lambda_k\in[0,1]$ and $\sum_{k\in\mathbb{N}_0}\lambda_k(1-\lambda_k)=+\infty$. Then $(x_k)_{k\in\mathbb{N}_0}$ converges to a point in $\operatorname{Fix}(T)$. In particular, if T is α-averaged with $\alpha\in(0,1)$, the sequence generated by $(\forall n\in\mathbb{N}_0)$ $x_{k+1}=T(x_k)$ converges to a point in $\operatorname{Fix}(T)$.

III. Properties of Problem 1 and iterative algorithm for wider applications

A. Overall convexity and existence of minimizer

We start with a sufficient condition for the convexity of the proposed cost function J in (2).

Proposition 1. *In Problem 1, set*

$$\mathfrak{d}:\mathscr{X}\to\mathbb{R}:x\mapsto f\circ A(x)-\frac{\mu}{2}\|B\mathfrak{L}x\|_{\widetilde{\mathscr{Z}}}^2.$$

Then the relation (C_1)*:* $\mathfrak{d}\in\Gamma_0(\mathscr{X})\implies(C_2)$*:* $J\in\Gamma_0(\mathscr{X})$ *holds. In particular, if* f *is twice continuously differentiable over* $\mathscr{Y}$*, then the condition* (C_1) *is equivalent to*

$$(C_1'):(\forall x\in\mathscr{X})\ A^*\circ\nabla^2 f(Ax)\circ A-\mu\mathfrak{L}^*B^*B\mathfrak{L}\succeq\mathrm{O}_{\mathscr{X}}.$$

We call (C_1) *the overall convexity condition.*

Remark 1 (Comparison with existing conditions)**.** For wider applicable conditions than (C_1) and (C_1'), see [8, Theorem 1

[2]The ℓ_1 norm $\|\cdot\|_1$ is the largest convex minorant of the ℓ_0 pseudo-norm in the vicinity of the zero vector. Thus, many models have used $\|\cdot\|_1$ as a sparsity-promoting function. See, e.g., [1], [18].

[3]Applications of LiGME type nonconvex enhancement are not limited to sparsity aware signal estimation. For example, it is applicable to discrete valued signal estimations [24].

and Corollary 1]. In a quadratic case $f := \frac{1}{2}\|\cdot - y\|_{\mathscr{Y}}^2$, the condition (C_1') reproduces the condition in [3, Theorem 1] which is a generalization of a special case in [5, Theorem 1] applicable for $(\Psi, \mathfrak{L}) := (\|\cdot\|_1, \mathrm{Id})$.

Remark 2 (An algebraic design of B enjoying (C_1'))**.** Given $\Lambda \succeq \mathrm{O}_{\mathscr{Y}}$ such that

$$(\forall x \in \mathscr{X})\ \nabla^2 f(Ax) - \Lambda \succeq \mathrm{O}_{\mathscr{Y}}, \tag{4}$$

design B enjoying

$$A^*\Lambda A - \mu \mathfrak{L}^* B^* B \mathfrak{L} \succeq \mathrm{O}_{\mathscr{X}}. \tag{5}$$

GME matrix B enjoying (5) can be designed via LDU decomposition of $\mathfrak{L}$ [9, Theorem 1]. Then B achieves the condition (C_1'). In particular, if $\mathscr{Y} := \mathbb{R}^m$ and $f \in \Gamma_0(\mathbb{R}^m)$ is a separable sum $f(u) := \sum_{i=1}^m f_i([u]_i)$ of twice continuously differentiable convex functions $f_i \in \Gamma_0(\mathbb{R})$ on real line, then a diagonal matrix $\Lambda \in \mathbb{R}^{m\times m}$ with i-th diagonal entry

$$[\Lambda]_{i,i} = \inf_{r\in\mathbb{R}} f_i''(r) \geq 0 \tag{6}$$

satisfies the condition (4).

Essentially based on [13, Theorem 2.4], we derived the following sufficient conditions to guarantee the existence of a minimizer of the model (2) under the overall convexity condition (C_1).

Theorem 1 (Existence of a minimizer)**.** *Consider Problem 1. Assume that the condition (C_1) in Proposition 1 is achieved. Then the model (2) has a minimizer if one of the following holds.*

(i) *f is coercive, and the constraint set $\mathfrak{C}^{-1}(\mathbf{C})$ is asymptotically multipolyhedral [13, Definition 2.3], i.e., the constraint set $\mathfrak{C}^{-1}(\mathbf{C})$ can be decomposed as $\mathfrak{C}^{-1}(\mathbf{C}) = S + K$ with a compact set $S \subset \mathscr{X}$ and a polyhedral cone $K \subset \mathscr{X}$, where the addition of two sets is understood as the Minkowski sum.*

(ii) *The constraint set $\mathfrak{C}^{-1}(\mathbf{C})$ is bounded.*

(iii) *f is coercive and $\mathrm{null} A \cap \mathrm{null}\,\mathfrak{L} = \{0_{\mathscr{X}}\}$.*

For a condition for $\mathfrak{C}^{-1}(\mathbf{C})$ to admit such a decomposition $\mathfrak{C}^{-1}(\mathbf{C}) = S + K$ in (i), see, e.g., [26]. The asymptotically multipolyhedral convex set in the condition (i) of Theorem 1 can cover a wide range of convex sets as follows.

Example 1 (Asymptotically multiprolyhedral sets)**.**

(a) (Polyhedral set). Any polyhedral set can be decomposed as the sum of a polytope (which is compact) and a polyhedral convex cone [26]. Therefore, any polyhedral convex set is asymptotically multipolyhedral.

(b) (Entire space). The entire space $\mathscr{X}$ is a typical example of polyhedral sets and thus asymptotically multipolyhedral as well. For the LiGME model, i.e., the model (2) with $f = \frac{1}{2}\|y - \cdot\|_{\mathscr{Y}}^2$ and $\mathfrak{C}^{-1}(\mathbf{C}) = \mathscr{X}$, the existence of its minimizer is guaranteed by the condition (i) in Theorem 1, while such an existence is assumed implicitly in [3].

(c) (Linearly involved compact set). Let $\mathbf{C}$ be a compact convex set. Then $\mathfrak{C}^{-1}(\mathbf{C})$ can be decomposed as $\mathfrak{C}^{-1}(\mathbf{C}) = \mathfrak{C}^{\dagger}(\mathbf{C}) + \ker\mathfrak{C}$, where $\mathfrak{C}^{\dagger}$ is the Moore-Penrose pseudo inverse of $\mathfrak{C}$, and hence $\mathfrak{C}^{\dagger}(\mathbf{C})$ is compact.

B. Proximal splitting type algorithm for Problem 1 with guaranteed convergence to a global minimizer

We propose an iterative algorithm for finding a global minimizer of the model (2) under the following assumption.

Assumption 1. In Problem 1, assume the following.

(i) $\nabla\mathfrak{d} = A^* \circ \nabla f \circ A - \mu\mathfrak{L}^* B^* B \mathfrak{L}$ is β-Lipschitz continuous over $\mathscr{X}$ for some $\beta \in \mathbb{R}_{++}$.

(ii) The overall convexity condition (C_1) in Proposition 1 is satisfied (see Remark 2 for choice of B enjoying (C_1)).

(iii) A minimizer of (2) exists (see Theorem 1 for sufficient conditions).

(iv) $\Psi \circ \mathfrak{L}$, and $\iota_{\mathbf{C}} \circ \mathfrak{C}$ in (2) satisfy[4]

$$\partial(\mu\Psi \circ \mathfrak{L} + \iota_{\mathbf{C}} \circ \mathfrak{C}) = \mu\mathfrak{L}^* \circ (\partial\Psi) \circ \mathfrak{L} + \mathfrak{C}^* \circ (\partial\iota_{\mathbf{C}}) \circ \mathfrak{C},$$

where the subdifferential $\partial\Psi : \mathscr{Z} \to 2^{\mathscr{Z}}$ of $\Psi \in \Gamma_0(\mathscr{Z})$ is defined as

$$\partial\Psi(z) := \{u \in \mathscr{H} \,|\, (\forall w \in \mathscr{H})\ \langle w - z, u\rangle_{\mathscr{H}} + \Psi(z) \leq \Psi(w)\}.$$

The following theorem shows that the set of all minimizers of the model (2) can be expressed in terms of the fixed point set of an averaged nonexpansive operator T in Theorem 2, and thus a global minimizer can be approximated iteratively by the Krasnosel'skiĭ-Mann iteration of T.

Theorem 2 (Fixed point translation of the solution set of (2))**.** *Consider Problem 1 under Assumption 1. Let $\mathscr{S}$ be the set of all global minimizers of the model (2). Set a product space $\mathscr{H} := \mathscr{X} \times \mathscr{Z} \times \mathscr{Z} \times \mathfrak{Z}$ and define an operator:*

$$T : \mathscr{H} \to \mathscr{H} : (x, v, w, z) \mapsto (\xi, \zeta, \eta, \varsigma)$$

with $(\sigma, \tau) \in \mathbb{R}_{++} \times \mathbb{R}_{++}$ by

$$\xi := \left(\mathrm{Id} - \frac{1}{\sigma}\nabla\mathfrak{d}\right)(x) - \frac{\mu}{\sigma}\mathfrak{L}^* B^* B v - \frac{\mu}{\sigma}\mathfrak{L}^* w - \frac{\mu}{\sigma}\mathfrak{C}^* z$$
$$\zeta := \mathrm{Prox}_{\frac{\mu}{\tau}\Psi}\left[\frac{2\mu}{\tau}B^* B\mathfrak{L}\xi - \frac{\mu}{\tau}B^* B\mathfrak{L}x + \left(\mathrm{Id} - \frac{\mu}{\tau}B^* B\right)(v)\right]$$
$$\eta := (\mathrm{Id} - \mathrm{Prox}_{\Psi})(2\mathfrak{L}\xi - \mathfrak{L}x + w)$$
$$\varsigma := (\mathrm{Id} - P_{\mathbf{C}})(2\mathfrak{C}\xi - \mathfrak{C}x + z).$$

Then the following hold.

(a) *The solution set $\mathscr{S}$ of (2) can be expressed as*

$$\mathscr{S} = \Xi(\mathrm{Fix}\,T) := \{\Xi(h) \in \mathscr{X} \mid h = T(h)\},$$

where $\Xi : \mathscr{H} \to \mathscr{X} : (x, v, w, z) \mapsto x$.

(b) *Choose $(\sigma, \tau) \in (0, \infty) \times (\frac{1}{2\rho}, \infty)$ satisfying*

$$\sigma > \mu\|\mathfrak{L}^*\mathfrak{L} + \mathfrak{C}^*\mathfrak{C}\|_{\mathrm{op}} + \frac{2\rho\mu^2\|B^* B\mathfrak{L}\|_{\mathrm{op}}^2 + \tau}{2\rho\tau - 1}, \tag{7}$$

where $\rho := \frac{1}{\max\{\beta, \mu\|B\|_{\mathrm{op}}^2\}} > 0$. Then

$$\mathfrak{P} := \begin{bmatrix} \sigma\mathrm{Id} & -\mu\mathfrak{L}^* B^* B & -\mu\mathfrak{L}^* & -\mu\mathfrak{C}^* \\ -\mu B^* B\mathfrak{L} & \tau\mathrm{Id} & \mathrm{O}_{\mathscr{Z},\mathscr{Z}} & \mathrm{O}_{\mathfrak{Z},\mathscr{Z}} \\ -\mu\mathfrak{L} & \mathrm{O}_{\mathscr{Z},\mathscr{Z}} & \mu\mathrm{Id} & \mathrm{O}_{\mathfrak{Z},\mathscr{Z}} \\ -\mu\mathfrak{C} & \mathrm{O}_{\mathscr{Z},\mathfrak{Z}} & \mathrm{O}_{\mathscr{Z},\mathfrak{Z}} & \mu\mathrm{Id} \end{bmatrix} \succ \mathrm{O}_{\mathscr{H}}.$$

[4]By [27, Proposition 6.19, Corollary 16.50 and Corollary 16.53], Assumption 1(iv) is satisfied if $\emptyset \neq \mathrm{ri}(\mathrm{dom}(\Psi \circ \mathfrak{L})) \cap \mathfrak{C}^{-1}(\mathbf{C})$ and $\emptyset \neq \mathrm{ri}(\mathbf{C}) \cap \mathrm{ran}\,\mathfrak{C}$. (For a convex set C, $\mathrm{ri}\,C$ denotes the relative interior of C. See, e.g., [27, Definition 6.9].)

Furthermore, T is $\frac{2}{4-\theta}$-averaged nonexpansive over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathfrak{P}}, \|\cdot\|_{\mathfrak{P}})$ with

$$\theta := \frac{\sigma+\tau-\mu\|\mathfrak{L}^*\mathfrak{L}+\mathfrak{C}^*\mathfrak{C}\|_{\mathrm{op}}}{\rho(\sigma\tau-\tau\mu\|\mathfrak{L}^*\mathfrak{L}+\mathfrak{C}^*\mathfrak{C}\|_{\mathrm{op}}-\mu^2\|B^*B\mathfrak{L}\|_{\mathrm{op}}^2)} \in (0,2).$$

(c) *For any initial point $h_0 \in \mathscr{H}$, the sequence $(h_k)_{k\in\mathbb{N}_0} \subset \mathscr{H}$ generated by the Krasnosel'skiĭ-Mann iteration:*

$$(\forall k \in \mathbb{N}_0)\ h_{k+1} := T(h_k) \tag{8}$$

converges to a point in $\mathrm{Fix}\,T$ (see Fact 1), which implies that the sequence $(\mathscr{X} \ni) x_k := \Xi(h_k)$ $(k \in \mathbb{N}_0)$ converges to a point in the solution set $\mathscr{S}$.

Remark 3 (Comparision with existing algorithms)**.** The proposed algorithm can be seen as an extension of [3, Algorithm 1] and [7, Algorithm 1] which are proposed for the quadratic data fidelity case, i.e., $f := \frac{1}{2}\|y-\cdot\|_{\mathscr{Y}}^2$. Moreover, the proposed algorithm is applicable to general cases where the conditions $\mathrm{dom}\,\Psi = \mathscr{X}$ and $\Psi\circ(-\mathrm{Id}) = \Psi$, imposed in [3, Algorithm 1] and [7, Algorithm 1], are no longer satisfied.

C. Remedy for lack of Lipschitz continuous gradient of data fidelity functions

For some applications, Lipschitz continuity assumption in Assumption 1(i) does not hold (see, e.g., the model (13) in Section IV). For such applications, we consider the model (2) under Assumption 1(ii)-(iv) and the following alternative assumption against Assumption 1(i). Hereafter, we denote the i-th component of a vector $u \in \mathbb{R}^m$ by $[u]_i \in \mathbb{R}$.

Assumption 2 (Alternative assumption against Assumption 1(i))**.** In Problem 1, let $\mathscr{Y} := \mathbb{R}^m$ and a closed convex set $\Pi \supset A(\mathfrak{C}^{-1}(\mathbf{C}))$ be decomposable as $\Pi := \times_{i=1}^m \Pi_i (\subset \mathscr{Y})$ with a closed convex set $\Pi_i \subset \mathbb{R}$. Assume that $f \in \Gamma_0(\mathbb{R}^m)$ is a separable sum:

$$(\forall u \in \mathbb{R}^m)\ f(u) = \sum_{i=1}^m f_i([u]_i)$$

of $f_i \in \Gamma_0(\mathbb{R})$ $(1 \le i \le m)$, where $f_i \in \Gamma_0(\mathbb{R})$ $(1 \le i \le m)$ are twice continuously differentiable over Π_i and f_i' $(1 \le i \le m)$ are Lipschitz continuous over Π_i (not necessarily Lipschitz continuous over $\mathbb{R}$).

Due to the lack of Lipschitz continuity of ∇f over $\mathscr{Y}$, Theorem 2(c) under Assumption 1(ii)-(iv) and Assumption 2 no longer guarantees that the proposed algorithm produces a convergent sequence to a global minimizer of the model (2). To circumvent this issue, we introduce an alternative function $\widetilde{f} \in \Gamma_0(\mathscr{Y})$, of f, enjoying desired properties as a data fidelity function applicable to the proposed algorithm.

Proposition 2 (Construction of alternative data fidelity function $\widetilde{f}$ of f)**.** *Consider an instance of the model* (2) *in Problem 1 satisfying Assumption 1(ii)-(iv) and Assumption 2. Then we have the following.*

(a) *Define $\widetilde{f} : \mathbb{R}^m \to \mathbb{R}$ by $\widetilde{f}(u) := \sum_{i=1}^m \widetilde{f}_i([u]_i)$ with univariate functions $(1 \le i \le m)$ $\widetilde{f}_i : \mathbb{R} \to \mathbb{R}$:*

$$r \mapsto \begin{cases} \frac{f_i''(c_r)}{2}(r-c_r)^2 + f_i'(c_r)(r-c_r) + f_i(c_r) & r \notin \Pi_i \\ f_i(r) & r \in \Pi_i, \end{cases}$$

where $c_r := P_{\Pi_i}(r) \in \mathbb{R}$. Then, for an optimization model:

$$\operatorname*{minimize}_{\mathfrak{C}x\in\mathbf{C}}\ \widetilde{f}\circ A(x) + \mu\Psi_B\circ\mathfrak{L}(x), \tag{9}$$

we have

$$\operatorname*{argmin}_{\mathfrak{C}x\in\mathbf{C}}\ f\circ A(x) + \mu\Psi_B\circ\mathfrak{L}(x) = \operatorname*{argmin}_{\mathfrak{C}x\in\mathbf{C}}\ \widetilde{f}\circ A(x) + \mu\Psi_B\circ\mathfrak{L}(x). \tag{10}$$

Moreover, the model (9) *is also an instance of Problem 1 and enjoys the following conditions:*

$$(\forall x \in \mathscr{X})\ A^*\circ\nabla^2\widetilde{f}(Ax)\circ A - \mu\mathfrak{L}^*B^*B\mathfrak{L} \succeq \mathrm{O}_{\mathscr{X}},$$

$$\nabla\left(\widetilde{f}\circ A - \frac{\mu}{2}\|B\mathfrak{L}\cdot\|_{\widetilde{\mathscr{Z}}}^2\right) \text{ is Lipschitz continuous over } \mathscr{Y}.$$

(b) *Generate the sequence $(h_k)_{k\in\mathbb{N}_0}$ by the proposed algorithm* (8) *with a replacement of $\mathfrak{d}$ with $\widetilde{\mathfrak{d}} := \widetilde{f}\circ A - \frac{\mu}{2}\|B\mathfrak{L}\cdot\|_{\widetilde{\mathscr{Z}}}^2$. Then, by Theorem 2, the sequence $x_k := \Xi(h_k)$ $(k \in \mathbb{N}_0)$ converges to a global minimizer $\bar{x} \in \mathfrak{C}^{-1}(\mathbf{C})$ of the model* (9)*. Moreover, from* (10)*, such a minimizer $\bar{x} \in \mathfrak{C}^{-1}(\mathbf{C})$ is also a minimizer of the model* (2)*.*

IV. Application to simultaneous declipping and denoising

A. Formulation via Problem 1

As an application of the proposed model (2), we consider a problem for simultaneous declipping and denoising. The task is to estimate the target signal $x^\star \in \mathbb{R}^m$ from

$$y = \mathrm{clip}_\vartheta(x^\star + \varepsilon) \in \mathbb{R}^m \tag{11}$$

with a priori knowledge that (i) $x^\star \in \mathbf{C}$ with a nonempty closed convex set $\mathbf{C}$ and (ii) $\mathfrak{L}x^\star$ is sparse with a certain linear operator $\mathfrak{L}$, where $\mathrm{clip}_\vartheta : \mathbb{R}^m \to \mathbb{R}^m$ is defined with $\vartheta \in \mathbb{R}_{++}$ as an entrywise operator:

$$[\mathrm{clip}_\vartheta(u)]_i := \begin{cases} [u]_i & |[u]_i| < \vartheta \\ \vartheta\cdot\mathrm{sign}([u]_i) & |[u]_i| \ge \vartheta \end{cases}$$

for every $u \in \mathbb{R}^m$ and $i \in \{1,2,\ldots,m\}$ and the noise values $[\varepsilon]_i$ follows Gaussian distribution with zero mean and known variance s^2 $(s > 0)$. Recently, [14] has proposed a data fidelity function $f : \mathbb{R}^m \to \mathbb{R} : u \mapsto \sum_{i=1}^m f_i([u]_i)$ with

$$f_i([x]_i) := \begin{cases} \frac{1}{2}\left(\frac{[y]_i-[x]_i}{s}\right)^2 & -\vartheta < [y]_i < \vartheta \\ -\log\left(\int_{\vartheta-[x]_i}^{\infty}\exp\left(-\frac{t^2}{2s^2}\right)\mathrm{d}t\right) & [y]_i = \vartheta \\ -\log\left(\int_{-\infty}^{-\vartheta-[x]_i}\exp\left(-\frac{t^2}{2s^2}\right)\mathrm{d}t\right) & [y]_i = -\vartheta \end{cases}$$

for the observation model (11), and formulated[5] an optimization model:

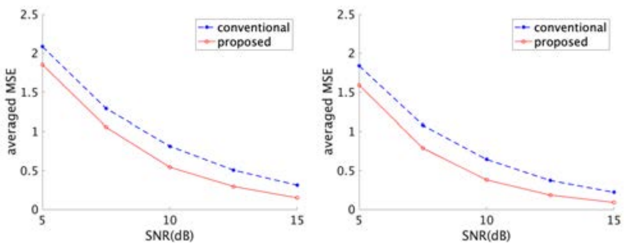
$$\operatorname*{minimize}_{x\in\mathbf{C}}\ f(x) + \mu\|\cdot\|_1\circ\mathfrak{L}(x), \tag{12}$$

where $\|\cdot\|_1 \in \Gamma_0(\mathbb{R}^m)$ is the ℓ_1 norm[6]. To enhance nonconvexly $\|\cdot\|_1$ in (12), we propose

$$\operatorname*{minimize}_{x\in\mathbf{C}}\ f(x) + \mu(\|\cdot\|_1)_B\circ\mathfrak{L}(x) \tag{13}$$

[5]The original model in [14] employed $f\circ A$ with a linear operator A as a data fidelity term and was given without constraints. To clarify the effectiveness of nonconvex enhancement, we consider the simplest case where $A := \mathrm{Id}$ although the proposed model and algorithm can also be applied to a general case $A \ne \mathrm{Id}$.

[6]$\|\cdot\|_1$ is prox-friendly. See, e.g., [27, Example 24.22].



(a) $\vartheta = 0.4$ (b) $\vartheta = 0.6$

Fig. 1: SNR versus Averaged MSE

as a special instance of the model (2). By employing a GME matrix B enjoying (C_1) in Proposition 1 and a compact set as $\mathbf{C}$, the model (13) enjoys Assumption 1(ii-iv) and Assumption 2. Therefore, we can approximate iteratively a minimizer of (13) by applying the proposed algorithm (8) to an alternative optimization model (9) in Proposition 2.

B. Numerical experiments

Following [14], we conducted numerical experiments on estimation of $x^\star \in \mathbb{R}^m$ $(m := 256)$ from its noisy observation y in (11), where $x^\star$ was given by the inverse Discrete Cosine Transform (DCT) of a randomly chosen sparse coefficient vector. The target signal $x^\star$ was normalized to satisfy $\|x^\star\|_\infty = 0.8$. We compared the conventional model (12) and the proposed model (13) for every $(\vartheta, s) \in \{0.4, 0.6\} \times \{s_5, s_{10}, s_{15}\}$, where s_5, s_{10} and s_{15} are standard deviations of Gaussian noise achieving respectively 5dB, 10dB and 15dB of SNR: $20\log_{10}\frac{\|x^\star\|_{\mathbb{R}^m}}{\mathbb{E}[\|\varepsilon\|_{\mathbb{R}^m}]}$. For both models (12) and (13), we employed $\mathbf{C} := [-10, 10]^m$ and $\mathfrak{L} := \mathfrak{L}_{\mathrm{DCT}}$ (DCT matrix [28]), and a simple GME matrix[7] $B := \sqrt{\frac{0.99}{\mu}}\sqrt{\Lambda}\mathfrak{L}_{\mathrm{DCT}}^{-1}$ for achieving the condition (C_1) in Proposition 1 for the model (13). For minimization of (12) and (13), we introduced an alternative data fidelity function $\widetilde{f}$ as in Proposition 2 with $\Pi := \mathbf{C}$ and then applied[8] the proposed algorithm[9] (8). We stopped the proposed algorithm (8) after the residual achieves $\|h_k - h_{k-1}\|_{\mathscr{H}} < 10^{-4}$.

Figure 1 shows the average of MSE: $\|x^\star - \bar{x}\|_2^2$ of estimates $\bar{x}$ by the existing model (12) and by the proposed model (13) over 100 realizations of Gaussian noise ε. For each (ϑ, s^2), we choose the best regularization parameter μ from $\{j \in \mathbb{N}_0 \mid 1 \le j \le 100\}$. From this figure, we see that the proposed model (13) outperforms the model (12) in all cases.

V. Conclusion

In this paper, we proposed a nonconvexly regularized convex model with a smooth data fidelity and the LiGME regularizer. Under the overall convexity condition, we propose sufficient conditions for existence of a minimizer for the proposed model. We also propose a proximal splitting type algorithm for finding a global minimizer of the proposed model. Numerical experiments demonstrate the effectiveness of the proposed model and algorithm in a scenario of simultaneous declipping and denoising.

[7] $\Lambda \in \mathbb{R}^{m\times m}$ was given by (6) with $\inf_{r\in\mathbb{R}} f_i''(r) = \begin{cases} \frac{1}{s^2} & -\vartheta < [y]_i < \vartheta \\ 0 & |[y_i]| = \vartheta. \end{cases}$

[8] The model (13) reproduces the model (12) by setting a zero matrix to B. Thus, we can apply the proposed algorithm (8) to the model (12).

[9] We used $\tau := \frac{5}{2\rho}$ and σ given by $1.001\times$(the value of RHS in (7)).

Acknowledgment: We thank Mr. Shih-Yang Lin in our laboratory for kind help on numerical experiments.

References

[1] M. Elad, *Sparse and Redundant Representations*. Springer, 2010.
[2] S. Theodoridis, *Machine Learning: From the Classics to Deep Networks, Transformers, and Diffusion Models*, 3rd ed. Academic Press, 2024.
[3] J. Abe, M. Yamagishi, and I. Yamada, "Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition," *Inverse Problems*, vol. 36, no. 3, 2020.
[4] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Annals of Statistics*, vol. 38, no. 2, 2010.
[5] I. Selesnick, "Sparse Regularization via Convex Analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, 2017.
[6] A. H. Al-Shabili, Y. Feng, and I. Selesnick, "Sharpening Sparse Regularizers via Smoothing," *IEEE Open Journal of Signal Processing*, vol. 2, 2021.
[7] W. Yata, M. Yamagishi, and I. Yamada, "A constrained LiGME model and its proximal splitting algorithm under overall convexity condition," *Journal of Applied and Numerical Optimization*, vol. 4, no. 2, 2022.
[8] Y. Zhang and I. Yamada, "A Unified Framework for Solving a General Class of Nonconvexly Regularized Convex Models," *IEEE Transactions on Signal Processing*, vol. 23, 2023.
[9] Y. Chen, M. Yamagishi, and I. Yamada, "A Unified Design of Generalized Moreau Enhancement Matrix for Sparsity Aware LiGME Models," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E106.A, no. 8, 2023.
[10] C. Bouman and K. Sauer, "A unified approach to statistical tomography using coordinate descent optimization," *IEEE Transactions on Image Processing*, vol. 5, 1996.
[11] P. D. Tao and L. T. H. An, "Convex analysis approach to dc programming: theory, algorithms and applications," *Acta mathematica vietnamica*, vol. 22, no. 1, 1997.
[12] H. A. Le Thi and T. Pham Dinh, "DC programming and DCA: thirty years of developments," *Mathematical Programming*, vol. 169, no. 1, 2018.
[13] A. Auslender, "Noncoercive Optimization Problems," *Mathematics of Operations Research*, vol. 21, no. 4, 1996.
[14] S. Banerjee, S. Peddabomma, R. Srivastava, and A. Rajwade, "A likelihood based method for compressive signal recovery under gaussian and saturation noise," *Signal Processing*, vol. 217, 2024.
[15] A. Blake and A. Zisserman, *Visual Reconstruction*. MIT Press, 1987.
[16] M. Nikolova, "Estimation of binary images by minimizing convex criteria," in *IEEE ICIP*, vol. 2, 1998.
[17] ——, "Markovian reconstruction using a GNC approach," *IEEE Transactions on Image Processing*, vol. 8, no. 9, 1999.
[18] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, 1996.
[19] Y. Feng, B. Ding, H. Graber, and I. Selesnick, "Transient Artifacts Suppression in Time Series via Convex Analysis," in *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, I. Obeid, I. Selesnick, and J. Picone, Eds. Springer, 2020.
[20] Q. Heng, X. Liu, and E. C. Chi, "Anderson Accelerated Operator Splitting Methods for Convex-nonconvex Regularized Problems," *arXiv preprint (2502.14269v1)*, 2025.
[21] D. Kitahara, R. Kato, H. Kuroda, and A. Hirabayashi, "Multi-Contrast CSMRI Using Common Edge Structures with LiGME Model," in *EUSIPCO*, 2021.
[22] W. Yata and I. Yamada, "Imposing Early and Asymptotic Constraints on LiGME with Application to Nonconvex Enhancement of Fused Lasso Models," in *IEEE ICASSP*, 2024, see arXiv:2309.14082v2 for revised version.
[23] H. Kuroda, "A Convex-Nonconvex Framework for Enhancing Minimization Induced Penalties," *arXiv preprint (2407.14819v2)*, 2024.
[24] S. Shoji, W. Yata, K. Kume, and I. Yamada, "An LiGME Regularizer of Designated Isolated Minimizers - An Application to Discrete-Valued Signal Estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2025.
[25] C. Groetsch, "A note on segmenting Mann iterates," *Journal of Mathematical Analysis and Applications*, vol. 40, no. 2, 1972.
[26] M. A. Goberna, E. González, J. E. Martínez-Legaz, and M. I. Todorov, "Motzkin decomposition of closed convex sets," *Journal of Mathematical Analysis and Applications*, vol. 364, no. 1, 2010.
[27] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Springer, 2017.
[28] K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic press, 1990.