J. Appl. Numer. Optim. 4 (2022), No. 2, pp. 245-271
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.4.2022.2.09

A CONSTRAINED LIGME MODEL AND ITS PROXIMAL SPLITTING ALGORITHM UNDER OVERALL CONVEXITY CONDITION

WATARU YATA, MASAO YAMAGISHI, ISAO YAMADA*

Department of Information and Communications Engineering, Tokyo Institute of Technology, Japan

Abstract. The convex optimization has been used for modeling of many estimation problems in data science and engineering, where convex constraint sets in such a model express respectively a priori knowledge regarding a certain unknown vector to be estimated. The LiGME model was established recently in [J. Abe, M. Yamagishi, I. Yamada, Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition, Inverse Probl. 36 (2020), 035012] for a sound utilization of linearly involved regularizers closer to certain ideal discrete measures, for sparsity as well as for low-rankness, than their convex envelopes. Despite of the nonconvexity of linearly involved regularizers, the LiGME model can keep the overall convexity of its optimization model with a strategic parameter tuning. In this paper, for flexible exploitation of multiple convex constraint sets, we propose a constrained LiGME (cLiGME) model as an enhancement of the original LiGME model. Within the frame of convex optimization, the proposed cLiGME model can promote such desired features more strategically than standard models using convex regularizers, as well as can admit multiple linearly involved convex indicator functions for hard constraints. We also propose a proximal splitting type algorithm for the cLiGME model and demonstrate its effectiveness with a simple numerical experiment. The cLiGME model can be seen as an integration of central ideas in the LiGME model and the set theoretic estimation.
Keywords. Moreau enhancement; Nonconvex penalties; Proximal splitting; Regularized least squares; Set theoretic estimation; Sparsity-rank-aware signal processing and machine learning.

1. INTRODUCTION

Many tasks in data science are formulated as the estimation of an unknown vector $x^\star \in \mathscr{X}$ from the observed vector $y \in \mathscr{Y}$, which follows the linear regression model:

$$y = Ax^\star + \varepsilon, \tag{1.1}$$

where $(\mathscr{X}, \langle\cdot,\cdot\rangle_{\mathscr{X}}, \|\cdot\|_{\mathscr{X}})$ and $(\mathscr{Y}, \langle\cdot,\cdot\rangle_{\mathscr{Y}}, \|\cdot\|_{\mathscr{Y}})$ are finite dimensional real Hilbert spaces, $A : \mathscr{X} \to \mathscr{Y}$ is a known linear operator, and ε is an unknown noise vector. To tackle such

*Corresponding author.
E-mail addresses: yata@sp.ict.e.titech.ac.jp (W. Yata), myamagi@sp.ict.e.titech.ac.jp (M. Yamagishi), isao @sp.ict.e.titech.ac.jp (I. Yamada).
Received September 1, 2021; Accepted May 2, 2022.

estimations, we commonly formulate the following regularized least-squares problem:

$$\underset{x\in\mathscr{X}}{\text{minimize}}\ J_{\Psi\circ\mathfrak{L}}(x) := \frac{1}{2}\|Ax-y\|_{\mathscr{Y}}^2 + \mu\Psi\circ\mathfrak{L}(x), \tag{1.2}$$

where a function $\Psi : \mathscr{Z} \to (-\infty,\infty]$ is designed over a finite dimensional real Hilbert space $(\mathscr{Z},\langle\cdot,\cdot\rangle_{\mathscr{Z}},\|\cdot\|_{\mathscr{Z}})$, $\mathfrak{L} : \mathscr{X} \to \mathscr{Z}$ is a linear operator, and $\mu \in \mathbb{R}_{++}$ is a regularization parameter. In problem (1.2), $\frac{1}{2}\|Ax-y\|_{\mathscr{Y}}^2$ is called the least-squares term, and $\Psi\circ\mathfrak{L}(x)$ is called a regularizer (or penalty), which is designed based on a priori knowledge regarding $x^\star \in \mathscr{X}$. The tuple $(\Psi,\mathfrak{L})$ is designed not only depending on each application but also based on mathematical tractability of the optimization model (1.2). Many models, such as ridge regression [1, 2], Tikhonov regularization [3, 4], l_1-regularization [5, 6, 7] and LASSO (Least Absolute Shrinkage and Selection Operator) [8], are special instances of the model (1.2).

To design $(\Psi,\mathfrak{L})$ in (1.2), one of the most crucial properties to be handled in modern signal processing is the sparsity of vectors or the low-rankness of matrices. The sparsity and the low-rankness have been used for many scenarios in data science, such as compressed sensing [9, 10] and related sparsity aware applications and machine learning [11, 12]. A naive measure of the sparsity is the l_0-pseudonorm $\|\cdot\|_0$, which stands for the number of nonzero components of a given vector. However, the problem (1.2) with $\Psi := \|\cdot\|_0$ is known to be in general NP-hard [13, 14]. To avoid this burden, the l_1-norm $\|\cdot\|_1$, which is the convex envelope of $\|\cdot\|_0$ in the vicinity of zero vector, has been used as Ψ in many applications, e.g., Lasso [8] with $(\Psi,\mathfrak{L}) := (\|\cdot\|_1,\mathrm{Id})$, the convex total variation (TV) [15] with $(\Psi,\mathfrak{L}) := (\|\cdot\|_1,D)$, and the wavelet-based regularization [16, 17] with $(\Psi,\mathfrak{L}) := (\|\cdot\|_1,W)$, where Id is the identity operator, D is the first order differential operator, and W is a wavelet transform matrix.

The convexity of $J_{\Psi\circ\mathfrak{L}}$ is a key property to obtain a global minimizer of $J_{\Psi\circ\mathfrak{L}}$ even by the state-of-the-art optimization techniques. If Ψ is a convex function, $J_{\Psi\circ\mathfrak{L}}$ remains to be convex because the sum of convex functions remains convex. However, the convexity of Ψ is just a sufficient condition for the overall convexity of $J_{\Psi\circ\mathfrak{L}}$. In fact, we can design a nonconvex penalty for $J_{\Psi\circ\mathfrak{L}}$ to be convex. Such penalties are called the *convexity-preserving nonconvex penalties*, which were introduced at latest in 80's by Blake and Zisserman [18] and followed by, for example, Nikolova [19, 20, 21]. For recent developments of the convexity-preserving nonconvex penalties, see [22, 23, 24, 25, 26, 27, 28, 29, 30] and the references therein. Most of these works rely on certain strong convexity assumptions in the least-squares term, which corresponds to the assumption for the nonsingularity of A^*A.

As an exceptional example of the convexity-preserving nonconvex penalty, which is free from the strong convexity condition of the least-squares term, Selesnick [31] proposed the generalized minimax concave (GMC) penalty:

$$(\|\cdot\|_1)_B(\cdot) := \|\cdot\|_1 - \min_{v\in\mathbb{R}^n}\left[\|v\|_1 + \frac{1}{2}\|B(\cdot - v)\|_{\mathbb{R}^q}^2\right] \tag{1.3}$$

with parameter $B \in \mathbb{R}^{q\times n}$. The GMC penalty in (1.3) is a nonseparable multidimensional extension of the minimax concave (MC) penalty [32] and it is in general nonconvex except for $B = \mathrm{O}_{q\times n}$, where $\mathrm{O}_{q\times n}$ stands for the zero matrix in $\mathbb{R}^{q\times n}$. However, the function $J_{(\|\cdot\|_1)_B\circ\mathrm{Id}}$ becomes convex if $A^*A - \mu B^\top B \succeq \mathrm{O}_n$ is achieved, where O_n stands for the zero matrix in $\mathbb{R}^{n\times n}$; see [31, Theorem 1]). Moreover, as seen in [33, Example 2(b)], the GMC penalty can serve as a parametric bridge between the naive discrete measure $\|\cdot\|_0$ and its convex envelope $\|\cdot\|_1$

and thus we can enjoy certain nonconvex enhancements of $\|\cdot\|_1$ to promote further the sparsity without losing the overall convexity of $J_{(\|\cdot\|_1)_B\circ\mathrm{Id}}$.

Abe, Yamagishi and Yamada [33] extended the GMC penalty in (1.3) to the Generalized Moreau Enhanced (GME) penalty:

$$\Psi_B(\cdot) := \Psi(\cdot) - \min_{v\in\mathscr{Z}} \left[\Psi(v) + \frac{1}{2}\|B(\cdot - v)\|^2_{\widetilde{\mathscr{Z}}}\right], \tag{1.4}$$

where (i) $\Psi : \mathscr{Z} \to (-\infty,\infty] \in \Gamma_0(\mathscr{Z})$ is a function, which is coercive (i.e., $\Psi(x) \to \infty$ as $\|x\|_{\mathscr{X}} \to \infty$) and satisfies the even symmetry (i.e., $\Psi \circ (-\mathrm{Id}) = \Psi$) with $\mathrm{dom}(\Psi) = \mathscr{Z}$, (ii) $B \in \mathscr{B}\left(\mathscr{Z}, \widetilde{\mathscr{Z}}\right)$ is a bounded linear operator from $\mathscr{Z}$ to a new finite dimensional real Hilbert space $(\widetilde{\mathscr{Z}}, \langle\cdot,\cdot\rangle_{\widetilde{\mathscr{Z}}}, \|\cdot\|_{\widetilde{\mathscr{Z}}})$. Moreover, [33] proposed the Linearly involved Generalized Moreau Enhanced (LiGME) model:

$$\underset{x\in\mathscr{X}}{\text{minimize}}\ J_{\Psi_B\circ\mathfrak{L}} = \frac{1}{2}\|Ax - y\|^2_{\mathscr{Y}} + \mu\Psi_B\circ\mathfrak{L}(x), \tag{1.5}$$

where $\mathfrak{L} : \mathscr{X} \to \mathscr{Z}$ is a bounded linear operator. In particular, if $A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L}$ is positive semi-definite (see [33, Proposition 2] for existence of such B), then the convexity of $J_{\Psi_B\circ\mathfrak{L}}$ is guaranteed (see [33, Proposition 1]). This implies that we can choose Ψ and $\mathfrak{L}$ freely in the LiGME model. Moreover, in [33, Theorem 1], a proximal splitting type algorithm was also presented for the minimization of $J_{\Psi_B\circ\mathfrak{L}}$.

In contrast, the goal of the set theoretic estimations [34, 35, 36, 37] is to find, if exists, a target vector $x^\star$ satisfying simultaneously $\mathfrak{C}_i x^\star \in C_i\ (i \in I)$, where I is a finite index set, $C_i\ (i \in I)$ is a nonempty closed convex subset of a finite dimensional real Hilbert space $\mathfrak{Z}_i$, and $\mathfrak{C}_i : \mathscr{X} \to \mathfrak{Z}_i$ is a linear operator. Remarkable effectiveness of the set theoretic estimations has been proven in many successful applications [36, 37], e.g., computerized tomography [34, 38], signal processing [35], and electron microscopy [39].

In this paper, for evolution of these strategies, we propose to integrate central ideas in the LiGME model and the set theoretic estimations into the model

$$\underset{(i\in I)\ \mathfrak{C}_i x\in C_i\subset\mathfrak{Z}_i}{\text{minimize}}\ J_{\Psi_B\circ\mathfrak{L}}(x). \tag{1.6}$$

Actually in order to deal with model (1.6), we propose its equivalent but seemingly simpler model (3.1) [and its further equivalent but constraint-free expression (3.3)] as the constrained LiGME (cLiGME) model (see Remark 3.1 and Proposition 3.1). In Proposition 3.2, we present an overall convexity condition for model (3.3). We also present a way to design such a matrix B satisfying the overall convexity condition (see Proposition 3.3 and Corollary 3.1). Moreover, as an extension of [33, Theorem 1], we present a novel proximal splitting type algorithm (Algorithm 1) of guaranteed convergence to a global minimizer of the model (3.3) under its overall convexity condition (see Assumption 3.1 and Theorem 3.1). Finally, to demonstrate the efficacy of the proposed model and algorithm, we present in Section 4 a rather simple but fairly convincing numerical experiment. A preliminary short version of this paper was prepared for a conference [40].

2. PRELIMINARIES

Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}_{++}$ be the set of nonnegative integers, real numbers, nonnegative real numbers, and positive real numbers, respectively. In this paper, we assume that $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}}, \|\cdot\|_{\mathscr{H}})$ and $(\mathscr{K}, \langle\cdot,\cdot\rangle_{\mathscr{K}}, \|\cdot\|_{\mathscr{K}})$ are finite dimensional real Hilbert spaces. For $S \subset \mathscr{H}$, $\mathrm{ri}(S)$ denotes the relative interior of S (see, e.g., [41, Definition 6.9]). $\mathscr{B}(\mathscr{H},\mathscr{K})$ denotes the set of all linear operators from $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}}, \|\cdot\|_{\mathscr{H}})$ to $(\mathscr{K}, \langle\cdot,\cdot\rangle_{\mathscr{K}}, \|\cdot\|_{\mathscr{K}})$. For $L \in \mathscr{B}(\mathscr{H},\mathscr{K})$, $\|L\|_{\mathrm{op}}$ denotes the operator norm of L (i.e., $\|L\|_{\mathrm{op}} := \sup_{x\in\mathscr{H}, \|x\|_{\mathscr{H}}\leq 1} \|Lx\|_{\mathscr{K}}$) 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 of general Hilbert spaces is denoted by Id and the zero operator from $\mathscr{H}$ to $\mathscr{K}$ is denoted by $\mathrm{O}_{\mathscr{B}(\mathscr{H},\mathscr{K})} \in \mathscr{B}(\mathscr{H},\mathscr{K})$. In particular, if $\mathscr{H} = \mathscr{K}$, the zero operator is denoted by $\mathrm{O}_{\mathscr{H}} \in \mathscr{B}(\mathscr{H},\mathscr{H})$. For $L \in \mathscr{B}(\mathscr{H},\mathscr{K})$, $\mathrm{ran}(L) := \{y \in \mathscr{K} \,|\, (\exists x \in \mathscr{H})\ Lx = y\}$ and $\mathrm{null}(L) := \{x \in \mathscr{H} \,|\, Lx = 0_{\mathscr{K}}\}$, where $0_{\mathscr{K}}$ stands for the zero vector in $\mathscr{K}$. 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 Hilbert space $(\mathscr{H}, \langle\cdot,\cdot\rangle_L, \|\cdot\|_L)$ equipped with an inner product $\langle\cdot,\cdot\rangle_L : \mathscr{H} \times \mathscr{H} \to \mathbb{R} : (x,y) \mapsto \langle x, Ly\rangle_{\mathscr{H}}$ and its induced norm $\|\cdot\|_L : \mathscr{H} \to \mathbb{R} : x \mapsto \sqrt{\langle x,x\rangle_L}$. For multiple real Hilbert spaces $(\mathscr{H}_i, \langle\cdot,\cdot\rangle_{\mathscr{H}_i}, \|\cdot\|_{\mathscr{H}_i})$ $(i = 1,2,\cdots,p)$, the product space $\mathscr{H} := \times_{1\leq i\leq p} \mathscr{H}_i$ can be seen as a real Hilbert space equipped with the vector addition $\mathscr{H} \to \mathscr{H} : (x,y) = ((x_i)_{1\leq i\leq p}, (y_i)_{1\leq i\leq p}) \mapsto (x_i + y_i)_{1\leq i\leq p}$, the scalar multiplication $\mathbb{R} \times \mathscr{H} \to \mathscr{H} : (\alpha, (x_i)_{1\leq i\leq p}) \mapsto (\alpha x_i)_{1\leq i\leq p}$ and the inner product $\mathscr{H} \times \mathscr{H} \to \mathbb{R} : (x,y) \mapsto \sum_i^p \langle x_i, y_i\rangle_{\mathscr{H}_i}$. In this paper, unless otherwise stated, the product space of multiple real Hilbert spaces $(\mathscr{H}_i, \langle\cdot,\cdot\rangle_{\mathscr{H}_i}, \|\cdot\|_{\mathscr{H}_i})$ $(i = 1,2,\cdots,p)$ is understood as a new real Hilbert space in this standard sense (The only exception is found in Theorem 3.1). For a matrix A, A^{T} denotes the transpose of A, and $A^\dagger$ the Moore-Penrose pseudo inverse of A. We use $\mathrm{I}_n \in \mathbb{R}^{n\times n}$ to denote the identity matrix for $\mathbb{R}^n$. We also use $\mathrm{O}_{m,n} \in \mathbb{R}^{m\times n}$ for the zero matrix. In particular, if $m = n$, we use the simplified notation $\mathrm{O}_n \in \mathbb{R}^{n\times n}$ for the zero matrix.

2.1. Tools in convex analysis, monotone operator theory and fixed point theory of nonexpansive operator.

A function $f : \mathscr{H} \to (-\infty,\infty]$ is called a proper lower semicontinuous convex function if f is proper (i.e., $\mathrm{dom}(f) := \{x \in \mathscr{H} \,|\, f(x) < \infty\} \neq \emptyset$), lower semicontinuous (i.e., the lower level set $\{x \in \mathscr{H} \,|\, f(x) \leq \alpha\}$ of f is closed for every $\alpha \in \mathbb{R}$), and convex. The set of all proper lower semicontinuous convex functions from $\mathscr{H}$ to $(-\infty,\infty]$ is is denoted by $\Gamma_0(\mathscr{H})$.

Fact 2.1 (Convexity preserving operations)**.**

(a) (Sum of convex functions [41, Corollary 9.4]). For $f_i \in \Gamma_0(\mathscr{H})$ $(i \in I$: finite index set) with $\bigcap_{i\in I} \mathrm{dom}(f_i) \neq \emptyset$, $\sum_{i\in I} f_i \in \Gamma_0(\mathscr{H})$ holds. In particular, for real Hilbert spaces $(\mathscr{H}_i, \langle\cdot,\cdot\rangle_{\mathscr{H}_i}, \|\cdot\|_{\mathscr{H}_i})$ and $f_i \in \Gamma_0(\mathscr{H}_i)$ $(i \in I)$,

$$f := \bigoplus_{i\in I} f_i : \mathscr{H} := \bigtimes_{i\in I} \mathscr{H}_i \to (-\infty,\infty] : x = (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i),$$

called the *separable sum* of $f_i \in \Gamma_0(\mathscr{H}_i)$ $(i \in I)$, satisfies $f \in \Gamma_0(\mathscr{H})$.

(b) (Composition of linear operator and convex function [41, Proposition 9.5]). For $f \in \Gamma_0(\mathscr{K})$ and $L \in \mathscr{B}(\mathscr{H},\mathscr{K})$, $f \circ L \in \Gamma_0(\mathscr{H})$ if $\mathrm{dom}(f) \cap \mathrm{ran}\, L \neq \emptyset$.

(**Subdifferential**) For $f \in \Gamma_0(\mathscr{H})$, the *subdifferential* of f is defined as the set valued operator

$$\partial f : \mathscr{H} \to 2^{\mathscr{H}} : x \mapsto \{u \in \mathscr{H} | (\forall y \in \mathscr{H})\ \langle y - x, u \rangle_{\mathscr{H}} + f(x) \leq f(y)\}.$$

For each $x \in \mathscr{H}$ satisfying $\partial f(x) \neq \emptyset$, the elements of $\partial f(x)$ are called the subgradients of f at x.

Fact 2.2 (Some properties of subdifferential)**.**

(a) (Fermat's rule [41, Theorem 16.3]). For $f \in \Gamma_0(\mathscr{H})$,

$$x^\star \in \operatorname*{argmin}_{x \in \mathscr{H}} f(x) \iff 0_{\mathscr{H}} \in \partial f(x^\star). \tag{2.1}$$

(b) (Sum rule [41, Corollary 16.48]). Let $f, g \in \Gamma_0(\mathscr{H})$ s.t. $\mathrm{dom}(g) = \mathscr{H}$. Then

$$\partial(f + g) = \partial f + \partial g. \tag{2.2}$$

(c) (Chain rule [41, Corollary 16.53]). Let $g \in \Gamma_0(\mathscr{K})$ and $L \in \mathscr{B}(\mathscr{H}, \mathscr{K})$. Suppose that $0_{\mathscr{K}} \in \mathrm{ri}(\mathrm{dom}\, g - \mathrm{ran}\, L)$. Then

$$\partial(g \circ L) = L^* \circ \partial g \circ L. \tag{2.3}$$

(d) (Subdifferential and Gâteaux differential [41, Proposition 17.31]). Let $f \in \Gamma_0(\mathscr{H})$ and $x \in \mathrm{dom}(f)$. Suppose that f is Gâteaux differentiable at x. Then $\partial f(x) = \{\nabla f(x)\}$.

(**Legendre-Fenchel Conjugate**) The (Legendre-Fenchel) conjugate of $f \in \Gamma_0(\mathscr{H})$ is defined as

$$f^* : \mathscr{H} \to [-\infty, \infty] : y \mapsto \sup_{x \in \mathscr{H}} \{\langle x, y \rangle_{\mathscr{H}} - f(x)\}.$$

For $f \in \Gamma_0(\mathscr{H})$, $f^* \in \Gamma_0(\mathscr{H})$ is guaranteed, and we have

$$(\forall (x, u) \in \mathscr{H} \times \mathscr{H})\ u \in \partial f(x) \iff x \in \partial f^*(u). \tag{2.4}$$

(**Nonexpansive Operator**) An operator $T : \mathscr{H} \to \mathscr{H}$ is *nonexpansive* if

$$(\forall x \in \mathscr{H})(\forall y \in \mathscr{H})\ \|T(x) - T(y)\|_{\mathscr{H}} \leq \|x - y\|_{\mathscr{H}}. \tag{2.5}$$

In particular, T is called α*-averaged nonexpansive* with $\alpha \in (0,1)$ if there exists a nonexpansive operator $\hat{T} : \mathscr{H} \to \mathscr{H}$ s.t.

$$T = (1 - \alpha)\,\mathrm{Id} + \alpha \hat{T}. \tag{2.6}$$

Fact 2.3 (Some properties of nonexpansive operator)**.**

(a) (Composition of averaged nonexpansive operators [42] [43, Proposition 2.6]). Suppose that each $T_i : \mathscr{H} \to \mathscr{H} (i = 1, 2)$ is α_i-averaged nonexpansive with $\alpha_i \in (0,1)$. Then $T_1 \circ T_2$ is α-averaged nonexpansive with $\alpha := \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in (0,1)$.

(b) (Krasnosel'skiĭ-Mann algorithm [41, Theorem 5.14]). Let $T : \mathscr{H} \to \mathscr{H}$ be a nonexpansive operator s.t. $\mathrm{Fix}(T) := \{x \in \mathscr{H} | x = T(x)\} \neq \emptyset$. Define $(x_n)_{n \in \mathbb{N}} \subset \mathscr{H}$ by

$$(\forall n \in \mathbb{N})\ x_{n+1} = [(1 - \lambda_n)\,\mathrm{Id} + \lambda_n T](x_n)$$

with any $x_0 \in \mathscr{H}$ and $\lambda_n \in [0,1]$ satisfying $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathrm{Fix}(T)$. In particular, if T is α-averaged with $\alpha \in (0,1)$, the sequence generated by

$$(\forall n \in \mathbb{N})\ x_{n+1} = T(x_n)$$

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

(**Monotone Operator**) A set-valued operator $A : \mathscr{H} \to 2^{\mathscr{H}}$ is *monotone* if

$$(\forall (x,u) \in \mathrm{gra}(A))(\forall (y,v) \in \mathrm{gra}(A))\ \langle x-y, u-v\rangle \geq 0,$$

where $\mathrm{gra}(A) := \{(x,u) \in \mathscr{H} \times \mathscr{H} | u \in A(x)\}$. In particular, A is *maximally monotone* if, for every $(x,u) \in \mathscr{H} \times \mathscr{H}$,

$$(x,u) \in \mathrm{gra}(A) \iff (\forall (y,v) \in \mathrm{gra}(A))\ \langle x-y, u-v\rangle \geq 0.$$

A set-valued operator $A : \mathscr{H} \to 2^{\mathscr{H}}$ is maximally monotone if and only if the *resolvent* $R_A := (\mathrm{Id}+A)^{-1} : \mathscr{H} \to 2^{\mathscr{H}} : u \mapsto \{x \in \mathscr{H} | u \in x + A(x)\}$ single-valued and $\frac{1}{2}$-averaged nonexpansive. For $f \in \Gamma_0(\mathscr{H})$, its subdifferential ∂f is maximally monotone.
(**Proximity Operator**) *The proximity operator* of $f \in \Gamma_0(\mathscr{H})$ is defined by

$$\mathrm{Prox}_f : \mathscr{H} \to \mathscr{H} : x \mapsto \underset{y\in\mathscr{H}}{\mathrm{argmin}} \left[f(y) + \frac{1}{2}\|x-y\|_{\mathscr{H}}^2 \right].$$

The proximity operator of $f \in \Gamma_0(\mathscr{H})$ is the resolvent of ∂f, i.e.,

$$\mathrm{Prox}_f = (\mathrm{Id} + \partial f)^{-1} = R_{\partial f}, \tag{2.7}$$

which implies that

$$\begin{aligned} \bar{x} \in \mathrm{Fix}(\mathrm{Prox}_f) &\iff \bar{x} = (\mathrm{Id}+\partial f)^{-1}(\bar{x}) \iff \bar{x} \in (\mathrm{Id}+\partial f)(\bar{x}) \\ &\iff \bar{x} \in \bar{x} + \partial f(\bar{x}) \iff 0_{\mathscr{H}} \in \partial f(\bar{x}) \iff \bar{x} \in \underset{x\in\mathscr{H}}{\mathrm{argmin}}\ f(x) \end{aligned}$$

with Fermat's rule (Fact 2.2(a)). The proximity operator of conjugate $f^* \in \Gamma_0(\mathscr{H})$ of $f \in \Gamma_0(\mathscr{H})$ can be can be expressed as $\mathrm{Prox}_{f^*} = \mathrm{Id} - \mathrm{Prox}_f$.

2.2. **LiGME model.**

Definition 2.1 (Linearly involved Generalized-Moreau-Enhanced (LiGME) Model [33, Definition 1]). Let $(\mathscr{X}, \langle\cdot,\cdot\rangle_{\mathscr{X}}, \|\cdot\|_{\mathscr{X}})$, $(\mathscr{Y}, \langle\cdot,\cdot\rangle_{\mathscr{Y}}, \|\cdot\|_{\mathscr{Y}})$, $(\mathscr{Z}, \langle\cdot,\cdot\rangle_{\mathscr{Z}}, \|\cdot\|_{\mathscr{Z}})$ and $(\widetilde{\mathscr{Z}}, \langle\cdot,\cdot\rangle_{\widetilde{\mathscr{Z}}}, \|\cdot\|_{\widetilde{\mathscr{Z}}})$ be finite dimensional real Hilbert spaces, $\Psi \in \Gamma_0(\mathscr{Z})$ coercive with $\mathrm{dom}(\Psi) = \mathscr{Z}$, $B \in \mathscr{B}\left(\mathscr{Z}, \widetilde{\mathscr{Z}}\right)$ and $(A, \mathfrak{L}, \mu) \in \mathscr{B}(\mathscr{X}, \mathscr{Y}) \times \mathscr{B}(\mathscr{X}, \mathscr{Z}) \times \mathbb{R}_+$. Then:

(a) GME penalty function $\Psi_B : \mathscr{Z} \to (-\infty, \infty]$ is defined as

$$\Psi_B(\cdot) := \Psi(\cdot) - \min_{v\in\mathscr{Z}} \left[\Psi(v) + \frac{1}{2}\|B(\cdot - v)\|_{\widetilde{\mathscr{Z}}}^2 \right]. \tag{2.8}$$

(b) Linearly involved Generalized-Moreau-Enhanced (LiGME) penalty is defined as $\Psi_B \circ \mathfrak{L} : \mathscr{X} \to (-\infty, \infty]$.

(c) LiGME model is defined as the minimization of

$$J_{\Psi_B \circ \mathfrak{L}} : \mathscr{X} \to (-\infty, \infty] : x \mapsto \frac{1}{2} \|Ax - y\|_{\mathscr{Y}}^2 + \mu \Psi_B \circ \mathfrak{L}(x). \tag{2.9}$$

Remark 2.1 (LiGME model can deal with multiple penalties [33, Example 3])**.** We can express multiple LiGME penalties $\left(\Psi^{\langle i \rangle}\right)_{B^{\langle i \rangle}} \circ \mathfrak{L}_i$ $(i = 1, 2, \cdots, p)$ as the single LiGME penalty $\Psi_B \circ \mathfrak{L}$. Let $\mathscr{Z}_i$ and $\widetilde{\mathscr{Z}}_i$ $(i = 1, 2, \cdots, p)$ be real Hilbert spaces, $\Psi^{\langle i \rangle} \in \Gamma_0(\mathscr{Z}_i)$ be coercive functions with $\mathrm{dom}\left(\Psi^{\langle i \rangle}\right) = \mathscr{Z}_i$, $\mu_i \in \mathbb{R}_{++}$, $B^{\langle i \rangle} \in \mathscr{B}\left(\mathscr{Z}_i, \widetilde{\mathscr{Z}}_i\right)$ and $\mathfrak{L}_i \in \mathscr{B}(\mathscr{X}_i, \mathscr{Z}_i)$. Then, by setting new real Hilbert spaces $\mathscr{Z} := \times_{i=1}^p \mathscr{Z}_i$ and $\widetilde{\mathscr{Z}} := \times_{i=1}^p \widetilde{\mathscr{Z}}_i$, a new function $\Psi := \bigoplus_{i=1}^p \mu_i \Psi^{\langle i \rangle}$ and new linear operators $B : \mathscr{Z} \to \widetilde{\mathscr{Z}} : (z_1, \cdots, z_p) \mapsto (\sqrt{\mu_1} B^{\langle 1 \rangle} z_1, \cdots, \sqrt{\mu_p} B^{\langle p \rangle} z_p)$ and $\mathfrak{L} : \mathscr{X} \to \mathscr{Z} : x \mapsto (\mathfrak{L}_1 x, \mathfrak{L}_2 x, \cdots, \mathfrak{L}_p x)$, we have

$$\Psi_B \circ \mathfrak{L} = \sum_{i=1}^p \mu_i {\Psi^{\langle i \rangle}}_{B^{\langle i \rangle}} \circ \mathfrak{L}_i.$$

The following fact shows the overall convexity condition for the LiGME model.

Fact 2.4 (Overall convexity condition for the LiGME model [33, Proposition 1])**.** The GME penalty defined in Definition 2.1 has the following properties:

(a) $\Psi_B \circ \mathfrak{L}(x) = \Psi(\mathfrak{L}x) - \left[\Psi(0_{\mathscr{Z}}) + \frac{1}{2} \|B\mathfrak{L}x\|_{\widetilde{\mathscr{Z}}}^2\right]$ if and only if $B^* B \mathfrak{L} x \in \mathrm{argmin}\,(\Psi^*)$.

(b) Let $(A, \mathfrak{L}, \mu) \in \mathscr{B}(\mathscr{X}, \mathscr{Y}) \times \mathscr{B}(\mathscr{X}, \mathscr{Z}) \times \mathbb{R}_{++}$. Then, for the three conditions (C_1) $A^*A - \mu \mathfrak{L}^* B^* B \mathfrak{L} \succeq \mathrm{O}_{\mathscr{X}}$, (C_2) $J_{\Psi_B \circ \mathfrak{L}} \in \Gamma_0(\mathscr{X})$ for any $y \in \mathscr{Y}$ and (C_3) $J^{(0)}_{\Psi_B \circ \mathfrak{L}}(x) := \frac{1}{2} \|A \cdot\|_{\mathscr{Y}}^2 + \mu \Psi_B \circ \mathfrak{L} \in \Gamma_0(\mathscr{X})$, the relation $(C_1) \implies (C_2) \iff (C_3)$ holds.

A proximal splitting type algorithm was also proposed in [33, Theorem 1] for the LiGME model under the overall convexity condition (C_1) in Fact 2.4.

3. Constrained LiGME Model

3.1. Constrained LiGME model and its overall convexity condition.

We start with simple reformulations of (1.6).

Problem 3.1 (cLiGME model)**.** Let $(\mathscr{X}, \langle \cdot, \cdot \rangle_{\mathscr{X}}, \|\cdot\|_{\mathscr{X}})$, $(\mathscr{Y}, \langle \cdot, \cdot \rangle_{\mathscr{Y}}, \|\cdot\|_{\mathscr{Y}})$, $(\mathscr{Z}, \langle \cdot, \cdot \rangle_{\mathscr{Z}}, \|\cdot\|_{\mathscr{Z}})$, $(\mathfrak{Z}, \langle \cdot, \cdot \rangle_{\mathfrak{Z}}, \|\cdot\|_{\mathfrak{Z}})$ and $\left(\widetilde{\mathscr{Z}}, \langle \cdot, \cdot \rangle_{\widetilde{\mathscr{Z}}}, \|\cdot\|_{\widetilde{\mathscr{Z}}}\right)$ be finite dimensional real Hilbert spaces and $\mathbf{C} (\subset \mathfrak{Z})$ be a nonempty closed convex set. Let $\Psi \in \Gamma_0(\mathscr{Z})$ be coercive with $\mathrm{dom}\,\Psi = \mathscr{Z}$. Let $(A, B, \mathfrak{L}, \mathfrak{C}, \mu) \in \mathscr{B}(\mathscr{X}, \mathscr{Y}) \times \mathscr{B}\left(\mathscr{Z}, \widetilde{\mathscr{Z}}\right) \times \mathscr{B}(\mathscr{X}, \mathscr{Z}) \times \mathscr{B}(\mathscr{X}, \mathfrak{Z}) \times \mathbb{R}_{++}$ satisfy $\mathbf{C} \cap \mathrm{ran}\,\mathfrak{C} \neq \emptyset$. Consider a constrained LiGME (cLiGME) model:

$$\text{find } x^* \in \mathscr{S} := \arg \min_{\mathfrak{C}x \in \mathbf{C}} J_{\Psi_B \circ \mathfrak{L}}(x) \text{ (see (2.9))} \tag{3.1}$$

Remark 3.1 (Model (3.1) is equivalent to model (1.6))**.** Clearly, model (3.1) can be seen as a specialization of the model (1.6) with $|I| = 1$. Conversely, by redefining a new real Hilbert space $\mathfrak{Z} := \times_{i \in I} \mathfrak{Z}_i$, a new linear operator $\mathfrak{C} : \mathscr{X} \to \mathfrak{Z} : x \mapsto (\mathfrak{C}_i x)_{i \in I}$ and a new closed convex set $\mathbf{C} := \times_{i \in I} C_i \subset \mathfrak{Z}$ with $(\mathfrak{Z}_i, \mathfrak{C}_i, C_i)$ $(i \in I)$ in (1.6), we confirm

$$\mathfrak{C}x \in \mathbf{C} \iff (\forall i \in I)\ \mathfrak{C}_i x \in C_i.$$

The following proposition shows that model (3.1) can be reformulated as a constraint-free model (3.3).

Proposition 3.1. *Let* $\mathscr{Z}_c := \mathscr{Z} \times \mathfrak{Z}$, $\widetilde{\mathscr{Z}}_c := \widetilde{\mathscr{Z}} \times \mathfrak{Z}$, $\mathfrak{L}_c : \mathscr{X} \to \mathscr{Z}_c : x \mapsto (\mathfrak{L}(x), \mathfrak{C}(x))$, $\boldsymbol{\Psi} := \Psi \oplus \iota_{\mathbf{C}} : \mathscr{Z}_c \to (-\infty, \infty]$ *and* $B_c := B \oplus \mathrm{O}_{\mathfrak{Z}}$, *where* $\iota_{\mathbf{C}} \in \Gamma_0(\mathfrak{Z})$ *is the indicator function*[1] *of* $\mathbf{C} \subset \mathfrak{Z}$*:*

$$\iota_{\mathbf{C}}(x) := \begin{cases} 0 & (x \in \mathbf{C}) \\ \infty & (otherwise). \end{cases}$$

Let

$$\boldsymbol{\Psi}_{B_c}(\cdot) := \boldsymbol{\Psi}(\cdot) - \inf_{v \in \mathscr{Z}_c} \left[\boldsymbol{\Psi}(v) + \frac{1}{2} \|B_c(\cdot - v)\|^2_{\widetilde{\mathscr{Z}}_c} \right] \tag{3.2}$$

and consider a constraint-free optimization problem:

$$\underset{x \in \mathscr{X}}{\text{minimize}} \ J_{\boldsymbol{\Psi}_{B_c} \circ \mathfrak{L}_c}(x) := \frac{1}{2} \|y - Ax\|^2_{\mathscr{Y}} + \mu \boldsymbol{\Psi}_{B_c} \circ \mathfrak{L}_c(x). \tag{3.3}$$

Then

(a)

$$\boldsymbol{\Psi} \in \Gamma_0(\mathscr{Z}_c).$$

(b)

$$\boldsymbol{\Psi}_{B_c} = \Psi_B \oplus \iota_{\mathbf{C}}. \tag{3.4}$$

(c)

$$\underset{\mathfrak{C}x \in \mathbf{C}}{\text{argmin}} \ J_{\Psi_B \circ \mathfrak{L}}(x) = \underset{x \in \mathscr{X}}{\text{argmin}} \ J_{\boldsymbol{\Psi}_{B_c} \circ \mathfrak{L}_c}(x).$$

Proof. See Appendix A. □

Remark 3.2 (Model (3.3) can deal with multiple LiGME penalties)**.** In a way similar to Remark 2.1, cLiGME model can deal with multiple penalties. For real Hilbert spaces $\mathscr{Z}_i$ and $\widetilde{\mathscr{Z}}_i$ $(i = 1, 2, \cdots, p)$, multiple penalties $\Psi^{\langle i \rangle} \in \Gamma_0(\mathscr{Z}_i)$ with $\mathrm{dom}(\Psi^{\langle i \rangle}) = \mathscr{Z}_i$, $\mu_i \in \mathbb{R}_{++}$, $B^{\langle i \rangle} \in \mathscr{B}\left(\mathscr{Z}_i, \widetilde{\mathscr{Z}}_i\right)$, and $\mathfrak{L}_i \in \mathscr{B}\left(\mathscr{X}, \widetilde{\mathscr{Z}}_i\right)$, define $(\mathscr{Z}, \widetilde{\mathscr{Z}}, \Psi, B, \mathfrak{L})$ as in Remark 2.1. Then, by (3.4), we have

$$\boldsymbol{\Psi}_{B_c} \circ \mathfrak{L}_c = \left(\sum_{i=1}^{p} \mu_i \Psi^{\langle i \rangle}{}_{B^{\langle i \rangle}} \circ \mathfrak{L}_i \right) + (\iota_{\mathbf{C}} \circ \mathfrak{C}).$$

The following proposition presents the overall convexity condition for the model (3.3).

[1]Let $C \subset \mathscr{H}$ be a nonempty closed convex set. For every $\gamma \in \mathbb{R}_{++}$, $\mathrm{Prox}_{\gamma \iota_C}$ coincides with the projection onto C, where the projection onto a closed convex set $C \subset \mathscr{H}$ is defined as

$$P_C : \mathscr{H} \to \mathscr{H} : x \mapsto \underset{y \in C}{\text{argmin}} \ \|x - y\|_{\mathscr{H}}.$$

.

Proposition 3.2 (Overall convexity condition for model (3.3))**.** *Let* $(A,\mathfrak{L},\mu)\in\mathscr{B}(\mathscr{X},\mathscr{Y})\times\mathscr{B}(\mathscr{X},\mathscr{Z})\times\mathbb{R}_{++}$. *For three conditions* (C_1) $A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L}\succeq \mathrm{O}_{\mathscr{X}}$, (C_2) $J_{\mathbf{\Psi}_{B_c}\circ\mathfrak{L}_c}\in\Gamma_0(\mathscr{X})$ *for any* $y\in\mathscr{Y}$, (C_3) $J^{(0)}_{\mathbf{\Psi}_{B_c}\circ\mathfrak{L}_c}:=\frac{1}{2}\|A\cdot\|^2_{\mathscr{Y}}+\mu\mathbf{\Psi}_{B_c}\circ\mathfrak{L}_c\in\Gamma_0(\mathscr{X})$, *the relation* $(C_1)\implies(C_2)\iff(C_3)$ *holds.*

Proof. See Appendix B. □

3.2. **A proximal splitting algorithm for the cLiGME model.**

We present a proximal splitting algorithm (Algorithm 1) of guaranteed convergence to a global minimizer for model (1.6), or equivalent model (3.3), under the following assumption.

Assumption 3.1. Suppose that $\Psi\in\Gamma_0(\mathscr{Z})$ satisfies even symmetry $\Psi\circ(-\mathrm{Id})=\Psi$ and is proximable (i.e., $\mathrm{Prox}_{\gamma\Psi}$ is available as a computable operator for every $\gamma\in\mathbb{R}_{++}$) and assume that $\mathbf{C}\subset\mathfrak{Z}$ is a closed convex subset onto which the metric projection $P_{\mathbf{C}}:\mathfrak{Z}\mapsto\mathbf{C}$ is computable. Suppose also that $(A,\mathfrak{L},\mathfrak{C},B_c,y,\mu)\in\mathscr{B}(\mathscr{X},\mathscr{Y})\times\mathscr{B}(\mathscr{X},\mathscr{Z})\times\mathscr{B}(\mathscr{X},\mathfrak{Z})\times\mathscr{B}(\mathscr{Z}_c,\widetilde{\mathscr{Z}}_c)\times\mathscr{Y}\times\mathbb{R}_{++}$ satisfies $A^*A-\mu\mathfrak{L}B^*B\mathfrak{L}\succeq\mathrm{O}_{\mathscr{X}}$ and $0_{\mathfrak{Z}}\in\mathrm{ri}(\mathbf{C}-\mathrm{ran}\,\mathfrak{C})$.

In the next theorem, (a) and (b) show that the set of all global optimal solutions of model (3.3) can be expressed in terms of the fixed-point set of an averaged nonexpansive operator in a certain Hilbert space, and (c) shows an iterative algorithm for model (3.3).

Theorem 3.1 (averaged nonexpansive operator T_{cLiGME} and its fixed point approximation)**.** *Consider Problem 3.1 under Assumption 3.1 with notations in Proposition 3.1. Define a real Hilbert space* $(\mathscr{H}:=\mathscr{X}\times\mathscr{Z}\times\mathscr{Z}_c,\langle\cdot,\cdot\rangle_{\mathscr{H}},\|\cdot\|_{\mathscr{H}})$ *as a product space. Define* $T_{\mathrm{cLiGME}}:\mathscr{H}\to\mathscr{H}:(x,v,w)\mapsto(\xi,\zeta,\eta)$ *with* $(\sigma,\tau)\in\mathbb{R}_{++}\times\mathbb{R}_{++}$ *by*

$$\xi:=\left[\mathrm{Id}-\frac{1}{\sigma}(A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})\right]x-\frac{\mu}{\sigma}\mathfrak{L}^*B^*Bv-\frac{\mu}{\sigma}\mathfrak{L}_c^*w+\frac{1}{\sigma}A^*y$$

$$\zeta:=\mathrm{Prox}_{\frac{\mu}{\tau}\Psi}\left[\frac{2\mu}{\tau}B^*B\mathfrak{L}\xi-\frac{\mu}{\tau}B^*B\mathfrak{L}x+\left(\mathrm{Id}-\frac{\mu}{\tau}B^*B\right)v\right]$$

$$\eta:=\mathrm{Prox}_{\mathbf{\Psi}^*}(2\mathfrak{L}_c\xi-\mathfrak{L}_cx+w),$$

where $\mathrm{Prox}_{\mathbf{\Psi}^*}(w_1,w_2)=(w_1-\mathrm{Prox}_{\Psi}(w_1),w_2-P_{\mathbf{C}}(w_2))$. *Then*

(a) *The solution set* $\mathscr{S}$ *of Problem 3.1 can be expressed as*

$$\mathscr{S}=\Xi(\mathrm{Fix}(T_{\mathrm{cLiGME}})):=\{\Xi(x,v,w)\,|\,(x,v,w)\in\mathrm{Fix}(T_{\mathrm{cLiGME}})\},$$

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

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

$$\begin{cases}\sigma\,\mathrm{Id}-\frac{\kappa}{2}A^*A-\mu\mathfrak{L}_c^*\mathfrak{L}_c\succ\mathrm{O}_{\mathscr{X}}\\ \tau\geq\left(\frac{\kappa}{2}+\frac{2}{\kappa}\right)\mu\,\|B\|^2_{\mathrm{op}}.\end{cases}\tag{3.5}$$

Then

$$\mathfrak{P}:=\begin{bmatrix}\sigma\,\mathrm{Id}&-\mu\mathfrak{L}^*B^*B&-\mu\mathfrak{L}_c^*\\-\mu B^*B\mathfrak{L}&\tau\,\mathrm{Id}&\mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\-\mu\mathfrak{L}_c&\mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)}&\mu\,\mathrm{Id}\end{bmatrix}\succ\mathrm{O}_{\mathscr{H}}\tag{3.6}$$

holds and T_{cLiGME} *is* $\frac{\kappa}{2\kappa-1}$*-averaged nonexpansive in the real Hilbert space* $(\mathscr{H},\langle\cdot,\cdot\rangle_{\mathfrak{P}},\|\cdot\|_{\mathfrak{P}})$.

(c) *Assume that* $(\sigma,\tau,\kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1,\infty)$ *satisfies* (3.5). *Then, for any initial point* $(x_0,v_0,w_0) \in \mathscr{H}$*, the sequence* $(x_k,v_k,w_k)_{k\in\mathbb{N}}$ *generated by*

$$(\forall k \in \mathbb{N})\ (x_{k+1},v_{k+1},w_{k+1}) = T_{\mathrm{cLiGME}}(x_k,v_k,w_k)$$

converges weakly to a point $(x^\diamond,v^\diamond,w^\diamond) \in \mathrm{Fix}(T_{\mathrm{cLiGME}})$ *and*

$$\lim_{k\to\infty} x_k = x^\diamond \in \mathscr{S}.$$

Proof. See Appendix C. □

Remark 3.3. For example, choose any $\kappa > 1$. By letting $\sigma := \left\|\frac{\kappa}{2}A^*A+\mu\mathfrak{L}_c^*\mathfrak{L}_c\right\|_{\mathrm{op}}+(\kappa-1)$ and $\tau := \mu\left(\frac{\kappa}{2}+\frac{2}{\kappa}\right)\|B\|_{\mathrm{op}}^2+(\kappa-1)$, it is not hard to verify that (σ,τ,κ) satisfies (3.5).

The proposed algorithm based on Theorem 3.1 is presented in Algorithm 1. Algorithm 1 can be seen as a generalization of [33, Algorithm 1] (Note: See [33, Remark 4] for comparisons between [33, Algorithm 1] and possibly related proximal splitting algorithms).

Algorithm 1 for cLiGME model

Choose $(x_0,v_0,w_0) \in \mathscr{H}$.
Let $(\sigma,\tau,\kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1,\infty)$ satisfying (3.5).
Define $\mathfrak{P}$ as in (3.6).
$k \leftarrow 0$.
do

$x_{k+1} \leftarrow \left[\mathrm{Id}-\frac{1}{\sigma}(A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})\right]x_k - \frac{\mu}{\sigma}\mathfrak{L}^*B^*Bv_k - \frac{\mu}{\sigma}\mathfrak{L}_c^*w_k + \frac{1}{\sigma}A^*y$

$v_{k+1} \leftarrow \mathrm{Prox}_{\frac{\mu}{\tau}\Psi}\left[\frac{2\mu}{\tau}B^*B\mathfrak{L}x_{k+1} - \frac{\mu}{\tau}B^*B\mathfrak{L}x_k + \left(\mathrm{Id}-\frac{\mu}{\tau}B^*B\right)v_k\right]$

$w_{k+1} \leftarrow \mathrm{Prox}_{\boldsymbol{\Psi}^*}\left(2\mathfrak{L}_cx_{k+1} - \mathfrak{L}_cx_k + w_k\right)$

$k \leftarrow k+1$

while $\|(x_k,v_k,w_k)-(x_{k-1},v_{k-1},w_{k-1})\|_{\mathfrak{P}}$ is not sufficiently small
return x_k

3.3. How to choose B to archive the overall convexity of $J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}$.

We present a choice of B to guarantee the overall convexity of $J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}$.

Proposition 3.3 (A design of B to ensure the overall convexity condition in Proposition 3.2)**.** *In Problem 3.1, let* $(\mathscr{X},\mathscr{Y},\mathscr{Z}) = (\mathbb{R}^n,\mathbb{R}^m,\mathbb{R}^l)$*,* $(A,\mathfrak{L},\mu) \in \mathbb{R}^{m\times n}\times\mathbb{R}^{l\times n}\times\mathbb{R}_{++}$ *and* $\mathrm{rank}(\mathfrak{L}) = l$*. Choose a nonsingular* $\widetilde{\mathfrak{L}} \in \mathbb{R}^{n\times n}$ *satisfying* $\begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_l\end{bmatrix}\widetilde{\mathfrak{L}} = \mathfrak{L}$*. Then*

$$B = B_\theta := \sqrt{\theta/\mu}\,\Lambda^{1/2}U^\mathsf{T} \in \mathbb{R}^{l\times l}, \theta \in [0,1] \tag{3.7}$$

ensures $J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c} \in \Gamma_0(\mathbb{R}^n)$*, where*

$$\begin{bmatrix}\widetilde{A}_1 & \widetilde{A}_2\end{bmatrix} := A\widetilde{\mathfrak{L}}^{-1} \tag{3.8}$$

and $U\Lambda U^\mathsf{T} := \widetilde{A}_2^\mathsf{T}\widetilde{A}_2 - \widetilde{A}_2^\mathsf{T}\widetilde{A}_1\left(\widetilde{A}_1^\mathsf{T}\widetilde{A}_1\right)^\dagger\widetilde{A}_1^\mathsf{T}\widetilde{A}_2$ *is an eigendecomposition.*

Proof. The proof is similar to that of [33, Proposition 2] (for completeness, see Appendix D). □

The following corollary shows the choice of $B^{\langle i\rangle}$ $(i=1,2,\cdots,p)$ in Remark 3.2 to ensure the condition for the overall convexity in Proposition 3.2. For self-containedness, we present the following corollary in the context of cLiGME model (3.3) although it is essentially same as [33, Corollary 1] in the context of LiGME model.

Corollary 3.1 (A design of $B^{\langle i\rangle}$ to ensure the overall convexity condition in Proposition 3.2)**.** *In Remark 3.2, let* $(\mathscr{X},\mathscr{Y},\mathscr{Z}_i)=(\mathbb{R}^n,\mathbb{R}^m,\mathbb{R}^{l_i})$, $(A,\mathfrak{L}_i,\mu)\in\mathbb{R}^{m\times n}\times\mathbb{R}^{l_i\times n}\times\mathbb{R}_{++}$ *and* $\operatorname{rank}(\mathfrak{L}_i)=l_i$ $(i=1,2,\cdots,p)$. *For each* $i=1,2,\cdots,p$, *choose nonsingular* $\widetilde{\mathfrak{L}}_i\in\mathbb{R}^{n\times n}$ *satisfying* $\left[\mathrm{O}_{l_i\times(n-l_i)}\;\;\mathrm{I}_{l_i}\right]\widetilde{\mathfrak{L}}_i=\mathfrak{L}_i$ *and* $\omega_i\in\mathbb{R}_{++}$ *satisfying* $\sum_{i=1}^p\omega_i=1$. *For each* $i=1,2,\cdots,p$, *apply Proposition 3.3 to* $\left(\sqrt{\frac{\omega_i}{\mu}}A,\mathfrak{L}_i,\mu_i\right)$ *to obtain* $B_{\theta_i}^{\langle i\rangle}\in\mathbb{R}^{l_i\times l_i}$ *satisfying* $\left(\sqrt{\frac{\omega_i}{\mu}}A\right)^{\mathsf{T}}\left(\sqrt{\frac{\omega_i}{\mu}}A\right)-\mu_i\mathfrak{L}_i^{\mathsf{T}}\left(B_{\theta_i}^{\langle i\rangle}\right)^{\mathsf{T}}B_{\theta_i}^{\langle i\rangle}\mathfrak{L}_i\succeq\mathrm{O}_n$. *Then* $B_\theta:\times_{i=1}^p\mathbb{R}^{l_i}\to\times_{i=1}^p\mathbb{R}^{l_i}:(z_1,\cdots,z_p)\mapsto\left(\sqrt{\mu_1}B_{\theta_1}^{\langle 1\rangle}z_1,\cdots,\sqrt{\mu_p}B_{\theta_p}^{\langle p\rangle}z_p\right)$ *ensures* $J_{\Psi_{B_c}\circ\mathfrak{L}_c}\in\Gamma_0(\mathbb{R}^n)$. *This is verified by*

$$
\begin{aligned}
A^{\mathsf{T}}A-\mu\mathfrak{L}^{\mathsf{T}}B^{\mathsf{T}}B\mathfrak{L}&=A^{\mathsf{T}}A-\mu\sum_{i=1}^p\mu_i\mathfrak{L}_i^{\mathsf{T}}\left(B_{\theta_i}^{\langle i\rangle}\right)^{\mathsf{T}}B_{\theta_i}^{\langle i\rangle}\mathfrak{L}_i\\
&=\mu\sum_{i=1}^p\left(\frac{\omega_i}{\mu}A^{\mathsf{T}}A-\mu_i\mathfrak{L}_i^{\mathsf{T}}\left(B_{\theta_i}^{\langle i\rangle}\right)^{\mathsf{T}}B_{\theta_i}^{\langle i\rangle}\mathfrak{L}_i\right)\succeq\mathrm{O}_n.
\end{aligned}
$$

4. Numerical Experiment

We conducted a numerical experiment based on a scenario of image restoration for piecewise constant N-by-N image, which is the same as [33, Sec. 4.2] but with considering additionally multiple convex constraints. Set $(\mathscr{X},\mathscr{Y},\mathscr{Z}_1,\mathscr{Z}_2)=\left(\mathbb{R}^{N^2},\mathbb{R}^{N^2},\mathbb{R}^{N(N-1)},\mathbb{R}^{N(N-1)}\right)$, $N=16$, $\Psi^{\langle 1\rangle}=\Psi^{\langle 2\rangle}=\|\cdot\|_1$, $\mu_1=\mu_2=1$ and $\mathfrak{L}=\bar{D}:=[D_{\mathrm{H}}^{\mathsf{T}},D_{\mathrm{V}}^{\mathsf{T}}]^{\mathsf{T}}$, where D_V is the vertical differential matrix:

$$
D_{\mathrm{V}}:=\begin{bmatrix}D&&&\\&D&&\\&&\ddots&\\&&&D\end{bmatrix}\in\mathbb{R}^{N(N-1)\times N^2}
$$

with

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

and D_{H} is the horizontal matrix:

$$
D_{\mathrm{H}}:=\begin{bmatrix}-1&0_{N-1}^{\mathsf{T}}&1&&\\&\ddots&\ddots&\ddots&\\&&-1&0_{N-1}^{\mathsf{T}}&1\end{bmatrix}\in\mathbb{R}^{N(N-1)\times N^2}.
$$

The blur matrix is designed by

$$
A:=\bar{A}\otimes\bar{A}\in\mathbb{R}^{N^2\times N^2},
$$

where $\otimes$ denotes the Kronecker product and $\bar{A} \in \mathbb{R}^{N\times N}$ is defined by

$$\bar{A}_{i,j} := \begin{cases} \frac{1}{\sqrt{1.62\pi}} \exp\left(-\frac{|i-j|^2}{1.62}\right) & |i-j| < 6 \\ 0 & (otherwise). \end{cases}$$

The observation vector $y \in \mathbb{R}^{N^2}$ is assumed to satisfy the linear regression model:

$$y = Ax^\star + \varepsilon,$$

where $x^\star \in \mathbb{R}^{N^2}$ is given by the vectorization of a piecewise constant image which is displayed in Figure 3(A) and $\varepsilon \in \mathbb{R}^{N^2}$ is additive white Gaussian noise. The signal-to-noise ratio (SNR) is set by

$$10\log_{10} \frac{\|x^\star\|_{\mathscr{X}}^2}{\|\varepsilon\|_{\mathscr{Y}}^2} = 20\text{dB}.$$

As available a priori knowledge on $x^\star$, we use (i) every entry in $x^\star$ belongs to $[0.25, 0.75]$, (ii) every entry in the background I_{back} is same but unknown valued (similar a priori knowledge is found, e.g., in a blind deconvolution [44]). By letting $\mathscr{X} =: \mathfrak{Z}_1 =: \mathfrak{Z}_2$, $\mathfrak{Z} := \mathfrak{Z}_1 \times \mathfrak{Z}_2$, $\mathfrak{C}_1 := \mathfrak{C}_2 := \mathrm{I}_{N^2}$, $\mathfrak{C} := \left[\mathfrak{C}_1^\mathsf{T}, \mathfrak{C}_2^\mathsf{T}\right]^\mathsf{T}$,

$$C_1 := \left\{ x \in \mathbb{R}^{N^2} = \mathfrak{Z}_1 \,\middle|\, (i = 1, 2 \cdots, N^2)\ 0.25 \leq x_i \leq 0.75 \right\},$$
$$C_2 := \left\{ \mathrm{vec}(X) \in \mathbb{R}^{N^2} = \mathfrak{Z}_2 \,\middle|\, X \in \mathbb{R}^{N\times N},\ (\forall (i_1, j_1) \in I_{\text{back}})(\forall (i_2, j_2) \in I_{\text{back}})\ X_{i_1,j_1} = X_{i_2,j_2} \right\},$$

where

$$I_{\text{back}} := [\{1,2,3,14,15,16\} \times \{1,2,\cdots,16\}] \cup [\{1,2,\cdots,16\} \times \{1,2,3,14,15,16\}],$$

we consider the following four cases:
(♣) $\mathfrak{C}x^\star \in \mathbf{C}_\clubsuit := \mathfrak{Z} := \mathbb{R}^{N^2} \times \mathbb{R}^{N^2}$
(♢) $\mathfrak{C}x^\star \in \mathbf{C}_\diamondsuit := C_1 \times \mathbb{R}^{N^2}$
(♡) $\mathfrak{C}x^\star \in \mathbf{C}_\heartsuit := \mathbb{R}^{N^2} \times C_2$
(♠) $\mathfrak{C}x^\star \in \mathbf{C}_\spadesuit := C_1 \times C_2$.
For each case ($s = \clubsuit, \diamondsuit, \heartsuit, \spadesuit$), we compared the following two models as instances of Problem 3.1:

(a) the anisotropic TV with constraint $\mathfrak{C}x \in \mathbf{C}_s$:

$$\underset{\mathfrak{C}x\in\mathbf{C}_s}{\text{minimize}}\ J_{(\|\cdot\|_1)_{\mathrm{O}_{2N(N-1)}}\circ\bar{D}}(x) (= J_{\|\cdot\|_1\circ\bar{D}}(x)), \tag{4.1}$$

where $\mathrm{O}_{2N(N-1)} \in \mathbb{R}^{2N(N-1)\times 2N(N-1)}$ is the zero matrix,

(b) cLiGME for $(\Psi, \mathfrak{L}) = (\|\cdot\|_1, \bar{D})$ with constraint $\mathfrak{C}x \in \mathbf{C}_s$:

$$\underset{\mathfrak{C}x\in\mathbf{C}_s}{\text{minimize}}\ J_{(\|\cdot\|_1)_{B_\theta}\circ\bar{D}}(x), \tag{4.2}$$

The vectorization of a matrix is the mapping:

$$\mathrm{vec} : \mathbb{R}^{m\times n} \to \mathbb{R}^{mn} : A \mapsto \left[a_1^\mathsf{T}, a_2^\mathsf{T} \cdots a_n^\mathsf{T}\right]^\mathsf{T},$$

where a_i is the i-th column vector of A for each $i \in \{1, 2, \cdots, n\}$.

where (i) the block diagonal matrix expression $B_\theta = \begin{bmatrix} B^{\langle 1\rangle}_{\theta_1} & \mathrm{O}_{N(N-1)} \\ \mathrm{O}_{N(N-1)} & B^{\langle 2\rangle}_{\theta_2} \end{bmatrix} \in \mathbb{R}^{2N(N-1)\times 2N(N-1)}$ is given from Corollary 3.1 with $\theta_1 = \theta_2 = 0.99$ and $\omega_1 = \omega_2 = \frac{1}{2}$, and (ii) $\widetilde{\mathfrak{L}}_1$ and $\widetilde{\mathfrak{L}}_2$ with $p = 2$ in Corollary 3.1 are given respectively by

$$\widetilde{\mathfrak{L}}_1 = \widetilde{D}_{\mathrm{V}} := \begin{bmatrix} E \\ D_{\mathrm{V}} \end{bmatrix} \in \mathbb{R}^{N^2\times N^2},\ \widetilde{\mathfrak{L}}_2 = \widetilde{D}_{\mathrm{H}} := \begin{bmatrix} [\mathrm{I}_N \quad \mathrm{O}_{N\times N(N-1)}] \\ D_{\mathrm{H}} \end{bmatrix} \in \mathbb{R}^{N^2\times N^2},$$

with $E = [E_{i,j}] \in \mathbb{R}^{N\times N^2}$ and

$$E_{i,j} = \begin{cases} 1 & ((i-1)N+1 = j) \\ 0 & (otherwise) \end{cases}$$

(Note: (i) $x^\star$ is not guaranteed to be a minimizer of the models (4.1) and (4.2), (ii) The case (♣) reproduces the experiment in [33, Sec. 4.2], (iii) For applications of LiGME model, as a special case of cLiGME model, to low-rankness promoting scenario, see [33, Sec 4.3, 4.4]).

For all cases, we applied Algorithm 1 to $(x_0, v_0, w_0) = (0_{\mathscr{X}}, 0_{\mathscr{Z}}, 0_{\mathscr{Z}_c})$ with $\kappa = 1.001$ and (σ, τ) in Remark 3.3. $\mathrm{Prox}_{\gamma\|\cdot\|_1}$ is available as the soft-thresholding

$$[\mathrm{Prox}_{\gamma\|\cdot\|_1}]_i : \mathbb{R}^{2N(N-1)} \to \mathbb{R} : (z_1, z_2, \cdots, z_{2N(N-1)})^\mathsf{T} \mapsto \begin{cases} 0 & (|z_i| \le \gamma) \\ (|z_i| - \gamma)\frac{z_i}{|z_i|} & (otherwise). \end{cases}$$

Projections onto C_1 and C_2 are realized by

$$[P_{C_1}(x)]_i = \begin{cases} 0.75 & (0.75 < x_i) \\ x_i & (0.25 \le x_i \le 0.75) \\ 0.25 & (x_i < 0.25), \end{cases}$$

$$[P_{C_2}(x)]_i = \begin{cases} \frac{1}{|I_{\mathrm{back}}|}\sum_{\Upsilon^{-1}(j)\in I_{\mathrm{back}}} x_j & (\Upsilon^{-1}(i) \in I_{\mathrm{back}}) \\ x_i & (otherwise), \end{cases}$$

where $\Upsilon(i, j) := i + j(N-1)$ (See Appendix E).

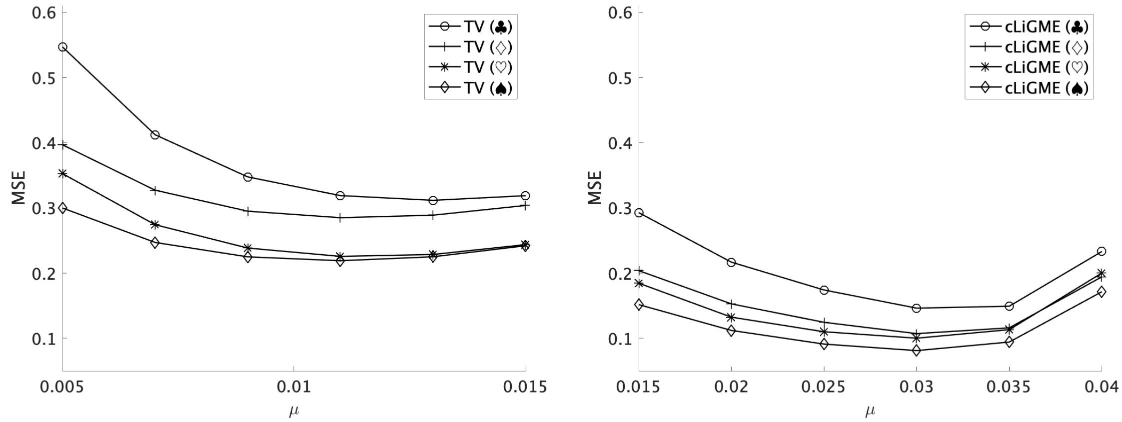
Figure 1 shows the dependency of recovery performance on the parameter μ in Problem 3.1. The recovery performance is measured by the mean squared error (MSE) defined by the average of

$$\text{squared error (SE): } \|x_k - x^\star\|^2_{\mathscr{X}}$$

over 100 independent realizations of the additive white Gaussian noise. We see that constraints are effective to improve the estimation and cLiGME model in (4.2) can achieve better estimation than TV model in (4.1).

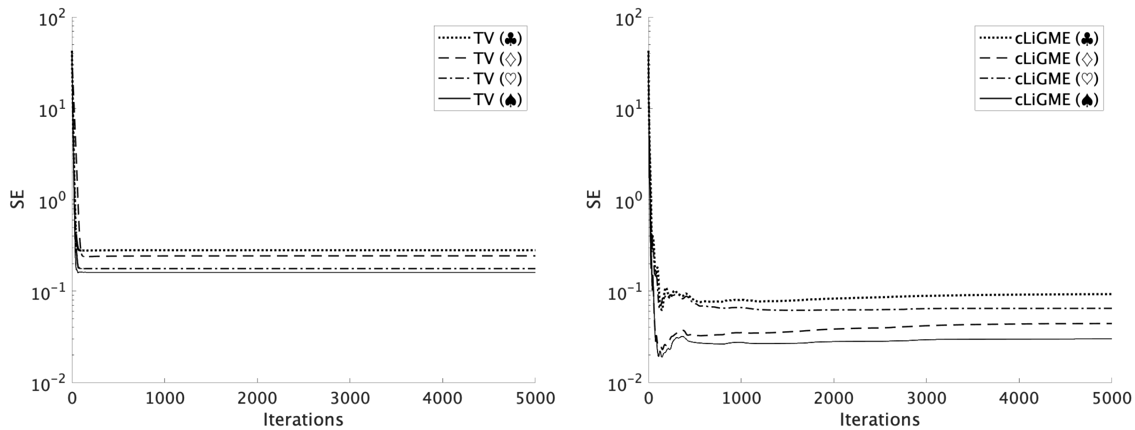
Figure 2 shows convergence performances observed over 5000 iterations. Common weights are used as $\mu = \mu_{\mathrm{TV}} := 0.013$ for (4.1) and $\mu = \mu_{\mathrm{cLiGME}} := 0.03$ for (4.2), where these values are best ones for the TV model and cLiGME model in the case of (♣). Even if we use C_1 and C_2, the model (4.1) cannot achieve the level 10^{-1} while the model (4.2) achieves this level only with C_1.

Figure 3 shows the original image, an observed image and recovered images by the models (4.1) and (4.2).



(A) (B)

FIGURE 1. MSE versus μ (around best μ) in Problem 3.1 just after 5000 iterations for (A) TV model in (4.1) and for (B) cLiGME model in (4.2)



(A) (B)

FIGURE 2. SE versus iterations for (A) TV model in (4.1) and for (B) cLiGME model in (4.2)

5. CONCLUSION

In this paper, we proposed the constrained LiGME (cLiGME) model, as an enhancement of LiGME model proposed in [33], for using closer nonconvex regularizers to naive discrete measures for desired properties than their convex envelopes in the constrained least square model without losing overall convexity. The proposed model covers multiple penalties and multiple convex constraint sets. Therefore, we can apply the proposed model to many problems in data science. We also proposed a proximal splitting algorithm for the proposed model under the overall convexity condition. The numerical experiment on a simple scenario shows the efficiency of the proposed cLiGME model and Algorithm 1 based on Theorem 3.1.

Acknowledgments

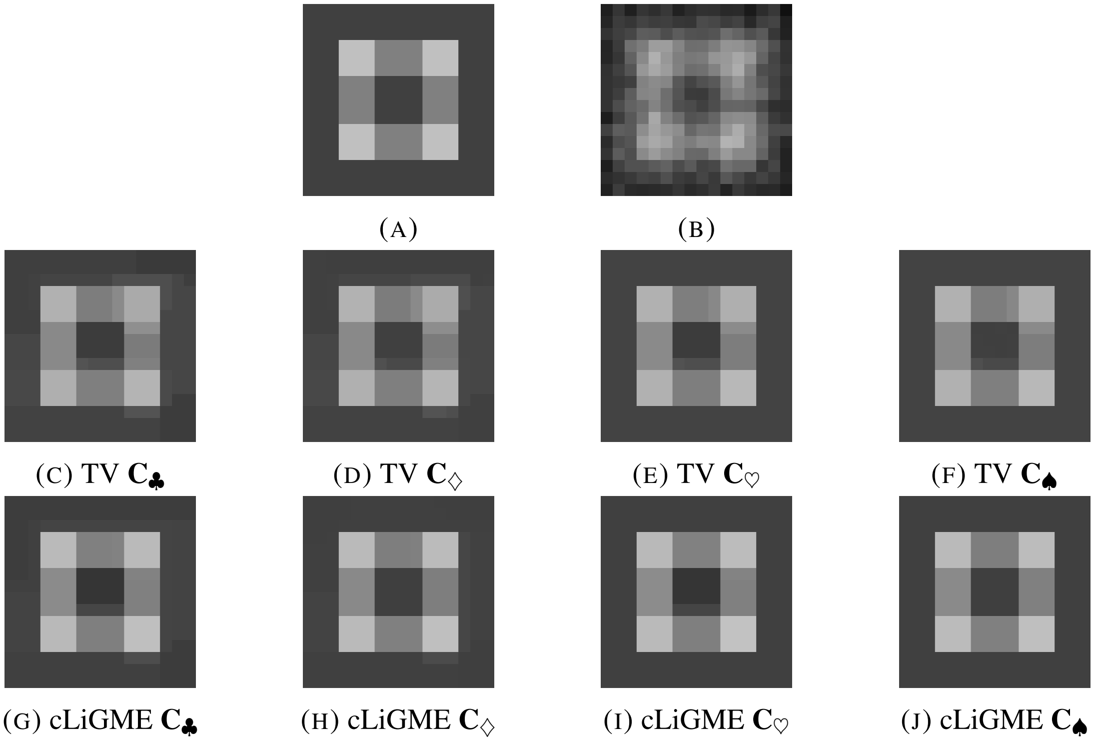


FIGURE 3. (A) original piecewise constant image of pixel values in $\{0.25, 0.5, 0.75\}$, (B) observed noisy blurred image, (C-J) recovered images. Each recovered image is obtained by Algorithm 1 at 5000 iteration and shown with the pixel value range $[0(\text{black}), 1(\text{white})]$.

This work was supported in part by JSPS Grants-in-Aid (19H04134) and by JST SICORP (JP-MJSC20C6).

REFERENCES

[1] A.E. Horel, Application of ridge analysis to regression problems, Chem. Eng. Progress 58 (1962), 54-59.
[2] A.E. Horel, R.W. Kennard, Ridge regression: Biased estimation for non-orthogonal problem, Technometrics 12 (1970), 55-67.
[3] A.N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Soviet. Math. Doctl 5 (1963), 1035-1038.
[4] A.N. Tikhonov, V.Y. Arsenin, Solutions of Ill-Posed Problems, Winston, 1977.
[5] J. Claerbout, F. Muir, Robust modeling with erratic data, Geophysics, 38 (1973), 826-844.
[6] F. Santosa, W.W. Symes, Linear inversion of band limited reflection seismograms, SIAM J. Sci. Comput. 7 (1986), 1307-1330.
[7] H. Taylor, S. Bank, J. McCoy, Deconvolution with the l_1-norm, Geo-physics, 44 (1979), 39-52.
[8] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, J. R. Stat. Soc. Ser. B. Stat. Methodol. 58 (1996), 267-288.
[9] E. J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (2006), 1207-1223.
[10] D.L. Donoho, Compressed sensing, IEEE Trans. Info. Theory, 52 (2006), 1289-1306.
[11] M. Elad, Sparse and Redundant Representations, Springer, New York, 2010.
[12] S. Theodoridis, Machine Learning: A Bayesian and Optimization Perspective, edition 2, Academic Press, 2020.

[13] B.K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (1995), 227-234.
[14] T.T. Nguyen, C. Soussen, J. Idier, E.H. Djermoune, NP-hardness of $\ell 0$ minimization problems: revision and extension to the non-negative setting, 2019 13th International conference on Sampling Theory and Applications (SampTA), pp. 1-4, 2019. doi: 10.1109/SampTA45681.2019.9030937.
[15] L.I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1992), 259-268.
[16] I. Daubechies, M. Defrise, C. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constrains, Commun. Pure Appl. Math. 57 (2004), 1413-1457.
[17] J.L. Starck, F. Murtagh, J. Fadili, Sparse Image and Signal Processing: Waveletes and Related Geometric Multiscale Analysis, edition 2, Cambidge University Press, 2015.
[18] A. Blake, A. Zisserman, Visual Reconstruction, MIT Press, 1987.
[19] M. Nikolova, Estimation of binary images by minimizing convex criteria, IEEE International Conference on Image Processing (ICIP), vol. 2, pp. 108-112, 1998. doi: 10.1109/ICIP.1998.723327.
[20] M. Nikolova, Markovian reconstruction using a GNC approach, IEEE Trans. Image Process. 8 (1999), 1204-1220.
[21] M. Nikolova, Energy minimization methods, In: O. Scherzer, (ed.) Handbook of Mathematical Methods in Imaging, pp. 157-204, Springer, 2015.
[22] İ. Bayram, On the convergence of the iterative shrinkage/thresholding algorithm with a weakly convex penalty, IEEE Trans. Signal Process. 64 (2016), 1597-1608.
[23] M. Carlsson, On convexification/optimization of functionals including an L2-misfit term, arXiv:1609.09378, 2016.
[24] Y. Ding, I. Selesnick, Artifact-free wavelet denoising: Non-convex sparse regularization, convex optimization, IEEE Signal Process. Lett. 22 (2015), 1364-1368,
[25] A. Lanza, S. Morigi, I. Selesnick, F. Sgallari, Nonconvex nonsmooth optimization via convex-nonconvex majorization-minimization, Numer. Math. 136 (2017), 343-381.
[26] A. Lanza, S. Morigi, F. Sgallari, Convex image denoising via non-convex regularization with parameter selection, J. Math. Imaging Vision, 56 (2016), 195-220.
[27] M. Malek-Mohammadi, C.R. Rojas, B. Wahlberg, A class of nonconvex penalties preserving overall convexity in optimization-based mean filtering, IEEE Trans. Signal Process. 64 (2016), 6650-6664.
[28] T. M̈ollenhoff, E. Strekalovskiy, M. Moeller, D. Cremers, The primaldual hybrid gradient method for semiconvex splittings, SIAM J. Imaging Sci. 8 (2015), 837-857.
[29] I. Selesnick, M. Farshchian, Sparse signal approximation via nonseparable regularization, IEEE Trans. Signal Process. 65 (2017), 2561-2575.
[30] E. Soubies, L. Blanc-Féraud, G. Aubert, A continuous exact L0 penalty (CEL0) for least squares regularized problem, SIAM J. Imaging Sci. 8 (2015), 1607-1639.
[31] I. Selesnick, Sparse regularization via convex analysis, IEEE Trans. Signal Process. 65 (2017), 4481-4494.
[32] C.-H. Zhang, Nearly unbiased variable selection under minimax concave penalty, Ann. Statist. 38 (2010), 894-942.
[33] J. Abe, M. Yamagishi, I. Yamada, Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition, Inverse Probl. 36 (2020), 035012.
[34] D.C. Youla, H. Webb, Image restoration by the method of convex projections: Part 1-theory, IEEE Trans. Med. Imag. 1 (1982), 81-94.
[35] H. Stark, ed., Image Recovery: Theory and Application, Academic Press, 1987.
[36] P.L. Combettes, Foundation of set theoretic estimation, Proc. IEEE 81 (1993), 182-208.
[37] Y. Censor, T. Elfving, N. Kopf, T. Bortfeld, The multiple-sets split feasibility problem and its applications for inverse problems, Inverse Probl. 21 (2005), 2071-2084.
[38] G.T. Herman, Image Reconstruction from Projections, Academic Press, 1980.
[39] M.I. Sezan, An overview of convex projections theory and its application to image recovery problems, Ultramicroscopy, 48 (1992), 55-67.
[40] W. Yata, M. Yamagishi, I. Yamada, A constrained linearly involved generalized Moreau enhanced model and its proximal splitting algorithm, IEEE International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-6, 2021.

[41] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, 2017,

[42] N. Ogura, I. Yamada, Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping, Numer. Funct. Anal. Optim. 23 (2002), 113-137.

[43] P.L. Combettes, I. Yamada, Compositions and convex combinations of averaged nonexpansive operators, J. Math. Anal. Appl. 425 (2015), 55-70.

[44] D. Kundur, D. Hatzinakos, A novel blind deconvolution scheme for image restoration using recursive filtering, IEEE Trans. Signal Process. 46 (1998), 375-390.

[45] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge Univ. Press, Cambridge, 2012.

[46] A. Arthur, Conditions for positive and nonnegative definiteness in terms of pseudoinverses, SIAM J. Appl. Math. 17 (1969), 434-440.

APPENDIX A. PROOF OF PROPOSITION 3.1

(a) $\Psi \in \Gamma_0(\mathscr{Z})$, $\iota_{\mathbf{C}} \in \Gamma_0(\mathfrak{Z})$ and Fact 2.1(a) yield $\boldsymbol{\Psi} \in \Gamma_0(\mathscr{Z}_c)$.

(b) For every $w = (w_1, w_2) \in \mathscr{Z}_c = \mathscr{Z} \times \mathfrak{Z}$ and $B_c = B \oplus \mathrm{O}_{\mathfrak{Z}} \in \mathscr{B}\left(\mathscr{Z}_c, \widetilde{\mathscr{Z}}_c\right)$, we have

$$\begin{aligned}
\boldsymbol{\Psi}_{B_c}(w) &= \boldsymbol{\Psi}(w) - \inf_{v \in \mathscr{Z}_c}\left[\boldsymbol{\Psi}(v) + \frac{1}{2}\|B_c(w-v)\|^2_{\widetilde{\mathscr{Z}}_c}\right] \\
&= \Psi(w_1) + \iota_{\mathbf{C}}(w_2) - \inf_{(v_1,v_2)\in\mathscr{Z}_c}\left[\Psi(v_1) + \iota_{\mathbf{C}}(v_2) + \|B_c((w_1,w_2)-(v_1,v_2))\|^2_{\widetilde{\mathscr{Z}}_c}\right] \\
&= \Psi(w_1) + \iota_{\mathbf{C}}(w_2) - \inf_{(v_1,v_2)\in\mathscr{Z}_c}\left[\Psi(v_1) + \iota_{\mathbf{C}}(v_2) + \|(B(w_1-v_1), \mathrm{O}_{\mathfrak{Z}}(w_2-v_2))\|^2_{\widetilde{\mathscr{Z}}_c}\right] \\
&= \Psi(w_1) + \iota_{\mathbf{C}}(w_2) - \inf_{v_1\in\mathscr{Z}}\left[\Psi(v_1) + \|(B(w_1-v_1)\|^2_{\widetilde{\mathscr{Z}}}\right] - \inf_{v_2\in\mathfrak{Z}} \iota_{\mathbf{C}}(v_2) \\
&= \Psi(w_1) - \min_{v_1\in\mathscr{Z}}\left[\Psi(v_1) + \|(B(w_1-v_1)\|^2_{\widetilde{\mathscr{Z}}}\right] + \iota_{\mathbf{C}}(w_2) \\
&= (\Psi_B \oplus \iota_{\mathbf{C}})(w),
\end{aligned}$$

where the 5th equality holds by the coercivity of Ψ (see, e.g., [41, Proposition 11.15]) and $\mathrm{dom}(\Psi) = \mathscr{Z}$.

(c) From (3.4), we have

$$\begin{aligned}
J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}(x) &= \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c(x) \\
&= \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu(\Psi_B \oplus \iota_{\mathbf{C}})(\mathfrak{L}x, \mathfrak{C}x) \\
&= \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu\Psi_B\circ\mathfrak{L}(x) + \iota_{\mathbf{C}}(\mathfrak{C}x),
\end{aligned}$$

which implies

$$\underset{x\in\mathscr{X}}{\mathrm{argmin}}\; J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}(x) = \underset{\mathfrak{C}x\in\mathbf{C}}{\mathrm{argmin}}\; \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu\Psi_B\circ\mathfrak{L}(x).$$

APPENDIX B. PROOF OF PROPOSITION 3.2

First, we prove $(C_1) \implies (C_2)$. Fix $y \in \mathscr{Y}$ arbitrarily. Then, for every $x \in \mathscr{X}$, we have

$$\begin{aligned}
J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}(x) &= \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c(x) \\
&= \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu(\Psi_B \oplus \iota_{\mathbf{C}})(\mathfrak{L}x, \mathfrak{C}x) \\
&= \frac{1}{2}\|y-Ax\|^2_{\mathscr{Y}} + \mu\Psi_B\circ\mathfrak{L}(x) + \iota_{\mathbf{C}}(\mathfrak{C}x) \qquad \text{(B.1)} \\
&= \frac{1}{2}\|Ax\|^2_{\mathscr{Y}} - \langle y, Ax\rangle_{\mathscr{Y}} + \frac{1}{2}\|y\|^2_{\mathscr{Y}} + \mu\Psi_B\circ\mathfrak{L}(x) + \iota_{\mathbf{C}}(\mathfrak{C}x) \qquad \text{(B.2)}
\end{aligned}$$

Let us focus on RHS of (B.1). Since Ψ and $\mathfrak{L}$ satisfy the condition for the LiGME model in Definition 2.1, Fact 2.4 yields $J_{\Psi_B\circ\mathfrak{L}} = \frac{1}{2}\|\cdot - y\|^2_{\mathscr{Y}} \circ A + \Psi_B\circ\mathfrak{L} \in \Gamma_0(\mathscr{X})$. On the other hand, $\mathbf{C}\cap\mathrm{ran}\,\mathfrak{C} \neq \emptyset$ and Fact 2.1(b) yield $\iota_{\mathbf{C}}\circ\mathfrak{C} \in \Gamma_0(\mathscr{X})$. Then, from $\mathrm{dom}(J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}) = \mathfrak{C}^{-1}(\mathbf{C}\cap\mathrm{ran}\,\mathfrak{C}) \neq \emptyset$, we can apply Fact 2.1(a) to RHS of (B.1), which yields $J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c} \in \Gamma_0(\mathscr{X})$.

Next, let us focus on RHS of (B.2). Since $\frac{1}{2}\|y\|_{\mathscr{Y}}^2 - \langle y, A\cdot\rangle_{\mathscr{Y}}$ is affine function, $(C_2) \iff \frac{1}{2}\|A\cdot\|_{\mathscr{X}}^2 + \mu\Psi_B\circ\mathfrak{L} + \iota_{\mathbf{C}}\circ\mathfrak{C} \in \Gamma_0(\mathscr{X}) \iff (C_3)$ holds.

Appendix C. Proof of Theorem 3.1

(a) We divide proof of (a) into two steps.

(Step 1) First, from Fermat's rule and the definition of $\mathscr{S}$, $\mathscr{S} = \left\{x^\diamond \in \mathscr{X} \,|\, 0_{\mathscr{X}} \in \partial J_{\mathbf{\Psi}_{B_c}\circ\mathfrak{L}_c}(x^\diamond)\right\}$ holds. In Step 1, we characterize $\mathscr{S}$ in terms of an affine operator and a set-valued operator.

Claim C.1. In Problem 3.1, for every $x^\diamond \in \mathscr{X}$, we have $x^\diamond \in \mathscr{S}$ if and only if there exists $(v^\diamond, w^\diamond) \in \mathscr{Z} \times \mathscr{Z}_c$ s.t.

$$(0_{\mathscr{X}}, 0_{\mathscr{Z}}, 0_{\mathscr{Z}_c}) \in F(x^\diamond, v^\diamond, w^\diamond) + G(x^\diamond, v^\diamond, w^\diamond),$$

where $F : \mathscr{H} \to \mathscr{H}$ and $G : \mathscr{H} \to 2^{\mathscr{H}}$ are defined by

$$\begin{aligned} F(x,v,w) &:= ((A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L})x - A^*y, \mu B^*Bv, 0_{\mathscr{Z}_c}), \\ G(x,v,w) &:= \{\mu\mathfrak{L}^*B^*Bv + \mu\mathfrak{L}_c^*w\} \times (-\mu B^*B\mathfrak{L}x + \mu\partial\Psi(v)) \\ &\qquad \times (-\mu\mathfrak{L}_c x + \mu\partial\mathbf{\Psi}^*(w)). \end{aligned}$$

(Step 2) Next, we prove that

$$(x^\diamond, v^\diamond, w^\diamond) \in \mathrm{Fix}(T_{\mathrm{cLiGME}}) \iff (\mathfrak{P} - F)(x^\diamond, v^\diamond, w^\diamond) \in (\mathfrak{P} + G)(x^\diamond, v^\diamond, w^\diamond) \tag{C.1}$$

holds for every $x^\diamond \in \mathscr{X}$, by using $\mathfrak{P}$ in (3.6).

Proof of (Step 1)

First, we decompose (B.1) as

$$\begin{aligned} J_{\mathbf{\Psi}_{B_c}\circ\mathfrak{L}_c}(x) &= \frac{1}{2}\|y - Ax\|_{\mathscr{Y}}^2 + \mu\Psi(\mathfrak{L}x) - \mu\min_{v\in\mathscr{Z}}\left[\Psi(v) + \frac{1}{2}\|B(\mathfrak{L}x - v)\|_{\widetilde{\mathscr{Z}}}^2\right] + \iota_{\mathbf{C}}(\mathfrak{C}x) \\ &= \left(\frac{1}{2}\|y\|_{\mathscr{Y}}^2 - \langle y, Ax\rangle_{\mathscr{Y}} + \frac{1}{2}\|Ax\|_{\mathscr{Y}}^2\right) + \mu\Psi(\mathfrak{L}x) + \iota_{\mathbf{C}}(\mathfrak{C}x) \\ &\quad - \mu\min_{v\in\mathscr{Z}}\left[\Psi(v) + \left(\frac{1}{2}\|B\mathfrak{L}x\|_{\widetilde{\mathscr{Z}}}^2 - \langle B\mathfrak{L}x, Bv\rangle_{\widetilde{\mathscr{Z}}} + \frac{1}{2}\|Bv\|_{\widetilde{\mathscr{Z}}}^2\right)\right] \\ &= \frac{1}{2}\left(\|Ax\|_{\mathscr{Y}}^2 - \mu\|B\mathfrak{L}x\|_{\widetilde{\mathscr{Z}}}^2\right) - \langle y, Ax\rangle_{\mathscr{Y}} + \frac{1}{2}\|y\|_{\mathscr{Y}}^2 + \mu\Psi(\mathfrak{L}x) \\ &\quad + \iota_{\mathbf{C}}(\mathfrak{C}x) + \mu\max_{v\in\mathscr{Z}}\psi_v(x) \\ &= \frac{1}{2}\langle x, (A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L})x\rangle_{\mathscr{X}} - \langle y, Ax\rangle_{\mathscr{Y}} + \frac{1}{2}\|y\|_{\mathscr{Y}}^2 \\ &\quad + \mu\Psi(\mathfrak{L}x) + \iota_{\mathbf{C}}(\mathfrak{C}x) + \mu\max_{v\in\mathscr{Z}}\psi_v(x), \end{aligned} \tag{C.2}$$

where

$$\psi_v : \mathscr{X} \to \mathbb{R} : x \mapsto -\left(\Psi(v) + \frac{1}{2}\|Bv\|_{\widetilde{\mathscr{Z}}}^2 - \langle Bv, B\mathfrak{L}x\rangle_{\widetilde{\mathscr{Z}}}\right). \tag{C.3}$$

We evaluate the subdifferential of the RHS of (C.2). At this point, we use

$$\begin{aligned}\max_{v\in\mathscr{Z}}\psi_v(x) &= \max_{v\in\mathscr{Z}}\left(-\Psi(v)-\frac{1}{2}\|Bv\|^2_{\widetilde{\mathscr{Z}}}+\langle Bv,B\mathfrak{L}x\rangle_{\widetilde{\mathscr{Z}}}\right)\\ &= \left[\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^*\circ B^*\right]\circ B\mathfrak{L}(x), \qquad (C.4)\\ 0_{\mathfrak{Z}} &\in \mathrm{ri}\left(\mathrm{dom}(\iota_{\mathbf{C}})-\mathrm{ran}(\mathfrak{C})\right) \qquad (C.5)\end{aligned}$$

and

$$\mathrm{dom}\left(\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^*\circ B^*\right)=\widetilde{\mathscr{Z}}, \qquad (C.6)$$

where (C.5) is in Assumption 3.1 and (C.6) is verified, with the coercivity of Ψ, by

$$\begin{aligned}(\forall z\in\widetilde{\mathscr{Z}})\ &\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^*(B^*z)\\ &= \sup_{v\in\mathscr{Z}}\left(\langle v,B^*z\rangle_{\mathscr{Z}}-\Psi(v)-\frac{1}{2}\|Bv\|^2_{\widetilde{\mathscr{Z}}}\right)\\ &\le \sup_{v\in\mathscr{Z}}(-\Psi(v))+\sup_{v\in\mathscr{Z}}\left(\langle v,B^*z\rangle_{\mathscr{Z}}-\frac{1}{2}\|Bv\|^2_{\widetilde{\mathscr{Z}}}\right)\\ &= \max_{v\in\mathscr{Z}}(-\Psi(v))+\max_{v\in\mathscr{Z}}\left(\langle Bv,z\rangle_{\widetilde{\mathscr{Z}}}-\frac{1}{2}\|Bv\|^2_{\widetilde{\mathscr{Z}}}\right)<\infty.\end{aligned}$$

In the RHS of (C.2), the first three terms are differentiable and we obtain

$$\begin{aligned}\partial J_{\Psi_{B_c}\circ\mathfrak{L}_c}(x) &= \nabla\left(\frac{1}{2}\langle x,(A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})x\rangle_{\mathscr{X}}+\frac{1}{2}\|y\|^2_{\mathscr{Y}}-\langle y,Ax\rangle_{\mathscr{Y}}\right)\\ &\quad+\partial\left(\mu\Psi\circ\mathfrak{L}+\mu\max_{v\in\mathscr{Z}}\psi_v+\iota_{\mathbf{C}}\circ\mathfrak{C}\right)(x)\\ &= (A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})x-A^*y+\mu\partial\left(\Psi\circ\mathfrak{L}+\max_{v\in\mathscr{Z}}\psi_v+\iota_{\mathbf{C}}\circ\mathfrak{C}\right)(x). \qquad (C.7)\end{aligned}$$

From $\mathrm{dom}(\Psi\circ\mathfrak{L})=\mathscr{X}$ and $\mathrm{dom}\left(\max_{v\in\mathscr{Z}}\psi_v\right)=\mathscr{X}$, the sum rule (2.2) decomposes (C.7) as

$$\begin{aligned}\partial J_{\Psi_{B_c}\circ\mathfrak{L}_c}(x) &= (A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})x-A^*y+\mu\partial(\Psi\circ\mathfrak{L})(x)\\ &\quad+\mu\partial\left(\max_{v\in\mathscr{Z}}\psi_v\right)(x)+\partial\left(\iota_{\mathbf{C}}\circ\mathfrak{C}\right)(x).\end{aligned} \qquad (C.8)$$

From [33, Lemma 1], we have

$$0_{\mathscr{Z}}\in\mathrm{ri}\left(\mathrm{dom}\left(\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^*\right)-\mathrm{ran}(B^*)\right). \qquad (C.9)$$

By applying the chain rule (2.3) to $\iota_{\mathbf{C}}\circ\mathfrak{C}$, (C.4) and $\Psi\circ\mathfrak{L}$, we have

$$\begin{aligned}\partial\left(\iota_{\mathbf{C}}\circ\mathfrak{C}\right)(x) &= \mathfrak{C}^*\partial\iota_{\mathbf{C}}\left(\mathfrak{C}x\right), \qquad (C.10)\\ \partial\left(\max_{v\in\mathscr{Z}}\psi_v\right) &= (B\mathfrak{L})^*\partial\left[\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)\circ B^*\right]\circ B\mathfrak{L}, \qquad (C.11)\end{aligned}$$

and

$$\partial\left(\Psi\circ\mathfrak{L}\right)(x)=\mathfrak{L}^{*}\partial\Psi(\mathfrak{L}x), \tag{C.12}$$

where (C.10) is verified by (C.5), (C.11) by (C.6) and (C.12) by dom$(\Psi)=\mathscr{Z}$. Applying again the chain rule (2.3) to (C.11) with (C.9) , we have

$$\partial\left(\max_{v\in\mathscr{Z}}\psi_v\right)=(B^{*}B\mathfrak{L})^{*}\,\partial\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^{*}(B^{*}B\mathfrak{L}). \tag{C.13}$$

Moreover, (C.10), (C.12), and the definition of $\boldsymbol{\Psi}$ in Proposition 3.1 yield

$$\mu\partial(\Psi\circ\mathfrak{L})(x)+\partial(\iota_{\mathbf{C}}\circ\mathfrak{C})(x)=\mu\mathfrak{L}^{*}\partial\Psi(\mathfrak{L}x)+\mathfrak{C}^{*}\partial\iota_{\mathbf{C}}(\mathfrak{C}x)=\mu\mathfrak{L}_c^{*}\partial\boldsymbol{\Psi}(\mathfrak{L}_c x). \tag{C.14}$$

Then we simplify (C.8) with (C.13) and (C.14) as

$$\begin{aligned}\partial J_{\boldsymbol{\Psi}_{B_c}\circ\mathfrak{L}_c}(x)=(A^{*}A-\mu\mathfrak{L}^{*}B^{*}B\mathfrak{L})x-A^{*}y+\mu\mathfrak{L}_c^{*}\partial\boldsymbol{\Psi}(\mathfrak{L}_c x)\\ +\mu\,(B^{*}B\mathfrak{L})^{*}\,\partial\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^{*}(B^{*}B\mathfrak{L}x).\end{aligned} \tag{C.15}$$

Furthermore, by

$$\begin{aligned}&w^{\diamond}\in\partial\boldsymbol{\Psi}(\mathfrak{L}_c x^{\diamond})\iff\mathfrak{L}_c x^{\diamond}\in\partial\boldsymbol{\Psi}^{*}(w^{\diamond}),\\ &v^{\diamond}\in\partial\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)^{*}(B^{*}B\mathfrak{L}x^{\diamond})\\ &\quad\iff B^{*}B\mathfrak{L}x^{\diamond}\in\partial\left(\Psi+\frac{1}{2}\|B\cdot\|^2_{\widetilde{\mathscr{Z}}}\right)(v^{\diamond})\\ &\quad\iff B^{*}B\mathfrak{L}x^{\diamond}\in\partial\Psi(v^{\diamond})+B^{*}Bv^{\diamond}\end{aligned}$$

due to the relation between the subdifferential and conjugate (2.4) and the sum rule (2.2) with dom$(\Psi)=\mathscr{Z}$, we deduce from (C.15)

$$\begin{aligned}&x^{\diamond}\in\mathscr{S}\\ &\iff\begin{cases}0_{\mathscr{X}}\in(A^{*}A-\mu\mathfrak{L}^{*}B^{*}B\mathfrak{L})x^{\diamond}-A^{*}y+\mu\mathfrak{L}_c^{*}w^{\diamond}+\mu\,(B^{*}B\mathfrak{L})^{*}\,v^{\diamond}\\ B^{*}B\mathfrak{L}x^{\diamond}\in\partial\Psi(v^{\diamond})+B^{*}Bv^{\diamond}\\ \mathfrak{L}_c x^{\diamond}\in\partial\boldsymbol{\Psi}^{*}(w^{\diamond})\end{cases}\\ &\iff\begin{cases}0_{\mathscr{X}}\in(A^{*}A-\mu\mathfrak{L}^{*}B^{*}B\mathfrak{L})x^{\diamond}-A^{*}y+\mu\mathfrak{L}_c^{*}w^{\diamond}+\mu\mathfrak{L}^{*}B^{*}Bv^{\diamond}\\ 0_{\mathscr{Z}}\in-\mu B^{*}B\mathfrak{L}x^{\diamond}+\mu B^{*}Bv^{\diamond}+\mu\partial\Psi(v^{\diamond})\\ 0_{\mathscr{Z}_c}\in-\mu\mathfrak{L}_c x^{\diamond}+\mu\partial\boldsymbol{\Psi}^{*}(w^{\diamond})\end{cases}\\ &\iff(0_{\mathscr{X}},0_{\mathscr{Z}},0_{\mathscr{Z}_c})\in F(x^{\diamond},v^{\diamond},w^{\diamond})+G(x^{\diamond},v^{\diamond},w^{\diamond})\end{aligned}$$

which completes the proof of Claim C.1.

Proof of (Step 2)
(C.1) is verified by

$$\begin{aligned}
&T_{\mathrm{cLiGME}}(x,v,w)=(\xi,\zeta,\eta)\\
&\iff \begin{cases} [\sigma\,\mathrm{Id}-(A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})]x-\mu\mathfrak{L}^*B^*Bv-\mu\mathfrak{L}_c^*w+A^*y=\sigma\xi \\ 2\mu B^*B\mathfrak{L}\xi-\mu B^*B\mathfrak{L}x^{\diamond}+(\tau\,\mathrm{Id}-\mu B^*B)\,v\in(\tau\,\mathrm{Id}+\mu\partial\Psi)\,(\zeta) \\ 2\mu\mathfrak{L}_c\xi-\mu\mathfrak{L}_cx+\mu w\in(\mu\,\mathrm{Id}+\mu\partial\boldsymbol{\Psi}^*)\,(\eta)=\mu(\mathrm{Id}+\partial\boldsymbol{\Psi}^*)(\eta)\end{cases}\\
&\iff (\mathfrak{P}-F)(x,v,w)\in(\mathfrak{P}+G)(\xi,\zeta,\eta),
\end{aligned} \tag{C.16}$$

where we use the relation between the proximity operator and the resolvent of subdifferential (2.7).

(b) First, under condition (3.5), we prove $\mathfrak{P}\succ\mathrm{O}_{\mathscr{H}}$. The Schur complement (see e.g. [45, Theorem 7.7.7]) yields

$$\begin{aligned}
&\mathfrak{P}\succ\mathrm{O}_{\mathscr{H}}\\
&\iff \sigma\,\mathrm{Id}-\begin{bmatrix}-\mu\mathfrak{L}^*B^*B & -\mu\mathfrak{L}_c^*\end{bmatrix}\begin{bmatrix}\tau\,\mathrm{Id} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mu\,\mathrm{Id}\end{bmatrix}^{-1}\begin{bmatrix}-\mu B^*B\mathfrak{L}\\ -\mu\mathfrak{L}_c\end{bmatrix}\succ\mathrm{O}_{\mathscr{X}}\\
&\iff \sigma\,\mathrm{Id}-\begin{bmatrix}-\mu\mathfrak{L}^*B^*B & -\mu\mathfrak{L}_c^*\end{bmatrix}\begin{bmatrix}\frac{1}{\tau}\mathrm{Id} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \frac{1}{\mu}\mathrm{Id}\end{bmatrix}\begin{bmatrix}-\mu B^*B\mathfrak{L}\\ -\mu\mathfrak{L}_c\end{bmatrix}\succ\mathrm{O}_{\mathscr{X}}\\
&\iff \sigma\,\mathrm{Id}-\frac{\mu^2}{\tau}\mathfrak{L}^*\left(B^*B\right)^2\mathfrak{L}-\mu\mathfrak{L}_c^*\mathfrak{L}_c\succ\mathrm{O}_{\mathscr{X}}\\
&\iff \left(\sigma\,\mathrm{Id}-\frac{\kappa}{2}A^*A-\mu\mathfrak{L}_c^*\mathfrak{L}_c\right)+\left(\frac{\kappa}{2}A^*A-\frac{\mu^2}{\tau}\mathfrak{L}^*\left(B^*B\right)^2\mathfrak{L}\right)\succ\mathrm{O}_{\mathscr{X}}.
\end{aligned}$$

From (3.5), it is sufficient to show $\left(\frac{\kappa}{2}A^*A-\frac{\mu^2}{\tau}\mathfrak{L}^*\left(B^*B\right)^2\mathfrak{L}\right)\succeq\mathrm{O}_{\mathscr{X}}$. Using (3.5), we have

$$\begin{aligned}
(\forall x\in\mathscr{X})\ \left\langle x,\frac{\mu^2}{\tau}\mathfrak{L}^*\left(B^*B\right)^2\mathfrak{L}x\right\rangle_{\mathscr{X}} &=\frac{\mu^2}{\tau}\left\|B^*B\mathfrak{L}x\right\|_{\mathscr{Z}}^2\\
&\le\frac{\mu^2}{\tau}\left\|B\right\|_{\mathrm{op}}^2\left\|B\mathfrak{L}x\right\|_{\widetilde{\mathscr{Z}}}^2\\
&\le\mu^2\left[\left(\frac{\kappa}{2}+\frac{2}{\kappa}\right)\mu\left\|B\right\|_{\mathrm{op}}^2\right]^{-1}\left\|B\right\|_{\mathrm{op}}^2\left\|B\mathfrak{L}x\right\|_{\widetilde{\mathscr{Z}}}^2\\
&\le\mu\frac{\kappa}{2}\left\|B\mathfrak{L}x\right\|_{\widetilde{\mathscr{Z}}}^2=\mu\frac{\kappa}{2}\left\langle x,\mathfrak{L}^*B^*B\mathfrak{L}x\right\rangle_{\mathscr{X}},
\end{aligned}$$

which yields

$$\begin{aligned}
(\forall x\in\mathscr{X})\ \left\langle x,\left(\frac{\kappa}{2}A^*A-\frac{\mu^2}{\tau}\mathfrak{L}^*\left(B^*B\right)^2\mathfrak{L}\right)x\right\rangle_{\mathscr{X}} &\ge\left\langle x,\frac{\kappa}{2}A^*Ax\right\rangle_{\mathscr{X}}-\mu\frac{\kappa}{2}\left\langle x,\mathfrak{L}^*B^*B\mathfrak{L}x\right\rangle_{\mathscr{X}}\\
&=\frac{\kappa}{2}\left\langle x,(A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L})x\right\rangle_{\mathscr{X}}\ge 0,
\end{aligned}$$

where the last inequality holds by the assumption $A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L}\succeq\mathrm{O}_{\mathscr{X}}$ in Assumption 3.1.

Next, we prove that T_{cLiGME} is $\frac{\kappa}{2\kappa-1}$-averaged nonexpansive over $(\mathscr{H},\langle\cdot,\cdot\rangle_{\mathfrak{P}},\|\cdot\|_{\mathfrak{P}})$. By applying $\mathfrak{P}\succ \mathrm{O}_{\mathscr{H}}$ to (C.16), we have

$$\begin{aligned} T_{\mathrm{cLiGME}}(x,v,w)=(\xi,\zeta,\eta) &\iff (\mathfrak{P}-F)(x,v,w)\in(\mathfrak{P}+G)(\xi,\zeta,\eta)\\ &\iff \mathfrak{P}^{-1}(\mathfrak{P}-F)(x,v,w)\in\mathfrak{P}^{-1}(\mathfrak{P}+G)(\xi,\zeta,\eta)\\ &\iff (\mathrm{Id}-\mathfrak{P}^{-1}\circ F)(x,v,w)\in(\mathrm{Id}+\mathfrak{P}^{-1}\circ G)(\xi,\zeta,\eta). \end{aligned} \tag{C.17}$$

Moreover, as will be shown later, $\mathfrak{P}^{-1}\circ G$ is maximally monotone over $(\mathscr{H},\langle\cdot,\cdot\rangle_{\mathfrak{P}},\|\cdot\|_{\mathfrak{P}})$, which implies its resolvent $(\mathrm{Id}+\mathfrak{P}^{-1}\circ G)^{-1}$ is single-valued and $\frac{1}{2}$-averaged nonexpansive over $(\mathscr{H},\langle\cdot,\cdot\rangle_{\mathfrak{P}},\|\cdot\|_{\mathfrak{P}})$. Therefore, from (C.17), we obtain the expression

$$\begin{aligned} T_{\mathrm{cLiGME}}&=(\mathrm{Id}+\mathfrak{P}^{-1}\circ G)^{-1}\circ(\mathrm{Id}-\mathfrak{P}^{-1}\circ F)\\ &=(\mathrm{Id}+\mathfrak{P}^{-1}\circ G)^{-1}\circ\left(\left(1-\frac{1}{\kappa}\right)\mathrm{Id}+\frac{1}{\kappa}\left(\mathrm{Id}-\kappa\mathfrak{P}^{-1}\circ F\right)\right), \end{aligned}$$

where the nonexpansiveness of $\mathrm{Id}-\kappa\mathfrak{P}^{-1}\circ F$ over $(\mathscr{H},\langle\cdot,\cdot\rangle_{\mathfrak{P}},\|\cdot\|_{\mathfrak{P}})$ is proven as follows by using

$$M:=\begin{bmatrix} A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L} & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{X})} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{X})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z})} & \mu B^*B & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{Z}_c}\end{bmatrix}\succeq \mathrm{O}_{\mathscr{H}}. \tag{C.18}$$

With $M^*=M$ and the expression $(\forall(x,v,w)\in\mathscr{H})$

$$F(x,v,w)=M\begin{bmatrix}x\\v\\w\end{bmatrix}+\begin{bmatrix}-A^*y\\0_{\mathscr{Z}}\\0_{\mathscr{Z}_c}\end{bmatrix},$$

we have for all $\mathbf{u}_1,\mathbf{u}_2\in\mathscr{H}$,

$$\begin{aligned} &\left\|\left(\mathrm{Id}-\kappa\mathfrak{P}^{-1}\circ F\right)\mathbf{u}_1-\left(\mathrm{Id}-\kappa\mathfrak{P}^{-1}\circ F\right)\mathbf{u}_2\right\|_{\mathfrak{P}}^2\\ &=\left\|(\mathbf{u}_1-\mathbf{u}_2)-\kappa\left[\left(\mathfrak{P}^{-1}\circ F\right)(\mathbf{u}_1)-\left(\mathfrak{P}^{-1}\circ F\right)(\mathbf{u}_2)\right]\right\|_{\mathfrak{P}}^2\\ &=\|\mathbf{u}_1-\mathbf{u}_2\|_{\mathfrak{P}}^2-2\kappa\langle\mathbf{u}_1-\mathbf{u}_2,F(\mathbf{u}_1)-F(\mathbf{u}_2)\rangle_{\mathscr{H}}+\kappa^2\left\|\left(\mathfrak{P}^{-1}\circ F\right)(\mathbf{u}_1)-\left(\mathfrak{P}^{-1}\circ F\right)(\mathbf{u}_2)\right\|_{\mathfrak{P}}^2\\ &=\|\mathbf{u}_1-\mathbf{u}_2\|_{\mathfrak{P}}^2-2\kappa\langle\mathbf{u}_1-\mathbf{u}_2,M(\mathbf{u}_1)-M(\mathbf{u}_2)\rangle_{\mathscr{H}}+\kappa^2\left\langle\mathfrak{P}^{-1}M\mathbf{u}_1-\mathfrak{P}^{-1}M\mathbf{u}_2,M\mathbf{u}_1-M\mathbf{u}_2\right\rangle_{\mathscr{H}}\\ &=\|\mathbf{u}_1-\mathbf{u}_2\|_{\mathfrak{P}}^2-2\kappa\left\langle\mathbf{u}_1-\mathbf{u}_2,\left(M-\frac{\kappa}{2}M\mathfrak{P}^{-1}M\right)(\mathbf{u}_1-\mathbf{u}_2)\right\rangle_{\mathscr{H}}, \end{aligned}$$

which implies, by the property of Schur complement,

$$\begin{aligned} \left(\mathrm{Id}-\kappa\mathfrak{P}^{-1}\circ F\text{ is nonexpansive}\right) &\iff M-\frac{\kappa}{2}M\mathfrak{P}^{-1}M\succeq\mathrm{O}_{\mathscr{H}}\\ &\iff \begin{bmatrix}M & M\\ M & 2\kappa^{-1}\mathfrak{P}\end{bmatrix}\succeq\mathrm{O}_{\mathscr{H}\times\mathscr{H}}. \end{aligned}$$

Moreover, by

$$(\forall(\mathbf{u}_1,\mathbf{u}_2)\in\mathscr{H}\times\mathscr{H})\ \left\langle\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix},\begin{bmatrix}M & M\\ M & 2\kappa^{-1}\mathfrak{P}\end{bmatrix}\begin{bmatrix}\mathbf{u}_1\\ \mathbf{u}_2\end{bmatrix}\right\rangle_{\mathscr{H}\times\mathscr{H}}$$
$$=\langle\mathbf{u}_1,M\mathbf{u}_1\rangle_{\mathscr{H}}+\langle\mathbf{u}_1,M\mathbf{u}_2\rangle_{\mathscr{H}}+\langle\mathbf{u}_2,M\mathbf{u}_1\rangle_{\mathscr{H}}+2\kappa^{-1}\langle\mathbf{u}_2,\mathfrak{P}\mathbf{u}_2\rangle_{\mathscr{H}}$$
$$=\langle\mathbf{u}_1+\mathbf{u}_2,M(\mathbf{u}_1+\mathbf{u}_2)\rangle_{\mathscr{H}}+2\kappa^{-1}\left\langle\mathbf{u}_2,\left(\mathfrak{P}-\frac{\kappa}{2}M\right)\mathbf{u}_2\right\rangle_{\mathscr{H}},$$

we have

$$\left(\mathrm{Id}-\kappa\mathfrak{P}^{-1}\circ F\text{ is nonexpansive}\right)\iff\mathfrak{P}-\frac{\kappa}{2}M\succeq\mathrm{O}_{\mathscr{H}},$$

where

$$\mathfrak{P}-\frac{\kappa}{2}M=\begin{bmatrix}\sigma\,\mathrm{Id} & -\mu\mathfrak{L}^*B^*B & -\mu\mathfrak{L}_c^*\\ -\mu B^*B\mathfrak{L} & \tau\,\mathrm{Id} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ -\mu\mathfrak{L}_c & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mu\,\mathrm{Id}\end{bmatrix}$$

$$-\frac{\kappa}{2}\begin{bmatrix}A^*A-\mu\mathfrak{L}^*B^*B\mathfrak{L} & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{X})} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{X})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z})} & \mu B^*B & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{Z}_c}\end{bmatrix}$$

$$=\begin{bmatrix}\sigma\,\mathrm{Id}-(\kappa/2)A^*A & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{X})} & -\mu\mathfrak{L}_c^*\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z})} & \mathrm{O}_{\mathscr{Z}} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ -\mu\mathfrak{L}_c & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mu\,\mathrm{Id}\end{bmatrix}$$

$$+\begin{bmatrix}(\kappa\mu/2)\mathfrak{L}^*B^*B\mathfrak{L} & -\mu\mathfrak{L}^*B^*B & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{X})}\\ -\mu B^*B\mathfrak{L} & \tau\,\mathrm{Id}-(\kappa\mu/2)B^*B & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{Z}_c}\end{bmatrix}.$$

By the property of Schur complement, we have

$$\begin{bmatrix}\sigma\,\mathrm{Id}-(\kappa/2)A^*A & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{X})} & -\mu\mathfrak{L}_c^*\\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z})} & \mathrm{O}_{\mathscr{Z}} & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})}\\ -\mu\mathfrak{L}_c & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mu\,\mathrm{Id}\end{bmatrix}\succeq\mathrm{O}_{\mathscr{H}}$$

$$\iff\sigma\,\mathrm{Id}-(\kappa/2)A^*A-\mu\mathfrak{L}_c^*\mathfrak{L}_c\succeq\mathrm{O}_{\mathscr{X}}\ (\Leftarrow(3.5))$$

and

$$\mathrm{O}_{\mathscr{H}} \preceq \begin{bmatrix} (\kappa\mu/2)\mathfrak{L}^*B^*B\mathfrak{L} & -\mu\mathfrak{L}^*B^*B & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{X})} \\ -\mu B^*B\mathfrak{L} & \tau\,\mathrm{Id}-(\kappa\mu/2)B^*B & \mathrm{O}_{\mathscr{B}(\mathscr{Z}_c,\mathscr{Z})} \\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{Z}_c)} & \mathrm{O}_{\mathscr{Z}_c} \end{bmatrix}$$

$$\iff \mathrm{O}_{\mathscr{X}\times\mathscr{Z}} \preceq \begin{bmatrix} \mathfrak{L}^* & \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{X})} \\ \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z})} & \mathrm{Id} \end{bmatrix} \begin{bmatrix} (\kappa\mu/2)B^*B & -\mu B^*B \\ -\mu B^*B & \tau\,\mathrm{Id}-(\kappa\mu/2)B^*B \end{bmatrix} \begin{bmatrix} \mathfrak{L} & \mathrm{O}_{\mathscr{B}(\mathscr{X},\mathscr{Z})} \\ \mathrm{O}_{\mathscr{B}(\mathscr{Z},\mathscr{X})} & \mathrm{Id} \end{bmatrix}$$

$$\Longleftarrow \mathrm{O}_{\mathscr{Z}\times\mathscr{Z}} \preceq \begin{bmatrix} (\kappa\mu/2)B^*B & -\mu B^*B \\ -\mu B^*B & \tau\,\mathrm{Id}-(\kappa\mu/2)B^*B \end{bmatrix}, \tag{C.19}$$

implying thus (RHS of (C.19)) $\implies \mathfrak{P} - \frac{\kappa}{2}M \succeq \mathrm{O}_{\mathscr{H}}$. The RHS of (C.19) is guaranteed by

$$\begin{aligned}
&(v_1, v_2 \in \mathscr{Z}) \quad \left\langle \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}, \begin{bmatrix} (\kappa\mu/2)B^*B & -\mu B^*B \\ -\mu B^*B & \tau\,\mathrm{Id}-(\kappa\mu/2)B^*B \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right\rangle_{\mathscr{Z}\times\mathscr{Z}} \\
&= \langle v_1, (\kappa\mu/2)B^*Bv_1\rangle_{\mathscr{Z}} - \langle v_1, \mu B^*Bv_2\rangle_{\mathscr{Z}} - \langle v_2, \mu B^*Bv_1\rangle_{\mathscr{Z}} \\
&\quad + \langle v_2, (\tau\,\mathrm{Id}-(\kappa\mu/2)B^*B)\, v_2\rangle_{\mathscr{Z}} \\
&= \frac{2\mu}{\kappa} \left\| \frac{\kappa}{2}Bv_1 - Bv_2 \right\|_{\widetilde{\mathscr{Z}}}^2 - \frac{2\mu}{\kappa} \|Bv_2\|_{\widetilde{\mathscr{Z}}}^2 + \langle v_2, (\tau\,\mathrm{Id}-(\kappa\mu/2)B^*B)\, v_2\rangle_{\mathscr{Z}} \\
&= \frac{2\mu}{\kappa} \left\| \frac{\kappa}{2}Bv_1 - Bv_2 \right\|_{\widetilde{\mathscr{Z}}}^2 + \tau \|v_2\|_{\mathscr{Z}}^2 - \mu\left(\frac{\kappa}{2} + \frac{2}{\kappa}\right) \|Bv_2\|_{\widetilde{\mathscr{Z}}}^2 \\
&\geq \tau \|v_2\|_{\mathscr{Z}}^2 - \mu\left(\frac{\kappa}{2} + \frac{2}{\kappa}\right) \|Bv_2\|_{\widetilde{\mathscr{Z}}}^2 \\
&\geq \left(\tau - \mu\left(\frac{\kappa}{2} + \frac{2}{\kappa}\right) \|B\|_{\mathrm{op}}^2\right) \|v_2\|_{\mathscr{Z}}^2 \geq 0.
\end{aligned}$$

Consequently, we have proven that $\mathrm{Id} - \mathfrak{P}^{-1} \circ F$ is $\frac{1}{\kappa}$-averaged nonexpansive and, by Fact 2.3(a), $T_{\mathrm{cLiGME}} = (\mathrm{Id} + \mathfrak{P}^{-1} \circ G)^{-1} \circ (\mathrm{Id} - \mathfrak{P}^{-1} \circ F)$ is $\frac{\kappa}{2\kappa-1}$-averaged nonexpansive over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathfrak{P}}, \|\cdot\|_{\mathfrak{P}})$.

Finally we show the maximal monotonicity of $\mathfrak{P}^{-1} \circ G$ over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathfrak{P}}, \|\cdot\|_{\mathfrak{P}})$. Let G_1 be a set-valued operator:

$$G_1 : \mathscr{H} \to 2^{\mathscr{H}} : (x, v, w) \mapsto \{0_{\mathscr{X}}\} \times (\mu\partial\Psi(v)) \times (\mu\partial\mathbf{\Psi}^*(w))$$

and G_2 be a linear operator:

$$G_2 : \mathscr{H} \to \mathscr{H} : (x, v, w) \mapsto (\mu\mathfrak{L}^*B^*Bv + \mu\mathfrak{L}_c^*w, -\mu B^*B\mathfrak{L}x, -\mu\mathfrak{L}_c x).$$

Then, from [41, Theorem 20.40, Proposition 16.9, 20.23], G_1 is maximally monotone over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}}, \|\cdot\|_{\mathscr{H}})$. Also, G_2 is bounded linear skew-symmetric operator (i.e. $G_2 \in \mathscr{B}(\mathscr{H}, \mathscr{K})$ and $G_2^* = -G_2$) and thus is maximally monotone over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}}, \|\cdot\|_{\mathscr{H}})$ from [41, Example 20.35]. Moreover, from $\mathrm{dom}(G_2) = \mathscr{H}$ and [41, Corollary 25.5 (i)],

$G = G_1 + G_2$ is maximally monotone over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathscr{H}}, \|\cdot\|_{\mathscr{H}})$. Thus, $\mathfrak{P}^{-1} \circ G$ is monotone over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathfrak{P}}, \|\cdot\|_{\mathfrak{P}})$. Next, we confirm the maximal monotonicity of $\mathfrak{P}^{-1} \circ G$ over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathfrak{P}}, \|\cdot\|_{\mathfrak{P}})$. Assume that there exists $(\mathbf{u},\mathbf{z}) \notin \mathrm{gra}\left(\mathfrak{P}^{-1} \circ G\right)$, which implies $(\mathbf{u},\mathfrak{P}\mathbf{z}) \notin \mathrm{gra}(G)$, s.t. $(\forall(\mathbf{u}',\mathbf{z}') \in \mathrm{gra}(\mathfrak{P}^{-1} \circ G))$ $\langle \mathbf{u}-\mathbf{u}', \mathbf{z}-\mathbf{z}'\rangle_{\mathfrak{P}} = \langle \mathbf{u}-\mathbf{u}', \mathfrak{P}(\mathbf{z}-\mathbf{z}')\rangle_{\mathscr{H}} \geq 0$. However, because of $(\mathbf{u}',\mathfrak{P}\mathbf{z}') \in \mathrm{gra}(G)$, it contradicts to the maximal monotonicity of G. Hence, $\mathfrak{P}^{-1} \circ G$ is maximally monotone over $(\mathscr{H}, \langle\cdot,\cdot\rangle_{\mathfrak{P}}, \|\cdot\|_{\mathfrak{P}})$.

(c) From (b), we can apply Krasnosel'skiĭ-Mann algorithm (Fact 2.3(b)) to T_{cLiGME}, which yields (c).

Appendix D. Proof of Proposition 3.3

If $\theta = 0$, we have $B_\theta = \mathrm{O}_l$ and $J_{\boldsymbol{\Psi}_{B_c} \circ \mathfrak{L}_c} \in \Gamma_0(\mathbb{R}^n)$ for all $y \in \mathbb{R}^m$ by Proposition 3.2. Let $\theta \in (0,1]$. From Proposition 3.2, it is sufficient to show

$$A^\top A - \mu \mathfrak{L}^\top B_\theta^\top B_\theta \mathfrak{L} \succeq \mathrm{O}_n.$$

$\begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix} \widetilde{\mathfrak{L}} = \mathfrak{L}$ and the definition (3.8) show

$$\begin{aligned}
& \mathrm{O}_n \preceq A^\top A - \mu \mathfrak{L}^\top B_\theta^\top B_\theta \mathfrak{L} \\
& \qquad = A^\top A - \mu \left(\begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix} \widetilde{\mathfrak{L}}\right)^\top B_\theta^\top B_\theta \left(\begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix} \widetilde{\mathfrak{L}}\right) \\
\iff & \mathrm{O}_n \preceq \left(A\widetilde{\mathfrak{L}}^{-1}\right)^\top \left(A\widetilde{\mathfrak{L}}^{-1}\right) - \mu \begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix}^\top B_\theta^\top B_\theta \begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix} \\
& \qquad = \begin{bmatrix}\widetilde{A}_1 & \widetilde{A}_2\end{bmatrix}^\top \begin{bmatrix}\widetilde{A}_1 & \widetilde{A}_2\end{bmatrix} - \mu \begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix}^\top B_\theta^\top B_\theta \begin{bmatrix}\mathrm{O}_{l\times(n-l)} & \mathrm{I}_n\end{bmatrix} \\
& \qquad = \begin{bmatrix}\widetilde{A}_1^\top \widetilde{A}_1 & \widetilde{A}_1^\top \widetilde{A}_2 \\ \widetilde{A}_2^\top \widetilde{A}_1 & \widetilde{A}_2^\top \widetilde{A}_2 - \mu B_\theta^\top B_\theta\end{bmatrix}.
\end{aligned} \tag{D.1}$$

$\widetilde{A}_1^\top \widetilde{A}_1 \succeq \mathrm{O}_l$ holds obviously and $\widetilde{A}_1^\top \widetilde{A}_1 \left(\widetilde{A}_1^\top \widetilde{A}_1\right)^\dagger \widetilde{A}_1^\top \widetilde{A}_2 = \widetilde{A}_1^\top \widetilde{A}_2$ due to $\mathrm{ran}\left(\widetilde{A}_1^\top \widetilde{A}_2\right) \subset \mathrm{ran}\left(\widetilde{A}_1\right) = \mathrm{ran}\left(\widetilde{A}_1^\top \widetilde{A}_1\right) = \mathrm{null}\left(\widetilde{A}_1^\top \widetilde{A}_1 \left(\widetilde{A}_1^\top \widetilde{A}_1\right)^\dagger - \mathrm{I}_{n-l}\right)$. Thus, by [46, Theorem 1], we have

$$(\text{RHS of (D.1)}) \iff \widetilde{A}_1^\top \widetilde{A}_2 - \mu B_\theta^\top B_\theta - \widetilde{A}_2^\top \widetilde{A}_1 \left(\widetilde{A}_1^\top \widetilde{A}_1\right) \widetilde{A}_1^\top \widetilde{A}_2 \succeq \mathrm{O}_l. \tag{D.2}$$

Since B_θ satisfies $\theta^{-1}\mu B_\theta^\top B_\theta = \widetilde{A}_2^\top \widetilde{A}_2 - \widetilde{A}_2^\top \widetilde{A}_1 \left(\widetilde{A}_1^\top \widetilde{A}_1\right)^\dagger \widetilde{A}_1^\top \widetilde{A}_2$, we have

$$(\text{RHS of (D.2)}) \iff \theta^{-1}\mu B_\theta^\top B_\theta - \mu B_\theta^\top B_\theta \succeq \mathrm{O}_l. \tag{D.3}$$

Since $\theta^{-1} \geq 1$ and $B_\theta^\top B_\theta \succeq \mathrm{O}_l$, RHS of (D.3) holds and therefore we have proven $J_{\boldsymbol{\Psi}_{B_c} \circ \mathfrak{L}_c} \in \Gamma_0(\mathbb{R}^n)$.

Appendix E. Proof of computation of P_{C_2}

First, from the definition of the projection, we have

$$P_{C_2}(x) = \operatorname*{argmin}_{y \in C_2} \|x-y\|_{\mathfrak{Z}_2} = \operatorname*{argmin}_{y \in C_2} \sum_{i=1}^{N^2} (x_i - y_i)^2.$$

Since C_2 influences only background entries $\left\{i \in \{1,\cdots,N^2\} | \Upsilon^{-1}(i) \in I_{\text{back}}\right\}$ with some constant level, say $c \in \mathbb{R}$, we have

$$[P_{C_2}(x)]_i = \begin{cases} c & (\Upsilon^{-1}(i) \in I_{\text{back}}) \\ x_i & (otherwise). \end{cases}$$

Moreover, by

$$\begin{aligned}
\Delta(c) &:= \sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} (x_i - c)^2 \\
&= |I_{\text{back}}| \left(c - \frac{1}{|I_{\text{back}}|} \sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} x_i \right)^2 - \frac{1}{|I_{\text{back}}|} \left(\sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} x_i \right)^2 + \sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} x_i^2 \\
&\geq -\frac{1}{|I_{\text{back}}|} \left(\sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} x_i \right)^2 + \sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} x_i^2,
\end{aligned}$$

which implies

$$\underset{c \in \mathbb{R}}{\text{argmin}}\ \Delta(c) = \frac{1}{|I_{\text{back}}|} \sum_{\Upsilon^{-1}(i) \in I_{\text{back}}} x_i.$$