A Douglas-Rachford Splitting Algorithm for the LiGME Model with Applications in Piecewise-Constant Signal Recovery

Masao Yamagishi
Department of Applied Informatics, *Graduate School of Science and Engineering, Hosei University*, Tokyo, Japan
yamagishi@hosei.ac.jp

***Abstract*—We propose a Douglas-Rachford splitting algorithm for the LiGME model which is a least squares estimation framework utilizing a nonconvex regularizer to promote sparsity while preserving the overall convexity of the problem. The proposed algorithm guarantees global convergence to the optimal solution under an overall convexity condition. By utilizing a slightly modified inner product, the matrix inversion in the updates of the proposed algorithm is reduced to sparse matrix inversion in certain situations. A numerical experiment demonstrates that the proposed algorithm achieves faster convergence compared to existing algorithms for the LiGME model.**

***Index Terms*—LiGME model, Douglas-Rachford splitting algorithm, Piecewise-Constant Signal Recovery.**

I. Introduction

Many tasks in inverse problems for data science and engineering, including signal processing and machine learning, involve estimating a vector $x^\star \in \mathcal{X}$ from observed data:

$$y = Ax^\star + \epsilon \in \mathcal{Y}, \tag{1}$$

where $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}}, \|\cdot\|_{\mathcal{X}})$ and $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}}, \|\cdot\|_{\mathcal{Y}})$ are finite-dimensional real Hilbert spaces, $A\colon \mathcal{X} \to \mathcal{Y}$ is a known bounded linear operator, and $\epsilon \in \mathcal{Y}$ represents an unknown noise vector. In many applications of this estimation problem, it is known that $x^\star$ has a sparse representation through a linear transformation $\mathfrak{L}\colon \mathcal{X} \to \mathcal{Z}$. Utilizing this prior information is key to constructing effective estimation algorithms.

A common approach to such estimation problems is to solve the regularized least squares problem:

$$\underset{x\in\mathcal{X}}{\text{Minimize}}\ \frac{1}{2}\|y - Ax\|_{\mathcal{Y}}^2 + \mu\Psi \circ \mathfrak{L}(x), \tag{2}$$

where $\Psi\colon \mathcal{Z} \to (-\infty, \infty]$ is utilized. Examples where $\Psi \circ \mathfrak{L}$ is a convex function include the Total Variation (TV) regularization [1] and wavelet-based sparsity regularization methods [2]. In these methods, the ℓ_1 norm $\|\cdot\|_1$, known as the best convex lower approximation of the ℓ_0 pseudo-norm, has been widely employed as Ψ, notably in the Lasso regression [3] (which utilizes $(\Psi, \mathfrak{L}) = (\|\cdot\|_1, \mathrm{Id})$). Despite its effectiveness, Lasso suffers from underestimation bias in the estimation of large coefficients, which leads us to explore alternative regularizations.

This work was supported in part by JSPS [grants-in-aid 23K11246] and JST SICORP [grant number JPMJSC20C6].

To address the bias issue, various nonconvex sparsity-promoting regularizations have been introduced (see, e.g., [4]–[6]). However, nonconvex regularization often makes the problem (2) difficult to find a globally optimal solution.

To circumvent these difficulties, convexity-preserving regularizers have been developed, i.e., the regularizer $\Psi \circ \mathfrak{L}$ is chosen to be nonconvex while ensuring the overall convexity of the optimization problem (2). One such method for the case of $(\Psi, \mathfrak{L}) = (\|\cdot\|_1, \mathrm{Id})$ is the minimax concave (MC) penalty [7], which ensures overall convexity by relying on the strong convexity of the quadratic function $\frac{1}{2}\|y - A(\cdot)\|_{\mathcal{Y}}^2$ (i.e., the nonsingularity of $A^\top A$). In cases where strong convexity is not satisfied, the generalized minimax concave (GMC) penalty has been proposed as an extension of the MC penalty to achieve overall convexity [8].

For general $\mathfrak{L}$, the LiGME model [9] employs a nonconvex penalty to promote sparsity in the transform domain while preserving overall convexity. This model modifies (2) by replacing Ψ with the GME penalty function Ψ_B (see Definition 1 in Sec. II-B), which can further emphasize sparsity:

$$\underset{x\in\mathcal{X}}{\text{Minimize}}\ \frac{1}{2}\|y - Ax\|_{\mathcal{Y}}^2 + \mu\Psi_B \circ \mathfrak{L}(x). \tag{3}$$

While the LiGME model (3) has favorable theoretical properties, its algorithmic solvers still leave room for improvement. In [9], an iterative algorithm (see also Fact 3(b); hereafter refered to as *Abe's algorithm*) was introduced, inspired by the Condat-Vu primal-dual algorithm [10], [11], which involves matrix multiplications and proximity operators in a high-dimensional Hilbert space. This approach is computationally efficient and has theoretical convergence guarantees to a globally optimal solution. On the other hand, for a generalization of the LiGME model, the DC (difference of convex functions) algorithm has been proposed [12], which solves more complex subproblems at each iteration. Although this algorithm converges faster in practice, it only guarantees subsequential convergence to an optimal solution, not whole sequence convergence. Thus, there remains a need for algorithms that ensure both global convergence and numerical efficiency.

In this paper, we propose a Douglas-Rachford splitting algorithm for the LiGME model (3), which numerically achieves faster convergence compared to Abe's algorithm while theoretically guaranteeing global convergence of the

iterates to the optimal solution.[1] Although the updates of the proposed algorithm involve inverse matrix computations, we reduce these computations to those for sparse matrices by introducing a special inner product on the Hilbert space under certain situations (more precisely, the matrix representation of A^*A, $\mathfrak{L}$, and $\mathfrak{L}^*$ are sparse). A numerical experiment in a piecewise-constant signal recovery demonstrates that the proposed method achieves significantly faster convergence.

II. Preliminaries

A. Douglas-Rachford algorithm

The Douglas-Rachford algorithm is a widely used method for finding a zero of the sum of two maximally monotone operators.[2] Its important building block is the *Douglas-Rachford splitting operator* [15, Proposition 26.1(c)] $T_{A,B}$ in Fact 1.

Fact 1: Let A and B be maximally monotone operators from $\mathcal{H} \to 2^{\mathcal{H}}$. Let $\gamma \in \mathbb{R}_{++}$, and let $\mathcal{P}$ be the set of solution of the problem to find $x \in \mathcal{H}$ such that $0_{\mathcal{H}} \in Ax + Bx$, and let[3]

$$T_{A,B} = \frac{R_A \circ R_B + \mathrm{Id}_{\mathcal{H}}}{2}. \tag{6}$$

Then the following hold:[4] (a) $T_{\gamma A,\gamma B}$ is firmly nonexpansive; (b) $\mathcal{P} = (\mathrm{Id} + \gamma B)^{-1}(\mathrm{Fix}\, R_{\gamma A} R_{\gamma B}) = (\mathrm{Id} + \gamma B)^{-1}(\mathrm{Fix}\, T_{\gamma A,\gamma B}) = T_{\mathrm{O}_{\mathcal{H}},\gamma B}(\mathrm{Fix}\, T_{\gamma A,\gamma B})$; (c) $\mathcal{P} \neq \varnothing \Leftrightarrow \mathrm{Fix}\, T_{\gamma A,\gamma B} \neq \varnothing$.

The Douglas-Rachford algorithm and its convergence properties are described in Fact 2, which is an adaptation of [15, Theorem 26.11] to the cases of finite dimensional Hilbert spaces instead of general real Hilbert spaces.

Fact 2: Let A, B, γ, and $\mathcal{P}$ as in Fact 1. Suppose $\mathrm{zer}(A+B) \neq \varnothing$. Let $(\lambda_k)_{k\in\mathbb{N}}$ be a sequence in $[0,2]$ such that $\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = +\infty$. Let $y_0 \in \mathcal{H}$, and

$$\text{for } k = 0, 1, \ldots \left\lfloor \begin{aligned} x_k &= T_{\mathrm{O}_{\mathcal{H}},\gamma B}(y_k), \\ z_k &= T_{\gamma A,\mathrm{O}_{\mathcal{H}}}(2x_k - y_k), \\ y_{k+1} &= y_k + \lambda_k(z_k - x_k). \end{aligned} \right. \tag{7}$$

Then there exists $y_\star \in \mathrm{Fix}\, R_{\gamma A} R_{\gamma B} = \mathrm{Fix}\, T_{\gamma A,\gamma B}$ such that $\lim_{k\to\infty} y_k = y_\star$. Now set $x_\star = T_{\mathrm{O}_{\mathcal{H}},\gamma B}(y_\star)$. Then the following hold: (i) $x_\star \in \mathcal{P}$. (ii) $\lim_{k\to\infty} \|x_k - z_k\|_{\mathcal{H}} = 0$. (iii) $\lim_{k\to\infty} x_k = x_\star$ and $\lim_{k\to\infty} z_k = x_\star$.

B. A brief review of LiGME model

Definition 1 (Linearly involved Generalized-Moreau-Enhanced (LiGME) Model [9]): Let $(\mathcal{X}, \langle\cdot,\cdot\rangle_{\mathcal{X}}, \|\cdot\|_{\mathcal{X}})$, $(\mathcal{Y}, \langle\cdot,\cdot\rangle_{\mathcal{Y}}, \|\cdot\|_{\mathcal{Y}})$, $(\mathcal{Z}, \langle\cdot,\cdot\rangle_{\mathcal{Z}}, \|\cdot\|_{\mathcal{Z}})$, and $(\widetilde{\mathcal{Z}}, \langle\cdot,\cdot\rangle_{\widetilde{\mathcal{Z}}}, \|\cdot\|_{\widetilde{\mathcal{Z}}})$ be finite dimensional real Hilbert spaces, $\Psi \in \Gamma_0(\mathcal{Z})$ coercive (i.e., $\|z\|_{\mathcal{Z}} \to \infty \;\Rightarrow\; \Psi(x) \to \infty$) with $\mathrm{dom}\,\Psi = \mathcal{Z}$, $B \in \mathcal{B}(\mathcal{Z}, \widetilde{\mathcal{Z}})$, $\mathfrak{L} \in \mathcal{B}(\mathcal{X}, \mathcal{Z})$, and $(A, \mathfrak{L}, \mu) \in \mathcal{B}(\mathcal{X}, \mathcal{Y}) \times \mathcal{B}(\mathcal{X}, \mathcal{Z}) \times \mathbb{R}_+$. Then:
(a) *GME penalty function* $\Psi_B \in \Gamma_0(\mathcal{Z})$ is defined as

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

(b) Linearly involved Generalized-Moreau-Enhanced (LiGME) penalty is defined as $\Psi_B \circ \mathfrak{L} \colon \mathcal{X} \to (-\infty, \infty]$.
(c) LiGME model is defined as the minimization of

$$J_{\Psi_B\circ\mathfrak{L}} \colon \mathcal{X} \to \mathbb{R} \colon x \mapsto \frac{1}{2}\|y - Ax\|^2_{\mathcal{Y}} + \mu\Psi_B \circ \mathfrak{L}(x). \tag{8}$$

Fact 3 (Selected properties of LiGME model [9]):
(a) (Overall convexity condition for the LiGME model [9, Proposition 1]) The GME penalty function Ψ_B in Definition 1 has the following properties: for the three conditions (C1) $A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L} \succeq \mathrm{O}_{\mathcal{X}}$, (C2) $J_{\Psi_B\circ\mathfrak{L}} \in \Gamma_0(\mathcal{X})$ for any $y \in \mathcal{Y}$, and (C3) $J^{(0)}_{\Psi_B\circ\mathfrak{L}} := \frac{1}{2}\|A\cdot\|^2_{\mathcal{Y}} + \mu\Psi_B \circ \mathfrak{L} \in \Gamma_0(\mathcal{X})$, the relation (C1)$\Rightarrow$(C2)$\Leftrightarrow$(C3) holds (Note: See [16] for details on the specific design method of B to satisfy (C1)).
(b) (Iterative approximation of a global minimizer of $J_{\Psi_B\circ\mathfrak{L}}$ [9, Theorem 1]) Let $\Psi \in \Gamma_0(\mathcal{Z})$ in Definition 1 be prox-friendly and even symmetry (i.e., $\Psi \circ (-\mathrm{Id}_{\mathcal{Z}}) = \Psi$). Assume the condition (C1) in (a) and the existence of a minimizer, of $J_{\Psi_B\circ\mathfrak{L}}$, in $\mathcal{X}$. Let $\mathcal{H} := \mathcal{X}\times\mathcal{Z}\times\mathcal{Z}$. Define $T_{\mathrm{LiGME}} \colon \mathcal{H} \to \mathcal{H} \colon (x, v, w) \mapsto (\xi + \frac{1}{\sigma}A^*y, \zeta, \eta)$ by

$$\begin{aligned} \xi &:= \left[\mathrm{Id} - \frac{1}{\sigma}(A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L})\right]x - \frac{\mu}{\sigma}\mathfrak{L}^*B^*Bv - \frac{\mu}{\sigma}\mathfrak{L}^*w, \\ \zeta &:= \mathrm{Prox}_{\frac{\mu}{\tau}\Psi}\left[\frac{2\mu}{\tau}B^*B\mathfrak{L}\xi - \frac{\mu}{\tau}B^*B\mathfrak{L}x + \left(\mathrm{Id} - \frac{\mu}{\tau}B^*B\right)v\right], \\ \eta &:= [\mathrm{Id} - \mathrm{Prox}_{\Psi}](2\mathfrak{L}\xi - \mathfrak{L}x + w), \end{aligned} \tag{9}$$

[1] The Douglas-Rachford splitting algorithm (see Sec. II-A) is known as one of the most prominent splitting algorithms [13]. In [8], the idea of using the Douglas-Rachford splitting algorithm as a solver for the least squares problem (2) with Ψ chosen as the MC penalty is mentioned. Inspired by this idea, we developed the proposed method.

[2] See Appendix for basic mathematical notations. The class of proper lower semicontinuous convex functions from $\mathcal{H}$ to $(-\infty, +\infty]$ is denoted by $\Gamma_0(\mathcal{H})$.
Let $A \colon \mathcal{H} \to 2^{\mathcal{K}}$. Denote its *graph* by $\mathrm{gra}\, A = \{(x,u) \in \mathcal{H}\times\mathcal{K} \mid u \in Ax\}$. The *inverse* of A, denoted by A^{-1}, is defined through its graph $\mathrm{gra}\, A^{-1} = \{(u,x) \in \mathcal{K}\times\mathcal{H} \mid (x,u) \in \mathrm{gra}\, A\}$. The set of zeros of A is $\mathrm{zer}\, A = A^{-1}0_{\mathcal{H}} = \{x \in \mathcal{H} \mid 0_{\mathcal{H}} \in Ax\}$.
Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$. The operator A is *monotone* if

$$(\forall (x,u) \in \mathrm{gra}\, A)\,(\forall (y,v) \in \mathrm{gra}\, A) \quad \langle x - y, u - v\rangle_{\mathcal{H}} \geq 0. \tag{4}$$

A monotone operator A is *maximally monotone* (or *maximal monotone*) if there exists no monotone operator $B \colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\mathrm{gra}\, B$ properly contains $\mathrm{gra}\, A$, i.e., for every $(x,u) \in \mathcal{H}\times\mathcal{H}$,

$$(x,u) \in \mathrm{gra}\, A \;\Leftrightarrow\; (\forall (y,v) \in \mathrm{gra}\, A)\, \langle x - y, u - v\rangle_{\mathcal{H}} \geq 0. \tag{5}$$

It is well-known that ∂f is maximally monotone for $f \in \Gamma_0(\mathcal{H})$ [14] (See Also [15, Theorem 20.25]), where ∂f denotes the subdifferential of f, defined as the set-valued operator

$$\partial f \colon \mathcal{H} \to 2^{\mathcal{H}} \colon x \mapsto \{u \in \mathcal{H} \mid (\forall z \in \mathcal{H})\langle z - x, u\rangle_{\mathcal{H}} + f(x) \leq f(z)\}.$$

Every element $u \in \partial f(x)$ is called a subgradient of f at x.

[3] Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$. The resolvent of A is $(\mathrm{Id}_{\mathcal{H}} + A)^{-1}$ and the refrected resolvent is $R_A \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto 2(\mathrm{Id}_{\mathcal{H}} + A)^{-1}x - x$.

[4] Let $T \colon \mathcal{H} \to \mathcal{H}$. The set of fixed points of T is denoted by $\mathrm{Fix}\, T$, i.e., $\mathrm{Fix}\, T = \{x \in \mathcal{H} \mid Tx = x\}$. The operator T is *nonexpansive* if it is Lipschitz continuous with constant 1, i.e., $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|Tx - Ty\|_{\mathcal{H}} \leq \|x - y\|_{\mathcal{H}}$; The operator T is *firmly nonexpansive* if $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|Tx - Ty\|^2_{\mathcal{H}} + \|(\mathrm{Id}_{\mathcal{H}} - T)x - (\mathrm{Id}_{\mathcal{H}} - T)y\|^2_{\mathcal{H}} \leq \|x - y\|^2_{\mathcal{H}}$. Note that the operator T is *firmly nonexpansive* if and only if $R := 2T - \mathrm{Id}$ is nonexpansive.

where[5] $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, +\infty)$ is chosen to satisfy[6]

$$\begin{cases} \sigma \mathrm{Id} - \dfrac{\kappa}{2} A^*A - \mu \mathfrak{L}^* \mathfrak{L} \succ \mathrm{O}_{\mathcal{X}}, \\ \tau \geq \left(\frac{\kappa}{2} + \frac{2}{\kappa}\right) \mu \|B\|_{\mathrm{op}}^2. \end{cases} \tag{11}$$

Then, for any initial point $(x_0, v_0, w_0) \in \mathcal{H}$, the sequence $(x_k)_{k\in\mathbb{N}} \subset \mathcal{X}$, in $(x_k, v_k, w_k)_{k\in\mathbb{N}} \subset \mathcal{H}$ generated by

$$(x_{k+1}, v_{k+1}, w_{k+1}) = T_{\mathrm{LiGME}}(x_k, v_k, w_k) \tag{12}$$

converges to a global minimizer of $J_{\Psi_B \circ \mathfrak{L}}$.

III. Proposed Method: A Douglas-Rachford Splitting Algorithm for the LiGME model

We shall derive an algorithmic solution for the following constrained problem.

Problem 1: For the function $J_{\Psi_B \circ \mathfrak{L}}$ in (8) and a closed convex constraint set $C \subset \mathcal{X}$, suppose that $\operatorname{argmin}(J_{\Psi_B \circ \mathfrak{L}} + \iota_C) \neq \varnothing$ holds[7]. Assume that the overall convexity condition $A^*A - \mu \mathfrak{L}^* B^* B \mathfrak{L} \succeq \mathrm{O}_{\mathcal{X}}$ (i.e., (C1) in Fact 3) hold. Then, find a solution of

$$\mathrm{Minimize}_{x\in\mathcal{X}} \ \ J_{\Psi_B \circ \mathfrak{L}}(x) + \iota_C(x). \tag{13}$$

For this problem, we denote the solution set by $\mathcal{S}$.

Remark 1 (On Problem 1):
(a) The existence of a solution to Problem 1 is guaranteed, for instance, for the cases where $J_{\Psi_B \circ \mathfrak{L}} + \iota_C$ is proper and coercive (See, e.g., [15, Section 11.4]). More specific conditions for the existence of a solution in the Euclidean spaces are discussed in [17]. Additionally, in the special case where $(\Psi, \mathfrak{L}) := (\|\cdot\|_1, \mathrm{Id}_{\mathcal{Z}})$, an analysis of the solution path is provided in [18].
(b) The simultaneous use of the LiGME penalty and closed convex constraints has been proposed in [19]–[21]. □

To derive a Douglas-Rachford algorithm for Problem 1, we reformulate Problem 1 as finding a zero of the sum of two carefully constructed maximally monotone operators, $\mathscr{A}$ and $\mathscr{B}$ (in Theorem 1 below) defined in a higher-dimensional space

$$(\mathcal{H}_\mu := \mathcal{X} \times \mathcal{Z} \times \mathcal{Z}, \langle\cdot,\cdot\rangle_{\mathcal{H}_\mu} := \langle\cdot,\cdot\rangle_{\mathrm{Id}_{\mathcal{X}} \times \mathrm{Id}_{\mathcal{Z}} \times \mu \mathrm{Id}_{\mathcal{Z}}}, \|\cdot\|_{\mathcal{H}_\mu})$$

with a slightly modified inner product. This inner product is designed to enhance the sparsity of the matrices that require matrix inversion in the iterative computations of the resulting algorithm, particularly in cases where A^*A, $\mathfrak{L}$, and $\mathfrak{L}^*$ are sparse (For further details, see Remark 2 below).

Theorem 1: For Problem 1, define an affine operator

$$\begin{aligned} \mathscr{A}\colon \mathcal{H}_\mu \to \mathcal{H}_\mu\colon (x, v, w) \\ \mapsto \big([A^*A - \mu \mathfrak{L}^* B^* B \mathfrak{L}]x - A^*y + \mu \mathfrak{L}^* B^* B v + \mu \mathfrak{L}^* w, \\ \mu B^* B(v - \mathfrak{L}x), -\mathfrak{L}x\big) \end{aligned} \tag{14}$$

and a set-valued operator[8] $\mathscr{B}\colon \mathcal{H}_\mu \to 2^{\mathcal{H}_\mu}$:

$$(x, v, w) \mapsto \partial \iota_C(x) \times \mu \partial \Psi(v) \times \partial \Psi^*(w). \tag{15}$$

Then

(a) $\mathscr{A}$ and $\mathscr{B}$ are maximally monotone in $(\mathcal{H}_\mu, \langle\cdot,\cdot\rangle_{\mathcal{H}_\mu}, \|\cdot\|_{\mathcal{H}_\mu})$.
(b) $\Xi_{\mathcal{X}} \operatorname{zer}(\mathscr{A} + \mathscr{B}) = \mathcal{S}$, where $\Xi_{\mathcal{X}}\colon \mathcal{H}_\mu \to \mathcal{X}\colon (x, v, w) \mapsto x$.

Theorem 1(b) clarifies that Problem 1 can be reduced to the problem of finding a point in $\operatorname{zer}(\mathscr{A} + \mathscr{B})$, i.e., the zero of a sum of maximally monotone operators.

Problem 2: For two maximally monotone operators $\mathscr{A}$ and $\mathscr{B}$ defined in Theorem 1, consider the problem to

$$\text{find } x \in \mathcal{H}_\mu \text{ such that } 0 \in \mathscr{A}x + \mathscr{B}x, \tag{16}$$

in other words, find a point in $\operatorname{zer}(\mathscr{A} + \mathscr{B})$ (Note: $\operatorname{zer}(\mathscr{A} + \mathscr{B}) \neq \varnothing$ is guaranteed if and only if Problem 1 has a solution, i.e., $\mathcal{S} \neq \varnothing$, because of Theorem 1(b)).

For Problem 2, we can utilize the Douglas-Rachford splitting operator in Fact 1 to characterize the solution set $\operatorname{zer}(\mathscr{A} + \mathscr{B})$ and the Douglas-Rachford algorithm in Fact 2 as an iterative algorithm. These ideas lead to the proposed operator $T_{\mathscr{A},\mathscr{B}}$ in Theorem 2 below and the proposed algorithm for Problem 2 in Theorem 3 below.

Theorem 2 (Douglas-Rachford splitting operator for Problem 2): Let $\gamma \in \mathbb{R}_{++}$. The Douglas-Rachford splitting operator $T_{\mathscr{A},\mathscr{B}}$ for Problem 2 satisfies:
(a) $T_{\gamma\mathscr{A},\gamma\mathscr{B}}$ is firmly nonexpansive in $(\mathcal{H}_\mu, \langle\cdot,\cdot\rangle_{\mathcal{H}_\mu}, \|\cdot\|_{\mathcal{H}_\mu})$;
(b) $\operatorname{zer}(\mathscr{A} + \mathscr{B}) = T_{\mathrm{O}_{\mathcal{H}_\mu},\gamma\mathscr{B}}(\operatorname{Fix} T_{\gamma\mathscr{A},\gamma\mathscr{B}})$ and $\mathcal{S} = \Xi_{\mathcal{X}} T_{\mathrm{O}_{\mathcal{H}_\mu},\gamma\mathscr{B}}(\operatorname{Fix} T_{\gamma\mathscr{A},\gamma\mathscr{B}})$;
(c) $\operatorname{zer}(\mathscr{A} + \mathscr{B}) \neq \varnothing \Leftrightarrow \operatorname{Fix} T_{\gamma\mathscr{A},\gamma\mathscr{B}} \neq \varnothing$;
(d) The operator $T_{\gamma\mathscr{A},\mathrm{O}_{\mathcal{H}_\mu}}$ can be represented as

$$T_{\gamma\mathscr{A},\mathrm{O}_{\mathcal{H}_\mu}}(x, v, w) = (\mathrm{Id}_{\mathcal{H}_\mu} + \gamma\mathcal{M})^{-1}(x + \gamma A^*y, v, w)$$

with the linear part of $\mathscr{A}$ in (14), i.e., $\mathcal{M}\colon \mathcal{H}_\mu \to \mathcal{H}_\mu\colon (x, v, w) \mapsto \mathscr{A}(x, v, w) + (A^*y, 0_{\mathcal{Z}}, 0_{\mathcal{Z}})$, and the operator $T_{\mathrm{O}_{\mathcal{H}_\mu},\gamma\mathscr{B}}$ can be represented as

$$\begin{aligned} &T_{\mathrm{O}_{\mathcal{H}_\mu},\gamma\mathscr{B}}(x, v, w) \\ &= \left(P_C(x), \operatorname{Prox}_{\mu\gamma\Psi}(v), w - \gamma \operatorname{Prox}_{\gamma^{-1}\Psi}(\gamma^{-1}w)\right). \end{aligned}$$

Remark 2: **(Sparsity promotion in the calculation of $(\mathrm{Id}_{\mathcal{H}_\mu} + \gamma\mathcal{M})^{-1}$)** In the cases where the matrix representation of A^*A, $\mathfrak{L}$, and $\mathfrak{L}^*$ are sparse (whose many entries are zero), the calculation of $(\mathrm{Id}_{\mathcal{H}_\mu} + \gamma\mathcal{M})^{-1}$, required for the resolvent

[5]The proximity operator of $f \in \Gamma_0(\mathcal{H})$ is defined by $\operatorname{Prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname{argmin}_{z\in\mathcal{H}} f(z) + \frac{1}{2}\|z - x\|^2$. It is also well known that Prox_f is nothing but the resolvent of ∂f, i.e., $\operatorname{Prox}_f = (\mathrm{I} + \partial f)^{-1}$.

[6]For example, for any $\kappa \in (1, +\infty)$ and $\delta > 0$, the selection

$$\begin{aligned} \sigma &= \left(\text{the maximum eigenvalue of } \frac{\kappa}{2} A^*A - \mu \mathfrak{L}^* \mathfrak{L}\right) + \delta, \\ \tau &= \left(\frac{\kappa}{2} + \frac{2}{\kappa}\right) \mu \|B\|_{\mathrm{op}}^2, \end{aligned} \tag{10}$$

satisfies the condition (11).

[7]For a nonempty closed convex set $C \subset \mathcal{H}$, the indicator function of C is defined by

$$\iota_C\colon \mathcal{H} \to (-\infty, \infty]\colon x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{otherwise,} \end{cases}$$

which belongs to $\Gamma_0(\mathcal{H})$.

[8]For $f\colon \mathcal{H} \to [-\infty, +\infty]$, its *conjugate* is

$$f^*\colon \mathcal{H} \to [-\infty, +\infty]\colon u \mapsto \sup_{x\in\mathcal{H}} \left(\langle x, u\rangle_{\mathcal{H}} - f(x)\right).$$

Algorithm 1: Douglas-Rachford algorithm for Problem 2

Input: A, $\mathfrak{L}$, μ, B, Ψ, C in Problem 1, $(\lambda_k)_{k\in\mathbb{N}} \subset [0,2]$ such that $\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = +\infty$, $\gamma > 0$, initial $(s_0, t_0, u_0) \in \mathcal{H}_\mu$.

for $k = 0, 1, \ldots$

1. Compute
   $x_k = P_C(s_k)$
   $v_k = \mathrm{Prox}_{\mu\gamma\Psi}(t_k)$
   $w_k = u_k - \gamma \,\mathrm{Prox}_{\gamma^{-1}\Psi}(\gamma^{-1}u_k)$
2. Find the unique solution $(s_{k+1/2}, t_{k+1/2}, u_{k+1/2})$ of a system of linear equations
   $(\mathrm{Id}_{\mathcal{H}_\mu} + \gamma\mathcal{M})(s_{k+1/2}, t_{k+1/2}, u_{k+1/2}) = (2x_k - s_k + \gamma A^* y, 2v_k - t_k, 2w_k - u_k)$,
   where $\mathcal{M} = \begin{pmatrix} A^*A - \mu\mathfrak{L}^*B^*B\mathfrak{L} & \mu\mathfrak{L}^*B^*B & \mu\mathfrak{L}^* \\ -\mu B^*B & \mu B^*B & \mathrm{O}_{\mathcal{Z}} \\ -\mathfrak{L} & \mathrm{O}_{\mathcal{Z}} & \mathrm{O}_{\mathcal{Z}} \end{pmatrix}$.
3. Compute
   $(s_{k+1}, t_{k+1}, u_{k+1}) = (s_k, t_k, u_k) + \lambda_k \left((s_{k+1/2}, t_{k+1/2}, u_{k+1/2}) - (x_k, v_k, w_k)\right)$.

operator $T_{\gamma\mathscr{A},\mathrm{O}_{\mathcal{H}_\mu}}$ in Theorem 2(d), reduces to the calculation of the inverse of

$$\begin{pmatrix} \mathrm{Id}_{\mathcal{X}} + \gamma A^*A & (\gamma^2\mu - 1)\mathfrak{L}^* & \gamma\mu\mathfrak{L}^* \\ \mathrm{O}_{\mathcal{B}(\mathcal{X},\mathcal{Z})} & \mathrm{Id}_{\mathcal{Z}} & -\mu B^*B \\ -\gamma\mathfrak{L} & \gamma\mathrm{Id}_{\mathcal{Z}} & \mathrm{Id}_{\mathcal{Z}} \end{pmatrix} =: Q \qquad (17)$$

whose the matrix representation is more sparse[9] compared to the matrix representation of $\mathrm{Id}_{\mathcal{H}_\mu} + \gamma\mathcal{M}$, through the factorization $\left(\mathrm{Id}_{\mathcal{H}_\mu} + \gamma\mathcal{M}\right)^{-1} = M_L \circ Q^{-1} \circ M_R$, where $M_L \colon \mathcal{H}_\mu \to \mathcal{H}_\mu \colon (x, v, w) \mapsto (x, v, \gamma v + w)$ and $M_R \colon \mathcal{H}_\mu \to \mathcal{H}_\mu \colon (x, v, w) \mapsto (x - \mathfrak{L}^*v, v - \mu B^*w, w)$. □

Theorem 3 (Proposed Algorithm: Douglas-Rachford algorithm for Problem 2): Consider Problem 2. Let $(\lambda_k)_{k\in\mathbb{N}} \subset [0,2]$ satisfy $\sum_{k\in\mathbb{N}} \lambda_k(2-\lambda_k) = +\infty$ and let $\gamma > 0$. For any initial $(s_0, t_0, u_0) \in \mathcal{H}_\mu$, set

for $k = 0, 1, \ldots$ (18)

$$\begin{aligned} &(x_k, v_k, w_k) = T_{\mathrm{O}_{\mathcal{H}_\mu},\gamma\mathscr{B}}(s_k, t_k, u_k), \\ &(s_{k+1/2}, t_{k+1/2}, u_{k+1/2}) \\ &\qquad = T_{\gamma\mathscr{A},\mathrm{O}_{\mathcal{H}_\mu}}(2x_k - s_k, 2v_k - t_k, 2w_k - u_k), \\ &(s_{k+1}, t_{k+1}, u_{k+1}) = (s_k, t_k, u_k) \\ &\qquad + \lambda_k \left((s_{k+1/2}, t_{k+1/2}, u_{k+1/2}) - (x_k, v_k, w_k)\right). \end{aligned}$$

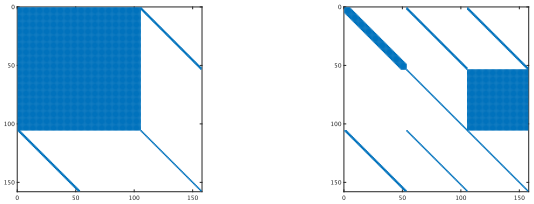
Then the sequence $(x_k, v_k, w_k)_{k\in\mathbb{N}}$ converges to a point in $\mathrm{zer}(\mathscr{A} + \mathscr{B})$ and the sequence $(x_k)_{k\in\mathbb{N}} \subset C$ converges to a point in $\mathcal{S}$.

IV. Numerical Experiment in Piecewise-Constant Signal Recovery

We conduct a numerical experiment to evaluate the convergence speed of the proposed algorithm in Theorem 3 for the LiGME model (3).

(Problem Settings) The goal is to recover a piecewise-constant signal $\boldsymbol{x}_{\mathrm{True}} \in \mathcal{X} = \mathbb{R}^{60}$ from blurred observations $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}_{\mathrm{True}} + \boldsymbol{\epsilon}$ with additve noise $\boldsymbol{\epsilon} \in \mathcal{Y} = \mathbb{R}^{56}$, where

[9]For sparse matrices, efficient methods for calculating the inverse matrix have been proposed. One class of such methods is the so-called direct method (see, e.g., [22] and references therein), which utilizes matrix factorization. For large-scale sparse matrices, Krylov subspace methods (see, e.g., [23] and references therein) are promising, as they compute the result of the inverse matrix multiplication without explicitly constructing the inverse.



(a) $\mathrm{Id} + \mathcal{M}$ in Algorithm 1. (b) Q in (17).

Fig. 1. The sparsity pattern of (a) $\mathrm{Id} + \mathcal{M}$ in Algorithm 1 and (b) Q in (17) where nonzero entries are colored blue while zero values are white. Q is more sparse compared with $\mathrm{Id} + \mathcal{M}$. In fact, Q has 3629 nonzero entries while $\mathrm{Id} + \mathcal{M}$ has 11285 nonzero ones.

$\boldsymbol{A} \in \mathcal{B}(\mathcal{X}, \mathcal{Y}) = \mathbb{R}^{56\times60}$ is a blurring matrix with blur kernel whose kernel size is 5 and entries of the kernel are drawn from the uniform distribution in the interval $(0, 1)$. The signal-to-noise ratio (SNR) is chosen to be 30dB, which is defined as $\mathrm{SNR} : 10\log_{10}\frac{\|\boldsymbol{A}\boldsymbol{x}_{\mathrm{True}}\|_{\mathcal{Y}}^2}{\|\boldsymbol{\epsilon}\|_{\mathcal{Y}}^2}$ [dB].

In order to emphasize the piecewise-constantness of the signal estimated by the LiGME model (3), let $\Psi = \|\cdot\|_1$ and $\mathfrak{L}$ be the first-order difference operator

$$\boldsymbol{D} := \begin{pmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{59\times60}.$$

In this case, where $\mathfrak{L} = \boldsymbol{D}$, $\boldsymbol{B}_\theta = \sqrt{\theta/\mu}\left(\boldsymbol{I}_{56} - \frac{\boldsymbol{A}\boldsymbol{1}\boldsymbol{1}^\top\boldsymbol{A}^\top}{\boldsymbol{1}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{1}}\right)\boldsymbol{A}\acute{\boldsymbol{R}}^{\boldsymbol{D}}$ satisfies the overall convexity condition[10] in (C1), i.e., $\boldsymbol{A}^\top\boldsymbol{A} - \mu\mathfrak{L}^\top\boldsymbol{B}_\theta^\top\boldsymbol{B}_\theta\mathfrak{L} \succeq \boldsymbol{O}_{60}$, where $\boldsymbol{1} := (1, 1, \ldots, 1)^\top \in \mathbb{R}^{60}$ and $\acute{\boldsymbol{R}}^{\boldsymbol{D}} \in \mathbb{R}^{60\times59}$ is a matrix whose diagonal and upper triangular elements are -1, and all other elements are 0. Here, assuming no other prior information on $\boldsymbol{x}_{\mathrm{True}}$, we set $C = \mathcal{X}$.

The parameter μ is chosen as $\mu = 1.2$ for the case of $\boldsymbol{B} = \boldsymbol{B}_\theta$ and $\mu = 0.1$ for the case of $\boldsymbol{B} = \boldsymbol{O}_{53\times53}$.

(Algorithm settings) We compare three algorithms: Proposed algorithm (in Theorem 3; $\lambda_k = 1, \gamma = 1$), Abe's algorithm (in Fact 3(b); $\kappa = 2$ and σ, τ in (10) with $\delta = 10^{-5}$; referred to as Abe (LiGME)), and Abe's algorithm (with the same settings except for $\boldsymbol{B} = \boldsymbol{O}_{53\times53}$; referred to as Abe (TV)) for the convex TV penalized LS problem, used as a benchmark in this numerical experiment. Their performance is evaluated based on the convergence speed and accuracy of the estimates.

Here, we present the effect of sparsity promotion as discussed in Remark 2 (Note: In this numerical experiment, $\boldsymbol{A}^\top\boldsymbol{A}$, $\mathfrak{L}$, and $\mathfrak{L}^\top$ are sparse, while $\boldsymbol{B}_\theta^\top\boldsymbol{B}_\theta$ is dense). Fig. 1 illustrates the locations of the non-zero elements in $\mathrm{Id} + \mathcal{M}$ and Q. Clearly, the block matrix containing $\boldsymbol{B}_\theta^\top\boldsymbol{B}_\theta$ becomes dense, while the other blocks remain sparse. Thus Q is more sparse compared with $\mathrm{Id} + \mathcal{M}$.

(Results) Fig. 2 shows that the proposed algorithm demonstrates significantly faster convergence compared to Abe

[10]This novel design of $\boldsymbol{B}_\theta$ is derived by [16, Theorem 1] with $\boldsymbol{L} = \boldsymbol{I}_{59}$ and $\boldsymbol{R} = \left(\begin{array}{c|c} \acute{\boldsymbol{R}}^{\boldsymbol{D}} & \boldsymbol{1} \end{array}\right)$. The proof that $\boldsymbol{B}_\theta$ satisfies the overall convexity condition will be reported elsewhere.

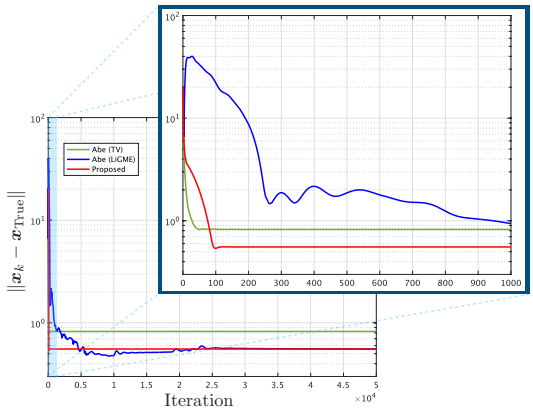


Fig. 2. Performance comparison of the proposed algorithm, Abe (LiGME), and Abe (TV), and the enlarged view of the first 1000 iterations.

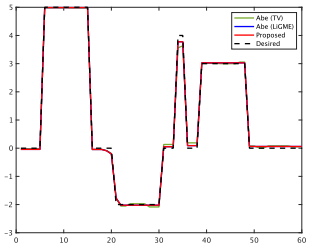


Fig. 3. Estimation results of $\boldsymbol{x}_{\rm true}$ by all the algorithms. The results of Abe (LiGME) and Proposed overlap.

(LiGME). It rapidly decreases the distance to the desired solution, reaching an accurate estimate in significantly fewer iterations. Abe (LiGME) shows a slower convergence, requiring more than 30,000 iterations to achieve a similar accuracy as the proposed algorithm.[11] While Abe (TV) converges quickly, its final accuracy is lower than both the proposed algorithm and Abe (LiGME).

Fig. 3 depicts the estimation results for all the algorithms after $50,000$ iterations. The results of the proposed algorithm and Abe (LiGME) are almost identical, and the two lines completely overlap in the figure. The proposed algorithm provides an estimate closer to the true signal $\boldsymbol{x}_{\rm True}$ compared to Abe (TV). While Abe (TV) exhibits fluctuations between 20–30th entries, and errors between 31–38th entries, these issues are resolved in the estimates obtained using the LiGME model, including the proposed algorithm.

Appendix: Basic Mathematical Notations

Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_+$, and $\mathbb{R}_{++}$ be the sets of natural numbers, real numbers, nonnegative real numbers, and positive real numbers, respectively. The superscript $(\cdot)^\top$ denotes transpose. For a vector $x := (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, we use $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ $(0 < p < \infty)$, and $\|x\|_0 := \#\{i \in \mathbb{N} \cap [1, n] \mid x_i \neq 0\}$.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ and $(\mathcal{K}, \langle\cdot,\cdot\rangle_{\mathcal{K}}, \|\cdot\|_{\mathcal{K}})$ be finite dimensional real Hilbert spaces. $\mathcal{B}(\mathcal{H}, \mathcal{K})$ denotes the set of all bounded linear operators[12] from $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}, \|\cdot\|_{\mathcal{H}})$ to $(\mathcal{K}, \langle\cdot,\cdot\rangle_{\mathcal{K}}, \|\cdot\|_{\mathcal{K}})$. For $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, we use $\|L\|_{\rm op} := \sup_{x\in\mathcal{H}:\, \|x\|_{\mathcal{H}}\leq 1} \|Lx\|_{\mathcal{K}}$. For $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, $L^* \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ denotes the adjoint of L, i.e., $\langle Lx, y\rangle_{\mathcal{K}} = \langle x, L^*y\rangle_{\mathcal{H}}$ $(\forall (x, y) \in \mathcal{H}\times\mathcal{K})$. We also use Id (or $\mathrm{Id}_{\mathcal{H}}$) to denote the identity operator for general Hilbert spaces. $\mathrm{O}_{\mathcal{B}(\mathcal{H},\mathcal{K})} \in \mathcal{B}(\mathcal{H}, \mathcal{K})$ and $\mathrm{O}_{\mathcal{H}} \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ stand for the zero operators. For $L \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, $L^\dagger \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ stands for the Moore-Penrose pseudo inverse of L, $\mathrm{ran}(L) := \{Lx \in \mathcal{K} \mid x \in \mathcal{H}\}$ denotes respectively the range spaces of L. The positive definiteness and positive semidefiniteness of a self-adjoint operator $L \in \mathcal{B}(\mathcal{H}, \mathcal{H})$ are expressed respectively as $L \succ \mathrm{O}_{\mathcal{H}}$ and $L \succeq \mathrm{O}_{\mathcal{H}}$. For any $L \succ \mathrm{O}_{\mathcal{H}}$, by defining an inner product $\langle\cdot,\cdot\rangle_L$: $\mathcal{H}\times\mathcal{H} \to \mathbb{R}\colon (x, y) \mapsto \langle x, Ly\rangle_{\mathcal{H}}$ and its induced norm $\|x\|_L := \sqrt{\langle x, x\rangle_L}$, $(\mathcal{H}, \langle\cdot,\cdot\rangle_L, \|x\|_L)$ becomes a real Hilbert space. We use $\boldsymbol{I}_n \in \mathbb{R}^{n\times n}$ to denote the identity matrix for $\mathbb{R}^n$. $\boldsymbol{O}_{m,n} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{O}_n \in \mathbb{R}^{n\times n}$ stand for the zero matrices.

[11] Abe (LiGME) passes through slightly better candidate solutions between 5000 and 20000 iterations, suggesting that the so-called early-stopping technique could be effectively utilized, which is beyond the scope of this paper.

[12] In a real finite dimensional Hilbert space, $\mathcal{B}(\mathcal{H}, \mathcal{K})$ is identical to the set of all linear operators.

References

[1] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20:89–97, 2004.

[2] J. L. Starck, F. Murtagh, and J. Fadili. *Sparse Image and Signal Processing: Wavelets and Related Geometric Multiscale Analysis*. Cambidge University Press, 2015.

[3] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Statist. Soc. B*, 58:267–288, 1996.

[4] J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[5] D. Peleg and R. Meir. A bilinear formulation for vector sparsity optimization. *Signal Processing*, 88(2):375–389, 2008.

[6] Z.-Y. Wang, H. C. So, and A. M. Zoubir. Robust low-rank matrix recovery via hybrid ordinary-Welsch function. *IEEE Trans. Signal Process.*, 71:2548–2563, 2023.

[7] C.-H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

[8] I. Selesnick. Sparse regularization via convex analysis. *IEEE Trans. Signal Process.*, 65(17):4481–4494, 2017.

[9] J. Abe, M. Yamagishi, and I. Yamada. Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition. *Inverse Problems*, 36(3):035012, 2020.

[10] L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158:460–479, 2013.

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

[12] Y. Zhang and I. Yamada. A unified framework for solving a general class of nonconvexly regularized convex models. *IEEE Trans. Signal Process.*, 71:3518–3533, 2023.

[13] W.M. Moursi. The range of the Douglas–Rachford operator in infinite-dimensional Hilbert spaces. *Arxiv*, 2022.

[14] J.J. Moreau. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93:273–299, 1965.

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

[16] Y. Chen, M. Yamagishi, and I. Yamada. A unified design of generalized Moreau enhancement matrix for sparsity aware LiGME models. *IEICE Trans. Fundam.*, E106-A(8):1025–1036, 2023.

[17] A. Lanza, S. Morigi, I. W. Selesnick, and F. Sgallari. Sparsity-inducing nonconvex nonseparable regularization for convex image processing. *SIAM Journal on Imaging Sciences*, 12(2):1099–1134, 2019.

[18] Y. Zhang and I. Yamada. Piecewise linearity of min-norm solution map of a nonconvexly regularized convex sparse model. *Arxiv*, page 53pp., 2024.

[19] D. Kitahara, R. Kato, H. Kuroda, and A. Hirabayashi. Multicontrast csmri using common edge structures with ligme model. In *EUSIPCO*, pages 2119–2123, 2021.

[20] W. Yata and I. Yamada. Imposing early and asymptotic constraints on LiGME with application to nonconvex enhancement of fused lasso models. In *IEEE ICASSP*, page 5pp., 2024.

[21] W. Yata, M. Yamagishi, and I. Yamada. A constrained LiGME model and its proximal splitting algorithm under overall convexity condition. *J. Appl. Numer. Optim.*, 4(2):245–271, 2022.

[22] T. A. Davis, S. Rajamanickam, and W. M. Sid-Lakhdar. A survey of direct methods for sparse linear systems. *Acta Numerica*, 25:383–566, 2016.

[23] Y. Saad. *Iterative methods for sparse linear systems*. SIAM, 2003.