2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)

# A Discrete-Valued Signal Estimation by Nonconvex Enhancement of SOAV with cLiGME Model

Satoshi Shoji, Wataru Yata, Keita Kume and Isao Yamada
Dept. of Information and Communication Engineering, Tokyo Institute of Technology
E-mail: {shoji, yata, kume, isao}@sp.ict.e.titech.ac.jp

***Abstract*—In this paper, for the discrete-valued signal estimation, we propose a regularized least squares model but with a nonconvex enhancement of the so-called SOAV convex regularizer. To design more contrastive regularizers whose minima correspond to desired discrete values, we propose a class of nonconvex functions with Generalized Moreau Enhancement (GME) of the weighted $\ell_1$-norm. Promisingly, by tuning properly the design parameters of the proposed GME regularizers, (i) we can make the nonconvexly-regularized least squares model convex; and (ii) we can use an iterative algorithm for finding a global minimizer of the proposed model. We also propose a pair of simple technical improvements, of the proposed algorithm, called respectively a generalized superiorization and an iterative reweighting. Numerical experiments demonstrate the effectiveness of the proposed model and algorithms in a scenario of MIMO signal detection.**

## I. Introduction

Many tasks in signal processing, including digital communication and discrete-valued control [1]–[7], have been formulated as the following discrete-valued estimation problem:

**Problem I.1** (A discrete-valued estimation problem)**.**

$$\text{Find } \mathbf{x}^\star \in \mathfrak{D} \text{ such that } \mathbf{y} = \mathbf{A}\mathbf{x}^\star + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathfrak{D} := \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{L^N}\} = \mathfrak{A}^N := \{a_1, a_2, \ldots, a_L\}^N \subset \mathbb{R}^N$, $\mathbf{y} \in \mathbb{R}^M$ is an observed vector, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is a known matrix, and $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ is noise (Note: the complex version of this problem can also be formulated as Problem I.1 essentially via simple $\mathbb{C} \rightleftarrows \mathbb{R}^2$ translation (see Appendix A)).

Problem I.1 is a special instance of the mixed integer programming [8], but a direct application of naive solvers for the mixed integer programming leads to exponential computational complexity in $N$. From a practical viewpoint, continuous optimization approaches [9]–[13] have been utilized for Problem I.1 as computationally efficient alternatives. For example, Problem I.1 has been tackled with projection of $\mathbf{x}^\diamond$ onto $\mathfrak{D}$ after solving a relaxed continuous optimization problem:

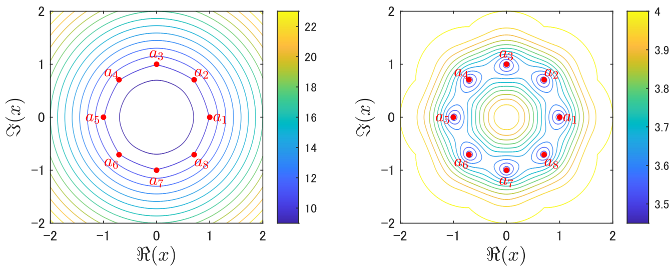
**Scheme 1** (A scheme for (1) via regularized least squares)**.**
**Step 1:**

$$\text{Find } \mathbf{x}^\diamond \in \underset{\mathbf{x}\in\widetilde{\mathfrak{D}}\subset\mathbb{R}^N}{\operatorname{argmin}}\ J_\Theta(\mathbf{x}) := \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mu\Theta(\mathbf{x}), \tag{2}$$

where the constraint set $\widetilde{\mathfrak{D}} \supset \mathfrak{D}$ is chosen usually as a connected subset (e.g., the convex hull of $\mathfrak{D}$) of $\mathbb{R}^N$, $\Theta : \mathbb{R}^N \to \mathbb{R}$ is a regularizer, and $\mu > 0$ is a regularization parameter.
**Step 2:** Assign $\mathbf{x}^\diamond \in \widetilde{\mathfrak{D}} \subset \mathbb{R}^N$ to the final estimate

$$\mathbf{x}^\natural = P_{\mathfrak{D}}(\mathbf{x}^\diamond) \in \underset{\mathbf{s}\in\mathfrak{D}}{\operatorname{argmin}}\|\mathbf{s}-\mathbf{x}^\diamond\|_2$$

of $\mathbf{x}^\star \in \mathfrak{D}$ in (1), where $P_{\mathfrak{D}} : \mathbb{R}^N \to \mathfrak{D} : \mathbf{x} \mapsto P_{\mathfrak{D}}(\mathbf{x})$ is defined to choose randomly one of nearest vectors from $\mathbf{x}$.



(a) (b)

Fig. 1: Illustrations of the values of $\Theta$ in the case where $\mathfrak{D} = \mathfrak{A} := \{a_l := \exp[j(l-1)\pi/4] \mid l = 1, 2, \ldots, 8 =: L\} \subset \mathbb{C} \equiv \mathbb{R}^2$, $\omega_{l,1} = 1/8$ $(l = 1, 2, \ldots, 8)$, and $N = 1$. (a) $\Theta^{\langle 1\rangle}_{\text{SOAV}}(\mathbf{x})$, (b) A proposed nonconvex enhancement $\Theta_{\text{GME}}$ of $\Theta^{\langle 1\rangle}_{\text{SOAV}}$ by GME (5) with $\mathbf{B}^{\langle l\rangle} = \mathbf{I}$ $(l = 1, 2, \ldots, 8)$.

In (2), the first term $\frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2$ is called a data-fidelity term evaluating the consistency with the linear regression model in (1), while the second term $\mu\Theta : \mathbb{R}^N \to \mathbb{R}$ is called a regularization term designed strategically based on a prior knowledge on $\mathbf{x}^\star$.

For achieving an acceptable estimation of a signal in Problem I.1, various prior knowledge, e.g., statistical properties, has been exploited for designing $\Theta$ in (2). For example, in a scenario of MIMO signal detection [14], the regularizer $\Theta$ in (2) has been found, e.g., as $\Theta = \|\cdot\|_p^p$ $(p = 1, 2)$ [9], [10] and $\Theta = \|\cdot\|_1 \circ D$ [11], where $D$ is a first order difference operator.

For the relaxed continuous constrained optimization problem in (2), it would be desired for the regularizer $\Theta$ to penalize any point not in $\mathfrak{D}$. Along this regularization strategy, the so-called SOAV function [13], [15]

$$\Theta^{\langle 1\rangle}_{\text{SOAV}}(\mathbf{x}) := \sum_{l=1}^{L}\|\mathbf{x}-a_l\mathbf{1}\|_{\boldsymbol{\omega}_l,1} := \sum_{l=1}^{L}\sum_{n=1}^{N}\omega_{l,n}|x_n - a_l|, \tag{3}$$

This work was supported by JSPS Grants-in-Aid (19H04134,23KJ0945) and by JST SICORP (JPMJSC20C6).

has been used with weighting vectors $\boldsymbol{\omega}_l := [\omega_{l,1}, \omega_{l,2}, \ldots, \omega_{l,N}]^\top \in \mathbb{R}_+^N$ $(l = 1, 2, \ldots, L)$ satisfying $\sum_{l=1}^{L} \omega_{l,n} = 1$ $(n = 1, 2, \ldots, N)$, where $\|\mathbf{x}\|_{\boldsymbol{\omega}_l,1} := \sum_{n=1}^{N} \omega_{l,n}|x_n|$ stands for the weighted $\ell_1$-norm associated with $\boldsymbol{\omega}_l$.

Indeed, the weighted SOAV (W-SOAV) model [13] used (2) with $\Theta = \Theta_{\text{SOAV}}^{\langle 1 \rangle}$ and $\widetilde{D} := \mathbb{R}^N$ for MIMO signal detection. Clearly, in this case, the model (2) is a convex model thanks to the convexity of $\Theta_{\text{SOAV}}^{\langle 1 \rangle}$, and therefore, a solution of (2) can be approximated iteratively by a convex optimization solver [13]. However, Fig. 1 (a) suggests that penalization by $\Theta_{\text{SOAV}}^{\langle 1 \rangle}$ is not contrastive enough for use as $\Theta$ in (2) because any point $\mathbf{s}_q \in \mathfrak{D} \subset \mathbb{C}^1$ is never unique minimizer of $\Theta_{\text{SOAV}}^{\langle 1 \rangle}$ over any neighborhood of $\mathbf{s}_q$. More contrastive regularizer than $\Theta_{\text{SOAV}}^{\langle 1 \rangle}$ has been certainly desired for use in (2) of Step 1 of Scheme 1 (see, e.g., Fig. 1 (b)).

We are interested in the following natural questions:

(Q1) Can we design a class of functions that contains fairly contrastive functions for use in (2) ?

(Q2) Can we choose any reasonable function $\Theta$, from such a function class, which is tractable for minimization of $J_\Theta$ ?

(Q1) has been examined in special cases, e.g., $\Theta_{\text{SOAV}}^{\langle p \rangle}(\mathbf{x}) := \sum_{l=1}^{L} \|\mathbf{x} - a_l \mathbf{1}\|_{\boldsymbol{\omega}_l,p}$ $(0 \leq p < 1)$ [16] as nonconvex variants of (3). However, any algorithm, of guaranteed to convergence to a global minimizer of $J_{\Theta_{\text{SOAV}}^{\langle p \rangle}}$, has not yet been established mainly because of the severe nonconvexity of $J_{\Theta_{\text{SOAV}}^{\langle p \rangle}}$. This situation tells us that computational tractability of (2) must be considered carefully from the beginning, i.e., (Q1) and (Q2) should be considered simultaneously.

In this paper, we present a positive answer to these questions

i) by designing a function class as

$$\Theta_{\text{GME}}(\mathbf{x}) := \sum_{l=1}^{L} (\|\cdot\|_{\boldsymbol{\omega}_l,1})_{\mathbf{B}^{\langle l \rangle}}(\mathbf{x} - a_l \mathbf{1}), \quad (4)$$

where $(\|\cdot\|_{\boldsymbol{\omega}_l,1})_{\mathbf{B}^{\langle l \rangle}}$ is a nonconvex enhancement (called *Generalized Moreau Enhancement (GME)*) of $\|\cdot\|_{\boldsymbol{\omega}_l,1}$ with a tunable matrix $\mathbf{B}^{\langle l \rangle}$ (see (5)) (Note: $\Theta_{\text{GME}}$ reproduces $\Theta_{\text{SOAV}}^{\langle 1 \rangle}$ with $\mathbf{B}^{\langle l \rangle} = \mathbf{O}$ (zero matrix));

ii) by exemplifying a fairly contrastive function (see Fig. 1 (b)) in the proposed class of $\Theta_{\text{GME}}$;

iii) by presenting a choice of $\mathbf{B}^{\langle l \rangle}$ $(l = 1, 2, \ldots, L)$ (see (9) and (10)) which achieves the overall convexity of $J_{\Theta_{\text{GME}}}$;

iv) by proposing an iterative algorithm (see Algorithm 1), based on a relaxation of even symmetric condition [17, Problem 1] required for the so-called seed convex function in cLiGME model (see Section II), with guaranteed to convergence to a global minimizer of $J_{\Theta_{\text{GME}}}$ over $\widetilde{\mathfrak{D}}$ under the overall convexity condition.

Indeed, via numerical experiments in a scenario of MIMO signal detection, we demonstrate the effectiveness of the proposed regularizer $\Theta_{\text{GME}}$ in the model (2).

We also propose a pair of simple technical improvements for the proposed iterative algorithm in Step 1 of Scheme 1 by exploiting adaptively the discrete information regarding $\mathfrak{D}$. More precisely, we propose (i) to use a strategic perturbation (13) to move the estimate closer to $\mathfrak{D}$ at each iteration $k \in \mathbb{N}$ (this idea is inspired by *superiorization* [18], [19]), and (ii) to update the weights $\omega_{l,n}$ $(l = 1, 2, \ldots, L; n = 1, 2, \ldots, N)$ in (4) adaptively (see (14)) by assigning larger weight to $\omega_{l,n}$ for $(l, n)$ such that the distance between $a_l \in \mathfrak{A}$ and $n$th coordinate $x_n \in \mathbb{R}$ of the latest estimate $\mathbf{x} \in \mathbb{R}^N$ is smaller (this idea is inspired by iterative reweighting of SOAV [13]). Experiment results demonstrate that these simple techniques improve numerical performance of the proposed Algorithm 1.

**Notation**. $\mathbb{N}, \mathbb{R}, \mathbb{R}_+, \mathbb{R}_{++}$ and $\mathbb{C}$ denote respectively the set of all nonnegative integers, all real numbers, all nonnegative real numbers, all positive real numbers and all complex numbers ($j$ stands for the imaginary unit, and $\Re(\cdot)$ and $\Im(\cdot)$ stand respectively for real and imaginary parts).

Let $\mathcal{H}, \mathcal{K}$ be finite dimensional real Hilbert spaces. The set of all *proper lower semicontinuous convex* functions[1] defined on $\mathcal{H}$ is denoted by $\Gamma_0(\mathcal{H})$. $f \in \Gamma_0(\mathcal{H})$ is said to be *prox-friendly* if $\text{Prox}_{\gamma f} : \mathcal{H} \to \mathcal{H} : x \mapsto \text{argmin}_{y \in \mathcal{H}}[f(y) + \frac{1}{2\gamma}\|y - x\|_{\mathcal{H}}^2]$ is available as a computable operator for any $\gamma \in \mathbb{R}_{++}$. A closed convex set $C \subset \mathcal{H}$ is said to be *simple* if the metric projection $P_C : \mathcal{H} \to \mathcal{H} : x \mapsto \text{argmin}_{y \in \mathcal{H}}\|x - y\|_{\mathcal{H}}$ is available as a computable operator. The set of bounded linear operators from $\mathcal{H}$ to $\mathcal{K}$ is denoted by $\mathcal{B}(\mathcal{H}, \mathcal{K})$. For $\mathfrak{L} \in \mathcal{B}(\mathcal{H}, \mathcal{K})$, $\mathfrak{L}^* \in \mathcal{B}(\mathcal{K}, \mathcal{H})$ denotes the adjoint operator of $\mathfrak{L}$ (i.e., $(\forall (x, y) \in \mathcal{H} \times \mathcal{K})$ $\langle \mathfrak{L}x, y \rangle_{\mathcal{K}} = \langle x, \mathfrak{L}^* y \rangle_{\mathcal{H}}$).

For discussion in Euclidean space, we use boldface letters to express vectors and general font letters to represent scalars. For a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, $\mathbf{X}^\top \in \mathbb{R}^{n \times m}$ denotes the transpose of $\mathbf{X}$. The symbols $\mathbf{I}$, $\mathbf{O}$ and $\mathbf{1}$ respectively stand for the identity matrix, the zero matrix and the all one vector.

## II. Brief Introduction to cLiGME

**Problem II.1** (cLiGME model [20], [21]). Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}_l$ $(l = 1, 2, \ldots, L)$, $\widetilde{\mathcal{Z}}_l$ $(l = 1, 2, \ldots, L)$ and $\mathfrak{Z}$ be finite dimensional real Hilbert spaces. Suppose that (a) $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, $y \in \mathcal{Y}$ and $\mu > 0$; (b) for each $l = 1, 2, \ldots, L$, $B^{\langle l \rangle} \in \mathcal{B}(\mathcal{Z}_l, \widetilde{\mathcal{Z}}_l)$, $\mathfrak{L}^{\langle l \rangle} \in \mathcal{B}(\mathcal{X}, \mathcal{Z}_l)$ and $\mu_l > 0$; (c) $C (\subset \mathfrak{Z})$ is a nonempty simple closed convex set and $\mathfrak{C} \in \mathcal{B}(\mathcal{X}, \mathfrak{Z})$; (d) $\Psi^{\langle l \rangle} \in \Gamma_0(\mathcal{Z}_l)$ is (i) coercive, (ii) $\text{dom}\,\Psi^{\langle l \rangle} = \mathcal{Z}_l$, (iii) even symmetry (i.e., $\Psi^{\langle l \rangle} \circ (-\text{Id}) = \Psi^{\langle l \rangle}$), (iv) prox-friendly. Then

i) With a tunable matrix $B^{\langle l \rangle} \in \mathcal{B}(\mathcal{Z}_l \widetilde{\mathcal{Z}}_l)$, the *Generalized Moreau Enhancement (GME)* of $\Psi^{\langle l \rangle}$ is defined by

$$\Psi_{B^{\langle l \rangle}}^{\langle l \rangle}(\cdot) := \Psi^{\langle l \rangle}(\cdot) - \min_{v \in \mathcal{Z}_l} \left[ \Psi^{\langle l \rangle}(v) + \frac{1}{2} \|B^{\langle l \rangle}(\cdot - v)\|_{\widetilde{\mathcal{Z}}_l}^2 \right]. \quad (5)$$

ii) The *constrained LiGME (cLiGME)* model is given as

$$\text{Find } x^\diamond \in \underset{\mathfrak{C}x \in C}{\text{argmin}} \frac{1}{2}\|y - Ax\|_{\mathcal{Y}}^2 + \mu \sum_{l=1}^{L} \mu_l \Psi_{B^{\langle l \rangle}}^{\langle l \rangle} \circ \mathfrak{L}^{\langle l \rangle}(x). \quad (6)$$

[1] A function $f : \mathcal{H} \to (-\infty, \infty]$ is (i) proper if $\text{dom}(f) := \{x \in \mathcal{H} \mid f(x) < \infty\} \neq \emptyset$, (ii) lower semicontinuous if $\{x \in \mathcal{H} \mid f(x) \leq \alpha\}$ is closed for $\forall \alpha \in \mathbb{R}$, (iii) convex if $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta) f(y)$ for $\forall x, y \in \mathcal{H}$, $0 < \theta < 1$.

The regularizer $\Psi^{\langle l\rangle}_{B^{\langle l\rangle}}$ was proposed originally in [17] as an extension of the so-called *GMC penalty* in [22] mainly for the sparsity aware estimation. Furthermore, although $\Psi^{\langle l\rangle}_{B^{\langle l\rangle}}$ with $B^{\langle l\rangle} \neq O$ is nonconvex, the convexity of the cost function in (6) is achieved by a strategic tuning of GME matrices $B^{\langle l\rangle}$ $(l = 1, 2, \dots, L)$ (see, e.g., [23]).

## III. Proposed regularizer and algorithm for discrete-valued signal estimation

### A. Proposed GME regularizer and cLiGME algorithm

In this section, we propose a class of regularizers $\Theta_{\text{GME}}$ in (4), which is designed with the GME functions (5) of $\|\cdot\|_{\boldsymbol{\omega}_l,1}$ $(l = 1, 2, \dots, L)$. Indeed, as we show in Fig. 1 (b), the class of the proposed regularizers $\Theta_{\text{GME}}$ contains fairly contrastive functions. Moreover, if $\mathbf{B}^{\langle l\rangle} = \mathbf{O}$ $(l = 1, 2, \dots, L)$, then $\Theta_{\text{GME}} = \Theta^{\langle 1\rangle}_{\text{SOAV}}$ holds. By using $\Theta_{\text{GME}}$ (see (4)) in Step 1 of Scheme 1, we propose the following model:

$$\text{Find } \mathbf{x}^{\diamond} \in \underbrace{\operatorname*{argmin}_{\mathbf{x}\in\widetilde{\mathfrak{D}}\subset\mathbb{R}^N} \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \mu\underbrace{\sum_{l=1}^{L}(\|\cdot\|_{\boldsymbol{\omega}_l,1})_{\mathbf{B}^{\langle l\rangle}}(\mathbf{x}-a_l\mathbf{1})}_{=\Theta_{\text{GME}}(\mathbf{x})}}_{=J_{\Theta_{\text{GME}}}(\mathbf{x})}, \tag{7}$$

where $\widetilde{\mathfrak{D}} \supset \mathfrak{D}$ is a closed convex set.

In order to solve (7), we consider the following problem which contains (7) as its special instance (see Remark III.2).

**Problem III.1.** Let $\mathcal{X}$, $\mathcal{Y}$ and $\widetilde{\mathcal{Z}}_l$ $(l = 1, 2, \dots, L)$ be finite dimensional real Hilbert spaces. Suppose that (a) $A \in \mathcal{B}(\mathcal{X}, \mathcal{Y})$, $y \in \mathcal{Y}$ and $\mu > 0$; (b) for each $l = 1, 2, \dots, L$, $B^{\langle l\rangle} \in \mathcal{B}(\mathcal{X}, \widetilde{\mathcal{Z}}_l)$, $z^{\langle l\rangle} \in \mathcal{X}$ and $\mu_l > 0$; (c) $C(\subset \mathcal{X})$ is a nonempty simple closed convex set; (d) $\Psi^{\langle l\rangle} \in \Gamma_0(\mathcal{X})$ is (i) coercive, (ii) $\operatorname{dom}\Psi^{\langle l\rangle} = \mathcal{X}$, (iii) even symmetry, (iv) prox-friendly. Then, consider

$$\text{Find } x^{\diamond} \in \underbrace{\operatorname*{argmin}_{x\in C} \frac{1}{2}\|y-Ax\|_{\mathcal{Y}}^2 + \mu\sum_{l=1}^{L}\mu_l\Psi^{\langle l\rangle}_{B^{\langle l\rangle}}(x-z^{\langle l\rangle})}_{=:J(x)}, \tag{8}$$

where $\Psi^{\langle l\rangle}_{B^{\langle l\rangle}} : \mathcal{X} \to \mathbb{R}$ is a GME function (5), with a tuning matrix $B^{\langle l\rangle}$, of a convex function $\Psi^{\langle l\rangle}$ $(l = 1, 2, \dots, L)$.

**Remark III.2.** The model (7) is a special instance of Problem III.1 with $\mathcal{X} = \mathbb{R}^N$, $\mathcal{Y} = \mathbb{R}^M$, $C = \widetilde{\mathfrak{D}}$, $\Psi^{\langle l\rangle} = \|\cdot\|_{\boldsymbol{\omega}_l,1}$ $(l = 1, 2, \dots, L)$, $z^{\langle l\rangle} = a_l\mathbf{1}$ $(l = 1, 2, \dots, L)$ and $\mu_l = 1$ $(l = 1, 2, \dots, L)$.

By tuning properly the design parameters of the proposed GME regularizers, we can make the nonconvexly-regularized least squares model convex.

**Fact III.3** (Overall convexity condition [17] for (8))**.** *Consider Problem III.1. Then $J$ in* (8) *is convex if $(B^{\langle l\rangle})_{l=1}^L$ satisfy*

$$A^*A - \mu\sum_{l=1}^{L}\mu_l {B^{\langle l\rangle}}^* B^{\langle l\rangle} \text{ is positive semidefinite.} \tag{9}$$

*For example, the following $B^{\langle l\rangle}$ [22] satisfy* (9)*:*

$$B^{\langle l\rangle} := \sqrt{\frac{\gamma_l}{\mu\mu_l}}A \in \mathcal{B}(\mathcal{X}, \mathcal{Y}) \quad (l = 1, 2, \dots, L), \tag{10}$$

*where $\gamma_l \in \mathbb{R}_+$ $(l = 1, 2, \dots, L)$ are chosen to satisfy $\sum_{l=1}^{L}\gamma_l \leq 1$.*

In a special case where $z^{\langle l\rangle} = 0 \in \mathcal{X}$ $(l = 1, 2, \dots, L)$, the model (8) is reduced to the cLiGME model (6) with $\mathcal{Z} = \mathfrak{Z} = \mathcal{X}$, $\mathfrak{L}^{\langle l\rangle} = \text{Id}$ $(l = 1, 2, \dots, L)$ and $\mathfrak{C} = \text{Id}$. For this special case under the condition (9), we can find a global minimizer of (8), by cLiGME algorithm [20], [21].

In the following, we present an iterative algorithm applicable even to general cases $z^{\langle l\rangle} \in \mathcal{X}$ $(l = 1, 2, \dots, L)$. The proposed algorithm is a variant of cLiGME algorithm [20], [21], [24] (see Remark III.5), and can be reduced to Algorithm 1 for the proposed model (7) (see also Remark III.2).

**Theorem III.4** (A relaxation of cLiGME algorithm for Problem III.1)**.** *Consider Problem III.1 under the overall convexity condition* (9)*. Assume $\operatorname{argmin}_{x\in C} J(x) \neq \emptyset$*[2]*. Define the operator $T : \mathcal{H} := \mathcal{X} \times (\mathcal{X})^L \times (\mathcal{X})^L \to \mathcal{H} : \big(x, (v^{\langle l\rangle})_{l=1}^L, (w^{\langle l\rangle})_{l=1}^L\big) \mapsto \big(\xi, (\zeta^{\langle l\rangle})_{l=1}^L, (\eta^{\langle l\rangle})_{l=1}^L\big)$ by*

$$\begin{cases} \xi := P_C\big[\big(\text{Id} - \frac{1}{\sigma}(A^*A - \mu\sum_{l=1}^{L}\mu_l {B^{\langle l\rangle}}^* B^{\langle l\rangle})\big)x \\ \qquad -\frac{\mu}{\sigma}\sum_{l=1}^{L}\big(\mu_l {B^{\langle l\rangle}}^* B^{\langle l\rangle} v^{\langle l\rangle} + w^{\langle l\rangle}\big) + \frac{1}{\sigma}A^*y\big], \\ \zeta^{\langle l\rangle} := z^{\langle l\rangle} + \operatorname{Prox}_{\frac{\mu\mu_l}{\tau}\Psi^{\langle l\rangle}}\big[\frac{\mu\mu_l}{\tau}{B^{\langle l\rangle}}^* B^{\langle l\rangle}(2\xi - x) \\ \qquad +\big(\text{Id} - \frac{\mu\mu_l}{\tau}{B^{\langle l\rangle}}^* B^{\langle l\rangle}\big)v^{\langle l\rangle} - z^{\langle l\rangle}\big], \\ \eta^{\langle l\rangle} := \big(\text{Id} - \operatorname{Prox}_{\mu_l\Psi^{\langle l\rangle}}\big)\big(2\xi - x + w^{\langle l\rangle} - z^{\langle l\rangle}\big), \end{cases}$$

*where $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ is chosen to satisfy*[3]

$$\begin{cases} (\sigma - \mu L)\text{Id} - \frac{\kappa}{2}A^*A \text{ is positive definite}, \\ \tau \geq (\frac{\kappa}{2} + \frac{2}{\kappa})\mu\max\{\mu_l\|B^{\langle l\rangle}\|_{\text{OP}}^2 \mid l = 1, 2, \dots, L\}. \end{cases} \tag{11}$$

*Then, (a) $T$ is an averaged nonexpansive operator*[4] *by defining a proper inner product on $\mathcal{H}$ (see, e.g., [17], [21]), and (b) for any initial point $\mathfrak{u}_0 \in \mathcal{H}$, the sequence $(\mathfrak{u}_k)_{k\in\mathbb{N}} \subset \mathcal{H}$ with $\mathfrak{u}_k := \Big(x_k, (v_k^{\langle l\rangle})_{l=1}^L, (w_k^{\langle l\rangle})_{l=1}^L\Big)$ generated by the following* Picard-type fixed point iteration*:*

$$(k \in \mathbb{N}) \quad \mathfrak{u}_{k+1} = T(\mathfrak{u}_k)$$

*converges to a fixed point $\bar{\mathfrak{u}} = \big(\overline{x}, (\overline{v}^{\langle l\rangle})_{l=1}^L, (\overline{w}^{\langle l\rangle})_{l=1}^L\big) \in$* $\operatorname{Fix}(T) := \{\mathfrak{u} \in \mathcal{H} \mid T(\mathfrak{u}) = \mathfrak{u}\} \subset \mathcal{H}$*, where $\overline{x} \in \mathcal{X}$ enjoys the condition as a global minimizer $x^{\diamond} \in C$, in* (8)*, of $J$.*

---

[2] $\operatorname{argmin}_{x\in C} J(x) \neq \emptyset$ is guaranteed in many cases, e.g., if $C$ is compact (not limited to this case).

[3] For example, choose $\kappa > 1$ and compute $(\sigma, \tau)$ by

$$\begin{cases} \sigma := \frac{\kappa}{2}\|A\|_{\text{OP}}^2 + \mu L + (\kappa - 1), \\ \tau := (\frac{\kappa}{2} + \frac{2}{\kappa})\mu\max\{\mu_l\|B^{\langle l\rangle}\|_{\text{OP}}^2 \mid l = 1, 2, \dots, L\} + (\kappa - 1), \end{cases}$$

where $\|B^{\langle l\rangle}\|_{\text{OP}}$ denotes the operator norm of $B^{\langle l\rangle}$ (i.e., $\|B^{\langle l\rangle}\|_{\text{OP}} := \sup_{x\in\mathcal{X}, \|x\|_{\mathcal{X}}\leq 1}\|B^{\langle l\rangle}x\|_{\widetilde{\mathcal{Z}}_l}$).

[4] An operator $T : \mathcal{X} \to \mathcal{X}$ is said to be *nonexpansive* if $(\forall x, y \in \mathcal{X})$ $\|T(x) - T(y)\| \leq \|x - y\|$, in particular, ($\alpha$-) *averaged nonexpansive* if there exists $\alpha \in (0, 1)$ and a nonexpansive operator $R : \mathcal{X} \to \mathcal{X}$ such that $T = (1 - \alpha)\text{Id} + \alpha R$.

**Algorithm 1** A relaxation of cLiGME algorithm for (7)

1: Choose $\left(\mathbf{x}_0,(\mathbf{v}_0^{\langle l\rangle})_{l=1}^L,(\mathbf{w}_0^{\langle l\rangle})_{l=1}^L\right)\in\mathbb{R}^N\times(\mathbb{R}^N)^L\times(\mathbb{R}^N)^L$.
2: Choose $(\sigma,\tau,\kappa)\in\mathbb{R}_{++}\times\mathbb{R}_{++}\times(1,\infty)$ satisfying (11).
3: **for** $k=0,1,2,\ldots$ **do**
4: { Insert modification in Section III-B if necessary. }
5: $\mathbf{x}_{k+1}\leftarrow P_{\widetilde{\mathfrak{D}}}\big[\big(\mathbf{I}-\frac{1}{\sigma}(\mathbf{A}^\top\mathbf{A}-\mu\sum_{l=1}^L{\mathbf{B}^{\langle l\rangle}}^\top\mathbf{B}^{\langle l\rangle})\big)\mathbf{x}_k$ $-\frac{\mu}{\sigma}\sum_{l=1}^L\big({\mathbf{B}^{\langle l\rangle}}^\top\mathbf{B}^{\langle l\rangle}\mathbf{v}_k^{\langle l\rangle}+\mathbf{w}_k^{\langle l\rangle}\big)+\frac{1}{\sigma}\mathbf{A}^\top\mathbf{y}\big]$
6: **for** $l=1,2,\ldots,L$ **do**
7: $\mathbf{v}_{k+1}^{\langle l\rangle}\leftarrow a_l\mathbf{1}+\mathrm{Prox}_{\frac{\mu}{\tau}\|\cdot\|_{\boldsymbol{\omega}_l,1}}\big[\frac{\mu}{\tau}{\mathbf{B}^{\langle l\rangle}}^\top\mathbf{B}^{\langle l\rangle}(2\mathbf{x}_{k+1}-\mathbf{x}_k)$ $+\big(\mathbf{I}-\frac{\mu}{\tau}{\mathbf{B}^{\langle l\rangle}}^\top\mathbf{B}^{\langle l\rangle}\big)\mathbf{v}_k^{\langle l\rangle}-a_l\mathbf{1}\big]$
8: $\mathbf{w}_{k+1}^{\langle l\rangle}\leftarrow\big(\mathrm{Id}-\mathrm{Prox}_{\|\cdot\|_{\boldsymbol{\omega}_l,1}}\big)\big(2\mathbf{x}_{k+1}-\mathbf{x}_k+\mathbf{w}_k^{\langle l\rangle}-a_l\mathbf{1}\big)$
9: **end for**
10: **end for**
11: **return** $\mathbf{x}_{k+1}$

**Remark III.5.** For Problem III.1, the condition (9) and Theorem III.4 are obtained by the following steps[5].

i) Define $\widetilde{\Psi}^{\langle l\rangle}:=\Psi^{\langle l\rangle}(\cdot-z^{\langle l\rangle})$ $(l=1,2,\ldots,L)$.
ii) We can verify that $\widetilde{\Psi}_{B^{\langle l\rangle}}^{\langle l\rangle}=\Psi_{B^{\langle l\rangle}}^{\langle l\rangle}(\cdot-z^{\langle l\rangle})$.
iii) By using $\widetilde{\Psi}^{\langle l\rangle}$, $J$ in (8) can be expressed equivalently as

$$J(x)=\frac{1}{2}\|y-Ax\|_{\mathcal{Y}}^2+\mu\sum_{l=1}^L\mu_l\widetilde{\Psi}_{B^{\langle l\rangle}}^{\langle l\rangle}(x). \tag{12}$$

iv) The overall convexity condition (9) for $J$ in (8) is derived by applying [17, Proposition 1] to (12).
v) Since $\widetilde{\Psi}^{\langle l\rangle}$ is not even symmetry in general, we cannot apply the cLiGME algorithm [20], [21] directly for (12). However, we can relax the even symmetric condition to $z^{\langle l\rangle}$-symmetric condition (i.e., $\widetilde{\Psi}^{\langle l\rangle}(z^{\langle l\rangle}+\cdot)=\widetilde{\Psi}^{\langle l\rangle}(z^{\langle l\rangle}-\cdot)$).

### *B. Two simple techniques for further improvement*

For further improvement of Algorithm 1, we introduce two simple techniques in Algorithm 1 to exploit adaptively the discrete information regarding $\mathfrak{D}$.

*1) Generalized superiorization of cLiGME algorithm:* *Superiorization* [18], [19] is known as a technique for an iterative algorithm, e.g., Picard-type fixed point iteration, to reduce a certain objective cost by adding strategic bounded perturbations to updated estimate $\mathbf{x}_k$ $(k\in\mathbb{N})$.

We propose to incorporate a superiorization technique into Algorithm 1 in order to move the estimate closer to $\mathfrak{D}$ at each iteration. More precisely, we use a modification

$$\mathbf{x}_k\leftarrow\mathbf{x}_k+\beta_k\underbrace{(P_{\mathfrak{D}}-\mathrm{Id})(\mathbf{x}_k)}_{=:\mathbf{d}_k} \tag{13}$$

[5] Even for Problem II.1 with a replacement of the seed convex function $\Psi^{\langle l\rangle}$ by its shifted seed convex function $\widetilde{\Psi}^{\langle l\rangle}=\Psi^{\langle l\rangle}(\cdot-z^{\langle l\rangle})$, via an essentialy same discussion, we can derive the overall convexity condition of the objective function and an iterative algorithm with guaranteed convergence to a global minimizer.

in line 4 of Algorithm 1, where $(\beta_k)_{k\in\mathbb{N}}\subset\mathbb{R}_+$, and $(\mathbf{d}_k)_{k\in\mathbb{N}}\subset\mathbb{R}^N$ is inspired by [19]. The global convergence guarantee is not violated even by the modification (13) if $(\beta_k)_{k\in\mathbb{N}}$ is summable and $(\mathbf{d}_k)_{k\in\mathbb{N}}$ is bounded (see Appendix B2). However, we dare to propose to use more general $(\beta_k)_{k\in\mathbb{N}}\subset\mathbb{R}_+$ which is not necessarily summable. We call such a modification *generalized superiorization*. As will be shown in numerical experiments (see Section IV), the proposed generalized superiorization is effective to guide the sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ to the discrete set $\mathfrak{D}$.

*2) Iterative reweighting of cLiGME algorithm:* The iterative reweighting technique, e.g., [25], has been used to enhance the effectiveness of the regularizer by updating the weights of the regularizer adaptively in an iterative algorithm. Iterative reweighting techniques are also used for Problem I.1 [13], [26]. To utilize such a technique in Algorithm 1, we propose to set $\omega_{l,n}$ $(l=1,2,\ldots,L;\ n=1,2,\ldots,N)$ in the seed functions $\|\cdot\|_{\boldsymbol{\omega}_l,1}$ $(l=1,2,\ldots,L)$ adaptively by using the latest estimate $\mathbf{x}:=[x_1,x_2,\ldots,x_N]^\top$ as [26]

$$\omega_{l,n}=\frac{(|x_n-a_l|+\epsilon)^{-1}}{\sum_{l'=1}^L(|x_n-a_{l'}|+\epsilon)^{-1}}. \tag{14}$$

where $\epsilon>0$ is a small number. If $|x_n-a_l|$ is small, then the corresponding $\omega_{l,n}$ becomes large and $x_n$ will be close to $a_l$. This iterative reweighting method can be realized by inserting

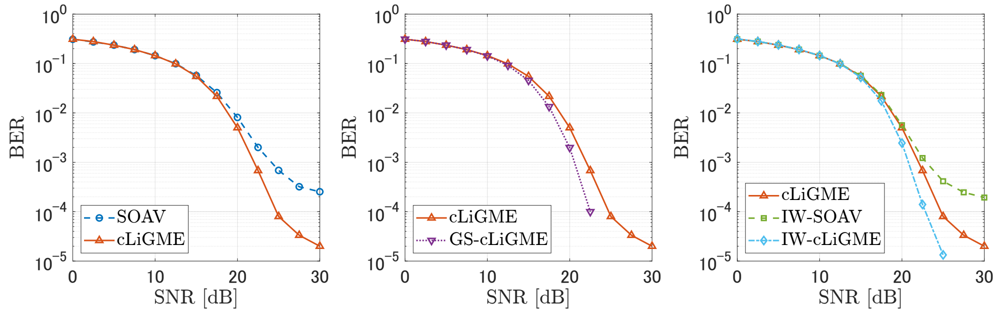
**if** $k \bmod K == 0$ **then**
Update $\boldsymbol{\omega}_l=[\omega_{l,1},\omega_{l,2},\ldots,\omega_{l,N}]^\top$ $(l=1,2,\ldots,L)$ as (14) with $\mathbf{x}=\mathbf{x}_k$. (15)
**end if**

in line 4 of Algorithm 1, where $K\in\mathbb{N}\backslash\{0\}$ controls the frequency of reweighting.

## IV. Numerical experiments

We conducted numerical experiments in a scenario of MIMO signal detection [12] with $N$-transmit antennas and $M$-receive antennas $(N=50,M=45)$ in 8PSK (phase shift keying) modulation with constellation set $\mathfrak{A}:=\{a_l:=\exp[j(l-1)\pi/4]\mid l=1,2,\ldots,8=:L\}\subset\mathbb{C}$. The task of this experiment is to estimate the transmitted signal $\mathbf{x}^\star\in\mathbb{C}^N$ from the received signal $\mathbf{y}=\mathbf{A}\mathbf{x}^\star+\boldsymbol{\varepsilon}\in\mathbb{C}^M$ with the channel matrix $\mathbf{A}\in\mathbb{C}^{M\times N}$ and a noise $\boldsymbol{\varepsilon}\in\mathbb{C}^M$. In this experiment, we chose randomly (i) $\mathbf{x}^\star\in\mathfrak{D}:=\mathfrak{A}^N$, (ii) $\mathbf{A}:=\sqrt{\mathbf{R}}\mathbf{G}\in\mathbb{C}^{M\times N}$, where each entry of $\mathbf{G}\in\mathbb{C}^{M\times N}$ was sampled from the complex gaussian distribution $\mathcal{CN}(0,1/M)$, and $\mathbf{R}\in\mathbb{R}^{M\times M}$ satisfies $(\mathbf{R})_{i,j}=0.5^{|i-j|}$ $(i=1,2,\ldots,M;\ j=1,2,\ldots,M)$, and (iii) each entry of $\boldsymbol{\varepsilon}\in\mathbb{C}^M$ was sampled from $\mathcal{CN}(0,\sigma_\varepsilon^2)$ with a variance $\sigma_\varepsilon^2>0$, which was chosen so that $10\log_{10}\frac{\mathrm{E}[\|\mathbf{x}^\star\|^2/N]}{\sigma_\varepsilon^2}$ achieved a given SNR (signal-to-noise ratio).

We consider to estimate $\mathbf{x}^\star\in\mathfrak{D}$ with Scheme 1 by employing the convex hull $\mathrm{conv}(\mathfrak{D})$ of $\mathfrak{D}$ as $\widetilde{\mathfrak{D}}$ in (2) via $\mathbb{C}\rightleftarrows\mathbb{R}^2$ translation (see Appendix A). In this experiment, we compared numerical performance of (i) the proposed cLiGME model (7), i.e., the model (2) with $\Theta=\Theta_{\mathrm{GME}}$ in (4), with that




(a) Basic cLiGME (b) Generalized superiorization (c) Iterative reweighting

Fig. 2: BER vs SNR (8PSK, $N = 50$, $M = 45$)

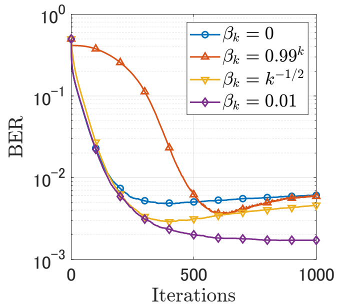



Fig. 3: BER vs iterations

of (ii) the SOAV model, i.e., the model (2) with $\Theta = \Theta_{\text{SOAV}}^{\langle 1 \rangle}$ in (3). For the cLiGME model, we used Algorithm 1 (denoted by 'cLiGME') by employing the tuning matrices in (7)

$$\mathbf{B}^{\langle l \rangle} = \sqrt{0.99/\mu L}\widehat{\mathbf{A}} \quad (l = 1, 2, \ldots, L)$$

to achieve the overall convexity condition (9), where $\mu$ is a predetermined regularization parameter, and $\widehat{\mathbf{A}}$ is obtained via $\mathbb{C} \rightleftarrows \mathbb{R}^2$ translation (see (16)). Since SOAV model can be reduced to the cLiGME model (7) with $\mathbf{B}^{\langle l \rangle} = \mathbf{O}$ $(l = 1, 2, \ldots, L)$, we used Algorithm 1 (denoted by 'SOAV') with $\mathbf{B}^{\langle l \rangle} = \mathbf{O}$ $(l = 1, 2, \ldots, L)$ for the SOAV model. For both 'cLiGME' and 'SOAV', we employed the same (i) stepsize $(\sigma, \tau, \kappa) \in \mathbb{R}_{++} \times \mathbb{R}_{++} \times (1, \infty)$ as footnote 3 in Theorem III.4 with $\kappa = 1.001$, and (ii) initial point $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{v}_0^{\langle l \rangle} = \mathbf{0}$ and $\mathbf{w}_0^{\langle l \rangle} = \mathbf{0}$ $(l = 1, 2, \ldots, L)$. Since $\widetilde{\mathfrak{D}}$ is compact, 'cLiGME' and 'SOAV' can find their global minimizers, respectively (see Theorem III.4). Algorithm 1 were terminated when the iteration number $k$ exceeded 500.

As a performance metric, we adopted averaged BER (bit error rate) over 1,000 independent realizations of $(\mathbf{x}^{\star}, \mathbf{A}, \boldsymbol{\varepsilon})$. The parameter $\mu$ was chosen to achieve the lowest BER from the set $\{10^i \mid i = -10, -9, \ldots, 2\}$ at each SNR.

Fig. 2 (a) shows BER of 'SOAV' and 'cLiGME' at each SNR, where $\omega_{l,n} = 1/8$ $(l = 1, 2, \ldots, 8;\ n = 1, 2, \ldots, N)$ in (7) were fixed. From Fig. 2 (a), 'cLiGME' achieves lower BER than 'SOAV', which implies the effectiveness of the proposed contrastive nonconvex regularizer $\Theta_{\text{GME}}$ compared with the convex regularizer $\Theta_{\text{SOAV}}$.

In the following, we verify the further performance improvements of 'cLiGME' by the proposed (i) generalized superiorization and (ii) iterative reweighting.

To examine the impact of choices of $(\beta_k)_{k\in\mathbb{N}}$ in generalized superiorization (13), we compared generalized superiorization of 'cLiGME' with (i) $\beta_k = 0$ (which reduces to the original 'cLiGME'), (ii) $\beta_k = 0.99^k$ ($(\beta_k)_{k\in\mathbb{N}}$ is summable), (iii) $\beta_k = k^{-1/2}$ ($(\beta_k)_{k\in\mathbb{N}}$ is nonsummable but $\beta_k \to 0$ $(k \to \infty)$), and (iv) $\beta_k = 0.01$. Fig. 3 shows history of BER achieved by generalized superiorization of 'cLiGME' with such $(\beta_k)_{k\in\mathbb{N}}$ in (13), where SNR = 20 dB, $\mu = 10^{-4}$ and $\omega_{l,n} = 1/8$ $(l = 1, 2, \ldots, 8;\ n = 1, 2, \ldots, N)$. From Fig. 3, $\beta_k = 0.01$ outperforms the others. Fig. 2 (b) shows BER, at each SNR, of 'cLiGME' and generalized superiorization of 'cLiGME' (denoted by 'GS-cLiGME') with $\beta_k = 0.01$. From Fig. 2 (b), we see 'GS-cLiGME' improves 'cLiGME'.

Fig. 2 (c) shows BER, at each SNR, of (i) 'cLiGME', (ii) iterative reweighting in (15) of 'cLiGME' (denoted by 'IW-cLiGME'), and (iii) iterative reweighting in (15) of 'SOAV' (denoted by 'IW-SOAV'), where the frequency period $K = 100$ in (15) is used (Note: the iterative reweighting of SOAV model was initially proposed [13], [26] with an ADMM-type algorithm). From Fig. 2 (c), 'IW-cLiGME' improves 'cLiGME', while even 'cLiGME' outperforms 'IW-SOAV'.

## V. Conclusion

We proposed a class of fairly contrastive regularizers for discrete-valued estimation problems, and presented an iterative algorithm with guaranteed convergence to a global minimizer of the nonconvexly-regularized least squares model. We also proposed two simple techniques for performance improvements. The numerical experiments demonstrate that the proposed model and algorithm have a great potential for challenging discrete-valued signal estimation problem, and that two simple techniques successfully contribute to performance improvements of the proposed algorithm.

## Appendix

### A. $\mathbb{C} \rightleftarrows \mathbb{R}^2$ *translation*

Consider the complex version of Problem I.1 where $\mathfrak{D}(\subset \mathbb{C}^N)$ is a finite set and $(\mathbf{x}^{\star}, \mathbf{y}, \mathbf{A}, \boldsymbol{\varepsilon}) \in \mathbb{C}^N \times \mathbb{C}^M \times \mathbb{C}^{M\times N} \times \mathbb{C}^M$. The $\mathbb{C} \rightleftarrows \mathbb{R}^2$ translation in this paper should be understood in the following sense:

$$\begin{aligned}
\widehat{\mathfrak{D}} &:= \left\{ \begin{bmatrix} \Re(\mathbf{s}) \\ \Im(\mathbf{s}) \end{bmatrix} \in \mathbb{R}^{2N} \;\middle|\; \mathbf{s} \in \mathfrak{D} \right\}, \\
\widehat{\mathbf{x}}^{\star} &:= \begin{bmatrix} \Re(\mathbf{x}^{\star}) \\ \Im(\mathbf{x}^{\star}) \end{bmatrix} \in \widehat{\mathfrak{D}} \subset \mathbb{R}^{2N}, \widehat{\mathbf{y}} := \begin{bmatrix} \Re(\mathbf{y}) \\ \Im(\mathbf{y}) \end{bmatrix} \in \mathbb{R}^{2M}, \\
\widehat{\mathbf{A}} &:= \begin{bmatrix} \Re(\mathbf{A}) & -\Im(\mathbf{A}) \\ \Im(\mathbf{A}) & \Re(\mathbf{A}) \end{bmatrix} \in \mathbb{R}^{2M\times 2N}, \widehat{\boldsymbol{\varepsilon}} := \begin{bmatrix} \Re(\boldsymbol{\varepsilon}) \\ \Im(\boldsymbol{\varepsilon}) \end{bmatrix} \in \mathbb{R}^{2M}.
\end{aligned} \tag{16}$$

Clearly, via (16), we can translate $\mathbf{y} = \mathbf{A}\mathbf{x}^{\star} + \boldsymbol{\varepsilon}$ into $\widehat{\mathbf{y}} = \widehat{\mathbf{A}}\widehat{\mathbf{x}}^{\star} + \widehat{\boldsymbol{\varepsilon}}$, and can estimate $\widehat{\mathbf{x}}^{\star}$ by applying Algorithm 1 to the translated real model.

### B. Bounded perturbation for Picard-type fixed point iteration

*1) Picard iteration:* Let $\mathcal{H}$ be a finite dimensional real Hilbert space. Suppose $T : \mathcal{H} \to \mathcal{H}$ is an averaged nonexpansive operator such that $\mathrm{Fix}(T) \coloneqq \{u \in \mathcal{H} \mid T(u) = u\} \neq \emptyset$. Then, a sequence $(u_k)_{k\in\mathbb{N}}$, generated by the so-called Picard iteration: $u_{k+1} = T(u_k)$ ($k \in \mathbb{N}$) with any initial point $u_0 \in \mathcal{H}$, is guaranteed to converge to a certain fixed point in $\mathrm{Fix}(T)$.

*2) Bounded perturbation resilience of Picard iteration [18], [19]:* Let $(\beta_k)_{k\in\mathbb{N}}$ be a summable sequence in $\mathbb{R}_+$ and $(d_k)_{k\in\mathbb{N}}$ be a bounded sequence in $\mathcal{H}$, where such a $(\beta_k d_k)_{k\in\mathbb{N}}$ is said to be a sequence of bounded perturbations. Then, with any initial point $u_0 \subset \mathcal{H}$, $(u_k)_{k\in\mathbb{N}}$ generated by

$$(\forall k \in \mathbb{N}) \; u_{k+1} = T(u_k + \beta_k d_k)$$

also converges to a point $\overline{u} \in \mathrm{Fix}(T)$.

## References


[1] V. Bioglio, G. Coluccia, and E. Magli, "Sparse image recovery using compressed sensing over finite alphabets," in *ICIP*, 2014.

[2] A.-J. Van der Veen, S. Talwar, and A. Paulraj, "Blind estimation of multiple digital signals transmitted over fir channels," *IEEE Signal Processing Letters*, 1995.

[3] B. Knoop, F. Monsees, C. Bockelmann, D. Peters-Drolshagen, S. Paul, and A. Dekorsy, "Compressed sensing k-best detection for sparse multi-user communications," in *EUSIPCO*, 2014.

[4] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, 1999.

[5] S. M. Fosson and M. Abuabiah, "Recovery of binary sparse signals from compressed linear measurements via polynomial optimization," *IEEE Signal Processing Letters*, 2019.

[6] B. Trotobas, A. Nafkha, and Y. Louët, "A review to massive mimo detection algorithms: Theory and implementation," *Advanced Radio Frequency Antennas for Modern Communication and Medical Systems*, 2020.

[7] P. Sarangi and P. Pal, "Measurement matrix design for sample-efficient binary compressed sensing," *IEEE Signal Processing Letters*, 2022.

[8] M. Toyoda and M. Tanaka, "Efficient iterative method for soav minimization problem with linear equality and box constraints and its linear convergence," *Journal of the Franklin Institute*, 2022.

[9] H. Zhu and G. B. Giannakis, "Exploiting sparse user activity in multiuser detection," *IEEE Transactions on Communications*, 2011.

[10] M. Wu, C. Dick, J. R. Cavallaro, and C. Studer, "High-throughput data detection for massive mu-mimo-ofdm using coordinate descent," *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2016.

[11] A. Kudeshia, A. K. Jagannatham, and L. Hanzo, "Total variation based joint detection and state estimation for wireless communication in smart grids," *IEEE Access*, 2019.

[12] J.-C. Chen, "Manifold optimization approach for data detection in massive multiuser mimo systems," *IEEE Transactions on Vehicular Technology*, 2018.

[13] R. Hayakawa and K. Hayashi, "Convex optimization-based signal detection for massive overloaded mimo systems," *IEEE Transactions on Wireless Communications*, 2017.

[14] M. A. Albreem, W. Salah, A. Kumar, *et al.*, "Low complexity linear detectors for massive mimo: A comparative study," *IEEE Access*, 2021.

[15] M. Nagahara, "Discrete signal reconstruction by sum of absolute values," *IEEE Signal Processing Letters*, 2015.

[16] R. Hayakawa and K. Hayashi, "Discrete-valued vector reconstruction by optimization with sum of sparse regularizers," in *EUSIPCO*, 2019.

[17] J. Abe, M. Yamagishi, and I. Yamada, "Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition," *Inverse Problems*, 2020.

[18] Y. Censor, R. Davidi, and G. T. Herman, "Perturbation resilience and superiorization of iterative algorithms," *Inverse Problems*, 2010.

[19] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Superiorized adaptive projected subgradient method with application to mimo detection," *IEEE Transactions on Signal Processing*, 2023.

[20] W. Yata, M. Yamagishi, and I. Yamada, "A constrained LiGME model and its proximal splitting algorithm under overall convexity condition," *Journal of Applied and Numerical Optimization*, 2022.

[21] W. Yata and I. Yamada, "Imposing early and asymptotic constraints on ligme with application to bivariate nonconvex enhancement of fused lasso models," 2024. arXiv: 2309.14082.

[22] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Transactions on Signal Processing*, 2017.

[23] Y. Chen, M. Yamagishi, and I. Yamada, "A unified design of generalized moreau enhancement matrix for sparsity aware ligme models," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, 2023.

[24] D. Kitahara, R. Kato, H. Kuroda, and A. Hirabayashi, "Multi-contrast CSMRI using common edge structures with LiGME model," in *EUSIPCO*, 2021.

[25] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted $\ell_1$ minimization," *Journal of Fourier Analysis and Applications*, 2008.

[26] R. Hayakawa and K. Hayashi, "Reconstruction of complex discrete-valued vector via convex optimization with sparse regularizers," *IEEE Access*, 2018.