# A Unified Class of DC-type Convexity-Preserving Regularizers for Improved Sparse Regularization

Yi Zhang, Isao Yamada
*Department of Information and Communications Engineering*
*Tokyo Institute of Technology*, Tokyo, Japan
Email: {yizhang, isao}@sp.ict.e.titech.ac.jp

***Abstract*—Recently, a novel class of nonconvex sparse regularizers which can preserve the convexity of the cost function has gained considerable attention. These regularizers are expressed as difference of two convex functions, where the parameterized subtrahend function can be adjusted flexibly to maintain the overall convexity. In this paper we propose a unified class of such DC-type (Difference-of-Convex) convexity-preserving regularizers. By selecting proper kernel functions, the proposed regularizer reproduces existing convexity-preserving models and opens the way to a large number of promising new regularizers. In order to solve the convex but involved regularization model, we propose a novel iterative algorithm based on DC programming. Unlike normal DC algorithms, the proposed method is guaranteed to converge to a global minimizer of the cost function. In addition, compared to algorithms for existing convexity-preserving models, the proposed algorithm makes less stringent assumptions on the kernel functions, thus is more general. Moreover, the proposed algorithm can be interpreted as a nested forward-backward splitting method with overrelaxed step size, which leads to its fast convergence. Numerical experiments are conducted to verify the efficiency of our algorithm with comparisons to existing methods.**

***Index Terms*—sparse recovery, nonconvex regularizers, DC programming, convexity-preserving models**

## I. Introduction

The reconstruction of sparse signals usually relies on solving the following type of regularized least-squares problems:

$$\underset{x}{\text{minimize}} \;\; J(x) := \frac{1}{2}\|y - Ax\|_2^2 + \lambda\Psi(x), \tag{1}$$

where $y \in \mathbb{R}^m$, $A \in \mathbb{R}^{m\times n}$ is the measurement matrix, $\lambda > 0$ is a tuning parameter, and $\Psi : \mathbb{R}^n \to \mathbb{R}$ is a regularizer that evaluates sparseness of the solution.

Ideally, $\Psi$ should be the $l_0$ pseudo-norm, i.e., the number of nonzero components in $x$. However, this discontinuous regularizer leads to NP-hard optimization problem [1]. To circumvent this difficulty, one usually resorts to some continuous approximation of the $l_0$-norm in practice. Earlier studies usually adopt convex regularizers (e.g., $l_1$-norm [2], Huber function [3]) to ensure efficient solution of (1). Nevertheless, due to coercivity, such regularizers usually overpenalize the $i$th component $x_i$ when $|x_i|$ is large, which causes underestimation of the true solution [4]. To overcome this problem, continuous nonconvex regularizers (e.g., SCAD [5], MCP [4]) have been proposed to pursue less biased estimation. But normal nonconvex regularizers yield computationally expensive nonconvex programs, and the estimate may get stuck in local minima, which in turn causes performance degradation.

This work was supported in part by Grant-in-Aid for JSPS Fellows (21J22393) and by JST SICORP (JPMJSC20C6).

To resolve this situation, an unusual class of nonconvex regularizers which can preserve the convexity of the cost function has recently been proposed [6]–[8]. The first regularizer of this type is the generalized minimax concave (GMC) penalty [6]:

$$\Psi_{\text{GMC}}(x) := l_1(x) - (l_1 \Box q_B)(x), \tag{2}$$

where $l_1(x) := \|x\|_1$ is the $l_1$-norm, $q_B(x) := \frac{1}{2}\|Bx\|_2^2$ is a quadratic smoothing function with $B \in \mathbb{R}^{p\times n}$, $\Box$ is the infimal convolution operator. For $f, g : \mathbb{R}^n \to \mathbb{R}$, $(f\Box g)$ is defined as:

$$(f\Box g)(x) := \inf_{z\in\mathbb{R}^n} \left(f(z) + g(x-z)\right), \tag{3}$$

which is known to be convex if $f$ and $g$ are both convex [9]. If $B$ is the identity matrix, $\Psi_{\text{GMC}}$ reproduces the minimax concave penalty (MCP [4]), hence the GMC penalty is a nonseparable generalization of MCP. Remarkably, in contrast to the standard MCP, the shape of $\Psi_{\text{GMC}}$ can be regulated flexibly via changing $B$. It has been proven [6] that if

$$A^TA \succeq \lambda B^TB,$$

then the concavity of $-\lambda(l_1\Box q_B)(x)$ is overpowered by the convexity of $\frac{1}{2}\|y - Ax\|_2^2$, and the cost function $J$ is convex. Therefore, the GMC penalty can achieve less biased estimation without losing the overall convexity of the problem.

The powerful idea of the GMC penalty has attracted considerable attention, and efforts have been made to broaden its applicability [7], [8], [10]. One notable extension is the linearly involved generalized Moreau enhanced (LiGME) model [7]:

$$\Psi_{\text{LiGME}}(x) = \psi(Lx) - (\psi\Box q_B)(Lx), \tag{4}$$

where $L \in \mathbb{R}^{q\times n}$ is an analysis matrix which encodes the sparsifying domain of the signal $x$ (e.g., wavelet matrix [11], discrete differential operator [12]); $\psi$ is a kernel function which is no longer restricted to the $l_1$-norm, but can be any convex function with computable proximal operator [13]. Accordingly, the LiGME model allows applying the construction technique of GMC to more general convex kernel functions. On the other hand, another useful extension termed Sharpening Sparse Regularizers (SSR) framework [8] considers the following formulation:

$$\Psi_{\text{SSR}}(x) := l_1(x) - ((l_1 \circ L)\Box(\Phi \circ B))(x), \tag{5}$$

where $\circ$ is the function composition operator, the analysis matrix $L$ is embedded at a different position, and $\Phi$ is a smoothing function which is not restricted to the $l_2$-norm. While the SSR model does not consider variability of the kernel function, it allows adopting different smoothing functions $\Phi$, thus is able to regulate the shape of the regularizer more delicately. For both extensions, overall-convexity conditions and solution algorithms based on proximal splitting methods [13], [14] have been derived independently so far.

In this paper, we are devoted to unifying and generalizing the aforementioned studies. We propose a novel class of convexity-preserving regularizers and present its overall-convexity condition. By selecting proper kernel functions, the proposed regularizer reproduces existing convexity-preserving models [6]–[8] and opens the way to a large number of promising new regularizers. In addition, despite the involved formulation of the cost function, we derive a unified DC-type (Difference-of-Convex [15]) algorithm for minimizing it. The proposed algorithm solves the difficult original problem via solution to a sequence of simpler subproblems. As long as the subproblems are solvable, the proposed algorithm is implementable, which poses a less stringent requirement compared to prior arts [6]–[8]. Although DC algorithms are locally convergent in general, we establish the convergence of the proposed algorithm to a global minimizer under the overall-convexity condition, which ensures its reliability. Moreover, we unravel the connection of our algorithm with the algorithm proposed in the original paper of the GMC model [6], whereby the proposed algorithm can be interpreted as a nested forward-backward splitting method with overrelaxed step size. The interpretation leads to the fast convergence of our algorithm. Numerical experiments demonstrate superior convergence speed of the proposed algorithm over existing algorithms.

### *Notation*

Let $\mathbb{N}, \mathbb{R}$ be the sets of nonnegative integers and real numbers. For $n$-dimensional Euclidean space $\mathbb{R}^n$, $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_p$ $(p \geq 1)$ denote respectively the inner product and the $l_p$-norm in $\mathbb{R}^n$. $0_n$ stands for the $n \times 1$ zero vector. $\mathrm{I}_n$ denotes the $n \times n$ identity matrix. For $A \in \mathbb{R}^{m \times n}$, $A^T \in \mathbb{R}^{n \times m}$ denotes its transpose. For a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ denotes its Hessian matrix at $x$. For a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the subdifferential of $f$ denoted as $\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is the set-valued operator

$$\partial f : x \mapsto \{u \in \mathbb{R}^n \mid (\forall z \in \mathbb{R}^n)\ \langle z - x, u \rangle + f(x) \leq f(z)\}.$$

We denote $\Gamma_0(\mathbb{R}^n)$ as the set of all proper lower semicontinuous convex functions from $\mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$ [9]. For $f$ in $\Gamma_0(\mathbb{R}^n)$, the proximal operator of $f$ is defined as

$$\mathrm{Prox}_f(x) = \underset{z \in \mathbb{R}^n}{\arg\min} \left\{ f(z) + \frac{1}{2}\|x - z\|_2^2 \right\}.$$

We say that $f$ is proximable if $\mathrm{Prox}_{\gamma f}$ can be computed efficiently for every $\gamma > 0$.

TABLE I
PRIOR ARTS AS INSTANCES OF THE PROPOSED REGULARIZER

| | $\psi_1(x)$ | $\psi_2(y)$ | $\phi(y)$ | $D$ |
|---|---|---|---|---|
| GMC [6] | $\|x\|_1$ | $\|y\|_1$ | $\frac{1}{2}\|By\|_2^2$ | $I_n$ |
| LiGME [7] | $\psi(Lx)$ | $\psi(y)$ | $\frac{1}{2}\|By\|_2^2$ | $L$ |
| SSR [8] | $\|x\|_1$ | $\|Ly\|_1$ | $\Phi(By)$ | $I_n$ |

## II. A Unified Class of Convexity-Preserving Regularizers

In this section, we present a unified class of convexity-preserving regularizers with its overall-convexity condition.

### *A. Abstract Formulation*

The proposed class of convexity-preserving regularizers is formulated as follows:

$$\Psi_{\mathrm{CP}}(x) := \psi_1(x) - (\psi_2 \square \phi)(Dx), \tag{6}$$

where $\psi_1 \in \Gamma_0(\mathbb{R}^n), \psi_2 \in \Gamma_0(\mathbb{R}^p)$ are kernel functions, $D \in \mathbb{R}^{p \times n}$, the smoothing function $\phi \in \Gamma_0(\mathbb{R}^p)$ is twice continuously differentiable everywhere on $\mathbb{R}^p$.

We note that different from the GMC [6], LiGME [7] and SSR [8] models, the formulation of (6) is abstract, i.e., it does not require $\psi_1, \psi_2$ to be sparseness-promoting or require $D$ to encode the sparsifying domain of the interested signal. Instead, every regularizer of the form (6) can be regarded as an instance of $\Psi_{\mathrm{CP}}$. In this light, existing convexity-preserving regularizers can be reproduced by selecting proper kernel functions and smoothing function, as summarized in Table I.

Moreover, considering that $(\psi_2 \square \phi)$ is a smooth approximation of $\psi_2$ [16], $\Psi_{\mathrm{CP}}$ can be regarded as a partially smoothed approximation of the DC function $\psi_1(x) - \psi_2(Dx)$. Since many nonconvex regularizers encountered in compressive sensing are DC functions [17] and can be reformulated into the form of $\psi_1(x) - \psi_2(Dx)$, the proposed regularizer certainly encompasses a large number of promising new regularizers, which will be a direction of future research.

### *B. Overall-Convexity Condition*

By designing the shape of $\phi$ properly, the overall-convexity of the cost function

$$J_{\mathrm{CP}}(x) := \frac{1}{2}\|y - Ax\|_2^2 + \lambda \Psi_{\mathrm{CP}}(x) \tag{7}$$

can be preserved; see the following proposition.

**Proposition 1.** If $A^T A \succeq \lambda D^T \nabla^2 \phi(z) D$ holds for every $z \in \mathbb{R}^p$, then $J_{\mathrm{CP}}$ is a convex function.

*Proof.* The cost function can be rewritten as follows,

$$\begin{aligned} & \frac{1}{2}\|y - Ax\|_2^2 + \lambda \Psi_{\mathrm{CP}}(x) \\ = \ & \frac{1}{2}\|y - Ax\|_2^2 + \lambda \psi_1(x) - \lambda \inf_{v \in \mathbb{R}^p} \{\psi_2(v) + \phi(Dx - v)\} \\ = \ & \sup_{v \in \mathbb{R}^p} \left\{ \frac{1}{2}\|y - Ax\|_2^2 + \lambda \left[\psi_1(x) - \psi_2(v) - \phi(Dx - v)\right] \right\} \\ = \ & \sup_{v \in \mathbb{R}^p} \left\{ M_v(x) + \lambda \left[\psi_1(x) - \psi_2(v)\right] \right\}, \end{aligned}$$

where we define $M_v(x) \coloneqq \frac{1}{2}\|y - Ax\|_2^2 - \lambda\phi(Dx - v)$. Since

$$\nabla^2 M_v(x) = A^T A - \lambda D^T \nabla^2 \phi(Dx - v)D$$

is positive semidefinite for every $x \in \mathbb{R}^n$ from assumption (cf. [18, A.4.4] for derivation of $\nabla^2 M_v(x)$), $M_v$ is convex [19, Thm 2.1.4]. Therefore, $J_{\text{CP}}$ is the supremum of a class of convex functions $\{M_v(\cdot) + \lambda[\psi_1(\cdot) - \psi_2(v)]\}_{v\in\mathbb{R}^p}$, which yields the convexity of $J_{\text{CP}}$ from [9, Prop 8.16]. □

It should be noted that Proposition 1 essentially embraces overall-convexity conditions of existing convexity-preserving models [6]–[8]. Especially, if $\phi(\cdot) \coloneqq q_B(\cdot)$, then the overall-convexity condition required in Proposition 1 is simplified into

$$A^T A \succeq \lambda D^T B^T B D,$$

which is easily verifiable in practice.

## III. A DC-type Solution Algorithm

Under Proposition 1, the minimization of $J_{\text{CP}}(x)$ can be regarded as a special DC program [15] with overall-convexity, which admits the following DC decomposition:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad J_{\text{CP}}(x) \coloneqq g(x) - h(x), \tag{8}$$

where convex functions $g, h$ are defined as

$$g(x) \coloneqq \frac{1}{2}\|y - Ax\|_2^2 + \lambda\psi_1(x), \tag{9}$$

$$h(x) \coloneqq \lambda(\psi_2 \Box \phi)(Dx). \tag{10}$$

(To see convexity of $h$, please refer to [9, Prop 12.11])

Despite convexity, solving (8) is difficult due to the involved formulation of $h$. In this section, by generalizing our previous work [20], we derive a DC-type algorithm for (8) and establish its convergence to a global minimizer under mild conditions.

### A. Derivation of the Proposed Algorithm

The proposed algorithm considers applying a standard approach for DC programs, termed the *basic DCA scheme* [15], to (8). It consists in repeating the following two steps

**step 1**: obtain $u_k \in \partial h(x_k)$,
**step 2**: compute $x_{k+1}$ by

$$x_{k+1} \in \underset{x\in\mathbb{R}^n}{\arg\min}\, g(x) - \langle u_k, x\rangle,$$

until the sequence of estimates $(x_k)_{k\in\mathbb{N}}$ converges.

For normal DC programs where $h$ has a closed-form expression, the step 1 is straightforward. However, the infimal convolution operator in (10) poses a computational difficulty in computing $u_k \in \partial h(x_k)$. Fortunately, we resolve this obstacle via the following proposition.

**Proposition 2.** Suppose that for every $x \in \mathbb{R}^n$, the following optimization program has at least one solution:

$$\min_{v\in\mathbb{R}^p} \ \psi_2(v) + \phi(Dx - v), \tag{11}$$

then $\partial h$ is single-valued and can be computed by

$$\partial h(x) = \{\lambda D^T \nabla\phi(Dx - v_x)\}, \tag{12}$$

where $v_x$ is an arbitrary solution of (11), and the RHS of (12) does not depend on the choice of such a solution $v_x$.

*Proof sketch.* According to [9, Thm 16.47],

$$\partial h(x) = \lambda\partial\left((\psi_2 \Box \phi) \circ D\right)(x) = \lambda D^T \partial(\psi_2 \Box \phi)(Dx). \tag{13}$$

Since the definition of $v_x$ implies that

$$(\psi_2 \Box \phi)(Dx) = \psi_2(v_x) + \phi(Dx - v_x),$$

we have the following from [9, Prop 16.61]:

$$\partial(\psi_2 \Box \phi)(Dx) = \partial\psi_2(v_x) \cap \partial\phi(Dx - v_x).$$

Smoothness of $\phi$ implies that

$$\partial\phi(Dx - v_x) = \{\nabla\phi(Dx - v_x)\},$$

hence $\partial(\psi_2 \Box \phi)(Dx) \subset \{\nabla\phi(Dx - v_x)\}$. Combining this with the nonemptiness of $\partial(\psi_2 \Box \phi)(Dx)$ yields

$$\partial(\psi_2 \Box \phi)(Dx) = \{\nabla\phi(Dx - v_x)\}. \tag{14}$$

Substituting (14) into (13) completes the proof. □

Applying Proposition 2 to the basic DCA scheme yields Algorithm 1. It should be noted that Algorithm 1 is implementable as long as (15) and (16) can be solved by some inner iterative algorithms. In particular, if $\psi_1, \psi_2$ are proximable functions or their composition with linear operators, which is the case of prior arts [6]–[8], then (15) and (16) can be solved efficiently via, e.g., ADMM [21]. Therefore, Algorithm 1 is more general than algorithms for existing convexity-preserving models.

---

**Algorithm 1:** The Proposed DC Algorithm

**Initialization:** $k = 0, x_0 \in \mathbb{R}^n$
Repeat the following steps until convergence.
**Step 1:** obtain $v_k$ by

$$v_k \in \underset{v\in\mathbb{R}^p}{\arg\min}\, \psi_2(v) + \phi(Dx_k - v), \tag{15}$$

and compute $u_k = \lambda D^T \nabla\phi(Dx_k - v_k)$.
**Step 2:** compute $x_{k+1}$ by

$$x_{k+1} \in \underset{x\in\mathbb{R}^n}{\arg\min}\, \frac{1}{2}\|y - Ax\|_2^2 + \lambda\psi_1(x) - \langle u_k, x\rangle, \tag{16}$$

and update $k \leftarrow k + 1$.

---

### B. Convergence Properties

It is well known that every limit point $x_*$ of the sequence $(x_k)_{k\in\mathbb{N}}$ generated by the basic DCA scheme is a critical point of $g - h$, i.e., $0_n \in \partial g(x_*) - \partial h(x_*)$ [22]. However, being a critical point of $g - h$ is a necessary but insufficient condition for being a local minimizer. Moreover, even if $g - h$ is convex, the convergence guarantee generally does not improve because $\partial(g - h)(x) \neq \partial g(x) - \partial h(x)$. Accordingly, one may concern that Algorithm 1 yields worse convergence guarantee in comparison with existing algorithms based on proximal splitting methods [6]–[8]. Fortunately, the following theorem establishes convergence of Algorithm 1 to a global minimizer of $J_{\mathrm{CP}}$, which dispels aforementioned concerns.

**Theorem 1.** Let $(x_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 1. Suppose that conditions required in Proposition 1 and 2 are satisfied, and that $\arg\min_{x\in\mathbb{R}^n} J_{\mathrm{CP}}(x)$ is nonempty and bounded, then every limit point of $(x_k)_{k\in\mathbb{N}}$ is a global minimizer of $J_{\mathrm{CP}}$.

*Proof sketch.* Since $J_{\mathrm{CP}} = g - h$, we have $J_{\mathrm{CP}} + h = g$. Then from [9, Thm 16.47], $\partial J_{\mathrm{CP}}(x) + \partial h(x) = \partial g(x)$ holds for any $x \in \mathbb{R}^n$. Since $\partial h(x)$ is single-valued from Proposition 2, we can guarantee that $\partial J_{\mathrm{CP}}(x) = \partial g(x) - \partial h(x)$.

Let $x_*$ be a limit point of $(x_k)_{k\in\mathbb{N}}$, then the condition that $x_*$ is a critical point of $g - h$ yields that

$$0_n \in \partial g\left(x_*\right) - \partial h\left(x_*\right) \implies 0_n \in \partial J_{\mathrm{CP}}\left(x_*\right).$$

Since $J_{\mathrm{CP}}$ is convex from Proposition 1, the condition above implies that $x_*$ is a global minimizer of $J_{\mathrm{CP}}$ [9, Thm. 16.3]. □

### C. On the Efficiency of Algorithm 1

Although it is difficult to analyse the convergence rate of Algorithm 1 theoretically, we can establish an intuitive interpretation for its empirically fast convergence. In this section, we unravel the connection of Algorithm 1 with an existing algorithm for solving the GMC model. The latter algorithm, firstly proposed in the pioneering work of Selesnick [6], is based on forward-backward splitting method [13].

Here we restrict our discussion to the GMC model, i.e., we assume that $\psi_1(\cdot) = \psi_2(\cdot) = \|\cdot\|_1$, $\phi(\cdot) = q_B(\cdot)$, $D = \mathrm{I}_n$. We consider a special implementation of Algorithm 1, assuming that in the basic DCA scheme, the subgradient $u_k$ is computed at $x_{k-1}$ instead of $x_k$. In addition, we solve (15) and (16) by forward-backward splitting (FBS) method introduced in [13]. For the $k$th outer iteration of Algorithm 1, applying FBS to (15) yields the following inner iterative steps:

$$\begin{aligned} s_{k,i+1} &= v_{k,i} - \rho_1 B^T B\left(v_{k,i} - x_{k-1}\right), \\ v_{k,i+1} &= \mathrm{soft}\left(s_{k,i+1}, \rho_1\right), \end{aligned}$$

where $v_{k,i}$ is the $i$th inner estimate of $v_k$, $s_{k,i}$ is an auxiliary variable. $\rho_1 \in (0, 2/\|B^T B\|_2)$ is the step size. $\mathrm{soft}(\cdot)$ is the soft thresholding operator. We adopt $v_{k,0} = v_{k-1}$ as the initial guess. Similarly, applying FBS to (16) yields

$$\begin{aligned} w_{k,j+1} &= x_{k+1,j} - \rho_2\left[A^T\left(Ax_{k+1,j} - y\right) - u_k\right] \\ x_{k+1,j+1} &= \mathrm{soft}\left(w_{k,j+1}, \rho_2\lambda\right), \end{aligned}$$

where $\rho_2 \in (0, 2/\|A^T A\|_2)$, $x_{k+1,0} = x_k$ is the initial guess. In addition, we assume that the inner iterative procedures for (15) and (16) are both terminated after only one inner iteration, i.e., we set $v_k = v_{k,1}$, $x_{k+1} = x_{k+1,1}$. Adopting the same design of $B$ as [6] (i.e., $\lambda B^T B = \gamma A^T A$ with $\gamma \in (0,1)$) yields the following implementation of Algorithm 1:

$$\begin{aligned} w_k &= x_k - \rho_2 A^T\left[A\left(x_k + \gamma\left(v_k - x_k\right)\right) - y\right], \\ s_{k+1} &= v_k - \rho_1' \gamma A^T A(v_k - x_k), \\ x_{k+1} &= \mathrm{soft}(w_k, \rho_1'\lambda), \\ v_{k+1} &= \mathrm{soft}(s_{k+1}, \rho_2\lambda), \end{aligned}$$

where $\rho_1' = \rho_1/\lambda \in \left(0, 2/\left(\gamma\|A^T A\|_2\right)\right)$.

It can be verified that the preceding is exactly the iterative steps of Selesnick's algorithm, except that in [6], $\rho_1'$ and $\rho_2$ are replaced by a common step size $\rho$ with the value range

$$\rho \in \left(0, \min\left\{2/\|A^T A\|_2, 2(1-\gamma)/\left(\gamma\|A^T A\|_2\right)\right\}\right).$$

If $\gamma \to 1$, it can be verified that the upper bound of $\rho$ goes to zero, whilst that of $\rho_1'$ and $\rho_2$ go to $2/\|A^T A\|_2$. Since in the GMC model, larger $\gamma$ generally leads to less biased estimation [7, Example 2], Algorithm 1 can be regarded as a nested forward-backward splitting method with overrelaxed step size in this case. Therefore, it is reasonable to expect Algorithm 1 to achieve fast convergence.

## IV. Numerical Experiments

To verify the efficiency of Algorithm 1, we conduct numerical experiments in a scenario of standard sparse recovery problems. Sparse signal $x_\star \in \mathbb{R}^{1000}$ is generated as follows: 20 out of 1000 components are uniformly selected to be nonzero, the value of which follow standard normal distribution. The observation is $y = Ax_\star + \epsilon$, where the entries of $A \in \mathbb{R}^{200\times 1000}$ follow the standard normal distribution, $\epsilon$ is additive white Gaussian noise. The signal-to-noise ratio (SNR) is 30dB, which is defined as $\mathrm{SNR} := 20\log_{10}\left(\|Ax_\star\|_2/\|\epsilon\|_2\right)$ (dB). We adopt GMC penalty [6] as the sparse regularizer. Since LiGME [7], SSR [8] and the proposed regularizer are all extensions of the GMC penalty, all of them can reproduce the GMC penalty by properly selecting kernel functions, whereby we can conduct a fair comparison between Algorithm 1 and the algorithms proposed in previous studies [6]–[8]. For Algorithm 1, we solve (15) and (16) by ISTA [23]. We consider the $l_1$-regularization model (solved by ISTA) as a baseline for conventional regularization models. All results are averaged over 100 Monte Carlo runs.

Fig. 1 shows dependency of Mean Squared Error (MSE) on the regularization parameter $\lambda$ for the $l_1$-regularization model and the convexity-preserving models (which amounts to the GMC model in this case). The results indicate over 3dB estimation performance gain of the convexity-preserving model over conventional $l_1$-regularization model.

Fig. 2 shows MSE versus computation time for various algorithms. The circle mark respectively represent 1000 iterations of GMC, 100 iterations of LiGME/SSR and 1 outer iteration of Algorithm 1. From Fig. 2, the ISTA algorithm (Fig. 2: magenta) for $l_1$-regularization model converges especially fast, but can only obtain a less satisfactory estimate. Algorithms for convexity-preserving models achieve more accurate estimation, with convergence speed: Algorithm 1 > LiGME

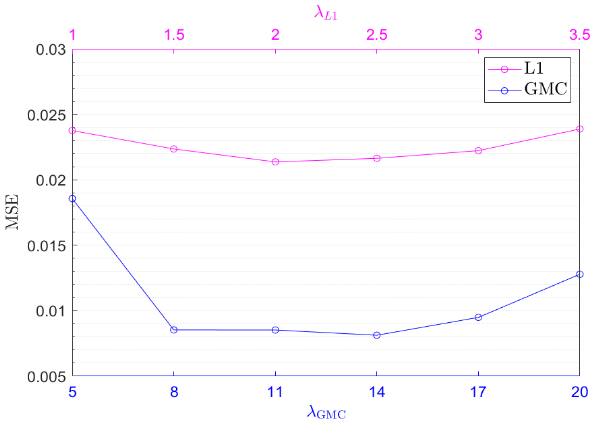


Fig. 1. MSE vs regularization parameter.

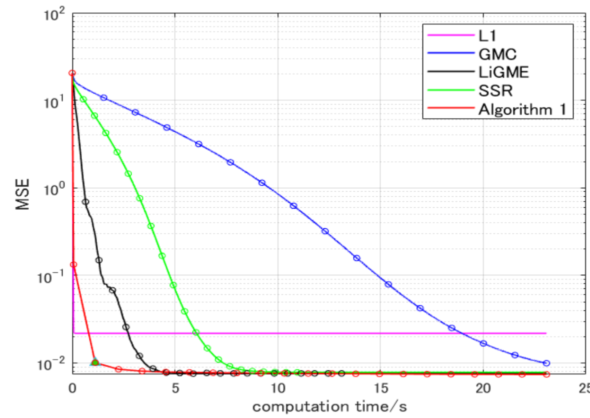


Fig. 2. MSE vs computation time.

> SSR > GMC. In addition, we remark that the triangle mark on the curve of Algorithm 1 represents its second outer iteration. This implies that as observed in our previous work [20], merely two iterations of Algorithm 1 suffice to produce a satisfactory solution, which serves as a useful stopping criterion in practice.

## V. Concluding Remarks

In this paper, we have proposed a unified class of convexity-preserving sparse regularizers, and have presented a general overall-convexity condition for it. In order to solve the resulting regularization model, we have proposed a unified DC-type solution algorithm. We have established the global convergence of the proposed algorithm, and have provided an intuitive interpretation for its fast convergence. Numerical experiments have demonstrated the efficiency of the proposed method.

## References

[1] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. New York, NY: Springer New York, 2013.

[2] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[3] J. A. Fessler, "Grouped coordinate descent algorithms for robust edge-preserving image restoration," in *Image Reconstruction and Restoration II*, vol. 3170. SPIE, 1997, pp. 184 – 194.

[4] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *The Annals of Statistics*, vol. 38, no. 2, 2010.

[5] J. Fan and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, 2001.

[6] I. Selesnick, "Sparse Regularization via Convex Analysis," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4481–4494, 2017.

[7] J. Abe, M. Yamagishi, and I. Yamada, "Linearly involved generalized Moreau enhanced models and their proximal splitting algorithm under overall convexity condition," *Inverse Problems*, vol. 36, no. 3, p. 035012, 2020.

[8] A. H. Al-Shabili, Y. Feng, and I. Selesnick, "Sharpening Sparse Regularizers via Smoothing," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 396–409, 2021.

[9] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, ser. CMS Books in Mathematics. Cham: Springer International Publishing, 2017.

[10] M. Yukawa, H. Kaneko, K. Suzuki, and I. Yamada, "Linearly-involved Moreau-Enhanced-over-Subspace Model: Debiased Sparse Modeling and Stable Outlier-Robust Regression," *arXiv preprint arXiv: 2201.03235*, 2022.

[11] J.-L. Starck, F. Murtagh, and J. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge: Cambridge University Press, 2010.

[12] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, "An Introduction to Total Variation for Image Analysis," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, M. Fornasier, Ed. De Gruyter, 2010, pp. 263–340.

[13] P. L. Combettes and J.-C. Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds. New York, NY: Springer New York, 2011, pp. 185–212.

[14] I. Yamada, M. Yukawa, and M. Yamagishi, "Minimizing the Moreau Envelope of Nonsmooth Convex Functions over the Fixed Point Set of Certain Quasi-Nonexpansive Mappings," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. New York, NY: Springer New York, 2011, pp. 345–390.

[15] H. A. Le Thi and T. Pham Dinh, "DC programming and DCA: thirty years of developments," *Mathematical Programming*, vol. 169, no. 1, pp. 5–68, 2018.

[16] A. Beck and M. Teboulle, "Smoothing and First Order Methods: A Unified Framework," *SIAM Journal on Optimization*, vol. 22, no. 2, pp. 557–580, Jan. 2012.

[17] H. Le Thi, T. Pham Dinh, H. Le, and X. Vo, "DC approximation approaches for sparse optimization," *European Journal of Operational Research*, vol. 244, no. 1, pp. 26–46, 2015.

[18] S. Boyd and L. Vandenberghe, *Convex Optimization*, 1st ed. Cambridge University Press, 2004.

[19] Y. Nesterov, *Lectures on Convex Optimization*, ser. Springer Optimization and Its Applications. Cham: Springer International Publishing, 2018, vol. 137.

[20] Y. Zhang and I. Yamada, "DC-LiGME: An Efficient Algorithm for Improved Convex Sparse Regularization," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, 2021.

[21] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[22] L. T. H. An and P. D. Tao, "The DC (Difference of Convex Functions) Programming and DCA Revisited with DC Models of Real World Nonconvex Optimization Problems," *Annals of Operations Research*, vol. 133, no. 1-4, pp. 23–46, 2005.

[23] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.