A Fast DC Algorithm for a Unified Class of Convexity-Preserving Sparse Regularizers

05001603 東京工業大学 *張 毅 ZHANG Yi
東京工業大学 山田 功 YAMADA Isao

1. Introduction

Obtaining sparse solutions to linear systems has been a long-standing challenge in data science, which is usually formulated as the following sparse regularization problem:

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

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 the sparseness-promoting regularizer which approximates the l_0 pseudo-norm (i.e., the cardinality of nonzero components in x).

Among conventional regularizers, convex ones such as the l_1-norm adopted in LASSO [1] ensure efficient solution of (1), whereas they usually suffer from estimation bias. Nonconvex regularizers, on the other hand, can yield debiased estimation, but lead to higher computational cost.

Distinct from conventional regularizers, a novel class of nonconvex regularizers which can maintain the convexity of the cost function J has been proposed recently. A pioneering work is the generalized minimax concave (GMC) penalty [2]

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

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

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

One can verify from Fig. 1 that Ψ_{GMC} is expressed as the difference between the l_1-norm and a smooth approximation of it. In the latter function $(l_1 \Box q_B)$, l_1 is a kernel regularizer which determines the function shape in the outer zone,

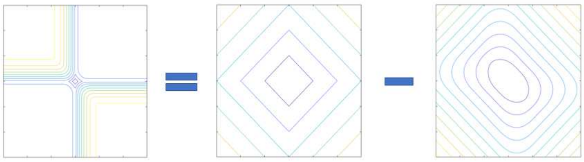

Fig. 1: Illustration of the GMC penalty

whereas $q_B(x)$ is a smoothing function that determines the shape in the inner zone. Remarkably, if the steering matrix B is properly selected, then the subtrahend function $(l_1 \Box q_B)$ is overpowered by the data fidelity $\frac{1}{2}\|y - Ax\|_2^2$, whereby Ψ_{GMC} can yield debiased estimation without losing the overall convexity of J [2]. To broaden the applicability of Ψ_{GMC}, useful extensions which enable variability of the kernel regularizer [3] and the smoothing function [4] have been proposed.

In this report, we introduce a unified class of convexity-preserving regularizers and develop an efficient solution algorithm for it based on difference-of-convex (DC) programming. We establish the global convergence of the proposed algorithm. Numerical experiments verify the efficiency of the proposed algorithm.

2. A Unified Class of Convexity-Preserving Regularizers

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

$$\Psi_{\mathrm{CP}}(x) := \psi_1(x) - (\psi_2 \Box \phi)(Dx), \quad (4)$$

where $\psi_1 \in \Gamma_0(\mathbb{R}^n)$[1], $\psi_2 \in \Gamma_0(\mathbb{R}^p)$ are kernel functions, $D \in \mathbb{R}^{p\times n}$, the smoothing function $\phi \in \Gamma_0(\mathbb{R}^p)$ is twice continuously differentiable.

[1] $\Gamma_0(\mathbb{R}^n)$ denotes the set of proper lower semicontinuous convex functions from $\mathbb{R}^n$ to $\mathbb{R} \cup \{+\infty\}$.

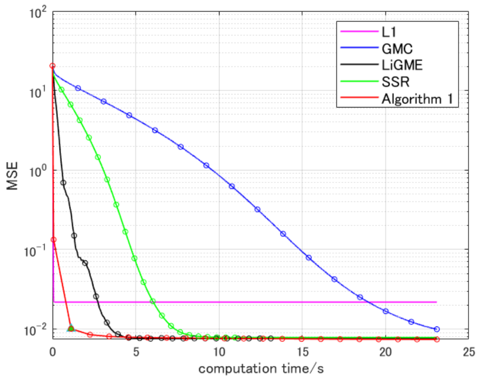

Fig. 2: MSE vs. computation time

Table 1: Prior arts as instances of Ψ_{CP}

	$\psi_1(x)$	$\psi_2(y)$	$\phi(y)$	D
GMC [2]	$\|x\|_1$	$\|y\|_1$	$\frac{1}{2}\|By\|_2^2$	I_n
LiGME [3]	$\psi(Lx)$	$\psi(y)$	$\frac{1}{2}\|By\|_2^2$	L
SSR [4]	$\|x\|_1$	$\|Ly\|_1$	$\Phi(By)$	I_n

By selecting proper building blocks, Ψ_{CP} reproduces existing convexity-preserving regularizers (cf. Table 1). Moreover, since Ψ_{CP} is a partially smoothed approximation of $\psi_1(x)-\psi_2(Dx)$ and many nonconvex sparse regularizers can be viewed as difference-of-convex (DC) functions [5], the proposed regularizer certainly encompasses a large number of promising new regularizers. The overall-convexity condition of Ψ_{CP} is as follows.

Proposition 1. If $A^TA \succeq \lambda D^T \nabla^2 \phi(z) D$ for every $z \in \mathbb{R}^p$, then $J_{\mathrm{CP}}(x) := \frac{1}{2}\|y-Ax\|_2^2+\lambda\Psi_{\mathrm{CP}}(x)$ is a convex function.

3. A DC-Type Solution Algorithm

J_{CP} admits the following DC decomposition:

$$J_{\mathrm{CP}}(x) = \underbrace{\frac{1}{2}\|y - Ax\|_2^2 + \lambda\psi_1(x)}_{g} - \underbrace{\lambda(\psi_2 \square \phi)(Dx)}_{h}.$$

With respect to this, we developed Alg. 1 based on a standard approach for minimizing DC functions named *the basic DCA scheme* [5]. Remarkably, we established the global convergence of Alg. 1, which makes a huge difference from normal DC algorithms that are locally convergent.

Theorem 1. Let $(x_k)_{k\in\mathbb{N}}$ be a sequence generated by Alg. 1. Suppose that $\arg\min_{x\in\mathbb{R}^n} J_{\mathrm{CP}}(x)$ is nonempty and bounded, and that the condition required in Prop. 1 holds, then every limit point of $(x_k)_{k\in\mathbb{N}}$ is a global minimizer of J_{CP}.

Algorithm 1: The Proposed Algorithm

Initialization: $k = 0, x_0 \in \mathbb{R}^n$.

Repeat the following until convergence.

Step 1: obtain v_k by

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

and compute $u_k = \lambda D^T \nabla\phi(Dx_k - v_k)$.

Step 2: solve

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

and update $k \leftarrow k + 1$.

4. Numerical Experiments

We conducted experiments in standard sparse recovery problems. The results are depicted in Fig. 2, which verified superior efficiency of Algorithm 1 in comparisons to existing methods.

References

[1] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc., B: Stat. Methodol.*, 1996.

[2] I. Selesnick, "Sparse Regularization via Convex Analysis," *IEEE Trans. Signal Process.*, 2017.

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

[4] A. H. Al-Shabili, Y. Feng, and I. Selesnick, "Sharpening Sparse Regularizers via Smoothing," *IEEE Open J. Signal Process.*, 2021.

[5] H. A. Le Thi and T. Pham Dinh, "DC programming and DCA: thirty years of developments," *Math. Program.*, 2018.