Computing an Entire Solution Path of a Nonconvexly Regularized Convex Sparse Model

Yi Zhang, Isao Yamada (email: {yizhang,isao}@sp.ce.titech.ac.jp)
Department of Information & Communications Engineering, Tokyo Institute of Technology

Tokyo Tech

Introduction

1. Sparse least-squares problems:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}}\ J(\boldsymbol{x}) \coloneqq \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\Psi(\boldsymbol{x}). \tag{1}$$

$\boldsymbol{y}\in\mathbb{R}^m$: measurement $\boldsymbol{A}\in\mathbb{R}^{m\times n}$: sensing matrix Ψ: sparse regularizer

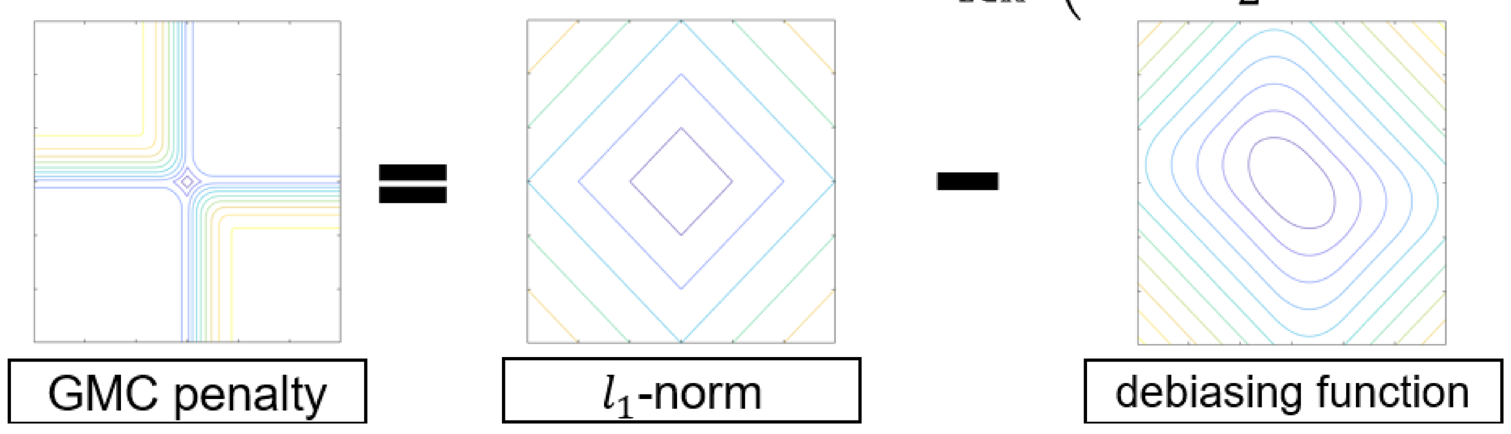
2. Conventional convex/nonconvex sparse regularizers:

(1) ℓ_1-norm (LASSO [1]): convex ☺ biased ☹

(2) the minimax concave (MC [2]) penalty: less biased ☺ nonconvex ☹

3. The generalized minimax concave (GMC [3]) penalty:

$$\Psi(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 - \min_{\boldsymbol{z}\in\mathbb{R}^n}\left(\|\boldsymbol{z}\|_1 + \frac{1}{2}\|\boldsymbol{B}(\boldsymbol{x}-\boldsymbol{z})\|_2^2\right)$$



(1) Overall convexity condition: if $\boldsymbol{B}\in\mathbb{R}^{p\times n}$ satisfies $\boldsymbol{A}^\top\boldsymbol{A}\succeq\lambda\boldsymbol{B}^\top\boldsymbol{B}$, then $J(\cdot)$ in (1) is convex despite nonconvexity of the GMC penalty.

(2) We study the scaled GMC (sGMC) model: $\boldsymbol{B}\coloneqq\sqrt{\frac{\rho}{\lambda}}\boldsymbol{A}$ with $\rho\in[0,1)$.

4. Key contributions: we prove that the minimum ℓ_2-norm solution path of the sGMC model is piecewise linear with respect to λ, and can be computed by an iterative algorithm within finite steps.

Theoretical Results

1. Optimality condition of the sGMC model:

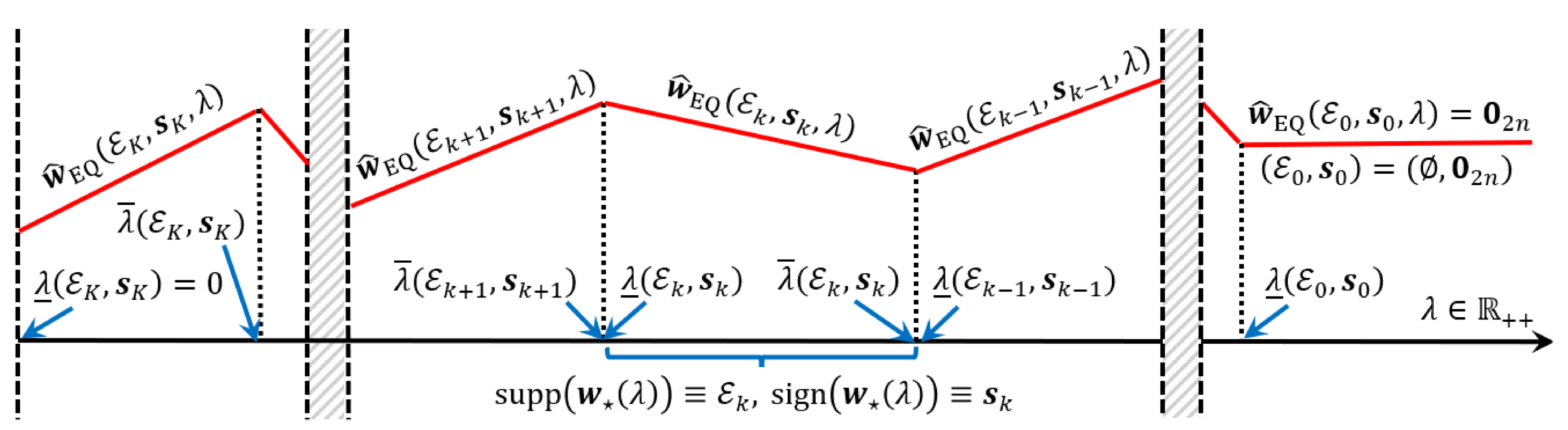
$\boldsymbol{x}\in\mathbb{R}^n$ is a solution of the sGMC model if and only if there exists $\boldsymbol{z}\in\mathbb{R}^n$, such that $\boldsymbol{w}\coloneqq\begin{bmatrix}\boldsymbol{x}^\top & \boldsymbol{z}^\top\end{bmatrix}^\top$ satisfies

$$\boldsymbol{0}\in\boldsymbol{C}^T(\boldsymbol{D}\boldsymbol{C}\boldsymbol{w}-\boldsymbol{b})+\lambda\partial(\|\cdot\|_1)(\boldsymbol{w}), \tag{2}$$

where $\boldsymbol{C}\coloneqq\operatorname{blkdiag}(\boldsymbol{A},\boldsymbol{A})$, $\boldsymbol{b}\coloneqq\begin{bmatrix}\boldsymbol{y}^\top & \boldsymbol{0}^\top\end{bmatrix}^\top$, $\boldsymbol{D}\coloneqq\begin{bmatrix}(1-\rho)\boldsymbol{I} & \rho\boldsymbol{I}\\ -\rho\boldsymbol{I} & \rho\boldsymbol{I}\end{bmatrix}$.

We call $\boldsymbol{w}\in\mathbb{R}^{2n}$ an *extended solution* if it satisfies (2). In this paper, we focus on the minimum ℓ_2-norm extended solution $\boldsymbol{w}_\star(\lambda)$.

2. Properties of the minimum ℓ_2-norm extended solution:



(1) $\boldsymbol{w}_\star(\lambda)$ is piecewise linear in λ with finite ($\leq 3^{2n}$) linear pieces.

(2) Within every linear piece of $\boldsymbol{w}_\star(\lambda)$, there exist uniquely a combination of $\mathcal{E}\subset\{1,2,\ldots,2n\}$ and $\boldsymbol{s}\in\{-1,0,1\}^{2n}$ such that:

- $\operatorname{supp}(\boldsymbol{w}_\star(\lambda))\equiv\mathcal{E}$, $\operatorname{sign}(\boldsymbol{w}_\star(\lambda))\equiv\boldsymbol{s}$ are constant.
- $\boldsymbol{w}_\star(\lambda)\equiv\hat{\boldsymbol{w}}_{\rm EQ}(\mathcal{E},\boldsymbol{s},\lambda)$, the latter is the least-squares solution of

$$\begin{cases}(\forall i\in\mathcal{E}) & \boldsymbol{c}_i^T(\boldsymbol{b}-\boldsymbol{D}\boldsymbol{C}\boldsymbol{w})=\lambda s_i, & \text{(EQ-a)}\\ (\forall j\in\neg\mathcal{E}) & w_j=0, & \text{(EQ-b)}\end{cases}$$

(3) The duration of λ for a linear piece is the set of $\lambda>0$ satisfying

$$\begin{cases}(\forall i\in\mathcal{E}) & s_i w_i\geq 0. & \text{(NQ-a)}\\ (\forall j\in\neg\mathcal{E}) & |\boldsymbol{c}_j^T(\boldsymbol{b}-\boldsymbol{D}\boldsymbol{C}\boldsymbol{w})|\leq\lambda, & \text{(NQ-b)}\end{cases}$$

Algorithmic Results

1. Least Angle Regression for the sGMC model:

The minimum ℓ_2-norm sGMC solution path $\boldsymbol{w}_\star(\lambda)$ can be computed by an extension of the well-known least angle regression (LARS [4]) algorithm:

(1) In each iteration, the proposed LARS-sGMC algorithm computes $(\mathcal{E},\boldsymbol{s})$ corresponding to one linear piece of $\boldsymbol{w}_\star(\lambda)$.

(2) Let $(\mathcal{E}_k,\boldsymbol{s}_k)$ corresponds to the kth linear piece, then:

- The LARS-sGMC iteration computes the lower breakpoint $\underline{\lambda}(\mathcal{E}_k,\boldsymbol{s}_k)$ of the current linear piece by solving (NQ).
- We conduct certain deletion and insertion operations on indices in $\mathcal{E}_k$ to obtain $\mathcal{E}_{k+1}$, and change the components of $\boldsymbol{s}_k$ in response to the change of $\mathcal{E}_k$ to yield $\boldsymbol{s}_{k+1}$.

See Algorithm 1 in the extended version of this paper [5] for details.

2. Properties of LARS-sGMC:

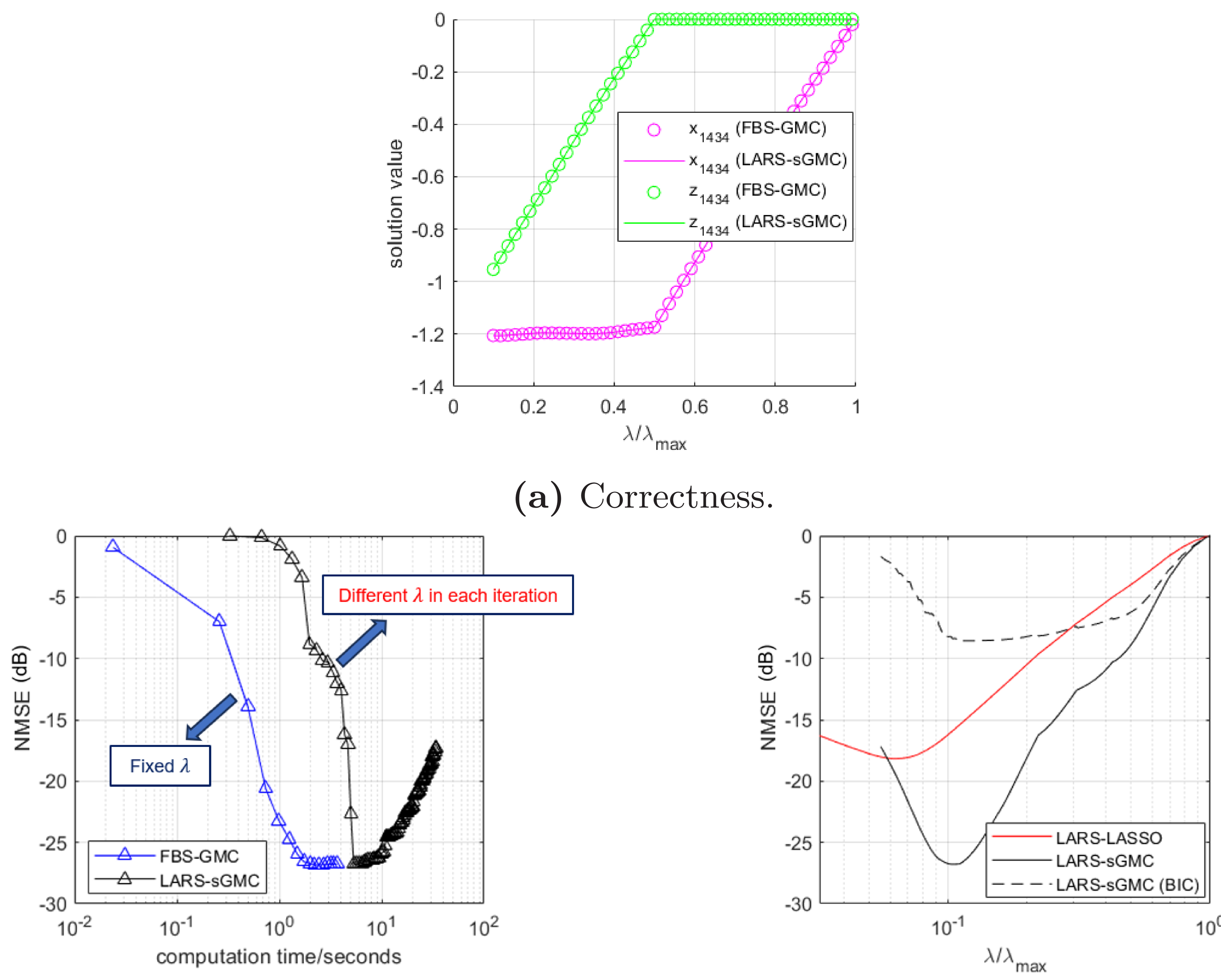
(1) We prove the correctness and finite termination of LARS-sGMC under a mild assumption.

(2) The complexity of the kth iteration is $\mathcal{O}\left(m|\mathcal{E}_k|^2+|\mathcal{E}_k|^3\right)$.

(3) If we set $\rho=0$ in the sGMC model, then LARS-sGMC reduces to the conventional LARS [4] algorithm for LASSO.

Experiments

We demonstrate the correctness, efficiency and practical utility for regularization parameter tuning of the LARS-sGMC algorithm.



(a) Correctness.

(b) Efficiency. **(c)** Practical utility.

Related Publications

[1] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc., B: Stat. Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[2] C. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Stat.*, vol. 38, no. 2, pp. 894 – 942, 2010.

[3] I. Selesnick, "Sparse Regularization via Convex Analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, 2017.

[4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Stat.*, vol. 32, no. 2, pp. 407 – 499, 2004.

[5] Y. Zhang and I. Yamada, "Solution-Set Geometry and Regularization Path of a Nonconvexly Regularized Convex Sparse Model," Mar 2024, arXiv: 2311.18438 [math].