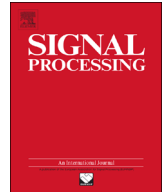




ELSEVIER

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Generalized LASSO with under-determined regularization matrices

Junbo Duan^a, Charles Soussen^b, David Brie^b, Jérôme Idier^c, Mingxi Wan^a, Yu-Ping Wang^{d,*}

^a Key Laboratory of Biomedical Information Engineering of Ministry of Education and Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China

^b Université de Lorraine and the CNRS at the CRAN (Centre de Recherche en Automatique de Nancy), UMR 7039, Campus Sciences, BP 70239, 54506 Vandoeuvre-lès-Nancy Cedex, France

^c L'UNAM Université, Ecole Centrale de Nantes and the CNRS at the Institut de Recherche en Communications et Cybernétique de Nantes (IRCCyN), UMR 6597, 1 rue de la Noë, BP 92101, F-44321 Nantes Cedex 3, France

^d Department of Biomedical Engineering, and the Department of Biostatistics and Bioinformatics, Lindy Boggs Center, Tulane University, New Orleans, LA 70118, USA



ARTICLE INFO

Article history:

Received 13 January 2015

Received in revised form

5 November 2015

Accepted 1 March 2016

Available online 10 March 2016

Keywords:

LASSO

Generalized LASSO

Robust LASSO

Deconvolution

Solution path

Diagonally dominant

Total variation

ABSTRACT

This paper studies the intrinsic connection between a generalized LASSO and a basic LASSO formulation. The former is the extended version of the latter by introducing a regularization matrix to the coefficients. We show that when the regularization matrix is even- or under-determined with full rank conditions, the generalized LASSO can be transformed into the LASSO form via the Lagrangian framework. In addition, we show that some published results of LASSO can be extended to the generalized LASSO, and some variants of LASSO, e.g., robust LASSO, can be rewritten into the generalized LASSO form and hence can be transformed into basic LASSO. Based on this connection, many existing results concerning LASSO, e.g., efficient LASSO solvers, can be used for generalized LASSO.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The least absolute shrinkage and selection operator (LASSO) [1] is one of the most popular approaches for sparse linear regression in the last decade, which is usually

formulated as

$$\mathbf{x}^*(\lambda) = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{x}\|_1 \right\}, \quad (\text{LASSO})$$

where $\mathbf{y} \in \mathbb{R}^n$ gathers n observed measurements; $\mathbf{A} \in \mathbb{R}^{n \times p}$ contains p predictors of dimension n ; $\mathbf{x} \in \mathbb{R}^p$ contains p coefficients; $\|\cdot\|$ and $\|\cdot\|_1$ stand for the ℓ_2 - and ℓ_1 -norm respectively; and $\lambda > 0$ is the regularization parameter, controlling the tradeoff between the data fidelity and the model complexity. The most attracting feature of LASSO is the use of ℓ_1 -norm regularization, which yields sparse coefficients. The ℓ_1 -norm regularization results in the piecewise linearity [2] of the solution path $\{\mathbf{x}^*(\lambda) | \lambda \in (0, +\infty)\}$ (i.e., the set of solutions with respect to continuous change of the regularization parameter λ), allowing for efficient

* Corresponding author.

E-mail addresses: junbo.duan@mail.xjtu.edu.cn (J. Duan),

charles.soussen@univ-lorraine.fr (C. Soussen),

david.brie@univ-lorraine.fr (D. Brie),

jerome.idier@ircyn.ec-nantes.fr (J. Idier),

mxwan@mail.xjtu.edu.cn (M. Wan), wyp@tulane.edu (Y.-P. Wang).

reconstruction of the whole solution path. Based on this property, famous path tracking algorithms such as the least angle regression (LARS) [3] and the homotopy [2,4] have been developed. LARS is a greedy algorithm working with decreasing λ value. At each iteration, a dictionary atom is selected and appended to the set previously selected, and the next critical λ value is computed. Homotopy is an extension of LARS, performing both forward and backward selection of the atoms already selected.

The generalized LASSO [5] (or analysis [6,7], least mixed norm [8]) extends the basic LASSO (or synthesis) by imposing a regularization matrix $\mathbf{D} \in \mathbb{R}^{m \times p}$ on the coefficient vector \mathbf{x} :

$$\mathbf{x}^*(\lambda) = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{D}\mathbf{x}\|_1 \right\}, \quad (\text{generalized LASSO})$$

where \mathbf{D} typically contains the prior knowledge (e.g., structure information) [9] about \mathbf{x} . For example, if \mathbf{x} is expected to be a piecewise constant signal (i.e., implying that its first order derivative is sparse), then \mathbf{D} is taken to be the first order derivative operator [10]. Some variants of LASSO can be regarded as the generalized LASSO by forming a structured matrix \mathbf{D} . For example, the fused LASSO proposed by Tibshirani et al. [10] imposes the ℓ_1 regularization on both the coefficients and their first order derivatives to encourage the solution to be locally constant. When one of the two regularization parameters is fixed, a fused LASSO problem can be rewritten as a generalized LASSO by cascading a scaled identity matrix with a first order derivative matrix to form the \mathbf{D} matrix. The graph-guided fused LASSO [11,12] incorporates the network prior information into \mathbf{D} for correlation structure. Application of generalized LASSO can be found in image restoration [8], visual recognition [13], electroencephalography (EEG) [14], bioinformatics [15], ultrasonics [16], etc.

Since the LASSO has been proposed, results concerning the recovery condition [3,17], solution path property [18], degree of freedom [19], model selection consistency [20], efficient algorithms as well as software development [2,4] have been widely studied. One may want to know whether these results are applicable to a generalized LASSO problem, e.g., solving a generalized LASSO problem with a LASSO solver. An immediate example is when \mathbf{D} is a full rank square (hence invertible) matrix. By a simple change of variables $\mathbf{u} = \mathbf{D}\mathbf{x}$, the original generalized LASSO problem can be transformed into the basic LASSO form with a predictor matrix $\mathbf{A}\mathbf{D}^{-1}$ and a coefficient vector \mathbf{u} . Therefore it can be solved by calling the LASSO subroutine.

Although the generalized LASSO and LASSO have been studied from various aspects [5,7,21], their connections are not fully explored, which is the main focus of this paper. Elad et al. [6] showed that they are equivalent, but confined the discussion to the denoising case, where \mathbf{A} is an identity matrix. Tibshirani and Taylor [5] showed that the former can be transformed to the latter; however, their method needs to introduce a matrix \mathbf{D}_0 (see Appendices), which brings other potential questions as discussed in the conclusion.

The paper is organized as follows: in Section 2, we show that when the regularization matrix is even- or

under-determined with full rank conditions, the generalized LASSO can be transformed into the LASSO form via the Lagrangian framework. Based on this formula, in Section 3 we show that some published results of LASSO can be extended to the generalized LASSO. In Section 4, two variants of LASSO, namely the regularized deconvolution and the robust LASSO are analyzed under the generalized LASSO framework. We conclude the paper in Section 5.

2. Condition and formula of transformation

The simplification of the generalized LASSO depends on the setting of \mathbf{D} [6]. For the even-determined case ($m=p$),¹ if \mathbf{D} has full rank, by a simple change of variables $\mathbf{u} = \mathbf{D}\mathbf{x}$, the original generalized LASSO problem can be transformed into the basic LASSO with a predictor matrix $\mathbf{A}\mathbf{D}^{-1}$ and a coefficient vector \mathbf{u} . Once the LASSO solution $\hat{\mathbf{u}}(\lambda)$ is known, the original generalized LASSO solution is immediate: $\hat{\mathbf{x}}(\lambda) = \mathbf{D}^{-1}\hat{\mathbf{u}}(\lambda)$.

For the over-determined case ($m > p$), if we change the variable \mathbf{x} to \mathbf{u} , then \mathbf{u} has a higher dimension than \mathbf{x} , or higher degree of freedom. If only a solution $\hat{\mathbf{u}}$ satisfies the constraint $\mathbf{D}\mathbf{D}^\dagger\hat{\mathbf{u}} = \hat{\mathbf{u}}$ where \mathbf{D}^\dagger is the Moore–Penrose pseudoinverse of \mathbf{D} , the original solution is guaranteed to be found [6]. Therefore, the generalized LASSO under this case cannot be transformed into a basic LASSO but a LASSO with equality constraint [22].

We consider the under-determined case ($m < p$). The following Theorem 1 states that a generalized LASSO problem can be transformed into a basic LASSO form under some conditions.

Theorem 1. If matrix $\Phi = \begin{bmatrix} \mathbf{A} \\ \mathbf{D} \end{bmatrix}$ has full column rank (implying $m+n \geq p$), and \mathbf{D} has full row rank (implying $m \leq p$), then the generalized LASSO problem can be transformed into the following LASSO form:

$$\mathbf{u}^*(\lambda) = \arg \min_{\mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{H}\mathbf{u}\|^2 + \lambda \|\mathbf{u}\|_1 \right\}, \quad (1)$$

where

$$\begin{cases} \mathbf{H} = \mathbf{Q}_1\mathbf{Q}_2^\dagger \\ \mathbf{z} = (\mathbf{I}_n - \mathbf{Q}_1(\mathbf{I}_p - \mathbf{Q}_2^\dagger\mathbf{Q}_2)\mathbf{Q}_1^\dagger)\mathbf{y} \end{cases} \quad (2)$$

and $\mathbf{Q}_1 \in \mathbb{R}^{n \times p}$ and $\mathbf{Q}_2 \in \mathbb{R}^{m \times p}$ are defined from the (unique) QR decomposition

$$\Phi = \mathbf{Q}\mathbf{R} = \begin{bmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{bmatrix} \mathbf{R}. \quad (3)$$

Proof. Since Φ has full column rank, its QR decomposition (3) is unique and the square matrix $\mathbf{R} \in \mathbb{R}^{p \times p}$ is invertible. Moreover, $\mathbf{Q}_2 \in \mathbb{R}^{m \times p}$ is full row rank because \mathbf{D} is full row rank. (3) rereads:

$$\mathbf{A} = \mathbf{Q}_1\mathbf{R}. \quad (4)$$

¹ In general, the terms even-, over- and under-determination are used to describe the predictor matrix \mathbf{A} . In this paper, we borrow these terms to characterize the regularization matrix \mathbf{D} .

$$\mathbf{D} = \mathbf{Q}_2 \mathbf{R}. \quad (5)$$

From $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_p$, we have

$$\mathbf{Q}_1^T \mathbf{Q}_1 + \mathbf{Q}_2^T \mathbf{Q}_2 = \mathbf{I}_p. \quad (6)$$

The generalized LASSO problem is equivalent to the following constrained optimization problem:

$$\begin{aligned} (\mathbf{x}^*(\lambda), \mathbf{u}^*(\lambda)) &= \arg \min_{\mathbf{x}, \mathbf{u}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{u}\|_1 \right\} \\ &\text{subject to } \mathbf{D}\mathbf{x} = \mathbf{u}. \end{aligned} \quad (7)$$

The Lagrange function associated with (7) reads

$$L(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 + \lambda \|\mathbf{u}\|_1 + \boldsymbol{\mu}^T (\mathbf{u} - \mathbf{D}\mathbf{x}),$$

where $\boldsymbol{\mu}$ gathers m Lagrange multipliers. The optimality condition reads

$$\frac{\partial L(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu})}{\partial \mathbf{x}} = \mathbf{A}^T (\mathbf{A}\mathbf{x} - \mathbf{y}) - \mathbf{D}^T \boldsymbol{\mu} = \mathbf{0},$$

$$\frac{\partial L(\mathbf{x}, \mathbf{u}, \boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \mathbf{u} - \mathbf{D}\mathbf{x} = \mathbf{0}.$$

From these two equations, we have the following system:

$$\begin{bmatrix} \mathbf{A}^T \mathbf{A} & -\mathbf{D}^T \\ \mathbf{D} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \mathbf{A}^T \mathbf{y} \\ \mathbf{u} \end{bmatrix}. \quad (8)$$

If matrix

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{A}^T \mathbf{A} & -\mathbf{D}^T \\ \mathbf{D} & \mathbf{0} \end{bmatrix} \quad (9)$$

is invertible, we can find the equivalent form of the generalized LASSO problem. By substituting \mathbf{A} and \mathbf{D} with (4) and (5) respectively, we have

$$\mathbf{M} = \begin{bmatrix} \mathbf{R}^T \mathbf{Q}_1^T \mathbf{Q}_1 \mathbf{R} & -\mathbf{R}^T \mathbf{Q}_2^T \\ \mathbf{Q}_2 \mathbf{R} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{R}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{Q}_1^T \mathbf{Q}_1 & -\mathbf{Q}_2^T \\ \mathbf{Q}_2 & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix},$$

where $\mathbf{I}_m \in \mathbb{R}^{m \times m}$ is the identity matrix. Since \mathbf{R} and \mathbf{I}_m are full rank, the invertibility of \mathbf{M} is equivalent to that of

$$\mathbf{N} \triangleq \begin{bmatrix} \mathbf{Q}_1^T \mathbf{Q}_1 & -\mathbf{Q}_2^T \\ \mathbf{Q}_2 & \mathbf{0} \end{bmatrix}.$$

From (6),

$$\mathbf{N} = \begin{bmatrix} \mathbf{I}_p - \mathbf{Q}_2^T \mathbf{Q}_2 & -\mathbf{Q}_2^T \\ \mathbf{Q}_2 & \mathbf{0} \end{bmatrix} = \mathbf{I}_{p+m} + \begin{bmatrix} \mathbf{Q}_2^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} -\mathbf{I}_m & -\mathbf{I}_m \\ \mathbf{I}_m & -\mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{Q}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}.$$

From Woodbury's matrix identity [23, p. 141], \mathbf{N} is invertible if and only if

$$\mathbf{W} \triangleq \begin{bmatrix} -\mathbf{I}_m & -\mathbf{I}_m \\ \mathbf{I}_m & -\mathbf{I}_m \end{bmatrix}^{-1} + \begin{bmatrix} \mathbf{Q}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{Q}_2^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_2 \mathbf{Q}_2^T - \frac{1}{2} \mathbf{I}_m & \frac{1}{2} \mathbf{I}_m \\ -\frac{1}{2} \mathbf{I}_m & \frac{1}{2} \mathbf{I}_m \end{bmatrix}$$

is invertible. Since

$$\det(\mathbf{W}) = \det\left(\frac{1}{2} \mathbf{Q}_2 \mathbf{Q}_2^T\right),$$

and $\det(\mathbf{Q}_2 \mathbf{Q}_2^T) \neq 0$ (\mathbf{Q}_2 is full rank), the invertibility of \mathbf{W} is guaranteed, and therefore \mathbf{M} is invertible.

From the block matrix inversion lemma [23, p. 108], we have

$$\mathbf{W}^{-1} = \begin{bmatrix} (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} & -(\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} \\ (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} & 2\mathbf{I}_m - (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} \end{bmatrix}.$$

Therefore,

$$\mathbf{N}^{-1} = \mathbf{I}_{p+m} - \begin{bmatrix} \mathbf{Q}_2^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \mathbf{W}^{-1} \begin{bmatrix} \mathbf{Q}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix},$$

and

$$\begin{aligned} \mathbf{M}^{-1} &= \begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \mathbf{N}^{-1} \begin{bmatrix} \mathbf{R}^{-T} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}^{-1} (\mathbf{I}_p - \mathbf{Q}_2^T (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} \mathbf{Q}_2) \mathbf{R}^{-T} & \mathbf{R}^{-1} \mathbf{Q}_2^T (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} \\ -(\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} \mathbf{Q}_2 \mathbf{R}^{-T} & (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} - \mathbf{I}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{R}^{-1} (\mathbf{I}_p - \mathbf{Q}_2^\dagger \mathbf{Q}_2) \mathbf{R}^{-T} & \mathbf{R}^{-1} \mathbf{Q}_2^\dagger \\ -(\mathbf{Q}_2^\dagger)^T \mathbf{R}^{-T} & (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} - \mathbf{I}_m \end{bmatrix}, \end{aligned}$$

where $\mathbf{Q}_2^\dagger = \mathbf{Q}_2^T (\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1}$. Finally, from (8) we have

$$\mathbf{x} = \mathbf{R}^{-1} (\mathbf{I}_p - \mathbf{Q}_2^\dagger \mathbf{Q}_2) \mathbf{Q}_1^T \mathbf{y} + \mathbf{R}^{-1} \mathbf{Q}_2^\dagger \mathbf{u}, \quad (10)$$

and

$$\boldsymbol{\mu} = -(\mathbf{Q}_2^\dagger)^T \mathbf{Q}_1^T \mathbf{y} + ((\mathbf{Q}_2 \mathbf{Q}_2^T)^{-1} - \mathbf{I}_m) \mathbf{u}.$$

By substituting (10) into (7), we obtain the results (1) and (2). \square

Once the solution path $\{\hat{\mathbf{u}}(\lambda) | \lambda \in (0, +\infty)\}$ is known, the corresponding solution path $\{\hat{\mathbf{x}}(\lambda) | \lambda \in (0, +\infty)\}$ can be calculated according to (10).

By substituting $\mathbf{Q}_1 = \mathbf{A}\mathbf{R}^{-1}$ and $\mathbf{Q}_2 = \mathbf{D}\mathbf{R}^{-1}$ into (2), and utilizing

$$\boldsymbol{\Psi} \triangleq \boldsymbol{\Phi}^T \boldsymbol{\Phi} = \mathbf{A}^T \mathbf{A} + \mathbf{D}^T \mathbf{D} = \mathbf{R}^T \mathbf{R}, \quad (11)$$

up to a few manipulations, an useful equivalent form of (2) without QR decomposition is obtained as

$$\begin{cases} \mathbf{H} = \mathbf{A}\boldsymbol{\Psi}^{-1} \mathbf{D}^T (\mathbf{D}\boldsymbol{\Psi}^{-1} \mathbf{D}^T)^{-1} \\ \mathbf{z} = (\mathbf{I}_n - \mathbf{A}\boldsymbol{\Psi}^{-1} \mathbf{A}^T + \mathbf{H}\mathbf{D}\boldsymbol{\Psi}^{-1} \mathbf{A}^T) \mathbf{y} \end{cases} \quad (12)$$

And from (10), \mathbf{x} can be recovered as

$$\mathbf{x} = \boldsymbol{\Psi}^{-1} (\mathbf{A}^T - \mathbf{D}^T \mathbf{H}^T) \mathbf{y} + \boldsymbol{\Psi}^{-1} \mathbf{D}^T (\mathbf{D}\boldsymbol{\Psi}^{-1} \mathbf{D}^T)^{-1} \mathbf{u} \quad (13)$$

Note that \mathbf{H} has the same dimensionality as $\mathbf{A}\mathbf{D}^\dagger$, but normally they are not equal unless $\boldsymbol{\Psi}$ is a diagonal matrix.

Remark 1. There are many real-world examples that satisfy this condition. e.g., the total variation denoising problem where \mathbf{A} and \mathbf{D} are an identity and a first order derivative matrix, respectively. By substituting \mathbf{A} and \mathbf{D} into \mathbf{M} in (9), one can check that \mathbf{M} is invertible, and hence satisfies the condition in Theorem 1. Another two examples are the regularized deconvolution problem and the robust LASSO, which will be presented in detail in Section 4.

Remark 2. Under the case $m+n=p$, $\boldsymbol{\Phi}$ is a full rank (square) matrix. Therefore, the row vectors of \mathbf{Q} are orthogonal, and $\mathbf{Q}_1 \mathbf{Q}_2^T = \mathbf{0}$. From (1) and (2), we have $\mathbf{H} = \mathbf{0}$ and the solution is obviously $\mathbf{u} = \mathbf{0}$. According to (10), the original solution is $\mathbf{x} = \mathbf{R}^{-1} \mathbf{Q}_1^T \mathbf{y}$. Note that this solution

does not depend on λ , and the solution path $\{\hat{\mathbf{x}}(\lambda)|\lambda \in (0, +\infty)\}$ is a singleton.

Another viewpoint is obtained by a change of variables. Defining $\mathbf{v} = \mathbf{R}\mathbf{x}$, the generalized LASSO problem rereads

$$\mathbf{v}^*(\lambda) = \arg \min_{\mathbf{v}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{Q}_1 \mathbf{v}\|^2 + \lambda \|\mathbf{Q}_2 \mathbf{v}\|_1 \right\}. \quad (14)$$

Since the rows of \mathbf{Q}_1 and \mathbf{Q}_2 form an orthonormal basis, it is obvious that the minimum ℓ_2 -norm least square solution $\hat{\mathbf{v}} = \mathbf{Q}_1^T \mathbf{y} = \mathbf{Q}_1^T \mathbf{y}$ satisfies $\mathbf{Q}_2 \hat{\mathbf{v}} = \mathbf{0}$. This means that $\hat{\mathbf{v}}$ minimizes both the ℓ_2 - and ℓ_1 -term simultaneously. Therefore, the cost function in (14) is equal to 0 whatever $\lambda > 0$.

Remark 3. Under the case $m+n < p$, $\text{Null}(\Phi)$ is nontrivial, therefore one can check that the generalized LASSO does not yield unique solution, and there are unlimited number of solutions $\{\mathbf{x}^* | \mathbf{x}^* = \Phi^T [\mathbf{y}]_0 + \mathbf{x}_0, \mathbf{x}_0 \in \text{Null}(\Phi)\}$ with $\mathbf{D}\mathbf{x}^* = \mathbf{0}$ and $\mathbf{A}\mathbf{x}^* = \mathbf{y}$ whatever $\lambda > 0$.

Note that this case does not satisfy the condition in Theorem 1 since Φ does not have full column rank. This case serves to complete the discussion.

3. Extension of existing LASSO results

Since the LASSO has been intensively studied during the last decade, many results concerning the computational issue have been published. In this section, we extend some of them into the generalized LASSO problem.

3.1. Monotonic solution path property

Our earlier result [18, Theorem 2] shows that if $(\mathbf{H}^T \mathbf{H})^{-1}$ is diagonally dominant, then the solution path of the corresponding LASSO problem changes monotonically with respect to λ . In other words, the solution path has at most m segments. Therefore, the ‘forward’ algorithm, LARS, yields the same solution path as the ‘forward-backward’ algorithm, homotopy, and both of them can recover the complete solution path within m iterations. Based on this fact, the computational complexity can thus be reduced. The following corollaries extend this result to the generalized LASSO.

Corollary 1. If a generalized LASSO problem satisfying condition in Theorem 1 also satisfies that $\left((\mathbf{D}\Psi^{-1}\mathbf{D}^T)^{-1} - \mathbf{I} \right)^{-1}$ is diagonally dominant, then the complete solution path can be recovered within m iterations of LARS and homotopy algorithm.

Proof. When a generalized LASSO problem satisfies condition in Theorem 1, it can be transformed into a LASSO problem with \mathbf{H} and \mathbf{z} being defined in (2). By utilizing Eqs. (4), (5), (6) and (11), we have

$$\mathbf{H}^T \mathbf{H} = \left(\mathbf{Q}_2 \mathbf{Q}_2^T \right)^{-1} - \mathbf{I} = \left(\mathbf{D}\Psi^{-1}\mathbf{D}^T \right)^{-1} - \mathbf{I},$$

therefore, $(\mathbf{H}^T \mathbf{H})^{-1}$ is diagonally dominant. From [18], this corollary is straightforward. \square

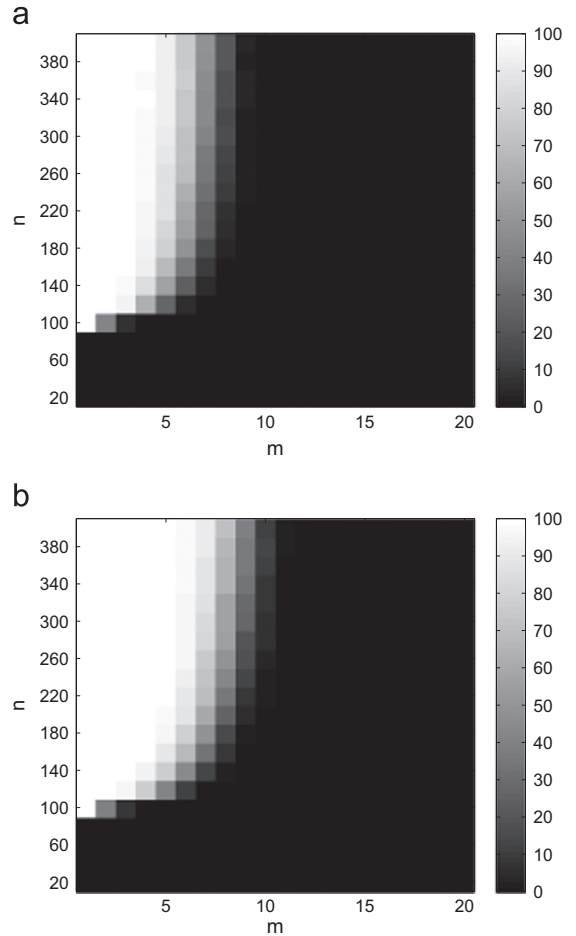


Fig. 1. The frequency (in percentage) of random matrices \mathbf{A} and \mathbf{D} satisfying condition in Corollary 1 with $p=100$. (a) Entries in \mathbf{A} and \mathbf{D} obey normal distribution and Bernoulli distribution with parameter $b=0.1$. (b) The distributions of \mathbf{A} and \mathbf{D} are exchanged.

There exist many matrices satisfying above condition. Obvious examples are the orthogonal dictionaries like Dirac basis or Hadamard basis. By Monte Carlo simulation, we study the probability of random matrices to satisfy the condition. p is fixed to 100, and for each given configuration (n, m) , 1000 trials \mathbf{A} and \mathbf{D} are generated, whose entries obey i.i.d. normal distribution or Bernoulli distribution with parameter $b=0.1$ (the probability for 1 is b , for 0 is $1-b$). The frequency of \mathbf{A} and \mathbf{D} satisfying condition in Corollary 1 is shown in Fig. 1. From the simulation results, it is shown that random matrices satisfy the condition when $n \gg m$.

When $\mathbf{A}^T \mathbf{A}$ forms an orthogonal design, i.e., $\mathbf{A}^T \mathbf{A} = \mathbf{I}$, above condition can be further simplified to the following corollary.

Corollary 2. If a generalized LASSO problem satisfying condition in Theorem 1 also satisfies that \mathbf{A} forms an orthogonal design, and $\mathbf{D}\mathbf{D}^T$ is diagonally dominant, then the complete solution path can be recovered within m iterations of LARS and homotopy algorithm.

Proof. Since \mathbf{D} has full row rank, $\mathbf{D}\mathbf{D}^T$ is invertible. By utilizing Searle's set of matrix identities [24, p. 151], we can verify that

$$\mathbf{D}(\mathbf{I} + \mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T = (\mathbf{I} + (\mathbf{D}\mathbf{D}^T)^{-1})^{-1}.$$

When $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, we have

$$\begin{aligned} (\mathbf{H}^T\mathbf{H})^{-1} &= ((\mathbf{D}\Psi^{-1}\mathbf{D}^T) - \mathbf{I})^{-1} \\ &= (\mathbf{D}(\mathbf{I} + \mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T - \mathbf{I})^{-1} = \mathbf{D}\mathbf{D}^T \end{aligned}$$

Therefore, condition in Corollary 1 can be simplified into checking the diagonal dominance of $\mathbf{D}\mathbf{D}^T$. \square

Corollaries 1 and 2 show that in the best case, the computational complexity of homotopy can reach as low as m iterations to compute the complete solution path on $\lambda \in (0, +\infty)$. It is shown [25] that in the worst case, the computational complexity can reach as high as $(3^m - 1)/2$. Therefore, the computational complexity highly varies depending on the specific problem. Instead of using a generic LASSO solver such as LARS, fast and efficient algorithms can be developed for some specific problems. e.g., the algorithm in [26] permits fast computation and low storage for the regularized deconvolution problem introduced later in Section 4.1 with $\alpha = 0$ and $\beta = -1$. The computational advantage is significant when the dimension is extremely large.

3.2. Decoupling into 1-dimensional subproblems

Efron et al. [3] and Friedman et al. [27] have shown that if the dictionary \mathbf{A} forms an orthogonal design (or uncorrelated design), i.e., $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, a LASSO problem can be decoupled into a set of subproblems, and then solved separately.

$$\left[\begin{array}{cccc|ccc} 1 & \beta - \alpha & -\alpha\beta & 0 & \dots & 0 \\ \beta - \alpha & 1 + (\beta - \alpha)^2 & (\beta - \alpha)(1 - \alpha\beta) & -\alpha\beta & \ddots & \vdots \\ -\alpha\beta & (\beta - \alpha)(1 - \alpha\beta) & 1 + \alpha^2\beta^2 + (\beta - \alpha)^2 & (\beta - \alpha)(1 - \alpha\beta) & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & -\alpha\beta \\ \vdots & \ddots & -\alpha\beta & (\beta - \alpha)(1 - \alpha\beta) & 1 + \alpha^2\beta^2 + (\beta - \alpha)^2 & (\beta - \alpha)(1 - \alpha\beta) \\ 0 & \dots & 0 & -\alpha\beta & (\beta - \alpha)(1 - \alpha\beta) & 1 + \alpha^2\beta^2 + (\beta - \alpha)^2 \end{array} \right] \quad (20)$$

To be more specific, when $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ the LASSO criterion can be rewritten as

$$\frac{1}{2} \|\mathbf{A}^T\mathbf{y} - \mathbf{x}\|^2 + \lambda \|\mathbf{x}\|_1 + \mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y}.$$

This shows that the p -dimensional LASSO problem can be decoupled into p' 1-dimensional subproblems, and each subproblem can be solved separately. The closed-form solution reads

$$\mathbf{x}^* = \eta_\lambda(\mathbf{A}^T\mathbf{y}), \quad (15)$$

where $\eta_\lambda(\cdot)$ is the componentwise soft thresholding function [28]:

$$\eta_\lambda(x) = \begin{cases} x - \lambda, & x > \lambda; \\ 0, & |x| \leq \lambda; \\ x + \lambda, & x < -\lambda. \end{cases}$$

Similarly, for generalized LASSO problem the following corollary provides a sufficient condition to guarantee decoupling.

Corollary 3. If a generalized LASSO problem satisfying condition in Theorem 1 also satisfies

$$\mathbf{D}\Psi^{-1}\mathbf{D}^T = \frac{1}{2}\mathbf{I}, \quad (16)$$

then it can be decoupled into p 1-dimensional LASSO problems, with the following closed-form solution:

$$\mathbf{u}^* = \eta_\lambda(2\mathbf{D}\Psi^{-1}\mathbf{A}^T\mathbf{y}). \quad (17)$$

Proof. Since \mathbf{A} and \mathbf{D} satisfy condition in Theorem 1, this generalized LASSO problem can be transformed into a LASSO problem, with \mathbf{H} , \mathbf{z} being defined in (2), and matrices \mathbf{Q}_1 , \mathbf{Q}_2 and \mathbf{R} being defined in the proof of Theorem 1.

By substituting Eqs. (4), (5) and (11) into (16), an equivalent condition reads

$$\mathbf{Q}_2\mathbf{Q}_2^T = \frac{1}{2}\mathbf{I}. \quad (18)$$

Therefore, $\mathbf{Q}_2^\dagger = 2\mathbf{Q}_2^T$.

From (2) and by utilizing (6), it is easy to prove that the Gramian matrix of \mathbf{H} is an identity matrix, i.e., $\mathbf{H}^T\mathbf{H} = \mathbf{I}$, so \mathbf{H} forms an orthogonal design, indicating the transformed LASSO problem can be decoupled.

Furthermore, with the help of (18), we can also prove

$$\mathbf{H}^T\mathbf{z} = 2\mathbf{Q}_2\mathbf{Q}_1^T\mathbf{y} = 2\mathbf{D}\Psi^{-1}\mathbf{A}^T\mathbf{y}. \quad (19)$$

Finally, by replacing \mathbf{A} and \mathbf{y} in (15) with \mathbf{H} and \mathbf{z} respectively, and substituting (19), the closed-form solution reads in (17). \square

Remark 4. If $\mathbf{A}^T\mathbf{A}$ is invertible, by utilizing Searle's set of matrix identities [24, p. 151], an alternative condition of (16) reads

$$\mathbf{D}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{D}^T = \mathbf{I}$$

Note that this condition is stronger than (16) since for under-determined \mathbf{A} , $\mathbf{A}^T \mathbf{A}$ is not invertible.

One can check that condition in Corollary 3 fulfills when \mathbf{D} is a square matrix with full rank and $\mathbf{A}^T \mathbf{A} = \mathbf{D}^T \mathbf{D}$.

4. Two examples on the analysis of LASSO variants

Several variants of LASSO can be unified under the generalized LASSO framework, such as the total variation regularized deconvolution for signal and image recovery [29], the robust LASSO for face recognition and sensor network [30], the adaptive LASSO for variable selection [31], the fused LASSO for gene expression data analysis [10], the ℓ_1 trend filtering for time series analysis [32], and the adaptive generalized fused LASSO for road-safety data analysis [33]. In this section, the first two will be discussed.

4.1. Regularized deconvolution

Deconvolution is a challenging problem, which can be formulated as a LASSO problem with a Toeplitz predictor matrix. Total variation [34] is a regularization tool to restore piecewise constant signals. This leads to the generalized LASSO form in which \mathbf{A} is a Toeplitz matrix containing the convolution kernel and \mathbf{D} is also a Toeplitz matrix containing the total variation kernel. For high dimensional problems, the computation burden is a major factor. The following theorem sheds some lights on this issue.

Corollary 4. Let \mathbf{A} and \mathbf{D} be square lower triangle Toeplitz matrices with the first column $[1, \alpha, \alpha^2, \alpha^3, \dots, \alpha^{n-1}]^T$ and $[1, \beta, 0, 0, \dots, 0]^T$, respectively. If α and β fall within the region displayed in Fig. 2, then the generalized LASSO problem can be solved for all λ by the LARS and homotopy algorithm in at most n iterations.

The total variation regularization corresponds to the special case when $\beta = -1$.

Proof. According to Theorem 1, since \mathbf{D} is invertible, this generalized LASSO problem is equivalent to the basic LASSO problem (1) with $\mathbf{H} = \mathbf{A}\mathbf{D}^{-1}$. Since \mathbf{A} is squared and invertible, the Gramian matrix inverse reads $(\mathbf{H}^T \mathbf{H})^{-1} = \mathbf{D}\mathbf{A}^{-1}\mathbf{A}^{-T}\mathbf{D}^T$. So in the following we will show that this matrix is diagonally dominant.

One can check that \mathbf{A}^{-1} is also a lower triangle Toeplitz matrix with the first column as $[1, -\alpha, 0, 0, \dots, 0]^T$. Then the Gramian matrix $\mathbf{D}\mathbf{A}^{-1}\mathbf{A}^{-T}\mathbf{D}^T$ reads as (20). In order to have the diagonal dominance property, the following three inequalities are required, which correspond to the first three rows in matrix (20) (The conditions related to the other rows are obviously implied by the third inequality):

$$\begin{aligned} 1 &\geq |\beta - \alpha| + |\alpha\beta|, \\ 1 + (\beta - \alpha)^2 &\geq |\beta - \alpha| + |\beta - \alpha||1 - \alpha\beta| + |\alpha\beta|, \\ 1 + \alpha^2\beta^2 + (\beta - \alpha)^2 &\geq 2|\alpha\beta| + 2|\beta - \alpha||1 - \alpha\beta|. \end{aligned}$$

Those inequalities yield three regions in the 2D plane (α, β) whose intersection is displayed in Fig. 2. □

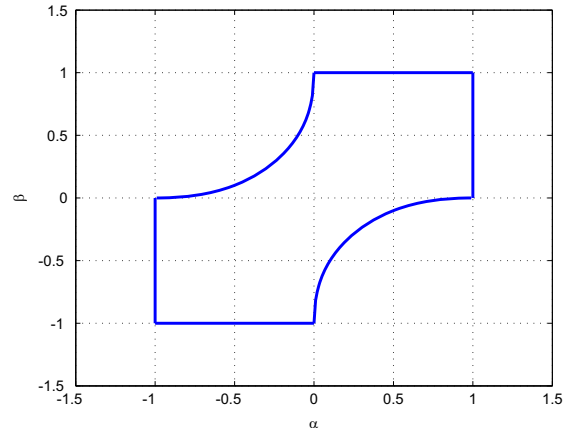


Fig. 2. The feasible region that admits matrix (20) to be diagonal dominant. The region is mirror symmetric with respect to line $\alpha = \beta$, and the curve in the south-east quadrant has expression $\beta = \frac{-(\alpha^2 + 1) + 2\sqrt{\alpha(\alpha^2 - \alpha + 1)}}{(\alpha - 1)^2}$.

Corollary 4 in this paper, Theorem 2 in [18], Theorem 1 and Corollary 2 in [17] jointly show that for deconvolution problem with total variation regularization ($\beta = -1$), if the convolution kernel is high pass ($-1 \leq \alpha < 0$) or all pass ($\alpha = 0$), but not low pass ($0 < \alpha \leq 1$), the true solution can be recovered via the generalized LASSO. If there is no total variation constraint ($\beta = 0$), then any convolution kernel having exponential attenuation $([1, \alpha, \alpha^2, \alpha^3, \dots, \alpha^{n-1}]^T, -1 \leq \alpha \leq 1)$ admits the perfect recovery.

4.2. Robust LASSO

LASSO with the first order total variation method can recover piecewise constant signals. However, if there are outliers in the signal \mathbf{y} , this approach tends to introduce false break points. A second component signal $\mathbf{s} \in \mathbb{R}^n$ is introduced to cope with the outliers in the observation [30]. Since outliers are usually due to burst error, which are assumed to obey the Laplace distribution, a second ℓ_1 regularization term is incorporated into the objective function, yielding the robust generalized LASSO as follows:

$$(\mathbf{x}^*, \mathbf{s}^*) = \arg \min_{\mathbf{x}, \mathbf{s}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x} - \mathbf{s}\|^2 + \lambda_1 \|\mathbf{D}\mathbf{x}\|_1 + \lambda_2 \|\mathbf{s}\|_1 \right\}.$$

This problem can be rewritten into a generalized LASSO form with $\mathbf{A}_{robust} = [\mathbf{A}, \mathbf{I}]$, $\mathbf{D}_{robust} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \tau \mathbf{I} \end{bmatrix}$, and $\mathbf{x}_{robust} = \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$, where $\tau = \lambda_2 / \lambda_1$. When $\Phi = \begin{bmatrix} \mathbf{A} \\ \mathbf{D} \end{bmatrix}$ has full column rank, \mathbf{D} has full column rank, and $\tau > 0$, one can check that $\Phi_{robust} = \begin{bmatrix} \mathbf{A}_{robust} \\ \mathbf{D}_{robust} \end{bmatrix}$ has full column rank, \mathbf{D}_{robust} has full row rank, and \mathbf{M} is invertible. Therefore, from Theorem 1 the robust generalized LASSO problem can be transformed into a basic LASSO form.

5. Conclusion

This paper discusses the simplification of a generalized LASSO problem into the basic LASSO form. When the regularization matrix \mathbf{D} is even- or under-determined

($m \leq p$), we showed that this simplification is possible. Otherwise, there is no guarantee that this simplification can be done. In the former case, optimization tools dedicated to LASSO can be straightforwardly applied to the generalized LASSO.

Tibshirani and Taylor [5] gave a simple way to transform a generalized LASSO to the basic LASSO form when \mathbf{D} is not a square matrix ($m < p$). As shown in Appendix A, they introduced a matrix \mathbf{D}_0 to form a square and invertible matrix $\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} \\ \mathbf{D}_0 \end{bmatrix}$. This poses the question on whether their results depend on \mathbf{D}_0 or not. Appendix B shows that their method yields the equivalent results as ours, and indicates that \mathbf{H} and \mathbf{z} in their formula do not depend on \mathbf{D}_0 . In addition, an improperly introduced \mathbf{D}_0 may incur potential numerical error.

The proposed formula reposes directly on \mathbf{A} and \mathbf{D} , therefore gives an insight into their interaction in the generalized LASSO problem. Based on the proposed formula, it is shown that existing results related to LASSO can be extended to the generalized LASSO. Since some variants of LASSO can be unified under the generalized LASSO framework, they can be transformed into the basic LASSO, and hence efficient LASSO solvers can be applied for the solution. Furthermore, under this framework, many different types of regression formulations such as the trend filtering in a recent study [35] can be unified.

Acknowledgement

This study was partially supported by National Science Foundation of China (No. 61401352), China Postdoctoral Science Foundation (No. 2014M560786), Shaanxi Postdoctoral Science Foundation, Fundamental Research Funds for the Central Universities (No. xjj2014060), National Science Foundation (No. 1539067), and National Institute of Health (Nos. R01MH104680, R01MH107354, and R01GM109068).

Appendix A. Tibshirani and Taylor's transformation method

Tibshirani and Taylor [5] proposed the following method to transform a generalized LASSO when $\text{rank}(\mathbf{D}) = m < p$.

First a matrix $\mathbf{D}_0 \in \mathbb{R}^{(p-m) \times p}$ is found, whose rows are orthogonal to these in \mathbf{D} . Then matrix $\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} \\ \mathbf{D}_0 \end{bmatrix} \in \mathbb{R}^{p \times p}$ is a square matrix with full rank, and thus invertible.

For further use, let us define $\tilde{\mathbf{D}}\mathbf{x} = \begin{bmatrix} \mathbf{D}\mathbf{x} \\ \mathbf{D}_0\mathbf{x} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$ with \mathbf{u} and \mathbf{v} are column vectors of length m and $p-m$ respectively, therefore $\mathbf{x} = \tilde{\mathbf{D}}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix}$, and the generalized LASSO criterion rereads

$$\frac{1}{2} \left\| \mathbf{y} - \tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1} \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \right\|^2 + \lambda \|\mathbf{u}\|_1.$$

Let us also define

$$\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-1} = [\mathbf{A}_1 \ \mathbf{A}_2],$$

where \mathbf{A}_1 and \mathbf{A}_2 are matrices of size $n \times m$ and $n \times (p-m)$

respectively. Above criterion rereads

$$\frac{1}{2} \|\mathbf{y} - \mathbf{A}_1\mathbf{u} - \mathbf{A}_2\mathbf{v}\|^2 + \lambda \|\mathbf{u}\|_1. \quad (\text{A.1})$$

which is a least square problem with respect to \mathbf{v} , and the solution is

$$\mathbf{v}^* = \mathbf{A}_2^\dagger(\mathbf{y} - \mathbf{A}_1\mathbf{u}),$$

$$\text{where } \mathbf{A}_2^\dagger = (\mathbf{A}_2^T\mathbf{A}_2)^{-1}\mathbf{A}_2^T.$$

Finally, by substituting \mathbf{v}^* into (A.1), the generalized LASSO can be transformed into form (1) with

$$\begin{cases} \mathbf{H} = \mathbf{A}_1 - \mathbf{A}_2\mathbf{A}_2^\dagger\mathbf{A}_1 \\ \mathbf{z} = (\mathbf{I}_n - \mathbf{A}_2\mathbf{A}_2^\dagger)\mathbf{y} \end{cases} \quad (\text{A.2})$$

Appendix B. The equivalence between Tibshirani and Taylor's method and the proposed method in this paper

In this appendix, we prove that \mathbf{H} and \mathbf{z} in form (2) and (A.2) are equal.

B.1. Equivalence of \mathbf{H}

First, for further use let us define $\mathbf{Q}_0 \triangleq \mathbf{D}_0\mathbf{R}^{-1}$ and

$$\begin{bmatrix} \mathbf{Q}_2 \\ \mathbf{Q}_0 \end{bmatrix}^{-1} \triangleq \begin{bmatrix} \mathbf{P}_2 & \mathbf{P}_0 \end{bmatrix}, \quad (\text{B.1})$$

where \mathbf{P}_2 and \mathbf{P}_0 are of size $p \times m$ and $p \times (p-m)$, respectively.

It is easy to verify that

$$\mathbf{P}_0^T\mathbf{Q}_2^T = \mathbf{0}, \quad (\text{B.2})$$

$$\mathbf{P}_2\mathbf{Q}_2 + \mathbf{P}_0\mathbf{Q}_0 = \mathbf{I}, \quad (\text{B.3})$$

$$\begin{cases} \mathbf{Q}_1\mathbf{P}_2 = \mathbf{A}_1, \\ \mathbf{Q}_1\mathbf{P}_0 = \mathbf{A}_2. \end{cases} \quad (\text{B.4})$$

From (B.4), (6) and (B.2), we have

$$\mathbf{A}_2^T\mathbf{Q}_1\mathbf{Q}_2^T = \mathbf{P}_0^T\mathbf{Q}_1^T\mathbf{Q}_1\mathbf{Q}_2^T = \mathbf{P}_0^T(\mathbf{I} - \mathbf{Q}_2^T\mathbf{Q}_2)\mathbf{Q}_2^T = \mathbf{0}. \quad (\text{B.5})$$

Left multiply both sides of (B.3) with \mathbf{Q}_1 , and substitute resultant equation into (B.5), we have

$$\mathbf{A}_2^T(\mathbf{A}_1\mathbf{Q}_2 + \mathbf{A}_2\mathbf{Q}_0)\mathbf{Q}_2^T = \mathbf{0}.$$

Move the second part to the right side of equals sign, and left multiply both sides with $(\mathbf{A}_2^T\mathbf{A}_2)^{-1}$, right multiply with $(\mathbf{Q}_2\mathbf{Q}_2^T)^{-1}$, we have

$$\mathbf{A}_2^\dagger\mathbf{A}_1 = -\mathbf{Q}_0\mathbf{Q}_2^\dagger.$$

Left multiply both sides of above equation with $-\mathbf{A}_2$, and add with \mathbf{A}_1 , we have

$$\mathbf{A}_1 - \mathbf{A}_2\mathbf{A}_2^\dagger\mathbf{A}_1 = \mathbf{A}_1 + \mathbf{A}_2\mathbf{Q}_0\mathbf{Q}_2^\dagger.$$

The left side of above equation is \mathbf{H} in (A.2), and the right side is equal to \mathbf{H} in (2), since left multiply both sides of (B.3) with \mathbf{Q}_1 , and right multiply with \mathbf{Q}_2^\dagger yields

$$\mathbf{A}_1 + \mathbf{A}_2\mathbf{Q}_0\mathbf{Q}_2^\dagger = \mathbf{Q}_1\mathbf{Q}_2^\dagger.$$

B.2. Equivalence of \mathbf{z}

To prove that \mathbf{z} in (2) and (A.2) are equal, we have to prove

$$\mathbf{Q}_1 \mathbf{Q}_1^T - \mathbf{Q}_1 \mathbf{Q}_2 \mathbf{Q}_2^T \mathbf{Q}_1^T = \mathbf{A}_2 \mathbf{A}_2^T.$$

From (B.1), one can verify that

$$\mathbf{I} - \mathbf{Q}_2^T \mathbf{Q}_2 = \mathbf{P}_0 \mathbf{P}_0^T.$$

Left multiply both sides of above equation with \mathbf{Q}_1 , and right multiply with \mathbf{Q}_1^T , we have

$$\begin{aligned} \mathbf{Q}_1 \mathbf{Q}_1^T - \mathbf{Q}_1 \mathbf{Q}_2 \mathbf{Q}_2^T \mathbf{Q}_1^T &= \mathbf{Q}_1 \mathbf{P}_0 \mathbf{P}_0^T \mathbf{Q}_1^T \\ &= \mathbf{A}_2 (\mathbf{P}_0^T \mathbf{P}_0)^{-1} \mathbf{A}_2^T \quad \text{using Eq. (B.4)} \\ &= \mathbf{A}_2 (\mathbf{P}_0^T \mathbf{P}_0 - \mathbf{P}_0^T \mathbf{Q}_2^T \mathbf{Q}_2 \mathbf{P}_0)^{-1} \mathbf{A}_2^T \quad \text{using Eq. (B.2)} \\ &= \mathbf{A}_2 (\mathbf{P}_0^T (\mathbf{I} - \mathbf{Q}_2^T \mathbf{Q}_2) \mathbf{P}_0)^{-1} \mathbf{A}_2^T \\ &= \mathbf{A}_2 (\mathbf{P}_0^T (\mathbf{Q}_1^T \mathbf{Q}_1) \mathbf{P}_0)^{-1} \mathbf{A}_2^T \quad \text{using Eq. (6)} \\ &= \mathbf{A}_2 \mathbf{A}_2^T \quad \text{using Eq. (B.4)} \end{aligned}$$

References

- [1] R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B* 58 (1) (1996) 267–288.
- [2] M.R. Osborne, B. Presnell, B.A. Turlach, A new approach to variable selection in least squares problems, *IMA J. Numer. Anal.* 20 (3) (2000) 389–403.
- [3] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, *Ann. Stat.* 32 (2) (2004) 407–499.
- [4] D.M. Malioutov, M. Cetin, A.S. Willsky, Homotopy continuation for sparse signal representation, in: *Proceedings of IEEE ICASSP*, Philadelphia, PA, March 2005, vol. 5, pp. 733–736.
- [5] R.J. Tibshirani, J. Taylor, The solution path of the generalized lasso, *Ann. Stat.* 39 (3) (2011) 1335–1371.
- [6] M. Elad, P. Milanfar, R. Rubinfeld, Analysis versus synthesis in signal priors, *Inverse Probl.* 23 (3) (2007) 947.
- [7] S. Vaiteer, G. Peyré, C. Dossal, J. Fadili, Robust sparse analysis regularization, *IEEE Trans. Inf. Theory* 59 (4) (2013) 2001–2016.
- [8] H. Fu, M.K. Ng, M. Nikolova, J.L. Barlow, Efficient minimization methods of mixed ℓ_2 - ℓ_1 and ℓ_1 - ℓ_1 norms for image restoration, *SIAM J. Sci. Comput.* 27 (6) (2006) 1881–1902.
- [9] Y.C. Eldar, G. Kutyniok (Eds.), *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, UK, May 2012.
- [10] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, *J. R. Stat. Soc. Ser. B* 67 (1) (2005) 91–108.
- [11] S. Kim, K.-A. Sohn, E.P. Xing, A multivariate regression approach to association analysis of a quantitative trait network, *Bioinformatics* 25 (2009) i204–i212. *ISMB* 2009.
- [12] X. Chen, Q. Lin, S. Kim, J.G. Carbonell, E.P. Xing, Smoothing proximal gradient method for general structured sparse regression, *Ann. Appl. Stat.* 6 (2) (2012) 719–752.
- [13] N. Morioka, S. Satoh, Generalized lasso based approximation of sparse coding for visual recognition, in: *Advances in Neural Information Processing Systems*, 2011, pp. 181–189.
- [14] M. Vega-Hernández, E. Martínez-Montes, J.M. Sánchez-Bornot, A. Lage-Castellanos, P.A. Valdés-Sosa, Penalized least squares methods for solving the EEG inverse problem, *Stat. Sin.* 18 (2008) 1535–1551.
- [15] R. Tibshirani, P. Wang, Spatial smoothing and hot spot detection for CGH data using the fused lasso, *Biostatistics* 9 (1) (2008) 18–29.
- [16] C. Yu, C. Zhang, L. Xie, A blind deconvolution approach to ultrasound imaging, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 59 (2) (2012) 271–280.
- [17] D.L. Donoho, Y. Tsaig, Fast solution of l_1 -norm minimization problems when the solution may be sparse, *IEEE Trans. Inf. Theory* 54 (November (11)) (2008) 4789–4812.
- [18] J. Duan, C. Soussen, D. Brie, J. Idier, Y.-P. Wang, On LARS/homotopy equivalence conditions for over-determined LASSO, *IEEE Signal Process. Lett.* 19 (12) (2012) 894–897.
- [19] H. Zou, T. Hastie, R. Tibshirani, On the “degrees of freedom” of the lasso, *Ann. Stat.* 35 (5) (2007) 2173–2192.
- [20] P. Zhao, B. Yu, On model selection consistency of lasso, *J. Mach. Learn. Res.* 7 (2006) 2541–2563.
- [21] Y. She, Sparse regression with exact clustering, *Electron. J. Stat.* 4 (2010) 1055–1096.
- [22] G.M. James, C. Paulson, P. Rusmevichientong, *The constrained lasso*, Technical Report, University of Southern California, 2013, (<http://www-bcf.usc.edu/~garth/research/ClassoFinal.pdf>).
- [23] D. Bernstein, *Matrix Mathematics*, Princeton University Press, Princeton, USA, 2009.
- [24] S.R. Searle (Ed.), *Matrix Algebra Useful for Statistics*, John Wiley & Sons, Inc., New York, USA, 1982.
- [25] J. Mairal, B. Yu, Complexity analysis of the lasso regularization path, in: *International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012, pp. 353–360.
- [26] J. Duan, J.-G. Zhang, J. Lefante, H.-W. Deng, Y.-P. Wang, Detection of copy number variation from next generation sequencing data with total variation penalized least square optimization, in: *IEEE International Conference on Bioinformatics and Biomedicine Workshops*, Atlanta, GA, USA, 2011, pp. 3–12.
- [27] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani, Pathwise coordinate optimization, *Ann. Appl. Stat.* 1 (2) (2007) 302–332.
- [28] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, *J. Am. Stat. Assoc.* 90 (432) (1995) 1200–1224.
- [29] A. Chambolle, P.-L. Lions, Image recovery via total variation minimization and related problems, *Numer. Math.* 76 (1997) 167–188.
- [30] N.H. Nguyen, T.D. Tran, Robust lasso with missing and grossly corrupted observations, *IEEE Trans. Inf. Theory* 59 (4) (2013) 2036–2058.
- [31] H. Zou, The adaptive lasso and its oracle properties, *J. Am. Stat. Assoc.* 101 (476) (2006) 1418–1429.
- [32] S.-J. Kim, K. Koh, S. Boyd, D. Gorinevsky, ℓ_1 trend filtering, *SIAM Rev.* 51 (2) (2009) 339–360.
- [33] V. Viallon, S. Lambert-Lacroix, H. Hoëfing, F. Picard, Adaptive generalized fused-lasso: asymptotic properties and applications, Technical Report, Université de Lyon, 2013, (<http://www.statistics.gov.hk/wsc/IPS012-P1-S.pdf>).
- [34] L. Rudin, S. Osher, C. Fatemi, Nonlinear total variation based noise removal algorithm, *Phys. D* 60 (1992) 259–268.
- [35] R.J. Tibshirani, Adaptive piecewise polynomial estimation via trend filtering, *Ann. Stat.* 42 (1) (2014) 285–323.