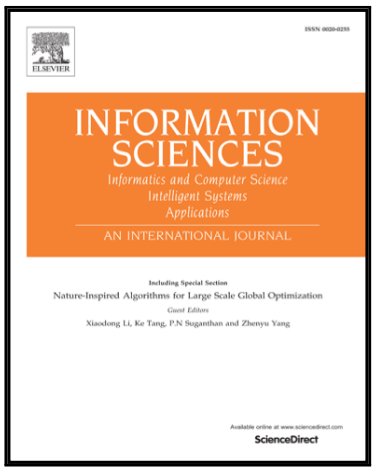

Sparse Regression with Output Correlation for Cardiac Ejection Fraction Estimation

Bin Gu, Yingying Shan, Victor S. Sheng, Shuo Li

PII: S0020-0255(17)30048-8
DOI: 10.1016/j.ins.2017.09.026
Reference: INS 13130

To appear in: *Information Sciences*

Received date: 4 January 2017
Revised date: 5 July 2017
Accepted date: 9 September 2017

Please cite this article as: Bin Gu, Yingying Shan, Victor S. Sheng, Shuo Li, Sparse Regression with Output Correlation for Cardiac Ejection Fraction Estimation, *Information Sciences* (2017), doi: 10.1016/j.ins.2017.09.026

Sparse Regression with Output Correlation for Cardiac Ejection Fraction Estimation

Bin Gu*[1,2,3,4], Yingying Shan[3], Victor S.Sheng[5], and Shuo Li[4,6]

[1]Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science & Technology, Nanjing, P.R.China
[2]Jiangsu-Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, P.R.China
[3]School of Computer & Software, Nanjing University of Information Science & Technology, Nanjing, P.R.China
[4]Department of Medical Biophysics, University of Western Ontario, London, Ontario, Canada
[5]Department of Computer Science, University of Central Arkansas, Conway, Arkansas, USA
[6]GE Health Care, London, Ontario, Canada

Traditional regression methods minimize the sum of errors of samples with various regularization terms such as the ℓ_1-norm and ℓ_2-norm. For the diagnosis of cardiovascular diseases, the cardiac ejection fraction (EF) represents an essential measure. However, existing regularization terms do not consider the output correlation (the correlation between ground truth volumes and estimated volumes w.r.t. each subject), which is beneficial in estimating the cardiac EF. In this paper, we first propose a sparse regression with two regularization terms of the ℓ_1-norm and output correlation (SROC). Then, we propose a one-dimensional solution path algorithm for quickly finding two good regulation parameters in the formulation of SROC. The solution path algorithm can effectively handle singularities and infinities in the key matrix. Finally, we conduct experiments on a clinical cardiac image dataset with 100 subjects. The experimental results show that our method produces competitive and strong results for estimating the cardiac EF based on quantitative and qualitative analyses.

1. Introduction

Traditional regression methods minimize the sum of errors of samples with certain regularization terms such as the ℓ_1-norm and the ℓ_2-norm. Different regularization terms generate solutions having different properties. For example, Lasso [23] uses the ℓ_1-norm to force solutions to be sparse. Any features that have non-zero regression coefficients are selected. Ridge regression [21] uses the ℓ_2-norm to shrink the coefficients of the correlated predictors toward each other. Elastic net [34] uses both the ℓ_1-norm and the ℓ_2-norm to achieve the properties of Lasso and ridge regression simultaneously.

Cardiovascular diseases are the most frequent cause of death in most countries [16]. In the diagnosis of these pathologies, the cardiac ejection fraction (EF) is an essential measure. The EF represents the volumetric fraction of blood pumped out of the left ventricle (LV) and right ventricle (RV) in a cardiac cycle [28]. The EF can be computed as $\frac{V^d - V^s}{V^d} \times 100\%$, where V^d and V^s denote the largest and smallest volumes of a ventricle, respectively, in a cardiac cycle. A normal EF is in the range 55-70%. The EF may be lower when a heart muscle is damaged due to a heart attack, a heart muscle disease, or other causes [28]. An EF of less than 40% may confirm a diagnosis of heart failure. An EF of less than 35% increases the risk of life-threatening irregular heartbeats that can cause a sudden cardiac arrest

*Corresponding author.
Bin Gu: *jsgubin@nuist.edu.cn*
Yingying Shan : *yyshan@nuist.edu.cn*
Victor S. Sheng: *sheng@uca.edu*
Shuo Li: *Shuo.Li@ge.com*

and a sudden cardiac death [28]. Effectively estimating the EF is highly significant in clinical diagnoses.

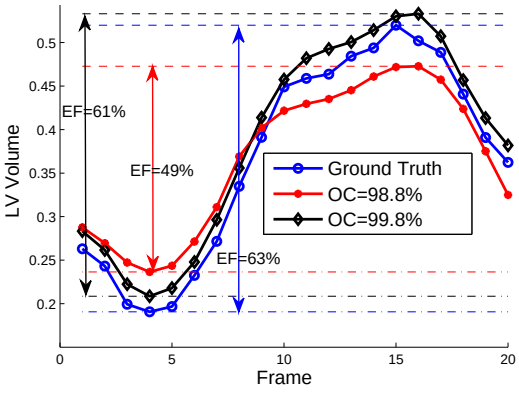

Figure 1. Regression analysis with the output correlation used to predict EF. The estimated volumes with OC=99.8% achieve a better estimation of the EF than the estimated volumes with OC=98.8%.

A clinical subject is scanned by a scanner (including MRI, CT, and ultrasound), which will produce n_k (normally 20, as shown in Fig. 1) frames throughout the cardiac cycle. Let $S_k = \{(x_i, y_i)\}_{i=1}^{n_k}$ denote these n_k frames, where $x_i \in \mathbb{R}^d$ is a cardiac image, which is the sum of cardiac images at different slices, and $y_i \in \mathbb{R}$ is the ground truth LV/RV volume, which is computed by adding the LV/RV areas at different slices [24]. Assume that a set of samples is $S = \bigcup_{k=1}^{\kappa} S_k$, where κ denotes the number of clinical subjects. We assume that the set S is generated by a linear regression model $y_i = x_i^T\beta + \varepsilon$, where β is the vector of model coefficients and ε is an i.i.d. random variable with mean 0 and variance σ^2. For clinical subject data S_k, we define the output correlation (OC) as follows.

Definition 1 *The output correlation for clinical subject data S_k is defined as the correlation between the ground truth volumes and the estimated volumes. Specifically, the OC coefficient ρ_k is computed as* $\rho_k = \frac{\sum_{i=1}^{|S_k|}\left(y_i - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} y_j\right)\left(x_i^T\beta - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} x_j^T\beta\right)}{\sqrt{\sum_{i=1}^{|S_k|}\left(y_i - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} y_j\right)^2 \sum_{i=1}^{|S_k|}\left(x_i^T\beta - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} x_j^T\beta\right)^2}}$.

OC is beneficial for estimating the cardiac EF. As shown in Fig. 1, the blue curve presents the ground truth LV volumes of a modality sequence with EF=63%. The red curve presents the estimated results by a sparse regression [23], which minimizes the sum of errors of the LV cavity areas with the ℓ_1-norm. From the figure, a poor EF value of 49% is produced. This causes a misdiagnosis in that a normal patient is suspected of having cardiovascular disease, even though OC=98.8%. A higher OC reflects a stronger dependency between the ground truth volumes and the estimated volumes. The EF is a special measure that only depends on both the largest and smallest volumes in a cardiac cycle. Thus, a higher OC would produce a higher estimation of the EF. In practice, the EF can be improved to 61%, which is very close to the ground truth EF of 63% if the OC is increased to 99.8%.

However, as previously mentioned, the regularization terms in traditional regression methods do not consider the OC. Recently, several works [1, 24, 31] have proposed to estimate the EF based on the LV and RV cavity areas. For example, Afshin et al. [1] used artificial neural networks to estimate the LV cavity area. Wang et al. [24] used a Bayesian formulation to estimate the LV and RV cavity areas, respectively. Zhen et al. [31] used a built-in feature selection regression method based on random forests to estimate the LV and RV cavity areas. It should be noted that the volume V is approximated by integrating the cavity areas as $V = \sum_i A_i$ [24]. Summarizing, all the above methods are dependent upon traditional regression methods, and they do not consider the OC.

As indicated in [31], feature selection plays an import role in estimating bi-cavity areas. There are three main categories of feature selection algorithms: wrappers, filters and embedded methods [13]. Embedded methods are a catch-all group of techniques that perform feature selection as part of the model construction process. The exemplar of embedded methods for regression is

Lasso regression [23], which is a classical sparse regression method. In addition to Lasso, there are several other state-of-the-art sparse methods. For example, Wright et al. [25] and Shao et al. [22] proposed sparse-representation-based classification methods. Zhu et al. [32] proposed a feature selection method by combining the graph regularizer and the $\ell_{2,1}$-norm regularizer. Zhu et al. [32] proposed a block-row sparse multiview multilabel learning framework for image classification. Liu et al. [15] proposed a sparse logistic regression method for multi-label learning. An et al. [2] proposed a sparse representation matching method based on the ℓ_1 and ℓ_2 norms. For a comprehensive survey on sparse methods, please refer to [29]. Because predicting the EF is a regression problem, this paper focuses on the classical sparse regression method, Lasso [7].

In this paper, we propose a new sparse regression with OC (SROC). We believe that this is important for machine learning and medical applications. For machine learning, SROC incorporates OC as a regularization term, which is lacking in traditional regression methods. For medical applications, SROC is better able to estimate the cardiac EF compared to traditional regression methods. More specifically, we first propose our SROC based on the ℓ_1-norm and the OC. Then, we propose a one-dimensional solution path algorithm for quickly finding two good regulation parameters in the formulation of SROC. The solution path algorithm can effectively handle the singularity and infinity in the key matrix. Finally, we conduct experiments on a clinical cardiac image dataset with 100 subjects. The experimental results show that our method produces competitive and strong results for estimating cardiac EF based on quantitative and qualitative analyses.

It should be noted that robust regression methods [26, 27, 19] have recently been proposed. For example, Yang et al. [27] proposed a regularized robust coding method that could perform robust regression with a regularized regression coefficient. Yang et al. [26] and Qian et al. [19] proposed nuclear-norm-based matrix regression methods that can be directly applied to images and that do not need the matrix-to-vector conversion step. Zhang et al. [30] proposed a non-convex sparse matrix regression method. In our experiments, we will compare the matrix regression method [26] with our SROC method.

The remainder of the paper is organized as follows. In Section 2, we propose our SROC. In Section 3, we propose the solution path algorithm corresponding to SROC. In Section 4, we present the experimental results on a cardiac MRI dataset. Finally, in Section 5, we give some concluding remarks.

Notation: To make the notations easier to follow, we give a summary of the notations in the following list.

β_i, g_i	The i-th element of the vector β and g.
$x_{i,j}$, $\widehat{x}_{k,i,j}$	The j-th dimension of the vector x_i and $\widehat{x}_{k,i}$.
$\lvert S_k\rvert$	The cardinality of a set S_k.
Δ	The amount of change in each variable.
$Q_{j_1 j_2}$	$\sum_{i=1}^{n} x_{i,j_1} x_{i,j_2}$.
$Q_{S_S S_S}$	The sub-matrix of Q, with the rows and columns indexed by S_S.
$\overline{\mathcal{E}^+}$, $\overline{\mathcal{E}^-}$	$\{1,\cdots,d\}-\mathcal{E}^+$, and $\{1,\cdots,d\}-\mathcal{E}^-$.

Abbreviations: To make the paper easier to follow, we also give a summary of the abbreviations in the following list.

EF	Ejection fraction.
LV	Left ventricle.
RV	Right ventricle.
OC	Output correlation.
SROC	Sparse regression with output correlation.

2. Sparse Regression with Output Correlation

As previously mentioned, the OC is beneficial for estimating the cardiac EF, and sparse regression is the focus of this paper. Thus, we

4Bin Gu[41,2,3,4], Yingying Shan[3], Victor S.Sheng[5], and Shuo Li[4,61]Jiangsu Engineering Center of Network Monitoring, Nanji

first briefly review a classical sparse regression method, i.e., Lasso, and then propose our SROC based on Lasso and the OC. Finally, we give the KKT conditions to the SROC.

2.1. Brief Review of Lasso

Let the set $S = \bigcup_{k=1}^{\kappa} S_k = \{(x_i, y_i)\}_{i=1}^{n}$, which is generated by a linear regression model $y_i = x_i^T\beta + \varepsilon$. ε is an i.i.d. random variable with mean 0 and variance σ^2. The model is assumed to be "sparse". Specifically, some of the regression coefficients β are exactly zero. The formulation of Lasso is presented as follows.

$$\min_{\beta} \sum_{i=1}^{n} \left(y_i - x_i^T\beta\right)^2 + \lambda \parallel \beta \parallel_1 \tag{1}$$

where $\parallel \cdot \parallel_1$ denotes the ℓ_1-norm of a vector, which induces sparsity in the solution, and $\lambda \geq 0$ is a regularization parameter controlling the amount of regularization applied to the estimate.

If $\lambda = 0$, Lasso degenerates to the minimization of the unregularized empirical risk. On the other hand, a very large λ will decrease β to zero, thus leading to an empty or null model. In general, moderate values of λ will cause some coefficients to be exactly zero, and the other features that have non-zero coefficients are 'selected' by Lasso.

As mentioned above, selecting a value of λ plays an essential role in Lasso. A global searching of λ can be efficiently implemented by a solution path algorithm [20, 12] because one solution in a solution path can act on an interval of the regularization parameter in which the solutions share the same linearity property. Thus, the solution path algorithm can effectively fit all the solutions based on a finite number of representative solutions.

2.2. Formulation of SROC

As discussed in Section 1, the OC is beneficial for estimating the cardiac EF. In this section, we will incorporate the OC into the standard formulation of Lasso to better predict the value of the cardiac EF.

To efficiently solve the new Lasso formulation, which incorporates the OC coefficient ρ_k, we use the standard deviation of the ground truth $\sigma_k = \sqrt{\sum_{i=1}^{|S_k|}\left(y_i - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} y_j\right)^2}$ to approximate the standard deviation of the estimated values $\sqrt{\sum_{i=1}^{|S_k|}\left(x_i^T\beta - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} x_j^T\beta\right)^2}$ in ρ_k. Thus, a new Lasso-type learning formulation SROC is proposed as follows.

$$\begin{aligned}\min_{\beta} \quad & \sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 + \lambda_1 \parallel \beta \parallel_1 \\ & -\lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \frac{\widehat{y}_{k,i}}{\sigma_k}\left(x_i^T\beta - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} x_j^T\beta\right)\end{aligned} \tag{2}$$

where $\widehat{y}_{k,i} = \frac{y_i - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} y_j}{\sigma_k}$. The formulation (2) can be further rewritten as

$$\min_{\beta} \sum_{i=1}^{n}\left(y_i - x_i^T\beta\right)^2 + \lambda_1 \parallel \beta \parallel_1 - \lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i}^T\beta \tag{3}$$

where $\widehat{x}_{k,i} = \frac{x_i^T - \frac{1}{|S_k|}\sum_{j=1}^{|S_k|} x_j^T}{\sigma_k}$. Obviously, the formulation (3) is not differentiable if a variable is zero because of the ℓ_1-norm. To design the solution path algorithm for SROC, (3) can be replaced by a differentiable function (4), similar to [20], where $\beta = \beta^+ - \beta^-$.

$$\begin{aligned}\min_{\beta^+,\beta^-} \quad & \sum_{i=1}^{n}\left(y_i - x_i^T(\beta^+ - \beta^-)\right)^2 + \lambda_1\sum_{i=1}^{d}(\beta_i^+ + \beta_i^-) \\ & -\lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i}^T(\beta^+ - \beta^-) \\ s.t. \quad & \beta^+ \geq 0, \quad \beta^- \geq 0\end{aligned} \tag{4}$$

Obviously, (4) is a box-constrained quadratic programming problem, where the upper bounds for β^+ and β^- are ∞. Similar to Lasso, selecting the values of λ_1 and λ_2 plays an essential role for the function (4).

This paper will present a one-dimensional solution path algorithm for (4) to fit all the solutions based on a finite number of representative solutions when $\Delta\lambda_2 = \delta\Delta\lambda_1$.[3] Then, we can use cross validation with a hybrid method for one-parametric solution path searching on λ_1 and grid

[3] δ is a constant used to denote the coefficient of variation between λ_2 and λ_1.

searching on δ to quickly find good values for λ_1 and λ_2.

2.3. KKT Conditions for SROC

As shown in [12, 20], each solution in a solution path is the optimal solution w.r.t. the corresponding value of the regularization parameter. This subsection presents the necessary and sufficient conditions (i.e., the KKT conditions) for the optimal solution of (4).

From the KKT theorem [3], we obtain the following equivalent conditions (5)-(6) for the optimal solution of (4):

$$g_j^+ \stackrel{\text{def}}{=} \sum_{i=1}^{n} -x_{i,j}\left(y_i - x_i^T(\beta^+ - \beta^-)\right) + \lambda_1 \quad (5)$$

$$-\lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i,j} \begin{cases} g_j^+ \geq 0 & \text{for} \quad \beta_j^+ = 0 \\ g_j^+ = 0 & \text{for} \quad \beta_j^+ > 0 \end{cases}$$

$$g_j^- \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_{i,j}\left(y_i - x_i^T(\beta^+ - \beta^-)\right) + \lambda_1 \quad (6)$$

$$+\lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i,j} \begin{cases} g_j^- \geq 0 & \text{for} \quad \beta_j^- = 0 \\ g_j^- = 0 & \text{for} \quad \beta_j^- > 0 \end{cases}$$

According to the value of g_j^+, a set $\{1, \cdots, d\}$ is partitioned as $\pi^+ = \left(\mathcal{E}^+, \overline{\mathcal{E}^+}\right)$, where

$$\mathcal{E}^+ = \{j \in \{1, \cdots, d\} \mid g_j^+ = 0, \quad \beta_j^+ > 0\} \quad (7)$$

$$\overline{\mathcal{E}^+} = \{j \in \{1, \cdots, d\} \mid g_j^+ \geq 0, \quad \beta_j^+ = 0\} \quad (8)$$

Accordingly, the set $\{1, \cdots, d\}$ possesses the partition $\pi^- = \left(\mathcal{E}^-, \overline{\mathcal{E}^-}\right)$ according to the value of g_j^-.

3. The Solution Path

Based on the KKT conditions (5)-(6), we propose a one-dimensional solution path algorithm (i.e., Algorithm 1) for the box-constrained quadratic programming problem (4), which includes two main parts:

1. establishing the initial solution for (4) when $\lambda_1 = 0$ and $\lambda_2 = 0$, which is described in Section 3.1,ad
2. exploring the solutions of (4) with $\Delta\lambda_2 = \delta\Delta\lambda_1$ for all values of $\lambda_1 \geq 0$, which is described in Section 3.2.

3.1. Initialization

When $\lambda_1 = 0$ and $\lambda_2 = 0$, it is easy to find that (3) degenerates to the minimization of the sum of squares errors, which corresponds to solving the linear system of equations $\sum_{i=1}^{n} x_i\left(y_i - x_i^T\beta\right) = \mathbf{0}$. The solution of the linear system is $\beta = \left(\sum_{i=1}^{n} x_i x_i^T\right)^{-1} \cdot \left(\sum_{i=1}^{n} x_i y_i\right)$.

Based on β, the solution of (4) with $\lambda_1 = 0$ and $\lambda_2 = 0$ can be obtained as follows: if $\beta_j \geq 0$, $\beta_j^+ = \beta_j$, $\beta_j^- = 0$; otherwise, $\beta_j^+ = 0$, $\beta_j^- = -\beta_j$.

3.2. Solution Path

When increasing λ_1, and λ_2, the weights β_j^+ indexed by $\mathcal{E}^+$ and the weights β_j^- indexed by $\mathcal{E}^-$ should also be adjusted accordingly. Thus, according to the KKT conditions (5)-(6), we have the following linear system:

$$\Delta g_j^+ \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_{i,j}\left(x_{i,\mathcal{E}^+}^T \Delta\beta_{\mathcal{E}^+}^+ - x_{i,\mathcal{E}^-}^T \Delta\beta_{\mathcal{E}^-}^-\right) + \Delta\lambda_1 - \Delta\lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i,j} = 0, \quad \forall j \in \mathcal{E}^+ \quad (9)$$

$$\Delta g_j^- \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_{i,j}\left(-x_{i,\mathcal{E}^+}^T \Delta\beta_{\mathcal{E}^+}^+ + x_{i,\mathcal{E}^-}^T \Delta\beta_{\mathcal{E}^-}^-\right) + \Delta\lambda_1 + \Delta\lambda_2 \sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i,j} = 0, \quad \forall j \in \mathcal{E}^- \quad (10)$$

If $\Delta\lambda_2 = \delta\Delta\lambda_1$, the linear system (9)-(10) can be written as

$$\underbrace{\begin{bmatrix} Q_{\mathcal{E}^+\mathcal{E}^+} & -Q_{\mathcal{E}^+\mathcal{E}^-} \\ -Q_{\mathcal{E}^-\mathcal{E}^+} & Q_{\mathcal{E}^-\mathcal{E}^-} \end{bmatrix}}_{Q} \begin{bmatrix} \xi_{\mathcal{E}^+}^+ \\ \xi_{\mathcal{E}^-}^- \end{bmatrix} \quad (11)$$

$$= \begin{bmatrix} -\mathbf{1} + \delta\sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i,\mathcal{E}^+} \\ -\mathbf{1} - \delta\sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|} \widehat{y}_{k,i}\widehat{x}_{k,i,\mathcal{E}^-} \end{bmatrix}$$

where $\xi_{\mathcal{E}^+}^+$ ($\xi_{\mathcal{E}^-}^-$) denotes the linear relationship between $\Delta\beta_{\mathcal{E}^+}^+$ ($\Delta\beta_{\mathcal{E}^-}^-$) and $\Delta\lambda_1$.

There are two issues in directly solving (11). As mentioned in [6, 17], solution path algorithms frequently encounter a singularity in the key matrix because the data set is characterized by the presence of linearly dependent points (in the kernel space), duplicate points, or nearly duplicate points. Although the intersection of $\mathcal{E}^+$ and $\mathcal{E}^-$

is empty[5], our proposed solution path algorithm would also encounter the singularity in Q when some features in $\mathcal{E}^+$ and $\mathcal{E}^-$ are linearly dependent or nearly duplicate. Another issue is that the determinant of Q is always near infinite when the training data size is large; thus, computing the inverse of Q in a numerical computing environment is not accurate.

To address these two issues, we use the LU decomposition with partial pivoting [18] to compute (11). Specifically, the matrix Q can be decomposed as $PQ = LU$, where L is a lower triangular matrix, U is an upper triangular matrix, and P is a permutation matrix that is multiplied with Q to record the rows of Q. Thus, we can obtain the following linear system according to (11) and solve it for z:

$$\underbrace{LU\begin{bmatrix}\xi^+_{\mathcal{E}^+}\\ \xi^-_{\mathcal{E}^-}\end{bmatrix}}_{z} = P\begin{bmatrix}-\mathbf{1}+\delta\sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|}\widehat{y}_{k,i}\widehat{x}_{k,i,\mathcal{E}^+}\\ -\mathbf{1}-\delta\sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|}\widehat{y}_{k,i}\widehat{x}_{k,i,\mathcal{E}^-}\end{bmatrix} \quad (12)$$

Note that we do not need to solve it for z by computing the inverse of L. Actually, the linear system (12) for z is very easy to solve by an iterative process called forward substitution [18]. Specifically, if $\text{rank}(L) < |\mathcal{E}^+ \cup \mathcal{E}^-|$, a general solution with $|\mathcal{E}^+ \cup \mathcal{E}^-| - \text{rank}(L)$ free parameters can be obtained. However, the solution path algorithm only needs one solution. Therefore, we choose a specific solution z and solve the linear system $U\begin{bmatrix}\xi^+_{\mathcal{E}^+}\\ \xi^-_{\mathcal{E}^-}\end{bmatrix} = z$, to obtain $\xi^+_{\mathcal{E}^+}$ and $\xi^-_{\mathcal{E}^-}$. Similar to (12), the linear system is also easily solved by an iterative process called back substitution [18].

Substituting $\xi^+_{\mathcal{E}^+}$ and $\xi^-_{\mathcal{E}^-}$ into (9) and (10), we can obtain the following linear relationship between Δg_j^+, Δg_j^- and $\Delta\lambda_1$:

$$\begin{aligned}\Delta g_j^+ &= \sum_{i=1}^{n} x_{i,j}\left(x^T_{i,\mathcal{E}^+}\xi^+_{\mathcal{E}^+} - x^T_{i,\mathcal{E}^-}\xi^-_{\mathcal{E}^-}\right)\Delta\lambda_1 + \Delta\lambda_1\\ &\quad -\Delta\lambda_2\sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|}\widehat{y}_{k,i}\widehat{x}_{k,i,j} \overset{\text{def}}{=} \gamma_j^+\Delta\lambda_1 \quad (13)\end{aligned}$$

$$\begin{aligned}\Delta g_j^- &= \sum_{i=1}^{n} x_{i,j}\left(-x^T_{i,\mathcal{E}^+}\xi^+_{\mathcal{E}^+} + x^T_{i,\mathcal{E}^-}\xi^-_{\mathcal{E}^-}\right)\Delta\lambda_1 + \Delta\lambda_1\\ &\quad +\Delta\lambda_2\sum_{k=1}^{\kappa}\sum_{i=1}^{|S_k|}\widehat{y}_{k,i}\widehat{x}_{k,i,j} \overset{\text{def}}{=} \gamma_j^-\Delta\lambda_1 \quad (14)\end{aligned}$$

When adjusting λ_1 while maintaining all the KKT conditions, the following constraints should be kept:

$$\beta_j^+ + \xi_j^+\Delta\lambda_1 \geq 0, \quad \forall j \in \mathcal{E}^+ \quad (15)$$

$$g_j^+ + \gamma_j^+\Delta\lambda_1 \geq 0, \quad \forall j \in \overline{\mathcal{E}^+} \quad (16)$$

$$\beta_j^- + \xi_j^-\Delta\lambda_1 \geq 0, \quad \forall j \in \mathcal{E}^- \quad (17)$$

$$g_j^- + \gamma_j^-\Delta\lambda_1 \geq 0, \quad \forall j \in \overline{\mathcal{E}^-} \quad (18)$$

Thus, while holding the partitions π^+ and π^- unchanged, we can compute the maximum adjustment of $\Delta\lambda_1$ (denoted as $\Delta\lambda_1^{max}$) according to (15), (16), (17), and (18).

After computing the maximum adjustment $\Delta\lambda_1^{max}$, we can easily fit β^+, β^-, g^+, and g^- w.r.t. λ_1 in the interval $[\lambda_1, \lambda_1 + \Delta\lambda_1^{max}]$, based on the linear coefficients $\xi^+_{\mathcal{E}^+}$, $\xi^-_{\mathcal{E}^-}$, γ_j^+, and γ_j^-. Thus, repeating this procedure for finding all such intervals by computing the maximum adjustment $\Delta\lambda_1^{max}$, we derive a solution path algorithm for problem (4) (i.e., Algorithm 1), which can fit the solution path of (4) for all values of λ_1 when $\Delta\lambda_2 = \delta\Delta\lambda_1$.

Algorithm 1 Solution path of SROC

Input: S, $\widetilde{S} = \bigcup_{k=1}^{\kappa} S_k$, δ.
Output: Solution path of β^+ and β^- w.r.t. λ_1.

1: Let $\lambda_1 = 0$, $\lambda_2 = 0$.
2: Initialize β^+, β^- according to Section 3.1.
3: Compute g^+, g^-, π^+ and π^-.
4: **while** $\sum_{i=1}^{d}(\beta_i^+ + \beta_i^-) > 0$ & $\lambda_1 < \infty$ **do**
5: Compute $\xi^+_{\mathcal{E}^+}$, $\xi^-_{\mathcal{E}^-}$ according to the LU decomposition with partial pivoting.
6: Compute γ_j^+, γ_j^- according to (13)-(14).
7: Compute $\Delta\lambda_1^{max}$ according to (15)-(18).
8: Update λ_1, β^+, β^-, g^+, g^-, π^+ and π^-.
9: **end while**

[5]When β_j is nonzero, without loss of generality, we let $\beta_j^+ > 0$. Then, we can easily derive that β_j^- must be zero. Thus, it is easy to obtain that the intersection of $\mathcal{E}^+$ and $\mathcal{E}^-$ is empty.

3.3. Computational Complexity

In this section, we provide the computational complexity of the solution path algorithm.

As mentioned in Algorithm 1, for each iteration, we need to solve the system of equations (11) of size $|\mathcal{E}^+| + |\mathcal{E}^-|$ to compute $\xi^+_{\mathcal{E}^+}$ and $\xi^-_{\mathcal{E}^-}$, which mainly involves an *LU* decomposition. Based on the Coppersmith-Winograd algorithm [8], the *LU* decomposition involves $O\left((|\mathcal{E}^+| + |\mathcal{E}^-|)^{2.37}\right)$ computations. According to (13)-(14), the computation of γ^+_j and γ^-_j requires $O\left((|\mathcal{E}^+| + |\mathcal{E}^-|)n\right)$ computations. According to (15)-(18), the computation of $\Delta\lambda_1^{max}$ requires $O\left(d\right)$ computations. The updates of λ_1, β^+, β^-, g^+, g^-, π^+ and π^- require $O\left(d\right)$ computations. Thus, the computational complexity of each iteration is $O\left((|\mathcal{E}^+| + |\mathcal{E}^-|)n + (|\mathcal{E}^+| + |\mathcal{E}^-|)^{2.37} + d\right)$.

Similar to [12, 9], our experience shows that the number of iterations in the solution path algorithm is some small multiple of d. If the average of $|\mathcal{E}^+| + |\mathcal{E}^-|$ is s, the computational complexity of the solution path algorithm is $O\left(snd + ds^{2.37} + d^2\right)$. Because $s \ll d$, it is clear that the computational complexity is mainly dominated by the dimension number d.

4. Experiments

In this section, we experimentally evaluate our SROC on a clinical cardiac image dataset. We first give the experimental setup, and then, we present our experimental results and discussion.

4.1. Experimental Setup

Data Acquisition As a reference modality for the non-invasive assessment of ventricular function, MRI supplies accurate information on morphology, muscle perfusion, and blood flow. In this paper, we mainly focus on MRI imagery to compute the EF. Specifically, 100 clinical subjects were scanned by a 1.5 T scanner. Each subject's data contains 20 frames throughout the cardiac cycle. In each frame, three representative slices [24] (i.e., apical, mid-cavity and basal; see Fig. 2) are selected following the standard AHA prescriptions [4]. This generates 6000 MRI images, and each image is 40×40 pixels. For each image, 1600

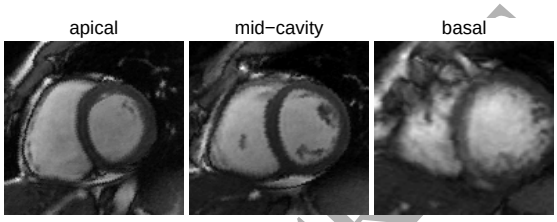

Figure 2. 3 representative cardiac slices in the short-axis view for each frame of an MRI scan.

features are extracted from the intensity to predict the cavity areas of LV and RV. Further, the volumes are computed by integrating the LV/RV cavity areas of the three representative slices in the sagittal direction, as mentioned in [24]. The manual segmentations are viewed as the benchmark.

Approaches Table 1 presents six approaches used in our experiments. The first five approaches are existing approaches, and the last approach is our proposed approach. Note that NN and RF are the abbreviations of Neural Networks and Random Forests, respectively, and HOG (the abbreviation of Histogram of Oriented Gradients) and PGF (the abbreviation of Pyramidal Gabor Features) are two types of features specifically used in [31]. NMR denotes the nuclear-norm-based matrix regression method [26].

Table 1
Approaches used in our experiments.

Approaches	Source	Sparse	Features
NN	[1]	no	image distance
Bayesian	[24]	no	template matching
RF	[31]	yes	intensity+HOG+PGF
NMR	[26]	yes	intensity
Lasso	[23]	yes	intensity
SROC	new	yes	intensity

Performance metrics We used two types of criteria to measure the performance of each approach: one is the quantitative description, and the other is the qualitative description. Let $\widehat{EF}_k$

8Bin Gu[81,2,3,4], Yingying Shan[3], Victor S.Sheng[5], and Shuo Li[4,6]1Jiangsu Engineering Center of Network Monitoring, Nanji

denote the predicted EF value of the k-th subject. We have the following quantitative description based metrics:

1. Mean Squared Error (MSE): $\frac{1}{\kappa}\sum_{k=1}^{\kappa}(\widehat{EF}_k - EF_k)^2$;
2. Mean Absolute Error (MAE): $\frac{1}{\kappa}\sum_{k=1}^{\kappa}|\widehat{EF}_k - EF_k|$.

A qualitative description of the EF is presented in Table 2, which is a comprehensive grading system designed to diagnose patients [14]. Suppose that $G(EF)$ denotes the grade of the EF value, as presented in Table 2. We have the following metrics [11] for measuring the errors of the predicted grades:

1. Mean Zero-one Error (MZE): $\frac{1}{\kappa}\sum_{k=1}^{\kappa}[G(\widehat{EF}_k) \neq G(EF_k)]$;[7]
2. MAE: $\frac{1}{\kappa}\sum_{k=1}^{\kappa}|G(\widehat{EF}_k) - G(EF_k)|$.

In a simplified case, both grades 1 and 2 are denoted as normal, and grades above 2 are marked as abnormal. Then, we have the standard binary classification error rate (BCER) [5].

Table 2
A comprehensive grading system for the EF.

Range of EF	Grade	Description
0-35%	5	Severe dysfunction
35-40%	4	Moderate dysfunction
40-50%	3	Mild dysfunction
50-70%	2	Normal
70-100%	1	Hyperdynamic

Implementation The performances in terms of both the quantitative description and the qualitative description were evaluated by a leave-one-subject-out validation approach, i.e., by training an algorithm using the dataset with 99 subjects and testing on the remaining one subject dataset, similar to the approach employed in [1, 24, 31].

[7]The boolean test $[\cdot]$ is 1 if the inner condition is true and 0 otherwise.

We implemented Algorithm 1 and NMR. In addition, we used the "lasso" function in the statistics toolbox to implement the solution path of Lasso.

To select a good parameter pair λ_1 and λ_2 for SROC on each training dataset, a two-layer 5-fold cross-validation procedure was employed as follows.

1. Essentially, Algorithm 1 is a one-dimensional solution path algorithm. Thus, an outer-layer 2-step grid searching is performed on δ. Specifically, the initial search is performed on a 20 coarse grid linearly spaced in the region $\{\log_2\delta| - 9 \leq \log_2\delta \leq 10\}$, followed by a fine search on a 20 uniform grid linearly spaced by 0.1 in the $\log_2\delta$ space.
2. For each value of δ, Algorithm 1 returns a corresponding solution path w.r.t. λ_1. Based on the solution path, an inner-layer grid search is performed on λ_1. Specifically, the search is performed on a 4000 uniform grid linearly spaced in the region $\{\log_2\lambda_1| - 39 \leq \log_2\lambda_1 \leq 40\}$.

The second step of the above two-layer 5-fold cross-validation procedure was also used to select the regularization parameter λ in Lasso.

4.2. Experimental Results and Discussion

Usefulness of Algorithm 1 To demonstrate the usefulness of Algorithm 1, we count the occurrences of singularities in $\widetilde{Q}$ (denoted as *Singularity*) and infinities in the determinant of $\widetilde{Q}$ (denoted as *Infinity*) during the running of Algorithm 1, and we investigate the number of iterations in Algorithm 1 (denoted as *Iterations*) over 20 trials. Table 3 presents the corresponding average numbers of *Singularity*, *Infinity*, and *Iterations* under different training set sizes (i.e., 480, 960, 1440, and 1920). From this table, we find that both *Singularity* and *Infinity* occur with a high probability. Algorithm 1 can handle *Singularity* and *Infinity* successfully and fit the entire solution path in a finite number of iterations. In addition, we also measure the average running time of Algorithm 1 under different training set sizes. The results confirm that the computational

Table 4
The performance of five approaches for predicting LV and RV EFs w.r.t. different metrics.

Approaches	Quantitative (LV)		Qualitative (LV)			Quantitative (RV)		Qualitative (RV)		
	MSE	MAE	MZE	MAE	BCER	MSE	MAE	MZE	MAE	BCER
NN	0.1185	0.3057	0.6044	1.253	0.23	0.0321	0.1558	0.7964	1.758	0.34
Bayesian	0.0098	0.0622	0.33	0.42	0.11	0.0109	0.0863	0.56	0.75	0.21
RF	0.0168	0.1057	0.55	0.91	0.16	0.0307	0.1491	0.75	1.48	0.24
NMR	0.0131	0.0836	0.416	0.585	0.15	0.0132	0.0912	0.5641	0.786	0.17
Lasso	0.0143	0.0895	0.446	0.625	0.15	0.0139	0.0933	0.571	0.804	0.17
SROC	**0.0092**	**0.0613**	**0.326**	**0.412**	**0.11**	**0.008**	**0.063**	**0.486**	**0.629**	**0.14**

complexity of Algorithm 1 is dominated by the dimension number d.

Table 3
The average numbers of *Singularity*, *Infinity*, and *Iterations* and the average running time of Algorithm 1.

Size	*Singularity*	*Infinity*	*Iterations*	*Time(s)*
480	1100	0	1100	232
960	920	0	920	240
1440	0	1082	1103	236
1920	0	1222	1263	243

Comparison with Other Methods The results of the six approaches for predicting LV and RV EFs on the metrics from the quantitative and qualitative analyses are presented in Table 4. From the results, we find that SROC achieves a better performance in predicting the LV and RV EFs than the other approaches (NN [1], Bayesian [24], RF [31], NMR [26] and Lasso [23]). It should be noted that the Bayesian method [24] uses template matching techniques, and the RF method [31] utilizes other advanced features (i.e., HOG and PGF). However, NMR, Lasso and SROC only use the basic feature (i.e., intensity). Fig. 3 presents the manual EF values and the predicted EF values of Lasso and SROC for LV and RV on the first 56 subjects. The results show that SROC produces a better prediction of the EF in most cases compared to Lasso. This is because SROC incorporates the correlation between the ground truth volumes and the estimated volumes with respect to each subject to improve the prediction of the EF, although it is a Lasso-type learning algorithm in nature.

5. Conclusion

In this paper, we proposed a new sparse regression with OC (SROC). We believe that this is important for machine learning and medical applications. For machine learning, SROC incorporates the OC as a regularization term, which is lacking in traditional regression methods. For medical applications, SROC is better able to estimate the cardiac EF than traditional regression methods. More specifically, we first proposed SROC based on the ℓ_1-norm and the OC. Then, we proposed a one-dimensional solution path algorithm for quickly finding two good regulation parameters in the formulation of SROC. Finally, we conduct experiments on a clinical cardiac image dataset with 100 subjects. The experimental results showed that SROC produces competitive and strong results for estimating the cardiac EF based on quantitative and qualitative analyses.

In the future, we plan to design a multi-task Lasso learning algorithm that can incorporate the strong correlation between the LV and RV volume outputs to estimate the LV and RV EFs simultaneously. We also plan to design independent solution paths of λ_1 and λ_2 and extend the method of bi-parameter space partitioning [10] to compute the solution surface w.r.t. both λ_1 and λ_2 for SROC.

Acknowledgments

This work was supported by the open Project (No. KXK1405) funded by Jiangsu Provincial Key Laboratory of big data analysis tech-

10Bin Gu[91,2,3,4], Yingying Shan[3], Victor S.Sheng[5], and Shuo Li[4,61]Jiangsu Engineering Center of Network Monitoring, Nanj

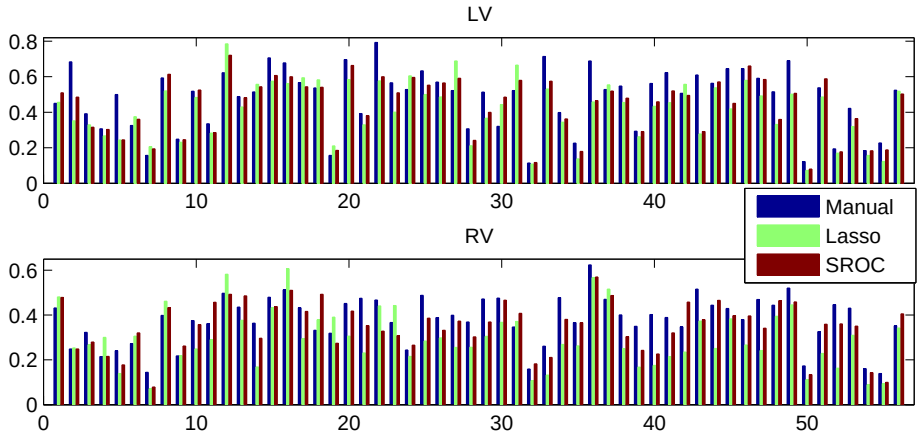

Figure 3. The manual LV and RV EF values vs. the predicted EF values of both Lasso and SROC on the first 56 subjects.

niques, the Project Funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions, the Natural Science Foundation of Jiangsu Province of China (No. BK20161534), and the National Natural Science Foundation of China (Nos. 61232016, 61573191, 61202137 and 61572259).

References

[1] M. Afshin, I. B. Ayed, A. Islam, A. Goela, T. M. Peters, and S. Li. Global assessment of cardiac function using image statistics in mri. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 535–543. Springer, 2012.

[2] Le An, Xiaojing Chen, Songfan Yang, and Bir Bhanu. Sparse representation matching for person re-identification. *Information Sciences*, 355:74–89, 2016.

[3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2009.

[4] M. D. Cerqueira, N. J. Weissman, V. Dilsizian, A. K. Jacobs, S. Kaul, W. K. Laskey, D. J. Pennell, J. A. Rumberger, T. Ryan, M. S. Verani, et al. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart a statement for healthcare professionals from the cardiac imaging committee of the council on clinical cardiology of the american heart association. *Circulation*, 105(4):539–542, 2002.

[5] Nello Cristianini and John Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[6] J. Dai, C. Chang, F. Mai, D. Zhao, and W. Xu. On the svmpath singularity. *Neural Networks and Learning Systems, IEEE Transactions on*, 24(11):1736–1748, Nov 2013.

[7] David L. Donoho and Yaakov Tsaig. Fast solution of l_1-norm minimization problems when the solution may be sparse. *IEEE Transactions on Information Theory*, 54 (11):4789–4812, 2008.

[8] François Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th international symposium on symbolic and algebraic computation*, pages 296–303. ACM, 2014.

[9] Bin Gu and Victor S Sheng. A robust regularization path algorithm for ν-support vector classification. *IEEE Transactions on neural networks and learning systems*, 2016.

[10] Bin Gu, Victor S. Sheng, and Shuo Li. Bi-parameter space partition for cost-sensitive SVM. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 3532–3539, 2015.

[11] B. Gu, V. Sheng, K. Tay, W. Romano, and S. Li. Incremental support vector learning for ordinal regression, 26(7):1403–1416, 2015.

[12] B. Gu, J.-D. Wang, G.-S. Zheng, and Y.-C. Yu. Regularization path for ν-support vector classification. *Neural Networks and Learning Systems, IEEE Transactions on*, 23(5):800–811, May 2012.

[13] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3:1157–1182, 2003.

[14] R. C. Hendel, M. J. Budoff, J. F. Cardella, C. E. Chambers, J. M. Dent, D. M. Fitzgerald, J. M. Hodgson, E. Klodas, C. M. Kramer, A. E. Stillman, et al. Acc/aha/acr/ase/asnc/hrs/nasci/rsna/saip/scai/scct/scmr/sir 2008 key data elements and definitions for cardiac imaging: A report of the american college of cardiology/american heart association task force on clinical data standards (writing committee to develop clinical data standards for cardiac imaging). *Journal of the American College of Cardiology*, 53(1):91–124, 2009.

[15] Huawen Liu, Shichao Zhang, and Xindong Wu. Mlslr: Multilabel learning via sparse logistic regression. *Information Sciences*, 281:310 – 320, 2014.

[16] S. Mendis, P. Puska, B. Norrving, et al. *Global atlas on cardiovascular disease prevention and control*. World Health Organization, 2011.

[17] C.-J. Ong, S. Shao, and J. Yang. An improved algorithm for the solution of the regularization path of support vector machine. *Neural Networks, IEEE Transactions on*, 21(3):451–462, 2010.

[18] D. Poole. *Linear algebra: A modern introduction*. Cengage Learning, 2014.

[19] Jianjun Qian, Lei Luo, Jian Yang, Fanlong Zhang, and Zhouchen Lin. Robust nuclear norm regularized regression for face recognition with occlusion. *Pattern Recognition*, 48(10):3145–3159, 2015.

[20] S. Rosset and J. Zhu. Piecewise linear regularized solution paths. *The Annals of Statistics*, pages 1012–1030, 2007.

[21] Craig Saunders, Alexander Gammerman, and Volodya Vovk. Ridge regression learning algorithm in dual variables. In *(ICML-1998) Proceedings of the 15th International Conference on Machine Learning*, pages 515–521. 1998.

[22] Changbin Shao, Xiaoning Song, Zhen-Hua Feng, Xiao-Jun Wu, and Yuhui Zheng. Dynamic dictionary optimization for sparse-representation-based face classification using local difference images. *Information Sciences*, 393:1–14, 2017.

[23] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[24] Z. Wang, M. Ben Salah, B. Gu, A. Islam, A. Goela, and S. Li. Direct estimation of cardiac biventricular volumes with an adapted bayesian formulation. *Biomedical Engineering, IEEE Transactions on*, 61(4):1251–1260, April 2014.

12Bin Gu[101,2,3,4], Yingying Shan[3], Victor S.Sheng[5], and Shuo Li[4,61]Jiangsu Engineering Center of Network Monitoring, Nar

[25] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, 31(2):210–227, 2009.

[26] Jian Yang, Lei Luo, Jianjun Qian, Ying Tai, Fanlong Zhang, and Yong Xu. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE transactions on pattern analysis and machine intelligence*, 39(1):156–171, 2017.

[27] Meng Yang, Lei Zhang, Jian Yang, and David Zhang. Regularized robust coding for face recognition. *IEEE Transactions on Image Processing*, 22(5):1753–1766, 2013.

[28] S. Yusuf, M. A. Pfeffer, K. Swedberg, C. B. Granger, P. Held, J. J. McMurray, E. L. Michelson, B. Olofsson, and J. Ostergren. Effects of candesartan in patients with chronic heart failure and preserved left-ventricular ejection fraction: the charm-preserved trial. *The Lancet*, 362(9386):777 – 781, 2003.

[29] Zheng Zhang, Yong Xu, Jian Yang, Xuelong Li, and David Zhang. A survey of sparse representation: algorithms and applications. *IEEE access*, 3:490–530, 2015.

[30] Hengmin Zhang, Jian Yang, Jianchun Xie, Jianjun Qian, and Bob Zhang. Weighted sparse coding regularized nonconvex matrix regression for robust face recognition. *Information Sciences*, 394:1–17, 2017.

[31] X. Zhen, Z. Wang, A. Islam, I. Chan, and S. Li. Direct estimation of cardiac biventricular volumes with regression forests. In *Accepted by Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. 2014.

[32] Xiaofeng Zhu, Xuelong Li, and Shichao Zhang. Block-row sparse multiview multilabel learning for image classification. *IEEE transactions on Cybernetics*, 46(2):450–461, 2016b.

[33] X. Zhu, X. Li, S. Zhang, C. Ju, and X. Wu. Robust joint graph sparse coding for unsupervised spectral feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, PP(99):1–13, 2016a. ISSN 2162-237X. doi:10.1109/TNNLS.2016.2521602.

[34] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.