Computational Statistics and Data Analysis 57 (2013) 429–434

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Short communication

A note on the lack of symmetry in the graphical lasso

Benjamin T. Rolfs [a,*], Bala Rajaratnam [b]

[a] Institute for Computational and Applied Mathematics, Stanford University, Stanford, CA 94395, United States

[b] Department of Statistics, Stanford University, Stanford, CA 94305, United States

ARTICLE INFO

Article history:
Received 9 February 2012
Received in revised form 30 June 2012
Accepted 10 July 2012
Available online 20 July 2012

Keywords:
Concentration model selection
Glasso
Graphical Gaussian models
Graphical lasso
ℓ_1 regularization

ABSTRACT

The graphical lasso (glasso) is a widely-used fast algorithm for estimating sparse inverse covariance matrices. The glasso solves an ℓ_1 penalized maximum likelihood problem and is available as an R library on CRAN. The output from the glasso, a regularized covariance matrix estimate $\hat{\Sigma}_{glasso}$ and a sparse inverse covariance matrix estimate $\hat{\Omega}_{glasso}$, not only identify a graphical model but can also serve as intermediate inputs into multivariate procedures such as PCA, LDA, MANOVA, and others. The glasso indeed produces a covariance matrix estimate $\hat{\Sigma}_{glasso}$ which solves the ℓ_1 penalized optimization problem in a dual sense; however, the method for producing $\hat{\Omega}_{glasso}$ after this optimization is inexact and may produce asymmetric estimates. This problem is exacerbated when the amount of ℓ_1 regularization that is applied is small, which in turn is more likely to occur if the true underlying inverse covariance matrix is not sparse. The lack of symmetry can potentially have consequences. First, it implies that $\hat{\Sigma}_{glasso}^{-1} \neq \hat{\Omega}_{glasso}$ and, second, asymmetry can possibly lead to negative or complex eigenvalues, rendering many multivariate procedures which may depend on $\hat{\Omega}_{glasso}$ unusable. We demonstrate this problem, explain its causes, and propose possible remedies.

1. Introduction

In modern applications, many data sets are simultaneously high-dimensional and low in sample size. Classic examples include microarray gene expression and SNP data. Dealing with such datasets has become an area of great interest in many fields such as biostatistics. Algorithms such as the graphical lasso (Friedman et al., 2008; Hastie et al., 2009) have been proposed to obtain regularized covariance estimators in the $n \ll p$ setting (where n is the sample size and p is the problem dimension) as well as perform graphical model selection.

In the case of the graphical lasso, graphical model selection involves inferring a concentration graph (or equivalently, a Markov model). A concentration graph encodes zeros in the inverse covariance (concentration) matrix, i.e., $i \not\sim j$ for $i, j \in \{1, \ldots, p\}$ in the graph implies that the partial correlation $\rho\left(X_i, X_j | X_{k \notin \{i,j\}}\right) = 0$. Along with inferring such a graph, the glasso provides $p \times p$ dimensional matrix estimators for both the covariance and concentration matrices, denoted $\hat{\Sigma}_\lambda$ and $\hat{\Omega}_\lambda$ respectively, for a given penalty parameter $\lambda > 0$. In particular, $\hat{\Omega}_\lambda$ is the solution to the convex maximization problem

$$\hat{\Omega}_\lambda = \hat{\Sigma}_\lambda^{-1} = \arg\min_{X \succ 0} \left[\log\det(X) - \operatorname{tr}(SX) - \lambda \|X\|_1\right], \tag{1}$$

where S is the sample covariance matrix, $X = \left\{x_{ij}\right\}_{i,j=1}^{p}$ is positive definite and $\|X\|_1 = \sum_{i,j} |x_{ij}|$. The non-zero elements of $\hat{\Omega}_\lambda$ correspond to edges in the estimated concentration graph.

* Corresponding author.
E-mail addresses: benrolfs@stanford.edu (B.T. Rolfs), brajarat@stanford.edu (B. Rajaratnam).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.07.013

In some applications, graphical model selection is the primary goal, where in other situations the estimators $\hat{\Sigma}_\lambda$ and $\hat{\Omega}_\lambda$ are used as inputs into other multivariate algorithms where a regularized covariance estimator is required. Typical examples include LDA, PCA, and MANOVA. Hence, it is often necessary that not only $\hat{\Sigma}_\lambda^T = \hat{\Sigma}_\lambda \succ 0$, but also that $\hat{\Omega}_\lambda^T = \hat{\Omega}_\lambda$, $\hat{\Omega}_\lambda \succ 0$, and $\hat{\Omega}_\lambda^{-1} = \hat{\Sigma}_\lambda$. We find that the output of the graphical lasso does not meet these conditions in certain situations, explain why, and discuss how to solve this problem. Such situations arise primarily when S is rank-deficient and λ is small. A low level of regularization is required when the true underlying concentration matrix is not sparse. It should however be noted that the glasso algorithm does indeed solve the dual problem corresponding to (1), so the above assertions should be interpreted in context.

2. Motivating examples

We now present two motivating examples, one in a classical setting and another in a high-dimensional setting, to illustrate the problem.

2.1. Example 1: low dimensional, large sample size inverse covariance estimation

Consider $n = 500$ *i.i.d.* samples drawn from a $p = 5$ dimensional multivariate Gaussian distribution with mean $\mu = 0$ and concentration matrix:

$$\Omega = \begin{bmatrix} 2.425 & 0.069 & -0.885 & 0 & 0 \\ 0.069 & 2.944 & -0.129 & 0.988 & 0 \\ -0.885 & -0.129 & 2.696 & 0.035 & -0.974 \\ 0 & 0.988 & 0.035 & 1.724 & 0.851 \\ 0 & 0 & -0.974 & 0.851 & 1.000 \end{bmatrix}.$$

The glasso algorithm was applied to this data set. A regularization parameter of $\lambda = 0.0033$, which is close to the cross-validated estimate, was chosen to demonstrate the problem. The glasso estimators for Ω and $\Sigma = \Omega^{-1}$ for a given λ are denoted $\hat{\Omega}_\lambda$ and $\hat{\Sigma}_\lambda$.

For reasons which are clarified in Section 3, the glasso produces estimators which are neither symmetric nor true inverses of one another, i.e., $\hat{\Omega}_\lambda^T \neq \hat{\Omega}_\lambda$ and $\hat{\Sigma}_\lambda^{-1} \neq \hat{\Omega}_\lambda$. To quantify the lack of symmetry, consider the matrix of relative errors between the elements of $\hat{\Omega}_\lambda$ and $\hat{\Omega}_\lambda^T$, as defined by $Err_{ij} = 100 \left| \frac{\hat{\Omega}_\lambda(i,j) - \hat{\Omega}_\lambda^T(i,j)}{\hat{\Omega}_\lambda(i,j)} \right| \%$.

For the numerical example above,

$$Err = \begin{bmatrix} 0 & 1.94 & 0.05 & 0 & 0.25 \\ 1.98 & 0 & 2.84 & 0.04 & \infty \\ 0.05 & 2.77 & 0 & 0.88 & 0.04 \\ 0 & 0.04 & 0.89 & 0 & 0.01 \\ 0.25 & 100.00 & 0.04 & 0.01 & 0 \end{bmatrix}$$

with the convention that if $\hat{\Omega}_\lambda(i,j) = 0 = \hat{\Omega}_\lambda(j,i)$ then $Err_{ij} = 0$. Note that the entries $Err_{5,2} = 100\%$ and $Err_{2,5} = \infty$ occur because $\hat{\Omega}_\lambda(5,2) \neq 0$ while $\hat{\Omega}_\lambda(2,5) = 0$.

Although the relative errors are small, i.e., on the order of 2%, there is a clear lack of symmetry in $\hat{\Omega}_\lambda$ and moreover the sparsity patterns in the upper and lower parts of $\hat{\Omega}_\lambda$ are different, and thus yield two different graphical models. In particular, $\hat{\Omega}_\lambda(5,2) \neq 0$ which indicates an edge between variables 2 and 5, while $\hat{\Omega}_\lambda(2,5) = 0$ indicates the absence of such. Furthermore, in high-dimensional examples, a graph is often calculated automatically when $\left(\hat{\Omega}_\lambda\right)_{ij} > \epsilon$ for some small ϵ. In such cases, a lack of symmetry may result, yielding two separate graphs.

2.2. Example 2: high dimensional, low sample size autoregressive model

The lack of symmetry in $\hat{\Omega}_\lambda$, and the resulting difference in the concentration graphs corresponding to the upper and lower parts of $\hat{\Omega}_\lambda$, often becomes more pronounced as the dimension p grows.

We now consider a high dimensional example with $n = 250$ *i.i.d.* samples drawn from a Gaussian AR(1) model such that $X_{t+1} = \phi X_t + \epsilon_t$ for $t = 2, \ldots, p$ and $X_1 = \epsilon_1$. Here, $p = 500$, $\phi = 0.75$, and $\epsilon_t \overset{i.i.d}{\sim} \mathcal{N}(0,1)$, $t = 1, \ldots, p$. The concentration matrix Ω is tridiagonal, with the diagonal entries equal to 1 and the off-diagonal entries equal to -0.75.

Given a glasso estimator $\hat{\Omega}_\lambda$, let E_1 and E_2 denote the edge sets corresponding to the upper and lower halves of $\hat{\Omega}_\lambda$, respectively. Then the symmetric difference $|E_1 \Delta E_2|$ is the number of edges which are present in the concentration graph encoded by one half of $\hat{\Omega}_\lambda$ but not in the graph encoded by the other half.

The glasso algorithm was applied to samples from the above model with the regularization parameter λ taking values between 0.001 and 0.03 in increments of 0.001. To put these values in perspective, note that when $\lambda = 0.03$, 102, 278 out