



Robust sparse Gaussian graphical modeling

Kei Hirose^{a,*}, Hironori Fujisawa^{b,c}, Jun Sese^d



^a Institute of Mathematics for Industry, Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka 819-0395, Japan

^b The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan

^c Nagoya University Graduate School of Medicine, 65 Tsurumai-cho, Showa-ku, Nagoya 466-8550, Japan

^d National Institute of Advanced Industrial Science and Technology, 2-4-7 Aomi, Koto-ku, Tokyo 135-0064, Japan

ARTICLE INFO

Article history:

Received 7 June 2017

Available online 6 August 2017

AMS 2010 subject classifications:

62H12

Keywords:

γ -divergence

Graphical lasso

Majorize-Minimization algorithm

Robust estimation

ABSTRACT

Gaussian graphical modeling is popular as a means of exploring network structures, such as gene regulatory networks and social networks. An L_1 penalized maximum likelihood approach is often used to learn high-dimensional graphical models. However, the penalized maximum likelihood procedure is sensitive to outliers. To overcome this problem, we introduce a robust estimation procedure based on the γ -divergence. The proposed method has a redescending property, which is a desirable feature in robust statistics. The parameter estimation procedure is constructed using the Majorize-Minimization algorithm, which guarantees that the objective function monotonically decreases at each iteration. Extensive simulation studies show that our procedure performs much better than the existing methods, in particular, when the contamination ratio is large. Two real data analyses are used for illustration purposes.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Gaussian graphical modeling is widely used to investigate conditional independence between two variables given other variables. Under a Gaussian assumption, conditional independence between two variables corresponds to a zero entry in the inverse covariance matrix [10]. A sparse estimation of this matrix, i.e., a method in which some of the elements of the inverse covariance matrix are exactly zero, is often used to obtain the conditional independence graph.

In many applications, the number of variables is much larger than the number of observations. This is the case, e.g., in the analysis of microarray gene expression data with a view to investigate the relation between pairs of genes. In such a case, the maximum likelihood estimate of the inverse covariance matrix does not exist. To overcome this problem, there has been a great deal of interest in L_1 regularization, such as the lasso [39], for estimating the sparse inverse covariance matrix. Meinshausen and Bühlmann [32] proposed fitting the lasso regression to each variable, in which one variable is a response and the other variables are predictors. Their method does not guarantee that the non-zero pattern of the inverse covariance matrix is symmetric. Peng et al. [34] introduced a joint regression, which is also based on the lasso regression and ensures the symmetry of the estimated inverse covariance matrix. Yuan and Lin [48] considered the graphical lasso, which involves maximizing the penalized log-likelihood function via the lasso. Among these methods, the graphical lasso has been the most popular because of its computational efficiency (see, e.g., [16,21,44]) and desirable statistical properties in high-dimensional settings [35,36].

In practical situations, however, outliers are often observed or the distribution is heavy-tailed [14,17]. In such cases, the conventional estimation procedure may produce an inappropriate graph structure. To overcome this problem, several

* Corresponding author.

E-mail address: hirose@imi.kyushu-u.ac.jp (K. Hirose).

researchers proposed robust estimation procedures with L_1 penalization. Liu et al. [29] proposed the *nonparanormal*, in which a truncated marginal empirical distribution was adopted to remove outliers, and a Gaussian copula was used to describe the conditional independence structure. The parameters of the model can then be estimated by a standard algorithm of the graphical lasso, such as the blockwise coordinate descent algorithm [16]. Finegold and Drton [14] introduced the *tlasso*, in which the underlying distribution was assumed to be the multivariate Student t distribution, which is heavy-tailed. The model parameters were estimated by the EM algorithm. Vinciotti and Hashem [42] compared the performance of various robust estimation procedures, including the nonparanormal and the *tlasso*, and they concluded that the nonparanormal performed well in many cases. Sun and Li [37] considered a modified likelihood approach based on the density power divergence [4] (hereafter referred to as the *dp-lasso*). The model parameter was estimated by the coordinate descent algorithm with a quadratic approximation [41].

Recently, many researchers also resorted to rank-based statistics for robust estimation in sparse Gaussian graphical modeling. Liu et al. [28] and Xue et al. [45] independently proposed to use rank correlation coefficients to estimate the correlation matrix in a robust way, and then plug the correlation matrix into the existing lasso-type sparse estimation methods in the graphical model, e.g., CLIME [6], the Dantzig Selector [7,47], and the graphical lasso. These approaches are regarded as a generalization of the nonparanormal [29]. Loh and Tan [30] also proposed robust estimation based on a rank correlation coefficient, essentially along the same lines as Liu et al. [28]. Barber and Kolar [3] established the asymptotic normality of the estimator based on Kendall's tau, and constructed robust confidence intervals; this is called the *ROCKET*, for *Robust Confidence Intervals via Kendall's Tau*. More generally, semi-parametric and non-parametric approaches have recently been proposed by several researchers; see, e.g., [13,43].

Nevertheless, the above procedures have drawbacks. The nonparanormal approach removes observations on both sides at the rate 2δ , i.e., observations that have extremely large positive and negative values are removed at the same rate δ . The truncation parameter δ corresponds to the contamination ratio and must be selected beforehand. In [29], δ was selected in such a way as to achieve a desired rate of convergence for the estimator. Nevertheless, the selected truncation parameter tends to be too small when the contamination ratio is large, because $\delta \rightarrow 0$ as $n \rightarrow \infty$, where n is the sample size. In addition, outliers may not be present on both sides in the same proportion δ . In fact, the outliers of yeast gene expression data described in Section 6.1 have only large negative values. The *tlasso* has the same drawback, because this heavy-tailed distribution generates both large and small outliers. Furthermore, the variance of the estimator derived from the *tlasso* tends to be large because a heavy-tail distribution often produces a small Fisher information. The *dp-lasso* approach, which has four tuning parameters used in the updating step of the algorithm, is often unstable in our experience. Rank-based approaches can eliminate the negative effect of outliers, but they cannot perform well when the contamination ratio is large. Moreover, in our simulation study, we observed that all of the above estimation procedures performed poorly when the contamination ratio was large and outliers were present on one side, and the estimators had a large mean squared error even when the sample size was sufficiently large.

To address these issues, we propose the γ -lasso, which is a robust sparse estimation procedure of the inverse covariance matrix based on the γ -divergence [9,18]. The γ -lasso regards an observation whose likelihood value is small as an outlier, unlike the nonparanormal. As a result, the γ -lasso can appropriately treat the outliers even when they exist on only one side. In addition, we do not need to know the contamination ratio in advance. The γ -lasso tends to yield a much smaller estimator variance than the *tlasso*, because the underlying distribution is assumed to be Gaussian. Parameter estimation is carried out using the Majorize-Minimization (MM) algorithm [22], which guarantees that the objective function monotonically decreases at each iteration. As a result, the parameter estimation is more stable than the *dp-lasso*. In addition, the γ -lasso has a redescending property, so that the bias of the estimator is expected to be sufficiently small when an outlier takes a large value [31]. We conducted extensive Monte Carlo simulations to investigate the performance of the proposed procedure. The result showed that our procedure performed better than existing methods in most cases. The proposed procedure is available through the R package `rsggm`.¹

This paper is organized as follows. In Section 2, we introduce a robust estimation of the sparse inverse covariance matrix via the γ -divergence. Section 3 provides a parameter estimation procedure via the MM algorithm. In Section 4, we compare the proposed procedure with several existing methods. Section 5 investigates the effectiveness of our proposed procedure via Monte Carlo simulations. Section 6 describes two real data analyses of the gene expression data. Concluding remarks are given in Section 7. Some technical proofs are collected in the *Appendices*.

2. Robust and sparse estimation of the inverse covariance matrix

2.1. Gaussian graphical model

Let $\mathbf{X} = (X_1, \dots, X_p)^\top$ be a $p \times 1$ multivariate normal random vector with mean $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^\top$ and covariance matrix $\boldsymbol{\Sigma} = (\sigma_{ij})$. Let the inverse of $\boldsymbol{\Sigma}$ be denoted $\boldsymbol{\Omega} = (\omega_{ij})$. It is well known that each variable can be written as $X_i = \sum_{j \neq i} \beta_{ij} X_j + \delta_i$, where $\beta_{ij} = -\omega_{ij}/\omega_{ii}$ and $\delta_i \sim \mathcal{N}(0, 1/\omega_{ii})$, and then the zero/non-zero element of the inverse covariance matrix corresponds to conditional independence/dependence given the other variables. The sparsity pattern of the inverse covariance matrix corresponds to the graph structure: there is an edge between vertices i and j if and only if $\omega_{ij} \neq 0$. We estimate the inverse covariance matrix by a sparse matrix to obtain a sparse graphical model.

¹ Available at <http://cran.r-project.org/web/packages/rsggm>.

Download English Version:

<https://daneshyari.com/en/article/5129321>

Download Persian Version:

<https://daneshyari.com/article/5129321>

[Daneshyari.com](https://daneshyari.com)