Journal of Multivariate Analysis 107 (2012) 119–140

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Model selection and estimation in the matrix normal graphical model

Jianxin Yin, Hongzhe Li *

School of Statistics and Center for Applied Statistics, Renmin University of China, No. 59 Zhongguancun Street, Haidian District, Beijing 100872, China
Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021, USA

ARTICLE INFO

Article history:
Received 1 June 2011
Available online 10 January 2012

AMS subject classifications:
62
92

Keywords:
Gaussian graphical model
Gene networks
High dimensional data
l_1 penalized likelihood
Matrix normal distribution
Sparsistency

ABSTRACT

Motivated by analysis of gene expression data measured over different tissues or over time, we consider matrix-valued random variable and matrix-normal distribution, where the precision matrices have a graphical interpretation for genes and tissues, respectively. We present a l_1 penalized likelihood method and an efficient coordinate descent-based computational algorithm for model selection and estimation in such matrix normal graphical models (MNGMs). We provide theoretical results on the asymptotic distributions, the rates of convergence of the estimates and the sparsistency, allowing both the numbers of genes and tissues to diverge as the sample size goes to infinity. Simulation results demonstrate that the MNGMs can lead to a better estimate of the precision matrices and better identifications of the graph structures than the standard Gaussian graphical models. We illustrate the methods with an analysis of mouse gene expression data measured over ten different tissues.

1. Introduction

Gaussian graphical models (GGMs) provide natural tools for modeling the conditional independence relationships among a set of random variables [23,37]. Many methods of estimating the standard GGMs have been developed in recent years, especially in high-dimensional settings. Meinshausen and Bühlmann [27] took a neighborhood selection approach to this problem by fitting a l_1 penalized regression or Lasso [33] to each variable using the other variables as predictors. They show that this neighborhood selection procedure consistently estimates the set of non-zero elements of the precision matrix. Other authors have proposed algorithms for the exact maximization of the l_1-penalized log-likelihood. Yuan and Lin [39], Banerjee et al. [4] and Dahl et al. [9] adapted an interior point optimization method for the solution to this problem. Based on the work of Banerjee et al. [4] and a block-wise coordinate descent algorithm, Friedman et al. [16] developed the graphical Lasso (glasso) for sparse inverse covariance estimation, which is computationally very efficient even when the dimension is greater than the sample size. Yuan [38] developed a linear programming procedure for high dimensional inverse covariance matrix estimation and obtained oracle inequalities for the estimation error in terms of several matrix norms. Some theoretical properties of this type of methods have also been developed by Yuan and Lin [39], Ravikumar et al. [30], Rothman et al. [31] and Lam and Fan [22]. Cai et al. [6] developed a constrained l_1 minimization approach to sparse precision matrix estimation, extending the idea of the Dantzig selector [7] developed for sparse high dimensional regressions.

The standard likelihood framework for building Gaussian graphical models assumes that samples are independent and identically distributed from a multivariate Gaussian distribution. This assumption is often limited in certain applications. For example, in genomics, gene expression data of p genes collected over q different tissues from the same subject are

* Corresponding author at: Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021, USA.
E-mail address: hongzhe@upenn.edu (H. Li).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.01.005

often correlated. For a given sample, let $\mathbf{Y}$ be the $p \times q$ matrix of the expression data, where the jth column corresponds to the expression data of p genes measured in the jth tissue, and the ith row corresponds to gene expressions of the ith gene over q different tissues. Instead of assuming that the columns or rows are independent, we assume that the matrix variate random variable $\mathbf{Y}$ follows a matrix normal distribution [10,23,18], where both row and column precision matrices can be specified. The matrix-variate normal distribution has been studied in analysis of multivariate linear model under the assumption of independence and homoscedasticity for the structure of the among-row and among-column covariance matrices of the observation matrix [15,34]. Such a model has also be applied to spatio-temporal data [26,21]. In genomics, Teng and Huang [32] proposed to use the Kronecker product matrix to model gene-experiment interactions, which leads to a gene expression matrix following a matrix-normal distribution. The gene expression matrix measured over multiple tissues is transposable, meaning that potentially both the rows and/or columns are correlated. Such matrix-valued normal distribution was also used in [2,12] for modeling gene expression data in order to account for gene expression dependency across different experiments. Dutilleul [11] developed the maximum likelihood estimation (MLE) algorithm for the matrix normal distribution. Mitchell et al. [28] developed a likelihood ratio test for separability of the covariances. Muralidharan [29] used a matrix normal framework for detecting column dependence when rows are correlated and estimating the strength of the row correlation.

The precision matrices of the matrix normal distribution provide the conditional independence structures of the row and column variables [23], where the non-zero off-diagonal elements of the precision matrices correspond to conditional dependencies among the elements in row or column of the matrix normal distribution. The matrix normal models with specified non-zero elements of the precision matrices define the matrix normal graphical models (MNGMs). This is analogous to the relationship between the Gaussian graphical model and the precision matrix of a multivariate normal distribution. Despite the flexibility of the matrix normal distribution and the MNGMs in modeling the transposable data, methods for model selection and estimation of such models have not been developed fully, especially in high dimensional settings. Wang and West [36] developed a Bayesian approach for the MNGMs using Markov Chain Monte Carlo sampling scheme that employs an efficient method for simulating hyper-inverse Wishart variates for both decomposable and nondecomposable graphs. Allen and Tibshirani [2,3] proposed penalized likelihood approaches for such matrix normal models, where both l_1-norm and l_2-norm penalty functions are used on the precision matrices.

The focus of this paper is to develop a model selection and estimation method for the MNGMs based on a l_1 penalized likelihood approach under the assumption of both row and column precision matrices being sparse. Our penalized estimation method is the same as that proposed in [2,1,3] when l_1 penalty is used. Allen and Tibshirani [2,3] only considered the setting when there is one observed matrix-variate normal data and used the estimated covariance matrices for imputing the missing data and for de-correlating the noise in the underlying data. We focus on evaluating how well such a l_1 penalized estimation method recovers the underlying graphical structures that correspond to the row and column precision matrices when we have n i.i.d. samples from a matrix normal distribution. In addition, we provide asymptotic justification of the estimates and show that the estimates enjoy similar asymptotic and oracle properties as the penalized estimates for the standard GGMs [13,22,39] even when the dimensions $p = p_n$ and $q = q_n$ diverge as the number of observations $n \to \infty$. In addition, if consistent estimates of the precision matrices are available and are used in the adaptive l_1 penalty functions, the resulting estimates have the property of sparsistency.

The rest of the paper is organized as follows. We introduce the MNGMs as motivated by analysis of gene expression data across multiple tissues in Section 2. In Section 3 we present a l_1 penalized likelihood estimate of such a MNGM and an iterative coordinate descent procedure for the optimization. We present the asymptotic properties of the estimates in Section 4 in both the classic setting when the dimensions are fixed and the setting allowing the dimensions to diverge as the sample size goes to infinity. In Section 5 we present simulation results and comparisons with the standard Gaussian graphical model. We present an application of the MNGM in Section 6 to an analysis of mouse gene expression data measured over 10 different tissues. Finally, in Section 7 we give a brief discussion. The proofs of all the theorems are given in the Appendix.

2. Matrix normal graphical model for multi-tissue gene expression data

We consider the gene expression data measured over different tissues. Let $\mathbf{Y}$ be the random $p \times q$ matrix of the gene expression levels of p genes over q tissues. Let vec($\mathbf{A}$) be the vectorization of a matrix $\mathbf{A}$ obtained by stacking the columns of the matrix $\mathbf{A}$ on top of one another. Instead of assuming that the expression levels are independent over different tissues, following [32], we can model this gene expression matrix as

$$\mathbf{Y} = \mathbf{G} + \mathbf{T} + \mathbf{I}_{GT} + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{G}$ and $\mathbf{T}$ are expected (constant) effects from the genes and tissues respectively, $\mathbf{I}_{GT}$ are the interaction effects that are assumed to be random with vec($\mathbf{I}_{GT}$) following a multivariate normal distribution with zero means and a covariance matrix $\mathbf{V} \otimes \mathbf{U}$, where the covariance matrices $\mathbf{U}$ and $\mathbf{V}$ respectively represent the gene and tissue dependencies, and $\boldsymbol{\epsilon}$ represents small random normal noises with zero means arising from all nuisance sources. With negligible nuisance effects, vec($\mathbf{Y}$) follows a multivariate normal distribution with means vec($\mathbf{M}$) = vec($\mathbf{G} + \mathbf{T}$) and a covariance matrix $\mathbf{V} \otimes \mathbf{U}$ [32].

Treating the data $\mathbf{Y}$ as a matrix-valued random variable, we say $\mathbf{Y}$ follows a matrix normal distribution, if $\mathbf{Y}$ has a density function

$$p(\mathbf{Y}|\mathbf{M}, \mathbf{U}, \mathbf{V}) = k(\mathbf{U}, \mathbf{V}) \exp(-\mathrm{tr}\{(\mathbf{Y} - \mathbf{M})^T \mathbf{U}^{-1} (\mathbf{Y} - \mathbf{M}) \mathbf{V}^{-1} / 2\}), \tag{2}$$