TEST
https://doi.org/10.1007/s11749-022-00807-0

ORIGINAL PAPER

Precision matrix estimation using penalized Generalized Sylvester matrix equation

Vahe Avagyan[1]

Received: 31 March 2021 / Accepted: 3 March 2022

Abstract
Estimating a precision matrix is an important problem in several research fields when dealing with large-scale data. Under high-dimensional settings, one of the most popular approaches is optimizing a Lasso or ℓ_1 norm penalized objective loss function. This penalization endorses sparsity in the estimated matrix and improves the accuracy under a proper calibration of the penalty parameter. In this paper, we demonstrate that the problem of minimizing Lasso penalized D-trace loss can be seen as solving a penalized Sylvester matrix equation. Motivated by this method, we propose estimating the precision matrix using penalized generalized Sylvester matrix equations. In our method, we develop a particular estimating equation and a new convex loss function constructed through this equation, which we call the generalized D-trace loss. We assess the performance of the proposed method using detailed numerical analysis, including simulated and real data. Extensive results show the advantage of the proposed method compared to other estimation approaches in the literature.

Keywords D-trace loss · Gaussian graphical models · Generalized Sylvester matrix equation · ℓ_1 Norm penalization · Linear discriminant analysis

Mathematics Subject Classification 62A09 · 15A24 · 62H30 · 90C06

1 Introduction

Precision or inverse covariance matrix has an important role in statistical learning and data analysis. Its applications span different research fields including genetics, brain studies, finance, psychology, etc. Moreover, under high-dimensional settings, the

✉ Vahe Avagyan
vahe.avagyan@wur.nl

[1] Biometris, Wageningen University and Research,
Droevendaalsesteeg 1 (Radix), 6708 PB Wageningen, The Netherlands

Published online: 09 April 2022

accurate estimation of a precision matrix is crucial for several statistical methodologies, including classification, forecasting, among others.

Precision matrix shows the partial correlations among normally distributed variables. Under the assumption of multivariate normality, the entry $\omega_{ij} = 0$ of the precision matrix $\Omega = [\omega_{ij}]_{1\leq i,j\leq p} \in R^{p\times p}$ indicates the conditional independence between the variables X^i and X^j, given all the other variables, i.e., $X^i \perp\!\!\!\perp X^j | X^{-(i,j)}$ (Dempster 1972). In this way, the precision matrix is closely related to the Gaussian graphical models (GGM) which is a useful technique to visualize the conditional independence of the variables (e.g., gene interaction networks). The GGM is an undirected graph $G = (N, E)$, where the set of nodes $N = \{1, \ldots, p\}$ represents the variables. The set of edges $E \subseteq N \times N$ consists of the pair indexes (i, j) corresponding to the "active" entries $\omega_{ij} \neq 0$ for $1 \leq i, j \leq p$ (Lauritzen 1996).

The estimation of large-scale precision matrices and corresponding GGMs has received a substantial attention in the extant literature. Under high-dimensional settings, one of the most commonly employed approaches in the literature is the Lasso (least absolute shrinkage and selection operator) or ℓ_1 norm penalization (Tibshirani 1996) of a certain loss function, which induces sparsity in the estimated precision matrix. In particular, Banerjee et al. (2006) introduce the ℓ_1 norm penalized log-likelihood maximization approach, also known as graphical Lasso or GLASSO estimator (see also Yuan and Lin 2007; Friedman et al. 2008; Scheinberg et al. 2010; Rothman et al. 2008; Ravikumar et al. 2011; Hsieh et al. 2014, for theoretical analyses and different solving algorithms for this estimator). Other methods that employ ℓ_1 norm penalization approach include the neighborhood selection for selecting the GGMs (Meinshausen and Bühlmann 2006), sparse partial correlation estimation or SPACE (Peng et al. 2009), constrained ℓ_1 norm minimization for inverse matrix estimation or CLIME (Cai et al. 2011), ℓ_1 norm penalized D-trace loss minimization (Zhang and Zou 2014), sparse column-wise inverse operator or SCIO (Liu and Luo 2015), among several others. Despite the popularity of ℓ_1 norm penalty, literature considers also other penalties such as Adaptive ℓ_1 norm (Fan et al. 2009; Avagyan et al. 2018), SCAD (Fan et al. 2009), Elastic-Net (Ryali et al. 2012), Ridge (i.e., squared Frobenius or ℓ_2 norm) (van Wieringen and Peeters 2016; Kuismin et al. 2017) and Generalized Ridge penalties (van Wieringen 2019). Note that the methods based on penalization framework require a proper selection (i.e., calibration) of the tuning parameter that controls the strength (i.e., intensity) of the employed penalty. This can be done empirically through several techniques such as penalized goodness-of-fit criteria (e.g., Bayesian Information Criterion) and cross-validation methods. In this paper, we employ both techniques for selecting the penalty parameters of the considered methods. For the further review on methods to estimate precision matrices, we refer to Fan et al. (2016) and Kuismin and Sillanpää (2017).

In line with the literature mentioned above, we propose a precision matrix estimation method using ℓ_1 norm penalized convex minimization. Our introduced method is motivated by the D-trace framework of Zhang and Zou (2014). First, we show that the D-trace loss function is closely related to the Sylvester matrix equation. In other words, minimizing a penalized D-trace loss function is equivalent to solving a penalized Sylvester equation. We note that the Sylvester equation is a particular case of matrix equation family called generalized Sylvester equations. Next, we discuss the estimation of the precision matrices through particular penalized generalized Sylvester equation. Furthermore, we construct a loss function based on the selected penalized equation. We call this function a generalized D-trace loss. Using extensive numerical analysis, we show that the estimated precision matrix obtained through solving the introduced penalized generalized Sylvester equation (i.e., minimizing ℓ_1 norm penalized generalized D-trace loss) provides more favorable performance in finite samples than the existing methods. We evaluate the performances of the considered estimators in terms of several statistical measures for different models (i.e., patterns) of the true precision matrix Ω.

The manuscript is organized as follows. In Sect. 2, we describe the proposed methodology. In Sect. 3, we evaluate the statistical performance of the proposed methodology and compare it with that of other approaches. In Sect. 4, we provide a real data application: classification of breast cancer patients using linear discriminant analysis. We provide our conclusions in Sect. 5. Finally, we provide technical details in “Appendix A” and the required solving algorithm in “Appendix B.”

2 Proposed methodology

We use the following notations throughout the paper. For any p-dimensional vector $\mathbf{a} = (a_1, \ldots, a_p)^T \in \mathbb{R}^p$, we define the ℓ_2 norm by $||\mathbf{a}||_2 = \sqrt{\sum_{j=1}^p a_j^2}$. For any symmetric matrix $\mathbf{A} = [a_{ij}]_{1\leq i,j\leq p} \in \mathbb{R}^{p\times p}$, we denote the Frobenius norm by $||\mathbf{A}||_2 = \sqrt{\sum_{i=1}^p \sum_{j=1}^p a_{ij}^2}$, the ℓ_∞ norm by $||\mathbf{A}||_\infty = \max_{1\leq i,j\leq p} |a_{ij}|$, the matrix ℓ_1 norm by $||\mathbf{A}||_{\ell_1} = \max_{1\leq j\leq p} \sum_{i=1}^p |a_{ij}|$, the componentwise ℓ_1 norm by $||\mathbf{A}||_1 = \sum_{i=1}^p \sum_{j=1}^p |a_{ij}|$ and off-diagonal ℓ_1 norm by $||\mathbf{A}||_{1,\text{off}} = \sum_{i=1}^p \sum_{j=1, j\neq i}^p |a_{ij}|$, the spectral norm by $||\mathbf{A}||_{\text{sp}} = \sup_{||x||_2\leq 1} ||Ax||_2$. Furthermore, we assume that $\mathbf{X}_{n\times p}$ is mean-centered observed sample data matrix, where each row $X_i = \left(X_{i1}, \ldots, X_{ip}\right)$ is a p-dimensional normal random vector, i.i.d. for $i = 1, \ldots, n$ and has a covariance matrix $\Sigma = \Omega^{-1}$.

Many studies focus on the estimation of a precision matrix under high-dimensional settings. Among the proposed approaches, graphical Lasso (or GLASSO) is one of the most popular and well-studied estimators. This estimator is obtained by minimizing the ℓ_1 norm penalized negative log-likelihood function of a multivariate normal distribution (Banerjee et al. 2006; Friedman et al. 2008):

$$\widehat{\Omega}_{\text{GLASSO}} = \arg\min_{\Omega} \; -\log\det(\Omega) + \text{trace}(\Omega S) + \nu||\Omega||_{1,\text{off}}, \tag{1}$$

where $S = \dfrac{1}{n}\sum_{i=1}^n X_i^T X_i$ is the sample covariance matrix and $\nu > 0$ is the associated tuning (or penalty) parameter that controls the accuracy and the sparsity of the precision matrix estimator. As an alternative to the log-likelihood function in (1), Zhang and Zou (2014) introduce a new loss function called D-trace:

$$f_{\mathrm{DT}}(\Sigma, \Omega) = \frac{1}{2}\mathrm{trace}(\Omega^2\Sigma) - \mathrm{trace}(\Omega), \tag{2}$$

which is a quadratic function of the precision matrix. Furthermore, the authors propose a precision matrix estimation method based on minimizing the ℓ_1 norm penalized D-trace loss function (hereafter, DT):

$$\widehat{\Omega}_{\mathrm{DT}} = \arg\min_{\Omega \succeq \epsilon I} \frac{1}{2}\mathrm{trace}(\Omega^2 S) - \mathrm{trace}(\Omega) + \tau ||\Omega||_{1,\mathrm{off}}, \tag{3}$$

where $\tau > 0$ is the associated penalty parameter. The constraint $\Omega \succeq \epsilon I$ guarantees that the solution of (3) is positive definite, where $\epsilon > 0$ is a small positive value (we set $\epsilon = 10^{-8}$ in the numerical analyses). This problem can be easily solved through an algorithm developed by the same authors, which is based on the alternating direction method (see also Wang and Jiang 2020, for a similar solving algorithm).

According to the first-order condition, the solution of (3) satisfies the following penalized matrix equation:

$$\frac{1}{2}\Omega S + \frac{1}{2}S\Omega - I + \mathrm{Pen}_\tau(\Omega) = 0. \tag{4}$$

Here, $\mathrm{Pen}_\tau(\Omega)$ is the penalty term of the estimating equation and is defined as $\mathrm{Pen}_\tau(\Omega) = \tau Z(\Omega) \in R^{p\times p}$, where $Z(\Omega) = \dfrac{\partial ||\Omega||_{1,\mathrm{off}}}{\partial \Omega} \in [-1, 1]$ is the subgradient of the (off-diagonal) ℓ_1 norm of a matrix.

Note that if there is no penalty, i.e., $\tau = 0$, (4) is known as the Sylvester equation. This is a matrix equation with the following definition:

$$\Omega R + L\Omega + C = 0, \tag{5}$$

where R, L and C are known matrices and Ω is the unknown. In this way, DT estimation can be seen as solving a Sylvester equation (using $R = L = \dfrac{1}{2}S$ and $C = -I$) with an imposed $\mathrm{Pen}_\tau(\Omega)$ penalty.

The Sylvester equation (5) is closely related to the matrix equation family known as generalized Sylvester equations, which are defined as

$$\sum_{i=1}^{k} L_i \Omega R_i + C = 0, \tag{6}$$

where L_i, R_i $(i = 1, \ldots, k)$ and C are known matrices (see, for instance, Li et al. 2010; De Terán and Iannazzo 2016). Here, we assume that these matrices guarantee that the resulting estimator of Ω is symmetric. Note that the classical Sylvester equation 5 is a spacial case of 6.

Motivated by the DT estimator (4), we focus on estimating the precision matrices through penalized generalized Sylvester equations. In other words, we induce an additional penalty term in (6), under the assumption of positive definiteness of the estimated precision matrix:

$$\sum_{i=1}^{k} L_i \Omega R_i + C + \mathrm{Pen}_\lambda(\Omega) = 0. \tag{7}$$

Note that different choices of those matrices will lead to different equations. In this paper, we restrict ourselves with one special case, by setting $L_1 = R_2 = I$, $L_2 = R_1 = \frac{1}{4}S$, $L_3 = R_3 = \frac{\sqrt{2}}{2}S^{1/2}$ and $C = -I$. We demonstrate the estimation of the precision matrix using the following penalized matrix equation:

$$\frac{1}{4}\Omega S + \frac{1}{4}S\Omega + \frac{2}{4}S^{1/2}\Omega S^{1/2} - I + \lambda Z(\Omega) = 0, \tag{8}$$

where $\lambda > 0$ is the penalty parameter and Ω is required to be positive definite. The term $Z(\Omega)$ is defined earlier. In contrast to (4), the equation (8) has an extra weighted component $S^{1/2}\Omega S^{1/2}$, which imposes additional balanced constraints on the estimand. We hypothesize that this may provide additional benefits on the estimated precision matrix. On the other hand, similar to (4), when $\lambda = 0$, the solution of (8) is S^{-1}.

Notice that the solution of the penalized matrix equation (8) is the minimizer of the following optimization problem:

$$\min_{\Omega \succeq \epsilon I} \frac{1}{4}\mathrm{trace}(\Omega^2 S) + \frac{1}{4}\mathrm{trace}(S^{1/2}\Omega S^{1/2}\Omega) - \mathrm{trace}(\Omega) + \lambda||\Omega||_{1,\mathrm{off}}. \tag{9}$$

Again, the constraint $\Omega \succeq \epsilon I$ is added to guarantee the positive definiteness of Ω. We call the proposed precision matrix estimator generalized D-trace estimator $\widehat{\Omega}_{\mathrm{GDT}}$. Correspondingly, we define a new loss function as

$$f_{\mathrm{GDT}}(\Sigma, \Omega) = \frac{1}{4}\mathrm{trace}(\Omega^2\Sigma) + \frac{1}{4}\mathrm{trace}(\Sigma^{1/2}\Omega\Sigma^{1/2}\Omega) - \mathrm{trace}(\Omega), \tag{10}$$

which we call generalized D-trace loss function. Notice that because of the additional term $\mathrm{trace}(\Sigma^{1/2}\Omega\Sigma^{1/2}\Omega)$, the f_{GDT} function is no longer quadratic of Ω (but rather "quasi-quadratic"). Nevertheless, $f_{\mathrm{GDT}}(\Sigma, \Omega)$ is a convex function of Ω and has a unique minimizer at Σ^{-1} (see "Appendix A" for more details on the proposed loss function). Finally, in order to solve the optimization problem (9), we use an algorithm based on the alternating direction method (see "Appendix B" for detailed description of the algorithm) similar to DT method.

The following remarks are in order. First, this paper does not aim at discussing the dominance of one loss function over the other one or which loss function should be used under different circumstances. Moreover, in this article, we discuss the advantages of employing penalized generalized Sylvester equation for estimating precision matrices by focusing on one particular case given in (8). As mentioned earlier, different generalized Sylvester equations can be introduced based on different choices of

L_i, R_i and C in (6). Notice that these matrices should be selected properly in order to guarantee the symmetry of the esimtating function (i.e., left side of Eq. (7)) and the corresponding precision matrix estimate. For example, in (8), the symmetry of the estimating function is ensured. We leave the selection of those matrices for the optimal estimating equation (7) for the future research. Second, in this article, we focus only on ℓ_1 norm penalized generalized Sylvester equation. We note that for instance, the adaptive (i.e., weighted) ℓ_1 norm penalization can also be employed in (8) (see, for instance Avagyan et al. 2018, for Adaptive ℓ_1 norm penalized D-trace loss minimization approach).

3 Simulation analysis

We demonstrate the numerical performance of the proposed approach based on simulated data generated using different models for the true precision matrix Ω in terms of several statistical measures. In our study, we consider the most popular techniques for selecting the penalty parameters for the considered methods. We compare $\hat{\Omega}_{\text{GDT}}$ with D-Trace estimator $\hat{\Omega}_{\text{DT}}$ (Zhang and Zou 2014), graphical Lasso estimator $\hat{\Omega}_{\text{GLasso}}$ (Banerjee et al. 2006) and CLIME estimator $\hat{\Omega}_{\text{CLIME}}$ (Cai et al. 2011).

3.1 Performance evaluation

We generate multivariate normal random samples with zero mean and covariance matrix $\Sigma = \Omega^{-1}$ for each model over 100 replications. We evaluate the quality of a precision matrix estimator based on popular statistical losses and sparsity pattern prediction measures, previously considered in the literature. In particular, we consider the Kullback–Leibler loss (KLL) or the entropy loss (see, for instance Yuan 2010; Yin and Li 2013, etc.), the reverse Kullback–Leibler loss (RKLL) (see Avagyan 2021), the relative trace error (RTE) and matrix losses (the Frobenius ℓ_2 norm, the spectral norm, and the matrix ℓ_1 norm) (see, for instance Cai et al. 2011; Zhang and Zou 2014; van Wieringen and Peeters 2016, etc.) defined as

$$\begin{aligned}
\text{KLL}(\widehat{\Omega}, \Omega) &= \text{trace}(\Omega^{-1}\widehat{\Omega}) - \log\det(\Omega^{-1}\widehat{\Omega}) - p,\\
\text{RKLL}(\widehat{\Omega}, \Omega) &= \text{trace}(\Omega\widehat{\Omega}^{-1}) - \log\det(\Omega\widehat{\Omega}^{-1}) - p,\\
\text{RTE}(\widehat{\Omega}, \Omega) &= \left|1 - \frac{\text{trace}(\widehat{\Omega})}{\text{trace}(\Omega)}\right|\\
\ell_2(\widehat{\Omega}, \Omega) &= ||\widehat{\Omega} - \Omega||_2,\\
\ell_{\text{sp}}(\widehat{\Omega}, \Omega) &= ||\widehat{\Omega} - \Omega||_{\text{sp}},\\
\ell_1(\widehat{\Omega}, \Omega) &= ||\widehat{\Omega} - \Omega||_{\ell_1}.
\end{aligned}$$

Furthermore, we evaluate the selection of GGM (i.e., sparsity pattern prediction) based on specificity, sensitivity and Matthews correlation coefficient (Matthews 1975) defined as

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}},$$
$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$
$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}.$$

Here, TP is the number of true positives (i.e., correctly selected nonzero entries), TN is the number of true negatives (i.e., correctly selected zero entries), FP is the number of false positives (i.e., incorrectly selected nonzero entries), and FN is the number of false negatives (i.e., incorrectly selected zero entries). The MCC is popular approach in statistics for measuring the binary classifications: the closer the MCC to one is, the better the overall classification is (see Chicco and Jurman 2020, for more details).

In this paper, we focus on both group of measures, i.e., statistical prediction and sparsity pattern selection. Usually, improving one measure may deteriorate the other one. Therefore, we aim at achieving a desirable performance in terms of both criteria, instead of focusing on either one. Moreover, this performance should remain consistent over different settings.

3.2 Simulation study 1

Our first simulation study is based on the following models.

- *Model* 1 $\omega_{ii} = 1$, $\omega_{i,i-1} = \omega_{i-1,i} = 0.45$ and other values are 0 (prevously considered in Yuan and Lin 2007, etc) friedman.
- *Model* 2 $\omega_{ii} = 1, \omega_{i,i-1} = \omega_{i-1,i} = 0.5, \omega_{i,i-2} = \omega_{i-2,i} = 0.35$ and other values are 0 (motivated by Kuismin et al. 2017)).
- *Model* 3 A block-diagonal matrix, with four equally sized blocks along the diagonal. Each block is defined as $\omega_{ij} = 0.6^{|i-j|}$ (prevously considered in Cai et al. 2011; Fan et al. 2009, etc).
- *Model* 4 A random positive definite matrix, with approximately 50% of nonzero entries, generated using MATLAB command `sprandsym`. The matrix is further standardized to have unit diagonal.

We set $n = 200$, $p = 200, 400$. For this study, we select the penalty parameters for the considered methods using Bayesian Information Criterion (BIC) (Yuan and Lin 2007).

3.3 Simulation study 2

Our second study is based on the following models previously considered in Zhang and Zou (2014).

- *Model* 5 $\omega_{i,i} = 1, \omega_{i,j} = 0.2$ for $1 \leq |i - j| \leq 2$ and other values are 0.
- *Model* 6 $\omega_{i,i} = 1, \omega_{i,j} = 0.2$ for $1 \leq |i - j| \leq 4$ and other values are 0.

- *Model* 7 $\omega_{i,i} = 1$, $\omega_{i,i+1} = 0.2$ for $\mathrm{mod}(i, p^{1/2}) \neq 0$, $\omega_{i,i+p^{1/2}} = 0.2$ and other values are 0. It is assumed that $p^{1/2}$ is an integer.

In line with Zhang and Zou (2014), we set $n = 400$, $p = 500$ for models 5 and 6 and $p = 484$ for model 7. For this study, we select the penalty parameters using fivefold cross-validation technique, consistent with the study of Zhang and Zou (2014).

3.4 Discussion of results

Tables 1, 2, 3, 4, 5, 6, and 7 report the averages (and standard deviations) of the measures over 100 replications. **Bold** letters indicate the best performance. First, we observe that GDT outperforms DT in terms of KLL, RKLL, RTE, ℓ_2 norm, ℓ_{sp} norm for all models, in terms of ℓ_1 norm for models 1, 2, 3, 4, in terms of Specificity for models 1, 2 (when $p = 400$), 3 (when $p = 200$), 4, 5, 6, 7, in terms of Sensitivity for models 3, 4, 5 and in terms of MCC for models 1, 2 (when $p = 400$), 3, 4, 5, 6, 7. Both GDT and DT provide similar Sensitivity for models 1 and 2. On the other hand, DT method performs better than GDT in terms of the matrix ℓ_1 norm for models 5, 6, 7, in terms of Specificity for models 2 (when $p = 200$), 3 (when $p = 400$), in terms of Sensitivity for models 6, 7 and in terms of MCC for model 2 (when $p = 200$).

Comparing our proposed method with GLASSO and CLIME methods, we see that in general, GDT provides better results, especially in terms of the statistical losses. However, GLASSO provides the best overall results in terms of the KLL for models

Table 1 Average measures (with standard deviations) over 100 replications for Model 1

	p	$\hat{\Omega}_{\mathrm{GDT}}$	$\hat{\Omega}_{\mathrm{DT}}$	$\hat{\Omega}_{\mathrm{GLASSO}}$	$\hat{\Omega}_{\mathrm{CLIME}}$
KLL	200	**7.920 (0.411)**	12.28 (0.973)	16.51 (0.893)	15.76 (0.665)
	400	**13.25 (0.612)**	33.86 (0.892)	37.26 (3.457)	33.12 (0.851)
RKLL	200	**6.895 (0.345)**	10.34 (0.862)	21.20 (1.285)	15.28 (0.660)
	400	**11.99 (0.539)**	29.02 (0.749)	48.42 (5.085)	30.82 (0.755)
RTE	200	**0.101 (0.009)**	0.156 (0.012)	0.340 (0.010)	0.235 (0.007)
	400	**0.070 (0.006)**	0.207 (0.005)	0.358 (0.016)	0.232 (0.005)
ℓ_2	200	**3.074 (0.116)**	3.905 (0.201)	6.620 (0.173)	5.133 (0.114)
	400	**4.068 (0.117)**	6.775 (0.107)	9.817 (0.392)	7.225 (0.098)
ℓ_{sp}	200	**0.597 (0.048)**	0.662 (0.043)	0.899 (0.023)	0.786 (0.038)
	400	**0.624 (0.063)**	0.765 (0.032)	0.936 (0.030)	0.786 (0.025)
ℓ_1	200	**0.759 (0.058)**	0.837 (0.054)	1.060 (0.034)	0.967 (0.056)
	400	**0.796 (0.077)**	0.920 (0.040)	1.108 (0.028)	0.983 (0.046)
Specificity	200	**0.992 (0.001)**	0.985 (0.002)	0.984 (0.003)	0.991 (0.001)
	400	**0.999 (0.001)**	0.998 (0.001)	0.991 (0.002)	0.993 (0.001)
Sensitivity	200	1 (0)	1 (0)	1 (0)	1 (0)
	400	1 (0)	1 (0)	1 (0)	1 (0)
MCC	200	**0.805 (0.012)**	0.709 (0.040)	0.695 (0.029)	0.797 (0.010)
	400	**0.924 (0.007)**	0.886 (0.009)	0.683 (0.053)	0.712 (0.006)

Table 2 Average measures (with standard deviations) over 100 replications for Model 2

	p	$\hat{\Omega}_{\mathrm{GDT}}$	$\hat{\Omega}_{\mathrm{DT}}$	$\hat{\Omega}_{\mathrm{GLASSO}}$	$\hat{\Omega}_{\mathrm{CLIME}}$
KLL	200	**14.73 (2.060)**	25.81 (0.760)	34.87 (2.476)	29.43 (0.943)
	400	**31.13 (1.034)**	52.62 (1.077)	80.76 (4.053)	89.76 (1.222)
RKLL	200	**13.35 (2.457)**	27.37 (0.887)	62.74 (6.118)	31.47 (1.124)
	400	**28.87 (0.927)**	54.83 (1.152)	152.3 (10.54)	128.2 (2.723)
RTE	200	**0.170 (0.030)**	0.310 (0.006)	0.471 (0.013)	0.309 (0.007)
	400	**0.180 (0.006)**	0.306 (0.004)	0.495 (0.009)	0.424 (0.003)
ℓ_2	200	**5.557 (0.607)**	8.552 (0.123)	11.85 (0.268)	8.911 (0.133)
	400	**8.157 (0.158)**	12.04 (0.111)	17.55 (0.253)	16.11 (0.088)
ℓ_{sp}	200	**1.123 (0.095)**	1.495 (0.043)	1.943 (0.039)	1.592 (0.055)
	400	**1.198 (0.058)**	1.506 (0.033)	2.025 (0.024)	1.994 (0.043)
ℓ_1	200	**1.480 (0.104)**	1.759 (0.063)	2.176 (0.031)	2.005 (0.091)
	400	**1.553 (0.088)**	1.796 (0.052)	2.257 (0.030)	2.387 (0.065)
Specificity	200	0.969 (0.007)	**0.985 (0.001)**	0.963 (0.007)	0.976 (0.001)
	400	0.993 (0.001)	0.989 (0.001)	0.982 (0.003)	**0.994 (0.001)**
Sensitivity	200	**1 (0)**	**1 (0)**	0.998 (0.002)	0.997 (0.002)
	400	**1 (0)**	**1 (0)**	0.996 (0.003)	0.960 (0.006)
MCC	200	0.669 (0.055)	**0.789 (0.008)**	0.630 (0.039)	0.705 (0.007)
	400	0.794 (0.006)	0.720 (0.006)	0.638 (0.033)	**0.797 (0.005)**

Table 3 Average measures (with standard deviations) over 100 replications for Model 3

	p	$\hat{\Omega}_{\mathrm{GDT}}$	$\hat{\Omega}_{\mathrm{DT}}$	$\hat{\Omega}_{\mathrm{GLASSO}}$	$\hat{\Omega}_{\mathrm{CLIME}}$
KLL	200	**19.72 (0.344)**	21.78 (0.352)	30.21 (0.394)	22.13 (0.357)
	400	**43.41 (2.509)**	52.62 (0.604)	61.35 (1.067)	52.21 (0.601)
RKLL	200	**31.44 (0.760)**	37.35 (0.847)	66.69 (0.955)	42.71 (0.996)
	400	**68.82 (7.742)**	98.66 (1.532)	136.2 (3.114)	111.7 (1.859)
RTE	200	**0.298 (0.007)**	0.332 (0.006)	0.452 (0.004)	0.375 (0.006)
	400	**0.299 (0.026)**	0.374 (0.004)	0.456 (0.004)	0.428 (0.004)
ℓ_2	200	**12.02 (0.085)**	12.64 (0.075)	14.41 (0.037)	13.06 (0.070)
	400	**17.32 (0.695)**	19.19 (0.067)	20.59 (0.082)	19.76 (0.066)
ℓ_{sp}	200	**2.817 (0.021)**	2.913 (0.017)	3.182 (0.010)	2.995 (0.018)
	400	**2.870 (0.085)**	3.074 (0.011)	3.219 (0.011)	3.146 (0.014)
ℓ_1	200	**3.157 (0.044)**	3.193 (0.035)	3.373 (0.023)	3.321 (0.043)
	400	**3.236 (0.071)**	3.310 (0.033)	3.418 (0.028)	3.420 (0.035)
Specificity	200	0.994 (0.001)	0.993 (0.001)	**0.995 (0.001)**	0.989 (0.001)
	400	0.998 (0.002)	**0.999 (0.001)**	0.995 (0.001)	0.996 (0.001)
Sensitivity	200	0.082 (0.002)	0.081 (0.002)	0.064 (0.001)	**0.094 (0.002)**
	400	0.035 (0.005)	0.032 (0.001)	0.035 (0.001)	**0.039 (0.001)**
MCC	200	**0.214 (0.005)**	0.203 (0.005)	0.185 (0.004)	0.203 (0.006)
	400	**0.148 (0.004)**	0.145 (0.002)	0.117 (0.004)	0.140 (0.002)

Table 4 Average measures (with standard deviations) over 100 replications for Model 4

	p	$\hat{\Omega}_{\text{GDT}}$	$\hat{\Omega}_{\text{DT}}$	$\hat{\Omega}_{\text{GLASSO}}$	$\hat{\Omega}_{\text{CLIME}}$
KLL	200	**60.04 (0.783)**	75.71 (3.654)	85.35 (2.673)	140.8 (3.190)
	400	**152.6 (1.330)**	176.9 (1.193)	213.3 (4.170)	343.1 (4.437)
RKLL	200	**175.1 (5.016)**	251.1 (30.43)	338.5 (21.70)	631.1 (27.20)
	400	**563.5 (13.8)**	653.5 (9.220)	997.2 (39.07)	2226 (83.42)
RTE	200	**0.412 (0.006)**	0.519 (0.020)	0.631 (0.006)	0.648 (0.005)
	400	**0.388 (0.005)**	0.568 (0.003)	0.688 (0.004)	0.740 (0.003)
ℓ_2	200	**15.01 (0.059)**	16.25 (0.318)	17.67 (0.100)	17.90 (0.069)
	400	**22.85 (0.067)**	24.29 (0.036)	26.25 (0.077)	26.93 (0.034)
ℓ_{sp}	200	**5.249 (0.024)**	5.367 (0.043)	5.380 (0.013)	5.479 (0.013)
	400	**4.821 (0.021)**	4.879 (0.014)	4.935 (0.009)	5.003 (0.010)
ℓ_1	200	**9.902 (0.072)**	10.09 (0.092)	10.17 (0.036)	10.34 (0.047)
	400	**11.94 (0.076)**	12.01 (0.048)	12.07 (0.020)	12.11 (0.025)
Specificity	200	0.965 (0.002)	0.961 (0.010)	0.967 (0.005)	**0.997 (0.001)**
	400	0.977 (0.001)	0.977 (0.001)	0.984 (0.002)	**0.999 (0.001)**
Sensitivity	200	0.123 (0.002)	0.116 (0.015)	**0.149 (0.011)**	0.044 (0.001)
	400	0.057 (0.001)	0.056 (0.001)	**0.068 (0.004)**	0.015 (0.001)
MCC	200	0.165 (0.006)	0.146 (0.006)	**0.203 (0.006)**	0.139 (0.004)
	400	0.087 (0.003)	0.085 (0.003)	**0.130 (0.004)**	0.081 (0.002)

Table 5 Average measures (with standard deviations) over 100 replications for Model 5

	$\hat{\Omega}_{\text{GDT}}$	$\hat{\Omega}_{\text{DT}}$	$\hat{\Omega}_{\text{GLASSO}}$	$\hat{\Omega}_{\text{CLIME}}$
KLL	**19.62 (0.451)**	20.52 (0.388)	20.88 (0.295)	25.99 (0.329)
RKLL	**19.96 (0.459)**	21.49 (0.410)	23.34 (0.321)	34.92 (0.540)
RTE	**0.045 (0.003)**	0.061 (0.003)	0.094 (0.003)	0.185 (0.003)
ℓ_2	**6.640 (0.074)**	6.966 (0.062)	7.223 (0.043)	8.645 (0.050)
ℓ_{sp}	**0.724 (0.023)**	0.745 (0.018)	0.759 (0.012)	0.917 (0.014)
ℓ_1	1.158 (0.061)	**1.106 (0.040)**	1.401 (0.061)	1.129 (0.025)
Specificity	0.984 (0.001)	0.983 (0.001)	0.963 (0.001)	**0.995 (0.001)**
Sensitivity	0.912 (0.008)	0.911 (0.009)	**0.919 (0.007)**	0.793 (0.010)
MCC	0.570 (0.006)	0.555 (0.006)	0.419 (0.005)	**0.707 (0.008)**

6, 7, in terms of Specificity for models 3 (when $p = 200$), 4, in terms of sensitivity for models 4, 5, 6, 7 and in terms of MCC for model 4. In addition, we observe that CLIME estimator provides the best overall results in terms of specificity for models 2 (when $p = 400$), 4, 5, 6, 7, in terms of sensitivity for model 3 and in terms of MCC for models 2 (when $p = 400$), 5, 6, 7. Note that these results are not surprising because GLASSO provides more dense solutions (thus, higher sensitivity but lower specificity), whereas CLIME provides more sparse solutions (thus, higher specificity but lower sensitivity), in general. This can be observed on the recorded numbers of nonzero entries of these

Table 6 Average measures (with standard deviations) over 100 replications for Model 6

	$\hat{\Omega}_{\text{GDT}}$	$\hat{\Omega}_{\text{DT}}$	$\hat{\Omega}_{\text{GLASSO}}$	$\hat{\Omega}_{\text{CLIME}}$
KLL	36.74 (0.417)	37.30 (0.375)	**35.80 (0.307)**	46.72 (0.347)
RKLL	**45.20 (0.520)**	47.36 (0.483)	47.52 (0.429)	78.13 (0.843)
RTE	**0.119(0.003)**	0.133 (0.003)	0.154 (0.002)	0.269 (0.003)
ℓ_2	**11.20 (0.046)**	11.48 (0.040)	11.42 (0.035)	13.39 (0.039)
ℓ_{sp}	**1.520 (0.016)**	1.560 (0.013)	1.544 (0.008)	1.800 (0.008)
ℓ_1	1.975 (0.048)	**1.947 (0.039)**	2.372 (0.059)	1.995 (0.020)
Specificity	0.983 (0.001)	0.982 (0.001)	0.945 (0.001)	**0.998 (0.001)**
Sensitivity	0.672 (0.007)	0.675 (0.007)	**0.736 (0.007)**	0.439 (0.008)
MCC	0.521 (0.006)	0.511 (0.005)	0.360 (0.004)	**0.580 (0.007)**

Table 7 Average measures (with standard deviations) over 100 replications for Model 7

	$\hat{\Omega}_{\text{GDT}}$	$\hat{\Omega}_{\text{DT}}$	$\hat{\Omega}_{\text{GLASSO}}$	$\hat{\Omega}_{\text{CLIME}}$
KLL	16.37 (0.431)	19.18 (0.415)	**15.34 (0.347)**	31.35 (0.533)
RKLL	**14.15 (0.362)**	16.15 (0.337)	15.48 (0.327)	35.02 (0.668)
RTE	**0.035 (0.003)**	0.058 (0.003)	0.093 (0.003)	0.217 (0.003)
ℓ_2	**4.577 (0.064)**	4.910 (0.059)	4.984 (0.053)	7.717 (0.064)
ℓ_{sp}	**0.506 (0.022)**	0.540 (0.018)	0.551 (0.011)	0.801 (0.012)
ℓ_1	0.927 (0.055)	**0.926 (0.037)**	1.295 (0.065)	1.030 (0.032)
Specificity	0.983 (0.001)	0.980 (0.001)	0.949 (0.001)	**0.999 (0.001)**
Sensitivity	0.994 (0.002)	0.995 (0.002)	**0.999(0.001)**	0.957 (0.006)
MCC	0.602 (0.004)	0.573 (0.004)	0.393 (0.003)	**0.933 (0.005)**

estimators applied on a real dataset (see Fig. 1). Moreover, the results indicate that good performance of GLASSO and CLIME in terms of GGM selection usually leads to deteriorated performances of matrix prediction (i.e., statistical losses).

In sum, the proposed GDT estimator in general provides better performance than DT, GLASSO, CLIME methods for most of the models in terms of most statistical losses and GGM prediction measures. Moreover, GDT shows a better trade-off between the matrix prediction and sparsity pattern identification, i.e., the outperformance in terms of one criterion does not diminish the other one.

In addition, we conduct a study with smaller off-diagonal nonzero entries in models 1-3 (not provided). The results support our discussion above (i.e., the comparison of the considered methods remains roughly the same) and shows the robustness of our simulation study.

In “Appendix A,” we briefly discuss the required theoretical assumption for the model selection, called the irrepresentability condition. Although we do not discuss theoretical properties of the considered method, we provide a simple numerical example (previously used in the literature) where the irrepresentability condition holds for the GDT estimator, but it fails for the DT estimator.

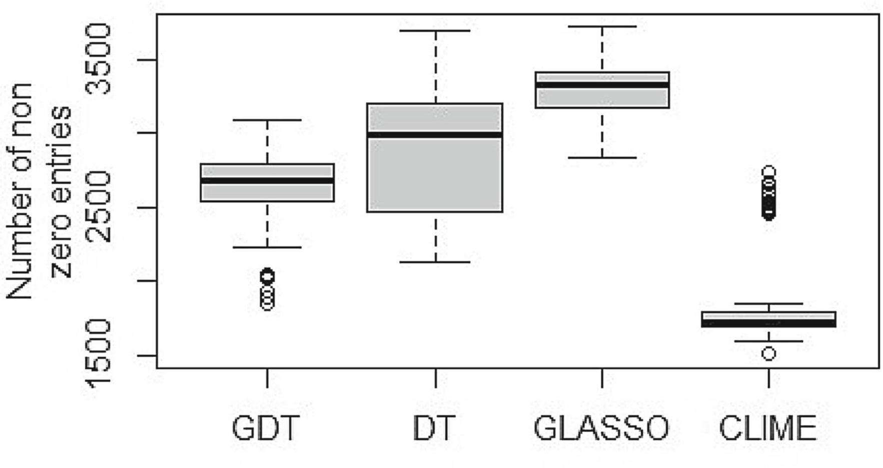

Fig. 1 Number of nonzero entries of the estimated precision matrix for different methods over 100 replications

4 Real data analysis

In this section, we demonstrate the performance of the considered methods on an empirical analysis. We focus on classifying breast cancer patients with pathological complete response (PCR) and patients with residual disease (RD). This is an important problem because the PCR condition after the treatment (e.g., chemotherapy) may potentially lead to a cancer-free life (Kuerer et al. 1999). We use a dataset (available at http://bioinformatics.mdanderson.org/pubdata.html) which contains 22283 gene expression levels of 133 patients (subjects) with breast cancer. Among these, 34 patients have PCR and 99 patients have RD.

Following Cai et al. (2011), we randomly divide the data into a training set and a testing set with sizes 112 and 21, respectively. The testing set consists of five subjects with PCR and 16 subjects with RD. The training set contains the remaining subjects. Next, we apply two sample t test between the two groups using the training set, and we select the most significant 150 genes with the smallest p-values. Using only the selected genes, we estimated the precision matrix Ω using the training set. For the sake of computational time, the penalty parameters are selected using the BIC technique. Finally, the estimated precision matrix is used in the linear discriminant analysis (LDA) score $\delta_t(Y) = Y^T \widehat{\Omega}\widehat{\mu}_t - \frac{1}{2}\widehat{\mu}_t^T \widehat{\Omega}\widehat{\mu}_t$, $t = 1$ for PCR and $t = 2$ for RD. Here, $\widehat{\mu}_t = \frac{1}{n_t}\sum_{i \in class_t} x_i$ is the within group average calculated using the training set. We use $\delta_t(Y)$ to classify the subject Y from the testing set. The classification rule is $\widehat{t} = \arg\max_t \delta_t(Y)$. We repeat this process 100 times.

In order to measure the overall classification accuracy, we use MCC defined earlier in Sect. 3.1. We consider TP and TN as the number of correctly predicted PCR and RD, respectively, and FP and FN as the number of wrongly predicted PCR and RD, respectively. Our calculation shows that the average MCC over 100 replications obtained using GDT is 0.445, whereas for DT, GLASSO, and CLIME, the average MCC values over 100 replications are 0.422, 0.338 and 0.406, respectively. This indicates that GDT provides better overall classification of the subjects with PCR condition compared to the other considered methods. In addition, we record the number of non-zero entries of the estimated precision matrix over 100 replications. Figure 1 shows that CLIME produces the sparsest precision matrix. On the other hand, our proposed GDT approach produces less nonzero entries on average than DT and GLASSO methods.

5 Conclusions

The current research presents a new method for estimating high-dimensional precision matrices. The proposed method is based on the ℓ_1 norm penalization of a new generalized D-trace loss function. Our introduced loss function is motivated by the D-trace loss. We show that D-trace loss function is based on the Sylvester equation, whereas our proposed loss function is constructed through the generalized Sylvester equations. In this article, we restrict ourselves with only one particular penalized equation, although different other possible versions can be proposed. Selecting the optimal equation and the corresponding loss function is an open question. We consider this as one of the future research direction. We provide an extensive numerical analysis using simulated data. Our study is based on several statistical measures and settings, including those previously used in the literature. The proposed method performs favorably compared to other estimators in the literature. Moreover, we demonstrate a better performance of our proposed method in an empirical application of breast cancer patient classification using linear discriminant analysis.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11749-022-00807-0.

Acknowledgements The author would like to thank the Editor, the Associate Editor, and two anonymous referees for their helpful comments that led to an improvement of this article.

Appendix A: Technical details

In this appendix, we confirm the properties of the GDT loss function. First, we show that $f_{\text{GDT}}(\Sigma, \Omega)$ is a convex function of Ω. We have the following:

$$\begin{aligned} & f_{\text{GDT}}(\Omega_1, \Sigma) + f_{\text{GDT}}(\Omega_2, \Sigma) - 2 f_{\text{GDT}}\left(\frac{\Omega_1 + \Omega_2}{2}, \Sigma\right) \\ & \quad = \frac{1}{8}\text{trace}((\Omega_1 - \Omega_2)^2 \Sigma) + \frac{1}{8}\text{trace}((\Omega_1 - \Omega_2)\Sigma^{1/2}(\Omega_1 - \Omega_2)\Sigma^{1/2}), \end{aligned}$$

which is nonnegative, given that for any symmetric matrix A and a positive definite matrix $B \succ 0$, the product ABA is positive semidefinite. This confirms the convexity of f_{GDT}. Next, we show that f_{GDT} has a unique minimizer at Σ^{-1}. We check the sign of the Hessian matrix of f_{GDT}:

$$\Gamma_{\text{GDT}}(\Sigma) = \frac{\partial^2 f_{\text{GDT}}}{\partial \Omega^2} = \frac{\Sigma \otimes I + I \otimes \Sigma + 2\Sigma^{1/2} \otimes \Sigma^{1/2}}{4},$$

where $\otimes$ is the Kronecker product. For $\Sigma \succ 0$, the Hessian matrix is always positive definite. Finally, we have

$$\frac{\partial f_{\text{GDT}}}{\partial \Omega} = \frac{\Sigma\Omega + \Omega\Sigma + 2\Sigma^{1/2}\Omega\Sigma^{1/2}}{4} - I.$$

By setting the first-order derivative to 0, we can see that the minimum of f_{GDT} occurs at Σ^{-1}.

Note that the Hessian matrix corresponding to the DT loss function (3) is defined as

$$\Gamma_{\text{DT}}(\Sigma) = \frac{\partial^2 f_{\text{DT}}}{\partial \Omega^2} = \frac{\Sigma \otimes I + I \otimes \Sigma}{2}.$$

In general, the Hessian matrix has an important role for the model selection consistency. More specifically, it is assumed that $\max_{e \in S^c} ||\Gamma_{e,S}\left(\Gamma_{S,S}\right)^{-1}||_1 < 1$, which is known as the irrepresentability condition. Here, we denote Γ as the value of the Hessian matrix at the true (unknown) covariance matrix (e.g., in case of DT estimator, $\Gamma = \Gamma_{\text{DT}}(\Sigma)$). Next, $S = \{(i, j)|\ \omega_{ij} \neq 0, \}$ is the set of nonzero entries (i.e., support) and S^c is its complement. We define $\Gamma_{S_1 S_2}$ as a sub-matrix of $\Gamma \in R^{p^2 \times p^2}$ with rows and columns indexing the subsets $S_1, S_2 \in \{1, \ldots, p\} \times \{1, \ldots, p\}$.

In this article, we do not provide theoretical properties for our proposed estimator. However, we suppose that in order to establish the model selection consistency of GDT estimator a similar irrepresentability condition would be required based on $\Gamma_{\text{GDT}}(\Sigma)$. In general, irrepresentability conditions are difficult to compare theoretically, i.e., show that a condition for one method is always weaker or stronger than the that for another method for a certain class of precision matrices. This remains an open question, and we plan to study this problem in separate paper. Nevertheless, we compare the irrepresentability conditions of DT and GDT estimators on a simple example, previously used by Ravikumar et al. (2011) and Zhang and Zou (2014).

Consider $\Omega = [\omega_{ij}]_{1\leq i,j\leq 4} \in R^{4\times 4}$, with $\omega_{11} = 1$, $\omega_{14} = \omega_{41} = 2c^2$, $\omega_{23} = \omega_{32} = 0$ and $\omega_{ij} = c$ otherwise, where we assume that the values of c guarantee the positive definiteness of the matrix. It can be checked that for DT estimator, the irrepresentability condition holds for $|c| \leq 0.315$. On the other hand, the irrepresentability condition based on $\Gamma = \Gamma_{\text{GDT}}(\Sigma)$ holds for $|c| \leq 0.335$, which shows that for $c \in (0.315, 0.335]$ the irrepresentability condition holds for GDT estimator, but it fails for DT estimator.

Appendix B: Algorithm

In this section, we describe an algorithm for obtaining the proposed estimator based on the alternating direction method. We modify the optimization problem (9) for matrices Ω_0 and Ω_1:

$$\begin{aligned}\widehat{\Omega}_{\text{GDT}} = \arg\min_{\Omega_1 \succ \epsilon I}\ & \frac{1}{4}\text{trace}(\Omega^2 S) + \frac{1}{4}\text{trace}(\Omega S^{1/2}\Omega S^{1/2}) - \text{trace}(\Omega) + \lambda||\Omega_0||_{1,\text{off}} \\ \text{subject to}\ & \{\Omega, \Omega\} = \{\Omega_0, \Omega_1\}.\end{aligned}$$

It is easy to see that the problem above is equivalent to (9). The Lagrangian of the new problem (B.1) is defined as

$$\begin{aligned}L(\Omega, \Omega_0, \Omega_1, \Lambda_0, \Lambda_1) = & \frac{1}{4}\text{trace}(\Omega^2 S) + \frac{1}{4}\text{trace}(\Omega S^{1/2}\Omega S^{1/2}) - \text{trace}(\Omega) \\ & +\lambda||\Omega_0||_{1,\text{off}} + h(\Omega_1 \succeq \epsilon I) + \text{trace}(\Lambda_0(\Omega - \Omega_0)) \\ & +\text{trace}(\Lambda_1(\Omega - \Omega_1)) + \frac{\rho}{2}||\Omega - \Omega_0||_2^2 + \frac{\rho}{2}||\Omega - \Omega_1||_2^2,\end{aligned}$$

where ρ, Λ_0, Λ_1 are the multipliers and h is an indicator function, which returns 0 if the statement $\Omega_1 \succeq \epsilon I$ is true and ∞, otherwise. For simplicity, we set $\rho = 1$. Let $(\Omega^t, \Omega_0^t, \Omega_1^t, \Lambda_0^t, \Lambda_1^t)$ is the solution at step $t = 0, 1, 2, \ldots$. The solution is updated according to the following steps:

$$\Omega^{t+1} = \arg\min_{\Omega=\Omega^T} L(\Omega, \Omega_0^t, \Omega_1^t, \Lambda_0^t, \Lambda_1^t), \tag{B.1}$$

$$\{\Omega_0^{t+1}, \Omega_1^{t+1}\} = \underset{\Omega_0=\Omega_0^T, \Omega_1\succeq\epsilon I}{\text{argmin}} L(\Omega^{t+1}, \Omega_0, \Omega_1, \Lambda_0^t, \Lambda_1^t), \tag{B.2}$$

$$\{\Lambda_0^{t+1}, \Lambda_1^{t+1}\} = \{\Lambda_0^t, \Lambda_1^t\} + \{\Omega^{t+1} - \Omega_0^{t+1}, \Omega^{t+1} - \Omega_1^{t+1}\}. \tag{B.3}$$

From (B.1), we write

$$\Omega^{t+1} = \frac{1}{4}\text{trace}(\Omega^2 A) + \frac{1}{4}\text{trace}(\Omega B \Omega B) - \text{trace}(\Omega C), \tag{B.4}$$

where $A = S + 4I$, $B = S^{1/2}$ and $C = I + \Omega_0^t + \Omega_1^t - \Lambda_0^t - \Lambda_1^t$. Let $A = UVU^T$ and $B = UWU^T$ are the eigendecomposition of matrices (it is easy to see that A and B have the same eigenvectors). Assume that eigenvalues are ordered: $V_1 \geq \cdots \geq V_p$ and $W_1 \geq \cdots \geq W_p$. It can be checked that the solution of (B.4) is given as $\hat{\Omega} = U\{(U^T BU) \circ C\}U^T$, where $\circ$ denotes the Hadamard product and $C_{ij} = \dfrac{4}{V_i + V_j + 2W_i W_j}$ for $1 \leq i, j \leq p$. Next, from (B.2), it follows that

$$\Omega_0^{t+1} = \underset{\Omega_0 = \Omega_0^T}{\operatorname{argmin}} \frac{1}{2}\operatorname{trace}({\Omega_0}^2) - \operatorname{trace}(\Omega_0(\Omega^{t+1} + \Lambda_0^t)) + \lambda ||\Omega_0||_{1,\text{off}}. \tag{B.5}$$

It is easy to check that the solution of (B.5) is given as $\Omega_0^{t+1} = ST(\Omega^{t+1} + \Lambda_0^t, \lambda)$, where ST is the soft-thresholding operator and is defined as $[ST(A, \lambda)]_{ij} = \operatorname{sign}(A_{ij}) \max(|A_{ij}| - \lambda, 0) I_{i \neq j} + A_{ij} I_{i=j}$.

Finally, from the second part of Eq. (B.2), it follows that

$$\Omega_1^{t+1} = \underset{\Omega_1 \succeq \epsilon I}{\operatorname{argmin}} \frac{1}{2}\operatorname{trace}(\Omega_1^2) - \operatorname{trace}(\Omega_1(\Omega^{t+1} + \Lambda_1^t)). \tag{B.6}$$

The solution of (B.6) is given as $\Omega_1^{t+1} = \left[\Omega^{t+1} + \Lambda_1^t\right]_+$, where for any symmetric matrix A with an eigendecomosition $A = U\operatorname{diag}(\alpha_1, \ldots, \alpha_p)U^T$ the operator $[A]_+$ is defined as

$$[A]_+ = U\operatorname{diag}(\max\{\alpha_1, \epsilon\}, \ldots, \max\{\alpha_p, \epsilon\})U^T$$

For the provided algorithm, we start with initial values at $t = 0$: $\Lambda_0^0 = \Lambda_1^0$, $\Omega_0^0 = \Omega_1^0$. We repeat the steps (B.1), (B.2), (B.3) until the following convergence conditions are satisfied:

$$\frac{||\Omega^{t+1} - \Omega^t||_2}{\max(1, ||\Omega^t||_2, ||\Omega^{t+1}||_2)} < 10^{-7}.$$

It is important to note that we can significantly reduce the computational time of the algorithm by discarding the constraint $\Omega \succeq \epsilon I$ in the initial optimization problem. This enables us to drop Ω_1 and omit the problem (B.6). If the reduced algorithm provides a positive definite outcome $\tilde{\Omega}$ (such that $\tilde{\Omega} \succeq \epsilon I$), then we stop the calculation and set $\hat{\Omega} = \tilde{\Omega}$. Otherwise, we repeat the complete algorithm provided earlier with an initial start $\hat{\Omega}$.

References

Avagyan V (2021) D-trace estimation of a precision matrix with eigenvalue control. Commun Stat Simul Comput 50(4):1231–1247

Avagyan V, Alonso AM, Nogales FJ (2018) D-trace estimation of a precision matrix using adaptive lasso penalties. Adv Data Anal Classif 12(2):425–447

Banerjee O, El Ghaoui L, d'Aspremont A, Natsoulis G (2006) Convex optimization techniques for fitting sparse Gaussian graphical models. In: Proceedings of the 23rd international conference on machine learning

Cai T, Liu W, Luo X (2011) A constrained ℓ_1 minimization approach to sparse precision matrix estimation. J Am Stat Assoc 106(494):594–607

Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over f1 score and accuracy in binary classification evaluation. BMC Genom 21(1):1–13

Dempster A (1972) Covariance selection. Biometrics 28(1):157–175

De Terán F, Iannazzo B (2016) Uniqueness of solution of a generalized Sylvester matrix equation. Linear Algebra Appl 493:323–335

Fan J, Feng J, Wu Y (2009) Network exploration via the adaptive LASSO and SCAD penalties. Ann Appl Stat 3(2):521–541

Fan J, Liao Y, Liu H (2016) An overview of the estimation of large covariance and precision matrices. Econom J 19(1):C1–C32

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. Biostatistics 9(3):432–441

Hsieh CJ, Sustik MA, Dhillon IS, Ravikumar P (2014) Quadratic approximation for sparse inverse covariance estimation. J Mach Learn Res 15:2911–2947

Kuerer HM, Newman LA, Smith TL, Ames FC, Hunt KK, Dhingra K, Theriault RL, Singh G, Binkley SM, Sneige N, Buchholz TA, Ross MI, McNeese MD, Buzdar AU, Hortobagyi GN, Singletary SE (1999) Clinical course of breast cancer patients with complete pathologic primary tumor and axillary lymph node response to doxorubicin-based neoadjuvant chemotherapy. J Clin Oncol 17(2):460–469

Kuismin MO, Sillanpää MJ (2017) Estimation of covariance and precision matrix, network structure, and a view toward systems biology. Wiley Interdiscip Rev Comput Stat 9(6):1–13

Kuismin M, Kemppainen J, Sillanpää M (2017) Precision matrix estimation with rope. J Comput Graph Stat 26(3):682–694

Lauritzen S (1996) Graphical models. Clarendon Press, Oxford

Li Z-Y, Wang Y, Zhou B, Duan G-R (2010) Least squares solution with the minimum-norm to general matrix equations via iteration. Appl Math Comput 215(10):3547–3562

Liu W, Luo X (2015) Fast and adaptive sparse precision matrix estimation in high dimensions. J Multivar Anal 135:153–162

Matthews BW (1975) Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochim Biophys Acta 405:442–451

Meinshausen N, Bühlmann P (2006) High-dimensional graphs and variable selection with the lasso. Ann Stat 34(2):1436–1462

Peng W, Wang P, Zhou N, Zhu J (2009) Partial correlation estimation by joint sparse regression models. J Am Stat Assoc 104(486):735–746

Ravikumar P, Wainwright M, Raskutti G, Yu B (2011) High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. Electron J Stat 5:935–980

Rothman A, Bickel P, Levina E, Zhu J (2008) Sparse permutation invariant covariance estimation. Electron J Stat 2:494–515

Ryali S, Chen T, Supekar K, Menon V (2012) Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. NeuroImage 59(4):3852–3861

Scheinberg K, Ma S, Goldfarb D (2010) Sparse inverse covariance selection via alternating linearization methods. In: Advances in neural information processing systems

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc 58(1):267–288

van Wieringen WN (2019) The generalized ridge estimator of the inverse covariance matrix. J Comput Graph Stat 28(4):932–942

van Wieringen WN, Peeters CF (2016) Ridge estimation of inverse covariance matrices from high-dimensional data. Comput Stat Data Anal 103:284–303

Wang C, Jiang B (2020) An efficient ADMM algorithm for high dimensional precision matrix estimation via penalized quadratic loss. Comput Stat Data Anal 142:106812

Yin J, Li J (2013) Adjusting for high-dimensional covariates in sparse precision matrix estimation by ℓ_1-penalization. J Multivar Anal 116:365–381

Yuan M (2010) High dimensional inverse covariance matrix estimation via linear programming. J Mach Learn Res 11:2261–2286

Yuan M, Lin Y (2007) Model selection and estimation in the Gaussian graphical model. Biometrika 94(1):19–35

Zhang T, Zou H (2014) Sparse precision matrix estimation via lasso penalized D-trace loss. Biometrika 88:1–18

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.