CROSS-VALIDATED BANDWIDTH SELECTION FOR PRECISION MATRIX ESTIMATION

Jun Tong, Jiangtao Xi, Yanguang Yu and Philip O. Ogunbona

Faculty of Engineering and Information Sciences
University of Wollongong, Wollongong, NSW 2522, Australia

ABSTRACT

Inverse covariance matrix, a.k.a. precision matrix, has wide applications in signal processing and is often estimated from training samples. The quality of estimation can be poor when the sample support is low. Banding/tapering are effective regularization approaches for covariance and precision matrix estimation but the bandwidth must be properly chosen. This paper investigates the bandwidth selection problem for banding/tapering-based precision matrix estimation. Exploiting a regression analysis interpretation of the precision matrix, we design a data-driven cross-validation (CV) method for automatically tuning the bandwidth. The effectiveness of the proposed method is demonstrated by numerical examples under a quadratic loss.

Index Terms— Banding, cross-validation, precision matrix, regression analysis, tapering

1. INTRODUCTION

Inverse covariance matrix, a.k.a. precision matrix, is used extensively in signal processing applications, such as filtering, beamforming, and correlation analysis [1, 2, 3]. In practice, precision matrix may be estimated from training samples. It is known that sample covariance matrix (SCM) is ill-conditioned and even singular when the number of training samples is not much larger than the dimensionality of the signal. In this case, a precision matrix constructed by directly inverting the SCM, referred to as sample precision matrix (SPM) below, may suffer from significant errors.

Regularization techniques, such as shrinkage [4, 5, 6], banding [7], and tapering [8], have been widely studied for covariance matrix estimation. Regularization generally imposes a priori assumptions on the structure of covariance matrix, and thus reduces the number of free parameters to be estimated. By properly tuning the regularization parameter for a good tradeoff between bias and variance, an improved covariance matrix estimate can be achieved, which can be subsequently inverted to produce a precision matrix estimate which improves SPM. Regularization can also be applied to the precision matrix itself for directly estimating the precision matrix from the training samples [9, 10, 11], which may outperform the approach based on the inversion of a regularized covariance matrix.

In order to optimize the performance of regularization-based precision matrix estimation, parameters, such as shrinkage factors [6] and bandwidth [7, 8], must be tuned properly. Data-driven methods that do not require a priori knowledge about the data distribution are often preferred [4, 7, 12, 13, 14]. However, most existing works focus on covariance matrix estimation. A regularization parameter optimized for covariance matrix estimation does not necessarily perform well for precision matrix estimation. This motivates the study of parameter tuning for optimizing precision matrix estimation. For shrinkage estimators, [15], [16] and [17] recently proposed solutions based on random matrix theory (RMT), but it is unclear how to extend their results to more general forms of regularization such as banding [7, 9] and tapering [8].

This paper introduces a simple method for choosing the bandwidth for precision matrix estimation based on banding/tapering. We follow the classical cross-validation (CV) principle [18, 19], which generally requires a proper choice of prediction error as the performance metric. Exploiting a regression analysis interpretation of the precision matrix, we propose an easy-to-compute, distribution-free metric for the CV. Numerical results show that the proposed technique can approach the oracle choice that minimizes a quardratic loss of the estimation.

2. CROSS-VALIDATED BANDWIDTH SELECTION

2.1. Precision matrix estimation

Consider an N-dimensional signal $\mathbf{y}$ with mean zero. Its covariance matrix is defined as $\mathbf{\Sigma} = \mathrm{E}\{\mathbf{y}\mathbf{y}^\dagger\}$, where $\mathrm{E}\{\cdot\}$ denotes expectation and † conjugate transpose. The precision matrix is defined as $\mathbf{\Omega} \triangleq \mathbf{\Sigma}^{-1}$. Both $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ have extensive applications in statistical signal processing and are often estimated from training samples. Suppose we have T training samples and let $\mathbf{y}_t$ be the t-th sample. The sample covariance matrix (SCM) is then computed as

$$\widehat{\mathbf{\Sigma}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t\mathbf{y}_t^\dagger. \tag{1}$$

If the SCM is nonsingular, the sample precision matrix (SPM) is computed as

$$\widehat{\mathbf{\Omega}} = \widehat{\mathbf{\Sigma}}^{-1}. \tag{2}$$

When the sample size T is not much larger than the dimensionality N, the SCM and SPM may suffer from significant errors.

Regularization techniques such as banding and tapering [7, 8, 9] have been suggested to improve the accuracy of covariance matrix estimation and may be generalized to precision matrix estimation in different ways. We take the tapering design as an example. With a bandwidth K, we can generate from SCM the following tapered covariance matrix estimate [7]

$$\widehat{\mathbf{\Sigma}}_K = \widehat{\mathbf{\Sigma}} \odot \mathbf{B}_K, \tag{3}$$

where $\odot$ denotes element-wise product and $\mathbf{B}_K$ is defined as [8]

$$[\mathbf{B}_K]_{i,j} = \begin{cases} 1, & \text{for } |i-j| \le K_h \\ 2 - \frac{i-j}{K_h}, & \text{for } K_h < |i-j| < K \\ 0, & \text{for } |i-j| \ge K \end{cases}, \tag{4}$$

where $K_h \triangleq K/2$. Note that the bandwidth K specifies the design.

A method for estimating the precision matrix is to directly invert the tapered covariance matrix estimate $\widehat{\mathbf{\Sigma}}_K$:

$$\widehat{\mathbf{\Omega}}_K^{(1)} = \widehat{\mathbf{\Sigma}}_K^{-1}. \tag{5}$$

The performance of the resulting precision matrix estimate depends critically on the bandwidth K. It is thus a fundamental issue to choose a proper bandwidth. Note that $\widehat{\mathbf{\Omega}}_K^{(1)}$ is generally not banded and the bandwidth K here actually refers to the bandwidth of the corresponding $\widehat{\mathbf{\Sigma}}_K$.

2.2. Automatic bandwidth selection

We now introduce a CV method that exploits a regression interpretation of the precision matrix. Let us partition the entries of the signal vector $\mathbf{y}$ as

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \mathbf{y}_{\sim 1} \end{bmatrix}, \tag{6}$$

where the lengths of y_1 and $\mathbf{y}_{\sim 1}$ are 1 and $N-1$, respectively. Accordingly, let us partition the covariance matrice of $\mathbf{y}$ as

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \boldsymbol{\sigma}_1^\dagger \\ \boldsymbol{\sigma}_1 & \mathbf{\Sigma}_{\sim 1} \end{bmatrix}, \tag{7}$$

where $\sigma_{11} \triangleq \mathrm{E}\{y_1 y_1^*\}$, $\boldsymbol{\sigma}_1 \triangleq \mathrm{E}\{\mathbf{y}_{\sim 1} y_1^*\}$, and $\mathbf{\Sigma}_{\sim 1} \triangleq \mathrm{E}\{\mathbf{y}_{\sim 1}\mathbf{y}_{\sim 1}^\dagger\}$. The precision matrix is then computed as

$$\mathbf{\Omega} \triangleq \mathbf{\Sigma}^{-1} = \begin{bmatrix} \omega_{11} & \boldsymbol{\omega}_1^\dagger \\ \boldsymbol{\omega}_1 & \mathbf{\Omega}_{\sim 1} \end{bmatrix}. \tag{8}$$

From the matrix inversion lemma, it can be shown that

$$\boldsymbol{\omega}_1 = -\frac{1}{\sigma_{11} - \boldsymbol{\sigma}_1^\dagger \mathbf{\Sigma}_{\sim 1}^{-1} \boldsymbol{\sigma}_1} \mathbf{\Sigma}_{\sim 1}^{-1} \boldsymbol{\sigma}_1, \tag{9}$$

$$\omega_{11} = \frac{1}{\sigma_{11} - \boldsymbol{\sigma}_1^\dagger \mathbf{\Sigma}_{\sim 1}^{-1} \boldsymbol{\sigma}_1}. \tag{10}$$

Therefore, from ω_{11} and $\boldsymbol{\omega}_1$ of the precision matrix $\mathbf{\Omega}$, we can construct an $(N-1) \times 1$ vector

$$\mathbf{w}_1 \triangleq -\frac{1}{\omega_{11}} \boldsymbol{\omega}_1. \tag{11}$$

It can be easily seen that

$$\mathbf{w}_1 = \mathbf{\Sigma}_{\sim 1}^{-1} \boldsymbol{\sigma}_1 \tag{12}$$

gives the coefficients for regressing y_1 on $\mathbf{y}_{\sim 1}$ and is the linear minimum mean squared error (LMMSE) estimator for estimating y_1 from $\mathbf{y}_{\sim 1}$.

The above interpretation links the precision matrix to regression analysis of the data. This has been exploited for deriving regularized precision matrix estimates. Given the training data, one can estimate $\mathbf{\Omega}$ by conducting a regression analysis of the training samples. Constraints on the regression coefficients can be imposed to obtain different regularized estimators [11, 20]. In this work, we exploit the above regression interpretation for determining the optimal bandwidth for the regularized precision matrix estimation. The rationale is that, if we have a good estimate $\widehat{\mathbf{\Omega}}$ of the true precision matrix $\mathbf{\Omega}$, then from it a linear predictor constructed as (12) (with $\mathbf{\Omega}$ replaced by $\widehat{\mathbf{\Omega}}$) should lead to a small error ξ_1 of predicting y_1 from $\mathbf{y}_{\sim 1}$, where

$$\xi_1 \triangleq y_1 - \mathbf{w}_1^\dagger \mathbf{y}_{\sim 1} \tag{13}$$

is obtainable from the training data and the first column of the estimate $\widehat{\mathbf{\Omega}}$ of the precision matrix.

We can generalize the above regression analysis for other entries of $\mathbf{y}$. For an arbitrary n, it can be shown that the estimator for estimating y_n from all the other entries can be found from the n-th column of $\mathbf{\Omega}$ as

$$\mathbf{w}_n = -\frac{1}{\omega_{nn}} \boldsymbol{\omega}_n, \tag{14}$$

where $\boldsymbol{\omega}_n$ denotes the n-th column of $\mathbf{\Omega}$ with its n-th entry excluded. The corresponding estimation error is then computed as

$$\xi_n = y_n - \mathbf{w}_n^\dagger \mathbf{y}_{\sim n}, \quad 1 \le n \le N. \tag{15}$$

We propose to use the above estimation errors to construct a performance metric for choosing the bandwidth K in a CV manner. In the time domain, the total training data $\mathbf{Y}$ is split into two disjoint subsets, i.e., $\mathbf{Y}^{(\sim i)}$ and $\mathbf{Y}^{(i)}$. The training subset $\mathbf{Y}^{(\sim i)}$ is used for constructing the predictors $\{\mathbf{w}_n\}$ in (14) from a regularized precision matrix estimate $\widehat{\mathbf{\Omega}}_K^{(i)}$ with

bandwidth K. The validation subset $\mathbf{Y}^{(i)}$ is used for evaluating the quality of precision matrix estimation using the estimation error $\{\xi_n\}$ in (15). The total estimation error will then be used for choosing the best bandwidth as

$$K^* = \arg\min_K J(K), \tag{16}$$

where

$$J(K) = \sum_{i=1}^{I}\sum_{n=1}^{N}\left\|\mathbf{y}_n^{(i)} - \mathbf{w}_{K,n}^{(i)\dagger}\mathbf{Y}_{\sim n}^{(i)}\right\|_F^2, \tag{17}$$

$\mathbf{y}_n^{(i)}$ denotes the n-th row of $\mathbf{Y}^{(i)}$, $\mathbf{Y}_{\sim n}^{(i)}$ corresponds to the entries of $\mathbf{Y}^{(i)}$ for predicting $\mathbf{y}_n^{(i)}$, $\mathbf{w}_{K,n}^{(i)}$ denotes the estimator for estimating the n-th entry of $\mathbf{y}$ constructed using the training subset and bandwidth K, and $\|\cdot\|_F$ denotes the Frobenius norm. In (17), we have assumed that the training data $\mathbf{Y}$ is split into $(\mathbf{Y}^{(\sim i)}, \mathbf{Y}^{(i)})$ for I times. Summarizing, the CV cost is given by

$$J_1(K) = \sum_{i=1}^{I}\sum_{n=1}^{N}\left\|\mathbf{y}_n^{(i)} + \frac{1}{\widehat{\omega}_{K,nn}^{(i)}}\widehat{\boldsymbol{\omega}}_{K,n}^{(i)\dagger}\mathbf{Y}_{\sim n}^{(i)}\right\|_F^2. \tag{18}$$

A grid search of K can be conducted to choose the minimizer of $J(K)$ as the optimal bandwidth. It can be shown that the performance metric (17) can be rewritten as

$$\boxed{J_1(K) = \sum_{i=1}^{I}\left\|[\mathbf{D}_{\widehat{\mathbf{\Omega}}_K^{(i)}}]^{-1}\widehat{\mathbf{\Omega}}_K^{(i)}\mathbf{Y}^{(i)}\right\|_F^2}, \tag{19}$$

where $\mathbf{D}_{\widehat{\mathbf{\Omega}}_K^{(i)}}$ denotes the diagonal matrix whose diagonal entries are the same as those of $\widehat{\mathbf{\Omega}}_K^{(i)}$.

The above bandwidth selection method is based on the regression analysis of the original signal $\mathbf{y}$. Alternatively, we can consider a treatment similar to generalized cross validation (GCV) [19]. Instead of conducting the regression analysis of the entries of $\mathbf{y}$, we can consider the regression analysis of the linearly transformed signal

$$\mathbf{y}' = \mathbf{V}^\dagger\mathbf{y}, \tag{20}$$

where $\mathbf{V} = \mathbf{UF}$, with $\mathbf{F}$ being the discrete Fourier transform matrix and $\mathbf{U}$ the eigenvector matrix of the covariance matrix $\mathbf{\Sigma}$. In this case, the precision matrix of $\mathbf{y}'$ is given by

$$\mathbf{\Omega}' = \mathbf{V}^\dagger\mathbf{\Omega}\mathbf{V} = \mathbf{F}^\dagger\mathbf{U}^\dagger\mathbf{\Omega}\mathbf{U}\mathbf{F}. \tag{21}$$

This is a circulant matrix with equal diagonal entries given by $\frac{1}{N}\mathrm{tr}(\mathbf{\Omega})$, i.e.,

$$\mathbf{D}_{\mathbf{\Omega}'} = \frac{1}{N}\mathrm{tr}(\mathbf{\Omega})\mathbf{I}. \tag{22}$$

Note also that for an arbitrary $\mathbf{y}$

$$||\mathbf{\Omega}'\mathbf{y}'||_F^2 = ||\mathbf{F}^\dagger\mathbf{U}^\dagger\mathbf{\Omega}\mathbf{U}\mathbf{F}\mathbf{F}^\dagger\mathbf{U}^\dagger\mathbf{y}||_F^2 = ||\mathbf{\Omega}\mathbf{y}||_F^2. \tag{23}$$

By replacing the true precision matrix as its estimate, the CV cost function of (19) applied to the transformed signal (20) can then be written as

$$\boxed{J_2(K) = N^2\sum_{i=1}^{I}\frac{\left\|\widehat{\mathbf{\Omega}}_K^{(i)}\mathbf{Y}^{(i)}\right\|_F^2}{(\mathrm{Tr}(\widehat{\mathbf{\Omega}}_K^{(i)}))^2}}. \tag{24}$$

Note that the result in (24) does not explicitly require the calculation of (20). In other words, (20) only serves as a proxy for deriving the generalized CV expression. We observe that (19) and (24) lead to similar performance of bandwidth selection.

3. NUMERICAL EXAMPLES

In order to better demonstrate the effectiveness of the proposed bandwidth selection method, let us consider one more precision matrix estimator which exploits the regression interpretation in (9)-(12) for directly producing a banded precision matrix estimate. With a bandwidth K, we can first generate from the SCM $\widehat{\mathbf{\Sigma}}$ a banded covariance matrix estimate $\mathbf{R}$ as [7] $\mathbf{R} = \widehat{\mathbf{\Sigma}} \odot \mathbf{B}_K$, where $\mathbf{B}_K$ is defined as

$$[\mathbf{B}_K]_{i,j} = \begin{cases} 1, & |i-j| \le K \\ 0, & |i-j| > K \end{cases}. \tag{25}$$

For each n, let $i_{\min} = \max(1, n-K)$, $i_{\max} = \min(N, n+K)$, $\mathbf{R}_{\sim n}$ be a submatrix consisting of rows $[i_{\min}, \cdots, i_{\max}]$ and columns $[i_{\min}, \cdots, i_{\max}]$ of $\mathbf{R}$ with the n-th row and n-th column excluded; $\mathbf{r}_n$ consisting of entries $[i_{\min}, \cdots, i_{\max}]$ of the n-th column of $\mathbf{R}$ with its n-th entry excluded; r_{nn} the (n,n)-th entry of $\mathbf{R}$. Let $\mathbf{w}_n = \mathbf{R}_{\sim n}^{-1}\mathbf{r}_n$. Then according to (9)-(12), the n-th diagonal entry of the precision matrix $\mathbf{\Omega}$ can be estimated by $\widehat{\omega}_{nn} = \frac{1}{r_{nn} - \mathbf{w}_n^\dagger\mathbf{r}_n}$. The remaining nonzero entries of the n-th column (excluding the diagonal entry) of $\mathbf{\Omega}$ is then set as the corresponding entries of $-\widehat{\omega}_{nn}\mathbf{w}_n$. Repeating this process for $n \in \{1, 2, \cdots, N\}$ will produce a banded precision matrix estimate $\widehat{\mathbf{\Omega}}$, which is generally not Hermitian. In order to produce a Hermitian precision matrix estimate, we set

$$\widehat{\mathbf{\Omega}}_K^{(2)} = \frac{1}{2}\left(\widehat{\mathbf{\Omega}} + \widehat{\mathbf{\Omega}}^\dagger\right). \tag{26}$$

We now present numerical examples of applying the proposed CV method to choose the bandwidth for the precision matrix estimates. An autoregressive (AR) model [7] is first assumed for the true covariance matrix $\mathbf{\Sigma}$ of $\mathbf{y}$, with its (i,j)-th entry given by

$$[\mathbf{\Sigma}]_{i,j} = \rho^{|i-j|}, \forall i, j, \tag{27}$$

where ρ is a constant. We assume zero-mean, Gaussian data but our methods do not rely on knowledge about the distribution. We use the normalized Frobenius norm of the estimation

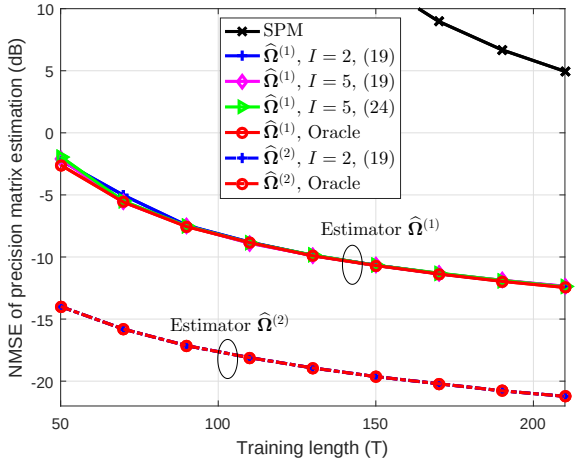

Fig. 1. NMSE of precision matrix estimates using the proposed CV bandwidth selection for different training lengths. The AR covariance matrix with $N = 100$ and $\rho = 0.7$ is assumed. The maximum bandwidth considered is $K_{max} = 20$.

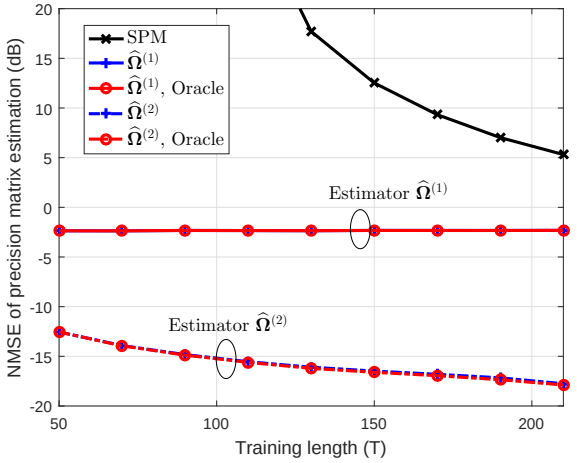

Fig. 2. NMSE of precision matrix estimates using the proposed CV bandwidth selection for different training lengths with $I = 2$ and the cost in (19). The FGN covariance matrix with $N = 100$ and $H = 0.9$ is assumed. The maximum bandwidth considered is $K_{max} = 20$.

error

$$\mathrm{L}(\widehat{\mathbf{\Omega}}) = \frac{||\widehat{\mathbf{\Omega}} - \mathbf{\Omega}||_F^2}{\|\mathbf{\Omega}\|_F^2} \tag{28}$$

to measure the accuracy of precision matrix estimation and define its average as the normalized MSE (NMSE). Fig. 1 demonstrates the performance of proposed CV method when applied to the precision matrix estimators $\widehat{\mathbf{\Omega}}_K^{(1)}$ and $\widehat{\mathbf{\Omega}}_K^{(2)}$ in (5) and (26), respectively. The results marked by "oracle" apply the bandwidths that minimize the Frobenius norm loss of (28), which can be obtained only when the true precision matrix is known. The "oracle" results are used to benchmark the performance of our proposed CV method. It can be seen that both the proposed precision matrix estimators significantly outperform the SPM, especially when the number of samples T is smaller than the dimension N. The estimator based on covariance matrix tapering, i.e., $\widehat{\mathbf{\Omega}}_K^{(1)}$, is less effective than the regression analysis-based estimator $\widehat{\mathbf{\Omega}}_K^{(2)}$. It can be seen that the proposed CV method with different implementations all achieve near-oracle choice of the bandwidth under the loss in (28).

We also test on the Fractional Gaussian noise (FGN) model considered in [7]:

$$[\mathbf{\Sigma}]_{ij} = \begin{cases} 1, & i = j \\ \frac{1}{2}\left[(|i-j|+1)^{2H} - 2|i-j|^{2H} + (|i-j|-1)^{2H}\right], & i \neq j \end{cases} \tag{29}$$

Fig. 2 shows the results of precision matrix estimation for $N = 100$ and $H = 0.9$. Note that this is an ill-conditioned case where the covariance matrix has a condition number of 222. The tapering-based design does not work well because the entries of $\mathbf{\Sigma}$ decays very slowly off diagonals. However, our proposed designs are still able to achieve near-optimal bandwidth tuning for the two precision matrix estimators considered.

4. CONCLUSIONS

This paper introduced a cross-validation method based on regression-analysis of the precision matrix for determining the bandwidth for regularized precision matrix estimation. This method is distribution-free, easy to implement and can approach the oracle bandwidth selection under a quadratic loss. Its effectiveness is evaluated with different structures of covariance matrix.

5. ACKNOWLEDGMENT

This work was supported in part by NSFC under Grant 61601325.

6. REFERENCES

[1] L. L. Scharf, *Statistical signal processing*. Addison-Wesley Reading, MA, 1991.

[2] Y. Song, P. J. Schreier, D. Ramírez, and T. Hasija, "Canonical correlation analysis of high-dimensional data with very small sample support," *Signal Processing*, vol. 128, pp. 449–458, 2016.

[3] X. Mestre and M. Á. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 69–82, 2006.

[4] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365 – 411, 2004.

[5] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for mmse covariance estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5016–5029, 2010.

[6] X. Chen, Z. J. Wang, and M. J. McKeown, "Shrinkage-to-tapering estimation of large covariance matrices," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5640–5656, Nov 2012.

[7] P. J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *The Annals of Statistics*, vol. 36, no. 1, pp. 199–227, 2008.

[8] T. T. Cai, C.-H. Zhang, H. H. Zhou, *et al.*, "Optimal rates of convergence for covariance matrix estimation," *The Annals of Statistics*, vol. 38, no. 4, pp. 2118–2144, 2010.

[9] W. B. Wu and M. Pourahmadi, "Nonparametric estimation of large covariance matrices of longitudinal data," *Biometrika*, pp. 831–844, 2003.

[10] N. Meinshausen and P. Buhlmann, "High-dimensional graphs and variable selection with the lasso," *The Annals of Statistics*, vol. 34, no. 3, pp. 1436–1462, 2006.

[11] J. Z. Huang, N. Liu, M. Pourahmadi, and L. Liu, "Covariance matrix selection and estimation via penalised normal likelihood," *Biometrika*, vol. 93, no. 1, pp. 85–98, 2006.

[12] F. Yi and H. Zou, "Sure-tuned tapering estimation of large covariance matrices," *Computational Statistics & Data Analysis*, vol. 58, pp. 339–351, 2013.

[13] Y. Qiu and S. X. Chen, "Bandwidth selection for high-dimensional covariance matrix estimation," *Journal of the American Statistical Association*, vol. 110, no. 511, pp. 1160–1174, 2015.

[14] J. Tong, P. J. Schreier, Q. Guo, S. Tong, J. Xi, and Y. Yu, "Shrinkage of covariance matrices for linear signal estimation using cross-validation," *IEEE Transactions on Signal Processing*, vol. 64, pp. 2965–2975, June 2016.

[15] T. Bodnar, A. K. Gupta, and N. Parolya, "Direct shrinkage estimation of large dimensional precision matrix," *Journal of Multivariate Analysis*, vol. 146, pp. 223–236, 2016.

[16] M. Zhang, F. Rubio, and D. P. Palomar, "Improved calibration of high-dimensional precision matrices," *IEEE Transactions on Signal Processing*, vol. 61, no. 6, pp. 1509–1519, 2013.

[17] C. Wang, G. Pan, T. Tong, and L. Zhu, "Shrinkage estimation of large dimensional precision matrix using random matrix theory," *Statistica Sinica*, vol. 25, no. 3, pp. 993–1008, 2015.

[18] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, 2010.

[19] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[20] M. Pourahmadi, "Covariance estimation: The glm and regularization perspectives," *Statistical Science*, vol. 26, no. 3, pp. 369–387, 2011.