Time-Series Analysis on Multiperiodic Conditional Correlation by Sparse Covariance Selection and Its Computational Method

Tokyo Institute of Technology Michael Lie
Tokyo Institute of Technology Suzuki Taiji

1 Introduction

In this research, we consider optimization problems where the objective is a negative Gaussian log-likelihood with two penalty terms. This minimization problem estimates the sparse inverse of its input covariance matrix; sparse covariance selection [Dem72]. The first penalty of ℓ_1-norm regularization term ensures a sparse solution, i.e., one with few nonzero entries, while the second penalty enhances block partitions in the parameter space. One of the well-known application of this estimation problem is Markowitz's portfolio selection [Mar52], where the inverse of covariance matrix is needed as its input. We propose to apply (a) Alternating Direction Method of Multipliers (ADMM) algorithm [Boy+11] and (b) Quadratic Approximation Method [Hsi+11] for solving the problem of sparse covariance selection.

2 Problem Setup

Given a set of i.i.d. data of $\mathbf{Y}_t = \{\mathbf{y}_{1,t}, \cdots, \mathbf{y}_{n,t}\}, t \in \{1, \cdots, T\}$ drawn from p-variate Gaussian distribution of $\mathcal{N}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let the empirical covariance matrix between $\mathbf{Y}_i$ and $\mathbf{Y}_j$ is given by

$$\mathbf{S}_{ij} := \frac{1}{n} \sum_{k,l} (\mathbf{y}_{k,i} - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_{l,j} - \hat{\boldsymbol{\mu}}_j)^\top,$$

where $\hat{\boldsymbol{\mu}}_i = \frac{1}{n}\sum_k \mathbf{y}_{k,i}$. Consider a stationary-time process such that the multiperiodic inverse covariance matrix $\mathbf{X}$ can be expressed as

$$\mathbf{X} = \underbrace{\begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \mathbf{X}_{13} & \cdots & \mathbf{X}_{1,T} \\ \mathbf{X}_{12}^\top & \mathbf{X}_{22} & \mathbf{X}_{23} & \cdots & \mathbf{X}_{2,T} \\ \mathbf{X}_{13}^\top & \mathbf{X}_{23}^\top & \mathbf{X}_{33} & \cdots & \mathbf{X}_{3,T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{1,T}^\top & \mathbf{X}_{2,T}^\top & \mathbf{X}_{3,T}^\top & \cdots & \mathbf{X}_{T,T} \end{bmatrix}}_{Tp \text{ columns}} \left.\vphantom{\begin{bmatrix} a\\a\\a\\a\\a \end{bmatrix}}\right\} Tp \text{ rows}.$$

We assume that $\mathbf{X}$ is a stationary time-process, such that $\mathbf{X}_{i,i+h} = \mathbf{X}_{j,j+h}$ for all i, j. Given some regularization penalty $\lambda_1 > 0, \lambda_2 > 0$, the ℓ_1 and ℓ_2 regularized Gaussian MLE for the inverse covariance matrix can be estimated by solving the following regularized *log-determinant* program:

$$\begin{aligned} &\underset{\mathbf{X} \succ 0}{\operatorname{argmin}} \Big\{ -\ln\det \mathbf{X} + \sum_{i,j} \operatorname{trace}\left(\mathbf{S}_{ij}^\top \mathbf{X}_{ij}\right) + \lambda_1 \sum_{i,j} \|\mathbf{X}_{ij}\|_1 \\ &\quad + \lambda_2 \sum_{i,j} \sum_{k>i,l>j} \|\mathbf{X}_{ij} - \mathbf{X}_{kl}\|_2^2 \Big\} =: \underset{\mathbf{X}_{ij} \succ 0}{\operatorname{argmin}} f(\mathbf{X}) \\ &\text{subject to} \quad \mathbf{X}_{i,i+h} = \mathbf{X}_{j,j+h}, \forall i, j. \end{aligned} \tag{2.1}$$

3 Proposed Algorithm

(a) ADMM algorithm

Firstly, the objective function f comprises of two parts, $f(\mathbf{X}) \equiv g(\mathbf{X}) + h(\mathbf{X})$, where

$$g(\mathbf{X}) = -\ln\det \mathbf{X} + \sum_{i,j} \operatorname{trace}\left(\mathbf{S}_{ij}^\top \mathbf{X}_{ij}\right),$$

$$h(\mathbf{X}) = \lambda_1 \sum_{i,j} \|\mathbf{X}_{ij}\|_1 + \lambda_2 \sum_{i,j} \sum_{k>i,l>j} \|\mathbf{X}_{ij} - \mathbf{X}_{kl}\|_F^2 .$$

The first component $g(\mathbf{X})$ is twice differentiable and strictly convex, while the second part $h(\mathbf{X})$ is convex but non-differentiable.

Then, we can write Equation (2.1) into the following optimization problem.

$$\begin{aligned} &\text{minimize} \quad g(\mathbf{X}) + h(\mathbf{Z}) \\ &\text{subject to} \quad \tilde{\mathbf{X}} = \tilde{\mathbf{Z}} \end{aligned} \tag{3.1}$$

where

$$h(\tilde{\mathbf{Z}}) = \lambda_1 \sum_{i,j} \|\mathbf{Z}_1\|_1 + \lambda_2 \sum_{i,j} \|\mathbf{Z}_2\|_F^2 ,$$

$$\tilde{\mathbf{X}} = \left[(\mathbf{X}')^\top, (\mathbf{D}\mathbf{X}')^\top, (\mathbf{H}\mathbf{X}')^\top\right]^\top, \tilde{\mathbf{Z}} = \left[(\mathbf{Z}_1)^\top, (\mathbf{Z}_2)^\top, (\mathbf{0})^\top\right]^\top$$

for appropriate matrices $\mathbf{D}$ and $\mathbf{H}$ that represent *set of time difference matrices* and *set of stationary time difference matrices*, respectively. The augmented Lagrangian of Equation (3.1) is

$$L_\rho(\tilde{\mathbf{X}}, \tilde{\mathbf{Z}}, \mathbf{Y}) = g(\mathbf{X}) + h(\tilde{\mathbf{Z}}) + (\rho/2) \left\| \tilde{\mathbf{X}} - \tilde{\mathbf{Z}} + \frac{\mathbf{Y}}{\rho} \right\|_F^2 .$$

Therefore, the iteration of ADMM is given as:

$\tilde{\mathbf{X}}$-update:

$$\tilde{\mathbf{X}}^{(k)} := \underset{\tilde{\mathbf{X}}}{\operatorname{argmin}} \left(g(\mathbf{X}) + \frac{\rho}{2} \left\| \tilde{\mathbf{X}} - \tilde{\mathbf{Z}}^{(k)} + \frac{\mathbf{Y}^{(k)}}{\rho} \right\|_F^2 \right),$$

$\tilde{\mathbf{Z}}$-update:

$$\tilde{\mathbf{Z}}^{(k)} := \underset{\tilde{\mathbf{Z}}}{\operatorname{argmin}} \left(h(\tilde{\mathbf{Z}}) + \frac{\rho}{2} \left\| \tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{Z}} + \frac{\mathbf{Y}^{(k)}}{\rho} \right\|_F^2 \right),$$

$\tilde{\mathbf{Y}}$-update:

$$\mathbf{Y}^{(k+1)} := \mathbf{Y}^{(k)} + \rho \left(\tilde{\mathbf{X}}^{(k+1)} - \tilde{\mathbf{Z}}^{(k+1)} \right).$$

(b) Quadratic Approximation

The summary of the algorithm is given as follows:

Input: Empirical covariance matrices $\mathbf{S}$, time difference T, scalar values of λ_1, λ_2, inner stopping tolerance ϵ

Output: Minimizer $\mathbf{X}$ of $f(\mathbf{X})$.

1: **for** $k = 0, 1, \cdots$ **do**
2: Compute $\mathbf{W}^{(t)} = \left(\mathbf{X}^{(t)}\right)^{-1}$.
3: Form the second order approximation of $f(\mathbf{X}^{(t)} + \boldsymbol{\Delta})$.
4: Partition the variables into free and fixed sets based on the gradient.
5: Use coordinate descent to find the Newton direction $\mathbf{D}^{(t)}$ over the free variable set.
6: Use an Armijo-rule based step-size selection to get α.
7: Update $\mathbf{X}^{(t+1)} = \mathbf{X}^{(t)} + \alpha \mathbf{D}^{(t)}$.
8: **end for**

4 Conclusion

Both algorithms show consistency on the optimal values and as we can see from the following table that Quadratic Approximation method runs much faster compared to the ADMM algorithm.

T	$n = 20$		$n = 30$	
	(a)	(b)	(a)	(b)
1	179.65	0.75	283	2.11
2	418.25	1.9	739.43	4.22
3	876.94	5.71	1725.67	12.84

Table 1: Runtime (in seconds) for both algorithms.

References

[Boy+11] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. "Distributed optimization and statistical learning via the alternating direction method of multipliers." In: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122.

[Dem72] A. P. Dempster. "Covariance selection." In: *Biometrics* (1972), pp. 157–175.

[Hsi+11] C.-J. Hsieh, I. S. Dhillon, P. K. Ravikumar, and M. A. Sustik. "Sparse inverse covariance matrix estimation using quadratic approximation." In: *Advances in Neural Information Processing Systems*. 2011, pp. 2330–2338.

[Mar52] H. Markowitz. "Portfolio selection." In: *The Journal of Finance* 7.1 (1952), pp. 77–91.