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 $\ell_1$-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 $Y_t = \{y_{1,t}, \ldots, y_{n,t}\}, t \in \{1, \ldots, T\}$ drawn from $p$-variate Gaussian distribution of $N(y; \mu, \Sigma)$. Let the empirical covariance matrix between $Y_{ij}$ and $Y_{ij}$ be given by

$$S_{ij} := \frac{1}{n} \sum_{t=1}^{T} (y_{ij,t} - \hat{\mu}_i)(y_{ij,t} - \hat{\mu}_j)^\top,$$

where $\hat{\mu}_i = \frac{1}{n} \sum_{t=1}^{T} y_{i,t}$. Consider a stationary-time process such that the multiperiodic inverse covariance matrix $X$ can be expressed as

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1T} \\ X_{21} & X_{22} & \cdots & X_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ X_{T1} & X_{T2} & \cdots & X_{TT} \end{pmatrix},$$

where $X_{ij}$ has $T_p$ columns.

We assume that $X$ is a stationary time-process, such that $X_{i,j+h} = X_{j,i+h}$ for all $i,j$. Given some regularization penalty $\lambda_1 > 0, \lambda_2 > 0$, the $\ell_1$ and $\ell_2$ regularized Gaussian MLE for the inverse covariance matrix can be estimated by solving the following regularized log-determinant program:

$$\arg\min_{X>0} \left\{ -\log \det X + \sum_{i,j} \text{tr}(S_{ij}X_{ij}) + \lambda_1 \sum_{i,j} \|X_{ij}\|_1 \right\},$$

subject to $\|X_{i,j+h} - X_{j,i+h}\|_2^2 = 0$ for all $i,j$. (2.1)

3 Proposed Algorithm

(a) ADMM algorithm

First, the objective function $f$ comprises of two parts, $f(X) \equiv g(X) + h(X)$, where

$$g(X) = -\log \det X + \sum_{i,j} \text{tr}(S_{ij}X_{ij}),$$

$$h(X) = \lambda_1 \sum_{i,j} \|X_{ij}\|_1 + \lambda_2 \sum_{i,j} \sum_{k,i,j,l,j} \|X_{ij} - X_{kl}\|_2^2.$$

The first component $g(X)$ is twice differentiable and strictly convex, while the second part $h(X)$ is convex but non-differentiable. Then, we can write Equation (2.1) into the following optimization problem.

$$\begin{align*}
\text{minimize} & \quad g(X) + h(Z) \\
\text{subject to} & \quad \tilde{X} = Z,
\end{align*}$$

where $h(Z) = \lambda_1 \sum_{i,j} \|Z_{ij}\|_1 + \lambda_2 \sum_{i,j} \|Z_{ij}\|_2^2$.

The summary of the algorithm is given as follows:

**Input:** Empirical covariance matrices $S$, time difference $T$, scalar values of $\lambda_1, \lambda_2$, inner stopping tolerance $\epsilon$

**Output:** Minimizer $X$ of $f(X)$.

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

<table>
<thead>
<tr>
<th>$T$</th>
<th>$n = 20$</th>
<th>$n = 30$</th>
<th>$n = 40$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>223.05</td>
<td>223.05</td>
<td>223.05</td>
</tr>
<tr>
<td>2</td>
<td>418.25</td>
<td>418.25</td>
<td>418.25</td>
</tr>
<tr>
<td>3</td>
<td>876.94</td>
<td>876.94</td>
<td>876.94</td>
</tr>
</tbody>
</table>

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