Sparse Gaussian Graphical Model Estimation Using $\ell_1$ Regularization

Ming Yuan

School of Industrial and Systems Engineering
Georgia Institute of Technology
myuan@isye.gatech.edu
http://www.isye.gatech.edu/~myuan
OUTLINE

• Introduction
• Methodology
• Asymptotic Properties
• Computation
• Numerical Examples
BACKGROUND
**UNDIRECTED GRAPHICAL MODEL**

- $X_V$ is represented by an undirected graph $G(V, E)$
  - $V = \{1, 2, 3, 4, 5, 6\}$ contains vertices corresponding to the random variables
  - the edges $E = \{(1, 2), (1, 3), \ldots, (5, 6)\}$

- Factorization of probability distribution

  $$p(x_V) = \frac{1}{Z} \psi_{12}(x_1, x_2) \psi_{13}(x_1, x_3) \psi_{24}(x_2, x_4) \psi_{25}(x_2, x_5) \psi_{26}(x_2, x_6) \psi_{35}(x_3, x_5) \psi_{56}(x_5, x_6)$$

- Conditional independence

  $$X_2 \perp X_3 | X_1, X_4, X_5, X_6$$
Gaussian Graphical Model

\[ X = (X^{(1)}, \ldots, X^{(p)}) \sim \mathcal{N}_p(\mu, \Sigma) \]

- Concentration matrix \( C = \Sigma^{-1} \)

\[
p(x_Y) = (2\pi)^{-p/2} |C|^{1/2} \exp \left\{ - \sum_{(i,j): c_{ij} \neq 0} c_{ij}(x_i - \mu_i)(x_j - \mu_j)/2 \right\}
\]

\[
= \frac{(2\pi)^{-p/2} |C|^{1/2}}{1/Z} \prod_{(i,j): c_{ij} \neq 0} \exp \left\{ -c_{ij}(x_i - \mu_i)(x_j - \mu_j)/2 \right\}
\]

- Graphical Model – \( E = \{(i, j) : c_{ij} \neq 0\} \):

\[
c_{ij} = 0 \implies X_i \perp X_j | X-\{i,j\}
\]

- Goal
  - Covariance Selection – estimate the graphical structure \( E \)
  - Covariance Estimation – estimate \( \Sigma \) – PCA, LDA and etc.
Time Series

- Markov Property

\[ C = \begin{bmatrix}
  c_{11} & c_{12} & 0 & \cdots & \cdots & 0 \\
  c_{21} & c_{22} & c_{23} & \ddots & \ddots & \vdots \\
  0 & c_{32} & c_{33} & c_{34} & \ddots & \vdots \\
  \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
  \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
  0 & \cdots & \cdots & c_{p-1,p-2} & c_{p-1,p-1} & c_{p-1,p} \\
  0 & \cdots & \cdots & 0 & c_{p,p-1} & c_{pp}
\end{bmatrix} \]
IMAGES

- Markov random fields

<table>
<thead>
<tr>
<th>$X^{(1)}$</th>
<th>$X^{(2)}$</th>
<th>$X^{(3)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(4)}$</td>
<td>$X^{(5)}$</td>
<td>$X^{(6)}$</td>
</tr>
<tr>
<td>$X^{(7)}$</td>
<td>$X^{(8)}$</td>
<td>$X^{(9)}$</td>
</tr>
</tbody>
</table>
GENE NETWORK AND PATHWAY

(a) Arabidopsis

(b) Gene Expression of 835 Genes

Isoprenoids ← Five-carbon intermediates (IPP, DMAPP) ← MEP Pathway
MVA Pathway
HOW TO DO IT

- Classical approach (Lauritzen, 1996; Whittaker, 1990)
  - Greedy stepwise forward-selection or backward-deletion
  - the edge selection or deletion through hypothesis testing
  - computational complexity
  - does not correctly adjust for the multiple comparisons

- Recent Advances
  - Drton and Perlman (2004)
    - produces conservative simultaneous $1 - \alpha$ confidence intervals
    - uses these confidence intervals to do model selection in a single step
    - Lasso for each node to identify its neighbours
    - combine the results to learn the structure of a Gaussian concentration graph model
  - Bayesian approaches: Liechty, Liechty and Müller (2004); Dobra and West (2005)
$\ell_1$ REGULARIZATION
**CONstrained MAXIMUM LIKELIHOOD ESTIMATE**

\[ C \equiv \Sigma^{-1} \Rightarrow e_{ij} = I(c_{ij} \neq 0) \]

- To find a sparse graph

\[
\min_{C > 0} \left[ -\ln |C| + \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)' C (X_i - \mu) \right] \text{ subject to } \sum_{i \neq j} |e_{ij}| \leq t
\]

- Lasso-type Estimate – Graphical Lasso

\[
\min_{C > 0} \left[ -\ln |C| + \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)' C (X_i - \mu) \right] \text{ subject to } \sum_{i \neq j} |c_{ij}| \leq t
\]

- Graphical Garrote – \( \tilde{C} \) a preliminary estimate \( \Rightarrow c_{ij} = d_{ij} \tilde{c}_{ij} \)

\[
\min_{C > 0} \left[ -\ln |C| + \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)' C (X_i - \mu) \right] \text{ subject to } \sum_{i \neq j} d_{ij} \leq t, d_{ij} \geq 0,
\]


**RELATIONSHIP TO SEMI-DEFINITE PROGRAMMING**

MaxDet Problem

$$\min_{x \in \mathbb{R}^m} \quad b'x - \ln |G(x)|$$

subject to

- $G(x)$ is positive definite
- $F(x)$ is positive semi-definite

where $b \in \mathbb{R}^m$ and the functions $G : \mathbb{R}^m \to \mathbb{R}^{l \times l}$ and $F : \mathbb{R}^m \to \mathbb{R}^{l \times l}$ are affine:

\[
G(x) = G_0 + x_1 G_1 + \ldots + x_m G_m, \\
F(x) = F_0 + x_1 F_1 + \ldots + x_m F_m
\]

where $F_i$ and $G_i$ are symmetric matrices such that

- $F_i, i = 1, \ldots, m$ are linearly independent
- $G_i, i = 1, \ldots, m$ are linearly independent
Properties
**LAGRANGE FORM**

$\bar{A}$ – MLE of the sample covariance matrix

- Lasso-type estimate

$$\min \left[ -\ln |C| + \text{trace} \left( C \bar{A} \right) + \lambda \| C \|_{\ell_1} \right]$$

where $\| C \|_{\ell_1} = \sum_{i \neq j} |c_{ij}|$.

- Garrote-type estimate

$$\min \left[ -\ln |C| + \text{trace} \left( C \bar{A} \right) + \lambda \sum_{i \neq j} \frac{c_{ij}}{\tilde{c}_{ij}} \right]$$

subject to $c_{ij}/\tilde{c}_{ij} \geq 0$. 
**Bivariate Normal**

\[
\hat{C}_0 = \tilde{A}^{-1} = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}
\]

**Theorem** In the case of the bivariate normal, the Lasso-type estimate is

\[
\hat{c}_{12} = \left( \frac{(1 - r^2)\{ |r| - \lambda (1 - r^2) \}}{1 - \{ |r| - \lambda (1 - r^2) \}^2} \right) \frac{1}{|r|} \sign(r),
\]

the Garrote-type estimate is

\[
\hat{c}_{12} = \left( \frac{(1 - r^2)\{ r^2 - \lambda (1 - r^2) \}}{|r| - \{ r^2 - \lambda (1 - r^2) \}^2 / |r|} \right) \frac{1}{|r|} \sign(r)
\]

and

\[
\hat{c}_{11} = \hat{c}_{22} = \frac{1}{2} \left[ (1 - r^2) + \sqrt{((1 - r^2)^2 + 4\hat{c}_{12}^2)} \right].
\]
both estimates of $c_{12}$ can be shrunken to exact zero

both estimates are continuous in the data

The garrote type estimate penalizes large $r'$s less heavily
ASYMPTOTICS – GRAPHICAL GARROTE

**Theorem** Denote by $\hat{C}$ the garrote-type estimate with initial estimator $\tilde{C} = \bar{A}^{-1}$. If $n\lambda \to \infty$ and $\sqrt{n}\lambda \to 0$ as $n \to \infty$, then $\Pr(\hat{c}_{ij} = 0) \to 1$ if $c_{ij} = 0$, and other elements of $\hat{C}$ have the same limiting distribution as the maximum likelihood estimator on the true graph structure.

- the garrote type estimate enjoys the so-called oracle property
- needs a good preliminary estimate
ASYMPTOTICS — GRAPHICAL LASSO

**Theorem** If $\sqrt{n} \lambda \to \lambda_0 \geq 0$ as $n \to \infty$, the lasso-type estimator is such that

$$\sqrt{n} \left( \hat{C} - C \right) \to \arg \min_{\hat{U} = U'} (V),$$

in distribution, where

$$V(U) = \text{trace} (U \Sigma U \Sigma) + \text{trace} (U W) + \lambda_0 \sum_{i \neq j} \{u_{ij} \text{sign}(c_{ij}) I(c_{ij} \neq 0) + |u_{ij}| I(c_{ij} = 0)\}$$

in which $W$ is a random symmetric $p \times p$ matrix such that $\text{vec}(W) \sim \mathcal{N}(0, \Lambda)$, and $\Lambda$ is such that

$$\text{cov}(w_{ij}, w_{i'j'}) = \text{cov}(X^{(i)} X^{(j)}, X^{(i')} X^{(j')}).$$
EXAMPLE – TRIVARIATE NORMAL

\[ C = \begin{pmatrix} 1 & \frac{1}{3} & 0 \\ \frac{1}{3} & 1 & \frac{2}{3} \\ 0 & \frac{2}{3} & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1.25 & -0.75 & 0.5 \\ -0.75 & 2.25 & -1.5 \\ 0.5 & -1.5 & 2 \end{pmatrix}. \]
COORDINATE DESCENT

• If \( C_{-i,-i} \succ 0 \), then \( C' \succ 0 \) iff

\[
    c_{ii} - C_{i,-i}C_{-i,i}^{-1}C_{-i,i} > 0
\]

• Given \( C_{-i,-i} \succ 0 \), then \( C' \succ 0 \),

\[
    \min_{C_{i,-i}=C',i} \left[ -\ln |C| + \text{trace} \left( C\bar{A} \right) + \lambda \sum_{i \neq j} |c_{ij}| \right]
\]

subject to

\[
    c_{ii} - C_{i,-i}C_{-i,i}^{-1}C_{-i,i} > 0
\]
BACKFITTING

• In terms of $C_{-i,i}$,

$$- \ln \left( c_{ii} - C'_{-i,i} C_{-i,i}^{-1} C_{-i,i} \right) + c_{ii} \bar{A}_{ii} + 2 \bar{A}_{i,-i} C_{-i,i} + 2\lambda \sum_{j \neq i} |c_{ij}|$$

• First order condition

  ◆ Diagonal element

$$c_{ii} = \frac{1}{\bar{a}_{ii}} + C'_{-i,i} C_{-i,i}^{-1} C_{-i,i}$$

  ◆ Off-diagonal elements

$$\frac{1}{2} C'_{-i,i} \left( \bar{a}_{ii} C_{-i,i}^{-1} \right) C_{-i,i} + A_{i,-i} C_{-i,i} + \lambda \sum_{j \neq i} |c_{ij}|$$
**Graphical Lasso Algorithm**

**Input:** $\bar{A}, \lambda \geq 0$ and an initial value for $C$

**Output:** Update for the $i$th row and column

**Repeat**

**for** $i = 1$ to $p$

update $C_{-i,i}$, or equivalently $C_{i,-i}$ by solving

$$\min_{C_{-i,i}=C'_{i,-i}} \frac{1}{2} C'_{-i,i} \left( \bar{a}_{ii} C^{-1}_{-i,-i} \right) C_{-i,i} + \bar{A}_{i,-i} C_{-i,i} + \lambda \sum_{j \neq i} |c_{ij}|$$

update $c_{ii} = 1/\bar{a}_{ii} + C'_{-i,i} C^{-1}_{-i,-i} C_{-i,i}$

**end**

**Until** certain convergence criterion is met
**Graphical Garrote Algorithm**

**Input:** \( \bar{A}, \lambda \geq 0 \) and an initial value for \( C \)

**Output:** Update for the \( i \)th row and column

Repeat

for \( i = 1 \) to \( p \)

update \( C_{-i,i} \), or equivalently \( C_{i,-i} \) by solving

\[
\min_{C_{-i,i}=C'_{-i,i}} \\frac{1}{2} C'_{-i,i} \left( \bar{a}_{ii} C_{-i,-i} \right) C_{-i,i} + A_{i,-i} C_{-i,i} + \lambda \sum_{j \neq i} \frac{c_{ij}}{\tilde{c}_{ij}} \quad \text{subject to} \quad c_{ij} / \tilde{c}_{ij} \geq 0
\]

update \( c_{ii} = 1 / \bar{a}_{ii} + C'_{-i,i} C_{-i,-i} C_{-i,i} \)

end

Until certain convergence criterion is met
Numerical Examples
AR(1) MODEL – $p = 25, n = 100$
AR(1) MODEL – $n = p = 25$
Mathmarks Data

graphLasso (50 lambda's)

graphLasso (200 lambda's)

graphLasso (approximate)

graphGarrote (50 lambda's)

graphGarrote (200 lambda's)

graphGarrote (approximate)
<table>
<thead>
<tr>
<th>Order</th>
<th>graphLasso</th>
<th>graphGarrote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Algebra – Analysis</td>
<td>Algebra – Analysis</td>
</tr>
<tr>
<td>Step 2</td>
<td>Algebra – Statistics</td>
<td>Algebra – Statistics</td>
</tr>
<tr>
<td>Step 3</td>
<td>Vector – Algebra</td>
<td>Vector – Algebra</td>
</tr>
<tr>
<td>Step 4</td>
<td>Analysis – Statistics</td>
<td>Mechanics – Vector</td>
</tr>
<tr>
<td>Step 5</td>
<td>Mechanics – Vector</td>
<td>Mechanics – Algebra</td>
</tr>
<tr>
<td>Step 6</td>
<td>Mechanics – Algebra</td>
<td>Analysis – Statistics</td>
</tr>
<tr>
<td>Step 7</td>
<td>Vector – Analysis</td>
<td>Vector – Analysis</td>
</tr>
<tr>
<td>Step 8</td>
<td>Vector – Statistics</td>
<td>Mechanics – Statistics</td>
</tr>
<tr>
<td>Step 9</td>
<td>Mechanics – Analysis</td>
<td>Vector – Statistics</td>
</tr>
<tr>
<td>Step 10</td>
<td>Mechanics – Statistics</td>
<td>Mechanics – Analysis</td>
</tr>
</tbody>
</table>
MATH MARKS DATA

graphLasso (steps)

graphLasso (tuning)

graphLasso (selected graph)

graphGarro (steps)

graphGarro (tuning)

graphGarro (selected graph)
HANDWRITTEN DIGITS CLASSIFICATION
HANDWRITTEN DIGITS CLASSIFICATION

![Handwritten Digits Classification Graph](image)

- **Proposed Method**
- **Sample Covariance**

**Misclassification Error**
THALIANA DATA

- DPPS3
- DXPS2\(\text{cla1}\)
- DXPS3
- GGPPS1mt
- GGPPS4
- GGPPS5
- GGPPS6
- HDS
- MPDC2
THALIANA DATA
THALIANA DATA
CONCLUSION

- We have introduced $\ell_1$ regularized likelihood methods for the problem of covariance selection and estimation.
- The implementation of our method takes advantage of recent advances in convex optimization.
- Numerical experiments demonstrate the merits of the new method.