Generalized likelihood approaches for scalable covariance selection

Kshitij Khare

Department of Statistics
University of Florida

Workshop on "High-Dimensional Covariance Matrices, Networks and Concentration Inequalities"

Outline

- Quick intro to covariance selection/graphical models
- Review of existing penalized covariance selection approaches
- CONCORD: Covariance selection using jointly convex generalized likelihood
- Illustrations/Applications

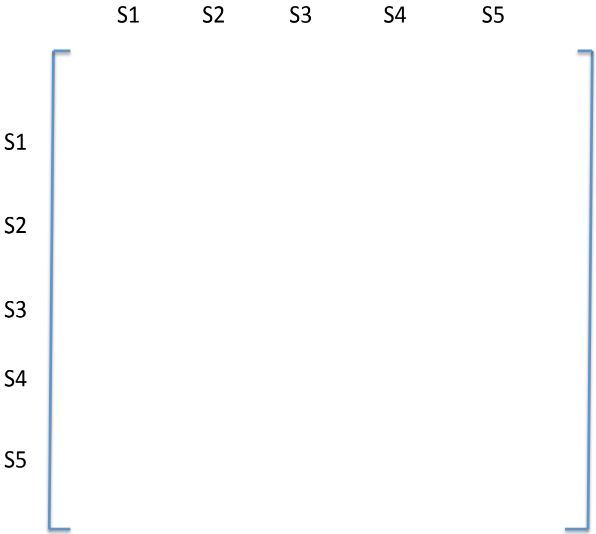
Motivation

- Availability of high-dimensional data from various applications
- Number of variables (p) much larger than (or sometimes comparable to) the sample size (n)
- Examples:
  - Biology: gene expression data
  - Environmental science: climate data on spatial grid
  - Finance: returns on thousands of stocks
- Common goals: Understand complex relationships & multivariate dependencies

Understanding relationships: The covariance matrix

- Covariance matrix Σ: a fundamental quantity to help understand multivariate relationships
- The covariance of two variables/features (say two stock prices) is a measure of linear dependence between these variables
- Positive covariance indicates similar behavior, Negative covariance indicates opposite behavior, zero covariance indicates lack of linear dependence
- Even if estimating the covariance matrix is not the end goal, it is a crucial first step before further analysis

Lets say we have five stock prices S1, S2, S3, S4, S5. The covariance matrix of these five stocks looks like



Challenges in high-dimensional estimation

- Covariance matrix (often denoted by Σ) has $O(p^2)$ unknown paramters
- If $p = 1000$, we need to estimate roughly 1 million parameters
- If sample size n is much smaller (or even same order) than p, this is not viable
- The sample covariance matrix (classical estimator) can perform very poorly in high-dimensional situations (not even invertible when $n < p$)
- Need novel methodology/tools to develop effective inferential procedures for analyzing high-dimensional data

Is there a way out?

- **Reliably estimate small number of parameters** in Σ or $\Omega = \Sigma^{-1}$
- Set insignificant parameters to zero
- Gives rise to sparse estimates of Σ or Ω
- Sparsity pattern can be represented by graphs/networks

Gaussian graphical models: Sparsity in Ω

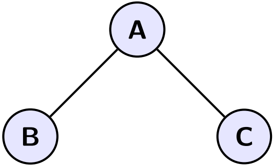
- Let $\mathbf{Y}$ be a p-dimensional random vector with a $N_p(\mathbf{0}, \Sigma = \Omega^{-1})$ distibution
- $\Omega = ((\omega_{ij}))_{1 \leqslant i,j \leqslant p}$
- $\omega_{ij} = Cov(Y_i, Y_j \mid \mathbf{Y}_{-(i,j)})$
- $\omega_{ij} = 0$ if and only if the i^{th} and j^{th} variables are conditionally independent given the other variables
- Zeros in Ω encode conditional independence under Gaussianity

Concentration Graphical Models: Connection with graphs

- Sparsity pattern in Ω can be represented by an undirected graph $G = (V, E)$
- $V = \{1, \ldots, p\}$ and set E of edges is such that $\omega_{ij} = 0 \Leftrightarrow (i, j) \in E$.
- Build a graph from sparse Ω

$$\Omega = \begin{array}{c} \\ \end{array}\begin{matrix} \text{A} & \text{B} & \text{C} \end{matrix}$$

$$\Omega = \begin{pmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 2 & 0 \\ 0.3 & 0 & 1.2 \end{pmatrix} \begin{matrix} \text{A} \\ \text{B} \\ \text{C} \end{matrix}$$



ℓ_1-regularized graphical model selection

- Suppose we have i.i.d. observations $\mathbf{Y}^1, \mathbf{Y}^2, \cdots, \mathbf{Y}^n$ from a p-dimensional distribution with mean $\mathbf{0}$ and covariance matrix $\Sigma = \Omega^{-1}$
- Graphical model selection: Selecting a sparsity pattern/graph for Ω based on observed data
- Two main approaches:
  1. ℓ_1-regularized **likelihood** methods
  2. ℓ_1-regularized **regression-based/pseudo-likelihood** methods

Regularized Gaussian likelihood graphical model selection: The Graphical Lasso

- Assume underlying distribution is multivariate normal
- All ℓ_1-regularized Gaussian-likelihood methods solve

$$\hat{\Omega} = \arg\max_{\Omega \succ 0} \{\log\det(\Omega) - \mathrm{tr}(\Omega S) - \lambda \|\Omega\|_1\}$$

- Adding ℓ_1-regularization term $\lambda \|\Omega\|_1$ introduces sparsity
- Penalty parameter λ controls level of sparsity
- Dependency on Gaussianity, sensitivity to outliers

Penalized Gaussian likelihood methods

- Block coordinate descent (COVSEL) [Banerjee et al. 2008]
- Graphical lasso (GLASSO) [Friedman, Hastie, & Tibshirani, 2008]
- Large-scale GLASSO [Mazumder & Hastie, 2012]
- QUIC [Hsieh et al., 2014]
- G-ISTA [Guillot, Rajaratnam et al., 2012]
- Graphical Dual Proximal Gradient Methods [Dalal & Rajaratnam, 2013]
- Others

Other approaches

- *Thresholding based approaches*: Hero & Rajaratnam (2011), Hero & Rajaratnam (2012), Zhang et al. (2020)
- Can be computational much faster but unstable

- *Bayesian approaches*: Wang (2015)
- Offer advantage of uncertainty quantification, but problems with computational scalability

Penalized pseudo-likelihood graphical model selection

- Based on a regression interpretation of entries of Ω
- Let $\mathbf{Y} \in \mathbb{R}^p$ have mean 0 and covariance matrix Ω^{-1}
- Let $\boldsymbol{\beta}^i = \arg\min_{\mathbf{u} \in \mathbb{R}^{p-1}} E\left[(Y_i - \mathbf{u}^t \mathbf{Y}_{-i})^2\right]$
- Then $\beta^i_j = -\frac{\omega_{ij}}{\omega_{ii}}$
- Does not assume normality
- Parametrize in terms of partial correlations

$$\rho^{ij} = -\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}$$

Neighborhood selection (Meinshausen and Buhlmann, 2006)

- Note that $\omega_{ij} = 0$ iff $\beta^i_j = 0$
- Separately use p lasso regression select sparsity patterns in the p rows of Ω
- Concern: Lack of symmetry in the resulting sparsity patterns
- β^i_j chosen as zero, but β^j_i may not be chosen as zero

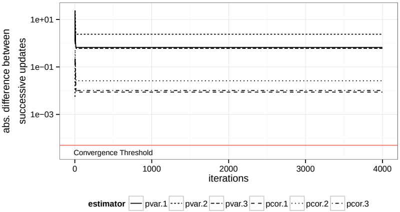
SPACE algorithm (Peng et al., 2009)

- **SPACE objective function**: $(w_i = 1 \text{ or } w_i = \omega_{ii})$

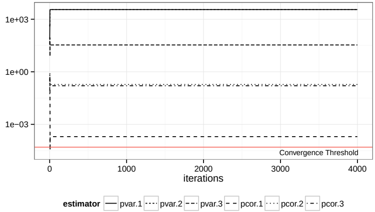
$$Q_{\text{spc}}(\Omega) =: \sum_{i=1}^{p} \left\{ \frac{-n \log \omega_{ii}}{2} + \frac{w_i}{2} \|\mathbf{Y}_i - \sum_{j \neq i} \underbrace{\rho^{ij} \sqrt{\frac{\omega_{jj}}{\omega_{ii}}}}_{\beta_j^i} \mathbf{Y}_j\|_2^2 \right\} + \lambda \sum_{1 \leqslant i < j \leqslant p} |\rho^{ij}|$$

- Use cyclic block coordinatewise minimization with blocks $\left\{\rho^{ij}\right\}_{1 \leqslant i < j \leqslant p}$ and $\{\omega_{ii}\}_{1 \leqslant i \leqslant p}$
- Q_{spc} is bi-convex **but not jointly convex**. No theoretical convergence guarantees

Non-converging example: $p = 3$ case



(a) SPACE ($w_i = \omega_{ii}$)



(b) SPACE ($w_i = 1$)

Figure: $\mathbf{Y}^{(\mathbf{i})} \sim \mathcal{N}_3(0, \Omega^{-1})$, (left) $n = 4$, (right) $n = 100$

In both cases, successive iterates generated by the SPACE algorithm eventually alternate between two matrices

Non-convergence of SPACE

- Investigate the nature and extent of convergence issues:
  1. Are such examples pathological? How widespread are they?
  2. When do they occur ?
- Consider a sparse 100×100 matrix Ω with edge density 4% and condition number of 100.
- Generate 100 multivariate Gaussian datasets (with $n = 100$), $\mu = 0$ and $\Sigma = \Omega^{-1}$.
- Record the number of times (out of 100) for which SPACE1 (uniform weights) and SPACE2 (partial variance weights) do not converge within 1500 iterations.
- Original implementation of SPACE by Peng et al. (2009) claims only 3 iterations are sufficient.

Non-convergence of SPACE

SPACE1 ($w_i = 1$)		SPACE2 ($w_i = \omega_{ii}$)	
λ^*	NC	λ^*	NC
0.026	92	0.085	100
0.099	100	0.160	0
0.163	100	0.220	0
0.228	100	0.280	0
0.614	0	0.730	97

Table: Number of simulations (out of 100) that do not converge within 1500 iterations (NC) for select values of penalty parameter ($\lambda^* = \lambda/n$). Average percentage of non-zeros (NZ) in $\hat{\Omega}$ are also shown.

- SPACE exhibits extensive non-convergence behavior
- Problem exacerbated when condition number is high
- Typical of high dimensional sample starved settings

Symmetric Lasso and SPLICE

SYMLASSO [Friedman, Hastie, & Tibshirani, 2010]:

$$Q_{\text{sym}}(\boldsymbol{\alpha}, \breve{\Omega}) = \frac{1}{2}\sum_{i=1}^{p}\left[n\log\alpha_{ii} + \frac{1}{\alpha_{ii}}\|\mathbf{Y}_i + \sum_{j\neq i}\omega_{ij}\alpha_{ii}\mathbf{Y}_j\|^2\right] + \lambda\sum_{1\leqslant i<j\leqslant p}|\omega_{ij}|,$$

where $\alpha_{ii} = 1/\omega_{ii}$.

SPLICE [Rocha et al. (2008)]:

$$Q_{\text{spl}}(\mathbf{B}, \mathbf{D}) = \frac{n}{2}\sum_{i=1}^{p}\log(d_{ii}^2) + \frac{1}{2}\sum_{i=1}^{p}\frac{1}{d_{ii}^2}\|\mathbf{Y}_i - \sum_{j\neq i}\beta_{ij}\mathbf{Y}_j\|^2 + \lambda\sum_{i<j}|\beta_{ij}|,$$

where $d_{ii}^2 = \omega_{ii}$.

Also, alternating (off-diagonal vs diagonal) iterative algorithms
No convergence guarantees

Regularized regression-based graphical model selection

- **Advantage**: Regression-based methods typically **computationally faster**
- **Advantage**: Regression-based methods are **more robust** and less restrictive than Gaussian likelihood-based methods
- **Disadvantage**: $\hat{\Omega}$ may not be positive definite (can be fixed)
- **Disadvantage**: Solution may not be computable. Lack of theoretical convergence guarantees

Regression-based methods: summary

	METHOD			
Property	NS	SPACE	SYMLASSO	SPLICE
Symmetry		$+$	$+$	$+$
Convergence guarantee	N/A			
Asymptotic consistency $(n, p \to \infty)$	$+$	$+$		

How can we obtain all of the good properties simultaneously?

Design goals of a new pseudo-likelihood approach

- Can we design a regression-based approach that guarantees existence of a solution?

- Is there a better chance of guaranteeing a well defined solution if a convex formulation is developed?

- Advantages of a convex formulation:
  - Easier analysis of theoretical properties
  - Better chance of algorithmic convergence
  - Global minimum is guaranteed to exist

Convex formulation of graphical model selection problem

Revisit the SPACE objective function

$$Q_{\text{spc}}(\Omega) =: -\frac{1}{2}\sum_{i=1}^{p} n\log\omega_{ii} + \frac{1}{2}\sum_{i=1}^{p} w_i\|\mathbf{Y}_i - \sum_{j\neq i}\underbrace{\rho^{ij}\sqrt{\frac{\omega_{jj}}{\omega_{ii}}}}_{\beta_{ij}}\mathbf{Y}_j\|_2^2 + \lambda\sum_{1\leqslant i<j\leqslant p}|\rho^{ij}|$$

- $Q_{\text{spc}}(\Omega)$ is not jointly convex in elements of Ω
- Since $\beta_{ij} = \rho^{ij}\sqrt{\frac{\omega_{jj}}{\omega_{ii}}} = -\frac{\omega_{ij}}{\omega_{ii}}$, regression term is not convex
- Since $|\rho^{ij}| = \left|-\frac{\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\right|$, penalty term is not convex

Convex formulation of graphical model selection problem

Consider,

$$
\begin{aligned}
w_i\|\mathbf{Y}_i - \sum_{j\neq i}\rho^{ij}\sqrt{\frac{\omega_{jj}}{\omega_{ii}}}\mathbf{Y}_j\|_2^2 &= w_i\|\mathbf{Y}_i + \sum_{j\neq i}\frac{\omega_{ij}}{\omega_{ii}}\mathbf{Y}_j\|_2^2 \qquad \left(\because \rho^{ij} = \frac{-\omega_{ij}}{\sqrt{\omega_{ii}\omega_{jj}}}\right) \\
&= w_i\|\frac{1}{\omega_{ii}}(\omega_{ii}\mathbf{Y}_i + \sum_{j\neq i}\omega_{ij}\mathbf{Y}_j)\|_2^2 \\
&= \frac{w_i}{\omega_{ii}^2}\|\sum_{j=1}^{p}\omega_{ij}\mathbf{Y}_j\|_2^2
\end{aligned}
$$

Now, let $w_i = \omega_{ii}^2$, then

$$
\|\sum_{j=1}^{p}\omega_{ij}\mathbf{Y}_j\|_2^2 = \omega_{\bullet i}'\mathbf{Y}'\mathbf{Y}\omega_{\bullet i} \geqslant 0, \text{ (quadratic form)}
$$

Therefore, Q_{con} below is jointly convex:

$$
Q_{\text{con}}(\Omega) =: -\sum_{i=1}^{p} n\log\omega_{ii} + \frac{1}{2}\sum_{i=1}^{p}\|\omega_{ii}\mathbf{Y}_i + \sum_{j\neq i}\omega_{ij}\mathbf{Y}_j\|_2^2 + \lambda\sum_{1\leqslant i<j\leqslant p}|\omega_{ij}|
$$

Establishing properties of CONCORD

Optimization properties

- Task 1: (**Optimization algorithm**) Can we find an effective algorithm to minimize the $Q_{\text{con}}(\Omega)$ so that a solution always exists and is computable?
- Task 2: (**Guarantee of convergence to global optimum**) Can we establish convergence? Do we have a globally optimal solution?
- Task 3: (**Computational complexity**) What is the computational complexity of the optimization method? Is it competitive with other methods?
- Task 4: (**Running time comparison**) How do the actual running times compare with other methods?

Establishing properties of CONCORD

Statistical Properties

- Task 5: (**Consistency and Large Sample properties**) Are Concord estimates guaranteed to recover the true underlying partial correlation graphs for data generated from such models?
- Task 6: (**Finite sample properties**) How does CONCORD perform in terms of recovering the partial correlation graph in finite sample settings?
- Task 7: (**Applications**) How does CONCORD perform in applications in comparison with other methods where high dimensional covariance estimates are required?

Goal: To investigate the above questions systematically

CONvex CORrelation selection methoD (CONCORD)

CONCORD objective function:

$$Q_{\text{con}}(\Omega) =: -\sum_{i=1}^{p} n \log \omega_{ii} + \frac{1}{2}\sum_{i=1}^{p} \|\omega_{ii}\mathbf{Y}_i + \sum_{j\neq i} \omega_{ij}\mathbf{Y}_j\|_2^2 + \lambda \sum_{1\leqslant i<j\leqslant p} |\omega_{ij}|$$

Cyclic coordinate-wise minimization algorithm

1. **Update** $[\omega_{ij}]$[1] (other coefficients held constant):

$$\omega_{ij} \leftarrow \frac{S_{\frac{\lambda}{n}}\left(-\left(\sum_{j'\neq j} \omega_{ij'} S_{jj'} + \sum_{i'\neq i} \omega_{i'j} S_{ii'}\right)\right)}{S_{ii} + S_{jj}}$$

2. **Update** $[\omega_{ii}]$ (other coefficients held constant):

$$\omega_{ii} \leftarrow \frac{-\sum_{j\neq i} \omega_{ij} S_{ij} + \sqrt{\left(\sum_{j\neq i} \omega_{ij} S_{ij}\right)^2 + 4S_{ii}}}{2S_{ii}}$$

[1] Soft-thresholding operator: $S_\lambda(x) = \text{sign}(x)(|x| - \lambda)_+$

Convergence guarantees for CONCORD algorithm

- CONCORD objective function is not in general strictly convex if $n < p$
- Existing convergence results in the optimization literature are not sufficient
- Tseng (2001): Every limit point of the sequence of iterates produced by the cyclic coordinatewise minimization algorithm is a stationary point (not enough)
- Tseng and Yun (2009): Convergence results for a quadratic approximation based gradient descent approach (does not apply)
- Others

Convergence guarantees for CONCORD algorithm

Theorem (K. and Rajaratnam (2014)) Let $\mathcal{A}_p$ denote space of $p \times p$ symmetric matrices. Also, let $\mathcal{M} \subset \mathcal{A}_p$ denote a subspace such that

$$\mathcal{M} := \{\Omega \in \mathcal{A}_p : \omega_{ii} > 0, \text{ for every } 1 \leqslant i \leqslant p\}.$$

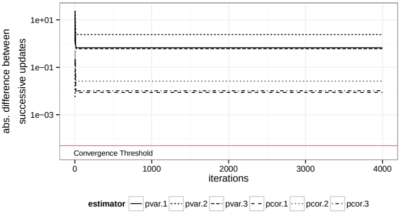
If $Y_i \neq 0$ for every $1 \leqslant i \leqslant p$, **the sequence of iterates $\{\hat{\Omega}^{(r)}\}_{r \geqslant 0}$ obtained by the CONCORD algorithm converges to a global minimum of $Q_{\text{con}}(\Omega)$.** More specifically,

$$\hat{\Omega}^{(r)} \to \hat{\Omega} \in \mathcal{M} \text{ as } r \to \infty$$

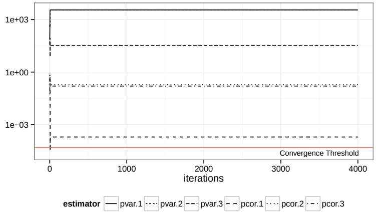
for some $\hat{\Omega}$, and furthermore

$$Q_{\text{con}}(\hat{\Omega}) \leqslant Q(\Omega) \text{ for all } \Omega \in \mathcal{M}.$$

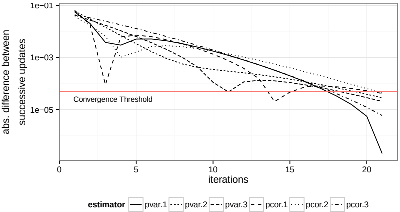
Non-converging example: $p = 3$ case



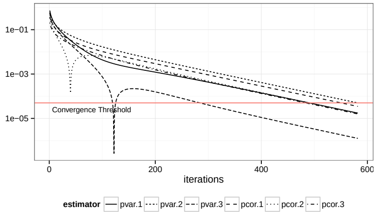
(a) SPACE ($w_i = \omega_{ii}$)



(b) SPACE ($w_i = 1$)



(c) CONCORD



(d) CONCORD

Figure: $\mathbf{Y}^{(\mathbf{i})} \sim \mathcal{N}_3(0, \Omega^{-1})$, (left) $n = 4$, (right) $n = 100$

Computational complexity of CONCORD algorithm

- GLASSO: $O(p^3)$
- SPACE: $\min(O(np^2), O(p^3))$
- SYMLASSO: $\min(O(np^2), O(p^3))$
- CONCORD: $\min(O(np^2), O(p^3))$

Running time of CONCORD: I

$\mathbf{p = 1000}$, $\mathbf{n = 200}$					
GLASSO			CONCORD		
λ	NZ	Time	λ^*	NZ	Time
0.14	4.77%	87.60	0.12	4.23%	6.12
0.19	0.87%	71.47	0.17	0.98%	5.10
0.28	0.17%	5.41	0.28	0.15%	5.37
0.39	0.08%	5.30	0.39	0.07%	4.00
0.51	0.04%	6.38	0.51	0.04%	4.76

$\mathbf{p = 1000}$, $\mathbf{n = 200}$					
SPACE1 ($w_i = 1$)			SPACE2 ($w_i = \omega_{ii}$)		
λ	NZ	Time	λ^*	NZ	Time
0.10	4.49%	101.78	0.16	100.00%	19206.55
0.17	0.64%	99.20	0.21	1.76%	222.00
0.28	0.14%	138.01	0.30	0.17%	94.59
0.39	0.07%	75.55	0.40	0.08%	108.61
0.51	0.04%	49.59	0.51	0.04%	132.34

Table: Timing comparison (seconds) for $p = 1000$, λ = penalty parameter, $\lambda^* = \lambda/n$ for CONCORD/SPACE. NZ = the percentage of non-zero entries

Running time of CONCORD: II

p = 3000, n = 600					
GLASSO			CONCORD		
λ	NZ	Time	λ^*	NZ	Time
0.09	2.71%	1842.74	0.09	2.10%	266.69
0.10	1.97%	1835.32	0.10	1.59%	235.49
0.10	1.43%	1419.41	0.10	1.19%	232.67

p = 3000, n = 900					
GLASSO			CONCORD		
λ	NZ	Time	λ^*	NZ	Time
0.09	0.70%	1389.96	0.09	0.64%	298.21
0.10	0.44%	1395.42	0.10	0.41%	298.00
0.10	0.27%	1334.78	0.10	0.26%	302.15

Table: Timing comparison (seconds) for $p = 3000$, λ = penalty parameter, $\lambda^* = \lambda/n$ for CONCORD. NZ = the percentage of non-zero entries

- CONCORD is highly competitive.
- Orders of magnitude faster in high dimensional settings.
- SPACE is slow to converge when $n \ll p$.

Large sample properties: Assumptions

For sample size n and number of feature $p = p_n$, assume
True inverse covariance matrix: $\bar{\Omega}_n = [\bar{\omega}_{n,ij}]$, $1 \leqslant i, j \leqslant p_n$, and
$\bar{\omega}_n^o$ denotes the off-diagonal elements.
Assumptions:

- A0: Accurate estimates of diagonals $\hat{\alpha}_{n,ii}$:

$$\max_{1 \leqslant i \leqslant p_n} |\widehat{\alpha}_{n,ii} - \bar{\omega}_{ii}| \leqslant C \left(\sqrt{\frac{\log n}{n}} \right),$$

  holds with probability larger than $1 - O(n^{-\eta})$.

- A1: Bounded eigenvalues: eigenvalues of $\bar{\Omega}_n$ are such that

$$\lambda_{min} > 0 \text{ and } \lambda_{max} < \infty \text{ , for all } n$$

- A2: Sub-Gaussianity,

- A3: Incoherence condition

Large sample properties: Theorem

Suppose that assumptions (A0)-(A3) are satisfied. Suppose $p_n = O(n^\kappa)$ for some $\kappa > 0$, $q_n = o\left(\sqrt{n}\log n\right)$, $\sqrt{\frac{q_n \log n}{n}} = o(\lambda_n)$, $\lambda_n \sqrt{n} \log n \to \infty$, and $\sqrt{q_n}\lambda_n \to 0$, as $n \to \infty$.

Then there exists a constant C such that for any $\eta > 0$, the following events hold with probability at least $1 - O(n^{-\eta})$.

- There exists a minimizer $\widehat{\omega}_n^o = ((\widehat{\omega}_{n,ij}))_{1 \leqslant i < j \leqslant p_n}$ of $Q_{\text{con}}(\omega^o, \widehat{\boldsymbol{\alpha}}_n)$.
- Any minimizer $\widehat{\omega}_n^o$ of $Q_{\text{con}}(\omega^o, \widehat{\boldsymbol{\alpha}}_n)$ satisfies

$$\|\widehat{\omega}_n^o - \bar{\omega}_n^o\|_2 \leqslant C\sqrt{q_n}\lambda_n \quad \textbf{(Parameter consistency)}$$

and

$$\text{sign}(\widehat{\omega}_{n,ij}) = \text{sign}(\bar{\omega}_{n,ij}),\ \forall\ 1 \leqslant i < j \leqslant p_n \quad \textbf{(Sign consistency)}.$$

CONCORD method: summary

- Optimization aspects
  - **Jointly convex formulation**
  - Theoretical **guarantee of convergence**
  - Converges to **globally optimal solution**
- Statistical properties
  - **Asymptotically consistent** estimator as $n, p \to \infty$
  - **Competitive with other pseudo-likelihood methods** in finite sample
- Computational efficiency
  - **Computationally complexity is competitive**
  - **Wall clock time performence is superior in high dimensions**

CONCORD method: summary

	METHOD				
Property	NS	SPACE	SYMLASSO	SPLICE	CONCORD
Symmetry		+	+	+	+
Convergence guarantee (fixed n)	N/A				+
Asymptotic consistency ($n, p \to \infty$)	+	+			+

Yes! CONCORD retains all good properties

Applications of graphical model selection and (inverse) covariance estimation

Biological application: gene co-expression of breast cancer

- Breast cancer gene expression study [Chang et al. (2005)]
- $n = 248$ and other clinical data (metastasis, tumor size, etc..)
- Reduce to $\sim$1100 genes by survival analysis (from $\sim$20000)
- Select λ such that 200 non-zero elements remain in $\hat{\Omega}$
- Identify most highly connected (hub) genes [Carter et al. (2004), Jeong et al. (2001), Han et al. (2004)]

Biological application: gene co-expression of breast cancer

Gene Symbol	CONCORD	SYMLASSO	SPACE1	SPACE2	Reference
HNF3A (FOXA1)	+	+	+	+	[Koboldt et al. (2012), Albergaria et al. (2009) , Davidson et al. (2011), Lacroix et al. (2004), Robinson et al. (2011)]
TONDU	+	+	+	+	
FZD9	+	+	+	+	[Katoh et al. (2008), Ronneberg et al. (2011)]
KIAA0481	+	+	+	+	[Gene record discontinued]
KRT16	+	+	+		[Glinsky et al. (2005), Joosse et al. (2012), Pellegrino et al. (1988)]
KNSL6 (KIF2C)	+			+	[Eschenbrenner et al. (2011), Shimo et al. (2007), Shimo et al. (2008)]
FOXC1	+	+	+	+	[Du et al. (2012), Sizemore et al. (2012), Wang et al. (2012), Ray et al. (2011), Tkocz et al. (2011)]
PSA	+	+		+	[Kraus et al. (2010), Mohajeri et al. (2011), Sauter et al. (2004), Yang et al. (2002)]
GATA3	+	+	+	+	[Koboldt et al. (2012), Davidson et al. (2011), Albergaria et al. (2009), Eeckhoute et al. (2007), Jiang et al. (2010), Licata et al. (2010), Yan et al. (2010)]
C20ORF1 (TPX2)	+				[Maxwell et al. (2011), Bibby et al. (2009)]
E48		+	+	+	
ESR1				+	[Zheng et al. (2012)]

Maxwell et al. (2011) identifies a regulatory mechanism involving TPX2, Aurora A, RHAMM and BRCA1 genes in breast cancer

TPX2 gene in breast cancer

- Maxwell et al. (2011) is an extensive study involving thousands of breast cancer patients

- Breast cancer type 1 susceptibility protein (BRCA1), a known gene related to breast cancer

- TPX2 gene is identified as having strong link to BRCA1

- **“Reorganization (of microtubules) is facilitated by BRCA1 and impaired by AURKA, which is regulated by negative feedback involving RHAMM and TPX2.”** [Maxwell et al., 2011]

Financial application: portfolio optimization

Dow-Jones Index:

- Index of 30 stocks
- Mean-variance portfolio (MVP) theory uses covariance matrix to hedge risk
- Simplest variant: minimum variance portfolio (given Σ)

$$\begin{array}{ll} \text{minimize} & w^T \Sigma w \\ \text{subject to} & \mathbf{1}^T w = 1 \end{array}$$

  Analytical solution: $w^\star = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \Sigma^{-1} \mathbf{1}$

- Due to non-stationarity, use rebalancing strategy:
  Every 4 weeks, use past N_{est} days for $\hat{\Sigma} = \hat{\Omega}^{-1}$

Financial application: portfolio optimization

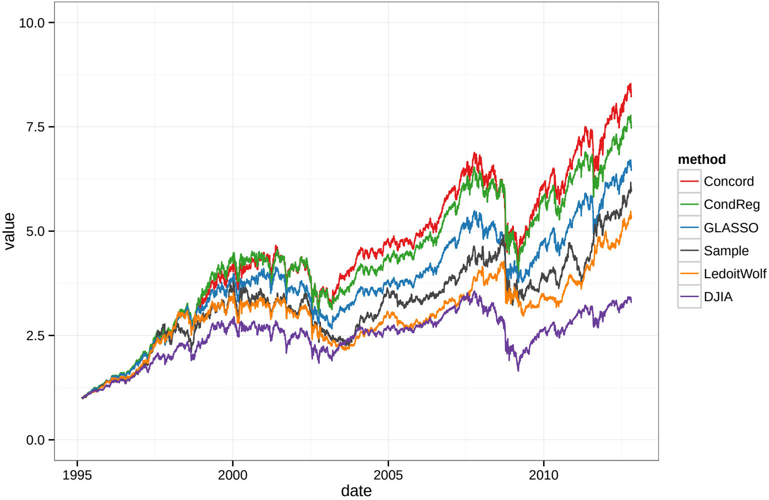


Figure: $N_{\text{est}} = 75$ days, rebalance every 4 weeks

Finance: Minimum variance portfolio returns

Return measure: mean excess return per unit of risk

$$\textbf{Sharpe ratio} = \frac{\mathbb{E}(R_t - R_f)}{\sqrt{\mathrm{Var}\,(R_t)}}, \text{ where } R_f = 3\% \text{ (annual) is chosen}$$

N_{est}	DJIA	Sample	GLASSO	Concord	CondReg	LedoitWolf
35	2.09	2.77	4.01	**4.12**	4.06	**4.10**
40	2.09	3.44	3.93	**4.10**	**3.98**	3.91
45	2.09	2.43	3.78	**3.98**	**3.85**	3.59
50	2.09	2.31	3.81	**4.06**	**3.89**	3.71
75	2.09	3.40	3.70	**4.04**	**3.89**	3.49
	References		Sparse models		Dense estimates	

Table: Penalty λ chosen with cross-validation to minimize RSS, (values multiplied by 100)

Applications: summary

- Biological example: hub gene discovery
  - Discovered empirically validated genes
  - Other methods are useful too!
- Finance example: minimum variance portfolio selection
  - CONCORD estimator yields best Sharpe ratio even better than Ledoit-Wolf
  - Graphical model selection methods adapt to changing covariance structure

Questions?