High-Dimensional Covariance Decomposition into Sparse Markov and Independence Domains

Majid Janzamin and Anima Anandkumar

U.C. Irvine

High-Dimensional Covariance Estimation

- n i.i.d. samples, p variables $\mathbf{X} := [X_1, \ldots, X_p]^T$.

- Covariance estimation:

$$\Sigma^* := \mathbb{E}[\mathbf{X}\mathbf{X}^T].$$

- High-dimensional regime: both $n, p \to \infty$ and $n \ll p$.

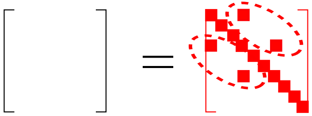
- Challenge: empirical (sample) covariance ill-posed when $n \ll p$:

$$\widehat{\Sigma}^n := \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}(k)\mathbf{x}(k)^T.$$

Solution: Imposing Sparsity for Tractable High-dimensional Estimation

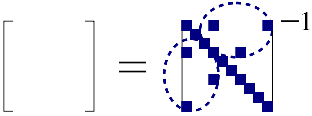
Incorporating Sparsity in High Dimensions

Sparse Covariance



Σ^* $\qquad$ Σ_R^*

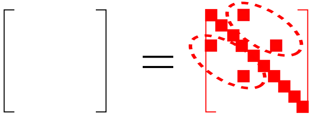
Sparse Inverse Covariance



Σ^* $\qquad$ ${J_M^*}^{-1}$

Incorporating Sparsity in High Dimensions

Sparse Covariance



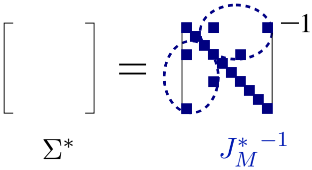
Σ^* $\quad$ Σ^*_R

Relationship with Statistical Properties (Gaussian)

- Sparse Covariance (Independence Model): marginal independence

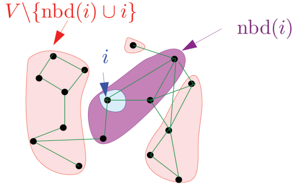
Incorporating Sparsity in High Dimensions

Sparse Inverse Covariance



Relationship with Statistical Properties (Gaussian)

- Sparse Inverse Covariance (Markov Model): conditional independence



Local Markov Property:

$$X_i \perp X_{V\backslash\{\mathrm{nbd}(i)\cup i\}} \mid X_{\mathrm{nbd}(i)}$$

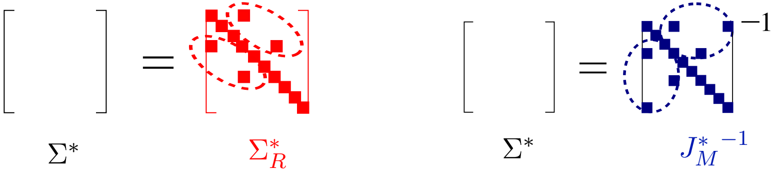
For Gaussian:

$$J_{ij} = 0 \Leftrightarrow (i,j) \notin E$$

Incorporating Sparsity in High Dimensions

Sparse Covariance

Sparse Inverse Covariance



Relationship with Statistical Properties (Gaussian)

- Sparse Covariance (Independence Model): marginal independence
- Sparse Inverse Covariance (Markov Model): conditional independence

Incorporating Sparsity in High Dimensions

Sparse Covariance

Sparse Inverse Covariance

Σ^* $\qquad$ ${\Sigma}_R^*$ $\qquad\qquad$ Σ^* $\qquad$ ${J_M^*}^{-1}$

Relationship with Statistical Properties (Gaussian)

- Sparse Covariance (Independence Model): marginal independence
- Sparse Inverse Covariance (Markov Model): conditional independence

---

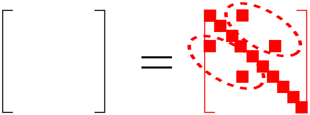
Guarantees under Sparsity Constraints in High Dimensions

Consistent Estimation when $\boxed{n = \Omega(\log p)} \Rightarrow n \ll p$.

Consistent: Sparsistent and Satisfying reasonable Norm Guarantees.

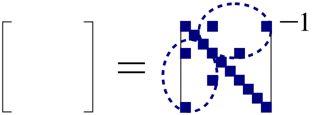
Incorporating Sparsity in High Dimensions

Sparse Covariance



Σ^* $\quad$ Σ_R^*

Sparse Inverse Covariance



Σ^* $\quad$ ${J_M^*}^{-1}$

Relationship with Statistical Properties (Gaussian)

- Sparse Covariance (Independence Model): marginal independence
- Sparse Inverse Covariance (Markov Model): conditional independence

---

Guarantees under Sparsity Constraints in High Dimensions

Consistent Estimation when $\boxed{n = \Omega(\log p)} \Rightarrow n \ll p$.

Going beyond Sparsity in High Dimensions?

Going Beyond Sparse Models

Motivation

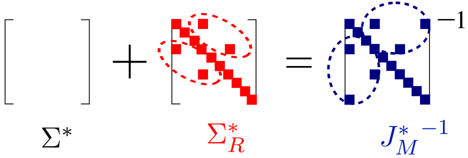
- Sparsity constraints restrictive to have faithful representation.
- Data not sparse in a single domain
- Solution: Sparsity in Multiple Domains.
- Challenge: Hard to impose sparsity in different domains

Going Beyond Sparse Models

Motivation

- Sparsity constraints restrictive to have faithful representation.
- Data not sparse in a single domain
- Solution: Sparsity in Multiple Domains.
- Challenge: Hard to impose sparsity in different domains

One Possibility (This Work): Proposing Sparse Markov Model by adding Sparse Residual Perturbation

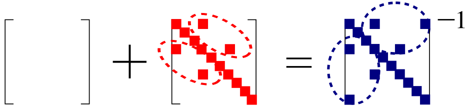


Going Beyond Sparse Models

Motivation

- Sparsity constraints restrictive to have faithful representation.
- Data not sparse in a single domain
- Solution: Sparsity in Multiple Domains.
- Challenge: Hard to impose sparsity in different domains

One Possibility (This Work): Proposing Sparse Markov Model by adding Sparse Residual Perturbation



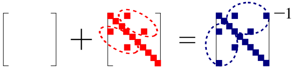
Σ^* $\qquad$ Σ_R^* $\qquad$ ${J_M^*}^{-1}$

Efficient Decomposition and Estimation in High Dimensions?

Unique Decomposition? $\quad$ Good Sample Requirements?

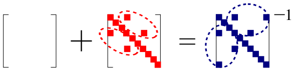
Summary of Results

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$



Summary of Results

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$

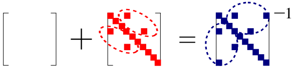


Contribution 1: Novel Model for Decomposition

- Decomposition into Markov and residual domains.
- Statistically meaningful model
- Unification of Sparse Covariance and Inverse Covariance Estimation.

Summary of Results

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$



Contribution 1: Novel Model for Decomposition

- Decomposition into Markov and residual domains.
- Statistically meaningful model
- Unification of Sparse Covariance and Inverse Covariance Estimation.

Contribution 2: Methods and Guarantees

- Conditions for unique decomposition (exact statistics).
- Sparsistency and norm guarantees in both Markov and independence domains (sample analysis)
- Sample requirement: no. of samples $\boxed{n = \Omega(\log p)}$ for p variables.

Related Works

Sparse Covariance/Inverse Covariance Estimation

- Sparse Covariance Estimation: Covariance Thresholding.
  - (Bickel & Levina) (Wagaman & Levina) (Cai et. al.)
- Sparse Inverse Covariance Estimation:
  - ℓ_1 Penalization (Meinshausen and Bühlmann) (Ravikumar et. al)
  - Non-Convex Methods (Anandkumar et. al) (Zhang)

Related Works

Sparse Covariance/Inverse Covariance Estimation

- Sparse Covariance Estimation: Covariance Thresholding.
  - (Bickel & Levina) (Wagaman & Levina) (Cai et. al.)
- Sparse Inverse Covariance Estimation:
  - ℓ_1 Penalization (Meinshausen and Bühlmann) (Ravikumar et. al)
  - Non-Convex Methods (Anandkumar et. al) (Zhang)

Beyond Sparse Models: Decomposition Issues

- Sparse + Low Rank (Chandrasekaran et. al) (Candes et. al)
- Decomposable Regularizers (Negahban et. al)

Related Works

Sparse Covariance/Inverse Covariance Estimation

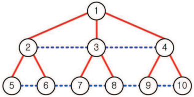
- Sparse Covariance Estimation: Covariance Thresholding.
  - (Bickel & Levina) (Wagaman & Levina) (Cai et. al.)
- Sparse Inverse Covariance Estimation:
  - ℓ_1 Penalization (Meinshausen and Bühlmann) (Ravikumar et. al)
  - Non-Convex Methods (Anandkumar et. al) (Zhang)

Beyond Sparse Models: Decomposition Issues

- Sparse + Low Rank (Chandrasekaran et. al) (Candes et. al)
- Decomposable Regularizers (Negahban et. al)

Multi-Resolution Markov+Independence Models (Choi et. al)

- Decomposition in inverse covariance domain
- Lack theoretical guarantees



Our contribution: Guaranteed Decomposition and Estimation

Outline

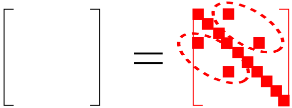
Some Intuitions and Ideas

Review Ideas for Special Cases: Sparse Covariance/Inverse Covariance

Some Intuitions and Ideas

Review Ideas for Special Cases: Sparse Covariance/Inverse Covariance

Sparse Covariance Estimation (Independence Model)

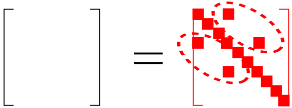


- $\Sigma^* = \Sigma^*_I$.
- $\widehat{\Sigma}^n$: sample covariance using n samples
- p variables: $p \gg n$.

Some Intuitions and Ideas

Review Ideas for Special Cases: Sparse Covariance/Inverse Covariance

Sparse Covariance Estimation (Independence Model)



- $\Sigma^* = \Sigma_I^*$.
- $\widehat{\Sigma}^n$: sample covariance using n samples
- p variables: $p \gg n$.

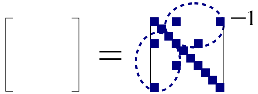
- Hard-thresholding the off-diagonal entries of $\widehat{\Sigma}^n$ (Bickel & Levina):

  threshold chosen as $\boxed{\sqrt{\frac{\log p}{n}}}$

- Sparsistency (support recovery) and Norm Guarantees when $\boxed{n = \Omega(\log p)} \Rightarrow n \ll p$.

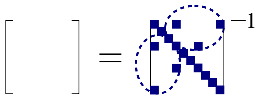
Recap of Inverse Covariance (Markov) Estimation

- $\Sigma^* = {J_M^*}^{-1}$
- $\widehat{\Sigma}^n$: sample covariance using n i.i.d. samples



Recap of Inverse Covariance (Markov) Estimation

- $\Sigma^* = {J_M^*}^{-1}$
- $\widehat{\Sigma}^n$: sample covariance using n i.i.d. samples



ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al. '08)

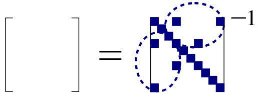
$$\widehat{J}_M := \operatorname{argmin} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \| J_M \|_{1,\text{off}}$$

where

$$\|J_M\|_{1,\text{off}} := \sum_{i \neq j} |\, (J_M)_{ij} \,|.$$

Recap of Inverse Covariance (Markov) Estimation

- $\Sigma^* = {J_M^*}^{-1}$
- $\widehat{\Sigma}^n$: sample covariance using n i.i.d. samples



ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al. '08)

$$\widehat{J}_M := \operatorname{argmin} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \| J_M \|_{1,\text{off}}$$

where

$$\|J_M\|_{1,\text{off}} := \sum_{i \neq j} |\, (J_M)_{ij} \,|.$$

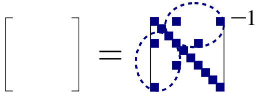
Max-entropy Formulation (Lagrangian Dual)

$$\widehat{\Sigma}_M := \underset{\Sigma_M \succ 0}{\operatorname{argmax}} \; \log\det \Sigma_M$$

$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d$$

Recap of Inverse Covariance (Markov) Estimation

- $\Sigma^* = {J_M^*}^{-1}$
- $\widehat{\Sigma}^n$: sample covariance using n i.i.d. samples



ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al. '08)

$$\widehat{J}_M := \operatorname{argmin} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\mathrm{off}}$$

where

$$\|J_M\|_{1,\mathrm{off}} := \sum_{i \neq j} |\,(J_M)_{ij}\,|.$$

Max-entropy Formulation (Lagrangian Dual)

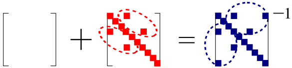
$$\widehat{\Sigma}_M := \operatorname*{argmax}_{\Sigma_M \succ 0} \log\det \Sigma_M$$

$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M\|_{\infty,\mathrm{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d$$

Consistent Estimation Under Certain Conditions, $\boxed{n = \Omega(\log p)}$

Extension to Markov+Independence Models?

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}}.$$



Sparse Covariance Estimation

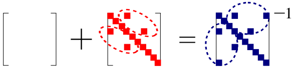
Hard-thresholding the off-diagonal entries of $\widehat{\Sigma}^n$.

Sparse Inverse Covariance Estimation

Add ℓ_1 penalty to maximum likelihood program (involving inverse covariance matrix estimation)

Extension to Markov+Independence Models?

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$



Sparse Covariance Estimation

Hard-thresholding the off-diagonal entries of $\widehat{\Sigma}^n$.

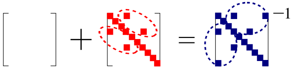
Sparse Inverse Covariance Estimation

Add ℓ_1 penalty to maximum likelihood program (involving inverse covariance matrix estimation)

Is it possible to unify above methods and guarantees?

Extension to Markov+Independence Models?

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$



Sparse Covariance Estimation

Hard-thresholding the off-diagonal entries of $\widehat{\Sigma}^n$.

Sparse Inverse Covariance Estimation

Add ℓ_1 penalty to maximum likelihood program (involving inverse covariance matrix estimation)

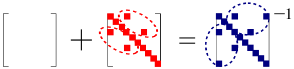
Is it possible to unify above methods and guarantees?

Challenges and Insights

- Penalties in above methods are in different domains

Extension to Markov+Independence Models?

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$



Sparse Covariance Estimation

Hard-thresholding the off-diagonal entries of $\widehat{\Sigma}^n$.

Sparse Inverse Covariance Estimation

Add ℓ_1 penalty to maximum likelihood program (involving inverse covariance matrix estimation)

Is it possible to unify above methods and guarantees?

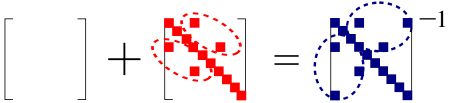
Challenges and Insights

- Penalties in above methods are in different domains
- Insight: Consider dual program of MLE

  Dual program is in covariance domain for Markov model.

Our Algorithm: Covariance Decomposition

- $\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}$.
- Extend ℓ_1-penalized MLE



Max-entropy Formulation

- Lagrangian dual of ℓ_1-penalized MLE

$$\widehat{\Sigma}_M \quad := \underset{\Sigma_M \succ 0}{\operatorname{argmax}} \ \log\det \Sigma_M$$

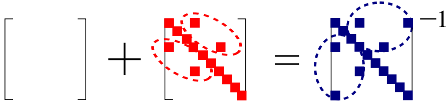
$$\text{s. t.} \ \ \|\widehat{\Sigma}^n - \Sigma_M \quad \|_{\infty,\text{off}} \leq \gamma, \ \ \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d \qquad .$$

ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

Our Algorithm: Covariance Decomposition

- $\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}$.
- Extend ℓ_1-penalized MLE



Max-entropy Formulation + ℓ_1-penalized Residuals (This work)

- Lagrangian dual of ℓ_1-penalized MLE

$$\widehat{\Sigma}_M := \underset{\Sigma_M \succ 0}{\operatorname{argmax}} \ \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

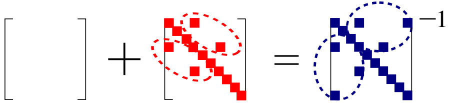
$$\text{s. t.} \quad \|\widehat{\Sigma}^n - \Sigma_M \|_{\infty,\text{off}} \leq \gamma, \ \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d .$$

ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

Our Algorithm: Covariance Decomposition

- $\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}$.
- Extend ℓ_1-penalized MLE



Max-entropy Formulation + ℓ_1-penalized Residuals (This work)

- Lagrangian dual of ℓ_1-penalized MLE

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \underset{\Sigma_M \succ 0, \Sigma_R}{\operatorname{argmax}} \; \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

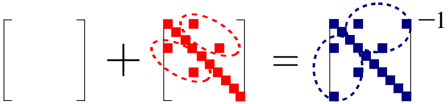
$$\text{s. t.} \quad \|\widehat{\Sigma}^n - \Sigma_M + \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \; \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

Our Algorithm: Covariance Decomposition

- $\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}$.
- Extend ℓ_1-penalized MLE



Max-entropy Formulation + ℓ_1-penalized Residuals (This work)

- Lagrangian dual of ℓ_1-penalized MLE

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \underset{\Sigma_M \succ 0, \Sigma_R}{\operatorname{argmax}} \ \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

$$\text{s.t.} \ \ \|\widehat{\Sigma}^n - \Sigma_M + \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \ \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

$\ell_1 - \ell_\infty$-penalized MLE (This work)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

$$\text{s.t.} \ \|J_M\|_{\infty,\text{off}} \leq \lambda.$$

Observations regarding the Proposed Method

$\ell_1 - \ell_\infty$-penalized MLE (Primal)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}, \ \text{s.t.} \ \|J_M\|_{\infty,\text{off}} \leq \lambda$$

Max-entropy Markov + ℓ_1-penalized Residuals (Dual)

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \underset{\Sigma_M \succ 0, \Sigma_R}{\operatorname{argmax}} \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M + \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

Observations regarding the Proposed Method

$\ell_1 - \ell_\infty$-penalized MLE (Primal)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}, \ \text{s.t.} \ \|J_M\|_{\infty,\text{off}} \leq \lambda$$

Max-entropy Markov + ℓ_1-penalized Residuals (Dual)

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \underset{\Sigma_M \succ 0, \Sigma_R}{\operatorname{argmax}} \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M + \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

Case: $\lambda \to 0$ (Sparse Covariance Estimation)

- $\lambda = \sqrt{\log p/n}$ reduces to approximate shrinkage estimator.

Observations regarding the Proposed Method

$\ell_1 - \ell_\infty$-penalized MLE (Primal)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}, \ \text{s.t.}\ \|J_M\|_{\infty,\text{off}} \le \lambda$$

Max-entropy Markov + ℓ_1-penalized Residuals (Dual)

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \underset{\Sigma_M \succ 0, \Sigma_R}{\operatorname{argmax}} \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

$$\text{s.t.}\quad \|\widehat{\Sigma}^n - \Sigma_M + \Sigma_R\|_{\infty,\text{off}} \le \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

Case: $\lambda \to 0$ (Sparse Covariance Estimation)

- $\lambda = \sqrt{\log p/n}$ reduces to approximate shrinkage estimator.

Case: $\lambda \to \infty$ (Sparse Inverse Covariance Estimator)

- Residual matrix $\widehat{\Sigma}_R = 0$: ℓ_1-penalized MLE of Ravikumar et. al

Observations regarding the Proposed Method

$\ell_1 - \ell_\infty$-penalized MLE (Primal)

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}, \ \text{s.t.} \ \|J_M\|_{\infty,\text{off}} \leq \lambda$$

Max-entropy Markov + ℓ_1-penalized Residuals (Dual)

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \underset{\Sigma_M \succ 0, \Sigma_R}{\operatorname{argmax}} \log\det \Sigma_M - \lambda \|\Sigma_R\|_{1,\text{off}}$$

$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M + \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

Case: $\lambda \to 0$ (Sparse Covariance Estimation)

- $\lambda = \sqrt{\log p/n}$ reduces to approximate shrinkage estimator.

Case: $\lambda \to \infty$ (Sparse Inverse Covariance Estimator)

- Residual matrix $\widehat{\Sigma}_R = 0$: ℓ_1-penalized MLE of Ravikumar et. al

---

Unification of Sparse Covariance & Inverse Covariance models.

Analysis under Exact Statistics

Analysis under Exact Statistics

Similar algorithm as sample statistics, only $\gamma = 0$ (no penalization):

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \Sigma^*, J_M \rangle - \log\det J_M$$
$$\text{s. t. } \|J_M\|_{\infty,\text{off}} \leq \lambda.$$

Analysis under Exact Statistics

Similar algorithm as sample statistics, only $\gamma = 0$ (no penalization):

$$\widehat{J}_M := \operatorname*{argmin}_{J_M \succ 0} \langle \Sigma^*, J_M \rangle - \log\det J_M$$
$$\text{s.t. } \|J_M\|_{\infty,\text{off}} \leq \lambda.$$

KKT conditions results identifiability conditions.

Analysis under Exact Statistics

Similar algorithm as sample statistics, only $\gamma = 0$ (no penalization):

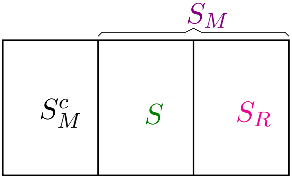
$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \Sigma^*, J_M \rangle - \log\det J_M$$
$$\text{s.t. } \|J_M\|_{\infty,\text{off}} \leq \lambda.$$

KKT conditions results identifiability conditions.

- The main identifiability condition: $\mathrm{Supp}(\Sigma_R^*) \subseteq \mathrm{Supp}(J_M^*)$.

Node pairs are partitioned as follows:

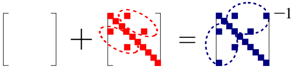
$$S_M := \mathrm{Supp}(J_M^*)$$
$$S_R := \mathrm{Supp}(\Sigma_R^*)$$
$$S := S_M \setminus S_R$$



Outline

Guarantees for High-Dimensional Estimation

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$

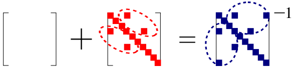


Conditions for Recovery

- Maximum degree Δ in the Markov graph (corresponding to J_M^*).
- Number of samples n, number of nodes p satisfy $\boxed{n = \Omega(\Delta^2 \log p).}$
- Mutual-Incoherence type conditions.

Guarantees for High-Dimensional Estimation

$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}.}$$



Conditions for Recovery

- Maximum degree Δ in the Markov graph (corresponding to J_M^*).
- Number of samples n, number of nodes p satisfy $\boxed{n = \Omega(\Delta^2 \log p).}$
- Mutual-Incoherence type conditions.

Theorem

The proposed method outputs estimates $(\widehat{J}_M, \widehat{\Sigma}_R)$ such that (w.h.p.)

- $(\widehat{J}_M, \widehat{\Sigma}_R)$ are sparsistent and sign consistent.
- satisfy norm guarantees.

$$\boxed{\|\widehat{J}_M - J_M^*\|_\infty, \|\widehat{\Sigma}_R - \Sigma_R^*\|_\infty = O\left(\sqrt{\log p/n}\right).}$$

Guarantee Sparsistency and Efficient Estimation in Both Domains

Observations

Corollary 1 (Sparse Covariance Estimation)

With $\boxed{\lambda = \Theta(\sqrt{\log p/n})}$, our method reduces to shrinkage estimator (comparable to Bickel & Levina which is hard-threshold estimator) and is sparsistent for covariance estimation.

Corollary 2 (Sparse Inverse Covariance Estimation)

With $\boxed{\lambda \to \infty}$, our method reduces to ℓ_1-penalized MLE (Ravikumar et. al) and is sparsistent for inverse covariance estimation.

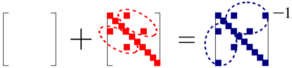
Conditions for Recovery

- Mutual incoherence-type conditions
- Sample complexity $\boxed{n = \Omega(\Delta^2 \log p)}$.
- Comparable to inverse covariance estimation (Ravikumar et. al).

Outline

Synthetic Data

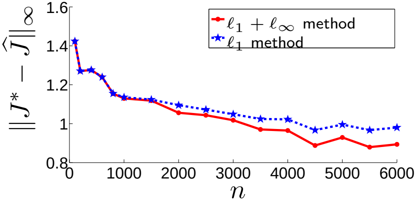
$$\boxed{\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}, \quad J^* = (\Sigma^*)^{-1}.}$$



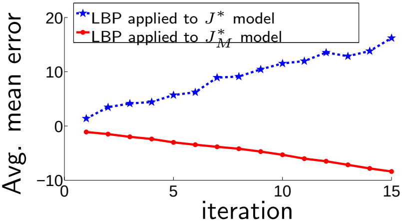
Setup

- 8×8 2-d grid for Markov model.
- Mixed Markov model (both positive and negative correlations).

J estimation



Performance under LBP



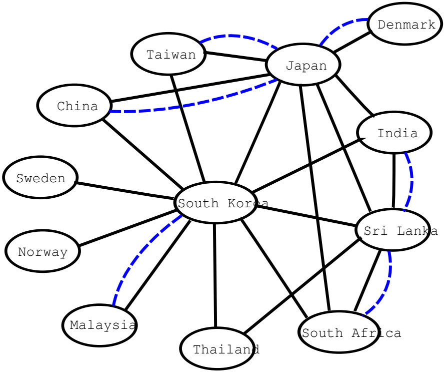
Learned model is amenable for efficient Inference.

Advantage over existing techniques.

Experiments on Foreign Exchange Rate Data

Setup

- Monthly Foreign Exchange Rates to US Dollar.
- Apply the proposed method.

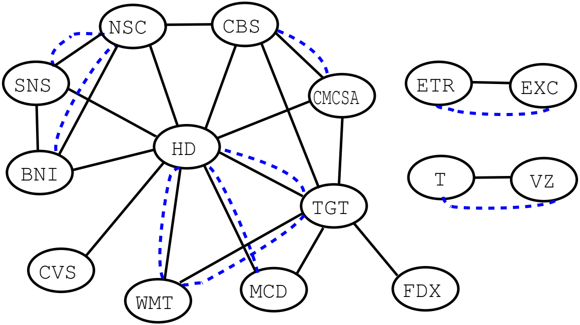


- Solid line: Markov graph. Dotted line: Independence graph.

Experiments on Stock Market Data

Setup

- Monthly stock returns of companies on S&P index.
- Companies in divisions E.Trans, Comm, Elec&Gas and G.Retail Trade.
- Apply the proposed method.



- Solid line: Markov graph. Dotted line: Independence graph.

Outline

Analysis under Sample Statistics

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

$$\text{s. t. } \|J_M\|_{\infty,\text{off}} \le \lambda.$$

Analysis under Sample Statistics

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$
$$\text{s. t. } \|J_M\|_{\infty,\text{off}} \leq \lambda.$$

- Challenges
  1) Sparsistency guarantee: hard to show $\operatorname{Supp}(\widehat{J}_M) \subseteq \operatorname{Supp}(J^*_M)$.

Analysis under Sample Statistics

$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$
$$\text{s. t. } \|J_M\|_{\infty,\text{off}} \leq \lambda.$$

- Challenges
  1) Sparsistency guarantee: hard to show $\operatorname{Supp}(\widehat{J}_M) \subseteq \operatorname{Supp}(J_M^*)$.
  2) Decoupling the errors: $\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}$.

Analysis under Sample Statistics

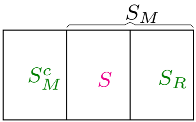
$$\widehat{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$
$$\text{s. t. } \|J_M\|_{\infty,\text{off}} \le \lambda.$$

- Challenges
  1) Sparsistency guarantee: hard to show $\operatorname{Supp}(\widehat{J}_M) \subseteq \operatorname{Supp}(J_M^*)$.
  2) Decoupling the errors: $\Sigma^* + \Sigma_R^* = {J_M^*}^{-1}$.

- Proposing a modified version which is easier to analyze.

Modified Program (Restricted and Relaxed)

$$\widetilde{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$
$$\text{s. t. } \big(J_M\big)_{S_M^c} = 0, \; \big(J_M\big)_{S_R} = \lambda \operatorname{sign}\Big(\big(J_M^*\big)_{S_R}\Big).$$



Primal-Dual Witness Method

$$\widetilde{J}_M := \operatorname*{argmin}_{J_M \succ 0} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

$$\text{s. t. } \left(J_M\right)_{S_M^c} = 0, \ \left(J_M\right)_{S_R} = \lambda \operatorname{sign}\Big(\left(J_M^*\right)_{S_R}\Big).$$

Primal-Dual Witness Method

$$\widetilde{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \ \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\mathrm{off}}$$

$$\text{s. t. } \ \big(J_M\big)_{S_M^c} = 0, \ \big(J_M\big)_{S_R} = \lambda \,\mathrm{sign}\Big(\big(J_M^*\big)_{S_R}\Big).$$

Sparsistency Guarantee

$\mathrm{Supp}(\widetilde{J}_M) \subseteq \mathrm{Supp}(J_M^*)$
$\mathrm{Supp}(\widetilde{\Sigma}_R) \subseteq \mathrm{Supp}(\Sigma_R^*)$

Primal-Dual Witness Method

$$\widetilde{J}_M := \operatorname*{argmin}_{J_M \succ 0} \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

$$\text{s. t.} \ \left(J_M\right)_{S_M^c} = 0, \ \left(J_M\right)_{S_R} = \lambda \operatorname{sign}\Big(\left(J_M^*\right)_{S_R}\Big).$$

Sparsistency Guarantee

$$\operatorname{Supp}(\widetilde{J}_M) \subseteq \operatorname{Supp}(J_M^*)$$
$$\operatorname{Supp}(\widetilde{\Sigma}_R) \subseteq \operatorname{Supp}(\Sigma_R^*)$$

Error Decoupling

$$\widetilde{\Delta}_J := \widetilde{J}_M - J_M^*, \quad \widetilde{\Delta}_R := \widetilde{\Sigma}_R - \Sigma_R^*$$

	$\underbrace{\hspace{4em}}_{S_M}$	
$(\widetilde{\Delta}_J) = 0$	$(\widetilde{\Delta}_R) = 0$	$(\widetilde{\Delta}_J) = \lambda_\delta$
S_M^c	S	S_R

Primal-Dual Witness Method

$$\widetilde{J}_M := \underset{J_M \succ 0}{\operatorname{argmin}} \; \langle \widehat{\Sigma}^n, J_M \rangle - \log\det J_M + \gamma \|J_M\|_{1,\text{off}}$$

$$\text{s.t.} \; (J_M)_{S_M^c} = 0, \; (J_M)_{S_R} = \lambda \,\text{sign}\Big((J_M^*)_{S_R} \Big).$$

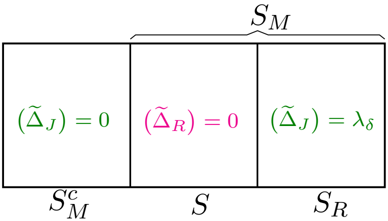
Sparsistency Guarantee

$\text{Supp}(\widetilde{J}_M) \subseteq \text{Supp}(J_M^*)$

$\text{Supp}(\widetilde{\Sigma}_R) \subseteq \text{Supp}(\Sigma_R^*)$

Error Decoupling

$$\widetilde{\Delta}_J := \widetilde{J}_M - J_M^*, \quad \widetilde{\Delta}_R := \widetilde{\Sigma}_R - \Sigma_R^*$$



- Sufficient Conditions for equivalence between the modified and original programs (Mutual Incoeherence): $(\widetilde{J}_M, \widetilde{\Sigma}_R) = (\widehat{J}_M, \widehat{\Sigma}_R)$.

Outline

Conclusion

Summary

- Combination of Markov and independence models
- Unifying sparse covariance/inverse covariance estimation methods
- Efficient method and guarantees for estimation in both domains

Conclusion

Summary

- Combination of Markov and independence models
- Unifying sparse covariance/inverse covariance estimation methods
- Efficient method and guarantees for estimation in both domains

---

Outlook

- Other forms of residuals (e.g. low rank)
- Discrete Model (via pseudo-likelihood)

http://arxiv.org/abs/1211.0919