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 $\mathbf{X} := [X_1, \ldots, X_p]^T$.
- High-dimensional regime: both $n, p \to \infty$ and $n \ll p$.
- Covariance estimation:

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

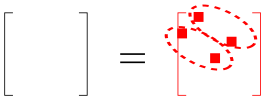
- Challenge: empirical (sample) covariance ill-posed when $n \ll p$:

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

Solution: Imposing Sparsity for Tractable High-dimensional Estimation

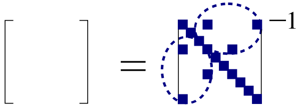
Incorporating Sparsity in High Dimensions

Sparse Covariance



Σ^* Σ_R

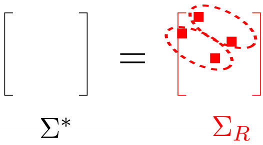
Sparse Inverse Covariance



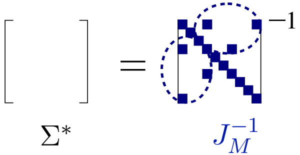
Σ^* J_M^{-1}

Incorporating Sparsity in High Dimensions

Sparse Covariance



Sparse Inverse Covariance

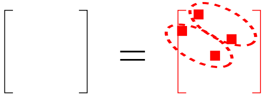


Relationship with Statistical Properties (Gaussian)

- Sparse Covariance= Independence Model: marginal independence.
- Sparse Inverse Covariance=Markov Model: conditional independence

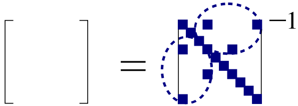
Incorporating Sparsity in High Dimensions

Sparse Covariance



Σ^* Σ_R

Sparse Inverse Covariance



Σ^* J_M^{-1}

Relationship with Statistical Properties (Gaussian)

- Sparse Covariance= Independence Model: marginal independence.
- Sparse Inverse Covariance=Markov Model: conditional independence

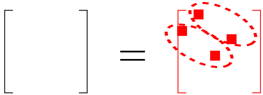
---

Guarantees under Sparsity Constraints in High Dimensions

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

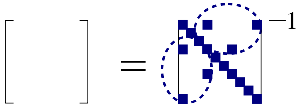
Incorporating Sparsity in High Dimensions

Sparse Covariance



Σ^* Σ_R

Sparse Inverse Covariance



Σ^* 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.

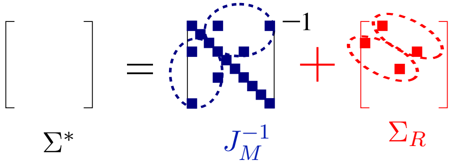
Going Beyond Sparse Models

Motivation

- Sparsity constraints restrictive to have faithful representation.
- Data not sparse in a single domain
- Solution: Sparsity in Multiple Domains.

One Possibility: Sparse Markov + Sparse Independence Models

- Sparsity in Multiple Domains: Multiple Statistical Relationships.



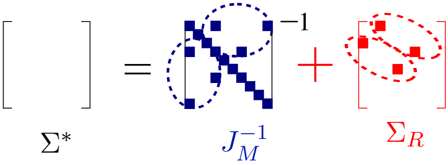
Going Beyond Sparse Models

Motivation

- Sparsity constraints restrictive to have faithful representation.
- Data not sparse in a single domain
- Solution: Sparsity in Multiple Domains.

One Possibility: Sparse Markov + Sparse Independence Models

- Sparsity in Multiple Domains: Multiple Statistical Relationships.



Efficient Decomposition and Estimation in High Dimensions?

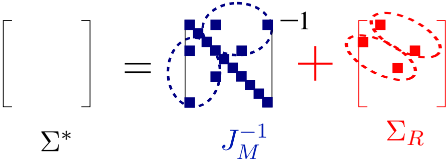
Going Beyond Sparse Models

Motivation

- Sparsity constraints restrictive to have faithful representation.
- Data not sparse in a single domain
- Solution: Sparsity in Multiple Domains.

One Possibility: Sparse Markov + Sparse Independence Models

- Sparsity in Multiple Domains: Multiple Statistical Relationships.

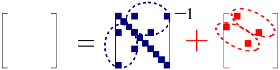


Efficient Decomposition and Estimation in High Dimensions?

Unique Decomposition? Good Sample Requirements?

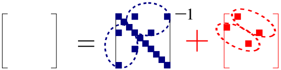
Summary of Results

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



Summary of Results

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

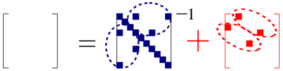


Contribution 1: Novel Method for Decomposition

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

Summary of Results

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



Contribution 1: Novel Method for Decomposition

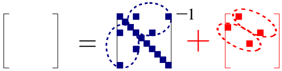
- Decomposition into Markov and residual domains.
- Unification of Sparse Covariance and Inverse Covariance Estimation.

Contribution 2: Guarantees for Estimation

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

Summary of Results

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



Contribution 1: Novel Method for Decomposition

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

Contribution 2: Guarantees for Estimation

- 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.

Efficient Method for Covariance Decomposition and Estimation

Related Works

Sparse Covariance/Inverse Covariance Estimation

- Sparse Covariance Estimation: Covariance Thresholding.
  - (Bickel & Levina) (Wagaman & Levina) (Cai 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)

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

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)

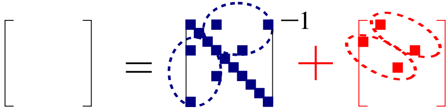
- Decomposition in inverse covariance domain
- Lack theoretical guarantees

Our contribution: Guaranteed Decomposition and Estimation

Outline

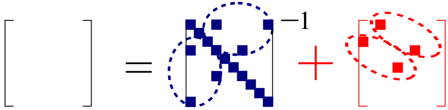
Some Intuitions and Ideas

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



Some Intuitions and Ideas

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

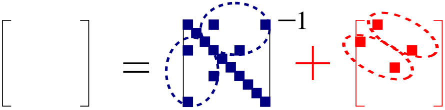


Review Ideas for Special Cases: Sparse Covariance/Inverse Covariance

---

Some Intuitions and Ideas

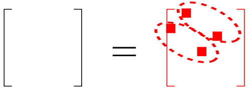
- $\Sigma^* = {J_M^*}^{-1} + \Sigma_R^*$.
- $\widehat{\Sigma}^n$: sample covariance using n i.i.d. samples



Review Ideas for Special Cases: Sparse Covariance/Inverse Covariance

Sparse Covariance Estimation (Independence Model)

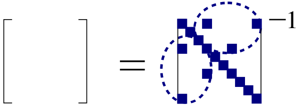
- $\Sigma^* = \Sigma_R^*$.
- $\widehat{\Sigma}^n$: sample covariance using n samples
- p variables: $p \gg n$.



- Thresholding estimator for off-diagonals (Bickel & Levina): threshold chosen as $\boxed{\sqrt{\frac{\log p}{n}}}$
- Sparsistency (support recovery) and Norm guarantees when $\boxed{n = \Omega(\log p)} \Rightarrow n \ll p$.

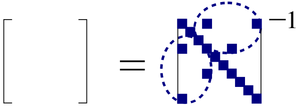
Recap of Inverse Covariance (Markov) Estimation

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



Recap of Inverse Covariance (Markov) Estimation

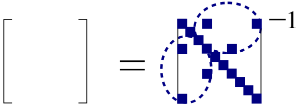
- $\Sigma^* = {J_M^*}^{-1}$
- $\widehat{\Sigma}^n$: sample covariance using n i.i.d. samples



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

Recap of Inverse Covariance (Markov) Estimation

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

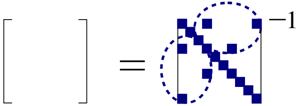


ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al. '08)

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

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)

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

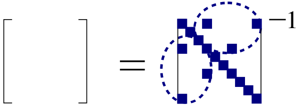
Max-entropy Formulation (Lagrangian Dual)

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

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

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)

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

Max-entropy Formulation (Lagrangian Dual)

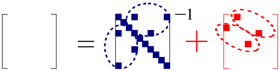
$$\widehat{\Sigma}_M := \underset{\Sigma_M \succ 0}{\operatorname{argmax}} \log\det \Sigma_M$$

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

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

Extension to Markov+Independence Models?

$$\Sigma^* = {J^*_M}^{-1} + \Sigma^*_R.$$



Sparse Covariance Estimation

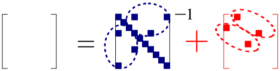
Threshold off-diagonal entries of $\widehat{\Sigma}^n$.

Sparse Inverse Covariance Estimation

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

Extension to Markov+Independence Models?

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



Sparse Covariance Estimation

Threshold off-diagonal entries of $\widehat{\Sigma}^n$.

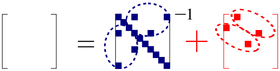
Sparse Inverse Covariance Estimation

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?

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



Sparse Covariance Estimation

Threshold off-diagonal entries of $\widehat{\Sigma}^n$.

Sparse Inverse Covariance Estimation

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

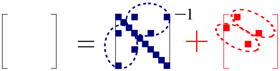
Is it possible to unify above methods and guarantees?

Challenges and Insights

- Penalties in above methods are in different domains

Extension to Markov+Independence Models?

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



Sparse Covariance Estimation

Threshold off-diagonal entries of $\widehat{\Sigma}^n$.

Sparse Inverse Covariance Estimation

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

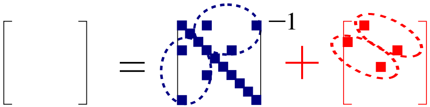
Is it possible to unify above methods and guarantees?

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

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



Max-entropy Formulation

- Lagrangian dual of ℓ_1-penalized MLE

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

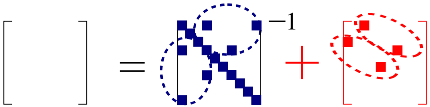
$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M\|_{\infty,\text{off}} \leq \gamma, \quad \big(\Sigma_M\big)_d = \big(\widehat{\Sigma}^n\big)_d$$

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

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

Our Algorithm: Covariance Decomposition

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



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

- Lagrangian dual of ℓ_1-penalized MLE

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

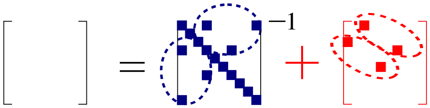
$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d$$

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

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

Our Algorithm: Covariance Decomposition

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



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

- Lagrangian dual of ℓ_1-penalized MLE

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

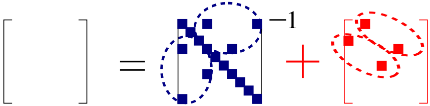
$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M - \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d$$

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

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



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

- Lagrangian dual of ℓ_1-penalized MLE

$$(\widehat{\Sigma}_M, \widehat{\Sigma}_R) := \operatorname*{argmax}_{\Sigma_M \succ 0, \Sigma_R} \log\det \Sigma_M$$

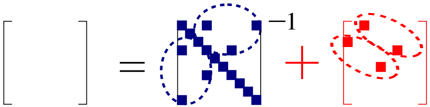
$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M - \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

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

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

Our Algorithm: Covariance Decomposition

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



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

- Lagrangian dual of ℓ_1-penalized MLE

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

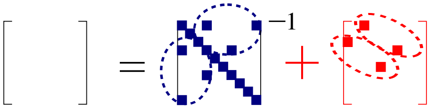
$$\text{s.t.} \quad \|\widehat{\Sigma}^n - \Sigma_M - \Sigma_R\|_{\infty,\text{off}} \leq \gamma, \quad \left(\Sigma_M\right)_d = \left(\widehat{\Sigma}^n\right)_d, \left(\Sigma_R\right)_d = 0.$$

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

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



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

- Lagrangian dual of ℓ_1-penalized MLE

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

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

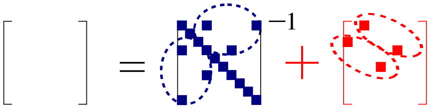
ℓ_1-MLE for Sparse Inverse Covariance (Ravikumar et. al)

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

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

Our Algorithm: Covariance Decomposition

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



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

- Lagrangian dual of ℓ_1-penalized MLE

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

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

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

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

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

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)

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

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)

- Threshold estimator for off-diagonals of Σ_R^* (under exact statistics)
- With samples, $\lambda = \sqrt{\log p/n}$ reduces to threshold estimator.

Observations regarding the Proposed Method

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

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

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

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

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

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

- Threshold estimator for off-diagonals of Σ_R^* (under exact statistics)
- With samples, $\lambda = \sqrt{\log p/n}$ reduces to threshold 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)

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

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

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

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

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

- Threshold estimator for off-diagonals of Σ_R^* (under exact statistics)
- With samples, $\lambda = \sqrt{\log p/n}$ reduces to threshold estimator.

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

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

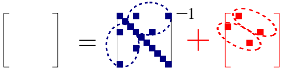
---

Unification of Sparse Covariance & Inverse Covariance Estimation

Outline

Guarantees for High-Dimensional Estimation

$$\Sigma^* = {J^*_M}^{-1} + \Sigma^*_R.$$



Conditions for Recovery

- Maximum degree Δ in the Markov graph (corresponding to J^*_M).
- Number of samples n, number of nodes p satisfy $n = \Omega(\Delta^2 \log p)$.
- Regularization constant: $\lambda = \max_{i \neq j} J^*_M(i,j) + \Theta(\sqrt{\log p/n})$.

Theorem

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

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

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

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 threshold estimator (Bickel & Levina) 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.

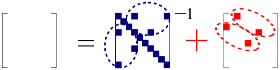
Conditions for Recovery

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

Outline

Synthetic Data

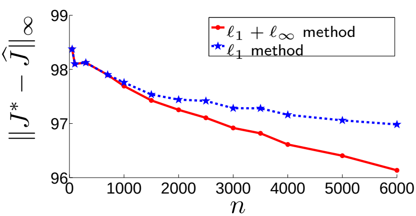
$$\Sigma^* = J_M^{*\,-1} + \Sigma_R^*, \quad J^* = (\Sigma^*)^{-1}.$$



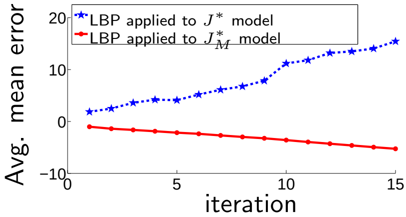
Setup

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

J estimation



Performance under LBP

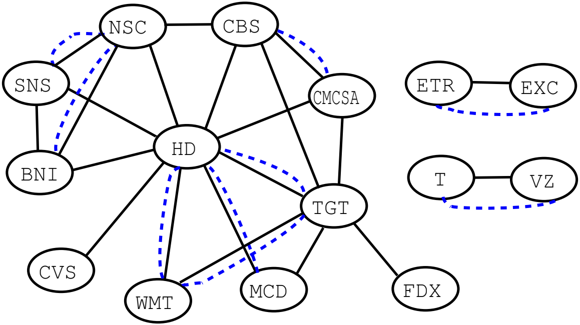


Advantage over existing techniques.

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

Conclusion

Summary

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

Conclusion

Summary

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

Not covered in this talk

- Analysis under Exact Statistics: Conditions for Unique Decomposition.
- Sample Analysis: Careful control of perturbations in both domains.

Longer version available on webpage.

Conclusion

Summary

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

Not covered in this talk

- Analysis under Exact Statistics: Conditions for Unique Decomposition.
- Sample Analysis: Careful control of perturbations in both domains.

Longer version available on webpage.

Outlook

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