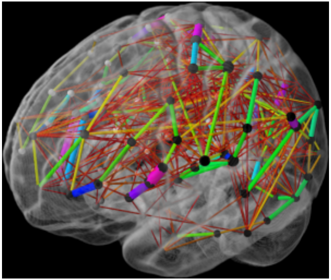
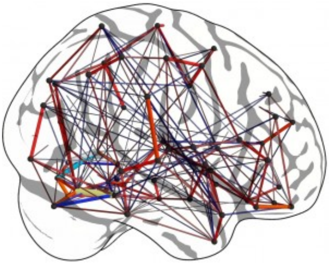
Graph Matching, Graph Inference and Applications

Marcelo Fiori - Guillermo Sapiro

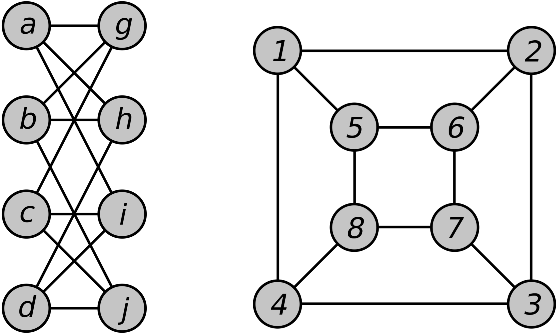
October 2013

Graphs in brain imaging

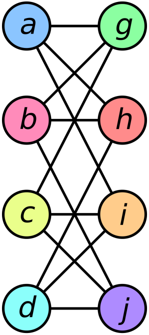


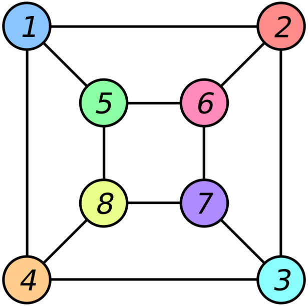


The Graph Matching Problem



The Graph Matching Problem





The Graph Matching Problem

Given two graphs G_A and G_B:

- Graph Isomorphism Problem (GIP):
  determine whether the two graphs are isomorphic
  - Complexity unsolved.
- Graph Matching Problem (GMP):
  find the isomorphism between the two graphs

The Graph Matching Problem

Given two graphs G_A and G_B:

- Graph Isomorphism Problem (GIP):
  determine whether the two graphs are isomorphic
  - Complexity unsolved.
- Graph Matching Problem (GMP):
  find the isomorphism between the two graphs

The Graph Matching Problem

Given two graphs G_A and G_B:

- Graph Isomorphism Problem (GIP):
  determine whether the two graphs are isomorphic
  - Complexity unsolved.
- Graph Matching Problem (GMP):
  find the isomorphism between the two graphs

The Graph Matching Problem

In terms of the adjacency matrices A and B:

- Find $P \in \mathcal{P}$ such that $A = PBP^T$

The Graph Matching Problem

In terms of the adjacency matrices A and B:

- Find $P \in \mathcal{P}$ such that $A = PBP^T$

$$\hat{P} = \arg\min_{P \in \mathcal{P}} ||AP - PB||_F^2,$$

The Graph Matching Problem

In terms of the adjacency matrices A and B:

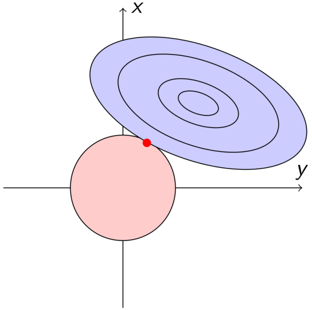
- Find $P \in \mathcal{P}$ such that $A = PBP^T$

$$\hat{P} = \arg \min_{P \in \underset{\mathcal{D}}{\cancel{\mathcal{P}}}} ||AP - PB||_F^2,$$

Convex relaxation

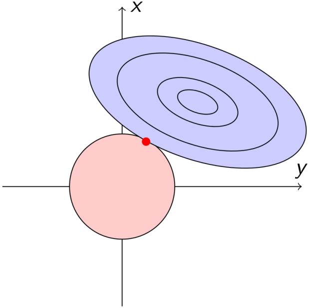
Two words on sparsity

$$\min_x f(x) \qquad s.t. ||x||_2 \leq k$$

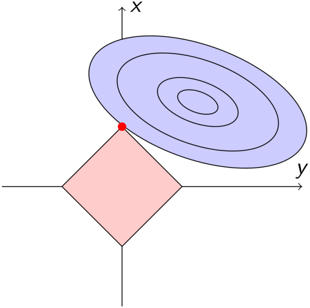


Two words on sparsity

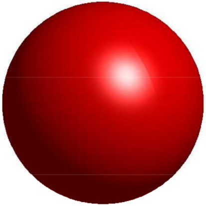
$$\min_x f(x) \quad s.t. ||x||_2 \leq k$$



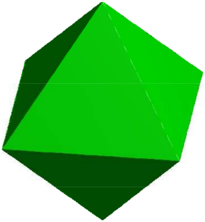
$$\min_x f(x) \quad s.t. ||x||_1 \leq k$$



Group Sparsity

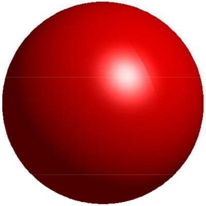


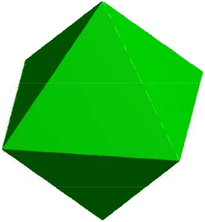
(a) ℓ_2-norm ball

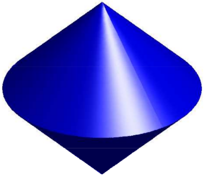


(b) ℓ_1-norm ball

Group Sparsity

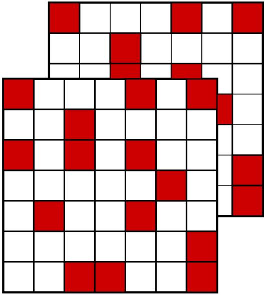

(a) ℓ_2-norm ball


(b) ℓ_1-norm ball


ℓ_1/ℓ_2-norm ball

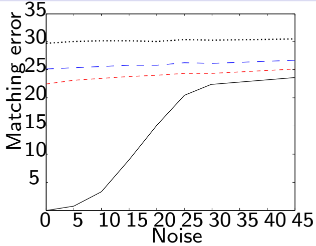
Graph Matching meets sparsity

Roubust multimodal graph matching formulation

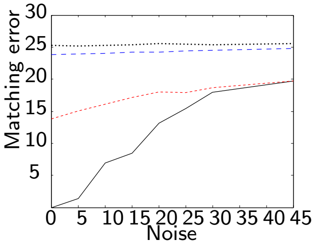


$$\tilde{P} = \arg\min_{P \in \mathcal{D}} \sum_{i,j} \left|\left|\big((AP)_{ij}, (PB)_{ij}\big)\right|\right|_2$$

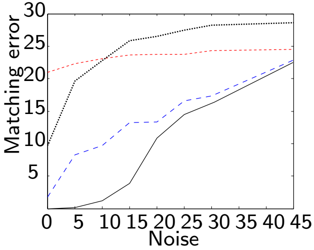
Multimodal results



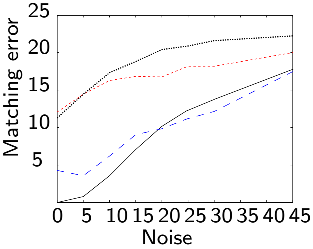
(a) Erdős-Rényi graphs



(b) Scale-free graphs



(c) Erdős-Rényi graphs



(d) Scale-free graphs

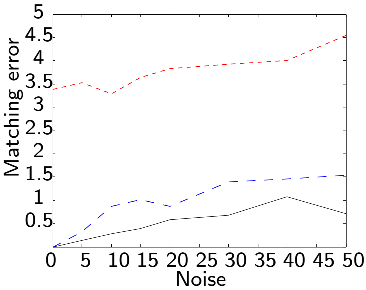
Figure: Graphs with $p = 100$ nodes. In (a) and (b), weights $\mathcal{N}(1, 0.4)$ and $\mathcal{N}(4, 1)$. In (c) and (d), weights $\mathcal{N}(1, 0.4)$ and uniform in $[1, 2]$.

Application: C. elegans connectome

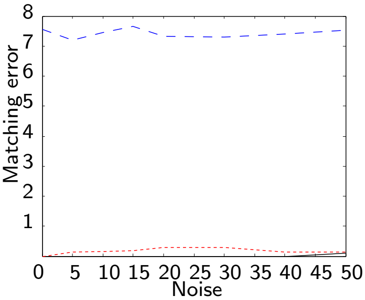
- Somatic nervous system consists of 279 neurons
- The two types of connections (chemical and electrical) between these 279 neurons have been mapped
- Corresponding adjacency matrices, A_c and A_e, are publicly available.

Application: C. elegans connectome

We match both the chemical and the electrical connection graphs against noisy artificially permuted versions of them.



(a) Electrical connection graph



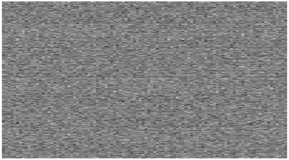
(b) Chemical connection graph

The Inverse Covariance Matrix

- $(X_1, \ldots, X_p) \sim N(0, \Sigma)$
- $k \times p$ data matrix X (k independent observations)
- Goal: infer the support of Σ^{-1}
- Property: If X_i y X_j are conditionally independent $\Rightarrow \Sigma^{-1}_{ij} = 0$
- Σ^{-1} known to be sparse in numerous applications

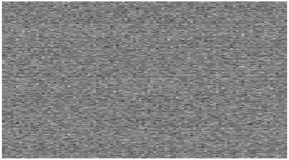
The Inverse Covariance Matrix

- $(X_1, \ldots, X_p) \sim N(0, \Sigma)$
- $k \times p$ data matrix X (k independent observations)
- Goal: infer the support of Σ^{-1}
- Property: If X_i y X_j are conditionally independent $\Rightarrow \Sigma^{-1}_{ij} = 0$
- Σ^{-1} known to be sparse in numerous applications



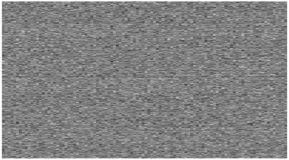
The Inverse Covariance Matrix

- $(X_1, \ldots, X_p) \sim N(0, \Sigma)$
- $k \times p$ data matrix X (k independent observations)
- Goal: infer the support of Σ^{-1}
- Property: If X_i y X_j are conditionally independent $\Rightarrow \Sigma^{-1}_{ij} = 0$
- Σ^{-1} known to be sparse in numerous applications



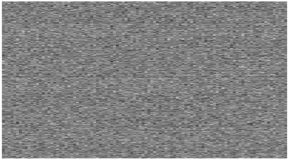
The Inverse Covariance Matrix

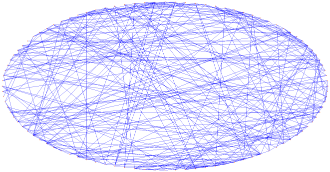
- $(X_1, \ldots, X_p) \sim N(0, \Sigma)$
- $k \times p$ data matrix X (k independent observations)
- Goal: infer the support of Σ^{-1}
- Property: If X_i y X_j are conditionally independent $\Rightarrow \Sigma^{-1}_{ij} = 0$
- Σ^{-1} known to be sparse in numerous applications



The Inverse Covariance Matrix

- $(X_1, \ldots, X_p) \sim N(0, \Sigma)$
- $k \times p$ data matrix X (k independent observations)
- Goal: infer the support of Σ^{-1}
- Property: If X_i y X_j are conditionally independent $\Rightarrow \Sigma^{-1}_{ij} = 0$
- Σ^{-1} known to be sparse in numerous applications





The Inverse Covariance Matrix

Why conditional dependence?

Suppose we have $\varepsilon_i \sim \mathcal{N}(0,1)$ and:

$$\begin{aligned} x &= z + \varepsilon_1 \\ y &= z + \varepsilon_2 \\ z &= \varepsilon_3 \end{aligned}$$

Then:

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$

The Inverse Covariance Matrix

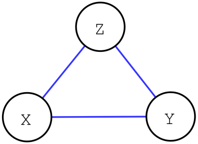
Why conditional dependence?

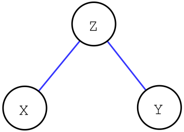
Suppose we have $\varepsilon_i \sim \mathcal{N}(0,1)$ and:

$$\begin{aligned} x &= z + \varepsilon_1 \\ y &= z + \varepsilon_2 \\ z &= \varepsilon_3 \end{aligned}$$

Then:

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$





Graphical Lasso

Maximum likelihood estimator for Σ^{-1} with an l_1 regularization:

$$\min_{\Theta \succ 0} \ \mathsf{tr}(S\Theta) - \log\det\Theta + \lambda \sum_{i,j} |\Theta_{ij}|$$

S: empirical covariance matrix

Collaborative Graphical Lasso

Goal: infer several graphs with the same structure

$$\min_{\substack{\Theta^A \succ 0 \\ \Theta^B \succ 0}} \mathsf{tr}(S^A \Theta^A) - \log\det\Theta^A + \mathsf{tr}(S^B \Theta^B) - \log\det\Theta^B + \lambda \sum_{i,j} ||(\Theta^A_{ij}, \Theta^B_{ij})||_2$$

Collaborative Graphical Lasso

Application to fMRI data

- rs-fMRI data from A. Hariri
- data matrix $X_i \in \mathcal{M}_{n \times p}$
- $i = 1 \ldots 155$ subjects
- n time points
- p regions or voxels

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph (A_M or A_F)

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph (A_M or A_F)

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph $(A_M$ or $A_F)$

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph $(A_M$ or $A_F)$

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph (A_M or A_F)

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph $(A_M$ or $A_F)$

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Collaborative Graphical Lasso

Proof of concept example

- split dataset: 105 training and 50 for testing
- build network for males (A_M) and females (A_F)
- for each subject in testing set:
  - build graph from fMRI
  - classify as M/F according to closest graph (A_M or A_F)

Compare to

- nearest neighbor w.r.t. subjects in training set

	Performance
NN	60%
CGL	80%

Combining graph matching with inference

- What if we have not aligned data?
- Jointly learn the graphs and the alignment

$$\min_{\substack{\Theta^A \succ 0 \\ \Theta^B \succ 0 \\ P \in \mathcal{P}}} \operatorname{tr}(S^A\Theta^A) - \log\det\Theta^A + \operatorname{tr}(S^B\Theta^B) - \log\det\Theta^B + \lambda \sum_{i,j} \left|\left|\left((\Theta^A P)_{ij}, (P\Theta^B)_{ij}\right)\right|\right|_2$$

- non-convex problem
- convex when minimized only over (Θ^A, Θ^B) or P leaving the other fixed.

Combining graph matching with inference

- What if we have not aligned data?
- Jointly learn the graphs and the alignment

$$\min_{\substack{\Theta^A \succ 0 \\ \Theta^B \succ 0 \\ P \in \mathcal{P}}} \text{tr}(S^A\Theta^A) - \log\det\Theta^A + \text{tr}(S^B\Theta^B) - \log\det\Theta^B + \lambda \sum_{i,j} \left|\left|\left((\Theta^A P)_{ij}, (P\Theta^B)_{ij}\right)\right|\right|_2$$

- non-convex problem
- convex when minimized only over (Θ^A, Θ^B) or P leaving the other fixed.

Combining graph matching with inference

- What if we have not aligned data?
- Jointly learn the graphs and the alignment

$$\min_{\substack{\Theta^A \succ 0 \\ \Theta^B \succ 0 \\ P \in \mathcal{P}}} \mathsf{tr}(S^A \Theta^A) - \log\det\Theta^A + \mathsf{tr}(S^B \Theta^B) - \log\det\Theta^B + \lambda \sum_{i,j} \left|\left| \left((\Theta^A P)_{ij}, (P\Theta^B)_{ij} \right) \right|\right|_2$$

- non-convex problem
- convex when minimized only over (Θ^A, Θ^B) or P leaving the other fixed.

Another toy example

Data

- same subject undergoing resting-state fMRI in two different sessions separated by a break.
- Each session: 10 minutes of data $\rightarrow$ 900 samples per study.
- two data matrices $X^A, X^B \in \mathbb{R}^{900\times 200}$, test/retest resp.

Using only part of the data in X^A and part of the data in a permuted version of X^B, we are able to infer a connectivity matrix almost as accurately as using the whole data

Another toy example

Data

- same subject undergoing resting-state fMRI in two different sessions separated by a break.
- Each session: 10 minutes of data $\rightarrow$ 900 samples per study.
- two data matrices $X^A, X^B \in \mathbb{R}^{900 \times 200}$, test/retest resp.

Using only part of the data in X^A and part of the data in a permuted version of X^B, we are able to infer a connectivity matrix almost as accurately as using the whole data

Another toy example

Data

- same subject undergoing resting-state fMRI in two different sessions separated by a break.
- Each session: 10 minutes of data $\rightarrow$ 900 samples per study.
- two data matrices $X^A, X^B \in \mathbb{R}^{900\times 200}$, test/retest resp.

Using only part of the data in X^A and part of the data in a permuted version of X^B, we are able to infer a connectivity matrix almost as accurately as using the whole data

Another toy example

Data

- same subject undergoing resting-state fMRI in two different sessions separated by a break.
- Each session: 10 minutes of data $\rightarrow$ 900 samples per study.
- two data matrices $X^A, X^B \in \mathbb{R}^{900\times 200}$, test/retest resp.

Using only part of the data in X^A and part of the data in a permuted version of X^B, we are able to infer a connectivity matrix almost as accurately as using the whole data

Another toy example

"Ground truth"

- the collaborative setting has already been proven successful,
- take as ground truth Θ_{GT}^{A} and Θ_{GT}^{B}
  result of the collaborative inference using the whole data

Mess up the data

- Throw away part of the data:
  - X_H^A: first 550 samples of X^A
  - X_H^B: first 550 samples of X^B
  - a little less than 6 minutes of study
- artificially permute columns in X_H^B:
  - $S_H^B = P_o^T S_H^B P_o$

Jointly infer connectivity graphs and alignment

Another toy example

“Ground truth”

- the collaborative setting has already been proven successful,
- take as ground truth Θ^A_{GT} and Θ^B_{GT} result of the collaborative inference using the whole data

Mess up the data

- Throw away part of the data:
  - X^A_H: first 550 samples of X^A
  - X^B_H: first 550 samples of X^B
  - a little less than 6 minutes of study
- artificially permute columns in X^B_H:
  - $S^B_H = P^T_\sigma S^B_H P_\sigma$

Jointly infer connectivity graphs and alignment

Another toy example

"Ground truth"

- the collaborative setting has already been proven successful,
- take as ground truth Θ_{GT}^A and Θ_{GT}^B
  result of the collaborative inference using the whole data

Mess up the data

- Throw away part of the data:
  - X_H^A: first 550 samples of X^A
  - X_H^B: first 550 samples of X^B
  - a little less than 6 minutes of study
- artificially permute columns in X_H^B:
  - $\tilde{S}_H^B = P_o^T S_H^B P_o$

Jointly infer connectivity graphs and alignment

Another toy example

"Ground truth"

- the collaborative setting has already been proven successful,
- take as ground truth Θ^A_{GT} and Θ^B_{GT}
  result of the collaborative inference using the whole data

Mess up the data

- Throw away part of the data:
  - X^A_H: first 550 samples of X^A
  - X^B_H: first 550 samples of X^B
  - a little less than 6 minutes of study
- artificially permute columns in X^B_H:
  - $\tilde{S}^B_H = P_o^T S^B_H P_o$

Jointly infer connectivity graphs and alignment

Another toy example

“Ground truth”

- the collaborative setting has already been proven successful,
- take as ground truth Θ_{GT}^{A} and Θ_{GT}^{B}
  result of the collaborative inference using the whole data

Mess up the data

- Throw away part of the data:
  - X_H^A: first 550 samples of X^A
  - X_H^B: first 550 samples of X^B
  - a little less than 6 minutes of study
- artificially permute columns in X_H^B:
  - $\tilde{S}_H^B = P_o^T S_H^B P_o$

Jointly infer connectivity graphs and alignment

Results

Compare:

- < 6 min of each study, variables not pre-aligned
  - Computation: Joint Graph Inference + Alignment
- only one of the 10 min studies (test and no retest)
  - Computation: inverse covariance matrix (Graphical Lasso)

Results

Compare:

- < 6 min of each study, variables not pre-aligned
  - Computation: Joint Graph Inference + Alignment
- only one of the 10 min studies (test and no retest)
  - Computation: inverse covariance matrix (Graphical Lasso)

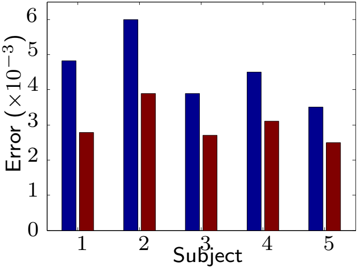


Figure: Blue: error using one complete 10 min study: $||\Theta^A_{GT} - \Theta^A_s||_F$. Red: error $||\Theta^A_{GT} - \Theta^A_H||_F$ with collaborative inference using < 6 min of each study, but solving for the node permutations at the same time.

Thank you!

Fiori, Marcelo, Musé, Pablo, Hariri, Ahamd, and Sapiro, Guillermo.
Multimodal graphical models via group lasso.
Signal Processing with Adaptive Sparse Structured Representations, 2013.

Fiori, Marcelo, Sprechmann, Pablo, Vogelstein, Joshua, Musé, Pablo, and Sapiro, Guillermo.
Robust Multimodal Graph Matching: Sparse Coding Meets Graph Matching.
Advances in Neural Information Processing Systems 26 (NIPS 2013).