Transposable Regularized Covariance Models with Applications to High-Dimensional Data

Genevera I. Allen

May 2010

Thesis Defense, Department of Statistics, Stanford University

Transposable Data

$$\mathbf{X}_{n \times p} =$$

$$\begin{pmatrix} X_{11} & \dots & X_{1p} \\ X_{21} & \dots & X_{2p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}$$

- **Multivariate Data**:
  - Rows: independent observations.
  - Column: features of interest.

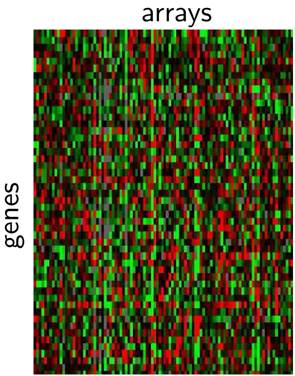
Transposable Data

$$\mathbf{X}_{n \times p} =$$

$$\begin{pmatrix} X_{11} & \dots & X_{1p} \\ X_{21} & \dots & X_{2p} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{np} \end{pmatrix}$$

- **Multivariate Data**:
  - Rows: independent observations.
  - Column: features of interest.
- **Transposable (Matrix) Data**:
  - Rows, columns or both are features.
  - Possible dependencies between and/or among rows and columns.

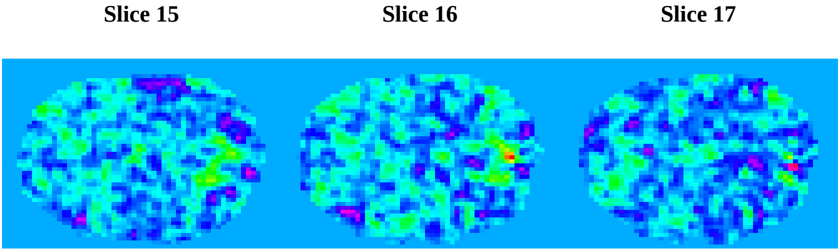
Example: Microarrays



- Rows: Genes ($\approx 10,000$).
- Columns: Arrays (subjects or samples).
- Measurement: Gene expression.

Example: Functional MRIs (fMRI)

- Rows: Voxels.
- Columns: Subjects (And/or replicates and times).
- Measurement: Hemodynamic response (change in blood flow).



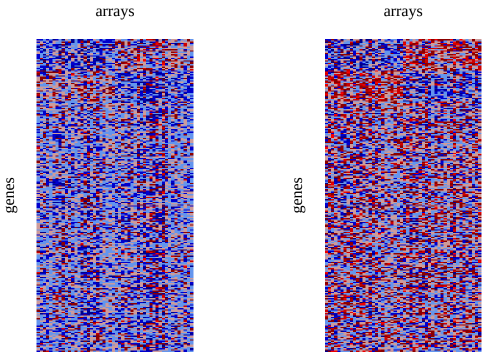
Example: Netflix Movie Rating Data

- Rows: Movies.
- Columns: Customers.
- Measurement: Movie ratings (scale of 1 - 5).

	Anne	Ben	Charlie	Doug	Eve	...
Star Wars	2	5	4	4	3	...
Harry Potter	3	4	5	3	?	...
Pretty Woman	4	?	2	?	5	...
Titanic	5	?	2	1	3	...
Lord of the Rings	?	5	5	4	4	...
⋮	⋮	⋮	⋮	⋮	⋮	⋱

Preview: De-Correlating Microarrays

Data $\quad$ Fit TRCM $\Longrightarrow$ $\quad$ Sphered Data



1. Allows one to reject more truly significant genes.
2. Obtain fewer false positives.

Objective

Goals

1. Develop *flexible* models for dependencies among rows and/or columns.
2. Develop *computational* approaches to fitting the models with high-dimensional data.

Objective

Goals

1. Develop *flexible* models for dependencies among rows and/or columns.
2. Develop *computational* approaches to fitting the models with high-dimensional data.

Approach

- Model: Matrix-variate normal.

My Contributions

1. Constraints on parameters:
   - Mean-restricted matrix-variate normal.
   - Joint covariance estimation via regularization.
     - Special case: Analytical solution.
2. Conditional distributions:
   - Algorithm and theoretical results.
3. Statistical applications:
   - Large-scale inference.
   - Missing data imputation.

My contributions make the matrix-variate model accessible for applications to high-dimensional data.

Matrix-variate Normal

Matrix extension of the multivariate normal:

- $\mathbf{X}_{n \times p} \sim N_{n,p}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Delta})$
  - Mean matrix: $\mathbf{M} \in \Re^{n \times p}$.
  - Column covariance: $\mathbf{\Delta} \in \Re^{p \times p}$.
  - Row covariance: $\mathbf{\Sigma} \in \Re^{n \times n}$

Matrix-variate Normal

Matrix extension of the multivariate normal:

- $\mathbf{X}_{n \times p} \sim N_{n,p}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Delta})$
  - Mean matrix: $\mathbf{M} \in \Re^{n \times p}$.
  - Column covariance: $\mathbf{\Delta} \in \Re^{p \times p}$.
  - Row covariance: $\mathbf{\Sigma} \in \Re^{n \times n}$
- $\mathrm{vec}(\mathbf{X}) \sim N\left(\mathrm{vec}(\mathbf{M}), \mathbf{\Omega}\right)$
  - $\mathbf{\Omega} = \mathbf{\Delta} \otimes \mathbf{\Sigma}$.

$$\mathbf{\Omega}_{np \times np} = \begin{pmatrix} \Delta_{11}\,\mathbf{\Sigma} & \Delta_{12}\,\mathbf{\Sigma} & \dots & \Delta_{1p}\,\mathbf{\Sigma} \\ \Delta_{21}\,\mathbf{\Sigma} & \Delta_{22}\,\mathbf{\Sigma} & & \\ \vdots & & \ddots & \vdots \\ \Delta_{p1}\,\mathbf{\Sigma} & & \dots & \Delta_{pp}\,\mathbf{\Sigma} \end{pmatrix}$$

(Gupta and Nagar, 1999)

Kronecker Product Covariance

Suppose is $\mathbf{X}$ multivariate with rows as features:

- $\mathbf{X}_{n \times p} \sim N\left(0, \mathbf{\Sigma}_{n \times n}\right)$.
- $\text{vec}(\mathbf{X}) \sim N\left(0, \mathbf{A}_{np \times np}\right)$.

$$\mathbf{A}_{np \times np} =$$

$$\begin{pmatrix} \mathbf{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} & & \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & & \dots & \mathbf{\Sigma} \end{pmatrix}$$

Kronecker Product Covariance

Suppose we want to allow **A** to be full.

- Too many parameters!

$$\mathbf{A}_{np \times np} =$$

$$\begin{pmatrix} A_{1,1} & A_{1,2} & \dots & A_{1,np} \\ A_{2,1} & A_{2,2} & & \\ \vdots & & \ddots & \vdots \\ A_{np,1} & & \dots & A_{np,np} \end{pmatrix}$$

Kronecker Product Covariance

Kronecker Product structure:

$$\mathbf{\Omega}_{np \times np} = \mathbf{\Delta} \otimes \mathbf{\Sigma}$$

$$\mathbf{\Omega}_{np \times np} =$$

$$\begin{pmatrix} \Delta_{11}\,\mathbf{\Sigma} & \Delta_{12}\,\mathbf{\Sigma} & \dots & \Delta_{1p}\,\mathbf{\Sigma} \\ \Delta_{21}\,\mathbf{\Sigma} & \Delta_{22}\,\mathbf{\Sigma} & & \\ \vdots & & \ddots & \vdots \\ \Delta_{p1}\,\mathbf{\Sigma} & & \dots & \Delta_{pp}\,\mathbf{\Sigma} \end{pmatrix}$$

Applying the Matrix-variate Normal

Challenges

1. No meaningful mean estimates, $\hat{\mathbf{M}}$.
2. Singular covariance estimates, $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Delta}}$.
3. Enormous number of parameters, $\boldsymbol{\Omega}_{np \times np}$:

Applying the Matrix-variate Normal

Challenges

1. No meaningful mean estimates, $\hat{\mathbf{M}}$.
2. Singular covariance estimates, $\hat{\boldsymbol{\Sigma}}$, $\hat{\boldsymbol{\Delta}}$.
3. Enormous number of parameters, $\boldsymbol{\Omega}_{np \times np}$:
   - Suppose medium-sized data: $\mathbf{X}$ is 100×100.
   - $\boldsymbol{\Omega}$ is $10,000 \times 10,000$.
   - Computing $\boldsymbol{\Omega}^{-1}$ is:

$$O(n^3 p^3) = 10^{12}!!$$

Restrictions on the Means

Mean-Restricted Matrix-variate Normal

- Recall:

  $$\mathbf{X}_{n \times p} \sim N_{n,p}(\mathbf{M}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}).$$

Restrictions on the Means

Mean-Restricted Matrix-variate Normal

- Recall:

  $\mathbf{X}_{n \times p} \sim N_{n,p}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Delta})$.
- $\mathbf{X}_{n \times p} \sim N_{n,p}(\nu, \mu, \mathbf{\Sigma}, \mathbf{\Delta})$
  - Column mean: $\nu \in \Re^n$.
  - Row mean: $\mu \in \Re^p$.

Restrictions on the Means

Mean-Restricted Matrix-variate Normal

- Recall:

  $\mathbf{X}_{n \times p} \sim N_{n,p}(\mathbf{M}, \mathbf{\Sigma}, \mathbf{\Delta})$.
- $\mathbf{X}_{n \times p} \sim N_{n,p}(\nu, \mu, \mathbf{\Sigma}, \mathbf{\Delta})$
  - Column mean: $\nu \in \Re^n$.
  - Row mean: $\mu \in \Re^p$.
  - $\mathbf{M} = \nu \mathbf{1}_{(p)}^T + \mathbf{1}_{(n)} \mu^T$.

$$\mathbf{M}_{n \times p} =$$

$$\begin{pmatrix} \nu_1 + \mu_1 & \dots & \nu_1 + \mu_p \\ \vdots & \ddots & \vdots \\ \nu_n + \mu_1 & \dots & \nu_n + \mu_p \end{pmatrix}$$

Transposable Regularized Covariance Model (TRCM)

$$\ell(\nu, \mu, \mathbf{\Sigma}, \mathbf{\Delta}) = \frac{p}{2}\log|\,\mathbf{\Sigma}^{-1}\,| + \frac{n}{2}\log|\,\mathbf{\Delta}^{-1}\,| \\ - \frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}(\mathbf{X} - \nu\mathbf{1}^T - \mathbf{1}\mu^T)\,\mathbf{\Delta}^{-1}(\mathbf{X} - \nu\mathbf{1}^T - \mathbf{1}\mu^T)^T\right) \\ -\rho_r J_r\left(\mathbf{\Sigma}^{-1}\right) - \rho_c J_c\left(\mathbf{\Delta}^{-1}\right).$$

- ρ_r and ρ_c are penalty parameters.
- $J_r : \Re^{n \times n} \rightarrow \Re$ and $J_c : \Re^{p \times p} \rightarrow \Re$ are convex functions.

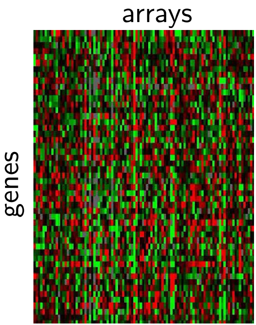
TRCM Contd.

$\ell(\nu, \mu, \mathbf{\Sigma}, \mathbf{\Delta})$ = log-likelihood mean-restricted matrix-variate normal distribution
$+$ penalties on inverse covariances (concentration matrices).

Why do we place penalties on inverse covariances?

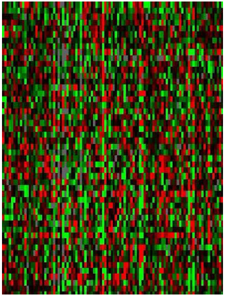
1. Gives non-singular estimates of $\mathbf{\Sigma}$ and $\mathbf{\Delta}$.
2. Log-likelihood concave in concentration matrix, not covariance matrix.
3. Separate penalties on $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Delta}^{-1}$ allow for simple maximization strategies.

Interpretation: Multivariate Marginals



- Each row and column: multivariate normal.
  - Rows: $\mathbf{X}_{i\cdot} \sim N\left(\mu_i + \nu, \Sigma_{ii}\, \mathbf{\Delta}\right)$.
  - Columns: $\mathbf{X}_{\cdot j} \sim N\left(\nu_j + \mu, \Delta_{jj}\, \mathbf{\Sigma}\right)$.

Interpretation: Multivariate Marginals



- Each row and column: multivariate normal.
  - Rows: $\mathbf{X}_{i\cdot} \sim N\left(\mu_i + \nu, \Sigma_{ii}\,\mathbf{\Delta}\right)$.
  - Columns: $\mathbf{X}_{\cdot j} \sim N\left(\nu_j + \mu, \Delta_{jj}\,\mathbf{\Sigma}\right)$.
- Multivariate Normal: Special Case.
  - If $\mathbf{\Sigma} = \mathbf{I}$ and $\nu = \mathbf{0}$ then, $\mathbf{X} \sim N\left(\mu, \mathbf{\Delta}\right)$.
  - If $\mathbf{\Delta} = \mathbf{I}$ and $\mu = \mathbf{0}$ then, $\mathbf{X} \sim N\left(\nu, \mathbf{\Sigma}\right)$.

Interpretation: Random Effects Model

If $\mathbf{\Sigma}$ and $\mathbf{\Delta}$ diagonal,

$$\mathbf{X}_{ij} = \nu_i + \mu_j + \epsilon_{ij} \qquad \text{where } \epsilon_{ij} \sim N(0, \mathbf{\Sigma}_{ii}\,\mathbf{\Delta}_{jj}).$$

Interpretation: Random Effects Model

If $\mathbf{\Sigma}$ and $\mathbf{\Delta}$ diagonal,

$$\mathbf{X}_{ij} = \nu_i + \mu_j + \epsilon_{ij} \qquad \text{where } \epsilon_{ij} \sim N(0, \mathbf{\Sigma}_{ii}\, \mathbf{\Delta}_{jj}).$$

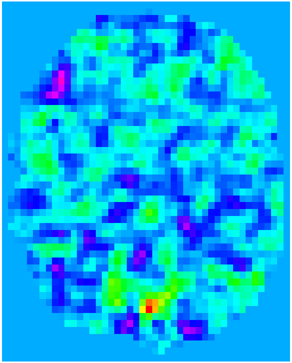
Otherwise,

$$(\mathbf{X}_{ij}, \mathbf{X}_{i'j'}) \sim N\left(\begin{pmatrix} \nu_i + \mu_j \\ \nu_{i'} + \mu_{j'} \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{ii}\, \mathbf{\Delta}_{jj} & \mathbf{\Sigma}_{ii'}\, \mathbf{\Delta}_{jj'} \\ \mathbf{\Sigma}_{i'i}\, \mathbf{\Delta}_{j'j} & \mathbf{\Sigma}_{i'i'}\, \mathbf{\Delta}_{j'j'} \end{pmatrix} \right)$$

Netflix Example:

- Rating = Customer mean + Movie mean
  + Variance/Covariance component depending on relationships with other customers and movies.

Interpretation: Tensor Product Gaussian Process



- Tensor product between row and column variables gives Kronecker covariance.
- Models *interaction* between rows and columns.
- Covariance function estimated directly from data (through penalties).

Maximizing the Likelihood

Maximum Likelihood Estimates

- Mean MLE's: row and column means.
- Covariance MLE's: more difficult ...

Covariance Estimation Challenges

- $\ell(0, 0, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ (written as $\ell(\boldsymbol{\Sigma}, \boldsymbol{\Delta})$) is *non-concave*.
- No theory supporting a global maximum, or a maximization strategy.

Covariance Estimation

But, can exploit concave properties:

- $\ell(\mathbf{\Sigma}, \mathbf{\Delta})$ a bi-concave function of $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Delta}^{-1}$.
  - $\ell(\mathbf{\Sigma}, \mathbf{\Delta})$ concave in $\mathbf{\Sigma}^{-1}$ with $\mathbf{\Delta}^{-1}$ fixed, and concave in $\mathbf{\Delta}^{-1}$ with $\mathbf{\Sigma}^{-1}$ fixed.
- Alternately maximize w.r.t. $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Delta}^{-1}$.
- Solves sub-gradient equations for $\mathbf{\Sigma}^{-1}$ and $\mathbf{\Delta}^{-1}$.

Proposition

Converges to a stationary point of $\ell(\mathbf{\Sigma}, \mathbf{\Delta})$!

Example Penalty Type: L_1

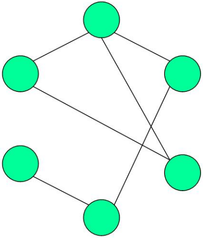
Penalty:

$$|| \mathbf{\Delta}^{-1} ||_1 = \sum_{i=1}^{p} \sum_{j=1}^{p} |\Delta_{ij}^{-1}|$$

- Related to sparse undirected graph estimation.
- Non-zeros in $\mathbf{\Delta}^{-1}$ correspond to *edges* in graph.

Netflix Example:

- Links between customers.
- Links between movies.



Example Penalty Type: L_2

$$\text{Penalty:} \quad ||\, \mathbf{\Delta}^{-1} \,||_2 = \sum_{i=1}^{p} \sum_{j=1}^{p} |\Delta_{ij}^{-1}|^2 = \text{tr}(\mathbf{\Delta}^{-2}).$$

Theorem
With L_2 penalties on $\mathbf{\Delta}^{-1}$ and $\mathbf{\Sigma}^{-1}$, $\text{argmax}_{\mathbf{\Sigma},\mathbf{\Delta}}\, \ell(\mathbf{\Sigma}, \mathbf{\Delta})$ has a *unique analytical solution* which is the *global maximum*.

- Solution a function of the singular value decomposition of $\mathbf{X}$.
- One of only a handful of known non-convex problems with an analytical solution!!

Transposable Data

Review: Matrix-variate Normal

Model

Transposable Regularized Covariance Model

Fitting the Model

Parameter Estimation

Conditional Expectations

Applications

Large-Scale Inference

Missing Data Imputation

Future Work

Background

$$\mathbf{X} = \begin{pmatrix} X_1 & X_4 \\ X_2 & X_5 \\ X_3 & X_6 \end{pmatrix}$$

Why do we need algorithms for conditional expectations?

1. Complete data.
   - How do we choose penalty parameters?
   - Cross-validation.
   - Remove elements from the matrix and predict them.
2. Incomplete data.
   - Predict scattered missing values from the observed values.

Computing Conditional Expectations

Problem:

- Recall: $\text{vec}(\mathbf{X}) \sim N\left(\text{vec}(\mathbf{M}), \boldsymbol{\Omega}\right)$.
- Use multivariate conditional expectation formulas with $\boldsymbol{\Omega}$:

$$\mathrm{E}(X_1 | X_2 \dots X_6) = \mathbf{M}_1 + \boldsymbol{\Omega}_{1,2:6}\, \boldsymbol{\Omega}_{2:6,2:6}^{-1} \left([X_2, \dots X_6] - \mathbf{M}_{2:6} \right)$$

- But, ... need $\boldsymbol{\Omega}^{-1}$ which is $O(n^3 p^3)$!

Computing Conditional Expectations

Problem:

- Recall: $\text{vec}(\mathbf{X}) \sim N\left(\text{vec}(\mathbf{M}), \boldsymbol{\Omega}\right)$.
- Use multivariate conditional expectation formulas with $\boldsymbol{\Omega}$:

$$\mathrm{E}(X_1 | X_2 \dots X_6) = \mathbf{M}_1 + \boldsymbol{\Omega}_{1,2:6}\, \boldsymbol{\Omega}_{2:6,2:6}^{-1} \left([X_2, \dots X_6] - \mathbf{M}_{2:6} \right)$$

- But, ... need $\boldsymbol{\Omega}^{-1}$ which is $O(n^3 p^3)$!

Goal:

- Find conditional expectation of scattered missing elements of a matrix in minimal computational time.
- Exploit structure of covariance matrix.

Algorithm

$$\mathbf{X} = \begin{pmatrix} X_1 & X_4 \\ X_2 & X_5 \\ X_3 & X_6 \end{pmatrix}$$

- Theorem: Conditional expectation of elements within a column (or row):
  1. Condition on other columns.
  2. Condition within column.

Algorithm

$$\mathbf{X} = \begin{pmatrix} X_1 & X_4 \\ X_2 & X_5 \\ X_3 & X_6 \end{pmatrix}$$

- Theorem: Conditional expectation of elements within a column (or row):
  1. Condition on other columns.
  2. Condition within column.
- Algorithm: Alternating Conditional Expectations.
  - Alternate sweeping through each row and column.
  - Missing values set to their conditional expectations.
  - Iterate until convergence.

Algorithm Contd.

$$\mathbf{X} = \begin{pmatrix} X_1 & X_4 \\ X_2 & X_5 \\ X_3 & X_6 \end{pmatrix}$$

- Theorem: Algorithm converges to conditional expectation of scattered elements!

$$\mathrm{E}(X_2, X_4, X_6 | X_1, X_3, X_5)$$

Algorithm Contd.

$$\mathbf{X} = \begin{pmatrix} X_1 & X_4 \\ X_2 & X_5 \\ X_3 & X_6 \end{pmatrix}$$

- Theorem: Algorithm converges to conditional expectation of scattered elements!

$$\mathrm{E}(X_2, X_4, X_6 | X_1, X_3, X_5)$$

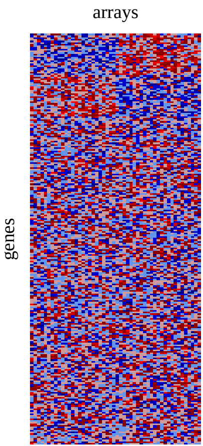
- For sparse or dense matrices, $\approx O(np)$ - linear time!

Time in Seconds:

	10×10	50×50	100×100	500×500	1000×1000
Naive Method	0.015	5.038	227.065	?	?
My Algorithm	0.001	0.037	0.078	0.082	0.339

Example: Two-Class Microarray

Signal + Independent Noise

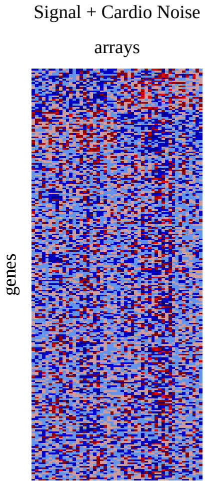


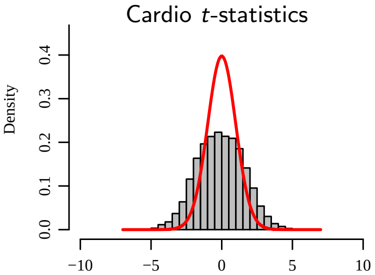
- Two classes, i.e. Diseased vs. Healthy.
- Goal: Find differentially expressed genes.
- Method: Two-sample t-test.

$$t_i = \frac{\bar{X}_{1,i} - \bar{X}_{2,i}}{S_{X_1 X_2,i}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- Use False Discovery Rate (FDR) to control false positives.
- Assumptions: 1. Array independence, 2. Limited gene dependencies

Incorrect Assumptions





- Arrays:
  - Measurement process: batch-effects, instrument drift, etc.
  - Latent variables: age, gender, family history, etc.

Null Distributions

Two-sample Z-test:

- Independent arrays:

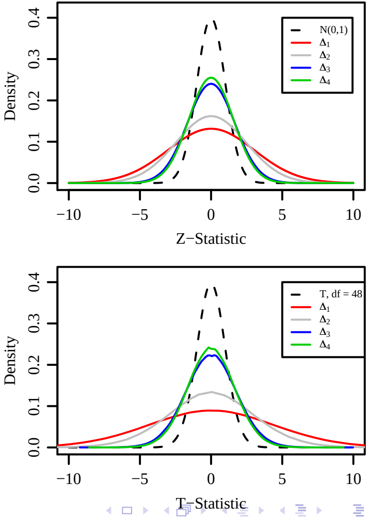
$$Z \sim N(0,1).$$

- Theorem: Under matrix-variate normal,

$$Z \sim N(0, \eta / c_n),$$

where $c_n = \frac{1}{n_1} + \frac{1}{n_2}$,

η is a function of $\mathbf{\Delta}$.



De-Correlating the Matrix

- Apply TRCM model to *sphere* the data matrix.

$$\mathbf{X} \quad \overset{\text{Apply TRCM}}{\Rightarrow} \quad \tilde{\mathbf{X}}$$

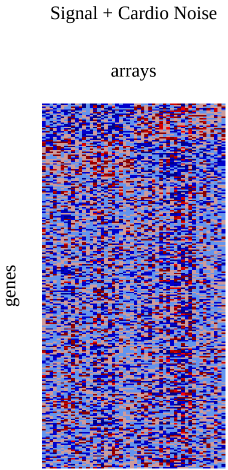
- $\tilde{\mathbf{X}}$ has approximately independent rows and columns.

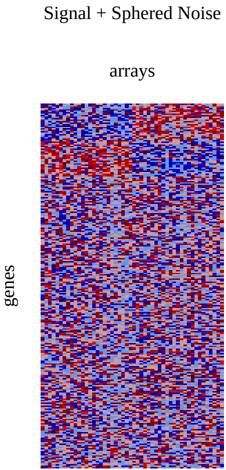
- Proposition: Under the null hypothesis (and certain assumptions),

$$\tilde{T} \sim \sqrt{\frac{\eta}{c_n}}\, t_{(n_1+n_2-2)}.$$

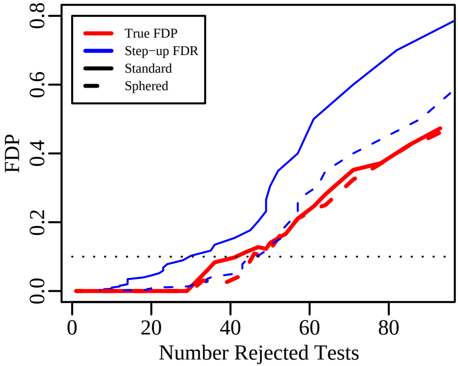
($\tilde{T}$ = two-sample T-statistic calculated from $\tilde{\mathbf{X}}$.)

Cardio Results: Data Images





Cardio Results: FDR Curves



Benefits of Sphering:

1. Increased statistical power.
2. Correct estimation of FDR.

Results: Other Models

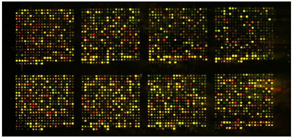
	Standard		Sphered	
	FDP	$\widehat{\text{FDR}}$	FDP	$\widehat{\text{FDR}}$
Latent Variable Model	0.189 (0.015)	0.383 (0.051)	0.167 (0.018)	0.166 (0.021)
Random Effects Model	0.52 (0.013)	0.0229 (0.0044)	0.154 (0.018)	0.207 (0.037)
Gene Correlations	0.169 (0.025)	0.19 (0.03)	0.141 (0.0.16)	0.185 (0.035)
Gene & Array Correlations	0.111 (0.011)	0.426 (0.04)	0.105 (0.0085)	0.124 (0.02)

True FDP and FDR estimated by the step-up method for 55/250 rejected tests averaged over 10 simulations.

Missing Values in Transposable Data

Missing due to measurement process:

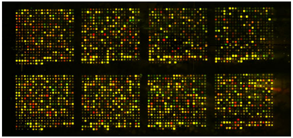
- Microarrays.
- fMRIs.



Missing Values in Transposable Data

Missing due to measurement process:

- Microarrays.
- fMRIs.



Predicting missing values as the main objective:

- Netflix.

	Ann	Ben	Chuck	Dee
The Shining	-	5	4	-
Saw	-	4	5	-
Carrie	-	5	-	-
Pretty Woman	4	-	2	5
Titanic	5	3	-	3

Penalized EM-type Algorithms

Goal: Maximize observed data log-likelihood.

- Covariance-regularized EM algorithm for multivariate models.
- TRCM: Penalized Multi-cycle ECM (Expectation Conditional Maximization) algorithm.
  - Groups computations according to rows and column separately.
- Bayesian Imputation: Blocked Gibbs sampler for sampling from the posterior of TRCM.

One-Step Approximation

Algorithm

1. Predict missing data via marginal models.
2. M Step: Estimate TRCM parameters.
3. E Step: Impute missing elements via Alternating Conditional Expectations algorithm.

One-Step Approximation

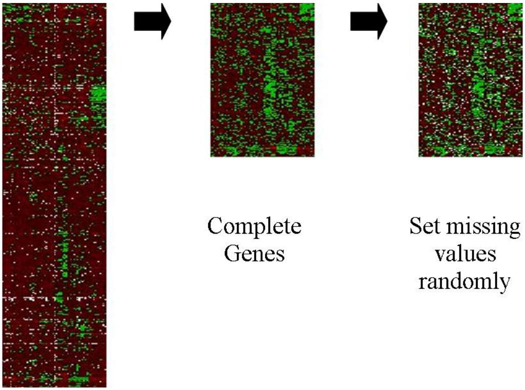
Algorithm

1. Predict missing data via marginal models.
2. M Step: Estimate TRCM parameters.
3. E Step: Impute missing elements via Alternating Conditional Expectations algorithm.

- Penalty parameters selected via CV.
- *Flexible:* CV used to determine multivariate row or column model or transposable model.

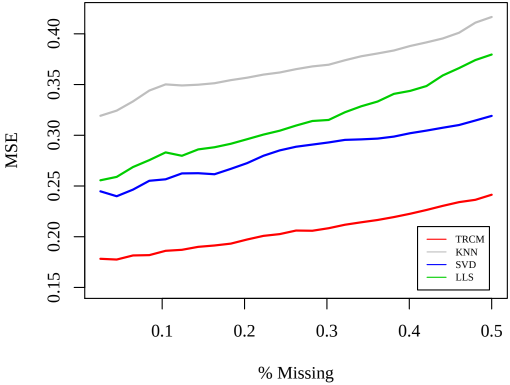
Results: Missing Microarray Data

- Kidney cancer microarray: 6,830 genes, 64 samples.
- Complete Genes: 2,069.
- Set missing values randomly.



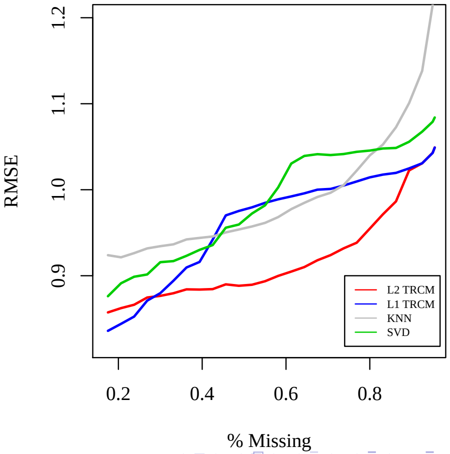
Results: Microarrays

- Mean Squared Error (MSE) of imputed values vs. proportion set to missing.
- Compared to 3 commonly used methods.



Results: Netflix Subset

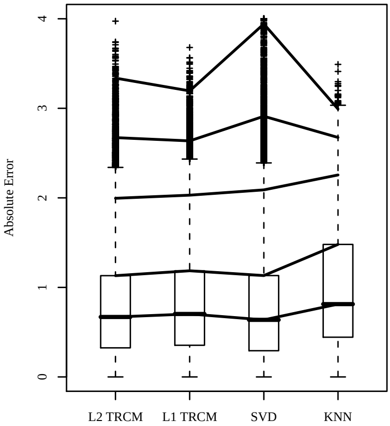
- Root MSE (RMSE) on a dense subset of Netflix data: 250 customers and 250 movies.
- With 95% missing, RMSE of SVD is 1.084 vs. 1.049 for TRCM.



Results: Netflix - Missing in Original Pattern

- 250 movies, each customer's ratings deleted in the pattern of 250 randomly selected customers - 74% missing.
- RMSE:

L_2, TRCM:	1.005,
L_1, TRCM:	1.029,
SVD:	1.032,
KNN:	1.184.



Transposable Data
- Review: Matrix-variate Normal

Model
- Transposable Regularized Covariance Model

Fitting the Model
- Parameter Estimation
- Conditional Expectations

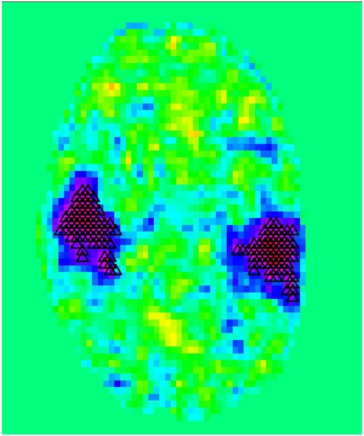
Applications
- Large-Scale Inference
- Missing Data Imputation

Future Work

Current Work

High-dimensional approximations to L_1 TRCM solution.

- Apply sphering algorithm to large data sets.
- Exploring thresholding approximations.
- Application: fMRIs.



Future Work

- Bi-clustering.
- Consistency: L_1 penalties.
- Extensions to large-scale matrix completion: convergence rates of Alternating Conditional Expectations algorithm.
- Regression and classification when $\mathbf{X}$ is transposable.
- Assessing TRCM model fit.
- Testing $\boldsymbol{\Delta} = \mathbf{I}$ and $\boldsymbol{\Sigma} = \mathbf{I}$.
- Extensions to discrete data.
- Extensions to three-way data.
- Other matrix-norm penalties.

Acknowledgments

A big thanks to my advisor, Rob Tibshirani, for all the help, support, and guidance during the development of this work.

I would also like to thank . . .

- Stephen Boyd for discussions and suggestions regarding bi-convex minimization.
- Jonathan Taylor for comments regarding de-correlating a matrix.
- Brad Efron for a discussion of large-scale inference and microarrays.
- Kshitij Khare for a suggestion for one of the proofs.
- Joe Romano for references on multiple testing dependence.
- My committee for their helpful comments during the proposal.

References

G. I. Allen and R. Tibshirani. Transposable regularized covariance models with an application to missing data imputation. (To appear) *Annals of Applied Statistics*, 2010.

G. I. Allen and R. Tibshirani. Inference with transposable data: modeling the effects of row and column correlations. (Submitted), 2010.

A. K. Gupta and D. K. Nagar. *Matrix variate distributions*. CRC Press, 1999.