Optimization Algorithms for Inference & Classification of Genetic Profiles from Undersampled Measurements

Belhassen Bayar

Rowan University

Jun. - 24 - 2014

Outline

Probabilistic Non-negative Matrix Factorization: Theory & Application to Microarray Data Analysis

Introduction to NMF

- The Non-negative Matrix Factorization (NMF) is a non supervised approach for data dimensionality reduction and clustering.
- The NMF is a constrained matrix decomposition method, i.e, $V \approx WH$, where W and H are respectively $n \times k$ and $k \times m$ non-negative matrices with $k < \min(n, m)$

The NMF Algorithm

- The optimal factors minimize the squared error and are solutions to the following constrained optimization problem,

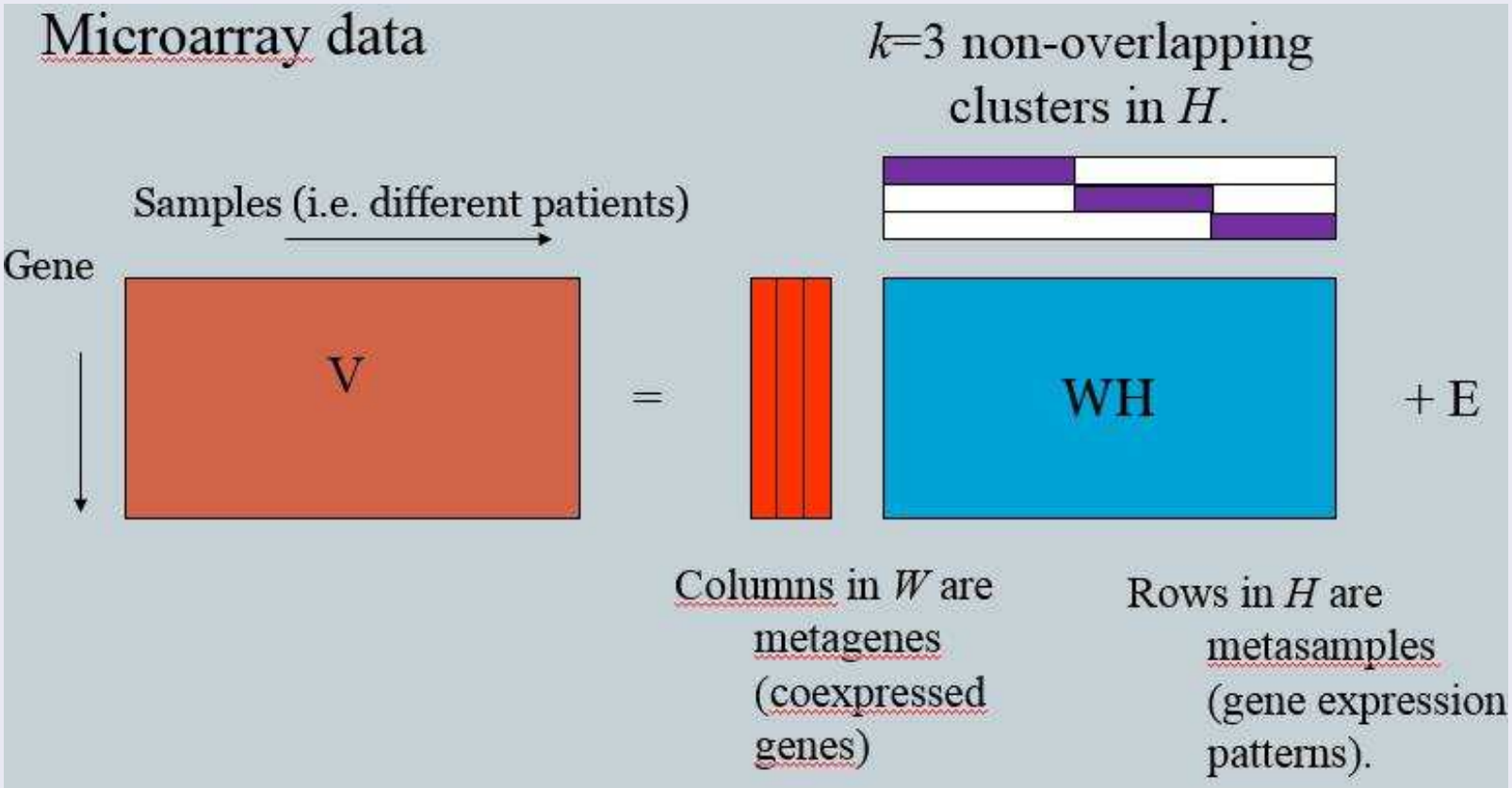
$$(W^*, H^*) = \underset{W,H \geq 0}{\arg\min} f(W,H) = \|V - WH\|_F^2, \tag{1}$$

- Lee and Seung showed that Eq. (1) converges to the local minimum under the following update rules:

$$\begin{cases} H_{ij}^{(t+1)} & \longleftarrow & H_{ij}^{(t)} \dfrac{(W^T V)_{ij}}{(W^T W H^{(t)})_{ij}} \\ \\ W_{ij}^{(t+1)} & \longleftarrow & W_{ij}^{(t)} \dfrac{(V H^T)_{ij}}{(W^{(t)} H H^T)_{ij}} \end{cases} \tag{2}$$

Matrix Factorization & Clustering

How NMF Clusters data



Motivation

- The NMF algorithm is a constrained approach: it conserves the positivity after processing the data.
- Brunet et al. (2004) PNAS- showed that NMF appears to have advantages over others popular clustering methods (i.e. HC and SOMs).
- However, the NMF is inherently a deterministic technique, whereas signals are generally noisy.

The Probabilistic NMF

- The data, represented by V, follows the normal distribution

$$p(V \mid W, H, \sigma^2) = \prod_{i=1}^{N} \prod_{j=1}^{M} [\mathcal{N}(V_{ij} \mid \mathbf{u}_i^T \mathbf{h}_j, \sigma^2)], \tag{3}$$

- We estimate factor matrices W and H using maximum a posteriori (MAP) criterion.
- The cost function for the stochastic model becomes:

$$\begin{aligned} (W^*, H^*) &= \arg\min_{W,H \geq 0} \| V - WH \|_F^2 + \lambda_W \| W \|_F^2 \\ &+ \lambda_H \| H \|_F^2, \end{aligned} \tag{4}$$

where $\lambda_W = \frac{\sigma^2}{\sigma_W^2}$ and $\lambda_H = \frac{\sigma^2}{\sigma_H^2}$.

Proposition: PNMF update rules

- The function

$$f(W,H) = \|V - WH\|_F^2 + \alpha\|W\|_F^2 + \beta\|H\|_F^2 \tag{5}$$

is non-increasing under the update rules

$$\begin{cases} H_{ij}^{(t+1)} & \longleftarrow & H_{ij}^{(t)} \dfrac{(W^T V)_{ij}}{(W^T WH^{(t)} + \beta H^{(t)})_{ij}} \\ W_{ij}^{(t+1)} & \longleftarrow & W_{ij}^{(t)} \dfrac{(VH^T)_{ij}}{(W^{(t)} HH^T + \alpha W)_{ij}^{(t)}} \end{cases} \tag{6}$$

PNMF-based Classification: Sparse Representation

- In a classification process, let A be the training data and Y is the testing data where A and Y are labeled.
- A can be written as: $A = [A_1, A_2, \cdots, A_k]$ where k is the number of classes.
- Each matrix A_i is a concatenation of r_i columns of the i^{th} class, $A_i = [c_{i,1}, c_{i,2}, \cdots, c_{i,r_i}]$.
- Then any testing vector $\boldsymbol{y} \in \mathbb{R}^n$ that belongs to the i^{th} class can be written as the following linear combination of the A_i columns, linear combination of the A_i columns,

$$\boldsymbol{y} = \alpha_{i,1}\boldsymbol{c}_{i,1} + \alpha_{i,2}\boldsymbol{c}_{i,2} + \ldots + \alpha_{i,r_i}\boldsymbol{c}_{i,r_i}, \tag{7}$$

PNMF-based Classification: Sparse Representation

- In terms of the training data A, a testing vector $\boldsymbol{y}$ can be re-written as $\boldsymbol{y} = A\boldsymbol{x}$ where

$$\boldsymbol{x} = [0, ...0, \alpha_{i,1}, \alpha_{i,2}, ..., \alpha_{i,r_i}, 0..., 0]^T \in \mathbb{R}^r, \tag{8}$$

  is the coefficient vector of the testing sample $\boldsymbol{y}$. $\boldsymbol{x}$ is a r_i-sparse vector whose nonzero entries are associated with the columns of the sub-matrix A_i.

- Thus, the vector $\boldsymbol{x}$ is determined by solving the following optimization problem suggested in [7]:

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}} \{ \|A\boldsymbol{x} - \boldsymbol{y}\|_2 + \lambda \|\boldsymbol{x}\|_1 \}, \tag{9}$$

PNMF-based Classification: Sparse Representation

- In our work, we propose to use the matrix $W = [W_1, W_2, .., W_k]$ computed using our suggested PNMF instead of A where $A_i = W_i \times H_i$.
- Thus, the optimization problem becomes:

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}} \{ \| W\boldsymbol{x} - \boldsymbol{y} \|_2 + \lambda \| \boldsymbol{x} \|_1 \}. \tag{10}$$

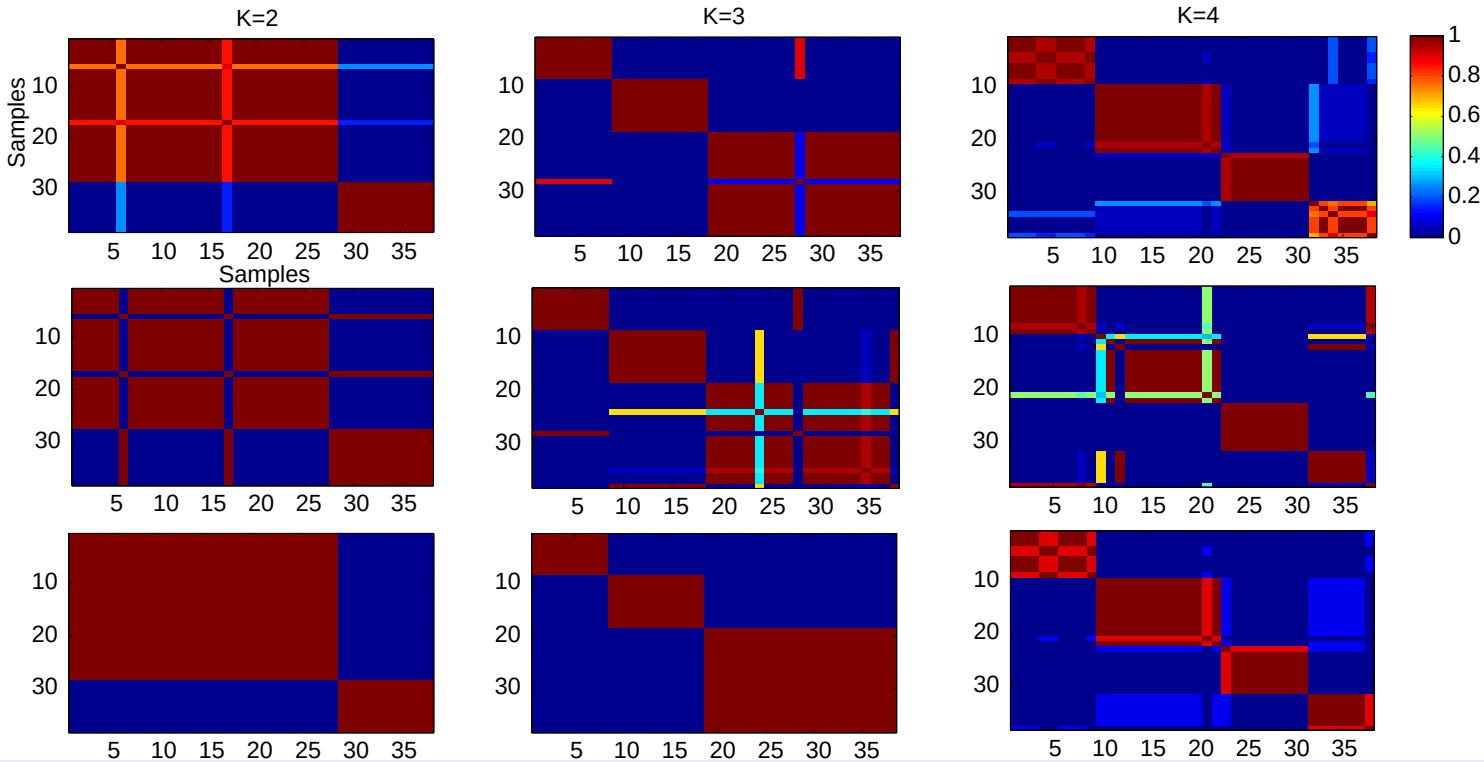
- Finally, associate class($\boldsymbol{y}$)$= \arg\min_i \| \boldsymbol{y} - W\delta_i(\boldsymbol{x}) \|_2$, $i = 1 : k$, where $\delta_i(\boldsymbol{x}) = [0, \cdots, 0, \boldsymbol{x}_i^T, 0, \cdots 0]^T$.

Simulation Results: Clustering

- We used the leukemia gene expression dataset which is a blood cancer and called ALLAML dataset as well.
- This gene expression dataset consists of 38 bone marrow samples with $5,000$ genes.
- The leukemia dataset contains 3 types: AML, ALL-T and ALL-B

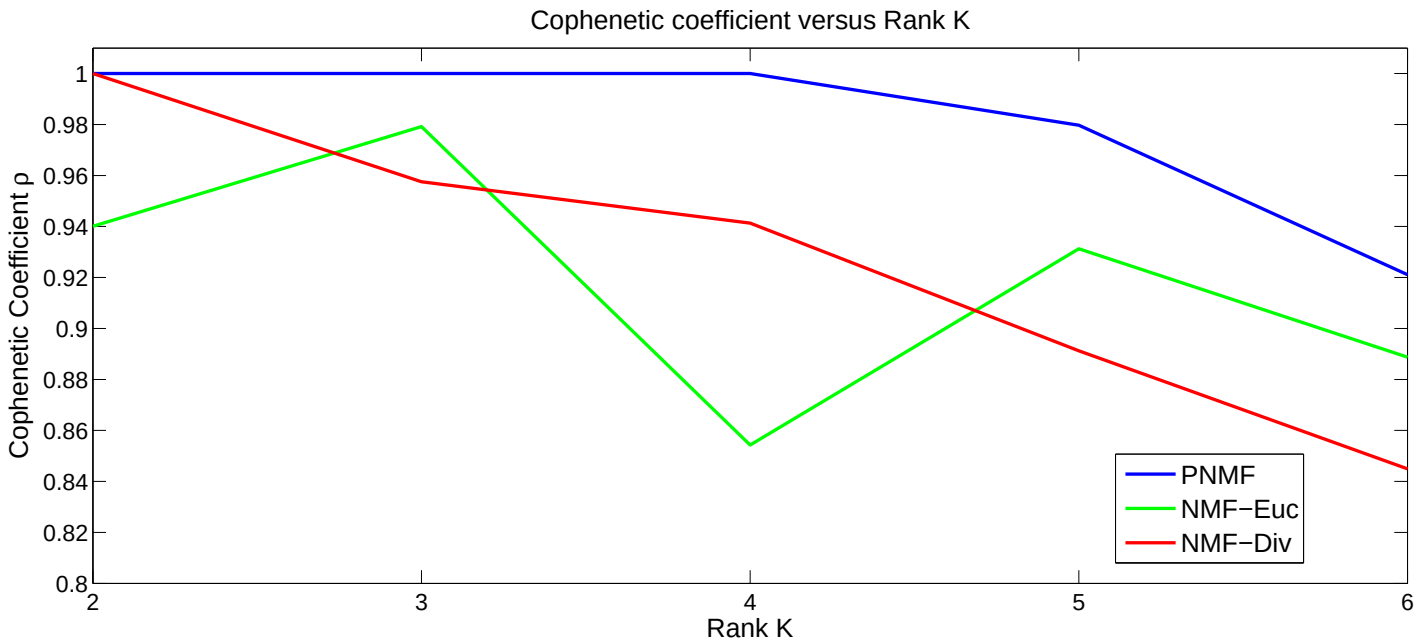
Simulation Results: Clustering

Consensus matrices: Top row NMF-Euc, Second row NMF-Div, bottom row: PNMF



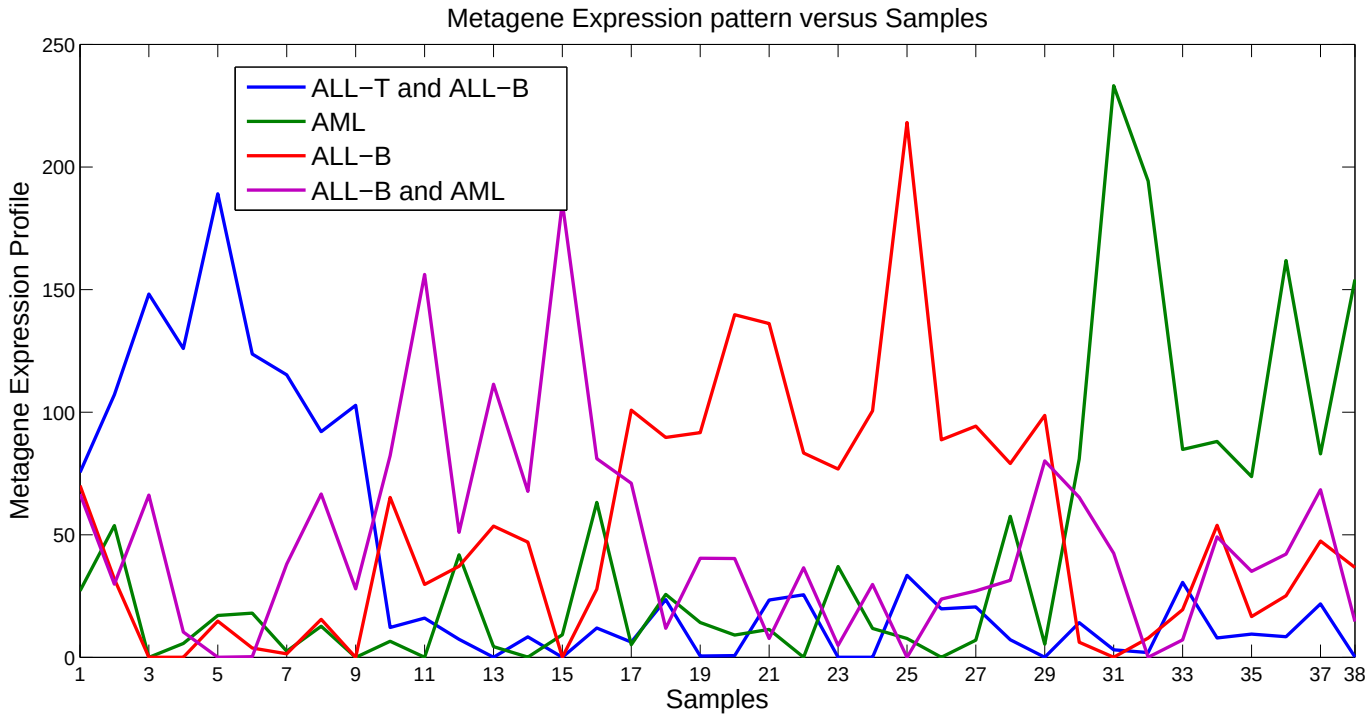
Simulation Results: Cophenetic Coefficient

Cophenetic coefficient $\rho_k = \frac{1}{m^2}\sum_{ij} 4(c_{ij} - \frac{1}{2})^2$ versus k



Simulation Results: Clustering

Metagenes expression patterns versus the samples for $k = 4$ in the Leukemia dataset



Simulation Results: Tumor Classification

PNMF-based Sparse Representation v.s. its homologues

Table: Classification accuracy

Data sets	*Nbr*. of classes	NMF-Euc	NMF-Div	SNMF	SVM	PNMF
Prostate	2	85.29%	86.27%	88.24%	99%	**92.16%**
Medulloblastoma	2	85.29%	91.18%	94.12%	79.16%	**94.12%**
Colon	2	85.48%	88.71%	90.32%	89.04%	**90.32%**
Breast-Colon	2	98.08%	95.19%	98.08%	84.63%	**98.08%**
Leukemia	3	97.37%	97.37%	97.37%	95.50%	**97.37%**
Lung	5	92.61%	90.64%	93.60%	85.54%	**94.09%**
Brain	5	76.19%	78.57%	83.33%	77%	**85.71%**

Conclusion

- In this work, we propose a new probabilistic approach for the NMF algorithm (PNMF).
- This algorithm has shown better clustering performance comparing to the deterministic NMF Algorithm.
- PNMF was able to recognize a fourth biological cluster that other clustering algorithms could not detect.
- We finally used PNMF as a feature extraction to classify tumors. The suggested PNMF-based classification algorithm has shown very high accuracy with Tumor data.

SMURC: High-Dimension Small-Sample Multivariate Regression with Covariance Estimation

Problem Formulation

We consider the under-determined multivariate regression problem with correlated noise in matrix notation:

$$Y = AX + E, \tag{11}$$

where Y, X are respectively $q \times n$ and $p \times n$ matrices. $E = [\boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_n]$ is the $q \times n$ random error matrix where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iq})^T$ and $\boldsymbol{\varepsilon_i} \sim \mathcal{N}(0, \Sigma)$. (A) is a $q \times p$ regression matrix and n is the sample size where $q > p > n$.

The model in (11) has been used in many applications such us:

- Modeling the gene expression (mRNA) as a function of copy number and microRNA.
- Modeling the dynamic of genetic networks

Problem Formulation

The lasso on A of the negative log-likelihood function of (A, Σ) can be expressed up to a constant as

$$-l(A,\Omega) = \mathrm{Tr}[\frac{1}{n}(Y-AX)^T\Omega(Y-AX)] - \ln|\Omega| + \sum_i\sum_j |a_{ij}|.$$

where $\Sigma^{-1} = \Omega$.

- Because the system is under-determined, there exist solutions satisfying $Y = AX$ and Σ infinitely small. Therefore, the negative log-likelihood tends to $-\infty$.
- Hence, the likelihood, as a function of the two variables (A, Ω), diverges.

Related Work: Sparse MRCE

- Rothman *et al.* [6] proposed MRCE: a regularized algorithm that simulatenously infers A and the inverse error covariance, $\Omega = \Sigma^{-1}$, by imposing sparsity constraints on Ω, i.e., l_1-norm penalty, to exclude exact solutions.

- However, in many applications, the assumption of a sparse inverse covariance matrix may not be reasonable or have any physical justification.

- MRCE relies on an iterative procedure that finds the maximum over A then over Ω since the problem is convex in each variable, A and Ω but not with respect to the pair (A, Ω).

Related Work: sCGGM

- Recently, Zhang *et al.* [8] proposed sCGGM: This algorithm minimizes the negative log-likelihood of the data with l_1 penalties on the autocorrelation and cross-correlation precision matrices.
- The main advantage of CGGM over MRCE is that CGGM leads to a convex problem, whereas the MRCE estimation problem is only bi-convex, not jointly convex.
- However, both algorithms solve an under-determined linear regression problem by maximizing the Gaussian likelihood subject to sparse constraints on the correlation structure.

Open Question

"How can we perform maximum likelihood estimation with unknown covariance structure for under-determined systems?"

- Solution: Modify the likelihood expression

Definition 1: The Normalized Likelihood

We define the normalized-likelihood of the under-determined ($n < p$) multivariate regression model under the Gaussian assumption, as

$$L_N(A,\Omega) = \frac{|(Y-AX)^T\Omega(Y-AX)|^{\frac{n}{2}}}{(2\pi)^{\frac{np}{2}}} \exp -\frac{1}{2}\mathrm{Tr}[(Y-AX)^T\Omega(Y-AX)]$$

It can be shown that the solution of $(A^*,\Omega^*) = \arg\max_{A,\Omega} L_N(A,\Omega)$ is

$$(Y-A^*X)^T\Omega^*(Y-A^*X) = n\,I$$

Uniqueness Issue

$$\boxed{(Y - A^*X)^T \Omega^* (Y - A^*X) = nI}$$

- There are (infinitely) many pairs (A, Ω) that satisfy the optimality condition.
- In order to obtain a unique solution, we need to further constrain the problem.

Error Constraint

Among all possible solutions we find those that minimize the regularized least-square error:

$$\| Y - AX \|_F^2 + \lambda \| \Omega \|_F^2, \tag{12}$$

where λ is a regularization parameter and $\| \cdot \|_F$ denotes the Frobenius norm. In addition, we consider a general constraint set, $\boldsymbol{A} \in \mathscr{A} \subset \mathbb{R}^{q \times p}$ as a prior knowledge. The constrained optimization problem, thus, becomes:

$$\begin{cases} \min\limits_{(\boldsymbol{A},\boldsymbol{\Omega})} \| \boldsymbol{Y} - \boldsymbol{AX} \|_F^2 + \lambda \| \boldsymbol{\Omega} \|_F^2 \\ \text{s.t.} \quad (\boldsymbol{Y} - \boldsymbol{AX})^T \boldsymbol{\Omega} (\boldsymbol{Y} - \boldsymbol{AX}) = n\boldsymbol{I}, \\ \boldsymbol{A} \in \mathscr{A}. \end{cases} \tag{13}$$

To solve (13), we use the polar decomposition of matrices.

Definition 2: Polar Decomposition

The polar decomposition of a matrix $B \in \mathbb{C}^{p \times n}$ is given by

$$B = U|B|, \tag{14}$$

where $|B| = (B^T B)^{1/2}$, $(\cdot)^{1/2}$ is the principal square root operator and $U : \mathbb{C}^p \longrightarrow \text{Range}(B)$ is a $\mathbb{C}^{p \times n}$ isometry such that $U^T U = I$.

Optimization Problem

$$\begin{cases} \min\limits_{(\boldsymbol{A},\boldsymbol{\Omega})} \|\boldsymbol{Y}-\boldsymbol{AX}\|_F^2+\lambda\|\boldsymbol{\Omega}\|_F^2 \\ \text{s.t.} \quad (\boldsymbol{Y}-\boldsymbol{AX})^T\boldsymbol{\Omega}(\boldsymbol{Y}-\boldsymbol{AX})=n\boldsymbol{I}, \\ \boldsymbol{A}\in\mathscr{A}. \end{cases} \tag{15}$$

Replacing the matrix $(Y-AX)$ by its polar decomposition, we obtain

$$\Omega_A = n\,U\,[(Y-AX)^T(Y-AX)]^{-1}U^T \tag{16}$$

where U is the isometry of $(Y-AX)$. Hence problem (15) becomes:

$$\begin{cases} \min\limits_{\boldsymbol{S}} \mathrm{Tr}(\boldsymbol{S}^2)+\lambda\ n^2\,\mathrm{Tr}(\boldsymbol{S}^{-4}) \\ \text{s.t.} \quad \boldsymbol{S}=\left((Y-AX)^T(Y-AX)\right)^{\frac{1}{2}}=|\boldsymbol{Y}-\boldsymbol{AX}|,\ \boldsymbol{A}\in\mathscr{A} \end{cases} \tag{17}$$

Sparsity constraint: Problem Approximation

Now we have

$$\begin{cases} \min\limits_{\boldsymbol{S}} \operatorname{Tr}(\boldsymbol{S}^2) + \lambda\ n^2 \operatorname{Tr}(\boldsymbol{S}^{-4}) \\ \text{s.t.} \quad \boldsymbol{S} = |\boldsymbol{Y} - \boldsymbol{A}\boldsymbol{X}|,\ \boldsymbol{A} \in \mathscr{A} \end{cases} \tag{18}$$

- The problem is not convex because the equality in the constraint is quadratic.
- We assume that the regression coefficient matrix A is sparse [1], therefore, we define the set $\mathscr{A} = \{\boldsymbol{A} : \|A\|_1 \leq \varepsilon\}$.
- We use the sparsity of A and approximate the problem by a convex optimization one.

Sparsity Constraint: Problem Approximation

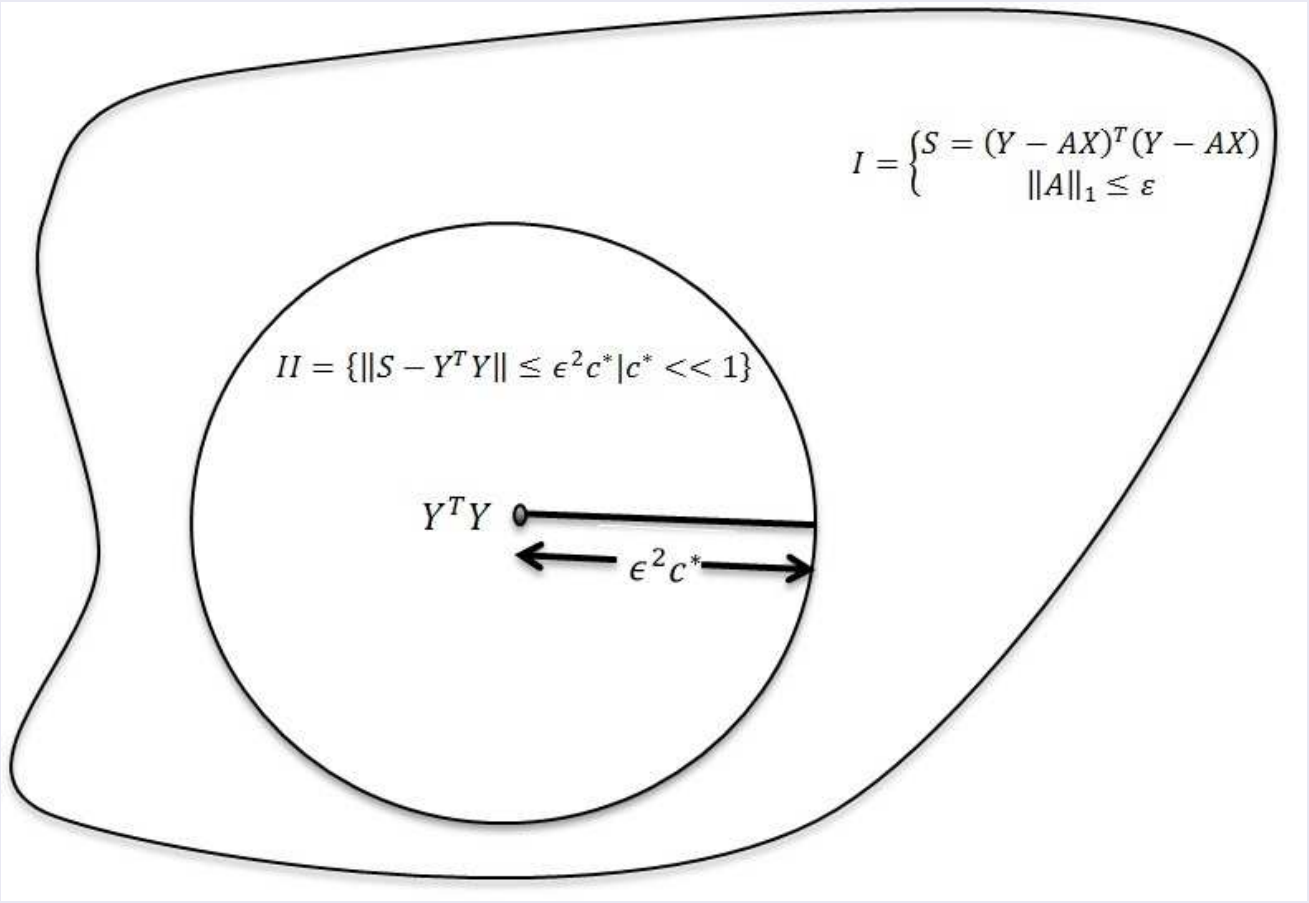
If $\mathscr{A} = \{\boldsymbol{A} : \|A\|_1 \leq \varepsilon\}$, then the solution to the optimization problem can be approximated by the solution to the following convex optimization

$$\begin{cases} \min\limits_{S} \operatorname{Tr}(S^2) + \lambda n^2 \operatorname{Tr}(S^{-4}) \\ \text{s.t.} \quad S \in \Lambda = \{S \in \mathbb{S}_{n,n} \,|\, \|S - |Y|\|_F \leq \varepsilon c^*\} \end{cases} \tag{19}$$

where $\mathbb{S}_{n,n}$ is the set of $n \times n$ symmetric positive semidefinite matrices and c^* is a small term which depends on X and Y.
Let S^* be the unique global solution of the convex optimization problem

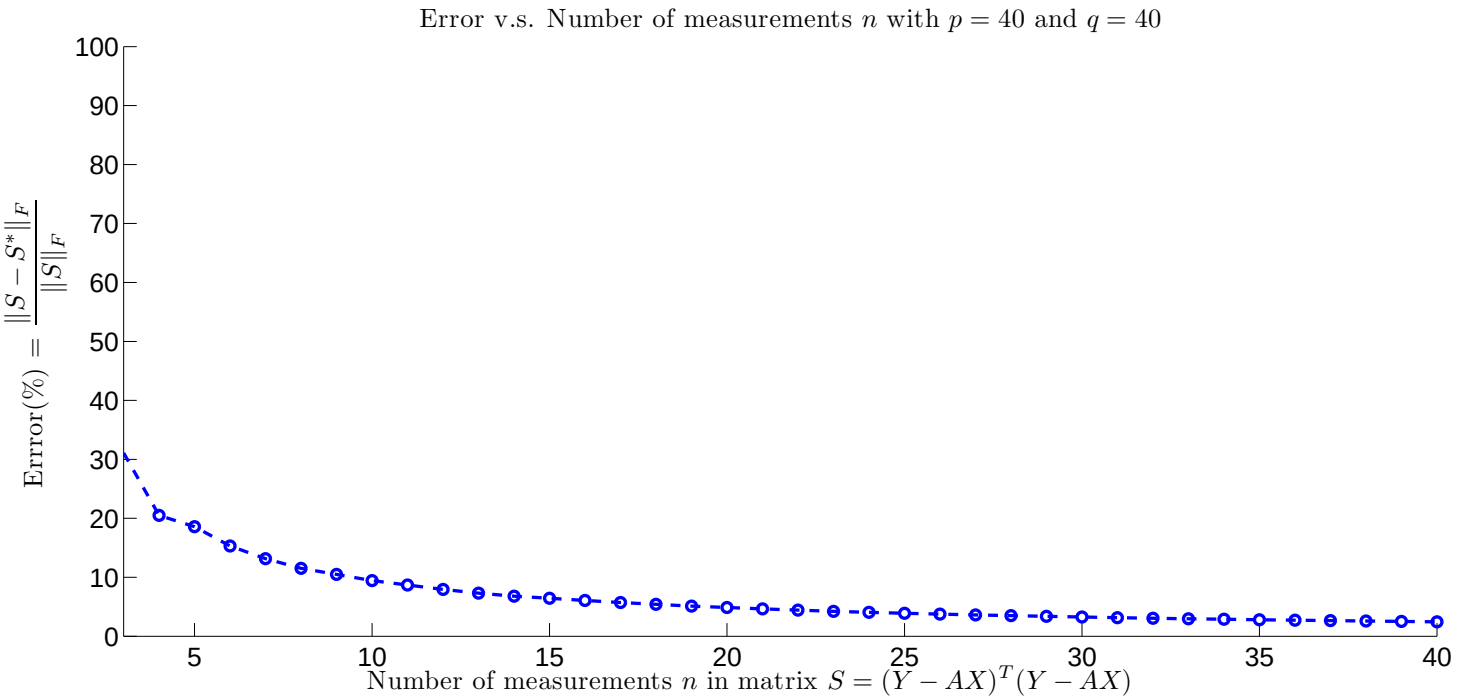
Sparsity Constraint: Problem Approximation

Geometrical Interpretation



Sparsity Constraint: Problem Approximation

Approximation Error $\|S - S^*\|_F / \|S\|_F$ v.s. $n = 1, \cdots, p$ for $p = 40$



Sparsity Constraint: Problem Approximation

- The optimal connectivity matrix, A^*, satisfies, $S^* = (Y - A^*X)^T(Y - A^*X)$.
- Since there are still many possible solutions to the optimal connectivity matrix, we propose to find the one with minimum l_1 norm:

$$\begin{cases} \min\limits_{A} \|A\|_1 \\ \\ \text{s.t.} \quad AX = Y - U(S^*)^{1/2}, \end{cases} \tag{20}$$

- However, the equality constraint is not affine with respect to A and U hence problem (20) is not anymore convex.

Solution Approximation

- Since A is sparse, we can approximate the solution constructed with U, the isometry of $(Y - AX)$, by a solution constructed by V, the isometry of Y.
- This approximation reduces the set over which we minimize but leads to an affine equality constraint and hence a convex problem. Thus, to find $\hat{A}$ we solve

$$\begin{cases} \min\limits_{A} \|A\|_1 \\ \text{s.t.} \quad AX = Y - V(S^*)^{1/2}, \end{cases} \tag{21}$$

The SMURC algorithm

Input: The matrices $X \in \mathbb{R}^{p\times n}$ and $Y \in \mathbb{R}^{q\times n}$ $(q > n)$ satisfying the under-determined regression model $Y = AX + E$ with an unknown covariance matrix.

Step 1 Solve the convex optimization problem

$$\begin{cases} \min\limits_{S} \mathrm{Tr}(S^2) + \lambda n^2 \mathrm{Tr}(S^{-4}) \\ \\ \text{s.t.} \quad S \in \Lambda = \{S \in \mathbb{S}_{n,n} \,|\, \|S - |Y|\|_F \leq \varepsilon c^*\} \end{cases} \tag{22}$$

Step 2 Given $S^* \in \mathbb{S}^{n\times n}$, the optimal connectivity matrix A^* is a solution of

$$\begin{cases} \min\limits_{A} \|A\|_1 \\ \\ \text{s.t.} \quad AX = Y - V(S^*)^{1/2} \end{cases} \tag{23}$$

Simulation Results: Synthetic Data

ODE Model

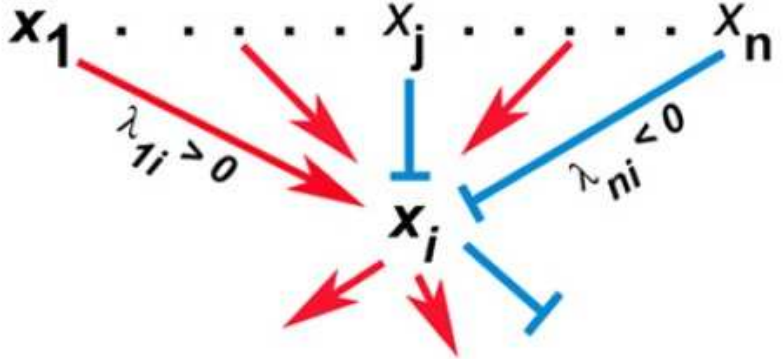
We consider the ODE model of genes interaction:

$$\frac{dx_i}{dt} = -a_{ii}x_i + \sum_{j\neq i}^{n} a_{ij}x_i + e_i, \quad (24)$$

The matrix notation of model (24)

$$Y = AX + E$$

The Gene Interaction Network

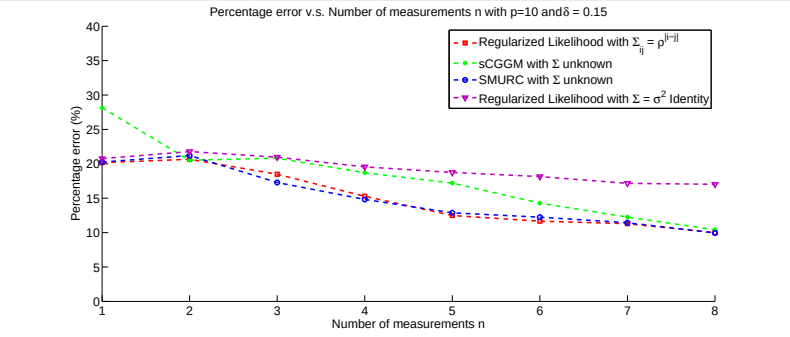


Simulation Results: Synthetic Data

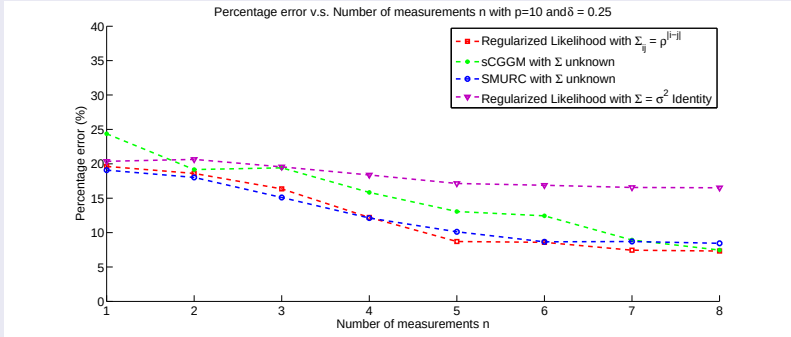
- We compare the proposed SMURC to sCGMM [8] and the regularized MLE algorithm, where the lasso penalty is imposed on A for given $\Sigma = \Sigma_{true}$ and $\Sigma = \sigma^2\ I$.
- we generate synthetic genomic networks with varying size p, number of measurements n and correlated structure Σ where $Y = AX + E$ and the entries of A, $a_{i,j} \in \mathcal{N}(0,1)$.
- We use a fixed correlation structure where $\Sigma_{i,j} = \rho^{|i-j|}$ and $\rho = 0.7$.
- We use two sparse models of the connectivity matrix, $\|A\|_0 = 0.2\ p^2$, where $\|\cdot\|_0$ is the number of non-zero elements.
- We define a False Positive interaction when $|a_{ij} - \hat{a}_{ij}| > \delta |a_{ij}|$

50 Monte Carlo simulations

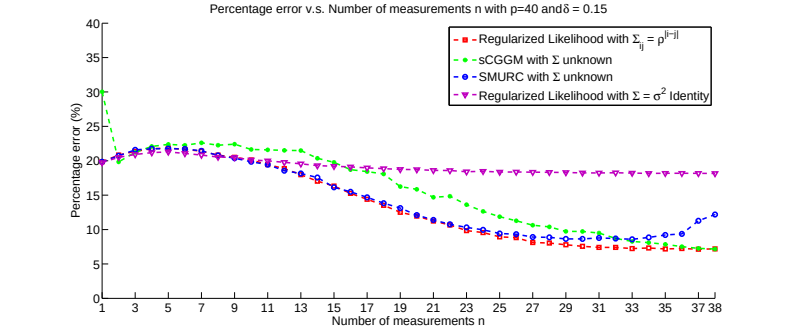
Performance comparison: SMURC v.s. sCGMM and RMLE for 80% sparsity



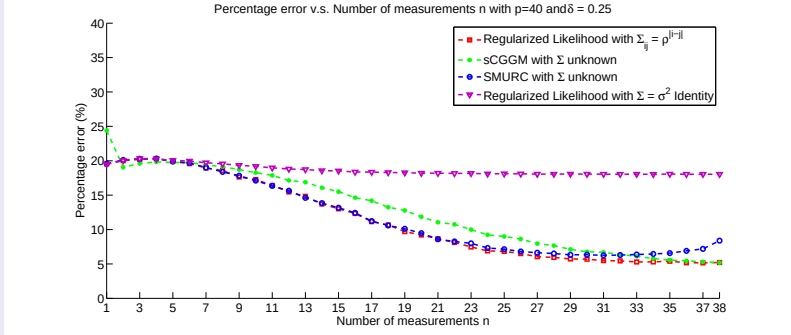
$p = 10$ and $\delta = 0.15$



$p = 10$ and $\delta = 0.25$



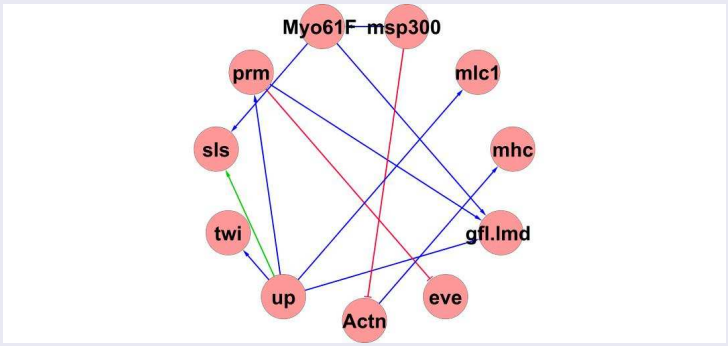
$p = 40$ and $\delta = 0.15$



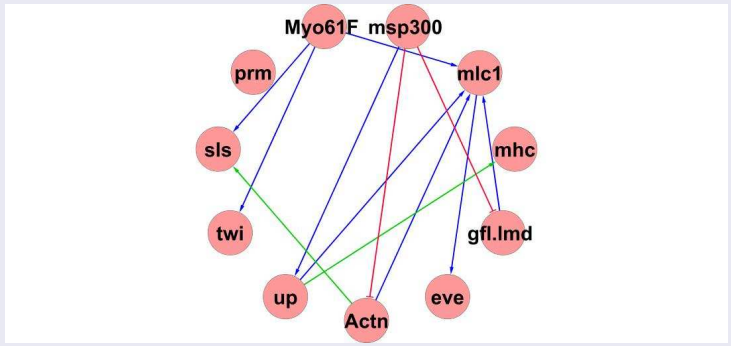
$p = 40$ and $\delta = 0.25$

Simulation Results: Real Data Application

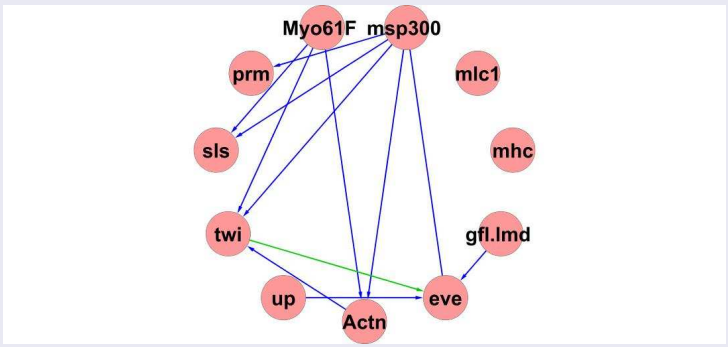
Estimated GRNs of the Drosophia using SMURC during four deveopmental phases



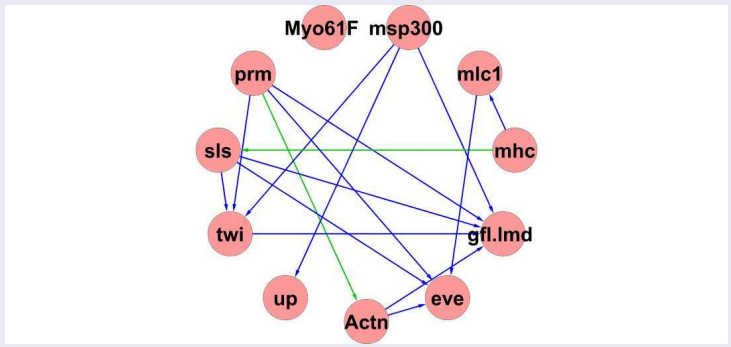
Embryonic



Larval



Pupal



Adulthood

Simulation Results: Real Data Application

Table: Detection of the known gene interactions in Flybase

	(*prm,Actn*)	(*sls,mhc*)	(*mhc,up*)	(*sls,Actn*)	(*sls,up*)	(*twi,eve*)
SMURC	✓ (A)	✓ (A)	✓ (L)	✓ (L)	✓ (E)	✓ (P)
Minimum description length [9]	✓	✓	×	×	×	✓
Random graph model [3]	×	×	✓ (E,L,P,A)	✓ (P,A)	✓ (E,L,P,A)	×
Dynamic Bayesian network [5]	×	✓ (E,L,P,A)	×	×	×	×
Nonparametric Bayesian regression [4]	×	×	×	×	×	✓ (E)

Conclusion

- The MLE of under-determined Gaussian systems with unknown covariance is senseless.
- We proposed a new algorithm, called SMURC, for under-determined multivariate regression model, that guarantees the convergence of the likelihood and keeps its Gaussian form.
- Our proposed SMURC algorithm outperforms sCGGM and the regularized maximum likelihood estimator with known covariance structure.
- SMURC was able to detect 6 out of 7 known interactions in Drosophila GRNs.

Kernel Reconstruction versus the Combinatorial ℓ_0-based Compressive Sensing Algorithm

What is Compressive Sensing?

Compressive Sensing

- Nyquist/Shannon Sampling Theorem: Fs must be at least 2$\times$ faster than the highest frequency in signal. Adresses limitations of signal reconstruction, i.e lossless.
- However, Data Acquisition rate has massively increased.
- It does not get benefits from signal Sparsity /Compressibility.
- Consequently, it affects and increases the resource consumption.

Compressive Sensing

- Unlike Nyquist theorem, CS exploits signals that admit a sparse representation in certain bases, e.g. Fourier..
- CS is used for acquiring and recovering compressible signals at a rate lower than Nyquist rate.
- It employs a few number of non adaptative-projections of the sparse signal to compute a set of M measurements, i.e. encoding.
- The ultimate goal is to reconstruct large N-dimensional signal from a small number of M measurements where $M << N$.

Compressive Sensing Model

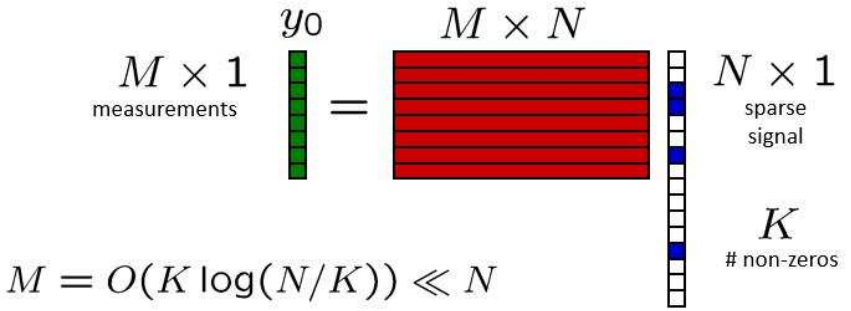
The Analytical Expression of a CS Model

Analytically, we have the following underdetermined system:

$$\boldsymbol{y} = \Phi \boldsymbol{x}$$

The CS Model

$$y_0 = \Phi x$$



Solution to CS

- Infering the sparse vector x is equivalent to solving the ℓ_0-minimization problem:

$$\min \|x\|_0 \text{ subject to } \quad y_0 = \phi \times x, \tag{25}$$

- Ideal CS reconstruction of K-sparse signal is to find the sparsest solution from the infinitely many solutions $\hat{x}$.
- If $M \leq K$, then there is no solution. Otherwise, there is high probability of exact reconstruction
- Donoho *et al.* [2] proposed two practical alternatives, i.e. ℓ_1-minimization leading to ℓ_2-minimization.

Measurement Matrix Φ

- The CS is an underdetrmined system, however not all underdetermined systems yield a solution.
- Therefore, some constraints are imposed on Φ:
  - The matrix Φ must satisfy the RIP condition.
  - The RIP implies stability and ℓ_2-recovery
  - The RIP implies the NSP.
  - The NSP implies ℓ_1-recovery.

Kernel Reconstruction Algorithm

- We consider the linear operator $\Phi : \mathbb{C}^N \longrightarrow \text{Range}(\Phi)$ where $\mathbb{C}^n = \text{Range}(\Phi^T) \oplus \text{Ker}(\Phi)$ and $\dim(\text{Ker}(\Phi)) = S$.
- Let $\boldsymbol{x}_0 \in \text{Range}(\Phi^T)$ be a particular solution such that:

$$\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{j=1}^{S} a_j \boldsymbol{b}_j, \tag{26}$$

where $\boldsymbol{b}_j$'s are the kernel vectors and a_j's are the coefficients of the linear combination in $\text{Ker}(\Phi)$.

- The matrix form of Eq. (26) is as follow

$$\boldsymbol{x} = \boldsymbol{x}_0 + B\boldsymbol{a}, \tag{27}$$

Thus, to find x we need to compute the entries in the vector $\boldsymbol{a}$.

Kernel Reconstruction Algorithm

- To find $\boldsymbol{a}$, we assume that we have AT LEAST $S = \dim(\mathrm{Ker}(\Phi))$ zeros in the vector x.
- $\mathrm{Rank}(B) = S$, therefore $\exists$ S linearly independent rows of B that span a space that we call L where these entries are equal to zero and let P_s be the projection matrix that projects x onto the space L.
- Finally, $\boldsymbol{a}$ can be computed as follow

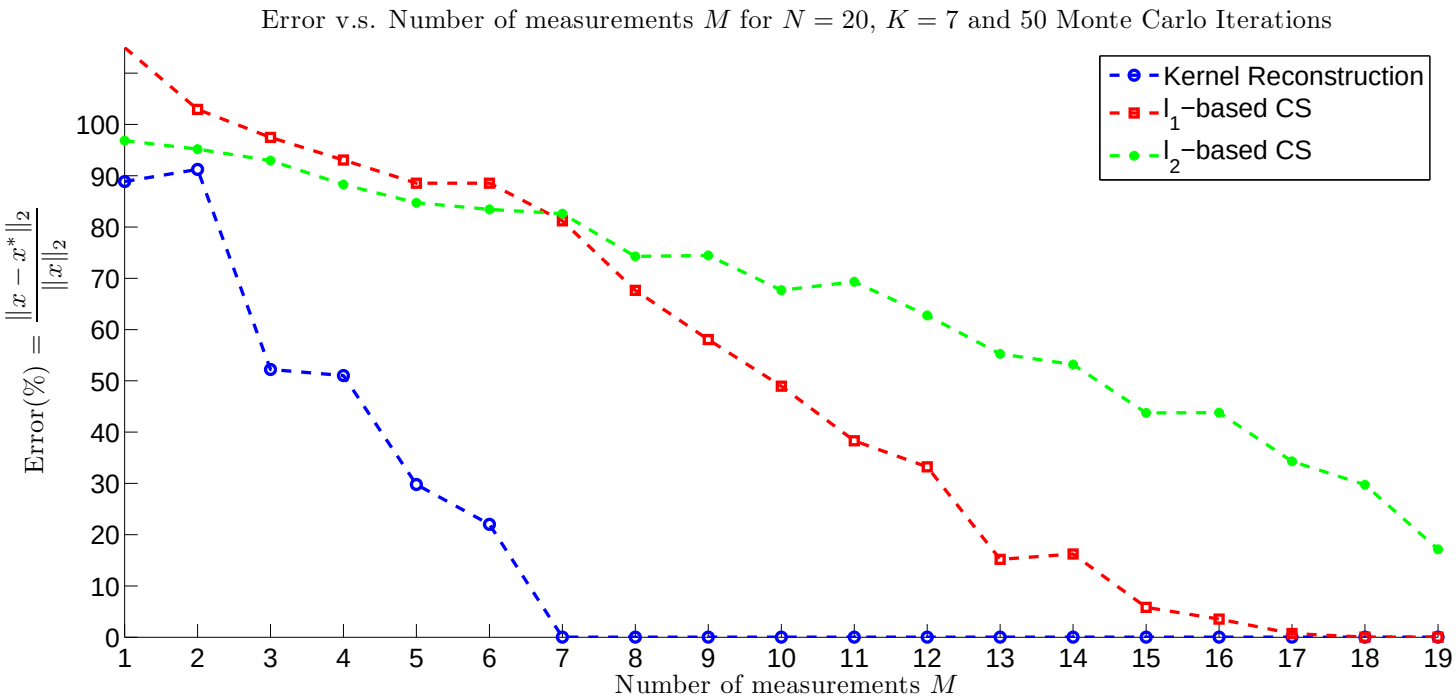
$$P_s\boldsymbol{x} = P_s\boldsymbol{x}_0 + P_sB\boldsymbol{a} = 0 \implies \boldsymbol{a} = -\left(P_sB\right)^{-1} P_s\boldsymbol{x}_0. \qquad (28)$$

Kernel Reconstruction v.s. ℓ_0-based CS

- The computational complexity of KR is $\mathscr{O}(N^S)$ since it requires $\binom{N}{S}$ operations in order to find the S linearly independent rows of B, hence the exact solution
- However, the computational complexity of the ℓ_0-based CS is $\mathscr{O}(N^K)$ since it requires $\sum_{k=1}^{K}\binom{N}{k}$ operations to find the exact solution.
- In order to have an exact solution, we should have $M \geq K+1$.
- For full rank matrix Φ, we have $S = N - M$. Therefore, we have $S \leq N-K-1$.
- Thus, for $K \geq \mathbb{E}(\frac{N-1}{2})$ we have $\mathscr{O}(N^S) << \mathscr{O}(N^K)$.

Simulation Results

Performance comparison of KR with ℓ_1-based and ℓ_2-based CS for $N = 20$



Conclusion

- In this work, we propose a new Algorithm that we call Kernel Reconstruction to infer K-sparse vectors from under-determined Systems.
- The KR guarantees an exact reconstruction for $M \geq K$.
- For $K \geq \mathbb{E}(\frac{N-1}{2})$, KR is computationally more efficient than the ℓ_0-based CS Algorithm.

Acknowledgement

- My most sincere thanks go to my advisor Dr. Nidhal Bouaynaya for her guidance, encouragement and support during the development of this work.
- This project is supported by Award Number R01GM096191 from the National Institute Of General Medical Sciences (NIH/NIGMS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute Of General Medical Sciences or the National Institutes of Health.

Bibliography

[1] Mark Davenport, Marco Duarte, Yonina Eldar, and Gitta Kutyniok. *Compressed Sensing: Theory and Applications*, chapter Introduction to compressed sensing. Cambridge University Press, 2012.

[2] David Donoho. Compressed sensing. *IEEE Trans. on Information Theory*, 52:1289–1306, April 2006.

[3] Fan Guo, Steve Hanneke, Wenjie Fu, and Eric P Xing. Recovering temporally rewiring networks: A model-based approach. In *Proceedings of the international conference on Machine learning*, pages 321–328, 2007.

[4] Hiroki Miyashita, Takuma Nakamura, Yasutoshi Ida, Takashi Matsumoto, and Takashi Kaburagi. Nonparametric Bayes-based heterogeneous "drosophila melanogaster" gene regulatory network inference: T-process regression. In *Proceedings of the International Conference on Artificial Intelligence and Applications*, pages 51–58, 2013.

[5] Joshua W Robinson and Alexander J Hartemink. Non-stationary dynamic Bayesian networks. In *Advances in Neural Information Processing Systems*, pages 1369–1376, 2008.

[6] Adam J Rothman, Elizaveta Levina, and Ji Zhu. Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, 2010.

[7] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, and Yi Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 31:210–227, February 2009.

[8] Lingxue Zhang and Seyoung Kim. Learning gene networks under SNP perturbations using eQTL datasets. *PLoS Computationaly Biology 10(4)*, 10, February 2014.

[9] Wentao Zhao, Erchin Serpedin, and Edward R Dougherty. Inferring gene regulatory networks from time series data using the minimum description length principle. *Bioinformatics*, 22(17):2129–2135, 2006.

Questions?