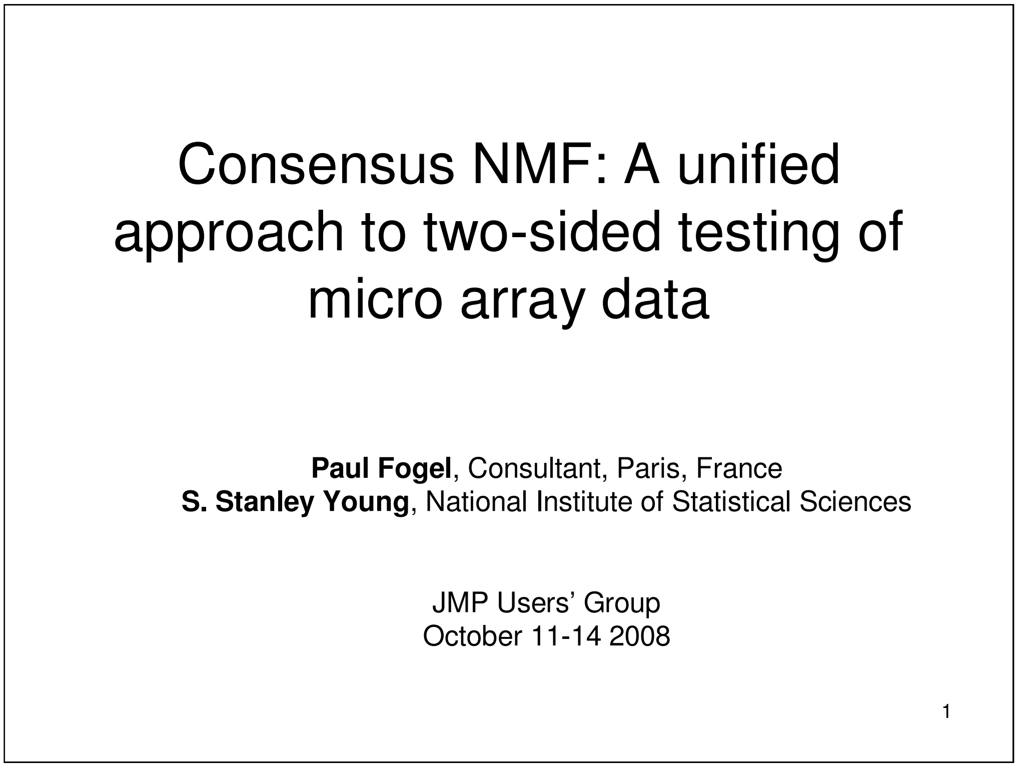
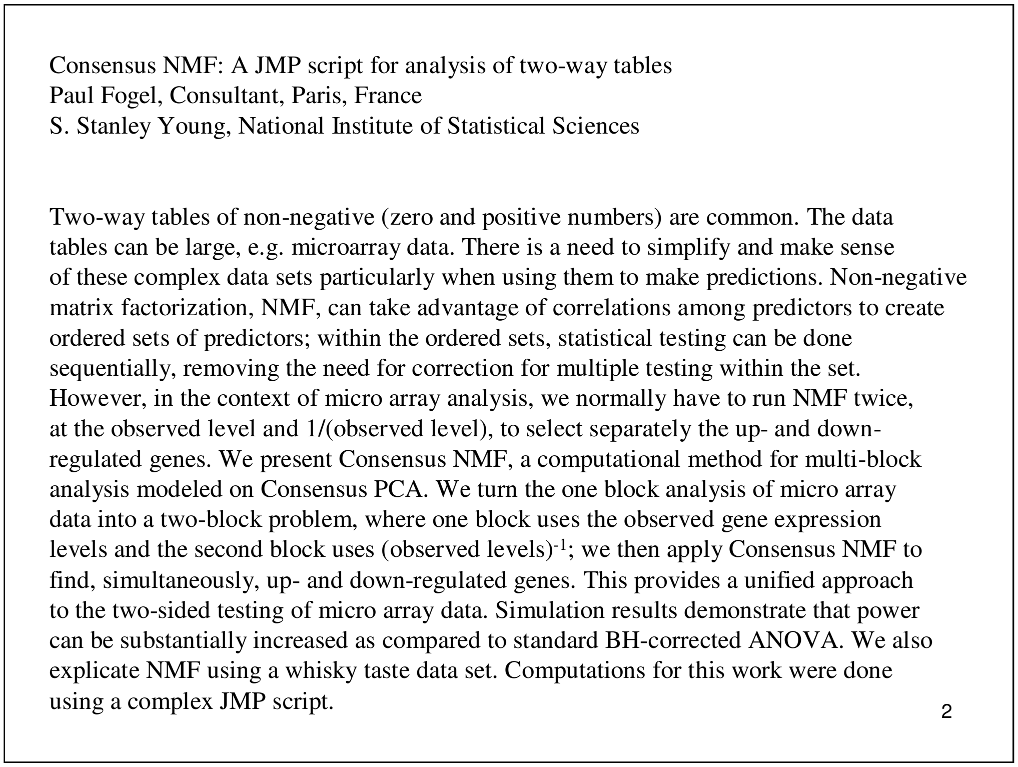


Consensus NMF: A unified approach to two-sided testing of micro array data

Paul Fogel, Consultant, Paris, France
S. Stanley Young, National Institute of Statistical Sciences

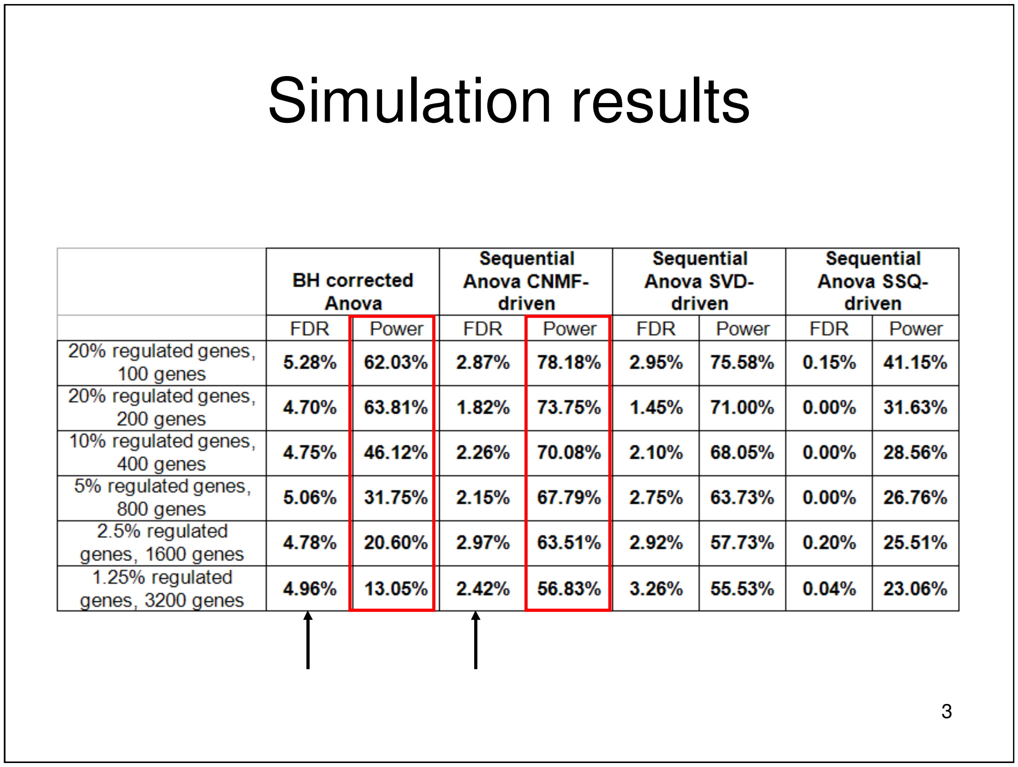
JMP Users' Group
October 11-14 2008



Consensus NMF: A JMP script for analysis of two-way tables

Paul Fogel, Consultant, Paris, France
S. Stanley Young, National Institute of Statistical Sciences

Two-way tables of non-negative (zero and positive numbers) are common. The data tables can be large, e.g. microarray data. There is a need to simplify and make sense of these complex data sets particularly when using them to make predictions. Non-negative matrix factorization, NMF, can take advantage of correlations among predictors to create ordered sets of predictors; within the ordered sets, statistical testing can be done sequentially, removing the need for correction for multiple testing within the set. However, in the context of micro array analysis, we normally have to run NMF twice, at the observed level and 1/(observed level), to select separately the up- and down-regulated genes. We present Consensus NMF, a computational method for multi-block analysis modeled on Consensus PCA. We turn the one block analysis of micro array data into a two-block problem, where one block uses the observed gene expression levels and the second block uses (observed levels)$^{-1}$; we then apply Consensus NMF to find, simultaneously, up- and down-regulated genes. This provides a unified approach to the two-sided testing of micro array data. Simulation results demonstrate that power can be substantially increased as compared to standard BH-corrected ANOVA. We also explicate NMF using a whisky taste data set. Computations for this work were done using a complex JMP script.

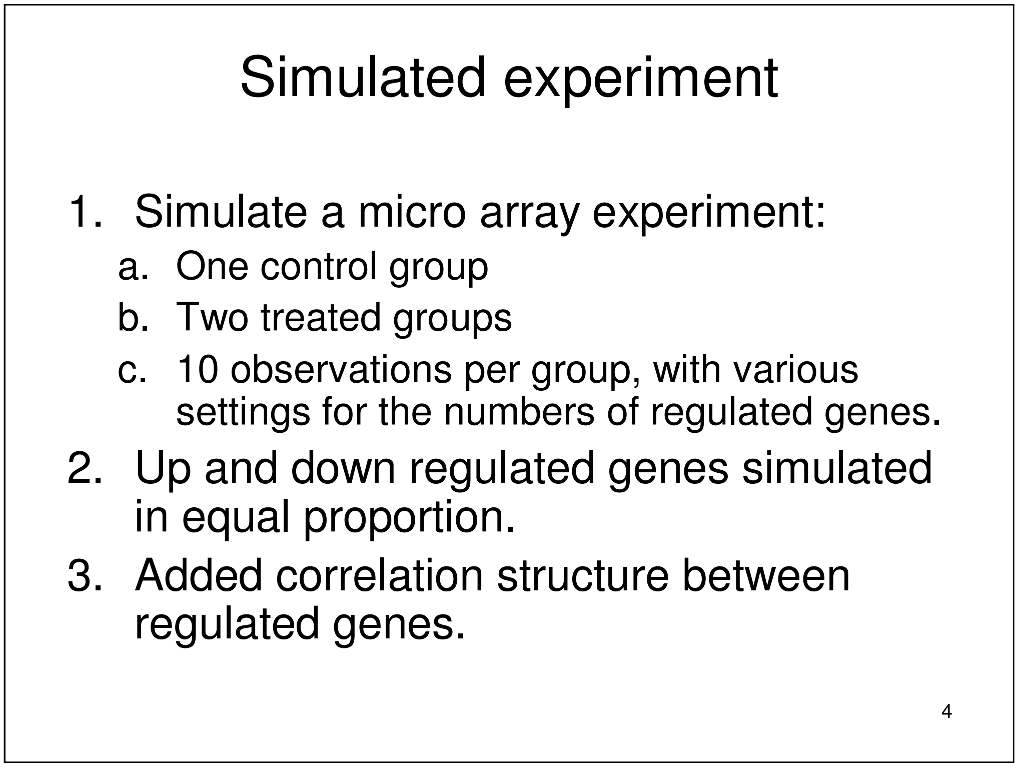


Simulation results

	BH corrected Anova		Sequential Anova CNMF-driven		Sequential Anova SVD-driven		Sequential Anova SSQ-driven	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power
20% regulated genes, 100 genes	**5.28%**	**62.03%**	**2.87%**	**78.18%**	**2.95%**	**75.58%**	**0.15%**	**41.15%**
20% regulated genes, 200 genes	**4.70%**	**63.81%**	**1.82%**	**73.75%**	**1.45%**	**71.00%**	**0.00%**	**31.63%**
10% regulated genes, 400 genes	**4.75%**	**46.12%**	**2.26%**	**70.08%**	**2.10%**	**68.05%**	**0.00%**	**28.56%**
5% regulated genes, 800 genes	**5.06%**	**31.75%**	**2.15%**	**67.79%**	**2.75%**	**63.73%**	**0.00%**	**26.76%**
2.5% regulated genes, 1600 genes	**4.78%**	**20.60%**	**2.97%**	**63.51%**	**2.92%**	**57.73%**	**0.20%**	**25.51%**
1.25% regulated genes, 3200 genes	**4.96%**	**13.05%**	**2.42%**	**56.83%**	**3.26%**	**55.53%**	**0.04%**	**23.06%**

We adopt a somewhat atypical presentation style. Here are the simulation results first. In the following slides we will fill you in on the method and background details.

First note that the power of our new procedure is much improved over Benjamini and Hockberg (2005). The gain in power is more substantial as the number of unregulated genes increases. More details later.



Simulated experiment

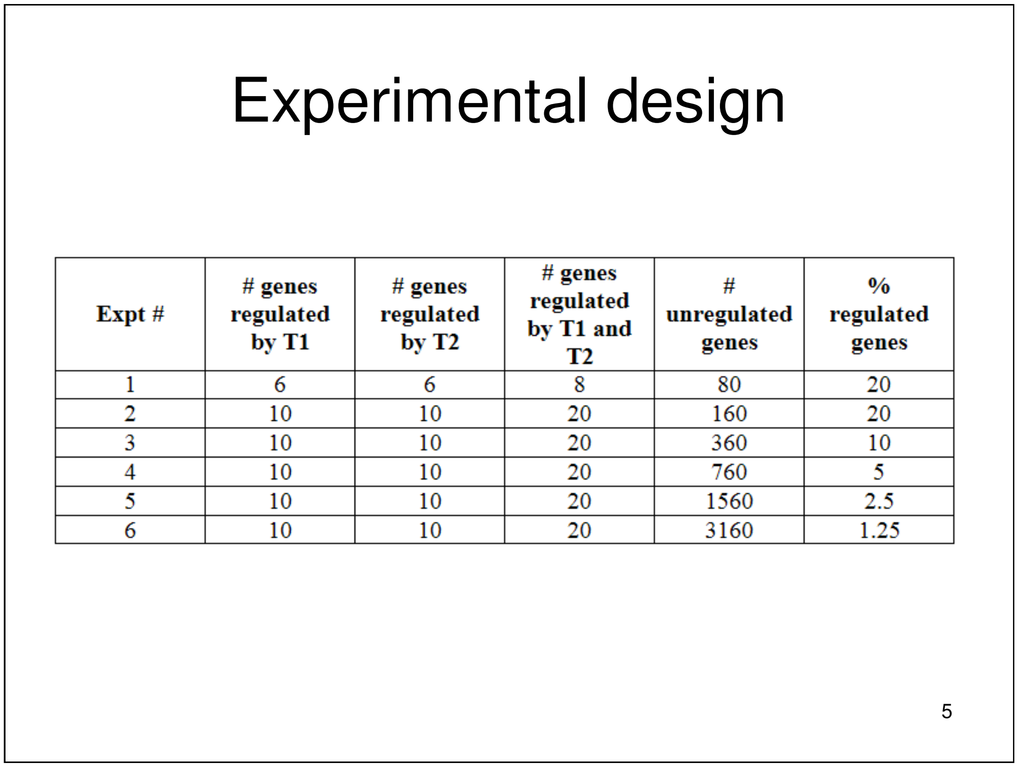
1. Simulate a micro array experiment:
   a. One control group
   b. Two treated groups
   c. 10 observations per group, with various settings for the numbers of regulated genes.
2. Up and down regulated genes simulated in equal proportion.
3. Added correlation structure between regulated genes.

4

Here is how the simulation was set up.

We have a control and two treatment groups, with 10 observations in each group.

We start with the proportion of treatment-related genes high and then increase the number of unregulated genes. Some genes increase with treatment, others decline. To make things more realistic, genes are correlated. There are some genes that are affected by both treatments.

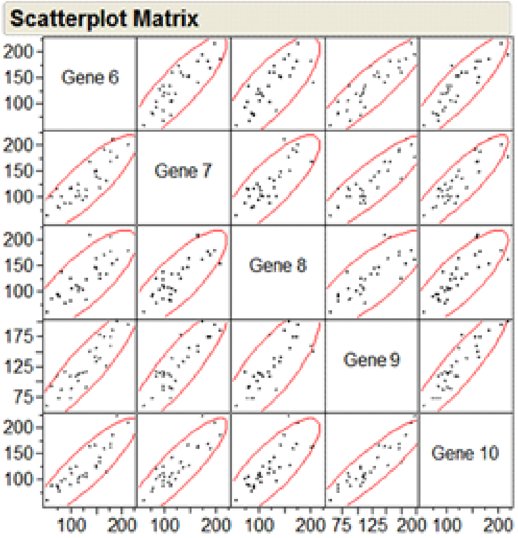


Experimental design

Expt #	# genes regulated by T1	# genes regulated by T2	# genes regulated by T1 and T2	# unregulated genes	% regulated genes
1	6	6	8	80	20
2	10	10	20	160	20
3	10	10	20	360	10
4	10	10	20	760	5
5	10	10	20	1560	2.5
6	10	10	20	3160	1.25

Here are details on the different simulated situations. We keep the number of regulated genes fairly constant and add more and more unregulated genes.

Simulation

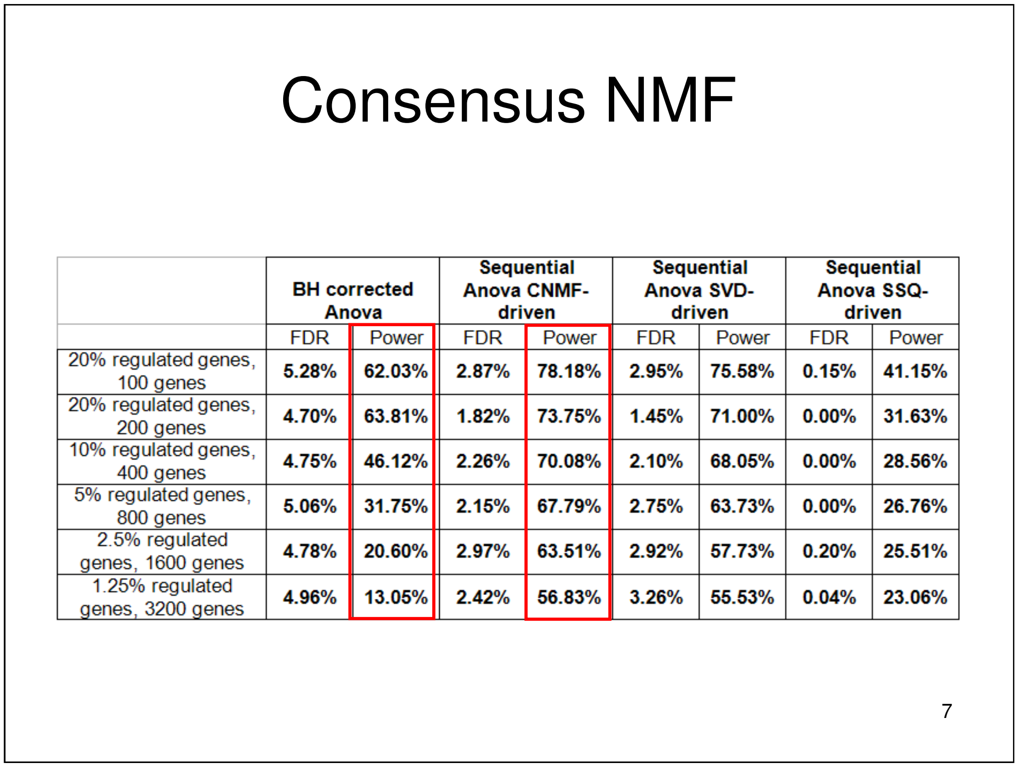


Genes 1-5: up-regulated by T1

Genes 6-10: up-regulated by T2

Genes 11-20: up-regulated by T1 and T2

This figure shows the added correlation structure.



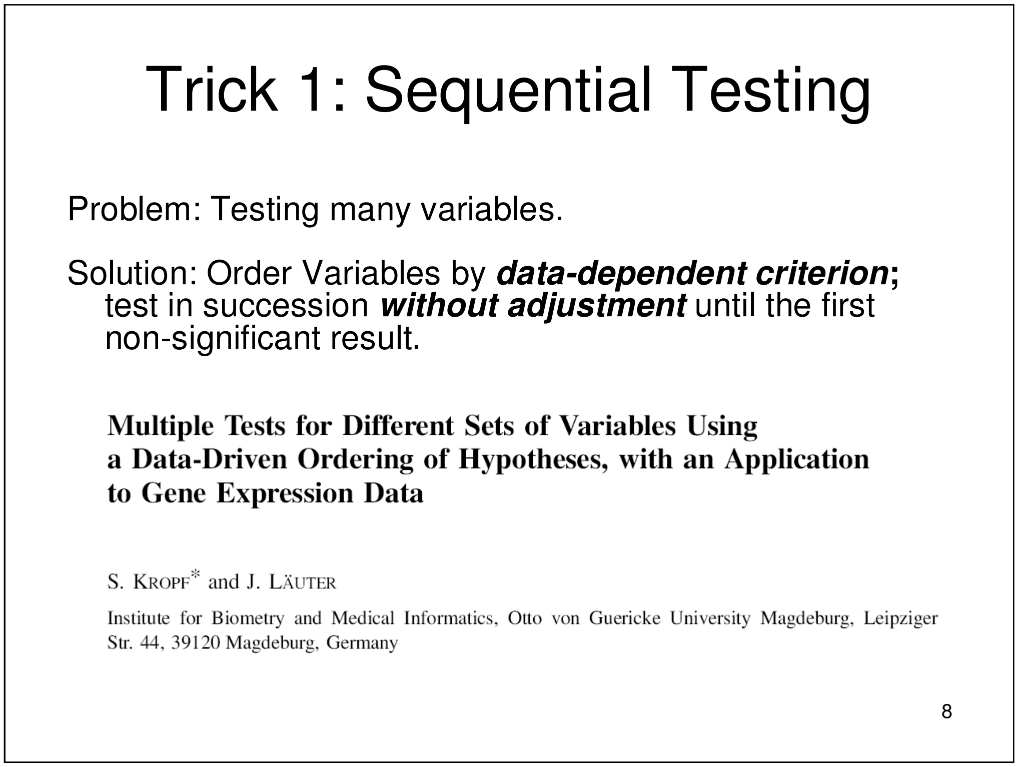
Consensus NMF

	BH corrected Anova		Sequential Anova CNMF-driven		Sequential Anova SVD-driven		Sequential Anova SSQ-driven	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power
20% regulated genes, 100 genes	**5.28%**	**62.03%**	**2.87%**	**78.18%**	**2.95%**	**75.58%**	**0.15%**	**41.15%**
20% regulated genes, 200 genes	**4.70%**	**63.81%**	**1.82%**	**73.75%**	**1.45%**	**71.00%**	**0.00%**	**31.63%**
10% regulated genes, 400 genes	**4.75%**	**46.12%**	**2.26%**	**70.08%**	**2.10%**	**68.05%**	**0.00%**	**28.56%**
5% regulated genes, 800 genes	**5.06%**	**31.75%**	**2.15%**	**67.79%**	**2.75%**	**63.73%**	**0.00%**	**26.76%**
2.5% regulated genes, 1600 genes	**4.78%**	**20.60%**	**2.97%**	**63.51%**	**2.92%**	**57.73%**	**0.20%**	**25.51%**
1.25% regulated genes, 3200 genes	**4.96%**	**13.05%**	**2.42%**	**56.83%**	**3.26%**	**55.53%**	**0.04%**	**23.06%**

Here we repeat the Power slide.

Again notice the increased power of the new method. BH controls the false discovery rate, as it should. Our new method has a lower FDR so in principle we could increase that and gain more power.

The other methods will be described later.



Trick 1: Sequential Testing

Problem: Testing many variables.

Solution: Order Variables by ***data-dependent criterion***; test in succession ***without adjustment*** until the first non-significant result.

Multiple Tests for Different Sets of Variables Using a Data-Driven Ordering of Hypotheses, with an Application to Gene Expression Data

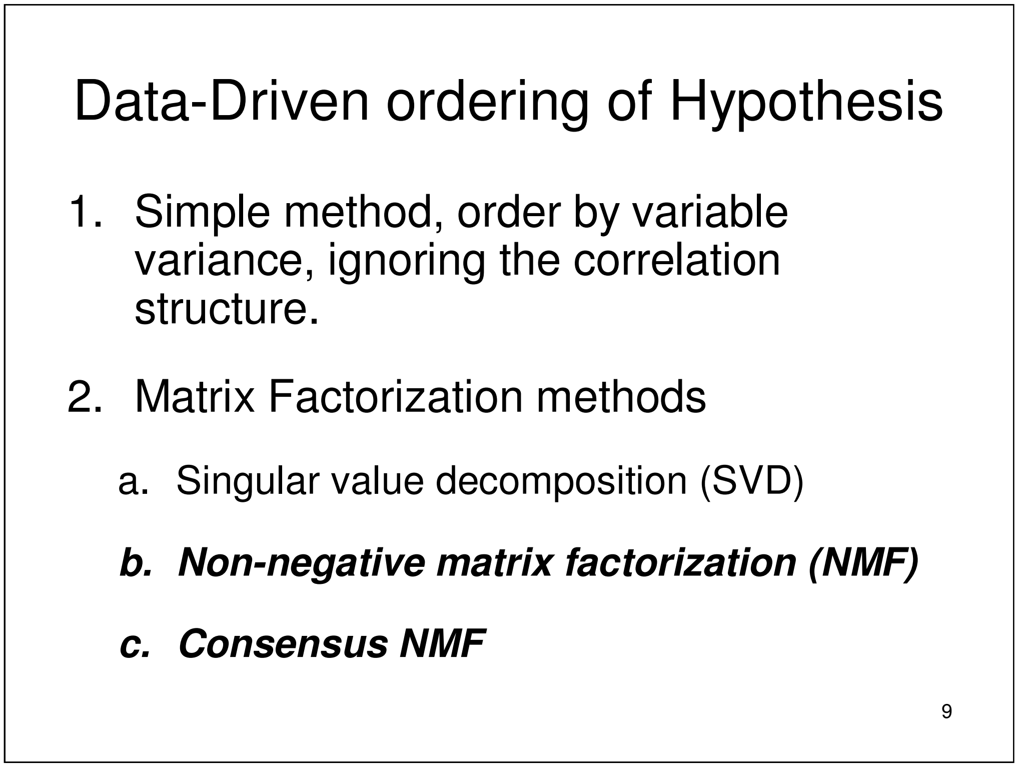
S. Kropf[*] and J. Läuter

Institute for Biometry and Medical Informatics, Otto von Guericke University Magdeburg, Leipziger Str. 44, 39120 Magdeburg, Germany

Three “tricks” are used in our method.

Note that if you test a series of hypotheses sequentially you can control the error rate by controlling the error rate for the first test. Basically you order the tests and test them sequentially, all at a specific level, say 0.05. You stop testing at the first non-significant result.

The question is How do you order the tests to take advantage of this idea? If you order the tests poorly, you will stop before you have found all the real effects.



Data-Driven ordering of Hypothesis

1. Simple method, order by variable variance, ignoring the correlation structure.

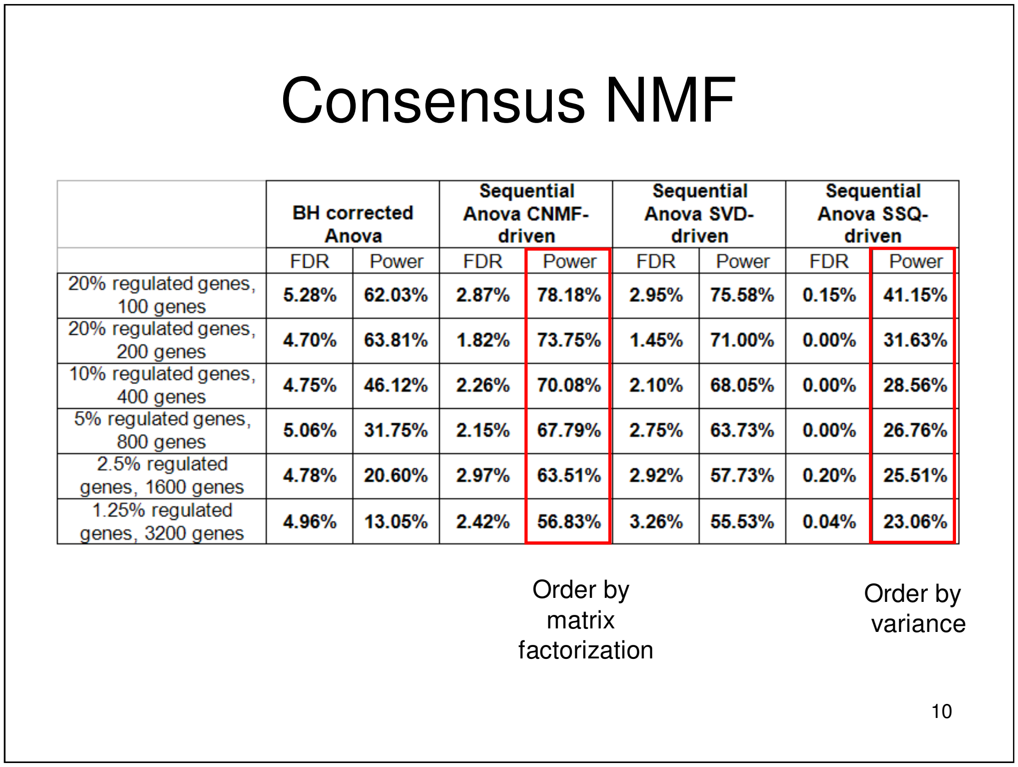
2. Matrix Factorization methods

    a. Singular value decomposition (SVD)

    b. Non-negative matrix factorization (NMF)

    c. Consensus NMF

We chose to order the variables by the elements of the vectors of a matrix factorization. We will look at two factorizations, singular value decomposition and non-negative matrix factorization. Note that the factorizations only look at the genes, not the treatments so no treatment information is used in the ordering of the genes.



Consensus NMF

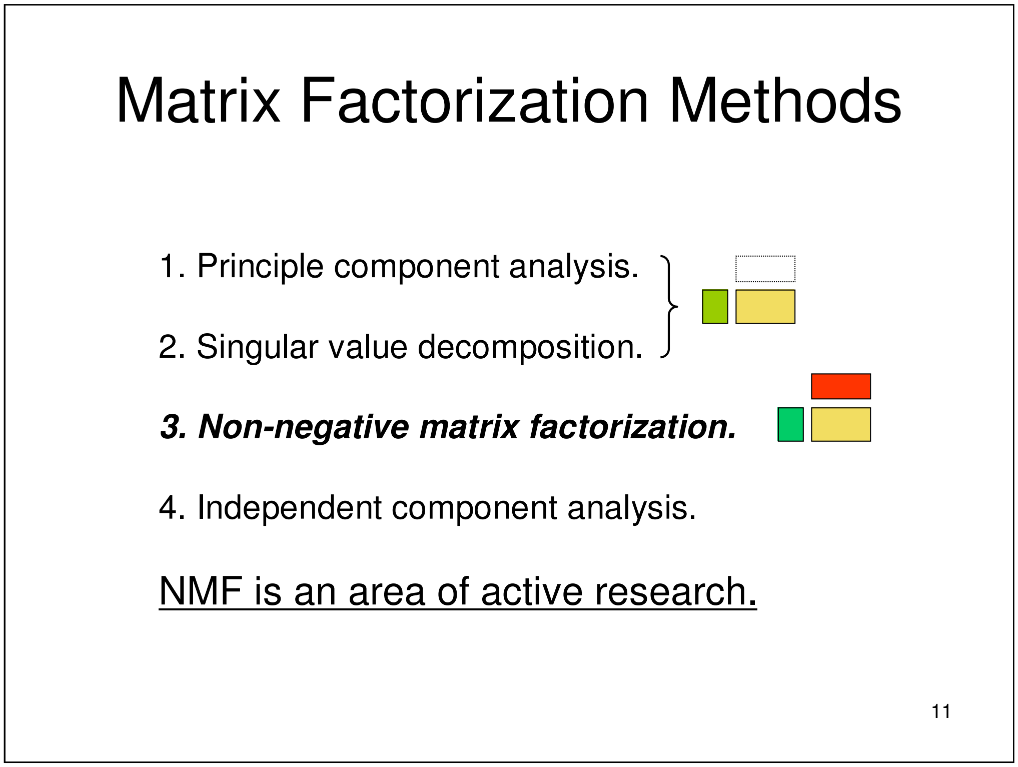
	BH corrected Anova		Sequential Anova CNMF-driven		Sequential Anova SVD-driven		Sequential Anova SSQ-driven	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power
20% regulated genes, 100 genes	**5.28%**	**62.03%**	**2.87%**	**78.18%**	**2.95%**	**75.58%**	**0.15%**	**41.15%**
20% regulated genes, 200 genes	**4.70%**	**63.81%**	**1.82%**	**73.75%**	**1.45%**	**71.00%**	**0.00%**	**31.63%**
10% regulated genes, 400 genes	**4.75%**	**46.12%**	**2.26%**	**70.08%**	**2.10%**	**68.05%**	**0.00%**	**28.56%**
5% regulated genes, 800 genes	**5.06%**	**31.75%**	**2.15%**	**67.79%**	**2.75%**	**63.73%**	**0.00%**	**26.76%**
2.5% regulated genes, 1600 genes	**4.78%**	**20.60%**	**2.97%**	**63.51%**	**2.92%**	**57.73%**	**0.20%**	**25.51%**
1.25% regulated genes, 3200 genes	**4.96%**	**13.05%**	**2.42%**	**56.83%**	**3.26%**	**55.53%**	**0.04%**	**23.06%**

Order by matrix factorization

Order by variance

The first idea for ordering the tests is to order them by the variance of each variable. If there is large variance, then presumably there is signal in addition to the noise. Variables with a lot of signal will be moved to the beginning of the sequential testing.

In our simulation, the new method has more power than sequential ANOVA (variables ordered by variance). The FDR for this method is quite low so the power could be improved by accepting a higher FDR. The new method appears to “win” but that may be because of the low FDR of Sequential ANOVA.



Matrix Factorization Methods

1. Principle component analysis.
2. Singular value decomposition.
3. ***Non-negative matrix factorization.***
4. Independent component analysis.

NMF is an area of active research.

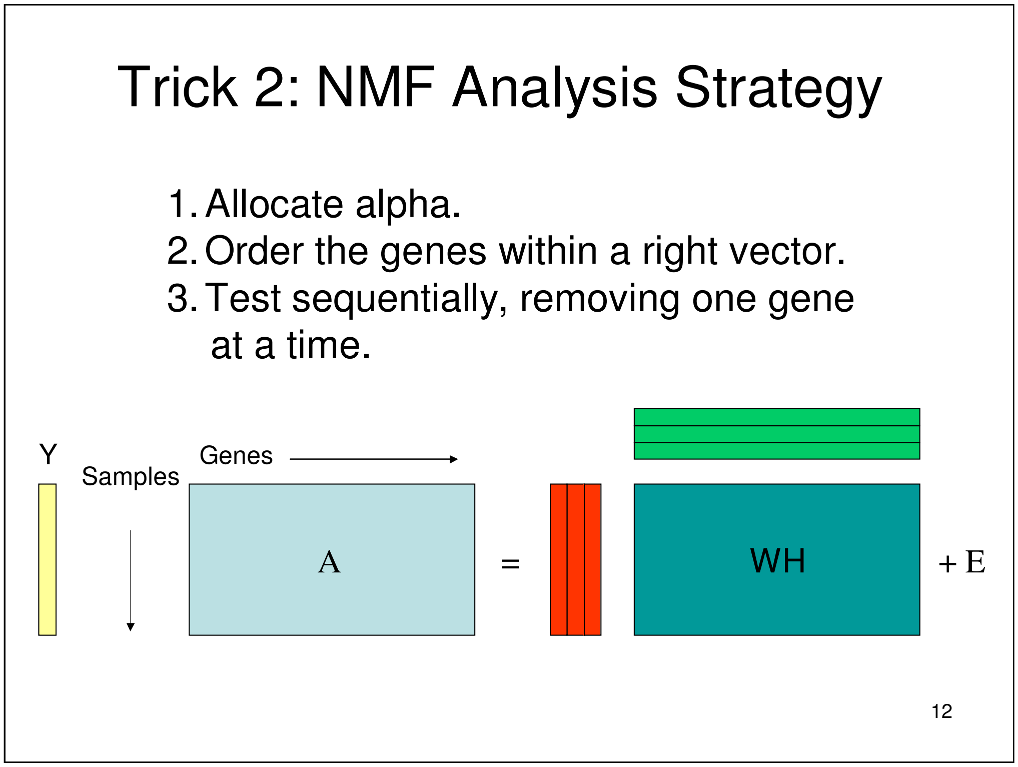
PCA is well known. A matrix can be factored into “loadings” (roughly measuring the importance of the variables) and the “scores” the projected positions in a lower dimension space.

PCA is just another term for singular value decomposition.

Non-negative matrix factorization is new to most statisticians. It looks like SVD. The difference is that the matrix to be factored has all non-negative elements and the elements of the factoring matrices all are constrained to be non-negative as well. NMF is a difficult computational problem that we will ignore.

NMF is a very active area of research for computer science.

Note that SVD is the basis for a lot of statistical methods. If the design matrix is positive, then NMF can be substituted for SVD creating a parallel analysis world.



Trick 2: NMF Analysis Strategy

1. Allocate alpha.
2. Order the genes within a right vector.
3. Test sequentially, removing one gene at a time.

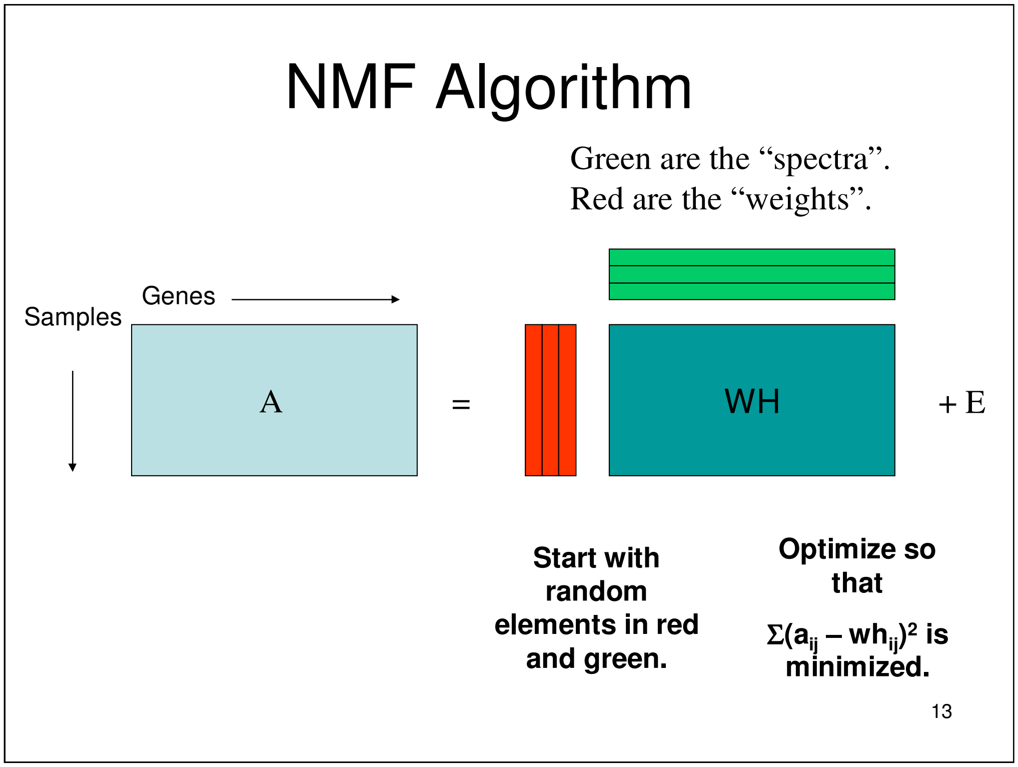
Y
Samples
Genes
A
=
WH
+ E

12

We come to the 2nd trick. We “alpha spend” dividing alpha over the rows of the factoring matix. If there are 3 rows, then you can allocate 0.05/3 to each row.

Remember the elements of the right factoring vectors are positive. We order them from largest (most important) to smallest and test them in the resulting order.

In practice, we re-compute the factorization after removing the columns of A in turn. This is complex and will not be discussed here.



NMF Algorithm

Green are the “spectra”.
Red are the “weights”.

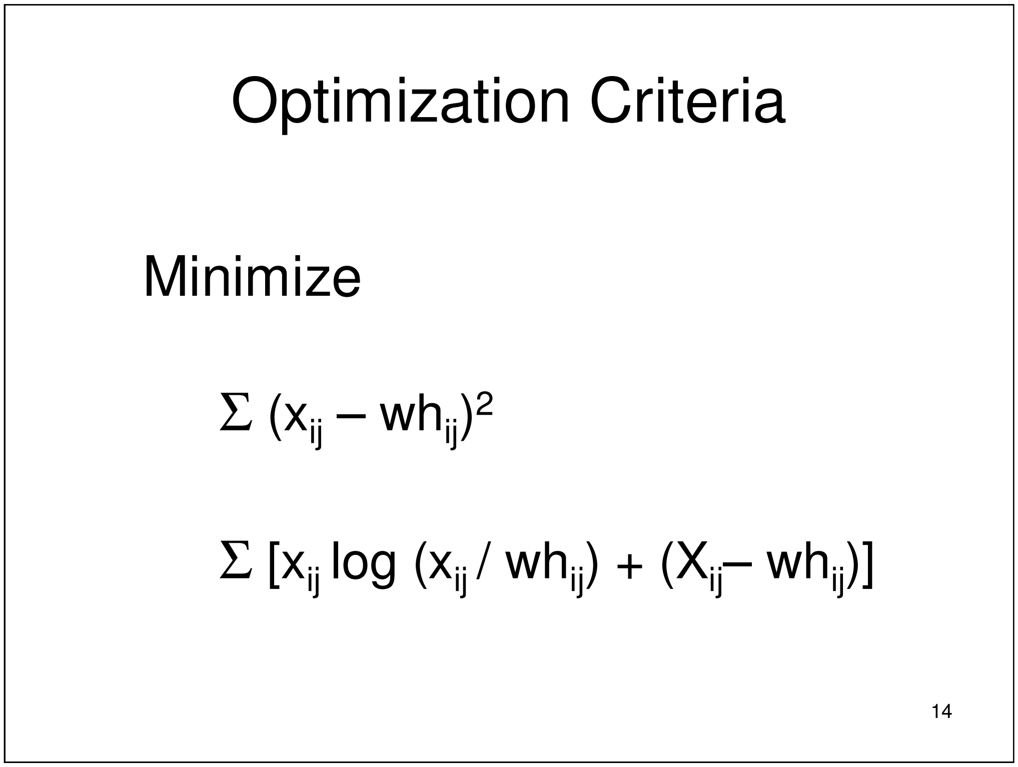
Start with random elements in red and green.

Optimize so that

$\Sigma(a_{ij} – wh_{ij})^2$ is minimized.

13

We give an idea of the computational details of NMF. The computations start with random positive elements in the right and left factoring vectors. The factoring vectors are re-computed seeking to minimize the squared element wise differences between A and WH.



Optimization Criteria

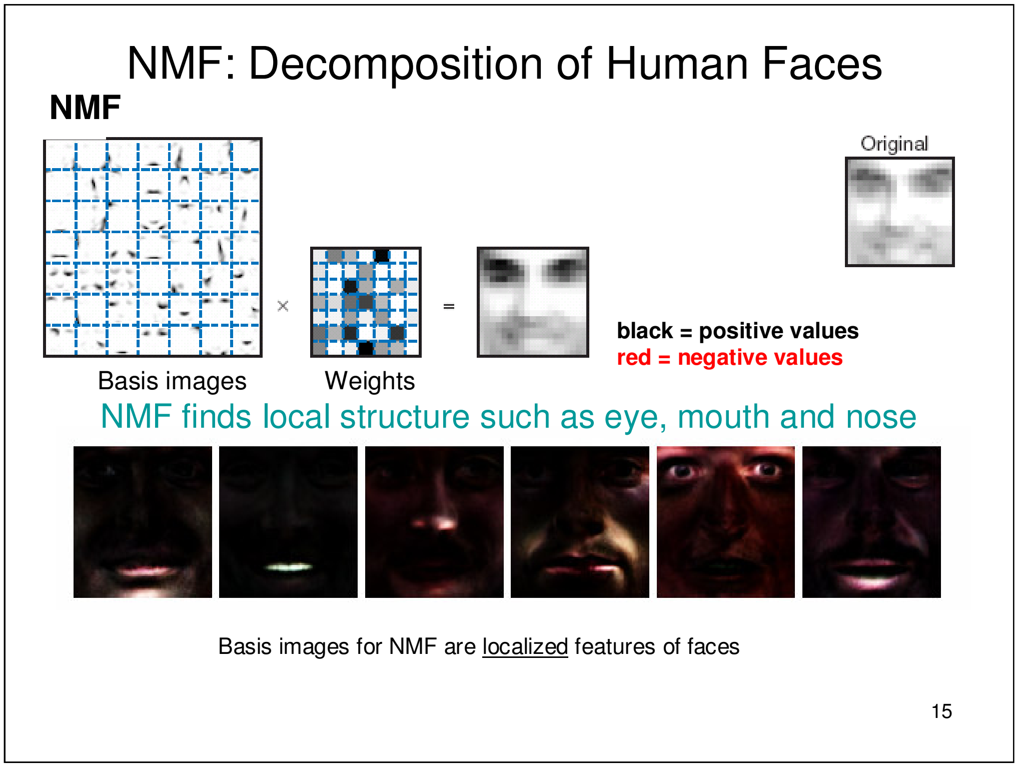
Minimize

$$\Sigma\ (x_{ij} - wh_{ij})^2$$

$$\Sigma\ [x_{ij} \log (x_{ij} / wh_{ij}) + (X_{ij} - wh_{ij})]$$

14

There are two optimization criteria commonly used, least squared difference and an “information” criterion. Note that both go toward zero the better the solution.



NMF: Decomposition of Human Faces

NMF

Original

× =

Basis images Weights

black = positive values

red = negative values

NMF finds local structure such as eye, mouth and nose

Basis images for NMF are <u>localized</u> features of faces

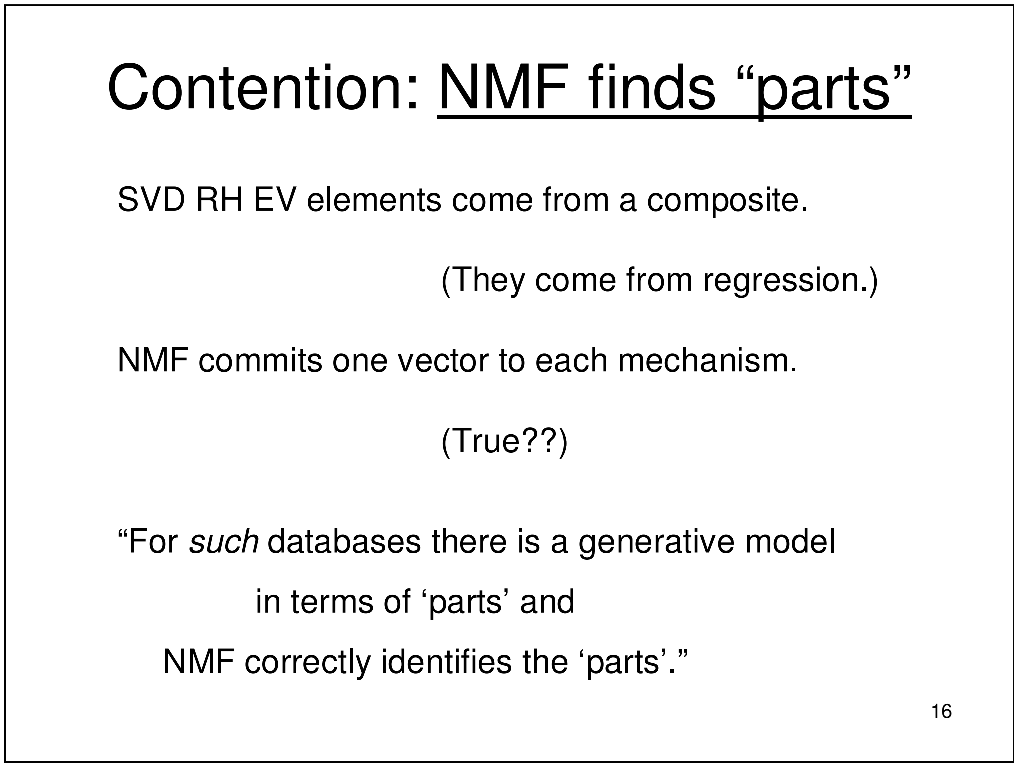
15

NMF is distinguished by its use of non-negative constraints on matrix decomposition to allow only additive combinations.

NMF basis is radically different than SVD, presented later. Its images are localized features that correspond better with intuitive notions of the parts of faces.

There was an intuitive leap that NMP finds “parts”. In biology land, groups of genes involved in a single mechanism are put together.

True? There is empirical support that this is happening. If true, it radically simplifies interpretation.



Contention: NMF finds "parts"

SVD RH EV elements come from a composite.

(They come from regression.)

NMF commits one vector to each mechanism.

(True??)

"For *such* databases there is a generative model
in terms of 'parts' and
NMF correctly identifies the 'parts'."

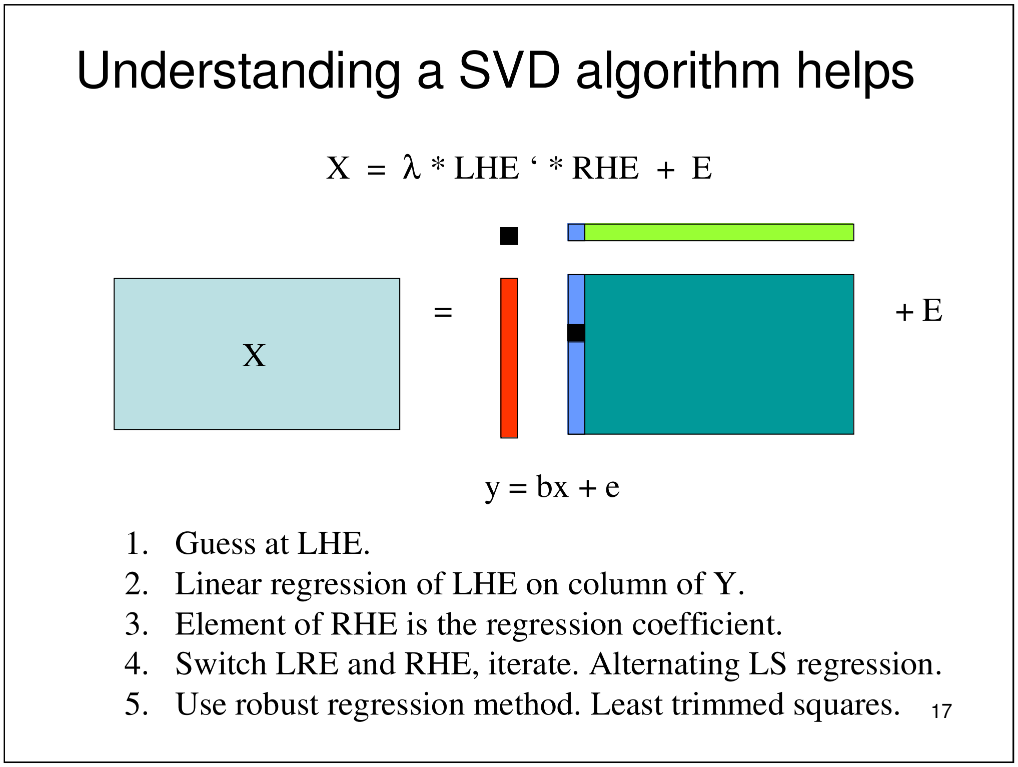
Why bother with NMF??

It is contended that if the starting matrix is the addition of matrices (mechanisms) then NMF will correctly put separate mechanisms in separate vectors of the factoring matrices.

This is a big deal. The elements of the factoring vectors point to the genes that are co-regulated, for example.

Some math and considerable empirical results support this contention for microarray data.

This area is still "Arts and Crafts".



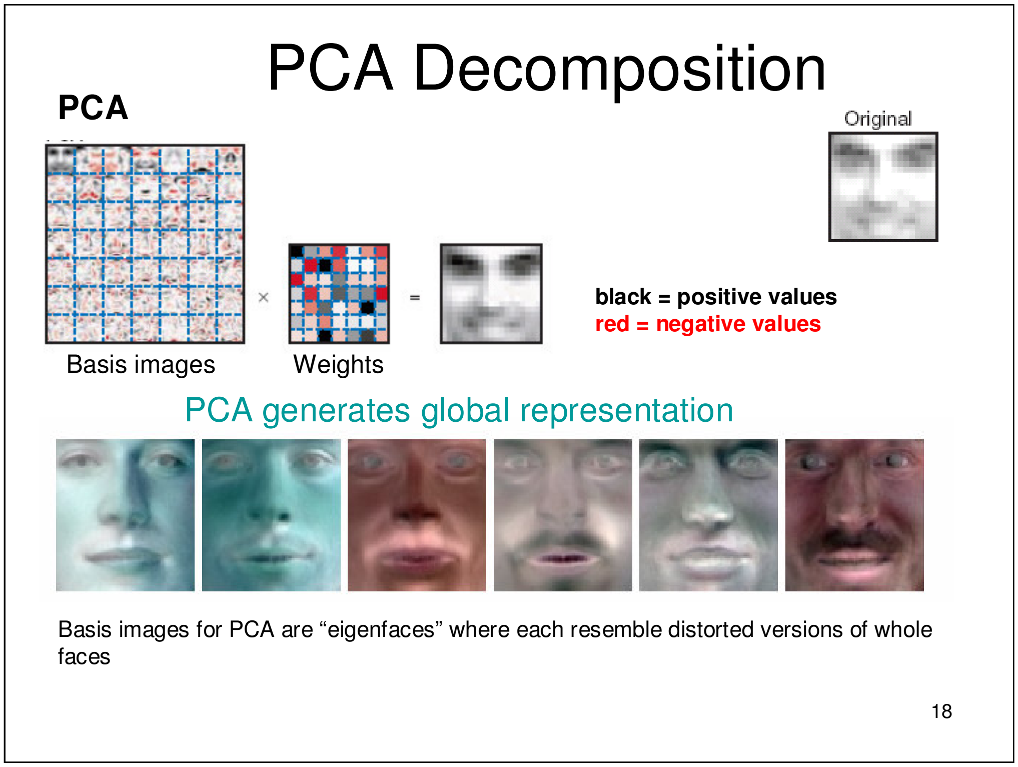
Understanding a SVD algorithm helps

$X = \lambda * LHE ' * RHE + E$

$y = bx + e$

1. Guess at LHE.
2. Linear regression of LHE on column of Y.
3. Element of RHE is the regression coefficient.
4. Switch LRE and RHE, iterate. Alternating LS regression.
5. Use robust regression method. Least trimmed squares.

Let’s backtrack and understand SVD. Understanding the alternating least squares SVD algorithm gives us insight into the interpretive problems with SVD.



PCA Decomposition

PCA

Original

Basis images × Weights =

black = positive values

red = negative values

PCA generates global representation

Basis images for PCA are “eigenfaces” where each resemble distorted versions of whole faces

18

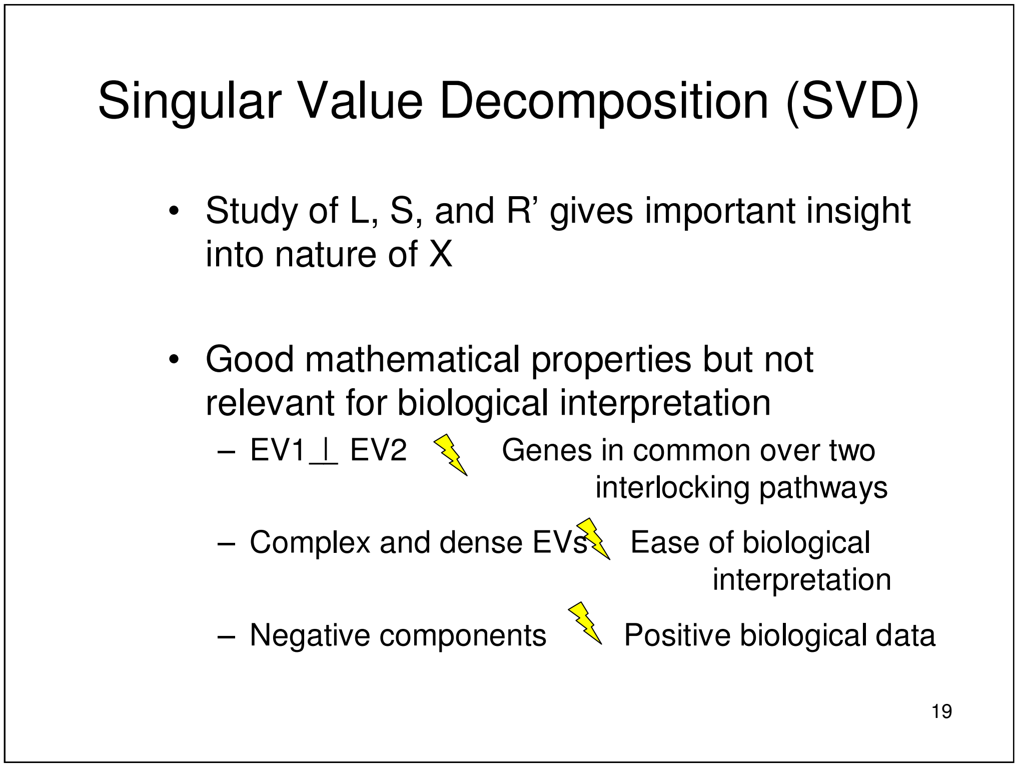
PCA allows the arbitrary signs in matrix decomposition. This involves complex cancellations between positive and negative numbers.

Basis images for PCA are eigenfaces some of which resemble distorted versions of whole faces

Black pixels=pos values red pixels = neg values

SVD comes of as the addition and subtraction of “holographic” faces.

If there are multiple mechanisms, then they are embedded in each factoring vector pair.



Singular Value Decomposition (SVD)

- Study of L, S, and R’ gives important insight into nature of X
- Good mathematical properties but not relevant for biological interpretation
  - EV1 $\perp$ EV2 ⚡ Genes in common over two interlocking pathways
  - Complex and dense EVs ⚡ Ease of biological interpretation
  - Negative components ⚡ Positive biological data

19

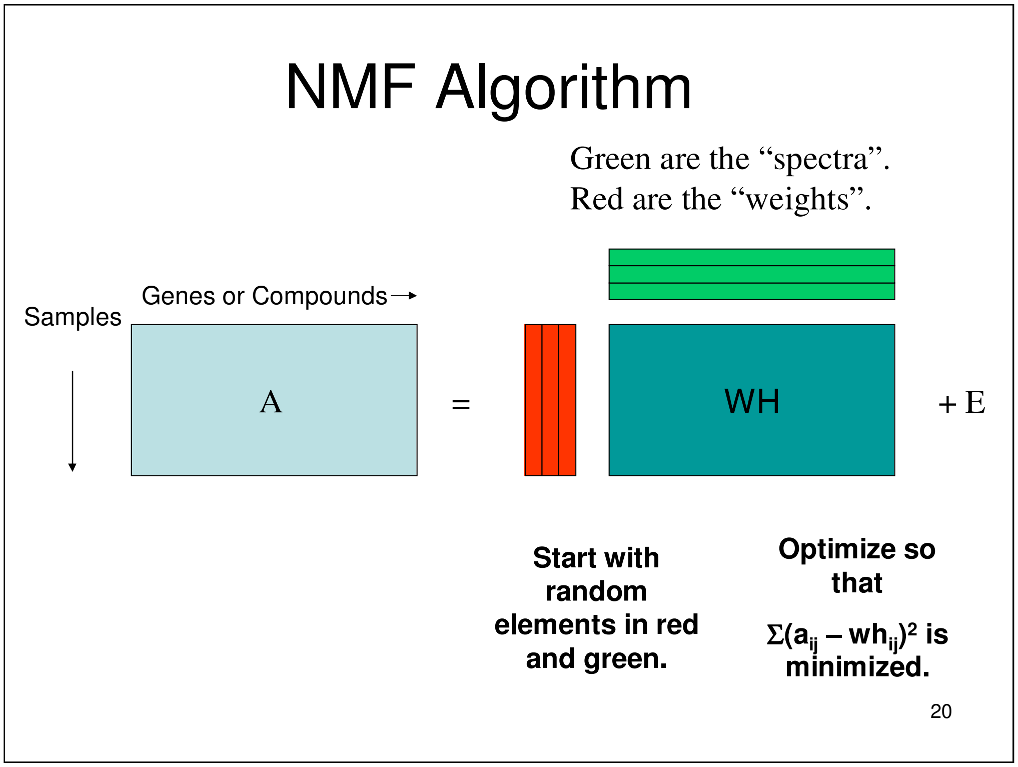
Interpretation of SVD is not easy at all. 10M number is nearly impossible. There are several problems in the reduction to 23k. Interpretation of L, S, and R’ is something of a high art that depends on both statistics and subject matter knowledge.

The mathematical interpretation is straightforward. The data is projected to a lower dimensional space. The weights use in the projection are orthogonal. The approximation of X is as good as it can be for a given k in a least squares sense. If you sum up the squared deviations between the observed data and the approximation it will be as small as possible. It is a triumph of mathematics.

There are problems. Genes can be in overlapping biochemical pathways so gene levels are unlikely to be orthogonal. Also, L and R’ can be quite “dense/compact” so that understanding what is going on can be (typically is) quite difficult.

Just as in simple arithmetic, it sometimes make no sense to have negative numbers so it is true for matrix multiplication. (–3x-2=6 and 3x2=6) The physics or science might dictate that the elements of X are positive and the elements of L and R’ should be positive as well.

Global vs local patterns. SVD seems to find global patterns when the problem calls for finding local patterns.



NMF Algorithm

Green are the “spectra”.
Red are the “weights”.

Samples

Genes or Compounds →

A

=

WH

+ E

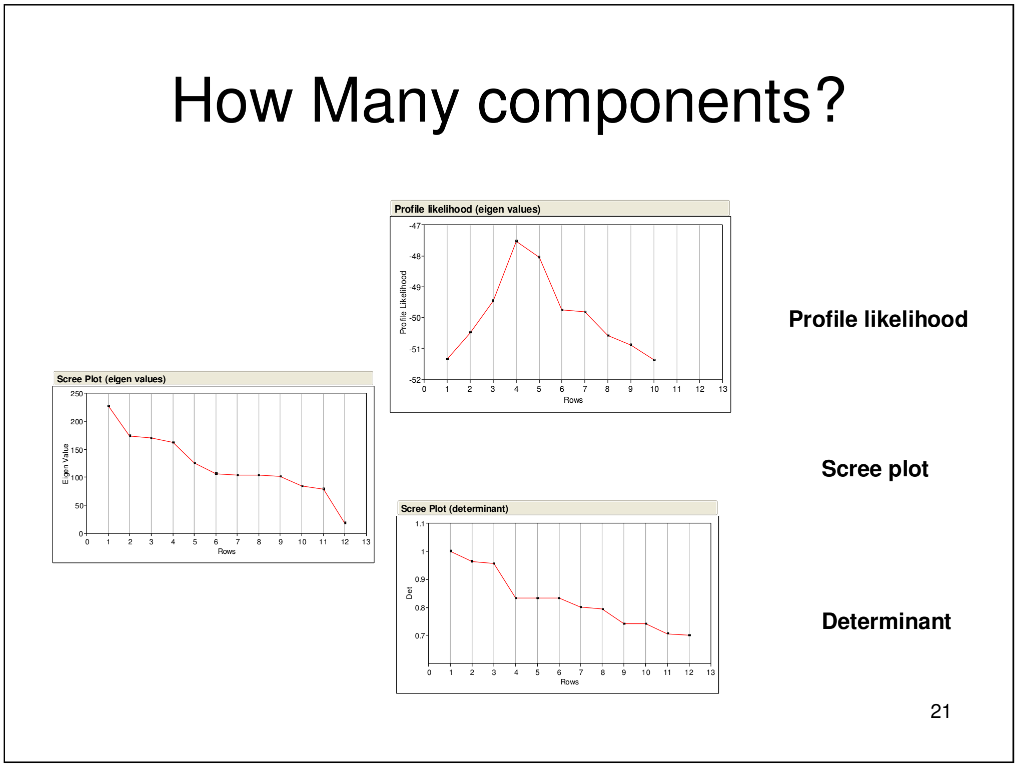
Start with random elements in red and green.

Optimize so that

$\Sigma(a_{ij} - wh_{ij})^2$ is minimized.

20

Here is the block diagram. Next we will see the NMF of a Scotch Whisky data set.



How Many components?

Profile likelihood

Scree plot

Determinant

21

How many flavor patterns are present?

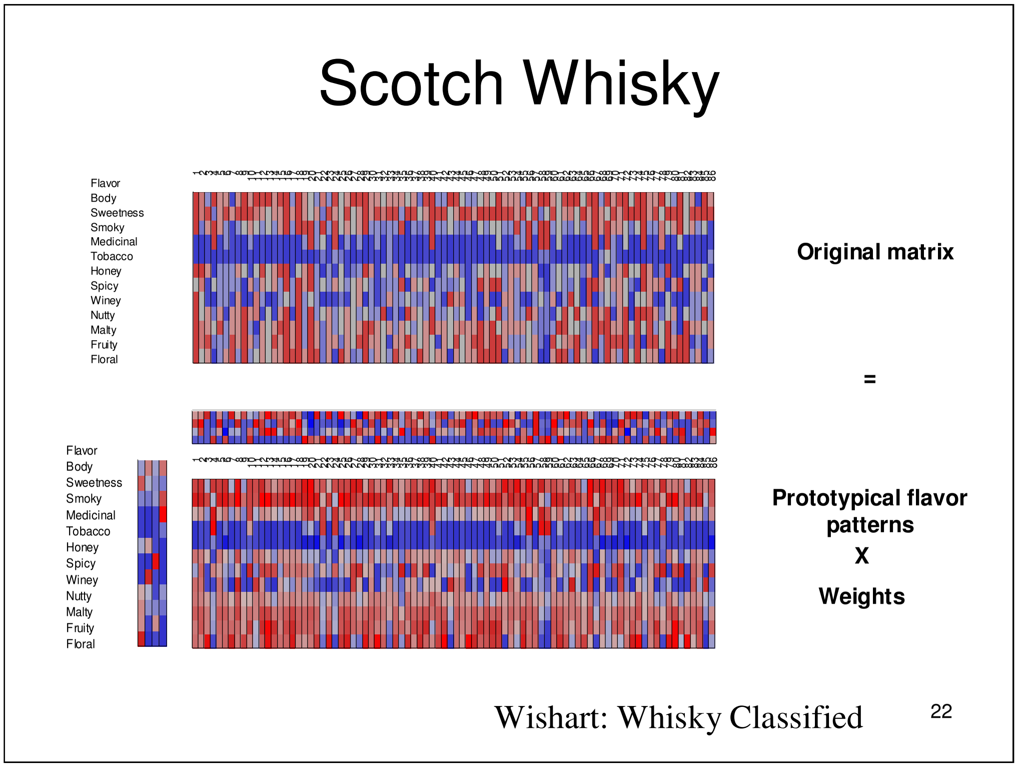
Obviously considerable care has been given to the various flavors chosen to characterize Scotch whisky as there is no dramatic clustering of the flavors. Even so, there appear to be “jumps” in the Scree plot when going from 10 to 9 factors and from 5 to 4.

Zhu and Ghodsi give a method defined as profile likelihood, for evaluation of a Scree plot that gives the likelihood of a mixture distribution – where the Scree plot can be cut so that noise components are to the right and signal components are to the left. The Scree plot for the profile likelihood method is shown in Figure 2.

How many flavor patterns are present?

Obviously considerable care has been given to the various flavors chosen to characterize Scotch whisky as there is no dramatic clustering of the flavors. Even so, there appear to be “jumps” in the Scree plot when going from 10 to 9 factors and from 5 to 4.

Zhu and Ghodsi give a method defined as profile likelihood, for evaluation of a Scree plot that gives the likelihood of a mixture distribution – where the Scree plot can be cut so that noise components are to the right and signal components are to the left. The Scree plot for the profile likelihood method is shown in Figure 2.



Scotch Whisky

Original matrix

=

Prototypical flavor patterns

X

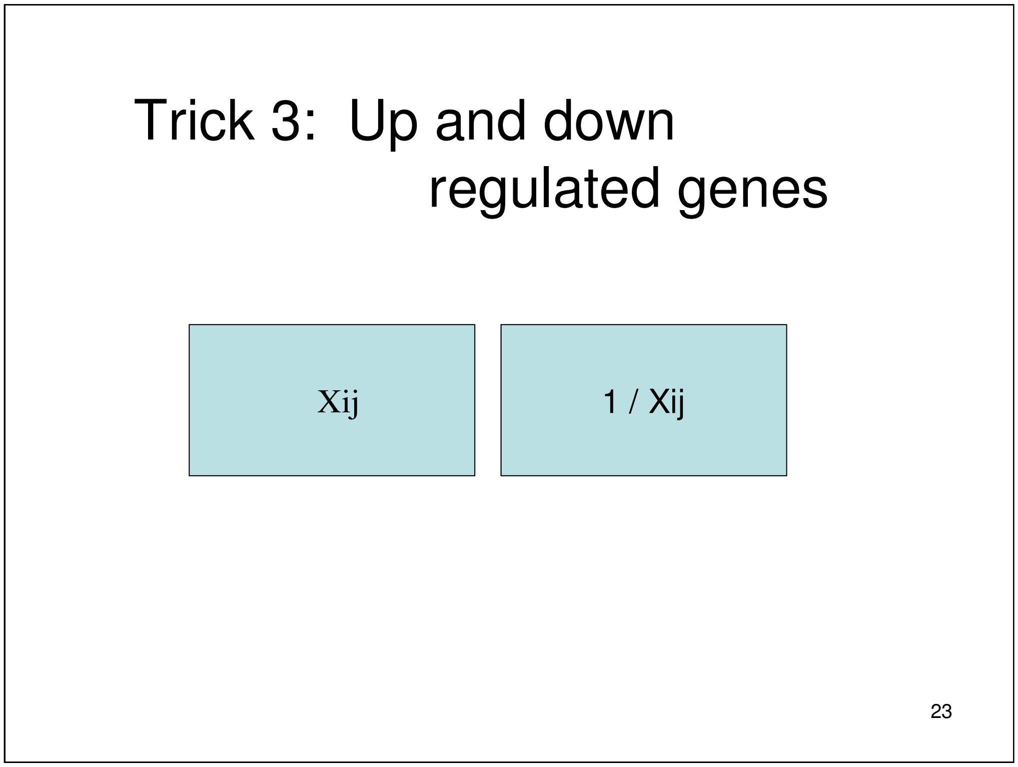
Weights

Wishart: Whisky Classified 22

The upper matrix is the raw data. The lower matrix is the NMF approximation. The factoring matrices are given above and to the right of the approximation.

In effect, four “prototypical” Scotch Whiskies could be used to approximate the 87 single malt whiskies!!!

Note well: Much of the favor of Scotch Whiskies comes from the barrel aging. There appear to be four aging strategies.



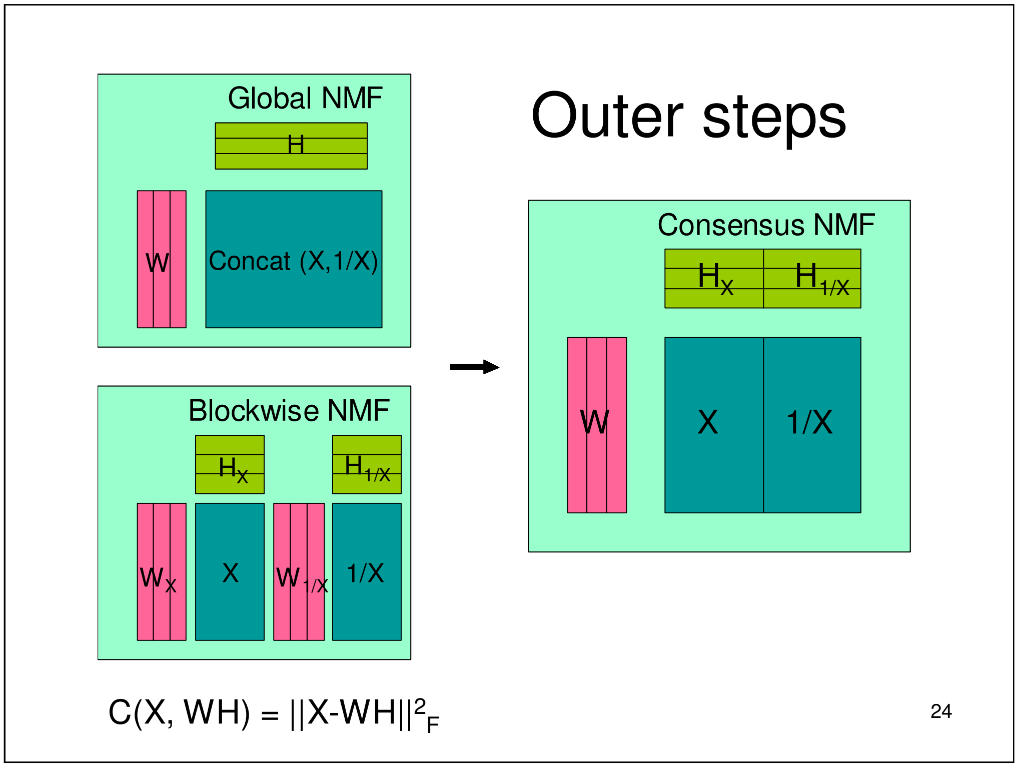
Trick 3: Up and down regulated genes

Xij

1 / Xij

23

We come to our third trick. We really want to find both up and down regulated genes. To do this, we append a matrix of element wise inverses to the original data matrix. This is a simple trick, once you see it, but it presents an analysis strategy problem. How do you weight the two matrices?



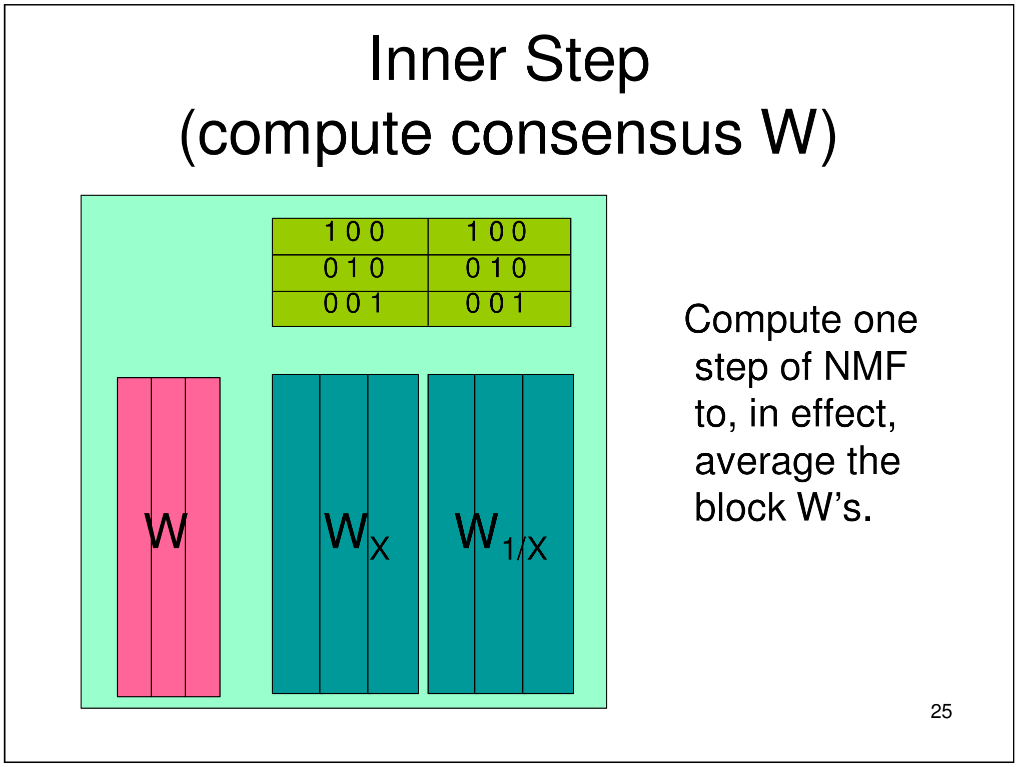
Outer steps

$C(X, WH) = ||X-WH||^2_F$

Here is the analysis strategy in blocks.

First factor the concatenated matrix without regard for the two parts to get an initial estimate of the factoring matrices, W and H.

Next factor each block separately.



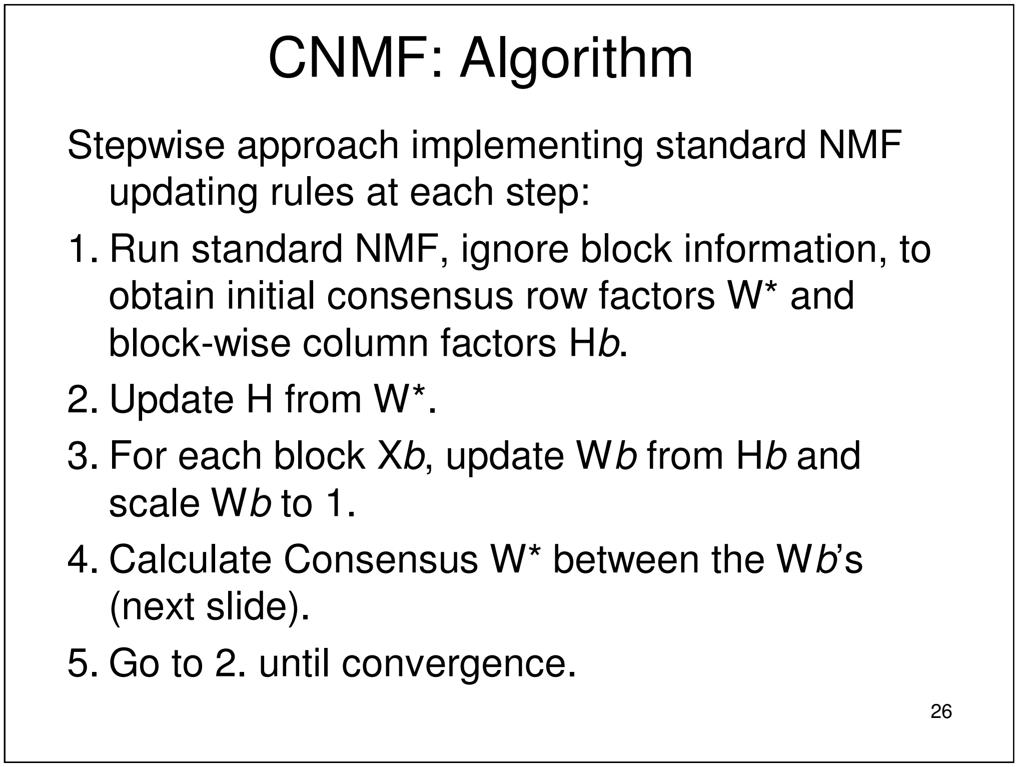
Inner Step (compute consensus W)

1 0 0	1 0 0
0 1 0	0 1 0
0 0 1	0 0 1

Compute one step of NMF to, in effect, average the block W’s.

25

Start with Wx and W1/x and use one step of NMF to compute a new W, the consensus. In effect we “average” Wx and W1/x.



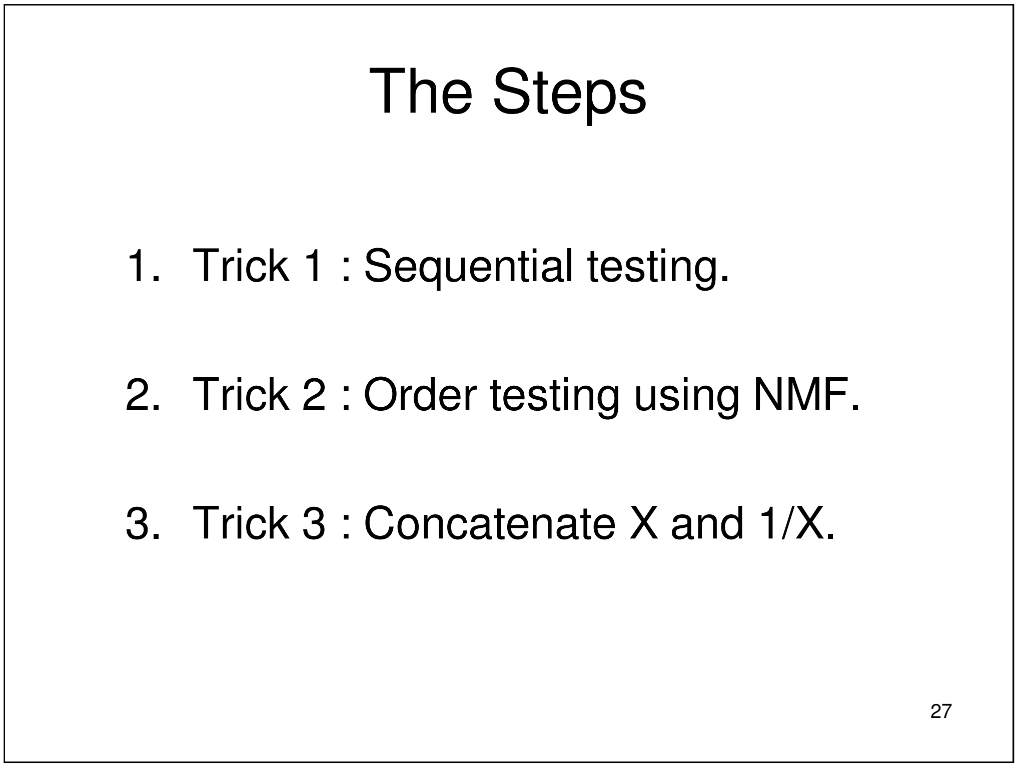
CNMF: Algorithm

Stepwise approach implementing standard NMF updating rules at each step:

1. Run standard NMF, ignore block information, to obtain initial consensus row factors W* and block-wise column factors H*b*.
2. Update H from W*.
3. For each block X*b*, update W*b* from H*b* and scale W*b* to 1.
4. Calculate Consensus W* between the W*b*’s (next slide).
5. Go to 2. until convergence.

Here is the algorithm in words.

Having factored each block separately, we need to reach a consensus of the relative importance of the two parts, up and down regulated genes.



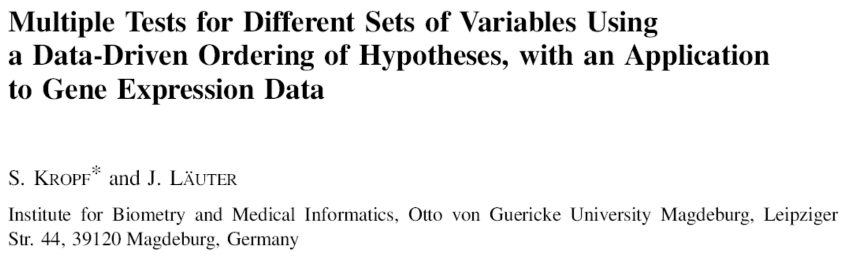
The Steps

1. Trick 1 : Sequential testing.
2. Trick 2 : Order testing using NMF.
3. Trick 3 : Concatenate X and 1/X.

27

Trick 1

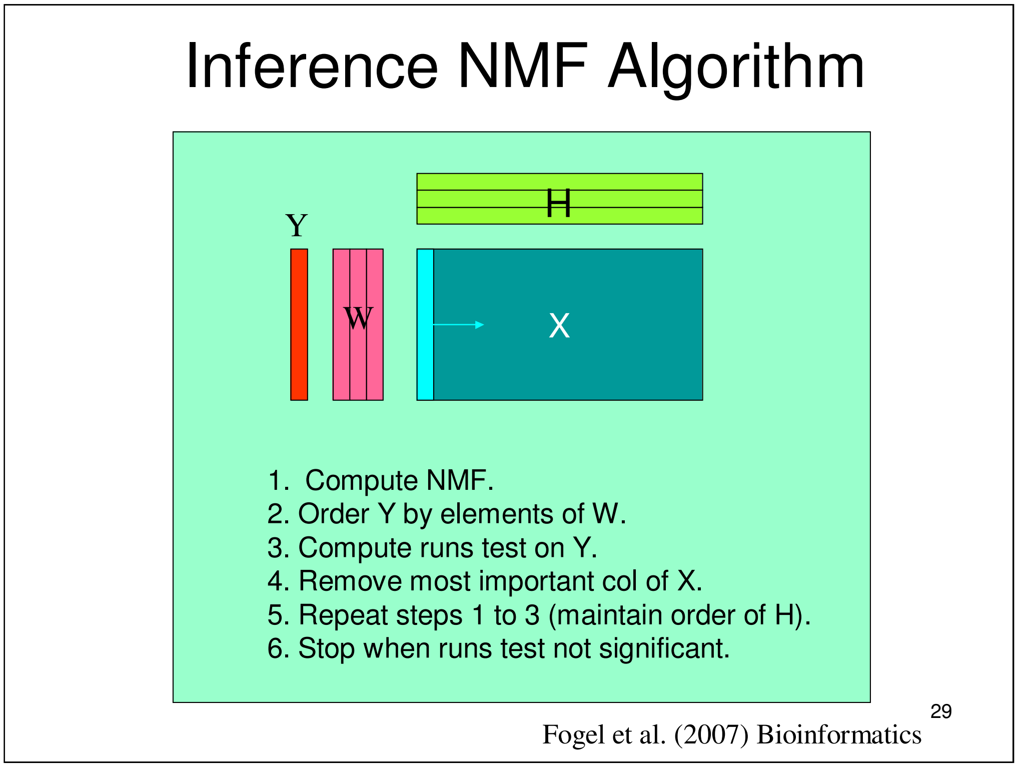
- Tests of means of many variables.
- Variables are ordered according to a ***data-dependent criterion*** and tested in this succession ***without alpha-adjustment until the first non-significant test*.**



Multiple Tests for Different Sets of Variables Using a Data-Driven Ordering of Hypotheses, with an Application to Gene Expression Data

S. Kropf* and J. Läuter

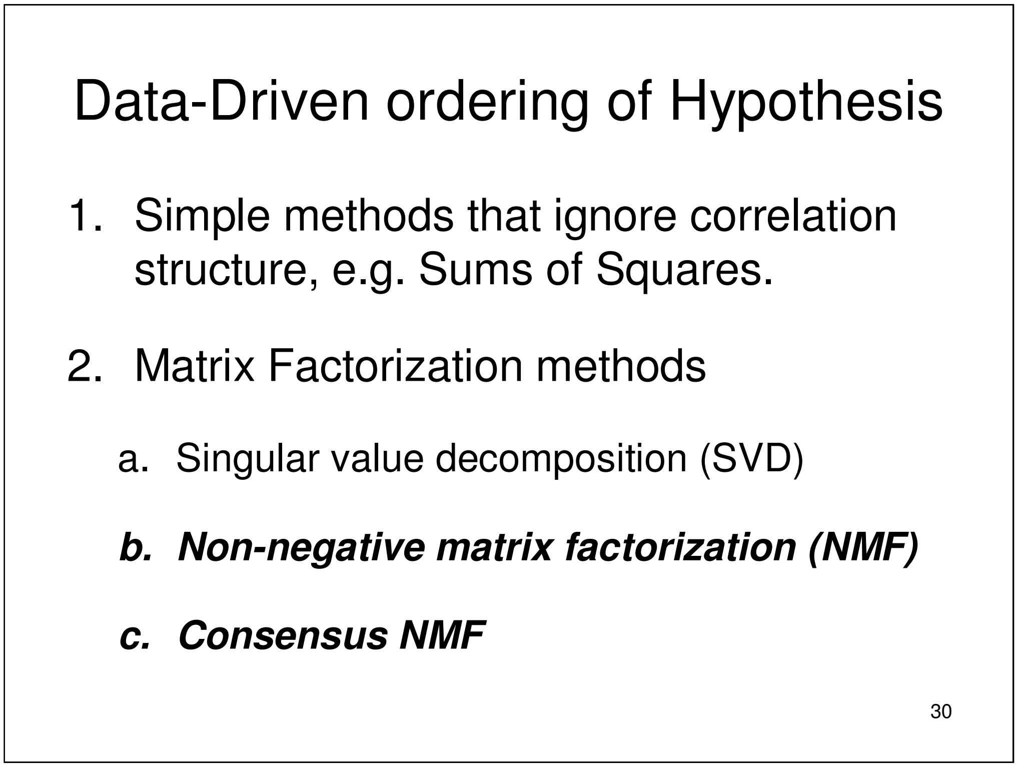
Institute for Biometry and Medical Informatics, Otto von Guericke University Magdeburg, Leipziger Str. 44, 39120 Magdeburg, Germany



Inference NMF Algorithm

1. Compute NMF.
2. Order Y by elements of W.
3. Compute runs test on Y.
4. Remove most important col of X.
5. Repeat steps 1 to 3 (maintain order of H).
6. Stop when runs test not significant.

Fogel et al. (2007) Bioinformatics



Data-Driven ordering of Hypothesis

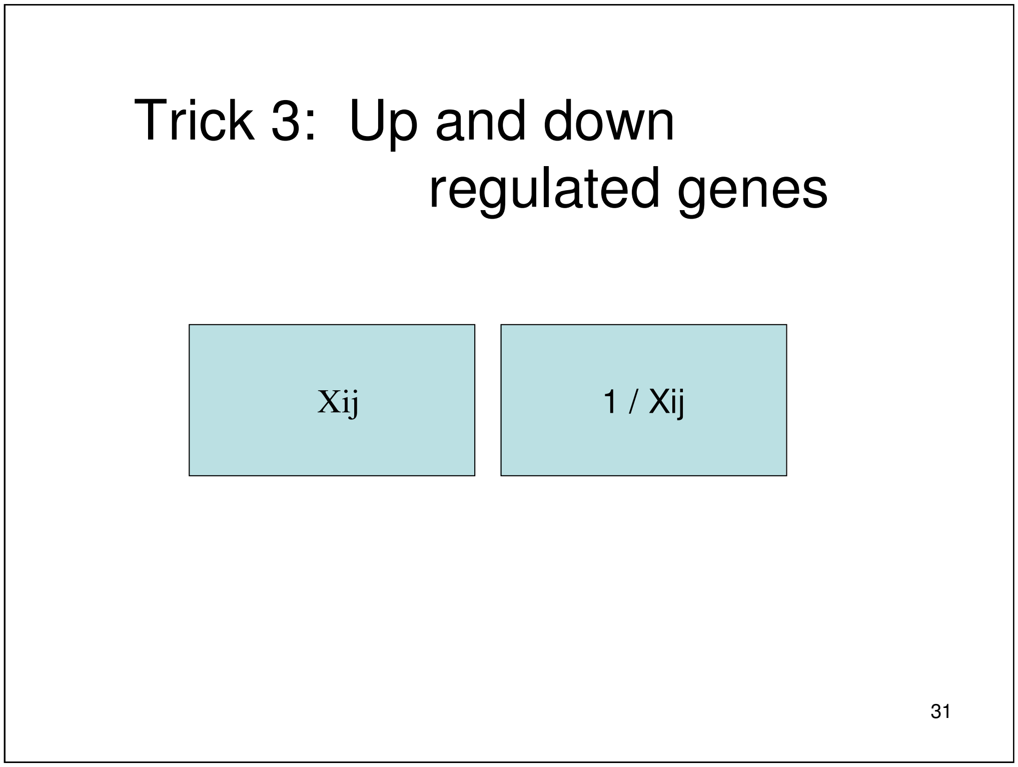
1. Simple methods that ignore correlation structure, e.g. Sums of Squares.

2. Matrix Factorization methods

   a. Singular value decomposition (SVD)

   b. Non-negative matrix factorization (NMF)

   c. Consensus NMF

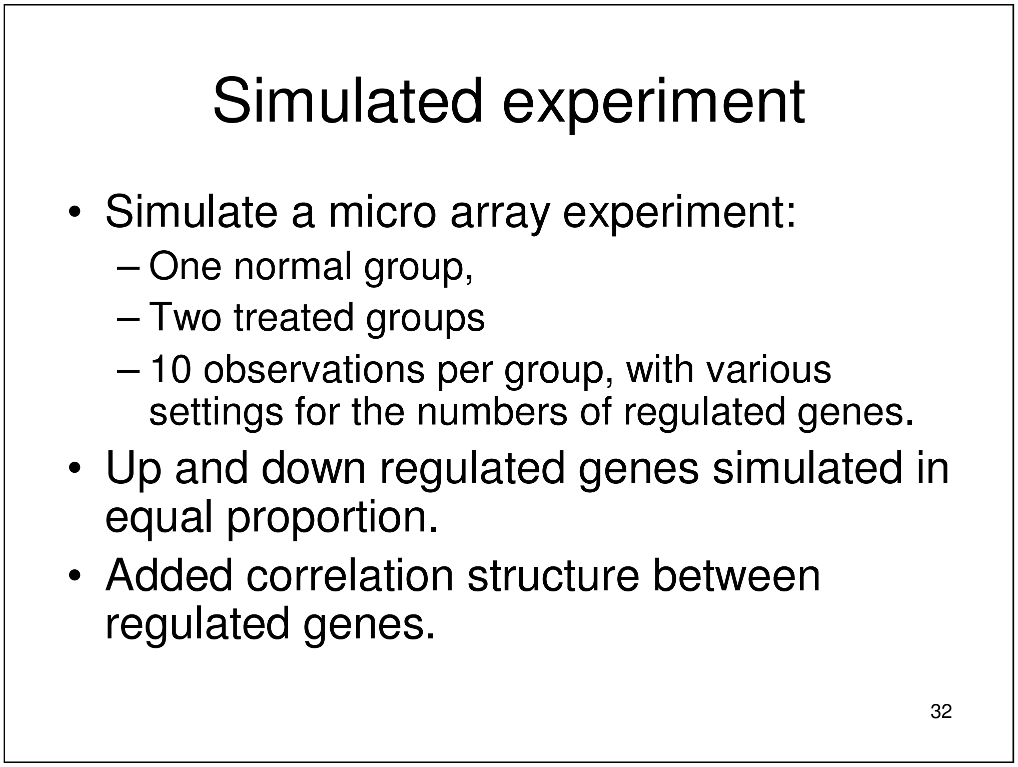


Trick 3: Up and down regulated genes

Xij

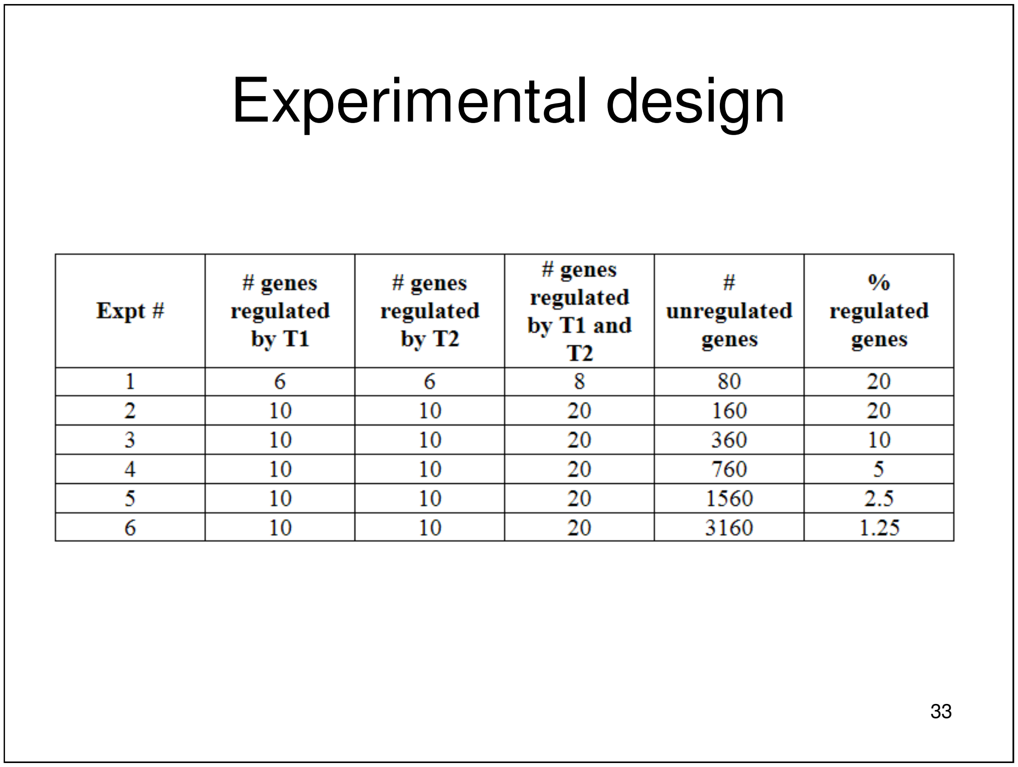
1 / Xij

We come to our third trick. We really want to find both up and down regulated genes. To do this, we append a matrix of element wise inverses to the original data matrix. This is a simple trick, once you see it, but it presents an analysis strategy problem. How do you weight the two matrices?



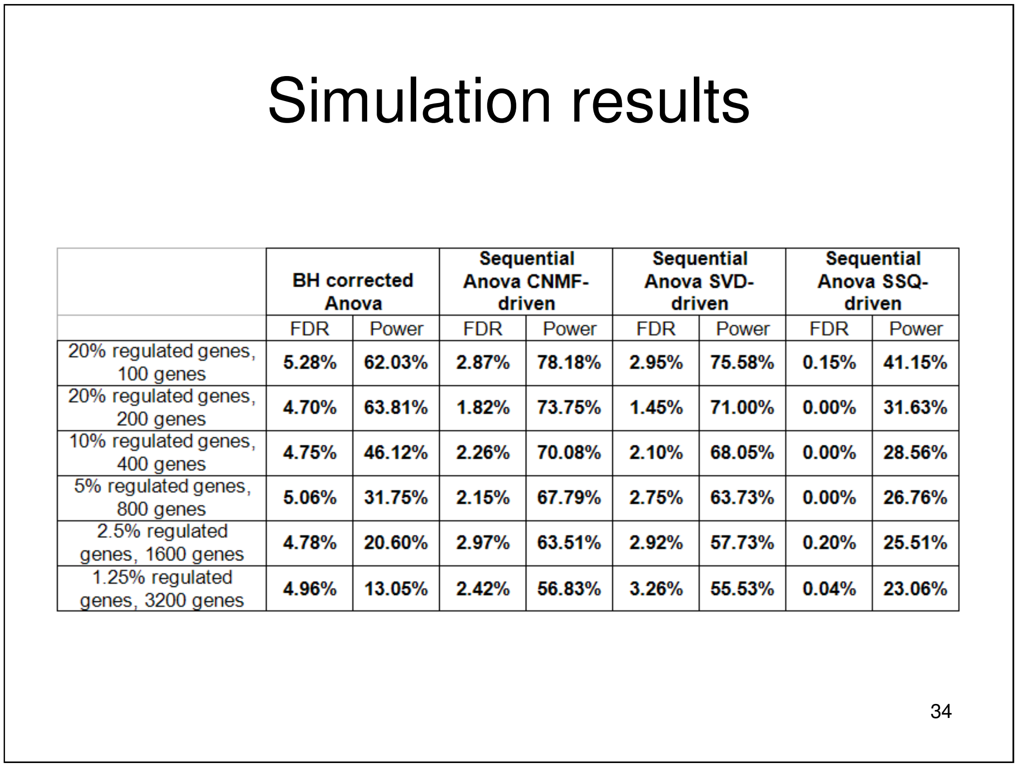
Simulated experiment

- Simulate a micro array experiment:
  - One normal group,
  - Two treated groups
  - 10 observations per group, with various settings for the numbers of regulated genes.
- Up and down regulated genes simulated in equal proportion.
- Added correlation structure between regulated genes.



Experimental design

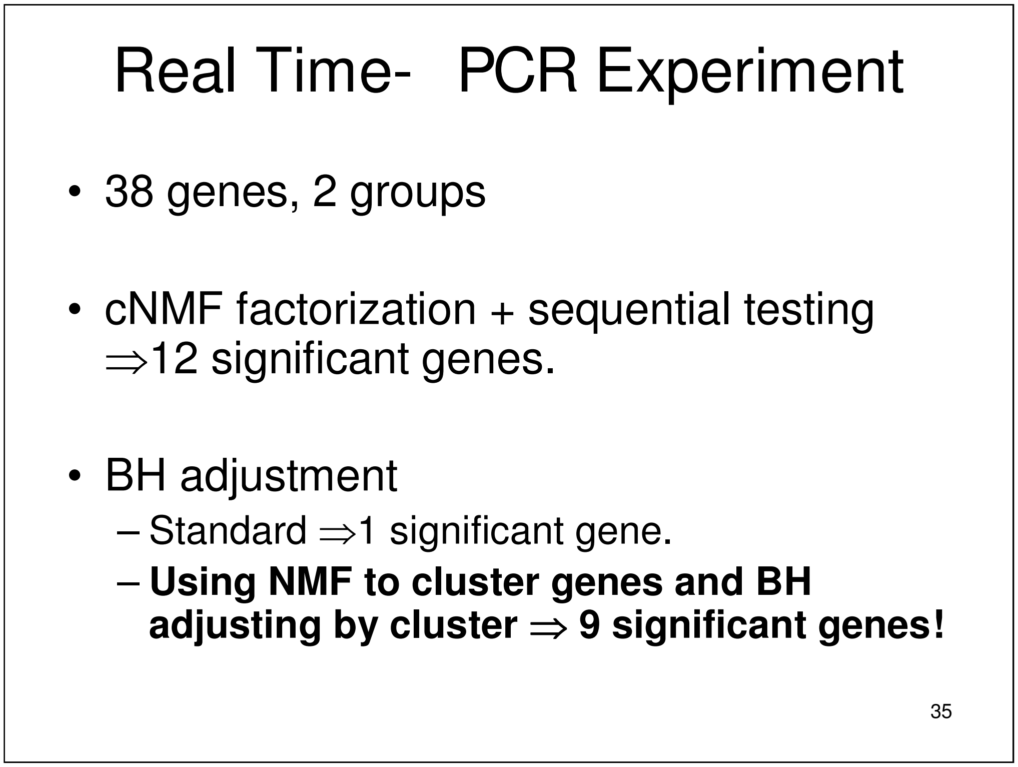
Expt #	# genes regulated by T1	# genes regulated by T2	# genes regulated by T1 and T2	# unregulated genes	% regulated genes
1	6	6	8	80	20
2	10	10	20	160	20
3	10	10	20	360	10
4	10	10	20	760	5
5	10	10	20	1560	2.5
6	10	10	20	3160	1.25



Simulation results

	BH corrected Anova		Sequential Anova CNMF-driven		Sequential Anova SVD-driven		Sequential Anova SSQ-driven	
	FDR	Power	FDR	Power	FDR	Power	FDR	Power
20% regulated genes, 100 genes	**5.28%**	**62.03%**	**2.87%**	**78.18%**	**2.95%**	**75.58%**	**0.15%**	**41.15%**
20% regulated genes, 200 genes	**4.70%**	**63.81%**	**1.82%**	**73.75%**	**1.45%**	**71.00%**	**0.00%**	**31.63%**
10% regulated genes, 400 genes	**4.75%**	**46.12%**	**2.26%**	**70.08%**	**2.10%**	**68.05%**	**0.00%**	**28.56%**
5% regulated genes, 800 genes	**5.06%**	**31.75%**	**2.15%**	**67.79%**	**2.75%**	**63.73%**	**0.00%**	**26.76%**
2.5% regulated genes, 1600 genes	**4.78%**	**20.60%**	**2.97%**	**63.51%**	**2.92%**	**57.73%**	**0.20%**	**25.51%**
1.25% regulated genes, 3200 genes	**4.96%**	**13.05%**	**2.42%**	**56.83%**	**3.26%**	**55.53%**	**0.04%**	**23.06%**

Note that we could order the genes using SVD. The method is good, but cNMF is just a little better. The difference for what it is worth is statistically significant for this simulated data set.



Real Time- PCR Experiment

- 38 genes, 2 groups
- cNMF factorization + sequential testing ⇒12 significant genes.
- BH adjustment
  - Standard ⇒1 significant gene.
  - **Using NMF to cluster genes and BH adjusting by cluster ⇒ 9 significant genes!**

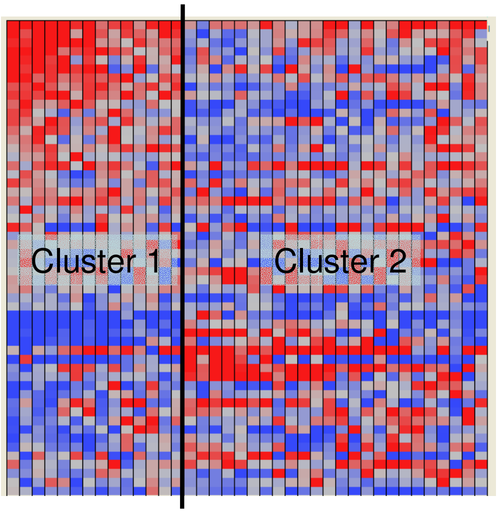
We have use this analysis strategy many times and consider it a success. Here is a real example. Real-time PCR was done with 38 genes and two treatment groups. Obviously the genes were suggested by previous work.

With cNMF 12 genes were significant. Standard BH only declared one significant gene.

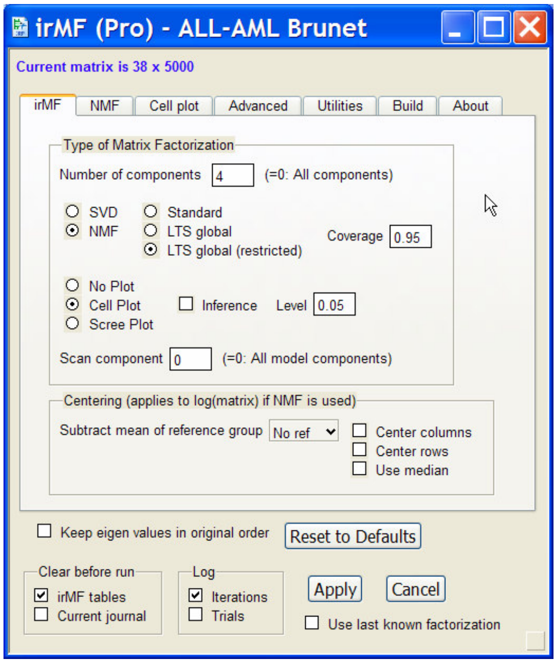
Heatmap view

- All significant genes were clustered into cluster 1.

- Cluster 1 is smaller ⇒ BH adjustment is less conservative.



irMF Dialog



Current matrix is 38 x 5000

Type of Matrix Factorization

Number of components 4 (=0: All components)

○ SVD ○ Standard
● NMF ○ LTS global
● LTS global (restricted) Coverage 0.95

○ No Plot
● Cell Plot □ Inference Level 0.05
○ Scree Plot

Scan component 0 (=0: All model components)

Centering (applies to log(matrix) if NMF is used)

Subtract mean of reference group No ref
□ Center columns
□ Center rows
□ Use median

□ Keep eigen values in original order

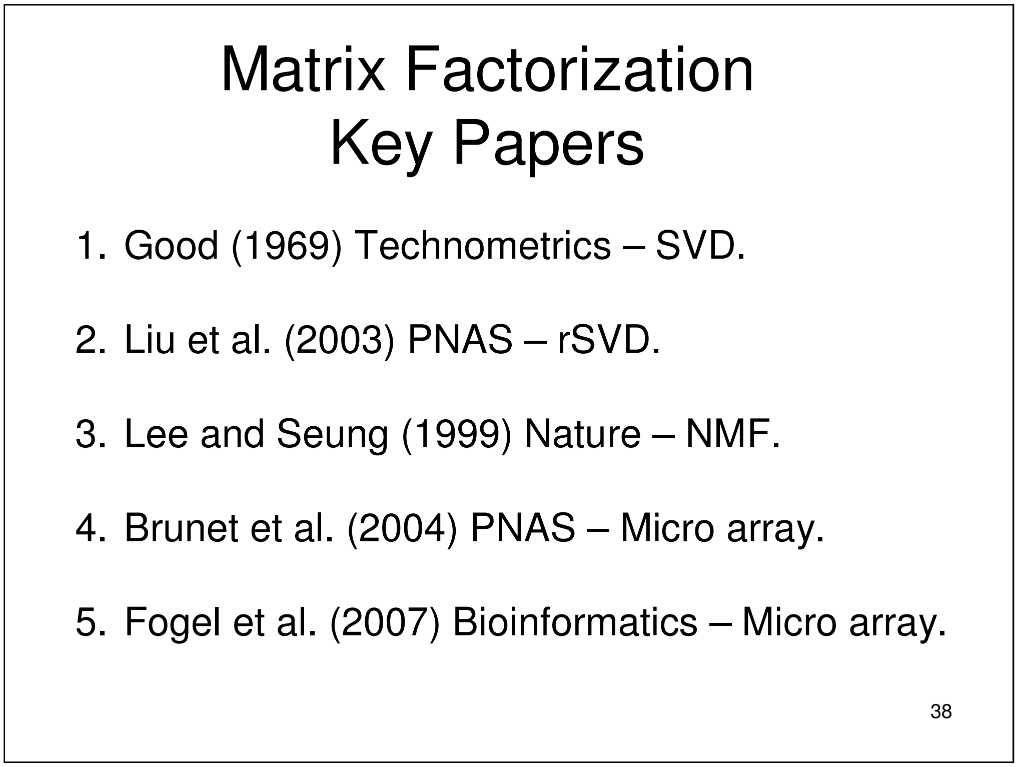
Clear before run
☑ irMF tables
□ Current journal

Log
☑ Iterations
□ Trials

□ Use last known factorization

37

Here is the dialog box for irMF. The user has a lot of control over the analysis process. We are still in the “Arts and Crafts” stage of using NMF for data analysis. As we gain more experience, the analysis should become simpler.



Matrix Factorization Key Papers

1. Good (1969) Technometrics – SVD.
2. Liu et al. (2003) PNAS – rSVD.
3. Lee and Seung (1999) Nature – NMF.
4. Brunet et al. (2004) PNAS – Micro array.
5. Fogel et al. (2007) Bioinformatics – Micro array.

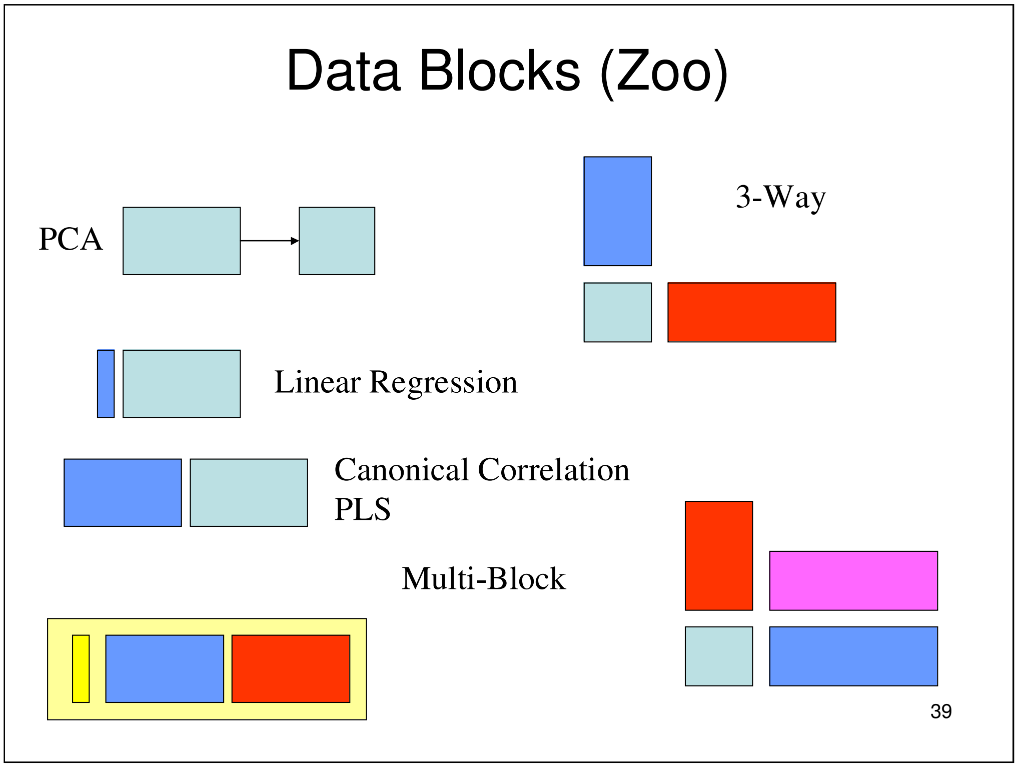
Here are the key papers for matrix factorization.

Most of linear statistics is a special case of SVD.

Liu et al. show how to do SVD for missing data and outliers.

Lee and Seung (ATT) have the most popular paper for NMF.

Kim and Tidor and Brunet et al. have applied NMF to micro array data.



Data Blocks (Zoo)

PCA

3-Way

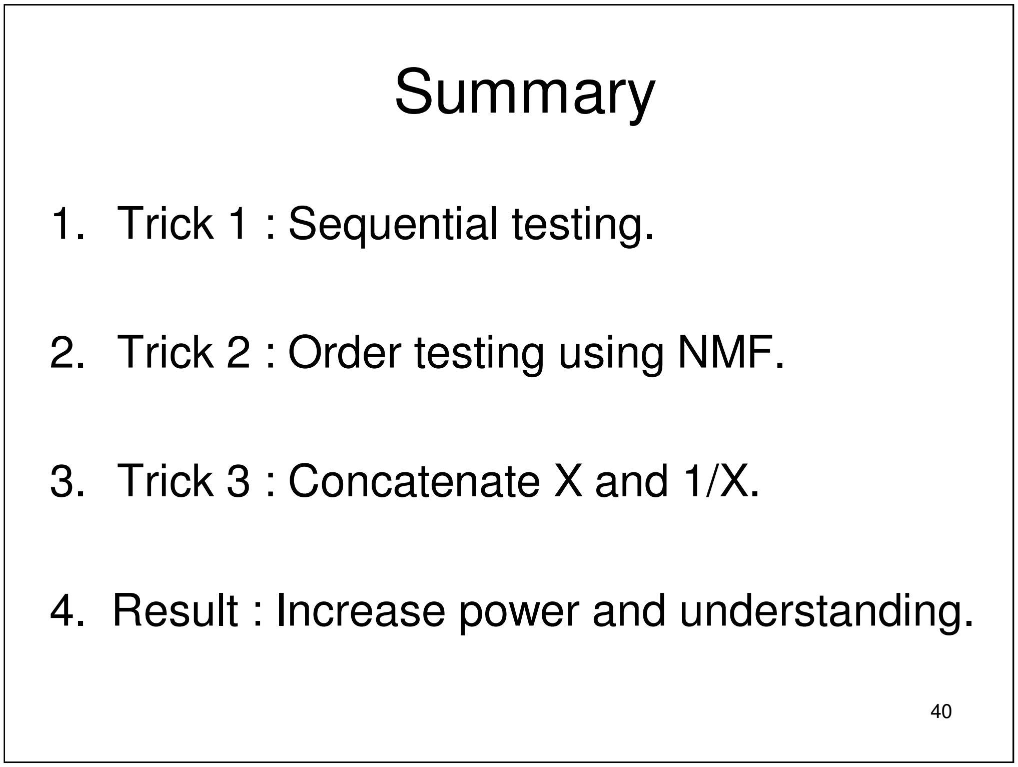
Linear Regression

Canonical Correlation
PLS

Multi-Block

There is a veritable data zoo out there. This talk focuses on two data blocks and the prediction of an outcome.

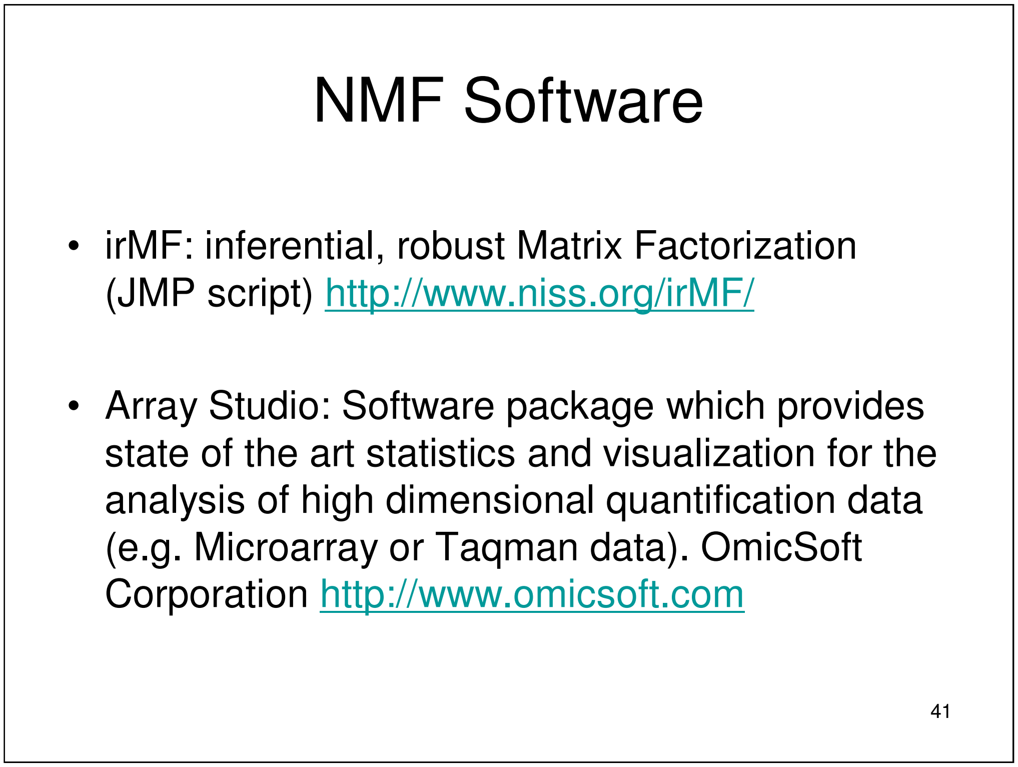
Where ever the data matrices have all positive elements, NMF should be considered.



Summary

1. Trick 1 : Sequential testing.
2. Trick 2 : Order testing using NMF.
3. Trick 3 : Concatenate X and 1/X.
4. Result : Increase power and understanding.

So we combined three tricks and as a result, we have greater statistical power and hopefully an easier interpretation.



NMF Software

- irMF: inferential, robust Matrix Factorization (JMP script) http://www.niss.org/irMF/
- Array Studio: Software package which provides state of the art statistics and visualization for the analysis of high dimensional quantification data (e.g. Microarray or Taqman data). OmicSoft Corporation http://www.omicsoft.com

Free NMF software can be found on the web, notably, BioNMF.

If you are a JMP user, then irMF is an obvious choice.

NMF is included in Array Studio, a sophisticated software system for the analysis of microarray data.