Factor Analysis

Leibny Paola García Perera.
Carnegie Mellon University.
Tecnológico de Monterrey, Campus Monterrey, Mexico
Universidad de Zaragoza, Spain.

Bhiksha Raj, Juan Arturo Nolazco Flores, Eduardo Lleida

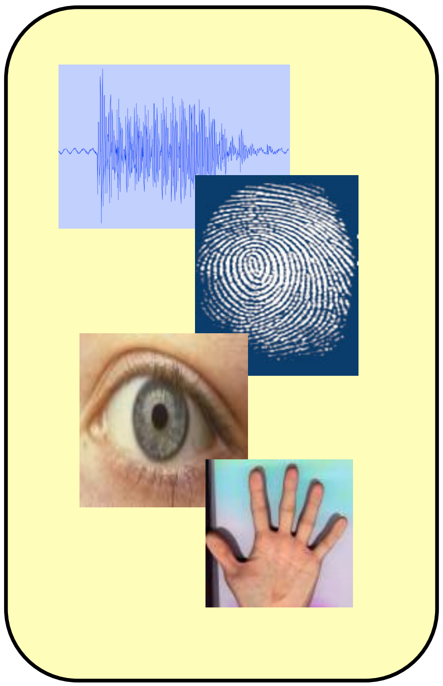
Agenda

- Introduction
  - Motivation:
    - Dimension reduction
    - Modeling: covariance matrix
- Factor Analysis (FA)
    - Geometrical explanation
    - Formulation (The Equations)
    - EM algorithm
    - Comparison with PCA and PPCA.
    - Example with numbers
- Applications
    - Speaker Verification: Joint Factor Analysis (JFA)
    - Some results
- References

Introduction

Problem: Lots of data with n-dimensions vectors.

Example:



$$Y = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1P} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2P} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3P} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & a_{N3} & \cdots & a_{NP} \end{bmatrix} \rightarrow \text{Feature Vectors}$$

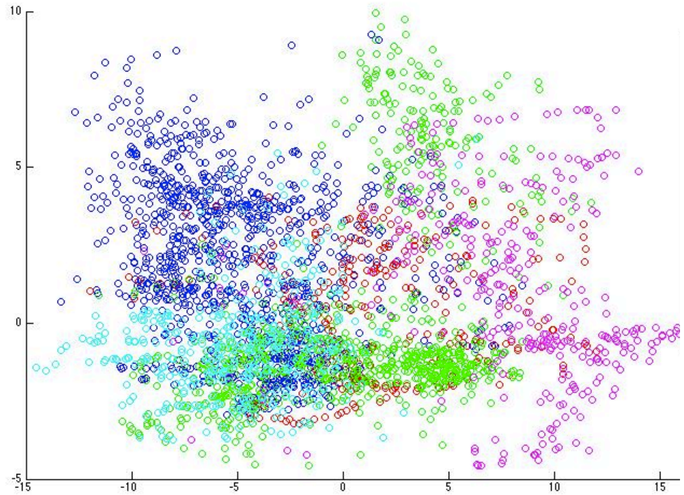
$$P \gg 1$$

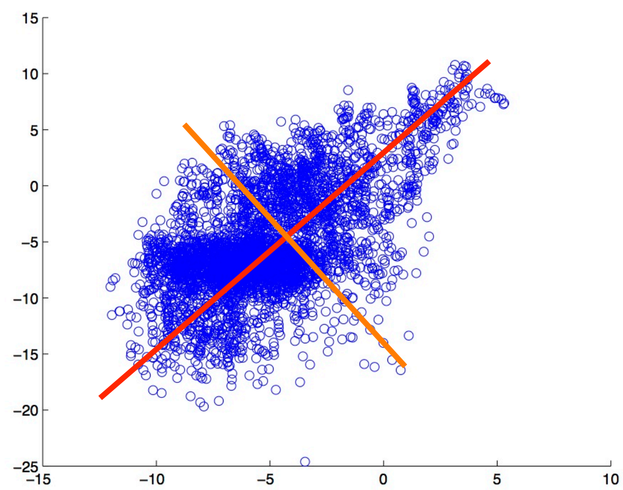
Can we reduce the number of dimensions? To reduce computing time, simplify process?

YES! ☺

Introduction: Covariance matrix

- What can give us information of the data? (Just for this special case)
    - The covariance matrix
        - Get rid of not important information.
        - Think of continuous factors that control the data.

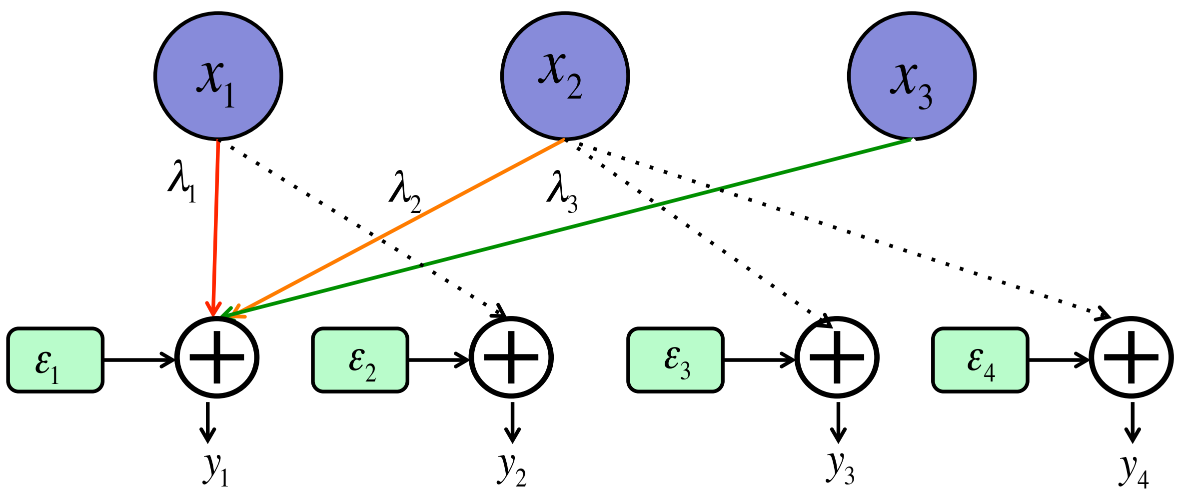




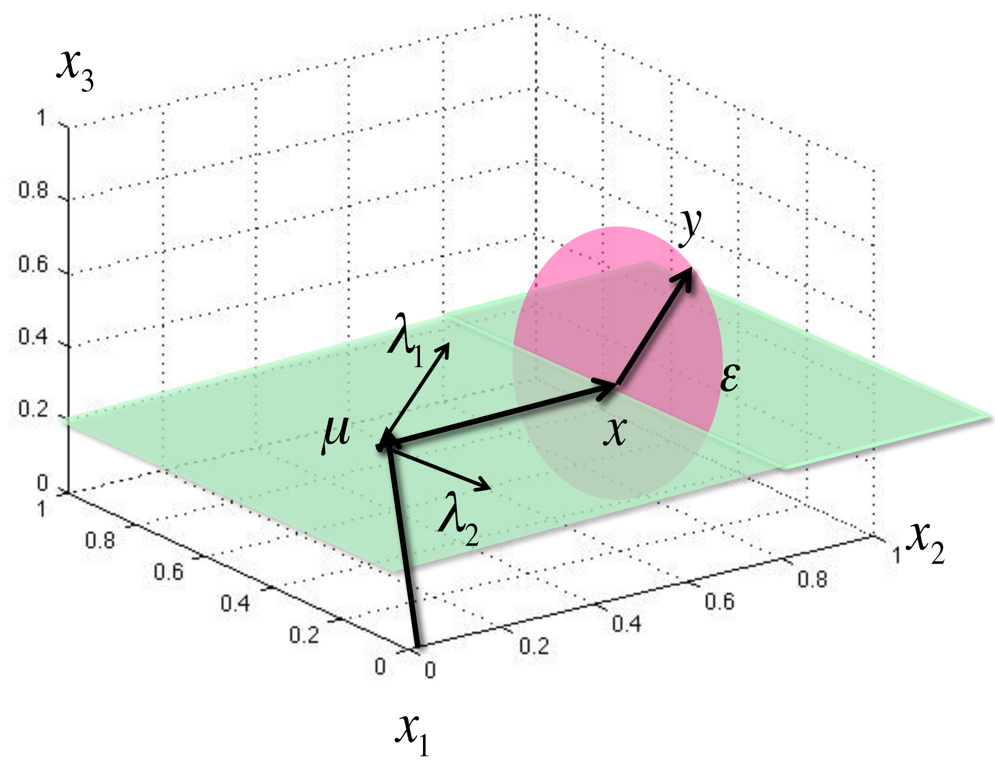
Factor Analysis (FA)

What is Factor Analysis?

- Analysis of the covariance in observed variables (Y).
- In terms of few (latent) common factors.
- Plus a specific error



Factor Analysis (FA): Geometrical Representation



Factor Analysis (FA): Formulation (the equations)

Form	Assumptions

Form

$$y - \mu = \Lambda x + \varepsilon$$

$$y = \Lambda x + \varepsilon$$

$y \rightarrow P \times 1$ data vector

$\mu \rightarrow P \times 1$ mean vector

$\Lambda \rightarrow P \times R$ loading Matrix

$x \rightarrow R \times 1$ factor vector

$\varepsilon \rightarrow P \times 1$ error vector

Assumptions

$$\mathrm{E}(x) = \mathrm{E}(\varepsilon) = 0$$

$$\mathrm{E}(\Lambda\Lambda^T) = \mathrm{I}$$

$$\mathrm{E}(\varepsilon\varepsilon^T) = \psi = \begin{bmatrix} \psi_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \psi_{PP} \end{bmatrix}$$

$$\mathrm{E}(y, x) = \Lambda$$

$$\Sigma = \mathrm{E}(yy^T) = \Lambda\Lambda^T + \Psi$$ Full rank!!

Factor Analysis (FA): Formulation (the equations)

Now that we have checked the matrices dimensions.

The model:

$$p(x) = N(x|0, I)$$

$$p(y|x,\theta) = N(y|\mu + \Lambda x, \Psi)$$

Quick notes:

$$\left.\begin{array}{l} p(x,y) \\ p(y) \\ p(x|y) \end{array}\right\}$$

Are Gaussians!!

Factor Analysis (FA): Formulation (the equations)

Now, we can compute:

$$p\left(y\middle|\theta\right) = \int_x p(x)p\left(y\middle|x,\theta\right)dx = N(y|u, \Lambda\Lambda^T + \Psi)$$

This marginal is… a Gaussian!!

Compute the expected value and covariance.

$$\mathrm{E}(y) = \mathrm{E}(\mu + \Lambda x + \varepsilon) = \mathrm{E}(\mu) + \Lambda\mathrm{E}(x) + \mathrm{E}(\varepsilon) = \mu$$

$$\mathrm{Cov}(y) = \mathrm{E}\left[(y-\mu)(y-\mu)^T\right]$$

$$= \mathrm{E}\left[(\mu + \Lambda x + \varepsilon - \mu)(\mu + \Lambda x + \varepsilon - \mu)^T\right] = \mathrm{E}\left[(\Lambda x + \varepsilon)(\Lambda x + \varepsilon)^T\right]$$

$$= \Lambda\mathrm{E}\left[xx^T\right]\Lambda^T + \mathrm{E}\left[\varepsilon\varepsilon^T\right] = \Lambda\Lambda^T + \Psi$$

Factor Analysis (FA): Formulation (the equations)

So, factor analysis is a constrained covariance Gaussian Model!!

$$p\left(y\middle|\theta\right) = N(y|\mu, \Lambda\Lambda^T + \Psi)$$

So, what is the covariance?

$$\mathrm{cov}(y) = \Lambda \; \Lambda^T + \begin{bmatrix} \psi_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \psi_{PP} \end{bmatrix}$$

Factor Analysis (FA): Formulation (the equations)

How can we compute the likelihood function?

$$\ell(\theta, D) = -\frac{N}{2}\log\left|\Lambda\Lambda^T + \Psi\right| - \frac{1}{2}\sum_n \left(y^n - \mu\right)^T \left(\Lambda\Lambda^T + \Psi\right)^{-1}\left(y^n - \mu\right)$$

$$\ell(\theta, D) = -\frac{N}{2}\log\left|\Sigma\right| - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}\sum_n \left(y^n - \mu\right)\left(y^n - \mu\right)^T\right)$$

$$\ell(\theta, D) = -\frac{N}{2}\log\left|\Sigma\right| - \frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}S\right)$$

S is the sample data covariance Matrix.

Conclusion:

Constrained model close to the Sample covariance!

Factor Analysis (FA): Formulation (the equations)

So we need sufficient statistics…

mean: $$\sum_n y^n$$

covariance: $$\sum_n \left(y^n - \mu\right)\left(y^n - \mu\right)^T$$

Factor Analysis (FA): Expectation Maximization

- How to estimate μ ?
  - Just compute the mean of the data.
- For the rest of the parameters Λ, Ψ ?
  - Expectation Maximization

Factor Analysis (FA): Expectation Maximization

- Advantages
  - Focuses on maximizing the likelihood
- Disadvantages
  - Need to know the distribution
  - No analytical solution

Factor Analysis (FA): Expectation Maximization

Remember EM algorithm?

- E-step:

$$q_n^{t+1} = p\left(x^n \middle| y^n, \theta^t\right)$$

- M-step

$$\theta^{t+1} = \underset{\theta}{\operatorname{argmax}} \sum_n \int_x q_n^{t+1}\left(x^n \middle| y^n\right) \log p\left(y^n, x^n \middle| \theta\right) dx^n$$

Factor Analysis (FA): Expectation Maximization

What do we need?

- E-step:

  Conditional probability!!!

$$q_n^{t+1} = p\left(x^n \middle| y^n, \theta^t\right) = N\left(x^n \middle| m^n, \Sigma^n\right)$$

- M-step:

  Log of the complete data for:

$$\Lambda^{t+1} = \underset{\Lambda}{\operatorname{argmax}} \sum_n \ell\left(x^n, y^n\right)\Big|_{q_n^{t+1}}$$

$$\Psi^{t+1} = \underset{\Psi}{\operatorname{argmax}} \sum_n \ell\left(x^n, y^n\right)\Big|_{q_n^{t+1}}$$

Factor Analysis (FA): Expectation Maximization

What else is needed? $p\left(x|y\right)$

Let's start with:

$$p\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = N\left(\begin{bmatrix} x \\ y \end{bmatrix} | \begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} I & \Lambda^T \\ \Lambda & \Lambda\Lambda^T + \Psi \end{bmatrix}\right)$$

Remember that,

$$\operatorname{cov}\left(x, y\right) = E\left((x-0)(y-u)^T\right) = E\left(x\left(\mu + \Lambda x + \varepsilon - u\right)^T\right)$$

$$= E\left(x\left(\Lambda x + \varepsilon\right)^T\right) = \Lambda^T$$

Factor Analysis (FA): Expectation Maximization

Now,

$$p(x|y) = N(x|m,V)$$

$$m = \Lambda\left(\Lambda\Lambda^T + \Psi\right)^{-1}(y-u)$$

$$V = I - \Lambda^T\left(\Lambda\Lambda^T + \Psi\right)^{-1}\Lambda$$

Remembering Gaussian conditioning formulas

$$p\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} \mid \begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$$

$$p(\mathbf{x}_1) = \mathcal{N}(\mu_1, \Sigma_{11})$$

$$p(\mathbf{x}_1|\mathbf{x}_2) = \mathcal{N}(\mathbf{x}_1|\mathbf{m}_{1|2}, \mathbf{V}_{1|2})$$

$$\mathbf{m}_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{x}_2 - \mu_2)$$

$$\mathbf{V}_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

Remember inversion lemma?

$$\Sigma^{-1} = \left(\Lambda\Lambda^T + \Psi\right)^{-1} = \Psi^{-1} + \Psi^{-1}\Lambda\left(I + \Lambda^T\Psi^{-1}\Lambda\right)^{-1}\Lambda^T\Psi^{-1}$$

Inverting this matrix is much more efficient O(MP) instead of O(P^2), thanks to the lemma.

Factor Analysis (FA): Expectation Maximization

We finally obtain:

$$p(x|y) = N(x|m, V)$$

$$V = \left(I - \Lambda^T \Psi^{-1} \Lambda\right)^{-1}$$

$$m = V \Lambda^T \Psi^{-1} (y - u)$$

Factor Analysis (FA): Expectation Maximization

Some nice observations:

$$p(x|y) = N(x|m, V)$$

$$V = \left(I - \Lambda^T \Psi^{-1} \Lambda\right)^{-1}$$

$$m = V\Lambda^T \Psi^{-1}(y - u)$$

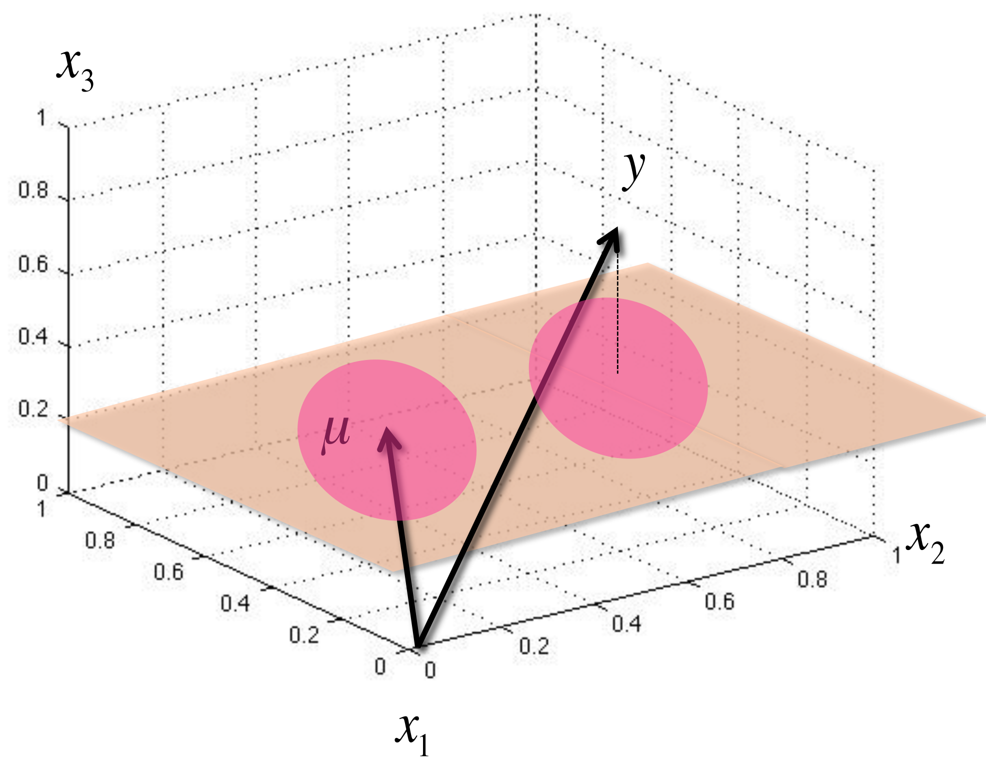
Means that the posterior mean is just a linear operation!!!

And the covariance does not depend on the observed data!!!

$$V = \left(I - \Lambda^T \Psi^{-1} \Lambda\right)^{-1}$$

Factor Analysis (FA): Expectation Maximization

How does it look?



Factor Analysis (FA): Expectation Maximization

Let's subtract the mean for our computation.

The likelihood for the complete data is:

$$\ell(\Lambda,\Psi) = \sum_n \log p\left(x^n, y^n\right)$$

$$\ell(\Lambda,\Psi) = \sum_n \log p\left(x^n\right) + \log p\left(x^n \middle| y^n\right)$$

$$\ell(\Lambda,\Psi) = -\frac{N}{2}\log|\Psi| - \frac{1}{2}\sum_n x^T x - \frac{1}{2}\sum_n \left(y^n - \Lambda x^n\right)^T \Psi^{-1}\left(y^n - \Lambda x^n\right)$$

$$\ell(\Lambda,\Psi) = -\frac{N}{2}\log|\Psi| - \frac{N}{2} tr(S\Psi^{-1})$$

$$S = \frac{1}{N}\sum_n \left(y^n - \Lambda x^n\right)^T \left(y^n - \Lambda x^n\right)$$

Factor Analysis (FA): Expectation Maximization

Now, let's compute the M step! (Almost there!)

We need to calculate the derivatives of the log likelihood

$$\frac{\partial \ell(\Lambda,\Psi)}{\partial \Lambda} = -\Psi^{-1}\sum_n y_n x_n^T + \Psi^{-1}\Lambda\sum_n x_n x_n^T$$

$$\frac{\partial \ell(\Lambda,\Psi)}{\partial \Psi^{-1}} = \frac{N\Psi}{2} - \frac{NS}{2}$$

And the expectations with respect to q^t

$$E\left[\ell'_\Lambda\right] = -\Psi^{-1}\sum_n y_n m_n^T + \Psi^{-1}\Lambda\sum_n V_n$$

$$E\left[\ell'_{\Psi^{-1}}\right] = \frac{N\Psi}{2} - \frac{N\cdot E[S]}{2}$$

Factor Analysis (FA): Expectation Maximization

Finally, set the derivatives to zero and solve!

$$\Lambda^{t+1} = \left(\sum_n y^n m^{nT}\right)\left(\sum_n V^n\right)^{-1}$$

$$\Psi^{t+1} = \frac{1}{N} diag\left(\sum_n y^n y^{nT} + \Lambda^{t+1} \sum_n m^n y^{nT}\right)$$

Factor Analysis (FA): Expectation Maximization

What are the final equations?

① $\mu \rightarrow$ Sample mean (Subtract the mean from data).

② E-step

$$q_n^{t+1} = p\left(x^n \middle| y^n, \theta^t\right) = N\left(x^n \middle| m^n, V^n\right)$$

$$V^n = \left(I - \Lambda^T \Psi^{-1} \Lambda\right)^{-1}$$

$$m^n = V^n \Lambda^T \Psi^{-1} \left(y - u\right)$$

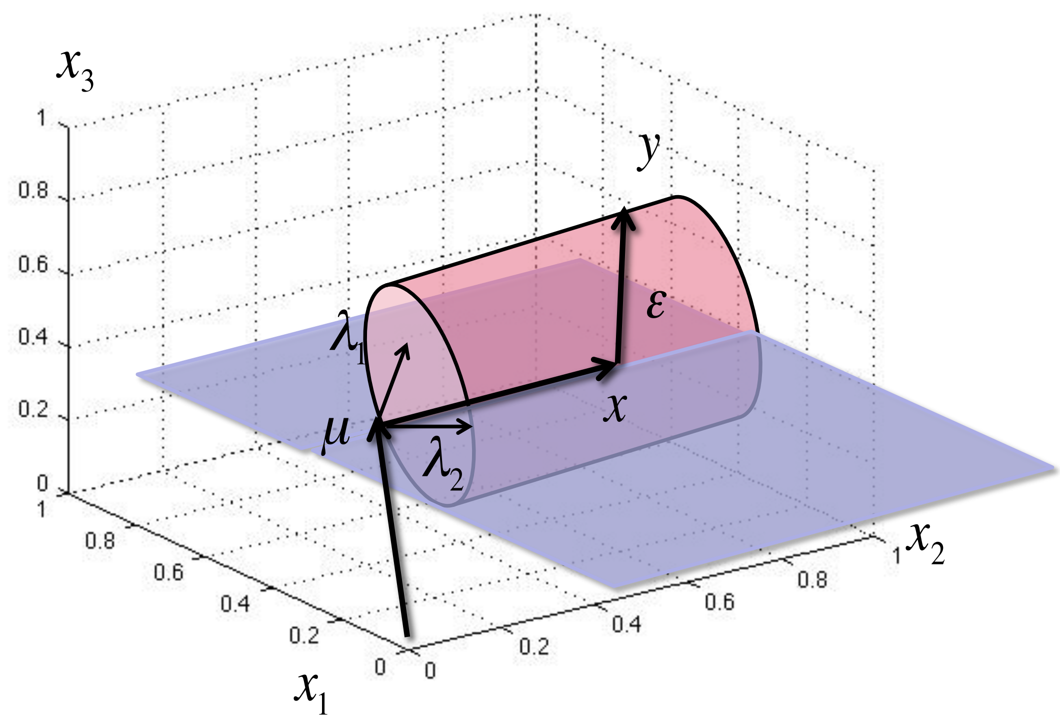
③ M-step

$$\Lambda^{t+1} = \left(\sum_n y^n m^{nT}\right)\left(\sum_n V^n\right)^{-1}$$

$$\Psi^{t+1} = \frac{1}{N} diag\left(\sum_n y^n y^{nT} + \Lambda^{t+1} \sum_n m^n y^{nT}\right)$$

Factor Analysis (FA): Geometrical Representation

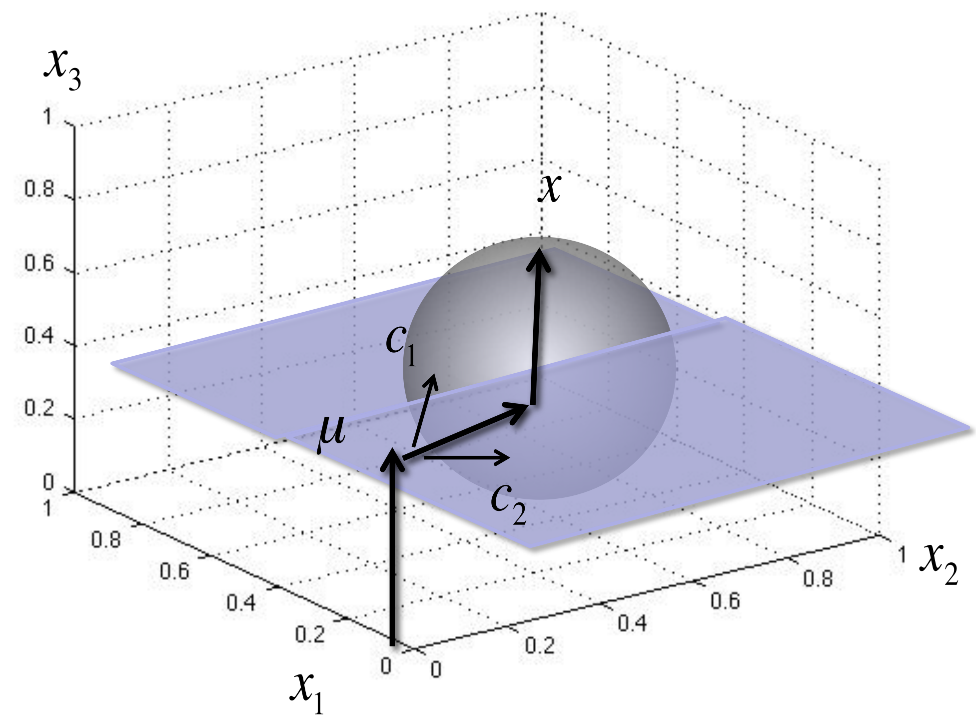
How does FA really look like?



Factor Analysis (FA): Comparison

What is PPCA? Just a quick intuition.

$p(x) = N(x|0, I)$ $\qquad$ $p(y|x,\theta) = N(y|u + \Lambda x, \sigma^2 I)$

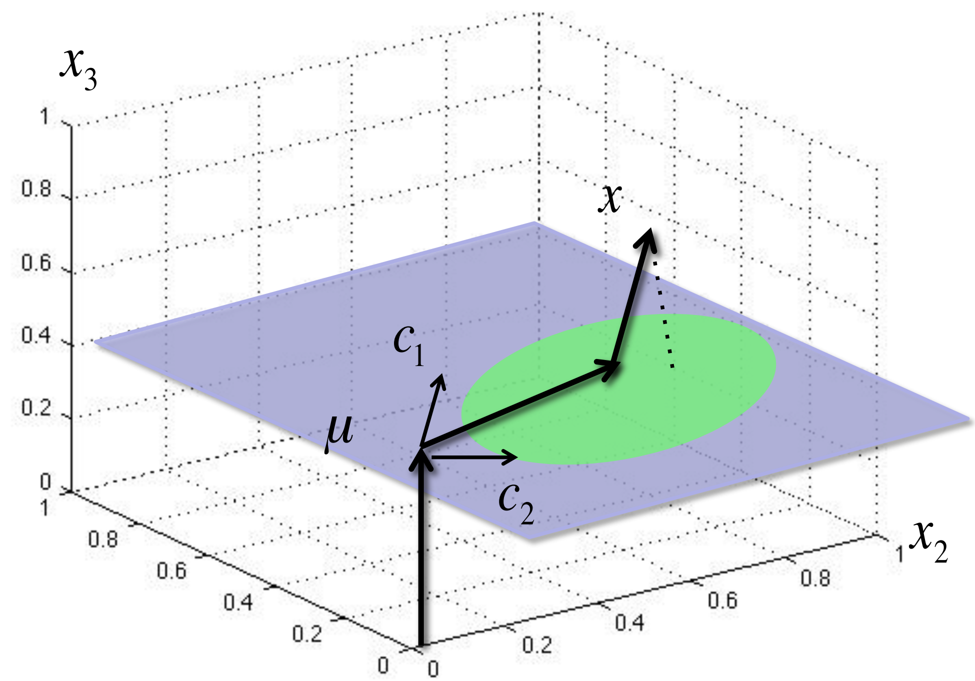


Nice, isn't it? ☺

Factor Analysis (FA): Comparison

What about PCA? Just a quick intuition.

$$p(x) = N(x|0,I) \qquad p\left(y\middle|x,\theta\right) = N(y|u+\Lambda x,0) \qquad \sigma^2 \rightarrow 0$$

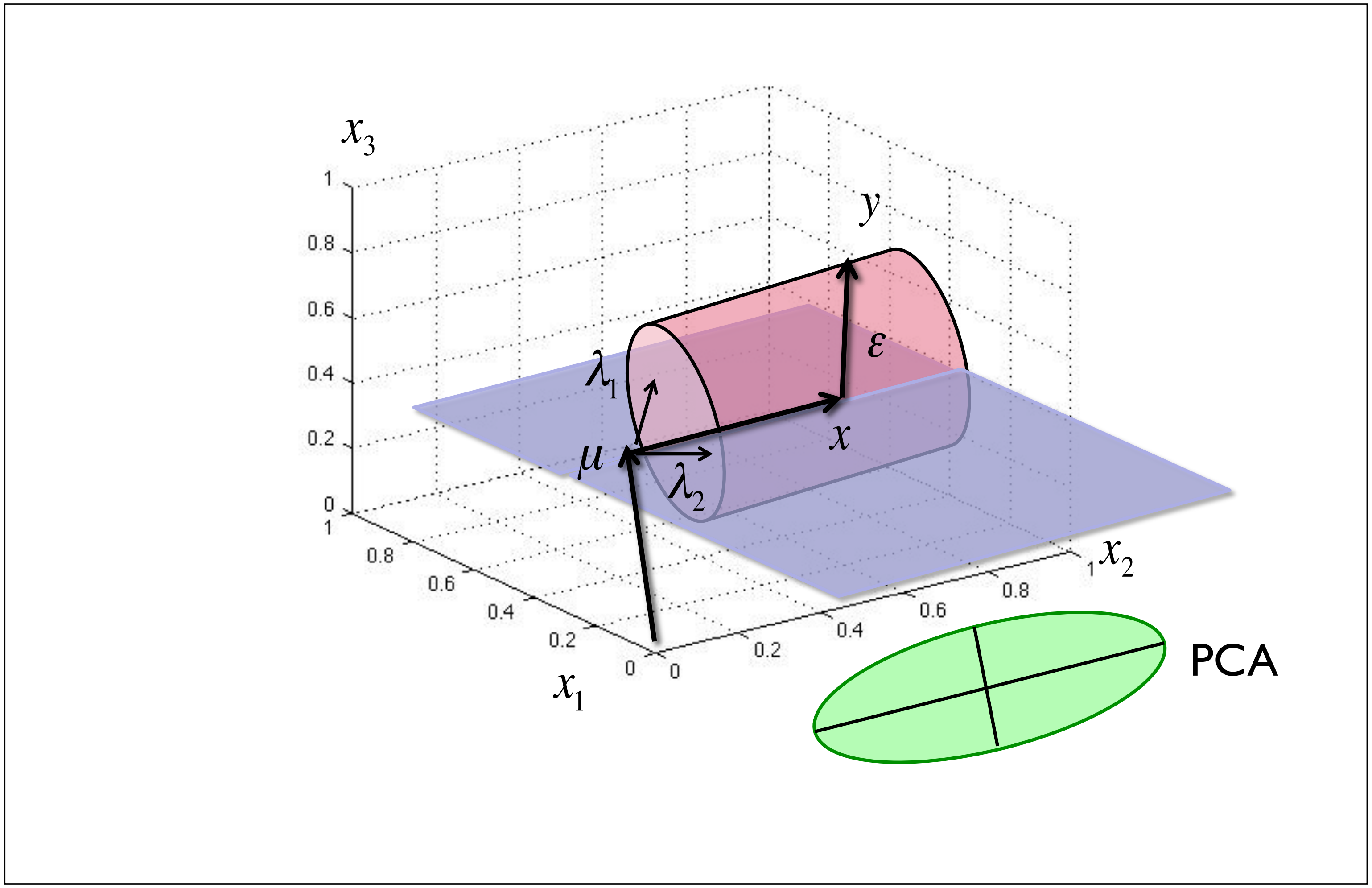


Nice! ☺

Factor Analysis (FA): Comparison

- Final notes:
  - FA is invariant if we change the scale.
  - FA looks for correlation of the data.
  - PCA is invariant if we rotate the data.
  - PCA looks for direction of large variance.

Factor Analysis (FA): Comparison



Factor Analysis (FA): Final notes

- More final notes:
  - Remember our initial goal?
    - Reduce dimensions

$$y = \Lambda x + \varepsilon$$

We can decide the value of R and compute a new set of features!!

$y \rightarrow P \times 1$ data vector

$\Lambda \rightarrow P \times R$ Loading Matrix

$x \rightarrow R \times 1$ factor vector

$\varepsilon \rightarrow P \times 1$ error vector

- Produce a suitable model to explain the data, based on constrained covariance Gaussian.

$$p(y|\theta) = N(y|\mu, \Lambda\Lambda^T + \Psi)$$

Factor Analysis (FA): The real algorithm

- Initialization
  - Give statistics a start value
- While (stop criteria)
  - Compute sufficient statistics and Expectation

    get V^n

    get m^n

  - Update the statistics (Maximization)

    update Λ

    update Ψ

Factor Analysis (FA): A practical example

```
Y = [

    2.5225   -1.6369   -3.6994   -5.7542   -2.4632
    3.8143    3.9840    3.3812   -4.5673   -1.9867
    1.8606    2.6580    1.0446   -9.2575   -1.1736
    0.6135    2.5380   -3.2632    0.1344   -1.4441
    2.1523    3.1987   14.3550   -8.3578   -1.8787
    1.3377    2.6883   -4.7846   15.0349   -2.5611 ]

ybar=mean(Y)

ybar =

    2.0501    2.2383    1.1723   -2.1279   -1.9179

S=cov(Y)
S =

    1.1907    0.1033    2.7165   -4.1557   -0.1477
    0.1033    3.8911    6.2666    1.8439    0.4391
    2.7165    6.2666   51.5143  -36.2420    0.9312
   -4.1557    1.8439  -36.2420   81.6850   -2.6748
   -0.1477    0.4391    0.9312   -2.6748    0.2992
```

Factor Analysis (FA): A practical example

Initialization

```
Psi=Psi0 % PCA obtained
V=V0 % PCA Obtained

Psi =

    0.9541
    2.1716
    0.1026
    0.0289
    0.2156


V =

   -0.4862   -0.0082
   -0.1978    1.2963
   -5.7194    4.3244
    8.5523    2.9180
   -0.2664   -0.1121
```

Factor Analysis (FA): A practical example

Sufficient Statistics

```
mu= ybar;
Psi=Psi0 % PCA obtained
V=V0 % PCA Obtained

Psi =

    0.9541
    2.1716
    0.1026
    0.0289
    0.2156


V =

   -0.4862   -0.0082
   -0.1978    1.2963
   -5.7194    4.3244
    8.5523    2.9180
   -0.2664   -0.1121
```

```
B=(V'*V)\V';%LSE

%expectation Y
for i=1:n
X(i,:)= B*((Y(i,:)-mu)') ;
end

X =

   -0.0232   -1.2666
   -0.3266    0.1622
   -0.5691   -0.7228
    0.4259   -0.4231
   -1.2140    1.3861
    1.7070    0.8642

Xbar=mean(X);
%conditional covariance
L=I+V'*IPsi*V;
Covx=eye(m)/L

Covx =

    0.0215   -0.0076
   -0.0076    0.0290
```

Factor Analysis (FA): A practical example

Deltas:

```
mu= ybar;
Psi=Psi0 % PCA obtained
V=V0 % PCA Obtained

Psi =

    0.9541
    2.1716
    0.1026
    0.0289
    0.2156


V =

   -0.4862   -0.0082
   -0.1978    1.2963
   -5.7194    4.3244
    8.5523    2.9180
   -0.2664   -0.1121
```

```
B=(V'*V)\V';%LSE

 %expectation X
   for i=1:n
       X(i,:)= B*((Y(i,:)-mu)') ;
   end

X =

   -0.0232   -1.2666
   -0.3266    0.1622
   -0.5691   -0.7228
    0.4259   -0.4231
   -1.2140    1.3861
    1.7070    0.8642

xbar=mean(x);
%conditional covariance
L=I+V'*IPsi*V;
Covx=eye(m)/L

Covx =

    0.0215   -0.0076
   -0.0076    0.0290
```

```
Dy= Y- ones(n,1)*mu
Dx= X- repmat(ybar,n,1);
```

Factor Analysis (FA): A practical example

```
mu= ybar;
Psi=Psi0 % PCA obtained
V=V0 % PCA Obtained

Psi =

    0.9541
    2.1716
    0.1026
    0.0289
    0.2156


V =

   -0.4862   -0.0082
   -0.1978    1.2963
   -5.7194    4.3244
    8.5523    2.9180
   -0.2664   -0.1121
```

```
B=(V'*V)\V';%LSE

 %expectation X
   for i=1:n
       X(i,:)= B*((Y(i,:)-mu)') ;
   end

X =

   -0.0232   -1.2666
   -0.3266    0.1622
   -0.5691   -0.7228
    0.4259   -0.4231
   -1.2140    1.3861
    1.7070    0.8642

xbar=mean(X);
%conditional covariance
L=I+V'*IPsi*V;
Covx=eye(m)/L

Covx =

    0.0215   -0.0076
   -0.0076    0.0290
```

```
Dy= Y- ones(n,1)*mu
Dx= X- repmat(xbar,n,1);

%maximize V/update
    V= (Dy'*Dx)/(Covx+(Dy'*Dx))

V =

    0.4858   -0.0088
    0.1960    1.2885
    5.7083    4.2920
   -8.5476    2.9118
    0.2663   -0.1118

%update mu
 mu=mean(Y- X*V');

% update Psi.
Psi=  (1/n)* diag((Dy'*Dy) - (Dy'*Dx)*V' )

Psi =

    0.7951
    1.8106
    0.1025
    0.0290
    0.1797
```

Factor Analysis (FA): A practical example

```
mu= ybar;
Psi=Psi0 % PCA obtained
V=V0 % PCA Obtained

Psi =

0.9541
2.1716
0.1026
0.0289
0.2156

V =

 0.4858   -0.0088
 0.1960    1.2885
 5.7083    4.2920
-8.5476    2.9118
 0.2663   -0.1118
```

```
B=(V'*V)\V';%LSE

%expectation X
for i=1:n
X(i,:)= B*((Y(i,:)-mu)') ;
end

X =

-0.0232   -1.2666
-0.3266    0.1622
-0.5691   -0.7228
 0.4259   -0.4231
-1.2140    1.3861
 1.7070    0.8642

xbar=mean(X);
%conditional covariance
L=I+V'*IPsi*V;
Covx=eye(m)/L

Covx =

 0.0215   -0.0076
-0.0076    0.0290
```

```
Dy= Y- ones(n,1)*mu
Dx= X- repmat(xbar,n,1);

%maximize V/update
V= (Dy'*Dx)/(Covx+(Dx'*Dx))

V =

 0.4858   -0.0088
 0.1960    1.2885
 5.7083    4.2920
-8.5476    2.9118
 0.2663   -0.1118

%update mu
mu=mean(Y- X*V');

% update Psi.
Psi=  (1/n)* diag((Dy'*Dy) - (Dy'*Dx)*V' )

Psi =

0.7951
1.8106
0.1025
0.0290
0.1797
```

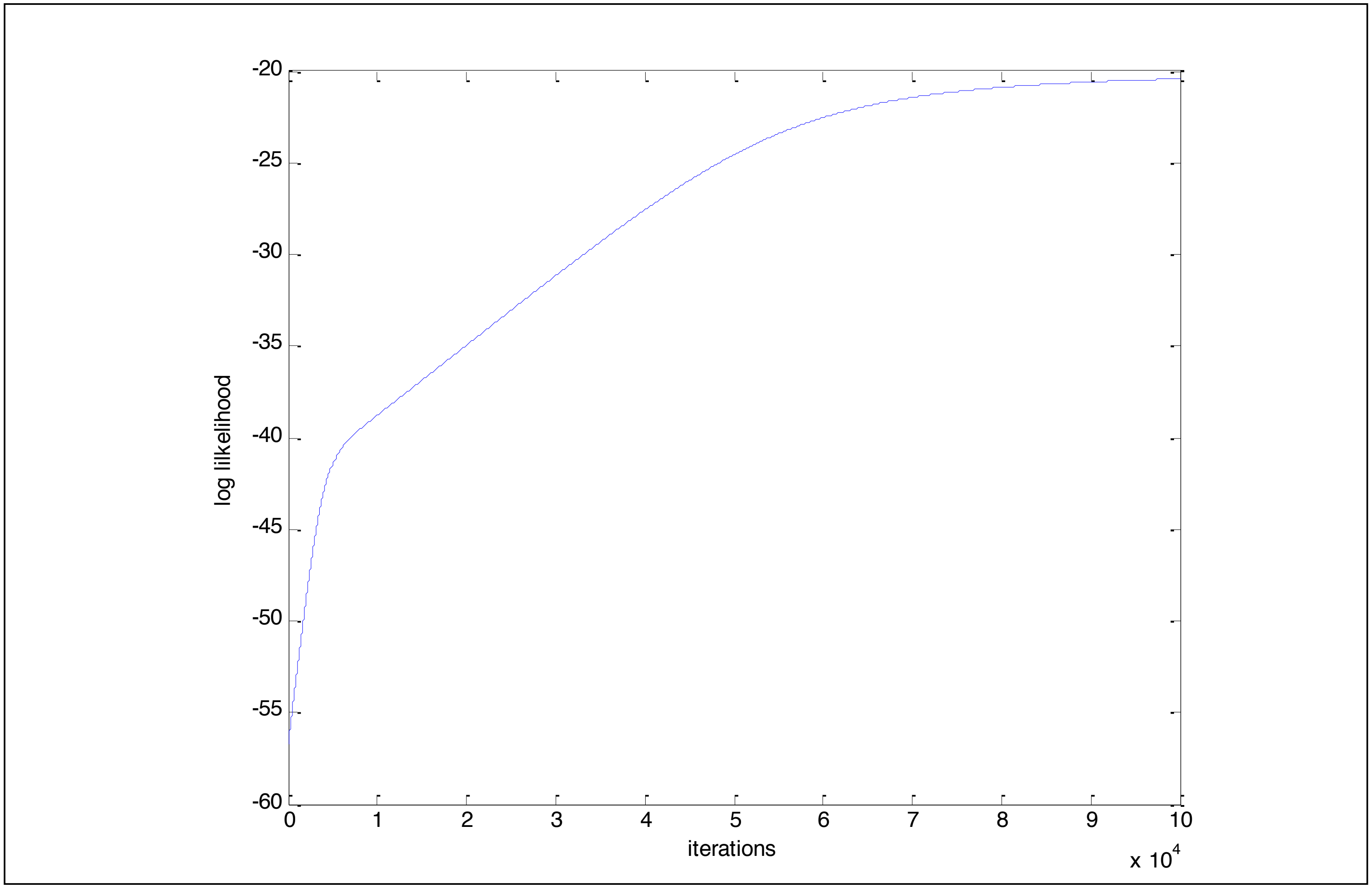
Factor Analysis (FA): A practical example

```
mu= ybar;
Psi=Psi0 % PCA obtained
V=V0 % PCA Obtained

Psi =

0.7951
1.8106
0.1025
0.0290
0.1797

V =

 0.4858   -0.0088
 0.1960    1.2885
 5.7083    4.2920
-8.5476    2.9118
 0.2663   -0.1118
```

```
B=(V'*V)\V';%LSE

%expectation X
for i=1:n
X(i,:)= B*((Y(i,:)-
mu)') ;
end

X =

-0.0232   -1.2666
-0.3266    0.1622
-0.5691   -0.7228
 0.4259   -0.4231
-1.2140    1.3861
 1.7070    0.8642

xbar=mean(X);
%conditional covariance
L=I+V'*IPsi*V;
Covx=eye(m)/L

Covx =

 0.0215   -0.0076
-0.0076    0.0290
```

```
Dy= Y- ones(n,1)*mu
Dx= X- repmat(xbar,n,1);

%maximize V/update
V= (Dy'*Dx)/(Covx+(Dx'*Dx))

V =

 0.4858   -0.0088
 0.1960    1.2885
 5.7083    4.2920
-8.5476    2.9118
 0.2663   -0.1118

%update mu
mu=mean(Y- X*V');

% update Psi.
Psi=  (1/n)* diag((Dy'*Dy) - (Dy'*Dx)*V' )

Psi =

0.7951
1.8106
0.1025
0.0290
0.1797
```

Factor Analysis (FA):
A practical example



Speaker Verification System

Speaker Verification

Speaker Verification: is a detection problem. Accepts or rejects a user as legitimate based on his speech signal.

- Input:
    - Speech signal X
    - Claimed identity i
- Output:
    - accept/reject

$$d = \begin{cases} \text{accept} & \phi(X,i) > \tau_i; \\ \text{reject} & \text{otherwise} \end{cases}$$

- A confidence measure $\phi(X,i)$

Speaker Verification

- Each speaker has its own model, known as target model λ_i
- And its antimodel $\overline{\lambda_i}$
- The target model is the prototype of each speaker in the training.
- The antimodel is the impostor's prototype.
- When all the impostors share the same model, the final model is called: UBM Universal Background Model.

Speaker Verification

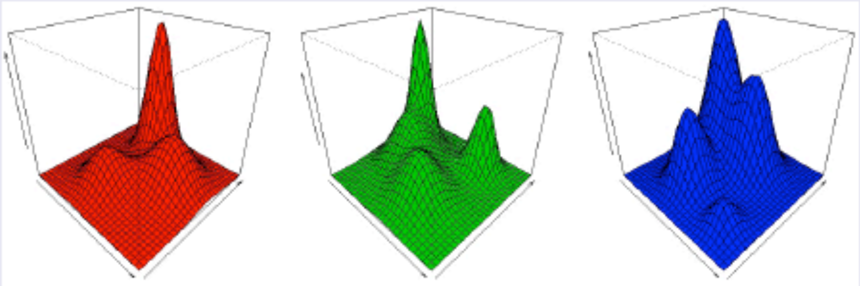
This is how we represent each speaker:

Gaussian Mixture Model (GMM)

- Let X be the acoustic feature vector.

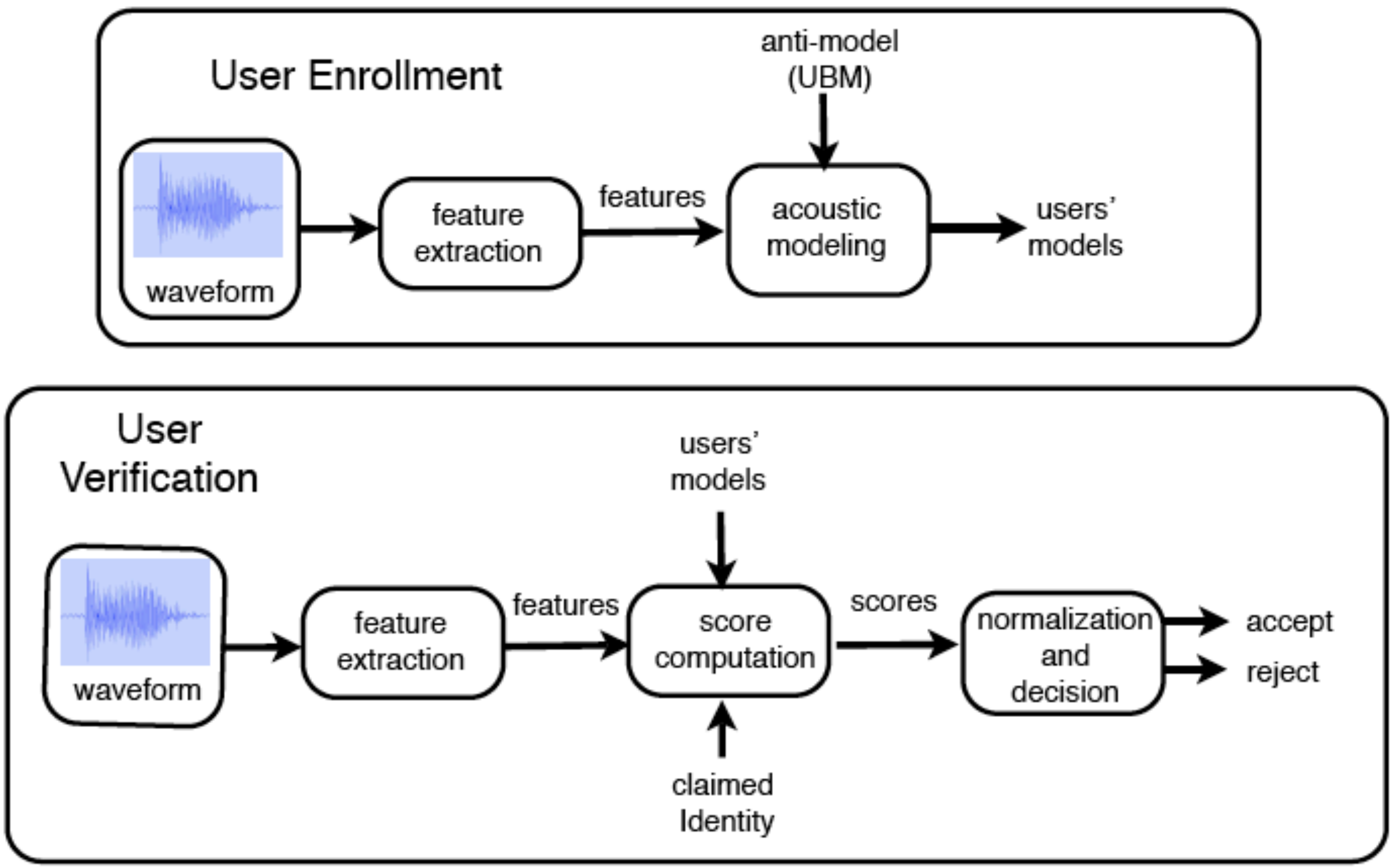
$$P(X;\lambda) = \sum_k w_k \mathcal{N}(X;\mu_k,\Sigma_k), \qquad \lambda = (w_k,\mu_k,\Sigma_k)$$

Three sample GMMs for a 3D feature vector:



- The GMM characterizes the set of mechanical configurations of a person's vocal tract.

Speaker Verification



Motivation of using JFA

Traditional systems are based on the estimation of the probability density functions (GMM in this case).

① UBM Generation

- We take all the data available and model a GMM (independent to the target speakers).
- The technique used is: Expectation Maximization (EM).

Motivation of using JFA

② Speaker model generation:

Problem:

- The amount of speech is quite small for an optimal estimation.
- It is not possible to use rely on EM

Solution: MAP (maximum aposteriori)

$$\theta_{MAP} = \underset{\lambda,\vartheta}{\operatorname{argmax}}\ p(X|\lambda;\vartheta)p(\lambda;\vartheta),$$

Motivation of using JFA

What is the real problem?

- ❑ Speaker data trained over different channels.
- ❑ MAP doesn't work. It does assume conventional conjugate priors.

What is the solution for non-ideal cases?

JFA!!!

- ❑ Provides priors for the parameters.
- ❑ Separates the speaker and the channel factors.
- ❑ The channel factors don't give information of the speaker so they can be marginalized out when computing score.

Motivation

② Speaker model generation JFA Joint Factor Analysis

- ❑ Is it possible to include a new latent variable? YES!!!
- ❑ What is the new model?

$$M = m + Vy + Ux + Dz,$$

$m \rightarrow CF \times 1$ supervector

$V \rightarrow$ low rank matrix eigenvoices

$y \rightarrow$ speaker factors

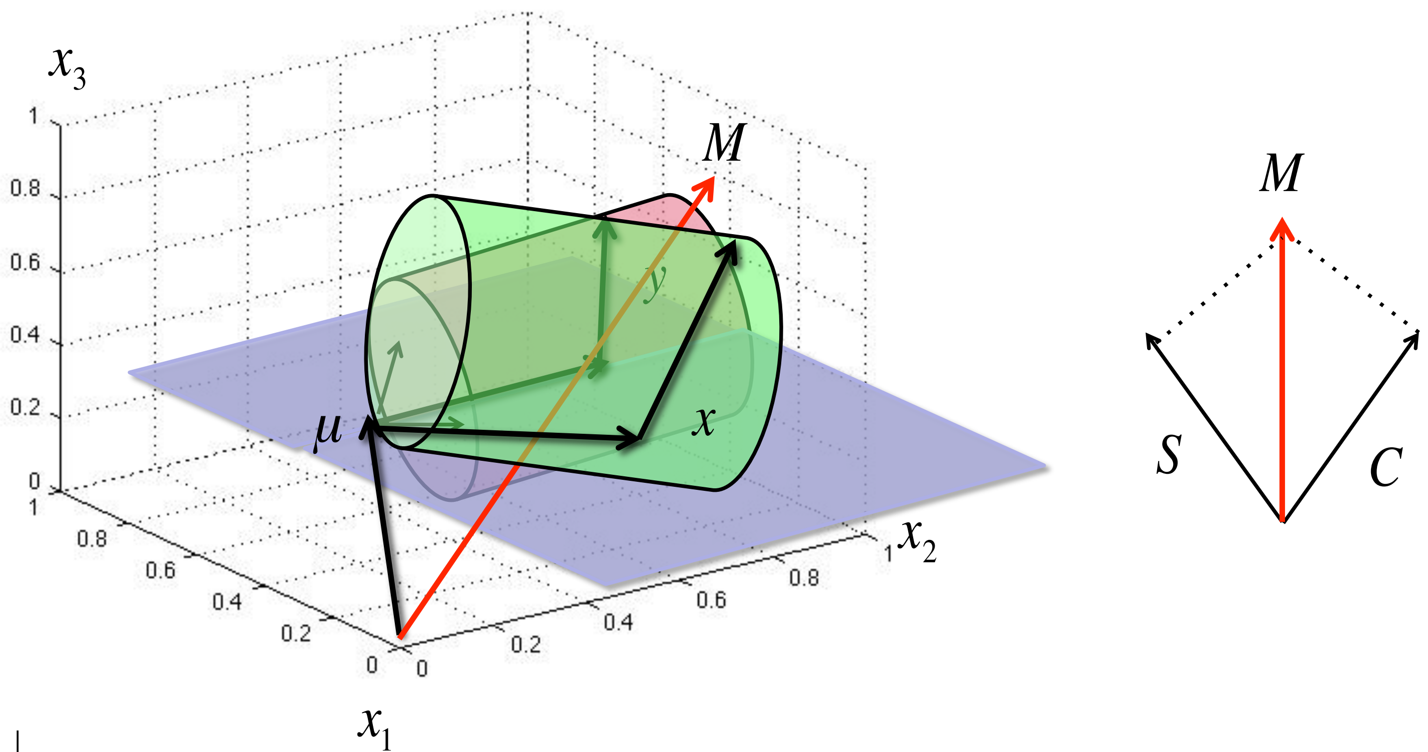
$U \rightarrow$ low rank matrix eigenchannels

$x \rightarrow$ channel factors

$D \rightarrow$ diagonal matrix

$z \rightarrow$ normally distributed random vector

Factor Analysis (FA): Geometrical representation

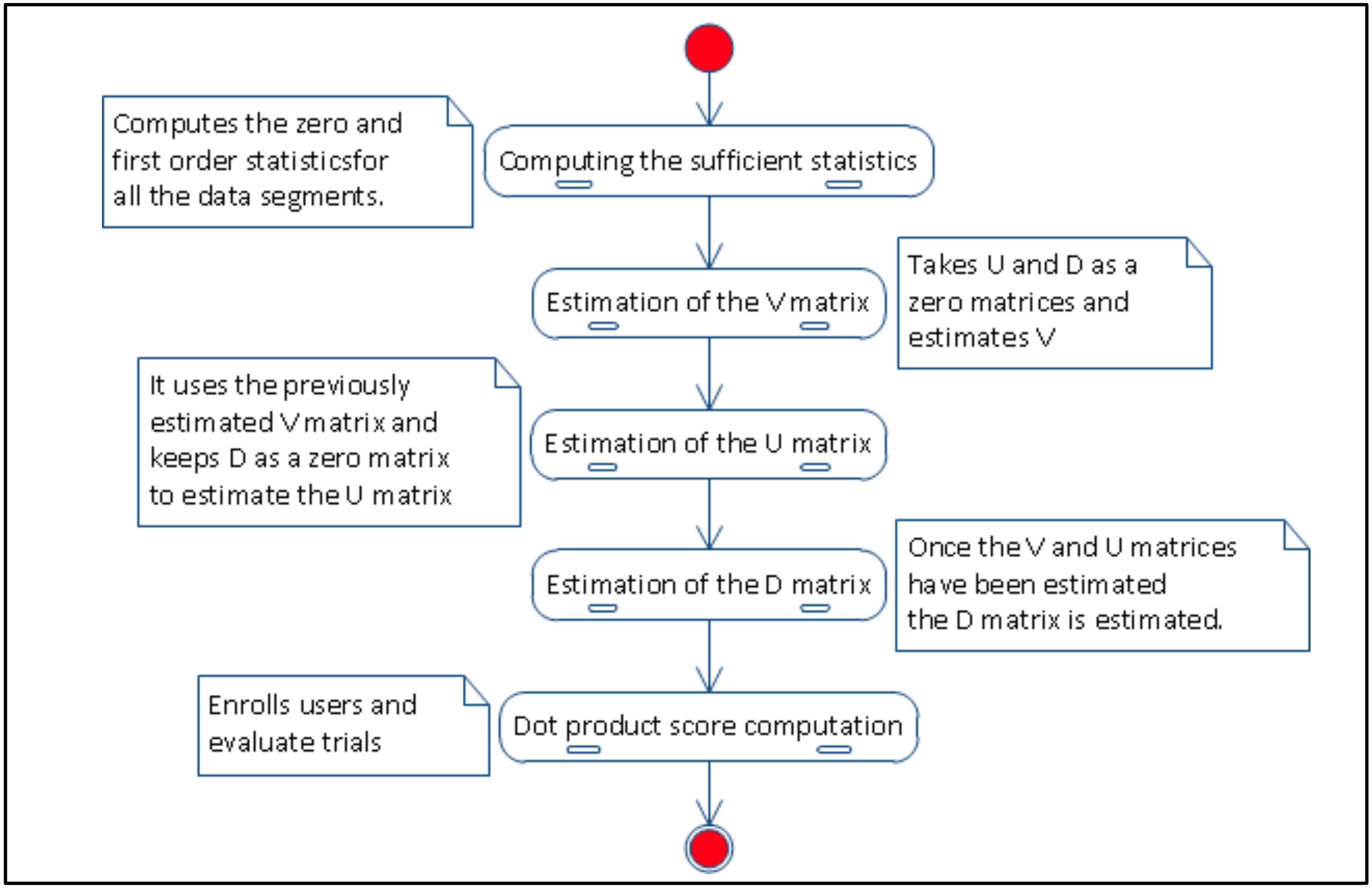


Algorithm

We may use a variable change in order to get an estimation of the VY, UX and DZ contributions with the Factor Analysis estimating methods.

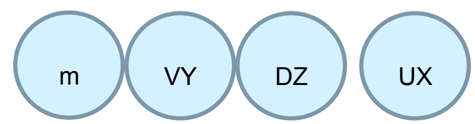
$$
\begin{aligned}
Data1' &= m + DZ + UX \\
Data &= Data1' + VY \\
Data2' &= m + VY + DZ \\
Data &= Data2' + UX \\
Data3' &= m + VY + UX \\
Data &= Data3' + DZ
\end{aligned}
$$

JFA Algorithm



Algorithm

1. Compute Sufficient Statistics
2. Compute V and Y
3. Compute U and X
4. Compute D and Z



History

What happened next?

Researchers discovered that the channel factors contained information of the speaker. ☺

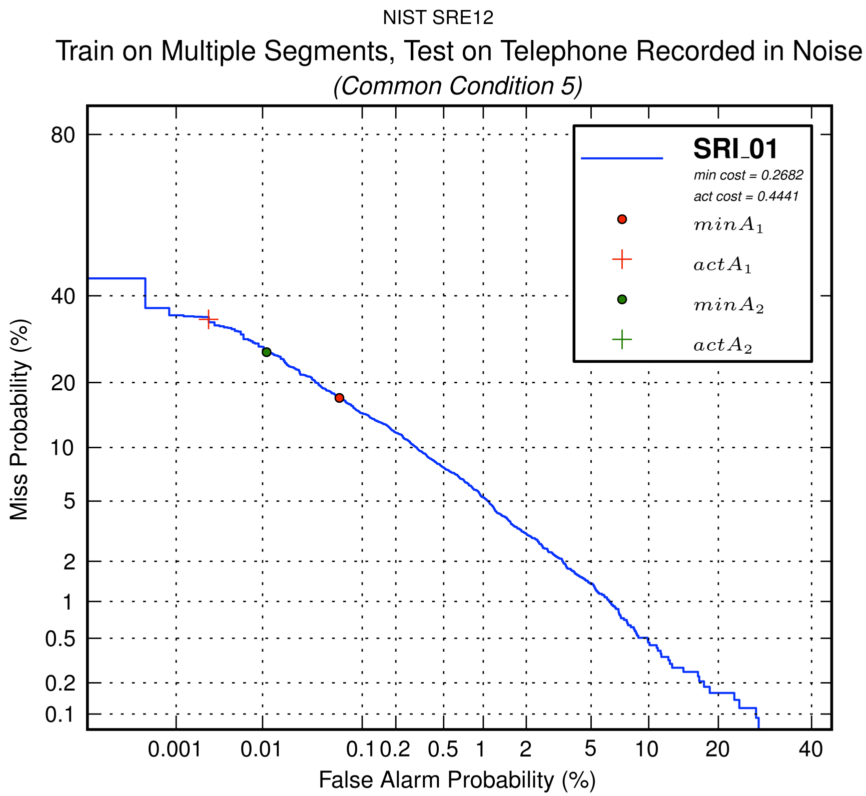
Go back to factor analysis!!! Now is called: I-vectors!!!

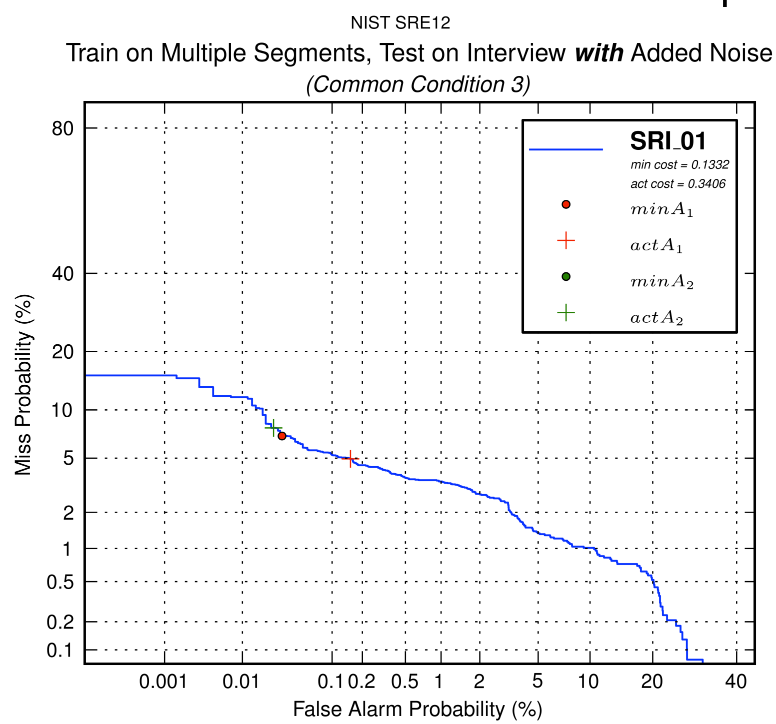
Important notes:

- JFA is actually used to build a model of the data
- I-vectors are used as feature extractor:

  Obtains the important information of the speakers and transforms it into vectors.

Some results… Last week. Best system!





References:

- **Saul and Rahim.** Maximum Likelihood and Minimum Classification Error Factor Analysis for Automatic Speech Recognition
- **D'Souza.** Derivation of Maximum Likelihood Factor Analysis using EM
- **Johnson and Wichern.** Applied Multivariate Statistical Analysis
- **Dehak, N., Kenny, P., Dehak, R., Dumouchel, P and Ouellet, P.** Front-End Factor Analysis for Speaker Verification
- **Kenny, P** Joint factor analysis of speaker and session variability : Theory and algorithms - Technical report CRIM-06/08-13 Montreal, CRIM, 2005