



# Αναγνώριση Προτύπων

Σχολή ΗΜ&ΜΥ, ΕΜΠ  
Πετρος Μαραγκος

**Compact Subspace Representations  
(SVD, PCA, KLT, LDA, ICA, CCA)**

***Συμπαγείς Αναπαραστάσεις Υποχώρων***

**<http://cvsp.cs.ntua.gr/courses/patrec>**

# Big (Multimodal) Data Challenges

## ■ Data are Voluminous:

- 24 hrs of TV = 430 Gb = 2.160.000 still (frame) images
- WWW: 300-hr videos are uploaded on YouTube per minute.
- 300 millions images are uploaded on FaceBook per day.
- Kinect sensor: 250 MB/sec (uncompressed RGB)

## ■ Data are Dynamic

- Temporal video, Website updating, News quickly get obsolete

## ■ Different Temporal Rates

- Video: 25-30 frames /second
- Audio: 44000 sound samples/sec,
- Speech: 100 feature-frames/sec, 4 syllables/sec

## ■ Cross-Media asynchrony

- image and audio scene boundaries are different

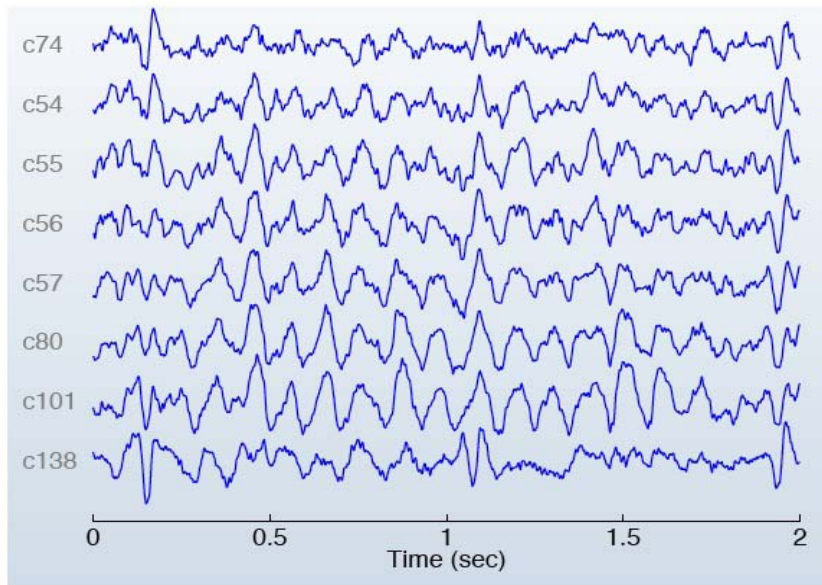
# Πολυδιάστατα Θορυβώδη Δεδομένα !

ΕΙΚΟΝΕΣ ΠΡΟΣΩΠΩΝ

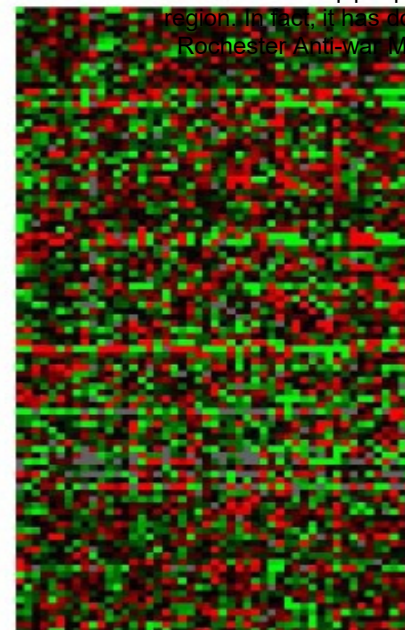


ΚΕΙΜΕΝΟ

Celebrations over the historic election of the first black president have been interspersed with emotional protests against a series of bans on gay marriage, and calls to remember the bigger picture. In California, voters approved a proposition by 52.5 percent to amend the state constitution to ban same sex marriage. Similar bills also passed in Arizona and Florida. Since then, protests against the measure have surged in Los Angeles and San Francisco, and civil rights organizations have banded together to attempt to block the implementation of the measure. On other fronts, animal rights activists are also calling the passage Proposition 2 in California a victory, and an anti-choice measure failed in California. Reports of celebrations over Barak Obama's victory continue to come in from Hollywood, New York, and New Orleans, while activists are working to keep the issues in the forefront: » Rochester, NY — a coalition of anti-war and veterans organizations banded together to make the following statement: "For nearly two years, major polls have shown that a majority of Americans want an end to the war in Iraq. The war has not "brought democracy" to the Middle East. It has not improved the lives of the Iraqi people, nor has it reduced the violence in the region. In fact, it has done exactly the opposite." Read More: Rochester Anti-War Movement Responds to 2008 Election

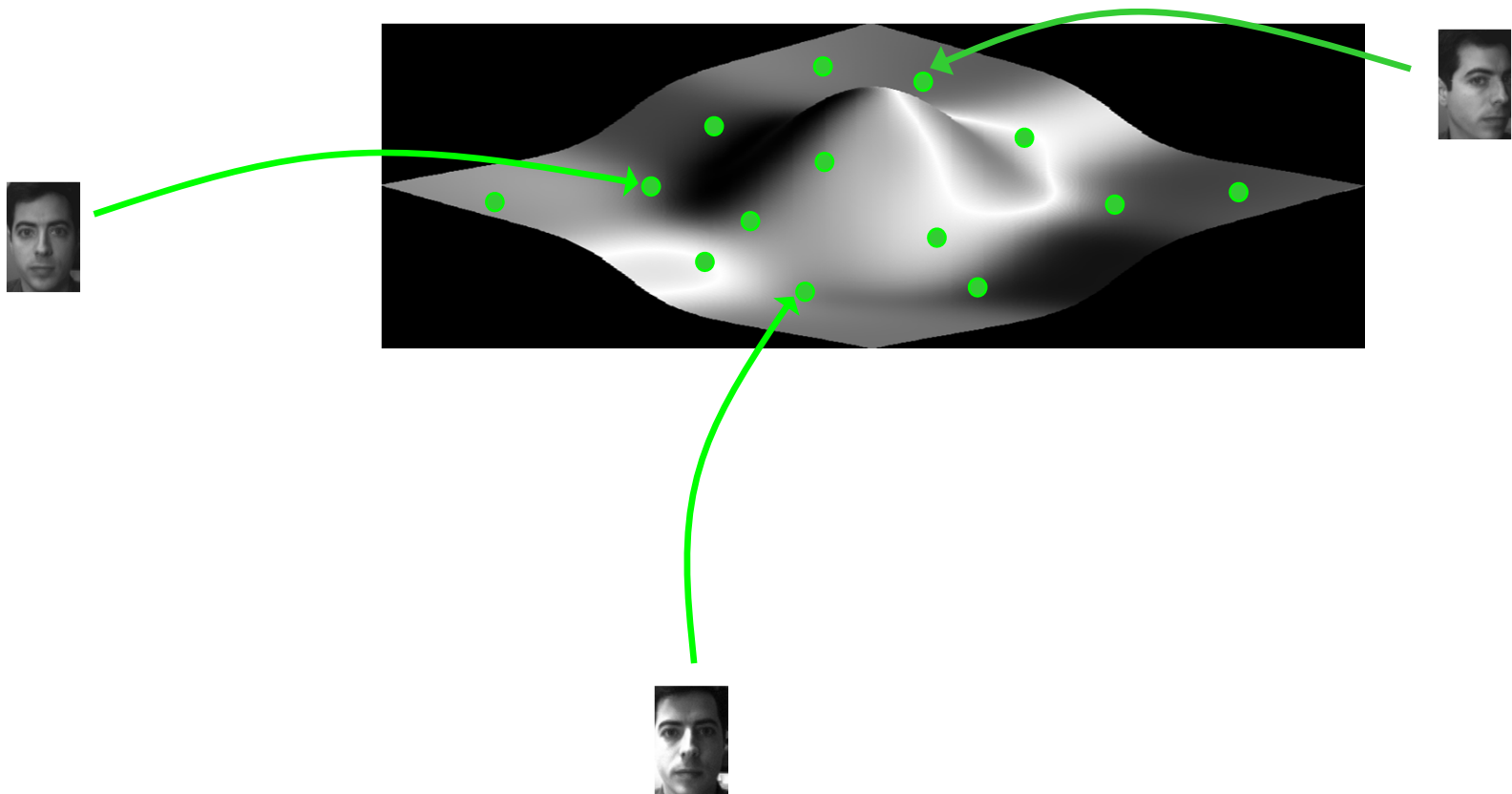


MEG



γονίδια

# Manifold of face category



# Singular Value Decomposition (SVD)

Ref: G. Strang, *Linear Algebra and Its Applications*, 1986.

**Singular Value Decomposition (SVD):** Any (real or complex)  $m \times n$  matrix  $\mathbf{A}$  can be factored as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad (7)$$

where the  $m \times m$  matrix  $\mathbf{U}$  is unitary and its columns  $\mathbf{u}_1, \dots, \mathbf{u}_m$  are the eigenvectors of  $\mathbf{A}\mathbf{A}^H$ , the  $n \times n$  matrix  $\mathbf{V}$  is unitary and its columns  $\mathbf{v}_1, \dots, \mathbf{v}_n$  are the eigenvectors of  $\mathbf{A}^H\mathbf{A}$ , and the  $m \times n$  matrix  $\mathbf{\Sigma}$  is real diagonal whose only nonzero elements are its  $r$  diagonal terms  $\sigma_1, \sigma_2, \dots, \sigma_r > 0$ , called *singular values*, with

$$r = \text{rank}(\mathbf{A}) \leq \min(m, n). \quad (8)$$

The singular values are the square roots of the nonzero eigenvalues  $\sigma_i^2$  of both  $\mathbf{A}\mathbf{A}^H$  and  $\mathbf{A}^H\mathbf{A}$ . Thus, the SVD of  $\mathbf{A}$  is related to the spectral decomposition of the Hermitian  $\mathbf{A}\mathbf{A}^H$  as follows:

$$\begin{aligned} \mathbf{A}\mathbf{A}^H &= \mathbf{U}\mathbf{\Sigma}\mathbf{\Sigma}^T\mathbf{U}^H = \sum_{i=1}^r \sigma_i^2 \mathbf{u}_i \mathbf{u}_i^H \\ \mathbf{A}^H\mathbf{A} &= \mathbf{V}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{V}^H = \sum_{i=1}^r \sigma_i^2 \mathbf{v}_i \mathbf{v}_i^H \end{aligned} \quad (9)$$

If  $\mathbf{A}$  is real, the only difference in its SVD (compared to the complex case) is that  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices. If  $\mathbf{A}$  is Hermitian and positive semidefinite, its SVD is identical to its spectral decomposition  $\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ . If  $\mathbf{A}$  is indefinite, then any negative eigenvalue in  $\mathbf{\Lambda}$  becomes positive in  $\mathbf{\Sigma}$ .

## Applications of SVD:

- (1) **Effective Rank:** Keep only the singular values above a threshold that determines the numerical precision.
- (2) *Image/Signal Compact Representation:* Use only a few large singular values to approximately represent  $\mathbf{A}$  using a truncated version of (7).
- (3) *Polar Decomposition:* Factorize a real square matrix  $\mathbf{A}$  as  $\mathbf{Q}\mathbf{S}$  where  $\mathbf{Q}$  is orthogonal and  $\mathbf{S}$  is symmetric and positive semidefinite. (If  $\mathbf{A}$  is invertible,  $\mathbf{S}$  is positive definite.)

$$\mathbf{A} = \mathbf{Q}\mathbf{S}, \quad \mathbf{Q} = \mathbf{U}\mathbf{V}^T, \quad \mathbf{S} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^T \quad (12)$$

This has applications in robotics where  $\mathbf{Q}$  represents rotation or reflection, and  $\mathbf{S}$  represents stretching or compression.

- (4) *Least Squares:* The **minimum length least squares solution** to the set of linear equations  $\mathbf{A}\mathbf{x} = \mathbf{b}$  is given by

$$\mathbf{x}^+ = \mathbf{A}^+\mathbf{b}, \quad \mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^H \quad (13)$$

where  $\mathbf{A}^+$  is called the **pseudoinverse** of  $\mathbf{A}$  and  $\mathbf{\Sigma}^+$  is a diagonal matrix with  $1/\sigma_1, \dots, 1/\sigma_r$  as its only nonzero diagonal terms.

## Προσεγγιση Πινακα (Εικονα, Δεδομενα) με SVD

Let a matrix  $\mathbf{A}$  be factorized using SVD and order the singular values as  $\sigma_1 \geq \dots \geq \sigma_r$ . If we approximate the matrix by keeping the  $p < r$  largest singular values

$$\hat{\mathbf{A}} = \sum_{k=1}^p \sigma_k \mathbf{u}_k \mathbf{v}_k^H \quad (14)$$

this approximation yields the smallest squared error

$$J_{svd} = \sum_{i=1}^m \sum_{j=1}^n |A(i, j) - \hat{A}(i, j)|^2 = (\|\mathbf{A} - \hat{\mathbf{A}}\|_{Frobenious})^2 \quad (15)$$

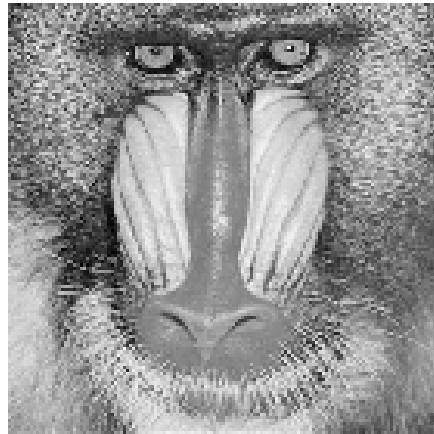
among all  $m \times n$  matrices with rank  $p$ . Thus, by keeping the  $p$  largest singular values the SVD gives the best rank- $p$  matrix approximation that minimizes the Frobenius norm of the error. The minimum error is

$$J_{svd} = \sum_{k=p+1}^r \sigma_k^2 \quad (16)$$

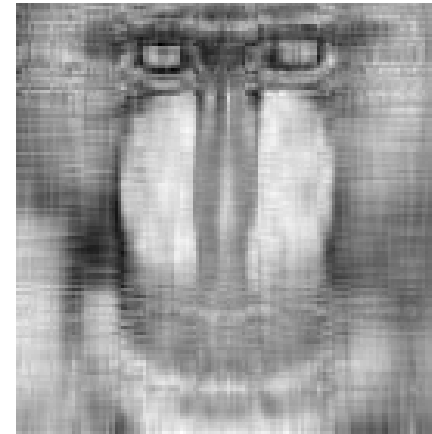


# Παραδειγμα SVD: Γκριζες Εικονες

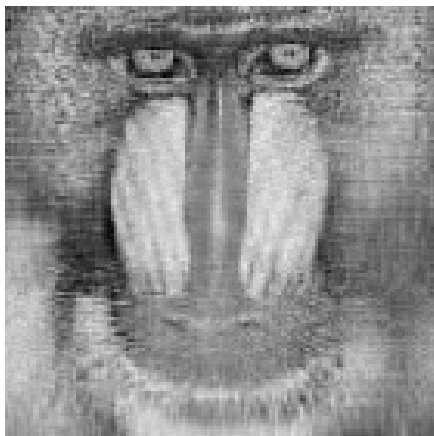
Original Image



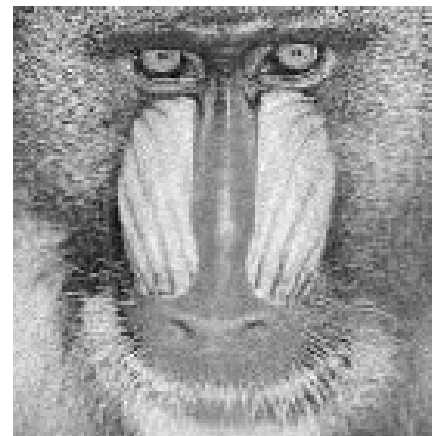
$k=10$



$k=20$



$k=50$



# Διακριτοί Ορθογωνιοί Μετασχηματισμοί

Εισοδος: Αρχικό σημα-διανυσμα:  $\mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T$

Εξοδος: Μετ/σμενο σημα-διανυσμα:  $\mathbf{y} = [y[0], y[1], \dots, y[N-1]]^T$

Ο  $N \times N$  Πινακας Μετασχηματισμου  $\mathbf{A} = [a[n, k]]$  είναι Unitary  
( $\mathbf{A}$  είναι Ορθογωνιος για Πραγματικους πινακες):

$$\mathbf{A}\mathbf{A}^H = \mathbf{I}$$

**Γραμμικός Μετασχηματισμός** (Πινακας x Διανυσμα εισοδου):

$$\mathbf{y} = \mathbf{A}^H \mathbf{x}, \quad \left( \begin{array}{l} y[k] = \sum_{n=0}^{N-1} a^*[k, n]x[n] \\ k = 0, 1, \dots, N-1 \end{array} \right)$$

**Αντιστροφος Μετασχηματισμος** (Αντιστροφος Πινακας x Διανυσμα Εξοδου):

$$\mathbf{x} = \mathbf{A}\mathbf{y}, \quad \left( \begin{array}{l} x[n] = \sum_{k=0}^{N-1} a[n, k]y[k] \\ n = 0, 1, \dots, N-1 \end{array} \right)$$

## Unitary Discrete Fourier Transform (DFT)

Αναλυση: 
$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] W_N^{kn}, \quad k = 0, 1, \dots, N-1$$

Συνθεση: 
$$x[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X[k] W_N^{-kn}, \quad n = 0, 1, \dots, N-1$$

Διανυσμα χρονικων δειγματων: 
$$\mathbf{x} = [x[0], \dots, x[N-1]]^T$$

Διανυσμα συχνοτικων δειγματων: 
$$\mathbf{y} = [X[0], \dots, X[N-1]]^T$$

DFT Πινακας: 
$$\mathbf{F} = \left[ \frac{1}{\sqrt{N}} W_N^{-kn} \right], \quad k, n = 0, 1, \dots, N-1$$

$\mathbf{F}$  ειναι Unitary ( $\mathbf{F}\mathbf{F}^H = \mathbf{I}$ ) και Συμμετρικος ( $\mathbf{F} = \mathbf{F}^T$ )

Ευθης: 
$$\mathbf{y} = \mathbf{F}^H \mathbf{x}, \quad \text{Αντιστροφος: } \mathbf{x} = \mathbf{F} \mathbf{y}$$

## Discrete Cosine Transform (DCT)

$$\mathbf{C} = [c[k, n]] = \begin{cases} \frac{1}{\sqrt{N}}, & k=0, 0 \leq n \leq N-1 \\ \sqrt{\frac{2}{N}} \cos \left[ \frac{\pi k (2n+1)}{2N} \right], & 1 \leq k \leq N-1, 0 \leq n \leq N-1 \end{cases}$$

**DCT** διανυσματος  $\mathbf{x} = [x[0], x[1], \dots, x[N-1]]^T$  :

$$y[k] = \beta[k] \sum_{n=0}^{N-1} x[n] \cos \left[ \frac{\pi k (2n+1)}{2N} \right], \quad 0 \leq k \leq N-1$$

$$\beta[0] = \sqrt{\frac{1}{N}}, \quad \beta[k] = \sqrt{\frac{2}{N}} \quad 1 \leq k \leq N-1$$

**Αντιστροφος DCT:**

$$x[n] = \sum_{k=0}^{N-1} \beta[k] y[k] \cos \left[ \frac{\pi k (2n+1)}{2N} \right], \quad 0 \leq n \leq N-1$$

# Principal Component Analysis (PCA)

## Karhunen Loeve Transform (KLT)

Ανάλυση σε Πρωτεύουσες  
Συνιστώσες

**KLT, PCA (1):** Assume (real or complex) random data vectors  $\mathbf{x} \in \mathbb{C}^d$  (which may represent a signal segment or some feature vector in a pattern recognition problem) We wish to find a unitary linear transformation (matrix)  $\mathbf{A}$  such that the transformed vectors

$$\mathbf{y} = \mathbf{A}^H \mathbf{x}, \quad \mathbf{A}^{-1} = \mathbf{A}^H,$$

should have two properties: 1) Orthogonal or uncorrelated components, and 2) if we keep the first  $p < d$  components to obtain a minimum Mean Squared Error (MSE). The solution is the KLT (a.k.a. PCA). Suppose we know the orthonormal vectors  $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$  that form the columns of  $\mathbf{A}$  and view them as an orthonormal basis for the whole space. Then,

$$\mathbf{x} = \sum_{i=1}^d y_i \mathbf{e}_i, \quad y_i = \langle \mathbf{x}, \mathbf{e}_i \rangle = \mathbf{e}_i^H \mathbf{x}$$

If we project this vector onto the subspace formed by the smaller basis  $\{\mathbf{e}_1, \dots, \mathbf{e}_p\}$ , then the best approximation  $\hat{\mathbf{x}}$  is

$$\hat{\mathbf{x}} = \sum_{k=1}^p y_k \mathbf{e}_k$$

and the corresponding MSE  $J$  is

$$\begin{aligned} J &= \mathcal{E}\{\|\mathbf{x} - \hat{\mathbf{x}}\|^2\} = \mathcal{E}\{\|\mathbf{x}\|^2\} - \sum_{k=1}^p \mathcal{E}\{|y_k|^2\} \\ &= \sum_{i=p+1}^d \mathcal{E}\{|y_i|^2\} = \sum_{i=p+1}^d \mathbf{e}_i^H \mathbf{R}_x \mathbf{e}_i \end{aligned}$$

where  $\mathbf{R}_x = \mathcal{E}\{\mathbf{x}\mathbf{x}^H\}$  is the input correlation matrix.

**KLT, PCA (2):** Now we want to find the optimal basis vectors  $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ , i.e. the optimum  $\mathbf{A}$ , and to select the  $p$  principal directions among them such that the MSE  $J$  is minimized. By minimizing  $J$  subject to the constraints  $\mathbf{e}_i^H \mathbf{e}_i = 1$ , we find that the best orthonormal basis  $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$  consists of the eigenvectors of  $\mathbf{R}_x$ :

$$\mathbf{R}_x \mathbf{e}_i = \lambda_i \mathbf{e}_i, \quad i = 1, 2, \dots, d$$

Hence,  $\mathbf{e}_i^H \mathbf{R}_x \mathbf{e}_i = \lambda_i$ . Thus, to minimize  $J$  we should order the eigenvalues as  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_d$  and choose as *principal directions*  $\{\mathbf{e}_1, \dots, \mathbf{e}_p\}$  those that correspond to the  $p$  largest eigenvalues. The corresponding transform values  $\{y_k : k = 1, \dots, p\}$  are called the *principal components*. Then, the minimum error equals the sum of the  $d - p$  smallest eigenvalues:

$$J_{min} = \sum_{i=p+1}^d \lambda_i$$

The above choices diagonalize the correlation matrix of the transformed vector:

$$\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\} = \mathbf{A}^H \mathbf{R}_x \mathbf{A} = \text{diag}[\lambda_1, \dots, \lambda_d]$$

Thus, the transformed vector components  $\{y_i\}$  are orthogonal and their variances equal the eigenvalues of  $\mathbf{R}_x$ ; i.e.,  $\mathcal{E}\{y_i y_j\} = \lambda_i \delta_{ij}$ .

**KLT for nonzero-mean data:** If the data have a *non-zero mean*  $\mathbf{m} = \mathcal{E}\{\mathbf{x}\}$ , an alternative version of KLT is to subtract it first and then proceed as above for the new data  $\mathbf{x}' = \mathbf{x} - \mathbf{m}$  that will be transformed into  $\mathbf{y}' = \mathbf{A}^H \mathbf{x}'$ . However, now the matrix  $\mathbf{A}$  consists of the eigenvectors  $\{\mathbf{u}_i\}$  of the covariance matrix  $\mathbf{C}_x = \mathcal{E}\{(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^H\}$ , and the minimum MSE approximation is

$$\hat{\mathbf{x}} = \mathbf{m} + \sum_{k=1}^p y'_k \mathbf{u}_k = \sum_{i=1}^p y_i \mathbf{u}_i + \sum_{i=p+1}^d \mathcal{E}\{y_i\} \mathbf{u}_i$$

The new principal values  $\{y'_k : k = 1, \dots, p\}$  are *uncorrelated* and their variances equal the  $p$  largest eigenvalues of  $\mathbf{C}_x$ .

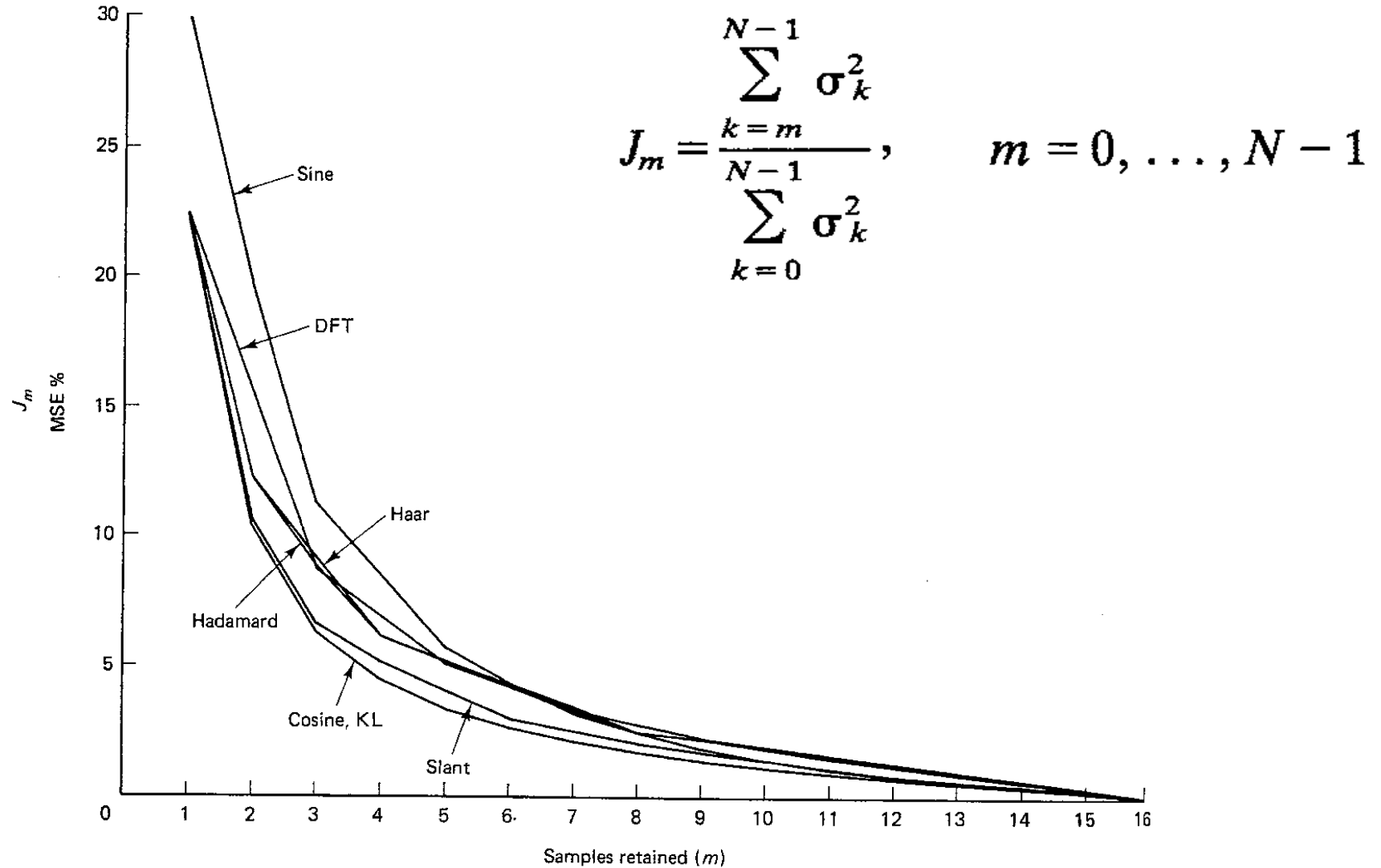


**Deterministic PCA:** The PCA problem also has a *deterministic* formulation where our data consist of  $N$  input vectors  $\mathbf{x}_n$ ,  $n = 1, \dots, N$ , and the statistical expectation  $\mathcal{E}$  above is replaced by the sample mean  $(1/N) \sum_{n=1}^N$ . Thus, the MSE error to minimize is

$$J = \frac{1}{N} \sum_{n=1}^N \left\| \mathbf{x}_n - \left( \sum_{k=1}^p y_{kn} \mathbf{e}_k \right) \right\|^2$$

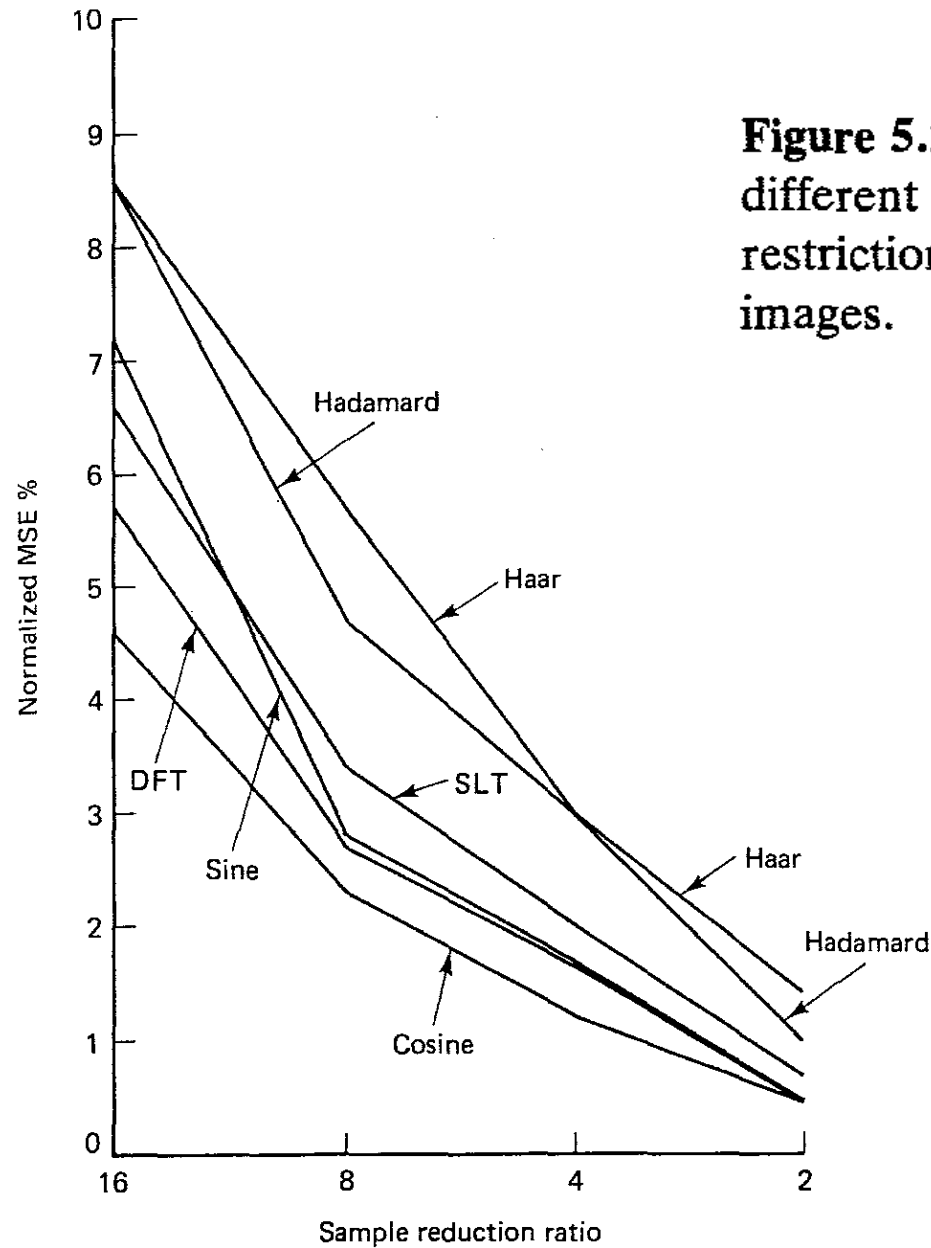
where  $y_{kn} = \langle \mathbf{x}_n, \mathbf{e}_k \rangle$ . The principal directions  $\{\mathbf{e}_i\}$  are the orthonormal eigenvectors of the sample correlation or covariance matrix.

# Λαθος για Αποκοπή Συντελεστών Μετ/σμών (για Markov ακολουθία)



**Figure 5.19** Performance of different unitary transforms with respect to basis restriction errors ( $J_m$ ) versus the number of basis ( $m$ ) for a stationary Markov sequence with  $N = 16$ ,  $\rho = 0.95$ .

# Εφαρμογή Ορθογωνίων Μετ/σμών σε Συμπιεση Εικονών



**Figure 5.23** Performance comparison of different transforms with respect to basis restriction zonal filtering for  $256 \times 256$  images.

## PCA-based Subspace Classifiers

- Given: training set of feature vectors.
- For each class  $i$ , estimate its correlation matrix and form a matrix  $\mathbf{A}_i$  whose columns are the  $p(i)$  principal eigenvectors.
- Classify unknown vector  $\mathbf{x}$  to class  $j$  if

$$\| \mathbf{A}_j^T \mathbf{x} \| > \| \mathbf{A}_i^T \mathbf{x} \| \quad \forall i \neq j$$

- Equivalent to classifying a vector in its nearest class subspace (*Πυθαγορειο θεωρημα*).
- If classification error is high, improve performance via **learning subspace methods**: iteratively rotate subspaces to adjust projection lengths of training vectors.
- Decision surfaces ?

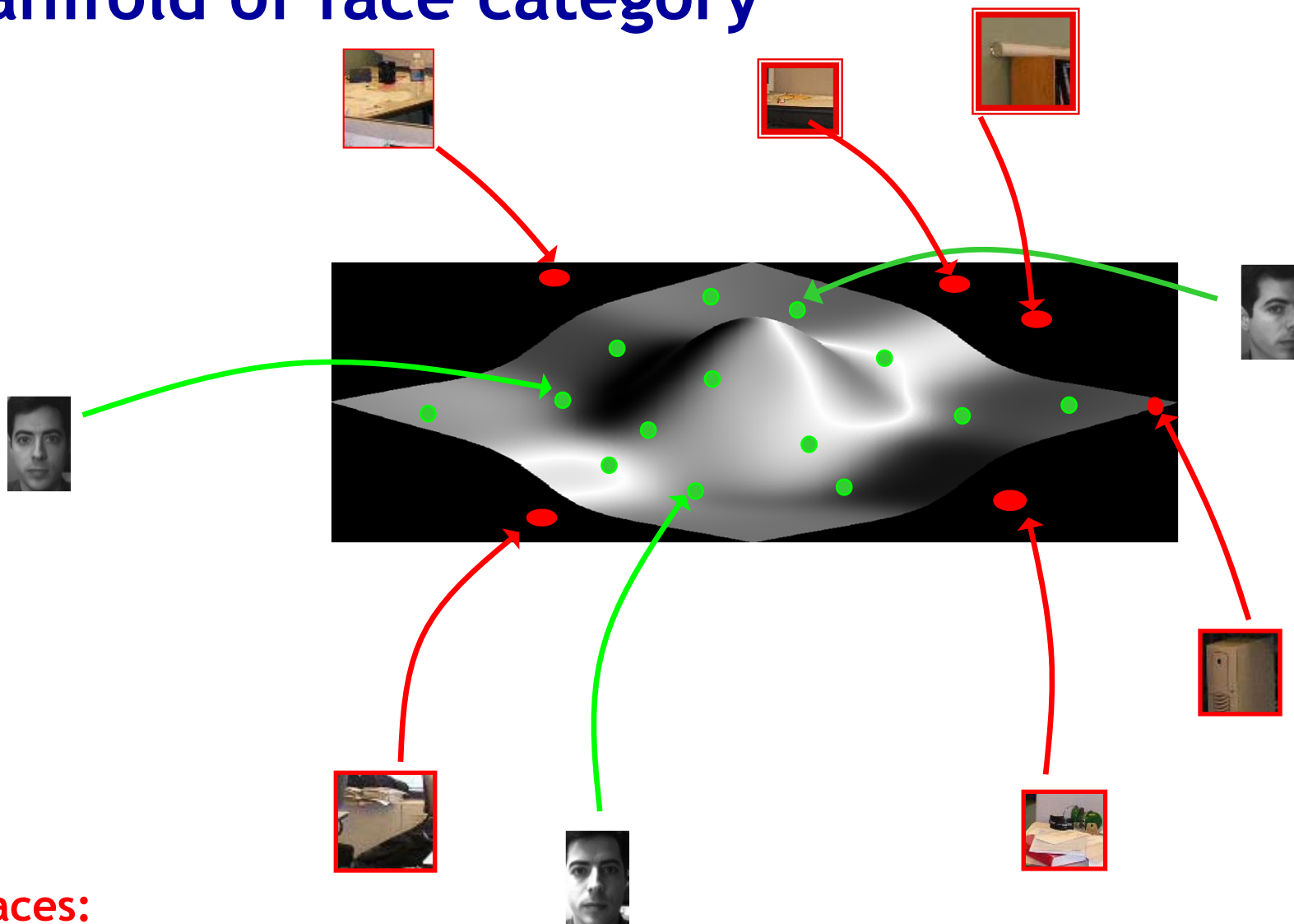
# Appearance modelling for faces

- When viewed as vectors of pixel values, face images are extremely high-dimensional
  - 100x100 image = 10,000 dimensions
- Very few vectors correspond to valid face images



- We want to model the subspace of face images

# Manifold of face category

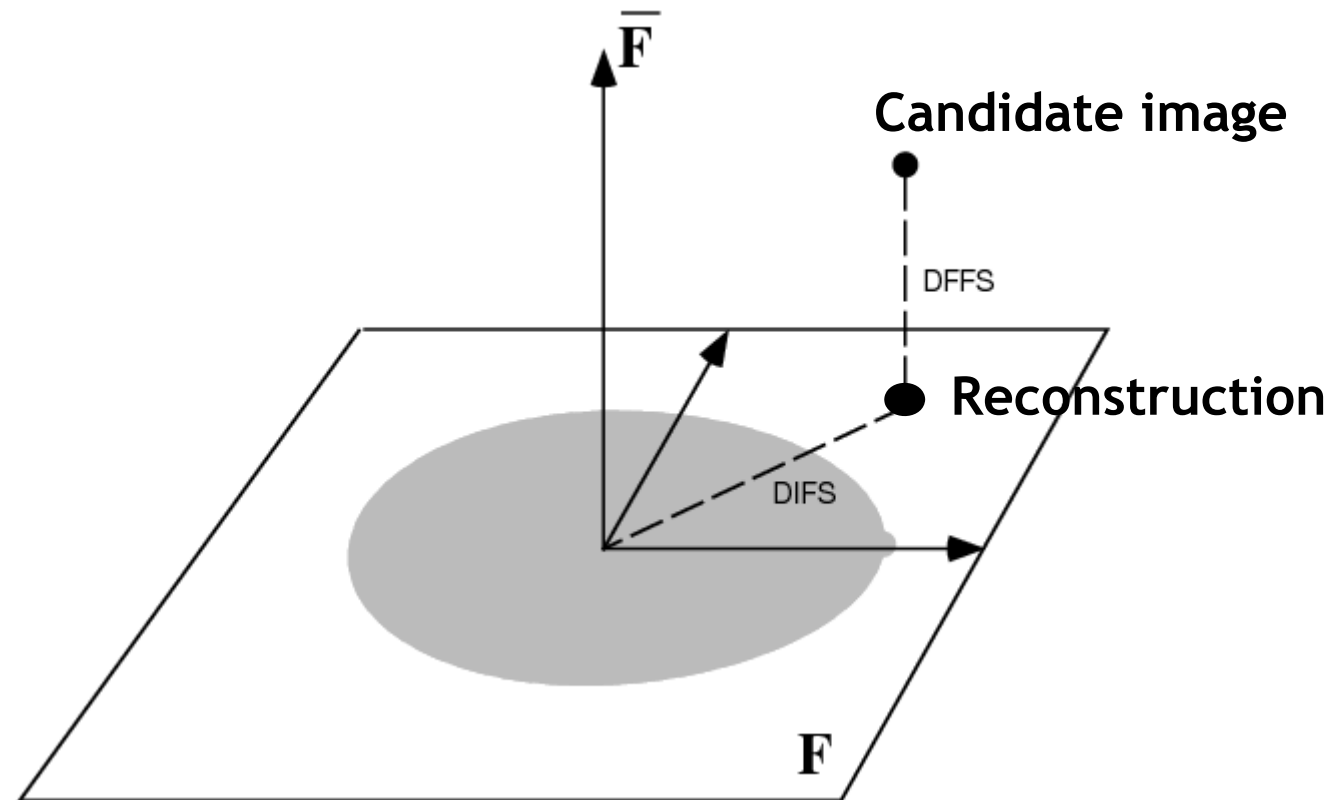


**Non-faces:**  
Lie far from manifold, or in  
underpopulated regimes

# Eigenfaces: Key idea

- Assume that most face images lie on a low-dimensional subspace determined by the first  $k$  ( $k < d$ ) directions of maximum variance
- Use PCA to determine the vectors or “eigenfaces”  $u_1, \dots, u_k$  that span that subspace
- Represent all face images in the dataset as linear combinations of eigenfaces

# Eigenfaces/PCA: linearize manifold



Linear subspace spanned by PCA model



# Eigenfaces example

- Training images
- $X_1, \dots, X_N$

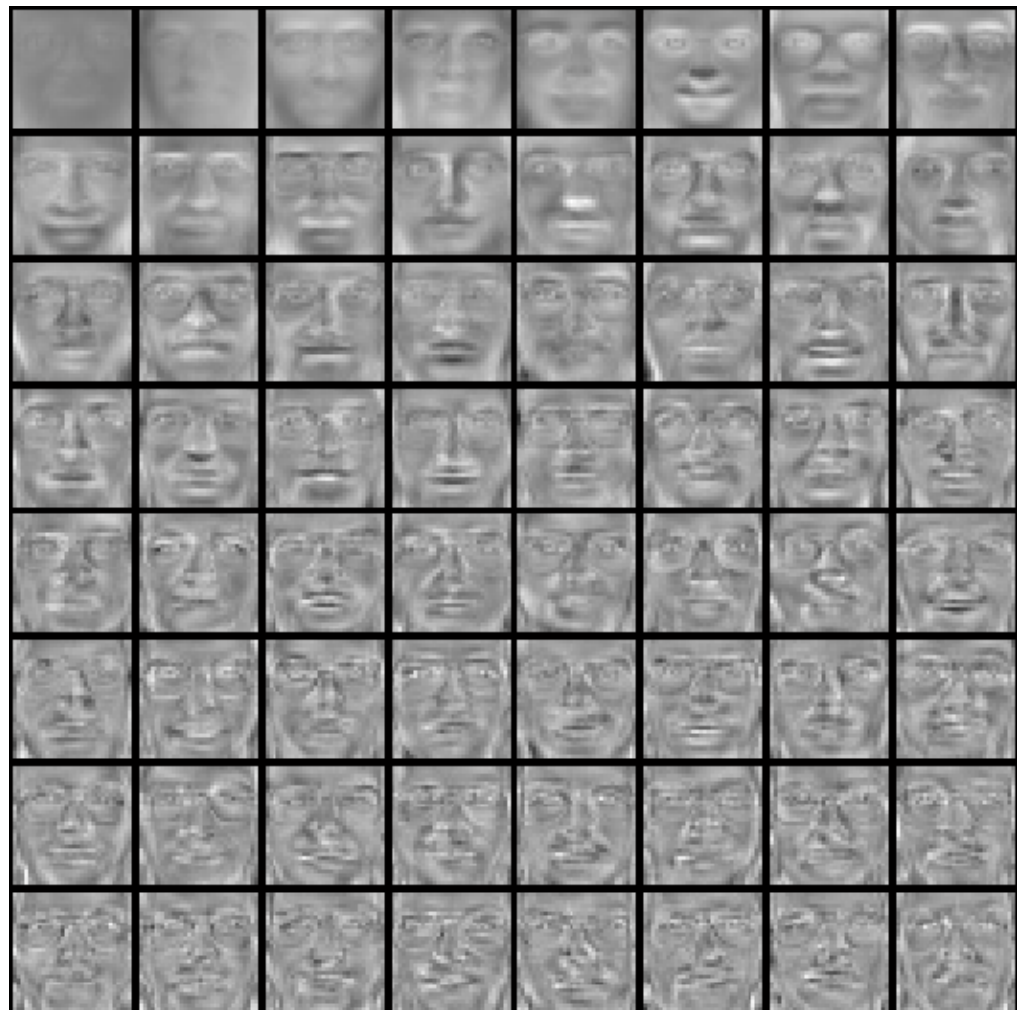


# Eigenfaces example

Mean:  $\mu$



Top eigenvectors:  $u_1, \dots, u_k$



# Eigenfaces example

Principal component (eigenvector)  $u_k$



$\mu + 3\sigma_k u_k$



$\mu - 3\sigma_k u_k$



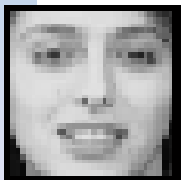
# Eigenfaces: orthonormal basis for faces

- Face  $\mathbf{x}$  in “face space” coordinates:

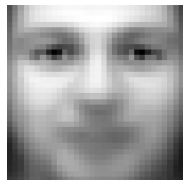


$$\mathbf{x} \rightarrow [\mathbf{u}_1^T (\mathbf{x} - \mu), \dots, \mathbf{u}_k^T (\mathbf{x} - \mu)]$$
$$= w_1, \dots, w_k$$

- Reconstruction:



=



+



$\hat{\mathbf{x}}$

=

$\mu$

+

$w_1 \mathbf{u}_1 + w_2 \mathbf{u}_2 + w_3 \mathbf{u}_3 + w_4 \mathbf{u}_4 + \dots$

## Eigenfaces (rewritten)

- New input to PCA:  $I(x)$
- Projection:

$$w_i = \sum_x (I(x) - \mu(x)) u_i(x)$$

- Reconstruction:

$$I(x) \simeq T(x; \mathbf{w}) = \mu(x) + \sum_{i=1}^N w_i u_i(x)$$

# Eigenfaces: Probabilistic formulation

- Probability of image, using Eigenface model

$$P(I)$$

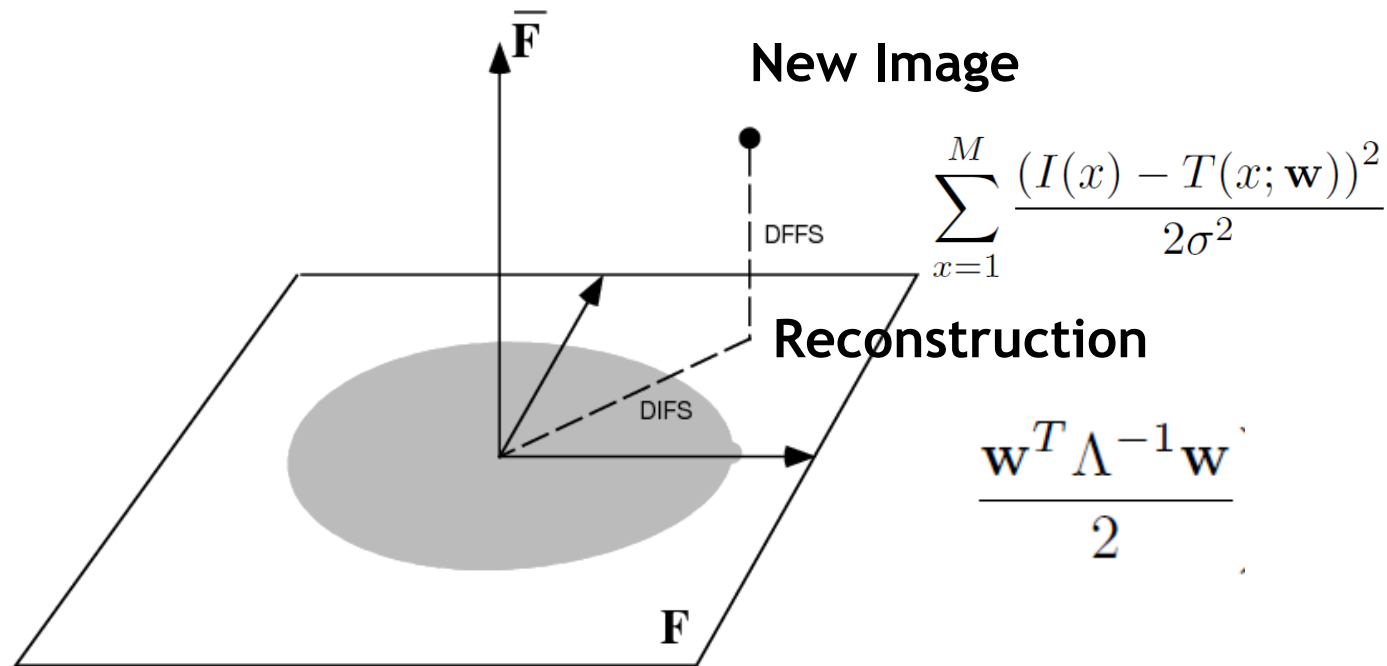
- Uncorrelated coefficients:

$$P(\mathbf{w}) = \frac{1}{\sqrt{2\pi|\Lambda|}} \exp\left(-\frac{\mathbf{w}^T \Lambda^{-1} \mathbf{w}}{2}\right) \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & & \\ 0 & 0 & \dots & \lambda_N \end{pmatrix}$$

- IID Gaussian noise assumption:

$$\begin{array}{l} \text{Image} \\ \text{Pixel} \end{array} P(I|\mathbf{w}) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\sum_{x=1}^M \frac{(I(x) - T(x; \mathbf{w}))^2}{2\sigma^2}\right)$$

# Eigenfaces for Detection



$$D(I) = DIFS(I) + DFFS(I) \begin{cases} < C & \text{face} \\ \geq C & \text{other} \end{cases}$$

- Assumptions:
  - Faces lie on a linear subspace
  - Gaussian distribution of parameters
  - Independent Identically distributed Gaussian noise

# Limitations (even if all assumptions hold)

- Global appearance method: not robust to misalignment, background variation





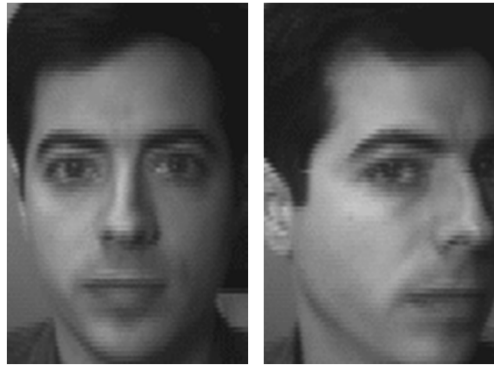
# Challenges addressed by Eigenfaces

## Short Term

### Expression



### Pose



### Illumination



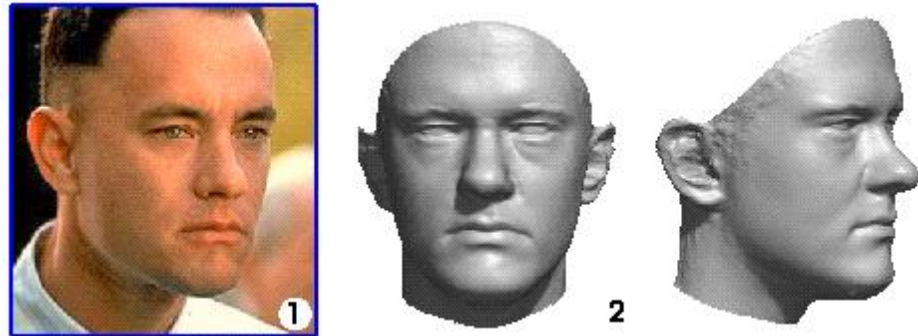
## Long Term

- Facial Hair
- Makeup
- Eyewear

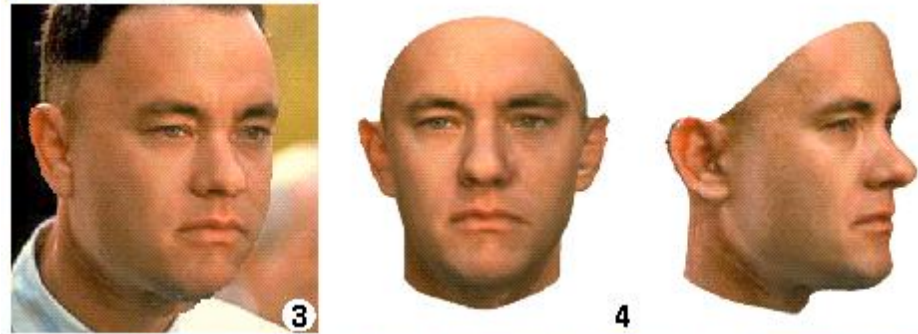
- Hairstyle
- Piercings
- Aging

# 3D Morphable models

Recover Shape



Synthesize new views



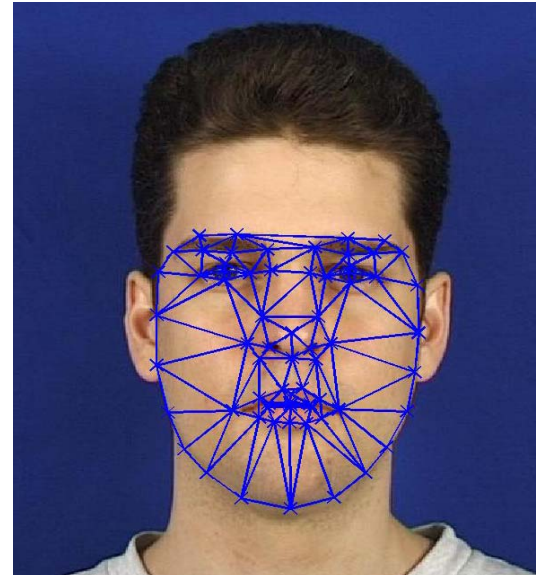
Synthesize new expressions



## PCA σε Active Appearance Models (AAMs)

**Ιδέα:** Αναπαράσταση Σχήματος και Υφής σε γραμμικούς χώρους

Παράδειγμα  
επισημείωσης 68  
σημείων-κλειδιών  
στο πρόσωπο.



Επισημειωμένο σύνολο  
εκπαίδευσης XM2VTS  
(295 πρόσωπα)

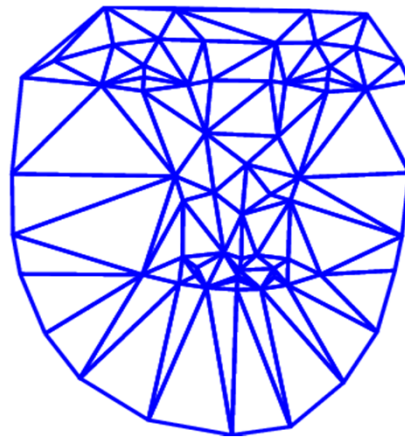


## PCA: Active Appearance Models

- **Εμφάνιση = (Σχήμα, Υφή)**
  - **Σχήμα:** συντεταγμένες 68 σημείων  $\rightarrow$  διάνυσμα 136Δ
  - **Υφή:** RGB τιμές 10000 σημεία (σε ουδετεροποιημένο ως προς το σχήμα πλαίσιο αναφοράς)  $\rightarrow$  διάνυσμα 30000Δ



Εμφάνιση  $I$



Σχήμα  $s$



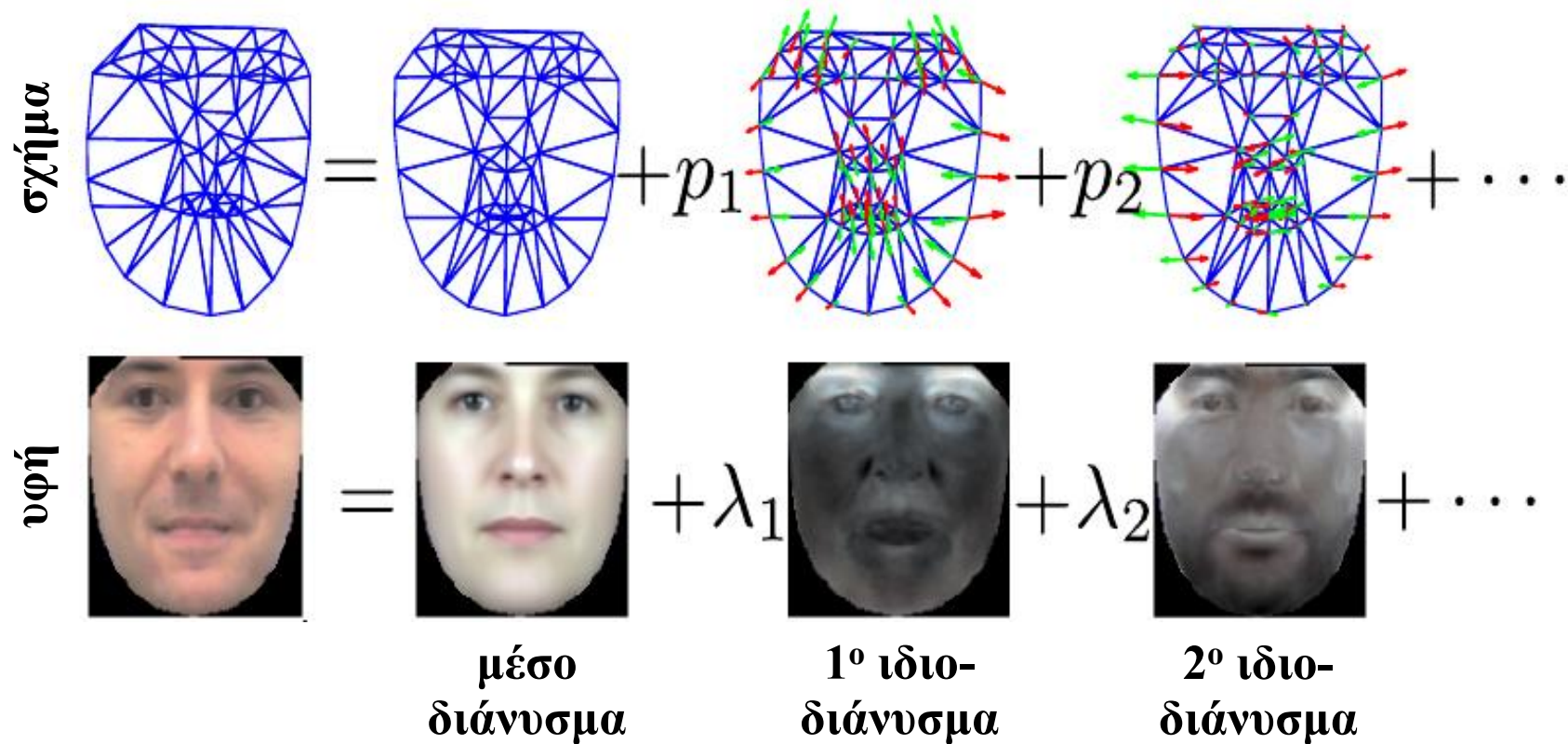
Υφή  $A(x)=I(W_s(x))$

$W_s: R^2 \rightarrow R^2$ , προβάλλει σημεία από ουδετεροποιημένο (μέσο) πλαίσιο-σχήμα αναφοράς σε σημεία στο σύστημα συντεταγμένων της εικόνας.



## PCA: Active Appearance Models

- Ελάττωση διάστασης με διπλή PCA (95% μεταβλητότητα)
- Σχήμα:  $s$  (136Δ)  $\rightarrow$   $p$  (10Δ)



- Συμπυκνωμένο διάνυσμα εμφάνισης για ερμηνεία προσώπου
- Διάνυσμα εμφάνισης (10+100Δ):  $\mathbf{q} = [\mathbf{p}^T, \boldsymbol{\lambda}^T]^T$

## Εφαρμογές PCA - AAM: Eigenfaces

- **Παράδειγμα:** Πρωτεύουσες συνιστώσες από εικόνες προσώπων:

Σύνολο  
εκπαίδευσης



Μέσο πρόσωπο



4 πρώτα ιδιοπρόσωπα



- Εφαρμογή σε αναγνώριση προσώπου ή αναγνώριση ομιλίας («διάβασμα χειλιών» ομιλητή) από χαμηλοδιάστατο διάνυσμα συντελεστών προβολής.

# Εφαρμογές PCA σε Active Appearance Models

- Audio-Visual Speech Recognition

AV



A

- Eye-gaze & facial pose estimation



- Sign language recognition



# Handshape Modeling: AAM, Dynamic & Static Priors

## ■ Shape-Appearance (SA) Representation



$$f(\mathbf{x}) = \begin{cases} g(I(\mathbf{x})), & \text{if } \mathbf{x} \in M \\ -C_b & \text{else} \end{cases}$$



## ■ Generative model

$$f(W_p(\mathbf{x})) \approx A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x})$$

$W_p(\mathbf{x})$  : 2D affine transform  
with parameters  $p \in \mathbb{R}^6$

## ■ Training of the Model

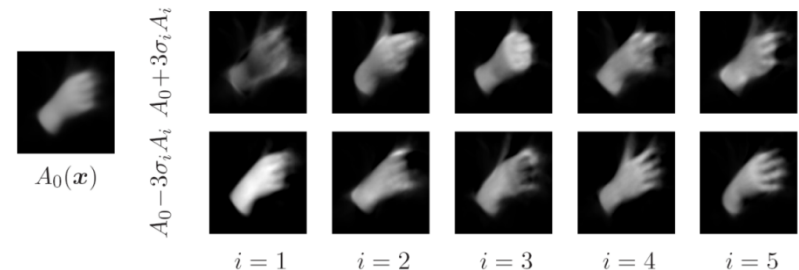
### □ Affine alignment of training set

- generalization of procrustes analysis
- iterative manual feedback



### □ PCA to learn $A_i(\mathbf{x})$

- keep only  $N_c=35$  components



## ■ Fitting : Find parameters $\lambda, p$ that minimize:

$$E(\lambda, p) = E_{rec}(\lambda, p) + w_S E_S(\lambda, p) + w_D E_D(\lambda, p)$$



# Αναφορές για Visual PCA-AAM

## Γενικές Αναφορές:

- T. F. Cootes, G. J. Edwards, and C. J. Taylor, “*Active Appearance Models*”, IEEE Trans. Pattern Analysis and Machine Intelligence, June 2001.

## Audio-Visual Speech Recognition:

- G. Papandreou, A. Katsamanis, V. Pitsikalis and P. Maragos, “*Adaptive Multimodal Fusion by Uncertainty Compensation With Application to Audiovisual Speech Recognition*”, IEEE Trans. Audio, Speech & Language Processing, 2009.

## Sign Language Recognition:

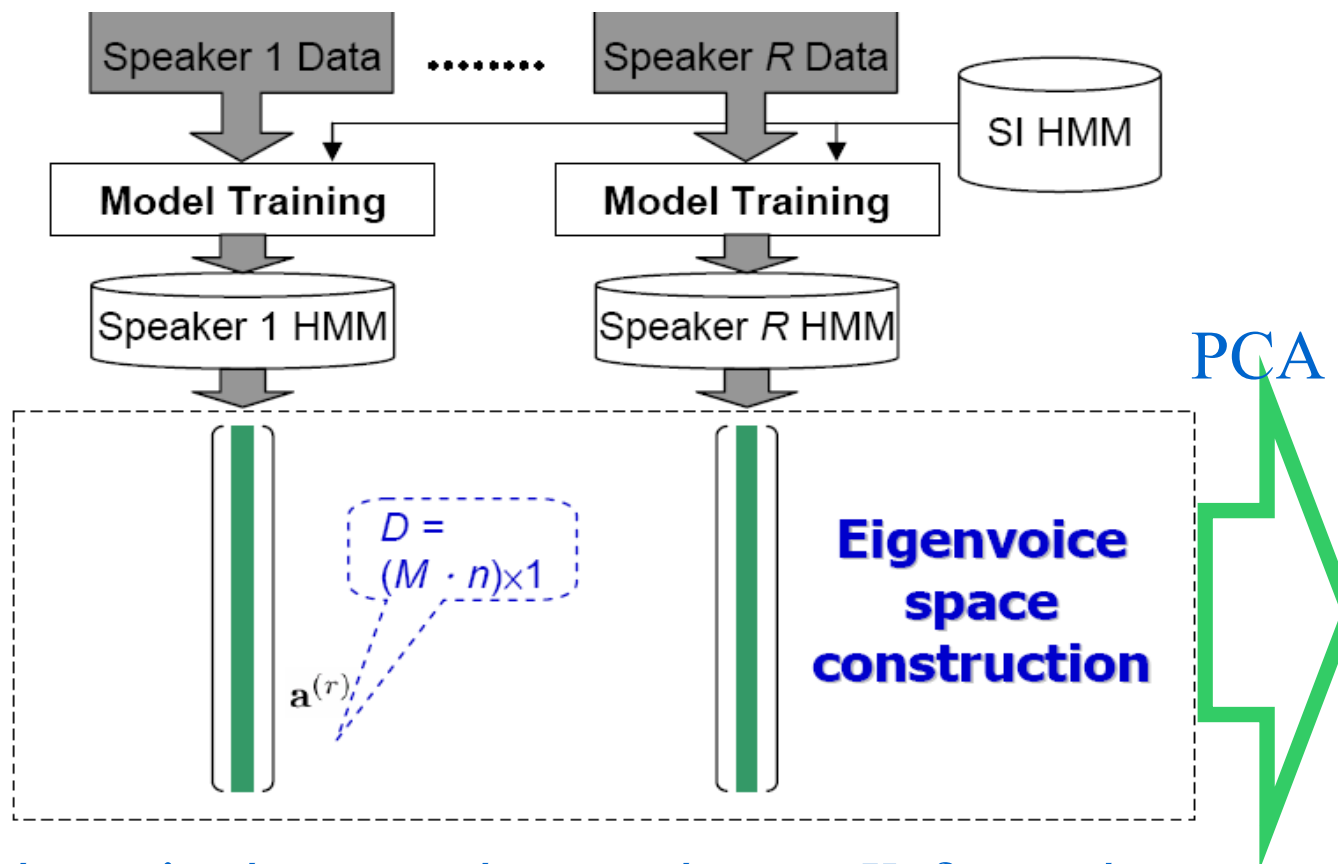
- A. Roussos, S. Theodorakis, V. Pitsikalis and P. Maragos, “*Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos*”, *Journal of Machine Learning Research*, 2013.

## Face detection/tracking, Eyegaze:

- E. Antonakos, V. Pitsikalis and P. Maragos, “*Classification of Extreme Facial Events in Sign Language Videos*”, EURASIP J. Image and Video Processing 2014.
- P. Koutras and P. Maragos, “*Estimation of Eye Gaze Direction Angles Based on Active Appearance Models*”, Proc. IEEE ICIP 2015.

# PCA: Ιδιοφωνές (eigenvoices)

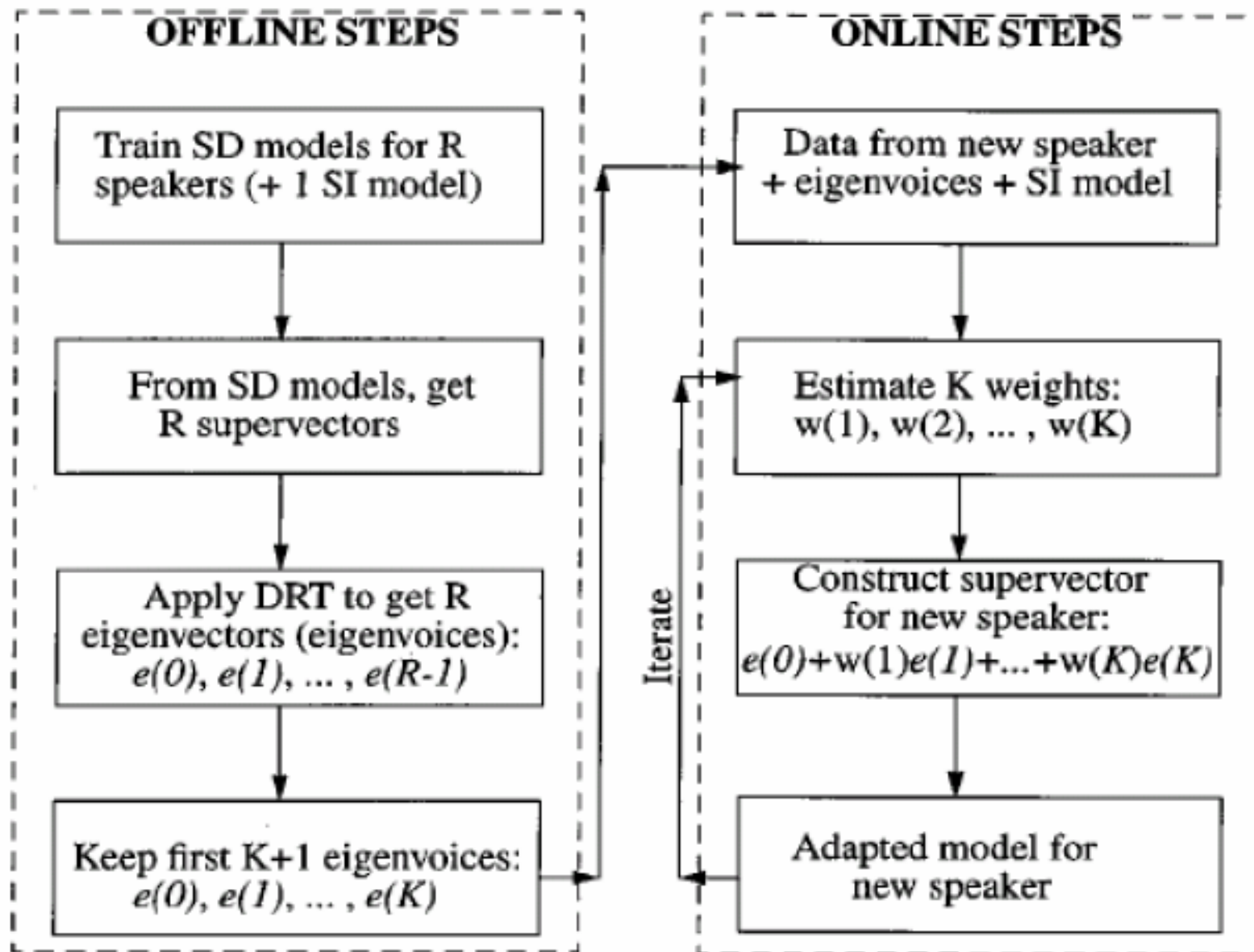
- εκπαίδευση μοντέλων εξαρτημένων από τον ομιλητή
- δημιουργία υπερ-διανυσμάτων ( $\mathbf{a}^{(r)}$ ) παραμέτρων / ομιλητή



- νέος ομιλητής ως σημείο στον χώρο των K-ιδιοφωνών

$$\mathbf{P}_i = \mathbf{e}(0) + w_{i,1} \mathbf{e}(1) + w_{i,2} \mathbf{e}(2) + \dots + w_{i,K} \mathbf{e}(K)$$

# PCA: Ιδιοφωνές και Προσαρμογή Ομιλητή



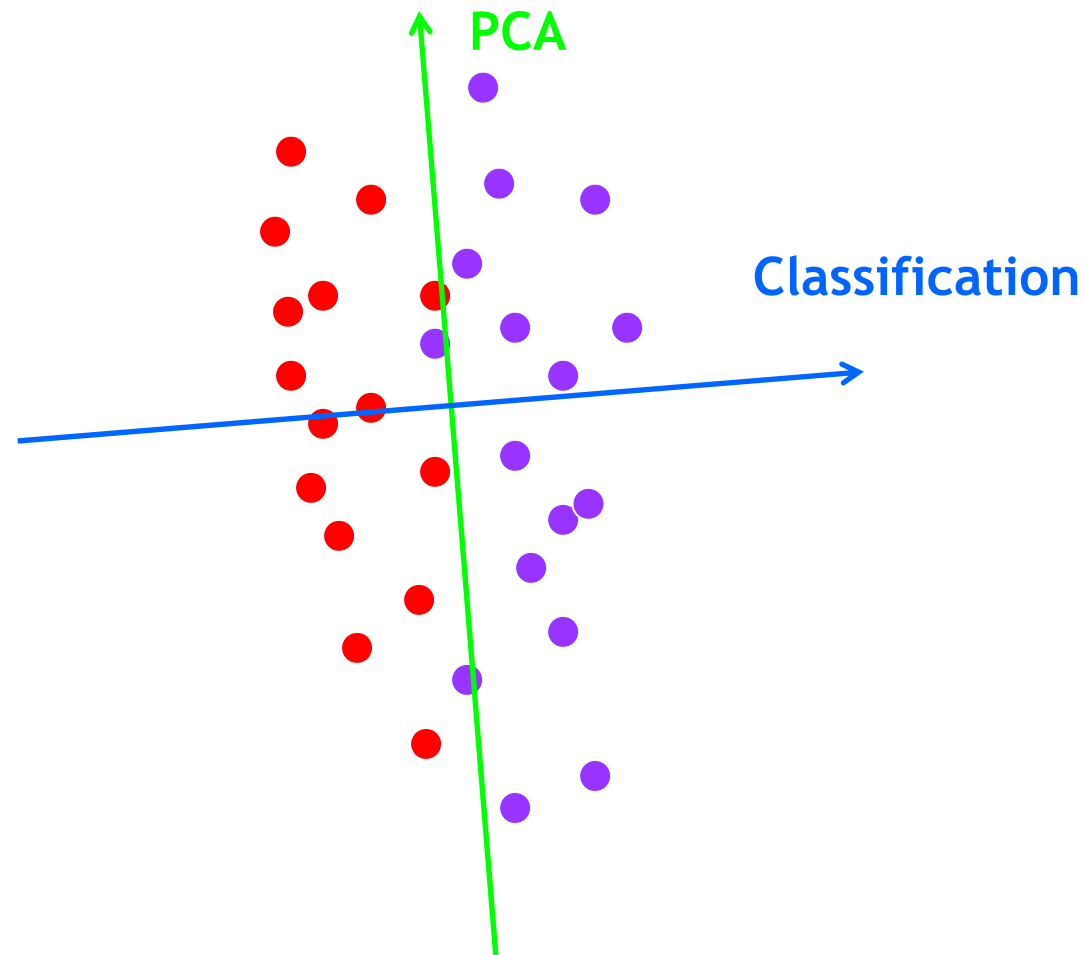
# PCA: ιδιοφωνές και προσαρμογή ομιλητή

## Κύριες Αναφορές:

- R. Kuhn, P. Nguyen, J.C. Junqua, L. Goldwasser, N. Niedzielski, S. Fincke, K. Field, and M. Contolini, “*Eigenvoices for Speaker Adaptation*”, Proc. 5th International Conference on Spoken Language Processing, 1998.
- R. Kuhn, J.C. Junqua, P. Nguyen and N. Niedzielski, “*Rapid Speaker Adaptation in Eigenvoice Space*”, IEEE Transactions on Speech and Audio Processing, 8(6), pp. 695-707, 2000.

# Inherent Limitation

- PCA does not find the best projection for classification
  - The direction of maximum variance is not necessarily useful



# Linear Discriminant Analysis (LDA)

-

Ανάλυση Γραμμικής  
{Διάκρισης ή Διαχωρισμού}

# Linear Discriminant Analysis:

## Περιεχόμενα

- ανάλυση σε υποχώρους: αιτίες
- χαρακτήρας και στόχοι
- αναπαραστάσεις
- διαφορές με ανάλυση πρωτευσών συνιστωσών
- σύγχρονες εφαρμογές

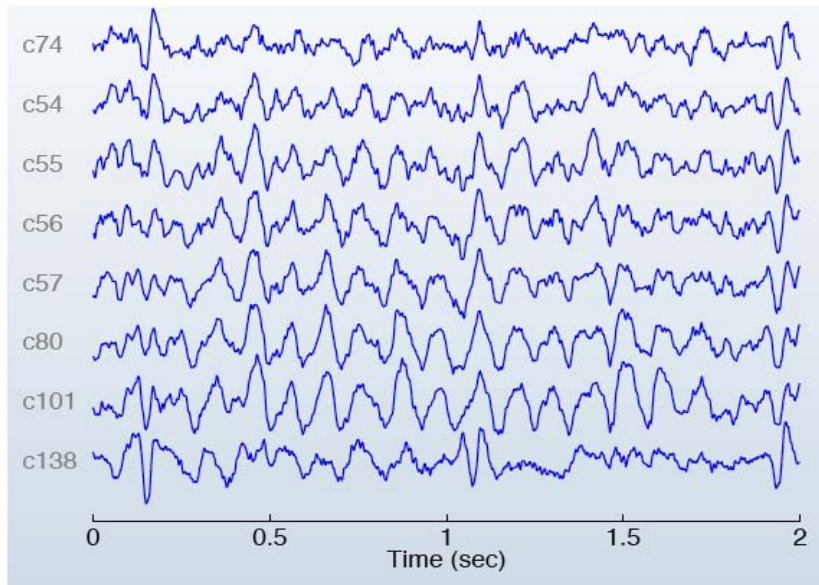
# Πολυδιάστατα Θορυβώδη Δεδομένα !

ΕΙΚΟΝΕΣ ΠΡΟΣΩΠΩΝ

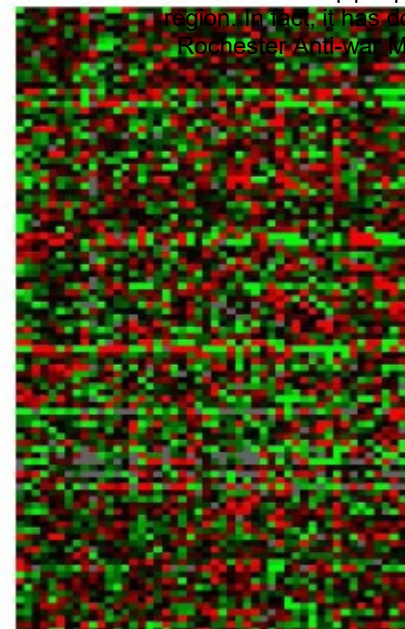


ΚΕΙΜΕΝΟ

Celebrations over the historic election of the first black president have been interspersed with emotional protests against a series of bans on gay marriage, and calls to remember the bigger picture. In California, voters approved a proposition by 52.5 percent to amend the state constitution to ban same sex marriage. Similar bills also passed in Arizona and Florida. Since then, protests against the measure have surged in Los Angeles and San Francisco, and civil rights organizations have banded together to attempt to block the implementation of the measure. On other fronts, animal rights activists are also calling the passage Proposition 2 in California a victory, and an anti-choice measure failed in California. Reports of celebrations over Barak Obama's victory continue to come in from Hollywood, New York, and New Orleans, while activists are working to keep the issues in the forefront: » Rochester, NY — a coalition of anti-war and veterans organizations banded together to make the following statement: "For nearly two years, major polls have shown that a majority of Americans want an end to the war in Iraq. The war has not "brought democracy" to the Middle East. It has not improved the lives of the Iraqi people, nor has it reduced the violence in the region. In fact, it has done exactly the opposite." Read More: Rochester Anti-War Movement Responds to 2008 Election



MEG



γονίδια



# LDA: Ανάλυση σε Υποχώρους

πολυδιάστατα θορυβώδη δεδομένα !

υψηλή διάσταση → αραιά δεδομένα + υπο-εκπαίδευση ...

- Μετασχηματισμός των δεδομένων έτσι ώστε:
  - να αναπαρίστανται με *οικονομία*
  - να διατηρούν τα πιο “*σημαντικά*” εκάστοτε χαρακτηριστικά(-πληροφορία)
- **Σημαντικά** σχετικά με την εφαρμογή, π.χ. :
  - αποθορυβοποίηση
  - ανάλυση σήματος
  - οπτικοποίηση και κατανόηση δεδομένων
  - επιλογή ή εξαγωγή χαρακτηριστικών
  - κατηγοριοποίηση
  - γενίκευση μοντέλου
  - συμπίεση

# LDA: Ορολογία και Ιστορία

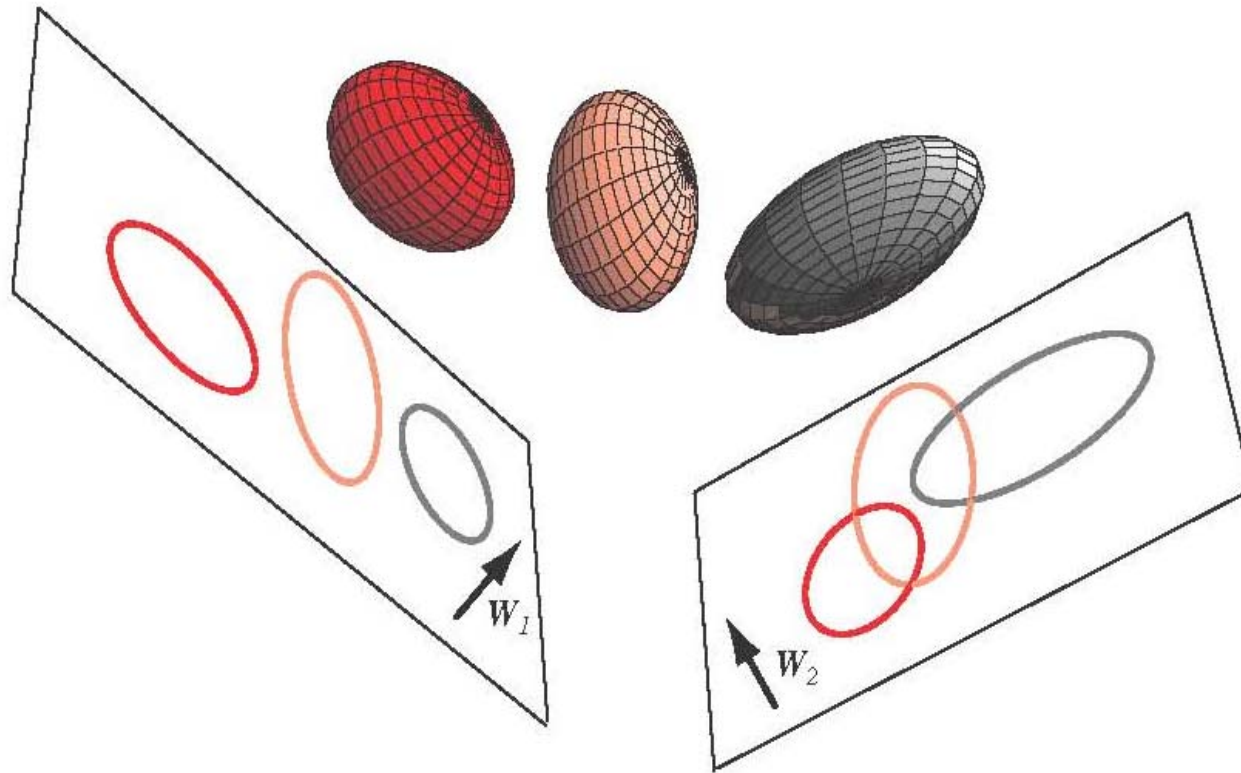
- **Fisher's Linear Discriminant Analysis**
- **Fisher-Rao Linear Discriminant Analysis**
- **Fisher (1936):** εισαγωγή μεθόδου για 2 τάξεις
- **Rao (1965):** επέκταση για πολλαπλές κατηγορίες
- **Πληθώρα σύγχρονων εκδοχών και επεκτάσεων :**
  - **Penalized Discriminant Analysis, Hastie 1995,**
  - **Generalized Discriminant Analysis, Baudat 2000.**

# Linear Discriminant Analysis:

## Ανάλυση Γραμμικής Διάκρισης: χαρακτηρισ & στόχοι

- **Μείωση διάστασης** διατηρώντας όσο το δυνατόν περισσότερο την διακριτική ικανότητα μεταξύ των τάξεων.
- **Εύρεση ιδιοδιανυσμάτων** κατά τις διευθύνσεις των οποίων οι τάξεις διαχωρίζονται καλύτερα.
- Λαμβάνει υπόψιν τις **αποστάσεις** τόσο **μεταξύ** των τάξεων όσο και **εσωτερικά** των τάξεων.
- Μείωση διάστασης δεδομένων **υπό επίβλεψη**, δηλ. με γνώση των κατηγοριών/τάξεων για κάθε δεδομένο.
- For the example of **face recognition**, LDA should be more capable of distinguishing image variation due to *identity* from variation due to other sources such as *illumination* and *expression*.

# LDA: Προβολή σε Υποχώρους



**FIGURE 3.6.** Three three-dimensional distributions are projected onto two-dimensional subspaces, described by a normal vectors  $W_1$  and  $W_2$ . Informally, multiple discriminant methods seek the optimum such subspace, that is, the one with the greatest separation of the projected distributions for a given total within-scatter matrix, here as associated with  $W_1$ . From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

**μεγιστοποίηση διάκρισης μεταξύ τάξεων**

# LDA: Φορμαλισμός (2 κατηγοριες)

- Κατηγορίες/Ταξεις:  $D_1, D_2$  με στοιχεία πληθους  $n_1, n_2$

- Μεσοι ταξεων:

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}, \quad i = 1, 2$$

- Πινακες Διασπορας ταξεων:

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T, \quad i = 1, 2$$

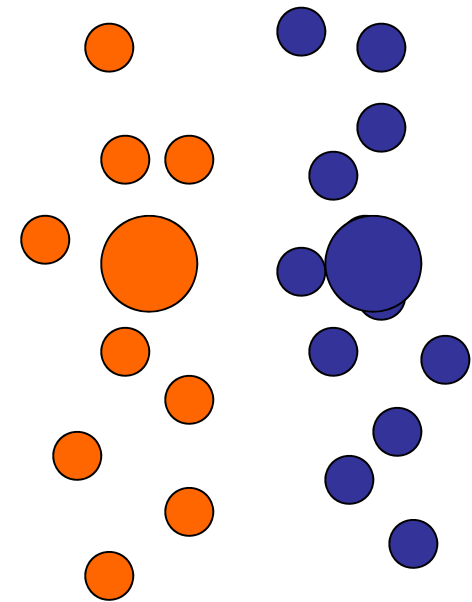
- Προβολη:  $\mathbf{y} = \mathbf{w}^T \mathbf{x}$

- Μεσοι προβολων:

$$\tilde{\mathbf{m}}_i = \mathbf{w}^T \mathbf{m}_i, \quad i = 1, 2$$

- Διασπορες προβολων:

$$(\tilde{S}_i)^2 = \sum_{\mathbf{x} \in D_i} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_i)^2 = \mathbf{w}^T \mathbf{S}_i \mathbf{w}, \quad i = 1, 2$$



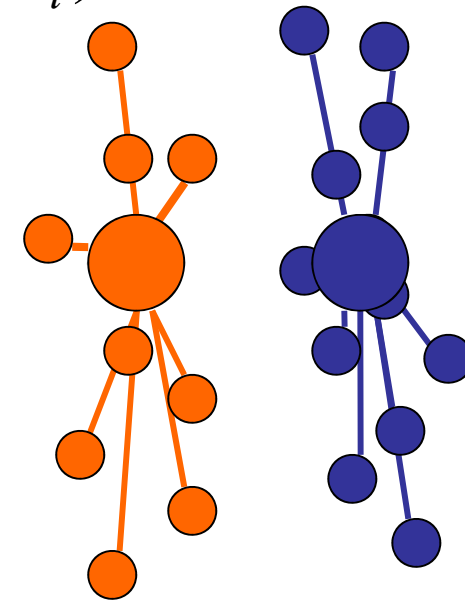
# LDA: Ενδοταξική Διασπορά

- Πίνακας Ενδοταξικής Διασποράς Τάξεων  
(**Within-class Scatter Matrix**)

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2 = \sum_{i=1}^2 \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

- Ενδοταξική Διασπορά Προβολών

$$\begin{aligned} (\tilde{S}_1)^2 + (\tilde{S}_2)^2 &= \sum_{i=1}^2 \sum_{\mathbf{x} \in D_i} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_i)^2 \\ &= \mathbf{w}^T \mathbf{S}_W \mathbf{w} \end{aligned}$$



— Within-class distance

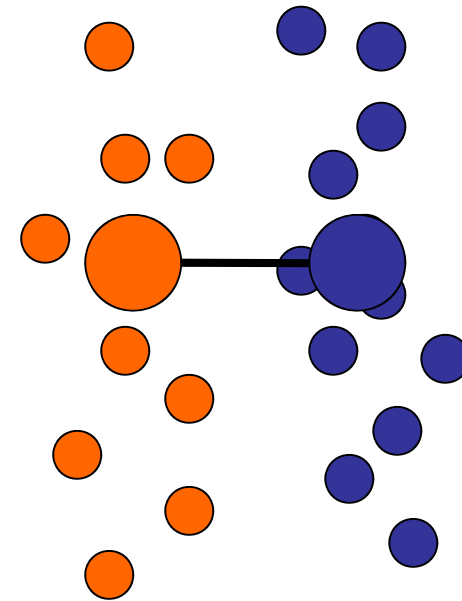
# LDA: Διαταξική Διασπορά

- Πίνακας Διαταξικής Διασποράς Τάξεων

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$

- Διαταξική Διασπορά Προβολών

$$(\tilde{m}_1 - \tilde{m}_2)^2 = \mathbf{w}^T \mathbf{S}_B \mathbf{w}$$



— Between-class distance

# LDA: Βέλτιστη λύση και μέγιστος διαχωρισμός 2 τάξεων

- Κριτήριο  $J$ : μεγιστοποίηση διαταξικών αποστάσεων - ελαχιστοποίηση ενδοταξικών αποστάσεων
- Προβολή/μετασχηματισμός αρχικών δεδομένων :  $\mathbf{y} = \mathbf{W}^T \mathbf{X}$
- Εύρεση μετασχηματισμού (διανύσματος)  $\mathbf{w}$ :

$$\operatorname{argmax}_{\mathbf{w}} J(\mathbf{w}) = \frac{(\tilde{m}_1 - \tilde{m}_2)^2}{(\tilde{s}_1)^2 + (\tilde{s}_2)^2} = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}}$$

- βέλτιστη λύση από πρόβλημα γενικευμένων ιδιοτιμών:

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}, \quad \left( \mathbf{S}_W^{-1} \mathbf{S}_B \mathbf{w} = \lambda \mathbf{w} \right)$$

- Βέλτιστη Λυση (2 ταξεις):

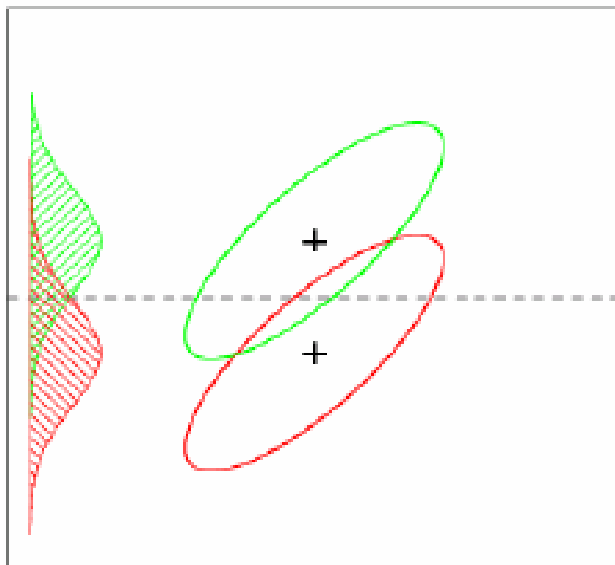
$$\mathbf{w} = \mathbf{S}_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2)$$



# LDA: διαφορά από ανάλυση πρωτεύουσων συνιστωσών

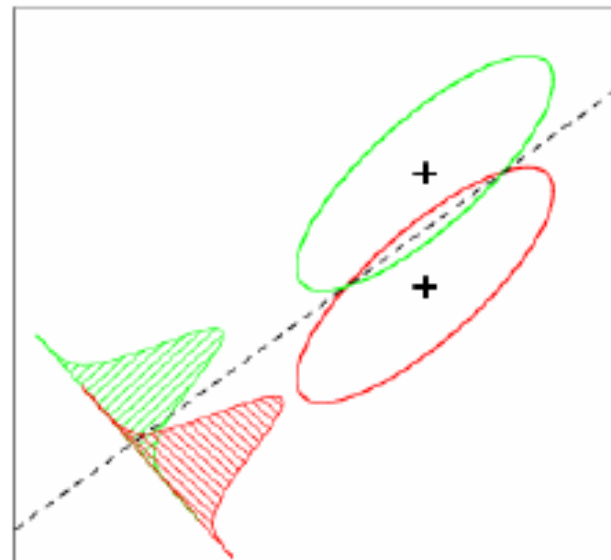
- μείωση διάστασης διατηρώντας όσο το δυνατόν την μεταβλητότητα των δεδομένων
- μέθοδος χωρίς επίβλεψη

PCA



- **ιδιοκατεύθυνση  
μεγιστοποίησης  
μεταβλητότητας**

LDA



- **ιδιοκατεύθυνση  
μεγιστοποίησης διάκρισης  
μεταξύ τάξεων**

# Linear Discriminant Analysis (C classes)

Within-class Scatter matrix  $\mathbf{S}_w = \sum_{i=1}^C \sum_{j=1}^{n_i} (\mathbf{x}_j - \mathbf{m}_i)(\mathbf{x}_j - \mathbf{m}_i)^T$

Between-class Scatter matrix  $\mathbf{S}_b = \sum_{i=1}^C n_i (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T$

**projection**  $\mathbf{y} = \mathbf{W}^T \mathbf{X}$

C-1 linear discriminants:  $y_i = \mathbf{w}_i^T \mathbf{X}$

- LDA computes a transformation  $\mathbf{W}$  that maximizes the between-class scatter while minimizing the within-class scatter:

$$\max \frac{|\tilde{\mathbf{S}}_b|}{|\tilde{\mathbf{S}}_w|} = \max \frac{|\mathbf{W}^T \mathbf{S}_b \mathbf{W}|}{|\mathbf{W}^T \mathbf{S}_w \mathbf{W}|}$$

products of eigenvalues = product of variances in principal directions

$\mathbf{S}_b \mathbf{w}_i = \lambda_i \mathbf{S}_w \mathbf{w}_i, \quad \mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{C-1}]$

$\tilde{\mathbf{S}}_b, \tilde{\mathbf{S}}_w$ : scatter matrices of the projected data  $\mathbf{y}$

# LDA (C classes)

- Does  $\mathbf{S}_w^{-1}$  always exist?
  - If  $\mathbf{S}_w$  is non-singular, we can obtain a conventional eigenvalue problem by writing:

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{w}_i = \lambda_i \mathbf{w}_i$$

- In practice,  $\mathbf{S}_w$  may often be singular (e.g. as in images) if the data vectors' dimensionality  $D$  is larger than the size  $N$  of the data set. (small sample size)
- Since  $\mathbf{S}_b$  has at most rank  $C-1$ , the max number of eigenvectors with non-zero eigenvalues is  $C-1$  (i.e., max dimensionality of subspace is  $C-1$ )

# LDA combined with PCA (C classes)

- Does  $\mathbf{S}_w^{-1}$  always exist? – cont.
  - To alleviate this problem, we can use PCA first:
    - 1) PCA is first applied to the data set to reduce its dimensionality.

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \dashrightarrow PCA \dashrightarrow \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_K \end{bmatrix}$$

- 2) LDA is then applied to find the most discriminative directions:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_K \end{bmatrix} \dashrightarrow LDA \dashrightarrow \begin{bmatrix} z_1 \\ z_2 \\ \dots \\ z_{C-1} \end{bmatrix}$$

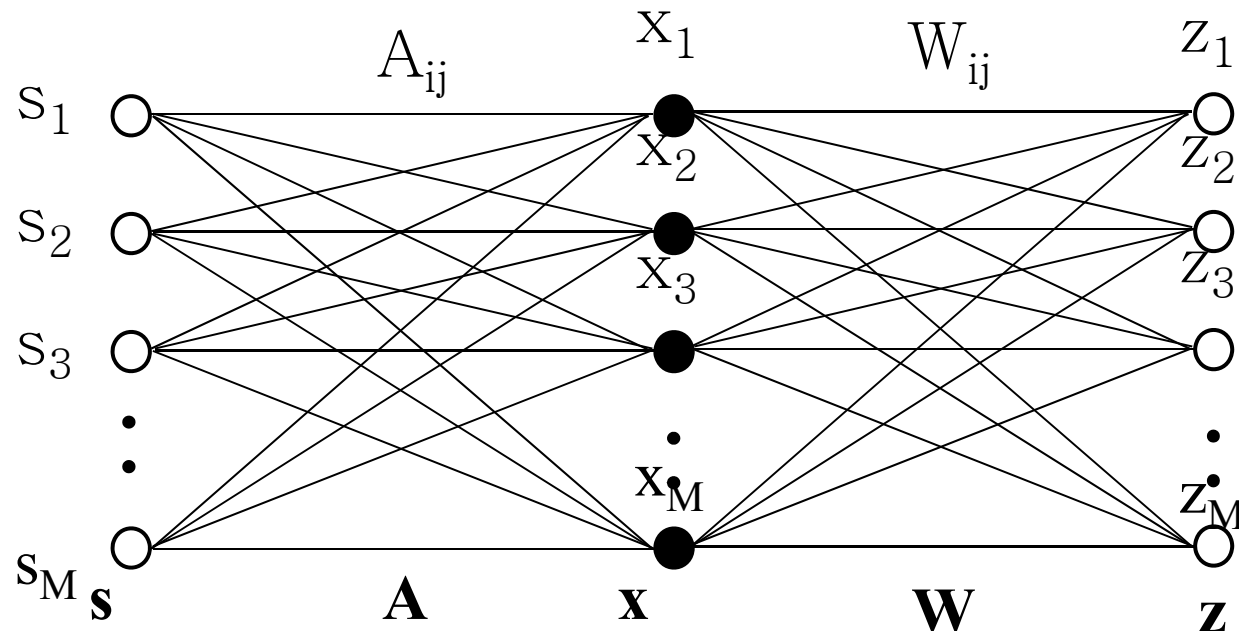
**Ανάλυση Ανεξάρτητων  
Συνιστωσών**

**Independent Component  
Analysis  
(ICA)**

# Independent Component Analysis

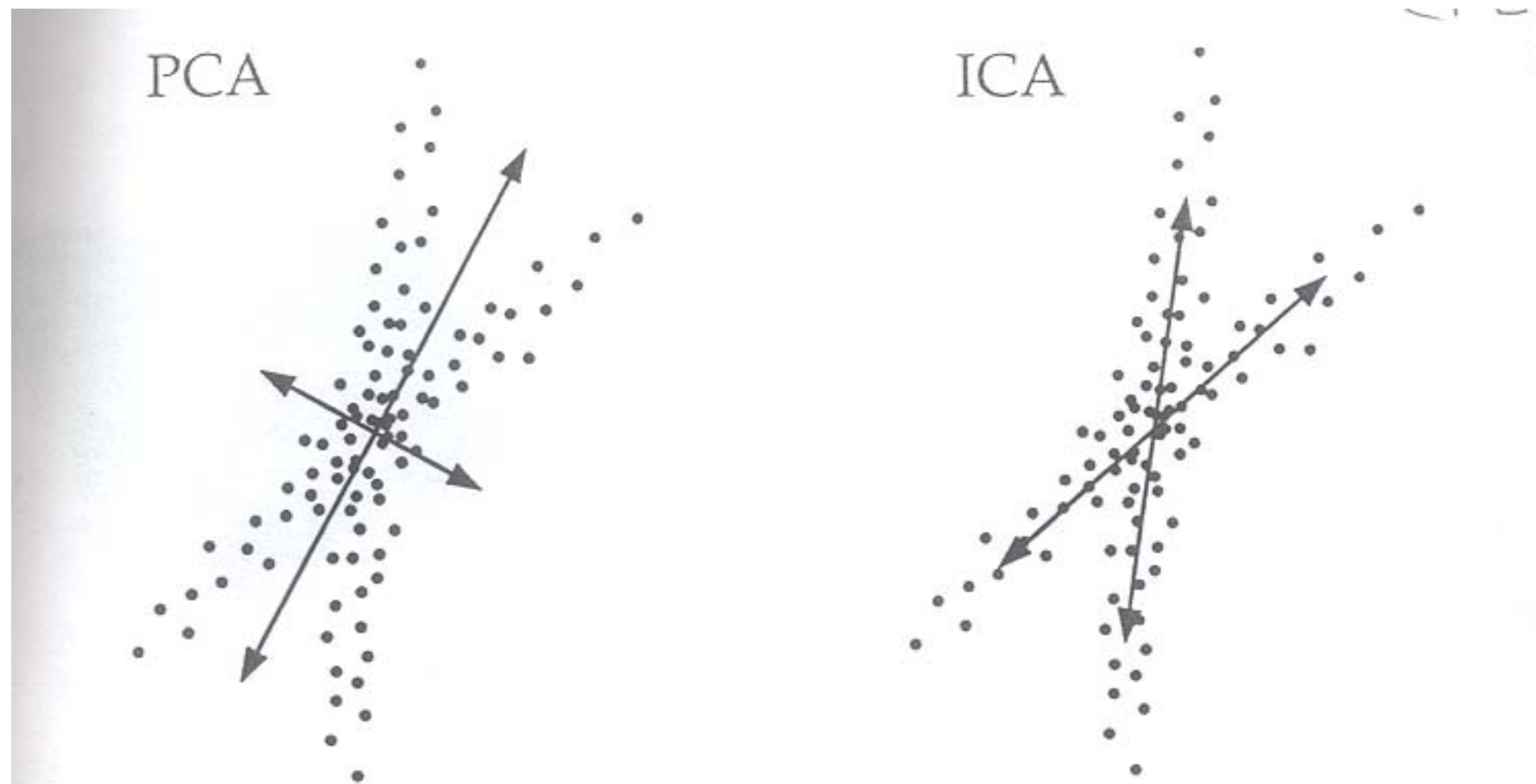
- **Concept of ICA**

- A given signal ( $\mathbf{x}$ ) is generated by linear mixing ( $\mathbf{A}$ ) of independent components ( $\mathbf{s}$ )
- ICA is a statistical analysis method to estimate those independent components ( $\mathbf{z}$ ) and mixing rule ( $\mathbf{W}$ )



$$\mathbf{z} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}$$

We do not know  
both unknowns  
→ Some optimization  
function is required



*Figure 2.3.* Example 2-D data distribution and the corresponding principal component and independent component axes. The data points could be, for example, grayvalues at pixel 1 and pixel 2. Figure inspired by Lewicki & Sejnowski (2000).

## ICA - Problem formulation (1)

### Mixture Components

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t) + a_{13}s_3(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t) + a_{23}s_3(t)$$

$$x_3(t) = a_{31}s_1(t) + a_{32}s_2(t) + a_{33}s_3(t)$$

### Independent Components

$$s_1(t) = w_{11}x_1(t) + w_{12}x_2(t) + w_{13}x_3(t)$$

$$s_2(t) = w_{21}x_1(t) + w_{22}x_2(t) + w_{23}x_3(t)$$

$$s_3(t) = w_{31}x_1(t) + w_{32}x_2(t) + w_{33}x_3(t)$$



## ICA - Problem formulation (2)

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = \mathbf{A} \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{pmatrix} \quad \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \mathbf{W} \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{pmatrix}$$

### **Ambiguities:**

- Cannot determine variances (energies) of independent components (assume unit variance).
- Ambiguity of sign: +/- 1.
- Cannot determine the order of independent components .

## Independence vs Uncorrelatedness

**Independent:**  $p(y_1, y_2) = p_1(y_1)p_2(y_2).$

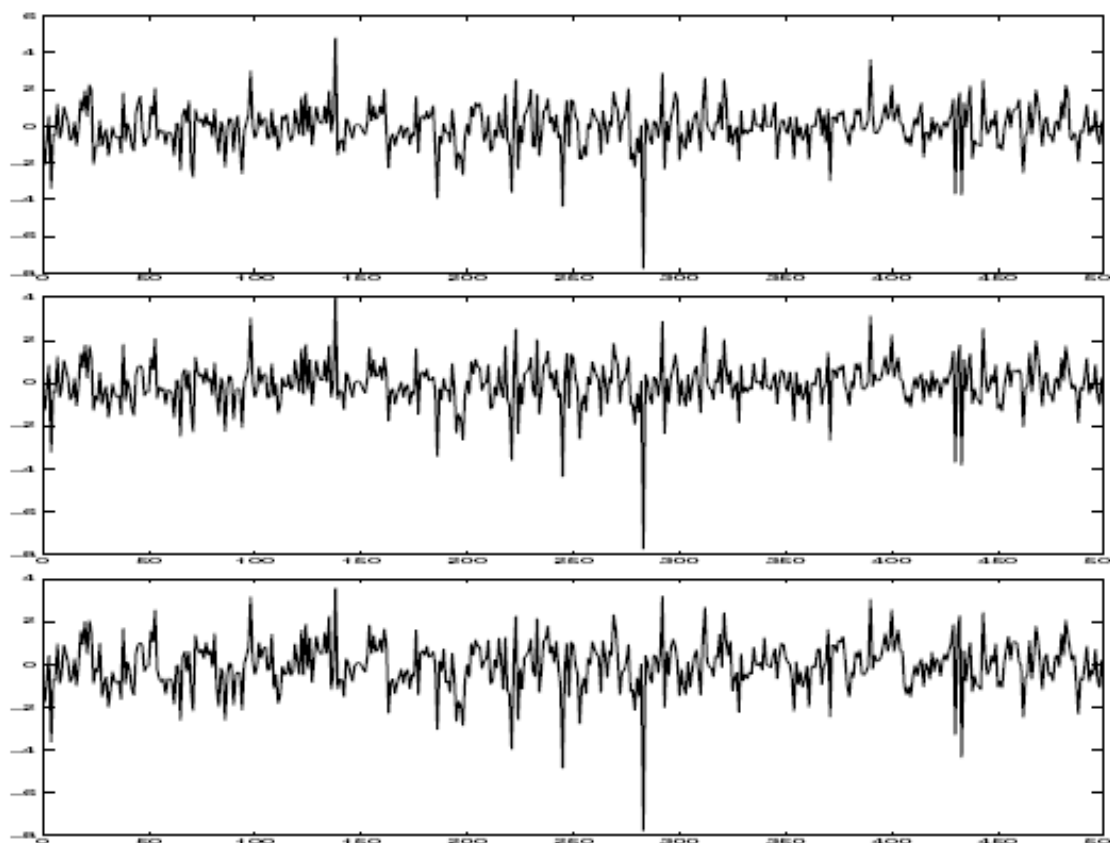
$$p_1(y_1) = \int p(y_1, y_2) dy_2$$

Necessary Condition of Independence:

$$E\{h_1(y_1)h_2(y_2)\} = E\{h_1(y_1)\}E\{h_2(y_2)\}$$

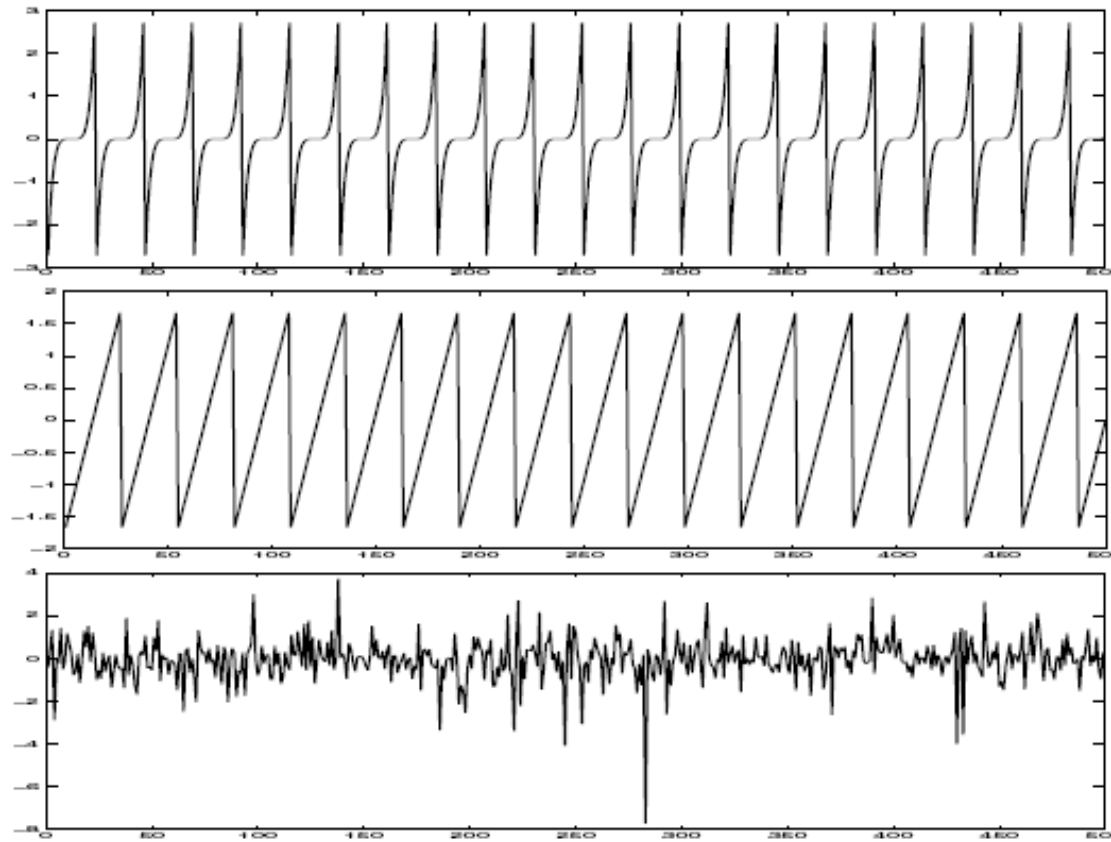
**Uncorrelated:**  $E\{y_1y_2\} - E\{y_1\}E\{y_2\} = 0$

# Mixture signals



**Fig. 1.2** The observed signals that are assumed to be mixtures of some underlying source signals.

# Independent Components



**Fig. 1.3** The estimates of the original source signals, estimated using only the observed mixture signals in Fig. 1.2. The original signals were found very accurately.

## Joint Distribution of Two Independent Components

$$p(s_1, s_2)$$

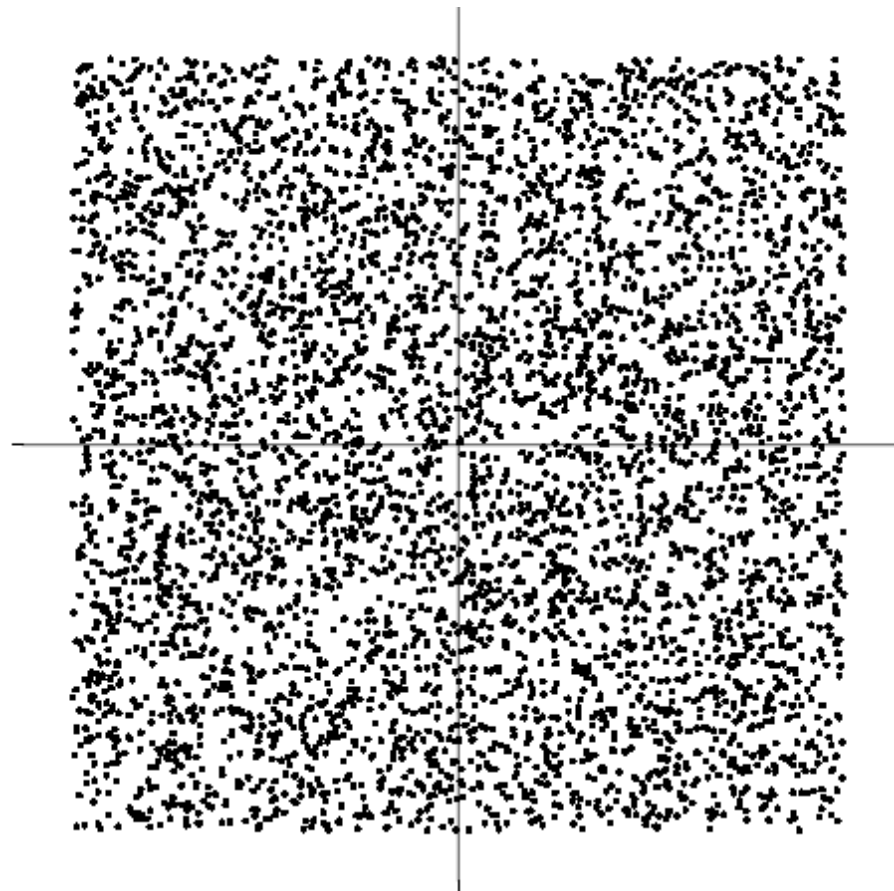


Figure 5: The joint distribution of the independent components  $s_1$  and  $s_2$  with uniform distributions. Horizontal axis:  $s_1$ , vertical axis:  $s_2$ .

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{if } |s_i| \leq \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

## Joint Distribution of Mix-2 Components

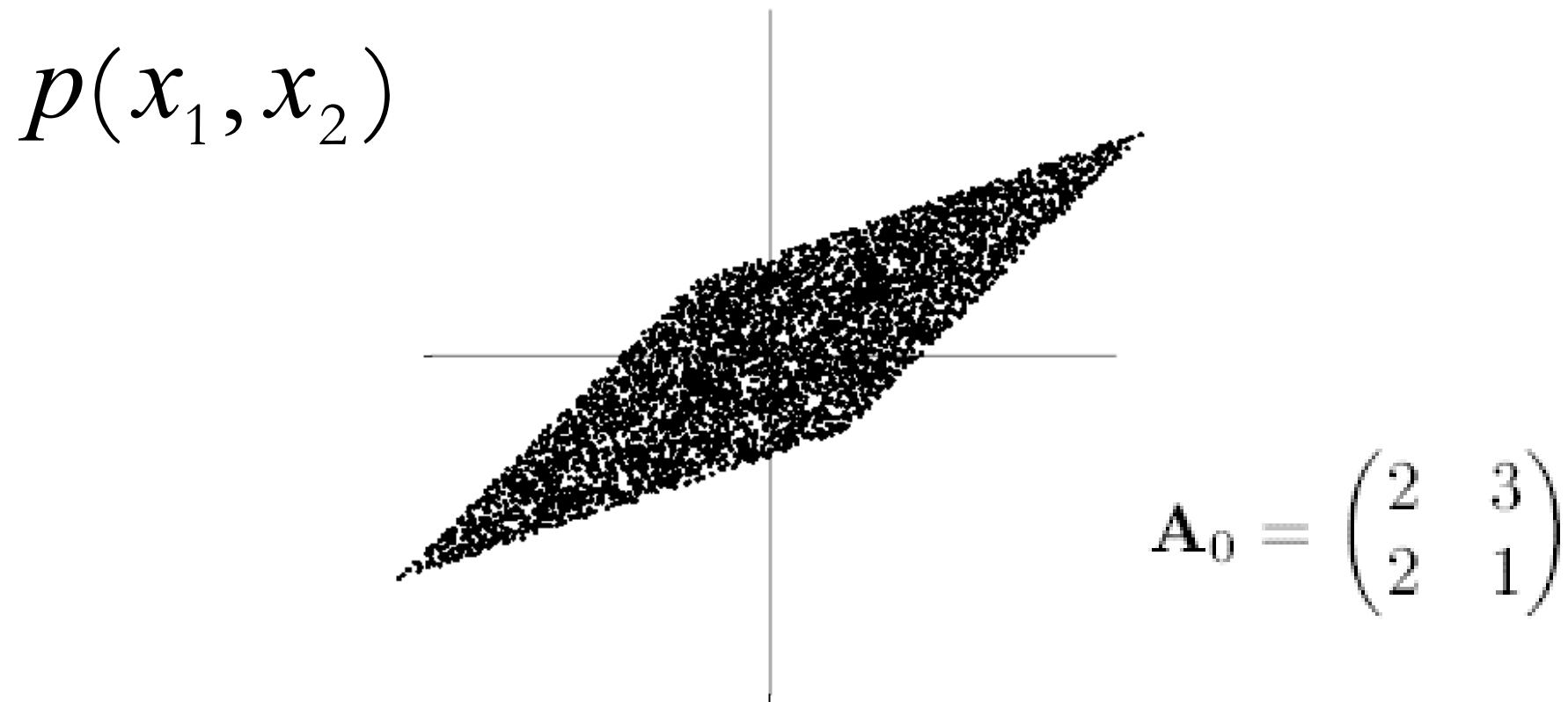
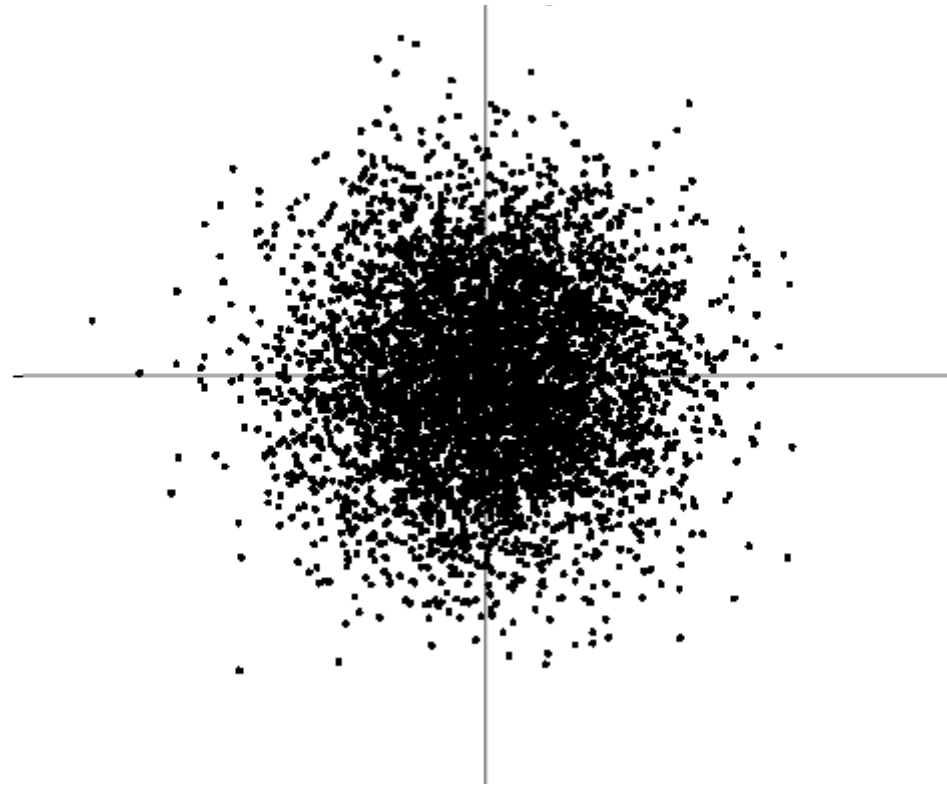


Figure 6: The joint distribution of the observed mixtures  $x_1$  and  $x_2$ . Horizontal axis:  $x_1$ , vertical axis:  $x_2$ .

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}} & \text{if } |s_i| \leq \sqrt{3} \\ 0 & \text{otherwise} \end{cases}$$

## Joint Distribution of Two Independent Gaussian R.V.

The multivariate distribution of two independent gaussian variables.



If mixing matrix  $\mathbf{A}$  is orthogonal, then the mixed components are gaussian, uncorrelated and of unit variance

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{x_1^2 + x_2^2}{2}\right)$$

# Higher – order Statistics

## First Characteristic Fcn

$$\Phi(j\omega) = E\{\exp(j\omega x)\} = \int_{-\infty}^{\infty} \exp(j\omega x) p_x(x) dx$$

## Moments

$$\Phi(s) = \int_{-\infty}^{\infty} \left( \sum_{n=0}^{\infty} \frac{x^n s^n}{n!} \right) p_x(x) dx = \sum_{n=0}^{\infty} E\{x^n\} \frac{s^n}{n!}$$

## Second Characteristic Fcn

$$\Psi(s) = \log \Phi(s) = \log(E\{\exp(sx)\})$$

## Cumulants

$$\Psi(s) = \sum_{n=0}^{\infty} \kappa_n \frac{s^n}{n!} \quad \kappa_n = \left. \frac{d^n \Psi(s)}{ds^n} \right|_{s=0}$$



# Cumulants

$$\kappa_1 = \mathbf{E}\{x\}$$

$$\kappa_2 = \mathbf{E}\{x^2\} - [\mathbf{E}\{x\}]^2$$

$$\kappa_3 = \mathbf{E}\{x^3\} - 3\mathbf{E}\{x^2\}\mathbf{E}\{x\} + 2[\mathbf{E}\{x\}]^3$$

$$\kappa_4 = \mathbf{E}\{x^4\} - 3[\mathbf{E}\{x^2\}]^2 - 4\mathbf{E}\{x^3\}\mathbf{E}\{x\} + 12\mathbf{E}\{x^2\}[\mathbf{E}\{x\}]^2 - 6[\mathbf{E}\{x\}]^4$$

## Cross-cumulants (zero-mean r.v.)

$$\text{cum}(x_i, x_j) = \mathbf{E}\{x_i x_j\}$$

$$\text{cum}(x_i, x_j, x_k) = \mathbf{E}\{x_i x_j x_k\}$$

$$\begin{aligned} \text{cum}(x_i, x_j, x_k, x_l) &= \mathbf{E}\{x_i x_j x_k x_l\} - \mathbf{E}\{x_i x_j\}\mathbf{E}\{x_k x_l\} \\ &\quad - \mathbf{E}\{x_i x_k\}\mathbf{E}\{x_j x_l\} - \mathbf{E}\{x_i x_l\}\mathbf{E}\{x_j x_k\} \end{aligned}$$

**ICA: all higher order cross-cumulants = 0**

## Kurtosis

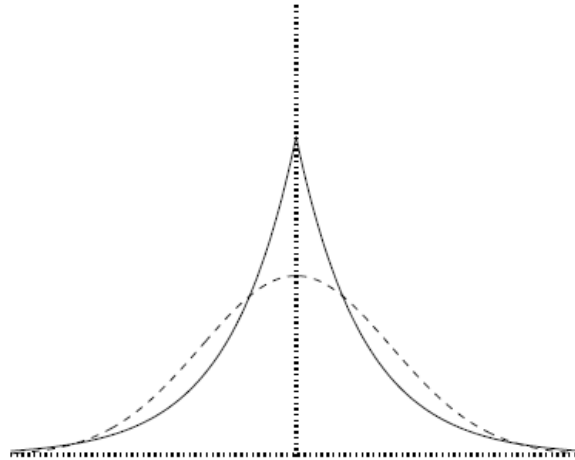
$$\text{kurt}(y) = E\{y^4\} - 3(E\{y^2\})^2$$

Independent:  $\text{kurt}(x_1 + x_2) = \text{kurt}(x_1) + \text{kurt}(x_2)$

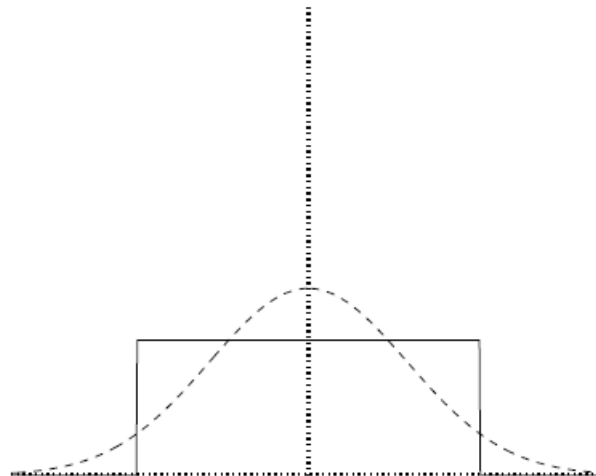
$$\text{kurt}(\alpha x_1) = \alpha^4 \text{kurt}(x_1)$$

$$\text{kurt}(y) \begin{cases} > 0, & \text{super-gaussian} \\ = 0, & y \sim \text{Gaussian} \\ < 0, & \text{sub-gaussian} \end{cases}$$

# Supergaussian and Subgaussian Distributions



**Fig. 8.9** The density function of the Laplacian distribution, which is a typical supergaussian distribution. For comparison, the gaussian density is given by a dashed curve. Both densities



**Fig. 8.10** The density function of the uniform distribution, which is a typical *subgaussian* distribution. For comparison, the gaussian density is given by a dashed line. Both densities are normalized to unit variance.

## Negentropy $J(\cdot)$

$$H(\mathbf{y}) = - \int p_y(\boldsymbol{\eta}) \log p_y(\boldsymbol{\eta}) d\boldsymbol{\eta}$$

$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$$

**Approximation - I: Higher Moments:**

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \text{kurt}(y)^2$$

**Approximation - II : non-quadratic  $G$ , standardized gaussian  $\nu$**

$$J(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2$$

$$G_1(y) = \frac{1}{a_1} \log \cosh a_1 y,$$

$$G_2(y) = - \exp(-y^2/2)$$

## Mutual Information $I(\cdot)$

$$I(\mathbf{y}) = \sum_i H(y_i) - H(\mathbf{y}) = \text{KL}(p(\mathbf{y}) \parallel \prod_i p_i(y_i))$$

$$\mathbf{y} = \mathbf{W}\mathbf{x} \Rightarrow$$

$$I(y_1, y_2, \dots, y_n) = \sum_i H(y_i) - H(\mathbf{x}) - \log |\det(\mathbf{W})|$$

$$E\{\mathbf{y}\mathbf{y}^T\} = \mathbf{I} = \mathbf{W}E\{\mathbf{x}\mathbf{x}^T\}\mathbf{W}^T \Rightarrow \det(\mathbf{W}) = \text{const}$$

$$I(y_1, y_2, \dots, y_n) = \text{constant} - \sum_i J(y_i)$$

Min M.I.  $\leftrightarrow$  Max sum of non-gaussianities of the estimates when the estimates are constrained to be uncorrelated.

## ICA Βιβλιογραφία

- A. Hyvarinen and E. Oja, “Independent Component Analysis: A Tutorial”, 1999.
- A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis*, Wiley, 2001.
- S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Acad. Press

**Ανάλυση Κανονικών  
Συσχετίσεων**

**Canonical Correlation Analysis  
(CCA)**

## Βασικά Χαρακτηριστικά Ανάλυσης CCA

- Ζεύγος τυχαίων διανυσμάτων:
  - $\mathbf{x} \in R^m$  και  $\mathbf{y} \in R^n$  με μηδενική μέση τιμή.
  - Γνώση ροπών 2<sup>ης</sup> τάξης:  
$$R_{xx} = E[\mathbf{x}\mathbf{x}^T], R_{yy} = E[\mathbf{y}\mathbf{y}^T], R_{xy} = E[\mathbf{x}\mathbf{y}^T]$$
- Απαντήσεις σε δύο συμπληρωματικά ερωτήματα:
  - Σε τι βαθμό είναι τα δύο σύνολα δεδομένων γραμμικά συσχετισμένα;
  - Πώς μπορούμε να μειώσουμε τη διάσταση των χώρων διατηρώντας παράλληλα τις μεταξύ τους συσχετίσεις;
- Σχέση μεταξύ CCA και PCA/LDA:
  - Μείωση διάστασης παρόμοια με PCA αλλά για την περίπτωση ζεύγους μεταβλητών.
  - Επιβλεπόμενη μέθοδος όπως η LDA.



## Φορμαλισμός CCA

- **Εύρεση προβολών που μεγιστοποιούν το συντελεστή συσχέτισης:**

- $\eta = \mathbf{a}^T \mathbf{x}, \phi = \mathbf{b}^T \mathbf{y}$ , με  $\mathbf{a} \in R^m, \mathbf{b} \in R^n$

- **Συντελεστής συσχέτισης:**

$$\rho(\mathbf{a}, \mathbf{b}) = \frac{E[\eta\phi]}{\sqrt{E[\eta^2]}\sqrt{E[\phi^2]}} = \frac{\mathbf{a}^T R_{xy} \mathbf{b}}{\sqrt{\mathbf{a}^T R_{xx} \mathbf{a}} \sqrt{\mathbf{b}^T R_{yy} \mathbf{b}}}$$

- **Πρόβλημα βελτιστοποίησης:**

$$\rho_i = \max \rho(\mathbf{a}, \mathbf{b}) \quad \text{s.t.} \quad \mathbf{a}_j^T R_{xx} \mathbf{a} = \mathbf{b}_j^T R_{yy} \mathbf{b} = 0, j = 1, \dots, i-1.$$

- **Λύση:**

- **Ζεύγη διανυσμάτων προβολών (κατευθύνσεις CCA):**

$$[\mathbf{a}_1, \dots, \mathbf{a}_r] \quad [\mathbf{b}_1, \dots, \mathbf{b}_r], \quad \rho_i = \rho(\mathbf{a}_i, \mathbf{b}_i)$$

- **Φθίνουσα ακολουθία συντελεστών συσχέτισης:**

$$1 \geq \rho_1 \geq \dots \geq \rho_r > 0, \quad r = \text{rank}(R_{xy}) \leq \min(m, n)$$

## Υπολογισμός Κατευθύνσεων μέσω Ανάλυσης SVD

- **Αλλαγή συστήματος συντεταγμένων:**

$$\boldsymbol{\alpha} = R_{xx}^{1/2} \mathbf{a}, \quad \boldsymbol{\beta} = R_{yy}^{1/2} \mathbf{b}, \quad C_{xy} = R_{xx}^{-1/2} R_{xy} R_{yy}^{-1/2}$$

$$\rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\boldsymbol{\alpha}^T C_{xy} \boldsymbol{\beta}}{\sqrt{(\boldsymbol{\alpha}^T \boldsymbol{\alpha})(\boldsymbol{\beta}^T \boldsymbol{\beta})}} = \frac{\boldsymbol{\alpha}^T C_{xy} \boldsymbol{\beta}}{\|\boldsymbol{\alpha}\| \|\boldsymbol{\beta}\|}$$

- **Μετασχηματισμένο πρόβλημα:**

$$\rho_i = \max \rho(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \max \boldsymbol{\alpha}^T C_{xy} \boldsymbol{\beta}$$

s.t.  $\|\boldsymbol{\alpha}\| = \|\boldsymbol{\beta}\| = 1$  and  $\boldsymbol{\alpha}_j^T \boldsymbol{\alpha} = \boldsymbol{\beta}_j^T \boldsymbol{\beta} = 0, j = 1, \dots, i-1.$

- **Λύση με τη βοήθεια της SVD του πίνακα συνεκτικότητας:**

$$C_{xy} = \sum_{k=1}^r \rho_k \boldsymbol{\alpha}_k \boldsymbol{\beta}_k^T, \quad \boldsymbol{\alpha}_k^T \boldsymbol{\alpha}_j = \boldsymbol{\beta}_k^T \boldsymbol{\beta}_j = \delta_{kj}, \quad \rho_k \geq \rho_{k+1} > 0$$

– Στο  $i$ -στο βήμα έχουμε:

Cauchy-Schwartz

Τετραγωνική μορφή, ελλειψοειδές

$$\boldsymbol{\alpha}^T C_{xy} \boldsymbol{\beta} = \sum_{k=i}^r \rho_k (\boldsymbol{\alpha}^T \boldsymbol{\alpha}_k) (\boldsymbol{\beta}_k^T \boldsymbol{\beta}) = \sum_{k=i}^r (\rho_k \alpha_k) \beta_k \leq \sqrt{\sum_{k=i}^r \rho_k^2 \alpha_k^2} \leq \rho_i$$

with  $\boldsymbol{\alpha} = \boldsymbol{\alpha}_i, \boldsymbol{\beta} = \boldsymbol{\beta}_i,$  thus  $\mathbf{a}_i = R_{xx}^{-1/2} \boldsymbol{\alpha}_i, \mathbf{b}_i = R_{yy}^{-1/2} \boldsymbol{\beta}_i$

## Ιδιότητες CCA

- Αμετάβλητο συντελεστών κανονικής συσχέτισης ως προς γραμμική αντιστρέψιμη αλλαγή συστήματος συντεταγμένων:

$$\text{If } \mathbf{x}' = U^T \mathbf{x}_i, \mathbf{y}' = V^T \mathbf{y}, \text{ then } \rho'_i = \rho_i, \mathbf{a}'_i = U^{-1} \mathbf{a}_i, \mathbf{b}'_i = V^{-1} \mathbf{b}_i$$

- Συμμετρία:

$$[\mathbf{a}, \mathbf{b}] = \text{CCA}(\mathbf{x}, \mathbf{y}) \quad \equiv \quad [\mathbf{b}, \mathbf{a}] = \text{CCA}(\mathbf{y}, \mathbf{x})$$

- Από κοινού πληροφορία (για Gaussian κατανομές):

$$I(\mathbf{x}; \mathbf{y}) = -\frac{1}{2} \sum_{i=1}^r \log(1 - \rho_i^2) \quad (\text{bits})$$

- MSE-βέλτιστη απεικόνιση  $\mathbf{y} = W\mathbf{x}$  (φίλτρο Wiener):

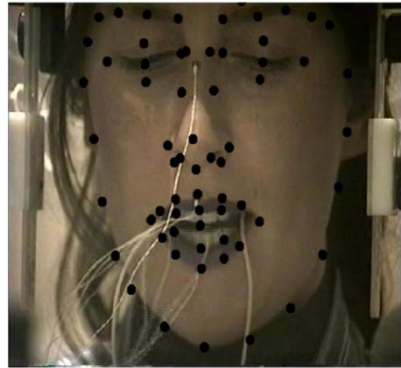
$$W = R_{yx} R_{xx}^{-1} = R_{yy} B P A^T$$

$$A = [\mathbf{a}_1 \dots \mathbf{a}_r], B = [\mathbf{b}_1 \dots \mathbf{b}_r], P = \text{diag}(\rho_1, \dots, \rho_r)$$

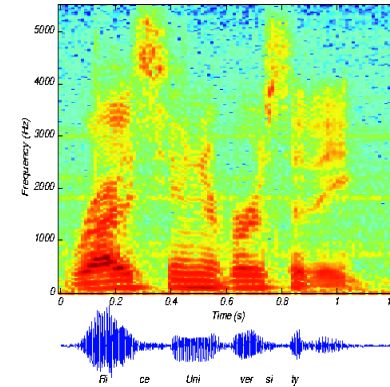
## Εφαρμογές CCA: Διατροφική Πρόβλεψη

- **Παράδειγμα:** Τα  $x$  και  $y$  αντιστοιχούν σε οπτικές και ακουστικές μετρήσεις που λαμβάνονται ταυτόχρονα από έναν ομιλητή:

$x=$



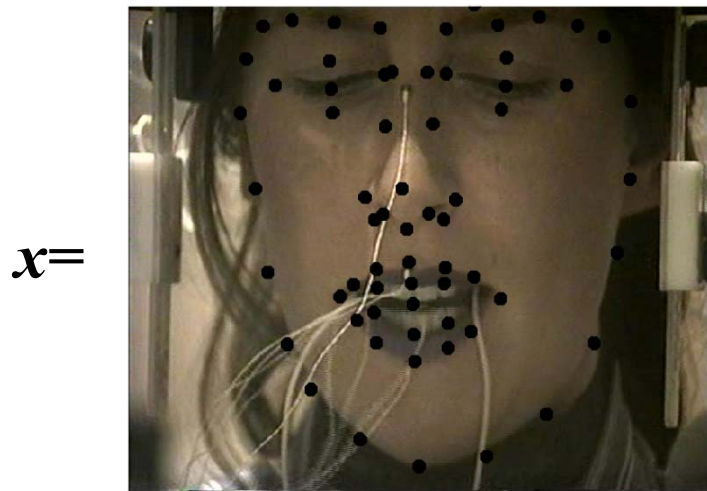
$y=$



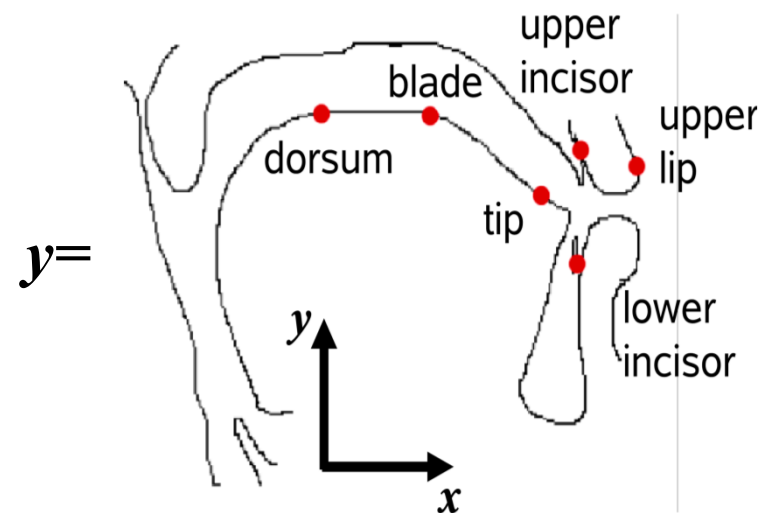
- **Εφαρμογή A2V: (Avatar)** Τεχνητό ομιλών πρόσωπο οδηγούμενο από ακουστικό σήμα.
- **Εφαρμογή V2A:** Αποθρομβοποίηση ακουστικού σήματος με χρήση οπτικής πληροφορίας.
- **Εφαρμογή AV inversion:** Ανάκτηση σχήματος φωνητικής οδού με σύμμιξη οπτικής και ακουστικής πληροφορίας.
- **Εφαρμογή σε συγχρονισμό AV καναλιών βίντεο:** Εύρεση χρονικής ολίσθησης που μεγιστοποιεί το συντ. συσχέτισης.

## Εφαρμογές CCA: Ανάκτηση γεωμετρίας φωνητικής οδού

- **Παράδειγμα:** Πρόβλεψη της θέσης σημείων επί της φωνητικής οδού από οπτική πληροφορία εξαγμένη από το πρόσωπο ομιλητή.



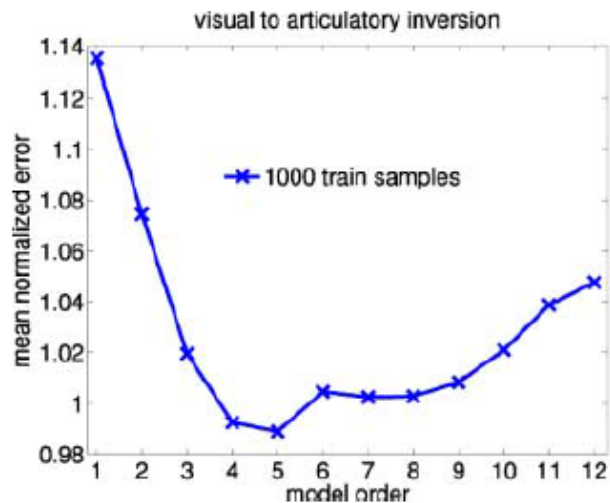
Θέση σημείων-κλειδιών στο πρόσωπο



Θέση αισθητήρων στη γλώσσα

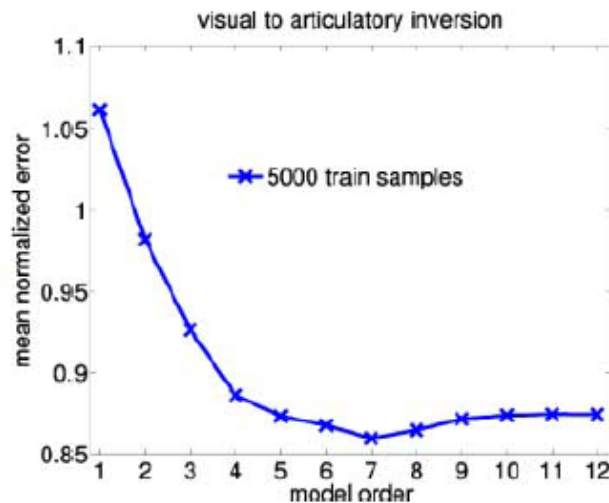
## Εφαρμογές CCA: Ανάκτηση Γεωμετρίας Φωνητικής Οδού

- Εκτίμηση στατιστικών 2<sup>ης</sup> τάξης από περιορισμένα δεδομένα:
  - Διατηρώντας μόνο τις κύριες CCA συνιστώσες αυξάνει τη γενικευτική ικανότητα του συστήματος



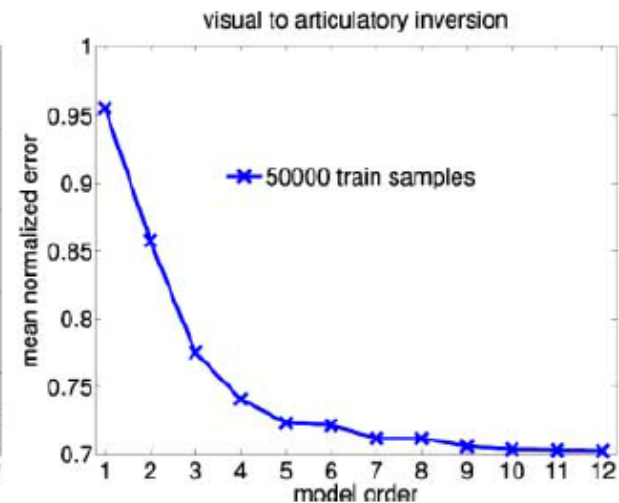
Λίγα δεδομένα  
εκπαίδευσης.

$N=1.000$



Περισσότερα δεδομένα  
εκπαίδευσης.

$N=5.000$

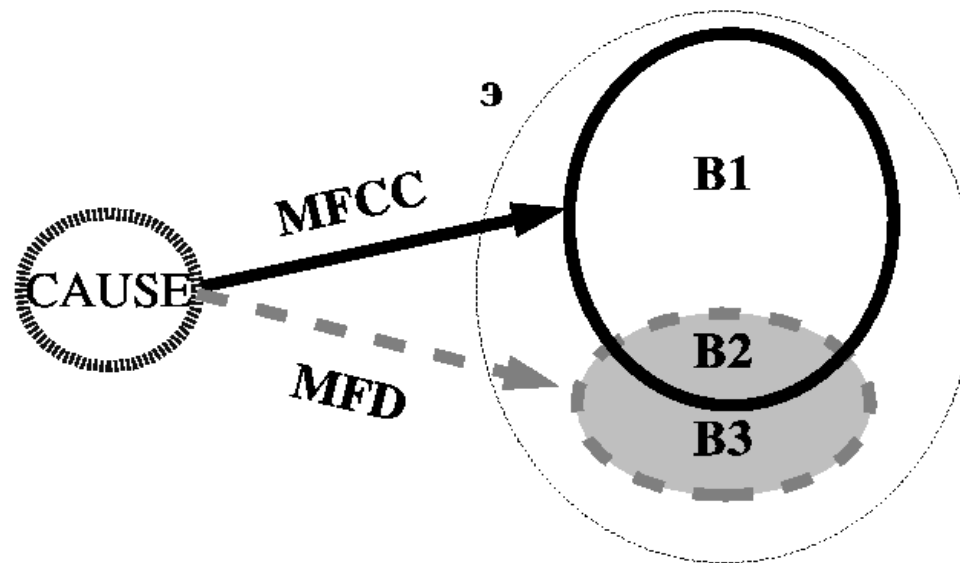


Άφθονα δεδομένα  
εκπαίδευσης.  $N=50.000$

Πλήρης τάξη CCA = 12.

# CCA μεταξύ Φράκταλ και MFCC χαρακτηριστικών: Αρχική θεώρηση

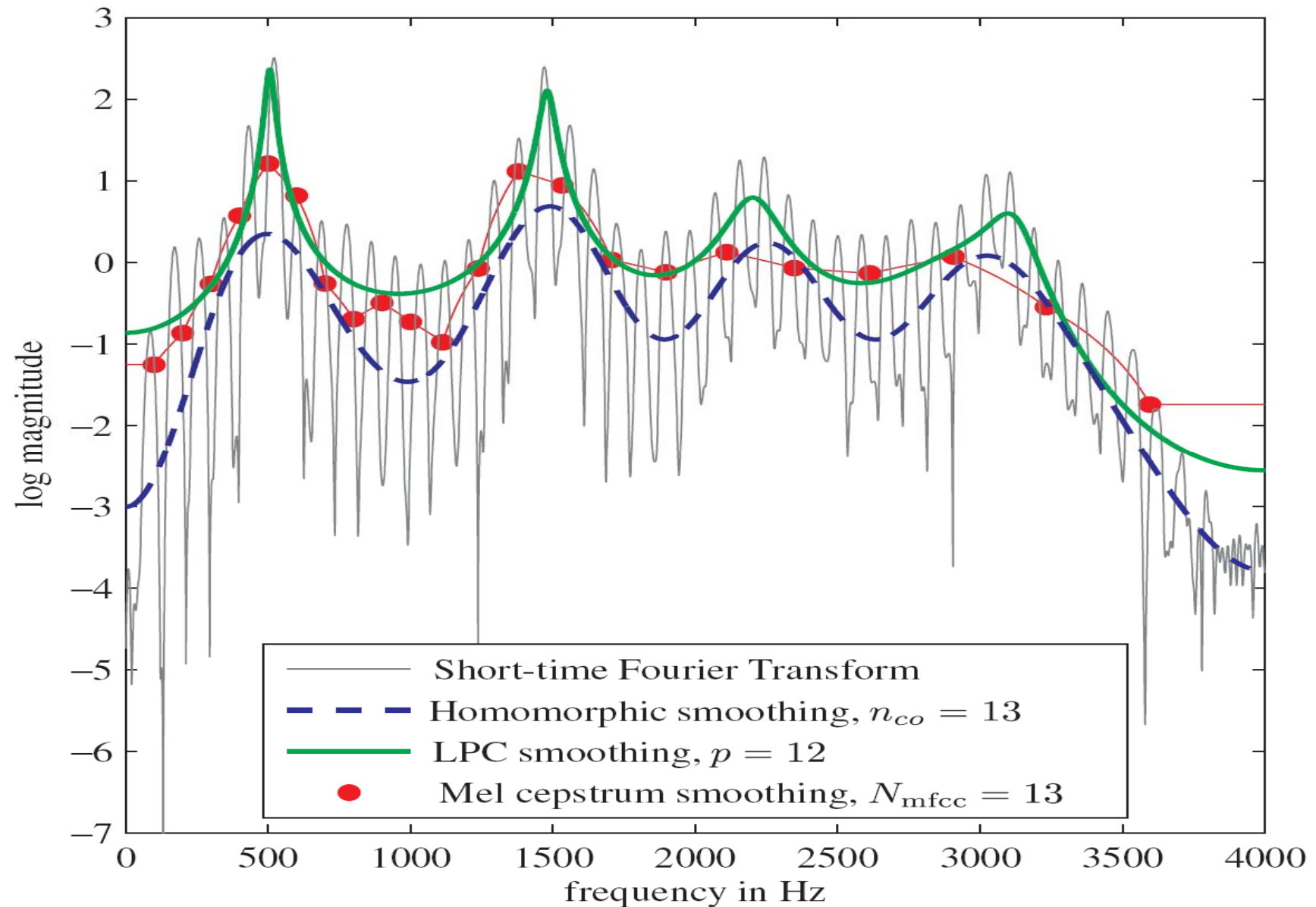
- κοινή αιτία φαινομένου – φωνητικό σήμα
- εναλλακτικές αναπαραστάσεις
  - π.χ. {LPC και Cepstrum} vs. Fractal Theory
- ποσοτικοποίηση συσχέτισης μεταξύ μετρήσεων των δύο διαφορετικών θεωρήσεων



Σχήμα 5.3: Σχηματική αναπαράσταση των υποχώρων μέγιστης και ελάχιστης συσχέτισης δυο πηγών πληροφορίας οι οποίες περιγράφουν-εξηγούν μερικώς διαφορετικές πτυχές του συνολικού χώρου ο οποίος οφείλεται σε ένα κοινό φαινόμενο-αιτία.



# Comparison of Spectral Smoothing Methods

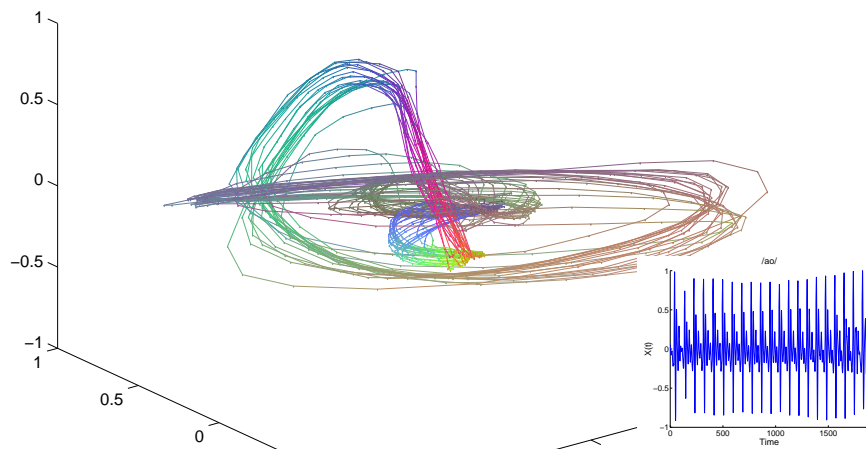




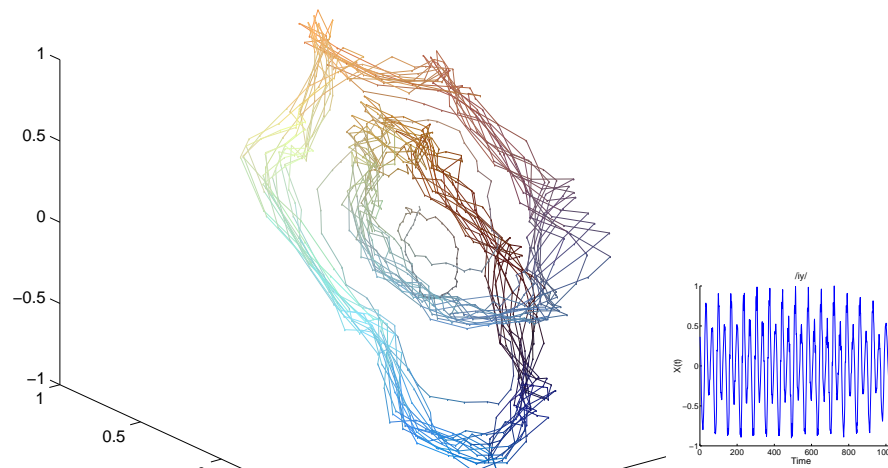
# Reconstructed Attractors

(Pitsikalis & Maragos 2002)

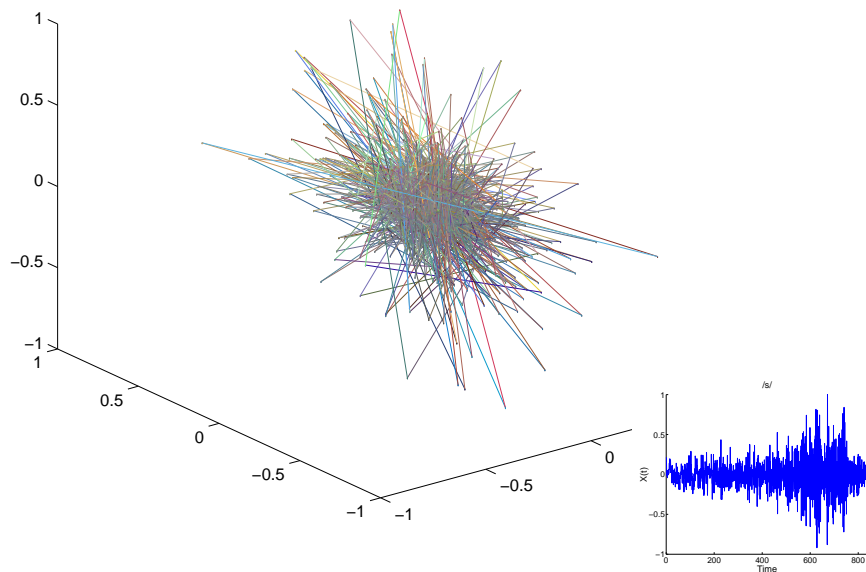
*/ao/*,  $D_E=6$ , #1846



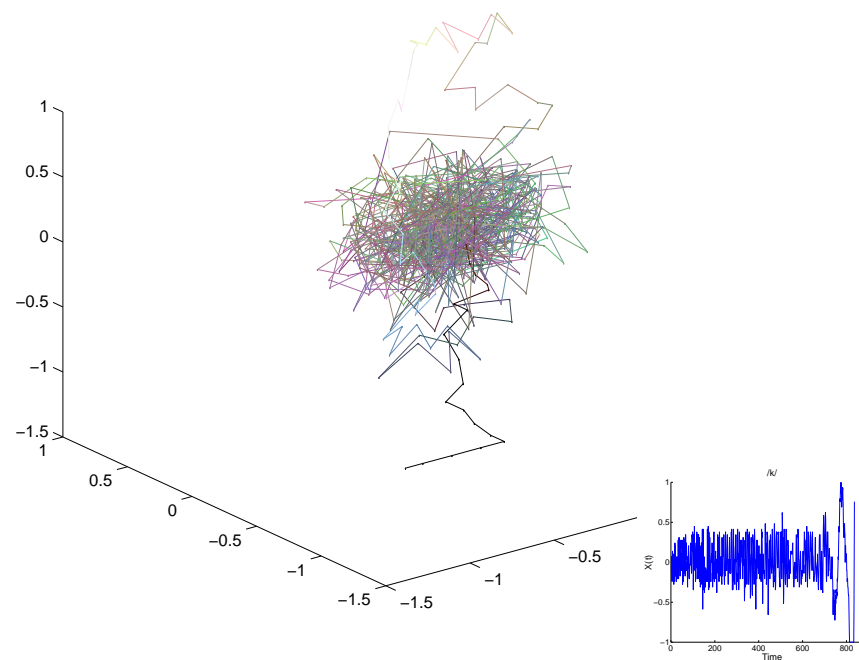
*/iy/*,  $D_E=5$ , #1068



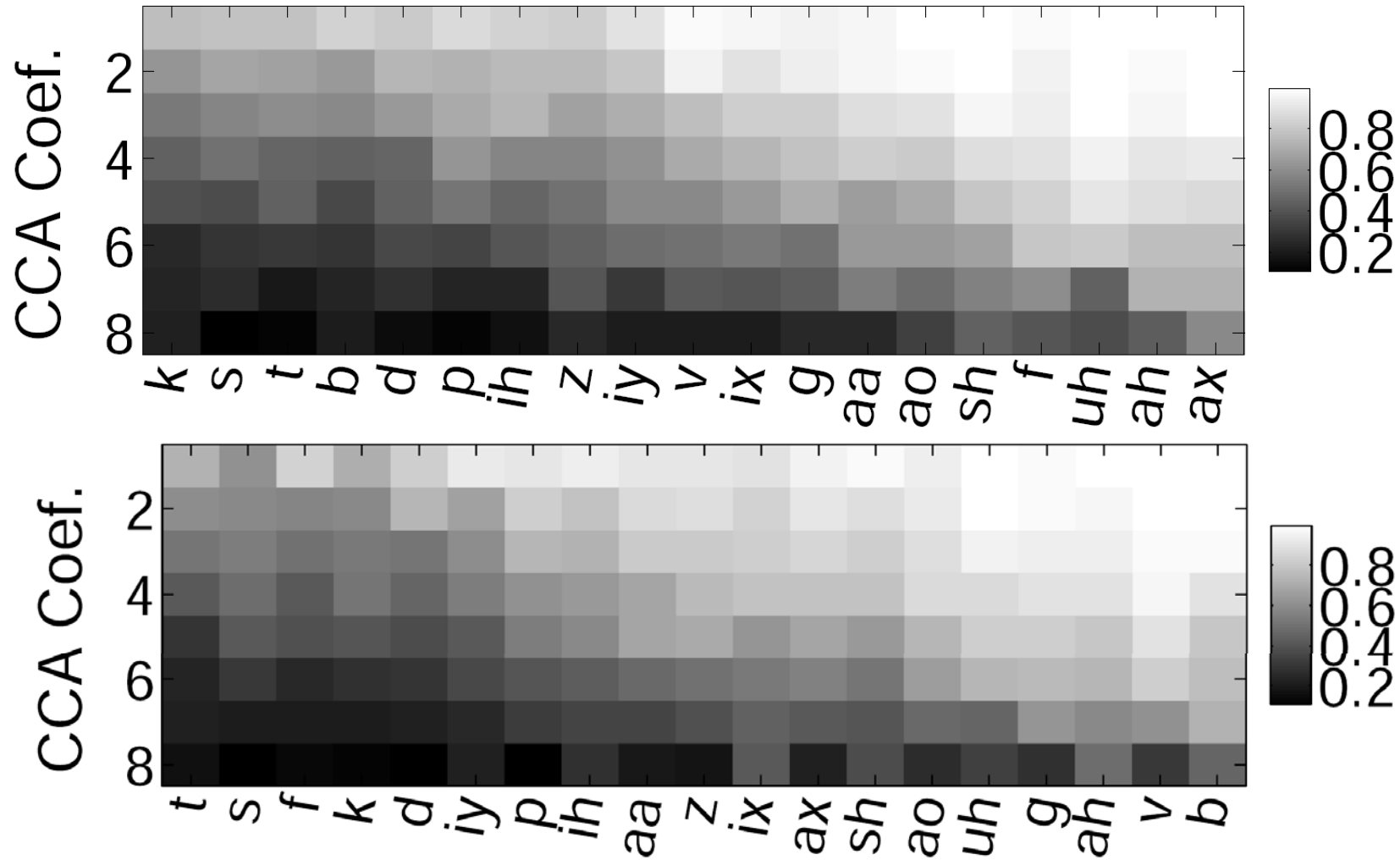
*/s/*,  $D_E=5$ , #829



*/k/*,  $D_E=6$ , #816

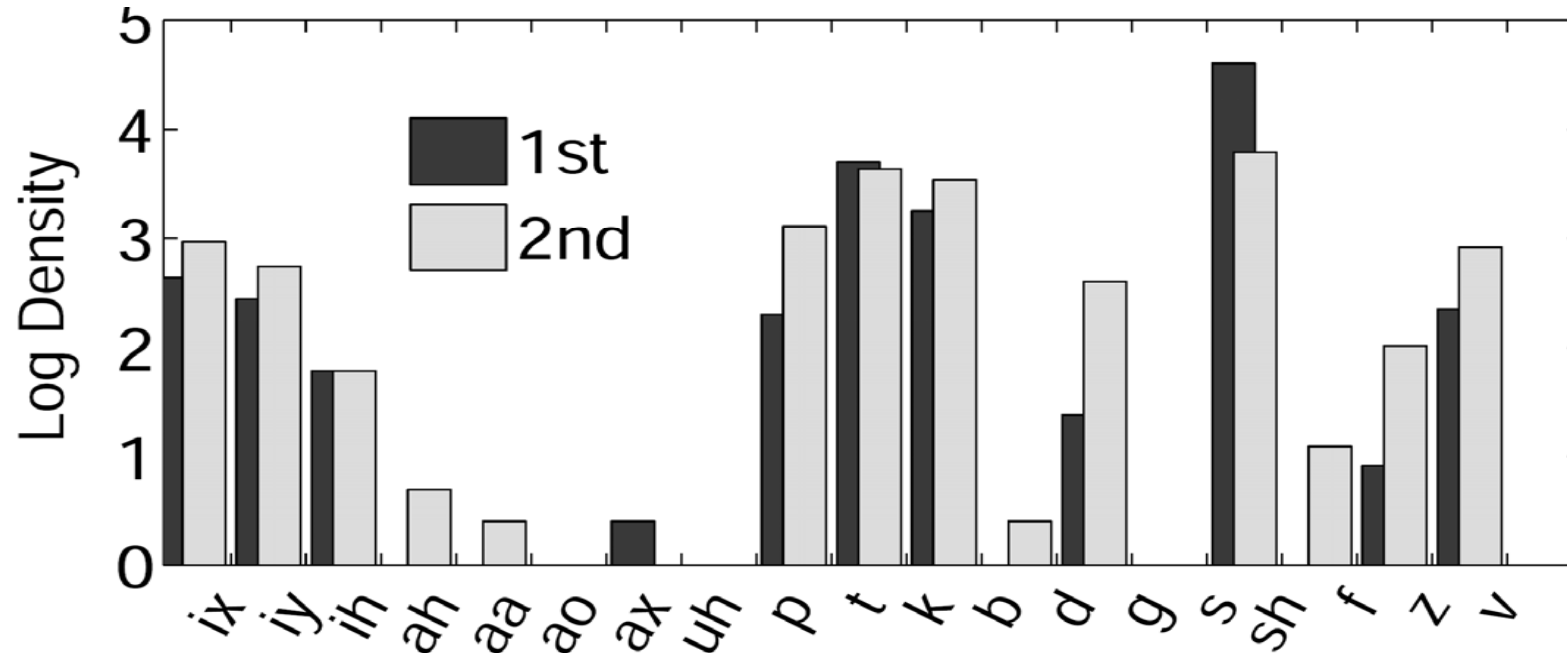


# CCA μεταξύ Φράκταλ και MFCC Χαρακτηριστικών: Κατάταξη Φωνημάτων



- κριτήριο: μέσος δείκτης συσχέτισης
- κατάταξη φωνήματος ανα ομιλητή

# CCA μεταξύ Φρακτάλ και MFCC χαρακτηριστικών: Ιστόγραμμα κατάταξης



- κριτήριο: μέσος δείκτης συσχέτισης
- κατάταξη φωνήματος ανεξαρτήτως ομιλητή
- πόσες φορές κάποιο φώνημα κατατάχθηκε 1ο ή 2ο

# Αναφορες για CCA

## Γενικες Αναφορες:

- L. L. Scharf and J. K. Thomas , “Wiener Filters in Canonical Coordinates for Transform Coding, Filtering, and Quantizing”, *IEEE Trans. Signal Processing*, March 1998.

## Audio-Visual Synchrony:

- M. Slaney and M. Covell, “FaceSync: A Linear operator for Measuring Synchronization of Video Facial Images and Audio Tracks”, *Proc. NIPS*, 2001.

## Audio-Visual Inversion of Speech to 3D Geometry of Vocal Tract:

- A. Katsamanis, G. Papandreou, and P. Maragos, “Face Active Appearance Modeling and Speech Acoustic Information to Recover Articulation”, *IEEE Trans. Audio, Speech and Language Processing*, March 2009.

## CCA μεταξύ Φράκταλ και MFCC χαρακτηριστικών Φωνής:

- V. Pitsikalis and P. Maragos, “Analysis and Classification of Speech Signals by Generalized Fractal Dimension Features”, *Speech Communication*, Dec. 2009.

## Ευχαριστιες για βοηθεια με τα slides:

- Ιασονας Κοκκινος
- Γιωργος Παπανδρεου
- Βασιλης Πιτσικάλης

## Ιστοσελίδα Μαθήματος:

<http://cvsp.cs.ntua.gr/courses/patrec/>