Αναγνώριση Προτύπων

Σχολή ΗΜ&ΜΥ, ΕΜΠ
Πετρος Μαραγκος

Compact Subspace Representations:

SVD, KLT, PCA, LDA, ICA, CCA, NMF

Συμπαγείς Αναπαραστάσεις Υποχώρων

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

Big (Multimodal) Data Challenges

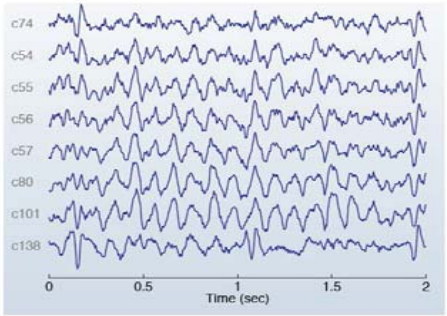
- **Data are Voluminous**:
  - 24 hrs of TV = 430 Gb = 2.160.000 still (frame) images
  - WWW: 300-hr videos are uploaded on YouTube per minute.
  - 300 millions images are uploaded on FaceBook per day.
  - Kinect sensor: 250 MB/sec (uncompressed RGB)
- **Data are Dynamic**
  - Temporal video, Website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

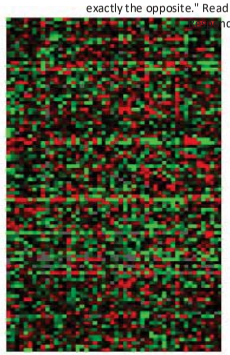
εικόνες προσώπων

κείμενο

Celebrations over the historic election of the first black president have been interspersed with emotional protests against a series of bans on bay 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 nds to 2008 Election

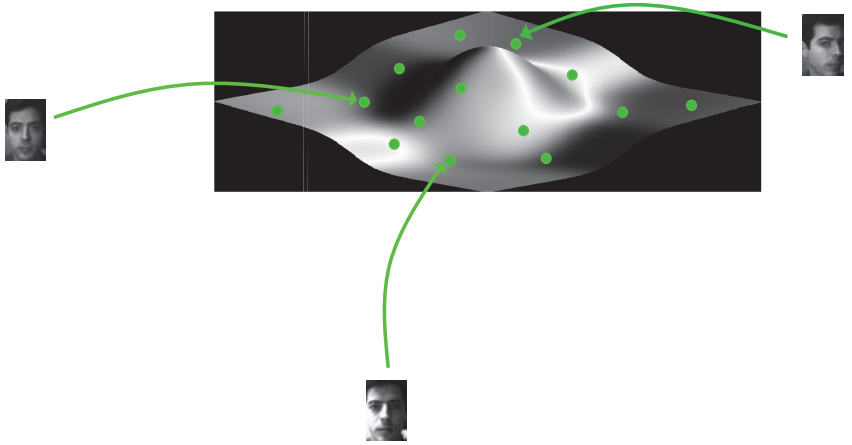


MEG



γονίδια

Manifold of face category



Χωροι Hilbert

- Ορισμοι, Εσωτερικο Γινομενο
- Μηκος, Ενεργεια
- Ορθογωνιοτητα
- Ορθοκανονικη Βαση
- Θεωρημα Προβολης
- Προσεγγιση με Ελαχιστα Τετραγωνα

Ref: A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science,* Springer-Verlag, 1982.

Χώροι Hilbert: Πληρεις Γραμμικοί Χώροι Εσωτερικού Γινομένου

- Χωροι Hilbert $\mathcal{H}$
  - Χωρος L_2 των σηματων συνεχους χρονου $f(t)$ με $\int |f(t)|^2 < \infty$
  - Χωρος ℓ_2 των σηματων διακριτου χρονου $x[n]$ με $\sum |x[n]|^2 < \infty$
  - Χωρος $\mathcal{C}^N$ των διανυσματων ή σηματων με N δείγματα.
  - Χωρος $\mathcal{V}$ τυχαιων μεταβλητων X με $E\{|X|^2\} < \infty$.
- **Γραμμικοι χωροι:**
  - Αθροισμα «διανυσματων»
  - Γινομενο «διανυσματος» με «βαθμωτο»
- **Εσωτερικο Γινομενο:**

$$L_2 : \langle \mathbf{f}, \mathbf{g} \rangle = \int f(t) g^*(t)\, dt$$

$$\ell_2,\ \mathcal{C}^N : \langle \mathbf{x}, \mathbf{y} \rangle = \sum_n x[n] y^*[n]$$

$$\mathcal{V} : \langle X, Y \rangle = E\{XY^*\}$$

- **Πληροτητα:** Περιεχουν τα οριακα σημεια ακολουθιων.

Χώροι Hilbert : Ιδιοτητες Εσωτερικου Γινομενου

Σε κάθε μιγαδικό γραμμικό χώρο εσωτερικού γινομένου, το εσωτερικό γινόμενο $\langle \cdot, \cdot \rangle$ ικανοποιεί τις ακόλουθες ιδιότητες. Για κάθε "διανύσματα" $\mathbf{x}, \mathbf{y}$, και κάθε "βαθμωτό" a :

1. $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$
2. $\langle a\mathbf{x}, \mathbf{y} \rangle = a \langle \mathbf{x}, \mathbf{y} \rangle$
3. $\langle \mathbf{x}, \mathbf{y} \rangle = (\langle \mathbf{y}, \mathbf{x} \rangle)^*$
4. $\langle \mathbf{x}, \mathbf{x} \rangle > 0 \quad \forall \mathbf{x} \neq \mathbf{0}$

Συνέπειες

5. $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle$
6. $\langle \mathbf{x}, a\mathbf{y} \rangle = a^* \langle \mathbf{x}, \mathbf{y} \rangle$
7. $\langle \mathbf{x}, \mathbf{y} \rangle = 0 \ \forall \mathbf{y} \quad \Rightarrow \quad \mathbf{x} = \mathbf{0}$

Χώροι Hilbert : Νόρμα και Ορθογωνιότητα

- **Νόρμα** ("Μήκος"):

$$L_2 : \|\mathbf{f}\| = \sqrt{\langle \mathbf{f}, \mathbf{f} \rangle} = \sqrt{\int |f(t)|^2\, dt}$$

$$\ell_2,\ \mathcal{C}^N : ||\mathbf{x}|| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_n |x[n]|^2}$$

$$\mathcal{V} : \|X\| = \sqrt{\langle X, X \rangle} = \sqrt{E\{|X|^2\}}$$

- **Ενεργεια = (Νορμα)^2**
- **Ορθογωνιότητα:**

$$\mathbf{x} \perp \mathbf{y} \Leftrightarrow \langle \mathbf{x}, \mathbf{y} \rangle = 0$$

- **Πυθαγορειο θεωρημα:**

$$\mathbf{x} \perp \mathbf{y} \Rightarrow ||\mathbf{x} + \mathbf{y}||^2 = ||\mathbf{x}||^2 + ||\mathbf{y}||^2$$

ΟΡΘΟΚΑΝΟΝΙΚΗ ΒΑΣΗ

Μια ακολουθια $(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, ..., \boldsymbol{\varphi}_n, ...)$ σε ενα Hilbert χωρο $\mathcal{H}$ λεγεται:

- **Ορθογωνια** (Orthogonal) αν $\boldsymbol{\varphi}_i \perp \boldsymbol{\varphi}_j \quad \forall\, i \neq j$
- **Ορθοκανονικη** (Orthonormal) αν $\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \rangle = \delta_{ij}$
- **Πληρης (maximal) Ορθοκανονικη** αν $B=\{\boldsymbol{\varphi}_n\}$

ειναι ορθοκανονικο συνολο και δεν υπαρχει μοναδιαιο $\mathbf{x}_0 \in \mathcal{H}$ ωστε $B \cup \{\mathbf{x}_0\}$ να ειναι ορθοκανονικο.

- **Ορθοκανονικη Βαση** (OrthoNormal Basis-ONB)

αν $(\boldsymbol{\varphi}_n)$ ειναι πληρης ορθοκανονικη ακολουθια.

Διασταση $(\mathcal{H})$ = cardinality (ONB)

Παραδειγματα Ορθοκανονικων Βασεων: Διακριτα Σηματα

(1) Χωρος: $\mathcal{H} = \mathbb{C}^N$

ONB: $\boldsymbol{\varphi}_i = (0, ...0, \underset{\uparrow}{1}, 0, ..., 0) \;,\; i = 1, 2, ..., N$

θεση i

(2) Χωρος: $\mathcal{H} = \ell_2(\mathbb{Z})$

ONB: $\varphi_i[n] = \delta[n-i] \;,\quad i \in \mathbb{Z}$

Παραδειγματα Ορθοκανονικων Βασεων: Συνεχη Σηματα

(3) Χωρος: $\mathcal{H} = L_2([0,1])$

ONB: $\varphi_n(t) = e^{j2\pi nt} \;,\; n = 0, \pm 1, \pm 2, ...$

(4) Χωρος: $\mathcal{H} = L_2[0, \infty)$

ONB: $\varphi_n(t) = \dfrac{1}{n!} e^{-t/2} L_n(t) \;,\; n = 0, 1, 2, ...$

Laguerre πολυωνυμα: $L_n(t) = e^t D^n (t^n e^{-t})$

(5) Χωρος: $\mathcal{H} = L_2(-\infty, \infty)$

ONB: $\varphi_n(t) = \dfrac{e^{-t^2/2}}{(2^n n! \sqrt{\pi})^{1/2}} H_n(t) \;,\; n = 0, 1, 2, ...$

Hermite πολυωνυμα: $H_n(t) = (-1)^n e^{t^2} D^n (e^{-t^2})$

Θεωρημα Γενικευμενων Σειρων Fourier

Σε ενα Hilbert χωρο $\mathcal{H}$ οι ακολουθες προτασεις ειναι ισοδυναμες:

(a) Η ακολουθια $(\boldsymbol{\varphi}_n)$ ειναι **ορθοκανονικη βαση**

(b) **Αναπαρασταση**: $\mathbf{x} = \sum_n \underbrace{\langle \mathbf{x}, \boldsymbol{\varphi}_n \rangle}_{\substack{\text{Γενικευμενοι} \\ \text{Fourier} \\ \text{Συντελεστες}}} \boldsymbol{\varphi}_n \;,\quad \forall \mathbf{x} \in \mathcal{H}$

(c) **Parseval** ισοτητα:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_n \langle \mathbf{x}, \boldsymbol{\varphi}_n \rangle \langle \mathbf{y}, \boldsymbol{\varphi}_n \rangle^* \;,\quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{H}$$

(d) **Plancherel** ισοτητα:

$$\|\mathbf{x}\|^2 = \sum_n |\langle \mathbf{x}, \boldsymbol{\varphi}_n \rangle|^2 \;,\quad \forall \mathbf{x} \in \mathcal{H}$$

Theorem: Generalized Fourier Series

Let $(\boldsymbol{\varphi}_n)$ be an orthonormal sequence in a Hilbert space $\mathcal{H}$.

Then the following are equivalent:

(a) The set $\{\boldsymbol{\varphi}_n\}$ is an **orthonormal basis**.

(b) **Representation**: $\mathbf{x} = \sum_n \underbrace{\langle \mathbf{x}, \boldsymbol{\varphi}_n \rangle}_{\text{Generalized Fourier Coefficients}} \boldsymbol{\varphi}_n \ , \quad \forall \mathbf{x} \in \mathcal{H}$

(b) implies: Unconditional convergence, Unique Coeffs

(c) **Parseval** equality:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_n \langle \mathbf{x}, \boldsymbol{\varphi}_n \rangle \langle \mathbf{y}, \boldsymbol{\varphi}_n \rangle^* \ , \quad \forall \mathbf{x}, \mathbf{y} \in \mathcal{H}$$

(d) **Plancherel** equality:

$$\|\mathbf{x}\|^2 = \sum_n |\langle \mathbf{x}, \boldsymbol{\varphi}_n \rangle|^2 \ , \quad \forall \mathbf{x} \in \mathcal{H}$$

(e) Any linear subspace $\mathcal{M}$ containing $\{\boldsymbol{\varphi}_n\}$ is dense in $\mathcal{H}$.

Namely, $\overline{\text{span}\{\boldsymbol{\varphi}_n\}} = \mathcal{H}$. Thus, each $\mathbf{x} \in \mathcal{H}$ can be written as $\mathbf{x} = \lim \mathbf{y}_k$ where $(\mathbf{y}_k)$ is some sequence in $\mathcal{M}$.

Διακριτοι Ορθογωνιοι Μετασχηματισμοι

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

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

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

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

Γραμμικος Μετασχηματισμος (Πινακας x Διανυσμα εισοδου):

$$\mathbf{y} = \mathbf{A}^H \mathbf{x} \ , \quad \begin{pmatrix} y[k] = \sum_{n=0}^{N-1} a^*[k,n] x[n] \\ k = 0,1,\ldots,N-1 \end{pmatrix}$$

Αντιστροφος Μετασχηματισμος (Αντιστροφος Πινακας x Διανυσμα Εξοδου):

$$\mathbf{x} = \mathbf{A}\mathbf{y} \ , \quad \begin{pmatrix} x[n] = \sum_{k=0}^{N-1} a[n,k] y[k] \\ n = 0,1,\ldots,N-1 \end{pmatrix}$$

Διακριτοι Ορθογωνιοι Μετασχηματισμοι με Ιδεες Hilbert Χωρων

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

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

Unitary Πινακας Μετασχηματισμου: $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$

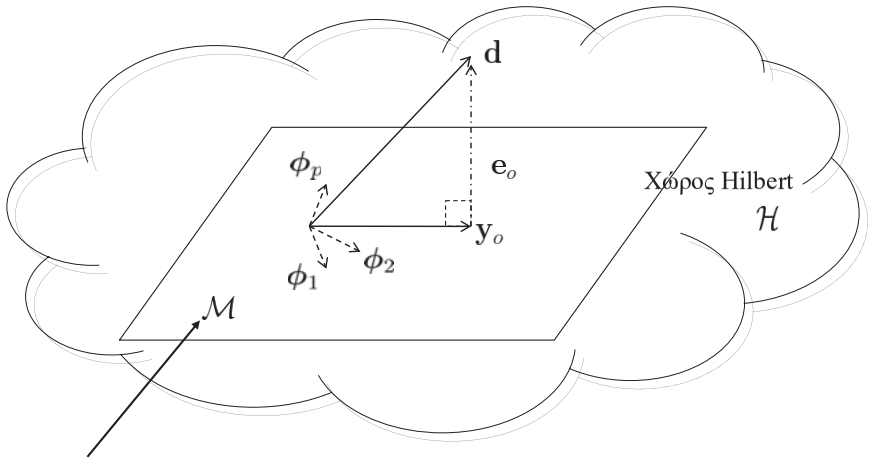
Γραμμικος Μετασχηματισμος: Τα στοιχεια της εξοδου **y** ειναι εσωτερικα γινομενα της εισοδου **x** με τις στηλες του **A**:

$$y[k] = \langle \mathbf{x}, \mathbf{a}_k \rangle \ , \quad k = 0,1,\ldots,N-1$$

Αντιστροφος Μετασχηματισμος: Αναπαρασταση του **x** σε νεα ορθοκανονικη βαση με διανυσματα βασης τις στηλες του **A**:

$$\mathbf{x} = \sum_{k=0}^{N-1} y[k] \mathbf{a}_k$$

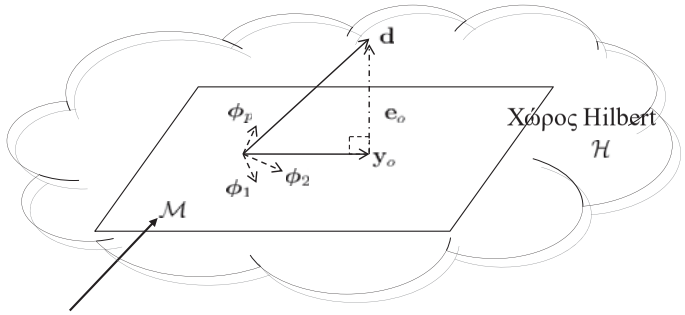
Θεώρημα Προβολής



Υποχώρος με διανύσματα βάσης $\phi_1, \phi_2, \ldots, \phi_p, \ldots$

- Θεώρημα : Υπάρχει ένα και μοναδικό διάνυσμα $\mathbf{y}_o \in \mathcal{M}$ που ελαχιστοποιεί την νόρμα λάθους. Το αντίστοιχο διάνυσμα λάθους $\mathbf{e}_o = \mathbf{d} - \mathbf{y}_o$ είναι ορθογώνιο προς τον υποχώρο $\mathcal{M}$. Επομένως $\mathbf{e}_o \perp \mathbf{y}_o$ και $\mathbf{e}_o \perp \phi_i \ \forall i$

Προσέγγιση Ελαχίστων Τετραγώνων



- Πρόβλημα : Προσέγγιση του $\mathbf{d}$ με γραμμικό συνδυασμό $\mathbf{y} = \sum_k w_k \phi_k$ ελαχιστοποιώντας την ενέργεια $\mathcal{E} = ||\mathbf{d} - \mathbf{y}||^2$
- Λύση : **Κανονικές Εξισώσεις**

$$\sum_k w_k \langle \phi_k, \phi_i \rangle = \langle \mathbf{d}, \phi_i \rangle , \quad \forall i$$

Ενέργεια Ελάχιστου Λάθους :

$$\mathcal{E}_{\min} = \langle \mathbf{d}, \mathbf{e}_o \rangle = \| \mathbf{d} \|^2 - \sum_k w_k^* \langle \mathbf{d}, \phi_k \rangle$$

Διαδοχικές Προσεγγισεις με Ορθοκανονικες Βασεις

Σε ενα Hilbert χωρο $\mathcal{H}$ η ακολουθια $(\boldsymbol{\varphi}_n)$ ειναι ορθοκανονικη βαση.

Απο το Θεωρημα Προβολης,

αν για καποιο $\mathbf{x} \in \mathcal{H}$

$\hat{\mathbf{x}}^{(k)}$ ειναι η καλυτερη προσεγγιση στον υποχωρο S_k με βαση $(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, ..., \boldsymbol{\varphi}_k)$,

τοτε η επομενη καλυτερη προσεγγιση στον υποχωρο S_{k+1} με βαση $(\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, ..., \boldsymbol{\varphi}_k, \boldsymbol{\varphi}_{k+1})$ ειναι

$$\hat{\mathbf{x}}^{(k+1)} = \hat{\mathbf{x}}^{(k)} + \langle \mathbf{x}, \boldsymbol{\varphi}_{k+1} \rangle \, \boldsymbol{\varphi}_{k+1}$$

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 A can be factored as

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

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

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

The singular values are the square roots of the nonzero eigenvalues σ_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:

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

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 $\mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$. If $\mathbf{A}$ is indefinite, then any negative eigenvalue in $\boldsymbol{\Lambda}$ becomes positive in $\boldsymbol{\Sigma}$.

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 $\mathbf{QS}$ where $\mathbf{Q}$ is orthogonal and $\mathbf{S}$ is symmetric and positive semidefinite. (If $\mathbf{A}$ is invertible, $\mathbf{S}$ is positive definite.)

$$\mathbf{A} = \mathbf{QS}, \quad \mathbf{Q} = \mathbf{UV}^T, \quad \mathbf{S} = \mathbf{V\Sigma V}^T \tag{12}$$

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 $\mathbf{Ax} = \mathbf{b}$ is given by

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

where $\mathbf{A}^+$ is called the **pseudoinverse** of $\mathbf{A}$ and $\mathbf{\Sigma}^+$ is a diagonal matrix with $1/\sigma_1, ..., 1/\sigma_r$ 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 $\sigma_1 \geq \cdots \geq \sigma_r$. If we approximate the matrix by keeping the $p < r$ largest singular values

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

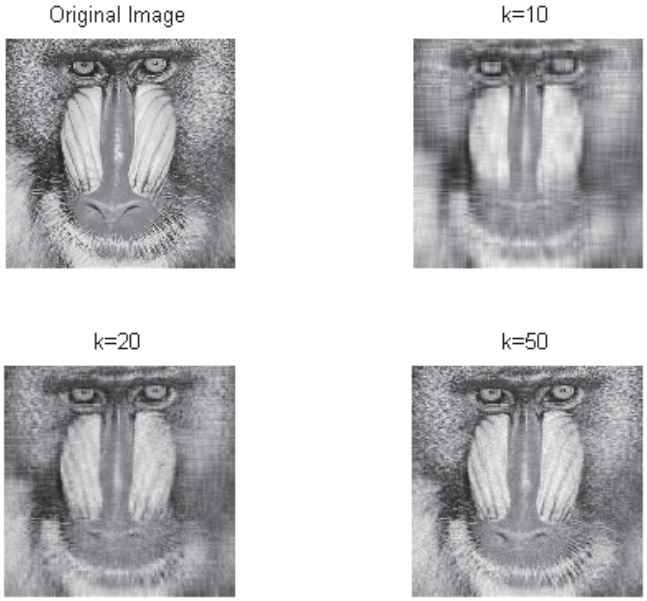
this approximation yields the smallest squared error

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

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

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

Παραδειγμα SVD: Γκριζες Εικονες



Διακριτοι Ορθογωνιοι Μετασχηματισμοι

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

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

Ο NxN Πινακας Μετασχηματισμου $\mathbf{A} = [a[n,k]]$ ειναι Unitary

($\mathbf{A}$ ειναι Ορθογωνιος για Πραγματικους πινακες):

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

Γραμμικος Μετασχηματισμος (Πινακας x Διανυσμα εισοδου):

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

Αντιστροφος Μετασχηματισμος (Αντιστροφος Πινακας x Διανυσμα Εξοδου):

$$\mathbf{x} = \mathbf{Ay}\ , \qquad \left(x[n] = \sum_{k=0}^{N-1} a[n,k]y[k], \quad n = 0,1,...,N-1 \right)$$

Unitary Discrete Fourier Transform (DFT)

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

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

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

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

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

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

Ευθυς: $\mathbf{y}=\mathbf{F}^H\mathbf{x}$, Αντιστροφος: $\mathbf{x}=\mathbf{F}\mathbf{y}$

Discrete Cosine Transform (DCT)

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

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

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

$$\beta[0]=\sqrt{1/N}, \qquad \beta[k]=\sqrt{2/N} \qquad 1\le k\le N-1$$

Αντιστροφος DCT:

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

Principal Component Analysis (PCA)

Karhunen Loeve Transform (KLT)

Αναλυση σε Πρωτευουσες Συνιστωσες

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 $\{\mathbf{e}_1,...,\mathbf{e}_d\}$ that form the columns of $\mathbf{A}$ and view them as an orthonormal basis for the whole space. Then,

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

If we project this vector onto the subspace formed by the smaller basis $\{\mathbf{e}_1,...,\mathbf{e}_p\}$, 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

$$J = \mathcal{E}\{\|\mathbf{x}-\hat{\mathbf{x}}\|^2\}=\mathcal{E}\{\|\mathbf{x}\|^2\}-\sum_{k=1}^{p}\mathcal{E}\{|y_k|^2\}$$

$$= \sum_{i=p+1}^{d}\mathcal{E}\{|y_i|^2\}=\sum_{i=p+1}^{d}\mathbf{e}_i^H\mathbf{R}_x\mathbf{e}_i$$

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 $\{\mathbf{e}_1, ..., \mathbf{e}_d\}$, 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 $\{\mathbf{e}_1, ..., \mathbf{e}_d\}$ consists of the eigenvectors of $\mathbf{R}_x$:

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

Hence, $\mathbf{e}_i^H \mathbf{R}_x \mathbf{e}_i = \lambda_i$. Thus, to minimize J we should order the eigenvalues as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ and choose as *principal directions* $\{\mathbf{e}_1, ..., \mathbf{e}_p\}$ those that correspond to the p largest eigenvalues. The corresponding transform values $\{y_k : k = 1, ..., p\}$ 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:

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

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 nozero-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 $\{y'_k : k = 1, ..., p\}$ 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 $\mathbf{x}_n$, $n = 1, ..., N$, 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 $y_{kn} =< \mathbf{x}_n, \mathbf{e}_k >$. The principal directions $\{\mathbf{e}_i\}$ are the orthonormal eigenvectors of the sample correlation or covariance matrix.

Λαθος για Αποκοπη Συντελεστων Μετ/σμων (για Markov ακολουθια)

$$J_m = \frac{\sum_{k=m}^{N-1} \sigma_k^2}{\sum_{k=0}^{N-1} \sigma_k^2}, \qquad m = 0, \ldots, N-1$$

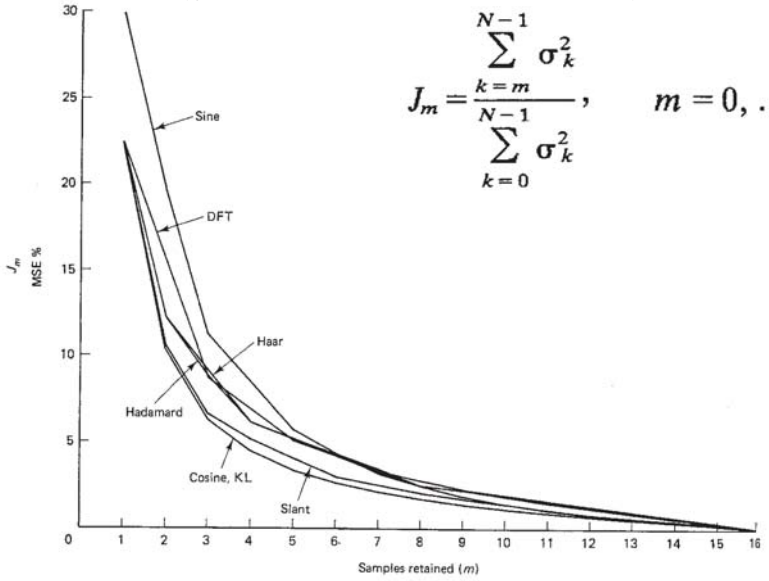


Figure 5.19 Performance of different unitary transforms with respect to basis restriction errors (J_m) versus the number of basis (m) for a stationary Markov sequence with $N = 16$, $\rho = 0.95$.

Εφαρμογη Ορθογωνιων Μετ/σμων σε Συμπιεση Εικονων

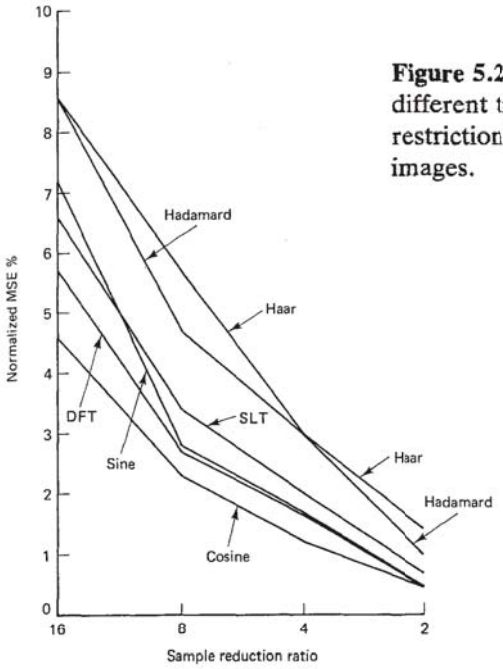


Figure 5.23 Performance comparison of different transforms with respect to basis restriction zonal filtering for 256×256 images.

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

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

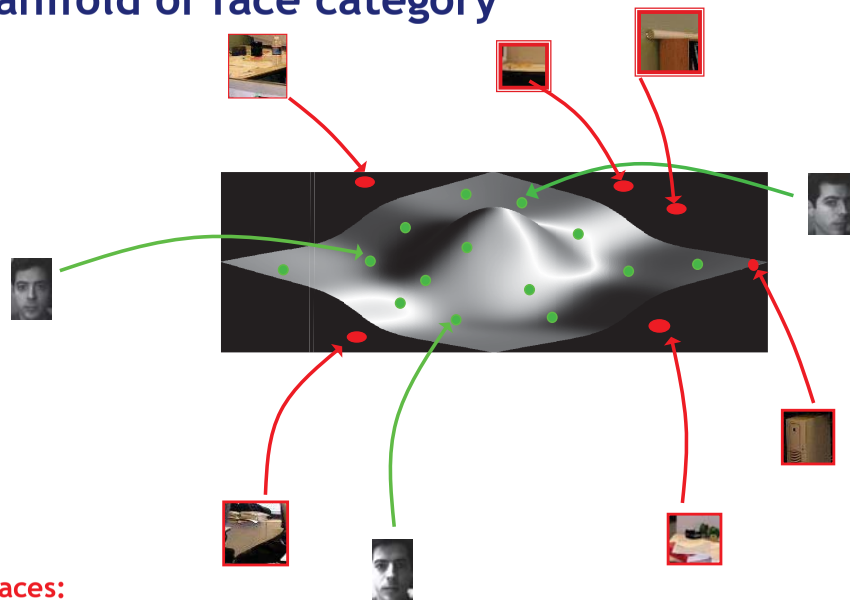
- Equivalent to classifying a vector in its nearest class subspace (*Πυθαγορειο θεωρημα*).
- If classification error is high, improve performance via learning subspace methods: iteratively rotate subspaces to adjust projection lengths of training vectors.
- Decision surfaces ?

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



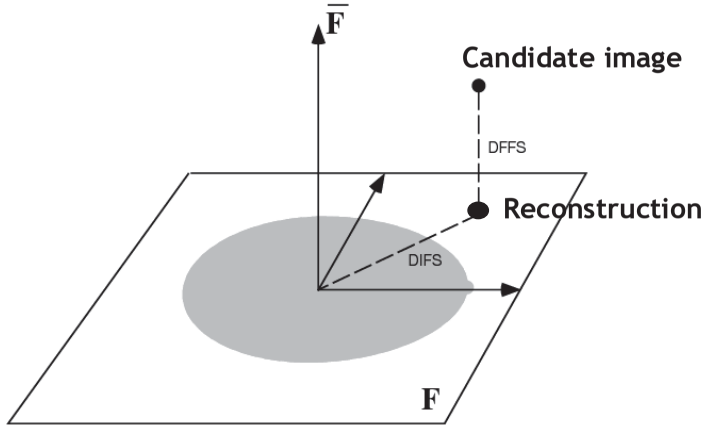
Non-faces:
Lie far from manifold, or in underpopulated regimes

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" $u_1,\ldots u_k$ that span that subspace
- Represent all face images in the dataset as linear combinations of eigenfaces

M. Turk and A. Pentland, Face Recognition using Eigenfaces, CVPR 1991

Eigenfaces/PCA: linearize manifold



Linear subspace spanned by PCA model

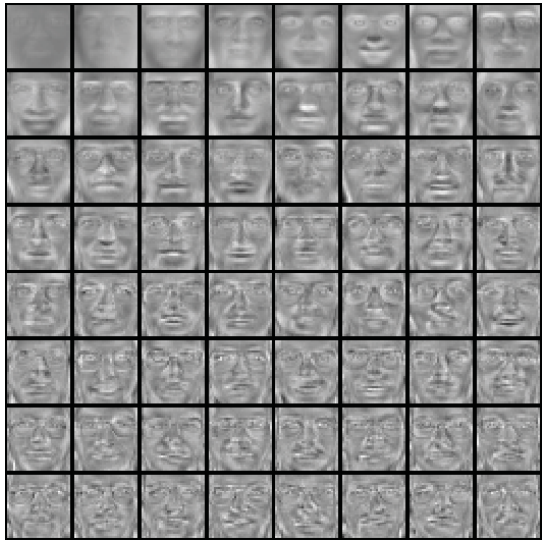
Eigenfaces example

- Training images
- $x_1,\ldots,x_N$

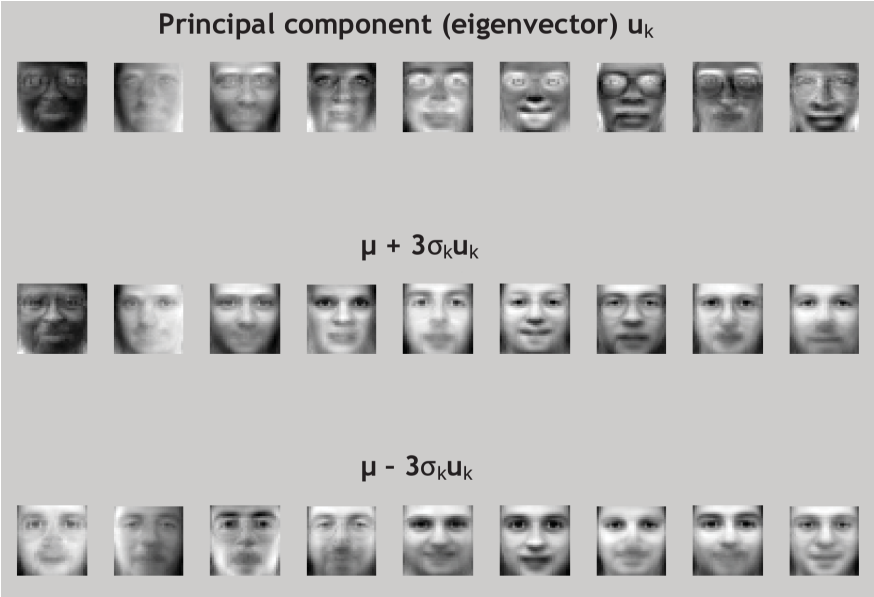
Eigenfaces example

Mean: μ

Top eigenvectors: $u_1,\ldots u_k$



Eigenfaces example



Eigenfaces: orthonormal basis for faces

- Face x in "face space" coordinates:

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

- Reconstruction:

$$\hat{\mathbf{x}} = \mu + w_1\mathbf{u}_1 + w_2\mathbf{u}_2 + w_3\mathbf{u}_3 + w_4\mathbf{u}_4 + \ldots$$

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:

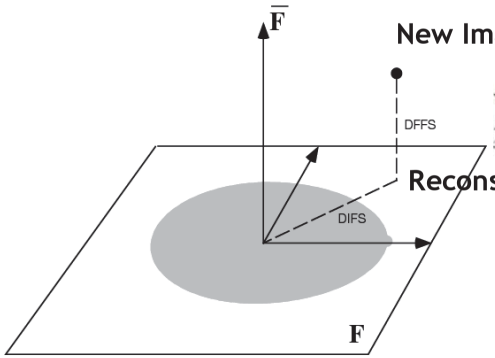
$$P(\mathbf{w}) = \frac{1}{\sqrt{2\pi|\Lambda|}} \exp\left(-\frac{\mathbf{w}^T\Lambda^{-1}\mathbf{w}}{2}\right) \qquad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & & & \\ 0 & 0 & \ldots & \lambda_N \end{pmatrix}$$

- IID Gaussian noise assumption:

Image Pixel

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

Eigenfaces for Detection



$$\sum_{x=1}^{M} \frac{(I(x) - T(x;\mathbf{w}))^2}{2\sigma^2}$$

$$\frac{\mathbf{w}^T \Lambda^{-1} \mathbf{w}}{2}$$

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

- 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

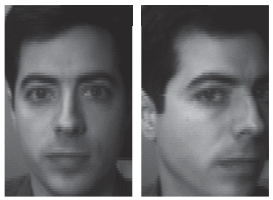


Challenges addressed by Eigenfaces

Short Term

Expression

Pose



Illumination

Long Term

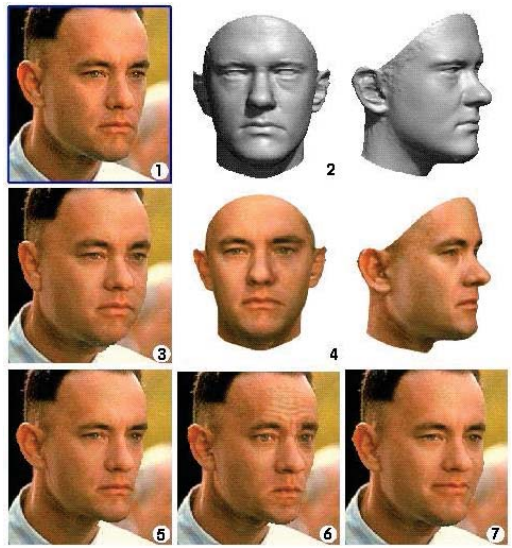
- Facial Hair
- Makeup
- Eyewear
- Hairstyle
- Piercings
- Aging

3D Morphable models

Recover Shape

Synthesize new views

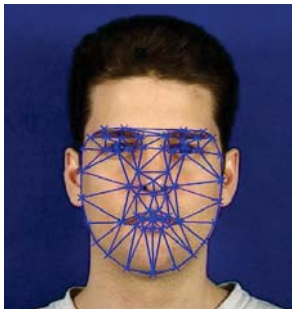
Synthesize new expressions



PCA σε Active Appearance Models (AAMs)

Ιδέα: Αναπαράσταση Σχήματος και Υφής σε γραμμικούς χώρους

Παράδειγμα επισημείωσης 68 σημείων-κλειδιών στο πρόσωπο.

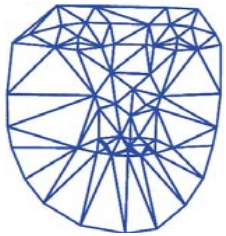


Επισημειωμένο σύνολο εκπαίδευσης XM2VTS (295 πρόσωπα)

PCA: Active Appearance Models

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

Εμφάνιση I



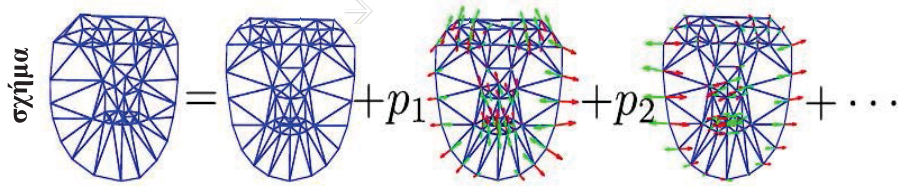
Σχήμα s

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) → p (10Δ)**



μέσο διάνυσμα | 1ο ιδιο-διάνυσμα | 2ο ιδιο-διάνυσμα

- **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

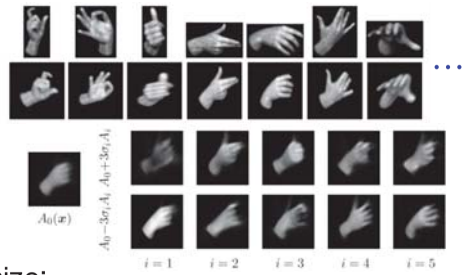
$$f(\boldsymbol{x}) = \begin{cases} g(\boldsymbol{I}(\boldsymbol{x})), & \text{if } \boldsymbol{x} \in M \\ -c_b & \text{else} \end{cases}$$

- Generative model

$$f(W_{\boldsymbol{p}}(\boldsymbol{x})) \approx A_0(\boldsymbol{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\boldsymbol{x})$$

$W_{\boldsymbol{p}}(\boldsymbol{x})$: *2D affine transform with parameters* $\boldsymbol{p} \in \Re^6$

- Training of the Model
  - Affine alignment of training set
    - generalization of procrustes analysis
    - iterative manual feedback
  - PCA to learn $A_i(\boldsymbol{x})$
    - keep only Nc=35 components



- Fitting : Find parameters $\boldsymbol{\lambda}, \boldsymbol{p}$ that minimize:

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

[Rousos, Theodorakis, Pitsikalis and Maragos, JMLR 2013]

Αναφορες για Visual PCA-AAM

Γενικες Αναφορες:

- 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:

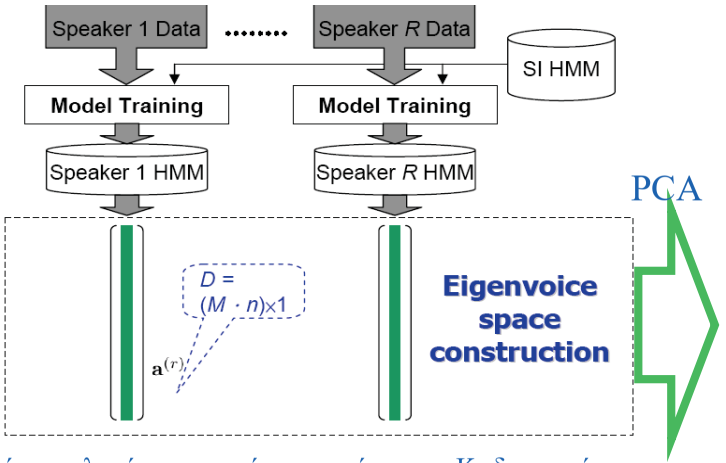
- A. Roussos, S. Theodorakis, V. Pitsikalis and P. Maragos, "Dynamic Affine-Invariant Shape-Appearance Handshape Features and Classification in Sign Language Videos", *Journal of Machine Learning Research*, 2013.

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)

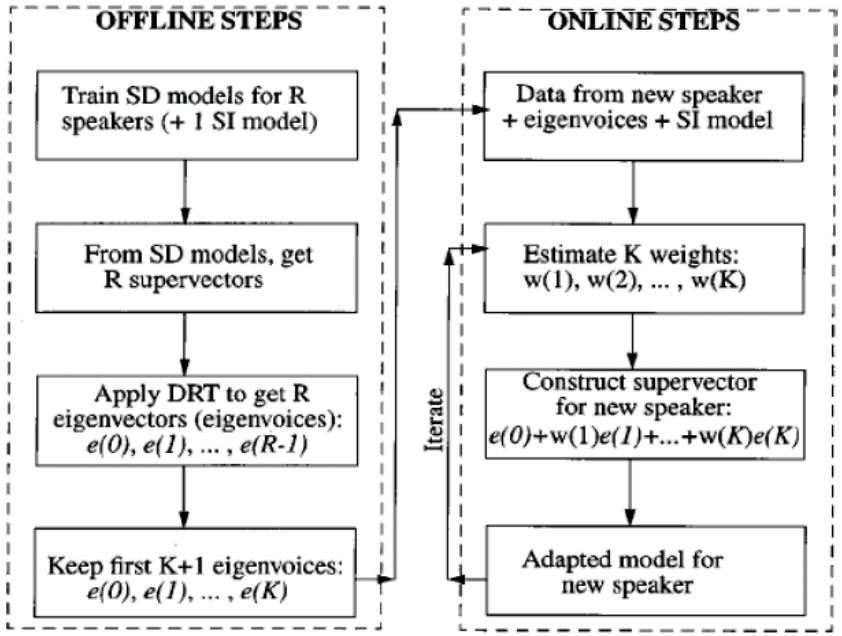
- **εκπαίδευση μοντέλων εξαρτημένων από τον ομιλητή**
- **δημιουργία υπερ-διανυσμάτων ($\boldsymbol{a}^{(r)}$) παραμέτρων / ομιλητή**



- νέος ομιλητής ως σημείο στον χώρο των K-ιδιοφωνών

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

PCA: Ιδιοφωνές και Προσαρμογή Ομιλητή



PCA: ιδιοφωνές και προσαρμογή ομιλητή

Κύριες Αναφορές:

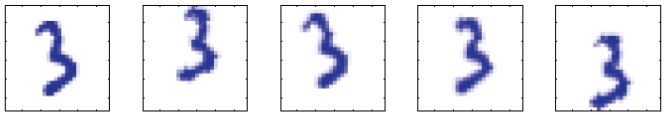
• R. Kuhn, P. Nguyen, J.C. Junqua, L. Goldwasser, N. Niedzielski, S. Fincke, K. Field, and M. Contolini, "***Eigenvoices for Speaker Adaptation***", Proc. 5th International Conference on Spoken Language Processing, 1998.

• 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.

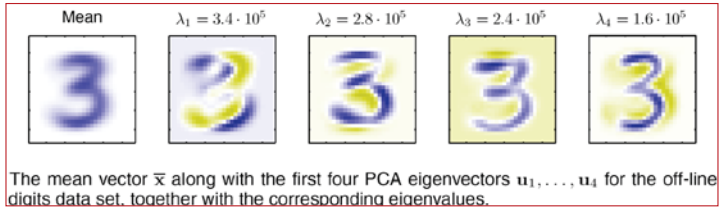
Probabilistic PCA (PPCA)

C. Bishop, Pattern Recognition and Machine Learning, Springer 2006

PCA on synthetic dataset with 3 latent variables



A synthetic data set obtained by taking one of the off-line digit images and creating multiple copies in each of which the digit has undergone a random displacement and rotation within some larger image field. The resulting images each have $100 \times 100 = 10,000$ pixels.



The mean vector $\bar{\mathbf{x}}$ along with the first four PCA eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_4$ for the off-line digits data set, together with the corresponding eigenvalues.

An original example from the off-line digits data set together with its PCA reconstructions obtained by retaining M principal components for various values of M. As M increases the reconstruction becomes more accurate and would become perfect when $M = D = 28 \times 28 = 784$.

PPCA: Generative viewpoint

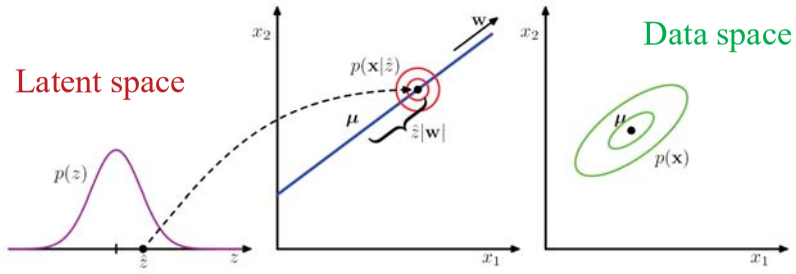


Figure 12.9 An illustration of the generative view of the probabilistic PCA model for a two-dimensional data space and a one-dimensional latent space. An observed data point $\mathbf{x}$ is generated by first drawing a value $\hat{z}$ for the latent variable from its prior distribution $p(z)$ and then drawing a value for $\mathbf{x}$ from an isotropic Gaussian distribution (illustrated by the red circles) having mean $\mathbf{w}\hat{z} + \boldsymbol{\mu}$ and covariance $\sigma^2\mathbf{I}$. The green ellipses show the density contours for the marginal distribution $p(\mathbf{x})$.

$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$$ Latent variable on principal-component subspace

$$\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}$$ Linear Gaussian Model

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$$ Gauss Conditional Distribution of Observed variable

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z} \Longrightarrow p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$$ Marginal Distribution

$$\mathbf{C} = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \sigma^2\mathbf{I}$$

Marginal (Predictive) and Posterior Distributions

Predictive Distribution $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$

$$\begin{aligned}\mathbb{E}[\mathbf{x}] &= \mathbb{E}[\mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\epsilon}] = \boldsymbol{\mu} \\ \mathrm{cov}[\mathbf{x}] &= \mathbb{E}\left[(\mathbf{W}\mathbf{z} + \boldsymbol{\epsilon})(\mathbf{W}\mathbf{z} + \boldsymbol{\epsilon})^{\mathrm{T}}\right] \\ &= \mathbb{E}\left[\mathbf{W}\mathbf{z}\mathbf{z}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\right] + \mathbb{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{\mathrm{T}}] = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \sigma^2\mathbf{I}\end{aligned}$$

$$\mathbf{C} = \mathbf{W}\mathbf{W}^{\mathrm{T}} + \sigma^2\mathbf{I} \Longrightarrow \mathbf{C}^{-1} = \sigma^{-1}\mathbf{I} - \sigma^{-2}\mathbf{W}\mathbf{M}^{-1}\mathbf{W}^{\mathrm{T}}$$

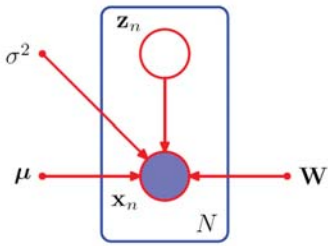
$$\mathbf{M} = \mathbf{W}^{\mathrm{T}}\mathbf{W} + \sigma^2\mathbf{I}$$

Posterior Distribution

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}\left(\mathbf{z}|\mathbf{M}^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}), \sigma^{-2}\mathbf{M}\right)$$

Maximum Likelihood PCA

The probabilistic PCA model for a data set of N observations of $\mathbf{x}$ can be expressed as a directed graph in which each observation $\mathbf{x}_n$ is associated with a value $\mathbf{z}_n$ of the latent variable.



$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2) = \sum_{n=1}^{N} \ln p(\mathbf{x}_n|\mathbf{W}, \boldsymbol{\mu}, \sigma^2)$$

$$= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\mathbf{C}| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{C}^{-1}(\mathbf{x}_n - \boldsymbol{\mu})$$

Set zero Log Likelihood Derivative wrt **μ** → **μ**=data mean →

$$\ln p(\mathbf{X}|\mathbf{W}, \boldsymbol{\mu}, \sigma^2) = -\frac{N}{2}\left\{D\ln(2\pi) + \ln|\mathbf{C}| + \mathrm{Tr}\left(\mathbf{C}^{-1}\mathbf{S}\right)\right\}$$

$$\mathbf{W}_{\mathrm{ML}} = \mathbf{U}_M(\mathbf{L}_M - \sigma^2\mathbf{I})^{1/2}\mathbf{R}$$

$$\sigma^2_{\mathrm{ML}} = \frac{1}{D-M}\sum_{i=M+1}^{D}\lambda_i$$

Inverse PPCA

PPCA = mapping from Latent space into Data space

Inverse PPCA = Summarize any point in Data space by its Posterior mean and covariance in Latent space.

$$\mathbb{E}[\mathbf{z}|\mathbf{x}] = \mathbf{M}^{-1}\mathbf{W}_{\mathrm{ML}}^{\mathrm{T}}(\mathbf{x} - \overline{\mathbf{x}})$$

$$\mathbf{W}\mathbb{E}[\mathbf{z}|\mathbf{x}] + \boldsymbol{\mu}$$ **Projection of posterior mean to a point in data space**

$$(\mathbf{W}_{\mathrm{ML}}^{\mathrm{T}}\mathbf{W}_{\mathrm{ML}})^{-1}\mathbf{W}_{\mathrm{ML}}^{\mathrm{T}}(\mathbf{x} - \overline{\mathbf{x}})$$ **Limit of posterior mean when variance → 0**

= Orthog. Proj. of data point onto latent space

EM Algorithm for PCA: E-step

Complete-Data Log Likelihood

$$\ln p\left(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2\right) = \sum_{n=1}^{N} \{\ln p(\mathbf{x}_n | \mathbf{z}_n) + \ln p(\mathbf{z}_n)\}$$

Expectation wrt posterior of latent distribution evaluated using "old parameters"

$$\mathbb{E}[\ln p\left(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \mathbf{W}, \sigma^2\right)] = -\sum_{n=1}^{N} \left\{ \frac{D}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^{\mathrm{T}}]\right) \right.$$
$$\left. + \frac{1}{2\sigma^2} \|\mathbf{x}_n - \boldsymbol{\mu}\|^2 - \frac{1}{\sigma^2} \mathbb{E}[\mathbf{z}_n]^{\mathrm{T}} \mathbf{W}^{\mathrm{T}} (\mathbf{x}_n - \boldsymbol{\mu}) + \frac{1}{2\sigma^2} \mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^{\mathrm{T}}] \mathbf{W}^{\mathrm{T}} \mathbf{W}\right) \right\}$$

$$\mathbb{E}[\mathbf{z}_n] = \mathbf{M}^{-1} \mathbf{W}^{\mathrm{T}} (\mathbf{x}_n - \overline{\mathbf{x}})$$

$$\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^{\mathrm{T}}] = \sigma^2 \mathbf{M}^{-1} + \mathbb{E}[\mathbf{z}_n] \mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}$$

M-step

Maximization of Expectation of Complete-Data Log Likelihood yields "new" parameter values

In the M step, we maximize with respect to $\mathbf{W}$ and σ^2, keeping the posterior statistics fixed. Maximization with respect to σ^2 is straightforward. For the maximization with respect to $\mathbf{W}$ we make use of (C.24), and obtain the M-step equations

$$\mathbf{W}_{\mathrm{new}} = \left[\sum_{n=1}^{N} (\mathbf{x}_n - \overline{\mathbf{x}}) \mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\right] \left[\sum_{n=1}^{N} \mathbb{E}[\mathbf{z}_n \mathbf{z}_n^{\mathrm{T}}]\right]^{-1}$$

$$\sigma^2_{\mathrm{new}} = \frac{1}{ND} \sum_{n=1}^{N} \left\{ \|\mathbf{x}_n - \overline{\mathbf{x}}\|^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}} \mathbf{W}_{\mathrm{new}}^{\mathrm{T}} (\mathbf{x}_n - \overline{\mathbf{x}}) \right.$$
$$\left. + \mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n \mathbf{z}_n^{\mathrm{T}}] \mathbf{W}_{\mathrm{new}}^{\mathrm{T}} \mathbf{W}_{\mathrm{new}}\right) \right\}.$$

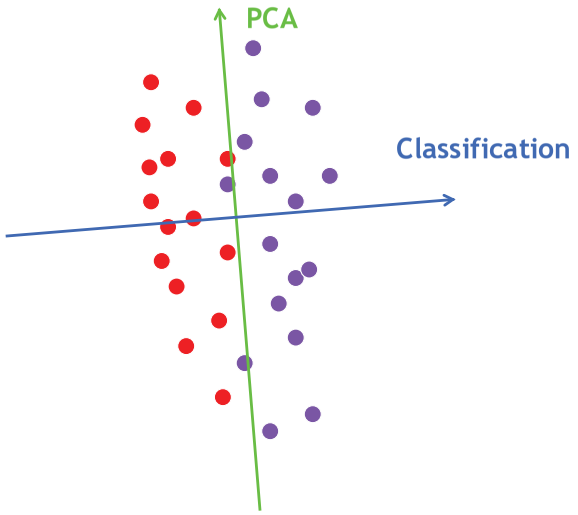
Linear Discriminant Analysis (LDA)

-

Ανάλυση Γραμμικής {Διάκρισης ή Διαχωρισμού}

Inherent Limitation of PCA

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



Linear Discriminant Analysis: Περιεχόμενα

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

LDA: Ανάλυση σε Υποχώρους

πολυδιάστατα θορυβώδη δεδομένα !

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

- **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: χαρακτηρας & 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

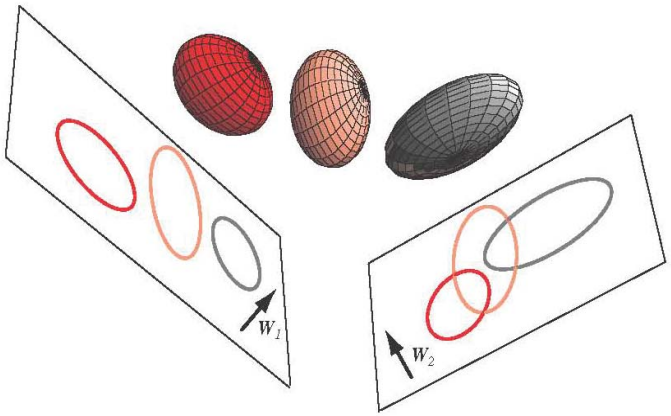


FIGURE 3.6. Three three-dimensional distributions are projected onto two-dimensional subspaces, described by a normal vectors $\mathbf{W}_1$ and $\mathbf{W}_2$. Informally, multiple discriminant methods seek the optimum such subspace, that is, the one with the greatest separation of the projected distributions for a given total within-scatter matrix, here as associated with $\mathbf{W}_1$. 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 στοιχεια 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:**

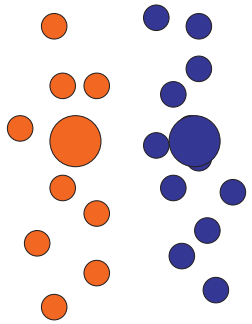
$$\mathbf{S}_i = \sum_{\mathbf{x}\in D_i}(\mathbf{x}-\mathbf{m}_i)(\mathbf{x}-\mathbf{m}_i)^T, \quad i=1,2$$

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

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

- **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: Ενδοταξικη Διασπορα

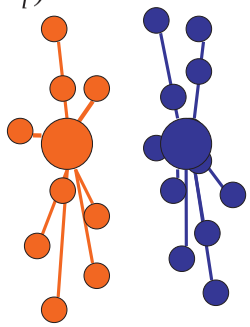
- **Πίνακας Ενδοταξικής Διασποράς Τάξεων (Within-class Scatter Matrix)**

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

- **Ενδοταξικη Διασπορα Προβολών**

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

$$= \mathbf{w}^T\mathbf{S}_W\mathbf{w}$$



Within-class distance

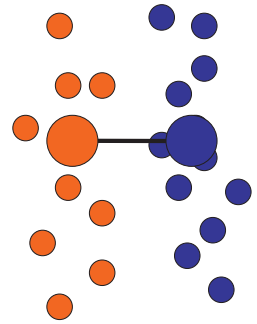
LDA: Διαταξικη Διασπορα

- **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**
- **Προβολη/μετασχηματισμος αρχικών δεδομένων :** $y = \mathbf{w}^T \mathbf{x}$
- **Εύρεση μετασχηματισμού (διανύσματος) w:**

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

- **βέλτιστη λύση από πρόβλημα γενικευμένων ιδιοτιμών:**

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

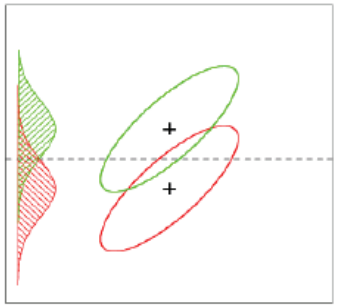
- **Βελτιστη Λυση (2 ταξεις):**

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

LDA: διαφορά από ανάλυση πρωτεύουσων συνιστωσών

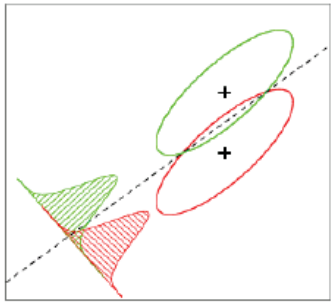
- μείωση διάστασης διατηρώντας όσο το δυνατόν την μεταβλητότητα των δεδομένων
- μέθοδος χωρίς επίβλεψη

PCA



- **ιδιοκατεύθυνση μεγιστοποίησης μεταβλητότητας**

LDA



- **ιδιοκατεύθυνση μεγιστοποίησης διάκρισης μεταξύ τάξεων**

Linear Discriminant Analysis (*C* classes)

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

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

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

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

$\tilde{\mathbf{S}}_b, \tilde{\mathbf{S}}_w$: 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 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.

$$\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \text{--->} PCA \text{--->} \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_K \end{bmatrix}$$

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

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

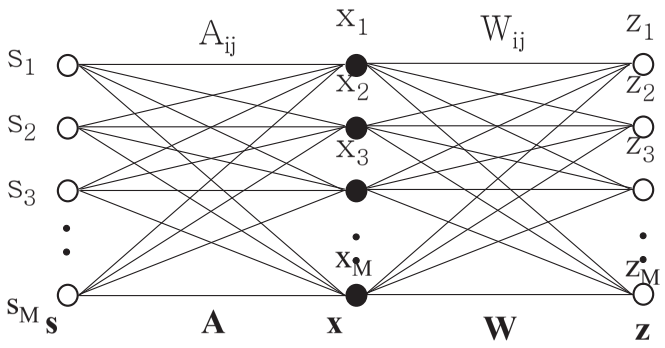
Ανάλυση Ανεξάρτητων Συνιστωσών

Independent Component Analysis

(ICA)

Independent Component Analysis

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



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

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

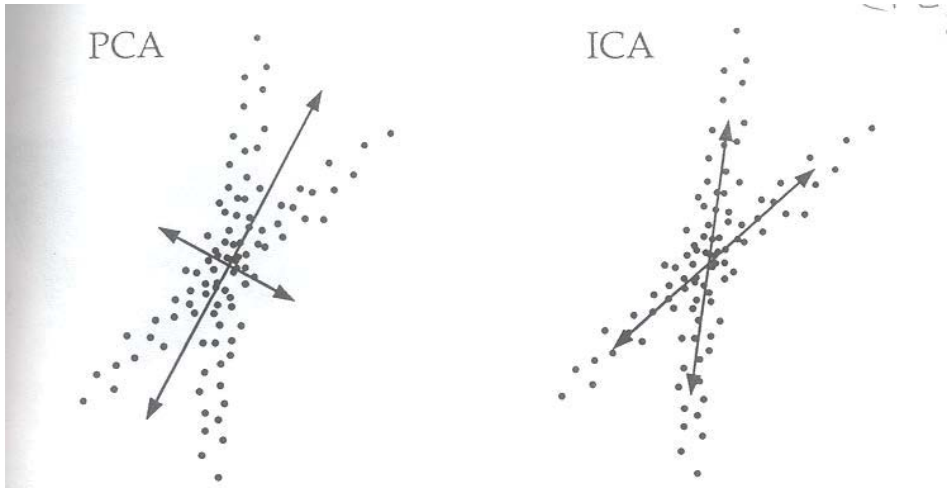


Figure 2.3. Example 2-D data distribution and the corresponding principal component and independent component axes. The data points could be, for example, grayvalues at pixel 1 and pixel 2. Figure inspired by Lewicki & Sejnowski (2000).

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)

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

Ambiguities:

- Cannot determine variances (energies) of independent components since both **A** and **s** are unknown → assume unit variances.
- Ambiguity of sign: +/-- 1.
- Cannot determine order of independent components: for any permutation matrix **P** → $\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{P}\mathbf{s}$

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

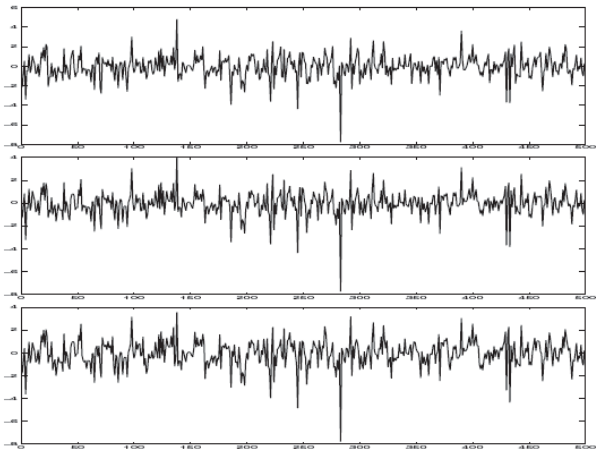


Fig. 1.2 The observed signals that are assumed to be mixtures of some underlying source signals.

Independent Components

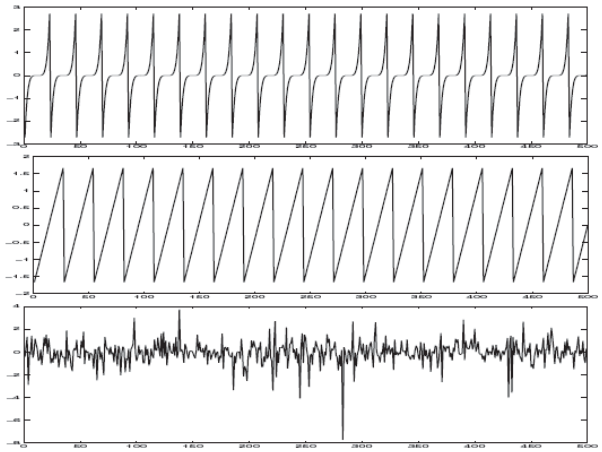


Fig. 1.3 The estimates of the original source signals, estimated using only the observed mixture signals in Fig. 1.2. The original signals were found very accurately.

Joint Distribution of Two Independent Components

$p(s_1, s_2)$

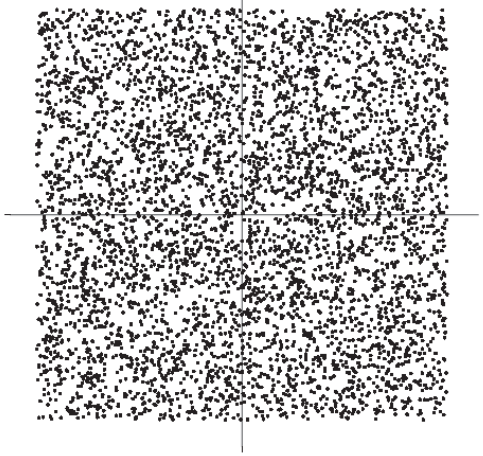
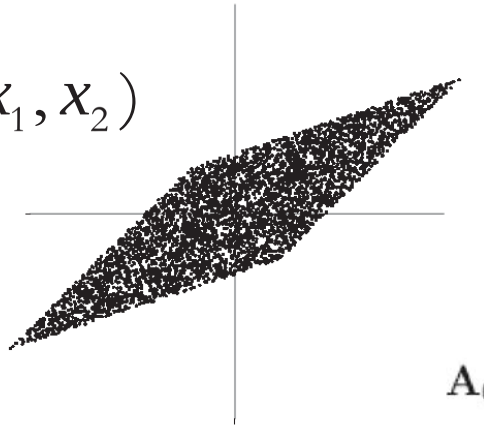


Figure 5: The joint distribution of the independent components s_1 and s_2 with uniform distributions. Horizontal axis: s_1, vertical axis: s_2.

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

Joint Distribution of Mix-2 Components

$p(x_1, x_2)$



$$\mathbf{A}_0 = \begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix}$$

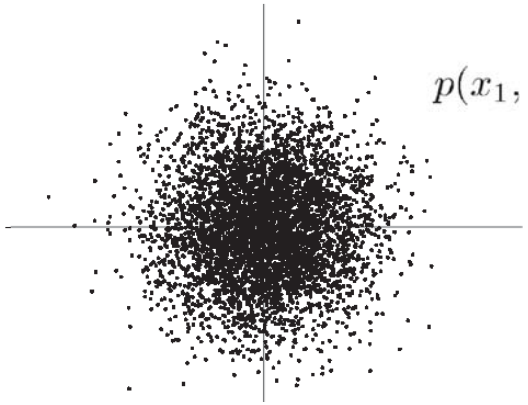
Figure 6: The joint distribution of the observed mixtures x_1 and x_2. Horizontal axis: x_1, vertical axis: x_2.

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

$$\mathbf{x} = \mathbf{A}_0 \mathbf{s}$$

Joint Distribution of Mix-2 Independent Gaussian R.V.

If mixing matrix **A** is orthogonal and original components are gaussian, then the mixed components are gaussian, uncorrelated and of unit variance



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

The multivariate distribution of two independent gaussian variables.

For gaussian r.v., we can only estimate the ICA model up to an orthogonal transformation; the mixing matrix **A** cannot be identified.

Central Limit Theorem: the distribution of a sum of independent r.v. tends to a gaussian, under certain conditions.

Thus, the sum of two independent r.v. has a distribution closer to a gaussian than any of the two original r.v.

To estimate one independent component:

$$y = \mathbf{w}^T\mathbf{x} = \mathbf{w}^T\mathbf{A}\mathbf{s} = \mathbf{z}^T\mathbf{s}$$

$\mathbf{z}^T\mathbf{s}$ is more gaussian than any of the s_i and becomes least gaussian when it in fact equals one of the s_i.

take as $\mathbf{w}$ a vector that *maximizes the nongaussianity* of $\mathbf{w}^T\mathbf{x}$

Maximizing the nongaussianity of $\mathbf{w}^T\mathbf{x}$ thus gives us one of the independent components. In fact, the optimization landscape for nongaussianity in the n-dimensional space of vectors $\mathbf{w}$ has $2n$ local maxima, two for each independent component, corresponding to s_i and $-s_i$ (recall that the independent components can be estimated only up to a multiplicative sign). To find several independent components, we need to find all these local maxima. This is not difficult, because the different independent components are uncorrelated:

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

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

(Auto- and Cross-) **Cumulants**

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

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

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

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

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

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

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

$$\mathrm{cum}(x_i, x_j, x_k, x_l) = \mathrm{E}\{x_i x_j x_k x_l\} - \mathrm{E}\{x_i x_j\}\mathrm{E}\{x_k x_l\} - \mathrm{E}\{x_i x_k\}\mathrm{E}\{x_j x_l\} - \mathrm{E}\{x_i x_l\}\mathrm{E}\{x_j x_k\}$$

PCA → 2nd-order cross-cumulant = 0

ICA → all cross-cumulants = 0

Kurtosis

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

Independent r.v. →

$$\mathrm{kurt}(x_1 + x_2) = \mathrm{kurt}(x_1) + \mathrm{kurt}(x_2)$$

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

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

nongaussianity is measured by the absolute value of kurtosis

Supergaussian and Subgaussian Distributions

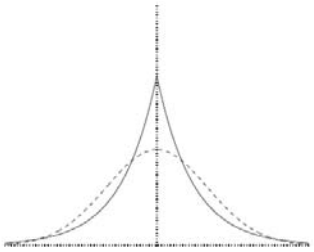


Fig. 8.9 The density function of the Laplacian distribution, which is a typical supergaussian distribution. For comparison, the gaussian density is given by a dashed curve. Both densities

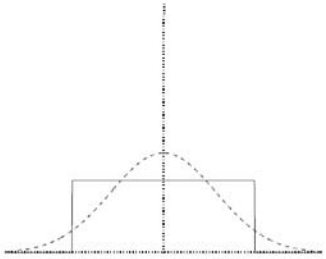


Fig. 8.10 The density function of the uniform distribution, which is a typical *sub*gaussian distribution. For comparison, the gaussian density is given by a dashed line. Both densities are normalized to unit variance.

Negentropy *J*(.)

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

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

Approximation - I: Higher Moments (r.v. ***y*** has zero mean and unit variance)

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

Approximation - II : non-quadratic *G*, standardized gaussian *v*

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

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

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

Mutual Information *I*(.)

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

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

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

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

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

Min M.I. ←→ Max sum of non-gaussianities of the estimates when the estimates are constrained to be uncorrelated.

ICA Βιβλιογραφια

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 2ης 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:
  $$\rho_i = \max \rho(\mathbf{a},\mathbf{b}) \quad \text{s.t.} \quad \mathbf{a}_j^T R_{xx}\mathbf{a} = \mathbf{b}_j^T R_{yy}\mathbf{b} = 0, j = 1,\ldots,i-1.$$
- Λύση:
  - Ζεύγη διανυσμάτων προβολών (κατευθύνσεις CCA):
    $[\mathbf{a}_1,\ldots,\mathbf{a}_r] \quad [\mathbf{b}_1,\ldots,\mathbf{b}_r], \quad \rho_i = \rho(\mathbf{a}_i,\mathbf{b}_i)$
  - Φθίνουσα ακολουθία συντελεστών συσχέτισης:
    $1 \geq \rho_1 \geq \cdots \geq \rho_r > 0, \quad r = \text{rank}(R_{xy}) \leq \min(m,n)$

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:
  $$\rho_i = \max \rho(\boldsymbol{\alpha},\boldsymbol{\beta}) = \max \boldsymbol{\alpha}^T C_{xy}\boldsymbol{\beta}$$
  $$\text{s.t. } \|\boldsymbol{\alpha}\| = \|\boldsymbol{\beta}\| = 1 \text{ and } \boldsymbol{\alpha}_j^T\boldsymbol{\alpha} = \boldsymbol{\beta}_j^T\boldsymbol{\beta} = 0, j = 1,\ldots,i-1.$$
- Λύση με τη βοήθεια της SVD του πίνακα συνεκτικότητας:
  $$C_{xy} = \sum_{k=1}^{r} \rho_k \boldsymbol{\alpha}_k \boldsymbol{\beta}_k^T, \quad \boldsymbol{\alpha}_k^T\boldsymbol{\alpha}_j = \boldsymbol{\beta}_k^T\boldsymbol{\beta}_j = \delta_{kj}, \ \rho_k \geq \rho_{k+1} > 0$$
  - Στο i-στο βήμα έχουμε: Cauchy-Schwartz, Τετραγωνική μορφή, ελλειψοειδές
    $$\boldsymbol{\alpha}^T C_{xy}\boldsymbol{\beta} = \sum_{k=i}^{r} \rho_k (\boldsymbol{\alpha}^T\boldsymbol{\alpha}_k)(\boldsymbol{\beta}^T\boldsymbol{\beta}_k) = \sum_{k=i}^{r} (\rho_k\alpha_k)\beta_k \leq \sqrt{\sum_{k=i}^{r} \rho_k^2\alpha_k^2} \leq \rho_i$$
    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:**

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

- **Συμμετρία:**

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

- **Από κοινού πληροφορία (για Gaussian κατανομές):**

$$I(\mathbf{x};\mathbf{y}) = -\frac{1}{2}\sum_{i=1}^{r}\log\left(1-\rho_i^2\right) \qquad \textbf{(bits)}$$

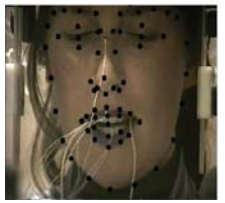
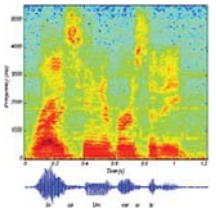
- **MSE-βέλτιστη απεικόνιση $y=Wx$ (φίλτρο Wiener):**

$$W = R_{yx}R_{xx}^{-1} = R_{yy}BPA^T$$

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

Εφαρμογές CCA: Διατροπική Πρόβλεψη

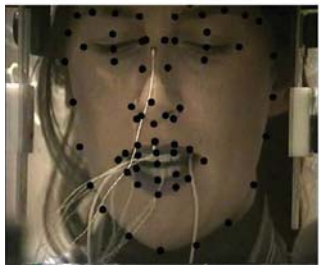
- **Παράδειγμα:** Τα ***x*** και ***y*** αντιστοιχούν σε οπτικές και ακουστικές μετρήσεις που λαμβάνονται ταυτόχρονα από έναν ομιλητή:

x=  ***y***= 

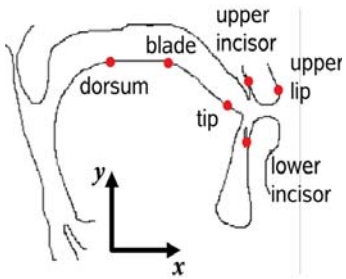
- **Εφαρμογή A2V: (Avatar)** Τεχνητό ομιλών πρόσωπο οδηγούμενο από ακουστικό σήμα.
- **Εφαρμογή V2A:** Αποθορυβοποίηση ακουστικού σήματος με χρήση οπτικής πληροφορίας.
- **Εφαρμογή AV inversion:** Ανάκτηση σχήματος φωνητικής οδού με σύμμιξη οπτικής και ακουστικής πληροφορίας.
- **Εφαρμογή σε συγχρονισμό 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.**

x= 

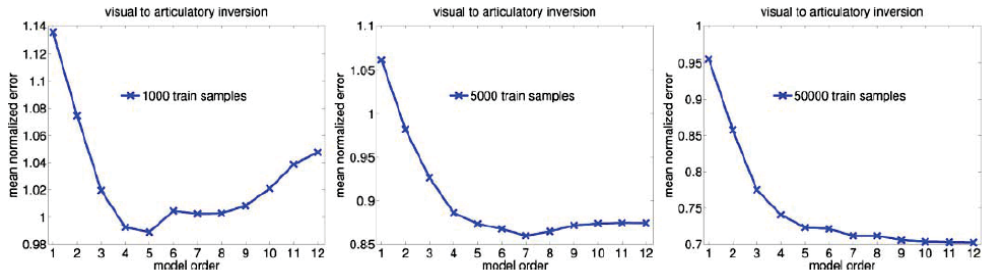
Θέση σημείων-κλειδιών στο πρόσωπο

y= 

Θέση αισθητήρων στη γλώσσα

Εφαρμογές CCA: Ανάκτηση Γεωμετρίας Φωνητικής Οδού

- Εκτίμηση στατιστικών 2ης 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.

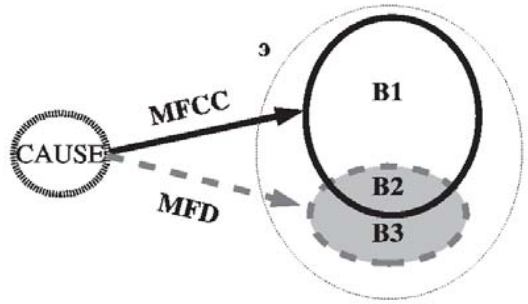
N=5.000

Άφθονα δεδομένα εκπαίδευσης. *N*=50.000

Πλήρης τάξη CCA = 12.

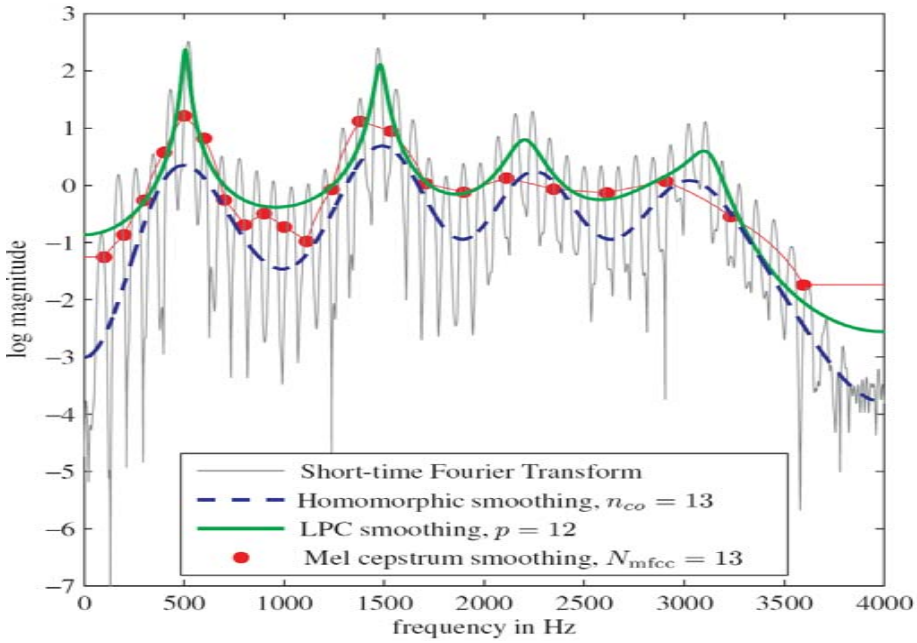
CCA μεταξύ Φράκταλ και MFCC χαρακτηριστικών: Αρχική θεώρηση

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



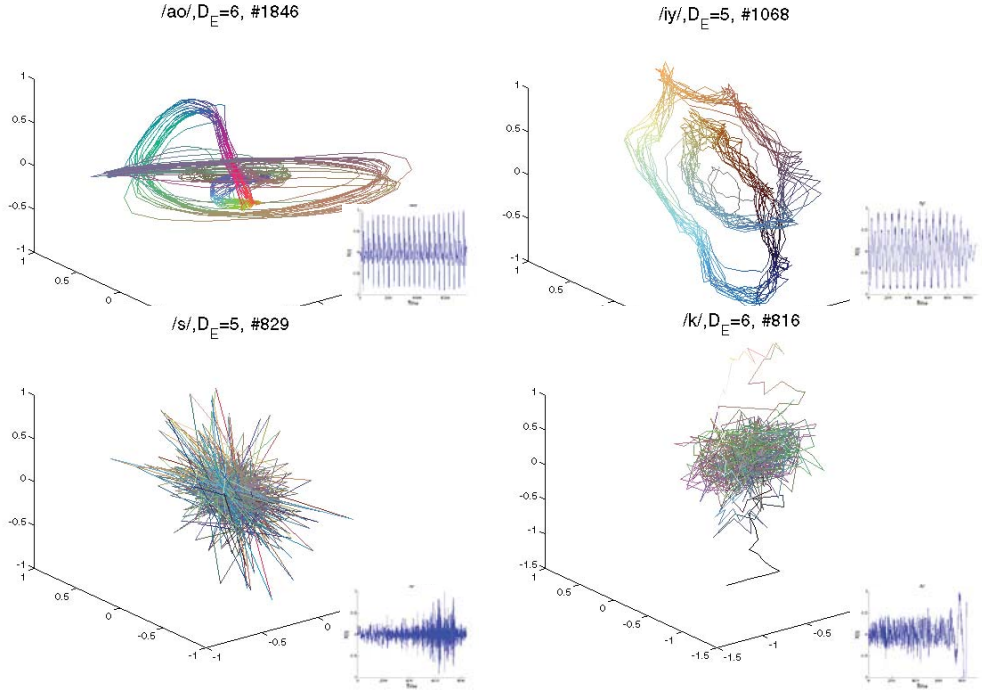
Σχήμα 5.3: *Σχηματική αναπαράσταση των υποχώρων μέγιστης και ελάχιστης συσχέτισης δυο πηγών πληροφορίας οι οποίες περιγράφουν-εξηγούν μερικώς διαφορετικές πτυχές του συνολικού χώρου ο οποίος οφείλεται σε ένα κοινό φαινόμενο-αιτία.*

Comparison of Spectral Smoothing Methods

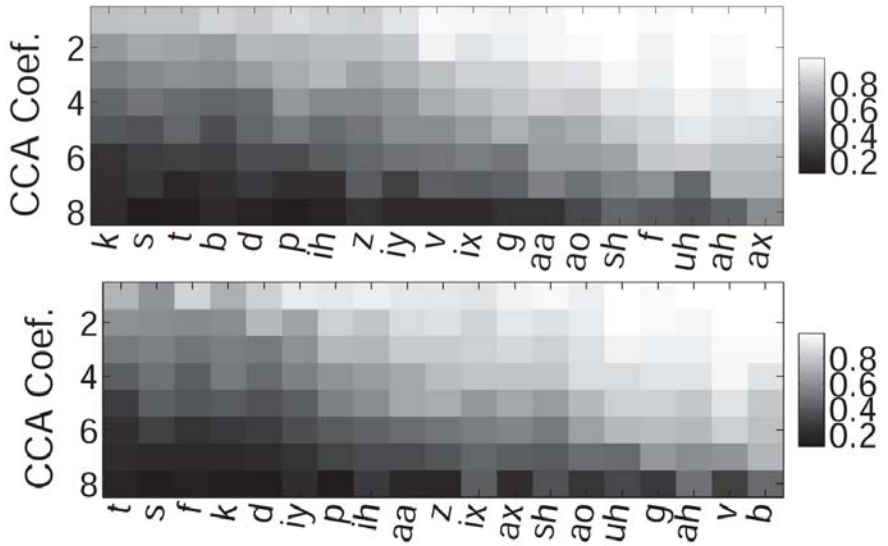


Rabiner & Schafer 2008

Reconstructed Attractors (Pitsikalis & Maragos 2002)

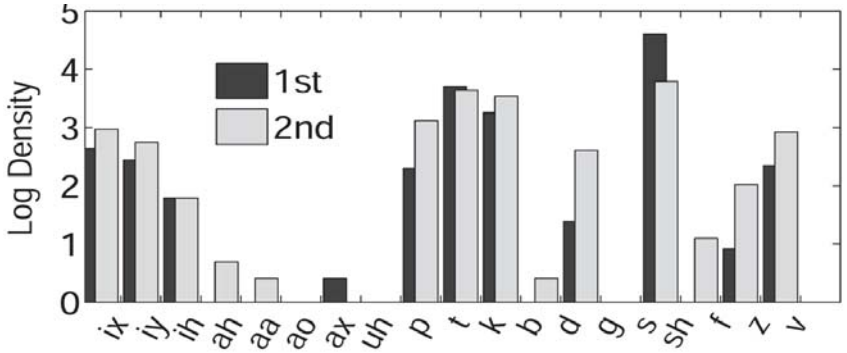


CCA μεταξύ Φράκταλ και MFCC Χαρακτηριστικών: Κατάταξη Φωνημάτων



- κριτήριο: μέσος δείκτης συσχέτισης
- κατάταξη φωνήματος ανα ομιλητή

CCA μεταξύ Φράκταλ 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.

Nonnegative Matrix Factorization

(NMF)

NMF: General Formulation

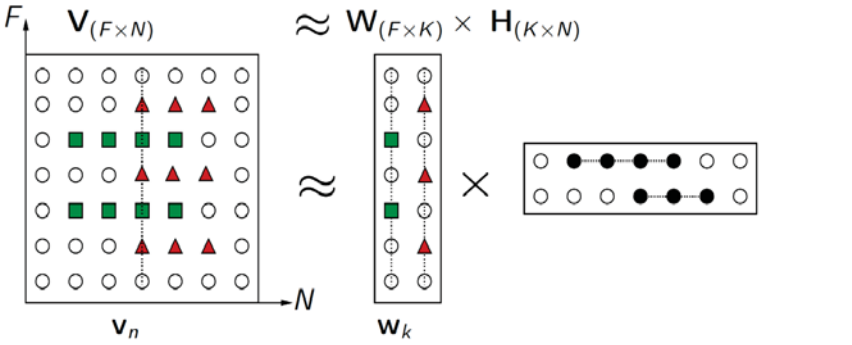


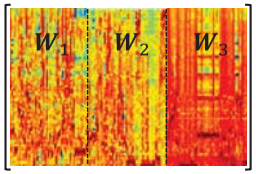
Illustration by C. Févotte

$$\mathbf{v}_n \approx \sum_{k=1}^{K} h_{kn}\mathbf{w}_k$$

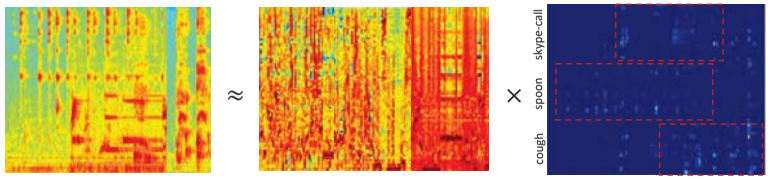
S. Essid and A. Ozerov, "Tutorial on NMF With Applications to AV Content Analysis, ICME 2014: https://perso.telecom-paristech.fr/essid/teach/NMF_tutorial_ICME-2014.pdf

Acoustic Event Detection - NMF

- **TASK:** Detection of 3 acoustic events of interest ("cough", "coffee-spoon", "skype-call")
- **Training stage:**
  - Build NMF dictionary $\boldsymbol{W}_i$ for each event-i using training data: $\boldsymbol{V}_i \approx \boldsymbol{W}_i \cdot \boldsymbol{H}_i$
  - Construct global dictionary from sub-dictionaries: $\boldsymbol{W} = [\boldsymbol{W}_1\, \boldsymbol{W}_2\, \boldsymbol{W}_3]$

$\boldsymbol{W} =$ 

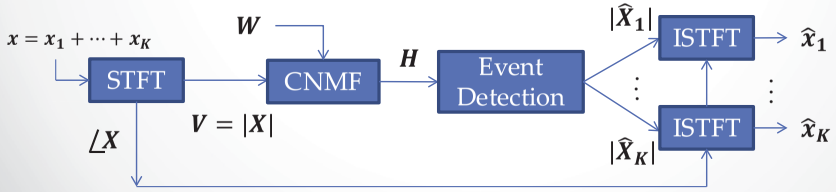
- **Testing stage:**
  - Given a test feature matrix $\boldsymbol{V}$ and global dictionary $\boldsymbol{W}$, compute activation matrix $\boldsymbol{H}$



$\boldsymbol{V}$ $\boldsymbol{W}$ $\boldsymbol{H}$

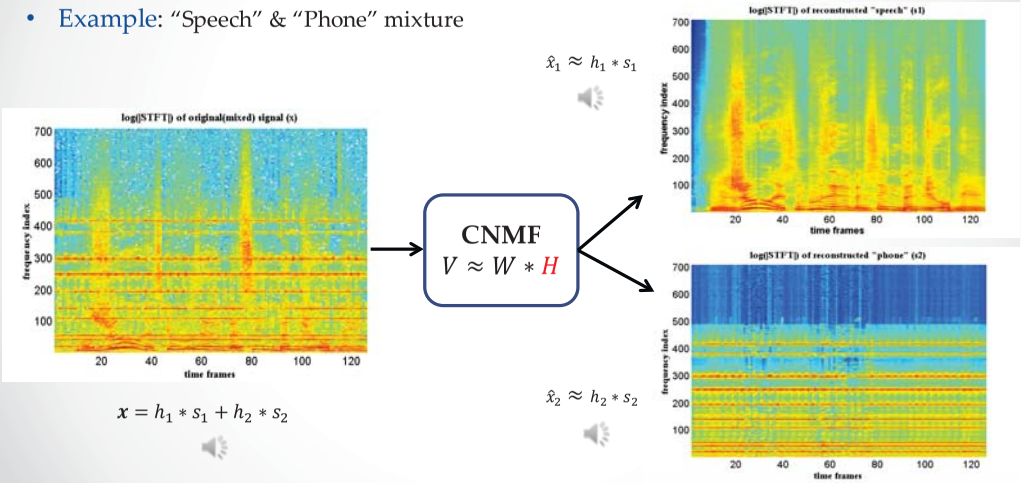
NMF in Audio Applications:
Acoustic Event
Detection & Separation (I)

- Convolutive NMF: Extension of NMF modeling temporal evolution of patterns.
- Approximate a Non-Negative matrix $\boldsymbol{V}$ as: $\mathbf{V} \approx \sum_{t=0}^{T-1} \mathbf{W}_t \cdot \overrightarrow{\mathbf{H}}^{t}$
- $\boldsymbol{W}$ is the **dictionary** containing spectral patches from different events.
- In **supervised**-CNMF, $\boldsymbol{W}$ is pre-built using **training** data of isolated acoustic events.
- $\boldsymbol{W}$ contains N sub-dictionaries, one for each event: ($\boldsymbol{W} = [\boldsymbol{W}_1\, \boldsymbol{W}_2\, \ldots\, \boldsymbol{W}_N]$).
- $\mathbf{H}$ contains the **activations** through time for each spectral patch in $\boldsymbol{W}$.
- Popular Applications:
  - **Overlapped** Acoustic Event **Detection**.
  - **Single-channel** Acoustic Event **Separation**.



NMF in Audio Applications:
Acoustic Event
Detection & Separation (II)

- Example: "Speech" & "Phone" mixture



References

- P. Smaragdis, J. C. Brown, "Non-Negative Matrix Factorization for Polyphonic Music Transcription", 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics.
  [http://www.ee.columbia.edu/~dpwe/e6820/papers/SmarB03-nmf.pdf]
  (*unsupervised NMF source separation for musical instruments*)
- P. Sprechmann, A. M. Bronstein, and G. Sapiro, "Supervised non-negative matrix factorization for audio source separation", *Excursions in Harmonic Analysis, Volume 4*. Springer International Publishing, 2015. 407-420.
  [http://vista.eng.tau.ac.il/publications/SprBroSapEHA15.pdf].
  (supervised NMF for source separation (speech from noises/speech enhancement))
- P. Giannoulis, G. Potamianos, P. Maragos, A. Katsamanis, "Improved dictionary selection and detection schemes in sparse-CNMF-based overlapping acoustic event detection", IEEE AASP Challenge on Detection and Classification of Acoustic Scenes and Events (DCASE), 2016.
  (*detection of overlapping acoustic events*)

NMF in Audio Applications: Music Source Separation (I)

Music Source Separation: The task of recovering the various vocal or instrumental sources that constitute a music signal.

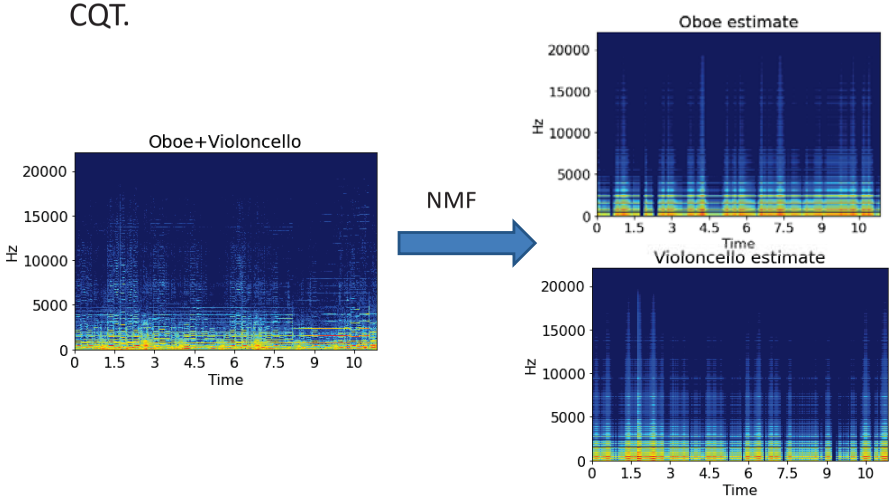
Applications:
- Audio demixing/remixing
- Automatic music transcription, instrument recognition, etc
- Suppression/removal of specific sources, eg. for karaoke systems

Non-negative matrix factorization (NMF) is suitable for source separation in music, provided the number of instruments/sources is known **a priori** since musical sources are by nature **sparse** in the frequency domain.

[1] A. Ozerov et al., Multichannel Nonnegative Matrix Factorization in Convolutive Mixtures For Audio Source Separation, in IEEE TASLP, vol. 18(3), 2010.
[2] J. Yoo et al., Nonnegative Matrix Partial Co-factorization for Drum Source Separation, ICASSP 2010.

NMF in Audio Applications: Music Source Separation (II)

We operate in a non-negative **time-frequency magnitude** signal representation (spectrogram), such as the STFT or the CQT.



NMF in Audio Applications: NMF in the era of Deep Learning

A number of **principles** presented in the original NMF algorithm have been adapted to fit into **deep learning-based methods** for audio source separation, such as:

- **Unfolding** the iterative problem of estimating the matrices W,H into a **non-negative** deep neural network architecture [1].
- Developing **non-negative autoencoder** neural networks, which combine the **unsupervised** generative approach of NMF and learning a suitable **non-negative representation** of the source signals [2].

[1] J. Le Roux et al., Deep NMF for Speech Separation, Proc. ICASSP 2015
[2] S. Venkataramani et al., End-To-End Non-Negative Autoencoders for Sound Source Separation, Proc. ICASSP 2020

PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION FOR UNSUPERVISED GRAPH CLUSTERING

Christos G. Bampis[1], Petros Maragos[2], Alan C. Bovik[1]
[1]University of Texas at Austin
[2]National Technical University of Athens

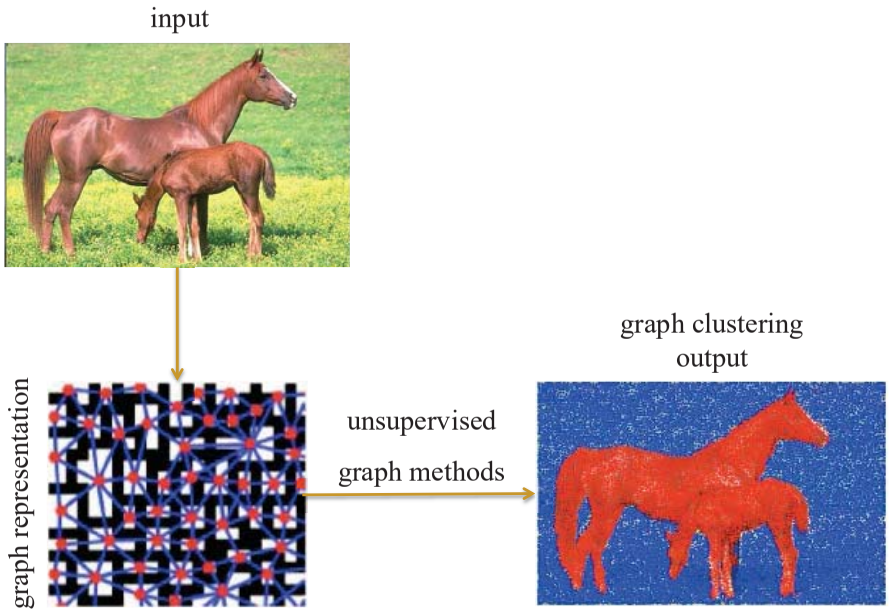
ICIP 2016

main references

- Non-Negative Matrix Factorization
  - D. D. Lee and S. H. Sebastian, "Algorithms for non-negative matrix factorization," in Proc. NIPS, 2001 (NMF)
  - Z. Yang and E. Oja, "Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization," IEEE Trans. Neural Netw., 2011. (PNMF)
  - C. Ding, X. He, and H. D. Simon, "On the equivalence of non-negative matrix factorization and spectral clustering." in Proc. of the SIAM Int'l Conf. on Data Mining (SDM), 2005.
  - C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," IEEE Trans. PAMI., 2010. (CNMF)
- Non-Negative Matrix Factorization for Image Segmentation
  - J. Yuan, D. Wang, and A. M. Cheriyadat, "Factorization-based texture segmentation," IEEE Trans. Imag. Proc., 2015. (FSeg)
- Regularized Graph-Based Factorization
  - D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," IEEE Trans. PAMI, 2011. (GNMF)

Unsupervised Graph Clustering

- seed-based segmentation methods are influenced by the seed quality (e.g. number and location).
- what if seeds are not available?
- develop unsupervised graph clustering methods for graph-based image segmentation
- similar setup as before: construct an image-driven graph
- take a step further: use matrix factorization schemes in the image feature matrix and regularized solution based on the graph structure



Non-Negative Matrix Factorization

- used for low rank data representations
- standard NMF: decompose feature matrix into non-negative matrix factors

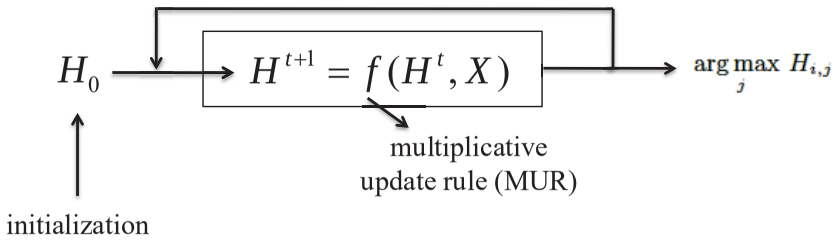
$$\min_{\mathbf{W},\mathbf{H}} \left\| \mathbf{X} - \mathbf{W}\mathbf{H}^{\top} \right\|^2, s.t.\ \mathbf{W} \geq 0, \mathbf{H} \geq 0$$

image data; cluster centroids; cluster membership

- Projective NMF - PNMF (useful for clustering):

$$\min_{\mathbf{H}} \left\| \mathbf{X} - \mathbf{X}\mathbf{H}\mathbf{H}^{\top} \right\|^2 \ s.t.\ \mathbf{H} \geq 0$$

- use Multiplicative Update Rules (MURs): assume an initial solution and solve iteratively



$\arg\max_j H_{i,j}$: each node i is assigned to the class j maximizing $H_{i,j}$

$f(.)$ implements the MUR update rule; determined using the auxiliary function technique

H^t current solution; repeat until convergence

Auxiliary Function Technique

Definition: $F(H^{t+1}, H^t)$ is an auxiliary function of $J(H^t)$ when:

- $F(H^{t+1}, H^t) \geq J(H^t)$
- $F(H^t, H^t) = J(H^t)$

Lemma: if F is an auxiliary function then $J(H)$ is non-increasing under the update rule:

$$H^{t+1} = \arg\min_H F(H^{t+1}, H^t)$$

Derivation Steps:

- define unconstrained minimization scheme using

$$\min_H \hat{J}(H) = \min_H \{J(H) + Tr(\Phi H^T)\}$$

- set $\frac{\vartheta \hat{J}}{\vartheta H} = 0$

- apply KKT conditions to determine the update rule

Graph Regularization

- images are governed by spatial regularities
- encourage those in the minimization scheme: neighboring nodes are more likely to have the same label
- (graph regularized) objective function (GR-PNMF):

$$\min_{\mathbf{H}} J(\mathbf{H}), \text{ with } J(\mathbf{H}) = J_{\text{data}}(\mathbf{H}) + \lambda J_{\text{reg}}(\mathbf{H}) \ \ s.t. \ \ \mathbf{H} \geq 0$$

data term $\left\|\mathbf{X} - \mathbf{X}\mathbf{H}\mathbf{H}^\top\right\|^2$

regularization term $\text{Tr}(\mathbf{H}^\top \mathbf{L}\mathbf{H})$

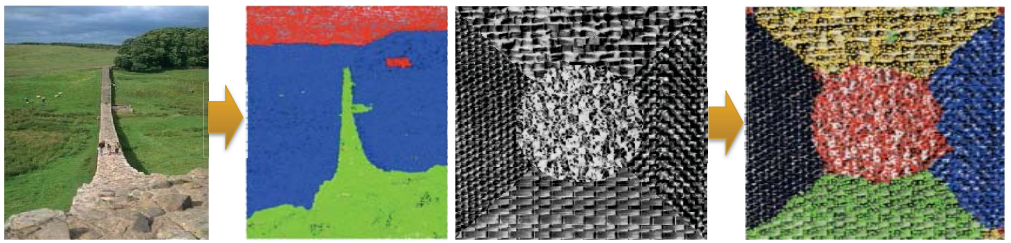
$$\mathbf{L} = \mathbf{D} - \mathbf{S}$$

graph laplacian — degree matrix — adjacency matrix

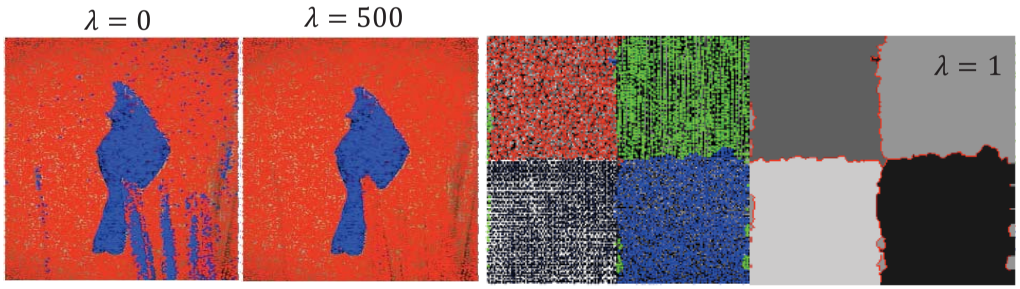
- using the auxiliary function technique to determine $f(H^t, X)$:

$$H_{ij} \leftarrow H_{ij} \sqrt[4]{\frac{[2\mathbf{X}^\top\mathbf{X}\mathbf{H} + \lambda\mathbf{S}\mathbf{H}]_{ij}}{[\mathbf{H}\mathbf{H}^\top\mathbf{X}^\top\mathbf{X}\mathbf{H} + \mathbf{X}^\top\mathbf{X}\mathbf{H}\mathbf{H}^\top\mathbf{H} + \lambda\mathbf{D}\mathbf{H}]_{ij}}}$$

- elementwise multiplies and adds, iterate until convergence
- for graph representation apply watershed and connect adjacent regions
- data matrix X: average RGB values over each region or Gabor features

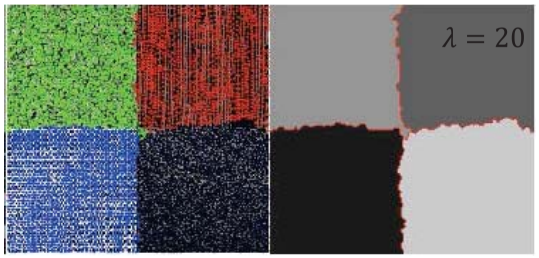


Effect of Regularization

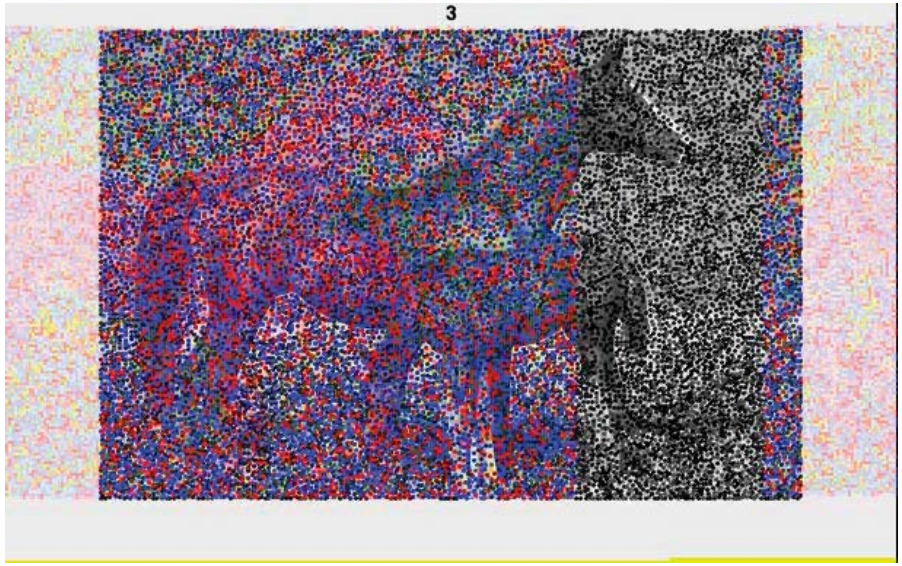


graph clustering pixel image segmentation

- larger λ enforces spatially smooth solutions
- color and/or texture segmentation
- graph and/or pixel segmentation



graph clustering pixel image segmentation



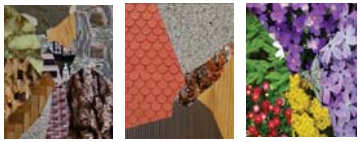
Demo: 2 classes, red and blue indicate the 2 classes

References of compared works

- Factorization-Based Texture Segmentation (**FSeg**)
  - J. Yuan, D. Wang, and A. M. Cheriyadat, "Factorization-based Texture Segmentation," IEEE Trans. Imag. Proc., 2015.
- Convex Non-negative Matrix Factorization (**CNMF**)
  - C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," IEEE Trans. PAMI, 2010.
- Orthogonal Non-negative Matrix Factorization (**ONMF**)
  - C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal non-negative matrix t-factorizations for clustering," in Proc. 12th ACM Int'l Conf. on Knowl. Discov. and Dat. Min., 2006
- Graph regularized Non-negative Matrix Factorization (**GNMF**)
  - D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," IEEE Trans. PAMI, 2011.
- Projective Non-negative Matrix Factorization (**PNMF**)
  - Z. Yang and E. Oja, "Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization," IEEE Trans. Neural Networks, 2011.
- **GRPNMF**
  - Projective non-negative matrix factorization for unsupervised graph clustering, "C. G. Bampis, P. Maragos, A. C. Bovik", in Proc. ICIP, 2016

Quantitative examples

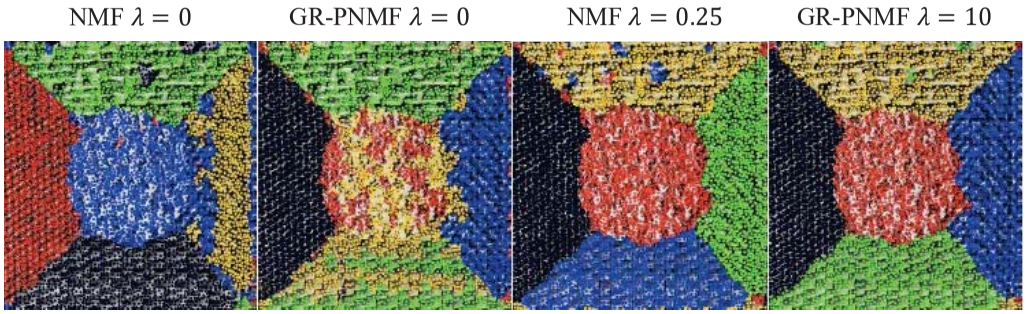
- focus on texture pixel segmentation
- compare GRPNMF with FSEG (also based on matrix factorization) and other methods on the Prague Texture Segmentation Dataset
- consists of computer generated textures:

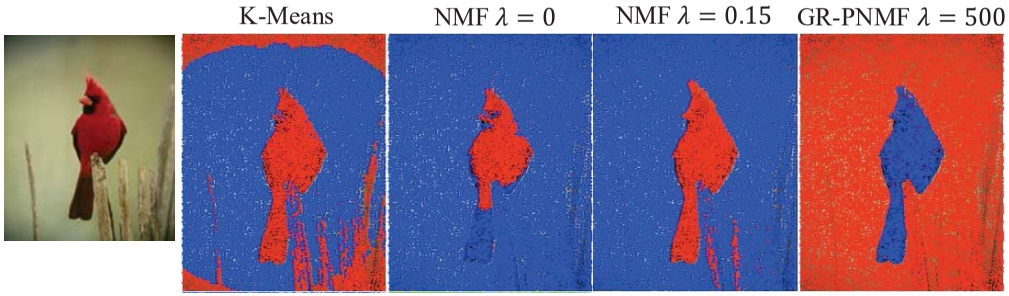


Method	CS(↑)	ME(↓)	NE(↓)	O(↓)	CA(↑)	CO(↑)	I.(↓)	EA(↑)	MS(↑)	RM(↓)	CI (↑)	GCE(↓)
FSeg [16]	69.02	6.28	5.66	10.79	77.50	**84.11**	**15.89**	83.99	78.25	4.51	84.71	10.82
CNMF [2]	49.32	**5.40**	**5.11**	36.81	60.07	70.48	29.52	66.69	57.25	11.01	67.87	**9.88**
ONMF [29]	68.33	5.81	6.30	**10.13**	77.20	83.19	16.81	84.23	78.03	4.54	85.18	11.59
GRPNMF (Our)	**69.50**	5.74	5.89	10.33	**77.92**	84.00	16.00	**84.39**	**78.57**	**4.34**	**85.24**	10.61

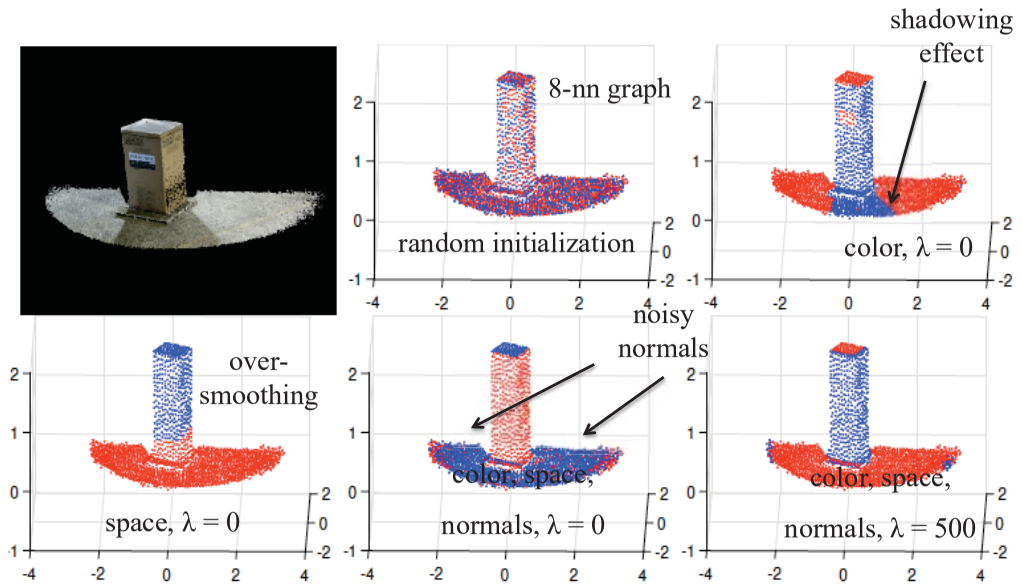
- better than other methods for some metrics but can be further improved

Qualitative Comparisons





Point Clouds



→ by combining all the features and regularizing, better results can be obtained

Ευχαριστιες για βοηθεια με τα slides:

- Ιασονας Κοκκινος
- Γιωργος Παπανδρεου
- Βασιλης Πιτσικάλης
- Παναγιωτης Γιαννουλης
- Χρηστος Γαρουφης
- Χρηστος Μπαμπης

Ιστοσελίδα Μαθήματος:

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