多變數分析
主要成分分析法、獨立成分分析法

盧家鋒 助理教授
國立陽明大學 物理治療暨輔助科技學系
alvin4016@ym.edu.tw

請先下載本週上課資料

- http://www.ym.edu.tw/~cflu
- 點選左欄〔課程資料〕
- 下載第**11**週上課資料〔**demodata_L9.zip**〕，檔案大小約**9MB**

本週課程內容

- 主要成分分析法**(principal component analysis, PCA)**
- 獨立成分分析法**(independent component analysis, ICA)**
- 生理訊號分析應用
  - 表面肌電訊號去除心跳訊號
  - 腦電波去除眼動訊號

訊號分析方法

資料輸入(data input)

濾波(filtering)
重新取樣(resampling)
多變數分析 去雜訊(denoising)
訊號切割 (epochs, blocks…)
閥值偵測 (thresholding)
訊號標準化 (normalizing)
事件偵測 (event detecting)
其他前處理 (pre-processing)

時頻分析 (time-frequency analysis)
相關性分析 (correlation/ coherence analysis)
統計分析 (statistics analysis)

資料儲存/輸出 (data output)

多變數分析

MULTIVARIATE ANALYSIS

多變數資料

- 用矩陣或向量形式表示多變數（多個量測訊號）
  - 例如：***m***個通道**(variables)**，
    資料長度包含***N***個量測值**(measurements)**的訊號資料可表示為

$$\mathbf{X}=\begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_m(1) \\ x_1(2) & x_2(2) & & x_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(N) & x_2(N) & \cdots & x_m(N) \end{bmatrix}=[x_1(t),x_2(t),\cdots,x_m(t)]$$

變數間的關係

- 變異數**(variance)**與共變異數**(covariance)**

Sample covariance (unbiased estimates)

$$\sigma(x_i,x_j)=\frac{1}{N-1}\sum_{t=1}^{N}(x_i(t)-\bar{x}_i)(x_j(t)-\bar{x}_j)$$

- 當***i*** **=** ***j***時即為單一變數的**variance**

$$\sigma(\mathbf{X})=\begin{bmatrix} \sigma(x_1,x_1) & \sigma(x_1,x_2) & \cdots & \sigma(x_1,x_m) \\ \sigma(x_2,x_1) & \sigma(x_2,x_2) & & \sigma(x_2,x_m) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(x_m,x_1) & \sigma(x_m,x_2) & \cdots & \sigma(x_m,x_m) \end{bmatrix}$$

變數間的關係

- 相關性**(correlation)**

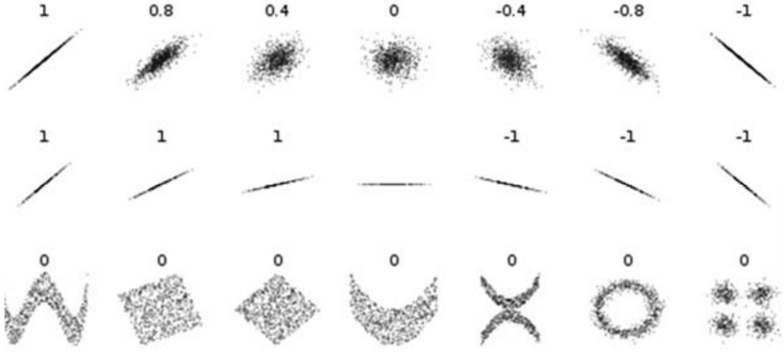
Sample correlation

$$r_{ij}=\frac{\sum_{t=1}^{N}(x_i(t)-\bar{x}_i)(x_j(t)-\bar{x}_j)}{\sqrt{\sum_{t=1}^{N}(x_i(t)-\bar{x}_i)^2\sum_{t=1}^{N}(x_j(t)-\bar{x}_j)^2}}=\frac{\sigma(x_i,x_j)}{(N-1)s_{x_i}s_{x_j}}$$

s **is the standard deviation of a variable**

雙變數資料分布與相關係數的關係

- Several sets of (x, y) points, with the Pearson correlation coefficient of x and y for each set. Note that the correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom). N.B.: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of Y is zero.



http://en.wikipedia.org/wiki/Correl

[MATLAB RULE]使用COV與CORRCOEF

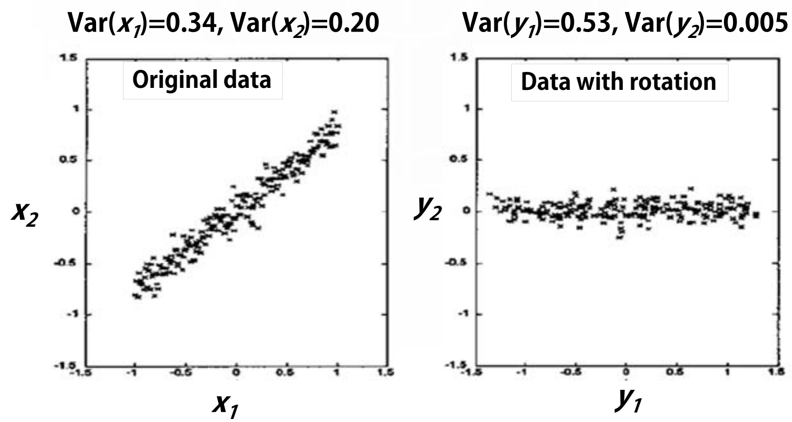
- For a matrix X, where each row is an observation, and each column is a variable, the covariance and correlation matrices can be computed as…
- help cov
  - C=cov(X)
- help corrcoef
  - [R, p]=corrcoef(X)

多變數分析 (MULTIVARIATE ANALYSIS)

- 在考量變數間與變數內的關係下，將資料重新呈現，以降低其資料量或更容易理解其意義。
  - hidden, latent data
- 透過線性轉換(linear transformation)對原始資料作處理
  - 例如旋轉或是縮放

$$\begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_k(t) \end{bmatrix} = \mathbf{W} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{bmatrix}$$

舉例說明



主要成分分析法

PRINCIPAL COMPONENT ANALYSIS, PCA

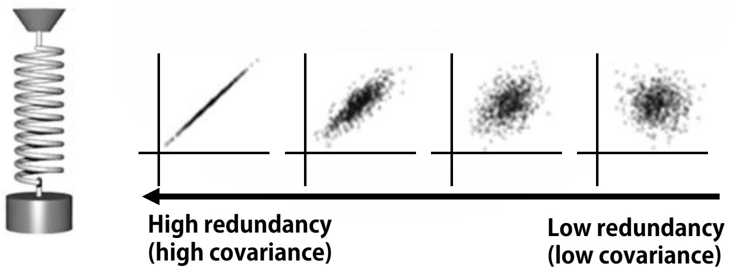
資料量(DATA)與資訊量(INFORMATION)

- 龐大的資料量不等於龐大的資訊量，通常資料中會夾帶著許多不必要的雜訊或是意義較低或重複性高的資訊。
- 主要成分分析法旨在
  - 將資料透過旋轉與線性混和的概念，找到新的**basis**來更有效地描述資訊
  - 在失去最少資訊的條件下，盡可能降低變數數目。
  - 將資料中資訊量較低的成分或雜訊剔除。

The "representation" is often sought as a linear transformation of the original data.

資料冗餘(REDUNDANCY)

- 舉例：以雙變數資料為例
  - 是否需要用兩個軸向(變數)的資料來描述一系統？



共變異數矩陣(COVARIANCE MATRIX)

- 我們希望經過線性轉換後的共變異數矩陣之…
  - 對角元素(**diagonal elements**)，數值越大➔包含越多資訊
  - 非對角元素(**off-diagonal elements**)，數值越小➔冗餘量越低(即各變數間互不相關)

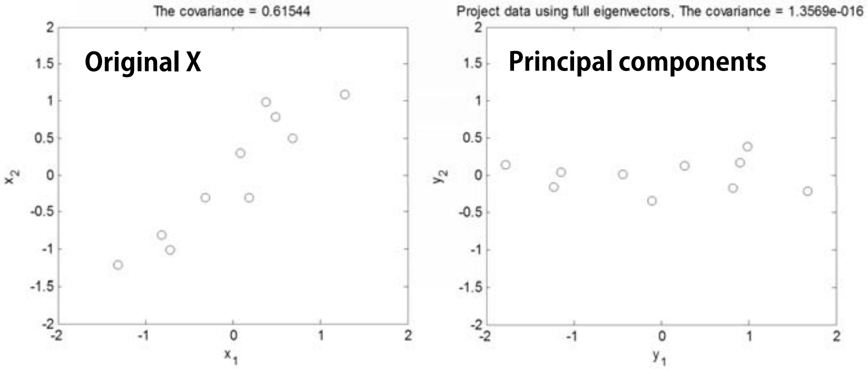
$$\sigma(\mathbf{Y}) = \begin{bmatrix} \sigma(y_1, y_1) & 0 & \cdots & 0 \\ 0 & \sigma(y_2, y_2) & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma(y_m, y_m) \end{bmatrix}$$

Diagonal matrix

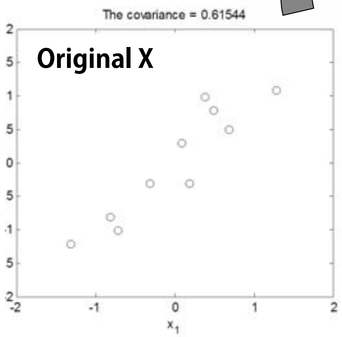
對角化➔使用奇異值分解(**eigendecomposition**)!!

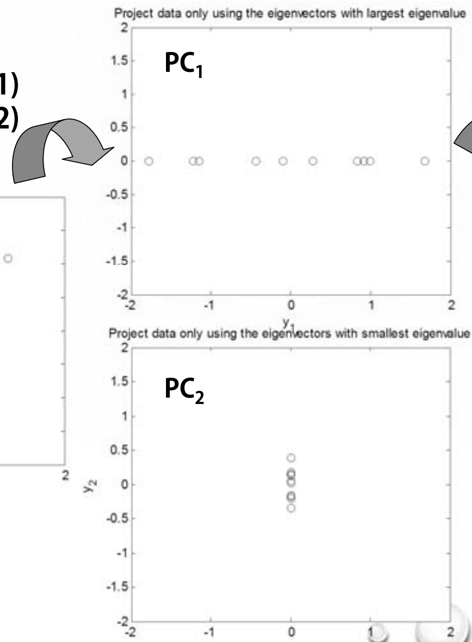
實例說明

- 請開啟並執行**demodata_L9\example_PCA.m**

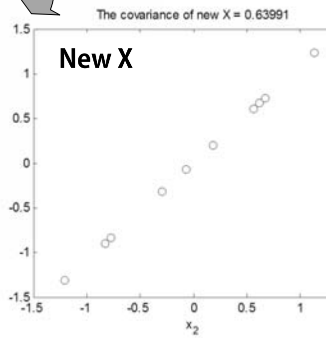


PC_1= X * eigVec(:,1)
PC_2= X * eigVec(:,2)





newX = PC_1 * eigVec(:,



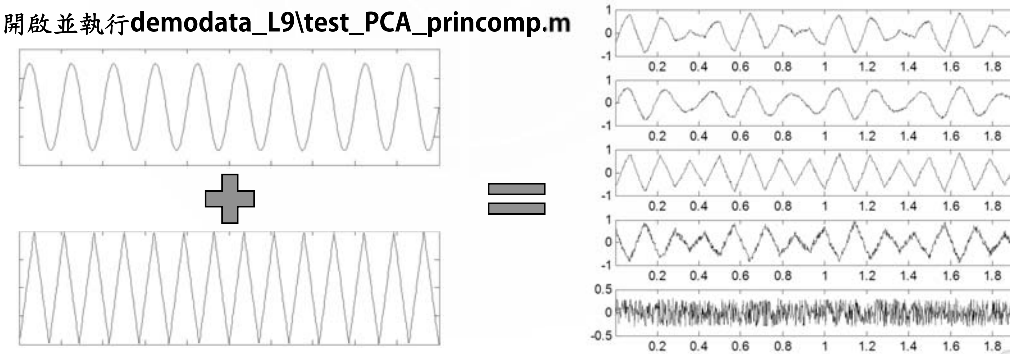
實例說明

[MATLAB RULE]使用PRINCOMP

- **help princomp**
  - **[eigenVec, PC, eigenVal]=princomp(X);**
    - **X is an N-by-P data matrix with N observations and P variables.**
    - **eigenVec: eigenvectors of cov(X)**
    - **eigenVal: eigenvalues of cov(X)**
    - **PC: principal components of X**

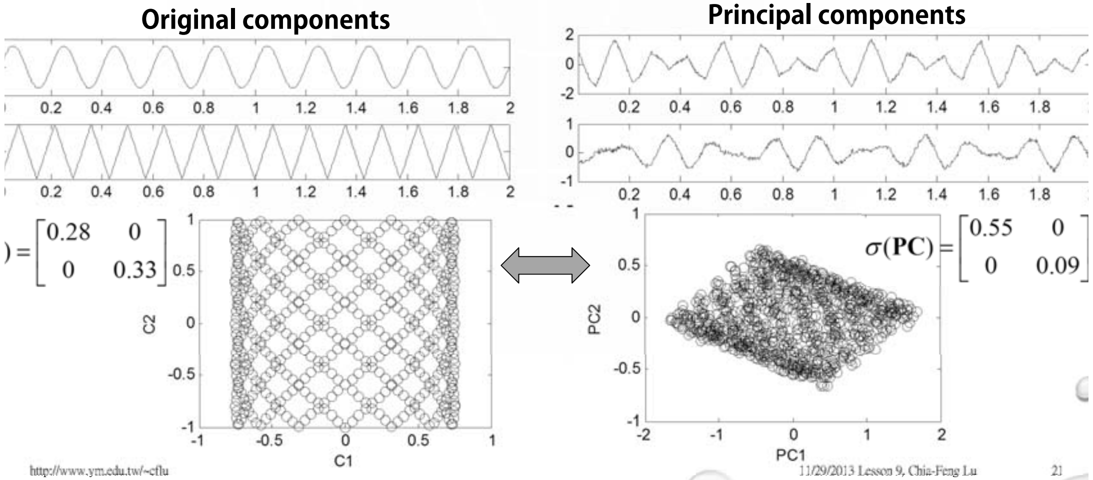
使用PRINCOMP進行PCA分析與去雜訊

- 請開啟並執行**demodata_L9\test_PCA_princomp.m**



- 請開啟並執行**demodata_L9\test_PCA_SVD.m (←有興趣的人可自行比較)**

比較UNCORRELATED與INDEPENDENT



獨立成分分析法

INDEPENDENT COMPONENT ANALYSIS, ICA

雞尾酒會模型(A COCKTAIL-PARTY PROBLEM)

$$x_1(t) = 0.7s_1(t) + 0.2s_2(t) + 0.1s_3$$
$$x_2(t) = 0.3s_1(t) + 0.4s_2(t) + 0.3s_3$$
$$x_3(t) = 0.1s_1(t) + 0.2s_2(t) + 0.7s_3$$

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{bmatrix} = \mathbf{A} \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_k(t) \end{bmatrix}$$

A is the mixing matrix

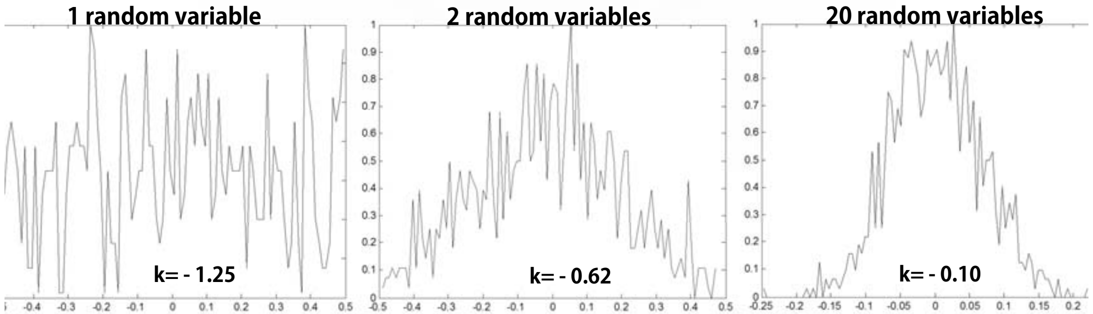
umber of microphones >= number of speakers

獨立成分分析法

- 與**PCA**不同，目標不在降低資料變數維度，而是盡可能從混合訊號中找出更具生理或物理意義的訊號來源。
- 假設：
  - 訊源**S**都是各自獨立的變數
  - 各個訊源變數的分布都是非高斯分布**(non Gaussian)**
- 限制：
  - 無法計算出訊源**S**的變異數或能量強度
    （如：可求解混和比例，但不知麥克風與人的距離）
  - 無法解出訊源**S**的正負號。

中央極限定理(CENTRAL LIMIT THEOREM)

- 數個非高斯分布的獨立變數之混和，其混和訊號的分布會趨向高斯分布。
- 請開啟並執行**demodata_L9\test_CenLimit.m**

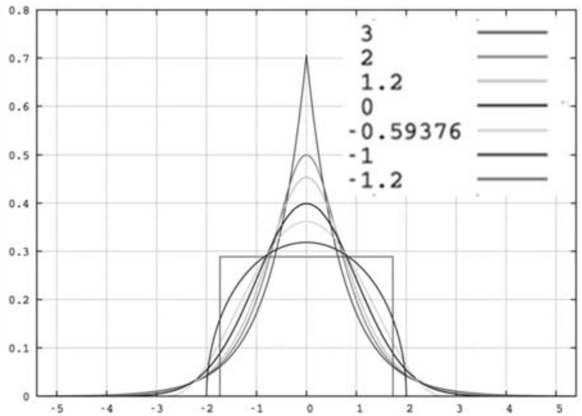


量測非高斯程度

- **Kurtosis = 0 ➔ Gaussian**
- **Kurtosis > 0 ➔ Super-Gaussian**
- **Kurtosis < 0 ➔ Sub-Gaussian**

$$k = \frac{E(x-\mu)^4}{\sigma^4} - 3$$

- 也可用**negentropy**等其他量測值



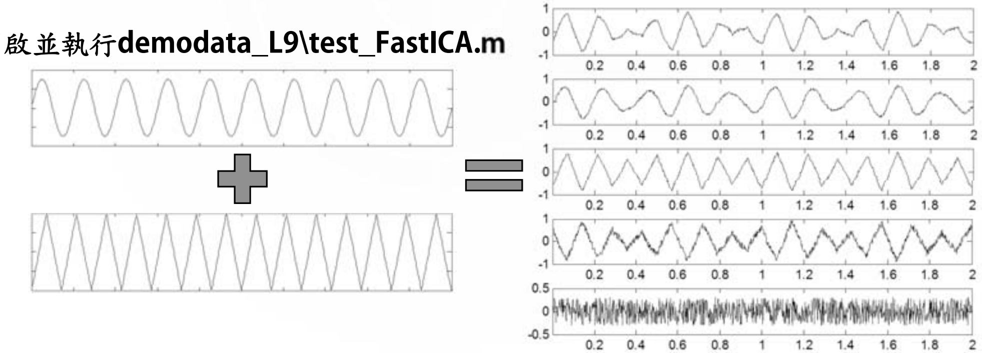
獨立成分分析法處理流程

- **Centering data (remove mean)**
- **Whitening process (sphere data)**
  - **Decorrelate variables**
  - **Scale variables so that their variance = 1**
- **Optimization algorithm to maximize non-Gaussianity of each source**

Whitening + non-Gaussianity ➔ independent

使用FASTICA找出獨立成分

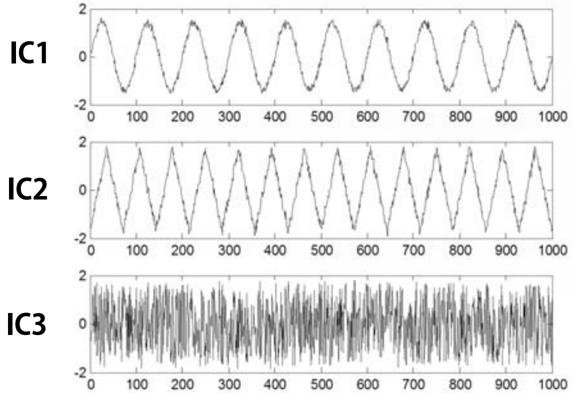
- 請開啟並執行**demodata_L9\test_FastICA.m**



- **[icasig, A, W] = fastica(X,'numOfIC',ICNo,'displayMode','off','firstEig',1,'lastEig', PCNo);**
- **fasticag(X)**

使用FASTICA找出獨立成分

- 請開啟並執行demodata_L9\test_FastICA.m



重複執行幾次程式，觀察

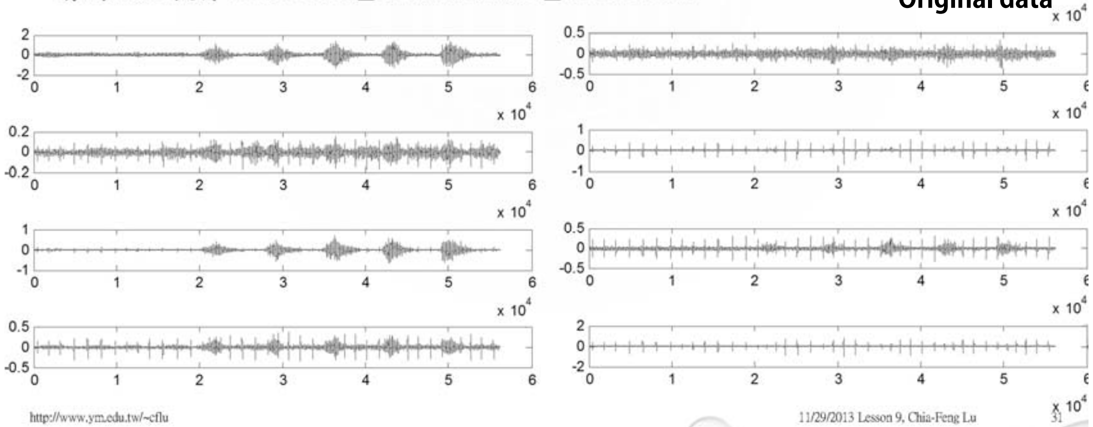
- PCNo對IC的影響
- IC的順序不會固定
- IC的正負值也不固定

生理訊號分析應用

以呼吸肌SURFACE EMG訊號去除雜訊為例

- 請開啟並執行demodata_L9\EMG\EMG_FastICA.m

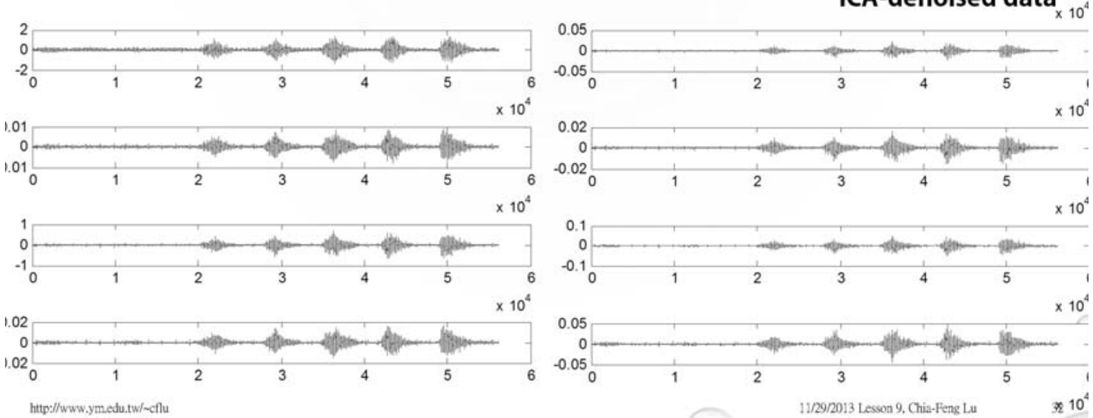
Original data



以呼吸肌SURFACE EMG訊號去除雜訊為例

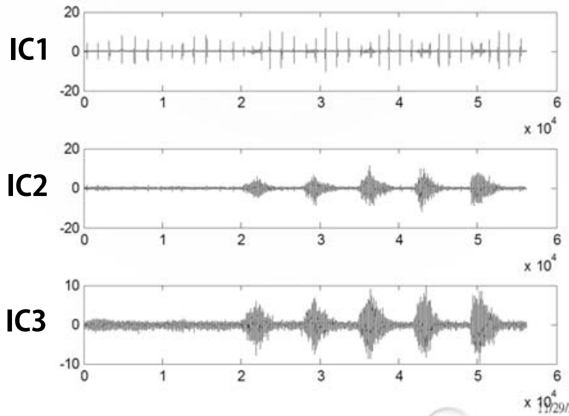
- 請開啟並執行demodata_L9\EMG\EMG_FastICA.m

ICA-denoised data



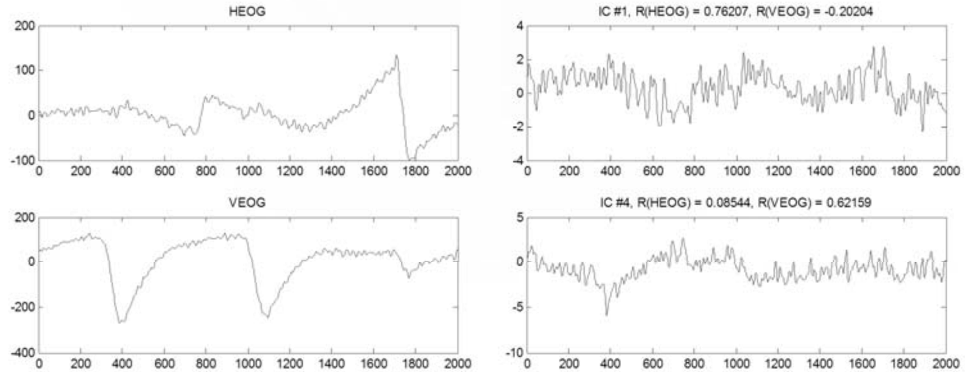
以呼吸肌SURFACE EMG訊號去除雜訊為例

- 請開啟並執行**demodata_L9\EMG\EMG_FastICA.m**



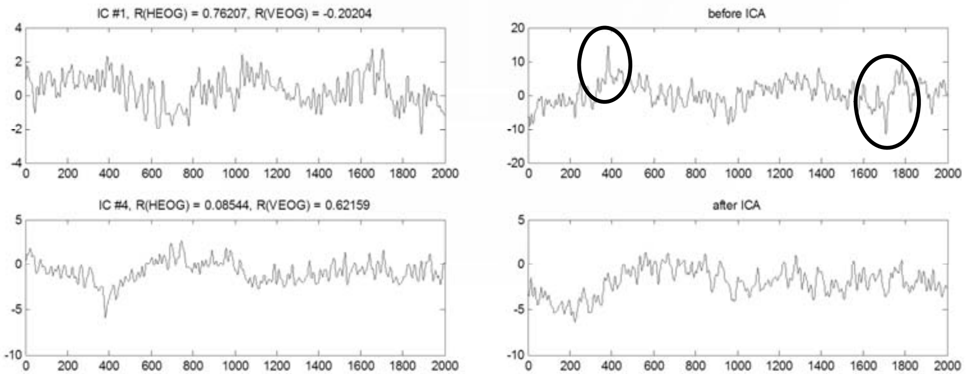
以EEG腦電波訊號去除眼動雜訊為例

- 請開啟並執行**demodata_L9\EEG_FastICA.m**



以EEG腦電波訊號去除眼動雜訊為例

- 請開啟並執行**demodata_L9\EEG_FastICA.m**



PCA & ICA參考資料

- **A. Hyvärinen, J. Karhunen, E. Oja (2001) Independent Component Analysis. John Wiley & Sons.**

THE END

http://www.ym.edu.tw/~cflu