

Técnicas de Análisis Estadístico Multivariado Basadas en Estadísticos de Segundo Orden

Tratamiento Avanzado de Señal en Comunicaciones

Curso 2009-2010



Contenido

- 1 Introducción
- 2 Análisis de Componentes Principales (PCA)
- 3 Mínimos Cuadrados Parciales (PLS)
- 4 Regresión Lineal Múltiple (MLR)
- 5 Análisis de Correlaciones Canónicas (CCA)
- 6 Resumen

Contenido

- 1 **Introducción**
- 2 Análisis de Componentes Principales (PCA)
- 3 Mínimos Cuadrados Parciales (PLS)
- 4 Regresión Lineal Múltiple (MLR)
- 5 Análisis de Correlaciones Canónicas (CCA)
- 6 Resumen

Técnicas Basadas en Estadísticos de Segundo Orden (SOS)

- Relaciones estadísticas en datos multidimensionales
 - Estructura interna
 - Relación entre diferentes tipos de datos
 - Asumimos datos reales. Extensión (**casi**) trivial a complejos
- Métodos basados en estadísticos de segundo orden (SOS)
 - Únicamente utilizaremos correlaciones
 - Óptimo para datos Gaussianos
 - A menudo resultan en técnicas lineales
- Formulación Estocástica
 - Asumimos datos sin media (correlación = covarianza)
 - Asumimos conocidas las matrices de correlación

$$\mathbf{R}_{\mathbf{x},\mathbf{y}} = E \left[\mathbf{xy}^T \right]$$

- Casi todos los resultados presentados se pueden reformular desde un punto de vista determinista

Herramientas Matemáticas

● Álgebra

- **SVD**: $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$
- **EVD**: $\mathbf{A} = \mathbf{X}\mathbf{\Sigma}\mathbf{X}^{-1}$

- Para matrices simétricas ($\mathbf{A}^T = \mathbf{A}$) semidefinidas positivas SVD y EVD coinciden

● Propiedades:

- Traza del producto

$$\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA})$$

- Optimización

$$\arg \max_{\mathbf{U}, \mathbf{V}} \text{Tr}(\mathbf{U}^T \mathbf{A} \mathbf{V}) \quad \text{suje to a} \quad \mathbf{U}^T \mathbf{U} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$$

La solución viene dada (hasta una rotación) por los vectores singulares de la matriz \mathbf{A}

● Derivadas Vectoriales (Operador Gradiente $\nabla_{\mathbf{A}}$)

$$\nabla_{\mathbf{A}} \text{Tr}(\mathbf{A}^T \mathbf{X} \mathbf{A}) = 2\mathbf{X} \mathbf{A}$$

$$\nabla_{\mathbf{A}} \text{Tr}(\mathbf{A}^T \mathbf{X}) = \mathbf{X}$$

Contenido

- 1 Introducción
- 2 Análisis de Componentes Principales (PCA)**
- 3 Mínimos Cuadrados Parciales (PLS)
- 4 Regresión Lineal Múltiple (MLR)
- 5 Análisis de Correlaciones Canónicas (CCA)
- 6 Resumen

Presentación del Problema

- Supongamos que disponemos de datos \mathbf{x} en un espacio n -dimensional ($\mathbf{x} \in \mathbb{R}^{n \times 1}$)
- Nuestro objetivo consiste en encontrar una proyección $\tilde{\mathbf{x}} \in \mathbb{R}^{p \times 1}$ p -dimensional ($p \leq n$) que permita recuperar los datos originales \mathbf{x} con el mínimo error cuadrático medio posible

● Definiciones

- Datos: $\mathbf{x} \in \mathbb{R}^{n \times 1}$
- Proyección: $\tilde{\mathbf{x}} = \mathbf{U}^T \mathbf{x} \quad \tilde{\mathbf{x}} \in \mathbb{R}^{p \times 1}$
- Proyector: $\mathbf{U} \in \mathbb{R}^{n \times p}$
- Reconstrucción de los datos:

$$\hat{\mathbf{x}} = \mathbf{V} \tilde{\mathbf{x}} \quad \hat{\mathbf{x}} \in \mathbb{R}^{n \times 1}$$

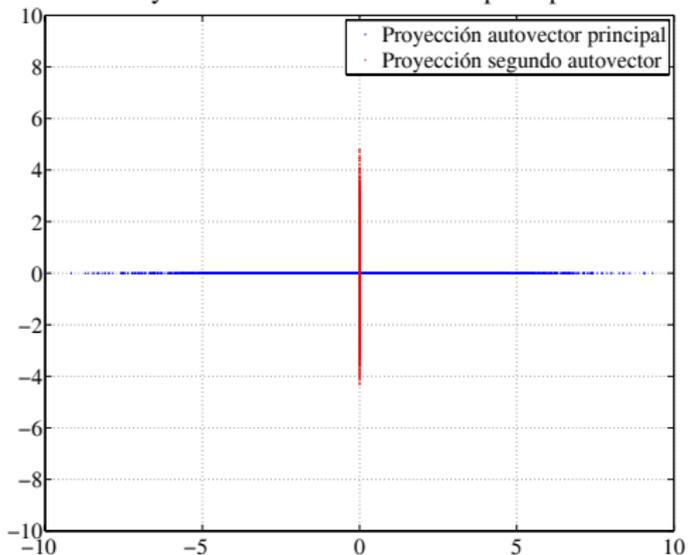
- Matriz de reconstrucción: $\mathbf{V} \in \mathbb{R}^{n \times p}$
- **Ambigüedad:** $\hat{\mathbf{x}} = \mathbf{V} \mathbf{U}^T \mathbf{x} = \mathbf{V} \mathbf{A} \mathbf{A}^{-1} \mathbf{U}^T \mathbf{x} \quad \forall \mathbf{A} \in \mathbb{R}^{p \times p} \text{ invertible}$
- Sin pérdida de generalidad asumiremos

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}$$

Ejemplo con datos bidimensionales

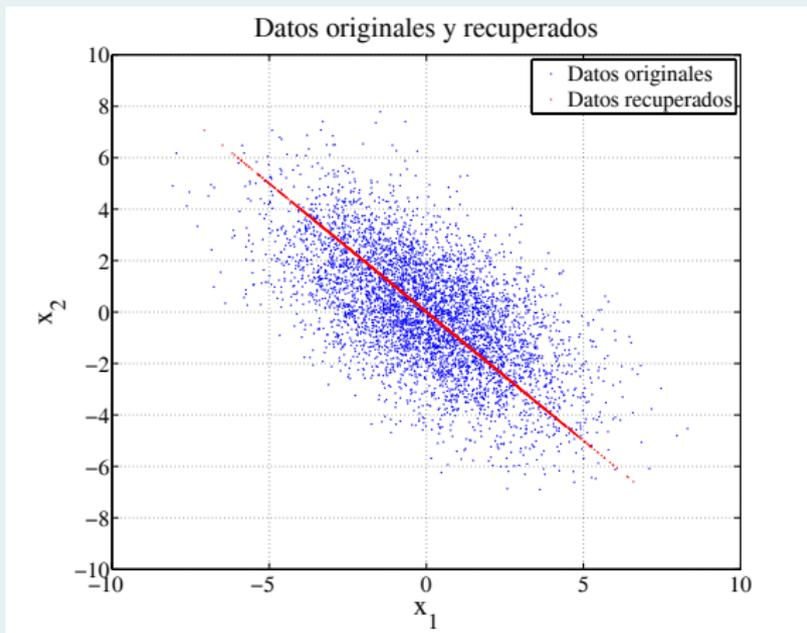
$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$$

Proyección sobre las direcciones principales



Ejemplo con datos bidimensionales

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix}$$



Interpretación

- Las proyecciones proporcionadas por PCA constituyen la mejor **representación p -dimensional** de los datos originales (n -dimensionales)
- Las proyecciones PCA son las de **máxima energía**
- **Ortogonalidad:**^a
 - **Proyectores ortogonales:** $\mathbf{u}_k^T \mathbf{u}_l = 0 \quad k \neq l$
 - **Proyecciones ortogonales:** $E[\tilde{x}_k \tilde{x}_l] = \mathbf{u}_k^T \mathbf{R}_{\mathbf{x}, \mathbf{x}} \mathbf{u}_l = 0$
- Gracias a la ortogonalidad los proyectores se pueden obtener de manera sucesiva (deflación)
 - Dada la solución PCA p -dimensional, para obtener la solución $p + 1$ -dimensional bastará con añadir un nuevo proyector: \mathbf{u}_{p+1}

^a \mathbf{u}_k representa la k -ésima columna de \mathbf{U} . $\tilde{x}_k = \mathbf{u}_k^T \mathbf{x}$ es el k -ésimo elemento de $\tilde{\mathbf{x}}$

Aplicación 1: Estima ML con Ruido Gaussiano

- Consideremos el modelo de señal

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}$$

donde $\mathbf{x} \in \mathbb{R}^{n \times 1}$ son la observaciones y $\mathbf{n} \in \mathbb{R}^{n \times 1} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$

- Queremos estimar $\mathbf{s} \in \mathbb{R}^{p \times 1}$, $\mathbf{A} \in \mathbb{R}^{n \times p}$ y la varianza σ^2
- Verosimilitud de las observaciones:

$$f(\mathbf{x}|\mathbf{A}, \mathbf{s}, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\|\mathbf{x} - \mathbf{A}\mathbf{s}\|^2}{2\sigma^2}}$$

- **Solución:**

- $\hat{\mathbf{A}} = \mathbf{U}\mathbf{B}$ ($\mathbf{B} \in \mathbb{R}^{p \times p}$ representa una ambigüedad)
- $\hat{\mathbf{s}} = \mathbf{B}^{-1}\mathbf{U}^T\mathbf{x} = \mathbf{B}^{-1}\tilde{\mathbf{x}}$
- $\hat{\mathbf{x}} = \hat{\mathbf{A}}\hat{\mathbf{s}}$ es la representación PCA p -dimensional de \mathbf{x}
- $\hat{\sigma}^2$ se obtiene promediando los $n - p$ menores autovalores de $\mathbf{R}_{\mathbf{x}, \mathbf{x}}$

Aplicación 2: Rate-Distortion Theory

- Consideremos $\mathbf{x} \in \mathbb{R}^{n \times 1} \sim \mathcal{N}(0, \mathbf{R}_{\mathbf{x}, \mathbf{x}})$
- Queremos averiguar cual es la tasa R (nº medio de bits) necesaria para recuperar los datos $\hat{\mathbf{x}}$ con una distorsión

$$J = E [\|\mathbf{x} - \hat{\mathbf{x}}\|^2] \leq D$$

- **Solución:**^a **Reverse Water-Filling** sobre las proyecciones PCA

$$R(D) = \sum_{k=1}^n \frac{1}{2} \log_2 \frac{\lambda_k}{D_k}$$

donde

$$D_k = \begin{cases} \nu, & \text{si } \nu < \lambda_k \\ \lambda_k, & \text{si } \nu \geq \lambda_k \end{cases}$$

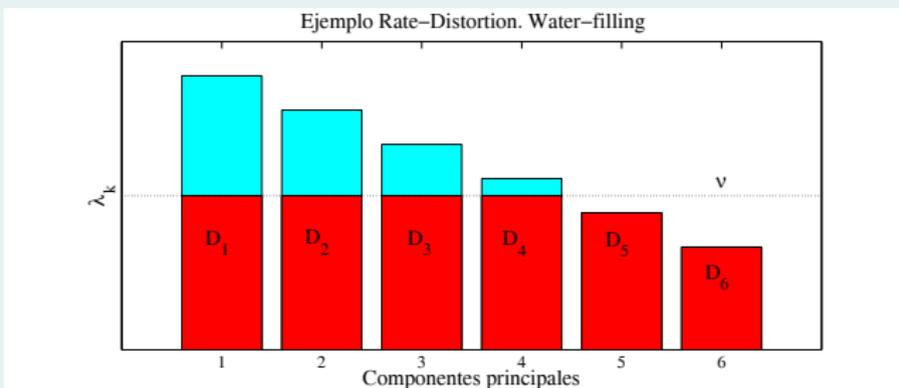
y ν se escoge de manera que $\sum_{k=1}^n D_k = D$

^aCover and Thomas, "Elements of information theory", Wiley-Interscience, 1991

Aplicación 2: Rate-Distortion Theory (Continuación)

$$R(D) = \sum_{k=1}^n \frac{1}{2} \log_2 \frac{\lambda_k}{D_k} \quad D_k = \begin{cases} \nu, & \text{si } \nu < \lambda_k \\ \lambda_k, & \text{si } \nu \geq \lambda_k \end{cases}$$

- λ_k es el k -ésimo autovalor de $\mathbf{R}_{\mathbf{x}, \mathbf{x}}$, es decir la varianza de $\tilde{\mathbf{x}}_k = \mathbf{u}_k^T \mathbf{x}$
- D_k es la distorsión asociada a $\tilde{\mathbf{x}}_k$ ($E[(\tilde{x}_k - \hat{\tilde{x}}_k)^2]$)
- ν establece una distorsión máxima
- $R_k = \frac{1}{2} \log_2 \frac{\lambda_k}{D_k}$ es el nº medio de bits con los que se representará \tilde{x}_k



Algoritmo adaptativo

- ¿Como realizar PCA cuando se dispone de los datos secuencialmente?
- **Regla de Oja**: Extracción del autovector principal

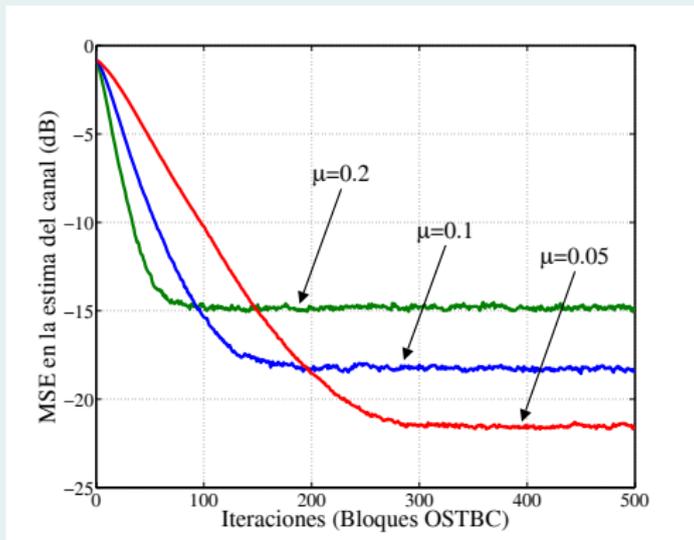
$$\mathbf{u}_1[n+1] = \mathbf{u}_1[n] + \mu \tilde{x}_1[n] [\mathbf{x}[n] - \tilde{x}_1[n] \mathbf{u}_1[n]]$$

- $\tilde{x}_1[n] = \mathbf{u}_1^T[n] \mathbf{x}[n]$
- El término $\tilde{x}_1^2[n] \mathbf{u}_1[n]$ fuerza la restricción $\|\mathbf{u}_1[n]\| = 1$
- El **learning rate** μ establece un compromiso entre la velocidad de convergencia/tracking y la precisión del algoritmo
- Filosofía análoga al **LMS** (least mean squares)
- Los siguientes autovectores se pueden extraer mediante deflación

$$\mathbf{u}_k[n+1] = \mathbf{u}_k[n] + \mu \tilde{x}_k[n] \left[\mathbf{x}[n] - \tilde{x}_k[n] \mathbf{u}_k[n] - 2 \sum_{l < k} \tilde{x}_l[n] \mathbf{u}_l[n+1] \right]$$

Ejemplo

- Estima ciega de canales MIMO-STBC:^a $\mathbf{x}[n] = \mathbf{W}(\mathbf{H})\mathbf{s}[n] + \mathbf{n}[n]$



^aJ. Vía, I. Santamaría, J. Pérez y D. Ramírez, "Blind decoding of MISO-OSTBC systems based on principal component analysis", IEEE Int. Conf. on Acoust., Speech, and Signal Processing (ICASSP 2006), Bordeaux, France, Mayo 2006

Contenido

- 1 Introducción
- 2 Análisis de Componentes Principales (PCA)
- 3 Mínimos Cuadrados Parciales (PLS)**
- 4 Regresión Lineal Múltiple (MLR)
- 5 Análisis de Correlaciones Canónicas (CCA)
- 6 Resumen

Mínimos Cuadrados Parciales (PLS)

- Supongamos que disponemos de los siguientes datos^a

$$\mathbf{x} \in \mathbb{R}^{n \times 1} \quad \mathbf{y} \in \mathbb{R}^{m \times 1}$$

- PLS busca las p **direcciones de máxima covarianza** $\mathbf{U} \in \mathbb{R}^{n \times p}$,
 $\mathbf{V} \in \mathbb{R}^{m \times p}$

$$\arg \max_{\mathbf{U}, \mathbf{V}} \text{Tr}(\mathbf{U}^T \mathbf{R}_{\mathbf{x}, \mathbf{y}} \mathbf{V}) \quad \text{sujeto a} \quad \mathbf{V}^T \mathbf{V} = \mathbf{U}^T \mathbf{U} = \mathbf{I}$$

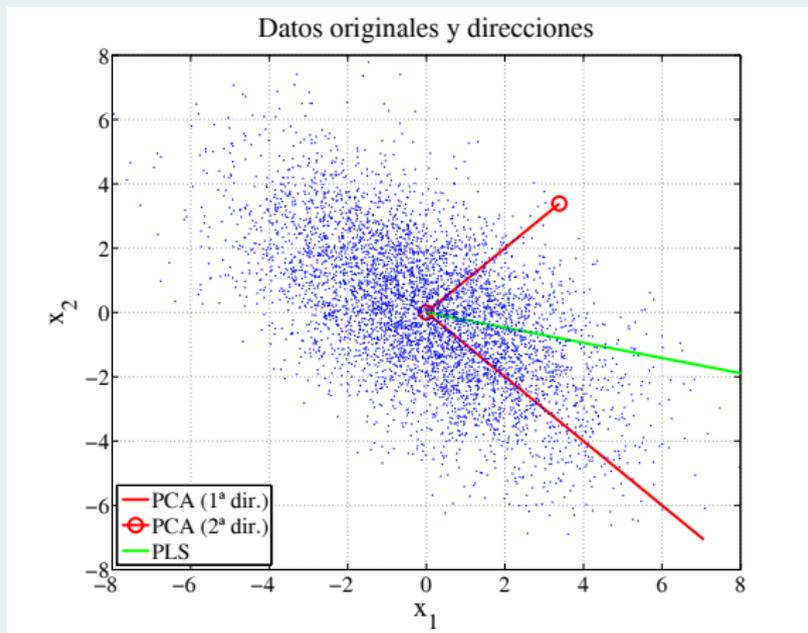
- Solución:** Vectores singulares de $\mathbf{R}_{\mathbf{x}, \mathbf{y}}$

$$\mathbf{R}_{\mathbf{x}, \mathbf{y}} \simeq \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T \quad \mathbf{\Lambda} = \text{diag}([\lambda_1, \dots, \lambda_p])$$

^aNormalmente \mathbf{x} se corresponde con las observaciones, mientras que \mathbf{y} contiene información que nos gustaría extraer

Ejemplo con datos bidimensionales

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{\mathbf{y},\mathbf{y}} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{\mathbf{x},\mathbf{y}} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



Contenido

- 1 Introducción
- 2 Análisis de Componentes Principales (PCA)
- 3 Mínimos Cuadrados Parciales (PLS)
- 4 Regresión Lineal Múltiple (MLR)**
- 5 Análisis de Correlaciones Canónicas (CCA)
- 6 Resumen

Presentación del Problema

- Supongamos que disponemos de los datos $\mathbf{x} \in \mathbb{R}^{n \times 1}$, $\mathbf{y} \in \mathbb{R}^{m \times 1}$
- Nuestro objetivo consiste en **estimar y a partir de una proyección p -dimensional de \mathbf{x}**
 - Una primera opción podría consistir en hacer una reducción de rango de \mathbf{x} mediante PCA, y posteriormente estimar \mathbf{y} a partir de dicha proyección $\tilde{\mathbf{x}} \in \mathbb{R}^{p \times 1}$
 - MLR resuelve este problema de manera óptima: Proporciona el proyector y la matriz de reconstrucción que minimizan el error cuadrático medio (MSE)
- **Definiciones:**
 - Proyector: $\mathbf{U} \in \mathbb{R}^{n \times p}$
 - Proyección: $\tilde{\mathbf{x}} = \mathbf{U}^T \mathbf{x} \in \mathbb{R}^{p \times 1}$
 - Matriz de reconstrucción: $\mathbf{V} \in \mathbb{R}^{m \times p}$
 - Reconstrucción: $\hat{\mathbf{y}} = \mathbf{V} \tilde{\mathbf{x}}$
- **Ambigüedad:** Dado que $\hat{\mathbf{y}} = \mathbf{V} \mathbf{U}^T \mathbf{x} = \mathbf{V} \mathbf{A} \mathbf{A}^{-1} \mathbf{U}^T \mathbf{x}$ podemos asumir

$$\mathbf{V}^T \mathbf{V} = \mathbf{I}$$

Formulación del Problema

- Error de reconstrucción

$$J = E \left[\| \mathbf{y} - \hat{\mathbf{y}} \|^2 \right] = \text{Tr} \left(E \left[(\mathbf{y} - \hat{\mathbf{y}})(\mathbf{y} - \hat{\mathbf{y}})^T \right] \right)$$

con

$$\hat{\mathbf{y}} = \mathbf{V} \tilde{\mathbf{x}} = \mathbf{V} \mathbf{U}^T \mathbf{x} \quad \mathbf{V} \in \mathbb{R}^{m \times p}, \quad \mathbf{U} \in \mathbb{R}^{n \times p}$$

- J se puede reescribir como

$$J = \text{Tr}(\mathbf{R}_{\mathbf{y},\mathbf{y}}) + \text{Tr}(\mathbf{R}_{\hat{\mathbf{y}},\hat{\mathbf{y}}}) - 2\text{Tr}(\mathbf{R}_{\hat{\mathbf{y}},\mathbf{y}})$$

con las matrices de correlación:

$$\mathbf{R}_{\mathbf{y},\mathbf{y}} = E \left[\mathbf{y} \mathbf{y}^T \right] \quad \mathbf{R}_{\hat{\mathbf{y}},\hat{\mathbf{y}}} = \mathbf{V} \mathbf{U}^T \mathbf{R}_{\mathbf{x},\mathbf{x}} \mathbf{U} \mathbf{V}^T \quad \mathbf{R}_{\hat{\mathbf{y}},\mathbf{y}} = \mathbf{V} \mathbf{U}^T \mathbf{R}_{\mathbf{x},\mathbf{y}}$$

- Finalmente, teniendo $\mathbf{V}^T \mathbf{V} = \mathbf{I}$ en cuenta:

$$J = \text{Tr}(\mathbf{R}_{\mathbf{y},\mathbf{y}}) + \text{Tr}(\mathbf{U}^T \mathbf{R}_{\mathbf{x},\mathbf{x}} \mathbf{U}) - 2\text{Tr}(\mathbf{U}^T \mathbf{R}_{\mathbf{x},\mathbf{y}} \mathbf{V})$$

Solución

- Nuestro problema es:

$$\arg \min_{\mathbf{U}, \mathbf{V}} \text{Tr}(\mathbf{U}^T \mathbf{R}_{\mathbf{x}, \mathbf{x}} \mathbf{U}) - 2\text{Tr}(\mathbf{U}^T \mathbf{R}_{\mathbf{x}, \mathbf{y}} \mathbf{V}) \quad \text{sujeto a} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}$$

- Solución en \mathbf{U} : Tomando el gradiente $\nabla_{\mathbf{U}}$ e igualando a cero

$$2\mathbf{R}_{\mathbf{x}, \mathbf{x}} \mathbf{U} - 2\mathbf{R}_{\mathbf{x}, \mathbf{y}} \mathbf{V} = 0 \quad \Rightarrow_{\mathbf{R}_{\mathbf{x}, \mathbf{x}} \succ \mathbf{0}} \quad \mathbf{U} = \mathbf{R}_{\mathbf{x}, \mathbf{x}}^{-1} \mathbf{R}_{\mathbf{x}, \mathbf{y}} \mathbf{V}$$

- Solución en \mathbf{V} : Con $\mathbf{U} = \mathbf{R}_{\mathbf{x}, \mathbf{x}}^{-1} \mathbf{R}_{\mathbf{x}, \mathbf{y}} \mathbf{V}$, el problema se reduce a

$$\arg \max_{\mathbf{V}} \text{Tr}(\mathbf{V}^T \mathbf{R}_{\mathbf{y}, \mathbf{x}} \mathbf{R}_{\mathbf{x}, \mathbf{x}}^{-1} \mathbf{R}_{\mathbf{x}, \mathbf{y}} \mathbf{V}) \quad \text{sujeto a} \quad \mathbf{V}^T \mathbf{V} = \mathbf{I}$$

cuya solución consiste en los p **autovectores principales de**

$$\mathbf{R}_{\mathbf{y}, \mathbf{x}} \mathbf{R}_{\mathbf{x}, \mathbf{x}}^{-1} \mathbf{R}_{\mathbf{x}, \mathbf{y}}$$

Solución (Continuación)

● Soluciones MLR:

- **V**: Autovectores principales de $\mathbf{R}_{y,x} \mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y}$

$$\mathbf{R}_{y,x} \mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y} \mathbf{V} = \mathbf{V} \Lambda \quad \Lambda = \text{diag}([\lambda_1, \dots, \lambda_p])$$

- **U**: Autovectores principales de $\mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y} \mathbf{R}_{y,x}$

$$\mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y} \mathbf{R}_{y,x} \mathbf{U} = \mathbf{U} \Lambda \quad \Lambda = \text{diag}([\lambda_1, \dots, \lambda_p])$$

- **U** y **V** se pueden obtener a partir de la SVD de $\mathbf{R}_{x,x}^{-\frac{1}{2}} \mathbf{R}_{x,y}$

● Formulación alternativa (GEV):

$$\begin{bmatrix} \mathbf{0} & \mathbf{R}_{x,y} \\ \mathbf{R}_{y,x} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} \Lambda^{-\frac{1}{2}} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{x,x} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{U} \Lambda^{-\frac{1}{2}} \\ \mathbf{V} \end{bmatrix} \Lambda^{\frac{1}{2}}$$

- **Error de reconstrucción**: $J = \text{Tr}(\mathbf{R}_{y,y}) - \underbrace{\text{Tr}(\mathbf{V}^T \mathbf{R}_{y,x} \mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y} \mathbf{V})}_{\text{Tr}(\Lambda) = \sum_{k=1}^p \lambda_k}$

Interpretación

- Para $y = x$, MLR se reduce a PCA
 - MLR se puede ver como una **generalización de PCA**
- MLR equivale a PLS tras preblanquear x :
 - Se está maximizando la **covarianza** entre las proyecciones de y ($V^T y$) y las proyecciones de la **versión preblanqueada** de x

- **Ortogonalidad:**

- Ortogonalidad (incorrelación) en las proyecciones \tilde{x} (no en U)

$$U^T R_{x,x} U = \Lambda \quad U^T U \neq \text{diagonal}$$

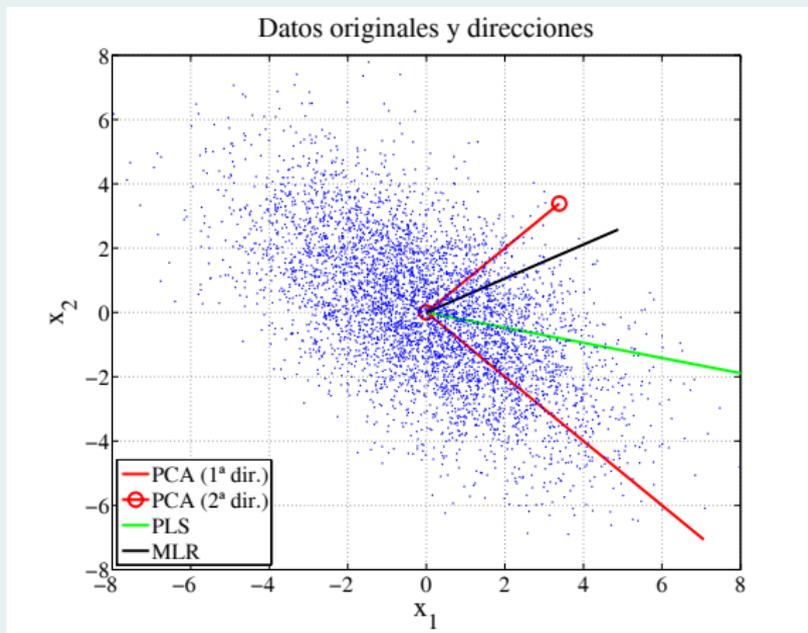
- Ortogonalidad en los proyectores V , pero no en las proyecciones

$$V^T V = I \quad V^T R_{y,y} V \neq \text{diagonal}$$

- MLR es también conocido como:
 - **Rank-reduced Wiener Filter**
 - Orthonormalized PLS (OPLS)
 - Half Canonical Correlation Analysis (Half-CCA)

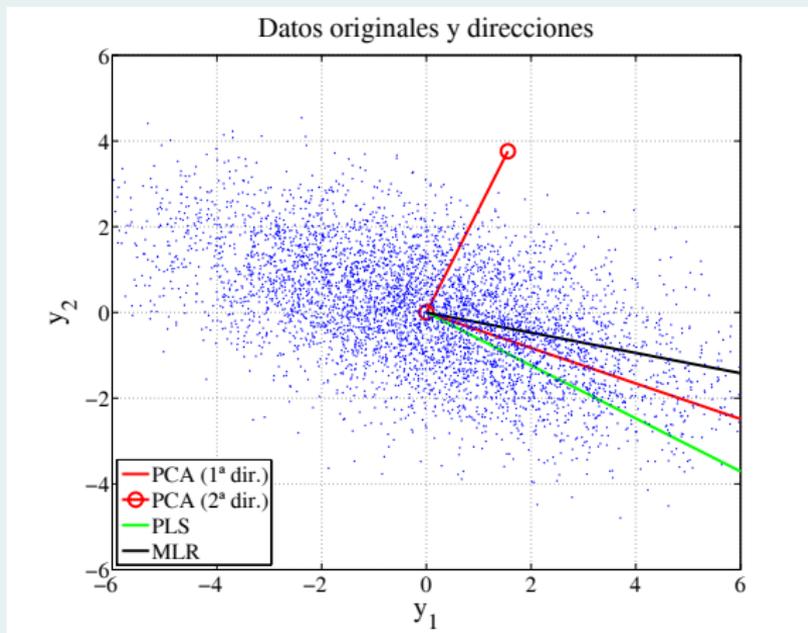
Ejemplo con datos bidimensionales

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{\mathbf{y},\mathbf{y}} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{\mathbf{x},\mathbf{y}} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



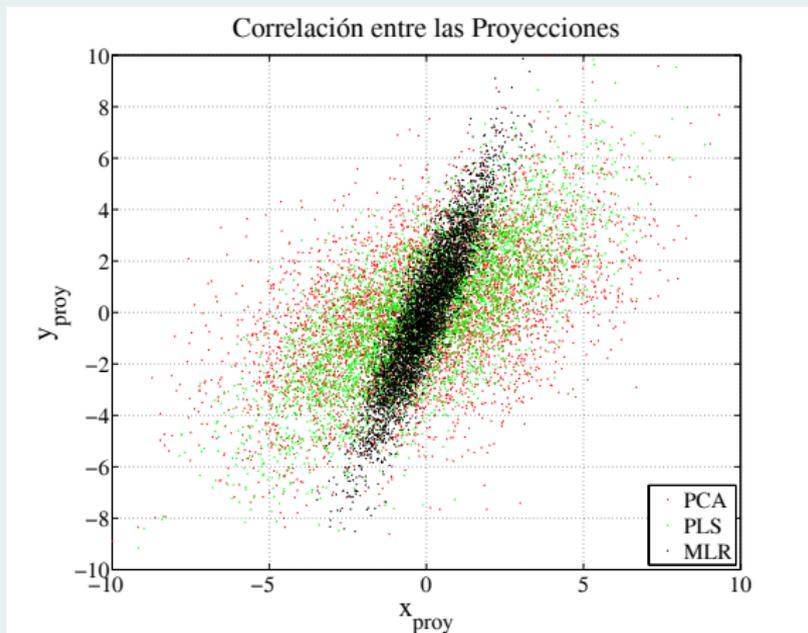
Ejemplo con datos bidimensionales

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{\mathbf{y},\mathbf{y}} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{\mathbf{x},\mathbf{y}} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



Ejemplo con datos bidimensionales

$$\mathbf{R}_{x,x} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{y,y} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{x,y} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



Contenido

- 1 Introducción
- 2 Análisis de Componentes Principales (PCA)
- 3 Mínimos Cuadrados Parciales (PLS)
- 4 Regresión Lineal Múltiple (MLR)
- 5 Análisis de Correlaciones Canónicas (CCA)**
- 6 Resumen

Solución

- **Soluciones a partir del GEV:**

$$\begin{bmatrix} \mathbf{0} & \mathbf{R}_{x,y} \\ \mathbf{R}_{y,x} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{x,x} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{y,y} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \Lambda$$

- Los autovalores en Λ son los coeficientes de correlación ρ_k
- Las soluciones \mathbf{z} se obtienen como $\mathbf{z} = \frac{\mathbf{U}^T \mathbf{x} + \mathbf{V}^T \mathbf{y}}{2}$
- **Formulaciones alternativas:**

- \mathbf{V} se puede obtener como

$$\mathbf{R}_{y,y}^{-1} \mathbf{R}_{y,x} \mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y} \mathbf{V} = \mathbf{V} \Lambda^2$$

- \mathbf{U} se puede obtener como

$$\mathbf{R}_{x,x}^{-1} \mathbf{R}_{x,y} \mathbf{R}_{y,y}^{-1} \mathbf{R}_{y,x} \mathbf{U} = \mathbf{U} \Lambda^2$$

- \mathbf{U} , \mathbf{V} también a partir de la SVD de $\mathbf{R}_{x,x}^{-\frac{1}{2}} \mathbf{R}_{x,y} \mathbf{R}_{y,y}^{-\frac{1}{2}}$

Propiedades

- **Ortogonalidad**: A diferencia de PLS, la ortogonalidad se da entre las proyecciones, y no entre los proyectores

$$\mathbf{U}^T \mathbf{R}_{\mathbf{x},\mathbf{x}} \mathbf{U} = \mathbf{V}^T \mathbf{R}_{\mathbf{y},\mathbf{y}} \mathbf{V} = \mathbf{I} \quad \mathbf{U}^T \mathbf{U} \neq \text{diagonal} \quad \mathbf{V}^T \mathbf{V} \neq \text{diagonal}$$

- Para datos **Gaussianos** CCA proporciona las proyecciones de **máxima información mutua**^a
 - Conexión con Teoría de la Información
 - Medida de la dependencia estadística
- Relación con otras técnicas
 - CCA es equivalente a PLS tras preblanquear \mathbf{x} e \mathbf{y}
 - CCA es equivalente a MLR tras preblanquear \mathbf{y}

^aScharf, L. L. and Mullis, C. T. Canonical coordinates and the geometry of inference, rate and capacity IEEE Transactions on Signal Processing, 2000, 48, 824-831

Aplicación 1: Estima ML con Ruido Gaussiano Coloreado

- Consideremos el modelo de señal

$$\mathbf{x} = \mathbf{A}_x \mathbf{s} + \mathbf{n}_x \quad \mathbf{y} = \mathbf{A}_y \mathbf{s} + \mathbf{n}_y$$

con

$$\mathbf{A}_x \in \mathbb{R}^{n \times p} \quad \mathbf{A}_y \in \mathbb{R}^{m \times p}$$

$$\mathbf{s} \in \mathbb{R}^{p \times 1} \sim \mathcal{N}(0, \mathbf{I}) \quad \mathbf{n}_x \in \mathbb{R}^{n \times 1} \sim \mathcal{N}(0, \mathbf{N}_x) \quad \mathbf{n}_y \in \mathbb{R}^{m \times 1} \sim \mathcal{N}(0, \mathbf{N}_y)$$

y con **ruidos independientes**

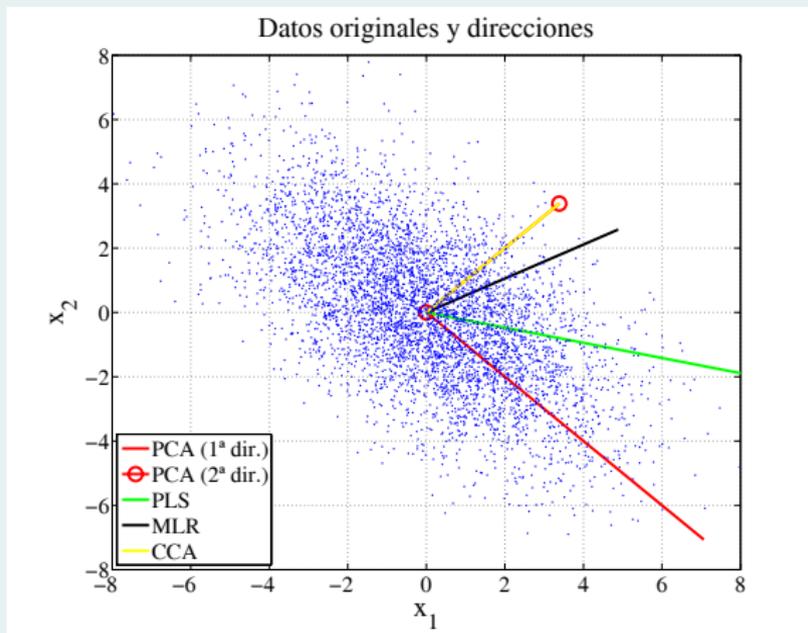
$$E[\mathbf{n}_x \mathbf{n}_y^T] = \mathbf{0}$$

- Queremos estimar \mathbf{A}_x , \mathbf{A}_y , \mathbf{N}_x y \mathbf{N}_y
- El estimador de **máxima verosimilitud** se obtiene a partir del análisis de correlaciones canónicas^a (de rango p) de \mathbf{x} e \mathbf{y}

^aWong, K. M.; Wu, Q. and Stoica, P. Generalized correlation decomposition applied to array processing in unknown noise environments, Prentice-Hall, Inc., 1995, 219-323

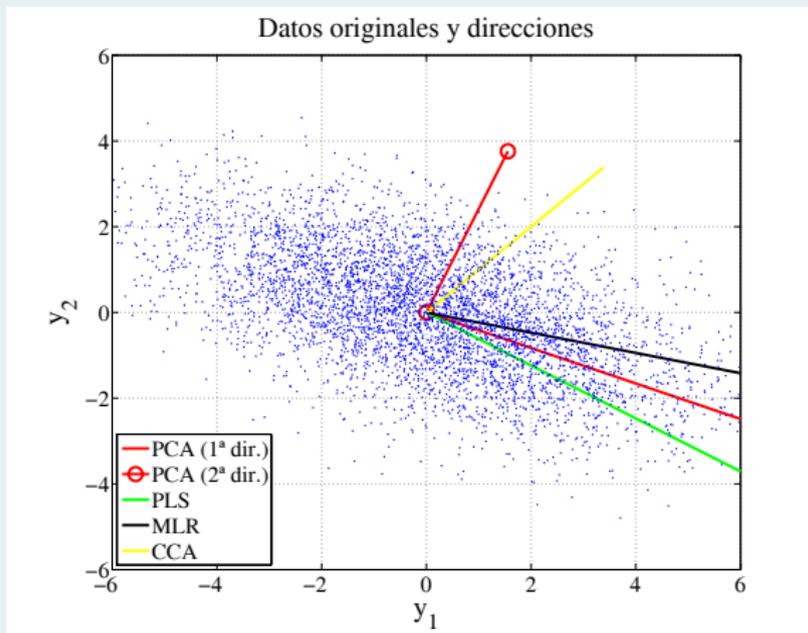
Ejemplo con datos bidimensionales

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{\mathbf{y},\mathbf{y}} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{\mathbf{x},\mathbf{y}} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



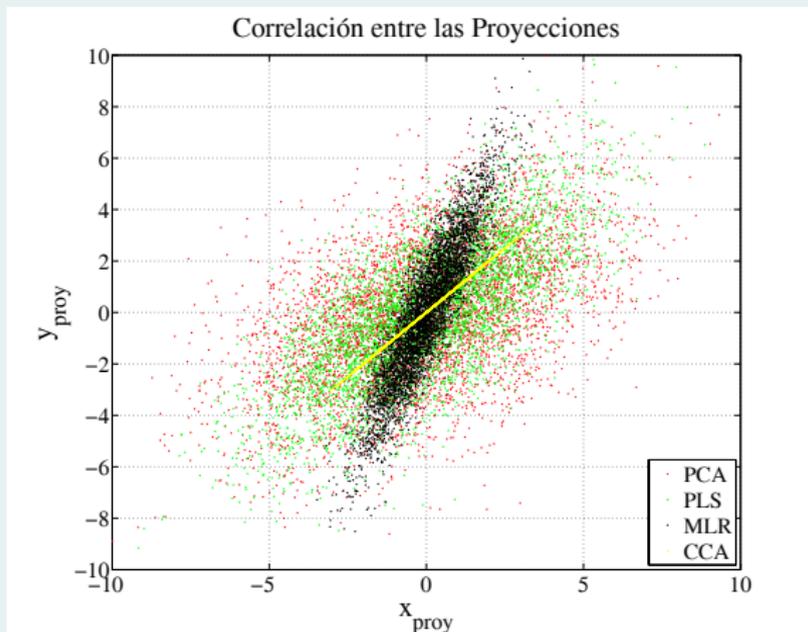
Ejemplo con datos bidimensionales

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{\mathbf{y},\mathbf{y}} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{\mathbf{x},\mathbf{y}} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



Ejemplo con datos bidimensionales

$$\mathbf{R}_{x,x} = \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} \quad \mathbf{R}_{y,y} = \begin{bmatrix} 6 & -2 \\ -2 & 2 \end{bmatrix} \quad \mathbf{R}_{x,y} = \begin{bmatrix} 4 & -2 \\ 0 & 2 \end{bmatrix}$$



Algoritmo adaptativo

- ¿Como obtener las soluciones CCA de manera adaptativa?
- **Idea:** Problemas de regresión LS acoplados^a

- Función de coste:

$$J = \underbrace{\|\mathbf{v}^T \mathbf{Y} - \mathbf{z}\|^2}_{\simeq E[\|\mathbf{v}^T \mathbf{y} - z\|^2]} + \underbrace{\|\mathbf{u}^T \mathbf{X} - \mathbf{z}\|^2}_{\simeq E[\|\mathbf{u}^T \mathbf{x} - z\|^2]}$$

- Obtención de \mathbf{v} y \mathbf{u} : Regresión LS

$$\hat{\mathbf{v}} = \underbrace{(\mathbf{Y}\mathbf{Y}^T)^{-1}}_{\simeq \mathbf{R}_{\mathbf{y},\mathbf{y}}^{-1}} \underbrace{\mathbf{Y}\mathbf{z}^T}_{\simeq \mathbf{R}_{\mathbf{y},z}} \quad \hat{\mathbf{u}} = \underbrace{(\mathbf{X}\mathbf{X}^T)^{-1}}_{\simeq \mathbf{R}_{\mathbf{x},\mathbf{x}}^{-1}} \underbrace{\mathbf{X}\mathbf{z}^T}_{\simeq \mathbf{R}_{\mathbf{x},z}}$$

- Obtención de \mathbf{z} : Promediado $\mathbf{z} = \frac{\hat{\mathbf{v}}^T \mathbf{Y} + \hat{\mathbf{u}}^T \mathbf{X}}{2}$
- Solución adaptativa mediante el **RLS**^b

^aVia, J.; Santamaria, I. and Pérez, J. A Learning Algorithm for Adaptive Canonical Correlation Analysis of Several Data Sets, Neural Networks, 2007, 20, 139-152

^bSayed, A. H. Fundamentals of Adaptive Filtering, John Wiley and Sons, 2003

Extensiones a varios conjuntos de datos

- Varias extensiones de CCA a varios conjuntos de datos^a

- Datos

$$\mathbf{x}_i \in \mathbb{R}^{n_i \times 1} \quad i = 1, \dots, M$$

- Se buscan las proyecciones

$$\tilde{x}_i = \mathbf{u}_i^T \mathbf{x} \quad \text{con} \quad E[\tilde{x}_i^2] = 1 \quad i = 1, \dots, M$$

- Definimos el vector $\tilde{\mathbf{x}} = [\tilde{x}_1, \dots, \tilde{x}_M]$ y la matriz

$$\mathbf{C} = E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] \in \mathbb{R}^{M \times M}$$

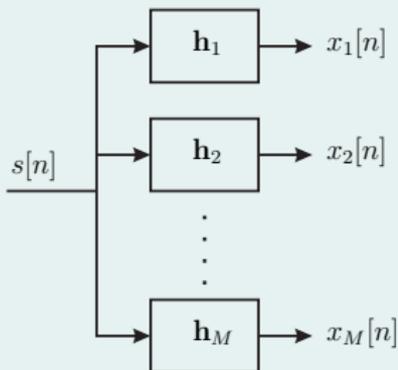
- Generalizaciones de CCA basadas en:

- Maximizar el Máximo Autovalor de \mathbf{C} (MAXVAR)
- Minimizar el Mínimo Autovalor de \mathbf{C} (MINVAR)
- Minimizar el Determinante de \mathbf{C} (MINDET)
- Maximizar la suma de los elementos en \mathbf{C} (SUMCORR)
- Maximizar $\|\mathbf{C}\|^2$ (SSQCORR)

^aKettenring, J. R. Canonical analysis of several sets of variables, *Biometrika*, 1971, 58, 433-451

Ejemplo

- Igualación ciega de canales MIMO-FIR:^a
- Ejemplo: Canal SIMO-FIR
 - Longitud conocida: L
 - Fuente desconocida $s[n]$
 - Canales desconocidos $\mathbf{h}_k[n] \in \mathbb{R}^{L \times 1}$



^aVia, J.; Santamaria, I. and Pérez, J. Deterministic CCA-Based Algorithms for Blind Equalization of FIR-MIMO Channels IEEE Transactions on Signal Processing, 2007, 55, 3867-3878

Ejemplo (Continuación)

- Modelo de señal:

$$\mathbf{y}[n] = \mathcal{T}(\mathbf{H})\mathbf{s}[n]$$

- Observaciones:

$$\mathbf{y}[n] = [\mathbf{x}^T[n] \dots \mathbf{x}^T[n - K + 1]]^T \quad \mathbf{x}[n] = [x_1[n], \dots, x_M[n]]^T$$

- Fuente:

$$\mathbf{s}[n] = [s[n], s[n - 1], \dots, s[n - K - L + 2]]$$

- Matriz de filtrado:

$$\mathcal{T}(\mathbf{H}) = \begin{bmatrix} \mathbf{h}[0] & \cdots & \mathbf{h}[L - 1] & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}[0] & \cdots & \mathbf{h}[L - 1] \end{bmatrix}$$

Ejemplo (Continuación)

- **Igualación:**

- Bajo condiciones sencillas, $\mathcal{T}(\mathbf{H})$ es completa en rango
- Por lo tanto, existe una matriz \mathbf{W} tal que

$$\mathbf{W}^H \mathbf{y} = \mathbf{s}$$

- Cada columna de \mathbf{W} es un igualador con un retardo diferente

$$\mathbf{w}_k^H \mathbf{y}[n] = s[n - k + 1]$$

- **¿Cómo encontrar los igualadores?**

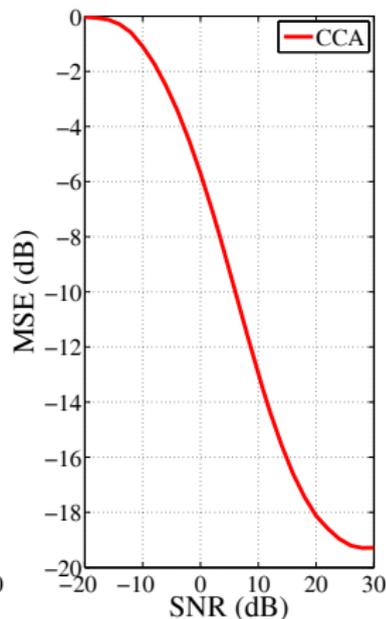
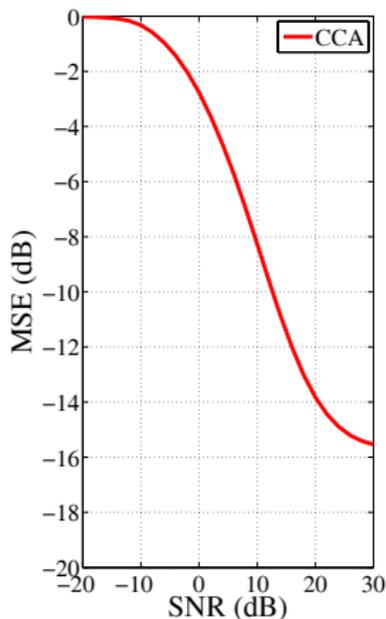
- **Idea:** CCA

$$\arg \min_{\mathbf{w}} \sum_{k,l=1}^M E \left[\left\| \mathbf{w}_k^H \mathbf{y}[n+k] - \mathbf{w}_l^H \mathbf{y}[n+l] \right\|^2 \right]$$

$$\text{sujeto a } \mathbf{w}_k^H \mathbf{R}_{\mathbf{y},\mathbf{y}} \mathbf{w}_k = 1$$

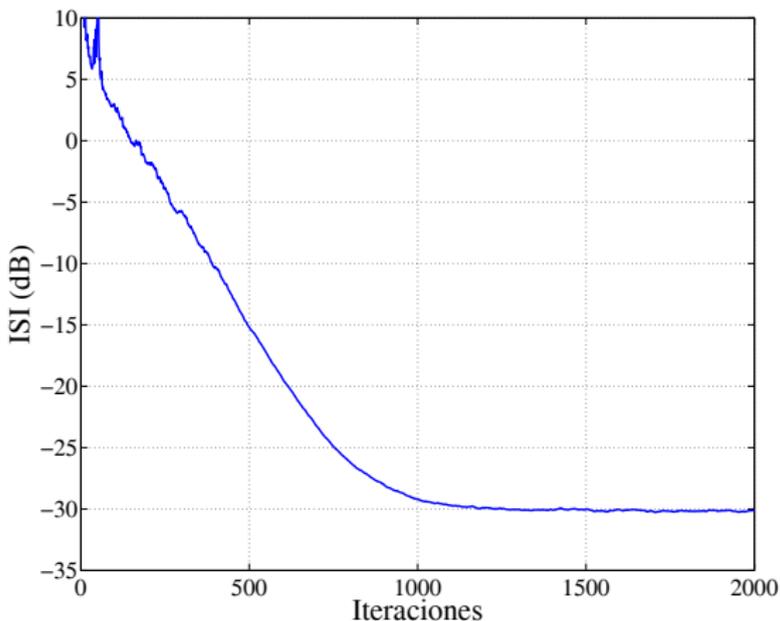
Ejemplo (Continuación)

- Canales SIMO 1×2 y 1×4 con $L = 2$



Ejemplo (Continuación)

- Canal SIMO 1×3 , $L = 7$, SNR=30dB: Algoritmo adaptativo (CCA-RLS)



Contenido

- 1 Introducción
- 2 Análisis de Componentes Principales (PCA)
- 3 Mínimos Cuadrados Parciales (PLS)
- 4 Regresión Lineal Múltiple (MLR)
- 5 Análisis de Correlaciones Canónicas (CCA)
- 6 **Resumen**

Análisis Estadístico Multivariado Basado en SOS

- **PCA:**
 - Direcciones de máxima varianza
 - Reducción de rango de mínimo MSE
- **PLS:**
 - Direcciones de máxima covarianza
- **MLR:**
 - Estimador de rango reducido de mínimo MSE (Wiener Filter)
 - Maximiza la covarianza entre y y la versión preblanqueada de x
 - Invariante a transformaciones lineales de x
- **CCA:**
 - Direcciones con mayor coeficiente de correlación
 - Maxima covarianza entre las versiones preblanqueadas de x e y
 - Invariante a transformaciones lineales de x e y

Extensiones

- Extensión sencilla a datos complejos
- Formulación determinista

$$\mathbf{X} = [\mathbf{x}[n] \cdots \mathbf{x}[N-1]] \quad \mathbf{Y} = [\mathbf{y}[n] \cdots \mathbf{y}[N-1]]$$

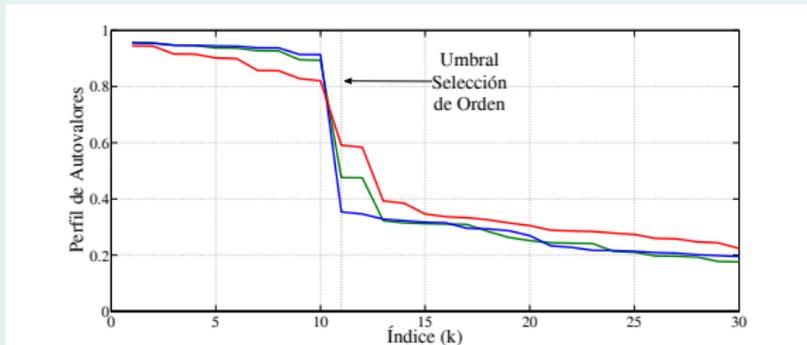
- Estima de las matrices de correlación

$$\hat{\mathbf{R}}_{\mathbf{x},\mathbf{x}} = \frac{1}{N} \mathbf{X}\mathbf{X}^H \quad \hat{\mathbf{R}}_{\mathbf{y},\mathbf{y}} = \frac{1}{N} \mathbf{Y}\mathbf{Y}^H \quad \hat{\mathbf{R}}_{\mathbf{x},\mathbf{y}} = \frac{1}{N} \mathbf{X}\mathbf{Y}^H$$

- Extensiones no lineales:
 - Redes Neuronales
 - **Métodos Kernel**

Un Problema no Tratado

- ¿Cómo seleccionar el orden p de la reducción de rango?
 - Criterio de Akaike^a
 - Criterio MDL (Minimum Description Length)^b
- Ejemplo



^aAkaike, H. A new look at the statistical model identification, IEEE Transactions on Automatic Control, 1974, 19, 716-723

^bRissanen, J. Universal coding, information, prediction, and estimation. IEEE Transactions on Information Theory, 1984, 30, 629-636