Kernel PCA

- Obtain nonlinear features from data
- Can form kernel PCA in primal space (R) or dual space (K).
- Problem closely related to LS SVM
- Must ensure feature data has zero mean
- Applications: Preprocessing data, denoising, compression, image interpretation

KPCA Formulation

- Kernel PCA uses kernels to max $E(0-w^T(\phi(x) - m_\phi))^2$.
- Use input data to approximate ensemble average to get the following quantities, $\Phi(x) = (\phi(x(1), \ldots, \phi(x(m))^T$, R is sample covariance matrix, and kernel matrix is

$$K = (\Phi(x) - 1/m\ 11^T \Phi(x))\,(\Phi(x) - 1/m\ 11^T \Phi(x))^T$$

- We can formulate as a QP problem where we

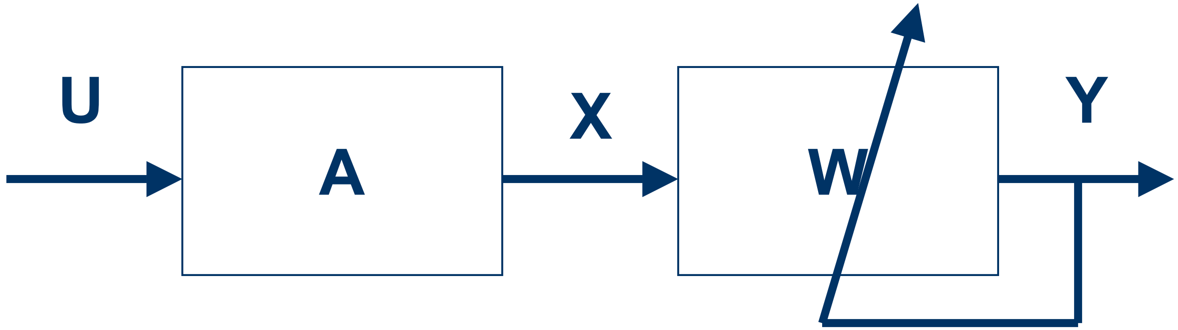
$$\max \tfrac{1}{2} w^T R w$$

$$\text{subject to } w^T w = 1 \text{ and } w = (\Phi(x) - 1/m\ 11^T \Phi(x))^T \alpha$$

- Can solve in primal or dual spaces. In dual space we have another eigenvector /eigenvalue problem $K\alpha = \lambda\alpha$.

Independent Component Analysis

PCA decorrelates inputs. However in many instances we may want to make outputs independent.



InpAuts U assumed independent and user sees X.
Goal is to find W so that Y is independent.

Applications of ICA

- **Speech Separation: several speech signals are mixed together (cocktail problem)**
- **Array antenna processing: several narrowband signals mixed together from unknown directions**
- **Hyperspectral Images:images at multiple wavelengths**
- **Biomedical information: Brain signals, EEG data, FMRI data**
- **Financial market data analysis: extract dominant signals**

ICA Solution

- Y = DPU where D is a diagonal matrix and P is a permutation matrix.
- Algorithm is unsupervised. What are assumptions where learning is possible? All components of U except possibly one are nongaussian.
- Establish criterion to learn from (use higher order statistics): information based criteria, kurtosis function.
- Kullback Leibler Divergence:

$$D(f,g) = \int f(x) \log (f(x)/g(x)) \, dx$$

ICA Information Criterion

- Kullback Leibler Divergence nonnegative
- Mutual Information I(X:Y) = H(X) – H(X|Y) nonnegative
- Set f to joint density of Y and g to products of marginals of Y then

$$D(f,g) = -H(Y) + \Sigma H(Y_i)$$

which is minimized when components of Y are independent.

- When outputs are independent they will be a permutation and scaled version of U.

ICA Preprocessing

- **Signal processing and filtering**
- **Center data (remove means)**
- **Decorrelate data (apply PCA). If data is jointly Gaussian cannot do any more**

Learning Algorithms

- Can learn weights by approximating divergence cost function established using contrast functions.
- Iterative gradient estimate algorithms can be used.
- Faster convergence can be achieved with fixed point algorithms that approximate Newton's methods.
- Algorithms have been shown to converge.

ICA Example

- **Three signals are linearly mixed**

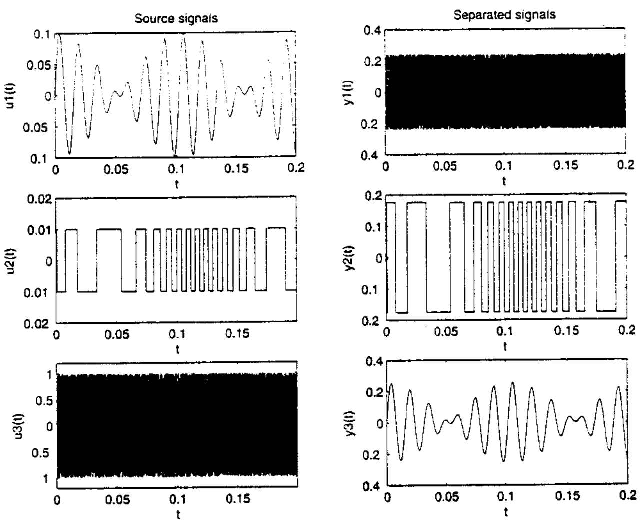


FIGURE 10.13 Waveforms on left-hand side: original source signals. Waveforms on right-hand side: separated source signals.