Projection of Data

- Principal Component Analysis (PCA) (maximize variance)
- Fisher Discriminant Analysis (FDA) (minimize in class scatter while maximizing distance between means)
- Canonical Correlation Analysis (CCA) (maximize covariance)
- Kernelize PCA, FDA, CCA
- Factor analysis, partial regression analysis
- Projection pursuit
- Independent Component Analysis
Adaptive Noise Cancellation

Given a signal $S_1$ mixed with additive impairment $S_2$ we can use an adaptive filter (using algorithms such as LMS) to cancel out the effects of the noise.
Modification to Noise Cancelling

Given $X_1$ and $X_2$ with Adaptive Filter 1 having access to $S_2$ and Adaptive Filter 2 having access to $S_1$, we can recover $S_1$ and $S_2$. 
Independent Component Analysis

- Let $X = AU$ where $A$ is a square mixing matrix, $U$ is a random $m$ vector, and $X$ is the observed random $m$ vector.
- Can we recover $U$ from $X$ if $A$ and $U$ are unknown?
- Under assumptions that $U$ are components of $U$ are independent random variables we can recover $U$ from $X$ under certain assumptions. We need to establish optimization criteria to recover $U$ from $X$.
- Closely related to projection pursuit and factor analysis
- Applications: Blind Source Separation, Blind Deconvolution, Feature Extraction
PCA decorrelates inputs. However in many instances we may want to make outputs independent.

Inputs 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:
  \[ D(f,g) = \int f(x) \log \left( \frac{f(x)}{g(x)} \right) \, dx \]
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

\[
D(f,g) = -H(Y) + \sum H(Y_i)
\]

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.