Overview: Wavelet Feature Extraction

- Orthogonality condition and Coifman Wickehauser's result
- Optimal Basis function decomposition
  - Best Basis algorithm
  - Best Basis for a group of signals
  - Basis Pursuit
  - Matching Pursuit
- Basis function decomposition for discrimination
  - LDB
  - Discriminant Pursuit
- Problems with these approaches

The Best Basis Algorithm
Coifman & Wickerhauser, 1992

- Generate an energy table from a WP by WP^2.
- At each level accumulate the energies of each time/frequency window (TFW).
- Accumulate the energies of the corresponding two TFW at the next level.
- Choose the TFW with higher energy.

BB for a class of signals

- Accumulate energies from a set of normalized signals into one wavelet packet energy table.
- Apply the best basis algorithm to the accumulated energy table.

Bases for Discrimination

- BDB, LDA attempt to extract discriminating features from wavelet packets

- However, wavelet packets and consequently Best Basis are very powerful for creating new efficient low dimensional representation of a high dim signal.

- One may then attempt to discriminate from the new efficient (invariant) BB representation

Discriminant Pursuit
Buckheit & Donoho, 1995

- Apply Fisher Linear Discriminant Analysis (1936) to each basis function:

$$D_i(X < Y) = \frac{E_X[WP_i(x)] - E_Y[WP_i(y)]}{\mathsf{STD}_X[WP_i(x)] + \mathsf{STD}_Y[WP_i(y)]}$$

- **Problem:**

$$E_X[WP_i(x)] = WP_i(E_X[x])$$

e.g., Let u, v be a uniform random phase, then $\sin(wt + u)$ can not be distinguished from $\sin(2wt + v)$.

Local Discriminant Basis (LDB)
Saito & Coifman, 1994

- For a two class problem: accumulate energies for each class of signals into a wavelet packet energy table.

- Find the absolute difference between the two energy tables.

- Apply the best basis algorithm to the new energy table.

Quadratic Discrimination Best Basis (QDB)

We create the set

$$\{Z_{ij} = (X_i - Y_j)\}_{i,j=1}^{n,m},$$

from observations: $\{X_i\}_{i=1}^{n}$ and $\{Y_j\}_{j=1}^{m}$,

Z **Shoud have a good discriminating power.**

Complexity: $W(Z)$ denotes the wavelet packet transform (WP) of Z, then $W^2(Z)$ is the WP of Z^2.
To construct a best basis of Z we need

$$\sum_{i=1}^{n} \sum_{j=1}^{m} W^2(Z_{ij}).$$

Consider a specific wavelet packet basis $wp_r(X_i - Y_j)$; Linearity gives:

$$wp_r(X_i - Y_j) = wp_r(X_i) - wp_r(Y_j).$$

Thus,

$$\begin{aligned} wp_r^2(X_i - Y_j) &= [wp_r(X_i) - wp_r(Y_j)]^2 \\ &= [wp_r^2(X_i) + wp_r^2(Y_j) \\ &\quad -2wp_r(X_i)wp_r(Y_j)]. \end{aligned}$$

Mean contribution of all observations to wp_r is

$$\begin{aligned} &\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} wp_r^2(X_i - Y_j) = \\ &= \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} [wp_r^2(X_i) + wp_r^2(Y_j) \\ &\quad -2wp_r(X_i)wp_r(Y_j)] \\ &= \frac{1}{n}\sum_i wp_r^2(X_i) + \frac{1}{m}\sum_j wp_r^2(Y_j) \\ &\quad -2[\frac{1}{n}\sum_i wp_r(X_i)][\frac{1}{m}\sum_j wp_r(Y_j)]. \end{aligned}$$

Thus, $O(max(n,m))$ calculations are required.

Note: This Best Basis is geared for discrimination in the original high dim space!

Discrimination from two Best Bases

- Find B_1 and best k coefficient locations from training set of class 1

- Find B_2 and best k coefficient locations from training set of class 2

- Project all the data through B_1 and B_2 and create a new $2k$ dimensional representation

- Train and test the classifier on the new $2k$ dim representation