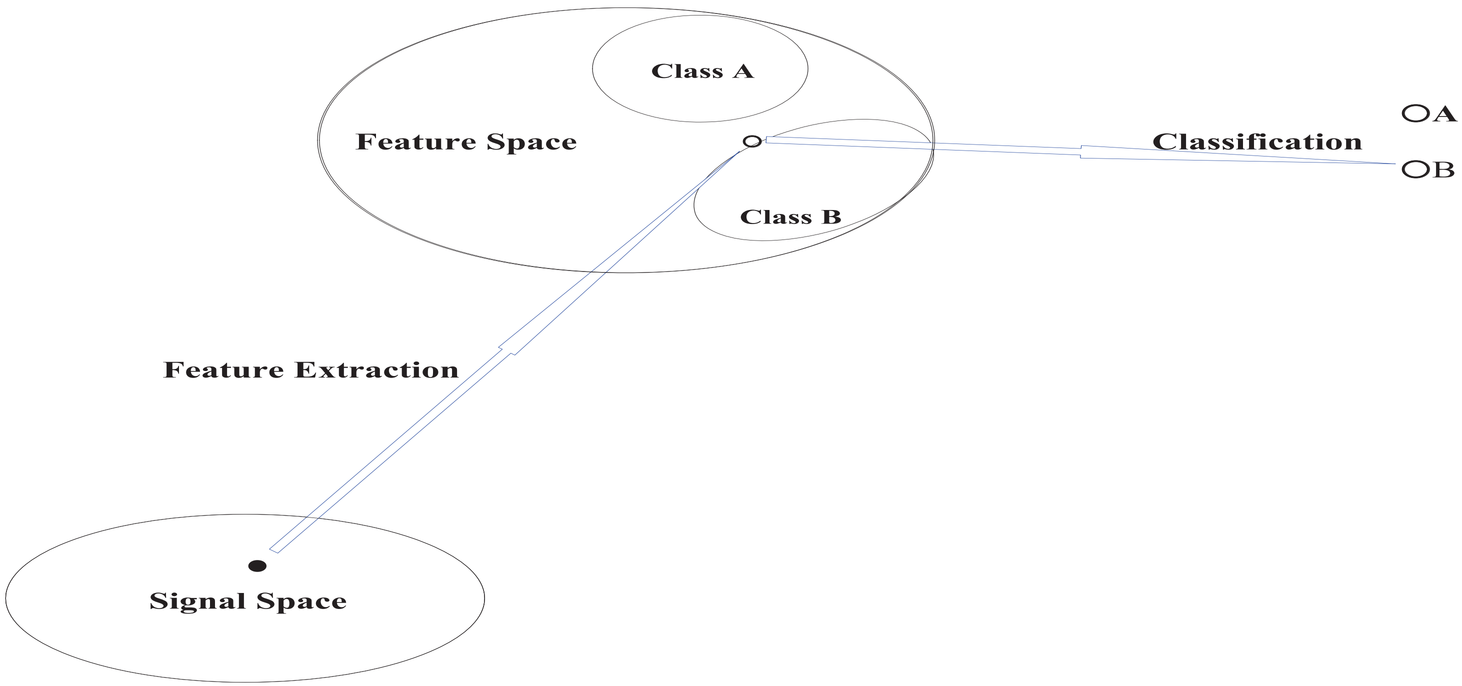
Signal Classification by Power Spectral Density: From Euclid to Riemann

K.M. Wong
McMaster University, Canada

An invited lecture delivered at
The Shun Hing Institute of Advanced Engineering, CUHK
October, 2011

Signal Classification

The concept of signal classification can be depicted as follows:



Process of Signal Classification

Examples of Features in Signal Classification

From daily life to high-tech systems, signal classification is a common problem.

Examples:

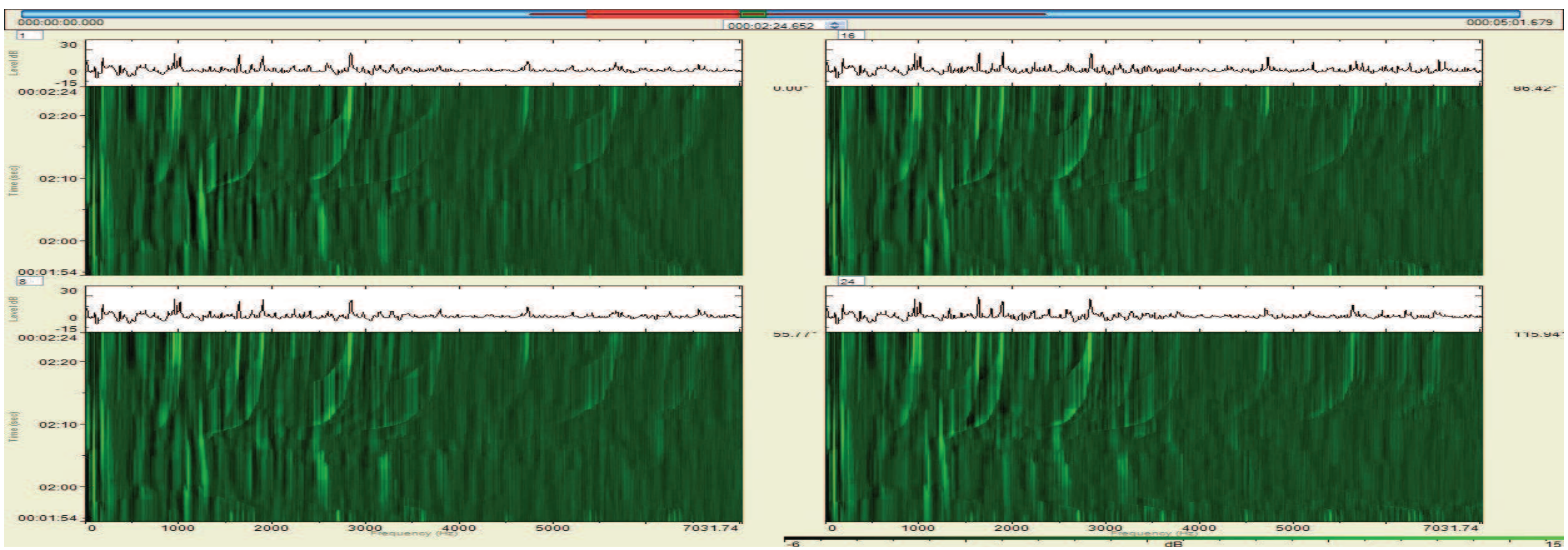
Table 1: Classification Examples

System	Signal	Classification	Feature
Sonar	Sonar array	Vessel class	Signal power spectrum
Medical exam 1	Stethoscope	Heart normal?	Heartbeat regularity
Medical exam 2	EEG	Sleep state	Signal power spectrum
Weather	Visual observation	Rain or dry	Colour of clouds
Family	Voice message	My wife annoyed?	Tone, choice of words

Power Spectral Density (PSD) as a Feature

It is not surprising that PSD is very often used as a feature for signal classification in engineering since many objects distinguish from each other by having different power in different frequency ranges which is what the PSD displays. PSD is also easily measurable and observable.

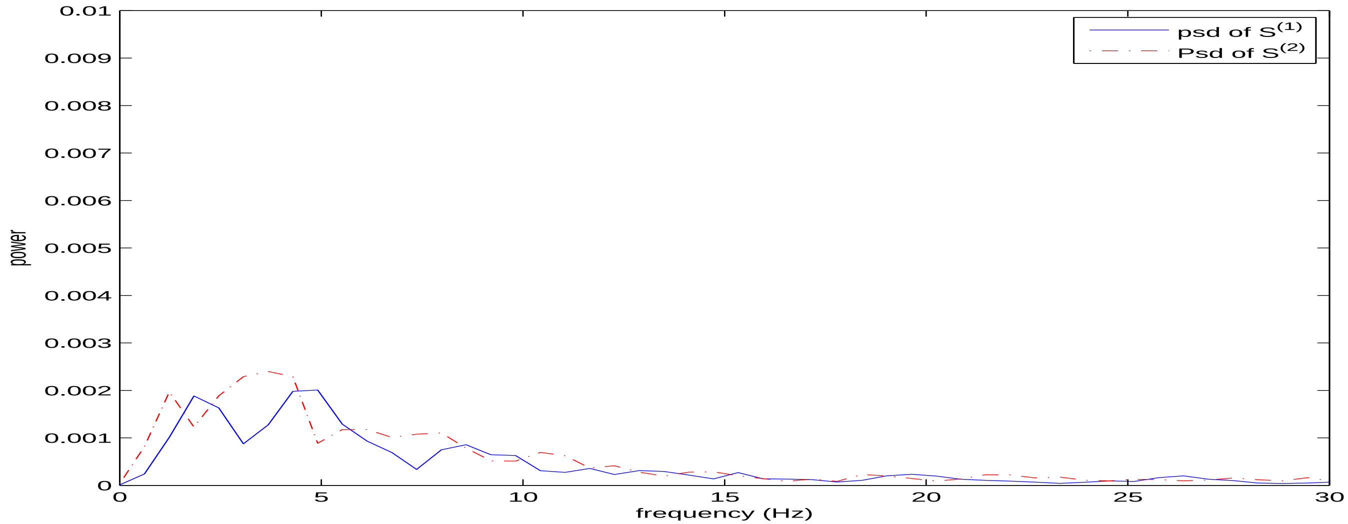
Example: Multi-sensor Sonar Signal Power Spectrum



Signature spectral lines of a vessel from different sonar sensors

PSD as a Feature (contd.)

Example: EEG Signal Power Spectrum



EEG signal PSD from different patients at same state

- Depending on where in the frequency domain the power of the EEG signal is located, the clinical staff decides on the state of the patient's brain.

In both examples of sonar and EEG signal classifications, even though multi-sensor measurements of the signals is employed, strangely, the cross-power spectral density functions between sensors have generally not been utilized!!

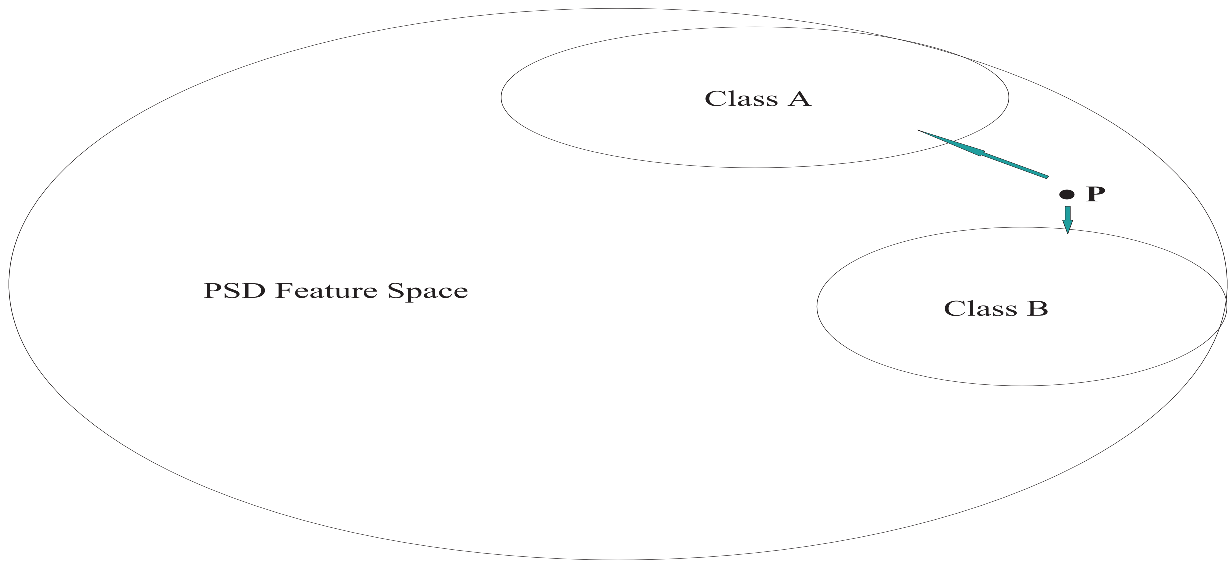
Extraction of PSD from Multi-channel Measurements

We have M channels of recording of the same signal measured by M different sensors. We extract the PSD matrices of different *epochs* of the measured signals as follows:

1. Clean up (noise filtering and artifact removal) the received signal.
2. Divide the cleaned M-channel signal into epochs, each representing a wide-sense stationary signal.
3. For the nth epoch, the M-channel data at different instants t form an $M \times T$ matrix
$$\mathbf{S'}_n = [\mathbf{s'}_n(0), \cdots, \mathbf{s'}_n(T-1)], \qquad n = 1, \cdots, N$$
which is then normalized by its Frobenius norm giving the nth epoch normalized data matrix $\mathbf{S}_n = \mathbf{S'}_n / \|\mathbf{S'}_n\|_F$.
4. Form the vector of the nth epoch $\tilde{\mathbf{s}}_n = \mathrm{vec}\,(\mathbf{S}_n)$ with mean $\tilde{\boldsymbol{\mu}}_n = \mathbb{E}[\tilde{\mathbf{s}}_n]$ so that its covariance matrix is $\tilde{\mathbf{R}}_n = \mathbb{E}[(\tilde{\mathbf{s}}_n - \tilde{\boldsymbol{\mu}}_n)(\tilde{\mathbf{s}}_n - \tilde{\boldsymbol{\mu}}_n)^T]$ which is $MT \times MT$ and contains the $M \times M$ matrices $\mathbf{R}_n(\tau), \tau = 0, \cdots, T-1$.
5. At any frequency ω, the power spectral density (PSD) matrix of the nth epoch signal is then the DFT[§]
$$\mathbf{P}_n(\omega) = \textstyle\sum_\tau e^{-j\omega\tau} \boldsymbol{R}_n(\tau)$$

[§] The Nuthall-Strand algorithm is usually used to obtain $\mathbf{P}_n(\omega)$.

Signal Classification in the Feature Space of PSD

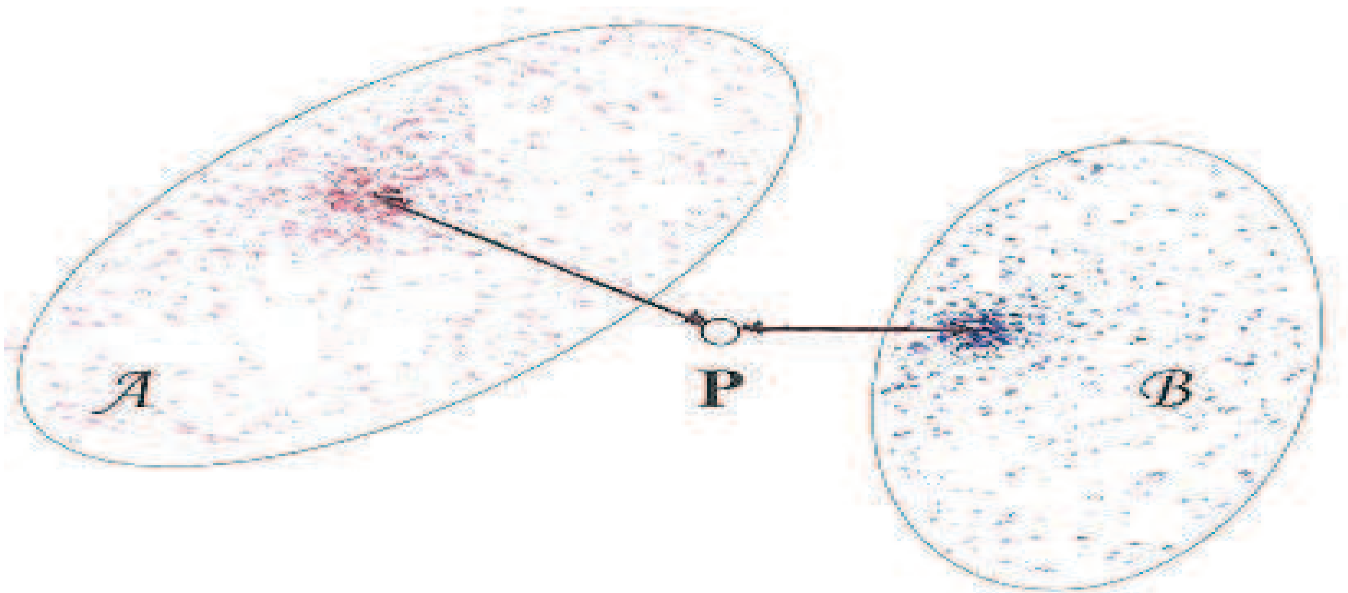


- Signal classification by PSD usually involves the comparison of the observed PSD, $\mathbf{P}$, with a library of collected PSD of different classes.
- The comparison is by examining to which class is the observed PSD "closer".
- The question is: *Closer to which part of the classes*?

Classification Methods

There are various ways of defining the measurement point of a group. Only mention two here:

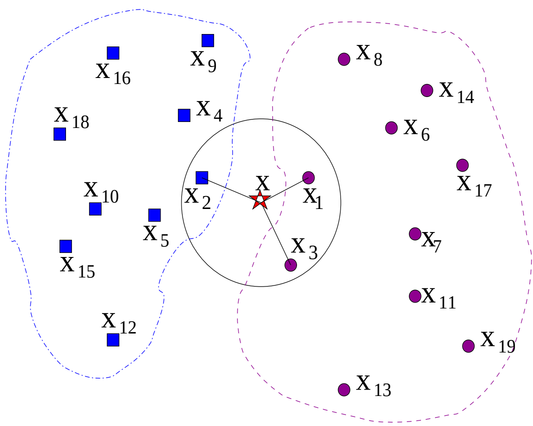
1. Mean of Class



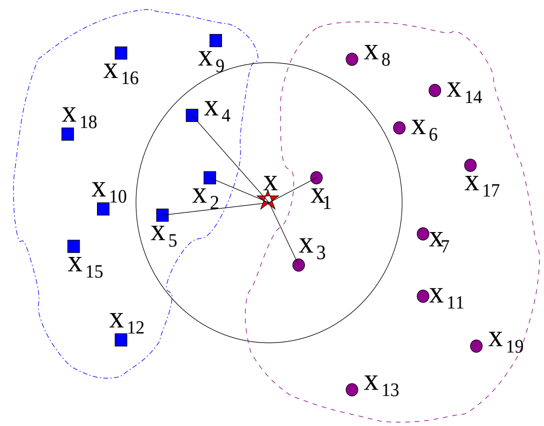
- One of the most intuitive strategies and is commonly used.
- Examine and compare the distances between the observed PSD to the means of the different classes.

Classification Methods (contd.)

2. k-Nearest Neighbour (kNN)



$k = 3$



k=5

- Here, we examine the distance of the observed PSD from the members of the different classes in the library.
- We then choose the k members which are the closest to the observed PSD.
- Out of the k closest members, we use the majority rule to decide to which class does the observed PSD belong.

Distance – An Essential Concept for Classification

Whichever method is used for PSD classification, the concept of distance is essential! The shorter is the distance from each other, the more similar are the features.

- A *distance* is a real number associated with a pair of points x and y. Such a real measure, $d : \{x, y\} \to \mathbb{R}$, is called a *metric* if it posses the following properties:

$$\left.\begin{array}{ll} d(x,y) \geq 0; \quad d(x,y) = 0 \text{ iff } x = y & \text{(positivity)} \\ d(x,y) = d(y,x) & \text{(symmetry)} \\ d(x,z) \leq d(x,y) + d(y,z) & \text{(triangular inequality)} \end{array}\right\} \quad (1)$$

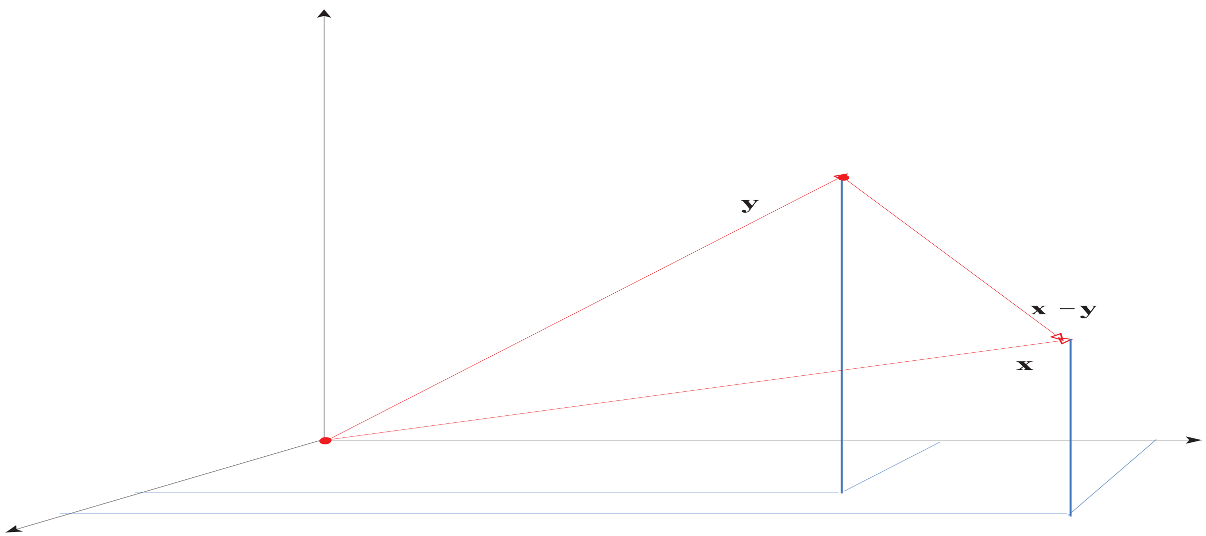
- In a signal space in which the elements are represented as n-dimensional vectors, the distance measure depends on the definition of the *norm* (size) of the vector. The most commonly used is the *inner-product* (Euclidean) norm,

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$$

The Euclidean Distance

The distance induced by the inner-product norm is called the *Euclidean distance*,

$$d_{\mathrm{E}}(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\| = \sqrt{\langle(\mathbf{x}-\mathbf{y}),(\mathbf{x}-\mathbf{y})\rangle} = \left(\sum_{i=1}^{n} |x_i - y_i|^2\right)^{1/2} \tag{2}$$



The Euclidean distance coincides with the usual concept of distance in a 3-dimensional space. It also represents many important physical quantities (MSE, energy difference, etc.)

Euclidean Distance for Matrices

- An $M \times M$ matrix can be looked upon as an element in the M^2 signal space so that the same idea of distance between two such matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ can also be applied:

$$d_{\text{Fo}}(\mathbf{A}, \mathbf{B}) = \left(\sum_{i=1}^{M} \sum_{j=1}^{M} |a_{ij} - b_{ij}|^2 \right)^{1/2} = \sqrt{\text{tr}[(\mathbf{A} - \mathbf{B})(\mathbf{A} - \mathbf{B})^H]} \quad (3)$$

  which is sometimes called the *Frobenius distance*.

- The Frobenius distance is in fact induced by the inner product norm. Thus, the Frobenius distance can be considered as the Euclidean distance between $\mathbf{A}$ and $\mathbf{B}$ since

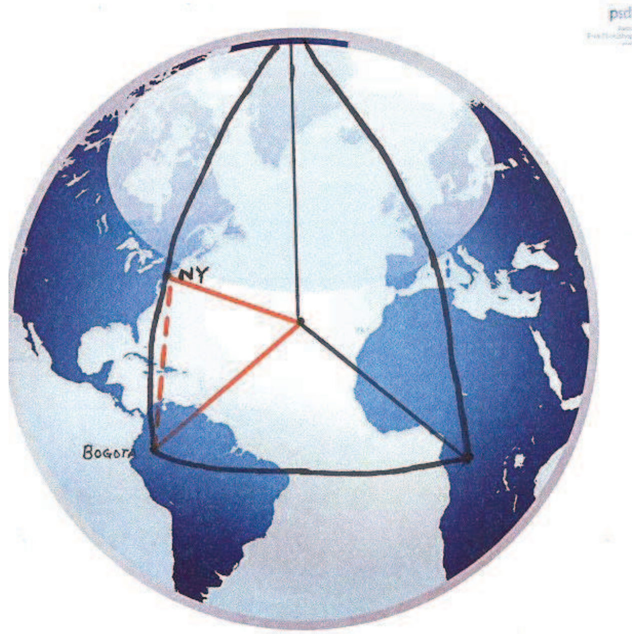
$$d_{\text{Fo}}(\mathbf{A}, \mathbf{B}) = \sqrt{\langle (\text{vec } \mathbf{A} - \text{vec } \mathbf{B}), (\text{vec } \mathbf{A} - \text{vec } \mathbf{B}) \rangle} = d_{\text{E}}(\text{vec } \mathbf{A}, \text{vec } \mathbf{B})$$

- Similar to the practice in applying the Euclidean distance, we can also weight the Frobenius distance by a positive definite matrix $\mathbf{W}$ to enhance certain parts of the feature matrices:

$$d_{\text{WFo}}(\mathbf{A}, \mathbf{B}) = \sqrt{\text{tr}[(\mathbf{A} - \mathbf{B})\mathbf{W}(\mathbf{A} - \mathbf{B})^H]} \quad (4)$$

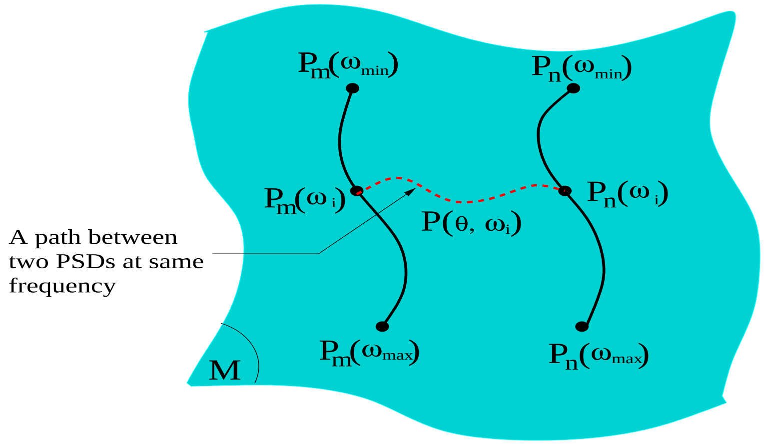
Euclidean Distance for PSD Matrices – Difficulties

PSD matrices are not freely structured, but are: 1) Hermitian and 2) Positive definite. Therefore, they are not just any point in the M^2-dimensional signal space, rather they form a *manifold* in the signal space. Thus, when we consider the distance between two PSD matrices, we should measure along the surface of the manifold. This concept is akin to finding the distance between two cities:



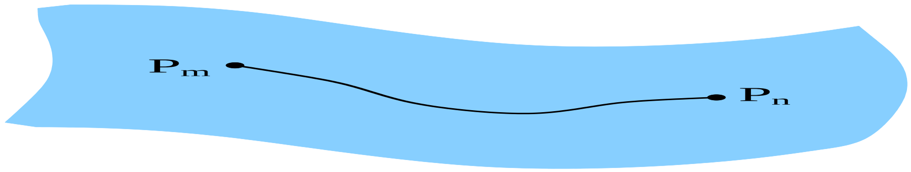
The Euclidean distance between two cities is neither informative nor accurate.

Distance Measure on the PSD Matrix Manifold $\mathcal{M}$



- Let the manifold described by *all* the feature PSD matrices be $\mathcal{M}$, $\therefore$ any point on $\mathcal{M}$ is denoted by $\mathbf{P}$.
- At different frequencies $\omega \in [\omega_{\min}, \omega_{\max}]$, two PSD matrices of the mth and nth epochs describe two sequences of points (two separate curves) on $\mathcal{M}$ denoted by $\mathbf{P}_m(\omega)$ and $\mathbf{P}_n(\omega)$ respectively (see Figure).
- We need to establish a distance which measures along the manifold of $\mathcal{M}$ between the two points $\mathbf{P}_m(\omega_i)$ and $\mathbf{P}_n(\omega_i)$, (or simply, $\mathbf{P}_m$ and $\mathbf{P}_n$ for short), on the two curves at the same frequency.

Some Basic Concepts of a Riemannian Manifold¶



- For a *smooth path* $\mathbf{P}(\theta)$ on $\mathcal{M}$ linking the two points $\mathbf{P}_m$ and $\mathbf{P}_n$ parameterized by θ, the length (real and positive) of the path between the two points is then given by

$$\ell(\mathbf{P}) = \int_{\theta_m}^{\theta_n} \sqrt{g(\dot{\mathbf{P}}(\theta))}\, d\theta, \qquad \left(\dot{\mathbf{P}} \triangleq d\mathbf{P}/d\theta\right)$$

- $g(\dot{\mathbf{P}}(\theta))$ is called a *Riemannian metric*.
- A differentiable manifold $\mathcal{M}$ in which a Riemannian metric is defined is called a *Riemannian manifold*.
- The curve on the Riemannian manifold linking $\mathbf{P}_m$ and $\mathbf{P}_n$ having the minimum length is called a *geodesic*, and the length of the geodesic is called the *Riemannian distance* between the two points, i.e.,

$$d_R(\mathbf{P}_m, \mathbf{P}_n) \triangleq \min_{\mathbf{P}(\theta):[\theta_1,\theta_2]\rightarrow\mathcal{M}} \{\ell(\mathbf{P}(\theta))\}$$

¶ After the great German mathematician Bernhard Riemann (1826-1866)

The Manifold $\mathcal{M}$ and its Associated Euclidean Spaces

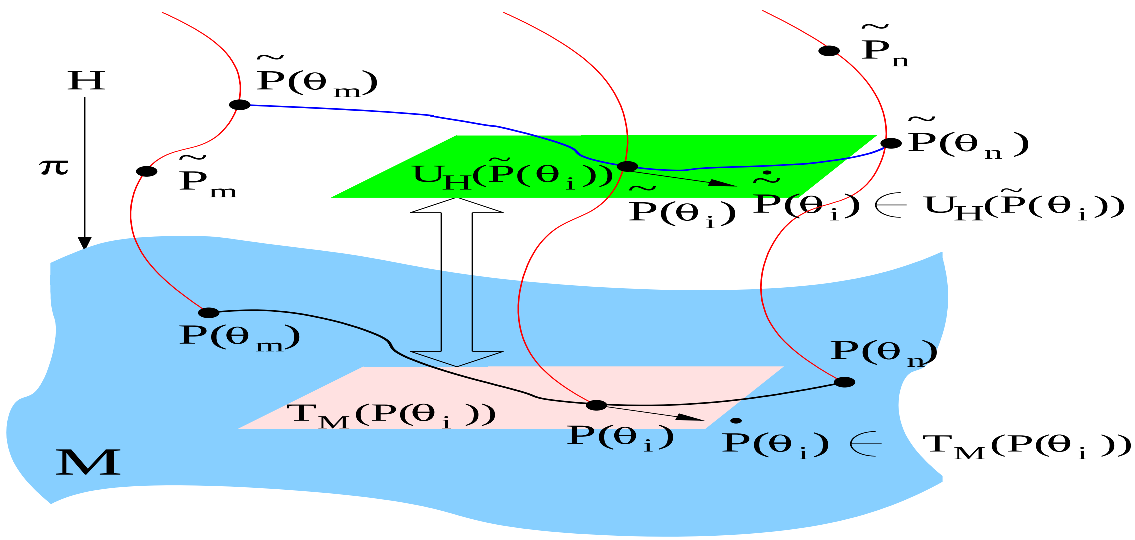
To obtain the Riemannian distance on $\mathcal{M}$, we must study the relationship between the manifold and the Euclidean spaces associated with it:

- Any $\mathbf{P}$ can be decomposed (*not uniquely*) as $\mathbf{P} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$ where $\tilde{\mathbf{P}}$ has no structural constraint, and is an $M \times M$ complex matrix in a Euclidean space $\mathcal{H}$. $\mathbf{P}$ is thus an "image" on $\mathcal{M}$ of the points $\tilde{\mathbf{P}}$ in $\mathcal{H}$. $\mathcal{M}$ and $\mathcal{H}$ are called the *base space* and the *total space*.
- The elements in $\mathcal{M}$ and $\mathcal{H}$ are related by $\mathbf{P} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$. This mapping links each point $\mathbf{P} \in \mathcal{M}$ with points $\tilde{\mathbf{P}} \in \mathcal{H}$, and this "link" is called the *fibre* above $\mathbf{P}$. Any point $\tilde{\mathbf{P}}$ along a fibre satisfies the mapping. A collection of these links is called a *fibre bundle*.
- The tangent space $\mathcal{T}_{\mathcal{H}}(\tilde{\mathbf{P}})$ at $\tilde{\mathbf{P}}$ in $\mathcal{H}$ is the collection of vectors tangent to any smooth curve passing through $\tilde{\mathbf{P}}$, and can be resolved into its horizontal and vertical subspaces $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$ and $\mathcal{V}_{\mathcal{H}}(\tilde{\mathbf{P}})$ respectively, i.e., $\mathcal{T}_{\mathcal{H}}(\tilde{\mathbf{P}}) = \mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}}) \oplus \mathcal{V}_{\mathcal{H}}(\tilde{\mathbf{P}})$.
- $\mathcal{H}$, $\mathcal{T}_{\mathcal{H}}$, $\mathcal{U}_{\mathcal{H}}$, and $\mathcal{V}_{\mathcal{H}}$ are Euclidean spaces, thus the Euclidean (Frobenius) norm of vectors applies making the distance measure to be:

$$d_E(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}) = \|\tilde{\mathbf{X}} - \tilde{\mathbf{Y}}\| = \sqrt{\mathrm{tr}(\tilde{\mathbf{X}}^H - \tilde{\mathbf{Y}}^H)(\tilde{\mathbf{X}} - \tilde{\mathbf{Y}})}$$

$\mathcal{M}$ and its Associated Euclidean Spaces (contd.)

- Consider a smooth curve $\mathbf{P}(\theta)$ on $\mathcal{M}$ with end points $\mathbf{P}_m = \mathbf{P}(\theta_m)$ and $\mathbf{P}_n = \mathbf{P}(\theta_n)$.
- A *lift* of the curve $\mathbf{P}(\theta)$ is a curve $\tilde{\mathbf{P}}(\theta)$ in $\mathcal{H}$ along the fibres above $\mathbf{P}$.
- A *horizontal lift* of the curve $\mathbf{P}(\theta)$ on $\mathcal{M}$ is a curve $\tilde{\mathbf{P}}(\theta)$ for which the tangent vector $\dot{\tilde{\mathbf{P}}}_i$ always lies in the horizontal subspace $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$ of the tangent space $\mathcal{T}_{\mathcal{H}}(\tilde{\mathbf{P}})$ at each point $\tilde{\mathbf{P}}_i$ along $\tilde{\mathbf{P}}(\theta)$.
- The concept of the horizontal lift of a PSD matrix is illustrated in the following figure.



Developing a Riemannian Distance on $\mathcal{M}$

Lemma 1. *Let* $\mathbf{P} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H;\ \tilde{\mathbf{P}} \in \mathcal{H},\ \mathbf{P} \in \mathcal{M}$. *If we choose the Riemannian metric on* $\mathcal{M}$ *to be* $g(\dot{\mathbf{P}}) = \frac{1}{2}\mathrm{tr}\dot{\mathbf{P}}\mathbf{K}$, *where* $\dot{\mathbf{P}} \in \mathcal{T}_{\mathcal{M}}(\mathbf{P})$, *and* $\mathbf{K}$ *is a Hermitian matrix such that* $\mathbf{KP} + \mathbf{PK} = \dot{\mathbf{P}}$, *then* $\mathcal{T}_{\mathcal{M}}(\mathbf{P})$ *and* $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$ *are isometric.* □

- By Lemma 1, if we use the particular Riemannian metric, then the tangent space $\mathcal{T}_{\mathcal{M}}(\mathbf{P})$ of $\mathcal{M}$ and the horizontal subspace $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$ of $\mathcal{H}$ are isometric. Thus, by lifting a curve between two points in $\mathcal{M}$ horizontally to $\mathcal{U}$, the measurement between the two points on $\mathcal{M}$ can be obtained from the corresponding measurement in $\mathcal{U}$.
- Hence, $\tilde{\mathbf{P}}(\theta)$ being a horizontal lift of $\mathbf{P}(\theta)$, $\mathbf{P}(\theta)$ is a geodesic (path of minimum length) between $\mathbf{P}(\theta_m)$ and $\mathbf{P}(\theta_n)$ in $\mathcal{M}$ iff $\tilde{\mathbf{P}}(\theta)$ is a geodesic between $\tilde{\mathbf{P}}(\theta_m)$ and $\tilde{\mathbf{P}}(\theta_n)$.
- Now, $\tilde{\mathbf{P}}_m$ and $\tilde{\mathbf{P}}_n$ along the two fibres above $\mathbf{P}(\theta_m)$ and $\mathbf{P}(\theta_n)$ are in a Euclidean space, thus the distance between them is a straight line, the length of which is given by $\|\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n\|$.

Developing a Riemannian Distance on $\mathcal{M}$ (contd.)

- If $\mathbf{P}(\theta)$ is a geodesic (minimum-length path) between $\mathbf{P}(\theta_m)$ and $\mathbf{P}(\theta_n)$ in $\mathcal{M}$, then its length must be equal to the length of the shortest straight line joining the two points $\tilde{\mathbf{P}}(\theta_m)$ and $\tilde{\mathbf{P}}(\theta_n)$ in $\mathcal{H}$, which, for $\mathbf{P}_m = \tilde{\mathbf{P}}_m\tilde{\mathbf{P}}_m^H$ and $\mathbf{P}_n = \tilde{\mathbf{P}}_n\tilde{\mathbf{P}}_n^H$, is given by

$$\min_{\tilde{\mathbf{P}}} \ \ell(\tilde{\mathbf{P}}) \ = \min \|\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n\| \ = \min \left[\mathrm{tr} \left\{ (\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)(\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)^H \right\} \right]^{1/2}$$

- Substituting $\tilde{\mathbf{P}}_m = \mathbf{P}_m^{1/2}\mathbf{U}_m$ and $\tilde{\mathbf{P}}_n = \mathbf{P}_n^{1/2}\mathbf{U}_n$ with $\mathbf{U}_m$ and $\mathbf{U}_n$ being unitary matrices, the Riemannian distance between $\mathbf{P}_m$ and $\mathbf{P}_n$ on the manifold $\mathcal{M}$ is

$$d_R(\mathbf{P}_m, \mathbf{P}_n) = \min_{\mathbf{U}_m, \mathbf{U}_n} \left[\mathrm{tr}\mathbf{P}_m + \mathrm{tr}\mathbf{P}_n - 2\Re \left\{ \mathrm{tr} \left(\mathbf{U}_n \mathbf{U}_m^H \mathbf{P}_m^{1/2} \mathbf{P}_n^{1/2} \right) \right\} \right]^{1/2}$$

- Now[♮], $\max_{\mathbf{U}_m, \mathbf{U}_n} \Re \left[\mathrm{tr}(\mathbf{U}_n \mathbf{U}_m^H \mathbf{P}_m^{1/2} \mathbf{P}_n^{1/2}) \right] = \mathrm{tr} \left[(\mathbf{P}_m^{1/2} \mathbf{P}_n \mathbf{P}_m^{1/2})^{1/2} \right]$ if $\mathbf{U}_n$ and $\mathbf{U}_m$ are chosen as the left and right singular vector matrices[♯] of $(\mathbf{P}_n^{1/2}\mathbf{P}_m^{1/2})$.

[♮] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[♯] SVD: $\mathbf{V}_\ell^H \mathbf{A} \mathbf{V}_r = \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$; $\mathbf{A}$ is $m \times n$, $\mathbf{V}_\ell$, $\mathbf{V}_r$ are $m \times m$ and $n \times n$ unitary, $\mathbf{\Sigma}$ is diagonal.

A Riemannian Distance on $\mathcal{M}$ (contd.)

- **Theorem 1.** *For $\mathbf{P}_m, \mathbf{P}_n \in \mathcal{M}$, the Riemannian distance[§] between $\mathbf{P}_m$ and $\mathbf{P}_n$ is:*

$$d_R(\mathbf{P}_m, \mathbf{P}_n) = \sqrt{\mathrm{tr}\mathbf{P}_m + \mathrm{tr}\mathbf{P}_n - 2\mathrm{tr}\left[(\mathbf{P}_m^{1/2}\mathbf{P}_n\mathbf{P}_m^{1/2})^{1/2}\right]} \quad (5)$$

  if $\mathbf{P} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$, and the Riemannian metric is chosen as in Lemma 1. □

- As in the case of Euclidean distance, we often weight the features to enhance their similarity/dissimilarity. This results in the following weighted distance:

  Corollary 1. *Let $\mathbf{W}$ be a positive definite Hermitian matrix. Then the weighted Riemannian distance[§] between $\mathbf{P}_m$ and $\mathbf{P}_n \in \mathcal{M}$ is given by*

$$d_{RW}(\mathbf{P}_m, \mathbf{P}_n) = \sqrt{\mathrm{tr}(\mathbf{W}\mathbf{P}_m) + \mathrm{tr}(\mathbf{W}\mathbf{P}_n) - 2\mathrm{tr}\left[(\mathbf{P}_n^{1/2}\mathbf{W}\mathbf{P}_m\mathbf{W}\mathbf{P}_n^{1/2})^{1/2}\right]} \quad (6)$$

  if $\mathbf{P} = \tilde{\mathbf{P}}\tilde{\mathbf{P}}^H$, and the chosen Riemannian metric is as in Lemma 1. □

- For the complete range of frequency, the distance between two curves is

$$d(\mathbf{P}_m(\omega), \mathbf{P}_n(\omega)) = \sum_i d_{R(W)}(\mathbf{P}_m(\omega_i), \mathbf{P}_n(\omega_i)) \quad (7)$$

[§] We have obtained closed forms for both d_R and d_{RW} which can be shown to satisfy the properties of a distance in Eq. (1).

Optimum Weighting for the Riemannian Distance

The purpose of having a weighted distance is to highlight certain parts of the feature and deemphasize other parts in order to enhance the efficiency of application. When applied to signal classification, our aim is to find the optimum weighting matrix which minimizes the distances between *similar* feature data while keeping the dissimilar members at a prescribed distance.

- We say that $\mathbf{P}_m(\omega)$ and $\mathbf{P}_n(\omega)$ are similar if they belong to the same class, and are dissimilar if they belong to different classes. Let
  $\mathcal{A}_s = \{(\mathbf{P}_m, \mathbf{P}_n); \mathbf{P}_m(\omega), \mathbf{P}_n(\omega) \in \mathcal{C}_\ell \ \forall\omega\}$ — the set of similar PSD matrix pairs
  $\mathcal{A}_d = \{(\mathbf{P}_m, \mathbf{P}_n); \mathbf{P}_m(\omega) \in \mathcal{C}_{\ell m}, \mathbf{P}_n(\omega) \in \mathcal{C}_{\ell n}, \ell_m \neq \ell_n \ \forall\omega\}$ — the set of pairs of dissimilar PSD matrices.
- The optimum $M \times M$ weighting matrix $\mathbf{W}$ may be found by maximizing the ratio of the sum of squared *interclass* distances and the sum of squared *intraclass* distances,

$$\max_{\mathbf{W}} \frac{\sum_{(\mathbf{P}_{mk}, \mathbf{P}_{nk}) \in \mathcal{A}_d} d_{RW}^2(\mathbf{P}_{mk}, \mathbf{P}_{nk})}{\sum_{(\mathbf{P}_{mk}, \mathbf{P}_{nk}) \in \mathcal{A}_s} d_{RW}^2(\mathbf{P}_{mk}, \mathbf{P}_{nk})} \qquad s.t. \quad \mathbf{W} = \mathbf{W}^H \succ \mathbf{0} \tag{8}$$

- Optimization of the quantity in Eq. (8) directly on the manifold $\mathcal{M}$ is difficult because of the relatively complicated expression of d_{RW}.

Optimum Weighting Matrix (contd.)

- However, $\mathcal{T}_{\mathcal{M}}(\mathbf{P})$ is *isometric* with $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$
  $\Rightarrow$ we can perform the optimization in $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$ using the *inner product metric*.
- Let $\mathbf{W} = \mathbf{\Omega}\mathbf{\Omega}^H$ where $\mathbf{\Omega}$ is $M \times K$, $K < M$. Isometry of $\mathcal{T}_{\mathcal{M}}(\mathbf{P})$ and $\mathcal{U}_{\mathcal{H}}(\tilde{\mathbf{P}})$ means
$$d_{RW}^2(\mathbf{P}_m, \mathbf{P}_n) = \mathrm{tr}[\mathbf{\Omega}^H(\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)(\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)^H\mathbf{\Omega}]$$
- Then, from our library, we gather all the similar $\tilde{\mathbf{P}}$ matrices in one group and the dissimilar $\tilde{\mathbf{P}}$ matrices in another and evaluate their mean difference matrices, i.e.,
$$\tilde{\mathbf{M}}_{\mathcal{A}_s} = \sum_{(\tilde{\mathbf{P}}_m, \tilde{\mathbf{P}}_n) \in \mathcal{A}_s} (\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)(\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)^H \quad \text{and}$$
$$\tilde{\mathbf{M}}_{\mathcal{A}_d} = \sum_{(\tilde{\mathbf{P}}_m, \tilde{\mathbf{P}}_n) \in \mathcal{A}_d} (\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)(\tilde{\mathbf{P}}_m - \tilde{\mathbf{P}}_n)^H.$$
- The optimization problem (8) becomes
$$\max_{\mathbf{\Omega}} \frac{\mathrm{tr}(\mathbf{\Omega}^H \tilde{\mathbf{M}}_{\mathcal{A}_d} \mathbf{\Omega})}{\mathrm{tr}(\mathbf{\Omega}^H \tilde{\mathbf{M}}_{\mathcal{A}_s} \mathbf{\Omega})} \approx \max_{\mathbf{\Omega}} \mathrm{tr}\left[\mathbf{\Omega}^H \tilde{\mathbf{M}}_{\mathcal{A}_d} \mathbf{\Omega}\right], \quad \text{s.t.} \quad (\mathbf{\Omega}^H \tilde{\mathbf{M}}_{\mathcal{A}_s} \mathbf{\Omega}) = \mathbf{I}_K$$
- The well-known solution is $\mathbf{\Omega}_{\mathrm{op}} = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_K]$ where $\mathbf{v}_1, \cdots, \mathbf{v}_K$ are the eigenvectors[†] corresponding to the K largest eigenvalues $\lambda_1 \geq \cdots \geq \lambda_K$ of $\tilde{\mathbf{M}}_{\mathcal{A}_s}^{-1} \tilde{\mathbf{M}}_{\mathcal{A}_d}$. Thus, $\mathbf{W}_{\mathrm{op}} = \mathbf{\Omega}_{\mathrm{op}} \mathbf{\Omega}_{\mathrm{op}}^H = [\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_K][\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_K]^H$.

[†] H. Lütkepohl, *Handbook of Matrices*, John Wiley and Sons, 1996. $\{\mathbf{v}_1, \quad \cdots . \quad \mathbf{v}_K\}$ are not necessarily orthonormal.

Sleep State Classification using EEG Signals

- EEG (electroencephalogram) is a valuable measure of the brain's electrical function and has been employed in many clinical areas such as administration of anaesthetics, detection and prediction of epileptic seizures, recognition of concussion, analysis of depression, etc. Here, we employ the EEG measurements to determine the *depth of natural sleep* (no anaesthetics) of a patient.
- The study of sleep is important in health care since sleep disorders affect the well-being of many individuals. The states of sleep are classified as shown in the following table†:

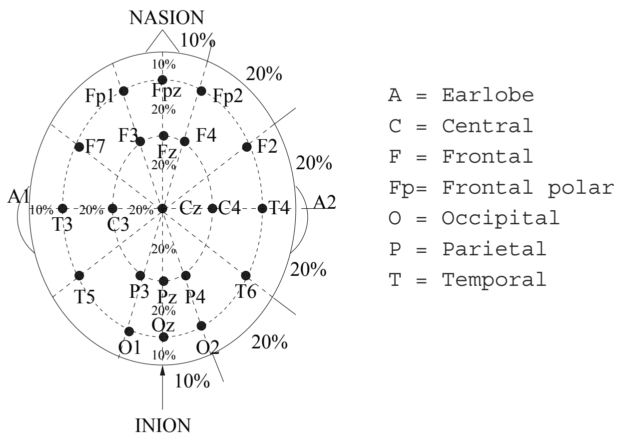
Rechtschaffen & Kales Sleep Classification *US Govt. Health Services* 1968

Sleep Stage	Main Frequency Range	Wave Patterns
Awake (relaxed)	$8 - 12$Hz or higher	α, β, or γ
REM	> 12Hz	β
Non-REM 1	$4 - 8$Hz	α, θ
Non-REM 2	$4 - 15$Hz	θ, spindles, K-complexes
Non-REM 3	$2 - 4$Hz	δ, θ
Non-REM 4	$0.5 - 2$Hz	δ, θ

† δ, θ, α, β, γ corresponds to $0.5-4$Hz, $4-8$Hz, $8-12$Hz, $12-30$Hz, above 30Hz. A *sleep spindle* is a burst of $12-16$ Hz waves that occur for $0.5-1.5$ sec., and a *K-complex* is a brief positive and negative spikes that occurs every $1.0 - 1.7$ min.

Sleep State Classification using EEG Signals (contd.)

- Collection of EEG data for the analysis of sleep requires putting the patient in a laboratory to acquire up to 8 hours of multiple channels of EEG recordings‡.



The 10-20 electrode placement system (only C_3,C_4,O_1,O_2 are needed for sleep analysis)

- The collected EEG signals is divided into epochs of 30 sec each and analysis is usually carried out by a trained clinical expert *visually* to determine the stage of sleep. The judgement of the expert is usually based on the amplitude and frequency of the signal fluctuation (equivalent to power at different frequencies)§.

‡There may be recordings of other physiological data used for determination of sleep state such as EMG, ECG, oxygen concentration in arterial blood, and breathing rate. Here, we only focus our attention on the use of EEG signals.
§This is a labour-intensive classification process, and its inter-rater reliability between two expert judges is typically around 77%.

Analytic Approaches to Sleep State Classification

- Many researchers have proposed sleep stage classification using analytic approaches:
  - Establish an AR model for the single-channel signals in each sleep state and determine their spectral properties using the DFT.
  - Employ the *multi-channel* EEG data to determine cross-correlation of the signals of different channels and evaluate the PSD of the different sleep-state data.
  - The suggested analytic methods propose the second order statistics (correlation, power spectrum, cross PSD) of the EEG data as the *feature* for sleep state determination is similar ro the judgments of the clinical experts being based mainly on the power distribution of the signal.
- However, for the measure of similarity/dissimilarity of the features, all these analytic methods employ the Euclidean distance, and all approaches have only limited success rate – in the range of 70% to 80% agreement with those of the experts.
- At present, the most common practice is still to have the EEG signals determined by the visual judgment of clinical experts on the EEG signals!

Sleep State Classification using Riemannian Distance

To test how effective is our proposed measure of Riemannian distance on EEG signal classification, we carried out experiments as follows:

1. Record several hours of 4-channel EEG data from different patients.
2. Remove artifacts and LP filter (cut-off 30Hz). Divide cleaned data into 30-sec "epochs" and normalize.
3. Pass epochs to clinical experts for sleep classification*.
4. Construct PSD matrix for each categorized epoch (N-S algorithm) at different frequencies. Using PSD matrices in all similar and dissimilar classes, calculate the optimum weighting matrix for the Riemannian distance.
5. For each trial, randomly choose *one* feature matrix from a state as the test signal. calculate the *R-distances* of the test signal from the library sets. Classify test signal according to the k-nearest neighbour algorithm.
6. Repeat trial Q times (Q-fold cross-validation), each time selecting a different test feature curve in a class. Record the number of correct classification for each state.

*We collected 75 epochs for each sleep state (450 epochs altogether). Since the true state of sleep measured from the signal epoch is not known, we therefore treat the library of signal epochs classified by clinical experts as the ground truth.

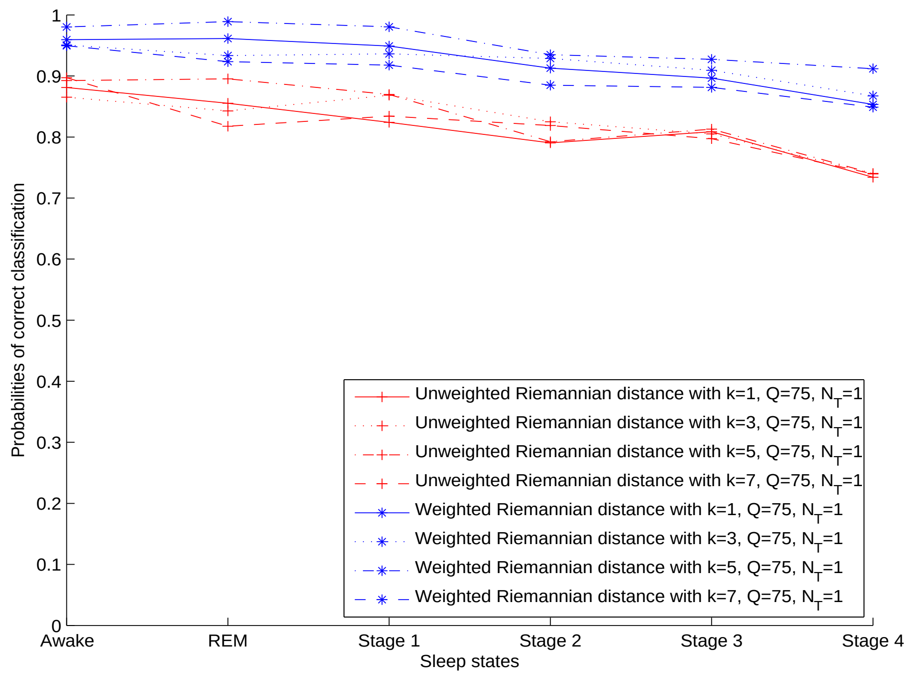
Classification of EEG Signals – Test Results

Example 1.

- We examine the performance of our classification algorithm using the unweighted and weighted Riemannian distances and observe the effect of optimum weighting on the accuracy of classification.
- For a selected PSD curve chosen from a particular class, we carry out our experiments using the parameter $k = 1, 3, 5$, and 7 for the nearest neighbour tests.
- Each test is repeated $Q = 75$ times using both weighted and unweighted Riemannian distances. The following figure shows the averaged results of the tests.
- It can be observed that having $k = 5$ as the number of nearest neighbours yields the best results for both distance measures.

Test Results – Example 1 (contd.)

- With the weighted Riemannian distance, the accuracy of classification ranges from about 94% for Stage 4 on Non-REM to over 99% for Stage 1 of Non-REM and REM.
- It is also clear that the optimum weighting improves the performance of the Riemannian distance by a margin of 8 to 12% in the accuracy of sleep state classification.



Comparison of Various Distances – Test Results

Example 2. We now compare the EEG classification performance using different distance measures. Included in the performance comparison with the unweighted and weighted Riemannian distances are the following commonly used measures:

a) Euclidean distances — For EEG classification, many researchers choose a vector quantity (such as the coefficients of the Fourier transform, or of the wavelet transform, etc.) of the single-channel[♮] EEG signal as the signal feature. The multi-channel PSD matrix provides extra information on the cross-power between channels. Here we will maintain the use of the multi-channel PSD matrix as the signal feature, and endowed it with the optimally[♯] weighted or unweighted Euclidean (Frobenius) distance[♭] as shown before.

[♮]For non-parametric processing of stationary signals, the different choices generally yield similar accuracies in EEG classification since these coefficients contain essentially the same information on the single-channel signal. For multi-channel measurements, the procedure usually involves the averaging of the chosen coefficients over the different channels.

[♯]Xing *et als. Neural Inf. Proc. Sys.*, 2003. Several variations of the same idea of optimal weighting for similarity of vector features exist in the literature. The performance of these differently optimized weighted distances however, are very similar in our tests.

[♭]There are other ways of writing a PSD matrix as a vector quantity such as writing it as a linear combination of the eigenvectors of $\mathbf{P}$ (Karhunen-Loéve expansion). However, most representations yield similar accuracies in EEG classification.

Test Results – Example 2 (con't)

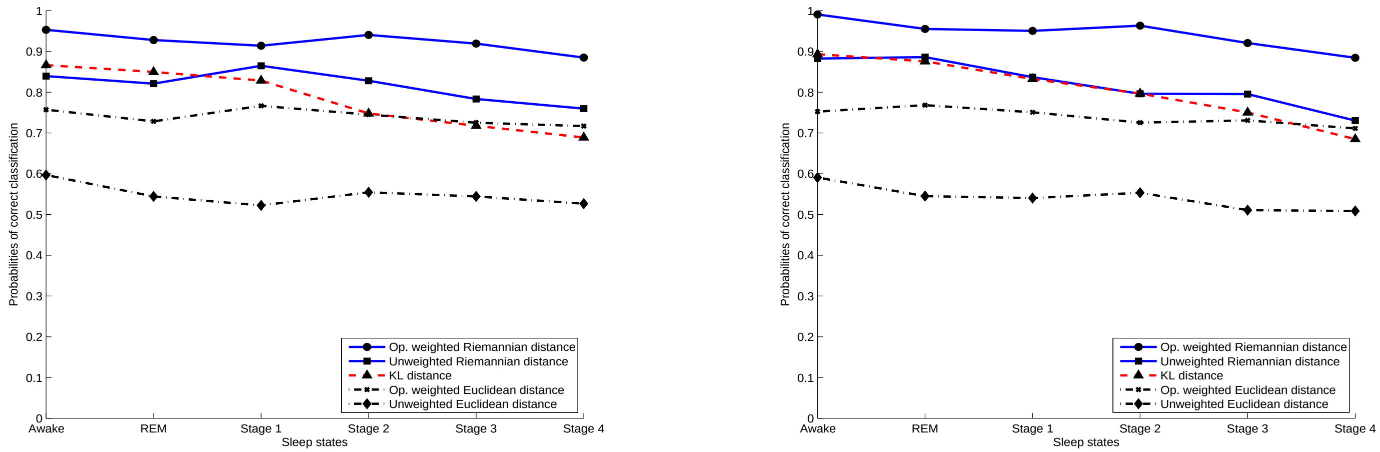
b) Kullback-Leibler (KL) "distance" – The K-L *divergence* was developed to measure the similarity between random variables of different distributions. For two zero-mean Gaussian random vectors, the KL divergence becomes a similarity measure between the two covariance matrices given by

$$d_{KL}(\mathbf{R}_1, \mathbf{R}_2) = \sqrt{\tfrac{1}{2}\,\mathrm{tr}(\mathbf{R}_1\mathbf{R}_2^{-1} + \mathbf{R}_1^{-1}\mathbf{R}_2 - 2\mathbf{I})}$$

The KL "distance" is invariant to weighting and does not satisfy the triangular inequality. Since it is a measure between the covariance matrices, the Fourier transforms of which are the feature PSD matrices, it has been applied to the classification of EEG signals.

We now apply the k-nearest neighbour algorithm to validate the EEG classification using all the above distance measures. In all cases, we choose $Q = 75$. We only show the performance comparison for $k = 3$ and 5 since these appear to yield the best results among all our tests. Again, each test follows the same procedure as outlined in Example 1, and the performance of the classification using different distance measures is evaluated.

Test Results – Example 2 (contd.)



Performance of various methods for: a) k = 3, b) k = 5

- Using the unweighted Euclidean distance d_E yields the worst performance by far, having accuracies between 50% to 60% only in both cases of $k = 3$ and $k = 5$.
- Optimum weighting improves the performance of the Euc. distance by approx. 15%.
- In the Awake, REM, and Stage 1 categories, the KL distance performs with accuracies at around 85% to 90%. matching that of the unweighted Riemannian distance, but markedly deteriorates in the deeper sleeping stages of 2, 3, and 4.
- The weighted Riemannian distance yields a performance superior to all, being 8 to 15% higher in accuracy over the closest competitor (the unweighted Riemannian distance).

Conclusions

- We study the problem of signal classification employing PSD as the feature.
- We examine the use of the Euclidean distance as a measure of similarity between PSD matrices and reason that it may not be a suitable measure since PSD describe a manifold in the signal space.
- We strongly suggest that, for PSD matrices, the measure of similarity should be the distance measured along the surface of the manifold — i.e., the Riemannian distance.
- We examine the geometry of the PSD manifold which has a measure of Riemannian distance and its associated total space which has a measure of Euclidean distance. We established isometry between the tangent space of the PSD manifold and the horizontal component of the total space tangent.
- From this, we obtained a Riemannian distance measure for the manifold, and we further developed an optimum weighting for the Riemannian distance for signal classification.
- Our objective here is not to compare classification methods be it the measurement from the mean or the k-NN measurement; neither is it our purpose to compare different optimum weightings. Our purpose here is to show that the employment of this new measure of Riemannian distance yields the advantage of superior accuracies in the application to signal classification by PSD over the conventional distance measures.

Thank you!