Explicit Signal to Noise Ratio in Reproducing Kernel Hilbert Spaces

Luis Gómez-Chova\textsuperscript{1}  Allan A. Nielsen\textsuperscript{2}  Gustavo Camps-Valls\textsuperscript{1}

\textsuperscript{1}Image Processing Laboratory (IPL), Universitat de València, Spain. luis.gomez-chova@uv.es, http://www.valencia.edu/chovago

\textsuperscript{2}DTU Space - National Space Institute. Technical University of Denmark.

IGARSS 2011 – Vancouver, Canada
Outline

1. Introduction
2. Signal-to-noise ratio transformation
3. Kernel Minimum Noise Fraction
4. Experimental Results
5. Conclusions and Open questions
Feature Extraction

- Feature selection/extraction is essential before classification or regression
  - to discard redundant or noisy components
  - to reduce the dimensionality of the data
- Create a subset of new features by combinations of the existing ones

Linear Feature Extraction

- Linear methods offer Interpretability $\sim$ knowledge discovery
  - PCA: projections maximizing the data set variance
  - PLS: projections maximally aligned with the labels
  - ICA: non-orthogonal projections with maximal independent axes
Feature Extraction

- Feature selection/extraction is essential before classification or regression
  - to discard redundant or noisy components
  - to reduce the dimensionality of the data
- Create a subset of new features by combinations of the existing ones

Linear Feature Extraction

- Linear methods offer Interpretability $\sim$ knowledge discovery
  - PCA: projections maximizing the data set variance
  - PLS: projections maximally aligned with the labels
  - ICA: non-orthogonal projections with maximal independent axes

Drawbacks

1. Most feature extractors disregard the noise characteristics!
2. Linear methods fail when data distributions are curved (nonlinear relations)
Objectives

- New nonlinear kernel feature extraction method for remote sensing data
- Extract features robust to data noise

Method

- Based on the Minimum Noise Fraction (MNF) transformation
- Explicit Kernel MNF (KMNF)
  - Noise is explicitly estimated in the reproducing kernel Hilbert space
  - Deals with non-linear relations between the noise and signal features jointly
  - Reduces the number of free parameters in the formulation to one

Experiments

- PCA, MNF, KPCA, and two versions of KMNF (implicit and explicit)
- Test feature extractors for real hyperspectral image classification
1 Introduction

2 Signal-to-noise ratio transformation

3 Kernel Minimum Noise Fraction

4 Experimental Results

5 Conclusions and Open questions
Signal and noise

Signal vs noise
- Signal: magnitude generated by an inaccessible system, $s_i$
- Noise: magnitude generated by the medium corrupting the signal, $n_i$
- Observation: signal corrupted by noise, $x_i$

Notation
- Observations: $x_i \in \mathbb{R}^N$, $i = 1, \ldots, n$
- Matrix notation: $X = [x_1, \ldots, x_n]^\top \in \mathbb{R}^{n \times N}$
- Centered data sets: assume $X$ has zero mean
- Empirical covariance matrix: $C_{xx} = \frac{1}{n}X^\top X$
- Projection matrix: $U$ (size $N \times n_p$) $\rightarrow X' = XU$ ($n_p$ extracted features)
Principal Component Analysis (PCA)

- Find projections of $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^\top$ maximizing the variance of data $\mathbf{XU}$

\[
\text{PCA: maximize: } \text{Trace}\{(\mathbf{XU})^\top(\mathbf{XU})\} = \text{Trace}\{\mathbf{U}^\top \mathbf{C}_{xx} \mathbf{U}\} \\
\text{subject to: } \mathbf{U}^\top \mathbf{U} = \mathbf{I}
\]

- Including Lagrange multipliers $\lambda$, this is equivalent to the eigenproblem

\[
\mathbf{C}_{xx} \mathbf{u}_i = \lambda_i \mathbf{u}_i \rightarrow \mathbf{C}_{xx} \mathbf{U} = \mathbf{UD}
\]

$\mathbf{u}_i$ are the eigenvectors of $\mathbf{C}_{xx}$ and they are orthonormal, $\mathbf{u}_i^\top \mathbf{u}_j = 0$
Principal Component Analysis Transformation

Principal Component Analysis (PCA)

- Find projections of $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^\top$ maximizing the variance of data $\mathbf{XU}$

  PCA: maximize: $\text{Trace}\{(\mathbf{XU})^\top(\mathbf{XU})\} = \text{Trace}\{\mathbf{U}^\top\mathbf{C}_{xx}\mathbf{U}\}$
  subject to: $\mathbf{U}^\top\mathbf{U} = \mathbf{I}$

- Including Lagrange multipliers $\lambda$, this is equivalent to the eigenproblem

  $\mathbf{C}_{xx}\mathbf{u}_i = \lambda_i \mathbf{u}_i \rightarrow \mathbf{C}_{xx}\mathbf{U} = \mathbf{UD}$

  $\mathbf{u}_i$ are the eigenvectors of $\mathbf{C}_{xx}$ and they are orthonormal, $\mathbf{u}_i^\top\mathbf{u}_j = 0$

PCA limitations

1. Axes rotation to the directions of maximum variance of data
2. It does not consider noise characteristics:
   - Assumes noise variance is low $\rightarrow$ last eigenvectors with low eigenvalues
   - Maximum variance directions may be affected by noise
Minimum Noise Fraction Transformation

The SNR transformation

Find projections maximizing the ratio between signal and noise variances:

\[ \text{SNR: maximize: } \frac{\text{Tr } U^\top C_{ss} U}{\text{Tr } U^\top C_{nn} U} \]

subject to: \[ U^\top C_{nn} U = I \]

Unknown signal and noise covariance matrices \( C_{ss} \) and \( C_{nn} \).
## Minimum Noise Fraction Transformation

### The SNR transformation

- **Find projections maximizing the ratio between signal and noise variances:**
  
  \[
  \text{SNR: maximize: } \text{Tr} \left\{ \frac{U^\top C_{ss} U}{U^\top C_{nn} U} \right\}
  \]
  
  subject to: \(U^\top C_{nn} U = I\)

- Unknown signal and noise covariance matrices \(C_{ss}\) and \(C_{nn}\)

### The MNF transformation

- Assuming additive \(X = S + N\) and orthogonal \(S^\top N = N^\top S = 0\) noise
- Maximizing SNR is equivalent to Minimizing \(NF = 1/(\text{SNR}+1)\):
  
  \[
  \text{MNF: maximize: } \text{Tr} \left\{ \frac{U^\top C_{xx} U}{U^\top C_{nn} U} \right\}
  \]
  
  subject to: \(U^\top C_{nn} U = I\)

- This is equivalent to solving the generalized eigenproblem:
  
  \[
  C_{xx} u_i = \lambda_i C_{nn} u_i \rightarrow C_{xx} U = C_{nn} UD
  \]
The MNF transformation

- Minimum Noise Fraction equivalent to solve the generalized eigenproblem:

\[ C_{xx} u_i = \lambda_i C_{nn} u_i \rightarrow C_{xx} U = C_{nn} U D \]

- Since \( U^T C_{nn} U = I \), eigenvalues \( \lambda_i \) are the SNR+1 in the projected space

- Need estimates of signal \( C_{xx} = X^T X \) and noise \( C_{nn} \approx N^T N \) covariances
Minimum Noise Fraction Transformation

The MNF transformation

- Minimum Noise Fraction equivalent to solve the generalized eigenproblem:

\[ C_{xx} u_i = \lambda_i C_{nn} u_i \rightarrow C_{xx} U = C_{nn} UD \]

- Since \( U^T C_{nn} U = I \), eigenvalues \( \lambda_i \) are the SNR+1 in the projected space.

- Need estimates of signal \( C_{xx} = X^T X \) and noise \( C_{nn} \approx N^T N \) covariances.

The noise covariance estimation

- Noise estimate: diff. between actual value and a reference ‘clean’ value

\[ N = X - X_r \]

\( X_r \) from neighborhood assuming a spatially smoother signal than the noise.

- Assume stationary processes in wide sense:
  - Differentiation: \( n_i \approx x_i - x_{i-1} \)
  - Smoothing filtering: \( n_i \approx x_i - \frac{1}{M} \sum_{k=1}^{M} w_k x_{i-k} \)
  - Wiener estimates
  - Wavelet domain estimates
1 Introduction

2 Signal-to-noise ratio transformation

3 Kernel Minimum Noise Fraction

4 Experimental Results

5 Conclusions and Open questions
Kernel methods for non-linear feature extraction

Kernel methods

1. Map the data to a high-dimensional feature space, $\mathcal{H} (d_\mathcal{H} \to \infty)$
2. Solve a linear problem there

Kernel trick

- No need to know $d_\mathcal{H} \to \infty$ coordinates for each mapped sample $\phi(x_i)$
- **Kernel trick**: “if an algorithm can be expressed in the form of dot products, its non-linear (kernel) version only needs the dot products among mapped samples, the so-called kernel function:”

  \[ K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \]

- Using this trick, we can implement K-PCA, K-PLS, K-ICA, etc
Kernel Principal Component Analysis (KPCA)

- Find projections maximizing variance of mapped data $[\phi(x_1), \ldots, \phi(x_N)]^T$

  KPCA: maximize: $\text{Tr}\{(\Phi U)^T (\Phi U)\} = \text{Tr}\{U^T \Phi^T \Phi U\}$

  subject to: $U^T U = I$

- The covariance matrix $\Phi^T \Phi$ and projection matrix $U$ are $d_H \times d_H$ !

KPCA through kernel trick

- Apply the representer’s theorem: $U = \Phi^T A$ where $A = [\alpha_1, \ldots, \alpha_N]$

  KPCA: maximize: $\text{Tr}\{A^T \Phi \Phi^T \Phi^T \Phi A\} = \text{Tr}\{A^T KKA\}$

  subject to: $U^T U = A^T \Phi \Phi^T A = A^T K A = I$

- Including Lagrange multipliers $\lambda$, this is equivalent to the eigenproblem

  \[ KK \alpha_i = \lambda_i K \alpha_i \rightarrow K \alpha_i = \lambda_i \alpha_i \]

  Now matrix $A$ is $N \times N$ !!! (eigendecomposition of $K$)

- Projections are obtained as $\Phi U = \Phi \Phi^T A = KA$
Kernel MNF Transformation

**KMNF through kernel trick**

- **Find projections maximizing SNR of mapped data** \([\phi(x_1), \ldots, \phi(x_N)]^T\)
  - Replace \(X \in \mathbb{R}^{n \times N}\) with \(\Phi \in \mathbb{R}^{n \times N_H}\)
  - Replace \(N \in \mathbb{R}^{n \times N}\) with \(\Phi_n \in \mathbb{R}^{n \times N_G}\)

\[
C_{xx}U = C_{nn}UD \Rightarrow \Phi^T \Phi U = \Phi_n^T \Phi_n UD
\]

- **Not solvable:** matrices \(\Phi^T \Phi\) and \(\Phi_n^T \Phi_n\) are \(N_H \times N_H\) and \(N_G \times N_G\)
- Left multiply both sides by \(\Phi\), and use representer’s theorem, \(U = \Phi^T A\):

\[
\Phi \Phi^T \Phi \Phi^T A = \Phi \Phi_n^T \Phi_n \Phi^T A D \rightarrow K_{xx}K_{xx}A = K_{xn}K_{xn}AD
\]

- Now matrix \(A\) is \(N \times N\) !!! (eigendecomposition of \(K_{xx}\) wrt \(K_{xn}\))
- \(K_{xx} = \Phi \Phi^T\) is symmetric with elements \(K(x_i, x_j)\)
- \(K_{xn} = \Phi \Phi_n^T = K_{nx}^T\) is non-symmetric with elements \(K(x_i, n_j)\)

- Easy and simple to program!
- Potentially useful when signal and noise are nonlinearly related
Implicit KMNF: noise estimate in the input space

- Estimate the noise directly in the input space: \( \mathbf{N} = \mathbf{X} - \mathbf{X}_r \)
- Signal-to-noise kernel:
  \[
  K_{xn} = \Phi \Phi_n^\top \rightarrow K(x_i, n_j)
  \]
  with \( \Phi_n^\top = [\phi(n_1), \ldots, \phi(n_n)] \)
- Kernels \( K_{xx} \) and \( K_{xn} \) dealing with objects of different nature → 2 params
- Two different kernel spaces → eigenvalues have no longer meaning of SNR

Explicit KMNF: noise estimate in the feature space

- Estimate the noise explicitly in the Hilbert space: \( \Phi_n = \Phi - \Phi_r \)
- Signal-to-noise kernel:
  \[
  K_{xn} = \Phi \Phi_n^\top = \Phi (\Phi - \Phi_r)^\top = \Phi \Phi^\top - \Phi \Phi_r^\top = K_{xx} - K_{xr}
  \]
- Again it is not symmetric \( K(x_i, r_j) \neq K(r_i, x_j) \)
- Advantage: same kernel parameter for \( K_{xx} \) and \( K_{xn} \)
Explicit KMNF: nearest reference

- Differentiation in feature space: \( \phi_{n_i} \approx \phi(x_i) - \phi(x_{i,d}) \)

\[
(K_{xn})_{ij} \approx \langle \phi(x_i), \phi(x_j) - \phi(x_{j,d}) \rangle = K(x_i, x_j) - K(x_i, x_{j,d})
\]

Explicit KMNF: averaged reference

- Difference to a local average in feature space (e.g. 4-connected neighboring pixels): \( \phi_{n_i} \approx \phi(x_i) - \frac{1}{D} \sum_{d=1}^{D} \phi(x_{i,d}) \)

\[
(K_{xn})_{ij} \approx \langle \phi(x_i), \phi(x_j) - \frac{1}{D} \sum_{d=1}^{D} \phi(x_{j,d}) \rangle = K(x_i, x_j) - \frac{1}{D} \sum_{d=1}^{D} K(x_i, x_{j,d})
\]

Explicit KMNF: autoregression reference

- Weight the relevance of each kernel in the summation:

\[
(K_{xn})_{ij} \approx \langle \phi(x_i), \phi(x_j) - \sum_{d=1}^{D} w_d \phi(x_{j,d}) \rangle = K(x_i, x_j) - \sum_{d=1}^{D} w_d K(x_i, x_{j,d})
\]
1 Introduction

2 Signal-to-noise ratio transformation

3 Kernel Minimum Noise Fraction

4 Experimental Results

5 Conclusions and Open questions
## Experimental results

### Data material
- AVIRIS hyperspectral image (220-bands): Indian Pine test site
- $145 \times 145$ pixels, 16 crop types classes, 10366 labeled pixels
- The 20 noisy bands in the water absorption region are intentionally kept

### Experimental setup
- PCA, MNF, KPCA, and two versions of KMNF (implicit and explicit)
- The 220 bands transformed into a lower dimensional space of 18 features
Visual inspection: extracted features in descending order of relevance

<table>
<thead>
<tr>
<th>Features</th>
<th>1–3</th>
<th>4–6</th>
<th>7–9</th>
<th>10–12</th>
<th>13–15</th>
<th>16–18</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MNF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPCA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>implicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KMNF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>explicit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KMNF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analysis of the eigenvalues: signal variance and SNR of transformed data

- Signal variance of the transformed data for PCA
- SNR of the transformed data for MNF and KMNF

The proposed approach provides the highest SNR!
LDA classifier: land-cover classification accuracy

- Best results: linear MNF and the proposed KMNF
- The proposed KMNF method outperforms MNF when the image is corrupted with non-additive noise
LDA classifier: land-cover classification maps

LDA-MNF

LDA-KMNF
1 Introduction

2 Signal-to-noise ratio transformation

3 Kernel Minimum Noise Fraction

4 Experimental Results

5 Conclusions and Open questions
Conclusions

- Kernel method for nonlinear feature extraction maximizing the SNR
- Good theoretical and practical properties for extracting noise-free features
  - Deals with non-linear relations between the noise and signal
  - The only parameter is the width of the kernel
  - Knowledge about noise can be encoded in the method
- Simple optimization problem → eigendecomposition of the kernel matrix
- Noise estimation in the kernel space with different levels of sophistication
- Simple feature extraction toolbox (SIMFEAT) soon at http://isp.uv.es

Open questions and Future Work

- Pre-images of transformed data in the input space
- Learn kernel parameters in an automatic way
- Test KMNF in more remote sensing applications: denoising, unmixing, ...
Explicit Signal to Noise Ratio in Reproducing Kernel Hilbert Spaces

Luis Gómez-Chova\textsuperscript{1} Allan A. Nielsen\textsuperscript{2} Gustavo Camps-Valls\textsuperscript{1}

\textsuperscript{1}Image Processing Laboratory (IPL), Universitat de València, Spain. luis.gomez-chova@uv.es, http://www.valencia.edu/chovago

\textsuperscript{2}DTU Space - National Space Institute. Technical University of Denmark.

IGARSS 2011 – Vancouver, Canada