Adaptation of multiscale function extension to inexact matching: Application to the mapping of individuals to a learnt manifold.

N. Duchateau\textsuperscript{1}, M. De Craene\textsuperscript{2}, M. Sitges\textsuperscript{1}, V. Caselles\textsuperscript{3}

\textsuperscript{1} Hospital Clínic – IDIBAPS - Universitat de Barcelona (ES)  
\textsuperscript{2} Philips Research, Medisys, Suresnes (FR)  
\textsuperscript{3} Universitat Pompeu Fabra, Barcelona (ES)

SEE-GSI’13 conference – 30/08/2013
How to quantify cardiac (ab)normal motion?

Healthy volunteer

CRT candidate with intra-ventricular dyssynchrony
How to quantify cardiac (ab)normal motion?

Duchateau et al. Med Image Anal (2011)
Then... how to compare dyssynchrony patterns?

Need for a “projection” distance
a) Manifold-learning from training set

\textsc{ISOMAP} = Tenenbaum et al. Science (2000)

**Synthetic dataset**

- Swiss roll ($P=3$)
- 2D embedding ($M=2$)

**CRT dataset**

- $N+1$ images, $P$ pixels
- 2D embedding (e.g.)

\textit{Duchateau et al. Med Image Anal (2012)}
b) Individual comparison to the whole population

```
High dim  \[\rightarrow\]  Low dim  \[\rightarrow\]  High dim
```

Duchateau et al. Med Image Anal (2012)

“projection” = element of the manifold

1) Image = high dim
2) Belongs to the manifold = parameterized by low dim
b) Individual comparison to the whole population

Analogy with:

“pre-image problem”

“denoising auto-encoders”
b) Individual comparison to the whole population

**Problem:**
- Often no explicit formulation of the mappings

**Solution:**
> Formulation of a two-steps interpolation problem + combination of f and g to complete mapping to the manifold:
  - Etyngier et al. NIPS (2007)
  - Gerber et al. Med Image Anal (2010)
Mapping points to the manifold

Combination of $f$ and $g$ to complete mapping to the manifold

$$\hat{\mathbf{I}} = g \circ f(\mathbf{I}).$$

Remark: reconstruction error

$$d_P(\mathbf{I}) = \|\hat{\mathbf{I}} - \mathbf{I}\|$$

Ambient space = 2D (x-y coord.)
Manifold = 1D (parametrization along curve)
Exact / inexact interpolation

\[
\begin{align*}
\arg\min_{f \in \mathcal{F}} \left( \frac{1}{2} \| f \|_{\mathcal{F}}^2 + \frac{\gamma_f}{2} \sum_{i=J+1}^{N} \| f(I_i) - x_i \|^2 \right),
\end{align*}
\]

under the constraint \( \forall j \in [0, J], f(I_j) = x_j \).

Also known as:
- Ridge regression
- Nyström extension
Exact / inexact interpolation

\[
\left\{ \begin{array}{l}
\underset{f \in \mathcal{F}}{\text{argmin}} \left( \frac{1}{2} \| f \|^2_\mathcal{F} + \frac{\gamma_f}{2} \sum_{i=J+1}^{N} \| f(I_i) - x_i \|^2 \right), \\
\text{under the constraint } \forall j \in [0, J], f(I_j) = x_j.
\end{array} \right.
\]

\[
f(I) = \sum_{i=0}^{N} k_{\mathcal{F}}(I, I_i) \cdot c_i \quad \text{with} \quad C = \left( K_{tt} + \frac{1}{\gamma_f} M \right)^{-1} \cdot X,
\]

Also known as:
• Ridge regression
• Nyström extension
Choice of optimal kernel scale

\[
f(I) = \sum_{i=0}^{N} k_{\mathcal{F}}(I, I_i) \cdot c_i
\]

\[
(k_{\mathcal{F}}(I_i, I_j))(i,j) \in [0, N]^2
\]

Example: clustering application (Gaussian kernel)

Blaschko. Machine learning course @Philips (2013)
Choice of optimal kernel scale

\[ f(I) = \sum_{i=0}^{N} k_{\mathcal{F}}(I, I_i) \cdot c_i \]

\[ k_{\mathcal{F}}(I, J) = \exp\left(-\|I - J\|^2 / \sigma_{\mathcal{F}}^2\right) \]

Generally set as:
- Gaussian kernel
- Bandwidth = average distance between neighbors:

\[ \sigma_{\mathcal{F}} = \frac{1}{N+1} \sum_{i=0}^{N} \|I_i - \text{nn}_k(I_i)\|^2 \]

Problem if non-uniform sampling !!!
First solution: locally adapted kernel bandwidth

Adapt the bandwidth of the kernel to the local neighborhood

\[
\sigma_F(I) = \frac{1}{K^2} \sum_{k=1}^{K} \sum_{l=1 \atop l \neq k}^{K} \|nn_k(I) - nn_l(I)\|^2
\]

Duchateau et al. Med Image Anal (2012)
Second solution: multiscale extension

Framework of “diffusion maps” to reach density invariance:

- For the manifold learning = diffusion maps

- For the out-of-sample extension = geometric harmonics
  
  = Coifman and Lafon, ACHA (2006)
Second solution: multiscale extension

Framework of “diffusion maps” to reach density invariance

In the manifold learning process:
Diffusion distance / diffusion maps

For the out-of-sample extension:
Geometric harmonics = Coifman and Lafon, ACHA (2006)

Definition 2. When $\lambda_j \neq 0$, the eigenfunction $\psi_j$ can be extended to $\tilde{x} \in \tilde{X}$ by

$$
\Psi_j(\tilde{x}) = \frac{1}{\lambda_j} \int_{X} k(\tilde{x}, y) \psi_j(y) \, d\mu(y).
$$

This extension is called geometric harmonic.

In our case = dyadic scales
Second solution: multiscale extension

Geometric harmonics
= Coifman and Lafon, ACHA (2006)

Bermanis et al. ACHA (2006)
Algorithm 4 Multiscale data sampling and function extension.

**Input:** A dataset $D = \{x_1, \ldots, x_n\}$ in $\mathbb{R}^d$, a positive number $T > 0$, a new data point $x_* \in \mathbb{R}^d \setminus D$, a function $f = [f_1 f_2 \ldots f_n]^T$ to be extended and an error parameter $err \geq 0$.

**Output:** An approximation $F = [F_1 F_2 \ldots F_n]^T$ of $f$ on $D$ and its extension $F_*$ to $x_*$. 

1: Set the scale parameter $s = 0$, $F^{(-1)} = 0 \in \mathbb{R}^n$ and $F_*^{(-1)} = 0$.

2: **while** $\| f - F^{(s-1)} \| > err$ **do**

3: Form the Gaussian kernel $G^{(s)}$ on $D$ (see Eqs. (1.11) and (4.3)), with $\epsilon_s = \frac{T}{2^s}$.

4: Estimate the numerical rank $k^{(s)}$ of $G^{(s)}$ using Eq. (3.20).

5: Apply Algorithm 2 to $G^{(s)}$ with the parameters $f^{(s)}$ and $l^{(s)}$ to get an $n \times k^{(s)}$ matrix $B^{(s)}$ and the sampled dataset $D_s$.

6: Apply Algorithm 3 to $B^{(s)}$ and $f$. We get the approximation $f^{(s)}$ to $f - F^{(s-1)}$ at scale $s$, and its extension $f_*^{(s)}$ to $x_*$. 

7: Set $F^{(s)} = F^{(s-1)} + f^{(s)}$, $f_*^{(s)} = F_*^{(s-1)} + f_*^{(s)}$, $s = s + 1$.

8: **end while**

9: $F = F^{(s-1)}$ and $F_* = F_*^{(s-1)}$.

---

**Geometric harmonics = Coifman and Lafon, ACHA (2006)**

**Taken from:** Bermanis et al., ACHA (2006)
Analogy with wavelets

Contribution: multiscale + inexact matching

Regularization weight AND until which scale to go

\[ k_{\mathcal{F}}(I, J) = \exp\left(-\frac{||I - J||^2}{\sigma_{\mathcal{F}}^2}\right) \]

\[
\left\{ \begin{array}{l}
\arg\min_{f \in \mathcal{F}} \left( \frac{1}{2} \|f\|_{\mathcal{F}}^2 + \frac{\gamma_f}{2} \sum_{i=J+1}^{N} \|f(I_i) - x_i\|^2 \right), \\
\text{under the constraint } \forall j \in [0, J], f(I_j) = x_j.
\end{array} \right.
\]
Contribution: multiscale + inexact matching

Regularization weight AND until which scale to go

Until the signal resolution

AND

Until the noise level
Contribution: multiscale + inexact matching

Regularization weight AND until which scale to go

Simple 1D-to-1D example (x-to-y)
Contribution: multiscale + inexact matching

Regularization weight AND until which scale to go

More complex:
Ambient space = 2D (x-y coord.)
Manifold = 1D (parametrization along curve)
Training = set of patients with Septal Flash (SF)
Testing = new subjects (healthy or not)
Back to the application: CRT candidates

- Now “projection” distance available = distance to the pattern
- Complemented by one along the manifold = distance to normality
Back to the application: CRT candidates

Varying bandwidth

Back to the application: CRT candidates

Same 2D-to-1D
REAL DATA

Varying bandwidth

Multiscale extension
Present work
Conclusions

- Modeling a pathological pattern as deviation from normality
- Non-linear analysis
- Comparison of motion patterns in their whole (and not just peak or time-to-event measurements)

**Contributions:**
1. Interpolation problem now well-solved
2. Combination of multiscale + inexact matching (and associated practical questions about parameters tuning)
3. A-posteriori check of the varying bandwidth kernel low bias
Acknowledgements


Previous collaborators:
- Universitat Pompeu Fabra, Barcelona (ES) G Piella, BH Bijnens
- University of Sheffield (UK) AF Frangi
- Hospital Clínic, Barcelona (ES) A Doltra, E Silva, MA Castel, L Mont, J Brugada

Projects:
- MICINN (MTM2012-3077)
- GRC (2009-SGR-773)
- GenCat (ICREA Academia)
Thanks !!! Any questions...?
Back to the application: CRT candidates

Evolution of abnormalities?

CRT #9
Septal flash

CRT #8
Septal flash

CRT #12
Left-right interaction

OFF

Follow-up
Distance analysis: evolution with CRT

- Tending to a relevant clustering at baseline
- With pacing ON and follow-up (all CRT candidates, **not only SF**)
  - Getting further from SF-like pattern
  - Getting closer to normality

Duchateau et al. FIMH (2013)
With pacing ON and follow-up (only SF vs CRT response)
- Disparition of SF-like pattern
- Getting closer to normality

At follow-up:
- Recovery of a “normal” motion = responders
- Others: may be responders / non-responders (probably additional factors)