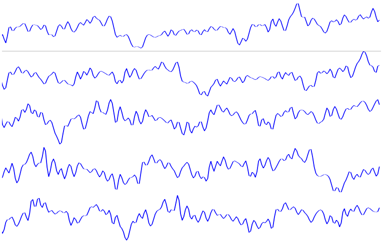
FuncICA for time series pattern discovery

Nishant Mehta and Alexander Gray

Georgia Institute of Technology

The problem

Given a set of inherently continuous time series (e.g. EEG)

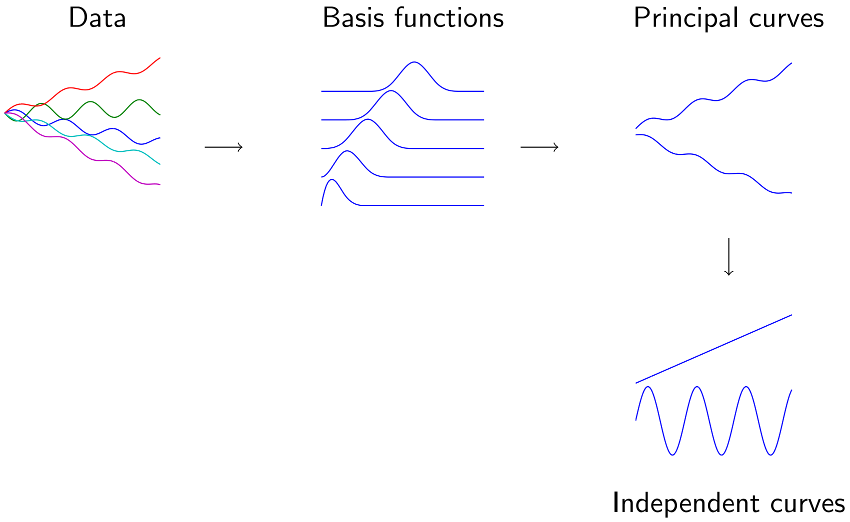


Find a set of patterns that vary independently over the data

Existing solutions:

1. Standard independent component analysis (ICA)
2. Functional principal component analysis

High level


Data
Basis functions
Principal curves
Independent curves

High level

FuncICA - Functional independent component analysis

- ICA for time series data

1. Express the data using a set of basis functions.
2. Functional PCA - Find the k curves of maximum variation across the data, subject to smoothness constraint
3. Rotate functional principal components to maximize independence and yield independent components Y
4. Optimize an independence objective $Q(Y)$ with respect to a smoothness regularization term

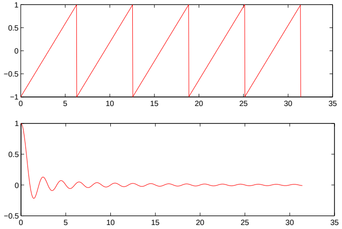
Towards automating science

- ICA used in pattern discovery in natural domains like neuroscience, genetics, and astrophysics, but patterns are often smooth
- Many domains created by humans, including financial markets and chemical processes, involve smooth variation
- FuncICA offers a way to find optimally smoothed patterns in these domains.
    - Automatic identification of event-related potentials in EEG
    - Automatic discovery of gene signaling mechanisms from microarray gene expression data
    - Identifying spatiotemporal activation patterns in fMRI

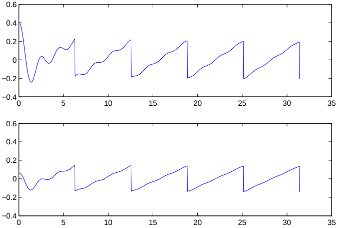
Independent component analysis

- Observe multiple signals $X^{(t)}$ over time
- Univariate signal $X_i^{(t)}$ is mixture of independent sources $S^{(t)}$
- Linear instantaneous mixing model: $X^{(t)} = AS^{(t)}$
- Find unmixing transformation that maximizes statistical independence of recovered sources

Sources



Mixtures



Independent component analysis

- Find unmixing matrix W such that $Y = WX = WAS$
- $Y \sim S$ up to a scaling and permutation of sources

Maximizing independence

$\Updownarrow$

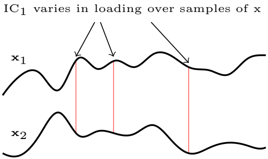
Minimizing difference between joint $\mathrm{P}[Y]$ and product of marginals $\prod \mathrm{P}[Y_i]$

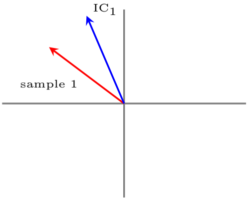
- Use the Kullback-Leibler Divergence:

$$\mathcal{H}(Y) = D_{\mathrm{KL}}(\mathrm{P}[Y] \| \prod_{i=1}^{n} \mathrm{P}[Y_i]) = \sum_{i=1}^{n} H(Y_i) - H(Y)$$

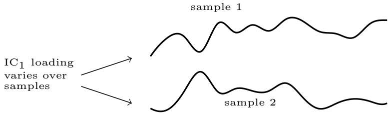
ICA duality

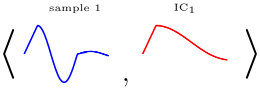
Primal





Dual





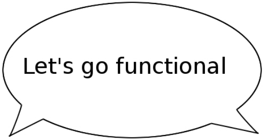
Why functional data?

- Higher sampling rate $\Rightarrow$ Higher dimensional data
- No principled way of dimensionality reduction
  - Subsampling?
- No principled way to handle asynchronous observations
  - Missing data from occasionally offline sensors
  - Each observation lives in different space
- Alternative?

  Generative (parametric) models - HMM, dynamic Bayesian network

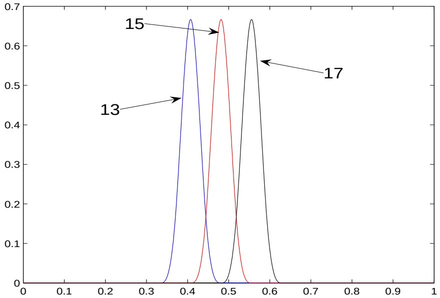
  OR

  Functional representation



Functional data

- Set of n curves $X = \{X_1(t), \ldots, X_n(t)\}, X_i \in \mathcal{X}$
- Set of m basis functions $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_m\}$
- Decompose data as $X_i(t) = \sum_{j=1}^{m} \psi_{i,j}\beta_j(t)$
- $\mathcal{X} \subset L^2$ (Hilbert space), with inner product $\langle f, g \rangle = \int f(t)g(t)dt$



Functional PCA

Functional PCA to get principal curves

$$E_i(t) = \sum_{j=1}^{m} \rho_{i,j}\beta_j(t)$$

Principal components (curve loadings) over the data

$$\sigma_{i,j}^{(E)} = \langle E_i, X_j \rangle = \sum_{k=1}^{m}\sum_{l=1}^{m} \rho_{i,k}\langle \beta_k, \beta_l \rangle \psi_{j,l}$$

$$\sigma^{(E)} = \rho B \psi^T$$

Functional ICA

Independent curves are rotation W of principal curves

$$Y_i(t) = \sum_{j=1}^{m} [W\rho]_{i,j} \beta_j(t)$$

Independent components

$$\sigma_{i,j}^{(Y)} = \langle Y_i, X_j \rangle = \sum_{k=1}^{m} \sum_{l=1}^{m} [W\rho]_{i,k} \langle \beta_k, \beta_l \rangle \psi_{j,l}$$

$$\sigma^{(Y)} = W\sigma^{(E)}$$

Now, just solve for W to find IC basis weights $\phi = W\rho$

Independence objective

KL-divergence objective

$$\mathcal{H}(Y) = \sum_{i=1}^{n} H(Y_i) - H(Y)$$

After FPCA, we have

$$\mathcal{H}(E) = \sum_{i=1}^{n} H(E_i) - H(E)$$

$Y = WE$ and W orthogonal yield minimum marginal entropy objective

$$\mathcal{H}^*(Y) = \sum_{i=1}^{n} H(Y_i)$$

Entropy estimator to evaluate $H(Y_i)$

Vasicek entropy estimator

- nonparametric entropy estimator that considers order statistics

1. Order samples in non-decreasing order $Z^{(1)} \leq Z^{(2)} \ldots \leq Z^{(N)}$
2. m-spacing is $Z^{(i+m)} - Z^{(i)}$
3. $\hat{H}_N(Z^1, Z^2, \ldots, Z^N) =$

$$\frac{1}{N} \sum_{i=1}^{N-m_N} \log \left(\frac{N}{m_N} (Z^{(i+m_N)} - Z^{(i)} \right)$$

Optimizer

Plug-in ICA estimator: **RADICAL ICA**
[Learned-Miller and Fisher 2003]

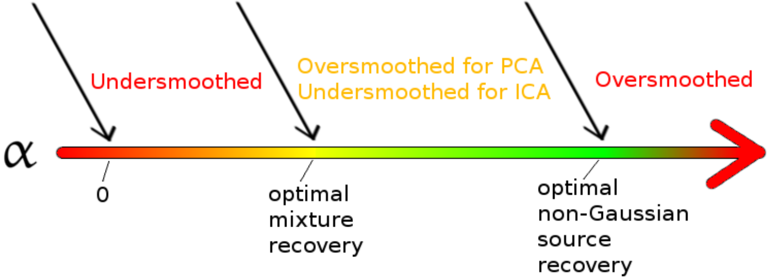
- RADICAL uses Vasicek entropy estimator
- For ICA in D dimensions (D eigenfunctions), do pairwise separation.
- 2-D rotation matrix is parameterized by one angle parameter θ:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Since $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, use brute-force to optimize $\hat{H}_N(\theta)$

Smoothing

- FPCA chooses smoothing level α_p^* minimizing leave-one-out cross-validation error
- α_p^* not optimal for source recovery
- FuncICA balances reconstruction error while avoiding Gaussian components



L_2 Smoothing (FPCA)

[Ramsay and Silverman 2002]
Penalize the second derivative so that functions are not "too wiggly"

$$\int \xi(t)^2 dt + \alpha \int (D^2 \xi(t))^2 dt = 1,$$

for $\alpha \geq 0$

- Smooths the principal curves directly
- Hence also smooths independent curves

Select optimal reconstruction error α_p^* via leave-one-out cross-validation

Inverse negentropy smoothing (FuncICA)

Motivation - Penalize Gaussian components
Why?

- ICA fails if there is more than 1 Gaussian component
- Oversmoothing $\Rightarrow$ components become noisy $\Rightarrow$ non-Gaussian components may become Gaussian

How? Inverse negentropy objective function:

$$Q(Y) = \sum_{i=1}^{p} \frac{1}{J(Y_i)}$$

where $J(Y_i) = H(\mathcal{N}(0,1)) - H(Y_i)$ is the negentropy of unit-variance Y_i

Optimal FuncICA inverse negentropy smoothing

Algorithm

1. $\alpha = \alpha_p^*,\ \tau = 0,\ Q^{(0)} = \infty$
2. **repeat**
3. $\quad \tau = \tau + 1$
4. $\quad (Y, Q^{(\tau)}) = \text{FuncICA}(X, \alpha^{(\tau)})$
5. $\quad \alpha^{(\tau+1)} = \gamma \alpha^{(\tau)}$
6. **until** $Q^{(\tau)} > Q^{(\tau-1)}$
7. **return** $\alpha^{(\tau-1)}$

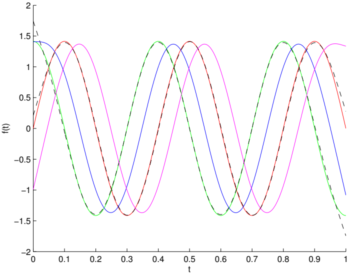
Intuition:

- α_p^* optimally smooths for reconstruction error
- Can further smoothing be beneficial for source recovery?

  Yes! Can effectively dampen Gaussian noise components

Synthetic data results

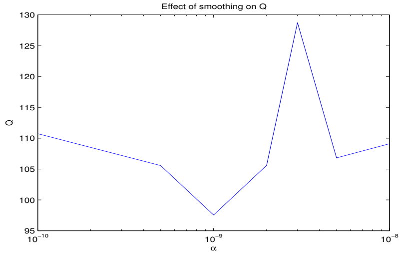
- Perfect source recovery for mixture of Laplace-distributed harmonics

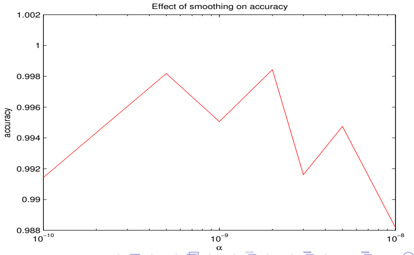


- Successful isolation of single high-frequency Gaussian source
  - FuncICA performs well for $\alpha \geq 0$
  - FPCA blends high-frequency source into all recovered curves
- Dampening of two high-frequency Gaussian sources
  - Q statistic performs well in recovering Laplacian sources

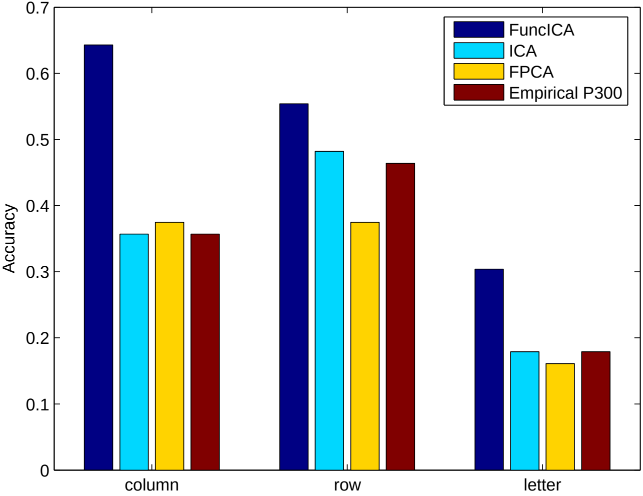
Synthetic data results

Source curves	Distributions
$S_1(t) = \frac{1}{\sqrt{2}}\sin(10\pi t)$	$\leftarrow$ Laplace
$S_2(t) = \frac{1}{\sqrt{2}}\cos(10\pi t)$	$\leftarrow$ Laplace
$S_3(t) = \sin(40\pi t)$	$\leftarrow$ Gaussian
$S_4(t) = \cos(40\pi t)$	$\leftarrow$ Gaussian



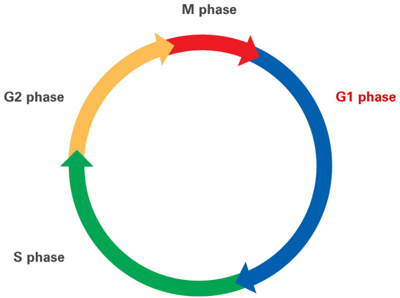


Event-related potential discovery results



Microarray gene expression results

- 6178 genes observed at 18 times in 7 minute increments
- Goal: identify co-regulated genes related to specific phases of the cell cycle
    - G_1 phase regulated vs non-G_1 phase regulated



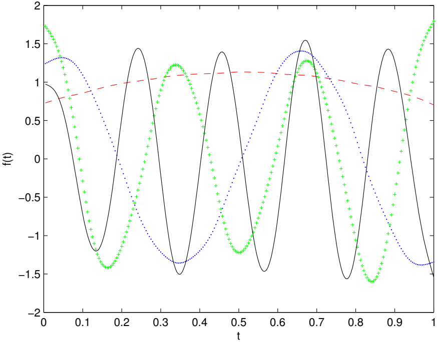
	IC	PC
Top 11	7.1%	9.2%
Filtered	6.2%	8.2%

Conclusion

- Functional ICA offers a way to find smooth, independent modes of variation in time series and other continuous-natured data
- Alternative to FPCA when components of interest may not be Gaussian
- Applicable for EEG, gene expression, finance, and other domains

Questions?

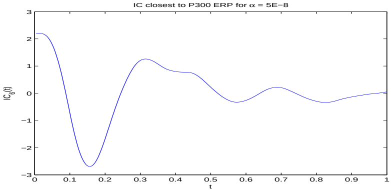
Let's see what FPCA does for our ERP data



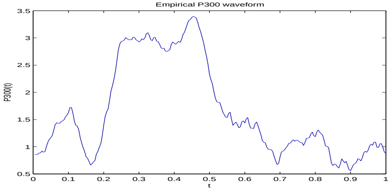
Looks like a Fourier basis
Makes sense

- α rhythm
- β rhythm
- μ rhythm

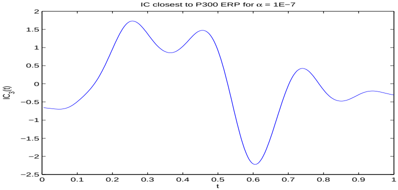
Let's see what Functional ICA (FuncICA) extracts



Closest IC to P300 for $\alpha = 5 \cdot 10^{-5}$

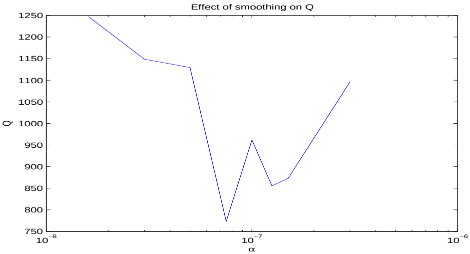


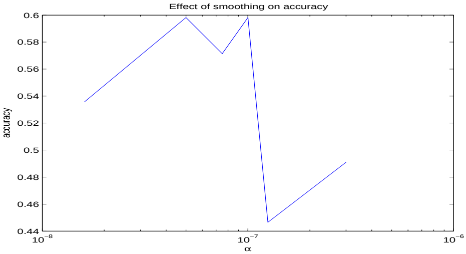
Empirical P300 waveform calculated from 2550 trials



Closest IC to P300 for $\alpha = 1 \cdot 10^{-7}$

Event-related potential discovery results





Smoothing

Choices to make

- Spline functional form?

  cubic b-spline - computationally efficient, common in statistics

- Number of knots?

  As many as we can use tractably

- Number of principal curves to retain?

  Use reconstruction error threshold

  OR

  Largest number that is tractable for FuncICA