POLITECNICO DI MILANO

MODELLISTICA E CALCOLO SCIENTIFICO

MOX

MODELING AND SCIENTIFIC COMPUTING

Distance and inference for covariance functions

Davide Pigoli

SuSTaIn Workshop

High dimensional and dependent functional data

Bristol, 11/09/2012

This work is in collaboration with:

Dr. J.A.D. Aston, Department of Statistics, University of Warwick

Prof. I.L. Dryden, School of Mathematical Sciences, University of Nottingham

Prof. P. Secchi, Department of Mathematics, Politecnico di Milano

Practical problem: exploring relationship among Romance languages

Spanish

French

Portuguese

American Spanish

Italian

Practical problem: exploring relationship among Romance languages

Until now, only textual comparisons between words have been considered

Spanish

"uno"

French

"un"

Portuguese

"um"

"uno"

American Spanish

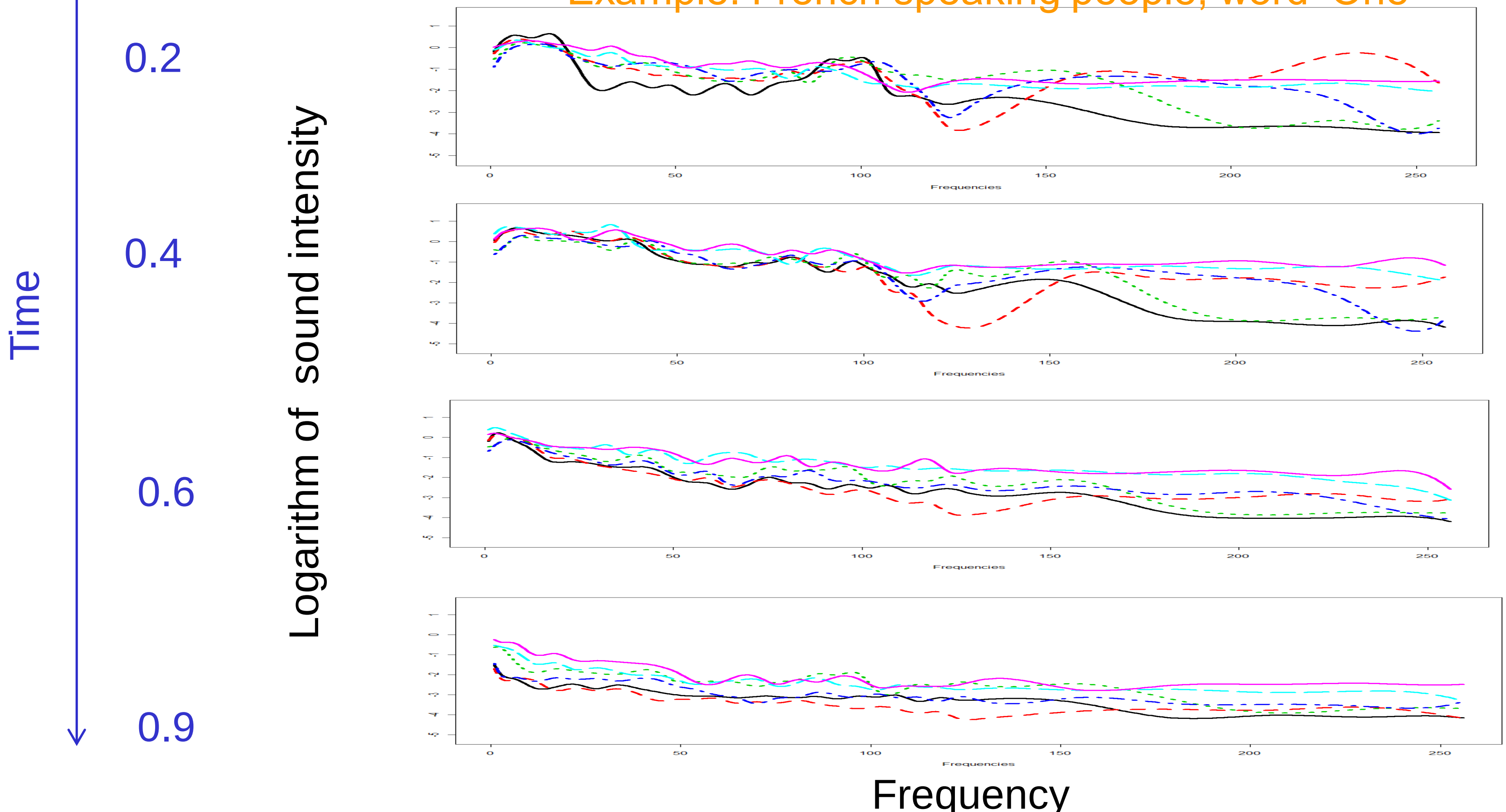
"uno"

Italian

This neglects pronunciation completely !!

Linguistic experts provide the log spectrogram of speech records for a sample of people of different languages… (courtesy of Prof. J. Coleman, Phonetic Laboratory, University of Oxford; preprocessing by P. Hadjipantelis, University of Warwick)

Example: French speaking people, word ‘One’



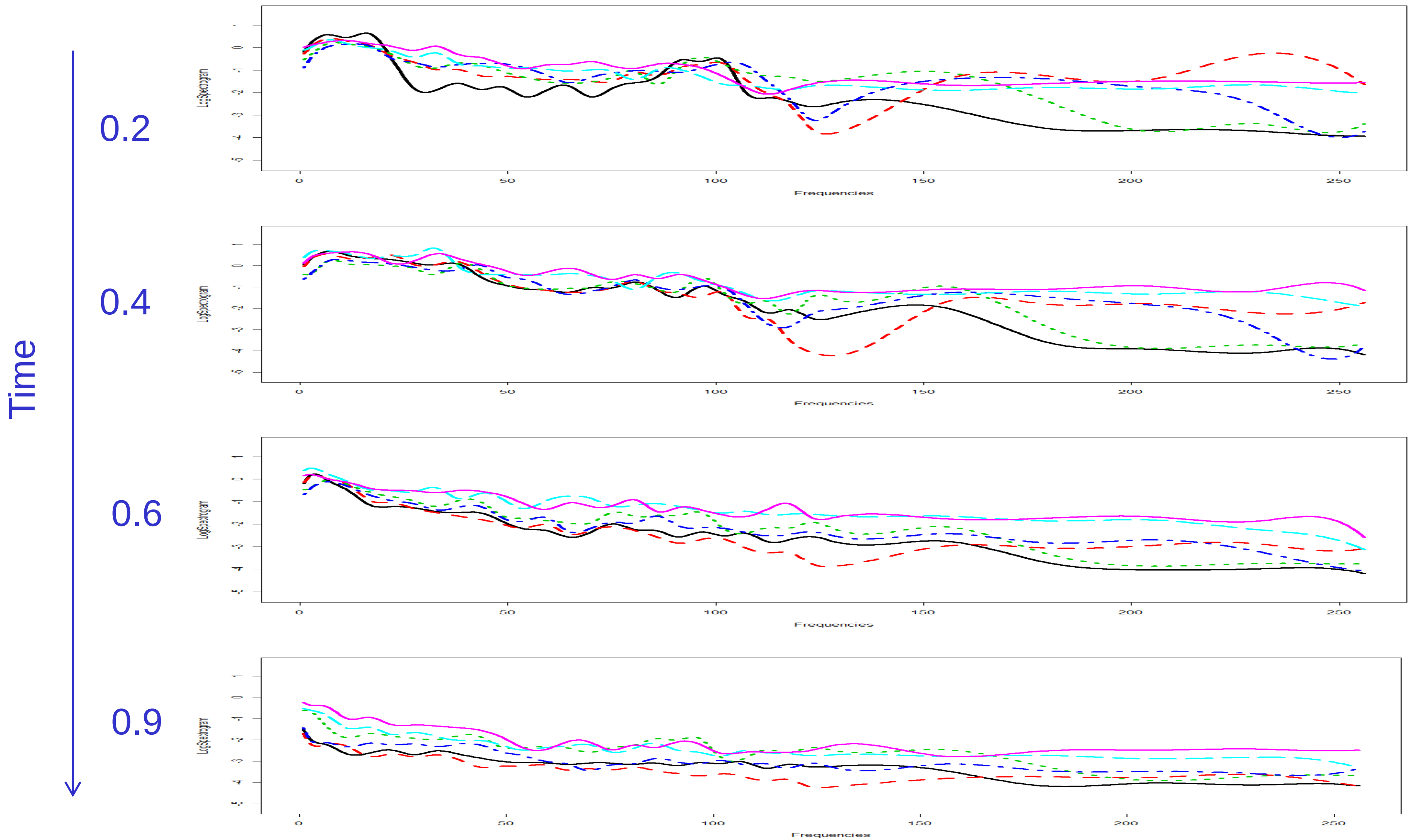
But how dealing with this large amount of information?

- Significant phonetic features of the language are caught by relationships among different frequencies.

- Working hypothesis: existence of a language frequency structure, common to all people.

- Different time instants as a sample from the same covariance operators population.

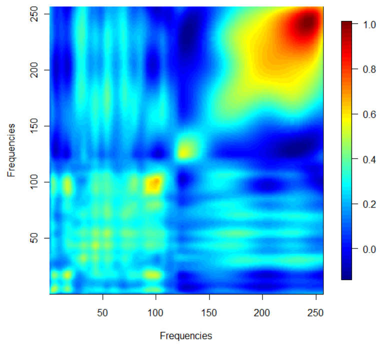
From individual curves to Language frequency covariance



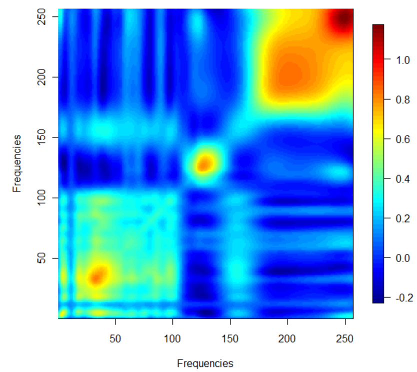
From individual curves to Language frequency covariance

Time

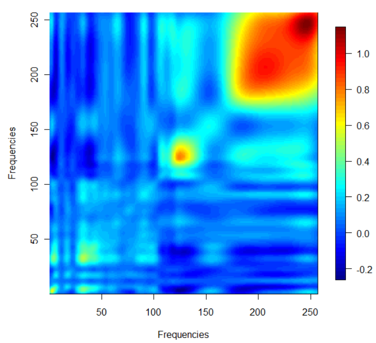
0.2



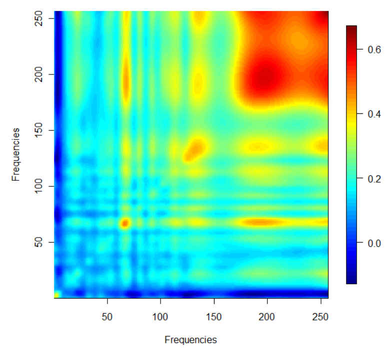
0.4



0.6



0.9



How can we deal with these covariance operators?

The focus is therefore on the covariance operator

$$C(s,t) = \mathrm{Cov}(f(s), f(t))$$

where $f \in L^2(\mathbb{R})$ is a functional random variable.

- How can we measure distance between covariance operators?
- Can we develop inferential techniques based on the chosen distance?

Some properties of covariance operators in Hilbert spaces

Definition *Let B_1 be the closed ball in $L^2(\Omega)$, i.e. it consists in all $f \in L^2(\Omega)$ so that $||f||_{L^2(\Omega)} <= 1$. A bounded linear operator $T : L^2(\Omega) \rightarrow L^2(\Omega)$ is compact if $T(B_1)$ is compact in the norm of $L^2(\Omega)$. A bounded linear operator T is self-adjoint if $T = T^*$*

[see, e.g., Zhu, 2007]

For every compact operator T, a canonical decomposition exists:

$$Tf = \sum_k \sigma_k \langle f, v_k \rangle u_k \quad \Longleftrightarrow \quad Tv_k = \sigma_k u_k$$

Scalar product

Singular values

where $\{u_k\}_k, \{v_k\}_k$ are two orthogonal bases for $L^2(\Omega)$

Some properties of covariance operators in Hilbert spaces

For a self-adjoint compact operator T, an orthogonal basis $\{v_k\}_k$ exists, so that :

$$Tf = \sum_k \lambda_k \langle f, v_k \rangle v_k \quad \Longleftrightarrow \quad T v_k = \lambda_k v_k$$

Eigenvalues

Some properties of covariance operators in Hilbert spaces

For a self-adjoint compact operator T, an orthogonal basis $\{v_k\}_k$ exists, so that :

$$Tf = \sum_k \lambda_k \langle f, v_k \rangle v_k \quad \Longleftrightarrow \quad T v_k = \lambda_k v_k$$

Eigenvalues

A compact operator T is said to be trace class if

$$\text{trace}(T) := \sum_k \langle T e_k, e_k \rangle < +\infty$$

for an orthonormal basis $\{e_k\}$

(trace value does not depend on the choice of the basis)

Some properties of covariance operators in Hilbert spaces

For a self-adjoint compact operator T, an orthogonal basis $\{v_k\}_k$ exists, so that :

$$Tf = \sum_k \lambda_k \langle f, v_k \rangle v_k \quad \Longleftrightarrow \quad T v_k = \lambda_k v_k$$

Eigenvalues

A compact operator T is said to be Hilbert – Schmidt if

$$||T||_{HS}^2 = \mathrm{trace}(T^*T) < +\infty$$

Some properties of covariance operators in Hilbert spaces

For a self-adjoint compact operator T, an orthogonal basis $\{v_k\}_k$ exists, so that :

$$Tf = \sum_k \lambda_k \langle f, v_k \rangle v_k \quad \Longleftrightarrow \quad T v_k = \lambda_k v_k$$

Eigenvalues

A compact operator T is said to be Hilbert – Schmidt if

$$||T||^2_{HS} = \text{trace}(T^*T) < +\infty$$

Covariance operator C(s,t) is a self- adjoint trace class operator on $L^2(\Omega)$,

if $\mathbb{E}[||\mathbf{f}||^2_{L^2(\Omega)}] < +\infty.$

Square root Distance

$$d_R(S_1, S_2) = ||(S_1)^{\frac{1}{2}} - (S_2)^{\frac{1}{2}}||_{HS}$$

where $$(S)^{\frac{1}{2}} f = \sum_k \lambda_k^{\frac{1}{2}} \langle f, v_k \rangle v_k$$

This generalizes to the operatorial case the metric proposed by Dryden et al., 2009, for positive definite matrices.

Procrustes size and shape Distance

It allows unitary transformations between operators

$$d_P(S_1, S_2)^2 = \inf_{R \in SO(L^2(\Omega))} ||L_1 - L_2 R||_{HS}^2 = \inf_{R \in SO(L^2(\Omega))} \operatorname{trace}((L_1 - L_2 R)^*(L_1 - L_2 R))$$

where $S_i = L_i L_i^*$ and $SO(L^2(\Omega))$ is the space of unitary operators R, such that $||Rf||_{L^2(\Omega)} = ||f||_{L^2(\Omega)}$.

Procrustes size and shape Distance

It allows unitary transformations between operators

$$d_P(S_1,S_2)^2 = \inf_{R\in SO(L^2(\Omega))} ||L_1 - L_2 R||_{HS}^2 = \inf_{R\in SO(L^2(\Omega))} \text{trace}((L_1 - L_2 R)^*(L_1 - L_2 R))$$

where $S_i = L_i L_i^*$ and $SO(L^2(\Omega))$ is the space of unitary operators R, such that $||Rf||_{L^2(\Omega)} = ||f||_{L^2(\Omega)}$.

Remark:
$(L_i R)(L_i R)^* = L_i R R^* L_i^* = L_i L_i^*$

Procrustes size and shape Distance

It allows unitary transformations between operators

$$d_P(S_1,S_2)^2 = \inf_{R\in SO(L^2(\Omega))} ||L_1 - L_2 R||^2_{HS} = \inf_{R\in SO(L^2(\Omega))} \text{trace}((L_1 - L_2 R)^*(L_1 - L_2 R))$$

where $S_i = L_i L_i^*$ and $SO(L^2(\Omega))$ is the space of unitary operators R, such that $||Rf||_{L^2(\Omega)} = ||f||_{L^2(\Omega)}$.

> Remark:
> $(L_i R)(L_i R)^* = L_i R R^* L_i^* = L_i L_i^*$

Proposition 1: it is minimized for the unitary operator $\widetilde{R}$ defined by

$$\widetilde{R} v_k = u_k \quad \forall k = 1, \ldots, +\infty$$

where u_k, v_k come from the SVD $L_2^* L_1 v_k = u_k$ for $k = 1, \ldots, +\infty$

Procrustes size and shape Distance

It allows unitary transformations between operators

$$d_P(S_1,S_2)^2 = \inf_{R\in SO(L^2(\Omega))} ||L_1 - L_2 R||_{HS}^2 = \inf_{R\in SO(L^2(\Omega))} \text{trace}((L_1-L_2R)^*(L_1-L_2R))$$

where $S_i = L_i L_i^*$ and $SO(L^2(\Omega))$ is the space of unitary operators R, such that $||Rf||_{L^2(\Omega)} = ||f||_{L^2(\Omega)}$.

Remark:
$(L_iR)(L_iR)^* = L_iRR^*L_i^* = L_iL_i^*$

Proposition 2:

$$d_P(S_1,S_2)^2 = ||L_1||_{HS}^2 + ||L_2||_{HS}^2 - 2\sum_{k=1}^{+\infty}\sigma_k$$

where σ_k are singular values of the operator $L_2^*L_1$

Other distances are not suitable...

For example all distance based on the product

$$S_1^{-1} S_2$$

are not well defined, since covariance operators on $L^2(\Omega)$ are not invertible.

(since $\lambda_k \to 0, \text{ for } k \to +\infty$)

Averaging of covariance operators

The need of averaging among covariance operators arises in many applications. In a finite dimensional setting, a possible estimator is

$$\widehat{\Sigma} = \frac{1}{n_1 + \cdots + n_g}(n_1 S_1 + \cdots + n_g S_g)$$

where $S_1, \ldots, S_g$ are the covariance operators of the g groups with $n_1,\ldots,n_g$ observations each.

This minimizes the sum of square Froebenius distance

$$\widehat{\Sigma} = \arg\min_P \sum_{i=1}^{g} n_i ||S_i - P||_F^2$$

We can therefore define a Frechét average for $S_1, \dots, S_g$ minimizing the appropriate distance:

$$\hat{\Sigma} = \arg\min_S \sum_{i=1}^{g} n_i d(S, S_i)^2$$

It depends on the choice of distance d(.,.)

We can therefore define a Frechét average for $S_1, \dots, S_g$ minimizing the appropriate distance:

$$\widehat{\Sigma} = \arg\min_{S} \sum_{i=1}^{g} n_i d(S, S_i)^2$$

It depends on the choice of distance d(.,.)

Square root Distance: we proved that $\widehat{\Sigma}_S = \left(\frac{1}{G}\sum_{i=1}^{g} n_i S_i^{\frac{1}{2}}\right)^2$

$G = n_1 + \dots + n_g$

Procrustes Distance: the minimum can be found using an iterative procedure, inspired by the algorithm proposed in Gower, 1975, for the matrix case.

Procrustes Averaging:

Initialization: $$\widehat{\Sigma}_P^{(0)} = \widehat{\Sigma}_S$$

Step 1: For each group, compute the unitary operator $\widetilde{R}_i$ that minimizes

$$||L^{(k)} - L_i R_i||_{HS}^2$$

where $\widehat{\Sigma}^{(k)} = L^{(k)} L^{(k)*}$ and $S_i = L_i L_i^*$

Step 2: Compute the average on the transformed operators $\widetilde{L}_i = L_i \widetilde{R}_i$

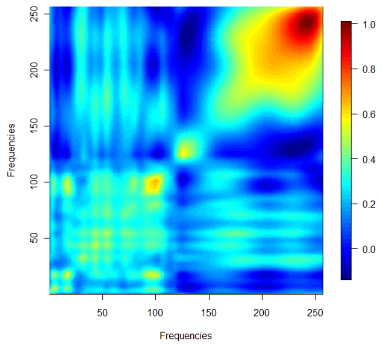
$$L^{(k)} = \frac{1}{G} \sum_i n_i \widetilde{L}_i$$

Step 3: Iterate steps 1-2 until convergence

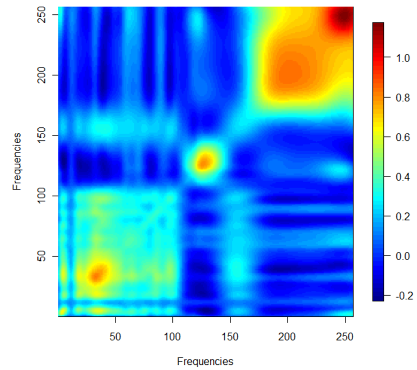
Example: French Language, word 'One'

Time

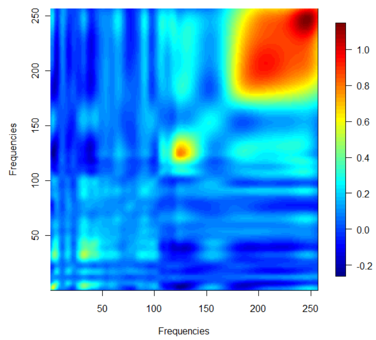
0.2



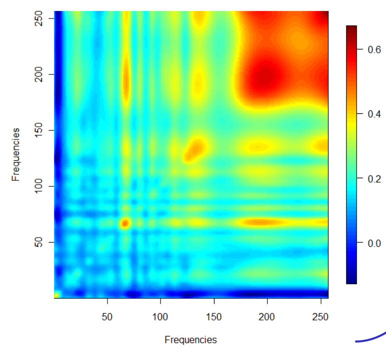
0.4



0.6

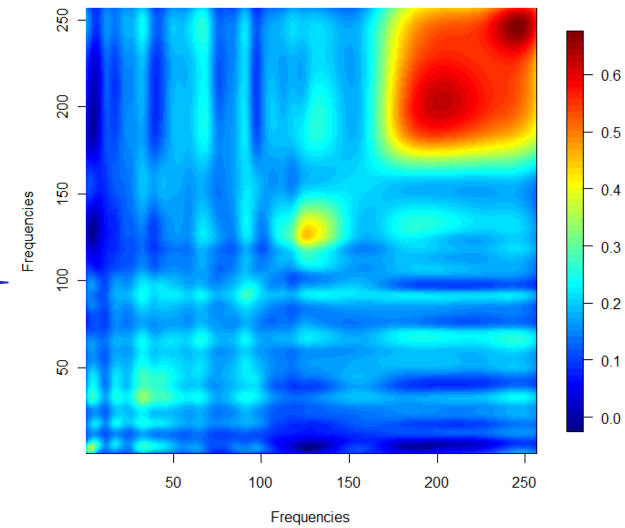


0.9

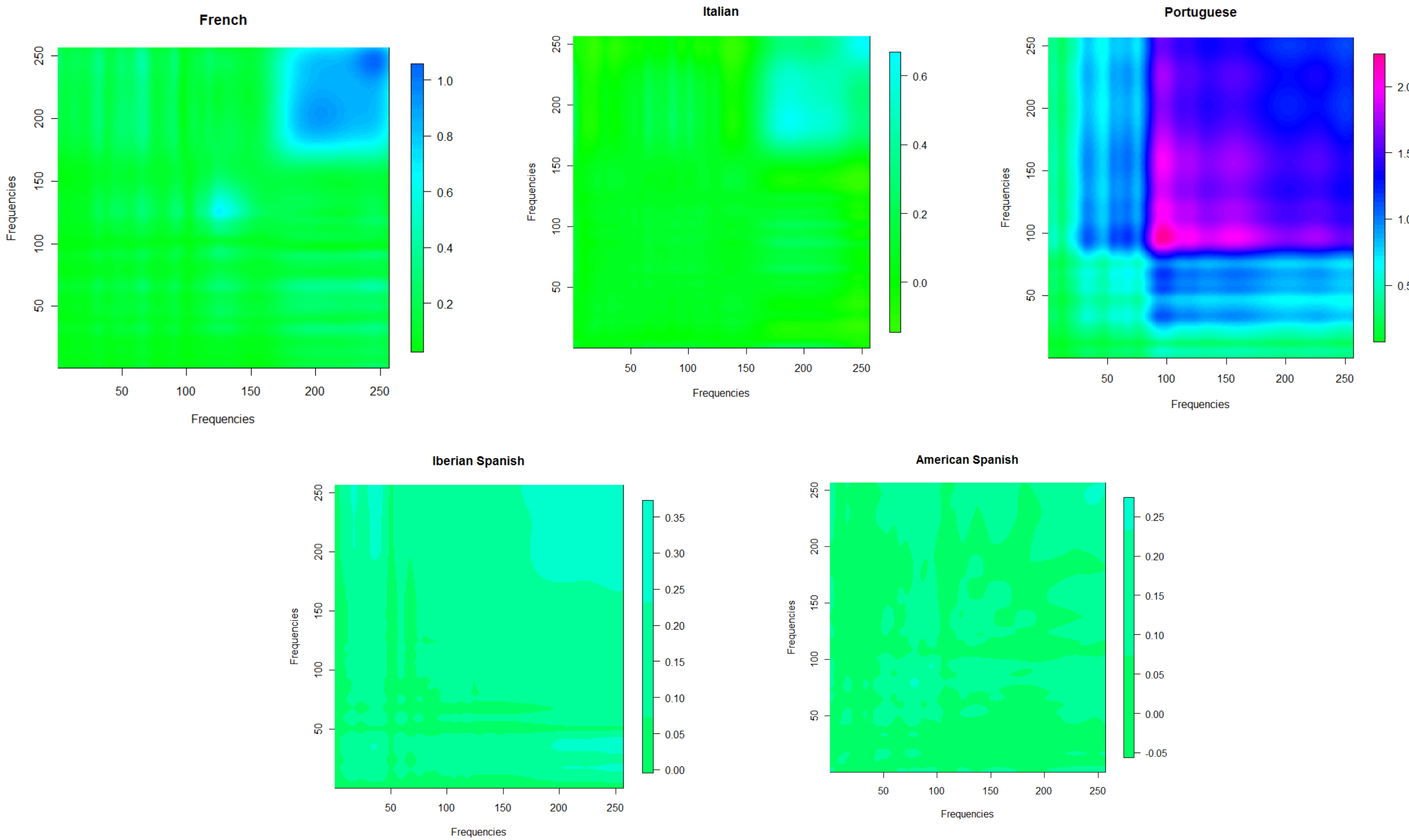


Frechét averaging

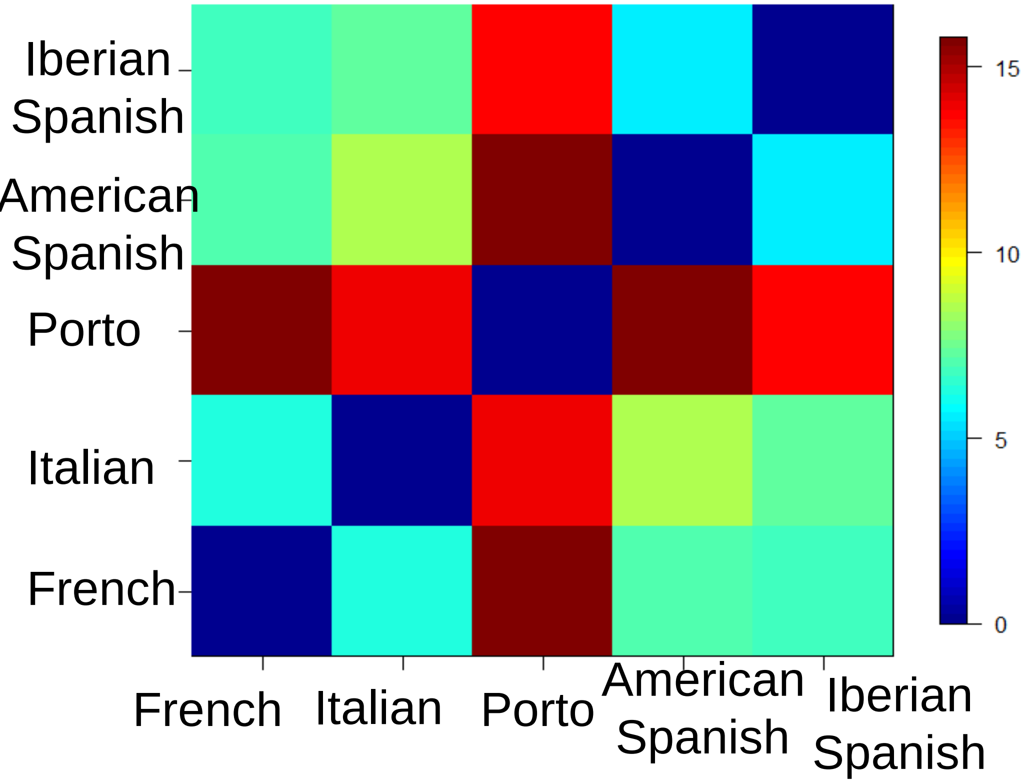
(Square root distance)

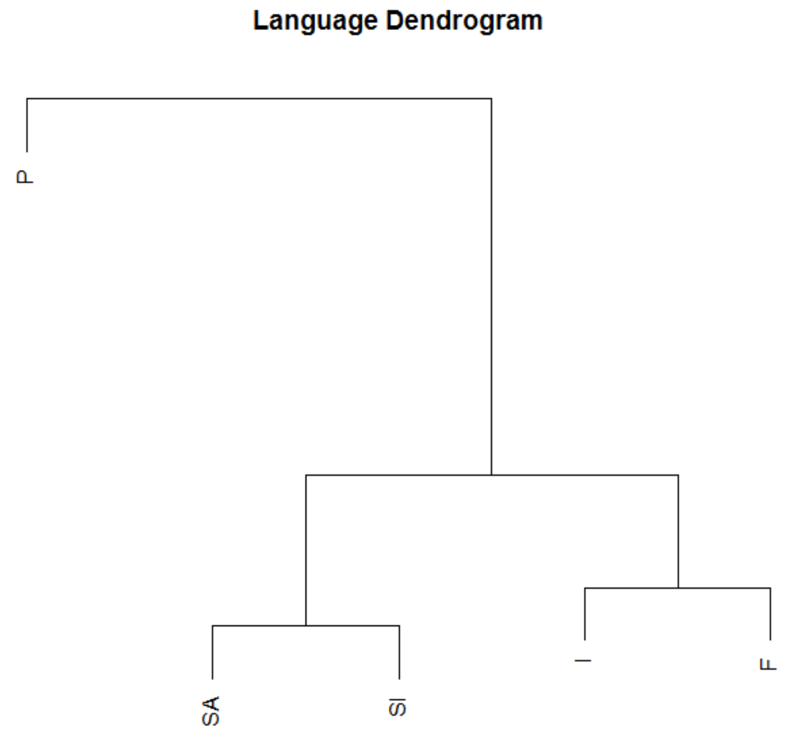


Estimates obtained with Square root distance for word 'One'

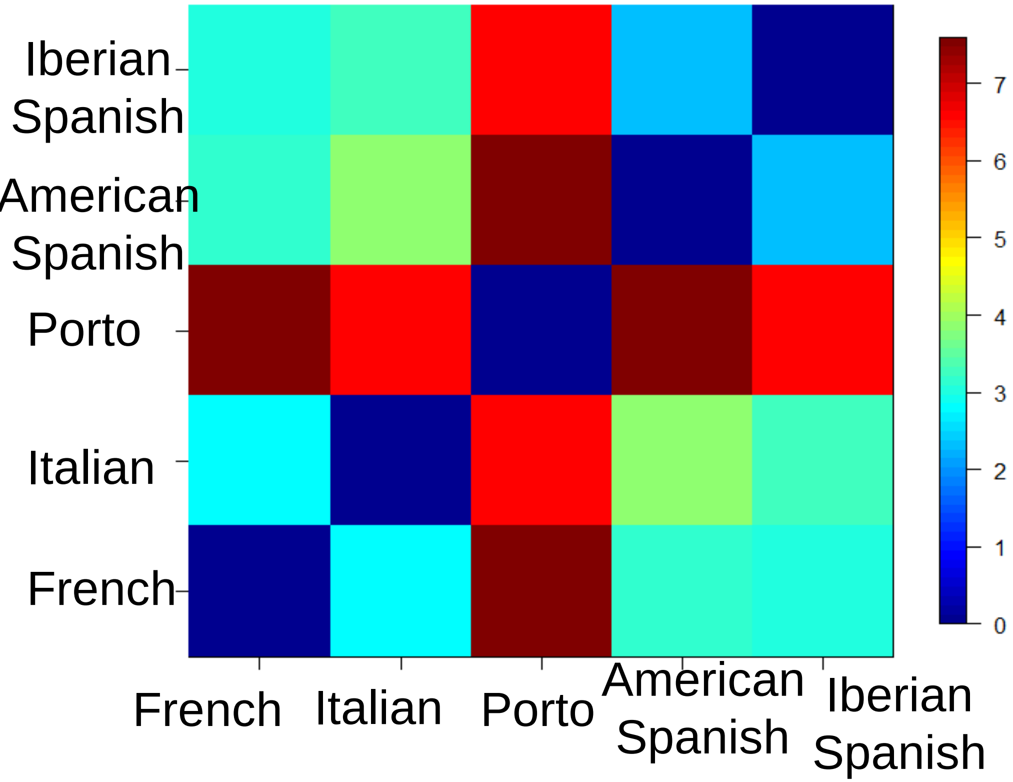


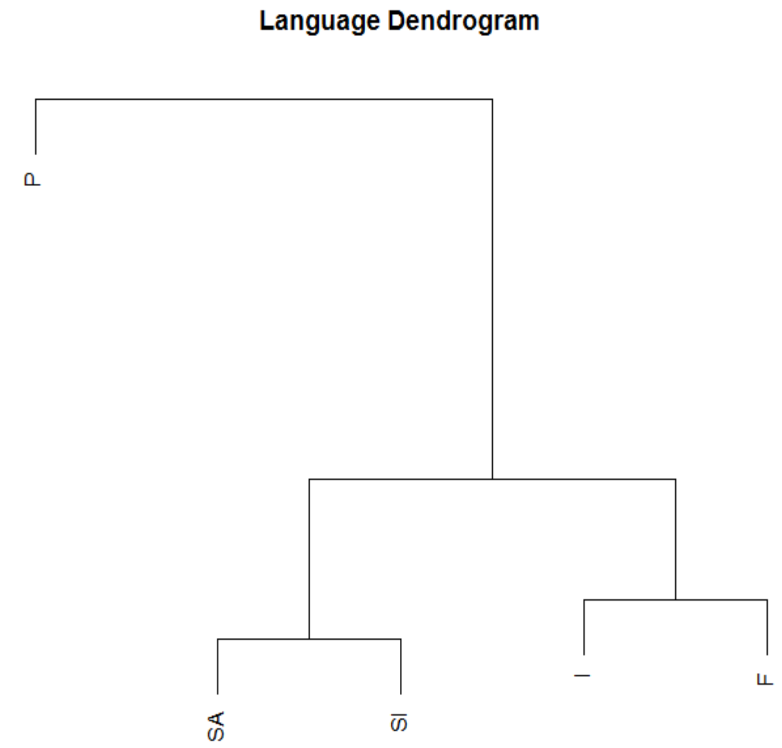
Distance matrix among languages (Square root distance)



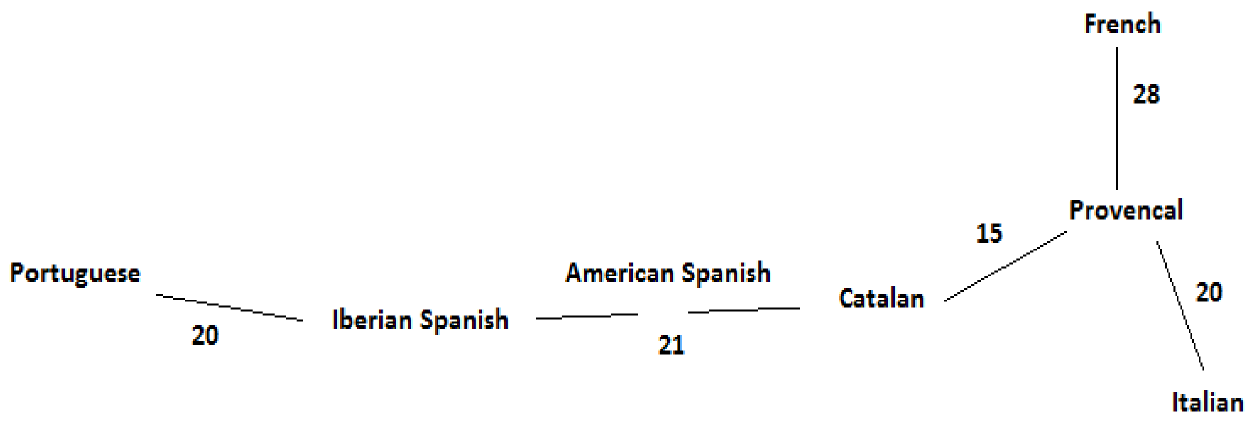


Distance matrix among languages (Procrustes distance)





Map suggested by linguistic knowledge
(Prof. John Coleman, personal communication):



Shortest path connecting covariance operators:

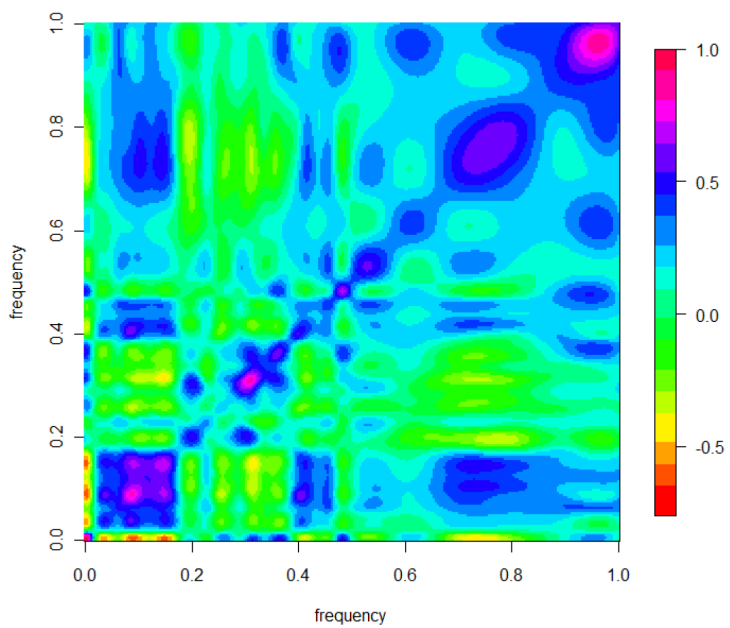
Geodesic extrapolation: Square Root Distance

Geodesic passing through American Spanish and Iberian Spanish covariance operators

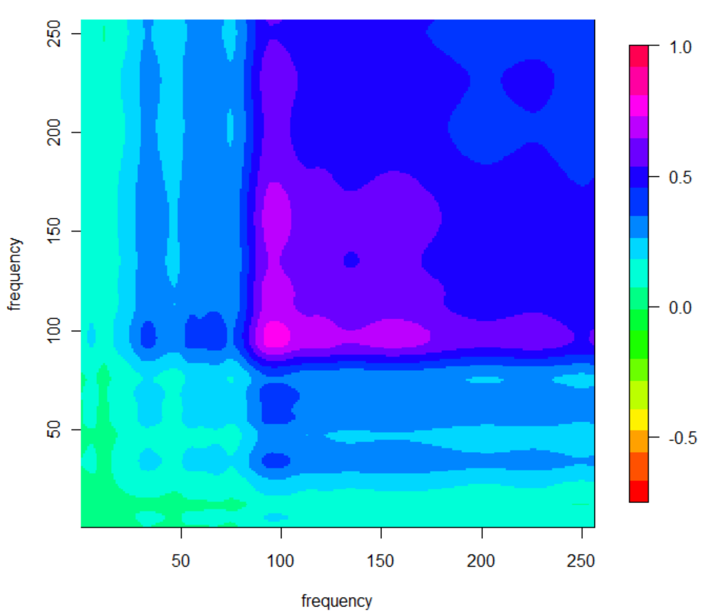
$$S(x) = \frac{1}{x}\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}})\}^*\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}})\}$$

Extrapolated covariance operator for

$$x = ||(S_P)^{\frac{1}{2}} - (S_{SA})^{\frac{1}{2}}||_{HS}$$



Portuguese Covariance operator



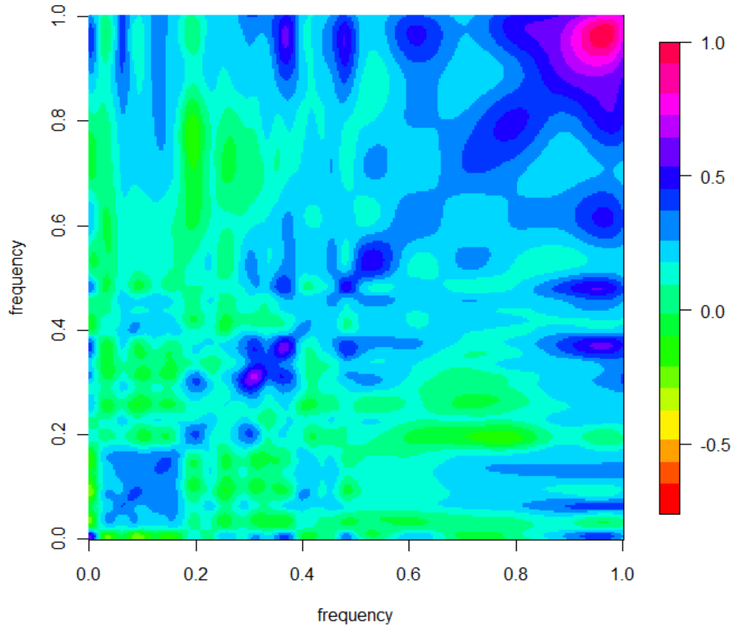
Geodesic extrapolation: Procrustes Distance

Geodesic passing through American Spanish and Iberian Spanish covariance operators

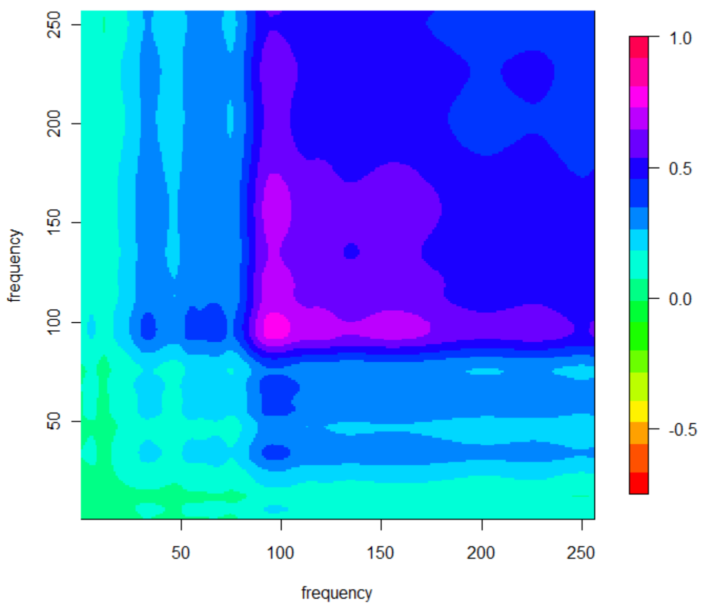
$$S(x) = \frac{1}{x}\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}}\tilde{R})\}\{(S_{SA})^{\frac{1}{2}} + x((S_{SA})^{\frac{1}{2}} - (S_{SI})^{\frac{1}{2}}\tilde{R})\}^*$$

Extrapolated covariance operator for

$$x = d_P(S_P, S_{SA})$$



Portuguese Covariance operator



Conclusions

- Novel distances for the comparison of infinite dimensional covariance operators have been illustrated.

- Estimators for the average covariance operator have been proposed.

- Covariance operator among frequency catches significant feature in human languages.

- Phonetic structure highlights that Portuguese language behaves differently from other languages.

Future perspectives:

- Hypothesis testing $\Longrightarrow$ distance–based permutation procedure
- Multi-words analysis

References:

- Dryden, I.L., Koloydenko, A., Zhou, D. (2009), Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging, The Annals of Applied Statistics, 3, 1102-1123.
- Gower, J. C. (1975), Generalized Procrustes analysis , Psychometrika, 40,33-50.
- Hadjipantelis, P. Z., Aston, J. A. D., and Evans, J. P. (2012), Characterizing fundamental frequency in Mandarin: A functional principal component approach utilizing mixed effect models, Journal of the Acoustical Society of America, in press.
- Pigoli, D., Aston, J.A.D., Dryden, I.L., Secchi, P. (2012) Distances and Inference for Covariance Functions, Technical Report Mox 35/2012.
- Zhu, K. (2007), Operator theory in function spaces (2nd ed.), American Mathematical Society.