VNIVERSITAT DE VALÈNCIA
Facultat de Ciències Matemàtiques
Departament d'Estadística i Investigació Operativa

Geoestadística en regiones heterogéneas con distancia basada en el coste

TESIS DOCTORAL

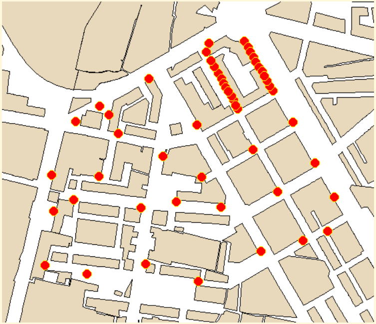
Facundo Martín Muñoz Viera

Director: Antonio López-Quílez

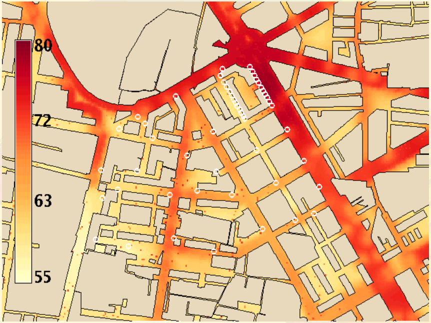
Febrero 2013

Outline

Motivation: acoustic maps and heterogeneous regions



observations



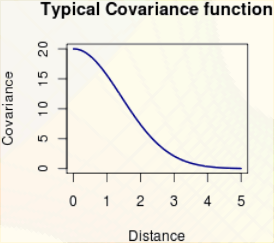
prediction

- Assessment of the uncertainty!

Covariance functions

$C(h) \overset{h=d(\boldsymbol{s}_1,\boldsymbol{s}_2)}{\overset{\downarrow}{=}} \mathbb{C}\left[Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)\right]$ represents the relationship between the *proximity* and the statistical correlation.

We restrict to isotropic functions.

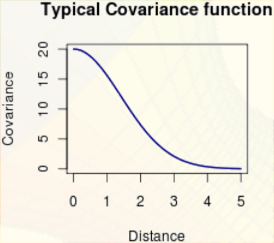


Valid covariance functions

Not all functions are permissible as covariance functions

Covariance functions

$C(h) \overset{\overset{h=d(\boldsymbol{s}_1,\boldsymbol{s}_2)}{\downarrow}}{=} \mathbb{C}\left[Z(\boldsymbol{s}_1), Z(\boldsymbol{s}_2)\right]$ represents the relationship between the *proximity* and the statistical correlation.
We restrict to isotropic functions.



Valid covariance functions

Not all functions are permissible as covariance functions

Positive-definite functions

Positivity condition

The covariance function must be positive-definite

$$\forall \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}, \quad \forall a_1, \ldots, a_n, \quad \sum_i \sum_j a_i a_j C(h_{ij}) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} \geq 0$$

In the Euclidean space $E_d = (\mathbb{R}^d, \cdot)$, the family of positive-definite functions is fully characterized by Schoenberg's (1938) theorem:

$$C(h) = \int_0^\infty \Omega_{\frac{d-2}{2}}(h\lambda)\, dG(\lambda),$$

where $\Omega_m(x) = \Gamma(m+1)(\frac{2}{x})^m J_m(x)$, J_m is the Bessel function of the first kind of order m, and G is a nondecreasing bounded measure on $[0, \infty)$.

Positive-definite functions

Positivity condition

The covariance function must be positive-definite

$$\forall\{\boldsymbol{s}_1,\ldots,\boldsymbol{s}_n\},\quad \forall a_1,\ldots,a_n,\quad \sum_i\sum_j a_i a_j C(h_{ij}) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} \geq 0$$

In the Euclidean space $E_d = (\mathbb{R}^d, \cdot)$, the family of positive-definite functions is fully characterized by Schoenberg's (1938) theorem:

$$C(h) = \int_0^\infty \Omega_{\frac{d-2}{2}}(h\lambda)\, dG(\lambda),$$

where $\Omega_m(x) = \Gamma(m+1)(\frac{2}{x})^m J_m(x)$, J_m is the Bessel function of the first kind of order m, and G is a nondecreasing bounded measure on $[0,\infty)$.

Bochner's Theorem (1933)

Characterizes the positive-definite (non-isotropic) functions as characteristic functions (a kind of Fourier Transform) of distribution functions in E_d.

$$\tilde{C}(\boldsymbol{h}) = \mathbb{E}\left[e^{i\boldsymbol{h}'\boldsymbol{X}}\right] = \int_{E_d} e^{i\boldsymbol{h}'\boldsymbol{x}} dF_X(\boldsymbol{x}), \quad \boldsymbol{h}, \boldsymbol{X} \in E_d \tag{1}$$

Sufficiency:

$$\begin{aligned}\sum_{i,j} a_i a_j \tilde{C}(\boldsymbol{s}_i - \boldsymbol{s}_j) &= \mathbb{E}\left[\sum_{i,j} a_i a_j e^{i(\boldsymbol{s}_i - \boldsymbol{s}_j)'\boldsymbol{x}}\right] \\ &= \mathbb{E}\left[\Big(\sum_i a_i e^{i\boldsymbol{s}_i'\boldsymbol{x}}\Big)\overline{\Big(\sum_j a_j e^{i\boldsymbol{s}_j'\boldsymbol{x}}\Big)}\right] \\ &= \mathbb{E}\left[\Big|\Big(a_i \sum_i e^{i\boldsymbol{s}_i'\boldsymbol{x}}\Big)\Big|^2\right] \geq 0.\end{aligned} \tag{2}$$

Bochner's Theorem (1933)

Characterizes the positive-definite (non-isotropic) functions as characteristic functions (a kind of Fourier Transform) of distribution functions in E_d.

$$\tilde{C}(\boldsymbol{h}) = \mathbb{E}\left[e^{i\boldsymbol{h}'\boldsymbol{X}}\right] = \int_{E_d} e^{i\boldsymbol{h}'\boldsymbol{x}} dF_X(\boldsymbol{x}), \quad \boldsymbol{h}, \boldsymbol{X} \in E_d \tag{1}$$

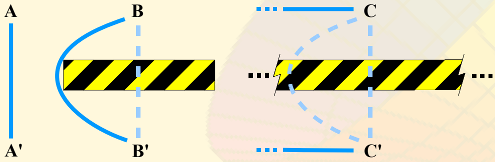
Sufficiency:

$$\begin{aligned}
\sum_{i,j} a_i a_j \tilde{C}(\boldsymbol{s}_i - \boldsymbol{s}_j) &= \mathbb{E}\left[\sum_{i,j} a_i a_j e^{i(\boldsymbol{s}_i - \boldsymbol{s}_j)'\boldsymbol{x}}\right] \\
&= \mathbb{E}\left[\Big(\sum_i a_i e^{i\boldsymbol{s}_i'\boldsymbol{x}}\Big)\overline{\Big(\sum_j a_j e^{i\boldsymbol{s}_j'\boldsymbol{x}}\Big)}\right] \\
&= \mathbb{E}\left[\Big|\Big(a_i \sum_i e^{i\boldsymbol{s}_i'\boldsymbol{x}}\Big)\Big|^2\right] \geq 0.
\end{aligned} \tag{2}$$

Heterogeneous regions

In the presence of barriers, the correlation is not directly associated with the Euclidean distance.

Heterogeneous regions

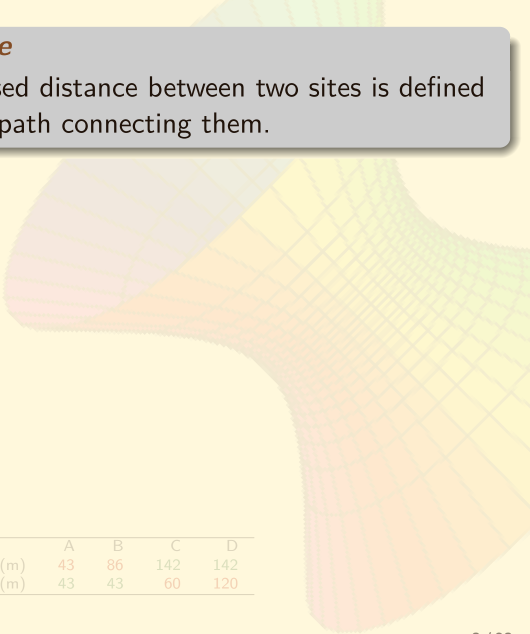


In the presence of barriers, the correlation is not directly associated with the Euclidean distance.

A practical approach

Definition: ***Cost-based distance***

Given a cost-surface, the cost-based distance between two sites is defined as the cost of the minimum-cost path connecting them.

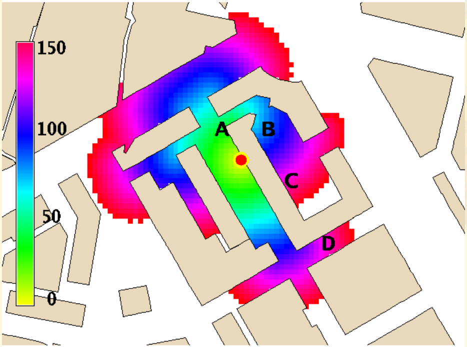


	A	B	C	D
Cost-based distance (m)	43	86	142	142
Euclidean distance (m)	43	43	60	120

A practical approach

Definition: *Cost-based distance*

Given a cost-surface, the cost-based distance between two sites is defined as the cost of the minimum-cost path connecting them.



	A	B	C	D
Cost-based distance (m)	43	86	142	142
Euclidean distance (m)	43	43	60	120

Cost-based geostatistics

- The cost-based distance generalizes the Euclidean distance, which is a particular case where the cost surface is flat
- It accounts not only for barriers but for general heterogeneous regions
- This definition and its implementation is an original contribution of the first part of the thesis project

Implementation

- Geographic computation of cost-based distances (GRASS GIS)
- Send covariates, observations and prediction locations with cost-based distance matrices to R
- Use (modified) geoR functions to perform cost-based geostatistical prediction
- Return results to GRASS GIS and produce prediction maps

Cost-based geostatistics

- The cost-based distance generalizes the Euclidean distance, which is a particular case where the cost surface is flat
- It accounts not only for barriers but for general heterogeneous regions
- This definition and its implementation is an original contribution of the first part of the thesis project

Implementation

- Geographic computation of cost-based distances (GRASS GIS)
- Send covariates, observations and prediction locations with cost-based distance matrices to R
- Use (modified) geoR functions to perform cost-based geostatistical prediction
- Return results to GRASS GIS and produce prediction maps

Cost-based geostatistics

- The cost-based distance generalizes the Euclidean distance, which is a particular case where the cost surface is flat
- It accounts not only for barriers but for general heterogeneous regions
- This definition and its implementation is an original contribution of the first part of the thesis project

Implementation

- Geographic computation of cost-based distances (GRASS GIS)
- Send covariates, observations and prediction locations with cost-based distance matrices to R
- Use (modified) geoR functions to perform cost-based geostatistical prediction
- Return results to GRASS GIS and produce prediction maps

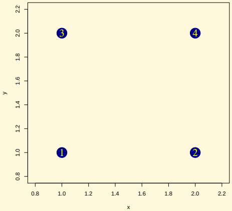
Cost-based geostatistics

- The cost-based distance generalizes the Euclidean distance, which is a particular case where the cost surface is flat
- It accounts not only for barriers but for general heterogeneous regions
- This definition and its implementation is an original contribution of the first part of the thesis project

Implementation

- Geographic computation of cost-based distances (GRASS GIS)
- Send covariates, observations and prediction locations with cost-based distance matrices to `R`
- Use (modified) `geoR` functions to perform cost-based geostatistical prediction
- Return results to GRASS GIS and produce prediction maps

Validity: a toy example



$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$$

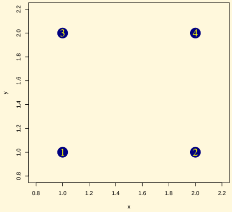
$$\mathbf{Z} = (Z_1, Z_2, Z_3, Z_4) \sim MVN(\mathbf{0}, \mathbf{\Sigma})$$

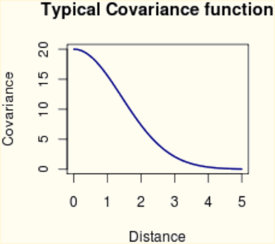
$$\mathbf{\Sigma} = \begin{pmatrix} 20.00 & 15.58 & 15.58 & 7.36 \\ 15.58 & 20.00 & 7.36 & 15.58 \\ 15.58 & 7.36 & 20.00 & 15.58 \\ 7.36 & 15.58 & 15.58 & 20.00 \end{pmatrix}$$

$$C(h) = 20e^{-\left(\frac{h}{2}\right)^2}$$

Eigenvalues: $\{58.52, 12.64, 12.64, -3.80\}$

Validity: a toy example





$$C(h) = 20e^{-\left(\frac{h}{2}\right)^2}$$

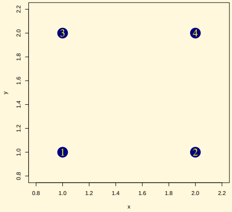
$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$$

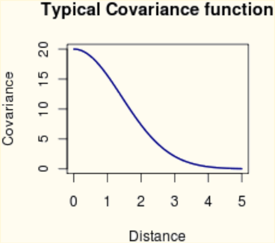
$$\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4) \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 20.00 & 15.58 & 15.58 & 7.36 \\ 15.58 & 20.00 & 7.36 & 15.58 \\ 15.58 & 7.36 & 20.00 & 15.58 \\ 7.36 & 15.58 & 15.58 & 20.00 \end{pmatrix}$$

Eigenvalues: $\{58.52, 12.64, 12.64, -3.80\}$

Validity: a toy example





$$C(h) = 20e^{-(\frac{h}{2})^2}$$

$$\mathbf{D} = \begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$$

$$\boldsymbol{Z} = (Z_1, Z_2, Z_3, Z_4) \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} 20.00 & 15.58 & 15.58 & 7.36 \\ 15.58 & 20.00 & 7.36 & 15.58 \\ 15.58 & 7.36 & 20.00 & 15.58 \\ 7.36 & 15.58 & 15.58 & 20.00 \end{pmatrix}$$

Eigenvalues: $\{58.52, 12.64, 12.64, -3.80\}$

Positive-definiteness with cost-based distances

- The positive-definite functions of the Euclidean space may not be valid with non-Euclidean distances are used
- If $\mathbf{\Sigma}$ is positive-semidefinite,
  - the kriging prediction is valid under the interpretation of a BLUP
  - it does not guarantee that C is positive-definite, thus, the underlying Gaussian field of the spatial statistical model might be invalid.
- The approach can be used safely, provided that the positive-definiteness of $\mathbf{\Sigma}$ is verified every time.

Second part of the thesis

Study the mathematical condition of positive-definiteness under cost-based distances

Positive-definiteness with cost-based distances

- The positive-definite functions of the Euclidean space may not be valid with non-Euclidean distances are used
- If $\mathbf{\Sigma}$ is positive-semidefinite,
  - the kriging prediction is valid under the interpretation of a BLUP
  - it does not guarantee that C is positive-definite, thus, the underlying Gaussian field of the spatial statistical model might be invalid
- The approach can be used safely, provided that the positive-definiteness of $\mathbf{\Sigma}$ is verified every time.

Second part of the thesis

Study the mathematical condition of positive-definiteness under cost-based distances

Positive-definiteness with cost-based distances

- The positive-definite functions of the Euclidean space may not be valid with non-Euclidean distances are used
- If $\mathbf{\Sigma}$ is positive-semidefinite,
  - the kriging prediction is valid under the interpretation of a BLUP
  - it does not guarantee that C is positive-definite, thus, the underlying Gaussian field of the spatial statistical model might be invalid
- The approach can be used safely, provided that the positive-definiteness of $\mathbf{\Sigma}$ is verified every time.

Second part of the thesis

Study the mathematical condition of positive-definiteness under cost-based distances

Positive-definiteness with cost-based distances

- The positive-definite functions of the Euclidean space may not be valid with non-Euclidean distances are used
- If $\mathbf{\Sigma}$ is positive-semidefinite,
  - the kriging prediction is valid under the interpretation of a BLUP
  - it does not guarantee that C is positive-definite, thus, the underlying Gaussian field of the spatial statistical model might be invalid
- The approach can be used safely, provided that the positive-definiteness of $\mathbf{\Sigma}$ is verified every time.

Second part of the thesis

Study the mathematical condition of positive-definiteness under cost-based distances

Positive-definiteness with cost-based distances

- The positive-definite functions of the Euclidean space may not be valid with non-Euclidean distances are used
- If $\mathbf{\Sigma}$ is positive-semidefinite,
  - the kriging prediction is valid under the interpretation of a BLUP
  - it does not guarantee that C is positive-definite, thus, the underlying Gaussian field of the spatial statistical model might be invalid
- The approach can be used safely, provided that the positive-definiteness of $\mathbf{\Sigma}$ is verified every time.

Second part of the thesis

Study the mathematical condition of positive-definiteness under cost-based distances

Positive-definiteness with cost-based distances

- The positive-definite functions of the Euclidean space may not be valid with non-Euclidean distances are used
- If $\boldsymbol{\Sigma}$ is positive-semidefinite,
  - the kriging prediction is valid under the interpretation of a BLUP
  - it does not guarantee that C is positive-definite, thus, the underlying Gaussian field of the spatial statistical model might be invalid
- The approach can be used safely, provided that the positive-definiteness of $\boldsymbol{\Sigma}$ is verified every time.

Second part of the thesis

Study the mathematical condition of positive-definiteness under cost-based distances

Riemannian model

- Define in $D \subseteq \mathbb{R}^d$ the following Riemannian metric

$$g_p(\boldsymbol{x}, \boldsymbol{y}) := \mathfrak{f}(p)^2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle, \qquad p \in D,\ \boldsymbol{x}, \boldsymbol{y} \in T_p D$$

where $\mathfrak{f}$ is the cost-surface and $\langle \cdot, \cdot \rangle$ the Euclidean inner product.

Now, given a curve α in D, its length is given by

$$L(\alpha) = \int_0^1 \sqrt{g_{\alpha(t)}(\alpha'(t), \alpha'(t))}\, dt = \int_0^1 \mathfrak{f}(\alpha(t)) \|\alpha'(t)\|\, dt.$$

This is, its *Euclidean* length weighted locally by the corresponding cost. The *metric* τ_g induced by this Riemannian metric is precisely the cost-based distance.

Riemannian model

- Define in $D \subseteq \mathbb{R}^d$ the following Riemannian metric

$$g_p(\boldsymbol{x}, \boldsymbol{y}) := \mathfrak{f}(p)^2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle, \qquad p \in D,\ \boldsymbol{x}, \boldsymbol{y} \in T_p D$$

where $\mathfrak{f}$ is the cost-surface and $\langle \cdot, \cdot \rangle$ the Euclidean inner product.

Now, given a curve α in D, its length is given by

$$L(\alpha) = \int_0^1 \sqrt{g_{\alpha(t)}(\alpha'(t), \alpha'(t))}\, dt = \int_0^1 \mathfrak{f}(\alpha(t)) \|\alpha'(t)\|\, dt.$$

This is, its *Euclidean* length weighted locally by the corresponding cost. The *metric* τ_g induced by this Riemannian metric is precisely the cost-based distance.

Riemannian model

- Define in $D \subseteq \mathbb{R}^d$ the following Riemannian metric

$$g_p(\boldsymbol{x}, \boldsymbol{y}) := \mathfrak{f}(p)^2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle, \qquad p \in D,\ \boldsymbol{x}, \boldsymbol{y} \in T_p D$$

where $\mathfrak{f}$ is the cost-surface and $\langle \cdot, \cdot \rangle$ the Euclidean inner product.

Now, given a curve α in D, its length is given by

$$L(\alpha) = \int_0^1 \sqrt{g_{\alpha(t)}(\alpha'(t), \alpha'(t))}\, dt = \int_0^1 \mathfrak{f}(\alpha(t)) \|\alpha'(t)\|\, dt.$$

This is, its *Euclidean* length weighted locally by the corresponding cost.
The *metric* τ_g induced by this Riemannian metric is precisely the cost-based distance.

Riemannian model

- Define in $D \subseteq \mathbb{R}^d$ the following Riemannian metric

$$g_p(\boldsymbol{x}, \boldsymbol{y}) := \mathfrak{f}(p)^2 \langle \boldsymbol{x}, \boldsymbol{y} \rangle, \qquad p \in D,\ \boldsymbol{x}, \boldsymbol{y} \in T_p D$$

where $\mathfrak{f}$ is the cost-surface and $\langle \cdot, \cdot \rangle$ the Euclidean inner product.

Now, given a curve α in D, its length is given by

$$L(\alpha) = \int_0^1 \sqrt{g_{\alpha(t)}(\alpha'(t), \alpha'(t))}\, dt = \int_0^1 \mathfrak{f}(\alpha(t)) \|\alpha'(t)\|\, dt.$$

This is, its *Euclidean* length weighted locally by the corresponding cost. The *metric* τ_g induced by this Riemannian metric is precisely the cost-based distance.

Positive definiteness in Riemannian manifolds

- We are interested in the family of positive-definite functions

$$\mathcal{P}(D, \mathfrak{d})$$

- In this framework the Vector Space (and group) structure is lost
- Generalizing Bochner's and Schoenberg's theorems in such an abstract context is extremely difficult
- Strategy: embedding into more structured spaces
- Embedding into an Euclidean (or Hilbert) space is not possible in general

Positive definiteness in Riemannian manifolds

- We are interested in the family of positive-definite functions

$$\mathcal{P}(D, \mathfrak{d})$$

- In this framework the Vector Space (and group) structure is lost
- Generalizing Bochner's and Schoenberg's theorems in such an abstract context is extremely difficult
- Strategy: embedding into more structured spaces
- Embedding into an Euclidean (or Hilbert) space is not possible in general

Positive definiteness in Riemannian manifolds

- We are interested in the family of positive-definite functions

$$\mathcal{P}(D, \mathfrak{d})$$

- In this framework the Vector Space (and group) structure is lost
- Generalizing Bochner's and Schoenberg's theorems in such an abstract context is extremely difficult
- Strategy: embedding into more structured spaces
- Embedding into an Euclidean (or Hilbert) space is not possible in general

Positive definiteness in Riemannian manifolds

- We are interested in the family of positive-definite functions

$$\mathcal{P}(D, \mathfrak{d})$$

- In this framework the Vector Space (and group) structure is lost
- Generalizing Bochner's and Schoenberg's theorems in such an abstract context is extremely difficult
- Strategy: embedding into more structured spaces
- Embedding into an Euclidean (or Hilbert) space is not possible in general

Positive definiteness in Riemannian manifolds

- We are interested in the family of positive-definite functions

$$\mathcal{P}(D, \mathfrak{d})$$

- In this framework the Vector Space (and group) structure is lost
- Generalizing Bochner's and Schoenberg's theorems in such an abstract context is extremely difficult
- Strategy: embedding into more structured spaces
- Embedding into an Euclidean (or Hilbert) space is not possible in general

Kuratowski embedding

The metric space D embeds isometrically in the Banach space $L^\infty(D)$ of bounded functions on D with the supremum norm. Fixing $x_0 \in D$, define

$$\begin{aligned} D &\hookrightarrow L^\infty(D) \\ x &\mapsto \phi_x : \, D \to \mathbb{R} \\ &\qquad\qquad y \mapsto \mathfrak{d}(x, y) - \mathfrak{d}(y, x_0). \end{aligned}$$

- ϕ_x are bounded (triangle ineq.)
- The norm $\|\cdot\|_\infty$ induces a distance in $L^\infty(D)$ compatible with the cost-based distance: $\|\phi_{x_1} - \phi_{x_2}\|_\infty = \mathfrak{d}(x_1, x_2)$
- Rudin (1991, Teo. 11.32) gives a generalization of Bochner's theorem in the context of Banach algebras
- Generalization of Schoenberg's theorem is open

Kuratowski embedding

The metric space D embeds isometrically in the Banach space $L^\infty(D)$ of bounded functions on D with the supremum norm. Fixing $x_0 \in D$, define

$$\begin{aligned} D &\hookrightarrow L^\infty(D) \\ x &\mapsto \phi_x : \; D \to \mathbb{R} \\ &\qquad\qquad y \mapsto \mathfrak{d}(x,y) - \mathfrak{d}(y,x_0). \end{aligned}$$

- ϕ_x are bounded (triangle ineq.)
- The norm $\|\cdot\|_\infty$ induces a distance in $L^\infty(D)$ compatible with the cost-based distance: $\|\phi_{x_1} - \phi_{x_2}\|_\infty = \mathfrak{d}(x_1, x_2)$
- Rudin (1991, Teo. 11.32) gives a generalization of Bochner's theorem in the context of Banach algebras
- Generalization of Schoenberg's theorem is open

Kuratowski embedding

The metric space D embeds isometrically in the Banach space $L^\infty(D)$ of bounded functions on D with the supremum norm. Fixing $x_0 \in D$, define

$$
\begin{aligned}
D &\hookrightarrow L^\infty(D) \\
x &\mapsto \phi_x : \; D \to \mathbb{R} \\
&\qquad\qquad y \mapsto \mathfrak{d}(x,y) - \mathfrak{d}(y,x_0).
\end{aligned}
$$

- ϕ_x are bounded (triangle ineq.)
- The norm $\|\cdot\|_\infty$ induces a distance in $L^\infty(D)$ compatible with the cost-based distance: $\|\phi_{x_1} - \phi_{x_2}\|_\infty = \mathfrak{d}(x_1, x_2)$
- Rudin (1991, Teo. 11.32) gives a generalization of Bochner's theorem in the context of Banach algebras
- Generalization of Schoenberg's theorem is open

Kuratowski embedding

The metric space D embeds isometrically in the Banach space $L^\infty(D)$ of bounded functions on D with the supremum norm. Fixing $x_0 \in D$, define

$$\begin{aligned} D &\hookrightarrow L^\infty(D) \\ x &\mapsto \phi_x : D \to \mathbb{R} \\ &\qquad\qquad y \mapsto \mathfrak{d}(x,y) - \mathfrak{d}(y,x_0). \end{aligned}$$

- ϕ_x are bounded (triangle ineq.)
- The norm $\|\cdot\|_\infty$ induces a distance in $L^\infty(D)$ compatible with the cost-based distance: $\|\phi_{x_1} - \phi_{x_2}\|_\infty = \mathfrak{d}(x_1, x_2)$
- Rudin (1991, Teo. 11.32) gives a generalization of Bochner's theorem in the context of Banach algebras
- Generalization of Schoenberg's theorem is open

Kuratowski embedding

The metric space D embeds isometrically in the Banach space $L^\infty(D)$ of bounded functions on D with the supremum norm. Fixing $x_0 \in D$, define

$$\begin{aligned} D &\hookrightarrow L^\infty(D) \\ x &\mapsto \phi_x : \; D \to \mathbb{R} \\ &\qquad\qquad y \mapsto \mathfrak{d}(x,y) - \mathfrak{d}(y,x_0). \end{aligned}$$

- ϕ_x are bounded (triangle ineq.)
- The norm $\|\cdot\|_\infty$ induces a distance in $L^\infty(D)$ compatible with the cost-based distance: $\|\phi_{x_1} - \phi_{x_2}\|_\infty = \mathfrak{d}(x_1, x_2)$
- Rudin (1991, Teo. 11.32) gives a generalization of Bochner's theorem in the context of Banach algebras
- Generalization of Schoenberg's theorem is open

Euclidean representation

An Euclidean representation of a distance matrix **D** $n \times n$ is a matrix **X** whose rows give the coordinates of a set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ that reproduce the distances.

- Not all distance matrices admit an exact Euclidean representation.
- The matrix **D** from the example does not.

Euclidean representation

An Euclidean representation of a distance matrix $\mathbf{D}$ $n \times n$ is a matrix $\mathbf{X}$ whose rows give the coordinates of a set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ that reproduce the distances.

- Not all distance matrices admit an exact Euclidean representation.
- The matrix $\mathbf{D}$ from the example does not.

Euclidean representation

An Euclidean representation of a distance matrix $\mathbf{D}$ $n \times n$ is a matrix $\mathbf{X}$ whose rows give the coordinates of a set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in \mathbb{R}^d$ that reproduce the distances.

- Not all distance matrices admit an exact Euclidean representation.
- The matrix $\mathbf{D}$ from the example does not.

Pseudo-Euclidean representation

A pseudo-Euclidean inner product in $\mathbb{R}^d$ of index k is of the form

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = (x_1 y_1 + \cdots + x_k y_k) - (x_{k+1} y_{k+1} + \cdots + x_d y_d).$$

Figure: Some points and their relative quadratic distances in the pseudo-Euclidean space $E_{(1,1)}$

Pseudo-Euclidean representation

A pseudo-Euclidean inner product in $\mathbb{R}^d$ of index k is of the form

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = (x_1 y_1 + \cdots + x_k y_k) - (x_{k+1} y_{k+1} + \cdots + x_d y_d).$$

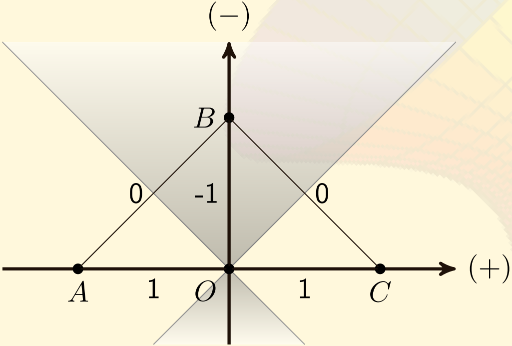


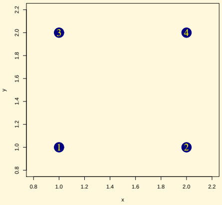
Figure: Some points and their relative quadratic distances in the pseudo-Euclidean space $E_{(1,1)}$

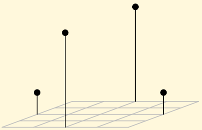
Pseudo-Euclidean representation

Theorem: All distance matrices D can be represented in a pseudo-Euclidean space

$$\mathbf{X} = \mathbf{\Gamma}(\mathbf{\Lambda}\mathbf{S}_k)^{1/2}, \quad \mathbf{HDH} = \mathbf{\Lambda}\mathbf{S}_k\mathbf{\Lambda},$$

where $\mathbf{S}_k$ is the signature of the space.





$$X = \begin{bmatrix} 0 & 1 & \frac{1}{2} \\ 1 & 0 & -\frac{1}{2} \\ -1 & 0 & -\frac{1}{2} \\ 0 & -1 & \frac{1}{2} \end{bmatrix}$$

$$d_{12}^2 = (0-1)^2 + (1-0)^2 - (\tfrac{1}{2} + \tfrac{1}{2})^2 = 1^2$$

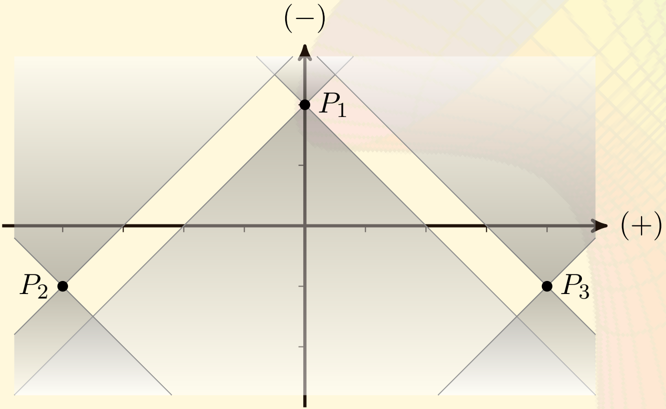
$$d_{13}^2 = (0+1)^2 + (1-0)^2 - (\tfrac{1}{2} + \tfrac{1}{2})^2 = 1^2$$

$$d_{14}^2 = (0-0)^2 + (1+1)^2 - (\tfrac{1}{2} - \tfrac{1}{2})^2 = 2^2$$

$$\ldots$$

Pseudo-Euclidean representation

- The pseudo-Euclidean embedding is not strict: there are configurations that are not representations of any cost-based problem (e.g., negative quadratic distances; violations of triangle ineq.)



Positive definiteness in pseudo-Euclidean spaces

- At least the trivial constant function is positive-definite in the pseudo-Euclidean space
- All cost-based problems can be represented in the pseudo-Euclidean space
- The Euclidean space is a particular case of cost-based manifold

$$\varnothing \neq \mathcal{P}(E_{(k,d-k)}) \subseteq \mathcal{P}(D) \subseteq \mathcal{P}(E).$$

Positive definiteness in pseudo-Euclidean spaces

- At least the trivial constant function is positive-definite in the pseudo-Euclidean space
- All cost-based problems can be represented in the pseudo-Euclidean space
- The Euclidean space is a particular case of cost-based manifold

$$\varnothing \neq \mathcal{P}(E_{(k,d-k)}) \subseteq \mathcal{P}(D) \subseteq \mathcal{P}(E).$$

Positive definiteness in pseudo-Euclidean spaces

- At least the trivial constant function is positive-definite in the pseudo-Euclidean space
- All cost-based problems can be represented in the pseudo-Euclidean space
- The Euclidean space is a particular case of cost-based manifold

$$\varnothing \neq \mathcal{P}(E_{(k,d-k)}) \subseteq \mathcal{P}(D) \subseteq \mathcal{P}(E).$$

Generalizations of Bochner's and Schoenberg's theorems

- Bochner's theorem remain valid in pseudo-Euclidean spaces!

$$\tilde{C}(\boldsymbol{h}) = \int_{E_d} e^{i\boldsymbol{h}'\boldsymbol{x}} dF_X(\boldsymbol{x}), \quad \boldsymbol{h}, \boldsymbol{X} \in E_d \tag{3}$$

- Schoenberg's theorem need to be adapted: integrate over the sphere

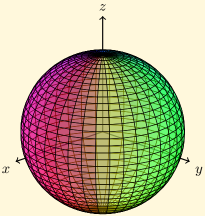
- isotropic positive-definite functions are not the same in the Euclidean and pseudo-Euclidean spaces

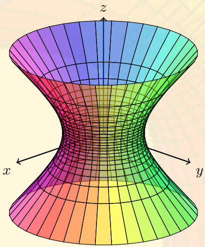
Generalizations of Bochner's and Schoenberg's theorems

- Bochner's theorem remain valid in pseudo-Euclidean spaces!

$$\tilde{C}(\boldsymbol{h}) = \int_{E_d} e^{i\boldsymbol{h}'\boldsymbol{x}} dF_X(\boldsymbol{x}), \quad \boldsymbol{h}, \boldsymbol{X} \in E_d \tag{3}$$

- Schoenberg's theorem need to be adapted: integrate over the sphere





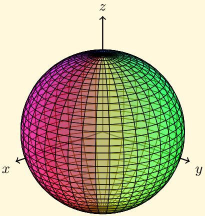
- isotropic positive-definite functions are not the same in the Euclidean and pseudo-Euclidean spaces

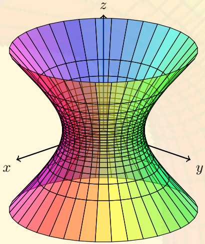
Generalizations of Bochner's and Schoenberg's theorems

- Bochner's theorem remain valid in pseudo-Euclidean spaces!

$$\tilde{C}(\boldsymbol{h}) = \int_{E_d} e^{i\boldsymbol{h}'\boldsymbol{x}} dF_X(\boldsymbol{x}), \quad \boldsymbol{h}, \boldsymbol{X} \in E_d \tag{3}$$

- Schoenberg's theorem need to be adapted: integrate over the sphere





- isotropic positive-definite functions are not the same in the Euclidean and pseudo-Euclidean spaces

Integrating on the hyperboloid

- The pseudo-Euclidean sphere has infinite surface, therefore the integration of a constant is divergent
- We can consider the *mean* value of the function $\tilde{C}(\boldsymbol{h})$ over the surface (which is $C(\rho)$, where $\rho = \|\boldsymbol{h}\|$).
- The mean of the right-hand side can be formally expressed as the quotient of two divergent integrals, and then change the integration order to express it as the integral of a function $M(\rho)$ with respect to the distribution F.

$$C(\rho) = \int_{S_\rho^+} \left(\int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}'\boldsymbol{x}} \, F(d\boldsymbol{\omega}) \right) s(d\boldsymbol{x}) \Bigg/ \int_{S_\rho^+} s(d\boldsymbol{x}) =$$

$$\int_{\mathbb{R}^d} \underbrace{\left(\int_{S_\rho^+} e^{i\boldsymbol{\omega}'\boldsymbol{x}} \, s(d\boldsymbol{x}) \Bigg/ \int_{S_\rho^+} s(d\boldsymbol{x}) \right)}_{M_{\|\boldsymbol{\omega}\|}(\rho)} F(d\boldsymbol{\omega}).$$

Integrating on the hyperboloid

- The pseudo-Euclidean sphere has infinite surface, therefore the integration of a constant is divergent
- We can consider the *mean* value of the function $\tilde{C}(\boldsymbol{h})$ over the surface (which is $C(\rho)$, where $\rho = \|\boldsymbol{h}\|$).
- The mean of the right-hand side can be formally expressed as the quotient of two divergent integrals, and then change the integration order to express it as the integral of a function $M(\rho)$ with respect to the distribution F.

$$C(\rho) = \int_{S_\rho^+} \left(\int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}'\boldsymbol{x}} \, F(d\boldsymbol{\omega}) \right) s(d\boldsymbol{x}) \Big/ \int_{S_\rho^+} s(d\boldsymbol{x}) =$$

$$\int_{\mathbb{R}^d} \underbrace{\left(\int_{S_\rho^+} e^{i\boldsymbol{\omega}'\boldsymbol{x}} \, s(d\boldsymbol{x}) \Big/ \int_{S_\rho^+} s(d\boldsymbol{x}) \right)}_{M_{\|\boldsymbol{\omega}\|}(\rho)} F(d\boldsymbol{\omega}).$$

Integrating on the hyperboloid

- The pseudo-Euclidean sphere has infinite surface, therefore the integration of a constant is divergent
- We can consider the *mean* value of the function $\tilde{C}(\boldsymbol{h})$ over the surface (which is $C(\rho)$, where $\rho = \|\boldsymbol{h}\|$).
- The mean of the right-hand side can be formally expressed as the quotient of two divergent integrals, and then change the integration order to express it as the integral of a function $M(\rho)$ with respect to the distribution F.

$$C(\rho) = \int_{S_\rho^+} \left(\int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}'\boldsymbol{x}} \, F(d\boldsymbol{\omega}) \right) s(d\boldsymbol{x}) \Bigg/ \int_{S_\rho^+} s(d\boldsymbol{x}) =$$

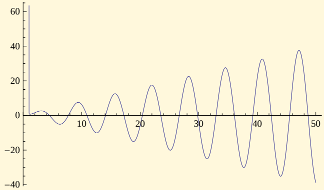
$$\int_{\mathbb{R}^d} \underbrace{\left(\int_{S_\rho^+} e^{i\boldsymbol{\omega}'\boldsymbol{x}} \, s(d\boldsymbol{x}) \Bigg/ \int_{S_\rho^+} s(d\boldsymbol{x}) \right)}_{M_{\|\boldsymbol{\omega}\|}(\rho)} F(d\boldsymbol{\omega}).$$

Divergence of the function M

- Defined formally as the mean value of the (bounded) complex exponential function over the (infinte) surface of the hyperboloid
- Integrate in pseudo-hyperspheric coordinates and reduce the problem to the quotient of one-dimensional integrals

$$\int_1^\infty x^{\frac{k}{2}} J_{\frac{k}{2}-1}(A_2 x)\,\frac{dx}{\sqrt{x^2-1}} \Bigg/ \int_1^\infty x^{k-1}\,\frac{dx}{\sqrt{x^2-1}},$$

  where J_ν denotes de Bessel function of the first kind, and A_2 is a constant.
- This is the quotient of two divergent functions. The numerator looks something like $(k=5)$



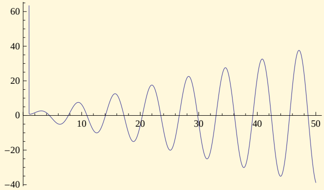
Divergence of the function M

- Defined formally as the mean value of the (bounded) complex exponential function over the (infinte) surface of the hyperboloid
- Integrate in pseudo-hyperspheric coordinates and reduce the problem to the quotient of one-dimensional integrals

$$\int_1^\infty x^{\frac{k}{2}} J_{\frac{k}{2}-1}(A_2 x) \frac{dx}{\sqrt{x^2-1}} \Bigg/ \int_1^\infty x^{k-1} \frac{dx}{\sqrt{x^2-1}},$$

where J_ν denotes de Bessel function of the first kind, and A_2 is a constant.

- This is the quotient of two divergent functions. The numerator looks something like $(k = 5)$



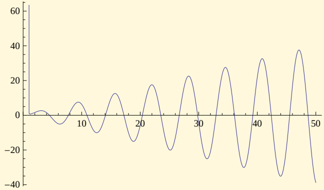
Divergence of the function M

- Defined formally as the mean value of the (bounded) complex exponential function over the (infinte) surface of the hyperboloid
- Integrate in pseudo-hyperspheric coordinates and reduce the problem to the quotient of one-dimensional integrals

$$\int_1^\infty x^{\frac{k}{2}} J_{\frac{k}{2}-1}(A_2 x) \frac{dx}{\sqrt{x^2-1}} \Bigg/ \int_1^\infty x^{k-1} \frac{dx}{\sqrt{x^2-1}},$$

where J_ν denotes de Bessel function of the first kind, and A_2 is a constant.

- This is the quotient of two divergent functions. The numerator looks something like ($k = 5$)



Spectral density functions of particular cases

- Isotropic correlation function in $E_{(2,1)}$

$$f(\boldsymbol{\omega}) = \begin{cases} \dfrac{1}{2\pi^2\|\boldsymbol{\omega}\|} \displaystyle\int_0^\infty R(\rho^2)\rho\Big(\cos(\|\boldsymbol{\omega}\|\rho) + e^{-\|\boldsymbol{\omega}\|\rho}\Big)\, d\rho, & (\boldsymbol{\omega},\boldsymbol{\omega}) > 0 \\ \dfrac{-1}{2\pi^2\|\boldsymbol{\omega}\|} \displaystyle\int_0^\infty R(\rho^2)\rho \sin(\|\boldsymbol{\omega}\|\rho)\, d\rho, & (\boldsymbol{\omega},\boldsymbol{\omega}) < 0 \end{cases}$$

- Exponential correlation function in $E_{(2,1)}$

$$f(\boldsymbol{\omega}) = \begin{cases} \dfrac{1}{2\pi^2\|\boldsymbol{\omega}\|}\left(\dfrac{\varphi^2 - \|\boldsymbol{\omega}\|^2}{(\varphi^2 + \|\boldsymbol{\omega}\|^2)^2} + \dfrac{1}{(\varphi + \|\boldsymbol{\omega}\|)^2}\right), & (\boldsymbol{\omega},\boldsymbol{\omega}) > 0 \\ \dfrac{-\varphi}{\pi^2(\varphi^2 + \|\boldsymbol{\omega}\|^2)^2}, & (\boldsymbol{\omega},\boldsymbol{\omega}) < 0 \end{cases}$$

where $\|\boldsymbol{\omega}\| = \sqrt{|(\boldsymbol{\omega},\boldsymbol{\omega})|}$. This goes negative for $\|\boldsymbol{\omega}\|$ large enough in $(\boldsymbol{\omega},\boldsymbol{\omega}) > 0$. The exponential function is not positive-definite in $E_{(2,1)}$.

Reparameterization of covariance matrices

- Model the elements of a reparameterization of the covariance matrix (e.g. Cholesky) as a function of the distances
- We still want covariances to be functions of the distances
- We need all possible covariance matrices to be positive-definite
- No significant progress on this line

Reparameterization of covariance matrices

- Model the elements of a reparameterization of the covariance matrix (e.g. Cholesky) as a function of the distances
- We still want covariances to be functions of the distances
- We need all possible covariance matrices to be positive-definite
- No significant progress on this line

Reparameterization of covariance matrices

- Model the elements of a reparameterization of the covariance matrix (e.g. Cholesky) as a function of the distances
- We still want covariances to be functions of the distances
- We need all possible covariance matrices to be positive-definite
- No significant progress on this line

Reparameterization of covariance matrices

- Model the elements of a reparameterization of the covariance matrix (e.g. Cholesky) as a function of the distances
- We still want covariances to be functions of the distances
- We need all possible covariance matrices to be positive-definite
- No significant progress on this line

Markov approximations of Matérn fields

- S different approach to irregular regions
- The resulting correlations structure is different from cost-based
- The approach works well, although is less general and has some other issues (e.g., border effects)

Markov approximations of Matérn fields

- S different approach to irregular regions
- The resulting correlations structure is different from cost-based
- The approach works well, although is less general and has some other issues (e.g., border effects)

Markov approximations of Matérn fields

- S different approach to irregular regions
- The resulting correlations structure is different from cost-based
- The approach works well, although is less general and has some other issues (e.g., border effects)

Conclusions

- Thesis topic: Geostatistical prediction in heterogeneous regions
- Main contribution 1: The cost-based methodology. A practical and applied approach, and its implementation.
- Main contribution 2: The mathematical framework of the problem of positive-definiteness with cost-based distances.
- Main contribution 3: Investigation of possible approaches. Pseudo-Euclidean embedding theorem. Formulas for the spectral density of an isotropic function in $E_{(2,1)}$.

Conclusions

- Thesis topic: Geostatistical prediction in heterogeneous regions
- Main contribution 1: The cost-based methodology. A practical and applied approach, and its implementation.
- Main contribution 2: The mathematical framework of the problem of positive-definiteness with cost-based distances.
- Main contribution 3: Investigation of possible approaches. Pseudo-Euclidean embedding theorem. Formulas for the spectral density of an isotropic function in $E_{(2,1)}$.

Conclusions

- Thesis topic: Geostatistical prediction in heterogeneous regions
- Main contribution 1: The cost-based methodology. A practical and applied approach, and its implementation.
- Main contribution 2: The mathematical framework of the problem of positive-definiteness with cost-based distances.
- Main contribution 3: Investigation of possible approaches. Pseudo-Euclidean embedding theorem. Formulas for the spectral density of an isotropic function in $E_{(2,1)}$.

Conclusions

- Thesis topic: Geostatistical prediction in heterogeneous regions
- Main contribution 1: The cost-based methodology. A practical and applied approach, and its implementation.
- Main contribution 2: The mathematical framework of the problem of positive-definiteness with cost-based distances.
- Main contribution 3: Investigation of possible approaches. Pseudo-Euclidean embedding theorem. Formulas for the spectral density of an isotropic function in $E_{(2,1)}$.

Open lines of work

- Combine the cost-based approach with the outcome of a Computer Model of noise diffusion
- Elaborate known results about positive-definite functions on Banach Algebras (Rudin, 1991; Berg et al., 1984)
- Elaborate the isotropy characterization of stationary functions under the action of a group over the manifold, considering a generalized Fourier transform with respect to the Hausdorff measure
- Mean value of a function over the d-dimensional hyperboloid
- Search positive-definite functions on the pseudo-Euclidean space using the formulas for the spectral density of isotropic functions
- Brute-force investigation of positive-definiteness for candidate functions

Open lines of work

- Combine the cost-based approach with the outcome of a Computer Model of noise diffusion
- Elaborate known results about positive-definite functions on Banach Algebras (Rudin, 1991; Berg et al., 1984)
- Elaborate the isotropy characterization of stationary functions under the action of a group over the manifold, considering a generalized Fourier transform with respect to the Hausdorff measure
- Mean value of a function over the d-dimensional hyperboloid
- Search positive-definite functions on the pseudo-Euclidean space using the formulas for the spectral density of isotropic functions
- Brute-force investigation of positive-definiteness for candidate functions

Open lines of work

- Combine the cost-based approach with the outcome of a Computer Model of noise diffusion
- Elaborate known results about positive-definite functions on Banach Algebras (Rudin, 1991; Berg et al., 1984)
- Elaborate the isotropy characterization of stationary functions under the action of a group over the manifold, considering a generalized Fourier transform with respect to the Hausdorff measure
- Mean value of a function over the d-dimensional hyperboloid
- Search positive-definite functions on the pseudo-Euclidean space using the formulas for the spectral density of isotropic functions
- Brute-force investigation of positive-definiteness for candidate functions

Open lines of work

- Combine the cost-based approach with the outcome of a Computer Model of noise diffusion
- Elaborate known results about positive-definite functions on Banach Algebras (Rudin, 1991; Berg et al., 1984)
- Elaborate the isotropy characterization of stationary functions under the action of a group over the manifold, considering a generalized Fourier transform with respect to the Hausdorff measure
- Mean value of a function over the d-dimensional hyperboloid
- Search positive-definite functions on the pseudo-Euclidean space using the formulas for the spectral density of isotropic functions
- Brute-force investigation of positive-definiteness for candidate functions

Open lines of work

- Combine the cost-based approach with the outcome of a Computer Model of noise diffusion
- Elaborate known results about positive-definite functions on Banach Algebras (Rudin, 1991; Berg et al., 1984)
- Elaborate the isotropy characterization of stationary functions under the action of a group over the manifold, considering a generalized Fourier transform with respect to the Hausdorff measure
- Mean value of a function over the d-dimensional hyperboloid
- Search positive-definite functions on the pseudo-Euclidean space using the formulas for the spectral density of isotropic functions
- Brute-force investigation of positive-definiteness for candidate functions

Open lines of work

- Combine the cost-based approach with the outcome of a Computer Model of noise diffusion
- Elaborate known results about positive-definite functions on Banach Algebras (Rudin, 1991; Berg et al., 1984)
- Elaborate the isotropy characterization of stationary functions under the action of a group over the manifold, considering a generalized Fourier transform with respect to the Hausdorff measure
- Mean value of a function over the d-dimensional hyperboloid
- Search positive-definite functions on the pseudo-Euclidean space using the formulas for the spectral density of isotropic functions
- Brute-force investigation of positive-definiteness for candidate functions

VNIVERSITAT DE VALÈNCIA
Facultat de Ciències Matemàtiques
Departament d'Estadística i Investigació Operativa

Geoestadística en regiones heterogéneas con distancia basada en el coste

TESIS DOCTORAL

Facundo Martín Muñoz Viera

Director: Antonio López-Quílez

Febrero 2013