

# Análisis de componentes principales de funciones indexadas espacialmente

Universidad Nacional de Colombia

Camilo Avellaneda

10 de junio de 2021

- 1 **Resumen**
- 2 **Introducción**
- 3 **Análisis espectral de campos aleatorios funcionales**
- 4 **Componentes principales funcionales espaciales**
- 5 **Estimación**
- 6 **Filtros lineales y proceso ARMA espacial funcional**

# Resumen

En el artículo se desarrolla una expansión, similar a la expansión de Karhunen-Loève, pero más adecuada para datos funcionales indexados en locaciones espaciales. Esta propuesta tiene en cuenta la dependencia espacial entre funciones, mediante lo cual se tiene una herramienta de reducción de dimensionalidad más eficiente. El artículo presenta la teoría asintótica y métodos de estimación correspondientes.

# Introducción

El análisis de componentes principales funcionales juega un papel relevante en el análisis de datos funcionales. Diferentes análisis en esta área trabajan con data sets de la forma  $X_s(u)$ ,  $\mathbf{s} \in \mathcal{S}$ ,  $u \in \mathcal{U}$ , donde  $\mathbf{s}$  es una ubicación espacial y  $\mathcal{U}$  denota un intervalo de tiempo. En adelante, este conjunto de datos se representará de la forma  $\{X_s, \mathbf{s} \in \mathbb{Z}^r\}$  y se denomina un campo aleatorio funcional.

# Introducción

Para definir los componentes principales funcionales, basta asumir que la distribución de  $X_s$  no depende de  $s$ , en cuyo caso los componentes principales funcionales (FPC)  $v_m$ ,  $m \geq 1$ , quedan determinadas por las funciones propias del operador de covarianza de las funciones  $X_s$ . La expresión de Karhunen-Loève se muestra en la ecuación (1)

$$X_s(t) = \sum_{m=1}^{\infty} \xi_{m,s} v_m(u), \quad u \in \mathcal{U}. \quad (1)$$

Por otro lado, el objetivo de este artículo es desarrollar una representación para datos funcionales sobre un dominio espacial que refleje la dependencia espacial.

# Análisis espectral de campos aleatorios funcionales

En esta sección se presentan resultados fundamentales para el desarrollo de una metodología del dominio de la frecuencia en el contexto de campos aleatorios funcionales definido sobre una grilla en  $\mathbb{Z}^r$ . Mientras se tienen diferentes resultados en el análisis de frecuencias en series de tiempo o en estadística espacial, no se tienen resultados correspondientes al dominio espectral para espacios de Hilbert en campos aleatorios.

# Notación y supuestos

Sea  $\{X_s, s \in \mathbb{Z}^r\}$  un campo aleatorio con una dimensión fija  $r \in \mathbb{N}$ . Cada  $X_s$  es una función aleatoria en un espacio separable de Hilbert  $H = \mathcal{L}([0, 1])$  de funciones complejas cuadrado integrables en el intervalo  $[0, 1]$ . El producto interno sobre  $H$  es definido, como se muestra en la ecuación (2). La norma se denota por  $\|x\| = \sqrt{\langle x, x \rangle}$ , mientras que el producto tensor se define sobre  $H$ , como se muestra en la ecuación (3).

$$\langle x, y \rangle = \int_0^1 x(u) \bar{y}(u) du \quad (2)$$

$$x \otimes y(\cdot) := \langle \cdot, y \rangle x. \quad (3)$$

# Notación y supuestos

Se trabaja con tres clases de operadores lineales  $A : H \rightarrow H'$ , donde  $H'$  es otro espacio separable de Hilbert.

- Clase de traza =  $S_1$ ,
- Hilbert Schmidt =  $S_2$  y
- Acotados =  $S_\infty$ .

Las correspondientes normas se definen como

- $\|A\|_{S_1} = \sum_{I \in \mathbb{N}} \langle (A^* A)^{1/2} \varphi_I, \varphi_I \rangle$ ,
- $\|A\|_{S_2} = \sum_{I \in \mathbb{N}} \|A \varphi_I\|^2$  y
- $\|A\|_{S_\infty} = \sup_{\|x\| \leq 1} \|Ax\|$ ,

donde  $\{\varphi_i, \ i \geq 1\}$  es una base ortonormal.



# Notación y supuestos

## Definición

Un campo aleatorio funcional  $(X_s)_{s \in \mathbb{Z}^r}$  se denomina estacionario en el sentido débil, si  $\forall s \in \mathbb{Z}^r$ ,

- $X_s \in \mathcal{L}^2$
- $E[X_s] = E[X_0]$
- $\forall s, h \in \mathbb{Z}^r$  y  $u, v \in [0, 1]$ , se tiene

$$c_h(u, v) := \text{Cov}(X_h(u), X_0(v)) = \text{Cov}(X_{s+h}(u), X_s(v)). \quad (4)$$

El operador integral definido por el kernel dado en la ecuación (4), se denota por  $C_h$ .

# Notación y supuestos

La definición de estacionariedad, considerada anteriormente será utilizada frecuentemente en lo que resta. Al igual, se trabajará con el supuesto de que los operadores de autocovarianza del proceso son absolutamente sumables, i.e.

$$\sum_{h \in \mathbb{Z}^r} \|C_h\|_{S_2} < \infty$$

# Notación y supuestos

- $\mathcal{L}_{\mathcal{S}(\mathcal{H}, \mathcal{H}')}^2([-\pi, \pi]^r)$  hace referencia al conjunto de operadores  $\mathcal{HS}$  definidos de  $\mathcal{H}$  a  $\mathcal{H}'$ .
- $\mathcal{L}_{\mathcal{H}}^2([-\pi, \pi]^r)$  es el espacio de funciones medibles  $x : [-\pi, \pi]^r \rightarrow \mathcal{H}$  cuadrado integrables.
- Usando el producto interno  $(x, y)_2 := \frac{1}{(2\pi)^r} \int_{[-\pi, \pi]^r} \langle x(\theta), y(\theta) \rangle d\theta$ , con su respectiva norma,  $\mathcal{L}_{\mathcal{H}}^2([-\pi, \pi]^r)$  es un espacio de Hilbert.

# Operador de la densidad espectral

## Definición

Sea  $(X_s)_{s \in \mathbb{Z}^r}$  un campo aleatorio funcional estacionario en el sentido débil. Sea

$$f_{\theta}^X(u, v) := \frac{1}{(2\pi)^r} \sum_{h \in \mathbb{Z}^r} c_h(u, v) e^{-i h^T \theta}, \quad \theta \in [-\pi, \pi]^r. \quad (5)$$

El operador integral definido a partir del anterior kernel se denomina el operador de densidad espectral de  $(X_s)$ ,  $\mathcal{F}_{\theta}^X$ , en una frecuencia espacial  $\theta$ .

# Operador de la densidad espectral

## Proposición

*Sea  $(X_s)_{s \in \mathbb{Z}^r}$  un campo aleatorio que cumple los supuestos de estacionariedad en el sentido débil y cuyos operadores de autocovarianza son absolutamente sumables, entonces  $\forall \theta$ , el operador  $\mathcal{F}_\theta^X$  es no negativo, autoadjunto, por lo cual admite la descomposición, que se muestra en la ecuación (6).*

$$\mathcal{F}_\theta^X = \sum_{m \geq 1} \lambda_m(\theta) \varphi_m(\theta) \otimes \varphi_m(\theta), \quad (6)$$

*donde  $\lambda_m(\theta)$  representa los valores propios del operador y  $\varphi_m(\theta) \in H$  su respectiva función propia.*

# Operador de la densidad espectral

## Proposición

*Adicionalmente,  $\theta \rightarrow \mathcal{F}_\theta^X$  es una función continua y*

*a) Los valores propios,  $\lambda_m(\theta)$  son funciones continuas en términos de  $\theta$ .*

*b) Suponiendo  $\sum_{\mathbf{h} \in \mathbb{Z}^r} \|h\|_\infty \|C_{\mathbf{h}}\|_{S_2} < \infty$ , entonces  $\lambda_m(\theta)$  son funciones Lipschitz continuas.*

*c) Si  $X_s$  toma valores en los reales, para cualquier frecuencia espacial  $\theta$  y  $m \geq 1$ , los valores propios  $\lambda_m(\theta)$  son funciones pares y las funciones propias son Hermitianas en  $\theta$ , i.e.,  $\lambda_m(\theta) = \lambda_m(-\theta)$  y  $\varphi_m(\theta) = \overline{\varphi_m(\theta)}$ .*

# Formula de inversión

## Proposición

Suponga que  $x \in \mathcal{L}_H^2([-\pi, \pi]^r)$ , definiendo el  $k$ -ésimo coeficiente de Fourier, como se muestra en la ecuación (7)

$$f_k := \frac{1}{(2\pi)^r} \int_{[-\pi, \pi]^r} x(\theta) e^{-ik^T \theta} d\theta \quad (7)$$

Mediante este coeficiente, la secuencia de sumas parciales

$$S_n(\theta) := \sum_{\|k\|_\infty \leq n} f_k e_k = \sum_{\|k\|_\infty \leq n} f_k e^{ik^T \theta}$$

tiene a  $x(\theta)$  como su cuadrado medio en el límite a.e., i.e.

$$\int_{[-\pi, \pi]^r} \|S_n(\theta) - x(\theta)\|^2 d\theta \rightarrow 0$$

# Dependencia espacial y filtros lineales

## Definición

- Sea  $\{\epsilon_s, s \in \mathbb{Z}^r\} \subset S$  un campo aleatorio i.i.d, que toma valores en un espacio medible  $S$  y sea  $f : S^{\mathbb{Z}^r} \rightarrow H$  una función medible.
- Suponga que  $\{X_s, s \in \mathbb{Z}^r \subset H\}$ , se puede reescribir como  $X_s = f(\{\epsilon_{s-k}, k \in \mathbb{Z}^r\})$ .



# Dependencia espacial y filtros lineales

## Definición

- Sea  $\{\epsilon'_s, s \in \mathbb{Z}^r\}$  una copia independiente de  $\{\epsilon_s, s \in \mathbb{Z}^r\}$ , se define

$$X_s^m = f((I_{\{\|k\| < m\}} \epsilon_{s-k} + I_{\{\|k\| \geq m\}} \epsilon'_{s-k})_{k \in \mathbb{Z}^r})$$

Para  $1 \leq p < \infty$ ,  $(X_s)_{s \in \mathbb{Z}^r}$  es aproximable en  $\mathcal{L}_m^p$  si

$$\sum_{h \in \mathbb{Z}^r} [E \|X_0 - X_0^{(\|h\|_\infty)}\|^p]^{1/p} < \infty \quad (8)$$

Lo que es equivalente a tener

$$\sum_{m \in \mathbb{N}} m^{r-1} V_p(X_0 - X_0^{(m)}) < \infty, \text{ con } V_p(Y) := [E(\|Y\|^p)]^{1/p} \quad (9)$$

# Dependencia espacial y filtros lineales

## Proposición

*Sea  $\{X_s, s \in \mathbb{Z}^r\}$  aproximable  $\mathcal{L}_m^2$ , entonces los operadores de autocovarianza son sumables y los operadores  $\mathcal{F}_\theta^X$ ,  $\theta \in [-\pi, \pi]^r$  son de clase de traza.*

# Dependencia espacial y filtros lineales

## Proposición

Sea  $\{X_s, s \in \mathbb{Z}^r\}$  un proceso lineal, i.e.

$$X_s = \sum_{k \in \mathbb{Z}^r} \Psi_k \epsilon_{s-k}, \quad (10)$$

donde  $\{\epsilon_s, s \in \mathbb{Z}^r\}$  es un campo aleatorio i.i.d. con  $V_p(\epsilon_0) < \infty$  con  $p \geq 1$  y  $\Psi_k$  son operadores en  $H$ . Si se cumple

$$\sum_{k \in \mathbb{N}} m^{r-1} \sum_{\|k\|_\infty \geq m} \|\Psi_k\| < \infty \quad (11)$$

entonces  $(X_s)_{s \in \mathbb{Z}^r}$  es aproximable en  $\mathcal{L}_m^p$ .

# Existencia y propiedades fundamentales

- Retomando de la ecuación (6), denotando  $\varphi_m(u|\boldsymbol{\theta})$  al valor de la función propia  $\varphi_m(\boldsymbol{\theta})$  evaluada en  $u \in [0, 1]$ .
- Como  $\varphi_m(\boldsymbol{\theta})$  es una función ortonormal, entonces  $\varphi_m(\boldsymbol{\theta}) \in \mathcal{L}^2([-\pi, \pi]^r)$ .
- Su respectivo coeficiente de Fourier, junto con su respectiva expansión se calculan como sigue

$$\phi_{m,\mathbf{l}} := \frac{1}{(2\pi)^r} \int_{[-\pi, \pi]^r} \varphi_m(u|\boldsymbol{\theta}) e^{-i\mathbf{l}^T \boldsymbol{\theta}} d\boldsymbol{\theta} \quad (12)$$

$$\varphi_m(u|\boldsymbol{\theta}) = \sum_{\mathbf{l} \in \mathbb{Z}^r} \phi_{m,\mathbf{l}}(u) e^{i\mathbf{l}^T \boldsymbol{\theta}} \quad (13)$$

# Existencia y propiedades fundamentales

## Definición

Sea  $(X_s)_{s \in \mathbb{Z}^r}$  un campo aleatorio estacionario, con operadores de covarianza absolutamente sumables. El score asociado al  $m$ -ésimo componente principal funcional espacial (SFPC) se define como se muestra en la ecuación (14).

$$Y_{m,s} := \sum_{l \in \mathbb{Z}^r} \langle X_{s-l}, \phi_{m,l} \rangle \quad (14)$$

# Existencia y propiedades fundamentales

## Proposición

Sea  $(X_s)_{s \in \mathbb{Z}^r}$  un campo aleatorio estacionario, con operadores de covarianza absolutamente sumables. Utilizando la definición dada en el anterior slide, entonces

- $E[Y_{m,s}] = 0$  y  $E[Y_{m,s}^2] = \sum_{\mathbf{k} \in \mathbb{Z}^r} \sum_{\mathbf{l} \in \mathbb{Z}^r} \langle C_{\mathbf{l}-\mathbf{k}}^X(\phi_{m,\mathbf{l}}), \phi_{m,\mathbf{k}} \rangle$
- Si  $X_s$  toma valores en los  $\mathbb{R}$  entonces los coeficientes  $\phi_{m,\mathbf{l}}(u)$  y  $Y_{m,s}$  son reales.
- Si  $C_h^X = 0 \ \forall \mathbf{h} \neq 0$  entonces  $Y_{m,s}$  coincide con los scores del FPC ordinario.
- Para  $m \neq m'$  y  $s, s' \in \mathbb{Z}^r$  los scores  $Y_{m,s}$  y  $Y_{m,s'}$  son no correlacionados.

# Expansión de Karhunen Loève espacial

Sea  $(X_s)_{s \in \mathbb{Z}^r}$  es un campo aleatorio funcional estacionario y con operadores de autocovarianza son absolutamente sumables. Sean  $\phi_{m,s}$  y  $Y_{m,s}$  como en las ecuaciones (13) y (14), entonces

$$X_s(u) = \sum_{m \geq 1} X_{m,s}(u) \text{ con } X_{m,s}(u) := \sum_{l \in \mathbb{Z}^r} Y_{m,s+l} \phi_{m,l}(u) \quad (15)$$

$$(16)$$

convergen en cuadrado medio. Las curvas  $X_{m,s}$  no están correlacionadas  $m \neq m'$

# Optimabilidad de los SPFC

Para un filtro funcional espacial  $\{\psi_{m,\mathbf{k}}, \mathbf{k} \in \mathbb{Z}^r\}$  y  $\{v_{m,\mathbf{k}}, \mathbf{k} \in \mathbb{Z}^r\}$  en  $\mathcal{H}$ .  
Tomando

$$\tilde{X}_{m,s}(u) = \sum_{\mathbf{l} \in \mathbb{Z}^r} \tilde{Y}_{m,s+\mathbf{l}} v_{m,\mathbf{l}}(u), \quad \tilde{Y}_{m,s} = \sum_{\mathbf{l} \in \mathbb{Z}^r} \langle X_{s-\mathbf{l}}, \psi_{m,\mathbf{l}} \rangle, \quad (17)$$

entonces,  $\forall p \geq 1$ ,

$$E \left[ \left\| X_s - \sum_{m=1}^p X_{m,s} \right\|^2 \right] = \sum_{m > p} \int_{[-\pi, \pi]^r} \lambda_m(\boldsymbol{\theta}) d\boldsymbol{\theta} \leq E \left[ \left\| X_s - \sum_{m=1}^p \tilde{X}_{m,s} \right\| \right]. \quad (18)$$



# Estimación del operador de densidad espectral

- La estimación del operador  $\mathcal{F}_\theta^X$  se realiza como sigue

$$\hat{F}_\theta^X := \frac{1}{(2\pi)^r} \sum_{\|\mathbf{h}\| \leq \mathbf{q}} w(\mathbf{h}/\mathbf{q}) \hat{C}_\mathbf{h} e^{-i\mathbf{h}^T \theta}, \quad (19)$$

donde  $w$  es una función de ponderaciones y el vector  $\mathbf{q}$  es un vector de parámetros a optimizar.

- El operador muestral de autocovarianzas se calcula como sigue

$$\hat{C}_\mathbf{h} := \frac{1}{N} \sum_{\mathbf{s} \in M_{\mathbf{h},n}} \quad (20)$$

donde  $M_{\mathbf{h},n} = \{\mathbf{s} \in \mathbb{Z}^r : 1 \leq s_i, s_i + h_i \leq n_i\}$ . Si  $M_{\mathbf{h},n} = \emptyset$  entonces  $\hat{C}_\mathbf{h} = 0$ .

# Scores muestrales

- $\hat{\phi}_{m,l}(u) := \frac{1}{(2\pi)^r} \int_{[-\pi,\pi]^r} \hat{\varphi}_m(u|\boldsymbol{\theta}) e^{-i\mathbf{l}^T \boldsymbol{\theta}} d\boldsymbol{\theta}$
- $\hat{Y}_{m,s} := \sum_{\|\mathbf{l}\|_\infty \leq L} \langle X_{\mathbf{s}-\mathbf{l}}, \hat{\phi}_{m,l} \rangle$
- $X_s(u) \approx \sum_{m=1}^p \hat{X}_{m,s}(u)$ , con  $\hat{X}_{m,s}(u) := \sum_{\|\mathbf{l}\|_\infty \leq L} \hat{Y}_{m,s+\mathbf{l}} \hat{\phi}_{m,l}(u)$

# Filtros lineales

Los filtros lineales y su respectiva representación del dominio espectral juegan un papel fundamental dentro de la propuesta realizada. Para ello, se define el operador rezago  $B^{\mathbf{k}}X_s := X_{s-\mathbf{k}}$  con  $\mathbf{k} \in \mathbb{Z}^r$ . El filtro funcional espacial  $SFF$ , representado por  $\Psi = \{\Psi_{\mathbf{l}}, \mathbf{l} \in \mathbb{Z}^r\}$ , es una secuencia de operadores lineales.

# Filtros lineales

El campo aleatorio funcional filtrado se define a continuación

$$Y_s = \Psi(B)X_s = \lim_{n \rightarrow \infty} \sum_{\|\mathbf{k}\|_\infty \leq n} \Psi_{\mathbf{k}}(X_{s-\mathbf{k}}). \quad (21)$$

La función de la frecuencia de la respuesta del *SFF* se define como

$$\Psi_\theta = \sum_{\mathbf{k} \in \mathbb{Z}^r} \Psi(e^{-i\mathbf{k}^T \theta}) \quad (22)$$

# Filtros lineales

De aquí en adelante  $\mathcal{S}(\mathcal{H}, \mathcal{H}')$  representa el espacio de operadores  $\mathcal{HS}$  definidos de  $\mathcal{H}$  a  $\mathcal{H}'$ .

## Proposición

Sea  $\Psi = \{\Psi_{\mathbf{k}}, \mathbf{k} \in \mathbb{Z}^r\}$  un SFF, con  $\Psi_{\mathbf{k}}$  operadores  $\mathcal{HS}$  y

- $$\int_{[-\pi, \pi]^r} \left\| \sum_{\|\mathbf{k}\| \leq n} \Psi_{\mathbf{k}} e^{-i\mathbf{k}^T \theta} - \Psi_{\theta} \right\|_{\mathcal{S}(\mathcal{H}, \mathcal{H}')}^2 d\theta \rightarrow 0 \quad (23)$$

- $$\text{esssup}_{\theta \in [-\pi, \pi]^r} \|\Psi_{\theta}\|_{\mathcal{S}_2} < \infty \quad (24)$$

- Sea  $Y_s$  calculado como en la ecuación (21), a partir de  $X_s$  un campo aleatorio estacionario, con

$$\text{esssup}_{\theta \in [-\pi, \pi]^r} \|\mathcal{F}_{\theta}^X\| < \infty \quad (25)$$

# Filtros lineales

## Proposición

Luego, se tiene

a)  $Y_s$  converge en  $\mathcal{L}_{\mathcal{H}}^2$

b)  $Y_s$  es estacionario en el sentido débil con operador de autocovarianza

$$C_h^Y = \sum_{k \in \mathbb{Z}^r} \sum_{l \in \mathbb{Z}^r} \Psi_k C_{h-k+l} \Psi_l \quad (26)$$

y operador de densidad espectral

$$\mathcal{F}_\theta^Y = \Psi_k \mathcal{F}_\theta^X \Psi_k \quad (27)$$

c)

$$\operatorname{esssup}_{\theta \in [-\pi, \pi]^r} \|\mathcal{F}_\theta^Y\| < \infty \quad (28)$$

# Proceso ARMA funcional espacial

## Definición

(SFARMA( $P, Q$ )) Sean  $P, Q \subset \mathbb{Z}^r$  dos índices finitos con  $(A_k)_{k \in P}$  y  $(B_k)_{k \in Q}$  operadores lineales acotados. Sea  $(\epsilon_s)_{s \in \mathbb{Z}^r}$  un campo aleatorio funcional de ruido blanco con media igual a 0. Sea  $(X_s)_{s \in \mathbb{Z}^r}$  un campo aleatorio funcional que satisface

$$X_s := \sum_{k \in P} A_k X_{s-k} + \sum_{k \in Q} B_k \epsilon_{s-k}. \quad (29)$$

Si  $X_s$  es estacionario en el sentido débil, este se denomina un proceso ARMA( $P, Q$ ) funcional espacial.

# Referencias I

Thomas Kuenzer, Siegfried Hörmann, and Piotr Kokoszka. Principal component analysis of spatially indexed functions. *Journal of the American Statistical Association*, pages 1–13, 2020.

Victor M Panaretos, Shahin Tavakoli, et al. Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41 (2):568–603, 2013.