Constructing functional linear filters

Denis Bosq

LSTA, Université Pierre et Marie Curie - Paris 6, France

September 2012, Bristol

Outline

1. Measurable linear transformations
2. Innovation of ARMAH processes
3. Inverse problems
4. Computing linear filters in Hilbert spaces
5. Statistics...

Linear prediction in large dimensions

Example: evolution of US Economy based on simultaneous observation of 500 series

Goal: Explicit expression of the Best Linear Predictor in a function space

Difficulty: The associated linear operator is, in general, NOT continuous

Linear prediction in large dimensions

Example: evolution of US Economy based on simultaneous observation of 500 series

Goal: Explicit expression of the Best Linear Predictor in a function space

Difficulty: The associated linear operator is, in general, NOT continuous

Introduction

Linear prediction in large dimensions

Example: evolution of US Economy based on simultaneous observation of 500 series

Goal: Explicit expression of the Best Linear Predictor in a function space

Difficulty: The associated linear operator is, in general, NOT continuous

Introduction

Linear prediction in large dimensions

Example: evolution of US Economy based on simultaneous observation of 500 series

Goal: Explicit expression of the Best Linear Predictor in a function space

Difficulty: The associated linear operator is, in general, NOT continuous

Hilbert spaces

H : real separable Hilbert space with norm $\|.\|$ and scalar product $\langle ., . \rangle$
$\mathscr{L}$: space of continuous linear operators from H to H with its usual norm $\|.\|_{\mathscr{L}}$

$L^2_H = L^2_H(\Omega, \mathscr{A}, P)$: Hilbert space of (classes of) random variables defined on the probability space $(\Omega, \mathscr{A}, P)$ and with values in $(H, \mathscr{B}_H)$, scalar product

$$[X, Y] = E \langle X, Y \rangle \ ; X, Y \in L^2_H.$$

In the following all the random variables are supposed to be **centered**.

Hilbert spaces

H : real separable Hilbert space with norm $\|.\|$ and scalar product $\langle .,.\rangle$
$\mathscr{L}$: space of continuous linear operators from H to H with its usual norm $\|.\|_{\mathscr{L}}$

$L^2_H = L^2_H(\Omega, \mathscr{A}, P)$: Hilbert space of (classes of) random variables defined on the probability space $(\Omega, \mathscr{A}, P)$ and with values in $(H, \mathscr{B}_H)$, scalar product

$$[X,Y] = E\langle X,Y\rangle \; ; X,Y \in L^2_H.$$

In the following all the random variables are supposed to be **centered**.

Linearly closed subspaces

A linear subspace $\mathscr{G}$ of L^2_H is said to be **linearly closed (LCS)** if $\mathscr{G}$ is closed in L^2_H and $X \in \mathscr{G}, l \in \mathscr{L}$ implies $l(X) \in \mathscr{G}$.

X and Y in L^2_H are said to be *weakly orthogonal* $(X \perp Y)$ if $E\langle X, Y\rangle = 0$ and *strongly orthogonal* if $C_{X,Y} = 0$ where

$$C_{X,Y}(x) = E(\langle X, x\rangle\ Y),\ x \in H$$

is the *cross-covariance operator of* X *and* Y.

Y weakly orthogonal to $\mathscr{G}$ implies Y strongly orthogonal to $\mathscr{G}$.

Linearly closed subspaces

A linear subspace $\mathscr{G}$ of L^2_H is said to be **linearly closed (LCS)** if $\mathscr{G}$ is closed in L^2_H and $X \in \mathscr{G}, l \in \mathscr{L}$ implies $l(X) \in \mathscr{G}$.

X and Y in L^2_H are said to be *weakly orthogonal* $(X \perp Y)$ if $E\langle X, Y\rangle = 0$ and *strongly orthogonal* if $C_{X,Y} = 0$ where

$$C_{X,Y}(x) = E(\langle X, x\rangle\ Y),\ x \in H$$

is the *cross-covariance operator of* X *and* Y.

Y weakly orthogonal to $\mathscr{G}$ implies Y strongly orthogonal to $\mathscr{G}$.

Measurable linear transformation

Let μ be a Probability on $(H, \mathscr{B}_H)$. An application λ is said to be a μ**−measurable linear tranformation (**μ**−MLT)** if λ is measurable and linear on a linear space S such that $\mu(S) = 1$.

It is equivalent to say that there exists a sequence $(l_k, k \geq 1)$ in $\mathscr{L}$ such that

$$l_k(x) \xrightarrow[k\to\infty]{} \lambda(x),\ x \in S.$$

(cf Mandelbaum (1984)).

λ is, in general, NOT continuous, example:

$$\lambda(x) = x'$$

In the following λ always denotes a MLT.

Measurable linear transformation

Let μ be a Probability on $(H, \mathscr{B}_H)$. An application λ is said to be a $\mu-$**measurable linear tranformation ($\mu-$MLT)** if λ is measurable and linear on a linear space S such that $\mu(S) = 1$.
It is equivalent to say that there exists a sequence $(l_k, k \geq 1)$ in $\mathscr{L}$ such that

$$l_k(x) \xrightarrow[k \to \infty]{} \lambda(x),\ x \in S.$$

(cf Mandelbaum (1984)).
λ is, in general, NOT continuous, example:

$$\lambda(x) = x'$$

In the following λ always denotes a MLT.

Measurable linear transformation

Let μ be a Probability on $(H, \mathscr{B}_H)$. An application λ is said to be a $\mu-$**measurable linear tranformation ($\mu-$MLT)** if λ is measurable and linear on a linear space S such that $\mu(S) = 1$.
It is equivalent to say that there exists a sequence $(l_k, k \geq 1)$ in $\mathscr{L}$ such that

$$l_k(x) \xrightarrow[k\to\infty]{} \lambda(x),\ x \in S.$$

(cf Mandelbaum (1984)).
λ is, in general, NOT continuous, example:

$$\lambda(x) = x'$$

In the following λ always denotes a MLT.

The gaussian case

In the gaussian case one has a more precise property:

Lemma

Let X be a $H-$valued gaussian random variable and let $\mathscr{G}_X$ be the LCS generated by X. If λ is $P_X - MLT$ there exists $(l_k, k \geq 1)$ in $\mathscr{L}$ such that

$$E\left\|l_k(X) - \lambda(X)\right\|^2 \underset{k\to\infty}{\longrightarrow} 0,$$

it follows that $\lambda(X) \in \mathscr{G}_X$.

An example

$H = L^2(\mathbb{R})$, $(h_j, j \geq 0)$ the orthonormal basis of Hermite functions, set

$$X = \sum_{j=0}^{\infty} \xi_j h_j$$

where the $\xi_j's$ are real independent and such that

$$P(\xi_j = -a_j) = P(\xi_j = a_j) = p_j, \ \ j \geq 1$$

with $p_j < \frac{1}{2}$, $\sum_j p_j < \infty$ and $a_j > 0$, $\sum_j p_j a_j^2 < \infty$. Then $P(X \in S) = 1$
where S is the **linear space of polynomials** with weight $\exp(-\frac{t^2}{2})$, $t \in \mathbb{R}$
and if $\lambda(x) = x'$ and $l_k(x)(t) = \frac{x(t+1/k)-x(t)}{1/k}$, $t \in \mathbb{R}$, $k \geq 1$,
then

$$2k \left\| l_k(x) - \lambda(x) \right\| \xrightarrow[k \to \infty]{} \left\| \lambda^2(x) \right\|, \, x \in S.$$

An example

$H = L^2(\mathbb{R})$, $(h_j, j \geq 0)$ the orthonormal basis of Hermite functions, set

$$X = \sum_{j=0}^{\infty} \xi_j h_j$$

where the $\xi_j' s$ are real independent and such that

$$P(\xi_j = -a_j) = P(\xi_j = a_j) = p_j, \;\; j \geq 1$$

with $p_j < \frac{1}{2}$, $\sum_j p_j < \infty$ and $a_j > 0$, $\sum_j p_j a_j^2 < \infty$. Then $P(X \in S) = 1$

where S is the **linear space of polynomials** with weight $\exp(-\frac{t^2}{2})$, $t \in \mathbb{R}$ and if $\lambda(x) = x'$ and $I_k(x)(t) = \frac{x(t+1/k) - x(t)}{1/k}$, $t \in \mathbb{R}$, $k \geq 1$, then

$$2k \left\| I_k(x) - \lambda(x) \right\| \xrightarrow[k \to \infty]{} \left\| \lambda^2(x) \right\|, \; x \in S.$$

Projection on a LCS

The link between MLT and LCS appears in the following statement

Proposition

Let $\mathscr{G}_X$ be the LCS generated by X and Π^X its orthogonal projector in L^2_H. Then, for each Y in L^2_H, there exists a P_X−MLT λ_0 such that

$$\Pi^X(Y) = \lambda_0(X).$$

Projection on a LCS

The link between MLT and LCS appears in the following statement

Proposition

Let $\mathscr{G}_X$ be the LCS generated by X and Π^X its orthogonal projector in L^2_H. Then, for each Y in L^2_H, there exists a P_X-MLT λ_0 such that

$$\Pi^X(Y) = \lambda_0(X).$$

Continuity

The next proposition underscores a special case where λ_0 is **continuous**:

Proposition

The following statements are equivalent

a) There exists $\alpha \geq 0$ *such that* $\left\| C_{X,Y}(x) \right\| \leq \alpha \left\| C_X(x) \right\|, \ x \in H,$

b) There exists $l_0 \in \mathscr{L}$ *such that* $C_{X,Y} = l_0 C_X$,

c) $\Pi^X(Y) = l_0(X)$.

Continuity

The next proposition underscores a special case where λ_0 is **continuous**:

Proposition

The following statements are equivalent

a) There exists $\alpha \geq 0$ *such that* $\|C_{X,Y}(x)\| \leq \alpha \|C_X(x)\|,\ x \in H,$

b) There exists $l_0 \in \mathscr{L}$ *such that* $C_{X,Y} = l_0 C_X$,

c) $\Pi^X(Y) = l_0(X)$.

Innovation

A $H-$ **white noise** is a sequence $(\varepsilon_n, n \in \mathbb{Z})$ of strongly orthogonal $H-$valued random variables such that $E\|\varepsilon_n\|^2 = \sigma^2 > 0$ and $E\varepsilon_n = 0,\ n \in \mathbb{Z}$.

A **weakly stationary process** *in* H satisfies

$$C_{X_{n+h},X_{m+h}} = C_{X_n,X_m},\ n,m,h \in \mathbb{Z}.$$

$(\varepsilon_n, n \in \mathbb{Z})$ is the **innovation** *of* $(X_n, n \in \mathbb{Z})$ *if*

$$X^*_{n+1} = X_n + \varepsilon_{n+1},\ n \in \mathbb{Z},$$

where X^*_{n+1} is the best linear predictor of X_{n+1} given $X_n, X_{n-1}, \ldots$

Innovation

A $H-$ **white noise** is a sequence $(\varepsilon_n, n \in \mathbb{Z})$ of strongly orthogonal $H-$valued random variables such that $E\|\varepsilon_n\|^2 = \sigma^2 > 0$ and $E\varepsilon_n = 0,\ n \in \mathbb{Z}$.

A **weakly stationary process** *in* H satisfies

$$C_{X_{n+h},X_{m+h}} = C_{X_n,X_m},\ n, m, h \in \mathbb{Z}.$$

$(\varepsilon_n, n \in \mathbb{Z})$ is the **innovation** *of* $(X_n, n \in \mathbb{Z})$ *if*

$$X^*_{n+1} = X_n + \varepsilon_{n+1},\ n \in \mathbb{Z},$$

where X^*_{n+1} is the best linear predictor of X_{n+1} given $X_n, X_{n-1}, \ldots$

Innovation

A $H-$ **white noise** is a sequence $(\varepsilon_n,\, n \in \mathbb{Z})$ of strongly orthogonal $H-$valued random variables such that $E\left\|\varepsilon_n\right\|^2 = \sigma^2 > 0$ and $E\varepsilon_n = 0,\ n \in \mathbb{Z}$.
A **weakly stationary process** *in* H satisfies

$$C_{X_{n+h},X_{m+h}} = C_{X_n,X_m},\ n,m,h \in \mathbb{Z}.$$

$(\varepsilon_n,\, n \in \mathbb{Z})$ is the **innovation** *of* $(X_n,\, n \in \mathbb{Z})$ *if*

$$X^*_{n+1} = X_n + \varepsilon_{n+1},\ n \in \mathbb{Z},$$

where X^*_{n+1} is the best linear predictor of X_{n+1} given $X_n, X_{n-1}, \ldots$

ARH(1)

Set $\mathscr{M}_n$ be the LCS generated by $X_n, X_{n-1}, \ldots$. A stationary process is an **autoregressive process of order 1** in H (ARH(1)) if

$$\Pi^{\mathscr{M}_{n-1}}(X_n) = \Pi^{\mathscr{G}_{X_{n-1}}}(X_n)$$

Hence

$$X_n = \lambda_n(X_{n-1}) + \varepsilon_n,\ n \in \mathbb{Z},$$

where λ_n is MLT and (ε_n) is the innovation.

ARH(1)

Set $\mathscr{M}_n$ be the LCS generated by $X_n, X_{n-1}, \ldots$. A stationary process is an **autoregressive process of order 1** in H (ARH(1)) if

$$\Pi^{\mathscr{M}_{n-1}}(X_n) = \Pi^{\mathscr{G}_{X_{n-1}}}(X_n)$$

Hence

$$X_n = \lambda_n(X_{n-1}) + \varepsilon_n,\ n \in \mathbb{Z},$$

where λ_n is MLT and (ε_n) is the innovation.

Innovation of an ARH(1)

Proposition

Suppose that the equation

$$X_n = \lambda(X_{n-1}) + \varepsilon_n,\ n \in \mathbb{Z} \tag{1}$$

has a solution such that $\lambda : S \longmapsto S$ *is* $P_{X_n} - MLT$ *for all* n, $\lambda^j(X_{n-j}) \in \mathscr{G}_{X_{n-j}}$ *and* $\lambda^j(\varepsilon_{n-j}) \in \mathscr{G}_{\varepsilon_{n-j}}, j \geq 1$, *then if*

$$\frac{1}{k}\sum_{j=1}^{k} E\left\|\lambda^j(X_{n-j})\right\|^2 \xrightarrow[k\to\infty]{} 0,\ n \in \mathbb{Z}$$

(1) *has a unique stationary solution given by*

$$X_n = \lim_{k\to\infty}(L_H^2)\sum_{j=0}^{k-1}\left(1-\frac{j}{k}\right)\lambda^j(\varepsilon_{n-j}),\ n \in \mathbb{Z}$$

and (ε_n) *is the innovation of* (X_n).

Innovation of an ARH(1)

Proposition

Suppose that the equation

$$X_n = \lambda(X_{n-1}) + \varepsilon_n,\ n \in \mathbb{Z} \tag{1}$$

has a solution such that $\lambda : S \longmapsto S$ *is* $P_{X_n} - MLT$ *for all* n, $\lambda^j(X_{n-j}) \in \mathscr{G}_{X_{n-j}}$ *and* $\lambda^j(\varepsilon_{n-j}) \in \mathscr{G}_{\varepsilon_{n-j}}, j \geq 1$, *then if*

$$\frac{1}{k}\sum_{j=1}^{k} E\left\|\lambda^j(X_{n-j})\right\|^2 \xrightarrow[k\to\infty]{} 0,\ n \in \mathbb{Z}$$

(1) *has a unique stationary solution given by*

$$X_n = \lim_{k\to\infty}(L_H^2)\sum_{j=0}^{k-1}\left(1-\frac{j}{k}\right)\lambda^j(\varepsilon_{n-j}),\ n \in \mathbb{Z}$$

and (ε_n) *is the innovation of* (X_n).

Innovation of an ARH(1)

Proposition

Suppose that the equation

$$X_n = \lambda(X_{n-1}) + \varepsilon_n, \; n \in \mathbb{Z} \tag{1}$$

has a solution such that $\lambda : S \longmapsto S$ *is* $P_{X_n} - MLT$ *for all* n, $\lambda^j(X_{n-j}) \in \mathscr{G}_{X_{n-j}}$ *and* $\lambda^j(\varepsilon_{n-j}) \in \mathscr{G}_{\varepsilon_{n-j}}, j \geq 1$*, then if*

$$\frac{1}{k}\sum_{j=1}^{k} E\left\|\lambda^j(X_{n-j})\right\|^2 \xrightarrow[k\to\infty]{} 0, \; n \in \mathbb{Z}$$

(1) *has a unique stationary solution given by*

$$X_n = \lim_{k\to\infty}(L_H^2)\sum_{j=0}^{k-1}(1 - \frac{j}{k})\lambda^j(\varepsilon_{n-j}), \; n \in \mathbb{Z}$$

and (ε_n) *is the innovation of* (X_n).

Proof

The proof is based on the relation

$$X_n = \sum_{j=0}^{k-1} (1 - \frac{j}{k}) \lambda^j (\varepsilon_{n-j}) + \frac{1}{k} \sum_{j=1}^{k} \lambda^j (X_{n-j}).$$

The above condition is strictly weaker than the classical conditions like:

"λ is continuous and there exists an integer j_0 such that $\|\lambda^j\|_{\mathscr{L}} < 1,\ j \geq j_0$." (cf Bosq-Blanke 2007)

Proof

The proof is based on the relation

$$X_n = \sum_{j=0}^{k-1} (1 - \frac{j}{k}) \lambda^j(\varepsilon_{n-j}) + \frac{1}{k} \sum_{j=1}^{k} \lambda^j(X_{n-j}).$$

The above condition is strictly weaker than the classical conditions like:

"λ is continuous and there exists an integer j_0 such that $\left\|\lambda^j\right\|_{\mathscr{L}} < 1,\, j \geq j_0$." (cf Bosq-Blanke 2007)

Innovation of a MAH(1)

Proposition

Suppose that (X_n) *is defined by*

$$X_n = \varepsilon_n - \lambda(\varepsilon_{n-1}),\ n \in \mathbb{Z}.$$

where $\lambda : H_1 \longmapsto H_1$ *is* $P_{\varepsilon_n} - MLT$ *for all* n*, with* $\lambda^j(X_{n-j}) \in \mathscr{G}_{X_{n-j}}, \lambda^j(\varepsilon_{n-j}) \in \mathscr{G}_{\varepsilon_{n-j}}, j \geq 1,\ n \in \mathbb{Z}$*, then, if*

$$\frac{1}{k^2} \sum_{j=1}^{k} E \left\| \lambda^j(\varepsilon_{n-j}) \right\|^2 \xrightarrow[k \to \infty]{} 0,$$

(ε_n) *is the innovation of* (X_n) *and*

$$\varepsilon_n = \lim_{k \to \infty} (L^2_H) \sum_{j=0}^{k-1} (1 - \frac{i}{k}) \lambda^j(X_{n-j}).$$

"Roots of modulus 1"

The condition is weak. In particular if λ is continuous and such that

$$\left\|\lambda^j\right\|_{\mathscr{L}} \leq 1,\, j \geq 1$$

the above Proposition holds. A simple example is

$$X_n = \varepsilon_n - \Pi^G(\varepsilon_{n-1}),\, n \in \mathbb{Z}$$

where G is a closed subspace of H and Π^G its orthogonal projector.

If the MA is real, it corresponds to **roots of modulus** 1.

"Roots of modulus 1"

The condition is weak. In particular if λ is continuous and such that

$$\left\|\lambda^j\right\|_{\mathscr{L}} \leq 1,\, j \geq 1$$

the above Proposition holds. A simple example is

$$X_n = \varepsilon_n - \Pi^G(\varepsilon_{n-1}),\, n \in \mathbb{Z}$$

where G is a closed subspace of H and Π^G its orthogonal projector.

If the MA is real, it corresponds to **roots of modulus** 1.

Example

Example

In $L^2[0,1]$ consider the white noise

$$\varepsilon_n(t) = \sum_{i=0}^{\infty} \xi_{ni} \frac{t^i}{i!},\ t \in [0,1],\ \ n \in \mathbb{Z}$$

where (ξ_{ni}) is a sequence of real independent random variables such that, for all n, $\xi_{ni} \sim \mathcal{N}(0, \sigma_i^2)$ where $0 < \sum_{i=1}^{\infty} \sigma_i^2 < \infty$. Set

$$X_n(t) = \varepsilon_n(t) - \varepsilon'_{n-1}(t)$$

then (ε_n) is the innovation.

Example

Example

In $L^2[0, 1]$ consider the white noise

$$\varepsilon_n(t) = \sum_{i=0}^{\infty} \xi_{ni} \frac{t^i}{i!}, \ t \in [0, 1], \ \ n \in \mathbb{Z}$$

where (ξ_{ni}) is a sequence of real independent random variables such that, for all n, $\xi_{ni} \sim \mathcal{N}(0, \sigma_i^2)$ where $0 < \sum_{i=1}^{\infty} \sigma_i^2 < \infty$. Set

$$X_n(t) = \varepsilon_n(t) - \varepsilon_{n-1}^{'}(t)$$

then (ε_n) is the innovation.

The mixed case

Proposition

Consider the ARMAH (1,1) process defined as

$$\varepsilon_n - l(\varepsilon_{n-1}) = X_n - \rho(X_{n-1}),\ n \in \mathbb{Z}$$

where (ε_n) *is a* $H-$*white noise and* l *and* ρ *belong to* $\mathscr{L}$*; suppose that*

$$\frac{1}{k^2}\sum_{j=0}^{k} \left\| l^j \right\|_{\mathscr{L}}^2 \xrightarrow[k\to\infty]{} 0 \text{ and that } \frac{1}{k}\sum_{j=0}^{k} \left\| \rho^j \right\|_{\mathscr{L}}^2 \xrightarrow[k\to\infty]{} 0$$

then, if that equation has a stationary solution, it is given by

$$X_n = \lim_{k\to\infty} (L_H^2) \sum_{j=0}^{k-1} (1 - \frac{j}{k})\,\rho^j(\varepsilon_{n-j} - l(\varepsilon_{n-j-1}),\ n \in \mathbb{Z}$$

and (ε_n) *is the innovation of* (X_n)*.*

Compound Ornstein-Uhlenbeck process

Example

Consider the Hilbert space $H = L^2([0,1], \mathscr{B}_{[0,1]}, \mu)$ where μ is the sum of Lebesgue measure and Dirac measure at the point 1. Set

$$\varepsilon_n(t) = \int_n^{n+t} \exp(-\theta(n+t-s))\, dW(s),\ t \in [0,1],\ n \in \mathbb{Z},\ (\theta > 0),$$

where W is a bilateral standard Wiener process. Put $I(x)(t) = x(t)$, and $\rho(x)(t) = \exp(-\theta t).x(1)\ t \in [0,1],\ x \in H$. Then the process

$$X_n(t) = \exp(-(\theta(n+t)) \int_{-\infty}^{n+t} \exp(\theta s)\, dW(s)$$
$$- \exp(-\theta(n-1+t)) \int_{-\infty}^{n-1+t} \exp(\theta s)\, dW(s),\ t \in [0,1],\ n \in \mathbb{Z}$$

is a stationary ARMAH (1,1).

Compound Ornstein-Uhlenbeck process

Example

Consider the Hilbert space $H = L^2([0,1], \mathscr{B}_{[0,1]}, \mu)$ where μ is the sum of Lebesgue measure and Dirac measure at the point 1. Set

$$\varepsilon_n(t) = \int_n^{n+t} \exp(-\theta(n+t-s))\, dW(s),\ t \in [0,1],\ n \in \mathbb{Z},\ (\theta > 0),$$

where W is a bilateral standard Wiener process. Put $I(x)(t) = x(t)$, and $\rho(x)(t) = \exp(-\theta t).x(1)\ t \in [0,1],\ x \in H$. Then the process

$$X_n(t) = \exp(-(\theta(n+t)) \int_{-\infty}^{n+t} \exp(\theta s)\, dW(s)$$
$$- \exp(-\theta(n-1+t)) \int_{-\infty}^{n-1+t} \exp(\theta s)\, dW(s),\ t \in [0,1],\ n \in \mathbb{Z}$$

is a stationary ARMAH (1,1).

Kalman-Bucy filter in H

Example

Consider the model

$$X_n = r(Y_n) + \varepsilon_n,\ n \geq 1$$

$$Y_n = \rho(Y_{n-1}) + \eta_n,\ n \geq 1$$

where (X_n) and (Y_n) are $H-$valued stationary processes and where (ε_n) and (η_n) are two strongly orthogonal white noises such that $C_{\varepsilon_n, Y_n} = C_{\eta_n, Y_{n-1}} = 0$; ρ and r belong to $\mathscr{L}$. Then, if $r\rho = \rho r$, (X_n) is an ARMAH (1,1).

Other examples of Kalman-Bucy filter in H appear in Ruiz-Medina et *al* in a spatial framework.

Kalman-Bucy filter in H

Example

Consider the model

$$X_n = r(Y_n) + \varepsilon_n,\ n \geq 1$$

$$Y_n = \rho(Y_{n-1}) + \eta_n,\ n \geq 1$$

where (X_n) and (Y_n) are $H-$valued stationary processes and where (ε_n) and (η_n) are two strongly orthogonal white noises such that $C_{\varepsilon_n, Y_n} = C_{\eta_n, Y_{n-1}} = 0$; ρ and r belong to $\mathscr{L}$. Then, if $r\rho = \rho r$, (X_n) is an ARMAH (1,1).

Other examples of Kalman-Bucy filter in H appear in Ruiz-Medina et *al* in a spatial framework.

MAH(2)

Proposition

Consider a MAH(2) admitting the decomposition

$$X_n = \varepsilon_n - (\alpha + \beta)(\varepsilon_{n-1}) + \beta\alpha(\varepsilon_{n-2}),\ n \in \mathbb{Z}$$

where (ε_n) *is a white noise and* $\alpha, \beta \in \mathscr{L}$ *and suppose that*

$$\frac{1}{k^2}\sum_{j=0}^{k} \left\|\alpha^j\right\|_{\mathscr{L}}^2 \xrightarrow[k\to\infty]{} 0$$

and

$$\frac{1}{k^2}\sum_{j=0}^{k} \left\|\beta^j\right\|_{\mathscr{L}}^2 \xrightarrow[k\to\infty]{} 0$$

then (ε_n) *is the innovation of* (X_n).

Constructing the innovation

What about the case where the noise associated with the process is NOT the innovation?

The case of a MAH(1)

Proposition

Consider the MAH(1) given by

$$X_n = e_n - l(e_{n-1}),\ n \in \mathbb{Z}$$

where $l \in \mathscr{L}$ *and* (e_n) *is a* $H-$ *white noise. We suppose that* l *is symmetric, invertible, such that* $\left\|(l^{-1})^{j_0}\right\|_{\mathscr{L}} < 1$ *for some* $j_0 \geq 1$*. Moreover* l *and* C_{e_0} *commute.*
Then, the innovation of (X_n) *is defined as* $\varepsilon_n = (I - l^{-1}B)^{-1}(I - lB)\, e_n$ *where* B *is the backward operator* $(B(x_n) = x_{n-1})$*, convergence takes place in* L^2_H*, and*

$$X_n = \varepsilon_n - l^{-1}(\varepsilon_{n-1}),\ n \in \mathbb{Z}.$$

In addition one has

$$C_{\varepsilon_0} = l^2\, C_{e_0}.$$

An example

Example

Suppose that

$$l = \sum_{i=1}^{\infty} a_i \, v_i \otimes v_i$$

where (v_i) is an orthonormal system in H and $1 < |a_1| \le |a_2| \le \ldots \le a < \infty$; and that

$$C_{e_0} = \sum_{i=1}^{\infty} c_i \, v_i \otimes v_i$$

then the above Proposition holds.

The case of an ARH(1)

Proposition

Consider the equation $X_n = r(X_{n-1}) + \eta_n \;\; n \in \mathbb{Z}$
where (η_n) *is a* $H-$*white noise and* $r \in \mathscr{L}$, *and suppose that*

$$\exists r^{-1} : \; \frac{1}{k} \sum_{j=1}^{k} \left\| r^{-j} \right\|_{\mathscr{L}}^{2} \xrightarrow[k \to \infty]{} 0,$$

then it has a stationary solution given by

$$X_n = - \lim_{k \to \infty} (L_H^2) \sum_{j=1}^{k} (1 - \frac{j-1}{k}) \, r^{-j}(\eta_{n+j}), \;\; n \in \mathbb{Z}.$$

If, in addition, $r^{-1} C_{X_0}$ *is symmetric and* $C_{X_0}(Is - (r^*)^{-2}) \neq 0$, *then the innovation of* (X_n) *is*

$$\varepsilon_n = X_n - r^{-1}(X_{n-1}) \;\; n \in \mathbb{Z}$$

Starting from the best predictor

Principle: Given the best linear predictor (BLP) find the associated model.
Choice: **Extended exponential smoothing** in H :

$$X^*_{n+1} = \alpha\left(\sum_{j=0}^{\infty} \beta^j(X_{n-j})\right),$$

where α and β belong to $\mathscr{L}$ and $\alpha\beta = \beta\alpha$. Then one has

$$X^*_{n+1} = \alpha(X_n) + \beta(X^*_n).$$

Associated model

Proposition

Suppose that $\left\|\beta^{j_0}\right\|_{\mathscr{L}} < 1$ *and* $\left\|(\alpha+\beta)^{j_0}\right\|_{\mathscr{L}} < 1$ *for some integer* j_0, *and that* $\alpha \neq 0$. *If* (X_n) *is a regular zero-mean stationary process with innovation* (ε_n) *and such that the BLP is*

$$X^*_{n+1} = \alpha\left(\sum_{j=0}^{\infty} \beta^j (X_{n-j})\right)$$

where $\alpha\beta = \beta\alpha$, *then* (X_n) *is an ARMAH (1,1):*

$$X_n - (\alpha+\beta)(X_{n-1}) = \varepsilon_n - \beta(\varepsilon_{n-1}), \qquad (2)$$

Conversely, if (X_n) *satisfies 2, then* X^*_{n+1} *is BLP for every* n.

(X, Y) in $G \times H$ real separable Hilbert spaces with spectral decompositions:

$$C_X = \sum_{i \in I} \alpha_i \, v_i \otimes v_i \;\; (\alpha_i > 0, \sum_{i \in I} \alpha_i < \infty)$$

and

$$C_Y = \sum_{j \in J} \beta_j \, w_j \otimes w_j \;\; (\beta_j > 0, \sum_{j \in J} \beta_j < \infty)$$

I and J are finite or infinite. Let $\mathscr{L}(G, H)$ be the space of continuous linear operators from G to H. Set

$$\mathscr{F}_X = \overline{sp}\{l(X), l \in \mathscr{L}(G, H)\}$$

where the closure is taken in $L^2_H = L^2_H(\Omega, \mathscr{A}, P)$.

(X, Y) in $G \times H$ real separable Hilbert spaces with spectral decompositions:

$$C_X = \sum_{i \in I} \alpha_i \, v_i \otimes v_i \ \ (\alpha_i > 0, \sum_{i \in I} \alpha_i < \infty)$$

and

$$C_Y = \sum_{j \in J} \beta_j \, w_j \otimes w_j \ \ (\beta_j > 0, \sum_{j \in J} \beta_j < \infty)$$

I and J are finite or infinite. Let $\mathscr{L}(G, H)$ be the space of continuous linear operators from G to H. Set

$$\mathscr{F}_X = sp\{l(X), l \in \mathscr{L}(G, H)\}$$

where the closure is taken in $L^2_H = L^2_H(\Omega, \mathscr{A}, P)$.

Let $\mathscr{L}(G,H)$ be the space of continuous linear operators from G to H. Set

$$\mathscr{F}_X = sp\{l(X), l \in \mathscr{L}(G,H)\}$$

where the closure is taken in $L^2_H = L^2_H(\Omega, \mathscr{A}, P)$.

The best linear predictor of Y given X is the orthogonal projection of Y on $\mathscr{F}_X$.

Let $\mathscr{L}(G,H)$ be the space of continuous linear operators from G to H. Set

$$\mathscr{F}_X = sp\{l(X), l \in \mathscr{L}(G,H)\}$$

where the closure is taken in $L^2_H = L^2_H(\Omega, \mathscr{A}, P)$.

The best linear predictor of Y given X is the orthogonal projection of Y on $\mathscr{F}_X$.

The best linear predictor

Proposition

The best linear predictor (BLP) of Y given X is

$$\lambda_0(X) = \sum_{i\in I, j\in J} \gamma_{ij}\,(v_i \otimes w_j)(X) \ \ (L^2_H)$$

where

$$\gamma_{ij} = \frac{E(\langle X, v_i\rangle_G \langle Y, w_j\rangle_H)}{E\,\langle X, v_i\rangle_G^2}, \ \ i\in I, j\in J$$

Proof

The proof uses the fact that

$$U_{ij} = \frac{\langle X, v_i \rangle_G}{\sqrt{\alpha_i}} . w_j \;\; i \in I, j \in J,$$

is an orthonormal system in L^2_H.

Continuity

λ_0 is a P_X-MLT. Continuity of λ_0 appears in the next statement

Proposition

If there exists $l_0 \in \mathscr{L}(G, H)$ such that

$$C_{X,Y} = l_0 \, C_X ,$$

then the best linear predictor takes the form

$$l_0(X) = \sum_{i \in I} \frac{\langle X, v_i \rangle_G}{\alpha_i} \, C_{X,Y}(v_i)$$

Converse

Proposition

If there exists $l_0 : G \longmapsto H$ *such that*

$$l_0(x) = \sum_{i=1}^{\infty} \frac{\langle x, v_i \rangle_G}{\alpha_i} C_{X,Y}(v_i), \ \ x \in H, \ \ (H)$$

then $l_0 \in \mathscr{L}(G, H)$ *and* $C_{X,Y} = l_0 \, C_X$.

The gaussian case

In the gaussian case a similar result can be obtained without continuity assumption:

Proposition

If $G = H$ and the vector (X, Y) is gaussian then the conditional expectation $E(Y \mid X)$ and the BLP coincide and have the form

$$E(Y \mid X) = \lambda_0(X) = \sum_{i \in I} \frac{\langle X, v_i \rangle_G}{\alpha_i} C_{X,Y}(v_i)$$

The gaussian case

In the gaussian case a similar result can be obtained without continuity assumption:

Proposition

If $G = H$ and the vector (X, Y) is gaussian then the conditional expectation $E(Y \mid X)$ and the BLP coincide and have the form

$$E(Y \mid X) = \lambda_0(X) = \sum_{i \in I} \frac{\langle X, v_i \rangle_G}{\alpha_i} C_{X,Y}(v_i)$$

Proof

The proof uses the fact that the sequence

$$E(\langle Y, y\rangle_H \,|(\langle X, v_1\rangle_G, \dots, \langle X, v_m\rangle_G)) = \sum_{i=1}^{m} \frac{E(\langle X, v_i\rangle_G \langle Y, y\rangle_H)}{E(\langle X, v_i\rangle_G^2)} \langle X, v_i\rangle_G ,$$

$m \geq 1$, is a martingale in L_H^2.

Basis of a LCS

The final statement is useful for computing a BLP

Proposition

The LCS $\mathscr{G}_X$ of L^2_G has the orthonormal basis

$$\mathscr{B} = \left\{ \frac{\langle X, v_i \rangle_G}{\alpha_i^{1/2}} v_j,\ i \in I, j \in I \right\} \cup \left\{ \frac{\langle X, v_i \rangle_G}{\alpha_i^{1/2}} u_j,\ i \in I, j \in J' \right\}$$

where

$$C_X = \sum_{i \in I} \alpha_i\, v_i \otimes v_i \ \ (\alpha_i > 0, \sum_{i \in I} \alpha_i < \infty)$$

and $(u_j, j \in J')$ is an orthonormal basis of the orthogonal complement of the closed subspace of G generated by $(v_i, i \in I)$.

(cf Bosq-Mourid (2012)).

Basis of a LCS

The final statement is useful for computing a BLP

Proposition

The LCS $\mathscr{G}_X$ of L^2_G has the orthonormal basis

$$\mathscr{B} = \left\{ \frac{\langle X, v_i \rangle_G}{\alpha_i^{1/2}} v_j,\ i \in I, j \in I \right\} \cup \left\{ \frac{\langle X, v_i \rangle_G}{\alpha_i^{1/2}} u_j,\ i \in I, j \in J' \right\}$$

where

$$C_X = \sum_{i \in I} \alpha_i\, v_i \otimes v_i \ \ (\alpha_i > 0, \sum_{i \in I} \alpha_i < \infty)$$

and $(u_j, j \in J')$ is an orthonormal basis of the orthogonal complement of the closed subspace of G generated by $(v_i, i \in I)$.

(cf Bosq-Mourid (2012)).

Applications: model with noise

Consider the model

$$X = r(Y) + \varepsilon$$

with $r \in \mathscr{L}(H, G)$ and $C_{Y,\varepsilon} = 0$, where only X is observed, Then

$$C_{X,Y} = C_Y \, r^*$$

hence

$$\lambda_0(X) = \sum_{i,j} \frac{\beta_j}{\alpha_i} \langle v_i, r(w_j) \rangle_H \langle X, v_i \rangle_G \, w_j.$$

Bayesian estimator

Modification of notation: (X, Θ) gaussian in $G \times H$, τ prior distribution for Θ, then the Bayesian estimator of θ is

$$E(\Theta \,|\, X) = \sum_{i,j} \frac{E(\langle X, v_i \rangle_G \langle \Theta, w_j \rangle_H)}{E(\langle X, v_i \rangle_G^2)} \langle X, v_i \rangle_G \, w_j$$

- Existence of density not required,
- G (resp. H) may be finite or infinite dimensional.

Bayesian estimator

Modification of notation: (X, Θ) gaussian in $G \times H$, τ prior distribution for Θ, then the Bayesian estimator of θ is

$$E(\Theta \mid X) = \sum_{i,j} \frac{E(\langle X, v_i \rangle_G \langle \Theta, w_j \rangle_H)}{E(\langle X, v_i \rangle_G^2)} \langle X, v_i \rangle_G \, w_j$$

- Existence of density not required,
- G (resp. H) may be finite or infinite dimensional.

Bayesian estimator

Modification of notation: (X,Θ) gaussian in $G \times H$, τ prior distribution for Θ, then the Bayesian estimator of θ is

$$E(\Theta\,|\,X) = \sum_{i,j} \frac{E(\langle X, v_i\rangle_G \langle \Theta, w_j\rangle_H)}{E(\langle X, v_i\rangle_G^2)} \langle X, v_i\rangle_G \, w_j$$

- Existence of density not required,
- G (resp. H) may be finite or infinite dimensional.

Tensorial product

Assume that (X, Y) is gaussian and such that

$$E(Y|X) = l_0(X)$$

where $l_0 \in \mathscr{L}(G, H)$. Thus

$$Y = l_0(X) + \eta$$

where η is strongly orthogonal to X. Then the tensorial product $Y \otimes Y$ has conditional expectation

$$E(Y \otimes Y|X) = l_0(X) \otimes l_0(X) + C_\eta,$$

with

$$l_0(X) = \sum_{i \in I} \frac{\langle X, v_i \rangle_G}{\alpha_i} C_{X,Y}(v_i).$$

Consider a sample (X_i, Y_i), $1 \leq i \leq n$ and suppose that X_{n+1} is observed. In order to "estimate" $\lambda_0(X_{n+1})$ the following steps are necessary

- Compute the empirical eigenvectors and eigenvalues from

$$C_{n,X} = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i$$

and

$$C_{n,Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \otimes Y_i$$

- Choose a double truncation index
- Find a doctoral student for the calculations.

Consider a sample (X_i, Y_i), $1 \leq i \leq n$ and suppose that X_{n+1} is observed. In order to "estimate" $\lambda_0(X_{n+1})$ the following steps are necessary

- Compute the empirical eigenvectors and eigenvalues from

$$C_{n,X} = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i$$

  and

$$C_{n,Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \otimes Y_i$$

- Choose a double truncation index
- Find a doctoral student for the calculations.

Consider a sample (X_i, Y_i), $1 \leq i \leq n$ and suppose that X_{n+1} is observed. In order to "estimate" $\lambda_0(X_{n+1})$ the following steps are necessary

- Compute the empirical eigenvectors and eigenvalues from

$$C_{n,X} = \frac{1}{n}\sum_{i=1}^{n} X_i \otimes X_i$$

and

$$C_{n,Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \otimes Y_i$$

- Choose a double truncation index
- Find a doctoral student for the calculations.

Consider a sample (X_i, Y_i), $1 \leq i \leq n$ and suppose that X_{n+1} is observed. In order to "estimate" $\lambda_0(X_{n+1})$ the following steps are necessary

- Compute the empirical eigenvectors and eigenvalues from

$$C_{n,X} = \frac{1}{n} \sum_{i=1}^{n} X_i \otimes X_i$$

and

$$C_{n,Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \otimes Y_i$$

- Choose a double truncation index
- Find a doctoral student for the calculations.

References I

Mandelbaum, A.
Linear estimators and measurable linear transformations on a Hilbert space.
Zfw, 65, 385-397, 1984.

Ruiz-Medina, M.D., R. Salmerón, Angulo, J.M.
Kalman filtering from PoP-based diagonalization of ARH(1).
Comp. Stat. Data Anal., 51, 4994-5008, 2007.

Hallin, M.
Talk, UPMC, 2012.

Cuevas, A.
A partial overview of the theory of statistics with functional data.
Preprint, 48 pages, 2012.

References II

Bosq, D. and Blanke, D.
Inference and prediction in large dimensions.
Wiley-Dunod, Chichester, 2007.

Bosq, D.
General linear processes in Hilbert spaces and prediction.
J.S.P.I., 137, 879-894, 2007.

Bosq, D.
Tensorial products of functional ARMA processes.
J.M.V.A., 101, 1352-1363.

Bosq, D.
Constructing functional linear filters.
Submitted, 23 pages, 2012.

Bosq, D. and Mourid, T.
In preparation, 2012.