Stochastic signals and systems

Lecture 2.

Prediction, innovation
and
the Wold decomposition
Version 2
September 27, 2020

2020.

REMINDER

Wide sense stationary processes

Random variables: defined over a probability space $(\Omega, \mathcal{F}, P)$.

A discrete time stochastic process: $y = (y_n)$, with $-\infty < n < +\infty$.

$y = (y_n)$ is wide sense stationary, w.s.st. for short, if

$$\mathrm{E} y_n = 0, \qquad \mathrm{E}(y_n^2) < +\infty$$

and for any fixed τ the auto-covariances below are independent of n:

$$r(\tau) = \mathrm{Cov}(y_{n+\tau}, y_n).$$

The auto-covariance matrix R

Let $\quad R := (R_{j,k}) = r(k-j) \qquad k, l = 1, \ldots, p.$

For any p he auto-covariance matrix R is a symmetric **Toeplitz matrix**:

$$R = \begin{pmatrix} r(0) & r(1) & r(2) & \ldots & r(p-1) \\ r(-1) & r(0) & r(1) & \ldots & r(p-2) \\ r(-2) & r(-1) & r(0) & \ldots & r(p-`3) \\ \vdots & \vdots & \ddots & \ddots & r(1) \\ r(-p+1) & r(-p+2) & \ldots & r(-1) & r(0) \end{pmatrix}$$

Note that the R is a Toeplitz matrix.

Recall: set $Y = (y_{n-1}, \ldots, y_{n-p})^T$. Show that $R = \mathbb{E}(YY^T) \geq 0$.

Prediction based on finite past

Predict y_n based on $y_{n-1}, \ldots, y_{n-p}$. Let $r := (r(1), \ldots, r(p))^T$.

Proposition. If R is nonsingular, then the LSQ linear prediction of y_n in terms of $y_{n-1}, \ldots, y_{n-p}$ is uniquely defined as

$$\widehat{y}_{n,n-p} := \sum_{k=1}^{p} \alpha_k y_{n-k},$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)^T$ is the solution of the normal equation

$$R\alpha = r.$$

Exercise. Show that the prediction $\widehat{y}_{n,n-p}$ is a w.s.st. process.

Question: what happens if we start increasing p and $p \to \infty$?

Linear operations on orthogonal processes

Let (e_n) be a real-valude w.s.st. orthogonal process and let us define

$$y_n = c_1 e_{n-1} + \cdots + c_p e_{n-p}, \qquad \text{with } c_k \text{ real.}$$

The process (y_n) is a moving average process. It is w.s.st. and

$$\mathbb{E}\, y_n^2 = \sum_{k=1}^{p} c_k^2 \sigma^2(e).$$

What happens if $p \to \infty$?

Exercise. Show that the process

$$y_n = \sum_{k=1}^{\infty} c_k e_{n-k} \quad \text{under} \quad \sum_{k=1}^{\infty} c_k^2 < \infty$$

is well-defined and it is a w.s.st. process. (*Hint*: See next exercise).

Exercise. Let $(v_{n,p})$ be a w.s.st. process for all p, and let $v_n = \lim_p v_{n,p}$ exist for all n. Show that (v_n) is w.s.st.

PREDICTION

Prediction based on the infinite past, I.

Consider the apparently impractical prediction problem with $p = \infty$.

Consider the infinite dimensional linear space spanned by $(y_{n-1}, y_{n-2} \ldots)$:

$$\mathcal{L}_{n-1} = \left\{ u : \sum_{k=0}^{p} \alpha_k y_{n-k}, \quad \alpha_k \in \mathbb{R} \quad \text{for some } p \right\}$$

Let H_{n-1} be the closure of $\mathcal{L}_{n-1}$ in the metric of $L_2(\Omega, \mathcal{F}, P)$:

$$H_{n-1} := \operatorname{cl} \mathcal{L}_{n-1}$$

Then the one-step ahead LSQ prediction of y_n is given by

$$\widehat{y}_n = (y_n | H_{n-1}),$$

the orthogonal projection of y_n on the subspace H_{n-1}.

Prediction based on the infinite past, II.

Consider now the linear space spanned by $(y_{n-1}, \dots, y_{n-p})$:

$$\mathcal{L}_{n-1,n-p} = \left\{ \sum_{k=1}^{p} \alpha_k y_{n-k} \;:\; \alpha_1, \dots, \alpha_p \in \mathbb{R}, \right\}.$$

Obviosuly, $\mathcal{L}_{n-1,n-p}$ is a finite dimensional subspace of H_{n-1}.

Hence it is a Hilbert subspace of H_{n-1}. Write $\mathcal{L}_{n-1,n-p} =: H_{n-1,n-p}$.

Proposition

We have, with convergence meant in $L_2(\Omega, \mathcal{F}, P)$,

$$(y_n | H_{n-1}) = \lim_{p \to \infty} (y_n | H_{n-1,n-p}).$$

Prediction based on finite past approximates $(y_n | H_{n-1})$ arbitrarily well.

Remarks on Hilbert space geometry, I.

In a more general setting: let G be a Hilbert space and let $G_p \subset G$ be a sequence of closed sub-spaces of G s.t. $G_p \subset G_{p+1}$ with $-\infty < p < +\infty$. I.e. (G_p) is monotone increasing. Let

$$G_\infty = \mathrm{cl}\,(\cup_p G_p).$$

with cl denoting the closure.

Lemma

Under the conditions above for any $x \in G$ we have

$$\lim_{p \to \infty} (x|G_p) = (x|G_\infty).$$

Apply: $G_p = H_{n-1,n-p}$ and $G_\infty = H_{n-1}$. Projection on expanding past.

Remarks on Hilbert space geometry, II.

Proof. First we show, replacing p by n, that for any $x \in G$

$$(x \,|\, G_n) = ((x \,|\, G_\infty) \,|\, G_n). \tag{1}$$

Cf. a projection on the (x, y) plane followed by projection to the x axis.

Indeed, let $(x|G_\infty) = y$, and take an orthogonal decomposition

$$x = y + \Delta x, \tag{2}$$

where $\Delta x \perp G_\infty$. Then also $\Delta x \perp G_n$ for any n.

Projecting both sides of the equality (2) onto G_n we get

$$(x|G_n) = (y|G_n).$$

Remarks on Hilbert space geometry, III.

Repeat: projecting both sides of the equality (2) onto G_n we get

$$(x|G_n) = (y|G_n).$$

But $y \in G_\infty = \mathrm{cl}\,(\cup G_n)$ implies that

$$\lim_{n\to\infty} (y - (y|G_n)) = 0.$$

Substituting $y = (x|G_\infty)$ and $(y|G_n) = (x|G_n)$ we get the claim:

$$\lim_{n\to\infty} ((x|G_\infty) - (x|G_n)) = 0.$$

A dual result, I.

Let G be and $G_m \subset G$ be as above, so that $G_m \subset G_{m+1}$, and let

$$G_{-\infty} = \cap_m G_{-m}.$$

Recall that $-\infty < m < +\infty$. Note that $G_{-\infty}$ is automatically closed.

Lemma

Under the conditions above for any $x \in G$ we have

$$\lim_{m \to \infty} (x|G_{-m}) = (x|G_{-\infty}).$$

Apply with $G = L_2(\Omega, \mathcal{F}, P)$, and $G_m = H_m$. Projection on shrinking past:

$$\lim_{p \to \infty} (x|H_{n-p}) = (x|H_{-\infty}). \tag{3}$$

Proof of the dual result

Let us denote the orthogonal complement of G_{-m} in G by $G^{\perp}_{-m}$, i.e.

$$G = G_{-m} \oplus G^{\perp}_{-m}.$$

From here we get

$$(x|G_{-m}) = x - (x|G^{\perp}_{-m}).$$

Obviously, $G^{\perp}_{-m}$ is monotone increasing. Let $F = \mathrm{cl}\,(\cup_m G^{\perp}_{-m})$. Then

$$\lim_{m\to\infty}(x|G^{\perp}_{-m}) = (x|F),$$

by Lemma 1. The lemma then follows from the next exercise.

Excercise. Show that $F = G^{\perp}_{-\infty}$. (*Hint*: Verify inclusion both way).

The innovation process, I.

Question: express the new information in y_n not contained in H_{n-1}

This is a key object in the theory of w.s.st. processes given below.

From now on we will use superscripts such as H^y_{n-1} for H_{n-1}.

Definition

The innovation process of (y_n) *is defined as:*

$$e_n := y_n - (y_n | H^y_{n-1}).$$

The innovation process, II.

Define the prediction error process based on finite past

$$e_{n,n-p} := y_n - (y_n | H_{n-1,n-p}).$$

Exercise. Prove that (e_n) is a w.s.st. process. (*Hint*: Recall that $\lim_{p\to\infty} e_{n,n-p} = e_n$.).

Exercise HW Prove that (e_n) is a w.s.st. orthogonal process.

Autoregressive processes, I.

Assume that a **finite segment** of past values is sufficient to compute $\widehat{y}_n$:

$$\widehat{y}_n = (y_n | H_{n-1}^y) = \sum_{k=1}^{p} a_k y_{n-k}.$$

Noting that $e_n = y_n - \widehat{y}_n$, we get

$$y_n = \sum_{k=1}^{p} a_k y_{n-k} + e_n.$$

Autoregressive processes II.

Repeat:

$$y_n = \sum_{k=1}^{p} a_k y_{n-k} + e_n. \tag{4}$$

Definition

A wide-sense stationary process $y = (y_n)$ satisfying (4), with (e_n) being its innovation process, is called an autoregressive or AR process.

If $a_p \neq 0$ then p is the **order** of the process, and y is an $\mathrm{AR}(p)$ process.

Q: Under what conditions on the coefficients a_k does such a y exist ?

WOLD DECOMPOSITION

Wold decomposition, I.

Let us consider a process $y = (y_n)$ and write

$$y_n = e_n + (y_n | H^y_{n-1}).$$

Obviously, e_n and H^y_{n-1} span H^y_n, and $e_n \perp H^y_{n-1}$. Thus we have

$$H^y_n = H(e_n) \oplus H^y_{n-1}$$

$\forall n$, where $H(e_n)$ denotes the 1-dimensional linear space spanned by e_n, assuming $e_n \neq 0$, and $\oplus$ denotes orthogonal direct sum of sub-spaces.

Consider now $(y_n | H^y_{n-1})$. It is an element of H^y_{n-1}. Hence we can decompose it as a sum of orthogonal elements using

$$H^y_{n-1} = H(e_{n-1}) \oplus H^y_{n-2}.$$

Wold decomposition, II.

Repeat: decompose $(y_n|H^y_{n-1})$ a sum of orthogonal elements using

$$H^y_{n-1} = H(e_{n-1}) \oplus H^y_{n-2}.$$

Note shift 1 time unit back. Now, $H(e_{n-1})$ is 1-dimensional, hence for any $v \in H(e_{n-1})$ we have $v = c\, e_{n-1}$ with some $c \in \mathbb{R}$.

Recall the identity on projecting on inclusive sub-spaces $H^y_{n-1} \supset H^y_{n-2}$:

$$((y_n|H^y_{n-1})|H^y_{n-2}) = (y_n|H^y_{n-2}).$$

Combining the two remarks above we get the decomposition

$$(y_n|H^y_{n-1}) = c_1 e_{n-1} + (y_n|H^y_{n-2}).$$

Wold decomposition, III.

Repeat: we obtain the following decomposition for $(y_n|H^y_{n-1})$:

$$(y_n|H^y_{n-1}) = c_1 e_{n-1} + (y_n|H^y_{n-2}).$$

Combining this with the definition of the innovation e_n we get

$$y_n = e_n + c_1 e_{n-1} + (y_n|H^y_{n-2}).$$

Iterating this decomposition we get, with $c_0 = 1,$

$$y_n = \sum_{k=0}^{p} c_k e_{n-k} + (y_n|H^y_{n-p-1}).$$

Wold decomposition, IV.

Repeat: iterating the one-step decomposition we get, with $c_0 = 1$,

$$y_n = \sum_{k=0}^{p} c_k e_{n-k} + (y_n | H^y_{n-p-1}).$$

To deal with the residual term define the prehistory of y as:

$$H^y_{-\infty} = \bigcap_{m \geq 0} H^y_{-m}.$$

Recall that by Lemma 1 we have, with convergence in $= L_2(\Omega, \mathcal{F}, P)$,

$$\lim_{m \to \infty} (y_n | H^y_{-m}) = (y_n | H^y_{-\infty}).$$

It follows that $\sum_{k=0}^{p} c_k e_{n-k}$ also converges when $p \to \infty$.

Wold decomposition, V.

Letting $p \to \infty$ we arrive at the following decomposition:

$$y_n = \sum_{k=0}^{\infty} c_k e_{n-k} + (y_n | H_{-\infty}^y). \tag{5}$$

Exercise. Show that $\sum_{k=0}^{\infty} c_k^2 < \infty$.

Consider the r.h.s. of (5) and define the processes

$$y_n^r = \sum_{k=0}^{\infty} c_k e_{n-k} \qquad \text{and} \qquad y_n^s = (y_n | H_{-\infty}^y).$$

Exercise. Show that both (y_n^r) and (y_n^s) are w.s.st. processes. (*Hint*: recall that $y_n^s = \lim_{p\to\infty} (y_n | H_{n-p})$).

Wold decomposition, VI.

Claim: the processes (y_n^r) and (y_n^s) are **orthogonal**, $y^s \perp y^r$, meaning:

$$y_n^r \perp y_m^s \qquad \text{for all} \quad n, \ m.$$

Proof: For any fixed p we have

$$\sum_{k=0}^{p} c_k e_{n-k} \perp H_{n-m}^y \quad \text{for} \quad m > p.$$

But $H_{n-m}^y \supset H_{-\infty}^y$, hence we get for any p:

$$\sum_{k=0}^{p} c_k e_{n-k} \perp H_{-\infty}^y.$$

Letting $p \to \infty$ we get the claim.

Wold decomposition, VII.

What can we say about $y_n^s = (y_n | H_{-\infty}^y)$?

Proposition

The history $H_n^{y^s} =: H_n^s$ is independent of n, hence $H_n^s = H_{n-1}^s$ for all n.

Proof: Fix n. Note that $e_n \perp H_{-\infty}^y$. (Why ?) But then the known equality $H_n^y = H(e_n) \oplus H_{n-1}^y$, implies, by projecting all three terms onto $H_{-\infty}^y$,

$$(H_n^y | H_{-\infty}^y) = (H_{n-1}^y | H_{-\infty}^y).$$

To end the proof it is sufficient to show that for any m

$$(H_m^y | H_{-\infty}^y) = H_m^s.$$

Wold decomposition, VIII.

Repeat: to end the proof it is sufficient to show that for any m

$$(H_m^y | H_{-\infty}^y) = H_m^s.$$

Recall the definition of H_m^y :

$$H_m^y = \mathrm{cl}\ \mathcal{L}_m^y = \left\{ u : \sum_{k=0}^{p} \alpha_k y_{m-k}, \quad \alpha_k \in \mathbb{R} \quad \text{for some } p \right\}.$$

Projecting all the elements of $\mathcal{L}_m^y$ onto $H_{-\infty}^y$ we get, due to linearity:

$$(\mathcal{L}_m^y | H_{-\infty}^y) = \left\{ u : \sum_{k=0}^{p} \alpha_k y_{m-k}^s, \quad \alpha_k \in \mathbb{R} \quad \text{for some } p \right\} = \mathcal{L}_m^s\,!$$

Take the closure of both sides. On the r.h.s we get $\mathrm{cl}\ \mathcal{L}_m^s = H_m^s$.

On the l.h.s. we get by the continuity of projections:

$$\mathrm{cl}\ (\mathcal{L}_m^y | H_{-\infty}^y) = (\mathrm{cl}\ \mathcal{L}_m^y | H_{-\infty}^y) = (H_m^y | H_{-\infty}^y), \quad \text{q.e.d.}$$

SINGULAR PROCESSES

Singular processes, I.

Let us have a closer look at $y_n^s = (y_n | H_{-\infty}^y)$. We have seen that for all n.

$$H_n^s = H_{n-1}^s.$$

Thus we get that the innovation process of y_n^s is identically 0:

$$e_n^s := y_n^s - (y_n^s | H_{n-1}^s) = 0\,!$$

Exercise. Verify the above equality. (*Hint*: Note that $y_n^s \in H_n^s$).

Intuition: a truly random process must have non-trivial innovations.

Definition

A w.s.st. stochastic process (v_n) is called singular, if its innovation process is identically 0: $\quad v_n - (v_n | H_{n-1}^v) = 0$ *for all n.*

Singular processes, II.

Exercise. Show that if $H_n^y = H_{n-1}^y$ for a single n then it is true for all n.

How can we construct a singular process? Example: consider the process

$$y_n = \xi e^{in\omega}, \qquad -\infty < n < +\infty,$$

where $\omega \in (0, 2\pi)$ is a fixed frequency, ξ is a complex-valued r.v. with

$$\mathrm{E}\xi = 0, \quad \mathrm{E}|\xi|^2 = \sigma^2 < +\infty.$$

Exercise. Show that the complex-valued process $y = (y_n)$ is w.s.st.:

$$r(\tau) := \mathrm{E} y_{n+\tau} \overline{y_n} = \sigma^2 e^{i\tau\omega}.$$

Note that $r(\tau) = \sigma^2 e^{i\tau\omega}$ does not decay in absolute value, as τ increases.

Singular processes III.

Consider now a finite sum of the above complex-valued w.s.st. processes:

$$y_n = \sum_{k=1}^{m} \xi_k e^{in\omega_k}. \tag{6}$$

Here $\omega_k \neq \omega_j$ for $k \neq j$, and

$$\mathrm{E}\xi_k = 0, \qquad \mathrm{E}|\xi_k|^2 = \sigma_k^2 < +\infty.$$

Assume in addition that $\xi_j \perp \xi_k$ for $j \neq k$, i.e.

$$\mathbb{E}\xi_j\bar{\xi}_k = 0, \quad \text{for} \quad k \neq j.$$

Singular processes, IV.

Repeat: consider the process $y = (y_n)$ given by $y_n = \sum_{k=1}^{m} \xi_k e^{in\omega_k}$.

Exercise. Show that the above complex-valued process y is w.s.st.:

$$r(\tau) := \mathbb{E} y_{n+\tau} \bar{y}_n = \sum_{k=1}^{m} \sigma_k^2 e^{i\tau\omega_k}.$$

The variance of y_n is obtained by setting $\tau = 0$:

$$\mathrm{E}|y_n|^2 = \sum_{k=1}^{n} \sigma_k^2.$$

In telecommunication $\mathbb{E}|y_n|^2$ is the energy of the signal.

The values σ_k^2 show how the energy of y_n is spread among frequencies.

Singular processes, V.

Proposition

The complex-valued process $y_n = \sum_{k=1}^{m} \xi_k e^{in\omega_k}$, *given above, is singular.*

We note in passing: all the arguments of the present lecture carry over to complex-valued processes with evident modifications.

An elegant way to establish singularity is given by the following exercise:

Exercise. Show that if $\dim(H_n^y) < \infty$ for some n, then y is singular.

(*Hint:* Prove that if y is not singular, then $\dim(H_n^y) = \infty$ for all n.)

Exercise. Using the last exercise prove the proposition above.

Singular processes, VI.

How can we construct a real-valued singular process?

Let us consider the complex-valued w.s.st. process $y = (y_n)$. Let

$$v_n := \operatorname{Re} y_n.$$

Now H_n^y is a subspace of $L_2^c(\Omega, \mathcal{F}, \mathcal{P})$, the set of complex-valued random variables ξ with $\mathbb{E}|\xi|^2 < \infty$. And $H_n^v \subset L^2(\Omega, \mathcal{F}, \mathcal{P})$.

Exercise. Prove that $\dim(H_n^y) < \infty$ for some n implies $\dim(H_n^v) < \infty$.

Singular processes, VII.

Example: Consider the w.s.st. processes as above:

$$y_n = \sum_{k=1}^{m} \xi_k e^{in\omega_k}. \quad \text{with} \quad \xi_k = \sigma_k e^{i\varphi_k}, \tag{7}$$

where the phases φ_k have uniform distribution on $[0, 2\pi]$, and are i.i.d.

We conclude that the real part of the process is given by

$$v_n = \sum_{k=1}^{m} \sigma_k \cos(in\omega_k + \varphi_k).$$

is singular. A simple special case for a real-valued singular process is:

$$y_n = \cos(\omega n + \varphi) \qquad \omega \neq 0,$$

where φ is a random phase with uniform distribution on $[0, 2\pi]$.