Basic Models & Methods

Basic Models & Methods

- **Introduction**

  We shall see some basic signal models used in the context of statistical signal processing. These are subdivided into two main classes: Parametric and non parametric models.

  Then we shall explore common statistical methods used to extract the information from signals, namely estimation and prediction. The way these methods are implemented depends on the type of signal model considered.

  Consequently, we will present estimation methods for parametric and non parametric signal models, and prediction methods for parametric and non parametric signal models.

Models

Non Parametric & Parametric

Types of Signal Models

- **Non Parametric Signal Models**

  We suppose the signal to have specific characteristics / properties but these cannot be explicitly represented by a finite set of parameters (or a function depending on them).

- **Parametric Signal Models** We suppose the signal to have specific characteristics / properties and these can be represented by a finite set of parameters (or a function depending on them).

Non Parametric Signal Models

- **Signal with i.i.d. behaviour**

  The signal is modelled as a discrete time stochastic process $\{X[n]\}_{n\in\mathbb{Z}}$ where the random variables $X[n]$ are independent and identically distributed.

  The most famous i.i.d. signal is probably the **White Noise**.

Non Parametric Signal Models

- **Wide Sense Stationary (w.s.s.) Processes**

  A discrete time stochastic process $\{X[n]\}_{n\in\mathbb{Z}}$ such that
  - $\mathrm{E}[X[n]]$ =constant, independent of n;
  - $\mathrm{E}\Big[X[k]\,X^*[l]\Big] = R(k-l)$ and $\mathrm{cov}\Big(X[k], X[l]\Big) = \Gamma(k-l),\ \forall\ k,l \in \mathbb{Z}$.

Property

w.s.s. processes admit a Power Spectral Density defined as

$$S_X(\omega) = \sum_{k=-\infty}^{\infty} R_X(k)\, e^{-i\omega k},$$

(if the correlation R_X is summable).

Parametric Signal Models

- **Parametric Law**

  The law of the process is described by a finite set of parameters.

- **Parametric Law: Gaussian Process**

  The most famous parametric signal is probably the Gaussian w.s.s. process.

  Call $\{X[n]\}_{n\in\mathbb{Z}}$ the process. Let $\boldsymbol{X}_N = [X[1], \ldots, X[N]]$ be the vector of N samples, $\boldsymbol{x}_N = [x_1, \ldots, x_N]$ be the vector of N sample values, and $\boldsymbol{m} = [m_1, \ldots, m_N]$ the vector of the probabilistic means.

  Assume a limited correlation, *e.g.*, $\mathrm{E}[X[n]X[m]] = \mathrm{E}[X[n]]\,\mathrm{E}[X[m]]$ for $|m-n| > K$.

  Then, the probability density is given by

$$f_{X[1],\ldots,X[N]}(x_1, \ldots, x_n) = \sqrt{\frac{1}{(2\pi)^N \det(\boldsymbol{C}_{X\,N})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_N - \boldsymbol{m}_N)\boldsymbol{C}_{X\,N}^{-1}(\boldsymbol{x}_N - \boldsymbol{m}_N)^t\right\}$$

  where (case $K = 2$)

$$\boldsymbol{C}_{X\,N} = \begin{pmatrix} \sigma^2 & C_{12} & C_{13} & 0 & 0 \\ C_{21} & \sigma^2 & C_{12} & C_{13} & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & C_{31} & C_{21} & \sigma^2 & C_{12} \\ 0 & 0 & C_{31} & C_{21} & \sigma^2 \end{pmatrix}, \qquad \begin{aligned} C_{ij} &= \mathrm{E}[X[n]X[m]] - \mathrm{E}[X[n]]\,\mathrm{E}[X[m]]\,, \\ \sigma^2 &= C_{ii}\,. \end{aligned}$$

Parametric Signal Models

- **Parametric Law: Stationary Markov Chain**

  Stochastic process $\{X[n]\}_{n\in\mathbb{Z}}$ (that we consider stationary) taking discrete values (in a set $\mathcal{D}$), such that

  $$\mathrm{P}\left(X[n] = i_n \mid X[n-1] = i_{n-1}, X[n-2] = i_{n-2}, \ldots\right) = \mathrm{P}\left(X[n] = i_n \mid X[n-1] = i_{n-1}\right) ,$$

  for $i_n, i_{n-1}, i_{n-2}, \ldots \in \mathcal{D}$.

  It is a process with a dependence limited to the nearest past.

  The process is completely characterized by:

  - **Transition probabilities** $\mathrm{P}\left(X[n] = i \mid X[n-1] = k\right) = p_{ki}$, ;
  - **Initial probabilities** $\mathrm{P}\left(X[n] = k\right) = \pi_k$

  A Markov chain $\{X[n]\}_{n\in\mathbb{Z}}$ can be generally expressed as

  $$X[n] = f(X[n-1], Z[n]) .$$

  where $Z[n]$ is a sequence of i.i.d. random variables independent of $X[n-1], X[n-2], \ldots$. $Z[n]$ represent the **innovation**.

Parametric Signal Models

- **Parametric Law: Stationary Markov Chain**

  Key tool to work with Markov chains: **Bayes' rule**.

  Given two random variables A and B with discrete values in $\mathcal{D}$, we have

  $$\mathrm{P}\left(A = k \mid B = l\right) = \frac{\mathrm{P}\left(A = k\,,\ B = l\right)}{\mathrm{P}\left(B = l\right)}\,, \quad k, l \in \mathcal{D}\,,$$

  supposing $\mathrm{P}\left(B = l\right) > 0$.

  Indeed, let $X[n]$ be a Markov chain. Using the Baye's rule, its probability (law of the process) $\mathrm{P}\left(X[n] = i_n, \ldots, X[n-k] = i_{n-k}\right)$, $i_n, \ldots, i_{n-k} \in \mathcal{D}$, can be expressed as transition probabilities p_{lj} and initial probabilities π_i

  $$\mathrm{P}\left(X[n] = i_n, \ldots, X[n-k] = i_{n-k}\right) = p_{i_{n-1} i_n} \ldots p_{i_{n-k} i_{n-k+1}} \pi_{i_{n-k}}\,, \quad i_n, \ldots, i_{n-k} \in \mathcal{D}\,.$$

Remark
Notice that Baye's rule is defined for probabilities (or cumulative distribution functions).

Parametric Signal Models

- **AutoRegressive (AR) Processes: Definition**

  An AR process is a w.s.s. process $X[n]$, taking continuous values in $\mathbb{R}$, satisfying the equation

  $$\sum_{k=0}^{M} p_k X[n-k] = W[n], \quad n \in \mathbb{Z},$$

  - $W[n]$, is a zero mean Gaussian white noise, with variance σ_W^2;
  - p_k, $k = 0, \ldots, M$ sequences of bounded coefficients (real or complex). Without loss of generality, we will assume $p_0 = 1$.

  The AR process is said to be of order M and it is characterized by the parameters $p_1, \ldots, p_M$, and σ_W^2.

  An AR of order 1

  $$X[n] + p_1 X[n-1] = W[n],$$

  is a particular case of continuous valued process with Markovian properties.

Parametric Signal Models

- **AutoRegressive (AR) Processes: Filtering Interpretation**

$$\sum_{k=0}^{M-1} p_k X[n-k] = W[n], \quad n \in \mathbb{Z},$$

Left term = filtering of $X[n]$ by a filter with impulse response $h[l] = p_l$, $l = 0, \ldots, M-1$ and z transform

$$P(z) = \sum_{k=0}^{M-1} p_k z^{-k}.$$

By interpreting z^{-1} as the delay operator, we have (**symbolically**)

$$P(z)\, X[n] = W[n].$$

When $P(z)$ is minimum phase with $p_0 = 1$ the AR is in its **canonical form**.

Remark

Such a symbolic notation is very handy to define AR processes and we will adopt it quite often. However, is has to be used carefully, by considering z^{-1} as a delay operator (and not by mixing time domain and the z-transform domain)!

Parametric Signal Models

- **AutoRegressive (AR) Processes: Synthesis and Analysis**

  When dealing with AR processes two problems naturally arise:

  - Generate an AR process given a white noise $W[n]$ (**synthesis problem**)
  - Check if $X[n]$ corresponds to a given AR process (**analysis problem**)

Parametric Signal Models

- **AutoRegressive (AR) Processes: Synthesis**

  $P(z)$ minimum phase filter, then $1/P(z)$ defines a fractional causal and stable ℓ_1 (BIBO) filter. Applying such filter to both sides of the AR equation yields

$$X[n] = \sum_{l=0}^{\infty} h_S[l] W[n-l],$$

  where $h_S[l]$, $l \in \mathbb{N}$, is the impulse response of the filter

$$H_S(z) = \frac{1}{P(z)}.$$

  $H_S(z)$ is called **synthesis filter**: it allows to generate an AR process $X[n]$ given a white noise $W[n]$.

Remark

Since $1/P(z)$ is causal and ℓ_1 stable, it is an analytical function that admits a power series expansion: we call $h_S[l]$, $l \in \mathbb{N}$, the coefficients of such a series. In particular, one can check that $h_S[0] = 1$ (since $p_0 = 1$).

Parametric Signal Models

- **AutoRegressive (AR) Processes: Analysis**

  In the case of an AR process the analysis filter is straightforwardly given by $P(z)$.

  $P(z)$ allows to check if a process $X[n]$ corresponds to a given AR model by testing if the output of the filter is a white noise.

Parametric Signal Models

- **AutoRegressive (AR) Processes: Correlation Structure**

  We assume that the filter $P(z)$ is **strictly** minimum phase

  Let $H_S(z)$ be the synthesis filter of the AR process and $h_S[k]$ its impulse response. Then, the correlation of the process has the following structure

$$R_X[m] + \sum_{k=1}^{M-1} p_k R_X[m-k] = {h_S}^*[-m]\sigma_W^2, , \quad m \geq 0 .$$

  Since the synthesis filter is causal, *i.e.*, $h_S[l] = 0$, for $l < 0$,

$$R_X[m] + \sum_{k=1}^{M-1} p_k R_X[m-k] = \delta_m \sigma_W^2 , \quad m \geq 0 .$$

  The above equation is of great importance: milestone of **Yule Waker Equations**, which play an important role in prediction and parametric spectral estimation.

Parametric Signal Models

- **AutoRegressive (AR) Processes: Correlation Structure**

$$R_X[m] + \sum_{k=1}^{M-1} p_k R_X[m-k] = \delta_m \sigma_W^2 \,, \quad m \geq 0 \,.$$

Remark

To prove the result multiply both sides of the AR equation by $X^*[n-m]$ and compute the expectation; To obtain $R_X[m]$ for $m < 0$ just recall that $R_X[-m] = R_X[m]^*$.

Parametric Signal Models

- **AutoRegressive (AR) Processes: Power Spectrum**

  We assume that the filter $P(z)$ is **strictly** minimum phase

  The power spectrum of an AR process is given by

$$S_X(\omega) = \frac{1}{|P(e^{j\omega})|^2}\sigma_W^2 .$$

Remark

The simplest proof of the above result is obtained by considering the filtering interpretation of an AR process. Then, by applying the fundamental filtering formula to both sides of the AR equation we have

$$S_X(\omega)\left|P\left(e^{j\omega}\right)\right|^2 = \sigma_W^2 .$$

Parametric Signal Models

- **Matlab Practice: AutoRegressive (AR) Processes**

  Generate an M samples of an AR process

  $$X[n] = a_1 X[n-1] + \ldots + a_N X[n-N] + W[n]\,.$$

  where $W[n]$ is a centered Gaussian white noise with variance σ_W^2 (choose the order, the values of the coefficients, and the power of the noise), using the Matlab command

  `model = arima('Constant',0,'AR',`a_1`,`a_2`,...,`a_N`,'Variance',`σ_W^2`);`
  `x = simulate(model,M);`

  Compute the empirical correlation of the AR samples (be careful how you do it!) and compare it with the empirical correlation of a realization of M samples of $W[n]$.

Parametric Signal Models

- **Matlab Practice: Markov Chain**

  Generate M samples of a N state Markov chain, using the Matlab command

```
mc = dtmc(P);
x = simulate(mc,M);
```

  where P is the $N \times N$ transition matrix (row-wise probability distribution).

  Compute the empirical correlation of the Markov chain samples (be careful how you do it!) and compare it with the empirical correlation of a realization of M samples of a Gaussian white noise.

  You can also use the command `graphplot(mc,'ColorEdges',true)` to visualize the state transitions of the Markov chain.

Methods

Estimation & Prediction

- Linear Estimation & Prediction;
- Estimation of Parametric Probability Densities;
- Estimation of Most Probable Realization.

Estimation & Prediction

- **Estimation**

  Obtain a certain quantity, such as a process $X[n]$, or a parameter, from a sequence of observations $Y[k] = f(X[k])$, $k \in \mathbb{Z}$ (**does not imply causality**).

- **Prediction**

  Predict the evolution of a quantity, such as a process $X[n]$, given the values of its past $X[n-1], X[n-2], \ldots$ (**it implies causality**).

  Common scenarios: Predict the weather or the stock value of tomorrow from the weather or the stock values of the past days; Predict the evolution of a model based signals (*e.g.*, predictor of an AR process).

Estimation & Prediction

- **Optimality in the Mean Square Error - MSE sense**

  Given a signal modeled as a w.s.s. process $X[n]$, its estimator or predictor $Y[n]$ is chosen so to minimize the MSE $\mathrm{E}\left[|X[n]-Y[n]|^2\right]$ (minimum energy of the error).

- **MSE Minimization and the Projection Theorem**

  W.s.s. processes are elements of a Hilbert Space.

  The estimator or the predictor contains less information than the process itself (we do not know the process we just estimate or predict it ..).

  Therefore, if we call $\mathcal{S}$ the space of the process and $\mathcal{G}$ the space of the estimator or predictor, we can say $\mathcal{G} \subset \mathcal{S}$. In many situations the estimator or the predictor is nothing but the projection of the process $X[n]$ onto the space $\mathcal{G}$.

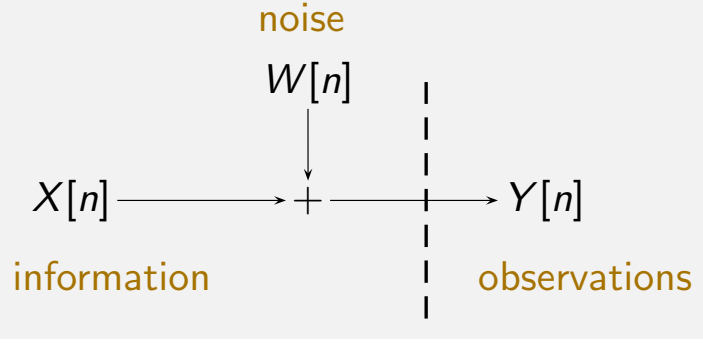
  By the projection theorem in Hilbert Spaces finding $Y[n]$ that minimizes $\mathrm{E}\left[|X[n]-Y[n]|^2\right]$ is equivalent to find $Y[n]$ such that $\mathrm{E}\left[(X[n]-Y[n])V[n]^*\right] = 0$ for every $V[n] \in \mathcal{G}$, that is for every $V[n]$ = linear combination of $Y[k]$, $k \in \mathbb{Z}$.

Linear Estimation of Non Parametric w.s.s. Models

- **Linear Estimation Based on Indirect Observations**

  Process $X[n]$ w.s.s. indirectly observed via the process $Y[$

  Typical scenario is a signal $X[n]$ corrupted by a white noise



$$Y[n] = X[n] + W[n]\,, \text{ or}$$

$$Y[n] = \sum_{k\in\mathbb{N}} g[k](X[n-k] + W[n-k])\,,$$

where:

- We assume that $X[n]$ and $Y[n]$ are jointly wide-sense stationary;
- We observe a realization $y[n]$ of $Y[n]$ (measurements);
- We estimate the corresponding realization $x[n]$ of $X[n]$ as a linear combination of the realization $y[n]$ of $Y[n]$, *i.e.*, using the model

$$\widehat{X}[n] = \sum_{k\in\mathbb{Z}} h_{n,k} Y[k]\,.$$

Notice that being *a priori* the linear combination of an infinite number of terms, its coefficients need to take the dependence w.r.t. to n into account.

Linear Estimation of Non Parametric w.s.s. Models

- **Linear Estimation as Filtering**

  Linearly estimate $X[n]$ given that we observe $Y[k]$, $k \in \mathbb{Z}$.

  $$\widehat{X}[n] = \sum_{k \in \mathbb{Z}} h_{n,k} Y[k] \,.$$

  By exploiting the wide sense stationarity we have

  $$\widehat{X}[n] = \sum_{k \in \mathbb{Z}} h_{n-k} Y[k] \,,$$

  which enables a **filtering interpretation** of the linear estimator

  $$\widehat{X}[n] = \sum_{k \in \mathbb{Z}} h[n-k] Y[k] \,.$$

- **Wiener Filter**

  As best linear estimator (filter) of $X[n]$ based on $Y[k]$, $k \in \mathbb{Z}$, we choose the one that minimizes the mean square error

  $$\mathrm{E}\left[\left| X[n] - \widehat{X}_N[n] \right|^2 \right] .$$

  The filter that minimize the mean square error is called **Wiener Filter**.

Linear Estimation of Non Parametric w.s.s. Models

- **Wiener Filter: Computation**

  The linear estimator minimizing the MSE in the $\mathrm{L}^2(P)$ space of w.s.s. processes is given by the projection theorem by projecting $X[n]$ onto the Hilbert space generated by $\{Y[k]\}_{k\in\mathbb{Z}}$.

  Projection computed by directly minimizing the mean square error $\mathrm{E}\left[\left|X[n]-\widehat{X}[n]\right|^2\right]$, or by applying the **orthogonality** between the error and any $Y[k]$, *i.e.*,

$$\mathrm{E}\left[(X[n]-\widehat{X}[n])Y^*[l]\right]=0\,,\quad \forall l\,.$$

  Using the latter $\mathrm{E}\left[(X[n]-\widehat{X}[n])Y^*[l]\right]=R_{XY}[n-l]-\sum_{k\in\mathbb{Z}}h[n-k]R_Y[k-l]=0$, *i.e.*,

$$R_{XY}[n-l]=\sum_{k\in\mathbb{Z}}h[n-k]R_Y[k-l]\,,$$

  or, by changing variables ($m=n-k$, $k=n-m$, $n-l=u$),

$$R_{XY}[u]=\sum_{m\in\mathbb{Z}}h[m]R_Y[u-m]\,,$$

  These are are called **Wiener-Hopf** (or **normal**) **equations**.

Linear Estimation of Non Parametric w.s.s. Models

- **Wiener Filter: Solution of the Wiener-Hopf equations**

  The are basically two approaches for the solution of the Wiener-Hopf equations

  a) Considering a finite impulse response (FIR) filter.

  We assume $h[k]$ being defined for $k = 0, \ldots, K$, and zero otherwise. Then, the Wiener-Hopf equations can be written as

$$R_{XY}[n] = \sum_{k=0}^{K} h[k] R_Y[n-k], \quad \text{that is} \quad \boldsymbol{r}_{XY} = \boldsymbol{R}_Y \boldsymbol{h} \quad \text{where}$$

$$\boldsymbol{r}_{XY} = \begin{bmatrix} R_{XY}[0] \\ \vdots \\ R_{XY}[K] \end{bmatrix} \quad \boldsymbol{R}_Y = \begin{bmatrix} R_Y[0] & \ldots & R_Y[-K] \\ \vdots & \ddots & \vdots \\ R_Y[K] & \ldots & R_Y[0] \end{bmatrix},$$

  and therefore $\boldsymbol{h} = \boldsymbol{R}_Y^{-1} \boldsymbol{r}_{XY}$

  b) Computing the Power Spectral Densities.

  Fourier transforming both sides of the Wiener-Hopf equations yields

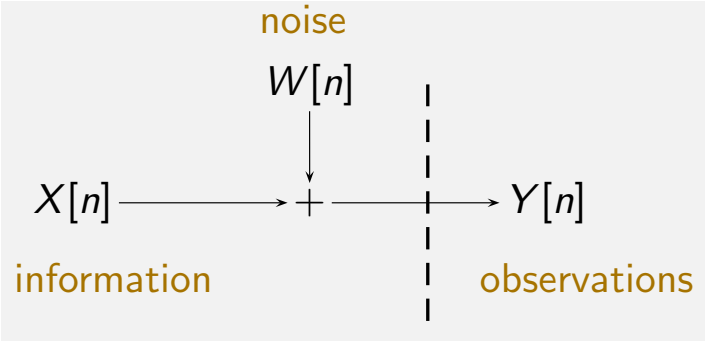
$$S_{XY}(\omega) = S_Y(\omega) H(e^{j\omega}), \quad \text{hence,} \quad H(e^{j\omega}) = \frac{S_{XY}(\omega)}{S_Y(\omega)},$$

  where $S_{XY}(\omega)$ is cross power spectral density between $X[n]$ and $Y[k]$ and $S_Y(\omega)$ is power spectral density of $Y[k]$.

Linear Estimation of Non Parametric w.s.s. Models

- **Wiener Filter: Signal Corrupted by Noise**

$$Y[n] = X[n] + W[n]\,,$$



Then, given the independence between X and W, the Wiener-Hopf equations read

$$R_X[n] = \sum_{k=0}^{K} h[k] R_Y[n-k]\,,$$

and:

a) The FIR filter is given by

$$\boldsymbol{h} = \boldsymbol{R}_Y^{-1} \boldsymbol{r}_X\,;$$

b) The Transfer function of the general filter is given by

$$H(e^{j\omega}) = \frac{S_X(\omega)}{S_Y(\omega)}\,.$$

Linear Estimation of Non Parametric w.s.s. Models

- **Wiener Filter: In Practice**

  While the theoretical derivation of the Wiener filter is very elegant, its practical use is quite limited. Indeed, by taking a closed look at the Wiener-Hopf equations

  $$R_{XY}[u] = \sum_{m \in \mathbb{Z}} h[m] R_Y[u-m],$$

  we can remark that:

  - Their solution requires the knowledge of the cross correlation between the signal we would like to reconstruct (in practice unknown) and the measured signal.
  - When considering the solution based on the power spectral de density, the estimates of the latter need to be chosen very wisely. Indeed they appear in a rational expression and estimation errors of $S_{XY}(\omega)$ and $S_Y(\omega)$ can have drastic effect on the computation of $H(e^{j\omega})$;

  In practice the Wiener filter given by the above equation is of no use for the reconstruction of the unknown signal $x[n]$, yet the **Wiener-Hopf** equations play a key role in

  - Adaptive filtering (as we shall see later);
  - System identification.

Linear Prediction of Non Parametric w.s.s. Models

- **Prediction as Orthogonal Projection**

  Predict the process w.s.s. $X[n]$ at the instant n, based on linear combination of a **finite** observation of its past,

$$\widehat{X}_N[n] = a_{1,N}X[n-1] + \ldots + a_{N,N}X[n-N]\,,$$

  $\widehat{X}_N[n]$ **linear predictor of order** N.

  We consider (as usual) best linear estimator in mean square sense $=$ orthogonal projection of $X[n]$ onto the Hilbert space generated by $X[n-1], \ldots, X[n-N]$.

Linear Prediction of Non Parametric w.s.s. Models

- **Error of The Orthogonal Projection (Prediction Error)**
  - The prediction error $\epsilon_N[n] = X[n] - \widehat{X}_N[n]$ is orthogonal to $X[n-1], \ldots, X[n-N]$

$$\mathrm{E}\left[\left(X[n] - \widehat{X}_N[n]\right) X[n-k]\right] = 0\,, \quad k = 1, \ldots, N\,.$$

  - The $\mathrm{L}^2(P)$ norm of the prediction error

$$\|\epsilon_N\|_2^2 = \mathrm{E}\left[\left|X[n] - \widehat{X}_N[n]\right|^2\right],$$

    is minimized.

Combining the above two equations with the predictor expression gives

$$\|\epsilon_N\|_2^2 = \mathrm{E}\left[\left(X[n] - \widehat{X}_N[n]\right) X[n]\right].$$

Linear Prediction of Non Parametric w.s.s. Models

- **Yule Walker Equations**

The projection theorem provides the best linear predictor in an algebraic way:

$$\textbf{Linear prediction eq.} \quad \widehat{X}_N[n] = a_{1,N} X[n-1] + \ldots + a_{N,N} X[n-N]\,,$$

$$+$$

$$\textbf{Orthogonality of the error} \quad \mathrm{E}\left[\left(X[n] - \widehat{X}_N[n]\right) X[n-k]\right] = 0\,, \quad k = 1, \ldots, N\,.$$

$$+$$

$$\textbf{Norm of the error} \quad \|\epsilon_N\|_2^2 = \mathrm{E}\left[\left(X[n] - \widehat{X}_N[n]\right) X[n]\right]\,.$$

$$=$$

System of $N+1$ equations with $N+1$ unknowns: $\|\epsilon_N\|_2^2$ and $a_{1,N}, \ldots, a_{N,N}$

Yule Walker equations

$$\begin{aligned}
\|\epsilon_N\|_2^2 + R_X[1]a_{1,N} + \ldots + R_X[N]a_{N,N} &= R_X[0] \\
R_X[0]a_{1,N} + \ldots + R_X[N-1]a_{N,N} &= R_X[1] \\
\vdots \qquad\qquad = \quad \vdots & \\
R_X[N-1]a_{1,N} + \ldots + R_X[0]a_{N,N} &= R_X[N]\,.
\end{aligned}$$

Linear Prediction of Non Parametric w.s.s. Models

- **Yule Walker Equations in Matrix Form**

$$\begin{aligned} \|\epsilon_N\|_2^2 + R_X[1]a_{1,N} + \ldots + R_X[N]a_{N,N} &= R_X[0] \\ R_X[0]a_{1,N} + \ldots + R_X[N-1]a_{N,N} &= R_X[1] \\ \vdots \qquad\qquad = \quad \vdots & \\ R_X[N-1]a_{1,N} + \ldots + R_X[0]a_{N,N} &= R_X[N] \end{aligned}$$

In matrix form, the N equations for coefficients estimation read $\boldsymbol{R}_X^N \boldsymbol{a}^{N,1} = \boldsymbol{R}_X^{N,1}$, where

$$\boldsymbol{R}_X^N = \begin{bmatrix} R_X[0] & R_X[1] & \cdots & R_X[N-1] \\ R_X[1] & R_X[0] & \cdots & R_X[N-2] \\ \vdots & & \ddots & \vdots \\ R_X[N-1] & R_X[N-2] & \cdots & R_X[0] \end{bmatrix}$$

$$\boldsymbol{a}^{N,1} = \begin{bmatrix} a_{1,N} \\ a_{2,N} \\ \vdots \\ a_{N,N} \end{bmatrix}, \quad \boldsymbol{R}_X^{N,1} = \begin{bmatrix} R_X[1] \\ R_X[2] \\ \vdots \\ R_X[N] \end{bmatrix}.$$

Linear Prediction of Non Parametric w.s.s. Models

- **Yule Walker Equations**

  Real scenario
  - I observe K samples $x[1], \ldots, x[K]$;
  - I compute the empirical correlation

$$\widehat{R}_X[k] = \alpha \sum_{l=1}^{K-k} x[l]x[l-k], \quad k = 1, \ldots, N < K,$$

Assignement

Should we use a biased or an unbiased correlation?

For next week, propose how to efficiently implement the Yule Walker equations so to reduce estimation issues like the one arising at extreme lags for the unbiased correlation, or the one arising when dealing with non zero mean signals and the biased correlation.

Linear Estimation AR Parametric Models

- **Yule Walker Equations (Once Again!)**

  Estimation of AR processes boils down into estimating the parameters of the process, that is, the coefficient of the linear regression and the variance of the white noise.

  Suppose you want to estimate the parameters of an AR model of order N

$$\sum_{k=0}^{N} p_k X[n-k] = W[n]\,,$$

  that is, $p_1, \ldots, p_N$, and σ_W^2.

  We are right within the framework of the Yule Walker equations. Indeed:

  - I observe K samples $x[1], \ldots, x[K]$ of the AR process of order N;
  - I compute the empirical correlation;
  - I exploit the correlation structure

$$R_X[m] + \sum_{n=1}^{N} p_n R_X[m-n] = {h_S}^*[-m]\sigma_W^2,, \quad m \geq 0\,,$$

  in order to write $N+1$ equations in the $N+1$ unknowns $p_1, \ldots, p_N$, and σ_W^2.

Linear Estimation of AR Parametric Models

- **Yule Walker Equations (Once Again!)**

  That is ...

$$\begin{aligned}
\sigma_W^2 + R_X[1]p_1 + \ldots + R_X[N]p_N &= R_X[0] \\
R_X[0]p_1 + \ldots + R_X[N-1]p_N &= R_X[1] \\
\vdots \qquad &= \quad \vdots \\
R_X[N-1]p_1 + \ldots + R_X[0]p_N &= R_X[N]
\end{aligned}$$

  In matrix form, the N equations for coefficients estimation read $\boldsymbol{R}_X^N \boldsymbol{p}^{N,1} = \boldsymbol{R}_X^{N,1}$, where

$$\boldsymbol{R}_X^N = \begin{bmatrix} R_X[0] & R_X[1] & \cdots & R_X[N-1] \\ R_X[1] & R_X[0] & \cdots & R_X[N-2] \\ \vdots & & \ddots & \vdots \\ R_X[N-1] & R_X[N-2] & \cdots & R_X[0] \end{bmatrix}$$

$$\boldsymbol{p}^{N,1} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix}, \quad \boldsymbol{R}_X^{N,1} = \begin{bmatrix} R_X[1] \\ R_X[2] \\ \vdots \\ R_X[N] \end{bmatrix}$$

Linear Estimation of AR Parametric Models

- **Recursive Solution of the Yule Walker Equations**

  The Yule Walker equations can efficiently solved iteratively, starting for a order 1 linear systems and then increasing the order.

  This is particularly interesting when the order of the process is not know: We keep increasing the model order until there is no significant change in the estimation error.

  The **Levinson's algorithm** enables to estimate the parameters of an AR of order N based on the parameters estimated for an AR of order N-1, with a significant computation burden reduction.

Linear Estimation of AR Parametric Models

- **Levinson's Algorithm: Backward and Forward Predictors**

  Forward predictor:
  Linear predictor of $X[n]$ based on the finite past $X[n-1], \ldots, X[n-N]$

$$\widehat{X}_N[n] = a_{1,N} X[n-1] + \ldots + a_{N,N} X[n-N]\,.$$

  Forward predictor error:

$$\epsilon_N[n] = X[n] - \widehat{X}_N[n]\,.$$

  Backward predictor:
  Linear predictor of $X[n]$ based on the finite future $X[n+1], \ldots, X[n+N]$

$$\widetilde{\widehat{X}}_N[n] = \widetilde{a}_{1,N} X[n+1] + \ldots + \widetilde{a}_{N,N} X[n+N]\,,$$

  Backward predictor error

$$\widetilde{\epsilon}_N[n] = X[n] - \widetilde{\widehat{X}}_N[n]$$

Linear Estimation of AR Parametric Models

- **Levinson's Algorithm: Backward and Forward Predictors**

  We can remark that by setting $Y[n] = X[-n]$, $\widetilde{\widehat{Y}}_N[n]$ is the forward predictor of the process $X[-n]$. Similarly, $\widehat{Y}_N[n]$ is its backward predictor.

  We shall consider two possible cases

  - $X[n]$ **is real:**
    $X[n]$ and $X[-n]$ have same correlation. Same solution of the Yule Waler equations for backward and forward predictor. Consequently $\widetilde{a}_{k,N} = a_{k,N}$, $k = 1, \ldots, N$.
  - $X[n]$ **is complex:**
    The correlations of $X[n]$ and $X[-n]$ are complex conjugate, therefore, $\widetilde{a}_{k,N} = a^*_{k,N}$.

  We can readily see that, since $|\epsilon_N[n]|^2$ and $|\widetilde{\epsilon}_N[n]|^2$ are both real,

$$\mathrm{E}\left[|\epsilon_N[n]|^2\right] = \mathrm{E}\left[|\widetilde{\epsilon}_N[n]|^2\right] .$$

  For notation ease, but without loss of generality, we shall restrict ourselves to the real case.

Linear Estimation of AR Parametric Models

- **Levinson's Algorithm: Recursion for the Backward and Forward Predictors**

The milestone of the Levinson's algorithm is the following property enabling to express the forward and backward predictors in a recursive form:

Property

We have

$$\textbf{Forward predictor} \quad \widehat{X}_N[n] = \widehat{X}_{N-1}[n] + K_N \left(X[n-N] - \widetilde{\widehat{X}}_{N-1}[n-N] \right)$$

$$\textbf{Forward predictor error} \quad \epsilon_N[n] = \epsilon_{N-1}[n] - K_N \widetilde{\epsilon}_{N-1}[n],$$

and

$$\textbf{Backward predictor} \quad \widetilde{\widehat{X}}_N[n-N] = \widetilde{\widehat{X}}_{N-1}[n-N] + K_N \left(X[n] - \widehat{X}_N[n] \right)$$

$$\textbf{Backward predictor error} \quad \widetilde{\epsilon}_N[n-N] = \widetilde{\epsilon}_{N-1}[n-N] - K_N \widetilde{\epsilon}_{N-1}[n-N],$$

where

$$\textbf{Reflection coefficient} \quad K_N = \frac{\mathrm{E}\left[\epsilon_{N-1}[n] X[n-N]\right]}{\mathrm{E}\left[\widetilde{\epsilon}_{N-1}[n-N] X[n-N]\right]}.$$

Linear Estimation of AR Parametric Models

- **Levinson's Algorithm: Recursion for the AR Coefficients**

  Inserting the forward predictor

$$\widehat{X}_N[n] = a_{1,N} X[n-1] + \ldots + a_{N,N} X[n-N] \,.$$

  into the recursive prediction formula

$$\widehat{X}_N[n] = \widehat{X}_{N-1}[n] + K_N \left(X[n-N] - \widetilde{\widehat{X}}_{N-1}[n-N] \right)$$

  we obtain

$$\begin{aligned} a_{k,N} &= a_{k,N-1} - K_N a_{N-k,N-1} \,, \quad k = 1, \ldots, N-1 \,, \\ a_{N,N} &= K_N \,, \end{aligned}$$

  where the reflection coefficient K_N can be expressed as

$$K_N = \frac{R_X[N] - \sum_{k=1}^{N-1} a_{k,N-1} R_X[N-k]}{R_X[0] - \sum_{k=1}^{N-1} a_{k,N-1} R_X[k]} \,.$$

  The above equations recursively provide the coefficients of the AR process.

Linear Estimation of AR Parametric Models

- **Levinson's Algorithm: Recursion for the AR Coefficients**

  1. Start the iteration with $K_1 = R_X[1]/R_X[0]$, and $a_{1,1} = K_1$ (1 multiplication);
  2. Recursively use

$$a_{k,N} = a_{k,N-1} - K_N a_{N-k,N-1}\,, \quad k = 1, \ldots, N-1\,,$$
$$a_{N,N} = K_N\,,$$

  where K_N can be expressed as

$$K_N = \frac{R_X[N] - \sum_{k=1}^{N-1} a_{k,N-1} R_X[N-k]}{R_X[0] - \sum_{k=1}^{N-1} a_{k,N-1} R_X[k]}\,.$$

  ($\mathcal{O}(N)$ multiplications each recursion step, where N is the step prediction order).

  In matrix form:

$$K_N = \frac{[R_X[N] \;\; \cdots \; R_X[1]] \begin{bmatrix} -1 & a_{1,N-1} & \cdots & a_{N-1,N-1} \end{bmatrix}'}{[R_X[0] \cdots R_X[N-1]] \begin{bmatrix} -1 & a_{1,N-1} & \cdots & a_{N-1,N-1} \end{bmatrix}'}$$

  and

$$\begin{bmatrix} a_{1,N} \\ \vdots \\ a_{N-1,N} \\ a_{N,N} \end{bmatrix} = \begin{bmatrix} a_{1,N-1} \\ \vdots \\ a_{N-1,N-1} \\ 0 \end{bmatrix} - K_N \begin{bmatrix} a_{N-1,N-1} \\ \vdots \\ a_{1,N-1} \\ -1 \end{bmatrix}$$

Linear Estimation of AR Parametric Models

- **Levinson's Algorithm: Computational Burden**

  Levinson's algorithm $= \mathcal{O}(N^2)$ multiplications.

  Yule-Walker equations $= \mathcal{O}(N^2)$ multiplications (Toeplitz systems).

  However, using the Levinson's algorithm, with $\mathcal{O}(N^2)$ multiplications we obtain not only the coefficient of order N but also the coefficients of all orders less than N as well as an estimation of the model order!

Linear Estimation of AR Parametric Models

- **Matlab Practice: Levinson's Algorithm as Order Estimation**

  Let's see how the Levinson's Algorithm can be used to estimate the order of an AR process.

  - Generate N samples of an AR process of order K (make sure the associated synthesis filter is stable, and that $N >> K$). Use the Matlab commands `arima` and `simulate`.
  - Estimate the coefficients of the AR process assuming its order to be $L > K$. Use the Matlab command `aryule` making sure to collect the estimated coefficients as well as the reflection coefficients and the estimation error.
  - Analyze the reflection coefficients. What can you remark?

Linear Prediction of AR Parametric Models

- **Prediction as Orthogonal Projection**

  Linear prediction of AR processes answers problem of the type:

  - Given $X[n]$, $X[n-1]$, $X[n-2]$, ..., how can I predict $X[n+k]$?
  - Which linear combination of $X[n]$, $X[n-1]$, $X[n-2]$, ..., best approximates $X[n+k]$?

  For AR process such a problem has an immediate solution, provided that $P(z)$ is minimum phase. How can that be so simple? **Hilbert Spaces do the job for you!**

  AR processes are w.s.s., therefore $\in L_2(P)$ (space of square integrable r.v.), the latter being a Hilbert space with norm $\|X\| = \mathrm{E}\left[|X|^2\right]$.

Linear Prediction of AR Parametric Models

- **Hilbert Subspaces of the Past**

  $H(A,n)$ Hilbert subspace of $\mathrm{L}_2(P)$ generated by the r.v. $A[n], A[n-1], A[n-2], \ldots$:

  $H(A,n)$ is composed by the linear combinations of the r.v. $\{A(n-k)\}_{k\in\mathbb{N}}$ and the limits of such linear combinations, *i.e.*, $H(A,n) = \text{span}\left(\{A(n-k)\}_{k\in\mathbb{N}}\right)$.

Theorem (Filtering the past)
Let $X[n]$ be a wide sense stationary process and $Y[n]$ the output of a stable causal filter when the input is $X[n]$. Then

$$Y[n] \in H(X,n)\,, \quad \forall\ n \in \mathbb{Z}\,.$$

In the framework of AR processes, the above theorem gives the following result:

Theorem
Let $X[n]$ be an AR process. Assume that the corresponding polynomial $P(z)$ is minimum phase. Then

$$H(X,n) = H(W,n)\,, \quad \forall\ n \in \mathbb{Z}\,.$$

Linear Prediction of AR Parametric Models

- **Prediction as Orthogonal Projection onto the Hilbert Subspace of the Past**

  The element of $H(X, n)$, denoted $\widehat{X}_{n+k,n}$, that best approximate (in the mean square sense) $X[n+k] \in H(X, n+k)$ is the orthogonal projection of $X[n+k]$ onto $H(X, n)$.

  In particular
  - $\widehat{X}_{n+k,n}$ exists, is unique, and it is such that $X[n+k] - \widehat{X}_{n+k,n}$ is orthogonal to $H(X, n)$.
  - $\widehat{X}_{n+k,n}$ is the orthogonal projection of $X[n+k]$ onto $H(X, n)$, and, in the prediction framework, it is called k-*steps predictor*.

- **Intuitive Property to Easily Compute the k-Steps Predictor**

  Let Y be a r.v. in $H(X, n+k)$. Suppose that we can write Y as

  $$Y = A + B\,.$$

  where A is a r.v. orthogonal to $H(X, n)$ and B is a r.v. in $H(X, n)$. Then, the orthogonal projection of Y onto $H(X, n)$ is given by B.

Linear Prediction of AR Parametric Models

- **Intuitive Property and Synthesis Filter**

  For the k-steps predictor of an AR processes express using the synthesis filter we have

$$X[n+k] = \sum_{l=0}^{\infty} h_S[l]W[n+k-l] = \underbrace{\sum_{l=0}^{k-1} h_S[l]W[n+k-l]}_{A} + \underbrace{\sum_{l=k}^{\infty} h_S[l]W[n+k-l]}_{B} .$$

  Then

$$\widehat{X}_{n+k,n} = \sum_{l=k}^{\infty} h_S[l]W[n+k-l] .$$

- 1-**Step Predictor**

  The case of a 1-step predictor $\widehat{X}_{n+1,n}$ is quite simple (no synthesis representation)

$$X[n+1] = -\sum_{k=1}^{M-1} p_k X[n+1-k] + W[n+1] , \text{ therefore } \widehat{X}_{n+1,n} = -\sum_{k=1}^{M-1} p_k X[n+1-k] .$$

  In a practical scenario, $p_1, \ldots, p_{M-1}$ are estimated using the Yule Walker equations.

Assignement

For next week, provide the expression of a 2-step predictor without using the synthesis filter.

Estimation of Parametric Probability Densities

- **Maximum Likelihood Estimation**

  The framework here is the following:

  - We observe the realization of a process $Y[n]$, say $y[1], \ldots, y[N]$;
  - We know that the probability density of the process $Y[n]$ has a particular form $f_Y(y, \boldsymbol{\theta})$ characterized by a set of K parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$;
  - We do not know the value of the parameters $\boldsymbol{\theta}$.

  Knowledge of $\boldsymbol{\theta}$ provides $f_Y(y, \boldsymbol{\theta})$ and information related to it.

  Having observed the samples $y[1], \ldots, y[N]$, among all the possible parameter values $\boldsymbol{\theta}$ we choose the one maximizing the probability of obtaining $y[1], \ldots, y[N]$, that is

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} f_Y(y[1], \ldots, y[N], \boldsymbol{\theta})$$

  The function $f_Y(y[1], \ldots, y[N], \boldsymbol{\theta})$ is called **Likelihood Function - LF** for the samples $y[1], \ldots, y[N]$, and $\widehat{\boldsymbol{\theta}}$ is the the **Maximum Likelihood Estimator - MLE**.

Estimation of Parametric Probability Densities

- **Maximum Likelihood Estimation (Example 1/2)**

  A thermistor probe (resistance varying according to the temperature) measures the skin temperature. The signal can be modeled as

  $$Y[n] = x[n] + W[n]$$

  where

  - $x[n]$ are the values of the temperature (deterministic);
  - $W[n]$ is a Gaussian, centered, white noise with unitary variance.

  $Y[n]$ is an i.i.d. sequence of Gaussian random variables (variance 1 and mean $x[n]$)

  $$f_Y(y[1], \ldots, y[N], \boldsymbol{\theta}) = \frac{1}{(2\pi)^{N/2}} \exp\left\{ -\frac{1}{2} \sum_{k=1}^{N} |y[k] - x[k]|^2 \right\}.$$

  where $\boldsymbol{\theta} = x[1], \ldots, x[N]$ (the mean vector of the joint Gaussian distribution).

  $f_Y(y[1], \ldots, y[N], \boldsymbol{\theta})$ represents the **Likelihood Function** and its maximization w.r.t. $x[1], \ldots, x$ provides the **most likely** values of the temperatures.

Estimation of Parametric Probability Densities

- **Maximum Likelihood Estimation (Example 2/2)**

  The maximization of the Likelihood Function is equivalent to the maximization of its logarithm, that is

$$\log\left(f_Y(y[1],\ldots,y[N],\boldsymbol{\theta})\right) = \log\left(\frac{1}{(2\pi)^{N/2}}\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}|y[k]-x[k]|^2\right\}\right)$$
$$= \log\left(\frac{1}{(2\pi)^{N/2}}\right) - \frac{1}{2}\sum_{k=1}^{N}|y[k]-x[k]|^2 \ .$$

  In this particular Gaussian case, the maximization of the Likelihood Function $f_Y(y[1],\ldots,y[N],\boldsymbol{\theta}$ corresponds to the minimization of the Mean Square Error

$$\sum_{k=1}^{N}|y[k]-x[k]|^2 \ .$$

Estimation of Most Probable Realization

- **A Posteriori Distribution**

  Process $X[n]$ indirectly observed via the process $Y[n]$.

  We assume here that $X[n]$ is a discrete valued process.

  Having observed the values $y[1], \ldots, y[N]$, the **a posteriori distribution** is given by

$$P\left(X[1], \ldots, X[N] \mid y[1], \ldots, y[N]\right) .$$

- **Maximum A Posteriori**

  Given the *a posteriori* distribution and the observed values $\boldsymbol{y} = y[1], \ldots, y[N]$ we estimate the values $\boldsymbol{x} = x[1], \ldots, x[N]$ of the process $\boldsymbol{X} = X[1], \ldots, X[N]$ as the ones maximizing the *a posteriori* distribution, *i.e.*,

$$\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} P\left(\boldsymbol{X} = \boldsymbol{x} \mid \boldsymbol{y}\right) .$$

Estimation of Most Probable Realization

- **Assignement**

Assignement
By next week propose a method to estimate σ_W^2 (hint: $X[n]$ is a correlated signal), and give a Matlab or Python example. Use a simple signal like a sum of K complex sinusoids (harmonic signal). Hint: Read about the Music algorithm in the old lecture notes.

Assignement
Should we use a biased or an unbiased correlation?

For next week, propose how to efficiently implement the Yule Walker equations so to reduce estimation issues like the one arising at extreme lags for the unbiased correlation, or the one arising when dealing with non zero mean signals and the biased correlation.

Assignement
For next week, provide the expression of a 2-step predictor without using the synthesis filter.

Chapter Summary

Chapter Summary

Basic Models

- **Wide Sense Stationary Processes (Signals)** (non parametric)
  They present a constant mean, a finite variance, a correlation depending only on the difference of the time lags, and they admin a power spectral density.
- **Processes (Signals) with i.i.d. Behaviour** (non parametric)
  They are a particular class of w.s.s. processes. The Gaussian white noise is the emblematic signal of such a class. The power spectral density of an i.i.d. process is constant.
- **Processes (Signals) Defined by a Parametric Law** (parametric)
  They admit a law (cumulative distribution / probability density) described by a finite number of parameters. As examples, we saw the discrete valued Markov chains, characterized by the initial and the transitions probabilities, and the Gaussian processes, characterized by the mean and the variance.
- **Processes (Signals) Defined by a Parametric Equation** (parametric)
  They can be described by a closed form equation dependent on a finite number of parameters. As example, we saw the Auto-Regressive (AR) processes of order M, characterized by the M parameters $p_1, \ldots, p_M$ and the noise power σ_W^2. AR are, by definition, w.s.s., they have a very particular form of the correlation, and they admin a power spectral density that can be expressed as a rational polynomial. AR process are easily represented using the associated synthesis filter, enabling to express them as a sum of orthogonal process.

Chapter Summary

Basic Methods

After collecting a set of data (signal samples), the common operations to extract information are:

- **Estimation**
  Obtain a certain quantity, such the values of a $x[k]$, or a set of parameters, from a sequence of observations $y[k]$ (**does not imply causality**).
  - Estimating the realization of a w.s.s. (non parametric signal model) processes giving observations related to it: **Wiener Filter**.
  - Estimating the parameters of an AR (parametric signal model) process giving a realization of the process: **Yule Walker Equations** (can be optimally solved using the Levinson algorithm).
  - Estimating the parameters of a probability density (parametric model) giving its outcome: **Maximum Likelihood Estimator**.
  - Estimating the most probable realization of a process given its distribution conditioned to the observations: **Maximum a Posteriori**.
- **Prediction**
  Predict the evolution of a quantity, such as the values of a process $x[n]$, given the values of its past $x[n-1], x[n-2], \ldots$ (**it implies causality**).
  - Predicting the future value of a w.s.s. processes (non parametric signal model) giving its past observations: **Yule Walker Equations** (can be optimally solved using the Levinson algorithm).
  - Predicting the future value of an Auto-Regressive processes (parametric signal model) giving its past observations: **Intuitive property** (projection/synthesis filter) and **Yule Walker Equations** (can be optimally solved using the Levinson algorithm).