Signal Processing
from Fourier to Machine Learning
Part 3 : Random signals

R. Flamary

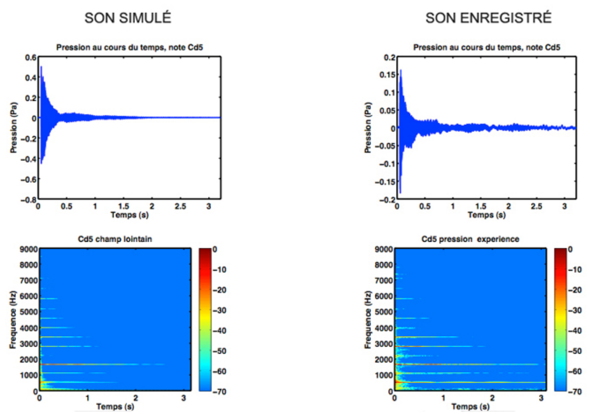
November 21, 2024

Full course overview

Course overview

Piano note

Propagation of the waves on the string of the piano and the air is modeled by Ordinary Differential Equation (LTI system, one note is an approximation of an impulse response).



Why are simulated and recorded signal different?

Source: Interstice, Images ©Juliette Chabassier 2012

Signal properties

Instantaneous power

The instantaneous power of signal $x(t)$

$$p_x(t) = |x(t)|^2 \tag{1}$$

Unit : Watt (W).

Energy of a signal

We define the energy of a signal $x(t)$ as :

$$E = \int_{-\infty}^{+\infty} |x(t)|^2 dt \tag{2}$$

the signal is said to be of finite energy if $E < \infty$ ($\|x\|_2 < \infty$ means $x \in L_2(\mathbb{R})$).
Unit: Joule, Calorie or Watt-hour (J, Cal ou Wh, 1 calorie = 4.2 J).

Average power

Average power of a signal

The average power of a signal is defined as

$$P_m = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{+\frac{T}{2}} |x(t)|^2 dt \tag{3}$$

- For a periodic signal, the average power can be computed on a unique period.
- Power is homogeneous to an energy divided by time.
- $P_{RMS} = \sqrt{P_m}$ is called the Root Mean Square power ("valeur efficace" in french).
- A finite energy signal has a n average power $P_m = 0$.
- Unit : Watt (W).

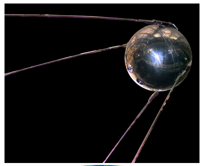
Notion of noise

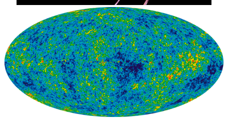
Definition

A natural process that can impede the measurement or interpretation of a signal.

Examples

- Satellite Signals and astrophysics
  - Telecommunications : Satellite signal is the signal of interest, astronomical background is the noise.
  - Astrophysics: Satellite signal is the noise, astronomical background is the signal of interest.
- Electricity grid EDF, spike at 50Hz when measuring low amplitude tensions.





Additive noise

Additive noise is a kind of noise that is added to the signal of interest.

$$y(t) = x(t) + b(t)$$

$y(t)$ is the observed signal, $x(t)$ the signal of interest and $b(t)$ is the noise.

Signal to Noise Ratio (SNR)

Definition

The Signal to Noise Ratio is defined as:

$$SNR = \frac{P_S}{P_N} \quad \text{ou} \quad SNR(dB) = 10\log_{10}(SNR) \tag{4}$$

where P_S is the power of the signal and P_N the power of the noise.

- An Analog-to-Digital conversion process should have the best possible SNR.
- The SNR is often used for additive noise models.
- Other measures such as Peak Signal to Noise Ratio (PSNR) can be used on specific data (images).
- One of the objective of filtering is to get a better SNR when the signal and the noise have different frequency contents..

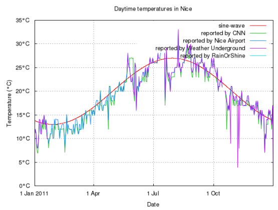
Random Signal

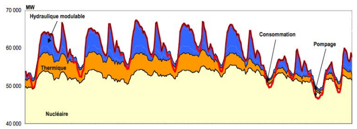
Motivation

- Hard to model **perfectly** physical signals measurement.
- Real signal are random or contain a random component.
- No exact prediction/reconstruction but **inference** si possible.

Examples

- Stock exchange.
- Temperature along the day.
- Instantaneous electrical energy usage.
- IP instruction pointer in processors.





Deterministic VS random/stochastic

$$x(t) = f(t, \dots)$$

Deterministic

- Signal modeled perfectly by a mathematical function of time.
- Knowledge of the signal at all time moment t .

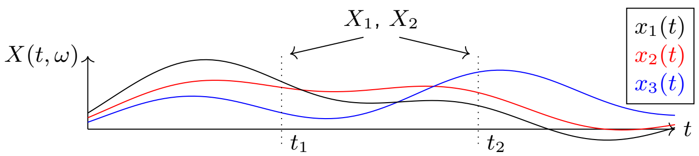
Random

- Part of uncertainty/stochasticity in the signal.
- Impossible to predict with certainty at time t.

Approach

- Uncertainty modeled by probability distributions.
- Using these probability distributions allows signal processing.
- Requires knowledge of both **signal processing** and **probability/statistics**.

Stochastic process



Definition

A stochastic process or random signal is a function of two variables. The first variable is the time t and $t \in \mathbb{R}$, the second variable is a random variable ω :

$$X(\underbrace{t}_{\text{time}}, \underbrace{\omega}_{\texttt{r.v.}}) \qquad (5)$$

- At $t = t_i$ fixed, $X(t_i, w) = X_i = X_i(\omega)$ is a random variable;
- For $\omega = \omega_i$ fixed (one realization), $X(t, \omega_i) = x_i(t)$ is a deterministic signal ;
- X can be discrete time with $X[n]$ and $n \in \mathbb{N}$

Example of stochastic process (1)

Random frequency

$$X(t, \omega) = \cos(\omega t)$$

- $t \in \mathbb{R}$ is time.
- $\omega \in [0, 2\pi]$ is a uniform random variable.
- X is a function of two variables.

Example of realizations (1)

Example of stochastic process (2)

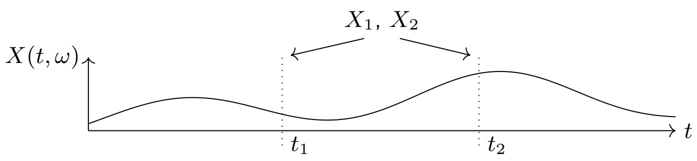
Random phase

$$X(t, \phi) = \cos(\pi * t + \phi)$$

- $t \in \mathbb{R}$ is time.
- $\phi \in [0, 2\pi]$ is a uniform random variable.
- X is a function of two variables.

Example of realizations

Description of a stochastic process



Complete description

The stochastic process $X(t, w)$ is completely known if $\forall t_1, t_2, \ldots, t_k$ and $\forall k$, we have access to the joint distribution :

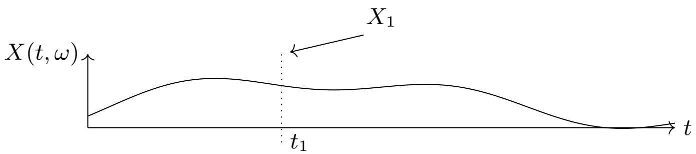
$$p_{X_1,\ldots,X_k}(x_1, \ldots, x_k) \tag{6}$$

where $X_1, \ldots, X_k$ are the random variables associated to time t_k.

- This means that you know all the possible probabilistic relation between any time samples.
- Often impossible in practice (except for Gaussian distributions).

$\Rightarrow$ Partial description.

Description with one time sample



Definition

We know the description of $X(t, w)$ for one time sample if $\forall t_1 \in \mathbb{R}$ we know the probability distribution of $X(t_1, w)$, *i.e.* the random variable X_1 at t_1.

Moments

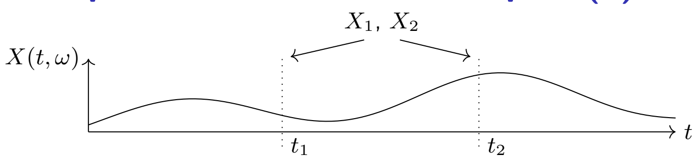
- The first moment of the stochastic process, its mean (and the expectation of X_1) is expressed as:

$$m_X(t_1) = E[X(t_1, w)] = \int x_1 p_{X_1}(x_1) d_{x_1} \tag{7}$$

- We can also define the moment of order n :

$$m_X^{(n)}(t_1) = E[X(t_1, w)^n] = \int x_1^n p_{X_1}(x_1) d_{x_1} \tag{8}$$

Description with two time samples (1)



Définition

$X(t, w)$ is known with two time samples $\forall t_1, t_2 \in \mathbb{R}^2$ if the joint distribution between the two random variables $X(t_1, w)$ and $X(t_2, w)$ is known:

$$p_{X_1,X_2}(x_1, x_2) \text{ is known } \forall t_1, t_2$$

Correlation between two time samples

$$R_X(t_1, t_2) = E[X(t_1, w) X^*(t_2, w)] = \int\int x_1 x_2^* p_{X_1,X_2}(x_1, x_2) d_{x_1} d_{x_2} \tag{9}$$

Covariance between two time samples

$$C_X(t_1, t_2) = E\left[(X(t_1, w) - m_X(t_1))(X^*(t_2, w) - m_X^*(t_2))\right] \tag{10}$$

Description with two time samples (2)

Remark

- $X_c(t,w) = X(t,w) - m_X(t)$ is the centered signal.
- Correlation and Covariance are properties of order 2 of the random signal.

Independence between X_1 and X_2

$$p_{X_1,X_2}(x_1,x_2) = p_{X_1}(x_1)p_{X_2}(x_2) \tag{11}$$

- The correlation becomes:

$$R_X(t_1,t_2) = E[X(t_1,w)X^*(t_2,w)] = E[X(t_1,w)]E[X^*(t_2,w)] = m_X(t_1)*m_X(t_2)^*$$

- If the signal is centered then $R_X(t_1,t_2) = 0$ for $t_1 \neq t_2$. This type of signal is called a *white noise*.

Stationarity

Definition

A signal is said to be stationary if its statistical properties are invariant to translation in time.

Strict stationarity

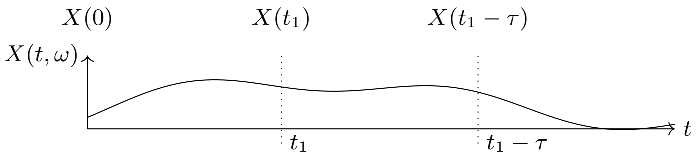
When the signal is completely known the strict stationarity implies that

$$p_{X(t_1),\dots,X(t_k)} = p_{X(t_1-\tau),\dots,X(t_k-\tau)} \quad \forall \tau \in \mathbb{R}. \tag{12}$$

- In order to simplify the notation we take $X(t_1) = X(t_1,w) = X_1$.
- The correlation $R_X(t_1,t_2)$ in this case depends only on the difference $t_1 - t_2$.
- Stationarity makes studying a random signal much simpler.
- Strict stationarity hard to check in practice.

$\Rightarrow$ Weak or wide-sense stationarity.

Weak or wide-sense stationarity (WSS) (1)



Stationarity for description with one time sample

$$p_{X(t_1)} = p_{X(t_1-\tau)} = p_{X(0)} \quad \forall \tau \in \mathbb{R} \tag{13}$$

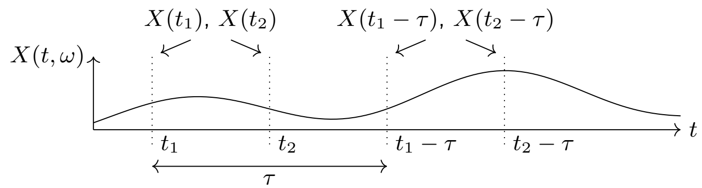
- All random variables $X(t)$ follow the same law $\forall t \in \mathbb{R}$.
- The moments

$$E(X(t_1)^n) = E(X(t_1-\tau)^n) = m_X^{(n)}(t) = m_X^{(n)} \quad \forall n \in \mathbf{N} \tag{14}$$

are constants and do not depend on the time.

- A signal is stationary of order 1 if its order 1 moment does not depend on time.

Weak or wide-sense stationarity (WSS) (2)



Stationarity for description with two time samples

The joint distribution between two time samples t_1, t_2 depends only on their difference $t_1 - t_2$

$$p_{X(t_1),X(t_2)} = p_{X(t_1-\tau),X(t_2-\tau)} = p_{X(t_1-t_2),X(0)} \quad \forall \tau \in \mathbf{R} \tag{15}$$

- The correlation is

$$R_X(t1,t2) = E[X(t_1)X^*(t_2)] = E[X(t_1-t_2)X^*(0)] = E[X(t)X^*(t-\tau)] \tag{16}$$

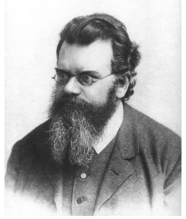
- We define the auto-correlation function as

$$R_X(\tau) = E(X(t)X^*(t-\tau)) \tag{17}$$

- A signal is said to be wide-sense stationary (WSS) if it is stationary at order two (two time samples) and order one and if $E[|X(t,w)|^2]$ is bounded for all $t \in \mathbb{R}$.

Ergodicity

Ergodic hypothesis



- The statistical properties of an ergodic process can be estimated on a unique realization if observed for a long time.
- Hypothesis cannot be tested in practice.
- Originally formulated by Boltzmann for his Kinetic theory of gases.

Temporal averaging

The time average (or order n) of a signal $x(t)$ is defined as

$$\overline{x^n} = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} x(t)^n dt \quad (18)$$

The time average of order 2 is the average power of the signal.

Erogodicity (2)

Time averaging of a stochastic process

$$\overline{X(t,w)^n} = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t,w)^n dt \quad (19)$$

In the general case this average is a random variable that has realizations $\overline{x_i^n}$ for random signal realizations $x_i(t)$

Temporal cross-correlation

$$\overline{X(t,w)X(t-\tau,w)} = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t,w)X(t-\tau,w)dt \quad (20)$$

Definition of ergodicity

The signal $X(t,w)$ is said to be ergodic if its temporal averaging and temporal cross correlation are certain (not random).
This property implies that

$$\overline{X(t,w)^n} = \overline{x_i^n} = \overline{x^n}, \forall \text{ realization } w_i$$

Stationary and ergodic signals

- With stationary and ergodic signals, one can compute a mathematical expectation with a temporal average.
- For first order moments it means :

$$\overline{x_i^n} = E[\overline{X(t,w)^n}] = E[X(t,w)^n] = m_X^{(n)} \quad (21)$$

- This is very important because having several realization of a random signal is seldom possible.
- One can measure/record long portion of a signal to have a good **estimation** of the mean, variance and covariance.
- For a second order WSS signal :

$$R_X(\tau) = E[X(t)X^*(t-\tau)] = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} X(t,w)X^*(t-\tau,w)dt \quad (22)$$

  The correlation of the signal with itself is also called auto-correlation.

Correlation and cross-correlation

The signals $X(t,x)$ and $Y(t,w)$ are both WSS and ergodic:

- **Correlation** or **autocorrelation**

$$R_X(\tau)=E[X(t,w)X^*(t-\tau,w)]=\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t,w)X^*(t-\tau,w)dt \quad (23)$$

  Measures a linear relation between two time instants of a signal.
- **Cross-correlation**

$$R_{XY}(\tau)=E[X(t,w)Y^*(t-\tau,w)]=\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t,w)Y^*(t-\tau,w)dt \quad (24)$$

  Measures a linear relation across two signals with a delay.
- Autocorrelation and cross-correlation have several nice properties.

Properties of Autocorrelation

Let $X(t, w)$ be a wide sens stationary signal.

Conjugate symmetric

$$R_X(\tau) = R_X^*(-\tau) \qquad (25)$$

Centering

$$R_X(\tau) = R_{X_c}(\tau) + m_X^2 \qquad (26)$$

Average power

$$P_X = R_X(0) \qquad (27)$$

By definition $R_X(0) \geq 0$

Non-negativity

$$\sum_{i,j} \lambda_i \lambda_j R_X(\tau_i - \tau_j) \geq 0, \quad \forall i, j \qquad (28)$$

Correlation coefficient

$$\rho(\tau) = \frac{R_X(\tau)}{R_X(0)} \qquad (29)$$

Memory

Finite memory process:

$\exists T_{max}$ such that $\rho(t) = 0$ for $|t| > T_{max}$

Properties of Cross-correlation

Let $X(t, x)$ and $Y(t, w)$ be two wide sens stationary signals

Hermitian Symmetry

$$R_{XY}(\tau) = R_{XY}^*(-\tau) \qquad (30)$$

Maximum

One can use the Schwartz inequality to proove that

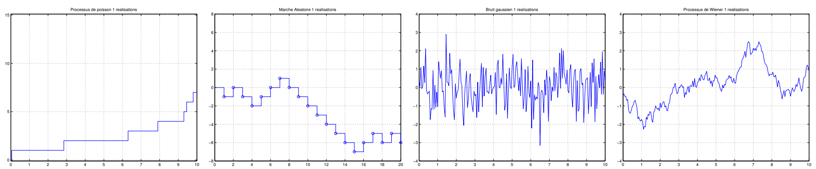
$$|R_{XY}(\tau)|^2 \leq R_{XX}(0) * R_{YY}(0) \qquad (31)$$

hence

$$|R_X(\tau)| \leq R_X(0) \qquad (32)$$

The auto correlation reaches its maximum in 0

Examples of stochastic processes



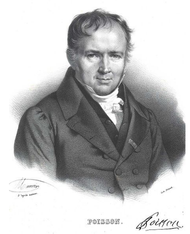
Examples of stochastic processes

1. Poisson process.
2. Random walk.
3. Gaussian Processes
4. Wiener process

Poisson distribution

Poisson distribution

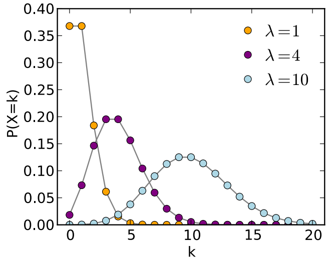
- Discrete distribution.
- Model the probability of a given number events of occurring in a given time interval.
- Parameter λ corresponds to the average number of occurences in the time interval.
- If λ_0 is an average number of occurrence per time unit then for a time interval Δt, $\lambda = \lambda_0 \Delta t$.

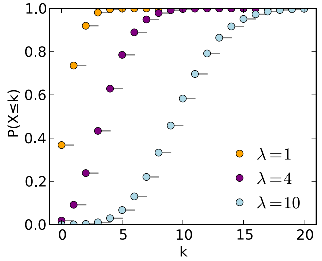


Examples of Poisson distribution

- Electronic : model the failure in electronic circuits.
- Waiting lines in stores : Model the arrival of clients at checkout and the time to handle them.
- Astronomy : Model the number of photon on CCD sensors in low light conditions.

Poisson distribution





Properties

- Probability density function (PDF)

$$p(k) = P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

- Expectation:

$$m_X = E(X) = \frac{1}{n}\sum_{k=0}^{\infty} k e^{-\lambda}\frac{\lambda^k}{k!}$$

$$= \lambda e^{-\lambda}\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

- Variance:

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= \sum_{k=1}^{+\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} - \lambda^2 \\ &= \lambda e^{-\lambda}\sum_{k=1}^{+\infty}\frac{d}{d\lambda}\frac{\lambda^k}{(k-1)!} - \lambda^2 \\ &= \lambda \end{aligned}$$

Poisson process

Count between two time instants

- Let $N(t, t')$ be the number of events t_i occurring during interval $\Delta t = t' - t$.
- $N(\cdot)$ is a Poisson random variable of parameter $\lambda_0 \Delta t$ and probability

$$P(n(t, t') = k) = \frac{e^{-\lambda_0 \Delta t}(\lambda_0 \Delta t)^k}{k!} \qquad (33)$$

- If the intervals $[t, t']$ and $[u, u']$ are disjoints then the random variables $N(t, t')$ and $N(u, u')$ are independent.

Poisson process

A Poisson process is a random signal defined as :

$$X_p(t) = N(0, t) \qquad \forall t \in \mathbb{R}^+ \qquad (34)$$

- Process corresponds to counting random events.
- The events are defined as their time instant of occurrence t_i.

Properties of the Poisson process

Mean $m_{X_p}(t)$

$$m_{X_p}(t) = E[X_p(t)] = E[N(0, t)] = \lambda t \qquad (35)$$

- Non stationary process.

Variance $Var(X_p(t))$

$$Var(X_p(t)) = E[(X_p(t) - m_{X_p}(t))^2] = \lambda t \qquad (36)$$

Autocorrelation $R_{X_p}(t_1, t_2)$ for $t_1 \leq t_2$

$$R_{X_p}(t_1, t_2) = \lambda t_1 + \lambda^2 t_1 t_2$$

Proof of autocorrelation $R_{X_p}(t_1, t_2)$

$$R_{X_p}(t_1, t_2) = \lambda t_1 + \lambda^2 t_1 t_2 \qquad (37)$$

Proof.

- If $t_1 = t_2$ Then Eq. (39) is true :

$$E[X_p(t)^2] = Var(X_p(t)) + E[X_p(t)]^2 = \lambda t + \lambda^2 t^2$$

- If $t_1 < t_2$, the r.v. $X_p(t_1)$ and $X_p(t_2) - X_p(t_1)$ are independent because they are Poisson with non overlapping intervals.
  - $X_p(t_1) \sim N(0, t_1)$
  - $X_p(t_2) - X_p(t_1) \sim N(0, t_2 - t_1)$
- We can deduce that

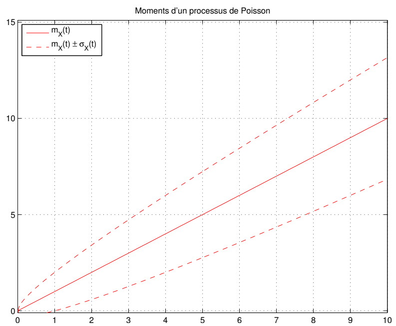
$$E[X_p(t_1)(X_p(t_2) - X_p(t_1)))] = E[X_p(t_1)]E[X_p(t_2) - X_p(t_1)] = \lambda t_1 \lambda(t_2 - t_1)$$

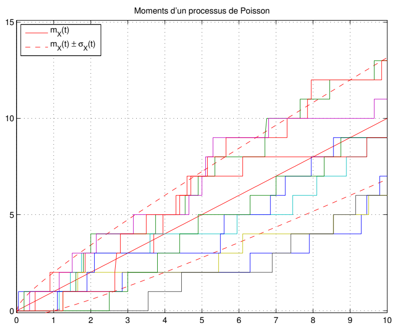
and since $$X_p(t_1)X_p(t_2) = X_p(t_1)(X_p(t_2) - X_p(t_1))) + X_p(t_1)^2$$

we find $$R_{X_p}(t_1, t_2) = \lambda t_1 + \lambda^2 t_1^2 + \lambda t_1 \lambda(t_2 - t_1)$$

□

Poisson process example realizations

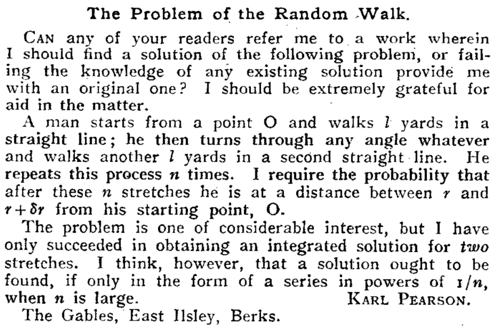




Poisson process

- $\lambda = 1$
- $m_{X_p}(t) = \lambda t$
- $Var(X_p(t)) = \lambda t$

History of the random walk



The Problem of the Random Walk.

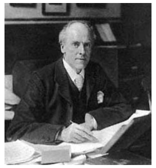
CAN any of your readers refer me to a work wherein I should find a solution of the following problem, or failing the knowledge of any existing solution provide me with an original one? I should be extremely grateful for aid in the matter.

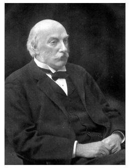
A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r+\delta r$ from his starting point, O.

The problem is one of considerable interest, but I have only succeeded in obtaining an integrated solution for *two* stretches. I think, however, that a solution ought to be found, if only in the form of a series in powers of $1/n$, when n is large.

KARL PEARSON.

The Gables, East Ilsley, Berks.





- Introduced by Karl Pearson in 1905 to model mosquitoes migration in a forest (Nature 72, 294; 318; 342 (1905)).
- Problem solved by Lord Rayleigh who provides him with the Gaussian approximation solution.
- Remark from Karl Pearson:

  "*The most probable place to find a drunken man who is at all capable of keeping on his feet is somewhere near his starting point!*".

Radom walk stochastic process

Définition

Discrete stochastic process described recursively by

$$X_{rw}[n] = X_{rw}[n-1] + sU[n] = s\sum_{i=1}^{n} U[i]$$

- $U[n] \sim \mathcal{U}(\{-1, 1\})$ are independent uniform random variables $U[n]$.
- Starting point : $X_{rw}[0] = 0$
- $s \in \mathbb{R}$ is the step.
- By construction the value of the stochastic process changes between two time instant n and $n+1$.
- Once can create a continuous version with T:

$$\tilde{X}_{rw}(t) = \{X_{rw}[n] | nT \leq t < (np+1)T\}$$

Properties of random walks (1)

Probability at time instant n

- At n, X_{rw} can take the values

$$X_{rw}[n] \in \{-ns, -(n+1)s, \ldots, (n-1)s, ns\}$$

- If the realizations of U_n gave k times $+1$ and $n-k$ times -1 then

$$X_{rw}[n] = ks - (n-k)s = ms, \qquad m = 2k - n$$

- We recognize a Binomial distribution with parameter $p = \frac{1}{2}$

$$P(X_{rw}[n] = ms) = \binom{n}{k} \frac{1}{2^n} = C_n^k \frac{1}{2^n} \qquad k = \frac{m+n}{2} \qquad (38)$$

Properties of random walks (2)

Moyenne $m_{X_{rw}}[n]$

$$m_{X_{rw}}(n) = E\left[\sum_{i=1}^{n} sU[i]\right] = s\sum_{i=1}^{n} E[U[i]] = 0 \quad (39)$$

Variance $Var(X_{rw}[n])$

$$Var(X_{rw}[n]) = E[X_{rw}[n]^2] = E[(\sum_{i=1}^{n} sU[i])^2] = \sum_{i=1}^{n} E[(sU[i])^2] = ns^2 \quad (40)$$

Autocorrelation $R_{X_{rw}}[n_1, n_2]$ pour $n_1 < n_2$

$$R_{X_{rw}}[n_1, n_2] = E[X_{rw}[n_1]X_{rw}[n_2]] = n_1 s^2$$

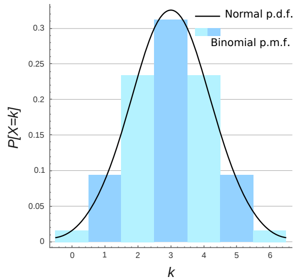
Stationary signal? NO !

Approximation for a large n

Approximating the Binomial law

For a large n, when k is in a neighborhood $\sqrt{npq}$ of np, then:

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq} \quad (41)$$

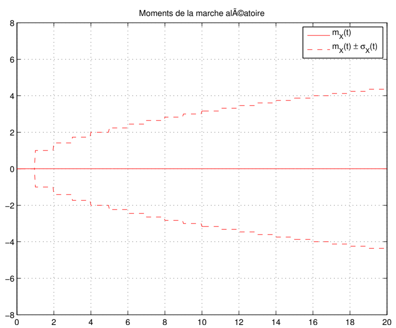


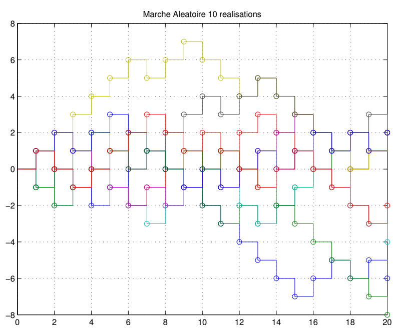
Application to random walk stochastic process

In our case $p = q = .5$ and for m=$2k - n$ then

$$P(X_{rw}[n] = ms) \approx \frac{1}{\sqrt{\pi n/2}} e^{-m^2/2n} \quad (42)$$

Random walk example realizations

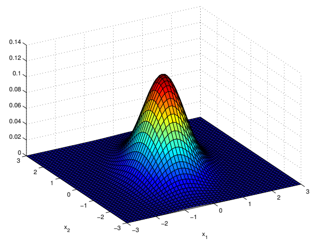


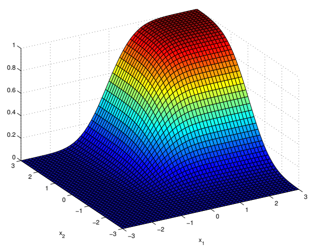


Random walk

- $s = 1$
- $m_{X_{rw}}(t) = 0$
- $Var(X_{rw}(t)) = \sqrt{n}$

Multivariate Gaussian distribution





Multivariate Gaussian distribution

- $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- Probability density function

$$p(x, y) = Ke^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

- Coefficient $K = \frac{1}{(2\pi)^{N/2}|\boldsymbol{\Sigma}|^{1/2}}$ where $|\cdot|$ is the determinant of the matrix.

- Espérance:

$$\mathbf{m}_X = E[\mathbf{X}] = \boldsymbol{\mu}$$

- Covariance:

$$Cov(\mathbf{X}) = E[(\mathbf{X} - \mathbf{m}_X)(\mathbf{X} - \mathbf{m}_X)^\top] = \boldsymbol{\Sigma}$$

Gaussian process (1)

Definition

A Gaussian process $X_G(t,w)$ is a random signal completely defined by a Gaussian distribution.

- Let $\mathbf{X}_G(w)$ the random vector:

$$\mathbf{X}_G(w)^T = [X_G(t_1,w), X_G(t_2,w), \dots, X_G(t_k,w)]^\top \tag{43}$$

- Then the Probability density function of the random vector is defined as

$$p_{\mathbf{X}_G}(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^k|\Sigma|}} e^{-\left(\frac{1}{2}(\mathbf{X}-m_\mathbf{X})^T\Sigma^{-1}(\mathbf{X}-m_\mathbf{X})\right)} \tag{44}$$

  Where the average vector $m_\mathbf{X}$ and the covariance matrix Σ are defined as

$$m_\mathbf{X} = [m_X(t_1), m_X(t_2), \dots, m_X(t_k)]^\top \tag{45}$$

$$(\Sigma)_{i,j} = E[X_c(t_i,w)X_c(t_j,w)^*] = C_X(t_i,t_j) \tag{46}$$

Gaussian process (2)

Stationary Gaussian process

- **Mean**

$$m_{\mathbf{X}_G} = m_X[1,1,\dots,1]^\top$$
$$m_{X_G}(t) = m_{X_G}$$

- **Covariance**

$$(\Sigma)_{i,j} = C_X(t_j - t_i) = C_X(\tau)$$

Discussion

- The Gaussian distribution is completely described by its two first order moments (means and covariance).
- For a stationary signal then only m_X and $R_{X_G}(\tau)/C_{X_G}(\tau)$ are necessary to completely know the random signal.

Gaussian noise

Definition

Gaussian noise is a random signal where every time instant follow a Gaussian distribution:

$$X_N(t,w) \sim \mathcal{N}(0,\sigma(t)^2) \quad \forall i \tag{47}$$

and for which all time instants are independent.

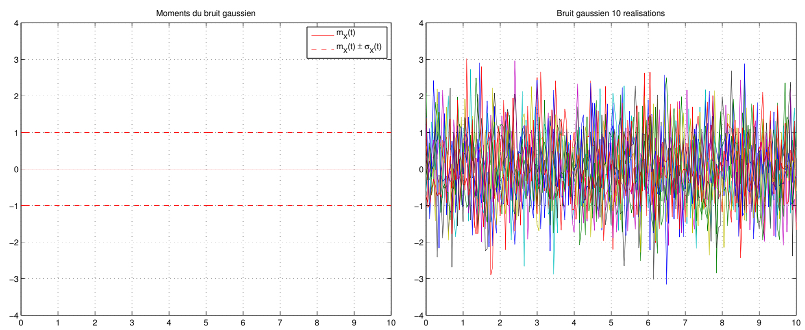
Independent and Identically Distributed (I.I.D.) Gaussian noise

- Independent means that $X_N(t_1,w)$ and $X_N(t_2,w)$ are indep. for $t_1 \neq t_2$.
- Identically Distributed means that the signal is stationary ($\sigma(t) = \sigma$)

$$m_{X_N}(t) = 0 \tag{48}$$
$$Var_{X_B}(t) = \sigma^2 \tag{49}$$
$$R_{X_N}(\tau) = C_{X_N}(\tau) = \sigma^2\delta(\tau) \tag{50}$$

- Signal is completely known because both mean and covariance are known.
- White noise : contains all the frequencies in the spectrum.
- Commonly used to model additive noise in discrete signals.

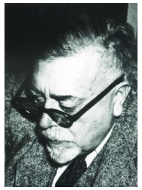
Gaussian noise example realizations



Gaussian noise

- $\sigma = 1$
- $m_{X_G}(t) = 0$
- $Var(X_G(t)) = \sigma^2$

Wiener Process



History

- Named in honor of Norbert Wiener, father of cybernetics.
- Also called Brownian motion (Robert Brown).
- Used in statistical physics, economy, finances.

Relation to random walks

The Wiener process can be constructed as

$$X_w(t,w) = \lim_{T \to 0} \tilde{X}_{rw}(t,w) \qquad \text{s.c.} \quad s = \sqrt{T}$$

It is a Gaussian process with the following 3 properties:

- $X(0,w) = 0$
- The random signal is almost surely continuous.
- Increments $X(t_2,w) - X(t_1,w)$ are independents and follow a Gaussian distribution $\mathcal{N}(0, t_2 - t_1)$.

Processus de Wiener (2)

Probability density function

$$p_{X_w(t,w)}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \tag{51}$$

Mean $m_{X_w}(t)$

$$m_{X_w}(t) = 0$$

Variance $Var_{X_w}(t)$

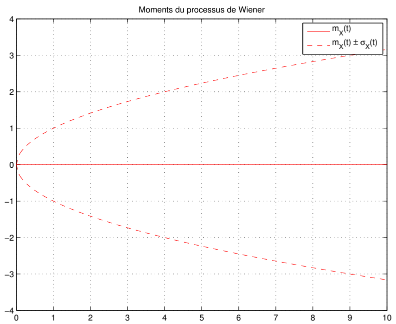
$$Var_{X_w}(t) = t$$

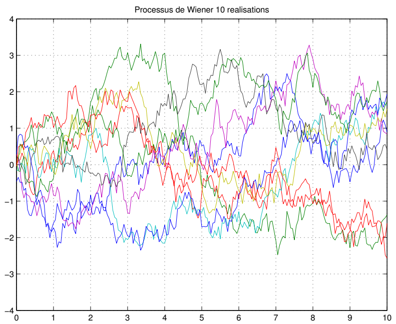
Autocorrelation $R_{X_w}(t_1,t_2)$ for $t_1 < t_2$

$$\begin{aligned} R_X(t_1,t_2) &= E(X(t_1,w)X(t_2,w)] && (52)\\ &= E(X(t_1,w)(X(t_2,w) - X(t_1,w) + X(t_1,w))) && (53)\\ &= E(X(t_1,w)(X(t_2,w) - X(t_1,w))) + E(X(t_1,w)^2) = t_1 && (54) \end{aligned}$$

Stationary signal? NO !

Wiener process example realizations





Wiener process

- $m_{X_w}(t) = 0$
- $Var(X_w(t)) = t$

Power Spectral Density (PSD)

Wiener-Kintchine Theorem

The Power Spectral Density (PSD) of a wide sens stationary stochastic process is the Fourier Transform of its autocorrelation function:

$$S_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau) e^{-2j\pi f\tau} d\tau \tag{55}$$

- It is a spectral representation of a random signal.
- **Inversion with inverse FT**

$$R_{XX}(\tau) = \int_{-\infty}^{\infty} S_{XX}(f) e^{2j\pi f\tau} df \tag{56}$$

- **Discrete signal** $X[k]$

$$S_{XX}(e^{2j\pi f}) = \sum_{k=-\infty}^{\infty} R_{XX}[k] e^{-2j\pi fk} = \sum_{k=-\infty}^{\infty} R_{XX}[k] z^{-k}$$

  - $z = e^{2j\pi f}$ we recover the Z-transform.
  - PSD is a periodic function.

Properties of the PSD

- The PSD is a real function since it is the FT of an even function.
- The PSD is a true density function in the probabilistic sens describing the repartition of the power across the frequency spectrum.
- **Positivity**
$$S_{XX}(f) \geq 0 \tag{57}$$
- **PSD and autocovariance**
$$S_{XX}(f) = \mathcal{F}\{C_x(\tau)\} + (m_X)^2\delta(f) \tag{58}$$
- **Average power**
$$P_X = R_{XX}(0) = \int_{-\infty}^{+\infty} S_{XX}(f)df = \lim_{T\to\infty}\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} |X(t,w)|^2 dt \tag{59}$$

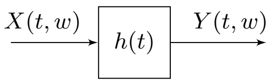
White noise

Definition

A white noise is a kind of random signal that has no correlation between two different time instants. Its Autocorrelation can be expressed as
$$R_{XX}(\tau) = \frac{N_0}{2}\delta(\tau) \tag{60}$$

- An I.I.D. signal (Independent et Identically Distributed) is a white noise.
- The Power Spectral Density of a White noise is
$$S_{XX}(f) = \mathcal{F}[R_{XX}(\tau)] = \frac{N_0}{2} \tag{61}$$
- For a discrete time signal the signal has an average power $P_X = \frac{N_0}{2}$.

Convolution of random signals



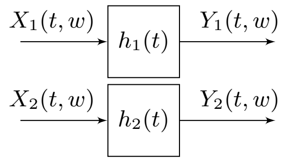
Transformation of the mean

If $X(t,w)$ is a wide sens stationary stochastic process then the mean of the output signal is
$$m_Y = E\left[\int_{-\infty}^{\infty} X(u,w)h(t-u)du\right] = m_X H(0)$$

Where $H(f)$ is the Fourier transform of $h(t)$:
$$H(f) = \int_{-\infty}^{\infty} h(t)e^{-2j\pi ft}dt \tag{62}$$

We recover $H(0)$ also known as the static gain.

Convolution, autocorrelation and PSD



- Let $X_1(t,w)$ and $X_2(t,w)$ be two WSS stochastic processes.
- Let $h_1(t)$ and $h_2(t)$ be the impulse response of two LTI systems.

$$Y_1(t,w) = (X_1 * h_1)(t,w)$$
$$Y_2(t,w) = (X_2 * h_2)(t,w)$$

Convolution formula for random signals

- The covariances between the filtered signals above can be expressed as
$$R_{Y_1Y_2}(\tau) = (h_1 * R_{X_1X_2} * h_2^{*-})(\tau) \tag{63}$$
where $h_2^-(t) = h_2(-t)$.
- Which gives in the Fourier domain the following relation
$$S_{Y_1Y_2}(f) = H_1(f)S_{X_1X_2}(f)H_2^*(f) \tag{64}$$
where $H_1(f)$ and $H_2(f)$ are respectively the FT of the impulse responses $h_1(t)$ and $h_2(t)$.

Proof of the convolution formula (1)

$$\begin{aligned} R_{Y_1Y_2}(\tau) &= E[Y_1(t,w)Y_2(t-\tau,w)] \\ &= E[(X_1 * h_1)(t,w)(X_2 * h_2)(t-\tau,w)] \end{aligned}$$

By definition

$$(X_1 * h_1)(t,w) = \int_{-\infty}^{\infty} X_1(u,w)h_1(t-u)du = \int_{-\infty}^{\infty} X_1(t-u,w)h_1(u)du$$

$$(X_2 * h_2)(t-\tau,w) = \int_{-\infty}^{\infty} X_2(v,w)h_2(t-\tau-v)dv = \int_{-\infty}^{\infty} X_2(t-\tau-v,w)h_2(v)dv$$

We can deduce that

$$\begin{aligned} R_{Y_1Y_2}(\tau) &= E\left[\int_{-\infty}^{\infty} X_1(t-u,w)h_1(u)du \int_{-\infty}^{\infty} X_2^*(t-\tau-v,w)h_2^*(v)dv\right] \\ &= E\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(u)X_1(t-u,w)X_2^*(t-\tau-v,w)h_2^*(v)dudv\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(u)E\left[X_1(t-u,w)X_2^*(t-\tau-v,w)\right]h_2^*(v)dudv \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(u)R_{X_1X_2}(\tau+v-u)h_2^*(v)dudv \end{aligned}$$

Proof of the convolution formula (2)

We found

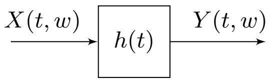
$$R_{Y_1Y_2}(\tau) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_1(u)R_{X_1X_2}(\tau+v-u)h_2^*(v)dudv$$

which corresponds to a convolution by h_1

$$\begin{aligned} R_{Y_1Y_2}(\tau) &= \int_{-\infty}^{\infty} (h_1 * R_{X_1X_2})(\tau+v)h_2^*(v)dv \\ &= \int_{-\infty}^{\infty} (h_1 * R_{X_1X_2})(u)h_2^*(u-\tau)du \\ &= \int_{-\infty}^{\infty} (h_1 * R_{X_1X_2})(u)h_2^{*-}(\tau-u)du \\ &= (h_1 * R_{X_1X_2} * h_2^{*-})(\tau) \end{aligned}$$

where $h_2^-(t) = h_2(-t)$.

Filtering of a random signal



Convolution formula

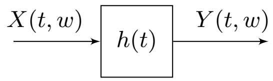
- Special case where $h_1(t) = h_2(t) = h(t)$ and $X_1(t,w) = X_2(t,w) = X(t,w)$
- The autocorrelation becomes

$$R_{YY}(\tau) = (h * R_{XX} * h^{*-})(\tau) \quad (65)$$

- In the frequency domain :

$$S_{YY}(f) = S_{XX}(f)|H(f)|^2 \quad (66)$$

Power in a frequency band



Power of $X(t,w)$ in the frequency band $[f_1, f_2]$

- Let the ideal filter

$$H_{[f_1,f_2]}(f) = \begin{cases} 1 & \text{if } f_1 < |f| < f_2 \\ 0 & \text{sinon} \end{cases} \quad (67)$$

- The power of $Y(t,w)$ after filtering is

$$\begin{aligned} P_Y &= \int_{-\infty}^{\infty} S_{YY}(f)df = \int_{-\infty}^{\infty} S_{XX}(f)|H(f)|^2 df \\ &= \int_{-f_2}^{-f_1} S_{XX}(f)df + \int_{f_1}^{f_2} S_{XX}(f)df = 2\int_{f_1}^{f_2} S_{XX}(f)df \end{aligned}$$

- **Power in the band** $[f_1, f_2]$**:**

$$P_{[f_1,f_2]} = 2\int_{f_1}^{f_2} S_{XX}(f)df \quad (68)$$

Discrete time signals

- We now focus on discrete time processes $X[n]$.
- The Autocovariance for WSS signal X in this case is

$$R_X[n] = E[X[k]X[k+n]]$$

- The periodic Power Spectral Density (PSD) of a WSS signal X is

$$S_X(e^{j2\pi f}) = \sum_{n=-\infty}^{+\infty} R_X[n]e^{-2j\pi fn}$$

- The Autocorrelation can be recovered with the inverse DTFT

$$R_X[n] = \int_0^1 S_X(e^{j2\pi f})e^{2j\pi fn}df$$

- A white noise process $W[n]$ has an autocorrelation and PSD respectively of

$$R_W[n] = \sigma^2\delta[n], \qquad S_W(e^{j2\pi f}) = \sigma^2$$

Autoregressive model

Definition

Let $W[n]$ be a WSS white noise of variance σ^2. The Autoregressive (AR) model is defined as the following recurrent relation

$$X[n] + \sum_{k=1}^{N} a[k]X[n-k] = W[n], \tag{69}$$

- This equation defined a linear regression for $X[n]$ using its N values $X[n-k]$ from the past.
- $W[n]$ is an additive IID noise often called the "innovation" term in french.
- The regression parameters $a[n]$ define the parameters of the model.
- When $a[0] = 1$, $a[n]$ defines a finite impulse response filter and Eq. 69 can be reformulated as

$$X \star a[n] = W[n] \tag{70}$$

- The filter $a[n]$ is said to perform a signal whitening (or decorrelation) of X.

Autoregressive model (2)

Transfer functions and convolution

The whitening filter $a[n]$ has the following transfer function

$$A[z] = \sum_{n=0}^{N} a[n]z^{-n}$$

and its inverse of impulse response $a_i[n]$ has the following transfer function

$$A_i[z] = \frac{1}{A[z]} = \frac{1}{\sum_{k=0}^{N} a[n]z^{-n}}$$

Stationary component

We can show that $\forall n \geq 0$

$$X[n] = a_i \star W[n] + Y[n]$$

- $a_i \star W[n]$ is the stationary component (convolution of a stationary signal).
- $Y[n]$ is a transient component such that

$$Y \star a[0] = 0$$

Autoregressive model (3)

Theorem

$$X[n] = a_i \star W[n] + Y[n]$$

The process $X[n]$ converges toward $a_i \star W[n]$ for large n if the zeros of $A(z)$ are strictly in the unit circle $|z| < 1$.

Proof

Note that the solutions of the homogeneous equation

$$Y \star a[n] = \sum_{k=0}^{N} a[k]Y[n-k] = 0. \tag{71}$$

can be expressed as

$$Y[n] = \sum_{k=0}^{N-1} A_k(c_k)^n \ ,$$

Where A_k depends only on initial values $Y[k]$ for $k = -N, \dots, 0$ and c_k are the zeros of the following equation

$$\hat{a}(z) = \sum_{k=0}^{N} a[k]z^{-k} = 0. \tag{72}$$

For $\lim_{n\to\infty} Y[n] = 0$ we need that $|c_k| < 1 \ \forall 0 \leq k \leq N$

Autoregressive model (4)

$$X[n] + \sum_{k=1}^{N} a[k]X[n-k] = X \star a[n] = W[n],$$
$$X[n] = a_i \star W[n] + Y[n],$$

Autocorrelation and PSD

- The autocorrelation of the stationary component of X can be expressed as

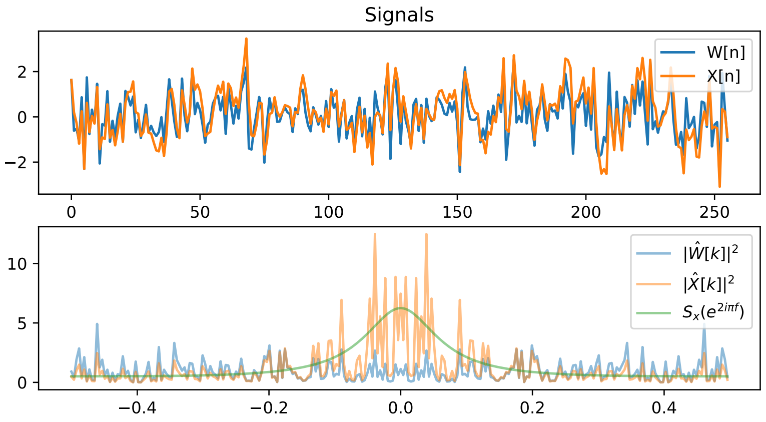
$$R_X[n] = a_i[n] \star R_W[n] \star a_i[-n] = \sigma^2 a_i[n] \star a_i[-n] \quad (73)$$

  That is the correlation (convolution with time mirror) of a_i.
- The Power Spectral Density of the stationary component of $X[n]$ can be expressed as

$$S_X(e^{2j\pi f}) = |A_i(e^{2j\pi f})|^2 S_W(e^{2j\pi f}) = \frac{\sigma^2}{|\sum_{n=0}^{N} a[n]e^{-2j\pi f n}|^2} \quad (74)$$

- In the following we will suppose that $X[n]$ is stationary and will suppose that the transient component is 0.

Autoregressive model example



- $\sigma^2 = 1$, $N = 2$
- $\mathbf{a} = [-.5, -.1]$ (low pass).
- Comparison between spectrum with finite number of samples $N_s = 256$ and theoretical spectrum.

AR model and covariance matrix

$$X[n] + \sum_{k=1}^{N} a[k]X[n-k] = X \star a[n] = W[n],$$

Autocovariance

Let us suppose that $X[n]$ is a stationary signal we get from the equation above:

$$X[n-l]X[n] + \sum_{k=1}^{N} a[k]X[n-k]X[n-l] = W[n]X[n-l]$$

By taking the expectation on each side of the equation we get

$$E[X[n-l]X[n]] + \sum_{k=1}^{N} a[k]E[X[n-k]X[n-l]] = E[W[n]X[n-l]]$$

Since $X[n]$ is supposed to be WSS we have for $l > 0$

$$R_X[l] + \sum_{k=1}^{N} a[k]R_X[l-k] = 0$$

AR model and Yule-Walker equations

Yule-Walker equations

$$R_X[l] + \sum_{k=1}^{N} a[k]R_X[l-k] = 0$$

- The equation above can be used to recover N equations with $l = 1, \dots, N$.
- This provides us with the following linear system:

$$\mathbf{R}_X \mathbf{a} = -\mathbf{r}_X$$

with

$$\mathbf{R}_X = \begin{bmatrix} R_X[0] & R_X[-1] & . & R_X[1-N] \\ R_X[1] & R_X[0] & . & R_X[2-N] \\ . & . & . & . \\ . & . & . & . \\ R_X[N-1] & R_X[N-2] & . & R_X[0] \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a[1] \\ a[2] \\ . \\ ,. \\ a[N] \end{bmatrix}, \quad \mathbf{r}_X = \begin{bmatrix} R_X[1] \\ R_X[2] \\ . \\ ,. \\ R_X[N] \end{bmatrix}$$

AR model estimation

Principle

The Yule-Walker equations provide us with the following linear system:

$$\mathbf{R}_X \mathbf{a} = -\mathbf{r}_X$$

If $\mathbf{R}_X$ is invertible then one can recover the coefficients in $\mathbf{a}$ with:

$$\mathbf{a} = -\mathbf{R}_X^{-1} \mathbf{r}_X$$

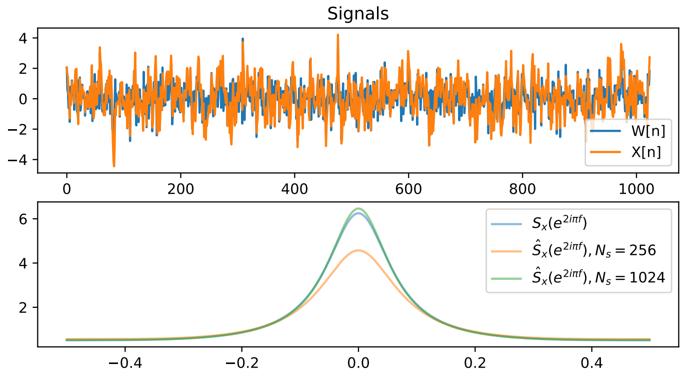
and the variance of the IID noise $W[n]$ can be estimated with

$$\sigma^2 = E[W[n]^2] = R_X[0] + \sum_{k=1}^{N} a[k] R_X[k]. \tag{75}$$

In practice

- In practice one does not have access to $\mathbf{R}_X$ and $\mathbf{r}_X$.
- When both stationarity and ergodicity are supposed one can estimate the auto-correlation $\hat{\mathbf{R}}_X$ and $\hat{\mathbf{r}}_X$ with temporal averaging.
- The computation is done on a finite number of samples N_s.
- Empirical estimation of autocorrelation in python: `np.signal.correlate`.
- Can be estimated with `scipy.linalg.solve` (but should not see Levinson-Durbin).

Example of AR estimation



- $\sigma^2 = 1,\ N = 2$
- $\mathbf{a} = [-.5, -.1]$ (low pass).
- Estimation of AR coefficient from finite signal of size N_s.
- $\hat{\mathbf{a}} = [-0.4307, -0.1016]$ for $N_s = 256$
- $\hat{\mathbf{a}} = [-0.4943, -0.1123]$ for $N_s = 1024$

Wiener filtering

Objective and linear model

- $Y[n]$ is a WSS stochastic process and $X[n]$ is an indirect observation of $Y[n]$ (usually additive noise).
- We want to find the optimal FIR filter $h[n]$ of support $[0, N]$ that reconstructs $\hat{Y}[n]$ from $X[n]$ with the linear model

$$\hat{Y}[n] = \sum_{k=0}^{N} h[k] X[n-k] = X \star h[n] \tag{76}$$

Optimization problem for Wiener filtering

The optimal filtering $\tilde{h}[n]$ is the one minimizing the expected square error:

$$\tilde{h} = \underset{h}{\text{argmin}}\, E\left[(Y[n] - X \star h[n])^2\right] \tag{77}$$

Classical linear regression problem : cf MAP 535 (Regression)

Geometric properties

$L_2(\mathbb{R})$ Space

- The scalar product between two random variables A and B in $L_2(\mathbb{R})$ can be expressed as :

$$< A, B >_{L_2(\mathbb{R})} = E[AB]$$

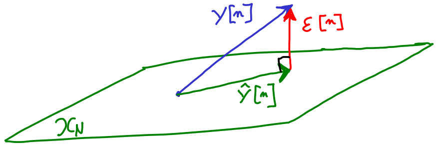
- The norm is this space can be expressed with the scalar product:

$$\|A\|^2_{L_2(\mathbb{R})} = E[A^2]$$

Geometry of stochastic processes and linear model

- We suppose that $X[n]$ and $Y[n]$ are centered WSS and belong to $L_2(\mathbf{R})$.
- This means that $E[X[n]^2] < \infty$ and $E[Y[n]^2] < \infty$
- Let $\mathcal{X}_N \subset L_2(\mathbf{R})$ the set of all linear combinations of $\{X[n-k]\}_{0 \leq k < N}$.
- By construction $\hat{Y}[n] = \sum_{l=0}^{N} \tilde{h}[l] X[n-l] \in \mathcal{X}_N$.

Wiener filter and orthogonal projection



Orthogonal Projection

- The residual for the optimal filter $\tilde{h}$ is

$$\mathcal{E}[n] = Y[n] - \hat{Y}[n] = Y[n] - \sum_{k=0}^{N} \tilde{h}[n]X[n-k] \tag{78}$$

- Minimizing $E\left[(Y[n] - X \star h[n])^2\right] = \|Y[n] - X \star h[n]\|^2_{L_2(\mathbb{R})}$ is an orthogonal projection of $Y[n]$ onto $\mathcal{X}_N$.
- Due to the projection, for any $0 \leq k \leq N$ we have

$$< \mathcal{E}[n], X[n-k] >_{L_2(\mathbf{R})} = E\left[\left(Y[n] - \sum_{l=0}^{N} \tilde{h}[l]X[n-l]\right) X[n-k]\right] = 0 \tag{79}$$

because $\mathcal{E}[n]$ is orthogonal to each of the random variables $\{X[n-k]\}_{0 \leq k < N}$.

Wiener-Hopf equations

- The orthogonality conditions can be expressed for WSS signals for $0 \leq k \leq N$ as:

$$\begin{aligned}
0 =& E\left[\left(Y[n] - \sum_{l=0}^{N} \tilde{h}[l]X[n-l]\right) X[n-k]\right] \\
=& E\left[Y[n]X[n-k]\right] - \sum_{l=0}^{N} \tilde{h}[l] E\left[X[n-k]X[n-l]\right] \\
=& R_{XY}[k] - \sum_{l=0}^{N} \tilde{h}[l] R_X[k-l]
\end{aligned}$$

- The orthogonality conditions define a linear system of size $N+1$ called the Wiener-Hopf equations:

$$\mathbf{R}_X \mathbf{h} = \mathbf{r}_{XY} \tag{80}$$

with

$$\mathbf{R}_X = \begin{bmatrix} R_X[0] & R_X[-1] & . & R_X[-N] \\ R_X[1] & R_X[0] & . & R_X[1-N] \\ . & . & . & . \\ . & . & . & . \\ R_X[N] & R_X[N-1] & . & R_X[0] \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} h[0] \\ h[1] \\ . \\ . \\ h[N] \end{bmatrix}, \quad \mathbf{r}_{XY} = \begin{bmatrix} R_{XY}[0] \\ R_{XY}[1] \\ . \\ . \\ R_{XY}[N] \end{bmatrix}$$

Wiener filter estimation

Estimation of $\tilde{\mathrm{h}}$

- When the autocorrelation $\mathbf{R}_X$ and the cross-correlation $\mathbf{r}_{XY}$ are known the optimal filter can be estimated by solving the linear system $\mathbf{R}_X \mathbf{h} = \mathbf{r}_{XY}$.
- When the random variables $\{X[n-k]\}_{0 \leq k < N}$ are linearly independent the matrix $\mathbf{R}_X$ is invertible and

$$\tilde{\mathbf{h}} = \mathbf{R}_X^{-1} \mathbf{r}_{XY} \tag{81}$$

Estimation in practice

- If $\mathbf{R}_X$ is not invertible, one usually decreases N until $\{X[n-k]\}_{0 \leq k < N}$ are linearly independent.
- Empirical estimation $\hat{\mathbf{R}}_X$ can be estimated from the observed signal $X[n]$.
- $\mathbf{r}_{XY}$ requires both $X[n]$ and $Y[n]$ to be "known" or can be estimated empirically from a finite sampling of both signals.

Estimation error and non centered signals

Estimation error

The estimation error of the Wiener filter can be expressed as

$$\begin{aligned}
E[(Y[n] - \hat{Y}[n])^2] =& E[Y[n]^2] - E[\tilde{Y}[n]^2] = E[Y[n]^2] - E[\tilde{Y}[n](Y[n] - E[n])] && (82) \\
=& E[Y[n]^2] - \sum_{l=0}^{N} h[l] E[Y[n]X[n-l]] && (83) \\
=& R_Y[0] - \sum_{l=0}^{N} h[l] R_{XY}[l] && (84)
\end{aligned}$$

Where (82) and (83) are due to the Pythagorean theorem and orthogonality respectively.

Non-centered signals

When $X[n]$ and $Y[n]$ are non centered WSS such that $E[X[n]] = \mu_X$ and $E[Y[n]] = \mu_Y$ then the linear relation becomes

$$\tilde{Y}[n] = \mu_Y + \sum_{l=0}^{N} h[l](X[n-l] - \mu_X) \tag{85}$$

and all the Wiener-Hopf equations still hold.

Example of Wiener filtering (1)

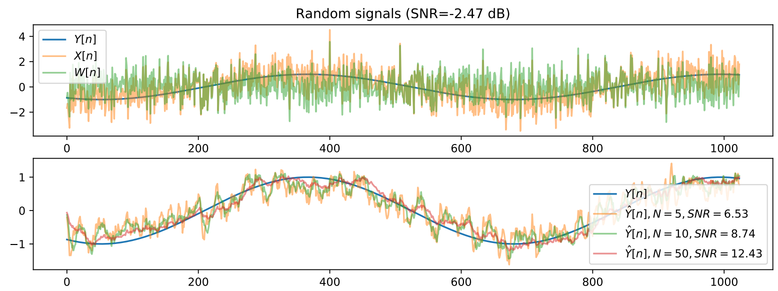
Known random signals

- $Y[n]$ is a stochastic process with known w_0 and a uniform random variable $\phi \sim \mathcal{U}(0, 2\pi)$ such that
$$Y[n] = \cos(w_0 n + \phi)$$
- The observed signal $X[n]$ contains an additive IID noise such that $W[n] \sim \mathcal{N}(0, \sigma^2)$ and can be expressed as
$$X[n] = Y[n] + W[n]$$
- The random variables ϕ and $W[n]\ \forall n$ are all considered independent.

Computation of the correlations

- $R_X[k] = E[X[n]X[n+k]] = \frac{1}{2}\cos(w_0 k) + \sigma^2\delta[k]$
- $R_{XY}[k] = E[X[n]Y[n+k]] = \frac{1}{2}\cos(w_0 k)$
- Here we have access to the exact correlations and do not need to estimate them.

Example of Wiener filtering (2)



- Realization for a finite signal of length $N_s = 1024$, noise level $\sigma^2 = 1$.
- Use of known theoretical auto and cross-correlations to estimate the wiener filter.
- Computation of filters for $N = \{5, 10, 50\}$ and corresponding SNR.

Linear prediction

Objective

- Predict $X[n]$ from a weighted sum of $\{X[n-k]\}_{1 \le k \le N}$ (note that $k \neq 0$).
- The model is the following:
$$\hat{X}[n] = \sum_{k=1}^{N} a_N[k] X[n-k] \tag{86}$$
- Optimal filter minimizes the expected square error.
- Similar problem to Wiener filtering with $Y[n] = X[n]$.

Estimation and equivalence to AR models

- The Wiener Hopf equations for the model are:
$$R_X[k] - \sum_{l=1}^{N} \tilde{a}_N[l] R_X[k-l] = 0$$
- We recover exactly the Yule-Walker equations for the AR model with a change in sign $a[l] = -a_N[l]$.

Solving the linear system (1)

Linear system with Toeplitz matrix

The linear methods discussed before (AR, Wiener, Linear prediction) exhibit a linear system of the form:
$$\mathbf{Rh} = \mathbf{r}$$
where $\mathbf{R}$ is a symmetric Toeplitz matrix.

Solving the system

- Solving general linear system of size N requires $\mathcal{O}(N^3)$ operation.
- Since $\mathbf{R}$ is Toeplitz, one can use a $\mathcal{O}(N^2)$ method called Levinson-Durbin recursion (see Sec. 5.2 [Mallat et al., 2015]) to diagonalize $\mathbf{R}$ with
$$\mathbf{D} = \mathbf{L}\mathbf{R}_X\mathbf{L}^T$$
whe $\mathbf{L}$ is a lower triangular matrix leading to the solution:
$$\hat{\mathbf{h}} = \mathbf{R}^{-1}\mathbf{r} = \mathbf{L}^{\mathbf{T}}\mathbf{D}^{-1}\mathbf{L}\mathbf{r}.$$
- Equivalent to a Gram-Schmidt orthogonalization.
- When the matrix $\mathbf{R}$ is not invertible, it means that some random variables are linearly dependent and suggests that N is too large (an infinite number of solution exist).

Solving the linear system (2)

Linear system from finite signals

- Finite signals of size $N+1$ (1D or 2D) are considered to be periodic.
- In this case causality might not be necessary.

Solving the system with FFT

- With periodic signals the covariance matrix $\mathbf{R}$ is Toeplitz circulant and can be diagonalized with the Discrete Fourier Transform.
- The solution can be computed with

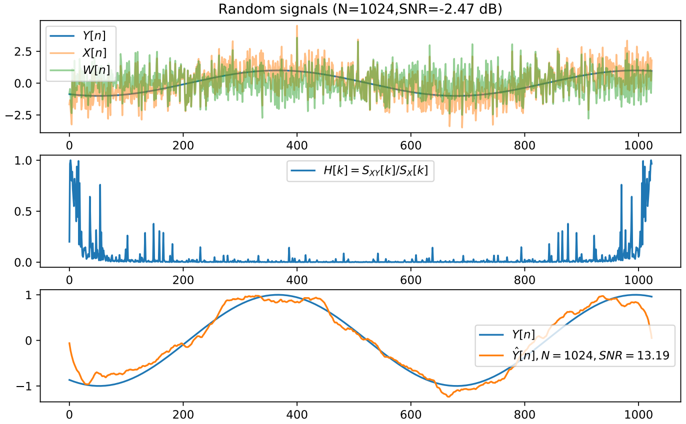
$$\hat{h}[n] = IFFT\left(\frac{S_{XY}[k]}{S_X[k]}\right) = IFFT\left(\frac{S_Y[k]}{S_Y[k]+S_W[k]}\right) \tag{87}$$

  where the second equality is true when $W[n]$ and $Y[n]$ are centered independent.
- The PSD can be estimated with FFT with the following formula:

$$S_X[k] = |FFT(X[n])|^2 \tag{88}$$

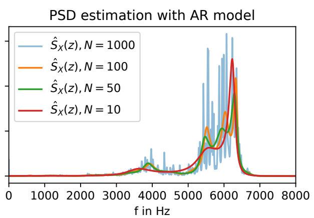
- The complexity of Wiener filtering (estimation+filtering) is $\mathcal{O}(N\log_2(N))$.

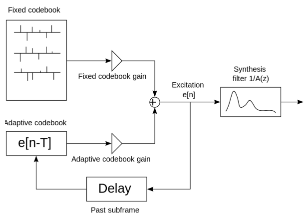
Example of Wiener filtering with FFT solver

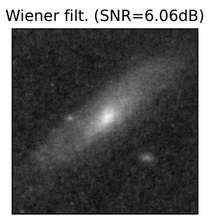


- Finite signals of length $N_s = 1024$, noise level $\sigma^2 = 1$.
- Computation of filters for $N = 1024$ with FFT and corresponding SNR.
- Border effects when the signal is not really periodic.

Applications of linear modeling



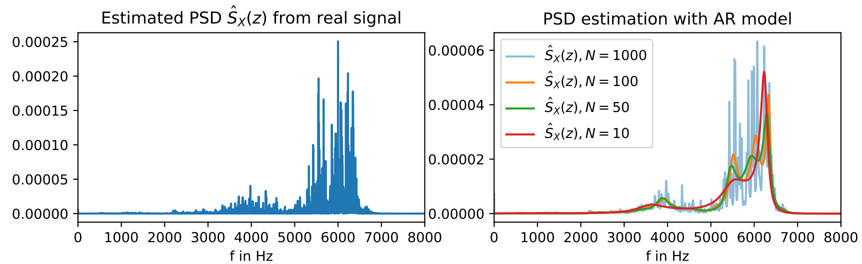




Applications for linear models on stochastic processes

- Smooth PSD estimation with AR modeling.
- Modeling speech signal with AR/linear prediction models.
- Denoising in 1D and 2D with Wiener filtering.
- Deconvolution in the presence of noise.

Smooth PSD estimation with AR modeling



Principle

- Autocorrelation and PSD estimation for finite signals of size N_s usually brings a lot of noise and makes them hard to interpret.
- One can estimate a small number of parameters N of an AR model to estimate a smooth version of the PSD:

$$\hat{S}_X(z) = \frac{\sigma^2}{\sum_{k=0}^{N} a[k]z^{-k}}$$

- The order N can tune the complexity of the PSD.

Speech modeling with linear models

Speech signals

- Speech signals are NOT stationary.
- But they can be supposed stationary on a small temporal window of a few ms.

Acoustic theory of speech production

- On small windows speech is modeled with a source-filter model [Fant, 1970].

$$X(z) = Y(z)H(z)$$

that is a (locally) LTI model where

  - $Y(z)$ is the Z-transform of a mixture of excitation signals.
  - $H(z)$ is the transfer function of the resonant filter due to the vocal, nasal, and pharyngeal tracts.

- $H(z)$ can be modeled as a linear prediction with

$$H(z) = \frac{1}{A(z)} = \frac{1}{\sum_{k=0}^{N} a[k]z^{-k}}$$

where N is relatively small (of the order of 10)

- Both $Y(z)$ and $H(z)$ change for each temporal window.

Voiced and unvoiced excitation signals

Voiced sounds

- Sounds made while pronouncing vowels (a, e, i, o, u).
- They have a pitch $1/T_0$ (fundamental frequency) that is tuned by the vocal chords vibrations.
- The excitation is modeled as as Dirac comb:

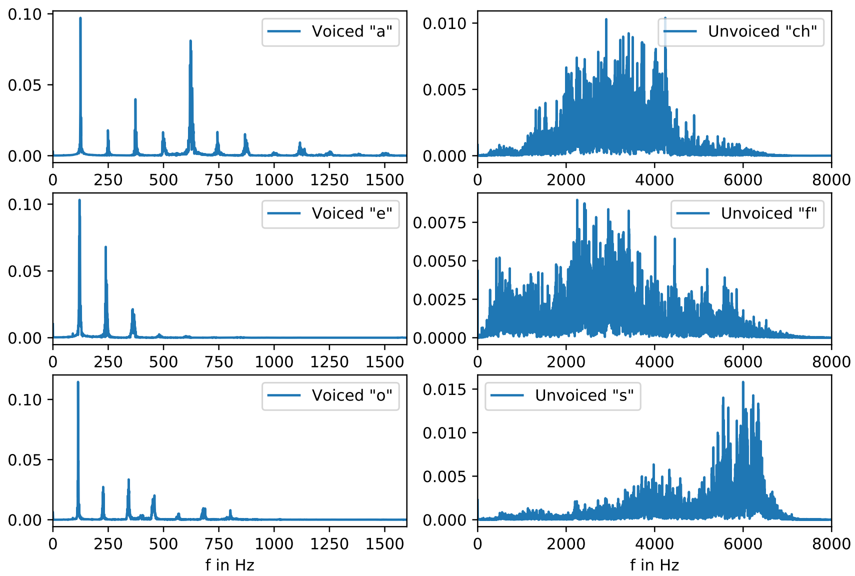
$$Y[n] = \sum_{k=-\infty}^{\infty} \delta[n - kT_0]$$

Unvoiced sounds

- Sounds made while pronouncing consonants (f, ch, s ,v).
- They are produced by the flow of air through the vocal track.
- The excitation is modeled as IID Gaussian noise:

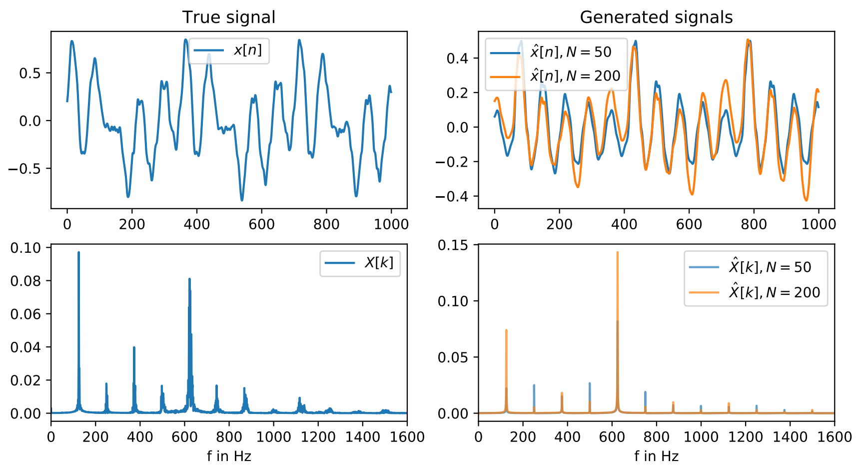
$$Y[n] \sim \mathcal{N}(0, \sigma^2)$$

Voiced and unvoiced sounds examples



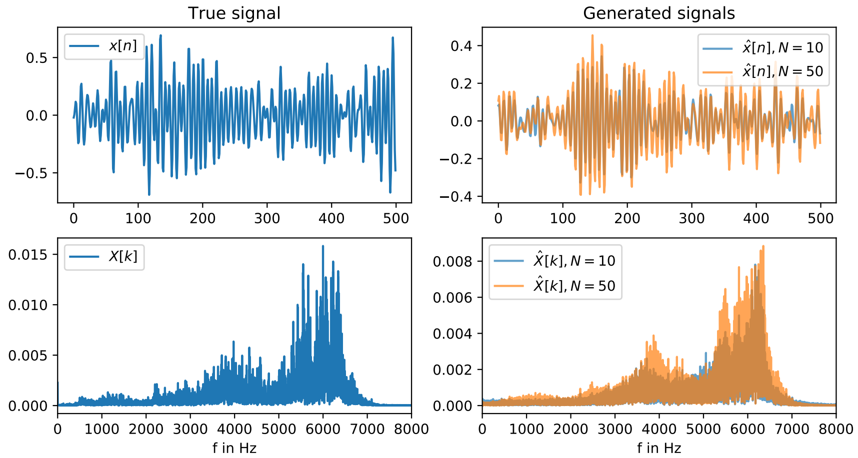
We represent the FFT of the signals with frequency in Hz.

Voiced signal generation example



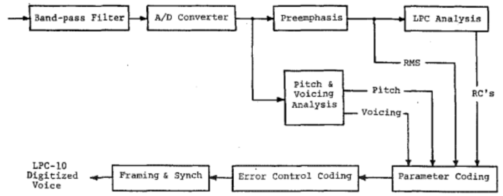
- Voiced sound "a", sampled at 44100Hz.
- Estimation of AR coefficients for $N = 50$ (1ms) and $N = 200$ (4ms).
- Generate signals by filtering a Dirac comb of period 8ms (inverse of 124.6Hz).

Unvoiced signal generation example



- Voiced sound "s", sampled at 44100Hz.
- Estimation of AR coefficients for $N = 10$ (0.2ms) and $N = 50$ (1ms).
- Generate signals by filtering gaussian noise.

Linear Predictive Coding Vocoder



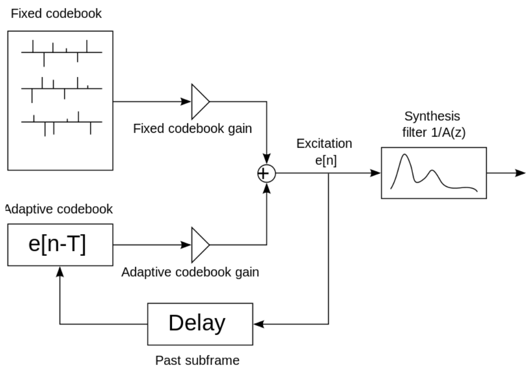
Principle

- Compress speech signal for telecommunications.
- Split the signal in small windows.
- Detect voiced or unvoiced speech and estimate AR coefficients.
- Perform quantization of all estimated values (pitch/AR coefficients)

Standard LPC-10 (FIPS 137 [FIPS, 1984])

- Bit rate of 2400bits/sec to store speech at 8000KHz.
- Windows of $N_s = 180$ samples (22.5ms) and and AR of order $N = 10$.
- Requires 20 MIPS of processing power, 2 kilobytes of RAM.

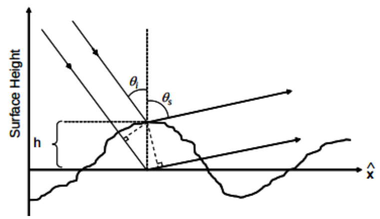
Code Excited Linear prediction (CELP)

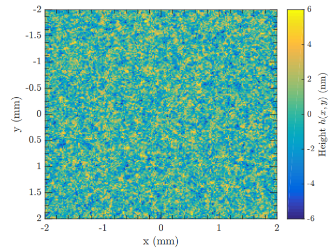


Principle [Schroeder and Atal, 1985]

- Similar to LPC but allows better quality with finer modelization of excitation.
- Excitation is a sum of fixed codebooks and adaptive pitch.
- Variants of CELP are used for GSM, internet VOIP.

Surface roughness in imperfect optics





Principle [Harvey et al., 2007]

- In optics lenses and mirrors are not perfect.
- Scattering of the light occurs due to micro-structures on the surface.
- Scattering of a planar wave can be modeled as a phase proportional to the micro-structure $H(x, y)$.
- One can fit a linear model such that $H = W \star F$ where W is a IID noise and F is a LTI filter describing the scattering.
- Used to model imperfect optics in telescopes and solar coronagraphs [Rougeot et al., 2019].

Denoising with Wiener filtering

Model for additive noise

- Observed signal $X[n]$ contains signal of interest $Y[n]$ and additive noise $W[n]$:

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

- Wiener filter aim at finding a FIR filter $h[n]$ of order N such that

$$\hat{Y}[n] = X[n] \star h[n] \approx Y[n]$$

- When $Y[n]$ and $W[n]$ and independent and $W[n]$ centered then

$$R_X[n] = R_Y[n] + R_W[n]$$

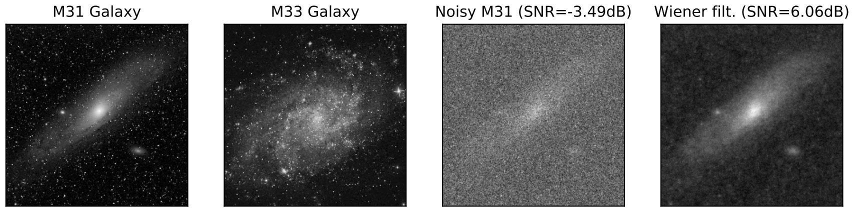
Simplification for additive IID noise

- When $W[n]$ is IID its autocorrelation is

$$R_W[n] = \sigma^2 \delta[n]$$

  and only σ^2 has to be known or estimated from the data.
- This corresponds to a correlation matrix $\mathbf{R}_X = \mathbf{R}_Y + \sigma^2 \mathbf{I}$ making it invertible for $\sigma^2 > 0$.

Wiener filtering for noisy images



- Observed X is a noisy (additive IID noise) $N \times N$ image of M31 Galaxy.
- Autocorrelation $\hat{R}_Y[n]$ estimated from clean version of the M33 Galaxy.
- Autocorrelation of noise is set to $\hat{R}_W[n] = \hat{\sigma}^2 \delta[n]$.
- Wiener filter obtained by FFT and circular convolution:

$$\hat{h}[n] = IFFT\left(\frac{\hat{S}_Y[k]}{\hat{S}_Y[k] + N^2\hat{\sigma}^2}\right)$$

Deconvolution in the presence of noise

Model for Wiener filtering

- Observed signal $X[n]$ contains signal of interest $Y[n]$ convolved by a filter $g[n]$ and additive noise $W[n]$:

$$X[n] = Y[n] \star g[n] + W[n]$$

- Wiener filter aim at finding a FIR filter $h[n]$ of order N such that

$$\hat{Y}[n] = X[n] \star h[n] = Y[n] \star g[n] \star h[n] + W[n] \star h[n] \approx Y[n]$$

- The filter $h[n]$ aim at both inverting the convolution by $g[n]$ but also limit the impact of the noise

Wiener filter in the Fourier domain

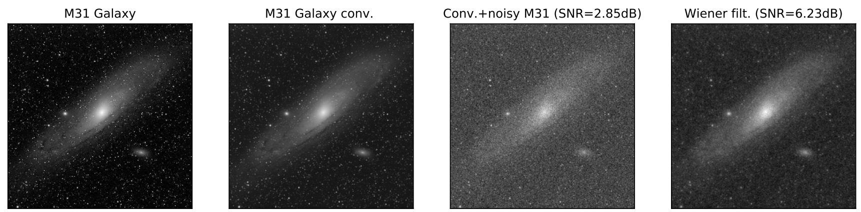
- The correlations can be expressed when $Y[n]$ and $W[n]$ are independent as

$$R_X[n] = (g \star R_Y \star g^-)[n] + R_W[n], \quad R_{XY}[n] = g \star R_Y[n]$$

- The optimal Wiener filter can be estimated in the Fourier domain as

$$\hat{H}[k] = \frac{S_Y[k]G[k]}{S_Y[k]|G[k]|^2 + S_W[k]}$$

Deconvolution with Wiener filtering



- Observed X is an image of M31 Galaxy convolved bu g and with additive noise.
- g is the Point Spread Function observation by a circular telescope.
- Autocorrelation of noise is set to $\hat{R}_W[n] = \hat{\sigma}^2 \delta[n]$.
- Autocorrelation $\hat{R}_Y[n]$ estimated from clean version of the M33 Galaxy.
- Wiener filter use the known operator g and the theoretical expression for noise PDF.

Bibliography

- Random variables and stochastic processes [Papoulis, 1965].
- Discrete-time signal processing [Oppenheim and Shafer, 1999].
- Signals and Systems [Haykin and Van Veen, 2007].
- Signals and Systems [Oppenheim et al., 1997].
- Signal Analysis [Papoulis, 1977].
- Fourier Analysis and its applications [Vretblad, 2003].

- Polycopiés from Stéphane Mallat and Éric Moulines [Mallat et al., 2015].
- Théorie du signal [Jutten, 2018].
- Distributions et Transformation de Fourier [Roddier, 1985]

References I

[Fant, 1970] Fant, G. (1970).
Acoustic theory of speech production.
Number 2. Walter de Gruyter.

[FIPS, 1984] FIPS, F. I. P. S. (1984).
Fips pub 137, analog to digital conversion of voice by 2,400 bit/second linear predictive coding.

[Harvey et al., 2007] Harvey, J. E., Krywonos, A., and Stover, J. C. (2007).
Unified scatter model for rough surfaces at large incident and scatter angles.
In *Advanced Characterization Techniques for Optics, Semiconductors, and Nanotechnologies III*, volume 6672, page 66720C. International Society for Optics and Photonics.

[Haykin and Van Veen, 2007] Haykin, S. and Van Veen, B. (2007).
Signals and systems.
John Wiley & Sons.

[Jutten, 2018] Jutten, C. (2018).
Théorie di signal.
Univ. Grenoble Alpes - Polytech' Grenoble.

References II

[Mallat et al., 2015] Mallat, S., Moulines, E., and Roueff, F. (2015).
Traitement du signal.
Polycopié MAP 555, École Polytechnique.

[Oppenheim and Shafer, 1999] Oppenheim, A. V. and Shafer, R. W. (1999).
Discrete-time signal processing.
Prentice Hall Inc.

[Oppenheim et al., 1997] Oppenheim, A. V., Willsky, A. S., and Nawab, S. H. (1997).
Signals and systems prentice hall.
Inc., Upper Saddle River, New Jersey, 7458.

[Papoulis, 1965] Papoulis, A. (1965).
Random variables and stochastic processes.

[Papoulis, 1977] Papoulis, A. (1977).
Signal analysis, volume 191.
McGraw-Hill New York.

[Roddier, 1985] Roddier, F. (1985).
Distributions et transformée de fourier.

References III

[Rougeot et al., 2019] Rougeot, R., Flamary, R., Mary, D., and Aime, C. (2019).
Influence of surface roughness on diffraction in the externally occulted lyot solar coronagraph-application to aspiics.
Astronomy & Astrophysics, 626:A1.

[Schroeder and Atal, 1985] Schroeder, M. and Atal, B. (1985).
Code-excited linear prediction (celp): High-quality speech at very low bit rates.
In *ICASSP'85. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 10, pages 937–940. IEEE.

[Vretblad, 2003] Vretblad, A. (2003).
Fourier analysis and its applications, volume 223.
Springer Science & Business Media.