

## Signaux et Systèmes

### Chapitre 12

# Propriétés statistiques des signaux

Mai 2022

## TABLE DES MATIÈRES

---

**12.1 PROBABILITÉS**

**12.2 INFORMATION**

**12.3 PROCESSUS ALÉATOIRES**

**12.4 FILTRAGE DE SIGNAUX BRUITÉS**

## 12.1 PROBABILITÉS

- Variables aléatoires
- Vecteurs aléatoires
- Probabilités conditionnelles – indépendance
- Propriétés de l'opérateur d'espérance
- Fonction caractéristique
- Détecteur optimal

## Variables aléatoires

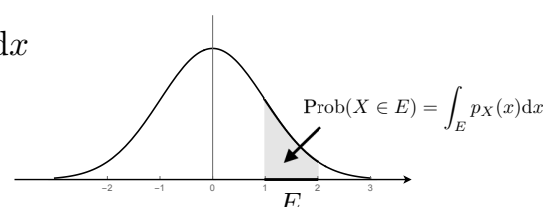
La notion d'aléatoire apparaît naturellement dans les systèmes de mesure

- parce que ce que l'on cherche à mesurer est susceptible de prendre plusieurs valeurs issues d'un processus (possiblement déterministe) inconnu de l'observateur
- parce que les systèmes physiques sont toujours affectés d'un **bruit**

La description mathématique repose sur les concepts de **variable aléatoire**, **vecteur aléatoire**, et **signal aléatoire**. Spécifiquement, on utilise les symboles  $X$ ,  $\mathbf{X}$ ,  $X(\cdot)$  et  $X[\cdot]$  (majuscules) pour désigner les contre-parties aléatoires des quantités déterministes  $x \in \mathbb{R}$ ,  $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$ ,  $x(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  et  $x[\cdot] : \mathbb{Z} \rightarrow \mathbb{R}$  (minuscules); ces dernières servent alors à décrire des **réalisations** du processus aléatoire.

Une variable aléatoire réelle  $X$  (scalaire) est caractérisée par sa **densité de probabilité**  $p_X : \mathbb{R} \rightarrow \mathbb{R}^+$ . Celle-ci permet de calculer la probabilité de tout événement  $E \subseteq \mathbb{R}$ .

$$\text{Prob}\{X \in E\} = \mathcal{P}(X \in E) = \int_E p_X(x) dx$$





# Espérance mathématique et produit scalaire

Soit  $f : \mathbb{R} \rightarrow \mathbb{R}$  une transformation (non-linéaire) de la variable aléatoire réelle  $X$ . Alors, la **valeur moyenne** (ou **espérance**) de  $f(X)$  est donnée par

$$\mathbb{E}_X \{f(X)\} = \mathbb{E}\{f(X)\} \triangleq \int_{\mathbb{R}} f(x) p_X(x) dx,$$

où l'intégrale peut aussi être interprétée comme un **produit scalaire**

$$\mathbb{E}\{f(X)\} = \langle p_X, f \rangle.$$

## ■ Relation avec les descripteurs statistiques fondamentaux

Fonction de répartition: la probabilité que  $X$  soit plus petit que  $x_0 \in \mathbb{R}$

$$\begin{aligned} x_0 \mapsto \mathcal{P}\{X \leq x_0\} &= \int_{-\infty}^{x_0} p_X(x) dx \\ &= (p_X * u)(x_0) = \langle p_X, u(x_0 - \cdot) \rangle = \mathbb{E}\{u(x_0 - X)\} \end{aligned}$$

Densité de probabilité

$$x \mapsto p_X(x) = \langle p_X, \delta(\cdot - x) \rangle = \mathbb{E}_X \{\delta(\cdot - x)\}$$

# Densité de probabilité et moments

## ■ Une densité de probabilité est **positive** et d'intégrale 1

$$p_X(x) \geq 0, \quad \int_{\mathbb{R}} p_X(x) dx = \mathbb{E}_X\{1\} = 1 = \mu_{X,0}$$

## ■ Valeur centrale ou moyenne

$$\mu_X = \mathbb{E}\{X\} = \int_{\mathbb{R}} x p_X(x) dx = \mu_{X,1}$$

## ■ Moment statistique d'ordre $n \in \mathbb{N}$

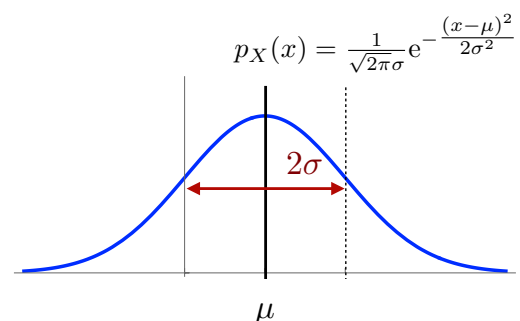
$$\mu_{X,n} = \mathbb{E}\{X^n\} = \int_{\mathbb{R}} x^n p_X(x) dx$$

## ■ Variance

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}\{(X - \mu_X)^2\} = \sigma_X^2 \\ &= \int_{\mathbb{R}} (x - \mu_X)^2 p_X(x) dx = \mu_{X,2} - \mu_{X,1}^2 \end{aligned}$$

## ■ Ecart type (ou dispersion)

$$\sigma_X = \sqrt{\mathbb{E}\{(X - \mu_X)^2\}}$$



# Exemples de densités de probabilité

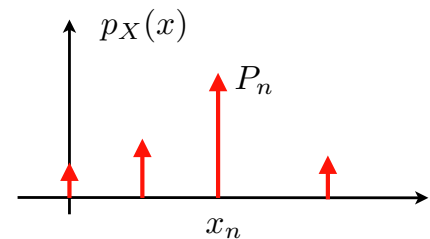
## ■ Densités de variables aléatoires de type “continues”

- loi uniforme sur  $[a, b]$  :  $p_{\text{uni}}(x; a, b) = \frac{1}{b-a} \text{rect}\left(\frac{x-a}{b-a} - \frac{1}{2}\right)$
- loi Gaussienne univariée:  $p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Leftrightarrow X \sim \mathcal{N}(\mu, \sigma)$
- loi exponentielle:  $p_{\text{exp}}(x; \lambda) = \lambda \cdot e^{-\lambda x} u(x)$

## ■ Variables aléatoires discrètes

Pour une variable aléatoire  $X$  qui ne prend que des valeurs **discrètes**  $(x_n)_{n=1}^N$ , on a

$$p_X(x) = \sum_{n=1}^N \mathcal{P}\{X = x_n\} \cdot \delta(x - x_n)$$



- loi binomiale:  $p_{\text{binom}}(x) = 2^{-N} \sum_{n=0}^N \binom{N}{n} \delta(x - x_n)$

# Extension: vecteurs aléatoires

$\mathbf{X} = (X_1, \dots, X_N)$  où  $X_1, \dots, X_N$  sont des variables aléatoires scalaires

## ■ Densité de probabilité $p_{\mathbf{X}} : \mathbb{R}^N \rightarrow \mathbb{R}^+$

- Probabilité d'un évènement:  $E$  est un sous-ensemble de  $\mathbb{R}^N$

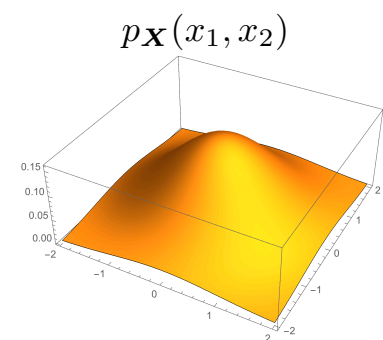
$$\mathcal{P}(E) = \text{Prob}(\mathbf{X} \in E) = \int_E p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

- Opérateur d'espérance ( $f$  est une fonction mesurable  $\mathbb{R}^N \rightarrow \mathbb{R}^M$ )

$$\mathbb{E}\{f(\mathbf{X})\} = \int_{\mathbb{R}^N} f(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

Vecteur moyenne:  $\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E}\{\mathbf{X}\} \in \mathbb{R}^N$

Matrice de covariance:  $\mathbf{C}_{\mathbf{X}} = \mathbb{E}\{(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T\} \in \mathbb{R}^{N \times N}$



## ■ Loi Gaussienne (ou normale) multivariée

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C}) \Leftrightarrow p_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{2\pi \det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

# Probabilités conditionnelles — indépendance

Soit  $p_{\mathbf{X}}(x, y)$  la densité de probabilité du vecteur aléatoire  $\mathbf{X} = (X, Y)$ . Alors  $X$  et  $Y$  sont de lois indépendantes ssi  $p_{\mathbf{X}}(x, y)$  peut se mettre sous la forme  $p_X(x)p_Y(y)$ ; c-à-d ssi

$$\frac{\partial^2 (\log p_{\mathbf{X}}(x, y))}{\partial x \partial y} = 0.$$

Ainsi, si  $X$  et  $Y$  sont indépendants, alors  $\mathbb{E}\{f(X) \cdot g(Y)\} = \mathbb{E}\{f(X)\} \cdot \mathbb{E}\{g(Y)\}$

## ■ Probabilités conditionnelles

On note  $p_{X|Y}(x|y)$  la densité de probabilité de  $X$  connaissant  $Y = y$ .

Les règles de Bayes donnent les formules suivantes avec  $\mathbf{X} = (X, Y)$

$$p_{\mathbf{X}}(x, y) = p_X(x) \cdot p_{Y|X}(y|x) = p_{X|Y}(x|y) \cdot p_Y(y)$$

$$\text{où } \begin{cases} p_X(x) = \int_{\mathbb{R}} p_{X,Y}(x, y) dy \\ p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x, y) dx \end{cases} \quad \text{sont les lois réduites de } X \text{ et } Y.$$

Bien sûr, si  $X$  et  $Y$  sont indépendants, alors  $p_{Y|X}(y|x) = p_Y(y)$  et  $p_{X|Y}(x|y) = p_X(x)$ .

# Propriétés de l'opérateur d'espérance

- Opérateur d'espérance:  $f : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathbb{E}\{f(X)\} = \int_{\mathbb{R}} f(x) p_X(x) dx = \langle f, p_X \rangle$$

- Espérance conjointe:  $f : \mathbb{R}^N \rightarrow \mathbb{R}$

$$\mathbb{E}\{f(X_1, \dots, X_N)\} = \int_{\mathbb{R}^N} f(x_1, \dots, x_N) p_{X_1, \dots, X_N}(x_1, \dots, x_N) dx_1 \cdots dx_N$$

- Linéarité

$$\mathbb{E}\{aX\} = a \mathbb{E}\{X\}, \quad \forall a \in \mathbb{C}$$

$$\mathbb{E}\{X_1 + X_2\} = \mathbb{E}\{X_1\} + \mathbb{E}\{X_2\} \quad (\text{les variables aléatoires } X_1 \text{ et } X_2 \text{ pouvant être dépendantes})$$

- Séparabilité:

$$\mathbb{E}\{\mathbf{X}\} = \mathbb{E}\{(X_1, \dots, X_N)\} = (\mathbb{E}\{X_1\}, \dots, \mathbb{E}\{X_N\})$$

## ■ Exemples d'utilisation

$$\mathbb{E}\{(aX - Y)^2\} = a^2 \mathbb{E}\{X^2\} + \mathbb{E}\{Y^2\} - 2a \mathbb{E}\{XY\}$$

$$\text{Cov}(X, Y) \triangleq \mathbb{E}\{(X - \mathbb{E}\{X\})(Y - \mathbb{E}\{Y\})\} = \mathbb{E}\{XY\} - \mathbb{E}\{X\} \mathbb{E}\{Y\}$$

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)$$

# Fonction caractéristique

Il est particulièrement utile de travailler avec la **fonction caractéristique** de la variable aléatoire. Il s'agit tout simplement de la transformée de Fourier de la densité de probabilité

$$P_X(\omega) = \mathbb{E} \{ e^{-j\omega X} \} = \int_{\mathbb{R}} p_X(x) e^{-j\omega x} dx = \mathcal{F}\{p_X\}(\omega)$$

## Théorème

Soient  $X_1, X_2, \dots, X_N$  des variables aléatoires **indépendantes** de lois  $p_1(x), p_2(x), \dots, p_N(x)$  et soit  $Y = X_1 + X_2 + \dots + X_N$ . Alors la densité de probabilité de  $Y$  est

$$p_Y(y) = (p_1 * p_2 * \dots * p_N)(y)$$

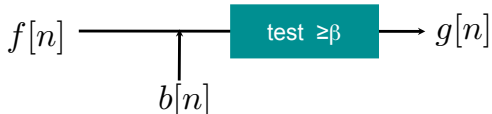
$$\begin{aligned} \text{Preuve: } P_Y(\omega) &= \mathbb{E} \{ e^{-j\omega Y} \} = \mathbb{E} \{ e^{-j\omega(X_1 + X_2 + \dots + X_N)} \} = \mathbb{E} \{ e^{-j\omega X_1} \} \cdot \mathbb{E} \{ e^{-j\omega X_2} \} \cdot \dots \cdot \mathbb{E} \{ e^{-j\omega X_N} \} \\ &= \underbrace{P_1(\omega) \cdot P_2(\omega) \cdot \dots \cdot P_N(\omega)}_{\mathcal{F}\{p_1 * p_2 * \dots * p_N\}(\omega)} \end{aligned}$$

On peut également calculer tous les **moments** de  $X$  en dérivant cette fonction caractéristique

$$\mu_{X,n} = \mathbb{E} \{ X^n \} = \int_{\mathbb{R}} x^n p_X(x) dx = j^n \frac{d^n P_X(\omega)}{d\omega^n} \Big|_{\omega=0}$$

## Exemple d'application

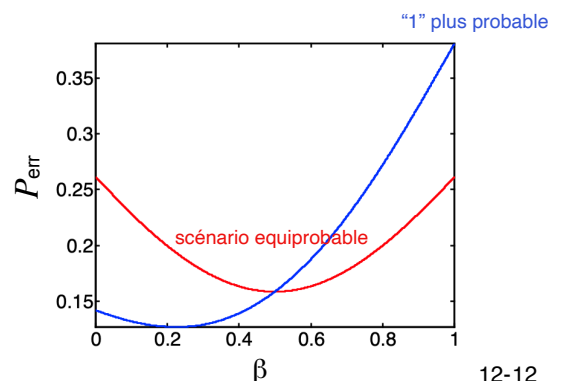
Soit un signal numérique  $f[n]$  composé de zéros avec une probabilité  $p_0$  et de 1 avec une probabilité  $(1 - p_0)$ , transmis à travers un système qui ajoute au signal un bruit  $b[n]$  de loi  $\mathcal{N}(0, \sigma^2)$ . On décide de la valeur reçue par seuillage

$$g[n] = \begin{cases} 1, & \text{si } f[n] + b[n] \geq \beta \\ 0, & \text{sinon} \end{cases}$$


La **probabilité d'erreur** est alors donnée par

$$\begin{aligned} P_{\text{err}} &= \underbrace{\mathcal{P}\{g[n] = 1 | f[n] = 0\}}_{\frac{1}{\sigma\sqrt{2\pi}} \int_{\beta}^{+\infty} e^{-\frac{(x-0)^2}{2\sigma^2}} dx} \cdot \underbrace{\mathcal{P}\{f[n] = 0\}}_{p_0} + \underbrace{\mathcal{P}\{g[n] = 0 | f[n] = 1\}}_{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\beta} e^{-\frac{(x-1)^2}{2\sigma^2}} dx} \cdot \underbrace{\mathcal{P}\{f[n] = 1\}}_{1-p_0} \\ &= p_0 \left( \frac{1}{2} - \frac{1}{2} \text{erf} \left( \frac{\beta}{\sigma\sqrt{2}} \right) \right) + (1 - p_0) \left( \frac{1}{2} - \frac{1}{2} \text{erf} \left( \frac{1-\beta}{\sigma\sqrt{2}} \right) \right) \end{aligned}$$

où  $\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$ .



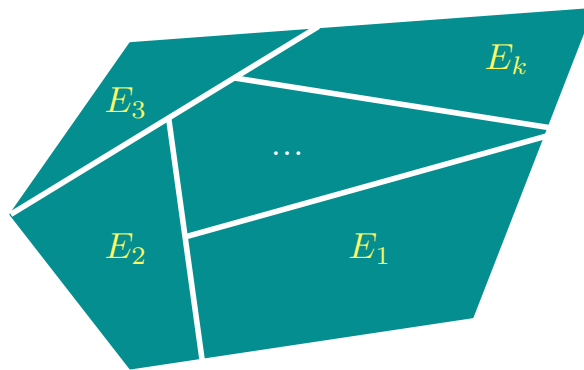
# Détecteur optimal

Soit une variable aléatoire  $X$  (éventuellement vectorielle) prenant  $N$  valeurs distinctes  $x_k$  avec la probabilité  $p_k = \mathcal{P}\{X = x_k\}$ . Elle est perturbée par un bruit  $B$  de densité de probabilité  $q(b) = p_B(b)$  et le récepteur fournit une réalisation  $y = x + b$ .

**Problème:** Quelle est la méthode qui permet d'associer le signal reçu  $y$  à l'une des valeurs  $x_k$  avec une probabilité d'erreur minimale?

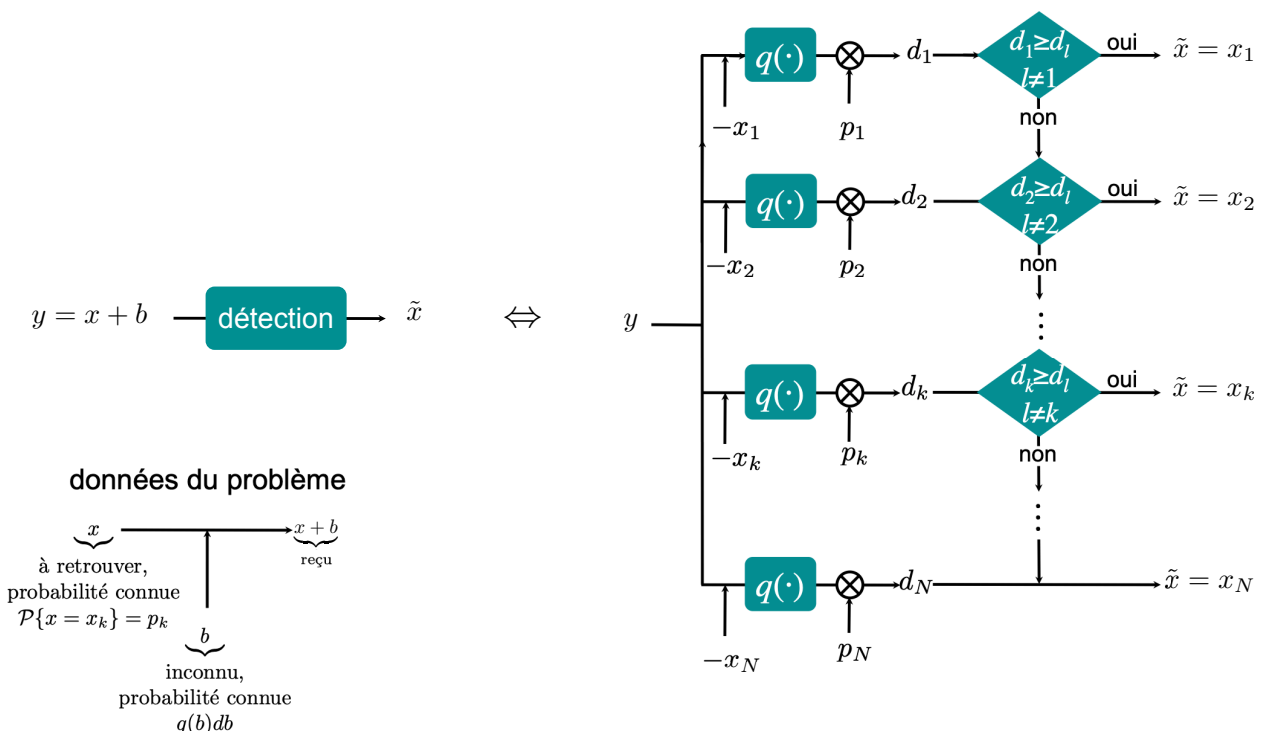
**Solution:** Le détecteur optimal est la fonction  $T : y \mapsto \sum_k x_k \cdot \mathbb{1}_{E_k}(y)$  où  $\mathbb{1}_{E_k}(y)$  est la fonction indicatrice de l'ensemble  $E_k$  qui est défini par les inégalités

$$y \in E_k \Leftrightarrow p_k \cdot q(y - x_k) - p_l \cdot q(y - x_l) \geq 0, \quad \forall l \neq k$$



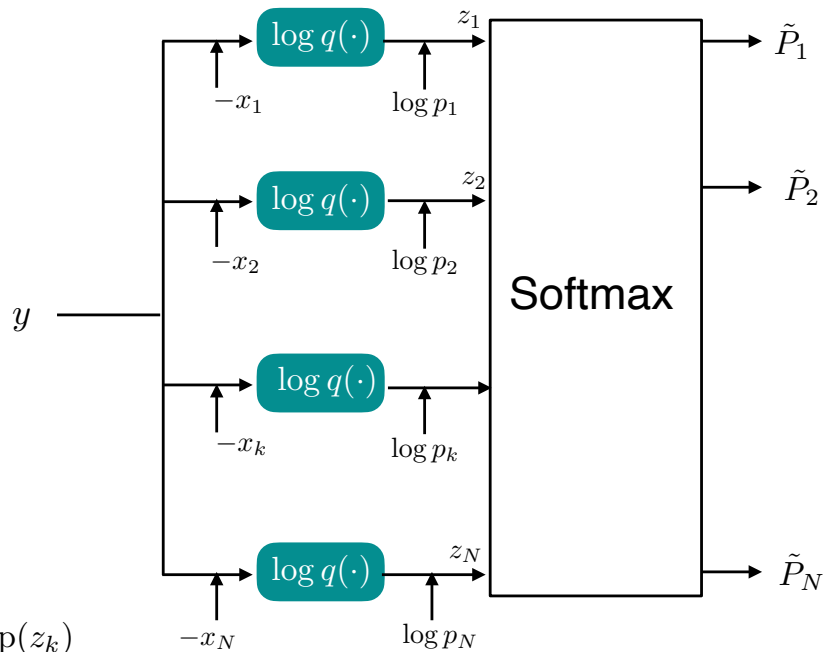
## Structure du détecteur optimal

$$\tilde{x} = x_k \Leftrightarrow y \in E_k \Leftrightarrow p_k \cdot q(y - x_k) \geq p_l \cdot q(y - x_l), \quad \forall l \neq k$$



## Détecteur optimal: version “Neural network”

$$y \in E_k \Leftrightarrow \log p_k + \log q(y - x_k) \geq \log p_l + \log q(y - x_l), \quad \forall l \neq k$$



$$\text{Softmax: } \tilde{P}_k = \frac{\exp(z_k)}{\sum_{k=1}^N \exp(z_k)}$$

$\Rightarrow$  transforme les mesures de vraisemblance en probabilités

## Détecteur optimal: preuve

Le détecteur attribue la valeur  $x_k$  chaque fois que  $y \in E_k$ . Alors, les ensembles  $E_k$  sont disjoints et sont tels que  $\bigcup_k E_k = E_{\text{tot}}$  où  $E_{\text{tot}}$  est l'espace de toutes les valeurs possibles de  $y$ .

La probabilité d'erreur s'exprime sous la forme

$$\begin{aligned} P_{\text{err}} &= \sum_k \mathcal{P}\{y \notin E_k | x = x_k\} \cdot \mathcal{P}\{x = x_k\} = 1 - \sum_k \mathcal{P}\{y \in E_k \text{ et } x = x_k\} \\ &= 1 - \mathcal{P}\{y \in E_l \text{ et } x = x_l\} - \sum_{k \neq l} \mathcal{P}\{y \in E_k \text{ et } x = x_k\} \\ &= 1 - p_l - \underbrace{\sum_{k \neq l} \left( \mathcal{P}\{y \in E_k \text{ et } x = x_k\} - \mathcal{P}\{y \in E_k \text{ et } x = x_l\} \right)}_{\int_{E_k} (p_k \cdot q(y - x_k) - p_l \cdot q(y - x_l)) dy} \end{aligned}$$

en choisissant  $l$  arbitrairement dans  $[1 \dots N]$ .

Les ensembles disjoints  $E_k$  qui minimisent  $P_{\text{err}}$  vérifient nécessairement

$$p_k \cdot q(y - x_k) - p_l \cdot q(y - x_l) \geq 0 \quad \text{pour tout } y \in E_k \text{ et } l \neq k$$

sinon il suffirait de transférer de  $E_k$  à  $E_l$  tous ses éléments  $y$  tels que  $p_k \cdot q(y - x_k) - p_l \cdot q(y - x_l) < 0$ .

Comme l'argument vaut pour tout  $l$ , les ensembles optimaux sous-jacents sont caractérisés par

$$y \in E_k \Leftrightarrow p_k \cdot q(y - x_k) - p_l \cdot q(y - x_l) \geq 0, \quad \forall l \neq k$$

et ils forment une partition de  $E_{\text{tot}}$ .

## ■ Exemple: le cas Gaussien

Si l'on suppose que le bruit suit une loi Gaussienne  $q(b) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{b^2}{2\sigma^2}}$   
alors le domaine de décision  $E_k$  est donné par

$$p_k \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-x_k)^2}{2\sigma^2}} \geq p_l \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-x_l)^2}{2\sigma^2}} \quad \forall l \neq k$$

$$\text{c-à-d: } (y - x_k)^2 \leq (y - x_l)^2 - 2\sigma^2 \frac{\log p_l}{\log p_k} \quad \forall l \neq k$$

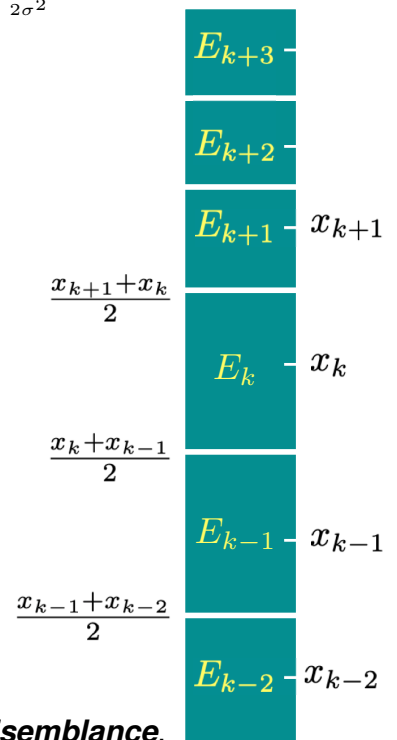
$\Updownarrow$

$$\max_{\substack{x_l \leq x_k \\ l \neq k}} \left( \frac{x_k + x_l}{2} + \frac{\sigma^2}{x_k - x_l} \log \frac{p_l}{p_k} \right) \leq y \leq \min_{\substack{x_l \geq x_k \\ l \neq k}} \left( \frac{x_k + x_l}{2} - \frac{\sigma^2}{x_l - x_k} \log \frac{p_l}{p_k} \right)$$

Le détecteur optimal est donc un simple seuil.

**Remarque:** si les symboles sont équiprobables ( $p_k = p_l$ )

le détecteur optimal correspond au choix naturel du **maximum de vraisemblance**.



## 12.2 INFORMATION

- Notion d'information
- Entropie d'une source
- Répartitions à entropie maximale
- Information commune

# Notion d'information

Dans les systèmes de communication, il est essentiel de quantifier l'information. Intuitivement, ce qui est quantifiable dans l'information est relié à la probabilité d'occurrence:

- un message dont la probabilité est grande (demain, il fera jour) véhicule peu d'information
- un message dont la probabilité est faible (la Suisse sera championne du monde de foot) en véhicule beaucoup plus

Donc l'information  $I(p)$  = fonction décroissante de  $p$ .

D'autre part, l'information véhiculée par un message constitué de deux messages **indépendants** (de probabilité  $p$  et  $q$ ) est bien sûr la somme des deux informations d'où

$$I(p \cdot q) = I(p) + I(q)$$

On peut montrer que la seule fonction **continue** vérifiant cette propriété est nécessairement proportionnelle à  $\log(p)$  d'où, si  $a > 1$ ,



$$I(p) = -\log_a(p).$$

## ■ Point de vue de l'ingénieur

L'information d'un message numérique ( $\in \mathbb{N}$ ) est le nombre minimal  $b$  de **bits** nécessaires pour le stocker. Par exemple, pour transmettre d'un message qui peut prendre  $n$  valeurs équiprobables avec  $2^{b-1} < n \leq 2^b$ ; il faut  $b = \lceil \log n \rceil$  bits. Comme la probabilité d'un tel message est  $1/n$ , on a

$$b = \lceil -\log_2 p \rceil \approx -\log_2 p$$

dans le cas d'un système de grande capacité.

## ■ Point de vue du physicien

La notion d'**information** est intimement liée à l'**entropie** d'un **système physique**, une quantité qui décrit le désordre d'un système ayant un grand nombre de degrés de liberté. Plus grand est le nombre d'états possibles, plus grande est l'**entropie**—le désordre—du système. En thermodynamique, l'entropie est précisément donnée par

$$S = k \log \Omega$$

où  $k$  est la constante de Boltzmann et  $\Omega$  le nombre d'états libres du système.



# Entropie d'une source

Soit une source de messages décrite par une variable aléatoire  $X$  (éventuellement vectorielle) prenant des valeurs discrètes  $x_k$  avec la probabilité  $p_k$ . Alors l'**entropie de la source** est définie comme l'**information moyenne** par message



$$H_X = - \sum_k p_k \log_2 p_k$$

Dans le cas d'une source de messages prenant des valeurs de nature continue, caractérisées par une densité de probabilité  $p_X(x)$ , la notion limite d'entropie perd son sens car

$$- \sum_k \underbrace{p_X(k\Delta x) \cdot \Delta x}_{p_k} \log_2 (p_X(k\Delta x) \cdot \Delta x) \xrightarrow{\Delta x \rightarrow 0} +\infty$$

La notion appropriée est alors l' **entropie différentielle**

$$\begin{aligned} H_X &= - \int_{\mathbb{R}} p_X(x) \log_2 (p_X(x)) dx \\ &= \mathbb{E} \{ -\log_2 (p_X(X)) \} \end{aligned}$$

## Répartitions à entropie maximale

On s'intéresse aux variables aléatoires (éventuellement vectorielles)  $X$  quantifiées qui stockent le plus d'information sous différentes contraintes:

- nombre fini de valeurs possibles:  $x \in \{x_1, x_2, \dots, x_N\}$
- énergie moyenne finie:  $\mathbb{E} \{ \|X\|^2 \} < \infty$
- positivité et moyenne finie:  $X \geq 0$  et  $\mathbb{E} \{ X \} < \infty$

### ■ Nombre fini de valeurs possibles

On cherche  $p_n$  maximisant  $\sum_{n=1}^N -p_n \log p_n$  avec la contrainte  $\sum_{n=1}^N p_n = 1$ .

Méthode des multiplicateurs de **Lagrange**: 
$$\max_{p_n} \underbrace{\sum_{n=1}^N -p_n \log p_n + \lambda \cdot \left( 1 - \sum_{n=1}^N p_n \right)}_{J(p_1, p_2, \dots, p_N)}$$

$$\frac{\partial J}{\partial p_n} = -\log p_n - 1 - \lambda = 0 \Rightarrow p_n = \text{constante} = 1/N$$

→ la solution est la loi de probabilité **uniforme**.

## ■ Énergie moyenne finie

On cherche  $p_n$  maximisant  $\sum_n -p_n \log p_n$  avec  $\sum_n \|x_n\|^2 p_n = E$  et  $\sum_n p_n = 1$

**Lagrange**  $\rightarrow \max_{p_n} \underbrace{\sum_n -p_n \log p_n + \lambda \cdot \left(1 - \sum_n p_n\right) + \mu \cdot \left(E - \sum_n \|x_n\|^2 p_n\right)}_{J(\dots, p_n, p_{n+1}, \dots)}$

$$\frac{\partial J}{\partial p_n} = -\log p_n - 1 - \lambda - \mu \|x_n\|^2 = 0 \Rightarrow p_n = C \cdot e^{-\mu \|x_n\|^2}$$

$\rightarrow$  la solution est la loi de probabilité **Gaussienne** discrète.

## ■ Positivité et moyenne finie

On cherche  $p_n$  maximisant  $\sum_n -p_n \log p_n$  avec  $\sum_n p_n = 1$  et  $\sum_n x_n p_n = M$  où  $x_n \geq 0$

**Lagrange**  $\rightarrow \max_{p_n} \underbrace{\sum_n -p_n \log p_n + \lambda \cdot \left(1 - \sum_n p_n\right) + \mu \cdot \left(M - \sum_n x_n p_n\right)}_{J(\dots, p_n, p_{n+1}, \dots)}$

$$\frac{\partial J}{\partial p_n} = -\log p_n - 1 - \lambda - \mu x_n = 0 \Rightarrow p_n = C \cdot e^{-\mu x_n} = p(x_n) \text{ avec } p(x) = C \cdot u(x) e^{-\mu x}$$

$\rightarrow$  la solution est la loi de probabilité **exponentielle** discrète.

# Information commune (ou mutuelle)

Soient deux sources aléatoires  $X$  et  $Y$ . On peut calculer l'entropie  $H_{X,Y}$  du couple  $(X, Y)$ , ou bien calculer l'entropie  $H_X$  de  $X$  et  $H_Y$  de  $Y$  **séparément**.

Si  $X$  et  $Y$  sont indépendantes alors on a  $H_{X,Y} = H_X + H_Y$ . De manière générale, on a toujours (égalité si et seulement si  $X$  et  $Y$  sont indépendantes)



$$H_{X,Y} = \mathbb{E}\{-\log_2 p_{X,Y}(X, Y)\} \leq H_X + H_Y$$

La quantité  $I_{X,Y} = H_X + H_Y - H_{X,Y} = H_X - H_{X|Y} = H_Y - H_{Y|X}$

$$= \mathbb{E} \left\{ \log_2 \left( \frac{p_{X,Y}(X, Y)}{p_X(X) \cdot p_Y(Y)} \right) \right\}$$

est alors une mesure de l'**information commune** entre  $X$  et  $Y$ . Elle est toujours positive ou nulle et vérifie

$$0 \leq I_{X,Y} \leq \max(H_X, H_Y)$$

avec égalité si et seulement si il existe une fonction  $f$  inversible telle que  $Y = f(X)$ .

L'information commune peut donc être vue comme une **corrélation généralisée**.

## 12.3 PROCESSUS ALÉATOIRES

---

- Statistique d'ordre 2
- Signaux aléatoires
- Stationnarité, ergodicité
- Temps continu/temps discret
- Densité spectrale de puissance
- Théorème de Wiener-Khintchine
- DSP d'un signal filtré
- Bruit blanc, mouvements Browniens

### Statistique d'ordre 2

---

En traitement des signaux, on se limite le plus souvent aux statistiques d'ordre 1 et 2 de vecteurs (ou signaux) aléatoires complexes dénotés ici par  $\mathbf{X}$

- moyenne simple:  $\boldsymbol{\mu}_{\mathbf{X}} = \mathbb{E} \{ \mathbf{X} \}$
- matrice de covariance:  $\mathbf{C}_{\mathbf{X}} = \mathbb{E} \{ (\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T \} = \mathbb{E} \{ \mathbf{X} \mathbf{X}^T \} - \boldsymbol{\mu}_{\mathbf{X}} \boldsymbol{\mu}_{\mathbf{X}}^T$

Il y a plusieurs raisons pour cela

1. On calcule essentiellement des **énergies** (ou des **puissances**) dans les systèmes physiques (lois de conservation).
2. Les calculs d'**optimisation** de formes quadratiques donnent des **systèmes linéaires** "faciles" à analyser et à résoudre.
3. La dualité des énergies dans la transformation de Fourier (**Parseval**) rendent ces statistiques particulièrement adaptées aux **systèmes convolutifs** (LIT).
4. Les bruits de nature physique se modélisent par des processus **Gaussiens** (justification: théorème central-limite) qui sont complètement décrits par leur statistique d'ordre 2.
5. L'**estimation** de paramètres linéaires par minimisation aux moindres carrés est **optimale** pour les processus Gaussiens.

# Signaux aléatoires

Un **signal aléatoire à temps discret**  $X[\cdot]$  peut se voir comme un vecteur de variables aléatoires  $(\dots, X[n-1], X[n], X[n+1], \dots)$  de dimension infinie.

Un **signal aléatoire à temps continu**  $X(\cdot)$  est une limite quand  $T \rightarrow 0$  du signal à temps discret constitué des échantillons  $(\dots, X((n-1)T), X(nT), X((n+1)T), \dots)$ .

On parle alors de **processus stochastique** (ou aléatoire).

L'idée philosophique est que, même si un signal est déterministe, l'absence d'information sur sa génération incite à le modéliser comme la réalisation d'un tirage aléatoire...

Sa statistique est donc caractérisée par une mesure de probabilité (que l'on n'exprime jamais) et donc par un opérateur d'**espérance**. Pour avoir une théorie prédictive, on peut faire diverses hypothèses probabilistes

- la **stationnarité** au sens strict
- l'**ergodicité**
- la stationnarité au sens large

## Stationnarité


Un signal aléatoire  $X(\cdot)$  est **stationnaire au sens strict** si et seulement si, quelle que soit l'espérance calculée, elle est indépendante du point de référence temporel.

Ainsi, si  $X(\cdot)$  est **stationnaire au sens strict**, alors pour tout  $n \in \mathbb{N}$  on a (signaux réels)

$$\begin{aligned}\mathbb{E}\{f(X(t_1), X(t_2), \dots, X(t_n))\} &= \mathbb{E}\{f(X(t_1 - a), X(t_2 - a), \dots, X(t_n - a))\} \quad \forall a \in \mathbb{R} \\ &= \mathbb{E}\{f(X(0), X(t_2 - t_1), \dots, X(t_n - t_1))\} \quad (a = t_1) \\ &= \text{fonctions de } (t_2 - t_1, t_3 - t_1, \dots, t_n - t_1)\end{aligned}$$

Définition semblable pour les signaux aléatoires complexes, ou discrets, ou les deux.

Un signal aléatoire  $X(\cdot)$  est **stationnaire au sens large** (SSL ou WSS) si et seulement si les moments d'ordre 1 et 2 sont indépendants du point de référence temporel. Donc, si et seulement si


$$\begin{aligned}\underbrace{\mathbb{E}\{X(t)\}}_{\text{moyenne statistique de } X(t)} &= \mathbb{E}\{X(0)\} = \text{constante} \quad \forall t \in \mathbb{R} \\ \underbrace{\mathbb{E}\{X(t)X^*(\tau)\}}_{\text{autocorrélation statistique de } X(t)} &= \mathbb{E}\{X(0)X^*(\tau - t)\} = \rho_X(\tau - t) \quad \forall t, \tau \in \mathbb{R}\end{aligned}$$

Caractérisation statistique par  $\rho_X : \mathbb{R} \rightarrow \mathbb{C}$

# Ergodicité

Un signal aléatoire  $X(\cdot)$  est **ergodique** ssi toutes ses moyennes statistiques peuvent être obtenues à l'aide de ses moyennes temporelles équivalentes d'une quelconque de ses réalisations  $x(\cdot)$ .

## Exemples

d'où l'utilité d'une notation distincte

$$\mathbb{E} \{ \underbrace{X(t)}_{\text{variable aléatoire}} \} = \lim_{A \rightarrow \infty} \frac{1}{A} \int_{-A/2}^{A/2} \underbrace{x(t+\tau)}_{\text{une réalisation du processus aléatoire}} d\tau$$

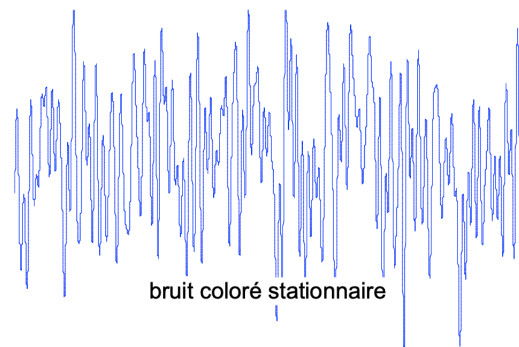
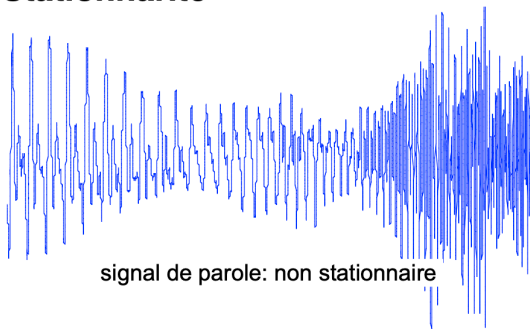
$$\mathbb{E} \{ X(t)X^*(t') \} = \lim_{A \rightarrow \infty} \frac{1}{A} \int_{-A/2}^{A/2} x(t+\tau)x^*(t'+\tau) d\tau$$



**Remarque:** un signal ergodique est automatiquement stationnaire au sens strict (l'inverse est faux). L'ergodicité est une propriété que l'on supposera toujours en pratique pour les signaux stationnaires.

## Exemples

### Stationnarité

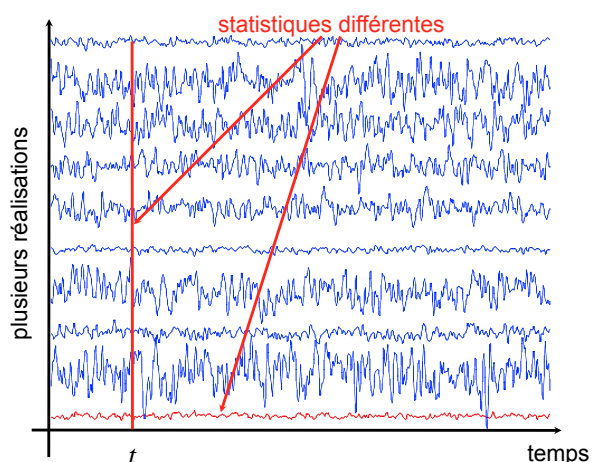


### Ergodicité

Signal stationnaire non ergodique

$$X(t) = AX_0(t)$$

où  $X_0(t)$  est un processus ergodique et  $A$  une variable aléatoire indépendante de  $X_0(t)$ .



# Temps continu/temps discret

Les **échantillons**  $X[n] = X(nT)$  d'un signal stationnaire  $X(\cdot)$  à temps continu forment un signal stationnaire à temps discret.

**Preuve:** 
$$\begin{aligned}\mathbb{E}\{f(\dots, X[n], X[n+1], \dots)\} &= \mathbb{E}\{f(\dots, X(nT), X((n+1)T), \dots)\} \\ &= \mathbb{E}\{f(\dots, X(nT-a), X((n+1)T-a), \dots)\} \quad \forall a \in \mathbb{R} \\ &= \mathbb{E}\{f(\dots, X(nT-n_0T), X((n+1)T-n_0T), \dots)\} \quad \forall n_0 \in \mathbb{Z} \\ &= \mathbb{E}\{f(\dots, X[n-n_0], X[n+1-n_0], \dots)\} \quad \forall n_0 \in \mathbb{Z}\end{aligned}$$

Au contraire, l'**interpolation**  $X_{\text{int}}(t) = \sum_n X[n]\varphi(t/T - n)$  d'un signal discret stationnaire  $X[n]$  n'est en général pas stationnaire, sauf si  $\varphi(t) = \text{sinc}(t)$  (SSL).

**Preuve:** 
$$\begin{aligned}\mathbb{E}\{X_{\text{int}}(t)X_{\text{int}}^*(t')\} &= \sum_{n,n'} \underbrace{\mathbb{E}\{X[n]X^*[n']\}}_{\rho_X[n'-n]} \varphi(t-n)\varphi(t'-n') \quad T=1 \text{ sans perte de généralité} \\ &= \sum_{m \in \mathbb{Z}} \rho_X[m] \sum_{n \in \mathbb{Z}} \varphi(t'-m-n)\varphi(t-n)\end{aligned}$$
  
Or, on sait que si  $x(t)$  and  $y(t)$  sont à bande limitée dans  $[-\pi, \pi]$ , alors  $\sum_{n \in \mathbb{Z}} x(n)y^*(n) = \int_{\mathbb{R}} x(\tau)y^*(\tau)d\tau$ .  
En prenant  $x(\tau) = \varphi(t'-m-\tau)$ ,  $y(\tau) = \varphi(t-\tau)$  et  $\varphi = \text{sinc}$ , on obtient alors  

$$\sum_{n \in \mathbb{Z}} \text{sinc}(t'-m-n)\text{sinc}(t-n) = \int_{\mathbb{R}} \text{sinc}(t'-m-\tau)\text{sinc}(t-\tau)d\tau = \underbrace{(\text{sinc} * \text{sinc})}_{=\text{sinc}}((t'-t)-m)$$
  
d'où finalement  $\mathbb{E}\{X_{\text{int}}(t)X_{\text{int}}^*(t')\} = \sum_{m \in \mathbb{Z}} \rho_X[m]\text{sinc}((t'-t)-m) = \text{fonction de } (t'-t).$

# Densité spectrale de puissance

Une réalisation  $x(\cdot)$  d'un signal aléatoire stationnaire  $X(\cdot)$  n'a pas de transformée de Fourier exploitable car, a priori,  $x(\cdot) \notin L_1(\mathbb{R})$ . Mais on peut rendre le signal à support borné en le multipliant par une fenêtre rectangulaire de taille  $A$

$$X_A(t) = \text{rect}(t/A) \cdot X(t)$$

On peut dès lors calculer la transformée de Fourier  $\hat{x}_A(\omega) = \mathcal{F}\{x_A\}(\omega)$  de toute réalisation  $x_A(\cdot)$  ce qui donne un sens à  $\hat{X}_A(\omega) = \mathcal{F}\{X_A\}$ . La densité spectrale de puissance (DSP)  $S_X(\omega)$  du processus  $X(t)$  est alors donnée par la limite (si elle existe)

$$S_X(\omega) = \lim_{A \rightarrow \infty} \frac{1}{A} \mathbb{E}\{|\hat{X}_A(\omega)|^2\} \geq 0$$

**Remarque:** définition équivalente pour les signaux à temps discret en remplaçant la transformée de Fourier par la DTFT.

**Interprétation:**  $P = \lim_{A \rightarrow \infty} \frac{1}{A} \int_{-A/2}^{+A/2} |x(t)|^2 dt$  est l'énergie moyenne du signal par unité de temps, c-à-d sa puissance moyenne. En utilisant Parseval

$$\mathbb{E}\{P\} = \lim_{A \rightarrow \infty} \frac{1}{2\pi A} \int_{\mathbb{R}} \mathbb{E}\{|\hat{X}_A(\omega)|^2\} d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} S_X(\omega) d\omega$$

De même,  $\int_{\omega_1}^{\omega_2} S_X(\omega) \frac{d\omega}{2\pi}$  est la contribution des fréquences  $[\omega_1, \omega_2]$  à la puissance moyenne du signal  $\rightarrow$  **densité spectrale de puissance**.

# Densité spectrale: généralisations

Une extension naturelle de la densité spectrale de puissance est la densité spectrale de puissance croisée entre deux signaux aléatoires

$$S_{X,Y}(\omega) = \lim_{A \rightarrow \infty} \frac{1}{A} \mathbb{E} \{ \hat{X}_A(\omega) \cdot \hat{Y}_A^*(\omega) \}$$

On peut aussi définir une version généralisée de la DSP dans le cas de signaux **vectoriels**  $\mathbf{X}(\cdot) = (X_1(\cdot), \dots, X_N(\cdot))$  (application aux **multicapteurs**)

$$S_{\mathbf{X}}(\omega) = \lim_{A \rightarrow \infty} \frac{1}{A} \mathbb{E} \left\{ \hat{\mathbf{X}}_A(\omega) (\hat{\mathbf{X}}_A(\omega)^T)^* \right\}$$

Dans ce cas bien sûr, la DSP devient une matrice. Si  $\mathbf{X}(t) = (X_1(t), X_2(t))$ , on a

$$\mathbf{S}_{\mathbf{X}}(\omega) = \begin{pmatrix} S_{X_1}(\omega) & S_{X_1, X_2}(\omega) \\ S_{X_2, X_1}(\omega) & S_{X_2}(\omega) \end{pmatrix}$$

## Théorème de Wiener-Khintchine

Autocorrelation statistique d'un signal SSL :  $\rho_X(t) = \mathbb{E}\{X(\tau)X^*(t+\tau)\}$ ,  $\forall \tau \in \mathbb{R}$

**Théorème** (Wiener-Khintchine, 1934)

La DSP d'un **signal stationnaire au sens large**  $X(t)$  (réel) est la **transformée de Fourier** de la **fonction d'autocorrélation** statistique de ce signal

$$S_X(\omega) = \int_{\mathbb{R}} \rho_X(t) e^{-j\omega t} dt = \mathcal{F}\{\rho_X\}(\omega).$$

**Preuve:** 
$$\begin{aligned} \frac{1}{A} \mathbb{E} \{ |\hat{X}_A(\omega)|^2 \} &= \frac{1}{A} \mathbb{E} \left\{ \int_{\mathbb{R}} X(t) \text{rect}(t/A) e^{-j\omega t} dt \int_{\mathbb{R}} X^*(t') \text{rect}(t'/A) e^{j\omega t'} dt' \right\} \\ &= \frac{1}{A} \int_{\mathbb{R}} \int_{\mathbb{R}} \text{rect}(t/A) \text{rect}(t'/A) \cdot \underbrace{\mathbb{E} \{ X(t) X^*(t') \}}_{\rho_X(t'-t)} e^{-j\omega(t-t')} dt dt' \\ &= \int_{\mathbb{R}} \rho_X(-u) e^{-j\omega u} \underbrace{\int_{\mathbb{R}} \text{rect}(u/A + u') \text{rect}(u') du'}_{=(\text{rect} * \text{rect})(u/A)} du \quad (\text{après ch. de variables } u = t - t', u' = t'/A) \\ &= \int_{\mathbb{R}} \text{tri}(u/A) \rho_X(-u) e^{-j\omega u} du \quad \text{avec } \rho_X(-u) = \rho_X(u) \quad (\text{signal réel}) \end{aligned}$$

$$\text{Or, } \lim_{A \rightarrow \infty} \text{tri}(u/A) = 1 \text{ d'où } \lim_{A \rightarrow \infty} \frac{1}{A} \mathbb{E} \{ |\hat{X}_A(\omega)|^2 \} = \int_{\mathbb{R}} \rho_X(u) e^{-j\omega u} du$$

De même, pour un signal à temps discret  $X[\cdot]$  avec  $\hat{X}_A(e^{j\omega}) = \sum_{n=-A/2}^{+A/2} X[n] e^{-j\omega n}$ , on a

$$\lim_{A \rightarrow \infty} \frac{1}{A} \mathbb{E} \{ |\hat{X}_A(e^{j\omega})|^2 \} = S_X(e^{j\omega}) = \sum_{n \in \mathbb{Z}} \rho_X[n] e^{-j\omega n} = \mathcal{F}_d\{\rho_X\}(\omega).$$

# Wiener-Khintchine: généralisations

Une généralisation de ce théorème est facile pour les densités spectrales de puissance croisées pour des signaux  $X(\cdot), Y(\cdot)$  SSL

$$S_{X,Y}(\omega) = \int_{\mathbb{R}} \rho_{X,Y}(t) e^{-j\omega t} dt$$

où  $\rho_{X,Y}(t) = \mathbb{E}\{X(\tau)Y^*(t+\tau)\}, \forall \tau \in \mathbb{R}$  (fonction d'intercorrélation).

En fait, on peut généraliser encore plus et considérer des signaux vectoriels  $\mathbf{X}(\cdot) = (X_1(\cdot), \dots, X_N(\cdot))$  SSL. Le théorème de Wiener-Khintchine s'écrit alors encore

$$\mathbf{S}_X(\omega) = \int_{\mathbb{R}} \mathbf{R}_X(t) e^{-j\omega t} dt = \int_{\mathbb{R}} \mathbb{E}\{\mathbf{X}(0)(\mathbf{X}(t)^T)^*\} e^{-j\omega t} dt$$

## ■ Fonction matricielle d'intercorrélation

$$\mathbf{R}_X : \mathbb{R} \rightarrow \mathbb{C}^{N \times N} \quad \text{avec} \quad [\mathbf{R}_X(t)]_{m,n} = \mathbb{E}\{X_n(0)X_m^*(t)\} = \rho_{X_n, X_m}(t).$$

# Combinaisons linéaires et covariances

## ■ En dimension fini (par linéarité)

$$\mathbb{E}\left\{\left(\sum_{m=1}^M a_m X_m\right)\left(\sum_{n=1}^N b_n Y_n\right)\right\} = \mathbb{E}\left\{\sum_{m=1}^M \sum_{n=1}^N a_m b_n X_m Y_n\right\} = \sum_{m=1}^M \sum_{n=1}^N a_m b_n \underbrace{\mathbb{E}\{X_m Y_n\}}_{\text{Cov}(X_m, Y_n)}$$

## ■ Filtrage et intercorrélation

- Fonction d'intercorrélation de deux processus stochastiques SSL  $X(\cdot)$  et  $Y(\cdot)$ :

$$\rho_{X,Y}(\tau) \triangleq \mathbb{E}\{X(0)Y^*(\tau)\} = \mathbb{E}\{X(t)Y^*(t+\tau)\}, \quad \forall t \in \mathbb{R}$$

- Signaux filtrés:  $U(t) = (h * X)(t), V(t) = (g * Y)(t)$

- Fonction d'intercorrélation après filtrage

$$\rho_{U,V}(\tau) = \mathbb{E}\{U(0)V^*(\tau)\} = (h^\vee * g^* * \rho_{X,Y})(\tau) \quad \text{où} \quad h^\vee(t) = h(-t)$$

Preuve:  $\rho_{U,V}(\tau) = \mathbb{E}\left\{\int_{\mathbb{R}} h(t)X(0-t)dt \int_{\mathbb{R}} g^*(t')Y^*(\tau-t')dt'\right\}$  Hypothèses:  $h, g, \rho_{X,Y} \in L_1(\mathbb{R})$

$= \int_{\mathbb{R}} \int_{\mathbb{R}} h(t)g^*(t') \underbrace{\mathbb{E}\{X(-t)Y^*(\tau-t')\}}_{\rho_{X,Y}(\tau-t'+t)} dt dt'$  changement de variable:  $u = t' - t$

$= \int_{\mathbb{R}} \left(\underbrace{\int_{\mathbb{R}} h(t'-u)g^*(t')dt'}_{c_{hg}(u)=(h^\vee * g^*)(u)}\right) \rho_{X,Y}(\tau-u)du = ((h^\vee * g^*) * \rho_{X,Y})(\tau)$

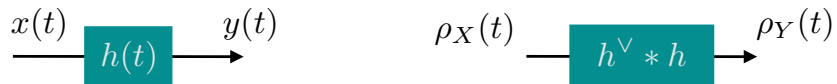


# Densité spectrale d'un signal filtré

Étant donné un signal SSL réel  $X(t)$ , le signal filtré  $Y(t) = (h * X)(t)$  est également SSL (exercice!). Les densités spectrales de puissance de  $X(t)$  et  $Y(t)$  sont alors reliées par

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega)$$

Graphiquement:



**Preuve:**  $\rho_Y(t) = \mathbb{E} \{Y(0)Y(t)\}$

$$\begin{aligned} &= \mathbb{E} \left\{ \int_{\mathbb{R}} h(\tau) X(-\tau) d\tau \int_{\mathbb{R}} h(\tau') X(t - \tau') d\tau' \right\} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(\tau) h(\tau') \mathbb{E} \{X(-\tau) X(t - \tau')\} d\tau d\tau' \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(\tau) h(\tau') \rho_X(t - \tau' + \tau) d\tau d\tau' \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} h(\tau) h(\tau') \left( \frac{1}{2\pi} \int_{\mathbb{R}} S_X(\omega) e^{j\omega(t - \tau' + \tau)} d\omega \right) d\tau d\tau' \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} h(\tau) e^{j\omega\tau} d\tau \right) \left( \int_{\mathbb{R}} h(\tau') e^{-j\omega\tau'} d\tau' \right) \frac{1}{2\pi} S_X(\omega) e^{j\omega t} d\omega \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} H^*(\omega) H(\omega) S_X(\omega) e^{j\omega t} d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} S_Y(\omega) e^{j\omega t} d\omega \end{aligned}$$

$$\rho_Y(t) = (h^\vee * h * \rho_X)(t)$$

$$\begin{aligned} \mathcal{F}\{h^\vee * h\}(\omega) &= H^\vee(\omega) H(\omega) \\ &= H^*(\omega) H(\omega) \\ &= |H(\omega)|^2 \end{aligned}$$

## Bruit blanc

### Cas du temps continu

Un **bruit blanc**  $B(\cdot)$  est l'idéalisation mathématique d'un signal stationnaire à moyenne nulle dont la DSP est constante

$$S_B(\omega) = \sigma_0^2 \Leftrightarrow \rho_B(t) = \sigma_0^2 \cdot \delta(t)$$

Un tel signal est donc d'**énergie infinie** (idéalisation) et ses échantillons sont **décor-rélés**, aussi proches soient-ils. Son utilité pratique est de pouvoir représenter/générer toutes sortes de processus stochastiques physiques par filtrage (bruit **coloré**).

Un **bruit blanc discret**  $B[\cdot]$  est un signal de moyenne nulle, stationnaire, caractérisé par l'**indépendance** de ses échantillons. En particulier, ceci implique

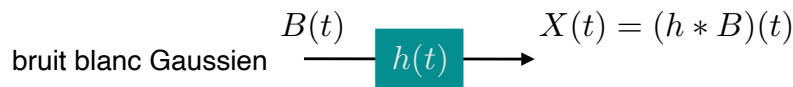
$$\rho_B[n] = \sigma_0^2 \cdot \delta[n] = \begin{cases} 0, & \text{si } n \neq 0 \\ \sigma_0^2, & \text{si } n = 0 \end{cases} \quad (\text{décorrélation})$$

Dans ce cas aussi, on a  $S_B(e^{j\omega}) = \mathcal{F}_d\{\rho_B\} = \sigma_0^2$ .

Caractériser le bruit à temps discret comme bruit blanc est souvent une hypothèse "relativement" réaliste qui simplifie énormément les calculs.

# Modélisation de processus Gaussiens stationnaires

- Génération de signal aléatoire par filtrage d'un bruit blanc



- Densités spectrales de puissance

$$S_B(\omega) = \sigma_0^2$$

$$S_X(\omega) = \sigma_0^2 \cdot |H(\omega)|^2 \quad \text{"spectral shaping"}$$

- Fonctions d'autocorrelation

$$\rho_B(t) = \mathbb{E}\{B(0)B(t)\} = \sigma_0^2 \delta(t)$$

$$\rho_X(t) = \mathbb{E}\{X(0)X(t)\}$$

$$= (h^\vee * h * \rho_B)(t) = \sigma_0^2 (h^\vee * h)(t) \quad \text{où} \quad h^\vee(t) = h(-t)$$

Généralisation pour processus non-Gaussiens et non-stationnaires:



Unser / Sig & Sys II

12-39

## Analyse spectrale généralisée d'un processus SSL

Transformée de Fourier (au sens des distributions):  $\hat{X}(\omega) = \mathcal{F}\{X\}(\omega) \in \mathcal{S}'(\mathbb{R})$

Fonction d'autocorrélation spectrale:  $(\omega, \xi) \mapsto \mathbb{E}\{\hat{X}(\omega)\hat{X}^*(\xi)\} \in \mathcal{S}'(\mathbb{R} \times \mathbb{R})$

### Théorème (Gelfand 1955)

Soit  $X(t)$  un processus stationnaire au sens large à valeur moyenne nulle. Alors

$$\mathbb{E}\{\hat{X}(\omega)\hat{X}^*(\xi)\} = 2\pi\delta(\omega - \xi) \cdot S_X(\omega).$$

La transformée de Fourier généralisée de  $X(\cdot)$  est donc **parfaitement décorrélée** mais avec une variance "infinie" proportionnelle à  $S_X(\omega) = \mathcal{F}\{\rho_X\}(\omega)$ .

**Interprétation:** La transformée de Fourier diagonalise la fonction d'autocorrélation d'un signal SSL. Ceci implique que les **composantes spectrales** d'un processus stationnaire Gaussien sont **indépendantes** (transformation de Karhunen-Loève).

Preuve formelle:

$$\begin{aligned} \mathbb{E}\{\hat{X}(\omega)\hat{X}^*(\xi)\} &= \mathbb{E}\left\{\int_{\mathbb{R}} X(t)e^{-j\omega t} dt \left(\int_{\mathbb{R}} X(\tau)e^{-j\xi\tau} d\tau\right)^*\right\} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-j\omega t} \mathbb{E}\{X(t)X^*(\tau)\} e^{j\xi\tau} dt d\tau = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{-j\omega t} \rho_X(\tau - t) e^{j\xi\tau} dt d\tau \\ &= \int_{\mathbb{R}} e^{j\xi\tau} e^{-j\omega\tau} d\tau \int_{\mathbb{R}} \rho_X(-u) e^{-j\omega u} du = \mathcal{F}\{e^{j\xi\cdot}\}(\omega) \mathcal{F}\{\rho_X^\vee\}(\omega) = 2\pi\delta(\omega - \xi) S_X(\omega) \end{aligned}$$

Changement de variable:  $u = t - \tau$

Unser / Sig & Sys II

12-40

# Mouvements Browniens

Le mouvement Brownien standard est l'**intégrale d'un bruit blanc** à temps continu. C'est un processus Gaussien tel que

$$\mathbb{E} \{ |X(t) - X(t')|^2 \} = C \cdot |t - t'|$$

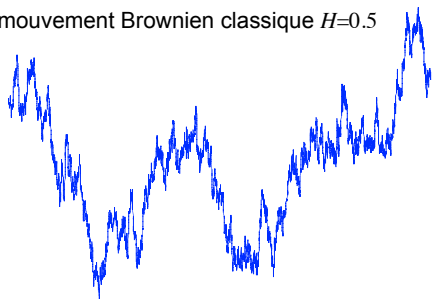
Il n'est pas stationnaire, mais ses accroissements le sont. Certains signaux naturels sont des mouvements Browniens (agitation des microparticules dans un liquide (cf. A. Einstein)).

Le mouvement Brownien **fractionnaire** est une extension du précédent: c'est toujours un processus Gaussien mais ses accroissements obéissent à une loi différente

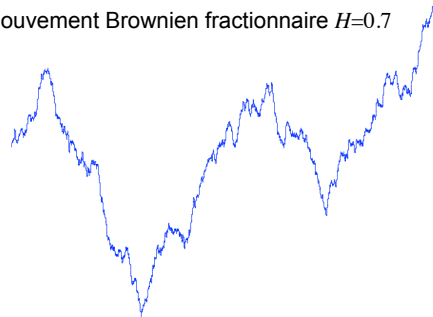
$$\mathbb{E} \{ |X(t) - X(t')|^2 \} = C \cdot |t - t'|^{2H}$$

où  $0 < H < 1$  est l'"exposant de Hurst". Nombreuses applications pour modéliser des phénomènes de croissance (e.g., fractales).

mouvement Brownien classique  $H=0.5$



mouvement Brownien fractionnaire  $H=0.7$



Unser / Sig & Sys II

12-41

## 12.4 FILTRAGE DE SIGNAUX BRUITÉS

- Estimation de signaux
- Filtrage accordé
- Minimisation de la puissance moyenne de l'erreur
- Filtre de Wiener-Hopf

# Rappel: Processus stochastiques

Le processus  $X(\cdot)$  (réel) est stationnaire au sens large (SSL) ssi:

- $\forall t \in \mathbb{R} : \mathbb{E}\{X(t)\} = \mu_X = \text{Constante}$
- $\forall t, t' \in \mathbb{R} : \mathbb{E}\{X(t)X(t')\} = \rho_X(t' - t) = \rho_X(t - t')$

■ Fonction d'autocorrélation statistique  $\rho_X : \mathbb{R} \rightarrow \mathbb{R}$

$$\rho_X(\tau) \triangleq \mathbb{E}\{X(0)X(\tau)\} = \mathbb{E}\{X(t)X(\tau + t)\}, \forall t \in \mathbb{R}$$

■ Densité spectrale de puissance

$$S_X(\omega) = \lim_{A \rightarrow \infty} \frac{1}{A} \mathbb{E}\{|\hat{X}_A(\omega)|^2\} \quad \text{avec} \quad \hat{X}_A(\omega) = \int_{-A/2}^{A/2} X(t)e^{-j\omega t} dt$$

■ Théorème de Wiener-Khintchine

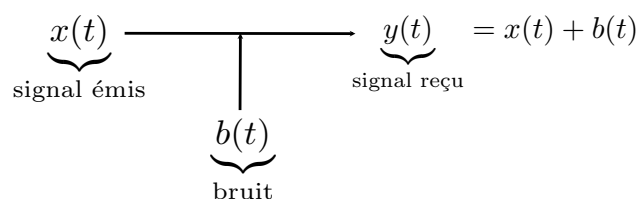
$$S_X(\omega) = \int_{\mathbb{R}} \rho_X(t)e^{-j\omega t} dt = \mathcal{F}\{\rho_X\}(\omega)$$

$$\Rightarrow \mathbb{E}\{|X(t)|^2\} = \frac{1}{2\pi} \int_{\mathbb{R}} S_X(\omega) d\omega \quad (\text{énergie moyenne du signal})$$

## Estimation de signaux

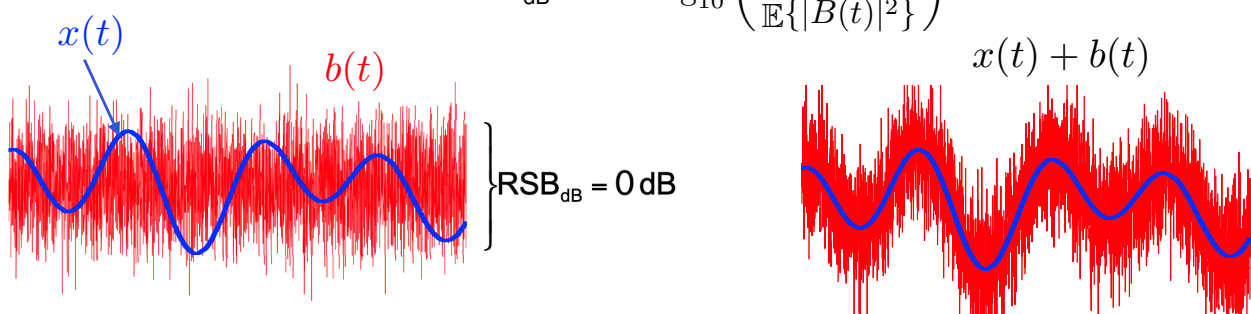
### Description du problème

Un récepteur reçoit un signal, constitué du *signal émis* à retrouver et d'un *bruit*, caractéristique du processus de transmission-réception. On supposera toujours ici que le bruit perturbe le signal de manière additive, qu'il est décorrélé du signal et de moyenne nulle.



La sévérité de la perturbation est quantifiée par son **rapport signal à bruit (RSB)**, souvent exprimé en décibels

$$\text{RSB}_{\text{dB}} = 10 \cdot \log_{10} \left( \frac{\mathbb{E}\{|X(t)|^2\}}{\mathbb{E}\{|B(t)|^2\}} \right)$$



# Filtrage accordé

Si le signal en entrée est de nature discrète, par exemple  $x(t) = \sum_{n=1}^N f[n]\varphi_n(t - t_n)$  alors on cherche un filtre qui **amplifie** le plus possible l'amplitude de l'impulsion  $\varphi_n(\cdot - t_n)$  en gardant le bruit à un niveau constant. Ceci permet de garantir une détection maximale par **seuillage** (cas de données quantifiées).

Pour  $x(t) = f[n]\varphi(t - t_n)$  (impulsion unique), le filtre optimal (à un facteur d'amplitude près) est

$$H(\omega) = \frac{\Phi^*(\omega)}{S_B(\omega)} \quad \text{avec} \quad \Phi = \mathcal{F}\{\varphi\}$$

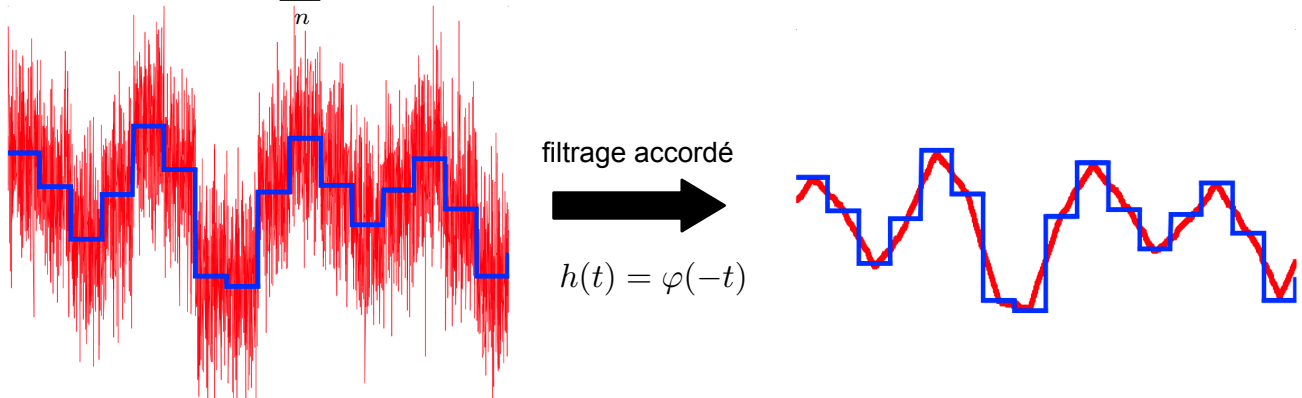
et si le bruit est blanc, alors (**filtre accordé**)  $\Rightarrow h(t) = \varphi(-t)$  = **corrélateur**

**Preuve:** puissance du bruit inchangée  $\Leftrightarrow \frac{1}{2\pi} \int_{\mathbb{R}} S_B(\omega) d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} |H(\omega)|^2 S_B(\omega) d\omega$

$$\begin{aligned} (h * \varphi(\cdot - t_n))(t_n) &= (h * \varphi)(0) = \frac{1}{2\pi} \int_{\mathbb{R}} H(\omega) \Phi(\omega) d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} H(\omega) \sqrt{S_B(\omega)} \cdot \left( \frac{\Phi^*(\omega)}{\sqrt{S_B(\omega)}} \right)^* d\omega \\ &\leq \frac{1}{2\pi} \underbrace{\sqrt{\int_{\mathbb{R}} |H(\omega)|^2 S_B(\omega) d\omega}}_{\sqrt{\int_{\mathbb{R}} S_B(\omega) d\omega}} \cdot \sqrt{\int_{\mathbb{R}} \frac{|\Phi(\omega)|^2}{S_B(\omega)} d\omega} \quad (\text{Cauchy-Schwarz}) \end{aligned}$$

avec égalité (**réponse max**) si et seulement si  $H(\omega) \sqrt{S_B(\omega)} = \lambda \frac{\Phi^*(\omega)}{\sqrt{S_B(\omega)}}$

signal bruité  $y(t) = \sum_n f[n]\varphi(t - n) + b(t)$

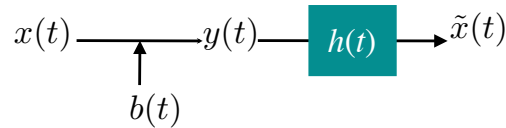


Le filtrage accordé assure qu'**aux instants d'échantillonnage**, le rapport signal à bruit sera maximisé, facilitant ainsi la **détection**. Mais il n'assure pas que les valeurs échantillonnées **reçues** égalent les valeurs échantillonnées émises: pour cela, il faut que  $(\varphi * \varphi^\vee)(t)|_{t=n} = \langle \varphi, \varphi(\cdot - n) \rangle = \delta[n]$ .

$$\Leftrightarrow \{\varphi(\cdot - n)\}_{n \in \mathbb{Z}} \text{ est une base orthonormale.}$$

## Minimisation de la puissance moyenne de l'erreur

On souhaite trouver un **filtre** qui, appliqué à une réalisation de  $Y(\cdot) = X(\cdot) + B(\cdot)$ , élimine le plus de bruit possible et conserve le mieux possible le signal.



On veut plus précisément minimiser la puissance moyenne de l'erreur  $E(t) = \tilde{X}(t) - X(t)$ , c-à-d

$$\mathbb{E} \{ |E(t)|^2 \} = \frac{1}{2\pi} \int_{\mathbb{R}} (|H(\omega) - 1|^2 S_X(\omega) + |H(\omega)|^2 S_B(\omega)) d\omega$$

**Dérivation:**  $\mathbb{E} \{ |\tilde{X}(t) - X(t)|^2 \} = \mathbb{E} \{ |(h * X)(t) - X(t) + (h * B)(t)|^2 \}$   
 $= \underbrace{\mathbb{E} \{ |(h * X)(t) - X(t)|^2 \}}_{\mathbb{E} \{ |(h(t) - \delta(t)) * X(t)|^2 \}} + \mathbb{E} \{ |(h * B)(t)|^2 \}$  (indépendance entre  $X(t)$  et  $B(t)$ )  
 $= \frac{1}{2\pi} \int_{\mathbb{R}} (|H(\omega) - 1|^2 S_X(\omega) + |H(\omega)|^2 S_B(\omega)) d\omega$  (Wiener-Khintchine)

## Filtre de Wiener-Hopf

On suppose maintenant que l'on connaît à la fois  $S_X(\omega)$  et  $S_B(\omega)$ , qui sont la DSP du signal et celle du bruit. Il s'agit donc de minimiser par rapport à  $H(\omega)$  l'intégrale

$$E = \frac{1}{2\pi} \int_{\mathbb{R}} (|H(\omega) - 1|^2 S_X(\omega) + |H(\omega)|^2 S_B(\omega)) d\omega$$

Par cela, on remarque que

$$\begin{aligned} 2\pi E &= \int_{\mathbb{R}} (S_X(\omega) + S_B(\omega)) \left| H(\omega) - \frac{S_X(\omega)}{S_X(\omega) + S_B(\omega)} \right|^2 d\omega + \int_{\mathbb{R}} \frac{S_X(\omega) S_B(\omega)}{S_X(\omega) + S_B(\omega)} d\omega \\ &\geq \int_{\mathbb{R}} \frac{S_X(\omega) S_B(\omega)}{S_X(\omega) + S_B(\omega)} d\omega \end{aligned}$$

avec égalité uniquement si  $H(\omega) = H_W(\omega)$  où

$$H_W(\omega) = \frac{S_X(\omega)}{S_X(\omega) + S_B(\omega)}$$

est le filtre de **Wiener-Hopf**, aussi appelé filtre de Wiener.

# Filtre de Wiener vs. filtre idéal

## ■ Filtre de Wiener

$$H_W(\omega) = \frac{S_X(\omega)}{S_X(\omega) + S_B(\omega)} \leq 1$$

Pour les fréquences  $\omega$  où  $S_X(\omega) \gg S_B(\omega)$ , on a  $H_W(\omega) \approx 1$ , et les fréquences où  $S_X(\omega) \ll S_B(\omega)$  alors  $H_W(\omega) \approx 0$ . Donc, l'avantage que l'on peut espérer par rapport à un filtrage idéal est limité aux cas où  $S_X(\omega)$  et  $S_B(\omega)$  sont du même ordre de grandeur (faible RSB).

# Processus stochastiques discrets

Le processus  $X[\cdot]$  (réel) est stationnaire au sens large (SSL) ssi:

- $\forall n \in \mathbb{Z} : \mathbb{E}\{X[n]\} = \mu_X = \text{Constante}$
- $\forall m, n \in \mathbb{Z} : \mathbb{E}\{X[m]X[n]\} \triangleq \rho_X[n - m] = \rho_X[m - n]$

## ■ Fonction d'autocorrélation statistique discrète $\rho_X : \mathbb{Z} \rightarrow \mathbb{R}$

$$\rho_X[n] \triangleq \mathbb{E}\{X[0]X[n]\} = \mathbb{E}\{X[m]X[m+n]\}, \forall m \in \mathbb{Z}$$

## ■ Densité spectrale de puissance ( $2\pi$ -périodique)

$$S_X(e^{j\omega}) = \lim_{K \rightarrow \infty} \frac{1}{2K} \mathbb{E}\{|\hat{X}_K(e^{j\omega})|^2\} \quad \text{avec} \quad \hat{X}_K(e^{j\omega}) = \sum_{n=-K}^K X[n]e^{-j\omega n}$$

## ■ Théorème de Wiener-Khintchine discret

$$S_X(e^{j\omega}) = \sum_{n \in \mathbb{Z}} \rho_X[n]e^{-j\omega n} = \mathcal{F}_d\{\rho_X\}(\omega)$$

$$\Rightarrow \mathbb{E}\{|X[n]|^2\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(e^{j\omega}) d\omega \quad (\text{énergie moyenne du signal})$$

# Filtrage statistique en temps discret

## ■ Filtrage réel d'un processus stochastique SSL : $Y[n] = (h * X)[n]$

- $X[\cdot]$  SSL  $\Rightarrow Y[\cdot]$  SSL

- $\rho_Y[n] = (h^\vee * h * \rho_X)[n]$  avec  $h^\vee[n] = h[-n]$

- $S_Y(e^{j\omega}) = |H(e^{j\omega})|^2 S_X(e^{j\omega})$

## ■ Bruit blanc discret: $B[\cdot]$ SSL à moyenne nulle avec $S_B(e^{j\omega}) = \sigma_0^2$

$$\Leftrightarrow \mathbb{E}\{B[n]\} = 0 \quad \text{et} \quad \mathbb{E}\{B[n]B[n']\} = \sigma_0^2 \delta[n - n']$$

## ■ Somme de processus aléatoires mutuellement indépendants

$X[\cdot]$  et  $Y[\cdot]$  SSL à moyenne nulle avec  $\mathbb{E}\{X[m]Y[n]\} = 0, \forall m, n \in \mathbb{Z}$

$\Rightarrow X[\cdot] + Y[\cdot]$  SSL à moyenne nulle avec DSP  $S_{X+Y}(e^{j\omega}) = S_X(e^{j\omega}) + S_Y(e^{j\omega})$

## ■ Filtre de Wiener discret: $H_W(e^{j\omega}) = \frac{S_X(e^{j\omega})}{S_X(e^{j\omega}) + S_B(e^{j\omega})}$

- Modèle stochastique de mesure:  $Y[n] = X[n] + B[n]$

- Estimateur:  $\tilde{X}[n] = (h_W * Y)[n]$  tel que  $\mathbb{E}\{|\tilde{X}[n] - X[n]|^2\}$  minimum