Chapitre 6. Canaux et signaux continus

1. Notions de base (survol rapide → voir notes à tête reposée)
   (a) Processus aléatoires en temps continu
   (b) Théorème d'échantillonnage
   (c) Entropies différentielles et théorème AEP
2. Canaux continus
   (a) Canal Gaussien (modèle abstrait, en temps discret)
   (b) Canaux à bande passante limitée (en temps continu)
   (c) Canaux parallèles et bruit coloré
   (d) Espaces de signaux (introduction au traitement du signal)

Processus aléatoires en temps continu

1. Fonction aléatoire (très général)

Espace probabilisé : $(\Omega, \mathcal{E}, P)$
Ensemble $\mathcal{T}$ d'indices (fini, infini, continu. . .)

Fonction aléatoire : une fonction de deux arguments

$$\begin{aligned} \mathcal{X}(\cdot,\cdot) \quad : \quad & \mathcal{T} \times \Omega \rightarrow \mathcal{F} \\ & (t,\omega) \longmapsto \mathcal{X}(t,\omega) \end{aligned} \tag{1}$$

2. Spécialisation (dans ce chapitre)

$\mathcal{F}=\mathbb{R}$ et $\mathcal{T}=\mathbb{R}$: valeurs et temps continu.

$(\Omega, \mathcal{E}, P)$: peut être discret ou continu (p.ex. : source discrète, canal continu)

Ressemble à un vecteur aléatoire de dimension infinie...

Si on fixe t : $\mathcal{X}(t,\cdot)$ devient une v.a. “classique” (on la note $\mathcal{X}(t)$)

Si on fixe ω : $\mathcal{X}(\cdot,\omega)$ devient une fonction “classique”

Modélisation probabiliste de processus aléatoires

NB. En principe une conséquence de la définition et de la loi $P(\cdot)$ définie sur Ω.

Point de vue descriptif :

Loi de probabilité fini-dimensionnelle du processus à l'ordre n = la loi du vecteur aléatoire $(\mathcal{X}(t_1), \mathcal{X}(t_2); \ldots, \mathcal{X}(t_n))^T$ donnée pour toute suite d'instants $(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n$.

Loi temporelle d'un processus aléatoire = l'ensemble de toutes les lois fini-dimensionnelles du processus à tout ordre.

Processus Gaussien = processus dont toutes les lois fini-dimensionnelles sont Gaussiennes.

Processus blanc = processus dont toutes les lois fini-dimensionnelles se factorisent (indépendance).

Processus Gaussien blanc = processus Gaussien avec matrices de variance-covariance diagonales.

Notions élémentaires

Moyenne du processus aléatoire (fonction du temps) :

$$m_{\mathcal{X}}(t) \stackrel{\triangle}{=} E\{\mathcal{X}(t)\}. \tag{2}$$

Processus centré si moyenne identiquement nulle. Sinon, version centrée =

$$\mathcal{X}_c(t) \stackrel{\triangle}{=} \mathcal{X}(t) - m_{\mathcal{X}}(t), \tag{3}$$

Fonction d'autocovariance (parfois d'autocorrélation) (deux arguments)

$$R_{\mathcal{X}\mathcal{X}}(t_1, t_2) \stackrel{\triangle}{=} E\{\mathcal{X}_c(t_1)\, \mathcal{X}_c(t_2)\}. \tag{4}$$

On a

1. $R_{\mathcal{X}\mathcal{X}}(t, t) \geq 0$,
2. $R_{\mathcal{X}\mathcal{X}}(t_1, t_2) = R_{\mathcal{X}\mathcal{X}}(t_2, t_1)$,
3. $|R_{\mathcal{X}\mathcal{X}}(t_1, t_2)|^2 \leq R_{\mathcal{X}\mathcal{X}}(t_1, t_1)\, R_{\mathcal{X}\mathcal{X}}(t_2, t_2)$ (inégalité de Schwarz).

Stationnarité/ergodicité

Voir notes pour les détails et précautions...

Stationnarité au sens large

1. $m_{\mathcal{X}}(t)$ constante.

2. $R_{\mathcal{X}\mathcal{X}}(t_1, t_2)$ ne dépend que de la différence $t_1 - t_2 = \tau$ (on note $R_{\mathcal{X}\mathcal{X}}(\tau)$)

Densité spectrale de puissance (processus stationnaire)

(Ici définition mathématique, interprétations : voir cours sur les proc.al.)

$$S_{\mathcal{X}\mathcal{X}}(f) \stackrel{\triangle}{=} \int_{-\infty}^{+\infty} R_{\mathcal{X}\mathcal{X}}(\tau)\, e^{-2i\pi f\tau} d\tau, \tag{5}$$

(transformée de Fourier de $R_{\mathcal{X}\mathcal{X}}(\tau)$.)

Interprétation : spectre de la répartition de puissance dans le signal (en moyenne).

Autres définition : espérance mathématique du carré du spectre des réalisations

Théorème d'échantillonnage (Shannon et Nyquist)

Trajectoire $\mathcal{X}(t)$ d'un processus (aléatoire ou non) limitée en fréquence : transformée de Fourier de $\mathcal{X}(\cdot)$

$$\mathcal{F}(f) \stackrel{\triangle}{=} \int_{-\infty}^{+\infty} \mathcal{X}(\tau)\, e^{-2j\pi f\tau} d\tau, \tag{6}$$

existe et soit telle que

$$\mathcal{F}(f) = 0, \forall |f| > f_0, \tag{7}$$

où f_0 désigne la largeur de bande du signal (en Hz).

Le théorème d'échantillonnage dit que la fonction $\mathcal{X}(t)$ est dans ces conditions complètement déterminée par des échantillons

$$\mathcal{X}\left(\frac{i}{2f_0}\right), \forall i = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots, +\infty.$$

Considérons la fonction sinc$(\cdot)$ définie par

$$\text{sinc}(x) \stackrel{\triangle}{=} \frac{\sin(x)}{x}. \tag{8}$$

Cette fonction vaut 1 en t=0 et $\mathrm{sinc}(i\pi)=0, \forall i \neq 0$.

A partir de cette fonction on peut construire une base orthonormée de signaux limités en fréquence dans la bande $[-f_0, +f_0]$ comme suit :

$$\phi_i(t)=\sqrt{2f_0}\,\mathrm{sinc}\left(2\pi f_0\left(t-\frac{i}{2f_0}\right)\right), \forall i=\ldots-2,-1,0,1,2,\ldots. \tag{9}$$

Cette base est orthonormée car

$$\langle\phi_i(\cdot), \phi_j(\cdot)\rangle=\int_{-\infty}^{\infty}\phi_i(t)\,\phi_j(t)\,dt=\delta_{i,j}. \tag{10}$$

Donc la fonction

$$g(t)=\sum_{n=-\infty}^{n=+\infty}\frac{\mathcal{X}(\frac{n}{2f_0})}{\sqrt{2f_0}}\phi_n(t) \tag{11}$$

est évidemment identique à $\mathcal{X}(t)$ aux instants d'échantillonnage. Comme cette fonction et une superposition de fonctions à spectre limité dans une même bande, elle est aussi à spectre limité dans cette bande. Le théorème d'échantillonnage garantit donc qu'elle doit être identique à $\mathcal{X}(t), \forall t \in \mathbb{R}$.

Entropies différentielles (Rappels)

$$H_d(\mathcal{X}) \stackrel{\triangle}{=} -\int \log\left[p_{\mathcal{X}}(x)\right] p_{\mathcal{X}}(x)\, dx, \tag{12}$$

(extension au cas de vecteurs aléatoires : intégrale multiple)

Invariante par translation de la v.a.

Produit par une matrice (non-singulière) : il faut ajouter $\log|\boldsymbol{A}|$

Exemples :

Loi uniforme

$$H_d(\mathcal{X}) = \log \mathrm{Vol}(S)\,. \tag{13}$$

Plus généralement, soit $\boldsymbol{\mathcal{X}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ une v.a. Gaussienne dans $\mathbb{R}^n$, alors

$$H_d(\boldsymbol{\mathcal{X}}) = \frac{1}{2}\log\left((2\pi e)^n\,|\boldsymbol{\Sigma}|\right) \tag{14}$$

où $|\boldsymbol{\Sigma}|$ désigne le déterminant de $\boldsymbol{\Sigma}$ (nécessairement non-négatif).

Propriétés de la loi Gaussienne

La loi Gaussienne est la loi qui maximise l'entropie différentielle sous les contraintes $E\{\mathcal{X}\}$=μ et Var$\{\mathcal{X}\}$=σ^2.

Comme $E\{\mathcal{X}^2\}$=Var$\{\mathcal{X}\}$ + μ^2, on en déduit immédiatement que la loi qui maximise l'entropie sous la seule contrainte d'égalité $E\{\mathcal{X}^2\}$=P est la loi Gaussienne $\mathcal{N}(0, P)$.

Donc, finalement : la loi qui maximise l'entropie sous la contrainte d'inégalité $E\{\mathcal{X}^2\} \leq P_0$ est la loi Gaussienne $\mathcal{N}(0, P_0)$.

Propriétés de la loi uniforme

La loi uniforme sur un ensemble S est la loi qui maximise l'entropie sous la contrainte $\int_S p(x)\, dx$=1.

Autres remarques

L'entropie différentielle de lois multidimensionnelles dégénérées (la probabilité étant concentrée sur un ensemble de volume nul) n'est pas définie.

Théorème AEP pour des v.a. continues

Le théorème AEP reste valable à condition d'effectuer les changements suivants :

- $P \to p$ (remplacement des probabilités par des densités),
- $H \to H_d$ (remplacement de l'entropie par l'entropie différentielle),
- $|\cdot| \to \mathrm{Vol}(\cdot)$ (remplacement des cardinalités par des volumes).

Il se formule donc de la manière suivante :

Soit $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n$ *une suite de v.a. i.i.d. selon* $p(\cdot)$. *Alors*

$$-\frac{1}{n}\log p(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_n) \xrightarrow{P} E\{-\log p(x)\} = H_d(\mathcal{X}) . \quad (15)$$

Ensembles typiques

Pour $\epsilon > 0$ et $\forall n$, ensemble typique $A_\epsilon^{(n)}$ par rapport à la densité $p(\cdot)$:

$$A_\epsilon^{(n)} \stackrel{\triangle}{=} \left\{ (X_1, \ldots, X_n) \in \mathcal{X}^n \ : \left| -\frac{1}{n} \log p(X_1, \ldots, X_n) - H_d(\mathcal{X}) \right| \leq \epsilon \right\}, \tag{16}$$

où $p(X_1, \ldots, X_n) = \prod_{i=1}^n p(X_i)$.

L'ensemble typique a les propriétés fondamentales suivantes :

1. $P(A_\epsilon^{(n)}) > 1 - \epsilon$, pour n suffisamment grand.
2. $\mathrm{Vol}(A_\epsilon^{(n)}) \leq 2^{n(H_d(\mathcal{X})+\epsilon)}$, pour tout n.
3. $\mathrm{Vol}(A_\epsilon^{(n)}) \geq (1-\epsilon)\, 2^{n(H_d(\mathcal{X})-\epsilon)}$, pour n suffisamment grand.

De plus, on montre que le volume minimal de tout sous-ensemble de $\mathcal{X}^n$ de probabilité supérieure à $1 - \epsilon$ est essentiellement le volume de l'ensemble typique.

Canaux continus (alphabet d'entrée et de sortie continu $\in \mathbb{R}$)

Un canal continu peut être continu ou discret en temps.

Un canal continu peut être utilisé avec des entrées "discrètes" (cf. modulation)

Modèle simple = bruit additif Gaussien : $\mathcal{Y}_i = \mathcal{X}_i + \mathcal{Z}_i, \quad \mathcal{Z}_i \sim \mathcal{N}(0, N)$.

NB. Capacité infinie si N=0, ou si $\mathcal{X}$ est illimité (puissance du signal illimitée).

Puissance moyenne (par symbole transmis) limitée : toute suite de symboles $(X_1, \ldots, X_n)$ transmise sur le canal vérifie

$$\frac{1}{n} \sum_{i=1}^{n} X_i^2 \leq P. \tag{17}$$

Si les signaux d'entrée sont ergodiques (ce que nous supposerons être le cas dans ce qui suit), cette contrainte devient lorsque $n \to \infty$ équivalente à la contrainte

$$E\{\mathcal{X}^2\} \leq P. \tag{18}$$

Capacité du canal Gaussien

Deux approches :

1. Alphabet d'entrée discret (p.ex. binaire) : $C = 1 - H_2(P_e) \leq 1$.

2. Alphabet d'entrée sans contrainte : $C = \frac{1}{2} \log(1 + \frac{P}{N})$.

Explications:

1. Alphabet binaire discret :

Sous la contrainte de puissance : $+\sqrt{P}$ et $-\sqrt{P}$.

$P_e = P(Z > \sqrt{P})$: canal binaire symétrique avec $p = P_e$.

2. Capacité sans restrictions supplémentaires :

Comme le canal est sans mémoire (bruit i.i.d.), on peut supposer que les entrées sont également indépendantes : on raisonne symbole par symbole (on suppose que les entrées sont i.i.d. selon une loi $p(x)$.)

Définition : capacité en information = $C = \max_{p(x):E\{\mathcal{X}^2\}\leq P} I(\mathcal{X};\mathcal{Y})$.

On montre (cf. notes) :

Capacité est réalisée pour $p(x) \sim \mathcal{N}(0, P)$ et vaut $\frac{1}{2}\log(1 + \frac{P}{N})$.

N.B :

1. Dans ce cas, la sortie est distribuée en loi $\mathcal{N}(0, P + N)$.

2. Résultat conceptuellement analogue au cas du canal symétrique binaire.

Discussion (intuitivement évident) :

L'utilisation d'un alphabet continu permet en principe d'augmenter la capacité.

D'autant plus que le rapport signal/bruit est élevé.

Montrons que la capacité en information est égale au débit maximum atteignable...

Code (M, n) Un code (M, n) pour le canal Gaussien avec limitation de puissance P consiste en :

1. Un ensemble d'indices $\{1, 2, \ldots, M\}$.
2. Une fonction d'encodage $x^n(\cdot) : \{1, 2, \ldots, M\} \to \mathcal{X}^n$, produisant les mots de code $x^n(1), \ldots, x^n(M)$ qui vérifient la contrainte de puissance moyenne pour chaque mot de code, i.e.

$$\sum_{i=1}^{n} x_i^2(w) \leq nP, \ \forall w \in \{1, 2, \ldots, M\}. \tag{19}$$

3. Une fonction de décodage $g(\cdot) : \mathcal{Y}^n \to \{1, 2, \ldots, M\}$.

Débit réalisable

Un débit R est dit réalisable pour le canal Gaussien avec limitation de puissance P s'il existe une suite de codes $(\lceil 2^{nR} \rceil, n)$ (respectant la limitation de puissance) telle que la probabilité d'erreur maximale $\lambda^{(n)}$ tende vers zéro.

Empilement de sphères (argument de plausibilité)

On se pace dans la situation qui réalise la capacité en information.

Si on émet un vecteur $\boldsymbol{x}^n$ donné de dimension n, le signal reçu est confiné (avec probabilité proche de 1) dans une boule centrée en $\boldsymbol{x}^n$ de rayon $\sqrt{n(N+\epsilon)}$.

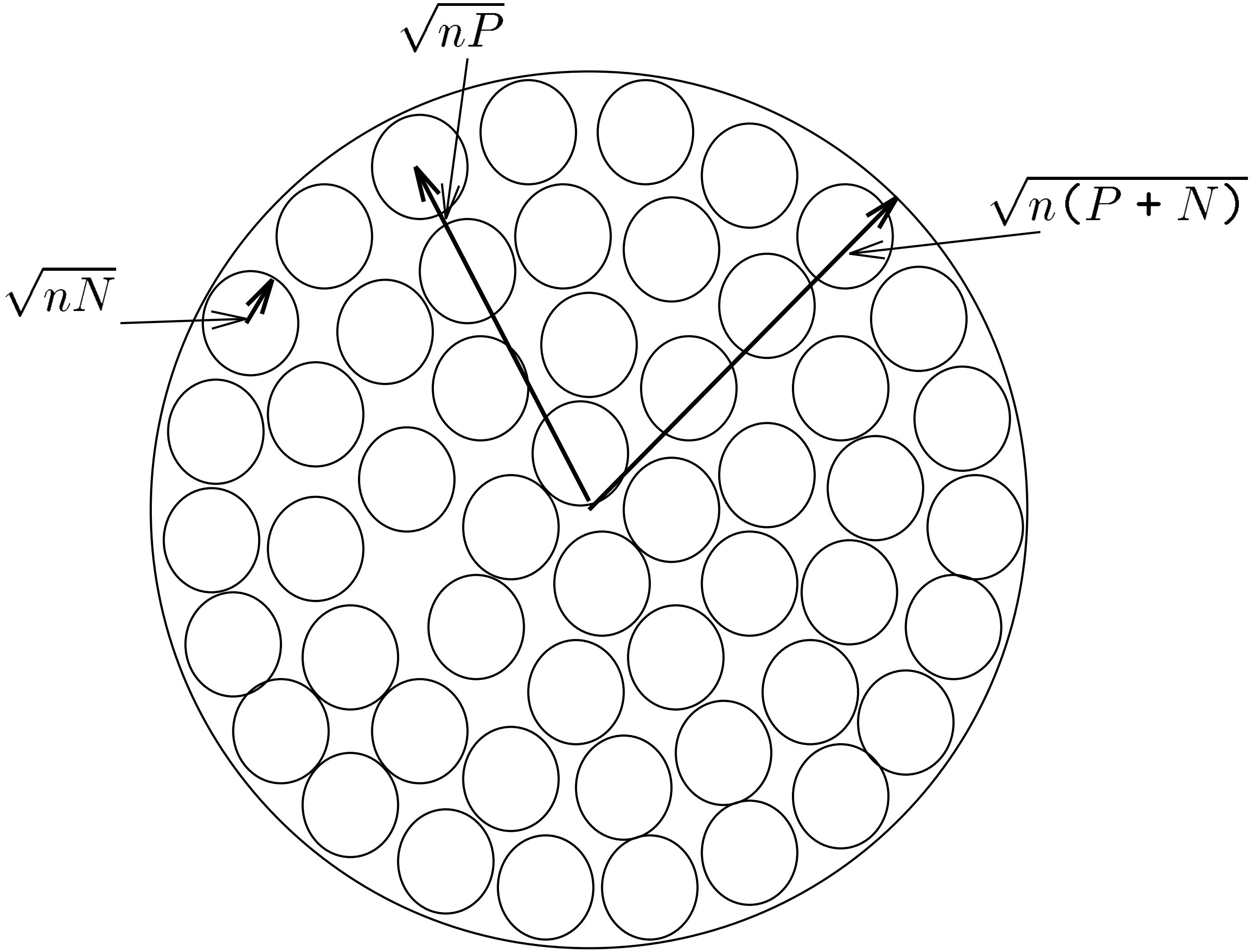
En fait, le signal tend à se concentrer à la surface de la sphère, car $\frac{1}{n}\sum_{i=1}^{n} z_i^2 \xrightarrow{P} N$.

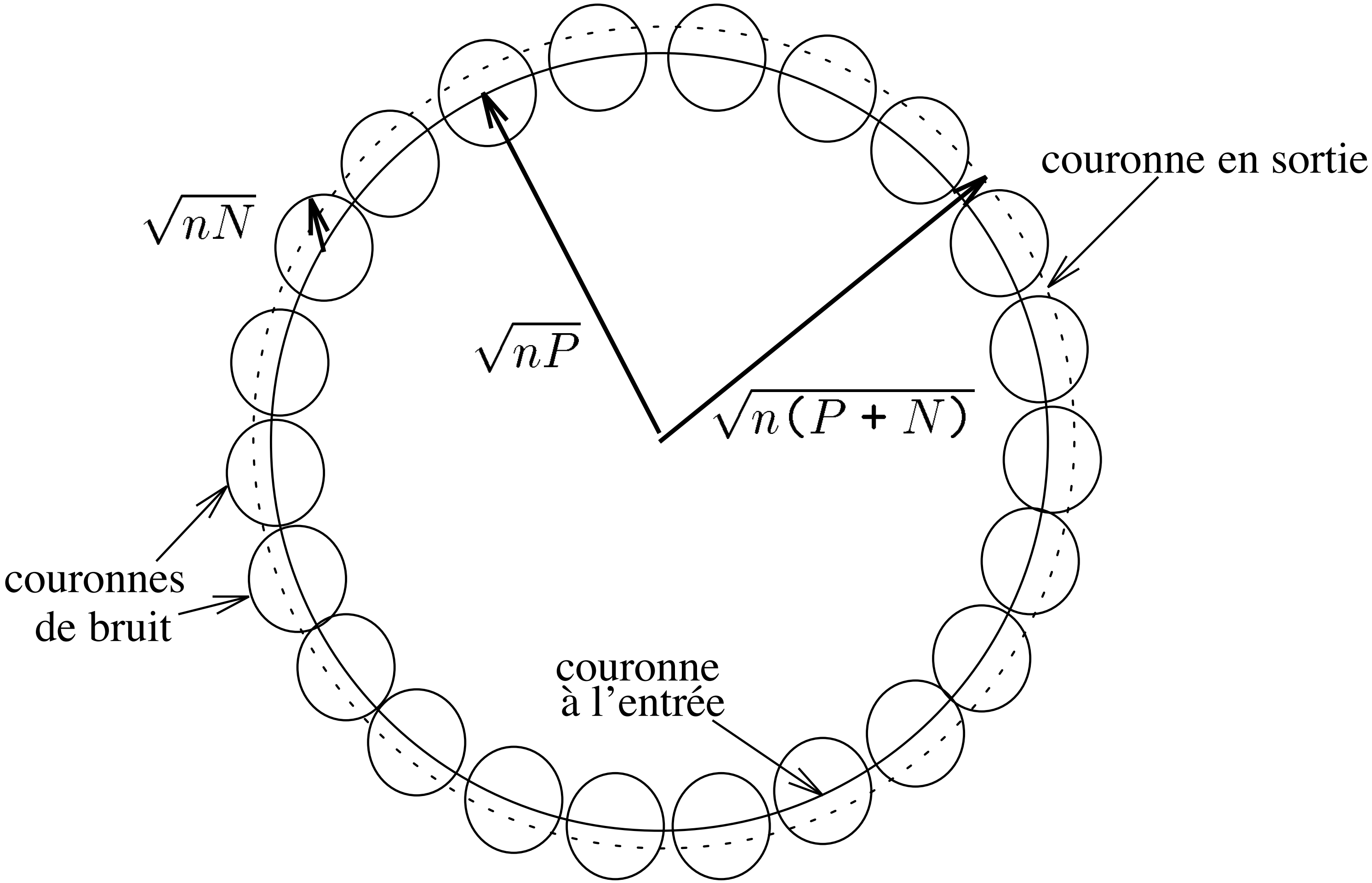
Plaçons les signaux d'entrée de manière à ce que les sphères ne se recouvrent pas : combien de sphères peut on ainsi empiler ?

Comme signaux limités en puissance : ils doivent se trouver à l'intérieur d'une sphère de rayon $\sqrt{nP}$.

On montre que les vecteurs reçus sont alors confinés (quel que soit le code utilisé) dans une sphère de rayon $\sqrt{n(P+N)}$.

Rapport des volumes : sphère de réception/sphères de bruit = $2^{\frac{n}{2}\log(1+\frac{P}{N})}$ qui fixe une borne à $M = \lceil 2^{nR} \rceil$.


$\sqrt{nP}$
$\sqrt{n(P+N)}$
$\sqrt{nN}$


couronne en sortie
$\sqrt{nN}$
$\sqrt{nP}$
$\sqrt{n(P+N)}$
couronnes
de bruit
couronne
à l'entrée

Démonstrations “réelles” $\rightarrow$ voir notes

Applications : canaux à bande passante limitée

Modèle :

$$Y(t) = (X(t) + Z(t)) * h(t)\,, \tag{20}$$

où $X(t)$ est le signal temporel d'entrée, $Z(t)$ une réalisation de bruit blanc Gaussien et $h(t)$ est la réponse impulsionnelle d'un filtre passe bande idéal qui coupe toutes les composantes fréquentielles au-delà d'un seuil f_0 (le symbole $*$ désigne le produit de convolution).

NB: un bruit blanc $\mathcal{X}(\omega, t)$ est un processus aléatoire stationnaire tel que $E\{\mathcal{X}(t)\} = 0, \forall t$, et tel que

$$R_{\mathcal{X}\mathcal{X}}(\tau) \stackrel{\triangle}{=} E\{\mathcal{X}(t)\,\mathcal{X}(t-\tau)\} = \frac{N_0}{2}\delta(\tau)\,, \tag{21}$$

où $\delta(\cdot)$ représente l'impulsion de Dirac (idéalisation mathématique) et $\frac{N_0}{2}$ la variance du bruit. Le bruit blanc est dit Gaussien si de plus, les $\mathcal{X}(t)$ suivent une loi Gaussienne $\forall t$.

Comme le signal en sortie est limité en fréquence, on peut considérer que les entrées et les sorties sont définies par les échantillons aux instants $t_i = \frac{i}{2f_0}$.

On montre que, pour un bruit blanc ces échantillons sont i.i.d. $\mathcal{N}(0, N_0 f_0)$.

Supposons que nous disposions d'une puissance d'émission P : on peut générer un signal aléatoire Gaussien limité en fréquence en choisissant les échantillons distribués selon une loi $\mathcal{N}(0, P)$.

On a donc une capacité (par utilisation du canal, i.e. par échantillon) qui vaut :

$$C = \frac{1}{2} \log \left(1 + \frac{P}{N_0 f_0}\right) \text{ bits par échantillon,} \tag{22}$$

et comme il y a $2f_0$ échantillons par seconde, on obtient

$$C = f_0 \log \left(1 + \frac{P}{N_0 f_0}\right) \text{ bits par seconde.} \tag{23}$$

En pratique, N_0 est une donnée physique. P et f_0 sont des grandeurs sur lesquelles l'ingénieur peut agir.

NB: la capacité croît avec la largeur de bande

$$\lim_{f_0 \to \infty} f_0 \log\left(1 + \frac{P}{N_0 f_0}\right) = \frac{P}{N_0} \log e \text{ bits par seconde.} \tag{24}$$

Si la largeur de bande est très large, on peut utiliser des schémas de modulation discrets (p.ex. binaire) : la capacité exploitable, à la limite vaut environ 63% de la capacité théoriquement atteignable (cela est lié à la décision brute à la sortie du démodulateur. . .).

Exemple

La largeur de bande d'une ligne téléphonique est limitée à 3300Hz (ce qui convient parfaitement à la parole). En supposant que le rapport signal bruit soit de 20dB (=$10 \log_{10} \frac{P}{N_0 f_0}$) on a

$$\frac{P}{N_0 f_0} = 100,$$

et on calcule que C=21.972 bits par seconde.

Conclusions générales

Si bande étroite et rapport signal bruit élevé : le codage de canal est peu utile.

Si rapport signal bruit faible, mais bande de fréquence large : codage de canal nécessaire.

Si bande de fréquence plus large, la plage de puissance ou la croissance de la capacité est plus large.

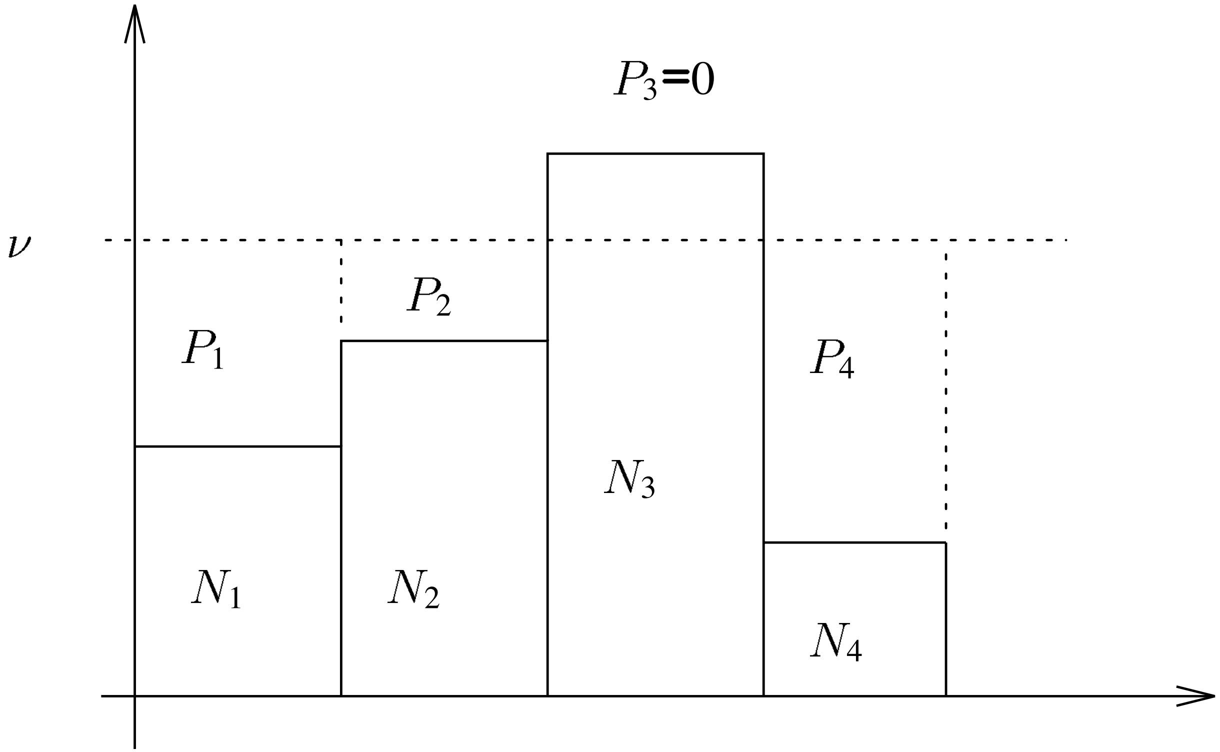
Si rapport signal bruit élevé: modulations discrètes pénalisées.

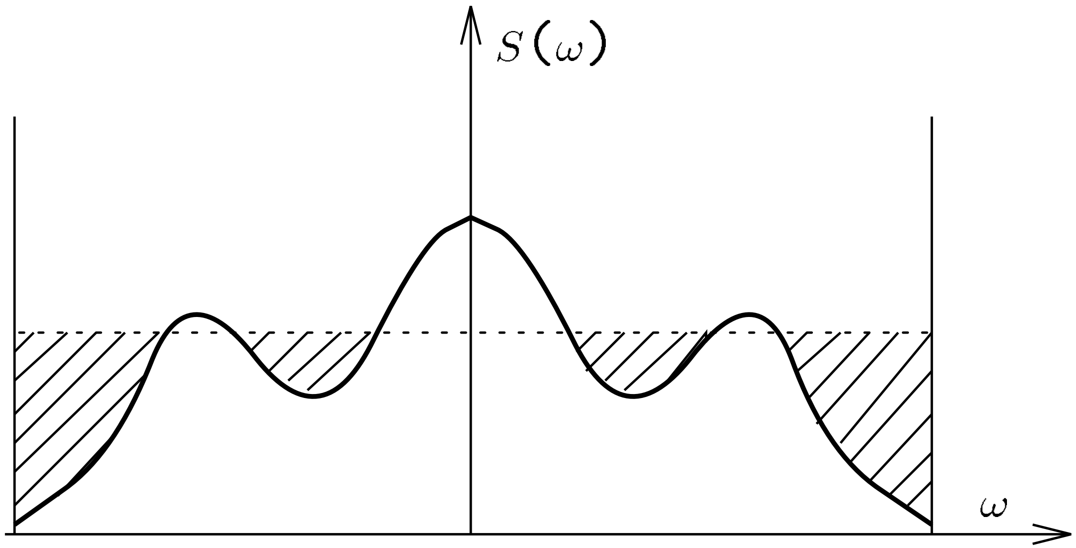
Autres résultats (voir notes)

Décodage : reconnaître le signal envoyé sur le canal.

Approximation : codage discret de signaux continus.

Canaux parallèles et canaux avec spectre de bruit quelconque :


P_3=0
ν
P_2
P_1
P_4
N_3
N_1
N_2
N_4


$S(\omega)$
ω

Moralité

Théorie de l'information dit : il faut exploiter un canal en maximisant la discernabilité des signaux envoyés, compte tenu des déformations introduites par le canal, et des limitations techniques imposées au concepteur.

On peut augmenter la dimensionnalité des espaces de signaux pour atteindre la capacité.

Dans certains cas, il faut augmenter l'alphabet de source (nombre de niveaux possibles pour chaque symbole).

Dans d'autres cas il faut coder des messages plus longs (réseaux de points dans un espace multidimensionnel).

Ce que ne nous n'avons pas pu voir:

- Codes sur un espace euclidien
  (Signaux et canaux continus)

- Cryptographie
  (Rendre le décodage difficile)

- Théorie de la distorsion
  (Compression irréversible)

- Théorie de l'information de réseaux de communication

- Relation entre théorie de l'information et physique statistique
  (Thermodynamique)

- Applications de la théorie de l'information
  (statistiques et apprentissage automatique)

- Complexité de Kolmogorov
  (relations avec l'informatique théorique : décidabilité, complexité. . .)