



Cours et Travaux Dirigés de
Traitement du Signal Déterministe

Benoît Decoux (benoit.decoux@wanadoo.fr)

Support de cours - v 1.1

1^{ère} partie : "Notions de base et études temporelles"

Sommaire

0) Introduction générale	4
0.1) Qu'est-ce qu'un signal ?	4
0.2) Qu'est-ce que le traitement de signal ?	5
0.3) Le traitement du signal en pratique	5
0.4) Applications	5
0.5) Plan du cours	5
I) Les bases du traitement de signal	6
I.1) Catégories de signaux.....	6
I.1.1) Déterministes vs aléatoires	6
I.1.2) Périodiques vs non-périodiques	6
I.2.3) Continus vs numériques	7
I.2) Signaux usuels.....	7
I.2.1) Signal sinusoïdal.....	7
a) Définition.....	7
b) Représentation vectorielle (de Fresnel)	8
c) Représentation complexe.....	8
d) Dérivation, intégration.....	8
I.2.2) Autres signaux périodiques	9
I.2.3) Signaux « mathématiques ».....	9
a) Fonction signe.....	10
b) Echelon unité	10
c) Fonction rectangulaire	10
d) Fonction triangulaire	11
e) Rampe	11
f) Sinus cardinal	11
g) Impulsion de Dirac	12
I.3) Caractéristiques des signaux	14
I.3.1) Période, fréquence.....	14
I.3.2) Valeur moyenne.....	14
I.3.3) Valeur efficace	14
I.3.4) Energie, puissance	15
a) Analogie électrique	15
b) Energie totale.....	15
c) Puissance moyenne totale	16
d) Rapport signal/bruit.....	16
e) Lien entre puissance et valeur efficace	17
f) Répartition de la puissance dans un signal.....	17
g) Mesure	17
I.4) Signaux et systèmes	17
I.4.1) Linéarité	17
I.4.2) Stationnarité	18
I.4.3) Causalité.....	18
I.4.4) Stabilité.....	18
II) Etude des signaux et systèmes dans le domaine temporel	18
II.1) Réponse impulsionnelle d'un système.....	18
II.1.1) Produit de convolution	18
II.1.2) Convolution avec réponse impulsionnelle	20
a) Expression de la relation entrée-sortie pour une entrée quelconque	21
b) Démonstration	21
c) Systèmes physiques	21
d) Exemple : circuit RC	22

c) Domaine discret	24
II.2) Corrélation.....	29
II.2.1) Définitions.....	29
II.2.2) Début d'interprétation	30
a) Lien avec produit scalaire	30
b) Comparaison avec convolution	30
c) Mesure de ressemblance	31
II.2.3) Quelques propriétés.....	31
II.2.4) Cas discret	32
a) Définitions	32
b) Domaines des statistiques et des probabilités.....	33
c) Corrélation bi-dimensionnelle : application aux images	34
d) Gestion des effets de bord	34
e) Exemples numériques.....	35
f) Utilisation de Scilab	36
g) Applications.....	37
d) Traitement de la parole : détection de voisement.....	41
e) Reconnaissance de caractères.....	42
f) Analyse de texture.....	43
II.3) Caractérisation des systèmes par équations différentielles.....	43
II.3.1) Résolution directe des équations différentielles linéaires à coefficients constants	44
a) Equations différentielles du 1 ^{er} ordre	44
b) Equations différentielles du 2 ^e ordre	46
II.3.2) Résolution par Transformée de Laplace	51
a) Rappels sur la Transformée de Laplace.....	51
b) Utilisation de la Transformée de Laplace pour la résolution des équations différentielles	53
c) Cas particulier du régime sinusoïdal	56

0) Introduction générale

0.1) Qu'est-ce qu'un signal ?

- une grandeur physique variant au cours du temps
- une fonction mathématique (variable : le temps)

mais également...

- une image (variables : les 2 dimensions spatiales)

Qu'est-ce que le traitement du signal ?

- A la fois très théorique et très appliqué

En physique

Un signal temporel est une grandeur physique variant au cours du temps. Il peut s'agir de n'importe quelle grandeur physique : courant, tension, pression, etc. Il est très souvent électrique, ou converti en une grandeur électrique (et la plupart du temps en tension). De plus, il est de plus en plus numérique.

Un signal peut être de dimension 1, 2 ou même plus. Dans le cas d'une seule variable, il s'agit le plus souvent du temps. Dans le cas de deux variables, on parle de signal bi-dimensionnel ; il s'agit le plus souvent des deux dimensions spatiales d'images. On entre alors dans le domaine du traitement d'images dans lequel les deux variables, les deux dimensions des images, sont spatiales. Si l'on effectue en plus des traitements utilisant plusieurs images consécutives, le temps est pris en compte et on peut considérer qu'il s'agit d'une troisième dimension, même si elle n'est pas de même nature que les deux dimensions spatiales.

En mathématiques

D'un point de vue mathématique, dans la plupart des cas un signal est une fonction, d'une ou plusieurs variable(s). Cette fonction peut être réelle ou complexe. Dans le 2^e cas, le signal ne correspond pas à une réalité physique.

Mais certains signaux ne sont pas dérivables du tout, comme une impulsion de durée infiniment petite et d'amplitude infiniment grande, appelée impulsion de Dirac, ou dérivable mais pas partout, comme un signal carré au niveau de ses transitions. La théorie des distributions permet l'analyse de ces signaux idéalisés.

On verra qu'un outil mathématique puissant d'analyse des signaux est la transformée de Fourier. Il s'agit d'une transformation faisant passer du domaine temporel au domaine des fréquences. Cet outil considère des fréquences négatives, symétriques des fréquences positives par rapport à l'axe des ordonnées. C'est le côté abstrait de cette théorie, mais elle décrit néanmoins très bien la réalité.

Signal déterministe vs signal aléatoire

Un signal est déterministe quand il est complètement prévisible, quand son évolution peut être décrite par un modèle mathématique ou physique.

Un signal physique ne peut pas être complètement prévisible. Il peut être constitué d'un signal déterministe auquel du bruit est ajouté, par exemple. Mais l'information elle-même peut avoir un aspect aléatoire, comme un signal de parole, par exemple.

0.2) Qu'est-ce que le traitement de signal ?

C'est une science physique qui utilise :

- des outils **mathématiques** : nombres complexes, transformée de Laplace, la transformée de Fourier (continue ou discrète), théorie des probabilités, formules de trigonométrie, variations de fonctions (dérivation, intégration), équations du 2^e degré, factorisation, le calcul différentiel, les séries et les suites, la transformée en z, etc.
- des outils **pratiques** : logiciels (Matlab, Scilab, etc.), matériels (DSP, FPGA, etc), voir ci-dessous.

C'est donc une discipline comportant un aspect théorique pouvant être très poussé, mais également un aspect très appliqué et concret (cf offres d'emplois "signal").

0.3) Le traitement du signal en pratique

Il existe un certain nombre d'outils, logiciels et matériels. Ces outils permettent de faire du traitement de signal en temps-réel ou en temps différé. Dans le premier cas, le signal traité de sortie est obtenu en "même temps" que le signal d'entrée. En même temps signifie que le décalage temporel n'est pas sensible pour l'application.

Logiciels

- Matlab, Scilab : ne permettent pas de faire du temps réel
- Programmation C, C++, Java

Matériel

- PC
- DSP (Digital Signal Processor : processeur spécialisé dans le traitement de signal)
- Autres circuits intégrés : FPGA, Microcontrôleurs PIC

Autrefois le traitement du signal était un domaine continu. En pratique, les systèmes de traitement de signal étaient des réalisations en électronique analogique. Exemple : analyseur de spectre, mesure de puissance, etc.

Aujourd'hui il est essentiellement numérique. En pratique, il consiste en de la programmation informatique sur ordinateurs classiques ou sur processeurs spécialisés (DSP).

0.4) Applications

Parmi les applications diverses du traitement du signal, on peut distinguer :

- Communication, téléphonie
- Vidéo
- Multimédia
- Reconnaissance et synthèse de la parole (conversion texte-parole, compression de données, etc)
- Audio, musique
- etc

0.5) Plan du cours

Les différentes parties abordées sont les suivantes :

- **aperçu général** du domaine, comportant de nombreux rappels : principaux types de signaux, **énergie et puissance**, etc ;
- étude des **systèmes** et leur lien avec le Traitement de Signal ;

- analyse dans le domaine temporel : la **convolution**, opération de base des systèmes (linéaires) ; la **corrélation**, méthode puissante d'analyse ;
- analyse des signaux en fréquence : développement en **séries de Fourier** ; généralisation à la **Transformée de Fourier** ; **fonctions de transfert**, **filtrage analogique**, comportant également de nombreux rappels ;
- étude de l'**échantillonnage**, de la transformée en z, un outil d'étude des signaux échantillonnés ;
- **filtrage numérique** : étude des deux grands types d'implémentations numériques : les filtres à réponse impulsionnelle infinie (RII) puis les filtres à réponse impulsionnelle finie (RIF). Analyse par corrélation ;
- **Transformée de Fourier discrète**, Transformée de Fourier rapide (TFR), un algorithme de calcul optimisé en nombre de calculs, techniques d'**analyse spectrale** ;

Chacune de ces parties comportera quelques **approfondissements mathématiques**, sur :

- les séries entières,
- la projection d'un signal sur une base de fonctions orthogonales
- la transformée de Laplace,
- la théorie des distributions,
- etc.

I) Les bases du traitement de signal

I.1) Catégories de signaux

I.1.1) Déterministes vs aléatoires

Un signal déterministe est complètement prévisible. Par exemple, une fois un signal sinusoïdal caractérisé au moyen de son amplitude, sa fréquence et sa phase, on peut prévoir sa valeur instantanée à n'importe quel instant dans le futur. L'aspect déterministe est décrit par des modèles mathématiques. Les signaux purement déterministes portent très peu d'information. Les signaux déterministes constituent une idéalisation de signaux réels.

L'aspect aléatoire est analysé par les modèles statistiques et les densités de probabilités (moyenne, variance, densité spectrale de puissance, etc). Dans cette catégorie de signaux on place le bruit, signaux parasites que l'on cherche en général à éliminer. Le bruit peut être blanc, rose, gaussien... Ces signaux aléatoires portent également peu d'information.

La plupart des signaux réels comportent un aspect aléatoire et un aspect périodique. Par exemple, un signal de parole comporte un aspect aléatoire dans la mesure où il comporte des parties bruitées mais également parce qu'il n'est pas prévisible (sons non-voisés : consonnes, etc). Il comporte également des parties périodiques : les sons voisés.

Il existe des méthodes pour repérer des signaux périodiques dans des signaux à l'aspect aléatoire : par exemple basées sur l'autocorrélation.

I.1.2) Périodiques vs non-périodiques

Un signal périodique se répète identique à lui-même à intervalles de temps réguliers T

$$x(t) = x(t + kT), \text{ k entier}$$

T est appelé période.

Dans les signaux non-périodiques, on peut trouver les signaux transitoires : les signaux « mathématiques » comme l'impulsion de Dirac, la fonction porte, etc.

On peut trouver également les signaux non-périodiques mais non-transitoires, comme les signaux aléatoires.

Les signaux réels sont souvent des combinaisons de signaux périodiques de fréquences multiples et variables, dans une certaine plage de valeur : la largeur de bande. Le spectre d'amplitude est une fonction continue.

I.2.3) Continus vs numériques

La plupart des résultats obtenus dans le domaine continu sont transposables dans le domaine discret. Les intégrales du premier correspondent alors à des sommes discrètes dans le deuxième.

L'opération permettant de passer du premier au deuxième est l'échantillonnage. Cette transposition est directe pourvu que les signaux soient correctement échantillonnés, c'est à dire que la fréquence d'échantillonnage soit suffisamment grande par rapport aux fréquences des signaux. Le principe de Shannon nous donnera la condition nécessaire, à savoir que la fréquence d'échantillonnage soit au moins égale au double de la fréquence maximale contenue dans le signal.

I.2) Signaux usuels

I.2.1) Signal sinusoïdal

On verra que ce signal périodique est fondamental, dans la mesure où tout signal périodique peut être décomposé en une somme de signaux sinusoïdaux d'amplitudes et de fréquences multiples entières celles du signal décomposé. Le terme "fondamental" est même le terme désignant la première composante de cette décomposition.

a) Définition

Un signal sinusoïdal est défini par :

$$s(t) = A \cos(\omega t + \varphi) \text{ ou } s(t) = A \sin(\omega t + \varphi)$$

A est la valeur maximale du signal, également appelée amplitude

ω la pulsation en radians par seconde :

$$\omega = 2\pi f$$

où f est la fréquence en Hertz. La période est l'inverse de la fréquence ; elle s'exprime en secondes.

φ est la phase à l'origine, ou plus simplement phase, exprimée en radians (rappel : $360^\circ = 2\pi \text{ rad}$).

Remarque 1 : les notations utilisées sont arbitraires : par exemple on peut trouver les notations ω_0 , f_0 et φ_0 pour bien mettre l'accent sur le fait que ce sont des constantes (même si elles peuvent varier !), que la variable de la fonction-signal est le temps t.

Remarque 2 : on parle de signal sinusoïdal, mais on pourrait très bien parler de signal cosinusoïdal. La seule différence entre les 2 est la phase. Pour simplifier, considérons le même signal que ci-dessus mais avec une phase nulle et une amplitude unitaire :

$$s(t) = \cos(\omega t)$$

On peut l'écrire également :

$$s(t) = \sin\left(\omega t + \frac{\pi}{2}\right)$$

$\sin(\theta)$ et $\cos(\theta)$ représentent les coordonnées d'un point sur un cercle de rayon égal à 1.

$\sin(\omega t)$ et $\cos(\omega t)$ représentent les coordonnées d'un point se déplaçant sur ce même cercle, à une vitesse angulaire de $\omega \text{ rad/s}$, ou de $f (= \omega/2\pi) \text{ Hertz}$.

b) Représentation vectorielle (de Fresnel)

Un signal sinusoïdal peut donc être représenté par la projection d'un vecteur tournant dans le sens positif conventionnel (=sens inverse des aiguilles d'une montre), sur l'axe des ordonnées. La longueur de ce vecteur est égale à l'amplitude du signal (=sa valeur maximale), et l'angle formé avec l'axe des abscisses la phase. La pulsation est liée à la vitesse de rotation du vecteur ; par exemple, si la pulsation vaut 2π rad/s (ce qui correspond à une fréquence de 1Hz), le vecteur mettra 1 s. pour faire un tour.

Cette représentation est utile quand on veut calculer la somme de 2 signaux sinusoïdaux, d'amplitudes et de phases différentes (par contre, la pulsation doit être la même). Par exemple, soit à additionner $a_1(t) = A_1 \cos(\omega t + \varphi_1)$ et $a_2(t) = A_2 \cos(\omega t + \varphi_2)$. Le signal résultant est un signal sinusoïdal tournant à la même pulsation que $a_1(t)$ et $a_2(t)$, mais pour connaître son amplitude et sa phase, une solution consiste à les calculer, en utilisant les formules de trigonométrie. La méthode vectorielle est une autre méthode possible, qui ne donnera que des résultats approchés mais qui a le mérite d'être rapide. En effet, l'opération somme de signaux sinusoïdaux est équivalente à effectuer une somme vectorielle des vecteurs de Fresnel correspondants.

c) Représentation complexe

Dans la représentation vectorielle de Fresnel, on a parlé uniquement de l'axe des abscisses. Dans la représentation complexe, l'axe des abscisses est l'axe réel et l'axe des ordonnées l'axe imaginaire. Dans le cas des complexes on parle de module et d'argument : le module désigne la longueur du vecteur, c'est à dire l'amplitude du signal ; l'argument désigne l'angle que fait le vecteur avec l'axe des abscisses, c'est à dire la phase du signal.

Un des avantages de la représentation complexe est de permettre l'utilisation de l'exponentielle (complexe) et de ses propriétés. L'exponentielle complexe est définie par :

$$e^{j\theta} = \cos \theta + j \sin \theta \text{ et } e^{-j\theta} = \cos \theta - j \sin \theta$$

$e^{j\theta}$ correspond à un point sur le cercle unité autour de l'origine. On l'appelle phaseur.

$e^{-j\theta}$ correspond au point symétrique de ce point, par rapport à l'axe des abscisses.

De même, $e^{j\omega t}$ et $e^{-j\omega t}$ correspondent à ces mêmes points tournant avec la pulsation ω , l'un dans le sens positif et l'autre dans le sens négatif.

Un signal sinusoïdal et un signal cosinusoidal peuvent être exprimés comme la somme et la différence, respectivement, d'exponentielles complexes (formules d'Euler) :

$$\cos \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2} \text{ et } \sin \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2j} = -j \frac{e^{j\omega t} - e^{-j\omega t}}{2}$$

Ces 2 signaux peuvent donc être vus comme la somme et la différence, respectivement, de 2 phaseurs. Dans la représentation complexe, $\cos(\omega t)$ est un signal réel pur et $\sin(\omega t)$ un signal imaginaire pur.

d) Dérivation, intégration

Soit $a(t)$ un signal sinusoïdal défini par :

$$s(t) = A \cos(\omega t + \varphi)$$

Sa dérivée est :

$$\frac{ds(t)}{dt} = -A\omega \sin(\omega t + \varphi) = A\omega \cos(\omega t + \varphi + \frac{\pi}{2})$$

Dans l'espace des vecteurs de Fresnel, le vecteur associé a pour module ωA . L'angle qu'il fait avec l'axe des abscisses (=sa phase) est $\varphi + \pi/2$.

Donc l'opération dérivation se traduit par une multiplication de l'amplitude par la pulsation, et une rotation de $+\pi/2$.

L'intégrale du signal est :

$$\int s(t)dt = \frac{A}{\omega} \sin(\omega t + \varphi) = \frac{A}{\omega} \cos(\omega t + \varphi - \frac{\pi}{2})$$

Dans l'espace des vecteurs de Fresnel, le vecteur associé a pour module A/ω . Sa phase est $\varphi - \pi/2$.

Donc l'opération intégration se traduit par une division de l'amplitude par la pulsation, et une rotation de $-\pi/2$.

On a vu que la dérivation dans l'ensemble des signaux sinusoïdaux se traduisait par une multiplication par ω et une rotation de $+\pi/2$: dans l'ensemble des nombres complexes, cela correspond à une multiplication par $j\omega$.

De même, on a vu que l'intégration dans l'ensemble des signaux sinusoïdaux se traduisait par une division par ω et une rotation de $-\pi/2$: dans l'ensemble des nombre complexes, cela correspond à une division par $j\omega$.

On peut le démontrer en dérivant un signal complexe :

$$\frac{d(e^{j\omega t})}{dt} = j\omega e^{j\omega t}$$

puis en l'intégrant :

$$\int e^{j\omega t} dt = \frac{1}{j\omega} e^{j\omega t}$$

I.2.2) Autres signaux périodiques

Les signaux de « laboratoire » les plus utilisés sont : signal carré ou rectangulaires, carré à rapport cyclique 1/2 ou autre, triangulaire, en dents de scie.

La décomposition par séries de Fourier montrera qu'il peuvent être décomposés en une somme pondérée de sinusoïdes/cosinusoïdes, avec des coefficients de pondération constants.

Les signaux carrés sont également très utilisés dans les techniques numériques de traitement du signal : à valeur moyenne nulle, signal logique (amplitude sans dimension ou en Volts ; signal TTL : amplitude 5V). Le rapport cyclique est le rapport de la durée de l'état haut sur la période.

I.2.3) Signaux « mathématiques »

Dans la formalisation et l'étude mathématique des signaux, on utilise plusieurs signaux particuliers qui constituent des idéalizations de signaux réels, périodiques ou non. On peut les appeler signaux mathématiques, dans la mesure où leur rôle est surtout de modéliser la réalité.

Un décalage d'un signal $s(t)$ vers la droite sur l'axe temporel, d'une quantité t_0 (positif) s'écrit sous la forme suivante :

$$s(t-t_0)$$

En effet, si on effectue le changement de variable $t'=t-t_0$, cela équivaut à considérer le signal $s(t')$, à la même position que le signal $s(t)$ mais dans un repère décalé. La position d'un point dans le repère original s'obtient par la relation :

$$t=t'+t_0$$

donc ce décalage est un décalage à droite.

a) Fonction signe

$$\text{sgn}(t) = \begin{cases} -1 & \text{pour } t < 0 \\ 0 & \text{pour } t = 0 \\ 1 & \text{pour } t > 0 \end{cases}$$

$$= \frac{t}{|t|} \text{ pour } t \neq 0$$

La valeur à l'origine est arbitraire, située entre -1 et $+1$. Par souci de symétrie on peut la prendre nulle par convention comme dans l'expression ci-dessus mais elle peut être prise différente.

Une fonction signe d'amplitude A et décalée de t_0 sera notée :

$$A \cdot \text{sgn}(t - t_0)$$

b) Echelon unité

Encore appelé échelon de Heaviside, il est défini par :

$$u(t) = \begin{cases} 1 & \text{pour } t > 0 \\ 0 & \text{pour } t \leq 0 \end{cases}$$

Remarque : la valeur à l'origine est conventionnelle. On aurait pu également la fixer à 1 ou à $1/2$, par exemple.

Cette fonction est utilisée par exemple pour rendre un signal causal. Soit $x(t)$ un signal défini de $-\infty$ à $+\infty$, donc non-causal. Le signal obtenu par la multiplication $x(t) \cdot u(t)$ est causal.

c) Fonction rectangulaire

Elle est également appelée fonction porte en mathématiques.

Conventionnellement, on utilise l'expression suivante :

$$x(t) = A \cdot \text{rect}[t/T]$$

pour décrire une fonction rectangulaire d'amplitude A et de durée T , centrée sur 0 .

Dans le cas plus général où la fonction est centrée sur t_0 :

$$x(t) = A \cdot \text{rect}[(t - t_0)/T]$$

La version normalisée possède une intégrale unité. Son amplitude est égale à $1/T$:

$$x(t) = \frac{1}{T} \text{rect}[(t - t_0)/T]$$

Remarque : on peut également définir la fonction rect avec une seule variable, mais telle que la fonction a été définie ci-dessus, cette variable est sans dimension. Soit t' cette variable :

$$\text{rect}(t') = \begin{cases} 1 & \text{pour } |t'| < \frac{1}{2} \\ 0 & \text{ailleurs} \end{cases}$$

Cette fonction est alors à intégrale unité.

Cette fonction est souvent utilisée pour "fenêtrer" un signal, par exemple pour rendre la durée du signal étudié finie, ou pour localiser un segment de ce signal. Cette opération consiste alors en une multiplication. Par exemple, le signal suivant est fenêtré sur une durée T autour de $t=0$:

$$x_T(t) = x(t) \cdot \text{rect}(t/T)$$

d) Fonction triangulaire

La fonction triangulaire est définie avec le même formalisme que pour la fonction porte. Conventionnellement, pour désigner une fonction triangulaire de durée T et d'amplitude A , on utilise l'expression suivante :

$$x(t) = A.\text{tri}[t/T]$$

On définit donc cette fonction avec une variable sans dimension t' , une intégrale unité, une durée égale à 2 et une amplitude égale à 1, par :

$$\text{tri}(t') = \begin{cases} 1 - |t'| & \text{pour } |t'| \leq 1 \\ 0 & \text{pour } |t'| > 1 \end{cases}$$

Dans le cas plus général où la fonction est centrée sur t_0 :

$$x(t) = A.\text{tri}[(t - t_0)/T]$$

e) Rampe

Le signal de rampe est défini par :

$$r(t) = \begin{cases} t & \text{pour } t \geq 0 \\ 0 & \text{pour } t < 0 \end{cases}$$

On peut l'obtenir en rendant causale la fonction t :

$$r(t) = t.u(t)$$

On peut remarquer qu'elle peut s'obtenir également à partir de l'intégrale de fonction échelon :

$$r(t) = \int_{-\infty}^t u(\tau) d\tau$$

L'intégrale d'une constante est en effet une rampe de pente égale à cette constante.

Inversement, on peut définir l'échelon comme la dérivée de la fonction rampe :

$$u(t) = \frac{dr(t)}{dt} \text{ pour } t \neq 0$$

La restriction $t \neq 0$ tient au fait que la fonction n'est pas dérivable en 0. On verra que cette dérivée pourra être définie dans le cadre de la théorie des distributions.

f) Sinus cardinal

Ce signal est défini par :

$$\text{sin c}(x) = \frac{\sin \pi x}{\pi x}$$

Cette fonction s'annule pour $\sin(\pi x) = 0$, soit $x = k$ avec k entier.

Il est important de constater que la fonction est bien définie pour $x = 0$. En effet :

$$\lim_{x \rightarrow 0} \frac{\sin \pi x}{\pi x} = 1$$

Pour s'en rendre compte il suffit de remplacer $\sin(\pi x)$ par son développement limité.

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} = \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}$$

Le sinus cardinal possède la propriété suivante :

$$\int_{-\infty}^{\infty} \text{sin c}(\tau) d\tau = 1$$

g) Impulsion de Dirac

Définition

L'impulsion de Dirac est un signal noté $\delta(t)$, définie par :

$$\delta(0) = +\infty, \quad \delta(t \neq 0) = 0 \quad \text{et} \quad \int_{-\infty}^{+\infty} \delta(t) dt = 1$$

Elle peut être vue comme la limite d'une fonction rectangulaire dont la durée tend vers 0 :

$$\delta(t) = \lim_{T \rightarrow 0} \frac{1}{T} \text{rect}\left[\frac{t}{T}\right]$$

Il ne s'agit pas à proprement parler d'une fonction mais d'une distribution. La théorie des distributions peut être vue comme une généralisation de celle des fonctions. On parle donc souvent de *distribution de Dirac*.

Une impulsion de Dirac centrée sur un instant quelconque t_0 s'écrit sous la forme :

$$\delta(t - t_0)$$

On aura également :

$$\int_{-\infty}^{+\infty} \delta(t - t_0) dt = 1$$

Il s'agit d'un signal très important car, comme nous le verrons plus loin, il permet de connaître la réponse impulsionnelle d'un système, et cette dernière permet de déduire sa réponse à n'importe quel signal.

En général, elle est représentée par une flèche verticale d'amplitude 1, qui désigne plutôt la surface qu'elle définit avec l'axe des abscisses que sa hauteur, qui n'est pas définie. Quand elle est pondérée par un coefficient c ($c \cdot \delta(t)$) appelé poids, elle est représentée par une flèche verticale d'amplitude c .

Propriétés

Multiplié avec un signal et intégré, il permet de connaître la valeur instantanée de ce signal, à l'instant de l'impulsion :

$$\begin{aligned} \int_{-\infty}^{+\infty} s(t) \cdot \delta(t) dt &= s(0) \\ \int_{-\infty}^{+\infty} s(t) \cdot \delta(t) dt &= \int_{-\infty}^{+\infty} s(0) \cdot \delta(t) dt = s(0) \int_{-\infty}^{+\infty} \delta(t) dt = s(0) \end{aligned}$$

Plus généralement, on peut démontrer de la même manière que l'on a :

$$\int_{-\infty}^{+\infty} s(t) \cdot \delta(t - t_0) dt = s(t_0)$$

en utilisant la propriété :

$$s(t) \cdot \delta(t - t_0) = s(t_0) \cdot \delta(t - t_0)$$

On verra que cette propriété fait de l'impulsion un outil mathématique d'étude de l'échantillonnage des signaux, consistant à relever la valeur instantanée d'un signal continu à des instants bien précis.

On peut remarquer que l'on n'accède à $s(t_0)$ pas directement par le produit de $s(t)$ et $\delta(t - t_0)$ mais par l'intermédiaire d'une intégrale. Contrairement à ce qu'on aurait pu penser au premier abord, on a :

$$s(t) \cdot \delta(t) \neq s(0)$$

car l'amplitude de l'impulsion étant infinie, la multiplication de l'impulsion par la valeur du signal en $t=0$ donnerait également l'infini.

Cette impulsion est généralement admise comme la dérivée de la fonction échelon :

$$\delta(t) = \frac{du(t)}{dt}$$

dans la théorie des distributions.

Cette propriété part de la propriété de l'impulsion de Dirac : $\int_{-\infty}^{+\infty} \delta(t) dt = 1$:

$$\int_{-\infty}^t \delta(\tau) d\tau = \begin{cases} 0 & \text{pour } t < 0 \\ 1 & \text{pour } t > 0 \end{cases}$$

Ce qui est la définition de $u(t)$. Donc si l'intégrale de $\delta(t)$ est $u(t)$, la dérivée de $u(t)$ est $\delta(t)$.

Remarque : l'impulsion a été obtenue à partir de la fonction rectangulaire, mais elle peut l'être tout aussi bien à partir d'autres fonctions, pourvu que ces dernières mènent aux mêmes propriétés de l'impulsion. Par exemple, la fonction triangulaire peut également être utilisée :

$$\delta(t) = \lim_{T \rightarrow 0} \frac{1}{T} \text{tri} \left[\frac{t}{T} \right]$$

ou encore la fonction *sinc* :

$$\lim_{T \rightarrow \infty} T \cdot \text{sinc}(Tt) = \delta(t)$$

En effet, $\text{sinc}(x)$ s'annule pour $\pi x = k\pi$ avec k entier, soit $x=k$. $\text{sinc}(Tt)$ s'annule pour $tT=k$, soit $t=k/T$.

Peigne de Dirac

Une suite d'impulsions de Dirac séparées par une durée constante T est définie par :

$$\delta_T(t) = \sum_{k=-\infty}^{+\infty} \delta(t - kT)$$

On appelle cette suite *peigne de Dirac*. On verra que ce signal est utilisé pour modéliser l'échantillonnage des signaux :

$$\sum_{k=-\infty}^{+\infty} x(t) \cdot \delta(t - kT) = \sum_{k=-\infty}^{+\infty} x(kT) \cdot \delta(t - kT)$$

Le résultat de cette opération est une suite de valeurs instantanées du signal $x(t)$, appelés échantillons. Les amplitudes de ces échantillons suivent celles du signal aux instants d'échantillonnage.

Domaine discret

Dans le domaine discret, le signal analogue à l'impulsion de Dirac porte le nom d'*impulsion de Kronecker*, définie par :

$$\delta_n = \begin{cases} 1 & \text{pour } n = 0 \\ 0 & \text{ailleurs} \end{cases}$$

I.3) Caractéristiques des signaux

I.3.1) Période, fréquence

Un signal périodique se répète identique à lui même au cours du temps. Il est défini par sa période T , le plus petit intervalle de temps au bout duquel il se répète. L'inverse de la période est la fréquence, notée par exemple f :

$$f = \frac{1}{T}$$

Un signal sinusoïdal est complètement déterminé par sa fréquence, son amplitude et sa phase.

Plus généralement, il faut bien comprendre que les deux aspects temporel et fréquentiel sont liés : une variation rapide de la grandeur physique correspond à une fréquence élevée.

I.3.2) Valeur moyenne

La valeur moyenne d'un signal $s(t)$ calculée sur l'intervalle $[t_1, t_2]$ est définie par :

$$S_{\text{moy}}(t_1, t_2) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} s(t) dt$$

On peut définir également une valeur moyenne totale :

$$S_{\text{moy}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s(t) dt$$

Dans le cas d'un signal périodique, sa valeur moyenne totale est identique à sa valeur moyenne sur une période. De plus, elle peut être calculée à partir de n'importe quelle origine : l'intervalle d'intégration doit posséder une largeur égale à T . Le début de cet intervalle peut être quelconque, donc on peut écrire indifféremment :

$$S_{\text{moy}}(T) = \frac{1}{T} \int_{t_0}^{t_0+T} s(t) dt = \frac{1}{T} \int_0^T s(t) dt = \frac{1}{T_0} \int_{-T/2}^{T/2} s(t) dt$$

car t_0 est un instant quelconque, et peut donc être pris égal par exemple à 0. Le résultat sera le même dans tous les cas. L'avantage de la 2^e forme est de simplifier les calculs (s'il y en a). Ceci sera valable pour toutes les autres expressions basées sur des intégrales.

La moyenne calculée ci-dessus donne une valeur en un point. Si on applique ce calcul en tout point d'un signal, on obtient ce que l'on appelle une moyenne mobile, qui est elle-même un signal. Soit $x(t)$ un signal quelconque, la définition de la moyenne mobile est la suivante :

$$y(t) = \frac{1}{T} \int_{t-T}^t x(\tau) d\tau$$

Selon la durée sur laquelle est calculée cette moyenne, cette opération réalise un lissage plus ou moins important du signal traité. On verra plus loin que ce moyennage est équivalent à un filtrage passe-bas.

I.3.3) Valeur efficace

Soit $s(t)$ un signal. Sa valeur efficace est définie par :

$$S_{\text{eff}} = \sqrt{\frac{1}{T} \int_{t_0}^{t_0+T} s^2(t) dt}$$

avec t_0 quelconque.

Physiquement, la valeur efficace correspond à la valeur du signal continu qui fournirait la même énergie dans une résistance (dissipation de chaleur par effet Joule).

I.3.4) Energie, puissance

a) Analogie électrique

L'énergie et la puissance d'un signal sont définies par analogie avec ces mêmes grandeurs dans un circuit électrique.

La puissance instantanée fournie par un générateur à un dipôle, traversé par un courant $i(t)$ et dont la tension aux bornes est $u(t)$, est définie par

$$p(t) = u(t) \cdot i(t)$$

L'énergie dissipée sur un intervalle $[t_1, t_2]$ avec $t_2 > t_1$ est l'intégrale de cette puissance instantanée :

$$E_{t_1, t_2} = \int_{t_1}^{t_2} p(t) dt$$

Elle s'exprime en Joules. L'énergie augmente au cours du temps.

La puissance s'obtient en divisant l'énergie par la durée de l'intervalle :

$$P_{t_1, t_2} = \frac{E_{t_1, t_2}}{t_2 - t_1}$$

Elle s'exprime en Watts. La puissance décrit l'énergie dissipée par unité de temps, la distribution de l'énergie dans le temps. Une grande puissance signifie beaucoup d'énergie en peu de temps.

Par exemple, dans le cas où le dipôle est une résistance :

$$p(t) = R \cdot i^2(t) = \frac{u^2(t)}{R}$$

$$E_{t_1, t_2} = R \int_{t_1}^{t_2} i^2(t) dt = \frac{1}{R} \int_{t_1}^{t_2} u^2(t) dt$$

$$P_{t_1, t_2} = \frac{R}{t_2 - t_1} \int_{t_1}^{t_2} i^2(t) dt$$

De la même manière, on définit l'énergie d'un signal $s(t)$ entre deux instants t_1 et t_2 en considérant que ce signal est un courant ou une tension, par celles qui seraient dissipées dans une résistance de 1 ohm :

$$E_{t_1, t_2} = \int_{t_1}^{t_2} s^2(t) dt$$

La puissance moyenne du signal sur l'intervalle $[t_1, t_2]$ est alors définie par :

$$P_{t_1, t_2} = \frac{E_{t_1, t_2}}{t_2 - t_1} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} s^2(t) dt$$

b) Energie totale

On peut également définir l'énergie totale du signal par :

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

Un signal est dit à énergie finie, lorsque son énergie totale est finie :

$$E < \infty$$

On dit également que l'intégrale est bornée, et que le signal est "de carré intégrable (ou sommable)". On verra plus loin que l'ensemble des signaux d'énergie finie est appelé L^2 .

En fait, tous les signaux réels sont à énergie finie, puisqu'on les observe sur une durée finie. Les bornes d'intégrations sont donc forcément finies (par opposition à infinies). Les signaux à énergie infinie sont des abstractions mathématiques, commodes pour les calculs. Par exemple, les signaux théoriques périodiques comme la sinusoïde, l'échelon unité, etc, sont à énergie infinie.

Les signaux permanents présentent une énergie infinie. Les signaux transitoires présentent une énergie finie.

De manière similaire au cas continu, dans le cas d'un signal discret l'énergie est définie par :

$$E = \sum_n s_n^2$$

où les s_n sont les échantillons du signal.

Remarque : si $s(t)$ est complexe, il faut remplacer $s^2(t)$ par $|s(t)|^2$ et s_n^2 par $|s_n|^2$, le carré des modules, dans les expressions. En effet, il faut remplacer $s^2(t)$ par :

$$s(t) \times s^*(t) = (a + jb)(a - jb) = a^2 + b^2 = |s(t)|^2$$

c) Puissance moyenne totale

Dans le cas d'un signal quelconque non limité dans le temps, on calcule sa puissance sur une durée la plus grande possible. On définit ainsi sa puissance moyenne totale :

$$P_{\text{moy}} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} s^2(t) dt$$

Dans le cas d'un signal périodique, sa puissance moyenne totale est la même que celle calculée sur sa période T (et ce quel que soit l'instant d'origine t_0) :

$$P_{\text{moy}} = \frac{1}{T} \int_{t_0}^{t_0+T} s^2(t) dt$$

La puissance instantanée d'un signal $s(t)$ est simplement définie par $s^2(t)$. En pratique, on utilise plutôt la puissance moyenne, car la puissance instantanée d'un signal varie en permanence !

Les signaux permanents présentent une puissance finie et une énergie infinie. Ils ne sont pas physiquement réalisables.

Un signal à énergie finie est à puissance moyenne nulle, car pour une puissance moyenne non-nulle on a :

$$E < \infty \quad \text{implique donc que} \quad P = \lim_{T \rightarrow \infty} \frac{E}{T} = 0$$

Inversement, si la puissance est finie, l'énergie est telle que :

$$E = \lim_{T \rightarrow \infty} T \times P = \infty \times \text{nombre fini} = \infty$$

elle est donc infinie.

Dans le cas d'un signal discret, la puissance moyenne totale est définie par :

$$P = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N |s_n|^2$$

d) Rapport signal/bruit

Un des intérêts de connaître l'énergie d'un signal est de pouvoir calculer le rapport signal/bruit d'un signal dans un système de transmission ou d'amplification, par exemple. Dans ce type d'applications, le signal informatif est entaché de bruit, et on va chercher à minimiser ce dernier pour que son influence soit négligeable par rapport à celle du bruit. Le rapport signal/bruit est défini par :

$$\xi = \frac{E_s}{E_b} = \frac{P_s}{P_b}$$

où E_s est l'énergie du signal et E_b l'énergie du bruit, P_s et P_b les puissances correspondantes.

En général, on s'intéresse à sa valeur en décibels (dB), définie par :

$$\xi_{\text{dB}} = 10 \log_{10} \xi$$

Rappel : la fonction logarithme de base 10 est la fonction inverse de la fonction puissance de base 10 :

$$y = 10^x \Leftrightarrow \log_{10} y = \log 10^x \Leftrightarrow \log_{10} y = x$$

e) Lien entre puissance et valeur efficace

En électricité, on utilise beaucoup la notion de valeur efficace. Elle est égale à la racine carrée de la puissance :

$$S_{\text{eff}} = \sqrt{P_{t_1, t_2}}$$

Remarque : la mise au carré du signal le rend complètement positif, ce qui a pour effet d'augmenter la surface calculée par l'intégrale. Une valeur moyenne nulle n'implique donc pas une valeur efficace nulle.

f) Répartition de la puissance dans un signal

On a vu que la représentation fréquentielle d'un signal quelconque était une fonction continue de la fréquence. La puissance peut être répartie plus dans certaines bandes de fréquences que dans d'autres. On définira la densité spectrale de puissance, qui caractérise cette répartition. On verra qu'elle s'obtient à partir de la fonction d'autocorrélation (dont elle est la transformée de Fourier). Dans le cas d'un bruit blanc, cette densité spectrale de puissance est constante.

g) Mesure

D'un point de vue pratique, en électronique analogique, la puissance peut être obtenue au moyen d'un redressement et d'une multiplication par lui même.

Dans le cas numérique, il s'agit d'une simple addition de valeurs d'échantillons mis au carré.

I.4) Signaux et systèmes

Dans l'expression Traitement du Signal, il y a signal mais également Traitement, ce qui traduit bien le fait que la plupart du temps, un signal est obtenu sur la sortie d'un système (ex. : sortie de filtres, etc).

I.4.1) Linéarité

Soit S un système. Il est linéaire si, à une entrée $a_1 e_1(t) + a_2 e_2(t)$, correspond une sortie $a_1 s_1(t) + a_2 s_2(t)$, $s_1(t)$ et $s_2(t)$ étant les sorties de $e_1(t)$ et $e_2(t)$, respectivement.

$$a_1 e_1(t) + a_2 e_2(t) \rightarrow a_1 s_1(t) + a_2 s_2(t)$$

Exemples de systèmes non-linéaires :

- Quadratureur : $s(t) = e^2(t)$;
- Saturateur : $s(t) = \text{signe}(e(t))$;
- etc

Exemples de systèmes linéaires :

- Filtre moyennneur, réalisant la fonction :

$$y(t) = \frac{1}{T} \int_{t-T}^t x(\tau) d\tau$$

- Sommateur ; $s(t) = e_1(t) + e_2(t)$;
- etc

Remarque : par extension, dans le cas général, a_1 et a_2 sont complexes, ainsi que les signaux d'entrée et de sortie et le système.

I.4.2) Stationnarité

Un système est stationnaire si ses propriétés ne changent pas au cours du temps. On dit également qu'il est invariant dans le temps (SLIT). Un système linéaire et stationnaire est appelé un filtre.

Concrètement, cela signifie qu'un décalage temporel en entrée se traduit par le même décalage en sortie. Soit t_0 ce décalage :

$$e(t-t_0) \rightarrow s(t-t_0)$$

Remarque : les propriétés de linéarité et de stationnarité sont des idéalizations. En général, les systèmes physiques comportent des non-linéarités, et leur comportement peut varier au cours du temps.

I.4.3) Causalité

Dans les systèmes dont la variable est le temps, ce principe énonce simplement que la réponse du système ne précède pas l'entrée, c'est à dire que l'effet ne peut pas précéder la cause. C'est bien sûr le cas de tout système réel, mais en théorie on pourra travailler sur des systèmes non causaux.

Un signal causal est un signal nul pour $t < 0$. Tous les signaux physiques sont causaux (on dit également "réalisables"). En théorie, on peut considérer des signaux non-causaux.

Par contre, dans le cas où la variable est une dimension spatiale (comme par exemple en traitement d'images), cette condition de causalité n'est pas indispensable pour décrire les systèmes réels.

I.4.4) Stabilité

Un système est stable si pour une entrée bornée la sortie est bornée.

Une condition suffisante est que la réponse impulsionnelle soit convergente en valeur absolue (on dit absolument convergente) :

$$\int_{-\infty}^{+\infty} |h(t)| dt < \infty$$

II) Etude des signaux et systèmes dans le domaine temporel

Avant d'étudier l'aspect fréquentiel des signaux et des systèmes, et notamment la transformée de Fourier, un des principaux outils d'étude et de traitement des signaux, nous allons étudier plusieurs de leur aspects temporels, comme la réponse impulsionnelle des systèmes, les propriétés des méthodes de corrélation, et des équations différentielles.

II.1) Réponse impulsionnelle d'un système

II.1.1) Produit de convolution

Dans le domaine continu, le produit de convolution entre 2 signaux $e_1(t)$ et $e_2(t)$ est défini par :

$$s(t) = e_1(t) * e_2(t) = \int_{-\infty}^{+\infty} e_1(\tau) \cdot e_2(t-\tau) \cdot d\tau$$

où «*» est le symbole du produit de convolution, et $s(t)$ est le signal de sortie.

Si l'on effectue le changement de variable $\tau' = t - \tau$, on obtient :

$$s(t) = - \int_{+\infty}^{-\infty} e_1(t - \tau') \cdot e_2(\tau') \cdot d\tau'$$

soit

$$s(t) = \int_{-\infty}^{+\infty} e_1(t - \tau') \cdot e_2(\tau') \cdot d\tau'$$

que l'on peut ré-écrire

$$s(t) = \int_{-\infty}^{+\infty} e_1(t - \tau) \cdot e_2(\tau) \cdot d\tau$$

L'intégrale sur un intervalle infini est théorique. Dans la pratique on n'aura pas un intervalle infini. En pratique, un signal (qui lui pourra être de très longue durée) est convolué avec un filtre qui aura une durée beaucoup plus petite que le signal.

Dans le cas de signaux physiques donc causaux, les conditions sont, pour $e_1(\tau)$: $\tau > 0$, et pour $e_2(t - \tau)$: $t - \tau > 0 \leftrightarrow \tau < t$. Donc finalement, les bornes de l'intégrale deviennent :

$$s(t) = \int_0^t e_1(\tau) \cdot e_2(t - \tau) \cdot d\tau$$

Propriétés

Le produit de convolution est commutatif, associatif et distributif.

Commutativité :

$$x(t) * y(t) = y(t) * x(t)$$

Associativité :

$$[x(t) * y(t)] * z(t) = x(t) * [y(t) * z(t)]$$

Distributivité :

$$[x(t) + y(t)] * z(t) = x(t) * z(t) + y(t) * z(t)$$

Elément neutre : impulsion de Dirac

$$x(t) * \delta(t) = x(t)$$

Translation (ou échantillonnage) :

$$x(t) * \delta(t - t_0) = x(t - t_0)$$

Démonstrations

Elément neutre

$$x(t) * \delta(t) = \int_{-\infty}^{+\infty} x(t - \tau) \cdot \delta(\tau) d\tau = \int_{-\infty}^{+\infty} x(\tau) \cdot \delta(t - \tau) d\tau$$

or $x(t - \tau) \cdot \delta(\tau) = x(t) \cdot \delta(\tau)$ car $x(t - \tau)$ est prélevé à la valeur de l'abscisse $\tau=0$ (car $\delta(\tau)$ est un pic à $t=0$), et comme $x(t)$ est constant par rapport à la variable d'intégration τ , on peut le sortir de l'intégrale :

$$x(t) * \delta(t) = \int_{-\infty}^{+\infty} x(t) \cdot \delta(\tau) d\tau = x(t) \int_{-\infty}^{+\infty} \delta(\tau) d\tau = x(t)$$

Translation

$$x(t) * \delta(t - t_0) = \int_{-\infty}^{+\infty} x(t - t_0 - \tau) \cdot \delta(\tau) d\tau = \int_{-\infty}^{+\infty} x(\tau) \cdot \delta(t - t_0 - \tau) d\tau$$

De même que précédemment, on a $x(t - \tau - t_0) \cdot \delta(\tau) = x(t - t_0) \cdot \delta(\tau)$, donc :

$$x(t) * \delta(t - t_0) = \int_{-\infty}^{+\infty} x(t - t_0) \delta(\tau) d\tau = x(t - t_0) \int_{-\infty}^{+\infty} \delta(\tau) d\tau = x(t - t_0)$$

Interprétation

La propriété d'associativité peut être vue comme la mise en cascade (également appelée "en série") de 2 systèmes. Soient 2 systèmes de réponses impulsionnelles $h_1(t)$ et $h_2(t)$. La sortie du premier système est définie par :

$$y_1(t) = x(t) * h_1(t)$$

La sortie du 2^e système est définie par :

$$y_2(t) = y_1(t) * h_2(t) = [x(t) * h_1(t)] * h_2(t)$$

Par la propriété d'associativité on a :

$$y_2(t) = x(t) * [h_1(t) * h_2(t)]$$

La mise en cascade est donc équivalente à un seul système de réponse impulsionnelle $h_1(t) * h_2(t)$. De plus, la propriété de commutativité nous indique que l'ordre des 2 systèmes n'a pas d'importance.

De même, la propriété de distributivité correspond à la mise en parallèle de systèmes, dont les sorties sont additionnées par un sommateur. Dans le cas de deux systèmes de réponses impulsionnelles $h_1(t)$ et $h_2(t)$:

$$y(t) = x(t) * h_1(t) + x(t) * h_2(t) = x(t) * [h_1(t) + h_2(t)]$$

Cette mise en parallèle est donc équivalente à un seul système de réponse impulsionnelle $h_1(t) + h_2(t)$.

Convolution bidimensionnelle

La convolution peut être bi-dimensionnelle. Soient $e_1(x,y)$ et $e_2(x,y)$ deux signaux dépendant de deux variables x et y :

$$e_1(x,y) * e_2(x,y) = \int_{\psi=-\infty}^{+\infty} \int_{\xi=-\infty}^{+\infty} e_1(\xi-x, \psi-y) \cdot e_2(\xi, \psi) d\xi d\psi$$

Dans les applications de traitement d'images, les 2 dimensions sont les 2 dimensions spatiales de l'image et sont donc discrétisées. Le filtrage spatial par convolution est très fréquent dans le domaine du traitement d'images. Elle permet de réaliser des opérations de lissage, ré-haussement des contrastes en vue d'extraire des contours, etc. L'un des deux signaux correspond à l'image et l'autre à des "masques de convolution", permettant d'effectuer divers traitements comme des lissages, des moyennages locaux, des réhaussements de contrastes, etc (voir plus bas : "Convolution discrète").

Remarque : comme dans la convolution temporelle, on peut intervertir e_1 et e_2 .

II.1.2) Convolution avec réponse impulsionnelle

Comme son nom l'indique, la réponse impulsionnelle du système est sa réponse à une impulsion. La réponse impulsionnelle est très importante, car elle permet de connaître la réponse du système à n'importe quelle autre entrée.

D'un point de vue un peu plus théorique, elle permet de savoir si un système est causal : une condition suffisante pour cela est que cette réponse soit nulle pour $t < 0$. Elle est souvent notée $h(t)$.

Elle permet également de savoir si un système est stable : un système est stable si sa réponse impulsionnelle est finie :

$$\int_{-\infty}^{+\infty} |h(t)| dt < \infty$$

On utilise parfois également l'expression "réponse percuSSIONnelle", qui a l'avantage d'être plus imagée : par exemple pour connaître l'acoustique d'une salle, c'est à dire en fait le traitement de réverbération appliqué sur les sons, on utilisera plutôt un claquement sec qu'un son prolongé ; dans ce dernier cas en effet, on ne pourra plus distinguer les réverbérations dues au début du son de ceux de la fin car tout est mélangé.

a) Expression de la relation entrée-sortie pour une entrée quelconque

Si l'on remplace l'un des 2 signaux par l'impulsion de Dirac $\delta(t)$, le résultat correspond à la réponse impulsionnelle du système, notée $h(t)$:

$$S(\delta(t)) = (\delta * e)(t) = \int_{-\infty}^{+\infty} \delta(t - \tau) e(\tau) d\tau = \int_{-\infty}^{+\infty} e(\tau) \delta(t - \tau) d\tau = h(t)$$

A partir de cette définition, on peut démontrer que la réponse du système à un signal d'entrée $e(t)$ quelconque est un produit de convolution entre le signal d'entrée $e(t)$ et sa réponse impulsionnelle $h(t)$:

$$s(t) = (e * h)(t) = \int_{-\infty}^{+\infty} h(t - \tau) e(\tau) d\tau = \int_{-\infty}^{+\infty} h(\tau) e(t - \tau) d\tau$$

Tout système linéaire et stationnaire (ou filtre) est un système de convolution.

b) Démonstration

La dernière relation peut être démontrée à partir de la définition du produit de convolution :

$\delta(t)$	$\rightarrow$	$h(t)$	
$\delta(t - \tau)$	$\rightarrow$	$h(t - \tau)$	(stationnarité)
$e(\tau) \cdot \delta(t - \tau)$	$\rightarrow$	$e(\tau) \cdot h(t - \tau)$	(linéarité, $e(\tau)$ est constant % t)
$\int_{-\infty}^{+\infty} e(\tau) \cdot \delta(t - \tau) d\tau = e(t)$	$\rightarrow$	$\int_{-\infty}^{+\infty} e(\tau) \cdot h(t - \tau) d\tau = s(t)$	(linéarité, par généralisation de l'additivité : somme vers intégrale)

L'expression en bas à gauche correspond à $e(t)$, car on peut décomposer le signal d'entrée en une suite d'impulsions de largeur Δt positionnées aux instants $0, \Delta t, 2\Delta t, \dots, i\Delta t, \dots$, chacune ayant pour amplitude la valeur du signal à ces instants. On peut définir ces impulsions à l'aide de la fonction porte, et pour que l'amplitude de l'impulsion soit celle du signal, il faut la multiplier par Δt , sinon cette amplitude tendrait vers l'infini quand $\Delta t \rightarrow 0$:

$$e(i\Delta t) \cdot \Pi_{\Delta t}(t - i\Delta t) \Delta t$$

Du fait de la linéarité du système, le signal d'entrée décomposé sous forme d'impulsions peut donc s'exprimer sous la forme :

$$e_{\Delta t}(t) = \sum_{i=-\infty}^{+\infty} e(i\Delta t) \cdot \Pi_{\Delta t}(t - i\Delta t) \Delta t$$

Le signal peut être considéré comme le passage à la limite de ces suites d'impulsions :

$$e(t) = \lim_{\Delta t \rightarrow 0} \{e_{\Delta t}(t)\}$$

car

$$i\Delta t \rightarrow \tau \quad \text{et} \quad \Delta t \rightarrow d\tau$$

et la fonction porte tend alors vers l'impulsion de Dirac :

$$\delta(t) = \lim_{\Delta t \rightarrow 0} \Pi_{\Delta t}(t - i\Delta t)$$

Donc on obtient bien l'expression de la réponse du système à n'importe quel signal à partir de sa réponse impulsionnelle.

c) Systèmes physiques

Comme on l'a vu dans le paragraphe précédent (sur le produit de convolution), dans le cas de signaux réels, les bornes de l'intégrale se réduisent à $[0, t]$:

$$s(t) = (e * h)(t) = \int_0^t h(t - \tau) e(\tau) d\tau$$

Si la réponse impulsionnelle possède une durée finie d , les bornes d'intégration deviennent :

$$s(t) = (e * h)(t) = \int_{t-d}^t h(t-\tau)e(\tau)d\tau$$

Interprétation

L'exemple le plus simple de convolution est celui où la fonction $h(t)$ est une fonction rectangle définie par :

$$h(t) = \frac{1}{T} \text{rect}\left[\frac{t-t_0}{T}\right]$$

Remarque : on trouve parfois la notation

$$h(t) = \frac{1}{T} \text{rect}_T(t-t_0)$$

$h(t)$ est centré sur t_0 ; on prend $t_0 = -T/2$.

$h(t)$ n'est pas forcément la réponse impulsionnelle d'un système, mais permet de réaliser un filtrage correspondant à un calcul de moyenne mobile. En effet, si l'on utilise la définition de la convolution :

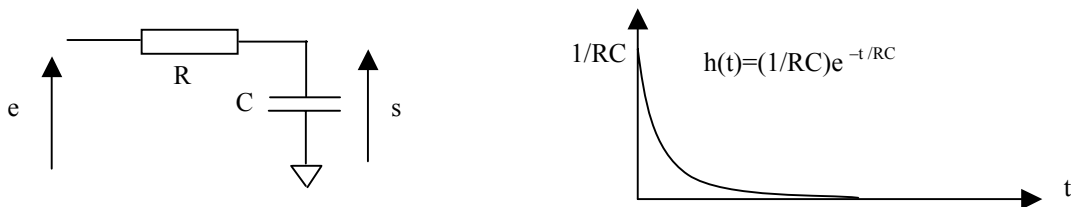
$$s(t) = \int_{-\infty}^{+\infty} h(t-\tau)e(\tau)d\tau = \frac{1}{T} \int_{-\infty}^{+\infty} \text{rect}\left[\frac{\tau-t-T/2}{T}\right]e(\tau)d\tau = \frac{1}{T} \int_{t-T}^t e(\tau)d\tau$$

qui est la définition de la moyenne mobile (voir plus haut).

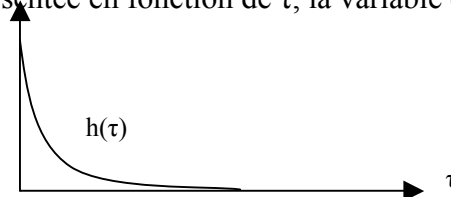
L'exemple de réponse impulsionnelle ci-dessous va nous permettre d'illustrer la réponse du système à une entrée quelconque à partir de sa réponse impulsionnelle.

d) Exemple : circuit RC

Démonstration graphique



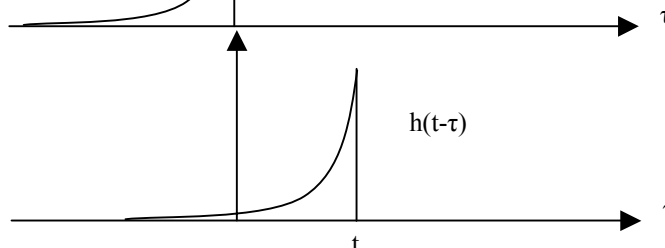
La réponse impulsionnelle est représentée en fonction de τ , la variable d'intégration :



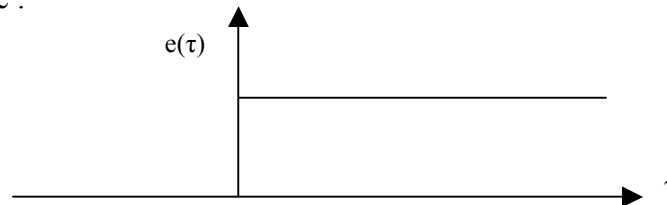
elle est retournée :



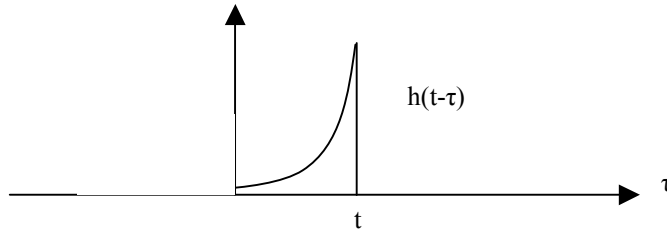
puis décalée :



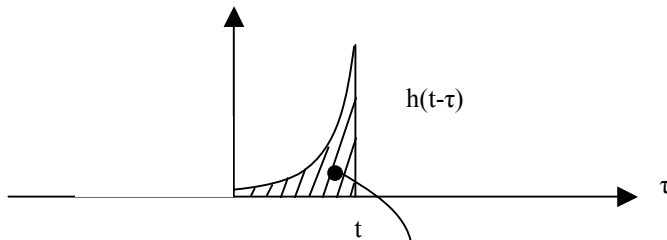
On définit le signal d'entrée :



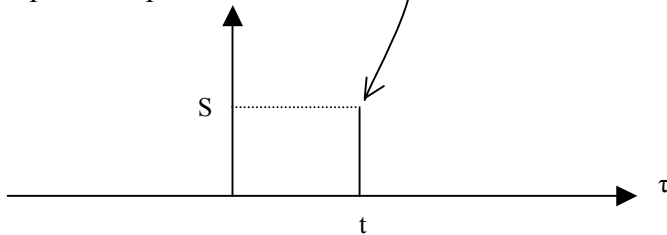
On effectue la multiplication entre les 2 :



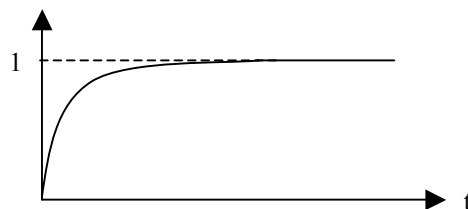
puis l'intégration :



ce qui donne un point de la réponse impulsionnelle :



En appliquant cette procédure pour toutes les valeurs de $t > 0$, on obtient la réponse du système à un échelon, appelée réponse indicielle :



Démonstration mathématique

La réponse impulsionnelle est définie par :

$$h(t) = \frac{1}{RC} \exp\left(-\frac{t}{RC}\right), t \geq 0 \text{ (0 pour } t < 0)$$

Remarque : pour démontrer cette réponse impulsionnelle, il faut utiliser l'équation différentielle du système, et la transformée de Laplace. Cette équation est définie par :

$$RC \frac{ds(t)}{dt} + s(t) = e(t)$$

$e(t)$ est remplacée par $\delta(t)$ et on utilise ensuite la transformée de Laplace. Cette étude sera détaillée plus loin.

Connaissant la réponse impulsionnelle, on peut calculer la réponse indicielle :

$$S(u(t)) = \int_{-\infty}^{\infty} u(\tau) h(t - \tau) d\tau$$

$$\begin{aligned}
&= \frac{1}{RC} \int_0^t e^{-(t-\tau)/RC} d\tau \\
&= \frac{1}{RC} \int_0^t e^{-t/RC} e^{\tau/RC} d\tau \\
&= \frac{e^{-t/RC}}{RC} \int_0^t e^{\tau/RC} d\tau \\
&= e^{-t/RC} \left[e^{\tau/RC} \right]_0^t \\
&= e^{-t/RC} \left[e^{t/RC} - 1 \right] \\
&= 1 - e^{-t/RC}
\end{aligned}$$

Il s'agit de l'expression de la courbe obtenue par démonstration graphique.

c) Domaine discret

Dans le domaine discret, l'opération de convolution est définie de manière analogue au cas continu, entre deux signaux ou un signal et une réponse impulsionnelle :

$$s_k = e_k * h_k = \sum_{i=-\infty}^{+\infty} e_i \cdot h_{k-i} = \sum_{i=-\infty}^{+\infty} e_{k-i} \cdot h_i \text{ avec } k = -\infty, \dots, +\infty$$

où e_k est une séquence d'entrée (constituée par exemple par les échantillons d'un signal continu $e(t)$), h_k une séquence représentant la réponse impulsionnelle d'un système numérique, ou plus généralement un filtre (on parle également de masque de convolution) et s_k la séquence résultante. En général, $M \gg N$, c'est à dire que h est un signal court et e un signal long. Le signal court est souvent appelé filtre, car la principale application de la convolution est le filtrage. Il s'agit ici d'un filtrage temporel (ou spatial dans le cas des images).

Les séquences d'entrée et de sortie seront notées dans la suite respectivement $\{e_n\}$ et $\{s_n\}$.

La convolution numérique avec une réponse impulsionnelle est à comparer à son équivalent dans le domaine continu :

$$s(t) = \int_0^t e(\tau) h(t-\tau) d\tau$$

Dans le cas de signaux physiques, le signal est causal, donc nul pour $i < 0$. De même, la réponse impulsionnelle h_{k-i} est nulle pour $k-i < 0$ car le système est également causal, soit $i > k$. L'indice de la somme doit donc varier de 0 à k :

$$s_k = e_k * h_k = \sum_{i=0}^k e_{k-i} \cdot h_i \text{ avec } k = 0, \dots, +\infty$$

En général, la réponse impulsionnelle h_k est de longueur finie N ; la borne supérieure est donc limitée à $N-1$. Pour $k-i < 0$, e_{k-i} est considéré comme nul, car le signal est causal :

$$s_k = e_k * h_k = \sum_{i=0}^{N-1} e_{k-i} \cdot h_i \text{ avec } k = 0, \dots, +\infty$$

En pratique, si la longueur de la séquence e est N_e et celle de la séquence h N_h , la longueur de la séquence résultat s est $N_e + N_h - 1$, ce qui équivaut à prolonger la séquence $\{e_n\}$ de ce nombre d'éléments, considérés comme nuls. Pour obtenir le premier élément de la séquence s , le chevauchement entre les deux séquences e et h est limitée à un seul élément. Les éléments manquant de la séquence e sont remplacés par des 0. On appelle le cas particulier du début de la séquence e "effet de bord".

Ce qui donne, si la longueur de la séquence e est M :

$$s_k = e_k * h_k = \sum_{i=0}^{N-1} e_{k-i} \cdot h_i \text{ avec } k = 0, 1, \dots, M + N - 1$$

Formalisation

L'équivalent discret de l'impulsion de Dirac est l'impulsion de Kronecker : une séquence, notée par exemple $\{\delta_n\}$, dont le premier terme est à 1 et tous les autres à 0. On notera une impulsion décalée de k éléments par l'expression $\{\delta_{n-k}\}$. La réponse du système à cette impulsion est appelée "réponse impulsionnelle", tout comme dans le domaine continu : la réponse à $\{\delta_n\}$ sera notée $\{h_n\}$; la réponse à $\{\delta_{n-k}\}$ sera donc $\{h_{n-k}\}$ car le système est considéré comme invariant.

La séquence d'entrée $\{e_n\}$ peut être décomposée sous la forme d'une somme d'impulsions décalées, de valeurs égales à celles de $\{e_n\}$ aux décalages correspondants :

$$\{e_n\} = \sum_{k=-\infty}^{+\infty} e_k \{\delta_{n-k}\}$$

La réponse du système à l'un des éléments $e_k \{\delta_{n-k}\}$ de la somme précédente est :

$$\{s_n\}_k = e_k \{h_{n-k}\}$$

par linéarité, et la réponse à la séquence d'entrée $\{e_n\}$, également par linéarité :

$$\{s_n\} = \sum_{k=-\infty}^{+\infty} \{s_n\}_k = \sum_{k=-\infty}^{+\infty} e_k \{h_{n-k}\}$$

Calcul rapide

On peut visualiser le processus de convolution sous la forme d'un balayage du signal par le filtre. On fait « glisser » le filtre du début à la fin du signal (constitué par un tableau de valeurs), et pour chaque échantillon de ce signal, on calcule le produit de convolution, les termes ajoutés étant les éléments en vis-à-vis.

0	0	0	1	1	1	1				
-1	1									

$$s(0) = h(0).e(0) + h(1).e(-1) = 0$$

$$s(1) = h(0).e(1) + h(1).e(0) = 0$$

etc.

Programmation Scilab

Voyons maintenant la programmation de cet exemple simple avec Scilab. On définit un filtre a et un signal b (le "-->" est le prompt de Scilab) :

```
-->a = [1 2 3]
a =
! 1. 2. 3. !
-->b = [1 0 0 0 0 1 0 0 0]
b =
! 1. 0. 0. 0. 0. 1. 0. 0. 0. !
-->c = convol(a,b)
c =
! 1. 2. 3. 0. 0. 1. 2. 3. 0. 0. 0. !
```

Ce résultat particulier est dû au fait que les '1' du signal soient suivis de '0'. On retrouve 2 fois la séquence a , qui représente la réponse impulsionnelle du système. Dans le cas plus général du signal :

```
-->a = [0 0 0 1 1 1 1 0 0 0]
a =
! 0. 0. 0. 1. 1. 1. 1. 0. 0. 0. !
```

et du filtre

```
-->b = [1 -1]
b =
! 1. -1. !
-->c = convol(a,b)
```

```

c =
!  0.    0.    0.    1.    0.    0.    0.   -1.    0.    0.    0. !

```

Ce résultat est très intéressant, et la notion de filtrage apparaît déjà : il n'y a des '1' qu'aux endroits des transitions du signal d'origine. La convolution agit comme un détecteur de transitions rapides, c'est à dire de hautes fréquences. On a donc réalisé un filtre passe-haut élémentaire. A la sortie du filtre le signal est devenu nul là où il n'y avait pas de transitions dans le signal d'origine.

D'une manière plus générale, selon la taille du filtre et son contenu, on peut effectuer différents types de filtrages (passe-bas, passe-bande, passe-haut...) avec différentes fréquences de coupure (=limite des fréquences éliminées). L'exemple suivant correspond à un filtrage moyenneur, qui est du type passe-bas :

```

-->a = [0 0 0 1 1 1 1 0 0 0];
-->b = [1/3 , 1/3, 1/3];
-->c = convol(a,b);

```

```

-->a'
ans =
!  0. !
!  0. !
!  0. !
!  1. !
!  1. !
!  1. !
!  1. !
!  1. !
!  0. !
!  0. !
!  0. !

```

```

-->c'
ans =
!  0.      !
!  0.      !
!  0.      !
!  0.333333 !
!  0.666667 !
!  1.      !
!  1.      !
!  0.666667 !
!  0.333333 !
!  0.      !
!  0.      !
!  0.      !

```

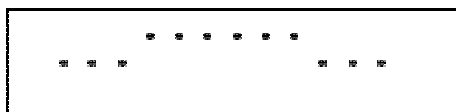
(la mise sous forme de colonne est uniquement pour faciliter l'affichage)

On voit que les transitions rapides du signal ont été "adoucies" : on peut alors dire que les hautes fréquences ont été éliminées ; les basses fréquences sont conservées, on peut s'en convaincre en constatant que la suite de 1, qui correspond à une fréquence très basse puisqu'il n'y a pas de variations, est conservée. Le code suivant permet de faire un affichage graphique :

```

-->x = [0 0 0 1 1 1 1 1 0 0 0];
-->xsetech([0,0,1,1/3],[-1,-2,15,2])
-->plot2d(1:length(x), x, -3, "002");
-->xsetech([0,1/3,1,1/3],[-1,-2,15,2])
-->x1 = convol([1, -1], x);
-->plot2d(1:length(x1), x1, -3, "002");
-->xsetech([0,2/3,1,1/3],[-1,-2,15,2])
-->x2 = convol([1/3, 1/3, 1/3], x);
-->plot2d(1:length(x2), x2, -3, "002");

```



Le résultat obtenu est :

Convolution à deux dimensions

La convolution est très utilisée dans le domaine du traitement d'images. Elle permet de révéler ou d'extraire des informations pertinentes des images, par exemple un rehaussement des contrastes, première étape en vue de la détection des contours des objets ; elle permet notamment de détecter l'orientation de ces contours. Le domaine des images est discret, puisque les images sont constituées de points : les pixels.

La convolution bi-dimensionnelle dans le domaine continu a déjà été définie plus haut. Dans le domaine discret, son expression devient :

$$s(i, j) = \sum_{m=-M}^M \sum_{n=-N}^{+N} e_1(m, n) \cdot e_2(i - m, j - n) = e_1(i, j) * e_2(i, j)$$

où e_1 et e_2 seraient 2 images de taille M lignes x N colonnes. Les images sont souvent carrées.

En général, l'une des 2 images est un filtre de taille petite et fixe (et également souvent carré) par rapport à l'autre, qui est l'image à filtrer. L'image résultant du traitement s'écrirait alors :

$$s(i, j) = \sum_{k=-M/2}^{M/2-1} \sum_{l=-M/2}^{M/2-1} h(k + M/2, l + M/2) \cdot e(i + k, j + l), \quad i, j = 0, \dots, N-1, \quad k \text{ entier}$$

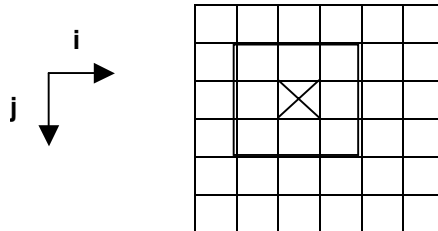
où N est la taille de l'image, et M est la taille du filtre, toutes les deux en nombre de pixels par côté (on suppose les images carrées).

De manière équivalente, on peut écrire :

$$s(i, j) = \sum_{k=0}^{M-1} \sum_{l=0}^{M-1} h(k, l) \cdot e(i + k - M/2, j + l - M/2)$$

La signification de l'application de cette expression est un balayage de l'image, pixel par pixel, par exemple de haut en bas et de droite à gauche. Pour chaque pixel de l'image e , on calcule la double somme de produits de l'expression. Chaque résultat est un pixel de l'image résultant du traitement.

Exemple : $M=3$, calcul de $s(3,3)$.



Gestion des effets de bord

Pour le calcul des pixels situés aux bords des images, le filtre sort de l'image. Il se pose alors le problème du choix de la valeur de ces pixels.

Comme dans le cas d'un signal, on peut les prendre nuls, mais cela va amener à détecter les contours de l'image.

On peut également considérer des valeurs miroir des pixels.

Souvent, on ne calcule que la partie de l'image pour laquelle le filtre se trouve complètement dans l'image, ce qui a pour inconvénient de donner une image plus petite que l'originale. Cette différence de taille est d'autant plus grande que le filtre est grand relativement à l'image.

Exemple

Soit l'image suivante (les cases vides contiennent des 0) :

			1	1	1				
			1	1	1				
			1	1	1				
			1	1	1				

et le filtre (masque de Prewitt) :

1	0	-1
1	0	-1
1	0	-1

Effectuons le calcul explicite pour un pixel, par exemple le pixel (2,2) :

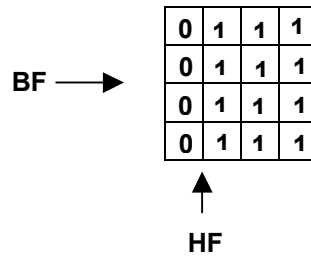
$$\begin{aligned}
 s(2,2) &= \sum_{k=-1}^1 \sum_{l=-1}^1 h(k+1, l+1) \cdot e(2+k, 2+l) \\
 &= h(0,0) \cdot e(1,1) + h(1,0) \cdot e(2,1) + h(2,0) \cdot e(3,1) \\
 &\quad + h(0,1) \cdot e(1,1) + h(1,1) \cdot e(2,2) + h(2,1) \cdot e(3,2) \\
 &\quad + h(0,2) \cdot e(1,2) + h(1,2) \cdot e(2,3) + h(2,3) \cdot e(3,3) \\
 &= -2
 \end{aligned}$$

Effectuons maintenant le filtrage pour toute l'image :

?	?	?							
?	0	-1	0	1					
?	0	-2	0	2					
	0	-3	0	3					
	0	-3	0	3					

Le filtrage est de type passe-haut, mais uniquement dans la direction verticale.

Il peut être appliqué à l'extraction des contrastes, à la base de la détection des contours des objets. Par exemple, un seuillage de l'image traitée permet de détecter les contours. Ces contrastes correspondent à des fréquences spatiales : dans sa dimension horizontale, cette image comporte une haute fréquence (HF) spatiale, alors que dans sa dimension verticale, elle ne comporte qu'une fréquence nulle (BF) puisqu'il n'y a pas de variations d'intensités de pixels :



Pour extraire des contours quelle que soit leur orientation, il existe le filtre Laplacien :

1	1	1
1	-8	1
1	1	1

La somme des éléments doit être égale à 0.

Pour réaliser un filtre moyenneur, on peut par exemple utiliser le filtre suivant :

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Ce filtrage a pour effet de rendre une image floue.

On peut également utiliser des filtres plus grands, pour détecter des variations d'intensité plus douces que celles détectées par le filtre 3x3, qui est le plus petit filtre possible.

II.2) Corrélation

Introduction

Dans le domaine continu, la corrélation est définie par une intégrale, et malgré sa ressemblance avec l'intégrale de convolution, on verra que sa signification et son utilité est très différente : comme il a été vu précédemment, la convolution avec la réponse impulsionnel d'un système permet de décrire complètement celui-ci, alors que la corrélation constitue un outils possédants de nombreuses applications : comparaison de signaux, recherche de motifs dans des signaux, mesure de retards, etc.

II.2.1) Définitions

Dans le domaine du traitement du signal, on parle d'autocorrélation ou d'intercorrélation selon que les 2 grandeurs comparées sont 2 parties différentes d'un même signal, ou 2 signaux différents, respectivement.

Dans le domaine **continu** (ou analogique), la fonction de corrélation entre 2 signaux $x(t)$ et $y(t)$ est définie par :

$$C_{xy}(\tau) = \int_{t=-\infty}^{+\infty} x(t).y(t-\tau)dt$$

Il s'agit d'une fonction de la variable τ , qui représente un décalage temporel. On parle également d'intercorrélation.

De la même manière, on définit la fonction d'autocorrélation d'un signal $x(t)$ par :

$$C_{xx}(\tau) = \int_{t=-\infty}^{+\infty} x(t).x(t-\tau)dt$$

Remarques :

- On peut rencontrer $y(t+\tau)$ au lieu de $y(t-\tau)$, ce qui ne change pas grand chose car par changement de variable $\tau \rightarrow -\tau$, on se rend compte que les 2 expressions sont équivalentes. Elles correspondent à une intégration dans un sens ou dans l'autre, mais le résultat est le même.

- On ne considérera que le cas de signaux réels. Dans le cas de signaux complexes, la définition serait :

$$C_{xx}(\tau) = \int_{t=-\infty}^{+\infty} x^*(t) \cdot y(t-\tau) dt$$

où $x^*(t)$ représente le signal conjugué du signal complexe $x(t)$.

- Dans le cas d'un signal de période T , l'intégration se limite à un intervalle de durée T :

$$C_{xy}(\tau) = \frac{1}{T} \int_{t=-T/2}^{+T/2} x(t) \cdot y(t-\tau) dt$$

La borne inférieure peut être quelconque, pourvu que l'intervalle soit de durée T , et notamment $[0, T]$.

- Théoriquement, la définition donnée ci-dessus est valable uniquement pour les signaux transitoires. La définition exacte pour les signaux permanents (à énergie infinie) est :

$$C_{xy}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t=-T/2}^{+T/2} x(t) \cdot y(t-\tau) dt$$

Opérons le changement de variable suivant : $t' = t - \tau$. On a : $t = t' + \tau$ et $dt' = dt$ (car $\frac{dt'}{dt} = \frac{dt - d\tau}{dt} = \frac{dt}{dt} - \frac{d\tau}{dt} = 1$)

$$C_{xy}(\tau) = \int_{t'=-\infty}^{+\infty} x(t'+\tau) \cdot y(t') dt'$$

que l'on peut alors écrire :

$$C_{xy}(\tau) = \int_{t=-\infty}^{+\infty} x(t+\tau) \cdot y(t) dt$$

Contrairement au produit de convolution, le produit de corrélation n'est donc pas commutatif :

$$\text{corr}(a,b) \neq \text{corr}(b,a)$$

alors que

$$\text{conv}(a,b) = \text{conv}(b,a)$$

Cette formule est théorique, car en pratique on n'utilisera jamais l'intégration sur un intervalle infini.

II.2.2) Début d'interprétation

a) Lien avec produit scalaire

Pour une valeur donnée de la variable τ , on effectue le produit entre 2 signaux "point par point", sur un intervalle donné. Dans le cas de l'autocorrélation, ce produit porte sur le même signal. $C_{xy}(\tau)$ est l'expression du produit scalaire dans le cas continu entre les 2 signaux considérés sur leur intervalle d'intégration. Il en est de même dans le cas discret.

b) Comparaison avec convolution

Pour rappel, la définition de la fonction de convolution est :

$$s(t) = e(t) * h(t) = \int_{-\infty}^{+\infty} h(t-\tau) e(\tau) d\tau = \int_{-\infty}^{+\infty} h(\tau) e(t-\tau) d\tau$$

Contrairement à l'utilisation du produit de convolution pour calculer la réponse d'un système à un signal quelconque, ici on garde la variable du signal (t) pour l'intégration. En effet, la fonction de corrélation ne doit pas être interprétée comme une fonction de t ou un signal, mais de τ , qui représente un décalage temporel, variable, d'un signal par rapport à l'autre (ou du signal par rapport à lui-même dans le cas de l'autocorrélation). De plus, on ne retourne pas l'un des 2 signaux, mais on le décale simplement.

Dans le cas de la convolution, l'un des 2 signaux est constitué par la réponse impulsionnelle du système. Dans le cas de la corrélation, l'un des 2 est constitué par un motif que l'on recherche.

Comme dans le cas de la convolution, on effectue un balayage de l'image avec un motif.

On peut constater que cette expression est très proche de celle de la convolution, étudiée précédemment dans le cadre du filtrage. Pour rappel, dans le domaine continu cette expression est :

$$s(t) = e(t) * h(t) = \int_{-\infty}^{+\infty} e(\sigma).h(t - \sigma).d\sigma$$

où $e(t)$ est le signal d'entrée, $h(t)$ un autre signal appelé filtre, et $s(t)$ le signal de sortie. $h(t)$ est également la réponse impulsionnelle du système dont l'entrée est $e(t)$ et la sortie $s(t)$; dans le domaine discret :

$$s_k = \sum_{i=0}^{N-1} e_{k-i}.h_i \quad \text{avec } k=0,1,\dots,M$$

où e_n sont les échantillons du signal d'entrée, h_n ceux du filtre, et s_n ceux du signal de sortie.

Si l'on compare cette opération à celle de convolution, on remarque que dans cette dernière, les signaux multipliés entre eux étaient parcourus en sens inverse, alors qu'ici ils sont parcourus dans le même sens.

c) Mesure de ressemblance

Dans le cas d'un signal quelconque, la fonction d'autocorrélation sera maximale quand, sur l'intervalle sur lequel est calculé le produit, la ressemblance entre les signaux sera maximale.

La corrélation possède donc des propriétés pouvant en faire un outil de comparaison entre signaux ou de comparaison de différentes parties d'un même signal. Elle permet notamment de rechercher des motifs dans des signaux. Par exemple, dans le cas des images, elle peut permettre de rechercher des formes comme des caractères ou des visages, précédant une étape de reconnaissance. De plus, elle pourra permettre de déterminer si un signal est périodique ou non.

II.2.3) Quelques propriétés

- L'autocorrélation d'un signal périodique et un signal périodique de même fréquence (par exemple, l'autocorrélation d'un signal carré donne un signal triangulaire).
- L'autocorrélation d'un bruit blanc est nulle partout sauf à l'origine (=Dirac). A l'origine il est égal à sa variance.
- Lien avec l'énergie : le calcul de la fonction d'autocorrélation pour un décalage $\tau=0$ donne

$$C_{xx}(0) = \int_{t=-\infty}^{+\infty} x(t).x(t-0)dt = \int_{t=-\infty}^{+\infty} x^2(t)dt = E$$

qui est l'expression de l'énergie du signal.

- La fonction d'autocorrélation est symétrique par rapport à l'axe des ordonnées.

Démonstration

$$C_{xx}(\tau) = \int_{t=-\infty}^{+\infty} x(t).x(t-\tau)dt$$

Une fonction symétrique par rapport à l'axe des ordonnées est paire ; elle possède la propriété $f(-x)=f(x)$. On effectue donc le changement de variable $\sigma' = -\sigma$, on a alors :

$$C_{xx}(\tau) = \int_{t=-\infty}^{+\infty} x(t).x(t+\tau)dt$$

On retrouve une autre expression du produit de la fonction de corrélation, telle que décrite plus haut (c'est à dire qu'on peut retrouver la forme de départ par changement de variable $t'=t+\tau$).

Par contre, on peut voir que la fonction d'intercorrélation n'est pas paire, car la forme obtenue par changement de variable $\sigma' = -\sigma$ est différente de celle obtenue par l'autre changement de variable $t=t'+\tau$, respectivement :

$$C_{xy}(-\tau) = \int_{t=-\infty}^{+\infty} x(t).y(t+\tau)dt \neq \int_{t=-\infty}^{+\infty} x(t+\tau).y(t)dt = C_{xy}(\tau)$$

II.2.4) Cas discret

a) Définitions

Dans le domaine **discret**, l'opération de convolution est définie par :

$$C_{xy}(k) = \sum_{i=-\infty}^{\infty} x_i \cdot y_{i-k} \quad \text{avec } k=-\infty, \dots, \infty \text{ entier}$$

Elle est à comparer avec son équivalent dans le cas continu :

$$C_{xy}(\tau) = \int_{t=-\infty}^{+\infty} x(t) \cdot y(t-\tau) dt$$

De même pour l'autocorrélation :

$$C_{xx}(k) = \sum_{i=-\infty}^{\infty} x_i \cdot x_{i-k} \quad \text{avec } k=-\infty, \dots, \infty \text{ entier}$$

Comme dans le cas continu, on peut changer la variable de la sommation :

$$i' = i - k \leftrightarrow i = i' + k$$

$$C_{xy}(k) = \sum_{i=-\infty}^{\infty} x_{i+k} \cdot y_i \quad \text{avec } k=-\infty, \dots, \infty \text{ entier}$$

En pratique, la sommation s'effectue toujours sur un nombre fini de points :

$$C_{xy}(k) = \sum_{i=0}^{N-1} x_i \cdot y_{i-k} \quad \text{et} \quad C_{xx}(k) = \sum_{i=0}^{N-1} x_i \cdot x_{i-k}$$

La plage de variation de k dépend de la stratégie utilisée pour gérer les effets de bord. Ces derniers concernent les cas où le signal décalé possède une partie qui n'est plus en vis-à-vis par rapport au signal fixe :

- si l'on ne prend pas en compte ces cas, le nombre d'éléments de la fonction de corrélation est $N-M+1$, où N est le nombre d'éléments du signal de départ et M celle du signal décalé, ce qui signifie que le signal résultant est plus court que le signal de départ ; cette stratégie n'est utilisée que lorsque l'un des 2 signaux est beaucoup plus court que l'autre (par exemple dans la recherche de motif), sinon la fonction d'autocorrélation se limiterait à un seul élément ;
- si on les prend en compte, le signal résultant est plus long que le signal de départ : sa taille est $N+M-1$. Il faut alors spécifier les position relatives des 2 signaux, par rapport à l'autre. Si ces 2 signaux sont cadrés sur la gauche, les valeurs de décalage seront alors comprises entre $-(M+1)$ et $N+M-1$. La fonction `corr` de Scilab impose que les 2 signaux possèdent le même nombre d'éléments. Si l'on souhaite un signal plus court que l'autre, il faut donc combler avec des 0. Le décalage ne s'effectue que vers la gauche.

Voir "Gestion des effets de bord", plus bas.

Dans ce cas, le signal de sortie comporte le même nombre d'échantillons que le signal d'entrée.

En général, on utilise la version normalisée de la fonction de corrélation :

$$C_{xy}(k) = \frac{1}{N} \sum_{i=0}^{N-1} x_i \cdot y_{i-k}$$

Sous cette forme "réaliste", on peut interpréter la signification physique de cette définition. On a une somme de produits avec un indice (k) qui va s'incrémenter pour les 2 signaux x et y , sur un nombre fini de valeurs. Donc on va parcourir une partie du signal x et une partie du signal y , comportant N échantillons. Cela revient à multiplier échantillon par échantillon, les 2 parties des 2 signaux, l'une étant décalée par rapport à l'autre. k est une valeur de décalage. Dans les applications, n prend plusieurs valeurs entre 2 limites, c'est pourquoi on obtient une fonction.

b) Domaines des statistiques et des probabilités

Dans le domaine des statistiques, l'étude de la corrélation entre 2 variables permet de quantifier si elles sont liées, comme son nom l'indique, la manière dont la variation de l'une dépend de l'évolution de l'autre.

Définitions

La corrélation est une notion très utilisée en statistiques. Définissons d'abord la variance d'une variable : elle est égale à la somme des carrés des différences entre les exemples de cette variable et leur moyenne. Soit une variable x prenant un certain nombre de valeurs $x_i=1, \dots, n$.

$$\text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

La variance caractérise donc la somme des écarts d'une variable par rapport à sa moyenne. L'écart-type de la variable x est définie par :

$$\sigma_x = \sqrt{\text{var}(x)}$$

où

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k \quad (\text{idem pour } y)$$

est sa moyenne.

Soient x et y deux variables et leurs moyennes respectivement $\bar{x}$ et $\bar{y}$. La covariance entre ces 2 variables est définie par :

$$\text{cov}(x, y) = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})(y_k - \bar{y})$$

où n est le nombre de réalisations des variables, x_k est la k^{e} réalisation de la variable x (et y_k est la k^{e} réalisation de la variable y).

Le produit de corrélation est la version normalisée (il est compris entre 0 et 1) de la covariance :

$$\text{cor}(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}$$

où

$$\sigma_x = \sqrt{\text{var}(x)}$$

est appelé écart-type de la variable x .

Une autre manière de calculer le produit de corrélation est de considérer les variables centrées, c'est à dire de remplacer chacune de leur n réalisation x_i par x_i' tel que :

$$x_i' = x_i - \bar{x}$$

Ainsi, en calculant

$$\frac{1}{n} \sum_{k=1}^n x_k' \cdot y_k'$$

on retrouve l'expression de la covariance statistique.

De plus, en considérant les variables réduites, c'est à dire ramenées à :

$$x_i' = \frac{x_i - \bar{x}}{\sigma_x}$$

on retrouve l'expression de la corrélation statistique.

Un signal peut être continu ou discret. Dans les 2 cas, il peut être considéré comme une variable aléatoire. Dans le cas du signal, la corrélation permet de comparer des parties de 2 signaux ou du même signal. Par rapport à la notation utilisée ci-dessus, x_i et x_j seraient les 2 zones comparées du ou des signaux, c'est à dire correspondraient à un décalage donné.

Pour que la corrélation des signaux puisse correspondre à la corrélation statistique et bénéficier de ses capacités de quantification des relations existants entre des variables, et donc servir d'outil de mesure de la ressemblance entre eux, il faut que ces signaux soient au moins centrés. La "corrélation" entre eux correspond alors à leur covariance statistique. Si en plus, ils sont réduits, cette même opération correspond à leur corrélation statistique. Si ça n'est pas le cas, le terme "corrélation" est un peu un abus de langage.

c) Corrélation bi-dimensionnelle : application aux images

De manière équivalente au cas monodimensionnel, on définit la corrélation bidimensionnelle :

$$\phi_{fg}(k, l) = \frac{1}{N^2} \sum_i \sum_j f(i, j) \cdot g(i - k, j - l)$$

où f et g sont 2 fonctions bi-dimensionnelles, par exemple 2 images carrées comportant N pixels par côté, c'est à dire avec $1 \leq i \leq N$ et $1 \leq j \leq N$. Cette fonction d'autocorrélation est également bi-dimensionnelle.

Comme dans le cas des signaux mono-dimensionnels, en général on utilise des images centrées et réduites, c'est à dire qu'on soustrait à tous les pixels l'image moyenne, puis on les divise par l'écart-type de l'image (complète), ce qui a pour effet de ramener l'expression de la fonction de corrélation à la corrélation statistique.

$$f_c(i, j) = \frac{f(i, j) - \mu_f}{\sigma_f} \quad \text{et} \quad g_c(i, j) = \frac{g(i, j) - \mu_g}{\sigma_g}$$

où μ et σ sont respectivement la moyenne et l'écart-type de chaque fonction. Cette opération a pour but de normaliser les 2 images, à savoir que toutes les pixels ont des valeurs comprises entre -1 et 1 , et que la valeur absolue maximale correspond à -1 ou $+1$.

Comme pour la corrélation mono-dimensionnelle, il faut gérer les effets de bord.

d) Gestion des effets de bord

On a vu plus haut que pour le calcul de la fonction de corrélation, on pouvait avoir besoin des échantillons situés en dehors de l'intervalle $[0, N-1]$. Il y a plusieurs manières de les traiter :

- Les prendre égaux à 0
- La fonction comprise dans l'intervalle est considérée comme une période de la fonction complète
- On peut utiliser les échantillons du signal, s'ils sont connus.

Le premier cas est à éviter car il apporte des distorsions par rapport aux autres.

L'autre solution, consistant à ne pas prendre en compte les cas où le signal décalé possède une partie qui n'est pas en vis-à-vis avec le signal fixe, mène à un signal résultat plus court que le signal de départ (le signal fixe). Cela n'est pas gênant dans les applications de recherche de motifs dans une image, par exemple. Chaque pixel de l'image-résultat représente une position possible du motif dans l'image testée. La solution décrite ici consiste donc à ne pas prendre en compte les positions où le motif ne serait pas complet sur l'image de test.


```

premier signal :      -1  0  2
deuxième signal :    2  0 -1
produit              0  0
somme                -----> 0    =Cxx(-1)

```

```

premier signal :      -1  0  2
deuxième signal :    2  0 -1
produit             -2  0 -2
somme                -----> -4    =Cxx(0)

```

```

premier signal :      -1  0  2
deuxième signal :    2  0 -1
produit              0  0  0
somme                -----> 0    =Cxx(1)

```

```

premier signal :      -1  0  1
deuxième signal :    2  0 -1
produit              1  0  0
somme                -----> 2    =Cxx(2)

```

Le résultat obtenu pour la corrélation est donc :

```
Résultat          1  0 -4  0  2 = Cxx(τ)
```

On constate que la fonction de corrélation n'était pas symétrique car il s'agit ici d'une intercorrélation entre 2 signaux différents.

Finalement on divise ce résultat par la longueur des 2 signaux :

```
Division par 3 (résultat final)      1/3  0  -4/3  0  2/3 = Cxx(τ) (version normalisée)
```

Exemple d'un signal court et d'un signal long

Voyons maintenant comment la corrélation peut servir à détecter des similarités entre un signal de test (long) et un signal de référence (court). Supposons que notre signal court se trouve représenté quelque part dans le signal long. On a alors :

```

signal long :      0  -1  0  1  0  1  1  0  0  -1  -1
signal court :
                                -1  0  1      --> -0.1
                              -1  0  1      --> -0.1
                             -1  0  1      --> -0.1
                            -1  0  1      --> 0.1
                           -1  0  1      --> 0
                          -1  0  1      --> 0
                         -1  0  1      --> 0.2
                        -1  0  1      --> 0
                       -1  0  1      --> -0.1
                      -1  0  1      --> 0
                     -1  0  1
                    -1  0  1
                   -1  0  1
                  -1  0  1
                 -1  0  1
                -1  0  1
               -1  0  1
              -1  0  1
             -1  0  1
            -1  0  1
           -1  0  1
          -1  0  1
         -1  0  1
        -1  0  1
       -1  0  1
      -1  0  1
     -1  0  1
    -1  0  1
   -1  0  1
  -1  0  1
 -1  0  1

```

On voit dans cet exemple, que le produit de corrélation est maximal lorsque les 2 signaux coïncident le plus. En effectuant un seuillage (par exemple avec un seuil égal à 0,1), on peut effectuer une détection automatique de ressemblance.

f) Utilisation de Scilab

On peut calculer automatiquement ces résultats avec la fonction *corr* de Scilab. Cette fonction calcule la covariance statistique : elle renvoie

$$C_{xy}(m) = \frac{1}{n} \sum_{k=1}^{n-m} (x(k) - \bar{x})(y(k+m) - \bar{y}), \quad m=0, \dots, N-1,$$

N étant le nombre de décalages effectués, que l'on peut choisir (en paramètre de la fonction),

n est la dimension des vecteurs x et y (qui doivent être égales), correspondant au nombre de réalisations des variables x et y .

Remarque sur la signification de cette expression : cette fonction ne calcule la corrélation que pour des décalages décroissants (quand m augmente, cela correspond à un décalage du signal y vers la gauche), ce qui signifie :

- 1) que la corrélation calculée ne représente que la moitié des valeurs calculées dans le cas général (en effet, dans le cas général, la corrélation est calculée avec des décalages positifs et négatifs) ;
- 2) que les résultats seront donnés dans l'ordre de τ décroissant.

Remarques sur l'utilisation de cette fonction :

- 1) les 2 vecteurs passés en argument doivent posséder la même taille ;
- 2) le nombre de décalages demandés (3^e argument) peut être quelconque, mais au maximum égal à la taille de ces 2 vecteurs.

```
-->a = [1 2 3]
a =
! 1. 2. 3. !

-->b = a
b =
! 1. 2. 3. !

-->corr(a,b,3)
ans =
! 0.6666667 0. - 0.3333333 !
```

On peut remarquer que la corrélation n'est pas commutative

```
-->corr(a,b,3)
ans =
! 0.3333333 0.3333333 - 0.3333333 !

-->corr(b,a,3)
ans =
! 0.3333333 - 0.3333333 0. !
```

On peut remarquer également qu'il y a des effets de bords : en se déplaçant, on est sorti des fenêtres des 2 signaux, et on a considéré qu'il y avait des échantillons égaux à 0.

```
-->a = [0 -1 0 1 0 1 1 0 0 -1]
a =
! 0. - 1. 0. 1. 0. 1. 1. 0. 0. - 1. !

-->b = [0 0 0 0 0 0 0 -1 0 1]
b =
! 0. 0. 0. 0. 0. 0. 0. - 1. 0. 1. !

-->corr(a,b,10)
ans =
! - 0.1 - 0.1 - 0.1 0.1 0. 0. 0.2 0. - 0.11 - 0.01 !
```

g) Applications

Ce paragraphe décrit quelques exemples d'applications concrètes de la fonction de corrélation :

- la détection d'un signal noyé dans du bruit, par autocorrélation et par détection synchrone ;
- la mesure de retard (appliquée par exemple à la localisation d'une fuite d'eau sur une canalisation) ;
- en traitement de la parole, la détection de voisement ;
- en traitement d'images, la recherche de formes (caractères, visages) et leur reconnaissance.

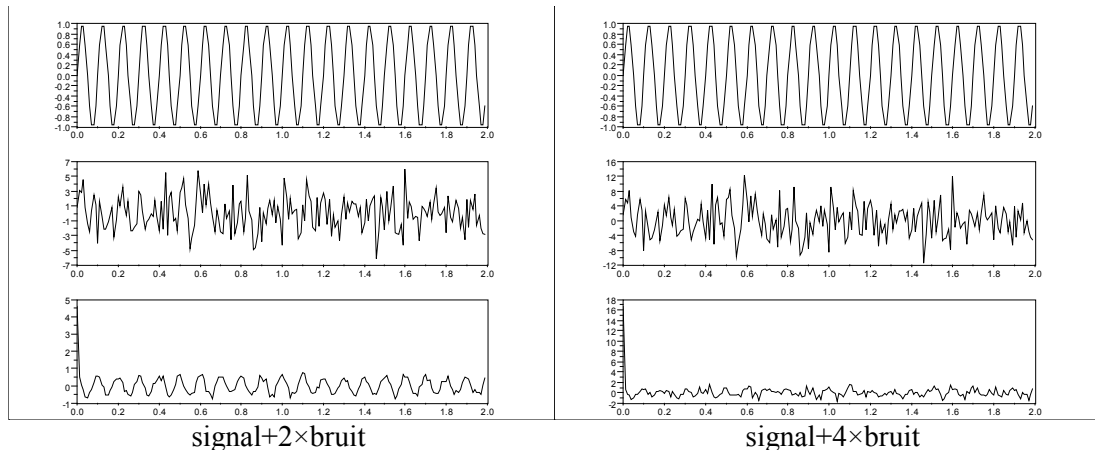
Des illustrations par programmation avec Scilab sont données dans certains cas.

g.1) Détection d'un signal noyé dans du bruit

g.1.a) Par autocorrélation

Une extension des propriétés de la corrélation de signaux périodiques et de signaux aléatoires concerne la corrélation de signaux additionnés à du bruit : selon l'amplitude du bruit elle montre une périodicité plus ou moins marquée, qui est celle du signal.

Dans les exemples suivants, l'amplitude du bruit est respectivement 2 fois puis 4 fois celle du signal. Dans chacune des figures ci-dessous, le cadre du haut représente le signal, celui du milieu le signal bruité et celui du bas la fonction d'autocorrélation.



Le signal est tellement bruité qu'on a du mal à le différencier du bruit seul à l'oeil, mais la fonction de corrélation montre clairement une périodicité, qui est celle du signal. La corrélation permet donc d'extraire le signal noyé dans du bruit.

Avec une amplitude du bruit 2 fois plus grande que celle du signal, la périodicité du signal est facilement **délectable**, par simple **seuillage à zéro** par exemple.

Avec une amplitude du bruit 4 fois plus grande que celle du signal, la périodicité du signal est plus difficilement mesurable.

En même temps, la hauteur du pic au niveau du zéro de l'axe des abscisse augment avec l'intensité du bruit, démontrant la nature de plus en plus aléatoire du signal.

Programme Scilab

```
//corr_bruit_1.sce
// détection de signal périodique par autocorrélation
rand('normal');
n=1000;
bruit=rand(1,n);
t=0:0.01:9.99;
signalbase = sin(2 * %pi * 10 * t);
rand('uniform')
rand('seed',0)
noise = rand(signalbase);
signalbruit = signalbase + 1*bruit;
xbasc()
corrsignal=corr(signalbruit,1000);
nb=200
xsetech([0,0,1,1/3])
plot2d(t(1:nb), signalbase (1:nb));
xsetech([0,1/3,1,1/3])
plot2d(t(1:nb), signalbruit (1:nb));
xsetech([0,2/3,1,1/3])
plot2d(t(1:nb), corrsignal (1:nb))
```

//essayer différentes valeurs du facteur de bruit : 4, 6...

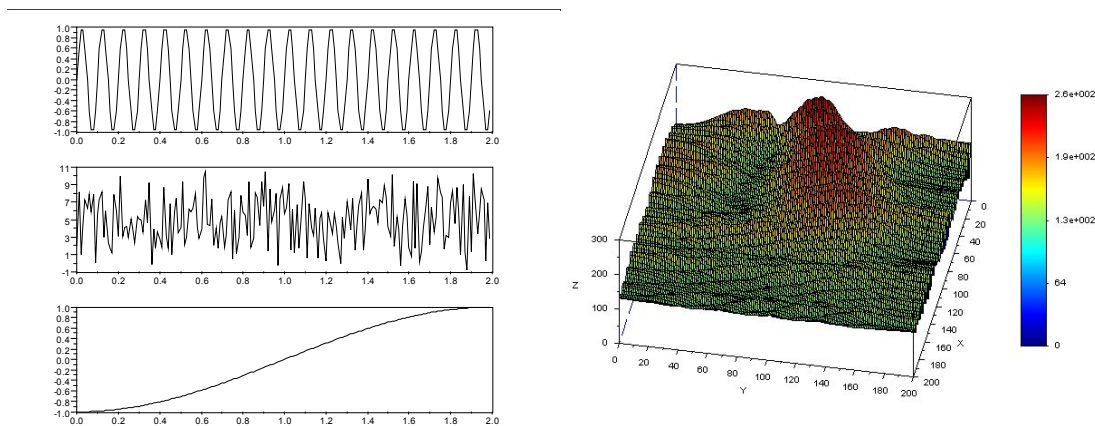
g.1.b) Par détection synchrone

La méthode dite de détection synchrone est plus puissante que l'autocorrélation pour détecter un signal noyé dans du bruit. Elle permet de détecter des signaux avec des rapports signal/bruit beaucoup plus grand qu'avec l'autocorrélation.

Plutôt que de calculer la fonction d'autocorrélation du signal bruité, elle consiste à calculer l'intercorrélation entre le signal bruité et un signal non bruité dont on fait varier la fréquence de manière alternative (on parle également de modulation, ou modulation).

Cette méthode est utilisée par exemple en radioastronomie pour détecter les pulsations des étoiles très lointaines, fortement entachées de bruit.

Pour tester cette méthode, on peut définir un signal sinusoïdal puis un 2^e signal composé de ce même signal additionné d'un signal aléatoire d'amplitude 10 fois supérieure, par exemple. Puis on écrit un algorithme qui calcule en boucle (pour un certain nombre d'itérations) la corrélation entre le signal noyé dans du bruit et un signal sinusoïdal dont la fréquence est variée graduellement (par l'intermédiaire d'un signal modulant, sinusoïdal, à basse fréquence). Un exemple de résultat obtenu de cette manière est illustré par la figure suivante :



La figure de gauche représente le signal initial (à retrouver dans le bruit), le même signal noyé dans le bruit, et le signal modulant.

La figure de droite représente le résultat de toutes ces mesures de corrélation, sous la forme d'une fonction à 2 dimensions. Ces 2 dimensions sont la fréquence du signal modulé, d'une part, et la variable de la fonction de corrélation, d'autre part.

En comparaison des résultats de la détection du signal dans du bruit par la méthode d'autocorrélation, pour laquelle on pouvait à la limite encore détecter le signal avec un niveau de bruit 4 fois supérieur à celui-ci, ici avec un niveau de bruit 10 fois supérieur au signal, la détection est encore facile. En effet, le pic de la fonction de corrélation correspond à la fréquence du signal recherché.

Programme Scilab

```
//corr_sync_1.sce
//Détection de signal noyé dans du bruit, par détection synchrone

nb_corr=200; //nombre de points pour la corrélation

//définition du signal à analyser (sinus bruité)
Te=0.01; //période d'échantillonnage
nech=nb_corr*Te; //nombre d'échantillons
t=0:Te:nech; //nb_corr points de temps « vrai »
fs=10; //fréquence du signal à analyser
signalbase=sin(2*pi*fs*t); //signal pur

rand('seed',0) ;
noise = rand(signalbase);
signalnoise = signalbase + 10*noise; //ajout de bruit -> signal bruité

//définition du signal modulant
```

```

fm=0.25; //fréquence du signal modulant
modulant=sin(2*pi*fm*t-pi/2); //signal modulant (déphasé pour commencer à -1)
nb_iter=nb_corr;
res=zeros(nb_corr,nb_iter); //matrice pour les résultats de la corrélation
for i=1:nb_iter //on calcule la corrélation pour 'nb_iter'
    //valeurs de fréquence du signal de test
    fsm=fs+modulant(i);
    signaltest=sin(2*pi*fsm*t); //signal de test : modulé en fréquence
    [corrsignal,moy]=corr(signalnoise,signaltest,nb_corr);
    //Inter-corrélation entre signal bruité et signal de test
    res(:,i)=corrsignal';
end

//Affichages
xbasec
set("current_figure",0)
n=nb_corr
xsetech([0,0,1,1/3])
plot2d(t(1:n), signalbase(1:n));
xsetech([0,1/3,1,1/3])
plot2d(t(1:n), signalnoise(1:n));
xsetech([0,2/3,1,1/3])
plot2d(t(1:n), modulant(1:n));

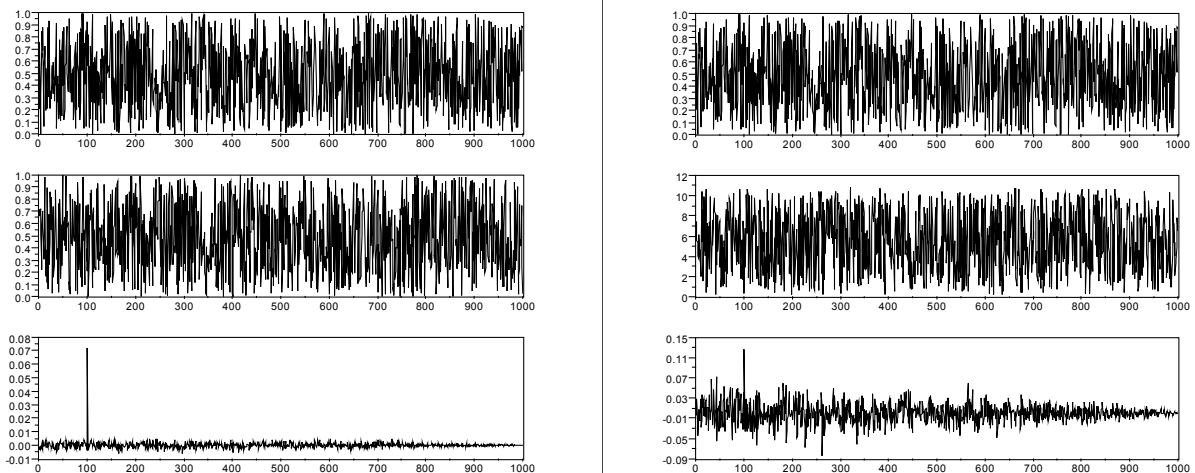
set("current_figure",1)
s1=size(res,1)
s2=size(res,2)
x=linspace(0,s1,s1);
y=linspace(0,s2,s2);
xset("colormap",jetcolormap(256));
colorbar(0,256);
res=256.*(res-min(res))./(max(res)-min(res)); //normalisation pour affichage
plot3d1(x,y,res);

```

g.2) Mesure de retards

Lorsque l'on dispose d'un signal quelconque (par exemple du bruit) et d'une version décalée dans le temps du même signal, la fonction d'autocorrélation possède un maximum à la position de ce décalage. Elle permet donc de déterminer ce décalage.

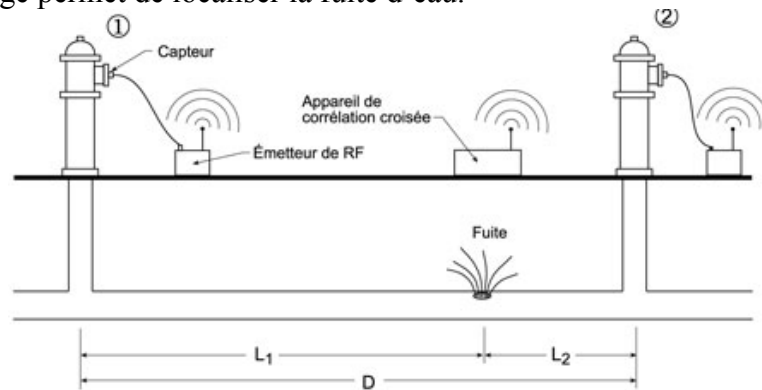
Dans l'exemple suivant, un signal aléatoire ("signal1"), à répartition uniforme, sous la forme d'une séquence de 1000 échantillons, est généré, ainsi qu'un autre, du même nombre d'échantillons, obtenu par retard du 1^{er} signal de 100 échantillons. Ses 100 premiers échantillons sont générés de la même manière que pour 1^{er}. Dans la fonction d'intercorrélation de ces signaux, on observe alors la propriété de détection de retard de la fonction de corrélation : celle-ci est nulle partout, comme pour un signal aléatoire, sauf au niveau du retard ; un pic clairement défini est présent au niveau de l'échantillon correspondant au retard, dans le signal résultant de la corrélation.



Dans l'exemple de droite, le signal décalé soit additionné d'un signal bruité d'amplitude 10 fois supérieure. Le décalage est toujours facilement détectable :

Application à la localisation de fuites d'eau

La propriété décrite ci-dessus peut être appliquée par exemple à la détection de fuites d'eau dans des canalisations : 2 microphones placés de part et d'autre de la fuite enregistrent le même signal mais décalé dans le temps. Ce décalage permet de localiser la fuite d'eau.



Temps mis par le signal ① pour arriver : $T_1 = L_1 / V$, où V représente la vitesse de propagation du son dans la conduite
Temps mis par le signal ② pour arriver : $T_2 = L_2 / V$
Décalage du signal ② par rapport au signal ① : $\Delta T = T_2 - T_1 = (L_2 - L_1) / V$
 $L_2 = D - L_1 \Rightarrow \Delta T = (D - 2L_1) / V \Rightarrow L_1 = (D - V\Delta T) / 2$

http://irc.nrc-cnrc.gc.ca/pubs/ctus/40_f.html

Le principe consiste à écouter le bruit causé par la fuite en 2 points différents. Si l'on peut identifier le même signal et le décalage de l'un par rapport à l'autre, on peut en déduire par calcul la distance de la fuite à l'un des 2 points d'écoute (indifféremment).

d) Traitement de la parole : détection de voisement

Le traitement automatique de la parole (reconnaissance, compression, synthèse...) est souvent basé sur la détection de voisement des signaux.

Les signaux de parole comportent des parties voisées et des parties non voisées. Les parties élémentaires sont appelées phonèmes. Les phonèmes voisés sont générés par les cordes vocales, comprennent les voyelles et plus généralement les sons périodiques ("on", "ou", etc). Les phonèmes non-voisés sont générés par le passage de l'air de différentes manières dans la bouche, comportent les consonnes et sont assimilables à du bruit. Ces derniers peuvent être entretenus mais transitoires ("ch", "s", etc), ou percussif ("t", "k", etc).

On a vu que l'autocorrélation d'un signal périodique est elle-même périodique et pour un signal non-périodique elle est nulle sauf à l'origine. Cette propriété peut donc être utilisée pour détecter une périodicité dans un signal de parole (et donc un voisement) : il suffit de détecter les parties des signaux où l'autocorrélation est très différente de 0.

Si le son est voisé, la fonction d'autocorrélation va présenter plusieurs pics. La valeur maximale est obtenue pour un décalage nul, c'est à dire à l'origine de la fonction. Il faut fixer un critère pour décider si une portion de signal est voisée ou non. Ce critère peut être basé sur la mesure de l'amplitude du premier pic de la fonction d'autocorrélation, en comparaison avec celle du pic à l'origine. Par exemple, un rapport de 0,4 est une valeur qui donne de bons résultats. On peut également ne pas prendre de décision et considérer ce rapport comme un taux de voisement.

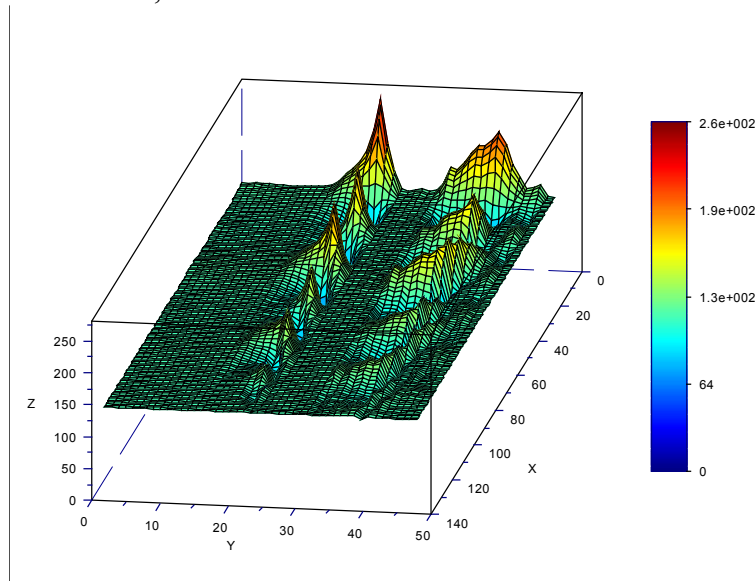
Une fois la décision de voisement éventuellement prise, il faut déterminer la fréquence du signal. On a vu avec l'exemple de détection de signal noyé dans du bruit, que la fonction d'autocorrélation est périodique de période égale à celle du signal. Le premier pic de la fonction d'autocorrélation, exprimé en nombre d'échantillons, correspond donc à une période du signal. Connaissant la fréquence d'échantillonnage, on peut convertir un nombre d'échantillons en durée, cette durée étant la période du signal. On peut donc retrouver sa période, donc sa fréquence :

$$T_0 = n_0 \times T_e$$

Pour retrouver la fréquence du son, il suffit de regarder la position centrale du pic, en nombre d'échantillons.

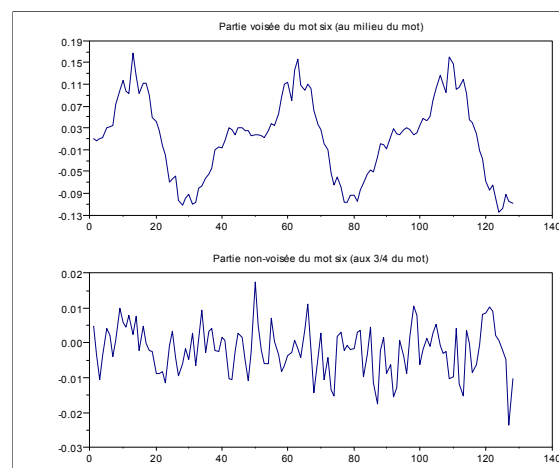
En pratique, on compare des morceaux correspondant à des durées de l'ordre de 20 à 30 ms.

Par exemple, pour le mot "zéro", le résultat obtenu est le suivant :



On détecte donc bien une périodicité aux endroits correspondant au "é" et au "o" du mot "zéro"

Visualisons une fenêtre de signal correspondant à une partie voisée et une autre correspondant à une partie non voisée. Avec le mot "six", en prenant une fenêtre de signal située à 1/4 du son complet ou aux 3/4 environ, on est sur des parties non-voisées.



On constate que l'on a bien signal d'apparence périodique qui doit correspondre au "i" pour la figure du haut, et pour celle du bas un signal d'apparence aléatoire qui doit correspondre au "s" de la fin du mot.

e) Reconnaissance de caractères

Par exemple, pour 2 images de taille 256x256 pixels, on va définir une zone centrale de 151x151 pixels, allant du pixel de coordonnées (50,50) jusqu'au pixel de coordonnées (201,201). Comme 201+50=251, on est sûr de ne pas sortir de l'image. La recherche va s'effectuer sur ± 20 pixels. L'expression de la fonction de corrélation (qui est une autocorrélation ici), va donc se préciser par :

$$\phi_{fg}(k,l) = \frac{1}{151^2} \sum_{i=50}^{200} \sum_{j=50}^{200} f(i,j).g(i-k,j-l) \text{ avec } k,l=[-20,20]$$

La fonction d'autocorrélation calculée sur la même image (c'est à dire avec $g=f$) présente un pic en son centre, égal à 1. Ce qui est normal puisque cette position correspond à un décalage nul de l'image à comparer par rapport à l'image de référence, c'est à dire la même image. La valeur unitaire du pic

provient d'une part de la normalisation apportée par le fait que les images soient centrées et réduites, et d'autre part que le résultat de la double somme soit divisée par N^2 .

Si la fonction d'autocorrélation entre les 2 images dépasse un certain seuil (par exemple 50% du maximum, c'est à dire 0,1), on peut décider qu'il y a suffisamment de ressemblance entre les 2 images pour qu'elles représentent la même chose, et donc que la ressemblance soit positive.

f) Analyse de texture

La texture d'une image peut être aléatoire (ex. sable, mur crépi...) ou présenter des régularités (mailles d'un tissu, mur de briques, etc). Dans le premier cas, la fonction d'autocorrélation va présenter un pic assez étroit à la position (0,0), puis être quasi-nul partout ailleurs. Par contre, dans le 2^e cas, elle présentera elle même une périodicité.

II.3) Caractérisation des systèmes par équations différentielles

Il a été vu précédemment qu'un système linéaire et stationnaire (ou filtre) pouvait être décrit par une intégrale de convolution avec sa réponse impulsionnelle. On va voir qu'il peut également être décrit par une équation différentielle linéaire à coefficients constants.

Cette équation différentielle va notamment permettre de connaître la réponse impulsionnelle du système.

En mathématique, les équations différentielles portent sur des fonctions, la variable étant souvent notée x . Par exemple, dans le cas du 1^{er} ordre, cette équation est définie par :

$$Ay'(x) + By(x) = u(x)$$

où u est une fonction de la variable x .

Dans le cas du 2^e ordre, l'équation est :

$$Ay''(x) + By'(x) + Cy(x) = u(x)$$

$y''(x)$ et $y'(x)$ représentent respectivement la dérivée et la dérivée seconde de y par rapport à la variable x :

$$y''(x) = \frac{d^2y}{dx^2} \quad \text{et} \quad y'(x) = \frac{dy}{dx}$$

L'équation avec membre de droite nul ("sans second membre") est appelée équation homogène associée ; respectivement, dans les 2 cas ci-dessus :

$$Ay'(x) + By(x) = 0$$

et

$$Ay''(x) + By'(x) + Cy(x) = 0$$

Ces deux équations sont appelées "à coefficients constants" car les coefficients A , B et C sont des constantes réelles.

Remarque : souvent, pour simplifier l'écriture on omet la variable. Par exemple dans le cas du 2^e ordre :

$$Ay'' + By' + Cy = u$$

Dans le cas des systèmes, y représentera la grandeur de sortie de ce système (par exemple une tension), u la grandeur d'entrée, et la variable x représente le temps. Elle est donc remplacée par t , $y(x)$ est remplacé par $s(t)$, et $u(x)$ par $e(t)$:

$$As''(t) + Bs'(t) + Cs(t) = e(t) \quad \text{ou} \quad A \frac{d^2s(t)}{dt^2} + B \frac{ds(t)}{dt} + Cs(t) = e(t)$$

Par exemple dans le cas d'un circuit RC, l'équation est du 1^{er} ordre, définie par :

$$RC \frac{ds(t)}{dt} + s(t) = e(t)$$

Le terme de gauche correspond au régime libre, c'est à dire au comportement du système indépendamment des signaux d'entrée. Il dépend des conditions initiales, sa solution disparaît au cours du

temps. Elle constitue le régime transitoire. Le terme de droite correspond au signal d'entrée imposé au système. La solution de l'équation complète correspond au régime entretenu.

En pratique, $e(t)$ pourra être une impulsion (pour étudier la réponse impulsionnelle), un échelon (pour étudier la réponse indicielle), ou un signal sinusoïdal (pour étudier la réponse en fréquence).

Les équations différentielles du 1^{er} ou du 2^e ordre peuvent être résolues directement, mais si l'ordre de l'équation augmente, cette méthode devient très difficile. Un outil puissant et simple permettant de les résoudre est la Transformée de Laplace.

II.3.1) Résolution directe des équations différentielles linéaires à coefficients constants

a) Equations différentielles du 1^{er} ordre

Définition

Une équation différentielle du 1^{er} ordre à coefficients constants est définie par :

$$Ay' + By = u$$

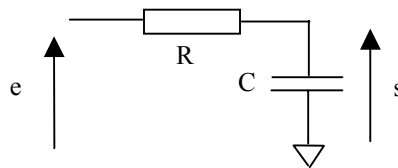
L'équation homogène associée est définie par :

$$Ay' + By = 0$$

où y est une fonction d'une variable x (par exemple), A est une constante réelle non nulle et B une constante réelle.

Exemple : circuit RC

On considère le circuit RC suivant :



Soit $i(t)$ le courant passant dans la résistance. Il est égal à :

$$i(t) = \frac{e(t) - s(t)}{R}$$

Ce courant est le même que celui passant dans le condensateur, donc :

$$i(t) = C \frac{ds(t)}{dt}$$

On en déduit l'équation différentielle :

$$RC \frac{ds(t)}{dt} + s(t) = e(t)$$

Ici, la variable qui remplace le x de l'équation précédente est le temps t .

Une fois une équation différentielle définie, il reste à déterminer ses solutions, c'est à dire, pour une fonction $u(x)$ donnée, la fonction $y(x)$ correspondante. Pour les systèmes étudiés en fonction du temps, cela revient à déterminer sa sortie $s(t)$ pour une entrée $e(t)$ donnée.

On peut montrer que cette résolution passe par la résolution de la même équation mais avec second membre nul, appelée équation homogène associée.

Dans le cas du 1^{er} ordre, l'équation complète est de la forme :

$$Ay' + By = u$$

Supposons que y_0 soit une solution particulière de l'équation complète. On a :

$$Ay' + By = u$$

et

$$Ay_0' + By_0 = u$$

donc, par soustraction de ces 2 équations :

$$A(y - y_0') + B(y - y_0) = 0$$

Ce qui signifie que $y - y_0'$ est une solution (particulière) de l'équation homogène. Si on additionne cette solution à la solution particulière de l'équation complète y_0' , on obtient y , qui est la solution générale de l'équation complète. D'où la propriété suivante :

La solution d'une équation différentielle du 1^{er} ordre est la somme de la solution générale de la même équation sans second membre et d'une solution particulière de l'équation complète.

a.1) Résolution de l'équation homogène associée

Puisque $A \neq 0$, on peut mettre l'équation sous la forme :

$$y' = ay \quad (1)$$

avec $a = -B/A$.

Si y ne s'annule pas sur l'intervalle d'étude, on peut l'écrire :

$$\frac{y'}{y} = a$$

La fraction y'/y est la dérivée de $\ln(y)$. On peut donc l'écrire :

$$(\ln y)' = a$$

En intégrant, on a :

$$\ln y = ax + b$$

où b est une constante réelle. On a donc :

$$y = e^{ax+b} = e^{ax} \times e^b$$

$\leftrightarrow$

$$y = \lambda e^{ax}$$

où λ est une constante réelle. En pratique, cette constante dépend des conditions initiales, c'est à dire de la valeur de $u(x)$ (ou $e(t)$) pour $x=0$ (resp. $t=0$). Par exemple, dans le cas d'un circuit RC, elles correspondraient à la tension de sortie, aux bornes du condensateur, à $t=0$, donc sa charge initiale.

Remarque : l'ensemble des solutions de cette équation est un espace vectoriel de dimension 1 (=une droite), de base e^{ax} . Il ne s'agit donc pas d'une solution unique, mais d'une infinité de solutions.

Exemple

$$y' - y = 0$$

L'ensemble des solutions est défini par λe^{ax} , avec $a=1$ et λ un réel quelconque.

On peut vérifier par exemple que e^x est solution :

$$y' - y = e^x - e^x = 0$$

De même que $3e^x$:

$$y' - y = 3e^x - 3e^x = 0$$

a.2) Résolution de l'équation complète (=avec second membre)

On recherche des solutions particulières de l'équation complète. Parfois, certaines solutions sont évidentes. Par exemple, pour l'équation :

$$y' - y = e^{2x},$$

une solution particulière "évidente" est e^{2x} . La solution générale de cette équation est donc (celle de l'équation homogène a été étudiée plus haut) :

$$y = ae^x + e^{2x},$$

où a est réel quelconque.

Il n'est pas toujours évident de trouver une solution particulière de l'équation complète. Il existe la méthode dite de "variation de la constante", qui permet de le faire. Elle consiste à considérer que la constante de la solution générale de l'équation homogène est une fonction de la variable (ici, x), de remplacer cette solution dans l'équation complète et d'identifier le terme de gauche avec celui de droite.

Exemple

$$y' - y = e^x \cos x$$

On a vu que la solution complète de l'équation homogène était $y = ae^x$ (a constante). On considère que a est fonction de x (on refait apparaître la variable x pour mettre en évidence cette dépendance) :

$$y(x) = a(x) \cdot e^x$$

Alors

$$y'(x) = a'(x)e^x + a(x)e^x$$

On remplace y' dans l'équation complète :

$$y' - y = a'e^x + ae^x - ae^x = a'e^x$$

On identifie alors le terme de droite avec celui de l'équation complète :

$$a'e^x = e^x \cos x \leftrightarrow a' = \cos x$$

d'où

$$a = \sin x$$

La solution générale de l'équation complète est donc :

$$y = e^x (a + \sin x)$$

avec a constante réelle. On peut vérifier ce résultat en remplaçant cette solution dans l'équation :

$$y' - y = e^x (a + \sin x) + e^x \cos x - e^x (a + \sin x) = e^x \cos x$$

b) Equations différentielles du 2^e ordre

Ces équations sont de la forme :

$$ay'' + by' + cy = u$$

avec a, b, c des constantes réelles ($a \neq 0$).

L'équation sans second membre (ou homogène) associée est :

$$ay'' + by' + cy = 0$$

Comme dans le cas du 1^{er} ordre, **la solution de l'équation complète est la somme d'une solution particulière cette équation, et de la solution générale de l'équation sans second membre associée.**

b.1) Recherche de la solution générale de l'équation homogène

La résolution de l'équation homogène n'est pas si simple que dans le cas du 1^{er} ordre.

La méthode employée part du constat qu'il peut exister des solutions de la forme :

$$e^{rx}$$

L'équation peut alors s'écrire :

$$ar^2 e^{rx} + br e^{rx} + ce^{rx} = 0$$

$$e^{rx} (ar^2 + br + c) = 0$$

Le terme e^{rx} n'étant jamais nul, la résolution de l'équation différentielle se ramène à celle du polynôme entre parenthèses, appelé polynôme caractéristique de l'équation :

$$ar^2 + br + c = 0$$

Si r est solution du polynôme caractéristique, e^{rx} est solution de l'équation homogène.

Exemple

$$y'' - 3y' + 2y = 0$$

L'équation caractéristique associée est :

$$r^2 - 3r + 2 = 0$$

Une solution est $r=1$. On peut vérifier que e^x est solution de l'équation homogène :

$$y'' - 3y' + 2y = e^x - 3e^x + 2e^x = 0$$

Si r_1 et r_2 sont solutions du polynôme caractéristique, $A_1 e^{r_1 x} + A_2 e^{r_2 x}$ sont solutions de l'équation homogène. Les solutions correspondantes de l'équation différentielle homogène sont, à chaque fois, des ensembles de fonctions.

De plus, on peut démontrer que si y_1 et y_2 sont 2 solutions particulières de l'équation homogène, l'ensemble des solutions de l'équation homogène est défini par :

$$y = A_1 y_1 + A_2 y_2$$

avec A_1 et A_2 constantes réelles. Cet ensemble définit un espace vectoriel de dimension 2 (=un plan).

La forme des solutions dépend de la valeur de Δ :

- $\Delta > 0$

$$y = A_1 e^{r_1 x} + A_2 e^{r_2 x}$$

où r_1 et r_2 sont les 2 racines du polynôme caractéristique associé à l'équation différentielle, A_1 et A_2 2 constantes réelles dépendant des conditions initiales.

- $\Delta = 0$

$$y = (Ax + B)e^{rx}$$

où r est la racine double du polynôme caractéristique, A et B 2 constantes réelles dépendant des conditions initiales.

- $\Delta < 0$

$$y = e^{\alpha x} (A_1 \cos \beta x + A_2 \sin \beta x)$$

où α et β sont les parties réelle et imaginaire des racines complexes du polynôme caractéristique, A_1 et A_2 2 constantes réelles. On peut l'écrire également sous la forme :

$$y = Be^{\alpha x} \cos(\beta x + \varphi)$$

La solution est une fonction sinusoïdale dont l'enveloppe est exponentielle décroissante : une sinusoïde amortie.

Exemple

Résolution de l'équation suivante dans le domaine des complexes :

$$y'' + 2y' + 2y = 0$$

On a vu que les 2 solutions de l'équation caractéristique associée étaient complexes :

$$r_1 = -1 - j \quad \text{et} \quad r_2 = -1 + j$$

Donc la solution générale de l'équation différentielle est :

$$y = e^{-x} (A \cos x + B \sin x)$$

A et B réels quelconques.

Exemple

Dans l'exemple précédent, on peut vérifier que $y_1 = e^{-x} \cos x$ et $y_2 = e^{-x} \sin x$ sont 2 solutions particulières de l'équation différentielle

$$y'' + 2y' + 2y = 0$$

Par exemple, pour y_1 :

$$y_1'' + 2y_1' + 2y_1 = (e^{-x} \cos x)'' + 2(e^{-x} \cos x)' + 2e^{-x} \cos x$$

$$\begin{aligned}
&= (-e^{-x} \cos x - e^{-x} \sin x)' - 2e^{-x} \cos x - 2e^{-x} \sin x + 2e^{-x} \cos x \\
&= \underline{e^{-x} \cos x} + \underline{e^{-x} \sin x} + \underline{e^{-x} \sin 2x} - \underline{e^{-x} \cos x} - 2e^{-x} \cos x - \underline{2e^{-x} \sin x} + 2e^{-x} \cos x \\
&= 0
\end{aligned}$$

Donc toute combinaison de la forme

$$y = e^{-x} (A_1 \cos x + A_2 \sin x)$$

est solution.

b.2) Recherche de solutions particulières de l'équation avec second membre

Forme des solutions particulières, en fonction de la forme du second membre

Selon la forme du second membre, on peut connaître la forme des solutions particulières :

- Cas d'un polynôme de degré n, les solutions particulières sont des polynômes de degré n. Par exemple, dans le cas où n=2 :

$$ay'' + by' + cy = P_n(x)$$

Une solution particulière est de la forme

$$\alpha x^2 + \beta x + \chi$$

Remarque : si des coefficients b et/ou c sont nuls, le degré du polynôme peut aller jusqu'à n+2.

Exemple

$$y'' - y' + y = x^2$$

Une solution particulière de cette équation est de la forme

$$y_0 = \alpha x^2 + \beta x + \chi$$

On remplace cette solution dans l'équation :

$$(2\alpha x + \beta)' - 2\alpha x - \beta\alpha + \alpha x^2 + \beta x + \chi = 2\alpha - 2\alpha x - \beta\alpha + \alpha x^2 + \beta x + \chi = \alpha x^2 + x(\beta - 2\alpha) + 2\alpha - \beta\alpha + \chi$$

Par identification :

$$\begin{aligned}
\alpha &= 1 \\
\beta - 2\alpha &= 0 \leftrightarrow \beta = 2 \\
2\alpha - \beta\alpha + \chi &= 0 \leftrightarrow \chi = 0
\end{aligned}$$

D'où la solution :

$$y_0 = x^2 + 2x$$

- $P_n(x)e^{sx}$
 - si s n'est pas racine du polynôme caractéristique :

$$y_0 = Q_n(x)e^{sx}$$
 - si s est racine simple du polynôme caractéristique :

$$y_0 = Q_{n+1}(x)e^{sx}$$
 - si s est racine double du polynôme caractéristique :

$$y_0 = Q_{n+2}(x)e^{sx}$$

- $\cos \omega x$ ou $\sin \omega x$

$$y_0 = A \sin \omega x + B \cos \omega x$$

- $f(x) + g(x)$

$$y_0 = y_1 + y_2$$

où : y_1 est solution de l'équation dont le 2nd membre est $f(x)$;
 y_2 est solution de l'équation dont le 2nd membre est $g(x)$

Exemple

$$y'' - y = x$$

La solution générale de l'équation homogène est :

$$y = Ae^x + Be^{-x},$$

A et B réels.

La solution générale de l'équation est donc :

$$y = -x + Ae^x + Be^{-x}$$

On peut vérifier par exemple que :

$$y = -x + e^x + e^{-x}$$

vérifie l'équation :

$$y'' = e^x + e^{-x} \text{ donc } y'' - y = e^x + e^{-x} - x - e^x - e^{-x} = x$$

De même :

$$y = x + 2e^x - e^{-x}$$

vérifie l'équation, car :

$$y'' = 2e^x - e^{-x}$$

Exemple : circuit RLC

Le fonctionnement entrée/sortie d'un circuit RLC est défini par l'équation différentielle :

$$\frac{d^2s(t)}{dt^2} + 2m\omega_0 \frac{ds(t)}{dt} + \omega_0^2 s(t) = e(t)$$

avec

$$\omega_0 = \frac{1}{\sqrt{LC}} \quad \text{et} \quad m = \frac{R}{2} \sqrt{\frac{C}{L}}$$

L'équation sans second membre associée est :

$$\frac{d^2s(t)}{dt^2} + 2m\omega_0 \frac{ds(t)}{dt} + \omega_0^2 s(t) = 0$$

Le déterminant de l'équation caractéristique associée est :

$$\Delta = b^2 - 4ac = 4m^2\omega_0^2 - 4\omega_0^2 = 4\omega_0^2(m-1)$$

Les solutions de cette équation dépendent du signe de Δ , donc de m :

- $m > 1$

La solution est la somme de 2 exponentielles :

$$x(t) = A_1 e^{a_1 t} + A_2 e^{a_2 t}$$

où a_1 et a_2 sont les racines de l'équation caractéristique associée :

$$a_1 = \omega_0(-m + \sqrt{m^2 - 1}) \quad \text{et} \quad a_2 = \omega_0(-m - \sqrt{m^2 - 1})$$

Les constantes A_1 et A_2 dépendent des conditions initiales. Elles concernent essentiellement la charge initiale du condensateur.

- $m = 1$

$$x(t) = (At + B)e^{-\omega_0 t}$$

- $m < 1$

La solution est un signal sinusoïdal pondéré (enveloppé) par une exponentielle décroissante, c'est à dire amorti.

$$y = e^{\alpha x} (A_1 \cos \beta x + A_2 \sin \beta x)$$

$$s(t) = A_0 e^{-m\omega_0 t} \cos(\omega t + \varphi)$$

avec

$$\omega = \omega_0 \sqrt{1 - m^2}$$

On peut remarquer que si $m=0$, $e^{-m\omega_0 t} = 1$, et le sinus n'est plus amorti. On a alors affaire à un oscillateur, dont la pulsation d'oscillation est la pulsation propre ω_0 . Mais la condition $m=0$ impose $R=0$, ce qui est impossible car dans un circuit réel il existe toujours des résistances non nulles (ne serait-ce que la bobine). Mais plus R est faible, plus l'amortissement est lent.

A , A_0 , A_1 , A_2 , B et ϕ sont des constantes déterminées à partir des conditions initiales. Par exemple, avec :

$$s(0^-) = 0$$

Supposons que R , L , C soient tels que le déterminant Δ soit positif. Du point de vue des conditions initiales, si on suppose le condensateur initialement chargé, avec une tension U_0 à ses bornes, on a :

$$s(0^+) = U_0 + A_1 + A_2$$

et donc :

$$A_1 = U_0 - A_2 \quad (1)$$

Pour relier A_1 et A_2 aux éléments caractéristiques du système, il faut considérer également une propriété électrique de la bobine : $i(0^-) = i(0^+)$ car le courant dans une bobine ne peut admettre de discontinuité. Or ce courant, qui passe également dans le condensateur, est défini par :

$$i(0^+) = C \frac{ds(t)}{dt} = C(A_1 a_1 + A_2 a_2) = 0$$

et

$$i(0^-) = 0$$

car le fonctionnement du système n'est pas entretenu. D'où :

$$A_1 = -\frac{a_2}{a_1} A_2 \quad (2)$$

$$s(0^+) = E + A_1 + A_2$$

et donc :

$$A_1 = -E - A_2$$

La résolution de ce système de 2 équations conduit au résultat :

$$A_1 = -\frac{a_2}{a_2 - a_1} E = -E \frac{m + \sqrt{m^2 - 1}}{2\sqrt{m^2 - 1}}$$

$$A_2 = -\frac{a_1}{a_1 - a_2} E = E \frac{m - \sqrt{m^2 - 1}}{2\sqrt{m^2 - 1}}$$

Remarque : si on avait pris $U_0=0$, la seule solution satisfaisant les conditions (1) et (2) aurait été :

$$A_1 = A_2 = 0$$

car $a_1 \neq a_2$. Ce qui est normal, car on cherche la solution de l'équation homogène, correspondant au système avec entrée nulle. Si le condensateur est initialement déchargé, aucun courant ni tension ne peuvent apparaître dans le circuit à $t=0$.

Dans le cas où le second membre n'est pas nul, il faut ajouter à la solution de l'équation homogène une solution particulière de l'équation complète. Si le second membre est un échelon E , on a $dE/dt = d^2E/dt^2 = 0$, donc il reste :

$$\omega_0^2 x(t) = E$$

soit

$$s(t) = \frac{E}{\omega_0^2}$$

donc on ajoute $\frac{E}{\omega_0^2}$ aux solutions de l'ESSM.

Méthode de la variation des constantes

Comme dans le cas du 1^{er} ordre, cette méthode consiste à considérer que les constantes de la solution générale de l'équation homogène sont des fonctions de la variable de l'équation (ici, x). On effectue la double dérivation de cette solution, on la remplace dans l'équation complète et on identifie terme à terme avec le membre de droite de l'équation complète.

Cette méthode donne des calculs relativement longs. Elle ne sera pas développée ici.

II.3.2) Résolution par Transformée de Laplace

a) Rappels sur la Transformée de Laplace

Introduction

La transformée de Laplace a été introduite pour faciliter le calcul des équations différentielles. En effet, elle permet de remplacer des dérivées par des puissances et donc ramener la résolution des équations différentielles à celle de polynômes dont la variable est notée p, variable de Laplace), dont l'étude est plus facile surtout quand l'ordre devient élevé.

La variable p est une variable complexe.

La transformée de Laplace ne s'applique qu'aux systèmes linéaires et stationnaires vérifiant le principe de causalité.

Définition

Soit une fonction s(t) causale, c'est à dire définie pour t>0. Sa transformée de Laplace est définie par :

$$L\{s(t)\} = S(p) = \int_0^{+\infty} s(t)e^{-pt} dt$$

Propriété importante : transformée d'une dérivée :

La transformée de Laplace s'obtient par :

$$L\left\{\frac{ds(t)}{dt}\right\} = \int_0^{+\infty} \frac{ds(t)}{dt} e^{-pt} dt$$

Par intégration par parties, on obtient :

$$= [s(t).e^{-pt}]_{0^-}^{+\infty} + p \int_{0^-}^{+\infty} s(t)e^{-pt} dt$$

On peut reconnaître dans le 2^e terme S(p), la transformée de s(t) :

$$L\left\{\frac{ds(t)}{dt}\right\} = pS(p) + \lim_{t \rightarrow \infty} s(t).e^{-pt} - s(0^-)$$

Dans le cas où

$$\lim_{t \rightarrow \infty} s(t).e^{-pt} = 0$$

il reste :

$$L\left\{\frac{ds(t)}{dt}\right\} = pS(p) - s(0^-)$$

Pour les dérivées d'ordre supérieur, le calcul permet d'obtenir :

$$L\left\{\frac{d^n s(t)}{dt^n}\right\} = p^n . S(p) - \left[p^{n-1} . s(0) + \sum_{i=2}^n p^{n-i} \cdot \left(\frac{d^{i-1} s(t)}{dt^{i-1}} \right)_{t=0^+} \right]$$

Ce résultat est une propriété intéressante : la transformée de Laplace de la dérivée d'une fonction dans le domaine temporel est égale à la multiplication de la transformée de cette fonction par p .

C'est cette propriété qui va permettre de ramener la résolution des équations différentielles à celles de polynômes.

Les conditions initiales sont souvent prises nulles.

Transformées de quelques signaux

Transformée de l'échelon unité

La transformée de Laplace de l'échantillon donne :

$$S(p) = \int_0^{+\infty} e^{-pt} dt$$

$$= \left[-\frac{1}{p} e^{-pt} \right]_0^{+\infty} = \frac{1}{p}, p > 0$$

Transformée de l'impulsion de Dirac

On reprend la définition et on fait tendre à la limite une fois l'intégrale calculée.

$$S(p) = \int_0^{+\infty} \delta(t) \cdot e^{-pt} dt$$

$$S(p) = \lim_{T \rightarrow +\infty} \int_0^T \frac{1}{T} \cdot e^{-pt} dt = \frac{1}{T} \int_0^T e^{-pt} dt = \frac{1}{T} \left[-\frac{1}{p} e^{-pt} \right]_0^T = \frac{1 - e^{-Tp}}{Tp}$$

La limite de cette fraction, quand $\varepsilon \rightarrow 0$, n'est pas évidente. La solution consiste à utiliser le développement limité de la fonction exponentielle. Changeons d'abord de variable pour simplifier l'écriture :

$$S(p) = \frac{1 - e^{-u}}{u}, u = Tp$$

Le développement de e^{-u} en $u=0$ (soit $\varepsilon=0$) est :

$$e(-u) = e(0) - ue(0) + \frac{u^2}{2!} e(0) + \dots + (-1)^n \frac{u^n}{n!} e(0) + \dots + (-1)^n u^n \varepsilon(u)$$

soit, si l'on s'arrête au 1^{er} ordre :

$$e(-u) = 1 - u + (-1)^n u^n \varepsilon(u)$$

$$S(p) = \frac{1 - e^{-u}}{u} = 1 + (-1)^n u^{n-1} \varepsilon(u)$$

donc, finalement :

$$L\{\delta(t)\} = 1$$

Ce résultat va nous permettre de calculer la réponse impulsionnelle d'un système.

Transformée d'un signal exponentiel

Soit un signal

$$s(t) = e^{at}$$

avec a constante. On a :

$$S(P) = \int_0^{+\infty} s(t) e^{-pt} dt = \int_0^{+\infty} e^{at} e^{-pt} dt = \int_0^{+\infty} e^{t(a-p)} dt = \frac{1}{a-p} \left[e^{t(a-p)} \right]_0^{+\infty}$$

$$= \frac{1}{a-p} \left[\lim_{t \rightarrow +\infty} e^{t(a-p)} - 1 \right]_0^{+\infty} = \frac{1}{p-a} \text{ si } a < p$$

b) Utilisation de la Transformée de Laplace pour la résolution des équations différentielles

L'utilisation de la transformée de Laplace pour déterminer les solutions des équations différentielles, par exemple dans le cas du 2^e degré, s'effectue de la façon suivante :

$$a \frac{d^2 s(t)}{dt^2} + b \frac{ds(t)}{dt} + cs(t) = e(t)$$

On supposera les conditions initiales nulles, c'est à dire

$$s(0^-) = \frac{ds(0^-)}{dt} = 0$$

En utilisant la transformée de Laplace, l'équation devient :

$$ap^2 S(p) + bpS(p) + cS(p) = E(p)$$

↔

$$S(p)[ap^2 + bp + cp] = E(p)$$

↔

$$\frac{S(p)}{E(p)} = \frac{1}{ap^2 + bp + cp}$$

Cette expression est appelée fonction de transfert de Laplace, ou transmittance.

Les racines (ou zéros) du polynôme du dénominateur sont également appelées pôles de la fonction de transfert.

On sait que le nombre de racines de ce polynôme dépend du signe de son déterminant. La variable p étant complexe, il y aura des racines pour les 3 cas possibles de la valeur du déterminant (négatif, nul ou positif), réelles ou complexes.

$$\frac{S(p)}{E(p)} = \frac{1}{(p-p_1)(p-p_2)} \quad (1)$$

Pour profiter des avantages de la transformée de Laplace et faciliter la résolution par rapport à l'étude directe, le principe consiste à décomposer cette transmittance du 2^e ordre en une somme de 2 transmittances du 1^{er} ordre. La TL étant linéaire, on pourra connaître la réponse temporelle du système en l'appliquant séparément à chacun des termes.

$$\frac{S(p)}{E(p)} = \frac{A_1}{(p-p_1)} + \frac{A_2}{(p-p_2)}$$

Rappels sur la décomposition d'une fraction rationnelle en éléments simples

Une fraction rationnelle est définie par un rapport de 2 polynômes. La transmittance de Laplace en constitue une.

Pour déterminer le pôle p_i , on multiplie les 2 membres de l'équation (1) par $(p-p_i)$, et on remplace p par p_i . Par exemple, pour p_1 :

$$(1) \leftrightarrow \frac{S(p)}{E(p)}(p-p_1) = \frac{1}{(p-p_2)}$$
$$\left[\frac{S(p)}{E(p)}(p-p_1) \right]_{p=p_1} = \frac{1}{(p_1-p_2)} = A_1$$

De même, on aura pour p_2 :

$$\left[\frac{S(p)}{E(p)}(p-p_2) \right]_{p=p_2} = \frac{1}{(p_2-p_1)} = A_2$$

On peut vérifier ce résultat, en remplaçant A_1 et A_2 dans l'équation (2) :

$$\frac{S(p)}{E(p)} = \frac{A_1(p-p_2) + A_2(p-p_1)}{(p-p_1)(p-p_2)} = \frac{p(A_1 + A_2) - A_1p_2 - A_2p_1}{(p-p_1)(p-p_2)}$$

On a bien

$$p(A_1 + A_2) = 0 \text{ car } A_1 = -A_2$$

et

$$-A_1p_2 - A_2p_1 = -\frac{p_2}{p_1 - p_2} - \frac{p_1}{p_2 - p_1} = \frac{p_2}{p_2 - p_1} - \frac{p_1}{p_2 - p_1} = 1$$

Ce qu'on a fait pour l'instant, c'est décomposer la transmittance du système du 2^e ordre en 2 transmittances du 1^{er} ordre.

Du point de vue des schémas-bloc, cela équivaut à une mise en parallèle de ces transmittances élémentaires, leur sortie étant combinées entre elles par un sommateur.

Dans le cas d'un pôle double, la transmittance peut s'écrire :

$$\frac{S(p)}{E(p)} = \frac{1}{(p-p_0)^2} \quad (1)$$

Cette expression est déjà décomposée en éléments simples. En effet, celle-ci s'écrirait :

$$\frac{S(p)}{E(p)} = \frac{A_1}{p-p_0} + \frac{A_2}{(p-p_0)^2} \quad (2)$$

et il est évident que l'on aurait $A_1=0$ et $A_2=1$. La situation serait différente s'il y avait un polynôme au numérateur.

Dans le cas général d'une fraction rationnelle ne comportant que des pôles simples :

$$\frac{S(p)}{E(p)} = \frac{A_1}{(p-p_1)} + \frac{A_2}{(p-p_2)} + \dots + \frac{A_n}{(p-p_n)}$$

Le coefficient A_i est une constante appelée résidu du pôle p_i . Ce résidu s'obtient par :

$$A_i = [F(p)(p-p_i)]_{p=p_i}$$

Un pôle peut être multiple. Par exemple, dans la fraction rationnelle suivante, p_2 est un pôle d'ordre 2 :

$$\frac{S(p)}{E(p)} = \frac{1}{(p-p_1)(p-p_2)^2 \dots (p-p_n)}$$

La décomposition en éléments simples donne :

$$\frac{S(p)}{E(p)} = \frac{A_1}{(p-p_1)} + \frac{A_{2,0}}{(p-p_2)^2} + \frac{A_{2,1}}{(p-p_2)} + \dots + \frac{A_n}{(p-p_n)}$$

Dans le cas général, un pôle multiple, d'ordre q , comporte q résidus :

$$A_{i,0} = [F(p)(p-p_i)^q]_{p=p_i}, \quad A_{i,1} = \frac{1}{1!} \left[\frac{d(F(p)(p-p_i)^q)}{dp} \right]_{p=p_i}, \quad A_{i,2} = \frac{1}{2!} \left[\frac{d^2(F(p)(p-p_i)^q)}{dp^2} \right]_{p=p_i}, \dots, \\ A_{i,q-1} = \frac{1}{(q-1)!} \left[\frac{d^{q-1}(F(p)(p-p_i)^q)}{dp^{q-1}} \right]_{p=p_i}$$

où i est l'indice du pôle. La formule générale est appelée formule des résidus.

L'ordre des résidus tels que formalisés ci-dessus correspond aux puissances décroissantes du dénominateur :

$$\dots + \frac{A_{i,0}}{(p-p_i)^q} + \frac{A_{i,1}}{(p-p_i)^{q-1}} + \frac{A_{i,2}}{(p-p_i)^{q-2}} + \dots + \frac{A_{i,q-1}}{p-p_i} + \dots$$

Rappel : $0! = 1! = 1$

Interprétations

Ce qu'on a fait pour l'instant, c'est décomposer la transmittance du système du 2^e ordre en 2 transmittances du 1^{er} ordre.

Du point de vue des schémas-bloc, cela équivaut à une mise en parallèle de ces transmittances élémentaires, leur sortie étant combinées entre elles par un sommateur.

Maintenant, pour connaître la réponse du système à une entrée donnée, il suffit de remplacer $E(p)$ par la transformée de Laplace de cette entrée. Connaissant $E(p)$, on peut connaître $S(p)$, et par transformée de Laplace inverse, $s(t)$, la réponse du système à cette entrée. On peut donc facilement connaître sa réponse à une impulsion de Dirac, un échelon, un signal sinusoïdal, etc.

Cette méthode peut être appliquée quel que soit l'ordre de l'équation différentielle, ce qui est un avantage par rapport à la méthode directe.

Il est intéressant de s'attarder sur le cas où l'entrée est une impulsion de Dirac. On a vu que la transmittance était définie par :

$$T(p) = \frac{S(p)}{E(p)}$$

et que la transformée de l'impulsion de Dirac était égale à 1 :

$$L\{\delta(t)\} = 1 = E(p)$$

Or, la réponse d'un système à cette entrée est appelée réponse impulsionnelle. On a vu l'importance de cette réponse, qui permet de déterminer la réponse du système à n'importe quel signal. Donc, $S(p)$ est ici la réponse impulsionnelle du système, et on a :

$$S(p) = T(p) \times E(p) = T(p) \times 1 = T(p)$$

D'où ce résultat très intéressant : **la transformée de Laplace de la réponse impulsionnelle d'un système est sa transmittance. A l'inverse, la réponse impulsionnelle d'un système est la transformée de Laplace inverse de sa transmittance.** Si l'on connaît la 2^e, on peut donc calculer la 1^{ère}.

Exemple

On reprend l'exemple du circuit RLC étudié plus haut. On cherche sa réponse à une impulsion de Dirac. Son fonctionnement est défini par l'équation différentielle :

$$\frac{d^2s(t)}{dt^2} + 2m\omega_0 \frac{ds(t)}{dt} + \omega_0^2 s(t) = e(t)$$

avec

$$\omega_0 = \frac{1}{\sqrt{LC}} \quad \text{et} \quad m = \frac{R}{2} \sqrt{\frac{C}{L}}$$

On supposera les conditions initiales nulles, c'est à dire

$$s(0^-) = \frac{ds(0^-)}{dt} = 0$$

En utilisant la transformée de Laplace, l'équation devient :

$$p^2 S(p) + 2m\omega_0 p S(p) + \omega_0^2 S(p) = E(p)$$

↔

$$S(p) [p^2 + 2m\omega_0 p + \omega_0^2] = E(p)$$

↔

$$\frac{S(p)}{E(p)} = \frac{1}{p^2 + 2m\omega_0 p + \omega_0^2}$$

Les pôles dépendent du signe de Δ , donc de m :

- $m > 0$

Les 2 pôles sont réels, définis par :

$$p_1 = -m\omega_0 + \omega_0\sqrt{m^2 - 1} \quad \text{et} \quad p_2 = -m\omega_0 - \omega_0\sqrt{m^2 - 1}$$

et donc

$$A_1 = \frac{1}{(p_1 - p_2)} = \frac{1}{2\omega_0\sqrt{m^2 - 1}} \quad \text{et} \quad A_2 = \frac{1}{(p_2 - p_1)} = -\frac{1}{2\omega_0\sqrt{m^2 - 1}} = -A_1$$

La réponse correspondante est :

$$s(t) = A_1 e^{a_1 t} + A_2 e^{a_2 t}$$

$$s(t) = \frac{1}{2\omega_0\sqrt{m^2 - 1}} (e^{p_1 t} - e^{p_2 t})$$

On peut remarquer qu'ici, tout est connu : on n'a pas un ensemble de solutions mais une seule solution. Cela vient du fait qu'on a fixé les conditions initiales.

- $m=0$

$$s(t) = (At + B)e^{-\omega_0 t}$$

- $m < 0$

Les 2 pôles sont complexes conjugués, définis par :

$$p_1 = -m\omega_0 + j\omega_0\sqrt{m^2 - 1} \quad \text{et} \quad p_2 = -m\omega_0 - j\omega_0\sqrt{m^2 - 1}$$

et donc, les coefficients

$$A_1 = \frac{1}{(p_1 - p_2)} = \frac{1}{2j\omega_0\sqrt{1 - m^2}} = \frac{-j}{2\omega_0\sqrt{1 - m^2}}$$

et

$$A_2 = \frac{1}{(p_2 - p_1)} = -\frac{1}{2j\omega_0\sqrt{1 - m^2}} = \frac{j}{2\omega_0\sqrt{1 - m^2}} = -A_1$$

sont imaginaires purs.

$$s(t) = \frac{-j}{2\omega_0\sqrt{1 - m^2}} (e^{p_1 t} - e^{p_2 t}) = \frac{e^{-j\frac{\pi}{2}}}{2\omega_0\sqrt{1 - m^2}} (e^{p_1 t} - e^{p_2 t})$$

car

$$\cos\left(-\frac{\pi}{2}\right) + j\sin\left(-\frac{\pi}{2}\right) = -j = e^{-j\frac{\pi}{2}}$$

donc

$$s(t) = \frac{1}{2\omega_0\sqrt{1 - m^2}} \left(e^{p_1 t - j\frac{\pi}{2}} - e^{p_2 t - j\frac{\pi}{2}} \right) = \frac{1}{2\omega_0\sqrt{1 - m^2}} \left(e^{(-m\omega_0 + j\omega_0\sqrt{m^2 - 1})t - j\frac{\pi}{2}} - e^{(-m\omega_0 - j\omega_0\sqrt{m^2 - 1})t - j\frac{\pi}{2}} \right)$$

$$= \frac{1}{2\omega_0\sqrt{1 - m^2}} \left(e^{(-m\omega_0 + j\omega_0\sqrt{m^2 - 1})t} - e^{(-m\omega_0 - j\omega_0\sqrt{m^2 - 1})t} \right) = \frac{1}{2\omega_0\sqrt{1 - m^2}} e^{-m\omega_0 t} \left(e^{(j\omega_0\sqrt{m^2 - 1})t - j\frac{\pi}{2}} - e^{(-j\omega_0\sqrt{m^2 - 1})t - j\frac{\pi}{2}} \right)$$

et finalement :

$$s(t) = \frac{1}{\omega_0\sqrt{1 - m^2}} e^{-m\omega_0 t} \cos\left(\left(\omega_0\sqrt{m^2 - 1}\right)t - \frac{\pi}{2}\right)$$

La solution est un signal sinusoïdal pondéré (enveloppé) par une exponentielle, c'est à dire amorti.

c) Cas particulier du régime sinusoïdal

Dans tout ce qui a été vu précédemment, le signal d'entrée des systèmes pouvaot être quelconque. Dans le cas particulier de signal sinusoïdal, on peut remplacer la variable de Laplace p par $j\omega$, où ω représente la pulsation ($=2\pi f$). La transmittance, alors appelée fonction de transfert harmonique, devient une fonction de la pulsation, ou de la fréquence, ce qui est équivalent (à une constante près).

Par exemple, dans le cas d'un circuit RC en régime sinusoïdal, la transmittance de Laplace :

$$T(p) = \frac{S(p)}{E(p)} = \frac{1}{1 + RCp}$$

devient

$$T(j\omega) = \frac{S(j\omega)}{E(j\omega)} = \frac{1}{1 + jRC\omega}$$

En général, les fonctions de transfert harmoniques sont étudiées dans le plan de Bode, l'étude portant séparément sur le module de $T(j\omega)$ et sur son argument.

On peut remarquer que dans le cas des systèmes électroniques, les fonctions de transfert harmoniques peuvent s'obtenir à partir des impédances complexes. Dans l'exemple du circuit RC, l'impédance complexe du condensateur est définie par $1/jC\omega$. Les grandeurs d'entrée et de sortie sont des tensions, et R et C forment un pont diviseur de tension. La tension de sortie est donc définie par :

$$\frac{S(j\omega)}{E(j\omega)} = \frac{Z_C}{R + Z_C} = \frac{1/jC\omega}{R + 1/jC\omega} = \frac{1}{1 + jRC\omega}$$