MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

OPTION: ACQUISITION ET TRAITEMENT DE L'INFORMATION

THESE

Pour l'obtention du grade de Magister
Présentée par Melle GUERAICHI RATIBA

SUJET

ELABORATION DE DICTIONNAIRES POUR LA QUANTIFICATION POLAIRE

Soutenue le 24 Janvier 1995 devant le jury composé de:

Mr B. DERRAS	*Maitre de conference*	*Président*
Mr D. BERKANI	*Maitre de conference*	*Rapporteur*
Mr Z TERRA	*Chargé de cours*	*Examinateur*
Mr B. BOUSSEKSOU	*Chargé de cours*	*Examinateur*
MR L. SAADAOUI	*Chargé de cours*	*Examinateur*

DEDICACES

Je dédie ce modeste travail:

A mes parents

A mes frères et sœurs

A mes neveux Sofiane, Riyad et Merouane

A mes nièces Hayat et Assia

A mon amie Souad

A toutes mes amies.

REMERCIEMENTS

Je tiens à exprmer mes vifs remerciements et ma profonde gratitude à Mr A. CHEKIMA professeur à l'ENP de m'avoir aider à accomplir cette thèse par ses recommandations, ses orientations et ses conseils tout au long de ce travail malgré ses nombreuses occupations.

J'exprime également ma vive gratitude et mes remerciements à Mr D. BERKANI maître de conférence à l'ENP pour son aide, son soutien moral, son amabilité et ses encouragements. Qu'il trouve ici l'expression de ma sincère reconnaissance et de mon profond respect.

Je remercie Mr B. DERRAS de m'avoir fait l'honneur de présider le jury.

Je remercie Mr P.F. SWASZEK professeur agrégé à l'université de Rhode Island New-Jersey, USA, pour la documentation qu'il m'a fournie ainsi que ses explications fructueuses et sa disponibilité même lors de ses occupations.

Enfin, je remercie tous ceux qui ont contribué de près ou de loin à la préparation de cette thèse notamment l'équipe du laboratoire de robotique du HCR pour leur aide et leur gentillesse.

SOMMAIRE

INTRODUCTION

Le traitement et la transmission des signaux numériques ont un rôle déterminant dans les systèmes de communication modernes.

Néanmoins, les origines physiques des signaux informationnels tels que la parole ou l'image sont intrinséquement analogiques et par nature temporellement continus.

Ainsi, la numérisation de tels signaux est éxigée. Celle-ci génère une grande quantité de données que les réseaux de communication et les équipements de traitement, actuellement disponibles, sont incapables de transmettre ou de manipuler.

Afin de résoudre cet important problème, des techniques de compression de données sont mises en oeuvre. L'une de ces techniques la plus intéressante et la plus utilisée est l'opération de quantification.

Celle-ci se fait sur des échantillons analogiques c'est à dire ayant des amplitudes continues. Elle réalise l'opération de discrétisation ou quantification des valeurs échantillonnées ou prélevées du signal analogique. Elle convertit ainsi des variables réelles considerées comme un alphabet infini en variables discrètes ou bien un alphabet fini. Ce qui introduit donc une erreur entre l'entrée et la sortie du quantificateur.

A cet effet, le problème de la performance du quantificateur a toujours été posé. C'est à dire, comment obtenir à la sortie, une approximation de la source en la représentant par un nombre de bits limité.

Aussi, beaucoup de chercheurs se sont attelés à résoudre cette difficulté [1, 2, 3, 6, 8, 9, 10, 12, 13, 15, 17, 23, 24, 25, 29, 31, 32, 42]. Ainsi, de nombreux procédés et algorithmes ont été mis au point dans le domaine de la quantification vectorielle [6, 7, 8, 10, 12, 16, 17, 24, 28, 29, 31, 33, 34, 38, 39, 41] ; principalement dans les cas de la quantification statistique et algébrique (où on utilise les réseaux réguliers) [24, 31, 41]. Cependant le premier type de quantification (statistique) demande, lors de l'exécution des algorithmes qui lui sont relatifs, un grand espace mémoire. Pour le second type, les algorithmes perdent de leur éfficacité dés lors que la source ne présente plus une distribution uniforme; ce qui est généralement le cas [9, 24].

C'est pourquoi, nous nous sommes intéressés à un autre aspect de la

quantification qui est la quantification sous-optimale appliquée à un système polaire qui ne donne pas d'aussi bons résultats obtenus par les deux précédentes méthodes, mais ils sont toutefois acceptables. En revanche, elle a l'avantage de présenter (dans le système polaire) le signal sous deux formes: en module et en phase laquelle est uniformément distribuée, donc peut-être quantifiée d'une façon optimale .En ce qui concerne le module, celui-ci est quantifié indépendemment de la phase, grace à des algorithmes simples à mettre en oeuvre.

Ainsi, notre travail a consisté à élaborer dans le systeme polaire, divers algorithmes qui nous permettaient d'avoir des structures et des constellations différentes. A chaque algorithme, nous avons essayer de s'approcher le plus possible de la structure de Dirichlet afin d'améliorer nos résultats.

La thèse s'articule autour de sept chapitres:

Le premier chapitre est consacré à l'étude de la théorie de la distorsion ainsi que des rappels mathématiques y inférent. Dans le deuxième chapitre, nous parlerons du principe de la quantification vectorielle ainsi qu'une retrospective sur les techniques de ce type de quantification connues jusqu'à présent, et d'un petit aperçu sur la quantification scalaire. Le troisième, quatrième, cinquième et sixième chapitres décrivent respectivement les quantificateurs polaires SPQ et UPQ, AUPQ, DPQ et DRPQ, et les algorithmes qui en découlent. Enfin, les commentaires sur les propriétés de notre quantificateur et sur les résultats obtenus sont dans la dernier chapitre de notre travail.

CHAPITRE 1

RAPPELS SUR LA THEORIE DE LA DISTORSION

Introduction:

La théorie de la distorsion occupe une place importante dans le domaine de la théorie de l'information. C'est l'étude théorique sur la meilleure façon de représenter la séquence aléatoire de la source. Autrement dit, on peut dire que l'on a, d'un côté une source X, de l'autre un alphabet fini; le but est de trouver la meilleure "table de correspondance" entre la source et cet alphabet fini, souvent appelé " *alphabet de reproduction*".

Cette dicipline a pris la dénomination de "la théorie de la distorsion" ou de "Rate distorsion theory" [10, 11] . On peut dire que c'est Shannon [10, 11], qui a énoncé les principes de base de cette théorie.

Le but essentiel de cette partie est de montrer que la fonction de la distorsion est l'expression mathématique de la meilleure performance que peut atteindre un quantificateur.

1.1. Cas discret sans memoire:

1.1.1. Source et Entropie:

Considerons une source X, discrete, stationnaire et ergodique sans mémoire. X est définie par le fait que chaque échantillon temporel x(i) est indépendant des échantillons x(i-1), x(i-2) ... et qu'il ne peut prendre que N valeurs différentes dans l'alphabet A={ x_1, x_2, ..., x_N }. x_1, x_2, ..., x_N sont appelés "lettres" de l'alphabet A.

Chaque échantillon possible est associé à une probabilité: $p(X(i)=x_i)= p(x_i) = p_i$. Cette probabilité est associée à une " auto-information" ou surprise $I(x_i)$ définie par:

$$I(x_i) = -\log p_i \qquad (1.1)$$

Plus un événement est rare ($p_i \rightarrow 0$), plus la surprise qui lui est associée est grande ($I(x_i)) \rightarrow \infty$). Chaque lettre de la source a donc une surprise propre. La surprise moyenne de la source est appelée *'entropie'* de la source:

$$H(x) = E(I(x)) = - \sum_{i=1}^{N} p_i \log p_i \qquad (1.2)$$

L'entropie mesure l'information moyenne d'une source X. Pour une source X donnée dont l'alphabet comporte N lettres, on peut montrer que:

$$0 \leq H(x) \leq \log N \qquad (1.3)$$

Il existe deux valeurs particulières pour l'entropie H(x):

- H(x) = 0 si et seulement si tous les éléments de l'alphabet X ont une probabilité nulle sauf un , qui possède une probabilité égale à 1. Cette source est totalement prédictible (surprise nulle).

- H(x) = log N si et seulement si quelque soit i compris entre 1 et N, $p_i = 1/N$. Toutes les lettres de l'alphabet ont la même probabilité: la source est totalement non-prédictible.

La double inégalité (1.3) peut se traduire comme suit: l'entropie d'une source discrète est toujours positive, et la source d'entropie maximale est la source dont les éléments ont une probabilité uniforme.

L'entropie et l'information ont une signification physique: ces deux grandeurs possèdent une unité qui dépend de la base utilisée pour le logarithme.

* Pour la base e, l'unité est le nat.
* Pour la base 2, l'unité est le bit.

Exemple:

Une source dont l'alphabet est {0, 1, 2, ..., 7}, chaque lettre ayant une probabilité uniforme, a une entropie:

$$H(x) = \sum_{i=1}^{8} \frac{1}{8} \log_2 \frac{1}{8} = \log_2 8$$

En base 2 (binaire) cette source a une entropie de 3 bits ($\log_2 8 = 3$)

Remarque:

Les entropies sont donc toujours définies à une constante multiplicative près, qui dépend de la base du logarithme:

$$\log_a b = \log_a c \, . \, \log_c b.$$

Exemple: H en nat= $\log_e 2$ x H en bit = 0.69 x H en bit.

L'entropie définie telle qu'en (1.2) est aisément généralisable à des variables vectorielles. La probabilité est alors la probabilité conjointe des composantes du vecteur.

On peut aussi définir une entropie conditionnelle: déduite de la notion de densité de probabilité conditionnelle, elle exprime l'informaton d'une source connaissant une autre source.

Ces différentes définitions peuvent se résumer par le schéma ci-dessous, qui permet en outre de dégager les principales propriétés de l'entropie:

Soient X et $\hat{X}$ deux sources schématisées par le diagramme de Venn suivant:

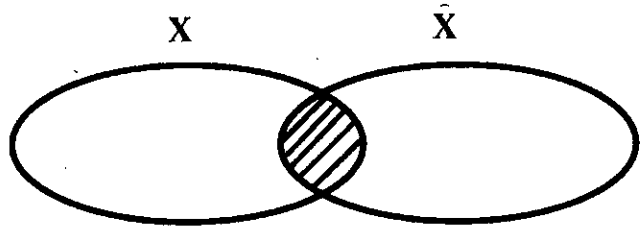


Figure 1.1: Variable à information commune

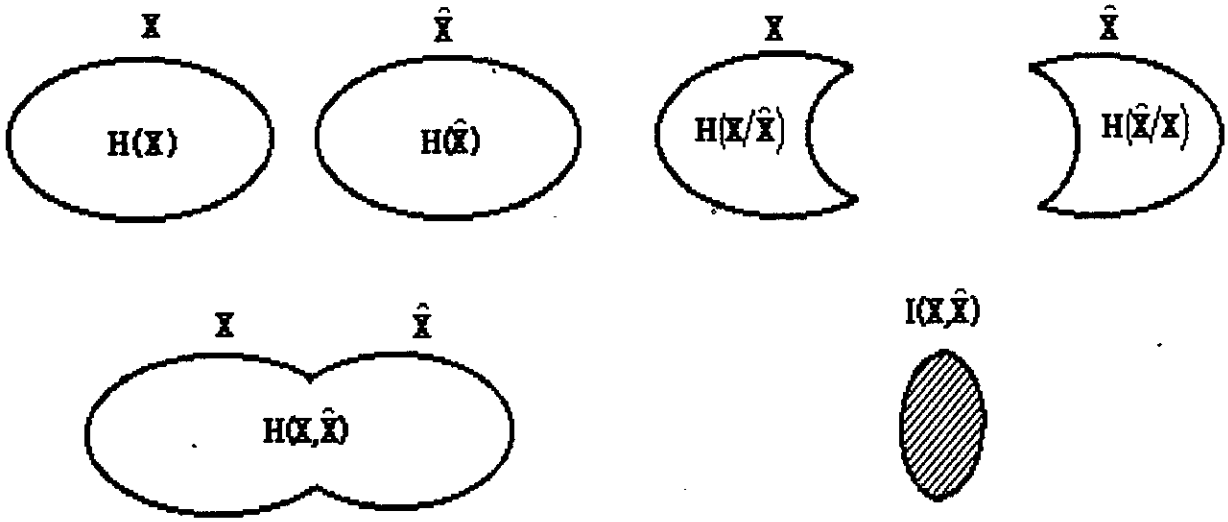


Figure 1.2: Entropie conditionnelle et information mutuelle

La partie hachurée s'appelle *"l'information mutuelle"* et est notée $I(X, \hat{X})$. Les propriétés de I sont évidentes à partir du schéma:

L'information mutuelle moyenne peut s'exprimer au moyen des entropies et entropies conditionnelles comme suit [41]:

$$
\begin{aligned}
I(X, \hat{X}) &= H(X) + H(\hat{X}) - H(X, \hat{X}) \\
&= H(\hat{X}) - H(X/\hat{X}) \\
&= H(\hat{X}) - H(\hat{X}/X) \\
&= I(\hat{X}, X)
\end{aligned}
\qquad (1.4)
$$

Si X et $\hat{X}$ sont indépendants, on a $I(X, \hat{X}) = 0$. La notion d'information mutuelle joue un rôle capital en théorie de la distorsion.

1.1.2. La courbe R(D):

Rappelons le schéma de notre quantificateur:

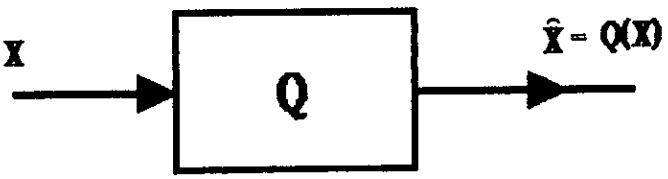


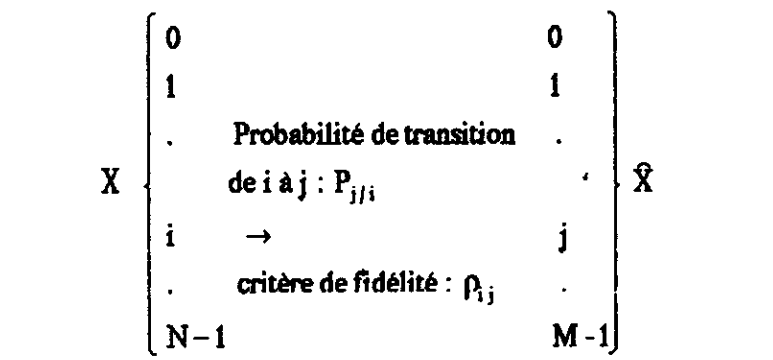
Figure 1.3. Quantificateur type

Ce quantificateur peut-être vu comme un canal de transmission qui corrompt X. Pour plus de simplicité, nous allons envisager le cas de la "transmission" d'une seule lettre. Soit X une source dont l'alphabet est $A = \{ 0, 1, \ldots, N-1 \}$, la lettre i étant associée à une probabilité p_i. Le quantificateur Q va être constitué en une table de correspondance qui à une lettre de la source X va associer une lettre de la reproduction $\hat{A}$, constituée des éléments $\{ 0, 1, \ldots, M-1 \}$.

Q est donc un tableau de taille (N x M). Comme en pratique $M \leq N$, Q ne peut être une correspondance biunivoque. On préfère donc en théorie de la distorsion passer du modèle de Q comme fonction (Q étant l'approche naturelle lorsqu'on notait $\hat{X} = Q(X)$), à un modèle de Q comme probabilité conditionnelle de $\hat{X}$ connaissant X . Cette approche est plus proche

d'une conception de Q comme canal de transmission entre X et $\hat{X}$ plutôt qu'à une " quantification" de X en $\hat{X}$. Néanmoins, elle permet de travailler en toute généralité sur la façon dont on passe de X à $\hat{X}$.

Notre quantificateur peut donc se voir comme ceci [41]:



Sur l'exemple ci-dessus, la lettre i de X est reproduite par la lettre j de $\hat{X}$ avec une probabilité $P_{j/i}$. A cette probabilité $P_{j/i}$ est associé un critère de fidélité ρ_{ij} qui estime l'erreur commise en choisissant la lettre j pour représenter la lettre i.

Notre quantificateur est donc défini par la donnée de la matrice:

$$P = \left[P_{j/i} \right]_{\substack{i \in [0, N-1] \\ j \in [0, M-1]}} \qquad (1.5)$$

Tous les éléments de cette matrice sont positifs ou nuls. La somme de tous les éléments d'une ligne vaut 1 (Il s'agit d'une probabilité). Si un seul élément par ligne de la matrice est non nul (auquel cas il vaut 1), le codeur est déterministe.

Le problème fondamental de la théorie de la distorsion est de trouver, pour un critère de fidélité ρ, et une distorsion D donnée, la meilleure densité de probabilité P.

Considérons la classe P_D des quantificateurs Q tels que $d(P) \leq D$:

$$P_D = \{ P / \ d(P) \leq D \}$$

A chaque quantificateur Q, on associe I(P), qui est l'information mutuelle entre X et $\hat{X}$ (entrée et sortie de Q).

$$I(P) = I(X, \hat{X}) = H(X) - H(X/\hat{X}) = \sum_{i,j} P_{j/i} \log_2 \frac{P_{j/i}}{P_j} \qquad (1.7)$$

où $P_j = \sum_i P_i P_{j/i}$

Le résultat central de la théorie de la distorsion est alors:

* pour une distorsion D donnée, le plus petit débit que peut avoir le codeur d'une source X est [41] :

$$R(D) = \min_{P \in P_D} (I(P)) \qquad (1.8)$$

La figure (1.5) illustre une courbe R(D) typique pour le cas discret.

R(D) possèdent des propriétés interressantes et qui sont les suivantes [31, 41]:

- R(D) est convexe U
- $R(0) = H(X) = - \sum_i p_i \log_2 p_i$
- $D_{max} = D(R=0) = \min_j \left(\sum_i p_i \rho_{ij} \right)$
- D_{max} existe toujours.

Il existent diverses techniques, qui permettent de trouver analytiquement R(D) dans le cas discret pour une source X donnée. Celles ci ont été etudiées par Berger [42] , et Blahut [43].

L'équation (1.8) définit une fonction R(D) qui donne pour une distorsion donnée, un débit minimale ou réciproquement pour un débit donné, une distorsion la plus faible possible. Ce qui nous ramène à parler du codage d'une source ou d'un vecteur.

Definition:

On appelle un code C de taille N et de longueur de bloc k,un ensemble de N séquences $\{ y_0, y_1, ..., y_{N-1} \}$ de dimension k chacune, c'est à dire que chaque séquence est un vecteur à k dimensions; elle s'écrit donc:

$$y_i = [\ y_{i1}, y_{i2}, ..., y_{ik}]$$

Chaque séquence de C est appelée mot de code qui est une séquence binaire de longueur $\log_2(N)$ au moins. Pour coder un vecteur X dans C, il faudra choisir le mot de code qui minimise la distorsion $\rho(x, y)$, notée:

$$\rho_k(x/C) = \min_{y \in C} \rho_k(x, y) \qquad (1.9)$$

On définit ainsi la distorsion moyenne:

$$\rho(C) = E\left[\rho_k(x/C)\right] = \sum_x P(x)\, \rho_k(x/C) \qquad (1.10)$$

Avec: $$P(x) = \prod_{i=0}^{k-1} P_{x_i}(x_i)$$

Pour un vecteur X de k échantillons, on définit le débit par échantillon du code C par la relation:

$$R = \frac{\log_2 N}{k} \text{ bits} \qquad (1.11)$$

La théorie sur le codage nous apprend que R(D) est une limite théorique des performances des systèmes de codage, et qu'il est possible d'atteindre une performance assez proche de cette limite; ce qui est dicté par l'équation suivante [31]:

$$\lim_{k \to \infty} \frac{1}{k} \log_2 N = R(D) \qquad (1.12)$$

Nous allons maintenant envisager le cas des sources continues pour lesquelles bon nombre des résultats précédents se généralisent.

1.2. Cas continu sans mémoire:

1.2.1 Definitions :

a/ La source:

Le cas d'une source continue est la généralisation du cas discret. La source est maitenant stationnaire , ergodique, sans mémoire, à alphabet continu, mais toujours à temps discret. Les notions introduites précédemment se généralisent aisément, à quelques détails près, comme nous allons le voir.

Les échantillons délivrés par la source X sont toujours indépendants statistiquement; mais au lieu d'appartenir à un alphabet discret, ils appartiennent à un "alphabet continu".

Les probabilites discrètes se généralisent aux densités de probabilités. La source est donc supposée produire des échantillons qui obéissent à une densité de probabilité p.

b/ L'entropie differentielle:

L'entropie généralisée à partir de la définition dans le cas discret est infinie [41]. Cela peut intuitivement se comprendre: alors qu'il faut un nombre fini de questions de type oui-non pour connaitre un élément d'un alphabet fini; il faut élucider un nombre infini de questions pour connaître entièrement un nombre réel. De façon plus formelle, reprenons la définition de l'entropie dans le cas discret.

$$H_d(p) = - \sum_j p_j \log p_j$$

Voyons de quelle façon nous pourrions étendre cette définition au cas où p est une densité de probabilité continue. Une méthode classique utilisée pour définir l'integrale de Riemann (l'integrale classique) consiste à approximer p par une fonction en escalier (discrète), comme le suggère la figure 1.4.

p est "échantillonnée" avec un pas Δx, et sa valeur en un pas centré en x_j est $p_j = p(x_j)$. Δx . L'entropie de cette nouvelle densité de probabilité s'écrirait:

$$H_{\Delta x}(p) = -\sum_i p(x_i)\,\Delta x \log\left(p(x_i)\,\Delta x\right) \qquad (1.13)$$
$$= -\sum_i p(x_i) \log\left(p(x_i)\right)\Delta x - \log(\Delta x)$$

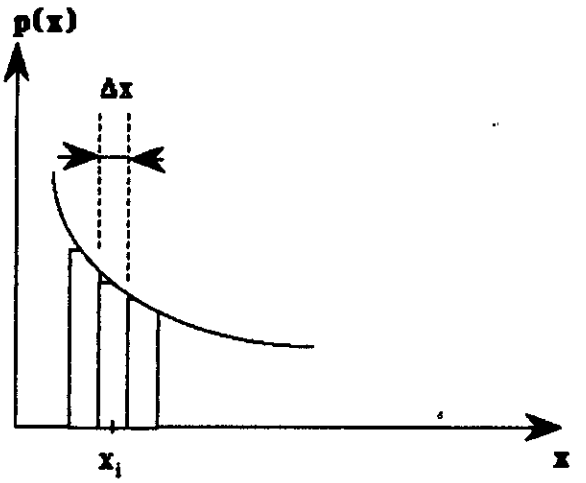


Figure 1.4: Approximation discrète de p(x)

En faisant tendre Δx vers zéro, pour obtenir l'entropie de p, on trouve les expressions suivantes pour les deux termes de (1.13).

Premier terme:

$$\lim_{\Delta x \to 0}\left[-\sum_j p(x_j)\log p(x_j)\,\Delta x\right] = -\int p(x)\log p(x)\,dx$$

(Si cette intégrale existe)

Deuxième terme:

$$\lim_{\Delta x \to 0} [-\log \Delta x] = +\infty$$

Seul le premier terme converge: on ne peut pas généraliser directement le concept d'entropie du cas discret. On définit donc une entropie differentielle h(p) composée de la limite du premier terme de (1.13).

$$h(p) = - \int p \log (p) \, dp$$

Cette entropie va être beaucoup moins "significative" que son homologue du cas discret. Son interprétation va demander quelques précautions que nous allons découvrir en énumérant quelques unes des propriétés de l'entropie différentielle.

* L'entropie différentielle peut être positive ou négative, contrairement au cas discret.

* L'entropie différentielle dépend du repère dans laquelle elle est exprimée. Cette propriété est fondamentale dans le cas d'un changement de variable.

Soient $x = (x_1 , x_2 , \ldots , x_N)$ et $y = (y_1 , \ldots , y_N)$ deux vecteurs aléatoires de densité de probabilité conjointes respectives $p_X (x_1 , x_2 , \ldots , x_N)$ et p_Y, y étant lié à x par une transformation f

$$y = f(x)$$

plus précisément: $y = (y_1 , \ldots , y_N) = (f_1(x), f_2(x), \ldots, f_N(x))$

$$|J_f| = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & & & \\ \vdots & & & \\ \frac{\partial f_N}{\partial x_1} & \cdots & \cdots & \frac{\partial f_N}{\partial x_N} \end{vmatrix}$$

Sachant par (1.14) comment sont liés p_X et p_Y, on en déduit le lien entre $h(p_X)$ et $h(p_Y)$:

$$h(p_Y) = h(p_X) + E(\log |J_f|) \qquad (1.15)$$

E étant l'espérance mathématique.

L'expression (1.15) expose une caractéristique fondamentale de l'entropie différentielle: elle ne peut servir de mesure absolue de l'information, puisqu'elle dépend du repère où on la calcule. Pour effectuer cette mise en ordre, plusieurs auteurs [42, 43, 44] ont cherché quelles densités de probabilités rendant maximales l'entropie suivant différents critères. Ils ont trouvé que la densité qui maximise h est la suivante:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

Le paramètre m désigne le moment du premier ordre ou moyenne, p est donc une loi normale. Donc, les variables gaussiennes sont des variables à entropie maximale. Ce sont donc, les variables les "pires" à coder.

L'entropie différentielle d'une variable gaussienne est:

$$h(x) = \log_2 \sqrt{2\pi e\sigma^2} \qquad (1.16)$$

Nous pouvons donc ordonner les densités de probabilités par rapport à la loi normale. Pour cela, on définit la puissance d'entropie Q_h d'une variable x comme:

$$Q_h = \frac{1}{2\pi e} 2^{2.h(x)} \qquad \text{h en bits} \qquad (1.17)$$

Q_h est la variance d'une source gaussienne qui aurait une entropie h. Q_h est donc comprise entre zéro et un. Voici quelques résultats connus [41] :

densité deprobabilité	h(x)	Q_h / σ^2
Uniforme	$\frac{1}{2} \log_2 12 \sigma^2$	0. 703
Gaussienne	$\frac{1}{2} \log_2 2 \pi e \sigma^2$	1
Laplace	$\frac{1}{2} \log_2 2 e^2 \sigma^2$	0. 865
Gamma	$\frac{1}{2} \log_2 4\pi e - e \frac{\sigma^2}{3}$	0. 874

c/ Information mutuelle:

On définit l'information mutuelle entre deux variables comme dans le cas discret [31]:

$$\begin{aligned} I(X, Y) &= h(X) + h(Y) - h(X, Y) \\ &= h(X) - h(X/Y) \\ &= h(Y) - h(Y/X) \end{aligned}$$

où les différentes entropies différentielles données par l'expression ci-dessus sont les suivantes:

- L'entropie différentielle relative à X est:

$$h(X) = - \int p_X(x) \log_2 p_X(x)\, dx \qquad (1.18)$$

- L'entropie conditionnelle est définie comme suit:

$$h(X/Y) = - \iint p_{XY}(x, y) \log_2 p_{X/Y}(x/y)\, dx\, dy \qquad (1.19)$$

L'information mutuelle est alors:

$$I(X; Y) = \iint p_{XY}(x, y) \log_2 \frac{p_{XY}(x, y)}{p_X(x)\, p_Y(y)}\, dx\, dy \qquad (1.20)$$

A partir des relations précédentes, si l'on considère que $\rho(x,y)$ est une mesure de distorsion et $p_{Y/X}(y/x)$ la densité de probabilité conditionnelle, on définit la distorsion moyenne suivante:

$$d(p_{Y/X}) = \iint p_X(x)\, p_{Y/X}(y/x)\, \rho(x,y)\, dx\, dy \qquad (1.21)$$

L'information mutuelle moyenne s'écrit:

$$I(p_{Y/X}) = \iint p_X(x)\, p_{Y/X}(y/x) \log_2 \frac{p_{Y/X}(y/x)}{p_Y(y)}\, dx\, dy \qquad (1.22)$$

d/ Courbe R(D):

On obtient finalement comme pour le cas discret, la limite théorique des performances ou la fonction R(D) [42] :

$$R(D) = \min_{p_{Y/X} \in F_D} I(p_{Y/X}) \tag{1.23}$$

où

$$F_D = \{ p_{Y/X} : d(p_{Y/X}) \leq D \} \tag{1.24}$$

Les proprietés de R(D) pour des sources continues, sont les mêmes que celles des sources discrètes [31], à savoir que :

- R(D) est une fonction convexe sur l'intervalle [0, Dmax [.
- Dmax est la distorsion maximale; elle existe toujours.
- La valeur extrêmale Dmax pour ce cas est donnée par:

$$D_{max} = \min_{y} \int p_X(x)\, \rho(x, y)\, dx \tag{1.25}$$

- R(D) est une fonction continue, monotone et décroissante.

La différence qui existe entre les sources discrètes et continues, (en ce qui concerne R(D)) est au niveau de l'origine de R(D), comme on peut le voir sur la figure 1.5.

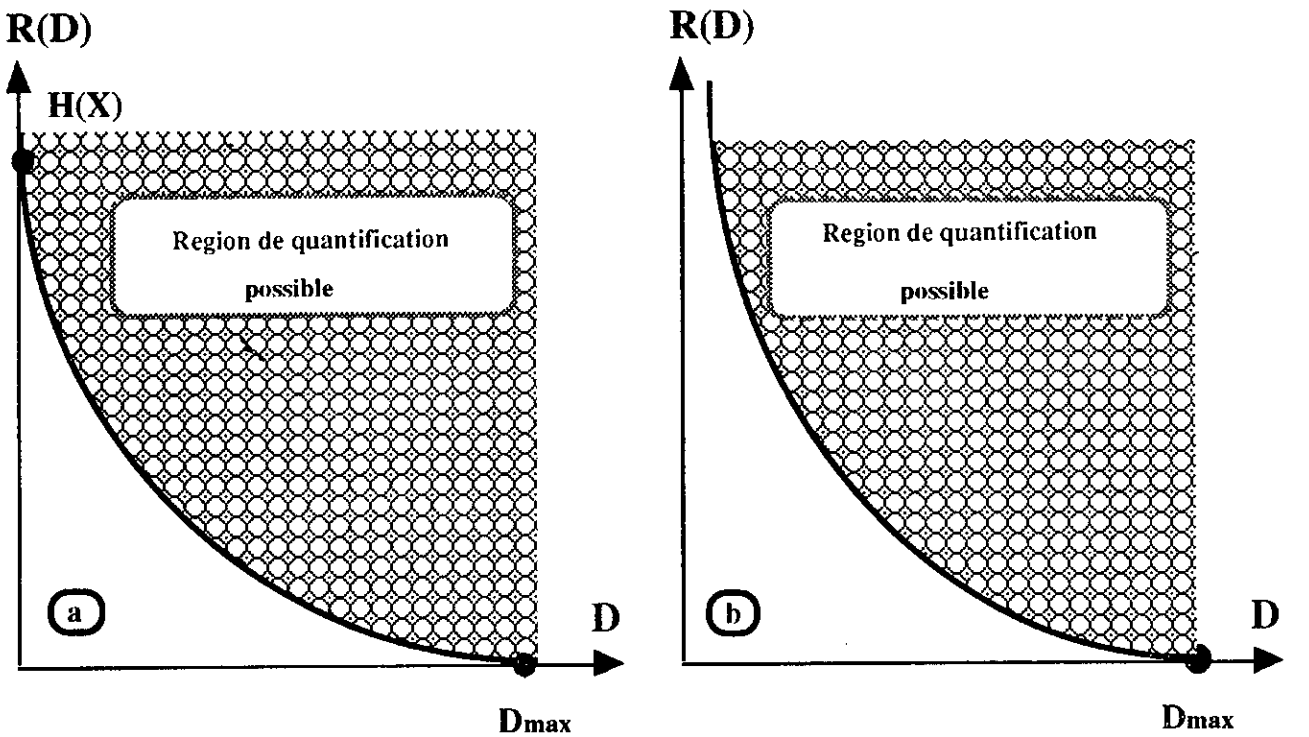


Figure 1.5: Comparaison de la fonction R(D) dans les cas:
a/ d'une source discrète
b/ d'une source continue

Le choix du critère de fidélité est très important lors de la conception d'un quantificateur. Dans la plupart des cas, le choix des chercheurs s'est porté sur l'erreur quadratique moyenne, notée aussi MSE (*mean-squared error*).

1.2.2. Exemple:

Si on considère la courbe (1.5.6) pour une loi gaussienne, on trouve que [41] :

$$D_{max} = \sigma^2 .$$

Ce résultat est tout à fait logique, car à un débit nul, il est normal de trouver une distorsion égale à l'energie du signal.

Comme nous l'avons vu précédemment, la loi normale a une entropie maximale et à ce titre, est la pire à coder. Ceci se résume par le théorème de Berger suivant [41]:

Quelle que soit la densité de probabilité, p, de moyenne nulle et de variance σ^2 , elle vérifie:

$$R(D) < 1/2. (\log (\sigma^2 / D))$$

avec égalité si et seulement si p suit une loi normale.

Pour une distorsion donnée, les variables gaussiennes vont demander le plus gros débit. Ce théorème peut s'exprimer de façon analogue par;

Quelle que soit la variable aléatoire X, la courbe D(R) (fonction inverse de R(D)) qui lui est associée vérifie [41]:

$$D_X(R) \geq D_X^{(L)}(R) = (1/2\pi e) . e^{2(h(x) - R)} \qquad R \geq 0$$

$D_X^{(L)}(R)$ est appelé limite inferieure de Shannon (le L entre parenthèses signifie "lower").

En prenant l'exemple d'une source gaussienne de variance σ^2 la fonction de distorsion s'écrit [41]:

$$R(D) = \begin{cases} \frac{1}{2}\log_2\frac{\sigma_x^2}{D} & 0 \le D \le \sigma_x^2 \\ 0 & D \ge \sigma_x^2 \end{cases} \qquad (1.26)$$

Quand la source suit une loi de Gauss (comme dans l'exemple). la fonction R(D) et la limite inferieure de Shannon notée SLB (*Shannon Lower bound*), forme la même expression, quand l'erreur quadratique moyenne est choisie comme critère de fidélité [31].

Remarque:

Dans le cas d'une source bidimensionnelle et en supposant que les deux composantes de cette source sont indépendantes statistiquement; la fonction de distorsion de cette source est donnée par [31] :

$$D(R) = \frac{1}{2}\left(D_1(R_1) + D_2(R_2)\right) \qquad (1.27)$$

R_1 et R_2 sont les debits des sources.

CHAPITRE 2

GENERALITES SUR LA QUANTIFICATION VECTORIELLE

INTRODUCTION:

Ce chapitre a pour objet de présenter les fondements de la quantification. On verra le principe de la quantification scalaire, puis vectorielle, où on montrera l'avantage du second par rapport au premier.

En dernier lieu, il sera question de faire une rétrospective de la quantification vectorielle optimale, en notant ses avantages et ses inconvenients. Enfin, on introduira la notion de quantification sous-optimale qu'on développera dans les chapitres suivants.

2.1. Notions fondamentales sur la quantification:

2.1.1. Principe:

La quantification est l'opération, qui consiste à remplacer une grandeur exacte d'entrée par une valeur choisie parmi un nombre fini de valeurs possibles. On substituera donc, une infinité de valeurs de la grandeur analogique x en un nombre fini de valeurs y_i de ce signal.
Il en résulte une erreur e, dûe à cette approximation et appelée bruit de quantification.

2.1.2. Quantification Scalaire:

La quantification scalaire est l'operation de quantification, appliquée à un signal unidimensionnel.

Un dispositif qui fait associer l'entrée X à la sortie Y, est appelé un quantificateur scalaire, s'il existe des constantes Xet Y tellesque:

$$X_{i-1} \leq X < X_i \quad \Rightarrow \quad Y = Y_i \qquad (2.1)$$

-Les valeurs **X** sont appelés seuils de quantification
-Les valeurs **Y** sont les niveaux de reconstruction

Ainsi, la région où se fait la quantification est un intervalle de données. La fonction de transfert correspondante est une caracteristique en marches d'escalier. Elle présente deux formes différentes à l'origine, selon que le nombre de niveaux de reconstruction N est pair ou impair (Figures 2.1.1, 2.1.2) [30].

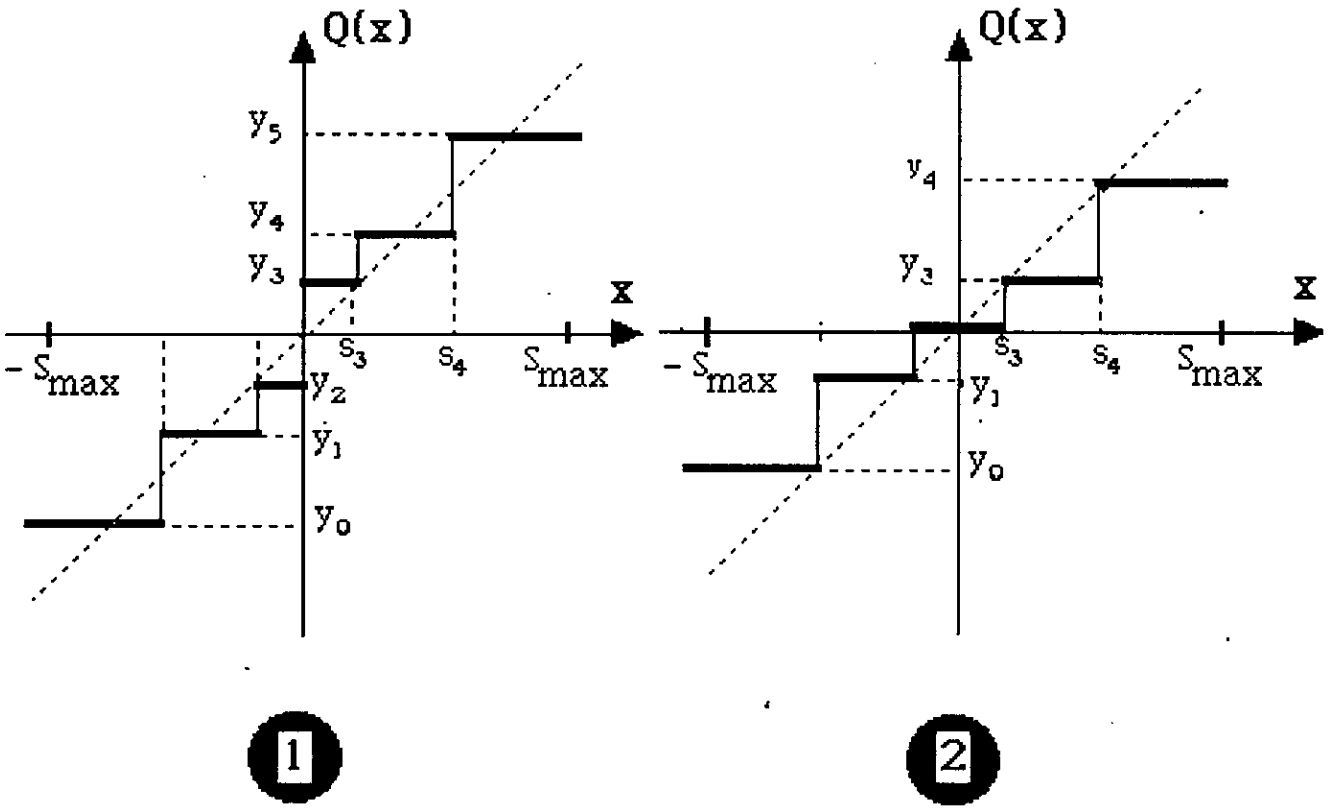


Figure 2.1 : Caracteristique d'un quantificateur scalaire pourun nombre de niveaux pair (1) et impair (2).

Les seuils et les niveaux optimaux ont été tabulés par Max [1]

2.1.3. Quantification Vectorielle:

Contraîrement à la quantification scalaire, la quantification vectorielle s'applique sur des vecteurs.

Ainsi, un quantificateur vectoriel (QV) fait correspondre à tout vecteur **X** décrit comme suit : $x = (x(0), x(1) , ... , x(k-1))$ un vecteur $y_i = (y_i(0) , y_i(1) , ... , y_i(k-1))$ parmi un ensemble fini **B** de N vecteurs de reproduction [31. Cet ensemble B est appelé dictionnaire.

L'ensemble B est décrit de la façon suivante:

$$B = \{ y_i \mid i = 0, \ldots, N\text{-}1 \} \qquad (2.\ 2)$$

On peut alors écrire la fonction de quantification sous la forme:

$$y_i = Q(x) \qquad \text{où} \quad i \in \{ 0, \ldots, N\text{-}1 \} \qquad (2.\ 3)$$

où x et y appartienent respectivement à l'espace euclidien $\mathcal{R}^k$ et à un sous-ensemble fini de $\mathcal{R}^k$.

Un quantificateur fournit une seule information à la sortie, mais celle-ci est disponible sous deux formes : la valeur y et l'indice i (figure 2-2).

Figure 2.2 : Schema type d'un quantificateur vectoriel

Ainsi, pour généraliser la notion de quantification à plus d'une variable, il est préférable de ne plus recourir au concept de seuil mais d'approcher le problème sous le seul angle de la table de décodage. La quantification devient alors une recherche de l'indice entre 0 et N-1 pour lequel la table donne la meilleure approximation; c'est à dire l'erreur de quantification la plus faible possible [24].

La figure 2.3 illustre un exemple de quantification vectorielle d'un vecteur d'entrée à deux dimensions. L'entrée x est représentée dans l'espace par un triangle. Le vecteur arrondi y le plus proche parmi les treize possibles est le vecteur d'indice i=2. .

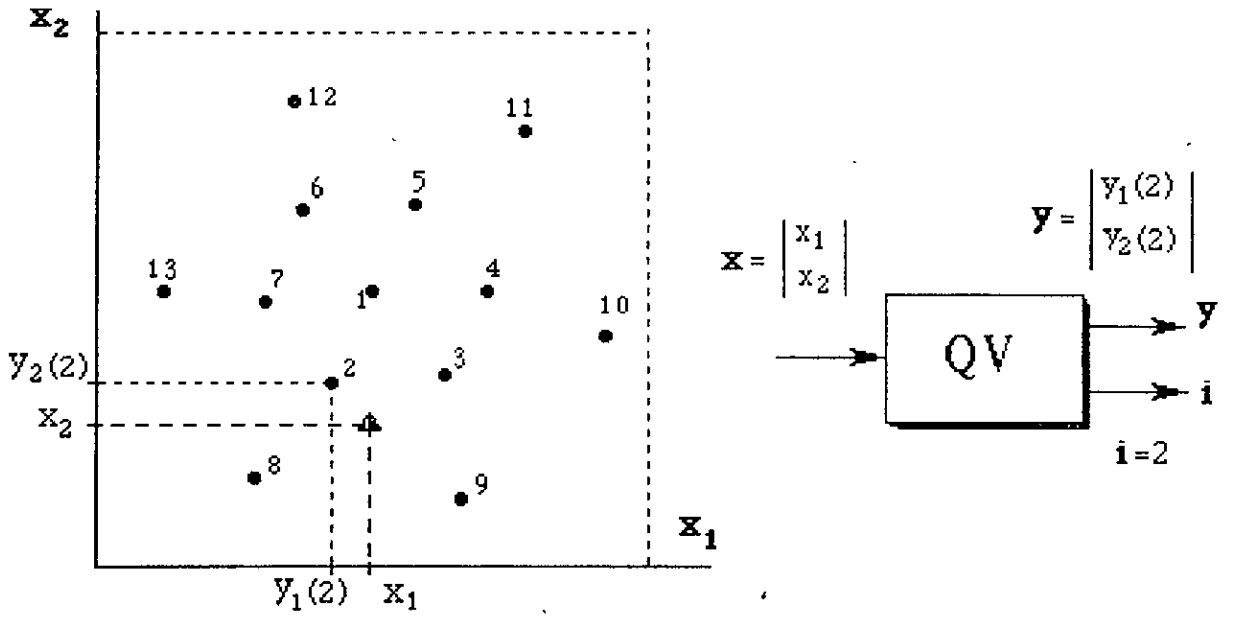


Figure 2.3 : Principe de la quantification vectorielle.

Dans les systèmes de communication, l'indice est transmis au décodeur qui choisit la séquence de sortie correspondante à cet indice comme le montre la figure 2-4.[31].

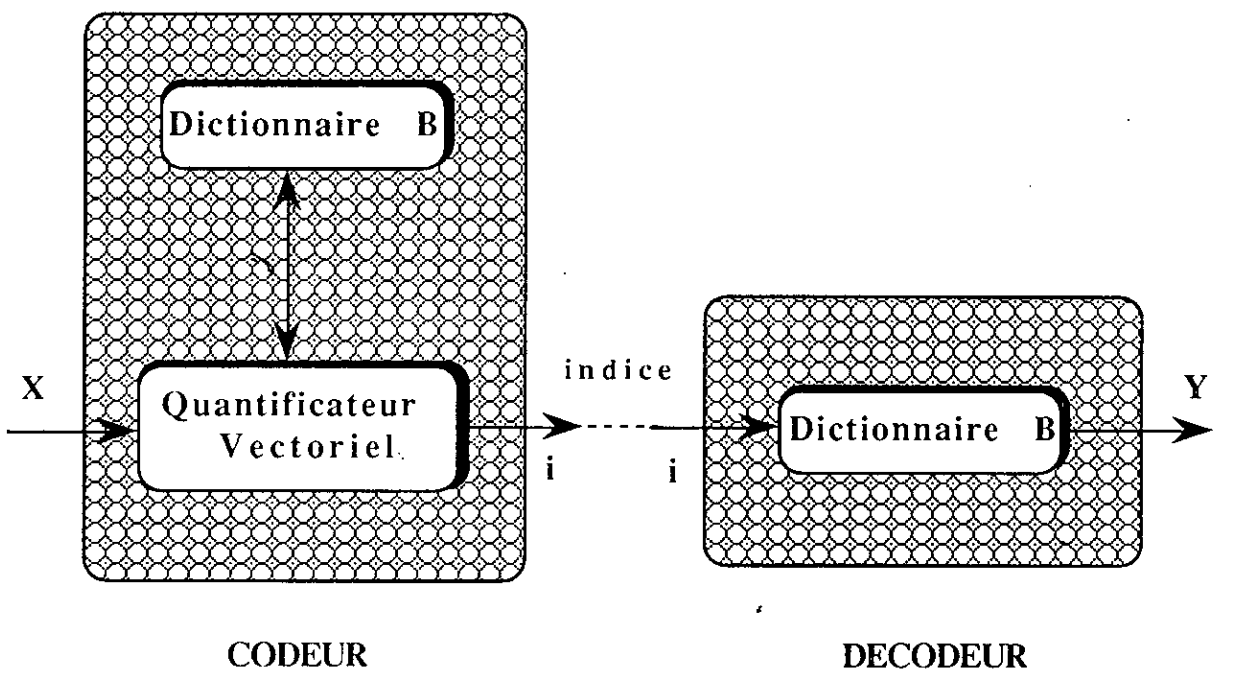


Figure 2.4 : Schéma d'un quantificateur vectoriel , x est la séquence à à coder, i l'indice à transmettre et y est la séquence reproduite.

2.1.4. Comparaison des performances de la quantification scalaire et vectorielle:

Il est convenable de parler du débit binaire d'un quantificateur. Il s'agit bien sûr de la transmission d'un indice i parmi N possible; ce qui représente $\log_2 N$ bits. Il est souvent pratique de ramener le débit par dimension comme par exemple pour permettre une comparaison entre quantificateurs operant sur des vecteurs de dimensions différentes [24].

Aprés avoir définit chacune des deux quantifications, il serait intéressant de comparer les performances de l'une par rapport à l'autre. On prendra pour cela, deux exemples simples à deux dimensions.

Premier exemple:

On y compare la quantification d'un vecteur x selon deux approches différentes. On suppose que les composantes du vecteur sont indépendantes et distribuées chacune selon la loi normale. Dans la première situation, on emploie une paire de quantificateurs scalaires que l'on applique à chaque composante séparément. Ces quantificateurs sont optimaux, selon les tables
Dans la deuxième situation, on utilise un quantificateur vectoriel obtenu par un algorithme iteratif [24] , ayant le même débit par dimension. La structure retrouvée est la structure (1, 6, 9) schématisée par la figure 2.5.b. C'est une structure optimale comme on le verra par la suite. Il apparait clairement qu'une paire de quantificateurs scalaires est équivalente à un quntificateur vectoriel dont les points, sont au sommet d'une grille cartesienne (figure 2.5.a).

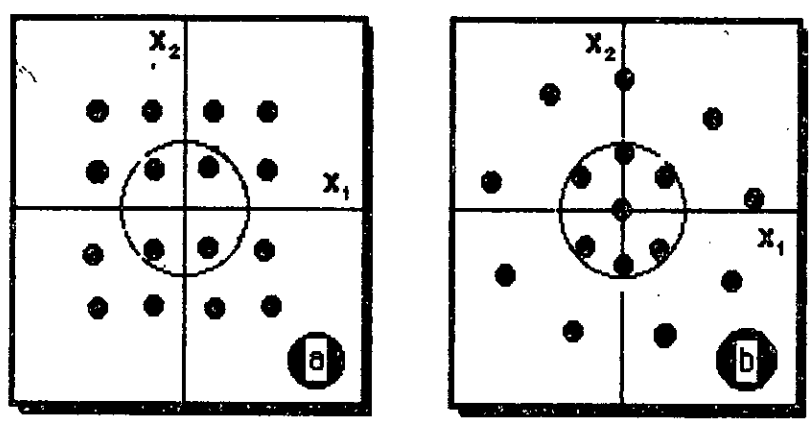


Figure 2.5 : schémas représentant les points quantifiés d'un vecteur à deux dimensions (dont les composantes sont indépendantes et distibuées selon une loi normale) par:

a/ une paire de quantificateurs scalaires

b/ un quantificateur vectoriel

Deuxième exemple:

Supposons que nous avons toujours affaire à un vecteur à deux dimensions, mais distribué cette fois de façon uniforme sur un certain domaine et que nous comparons à nouveau les performances d'une paire de quantificateurs scalaires et un quantificateur vectoriel. On notera que les deux quantificateurs sont à débits comparables; par conséquent, ils ont le même nombre de points et par suite la même densité moyenne.
Ces deux motifs (figure 2.6), forment des réseaux réguliers.

Dans la theorie des réseaux [24]; le premier est couramment appelé le réseau cubique Zm (pour m=2 dimensions) et le second réseau hexagonal A_2. Pour conserver des densités de points comparables, il faut poser :

$$d_2 / d_1 = \sqrt{2} / \sqrt[4]{3} = 1,075 \qquad (2.4)$$

Le point le plus mal quantifié est représenté dans les deux cas par un point gris à distance r_1 et r_2 respectivement. Le rapport :

$$r_2 / r_1 = \sqrt{2/3}\, d_2/d_1 = 0,877 \qquad (2.5)$$

nous permet de voir la supériorité du second quantificateur par rapport au premier.
Par ces deux exemples pratiques, nous voyons l'éfficacité de l'utilisation de la quantification vectorielle lors du codage d'une source.

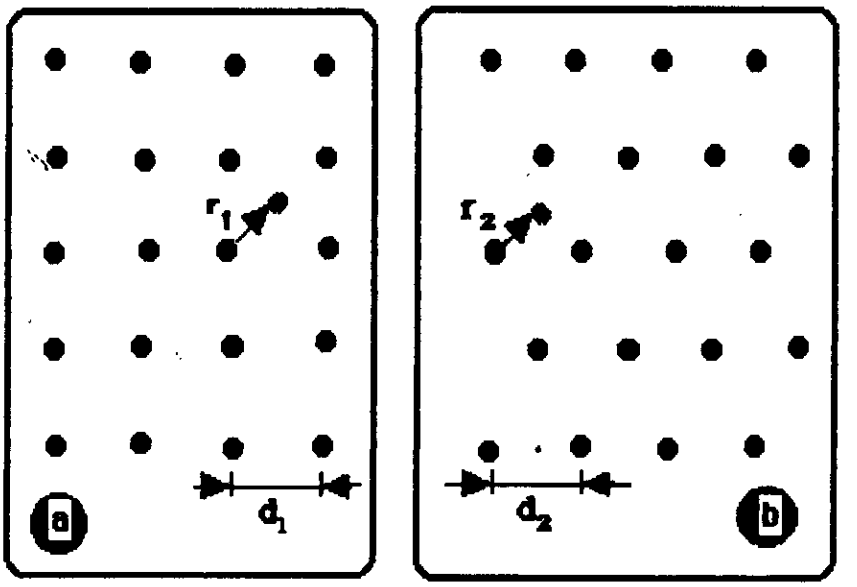


Figure 2.6 : Schémas représentant un vecteur à deux dimensions (dont les composantes sont uniformément réparties) quantifié par : Deux quantificateurs scalaires (a),et par un quantificateur vectoriel à motif hexagonal (b).

2.2. Quantification vectorielle optimale:

2.2.1. Mesure de performances d'un quantificateur:

Ainsi, comme on a vu antérieurement, un quantificateur vectoriel est caractérisé par son dictionnaire qui influe considérablement sur ses performances techniques. Ce dernier est caractérisé par sa taille N et sa dimension k; ces deux paramètres permettent de déterminer le débit binaire:

$$R = (\log_2 N)/k \qquad (2.6)$$

par échantillon. Une telle formulation permet de faire des études comparatives de quantificateurs opérant sur des vecteurs de dimensions différentes.

Mais un quantificateur n'est completement défini, que si on prend un autre critère (en plus de celui du dictionnaire) en considération: C'est la distance définissant le degré de distorsion qui s'introduit lors de l'approximation d'un vecteur $\mathbf{x}$ par un vecteur approché $\mathbf{y_i}$ notée par $\mathbf{d(x, y_i)}$. Cette distance est une fonction réelle, non-négative qui est utilisée seulement aux fins de comparaison. Elle est souvent utilisée sous la forme suivante [24, 31].

$$d(x, y_i) = |x-y_i|^r \qquad (2.7)$$

En particulier pour r=2, d devient le carré de la distance Euclidienne. Les chercheurs [6, 8, 13, 16, 24] utilisent souvent cette distance qui dans plusieurs situations, s'interprète directement comme une puissance d'erreur de quantification.

2.2.2. Définition de quantification optimale:

Un quantificateur Q est entièrement défini par son ensemble **B** (Dictionnaire) et la distance associée $\mathbf{d(x,y)}$:

$$Q(x) = y_i \quad \text{tel que} \quad d(x, y_i) \leq d(x, y_j) \quad \forall j \neq i \qquad (2.8)$$

$$\text{où} \quad (y_i, y_j) \in B \times B$$

Par ailleurs, si l'on connaît le nombre N de vecteurs arrondis $\mathbf{y_i}$, la distance d(x,.) et la distribution conjointe du vecteur d'entrée, p(x), on peut reconnaître le quantificateur optimal. C'est de tous les quantificateurs, ou les ensembles S possibles, celui qui minimise l'espérance mathématique de la distance d(x, .) [24, 29, 31], soit :

$$E(d) = \int d(x, Q(x)).\ p(x)\ dx$$

Cependant approcher les performances d'un tel quantificateur n'est pas une chose aisée, aussi, pour des cas particuliers en quantification scalaire, Lloyd [25], a proposé un algorithme iteratif appelé algorithme de Lloyd-Max pour construire un quantificateur optimal et Max a tabulé les valeurs arrondies [1]. Paez et Glisson ont étendu ces résultats aux densités Laplace et Gamma [31], et Pearlman à la distribution de Rayleigh [26].

Dans le cas d'un signal multidimensionnel, c'est à dire pour une quantification vectorielle, Zador a fait une étude théorique où il a utilisé l'integrale de Bennett géneralisée comme fonction de compression appliquée pour un nombre de niveaux de quantification trés élevé [32]. La figure 2.7 montre le modèle de compression appliqué, lors de la quantification vectorielle.

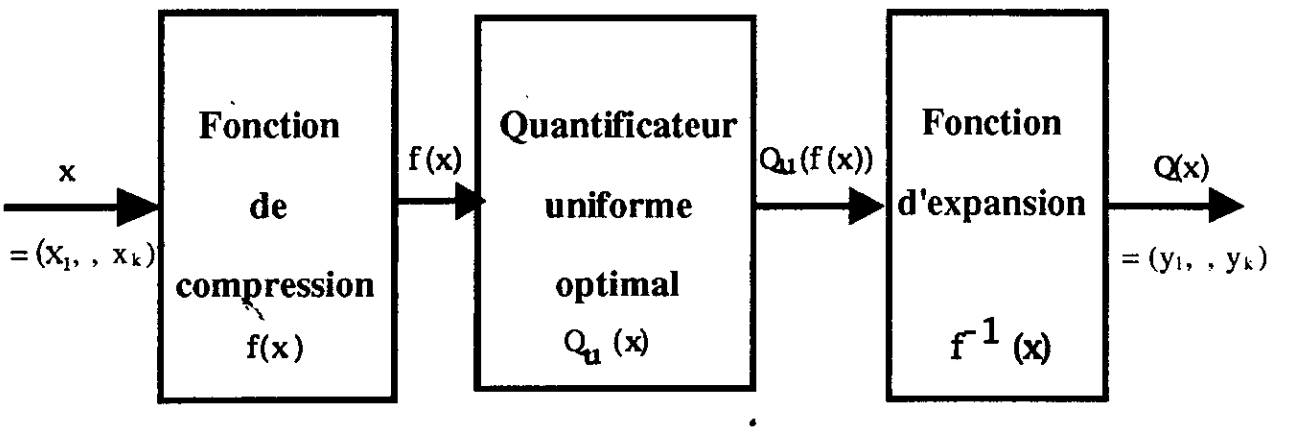


Figure 2.7-Schéma montrant un modèle de compression dans le cas d'une quantification vectorielle.

A partir de la figure ci-dessus donnée par Gersho [9], on peut trouver la sortie du quantificateur donnée par l'expression suivante:

$$Q(x) = f^{-1} \circ Q \circ f(x) \qquad (2.9)$$

Les expressions qu'il a obtenu [9], montre que le problème de la quantification peut-être divisé en deux aspects separés :

1°/ Trouver la meilleure fonction de compression d'un espace d'entrée multi-dimensionnel.
2°/ Implémenter le quantificateur uniforme multidimensionnel optimal dans l'hypercube unité.

En deux dimensions, le quantificateur optimal uniforme a une structure ou une constellation d'hexagones réguliers. En cherchant la fonction de compression inverse, la structure du quantificateur sera formée d'hexagones distordus ou irréguliers [9] . En pratique, il est trés difficile voire impossible (à partir de k=4), de trouver la fonction inverse de compression, c'est pourquoi les chercheurs ont pris une autre voie afin de trouver de plus amples résultats dans ce domaine. Il s'agit d'utiliser les conditions d'optimalité d'un quantificateur vectoriel.

a/ Conditions d'optimalité:

En termes mathématiques, une quantification vectorielle est défini comme étant une transformation d'un espace Euclidien à k dimensions ($\Re^k$), vers un sous-ensemble **Y** fini de $\Re^k$. Donc :

$$Q : \mathcal{R}^k \rightarrow Y$$

où

$$Y = \{ y_1 , y_2 , \ldots , y_N \} \qquad (2.10)$$

où Y est un sous-ensemble de $\Re^k$. Tel que y_i est un point de la sortie, qui appartient à $\Re^k$ pour tout i.
A chaque quantificateur à N points est associée une partition $S_1, S_2, \ldots, S_N$ où:

$$S_i = Q^{-1}(y_i) = \left\{ x \in \mathcal{R}^k \mid Q(x) = y_i \right\} \qquad (2.11)$$

Ce qui implique que :

$$\bigcup_{i=1}^{N} S_i = \mathcal{R}^k \qquad \text{et} \qquad S_i \cap S_j = \varnothing \quad \text{pour } i \neq j \qquad (2.12)$$

Ainsi le quantificateur Q est défini d'une manière unique en spécifiant l'ensemble de sortie Y et la partition correspondante { S_i }.

L'erreur absolue quadratique moyenne (mean absolute squared error: MASE) liée à ce quantificateur, est donnée par [31] :

$$D = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left| x - Q_N(x) \right|^2 . p(x). dx \qquad (2.13)$$

où p(x) est la fonction densité de probabilité conjointe.

En minimisant D par rapport à S_i et y_i , on obtient les deux conditions d'optimalité suivantes:

$$y_i = \int_{S_i}\int x . p(x) . dx \Big/ \int_{S_i}\int p(x) . dx \qquad (2.14)$$

y_i est le centroïde de la région S_i.

$$S_i = \bigcap_{i=1, j\neq i}^{N} \left\{ x: |x - y_i| < \left| x - y_j \right| \right\} \qquad (2.15)$$

L'expression ci-dessous indique que les régions S_i sont formées en prenant l'intersection des partitions de Dirichlet, des y_i et des autres points de la sortie[9].
Il faut savoir qu'une partition de Dirichlet, est formée par la bissection perpendiculaire à un segment connectant une paire de points de la sortie [9, 24].
A partir de ces deux conditions, on montre que l'erreur quadratique moyenne sera sous la forme [31]

$$D = \sigma_x^2 - \sum_{i=1}^{N} \left| y_i \right|^2 \int_{S_i}\int p(x) . dx \qquad (2.16)$$

Où σ_x^2 est la puissance du signal.

b/.Exemples d'un quantificateur optimal:

Pour une source de densité uniforme, le quantificateur qui lui correspond aura une partition de Dirichlet [9]. Il existe pour ce type de densité, un quantificateur optimal qui vérifie les deux conditions (2.14) et (2.15).

On notera qu'en général, les points générants une partition de Dirichlet ne sont pas nécéssairement les centroïdes de leurs régions respectives. La figure 2.8 montre bien cet aspect là [24].

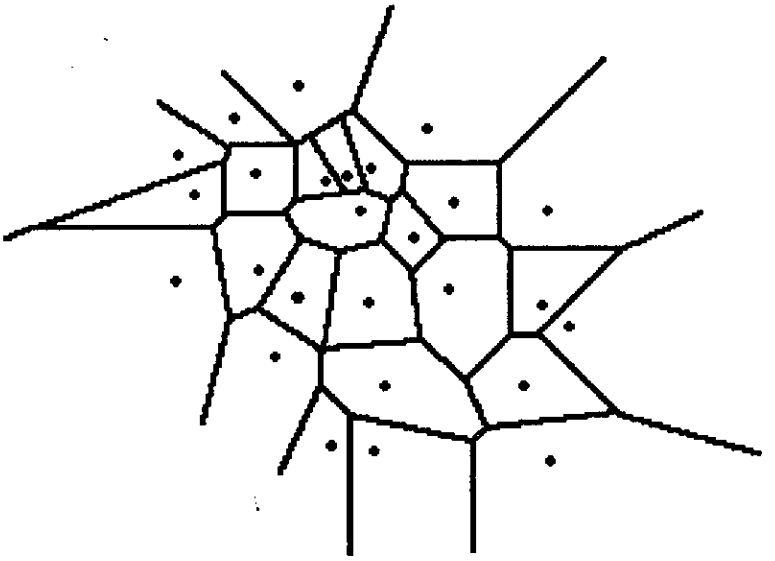


Figure 2.8 : Schéma représentant une partition de Dirichlet [9]

Pour un quantificateur optimal dont la source est uniformément distribuée, la partition de Dirichlet sera formée par une constellation d'hexagones réguliers; comme il est illustré sur la figure 2.9. Dans ce cas là, les points de la sortie formant la partition de Dirichlet seront bien les centroïdes de celle-ci.

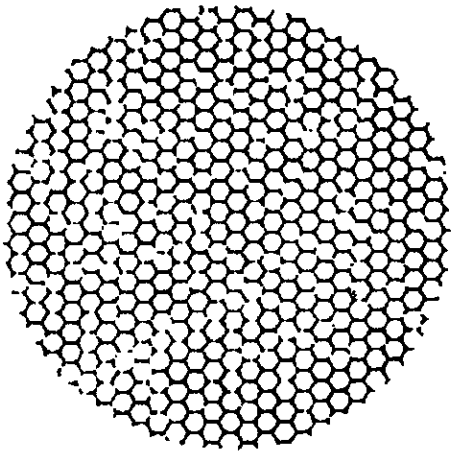


Figure 2.9 : schéma représentant une constellation d'héxagones réguliers

Remarque:

Pour des sources ayant des densités non-uniformes, une méthode iterative est indispensable (en utilisant les conditions d'optimisation) pour converger vers un minimum local. Pour ce cas là, Fejes Toth [9] (figure 2.10) a trouvé une partition sous forme d'une constellation d'héxagones irréguliers.

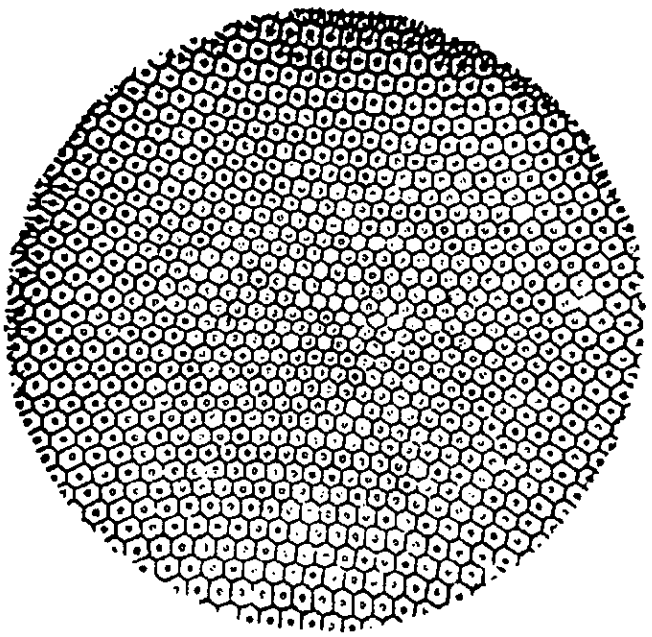


Figure 2.10 : schéma représentant une constellation d'héxagones irréguliers

Comme on l'a déja dit, il n'est pas facile de trouver une source qui permet d'obtenir de manière unique, les conditions d'optimalité.C'est pourquoi, d'un point de vue général, il serait intéressant de donner ci-dessous quelques aspects sur les limites des performances que peut atteindre un quantificateur sur la base des résultats théoriques obtenus par plusieurs chercheurs [9, 12, 14, 15, 31].

2.2.3. Performances asymptotiques des QV:

a/ Préliminaire:

Soit X un vecteur aléatoire tel que: $X = \{ x_1, x_2, ..., x_k \}$ appartient à l'espace $\Re^k$. Les composantes xi sont décrites par une densité de probabilité conjointe: p(x).

Soit un quantificateur Q à k dimensions décrit par:

- Une collection de N vecteurs de reproduction $y_1, y_2, ..., y_N$ appartenant à $\Re^k$; appelés alphabet de reproduction.
- Une partition $\{ S_1, S_2, ... S_N\}$ de $\Re^k$ où:

$$S_i \cap S_j = \varnothing \quad \text{pour } i \neq j$$

et

$$\bigcup_{i=1}^{N} S_i = \mathcal{R}^k$$

On définit un quantificateur par [9]:

$$Q(x) = y_i \qquad \text{Si } x \in S_i$$

La performance d'un tel quantificateur est mesurée au moyen de la distorsion moyenne donnée par:

$$D = E\left[d(x,Q(x))\right] = \sum_{i=0}^{N-1} \int_{S_i} p(x)\, \|x-y\|^2 dx \qquad (2.17)$$

L'entropie différentielle relative à ce quantificateur est la suivante:

$$H_Q = -\sum_{i=0}^{N-1} P_i \log_2 P_i \qquad (2.18)$$

Avec

$$P_i = \int_{S_i} p(x)dx = P_X(x \in S_i) \qquad (2.19)$$

On définit aussi le volume V_G de la région G, telle que $G \subset \Re^k$; exprimé par [31]:

$$V(G) = \int_G dx \qquad (2.20)$$

Soit V_k le volume de la sphère unité de dimension k [31]:

$$V_k = \mathrm{Vol}(\{ u : \| u \| \leq 1 \}) = \frac{2\Gamma(1/2)^k}{k\Gamma(k/2)} \qquad (2.21)$$

où

$$\|u\| = \left(\sum_{j=0}^{k-1} u^2(j) \right)^{1/2} \qquad (2.22)$$

et Γ est la fonction Gamma $(\Gamma(1/2) = \sqrt{\pi})$

b/ Bornes asymptotiques de performance:

Dans cette section, notre but est d'obtenir une borne inferieure à la distorsion moyenne asymptotique D pour les mesures de distorsion d(x, y):

L' hypothèse fondamentale dans l'étude de la quantification asymptotique (faite par Gersho [9]), est que la densité de probabilité p(x) soit constante pour des régions de quantification étroites; en d'autres termes, pour un nombre N assez grand, la probabilité p(x) varie peu à l'interieur des contours de la région de quantification Si; On aura donc:

$$p(x) \cong p_i \qquad \text{pour } x \in S_i \qquad (2.23)$$

L'expression (2. 19) deviendra équivalente à:

$$P_i = \int_{S_i} p(x)\, dx \cong p_i \int_{S_i} dx = p_i\, V(S_i) \qquad (2.24)$$

Donc, la distorsion moyenne donnée par l'expression(2. 17) se réécrit:

$$D = \sum_{i=0}^{N-1} \frac{P_i}{V(S_i)} \int_{S_i} \|x - y_i\|^2 \, dx \qquad (2.25)$$

On montre [31] , (en prenant $L(x-y) = (x - y)^2$) que:

$$\int_S d(x,y)dx = \int_{S_i} \|x - y_i\|^2 dx \geq \frac{V(S)}{V_k} \int_{u:\|u\|<1} \left\| \left(\frac{V(S)}{V_k} \right)^{1/k} . u \right\|^2 du \qquad (2.26)$$

Soit la fonction [31] :

$$F_k(v) = \frac{1}{V_k} \int_{u:\|u\| \le 1} \left\| \left(\frac{V(S)}{V_k} \right)^{1/k} . u \right\|^2 .du \qquad (2.27)$$

L'expression (2. 26) peut se réecrire (en utilisant la fonction donnée en (2. 27)), de la façon suivante:

$$\int_S d(x,y)dx = \int_{S_i} \|x - y_i\|^2 dx \geq V(S).F_k\left(V(S)^{1/k}\right) \qquad (2.28)$$

A partir des équations (2. 24) et (2. 28), on peut écrire:

$$D \geq \sum_{i=0}^{N-1} P_i F_k\left(V(S_i)^{1/k}\right) \approx E\left[F_k\left(V(X)^{1/k}\right)\right] = D_L \qquad (2.29)$$

Sachant que F_k est une fonction convexe [31], on aura alors en appliquant l'inégalité de Jensen:

$$D \geq D_L \geq F_k(E[V(X)^{1/k}]) \qquad (2.30)$$

d'où:

$$F_k^{-1}(D) \leq F_k^{-1}(D_L) \leq E\left[V(X)^{1/k} \right] \qquad (2.31)$$

En utilisant les expressions (2. 18) et (2. 24), l'expression de l'entropie sera de la forme :

$$H_Q \approx h(p) - E[\log_2 V(X)] \qquad (2.32)$$

où $h(p) = \int p(x) \log_2 p(x)\, dx$

En appliquant une deuxième fois, l'inégalité de Jensen à la fonction log, on aura:

$$H_Q = h(p) - k\, E\left[\log_2 V(X)^{1/k}\right] \geq h(p) - k \log_2\left(E\left[V(X)^{1/k}\right]\right) \qquad (2.33)$$

En utilisant la relation (2. 31), on obtient:

$$H_Q \geq h(p) - k\log_2\left(F_k^{-1}(D_L)\right) \geq h(p) - k\log_2\left(F_k^{-1}(D)\right) \qquad (2.34)$$

Finalement, on obtient la borne inferieure asymptotique de la distorsion moyenne exprimée par l'inégalité suivante:

$$D \geq D_L \geq F_k\left(2^{-(H_Q - h(p))/k}\right) \qquad (2.\ 35)$$

En remplaçant F_k par la relation (2. 27), on aura:

$$D \geq D_L \geq D_Q = \frac{k\, V_k^{2/k}}{k+2} 2^{-(2/k)(H_Q - h(p))} \qquad (2.\ 36)$$

Ainsi , pour un quantificateur Q ayant une entropie fixe H_Q, il est toujours possible de trouver une borne inferieure asymptotique de la distorsion moyenne de ce quantificateur.

c/ Exemple [31] :

Dans cette exemple, on va donner les résultats de comparaison (faite par échantillon) entre la borne inferieure asymptotique de D et celle de Shannon (pour une source gaussienne sans mémoire).

$$\bar{D} \geq \bar{D}_Q^{(k)}(R) = \left[\left(\frac{e}{1+k/2}\right)\Gamma\left(1+\frac{k}{2}\right)^{2/k}\right]\bar{D}_{SLB}^{(k)}(\bar{R}) \qquad (2.\ 37)$$

où $\bar{D} = \frac{D}{k}$; $\bar{D}_Q^{(k)} = \frac{D_Q}{k}$

$H_Q/k = H_k \leq \bar{R}$; $h_k = \frac{h_p}{k}$

et $\bar{D}_{SLB}^{(k)}(\bar{R})$ exprime la borne de Shannon par échantillon

$$\bar{D}_{SLB}^{(k)}(R) = \frac{D_{SLB}}{k} = \left[2e\, V_k^{2/k}\Gamma\left(1+\frac{k}{2}\right)^{2/k}\right]^{-1} .2^{-2(\bar{R}-h_k)}$$

On peut toujours exprimer ces mêmes performances en fonction du rapport signal à bruit RSB tel que:

$$RSB = 10 \log_{10} \frac{\sigma_x^2}{D} \qquad (2.38)$$

et

$$RSB_Q^{(k)}(R) = RSB_{max}(R) - 10\log_{10}\left[\left(\frac{e}{1+k/2}\right)\Gamma\left(1+\frac{k}{2}\right)^{2/k}\right] 2^{-2R} \quad (dB) \qquad (2.39)$$

tel que:

$$RSB_{max}(R) = 6.02 \times R$$

En comparant maintenant la borne inferieure de D à la fonction de distorsion, on trouve [31] :

$$\overline{D}_Q^{(k)}(R) = \left[\left(\frac{e}{1+k/2}\right)\Gamma\left(1+\frac{k}{2}\right)^{2/k}\right] 2^{-2R} \qquad (2.40)$$

avec $D(R) = 2^{-2R}$ pour une source gaussienne

Cette borne donne de meilleurs résultats que ceux de la fonction D(R) pour k fixe. Ainsi, elle tend vers la limite de D(R) quand k tend vers l'infini [31].

2.2.4. Techniques de calcul de quantificateurs vectoriels optimaux:

Soit un quantificateur Q défini par son dictionnaire $B = \{ y_0, y_1, ..., y_{N-1} \}$ et sa partition $S = \{ S_0, S_1, ..., S_{N-1} \}$ de l'espace $\Re^k$ (figure 2.11 pour k=2).

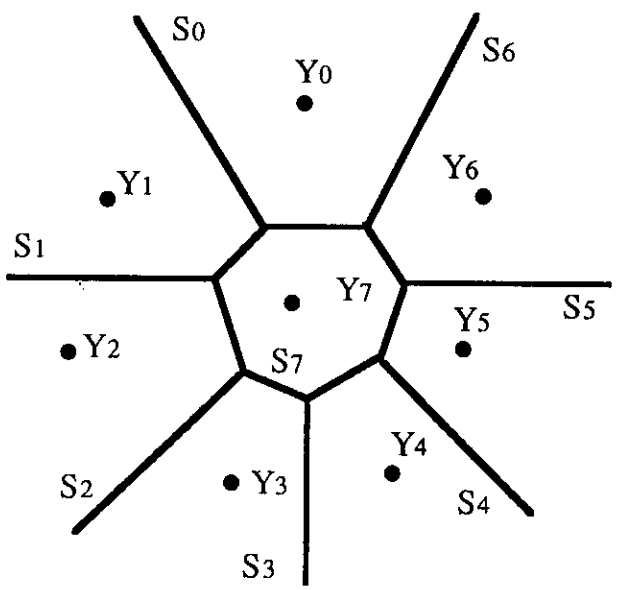


Figure 2.11 : Exemple de partition dans R^2 pour N=8

En théorie, trois opérations successives sont nécéssaires pour quantifier un vecteur donné X [29].

1- Trouver la région de quantification contenant x, c'est à dire la région de Voronoï S_i telle que $x \in S_i$ ce qui revient à determiner le plus proche voisin de x dans le dictionnaire B.

2- Chercher l'indice i de la région de Voronoï contenant x qui est aussi l'indice du mot de code y le plus proche de x.

3- Régénérer à la réception le mot de code y_i à partir de l'indice i.

Ces trois opérations sont représentées par la figure ci-dessous (figure 2.12) [31].

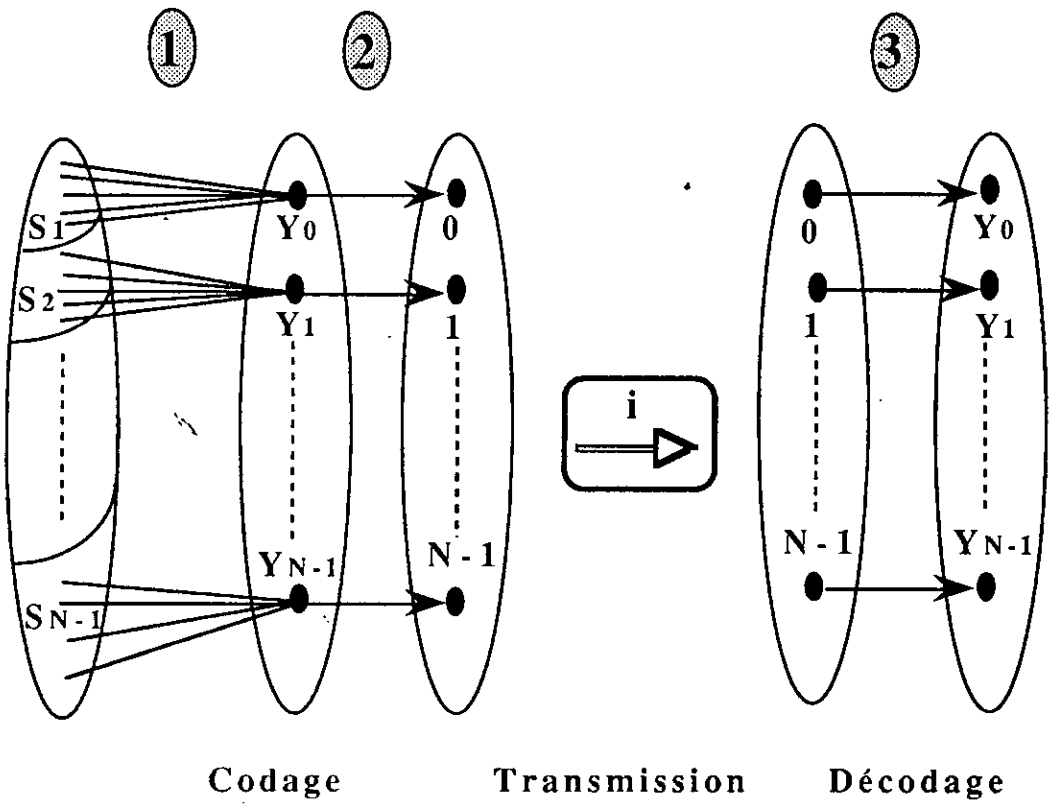


Figure 2.12 : Schéma representant les trois operations nécéssaires pour une quantification vectorielle.

Remarque: Rappel sur la région de Voronoi

On rappelera que pour un quantificateur, on définit la région de Voronoi (ou encore Région de Derichlet) autour d'un point y_i, dénotée V_i, comme l'ensemble des points de $\Re^k$; qui sont quantifiés par y_i[44].

$$V_i = \{ x / x \in \Re^k \text{ et } d(x-y_i) \leq d(x-y_j) \} \quad \forall i \neq j$$

La figure suivante (figure 2.13) illustre un exemple de région de Voronoi pour les réseaux réguliers hexagonaux.

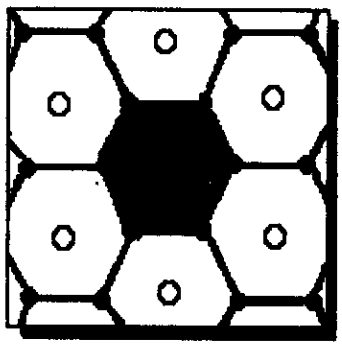


Figure 2.13: Schéma représentant des régions de Voronoi
(la région contrastée est la région V_0)

En pratique, le problème de la conception du meilleur quantificateur a focalisé les recherches sur la quantification scalaire puis vectorielle. Dans les deux cas, les résultats théoriques ont montré que les distributions uniformes sont les mieux représentées; dans le sens d'une plus faible distorsion.

La théorie de la distorsion a même montré que la distorsion peut-être réduite par un codage de blocs toujours plus long [9, 14].

L'exploitation de la propriété des composantes gaussiennes du vecteur, dont la distribution possède une symetrie sphérique; a conduit à trouver une méthode de quantification appelée quantification vectorielle sphérique. Cette méthode consiste à quantifier séparement la norme (le gain) et la phase (orientation). En général, elle se base sur les réseaux réguliers en tirant profit de leurs propriétés [24].

Dans cet ordre là, il existe deux approches essentielles qui sont: L'approche statistique et l'approche algébrique.

a/ L'approche statistique:

Cette approche consiste à appliquer l'algorithme LBG (du nom de ses auteurs: Lynde, Buzo et Gray) [117]. L'algorithme cité est désigné aussi par le vocable de la K-moyenne. Certains chercheurs [24, 29, 31] le présentent comme une extension de l'algorithme Lloyd-Max [1, 25].

C'est un algorithme iteratif; il transforme les représentants d'une partition pour obtenir une nouvelle partition dont l'erreur est inférieure à celle de la première. En faisant plusieurs iterations, celui-ci va tendre vers un minimum local qui n'implique nécessairement pas sa convergence vers un minimum absolu.

En général, l'algorithme en question est appliqué dans le cas d'une source dont la densité est non-uniforme.

Cette approche est recommandée quand la distribution des orientations sur la sphère unité est franchement non-uniforme. L'algorithme de la K-moyenne permet de tirer parti au maximum de cette non-uniformité [24].

b/ L'approche algébrique:

Elle consiste à faire usage des propriétés des réseaux réguliers dans l'étude des quantificateurs, et ainsi constituer des dictionnaires de points sur une hypersphère unité. Les dictionnaires obtenus à partir de cette technique sont trés bien structurés. Les algorithmes relatifs à cette approche sont éfficaces et rapides.

Cette deuxième approche est appropriée dans le cas où les orientations sont raisonnablement uniformément distribuées sur la sphère [24].

Rappel: Definition d'un réseau [24]:

Un réseau régulier dans $\Re^k$ est l'ensemble de points x qui s'obtiennent par combinaison linéaire de k vecteurs de base indépendants $z_1, z_2, ..., z_k$ avec des coefficients de probabilité entiers $m_1, m_2, ..., m_k$.

$$L_k = \{ x / x = \Sigma\, m_i z_i \} ; \qquad m_i \text{ entier}$$

Exemple:

Le réseau hexagonal A_2, de la figure (2. 13) est obtenu à partir des vecteurs de base

$$z_1 = [\, 1, \sqrt{3}\,] \text{ et } z_2 = [\, 2, 0] \,.$$

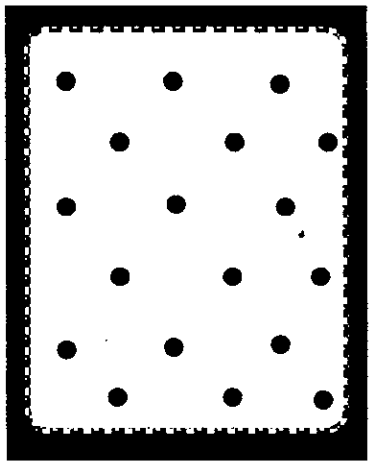


Figure 2.13 : Structure du réseau héxagonal A_2

Ainsi, un réseau régulier conduit au quantificateur optimal [24]. Une hypothèse émise par Gersho [9] veut que pour toute dimension, il existe un réseau régulier qui atteint ainsi les performances optimales. Cette hypothèse jamais demontré n'est à présent demontrée que pour les dimensions 2 et 3 [9]. Il existe cependant des réseaux particulièrement bons en dimensions 8 et 24 [24].

c/ Quantification par treillis :

Dans ce même ordre d'idée, les réseaux réguliers ont été largement exploités en modulation ou codage de canal par Ungerboeck [31]. Ainsi, ce dernier a obtenu de bons résultats dans le codage de canal à bande limitée en utilisant une technique appelée " Mapping by Set Partitioning" (ou treillis) pour des constellations de dimensions 1 et 2. Cette méthode se base sur les codes convolutifs [10,31]. Profitant de la dualité entre le codage de canal (modulation) et le codage de source (quantification) [10] , Fischer et Marcellin ont généralisé cette fois la notion de codage par treillis pour le codage de source. Cette nouvelle technique s'appelle " Treillis coded quantization" (TCQ) [46].

** Principe:*

Dans le cas du codeur TCM (Treillis coded modulation), la figure suivante (fig 2. 14) montre son principe de base.

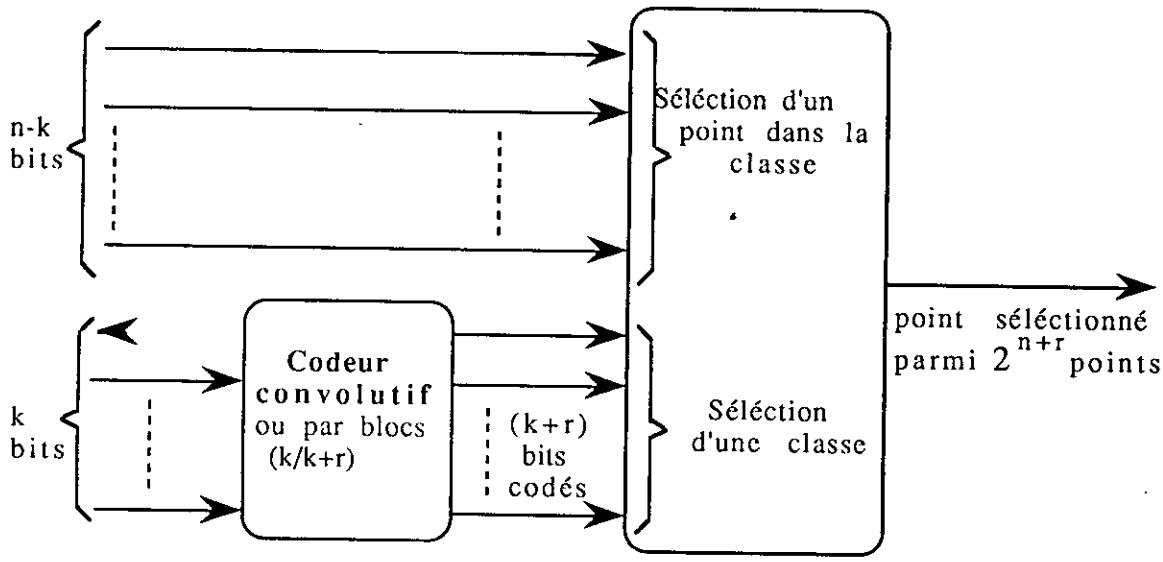


Fig 2. 14 : Structure d'un codeur TCM [31].

En ce qui concerne le TCQ, la taille de l'alphabet vectoriel va être doublé, tout en conservant le débit initial intact; cela ne se fait qu'en imposant quelques règles à l'alphabet.

Ainsi si l'alphabet vectoriel Y est composé de 2^m vecteurs à k dimensions, en doublant la taille de cet alphabet, on obtient un nouvel alphabet Y' contenant 2^{m+1} vecteurs. On divise ce nouveau alphabet en $2^{m'+1}$ classes (ou sous alphabets) { Q_0, Q_1, ..., Q_m} où chacune de ses classes est constituée de $2^{m-m'}$ vecteurs.

Ainsi, dans le cas d'une utilisation d'un quantificateur vectoriel "classique", il faudrait m'+1 bits et m-m' bits pour identifier respectivement la classe et le vecteur type; ce qui donnerait un débit total de m+1 bits. Cependant, en faisant appel à un treillis, il nous offrira la possibilité de ramener le débit à R bits.

Dans la technique du TCQ, chaque classe représente un quantificateur dont les performances dépendent du choix du type de quantification à appliquer. Fischer et Marcellin avaient utilisé une quantification scalaire.

*** Exemple de TCQ selon Marcellin et Fischer:*

Nous allons voir d'une façon brève un exemple de TCQ afin de comprendre son principe.

Considérons un codeur convolutif de taux d'émission $\tau = 1/2$ (c'est à dire qu'à l'entrée, on a 1 bit et on reçoit 2 bits à la sortie). Le codeur est composé d'un regitre à décalage à

deux cellules C_1 et C_2 (il comporte donc, $2^2 = 4$ etats), comme le montre la figure 2. 15 [45].

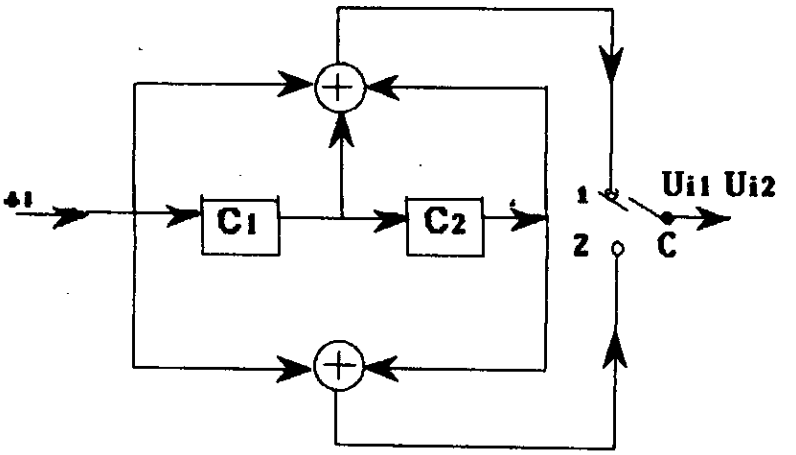


Fig 2. 15: Codeur convolutif à deux cellules

A partir du schéma ci-dessus, on trouve les sorties d'aprés le tableau suivant:

Tableau 2.1: Table de codage du codeur convolutif

t_i	x_i	C_1C_2	U_{i1} U_{i2}
1	1	00	11
2	1	10	01
3	1	11	10
4	0	11	01
5	1	01	00
6	0	10	10
7	0	01	11
8	0	00	00

En utilisant ce tableau, on peut dresser le diagramme des etats comme l'illustre la figure suivante (figure 2. 16).

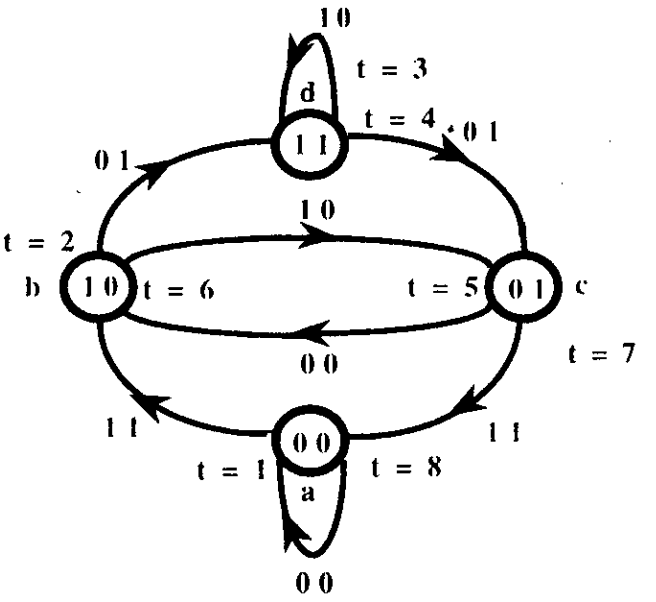


Fig 2. 16 : Diagramme des états [45]

Comme dans la théorie du TCQ il est question de classe et de leurs centres, donc dans la figure 2. 16, il est préférable de représenter les classes au lieu des éléments binaires. On aura alors le diagramme suivant:

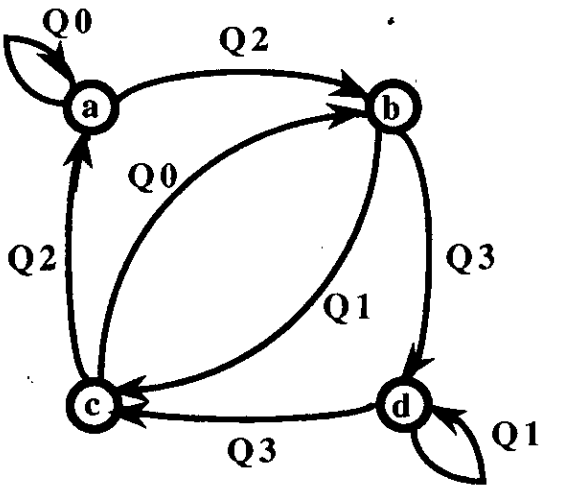


Fig 2. 17 : Diagramme de transition

On peut donner une autre forme au diagramme de transition et qui est la suivante:

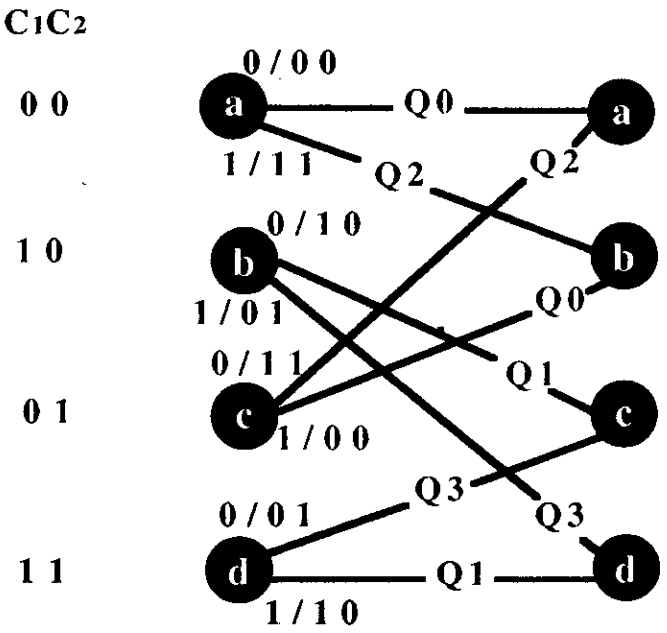


Fig 2. 18: Exemple de treillis à quatre états.

Dans le treillis, chaque noeud représente un état et les branches correspondent aux transitions. Des deux chemins arrivant à chaque état, un seul présente un certain intérêt. Ainsi, le chemin qui aura accumulé la plus faible erreur sera le chemin restant ou le survivant. il représente, pour les chemins arrivant à un état donné la meilleure stratégie de quantification pour les vecteurs passés.

Pour minimiser la distance euclidienne, on utilise l'algorithme de Viterbi [10]:

En premier lieu, on attribue une erreur cumulée nulle à l'un des états du treillis et une erreur infinie aux autres états. Le codeur et le décodeur s'entendent préalablement sur le choix de l'état ayant une erreur nulle.

1°/ On calcule les distances euclidiennes pour chaque étage.

$$\left(d_j^i\right)^2 = \left(x_i - q_j\right)^2 \qquad \text{avec} \qquad q_j \in Q_j\ ,\ j = 0, 1, \ldots, 3$$

2°/ Puis, on additionne les distorsions trouvées à chaque branche et à un noeud donné, nous gardons la branche ayant la distorsion minimale.

3°/ Enfin, nous comparons les quatres états dont nous choisirons celui qui possède la distorsion cumulée minimale.

d/ La quantification sous-optimale:

* Quantificateurs polaires:

Comme nous avons vu plus haut, les conditions énumerées par les expressions (2.14) et (2.15) sont des conditions sévères pour le calcul des centroïdes et des régions avoisinantes d'un quantificateur vectoriel optimal.

Dans les algorithmes de design ou de conception, lors du calcul du centroïde de l'une des N régions de quantification S_i et qui est donné par la relation (2. 15); il est à remarquer que la dimension k des coordonnées d'un point implique k divisions scalaires et k+1 intégrations de dimension k, soit pour les N régions à considerer un total de kN divisions et kN intégrations.

La complexité s'accroit lors de la définition des régions avoisinantes de quantification S_i pour chaque valeur arrondie de sortie.

La relation exprimant la région S_i montre que la détermination de chaque région S_i nécéssite la recherche de N-1 demi-espaces de dimension k, soit un total, pour N régions, de N^2 intersections.

D'autres difficultés surgissent lors de l'estimation de la distorsion. C'est pourquoi, la conception d'un quantificateur vectoriel optimal reste un art en soi [24].
En fait, une étude comparative relative à la complexité des quantificateurs vectoriels et scalaires, a été faite par Swaszek [38] et a été reprise par Berkani [31]..

Aussi, pour remédier à la complexité des quantificateurs vectoriels, les chercheurs se sont orientés vers la quantification sous-optimale. Chacun tente par de nouvelles études, méthodes ou techniques de récolter un certain gain en performances sans exiger des conditions contraignantes ou supplémentaires; ce qui revient à dire qu'il y a réduction de la complexité de façon indirecte.

Ainsi, ces derniers préfèrent présenter le système quantificateur comme un ensemble de quantificateurs élémentaires scalaires déja connus [38]. Mais il est plus judicieux d'utiliser un repère polaire de coordonnées qu'un système cartésien. En effet, un signal exprimé sous une forme complexe peut être obtenu et traité plus facilement dans un repère polaire où son amplitude et sa phase sont statistiquement indépendants et peuvent donc être quantifiées indépendemment l'une de l'autre. Ce type de procédé donne un nouvelle sorte de quantificateurs appelés quantificateurs polaires qui renferment en soi plusieurs catégories de quantificateurs selon les contraintes utilisées comme on le verra dans les prochains chapitres.

**Quantificateur en spirale:

De manière similaire, des travaux récents ont été faits dans le domaine de la quantification sous-optimale par Berkani [31]: cette méthode consiste en la discrétisation de la spirale dans un système polaire. Elle est basée sur l'optimisation des différents paramètres de la spirale, que nous verrons ultérieurement.

Description:

Soit une variable aléatoire Z Gaussienne à deux dimensions. Dans le système cartésien, elle est exprimée par:

$$Z = x + j\,y \qquad \text{où} \quad j^2 = -1$$

Dans le système polaire, elle est de la forme:

$$Z = r\,e^{j\theta}$$

où:

$$\begin{cases} r = \sqrt{x^2 + y^2} \\ \theta = \text{arctg}\ \dfrac{y}{x} \end{cases}$$

La théorie du quantificateur en spirale repose sur celle de la spirale d'Archimède représentée par la figure 2. 19:

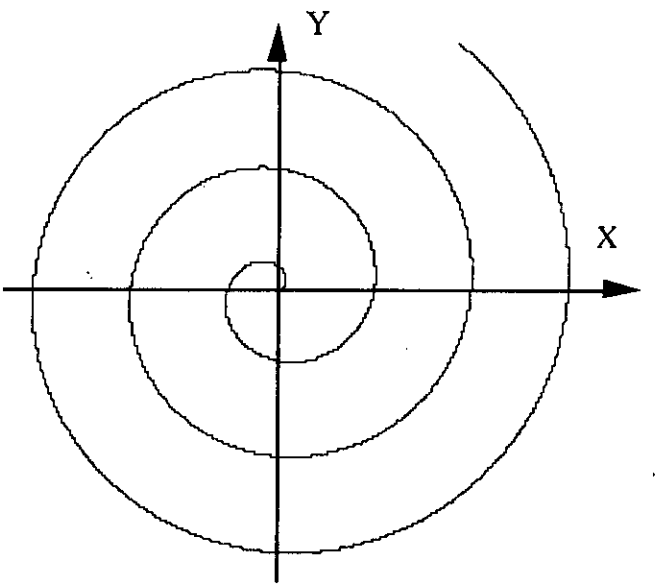


Fig 2. 19: Spirale d'Archimède dans le système cartésien [31]

Les points paramétriques régissant les points de la spirale sont les suivants:

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases}$$

où:

$$r = G\varphi \qquad \text{et} \qquad \varphi = \theta + 2k\pi \qquad (k \in Z) \qquad (2.41)$$

Comme première étape, la discrétisation de la spirale d'Archimède s'applique sur l'angle θ. Cela se fait en définissant un pas d'incrémentation "delta" (Δ) à l'angle θ. Δ constituera un des paramètres à optimiser.

Pour pouvoir mieux discrétiser l'angle θ, il est impératif de définir un autre paramètre (à optimiser aussi), et qui permet à la constellation de points de la spirale de s'éloigner de l'origine dans un mouvement de révolution: C'est l'angle initial θ_0.

Enfin, il ya un autre paramètre qui doit faire l'objet d'une optimisation: c'est le gain de spirale.

D'aprés l'expression (2. 41), si on suppose que $|G| = 1$, alors:

$$r = \varphi = \theta + 2k\pi$$

d'où

$$\theta = \varphi \ [2\pi] = r \ [2\pi]$$

Les crochets ([ ]) désigne la fonction modulo.

Ces relations montrent que toute l'information de la phase pour une spirale, est implicite dans l'amplitude. Ainsi, l'un des avantages essentiels de la spirale, est qu'il y a une réduction du débit de stockage, ce qui lui permet d'avoir de meilleures performances par rapport à celles d'autres quantificateurs.

Il existe d'autres avantages de la spirale que l'on peut trouver dans le travail de Berkani [31].

CHAPITRE 3

QUANTIFICATEUR POLAIRE SPQ

Introduction :

La quantification sous-optimale comme son nom l'indique ne permet pas d'obtenir des résultats aussi performants que ceux de la quantification optimale. Mais en revanche les méthodes générées par ce type de quantification sont plus simples à appliquer bien qu'elles soient liées à un certain nombre de contraintes. Afin d'améliorer les performances de tels quantificateurs le plus possible; on peut jouer sur certains de leurs paramètres.

Il serait intéressant d'utiliser un repére polaire pour un signal bidimensionnel en particulier si la densité de probabilité du signal d'entrée est de symétrie circulaire. Par la suite, on verra que ce système permet d'obtenir de meilleures performances que pour un système cartésien de plus, il existe deux approches pour la quantification polaire : celle du quantificateur polaire SPQ qui veut dire en anglais "*strictly polar quantizer* " et du quantificateur UPQ qui veut dire "*unrestrected polar quantizer* ", selon le débit du signal comme on le verra dans le prochain chapitre .

3.1 Quantificateur SPQ non-uniforme:

3.1.1 Formulation Mathématique :

Considérons une source aléatoire gaussienne bidimensionnelle , l'expression d'un point z, dans l'espace de coordonées x et y , dans un système de coordonnées polaires satisfait aux relations suivantes:

$$\begin{cases} x = r\cos\phi \\ y = r\sin\phi \end{cases} \qquad (3.1)$$

En d'autres termes, on aura :

En d'autres termes, on aura :

$$\begin{cases} r = \sqrt{x^2 + y^2} \\ \phi = \text{Arctg}\left(\frac{y}{x}\right) \end{cases} \qquad (3.2)$$

En notation complexe , on aura :

$$Z = r\, e^{j\phi} \qquad (3.3)$$

Comme la source est Gaussienne de variance $2\sigma^2$, alors la densité de probabilité conjointe sera donnée par :

$$P_{X,Y}(x,y) = \frac{1}{2\pi}\exp -\left(x^2 + y^2\right)/2 \qquad (3.4)$$

pour $-\infty < x \;\; ; \;\; y < +\infty$

En considérant les variables r et ϕ statistiquement indépendantes , alors on aura :

$$p(r,\phi) = f(r).\, g(\phi)$$

Utilisant cette hypothèse , on obtient [30]:

$$\begin{cases} f(r) = r.\exp -\left(r^2/2\right) & 0 \le r < +\infty \qquad (3.5) \\ g(\phi) = \frac{1}{2\pi} & 0 \le \phi < 2\pi \qquad (3.6) \end{cases}$$

Ainsi , on trouve que le module (ou l'amplitude) suit une loi de Rayleigh alors que la phase est uniformément distribuée.

3.1.2 Optimisation du quantificateur SPQ:

Si on divise le plan en N région $\Re_{i,j}$ telle que

$$\Re_{i,j} = \left\{ Z = r\, e^{j\phi} \,/\, r_{i-1} \le r < r_i \,;\, \phi_{j-1} \le \phi < \phi_j \right\} \qquad (3.7)$$

Pour i=1, ..., N_r et j=1, ... , N_ϕ où N_r et N_ϕ représentent respectivement le nombre de niveaux de quantification pour l'amplitude r et la phase ϕ . L'une de ces régions est représentée par la figure suivante :

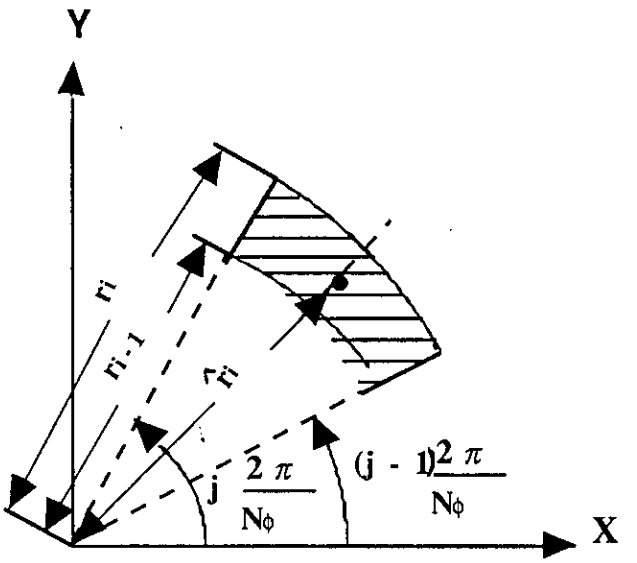


Figure 3.1: Représentation d'une région de quantification polaire.

L'erreur quadratique moyenne relative à ce quantificateur est donnée par.

$$D = \sum_{i=1}^{N_r} \sum_{j=1}^{N_\phi} \int_{r_{i-1}}^{r_i} \int_{\phi_{j-1}}^{\phi_j} \left| re^{j\phi} - \hat{r}_i e^{j\hat{\phi}_j} \right|^2 f(r).\ g(\phi)\ d\phi\ dr \qquad (3.8)$$

La contrainte relative à ce quantificateur est donnée par la représentation du nombre de niveaux N du quantificateur par le produit du nombre de niveaux de quantification de l'amplitude par celui de phase. Autrement dit:

$$N = N_r \times N_\phi \qquad (3.9)$$

Maintenant, il serait intéressant de trouver les seuils de decision et les niveaux de quantification pour lesquels le quantificateur SPQ serait optimal. Pour cela, il faut trouver la solution aux conditions nécessaires suivantes [28]:

$$\frac{\partial D}{\partial \phi_j} = 0 \tag{3.10}$$

$$\frac{\partial D}{\partial \tilde{\phi}_j} = 0 \tag{3.11}$$

$$\frac{\partial D}{\partial r_i} = 0 \tag{3.12}$$

$$\frac{\partial D}{\partial \tilde{r}_i} = 0 \tag{3.13}$$

a. Optimisation de D par rapport à la phase:

Avant d'optimiser D par rapport à la phase. Il est nécessaire de developper son expression donnée par la formule (3.8). Ainsi, de premier abord on obtient:

$$D = \sum_{i=1}^{N_r} \sum_{j=1}^{N_\phi} \int_{r_{i-1}}^{r_i} \int_{\phi_{j-1}}^{\phi_j} \left[r^2 + \tilde{r}_i^2 - 2\, r.\, \tilde{r}_i .\, \cos\left(\phi - \tilde{\phi}_j\right) \right] \frac{f(r)}{2\pi}\, d\phi.\, dr \tag{3.14}$$

Aprés avoir intégrer par rapport à la phase, l'expression de la distorsion D sera de la forme:

$$D = \sum_{j=1}^{N_\phi} \sum_{i=1}^{N_r} \int_{r_{i-1}}^{r_i} \left[\left(r^2 + \tilde{r}_i^2 \right)\left(\phi_j - \phi_{j-1} \right) - 2\, r.\, \tilde{r}_i \left(\sin\left(\phi_j - \tilde{\phi}_j \right) - \sin\left(\phi_{j-1} - \tilde{\phi}_j \right) \right) \right] \frac{f(r)}{2\pi} .\, dr \tag{3.15}$$

Pour trouver les niveaux de phase optimaux, il faut que:

$$\frac{\partial D}{\partial \tilde{\phi}_j} = 0 \quad \Leftrightarrow \quad 2r.\tilde{r}_i \left[\cos\left(\phi_j - \tilde{\phi}_j \right) + \cos\left(\phi_{j-1} - \tilde{\phi}_j \right) \right] = 0$$

Ce qui veut dire:

$$\phi_j - \tilde{\phi}_j = \phi_{j-1} - \tilde{\phi}_j \tag{3.16}$$

Ces équations représentent celles d'un quantificateur uniforme, donc

En supposant $\phi_0 = 0$, on peut donc écrire :

$$\phi_1 - \phi_0 = \frac{2\pi}{N_\phi}$$

$$\phi_2 - \phi_0 = (\phi_2 - \phi_1) + (\phi_1 - \phi_0) = 2\frac{2\pi}{N_\phi}$$

$$\phi_3 - \phi_0 = 3\frac{2\pi}{N_\phi}$$

.

.

.

.

$$\phi_j - \phi_0 = j\frac{2\pi}{N_\phi}$$

Comme :

$$\phi_j - \hat{\phi}_j = \frac{\pi}{N_\phi} \quad \Rightarrow \quad \hat{\phi}_j = (2j-1)\frac{\pi}{N_\phi}$$

Les seuils et les niveaux optimaux sont donc donnés par les expressions suivantes :

$$\begin{cases} \phi_j = j\,\dfrac{2\pi}{N_\phi} & j = 1, \cdots, N_\phi - 1 \qquad (3.17.a) \\ \hat{\phi}_j = (2j-1)\dfrac{\pi}{N_\phi} & j = 1, \cdots, N_\phi \qquad (3.17.b) \end{cases}$$

b.Optimisation par rapport à l'amplitude :

En remplaçant les seuils et les niveaux de la phase par leurs expressions optimales données en (3.17) dans l'expression (3.15) de la distorsion, on trouve :

$$D=\sum_{i=1}^{N_r}\int_{r_{i-1}}^{r_i}\left[r^2+\hat{r}_i^2-2r.\hat{r}_i\,\mathrm{sinc}\left(\frac{1}{N_\phi}\right)\right]f(r).dr \qquad (3.18)$$

où

$$\mathrm{sinc}\left(\frac{1}{N_\phi}\right)=\frac{N_\phi}{\pi}\cdot\sin\left(\frac{\pi}{N_\phi}\right)$$

Pour trouver les seuils et les niveaux optimaux pour l'amplitude, il faut appliquer respectivement les expressions (3.12) et (3.13).
A partir de l'équation (3.13), on trouve :

$$\hat{r}_i=\mathrm{sinc}\left(\frac{1}{N_\phi}\right)\frac{\int_{r_{i-1}}^{r_i} r.f(r)dr}{\int_{r_{i-1}}^{r_i} f(r)dr} \qquad i=1,\cdots,N_r \qquad (3.19)$$

Pour trouver les seuils optimaux, il faut donc appliquer l'équation (3.12). Ainsi, pour N_r fixe, pour minimiser la distorsion D, il faut satisfaire la condition suivante [6]:

$$\frac{\partial D}{\partial r_i}=\left[g(r_i-\hat{r}_{i+1})-g(r_i-\hat{r}_i)\right]f(r_i)=0 \qquad i=1,\ldots,N_r-1 \qquad (3.20)$$

avec

$$g(r_i-\hat{r}_{i+1})=r_i^2+\hat{r}_{i+1}^2-2r_i.\hat{r}_{i+1}\mathrm{sinc}\left(\frac{1}{N_\phi}\right)$$

et

$$g(r_i-\hat{r}_i)=r_i^2+\hat{r}_i^2-2r_i\hat{r}_i\mathrm{sinc}\left(\frac{1}{N_\phi}\right)$$

Ce qui donne :

$$r_i = \frac{\bar{r}_i + \bar{r}_{i+1}}{2\,\mathrm{sinc}\left(\frac{1}{N_\phi}\right)} \qquad i = 1,\ldots,N_r - 1 \tag{3.21}$$

Si on définit :

$$\bar{r}_i = \frac{1}{\mathrm{sinc}\left(\frac{1}{N_\phi}\right)} r_i \qquad i = 1,\ldots,N_r \tag{3.22}$$

On peut donc remplacer les expressions (3.21) et (3.19) par :

$$\begin{cases} r_i = \dfrac{\bar{r}_i + \bar{r}_{i+1}}{2} & i = 1,\cdots,N_r - 1 \quad (3.23.a) \\ \bar{r}_i = \dfrac{\int_{r_{i-1}}^{r_i} r.f(r)dr}{\int_{r_{i-1}}^{r_i} f(r)dr} & i = 1,\cdots,N_r \quad (3.23.b) \end{cases}$$

c. Calcul de la distorsion optimale :

Les niveaux de sortie $\bar{r}_i$ sont les centroïdes relatifs à la densité de Rayleigh, dans les régions de décision $r_{i-1} \le r < r_i$. Pour un nombre de niveaux d'amplitude N_r fixe, on peut résoudre le problème des seuils r_i et des niveaux $\bar{r}_i$ sans prendre connaissance du nombre de niveaux de phase N_ϕ données dans les équations (3.19) et (3.21). Donc, quelque soit le nombre N_ϕ donné, on peut toujours déterminer les niveaux d'amplitude à partir de l'équation (3.23). En effet, les expressions (où $r_0=0$ et $r_{N_r} = +\infty$) sont les mêmes équations que celles d'un quantificateur scalaire dont la source suit une loi de Rayleigh. Or, ce type de quantificateur a été calculé par Pearlman et Senge [26] qui ont tabulé leurs résultats pour Nr=1 à 64. Donc, ces résultats peuvent être directement utilisés pour r_i et $\bar{r}_i$. On peut aussi utiliser les distorsions minimales relatives à la loi de Rayleigh dans l'expression de la distorsion du quantificateur SPQ.

L'expression de la distorsion du quantificateur donné par Pearlman est la suivante [28] :

$$D_r = \sum_{i=1}^{N_r} \int_{r_{i-1}}^{r_i} (r - \bar{r}_i)^2 f(r) dr \qquad (3.24)$$

Donc, en utilisant l'expression ci-dessus ainsi que la formile (3.22) dans l'expression (3.18), on trouve :

$$C(N_r) = \sum_{i=1}^{N_r} \bar{r}_i^2 \int_{r_{i-1}}^{r_i} f(r) dr$$

Remarque :

Il est maintenant évident que pour un Nr donné, il est seulement nécessaire de calculer les seuils, les niveaux de sortie, la distorsion de l'amplitude et la quantité C(Nr) pour la densité de Rayleigh f(r) seulement. dans la limite où Nr approche l'infini, on a:

$$\lim_{N_r \to +\infty} C(N_r) = \int_0^{+\infty} r^2 f(r) dr = 2\sigma^2$$

Ainsi, quelque soit N_ϕ tel que $N_\phi = N/N_r$, on peut substituer dans l'équation (3.25) pour trouver la distorsion totale pour la quantification polaire d'un signal gaussien. Fleisher [4] a pu trouver la condition suffisante de minimisation de Dr, selon laquelle f(r) doit être convexe.

La condition suffisante de minimalité est assurée par la matrice des dérivés du second ordre par rapport aux seuils et aux niveaux qui doit être définie positive.

Finalement, on peut présenter l'organigramme qui permet de calculer les différents paramètres du quantificateur à la fin de ce chapitre.

3.2. Quantificateur SPQ uniforme :

Le quantificateur SPQ uniforme utilise un pas de quantification uniforme pour chaque seuil et niveau d'amplitude: en d'autres termes, les seuils et les niveaux sont espacés d'une façon égale les uns des autres à l'exception du dernier intervalle qui est semi-infini.

Quant à la phase, elle est déjà quantifiée de manière uniforme d'après l' équation (3.18).

L'expression de la distorsion du quantificateur donné par Pearlman est la suivante [28] :

$$D_r = \sum_{i=1}^{N_r} \int_{r_{i-1}}^{r_i} (r - \bar{r}_i)^2 f(r) dr \qquad (3.24)$$

Donc, en utilisant l'expression ci-dessus ainsi que la formile (3.22) dans l'expression (3.18), on trouve :

$$D = D_r + \left(1 - \mathrm{sinc}^2\left(1/N_\phi\right)\right) C(N_r) \qquad (3.25)$$

où $$C(N_r) = \sum_{i=1}^{N_r} \bar{r}_i^2 \int_{r_{i-1}}^{r_i} f(r) dr$$

Remarque :

Il est maintenant évident que pour un Nr donné, il est seulement nécessaire de calculer les seuils, les niveaux de sortie, la distorsion de l'amplitude et la quantité C(Nr) pour la densité de Rayleigh f(r) seulement. dans la limite où Nr approche l'infini, on a:

$$\lim_{N_r \to +\infty} C(N_r) = \int_0^{+\infty} r^2 f(r) dr = 2\sigma^2$$

Ainsi, quelque soit N_ϕ tel que $N_\phi = N/N_r$, on peut substituer dans l'équation (3.25) pour trouver la distorsion totale pour la quantification polaire d'un signal gaussien. Fleisher [4] a pu trouver la condition suffisante de minimisation de Dr, selon laquelle f(r) doit être convexe.

La condition suffisante de minimalité est assurée par la matrice des dérivés du second ordre par rapport aux seuils et aux niveaux qui doit être définie positive.

Finalement, on peut présenter l'organigramme qui permet de calculer les différents paramètres du quantificateur à la fin de ce chapitre.

3.2. Quantificateur SPQ uniforme :

Le quantificateur SPQ uniforme utilise un pas de quantification uniforme pour chaque seuil et niveau d'amplitude; en d'autres termes, les seuils et les niveaux sont espacés d'une façon égale les uns des autres à l'exception du dernier intervalle qui est semi-infini.

Quant à la phase, elle est déjà quantifiée de manière uniforme d'après les équations (3.18).

La minimisation de D par rapport à h revient donc à trouver la solution à l'équation (3.18):

$$\frac{\partial^2 D}{\partial h^2} = 2\sum_{i=1}^{N_r}\left(i-\frac{1}{2}\right)^2 Q(i) - \left(2-\mathrm{sinc}\left(\frac{1}{N_\phi}\right)\right) h \sum_{i=1}^{N_r-1} i^2 . f(ih) \qquad (3.28)$$

soit positive.

3.2.2 Commentaire :

Le pas h optimal dans l'équation (3.27) peut être trouvé ou calculé par la méthode de Newton-Raphson.

Bien que l'implémentation du quantificateur uniforme est simple à faire, car il suffit de trouver le pas h pour trouver tous les paramètres du quantificateur; elle est plus difficile que celle du quantificateur non uniforme suivant un aspect : il s'agit du fait que le pas optimal est une fonction aussi bien de N_r que de N_ϕ comme on peut le voir dans l'équation (3.27). Pour des valeurs élevées de N_ϕ, on peut faire l'approximation suivante :

$$\mathrm{sinc}\left(\frac{1}{N_\phi}\right) \approx 1$$

Avec cette approximation, la précision est de l'ordre de 5.10^{-3}, alors que sans approximation, la précision peut aller jusqu'à 5.10^{-7}.

Comme on a déjà vu le quantificateur SPQ non-uniforme, on peut calculer les seuils et les niveaux de sortie pour un certain N_r sans prendre connaissance du nombre N_ϕ. Alors que pour le quantificateur SPQ uniforme, à chaque valeur de N_ϕ, on peut trouver la valeur optimale de h; ce qui demande un temps de calcul énorme. C'est pourquoi, il serait plus intéressant de calculer le pas h optimal pour des valeurs nécessaires de N_r et N_ϕ comme on le verra dans le prochain paragraphe. Enfin, quand le pas h optimal est calculé pour N_ϕ et N_r donnés, la distorsion est calculée à partir de l'expression (3.28), celle-ci ne dépend pas explicitement de N_ϕ.

3.3 Répartition optimale des débits de phase et d'amplitude :

Les méthodes ayant permis de trouver les seuils et les niveaux optimaux pour le

quantificateur SPQ uniforme et non-uniforme, ne révèlent pas comment sélectionner le nombre de niveaux de phase N_ϕ et le nombre de niveaux d'amplitude Nr, qui produisent la plus petite distorsion possible; quelque soit le nombre total de niveaux N.
On notera qu'on donne à la répartition des débits entre la phase et l'amplitude le nom de factorisation.

Notre but maintenant est de trouver ces choix optimaux de N_ϕ et N_r. Le seul résultat exact publié par Gallengher [6] est le cas où N=396. Ce dernier a trouvé que N_r=12 et N_ϕ=33 (c-à-d: N_ϕ/N_r = 2.76). La distorsion minimale normalisée trouvée à partir de ce cas est égale à 5.96×10^{-3} pour le quantificateur non-uniforme et 6.67×10^{-3} pour le quantificateur uniforme. Ces valeurs de la distorsion sont plus faibles que celles obtenues par Max[1] pour les parties réelles et imaginaires. Plusieurs auteurs [5, 6, 8, 41] ont noté dans leurs écrits, qu'en quantification la phase est quantifiée plus finement que l'amplitude. Parmi eux, il y a Powers [28] qui a obtenu des solutions sous-optimales (pour des seuils et des niveaux) telles que les valeurs de N ne dépassent pas 400; pour cela, il a divisé les seuils d'amplitude de manière équiprobable. Il a pu trouver que le nombre de niveaux de phase N_ϕ est plus grand que N_r et que le rapport minimal est de 2.46.

Nous allons maintenant démontré une méthode formulée par Pearlman et Gray [8] qui utilise le théorie de la distorsion.

Démonstration de la méthode :

Nous allons noter quelques points importants pour la compréhension de ce qui suit:

1°/ Comme nous le savons, le vecteur que nous avons à quantifier est la variable complexe $z=r.e^{j\phi}$. La mesure de distorsion associée à cette quantification est la distance euclidienne :

$$d(z,\hat{z})=\left|z-\hat{z}\right|^2=r^2+\hat{r}^2-2r.\hat{r}\cos(\phi-\hat{\phi}) \qquad (3.29)$$

2°/ Les résultats classiques de la théorie de la distorsion exigent la mesure de distorsion utilisée par un vecteur z(x,y) puisse se mettre sous la forme de la somme de deux mesures de type différence.

$$d(z,\hat{z})=d((x,\hat{x}),(y,\hat{y}))=d_x(\hat{x}-x)+d_y(\hat{y}-y) \qquad (3.30)$$

Comme nous pouvons le constater, les points 1 et 2 sont incompatibles. La distance

La minimisation de D par rapport à h revient donc à trouver la solution à l'équation (3.27) : Pour que cette dernière soit nécessaire et suffisante, il faut que:

$$\frac{\partial^2 D}{\partial h^2} = 2\sum_{i=1}^{N_p}\left(i-\frac{1}{2}\right)^2 Q(i) - \left(2-\mathrm{sinc}\left(\frac{1}{N_\phi}\right)\right) h\sum_{i=1}^{N_r-1} i^2 . f(ih) \qquad (3.28)$$

soitpositive.

3.2.2 Commentaire :

Le pas h optimal dans l'équation (3.27) peut être trouvé ou calculé par la méthode de Newton Raphson.

Bien que l'implémentation du quantificateur uniforme est simple à faire, car il suffit de trouver le pas h pour trouver tous les paramètres du quantificateur; elle est plus difficile que celle du quantificateur non uniforme suivant un aspect : il s'agit du fait que le pas optimal est une fonction aussi bien de N_r que de N_ϕ comme on peut le voir dans l'équation (3.27). Pour des valeurs élevées de N_ϕ, on peut faire l'approximation suivante :

$$\mathrm{sinc}\left(\frac{1}{N_\phi}\right) \approx 1$$

Avec cette approximation, la précision est de l'ordre de 5.10^{-3}, alors que sans approximation, la précision peut aller jusqu'à 5.10^{-7}.

Comme on a déjà vu le quantificateur SPQ non-uniforme, on peut calculer les seuils et les niveaux de sortie pour un certain N_r sans prendre connaissance du nombre N_ϕ. Alors que pour le quantificateur SPQ uniforme, à chaque valeur de N_ϕ, on peut trouver la valeur optimale de h; ce qui demande un temps de calcul énorme. C'est pourquoi, il serait plus intéressant de calculer le pas h optimal pour des valeurs nécessaires de N_r et N_ϕ comme on le verra dans le prochain paragraphe. Enfin, quand le pas h optimal est calculé pour N_ϕ et N_r donnés, la distorsion est calculée à partir de l'expression (3.18), celle-ci ne dépend pas explicitement de N_ϕ.

3.3 Répartition optimale des débits de phase et d'amplitude :

Les méthodes ayant permis de trouver les seuils et les niveaux optimaux pour le

a. Calcul des limites inférieures de Shannon :

• Pour la variable U=Ln r :

Nous allons utiliser dans toute cette partie la formule donnant la limite inférieure de Shannon associée à un critère quadratique (mesure de distorsion euclidienne) [41].

$$D_U^{(L)}(R_U) = \frac{1}{2\pi e} e^{2\left|h(U) - R_U\right|} \qquad R_U \geq 0$$

Or, nous savons calculer, grâce au chapitre 1, l'entropie différentielle qui est :

$$h(U) = 1 + C - \ln 2 - R_U$$

où C est une constante telle que [8] : C=0.5772156649

Il vient donc :

$$D_U^{(L)}(R_U) = \frac{1}{2\pi e} e^{2\left|1 + C - \ln 2 - R_U\right|}$$

En introduisant $k_C = eC$, on a :

$$D_U^{(L)}(R_U) = \frac{k_C^2 \, e}{8\pi} e^{2-R_U} \qquad R \geq 0 \qquad (3.35)$$

• Pour la phase :

L'entropie différentielle de ϕ s'écrit :

$$h(\phi) = \ln(2\pi)$$

La formule de la limite inférieure de Shannon donnée dans le chapitre 1 et appliquée à la phase s'écrit donc :

$$D_\phi^{(L)}(R_\phi) = \frac{2\pi}{e} e^{2 - R_\phi} \qquad R_\phi \geq 0 \qquad (3.36)$$

donnée au point 1 ne peut se mettre sous la forme prescrite au point 2.
Pour lever cette contradiction, Pearlman et Gray [8] ont proposé la quantification de Log z au lieu de celui de $z=r.e^{j\phi}$.
Le logarithme complexe de z s'écrit en effet :

$$\begin{cases} \log z = \log r + j\phi \\ \phi = \theta + 2m\pi \qquad (m \in N) \end{cases} \tag{3.31}$$

L'idée consiste donc à remplacer le couple (r,ϕ) par le couple (U,ϕ). La distinction entre ϕ et θ n'apparaissant pas fondamentale, nous n'utiliserons par la suite que ϕ. L'avantage immédiat de cet artifice est que la mesure euclidienne assurée à log z est simple :

$$\left|\log z - \log \hat{z}\right|^2 = \left|U - \hat{U}\right|^2 + \left|\phi - \hat{\phi}\right|^2 \tag{3.32}$$

Voila donc les points 1 et 2 reconciliés.

Comme on a vu, r et ϕ sont deux variables aléatoires statistiquement indépendantes. Nous savons que le quantificateur du couple (r,ϕ) peut se mettre sous la forme de deux quatificateurs indépendants : un pour r et un pour ϕ.
La donnée ci-dessus jointe à l'expression (3.30) rend le théorème suivant [41] :

Théorème

La limite inférieure de Shannon pour le couple (U,ϕ) associée à la mesure de distorsion (3.30) est la somme de la limite de Shannon de U considérée seule et de ϕ considérée seule

$$D^{(L)}(R_U + R_\phi) = D^{(L)}(R_U) + D^{(L)}(R_\phi) \tag{3.34}$$

R_U et R_ϕ représentent respectivement les débits de l'amplitude U et de la phase ϕ.

Grâce à cette notable simplification, il va nous être facile de calculer explicitement les termes de la formule (3.33) afin de déboucher directement sur le résultat cherché, à savoir les débits optimaux R_U et R_ϕ.

$$\left\{\frac{\partial D^{(1)}(R)}{\partial t}\right\}_{t=t_0} = 0 \tag{3.39}$$

Il vient :

$$t_0 = \frac{1}{2R} Ln\left[\frac{k_c e}{4\pi}\right] + \frac{1}{2} \tag{3.40}$$

En reportant cette valeur dans (3.38), on obtient :

$$\begin{cases} R_U = t_0 \cdot R = \frac{1}{2}|R - (Ln(4\pi) - C - 1)| \\ R\phi = (1 - t_0) \cdot R = \frac{1}{2}|R + (Ln(4\pi) - C - 1)| \end{cases} \tag{3.41}$$

Comme les débits sont positifs, cela impose une condition au débit total R.

$$R > Ln\,(4\pi) - C - 1 = 0.954 \text{ nats/v.c} = 1.37 \text{ bits/v.c}$$

En d'autres termes, pour un débit inférieur à 1.37 bits/v.c, le débit à allouer aux amplitudes est nul.
A partir du système donné en (3.41), on peut déduire : $R_U - R_\phi = 1.37$ bits/v.c, ce qui donne :

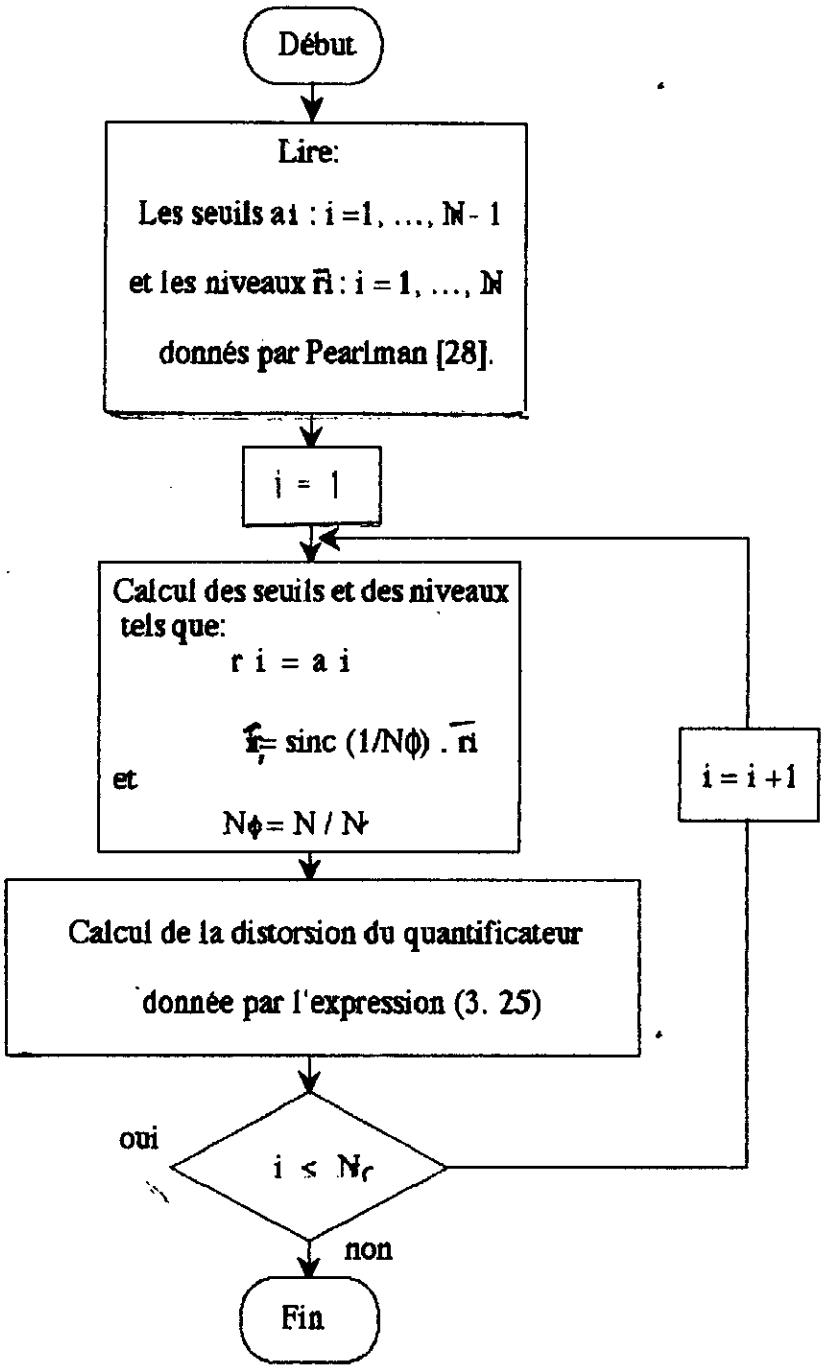
$$N_\phi \approx 2.6\, N_r \tag{3.42}$$

Donc le rapport optimal entre le nombre de niveaux de quantification de la phase et de l'amplitude est 2.6

Conclusion :

On peut dire que la quantificateur SPQ est un type de quantificateur polaire fondamental qu'on ne peut ignorer, bien que les résultats fournis par son algorithme (comme on le verra ultérieurement) sont faibles par rapport à ceux du quantificateur rectangulaire [1] à bas débit. En contre-partie, le calcul de son algorithme est très simple; ce qui a fait que le temps d'exécution est très faible.

organigramme de calcul d'un quantificateur SPQ non-uniforme:



CHAPITRE 4

QUANTIFICATEUR POLAIRE UPQ

INTRODUCTION:

Comme on a vu dans le précédent chapitre, une partie des difficultés pour les quantificateurs polaires de type SPQ est la factorisation; c'est à dire le choix de N_r et N_ϕ sachant que $N=N_r.N_\phi$. La caractéristique de ce quantificateur est qu'à chaque niveau de quantification du module, on a un nombre fixe de niveaux de phase. Des chercheurs ont exprimé que ce type de contrainte ne constitue pas une utilisation optimale, particulièrement vers l'origine où la densité relative aux régions de quantification est plus grande que vers l'exterieur.

C'est pourquoi l'approche qu'a trouvé Wilson [13] est de donner plus de liberté dans le choix du nombre de niveaux de phase pour chaque niveau de quantification du module; ce qui permet à l'erreur quadratique moyenne d'être plus faible au frais d'un processus de quantification plus compliqué. C'est l'une des raisons qui fait que cet algorithme est utilisé pour des faibles débits.

4.1. DESCRIPTION DU QUANTIFICATEUR UPQ:

Soit un coûple de variables indépendantes Gaussiennes de moyenne nulle et de variance unité telle que sa densité de probabilité est la suivante:

$$P_{X,Y}(x,y) = \frac{1}{2\pi} \exp[-(x^2+y^2)/2] \qquad (4.1)$$

$$\text{pour } -\infty < x \ , \ y < +\infty$$

La densité polaire correspondante est donnée par :

$$P_{r,\phi}(r,\phi) = \frac{r}{2\pi}\exp(-r^2/2) \qquad 0 \le r < +\infty \qquad (4.2)$$

$$0 \le \phi < 2\pi$$

où

$$r = \left(x^2 + y^2 \right)^{1/2}$$

$$\phi = \text{Arctg}\left(\frac{y}{x}\right) \qquad (4.3)$$

Suggeré par la symétrie circulaire prouvée par (4.1), le plan est divisé en N régions. Ces régions sont formées d'anneaux divisés en secteurs équiangulaires, comme le montre la figure (4.1). Les limites de chaque régions sont formées par des courbes de rayons constants et d'angles constants.

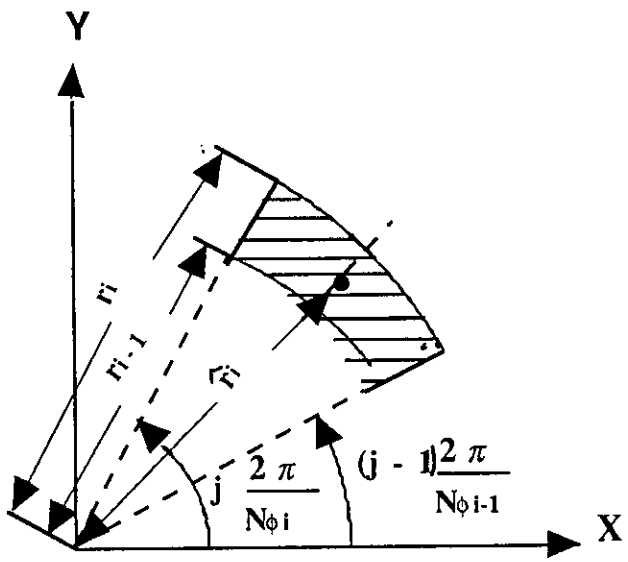


Figure 4.1- Schéma représentant la région de quantification et les contours la délimitant.

Pour obtenir les N régions de quantifications, on doit premièrement diviser l'intervalle du module ([0 , +∞ [ ) en N_r niveaux. Les rayons de ces anneaux sont tels que:

$$0 < r_1 < r_2 < \ldots < r_{Nr} = +\infty$$

Pour chaque anneau relatif au module, indexé par i=1, 2, .., N_r . on permet d'avoir $N_{\phi i}$ régions de phase égale dont la limite est donnée par:

$$(j-1)\frac{2\pi}{N_{\phi i}} \le \phi \le j\frac{2\pi}{N_{\phi i}} \qquad j=1,..,N_{\phi i} \qquad (4.4)$$

La contrainte est la suivante:

$$\sum_{i=1}^{N_r} N_{\phi_i} = N \qquad (4.5)$$

Avec cette contrainte, une liberté complète est permise dans l'attribution du nombre de régions pour chaque anneau. On appelle ce type de conception *unrestrected polar quantizer* (UPQ), inversement au SPQ (*Strictly polar quantizer)* étudié précédemment.

Ce type de quantificateur impose $N_{\phi i} = N_{\phi}$ pour chaque niveau du module tel que $N=N_r.N_{\phi}$. Cette conception a montré ses limites.Par exemple, N=16 peut-être factorisé pour $N_r = 1, 2, 4, 8, 16$ et respectivement $N_{\phi}=16, 8, 4, 2, 1$.Cependant, comme on le verra, le quantificateur UPQ avec trois niveaux de quantification (pour le module) se révèle optimal. il s'agit de la constellation (1, 6, 9).

On remarquera que pour cet algorithme, quand on permet plus de liberté, cela n'impliquera pas automatiquement que le quantificateur est optimal; quelque soit le nombre total de niveaux N. Néanmoins, il est plus facile à configurer et à implementer pour une certaine valeur de N que pour une autre.

Pour ce faire, la détermination de la région indexée n est obtenue comme suit:

Examinons une table de couples (r_n , ϕ_n) ordonnée par ordre de croissance de r_n. Pour un r_n fixe, la table est donnée par ordre de croissance de ϕ_n

Soit les couples (r_n , ϕ_n) représentants respectivement les limites supérieures du module et de la phase pour la région n. Tant que $r_n \geq r$, nous avons localisé l'anneau. Par la suite, à l'intérieur de cette section, nous notons les ϕ_n jusqu'à ce que $\phi \geq \phi_n$. Sur quoi, la région escomptée est localisée.

La notation utilisée pour décrire une configuration particulière est, $(N,N_r, N_{\phi 1},..., N_{\phi Nr})$. Elle dénote le quantificateur à N régions, dont N_r niveaux de module, et $N_{\phi 1},..., N_{\phi Nr}$ nombres de niveaux de phase pour les differents niveaux de module. Ainsi, la constellation (1,6,9) sera notée comme étant la configuration (16,3,1,6,9) qui dénote que le quantificateur a seize régions dans le plan, avec une région au centre, six régions dans le prochain anneau, et neuf régions dans le niveau exterieur comme le montre la figure (4.2).

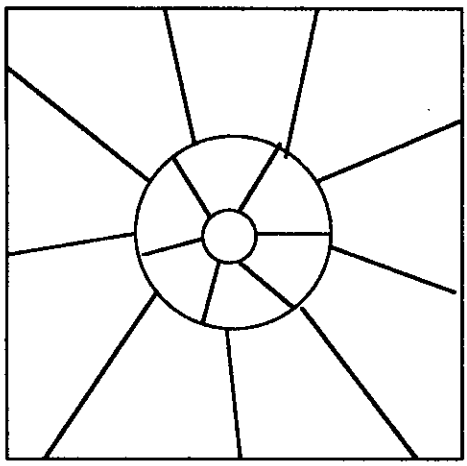


Figure 4.2- Représentation de la configuration optimale (16, 3, 1, 6, 9)

4.2. OPTIMISATION DU QUANTIFICATEUR UPQ

Pour un nombre N donné, il ya évidemment beaucoup de selections possibles pour, N_r et N_{ϕ_i}, correspondants tels que :

$$\sum_{i=1}^{N_r} N_{\phi_i} = N$$

C'est pourquoi, le problème devient embarrassant quand N augmente. Plusieurs de ces sélections sont rejetées carrément car elles ne donnent pas de résultats optimaux. Par exemple, la sélection (5,2.3,2) représente un mauvais choix parmi les autres sélections pour N=5.

La stratégie d'optimisation est simple. On s'attribue un ensemble $(N, N_r, N_{\phi 1}, \ldots, N_{\phi_{Nr}})$ qui vérifie la contrainte. On sectionne les N_r-1 seuils du module ($r_{Nr} = \infty$), les niveaux r_i (i = 1, 2, ..., N_r,) et on fait de même pour la phase. Finallement, on minimise l'erreur quadratique moyenne D pour cette configuration particulière. Une autre sélection est choisie et optimisée. Par une recherche exhaustive, le quantificateur UPQ, optimal est déterminé.

Maintenant, considérons une certaine configuration$(N, N_r, N_{\phi_1}, \ldots, N_{\phi_{Nr}})$. L'erreur quadratique moyenne est de la forme:

$$D = \sum_{n=1}^{N} \int_{\phi} \int_{r}^{} \left| re^{j\phi} - \hat{r}_n e^{j\hat{\phi}_n} \right|^2 p(r,\phi)\, dr\, d\phi \qquad (4.6)$$

où $(\hat{r}_n , \hat{\phi}_n)$ sont respectivement les niveaux de quantification du module et de la phase relatifs à la région $\Re_n$.

Le choix des limites de phase est fixé par la spécification de N_{ϕ_i}, c'est à dire de la condition (4.4), de plus on sait que $\hat{\phi}_n$ doit-être à égale distance entre les deux limites de phase. Le nombre de variables indépendantes est donc réduit à $N_r - 1$ valeurs de r_i et à N_r valeurs de $\hat{r}_i$.
Les conditions nécessaires d'optimisation sont données par la différenciation par rapport à r_i et $\hat{r}_i$ [13]. Le résultat est généralisé au cas du quantificateur UPQ. On aura donc:

$$\hat{r}_i = \frac{N_{\phi_i}}{\pi} \left(\sin \frac{\pi}{N_{\phi_i}} \right) \frac{\int_{r_{i-1}}^{r_i} r\, f(r)\, dr}{\int_{r_{i-1}}^{r_i} f(r)\, dr} \qquad (4.7)$$

pour: $i = 1 , \ldots , N_r$

$$\frac{\partial D}{\partial r_i} = 2 r_i \left[\left(\frac{N_{\phi_{i+1}}}{\pi} \sin \frac{\pi}{N_{\phi_{i+1}}} \right) \hat{r}_{i+1} - \left(\frac{N_{\phi_i}}{\pi} \sin \frac{\pi}{N_{\phi_i}} \right) \hat{r}_i \right] - \hat{r}_{i+1}^2 + \hat{r}_i^2 = 0 \qquad (4.8)$$

pour: $i = 1 , \ldots , N_r - 1$

Pour le cas d'un quantificateur polaire de type SPQ, où $N_{\phi_i} = N_\phi$, pour trouver les seuils et les niveaux, il suffisait d'exploiter les résultats du quantificateur polaire à une dimension; or dans notre cas, cette méthode ne peut plus être appliquée car N_{ϕ_i} varie en fonction de i. C'est pourquoi, wilson a fait un algorithme numerique iteratif.

Au début, on prend un vecteur d'essai $(r_1, \ldots , r_{Nr-1})$. En utilisant (4.7), on calcule les valeurs de r_i correspondantes qui sont optimales pour ce vecteur d'essai. On tiend à remarquer que le calcul des intégrales s'est fait par la méthode de quadrature de Gauss (annexe 3); celle-ci donne quatre chiffres décimaux exacts.

L'équation (4-8) est utilisée comme un gradient de correction du vecteur $(r_1, \ldots , r_{Nr-1})$ On aura donc:

$$\begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_{N_r-1} \end{bmatrix}_{k+1} = \begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_{N_r-1} \end{bmatrix}_{k} + C \begin{bmatrix} \frac{\partial D}{\partial r_1} \\ \cdot \\ \cdot \\ \frac{\partial D}{\partial r_{N_r-1}} \end{bmatrix} \tag{4.9}$$

C est une constante choisie empiriquement.

Aprés la convergence de l'algorithme, on calcule l'erreur quadratique moyenne qui est exprimée par:

$$D = 2 + \hat{r}_1^2\left[1-e^{-r_1^2/2}\right] + \hat{r}_2^2\left[e^{-r_1^2/2} - e^{-r_2^2/2}\right] + \ldots + \hat{r}_{N_r-1}^2\left[e^{-r_{N_r-1}^2/2}\right]$$
$$+ \frac{-2\,N_{\phi_1}\,\hat{r}_1}{\pi}\sin\left(\frac{\pi}{N_{\phi_1}}\right)\int_0^{r_1} r^2\,e^{-r^2/2}\,dr$$
$$+ \ldots + \frac{-2\,N_{N_r}\,\hat{r}_{N_r}}{\pi}\sin\left(\frac{\pi}{N_{\phi_{N_r}}}\right)\int_{r_{N_r-1}}^{+\infty} r^2\,e^{-r^2/2}\,dr \tag{4.10}$$

On tire alors l'organigramme se trouvant à la fin du chapitre.

CONCLUSION

Comme on l'a étudié dans le présent chapitre pour le cas du quantificateur UPQ, le nombre de région de quantification varie d'un niveau (de module) à un autre. Pour avoir un nombre total de N régions, il suffissait qu'il satisfasse à la contrainte:

$$\sum_{i=1}^{N_r} N_{\phi_i} = N$$

C'est ainsi que le quantificateur UPQ est une forme générale du quantificateur SPQ. Donc, l'UPQ doit être plus performant que le SPQ. La figure 4.3 montre les constellations respectives des deux types de quantificateurs pour le cas où N=10.

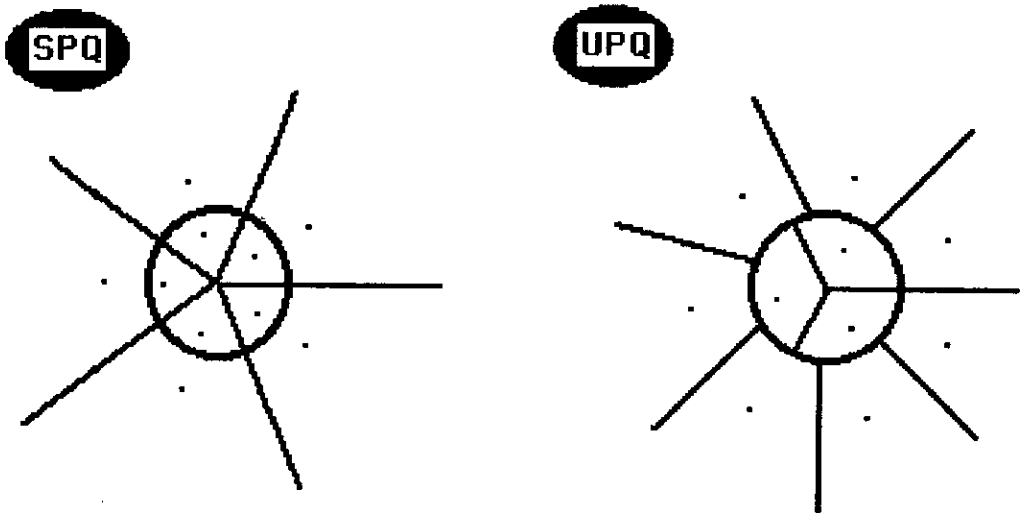
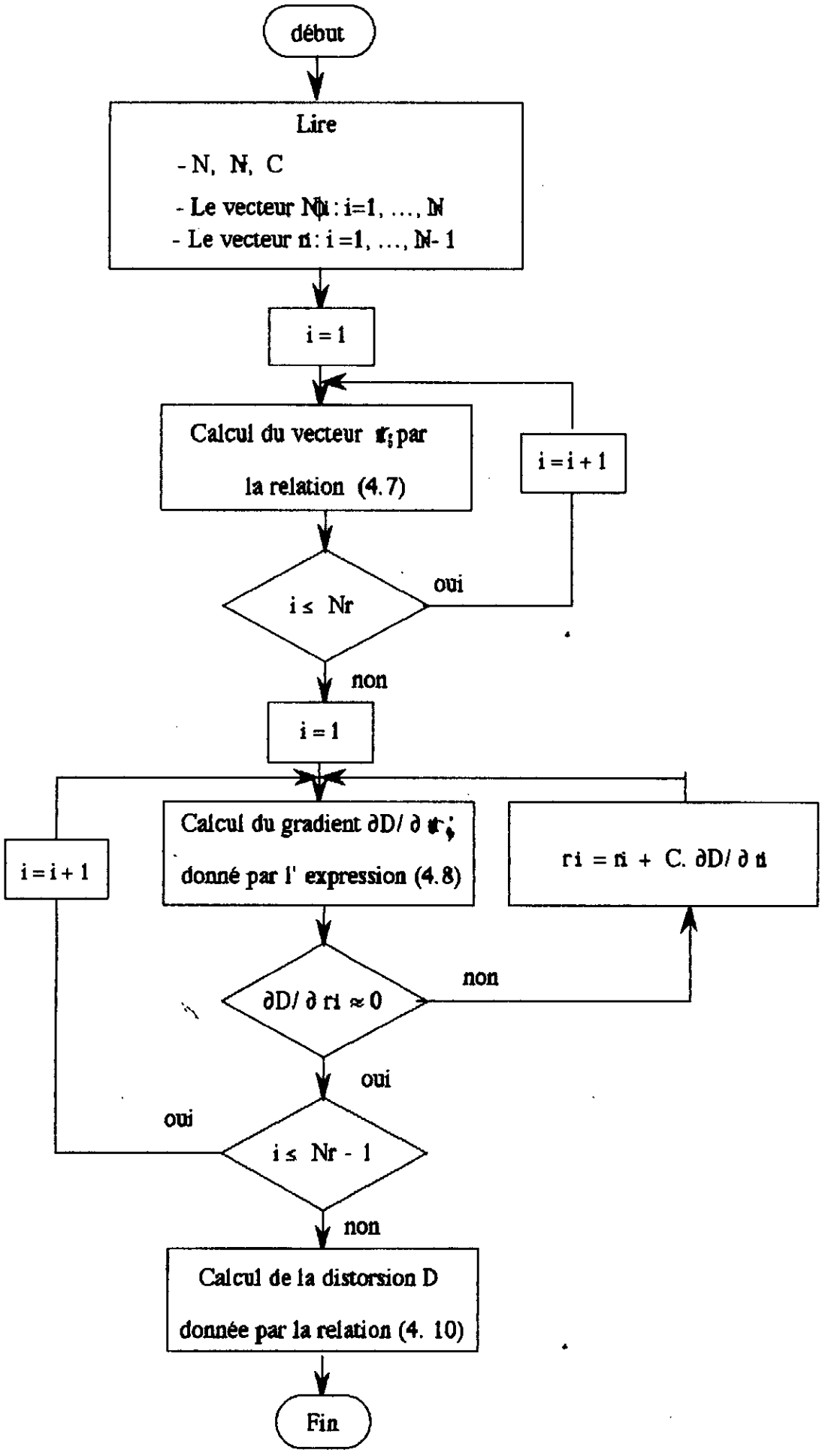


Figure 4.3- Représentation respective des constellations des quantificateurs SPQ et UPQ pour N=10.


début
Lire
- N, N, C
- Le vecteur N : i=1, ..., N
- Le vecteur r : i =1, ..., N- 1
i = 1
Calcul du vecteur par
la relation (4.7)
i = i + 1
i ≤ Nr
oui
non
i = 1
Calcul du gradient ∂D/ ∂
donné par l' expression (4.8)
ri = ri + C. ∂D/ ∂ r
i = i + 1
∂D/ ∂ ri ≈ 0
non
oui
oui
i ≤ Nr - 1
non
Calcul de la distorsion D
donnée par la relation (4. 10)
Fin

CHAPITRE 5

ANALYSE ASYMPTOTIQUE DES PERFORMANCES D'UN QUANTIFICATEUR POLAIRE DE TYPE UPQ

Introduction:

Comme on a déja vu dans les précédents chapitres que le quantificateur polaire de type SPQ est caracterisé par la contrainte $N = N_\phi \times N_r$ où N, N_ϕ et N_r sont respectivement le nombre de niveaux de quantification total, le nombre de niveaux de phase et le nombre de niveaux de quantification de l'amplitude.

Dans le cas du quantificateur polaire de type UPQ, sa contrainte est donnée par $N = \sum N_{\phi i}$ $(i = 1, \ldots, N_r)$.

On peut voir la structure (ou la constellation) de chacun des deux quantificateurs à la figure (4. 3) du précédent chapitre.

On a vu aussi que le quantificateur SPQ ne dépassait pas les performances d'un quantificateur rectangulaire (formé de deux quantificateurs scalaires pour chaque dimension) qu'à haut débit (pour un débit superieur à 6 bits, c'est à dire pour $N > 64$). Pour remédier à cet inconvenient, Wilson [13] a proposé le quantificateur UPQ mais qui possède lui aussi un autre inconvenient; celui-ci se résume au fait que l'algorithme élaboré pour ce type de quantificateur est difficile à éxecuter pour des débits superieurs à 6 bits.

On peut se dire dans ce cas, qu'il serait interessant de commuter entre ces deux types de quantificateurs en passant de bas débit en haut débit; mais il est encore plus interressant d'étudier les caracteristiques asymptotiques du quantificateur UPQ tout en changeant l'algorithme de calcul [34]. On obtiendra ainsi un quantificateur UPQ "asymptotique" qu'on appelera en anglais quantificateur AUPQ : *"Asymptotic unrestrected polar quantizer"*.

5.1. Formulation mathématique:

5.1.1. Rappels:

Pour cette analyse, on suppose que la source suit une densité continue, symmétrique et circulaire dont les densités marginales en coordonnées rectangulaires sont de variance unité. La fonction densité conjointe est:

$$f(x, y) = p(\sqrt{x^2 + y^2}) \qquad (5.1)$$

En transformant en coordonnées polaires, la phase ϕ = Arctg (y / x) est uniformément distribuée sur l'intervalle [0, 2π [alors que l'amplitude $r = (x^2 + y^2)^{1/2}$ est distribué sur [0,$+\infty$ [avec une fonction densité:

$$f(r) = 2\pi\, p(r) \qquad (5.2)$$

on a vu que le quantificateur UPQ emploie des quantificateurs scalaires séparés pour r et ϕ : un quantificateur non-uniforme pour r sur [0,$+\infty$ [et un quantificateur uniforme pour ϕ sur [0, 2π [.

5.1.2. Positionnement du problème:

Le modèle qu'on utilisera pour l'implémentation du quantificateur AUPQ, est le modèle du compresseur qu'on a déja présenté au chapitre 2 (section 2.2.2).

On rappelera qu'en général avec ce modèle, on remplace chaque quantificateur par une serie de connections de trois éléments (comme on peut le voir sur la figure 5. 1): un compresseur g, un quantificateur uniforme Q_U et un expanseur g^{-1}.

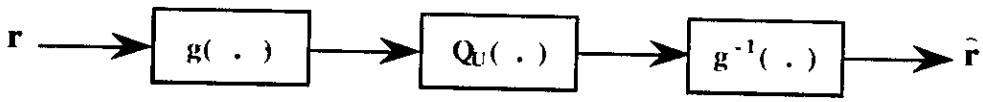


Figure 5. 1 Modèle du compresseur pour un quantificateur scalaire.

Dans notre cas, g représentera le compresseur de l'amplitude. On ne parlera pas de compresseur de phase car le quantificateur relatif à la phase est déja un quantificateur uniforme.

Soient r_i et $\hat{r}_i$ représentant respectivement le $i^{ème}$ seuil et le $i^{ème}$ niveau de sortie d'amplitude où, i = 1, . . . N_r et $r_{N_r+1} = \infty$. Le pas non-uniforme du quantificateur d'amplitude est décrit à travers le compresseur g(r), où:

$$\begin{cases} r_i = g^{-1}\left[\dfrac{(i-1)}{N_r} \right] \\ \hat{r}_i = g^{-1}\left[\dfrac{(2i-1)}{2N_r} \right] \end{cases} \qquad (5.3)$$

De la même manière, soient $\phi_{j,i}$ et $\hat{\phi}_{j,i}$ respectivement le $j^{ième}$ seuil et le $j^{ième}$ niveau de sortie de la phase correspondant (ou se trouvant) au $i^{ème}$ anneau d'amplitude (figure 5. 2) pour j = 1, . . . N_{ϕ_i} et $\phi_{N_{\phi_{i+1}}} = 2\pi$.

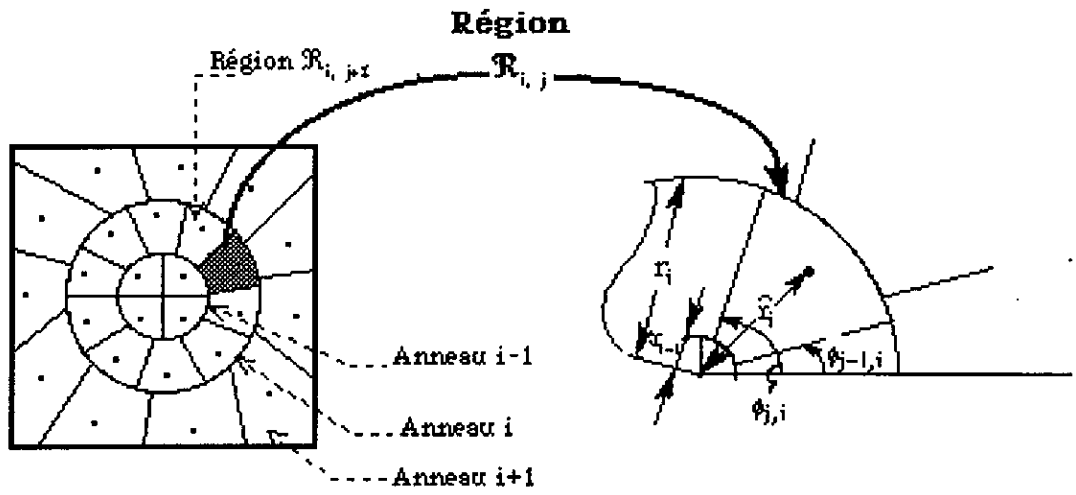


Figure 5.2: Exemple d'une région de quantification $\Re_{i,j}$.

Etant donné que le nombre de niveaux de phase varie d'un anneau à un autre, donc d'une amplitude à un autre, alors on le définira comme une fonction de cette amplitude; ainsi on le notera $N_\phi(r)$. En évaluant $N_\phi(r)$ au $i^{ème}$ anneau de l'amplitude, alors on écrit: $N_\phi(\hat{r}_i) = N_{\phi_i}$ En utilisant cette fonction, on peut écrire que:

$$\begin{cases} \phi_{j,i} = (j-1)\dfrac{2\pi}{N_\phi(\hat{r}_i)} \\ \hat{\phi}_{j,i} = (2j-1)\dfrac{\pi}{N_\phi(\hat{r}_i)} \end{cases} \qquad (5.4)$$

5.1.3. Calcul de la distorsion:

L'expression de l'erreur quadratique moyenne (MSE) pour le quantificateur AUPQ est similaire à celle des autres quanticateurs polaires: donc elle est donnée sous la forme suivante:

$$D = \sum_{i=1}^{Nr} \sum_{i=1}^{N_\phi(\bar{r}_i)} \int_{\phi_{j,i}}^{\phi_{i+1,i}} \int_{r_i}^{r_{i+1}} \left[r^2 + \bar{r}_i^2 - 2 r \bar{r}_i \cos(\phi - \bar{\phi}_{j,i}) \right] \frac{f(r)}{2\pi} \, dr \, d\phi \qquad (5. 5)$$

En supposant N élevé, on peut faire plusieurs approximations pour le calcul de D. Premièrement, on peut écrire que $f(r) = f(\bar{r}_i)$ car chaque intervalle de variation est trés petit (N est élevé); aussi $f(\bar{r}_i)$ varie trés peu pour chaque intervalle.

De même, ϕ diffère trés peu de $\bar{\phi}$ sur chaque région, d'où on peut remplaçer le cosinus par le développement de Taylor:

$$\cos(\phi - \bar{\phi}_{j,i}) \approx 1 - \frac{(\phi - \bar{\phi}_{j,i})^2}{2}$$

On trouvera alors:

$$D = \sum_{i=1}^{N_r} f(\bar{r}_i) \sum_{j=1}^{N\phi(\bar{r}_i)} \int_{\phi_{j,i}}^{\phi_{j+1,i}} \left[\left(r - \bar{r}_i \right)^2 + r \, \bar{r}_i \left(\phi - \bar{\phi}_{j,i} \right)^2 \right] \frac{1}{2\pi} \, dr \, d\phi \qquad (5. 6)$$

En intégrant suivant r et ϕ et connaissant, d'aprés les équations (5. 4) que:

$$\phi_{j+1,i} - \bar{\phi}_{j,i} = \frac{\pi}{N_\phi(\bar{r}_i)}$$

et

$$\phi_{j,i} - \bar{\phi}_{j,i} = - \frac{\pi}{N_\phi(\bar{r}_i)}$$

L'expression de D sera sous la forme suivante :

$$D = \sum_{i=1}^{N_r} f(\bar{r}_i) \left[\frac{\left(r_{i+1} - \bar{r}_i \right)^3 - \left(r_i - \bar{r}_i \right)^3}{3} + \frac{\pi^2}{3 N_\phi^2(\bar{r}_i)} \left(r_{i+1}^2 - r_i^2 \right) \frac{\bar{r}_i}{2} \right] \qquad (5. 7)$$

Pour un nombre de niveaux N_r d'amplitude élevé, on peut approximer:

$$r_{i+1} - \hat{r}_i \approx \hat{r}_i - r_i$$

et

$$r_{i+1} - \hat{r}_i = \frac{g(r_{i+1}) - g(\hat{r}_i)}{g'(\hat{r}_i)}$$

D'aprés les équations (5. 3), on peut écrire donc:

$$\begin{cases} g(r_{i+1}) = \dfrac{i}{N_r} \\ g(\hat{r}_i) = \dfrac{2i - 1}{2N_r} \end{cases}$$

Donc
$$r_{i+1} - \hat{r}_i \approx \hat{r}_i - r_i \approx \frac{1}{2 N_r g'(\hat{r}_i)} \qquad (5.8)$$

Le second terme de D peut-être approximé de la façon suivante:

$$\begin{aligned} r_{i+1}^2 - r_i^2 &= (r_{i+1} - r_i)(r_{i+1} + r_i) \\ &= (r_{i+1} - \hat{r}_i - r_i + \hat{r}_i)(r_{i+1} - \hat{r}_i + r_i - \hat{r}_i + 2\hat{r}_i) \\ &\approx \frac{1}{N_r g'(\hat{r}_i)} 2\hat{r}_i \end{aligned} \qquad (5.9)$$

D'où l'expression de D:

$$D = \sum_{i=1}^{N_r} f(\hat{r}_i) \left[\frac{1}{12 N_r^2 [g'(\hat{r}_i)]^2} + \frac{\pi^2 \hat{r}_i^2}{3 N_\phi^2(\hat{r}_i)} \right] \frac{1}{g'(\hat{r}_i) N_r} \qquad (5.10)$$

A partir de la définition du compresseur de l'amplitude, chaque région de quantification (r_{i+1}, r_i) se transforme en un intervalle de largeur $1/N_r$ dans la région d'un quantificateur uniforme, on aura alors:

$$\frac{1}{N_r} = g(r_{i+1}) - g(r_i) \approx g'(\hat{r}_i)(r_{i+1} - r_i) \approx g'(\hat{r}_i)\Delta_i \qquad (5.11)$$

où Δ_i represente la largeur de la $i^{ème}$ région de quantification de l'amplitude et est donnée par:

$$\Delta_i = \frac{1}{g'(\hat{r}_i) N_r}$$

En approximant les sommes dans D (donnée par l'expression 5. 10) par des intégrales (car $\Delta_i \approx dr$), on aura donc:

$$D = \frac{1}{12\ N_r^2} \int_0^{\infty} \frac{f(r)}{[g'(r)]^2}\ dr + \frac{\pi^2}{3} \int_0^{\infty} \frac{r^2\ f(r)}{N_\phi^2(r)}\ dr \qquad (5.\ 12)$$

Comme le quantificateur AUPQ est un quantificateur UPQ à haut débit, donc la contrainte du quantificateur UPQ est toujours valable; et qui est:

$$N = \sum_{i=1}^{N_r} N_\phi(\bar{r}_i) \qquad (5.\ 13)$$

En divisant les deux membres de (5. 13) par N_r , et en utilisant l'expression (5. 11), on trouve:

$$\frac{N}{N_r} = \sum_{i=1}^{N_r} N_\phi(\bar{r}_i)\ g'(\bar{r}_i)\ \Delta_i \qquad (5.\ 14)$$

Quand N_r est élevé $\Delta_i \approx dr$, alors l'équation (5. 14) deviendrait équivalente à:

$$\frac{N}{N_r} = \int_0^{\infty} N_\phi(r)\ g'(r)\ dr \qquad (5.\ 15)$$

Ainsi, l'équation (5. 15) exprime la contrainte du quantificateur AUPQ.

5. 1. 4. Optimisation de la distorsion:

a/ Optimisation de D par rapport à $N_\phi(r)$:

L'optimisation de la distorsion D donnée par l'expression (5. 12) par rapport au nombre de niveaux de phase est un problème de calcul de variations [19, 20]. Puisque la distorsion D est liée à la contrainte donnée par l'équation (5. 15), il s'agit donc d'appliquer un point particulier de la méthode de calcul de variations et qui est le calcul des extrema liés. Cette méthode sera détaillée en annexe 2. De plus, dans notre cas les fonctions à optimiser sont sous forme de fonctions d'intégrales; alors la méthode qu'on

doit appliquer exactement se trouve dans le dernier paragraphe de l'annexe 2. Celle-ci s'appelle : "*Problème isopérimétrique*".

Ainsi d'aprés ce paragraphe (expression 5.12), l'optimisation (ou la minimisation) de D par rapport à $N_\phi(r)$ revient à chercher l'extremum (ou le minimum) de l'intégrale suivante:

$$I^* = \int_{t_1}^{t_2} F^*(t, x, x')\, dt$$

où $F^* = F + \lambda G$

λ est le multiplicateur de Lagrange.

Dans notre cas, on pose:

- $t \equiv r$
- $x(t) \equiv N_\phi(r)$
- $x'(t) \equiv N'_\phi(r)$
- $t_1 \equiv 0$
- $t_2 \equiv \infty$
- $F \equiv \dfrac{1}{12\, N_r^2} \dfrac{f(r)}{[g'(r)]^2} + \dfrac{\pi^2}{3} \dfrac{r^2\, f(r)}{N_\phi^2(r)}$
- $G \equiv N_\phi(r)\, g'(r)$

On aura alors:

$$I^* = \int_0^{+\infty} \left(\frac{1}{12\, N_r^2} \frac{f(r)}{[g'(r)]^2} + \frac{\pi^2}{3} \frac{r^2\, f(r)}{N_\phi^2(r)} + \lambda\, N_\phi(r)\, g'(r) \right) dr \qquad (5.16)$$

La solution à ce problème est donnée en cherchant la solution à l'équation d'Euler-Lagrange (annexe 2) donnée par l'expression générale suivante :

$$\frac{\partial F^*}{\partial x} - \frac{d}{dt}\left(\frac{\partial F^*}{\partial x'}\right) = 0$$

En voulant appliquer cette équation à notre cas, nous avons remarqué que F^* ne dépendait pas de x' (en d'autres termes de $N'_\phi(r)$); donc l'équation d'Euler-Lagrange se réduit dans notre cas à:

$$\frac{\partial F^*}{\partial x} = 0$$

où $$F^* = \frac{1}{12 N_r^2} \frac{f(r)}{\left[g'(r)\right]^2} + \frac{\pi^2}{3} \frac{r^2 f(r)}{N_\phi^2(r)} + \lambda N_\phi(r) g'(r)$$

et $$x = N_\phi(r)$$

Finalement, la minimisation de D par rapport à $N_\phi(r)$ revient à trouver une solution à l'équation suivante :

$$\frac{\partial}{\partial N_\phi}\left(\frac{1}{12 N_r^2} \frac{f(r)}{\left[g'(r)\right]^2} + \frac{\pi^2}{3} \frac{r^2 f(r)}{N_\phi^2(r)} + \lambda N_\phi(r) g'(r) \right) = 0 \qquad (5.17)$$

La résolution de l'équation (5. 17) donne l'expression de $N_\phi(r)$ suivante:

$$N_\phi(r) = \lambda \frac{r^{2/3} \left[f(r)\right]^{1/3}}{\left[g'(r)\right]^{1/3}} \qquad (5.18)$$

La constante λ (multiplicateur de Lagrange) est calculé à partir de l'équation (5. 15) de la contrainte; on trouve alors:

$$\lambda = \frac{N}{N_r} \frac{1}{\int_0^\infty r^{2/3} \left[f(r)\right]^{1/3} \left[g'(r)\right]^{2/3} dr} \qquad (5.19)$$

d'où:

$$N_\phi(r) = \frac{N}{N_r} \frac{r^{2/3} \left[f(r)\right]^{1/3} \left[g'(r)\right]^{-1/3}}{\int_0^\infty r^{2/3} \left[f(r)\right]^{1/3} \left[g'(r)\right]^{2/3} dr} \qquad (5.20)$$

La formule (5. 20) représente la fonction optimale du nombre de niveaux de phase $N_\phi(r)$ quelque soit la valeur de l'amplitude choisie (pour des débits élevés).

En remplaçant $N_\phi(r)$ par son expression finale dans celle de la distorsion D donnée par (5. 12), on aura:

$$D \approx \frac{1}{12 N_r^2} \int_0^{+\infty} \frac{f(r)}{\left[g'(r)\right]^2} dr + \frac{\pi^2}{3} \frac{N_r^2}{N^2} \left[\int_0^{+\infty} r^{2/3} \left[f(r)\right]^{1/3} \left[g'(r)\right]^{2/3} dr \right]^3 \qquad (5.21)$$

b/ Optimisation de D par rapport au nombre de niveaux d'amplitude N_r:

Pour cela, on doit mettre:

$$\frac{\partial D}{\partial N_r} = 0$$

Comme les expressions sous le signes sommes dans D ne dépendent pas de N_r, alors on peut poser:

$$A = \frac{1}{12} \int_0^{+\infty} \frac{f(r)}{\left[g'(r)\right]^2} dr$$

et

$$B = \frac{\pi^2}{3} \left[\int_0^{+\infty} r^{2/3} \left[f(r)\right]^{1/3} \left[g'(r)\right]^{2/3} dr \right]^3$$

Ainsi D aura la forme suivante:

$$D \approx A N_r^{-2} + B N_r^2 N^{-2} \qquad (5.22)$$

En optimisant la formule (5. 22) par rapport à N_r, on tire l'expression optimale de N_r et qui est:

$$N_r = \left[\frac{A}{B} \right]^{1/4} N^{1/2} \qquad (5.23)$$

En substituant la valeur de N_r dans celle de D donnée en (5. 22), on aura:

$$D \approx 2 A^{1/2} B^{1/2} N^{-1}$$

En remplaçant A et B par leurs expressions respectives, on trouve:

$$D \approx \frac{\pi}{3N} \left[\int_0^{+\infty} \frac{f(r)}{\left[g'(r)\right]^2} dr \right]^{1/2} \left[\int_0^{+\infty} r^{2/3} \left[f(r)\right]^{1/3} \left[g'(r)\right]^{2/3} dr \right]^{3/2} \qquad (5.24)$$

c/ Optimisation de D par rapport à la fonction de compression g(r):

On remarque d'aprés l'équation (5. 24) que la distorsion D possède une forme assez compliquée et est de plus fonction de la dérivée de la fonction de compression g(r); c'est pourquoi, il est difficile d'optimiser D par rapport à g(r) comme on l'a fait pour

$N_\phi(r)$ et N_r. Ainsi, l'idée est de chercher une borne inferieure à D et de calculer la fonction de compression g(r) lorsque D atteint cette borne inferieure; On aura alors trouver la fonction g(r) qui optimise D pour cette borne inferieure.

A cet effet, et d'aprés la forme de l'expression de la distorsion, l'inégalité de Hölder [38] (qui est détaillée en annexe 3) nous a permit de trouver une borne inferieure à D; telle que:

$$D \geq \frac{\pi}{3N}\left[\int_0^{+\infty} r^{1/2}\left[f(r)\right]^{1/2}\,dr\right]^2 \qquad (5.25)$$

Comme on l'a dit précédemment, D sera optimale si elle est égale à sa borne inferieure. C'est à dire:

$$D = \frac{\pi}{3N}\left[\int_0^{+\infty} r^{1/2}\left[f(r)\right]^{1/2}\,dr\right]^2 \qquad (5.26)$$

D'aprés le théorème sur l'inégalité de Hölder (voir annèxe 3), l'inégalité ne devient égalité que si et seulement si:

$$\left[r^{2/3}\left|f(r)\right|^{1/3}\left|g'(r)\right|^{2/3}\right]^{3/4} = k\left[\frac{f(r)}{\left|g'(r)\right|^2}\right]^{1/4} \qquad (5.27)$$

où k est une constante de proportionalité.
A partir de l'équation (5. 27), on aura:

$$g'(r) = k\, r^{-1/4}\left[f(r)\right]^{1/4}$$

Sachant que la fonction g(r) est une fonction de compression donc elle est définie par:

$$g : \left[0, +\infty\right[\mapsto \left[0, 1\right]$$

On peut retrouver l'expression de g(r) et qui est la suivante:

$$g(r) = \frac{\int_0^r s^{-1/4}\left[f(s)\right]^{1/4}\,ds}{\int_0^{+\infty} s^{-1/4}\left[f(s)\right]^{1/4}\,ds} \qquad (5.28)$$

Ainsi, l'expression ci-dessus de g(r) représente sa forme optimale.

En substituant l'expression de g'(r) dans celles de $N_\phi(r)$ et N_r, on trouve respectivement:

$$N_\phi(r) = \sqrt{2\pi}\, N^{1/2}\, \frac{r^{3/4}\,[f(r)]^{1/4}}{\left[\int_0^{+\infty} s^{-1/2}\,[f(s)]^{1/2}\,ds\right]^{1/2}} \qquad (5.29)$$

et

$$N_r = \frac{1}{\sqrt{2\pi}}\, N^{1/2}\, \frac{\int_0^{+\infty} s^{-1/4}\,[f(s)]^{1/4}\,ds}{\int_0^{+\infty} s^{-1/2}\,[f(s)]^{1/2}\,ds} \qquad (5.30)$$

5. 3. Exemple: Application des résultats obtenus à une source Gaussienne:

Si on applique les résultats obtenus à une source bidimensionnelle qui suit une densité de Gauss:

$$f(x, y) = \frac{1}{2\pi}\, e^{-(x^2 + y^2)/2} \qquad (5.31)$$

Les paramètres relatifs au quantificateur AUPQ seront donnés par les expressions suivantes:

5. 3. 1. Le nombre optimal de niveaux de phase $N_\phi(r)$:

$$N_\phi(r) = \sqrt{\pi}\, N^{1/2}\, r\, e^{-r^2/2} \qquad (5.32)$$

5. 3. 2. Le nombre optimal de niveaux d'amplitude N_r:

$$N_r = \left(\frac{N}{2}\right)^{1/2} \qquad (5.33)$$

5. 3. 3. La fonction de compression g(r):

$$g(r) = \int_0^r \frac{1}{\sqrt{2\pi}} e^{-s^2/8}\, ds \qquad (5.34)$$

On remarque que g(r) est une fonction cumulative de la distribution de Gauss de moyenne nulle et de variance égale à 4. On peut donc l'écrire sous la forme de:

$$g(r) = -1 + 2 \int_{-\infty}^{r/2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \qquad (5.35)$$

5.3.4. La distorsion:

La distorsion sera donnée par l'expression suivante:

$$D \approx \frac{4\pi}{3} N^{-1} = 4.18\, N^{-1} \qquad (5.36)$$

5. 4. Comparaison du quantificateur AUPQ au quantificateur optimal:

Pour pouvoir comparer les performances du quantificateur AUPQ à celles du quantificateur optimal, il suffit de comparer leurs distorsions respectives. Ainsi, le quantificateur optimal à deux dimensions possédant des performances asymptotiques, a pour distorsion l'expression suivante [9] :

$$D_{opt} \approx \frac{5}{36\sqrt{3}\, N} \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[f(x, y) \right]^{1/2} dx\, dy \right]^2 \qquad (5.37)$$

En coordonnées polaires et aprés avoir intégrer par rapport à la phase, D_{opt} devient:

$$D_{opt} \approx \frac{5\pi}{18\sqrt{3}\, N} \left[\int_0^{+\infty} r^{1/2} \left[f(r) \right]^{1/2} dr \right]^2 \qquad (5.38)$$

Si on compare la distorsion du quantificateur optimal à celle du quantificateur AUPQ, donnée par la formule (5. 26), on aura:

$$\frac{D_{AUPQ}}{D_{opt}} = \frac{3\sqrt{3}}{5} = 1.039 = 0.167\ dB \qquad (5.39)$$

En conclusion, on peut dire (d'aprés la formule (5. 39)) que les performances du quantificateur AUPQ sont inférieures à celles du quantificateur optimal de 0.167 dB,

quelque soit le type de sources symmétriques circulaires appliquées à l'entrée du quantificateur.

5. 5. Description de l'algorithme:

Etape 1: Entrée d'une seule donnée qui est le nombre de quantification N.

Etape 2: Calcul de la valeur du nombre de niveaux de l'amplitude N_r.

Etape 3: Calcul des points de sortie r_i et $\bar{r}_i$ à l'aide des équations (5. 3) et telle que:

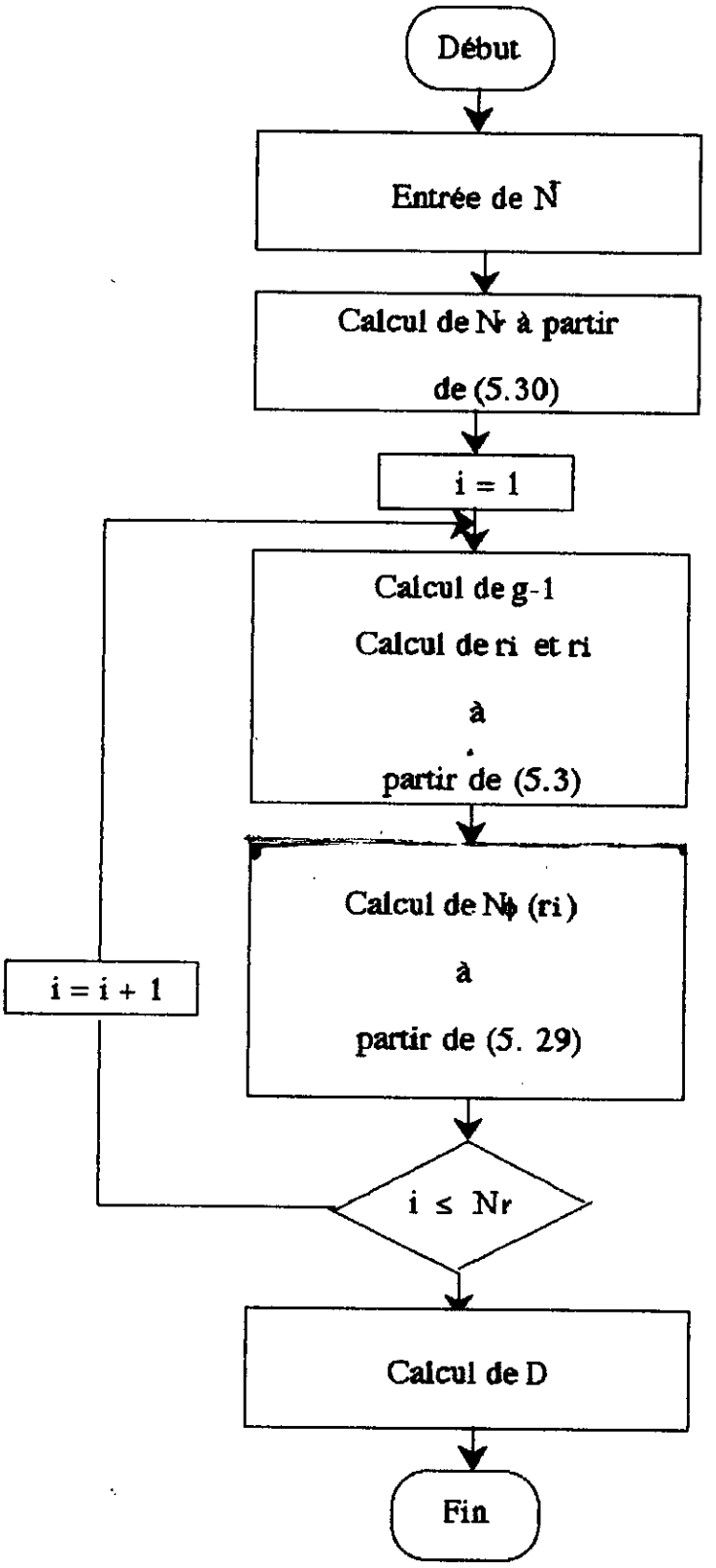
$$g^{-1}(r) = \sqrt{2\pi} \int_0^r e^{s^2/8}\, ds \qquad (5.40)$$

Etape 4: Calcul de la valeur du nombre de niveaux de phase $N_\phi(\bar{r}_i)$ et la distorsion D.

(Cf. organigramme page suivante).

Remarque:

On notera qu'il est possible à la fin du calcul de trouver que la somme des nombres de niveaux $N_\phi(\bar{r}_i)$ soit légèrement différente de la valeur de N. Cela est dû aux approximations introduites dans les différentes étapes de calcul des expressions de N_r et $N_\phi(\bar{r}_i)$. Pour remédier à ce petit problème, on peut ajuster les valeurs de $N_\phi(\bar{r}_i)$ (en ajoutant ou en retranchant autour de ces valeurs) de telle façon que leur somme soit égale à N.


Début
Entrée de N
Calcul de Nr à partir de (5.30)
i = 1
Calcul de g-1
Calcul de ri et ri à partir de (5.3)
Calcul de Nϕ (ri) à partir de (5. 29)
i = i + 1
i ≤ Nr
Calcul de D
Fin

CHAPITRE 6

LES QUANTIFICATEURS OPTIMAUX CIRCULAIRES ET SYMETRIQUES

INTRODUCTION :

La quantification polaire d'une source bidimensionnelle, Gaussienne ou un autre type de sources symétriques circulaires a été déjà vue dans les chapitres précédents.

Dans le présent chapitre, on verra d'autres types nouveaux de quantificateurs polaires appelés quantificateurs optimaux circulaires symmétriques et nommés aussi quantificateurs polaires de Dirichlet [34]..

6.1. RAPPEL :

Il est bien connu comme nous l'avons vu au chapitre 2 section 2.2.2, que deux conditions sont nécéssaires pour un minimum local de l'erreur quadratique moyenne (MSE) et qui sont : les calculs des centroïdes et les contours des partitions de Dirichlet. Les quantificateurs polaires ne satisfont pas ces deux conditions, les seules méthodes conformes à ces deux dernières sont les quantificateurs polaires de Dirichlet [34].
Soit un quantificateur Q_N à N niveaux défini sur un plan $\{ S_i , x_i : i = 1,2,...,N \}$ où S_i représentent les régions disjointes telles que leur union forme le plan c'est à dire :

$$S_i \bigcap_{i \neq j} S_j = \varnothing$$

et (6.1)

$$\bigcup_{i=1}^{N} S_i = \Re$$

Pour un vecteur d'entrée x, on a :

$$Q(x) = x_i \; ; \; x \in S_i \tag{6.2}$$

Pour une densité de probabilité p(x), où x est un vecteur d'entrée, l'erreur quadratique moyenne est telle que :

$$D = \int\int |x - Q_N(x)|^2 \; p(x) \, dx \tag{6.3}$$

En minimisant D par rapport à S_i et x_i, on trouve respectivement les conditions suivantes [34] :

$$x_i = \frac{\int\int_{S_i} x \, p(x) \; dx}{\int\int_{S_i} p(x) \; dx} \tag{6.4}$$

$$S_i = \bigcap_{j=1, j\neq i}^{N} \left\{ x : |x - x_i| \leq |x - x_j| \right\} \tag{6.5}$$

L'équation (6.4) exprime que $\hat{x}_i$ est le centroïde de la région S_i avec une densité p(x). La formule (6.5) exprime que S_i est formée par l'intersection des partitions de Dirichlet de $\hat{x}_i$ et des autres points de sortie. On rappelera que la partition de Dirichlet est formée par l'intersection des bissectrices perpendiculaires aux segments formée par une paire de points de sortie. A partir de (6.5), on montre que les S_i sont toutes convexes et sont connectées les unes aux autres[9].
L'erreur quadratique moyenne résultante pour ce quantificateur optimal est :

$$D = \sigma_X^2 - \sum_{i=1}^{N} |x_i|^2 \int\int_{S_i} p(x) \, dx$$

où σ_X^2 est la puissance du signal.

On remarquera que pour d'autres densités, les conditions (6.4) et (6.5) peuvent être utilisées iterativement pour converger vers un minimum local de la MSE.

Notons que si les régions S_i sont fixées, (6.4) est nécéssaire et suffisante pour minimiser D. Quand les points de sortie sont fixes, la condition (6.5) est nécéssaire et suffisante pour minimiser D.

Comme on l'a vu précédemment, l'algorithme de calcul de (6.4) et (6.5) est de selectionner un ensemble de points de sortie $\{x_i\}$ et d'employer (6.5) pour le calcul de S_i optimale. Ce calcul aboutit à une distorsion D_1. La relation (6.4) n'est pas toujours satisfaite car les $\{x_i\}$ peuvent ne pas être optimaux pour ces régions, ce qui fait que le calcul des $\{x_i\}$ par (6.4) va faire décroitre la distorsion D vers une valeur plus petite que D_1 nommée D_2. De la même manière , (6.5) n'est pas toujours satisfaite, aussi le calcul des régions S_i a pour conséquence une décroissance accrue de la distorsion D. Ce schéma iteratif va aboutir à un minimum local.

6.2. LES QUANTIFICATEURS POLAIRES DE DIRICHLET

6.2.1. INTRODUCTION

Comme nous l'avons vu dans les trois chapitres précédents, les quantificateurs polaires sont classés en deux types : le quantificateur SPQ et UPQ.
Toutes les régions de quantification polaire sont sous forme d'anneaux délimités par des rayons d'angles constant et par des arcs de rayon constant comme le montre la figure (6.1).

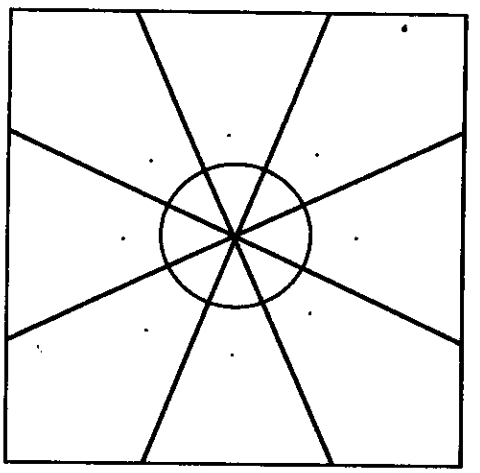


Figure 6.1: Structure d'un quantificateur SPQ.

Malheureusement, ces quantificateurs ne satisfont pas les conditions d'optimalité données par les expressions (6.4) et (6.5). En particulier, les contours de l'amplitude ne forment pas des partitions de Dirichlet. A partir de (6.5), on montre [34] que chaque S_i est un polygone convexe; ce qui n'est pas le cas des régions polaires.

La méthode iterative comme expliquée au début de ce chapitre où on utilise les conditions (6.4) et (6.5) peut être employée pour le quantificateur SPQ afin de réduire la distorsion, ainsi, cette dernière converge vers un minimum local. Aprés avoir selectionner une factorisation de N soit : $N = N_r . N_\phi$; en appliquant la condition (6.5), la constellation aura la forme donnée par la figure (6.2).

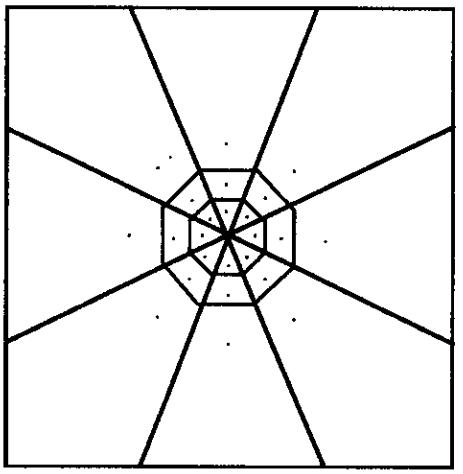


Figure 6.2: Structure du quantificateur DPQ.

La symétrie inhérente de cette constellation permet d'étudier une de ces parties de largeur $2\pi/N\phi$ (par rapport à la phase).

L'utilisation de la méthode iterative ne change pas les limites (ou les contours) de la phase; mais modifie les contours de l'amplitude qui se déplacent alors. De même, les points de la sortie vont varier le long de la bissectrice de la phase. Ainsi, on obtient un nouveau type de quantificateur polaire qui est le quantificateur DPQ (Dirichlet polar Quantizer) à partir du quantificateur SPQ.

6.2.2 FORMULATION MATHEMATIQUE :

A partir de la figure (6.2), on montre que le quantificateur DPQ peut être implémenté de la façon suivante :

La deuxième coordonnée à quantifier est la distance s le long du rayon de la phase quantifiée. L'équation de la coordonnée s aura donc la forme :

$$s = r\cos(\phi - \hat{\phi}) \tag{6.6}$$

où ϕ est la phase quantifiée de ϕ

Sachant que la source est distribuée suivant une loi de Gauss. et que r et ϕ sont des variables aléatoires indépendantes alors r et $\cos(\phi - \hat{\phi})$ sont aussi indépendants. on peut donc trouver la distribution de s qui est la suivante :

$$f(s) = \frac{2N_\phi}{\sqrt{2\pi}} e^{-x^2/2} \left[\Psi\left(s.\tan\left(\frac{\pi}{N_\phi}\right)\right) - \frac{1}{2}\right] \tag{6.7}$$

où :

$$\Psi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \tag{6.8}$$

Quand $N_\phi \rightarrow +\infty$, f(s) approche la densité de Rayleigh [34].

On peut toujours appliquer les expressions données par Max [1], pour définir l'erreur quadratique moyenne minimale où on aura :

$$s_i = \frac{\hat{s}_{i-1} + \hat{s}_i}{2} \tag{6.9}$$

Avec

$$\hat{s}_i = \frac{\int_{s_{i-1}}^{s_i} t.f(t)\, dt}{\int_{s_{i-1}}^{s_i} f(t)\, dt} \tag{6.10}$$

Où $\hat{s}_i$ sont les points de sortie et les s_i sont les limites des régions de quantification. Le quantificateur trouvé est unique (ou encore les valeurs du quantificateur sont uniques) si la condition de Fleischer[4] est vérifiée.
Cette condition est la suivante :

Cette condition est la suivante :

$$\frac{\partial^2}{\partial s^2}\left(\log(f(s))\right) < 0 \qquad (6.11)$$

6.2.3 REMARQUE

On peut toujours utiliser la méthode iterative dans le cas du quantificateur UPQ pour N = 1,2,....32; mais on remarque que pour N = 1, 2, 3 et 4, le quantificateur est déjà optimal. Pour les cas de N = 5, 6, 7, et 8, il est façile de les tendre à ces valeurs.

Malheureusement pour N > 8, il est difficile d'utiliser cette méthode. C'est pourquoi nous n'avons tenu compte que des faibles valeurs de N (N inférieur ou égal à 8), car c'est pour ces seules valeurs que le quantificateur rectangulaire dépasse en performances celles du quantificateur polaire.

6.3 QUANTIFICATEUR POLAIRE ROTATIF DE DIRICHLET (DRPQ)

On a vu dans la section précédente que les conditions d'optimalité (6.4) et (6.5) peuvent être appliquées d'un point de vue iteratif jusqu'à l'obtention d'un minimum local de la distorsion. Le quantificateur résultant dépend du point de sortie initial.

Une grille rectangulaire initiale produit un quantificateur rectangulaire à partir d'une paire de quantificateurs de Max [1]. Les partitions se déplacent toujours perpendiculairement. Une partition polaire initiale produit un quantificateur polaire de Dirichlet que l'on a déjà introduit.

Considérons le quantificateur polaire (SPQ) où l'amplitude et la phase sont quantifiées indépendemment l'une de l'autre. La rotation de chaque anneau de l'amplitude, comme on peut le voir dans la figure (6.3) ne change pas la valeur de la distorsion [34].

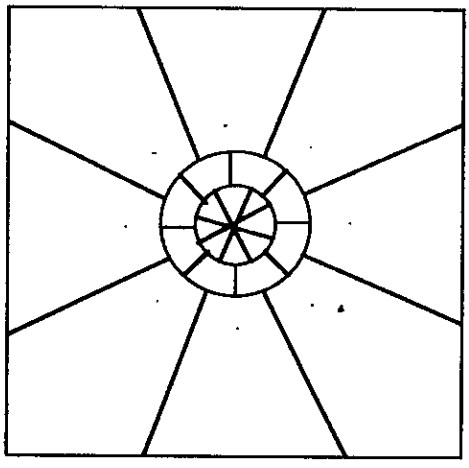


Figure 6.3: Structure du Quantificateur SPQ
, aprés la rotation.

Quand on applique ce nouveau modèle comme valeur initiale pour la méthode iterative (expressions (6.4) et (6.5)), donnera un nouveau motif presenté par la figure (6.4), qui est assez différent du quantificateur polaire de Dirichlet. Ce nouveau quantificateur est nommé : Quantificateur polaire rotatif de Dirichlet ou DRPQ (Dirichlet rotated polar quantizer).

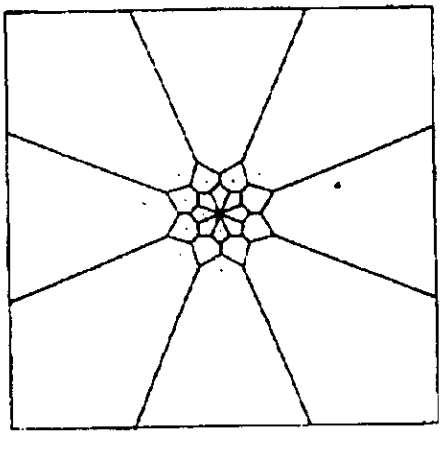


Figure 6.4: Structure du quantificateur DRPQ.

6.3.1 IMPLEMENTATION DU QUANTIFICATEUR DRPQ

Le quantificateur DRPQ est plus difficile à implémenter que le quantificateur DPQ. Nous donnons cependant dans cette partie l'algorithme de calcul de ce quantificateur. On rappellera que la factorisation est la même que celle du quantificateur SPQ; c'est à dire : $N = N_r . N\phi$.

Première étape :

Convertir l'entrée x en coordonnées polaires, r et ϕ.

Deuxième étape :

Calcul du quantificateur uniforme de la phase ϕ avec $2N\phi$ niveaux sur $[0,2\pi[$. La sortie $\hat{\phi}_i$ est de la forme :

$$\hat{\phi}_j = \frac{\pi(2j - 1)}{2N_\phi} \qquad j \in [1,2,...,2N_\phi]$$

Troixième étape :

Calcul du quantificateur d'amplitude à N_r niveaux.

Quatrième étape :

Pour l'amplitude de niveau i et la phase de niveau j, la sortie est l'une des deux expressions:

$$\hat{r}_i = \exp(\hat{\phi}_{i,j})$$

$$\hat{r}_{i+1} = \exp(\hat{\phi}_{i,j-1})$$

Où :

$$\hat{\phi}_{i,j} = \frac{\pi}{N_\phi}\frac{2j - 1}{2} \pm \frac{\pi}{2N_\phi}\Bigg\} \qquad + \text{ si } |i-j| \text{ est paire, } - \text{ si } |i-j| \text{ est impaire}$$

Caculer les distances à partir de ces points et prendre les plus proches de $r.e^{j\phi}$, comme points de sortie.

On utilise ces points de sortie comme séquence d'apprentissage pour l'algorithme LBG. Cet algorithme permet le calcul iteratif des conditions (6.4) et (6.5). Nous rappelerons son organigramme à la fin de ce chapitre.

6.4. REPARTITION DES DEBITS POUR LES DIFFERENTS QUANTIFICATEURS

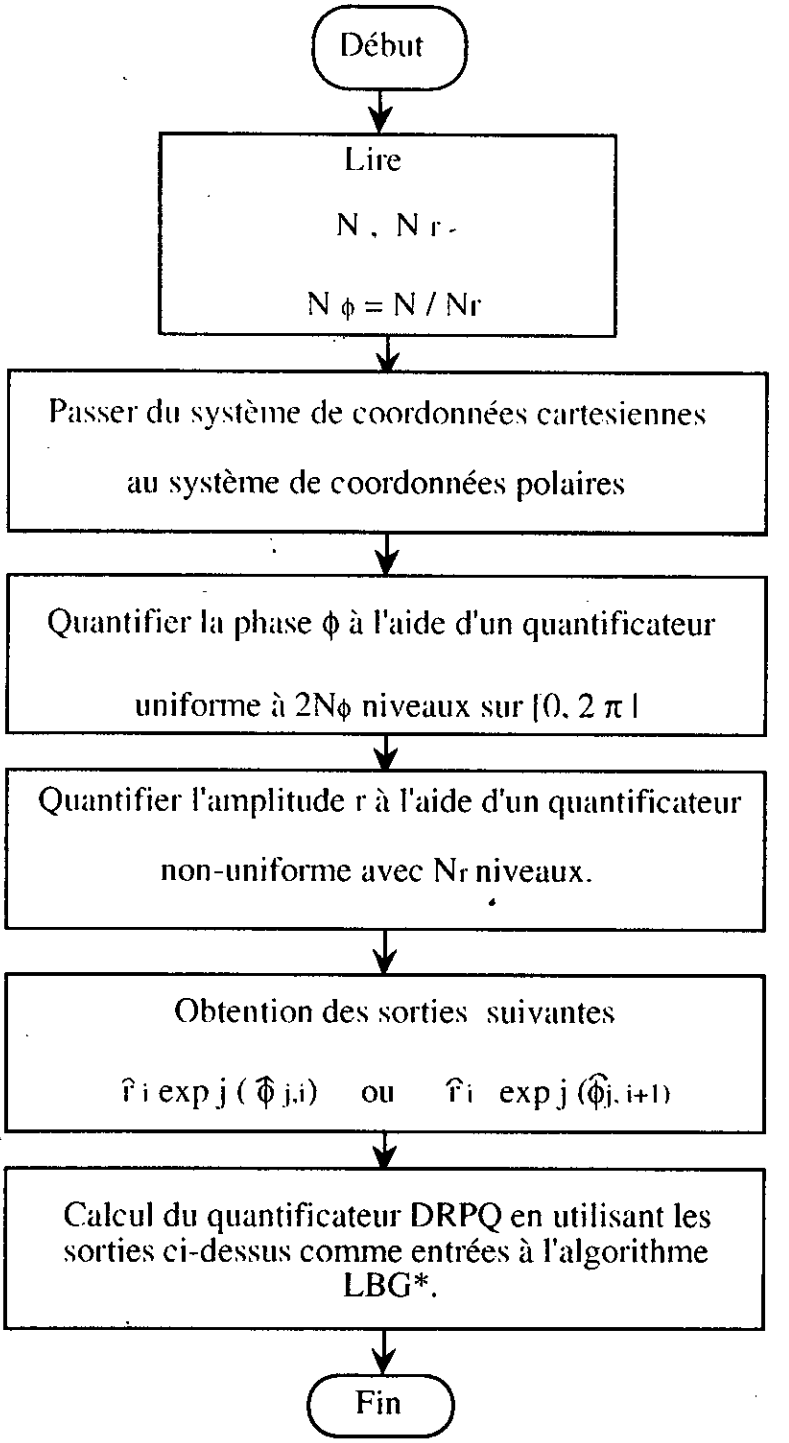
Cet exemple donne la factorisation des différents quantificateurs étudiés jusqu'alors dans le tableau ci-dessous [38] :

Quantificateurs	Factorisation
Rectangulaire	$N_X = N_Y$ [37]
SPQ - UPQ - AUPQ	$N_\phi \approx 2.6\, N_r$ [28]
DPQ	$N_\phi \approx 2.6\, N_s$ [38]
DRPQ	$N_\phi \approx N_r$ [38]

CONCLUSION:

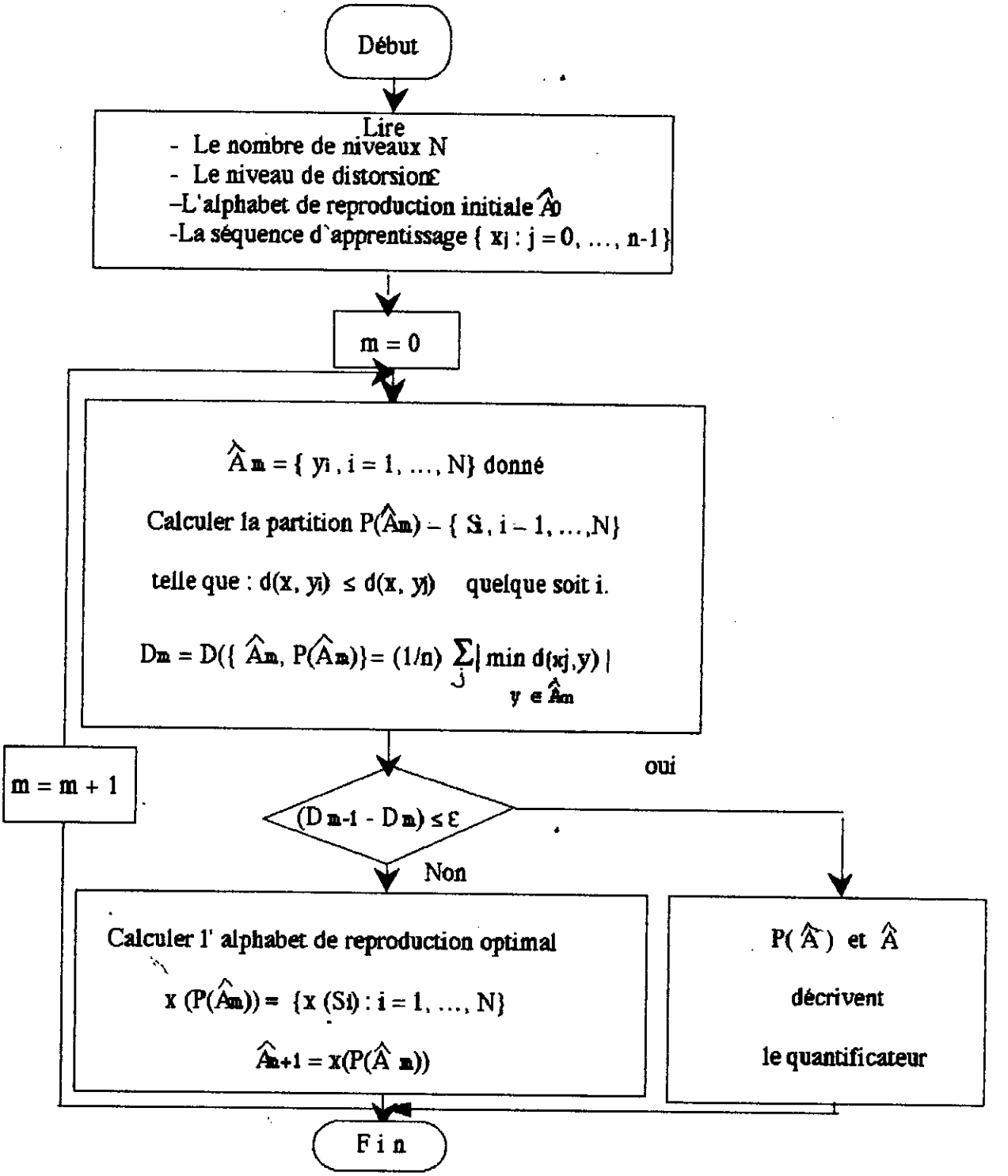
Comme nous pouvons le constater, les constellations des quantificateurs DPQ et DRPQ tendent respectivement au fur et à mesure vers l'hexagonalité des régions de quantification; ce qui prouve que ces quantificateurs tendent vers l'optimalité plus que tout autre quantificateur polaire, car comme nous avons vu au chapitre 2 , la structure ou la constellation du quantificateur optimal est formée d'hexagones réguliers.

Organigramme de calcul du quantificateur DRPQ



(*) voir page suivante

AlgorithmeLBG:



CHAPITRE 7

RESULTATS ET INTERPRETATIONS

INTRODUCTION

Dans cette partie, nous allons présenter les résultats obtenus à partir des algorithmes des différents quantificateurs, soit sous formes de tableaux de graphes, ou (et) de schemas. nous comparerons ces résultats les uns par rapport aux autres, et en dernier lieu à la borne de shannon.

Il serait intéressant de faire deux remarques avant la présentation des résultats: Premièrement, les résultats obtenus sont relatifs à une source qui suit une distribution de Gauss; en second lieu, les algorithmes ont été traités sur un micro-ordinateur Macintosh II dont le système est de type Unix, version 6.1 et d'horloge égale à 25 MHz.

7.1.Comparaison des résultats du quantificateur SPQ à ceux du quantificateur rectangulaire:

Nous souhaitons maintenant comparer les résultats du quantificateur polaire SPQ à ceux du quantificateur rectangulaire (ou cartésien). Les résultats du quantificateur rectangulaire bidimensionnel ont été pris à partir de la thèse [30] où on a utilisé la méthode de Max [1] . Chaque coordonnée cartésienne suit une loi de Gauss où elle est quantifiée comme on l'a déjà dit à l'aide du quantificateur de Max [1].
Pour le quantificateur SPQ', nous avons calculé les distorsions minimales pour chaque factorisation de N variant de 1 à 1024. Ainsi, on a cherché la distorsion minimale ainsi que la factorisation qui lui est associée pour chaque valeur de N.

Avant de comparer le quantificateur polaire (SPQ) aux autres types de quantificateurs, il serait intéressant de comparer les résultats du quantificateur SPQ non-uniforme à ceux du quantificateur SPQ uniforme. D'aprés la figure 7.1, on remarque qu'au fur et à mesure que le nombre de niveaux N augmente, la distorsion introduite décroit. On constate aussi d'aprés la même figure 7.1 , que les performances du

quantificateur SPQ non-uniforme dépasse celles du quantificateur SPQ uniforme lorsque le débit dépasse 2.5 bits/v.c. Pour un débit inférieur à 2.5 bits/v.c, leurs performances respectives sont trés proches l'une de l'autre. Cela s'explique par ce qu'on a vu au chapitre 2, où quelque soit la valeur du débit, la distorsion du quantificateur SPQ uniforme est dans le meilleur des cas égale à celle du quantificateur SPQ non-uniforme. De plus, le quantificateur uniforme est optimal lorsque la source suit une loi uniforme, où sa structure est une partition de Dirichlet (formée par des hexagones réguliers). Mais lorsque la source suit une loi quelconque (autre que la loi uniforme), comme dans notre cas, ce quantificateur perd de ses performances; c'est la raison pour laquelle on quantifie la source par un quantificateur non-uniforme qui divise les régions de quantification de façon non-uniforme.

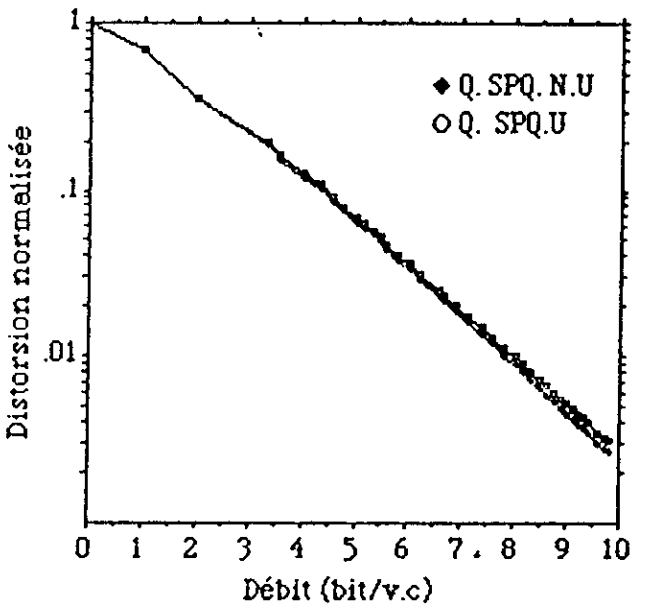


Figure 7.1: représentation des performances du quantificateur SPQ uniforme (designé par la lettre U) et non-uniforme (designé par la lettre NU).

Pour comparer les performances de la quantification polaire et rectangulaire, nous avons représenté dans le graphe de la figure 7.2, le $\log_2$ de la distorsion minimale en fonction du débit ($Log_2N = R$) des quantificateurs polaire et rectangulaire pour les cas uniforme et non-uniforme. La remarque à faire à partir de ce graphe est que le quantificateur polaire (aussi bien uniforme que non-uniforme) dépasse en performances le quantificateur rectangulaire à haut débit et est comparable ou inférieur (d'une maniére insignifiante) à bas débit. Ce dernier se situe approximativement entre 0 et 6 bits/v.c , alors que le haut débit va de 6 bits/v.c à au delà. Cela voudrait dire que pour

un N supérieur à 64, la distorsion du quantificateur rectangulaire dépasse celle du quantificateur polaire SPQ.

Quand le débit augmente, le surpassement du quantificateur SPQ sur le quantificateur rectangulaire pour le cas uniforme devient plus significatif que pour le cas non-uniforme.

Par exemple, à 9.8 bit/v.e (N=900), le taux de différence de distorsion (entre le quantificateur SPQ et le quantificateur rectangulaire) est pour le cas uniforme de 1.27 alors que pour le non-uniforme, il est de 1.07 seulement.

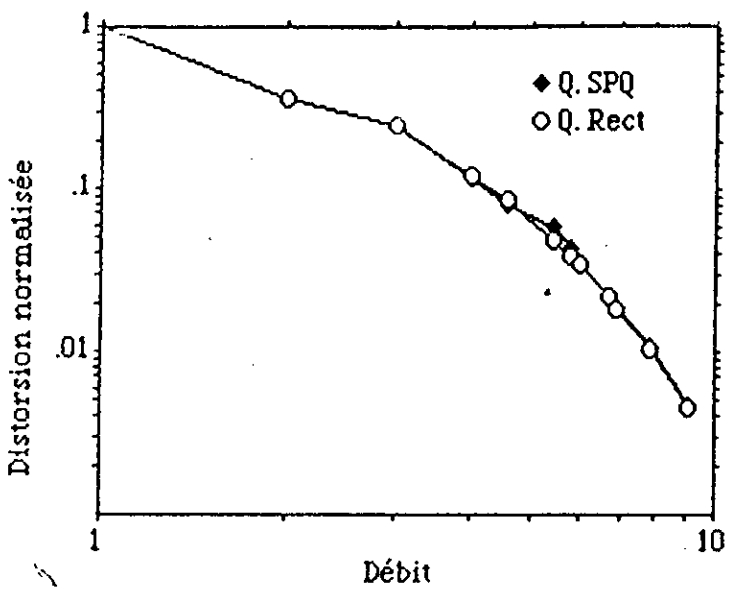


Figure 7.2: représentation des performances des quantificateurs SPQ et rectangulaire non-uniformes.

Lorsque le débit croit, les courbes obtenues à la figure 7.2 approchent des droites; c'est ce qui a poussé Pearlman [28] à trouver les équations régissant ces droites; pour cela, il a utilisé la méthode des moindres carrés.

Leurs équations sont données par le tableau suivant:

Tableau 7.1: Résumé des différentes équations de droites obtenues par Pearlman

Quantificateur et borne	Equations $D_z/2\sigma^2$	Intervalle de definition
Polaire non-uniforme	$(2.056)\,2^{-.977R}$	$7.69 \leq R \leq 10$
Rectangulaire non- uniforme	$(1.887)\,2^{-.955R}$	$6.64 \leq R \leq 10$
Polaire uniforme	$(1.849)\,2^{-.939R}$	$7.26 \leq R \leq 10$
Rectangulaire uniforme	$(1.358)\,2^{-.861R}$	$6.64 \leq R \leq 10$
Borne polaire Pearlman - Gray	$1 - \exp\{-(1.781)\,2^{-R}\}$	$R > 1.376$
Borne de Shannon	2^{-R}	$R > 0$

7. 2. Comparaison entre la répartition des débits de la phase et du module:

Nous desirons maintenant faire quelques remarques en ce qui concerne la répartition des débits entre la phase et l'amplitude dans la pratique pour la quantification polaire.
A partir des résultats que nous avons obtenu et qui donnent les N_ϕ (nombre de niveaux de phase) et N_r (nombre de niveaux d'amplitude) optimaux; on a facilement calculé et représenté (figure 7.3) les débits correspondants et la différence des débits.

D'aprés l'expression (3. 41) du chapitre 3, nous avons trouvé théoriquement la différence optimale des débits entre la phase et l'amplitude et qui est:

$$R_\phi - R_r = 1.376 \text{ bits/s.}$$

Comme le montre la figure (7.3), les valeurs pratiques trouvées sont voisines de la valeur optimale qui est 1.376 bits/v.c

La moyenne de la différence des débits peut-être un bon indicateur pour la quantification optimale étant donné que cette moyenne lisse les effets de la contrainte. Les moyennes de toutes les différences des débits pour tous les débits supérieurs à 1.376 bits, sont respectivement 11.52 et 11.47 bits (soit N_ϕ/N_r égal respectivement à 2.89 et 2.77) pour les quantificateurs non-uniforme et uniforme. Considérons que l'usage de 1.376 bits/v.c appliqué au codage optimal donne une limite inférieure qui est une approximation précise de la dernière performance pour des débits superieurs à 4.12 bits/v.c.
Cette limite est un bon indicateur pour la manière de diviser le débit entre la phase et l'amplitude dans le cas de la quantification polaire.

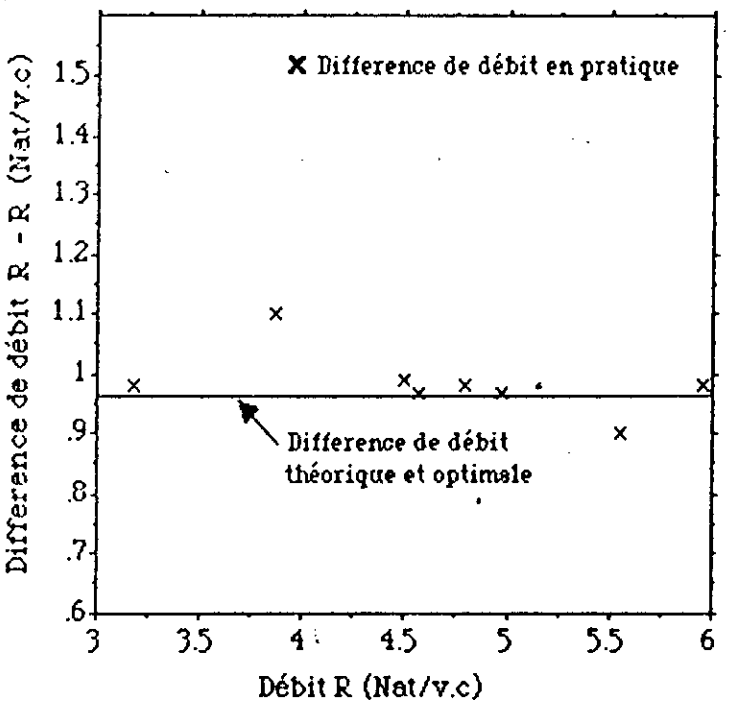


Figure 7.3: Répartition des débits entre la phase et le module.

7.3, Comparaison des quantificateurs SPQ et rectangulaire au quantificateur UPQ:

L´optimisation du quantificateur UPQ a été faîte pour N = 1, 2, . . . , 16, 25, 32 et 36. car, comme on a vu précédemment, ce quantificateur s´intéresse aux faibles valeurs de N pour deux raisons:

- Le quantificateur SPQ est moins performant que le quantificateur rectangulaire à bas débit, c'est pourquoi il est interressant de chercher un quantificateur polaire qui dépasse le quantificateur rectangulaire en performances.

- Le temps de calcul du quantificateur UPQ devient trés long à haut débit.

Enfin, une trosième remarque s'impose dans cette introduction: On ne comparera le quantificateur UPQ au quantificateur rectangulaire que pour des valeurs de N (nombres de niveaux) ayant une propriété de carrés parfaits.

D'aprés la figure (7.4) où on a représenté la distorsion minimale en fonction du débit pour les trois types de quantificateurs, plusieurs conclusions peuvent être tirées:

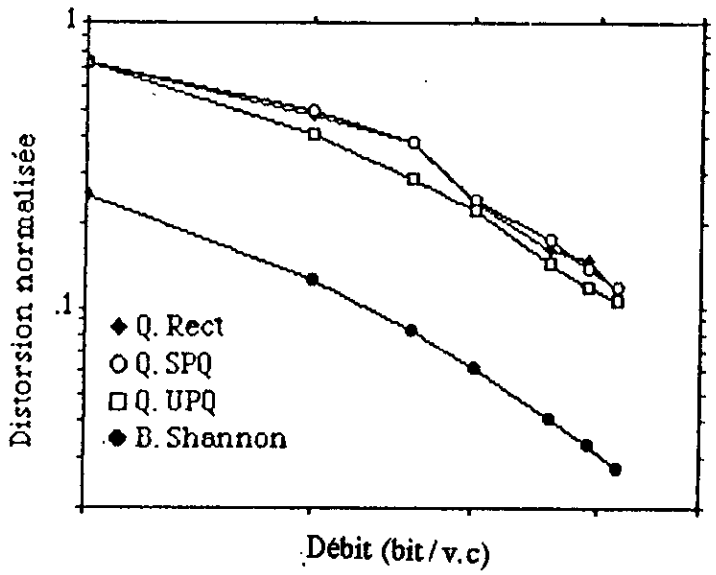


Figure 7.4: Comparaison des performances des quantificateurs rectangulaire, SPQ, et UPQ avec la borne de Shannon à bas débit (N < 64).

- En comparant le quantificateur UPQ au quantificateur rectangulaire à deux dimensions, on constate que le quantificateur UPQ est aussi bon ou meilleur pour tous les cas où N est un carré parfait; quand N n'est pas un carré parfait, moins de symmétrie existe, mais les gains ne sont pas dramatiques: 0.06 dB pour N=3, 0.3 dB pour N= 16, 0.34 dB pour N= 25 et 0.4 dB pour N=36. Donc, les quantificateurs UPQ améliorent les performances des quantificateurs SPQ lesquels sont meilleurs que les quantificateurs rectangulaires pour N > 64.

- La deuxième remarque que l'on peut faire est que les résultats du quantificateur UPQ sont meilleurs ou similaires à ceux du quantificateur SPQ (quand N<4) ; cela n'est pas un cas surprenant puisque le type des quantificateurs SPQ est une sous-classe des quantificateurs UPQ .

Cependant, l'amélioration introduite pour un quantificateur est parfaite quand la meilleure factorisation cherchée de $N = N_\phi \cdot N_r$ demeure assez modeste. Ce qui est toujours le cas quand N est un nombre premier, mais parfois pour N quelconque [13] . Par exemple, pour N=8 (c'est à dire attribuant 3 bits pour deux échantillons), le quantificateur SPQ optimal (8, 2, 4, 4) obtient une distorsion D=0.125, alors que le quantificateur UPQ optimal (8, 2, 1, 7) réalise une distorsion D= 0.102. On peut donc conclure que la contrainte de la somme est plus flexible et donne de meilleurs résultats que celle de la contrainte de factorisation.

Bien que plusieurs améliorations ont été obtenues pour la quantification à deux dimensions: la limite de Shannon reste relativement loin. Il peut y exister des algorithmes pour deux dimensions qui soient légèrement meilleurs probablement avec moins de simplicité.

Cependant, il est évident que l'obtention de performances prés de R(D) (limite de Shannon) demande des blocs de mots plus longs pour les quantificateurs multidimensionnels analysés par Gersho [9] ou encore un autre type de codage tel que le codage par treillis.

7.4. Comparaison des quantificateurs SPQ, UPQ au quantificateur AUPQ:

Comme on a vu (chapitre 5), le quantificateur AUPQ possède les mêmes propriétés que le quantificateur UPQ, mais il est appliqué aux hauts débits vu les inconvénients que possède le quantificateur UPQ. Ainsi, le calcul du quantificateur polaire à haut débit ne se fait qu'à partir de l'algorithme de calcul du quantificateur SPQ. Les performances de ce dernier ne sont pas trés satisfaisantes, c'est pourquoi nous avons opté pour le quantificateur AUPQ.

Nous remarquons que ce quantificateur dépasse en performances les quantificateurs SPQ et rectangulaire à haut débit. On peut donc envisager à première vue qu'au lieu d'utiliser oomme quantificateur polaire le quantificateur SPQ, il serait plus intéressant d'utiliser le quantificateur UPQ à bas débit et le quantificateur AUPQ à haut débit.

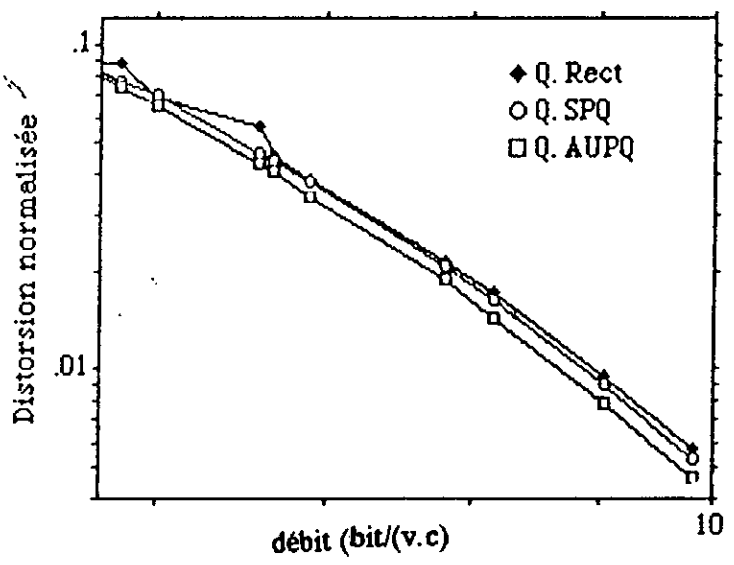


Figure 7.5: Comparaison des performances des quantificateurs rectangulaire, SPQ, et AUPQ à haut débit (N > 64).

7.5. Comparaison des quantificateurs rectangulaire, polaires (SPQ, UPQ) aux quantificateurs de Dirichlet (DPQ et DRPQ) et conclusion:

D'aprés le tableau (donné ci-dessous), où on donne dans la première colonne les nombres de niveaux N qui sont des carrés parfaits car à ces valeurs de N seulement que les quantificateurs rectangulaire et DRPQ obtiennent les plus faibles distorsions , alors que les autres colonnes sont constituéesrespectivement des distorsions des quantificateurs rectangulaire, SPQ, UPQ, AUPQ, DPQ, et DRPQ.

Les valeurs de N mises entre parenthèses sont représentées ainsi, dans le cas où le nombre N (qui donne la distorsion correspondante) soit différent de celui donné par la première colonne.

La figure (7.5) donne alors les courbes obtenues à partir du tableau.

Tableau 7.2: Résumé des résultats finaux pour tous les types de quantificateurs

N	Q. Rect	Q. SPQ	Q. UPQ	Q. AUPQ	Q. DPQ	Q. DRPQ
16	0.2350	0.2396	0.2180		0.2391	0.2224
25	0.1599	0.1709 (24)	0.1438		0.1702 (24)	0.1462
36	0.1159	0.1176	0.1056		0.1174	0.1052
49	0.0880	0.08889 (48)	0.8064		0.08882 (48)	0.07899
64	0.06908	0.06973		0.06520	0.06967	0.60134
100	0.04586	0.04392 (102)		0.4120	0.04387 (102)	0.04003
144	0.03268	0.03244 (140)		0.02902	0.03241 (140)	0.02816
225	0.02146	0.02056		0.01857	0.02055	0.01822
324	0.01519	0.01468 (320)		0.01290	0.01467 (320)	0.01280
529	0.009482	0.008904 (352)		0.008788	0.008899 (532)	0.0080046
900	0.005668	0.005310		0.004740	0.005308	0.004684

A partir de ces graphes, on constate en premier lieu que les performances du quantificateur DRPQ sont meilleures que celles du quantificateur DPQ. Ce dernier dépasse (en performances toujours) faiblement le quantificateur SPQ à bas débit. Cependant, on remarque que pour un certain débit (N =16), le quantificateur UPQ est meilleur que tout autre quantificateur: cela est dû vraisemblablement au fait que la structure (16, 3, 1, 6, 9) est une structure optimale [34]

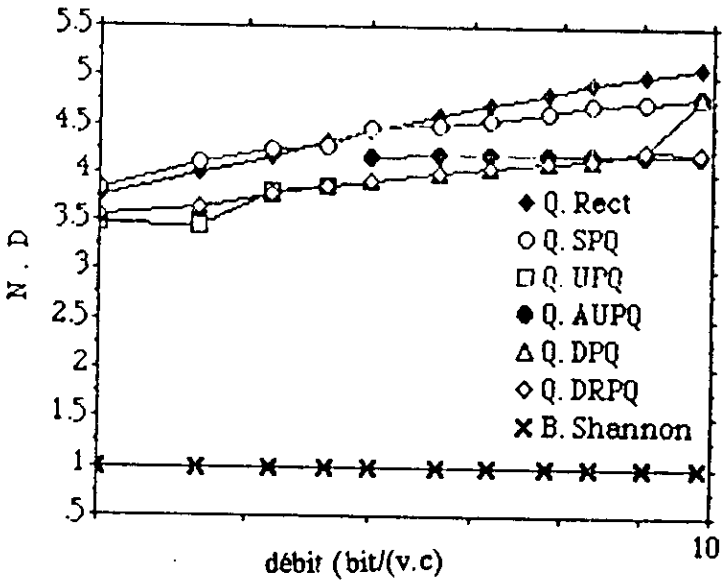


Figure 7.6: Performances des quantificateurs (étudiés) comparées à la borne de Shannon.

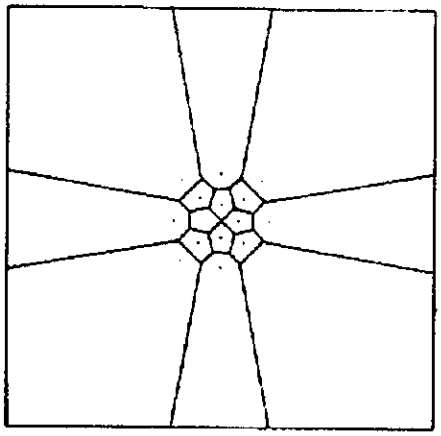
Quand N devient trés grand, les performances du quantificateur SPQ s'alignent sur celles du quantificateur DPQ. En ce qui concerne le quantificateur DRPQ, celui-ci dépasse d'une manière considérable tous les autres quantificateurs et s'approche plus de la limite de Shannon: il se trouve entre le quantificateur polaire et le quantificateur optimal. Par exemple pour N=100, le rapport SNR (signal/ bruit) dépasse de 0.6 dB le quantificateur rectangulaire et de 0.4 dB le quantificateur polaire.

Enfin, comme on peut le voir sur la figure (7.6) la structure des régions de quantification du quantificateur DRPQ tend vers une certaine hexagonalité. Cette dernière s'accroit d'une manière importante lorsque le nombre de niveaux augmente.

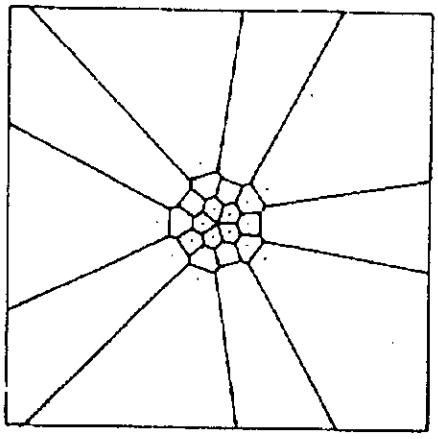
En ce qui concerne la répartition des débits pour les quantificateurs de Dirichlet, on a vu au chapitre 6 que l'erreur quadratique moyenne est mieux minimisée si le nombre de niveaux de quantification est un carré parfait. On a aussi montré que la factorisation polaire optimale des nombres de niveaux et de phase est telle que $N_\phi = 2.6\ N_r$.

Quand le nombre de niveaux devient élevé, les quantificateurs DPQ et SPQ deviennent équivalents, donc les factorisations asymptotiques sont les mêmes. Pour le quantificateur DRPQ, toutes les combinaisons ont été essayées pour N< 144.

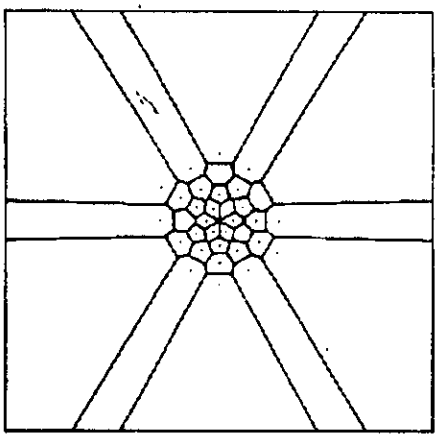
On a trouvé les meilleurs résultats lorsque le nombre de niveaux (pour chaque dimension) est identique. Pour N > 144, les valeurs de N prises sont celles où N est un carré parfait.



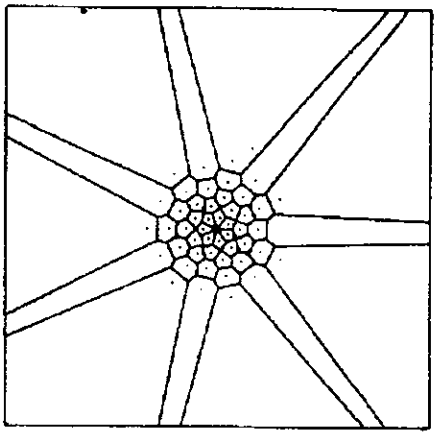
N = 16



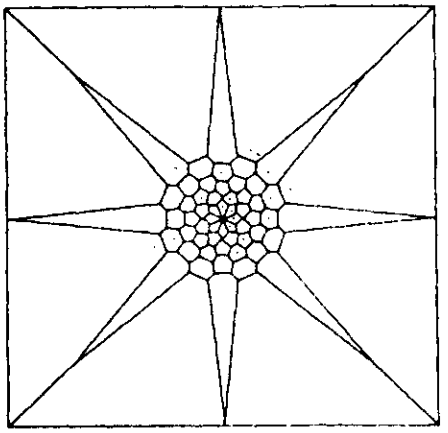
N = 25



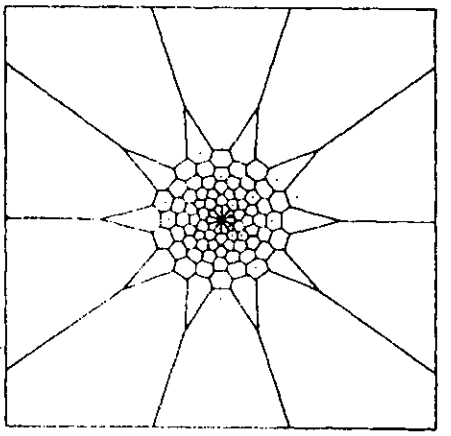
N = 36



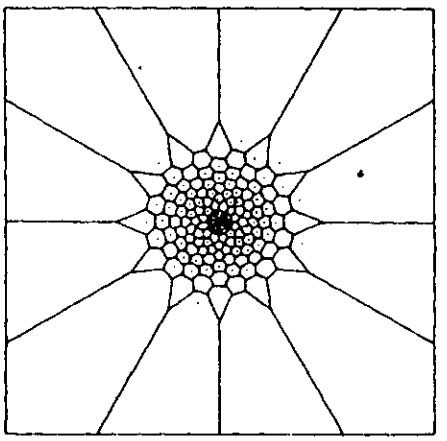
N = 49



N = 64



N = 100



N = 144

Figure 7.6: Représentation des constellations du quantificateur DRPQ pour N = 16, 25, 36, 49, 64, 100 et 144.

Conclusion:

On peut donc conclure que plus la distorsion (dont on a choisi l'erreur quadratique moyenne comme critère) diminue, plus les limites des régions de quantification tendent à être hexagonales.

Dans ce cas, nous nous sommes intéressés à une source Gaussienne. Les trapézoides du quantificateur DPQ les polygones du quantificateur DRPQ deviennent des polytopes pour de grandes dimensions [9].

CONCLUSION

Le but de ce travail a été d'étudier la quantification polaire dans tous ses aspects pour des signaux à deux dimensions.

Cette étude nous a permis de cerner le problème de la quantification polaire et cela en introduisant tous les types de quantificateurs polaires, passant de l'un à l'autre, en changeant l'une des coordonnées; où la principale propriété de la quantification polaire est que ses coordonnées respectives sont des variables aléatoires statistiquement indépendantes.

Pour ce qui est des quantificateurs polaires SPQ et UPQ, le passage de l'un à l'autre se fait par le changement de la contrainte; donc de la structure du quantificateur.

En ce qui concerne le quantificateur DPQ, celui-ci a été introduit en passant de la coordonnée du rayon à la ccordonnée de la corde. Son algorithme est simplement l'algorithme de Max[1]. Pour ce qui est du quantificateur DRPQ, ce dernier a été trouvé en utilisant l'algorithme LBG[17] et les résultats du quantificateur SPQ comme séquence d'apprentissage.

Ainsi, la conclusion qui s'impose à partir de ce travail est la suivante: tous les quantificateurs vus jusqu'à présent (rectangle, polaire, DPQ et DRPQ) minimisent l'erreur quadratique moyenne (MSE) selon leur contrainte respective. Le quantificateur rectangulaire satisfait aux conditions nécessaires (centroïde et les partitions de Dirichlet données au chapitre 6) , mais perd la symétrie.

Les quantificateurs polaires (SPQ, UPQ, AUPQ) préservent le problème de la symétrie mais ne vérifient plus les conditions nécessaires.

Quant aux quantificateurs DPQ et DRPQ, ces derniers possèdent les deux propriétés. En d'autres termes, ils vérifient les deux conditions nécessaires et possèdent la symétrie dans leurs structures.

Toutefois, le quantificateur DRPQ tout en réduisant plus l'erreur quadratique moyenne que tout autre quantificateur précédemment vu, augmente en complexité lors

de son implémentation. Ainsi, il existe un compromis à faire entre l'optimalité du quantificateur et la complexité à le concevoir.

Enfin, on a pu optimiser la répartition des débits entre les différentes variables à quantifier, pour chaque quantificateur, ce qui a permis la réduction du temps d'exécution d'une manière considérable. On tient à souligner aussi que les méthodes vues dans ce travail ont été réalisées pour une source gaussienne, et on peut toujours les étendre à d'autres types de sources.

Bien que les résultats trouvés à partir des méthodes données par la quantification sous-optimale sont assez loin de ceux trouvés par la quantification optimale, le travail réalisé dans cette thèse a pu donner des dictionnaires qui peuvent être utilisés comme des valeurs d'entrée pour d'autres méthodes de calcul afin d'approcher de plus en plus le quantificateur optimal.

De plus, la quantification sous-optimale reste l'objet de plusieurs travaux de recherche (telle que la spirale récemment mise en oeuvre) car les chercheurs ont trouvé des difficultés (telles qu'une très grande capacité mémoire et aussi un temps d'exécution énorme) à utiliser les méthodes optimales malgré leurs bonnes performances. C'est pourquoi, ils se sont attelés à trouver de nouvelles méthodes efficaces dans le domaine de la quantification sous-optimale, c'est-à-dire des méthodes assez bonnes du point de vue performance et qui ne demandent pas une mémoire et un temps d'exécution importants. Ainsi, ce travail reste un premier pas dans la quantification sous-optimale par lequel on peut s'inspirer pour d'autres travaux.

BIBLIOGRAPHIE

[1] J. Max, " Quantizing for minimum distorsion", IRE Tran. Inform. Theory, vol IT-6, pp. 7-12, March 1960.

[2] T. Berger, "Optimum Quantizers and permutation codes, " IEEE Trans. Inform. Theory, vol. IT-18, pp. 759-765, Nov. 1972.

[3] J. A. Bucklew and N. C. Gallagher, "Quantization Schemes for bivariate Gaussian Random Variables," IEEE Trans. Inform. Theory, July 1978.

[4] P.E. Fleischer,"Sufficient Conditions for Achieving Minimum distorsion in a Quantizer". IEEE Int. Conv. Rec., Part I, pp. 104-111, 1964.

[5] A. G. Tescher, "The Role of Phase in Adaptative Image Coding," Technical Repport N°. 510, Image Processing Institute, Electronic Sciences Laboratory, University of Southern California, Los Angeles, CA, Dec. 1973.

[6] N. C. Gallagher, Jr., "Quantizing Schemes for the Discrete Fourier Transform of a Random Time Series," IEEE Trans. Inform. Theory, vol. IT-24, pp. 156-163, March 1978.

[7] J. J. Huang and P. M. Schultheiss, "Block Quantization of correlated Gaussian Random Variables," IEEE Trans. Commun. Syst., vol. CS-11, pp. 289-296, Sept. 1963.

[8] W. A. Pearlman and R. M. Gray, "Source coding of the Discrete Fourier Transform," IEEE. Trans. Inform. Theory, vol. IT-24, pp. 683-692, Nov. 1978.

[9] A. Gersho "Asymptotically Optimal Block Quantization,"IEEE Trans. Inform. Theory, vol. IT-25, july 1979.

[10] Andrew J. Viterbi, Jim K. Omura: "Principles of digital communication and coding". McGraw-Hill Book Company 1979.

[11] N. S. Jayant, P. Noll: " Digital coding of waveforms ", Prentice-Hall, INC, 1984.

[12] P. F. Swaszek and J ohn B. Thomas: " Multidimensional Spherical Coordinates Quantization," IEEE Trans. Inform. Theory, vol. IT-29, pp. 272-278, April 1983.

[13] Wilson: " Block Quantization of Correlated Gaussian Random Variables," IEEE Trans. Commun. Syst. vol. CS-11, pp. 237-243, April 1963.

[14] A. Gersho, "On the structure of vector quantization," IEEE Trans. Inform. Theory, vol. IT-28, pp.157-166, March 1982.

[15] Y. Y. S. Tazaki and R. M. Gray, " Asymptotic performance of a block quantizer with difference distorsion measure," IEEE Trans. Inform. Theory, vol. IT-26, pp. 6-14, Jan. 1980.

[16] J. A. Bucklew and J. L. Wise, "Multidimensional Asymptotic quantization theory with rth power distorsion measures," IEEE Trans. Inform. Theory, vol. IT-29, pp. 239-247, April 1983.

[17] Y. Linde, A. Buzo, and R. M. Gray, " an Algorithm for vector quantizer design," IEEE Trans. Commun. Technol. , vol. COM-28, pp. 84-95, Jan. 1980.

[18] G.H.Hardy, J.E.Littlewood, G.Polya: " Inequalities," Cambridge at the university press, 1967.

[19] J.Bass: "Cours de mathématiques," Tomes I et II, Masson et C^{ie}, 1968.

[20] R.Weinstocks," Calculus of variations with applications to phisics and engineering," McGraw-Hill Book company, 1952.

[21] N.Piskounov: "Calcul différentiel et intégral." tome II, Editions Mir. Moscou 1980.

[22] R. M. Gray: " Source coding theory ," Kluwer, Academic press, 1990.

[23] A. Gersho: " Principles of quantization," IEEE Tran. on curcuits and syst., vol. CAS.25, pp 427-436, july 1978.

[24] J. P. Adoul: " La quantification vectorielle des signaux: Approche algebrique," Annales de Telecommunications, vol. 41,1986.

[25] S. P. Lloyd: " Least square quantization in PCM," IEEE Trans. Inform. Theory, vol. IT-28, pp. 129-137, March 1982.

[26] W. A. Pearlman and G. H. Senge: " Optimal quantization of the Rayleigh probability distribution," IEEE Trans. on Communication, vol. COM-2, pp. 101-112, Jan 1979.

[27] T. Berger: "Optimum quantizer and permutation codes," IEEE Trans. Inform. Theory, vol. IT-28, pp. 149-166, March 1982.

[28] [8] W. A. Pearlman:"Polar quantization of a complex random variabl," IEEE. Trans. Communications, vol. COM-27, pp. 892-897, Jun. 1979.

[29] R. M. Gray: "Vector quantization, " IEEE ASSP magazine, vol 1, pp 4-29, April 1984.

[30] S. Ghernaouti, R. Gueraichi: "Quantificateurs optimaux des signaux bidimensionnels, " thèse de PFE, Electronique ENP, Juin 1989.

[31] D. Berkani: "Design du quantificateur en spirale," thèse de Doctorat d'Etat, Electronique. ENP, Sept 1991

[32] W. R. Bennett: "Spectra of quantized signals," Bell System Technical Journal, pp. 446-471 , 1948.

[33] P. F. Swaszek, J. B. Thomas: "Optimal circularly symmetric quantizers, " Franklin Institute Journal, pp. 279-290, 1982.

[34] P. F. Swaszek, Tauwei Ku: "Asymptotic performance of unrestrected polar quantizer," from the conference on information sciences and systems, pp. 266-271, March 1984.

[35] P. F. Swaszek, Tsu Wei Ku: "Asymptotic performance of unrestrected polar quantizer," IEEE Trans. Inform. Theory, vol. IT-32, pp. 330-333, March 1986.

[36] R. M. Gray, A. H. Gray: " Asymptotically optimal quantizers, " IEEE Trans. Inform. Theory, pp. 143-144, Jan 1977.

[37] F. Lu, G. L. Wise: "A further investigation of Max's algorithm for optimum quantization, " IEEE Trans. Communications, vol. COM-33, pp. 746-750, July 1985.

[38] P. F. Swaszek, J. B. Thomas: "K-Dimensional polar quantizers for Gaussian sources, " Princeton University Journal, pp.89-97, Sept 1981.

[39] G. Ungerboeck: "Channel coding with multilevel / phase signals," IEEE Trans. Inform. Theory, vol. IT-28, pp. 55-67, Jan 1982.

[40] L. F. Wei: "Treillis-Coded modulation with multidimensional constellations, " IEEE Trans. Inform. Theory, vol. IT-33, pp. 483-496, July 1987.

[41] X. A. Lebrun: "Codage bas débit d'une variable Gaussienne complexe en coordonnées polaires, " Thèse de Master, Departement de Génie Electrique, Univ. Sherbrook, CANADA, Juin 1989.

[42] R. E. Blahut: "Principles and practice of information theory," Addison Wesley, 1987.

[43] T. Berger: "Rate distortion theory, a mathematical basis for a data compression, " Englewood Cliffs, N.J, Prentice - Hall, 1971.

[44]B. Demidovitch et I. Maron, " Eléments de calcul numérique". Editions Mir. Moscou 1980.

[45] A. Hellil: "Quantification vectorielle des signaux et codage par treillis" , Thèse de PFE, Electronique ENP, Juin 1990.

ANNEXE 1

INTEGRATION PAR LA METHODE DE QUADRATURE DE GAUSS

1. GENERALITES

Si la fonction f(x) est continue sur le segment [a, b] et si l'on connait sa primitive F(x), l'intégrale définie de cette fonction dans les limites de a à b peut être calculée d'aprés la formule de *Newton-Leibniz:*

$$\int_a^b f(x)\,dx = F(b) - F(a) \qquad (1)$$

où $F'(x) = f(x)$

Pourtant dans de nombreux cas la primitive F(x) est trop compliquée ou ne peut s'obtenir à l'aide de procédés élémentaires; il en résulte que le calcul de l'intégrale définie d'aprés la formule (1) peut être trop difficile ou même pratiquement impossible.

Par ailleurs, dans la pratique, l'expression sous le signe somme f(x) est donnée souvent tabulairement et la notion même de primitive perd alors tout son sens. C'est pourquoi les méthodes approchées et, en premier lieu, les *méthodes numériques* de calcul des intégrales définies acquièrent une grande importance.

Le problème de l'intégration numérique d'une fonction consiste à rechercher la valeur de l'intégrale définie à partir de plusieurs valeurs de la fonction sous le signe somme.

Le calcul numérique d'une intégrale simple s'appelle *Quadrature mécanique.* Le procédé usuel pour réaliser une quadrature consiste à remplaçer la fonction donnée f(x) sur le segment concerné [a, b] par une fonction d'interpolation ou d'approximation $\phi(x)$ simple par un polynôme, par exemple, pour admettre approximativement ensuite:

$$\int_a^b f(x)\,dx = \int_a^b \phi(x)\,dx \qquad (2)$$

La fonction $\phi(x)$ doit être telle que le calcul de l'intégrale $\int_a^b \phi(x)\,dx$ soit immédiat.

2. FORMULE DE QUADRATURE DE GAUSS

Dans ce paragraphe nous appliquerons certains renseignements sur les polynômes de Legendre. On appelle polynômes de Legendre les expressions de la forme:

$$P_n(x) = \frac{1}{2^n\, n!}\,\frac{d^n}{dx^n}\left[\left(x^2 - 1\right)^n\right] \qquad (n = 0, 1, 2, \ldots). \qquad (3)$$

Voici les propriétés fondamentales de ces polynômes [46] :

1) $P_n(1) = 1,\ P_n(-1) = (-1)^n \qquad (n = 0,1, 2, \ldots)$

2) $\int_{-1}^{1} P_n(x) \cdot Q_k(x)\,dx = 0 \qquad (k < n),$

où $Q_k(x)$ est un polynôme queconque de degré k.

3) Le polynôme de Legendre $P_n(x)$ possède n racines distinctes et réelles comprises dans l'intervalle [-1,1].

Ci-dessous nous donnons cinq polynômes de Legendre :

$$P_0(x) = 1$$
$$P_1(x) = x$$
$$P_2(x) = \frac{1}{2}\left(3x^2 - 1\right)$$
$$P_3(x) = \frac{1}{2}\left(5x^2 - 3x\right)$$
$$P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right)$$

Déduisons maintenant la *formule de quadrature de Gauss.* Considérons d'abord la fonction y = f(t) définie sur le segment usuel [-1, 1]. Le cas général s'applique aisément à notre cas par substitution linéaire de la variable indépendante.

Voici la formulation du problème: comment sélectionner les points $t_1, t_2 \ldots, t_n$ et les coefficients $A_1, A_2, \ldots, A_n$ pour que la formule de quadrature:

$$\int_{-1}^{1} f(t)\, dt = \sum_{i=1}^{N} A_i f(t_i) \qquad (4)$$

soit exacte pour tout polynôme f(t) de degré N le plus grand possible.

Puisque nous avons 2n constantes t_i et A_i (i = 1, 2, . . . , n), alors que le polynôme de degré 2n-1 est défini par 2n coefficients, ce degré maximal dans le cas général est évidemment N = 2n - 1.

Pour garantir l'égalité (1) il faut et il suffit qu'elle soit vérifiée pour:

$$f(t) = 1, t, t^2, \ldots, t^{2n-1}.$$

En effet, en posant:

$$\int_{-1}^{1} f(t)\, dt = \sum_{i=1}^{n} A_i t_i^k \qquad (k = 0, 1, 2, \ldots, 2n - 1) \qquad (5)$$

et

$$f(t) = \sum_{k=0}^{2n-1} C_k t^k$$

$$\int_{-1}^{1} f(t)\, dt = \sum_{k=0}^{2n-1} C_k \int_{-1}^{1} t^k\, dt = \sum_{k=0}^{2n-1} C_k \sum_{i=1}^{n} A_i t_i^k = \sum_{i=1}^{n} A_i \sum_{k=0}^{2n-1} C_k t_i^k = \sum A_i f(t_i).$$

Ainsi, en tenant compte des relations:

$$\int_{-1}^{1} t^k\, dt = \frac{1 - (-1)^{k+1}}{k + 1} = \begin{cases} \dfrac{2}{k + 1} & \text{avec k pair ;} \\ 0 & \text{avec k impair ;} \end{cases}$$

on peut donc conclure que pour résoudre le problème posé [44], il suffit de déterminer t_i et A_i à partir du système de 2n équations:

$$\left.\begin{aligned}
&\sum_{i=1}^{n} A_i = 2 \\
&\sum_{i=1}^{n} A_i t_i = 0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\sum_{i=1}^{n} A_i t_i^{2n-2} = \frac{2}{2n-1} \\
&\sum_{i=1}^{n} A_i t_i^{2n-1} = 0
\end{aligned}\right\} \qquad (6)$$

Le système (6) est non-linéaire et sa résolution par la voie usuelle présente de grandes difficultés. Toutefois, on peut appliquer à cet effet l'artifice suivant:

Considérons les polynômes:

$$f(t) = t^k P_n(t) \qquad (k = 0, 1, \ldots, n-1)$$

Où $P_n(t)$ est le polynôme de Legendre.

Les degrés de ce polynôme ne dépassant pas $2n - 1$, ces polynômes doivent vérifier en vertu du système (6) la formule (4) est:

$$\int_{-1}^{1} t^k P_n(t)\, dt = \sum_{i=1}^{n} A_i t_i^k P_n(t_i) \qquad (k = 0, 1, \ldots, n-1) \qquad (7)$$

D'autre part, l'orthogonalité des polynômes de Legendre (propriété (2)) rend vraies les égalités:

$$\int_{-1}^{1} t^k P_n(t)\, dt = 0 \qquad \text{pour } k < n$$

$$\sum_{i=1}^{n} A_i t_i^k P_n(t_i) = 0 \qquad (k = 0, 1, \ldots, n-1) \qquad (8)$$

Par ailleurs, si l'on pose:

$$P_n(t_i) = 0 \qquad (i = 0, 1, \ldots, n) \qquad (9)$$

Les égalités (8) seront nécessairement vraies quelles que soient les valeurs A_i. On sait

(propriété (6)) que ces zéros sont réels, distincts et compris dans l'intervalle (-1, 1). Si l'on connait les abscisses t_i, on trouve facilement à partir du système linéaire des n premières équations du système (6) les constantes A_i (i = 1, 2, ... n). Le déterminant de ce sous-système est un déterminant de Vandermonde:

$$D = \prod_{i>j} (t_i - t_j) \neq 0$$

et, par suite, la détermination des A_i est univoque.

On peut montrer [44] que la formule (4) à coefficients ainsi déterminés est exacte pour tout polynôme de degré égal ou inférieur à 2n - 1.
La formule où les t_i sont les zéros du polynôme de Legendre $P_n(t)$ et où les A_i (i =1, 2,.. .n) sont définies à partir du système (6) s'appelle *formule de quadrature de Gauss.*

3.Calcul de l'intégrale généralisée par la formule de Gauss:

Examinons maintenant l'utilisation de la formule de Gauss pour calculer l'intégrale généralisée:

$$\int_a^b f(x)\, dx$$

En changeant la variable:

$$x = \frac{b + a}{2} + \frac{b - a}{2} t$$

on obtient:

$$\int_a^b f(x)\, dx = \frac{b - a}{2} \int_{-1}^{1} f\left(\frac{b + a}{2} + \frac{b - a}{2} t\right) dt$$

Appliquons à la dernière intégrale la formule de Gauss (4), on aura:

$$\int_a^b f(x)\,dx = \frac{b-a}{2} \sum_{i=1}^{n} A_i\, f(x_i) \qquad (10)$$

Remarque:

Lors du calcul, la formule de Gauss présente l'inconvénient que les abscisses des points t_i et les coefficients A_i sont en général des nombres irrationnels. Cet inconvénient est en partie compensé par une précision élevée en présence d'un nombre d'ordonnées relativement petit.

ANNEXE 2

Méthodes de calcul de variations

Introduction

La recherche des maxima et minima des fonctions d'une variable est un problème bien connu. Si f(x) est une fonction continue et admettant une dérivée continue dans un intervalle [a, b], les extrema relatifs de f(x) à l'interieur de cet intervalle sont atteints en des points où f'(x) = 0.

Nous allons généraliser ce problème en conservant les hypothèses de régularité, mais en remplaçant la fonction d'une variable f(x) d'abord par une fonction de plusieurs variables, puis par une fonction d'une infinité de variables dans des conditions qui seront precisées.

1. Extrema de fonctions à n variables:

Soit $f(x_1, x_2, ..., x_n)$ une fonction à n variables, continue et pourvue de derivées continues dans un certain domaîne $\mathcal{D}$.

Pour que f admette un extremum en un point interieur au domaine (frontières exclues), il est nécéssaire qu'en ce point:

$$\frac{\partial f}{\partial x_1} = \ldots = \frac{\partial f}{\partial x_n} = 0 \qquad (1)$$

2. Extrema liés de fonctions à n variables:

Bien souvent le problème de la détermination des extremums d'une fonction se ramène à la recherche d'extremums d'une fonction de plusieurs variables qui ne sont pas indépendantes, mais liées entres elles par certaines conditions supplémentaires (par exemple, assujetties à vérifier certaines équations).

Considérons tout d'abord le problème de l'extremum lié d'une fonction de deux variables quand elles ne sont liées entre elles que par une seule condition. Soit à calculer les extremums de la fonction:

$$u = f(x, y) \qquad (2)$$

où x et y sont liés par l'équation:

$$\phi(x, y) = 0 \qquad (3)$$

La condition (2) implique que seule l'une des variables x et y est indépendante, par exemple x, car y est déterminé à partir de l'égalité (2) comme fonction de x. Si l'on résoud l'équation (2) par rapport à y, et si l'on substitue dans l'égalité (1) l'expression trouvée pour y, u sera fonction d'une seule variable x et le problème sera ainsi ramené à l'étude de l'extremum d'une seule variable indépendante x.

Mais on peut résoudre le problème posé sans qu'il soit nécéssaire de résoudre l'équation (2) par rapport à x ou à y. La dérivée de u par rapport à x doit s'annuler pour les valeurs de x telles que la fonction u est susceptible d'admettre un extremum.

Calculons $\frac{du}{dx}$ à partir de (2), sachant que y est une fonction de x:

$$\frac{du}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x}$$

Par conséquent, aux points d'extremum

$$\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0 \qquad (4)$$

On trouve de l'égalité (3):

$$\frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y}\frac{dy}{dx} = 0 \qquad (5)$$

Cette équation est satisfaite pour tous les x et y vérifiant l'équation (3).

Multiplions tous les termes de l'égalité (5) par un coefficient indéterminé λ et ajoutons-les aux termes correspondants de l'égalité (4). Nous trouvons:

$$\left(\frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx}\right) + \lambda\left(\frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y}\frac{dy}{dx}\right) = 0$$

ou

$$\left(\frac{\partial f}{\partial x} + \lambda\frac{\partial \phi}{\partial x}\right) + \left(\frac{\partial f}{\partial y} + \lambda\frac{\partial \phi}{\partial y}\right)\frac{dy}{dx} = 0 \qquad (6)$$

Cette égalité a eu lieu pour tous les points où il y a un extremum. Choisissons λ de manière que pour les valeurs de x et y telles que la fonction u présente un extremum, la seconde parenthèse de l'égalité (6) s'annule:

$$\frac{\partial f}{\partial y} + \lambda\frac{\partial \phi}{\partial y} = 0$$

Mais alors pour ces valeurs de x et de y il vient de l'égalité (6) que:

$$\frac{\partial f}{\partial x} + \lambda \frac{\partial \phi}{\partial x} = 0$$

Ainsi aux points d'extremum les trois équations:

$$\begin{cases} \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial \phi}{\partial x} = 0 \\ \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial \phi}{\partial x} = 0 \\ \phi(x, y) = 0 \end{cases} \qquad (7)$$

à trois inconnues x, y, λ sont vérifiées. La résolution de ces équations nous donne les inconnues x, y et λ qui n'a joué qu'un rôle auxilliaire et dont nous n'aurons pas besoin.

Il est clair que les équations (7) sont les conditions nécéssaires pour l'existence d'un extremum lié, c'est-à-dire en tout point d'extremum les équations (7) sont vérifiées. La réciproque n'est pas vraie, car la fonction peut ne pas avoir d'extremum lié pour les valeurs correspondantes de x, y et λ tirées des équations (7). On est donc amené à entreprendre une étude détaillée de la nature du point critique. En résolvant des problèmes concrets, on peut parfois déterminer la nature du point critique d'aprés le caractère même du problème. Remarquons que les premiers membres des équations (7) sont les dérivées partielles par rapport aux variables x,y , λ de la fonction:

$$F(x, y, \lambda) = f(x, y) + \lambda \phi(x, y) \qquad (8)$$

Ainsi, pour trouver les valeurs de x et y vérifiant la condition (3) pour lesquelles la fonction u = f(x, y) admet un extremum lié, il faut former la fonction auxilliaire (8), égaler à zéro ses dérivées partielles par rapport à x, y, λ et déterminer les inconnues x, y (ainsi que le facteur auxillaire λ) des trois équations (7) ainsi obtenues. Cette méthode peut être aisément étendue à la recherche des extremums liés d'une fonction d'un nombre quelconque de variables.

Soit à déterminer les extremums de la fonction $u = f(x_1, x_2, ..., x_n)$ à n variables $x_1, x_2, ..., x_n$ assujetties à vérifier les m équations ($m < n$) :

$$\begin{cases} \phi_1(x_1, x_2, ..., x_n) = 0 \\ \phi_2(x_1, x_2, ..., x_n) = 0 \\ \cdots\cdots\cdots\cdots\cdots \\ \phi_m(x_1, x_2, ..., x_n) = 0 \end{cases} \qquad (9)$$

Pour trouver les valeurs de $x_1, x_2, ..., x_n$ susceptibles de donner des extremums liés de cette fonction, on doit former la fonction auxillaire:

$$F(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = f(x_1, x_2, ..., x_n) + \lambda_1 \phi_1(x_1, x_2, ..., x_n) + \lambda_2 \phi_2(x_1, x_2, ..., x_n) + ... + \lambda_m \phi_m(x_1, x_2, ..., x_n)$$

Egaler à zéro ses dérivées partielles par rapport à $x_1, x_2, ..., x_n$:

$$\begin{cases} \dfrac{\partial \phi}{\partial x_1} + \lambda_1 \dfrac{\partial \phi}{\partial x_1} + ... + \lambda_m \dfrac{\partial \phi}{\partial x_1} = 0 \\ \dfrac{\partial \phi}{\partial x_2} + \lambda_1 \dfrac{\partial \phi}{\partial x_2} + ... + \lambda_m \dfrac{\partial \phi}{\partial x_2} = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \dfrac{\partial \phi}{\partial x_n} + \lambda_1 \dfrac{\partial \phi}{\partial x_n} + ... + \lambda_m \dfrac{\partial \phi}{\partial x_n} = 0 \end{cases} \qquad (10)$$

et déterminer des m+n équations (9) et (10), $x_1, x_2, ..., x_n$ et les inconnues auxilliaires $\lambda_1, \lambda_2, ..., \lambda_m$. Tout comme pour une fonction de deux variables, la question de savoir si aux valeurs trouvées des variables correspond véritablement un extremum de la fonction ou si cette dernière n'admet pas d'extremum en ce point reste sans réponse dans le cas général. Cette question sera résolue à l'aide de considérations particulières découlant de chaque problème concret.

3. Lemme fondamental du calcul de variation:

Au lieu d'extrema de fonctions d'un nombre fini de variables, on a parfois à se poser des problèmes d'extrema plus compliqués, où les inconnues sont elles-mêmes des fonctions et où l'on à rendre extrema une certaine fonctionnelle de ces fonctions. De tels problèmes sont dans une certaine mesure, l'extension directe de ceux que nous avons traités aux paragraphes précédents.

Théorème:

Si l'intégrale

$$\int_{t_1}^{t_2} \xi(t)\, x(t)\, dt$$

est nulle, quelle que soit la fonction x(t) *continue et nulle pour* $t = t_1$ *et* $t = t_2$ *, et si* x(t) *est une fonction continue, alors* x(t) *est identiquement nulle.*

4. Variations d'une intégrale à limites fixes:

4.1 Positionnement du problème:

On considère dans le plan des x, t une famille de courbes Γ dont l'équation est donnée par:

$$x = \varphi(t, \varepsilon)$$

Cette fonction dépend du paramètre ε. φ(t, ε) doit être une fonction continue et pourvue de dérivées partielles d'ordre 1 et 2 , continues pour un certain ensemble de valeurs de t et ε.

Soit maintenant f(t,x,x') une fonction de trois variables, continue et pourvue de dérivées premières et secondes continues dans un domaine convenable des t, x, x'. L'expression " domaine convenable " implique qu'on pourra donner à t, x, x' des valeurs qui soient celles que prennent t, φ, ∂φ/∂t.

Considérons l'intégrale:

$$I(\varepsilon) = \int_{t_1}^{t_2} f\left(t, \varphi, \frac{\partial \varphi}{\partial t}\right) dt$$

où $x = \varphi(t, \varepsilon)$ et $x' = \frac{\partial \varphi}{\partial t}$

Cette intégrale dépend de ε car φ et $\frac{\partial \varphi}{\partial t}$ dépendent de ε. On calcule sa dérivée par dérivation sous le signe somme :

$$I'(\varepsilon) = \int_{t_1}^{t_2} \left(\frac{\partial f}{\partial x} \frac{\partial \phi}{\partial \varepsilon} + \frac{\partial f}{\partial x'} \frac{\partial^2 \varphi}{\partial t\, \partial \varepsilon} \right) dt$$

Or une intégration par partie montre que :

$$\int_{t_1}^{t_2} \frac{\partial f}{\partial x'} \frac{\partial^2 \varphi}{\partial t \, \partial \varepsilon} dt = \left(\frac{\partial f}{\partial x'} \frac{\partial \phi}{\partial \varepsilon} \right)_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial \phi}{\partial \varepsilon} \frac{d}{dt} \left(\frac{\partial f}{\partial x'} \right) dt$$

Comme aux points t_1 et t_2 , φ ne dépend pas de ε, la dérivée $\frac{\partial \phi}{\partial \varepsilon}$ s'annule aux limites, et le terme intégré disparait, il reste:

$$I'(\varepsilon) = \int_{t_1}^{t_2} \frac{\partial \phi}{\partial \varepsilon} \left[\frac{\partial f}{\partial x} - \frac{d}{dt} \left(\frac{\partial f}{\partial x'} \right) \right] dt \qquad (11)$$

4.2. Equation d'Euler-Lagrange:

Nous nous posons le problème suivant: Trouver les fonctions x = x(t), continues et pourvues de dérivées premières et secondes continues; qui prennent pour $t = t_1$ et $t = t_2$ des valeurs données x_1 et x_2, et qui rendent extrema l'intégrale:

$$I = \int_{t_1}^{t_2} f(t, x, x') \, dt$$

où f vérifie les conditions données au paragraphe 4.1.

Supposons que le problème posé ait une solution x(t). Remplaçons dans l'intégrale I la fonction x(t) par la fonction $x(t) + \varepsilon \xi(t)$, où $\xi(t)$ est une fonction arbitraire (satisfaisant aux hypothèses de dérivabilité) et où ε est un paramètre. Une condition nécéssaire pour que x(t) soit une solution, est que $\partial I/\partial \varepsilon$ soit nulle pour $\varepsilon = 0$ | 19|, quelle que soit $\xi(t)$ s'annulant pour $t = t_1$ et $t = t_2$. Connaissant l'expression de $\partial I/\partial \varepsilon$ donnée par la formule (11), et sachant que :

$$\varphi(t, \varepsilon) = x(t) + \varepsilon \, \xi(t)$$

Donc $$\frac{\partial \varphi}{\partial \varepsilon} = \xi(t)$$

Alors $$I'(\varepsilon) = \frac{\partial I}{\partial \varepsilon} = \int_{t_1}^{t_2} \xi(t) \left[\frac{\partial f}{\partial x} - \frac{d}{dt} \left(\frac{\partial f}{\partial x'} \right) \right] dt \qquad (12)$$

L'intégrale $\partial I/\partial \varepsilon$ doit être nulle quelque soit la fonction $\xi(t)$, continue et nulle aux limites. D'aprés le lemme fondamental du calcul de variations (donné au paragraphe 3), la quantité:

$$\frac{\partial f}{\partial x} - \frac{d}{dt}\left(\frac{\partial f}{\partial x'}\right) = 0 \qquad (13)$$

L'expression (13) représente l'équation d'Euler-Lagrange. Son éventuelle solution x(t) représentera l'extremum à l'intégrale **I**.

4.2. Extrema liés : Problèmes isopérimétriques

Le problème des extrema liés dans le cas de variations d'une intégrale s'énonce de la façon suivante:

Trouver les fonctions x(t) qui rendent extrema l'intégrale:

$$I = \int_{t_1}^{t_2} f(t, x, x')\, dt$$

et qui prennent pour $t = t_1$ et $t = t_2$ des valeurs données x_1 et x_2 de telle sorte qu'une seconde intégrale:

$$J = \int_{t_1}^{t_2} g(t, x, x')\, dt$$

ait une valeur donnée **a.**

Pour résoudre ce problème, on écrit que l'intégrale I plus petite (ou plus grande) pour la fonction cherchée x(t), que pour toute fonction voisine de x(t), et prenant les mêmes valeurs que x(t) pour $t = t_1$ et $t = t_2$. Dans le précédent paragraphe où il ne s'agissait pas d'extrema liés, il a suffit de faire appel à des fonctions de la forme:

$$\varphi(t, \varepsilon) = x(t) + \varepsilon\, \xi(t)$$

dépendant d'un paramètre. Mais ici, ces fonctions doivent vérifier la condition:

$$\int_{t_1}^{t_2} g(t, x + \varepsilon\xi, x' + \varepsilon\xi')\, dt = \mathbf{a}$$

de sorte que $\xi(t)$ n'est plus une fonction arbitraire même en se fixant ε. On est conduit à élargir le champ des fonctions soumises à la comparaison en y introduisant des fonctions qui dépendent de deux paramètres:

$$\varphi(t, \varepsilon_1, \varepsilon_2) = x(t) + \varepsilon_1\, \xi_1(t) + \varepsilon_2\, \xi_2(t) \qquad (14)$$

où $\xi_1(t)$ et $\xi_2(t)$ sont deux fonctions arbitraires, nulle pour $t = t_1$ et $t = t_2$. I devient une fonction de ε_1 et ε_2, qui doit être extremum pour $\varepsilon_1 = 0$, $\varepsilon_2 = 0$, moyennant que ε_1 et ε_2 soient liés par la condition:

$$J(\varepsilon_1, \varepsilon_2) = \int_{t_1}^{t_2} g(t, x + \varepsilon_1 \xi_1 + \varepsilon_2 \xi_2, x' + \varepsilon_1 \xi_1' + \varepsilon_2 \xi_2') \, dt = a$$

On cherche donc les extrema liés d'une fonction de deux variables ε_1, ε_2. D'aprés le paragraphe 2, le problème se résoud par la méthode des multiplicateurs de Lagrange. On cherche donc les extrema de la fonction :

$$I^* = I(\varepsilon_1, \varepsilon_2) + \lambda J(\varepsilon_1, \varepsilon_2) = \int_{t_1}^{t_2} f^*(t, x, x') \, dt \qquad (15)$$

Où

$f^* = f + \lambda g$

Avec λ le multiplicateur de Lagrange

Pour obtenir les extrema de cette fonction, il faut que:

$$\frac{\partial f}{\partial \varepsilon_1} = \frac{\partial f}{\partial \varepsilon_2} = 0 \qquad \text{quand} \quad \varepsilon_1 = \varepsilon_2 = 0$$

A partir de (15) et en utilisant (14), on aura:

$$\frac{\partial f^*}{\partial \varepsilon_j} = \int_{t_1}^{t_2} \left\{ \frac{\partial f^*}{\partial \phi} \frac{\partial \varphi}{\partial \varepsilon_j} + \frac{\partial f^*}{\partial \phi'} \frac{\partial \varphi'}{\partial \varepsilon_j} \right\} dt = \int_{t_1}^{t_2} \left\{ \frac{\partial f^*}{\partial \phi} \xi_j + \frac{\partial f^*}{\partial \phi'} \xi_j' \right\} dt \qquad (j = 1, 2)$$

En mettant $\varepsilon_1 = \varepsilon_2 = 0$ et suivant l'expression (14), (φ, φ') est remplacée par (x, x') d'où on a:

$$\left. \frac{\partial f^*}{\partial \varepsilon_j} \right|_0 = \int_{t_1}^{t_2} \left\{ \left\{ \frac{\partial f^*}{\partial x} \xi_j + \frac{\partial f^*}{\partial x'} \xi_j' \right\}_j \right\} dt = 0 \qquad (j = 1, 2) \qquad (16)$$

on notera que l'indice 0 indique que $\varepsilon_1 = \varepsilon_2 = 0$. En intégrant par parties :

$$(16) \quad \Leftrightarrow \quad \frac{\partial f^*}{\partial \varepsilon} = \int_{t_1}^{t_2} \xi_j(t) \left[\frac{\partial f^*}{\partial x} - \frac{d}{dt} \left(\frac{\partial f^*}{\partial x'} \right) \right] dt \quad (j = 1, 2) \qquad (17)$$

En utilisant le lemme de base de calcul de variations donné au paragraphe 3, on trouve que :

$$\frac{\partial f^*}{\partial x} - \frac{d}{dt}\left(\frac{\partial f^*}{\partial x'}\right) = 0 \qquad (18)$$

On retrouve l'équation d'Euler-Lagrange, mais dans ce cas donnée pour le calcul d'extrema liés où:

$$f^* = f + \lambda\ g$$

Remarque:

Nous remarquons qu'un problème similaire s'est posé au chapitre 5 lors de l'optimisation de la distorsion D par rapport au nombre de niveaux de phase.

ANNEXE 3

Rappels sur les inégalités de Hölder

Introduction

L'inégalité de Hölder s'applique dans deux cas différents:

- Dans le cas de variables discrètes : c'est à dire lorsqu'il s'agit de de somme de valeurs élémentaires telles que les séries de nombres comme on le verra un plus loin.

- Dans le cas de variables continues: C'est à dire lorsque l'inégalité s'applique à des intégrales de fonctions.

1. Cas de valeurs discrètes:

1.1.Théorème

Si $\alpha, \beta, \ldots, \lambda$ sont des nombres positifs, et $\alpha + \beta + \ldots + \lambda = 1$, alors on définit l'inégalité de Hölder par:

$$\sum a^{\alpha} b^{\beta} \ldots l^{\lambda} < \left(\sum a\right)^{\alpha} \left(\sum b\right)^{\beta} \ldots \left(\sum l\right)^{\lambda} \qquad (1)$$

L'inégalité se transforme en égalité dans deux cas:

- *Si tous les nombres a, b . . . l sont proportionnels entre eux.*
- *ou l'un des nombres soit nul.*

1.2. Extension de l´inégalité de Holder:

Supposons qu´il existe un nombre $k \neq 0$ et $k \neq 1$, et un nombre k´ (conjugué) de k. k' est dit conjugué de k si :

$$k' = \frac{k}{k-1}$$

ou encore : $$\frac{1}{k} + \frac{1}{k'} = 1$$

On parle d'inégalité de Holder si pour k >1 , on a :

$$\sum ab \rangle \left(\sum a^k\right)^{1/k} \left(\sum b^{k'}\right)^{1/k'}$$

Sinon il y a égalité si et seulement si (a^k) est proportionnel à ($b^{k'}$) ou k < 1:

$$\sum ab \rangle \left(\sum a^k\right)^{1/k} \left(\sum b^{k'}\right)^{1/k'}$$

Il y a égalité si (a^k) est proportionnel à ($b^{k'}$) :

On montre que cette égalité est nécéssaire et suffisante [18].

2. Inégalité de Holder appliquée aux intégrales :

2.1. Théorème :

Soient $\alpha,\beta,\cdots,\lambda$ un nombre fini de valeurs positives avec $\alpha + \beta + \cdots + \lambda = 1$, alors on a :

$$\int f^{\alpha} g^{\beta} \cdots l^{\lambda} dx \leq \left(\int f dx\right)^{\alpha} \left(\int g dx\right)^{\beta} \cdots \left(\int l dx\right)^{\lambda}$$

Sinon on a égalité si l'une des fonctions est nulle ou si toutes les fonctions sont proportionnelles les unes aux autres.

2.2. Corollaire :

Si k ⟩ 1 alors : $$\frac{1}{k} + \frac{1}{k'} = 1$$

et

$$\int fg dx \langle \left(\int f^k dx\right)^{1/k} \left(\int g^{k'} dx\right)^{1/k'}$$