

Mémoire d'habilitation à diriger des recherches

Entropies et radiotechnique

Jean-François Bercher

Adresse

Suite de l'adresse

Version du 16 avril 2009

Le 16 avril 2009

Soutenance un jour de 2009, devant le jury constitué de :

M	Prénom	NOM	Rapporteur
M.	Prénom	NOM	Rapporteur
M.	Prénom	NOM	Rapporteur
M.	Prénom	NOM	Examineur
M.	Prénom	NOM	Examineur
M.	Prénom	NOM	Examineur

Table des matières

0	Préambule	1
1	Bilan succinct	7
1.1	Traitement et théorie de l'information	7
1.2	Radiotechnique	10
1.3	Publications	13
2	Entropie et maximum d'entropie	19
2.1	Maximum d'entropie sur la moyenne	19
2.2	Maximum d'entropie sur la moyenne itératif	21
2.3	Maximum d'entropie avec contraintes aléatoires	22
2.4	Estimation de l'entropie	23
2.5	Inégalité de convolution pour l'entropie de Rényi	24
3	Entropies de Rényi-Tsallis	27
3.1	Contexte	27
3.2	L'entropie de Rényi-Tsallis comme solution d'un problème standard de maximum d'entropie	29
3.3	Distributions de Tsallis et distribution des excès	31
3.4	Fonctionnelles d'entropie dérivées de la divergence de Rényi	33
4	Information de Fisher	37
4.1	Quelques éléments de contexte	37
4.2	Inégalités sur l'information de Fisher	39
4.3	Plan d'information de Fisher-Shannon	39
4.4	Minimum de l'information de Fisher	41
5	Problèmes d'architecture Radio	45
5.1	Une architecture d'émetteur pour les modulations à enveloppe variable	46
5.2	Boucle à Verrouillage de phase numérique	48
5.2.1	Simulation	48
5.2.2	Contributions adaptatives	50
5.3	Référence très haute fréquence stabilisée à base de BAW	51
6	Synchronisation des voies dans les architectures polaires	53
6.1	Effet d'un mauvais alignement	53
6.2	Faisabilité de l'alignement entre voies	54
6.3	Un algorithme de correction	55
6.4	Extensions	58
7	Points de suspension	59
8	Recueil de publications	67

8.1	J.-F. Bercher, “Maximum entropy with fluctuating constraints - The example of K-distributions,” <i>Physics Letters A</i> , vol. 372, no. 24, pp. 4361–4363, Jun. 2008.	67
8.2	J.-F. Bercher and C. Vignat, “Estimating the entropy of a signal with applications,” <i>IEEE Transactions on Signal Processing</i> , vol. 48, no. 6, pp. 1687–1694, 2000.	73
8.3	J.-F. Bercher, “Tsallis distribution as a standard maximum entropy solution with ‘tail’ constraint,” <i>Physics Letters A</i> , vol. 372, no. 35, pp. 5657–5659, Aug. 2008.	85
8.4	J.-F. Bercher and C. Vignat, “A new look at q -exponential distributions via excess statistics,” <i>Physica A</i> , vol. 387, no. 22, pp. 5422–5432, Sep. 2008.	91
8.5	J.-F. Bercher, “On some entropy functionals derived from Rényi information divergence,” <i>Information Sciences</i> , vol. 178, pp. 2489–2506, Jun. 2008.	103
8.6	C. Vignat and J.-F. Bercher, “On Fisher information inequalities and score functions in non-invertible linear systems,” <i>Journal of Inequalities in Pure and Applied Mathematics</i> , vol. 4, no. 4, p. Article 71, 2003.	123
8.7	C. Vignat and J. F. Bercher, “Analysis of signals in the Fisher-Shannon information plane,” <i>Physics Letters A</i> , vol. 312, no. 1-2, pp. 27–33, Jun. 2003.	135
8.8	J.-F. Bercher and C. Vignat, “A note on minimum Fisher information distributions with restricted support and fixed variance,” <i>Submitted to Information Sciences</i> , Mar. 2009. . .	145
8.9	C. Berland, I. Hibon, J.-F. Bercher et al. « New transmitter architecture for nonconstant envelope modulation », <i>IEEE Trans. on Circuits and Systems</i> , vol. 53, Issue 1, pp 13 - 17, 2006.	161
8.10	C. Joubert, J.-F. Bercher, and G. Baudoin, « Contribution to the study of a Phase-Domain ADPLL », version longue de l’article , “Contributions to the analysis and design of an ADPLL,” in <i>13th IEEE International Conference on Electronics, Circuits and Systems, 2006. (ICECS) 2006.</i> , 2006, pp. 322–325.	169
8.11	P. Guillot, P. Philippe, C. Berland, and J. Bercher, “A 2GHz 65nm CMOS digitally-tuned BAW oscillator,” in <i>15th IEEE International Conference on Electronics, Circuits and Systems, 2008. ICECS 2008.</i> , Sep. 2008, pp. 722–725.	177
8.12	J.-F. Bercher and C. Berland, “Envelope and phase delays correction in an EER radio architecture,” <i>Analog Integrated Circuits and Signal Processing</i> , vol. 55, pp. 21–35, Apr. 2008.	185
8.13	J.-F. Bercher and C. Berland, “Adaptive delays alignment in polar transmitter architecture,” submitted to <i>IEEE trans. on Vehicular Technologies</i> , april 2009.	203

À Rosalind Franklin

Chapitre 0

Préambule

Ce document présente une synthèse de mes contributions dans le domaine de la recherche depuis la soutenance de ma thèse de doctorat en 1995. L'activité rapportée ici est celle d'un enseignant-chercheur, et s'est donc déroulée en parallèle d'une activité d'enseignement comprenant également certaines charges pédagogiques, d'organisation ou d'ordre administratif. Je suis enseignant-chercheur à l'ESIEE depuis 1996 (professeur assistant, puis professeur associé (2002) au sens anglo-saxon). L'ESIEE est la grande école technique de la Chambre de Commerce et d'Industrie de Paris (la CCIP est un établissement public, l'école étant sous tutelle du ministère de l'industrie).

Lors de la préparation de ma thèse de doctorat, je m'étais intéressé à une méthode de résolution problèmes inverses reposant sur une approche à maximum d'entropie. C'est assez naturellement que j'ai continué à investir sur le thème du traitement et la théorie de l'information, et notamment sur l'étude de mesures d'information ou de complexité pour l'analyse, le filtrage et la restauration de signaux et images, avec des applications aux télécommunications et à la physique statistique. D'un autre côté, je n'ai pas complètement oublié que j'ai débuté mes études supérieures dans le domaine de l'électronique (bien qu'assez fortuitement). C'est par l'interaction avec les collègues de l'ESIEE, et en coordination avec des activités d'enseignement, que j'ai été amené à m'intéresser à certains problèmes d'architecture des émetteurs-récepteurs radio. Cette activité, que j'appelle "Radiotechnique" (en écho au titre du livre fascinant [1]), s'est beaucoup développée ces dernières années.

Ces deux volets de mes recherches sont menés tantôt alternativement, parfois conjointement. Les aspects concernant le traitement de l'information et ses applications sont présentés dans les chapitres 2 à 4, quand les aspects sur la radiotechnique sont décrits dans les chapitres 5 et 6. Il aurait été possible, à l'occasion de ce mémoire, de relier ces deux thématiques ; par exemple par le biais du codage, ou de l'égalisation. En réalité, ces deux axes sont indépendants, si ce n'est qu'ils partagent un même acteur, et je ne souhaite pas afficher ici de montage artificiel. Il est bien possible que ces deux axes puissent se rejoindre ou échanger dans le futur. Le présent mémoire n'est ni une finalité, ni un achèvement, bien heureusement, et en recherche comme ailleurs, la cohérence s'imposera par ses propres qualités.

L'exercice de rédaction et soumission d'un manuscrit de HdR permet de faire le point sur un bout de carrière, et de recueillir l'appréciation d'autrui (mais on sait bien, sauf peut-être dans les sommets de l'organisation de l'État, que l'évaluation de la recherche est permanente). L'exercice constitue un jalon dans de devenir d'un enseignant chercheur. Incidemment, le grade de HdR est important pour les autres : les dossiers de reconnaissance, de demande de crédit, les réponses aux appels d'offre, les évaluations d'équipe ou d'établissement comptabilisent l'HdR ou le nombre de HdR comme un indicateur de qualité. C'est également un jalon en termes personnels, parce qu'il permet sans doute d'aller plus loin, en gagnant un peu en stature (encore le regard des autres) et le droit de développer des recherches autonomes. En réalité, cela fait bien longtemps que j'assume des recherches personnelles. Celles-ci ne sont toutefois pas sans interaction avec autrui. En particulier C. Vignat et C. Berland auront contribué largement aux thématiques et résultats présentés ici. La possibilité de diriger en direct des travaux sur le thème de la Radiotechnique permettra de développer plus avant ces aspects. La possibilité de pouvoir rechercher des

étudiants pour des travaux sur les thèmes plus amont, liés aux aspects entropiques est également fort intéressante.

Ma démarche de recherche est de type spagyriste – l’expression est de [2], et consiste à extraire, assembler, en vue de construire autre chose. Je conçois et profite de l’activité de recherche pour continuer à apprendre, et pas simplement à produire.

En rapport avec la dédicace en début de ce mémoire, je rappellerai les grands principes, qui selon J. Watson (co-découvreur de la structure de l’ADN), devrait conduire une carrière de chercheur. Si ceux-ci peuvent être un peu critiquables (notamment en raison de la personnalité de leur auteur), ils me semblent un bon guide (et une bonne motivation au présent manuscrit) pour aller vers une recherche originale et de qualité. Les voici : (i) tuer le père (au sens figuré, s’entend), le quitter et mener des recherches hors de son haut patronage (ii) ne fréquenter que des gens d’un niveau scientifique supérieur, ou au pire égal, de manière à profiter d’une dynamique et progresser (iii) refuser de faire ce que l’on n’a pas envie de faire (iv) trouver une position, ou se mettre dans une situation telle qu’on puisse mettre en application les points précédents.

La suite du document est organisée comme suit. Les pages suivantes présentent un bref curriculum-vitæ, quelques mots sur mon activité d’enseignement, sur la formation doctorale, l’encadrement de stages et thèses. Dans le premier chapitre, je présente rapidement un bilan de mes différentes contributions, le chapitre se terminant en page 13 par la liste des publications. Les cinq chapitres suivants reprennent une grande partie des résultats que j’ai pu obtenir, en s’appuyant sur les articles placés en annexe. J’essaie de donner l’essentiel du contexte, de la démarche et des résultats concernant chacun des travaux présentés. Lorsque les articles ne sont pas annexés, je donne parfois un peu plus de détails et d’illustration. Bien entendu et comme il se doit, le mémoire se termine sur quelques perspectives.

BREF CURRICULUM VITÆ

Nom patronymique : *BERCHER*

Prénom : *Jean-François*

Date et lieu de naissance : *12 août 1965 à Marseille*

Nationalité : *Française*

Situation de famille : *Marié, trois enfants.*

Fonctions actuelles : *Enseignant-chercheur à l'École Supérieure d'Ingénieurs en Électrotechnique et Électronique (ESIEE).
Cité Descartes, BP 99, 93 162 NOISY LE GRAND Cedex
Tél. 01-45-92-65-15
email : jf.bercher@esiee.fr
et Laboratoire d'Informatique Gaspard Monge, LIGM, UMR 8049, Équipe Signal et Communications*

Titres universitaires

Doctorat de l'Université Paris-Sud (UPS),

Spécialité : « Automatique et traitement du signal »

Intitulé : « Développement de critères de nature entropique pour la résolution de problèmes inverses linéaires »

Directeur de thèse : M. Guy DEMOMENT, professeur à l'Université de Paris-Sud,

Date et lieu de soutenance : le 02 février 1995 à l'École Supérieure d'Électricité,

Mention : mention très honorable, avec les félicitations du jury.

Diplôme d'Études Approfondies (DEA) en « Signal Image Parole » obtenu mention Très Bien en septembre 1990 à l'Institut National Polytechnique de Grenoble (INPG).

Diplôme d'Ingénieur Électricien spécialité Traitement du Signal obtenu mention Bien en juin 1990 à l'École Nationale Supérieure d'Ingénieurs Électriciens de Grenoble (ENSIEG).

Maîtrise des Sciences et Techniques en Génie médical et Instrumentation obtenue en septembre 1988 mention Bien à l'Université Paris XII.

Diplôme Universitaire en Technologie (DUT) option Électronique et informatique industrielle obtenu en Juillet 1986 à l'IUT de Cachan.

Activités professionnelles

- septembre 2007-septembre 2008 - congé de perfectionnement au laboratoire des signaux et systèmes (Orsay) et dans l'équipe Signal et Communication du LabInfo (IGM-Marne la Vallée)
- depuis septembre 1996 : enseignant-chercheur à l'ESIEE (départements “signaux et télécoms” et “modélisation et simulation”). Ancien responsable de l'option Traitement du signal et Télécommunications, membre élu de la commission d'évaluation (commission de spécialistes)
- Attaché temporaire d'enseignement et de recherche - institut Galilée, Paris 13 (1995-96), université Paris-Sud (1994-95).

Activité de recherche

Depuis 2008, je suis membre de l'équipe "Signal et Communications" de LabInfo-IGM, UMR 8049. Auparavant, j'étais attaché thème "radiocommunications numériques" de l'équipe Esycom, EA 2552.

Les recherches que je développe comprennent deux axes :

1. le traitement et la théorie de l'information, et notamment l'étude de mesures d'information ou de complexité appliqué à l'analyse, le filtrage et la restauration de signaux et images, avec des applications aux télécommunications et à la physique statistique.
2. La radiotechnique (statistique) : analyse statistique et calcul de performances pour les émetteurs-récepteurs numériques, développement et caractérisation statistique d'algorithmes de correction, calibration ou contrôle.

Les principaux résultats obtenus sur ces deux thèmes sont décrits dans le chapitre 1 et dans la suite du mémoire.

Formation doctorale

- Intervention durant plusieurs années dans le DEA IFA (Informatique Fondamentale et Applications) pour un cours de traitement et compression d'images.
- J'ai participé récemment à l'aménagement et la proposition d'une nouvelle spécialité, "Image Signal et Son (SIS)" dans le cadre de la filière Image du Master d'informatique.

Encadrement

Master et Post-Doc

- PostDoc 6 mois en 2003 - Mlle Pavla Holejšovská - méthodes de reconstruction entropiques ; 3 communications
- 2007 : co-encadrement, avec A. Chevreuil, du stage de dernière année de l'ESIEE de Leire Azpilicueta et Pascal Vallet (ce dernier poursuit actuellement en thèse dans l'équipe sous la direction de P. Loubaton). Sujet : estimation d'un signal d'émission pour un procédé de diagnostic de la maladie de Menière. Financement par le contrat DREAMM.
- 2008 : co-encadrement, avec A. Djafari, du stage de Master de M. Boris Pougaza (au laboratoire des Signaux et Systèmes) - Using the notion of Copula in tomography.

Thèses

- Thèse Cifre de Cyril Joubert coencadrée avec G. Baudoin (directrice de thèse), 2004-2007, "contribution à l'analyse d'une boucle à verrouillage de phase (PLL) à traitement de signal numérique dans les émetteurs radio" 3 communications. Embauché par ST Microelectronics, Cyril se fait attendre pour soutenir effectivement sa thèse. Les résultats sont présentés en section 5.2.
- Thèse Cifre de Pierre Guillot, coencadrée avec C. Berland (directrice de thèse), 2006-2008, "Architecture des systèmes de radiocommunication à références d'horloges hautes fréquences : application des résonateurs BAW à la génération de fréquence de référence dans les systèmes de communication mobile", 2 communications, un article soumis. Cette thématique et les résultats sont présentés en 5.3.
- Thèse Cifre de Mlle Amandine Lesellier, débutée en octobre 2008, sur le sujet "contribution à l'étude des architectures de récepteurs large bande multicanaux". Pour la direction de cette thèse (coencadrée avec O. Venard), j'ai obtenu l'accord du conseil scientifique de l'université en septembre 2008.

Participation à d'autres encadrements

Je citerai également ici deux autres thèses pour lesquelles j'ai eu des contributions certaines à l'encadrement et aux résultats obtenus :

- Thèse d'Isabelle Jamin-Hibon "Analyse et conception d'une architecture numérique innovante d'émetteur universel pour radiocommunication numérique" (direction C. Berland et M. Villegas), soutenue en novembre 2006. Un article de revue et un brevet. Je partage avec Corinne Berland l'idée initiale qui sous-tend le travail de thèse, a été déposé comme brevet et a fait l'objet de l'article de revue (discuté au paragraphe 5.1 et annexé en page 164).

- Thèse d'Antoine Diet, "Etude des architectures EER pour les émetteurs de radiocommunications numériques 3G et au-delà" (direction C. Berland et M. Villegas), soutenue en mars 2005. J'ai ici contribué aux recherches sur l'évaluation, par méthode de Monte-Carlo, des performances obtenues en présence de défauts d'alignement. Ceci est présenté dans le paragraphe 6.1, a fait l'objet d'une communication et du chapitre 5 de sa thèse.

Relecture

J'ai effectué ou effectuée des relectures d'articles pour IEEE transactions on Signal Processing, Signal Processing (version Elsevier), Traitement du signal, Physics Letters A, Information Sciences, Electronics Letters

Activité d'enseignement

A l'ESIEE, j'interviens pour des cours, travaux dirigés et pratiques dans les 3 années du cycle ingénieur, dans le domaine du traitement du signal et des télécommunications. Je suis également intervenu dans l'école de techniciens supérieurs (à Bac +3), ESTE, appartenant au groupe ESIEE. Je suis également intervenu dans le programme N+I, et j'ai fait parti du comité de pilotage de l'ISBS, Institut Supérieur des Biosciences de Paris, formation créée par la CCIP et en cohabilitation entre l'université Paris12 et l'ESIEE.

Durant trois années, j'ai été responsable de la majeure télécommunications et traitement du signal, spécialisation couvrant les deux dernières années du cycle ingénieur. J'ai également été co-responsable du bloc électronique de tronc commun.

J'assume actuellement et ai assuré la responsabilité de quelques unités d'enseignement (programme, coordination, vacataires, etc).

- unité EL201 bases d'électronique et de traitement du signal (2^e année — depuis septembre 2000),
- unité EL301 filtrage numérique, signaux aléatoires (3^e année — co-responsable, responsabilité Gsignal 1997-2001),
- bloc traitement du signal et dsp, devenu Signaux et Systèmes 1, SIS1, (4^e année, majeure « systèmes embarqués »), depuis 2000 (au moins),
- bloc théorie du signal SIG1 (4^e année, majeure « télécommunications et traitement du signal »), depuis 2004,
- unité techniques avancées de communications numériques (5^e année) jusqu'à 2002,
- unité traitement satellite et radar (5^e année), entre 2002 et 2003,
- unité égalisation et synchronisation (5^e année), depuis 2005
- unité traitement du signal (1^{re} année ISBS) depuis 2004
- unité théorie de l'information et applications (2^e année ISBS) depuis 2004
- unité N+I EL01, bases du traitement du signal, entre 2003 et 2004

Les enseignements que j'ai effectué ou effectuée sont les suivants (pour 200 heures équivalent TD) : communications numériques, bases du traitement du signal, filtrage numérique, filtrage adaptatif, compression d'images, analyse spectrale, estimation et détection, traitement d'antenne, traitement satellite et radar, théorie de l'information, compression de données

Chapitre 1

Bilan succinct

Depuis 1996 je suis enseignant-chercheur à l'ESIEE, affecté au département Signaux et Télécommunications, (avec un passage de deux années dans le département Modélisation et Simulation Numérique).

Depuis début 2008, je suis membre de l'équipe "Signal et Communications" du laboratoire d'Informatique de l'Institut Gaspard Monge (LIGM, UMR 8049, CNRS). Auparavant, je faisais partie du thème "radiocommunications numériques" de l'équipe Esycom (EA 2552).

J'ai bénéficié en 2007-08 d'une "période de perfectionnement" d'une année, possibilité prévue au statut des personnels de la CCIP. Cette période de perfectionnement s'est déroulée en partie au laboratoire des signaux et systèmes LSS, UMR 8506, (situé à Gif-sur-Yvette dans les locaux de Supelec) dans l'équipe problèmes inverses ; et en partie dans l'équipe "signal et communications" du LIGM.

Comme je le disais en préambule, les recherches que je développe comprennent deux axes très distincts, qui sont menés parfois alternativement, parfois simultanément.

1. Le traitement et la théorie de l'information, et notamment l'étude de mesures d'information ou de complexité appliqué à l'analyse, le filtrage et la restauration de signaux et images, avec des applications aux télécommunications et à la physique statistique.
2. La radiotechnique (statistique) : analyse statistique et calcul de performances pour les émetteurs-récepteurs numériques, développement et caractérisation statistique d'algorithmes de correction, calibration ou contrôle.

Les paragraphes suivants donnent un bilan succinct de ces activités, en suivant l'organisation, somme toute logique, de la suite du mémoire, qui développe plus avant ces différents thèmes. Les références bibliographiques correspondent à la liste de publications présentée page 13.

— o 0 O 0 o —

1.1 Traitement et théorie de l'information

Entropie et maximum d'entropie

La question de la résolution de problèmes inverses à l'aide de techniques entropiques était l'objet de mon travail de thèse. Ce travail s'est prolongé ensuite, avec quelques développements (prise en compte de corrélation, opérateur non-linéaire). Un article de synthèse [R2] a été publié en 1999. J'ai ensuite proposé une méthode de reconstruction itérative prolongeant l'idée du maximum d'entropie sur la moyenne, et s'apparentant à un algorithme EM et accueilli une étudiante en post-doctorat sur ce sujet. Cette méthode a été présentée en [C18,C19,C21], et mériterait qu'on y revienne. La technique du maximum d'entropie sur la moyenne, et sa prolongation itérative, sont présentées dans le chapitre 2, sections 2.1 et 2.2.

Toujours concernant l'entropie de Shannon, et en vue de pouvoir traiter des données, nous nous sommes intéressés, avec C. Vignat au problème d'estimation de l'entropie. Nous avons envisagé une approche "signal", en passant par le biais d'une technique d'estimation spectrale par modèles autorégressifs régu-

larisés. Nous nous sommes attachés à construire un estimateur de l'entropie de Shannon, qui possède des caractéristiques récursives, ce qui permet de l'employer « en ligne ». Nous avons envisagé quelques applications (séparation de sources, déconvolution, détection de ruptures) et décrit l'apport d'un tel estimateur pour ces applications. [R3,C14]. Un autre estimateur, reposant sur une manipulation des histogrammes, a été présenté en [C15].

Comme il est parfois difficile de manipuler l'entropie de Shannon, nous nous sommes penchés sur l'entropie de Rényi comme alternative pour une utilisation pratique en traitement de signal. Nous avons pu donner une inégalité sur la transformation de l'entropie de Rényi par un système linéaire et proposer une application de cette inégalité sur un problème de déconvolution MIMO. Ceci est présenté dans la communication [C17] et décrit dans la section 2.5.

Une application des mesures d'information pour la caractérisation de séquences d'ADN, et plus exactement la localisation de séquences de promoteurs de transcription a été présentée dans [C24]. Nous y avons étudié notamment un classificateur de Bayes naïf qui permet de détecter les séquences promotrices. Ce classificateur peut être relié à des tests sur des divergences de Kullback. Une suite indirecte de ce travail a été ma contribution à [R7], où j'ai effectué différents calculs statistiques qui ont permis de corroborer le modèle présenté.

Entropies de Rényi-Tsallis

Je me suis ensuite intéressé de manière plus approfondie aux entropies de Rényi, Tsallis, à leurs interactions, aux justifications que l'on peut leur trouver et aux propriétés des solutions à maximum d'entropie de Rényi-Tsallis. Ces travaux sont présentés dans le chapitre 3. Resituons d'abord en quelques mots le contexte.

Pour deux distributions univariées de densités P et Q , la divergence de Rényi (1961) est

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \int_{\mathcal{D}} P(x)^\alpha Q(x)^{1-\alpha} dx, \quad (1.1)$$

où α est un réel positif.

Les applications et domaines d'intérêt pour l'entropie de Rényi sont nombreux : communication et codage, data mining, détection, segmentation, classification, test d'hypothèses, alignement d'images, etc. L'entropie de Rényi joue aussi un rôle central dans la théorie des multifractales. En physique statistique, suivant un article fondateur de Tsallis (1988), une communauté en “nonextensive statistics” s'est développée pour étudier une entropie comparable, et les distributions à Maximum Entropie associées, dites distributions de Tsallis

$$P_\nu(x) = \frac{[1 + \gamma(x - \bar{x})]^\nu}{Z_\nu(\gamma, \bar{x})} Q(x), \quad (1.2)$$

avec $Z_\nu(\gamma, \bar{x})$ la fonction de partition. Ces distributions, qui exhibent un comportement en loi puissance coïncident remarquablement avec certains jeux de données expérimentales, et semblent apparaître dans des situations à la limite de l'équilibre, ou avec un équilibre déplacé. Elles apparaissent également dans d'autres domaines : réseaux, fiabilité, climatologie, imagerie radar, économétrie...

Dans [C23], j'ai montré que les distributions de Tsallis peuvent être déduites à partir d'une approche classique de maximum d'entropie, en prenant en compte une contrainte figurant un système dont l'équilibre est déplacé. Une version dérivée de cette approche a également été présentée dans une revue de physique [R11]. Le résultat me semble avoir une certaine importance car il permet de relier les distributions de Tsallis au maximum d'entropie habituel, en introduisant une contrainte qui peut d'ailleurs avoir un sens en théorie du codage. Ceci est présenté et argumenté plus avant dans la section 3.2.

Il est également intéressant de noter que dans ce contexte, la divergence de Rényi apparaît en sous-produit et la nouvelle solution d'équilibre est solution du problème

$$\mathcal{F}_\alpha^{(C \text{ resp. } G)}(m) = \begin{cases} \min_P D_\alpha(P||Q) \\ \text{s.c. } (C) \ m = E_P[X] \\ \text{ou } (G) \ m = E_{P^*}[X] \\ \text{et } \int_{\mathcal{D}} P(x) dx = 1 \end{cases} \quad (1.3)$$

On a considéré une contrainte de moyenne classique (C) et une contrainte “généralisée” (G) telle que celle qui est employée par les physiciens de la communauté de physique statistique nonextensive. A l’optimum, la solution du problème est une fonctionnelle de m , $\mathcal{F}_\alpha^{(C)}(m)$ ou $\mathcal{F}_\alpha^{(G)}(m)$.

Il est alors légitime de s’intéresser aux propriétés et caractérisations de ces “fonctionnelles entropiques”. Ceci a fait l’objet du travail présenté en [R9]. J’y ai étudié et donné les formes et propriétés des solutions à maximum d’entropie, et des deux fonctionnelles associées, indexées par le paramètre α . La forme exacte (1.2) des distributions solution a été donnée et discutée. On a montré que les fonctionnelles $\mathcal{F}_\alpha^{(\cdot)}(m)$ sont simplement liées aux fonctions de partition. Les propriétés générales de ces fonctionnelles, comprenant la non-négativité, la convexité, ont été établies. On a aussi montré comment les solutions peuvent être obtenues numériquement, quand bien même la formulation fait intervenir des fonctions implicites. Ces fonctionnelles entropiques peuvent servir de fonction objectif à optimiser (où l’on peut utiliser le logiciel [V1]), ou de potentiels dans différentes applications. Une divergence dans l’espace objet, qui se réduit à une divergence de Bregman quand $\alpha \rightarrow 1$ a été introduite. Différents exemples ont été traités, et on a vérifié que lorsque $\alpha \rightarrow 1$, on retrouve des entropies classiques. Ce travail est présenté dans la section 3.4, et l’article est annexé page 105 et suivantes.

La distribution de Tsallis (1.2) peut être aussi obtenue comme un mélange de distributions en faisant varier, suivant une loi gamma, le paramètre naturel d’une loi exponentielle. A partir de cette remarque, une théorie des ‘Superstatistics’ a fait l’objet d’une série de publications depuis 2003. Pour expliquer comment le paramètre naturel d’une loi de Boltzmann-Gibbs pouvait fluctuer, j’ai proposé de décrire l’équilibre comme une solution à Maximum d’entropie avec contraintes aléatoires. Dans ce cadre, les fluctuations des contraintes entraînent d’autres fluctuations sur le paramètre de Lagrange associé, ce qui ramène aux Superstatistics. J’ai ensuite considéré l’exemple naturel dans lequel les fluctuations sur les contraintes se font suivant une loi Gamma. Dans ce cas, la distribution résultante est une K-distribution, dont il se trouve qu’elle a des propriétés intéressantes et de nombreuses applications en imagerie radar, ultrasons, scintillation optique, etc. Ce travail a été décrit dans [R10].

Une caractéristique importante de la distribution de Tsallis (1.2) est son comportement en loi puissance. Compte tenu de l’importance de ces lois, souvent utilisées pour modéliser les queues de distribution, nous nous sommes intéressé aux relations entre une distribution et sa queue. Plus exactement, nous avons recherché quelles seraient les queues à maximum d’entropie, sous une certaine contrainte de distance à la distribution initiale. Le lecteur ne sera pas surpris par le fait que ces queues à maximum d’entropie sont justement les distributions de Tsallis. Ceci est décrit dans [R11].

En réalité, ceci n’est pas complètement fortuit : avec Christophe Vignat, nous avons observé qu’un théorème sur la distribution des excès au dessus d’un seuil, le théorème de Pickands (1975), montre que dès que les distributions sont dans le domaine d’attraction des distributions extrêmes, alors la distribution des excès converge vers la distribution de Pareto Généralisée, qui n’est autre que la distribution de Tsallis. Dans [C30], nous montrons que la distribution des excès converge vers une distribution de Tsallis, et que dans le même temps, la solution d’un problème de maximum d’entropie généralisée converge vers cette même solution, ce qui justifie asymptotiquement de l’utilisation des entropies de Rényi-Tsallis.

Ceci fournit en outre une explication possible pour l’apparition des distributions de Tsallis dans un certain nombre de domaines [R12] : la distribution des excès converge vers une distribution de Tsallis et cette distribution est de plus stable par seuillage. Dès qu’un système comprend un seuillage, ou une série de seuillages d’une distribution sous-jacente, alors la loi limite converge vers une distribution de Tsallis, et cette loi d’équilibre pourra être décrite comme une distribution à maximum d’entropie de Rényi-Tsallis.

La poursuite de ces travaux comprend le prolongement de ces résultats au cas multivariable, avec en particulier l’étude des dépendances implicites qui apparaissent dans les lois à maximum de Rényi entropie. Ceci pourrait rejoindre les problématiques de “longues dépendances”. À partir de cette idée d’analyse de dépendances statistiques dans les lois, j’ai encadré (avec Ali Djafari) un étudiant de Master Recherche sur le thème “apport de la théorie des copulas pour la reconstruction en tomographie” (les copulas permettant de mesurer des dépendances statistiques).

L’existence d’une géométrie particulière sous-jacente à la famille (1.2), à l’image de la géométrie des familles exponentielles, est également un sujet d’intérêt. La divergence de Bregman déformée que l’on

a obtenue dans [R9] est un premier élément. La communication *invitée* [C29] présente une synthèse de quelques idées sur des raisons possibles de l'apparition des entropies de Rényi-Tsallis et des distributions associées. D'autres perspectives seront évoquées dans le corps du texte et dans le chapitre 7.

Information de Fisher

Plusieurs contributions concernent une autre mesure d'information, historiquement la première : l'information de Fisher. Ces contributions sont données dans le chapitre 4 et les articles en rapport [R4,R5,RS1] sont annexés. L'information de Fisher est d'une importance fondamentale en théorie de l'estimation. Elle sert également dans des problèmes d'inférence et d'interprétation de nombreux processus physique. Avec Christophe Vignat, nous avons considéré les relations entre l'information de Fisher et l'entropie de Shannon et défini le fameux "Fisher-Shannon information plane", qui a fait école. Ce plan d'information permet de caractériser les distributions sous-jacentes à des signaux, par exemple, à partir de l'analyse conjointe des informations de Fisher et de Shannon. Ceci est discuté dans l'article [R4].

Nous nous sommes également intéressés à la transformation de l'information de Fisher par des systèmes singuliers. Plus spécifiquement, nous avons étudié l'extension au cas multivarié, et aux transformations non inversibles, de l'inégalité sur l'information de Fisher. Nous avons alors pu retrouver, avec des démonstrations simplifiées, les principaux résultats de Zamir sur le sujet. Ces contributions ont été publiées dans [R5].

Plus récemment, nous nous sommes intéressés à la minimisation de l'information de Fisher sous contrainte de variance et de support. Ceci revient à rechercher et caractériser les distributions de support limité, dont l'information de Fisher (relative au paramètre de localisation), soit la plus faible, ce qui indique que le paramètre correspondant, par l'inégalité de Cramér-Rao, est le plus délicat à estimer. L'intérêt est bien sûr que dans nombre de cas pratiques, on sait que les distributions sont à support limité, par exemple à un intervalle, ou aux réels positifs. Dans l'article actuellement soumis [RS1], nous présentons les expressions explicites des solutions sur \mathbb{R}^+ et sur un intervalle, on décrit leur comportement, et celui de l'information de Fisher associée.

1.2 Radiotechnique

Je participe depuis quelques années à différents travaux autour des architectures radio. Mes intérêts et apports dans ces recherches sont en particulier les aspects modélisation statistique, et le développement de solutions issues des techniques de traitement du signal pour la correction et la synchronisation.

Différents problèmes d'architecture

Une des possibilités est d'aller vers des architectures entièrement numériques. Dans cette direction, j'ai coencadré, avec G. Baudoin (directrice de thèse), la thèse Cifre de C. Joubert en convention Cifre avec St Microelectronics, (RF Expertise Center - Cellular Terminal Division), sur l'analyse et l'optimisation d'architectures fondées sur une boucle à verrouillage de phase numérique. Nous avons en particulier proposé un modèle comportemental complet du système, qui permet de le simuler avec une charge en calcul très raisonnable. L'article de congrès [C22] présente ce modèle. Nous avons également développé des algorithmes rapides permettant d'alléger l'implantation [C26], et calculé explicitement la dsp du bruit de phase correspondant à la structure [C27]. Un prototype d'un émetteur FM employant cette PLL et implantant sur silicium les algorithmes développés a été réalisé et validé. Quelques éléments sur ces différentes contributions sont données dans la section 5.2 du chapitre 5.

Les contraintes posées pour la réalisation des émetteurs radio amènent à sélectionner des solutions de type polaire, où le signal est décomposé en un module (l'enveloppe) et une phase (le cosinus de la phase). Cette décomposition polaire, plutôt que la classique décomposition en phase et quadrature, permet d'utiliser des amplificateurs à haut rendement, puisque le signal présenté à l'amplificateur est de type enveloppe constante (l'enveloppe est restaurée en sortie en modulant la tension d'alimentation).

Avec Corinne Berland, nous avons proposé d'utiliser un codage sigma-delta qui permet d'obtenir des performances comparables à l'aide de dispositifs existants. Cette modification de l'architecture l'EER a abouti à une nouvelle architecture d'émetteur complètement numérique. (C. Berland a encadré une étudiante en convention Cifre avec STMicroelectronics Crolles pour la réalisation matérielle de cette architecture dans le cadre d'un financement CIFRE). Un brevet (en une version nationale 2005 et une internationale 2006) a été déposé [V1]. Dans le cadre de ce travail, le but étant de retrouver un signal modulé « normal », la restauration de la variation de l'enveloppe se fait en sortie de l'amplificateur de puissance par une simple opération de filtrage. L'avantage de cette architecture réside dans le fait que le signal à l'entrée de l'amplificateur est à enveloppe constante. L'amplification peut être réalisée par tout type d'amplificateur (A, B, AB, C, E) et permet d'optimiser les rendements. Ceci est décrit dans l'article [R6] qui est présenté en annexe page 164 et suivantes, ainsi que dans la section 5.1 du chapitre 5.

Du point de vue des émetteurs, il convient de contrôler les différentes dérives possibles (température, alimentation, vieillissement) et stabiliser le fonctionnement, au cours du cycle de vie du mobile, et ce, d'autant plus que les fréquences de fonctionnement sont élevées. Il convient également de concevoir des solutions qui prennent en compte la variabilité statistique des composants et réalisations, et comprennent une dimension fiabilité. Des solutions sont étudiées pour un oscillateur à très hautes fréquences réalisé chez Nxp par Pierre Guillot dans le cadre de sa thèse, à l'encadrement de laquelle je participe (Coencadrement C. Berland/J.-F. Bercher). Ces aspects sont présentés ici en 5.3.

Une première étude sur la réalisation d'une référence très haute fréquence – 2GHz, construite autour d'un BAW, est présentée dans [C31], où l'on décrit les caractéristiques du système, notamment le dimensionnement d'une banque de capacités qui servira à contrôler la fréquence de l'oscillateur. La réalisation correspondante est décrite dans [C33], ainsi que dans une publication en préparation où l'on présentera également des résultats de mesures. La caractérisation du système impose d'estimer ses différents paramètres, et de poursuivre leur évolution éventuelle. Ensuite, il s'agit, via un modèle d'évolution du système, de contrôler celui-ci. Pour ces étapes d'estimation et de contrôle, nous développons une solution fondée sur une modélisation d'état et un filtrage de Kalman.

Du point de vue des récepteurs, dans le front-end avant le filtre de Nyquist, c'est-à-dire en amont des algorithmes de traitement bien connus, différents problèmes d'optimisation peuvent être abordés. Il est possible technologiquement d'échantillonner très tôt le signal RF et de disposer de capacités de traitement numérique du signal. Les points qui sont d'intérêt sont : l'échantillonnage RF, avec ou sans parallélisme, l'optimisation de la quantification, la suppression d'interférences, etc. Une thèse Cifre avec Nxp a débuté en octobre dernier sur ce thème (thèse d'Amandine Lesellier - coencadrement J.-F. Bercher/O. Venard), avec une application aux récepteurs de télévision numérique multicanaux. Le problème est de pouvoir accéder simultanément à plusieurs canaux, 5 typiquement, situés quelque part dans une bande de 1 GHz. Le problème est donc de trouver une procédure d'échantillonnage très large bande, permettant de préserver plusieurs canaux de localisations aléatoires. Pour ce travail, on a commencé à étudier des solutions autour de l'utilisation de bancs de filtres hybrides analogique/numérique, avec une condition de reconstruction parfaite. On s'intéresse également aux possibilités d'utiliser des résultats sur l'échantillonnage non-uniforme, ou aléatoire, qui permet de dépasser la condition d'échantillonnage de Shannon (Whittaker).

Synchronisation des voies dans les architectures polaires

Dans l'architecture polaire que l'on a évoquée plus haut, le signal d'enveloppe et le signal de phase subissant des traitements différents, une désynchronisation entre l'enveloppe et la phase peut apparaître. Cette désynchronisation entraîne une remontée du plancher de bruit et une augmentation de l'EVM (Error Vector Magnitude). Ce problème devient d'autant plus crucial que la période symbole diminue, avec l'augmentation des débits. Par ailleurs, les caractéristiques analogiques sont affectées par des dérives dans le temps, par une variation avec la température et par un effet de vieillissement. Enfin, les solutions sont de plus en plus multimodes et les caractéristiques peuvent varier lorsque l'on passe d'un mode à un autre.

Avec Corinne Berland, nous avons travaillé sur les possibilités de compenser ces retards entre voies. Ceci est l'objet du chapitre 6, à partir de la page 53.

Nous avons proposé un premier algorithme de correction dans [C25]. En nous appuyant sur les statistiques de signaux, nous avons montré d'une part que la synchronisation ne peut pas être effectuée en aveugle et nécessite une boucle de retour. Nous avons également discuté d'une procédure d'identification dans [C28]. Le principe de correction est simplement d'ajuster des paramètres afin de minimiser l'erreur quadratique d'entrée sortie, à l'aide d'un algorithme stochastique. La procédure emploie une étape d'interpolation numérique qui a été discutée. Nous avons caractérisé analytiquement le comportement de l'algorithme (vitesse de convergence, biais, variance) à l'aide d'un 'toy model', et ces résultats confrontés à des simulations numériques. L'ensemble a été validé sur un modèle de modulation OFDM [R8].

Nous avons ensuite développé une version sous optimale, qui permet d'alléger notablement l'implantation numérique comme la réalisation en hardware, en évitant de devoir effectuer une démodulation en quadrature qui peut être remplacée par une conversion directe. Le domaine de convergence de ce nouvel algorithme a été caractérisé. Nous avons également vérifié analytiquement que ces solutions sont robustes à une mauvaise connaissance du gain ou à un offset de phase ; mais bien entendu au prix d'une augmentation de la variance d'estimation. Nous avons alors développé une extension de la procédure permettant de prendre en compte et compenser ces méconnaissances. Enfin, nous avons étudié le comportement de cette solution en contexte non stationnaire. Un article reprenant ces différents résultats est actuellement soumis [RS2]. Les deux publications [R8], et [RS2] sont annexées.

Nous avons d'ores et déjà prolongé ces résultats, analytiquement et en simulation, à l'architecture LINC, potentiellement très performante, mais qui confrontée aux mêmes types de désynchronisation, peut-être de manière plus aigüe encore. Une autre perspective consistera à également prendre en charge des onduations résiduelles du gain dans la bande passante.

La conception et l'optimisation des émetteurs numériques semble ainsi pouvoir bénéficier de l'apport d'outils de traitement du signal pour l'optimisation de la "dirty RF" : c'est le point de vue présenté dans [C32].

1.3 Publications

Articles de revue avec comité de lecture

Articles soumis :

[RS2] J.-F. Bercher and C. Berland, “**Adaptive delays alignment in polar transmitter architecture**,” *IEEE trans. on Vehicular Technologies*, submitted, April 2009.

[RS1] J.-F. Bercher and C. Vignat, “**A note on minimum Fisher information distributions with restricted support and fixed variance**,” *Information Sciences*, submitted, Mar. 2009.

Articles publiés :

[R12] J.-F. Bercher and C. Vignat, “**A new look at q -exponential distributions via excess statistics**”, *Physica A*, vol. 387, n°22, pp 5422-5432, september 2008.

[R11] J.-F. Bercher, “**Tsallis distribution as a standard maximum entropy solution with ‘tail’ constraint**”, *Physics Letters A*, vol. 372, n°35, pp 657-665, august 2008.

[R10] J.-F. Bercher, “**Maximum entropy with fluctuating constraints : The example of K-distributions**”, *Physics Letters A*, vol. 372, n°24, pp 4361-4363, june 2008.

[R9] J.-F. Bercher, “**On some entropy functionals derived from Rényi information divergence**”, *Information Sciences*, vol. 178, pp. 2489-2506, june 2008.

[R8] J.-F. Bercher and C. Berland, “**Envelope and phase delays correction in an EER radio architecture**”, *Analog Integrated Circuits and Signal Processing*, vol. 55, pp. 21-35, Apr. 2008.

[R7] B. Duriez, P. Duquesnoy, E. Escudier, A.-M. Bridoux, D. Escalier, I. Rayet, E. Marcos, A.-M. Vojtek, J.-F. Bercher and S. Amselem, “**A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia**”, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, n°9, pp. 3336-3341, february 2007 ; and “**Corrections**”, vol. 104, n°15, p. 6490, april 2007.

[R6] C. Berland, I. Hibon, J.-F. Bercher, M. Villegas, D. Belot, D. Pache, V. Le Goasoz “**New transmitter architecture for nonconstant envelope modulation**”, *IEEE Trans. on Circuits and Systems*, vol. 53, Issue 1, pp 13 - 17, january 2006.

[R5] C. Vignat, J.-F. Bercher, “**On Fisher information inequalities and score functions in non-invertible linear systems**”, *JIPAM. Journal of Inequalities in Pure and Applied Mathematics*, vol. 4, no. 4, Article 71, 9 pp., november 2003.

[R4] C. Vignat, J.-F. Bercher, “**Analysis of signals in the Fisher–Shannon information plane**”, *Physics Letters A*, vol. 312, pp. 27–33, june 2003.

[R3] J.-F. Bercher, C. Vignat, “**Estimating the entropy of a signal with applications**”, *IEEE Trans. on Signal Processing*, vol. 48, n° 6, june 2000.

[R2] G. Le Besnerais, J.-F. Bercher, G. Demoment, “**A new look at the entropy for solving linear inverse problems**”, *IEEE Trans. on Information Theory*, vol. 45, n°5, pp.1565-1578, july 1999.

[R1] J.-F. Bercher et P. Larzabal, « **Traitement d’antenne avec adéquation de modèle**, » *Traitement du Signal*, numéro spécial *Calibration d’antenne*, volume 10, n° 5-6, pp. 349–359, 1995.

Colloques avec comité de lecture et actes

[CS1] C. Vignat and J.-F. Bercher, “Geometric aspects of the non-extensive statistical theory”, submitted to MaxEnt 2009, July 2009, Oxford, Mississippi.

[C33] P. Guillot, P. Philippe, C. Berland, and J.-F. Bercher, “**Faisabilité de référence haute fréquence pour les architectures RF**,” in *Journées Nationales Microondes 2009*, may 2009.

- [C32] C. Berland, J.-F. Bercher, O. Venard, “**Digital Signal Processing techniques to compensate for RF imperfections in Advanced Transmitter Architectures**”, European Wireless Technology Conference (EuWit2008), Amsterdam, october 2008.
- [C31] P. Guillot, C. Berland, J.-F. Bercher, P. Philippe, “**A 2GHz 65nm CMOS digitally-tuned BAW oscillator**”, 15th IEEE International Conference on Electronics (ICECS2008), Circuits and Systems, Malta, september 2008.
- [C30] J.-F. Bercher and C. Vignat, “**An entropic view of Pickands’ theorem**”, IEEE International Symposium on Information Theory (ISIT2008), Toronto, july 2008
- [C29] J.-F. Bercher, “**Possible rationales for Rényi-Tsallis entropy maximization**”, International Workshop on Applied Probabilities (IWAP2008), Compiègne, july 2008.
- [C28] C. Berland and J.-F. Bercher, “**Adaptive time mismatches identification and correction in polar transmitter architecture**”, European MicroWave 2007, Munich, 8-12 october 2007.
- [C27] G. Baudoin, J.-F. Bercher and C. Joubert, « **Calcul théorique du bruit de phase dû au dithering dans un oscillateur contrôlé numériquement** », Journées Nationales Microondes, Toulouse, 23-25 Mai 2007
- [C26] C. Joubert, J.-F. Bercher, G. Baudoin, T. Divel, S. Ramet, P. Level, “**Contributions to the analysis and design of an ADPLL**”, Proceedings of the 13th IEEE International Conference on Electronics, Circuits, and Systems, ICECS2006, pp. 322-325, december 2006.
- [C25] J.-F. Bercher and C. Berland, “**Envelope/phase delays correction in an EER radio architecture**”, Proceedings of the 13th IEEE International Conference on Electronics, Circuits, and Systems, ICECS2006, pp. 443-446, december 2006.
- [C24] J.-F. Bercher, P. Jardin et B. Duriez, “**Bayesian classification and entropy for promoter prediction in human DNA sequences**”, AIP Conference Proceedings, vol. 872, pp. 235-242, Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 2006
- [C23] J.-F. Bercher, “**An amended MaxEnt formulation for deriving Tsallis factors, and associated issues**”, AIP Conference Proceedings, vol. 872, pp. 441-448, Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, 2006, [ArXiv.org](https://arxiv.org/abs/2006.04411)
- [C22] C. Joubert, J.-F. Bercher, G. Baudoin, T. Divel, S. Ramet, P. Level, “**Time Behavioral model for Phase Domain ADPLL based frequency synthesizer**” IEEE Radio and Wireless Symposium (RWS 2006), january 2006.
- [C21] P. Holejšovská, Z. Peroutka, J. Cengery, J.-F. Bercher, “Continuous Non-Invasive Blood Pressure Measuring Based on the Reconstruction by Maximum Entropy on the Mean Method” In Jirí Jan, Jirí-Kozumplik, and Ivo Provazník, editors, BIOSIGNAL 2004 : 17-th Biennial International Eurasip Conference Proceedings, pages 121-123, Brno, Czech Republic, June 2004.
- [C20] J.-F. Bercher, A. Diet, C. Berland, G. Baudoin, M. Villegas, “**Monte carlo Estimation of Time mismatch Effect in an OFDM EER Architecture**”, IEEE Radio & Wireless Conference RAWCON 2004, sept 2004.
- [C19] P. Holejšovská, Z. Peroutka, J.-F. Bercher, “Adaptive signal processing using maximum entropy on the mean method and Monte Carlo analysis” In : MATLAB 2003, pp 191-196, ISBN 80-7080-526-9.
- [C18] P. Holejšovská, J.-F. Bercher, “**Elimination of the Measurement System Impact on the Measured Data by Maximum Entropy on the Mean Method**”, Diagnostika ‘03, Plzen, pp. 192-195, september 2003, ISSN 80-7082-952-4.
- [C17] J.-F. Bercher, C. Vignat, “**Rényi entropy convolution inequality with Applications**”, Eusipco 2002, Toulouse, France, September 2002.

- [C16] C. Vignat, J.-F. Bercher, “**Matrix Fisher Information Inequalities for Non-Invertible Linear Systems**”, IEEE International Symposium on Information Theory, Lausanne, Switzerland, June 30-July 5, 2002.
- [C15] C. Vignat, J.-F. Bercher, “**Un estimateur récursif de l’entropie**”, *actes du 17^e Colloque GRETSI sur le Traitement du Signal et ses Applications*, Vannes, septembre 1999.
- [C14] J.-F. Bercher, C. Vignat, “**Estimating the entropy of a signal with applications**”, *Proceedings. ICASSP99 (Cat. No.99CH36258)*, 1999, Page 1705-1708 Volume 3.
- [C13] C. Leal, C. Meihlac, A. Pesme, J.-F. Bercher, C. Vignat, “**Recovering binary data transmitted over unknown communication channels**”, in *Proc. of DSPCONF Texas-Instrument*, Noisy-le-Grand, sept. 1998, pp. 119-124.
- [C12] D. Janu, G. Baudoin, J.-F. Bercher, O. Venard, “**Design of a CDMA simulator and implementation on a TMS320601**”, in *Proc. of DSPCONF Texas-Instrument*, Noisy-le-Grand, Sept. 1998, pp. 101-106.
- [C11] C. Heinrich, J.-F. Bercher et G. Demoment, « **The Maximum Entropy on the Mean Method, Correlations, and Implementation Issues**, » 16th Conference on Maximum Entropy and Bayesian Methods, South Africa, août 1996, 10 pages.
- [C10] J.-F. Bercher et C. Heinrich, « **Construction de mesures de divergence. Application à la résolution de problèmes inverses linéaires**, » 15^e Colloque GRETSI sur le Traitement du Signal et ses Applications, Juan-les-Pins, septembre 1995, 4 pages, et une **version longue**.
- [C9] C. Heinrich, J.-F. Bercher, G. Le Besnerais et G. Demoment, « **Méthode du maximum d’entropie sur la moyenne et mélanges de distributions**, » 15^e Colloque GRETSI sur le Traitement du Signal et ses Applications, Juan-les-Pins, septembre 1995, 4 pages.
- [C8] C. Heinrich, J.-F. Bercher, G. Le Besnerais et G. Demoment, “**Restoration of spiky signals : a new optimal estimate and a comparison**,” IEEE International Conference on Acoustic Speech and Signal Processing, mai 1995, 4 pages.
- [C7] J.-F. Bercher, G. Le Besnerais et G. Demoment, « **Building Convex Criteria for solving linear inverse problems**,” *Inverse Problems and Applications*, HoChiMinh City, Vietnam, 17-19 janvier 1995, 12 pages.
- [C6] J.-F. Bercher, G. Le Besnerais et G. Demoment, « **The Maximum Entropy on the Mean Method, Noise, and Sensitivity**,” 14th Conference on Maximum Entropy and Bayesian Methods, Cambridge, 1-5 août 1994. Publié dans *Maximum entropy and Bayesian Methods*, Kluwer Academic Publishers, The Netherlands, 1996, pp. 223-232.
- [C5] G. Le Besnerais, J.-F. Bercher et G. Demoment, « Probabilistic Issues in Fourier Synthesis,” *First School on Scientific Computation in Systems*, Jouy-en-Josas, 6-8 décembre 1993, 24 pages.
- [C4] J.-F. Bercher, « **Un algorithme de factorisation bispectrale**, » 14^e Colloque GRETSI sur le Traitement du signal et ses Applications, Juan-les-Pins, septembre 1993, pp 117–120.
- [C3] B. Lumeau, J.-C. Pesquet, J.-F. Bercher, L. Louveau, “Optimization of Bias-variance Trade-Off in Non parametric Spectral Analysis by decomposition into Wavelets Packets,” *International Conference Wavelets and Applications*, Toulouse, 8-12 juin 1992, 6 pages. Publié dans “Progress in Wavelet Analysis and Applications”, Y. Meyer and S. Roques ed., Éditions Frontières, 1993.
- [C2] B. Lumeau et J.-F. Bercher, « **Extension de la méthode de Pisarenko et ses dérivées au cas de signaux vectoriels et applications à la localisation de sources**, » 13^e Colloque GRETSI sur le Traitement du Signal et ses Applications, Juan-les-Pins, 16-20 septembre 1991, pp 361–364.
- [C1] B. Lumeau et J.-F. Bercher, « **Critère optimisant un compromis biais-variance pour estimer l’ordre d’un processus autorégressif mono ou multivariable**, » 13^e Colloque GRETSI sur le Traitement du Signal et ses Applications, Juan-les-Pins, 16-20 septembre 1991, pp 321–324.

Autres publications

- [1] P. Holejovská, J.-F. Bercher. “Adaptive signal processing : maximum entropy on the mean method”
In : Elektrotechnika a informatika 2003, pp. 35-38, ISBN 80-7082-992-3

Valorisation

- [V2] C. Berland, J. F. Bercher, I. Hibon, M. Villegas, D. Belot, D. Pache, V. Le Goasoz, “**Digital Transmitter Architecture**”, Brevet FR0405636, publié le 25/05/2004 ; Brevet international US2006034391A1, publié le 16/02/2006.
- [V1] J.-F. Bercher, S. Brette, H. Carfantan, J.-F. Giovannelli, C. Heinrich, J. Idier, T. Martin, C. Soussen, « **Gradient à pas Adaptatif avec Corrections** », logiciel déposé par le CNRS auprès de l’Agence de Protection des Programmes sous le N°00258-01, juin 2005.

Thèse

- [T1] J.-F. Bercher, « **Développement de critères de nature entropique pour la résolution de problèmes inverses linéaires**, » Thèse de doctorat, Université Paris-Sud, 1995.

Livres

Participation à l’ouvrage collectif d’enseignement (à paraître) :

- [L2] « *Antennes* », sous la direction d’O. Picon, Dunod 2009. Première édition, 2009.

Auteur du chapitre **traitement d’antenne**.

Participation à l’ouvrage collectif d’enseignement :

- [L1] « *Systèmes de communications numériques – Modélisation et simulation* », sous la direction de G. Baudoin, Dunod 2002. Seconde édition, 2007.

Auteur ou coauteur des chapitres :

- **Récepteur optimal**,
- **Égalisation**,
- **Synchronisation** (avec J.-M. Brossier).

Colloques sans actes, journées scientifiques

- J.-F. Bercher et C. Vignat, « **Levy distributions and the maximization of Rényi-Tsallis entropy** », in “Power laws in probabilities and statistics”, CIRM, march 22-26, 2004.
- J.-F. Bercher et C. Vignat, « Estimation de l’entropie d un signal et applications, » GdR Information, signal et images (GT 1 : Méthodes Entropiques en Signal et Image), 10 décembre 1998 ; et sur le même thème (GT2 Séparation de sources) 29 avril 1999.
- J.-F. Bercher, « Critères de type entropique, » Journée thématique du GRCE-AFCET (Groupe de Reconnaissance en Communication Ecrite), 06 février 1996.
- J.-F. Bercher, « Critères entropiques et problèmes inverses, » Journée scientifique « Information & signal » du GdR Information, signal et images, 10 novembre 1995.
- J.-F. Bercher, « Critères entropiques et optimisation convexe pour la résolution de problèmes inverses linéaires, » journée scientifique « optimisation », École Nationale Supérieure des Télécommunications de Bretagne, 18 mai 1995.

- J.-F. Bercher, C. Heinrich, G. Le Besnerais et G. Demoment, « Utilisation de la méthode du maximum d'entropie sur la moyenne en traitement du signal et de l'image, » journée maximum d'entropie sur la moyenne et grandes déviations, Université d'Évry, 1^{er} avril 1995.
- J.-F. Bercher et C. Heinrich, « Entropies et reconstruction d'images, » 3^e Colloque jeunes chercheurs en physique, Campus d'Orsay, 27-28 mars 1995.

Chapitre 2

Entropie et maximum d'entropie

L'entropie et le principe du maximum d'entropie sont parmi mes sujets de réflexion majeurs depuis mon travail de thèse. Le rôle central de l'entropie de Shannon en théorie de la communication, le rôle de l'entropie de Boltzmann-Gibbs en physique statistique et en thermodynamique également. Dans deux articles fondamentaux [3, 4], Jaynes discute de l'interaction entre la théorie de l'information et la mécanique statistique. En particulier, il introduit le principe du maximum d'entropie : *Information theory provides a constructive criterion for setting up probability distributions on the basis of partial knowledge, and leads to a type of statistical inference which is called the maximum entropy estimate. It is the least biased estimate possible on the given information ; i.e., it is maximally noncommittal with regard to missing information.* Dans ce chapitre, on discute de méthodes d'inférence s'appuyant sur le maximum d'entropie, de caractérisations du principe lui-même, puis finalement de l'intervention de l'entropie comme mesure de la complexité d'une distribution permettant le traitement de données. Le premier paragraphe, 2.1 reprend le travail [5] sur le maximum d'entropie sur la moyenne débuté lors de la préparation de ma thèse de doctorat. Le second paragraphe 2.2 présente une technique de maximum d'entropie sur la moyenne itératif [6, 7]. Dans le paragraphe 2.3 on rapporte quelques idées sur le cas du maximum d'entropie avec des contraintes aléatoires, et l'apparition de K -distributions [8]. Dans le paragraphe suivant, section 2.4, on présente un estimateur de l'entropie de Shannon et des exemples d'application [9]. Enfin, dans le dernier paragraphe, on présente une inégalité sur l'entropie de Rényi que l'on utilise dans le cadre de problèmes de déconvolution [10].

2.1 Maximum d'entropie sur la moyenne

Le travail sur le maximum d'entropie sur la moyenne correspond au départ à mon travail de doctorat, soutenu en 1995. Ce travail s'est prolongé au delà, avec en particulier la parution d'un article [5] en 1999, qui se trouve un peu à la frontière d'une recherche du travail de thèse et le début d'une recherche autonome. Par rapport au travail de thèse, les mathématiques ont été clarifiées, et de nouveaux exemples et développements insérés (le cas d'un opérateur non linéaire, la prise en compte d'un bruit poissonnien, l'introduction de corrélations). D'autres idées et travaux ultérieurs, exposés dans la suite du document, se rattachent également à cette thématique. Je reprends donc rapidement ici le principe de la méthode et les principaux résultats obtenus, sans m'appesantir outre mesure sur les subtilités. Les lecteurs intéressés pourront consulter par exemple [5], ou l'excellent article de H. Gzyl [11]. On considère un problème inverse linéaire, comme il en apparaît dans de nombreuses situations expérimentales, qui se pose donc sous la forme simple

$$\mathbf{y} = \mathbf{Ax} + \mathbf{b}. \quad (2.1)$$

On cherche souvent à trouver une solution acceptable, par exemple en incluant des *desiderata*, ou des contraintes sur \mathbf{x} (ou sur une représentation de \mathbf{x}) en minimisant un critère composite de la forme

$$\mathcal{F}(\mathbf{x}) + \mathcal{G}(\mathbf{y} - \mathbf{Ax}). \quad (2.2)$$

Ce type de critère peut bien entendu provenir ou être interprété comme un maximum a posteriori. Les propriétés des fonctionnelles \mathcal{F} et \mathcal{G} ont bien sûr de l'importance, du point de vue pratique ou pragmatique : unimodalité, convexité, support, dualité, etc. La technique du maximum d'entropie sur la moyenne permet de construire des critères convexes $\mathcal{F}(\mathbf{x})$ incorporant des caractéristiques a priori, notamment des contraintes de support et possédant une formulation duale attrayante. Considérons pour le moment le problème sans bruit, c'est-à-dire simplement le problème $\mathbf{y} = \mathbf{A}\mathbf{x}$, typiquement sous déterminé. L'idée est, comme en physique, de considérer l'objet recherché \mathbf{x} comme la moyenne sous une loi. Dans la mesure où plusieurs distributions peuvent être compatibles avec les observations, on choisit de sélectionner la distribution à maximum d'entropie, dont la moyenne vérifie $\mathbf{y} = \mathbf{A}\mathbf{E}[\mathbf{X}]$. En réalité, on peut savoir que l'objet recherché possède des propriétés de concentration, ou de support (\mathbb{R}^+ ou un intervalle par exemple). Plutôt que de maximiser l'entropie standard, on maximisera une μ -entropie, où μ est une mesure sur le domaine concerné, ou, de manière équivalente, on cherche à minimiser la distance de Kullback-Leibler à une loi de référence, disons μ , que l'on note $D(P||\mu)$. Notons que cette démarche trouve une argumentation dans le théorème de Sanov. Si P_N est la distribution empirique correspondant à un tirage de N variables aléatoires selon une loi μ , alors la probabilité de trouver P_N dans un ensemble \mathcal{B} est, grosso-modo,

$$Q(P_N \in \mathcal{B}) \approx \exp\left(-N \inf_{P \in \mathcal{B}} D(P||\mu)\right). \quad (2.3)$$

On en déduit donc, en itérant le raisonnement sur des sous ensembles de \mathcal{B} , que la distribution absolument prépondérante dans \mathcal{B} est celle qui réalise le minimum de la distance de Kullback à μ : on a concentration de toute la probabilité sur la distribution la plus proche de μ . En réalité ce qui nous intéresse c'est plutôt la probabilité de trouver une moyenne empirique x_N , c'est-à-dire la moyenne sous P_N , dans un ensemble \mathcal{C} . Dans ce cas, on a un résultat de grandes déviations de niveau 1, qui indique que

$$Q(x_N \in \mathcal{C}) \approx \exp\left(-N \inf_{x \in \mathcal{C}} \mathcal{F}(\mathbf{x})\right), \quad (2.4)$$

où $\mathcal{F}(\mathbf{y})$ est la fonction de taux. Ce résultat suggère maintenant de sélectionner l'élément le plus probable, celui qui réalise le minimum de $\mathcal{F}(\mathbf{x})$ sur \mathcal{C} . Le passage d'une problématique sur les distributions à une problématique sur les moyennes est connu comme le principe de contraction. Si on revient au problème initial, on voit que la minimisation de l'information de Kullback-Leibler sous la contrainte $\mathbf{y} = \mathbf{A}\mathbf{E}[\mathbf{X}]$ peut se traiter comme une minimisation en deux temps : d'abord une minimisation sous $\mathbf{x} = \mathbf{E}[\mathbf{X}]$, puis une minimisation sous la contrainte $\mathbf{y} = \mathbf{A}\mathbf{x}$:

$$\inf_{P: \mathbf{y}=\mathbf{A}\mathbf{E}[\mathbf{X}]} D(P||\mu) = \inf_{\mathbf{x}: \mathbf{y}=\mathbf{A}\mathbf{x}} \left[\inf_{P: \mathbf{x}=\mathbf{E}[\mathbf{X}]} D(P||\mu) \right]. \quad (2.5)$$

On voit que la minimisation de droite sur P définit en fait une fonction de \mathbf{x} :

$$\mathcal{F}(\mathbf{x}) = \inf_{P: \mathbf{x}=\mathbf{E}[\mathbf{X}]} D(P||\mu) \quad (2.6)$$

Il n'est pas difficile de voir que la fonction $\mathcal{F}(\mathbf{x})$ ainsi définie est strictement convexe (minimisation d'une fonction strictement convexe sous une contrainte linéaire). Le problème initial, sur l'espace des distributions de probabilité, devient ainsi équivalent à une minimisation sur l'espace objet

$$\inf_{P: \mathbf{y}=\mathbf{A}\mathbf{E}[\mathbf{X}]} D(P||\mu) = \inf_{\mathbf{x}: \mathbf{y}=\mathbf{A}\mathbf{x}} \mathcal{F}(\mathbf{x}) \quad (2.7)$$

Considérons encore un instant (2.6). On sait que la solution de la minimisation est dans la famille exponentielle engendrée par μ : $dP(\mathbf{x}) = \exp(\lambda^T \mathbf{x} - \log(Z(\lambda^T))) d\mu(\mathbf{x})$, où $Z(\lambda^T)$ est la fonction de partition qui sert à normaliser $dP(\mathbf{x})$. Dans ce cas, et si tout va bien, on a égalité avec le problème dual associé, et

$$\mathcal{F}(\mathbf{x}) = \sup_{\lambda} \{\lambda^T \mathbf{x} - \log(Z(\lambda^T))\} = \sup_{\lambda} \{\lambda^T \mathbf{x} - \mathcal{F}^*(\lambda^T)\}. \quad (2.8)$$

La seconde égalité ci-dessus introduit la fonction $\mathcal{F}^*(\lambda^T)$. Sous cette forme, la fonction $\mathcal{F}(\mathbf{x})$ apparaît comme la conjuguée convexe, ou la transformée de Legendre-Fenchel de $\mathcal{F}^*(\lambda^T)$, et \mathbf{x} et λ sont les variables conjuguées. Dans le cas présent, on identifie $\mathcal{F}^*(\lambda^T)$ au logarithme de la transformée de Laplace de la mesure μ : la transformée de Legendre-Fenchel correspondante est la transformée de Cramér de μ . On sait en outre que les variables duales sont reliées par les relations de « dualité de Legendre » : $\mathbf{x} = \mathcal{F}^{*'}(\lambda^T)$ et $\lambda_* = \mathcal{F}'(\mathbf{x})$. D'un autre côté, on retrouve aussi le fait que la dérivée de la log-fonction de partition, par rapport au paramètre naturel λ , n'est autre que la moyenne sous la loi.

Revenons maintenant à la résolution du problème et à l'équation (2.7). Les deux problèmes d'optimisation, dans l'espace des mesures de probabilité et dans l'espace objet, possèdent la même formulation duale :

$$\sup_{\lambda} \{ \lambda^T \mathbf{y} - \mathcal{F}^*(\lambda^T \mathbf{A}) \}, \quad (2.9)$$

où \mathcal{F}^* est comme précédemment la log-fonction de partition de la famille exponentielle de paramètre naturel $\mathbf{A}^T \lambda$. En résolvant (2.9), on obtient ainsi à la fois la distribution à maximum d'entropie et la moyenne correspondante, qui est aussi la solution du problème de minimisation de $\mathcal{F}(\mathbf{x})$ sous contrainte, et qui rappelons le, n'est autre que la dérivée de \mathcal{F}^* à l'optimum. Il est utile de remarquer que \mathcal{F}^* est également une fonction convexe. On note que finalement il n'est pas nécessaire de travailler dans l'espace direct, et que la résolution se fait dans le domaine dual, généralement sans contrainte.

Associé à la mesure de référence μ , on peut ainsi trouver un critère entropique $\mathcal{F}(\mathbf{x})$, strictement convexe, qui encode certaines caractéristiques de μ . Notamment, lorsqu'on utilise une mesure à support borné, loi uniforme ou loi de Bernoulli, ou une loi sur \mathbb{R}_+^N , le critère n'est défini que sur l'enveloppe convexe de la mesure, et la solution, forcément, appartient à ce domaine. Un certain nombre d'exemples de critères sont donnés et interprétés ainsi dans la publication [5], en remarquant cependant qu'il n'est pas nécessaire d'avoir une expression explicite pour \mathcal{F} puisque la résolution se fait dans le domaine dual.

Disons encore un mot sur les techniques de prise en compte du bruit d'observation : lorsqu'on relaxe la contrainte égalité $\mathbf{y} = \mathbf{A}\mathbf{x}$ à une contrainte du type $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 < c$, on peut résoudre le problème par $\inf_{\mathbf{x}} \mathcal{F}(\mathbf{x}) + \alpha \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2$, où α est un paramètre de Lagrange. On peut également utiliser une approche duale, dans laquelle le terme quadratique précédent se traduit simplement par l'ajout d'un terme en $\alpha \|\lambda\|^2$. Une autre approche consiste à définir un objet étendu $\mathbf{w} = [\mathbf{x}; \mathbf{b}]^T$, qui vérifie alors $\mathbf{y} = \mathbf{H}\mathbf{w}$ avec $\mathbf{H} = [\mathbf{A}; \mathbf{1}]$. On peut alors introduire une mesure de référence $\mu_w = \mu_x \mu_b$ sur l'objet global et utiliser la démarche du maximum d'entropie sur la moyenne. Celle-ci conduit alors à un critère composite dans l'espace objet

$$\mathcal{F}(\mathbf{x}) + \mathcal{G}(\mathbf{y} - \mathbf{A}\mathbf{x}), \quad (2.10)$$

où \mathcal{F} et \mathcal{G} sont respectivement les critères entropiques associés aux mesures μ_x et μ_b .

2.2 Maximum d'entropie sur la moyenne itératif

Il n'existe pas de connexion naturelle, sinon asymptotique [12], entre la construction du maximum d'entropie sur la moyenne et une approche bayésienne. Les techniques à maximum d'entropie sont parfois, souvent même, insérées dans une démarche bayésienne en prenant comme a priori une forme $\exp \mathcal{F}(\mathbf{x})$, et la recherche d'un MAP conduit alors à la maximisation d'un critère régularisé classique. On présente ici une alternative aux techniques de prise en compte du bruit exposées dans le paragraphe précédent, avec une approche qui fait coopérer l'approche bayésienne et celle du maximum d'entropie sur la moyenne.

L'idée de base est la suivante. On a vu que l'on sait bien résoudre le problème sans bruit $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x}$, pour lequel il est respectable de postuler que l'on recherche une solution comme la moyenne d'une distribution à maximum d'entropie. En réalité, l'observation est bruitée, et l'on a $\mathbf{y} = \bar{\mathbf{y}} + \mathbf{b}$.

Supposons que l'on dispose d'une estimée, disons $\bar{\mathbf{y}}_k$ de la moyenne des observations. Dans ces conditions, on peut résoudre suivant

$$\bar{\mathbf{x}}_k = \arg\min_{\bar{\mathbf{y}}_k = \mathbf{A}\mathbf{x}} \mathcal{F}(\mathbf{x}), \quad (2.11)$$

ou de manière équivalente, le problème sur le paramètre naturel de la loi à maximum d'entropie :

$$\bar{\lambda}_k = \operatorname{argmax}_{\lambda} \{ \lambda^T \bar{\mathbf{y}}_k - \mathcal{F}^*(\lambda^T \mathbf{A}) \}. \quad (2.12)$$

La description de notre objet \mathbf{x} , $\bar{\mathbf{y}}_k$ étant connue, est ainsi donnée par la loi à maximum d'entropie, une loi exponentielle par rapport à la densité de référence $\mu(x)$, de paramètre $\bar{\lambda}_k$. Il est alors parfaitement légitime d'utiliser cette densité, que l'on pourra noter $p(\mathbf{x}|\bar{\mathbf{y}}_k)$,

$$p(\mathbf{x}|\bar{\mathbf{y}}_k) = \exp(\bar{\lambda}_k^T \mathbf{A} \mathbf{x} - \log Z(\bar{\lambda}_k^T \mathbf{A})) \mu(\mathbf{x}) \quad (2.13)$$

comme densité a priori et d'appliquer la mécanique bayésienne pour aboutir à la loi a posteriori selon

$$p(\mathbf{x}|\mathbf{y}, \bar{\mathbf{y}}_k) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\bar{\mathbf{y}}_k). \quad (2.14)$$

A partir de cette loi a posteriori, qui a donc intégré les véritables observations \mathbf{y} , on peut calculer la moyenne a posteriori $\check{\mathbf{x}}_{k+1} = \mathbb{E}[\mathbf{X}|\mathbf{y}, \bar{\mathbf{y}}_k]$, et définir une nouvelle estimée de la moyenne des observations par

$$\bar{\mathbf{y}}_{k+1} = \mathbf{A} \check{\mathbf{x}}_{k+1} = \mathbf{A} \mathbb{E}[\mathbf{X}|\mathbf{y}, \bar{\mathbf{y}}_k]. \quad (2.15)$$

Il ne reste plus alors qu'à itérer le processus, tant que, par exemple, $\|\bar{\lambda}_k - \bar{\lambda}_{k-1}\|^2 > \epsilon$. On pourra alors retenir comme estimée, selon les cas, soit la moyenne a posteriori, soit la moyenne de la loi à maximum d'entropie. Comme on dispose de la loi a posteriori, il est également possible de calculer une covariance a posteriori, ou au moins des variances a posteriori.

En fait, cette procédure est analogue à un algorithme EM pour une famille exponentielle engendrée par la densité de référence initiale $\mu(\mathbf{x})$. C'est le choix de la loi a priori à maximum d'entropie (2.13) qui confère à la loi a posteriori complète cette nature de famille exponentielle (l'ajout de la vraisemblance ne modifie rien sur cet aspect). Le paramètre naturel à estimer est alors $\Psi = \mathbf{A} \lambda$, ou plus simplement λ , et dans notre cas la statistique suffisante est simplement $t(\mathbf{x}) = \mathbf{x}$. Dans ces conditions les deux étapes de l'EM sont

$$(E) : \check{\mathbf{x}}_{k+1} = \mathbb{E}[\mathbf{X}|\mathbf{y}, \bar{\mathbf{y}}_k] \quad (2.16)$$

$$(M) : \bar{\lambda}_{k+1} = \operatorname{argmax}_{\lambda} \{ \lambda^T \mathbf{A} \check{\mathbf{x}}_{k+1} - \mathcal{F}^*(\lambda^T \mathbf{A}) \}. \quad (2.17)$$

Ces deux étapes correspondent bien à la démarche que nous avons adoptée, équations (2.12) et (2.15). Par cette identification avec l'algorithme EM, on peut prévoir la convergence de la procédure. Notons cependant que l'étape délicate va être ici l'étape E, avec le calcul de la moyenne a posteriori pour lequel il faut, hors cas spécifique, échantillonner suivant la loi a posteriori. L'étape M correspond quant à elle au problème de maximum d'entropie sur la moyenne, pour lequel la solution est obtenue par un simple algorithme de gradient, d'autant plus rapide que l'on peut utiliser la solution à l'itération précédente comme condition initiale.

L'originalité de la démarche est ainsi d'introduire un a priori à maximum d'entropie sur la moyenne, dans une approche bayésienne, « réconciliant » ainsi les deux approches. La procédure résultante s'interprète ensuite simplement comme un algorithme EM dont les propriétés sont bien documentées.

Sur ce sujet, j'ai encadré sur quelques mois un travail de post-doctorat d'une étudiante tchèque, Pavla Holejšovská, qui, en dépit d'un investissement modeste, a obtenu des résultats numériques encourageants qui ont été présentés lors de conférences [6, 13, 7]. Il fait partie de mes projets de reprendre ce travail.

2.3 Maximum d'entropie avec contraintes aléatoires

Quittons maintenant le champ du maximum d'entropie sur la moyenne pour revenir à la problématique standard du maximum d'entropie : trouver la distribution la plus générale, la moins compromettante, compatible avec les contraintes. Le problème qui m'intéressait dans ce cadre était d'examiner ce que l'on peut dire lorsque les contraintes, justement, ne sont pas bien définies. Cette réflexion étant alimentée,

bien entendu, par l'expérience selon laquelle les données sont parfois bruitées. Un autre guide, il faut le reconnaître, était la connaissance de différents articles de physique définissant des “superstatistiques” [14, 15, 16, 17]. Dans ces travaux, les auteurs considèrent la loi d'équilibre d'un système, typiquement la distribution de Boltzmann-Gibbs – une distribution exponentielle, et imaginent que le paramètre naturel β est susceptible de varier selon une certaine distribution de probabilité. Dans ces conditions, on est face à une loi jointe, qu'il convient d'intégrer par rapport à la loi sur β pour obtenir la nouvelle description du système. Il est connu que l'on peut retrouver ainsi la distribution de Tsallis en choisissant une loi Gamma sur β . Ce fait, qui a suscité une abondante littérature en statistiques non-extensive, avait été noté depuis bien longtemps [18].

Dans le petit article [8], annexé au document en page 70, je considère des fluctuations non sur le paramètre de la loi de Boltzmann, mais sur les observables. Celles-ci servent classiquement de contraintes dans le cadre d'une approche à maximum d'entropie. Si \bar{E} est la valeur de la contrainte (énergie moyenne), alors l'entropie correspondante s'écrit $S(\bar{E}) = \max_{\beta} \{-\beta \bar{E} - \log Z(\beta)\}$, et le paramètre β et la valeur moyenne sont deux variables duales reliées par une transformée de Legendre. Ces deux variables ne sont donc pas indépendantes. Aussi, lorsque les observables sont affectées par des fluctuations suivant une distribution $f_{\bar{E}}(\bar{E})$, le paramètre dual β est lui même affecté par des fluctuations suivant une loi $f_{\beta}(\beta)$ image de la précédente par la transformation. Ceci montre que l'on peut donc interpréter les fluctuations sur le paramètre naturel comme résultant d'une incertitude sur les contraintes.

A titre d'exemple, on considère le modèle canonique, avec une contrainte sur la valeur de l'énergie moyenne. Dans ce cadre, il est intéressant de considérer une loi Gamma $\Gamma(a, b)$ comme modèle de fluctuations de cette énergie moyenne. Après intégration, il en résulte alors la distribution suivante

$$P(E) = \frac{2}{E} \frac{1}{\Gamma(a) \Gamma(3/2)} (bE)^{\frac{3/2+a}{2}} K_{-\frac{3}{2}+a} \left(2\sqrt{bE} \right). \quad (2.18)$$

où K_{ν} est une fonction de Bessel modifiée de seconde espèce et d'ordre ν . On souligne que cette distribution, dite K -distribution a trouvé des applications importantes dans le domaine de l'imagerie radar. On discute ensuite de sa caractérisation, de ses principales propriétés et de son comportement asymptotique (dont son caractère “heavy-tailed”). Les lecteurs pourront retrouver plus de détails dans l'article annexé.

2.4 Estimation de l'entropie

Pour pouvoir utiliser l'entropie directement dans des applications de traitement du signal ou de traitement de données, il est nécessaire de pouvoir évaluer l'entropie attachée aux données considérées. Des estimateurs d'entropie sont présentés dans la synthèse [19] : on y trouve des estimateurs de type “plug-in” à partir d'estimées de la densité (typiquement à noyaux), des estimateurs reposant sur les statistiques d'ordre (dont l'estimateur de Vasicek), et un estimateur fondé sur les plus proches voisins (estimateur de Leonenko, 1987). Dans un travail mené avec Christophe Vignat, nous avons envisagé une approche “signal” et nous nous sommes attachés à construire un estimateur de l'entropie de Shannon, qui possède des caractéristiques récursives, ce qui permet de l'employer « en ligne ». En outre, nous avons envisagé quelques applications et décrit l'apport d'un tel estimateur pour ces applications. Nous avons proposé deux estimateurs de l'entropie. Le premier repose sur une modélisation paramétrique des densités de probabilité et a fait l'objet d'une communication [20] et d'un article de revue [9]. Cet article est disponible en annexe page 76. Le second estimateur est un estimateur classique, reposant sur une technique de type histogramme, pour lequel nous avons trouvé une formulation récursive sur le temps, et a été exposé dans une communication [21].

L'originalité du premier estimateur est de proposer d'assimiler la densité de probabilité à un spectre de puissance (normalisé), et d'utiliser un modèle autorégressif régularisé, cher à Jean-François Giovannelli, pour estimer cette densité de probabilité. Les exemples donnés dans l'article, que l'on retrouvera ici page 79 indiquent la pertinence et l'applicabilité du modèle. En réalité l'idée d'utiliser des méthodes d'estimation spectrale dans le cadre d'estimation de densités de probabilité avait été exposée un peu auparavant par Pagès & Lagunas d'une part et S. Kay d'autre part (mais les résultats numériques sont moins inté-

ressants qu'avec des AR longs). Au passage, nous avons introduit un processus sous-jacent, défini par $W(n) = \exp(jX(n) + \phi)$ dont la densité spectrale de puissance $S_W(x)$ est précisément la densité de probabilité $f_X(x)$ de X (il suffit d'observer que sa fonction d'autocorrélation est la fonction caractéristique). A partir d'un modèle de la densité de probabilité sous la forme d'une fraction rationnelle, on est alors capable d'établir une expression explicite de l'entropie, en fonction des pôles de la fonction de transfert et des résidus associés ; ce qui n'est tout de même pas très pratique. En poursuivant avec une approche signal, on utilise simplement la relation de Plancherel-Parseval qui indique ici que

$$\hat{H}(X) = - \int_{-\frac{1}{2}}^{+\frac{1}{2}} S_W(x) \log_2 S_W(x) dx = - \sum_{k=-\infty}^{+\infty} R_W(k) C_W^*(k) \quad (2.19)$$

où R_W et C_W sont respectivement la fonction d'autocorrélation et cepstre de $W(n)$. Grace à la structure AR du signal, les séquences de corrélation et le cepstre peuvent être exprimées récursivement à partir des coefficients AR ou de la réponse impulsionnelle. Ainsi, il suffit d'estimer p coefficients de corrélation pour en déduire les coefficients AR, donc la réponse impulsionnelle, puis étendre les séquences de corrélation et de cepstre autant que souhaité. La formule (2.19) permet ensuite d'éviter le calcul explicite de la densité et l'intégration, en résumant le calcul de l'estimateur de l'entropie à un simple produit scalaire. De plus, on peut calculer récursivement la matrice de corrélation, en introduisant un facteur d'oubli, ce qui permet d'utiliser l'estimateur dans un cadre non stationnaire. On peut également alléger la charge calculatoire en utilisant un algorithme du gradient pour déterminer les paramètres AR plutôt que d'inverser une matrice de grande taille. Comparé aux estimées obtenues avec un estimateur par noyaux de la densité de probabilité, à l'estimateur de Vasciceck modifié, on obtient que notre estimateur se comporte agréablement en termes de biais et de variance. La charge en calcul est également réduite. On a ensuite envisagé quelques applications pour lesquelles on peut utiliser directement un estimateur d'entropie : détection de ruptures dans un signal, déconvolution aveugle et séparation de sources. Dans les trois cas, l'utilisation de cet estimateur d'entropie se révèle pertinente. Signalons ici que nous retrouverons plus loin, en §4.3 page 39 le problème de détection ou de modification de la loi instantanée d'un signal, que nous avons également étudié en surveillant conjointement l'entropie et l'information de Fisher. On notera aussi que nous avons abordé le problème de déconvolution aveugle en nous appuyant sur le fait, conséquence de l'inégalité sur la puissance entropique (Entropy Power Inequality - EPI), que l'entropie du signal de sortie d'un système linéaire, alimenté par une séquence i.i.d. non gaussienne est toujours supérieure à l'entropie de l'entrée, si la norme du filtre est unitaire. Ce résultat se comprend intuitivement : sous une contrainte de puissance constante, le mélange des composantes d'un signal iid le rapproche de la gaussienne. Pour déconvoluer, on peut alors penser à ajuster les paramètres d'un filtre inverse de sorte à minimiser l'entropie de sortie. Le minimum absolu sera atteint lorsque l'on aura retrouvé la séquence i.i.d. de départ.

Dans la même lignée, nous avons étudié un second estimateur, reposant sur les simples histogrammes des données. Dans ce cadre, et dans la communication associée [21], nous avons proposé une interprétation des histogrammes comme l'estimation de la densité spectrale de puissance du « processus sous-jacent » à l'aide d'un banc de filtres ; donné une formulation récursive sur le temps d'un estimateur de l'entropie de Shannon fondé sur une analyse par histogramme, et enfin décrit une application pour la détection de changement de loi, et une application en séparation de sources.

2.5 Inégalité de convolution pour l'entropie de Rényi

Une difficulté posée par l'entropie de Shannon est que des résultats analytiques sont difficiles à obtenir dans le cas où l'on analyse des signaux subissant des transformations linéaires, comme par exemple dans la procédure de déconvolution évoquée dans le paragraphe précédent. Avec Christophe Vignat, nous avons cherché à exploiter la famille des entropies de Rényi, pour des applications de déconvolution.

Dans la communication [10], nous avons donné une inégalité sur l'entropie de Rényi qui étend (faiblement) l'inégalité de convolution pour l'entropie de Shannon (l'EPI). Cette inégalité repose sur une

version étendue de l'inégalité de Young et sur l'aide de Franck Barthe. Nous avons montré que si \mathbf{X} et \mathbf{Y} sont deux vecteurs aléatoires indépendants, alors

$$H_r(\mathbf{X} + \mathbf{Y}) \geq \max(H_r(\mathbf{X}), H_r(\mathbf{Y})), \quad (2.20)$$

avec égalité si et seulement si \mathbf{X} ou \mathbf{Y} est un vecteur déterministe. A partir de cette inégalité, en considérant Z la sortie d'un système alimenté par une séquence de vecteurs iid X_i , alors

$$H_r(Z) \geq H_r(X) + \frac{1}{2} \log |f_i|^2, \quad \forall i \quad (2.21)$$

avec égalité si et seulement si le filtre est un retard pur. Cette inégalité est à rapprocher de celle que l'on obtient pour l'entropie de Shannon :

$$H_1(Z) \geq H_1(X) + \frac{1}{2} \log \sum_i |f_i|^2, \quad (2.22)$$

avec égalité si et seulement si $X(n)$ est gaussien ou le filtre est un retard pur. En comparant les deux relations, on voit que l'inégalité pour l'entropie de Rényi reste plus faible, car elle fait apparaître la norme infinie du filtre et un seul de ses coefficients. Cependant le résultat n'est pas inintéressant puisqu'il permet d'obtenir des résultats explicites. Il est amusant de remarquer que le même résultat a été obtenu pratiquement simultanément (et en parfaite indépendance) par Hild, Erdogmus et Principe [22]. Il a été repris comme un résultat original dans plusieurs publications parmi lesquelles [23, 24, 25, 26, 27]. La dernière référence [27] met en garde sur un mauvais usage de l'entropie de Rényi en séparation de sources mais valide notre procédure de déconvolution.

En utilisant l'inégalité de convolution pour l'entropie de Rényi, on peut utiliser la procédure suivante : rechercher les coefficients d'un égaliseur tel que l'entropie en sortie du système global soit minimale. Il faut ajouter à cela une contrainte fixant un des coefficients, disons f_k à un : $|f_k| = 1$. Cette contrainte évite que la minimisation ne conduise au filtre et à une sortie nulle pour laquelle on a effectivement égalité dans (2.21). Dans la mesure où les f_k dépendent à la fois de l'égaliseur et du filtre inconnu, la contrainte précédente ne peut pas être assurée. Par contre, on peut fixer l'un des coefficients à une constante. Une solution est de fixer le premier coefficient de l'égaliseur, h_0 , ce qui fixe le premier coefficient $f_0 = h_0 g_0$ de la réponse impulsionnelle globale. Dans le cas d'une entropie quadratique, et en utilisant un estimateur par noyaux, nous avons pu exprimer analytiquement l'entropie de Rényi, son gradient (et son hessien) par rapport aux coefficients matriciels d'un égaliseur (et les calculs correspondants furent formateurs). Des résultats de simulation nous ont montré que ce critère permettait, en dépit certainement de minima locaux, d'obtenir une déconvolution parfaite en aveugle, [10]. Terminons ce chapitre par une petite illustration : on considère un système MIMO avec $p = 2$ entrées et $q = 3$ sorties, avec la fonction de transfert suivante :

$$g(z) = \begin{bmatrix} 1 & -0.7 \\ 0.82 & 1 \\ 0.7 & 0.3 \end{bmatrix} + \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & -0.3 \\ 0.1 & -0.2 \end{bmatrix} z^{-1} + \begin{bmatrix} 0.3 & -0.5 \\ 0.6 & -0.5 \\ 1 & 0.8 \end{bmatrix} z^{-2}.$$

Avec un égaliseur de longueur $L = 12$, la procédure précédente converge vers une solution intéressante. La réponse impulsionnelle globale du système, qui est très proche de l'identité, est donnée figure 2.1.

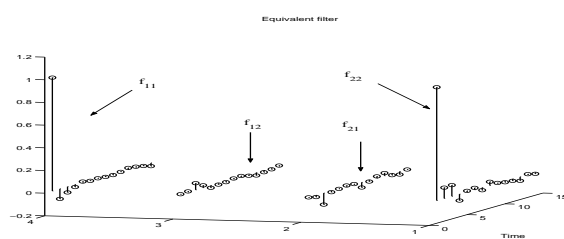


FIG. 2.1 – Equivalent impulse responses f for experiment 3.

Chapitre 3

Entropies de Rényi-Tsallis

Le dernier paragraphe du chapitre précédent, sur une inégalité de convolution pour l'entropie de Rényi, nous permet une transition douce. Nous nous étions intéressés à l'entropie de Rényi pour examiner si, du point de vue pratique, on pouvait obtenir des résultats plus aisément manipulables qu'avec l'entropie de Shannon. Les entropies alternatives, telle que celle de Rényi, peuvent également avoir des fondements axiomatiques qui en justifient l'utilisation. Parente de l'entropie de Rényi, l'entropie de Tsallis a suscité ces 20 dernières années un nombre important de recherches en physique statistique. L'une des raisons, au moins d'un point de vue pragmatique, est qu'elle permet de faire apparaître des lois puissance, et que des lois de ce type semblent très importantes en pratique, voir par exemple la synthèse [28]. La thermodynamique associée, qui recouvre la thermodynamique habituelle comme cas particulier semble également en mesure de mieux expliquer certains phénomènes. En m'investissant sur ces questions, j'ai cherché à comprendre les raisons qui peuvent justifier ces entropies alternatives, quelles sont leurs propriétés remarquables, et comment on peut les utiliser. Il est connu par exemple que les distributions de Tsallis peuvent résulter d'une fluctuation selon un loi gamma du paramètre naturel d'une loi exponentielle. Nous avons discuté ce point en section 2.3 et proposé de relier ces fluctuations à des fluctuations de la contrainte.

En observant que les distributions de Tsallis apparaissent dans des contextes avec un équilibre déplacé ou perturbé, j'ai proposé un modèle qui permet d'interpréter les distributions de Tsallis comme résultant d'une procédure standard de maximum d'entropie (de Shannon), où l'on a ajouté une contrainte pour modéliser un déplacement de l'équilibre. Cette procédure, présentée ici en 3.2, a fait l'objet des publications [29, 30].

J'ai ensuite réalisé que les distributions de Tsallis sont présentes dans nombre de domaines, et coïncident avec la distribution de Pareto généralisée. Or cette distribution est la distribution limite qui décrit les queues de distribution en théorie des excès. Il s'agit probablement là d'une remarque importante qui peut expliquer une certaine ubiquité des distributions de Tsallis. Ce résultat a été présenté, avec Christophe Vignat, dans la référence [31]. Au passage, nous avons montré comment on pouvait connecter la distribution des excès, asymptotiquement, avec la solution d'un problème de maximum d'entropie [32]. Ces deux points sont présentés dans la section 3.3.

Une procédure de maximum d'entropie (de Rényi-Tsallis), ou de minimum de la divergence associée, telle qu'elle apparaît notamment comme résultante des modèles précédents, définit implicitement une entropie contractée, fonction de la contrainte et non plus de la distribution. Il est alors légitime de s'intéresser aux propriétés et caractérisations de ces "fonctionnelles entropiques". Ceci a fait l'objet du travail présenté dans la section 3.4 et dans la publication [33] placée ici en annexe.

3.1 Contexte

Dans son article de 1961 [34] Rényi présente un ensemble de postulats qui permet d'obtenir l'entropie de Shannon comme unique mesure d'information. L'un des postulats décrit que l'entropie de l'union

de deux distributions incomplètes est la moyenne arithmétique des entropies pondérée par les poids des distributions. En remplaçant dans ce postulat la moyenne arithmétique par une moyenne géométrique, Rényi obtient alors l'entropie de Rényi attachée à une distribution \mathcal{P} :

$$H_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \log \left(\frac{\sum_{k=1}^n p_k^\alpha}{\sum_{k=1}^n p_k} \right). \quad (3.1)$$

Avec la même démarche, il donne également l'expression de l'information d'ordre α , $D_\alpha(\mathcal{P}||\mathcal{Q})$, dont nous reparlerons abondamment plus loin, qui est l'information qui apparaît lorsqu'une distribution \mathcal{P} est remplacée par une distribution \mathcal{Q} .

Un autre résultat, apparemment bien peu connu et qu'il faudrait enseigner, qui justifie pleinement l'entropie de Rényi est un théorème de codage de source dû à Campbell (1965) [35]. Dans cette contribution, Campbell propose d'introduire une mesure de longueur des mots codes pénalisant les codes de longueur élevée. En effet, dans la définition classique, le coût associé à un mot code croît linéairement avec sa longueur. En notant D la taille de l'alphabet, n_i les longueurs élémentaires associées aux mots codes de probabilités p_i , Campbell définit alors

$$L(t) = \frac{1}{t} \log_D \sum_i p_i D^{-tn_i} \quad (3.2)$$

qui pénalise la longueur de chacun des mots codes de manière exponentielle (il s'agit d'une transformation monotone simple de la longueur plus intuitive $\sum_i p_i D^{-tn_i}$). Dans ces conditions, L.L. Campbell montre que

$$L(t) \geq H_\alpha, \quad (3.3)$$

où H_α est l'entropie de Rényi définie en (3.1), avec $\alpha = 1/(1+t)$. L'égalité est atteinte lorsque les longueurs des mots codes sont

$$n_i = \log_D \left(\frac{p_i^\alpha}{\sum_i p_i^\alpha} \right), \quad (3.4)$$

où le terme de droite fait apparaître une “escort distribution” que nous retrouverons plus bas. Ce théorème étend ainsi de manière naturelle le théorème standard du codage de source, à nouveau par la substitution d'une moyenne arithmétique par une moyenne géométrique, et légitime l'entropie de Rényi. Resterait bien entendu à développer des codeurs à longueur (au sens de Campbell) variable.¹

L'entropie de Tsallis [36] a été introduite en 1988 dans le champ de la physique statistique, à l'origine pour la description des multifractales. Elle est définie par

$$S_\alpha(f_X) = \frac{1}{1-\alpha} \left(\int f_X(x)^\alpha dx - 1 \right) \quad (3.5)$$

où α est un réel positif et f_X une distribution univariée. Comme dans le cas de l'entropie de Rényi, l'entropie de Shannon est obtenue, par la règle de L'Hospital, dans le cas limite $\alpha = 1$.

Notons au passage, car cela semble largement ignoré, que l'entropie de Tsallis est suggérée dans l'article initial de Rényi [34, eq. 4.20, p 561] :

It is clear from the above proof that instead of the quantities (4.4) we could have used the analogous sums

$$(4.20) \quad \sum p_{jk}^{(n)} f \left(\frac{p_{jk}^{(n)}}{p_k} \right)$$

where $f(x)$ is any function such that $xf(x)$ is strictly convex. Thus for instance we could have taken $f(x) = x^{\alpha-1}$ with $\alpha > 1$ or $f(x) = -x^{\alpha-1}$ with $0 < \alpha < 1$.

¹Campbell a introduit une pénalisation exponentielle de la longueur des mots codes. Une autre manière d'introduire une fonction de coût, ou longueur, pénalisant les mots-codes de grande taille associée à des probabilités faibles serait de déformer la distribution de probabilité pour tenir davantage compte des indices de probabilités (initialement) faible. Dans cet esprit il serait intéressant de rechercher un résultat de type codage de source pour une moyenne généralisée définie comme $L(\alpha) = \sum_i n_i P_i$, avec P_i une “escort distribution” $P_i = p_i^\alpha / \sum_k p_k^\alpha$.

Les distributions de Tsallis peuvent être obtenues par maximisation de l'entropie de Tsallis (3.5), sous contrainte de moyenne et de normalisation. Bien entendu, comme l'entropie de Rényi et de Tsallis sont liées par une transformation monotone, la maximisation de l'une fournit le même argument que la maximisation de l'autre. Deux types de contraintes ont été considérées dans la littérature : la moyenne statistique standard $E[X] = m$, et une moyenne généralisée $E_\alpha[X] = m$, qui est prise par rapport à une 'escort' ou 'zooming distribution' $p(x)^\alpha / \int p(x)^\alpha dx$. Ces deux contraintes conduisent à un résultat de la forme

$$f_X(x) = \frac{1}{\sigma} \left(1 + \frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}-1} \text{ for } x_F > x \geq 0, \quad (3.6)$$

où σ et γ sont respectivement les paramètres d'échelle et de forme. Pour $\gamma < 0$, la densité a un support fini, avec $x_F = -\sigma/\gamma$ si $\gamma < 0$ un support infini dans le cas contraire. Dans le premier cas, l'exposant est $1/(\alpha - 1)$ et $\sigma = 1/\beta$; dans le second cas, l'exposant est $1/(1 - \alpha)$ et $\sigma = 1/\beta(2 - \alpha)$. Pour $\gamma = 0$ ($\alpha = 1$), la solution se réduit à la distribution exponentielle standard $f_X(x) = 1/\sigma \exp(-x/\sigma)$ for $\gamma = 0$. Un des gros intérêts des distributions de Tsallis est de pouvoir rendre compte d'un comportement en loi puissance, qui est souvent rencontré dans des applications.

Depuis l'introduction de l'entropie de Tsallis en physique statistique, un travail considérable, et un nombre non moins considérable de publications (plus de 1000 références dans la bibliographie tenue par Tsallis), se sont penchés sur l'étude des entropies alternatives et des thermodynamiques associées. C. Tsallis a publié très récemment un livre de synthèse [37]. En effet, il apparaît que dans le cas de certains systèmes complexes, la thermodynamique standard semble insuffisante pour une description complète. Les applications incluent les problèmes de turbulence, les statistiques des rayons cosmiques, l'économetrie, etc. Dans nombre d'exemples, les résultats d'expérimentation et les résultats numériques coïncident très bien avec une description par une distribution de Tsallis.

3.2 L'entropie de Rényi-Tsallis comme solution d'un problème standard de maximum d'entropie

Le principe du maximum d'entropie peut être vu, interprété, comme une conséquence du théorème de Sanov, comme nous l'avons déjà indiqué dans le chapitre 2. L'essence de celui-ci est que dans l'ensemble de toutes les distributions issues d'une distribution q , compatibles avec une contrainte, celle qui devient absolument prépondérante est la distribution la plus proche de q au sens de l'information de Kullback-Leibler. Dans ces conditions, et dans la mesure où les distributions de Tsallis sont différentes des distributions obtenues par une approche de maximum entropie classique, celles-ci seront avec les mots de Grendar [38] *will be asymptotically improbable*. Compte-tenu du succès rencontré dans les applications, il est légitime de se demander si on peut trouver des justifications probabilistes pour l'entropie de Tsallis ou pour sa maximisation qui conduit aux distributions de Tsallis. Un essai de synthèse a été présenté lors du workshop Iwap 2008 [39]. Une première approche est proposée ici. L'idée sous jacente a été présentée au colloque MaxEnt2006 [29], puis après une métamorphose dans la publication [30], qui est annexée ici page 88.

Le point de départ pour rechercher une justification de l'apparition des distributions de Tsallis est de noter qu'elles semblent apparaître dans le cas d'un équilibre de Boltzmann-Gibbs modifié, perturbé, ou déplacé. On peut essayer de rendre compte de cela en introduisant dans la formulation du maximum d'entropie une contrainte supplémentaire dont l'effet est de déplacer l'équilibre. Cet équilibre déplacé peut être imaginé comme résultant de l'équilibre entre deux distributions, R et Q . Au lieu de sélectionner la distribution la plus proche de la référence Q sous une contrainte de moyenne, on recherche une distribution P^* proche, simultanément, des deux références. Une telle distribution sera localisée quelque part entre R and Q . Par exemple, on considèrera un système global composé de deux sous-systèmes caractérisés par deux distributions distinctes. La distribution d'équilibre sera atteinte pour un intermédiaire entre les deux distributions de référence, et l'observable pourra être, selon les cas, soit la moyenne sur un sous système, soit la moyenne sous la distribution d'équilibre. Un tel modèle peut rendre compte d'un processus de fragmentation : un système $\Sigma(A, B)$ se fragmente en A , avec une distribution R et

B avec une distribution Q , et le système complet est vu avec la distribution d'équilibre P^* . On peut aussi voir une transition de phase : un système quitte l'état Q vers un état R , et présente une distribution intermédiaire P^* .

Ceci peut-être posé de la manière suivante : trouver une distribution P^* telle que la divergence de Kullback-Leibler divergence à Q , $D(P||Q)$ soit minimum, sous la contrainte d'équilibre entre R et Q , par exemple $D(P||Q) - D(P||R) \leq \theta$, ou de manière équivalente, une contrainte $D(P||R) \leq \theta'$. On peut remarquer que la première contrainte correspond à un log-rapport de vraisemblance moyen. Ceci s'écrit

$$\begin{cases} \min_P D(P||Q) = \min_P \int P(x) \log \frac{P(x)}{Q(x)} dx \\ s.t \ \theta = D(P||Q) - D(P||R) = \int P(x) \log \frac{R(x)}{Q(x)} dx \leq \theta. \end{cases} \quad (3.7)$$

En fait, ce petit problème a été considéré, et résolu, par Kullback [40, page 39] comme une illustration de son théorème général sur la minimisation de $D(P||Q)$ sous contrainte, et l'on obtient :

$$P^*(x) = \frac{R(x)^\alpha Q(x)^{1-\alpha}}{\int R(x)^\alpha Q(x)^{1-\alpha} dx}, \quad (3.8)$$

expression qui n'est rien d'autre que l'"escort distribution" de la physique nonextensive (avec en plus ici l'apparition du terme en Q). Le paramètre α est simplement le paramètre de Lagrange associé à la contrainte. La distribution P^* est simplement la moyenne géométrique (pondérée) entre R et Q et réalise un compromis, ajusté par α , entre les deux références. Par dualité de Lagrange, on a ici l'égalité entre le problème initial et son dual :

$$\begin{cases} \min_P D(P||Q) \\ s.t \ \theta = D(P||Q) - D(P||P_1) \end{cases} = \sup_\alpha \left(\alpha\theta - \log \left(\int P_1(x)^\alpha Q(x)^{1-\alpha} dx \right) \right). \quad (3.9)$$

On peut alors noter que dans cette dernière relation, le terme $\log \left(\int P_1(x)^\alpha Q(x)^{1-\alpha} dx \right)$, qui est strictement concave en α , est directement proportionnel à la divergence de Rényi. En d'autres termes, la solution du problème initial, $D(P^*||Q)$ s'exprime simplement à partir de la divergence de Rényi d'ordre α :

$$D(P^*||Q) = \alpha\theta - (\alpha - 1)D_\alpha(R||Q). \quad (3.10)$$

Il n'est pas difficile de vérifier que la solution satisfait une relation de Pythagore :

$$D(P||Q) = D(P||P^*) + D(P^*||Q) \quad (3.11)$$

pour toute distribution P telle que $D(P||Q) - D(P||R) = \int P \log \frac{R}{Q} = \theta$. La solution optimale P^* s'interprète alors comme la projection orthogonale sur la ligne joignant R et Q , ce qui minimise à la fois la distance à R et à Q .

On peut utiliser différentes variantes de l'argumentation, en considérant par exemple un problème de maximisation de l'entropie, avec une contrainte sur la distance à une loi de référence Q . C'est l'approche adoptée dans [30], où on imagine que Q représente la queue de la distribution. On cherche alors une distribution à maximum d'entropie dont la « queue » est à une distance (de Kullback) θ de la distribution de départ.

Remarques

Il est fort intéressant également de noter que la contrainte sur le log rapport de vraisemblance n'est pas totalement déraisonnable. Dans son mémoire de doctorat, Marcel-Paul Schützberger [41] introduit en pages 57 à 61 de son mémoire, une information qu'il appelle information de Wald [42] qui inclut notamment le log-rapport de vraisemblance moyen entre deux hypothèses. On la rencontre en analyse séquentielle. Ainsi, notre problème initial s'interprète comme la recherche de la distribution la plus proche d'une distribution de référence, au sens de la divergence de Kullback, sous contrainte que l'information de Wald-Schützenberger soit bornée (ou fixée). Notons également que dans ce mémoire tout-à-fait précurseur, Marcel-Paul Schützberger propose une démarche unificatrice des « théories des informations

». Entre autres, il définit, page 65, la fonction génératrice des cumulants c'est-à-dire la log-fonction de partition, rencontrée dans le chapitre 2 dans le problème de maximum d'entropie standard, comme « pseudo-information ».

Poursuivant sur l'interprétation du log rapport de vraisemblance, on peut encore noter que les décodeurs itératifs modernes produisent des sorties souples qui sont des log-rapports de vraisemblance. On comprend donc que les probabilités d'erreurs doivent alors pouvoir être obtenues par l'application d'un principe de grande déviation sur la somme des log-rapports de vraisemblance, i.e. leur moyenne empirique. Comme le principe de grandes déviations fait intervenir comme fonction de taux le minimum de l'information de Kullback sous la contrainte, ici de log-rapport de vraisemblance, on voit émerger l'entropie de Rényi de paramètre α , comme en (3.10). Ce paramètre α est le paramètre de Lagrange associé à la contrainte et lié à celle-ci par une opération de dérivation. Bien qu'assez fier de cette interprétation, je dois renvoyer les lecteurs sur l'article [43] à qui revient la primeur de cette remarque. Curieusement, il ne me semble pas l'avoir vu reprise par ailleurs.

Observables

Les valeurs observables sont habituellement la valeur moyenne sous une certaine distribution. Ici, selon le point de vue, l'observable pourrait être soit la moyenne sous R , la distribution d'un sous système, ou encore sous la distribution d'équilibre P^* . Ainsi, on peut chercher à minimiser plus avant la divergence de Kullback-Leibler $D(P||Q)$, une fois la contrainte de log-vraisemblance moyenne atteinte, en ajustant R de sorte à vérifier la contrainte de moyenne. Ceci peut s'écrire

$$K = \left\{ \min_{P_1} \left\{ \begin{array}{l} \min_P D(P||Q) = \min_P \int P(x) \log \frac{P(x)}{Q(x)} dx \\ \text{s.c. } \theta = \int P(x) \log \frac{R(x)}{Q(x)} dx \\ \text{s.c. } m = E_R[X] \text{ or } m = E_{P^*}[X] \end{array} \right. \right\}, \quad (3.12)$$

En prenant en compte (3.9), on obtient alors

$$K = \sup_{\alpha} \left[\alpha \theta - \left\{ \begin{array}{l} \max_R (\alpha - 1) D_{\alpha}(R||Q) \\ \text{s.c. } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. \right] \quad (3.13)$$

qui revient simplement à la recherche d'un extremum de la divergence de Rényi sous contrainte de moyenne. On notera également que la moyenne sous la distribution P^* n'est rien d'autre que la moyenne généralisée utilisée en « nonextensive statistics ». Bien entendu, les solutions correspondantes sont les distributions de Tsallis que nous avons indiquées en (3.6).

3.3 Distributions de Tsallis et distribution des excès

C'est en étudiant les domaines d'apparition, ou d'utilisation, de modèles de distribution de lois puissance, et notamment à la lecture de [44] que je vis apparaître une possibilité de connection des “statistiques non extensives” et des résultats de la théorie des excès (Extreme value theory). En effet, on s'aperçoit que la distribution de Tsallis a exactement la même forme (3.6) que la distribution de Pareto généralisée (GPD) qui apparaît en théorie des excès. Le point intéressant est que ces distributions de Pareto généralisée sont utilisées hors du champ de la physique statistique pour modéliser des distributions à queues lourdes. Les exemples sont nombreux, allant de la théorie de la fiabilité, au trafic dans les réseaux, à l'hydrologie, la climatologie, la géophysique, la science des matériaux, l'imagerie radar et l'économie.

La pertinence, sinon toujours une justification, de ce modèle de distribution de Pareto généralisée dans ces applications peut être liée au théorème de Balkema-de Haan-Pickands [45, 46] qui indique que la distribution des excès au dessus d'un seuil élevé est souvent bien approchée par une GPD. Le résultat de Pickands et Balkema-De Haan relie l'appartenance au domaine d'attraction d'une des distributions extrême au comportement de la queue de la distribution considérée. Il montre de plus que la convergence vers une des trois distributions limite n'a lieu que si la queue de la distribution, vue au delà d'un seuil de

plus en plus élevé, est décrit par une GPD, avec une erreur aussi faible que l'on veut. Plus spécifiquement, on considère une variable aléatoire X , un seuil u et on note X_u la variable conditionnelle représentant les excès au dessus du seuil u

$$X_u = X - u | X > u.$$

La fonction de survie correspondante s'écrit alors

$$\bar{F}_{X_u}(x) = \Pr(X > x + u | x > u) = \frac{\bar{F}_X(x + u)}{\bar{F}_X(u)}, \quad x \geq 0. \quad (3.14)$$

Le théorème de Pickands indique que la loi de X appartient au domaine de convergence d'une distribution extrême si et seulement si on peut trouver γ et σ tel que

$$\lim_{u \rightarrow x_\infty} \inf_{\sigma > 0} \|\bar{F}_{X_u}(x) - \bar{G}_{\gamma, \sigma}(x)\|_\infty = 0$$

où $\bar{G}_{\gamma, \sigma}$ est la fonction de survie associée à la densité de la GPD en (3.6) :

$$\bar{G}_{\gamma, \sigma} = \left(1 + \frac{\gamma}{\sigma}x\right)_+^{-\frac{1}{\gamma}} \quad (3.15)$$

avec la notation $(x)_+ = \max(0, x)$. Le cas $\gamma > 0$ correspond au domaine de Fréchet, avec des lois à queues lourdes (Cauchy, t , Pareto), le cas $\gamma < 0$ à des lois à support borné (x appartient à l'intervalle $[0, -\sigma/\gamma]$) du domaine de Weibull, et enfin le cas $\gamma = 0$ correspond à une distribution exponentielle (domaine de Gumbel – exponentielle, gamma, normale, K, etc).

Jusqu'alors, les distributions de Tsallis, qui sont des GPD, étaient justifiées soit par le recours au principe de Maximum d'entropie de Tsallis, soit par fluctuations du paramètre de la loi de Boltzmann (suivant une loi gamma - de manière analogue à ce que nous avons décrit en section 2.3). En notant que ces distributions sont des GPD, et en nous appuyant sur le résultat de Pickands, nous voyons donc que ces distributions peuvent également être interprétées statistiquement comme résultant d'un seuillage ou d'une série de seuillages. C'est le point de vue original que nous avons exposé et illustré dans l'article [31] qui est annexé ici en pages 93 à 102.

Dans cet article, nous avons également fait un parallèle entre le théorème de Pickands et le théorème central limite et souligné la propriété de stabilité des GPD par seuillage (au sens où une GPD reste une GPD avec le même exposant). Nous avons présenté un certain nombre de procédures d'estimation des paramètres de la GPD. En effet, pour l'estimation des paramètres de distributions à queues lourdes, il convient d'être précautionneux, ce qui n'est pas forcément toujours le cas dans les applications en statistiques nonextensives [47, 48]. Enfin, nous avons repris brièvement les résultats que nous avons exposés à la conférence ISIT [32] dans lesquels nous établissons une autre connection avec le maximum d'entropie de Tsallis.

Dans cet autre travail, on commence par montrer que la GPD peut être obtenue comme solution d'un problème de maximum d'entropie de Rényi-Tsallis, sous des contraintes de moyenne et de normalisation. Dans un second temps, on considère les distributions appartenant au domaine d'attraction de Fréchet, et on caractérise les normes 1 et q (la norme q étant proportionnelle à l'entropie de Tsallis) de la fonction de survie associée à la variable d'excès $X - u | X > u$. En utilisant une normalisation appropriée, on définit une nouvelle variable, qui peut être aussi simple que X_u/u , dont la moyenne et les normes 1 et q de la fonction de survie convergent vers des valeurs constantes. Le même travail est fait pour une sous-classe des distributions du domaine de Gumbel. On montre alors que la distribution des excès est une distribution de Tsallis. Dans le cas Fréchet, on identifie le paramètre q comme une fonction simple du paramètre de la famille de Fréchet, de sorte à ce que la moyenne, les moments d'ordre 1 et q coïncident avec ceux de la variable des excès normalisée. Dans la mesure où la solution d'un problème de maximum d'entropie sous contraintes est unique, on obtient que la variable des excès obéit asymptotiquement à une loi de Pareto généralisée, qui est solution d'un problème de maximum d'entropie de Rényi-Tsallis.

3.4 Fonctionnelles d'entropie dérivées de la divergence de Rényi

Dans la référence [29], reprise dans le paragraphe 3.2, nous avons vu que l'on peut ramener un problème de minimisation de l'information de Kullback, sous contrainte de log-vraisemblance et de moyenne, à un problème de minimisation de la divergence de Rényi. Le minimum de l'information sous contrainte de moyenne m est bien-sûr une fonction de cette moyenne, qui est une entropie dans l'espace des observations. Dans le cas de l'information de Kullback, ceci correspond au principe de contraction qui amène aux fonctions de taux (entropie de niveau 1) dans la théorie des grandes déviations. Dans la lignée des travaux précédents, il est tout-à-fait naturel de se pencher sur les propriétés de la minimisation de la divergence de Rényi et des fonctionnelles entropiques associées. L'article [33] qui est annexé ici en pages 105 à 121 reprend les résultats qui ont été obtenus.

On a considéré une minimisation de la divergence de Rényi sous une contrainte de moyenne standards, mais aussi, comme en physique statistique, sous une contrainte de moyenne généralisée. Celle-ci est, on l'a vu plus haut, la moyenne prise par rapport à l'"escort distribution"

$$P^*(x) = \frac{P(x)^\alpha Q(x)^{1-\alpha}}{\int_{\mathcal{D}} P(x)^\alpha Q(x)^{1-\alpha} dx}. \quad (3.16)$$

Celle-ci se réduit d'ailleurs à la distribution P quand α tend vers un, et la moyenne généralisée redevient la moyenne habituelle. Le problème de minimisation de la divergence de Rényi $D_\alpha(P||Q)$ d'une distribution P à une distribution Q , sujet à une contrainte de normalisation et à une contrainte de moyenne classique (C) ou généralisée (G) s'écrit comme suit :

$$\mathcal{F}_\alpha^{(C \text{ resp. } G)}(m) = \begin{cases} \min_P D_\alpha(P||Q) \\ \text{s.c. } (C) \quad m = E_P[X] \\ \text{ou } (G) \quad m = E_{P^*}[X] \\ \text{et } \int_{\mathcal{D}} P(x) dx = 1 \end{cases} \quad (3.17)$$

où $\mathcal{F}_\alpha^{(C)}(m)$ et $\mathcal{F}_\alpha^{(G)}(m)$ sont les entropies de niveau 1 associées à la divergence de Rényi pour les contraintes classique et généralisée.

Dans le travail présenté dans l'article annexé [33], on étudie les formes et propriétés des solutions des solutions du problème ci-dessus, sous les deux types de contraintes. Les valeurs prises à l'optimum, vues comme fonctions de la contrainte, définissent des fonctionnelles entropiques $\mathcal{F}_\alpha^{(\cdot)}(m)$ associées à chaque choix de mesure de référence Q et indexées par le paramètre α . Il me semble que cette étude est complètement originale.

On commence par établir la forme exacte des distributions P qui réalisent le minimum de l'information de Rényi dans le terme de droite de (3.17), et on montre que ces distributions, définies sur un sous-ensemble \mathcal{D} de \mathbb{R} (avec notamment \mathcal{D} tel que l'intégrale soit finie et la densité de probabilité toujours réelle positive), sont de la forme

$$P_\nu(x) = \frac{[1 + \gamma(x - \bar{x})]^\nu}{Z_\nu(\gamma, \bar{x})} Q(x). \quad (3.18)$$

La normalisation $Z_\nu(\gamma, \bar{x})$ est une fonction de partition vérifiant

$$Z_\nu(\gamma, \bar{x}) = \int_{\mathcal{D}} [1 + \gamma(x - \bar{x})]^\nu Q(x) dx \quad (3.19)$$

Plus spécifiquement, on montre que dans le cas de la contrainte (C), on a $\nu = \xi = \frac{1}{\alpha-1}$ et $\bar{x} = E_P[X] = E_\xi[X]$; et dans le cas de la contrainte généralisée, on a $\nu = -\xi = \frac{1}{1-\alpha}$ et $\bar{x} = E_{P^*}[X] = E_{-(\xi+1)}[X]$, la moyenne généralisée. Il est important de voir que \bar{x} est dans les deux cas une moyenne statistique et non la contrainte m , et par suite une fonction du paramètre γ . C'est uniquement dans le cas du paramètre optimal γ^* que l'on a bien entendu $\bar{x} = m$.

On liste quelques propriétés des distribution P_ν en (3.18) et de leurs fonctions de partition (3.19). En particulier on montre que la divergence de Rényi, et par suite les fonctionnelles entropiques $\mathcal{F}_\alpha^{(\cdot)}(m)$, s'expriment simplement comme le logarithme de ces fonctions de partition. On établit ensuite des propriétés générales de ces fonctionnelles, dont le fait qu'elles sont toujours non-négatives et présentent un minimum unique en m_Q , la moyenne sous la référence Q . De plus, on montre que $\mathcal{F}_\alpha^{(C)}(m)$ est strictement convexe, pour $\alpha \leq 1$. On montre également que les fonctionnelles $\mathcal{F}_\alpha^{(\cdot)}(x)$ vérifient une symétrie spécifique : $\mathcal{F}_\alpha^{(C)}(x) = \mathcal{F}_{\frac{1}{\alpha}}^{(G)}(x)$. A partir de ces fonctionnelles, on définit aussi une divergence dans l'espace des moyennes possibles, cette divergence se réduisant à une divergence de Bregman lorsque $\alpha \rightarrow 1$.

Dans la mesure où les distributions optimales (3.18) sont définies implicitement, puisqu'elles dépendent de leur moyenne, la détermination des paramètres n'est pas évidente (il existe des techniques itératives, mais qui supposent que la solution est un attracteur). Pour résoudre ceci, on définit deux fonctionnelles dont la maximisation fournit le paramètre γ , pour les deux types de contraintes. On vérifie que les valeurs γ^* du paramètre γ des distributions optimales solution de (3.17) sont les minimum des maximiseurs de

$$D_C(\gamma) = -\log Z_{\xi+1}(\gamma, m) \quad (3.20)$$

$$D_G(\gamma) = -\log Z_{-\xi}(\gamma, m) \quad (3.21)$$

où les deux fonctions de partition sont convexes, mais éventuellement définies sur plusieurs intervalles. Les fonctionnelles entropiques associées $\mathcal{F}_\alpha^{(\cdot)}$ sont alors données par

$$\mathcal{F}_\alpha^{(C \text{ resp. } G)}(m) = D_{C \text{ resp. } G}(\gamma^*). \quad (3.22)$$

Remarque - C'est dans le cas discret et pour $\nu < 0$ (ce qui correspond à la contrainte standard (C) avec $\alpha \in (0, 1)$), que la fonction de partition $Z_\nu(\gamma, m)$ présente des singularités sur son support, et n'est définie que sur des segments. Dans ce cas de figure, $-\log Z_\nu(\gamma, m)$ peut présenter plusieurs maxima. Cela indique que pour une moyenne donnée, il existe plusieurs distributions de la famille P_ν qui admettent cette moyenne. Réciproquement, on peut aussi montrer que pour un paramètre gamma donné (et γ est homogène à un « inverse de température »), il peut exister plusieurs moyennes (états) associées. On peut même définir une fonction potentiel dont les maxima fournissent les valeurs de ces moyennes. Ces différentes solutions semblent correspondre à des distributions dont le support est différent, ou s'étend. Ce phénomène qui reflète le fait que la transformation $\gamma \leftrightarrow \bar{x}(\gamma)$ n'est pas bijective, peut être imaginé comme une transition de phase. Il y a sur ces points un travail à effectuer.

On présente ensuite 4 cas particuliers de mesure de référence $Q(x)$: des distributions uniforme et exponentielle dans le cas de systèmes à états continu, puis une distribution de Bernoulli et une distribution de Poisson. Dans chaque cas, on calcule la fonction de partition et les fonctions duales associées, et on cherche à déterminer une expression des fonctionnelles $\mathcal{F}_\alpha^{(\cdot)}(x)$. On vérifie que l'on obtient des fonctionnelles connues (Burg, Fermi-Dirac, Kullback), lorsque α tend vers 1. Dans tous les cas, on est fourni des évaluations numériques des fonctionnelles entropiques. Les figures Fig. 3.1 et 3.2, extraites de l'article, illustrent les comportements que l'on peut obtenir. L'exploitation de ces fonctions pour des problèmes de sélection, de résolution de problèmes inverses peut faire l'objet d'une poursuite de ce travail.

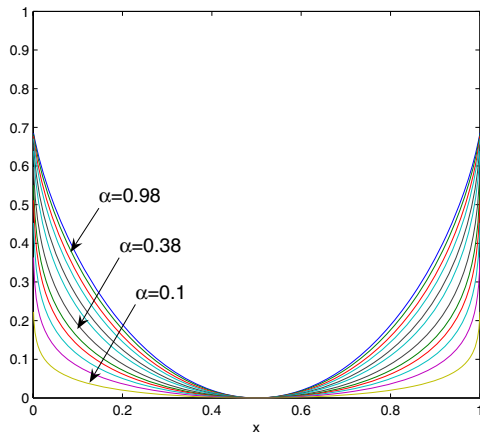


FIG. 3.1 – Fonctionnelle entropique $\mathcal{F}_\alpha^{(C)}(x)$ dans le cas d'une mesure de Bernoulli et $\alpha \in (0, 1)$.

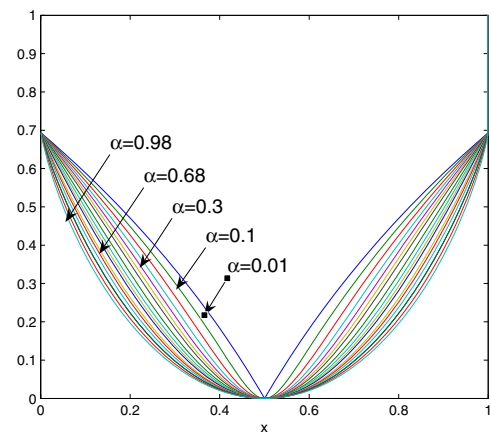


FIG. 3.2 – Fonctionnelle entropique $\mathcal{F}_\alpha^{(G)}(x)$ dans le cas d'une mesure de Bernoulli et $\alpha \in (0, 1)$.

Chapitre 4

Information de Fisher

L'information de Fisher est d'une importance fondamentale en théorie de l'estimation, notamment par le théorème de Cramér-Rao sur la variance d'un estimateur. Elle est également dans des problèmes d'inférence et d'interprétation en physique, avec le principe d'*Extreme Physical Information* défendu par Roy Frieden, avec de nombreux exemples [49, 50, 51]. L'information de Fisher est encore utilisée comme un outil de caractérisation de signaux et systèmes [52, 53], avec des applications en géophysique, [54, 55, 56], en biologie [57], en reconstruction [58, 59] en traitement du signal [60, 61].

Avec Christophe Vignat, nous nous sommes intéressés à trois problèmes liés à l'information de Fisher, et ce chapitre reprend les contributions correspondantes. La première concerne la transformation de l'information de Fisher par des systèmes singuliers [62]. La seconde étudie l'interaction de l'information de Fisher et de l'entropie de Shannon et introduit le fameux "Fisher-Shannon information plane" [63]. Enfin, dans un travail très récent, nous étudions les distributions d'information de Fisher minimale, avec une variance fixée et sous contrainte de support borné [64].

4.1 Quelques éléments de contexte

L'information de Fisher est une mesure de l'information portée par la loi d'une variable aléatoire X sur un paramètre, généralement noté θ . Si $f(x, \theta)$ est la vraisemblance, la fonction score est définie par

$$\phi_\theta(X) = \frac{\partial}{\partial \theta} \log f(X; \theta) = \frac{1}{f(\theta; X)} \frac{\partial f(\theta; X)}{\partial \theta}.$$

et l'information de Fisher associée à θ est la variance de la fonction score :

$$\mathcal{I}(\theta) = \mathbb{E} \left\{ \left[\frac{\partial}{\partial \theta} \ln f(X; \theta) \right]^2 \middle| \theta \right\},$$

où l'on a utilisé le fait que la moyenne de la fonction score est nulle.

Si $T(X)$ est une statistique pour θ , alors

$$\mathcal{I}_T(\theta) \leq \mathcal{I}_X(\theta),$$

avec égalité uniquement si $T(X)$ est suffisante.

L'inégalité de Cramér-Rao fournit une borne sur l'erreur quadratique moyenne. Si $T(X)$ est un estimateur de θ , alors on a toujours

$$\mathbb{E} [(T(X) - \theta)^2] \geq \frac{(1 + B(\theta)')^2}{\mathcal{I}(\theta)} + B(\theta)^2, \quad (4.1)$$

avec $B(\theta) = \mathbb{E} [T(X) - \theta]$. Dans le cas non biaisé, l'inégalité se réduit à

$$\text{Var} \{T(X)\} \geq \frac{1}{\mathcal{I}(\theta)}. \quad (4.2)$$

On sait également que l'information de Fisher apparaît dans l'expression de l'information de Kullback-Leibler entre deux densités de probabilité, où le paramètre de l'une est le perturbé de l'autre. Dans le cas d'une perturbation infinitésimale, on obtient que l'information de Kullback-Leibler devient une forme quadratique pondérée par l'information de Fisher. L'information de Fisher est alors une métrique sur la variété des distributions de probabilité paramétrées.

Dans le cas multivarié, si θ est le vecteur $\theta = [\theta_1, \theta_2, \dots, \theta_N]$, alors l'information de Fisher est une matrice de terme générique

$$(\mathcal{I}(\theta))_{i,j} = E \left[\frac{\partial}{\partial \theta_i} \ln f(X; \theta) \frac{\partial}{\partial \theta_j} \ln f(X; \theta) \right]. \quad (4.3)$$

L'inégalité de Cramér-Rao s'écrit alors $\text{Var} \{ \mathbf{T}(\mathbf{X}) \} \geq \mathcal{I}(\theta)^{-1}$, et l'inégalité s'entend au sens où la différence des deux matrices est définie non-négative.

L'information de Fisher est définie pour une famille paramétrique, contrairement à l'entropie, qui est définie pour toute distribution. On peut en fait définir une famille paramétrée associée à toute distribution en considérant le paramètre de localisation, et définir l'information de Fisher associée à la famille $f(x - \theta)$. Dans ce cas, la dérivation par rapport à θ est équivalente à la dérivation par rapport à x , et on obtient alors

$$\mathcal{I}(X) = E \left\{ \left[\frac{\partial}{\partial x} \ln f(X) \right]^2 \right\} = \int_{-\infty}^{+\infty} f(x) \left[\frac{\partial}{\partial x} \ln f(x) \right]^2 dx. \quad (4.4)$$

Il s'agit de l'information de Fisher de la distribution. Lorsque x est vectoriel et le paramètre de translation scalaire, l'information de Fisher attachée à la distribution est scalaire. Si le paramètre de translation est vectoriel, alors c'est une matrice d'information de Fisher de la distribution qui apparaît en (4.4). Dans le cas d'un paramètre d'échelle σ , on obtient $\mathcal{I}(X) = \int_0^{+\infty} (1 + xp(x)/p(x))^2 p(x) dx$, en $\sigma = 1$.

C'est essentiellement à l'information de Fisher sur la distribution (4.4) que nous nous sommes intéressés (jusqu'ici). Pour cette information, on dispose d'une inégalité de convolution, la Fisher Information Inequality, FII [65] :

$$\mathcal{I}(X + Y)^{-1} \geq \mathcal{I}(X)^{-1} + \mathcal{I}(Y)^{-1}, \quad (4.5)$$

qui est aussi valable dans le cas où X et Y sont des vecteurs. En utilisant cette inégalité et le théorème de De Bruijn, on peut obtenir l'inégalité de la puissance entropique, l'Entropy Power Inequality EPI. Si on pose

$$N(\mathbf{X}) = \frac{1}{2\pi e} \exp \frac{2}{d} H(\mathbf{X})$$

où $H(\mathbf{X})$ est l'entropie de Shannon et d la dimension de X , alors l'EPI s'écrit

$$N(\mathbf{X} + \mathbf{Y}) \geq N(\mathbf{X}) + N(\mathbf{Y}).$$

On peut encore compléter ces inégalités par des relations d'incertitude (isoperimetric inequality), cf [65]

$$\frac{1}{d} \mathcal{I}(\mathbf{X}) N(\mathbf{X}) \geq 1 \text{ et } N(\mathbf{X}) |\mathcal{I}(\mathbf{X})|^{\frac{1}{d}} \geq 1 \quad (4.6)$$

où $\mathcal{I}(\mathbf{X})$ est la matrice d'information de Fisher. On notera que si on applique un facteur d'échelle sur le vecteur aléatoire, $\mathbf{Y} = \mathbf{A}\mathbf{X}$, l'inégalité de droite reste vraie.

Enfin, pour un paramètre de translation, l'inégalité de Cramér-Rao est

$$\mathcal{I}(\mathbf{X}) \mathbf{K}(\mathbf{X}) \geq \mathbf{1}, \quad (4.7)$$

où \mathbf{K} est la matrice de covariance de la loi de X .

4.2 Inégalités sur l'information de Fisher

Les premiers résultats sur la transformation de l'information de Fisher ont été donnés dans les années 60 par [66] et Stam [67]. Il s'agit en particulier de l'inégalité

$$\mathcal{I}^{-1}\left(\sum_i a_i X_i\right) \geq \sum_i a_i^2 \mathcal{I}^{-1}(X_i).$$

En 1993, Papathanasiou [68] donna une importante série d'inégalités sur l'information de Fisher, avec des applications à la caractérisation de la normalité. En parallèle et indépendamment, Zamir [69] établit ce même résultat et discutât les cas d'égalité [70, 71]. Il s'agit de l'inégalité

$$\mathcal{I}^{-1}(\mathbf{Y}) \geq \mathbf{A} \mathcal{I}^{-1}(\mathbf{X}) \mathbf{A}^T \quad (4.8)$$

où \mathbf{A} est une matrice de rang lignes plein, \mathbf{X} est un vecteur aléatoire à composantes indépendantes, et $\mathbf{Y} = \mathbf{A}\mathbf{X}$. Une autre inégalité est trouvée dans ce travail :

$$\mathcal{I}(\mathbf{X}) \geq \mathbf{A}^T \mathcal{I}(\mathbf{Y}) \mathbf{A}. \quad (4.9)$$

Ces deux inégalités constituent la version de FII multivariée, donnée par Zamir. On note par ailleurs que dans le cas inversible, on a simplement $J(\mathbf{A}\mathbf{X}) = \mathbf{A}^{-T} J(\mathbf{X}) \mathbf{A}^{-1}$, ce qui permet de retrouver le membre de droite dans (4.8). Cette borne fournit une fonction de contraste possible pour les problèmes de déconvolution et séparation de sources dans le cas convolutif.

Cependant, les démonstrations fournies dans les articles et complétées dans le rapport technique [71] sont assez ardues (représentation conditionnelle de la fonction score et extension de l'inégalité de Stam au cas vectoriel), et plus encore pour ce qui concerne l'étude des cas d'égalité. Nous avons pu trouver une preuve très simple de la FII de Zamir, en nous appuyant simplement sur une inégalité sur les matrices partitionnées. Donnons l'idée de la démarche. On considère la factorisation LDL d'une matrice symétrique, définie non-négative, partitionnée selon $U = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$, avec $B = C^T$ et D carrée de rang plein. La condition de définie non-négativité entraîne $A - BD^{-1}C \geq 0$, avec égalité si U a le même rang que D . Il ne reste plus qu'à identifier les matrices qui permettent de retrouver ((4.8) et ((4.9) : dans le premier cas on utilise la matrice de covariance de $[\phi_X(X)\phi_Y(Y)]^T$, et dans le second cas celle de $[\phi_Y(Y)\phi_X(X)]^T$, où les ϕ sont les fonctions score. Enfin pour conclure, il faut utiliser le fait que $E[\phi_X(X)\phi_Y(Y)^T] = \mathbf{A}^T \mathcal{I}(\mathbf{Y})$, une égalité qui demande un peu de réflexion. Les cas d'égalité se déduisent ensuite à partir de la condition de rang, et on a relié ceux-ci aux concepts de pseudoinverse et de gaussianité.

Par ailleurs, dans sa démonstration originale, Zamir utilise l'expression de la fonction score de l'observation $\mathbf{Y} = \mathbf{A}\mathbf{X}$ en fonction de la moyenne conditionnelle de \mathbf{X} . On cherche à exprimer $\phi_X(X)$ en fonction de $\phi_Y(Y)$, sachant que $Y = AX$, avec A non inversible. Nous avons formulé le problème comme un problème d'estimation au sens MMSE : $\arg\min_w E[|\phi_X(X) - w(Y)|^2]$ sous la contrainte $Y = AX$. On obtient alors que l'estimateur s'écrit $E[\phi_X(X)|Y] = A^T \phi_Y(Y)$, retrouvant ainsi les résultats de Zamir, avec même l'économie d'une hypothèse. Ces différents résultats ont été présentés à ISIT, [72], où nous avons rencontré Ram Zamir. Un article [62], a ensuite été publié dans Journal of Inequalities in Pure and Applied Mathematics. Cet article est annexé en pages 126 à 133. Notons enfin que ces résultats ont été repris dans une entrée de Mathworld <http://mathworld.wolfram.com/FisherInformationMatrix.html>.

4.3 Plan d'information de Fisher-Shannon

Dans le travail précédent, nous nous étions intéressés à l'extension de l'inégalité sur l'information de Fisher (4.5). Nous nous sommes ensuite attardés sur la relation d'incertitude (4.6), dans le cas scalaire cette fois :

$$\mathcal{I}(X)N(X) \geq 1. \quad (4.10)$$

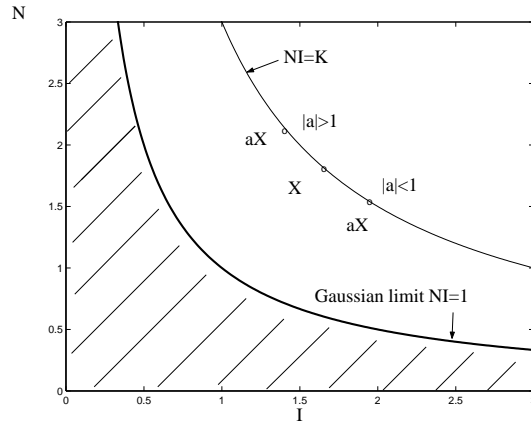


FIG. 4.1 – Plan d'information de Fisher-Shannon

En remarquant que par une transformation d'échelle a , on a $\mathcal{I}(aX) = a^{-2}\mathcal{I}(X)$ et $N(aX) = a^2N(X)$, on voit que le produit $\mathcal{I}(X)N(X)$ est invariant par cette transformation. Ces deux propriétés, relation d'incertitude et invariance montre que l'information de Fisher et l'entropie power sont liées. Nous avons alors introduit la notion de plan d'information de Fisher-Shannon et proposé de caractériser un phénomène physique (ou plutôt la loi sous-jacente) par une position dans le plan d'information de Fisher-Shannon. Dans ce plan, la transformation d'échelle précédente se traduit par l'évolution sur une courbe $\mathcal{I}(X)N(X) = K$, où K est une constante. Ceci est illustré sur la figure Fig. 4.1.

Nous avons étudié deux familles de densité de probabilité : les Student-t, qui sont les distributions de Tsallis obtenues sous contrainte de variance, et les gaussiennes généralisées qui sont également souvent utilisées en physique (et même en signal pour modéliser les statistiques des coefficients d'ondelettes). Dans les deux cas, nous avons donné les expressions de l'entropie et de l'information de Fisher, et nous avons montré qu'il existe une bijection entre les paramètres de la distribution et la position dans le plan Fisher-Shannon.

Une évolution des paramètres des lois se traduit alors par une trajectoire dans le plan Fisher-Shannon. Il est alors possible de construire des trajectoires arbitraires dans le plan Fisher-Shannon ; nous avons ainsi construit une trajectoire en escalier, où l'information de Fisher reste constante tandis que l'entropie power croît, avant de basculer sur une croissance de l'information de Fisher tandis que l'entropie power est fixée. Ceci signifie donc que l'étude d'un seul des deux indicateurs peut être insuffisante et qu'aucun des indicateurs n'est plus adéquat que l'autre. Une autre trajectoire arbitraire est une trajectoire circulaire dans le plan d'information. Cependant, la représentation reste incomplète : on peut par exemple trouver des distributions de variances différentes qui partagent la même coordonnée dans le plan Fisher-Shannon, et il peut être alors judicieux de s'intéresser à une représentation Fisher-Shannon-variance (notons que dans le cas des deux familles que nous avons étudié, un point de même coordonnée en Fisher-Shannon-variance correspond à une loi gaussienne qui est l'intersection de nos deux familles). Bien entendu, on pourrait aussi trouver des distributions différentes qui partageraient une même coordonnée dans la représentation Fisher-Shannon-variance : la bonne caractérisation, c'est la connaissance de toute la loi. La proposition est ici d'une caractérisation partielle, mais qui présente une certaine cohérence en terme d'information. Un article issu de ce travail a été publié dans Physics Letters A, et est ici proposé en annexe, pages 137 à 143, où l'on pourra retrouver des représentations de différentes réalisations suivant les trajectoires décrites ci-dessus.

Cette idée de représentation conjointe a rencontré un certain succès, notamment comme outil d'analyse de systèmes d'atomes [73, 74, 75], de systèmes électroniques [76], (dans ces deux cas, on possède des modèles analytiques des lois de comportement et on peut ainsi comparer des systèmes à l'aide de cet indicateur simple), pour l'analyse de processus d'ionisation [55, 77]. D'autres plans d'information ont été également proposés : Cramér-Rao-Variance [78], Fisher-Rényi [79], ou encore Fisher-Complexité [77, 80].

4.4 Minimum de l'information de Fisher

Une des caractéristiques de l'information de Fisher est bien sûr de faire intervenir explicitement, dans sa définition même, la dérivée de la densité de probabilité. Tout problème d'extrémisation de l'information de Fisher va alors faire apparaître une équation différentielle du second ordre. C'est pour cette raison, au moins du point de vue opérationnel, que l'on peut retrouver la plupart des équations de la physique à partir de l'information de Fisher. Ceci est discuté notamment dans les livres de Frieden [51, 81, 82] et dans de nombreux articles, où celui-ci définit un principe d'extrémisation de l'information EPI Extreme Physical Information, qui préconise d'extrémiser le « flot d'information », la différence de deux informations de Fisher associées respectivement à la source et aux données.

Dans les travaux exposés ci-dessus, nous avons étudié quelques aspects relatifs à l'inégalité sur l'information de Fisher FII (4.5), en examinant l'évolution de l'information lors de la transformation d'un vecteur aléatoire par un système singulier. Nous nous sommes ensuite intéressé à la relation d'incertitude (4.6) et avons introduit la notion de plan d'information de Fisher-Shannon pour l'analyse de signaux. Une autre inégalité fondamentale sur l'information de Fisher est bien entendu l'inégalité de Cramér-Rao (4.7).

Il est bien connu que l'inégalité de Cramér-Rao est saturée par la distribution gaussienne. Ainsi, la distribution sur de variance fixée qui minimise l'information de Fisher est la gaussienne. Ceci est vrai sur \mathbb{R} . Cependant, dans de nombre de cas pratiques on sait que la variable d'intérêt appartient à un sous ensemble de \mathbb{R} (positivité, normalisation, acquisition avec un support fini par exemple). Il est donc intéressant de rechercher, comme modèle de la distribution des données ou pour améliorer l'inégalité de Cramér-Rao, les distributions compatibles avec les données – nous choisissons ici une contrainte de variance, présentant une information de Fisher minimale. Ceci revient à rechercher et caractériser les distributions de support limité, dont l'information de Fisher (relative au paramètre de localisation), soit la plus faible, ce qui indique que le paramètre correspondant, par l'inégalité de Cramér-Rao, est le plus délicat à estimer. Cette étude a été également motivée par la lecture de l'article de Elke Uhrmann-Klingen [83] paru dans *Sankhya : The Indian Journal of Statistics* (1995), qui a analysé les propriétés générale des solutions sur un support compact, mais sans obtenir d'expression explicite. Elke Uhrmann-Klingen a eu l'amabilité de nous envoyer sa thèse de doctorat [84] (1992 - en allemand).

Dans le travail, qui m'a bien occupé ces derniers mois, nous avons obtenu les expressions explicites des densités probabilité d'information de Fisher minimale avec une variance fixée, sur \mathbb{R}^+ et sur un intervalle, on a étudié leur comportement et celui de l'information de Fisher associée. Ces résultats sont actuellement soumis à *Information Sciences* [64] et annexés ici entre les pages 148 et 159. On considère ainsi le problème variationnel

$$I(\sigma^2) = \inf_f \{ I[f] : \text{Supp}[f] = \mathcal{S}, f \in \mathcal{D} \text{ and } \text{Var}[f] = \sigma^2 \} \quad (4.11)$$

où \mathcal{S} désigne le support de la distribution. Une des difficultés rencontrées est que bien que $I[f]$, l'information de Fisher attachée à la distribution, soit convexe, la contrainte $\text{Var}[f] = \sigma^2$ ne définit pas un ensemble convexe. En posant $f = u^2$ et en utilisant le calcul de variations, on obtient l'équation d'Euler-Lagrange

$$u''(z) - (\alpha + \beta z^2) u(z) = 0 \quad (4.12)$$

qui est une équation différentielle parabolique, établie par Weber pour l'étude des solutions de l'équation de Laplace pour des coordonnées cylindriques. Cette équation différentielle est également analogue à celle de l'oscillateur harmonique.

Les solutions de l'équation de Weber peuvent être exprimées à l'aide des fonctions parabolique cylindriques $D_\nu(x)$ et $D_{-\nu-1}(ix)$ [85]. On peut également se ramener à l'équation différentielle de Whittaker par changement de variable, dont les solutions s'expriment à partir des fonctions de Whittaker $M_{\lambda,\mu}(z)$ et $M_{\lambda,-\mu}(z)$, ou encore à l'aide de la fonction $W_{\lambda,\mu}(z)$ de Whittaker qui est reliée aux précédentes. Certaines de ces fonctions peuvent s'exprimer comme une série, ce qui permet d'étudier assez simplement les cas limite.

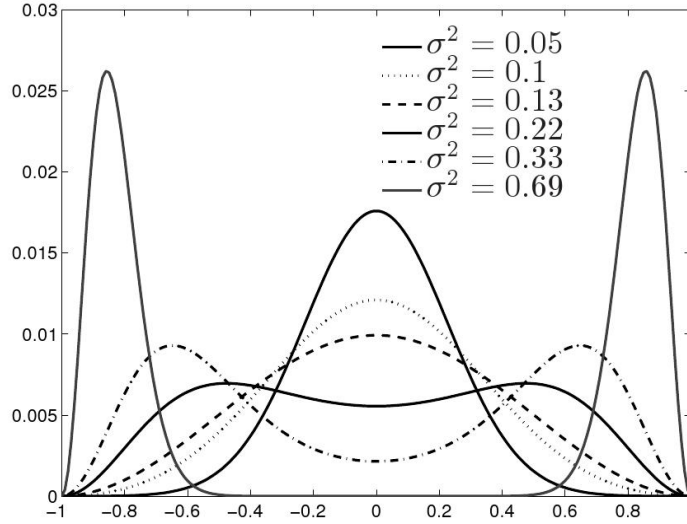


FIG. 4.2 – Distributions d'information de Fisher minimale sur l'intervalle $[-1, 1]$

En prenant pour support les réels positifs, on a alors obtenu comme solution la famille

$$u_{\alpha,\beta}(z) \propto D_{-\frac{\alpha}{2\sqrt{\beta}} - \frac{1}{2}}((2\sqrt{\beta})^{\frac{1}{2}}z) \quad (4.13)$$

où le dénominateur sert à normaliser à une densité $f(z) = u(z)^2$. Pour ces densités, on a établi que

$$I_{\nu}(\sigma^2) = \frac{K_{\nu}}{\sigma^2}, \quad (4.14)$$

avec $K_{\nu} = I_{\nu,1}V_{\nu,1}$, où on a fixé $\xi = 1$. On a ensuite recherché la valeur ν^* de ν qui minimise K_{ν} , de sorte à sélectionner le représentant de la famille d'information de Fisher minimale. Nous avons pu déterminer analytiquement les expressions de l'information de Fisher et de la variance, ce qui a permis d'étudier K_{ν} en fonction de ν . On a ainsi obtenu que $\nu^* = 1$ et ainsi que la distribution qui minimise l'information de Fisher sur \mathbb{R}^+ , avec une contrainte de variance, est

$$f(x) = \sqrt{\frac{2}{\pi}} x^2 e^{-\frac{1}{2}x^2}, \quad (4.15)$$

la chi-distribution à trois degrés de liberté. On obtient alors que le produit Fisher-variance est $K_1 = 9 - 24/\pi \approx 1.3606$. Au passage, on en déduit donc que sur \mathbb{R}^+ , on a toujours

$$I[f] \geq \frac{K_1}{\sigma^2}. \quad (4.16)$$

Pour un support compact (ici $[-1, 1]$), nous avons obtenu l'expression analytique de la solution :

$$u(x) = \frac{\frac{1}{\sqrt{x}} M_{-\frac{\alpha}{4\sqrt{\beta}}, -\frac{1}{4}}(\sqrt{\beta}x^2)}{\left(\int_{-1}^1 \frac{1}{\sqrt{x}} M_{-\frac{\alpha}{4\sqrt{\beta}}, -\frac{1}{4}}(\sqrt{\beta}x^2) dx \right)^{\frac{1}{2}}}. \quad (4.17)$$

où $\sqrt{\beta}$ est le premier zéro $M_{-\frac{\alpha}{4\sqrt{\beta}}, -\frac{1}{4}}$. Cette dernière condition imposant l'annulation de la densité sur les frontières du domaine. Malheureusement, il n'existe pas de solution analytique donnant ces zéros, et les paramètres doivent être obtenus numériquement. La figure 4.2 donne quelques exemples du comportement de cette solution pour différentes contraintes de variance. On note que deux classes de solutions apparaissent : des solutions unimodales et d'autres pour une variance plus élevée, qui sont bimodales.

En fait, on a montré que la fonction $I(\sigma^2)$ est une fonction strictement convexe de σ^2 , avec un minimum (forcément unique) en un certain σ_* , et que pour $\sigma < \sigma_*$, alors $\beta > 0$; pour $\sigma > \sigma_*$, $\beta < 0$ et finalement $\sigma = \sigma_*$, $\beta = 0$; où β est le paramètre de Lagrange de l'équation différentielle (4.12). Dans le cas $\beta < 0$, l'argument de la solution (4.17) devient complexe et cela provoque l'apparition des deux modes. Les cas limite, pour $\sigma \rightarrow 0$ et $\sigma \rightarrow 1$ sont respectivement une densité gaussienne et deux masses sur les frontières de l'intervalle. Finalement, on obtient aussi facilement l'expression de la densité d'information de Fisher minimale sur l'intervalle, correspondant à la variance σ_* :

$$f(x) = \cos(\pi x/2)^2 \quad (4.18)$$

$$\text{avec } \sigma_*^2 = \frac{1}{3} - \frac{2}{\pi^2} \text{ et } I(\sigma_*^2) = \pi^2. \quad (4.19)$$

Des prolongations de ce travail, sont possibles, dans le cas multivarié par exemple, ou encore pour l'analyse des distributions conjointement difficiles, au sens de l'information de Fisher, pour le paramètre de localisation et le paramètre d'échelle.

On a vu que l'on peut soupçonner une « géométrie » autour des lois de Tsallis, qui étendent la notion de famille exponentielle. Cependant, lorsque l'on étudie le développement au second ordre des divergences de Rényi et Tsallis, on obtient l'information de Fisher standard, comme dans le cas de l'information de Kullback-Leibler : la géométrie différentielle associée semble inchangée. Il n'est toutefois pas illégitime d'étudier la correspondance distribution \rightarrow escort distribution et de s'intéresser alors à l'information de Fisher associée aux escort distributions, définie par :

$$I_\alpha[f] = \frac{\alpha^2}{\int f(x)^\alpha dx} \int f(x)^\alpha \left(\frac{df}{dx}\right)^2 dx. \quad (4.20)$$

Il est facile de voir que l'on a une inégalité de Cramér-Rao reliant cette information de Fisher et la variance généralisée. En introduisant l'entropy power associé à l'entropie de Rényi selon

$$V_\alpha[f] = \left[\int f(x)^\alpha dx \right]^{\frac{2}{\alpha} \frac{\alpha'}{\alpha}} \quad (4.21)$$

avec $1/\alpha + 1/\alpha' = 1$, cf [86] et la définition classique dans le cas $\alpha = 1$, il est également possible de vérifier que le produit $V_\alpha[f]I_\alpha[f]$ est invariant par changement d'échelle. Existe-t-il une relation d'incertitude associée ? En tout état de cause, on a le lien suivant [87] entre l'information de Fisher généralisée et l'entropie de Rényi :

$$I_\alpha[f] \geq (\alpha - 1)^2 \left(\frac{\partial H_\alpha}{\partial \theta} \right)^2. \quad (4.22)$$

Une autre question serait de déterminer la forme des densités à minimum d'information de Fisher, sous contrainte de variance ou sous contrainte de variance généralisée. Enfin, on peut noter que les distributions de Tsallis, les Student-t et r, vérifient l'équation différentielle $f^{2\alpha-1}(x)f''(x) = [ax^2 + b]$ simple et étudier les problèmes qui feraient apparaître une équation différentielle de ce type.

Chapitre 5

Problèmes d'architecture Radio

La demande des usagers, la force commerciale qui conduit chacun à imaginer regarder des images animées en prenant le métro, les nouveaux usages (localisation gps, lecture mp3, internet mobile, photographie, vidéo, etc) conduisent à un développement extraordinaire de la complexité, des applications et des systèmes qui doivent faire face à de multiples schémas de modulation, protocoles, bandes de fréquence multiples. Bien entendu les applications citées ci-dessus demandent une très forte intégration et des capacités de traitement du signal très importantes. Les usagers entendent aussi disposer de “multifonctions” (le terme est de Rufin dans son livre Globalia [88]) légers en poids, en prix et en consommation. En ce qui concerne l'aspect consommation, une des clés est de rechercher des architectures permettant d'utiliser un amplificateur dans un mode le plus efficace possible. Dans cet ordre d'idée, j'ai participé à un travail sur une évolution de l'architecture EER (travail avec I. Jamin et C. Berland), qui est commenté ici 5.1. Une autre clé va être d'intégrer au maximum la partie des composants analogiques radiofréquence (RF). Il semble que ce soit cependant une tâche particulièrement complexe, et que s'il est certainement possible de développer des composants analogiques hautement intégrés, le coût en étude, réalisation et consommation ne soit pas raisonnable. Une autre possibilité est de chercher à développer des alternatives numériques (avec un coût de production plus faible) aux composants analogiques. C'est le chemin qui a été initié notamment chez Texas Instrument par l'équipe de R. Staszewski avec le développement d'un “*digital RF processor*” [89], qui permet en outre d'aller vers des systèmes reconfigurables. Dans un travail avec C. Joubert, G. Baudoin, T. Divel, nous avons apporté quelques contributions à l'étude des boucles à verrouillage de phase entièrement numérique. Ces contributions sont abordées section 5.2. Avec C. Berland, nous pensons aussi qu'il faut profiter du traitement du signal, ainsi de plus en plus intégré dans l'émetteur (et le récepteur) pour compenser les défauts des composants analogiques, voire pour relâcher les contraintes sur ces composants. Ceci sera discuté plus avant dans le chapitre suivant.

La montée en fréquence nécessite aussi des composants alternatifs aux composants actuels. Ainsi, avec P. Guillot, C. Berland et P. Philippe, nous travaillons à la conception d'un oscillateur stabilisé à très haute fréquence mettant en jeu un BAW, et pouvant servir d'alternative aux systèmes à quartz ou évitant l'emploi d'une PLL. Ce travail est rapidement présenté section 5.3.

D'un autre côté, si on veut essayer de faire intervenir le traitement du signal au plus tôt dans la chaîne de traitement, se pose le problème d'un échantillonnage précoce du signal. Avec A. Lesellier, O. Venard, C. Berland et O. Jamin, nous débutons une étude sur les solutions pour un échantillonnage et traitement de signaux très large bande dans le cadre d'une application de réception de télévision numérique.

Dans ces différents travaux, il s'agit de travaux conjoints dans lesquels les compétences de mes différents collaborateurs, seniors ou étudiants sont essentielles. Corinne Berland notamment m'a fait bénéficier de ses connaissances et entraîné sur ses thématiques d'architecture EER. C'est par l'interaction recherchée entre différentes visions que nous avons pu avoir quelques idées.

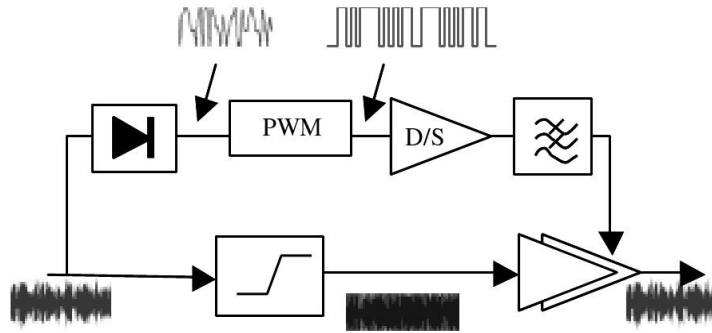


FIG. 5.1 – Architecture EER : version classique

5.1 Une architecture d'émetteur pour les modulations à enveloppe variable

Les systèmes modernes de radiocommunication, du fait des débits très importants qui sont visés, reposent sur des schémas de modulation où les modulations de type QAM sont préférées aux simples modulations de phase ou de fréquence. Alors que le standard 3GPP met en jeu des QPSK et 16QAM, des débits plus élevés, dans les applications de type WLAN ou DVB, sont atteints avec des modulations à multiporteuses. La contrepartie des débits élevés est que la dynamique de l'enveloppe du signal s'accroît, ce qui pose des contraintes pour la conception et la réalisation de la partie analogique du dispositif. En effet, dans la conception d'un émetteur, il est impératif de rechercher à la fois l'efficacité et la linéarité. La variation de l'enveloppe, combinée aux conversions AM/AM et AM/PM de l'amplificateur de puissance dégradent l'EVM (Error Vector Magnitude), le spectre du signal transmis, et l'ACPR (Adjacent Channel Power Ratio). Des méthodes de linéarisation doivent être mises œuvre. Ces méthodes peuvent s'intéresser à corriger les défauts de linéarité des composants, notamment de l'amplificateur, où changer le problème en convertissant le signal en un signal équivalent à enveloppe constante. C'est l'idée initiale de l'EER (Envelope Elimination and Restoration) introduite initialement par Kahn [90]. Il s'agit d'une solution analogique où le signal est séparé en un signal d'enveloppe et un signal de phase. Le signal de phase est amplifié par un amplificateur de type classe C et l'enveloppe est restaurée en sortie du dispositif en modulant l'alimentation de l'amplificateur, voir la figure Fig. 5.1.

L'architecture classique de l'EER peut être modifiée de sorte à ce que la restauration intervienne par multiplication avant l'amplificateur de puissance. Plus exactement, on sépare l'enveloppe et la phase, et on code l'enveloppe par une modulation à largeur d'impulsions et on multiplie l'enveloppe ainsi codée par les signaux I et Q. On obtient alors un signal qui alimente un amplificateur fonctionnant en mode commuté. La variation d'enveloppe est restaurée après l'amplificateur par filtrage passe-bande, comme illustré Fig. 5.2.

Avec C. Berland, nous réfléchissions au problème de détermination d'une forme d'onde optimale pour coder l'entrée à l'aide d'une modulation à largeur d'impulsions PWM. J'ai suggéré d'utiliser un codage Sigma-Delta qui permet d'obtenir des performances comparables à l'aide de dispositifs existants. D'une part cette modification a pu permettre d'améliorer les performances par rapport au codage PWM, et d'autre part nous avons compris que cette modification, en utilisant un $\Sigma\Delta$ à sorties signées $\pm a$ permettait de reconstruire un simple signal à enveloppe constante portant à la fois les informations d'enveloppe et de phase du signal initial : nous avons isolé une procédure permettant de transformer n'importe quel signal en un signal à enveloppe constante, sans perte d'information. Plus encore, la restauration du signal initial se fait simplement par filtrage passe-bande. Du fait de la propriété d'enveloppe constante, le choix d'un amplificateur de puissance n'est plus limité aux amplificateurs commutés. Le schéma de principe est présenté figure Fig. 5.3.

Cette architecture a été étudiée et implantée lors du travail de thèse d'Isabelle Jamin, en convention Cifre avec STMicroelectronics Crolles pour la réalisation matérielle de cette architecture. L'architecture et

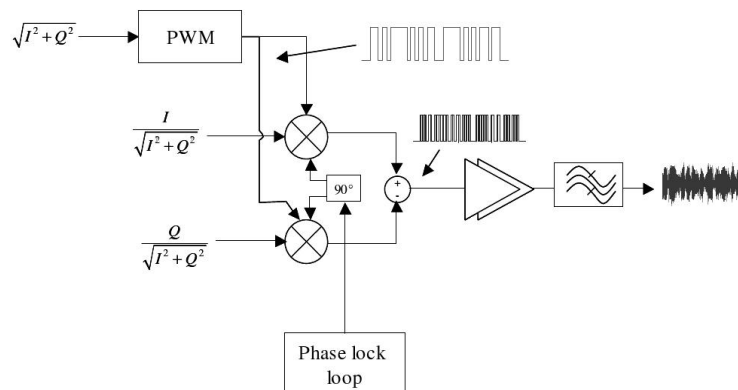
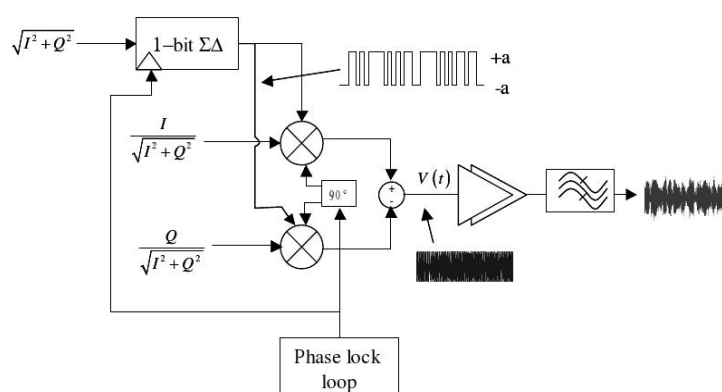


FIG. 5.2 – Nouvelle architecture EER, version commutée :

FIG. 5.3 – Nouvelle architecture EER à modulateur $\Sigma\Delta$ symétrique : version enveloppe constante

ses possibilités d'extension a fait l'objet d'un brevet, en versions nationale et internationale [91] (2006). Dans l'article associé à ce travail, annexé page 164, nous présentons cette architecture, qui peut être adoptée pour tout type de modulation, sur une large gamme de fréquences et avec une grande efficacité et linéarité. Nous illustrons les performances de la solution à PWM, qui transforme le signal initial en un signal commuté, et souffre de termes d'intermodulation et peut présenter des performances insuffisantes. Le second schéma, à modulateur $\Sigma\Delta$ permet de transformer le signal initial en un signal à enveloppe constante. Il met en jeu un modulateur de très haute fréquence. Le choix des filtres passe-bande est analysé, et on montre que la solution permet d'obtenir les performances escomptées en terme de spectre de sortie et d'EVM, pour une modulation sévère comme l'OFDM.

5.2 Boucle à Verrouillage de phase numérique

Pour le développement des systèmes modernes de communication, une des possibilités, comme indiqué dans l'introduction, est d'aller vers des architectures entièrement numériques. Dans cette direction, j'ai coencadré, avec G. Baudoin (directrice de thèse), la thèse Cifre de C. Joubert en convention Cifre avec St Microelectronics, (RF Expertise Center - Cellular Terminal Division), sur l'analyse et l'optimisation d'architectures fondées sur une boucle à verrouillage de phase numérique. L'intitulé exact du sujet était sujet suivant : « Contribution à l'analyse d'une boucle à verrouillage de phase à traitement de signal numérique : stabilité et cycles limites, sources de bruit de phase. ».

En effet, une architecture (et même plusieurs) de boucle à verrouillage de phase entièrement numérique a été proposée par R. Staszewski et son équipe (Texas Instrument). Une synthèse des approches et travaux de Staszewski et al. est présentée dans [92], ainsi que chapitre 10 du livre [89]. Cette architecture était très récente au début de ce travail, la publication « complète » datant de 2005 [93] (mais les éléments étaient disponibles avant dans différents articles de conférence). L'idée du travail de thèse est d'analyser cette architecture, de proposer éventuellement des aménagements, et de l'adapter pour un fonctionnement à plus haute fréquence, pour arriver jusqu'à la réalisation d'un prototype.

Nous avons en particulier proposé un modèle comportemental complet du système, qui permet de le simuler avec une charge en calcul très raisonnable. L'article de congrès [94] présente ce modèle. Nous avons également développé des algorithmes rapides permettant d'alléger l'implantation [95], et calculé explicitement la dsp du bruit de phase correspondant à la structure [96]. Un prototype d'un émetteur FM employant cette PLL et implantant sur silicium les algorithmes développés a été réalisé et validé. Nous décrivons ici les principaux résultats obtenus et annexons au document, page 169, une version longue et plus détaillée de l'article [95].

Un bloc diagramme de la PLL numérique est présenté Figure 5.4. Un oscillateur contrôlé numériquement, le DCO (digitally controlled oscillator) permet d'implanter cette PLL de manière complètement numérique. Les accumulateurs de phase sont utilisés pour compter les périodes de la référence et de la sortie du DCO. Une horloge FS synchronisée avec Fdco, sous-échantillonne la sortie du DCO de sorte que les comparaisons des deux phases soient réalisées avec la même horloge.

Une ADPLL de haute précision est obtenue en utilisant une erreur de phase fractionnaire. Cette erreur est proportionnelle au retard entre les signaux Fdco et Fref [notations étranges de chez TI, mais ce sont bien des signaux temporels]. Le TDC (Time to Digital Converter) est utilisé pour convertir ce retard en une quantité numérique, avec une résolution Δ égale au plus petit retard de propagation au travers d'une porte inverseuse, constituant élémentaire du TDC. Le mot de commande FCW (Frequency Command Word) est donné comme entrée de l'accumulateur de phase et permet de régler la fréquence de sortie selon

$$F_{DCO} = FCW \times F_{REF}.$$

5.2.1 Simulation

Outre la compréhension du fonctionnement, un premier travail est la *simulation* de l'architecture. Compte tenu des fréquences mises en jeu, cette simulation directe est difficile à envisager car elle conduirait à des

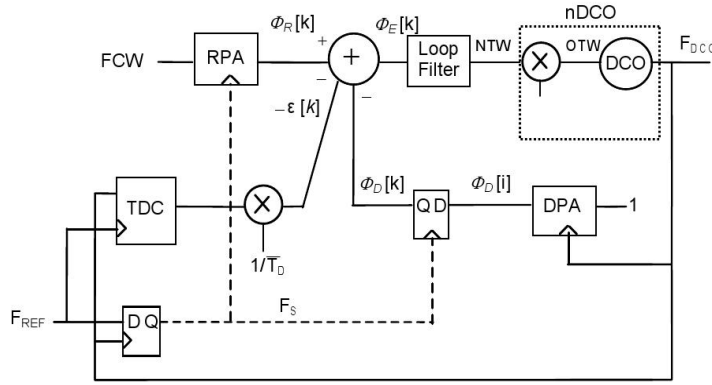


FIG. 5.4 – Synthèse de fréquence RF par ADPLL : principaux blocs fonctionnels et synchronisations.

temps de simulation considérables et des quantités de données énormes. Pour traiter ce point, et disposer d'un moyen d'ajuster et d'optimiser les différents paramètres de la boucle, nous avons développé un *modèle comportemental complet* du système, qui permet de le simuler avec une charge en calcul très raisonnable.

Les composants limitants sont l'accumulateur de phase du DCO et le TDC (avec une résolution de 10 à 20 ps, qui requièreraient une fréquence de simulation extrêmement élevée. L'astuce du modèle de simulation est le fait que les sorties de ces composants peuvent être exprimées uniquement au rythme Fref (ce qui se comprend comme une suppression de porteuse). L'idée est que la variable intéressante est le décalage de phase entre l'horloge de référence Fref et la sortie du DCO Fdco, et de rechercher à ne travailler que sur ce retard. On a alors pu établir une récurrence permettant d'exprimer directement ce décalage τ en fonction de l'état précédent et des caractéristiques du système au rythme Fref. Ce modèle est décrit dans l'article de congrès [94]. A ce sujet, B. Staszewski nous a écrit : « *I read your paper with great interest. Certain passages looked like something I would have written but could not due to the confidential nature of the design and application. I welcome a new and fresh angle on the stuff that I have been working on for quite a long time. I wish we could co-operate on this research. This topic requires a special set of skill and you have excelled.* »

Cependant, le modèle comportemental n'était valide que pour les modes de calibration et d'acquisition. En mode de poursuite, un modulateur $\Sigma\Delta$ est activé pour introduire un dithering à l'entrée du DCO de sorte à atteindre la résolution voulue. Dans ce cas, on a dû (et pu) modifier les équations précédentes avec l'introduction de deux variables intermédiaires, dont une seule est évaluée à la fréquence de fonctionnement du $\Sigma\Delta$, les autres restant évaluées au rythme Fref. Dans la référence [95], on a également intégré une modélisation plus fine du TDC, prenant en compte les effets de quantification temporelle. Ceci permet de simuler le modèle complet avec une complexité très inférieure à celle, déraisonnable, que l'on obtiendrait par simulation directe. Un dernier point est que l'on cherche, bien sûr, à déterminer la densité spectrale de la forme d'onde en sortie du DCO. Si $f_i(t) = f_0 + \Delta f g(t)$ désigne la fréquence instantanée, $\theta_i(t)$ la phase associée, le signal de sortie s'écrit $s(t) = \cos(\theta_i(t))$. On montre simplement que le spectre correspondant peut s'exprimer directement à partir du signal en bande de base :

$$S(f) = \frac{\Delta f}{2} \left(\frac{G(f - f_0)}{(f - f_0)} + \frac{G(f + f_0)}{(f + f_0)} \right)$$

avec $S(f_0) = 1/2 - \pi \Delta f / T \int_0^T \int_0^t g(\tau) d\tau dt$. Ceci permet, à nouveau, de se passer de l'évaluation du signal de sortie, à la fréquence la plus élevée Fdco. À partir de l'ensemble de ces résultats, le modèle comportemental permet une simulation rapide, avec des résultats identiques à ceux délivrés par le modèle de simulation "circuits" développé en VHDL, tout en permettant l'accès aux variables d'intérêt.

Une autre contribution, essentiellement due à Geneviève, est présentée en [96], est le calcul théorique de la densité spectrale de puissance de la sortie d'un DCO soumis à un dithering (celui-ci consiste à commuter deux valeurs de capacité à une fréquence élevée, ce qui permet d'améliorer la résolution

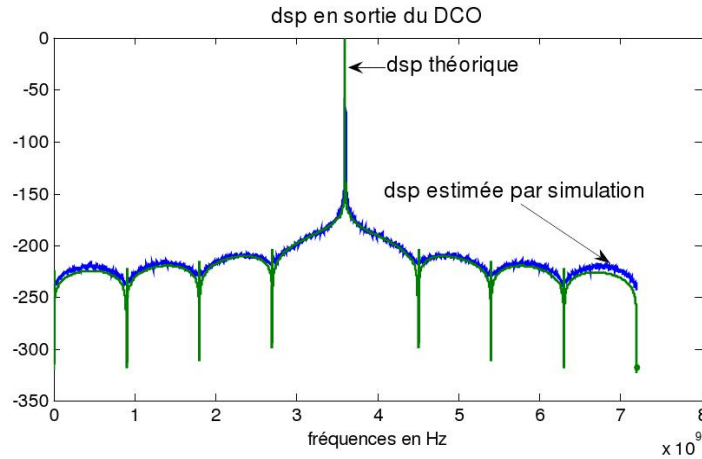


FIG. 5.5 – Comparaison obtenue entre la formule exacte et un résultat de simulation

fréquentielle). En supposant une commutation entre deux capacités suivant une loi binaire non équilibrée, on obtient l'expression exacte de la dsp : l'opération de dithering est assimilée à une modulation de fréquence numérique dans laquelle l'horloge des données est la fréquence de dithering $f_{dither} = 1/T$ et l'excursion de fréquence Δf est égale à la différence entre les deux valeurs de fréquence f_0 et f_1 correspondant aux deux capacités commutées. La fréquence instantanée est alors un signal de type NRZ à deux états $f_c \pm \Delta f/2$. L'indice de modulation correspondant est égal à : $h = \Delta f / f_{dither}$.

En utilisant le principe de la décomposition de Laurent pour les signaux modulés en fréquence à phase continue, on peut récrire le signal généré comme un signal PAM (Pulse Amplitude Modulation), où l'impulsion de mise en forme est une cloche de Laurent. Dès lors, on peut calculer la densité spectrale par la formule de Benett. On retrouve en particulier une constation expérimentale : le fait que la dsp du signal de sortie de l'ADPLL présente une décroissance en $1/f^4$ quel que soit le type de filtre de boucle.

5.2.2 Contributions adaptatives

D'autre part, le système comporte deux calculs d'inversion sur des données qui peuvent être variables dans le temps. Nous avons construit deux algorithmes adaptatifs permettant de suivre ces variations et surtout d'alléger le coût de calcul global. Le premier permet d'éviter une inversion en passant par le biais d'une estimation récursive de l'inverse. Le second permet de calculer la normalisation du gain du DCO. La figure Fig. 5.4 fait apparaître que le calcul de l'erreur fractionnaire $\epsilon(k)$ demande la multiplication par l'inverse de la période (moyenne) du DCO. En posant $\alpha = 1/T$, et en utilisant le critère quadratique $E[\alpha T - 1]$, il est aisé de dériver un algorithme du gradient permettant le calcul de cet inverse. On a aussi proposé au passage de calculer la période moyenne par moyenne exponentielle plutôt que par fenêtre glissante, ce qui permet d'économiser un peu en mémoire. Cette approche a été implanté "en vrai" dans un prototype de PLL numérique pour émetteur FM réalisé et testé par Cyril Joubert chez ST.

Pour introduire une robustesse vis-à-vis des variations du gain du DCO, et une facilité de commande, il est utile d'introduire une normalisation par le gain du DCO. Nous avons proposé également une structure adaptative permettant d'estimer le facteur de normalisation. Simultanément, une formule équivalente a été publiée et évaluée par Staszewski et al. [97]. Notre petit développement a toutefois le mérite d'apporter une justification de l'algorithme ad-hoc proposé. Ces deux algorithmes sont présentés page 174 et suivantes.

5.3 Référence très haute fréquence stabilisée à base de BAW

Les systèmes de communication modernes réclament des oscillateurs de référence à très haute fréquence, notamment pour l'étape de synchronisation entre systèmes. Les nouvelles applications qui requièrent des circuits bas coût, précis et stables peuvent profiter de solutions alternatives aux quartz conventionnels. En effet, un quartz est un composant relativement coûteux qui est difficile à intégrer. Bien que les oscillateurs à quartz présentent une très grande pureté spectrale, ils opèrent à une fréquence basse (26 MHz) et requièrent des multiplicateurs de fréquence nombreux. Une alternative intéressante est la possibilité d'utiliser un résonateur à onde acoustique volumique BAW (Bulk Acoustic Wave) dans une structure d'oscillateur. Le facteur de qualité élevé, couplé à la très haute fréquence de résonance des résonateurs BAW permet de réaliser des résonateurs RF présentant une pureté spectrale équivalente à celle des oscillateurs à quartz. Un des intérêts des résonateurs à BAW est leur très faible taille, qui permettra leur intégration avec le transceiver dans un seul "chip" radio. C'est sur cette problématique que Nxp nous a proposé de travailler, dans le cadre du suivi de la thèse en contrat Cifre de Pierre Guillot (Coencadrement C. Berland/J.-F. Bercher).

Cependant, comparé aux oscillateurs à quartz, les oscillateurs à BAW ont une précision de fréquence plus faible, et il convient donc de disposer d'une possibilité d'ajuster la fréquence de l'oscillateur, sur une gamme de fréquence relativement large. A partir d'une préétude sur la dispersion des composants, on a fixé une contrainte d'un ajustement possible sur une bande de 500 kHz, avec un pas de 0.4ppm, soit 800 Hz pour une fréquence d'oscillation cible de 2 GHz.

Par ailleurs, l'imprécision en fréquence dépend de

- l'erreur initiale sur le résonateur BAW (dispersion des caractéristiques nominales),
- le glissement de fréquence avec les variations de température, qui peuvent être particulièrement sévères dans les mobiles,
- variation de la fréquence avec la valeur de la tension d'alimentation,
- phénomènes de vieillissement.

Il convient ainsi de contrôler les différentes dérives possibles (température, alimentation, vieillissement) et stabiliser le fonctionnement, au cours du cycle de vie du mobile, et ce, d'autant plus que les fréquences de fonctionnement sont élevées. Il convient également de concevoir des solutions qui prennent en compte la variabilité statistique des composants et réalisations, et comprennent une dimension fiabilité.

Une première étude sur la réalisation d'une référence à 2GHz, construite autour d'un BAW, est présentée dans [98] (septembre 2008), où l'on décrit le principe et les caractéristiques du système, avec notamment le dimensionnement d'une banque de capacités qui servira à contrôler la fréquence de l'oscillateur. Nous avons choisi d'utiliser deux banques de capacité, recouvrantes. La première, grossière, permet un réglage sur 1024 ppm, avec un pas de 32 ppm. La banque fine, quant-à-elle permet un réglage sur 50 ppm avec un pas de 0.2 ppm. Ces deux banques, qui font finalement écho à la solution adoptée pour la partie DCO de l'ADPLL, permettront un réglage et un contrôle numérique de l'oscillateur, pour lesquels il faudra développer les stratégies adéquates.

Dans ce premier article [98], que nous rapportons en annexe, pages 179 à 183, on aborde la faisabilité de principe du dispositif, et les différents résultats sont établis par modélisation et simulation. Depuis, Pierre a fait réaliser la référence de fréquence en technologie CMOS 65nm. Une photographie est présentée Fig. 5.6.

Cette réalisation a permis de mesurer pratiquement les performances obtenues et de les confronter aux différents modèles de simulation. En pratique, on a obtenu une grande pureté spectrale (-128dBc/Hz à 100 kHz de la fréquence porteuse). On a noté un décalage de la fréquence d'oscillation. Une faible consommation (0.9 mW) a été mesurée. Sur cette réalisation et les mesures, où ma contribution est essentiellement d'ordre méthodologique, nous envisageons un petit article de revue pour présenter la réalisation, les résultats de mesure et leur comparaison à ceux obtenus par les modèles numérique. Une communication en congrès est soumise [99].

Ainsi, les objectifs en terme de pureté et de consommation ont été atteints et on évalue maintenant l'aspect calibration et contrôle. Un modèle physique du système a été réalisé, notamment concernant la

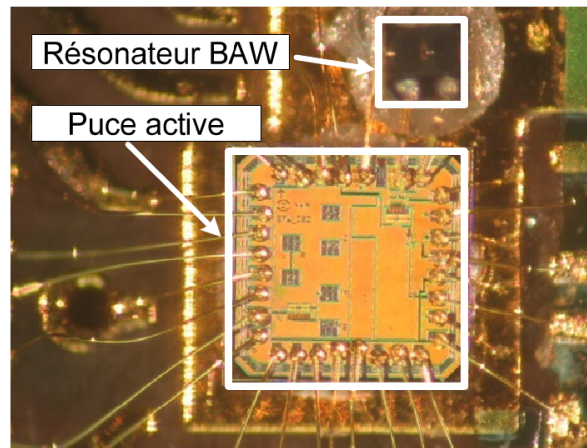


FIG. 5.6 – Réalisation : Résonateur BAW et partie active

variation de la fréquence de sortie avec les capacités. Il a été complété par deux modèles des dérives en fonction de la température et de la tension d'alimentation. Pour cela, nous avons simplement ajusté les paramètres de polynômes du second ordre à l'aide de mesures. Au total, on obtient un modèle direct dans lequel la fréquence de sortie dépend des conditions de réalisation, de la température, de la tension d'alimentation et des valeurs prises par les capacités de contrôle (toutes ces grandeurs sont connues ou mesurables). La caractérisation du système impose d'estimer les différents paramètres, à savoir les gains des banques de capacité et les coefficients des polynômes, et de poursuivre leur évolution éventuelle.

Pour les étapes d'estimation et de contrôle, nous développons une solution fondée sur une modélisation d'état et un filtrage de Kalman. Le problème est difficile car il est très sous déterminé (seule la fréquence de sortie est observable), mais on a une bonne idée de l'état moyen qui servira de condition initiale et des incertitudes associées qui permettront de définir la matrice de covariance a priori. Le filtre de Kalman permet alors d'ajuster les paramètres du modèle d'état et de contrôler la précision des estimées (via les matrices de covariance a posteriori). Deux modes de fonctionnement seront utilisés : une étape de calibration initiale, où l'on disposera d'une fréquence de référence externe, et une possibilité de poursuite des paramètres, en profitant de la fréquence de référence délivrée par le réseau. Pour ce faire, un dispositif de mesure de l'écart en fréquence a été développé. La partie traitement et contrôle sera développée en VHDL et implantée sur FPGA.

Chapitre 6

Synchronisation des voies dans les architectures polaires

Dans le chapitre précédent, lors de l'exposé de la nouvelle architecture d'émetteur, nous avons déjà abordé le principe des architectures EER, dites aussi polaires, qui consistent à séparer les traitements sur l'enveloppe et la phase du signal de communication.

Dans une architecture polaire, le signal d'enveloppe et le signal de phase subissant des traitements différents, une désynchronisation entre l'enveloppe et la phase peut apparaître. Cette désynchronisation est liée à la différence de temps de traitement entre les deux voies, au retard de groupe des filtres, tout autant que les petites contributions de tous les petits éléments analogiques. Cette désynchronisation entraîne une remontée du plancher de bruit et une augmentation de l'EVM (Error Vector Magnitude). Ce problème, qui n'était pas réellement considéré jusqu'à ces dernières années, devient maintenant d'autant plus crucial que la période symbole diminue, avec l'augmentation des débits [100, 101]. Cela devient un sujet de préoccupation qui est évoqué dans plusieurs publications, par exemple [102, 103, 104]. En production, même avec la connaissance globale des retards apportés par les différents éléments, les retards doivent être ajustés au cas par cas pour tenir compte de la dispersion des éléments. Par ailleurs, les caractéristiques analogiques sont affectées par des dérives dans le temps, par une variation avec la température, de la tension d'alimentation, et par un effet de vieillissement. Enfin, les solutions sont de plus en plus multimodes et les caractéristiques peuvent varier rapidement lorsque l'on passe d'un mode à un autre. On pourrait imaginer que le passage à des architectures numériques efface ce problème. En fait il n'en est rien, comme en témoigne la publication récente [105] (ISCAS - mai 2008) de K. Waheed, R. B. Staszewski and S. Rezek intitulée de manière évocatrice « *Curse of Digital Polar Transmission : Precise Delay Alignment in Amplitude and Phase Modulation Paths* ». Dans ce travail, les auteurs montrent que l'assymétrie du traitement, l'indépendance des horloges sous-jacentes et les interfaces analogiques imposent, pour satisfaire aux contraintes strictes des standardisations, un alignement extrêmement précis des deux voies. On peut noter que les courbes sur la dégradation des performances fournies dans ce papier confirment celles que nous avons données dans [106].

Dans un premier paragraphe, nous examinons l'effet d'un mauvais alignement sur les performances du système, en introduisant un modèle de simulation par technique de Monte-Carlo [107]. Dans la section, nous discutons de la faisabilité de cet alignement, avant de décrire rapidement, en section 6.3, un algorithme de correction de ces défauts [108, 106]. Enfin, dans la section 6.4, nous proposons une version sous-optimale de cet algorithme, ainsi que des extensions permettant de prendre en compte des méconnaissances du gain et de la phase en sortie de l'émetteur [109].

6.1 Effet d'un mauvais alignement

Pour des modulations à grand nombre d'états, ou à multiporteuses, le signal émis $x(t)$ peut souvent être modélisé comme gaussien complexe (et c'est bien d'ailleurs ce qui pose des problèmes car les grandes

variations d'amplitude entraînent des difficultés et des distorsions pour tous les éléments non-linéaires). Dans le cas d'un processus complexe circulaire (possédant donc un spectre passe bande symétrique autour de sa fréquence centrale et dont les composantes I et Q sont distribuées avec une loi gaussienne), il est bien connu que l'enveloppe et la phase, au même instant, sont indépendantes et distribuées selon des lois respectives Rayleigh et uniforme. Ce résultat est également vrai à deux instants différents, comme nous l'avons montré dans [110]. Ceci signifie, en clair, que pour un processus gaussien, l'enveloppe et la phase sont toujours indépendantes, sans aucune référence au coefficient de corrélation, et ce quelque soit le décalage temporel entre les deux composantes. Ainsi, les statistiques à un instant ne portent aucune information sur le décalage entre l'enveloppe et la phase. Il faudra donc au moins s'intéresser aux statistiques d'ordre 2, et c'est justement celles qui nous importent, puisqu'une des conséquences importantes d'un mauvais alignement est une remontée du plancher de bruit.

Dans un premier temps, j'ai cherché à étudier les propriétés du processus

$$x_{\Delta}(t) = \rho(t)\cos(\phi(t - \Delta)),$$

où ρ et ϕ sont l'enveloppe et la phase issus d'un processus gaussien connu. En effet, aussi bien pour la détermination analytique de la densité spectrale $S_{x_{\Delta}x_{\Delta}}(f)$ que pour les différentes propriétés des algorithmes du second ordre que l'on peut avoir à mettre en œuvre, il est utile de caractériser le signal précédent, au moins au second ordre. En fait, on voit que pour déterminer la fonction de corrélation $R_{x_{\Delta}x_{\Delta}}(\tau)$ associée à $x_{\Delta}(t)$, il faut utiliser la loi du processus gaussien aux quatre instants $\{t, t - \tau, t - \Delta, t - \tau - \Delta\}$, et l'intégrer par rapport à 4 des 8 variables après passage en coordonnées polaires. On s'aperçoit, dans mon cas après beaucoup d'efforts, mais avec la lecture fascinante de l'ouvrage de Levine [1], qu'il n'y a pas d'expression utilisable. Tout juste existe-t-il des développements en série (où les coefficients sont des intégrales de fonctions de Bessel) pour la fonction d'autocorrélation du module et pour celle de la phase. Notons qu'il est aisé de vérifier que même dans le cas $\Delta = 0$, on n'a pas $p(\rho(t), \phi(t), \rho(t - \tau), \phi(t - \tau)) = p(\rho(t), \rho(t - \tau))p(\phi(t), \phi(t - \tau))$, ce qui montre que les processus d'enveloppe et de phase ne sont pas indépendants.

Avec Antoine Diet, dont j'ai dirigé les travaux sur cette partie de sa thèse de doctorat, nous avons cherché à évaluer simplement, forcément numériquement, l'effet du décalage entre les deux voies sur la densité spectrale de sortie. Bien sûr il est possible de simuler une chaîne d'émission complète et d'évaluer, par méthode de Monte-Carlo, les performances spectrales obtenues en faisant varier tel ou tel défaut. Un premier inconvénient est que les résultats peuvent dépendre d'autres caractéristiques ou défauts du modèle de simulation. Un second inconvénient est la lourdeur du processus et le temps de simulation associé. Dans ce travail, nous avons simplement choisi de générer des vecteurs aléatoires gaussiens complexes aux 4 instants voulus, avec une structure de corrélation est connue (c'est la carré de convolution de la réponse impulsionnelle du filtre d'émission) puis de calculer la fonction de corrélation empirique associée :

$$\hat{R}_{x_{\Delta}x_{\Delta}}(\tau) = \frac{1}{K} \sum_{i=1}^K \rho_1(i)\phi_2(i)\rho_3(i)\phi_4(i),$$

où les ρ_k et ϕ_k , $k = 1 \dots 4$, sont les variables aléatoires d'enveloppe et de phase aux 4 instants. Nous avons pu également calculer la variance d'estimation, qui décroît évidemment en $1/K$, ce qui a permis d'affecter des intervalles de confiance aux résultats. Ce travail a été publié en [107], et présenté de manière plus détaillée dans le manuscrit de thèse d'Antoine Diet (Partir V). La figure Fig. 6.1 montre le comportement de la densité spectrale de puissance dans le cas d'un décalage entre voies dans le cas de l'Hiperlan2. On constate que les spécifications spectrales ne sont plus respectées dès que le décalage est supérieur ou égal à 2 ns, ce qui représente 10% de la période symbole. Il apparaît donc impératif, surtout avec la montée en fréquence, de chercher à corriger les défauts d'alignement.

6.2 Faisabilité de l'alignement entre voies

Discutons encore un instant de la faisabilité de l'alignement. Comme on l'a vu, on doit s'intéresser à une statistique à deux instants, au moins, sur la sortie. Si $z(t)$ est la sortie réelle, où l'enveloppe a subi un

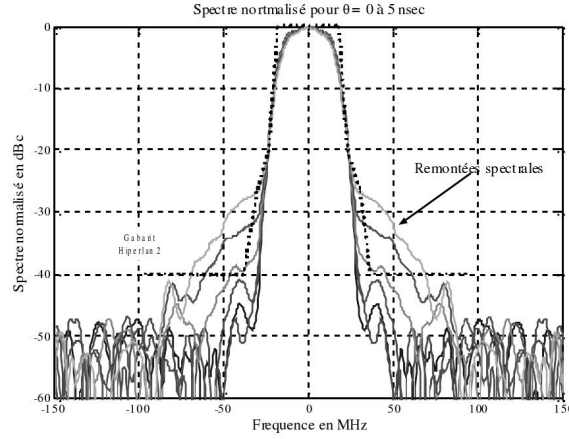


FIG. 6.1 – Remontées spectrales dans le cas d'un décalage enveloppe-phase, de 0 à 5 ns

décalage de Δ_1 et la phase un décalage de Δ_2 , la fonction d'autocorrélation s'écrit

$$R_{zz}(\tau) = E \{ \rho(t - \Delta_1) \cos(\phi(t - \Delta_2)) \rho(t - \tau - \Delta_1) \cos(\phi(t - \tau - \Delta_2)) \}, \quad (6.1)$$

dont, on l'a déjà indiqué, on n'a pas d'expression explicite. Il est donc difficile d'imaginer une procédure qui reposerait sur cette seule fonction d'autocorrélation. Cependant, si on introduit un retard μ sur la voie enveloppe, par exemple, en notant $\Delta = \Delta_2 - \Delta_1$ le retard relatif et en utilisant la stationnarité du signal,

$$R_{zz}(\tau, \mu) = E \{ \rho(t - \mu) \cos(\phi(t - \Delta)) \rho(t - \mu - \tau) \cos(\phi(t - \Delta - \tau)) \}. \quad (6.2)$$

Évidemment cette nouvelle fonction de corrélation coïncidera avec la fonction de corrélation $R_{xx}(\tau)$ du signal initial, aligné, quand $\mu = \Delta$. Dès lors, une procédure possible consiste à rechercher une valeur de μ qui minimise une distance entre $R_{zz}(\tau, \mu)$ et $R_{xx}(\tau)$, pour tout τ , par exemple

$$\int w(\tau) (R_{zz}(\tau, \mu) - R_{xx}(\tau))^2 d\tau, \quad (6.3)$$

ou $w(\tau)^2$ est une fonction de fondération, ou, de manière équivalente, dans le domaine fréquentiel,

$$\int |W(f)|^2 (S_{zz}(f, \mu) - S_{xx}(f))^2 df. \quad (6.4)$$

À partir de ces idées, on note que l'on a évidemment besoin d'un retour sur la sortie, c'est-à-dire d'une manière d'acquérir ou de calculer une quantité caractéristique à partir des données de sortie, et de comparer celle-ci à une statistique connue, ou estimée, de l'entrée (du signal aligné). On devra donc implanter une boucle de retour, qui mettra en œuvre une démodulation et numérisation d'une image du signal en sortie de l'émetteur. Notons qu'une telle boucle de commande est de toute façon souvent déjà présente, pour des problèmes de calibration et contrôle (linéarisation de l'amplificateur de puissance par exemple). En réalité, ces deux possibilités sont difficiles à mettre en œuvre pratiquement, car elles nécessitent l'estimation de sommes de fonction de corrélations, ou de spectres, (et sans doute de leurs dérivées) ce qui est lourd et statistiquement délicat. De plus l'algorithme de contrôle ne peut pas être caractérisé précisément, s'il peut être établi, du fait du manque d'expression explicite pour la corrélation.

6.3 Un algorithme de correction

La procédure que nous avons proposée, cf Fig. 6.2, consiste à comparer directement la sortie au signal numérique initial, bien sûr après démodulation et numérisation (suivant les cas par sous-échantillonnage, par démodulation simple ou en quadrature). Cette comparaison est effectuée en moyenne quadratique, et deux retards μ_1, μ_2 sont ajustés de sorte à identifier et compenser les retards inconnus Δ_1 et Δ_2 introduits par le système.

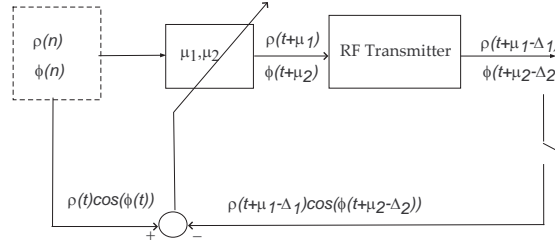


FIG. 6.2 – Principe de correction des retards.

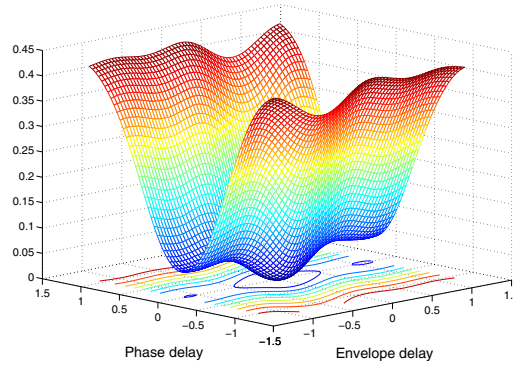
Après l'introduction des deux avances μ_1 et μ_2 , la sortie du système s'écrit $z(t) = \rho(t+\mu_1-\Delta_1) \cos(\phi(t+\mu_2-\Delta_2))$. La distance quadratique au signal $x(t)$ fournit alors le critère

$$J(\mu_1, \mu_2) = E[|x(t) - \rho(t+\mu_1-\Delta_1) \cos(\phi(t+\mu_2-\Delta_2))|^2],$$

En tenant compte de l'indépendance entre l'enveloppe et la phase à un instant, le critère se réduit simplement à

$$J(\mu_1, \mu_2) = 4R(0, 0) - 4R(\mu_1 - \Delta_1, \mu_2 - \Delta_2) \quad (6.5)$$

où $R(\tau_1, \tau_2) = E[\rho(t) \cos(\phi(t)) \rho(t - \tau_1) \cos(\phi(t - \tau_2))]$ est une sorte de fonction de corrélation. On pourra noter que par l'inégalité de Schwartz $R(0, 0) \geq |R(\tau_1, \tau_2)|$. La figure Fig. 6.3 donne l'allure du critère, obtenue par l'évaluation numérique de (6.5). On observe un minimum global clair, qui pourra être atteint par une méthode de descente, pour des retards inférieurs à une période symbole ; au delà, des minima secondaires apparaissent.

FIG. 6.3 – Critère quadratique entrée-sortie en fonction des décalages Δ_1 et Δ_2 sur les deux voies.

Nous avons proposé un premier algorithme de correction, reposant sur ce critère, dans [108]. On propose simplement de minimiser le critère à l'aide d'un algorithme de gradient stochastique, ce qui conduit à des itérations du type

$$\begin{aligned} \mu_1(n+1) &= \mu_1(n) + \gamma_1(n) \left. \frac{d\rho(u)}{du} \right|_{t_1(n)} \cos(\phi(t_2(n))) e(t) \\ \mu_2(n+1) &= \mu_2(n) + \gamma_2(n) \rho(t_1(n)) \left. \frac{d\cos\phi(u)}{du} \right|_{t_2(n)} e(t) \end{aligned} \quad (6.6)$$

avec $t_1(n) = nT_s + \mu_1(n) - \Delta_1$ et $t_2(n) = nT_s + \mu_2(n) - \Delta_2$, et $e(t) = x(t) - z(t)$.

A la suite de cette communication, nous avons été invités à compléter ce travail et soumettre une version étendue sous forme d'un article de revue [106], qui est proposé ici en annexe, pages 188 à 201. Bien

sûr, comme dans les algorithmes de synchronisation intervenant au niveau du récepteur, on doit disposer, et reconstituer au final des valeurs entre les échantillons disponibles : la procédure doit alors employer soit un taux de suréchantillonnage complètement déraisonnable, soit des dispositifs d'interpolation numérique. Le rôle, le choix et les conséquences de l'interpolateur a été étudié avec quelques détails dans l'article en question. On a également étudié un modèle approché de notre problème initial, et simplement considéré $z(t) = x(t - \Delta)$, une version retardée d'un signal original $x(t)$. Afin d'identifier et corriger le retard, on utilise

$$\mu(n+1) = \mu(n) + \gamma \left. \frac{dx(u)}{du} \right|_{u=t-\Delta+\mu(n)} [x(t) - \tilde{x}(t - \Delta + \mu(n))] \quad (6.7)$$

avec $\tilde{x}(t - \Delta + \mu(n))$ la valeur interpolée au temps $(t - \Delta + \mu(n))$. En utilisant ce modèle simplifié, on a pu établir, et on l'a vérifié numériquement sur le problème « complet », que

- le temps de convergence décroît exponentiellement avec le pas d'adaptation,
- un biais d'estimation existe, lié à l'ordre de l'interpolateur et au fait que l'erreur quadratique totale est minimum avec un peu de biais,
- la variance est une fonction (approximativement) linéaire du pas d'adaptation,
- l'EVM s'accroît linéairement avec le pas d'adaptation.

On a aussi montré que l'algorithme proposé permet d'augmenter considérablement les performances, que ce soit en terme spectral ou d'EVM. Les figures Fig. 6.4 et Fig. 6.5 en donnent une illustration.

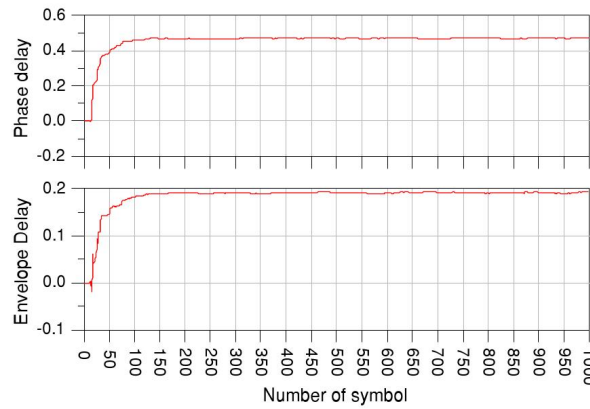


FIG. 6.4 – Convergence de μ_1 et μ_2 vers les vrai retards $\Delta_1 = 0.47T_s$ et $\Delta_2 = 0.18T_s$.

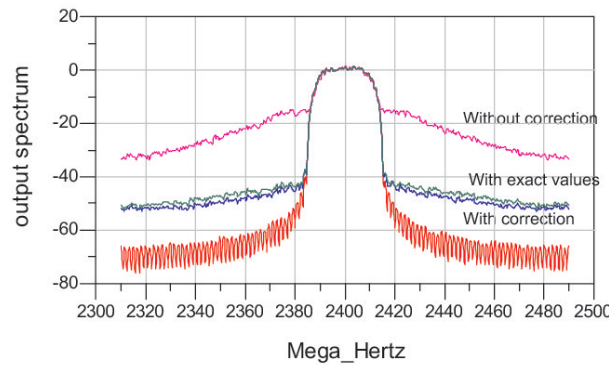


FIG. 6.5 – Comparaison des spectres idéal, non corrigé et corrigé, dans le cas de retards $\Delta_1 = 0.47T_s$ et $\Delta_2 = 0.18T_s$.

6.4 Extensions

Un défaut de la procédure précédente est que le terme d'adaptation de l'algorithme est calculé sur le signal de sortie et nécessite de disposer séparément de l'enveloppe et de la phase de la sortie. Ceci impose nécessairement de procéder à une démodulation en quadrature, puis à une conversion en représentation polaire. Dans la communication [110], nous avons considéré une procédure d'identification des retards, puis de correction a posteriori. Dans ce cas, le terme correctif dépend uniquement de l'entrée, qui par ailleurs est disponible en module et phase (on a tout fait pour). Par conséquent, on peut alors se contenter d'une démodulation simple, ou même d'une conversion directe par sous-échantillonnage, ce qui allège beaucoup le coût d'implantation.

C'est à partir de ces observations que nous avons ensuite développé une version sous-optimale de l'algorithme de correction, qui permet elle-aussi d'éviter une démodulation en quadrature dans la boucle de retour. L'idée est simplement se substituer aux quantités qui nécessiteraient la séparation de l'enveloppe et de la phase, en sortie, les mêmes quantités exprimées aux instants d'entrée (i.e. sans prendre en compte le décalage temporel) : il est clair que lorsque l'algorithme de correction approche la solution correcte $\mu_i \rightarrow \Delta_i$, on a bien sûr $t_i = t + \mu_i - \Delta_i \rightarrow t$, et on peut approximer, par exemple

$$\left. \frac{d\rho(u)}{du} \right|_{t_1} \cos \phi(t_2) \approx \left. \frac{d\rho(u)}{du} \right|_t \cos \phi(t). \quad (6.8)$$

Ceci entraîne une réduction du domaine de convergence, qui a été décrit, mais la procédure reste fonctionnelle et performante. En pratique, la sortie de l'émetteur est entachée d'un gain et d'une rotation de phase, ce qui n'a pas beaucoup d'importance habituellement, compte tenu du fait que le signal va subir d'autres avanies lors de la transmission et que le récepteur est équipé pour compenser ces défauts. Par contre, dans le cas qui nous occupe, cela a certains effets puisque l'on compare le signal de sortie à celui d'entrée. On montre analytiquement que l'argument du minimum critère est inchangé, c'est-à-dire que la procédure de correction est robuste à ces défauts. Cependant, l'erreur quadratique à l'optimum n'est plus nulle, ce qui entraîne une augmentation de la variance sur les estimées des retards. Nous avons alors développé une extension de la procédure, par l'introduction d'une étape d'identification du gain et de la phase dans la boucle de retour, qui permet de prendre en compte et compenser ces méconnaissances. Les résultats de simulation montrent une nette augmentation des performances. Nous avons aussi étudié le comportement de cette solution en contexte non stationnaire. Enfin, le problème a été simulé, et les performances mesurées, en prenant en compte les effets de quantification. Pour le détail de ces résultats, les lecteurs pourront se reporter en pages 206 à 213 où est reproduit un article actuellement soumis à *IEEE trans. on Vehicular Technologies* [109].

Nous avons d'ores et déjà prolongé ces résultats, analytiquement et en simulation, à l'architecture LINC, potentiellement très performante, mais qui confrontée aux mêmes types de désynchronisation, peut-être de manière plus aigüe encore. Rappelons très rapidement que l'idée du LINC est d'écrire le signal de départ comme la somme de deux signaux à enveloppe constante, selon

$$X(t) = \rho(t)e^{j\phi(t)} = R \cos \theta(t)e^{j\phi(t)} = R \left(e^{j(\phi(t)+\theta(t))} + e^{j(\phi(t)-\theta(t))} \right)$$

puis d'amplifier séparément les deux signaux avant de les recombinaer au niveau de l'antenne.

Du fait que l'architecture comporte une opération très non linéaire, le SCS (Signal Components Separation), il est délicat d'effectuer une correction directe et notre démarche est une démarche d'identification des retards puis de correction a posteriori. Les résultats obtenus sont excellents. Une autre perspective consistera à également prendre en charge des ondulations résiduelles du gain dans la bande passante (ce qui revient en fait à remplacer la correction du retard par un égaliseur, mais au niveau de l'émetteur).

Par ces travaux, on confirme que la conception et l'optimisation des émetteurs numériques peut ainsi bénéficier de l'apport d'outils de traitement du signal pour l'optimisation de la "dirty RF" : c'est le point de vue que nous avons présenté dans [111] (Conférence invitée - C. Berland).

Chapitre 7

Points de suspension

Comme je l'écrivais en préambule, la présentation de ce mémoire n'est ni une finalité ni un achèvement mais simplement un signe de ponctuation dans une histoire que je souhaite plus longue. J'ai montré dans les pages précédentes que j'ai développé des recherches suivant deux sillons parallèles dans ma géométrie personnelle. Le premier correspond à mes intérêts historiques en théorie et traitement de l'information, avec en particulier l'étude et la caractérisation d'entropies et autres mesures d'information, et leurs applications à divers problèmes. Le second axe concerne les apports du traitement du signal pour la conception et l'évaluation des émetteurs récepteurs de communication. Des perspectives naturelles se dégagent sur les deux axes.

Comme j'ai débuté le mémoire par la partie entropies, je terminerai, en forme de boucle, en envisageant d'abord les perspectives en radiotechnique.

L'un des points de départ de ces travaux menés avec C. Berland est la recherche d'une structure d'émetteur universel, capable de prendre en charge des signaux de nature différente. La première idée étant ainsi de s'intéresser aux structures décomposant les signaux en deux composantes d'enveloppe et de phase. C'est sur cette structure que l'on a proposé des évolutions et des dispositifs de correction afin d'en améliorer les performances. Il est possible de développer d'autres méthodes de compensation des défauts. Une idée fort simple, et qu'il faudra évaluer prochainement, analytiquement et en pratique, est de chercher à minimiser la puissance, et donc les remontées spectrales dans une bande d'intérêt. Une autre perspective consistera, en plus des corrections de retards, à également prendre en charge des ondulations résiduelles du gain dans la bande passante, ce qui revient en fait à intégrer un égaliseur, qu'il faudra imaginer léger, au niveau de l'émetteur. Une autre prolongation immédiate est de considérer l'architecture LINC qui consiste elle-aussi à décomposer le problème difficile en deux problèmes plus simples, mais qui souffre également de problèmes de désynchronisation. Il est tout-à-fait envisageable que les structures d'algorithme de correction que nous avons étudiées puissent apporter ici des résultats intéressants. En poursuivant sur cette idée de simplifier la réalisation, il serait également intéressant de rechercher d'autres types de représentation des signaux, par décompositions sur des ensemble de fonctions, ou par des dispositifs de codage simple qui permettraient d'envisager d'autres structures de calibration au niveau de l'émetteur. Le cheminement des signaux comprend des non-linéarités, notamment l'étalement de puissance, auxquelles je souhaite m'intéresser, tout en étant bien conscient que le sujet a été largement exploré, mais tout le monde n'a pas la chance de pouvoir étudier [1]. Toujours au niveau de l'émetteur, les travaux menés avec P. Guillot concernant la réalisation d'une fréquence de référence à très haute fréquence nous conduisent à l'étude d'un problème direct, et au développement de solutions d'estimation et de contrôle à base de modélisation d'état et de filtrage de Kalman. À trop promouvoir l'intégration du traitement du signal tôt dans les traitements, on se retrouve à chercher à numériser au plus tôt, éventuellement avec des défauts qui devront être compensés par la suite. C'est sur cette problématique que débute un autre travail avec A. Lesellier, sur laquelle on envisage des solutions à base de bancs de filtres hybrides analogique/numérique, avec des contraintes sur la réalisabilité des solutions. D'autres possibilités, comme l'utilisation de résultats sur l'échantillonnage non-uniforme demandent à être explorées.

Du côté des entropies, on a des possibilités immédiates de prolongation des travaux exposés. Il ne serait pas difficile par exemple d'exploiter les fonctionnelles entropiques associées à la divergence de Rényi dans le cadre des problèmes inverses. Sur ce sujet, il serait intéressant d'imaginer des algorithmes de relaxation sur le paramètre α . Sur ce thème, un autre travail à reprendre est celui qui concerne la technique de maximum d'entropie itérative exposée en §2.2. On a noté, dans le paragraphe §3.4 que pour les distributions de Tsallis que nous avons étudiées, la transformation paramètre-moyenne n'est pas bijective, ce qui peut mener à des interprétations sympathiques de type transition de phase. Une étude précise sur la caractérisation du phénomène est à compléter. Le rôle de la contrainte de log-vraisemblance moyen, que nous avons utilisé dans [29], cf. §3.2, notamment le lien avec le codage, est encore un point à explorer.

Dans les constructions que nous avons présentées apparaît souvent l'"escort distribution" associée à la distribution initiale, et il est possible que celle-ci joue un rôle fondamental dans ces constructions. J'envisage ainsi de me pencher plus précisément sur cette correspondance entre distributions et sur les propriétés de ces escort distribution, qui peuvent être éventuellement à relier aux notions de distributions incomplètes utilisées par Rényi, cf l'excellent livre de Pierre Hammad [112]. Il est possible par exemple de s'intéresser à l'information de Fisher attachée aux escorts, qui semble présenter quelques caractéristiques intéressantes. On pourra noter avec intérêt que la question de la justification des escort distributions fait partie des quelques questions ouvertes relevées dans [37]. L'existence d'une géométrie particulière sous-jacente à la famille des exponentielles déformées, à l'image de la géométrie des familles exponentielles, est également un sujet d'intérêt. La divergence de Bregman déformée que l'on a obtenue est un premier élément. Je souhaite aussi continuer à explorer les domaines d'apparition des entropies et distributions de Rényi-Tsallis et en extraire les raisons ou motivations sous-jacentes. Il serait ainsi intéressant d'explorer les connections avec les théories des multifractales. Les lois d'échelle, et sous une certaine forme les lois de Tsallis, apparaissent également dans certains processus multiplicatifs.

Bibliographie

- [1] B. R. Levin, *Fondements théoriques de la radiotechnique statistique*. Moscow : Éditions Mir, 1979, vol. 1.
- [2] L. Bercher, *L'homœopathie sans mystère*. Paris : G. Doin & Cie, 1937.
- [3] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, pp. 620–630, May 1957.
- [4] E. T. Jaynes, "Information theory and statistical mechanics. II," *Physical Review*, vol. 108, no. 2, pp. 171–190, Oct. 1957.
- [5] G. L. Besnerais, J.-F. Bercher, and G. Demoment, "A new look at entropy for solving linear inverse problems," *Information Theory, IEEE Transactions on*, vol. 45, no. 5, pp. 1565–1578, 1999.
- [6] P. Holejšovská and J.-F. Bercher, "Elimination of the measurement system impact on the measured data by maximum entropy on the mean method," Plzen, Sep. 2003, pp. 192–195.
- [7] P. Holejšovská, Z. Peroutka, J. Cengery, and J.-F. Bercher, "Continuous Non-Invasive blood pressure measuring based on the reconstruction by maximum entropy on the mean method," in *17-th Biennial International Eurasip Conference Proceedings*. Brno : Jirí Jan, Jirí-Kozumplík, and Ivo Provazník, Jun. 2004, pp. 121–123.
- [8] J.-F. Bercher, "Maximum entropy with fluctuating constraints - The example of K-distributions," *Physics Letters A*, vol. 372, no. 24, pp. 4361–4363, Jun. 2008. [PDF](#)
- [9] J.-F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1687–1694, 2000.
- [10] J.-F. Bercher and C. Vignat, "A renyi entropy convolution inequality with application," in *EUSIP-CO' 2002, XI European Signal Processing Conference*, Toulouse, France., sep 2002, pp. 111–114.
- [11] H. Gzyl, "Ill-posed linear inverse problems and maximum entropy in the mean," *Acta Científica Venezolana*, vol. 53, no. 2, pp. 74–93, 2002.
- [12] F. Gamboa and E. Gassiat, "Bayesian methods and maximum entropy for ill-posed inverse problems," *The Annals of Statistics*, pp. 328–350, 1997.
- [13] P. Holejšovská, Z. Peroutka, and J.-F. Bercher, "Adaptive signal processing using maximum entropy on the mean method and monte carlo analysis," in *MATLAB 2003*, 2003, pp. 191–196.
- [14] G. Wilk and Z. Włodarczyk, "Interpretation of the nonextensivity parameter q in some applications of Tsallis statistics and Lévy distributions," *Phys.Rev.Lett.*, vol. 84, p. 2770, Aug. 2000.
- [15] C. Beck and E. G. D. Cohen, "Superstatistics," *Physica A*, vol. 322, pp. 267–275, May 2003.
- [16] F. Sattin, "Non-Gaussian probability distribution functions from maximum-entropy-principle considerations," *Physical Review E*, vol. 68, no. 3, p. 32102, 2003.
- [17] H. Touchette and C. Beck, "Asymptotics of superstatistics," *Physical Review E (Statistical, Non-linear, and Soft Matter Physics)*, vol. 71, no. 1, pp. 016 131–6, 2005.
- [18] B. A. Maguire, E. S. Pearson, and H. A. Wynn, "The time intervals between industrial accidents," *Biometrika*, vol. 39, pp. 168–180, 1952.

- [19] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. van der Meulen, “Nonparametric entropy estimation : An overview,” *International Journal of Mathematical and Statistical Sciences*, vol. 6, pp. 17–39, 1997.
- [20] J.-F. Bercher and C. Vignat, “Estimating the entropy of a signal with applications,” in *Proceedings of the .IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. ICASSP '99*, vol. 3, 1999, pp. 1705–1708 vol.3.
- [21] C. Vignat and J.-F. Bercher, “Un estimateur récursif de l’entropie,” in *17e Colloque sur le traitement du signal et des images*. Vannes : GRETSI., 1999, pp. 701–704. [PDF](#)
- [22] K. E. Hild, D. Erdogmus, and J. Principe, “Blind deconvolution with rényi’s minimum entropy,” in *Proceedings EUSIPCO 2002*, 2002, pp. 557–560.
- [23] D. Erdogmus, K. E. Hild, J. C. Principe, M. Lazaro, and I. Santamaria, “Adaptive blind deconvolution of linear channels using renyi’s entropy with parzen window estimation,” *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1489–1498, 2004.
- [24] D. Han, S. Kim, and J. Principe, “Sparse channel estimation with regularization method using convolution inequality for entropy,” in *Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on*, vol. 4, 2005, pp. 2359–2362 vol. 4.
- [25] S. Cruces and I. Duran, “The minimum support criterion for blind signal extraction : A limiting case of the strengthened young’s inequality,” *Lecture notes in computer science*, pp. 57–64, 2004.
- [26] S. Cruces, A. Sarmiento, and I. Duran, “The complex version of the minimum support criterion,” *Lecture Notes in Computer Science*, vol. 4666, p. 17, 2007.
- [27] D. T. Pham, F. Vrins, and M. Verleysen, “On the risk of using rényi’s entropy for blind source separation,” *IEEE Transactions on Signal Processing*, vol. 56, no. 10 Part 1, pp. 4611–4620, 2008.
- [28] M. E. J. Newman, “Power laws, Pareto distributions and Zipf’s law,” *Contemporary Physics*, vol. 46, no. 5, pp. 323–351, 2005.
- [29] J.-F. Bercher, “An amended MaxEnt formulation for deriving Tsallis factors, and associated issues,” in *AIP Conference Proceedings - Twenty sixth International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, A. Mohammad-Djafari, Ed., vol. 872, Nov. 2006, pp. 441–448. [PDF](#)
- [30] J.-F. Bercher, “Tsallis distribution as a standard maximum entropy solution with ‘tail’ constraint,” *Physics Letters A*, vol. 372, no. 35, pp. 5657–5659, Aug. 2008. [PDF](#)
- [31] J.-F. Bercher and C. Vignat, “A new look at q -exponential distributions via excess statistics,” *Physica A*, vol. 387, no. 22, pp. 5422–5432, Sep. 2008. [PDF](#)
- [32] J. F. Bercher and C. Vignat, “An entropic view of Pickands’ theorem,” in *International Symposium on Information Theory ISIT2008*, Toronto, Jul. 2008. [PDF](#)
- [33] J.-F. Bercher, “On some entropy functionals derived from Rényi information divergence,” *Information Sciences*, vol. 178, pp. 2489–2506, Jun. 2008. [PDF](#)
- [34] A. Rényi, “On measures of entropy and information,” in *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*. Berkeley, Calif. : Univ. California Press, 1961, pp. 547–561.
- [35] L. L. Campbell, “A coding theorem and renyi’s entropy,” *Information and Control*, vol. 8, no. 4, p. 423 7429, 1965.
- [36] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *Journal of Statistical Physics*, vol. 52, no. 1, pp. 479–487, Jul. 1988.
- [37] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics*, illustrated edition ed. Springer, Apr. 2009.
- [38] M. Grendar and M. Grendar, “Maximum probability and maximum entropy methods : bayesian interpretation,” in *AIP Conference Proceedings*, vol. 707. IOP Institute of Physics publishing ltd, 2004, pp. 490–494.

- [39] J.-F. Bercher, "Possible rationales for Rényi-Tsallis entropy maximization," in *International Workshop on Applied Probabilities (IWAP2008)*, Compiègne, Jul. 2008. [PDF](#)
- [40] S. Kullback, *Information Theory and Statistics*. Wiley, New York, 1959.
- [41] M.-P. Schützenberger, *Contributions aux applications statistiques de la théorie de l'information, thèse de doctorat*, ser. Institut de Statistique de l'Université de Paris. Institut de Statistique de l'Université de Paris, 1954, vol. 3.
- [42] A. Wald, "Sequential tests of statistical hypotheses," *Annals of Mathematical Statistics*, vol. 16, no. 2, p. 117–186, Jun. 1945.
- [43] I. Csiszar, "Generalized cutoff rates and renyi's information measures," *Information Theory, IEEE Transactions on*, vol. 41, no. 1, pp. 26–34, 1995.
- [44] B. L. Cour, "Statistical characterization of active sonar reverberation using extreme value theory," *Oceanic Engineering, IEEE Journal of*, vol. 29, no. 2, pp. 310–316, 2004.
- [45] J. Pickands, "Statistical inference using extreme order statistics," *The Annals of Statistics*, vol. 3, pp. 119–131, 1975.
- [46] A. A. Balkema and L. de Haan, "Residual life time at great age," *The Annals of Probability*, vol. 2, pp. 792–804, Oct. 1974.
- [47] N. Markovich, *Nonparametric Analysis of Univariate Heavy-Tailed Data*, ser. Series in Probability and Statistics. Wiley, 2007.
- [48] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *0706.1062*, Jun. 2007.
- [49] B. R. Frieden, "Fisher information, disorder, and the equilibrium distributions of physics," *Physical Review A*, vol. 41, no. 8, pp. 4265–4276, 1990.
- [50] B. R. Frieden and P. M. Binder, "Physics from Fisher information : A unification," *American Journal of Physics*, vol. 68, p. 1064, 2000.
- [51] B. R. Frieden, *Science from Fisher Information : A Unification*. Cambridge University Press, 2004.
- [52] M. T. Martin, F. Pennini, and A. Plastino, "Fisher's information and the analysis of complex signals," *Physics Letters A*, vol. 256, no. 2-3, pp. 173–180, 1999.
- [53] A. Nagy, "Fisher information in density functional theory," *The Journal of Chemical Physics*, vol. 119, no. 18, pp. 9401–9405, Nov. 2003.
- [54] L. Telesca, V. Lapenna, and M. Lovallo, "Fisher information analysis of earthquake-related geoelectrical signals," *Natural Hazards and Earth System Sciences*, vol. 5, no. 4, pp. 561–564, 2005.
- [55] K. D. Sen, J. Antolín, and J. C. Angulo, "Fisher-Shannon analysis of ionization processes and isoelectronic series," *Physical Review A*, vol. 76, no. 3, p. 32502, 2007.
- [56] M. Balasco, V. Lapenna, M. Lovallo, G. Romano, A. Siniscalchi, and L. Telesca, "Fisher information measure analysis of earth's apparent resistivity," *International Journal of Nonlinear Science*, vol. 5, no. 3, pp. 230–236, 2008.
- [57] S. A. Frank, "Natural selection maximizes Fisher information," *Journal of Evolutionary Biology*, vol. 22, no. 2, pp. 231–244, 2009.
- [58] J. M. Borwein, A. S. Lewis, M. N. Limber, and D. Noll, "Maximum entropy reconstruction using derivative information part 2 : computational results," *Numerische Mathematik*, vol. 69, no. 3, pp. 243–256, 1995.
- [59] J. M. Borwein, A. S. Lewis, and D. Noll, "Maximum entropy reconstruction using derivative information, part 1 : Fisher information and convex duality," *Mathematics of Operations Research*, vol. 21, pp. 442–468, 1996.
- [60] V. Zivojnovic and D. Noll, "Minimum Fisher information spectral analysis," in *ICASSP-97, IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, 1997, pp. 3957–3960.

- [61] V. Zivojnovic, "Minimum Fisher information of moment-constrained distributions with application to robust blind identification," *Signal Processing*, vol. 65, no. 2, pp. 297–313, Mar. 1998.
- [62] C. Vignat and J.-F. Bercher, "On Fisher information inequalities and score functions in non-invertible linear systems," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 4, no. 4, p. Article 71, 2003. [PDF](#)
- [63] C. Vignat and J. F. Bercher, "Analysis of signals in the Fisher-Shannon information plane," *Physics Letters A*, vol. 312, no. 1-2, pp. 27–33, Jun. 2003. [PDF](#)
- [64] J.-F. Bercher and C. Vignat, "A note on minimum Fisher information distributions with restricted support and fixed variance," *Submitted to Information Sciences*, Mar. 2009. [PDF](#)
- [65] A. Dembo, T. Cover, and J. Thomas, "Information theoretic inequalities," *Information Theory, IEEE Transactions on*, vol. 37, no. 6, pp. 1501–1518, 1991.
- [66] N. Blachman, "The convolution inequality for entropy powers," *Information Theory, IEEE Transactions on*, vol. 11, no. 2, pp. 267–271, 1965.
- [67] A. Stam, "Some inequalities satisfied by the quantities of information of fisher and shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959.
- [68] V. Papathanasiou, "Some characteristic properties of the fisher information matrix via Cacoullos-Type inequalities," *Journal of Multivariate Analysis*, vol. 44, no. 2, pp. 256–265, Feb. 1993.
- [69] R. Zamir and M. Feder, "A generalization of the entropy power inequality with applications," *IEEE transactions on Information Theory*, vol. 39, pp. 1723–1728, 1993.
- [70] R. Zamir, "A proof of the fisher information inequality via a data processing argument," *IEEE transactions on Information Theory*, vol. 44, pp. 1246–1250, 1998.
- [71] R. Zamir, "A necessary and sufficient condition for equality in the matrix Fisher-Information-Inequality," Tech. Rep. EE-S-97-11, 1997.
- [72] C. Vignat and J.-F. Bercher, "Matrix fisher inequalities for non-invertible linear systems," in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, 2002, p. 237.
- [73] E. Romera and J. S. Dehesa, "The Fisher-Shannon information plane, an electron correlation tool," *The Journal of Chemical Physics*, vol. 120, no. 19, pp. 8906–12, May 2004.
- [74] J. Szabó, K. Sen, and A. Nagy, "The Fisher-Shannon information plane for atoms," *Physics Letters A*, vol. 372, no. 14, pp. 2428–2430, Mar. 2008.
- [75] J. Sanudo and R. Lopez-Ruiz, "Alternative evaluation of statistical indicators in atoms : the non-relativistic and relativistic cases," *0904.1144*, Apr. 2009.
- [76] H. M. Jr. and K. Sen, "Statistical complexity and Fisher-Shannon information measure of $h+2$," *Physics Letters A*, vol. 372, no. 13, pp. 2271–2273, Mar. 2008.
- [77] J. Antolín and J. C. Angulo, "Complexity analysis of ionization processes and isoelectronic series," *International Journal of Quantum Chemistry*, vol. 109, no. 3, pp. 586–593, 2009.
- [78] J. Dehesa, P. Sánchez-Moreno, and R. Y. nez, "Cramer-Rao information plane of orthogonal hypergeometric polynomials," *Journal of Computational and Applied Mathematics*, vol. 186, no. 2, pp. 523–541, Feb. 2006.
- [79] E. Romera and A. Nagy, "Fisher-Rényi entropy product and information plane," *Physics Letters A*, vol. 372, no. 46, pp. 6823–6825, Nov. 2008.
- [80] G. L. Ferri, F. Pennini, and Plastino, "LMC complexity and various cahotic regimes," *Physics Letters A*, vol. To Appear, 2009.
- [81] B. R. Frieden, *Physics from Fisher Information*. Cambridge University Press, 1998.
- [82] R. B. Frieden and R. A. Gatenby, *Exploratory Data Analysis Using Fisher Information*, 1st ed. Springer, Nov. 2006.
- [83] E. Uhrmann Klingens, "Minimal Fisher information distributions with compact-supports," *Sankhya : The Indian Journal of Statistics*, vol. 57, no. 3, pp. 360–374, 1995.

- [84] E. Uhrmann-Klingen, “Fisher-Minimale Dichten auf Kompacten intervallen (in german),” Ph.D. dissertation, Essen, 1992.
- [85] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed. Cambridge University Press, 1927.
- [86] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, Aug. 1991.
- [87] P. Hammad, “Mesure d’ordre α de l’information au sens de fisher,” *Revue de Statistique Appliquée*, vol. 26, no. 1, pp. 73–84, 1978.
- [88] J. Rufin, *Globalia*. Editions Gallimard, Jun. 2005.
- [89] K. Iniewski, *Wireless Technologies : Circuits, Systems, and Devices*, 1st ed. CRC, Oct. 2007.
- [90] L. Kahn, “Single-Sideband transmission by envelope elimination and restoration,” *Proceedings of the IRE*, vol. 40, no. 7, pp. 803–806, 1952.
- [91] C. Berland, J.-F. Bercher, I. Hibon, M. Villegas, D. Belot, D. Pache, V. Le Goasoz, “US Patent US2006034391A1 : Digital Transmitter Architecture,” Feb. 2006. [PDF](#)
- [92] R. B. Staszewski and P. T. Balsara, *All-Digital Frequency Synthesizer in Deep-Submicron CMOS*. Wiley-Interscience, Sep. 2006.
- [93] R. Staszewski, J. Wallberg, S. Rezeq, C. Hung, O. Eliezer, S. Vemulapalli, C. Fernando, K. Maggio, R. Staszewski, N. Barton, M. Lee, P. Cruise, M. Entezari, K. Muhammad, and D. Leipold, “All-digital PLL and transmitter for mobile phones,” *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 12, pp. 2469–2482, 2005.
- [94] C. Joubert, J.-F. Bercher, G. Baudoin, T. Divel, S. Ramet, and P. Level, “Time behavioral model for phase-domain ADPLL based frequency synthesizer,” in *Radio and Wireless Symposium, 2006 IEEE*, 2006, pp. 167–170. [PDF](#)
- [95] C. Joubert, J.-F. Bercher, and G. Baudoin, “Contributions to the analysis and design of an ADPLL,” in *13th IEEE International Conference on Electronics, Circuits and Systems, 2006. (ICECS) 2006.*, 2006, pp. 322–325. [PDF](#)
- [96] G. Baudoin, J.-F. Bercher, and C. Joubert, “Calcul théorique du bruit de phase dû au dithering dans un oscillateur contrôlé numériquement,” in *Journées Nationales Microondes 2007*, may 2007. [PDF](#)
- [97] R. Staszewski, J. Wallberg, C. Hung, G. Feygin, M. Entezari, and D. Leipold, “LMS-based calibration of an RF digitally controlled oscillator for mobile phones,” *Circuits and Systems II : Express Briefs, IEEE Transactions on*, vol. 53, no. 3, pp. 225–229, 2006.
- [98] P. Guillot, C. Berland, J.-F. Bercher, and P. Philippe, “A 2GHz 65nm CMOS digitally-tuned BAW oscillator,” in *International Conference on Electronics Circuits and Systems (ICECS2008)*, Malta, Sep. 2008. [PDF](#)
- [99] P. Guillot, P. Philippe, C. Berland, and J.-F. Bercher, “Faisabilité de référence haute fréquence pour les architectures RF,” in *Journées Nationales Microondes 2009*, may 2009. [PDF](#)
- [100] J. Groe, “Polar transmitters for wireless communications,” *IEEE Communications Magazine*, vol. 45, pp. 58–63, 2007.
- [101] J. Groe, “A multimode cellular radio,” *IEEE Transactions on Circuits and Systems II : Express Briefs*, vol. 55, pp. 269–273, 2008.
- [102] M. Markovic and H. Modi, “Feasibility of EER transmitters for 3GPP applications,” in *IEEE Annual Wireless and Microwave Technology Conference, WAMICON '06*, 2006, pp. 1–2.
- [103] W.-F. Loke, M.-W. Chia, and P.-Y. Chee, “Design considerations for multi-band OFDM polar transmitter of UWB system,” *Electronics Letters*, vol. 43, 2007.
- [104] B. Priyanto, T. Sorensen, O. Jensen, T. Larsen, T. Kolding, and P. Mogensen, “Impact of polar transmitter imperfections on UTRA LTE uplink performance,” in *Norchip, 2007*, 2007, pp. 1–4.

- [105] K. Waheed, R. B. Staszewski, and S. Rezek, "Curse of digital polar transmission : Precise delay alignment in amplitude and phase modulation paths," in *Circuits and Systems, 2008. ISCAS 2008. IEEE International Symposium on*, 2008, pp. 3142–3145.
- [106] J.-F. Bercher and C. Berland, "Envelope and phase delays correction in an EER radio architecture," *Analog Integrated Circuits and Signal Processing*, vol. 55, pp. 21–35, Apr. 2008. [PDF](#)
- [107] J. Bercher, A. Diet, C. Berland, G. Baudoin, and M. Villegas, "Monte-Carlo estimation of time mismatch effect in an OFDM EER architecture," in *Radio and Wireless Conference, 2004 IEEE*, 2004, pp. 283–286.
- [108] J.-F. Bercher and C. Berland, "Envelope/phase delays correction in an EER radio architecture," in *Electronics, Circuits and Systems, 2006. ICECS '06. 13th IEEE International Conference on*, 2006, pp. 443–446. [PDF](#)
- [109] J.-F. Bercher and C. Berland, "Adaptive delays alignment in polar transmitter architecture," *submitted to IEEE trans. on Vehicular Technologies*, 2009. [PDF](#)
- [110] J.-F. Bercher and C. Berland, "Adaptive time mismatches identification and correction in polar transmitter architecture," in *Wireless Technologies, 2007 European Conference on*, 2007, pp. 78–81.
- [111] C. Berland, J.-F. Bercher, and O. Venard, "Digital signal processing techniques to compensate for RF imperfections in advanced transmitter architectures," in *European Wireless Technology Conference (EuWit2008)*, Amsterdam, Oct. 2008. [PDF](#)
- [112] P. Hammad, *Information, systèmes et distribution*. Editions Cujas, Jul. 1987.

Chapitre 8

Recueil de publications

- 8.1 J.-F. Bercher, “Maximum entropy with fluctuating constraints - The example of K-distributions,” *Physics Letters A*, vol. 372, no. 24, pp. 4361–4363, Jun. 2008.**

J.-F. Bercher, “Maximum entropy with fluctuating constraints - The example of K-distributions,” *Physics Letters A*, vol. 372, no. 24, pp. 4361–4363, Jun. 2008.