



HAL
open science

Fonction d'autocorrélation partielle et estimation autorégressive dans le domaine temporel

Serge Dégerine

► **To cite this version:**

Serge Dégerine. Fonction d'autocorrélation partielle et estimation autorégressive dans le domaine temporel. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1988. Français. NNT : . tel-00243761

HAL Id: tel-00243761

<https://theses.hal.science/tel-00243761>

Submitted on 7 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THESE

présentée à

l'Université Joseph Fourier - Grenoble I

pour obtenir le grade de
DOCTEUR ES SCIENCES
"Mathématiques"

par

DEGERINE Serge

FONCTION D'AUTOCORRELATION PARTIELLE
ET
ESTIMATION AUTOREGRESSIVE DANS LE DOMAINE TEMPOREL

Thèse soutenue le 12 juillet 1988 devant la commission d'examen

B. VAN CUTSEM	Président
C. GUEGUEN	
M. HALLIN	
A. LE BRETON	Examineurs
G. OPPENHEIM	
D. T. PHAM	

Thèse préparée au sein du Laboratoire TIM 3, Institut IMAG

UNIVERSITE Joseph FOURIER (GRENOBLE I)

Président de l'Université :
M. PAYAN Jean Jacques

Année Universitaire 1987 - 1988

MEMBRES DU CORPS ENSEIGNANT DE SCIENCES ET DE GEOGRAPHIE

PROFESSEURS DE 1ère Classe

ARNAUD Paul	Chimie Organique
ARVIEU ROBERT	Physique Nucléaire I.S.N.
AUBERT Guy	Physique C.N.R.S
AURIAULT Jean-Louis	Mécanique
AYANT Yves	Physique Approfondie
BARBIER Marie-Jeanne	Electrochimie
BARJON Robert	Physique Nucléaire ISN
BARNOUD Fernand	Biochimie Macromoléculaire Végétale
BARRA Jean-René	Statistiques-Mathématiques Appliquées
BECKER Pierre	Physique
BEGUIN Claude	Chimie Organique
BELORISKY Elie	Physique
BENZAKEN Claude	Mathématiques Pures
BERARD Pierre	Mathématiques Pures
BERNARD Alain	Mathématiques Pures
BERTRANDIAS Françoise	Mathématiques Pures
BERTRANDIAS Jean-Paul	Mathématiques Pures
BILLET Jean	Géographie
BOELHER Jean-Paul	Mécanique
BONNIER Jane Marie	Chimie Générale
BOUCHEZ Robert	Physique Nucléaire ISN
BRAVARD Yves	Géographie
CARLIER Georges	Biologie Végétale
CAUQUIS Georges	Chimie Organique
CHARDON Michel	Géographie
CHIBON Pierre	Biologie Animale
COHEN ADDAD Jean-Pierre	Physique
COLIN DE VERDIERE Yves	Mathématiques Pures
CYROT Michel	Physique du Solide
DEBELMAS Jacques	Géologie Générale
DEGRANGE Charles	Zoologie
DEMAILLY Jean-Pierre	Mathématiques Pures
DENEUVILLE Alain	Physique
DEPORTES Charles	Chimie Minérale
DOLIQUE Jean-Michel	Physique des Plasmas
DOUCE Roland	Physiologie Végétale
DUCROS Pierre	Cristallographie
FONTAINE Jean-Marc	Mathématiques Pures
GAGNAIRE Didier	Chimie Physique
GERMAIN Jean-Pierre	Mécanique,
GIRAUD Pierre	Géologie
HICTER Pierre	Chimie
IDELMAN Simon	Physiologie Animale
JANIN Bernard	Géographie
JOLY Jean-René	Mathématiques Pures
KAHANE André, détaché	Physique
KAHANE Josette	Physique
KRAKOWIAK Sacha	Mathématiques Appliquées

LAJZEROWICZ Jeanine
 LAJZEROWICZ Joseph
 LAURENT Pierre-Jean
 LEBRETON Alain
 DE LEIRIS Joël
 LHOMME Jean
 LLIBOUTRY Louis
 LOISEAUX Jean-Marie
 LUNA Domingo
 MACHE Régis
 MASCLE Georges
 MAYNARD Roger
 OMONT Alain
 OZENDA Paul
 PAYAN Jean-Jacques
 PEBAY-PEYROULA Jean-Claude
 PERRIER Guy
 PIERRARD Jean-Marie
 PIERRE Jean-Louis
 RENARD Michel
 RINAUDO Marguerite
 ROSSI André
 SAXOD Raymond
 SENDEL Philippe
 SERGERAERT Francis
 SOUCHIER Bernard
 SOUTIF Michel
 STUTZ Pierre
 TRILLING Laurent
 VALENTIN Jacques
 VAN CUTSEM Bernard
 VIALON Pierre

Physique
 Physique
 Mathématiques Appliquées
 Mathématiques Appliquées
 Biologie
 Chimie
 Géophysique
 Sciences Nucléaires I.S.N.
 Mathématiques Pures
 Physiologie Végétale
 Géologie
 Physique du Solide
 Astrophysique
 Botanique (Biologie Végétale)
 Mathématiques Pures
 Physique
 Géophysique
 Mécanique
 Chimie Organique
 Thermodynamique
 Chimie CERMAV
 Biologie
 Biologie Animale
 Biologie Animale
 Mathématiques Pures
 Biologie
 Physique
 Mécanique
 Mathématiques Appliquées
 Physique Nucléaire I.S.N.
 Mathématiques Appliquées
 Géologie

PROFESSEURS de 2^{ème} Classe

ADIBA Michel
 ANTOINE Pierre
 ARMAND Gilbert
 BARET Paul
 BLANCHI J.Pierre
 BLUM Jacques
 BOITET Christian
 BORNAREL Jean
 BRUANDET J.François
 BRUGAL Gérard
 BRUN Gilbert
 CASTAING Bernard
 CERFF Rudiger
 CHIARAMELLA Yves
 COURT Jean
 DUFRESNOY Alain
 GASPARD François
 GAUTRON René
 GENIES Eugène
 GIDON Maurice
 GIGNOUX Claude
 GILLARD Roland
 GIORNI Alain
 GONZALEZ SPRINBERG Gérardo
 GUIGO Maryse
 GUMUCHAIN Hervé
 GUITTON Jacques

Mathématiques Pures
 Géologie
 Géographie
 Chimie
 STAPS
 Mathématiques Appliquées
 Mathématiques Appliquées
 Physique
 Physique
 Biologie
 Biologie
 Physique
 Biologie
 Mathématiques Appliquées
 Chimie
 Mathématiques Pures
 Physique
 Chimie
 Chimie
 Géologie
 Sciences Nucléaires
 Mathématiques Pures
 Sciences Nucléaires
 Mathématiques Pures
 Géographie
 Géographie
 Chimie

HACQUES Gérard
 HERBIN Jacky
 HERAULT Jeanny
 JARDON Pierre
 JOSELEAU Jean-Paul
 KERCKHOVE Claude
 LONGEQUEUE Nicole
 LUCAS Robert
 MANDARON Paul
 MARTINEZ Francis
 NEMOZ Alain
 OUDET Bruno
 PECHER Arnaud
 PELMONT Jean
 PERRIN Claude
 PFISTER Jean-Claude
 PIBOULE Michel
 RAYNAUD Hervé
 RICHARD Jean-Marc
 RIEDTMANN Christine
 ROBERT Gilles
 ROBERT Jean-Bernard
 SARROT-REYNAULD Jean
 SAYETAT Françoise
 SERVE Denis
 STOECKEL Frédéric
 SCHOLL Pierre-Claude
 SUBRA Robert
 VALLADE Marcel
 VIDAL Michel
 VIVIAN Robert
 VOTTERO Philippe

Mathématiques Appliquées
 Géographie
 Physique
 Chimie
 Biochimie
 Géologie
 Sciences Nucléaires I.S.N.
 Physique
 Biologie
 Mathématiques Appliquées
 Thermodynamique CNRS - CRTBT
 Mathématiques Appliquées
 Géologie
 Biochimie
 Sciences Nucléaires I.S.N.
 Physique du Solide
 Géologie
 Mathématiques Appliquées
 Physique
 Mathématiques Pures
 Mathématiques Pures
 Chimie Physique
 Géologie
 Physique
 Chimie
 Physique
 Mathématiques Appliquées
 Chimie
 Physique
 Chimie Organique
 Géographie
 Chimie

MEMBRES DU CORPS ENSEIGNANT DE L' IUT 1

PROFESSEURS de 1^{ère} Classe

BUISSON Roger	Physique IUT 1
DODU Jacques	Mécanique Appliquée IUT 1
NEGRE Robert	Génie Civil IUT 1
NOUGARET Marcel	Automatique IUT 1
PERARD Jacques	EEA. IUT 1

PROFESSEURS de 2^{ème} classe

BOUTHINON Michel	EEA. IUT 1
CHAMBON René	Génie Mécanique IUT 1
CHEHIKIAN Alain	EEA. IUT 1
CHENAVAS Jean	Physique IUT 1
CHOUTEAU Gérard	Physique IUT 1
CONTE René	Physique IUT 1
GOSSE Jean-Pierre	EEA.IUT 1
GROS Yves	Physique IUT 1
KUHN Gérard, (Détaché)	Physique IUT 1
MAZUER Jean	Physique IUT 1
MICHOULIER Jean	Physique IUT 1
MONLLOR Christian	EEA.IUT 1
PEFFEN René	Métallurgie IUT 1
PERRAUD Robert	Chimie IUT 1
PIERRE Gérard	Chimie IUT 1
TERRIEZ Jean-Michel	Génie Mécanique IUT 1
TOUZAIN Philippe	Chimie IUT 1
VINCENDON Marc	Chimie IUT 1

PROFESSEURS DE PHARMACIE

AGNIUS-DELORD Claudine	Physique	Faculté La Tronche
ALARY Josette	Chimie Analytique	Faculté La Tronche
BERIEL Hélène	Physiologie et Pharmacologie	Faculté La Tronche
CUSSAC Max	Chimie Therapeutique	Faculté La Tronche
DEMENGE Pierre	Pharmacodynamie	Faculté La Tronche
FAVIER Alain	Biochimie	C.H.R.G.
JEANNIN Charles	Pharmacie Galénique	Faculté Meylan
LATURAZE Jean	Biochimie	Faculté La Tronche
LUU DUC Cuong	Chimie Générale	Faculté La Tronche
MARIOTTE Anne-Marie	Pharmacognosie	Faculté La Tronche
MARZIN Daniel	Toxicologie	Faculté Meylan
RENAUDET Jacqueline	Bactériologie	Faculté La Tronche
ROCHAT Jacques	Hygiène et Hydrologie	Faculté La Tronche
SEIGLE-MURANDI Françoise	Botanique et Cryptogamie	Faculté Meylan
VERAIN Alice	Pharmacie Galénique	Faculté Meylan

MEMBRES DU CORPS ENSEIGNANT DE MEDECINE

PROFESSEURS CLASSE EXEPTIONNELLE ET 1ère CLASSE

AMBLARD Pierre	Dermatologie	C.H.R.G.
AMBROISE-THOMAS Pierre	Parasitologie	C.H.R.G.
BEAUDOING André	Pédiatrie-Pucriculture	C.H.R.G.
BEZEZ Henri	Orthopédie-Traumatologie	Hopital SUD
BONNET Jean-Louis	Ophthalmologie	C.H.R.G.
BOUCHET Yves	Anatomie	Faculté La Merci
	Chirurgie Générale et Digestive	C.H.R.G.
BUTEL Jean	Orthopédie-Traumatologie	C.H.R.G.
CHAMBAZ Edmond	Biochimie	C.H.R.G.
CHAMPETIER Jean	Anatomie-Topographique et Appliquée	
	O.R.L.	C.H.R.G.
CHARACHON Robert	Immunologie	C.H.R.G.
COLOMB Maurice	Anatomie-Pathologique	Hopital sud
COUDERC Pierre	Pneumophtisiologie	C.H.R.G.
DELORMAS Pierre	Cardiologie	C.H.R.G.
DENIS Bernard	Pharmacologie	C.H.R.G.
GAVEND Michel	Hématologie	Faculté La Merci
HOLLARD Daniel	Chirurgie Thoracique et Cardiovasculaire	C.H.R.G.
LATREILLE René	Bactériologie-Virologie	C.H.R.G.
	Gynécologie et Obstétrique	C.H.R.G.
LE NOC Pierre	Médecine du Travail	C.H.R.G.
MALINAS Yves	Clinique Médicale et Maladies Infectieuses	C.H.R.G.
MALLION Jean-Michel	Histologie	Faculté La Merci
MICOUD Max	Pneumologie	C.H.R.G.
	Neurologie	C.H.R.G.
MOURIQUAND Claude	Hépto-Gastro-Entérologie	C.H.R.G.
PARAMELLE Bernard	Neurochirurgie	C.H.R.G.
PERRET Jean	Clinique Chirurgicale	C.H.R.G.
RACHAIL Michel	Anestésiologie	C.H.R.G.
DE ROUGEMONT Jacques	Physiologie	Faculté La Merci
SARRAZIN Roger	Biochimie	Faculté La Merci
STIEGLITZ Paul		
TANCHE Maurice		
VIGNAIS Pierre		

PROFESSEURS 2ème CLASSE

BACHELOT Yvan	Endocrinologie	C.H.R.G.
BARGE Michel	Neurochirurgie	C.H.R.G.
BENABID Alim Louis	Biophysique	Faculté La Merci
BENSA Jean-Claude	Immunologie	Hopital Sud
BERNARD Pierre	Gynécologie-Obstétrique	C.H.R.G.
BESSARD Germain	Pharmacologie	ABIDJAN
BOLLA Michel	Radiothérapie	C.H.R.G.
BOST Michel	Pédiatrie	C.H.R.G.
BOUCHARLAT Jacques	Psychiatrie Adultes	Hopital Sud
BRAMBILLA Christian	Pneumologie	C.H.R.G.
CHIROSEL Jean-Paul	Anatomie-Neurochirurgie	C.H.R.G.
COMET Michel	Biophysique	Faculté La Merci
CONTAMIN Charles	Chirurgie Thoracique et Cardiovasculaire	C.H.R.G.
CORDONNIER Daniel	Néphrologie	C.H.R.G.
COULOMB Max	Radiologie	C.H.R.G.
CROUZET Guy	Radiologie	C.H.R.G.
DEBRU Jean-Luc	Médecine Interne et Toxicologie	C.H.R.G.
DEMONGEOT Jacques	Biostatistiques et Informatique Médicale	Faculté La Merci
DUPRE Alain	Chirurgie Générale	C.H.R.G.
DYON Jean-François	Chirurgie Infantile	C.H.R.G.
ETERRADOSSI Jacqueline	Physiologie	Faculté La Merci
FAURE Claude	Anatomie et Organogénèse	C.H.R.G.
FAURE Gilbert	Urologie	C.H.R.G.
FOURNET Jacques	Hépto-Gastro-Entérologie	C.H.R.G.
FRANCO Alain	Médecine Interne	C.H.R.G.
GIRARDET Pierre	Anesthésiologie	C.H.R.G.
GUIDICELLI Henri	Chirurgie Générale et Vasculaire	C.H.R.G.
GUIGNIER Michel	Thérapeutique et Réanimation Médicale	C.H.R.G.
HADJIAN Arthur	Biochimie	Faculté La Merci
HALIMI Serge	Endocrinologie et Maladies Métaboliques	C.H.R.G.
HOSTEIN Jean	Hépto-Gastro-Entérologie	C.H.R.G.
HUGONOT Robert	Médecine Interne	C.H.R.G.
JALBERT Pierre	Histologie-Cytogénétique	C.H.R.G.
JUNIEN-LAVILLAULOY Claude	O.R.L.	C.H.R.G.
KOLODIE Lucien	Hématologie Biologique	C.H.R.G.
LETOUBLON Christian	Chirurgie Générale	C.H.R.G.
MACHECOURT Jacques	Cardiologie et Maladies Vasculaires	C.H.R.G.
MAGNIN Robert	Hygiène	C.H.R.G.
MASSOT Christian	Médecine Interne	C.H.R.G.
MOUILLON Michel	Ophthalmologie	C.H.R.G.
PELLAT Jacques	Neurologie	C.H.R.G.
PHELIP Xavier	Rhumatologie	C.H.R.G.
RACINET Claude	Gynécologie-Obstétrique	Hopital Sud
RAMBAUD Pierre	Pédiatrie	C.H.R.G.
RAPHAEL Bernard	Stomatologie	C.H.R.G.
SCHAERER René	Cancérologie	C.H.R.G.
SEIGNEURIN Jean-Marie	Bactériologie-Virologie	Faculté La Merci
SELE Bernard	Cytogénétique	Faculté La Merci
SOTTO Jean-Jacques	Hématologie	C.H.R.G.
STOEBNER Pierre	Anatomie Pathologique	C.H.R.G.
VROUSOS Constantin	Radiothérapie	C.H.R.G.

C'est à bien plus d'un titre que je remercie ici B. VAN CUTSEM, Professeur à l'Université J. Fourier de Grenoble. Je lui dois en effet le privilège de pouvoir consacrer une partie de mon temps à la recherche. Tout d'abord parce qu'il m'a permis de débiter dans cette voie en me conduisant à un Doctorat de 3^{ème} Cycle consacré aux statistiques directionnelles. Ensuite parce que son dévouement à la gestion des structures, sans cesse en évolution, dans lesquelles j'ai effectué ce travail m'a toujours préservé de ces tâches indispensables au bon fonctionnement d'un laboratoire ou d'une équipe. Sa présence amicale et l'intérêt qu'il a toujours témoigné pour mon travail sont ici confirmés par l'honneur qu'il me fait en présidant mon jury.

J'apprécie la présence dans mon jury de C. GUEGUEN, Professeur à l'Ecole Nationale Supérieure des Télécommunications de Paris, qui a accepté de rapporter sur les orientations de cette thèse en Traitement du Signal. En m'invitant à présenter certains aspects de mon travail au groupe de recherche qu'il anime, et par l'accueil qu'il m'a toujours réservé lors de mes passages à l'Ecole, en particulier à travers l'enthousiasme qu'il a souvent manifesté au cours de nos discussions, il a contribué pour une grande part à me donner la confiance nécessaire pour mener à bien ce travail et je l'en remercie sincèrement.

C'est au cours de la 3^{ème} Rencontre Franco-Belge de Statisticiens à Rouen en 1982 que j'ai fait la connaissance de M. HALLIN, Professeur à l'Université Libre de Bruxelles, alors que je défendais, pour la première fois, les "couleurs" de la fonction d'autocorrélation partielle. Le hasard a voulu qu'il préside la session des Journées de Statistique de Lille en 1986 au cours de laquelle je poursuivais mon "combat", et que l'on se retrouve dans la même session, cette année, pendant les Journées de Statistique de Grenoble. Ce fut pour moi l'occasion d'apprécier ses qualités scientifiques et humaines ainsi que ses compétences dans le domaine des séries temporelles. C'est pourquoi, en ayant accepté très tôt de faire un rapport sur ma thèse, il m'a redonné le goût et la volonté d'en achever la rédaction, ce dont je le remercie vivement.

Je remercie G. OPPENHEIM, Professeur à l'Université Paris V, d'avoir pris sur son temps en cette période très chargée de l'année, pour s'intéresser à mon travail et participer à mon jury.

C'est sous l'impulsion de A. LE BRETON, Professeur à l'Université J. Fourier de Grenoble, que je me suis intéressé à l'étude de la statistique dans les processus. Il a depuis assuré la direction de mon travail malgré mon orientation dans le domaine plus spécifique des séries chronologiques. L'exemplarité de son dévouement à l'intérêt commun de l'Equipe de Statistique, dont il a pris la charge pendant plusieurs années, a été pour beaucoup dans ma détermination à aller au bout de ce travail. En dehors d'une amitié extra-professionnelle que je me permets d'évoquer ici, je lui suis très reconnaissant de la confiance qu'il m'a toujours témoignée et de la patience dont il a fait preuve à mon égard. Les incitations et encouragements qu'il m'a sans cesse prodigués sont à l'origine de bien des concrétisations de ce travail.

Beaucoup comprendront la chance qu'il m'a été donné d'avoir à mes côtés D. T. PHAM, Chargé de Recherche au CNRS, qui a assuré la co-direction de cette thèse. Son rayonnement et sa compétence dans le domaine des séries chronologiques, reconnus sur le plan international, font de lui un interlocuteur privilégié. Ainsi ses suggestions et ses critiques m'ont sans cesse stimulé et ont été bien souvent déterminantes dans l'avancement de mon travail. Je le remercie de l'intérêt qu'il a toujours manifesté à mon égard et de la collaboration qu'il m'a apportée.

Je remercie tous mes amis et collègues de l'Equipe de Statistique, et plus largement du Laboratoire TIM3, pour leurs marques de sympathie et l'aide que les uns et les autres ont pu m'apporter dans l'élaboration de mes programmes de simulation ou lors de l'utilisation de logiciels de traitement de texte pour la réalisation de la dactylographie de ce manuscrit.

Je remercie également, à travers J. L. LACOUME, Directeur du Centre d'Etude des Phénomènes Aléatoires et Géophysiques de Grenoble, les personnes qui, dans son laboratoire, m'ont sensibilisé aux problèmes statistiques en Traitement du Signal.

Je remercie enfin le personnel du Service de Reprographie de l'Institut IMAG à travers D. IGLESIAS, qui a réalisé le tirage de cette thèse, pour l'accueil qu'il m'a toujours réservé.

TABLE DES MATIERES

	pages
INTRODUCTION GENERALE	13
PARTIE A FONCTION D'AUTOCORRELATION PARTIELLE	23
INTRODUCTION	25
CHAPITRE I FONCTION D'AUTOCORRELATION PARTIELLE ET INNOVATIONS	29
§ I.1. Innovations partielles	30
§ I.2. Equivalence entre $\beta(\cdot)$ et $\rho(\cdot)$	33
§ I.3. Autres paramétrisations	35
§ I.4. Processus harmoniques	41
CHAPITRE II DECOMPOSITION DE CHOLESKY DES MATRICES D'AUTOCOVARIANCE ET DE LEURS INVERSEES	45
§ II.1. Les facteurs des décompositions de Cholesky de Λ et Λ^{-1}	46
§ II.2. Algorithmes de calcul récursifs des facteurs B et A	48
§ II.3. Paramétrisations de Λ liées aux décompositions de Cholesky	51
§ II.4. Comportement limite des éléments des facteurs A et B	55
CHAPITRE III APPROXIMATIONS PAR LES MODELES AUTOREGRESSIFS	61
§ III.1. Modèle autorégressif d'ordre p : $AR(p)$	61
§ III.2. Représentation autorégressive	64
§ III.3. Les critères d'approximation	76
CHAPITRE IV FONCTIONS D'AUTOCORRELATION PARTIELLE D'UN PROCESSUS VECTORIEL	83
§ IV.1. Innovations partielles	84
§ IV.2. Fonctions d'autocorrélation partielle standards	88
§ IV.3. Fonction d'autocorrélation partielle canonique	91

PARTIE B	APPORT DES AUTOCORRELATIONS PARTIELLES A L'ESTIMATION DANS LE DOMAINE TEMPOREL	99
	INTRODUCTION	101
CHAPITRE I	ESTIMATION DE MAXIMUM DE VRAISEMBLANCE D'UNE MATRICE DE COVARIANCE	107
§ I.1.	Généralités	108
§ I.2.	Comportement de la fonction de vraisemblance au bord de son domaine	110
§ I.3.	Une caractérisation du maximum de vraisemblance	114
§ I.4.	Moindres carrés sous contrainte	117
CHAPITRE II	ESTIMATION DE MAXIMUM DE VRAISEMBLANCE D'UNE MATRICE DE TŒPLITZ	127
§ II.1.	Le problème d'existence	127
§ II.2.	Décompositions spectrales des matrices de Tœplitz	133
§ II.3.	Une méthode de relaxation	138
§ II.4.	Algorithme et simulations	142
§ II.5.	Maxima locaux de la fonction de vraisemblance	146
CHAPITRE III	ESTIMATION DE LA STRUCTURE DE COVARIANCE D'UNE SEQUENCE STATIONNAIRE	153
§ III.1.	Prolongements presque-périodiques d'une séquence	154
§ III.2.	Le maximum de vraisemblance et ses singularités	161
§ III.3.	Autocorrélations partielles empiriques	168
CHAPITRE IV	ESTIMATION AUTOREGRESSIVE	175
§ IV.1.	Présentation des méthodes	175
§ IV.2.	Comparaison des méthodes par simulation	184
§ IV.3.	Extension des méthodes à un échantillon	193
	REFERENCES	201

INTRODUCTION GENERALE

Ce travail est consacré à l'étude, dans un cadre probabiliste et statistique, de la fonction d'autocorrélation partielle d'un processus scalaire réel, à temps discret, stationnaire au second ordre et centré. Cette fonction, que l'on notera $\beta(\cdot)$, constitue un paramètre équivalent à la fonction d'autocorrélation $\rho(\cdot)$ et caractérise ainsi la structure au second ordre du processus $X(\cdot)$, au facteur d'échelle près donné par la variance des composantes. Comme $\rho(\cdot)$, le paramètre $\beta(\cdot)$ est temporel au sens où la structure de la séquence $X(0), \dots, X(n)$ ne dépend de $\beta(\cdot)$ que par ses valeurs sur $\{1, \dots, n\}$. L'avantage de $\beta(\cdot)$, a priori, est qu'il s'agit d'un paramètre libre, ses valeurs successives sont quelconques dans $] -1, 1[$, alors que $\rho(\cdot)$ est une fonction de type positif et vérifie, par exemple, $\rho(2) > 2\rho^2(1) - 1$. Ceci laisse présager, sur le plan statistique, un meilleur comportement des méthodes axées sur l'estimation de $\beta(\cdot)$ par rapport à celles basées sur le paramètre $\rho(\cdot)$.

Cependant la fonction d'autocorrélation partielle ne semble pas avoir beaucoup retenu l'attention des auteurs statisticiens. Ainsi, par exemple, l'équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$ n'est pas mentionnée dans des ouvrages récents consacrés aux séries temporelles, qu'ils aient un caractère fondamental (Azencott et Dacunha-Castelle, 1984) ou plus appliqué (Priestley, 1981). Ceci peut s'expliquer par le fait que, sur le plan statistique, le recours à $\beta(\cdot)$ se justifie surtout dans des situations spécifiques : série courte ou modèle proche de la singularité. L'aspect "échantillon de séries courtes" élargit aussi le champ d'application d'autant plus que la notion de stationnarité qui, le plus souvent, fait

référence à un point de vue temporel, doit aussi être entendue au sens spatial. Un domaine d'application privilégié est constitué par certains aspects du traitement du signal (parole, géophysique, antenne,...) où $-\beta(k)$ correspond à une notion physique et est souvent appelé coefficient de réflexion. C'est pourquoi bien des contributions liées à l'utilisation de $\beta(\cdot)$ sont issues de cette discipline.

Bien évidemment $\beta(\cdot)$ n'est pas absent de la littérature statistique. Le coefficient $\beta(k)$ représente la corrélation partielle entre $X(t)$ et $X(t-k)$ dans l'ensemble $\{X(t), \dots, X(t-k)\}$ (cf. Kendall et Stuart, 1966). C'est aussi le coefficient de $X(t-k)$ dans la régression de $X(t)$ sur $\{X(t-1), \dots, X(t-k)\}$, cette propriété étant à la base du test de Quenouille (1949). Le modèle $AR(p)$ se caractérisant par la nullité de $\beta(\cdot)$ au-delà de p , Box et Jenkins (1970) suggèrent d'utiliser la représentation graphique de $\beta(\cdot)$ à des fins de modélisation. Il aura néanmoins fallu attendre 1973 pour que Barndorff-Nielsen et Schou montrent que $\{\beta(1), \dots, \beta(p)\}$ constitue une paramétrisation du modèle $AR(p)$. L'équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$, pour une situation quelconque, est établie en 1974 par Ramsey. En fait il s'agit d'un résultat classique de la théorie des polynômes orthogonaux sur le cercle unité (cf. Geronimus, 1960) via l'isométrie habituelle entre les approches temporelle et spectrale. Les coefficients $\beta(k)$, $k=1, \dots, p$, sont également les éléments clefs de l'algorithme de Levinson (1947) et Durbin (1960) pour la résolution des équations de Yule-Walker qui déterminent les paramètres du modèle $AR(p)$ associés à une structure de corrélation $\rho(\cdot)$ connue sur $\{1, \dots, p\}$. Cet algorithme passe par la construction récursive des modèles $AR(k)$, $k=1, \dots, p$, "sous-jacents". Ce rôle privilégié du modèle autorégressif est remarqué par Burg (1967) qui introduit la notion d'estimation spectrale autorégressive souvent dénommée "méthode d'entropie maximale". De plus Burg propose en 1968 une technique originale pour l'estimation du modèle $AR(p)$ qui repose sur l'estimation directe de $\beta(\cdot)$: les coefficients $\beta(k)$, $k \leq p$, de l'algorithme de Levinson-Durbin sont obtenus récursivement en minimisant, à chaque étape, les erreurs de prédiction progressive et rétrograde. Il notera également l'équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$ en 1975.

Ce travail est organisé en deux parties. La première, de nature probabiliste, présente les aspects de la fonction d'autocorrélation partielle liés à la description de la structure du processus. La deuxième montre que cette fonction joue un rôle essentiel dans l'estimation de cette structure, tout au moins dans les situations non standards. Nous présentons ci-dessous les différentes contributions que nous pensons avoir apportées au cours de cette étude et renvoyons le lecteur à l'introduction de chaque partie pour une description plus structurée de ce travail.

C'est un problème de modélisation d'une séquence stationnaire $\{X(0), \dots, X(n)\}$ qui, il y a une dizaine d'années, nous a conduit à aborder l'étude des séries temporelles. Partant d'un point de vue de statisticien formé à la statistique classique, qui n'a pas encore acquis le réflexe de l'utilisation des autocorrélations empiriques, nous avons posé le problème comme celui de l'estimation pour une structure particulière dans l'espace \mathbb{R}^{n+1} . Constatant que les contraintes qui pèsent sur $\rho(k)$, $k = 1, \dots, n$, constituaient une difficulté majeure dans la mise en œuvre de méthodes d'estimation, nous avons cherché une nouvelle paramétrisation. Nous avons alors remarqué que la structure de la séquence pouvait se décrire en orthogonalisant par le procédé de Gram-Schmidt la suite $X(0), \dots, X(n)$ et que ce procédé conduisait à un paramètre sans contrainte dans $]-1, 1[^n$ que nous avons noté $\{\beta(1), \dots, \beta(n)\}$. Nous avons conservé depuis cette notation $\beta(\cdot)$ alors que dans la littérature le plus souvent la fonction d'autocorrélation partielle est notée $\pi(\cdot)$ ou $\phi(\cdot)$.

Nos recherches bibliographiques nous ont permis de constater que $\beta(\cdot)$ joue un rôle plus important en traitement du signal qu'en statistique. De plus, bien avant leur parution dans la littérature spécialisée de l'une ou l'autre de ces disciplines, de nombreux résultats, comme l'équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$, étaient connus en théorie des polynômes orthogonaux. L'intérêt de ce rapprochement a d'ailleurs été souligné par Kailath en 1974. Notre premier travail a consisté à effectuer une étude complète de la fonction d'autocorrélation partielle, dans le cadre le plus simple des processus réels scalaires, afin de mettre en relief les différents aspects de ce paramètre. Notre approche est essentiel-

lement temporelle. Cependant, une présentation géométrique, dans l'espace de Hilbert engendré par les composantes du processus, facilite le lien avec le domaine spectral tout en indiquant l'apport de la théorie des polynômes orthogonaux. Ce point de vue a été exposé à la 3ème Rencontre Franco-Belge de Statisticiens (Dégerine, 1982b). Cette formulation géométrique permet de souligner et de préciser diverses propriétés à l'aide de démonstrations souvent simplifiées. En particulier nous obtenons l'équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$ par un procédé constructif d'une séquence stationnaire (Proposition A.I.2.1), utile en simulation, qui s'étend immédiatement au cas vectoriel (Proposition A.VI.3.3). Certains points, assez mal connus, apparaissent ici de façon naturelle : c'est le cas de l'utilisation de l'algorithme de Levinson-Durbin comme correspondance entre $\beta(\cdot)$ et $\rho(\cdot)$. Il fournit un test de positivité pour $\rho(\cdot)$, mais contient aussi un test de stabilité pour un polynôme (Remarque A.II.3.2) lorsque $\beta(\cdot)$ est déterminé à partir du paramètre $b(\cdot)$ du modèle autorégressif. D'autre part si A , triangulaire inférieure unitaire, et Σ^2 , diagonale, désignent les éléments de la décomposition de Cholesky d'une matrice d'autocorrélation, $\Gamma = A\Sigma^2A'$, l'algorithme construit récursivement les lignes de $B=A^{-1}$, qui représentent les paramètres des modèles AR(k) "sous-jacents", et fournit ainsi un procédé d'inversion de Γ . Le Roux et Guéguen (1977) proposent un autre algorithme de calcul de $\beta(\cdot)$ en fonction de $\rho(\cdot)$. Nous montrons qu'il est lié à la construction récursive des colonnes ou des lignes de $A\Sigma^2$ (Propositions A.II.2.2 et A.II.2.3). Les matrices A , B et Σ d'ordres successifs sont emboîtées, nous établissons que la convergence dans \mathcal{P}^2 des lignes de A , lorsque l'ordre augmente, caractérise les processus réguliers (Proposition A.II.4.2). Nous suggérons aussi, au paragraphe A.I.3, d'autres paramétrisations possibles de la structure de $X(\cdot)$, dont la "fonction d'autorégression orthogonale" $\gamma(\cdot)$, paramètres que l'on retrouve au paragraphe A.II.3 comme caractérisations des matrices de Tœplitz et qui aboutissent, par identification, à l'estimation autorégressive par "Rapports d'Energies Résiduelles" proposée par Dickinson (1978) (cf. chapitre B.IV).

Le modèle autorégressif, que nous avons déjà évoqué à plusieurs reprises, joue un rôle central dans tout notre travail. Cependant la représentation autorégressive (infinie) d'un processus stationnaire, contrairement à la représentation sous forme de moyenne mobile, pose encore de nombreuses questions. Grâce à des résultats souvent méconnus, nous pensons avoir apporté des éléments de réponse à certaines d'entre elles. La définition que nous proposons (Définition A.III.2.1) est qu'il soit possible d'extraire un bruit blanc $\varepsilon(\cdot)$ du processus $X(\cdot)$, par un filtre linéaire réalisable $b(\cdot)$:

$$\varepsilon(t) = \sum_{k=0}^{\infty} b(k)X(t-k), t \in \mathbb{Z}. \quad (1)$$

La série est définie au sens de la convergence en moyenne quadratique des sommes partielles, $b(\cdot)$ n'est pas nécessairement dans \mathcal{L}^2 et nous n'exigeons pas que $\varepsilon(\cdot)$ soit le processus d'innovation. On dira que $X(\cdot)$ est autorégressif lorsqu'il satisfait (1) et que $\varepsilon(\cdot)$ est le processus d'innovation. Une condition suffisante classique est que $X(\cdot)$ soit minimal (régulier et $b(\cdot) \in \mathcal{L}^2$) et que sa densité spectrale soit bornée (Corollaire A.III.2.6) mais ce dernier point n'est pas nécessaire (Corollaire A.III.2.8). Nous prouvons, par des exemples, que la condition $b(\cdot) \in \mathcal{L}^2$ n'est pas nécessaire et que $X(\cdot)$ peut ne pas être régulier (présence d'une composante discrète). Néanmoins certaines questions sont restées sans réponse : la représentation (1), avec $\varepsilon(\cdot)$ non nul, implique-t-elle que $X(\cdot)$ soit non singulier? Si $X(\cdot)$ est régulier et satisfait (1), est-il possible d'obtenir une représentation analogue en fonction du processus d'innovation? Nous précisons, dans le même chapitre, les critères d'approximation de $X(\cdot)$ par les modèles $AR(p)$ qui lui sont sous-jacents. Nous montrons, en utilisant des résultats de la théorie des polynômes orthogonaux, que le comportement limite des densités spectrales f_p correspondantes, lorsque l'ordre p du modèle tend vers l'infini, est cohérent avec la mesure spectrale F de $X(\cdot)$: convergence uniforme vers la densité spectrale f dans tout intervalle où F est absolument continue, f minorée par un nombre positif et suffisamment régulière (Proposition A.III.3.5), convergence vers zéro là où f s'annule, en étant régulière, et vers l'infini là où F présente une masse.

L'extension de la fonction d'autocorrélation partielle aux processus vectoriels est un travail récent. Les résultats ont été annoncés dans une note aux Comptes Rendus de l'Académie des Sciences (Dégerine, 1987) et seront présentés au congrès EUSIPCO-88 (Dégerine, 1988). Morf et coll. (1978) ont proposé une définition de $\beta(\cdot)$ à partir d'une normalisation de l'algorithme de Levinson-Durbin vectoriel (Whittle, 1963). Celle que nous proposons est intrinsèque au processus $X(\cdot)$. Notons $\varepsilon(t;n)$ et $\varepsilon^*(t;n)$, $n \geq 0$, les innovations progressive et rétrograde d'ordre n . Elles sont définies indépendamment de toute représentation en fonction des composantes de $X(\cdot)$. Il en est alors de même des variances résiduelles $\sigma^2(n) = E\{\varepsilon(t;n)\varepsilon(t;n)'\}$ et $\sigma^{*2}(n) = E\{\varepsilon^*(t;n)\varepsilon^*(t;n)'\}$ ainsi que de la fonction de covariance partielle $\delta(n) = E\{\varepsilon(t;n-1)\varepsilon^*(t-n;n-1)'\}$ (Définition A.IV.1.1). Ces quantités satisfont une récurrence (Proposition A.IV.1.2), analogue à celle du cas scalaire, dans laquelle intervient un inverse généralisé des variances résiduelles lorsqu'elles sont singulières. Cela permet d'étendre l'algorithme de Levinson-Durbin. On appelle "procédé d'orthogonalisation standard" toute procédure qui, à une matrice symétrique positive σ^2 , associe de façon unique deux matrices σ et σ^- où σ est une racine carrée de σ^2 et σ^- un inverse généralisé de σ tel que $\sigma^- \sigma$ soit diagonale et idempotente. Un tel procédé définit une fonction d'autocorrélation partielle standard $\beta(n) = \sigma(n-1)^- \delta(n) [\sigma^*(n-1)^-]'$ (Proposition A.IV.2.2). Celui de Gram-Schmidt correspond à la définition de Morf et coll. (1978). Celui que nous proposons conduit à la notion de fonction d'autocorrélation partielle canonique (cf. § A.IV.3). Dans ce cas $\beta(n)$ admet une décomposition en valeurs singulières $L'_n \Delta_n \Theta_n M_n$ dans laquelle L_n , M_n et Θ_n sont des matrices orthogonales telles que Θ_n commute avec Δ_n et peut s'interpréter comme un "signe généralisé". Les termes de la matrice diagonale Δ_n représentent les corrélations canoniques entre $\varepsilon(t;n-1)$ et $\varepsilon^*(t-n;n-1)$ qui, lorsqu'elles sont inférieures à un, sont aussi celles entre $\varepsilon(t;n)$ et $\varepsilon^*(t-n;n)$, les variables canoniques étant alors les composantes de $\sigma(n)^- \varepsilon(t;n)$ et de $\sigma^*(n)^- \varepsilon^*(t-n;n)$. Cette fonction canonique permet de mieux décrire la structure de $X(\cdot)$. Par exemple $\{\sigma^2(0); (L_n, \Delta_n), n \geq 1\}$ est caractéristique de la suite des variances résiduelles progressives $\{\sigma^2(n), n \geq 0\}$.

L'estimation de la structure de $X(\cdot)$ n'est envisagée que d'un point de vue temporel. L'usage de l'autocovariance empirique, ou du maximum de vraisemblance approché, n'est pas acceptable lorsque la série observée est courte, y compris dans le cas d'un échantillon, ou encore lorsque la structure de covariance du modèle est proche de la singularité, auquel cas on le dira "singulier". De même les méthodes de gradient, malgré un calcul exact des dérivées premières et secondes rendu possible par l'emploi du paramètre $\beta(\cdot)$, se heurtent à la non convexité de la fonction de vraisemblance (Dégerine et Pham, 1987). Nous verrons que le maximum de vraisemblance peut être obtenu par des méthodes de relaxation. D'autre part l'estimation empirique naturelle de $\beta(\cdot)$ n'est pas celle qui résulte des relations de Yule-Walker appliquées à l'autocovariance empirique. Dans celle que nous proposons, la valeur $\beta(k)$ apparaît à la fois comme une corrélation partielle empirique et comme un coefficient de régression. Elle s'avère être plus efficace que la méthode de Burg mais pour un coût de calcul plus important.

Nous accordons la plus grande place à la méthode du maximum de vraisemblance sous hypothèse gaussienne. Si x est la matrice $m \times n$ constituée par l'observation d'un échantillon de n séquences ($n \geq 1$) de longueur m , maximiser la vraisemblance équivaut à minimiser, par rapport à la matrice d'autocovariance Λ , la fonction

$$L(\Lambda; X) = \ln |\Lambda| + \text{tr} (\Lambda^{-1} X),$$

où $X = xx'/n$ est la matrice de covariance empirique. Ainsi X est une matrice positive d'ordre m , $X \in \mathcal{P}^+$ (en général singulière), et Λ parcourt une hypothèse \mathcal{H} incluse dans le cône positif \mathcal{C}^+ formé par l'ensemble des matrices de Toeplitz symétriques définies positives.

Le chapitre B.I est consacré au cas d'une hypothèse \mathcal{H} quelconque dans \mathcal{P}^+ . Nous établissons que la condition $\ker(\Lambda) \cap \ker(X) = \{0\}$, pour toute matrice Λ de la frontière de \mathcal{H} , est suffisante pour l'existence d'une solution (§ B.I.2). Ce résultat, ainsi qu'une caractérisation géométrique des solutions de l'équation de vraisemblance et de celle réalisant le maximum (§ B.I.3), ont été publiés dans Dégerine, 1985. Le cas stationnaire permet de vérifier que cette condition suffisante est minimale sans être néces-

saire (cf. § B.III.2). La minimisation de $L(\Lambda; X)$ équivaut à celle de $\text{tr}(\Lambda^{-1}X)/m$ (resp. $|\Lambda|$) sous la contrainte $|\Lambda| \leq 1$ (resp. $\text{tr}(\Lambda^{-1}X)/m \leq 1$). Ces problèmes conduisent à interpréter la covariance empirique X , dans le cas standard $\mathfrak{H} = \mathcal{S}^+$, comme une solution de moindres carrés dans un modèle à erreurs multiplicatives (§ B.I.4) et ont été présentés (Dégerine, 1982a) avec des applications possibles en analyse des données.

C'est dans le cadre d'un échantillon de séries courtes, $\mathfrak{H} = \mathcal{C}^+$, que nous avons proposé une méthode de calcul du maximum de vraisemblance exact (Dégerine, 1987). Il s'agit d'une méthode de relaxation basée sur la notion de décomposition spectrale spécifique aux matrices de Toeplitz que nous avons précédemment introduite (Dégerine, 1983). Ces résultats, auxquels nous avons ajouté la preuve de l'existence de maxima locaux pour la fonction de vraisemblance, font l'objet du chapitre B.II. Notons que les propriétés de convergence de la méthode sont clairement établies (§ B.II.3) et que, pour sa mise en œuvre, nous donnons un procédé de calcul de la structure de covariance d'une fonction presque-périodique (§ B.II.4).

Dans le cas d'une seule séquence, traité au chapitre B.III, la vraisemblance n'est pas presque sûrement bornée et la méthode du maximum de vraisemblance ne peut s'appliquer que sous une forme pénalisée. L'estimation autorégressive en est un exemple et fait l'objet du chapitre B.IV. L'autocovariance empirique $\Lambda_e(\cdot)$ est certainement la méthode la moins efficace. Son défaut, dû au biais, pour une série courte n'est pas surprenant. Cependant, pour une série longue ($m=1000$) d'un modèle AR(4) "singulier", on constate par simulation que, bien qu'ayant les mêmes propriétés (biais, variance) que les autres méthodes dans l'estimation des autocorrélations $\rho(\cdot)$, elle est la seule à réaliser de très mauvais résultats pour $\beta(\cdot)$ ainsi que pour le paramètre $b(\cdot)$ du modèle autorégressif, alors que les autres sont identiques de ce point de vue. On montre que cette estimation consiste, en un certain sens, à prolonger la séquence observée par des zéros alors que celle du maximum de vraisemblance, lorsqu'elle existe, la prolonge en annulant les erreurs de prédiction non observées. L'autocorrélation partielle empirique

$\beta_e(k)$ naturelle est la corrélation partielle entre les séquences $\underline{x}_k(1) = [x_1, \dots, x_{m-k}]'$ et $\underline{x}_k(k+1) = [x_{k+1}, \dots, x_m]'$ dans l'ensemble $\underline{x}_k(t) = [x_t, \dots, x_{t+m-k-1}]'$, $t=1, \dots, k+1$, calculée dans \mathbb{R}^{m-k} . Mais $\beta_e(k)$ n'est pas le coefficient de $\underline{x}_k(1)$ dans la régression de $\underline{x}_k(k+1)$ sur $\{\underline{x}_k(1), \dots, \underline{x}_k(k)\}$. La version que nous proposons, dite "symétrique" et notée $\beta_{es}(k)$, a cette propriété.

Les méthodes d'estimation autorégressive, présentées pour une seule séquence dans le paragraphe B.IV.1, sont liées au maximum de vraisemblance ou à des techniques de moindres carrés. Considérons les problèmes de minimisation des erreurs de prédiction progressive et rétrograde d'ordre k :

$$\text{Min}_{b(\cdot)} \sum_{t=k+1}^m \{\varepsilon^2(t;k) + \varepsilon^{*2}(t-k;k)\}, \quad \varepsilon(t;k) = \sum_{j=0}^k b(j)x_{t-j}, \quad \varepsilon^*(t;k) = \sum_{j=0}^k b(j)x_{t+j}.$$

L'estimation du modèle $\text{AR}(p)$, proposée par Burg, est obtenue de façon récursive sur l'ordre en imposant que le modèle d'ordre k soit sous-jacent à la solution d'ordre $(k+1)$. L'autocorrélation partielle empirique "symétrique" (ACPE) équivaut à poser $\beta_{es}(k) = -b(k)$ dans les solutions obtenues pour $k = 1, \dots, p$. La méthode des moindres carrés symétriques (MCS) retient le modèle $\text{AR}(p)$ dont la densité spectrale est définie par le carré du module du filtre, éventuellement instable, donné à l'ordre p . Celle des "rapports d'énergies résiduelles" (RER) est une approximation d'ACPE qui revient à ne retenir, dans la minimisation à l'ordre k ci-dessus, que les erreurs intervenant dans la formulation à l'ordre p . La méthode du maximum de vraisemblance exact (MVE), sous l'hypothèse d'un modèle $\text{AR}(p)$, a été réalisée initialement, en collaboration avec Pham, par une technique de gradient s'appuyant sur le paramètre $\beta(\cdot)$. Les simulations reportées ici utilisent une méthode de relaxation, portant également sur $\beta(\cdot)$, obtenue par Pham (1988). Elle consiste à itérer la méthode du maximum de vraisemblance récursif (MVR) proposée par Kay (1983) qui procède comme celle de Burg mais selon le critère de vraisemblance. Toutes ces méthodes s'étendent immédiatement au cas d'un échantillon de séquences (§ B.IV.3). Elles sont comparées par simulation (§ B.IV.2). Il en ressort que la qualité d'une méthode peut varier selon le modèle "singulier" simulé, ou encore

entre le cas d'une séquence et celui d'un échantillon, mais surtout selon le paramètre estimé, $\rho(\cdot)$, $\beta(\cdot)$ ou $b(\cdot)$. Les seules à ne pas présenter d'incohérence notable entre ces trois paramètres, MVE et ACPE, sont également celles qui réalisent simultanément des estimations correctes des variances du processus et de l'innovation d'ordre p . Elles apparaissent aussi le plus souvent comme les plus performantes.

Il nous reste à indiquer, en conclusion, d'autres applications en rapport avec la fonction d'autocorrélation partielle et que nous n'avons pas abordées ici. Tout d'abord l'extension au cas d'un processus scalaire complexe est immédiate pour la première partie car la théorie des polynômes orthogonaux (Geronimus, 1960) est présentée dans ce cadre; elle ne devrait pas présenter de difficulté pour les méthodes d'estimation du type "moindres carrés"; notons que celle du maximum de vraisemblance a été réalisée par Le Breton (1987), Pham, Le Breton et Tong (1988). La fonction d'autocorrélation partielle inverse (Bhansali, 1983a) correspond à la fonction d'autocorrélation inverse introduite par Cleveland (1972) et constitue un paramètre libre pour les moyennes mobiles inversibles. Pham et Le Breton (1988) retrouvent certains aspects de $\beta(\cdot)$ dans un algorithme du type Levinson-Durbin pour les modèles autorégressifs en temps continu. Kailath, Vieira et Morf (1978) soulignent des analogies entre le cas discret et le cas continu via la théorie des polynômes orthogonaux. Notons enfin que $\beta(\cdot)$ peut être considérée dans le cas non stationnaire (cf. Ramsey, 1974).

PARTIE A**FONCTION D'AUTOCORRELATION****PARTIELLE****INTRODUCTION**

- Chapitre I** **FONCTION D'AUTOCORRELATION PARTIELLE ET
INNOVATIONS**
- Chapitre II** **DECOMPOSITIONS DE CHOLESKY DES MATRICES
D'AUTOCOVARIANCE ET DE LEURS INVERSEES**
- Chapitre III** **APPROXIMATIONS PAR LES MODELES AUTOREGRESSIFS**
- Chapitre IV** **FONCTIONS D'AUTOCORRELATION PARTIELLE
D'UN PROCESSUS VECTORIEL**

INTRODUCTION

Dans cette première partie, à l'exclusion du dernier chapitre, $X(\cdot)$ désigne un processus scalaire réel, à temps discret, stationnaire au second ordre et centré. Sa structure (au second ordre) est donc caractérisée par sa fonction d'autocovariance $\Lambda(\cdot)$ ou, de façon équivalente, par sa mesure spectrale F . La variance de $X(t)$, $\sigma^2 = \Lambda(0)$, est un paramètre d'échelle et la forme de $\Lambda(\cdot)$ est donnée par la fonction d'autocorrélation $\rho(\cdot) = \Lambda(\cdot)/\sigma^2$. Cette fonction est la transformée de Fourier de la mesure spectrale normalisée $G = F/\Lambda(0)$. D'un point de vue mathématique, $\rho(\cdot)$ et G sont deux paramètres équivalents. Ce sont cependant deux concepts très différents pour l'étude des processus stationnaires. Le paramètre $\rho(\cdot)$ est temporel au sens où la séquence $X(0), \dots, X(n)$ ne dépend de $\rho(\cdot)$ que par ses valeurs sur $\{0, \dots, n\}$. Le domaine de variation des valeurs successives de $\rho(\cdot)$ est difficilement descriptible car $\rho(\cdot)$ est une fonction de type positif (transformée de Fourier de G). On a, par exemple, $\rho(2) \geq 2\rho^2(1) - 1$. L'aspect temporel n'apparaît pas dans la mesure spectrale G , mais G est un paramètre libre au sens où toute mesure positive symétrique de masse un sur $[-\pi, +\pi]$ peut être considérée comme la mesure spectrale normalisée d'un processus stationnaire $X(\cdot)$. La fonction d'autocorrélation partielle $\beta(\cdot)$ constitue un troisième paramètre équivalent à $\rho(\cdot)$ ou G . Cette fonction est, comme $\rho(\cdot)$, un paramètre temporel mais ses valeurs successives sont libres dans l'intervalle $] -1, 1[$. En contre partie la relation simple entre $\rho(\cdot)$ et G (transformée de Fourier) est remplacée, pour la correspondance entre $\beta(\cdot)$ et G , par un ensemble de relations faisant appel à un système de polynômes orthogonaux. En effet l'orthogonalisation de la suite $\{e^{in\lambda}, n \in \mathbb{N}\}$ dans $L^2\{[-\pi, +\pi], G\}$ selon le procédé de Gram-Schmidt construit un système orthogonal $\{\phi_n(e^{i\lambda}), n \in \mathbb{N}\}$ qui caractérise G et qui est entièrement défini par $\beta(\cdot)$. Cependant cette correspondance souligne le rôle des modèles autorégressifs dans la relation entre les aspects spectral et temporel.

Ainsi $\beta(k)$, $k \in \mathbb{N}^*$, désigne la corrélation partielle entre $X(t)$ et $X(t-k)$ dans l'ensemble $\{X(t), X(t-1), \dots, X(t-k)\}$ telle qu'elle est définie dans Kendall et Stuart (1966), c'est-à-dire la corrélation qui subsiste entre $X(t)$ et $X(t-k)$ lorsque l'on a éliminé l'effet (linéaire) des variables intermédiaires $X(t-1), \dots, X(t-k+1)$. Mais $\beta(k)$ est aussi le coefficient de $X(t-k)$ dans la régression linéaire multiple de $X(t)$ sur $\{X(t-1), \dots, X(t-k)\}$ et c'est certainement sous cette forme que $\beta(\cdot)$ a été utilisée pour la première fois dans le cadre des séries chronologiques (cf. test de Quenouille, 1949). Plus tard Box et Jenkins (1970) suggèrent d'utiliser $\beta(\cdot)$ à des fins de modélisation. En 1973, Barndoff-Nielsen et Schou montrent que $\{\beta(1), \dots, \beta(p)\}$ constitue une autre paramétrisation des modèles autorégressifs (AR) d'ordre p . L'équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$ est établie, pour une situation quelconque, par Ramsey en 1974. Elle est également observée par Burg en 1975, dans le cadre du traitement du signal, où $-\beta(k)$ est appelé coefficient de réflexion. En fait cette équivalence est un résultat classique en théorie des polynômes orthogonaux (cf. Geronimus, 1960).

Nous présentons ici les propriétés de la structure au second ordre de $X(\cdot)$ qui sont clairement liées à la fonction d'autocorrélation partielle. Notre approche est essentiellement temporelle. Cependant, utilisant la terminologie de la théorie des espaces de Hilbert, la correspondance avec le domaine spectral est immédiate. Nous insistons en particulier sur le lien, dont l'intérêt est souligné par Kailath (1974), avec des résultats connus en théorie des polynômes orthogonaux. Nous nous sommes restreint au cas le plus simple des processus scalaires réels à temps discret, néanmoins l'extension au cas vectoriel est abordée dans le dernier chapitre.

Le premier chapitre est consacré à la définition et aux propriétés immédiates de la fonction d'autocorrélation partielle $\beta(\cdot)$ liées aux innovations progressives et rétrogrades d'ordre fini ou non. Le langage géométrique, dans l'espace de Hilbert réel associé à $X(\cdot)$, permet de souligner la simplicité avec laquelle peuvent être obtenus les différents résultats : algorithme de Levinson-Durbin, équivalence entre $\beta(\cdot)$ et $\rho(\cdot)$, caractérisation

des processus non singuliers. Nous introduisons d'autres paramétrisations proches de $\beta(\cdot)$, dont la fonction d'autorégression orthogonale $\gamma(\cdot)$, qui nous seront utiles dans l'étude des techniques d'estimation.

Nous étudions, dans le second chapitre, les facteurs A et B des décompositions de Cholesky, $\Lambda = A\Sigma^2A'$ et $\Lambda^{-1} = B'\Sigma^{-2}B$, de la matrice d'autocovariance et de son inverse. La détermination récursive de $\beta(\cdot)$ en fonction de $\rho(\cdot)$ passe par la construction, elle-même récursive, de B (algorithme de Levinson-Durbin) ou de A (Le Roux et Guéguen, 1977). Ces facteurs permettent de souligner le rôle de plusieurs paramétrisations possibles de Λ et celle liée à $\gamma(\cdot)$ associe de façon naturelle une matrice d'autocovariance à toute matrice de covariance. Les matrices A (resp. B) d'ordres successifs sont emboîtées et la matrice d'ordre infini correspondante caractérise la structure de $X(\cdot)$. On montre en particulier que $X(\cdot)$ est régulier si et seulement si les lignes de A convergent dans \mathcal{L}^2 .

Une moyenne mobile (MA) d'ordre p se caractérise par la nullité de $\rho(\cdot)$ au-delà de p . Cependant la restriction de $\rho(\cdot)$ à $\{0, \dots, p\}$ ne correspond pas nécessairement à un modèle $MA(p)$. Nous verrons, dans le troisième chapitre, que le modèle $AR(p)$ se caractérise par la nullité de $\beta(\cdot)$ au-delà de p , définissant ainsi toujours un prolongement possible de $\rho(\cdot)$. La majeure partie de ce chapitre est consacrée au problème de la représentation autorégressive (infinie) d'un processus $X(\cdot)$. A notre connaissance il n'existe pas de solution satisfaisante au problème de l'existence d'une telle représentation, en termes de condition nécessaire et suffisante portant sur la mesure spectrale, analogue à celle connue pour une représentation $MA(\infty)$. Une condition suffisante, peu usitée, est la positivité de l'angle entre le passé et le futur de $X(\cdot)$, notion introduite par Helson et Szegö en 1960. Elle nous permet de montrer qu'un processus $AR(\infty)$ peut comporter une partie singulière, que la densité spectrale de la partie régulière peut s'annuler ou ne pas être bornée, et surtout que la minimalité (densité spectrale d'inverse intégrable) n'est pas une condition nécessaire dans le cas régulier. Le chapitre se termine par la présentation de différents critères d'approximation d'un processus $X(\cdot)$ par les modèles $AR(p)$ sous-

jacents dont un que l'on introduit en liaison avec le principe du maximum de vraisemblance. Le comportement limite de la densité spectrale du modèle $AR(p)$, lorsque p tend vers l'infini, est très cohérent avec la structure initiale de $X(\cdot)$. A cette occasion, les références systématiques au livre de Geronimus (1960) montrent la richesse de la théorie des polynômes orthogonaux sur le sujet.

L'extension de la notion de fonction d'autocorrélation partielle aux processus vectoriels stationnaires est présentée dans le dernier chapitre. La difficulté réside dans le choix de la normalisation à adopter pour définir une matrice de corrélation entre deux vecteurs aléatoires. Nous montrons ainsi que tout procédé d'orthogonalisation, sous réserve de quelques contraintes, conduit à une définition possible de $\beta(\cdot)$. Nous proposons cependant la notion de fonction d'autocorrélation partielle canonique : basée sur l'analyse canonique des innovations progressives et rétrogrades, elle a un caractère intrinsèque au processus que n'ont pas les autres.

CHAPITRE I

FONCTION D'AUTOCORRELATION PARTIELLE ET INNOVATIONS

Soit $X(\cdot)$ un processus scalaire réel, à temps discret, stationnaire au second ordre et centré. C'est-à-dire une famille de variables aléatoires réelles, $\{X(t), t \in \mathbb{Z}\}$, définies sur un espace probabilisé $(\Omega, \mathfrak{F}, P)$, dont les deux premiers moments existent et vérifient

$$E\{X(t)\}=0, \quad E\{X^2(t)\}=\sigma^2, \quad t \in \mathbb{Z}, \quad E\{X(t)X(s)\}=\Lambda(t-s), \quad (t,s) \in \mathbb{Z}^2.$$

Si $X(\cdot)$ est gaussien, la fonction d'autocovariance $\Lambda(\cdot)$ définit la loi de probabilité du processus. Sans l'hypothèse gaussienne, $\Lambda(\cdot)$ caractérise l'ensemble des propriétés du sous-espace de Hilbert réel \mathcal{M} de $L^2(\Omega, \mathfrak{F}, P)$ engendré par la famille $\{X(t), t \in \mathbb{Z}\}$. L'espace \mathcal{M} est formé de l'ensemble des combinaisons linéaires finies $\sum \alpha(t)X(t)$, $\alpha(t) \in \mathbb{R}$, ainsi que des limites, dans $L^2(\Omega, \mathfrak{F}, P)$, de suites de telles combinaisons. Notons que \mathcal{M} est en général strictement inclus dans $L^2(\Omega, \mathfrak{F}_X, P)$ où \mathfrak{F}_X désigne la sous-tribu de \mathfrak{F} engendrée par $X(\cdot)$ et cela même dans le cas gaussien. On écrira plus simplement $\mathcal{M} = \overline{\mathcal{L}}(X(t), t \in \mathbb{Z})$.

Seules les propriétés au second ordre de $X(\cdot)$ étant considérées, nous nous plaçons systématiquement dans l'espace de Hilbert \mathcal{M} . Les éléments de \mathcal{M} sont des variables aléatoires réelles centrées, le produit scalaire, noté $\langle \cdot, \cdot \rangle$, représente une covariance et la variance est donc notée $\|\cdot\|^2$. L'égalité, la somme (finie ou non), les projections et les notions qui s'y rattachent sont donc prises au sens L^2 .

L'aspect spectral est alors immédiat. Soit F la mesure spectrale liée à $\Lambda(\cdot)$ par la transformée de Fourier

$$\Lambda(n) = \int_{-\pi}^{+\pi} e^{in\lambda} dF(\lambda), n \in \mathbb{Z}.$$

Les relations $X(t) \leftrightarrow e^{it\lambda}$, $t \in \mathbb{Z}$, se prolongent en une isométrie entre les espaces de Hilbert \mathcal{M} et $L^2\{[-\pi, +\pi], F\}$. Nous ferons systématiquement référence au livre de Geronimus (1960) pour indiquer l'analogie, dans $L^2\{[-\pi, +\pi], F\}$, des résultats énoncés dans \mathcal{M} . En fait cet ouvrage correspond au cas complexe puisque F y est réelle mais pas nécessairement symétrique.

On introduit, dans le premier paragraphe, les innovations progressives et rétrogrades partielles en fonction desquelles on précise la définition de $\beta(\cdot)$. On montre alors, dans le paragraphe suivant, la correspondance bijective entre $\beta(\cdot)$ et $\rho(\cdot)$ à l'aide d'un procédé construisant $X(\cdot)$ à partir de $\beta(\cdot)$. Puis on propose, au paragraphe 3, d'autres paramétrisations liées à $\beta(\cdot)$ dont la fonction d'autorégression orthogonale $\gamma(\cdot)$. Le dernier paragraphe est consacré aux processus harmoniques en liaison avec les processus singuliers.

§ I.1. Innovations partielles

Pour $t \in \mathbb{Z}$ on introduit les sous-espaces de \mathcal{M} suivants :

$$\mathcal{M}(t;n) = \mathfrak{L}\{X(t), X(t-1), \dots, X(t-n+1)\}, n \in \mathbb{N}^*, \mathcal{M}(t;0) = \{0\},$$

$$\mathcal{M}(t;\infty) = \overline{\mathfrak{L}\{X(s), s \leq t\}}, \mathcal{M}(-\infty) = \bigcap \mathcal{M}(t;\infty).$$

Si d est la dimension de \mathcal{M} , avec $d=+\infty$ lorsque cette dimension n'est pas finie, celle de $\mathcal{M}(t;n)$ est égale à $n \wedge d$ ($\min(n, d)$) compte tenu de la stationnarité de $X(\cdot)$. Ainsi la projection orthogonale $X(t;n)$ de $X(t)$ sur $\mathcal{M}(t-1;n)$, $n \in \mathbb{N}^*$, se décompose de façon unique sur $\{X(t-1), \dots, X(t-n)\}$, les coefficients étant indépendants de t :

$$X(t;n) = -\sum_{k=1}^n b(n,k)X(t-k),$$

si l'on impose la condition $b(n,k)=0$, $d < k \leq n$.

On appelle **innovation progressive partielle d'ordre n** , $\varepsilon(t;n)$, l'erreur de prévision de $X(t)$ par son passé immédiat de longueur n . Avec la convention $X(t;0)=0$ et posant $b(n,0)=1$ on a :

$$\varepsilon(t;n) = X(t) - X(t;n) = \sum_{k=0}^n b(n,k)X(t-k), \quad n \in \mathbb{N}, t \in \mathbb{Z}. \quad (1)$$

On note également $X(t;\infty)$ la projection orthogonale de $X(t)$ sur $\mathcal{M}(t-1;\infty)$ et $\varepsilon(t;\infty)$ l'innovation correspondante. Lorsqu'il n'est pas nul $\varepsilon(.,\infty)$ est le **processus d'innovation (progressive)** et sera noté $\varepsilon(.,\infty)$.

Les notions analogues obtenues par inversion du temps seront signalées par un astérisque. Par exemple l'**innovation rétrograde partielle d'ordre n** est notée $\varepsilon^*(t;n)$, elle est donnée par :

$$\varepsilon^*(t;n) = X(t) - X^*(t;n) = \sum_{k=0}^n b(n,k)X(t+k), \quad n \in \mathbb{N}, t \in \mathbb{Z}.$$

La variance de $\varepsilon(t;n)$ ne dépend pas de t et est égale à celle de $\varepsilon^*(t;n)$:

$$\sigma^2(n) = \|\varepsilon(t;n)\|^2 = \|\varepsilon^*(t;n)\|^2, \quad n \in \mathbb{N}, t \in \mathbb{Z}.$$

Lorsque la dimension d de \mathcal{M} est finie le processus $X(.,\infty)$ est dit **linéairement singulier d'ordre d** . Il satisfait alors l'équation aux différences d'ordre (minimal) d traduisant la nullité des innovations d'ordre supérieur ou égal à d :

$$\sum_{k=0}^d b(k)X(t-k) = \sum_{k=0}^d b(k)X(t+k) = 0, \quad t \in \mathbb{Z}. \quad (2)$$

C'est à dire que l'on a posé $b(k) = b(n,k)$, $k \leq d$, $n \geq d$.

I.1.1. Définition.—La fonction d'autocorrélation partielle de $X(.,\infty)$, notée $\beta(.,\infty)$, est définie en tout n de \mathbb{Z} comme la corrélation partielle entre $X(t)$ et $X(t+n)$ dans l'ensemble $\{X(t), \dots, X(t+n)\}$. De façon précise :

$$\beta(-n) = \beta(n) = \langle \varepsilon(t;n-1), \varepsilon^*(t-n;n-1) \rangle / \sigma^2(n-1), \quad 1 \leq n \leq d, \quad (3)$$

$$\beta(0) = 1, \quad \beta(-n) = \beta(n) = \beta(d), \quad n > d.$$

La convention retenue pour $n > d$ est celle de Ramsey (1974). Elle consiste à

poser égale à ± 1 la corrélation entre deux variables nulles, elle simplifie l'énoncé de la caractérisation des processus non singuliers en fonction de $\beta(\cdot)$ (cf. Corollaire I.3.3).

I.1.2. Proposition.—*Les innovations partielles satisfont les récurrences suivantes :*

$$\varepsilon(t;n) = \varepsilon(t;n-1) - \beta(n)\varepsilon^*(t-n;n-1) \quad n \in \mathbb{N}^*, \quad t \in \mathbb{Z}, \quad (4)$$

$$\varepsilon^*(t;n) = \varepsilon^*(t;n-1) - \beta(n)\varepsilon(t+n;n-1)$$

et leurs variances sont données par :

$$\sigma^2(n) = [1 - \beta^2(n)]\sigma^2(n-1) = \sigma^2 \prod_{k=1}^n [1 - \beta^2(k)], \quad n \in \mathbb{N}^*. \quad (5)$$

Démonstration.— Pour $n \leq d$, $\mathcal{M}(t-1;n)$ se décompose selon la somme directe :

$$\mathcal{M}(t-1;n) = \mathcal{M}(t-1;n-1) \oplus \mathfrak{L}\{\varepsilon^*(t-n;n-1)\}.$$

Ainsi la projection orthogonale de $X(t)$ sur $\mathcal{M}(t-1;n)$ s'écrit :

$$X(t;n) = X(t;n-1) + \langle X(t), \varepsilon^*(t-n;n-1) \rangle \varepsilon^*(t-n;n-1) / \sigma^2(n-1),$$

et la première relation de la récurrence (4) est donnée par la définition (3) de $\beta(n)$:

$$\langle X(t), \varepsilon^*(t-n;n-1) \rangle = \langle \varepsilon(t;n-1), \varepsilon^*(t;n-1) \rangle = \beta(n)\sigma^2(n-1).$$

La deuxième relation de cette récurrence s'obtient de la même façon. Les égalités (5) sont une conséquence simple de l'orthogonalité entre $\varepsilon(t;n)$ et $\varepsilon^*(t-n;n-1)$ dans (4). Pour $n > d$ les relations deviennent "0=0" et $\sigma^2(n)=0$. ♦

Les récurrences (4) traduisent, dans le domaine temporel, des relations qui, d'un point de vue spectral, sont classiques en théorie des polynômes orthogonaux. L'orthogonalisation, selon le procédé de Gram-Schmidt, de la suite $\{e^{in\lambda}, n \in \mathbb{N}\}$ dans $L^2([-\pi, +\pi], F)$ définit un système de polynômes orthogonaux $\{\phi_n, n \leq d\}$ (polynômes de Szegö). Le polynôme ϕ_n est normalisé en posant égal à 1 le coefficient du terme de plus haut degré. Ces polynômes se construisent alors par la double récurrence :

$$\phi_n(z) = z\phi_{n-1}(z) - \beta(n)\phi_{n-1}^*(z) \quad 1 \leq n \leq d, \quad (6)$$

$$\phi_n^*(z) = \phi_{n-1}^*(z) - \beta(n)z\phi_{n-1}(z)$$

où $\phi_n^*(z) = z^n \phi_n(1/z)$, $n \leq d$, et $\phi_0(z) = 1$ (Geronimus, Ch.VIII).

On a, selon l'isométrie habituelle :

$$\varepsilon(n;n) \leftrightarrow \phi_n, \quad \varepsilon^*(0;n) \leftrightarrow \phi_n^*, \quad n \leq d,$$

et donc, avec les mêmes coefficients que dans (1) :

$$\phi_n(z) = \sum_{k=0}^n b(n,k) z^{n-k}, \quad n \leq d.$$

En théorie des polynômes orthogonaux les $\beta(n)$, $1 \leq n \leq d$, sont appelés paramètres du système $\{\phi_n, n \leq d\}$. Nous ferons parfois l'abus d'écrire $\phi_n(\lambda)$ pour $\phi_n(e^{i\lambda})$.

§ I.2. Equivalence entre $\beta(\cdot)$ et $\rho(\cdot)$

La fonction d'autocorrélation $\rho(\cdot)$, proportionnelle à la fonction d'autocovariance, $\rho(\cdot) = \Lambda(\cdot) / \sigma^2$, est la transformée de Fourier de la mesure spectrale normalisée $G = F / \sigma^2$:

$$\rho(n) = \int_{-\pi}^{+\pi} e^{in\lambda} dG(\lambda), \quad n \in \mathbb{Z}. \quad (7)$$

cette fonction est évidemment symétrique, $\rho(-n) = \rho(n)$, avec $\rho(0) = 1$ et est donc caractérisée par ses valeurs sur \mathbb{N}^* .

De même $\beta(\cdot)$ est caractérisée par sa restriction à \mathbb{N}^* . Les relations (5), ainsi que la convention de la Définition I.1.1, montrent qu'elle est de l'un des deux types suivants :

- (i) $|\beta(n)| < 1$, $n \in \mathbb{N}^*$, $(d = +\infty)$
(ii) $|\beta(n)| < 1$, $1 \leq n < d < +\infty$, $|\beta(d)| = 1$, $\beta(n) = \beta(d)$, $n > d$.
(8)

Dans le chapitre VIII de Geronimus l'équivalence entre $\rho(\cdot)$ satisfaisant (7) et $\beta(\cdot)$ du type (8) (i) est obtenue par l'intermédiaire de G où (7) est caractérisée par la stricte positivité des déterminants des matrices de Toeplitz

$$\Delta_n = |\rho(i-j), \quad i, j = 0, \dots, n|, \quad n \in \mathbb{N},$$

et vérifiée à partir des relations (cf. Remarque II.1.3) :

$$\Delta_n / \Delta_{n-1} = \prod_{k=1}^n [1 - \beta^2(k)], \quad n \in \mathbb{N}^*.$$

Les relations algébriques entre $\rho(\cdot)$ et $\beta(\cdot)$ montrent que la correspondance biuni-

voque est satisfaite par les restrictions de ces fonctions à $\{1, \dots, n\}$ pour tout n de \mathbb{N}^* . Cette correspondance est assurée par l'algorithme de Levinson-Durbin lorsque n est inférieur à d .

Algorithme de Levinson-Durbin

$$\begin{aligned} & \tilde{\sigma}^2(0)=1 \\ & n \leq d \left\{ \begin{aligned} & \beta(n)\tilde{\sigma}^2(n-1) = \rho(n) + \sum_{k=1}^{n-1} b(n-1,k)\rho(n-k) \\ & \tilde{\sigma}^2(n) = [1-\beta^2(n)]\tilde{\sigma}^2(n-1) \\ & b(n,n) = -\beta(n), \quad b(n,k) = b(n-1,k) - \beta(n)b(n-1,n-k), \quad 1 \leq k \leq n-1 \end{aligned} \right. \\ & n > d \left\{ \begin{aligned} & 0 = \rho(n) + \sum_{k=1}^d b(k)\rho(n-k) \\ & b(k) = b(d,k), \quad 1 \leq k \leq d \\ & \beta(n) = \beta(d) \end{aligned} \right. \end{aligned}$$

Ces relations sont une conséquence directe de la Proposition I.1.2 et des égalités suivantes :

$$\begin{aligned} \varepsilon(n;n-1) &= X(n) + \sum_{k=1}^{n-1} b(n-1,k)X(n-k), \\ \langle X(0), \varepsilon(n;n-1) \rangle &= \langle \varepsilon^*(0;n-1), \varepsilon(n;n-1) \rangle = \beta(n)\sigma^2(n-1), \quad \sigma^2(n) = \sigma^2\tilde{\sigma}^2(n). \end{aligned}$$

Afin de montrer que cet algorithme définit une correspondance bijective entre $\beta(\cdot)$ du type (8) et $\rho(\cdot)$ satisfaisant (7) on peut, comme Ramsey (1974) ou Burg (1975), utiliser la même caractérisation de (7) que ci-dessus. Nous proposons le procédé constructif suivant (Dégerine, 1982b) :

I.2.1. Proposition.—*Soit $\{Y(n), n \in \mathbb{N}\}$ une suite orthogonale telle que :*

$$\|Y(0)\|^2 = \sigma^2(0) = \sigma^2, \quad \|Y(n)\|^2 = \sigma^2(n) = [1-\beta^2(n)]\sigma^2(n-1), \quad n \in \mathbb{N}^*,$$

où $\beta(\cdot)$ est donné sur \mathbb{N}^* et satisfait :

$$|\beta(n)| \leq 1, \quad |\beta(n)| = 1 \Rightarrow \beta(n+1) = \beta(n).$$

Alors la suite $\{X(n), n \in \mathbb{N}\}$ construite par

$$X(0) = Y(0),$$

$$X(n) = X(n;n-1) + \beta(n)\varepsilon^*(0;n-1) + Y(n), n \in \mathbb{N}^*,$$

est stationnaire et sa fonction d'autocorrélation partielle est égale à $\beta(\cdot)$.

Démonstration.— Il est facile de vérifier que les variances de $X(0)$ et de $X(1) = \beta(1)X(0) + Y(1)$ sont égales à σ^2 et que $\langle X(0), X(1) \rangle = \beta(1)\sigma^2$. Supposons que $\{X(0), \dots, X(n-1)\}$ est stationnaire et admet $\beta(1), \dots, \beta(n-1)$ comme autocorrélations partielles. Les éléments $X(n;n-1)$ et $\varepsilon^*(0;n-1)$ sont alors parfaitement définis par $\beta(1), \dots, \beta(n-1)$ et on a, pour $1 \leq k \leq n-1$:

$$\begin{aligned} \langle X(n), X(n-k) \rangle &= \langle X(n;n-1), X(n-k) \rangle = \langle X(n-1;n-1), X(n-1-k) \rangle \\ &= \langle X(n-1), X(n-1-k) \rangle. \end{aligned}$$

La variance de $X(n)$ est donnée par :

$$\begin{aligned} \|X(n)\|^2 &= \|X(n;n-1)\|^2 + \beta^2(n)\|\varepsilon^*(0;n-1)\|^2 + \|Y(n)\|^2 \\ &= \sigma^2 - \sigma^2(n-1) + \beta^2(n)\sigma^2(n-1) + [1-\beta^2(n)]\sigma^2(n-1) = \sigma^2. \end{aligned}$$

Donc $\{X(0), \dots, X(n)\}$ est stationnaire. L'expression de $X(n)$ permet alors d'écrire :

$$\varepsilon(n;n-1) - \beta(n)\varepsilon^*(0;n-1) = Y(n).$$

Enfin l'orthogonalité de $Y(n)$ à $\mathcal{M}(n-1;n-1)$ donne :

$$\langle \varepsilon(n;n-1), \varepsilon^*(0;n-1) \rangle - \beta(n)\sigma^2(n-1) = 0.$$

Lorsque $\beta(\cdot)$ est du type (8) (ii) la construction ci-dessus se poursuit jusqu'à l'étape d . Elle aboutit à $X(d)=X(d;d)$ et le processus est singulier d'ordre d . ♦

Cette proposition prouve la surjectivité de l'application $\rho(\cdot) \rightarrow \beta(\cdot)$. Les relations de l'algorithme de Levinson-Durbin restent cependant nécessaires pour assurer l'injectivité.

§ I.3. Autres paramétrisations

L'attrait essentiel de la fonction d'autocorrélation partielle $\beta(\cdot)$ par rapport à la fonction d'autocorrélation $\rho(\cdot)$ est la simplicité de son domaine de variation. Cependant il est parfois avantageux de disposer de variables à valeurs sur \mathbb{R} plutôt que sur $]-1,1[$. Ceci

peut être réalisé par le biais de diverses transformations. Si de plus le nouveau paramètre présente une interprétation simple par rapport au processus $X(\cdot)$, il peut conduire à de nouvelles méthodes d'estimation.

Pour $1 \leq k \leq d$, $\beta(k)$ est le cosinus de l'angle entre les innovations $\varepsilon(t; k-1)$ et $\varepsilon^*(t-k; k-1)$. La cotangente $\delta(k)$ de cet angle est sans doute la transformation la plus naturelle conduisant à un paramètre à valeurs sur \mathbb{R} ($k < d$). La correspondance entre $\beta(\cdot)$ et $\delta(\cdot)$ est simple :

$$\delta(k) = \beta(k)[1 - \beta^2(k)]^{-1/2}, \quad \beta(k) = \delta(k)[1 + \delta^2(k)]^{-1/2}, \quad k < d.$$

Les variances $\sigma^2(k)$, $k < d$, des innovations sont données par

$$\sigma^2(k) = \sigma^2(k-1) / [1 + \delta^2(k)] = \sigma^2 / \prod_{j=1}^k [1 + \delta^2(j)].$$

On a ainsi :

$$\delta^2(k) = [\sigma^2(k-1) - \sigma^2(k)] / \sigma^2(k), \quad k < d.$$

Sachant que $\sigma^2(k)$ mesure l'erreur de prévision de $X(t)$ par son passé immédiat de longueur k , l'interprétation de $\delta(k)$ est analogue à celle de $\beta(k)$ puisque l'on a :

$$\beta^2(k) = [\sigma^2(k-1) - \sigma^2(k)] / \sigma^2(k-1), \quad k \leq d.$$

Lorsqu'ils sont proches de zéro $\beta(k)$ et $\delta(k)$ sont équivalents. Par contre $\delta(k)$ marque plus nettement la proximité de $X(\cdot)$ à la singularité. Cependant $\delta(k)$ reste, par rapport à la structure de $X(\cdot)$, une transformation artificielle de l'autocorrélation partielle $\beta(k)$.

D'autres paramétrisations sont suggérées par certaines décompositions orthogonales de $X(t; n)$ et de $\varepsilon(t; n)$ basées sur la remarque suivante :

I.3.1. Remarque.— Pour tout n de $\bar{\mathbb{N}}^* = \mathbb{N}^* \cup \{+\infty\}$, les innovations rétrogrades $\{\varepsilon^*(t-k; k), 0 \leq k < n \wedge d\}$ forment une base orthogonale de l'espace $\mathcal{M}(t; n)$.

I.3.2. Proposition.— La décomposition orthogonale de $X(t; n)$ sur $\{\varepsilon^*(t-1-k; k), 0 \leq k < n \wedge d\}$ est donnée par

$$X(t; n) = \sum_{k=1}^n \beta(k) \varepsilon^*(t-k; k-1), \quad n \in \bar{\mathbb{N}}^*. \quad (9)$$

La variance $\sigma^2(\infty)$ de $\varepsilon(t;\infty)$ satisfait alors :

$$\sigma^2(\infty) = \sigma^2 \prod_{k=1}^{\infty} [1 - \beta^2(k)].$$

Démonstration.—Pour $n \in \mathbb{N}^*$, l'égalité (9) s'obtient en itérant la première relation de récurrence de la Proposition I.1.2. Pour $d < +\infty$ et $n \geq d$, cette égalité donne l'expression de $X(t)$:

$$X(t) = \sum_{k=1}^d \beta(k) \varepsilon^*(t-k; k-1).$$

Lorsque $d = +\infty$ la projection de $X(t; \infty)$ sur $\mathcal{M}(t-1; n)$ est égale à $X(t; n)$ et la relation (9) avec $n = +\infty$ traduit la limite en moyenne quadratique $X(t; \infty) = \text{l.i.m. } X(t; n)$. On a alors $\varepsilon(t; \infty) = \text{l.i.m. } \varepsilon(t; n)$ et par suite

$$\sigma^2(\infty) = \lim \sigma^2(n) = \sigma^2 \prod_{k=1}^{\infty} [1 - \beta^2(k)]. \blacklozenge$$

Le processus $X(\cdot)$ est dit **singulier** lorsque $\mathcal{M} = \mathcal{M}(-\infty)$, ceci équivaut à $\mathcal{M}(t; \infty) = \mathcal{M}(t-1; \infty)$ ou encore $\sigma^2(\infty) = 0$. La caractérisation des processus non singuliers ($\sigma^2(\infty) > 0$) est alors immédiate (Ramsey, 1974).

I.3.3 Corollaire.—*Le processus $X(\cdot)$ est non linéairement singulier si et seulement si sa fonction d'autocorrélation partielle $\beta(\cdot)$ est dans \mathcal{F}^2 .*

$$\sum_{k=1}^{\infty} \beta^2(k) < +\infty.$$

Dans la littérature on utilise parfois le terme "régulier" pour désigner un processus non singulier, nous le réserverons aux processus ne comportant pas de partie singulière. Pour les processus non singuliers on notera σ_ε^2 la variance $\sigma^2(\infty)$ de l'innovation. Les **innovations normalisées** (de norme 1) sont désignées en remplaçant ε par η dans les différentes expressions avec la convention $\eta(\cdot) = 0$ lorsque $\varepsilon(\cdot) = 0$.

I.3.4. Définition.—On appelle fonction d'autorégression orthogonale de $X(\cdot)$, la fonction $\gamma(\cdot)$ définie sur \mathbb{N} par la décomposition orthogonale de $X(t)$ suivante :

$$X(t) = \sum_{k=1}^{\infty} \gamma(k) \eta^{*(t-k; k-1)} + \gamma(0) \eta(t), \quad (10)$$

où $\gamma(k)=0$ lorsque l'innovation correspondante est nulle.

On a $\gamma(0) = \sigma_{\varepsilon}$ (0 dans le cas singulier) et la série, dans l'égalité (10), est la décomposition orthogonale normalisée de $X(t; \infty)$ donnée par la relation (9).

I.3.5. Proposition.—Les relations

$$\begin{aligned} \gamma(0) &= \sigma \prod_{k=1}^{\infty} [1 - \beta^2(k)]^{1/2}, \quad \gamma(1) = \sigma \beta(1), \quad \gamma(k) = \sigma \beta(k) \prod_{j=1}^{k-1} [1 - \beta^2(j)]^{1/2}, \quad k \geq 2, \\ \sigma^2 &= \sum_{k=1}^{\infty} \gamma^2(k), \quad \beta(k) = \gamma(k) [\gamma(0) + \sum_{j=k}^{\infty} \gamma^2(j)]^{-1/2}, \quad k \in \mathbb{N}^*, \quad k \leq d, \end{aligned}$$

établissent une correspondance biunivoque entre le paramètre $\{\sigma^2; \beta(\cdot)\}$ et la fonction de régression orthogonale $\gamma(\cdot)$. Le domaine de variation de $\gamma(\cdot)$ est l'ensemble des fonctions définies sur \mathbb{N} , appartenant à \mathcal{P}^2 , et vérifiant $\gamma(0) \geq 0$.

Démonstration.—La Définition I.3.4 montre que $\gamma(\cdot)$ appartient à \mathcal{P}^2 , satisfait $\gamma(0) \geq 0$, et que son expression en fonction de $\{\sigma^2; \beta(\cdot)\}$ est égale à :

$$\gamma(0) = \sigma(\infty), \quad \gamma(k) = \beta(k) \sigma(k-1), \quad k \in \mathbb{N}^*.$$

Ceci correspond aux relations annoncées compte tenu des expressions de $\sigma^2(k-1)$ et de $\sigma^2(\infty)$ (Propositions I.1.2 et I.3.2). Elle montre aussi que l'on a :

$$\sigma^2(k-1) = \gamma^2(0) + \sum_{j=k}^{\infty} \gamma^2(j),$$

d'où la nécessité des relations inverses. Enfin ces relations associent à toute fonction $\gamma(\cdot)$ de \mathcal{P}^2 , vérifiant $\gamma(0) \geq 0$, un couple $\{\sigma^2; \beta(\cdot)\}$ dans lequel $\beta(\cdot)$ satisfait (8). ♦

Le paramètre $\gamma(\cdot)$ n'est pas rigoureusement temporel car la structure de la séquence $\{X(0), \dots, X(n)\}$ est caractérisée par $\{\sigma^2(n); \gamma(1), \dots, \gamma(n)\}$ où $\sigma^2(n)$ est, comme ci-dessus, la somme des carrés des autres valeurs de $\gamma(\cdot)$. Son interprétation est

clairement exprimée par sa définition mais on peut souligner les relations :

$$\gamma^2(k) = \sigma^2(k-1) - \sigma^2(k), k \in \mathbb{N}^*.$$

Notons l'importance du rôle tenu par $\gamma(0)$: quelle que soit $\gamma(\cdot)$, sa restriction à \mathbb{N}^* caractérise la structure d'un processus singulier; par contre pour un processus non singulier, $\gamma(0) > 0$, les paramètres $\gamma(\cdot)$ et $\beta(\cdot)$ ont le même comportement à l'infini.

I.3.6. Proposition.—La décomposition orthogonale de $\varepsilon(t;n)$, $n < d$, sur $\{\varepsilon^*(t-k;k), 0 \leq k \leq n\}$ est donnée par :

$$\varepsilon(t;n) = \sigma^2(n) \left\{ X(t)/\sigma^2 - \sum_{k=1}^n \beta(k) \varepsilon^*(t-k;k)/\sigma^2(k) \right\}.$$

La série

$$X(t)/\sigma^2 - \sum_{k=1}^{\infty} \beta(k) \varepsilon^*(t-k;k)/\sigma^2(k) \quad (11)$$

est convergente (au sens L^2) si et seulement si $X(\cdot)$ est non singulier. Sa somme est alors égale à $\varepsilon(t)/\sigma^2$.

Démonstration.—La décomposition de $\varepsilon(t;n)$ sur la base considérée s'écrit :

$$\varepsilon(t;n) = \sum_{k=0}^n \langle \varepsilon(t;n), \varepsilon^*(t-k;k) \rangle \varepsilon^*(t-k;k)/\sigma^2(k).$$

On a pour $0 \leq k \leq n$:

$$\begin{aligned} \langle \varepsilon(t;n), \varepsilon^*(t-k;k) \rangle &= \sum_{j=0}^k b(k,j) \langle \varepsilon^*(t;n), X(t-k+j) \rangle \\ &= b(k,k) \langle \varepsilon^*(t;n), X(t) \rangle = b(k,k) \sigma^2(n). \end{aligned}$$

D'où l'expression de $\varepsilon(t;n)$ puisque $b(0,0)=1$ et $b(k,k) = -\beta(k)$ pour $1 \leq k \leq n$ (cf. algorithme de Levinson-Durbin). La série (11) est convergente si et seulement si la série des coefficients, après normalisation des $\varepsilon^*(t-k;k)$, est dans \mathcal{L}^2 . D'après la décomposition de $\varepsilon(t;n)$ on a

$$1/\sigma^2 + \sum_{k=1}^n \beta^2(k)/\sigma^2(k) = \|\varepsilon(t;n)\|^2 / \sigma^4(n) = 1/\sigma^2(n).$$

La condition équivaut à $\sigma^2(\infty) > 0$, c'est-à-dire $X(\cdot)$ non singulier, et la limite est $\varepsilon(t)/\sigma^2$ car $\text{l.i.m. } \varepsilon(t;n) = \varepsilon(t)$. ♦

En normalisant les innovations dans les expressions de la Proposition I.3.6 on obtient :

$$\eta(t;n)/\sigma(n) = \sum_{k=0}^n \alpha(k)\eta^*(t-k;k), \quad n < d, \quad (12)$$

et dans le cas non singulier :

$$\eta(t)/\sigma_{\varepsilon} = \sum_{k=0}^{\infty} \alpha(k)\eta^*(t-k;k), \quad (13)$$

où l'on a posé $\alpha(0) = 1/\sigma$ et $\alpha(k) = -\beta(k)/\sigma(k)$, $k < d$.

I.3.7. Proposition.—*Les relations, compte tenu de conventions évidentes,*

$$\alpha(0) = 1/\sigma, \quad \alpha(k) = -\beta(k)/\left[\sigma^2 \prod_{j=1}^k [1-\beta^2(j)]\right]^{1/2}, \quad k \in \mathbb{N}^*,$$

$$\sigma = 1/\alpha(0), \quad \beta(k) = -\alpha(k)\left[\sum_{j=0}^k \alpha^2(j)\right]^{-1/2}, \quad k \in \mathbb{N}^*,$$

définissent un paramètre $\alpha(\cdot)$ en correspondance biunivoque avec $\{\sigma^2; \beta(\cdot)\}$. Le domaine de variation de $\alpha(\cdot)$ est l'ensemble des fonctions définies sur \mathbb{N} , à valeurs dans $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty; +\infty\}$, de la forme :

$$|\alpha(k)| = +\infty \Rightarrow \alpha(k+1) = \alpha(k).$$

Le processus $X(\cdot)$ est alors non singulier si et seulement si $\alpha(\cdot)$ appartient à \mathcal{P}^2 .

La démonstration est analogue à celle de la Proposition I.3.5 à partir de l'égalité issue de la relation (12) :

$$1/\sigma^2(k) = \sum_{j=0}^k \alpha^2(j), \quad k < d.$$

Le paramètre $\alpha(\cdot)$ est temporel et représente une mesure d'information :

$$\alpha^2(k) = 1/\sigma^2(k) - 1/\sigma^2(k-1), \quad k < d.$$

Dans le cas non singulier les paramètres $\alpha(\cdot)$ et $\beta(\cdot)$ ont le même comportement à l'infini.

Les différentes décompositions considérées dans ce paragraphe ont évidemment une transcription dans le domaine spectral. Dans Geronimus (1960) on reconnaîtra (9), (12), (13) et le Corollaire I.3.3 dans (1.6), (1.9), (2.1) et dans le Théorème 8.2.

§ I.4. Processus harmoniques

Lorsque la dimension d de \mathcal{M} est finie le processus $X(\cdot)$ est dit linéairement singulier d'ordre d et satisfait l'équation aux différences (cf. (2) § I.1) :

$$\sum_{k=0}^d b(k)X(t-k) = 0, t \in \mathbb{Z}.$$

Ce sont les processus dont la fonction d'autocorrélation partielle $\beta(\cdot)$ est du type (ii) de (8). Précisons la forme de leur mesure spectrale notée par Ramsey (1974).

I.4.1. Proposition.—*Une mesure F , réelle positive symétrique sur $[-\pi, +\pi]$, est la mesure spectrale d'un processus $X(\cdot)$ linéairement singulier d'ordre d si et seulement si son support comporte exactement d fréquences distinctes modulo 2π , $-\pi < \lambda_1 < \dots < \lambda_d \leq \pi$. Ces fréquences sont alors définies par les racines du polynôme ϕ_d ,*

$$\phi_d(z) = \sum_{k=0}^d b(k)z^{d-k} = 0 \Leftrightarrow z = e^{i\lambda_k}, k=1, \dots, d,$$

associé à une fonction $\beta(\cdot)$ vérifiant $|\beta(k)| < 1$, $k < d$, et $|\beta(d)| = 1$. De plus la fréquence $\lambda=0$ appartient au support de la mesure F si et seulement si $\beta(d) = +1$.

Démonstration.—Toute mesure F , réelle positive symétrique sur $[-\pi, +\pi]$, est la mesure spectrale d'un processus $X(\cdot)$ pour lequel la norme des polynômes ϕ_k , $k \in \mathbb{N}^*$, est donnée par :

$$\|\phi_k\|^2 = \int_{-\pi}^{+\pi} |\phi_k(e^{i\lambda})|^2 dF(\lambda) = \sigma^2(k) = \sigma^2 \prod_{j=1}^k [1 - \beta^2(j)].$$

Si le processus $X(\cdot)$ est linéairement singulier d'ordre d , la fonction $\beta(\cdot)$ est de la forme annoncée et le support de la mesure F est contenu dans l'ensemble des zéros du polynôme ϕ_d car la fonction $|\phi_d(e^{i\lambda})|^2$, positive continue, satisfait $\|\phi_d\|^2 = \sigma^2(d) = 0$. Réciproquement si le support de la mesure F comporte exactement d fréquences, $-\pi < \lambda_1 < \dots < \lambda_d \leq \pi$, les fonctions $e^{ik\lambda}$ s'identifient, dans $L^2\{[-\pi, +\pi], F\}$, aux vecteurs $(e^{ik\lambda_1}, \dots, e^{ik\lambda_d})'$, $k \in \mathbb{N}$, et le procédé d'orthogonalisation se termine avec un polynôme ϕ_p , $p \leq d$, de norme nulle. La fonction $\beta(\cdot)$ vérifie $|\beta(k)| < 1$, $k < p$, et

$|\beta(p)|=1$. Le processus $X(\cdot)$ est linéairement singulier d'ordre p et $p=d$ d'après la condition nécessaire. Enfin la double récurrence (6) pour $z=1$ conduit à

$$\phi_k(1) = \phi_k^*(1) = \prod_{j=1}^k [1-\beta(j)], \quad k \leq d,$$

et $\beta(d)=1$ équivaut à $\phi_d(e^{i0})=0$. ♦

Notons qu'une équation aux différences

$$\sum_{k=0}^n \tilde{b}(k)X(t-k) = 0, \quad \tilde{b}(0)=1, \quad t \in \mathbb{Z},$$

admet une solution stationnaire non triviale ($X(\cdot) \neq 0$) si et seulement si le polynôme

$$\tilde{\phi}(z) = \sum_{k=0}^n \tilde{b}(k)z^{n-k}$$

possède des racines sur le cercle unité. Le processus $X(\cdot)$ satisfait alors l'équation aux différences d'ordre minimal d définie par les seules racines du polynôme $\tilde{\phi}$ situées sur le cercle unité $|z|=1$. Par la suite seule cette équation sera utilisée et le terme "minimal" sera omis.

L'écriture du polynôme ϕ_d sous la forme

$$\phi_d(z) = \sum_{k=0}^d b(k)z^{d-k} = \prod_{k=1}^d (z - e^{i\lambda_k})$$

montre clairement qu'il caractérise uniquement le support de la mesure F . Pour un choix défini par $d_k^2 = F(\{\lambda_k\})$, $k=1, \dots, d$, le processus $X(\cdot)$ admet la représentation spectrale

$$X(t) = \sum_{k=1}^d Z_k d_k e^{it\lambda_k}, \quad t \in \mathbb{Z},$$

où Z_k , $k=1, \dots, d$, sont des variables aléatoires complexes (réelles si $\lambda_k \in \{0, \pi\}$), centrées, de variance 1, non corrélées et vérifiant $Z_k = \bar{Z}_l$ lorsque $\lambda_k = -\lambda_l$. On peut représenter le processus $X(\cdot)$ sous la forme réelle,

$$X(t) = \sum_k A_k \sin(\lambda_k t + \Phi_k), \quad t \in \mathbb{Z},$$

où la somme porte sur les valeurs k pour lesquelles $\lambda_k \in [0, \pi]$. Le processus $X(\cdot)$

apparaît ainsi comme une superposition de sinusôides dont les amplitudes (positives) A_k et les phases $\Phi_k (\in [-\pi, +\pi])$ sont des variables aléatoires. Il est clair que l'observation d'une trajectoire du processus $X(\cdot)$, y compris sur \mathbb{Z} tout entier, ne détermine, de façon presque sûre, que le support de la mesure F .

I.4.2. Définition.—On appelle processus harmonique tout processus $X(\cdot)$ de la forme

$$X(t) = \sum_{k=1}^p a_k \sin(\lambda_k t + \Phi_k), \quad t \in \mathbb{Z},$$

dans lequel seules les phases $\Phi_k, k=1, \dots, p$, sont aléatoires et sont de plus indépendantes et de loi uniforme sur $[-\pi, +\pi]$.

Un tel processus est évidemment stationnaire, centré, de fonction de covariance

$$\Lambda(t) = \frac{1}{2} \sum_{k=1}^p a_k^2 \cos t \lambda_k, \quad t \in \mathbb{Z}.$$

On peut supposer, sans aucune restriction, que les fréquences λ_k sont dans $[0, \pi]$ avec $0 \leq \lambda_1 < \dots < \lambda_p \leq \pi$. Le processus est linéairement singulier d'ordre d ($2(p-1) \leq d \leq 2p$) et sa mesure spectrale F charge les fréquences λ de la forme $\lambda = \pm \lambda_k \pmod{2\pi}$ avec

$$F(\{\lambda_k\}) = a_k^2 / 4 \quad \text{si } \lambda_k \in]0, \pi[, \quad a_k^2 / 2 \quad \text{sinon.}$$

Un processus harmonique est presque sûrement identifié au vu de son observation sur une durée finie de longueur suffisante. Ses trajectoires sont uniformément bornées par $\sum a_k$ et, par conséquent, il ne peut être gaussien.

Ainsi les modèles harmoniques constituent une classe particulière de processus singuliers qui ne peut se distinguer du cas général à partir d'une seule trajectoire. Il n'en est plus de même lorsque l'on utilise plusieurs trajectoires ou encore lorsque l'on fait des hypothèses, par exemple d'égalité, sur les "énergies" a_k^2 .

CHAPITRE II

DECOMPOSITION DE CHOLESKY DES MATRICES D'AUTOCOVARIANCE ET DE LEURS INVERSES

Les lignes et les colonnes des matrices considérées ici sont indexées à partir de 0. La notation $M = \{f(n,k)\}$ signifie que l'élément $M(n,k)$ de la matrice M est égal à la valeur de f en (n,k) . La matrice transposée de M est notée M' . L'indice m dans M_m rappelle que la matrice est carrée d'ordre m , mais le plus souvent il sera omis, l'ordre étant précisé par le contexte. Ainsi I désigne la matrice identité.

On note $\mathcal{S}(m)$ l'espace vectoriel des matrices réelles symétriques d'ordre m . $\mathcal{S}^+(m)$ et $\overline{\mathcal{S}}^+(m)$ désignent alors respectivement le cône de $\mathcal{S}(m)$ formé des matrices définies positives ($x'Mx > 0, x \in \mathbb{R}^m - \{0\}$) et celui formé des matrices positives ($x'Mx \geq 0, x \in \mathbb{R}^m$). Là encore on omettra de préciser l'ordre m . Si $M \in \mathcal{S}(m)$, $(M)_n, n \leq m$, est l'élément de $\mathcal{S}(n)$ défini par le coin supérieur gauche de M .

Le sous-espace vectoriel de \mathcal{S} constitué par les matrices de Toeplitz, $M(n,k)$ n'est fonction que de $|n-k|$, est noté \mathcal{C} et les cônes associés sont désignés par \mathcal{C}^+ et $\overline{\mathcal{C}}^+$.

On appelle décomposition de Cholesky d'un élément M de \mathcal{S}^+ l'unique factorisation $M = TT'$ dans laquelle T est une matrice triangulaire inférieure à éléments diagonaux positifs. Posons $T = SD$ où D est la matrice diagonale définie par les éléments diagonaux de T (on écrira $D = \text{diag}(T)$). La matrice S est triangulaire inférieure unitaire, $\text{diag}(S) = I$, et la factorisation $M = SD^2S'$ sera, par abus, appelée décomposition de Cholesky de M . De même on désignera ainsi les factorisations de M

qui résultent des précédentes appliquées à M^{-1} et dans lesquelles les matrices triangulaires inférieures sont remplacées par des matrices triangulaires supérieures. Enfin la terminologie est étendue à certaines décompositions canoniques de la même forme d'éléments de \mathcal{P}^+ .

On note Λ la matrice d'autocovariance d'ordre m , $m \in \mathbb{N}^*$, associée au processus $X(\cdot)$:

$$\Lambda = \{\Lambda(n-k), 0 \leq n < m, 0 \leq k < m\}.$$

On précise, dans le premier paragraphe, les décompositions de Cholesky $\Lambda = A \Sigma^2 A'$ et $\Lambda^{-1} = B' \Sigma^{-2} B$, dans lesquelles A est triangulaire inférieure unitaire et $B = A^{-1}$. On considère, dans le paragraphe suivant, les algorithmes de calcul récursifs des facteurs A et B , celui de B étant la base de l'algorithme de Levinson-Durbin. Les factorisations de Λ ou de Λ^{-1} mettent en relief diverses paramétrisations de Λ que nous analysons au paragraphe 3 en soulignant la spécificité du cas des matrices de Toeplitz. Enfin on étudie dans un dernier paragraphe le comportement limite des éléments des facteurs A et B .

§ II.1. Les facteurs des décompositions de Cholesky de Λ et de Λ^{-1}

Les innovations partielles introduites au chapitre précédent conduisent naturellement à une décomposition de Cholesky canonique de la matrice d'autocovariance Λ .

II.1.1. Proposition.—Soit $B = \{b(n, n-k)\}$ la matrice d'ordre m , triangulaire inférieure unitaire, définie par les relations

$$\varepsilon(n; n) = \sum_{k=0}^n b(n, k) X(n-k), \quad 0 \leq n < m, \quad (1)$$

et la convention $b(n, k) = 0$ si $d < k \leq n$. La matrice d'autocovariance Λ d'ordre m se factorise de façon canonique sous la forme $\Lambda = A \Sigma^2 A'$ dans laquelle A est la matrice triangulaire inférieure égale à B^{-1} et Σ^2 est la matrice diagonale associée aux variances des innovations partielles $\sigma^2(n)$, $0 \leq n < m$. Posant $A = \{a(n, n-k)\}$, les

lignes de A conduisent aux relations

$$X(n) = \sum_{k=0}^n a(n,k)\varepsilon(n-k;n-k), 0 \leq n < m. \quad (2)$$

Démonstration.—Notons $\underline{\varepsilon}$ et \underline{X} les vecteurs aléatoires de dimension m :

$$\underline{\varepsilon} = \{\varepsilon(0;0), \varepsilon(1;1), \dots\}', \quad \underline{X} = \{X(0), X(1), \dots\}'.$$

Les relations (1) s'écrivent $\underline{\varepsilon} = B\underline{X}$ et donnent les relations (2) sous la forme $\underline{X} = A\underline{\varepsilon}$ puisque B est inversible. Par construction la suite $\{\varepsilon(0;0), \varepsilon(1;1), \dots\}$ est orthogonale. Elle vérifie $E(\underline{\varepsilon}, \underline{\varepsilon}') = \Sigma^2$ et ainsi $\Lambda = E(\underline{X}, \underline{X}') = A\Sigma^2 A'$. ♦

Les relations (2) caractérisent la décomposition $A\Sigma^2 A'$ où, lorsque Λ est singulière ($m > d$), seules les d premières colonnes de A alors définies contribuent à la reconstitution de Λ . Dans le cas non singulier la décomposition est unique ainsi que celle qu'elle induit pour Λ^{-1} ($m < +\infty$) :

$$\Lambda^{-1} = B'\Sigma^{-2}B.$$

Dans cette décomposition B' est triangulaire supérieure; il ne s'agit donc pas de la décomposition de Cholesky habituelle. En fait celle-ci est une transformation simple de $B'\Sigma^{-2}B$. Pour cela considérons l'opérateur d'inversion J défini par la matrice carrée

$$J = \begin{pmatrix} & & & 1 \\ (0) & \cdot & & \\ & \cdot & & \\ & & (0) & \\ 1 & & & \end{pmatrix}.$$

L'ordre de J sera toujours supposé adéquat dans les opérations effectuées. Pour une matrice M quelconque, dans JM (resp. MJ), l'ordre des lignes (resp. colonnes) est inverse de ce qu'il est dans M et $J^2 = I$. Une matrice carrée M est dite **centrosymétrique** lorsqu'elle vérifie $JMJ = M$. Les matrices qui sont symétriques et centrosymétriques sont **doublement symétriques** (symétriques par rapport à chaque diagonale). Elles forment alors un sous-espace vectoriel de \mathcal{S} stable par inversion que l'on note \mathcal{C} . Soit SD^2S' la décomposition de Cholesky de $M \in \mathcal{C}^+$ où S est triangulaire inférieure unitaire. La deuxième décomposition est alors donnée par

$$M = JMJ = (JSJ)(JD^2J)(JSJ)'$$

L'espace \mathcal{C} étant inclus dans \mathcal{E} , nous pouvons donc énoncer :

II.1.2. Corollaire.—Les décompositions de Cholesky des matrices Λ et Λ^{-1} s'expriment en fonction de A , B et Σ^2 par

$$\Lambda = A\Sigma^2A' = (JAJ)(J\Sigma^2J)(JAJ)', \quad \Lambda^{-1} = B'\Sigma^{-2}B = (JB'J)(J\Sigma^{-2}J)(JB'J)'$$

Les matrices $JB'J$ et JAJ sont évidemment les matrices de passage entre $J\mathcal{X}$ et $J\mathcal{E}$. Ce sont aussi celles que l'on aurait obtenues entre \mathcal{X} et son orthogonalisation basée sur les innovations rétrogrades $\mathcal{E}^* = \{\varepsilon^*(0;m-1), \varepsilon^*(1;m-2), \dots, \varepsilon^*(m-1;0)\}'$.

II.1.3. Remarque.—La matrice A étant triangulaire inférieure unitaire, le déterminant de Λ est égal à celui de Σ^2 et s'exprime simplement en fonction de σ^2 et de $\beta(\cdot)$:

$$|\Lambda_m| = \prod_{n=0}^{m-1} \sigma^2(n) = \sigma^{2m} \prod_{n=1}^{m-1} [1 - \beta^2(n)]^{m-n}.$$

§ II.2. Algorithmes de calcul récurrents des facteurs B et A

Les matrices Λ , A , B et Σ^2 d'ordres successifs sont emboîtées, $(\Lambda_m)_n = \Lambda_n$, $n \leq m$. Ce n'est évidemment pas le cas de Λ^{-1} mais son calcul est obtenu sous la forme $B'\Sigma^{-2}B$. L'algorithme de Levinson-Durbin construit, à partir de $\Lambda(\cdot)$, la fonction $\beta(\cdot)$ et simultanément les matrices B et Σ^2 . La construction de B en fonction de $\beta(\cdot)$ est la conséquence directe des relations de récurrence satisfaites par les innovations partielles.

II.2.1. Proposition.—La matrice $B = [b(n, n-k)]$ se construit en fonction de $\beta(\cdot)$ selon la récurrence en ligne : $b(0,0) = 1$,

$$\begin{aligned} 1 \leq n \leq d & & b(n,0) &= 1, \quad b(n,n) = -\beta(n), \\ & & b(n,k) &= b(n-1,k) - \beta(n)b(n-1,n-k), \quad 1 \leq k < n, \\ n > d & & b(n,k) &= b(d,k), \quad 0 \leq k \leq d, \\ & & b(n,k) &= 0, \quad d < k \leq n. \end{aligned}$$

Soulignons que les valeurs de $\beta(\cdot)$ apparaissent en première colonne de B , $b(n,n) = -\beta(n)$, $1 \leq n \leq d$. Pour $1 \leq n \leq d$, la récurrence ci-dessus est plus expressive

sous la forme vectorielle

$$\begin{pmatrix} 1 \\ b(n,1) \\ \cdot \\ \cdot \\ b(n,n) \end{pmatrix} = \begin{pmatrix} 1 \\ b(n-1,1) \\ \cdot \\ b(n-1,n-1) \\ 0 \end{pmatrix} - \beta(n) \begin{pmatrix} 0 \\ b(n-1,n-1) \\ \cdot \\ b(n-1,1) \\ 1 \end{pmatrix}$$

soit, en posant $b(n,.) = [1, b(n,1), \dots, b(n,n)]'$ et $\tilde{b}(n,.) = [b(n,.)', 0]'$,

$$b(n,.) = \tilde{b}(n-1,.) - \beta(n) \mathcal{J} \tilde{b}(n-1,.)$$

Le vecteur $b(n,.)$ apparaît en dernière ligne de B_{n+1} et constitue les paramètres du modèle AR(n) associé à Λ_{n+1} (cf. Ch. III). La finalité de l'algorithme de Levinson-Durbin est précisément de calculer $b(n,.)$ défini par les relations de Yule-Walker :

$$\Lambda b(n,.) = (\sigma^2(n), 0, \dots, 0)'$$

La construction de B est évidemment récursive par rapport à l'ordre de la matrice et permet d'obtenir aisément A par inversion. Le Roux et Guéguen (1977) donnent un algorithme pour construire, à partir de $\Lambda(\cdot)$, la fonction $\beta(\cdot)$ en introduisant un autre système de variables que B . Cela équivaut à construire $\beta(\cdot)$ et simultanément les matrices A et Σ^2 .

II.2.2. Proposition.—Soit E^* la matrice triangulaire inférieure d'ordre m , $E^* = A\Sigma^2$.

Elle se construit en fonction de $\Lambda(\cdot)$ selon la récurrence en colonne :

$$E^*(n,0) = E(n,0) = \Lambda(n), \quad n=0, \dots, m-1,$$

et pour $k=1, \dots, d_\Lambda(m-1)$:

$$\beta(k) = E(k,k-1) / E^*(k-1,k-1),$$

$$E^*(n,k) = E^*(n-1,k-1) - \beta(k)E(n,k-1)$$

$$n = k, \dots, m-1.$$

$$E(n,k) = E(n,k-1) - \beta(k)E^*(n-1,k-1)$$

Démonstration.—De la relation $\Lambda = A\Sigma^2A'$ on obtient $B\Lambda = \Sigma^2A' = (E^*)'$. Ainsi la colonne $E^*(.,k)$ de E^* est donnée par

$$E^*(n,k) = \sum_{j=0}^k b(k,k-j)\Lambda(n-j) = \sigma^2(k)A(n,k) = \sigma^2(k)a(n,n-k), \text{ si } k \leq n < m, 0 \text{ sinon.}$$

Soit, avec des notations inspirées de celles des innovations partielles,

$$E^*(n,k) = \sum_{j=0}^k b(k,j)\Lambda(n-k+j) = e^*(n-k;k).$$

On a $E^*(k,k)=e^*(0;k)=\sigma^2(k)$ ainsi que les récurrences

$$e^*(n-k;k) = e^*(n-k;k-1) - \beta(k)e(n;k-1)$$

$$e(n;k) = e(n;k-1) - \beta(k)e^*(n-k;k-1)$$

à condition de poser

$$e(n,k) = \sum_{j=0}^k b(k,j)\Lambda(n-j). \quad (3)$$

Ce sont les relations annoncées avec $E(n,k)=e(n;k)$ et $\beta(k)$ est obtenu à partir de l'étape précédente en notant, par exemple, la relation

$$0 = e(k,k) = e(k;k-1) - \beta(k)e^*(0;k-1). \spadesuit$$

Cette construction de E^* n'est évidemment pas récursive selon l'ordre de la matrice. La construction en ligne qui lui correspond est récursive mais utilise, à chaque étape, l'ensemble des valeurs de $\beta(\cdot)$ calculées aux étapes précédentes.

II.2.3. Corollaire.—*La matrice $E^*=A\Sigma^2$ se construit à partir de $\Lambda(\cdot)$ de manière récursive par rapport à son ordre selon la récurrence en ligne, $n \in \mathbb{N}$:*

$$E^*(n,0) = E(n,0) = \Lambda(n),$$

pour $1 \leq k < n \wedge d$:

$$E^*(n,k) = E^*(n-1,k-1) - \beta(k)E(n,k-1)$$

$$E(n,k) = E(n,k-1) - \beta(k)E^*(n-1,k-1),$$

avec, si $n < d$:

$$\beta(n) = E(n,n-1) / E^*(n-1,n-1),$$

$$E^*(n,n) = E^*(n-1,n-1) - \beta(n)E(n,n-1).$$

Les paramètres introduits par Le Roux et Guéguen sont les valeurs de $e(n;k)$ définies par (3) pour $k \in \mathbb{N}$ et $n \in \mathbb{Z}$ et on a

$$e(n,k) = \langle \varepsilon(n;k), X(0) \rangle = \begin{cases} E^*(k-n,k), & n \leq 0 \\ 0, & n=1, \dots, k \\ E(n,k), & n > k. \end{cases}$$

§ II.3. Paramétrisations de Λ liées aux décompositions de Cholesky

On suppose ici que la matrice d'autocovariance Λ est d'ordre $(p+1)$ et qu'elle est non singulière ($p < d$). L'objet de ce paragraphe est d'analyser diverses paramétrisations de Λ par rapport à la structure même de la matrice en soulignant les points spécifiques à la forme de Tœplitz. Pour cela remarquons que les lignes de la matrice $A\Sigma$, dans la décomposition de Cholesky de Λ , définissent, dans \mathbb{R}^{p+1} , une représentation de l'espace vectoriel $\mathfrak{L}(X(0), \dots, X(p))$ selon la correspondance

$$(A\Sigma)(n,.) \leftrightarrow X(n), n=0, \dots, p.$$

Pour ne pas alourdir les notations nous utiliserons systématiquement cette correspondance. Par exemple $X(n;n)$ désigne la projection orthogonale, dans \mathbb{R}^{p+1} , de $(A\Sigma)(n,.)$ sur $\mathfrak{L}\{(A\Sigma)(j,.), j=n-1, \dots, 0\}$ et l'erreur $\varepsilon(n;n) = X(n) - X(n;n)$ est donnée par

$$\varepsilon(n;n) = \sum_{k=0}^n b(n,k)X(n-k) = (0, \dots, \sigma(n), 0, \dots)',$$

où $\{b(n,n), \dots, b(n,0)\}$ est la ligne $B(n,.)$ de la matrice $B=A^{-1}$. Plus généralement $X(0), \dots, X(p)$ pourra représenter tout système de vecteurs admettant Λ comme matrice de produits scalaires.

a) $\{\sigma; \rho(1), \dots, \rho(p)\}$. Dans cette paramétrisation usuelle de Λ , le paramètre $\rho(\cdot)$ est assujéti à la contrainte de positivité, il constitue la première colonne de A sous la forme

$$\rho(n) = a(n,n) = \langle X(0), X(n) \rangle / \|X(0)\|^2, \quad n = 1, \dots, p.$$

La contrainte n'est évidemment pas satisfaite en général lorsque Λ est une matrice de covariance ou de corrélation.

b) $\{\sigma; \beta(1), \dots, \beta(p)\}$. Ici $\beta(n)$ est libre dans $] -1, 1[$ et est donné par la première colonne de B puisque $\beta(n) = -b(n,n)$, $n=1, \dots, p$. Ainsi $\beta(n)$ est le coefficient

de $X(0)$ dans la régression linéaire multiple de $X(n)$ sur $\{X(n-1), \dots, X(0)\}$ et se définit donc par

$$\beta(n) = \langle \varepsilon(n;n-1), \varepsilon^*(0;n-1) \rangle / \|\varepsilon^*(0;n-1)\|^2, \quad n=1, \dots, p.$$

Là encore l'appartenance de $\beta(n)$ à $] -1, 1[$ n'est pas garantie.

c) $\{\alpha(0), \dots, \alpha(p)\}$. Cette paramétrisation forme la première colonne de $\Sigma^{-1}B$ et la contrainte $\alpha(0) > 0$ est satisfaite pour toute matrice Λ de \mathcal{P}^+ .

II.3.1. Proposition.—Soit $B'\Sigma^{-2}B = \Lambda^{-1}$, la décomposition de Cholesky de l'inverse d'une matrice Λ de \mathcal{P}^+ . La première colonne $\{\alpha(0), \dots, \alpha(p)\}$ de $\Sigma^{-1}B$ se caractérise par la relation

$$\varepsilon^*(0;p) / \|\varepsilon^*(0;p)\|^2 = \sum_{n=0}^p \alpha(n) \eta(n;n).$$

Dans la paramétrisation $\{\sigma; \beta(1), \dots, \beta(p)\}$ correspondante, $\sigma = 1/\alpha(0) = \|X(0)\|$, et σ^2 est la "variance" de $X(0)$; celle de $\varepsilon^*(0;k)$ est donnée par :

$$\|\varepsilon^*(0;k)\|^2 = \left[\sum_{n=0}^k \alpha^2(n) \right]^{-2} = \sigma^2 \prod_{n=1}^k [1 - \beta^2(n)], \quad k=1, \dots, p,$$

où $\beta(k)$ représente la "corrélation partielle" entre $X(0)$ et $X(k)$ dans l'ensemble $\{X(0), \dots, X(k)\}$,

$$\beta(k) = \langle \eta^*(0;k-1), \eta(k;k-1) \rangle, \quad k=1, \dots, p.$$

Démonstration.—La relation $(\Sigma^{-1}B)(A\Sigma) = I$ montre que $\alpha(n)$ est le coefficient de $X(0)$ dans la décomposition de $\eta(n;n)$ sur $\{X(0), \dots, X(n)\}$. Ainsi pour $k \geq n$ on a

$$\langle \varepsilon^*(0;k), \eta(n;n) \rangle = \alpha(n) \langle \varepsilon^*(0;k), X(0) \rangle = \alpha(n) \|\varepsilon^*(0;k)\|^2,$$

d'où l'expression de $\varepsilon^*(0;k)$:

$$\varepsilon^*(0;k) / \|\varepsilon^*(0;k)\|^2 = \sum_{n=0}^k \alpha(n) \eta(n;n), \quad k=0, \dots, p.$$

Ceci justifie la caractérisation de $\alpha(\cdot)$ et les expressions de $\|\varepsilon^*(0;k)\|$, $k=0, \dots, p$. La correspondance avec $\beta(\cdot)$ donne

$$\begin{aligned}\beta(k) &= -\alpha(k) \left[\sum_{n=0}^k \alpha^2(n) \right]^{1/2} = -\alpha(k) \|\varepsilon^*(0;k)\| \\ &= -\langle \varepsilon^*(0;k), \eta(k;k) \rangle / \|\varepsilon^*(0;k)\| = -\langle \eta^*(0;k), \eta(k;k) \rangle.\end{aligned}$$

Il suffit donc de vérifier les égalités

$$\langle \eta^*(0;k-1), \eta(k;k-1) \rangle = -\langle \eta^*(0;k), \eta(k;k) \rangle, \quad k = 1, \dots, p.$$

En notant β_k la "corrélation partielle" que représente le premier membre, ces égalités sont une conséquence des relations générales

$$\begin{aligned}\eta^*(0;k) &= \{\eta^*(0;k-1) - \beta_k \eta(k;k-1)\} / (1-\beta_k^2)^{1/2}, \\ \eta(k;k) &= \{\eta(k;k-1) - \beta_k \eta^*(0;k-1)\} / (1-\beta_k^2)^{1/2}. \quad \blacklozenge\end{aligned}$$

d) $\{\sigma(p); b(1), \dots, b(p)\}$. Nous notons ainsi la paramétrisation du modèle AR(p) qui admet Λ comme matrice d'autocovariance d'ordre ($p+1$) (cf. Ch. III). Elle apparaît en dernière ligne de B :

$$\varepsilon(p;p) = \sum_{k=0}^p b(k)X(p-k), \quad \|\varepsilon(p;p)\| = \sigma(p),$$

et constitue la solution du problème de minimisation sous contrainte,

$$\sigma^2(p) = \min \{ \tilde{\mathcal{B}}' \Lambda \tilde{\mathcal{B}} ; \tilde{\mathcal{B}} = [\tilde{\mathcal{B}}(p), \dots, \tilde{\mathcal{B}}(1), 1] \}.$$

II.3.2. Remarque.—Le paramètre $b(\cdot)$ doit satisfaire la contrainte de stabilité

$$\sum_{k=0}^p b(k)z^k = 0 \Rightarrow |z| > 1,$$

qui équivaut à ce que la récurrence

$$b(p,k) = b(k), \quad k = 0, \dots, p, \quad \beta(p) = -b(p,p),$$

et pour $n = p, \dots, 2$

$$b(n-1,k) = [b(n,k) + \beta(n)b(n,n-k)] / [1-\beta^2(n)], \quad k=1, \dots, n-1,$$

$$\beta(n-1) = -b(n-1,n-1),$$

construise $\beta(\cdot)$ à valeurs dans $] -1, 1[$.

Cette condition n'est pas garantie lorsque Λ est dans \mathcal{Y}^+ .

e) $(\sigma(p); \gamma(1), \dots, \gamma(p))$. Soit P l'opérateur défini par la matrice

$$P = \begin{pmatrix} 0 & 1 & & & \\ \cdot & \cdot & \cdot & (0) & \\ \cdot & & \cdot & \cdot & \\ 0 & & & \cdot & 1 \\ 1 & 0 & \cdot & \cdot & 0 \end{pmatrix},$$

Px effectue une permutation circulaire des composantes de x , la première devenant la dernière et $PP'=I$.

II.3.3. Proposition.—Pour $\Lambda \in \mathcal{J}^+$ les coefficients de la décomposition orthogonale

$$X(p) = \sum_{k=1}^p \gamma(k) \eta^{*(p-k; k-1)} + \sigma(p) \eta(p; p)$$

constituent la dernière ligne de la matrice triangulaire inférieure $\tilde{A}\Sigma$ de la décomposition de Cholesky $PJ\Lambda JP' = \tilde{A}\Sigma^2\tilde{A}'$. Dans la paramétrisation $\{\sigma; \beta(1), \dots, \beta(p)\}$ correspondante, σ^2 est la "variance" de $X(p)$,

$$\sigma^2 = \sum_{k=1}^p \gamma^2(k) + \sigma^2(p) = \|X(p)\|^2,$$

celle de $\varepsilon(p; k)$ est donnée par

$$\|\varepsilon(p; k)\|^2 = \sum_{n=k+1}^p \gamma^2(n) + \sigma^2(p) = \sigma^2 \prod_{n=1}^k [1 - \beta^2(n)], \quad k=1, \dots, p,$$

où $\beta(k)$ représente la "corrélation partielle" entre $X(p)$ et $X(p-k)$ dans l'ensemble $\{X(p), \dots, X(p-k)\}$,

$$\beta(k) = \langle \eta(p; k-1), \eta^{*(p-k; k-1)} \rangle, \quad k=1, \dots, p.$$

Démonstration.—Par définition $\eta = \{\eta^{*(p-1; 0)}, \dots, \eta^{*(0; p-1)}, \eta(p; p)\}'$ et $X = \{X(0), \dots, X(p)\}'$ sont liés par la relation $PJX = T\eta$ où T est une matrice triangulaire inférieure. Ainsi $PJ\Lambda JP' = TT' = \tilde{A}\Sigma^2\tilde{A}'$ et $\{\gamma(1), \dots, \gamma(p), \sigma(p)\}$ est la dernière ligne de $\tilde{A}\Sigma$. Cette décomposition orthogonale donne

$$\varepsilon(p; k) = \sum_{n=k+1}^p \gamma(n) \eta^{*(p-n; n-1)} + \sigma(p) \eta(p; p), \quad k=0, \dots, p,$$

d'où les expressions de σ^2 et de $\|\varepsilon(p; k)\|^2$, $k=1, \dots, p$, avec

$$\begin{aligned}\beta(k) &= \gamma(k) / \left\{ \sum_{n=k}^p \gamma^2(n) + \sigma^2(p) \right\}^{1/2} = \langle \varepsilon(p; k-1), \eta^*(p-k; k-1) \rangle / \|\varepsilon(p; k-1)\| \\ &= \langle \eta(p; k-1), \eta^*(p-k; k-1) \rangle. \quad \blacklozenge\end{aligned}$$

La factorisation $P\Lambda P' = \tilde{A}\Sigma^2\tilde{A}'$ correspond à la décomposition orthogonale

$$X(0) = \sum_{k=1}^p \gamma(k)\eta(k; k-1) + \sigma(p)\eta^*(0; p),$$

et conduit à la même paramétrisation $\{\sigma; \beta(1), \dots, \beta(p)\}$ que celle donnée par $\alpha(\cdot)$ dans la Proposition II.3.1. Les deux propositions conduisent également au même résultat lorsque $\Lambda \in \mathcal{E}^+$ puisque $J\Lambda J = \Lambda$. Lorsque Λ appartient à \mathcal{E}^+ la dernière ligne de \tilde{A} est égale à $\{\beta(1), \dots, \beta(p), 1\}$ et $\{b(1), \dots, b(p), 1\}$ constitue la dernière ligne de \tilde{A}^{-1} , les contraintes n'étant plus garanties pour Λ dans \mathcal{S}^+ .

§ II.4. Comportement limite des éléments des facteurs A et B

Considérons tout d'abord le cas d'un processus $X(\cdot)$ singulier d'ordre d . La convention adoptée assure la convergence des suites $b(n, \cdot)$ avec

$$\lim_n b(n, k) = \begin{cases} b(k), & 0 \leq k \leq d, \\ 0, & k > d. \end{cases}$$

Les suites $a(n, \cdot)$ sont alors définies par $AB=I$ et satisfont pour $k \in \mathbb{N}^*$,

$$a(n, k) = -\sum_{j=1}^{k \wedge d} b(n-k+j, j)a(n, k-j), \quad n > k.$$

Un raisonnement par récurrence montre que $a(n, \cdot)$ converge simplement vers la fonction pseudo-périodique $a(\cdot)$ définie par

$$\sum_{j=0}^d b(j)a(k-j) = 0, \quad k \geq d,$$

et les conditions initiales

$$a(k) = -\sum_{j=1}^k b(j)a(k-j), \quad 1 \leq k < d, \quad a(0) = 1.$$

Rappelons que le polynôme

$$\mathfrak{B}(z) = \sum_{k=0}^d b(k)z^k$$

a ses racines sur le cercle unité (cf. § 1.4). Il ne s'annule donc par dans $|z| < 1$ et les relations définissant $a(\cdot)$ ci-dessus équivalent à

$$1 / \mathfrak{B}(z) = \mathfrak{A}(z) = \sum_{k=0}^{\infty} a(k)z^k, \quad |z| < 1.$$

Ainsi, dans cette situation particulière, $b(n, \cdot)$ converge dans \mathcal{P}^2 vers $b(\cdot)$. L'"inverse" $a(\cdot)$, limite simple de $a(n, \cdot)$, est bornée dans \mathcal{P}^∞ mais n'appartient pas à \mathcal{P}^2 .

On suppose désormais que $d = +\infty$. Les suites $a(n, \cdot) = \{a(n, k), k \in \mathbb{N}\}$ d'une part et $b(n, \cdot) = \{b(n, k), k \in \mathbb{N}\}$ d'autre part, dans lesquelles on a posé $a(n, k) = b(n, k) = 0$ pour $k > n$, caractérisent la structure de $X(\cdot)$. L'égalité $AB = I$ implique, pour $k \in \mathbb{N}^*$,

$$\sum_{j=0}^k a(n, j)b(n-j, k-j) = 0, \quad n > k.$$

Ainsi, quelque soit K dans \mathbb{N} , l'existence d'une limite sur $\{0, \dots, K\}$ de $a(n, \cdot)$ équivaut à celle de $b(n, \cdot)$ sur le même domaine et les deux limites sont alors liées par

$$\sum_{j=0}^k a(j)b(k-j) = 0, \quad 1 < k \leq K. \quad (4)$$

La récurrence satisfaite par $b(n, \cdot)$ donne

$$b(n, 1) = b(n-1, 1) + \beta(n)\beta(n-1), \quad n \geq 2.$$

Donc $b(n, 1)$ peut tendre vers l'infini ou encore osciller et il en sera alors de même de $a(n, 1) = -b(n, 1)$.

La décomposition de Wold d'un processus non singulier $X(\cdot)$ est notée

$$X(t) = U(t) + V(t) = \sum_{k=0}^{\infty} a(k)\epsilon(t-k) + V(t), \quad a(0) = 1, \quad t \in \mathbb{Z}.$$

Le processus d'innovation $\epsilon(\cdot)$ ne dépend que de la partie régulière $U(\cdot)$. Cette décomposition est unique et la fonction

$$\mathfrak{A}(z) = \sum_{k=0}^{\infty} a(k)z^k$$

ne s'annule pas dans $|z| < 1$ définissant ainsi la suite $b(\cdot)$ par

$$1 / \mathfrak{A}(z) = \mathfrak{B}(z) = \sum_{k=0}^{\infty} b(k)z^k, \quad |z| < 1.$$

II.4.1. Proposition.—*Pour un processus non singulier $X(\cdot)$ les lignes $a(n, \cdot)$ de A sont bornées dans \mathfrak{P}^2 par $\sigma^2 / \sigma_\varepsilon^2$ et convergent simplement vers la suite $a(\cdot)$ de la décomposition de Wold. Les suites $b(n, \cdot)$ convergent alors simplement vers l'«inverse» $b(\cdot)$ qui vérifie aussi*

$$b(k) = -\sum_{n=k}^{\infty} \beta(n)b(n-1, n-k), \quad k \in \mathbb{N}^*,$$

et les colonnes de B sont dans \mathfrak{P}^2 .

Démonstration.—Considérons les décompositions orthogonales de $X(t)$ données par la Proposition II.1.1:

$$X(t) = \sum_{k=0}^n a(n, k)\varepsilon(t-k; n-k), \quad n \in \mathbb{N}.$$

On a

$$\sigma^2 = \sum_{k=0}^n a^2(n, k)\sigma^2(n-k) \geq \sigma^2(n) \sum_{k=0}^n a^2(n, k),$$

d'où la majoration annoncée. La variable $\varepsilon(t-k; n-k)$, k fixé dans \mathbb{N} , converge au sens L^2 vers $\varepsilon(t-k)$ lorsque n tend vers l'infini et par suite

$$\lim a(n, k) = \lim \langle X(t), \varepsilon(t-k; n-k) \rangle / \sigma^2(n-k) = \langle X(t), \varepsilon(t-k) \rangle / \sigma_\varepsilon^2 = a(k).$$

Ainsi $b(n, k)$ converge vers $b(k)$ satisfaisant la relation (4). L'expression de $X(t; m)$ donnée par la Proposition I.3.2 se développe sous la forme

$$X(t; m) = -\sum_{k=1}^m b(m, k)X(t-k) = \sum_{n=1}^m \beta(n) \sum_{j=0}^{n-1} b(n-1, j)X(t-n+j).$$

Identifions le coefficient de $X(t-k)$:

$$b(m, k) = -\sum_{n=k}^m \beta(n)b(n-1, n-k).$$

Par suite la limite du second membre ci-dessus existe et donne l'expression de $b(k)$. Cette expression apparaît comme le produit scalaire dans \mathfrak{P}^2 de la $(k+1)$ ème colonne de B avec $(\beta(1), \beta(2), \dots)$. En effet par hypothèse $\beta(\cdot) \in \mathfrak{P}^2$ et donc la première

colonne de B aussi. Supposons qu'il en soit ainsi pour les k premières colonnes. La suivante est donnée par $b(k,0)=1$ et

$$b(n,n-k) = b(n-1,n-k) - \beta(n)b(n-1,k), \quad n > k.$$

La suite $\{b(n-1,k), n > k\}$ converge vers $b(k)$, elle est donc bornée et on peut écrire

$$\sum_{n=k}^{\infty} b^2(n,n-k) \leq 2 \sum_{n=k}^{\infty} b^2(n-1,n-k) + 2(\sup_{n>k} b^2(n-1,k)) \sum_{n=k+1}^{\infty} \beta^2(n)$$

en utilisant l'inégalité $(a+b)^2 \leq 2(a^2+b^2)$. Ceci prouve, par récurrence, que les colonnes de B sont dans \mathcal{P}^2 . ♦

Cette proposition indique le comportement limite des éléments de A et de B lorsque l'on se déplace parallèlement à la diagonale principale. L'appartenance à \mathcal{P}^2 des colonnes de B implique que la limite sur chaque colonne est nulle :

$$\lim_n b(n,n-k) = 0, \quad k \in \mathbb{N}.$$

Cependant $b(\cdot)$ n'est pas nécessairement dans \mathcal{P}^2 . En particulier si $X(\cdot)$ est une moyenne mobile d'ordre d

$$X(t) = \sum_{k=0}^d a(k)\varepsilon(t-k), \quad t \in \mathbb{Z},$$

et si le polynôme

$$\sum_{k=0}^d a(k)z^k, \quad a(0)=1,$$

a ses racines sur le cercle unité, $b(\cdot)$ est la fonction pseudo-périodique associée à $a(\cdot)$.

Le résultat ci-après améliore la Proposition 3.2 de Dégerine, 1982b.

II.4.2. Proposition.—Une condition nécessaire et suffisante pour qu'un processus stationnaire $X(\cdot)$ soit régulier est que $a(n,\cdot)$ admette une limite $a(\cdot)$ au sens \mathcal{P}^2 :

$$\lim_{n \rightarrow +\infty} \sum_{k=0}^n [a(k) - a(n,k)]^2 = 0.$$

Démonstration.—Considérons tout d'abord le cas d'un processus non singulier et montrons que $V_n(t)$ défini par

$$V_n(t) = \sum_{k=0}^n [a(n,k) - a(k)] \varepsilon(t-k; n-k)$$

converge au sens L^2 vers la partie singulière $V(t)$ de la décomposition de Wold. Pour tout N fixé dans \mathbb{N} et $n > N$ on a

$$V(t) - V_n(t) = \sum_{k=0}^N a(k) [\varepsilon(t-k; n-k) - \varepsilon(t-k)] + \sum_{k=N+1}^n a(k) \varepsilon(t-k; n-k) - \sum_{k=N+1}^{\infty} a(k) \varepsilon(t-k).$$

Chaque terme du second membre se majore comme suit :

$$\begin{aligned} \left\| \sum_{k=0}^N a(k) [\varepsilon(t-k; n-k) - \varepsilon(t-k)] \right\| &\leq \sum_{k=0}^N |a(k)| \|\varepsilon(t-k; n-k) - \varepsilon(t-k)\|, \\ \left\| \sum_{k=N+1}^n a(k) \varepsilon(t-k; n-k) \right\|^2 &= \sum_{k=N+1}^n a^2(k) \sigma^2(n-k) \leq \sigma^2 \sum_{k=N+1}^{\infty} a^2(k), \\ \left\| \sum_{k=N+1}^{\infty} a(k) \varepsilon(t-k) \right\|^2 &= \sigma_\varepsilon^2 \sum_{k=N+1}^{\infty} a^2(k) \leq \sigma_\varepsilon^2 \sum_{k=N+1}^{\infty} a^2(k). \end{aligned}$$

Ainsi, pour $\eta > 0$ donné, choisissant N tel que $\sigma^2 \sum_{k=N+1}^{\infty} a^2(k) \leq \eta^2$, la convergence au sens L^2 de $\varepsilon(t-k)$, $k=0, \dots, N$, assure l'existence de $Q(N)$ pour lequel

$$\sum_{k=0}^N |a(k)| \|\varepsilon(t-k; n-k) - \varepsilon(t-k)\| < \eta$$

dès que $n > Q(N)$. On a alors $\|V(t) - V_n(t)\| < 3\eta$ pour $n > Q(N)$ et donc la convergence de $V_n(t)$ vers $V(t)$. La majoration de la norme de $V_n(t)$,

$$\|V_n(t)\|^2 = \sum_{k=0}^n [a(n,k) - a(k)]^2 \sigma^2(n-k) \leq \sigma^2 \sum_{k=0}^{\infty} [a(n,k) - a(k)]^2,$$

prouve la condition suffisante de la proposition. Pour la condition nécessaire on écrit

$$\begin{aligned} \sum_{k=0}^{\infty} [a(n,k) - a(k)]^2 &\leq (1/\sigma_\varepsilon^2) \sum_{k=0}^n \sigma^2(n-k) [a(n,k) - a(k)]^2 + \sum_{k=n+1}^{\infty} a^2(k) \\ &= \|V_n(t)\|^2 / \sigma_\varepsilon^2 + \sum_{k=n+1}^{\infty} a^2(k). \end{aligned}$$

Si $V(t)=0$, cette dernière expression tend vers zéro puisque $\|V_n(t)\|^2$ tend vers $\|V(t)\|^2$.

Ainsi si $X(\cdot)$ est régulier, il est non singulier et la condition nécessaire est démontrée.

Soit $a(\cdot)$ la limite dans \mathcal{L}^2 de $a(n, \cdot)$. Alors $d=+\infty$ sinon $a(n, \cdot)$ converge simplement

vers une fonction pseudo-périodique qui ne peut appartenir à \mathcal{P}^2 . Quel que soit N dans \mathbb{N} , on a

$$\begin{aligned}\sigma^2 &= \sum_{k=0}^N a^2(n,k) \sigma^2(n-k) + \sum_{k=N+1}^n a^2(n,k) \sigma^2(n-k) \\ &\leq \sum_{k=0}^N a^2(n,k) \sigma^2(n-k) + \sigma^2 \sum_{k=N+1}^n a^2(n,k), \quad n > N.\end{aligned}$$

Si $X(\cdot)$ était singulier on aurait pour tout N ,

$$\sigma^2 \leq \sigma^2 \sum_{k=N+1}^{\infty} a^2(k),$$

ce qui est impossible puisque $a(\cdot) \in \mathcal{P}^2$. Le processus $X(\cdot)$ est alors non singulier et donc régulier d'après la première partie de la démonstration. ♦

La convergence, autre que simple, de $b(n, \cdot)$ vers $b(\cdot)$ est liée à la représentation autorégressive de $X(\cdot)$ et sera abordée dans le prochain chapitre.

CHAPITRE III

APPROXIMATIONS PAR LES MODELES AUTOREGRESSIFS

Le modèle autorégressif est essentiel dans l'étude statistique des processus stationnaires. L'estimation spectrale autorégressive se justifie, selon Parzen (1974), par un critère L^2 sur la fonction de transfert. Burg (1967) la justifie par un critère d'entropie dont l'aspect est plutôt temporel. Le modèle autorégressif est la solution des moindres carrés dans la prévision temporelle. Il est aussi utilisé comme étape préalable à l'ajustement d'une moyenne mobile (Bhansali, 1983b).

Nous ne considérons pas, dans cette partie, les aspects statistiques. En effet les critères sont souvent développés dans le cadre de la structure probabiliste de $X(\cdot)$ pour être appliqués ensuite à une estimation empirique de cette structure. Notons toutefois que Burg (1968) propose une méthode directe échappant ainsi à l'utilisation habituelle de la covariance empirique ou, de façon équivalente, à celle du périodogramme dans le domaine spectral.

Nous détaillons, dans le premier paragraphe, la présentation du modèle autorégressif d'ordre fini afin de préparer l'étude de la représentation autorégressive infinie abordée au paragraphe suivant. Le dernier paragraphe est consacré à l'analyse des différents critères justifiant l'approximation autorégressive.

§ III.1. Modèle autorégressif d'ordre p : $AR(p)$

On appelle **bruit blanc** une suite de variables aléatoires non corrélées, centrées et de même variance.

III.1.1. Définition.— On dit d'un processus $X(\cdot)$ qu'il est autorégressif d'ordre p lorsqu'il satisfait une équation aux différences stochastique

$$\tilde{\varepsilon}(t) = \sum_{k=0}^p \tilde{b}(k)X(t-k), \quad t \in \mathbb{Z}, \tilde{b}(0)=1, \quad (1)$$

dans laquelle $\tilde{\varepsilon}(\cdot)$ est un bruit blanc et $\tilde{b}(p)$ est différent de zéro.

Dans cette définition il n'est fait aucune hypothèse de non corrélation entre $\tilde{\varepsilon}(t)$ et $X(s)$, $s < t$. La mesure spectrale F d'une solution stationnaire de (1) doit satisfaire

$$\tilde{\sigma}_{\varepsilon}^2 / 2\pi = |\tilde{\phi}^*(e^{i\lambda})|^2 F(d\lambda), \quad \lambda \text{ p.p.},$$

où $\tilde{\sigma}_{\varepsilon}^2$ est la variance du bruit blanc et

$$\tilde{\phi}^*(z) = \sum_{k=0}^p \tilde{b}(k)z^k.$$

Ainsi (1) admet toujours une solution stationnaire. Cependant, si $\tilde{\phi}^*$ s'annule sur le cercle unité $|z|=1$, la variance $\tilde{\sigma}_{\varepsilon}^2$ est nulle et les solutions sont constituées de processus singuliers (cf. § I.4). Nous excluons dorénavant cette possibilité. Alors l'unique solution est celle dont la mesure spectrale, absolument continue, admet pour densité

$$f(\lambda) = (\tilde{\sigma}_{\varepsilon}^2 / 2\pi) |\tilde{\phi}^*(e^{i\lambda})|^{-2}.$$

Le processus $X(\cdot)$ s'exprime comme moyenne mobile de $\tilde{\varepsilon}(\cdot)$,

$$X(t) = \sum_{k=-\infty}^{+\infty} \tilde{a}(k)\tilde{\varepsilon}(t-k), \quad t \in \mathbb{Z},$$

où les coefficients $\tilde{a}(k)$, $k \in \mathbb{Z}$, sont ceux du développement de Laurent de $\tilde{\phi}^{*-1}$:

$$1 / \tilde{\phi}^*(z) = \sum_{k=-\infty}^{+\infty} \tilde{a}(k)z^k.$$

Soit α une racine de $\tilde{\phi}^*$ située dans le disque unité $|z| < 1$. Le facteur $(z-\alpha)$ a même module sur le cercle unité que $(1-\alpha z)$. Ainsi il est possible de modifier $\tilde{\phi}^*$ de sorte qu'il n'ait plus de racine à l'intérieur du disque sans modifier la densité spectrale f . Pour chaque racine α ainsi modifiée la variance $\tilde{\sigma}_{\varepsilon}^2$ est multipliée par $|\alpha|^2$ afin de conserver $b(0)=1$. Une racine non réelle est accompagnée de la racine conjuguée et le polynôme $\tilde{\phi}^*$ reste à coefficients réels.

III.1.2. Proposition.— *Un processus stationnaire régulier $X(\cdot)$, autorégressif d'ordre p , satisfait l'équation aux différences canonique,*

$$\varepsilon(t) = \sum_{k=0}^p b(k)X(t-k), \quad t \in \mathbb{Z}, \quad b(0)=1, \quad (2)$$

caractérisée par l'une des conditions équivalentes suivantes :

(i) $\phi^*(z) = \sum_{k=0}^p b(k)z^k$ n'a pas de racine dans $|z| \leq 1$,

(ii) $X(\cdot)$ est une moyenne mobile unilatérale de $\varepsilon(\cdot)$,

$$X(t) = \sum_{k=0}^{\infty} a(k)\varepsilon(t-k), \quad t \in \mathbb{Z}, \quad a(0)=1,$$

(iii) $\varepsilon(\cdot)$ est le processus d'innovation i.e. $\langle \varepsilon(t), X(s) \rangle = 0$ si $s < t$,

(iv) $\varepsilon(\cdot)$ est le bruit blanc de variance minimum avec lequel $X(\cdot)$ satisfait une telle équation.

Démonstration.— Soit $X(\cdot)$ un processus stationnaire régulier solution de (1). Nous pouvons construire un polynôme ϕ^* satisfaisant (i) auquel (2) associe un bruit blanc $\varepsilon(\cdot)$. Ce polynôme ϕ^* s'inverse pour donner (ii) qui avec (2) caractérise le processus d'innovation. Ainsi (ii) est la décomposition de Wold de $X(\cdot)$ et les suites $a(\cdot)$ et $b(\cdot)$ sont celles définies au paragraphe II.4. Enfin (iv) résulte de la construction de ϕ^* ou encore de la relation

$$\tilde{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2 + \|\tilde{\varepsilon}(t; \infty)\|^2,$$

où $\tilde{\varepsilon}(t; \infty)$ désigne la projection orthogonale de $\tilde{\varepsilon}(t)$ sur $\mathcal{M}(t-1; \infty)$. ♦

Les coefficients $b(1), \dots, b(p)$ sont les paramètres du modèle AR(p) et il est clair que $\phi^* = \phi^*_k$, $k \geq p$, dans la récurrence (6) du § I.1, d'où la caractérisation en fonction de $\beta(\cdot)$ (Ramsey, 1974).

III.1.3. Proposition.— *Un processus stationnaire $X(\cdot)$ est autorégressif d'ordre p si et seulement si sa fonction d'autocorrélation partielle $\beta(\cdot)$ satisfait*

$$0 < |\beta(p)| < 1, \quad \beta(k) = 0, \quad k > p, \quad \beta(p) \neq 0.$$

§ III.2. Représentation autorégressive

Le problème d'existence d'une représentation autorégressive pour un processus stationnaire $X(\cdot)$ sous forme d'une condition nécessaire et suffisante portant sur sa mesure spectrale n'est pas résolu. Déjà en 1957 Wiener et Masani soulignaient l'absence de solution satisfaisante au problème analogue qu'est celui de l'expression de la prédiction de $X(t)$, $t > 0$, en fonction de $\{X(s), s \leq 0\}$.

Tout d'abord nous rappelons les résultats concernant la représentation de $X(\cdot)$ sous forme de moyenne mobile. Un processus stationnaire $X(\cdot)$ admet une représentation sous forme de moyenne mobile,

$$X(t) = \sum_{-\infty}^{+\infty} \tilde{a}(k) \tilde{\varepsilon}(t-k), \quad t \in \mathbb{Z}, \quad \tilde{a}(0) = 1, \quad (3)$$

où $\tilde{\varepsilon}(\cdot)$ est un bruit blanc non nul, si et seulement si sa mesure spectrale F est absolument continue par rapport à la mesure de Lebesgue. La densité est alors donnée par

$$f(\lambda) = (\tilde{\sigma}_\varepsilon^2 / 2\pi) |\tilde{A}(\lambda)|^2,$$

où $\tilde{A}(\lambda)$ est définie dans L^2 par sa série de Fourier,

$$\tilde{A}(\lambda) = \text{l.i.m.} \sum_{-n}^n \tilde{a}(k) e^{ik\lambda}.$$

Dans (3) le bruit blanc $\tilde{\varepsilon}(\cdot)$ est réel et il en est alors de même pour les coefficients de Fourier de \tilde{A} . Cependant \tilde{A} est le plus souvent à valeurs complexes. Notons le cas réel

$$\tilde{A}(\lambda) = [2\pi f(\lambda)]^{1/2} / \tilde{\sigma}_\varepsilon, \quad \tilde{\sigma}_\varepsilon = (2\pi)^{-1/2} \left| \int_{-\pi}^{+\pi} f^{1/2}(\lambda) d\lambda \right|,$$

qui est celui pour lequel la corrélation entre $X(t)$ et $\tilde{\varepsilon}(t)$ est maximum ou encore $\tilde{\sigma}_\varepsilon$ maximum compte tenu de la convention $\tilde{a}(0)=1$.

Le processus $X(\cdot)$ admet une représentation sous forme de moyenne mobile unilatérale

$$X(t) = \sum_0^{+\infty} \tilde{a}(k) \tilde{\varepsilon}(t-k), \quad t \in \mathbb{Z}, \quad \tilde{a}(0) = 1, \quad (4)$$

si et seulement si il est régulier. Cela équivaut à ce que F soit absolument continue et que le logarithme de f soit intégrable. On a l'inclusion

$$\mathcal{M}_X(t) = \mathfrak{F}\{X(s), s \leq t\} \subset \mathfrak{F}\{\tilde{\varepsilon}(s), s \leq t\} = \tilde{\mathcal{M}}_{\varepsilon}(t),$$

avec égalité si et seulement si $\tilde{\varepsilon}(\cdot)$ est le processus d'innovation $\varepsilon(\cdot)$, et (4) correspondant alors à la décomposition de Wold (cf. § II.4). Le processus $\varepsilon(\cdot)$ est aussi le bruit blanc de variance maximum parmi ceux satisfaisant (4).

Nous proposons la définition suivante :

III.2.1. Définition.— *On dit qu'un processus $X(\cdot)$ admet une représentation autorégressive lorsqu'il existe un bruit blanc $\tilde{\varepsilon}(\cdot)$ tel que*

$$\tilde{\varepsilon}(t) = \text{l.i.m.}_n \sum_0^n \tilde{b}(k)X(t-k), \quad t \in \mathbb{Z}, \quad \tilde{b}(0) = 1. \quad (5)$$

L'hypothèse restrictive dans (5) est que le bruit blanc $\tilde{\varepsilon}(t)$ s'exprime en fonction de $X(s)$, $s \leq t$, sous forme d'une série, car pour tout processus stationnaire non singulier on a (cf. § I.3)

$$\varepsilon(t) = \text{l.i.m.}_n \sum_0^n b(n,k)X(t-k),$$

alors que la série associée à $\varepsilon(t)$,

$$\varepsilon(t) \sim \sum_0^{\infty} b(k)X(t-k), \quad (6)$$

n'est pas nécessairement convergente en moyenne quadratique. Cependant on sait que (6) caractérise la structure de la partie régulière de $X(\cdot)$. D'autre part les sommes partielles satisfont

$$\varepsilon(t) - \sum_0^n b(k)X(t-k) = E_n(t), \quad E_n(t) \in \mathcal{M}_X(t-n-1).$$

Ainsi, pour un processus régulier, on dit qu'il y a convergence au sens où les différences $E_n(t)$ appartiennent aux sous-espaces $\mathcal{M}_X(t-n-1)$ et que ceux-ci vérifient (cf. Rozanov, 1967, § 11, Ch. II)

$$\bigcap_n \mathcal{M}_X(t-n) = \{0\}. \quad (7)$$

En remplaçant, dans la Définition III.2.1, la limite en moyenne quadratique par la condition de convergence ci-dessus, les processus autorégressifs stationnaires s'identifient aux processus réguliers. La notion est alors en dualité parfaite avec celle de moyenne mobile unilatérale. Dans ce cas on a l'inclusion inverse,

$$\tilde{\mathcal{M}}_\varepsilon(t) \subset \mathcal{M}_X(t),$$

avec égalité si et seulement si $\tilde{\varepsilon}(\cdot)$ est le processus d'innovation $\varepsilon(\cdot)$, lequel se caractérise comme le bruit blanc de variance minimum satisfaisant une telle représentation. Ici l'hypothèse restrictive est la condition (7) puisqu'elle équivaut à la régularité de $X(\cdot)$. Il est clair que ce mode de convergence n'implique pas la convergence en moyenne quadratique, ni même la convergence faible. En effet considérons la moyenne mobile

$$X(t) = \varepsilon(t) - 2\varepsilon(t-1) + \varepsilon(t-2), \quad t \in \mathbb{Z}.$$

Le processus $X(\cdot)$ est régulier et les sommes partielles de la représentation

$$\varepsilon(t) \sim \sum_0^\infty (k+1)X(t-k)$$

ont pour expression

$$\sum_0^{n-1} (k+1)X(t-k) = \varepsilon(t) - (n+1)\varepsilon(t-n) + n\varepsilon(t-n-1).$$

Elles ne sont pas bornées dans \mathcal{M}_X et il n'y a donc pas convergence faible.

Cela ne signifie pas pour autant que la Définition III.2.1 soit plus exigeante. D'une part nous verrons qu'un processus satisfaisant une telle représentation n'est pas nécessairement régulier y compris lorsque le processus $\tilde{\varepsilon}(\cdot)$ dans (5) est l'innovation. D'autre part, pour un processus régulier, nous ne savons pas si (5) implique une représentation canonique en fonction du processus d'innovation. Considérons par exemple la moyenne mobile

$$X(t) = \varepsilon(t) - \varepsilon(t-1), \quad t \in \mathbb{Z}.$$

La série associée à $\varepsilon(\cdot)$,

$$\varepsilon(t) \sim \sum_0^{\infty} X(t-k),$$

ne converge pas en moyenne quadratique. Cependant l'existence d'un bruit blanc $\tilde{\varepsilon}(\cdot)$ satisfaisant (5) équivaut à celle d'une fonction de L^2

$$C(\lambda) = \sum_0^{\infty} c(k)e^{ik\lambda}, \quad c(0) = 1,$$

de module constant et dont la série des coefficients de Fourier (réels) soit égale à zéro,

$$\lim_n \sum_0^n c(k) = 0.$$

En effet posant

$$\tilde{b}(k) = \sum_{j=0}^k c(j), \quad k \in \mathbb{N},$$

il vient

$$\sum_0^n \tilde{b}(k)X(t-k) = \sum_0^n c(k)\varepsilon(t-k) - \tilde{b}(n)\varepsilon(t-n-1), \quad n \in \mathbb{N}.$$

D'où la convergence vers un processus $\tilde{\varepsilon}(\cdot)$ de densité spectrale

$$(\sigma_{\varepsilon}^2/2\pi)|C(\lambda)|^2 = \tilde{\sigma}_{\varepsilon}^2/2\pi, \quad \lambda \text{ p.p.}$$

La fonction $(\sigma_{\varepsilon}/\tilde{\sigma}_{\varepsilon})C(\lambda)$, de module 1 p.p., est une fonction intérieure de H^{∞} (cf. Hoffman, 1962). La contrainte exprime que sa série de Fourier est sommable et égale à zéro en $\lambda=0$. Nous ignorons si il existe de telles fonctions. Plus généralement, toute fonction intérieure dont la série de Fourier (à coefficients réels) est sommable et égale à zéro en $\pm\lambda_0$ conduirait au constat analogue pour la moyenne mobile

$$X(t) = \varepsilon(t) - 2\cos\lambda_0\varepsilon(t-1) + \varepsilon(t-2), \quad t \in \mathbb{Z}.$$

Ce dernier point peut être vérifié à partir de la proposition suivante.

III.2.2. Proposition.—*Soit $X(\cdot)$ un processus obtenu comme moyenne mobile d'un processus stationnaire $Y(\cdot)$,*

$$X(t) = \sum_0^p e(k)Y(t-k), \quad e(0)=1,$$

où les racines du polynôme $\psi(z)=\sum e(k)z^k$ sont simples et sur le cercle unité. Une

condition nécessaire et suffisante pour que $X(\cdot)$ admette une représentation autorégressive (5), est que $Y(\cdot)$ satisfasse, avec le même bruit blanc $\tilde{\varepsilon}(\cdot)$, une représentation de même nature,

$$\tilde{\varepsilon}(t) = \sum_0^{\infty} h(k)Y(t-k), \quad h(0)=1,$$

telle que, pour chaque racine $e^{i\lambda}$ de $\psi(\cdot)$, on ait

$$\sum_0^{\infty} h(k)e^{ik\lambda} = 0.$$

Démonstration.—Supposons que $X(\cdot)$ satisfasse (5) et posons

$$h(k) = \sum_{j=0}^{p \wedge k} e^{(j)}\tilde{b}(k-j), \quad k \in \mathbb{N}.$$

La différence entre les sommes partielles des deux séries s'écrit

$$\sum_0^n \tilde{b}(k)X(t-k) - \sum_0^n h(k)Y(t-k) = \sum_{k=1}^p \left[\sum_{j=k}^p e^{(j)}\tilde{b}(n+k-j) \right] Y(t-n-k).$$

La convergence de (5) implique la convergence vers zéro de la suite $(\tilde{b}(k), k \in \mathbb{N})$.

Cela suffit à assurer la convergence en moyenne quadratique vers zéro de la différence des sommes partielles. De plus, pour toute racine $e^{i\lambda}$ de $\psi(\cdot)$ et pour $n > p$, on a

$$\sum_{k=0}^n h(k)e^{ik\lambda} = \sum_{k=n-p+1}^n \tilde{b}(k) \left[\sum_{j=0}^{(n-k) \wedge p} e^{(j)}e^{ij\lambda} \right] e^{ik\lambda}.$$

Ceci prouve la condition nécessaire. Réciproquement, inversons la correspondance entre $h(\cdot)$ et $\tilde{b}(\cdot)$. L'inverse de $\psi(\cdot)$ est défini dans $|z| < 1$ par

$$1/\psi(z) = \sum_0^{\infty} c(k)z^k$$

où $c(\cdot)$ est associé à $e(\cdot)$ par $c(0)=1$ et

$$\sum_{j=0}^{p \wedge k} e^{(j)}c(k-j) = 0, \quad k \in \mathbb{N}^*.$$

Ainsi $C(\cdot)$ est presque-périodique et s'exprime en fonction des racines du polynôme $\psi(\cdot)$:

$$c(k) = \sum_{j=1}^p \gamma(j)e^{ik\lambda_j}, \quad k \in \mathbb{N}, \psi(e^{i\lambda_j})=0, \quad j=1, \dots, p,$$

les valeurs de $\gamma(\cdot)$ étant données par les conditions initiales $c(k), k=0, \dots, p-1$.

L'expression de $\tilde{b}(\cdot)$ en fonction de $h(\cdot)$ se met sous la forme :

$$\tilde{b}(n) = \sum_{j=1}^p \left[\sum_{k=0}^n h(k) e^{-ik\lambda_j} \right] \gamma(j) e^{in\lambda_j}, \quad n \in \mathbb{N}.$$

La contrainte sur $h(\cdot)$ entraîne la convergence vers zéro de $\tilde{b}(n)$ et donc la représentation (5) d'après la partie nécessaire. ♦

La Définition III.2.1 présente encore d'autres difficultés. Considérons la décomposition $F = F^a + F^s + F^d$ de la mesure spectrale F de $X(\cdot)$, où F^a est la partie absolument continue par rapport à la mesure de Lebesgue,

$$F^a(\lambda) = \int_{-\pi}^{\lambda} f(u) du,$$

où F^d est la mesure discrète définie par les sauts de F et où F^s désigne la partie singulière, $F^s = F - F^a - F^d$. Cela induit une décomposition correspondante de $X(\cdot)$ sous forme de somme orthogonale $X(\cdot) = X^a(\cdot) \oplus X^s(\cdot) \oplus X^d(\cdot)$. Lorsque $X(\cdot)$ admet une représentation autorégressive, chaque composante satisfait cette représentation avec la même suite $\tilde{b}(\cdot)$ et un bruit blanc nul pour $X^s(\cdot)$ et $X^d(\cdot)$, le bruit blanc $\tilde{\varepsilon}(\cdot)$ étant associé à $X^a(\cdot)$. Si la densité f est suffisamment régulière, les deux autres composantes seront absentes. Cependant il est possible d'avoir simultanément $X^a(\cdot)$ et $X^d(\cdot)$ et, en toute généralité, nous ne savons pas si l'existence de cette représentation suffit pour exclure la composante singulière $X^s(\cdot)$. D'autre part nous ne savons pas non plus si elle implique la régularité de $X^a(\cdot)$, c'est-à-dire l'intégrabilité du logarithme de f .

III.2.3. Proposition.—*Soit $X(\cdot) = X^a(\cdot) \oplus X^s(\cdot) \oplus X^d(\cdot)$ la décomposition d'un processus stationnaire satisfaisant la représentation autorégressive (5).*

(i) $X^a(\cdot)$ satisfait la même représentation. La mesure spectrale F^a est équivalente à la mesure de Lebesgue et la densité f vérifie

$$f(\lambda) = (\sigma_\varepsilon^2 / 2\pi) |\tilde{B}(\lambda)|^{-2}$$

où $\tilde{B}(\cdot)$ est la limite dans $L^2(F^a)$ des sommes partielles

$$\tilde{B}_n(\lambda) = \sum_0^n \tilde{b}(k)e^{ik\lambda}, \quad n \in \mathbb{N}.$$

(ii) $X^d(\cdot)$ et $X^s(\cdot)$ satisfont (5) avec $\tilde{\varepsilon}(\cdot) \equiv 0$. Les sommes partielles $\tilde{B}_n(\cdot)$ convergent vers zero dans $L^2(Fs)$ et pour tout λ appartenant au support de Fd on a

$$\lim_n \tilde{B}_n(\lambda) = 0.$$

Cette proposition traduit sur le plan spectral la convergence en moyenne quadratique de (5) en séparant les différentes composantes de F . Notons que $X^a(\cdot)$ représente la projection orthogonale de $X(\cdot)$ sur $\mathcal{M}_\varepsilon = \mathcal{F}[\varepsilon(s), s \in \mathbb{Z}]$,

$$X^a(t) = \sum_{-\infty}^{+\infty} \tilde{a}(k)\varepsilon(t-k), \quad t \in \mathbb{Z}.$$

L'expression de f montre qu'elle est strictement positive presque partout mais ceci n'implique évidemment pas l'intégrabilité de son logarithme. Soulignons qu'il n'est pas supposé dans (5) que $\tilde{b}(\cdot)$ appartienne à \mathcal{P}^2 . Nous verrons que cette condition, qui impliquerait alors que l'inverse de f , et a fortiori son logarithme, soit intégrable, n'est pas nécessaire pour que $X(\cdot)$ satisfasse une telle représentation.

Par la suite, les processus pour lesquels nous avons obtenu l'existence d'une représentation autorégressive, sont non singuliers, sans composante singulière $X^s(\cdot)$, et satisfont la représentation canonique en fonction du processus d'innovation. Celle-ci se caractérise de façon analogue au cas $AR(p)$ de la Proposition III.1.2.

III.2.4. Proposition.—Soit $X(\cdot)$ un processus stationnaire satisfaisant la représentation autorégressive canonique avec son processus d'innovation $\varepsilon(\cdot)$,

$$\varepsilon(t) = \sum_0^\infty b(k)X(t-k), \quad t \in \mathbb{Z}, \quad b(0) = 1. \quad (8)$$

Parmi les représentations (5), celle-ci se caractérise par l'une des conditions équivalentes suivantes

(i) la fonction $\phi^*(z) = \sum_0^\infty b(k)z^k$, analytique dans $|z| < 1$, n'a pas de zéro dans ce

domaine et son inverse

$$1/\phi^*(z) = \mathfrak{A}(z) = \sum_0^{\infty} a(k)z^k, \quad |z| < 1,$$

vérifie $\sum_0^{\infty} a^2(k) < +\infty,$

(ii) la composante régulière $X^a(\cdot)$ est une moyenne mobile unilatérale de $\varepsilon(\cdot)$,

$$X^a(t) = \sum_0^{\infty} a(k)\varepsilon(t-k), \quad t \in \mathbb{Z}, \quad a(0) = 1,$$

(iii) $\varepsilon(\cdot)$ est le bruit blanc de variance minimum avec lequel $X(\cdot)$ satisfait (5).

La démonstration de cette proposition est immédiate et seul le point (i) appelle un commentaire plus général. Lorsqu'un processus $X(\cdot)$ satisfait (5) la suite $\{\tilde{b}(k), k \in \mathbb{N}\}$ tend vers zéro et définit ainsi une fonction analytique dans $|z| < 1$,

$$\tilde{\phi}^*(z) = \sum_0^{\infty} \tilde{b}(k)z^k.$$

Dans tout compact $|z| \leq r < 1$, $\tilde{\phi}^*$ possède un nombre fini de zéros, $\alpha_1, \dots, \alpha_n$. Ceux-ci peuvent être éliminés donnant lieu à une nouvelle représentation dans laquelle la variance du bruit blanc est égale à

$$\tilde{\sigma}_{\varepsilon}^2 \prod_1^n |\alpha_k|^2.$$

D'où une condition nécessaire évidente pour que $X(\cdot)$ soit non singulier. Lorsque le produit des modules des zéros de $\tilde{\phi}^*$ est convergent, la nouvelle fonction de transfert $\tilde{\phi}^*$ est bien définie sans que l'on sache pour autant si la série (5) correspondante est convergente. D'autre part si (5) est vérifié avec $\tilde{\phi}^*$ sans zéro dans $|z| < 1$, le bruit blanc n'est pas nécessairement l'innovation et nous ne savons pas si $X(\cdot)$ est non singulier.

Désormais nous considérons uniquement la représentation canonique (8) et nous dirons, dans ce cas, que $X(\cdot)$ est autorégressif. Les conditions suffisantes d'existence de (8) que l'on trouve dans la littérature sont en fait des conséquences de conditions plus générales portant sur la notion de base que forme, selon différents critères, le système $\{X(t), t \in \mathbb{Z}\}$ dans \mathcal{M} .

Nous présentons, tout d'abord, l'approche de Rozanov (1960) en insistant sur l'aspect temporel, précisant ainsi certains points. On dit que le système $\{X(t), t \in \mathbb{Z}\}$ est **minimal** dans \mathcal{M} lorsque, pour tout t de \mathbb{Z} , $X(t)$ n'appartient pas à $\overline{\mathcal{L}}\{X(s), s \in \mathbb{Z}, s \neq t\}$. Pour cela une condition nécessaire et suffisante évidente est qu'il existe un système **conjugué** $\{Y(t), t \in \mathbb{Z}\}$ dans \mathcal{M} vérifiant $\langle X(t), Y(s) \rangle = 1$ si $t=s$, 0 sinon. Cette notion peut être considérée dans l'espace $\mathcal{M}(t)$ pour le système $\{X(s), s \leq t\}$. Dans ce cas le système est minimal si et seulement si $X(\cdot)$ est non singulier et un simple calcul algébrique montre que le système conjugué est donné par

$$Y_{t}(t-s) = (1/\sigma_{\varepsilon}^2) \sum_{k=0}^s b(k) \varepsilon(t-s+k), \quad s \geq 0.$$

Il s'en suit que $\{X(t), t \in \mathbb{Z}\}$ est minimal dans \mathcal{M} si et seulement si $X(\cdot)$ est non singulier et $b(\cdot)$ est dans ℓ^2 , le système conjugué étant

$$Y(t) = (1/\sigma_{\varepsilon}^2) \sum_{k=0}^{\infty} b(k) \varepsilon(t+k).$$

Si de plus $\{Y(t), t \in \mathbb{Z}\}$ est complet dans \mathcal{M} , alors tout élément Z de \mathcal{M} est caractérisé par la série

$$Z \sim \sum_{k=-\infty}^{+\infty} c(k) X(k), \quad c(k) = \langle Z, Y(k) \rangle, \quad k \in \mathbb{Z}, \quad (9)$$

dont la somme, lorsqu'elle converge, est égale à Z . Dans ce cas on dira que le processus $X(\cdot)$ est **minimal**. L'expression de $Y(t)$ montre que le système $\{Y(t), t \in \mathbb{Z}\}$ est minimal dans $\overline{\mathcal{L}}\{\varepsilon(t), t \in \mathbb{Z}\}$ d'où la caractérisation d'un processus minimal.

III.2.5. Proposition.—*Une condition nécessaire et suffisante pour qu'un processus stationnaire $X(\cdot)$ soit minimal est qu'il soit régulier et que l'inverse de sa densité spectrale soit intégrable.*

Le système minimal $\{X(t), t \in \mathbb{Z}\}$ est dit **Hilbertien** si à tout $c(\cdot)$ de ℓ^2 correspond un élément Z de \mathcal{M} satisfaisant (9). La condition nécessaire et suffisante pour qu'il en soit ainsi est que la densité spectrale soit bornée; de plus la série (9) associée

est alors convergente.

III.2.6. Corollaire.—*Un processus stationnaire $X(\cdot)$ est autorégressif lorsqu'il est minimal et que sa densité spectrale est bornée.*

Lorsque la densité spectrale d'un processus régulier est bornée, le système $\{X(s), s \leq t\}$ est Hilbertien dans $\mathcal{M}(t)$ et réciproquement; mais $X(\cdot)$ n'est pas nécessairement autorégressif.

Dans leur article, Wiener et Masani (1958) présument que l'hypothèse $X(\cdot)$ minimal serait suffisante pour qu'il soit autorégressif. Nous ignorons si la condition plus exigeante d'une densité spectrale minorée par un nombre strictement positif est elle-même suffisante mais nous montrerons que $X(\cdot)$ minimal n'est pas une condition nécessaire.

Enfin, pour un processus minimal, il nous semble intéressant de préciser certains aspects du système conjugué $\{Y(t), t \in \mathbb{Z}\}$. Il est clair que $Y(\cdot)$ est égal, à un facteur près, au processus d'erreur d'interpolation de $X(\cdot)$, $\hat{e}(t) = X(t) - \hat{X}(t)$, où $\hat{X}(t)$ désigne la projection de $X(t)$ sur $\mathcal{F}\{X(s), s \neq t\}$. La réversibilité de $X(\cdot)$ conduit à

$$\hat{e}(t) = (1/b^2) \sum_0^{\infty} b(k) \varepsilon^*(t-k), \quad t \in \mathbb{Z}, \quad b^2 = \sum_0^{\infty} b^2(k),$$

où $\varepsilon^*(\cdot)$ est le processus d'innovation rétrograde. La densité spectrale de $\hat{e}(\cdot)$ est égale à

$$(2\pi\sigma_{\varepsilon}^2/b^2)^2 1/f(\lambda).$$

Ainsi le processus d'erreur d'interpolation d'un processus minimal est minimal et son innovation est proportionnelle à l'innovation rétrograde.

L'article de Rozanov (1960) se termine par la notion de base inconditionnelle. Le système $\{X(t), t \in \mathbb{Z}\}$ forme une base inconditionnelle de \mathcal{M} (base de Riesz) si pour tout Z de \mathcal{M} la série (9) converge quel que soit l'ordre de sommation. Cela équivaut, sur le plan spectral, à une densité satisfaisant

$$0 < m < f(\lambda) < M < +\infty, \quad \lambda \text{ p.p.} \quad (10)$$

L'approche de Helson et Szegö (1960) correspond à la notion plus faible de base

de Hamel où l'on demande seulement que la série (9) soit convergente lorsque le procédé de sommation est symétrique

$$Z = \text{l.i.m.} \sum_{-n}^n c(k)X(k).$$

En fait l'article concerne la positivité de l'angle entre le passé $\mathcal{M}(0) = \mathfrak{L}\{X(t), t \leq 0\}$ et le futur $\mathcal{M}^*(1) = \mathfrak{L}\{X(t), t > 0\}$:

$$\sup \langle Z, W \rangle < 1, \quad Z \in \mathcal{M}(0), \quad W \in \mathcal{M}^*(1), \quad \|Z\| = \|W\| = 1.$$

III.2.7. Proposition. — *Le passé $\mathcal{M}(0)$ et le futur $\mathcal{M}^*(1)$ forment un angle positif dans \mathcal{M} si et seulement si la mesure spectrale est absolument continue avec*

$$f(\lambda) = e^{u(\lambda) + \tilde{v}(\lambda)} \quad (11)$$

où u est une fonction réelle bornée, $\|u\|_\infty < +\infty$, et \tilde{v} est la fonction conjuguée d'une fonction réelle v satisfaisant $\|v\|_\infty < \pi/2$.

La fonction conjuguée \tilde{v} d'une fonction v définie par la série trigonométrique $\sum a_n e^{in\lambda}$ est celle associée à $\sum (-i\varepsilon_n) a_n e^{in\lambda}$ où ε_n désigne le signe de n avec $\varepsilon_0 = 0$. La densité f de la forme (11) se caractérise aussi par (cf. Jewell et Bloomfield, 1983)

$$\sup_I \left\{ \frac{1}{\text{III}} \int_I f(\lambda) d\lambda \frac{1}{\text{III}} \int_I [1/f(\lambda)] d\lambda \right\} < +\infty$$

où I représente un intervalle de $[-\pi, +\pi]$ et III sa mesure de Lebesgue. La condition d'angle positif entre le passé et le futur entraîne l'autorégressivité de $X(\cdot)$ puisqu'elle équivaut à ce que $\{X(t), t \in \mathbb{Z}\}$ forme une base (de Hamel) de \mathcal{M} . Cela peut se vérifier directement en notant que la somme partielle

$$\sum_0^n b(k)X(t-k)$$

est la projection de $X(t)$ sur $\mathfrak{L}\{X(t-k), 0 \leq k \leq n\}$ parallèlement à $\mathcal{M}(t-n-1)$. La condition implique aussi la minimalité de $X(\cdot)$.

III.2.8. Corollaire. — *Un processus stationnaire $X(\cdot)$ est autorégressif lorsqu'il est*

minimal et que sa densité spectrale est de la forme (11)

$$f(\lambda) = e^{u(\lambda) + \tilde{v}(\lambda)}$$

Lorsque f est de la forme (11), il en est de même de $1/f$ mais f n'est pas nécessairement bornée comme le montre l'exemple de Helson et Szegö.

III.2.9. Corollaire.— $\mathcal{M}(0)$ et $\mathcal{M}^*(1)$ forment un angle positif dans \mathcal{M} , où

$$f_\alpha(\lambda) = \prod_1^n |e^{i\lambda_j} - e^{i\lambda \alpha_j}|,$$

les points $e^{i\lambda_j}$ étant distincts sur le cercle, si et seulement si $|\alpha_j| < 1$, $j=1, \dots, n$.

Avec la Proposition III.2.2 et ce corollaire on constate que $X(\cdot)$ est autorégressif lorsque sa densité spectrale est de la forme f_α ci-dessus avec $-1 < \alpha_j < 2$, $\alpha_j \neq 1$, $j=1, \dots, n$. Ainsi $X(\cdot)$ minimal n'est pas une condition nécessaire pour qu'il soit autorégressif. D'autre part la représentation (8) associée à f_α est telle que

$$\sum_0^\infty b(k) e^{ik\lambda_j} = 0$$

en tout λ_j où α_j est négatif. Il est alors possible d'adjoindre à $X(\cdot)$ une composante discrète $X^d(\cdot)$ tout en conservant (8). Le cas où, dans f_α , certains α_j sont égaux à 1 n'est pas résolu en toute généralité. On peut montrer directement que $X(\cdot)$ est autorégressif lorsque $f(\lambda) = |1 - e^{i\lambda}|$. En effet on a, pour tout $n \geq 0$,

$$\sum_0^n b(k) X(t-k) = \varepsilon(t) + \sum_{n+1}^\infty c_n(k) \varepsilon(t-k),$$

où $\{c_n(k), k > n\}$ est une suite vérifiant

$$\sup_k |c_n(k)| = |c_n(n+1)| \xrightarrow{n \rightarrow \infty} 0, \quad \sum_{k=n+1}^\infty |c_n(k)| = 1.$$

D'où la majoration

$$\sum_{n+1}^\infty c_n^2(k) < |c_n(n+1)|.$$

La même démonstration vaut pour $|1 + e^{i\lambda}|$.

§ III.3. Les critères d'approximation

La structure du processus stationnaire $X(\cdot)$ étant caractérisée par $\{\sigma^2; \beta(\cdot)\}$ il est naturel d'envisager son approximation par le modèle, noté $AR(\leq p)$, obtenu en annulant les valeurs de $\beta(k)$ pour $k > p$ (cf. Proposition III.1.3). Les deux modèles coïncident lorsque $X(\cdot)$ est singulier d'ordre d avec $d \leq p$ ou autorégressif d'ordre k avec $k \leq p$. Dans le cas général le modèle $AR(\leq p)$ apparaît comme la meilleure approximation selon plusieurs critères.

A) Critère d'entropie maximale

La méthode d'estimation spectrale selon le critère d'entropie maximale a été proposée par Burg en 1967 (cf. Priestley, 1981, p. 604). Elle consiste à choisir la densité spectrale f qui maximise la quantité

$$\int_{-\pi}^{+\pi} \ln f(\lambda) d\lambda = 2\pi \ln(\sigma_{\varepsilon}^2 / 2\pi),$$

sous les contraintes

$$\int_{-\pi}^{+\pi} e^{ik\lambda} f(\lambda) d\lambda = \Lambda(k), k=0, \dots, p,$$

où $\Lambda(k)$, $k=0, \dots, p$ sont des autocovariances données (par exemple empiriques). La quantité σ_{ε}^2 représente la variance de l'innovation $\varepsilon(\cdot)$. L'entropie de la variable $\varepsilon(t)$, centrée et de variance σ_{ε}^2 , est maximisée par la loi gaussienne et égale à $(1/2)\ln(2\pi e \sigma_{\varepsilon}^2)$, d'où le nom du critère. Cependant Burg (1975) constate que la méthode, en dehors de toute considération sur sa loi de probabilité, consiste à maximiser la variance de $\varepsilon(\cdot)$. La formulation suivante est la plus générale possible et sa démonstration est directe.

III.3.1. Proposition.—*Parmi tous les processus stationnaires qui ont même fonction d'autocovariance que $X(\cdot)$ sur $\{0, \dots, p\}$, le modèle $AR(\leq p)$ est celui pour lequel la variance de $\varepsilon(\cdot; \infty)$ est maximum.*

Démonstration.—La connaissance de $\Lambda(\cdot)$ sur $\{0, \dots, p\}$ équivaut à celle de $\{\sigma^2; \beta(1), \dots, \beta(p)\}$. On doit supposer $p < d$, c'est-à-dire $|\beta(p)| < 1$, sans quoi la contrainte définit entièrement $\Lambda(\cdot)$. On a alors (cf. Proposition I.3.2)

$$\sigma^2(\infty) = \sigma^2 \prod_1^{\infty} [1 - \beta^2(k)] \leq \sigma^2 \prod_1^p [1 - \beta^2(k)]$$

avec égalité si et seulement si $\beta(k) = 0$ pour $k > p$. ♦

Notons que le modèle $\text{AR}(\leq p)$ maximise la variance de $\varepsilon(\cdot; n)$ pour tout $n \geq p$.

B) Critère $L^2(F)$ sur la fonction de transfert

Parzen (1974) considère un processus stationnaire $X(\cdot)$ dont la densité spectrale f est continue et ne s'annule pas. Ainsi $X(\cdot)$ est autorégressif et f est caractérisée par σ_ε^2 et la fonction de transfert ϕ_∞^* sous la forme.:

$$f(\lambda) = (\sigma_\varepsilon^2 / 2\pi) |\phi_\infty^*(e^{i\lambda})|^{-2}, \quad \phi_\infty^*(e^{i\lambda}) = \sum_0^{\infty} b(k) e^{ik\lambda}.$$

On a

$$\min_{\Pi_p} (1/2\pi) \int_{-\pi}^{+\pi} \left| \frac{\Pi_p(e^{i\lambda}) - \phi_\infty^*(e^{i\lambda})}{\phi_\infty^*(e^{i\lambda})} \right|^2 d\lambda = 1 - \sigma_\varepsilon^2 / \sigma^2(p),$$

où $\Pi_p(e^{i\lambda}) = \sum_0^p \alpha_k e^{ik\lambda}$ parcourt l'ensemble des polynômes de degré inférieur ou égal à p .

Le minimum est atteint par le polynôme

$$[\sigma_\varepsilon^2 / \sigma^2(p)] \phi_p^*(e^{i\lambda}), \quad \phi_p^*(e^{i\lambda}) = \sum_0^p b(p, k) e^{ik\lambda}.$$

En remplaçant, dans l'expression de f , $\{\sigma_\varepsilon; \phi_\infty^*\}$ par $\{\sigma(p); \phi_p^*\}$ on obtient la densité spectrale f_p du modèle $\text{AR}(\leq p)$. Il est clair, au vu de f , que le problème équivaut à approcher la fonction de transfert par un polynôme de degré inférieur ou égal à p dans $L^2(F)$. La théorie des polynômes orthogonaux (cf. Geronimus, 1960, Ch. II) conduit au même résultat pour un processus non singulier. Dans ce cas la fonction de transfert est définie presque partout sur $[-\pi, +\pi]$ comme limite radiale de sa transformée en z

$$\mathfrak{B}(z) = \sum_0^{\infty} b(k)z^k, |z| < 1.$$

La densité f est celle de la partie absolument continue F^a de F . La fonction ϕ_{∞}^* est dans $L^2(F)$ et le critère est une application directe du développement de Fourier-Chebyshev de ϕ_{∞}^* . On reconnaît, dans la formulation temporelle,

$$\min_{\alpha_0, \dots, \alpha_p} \|\varepsilon(t) - \sum_0^p \alpha_k X(t-k)\|^2 = \|\varepsilon(t) - [\sigma_{\varepsilon}^2 / \sigma^2(p)] \varepsilon(t;p)\|^2 = \sigma_{\varepsilon}^2 [1 - \sigma_{\varepsilon}^2 / \sigma^2(p)],$$

une conséquence de la Proposition I.3.6. En toute généralité on peut écrire (cf. Proposition I.3.2)

$$\min_{\alpha_1, \dots, \alpha_p} \|X(t;\infty) - \sum_1^p \alpha_k X(t-k)\|^2 = \|X(t;\infty) - X(t;p)\|^2 = \sigma^2(p) - \sigma^2(\infty).$$

C) Critère de moindres carrés sur les matrices d'autocovariance

Toute matrice d'autocovariance Λ , non singulière d'ordre n , est définie par $\{\sigma^2; \beta(1), \dots, \beta(n-1)\}$. Pour $p=0, \dots, n-1$, la matrice Λ_p désigne celle d'ordre n associée à $\{\sigma^2; \beta(1), \dots, \beta(p), 0, \dots, 0\}$ ($\Lambda_0 = \sigma^2 I_n$, $\Lambda_{n-1} = \Lambda$).

Si $\Lambda(X)$ est la matrice de covariance de $\{X(0), \dots, X(n-1)\}$, supposée non singulière ($n \leq d$), les matrices $\Lambda_p(X)$, $p=0, \dots, n-1$, sont caractérisées par un problème de minimisation sous contraintes traduisant un problème de moindres carrés (cf. Ch. B.I). Pour une matrice carrée, $\text{tr}(M)$ et $|M|$ désignent respectivement la trace et le déterminant de M .

III.3.2. Proposition.—(Dégerine, 1982b) *Pour toute matrice d'autocovariance $\Lambda(X)$, non singulière d'ordre n , et pour tout p de $\{0, \dots, n-1\}$, le minimum dans*

$$\min_{\Lambda_p} \text{tr}[\Lambda_p^{-1} \Lambda(X)] = n |\Lambda_p(X)|^{1/n},$$

sous la contrainte $|\Lambda_p|=1$, est réalisé uniquement par la matrice

$$|\Lambda_p(X)|^{-1/n} \Lambda_p(X).$$

Démonstration.—Il s'agit d'une application directe d'un résultat général (Lemme B.I.4.1) à condition de noter l'égalité

$$\text{tr}[\Lambda_p^{-1} \Lambda(X)] = \text{tr}[\Lambda_p^{-1} \Lambda_p(X)].$$

Posons $X' = \{X(0), \dots, X(n-1)\}$ et soit $B\Sigma^{-2}B'$ la décomposition de Cholesky de Λ_p^{-1} . On a

$$E\{\|\Sigma^{-1}BX\|^2\} = E\{(\Sigma^{-1}BX)'(\Sigma^{-1}BX)\} = E\{X'\Lambda_p^{-1}X\} = \text{tr}[\Lambda_p^{-1} \Lambda(X)].$$

Chaque composante de $\Sigma^{-1}BX$ dépend de $\{X(0), \dots, X(n-1)\}$ seulement à travers une séquence de longueur inférieure ou égale à $p+1$. Donc $\text{tr}[\Lambda_p^{-1} \Lambda(X)]$ n'est fonction que des valeurs $\sigma^2, \beta(1), \dots, \beta(p)$ de $\Lambda(X)$, d'où l'égalité ci-dessus. ♦

La minimisation sous contrainte considérée dans cette proposition équivaut aussi à la minimisation, sans contrainte sur le déterminant de Λ_p , de la fonction (cf. Ch. B.I)

$$L\{\Lambda_p; \Lambda(X)\} = \ln |\Lambda_p| + \text{tr}[\Lambda_p^{-1} \Lambda(X)].$$

Or, si $\hat{\Lambda}(X)$ désigne la matrice de covariance empirique associée à un échantillon de X supposé $\text{AR}(\leq p)$ et gaussien, la valeur réalisant le minimum de $L\{\Lambda_p; \hat{\Lambda}(X)\}$ est l'estimateur de maximum de vraisemblance de $\Lambda(X) = \Lambda_p(X)$. Ainsi $\Lambda_p(X)$ est la meilleure approximation de $\Lambda(X)$ dans l'ensemble des matrices Λ_p selon le critère de maximum de vraisemblance $L\{\Lambda_p; \Lambda(X)\}$.

Ces trois critères sont équivalents au sens où ils conduisent au même modèle $\text{AR}(\leq p)$. Ils diffèrent cependant dans leur conception. Dans A) on prolonge la fonction de covariance $\Lambda(\cdot)$ au-delà de p et le prolongement associé au modèle $\text{AR}(\leq p)$ apparaît comme étant le plus neutre possible (variance de $\varepsilon(\cdot)$ maximum). Dans les deux autres cas, $\Lambda(\cdot)$ étant fixé, on montre que le modèle $\text{AR}(\leq p)$ le plus proche est celui dont la fonction de covariance coïncide avec $\Lambda(\cdot)$ sur $\{0, \dots, p\}$.

Il est clair que ces considérations plaident en faveur du modèle autorégressif en l'absence d'hypothèse a priori sur la structure de $X(\cdot)$. De plus le comportement limite du modèle $\text{AR}(\leq p)$, lorsque p tend vers l'infini, est en général cohérent avec cette structure, y compris lorsque $X(\cdot)$ n'admet pas de représentation autorégressive. Le Théorème 8.5

de Geronimus (1960) règle le cas de la plupart des situations régulières classiques.

III.3.3. Théorème.—*Si la fonction d'autocorrélation partielle $\beta(\cdot)$ est dans \mathcal{P}^1 ,*

$$\sum_{k=0}^{\infty} |\beta(k)| < +\infty,$$

la mesure spectrale F est absolument continue. La densité spectrale f est continue, minorée par un nombre strictement positif et on a

$$|\phi_{\infty}^*(e^{i\lambda}) - \phi_p^*(e^{i\lambda})| \leq \prod_{k=1}^{\infty} [1 + |\beta(k)|] \sum_{k=p+1}^{\infty} |\beta(k)|.$$

Ici la fonction de transfert ϕ_{∞}^* est définie comme la somme de la série absolument convergente

$$\phi_{\infty}^*(e^{i\lambda}) = \sum_{k=0}^{\infty} b(k)e^{ik\lambda}.$$

En effet $b(\cdot)$ est la limite dans \mathcal{P}^1 des suites $\{b(p, \cdot), p \in \mathbb{N}\}$ et est borné dans \mathcal{P}^1 par

$$\sum_{k=0}^{\infty} |b(k)| \leq \prod_{k=1}^{\infty} [1 + |\beta(k)|].$$

Ceci peut être obtenu en remarquant que la suite $b(p, \cdot)$ est majorée en module par la suite $\tilde{b}(p, \cdot)$ associée à la fonction $\tilde{\beta}(\cdot)$ définie par

$$\tilde{\beta}(k) = -|\beta(k)|, \quad k \in \mathbb{N}^*.$$

Or, pour tout p de \mathbb{N}^* , on a

$$\sum_{k=0}^p b(p, k) = \prod_{k=1}^p [1 - \beta(k)],$$

d'où

$$\sum_{k=0}^p \tilde{b}(p, k) = \prod_{k=1}^p [1 + |\beta(k)|] \leq \prod_{k=1}^{\infty} [1 + |\beta(k)|] < +\infty,$$

et la récurrence $\tilde{b}(p, k) = \tilde{b}(p-1, k) + |\beta(p)|\tilde{b}(p-1, p-k)$ montre que $\tilde{b}(p, \cdot)$ converge vers $\tilde{b}(\cdot)$ en croissant.

Ainsi, lorsque $\beta(\cdot)$ est dans \mathcal{P}^1 , la densité spectrale f_p du modèle $\text{AR}(\leq p)$ converge uniformément vers f et la vitesse de convergence dépend de celle avec laquelle

$\beta(\cdot)$ tend vers zéro

$$\prod_{k=p+1}^{\infty} \frac{1 - |\beta(k)|}{1 + |\beta(k)|} \leq \frac{f(\lambda)}{f_p(\lambda)} \leq \prod_{k=p+1}^{\infty} \frac{1 + |\beta(k)|}{1 - |\beta(k)|}.$$

Les bornes inférieures (resp. supérieures) étant atteintes en $\lambda=0$, lorsque $\beta(\cdot)$ est constamment négatif (resp. positif), cet encadrement ne peut être amélioré.

Un cas particulier important est celui des processus ARMA **inversibles**, c'est-à-dire ceux dont le polynôme MA n'a pas de racine sur le cercle unité. Dans ce cas $\beta(\cdot)$ est dans \mathcal{P}^1 et tend vers zéro de façon exponentielle. Plus généralement c'est le cas de tout processus régulier pour lequel le rayon de convergence R de la série

$$\mathfrak{B}(z) = \sum_{k=0}^{\infty} b(k)z^k$$

est plus grand que 1. En effet la fonction $\mathfrak{B}(z)$ se développe également sur la base des polynômes ϕ_k (cf. Geronimus, Ch. V)

$$\mathfrak{B}(z) = \sigma_{\varepsilon}^2 \left\{ [1/\sigma^2(0)]\phi_0(z) - \sum_{k=1}^{\infty} [\beta(k)/\sigma^2(k)]\phi_k(z) \right\},$$

et son rayon de convergence coïncide avec celui de la série entière (cf. Théorèmes 6.2 et 7.8, Geronimus) donnant ainsi

$$\limsup_n \sqrt[n]{|b(n)|} = \limsup_n \sqrt[n]{|\beta(n)|} = 1/R < 1.$$

Il est essentiel ici de supposer $X(\cdot)$ régulier de sorte que $\beta(\cdot)$ soit l'unique fonction associée à $b(\cdot)$. On peut donc énoncer :

III.3.4. Proposition.—*La fonction d'autocorrélation partielle $\beta(\cdot)$ d'un processus stationnaire décroît vers zéro de façon exponentielle si et seulement si le processus est régulier, autorégressif, et tel que sa fonction d'autorégression $b(\cdot)$ décroît vers zéro de façon exponentielle.*

Lorsque le processus ARMA n'est pas inversible, la convergence de f_p vers f reste uniforme dans tout intervalle $[\alpha, \beta]$ ne contenant pas de zéro de f . Là encore le résultat est une conséquence directe du Théorème 5.5 de Geronimus que l'on peut

formuler ainsi :

III.3.5. Proposition.—Soit $X(\cdot)$ un processus non singulier dont la mesure spectrale F est absolument continue dans l'intervalle $[\alpha, \beta]$. On suppose que la densité spectrale f de la partie régulière vérifie $f(\lambda) \geq m > 0$ dans $[\alpha, \beta]$ et qu'elle y admet une dérivée seconde continue satisfaisant la condition de Dini-Lipschitz

$$|f'(\lambda) - f'(\lambda+h)| \leq C[\ln(1/h)]^{-\alpha}, \quad \alpha > 1, \quad (\lambda, \lambda+h) \in [\alpha, \beta]^2.$$

Alors, pour tout $\varepsilon > 0$, on a

$$\lim_p f_p(\lambda) = f(\lambda), \quad \lambda \in [\alpha+\varepsilon, \beta-\varepsilon],$$

la convergence étant uniforme.

Il reste à considérer le comportement de f_p là où f s'annule ou encore là où F présente un saut (partie discrète de F).

Dans le premier cas, si F est absolument continue au voisinage de λ_0 et que sa densité, nulle en λ_0 , satisfait

$$|f(\lambda_0+h) - f(\lambda_0-h)| \leq Ch^\alpha, \quad \alpha > 0,$$

alors f_p vérifie (cf. Théorème 4.12, Geronimus)

$$\liminf_p f_p(\lambda_0) = 0.$$

Dans le second cas, si F présente une masse en λ_0 , la convergence de $\varepsilon(t;n)$ vers $\varepsilon(t)$ (cf. Proposition I.3.6) implique que l'on a

$$\lim_p f_p(\lambda_0) = +\infty.$$

De plus la masse en λ_0 est donnée par (Geronimus, p. 188)

$$F^d(\{\lambda_0\}) = 2\pi \left\{ \sum_{p=0}^{\infty} [1/f_p(\lambda_0)] \right\}^{-1}.$$

CHAPITRE IV

FONCTIONS D'AUTOCORRELATION PARTIELLE D'UN PROCESSUS VECTORIEL

Nous désignons ici par $X(\cdot)$ un processus vectoriel réel, à temps discret, stationnaire au second ordre et centré; c'est-à-dire une famille $\{X(t), t \in \mathbb{Z}\}$ de vecteurs aléatoires à valeurs dans \mathbb{R}^m , définis sur un espace probabilisé (Ω, \mathcal{F}, P) , dont les deux premiers moments existent et satisfont

$$E\{X(t)\} = 0, t \in \mathbb{Z}, \quad E\{X(t)X(s)'\} = \Lambda(t-s), (t,s) \in \mathbb{Z}^2.$$

Ainsi les composantes de $X(\cdot)$, $X_j(\cdot)$, $j=1, \dots, m$, constituent des processus scalaires stationnaires qui sont stationnairement corrélés entre eux. La fonction de covariance $\Lambda(\cdot)$, à valeurs dans l'espace des matrices carrées réelles d'ordre m , caractérise la structure au second ordre de $X(\cdot)$. Elle est aussi caractéristique de la structure de l'espace de Hilbert

$$\mathcal{M} = \overline{\mathcal{L}}\{X_j(t), j=1, \dots, m, t \in \mathbb{Z}\}$$

dont les éléments sont des variables aléatoires réelles centrées, le produit scalaire étant défini par $\langle U, V \rangle = E\{UV\}$.

Une géométrie plus proche du cas scalaire peut être obtenue en considérant l'espace de Hilbert

$$\mathcal{H} = \overline{\mathcal{L}}\{X(t), t \in \mathbb{Z}\}$$

dont les éléments sont des vecteurs aléatoires à valeurs dans \mathbb{R}^m , le produit scalaire étant défini par $\langle U, V \rangle = \text{tr}[E\{UV'\}]$. Cependant, afin de récupérer la richesse de la structure de \mathcal{M} , il est nécessaire d'enrichir celle de \mathcal{H} en le munissant de la structure de Gram, d'indice m , définie par

a) la multiplication par des matrices : $(A, U) \rightarrow AU \in \mathcal{H}$, où $U \in \mathcal{H}$ et où A est

une matrice carrée réelle d'ordre m , jouissant des mêmes propriétés que la multiplication par un scalaire,

b) le produit intérieur matriciel : $(U, V) \in \mathfrak{R}^2 \rightarrow [U, V]$, où $[U, V] = E\{UV'\}$ est une matrice réelle d'ordre m , ayant les propriétés

- (i) $[U, V] = [V, U]'$,
- (ii) $[\sum A_i U, V] = \sum A_i [U, V]$,
- (iii) $[U, U]$ est positive ($\in \mathfrak{P}^+(m)$),
- (iv) $\langle U, V \rangle = \text{tr}[U, V]$.

La structure ainsi obtenue est proche du cas scalaire mais la multiplication des "scalaires", constitués ici par des matrices, n'est pas commutative et cela est la cause des différences importantes entre le cas vectoriel et le cas scalaire.

On retrouve, dans le premier paragraphe, les innovations progressives et rétrogrades partielles en liaison avec la fonction d'autocovariance partielle. Cette dernière, par normalisation, conduit aux fonctions d'autocorrélation partielle standards présentées au paragraphe suivant. Nous introduisons, dans le dernier paragraphe, la notion de fonction d'autocorrélation partielle canonique.

§ IV.1. Innovations partielles

Nous conserverons, dans la mesure du possible, les mêmes notations que celles introduites au Chapitre I. Pour $t \in \mathbb{Z}$ on considère les sous-espaces de \mathcal{M} suivants :

$$\mathcal{M}(t; n) = \mathfrak{L}\{X_j(s), j=1, \dots, m, t-n < s \leq t\}, \quad n \in \mathbb{N}^*, \quad \mathcal{M}(t; 0) = \{0\}.$$

Lorsque la dimension de \mathcal{M} est finie, $X(\cdot)$ est dit **linéairement singulier d'ordre d** où d est le plus petit entier k pour lequel $X(\cdot)$ satisfait l'équation aux différences

$$\sum_{j=0}^k b(j)X(t-j) = 0, \quad t \in \mathbb{Z}, \quad b(0) = I_m.$$

Le vecteur $X(t; n)$ désigne la "meilleure approximation" de $X(t)$ par son passé immédiat de longueur n : chaque composante $X_j(t; n)$, $j=1, \dots, m$, est la projection orthogonale de $X_j(t)$ sur $\mathcal{M}(t-1; n)$. Avec la convention $X(t; 0)=0$, l'innovation

progressive partielle d'ordre n est définie par

$$\varepsilon(t;n) = X(t) - X(t;n) = \sum_{k=0}^n b(n,k)X(t-k), \quad n \in \mathbb{N}, t \in \mathbb{Z}, \quad (1)$$

où $b(n,0) = I_m$ et où les matrices $b(n,k)$, $k=1, \dots, n$, sont une solution du système traduisant la définition de $X(t;n)$

$$\sum_{k=0}^n b(n;k)\Lambda(j-k) = 0, \quad j=1, \dots, n. \quad (2)$$

Ces écritures matricielles sont une façon commode de regrouper les conditions portant sur les différentes composantes; on minimise la variance de chaque composante $\varepsilon_j(t;n)$ et les lignes des matrices $b(n;k)$ sont définies indépendamment les unes des autres. Cependant, en utilisant la structure de Gram, (2) équivaut aux conditions d'"orthogonalité" $[\varepsilon(t;n), X(t-k)] = 0$, $k=1, \dots, n$, qui résultent de la minimisation de la norme $\|\varepsilon(t;n)\| = \langle \varepsilon(t;n), \varepsilon(t;n) \rangle^{1/2}$. Par inversion du temps on obtient l'innovation rétrograde partielle d'ordre n ,

$$\varepsilon^*(t;n) = X(t) - X^*(t;n) = \sum_{k=0}^n b^*(n,k)X(t+k), \quad n \in \mathbb{N}, t \in \mathbb{Z},$$

où $b^*(n,0) = I_m$ et où les matrices $b^*(n,k)$, $k=1, \dots, n$, sont une solution du système

$$\sum_{k=0}^n b^*(n;k)\Lambda(k-j) = 0, \quad j=1, \dots, n$$

Notons que les coefficients $b^*(n,k)$ ne sont pas directement liés à leurs homologues $b(n,k)$; les matrices étant réelles le symbole $*$ ne sera pas utilisé comme opérateur sur l'espace des matrices. Les matrices de covariance résiduelle sont également différentes et satisfont

$$\sigma^2(n) = E\{\varepsilon(t;n)\varepsilon(t;n)'\} = \sum_{k=0}^n b(n;k)\Lambda(-k),$$

$$\sigma^{*2}(n) = E\{\varepsilon^*(t;n)\varepsilon^*(t;n)'\} = \sum_{k=0}^n b^*(n;k)\Lambda(k).$$

L'unicité des coefficients $b(n,k)$ ou $b^*(n,k)$ n'est pas garantie et peut être obtenue par certaines conventions sur lesquelles nous reviendrons. Cependant la définition des

innovations partielles conserve un caractère intrinsèque qui permet de poser :

IV.1.1. Définition.—La fonction d'autocovariance partielle de $X(\cdot)$, notée $\delta(\cdot)$, est définie en tout n de \mathbb{Z} comme la covariance partielle entre $X(t)$ et $X(t-n)$ dans l'ensemble $\{X(t-n), \dots, X(t)\}$. De façon précise on a $\delta(0) = E\{X(t)X(t)'\}$ et pour tout n de \mathbb{N}^*

$$\delta(-n)' = \delta(n) = E\{\varepsilon(t;n-1)\varepsilon^*(t-n;n-1)'\}.$$

Pour une matrice A quelconque A^- désigne n'importe quel inverse généralisé de A , c'est-à-dire toute matrice satisfaisant $AA^-A=A$ (cf. Rao, 1965, p. 24).

IV.1.2. Proposition.—(Dégerine, 1987) Les innovations partielles satisfont les récurrences suivantes :

$$\varepsilon(t;n) = \varepsilon(t;n-1) - \delta(n)\sigma^{*2}(n-1)^{-}\varepsilon^*(t-n;n-1) \quad n \in \mathbb{N}^*, \quad t \in \mathbb{Z}. \quad (3)$$

$$\varepsilon^*(t;n) = \varepsilon^*(t;n-1) - \delta(n)'\sigma^2(n-1)^{-}\varepsilon(t+n;n-1)$$

Les matrices de covariance résiduelle s'expriment alors en fonction de $\delta(\cdot)$ par la double récurrence

$$\sigma^2(n) = \sigma^2(n-1) - \delta(n)\sigma^{*2}(n-1)^{-}\delta(n)' \quad n \in \mathbb{N}. \quad (4)$$

$$\sigma^{*2}(n) = \sigma^{*2}(n-1) - \delta(n)'\sigma^2(n-1)^{-}\delta(n)$$

Démonstration.— Supposons les propriétés vraies jusqu'à l'ordre $n-1$, $n \in \mathbb{N}^*$. Le second membre de la première égalité dans (3) est donc non corrélé avec $X(t-k)$, $k = 1, \dots, n-1$. Il sera alors caractéristique de l'innovation $\varepsilon(t;n)$, compte tenu de $\varepsilon(t;n-1) = X(t) - X(t;n-1)$, si l'on montre qu'il n'est pas corrélé avec $X(t-n)$. Cette condition équivaut à

$$[\varepsilon(t;n-1), X(t-n)] = \delta(n)\sigma^{*2}(n-1)^{-}[\varepsilon^*(t-n;n-1), X(t-n)],$$

et donc à $\delta(n) = \delta(n)A^-A$ en posant $A = \sigma^{*2}(n-1)$. Cette matrice étant symétrique, $(A^-)'$ est aussi un inverse généralisé de A et par conséquent $(A^-)'\delta(n)'$ sera une solution de l'équation $AX = \delta(n)'$ si celle-ci est cohérente. Dans ce cas on aura $A(A^-)'\delta(n)' = \delta(n)'$ d'où la condition requise par transposition. L'existence d'une

solution à l'équation $AX = \delta(n)'$ provient de la définition de $\delta(n)$. Soit $A = V\Delta^2V'$ une décomposition spectrale de $\sigma^{*2}(n-1)$ dans laquelle Δ^2 est la matrice carrée diagonale formée par les valeurs propres de A , avec leur multiplicité et rangées par ordre décroissant. On peut écrire $\varepsilon^*(t-n; n-1) = V\Delta\eta$, où Δ est la matrice diagonale, pas nécessairement carrée, formée par les racines carrées des éléments non nuls de Δ^2 . Notre équation est donc de la forme $V\Delta\Delta'V'X = V\Delta Y$, avec $Y = E\{\eta\varepsilon(t; n-1)\}$, qui admet comme solution $X = V(\Delta^-)'Y$ où Δ^- est la matrice obtenue en inversant les éléments non nuls de Δ . L'expression de $\varepsilon^*(t; n)$ s'obtient de la même manière. La non corrélation entre $\varepsilon(t; n)$ et $\varepsilon^*(t-n; n-1)$ dans (3) conduit à

$$\sigma^2(n) = \sigma^2(n-1) - \delta(n)A^{-1}A^{-1}\delta(n)',$$

qui correspond à la première égalité de (4) puisque $\delta(n)A^{-1}A^{-1} = \delta(n)$. Il en est de même pour la seconde. ♦

Lorsque les matrices de covariance résiduelle sont non singulières, les coefficients dans (3) sont uniques et c'est alors aussi le cas des coefficients d'autorégression. Ceux-ci s'expriment en fonction de $\Lambda(\cdot)$ selon l'algorithme de Levinson-Durbin généralisé par Whittle (1963) qui est une conséquence directe de la proposition précédente :

$$b(n,0) = b^*(n,0) = I, \quad b(n,n) = -\delta(n)\sigma^{*2}(n-1)^-, \quad b^*(n,n) = -\delta(n)'\sigma^2(n-1)^-,$$

pour $k=1, \dots, n-1$:

$$b(n,k) = b(n-1,k) + b(n,n)b^*(n-1,n-k),$$

$$b^*(n,k) = b^*(n-1,k) + b^*(n,n)b(n-1,n-k),$$

où

$$\delta(n) = \sum_{k=0}^{n-1} b(n-1,k)\Lambda(n-k) = \left\{ \sum_{k=0}^{n-1} b^*(n-1,k)\Lambda(k-n) \right\}',$$

les matrices de covariance résiduelle étant incrémentées selon (4) avec, comme valeur initiale, $\sigma^2(0) = \sigma^{*2}(0) = \Lambda(0)$.

L'algorithme s'étend donc aux cas singuliers dès que l'on donne un caractère intrinsèque aux inverses généralisés. On peut, par exemple, utiliser l'inverse de Moore. S'agissant ici de matrices A symétriques positives cet inverse, noté A^+ , est défini par

$A^+ = V'(\Delta^-)' \Delta^- V$ avec les mêmes notations que celles introduites dans la démonstration ci-dessus. L'inverse A^+ ainsi défini est unique (cf. Rao, 1965, p. 25).

Il ne serait pas difficile de montrer que cet algorithme établit une correspondance biunivoque entre la fonction d'autocovariance $\Lambda(\cdot)$, de type positif, et la fonction d'autocovariance partielle $\delta(\cdot)$. Cependant $\delta(\cdot)$ doit satisfaire les contraintes garantissant le caractère positif des matrices de covariance résiduelle construites selon (4). La notion de fonction d'autocorrélation partielle élimine ce type de contrainte. L'idée consiste à normaliser, en un certain sens, les innovations partielles.

§ IV.2. Fonctions d'autocorrélation partielle standards

On se fixe un procédé d'orthogonalisation. C'est-à-dire qu'à tout élément aléatoire centré ε , de matrice de covariance σ^2 , on associe, de façon unique, deux matrices σ et σ^- telles que $\eta = \sigma^- \varepsilon$ soit une orthogonalisation des composantes de ε , $E\{\eta\eta'\} = h$ où h est diagonale et idempotente, et $\varepsilon = \sigma\eta$. De plus on suppose que σ satisfait $\sigma\sigma' = \sigma^2$, ce qui équivaut à $\sigma h = \sigma$ et implique $\sigma^- \sigma = h$. Il n'est fait aucune hypothèse sur les dimensions des matrices si ce n'est qu'elles sont cohérentes avec les écritures proposées.

IV.2.1. Définition.—(Dégerine, 1987) *On appelle fonction d'autocorrélation partielle standard de $X(\cdot)$ la fonction $\beta(\cdot)$ définie par $\beta(0) = \Lambda(0)$ et*

$$\beta(-n)' = \beta(n) = E\{\eta(t;n-1)\eta^*(t-n;n-1)'\}, \quad n \in \mathbb{N}^*,$$

où $\eta(t;n-1) = \sigma(n-1)^- \varepsilon(t;n-1)$ et $\eta^*(t;n-1) = \sigma^*(n-1)^- \varepsilon^*(t;n-1)$.

IV.2.2. Proposition.—(Dégerine, 1987) *Tout procédé d'orthogonalisation établit une correspondance biunivoque entre la fonction d'autocovariance partielle $\delta(\cdot)$ et la fonction d'autocorrélation partielle standard $\beta(\cdot)$ selon la récurrence, $n \in \mathbb{N}^*$:*

$$\beta(n) = \sigma(n-1)^- \delta(n) [\sigma^*(n-1)^-]', \quad \delta(n) = \sigma(n-1) \beta(n) \sigma^*(n-1)',$$

$$\sigma^2(n) = \sigma(n-1) [I - \beta(n) \beta(n)'] \sigma(n-1)', \quad \sigma^{*2}(n) = \sigma^*(n-1) [I - \beta(n) \beta(n)] \sigma^*(n-1)',$$

avec les conditions initiales $\sigma^2(0) = \sigma^{*2}(0) = \beta(0) = \delta(0)$.

Démonstration.—Les relations entre $\beta(n)$ et $\delta(n)$ sont immédiates. A partir de (4) et de $\sigma(n-1)\sigma(n-1)' = \sigma^2(n-1)$ on obtient

$$\sigma^2(n) = \sigma(n-1)[I - \beta(n)\sigma^*(n-1)'\sigma^{*2}(n-1)^{-1}\sigma^*(n-1)\beta(n)']\sigma(n-1)'.$$

Posant $h = \sigma(n-1)^{-1}\sigma(n-1)$ on a $h\eta(t;n-1) = \eta(t;n-1)$ et par suite $h\beta(n) = \beta(n)$. De même $\beta(n)h^* = \beta(n)$ où $h^* = \sigma^*(n-1)^{-1}\sigma^*(n-1)$. En omettant d'indiquer les arguments $(n-1)$ on peut écrire les égalités suivantes :

$$\begin{aligned} \beta(n)\sigma^*\sigma^{*2}\sigma^* &= \beta(n)h^*\sigma^*\sigma^{*2}\sigma^* = \beta(n)\sigma^*\sigma^{*2}\sigma^* \\ &= \beta(n)\sigma^*\sigma^* = \beta(n)h^* = \beta(n). \end{aligned}$$

D'où l'expression de $\sigma^2(n)$; celle de $\sigma^{*2}(n)$ se vérifie de la même façon. ♦

La contrainte essentielle que doit satisfaire $\beta(\cdot)$ est que, pour tout n de \mathbb{N}^* , les valeurs singulières de $\beta(n)$, racine carrée des valeurs propres de $\beta(n)\beta(n)'$ (ou de $\beta(n)'\beta(n)$), restent inférieures ou égales à un. Cependant si $\tilde{\beta}(\cdot)$ est une suite de matrices satisfaisant cette contrainte, et dont les dimensions sont cohérentes avec le procédé retenu pour définir une fonction $\delta(\cdot)$ par la récurrence de la proposition ci-dessus, la fonction standard $\beta(\cdot)$ peut être distincte de $\tilde{\beta}(\cdot)$ et est obtenue par

$$\beta(n) = \sigma(n-1)^{-1}\sigma(n-1)\tilde{\beta}(n)[\sigma^*(n-1)^{-1}\sigma^*(n-1)].$$

Ainsi, selon le procédé d'orthogonalisation utilisé, le domaine de variation de $\beta(\cdot)$ n'est pas nécessairement descriptible de façon simple. Néanmoins $\beta(\cdot)$ est unique et permet de construire les innovations normalisées selon la récurrence issue de (3) :

$$\eta(t;n) = \sigma(n)^{-1}\sigma(n-1)\{\eta(t;n-1) - \beta(n)\eta^*(t-n;n-1)\}, \quad (5)$$

$$\eta^*(t;n) = \sigma^*(n)^{-1}\sigma^*(n-1)\{\eta^*(t;n-1) - \beta(n)'\eta(t+n;n-1)\}.$$

Il s'en suit une décomposition standard des innovations partielles $\eta(t;n)$ ou $\varepsilon(t;n)$ en fonction des éléments $X(t-k)$, $k=0, \dots, n$, du processus, le passage de l'une à l'autre résultant des relations $\varepsilon(t;n) = \sigma(n)\eta(t;n)$ et $\eta(t;n) = \sigma(n)^{-1}\varepsilon(t;n)$. On peut aussi écrire une version normalisée de l'algorithme de Levinson-Durbin.

Le domaine de variation de $\beta(\cdot)$ peut être simple lorsque le procédé d'orthogonalisation est bien choisi. C'est le cas des exemples suivants.

a) *Orthogonalisation de Gram-Schmidt*

Soit ϵ un vecteur aléatoire à valeurs dans \mathbb{R}^m , centré, de matrice de covariance σ^2 . L'orthogonalisation des composantes successives $\epsilon_j, j=1, \dots, m$, de ϵ , par le procédé de Gram-Schmidt et selon le produit scalaire $\langle \epsilon_j, \epsilon_k \rangle = E\{\epsilon_j \epsilon_k\}$, définit de nouvelles variables $\eta_j, j=1, \dots, r$, où r est le rang de σ^2 . Le vecteur aléatoire $\eta = (\eta_1, \dots, \eta_r)'$ est centré avec $E\{\eta \eta'\} = I_r$. On a alors $\epsilon = \sigma \eta$ où σ est une matrice $m \times r$, unique, vérifiant $\sigma \sigma' = \sigma^2$. De même η s'exprime en fonction de ϵ par $\eta = \sigma^- \epsilon$ où σ^- est une matrice $r \times m$ rendue unique en exprimant η en fonction des r composantes de ϵ qui, au cours du procédé d'orthogonalisation, ont conduit aux variables $\eta_j, j=1, \dots, r$.

b) *Orthogonalisation selon les composantes principales*

Soit $\sigma^2 = V \Delta^2 V'$ une décomposition spectrale de la matrice de covariance σ^2 de ϵ où Δ^2 , et par suite Δ et Δ^- , sont définis de manière unique comme dans la démonstration de la Proposition IV.1.2. Les vecteurs propres associés, colonnes de V , sont rendus uniques en imposant que la première composante non nulle soit positive. Lorsqu'une valeur propre est multiple, le premier vecteur propre associé est celui pour lequel le nombre de composantes nulles successives, à partir de la dernière, est maximum; les suivants sont construits selon le même principe tout en conservant les conditions d'orthogonalité. L'orthogonalisation est réalisée en posant $\sigma = V \Delta$ et $\sigma^- = \Delta^- V'$.

Dans ces deux situations, la récurrence (5) entre les innovations partielles, montre que le domaine de variation de $\beta(\cdot)$, noté par la suite $\mathfrak{D}_\beta(m)$, se présente ainsi : la matrice $\beta(0)$ est symétrique positive d'ordre m ; la suite $\{\beta(n), n=1, 2, \dots\}$ est constituée de matrices carrées dont les valeurs singulières sont inférieures ou égales à un et telle que l'ordre de $\beta(n+1)$ soit égal au rang de $I - \beta(n)\beta(n)'$, celui de $\beta(1)$ étant égal au rang de la matrice $\beta(0)$. Ainsi soit la suite est infinie et l'ordre de $\beta(n)$ est constant à partir d'un certain rang, soit $X(\cdot)$ est linéairement singulier d'ordre d et $\beta(\cdot)$ se termine par une matrice $\beta(d)$ dont toutes les valeurs singulières sont égales à un.

La situation a) correspond à l'extension de la fonction d'autocorrélation partielle au cas vectoriel proposée par Morf et coll. (1978). Cependant les auteurs considèrent uniquement le cas où les matrices de covariance résiduelle sont non singulières à l'exception éventuelle de la dernière. Il est clair que de nombreuses situations, y compris celles de processus non singuliers, n'entrent pas dans ce cadre là.

Tout procédé d'orthogonalisation équivaut à choisir les variables d'innovation que constituent les composantes de $\eta(t;n)$. On peut aussi envisager un procédé différent selon que l'on normalise l'innovation progressive ou rétrograde. L'orthogonalisation selon Gram-Schmidt avec σ triangulaire supérieure et σ^* triangulaire inférieure conduit aux mêmes variables d'innovation que celles introduites par Sakai (1983). Dans cet article l'auteur définit la fonction d'autocorrélation partielle sous forme de m suites à valeurs dans $]-1,+1[$. Là encore les matrices de covariance résiduelle sont supposées non singulières.

§ IV.3. Fonction d'autocorrélation partielle canonique

Nous introduisons, ci-après, la notion de fonction d'autocorrélation partielle canonique. Pour cela il est nécessaire de disposer, pour toute matrice carrée β dont les valeurs singulières sont inférieures ou égales à un, d'une décomposition en valeurs singulières qui soit unique. Nous écrivons donc β sous la forme

$$\beta = L'\Delta\Theta M = L'\Theta\Delta M, \quad (6)$$

où L et M sont les matrices orthogonales des décompositions spectrales

$$I - \beta\beta' = L'(I - \Delta^2)L, \quad I - \beta'\beta = M'(I - \Delta^2)M,$$

dont l'unicité est obtenue comme dans le procédé d'orthogonalisation b) plus haut. Alors la matrice diagonale Δ est carrée et ses éléments diagonaux sont les valeurs singulières de β , avec leur multiplicité; ils sont donc rangés par ordre croissant. La matrice Θ est définie par $L\beta M' = \Delta\Theta$ en convenant de poser égaux à zéro les éléments des lignes et colonnes correspondant aux valeurs nulles de Δ . C'est une matrice diagonale par bloc,

chaque bloc est associé à chacune des valeurs distinctes de Δ et constitue, dans le cas non nul, une transformation orthogonale dans les sous-espaces propres correspondants. Lorsque les valeurs singulières, éléments de Δ , sont distinctes, Θ est diagonale et indique le signe de $\Delta\Theta$. Notons que les lignes de L , ou de M , associées aux valeurs nulles de Δ n'entrent pas dans la reconstitution de β ; elles sont d'ailleurs caractérisées par les autres lignes de la matrice; elles sont par contre nécessaires si l'on veut préserver la relation $I-\beta\beta' = L'(I-\Delta^2)L$ dans laquelle ce sont les lignes associées aux valeurs singulières égales à un qui deviennent inutiles.

Nous utiliserons également les notations suivantes : $(I-\Delta^2)^{1/2}$ et $(I-\Delta^2)^{1/2-}$ sont les matrices diagonales, éventuellement rectangulaires, définies, à partir de $I-\Delta^2$, comme dans la démonstration de la Proposition IV.1.2. Dans le cas particulier de la décomposition spectrale $\sigma^2(0)=L_0'\Delta^2(0)L_0$, les matrices rectangulaires Δ_0 et Δ_0^- sont construites, en fonction de $\Delta^2(0)$, comme ci-dessus. Notons que l'unicité des décompositions spectrales des matrices symétriques positives est toujours obtenue selon le procédé b).

IV.3.1. Proposition.—(Dégerine, 1987) Soit $\sigma^2(n)$, $n \in \mathbb{N}$, la suite des matrices de covariance résiduelle progressives d'un processus $X(\cdot)$. A partir de la décomposition spectrale $\sigma^2(0) = L_0'\Delta^2(0)L_0$ on définit les conditions initiales $\sigma(0) = L_0'\Delta_0$ et $\sigma(0)^- = \Delta_0^-L_0$ de la double récurrence

$$\begin{aligned}\sigma(n) &= \sigma(n-1)L_n'[I - \Delta^2(n)]^{1/2}, \\ \sigma(n)^- &= [I - \Delta^2(n)]^{1/2-}L_n\sigma(n-1)^-, \end{aligned}$$

où L_n et $\Delta^2(n)$ sont les éléments de la décomposition spectrale

$$\sigma(n-1)^-\sigma^2(n)[\sigma(n-1)^-]' = L_n'[I - \Delta^2(n)]L_n.$$

Alors les matrices $\sigma(n)$ et $\sigma(n)^-$, $n \in \mathbb{N}$, ainsi construites sont rectangulaires d'ordres respectifs $m \times r(n)$ et $r(n) \times m$ où $r(n)$ est le rang de $\sigma^2(n)$ et satisfont

- (i) $\sigma(n)\sigma(n)' = \sigma^2(n)$,
- (ii) $\sigma(n)^-\sigma(n) = I_{r(n)}$.

Démonstration.— Notons, tout d'abord, les conséquences des propriétés (i) et (ii). La

seconde implique que $\sigma(n)^-$ et $\sigma(n)$ sont des inverses généralisés de $\sigma(n)$ et de $\sigma(n)^-$. Jointe à la première, il s'en suit que $[\sigma(n)^-]'\sigma(n)^-$ est un inverse généralisé de $\sigma^2(n)$. D'autre part cette proposition a son homologue pour la suite des matrices de covariance résiduelle rétrogrades $\sigma^{*2}(n)$, $n \in \mathbb{N}$, et sa démonstration se conduit par récurrence sur l'ensemble des deux suites. Il est clair que $\sigma(0) = L_0'\Delta_0$ et $\sigma(0)^- = \Delta_0^-L_0$ satisfont (i) et (ii). Puisque $\sigma^{*2}(0) = \sigma^2(0)$, on dispose d'un inverse généralisé $\sigma^{*2}(0)^-$. Supposons donc que $\sigma(n-1)$ et $\sigma(n-1)^-$ satisfont (i) et (ii) et que $\sigma^{*2}(n-1)^-$ soit un inverse généralisé de $\sigma^{*2}(n-1)$. D'après l'expression (4) de $\sigma^2(n)$ donnée dans la Proposition IV.1.2, on peut écrire

$$\sigma(n-1)^-\sigma^2(n)[\sigma(n-1)^-]' = I_{\alpha(n-1)} - \sigma(n-1)^-\delta(n)\sigma^{*2}(n-1)^-\delta(n)[\sigma(n-1)^-]'$$

Le premier membre est une matrice symétrique positive et le second montre que sa décomposition sous la forme $L_n'[I - \Delta^2(n)]L_n$ est légitime. La récurrence permet de construire $\sigma(n)$ et $\sigma(n)^-$ et la vérification de (ii) est immédiate. Pour (i) on obtient

$$\sigma(n)\sigma(n)' = \sigma(n-1)\sigma(n-1)^-\sigma^2(n)[\sigma(n-1)^-]'\sigma(n-1)'$$

$\sigma(n-1)\sigma(n-1)^-$ est le projecteur, dans \mathbb{R}^m , sur l'espace engendré par les colonnes (ou lignes) de $\sigma^2(n-1)$. Celui-ci contient l'espace engendré par les colonnes de $\sigma^2(n)$ puisque le noyau de la forme quadratique associée à $\sigma^2(n-1)$ est inclus dans celui associé à $\sigma^2(n)$ comme le montre l'égalité (4) sous la forme

$$\sigma^2(n-1) = \sigma^2(n) + \delta(n)\sigma^{*2}(n-1)^-\delta(n)'$$

On peut donc éliminer $\sigma(n-1)\sigma(n-1)^-$ et sa transposée dans l'expression ci-dessus de $\sigma(n)\sigma(n)'$. La présence, dans l'ordre des matrices, du rang $r(n)$ de $\sigma^2(n)$ est une conséquence de l'inclusion des noyaux évoquée précédemment et des conventions adoptées plus haut. On dispose également d'un inverse généralisé $\sigma^2(n)^-$, ce qui permet de poursuivre la récurrence menée parallèlement sur le cas rétrograde. ♦

Cette proposition conduit à un procédé d'orthogonalisation différent des précédents par son caractère récurrent : $\sigma(n)$ et $\sigma(n)^-$ sont construits à partir de $\sigma(n-1)$ et de $\sigma(n-1)^-$. Dans le cas rétrograde on utilise les notations suivantes :

$$\sigma^{*2}(0) = M_0' \Delta^2(0) M_0, M_0 = L_0,$$

$$\sigma^{*(n-1)-} \sigma^{*2}(n) [\sigma^{*(n-1)-}]' = M_n' [I - \Delta^2(n)] M_n, n \in \mathbb{N}^*.$$

Le fait d'obtenir les mêmes matrices $\Delta^2(n)$ dans le sens progressif ou rétrograde est une conséquence de la Proposition IV.2.2 appliquée à ce procédé. Par la suite, sauf expression du contraire, il ne sera question que de ce procédé que l'on dira canonique pour des raisons qui apparaîtront plus loin. La définition de $\beta(\cdot)$ selon IV.2.1, qui résulte de ce procédé, et sa paramétrisation par (6) donnent une forme particulière à la récurrence (5) entre les innovations normalisées.

IV.3.2. Définition.—(Dégerine, 1987) *On appelle fonction d'autocorrélation partielle canonique de $X(\cdot)$ la fonction $\beta(\cdot)$ définie par*

$$\beta(0) = E\{X(t)X(t)'\} = L_0' \Delta^2(0) L_0,$$

$$\beta(-n)' = \beta(n) = E\{\eta(t;n-1)\eta^{*(t-n;n-1)'}\} = L_n' \Delta_n \Theta_n M_n, n \in \mathbb{N}^*,$$

où les innovations sont construites selon la récurrence

$$\eta(t;n) = [I - \Delta^2(n)]^{1/2-} \{L_n \eta(t;n-1) - \Delta_n \Theta_n M_n \eta^{*(t-n;n-1)}\},$$

$$\eta^{*(t;n)} = [I - \Delta^2(n)]^{1/2-} \{M_n \eta^{*(t;n-1)} - \Delta_n \Theta_n L_n \eta(t+n;n-1)\},$$

avec, pour conditions initiales, $\eta(t;0) = \eta^{*(t;0)} = \Delta_0^- L_0 X(t)$.

Les innovations normalisées ainsi définies s'expriment en fonction des valeurs de $X(\cdot)$ sous la forme

$$\eta(t;n) = \sum_{k=0}^n B(n;k) X(t-k), \eta^{*(t;n)} = \sum_{k=0}^n B^*(n;k) X(t+k).$$

Les coefficients $B(n,k)$ et $B^*(n,k)$ sont ceux définis, en fonction de $\beta(\cdot)$, par la récurrence ci-dessous.

$$B(n,0) = [I - \Delta^2(n)]^{1/2-} L_n B(n-1,0), B^*(n,0) = [I - \Delta^2(n)]^{1/2-} M_n B^*(n-1,0),$$

$$B(n,n) = -[I - \Delta^2(n)]^{1/2-} \Delta_n \Theta_n M_n B^*(n-1,0),$$

$$B^*(n,n) = -[I - \Delta^2(n)]^{1/2-} \Delta_n \Theta_n L_n B(n-1,0),$$

pour $k=1, \dots, n-1$:

$$B(n,k) = [I - \Delta^2(n)]^{1/2-} \{L_n B(n-1,k) - \Delta_n \Theta_n M_n B^*(n-1,n-k)\},$$

$$B^*(n,k) = [I - \Delta^2(n)]^{1/2-} \{M_n B^*(n-1,k) - \Delta_n \Theta_n L_n B(n-1,n-k)\},$$

et les conditions initiales $B(0,0) = \Delta_0^- L_0$, $B^*(0,0) = \Delta_0^- M_0$.

La version normalisée de l'algorithme de Levinson-Durbin est obtenue en ajoutant à ces relations l'une des deux suivantes :

$$\beta(n) = \sum_0^{n-1} B(n-1,k)\Lambda(n-k)B^*(n-1,0)' = \left\{ \sum_0^{n-1} B^*(n-1,k)\Lambda(k-n)B(n-1,0)' \right\}'.$$

Cet algorithme donne la correspondance biunivoque entre $\Lambda(\cdot)$ et $\beta(\cdot)$ car les coefficients $B(n,0)$ et $B^*(n,0)$ ne sont autre que $\sigma(n)^-$ et $\sigma^*(n)^-$. On a ainsi

$$\Lambda(n) = \sigma(n-1)\beta(n)[\sigma^*(n-1)]' - \sum_1^{n-1} \sigma(n-1)B(n-1,k)\Lambda(n-k),$$

les valeurs successives de $\sigma(n)$ et $\sigma^*(n)$ étant obtenues selon la récurrence de la Proposition IV.3.1.

Le domaine de variation de $\beta(\cdot)$ est le même que celui obtenu dans les procédés a) et b) décrits plus haut et que nous avons noté $\mathfrak{D}_\beta(m)$. Notons, dans la décomposition $\beta(n) = L_n' \Delta_n \Theta_n M_n$, que la fonction $\beta(\cdot)$, à l'exception du "signe" Θ_n , est caractérisée par les deux suites de matrices de covariance résiduelle $\sigma^2(n)$ et $\sigma^{*2}(n)$, $n \in \mathbb{N}$, la correspondance étant assurée par les relations de la Proposition IV.3.1. La surjectivité de l'application $\Lambda(\cdot) \rightarrow \beta(\cdot)$ s'établit de façon directe par le procédé constructif analogue à celui obtenu dans le cas scalaire (Proposition I.2.1).

IV.3.3. Proposition.— Soit $\{Y(n), n \in \mathbb{N}\}$ une suite de vecteurs aléatoires centrés, non corrélés, et dont les matrices de covariance, $E\{Y(n)Y(n)'\} = \sigma^2(n)$, sont celles associées à une fonction $\beta(\cdot)$ de $\mathfrak{D}_\beta(m)$. Alors la suite $\{X(n), n \in \mathbb{N}\}$ construite par $X(0) = Y(0)$ et la récurrence

$$X(n) = X(n;n-1) + \sigma(n-1)\beta(n)\sigma^*(n-1)^- \varepsilon^*(0;n-1) + Y(n), \quad n \in \mathbb{N}^*,$$

est stationnaire et sa fonction d'autocorrélation partielle canonique est égale à $\beta(\cdot)$.

Démonstration.— Vérifions l'égalité des matrices de covariance de $X(0)$ et de $X(1) = \sigma(0)\beta(1)\sigma^*(0)^- X(0) + Y(1)$:

$$E\{X(1)X(1)'\} = \sigma(0)\beta(1)\beta(1)'\sigma(0)' + \sigma(0)[I - \beta(1)\beta(1)']\sigma(0)' = \sigma^2(0).$$

L'autocorrélation partielle canonique entre $X(1)$ et $X(0)$ est égale à $\beta(1)$ puisque l'on a

$$E\{X(1)X(0)'\} = \sigma(0)\beta(1)\sigma^*(0)'$$

Supposons que $\{X(0), \dots, X(n-1)\}$ soit stationnaire et admette $\{\beta(0), \dots, \beta(n-1)\}$ comme autocorrélations partielles canoniques. Les éléments $X(n;n-1)$ et $\varepsilon^*(0;n-1)$ sont alors parfaitement définis et on a, pour $1 \leq k \leq n-1$:

$$\begin{aligned} E\{X(n)X(n-k)'\} &= E\{X(n;n-1)X(n-k)'\} \\ &= E\{X(n-1;n-1)X(n-1-k)'\} = E\{X(n-1)X(n-1-k)'\}. \end{aligned}$$

La matrice de covariance de $X(n)$ est égale à $\sigma^2(0)$ car

$$\begin{aligned} E\{X(n)X(n)'\} &= E\{X(n;n-1)X(n;n-1)'\} + \sigma(n-1)\beta(n)\beta(n)'\sigma(n-1)' + \sigma^2(n) \\ &= \sigma^2(0) - \sigma^2(n-1) + \sigma(n-1)\beta(n)\beta(n)'\sigma(n-1)' + \sigma(n-1)[I - \beta(n)\beta(n)']\sigma(n-1)'. \end{aligned}$$

La séquence $\{X(0), \dots, X(n)\}$ est donc stationnaire. L'expression de $X(n)$ donne

$$\varepsilon(n;n-1) - \sigma(n-1)\beta(n)\sigma^*(n)^{-1}\varepsilon^*(0;n-1) = Y(n).$$

Le vecteur $Y(n)$ est non corrélé avec $\varepsilon^*(0;n-1)$ d'où

$$E\{\varepsilon(n;n-1)\varepsilon^*(0;n-1)'\} = \sigma(n-1)\beta(n)\sigma^*(n-1)',$$

qui exprime que $\beta(n)$ est l'autocorrélation partielle canonique d'ordre n . ♦

Considérons les variables :

$$L_n \eta(t;n-1) = L_n \sigma(n-1)^{-1} \varepsilon(t;n-1),$$

$$M_n \eta^*(t-n;n-1) = M_n \sigma^*(n-1)^{-1} \varepsilon^*(t-n;n-1).$$

Leurs matrices de covariance sont égales à l'identité et leur corrélation est donnée par

$$E\{L_n \eta(t;n-1) \eta^*(t-n;n-1)' M_n'\} = L_n \beta(n) M_n' = \Delta_n \Theta_n.$$

Ainsi Δ_n représente les corrélations canoniques entre $\varepsilon(t;n-1)$ et $\varepsilon^*(t-n;n-1)$.

$L_n \eta(t;n-1)$ et $M_n \eta^*(t-n;n-1)$ sont les variables canoniques correspondantes rendues uniques par les conventions adoptées, identiques dans le sens progressif ou rétrograde, d'où la présence supplémentaire du "signe" Θ_n . Cependant le rôle de Θ_n dans $\beta(n)$, et donc dans la structure de $X(\cdot)$, est essentiel. On a déjà remarqué que $\{(L_n, \Delta_n), n \in \mathbb{N}\}$ caractérise $\{\sigma^2(n), n \in \mathbb{N}\}$; de même $\{(M_n, \Delta_n), n \in \mathbb{N}\}$ caractérise $\{\sigma^{*2}(n), n \in \mathbb{N}\}$, ce qui rend plus transparente la liaison entre les deux suites.

Cette présentation montre aussi les points suivants :

(i) $\sigma^2(n)$ et $\sigma^{*2}(n)$ ont même rang ou encore les dimensions des supports de $\varepsilon(t;n)$ et de $\varepsilon^*(t;n)$ sont égales,

(ii) le support de $\varepsilon(t;n)$ est inclus dans celui de $\varepsilon(t;n-1)$,

(iii) Les colonnes de $\sigma(n-1)L_n'$ sont les vecteurs propres normalisés associés aux valeurs propres données par la diagonale de $[I-\Delta^2(n)]$ pour la matrice $\sigma^2(n)$ et relativement au produit scalaire défini par $\sigma^2(n-1)^{-}$. C'est-à-dire que l'on a :

$$\sigma^2(n)\sigma^2(n-1)^{-}\sigma(n-1)L_n' = \sigma(n-1)L_n'[I - \Delta^2(n)],$$

$$[\sigma(n-1)L_n']'\sigma^2(n-1)^{-}[\sigma(n-1)L_n'] = I.$$

Tout autre choix de normalisation conduit à une suite de matrices $\tilde{\beta}(n)$, $n \geq 1$, qui ont les mêmes valeurs singulières Δ_n , $n \geq 1$. Cependant tout choix se traduit par une innovation normalisée $\eta(t;n)$ qui a sa propre interprétation. Dans celui que nous préconisons, les composantes de $\eta(t;n)$ et celles de $\eta^*(t-n;n)$ sont les variables canoniques associées au couple $\{\varepsilon(t;n), \varepsilon^*(t-n;n)\}$.

IV.3.4. Proposition.—(Dégerine, 1987) *La matrice de corrélation entre les innovations normalisées $\eta(t;n) = \sigma(n)^{-}\varepsilon(t;n)$ et $\eta^*(t-n;n) = \sigma^*(n)^{-}\varepsilon^*(t-n;n)$ est donnée par :*

$$E\{\eta(t;n)\eta^*(t-n;n)\} = -[I - \Delta^2(n)]^{1/2} \Delta_n \Theta_n [I - \Delta^2(n)]^{1/2}.$$

Démonstration.— Une suite d'égalités, dans lesquelles on vérifie que les produits matriciels effectués sont autorisés, conduit au résultat.

$$\begin{aligned} E\{\eta(t;n)\eta^*(t-n;n)'\} &= \sum_{k=0}^n E\{B(n,k)X(t-k)\eta^*(t-n;n)'\} = E\{B(n,n)X(t-n)\eta^*(t-n;n)'\} \\ &= E\{B(n;n)\varepsilon^*(t-n;n)\eta^*(t-n;n)'\} = E\{B(n;n)\varepsilon^*(t-n;n)\varepsilon^*(t-n;n)[\sigma^*(n)^{-}]\} \\ &= -[I - \Delta^2(n)]^{1/2} \Delta_n \Theta_n M_n B^*(n-1;0)\sigma^*(n)\sigma^*(n)'[\sigma^*(n)^{-}]' \\ &= -[I - \Delta^2(n)]^{1/2} \Delta_n \Theta_n M_n \sigma^*(n-1)^{-}\sigma^*(n-1)'M_n'[I - \Delta^2(n)]^{1/2} \\ &= -[I - \Delta^2(n)]^{1/2} \Delta_n \Theta_n [I - \Delta^2(n)]^{1/2}. \spadesuit \end{aligned}$$

La matrice de corrélation ci-dessus est égale à la matrice obtenue en supprimant dans $-\Delta_n \Theta_n$ les lignes et colonnes correspondant aux éléments de Δ_n égaux à un.

Il est clair que $\{\Lambda(0), \dots, \Lambda(n)\}$ est paramétré par $\{\sigma^2(0); \beta(1), \dots, \beta(n)\}$ mais, contrairement au cas scalaire, il n'est pas possible, en général, de remplacer $\sigma^2(0)$ par $\sigma^2(n)$. La matrice de covariance résiduelle $\sigma^2(n)$ est non singulière si et seulement si les corrélations canoniques définissant $\Delta_k, k=1, \dots, n$, sont strictement inférieures à un, $\sigma^2(0)$ étant supposée non singulière. Dans ce cas la correspondance $\sigma(0) \leftrightarrow \sigma(n)$, pour $\beta(\cdot)$ fixé, est biunivoque. Par contre celle qui associe $\sigma^2(0)$ et $\sigma^2(n)$, dans l'ensemble des matrices non singulières, n'est pas nécessairement surjective ni même injective comme le montre l'exemple suivant. Pour obtenir $\sigma^2(1) = I$ on doit avoir $L_1 = J$ (J est la matrice unitaire dont les seuls éléments non nuls, alors égaux à un, sont sur l'antidiagonale), ce qui prouve la non surjectivité, et alors $\Delta_0 = J'[I - \Delta^2(1)]^{1/2} J$, mais L_0 reste unitaire quelconque, d'où la non injectivité, et $\sigma(1) = L_0' J'$.

PARTIE B**APPORT DES AUTOCORRELATIONS PARTIELLES****A L'ESTIMATION DANS LE DOMAINE TEMPOREL****INTRODUCTION**

- Chapitre I ESTIMATION DE MAXIMUM DE VRAISEMBLANCE
D'UNE MATRICE DE COVARIANCE**
- Chapitre II ESTIMATION DE MAXIMUM DE VRAISEMBLANCE
D'UNE MATRICE DE TOEPLITZ**
- Chapitre III ESTIMATION DE LA STRUCTURE DE COVARIANCE
D'UNE SEQUENCE STATIONNAIRE**
- Chapitre IV ESTIMATION AUTOREGRESSIVE DANS LE DOMAINE
TEMPOREL**

INTRODUCTION

Cette deuxième partie est consacrée à l'estimation de la structure temporelle d'un processus scalaire réel $X(\cdot)$, à temps discret, stationnaire au second ordre et centré. La nécessité de recourir à d'autres méthodes que l'autocovariance empirique usuelle $\Lambda_e(\cdot)$ apparaît lorsque la séquence observée est courte ou lorsque le modèle est "singulier" au sens où sa matrice de covariance est proche de la singularité. La méthode du maximum de vraisemblance est la solution privilégiée pour un échantillon de séries courtes et ce d'autant plus lorsque se rajoute le facteur "singulier". Elle l'est aussi pour l'estimation autorégressive dans le domaine temporel lorsque se conjuguent les deux facteurs précédents. Cette situation se rencontre principalement en traitement du signal, d'où l'origine de la plupart des méthodes.

Notre ambition ici est évidemment de démontrer que la fonction d'autocorrélation partielle $\beta(\cdot)$ constitue à bien des égards un atout important dans les problèmes d'estimation dans le domaine temporel. Elle s'impose dans la méthode du maximum de vraisemblance, tant en ce qui concerne le souci d'existence et d'unicité de la solution, que la nécessité de sa mise en œuvre. Pour ce dernier point, elle facilite en effet l'utilisation des méthodes de gradient, mais elle conduit surtout à des techniques de relaxation qui s'avèrent plus efficaces. D'autre part l'estimation empirique naturelle de $\beta(\cdot)$ n'est pas celle qui résulte des relations de Yule-Walker appliquées à $\Lambda_e(\cdot)$. En particulier elle ne nécessite aucun artifice pour assurer la positivité de la covariance correspondante, ce qui explique en partie sa supériorité là où le biais de $\Lambda_e(\cdot)$ est important (série courte). De plus elle est étroitement liée à un critère de moindres carrés analogue à celui de la méthode de Burg avec laquelle elle reste très comparable sur le plan de la qualité des estimateurs. Les méthodes d'estimation autorégressive sont d'ailleurs très souvent construites sur une estimation directe de $\beta(\cdot)$. Leur performance par rapport à $\Lambda_e(\cdot)$,

y compris pour une série longue, dans le cas "singulier" montre l'avantage de cette paramétrisation dans l'investigation de la structure du processus.

Soit x la matrice $m \times n$ constituée de l'observation d'un échantillon de taille n d'un vecteur gaussien centré. Maximiser la vraisemblance, par rapport à la matrice de covariance Λ parcourant une certaine hypothèse \mathcal{H} , équivaut à minimiser la fonction

$$L(\Lambda; X) = \ln |\Lambda| + \text{tr} (\Lambda^{-1}X),$$

où $X = xx'/n$ est la matrice de covariance empirique. Il n'existe donc pas de différence fondamentale, pour le calcul d'une solution, entre le cas d'une seule séquence ($n=1$) et celui d'un échantillon. Par contre l'hypothèse \mathcal{H} est primordiale. Dans le cas des séries chronologiques, elle est une partie du cône positif $\mathcal{C}^+(m)$ constitué des matrices de Toeplitz symétriques définies positives d'ordre m . La complexité de \mathcal{H} conduit le plus souvent à utiliser une approximation de $L(\Lambda; X)$. Hasza (1980) donne une solution analytique pour une séquence du modèle AR(1). Le Cadre et Lopez (1984) présentent un algorithme de gradient à pas optimal pour l'estimation d'une matrice interspectrale ($\mathcal{H} = \mathcal{C}^+(m)$, n grand devant m). La méthode de gradient est issue d'une situation plus générale étudiée par Burg et coll. (1982) et repose sur une paramétrisation de $\mathcal{C}^+(m)$ par une base de matrices. Pham (1987) résoud le cas d'une séquence ARMA par la méthode de Newton-Rhaphson appliquée aux paramètres du modèle en utilisant une approximation de la matrice des dérivées secondes. Dans le cas AR l'utilisation de $\beta(\cdot)$ permet un calcul exact de la dérivée seconde et un meilleur contrôle de la stabilité du filtre autorégressif nécessaire en situation "singulière" (Dégerine et Pham, 1987). Cependant les méthodes de gradient se heurtent à la non convexité de la fonction $L(\Lambda; X)$. Nous proposons (Dégerine, 1987) une méthode de relaxation, pour un échantillon de séries courtes, $\mathcal{H} = \mathcal{C}^+(m)$, basée sur une décomposition spectrale spécifique aux matrices de Toeplitz. Pham (1988), en itérant la méthode du maximum de vraisemblance récursif de Kay (1983), aboutit à une nouvelle méthode dans laquelle la relaxation porte directement sur les composantes de $\beta(\cdot)$. Elle s'applique à l'ensemble des situations autorégressives : modèle AR(p), avec $p < m$, pour une seule séquence ou pour un échantillon. Les

problèmes d'existence et d'unicité de la solution sont rarement abordés, citons Burg et coll. (1982), A (1984), Miller et Snyder (1987).

L'estimation autorégressive est proposée par Burg (1967) et souvent référencée comme méthode d'entropie maximale. De plus Burg (1968) introduit une nouvelle méthode d'estimation des paramètres du modèle autorégressif. Elle consiste à minimiser, de façon récursive par rapport à l'ordre du modèle, les erreurs de prédiction progressive et rétrograde. La version "symétrique" de la méthode des moindres carrés (Nutall ; Ulrych et Clayton, 1976) et celle des "rapports d'énergies résiduelles" (Dickinson, 1978) sont basées sur la minimisation de ces mêmes erreurs.

Nous présentons, dans le premier chapitre, quelques propriétés générales de la méthode du maximum de vraisemblance. La difficulté tient à ce que $L(\Lambda; X)$ se présente comme la différence de deux fonctions convexes. La condition $\ker(\Lambda) \cap \ker(X) = \{0\}$, pour toute matrice Λ de la frontière de \mathfrak{H} , assure l'existence d'une solution. Nous caractérisons les solutions de l'équation de vraisemblance, et celle réalisant le maximum, de façon géométrique. Pour faire abstraction de l'hypothèse gaussienne, on peut remarquer que la minimisation de $L(\Lambda; X)$ équivaut à celle de $|\Lambda|$ (resp. $\text{tr}(\Lambda^{-1}X)/m$) sous la contrainte $\text{tr}(\Lambda^{-1}X)/m \leq 1$ (resp $|\Lambda| \leq 1$). Nous montrons que ces problèmes conduisent à interpréter la covariance empirique X comme une solution de moindres carrés dans un modèle à erreurs multiplicatives.

Nous résolvons de façon détaillée le cas d'un échantillon de séries courtes, $\mathfrak{H} = \mathcal{C}^+(m)$, dans le deuxième chapitre. Nous reconsidérons de façon spécifique le problème d'existence : elle est assurée presque sûrement pour $n \geq m/2$ et, dans le cas d'une seule séquence d'un modèle $AR(p)$, pour $p < 2m/3$. Nous introduisons la notion de décomposition spectrale d'une matrice de Toeplitz sous la forme $\Lambda = MDM^*$ où M est une matrice liée à un ensemble de fréquences dont les énergies sont précisées par la matrice diagonale D . Pour Λ dans $\mathcal{C}^+(m)$ il n'existe que deux décompositions, notées ${}_+M {}_+D {}_+M^*$ et ${}_-_M {}_-D {}_-M^*$. Le couple $\{ {}_+M, {}_-M \}$ caractérise la matrice

d'autocorrélation correspondante. La minimisation de $L(MDM^*;X)$ par rapport à D est explicite et conduit à une méthode de relaxation portant sur les deux paramètres $+M$ et $-M$. On montre que la méthode converge vers une solution de l'équation de vraisemblance. Son avantage sur l'autocovariance empirique est illustrée par simulation. Nous prouvons aussi l'existence de minima locaux pour $L(\Lambda;X)$.

Dans le cas d'une seule séquence, la vraisemblance, sans contrainte supplémentaire, est non bornée avec probabilité positive. Nous comparons cependant, dans le troisième chapitre, l'autocovariance empirique avec la solution du maximum de vraisemblance, lorsqu'elle existe, pour mieux illustrer la différence de principe entre les deux méthodes. En effet la corrélation obtenue par les deux solutions peut s'interpréter comme celle d'un prolongement presque-périodique de l'observation. Dans le cas empirique la séquence est prolongée par des zéros alors que, plus naturellement, dans le cas du maximum de vraisemblance ce sont les erreurs de prédiction qui sont prises égales à zéro. Une alternative à l'autocovariance empirique $\Lambda_e(\cdot)$ est l'autocorrélation partielle empirique notée $\beta_e(\cdot)$. Sa définition est naturelle et invariante par l'inversion du temps mais $\beta_e(k)$, défini pour $k < m/2$, ne s'interprète pas comme un coefficient de régression. La version symétrique $\beta_{es}(\cdot)$ a cette propriété. Elle est définie pour $k < 2m/3$, comme celle du maximum de vraisemblance d'un modèle $AR(k)$. Elle est aussi liée aux minimisations successives des erreurs de prédiction progressive et rétrograde, comme celle de Burg. Pour une séquence de longueur 1000 d'un modèle $AR(4)$ "singulier", elle est équivalente à l'autocovariance empirique dans l'estimation des autocorrélations $\rho(\cdot)$, mais la différence est spectaculaire dans l'estimation des autocorrélations partielles $\beta(\cdot)$ ou celle des coefficients $b(\cdot)$ du modèle.

Les méthodes d'estimation autorégressive sont présentées et comparées par simulation dans le dernier chapitre. Elles sont habituellement proposées dans le cas d'une seule séquence, mais leur extension au cas d'un échantillon est immédiate. Nous donnons un nouvel algorithme de calcul pour la méthode de Burg. La méthode du

maximum de vraisemblance permet de mieux juger la qualité des autres méthodes. Dans l'ensemble elle leur est préférable et plus particulièrement dans l'estimation de $\rho(\cdot)$. Cependant le maximum de vraisemblance récursif (première itération dans le calcul du maximum de vraisemblance par relaxation) réalise de meilleurs performances pour l'estimation de $\beta(\cdot)$, et surtout pour celle de $b(\cdot)$, dans le cas d'une séquence du modèle AR(4) "singulier". Ceci n'est plus vrai pour une séquence d'un modèle AR(8) "singulier" ni pour un échantillon du modèle AR(4)!

CHAPITRE I

ESTIMATION DE MAXIMUM DE VRAISEMBLANCE

D'UNE MATRICE DE COVARIANCE

Soit $X=(X_{kj}; k=1, \dots, m; j=1, \dots, n)$, la matrice aléatoire constituée par un échantillon de taille n d'un vecteur aléatoire réel gaussien, centré, de matrice de covariance Λ . Lorsque Λ est non singulière ($\in \mathcal{S}^+(m)$), la loi de X est définie par la densité

$$f(x) = (2\pi)^{-nm/2} |\Lambda|^{-n/2} e^{-n\text{tr}(\Lambda^{-1}X)},$$

où $X = (xx')/n$ désigne la matrice de covariance empirique. Notons que la statistique $(XX')/n$ est exhaustive et constitue un estimateur sans biais de Λ . Maximiser la vraisemblance, par rapport à Λ satisfaisant une hypothèse \mathfrak{H} , équivaut à minimiser la fonction

$$L(\Lambda; X) = \ln |\Lambda| + \text{tr}(\Lambda^{-1}X).$$

Ainsi $\Lambda \longrightarrow L(\Lambda; X)$ varie comme l'opposé de la vraisemblance; la convergence de la vraisemblance vers zéro au bord de son domaine équivaut donc à une convergence vers l'infini de la fonction $L(\cdot; X)$.

Nous présentons, dans le premier paragraphe, quelques résultats généraux sur la fonction $L(\cdot; X)$. L'étude de son comportement au bord de son domaine de définition, au paragraphe suivant, permet de donner des conditions suffisantes pour l'existence d'une solution au problème du maximum de vraisemblance. Le troisième paragraphe est consacré à une caractérisation de cette solution et à sa comparaison avec la solution de problèmes voisins. Le chapitre se termine sur des interprétations possibles de la solution du maximum de vraisemblance comme solution de moindres carrés sous contraintes dans des modèles avec erreurs multiplicatives.

§ I.1. Généralités

On suppose que l'hypothèse \mathfrak{H} est conique : lorsque $\Lambda \in \mathfrak{H} \subset \mathcal{S}^+(m)$, il en est de même des matrices $\lambda\Lambda$, $\lambda > 0$. Pour toute matrice M , carrée d'ordre m , on note \bar{M} la matrice, de déterminant égal à un, $\bar{M} = |M|^{-1/m} M$. La variation radiale de $L(\cdot; X)$ est donnée, pour tout Λ dans \mathfrak{H} , par la variation, par rapport à λ , de la fonction

$$L(\lambda\bar{\Lambda}; X) = m[\ln \lambda + \lambda_0 / \lambda], \quad \lambda_0 = \text{tr}(\bar{\Lambda}^{-1} X) / m,$$

où λ_0 est strictement positif si X est non nul. Cette fonction tend vers $+\infty$ en 0^+ et en $+\infty$. Elle atteint son minimum en λ_0 (seule valeur annulant sa dérivée) et ce minimum vaut $m[1 + \ln \lambda_0]$. Ainsi $L(\cdot; X)$ n'admet pas de maximum local dans \mathfrak{H} et les trois problèmes suivants, (C) désignant une contrainte, sont équivalents :

$$\text{problème 1: } \underset{\Lambda \in \mathfrak{H}}{\text{Min}} L(\Lambda; X) = \underset{\Lambda \in \mathfrak{H}}{\text{Min}} \{ \ln |\Lambda| + \text{tr}(\Lambda^{-1} X) \};$$

$$\text{problème 2: } \underset{\Lambda \in \mathfrak{H}}{\text{Min}} \text{tr}(\Lambda^{-1} X) / m, \quad (\text{C}): |\Lambda| = 1 \text{ (ou } |\Lambda| \leq 1);$$

$$\text{problème 3: } \underset{\Lambda \in \mathfrak{H}}{\text{Min}} |\Lambda|, \quad (\text{C}): \text{tr}(\Lambda^{-1} X) / m = 1 \text{ (ou } \text{tr}(\Lambda^{-1} X) / m \leq 1).$$

Pour $\Lambda \in \mathcal{S}^+$ et $\tilde{\Lambda} \in \mathcal{S}$ il existe une matrice Q et une matrice diagonale Δ telles que $Q' \Lambda Q = I$ et $Q' \tilde{\Lambda} Q = \Delta$. On dit que (Q, Δ) est une diagonalisation simultanée des matrices Λ et $\tilde{\Lambda}$. On note $\delta_1, \dots, \delta_m$ les éléments diagonaux de Δ , $Q_{.k}$ (resp. Q_k) le vecteur colonne défini par la $k^{\text{ième}}$ colonne (resp. ligne) de Q .

H désigne le sous-espace vectoriel de \mathcal{S} engendré par l'hypothèse \mathfrak{H} , H^+ et \bar{H}^+ sont alors les cônes de H formés respectivement par les matrices définies positives et les matrices positives.

I.1.1. Lemme.—Les deux premières dérivées de la fonction $\Lambda \longrightarrow \ln |\Lambda|$, définie sur H^+ , sont données par :

$$dA \in H \longrightarrow \text{tr}(\Lambda^{-1} dA),$$

$$(dA, dB) \in H^2 \longrightarrow -\text{tr}(\Lambda^{-1}dA\Lambda^{-1}dB).$$

Cette fonction est concave et strictement croissante au sens où

$$\forall A \in \bar{H}^+, \quad \ln |\Lambda + A| \geq \ln |\Lambda|,$$

l'égalité n'étant vérifiée que pour $A = 0$.

Démonstration.—Soit (Q, Δ) une diagonalisation des matrices Λ et dA . Posant $g(\lambda) = \ln |\Lambda + \lambda dA|$, on a $g(\lambda) = \ln |\Lambda| + \ln |I + \lambda \Delta|$ et, par dérivation :

$$g'(\lambda) = \text{tr}\{\Delta(I + \lambda \Delta)^{-1}\}, \quad g'(0) = \text{tr}(\Delta) = \text{tr}(Q'dAQ) = \text{tr}(QQ'dA) = \text{tr}(\Lambda^{-1}dA).$$

$dB \longrightarrow \text{tr}(\Lambda^{-1}dB)$ est donc la dérivée première. Avec la même diagonalisation, on pose $h(\lambda) = \text{tr}\{(\Lambda + \lambda dA)^{-1}dB\}$; on a $h(\lambda) = \text{tr}\{(I + \lambda \Delta)^{-1}Q'dBQ\}$ et par suite :

$$h'(\lambda) = -\text{tr}\{\Delta(I + \lambda \Delta)^{-2}Q'dBQ\},$$

$$h'(0) = -\text{tr}(\Delta Q'dBQ) = -\text{tr}(Q\Delta Q'dB) = -\text{tr}(QQ'dAQQ'dB) = -\text{tr}(\Lambda^{-1}dA\Lambda^{-1}dB).$$

On obtient ainsi la dérivée seconde et de plus $\Lambda \longrightarrow \ln |\Lambda|$ est concave puisque pour tout dA de H on a $-\text{tr}(\Lambda^{-1}dA\Lambda^{-1}dA) < 0$. Enfin la croissance stricte de la fonction s'obtient en remarquant que pour A dans \bar{H}^+ et une diagonalisation $(Q, \tilde{\Delta})$ du couple (Λ, A) on a $\ln |\Lambda + A| = \ln |\Lambda| + \ln |I + \tilde{\Delta}|$. ♦

I.1.2. Lemme.—Les deux premières dérivées de la fonction $\Lambda \longrightarrow \text{tr}(\Lambda^{-1}X)$, définie sur H^+ , sont données par

$$dA \in H \longrightarrow -\text{tr}(\Lambda^{-1}X\Lambda^{-1}dA),$$

$$(dA, dB) \in H^2 \longrightarrow 2\text{tr}(\Lambda^{-1}dA\Lambda^{-1}X\Lambda^{-1}dB).$$

Cette fonction est convexe et strictement décroissante au sens où

$$\forall A \in \bar{H}^+, \quad \text{tr}\{(\Lambda + A)^{-1}X\} \leq \text{tr}(\Lambda^{-1}X),$$

l'égalité n'étant vérifiée que pour les matrices A satisfaisant $\text{tr}(\Lambda^{-1}A\Lambda^{-1}X) = 0$.

Démonstration.—Elle est analogue à celle du Lemme I.1.1 sauf pour le dernier point. Mais si (Q, Δ) est une diagonalisation des matrices Λ et A on a

$$\text{tr}\{(\Lambda + A)^{-1}X\} = \text{tr}\{(I + \Delta)^{-1}Q'XQ\} = \sum_{k=1}^m \frac{y_k^2}{1 + \delta_k},$$

$$\text{tr}(\Lambda^{-1}X) = \text{tr}(Q'XQ) = \sum_{k=1}^m y_k^2,$$

où y_k^2 , $k=1, \dots, m$, sont les éléments diagonaux de $Q'XQ$. Comme A est positive si et seulement si $\delta_k \geq 0$, $k=1, \dots, m$, l'égalité

$$\sum_{k=1}^m \frac{y_k^2}{1+\delta_k} = \sum_{k=1}^m y_k^2$$

est équivalente à $\delta_k y_k^2 = 0$, $k=1, \dots, m$, ou encore à $\sum \delta_k y_k^2 = 0$, c'est-à-dire

$$0 = \text{tr}(\Delta Q'XQ) = \text{tr}(Q'AQQ'XQ) = \text{tr}(\Lambda^{-1}A\Lambda^{-1}X). \spadesuit$$

Notons l'implication : $A \in \bar{H}^+$, $\text{tr}(\Lambda^{-1}A\Lambda^{-1}X) = 0 \Rightarrow \ker(A) + \ker(X) = \mathbb{R}^m$, où $\ker(A)$ désigne le noyau de la forme quadratique associée à A .

Ainsi la fonction $L(\cdot; X)$ se présente comme la différence de deux fonctions convexes et sa minimisation (existence, unicité) n'est donc pas résolue par les procédures standards.

§ I.2. Comportement de la fonction de vraisemblance au bord de son domaine

On étudie en fait le comportement de $L(\cdot; X)$ au bord du cône H^+ associé à l'hypothèse \mathcal{H} . La proposition suivante fournit le comportement non tangentiel de $L(\cdot; X)$ au voisinage d'un Λ_0 de la frontière $\bar{H}^+ \setminus H^+$.

I.2.1. Proposition.—(Dégerine, 1985) *Pour tout Λ_0 de $\bar{H}^+ \setminus H^+$ et tout Λ de H^+ , posant $\Lambda_\varepsilon = \Lambda_0 + \varepsilon(\Lambda - \Lambda_0)$, $\varepsilon > 0$, la limite de $L(\Lambda_\varepsilon; X)$, lorsque ε tend vers zéro, existe; elle vaut $-\infty$ si $\ker(\Lambda_0)$ est inclus dans $\ker(X)$ et $+\infty$ sinon.*

Démonstration.— (Q, Δ) désignant une diagonalisation des matrices Λ et Λ_0 on a

$$L(\Lambda_\varepsilon; X) = \ln |\Lambda| + \sum_{k=1}^m \ln [\delta_k + \varepsilon(1-\delta_k)] + \sum_{k=1}^m \frac{Q'_{\cdot k} X Q_{\cdot k}}{\delta_k + \varepsilon(1-\delta_k)}.$$

$\delta_k = 0$ équivaut à $Q_{\cdot k} \in \ker(\Lambda_0)$; si $\ker(\Lambda_0)$ est inclus dans $\ker(X)$, pour tout δ_k nul on a

$Q'_{.k}XQ_{.k}=0$ et par conséquent

$$\lim_{\varepsilon \rightarrow 0} \operatorname{tr}(\Lambda_{\varepsilon}^{-1}X) = \sum_{k:\delta_k > 0} \frac{Q'_{.k}XQ_{.k}}{\delta_k} < +\infty.$$

Comme la limite de $\ln |\Lambda_{\varepsilon}|$, lorsque ε tend vers zéro, est égale à $+\infty$, l'implication

$$\ker(\Lambda_0) \subset \ker(X) \Rightarrow \lim_{\varepsilon \rightarrow 0} L(\Lambda_{\varepsilon}; X) = -\infty$$

est démontrée. Si $\ker(\Lambda_0)$ n'est pas inclus dans $\ker(X)$, il existe un k pour lequel δ_k est nul et $Q'_{.k}XQ_{.k}$ strictement positif. Alors $L(\Lambda_{\varepsilon}; X)$ est équivalent, lorsque ε tend vers zéro, à $r \ln \varepsilon + C/\varepsilon$, où r est la multiplicité de la valeur propre nulle de Λ_0 et C une constante strictement positive. Sa limite sera donc égale à $+\infty$. ♦

L'exemple qui suit montre que cette proposition ne suffit pas pour préciser le comportement de $L(\cdot; X)$ à la frontière de H^+ en toute généralité :

$$\Lambda_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, X = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \Lambda_{\varepsilon} = \begin{pmatrix} -2/\varepsilon & 0 & 0 \\ \varepsilon & \varepsilon & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

lorsque ε tend vers zéro, Λ_{ε} tend vers Λ_0 et $L(\Lambda_{\varepsilon}; X) = \ln \varepsilon - 1/\varepsilon$ tend vers $-\infty$.

La proposition ci-dessus montre qu'une condition nécessaire pour que la vraisemblance soit bornée dans H^+ , X étant fixé, est que l'on ait $\ker(\Lambda_0) \not\subset \ker(X)$ pour tout Λ_0 de $\bar{H}^+ \setminus H^+$. Le résultat ci-après donne une condition suffisante pour qu'elle le soit au voisinage d'un Λ_0 de $\bar{H}^+ \setminus H^+$.

I.2.2. Proposition.—(Dégerine, 1985) *Pour tout Λ_0 de $\bar{H}^+ \setminus H^+$ on a*

$$\ker(\Lambda_0) \cap \ker(X) = \{0\} \Rightarrow \lim_{\Lambda \rightarrow \Lambda_0} L(\Lambda; X) = +\infty.$$

Démonstration.—Soient $P'\Delta P = \Lambda$ et $Q'\Delta_0 Q = \Lambda_0$ des décompositions spectrales des matrices Λ et Λ_0 dans lesquelles les éléments des matrices diagonales Δ et Δ_0 , valeurs propres de Λ et de Λ_0 , sont rangés par ordre croissant. Les matrices Δ et Δ_0 , ainsi

associées à Λ et Λ_0 , sont uniques et Δ converge vers Δ_0 lorsque Λ tend vers Λ_0 . Soit r la multiplicité de la valeur propre nulle de Λ_0 et décomposons la matrice $A = P'Q$ sous la forme :

$$A = \begin{bmatrix} A^r & A^{rq} \\ A^{qr} & A^q \end{bmatrix},$$

où A^r et A^q sont des blocs carrés $r \times r$ et $q \times q$, $q+r=m$. La matrice A^{qr} tend vers $[0]$ car, comme $A'\Delta A = Q'\Lambda Q$ tend vers Δ_0 , on a pour $i \leq r$:

$$\sum_k a_{ki}^2 \delta_k \longrightarrow 0 \Rightarrow a_{ki} \longrightarrow 0 \text{ si } k > r.$$

Alors $(A^r)A^r$ tend vers I_r , car $A'A = I$, et par suite A^{rq} tend vers $[0]$ car $AA' = I$. On pourrait montrer que la matrice A^q "tend vers" une matrice diagonale par blocs; ces blocs correspondent aux différentes valeurs propres de Λ_0 et représentent des transformations orthogonales dans les sous-espaces propres correspondants. Si Δ_r désigne la matrice diagonale d'ordre r formée par les r plus petites valeurs propres de Λ et Q_r la matrice $m \times r$ formée par les r premières colonnes de Q on a, lorsque Λ tend vers Λ_0 :

$$\text{tr}(\Lambda^{-1}X) = \text{tr}(\Lambda^{-1}AQ'XQA) \sim \text{tr}[\Delta_r^{-1}A^rQ_r'XQ_r(A^r)].$$

La forme quadratique définie par $Q_r'XQ_r$ est fixée car elle ne dépend que de X et de Λ_0 . Elle est définie positive; en effet, pour tout $z \neq 0$ de \mathbb{R}^r , on a

$$z'Q_r'XQ_rz = (Q_rz)'X(Q_rz) > 0,$$

puisque Q_rz , qui appartient à $\ker(\Lambda_0)$, ne peut appartenir à $\ker(X)$ selon l'hypothèse $\ker(\Lambda_0) \cap \ker(X) = \{0\}$. Désignant par δ sa plus petite valeur propre, le comportement de $L(.,X)$ au voisinage de Λ_0 est le même que celui de la fonction

$$\sum_{k=1}^r (\ln \delta_k + \delta/\delta_k)$$

qui tend vers $+\infty$ lorsque les δ_k tendent vers zéro. ♦

Enfin pour une norme donnée quelconque sur H le comportement à l'infini de $L(.,X)$ est précisé par la proposition suivante (Dégerine, 1985).

I.2.3. Proposition.—Si pour tout Λ_0 de \bar{H}^+ on a $\ker(\Lambda_0) \cap \ker(X) = \{0\}$, alors :

$$\lim_{\Lambda: \|\Lambda\| \rightarrow +\infty} L(\Lambda; X) = +\infty.$$

Démonstration. Notons $\Lambda = \lambda \tilde{\Lambda}$, $\|\tilde{\Lambda}\| = 1$, le terme général d'une suite telle que λ tende vers $+\infty$ et posons $\tilde{\lambda} = \text{tr}(\tilde{\Lambda}^{-1}X)$. La fonction $L(\Lambda; X)$ est minorée :

$$L(\Lambda; X) = \ln |\tilde{\Lambda}| + m \ln \lambda + \tilde{\lambda} / \lambda \geq \ln |\tilde{\Lambda}| + m[\ln(\tilde{\lambda}/m) + 1].$$

Soit $\tilde{\Lambda}_0$ un point d'accumulation de la suite de terme général $\tilde{\Lambda}$. Si $\tilde{\Lambda}_0$ est non singulière $L(\Lambda; X)$ est équivalent à $m \ln \lambda$ et tend vers $+\infty$. Sinon, en utilisant les notations de la démonstration précédente pour le couple $(\tilde{\Lambda}, \tilde{\Lambda}_0)$, $\tilde{\lambda}$ s'écrit :

$$\tilde{\lambda} = \text{tr}(\tilde{\Lambda}^{-1}X) = \text{tr}(\Delta^{-1}PXP) = \sum_{k=1}^m y_k^2 / \delta_k$$

où (y_1^2, \dots, y_m^2) [resp. $(\delta_1, \dots, \delta_m)$] est la suite des éléments diagonaux de PXP (resp.

Δ) On sait d'après la proposition ci-dessus que, puisque $\tilde{\Lambda}$ tend vers la matrice singulière $\tilde{\Lambda}_0$, $\tilde{\lambda}$ tend vers $+\infty$. En minorant $\delta_k \tilde{\lambda}$ par y_k^2 , $k=1, \dots, r$, on obtient :

$$L(\Lambda; X) \geq \sum_{k=1}^m \ln(\delta_k \tilde{\lambda} / m) + m \geq \sum_{k \leq r} \ln(y_k^2 / m) + \sum_{k > r} \ln(\delta_k \tilde{\lambda} / m).$$

Puisque les y_k^2 , $k=1, \dots, r$ sont minorés par une quantité qui tend vers $\delta > 0$, lorsque $\tilde{\Lambda}$ tend vers $\tilde{\Lambda}_0$ (cf. Proposition I.2.2), et que $\delta_k \tilde{\lambda}$ tend vers $+\infty$ pour $k > r$, le résultat annoncé est démontré. ♦

Ainsi la condition $\ker(\Lambda_0) \cap \ker(X) = \{0\}$, pour tout Λ_0 de \bar{H}^+ , assure que la vraisemblance tend vers zéro au bord de H^+ et, par conséquent, que son maximum est atteint par une matrice Λ non singulière. Lorsque le maximum de vraisemblance est atteint en un point de \mathcal{H} qui admet, dans \mathcal{H} , un voisinage de même dimension que H , il est réalisé par une matrice Λ satisfaisant l'équation de vraisemblance :

$$\text{grad}\{L(\Lambda; X)\} = 0 \Leftrightarrow \forall dA \in H, \text{tr}(\Lambda^{-1}dA) - \text{tr}(\Lambda^{-1}X\Lambda^{-1}dA) = 0.$$

Cependant des situations classiques sont souvent très complexes. Considérons le

cas d'un échantillon de séries de longueur m d'un modèle $MA(r)$ avec $r \leq q$, où q est fixé dans $\{1, \dots, m-1\}$. \mathcal{H} est un cône convexe, dans l'espace \mathcal{C} des matrices de Toeplitz, qui engendre le sous-espace vectoriel H des matrices de \mathcal{C} vérifiant $\Lambda(i, i+j)=0$ pour $j > q$. Mais \mathcal{H} est strictement inclus dans le cône positif H^+ . Il est vraisemblable que $L(\cdot; X)$ puisse atteindre son minimum global dans $H^+ \setminus \mathcal{H}$; dans ce cas on peut envisager que le minimum dans \mathcal{H} soit atteint sur la frontière sans satisfaire l'équation de vraisemblance.

§ I.3. Une caractérisation du maximum de vraisemblance

Soit $E = \mathcal{L} \{ \Lambda^{-1}; \Lambda \in \mathcal{H} \}$ le sous-espace vectoriel de \mathcal{S} engendré par les inverses des éléments de l'hypothèse \mathcal{H} . Posant $M = \Lambda^{-1}$, la fonction $L(\cdot; X)$ définit sur E une fonction $\tilde{L}(\cdot; X)$:

$$\tilde{L}(M; X) = L(\Lambda; X) = -\ln |M| + \text{tr}(MX), M \in E,$$

dont la dérivée est donnée par

$$dA \in E \longrightarrow -\text{tr}(M^{-1}dA) + \text{tr}(XdA).$$

La statistique $X \longrightarrow \hat{X}$, où \hat{X} désigne la projection de X sur E selon le produit scalaire $\langle A, B \rangle = \text{tr}(AB)/m$, est exhaustive car $\text{tr}(MX) = \text{tr}(M\hat{X})$. Lorsqu'elle appartient à \mathcal{H} , la statistique \hat{X} est l'unique solution du problème du maximum de vraisemblance. Ainsi, dès que le cône positif E^+ de E est égal à \mathcal{H} , le problème est résolu. En effet, lorsqu'elle existe, et c'est le cas par exemple si $\ker(\Lambda_0) \cap \ker(\hat{X}) = \{0\}$ pour tout Λ_0 de \bar{H}^+ , l'estimation de maximum de vraisemblance satisfait l'équation de vraisemblance dans E^+ , elle est donc égale à \hat{X} puisque c'est l'unique solution dans E . Cette situation est exceptionnelle, mais c'est évidemment le cas lorsque $\mathcal{H} = \mathcal{S}^+$ ou encore celui des matrices doublement symétriques $\mathcal{H} = \mathcal{C}^+$ considéré par A (1984).

On suppose désormais que $\mathcal{H} = H^+$. Cette hypothèse est équivalente à un ensemble de contraintes linéaires homogènes et symétriques sur les éléments $\Lambda(i, j)$ de la matrice Λ . En effet $\Lambda \in H$ peut se caractériser par des conditions d'orthogonalité du type

$$\langle \Lambda, A \rangle = (\sum_{i,j} A(i,j)\Lambda(i,j))/m = 0, \quad A \in \mathcal{S}.$$

Un exemple important est celui des matrices de Toeplitz que nous étudions dans le chapitre suivant. Comme nous l'avons remarqué plus haut, les modèles MA n'entrent pas dans ce cas; il en est de même pour les modèles AR ou ARMA.

On suppose aussi que le maximum de vraisemblance est atteint dans H^+ . Il est également clair que X peut être remplacé par \hat{X} , y compris pour les problèmes d'existence considérés au paragraphe précédent. Pour tout Λ de \mathcal{S}^+ on note $\langle \cdot, \cdot \rangle_\Lambda$ le produit scalaire défini sur \mathcal{S} par $\langle A, B \rangle_\Lambda = \text{tr}(\Lambda^{-1}A\Lambda^{-1}B)/m$ et $\|\cdot\|_\Lambda$ désigne la norme correspondante. L'équation de vraisemblance se traduit par:

I.3.1. Proposition.—(Dégerine, 1985) *La matrice Λ est solution de l'équation de vraisemblance $\text{grad}\{L(\Lambda; X)\} = 0$, si et seulement si elle est la projection de X sur H selon le produit scalaire $\langle \cdot, \cdot \rangle_\Lambda$*

Remarquons que, si Λ_0 est la vraie valeur de Λ , la projection de X sur H selon $\langle \cdot, \cdot \rangle_{\Lambda_0}$ est un "estimateur" sans biais de Λ_0 . Cela laisse espérer de bonnes propriétés pour l'estimation de maximum de vraisemblance. D'autre part cette proposition suggère une technique de calcul itérative; cependant il est nécessaire, à chaque itération, que la projection de X sur H soit dans H^+ et prudent de vérifier l'augmentation de la vraisemblance.

Notons que la forme quadratique définie par les dérivées secondes de $L(\cdot; X)$ s'écrit

$$(dA, dB) \longrightarrow m\{2\langle X, dA\Lambda^{-1}dB \rangle_\Lambda - \langle dA, dB \rangle_\Lambda\}.$$

Une condition suffisante pour que Λ_0 , solution de l'équation de vraisemblance, réalise un minimum, éventuellement local, de $L(\cdot; X)$ est que l'on ait :

$$\forall A \in H, \|A\|_{\Lambda_0} = 1 \Rightarrow 2\langle X, A\Lambda_0^{-1}A \rangle_{\Lambda_0} > 1,$$

ou, avec les mêmes contraintes sur A :

$$2\langle X - \Lambda_0, A\Lambda_0^{-1}A \rangle_{\Lambda_0} < 1.$$

Ainsi si X est "proche" de Λ_0 (grands échantillons), la recherche du maximum de vraisemblance par des méthodes de gradient peut être envisagée. Il est cependant des situations (cf. Ch. IV) où ces méthodes sont difficilement utilisables.

La proposition suivante donne une caractérisation d'une matrice réalisant le maximum de vraisemblance; selon la notation déjà introduite $\bar{\Lambda}$ désigne la matrice proportionnelle à Λ de déterminant égal à un.

I.3.2. Proposition.—(Dégerine, 1985) *La matrice Λ_0 réalise le maximum de la vraisemblance si et seulement si, pour la métrique définie par $\bar{\Lambda}_0$, la norme de la projection de X sur H est minimale parmi celles obtenues pour les métriques définies par les matrices $\bar{\Lambda}$ de H^+ :*

$$\begin{aligned} \langle \bar{\Lambda}_0, X \rangle_{\bar{\Lambda}_0} &= \text{Min} \{ \langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}}, \bar{\Lambda} \in H^+ \} \\ &= \text{Min} \{ \text{Max} [\langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}} / \|\bar{\Lambda}\|_{\bar{\Lambda}}], \bar{\Lambda} \in H \}, \bar{\Lambda} \in H^+ \}. \end{aligned}$$

Démonstration.—Si Λ_0 réalise le maximum de vraisemblance, le problème 2 se formule comme suit

$$\text{Min} \{ \text{tr}(\bar{\Lambda}^{-1}X), \bar{\Lambda} \in H^+ \} = \text{Min} \{ \langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}}, \bar{\Lambda} \in H^+ \} = \langle \bar{\Lambda}_0, X \rangle_{\bar{\Lambda}_0}.$$

Pour tout Λ de H^+ notons $\hat{\Lambda}$ la projection de X sur H selon $\langle, \rangle_{\Lambda}$ ou de manière équivalente selon $\langle, \rangle_{\bar{\Lambda}}$. D'après la Proposition I.3.1 Λ_0 vérifie :

$$\langle \bar{\Lambda}_0, X \rangle_{\bar{\Lambda}_0} = \langle \bar{\Lambda}_0, \hat{\Lambda}_0 \rangle_{\bar{\Lambda}_0} = \langle \bar{\Lambda}_0, \Lambda_0 \rangle_{\bar{\Lambda}_0} = \|\hat{\Lambda}_0\|_{\bar{\Lambda}_0}.$$

On a d'autre part

$$\text{Max} [\langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}} / \|\bar{\Lambda}\|_{\bar{\Lambda}}, \bar{\Lambda} \in H] = \langle \hat{\Lambda}, X \rangle_{\bar{\Lambda}} / \|\hat{\Lambda}\|_{\bar{\Lambda}} = \|\hat{\Lambda}\|_{\bar{\Lambda}},$$

et par conséquent :

$$\text{Min} [\|\hat{\Lambda}\|_{\bar{\Lambda}}, \bar{\Lambda} \in H^+] \leq \|\hat{\Lambda}_0\|_{\bar{\Lambda}_0}.$$

L'autre inégalité s'obtient par passage au minimum sur $\bar{\Lambda} \in H^+$ dans la suivante :

$$\langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}} = \langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}} / \|\bar{\Lambda}\|_{\bar{\Lambda}} \leq \text{Max} [\langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}} / \|\bar{\Lambda}\|_{\bar{\Lambda}}, \bar{\Lambda} \in H]. \blacklozenge$$

Il est clair que l'on peut caractériser de façon analogue tout minimum local de

$L(\cdot; X)$ dans H^+ , lequel pouvant être la solution du maximum de vraisemblance lorsque \mathfrak{H} n'est pas égal à H^+ . Il suffit de substituer la contrainte $\bar{\Lambda} \in H^+ \cap \mathcal{V}$, où \mathcal{V} est un voisinage adéquat de $\bar{\Lambda}_0$, à $\bar{\Lambda} \in H^+$ dans les deux problèmes de minimisation.

La projection de X sur H selon \langle, \rangle , lorsqu'elle est dans \bar{H}^+ , est la solution au problème des moindres carrés :

$$\text{Min} \{ \|X - \Lambda\|, \Lambda \in \bar{H}^+ \}.$$

Elle ne correspond donc pas en général au maximum de vraisemblance.

De même le problème

$$\text{Min} \{ \|\Lambda^{-1/2} X \Lambda^{-1/2} - I\|, \Lambda \in H^+ \} = \text{Min} \{ \|X - \Lambda\|_{\Lambda}, \Lambda \in H^+ \},$$

de recherche, parmi les structures de covariance empirique associées à $\Lambda^{-1/2} X$, $\Lambda \in H^+$, de celle qui est la plus proche de I équivaut à rechercher $\bar{\Lambda}$ dans H^+ tel que, pour la métrique $\langle, \rangle_{\bar{\Lambda}}$, l'angle entre X et $\bar{\Lambda}$ soit minimal. La dérivée de $\|X - \Lambda\|_{\Lambda}^2$ est donnée par

$$dA \in H \longrightarrow 2 \langle X - X \Lambda^{-1} X, A \rangle_{\Lambda},$$

et donc une solution Λ est telle que les matrices X et $X \Lambda^{-1} X$ ont même projection sur H selon $\langle, \rangle_{\Lambda}$. Une telle métrique ne minimise pas en général la quantité $\langle \bar{\Lambda}, X \rangle_{\bar{\Lambda}}$

En dehors de la situation évoquée plus haut, l'unicité de la solution de l'équation de vraisemblance doit être examinée en fonction de \mathfrak{H} mais aussi de X . Nous verrons, au chapitre suivant, que déjà dans le cas des matrices de Toeplitz elle n'est pas garantie et que $L(\cdot; X)$ peut présenter plusieurs minima locaux.

§ I.4. Moindres carrés sous contrainte

Considérons la situation classique d'un vecteur aléatoire X à valeurs dans \mathbb{R}^p , d'espérance m_X et de matrice de covariance Λ_X . A un échantillon X_1, \dots, X_n , de taille n , $n > p$, issu de X on associe la moyenne et la matrice de covariance empiriques :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \hat{\Lambda}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'.$$

Il est bien connu que \bar{X} et $\hat{\Lambda}_X$ sont des estimateurs de m_X et Λ_X qui possèdent de bonnes propriétés : ils sont sans biais ($n\hat{\Lambda}_X/(n-1)$), convergents et de maximum de vraisemblance dans le cas gaussien. On sait de plus que \bar{X} est un estimateur des moindres carrés; nous montrons ici qu'il est possible d'interpréter $\hat{\Lambda}_X$ de façon analogue.

Considérons le modèle linéaire

$$X = m_X + \varepsilon$$

dans lequel ε est un vecteur aléatoire à valeurs dans \mathbb{R}^p , centré, de matrice de covariance Λ_X . Dans ce modèle ε s'interprète comme une erreur additive sur la mesure de m_X que représente X . Ainsi, pour estimer m_X , il est naturel d'utiliser la méthode des moindres carrés, c'est-à-dire de retenir le vecteur m de \mathbb{R}^p qui minimise la somme des carrés des normes des écarts entre les différents X_i et m . Il reste cependant à choisir une norme sur \mathbb{R}^p ; en se restreignant aux normes euclidiennes et en notant M la matrice de \mathcal{S}^+ correspondante on a :

$$\sum_{i=1}^n (X_i - m)' M (X_i - m) = \sum_{i=1}^n (X_i - \bar{X})' M (X_i - \bar{X}) + n(\bar{X} - m)' M (\bar{X} - m).$$

Ceci prouve que \bar{X} est l'estimateur des moindres carrés de m_X indépendamment du choix de M et de la matrice de covariance Λ_X . La relation

$$E\{(X - m)' M (X - m)\} = E\{(X - m_X)' M (X - m_X)\} + (m_X - m)' M (m_X - m),$$

montre que m_X est caractérisé, dans le modèle, comme le vecteur m de \mathbb{R}^p qui minimise l'espérance du carré de la norme de l'écart entre X et m . Ainsi, l'estimateur des moindres carrés \bar{X} possède relativement à la loi empirique de l'échantillon la même propriété que m_X relativement à la loi de X .

Dans la relation ci-dessus le minimum de $E\{(X - m)' M (X - m)\}$ n'est atteint que par $m = m_X$ et vaut $\text{tr}(M\Lambda_X)$. On peut envisager alors de minimiser $\text{tr}(M\Lambda_X)$ par le choix de M mais on constate que, sans contrainte sur M , le problème n'a pas d'intérêt

car $\text{Inf} \{ \text{tr}(M\Lambda_X), M \in \mathcal{J}^+ \} = 0$.

La formulation de la méthode du maximum de vraisemblance, dans le premier paragraphe, selon le problème 2 invite à utiliser une contrainte portant sur le déterminant de M^{-1} . Le Lemme 4, Ch. II, de Romeder (1973) peut s'énoncer ainsi :

I.4.1. Lemme.—*Soit Λ une matrice symétrique définie positive. On a :*

$$\text{Min} \{ \text{tr}(M\Lambda); M \in \mathcal{J}^+, |M^{-1}| \leq 1 \} = p[|M|]^{1/p},$$

le minimum n'étant réalisé que par $M = |\Lambda|^{1/p} \Lambda^{-1}$.

Il est également possible d'utiliser une contrainte portant sur la trace de M^{-1} .

I.4.2. Lemme.—(Dégerine, 1982a) *Soit Λ une matrice symétrique définie positive. On a :*

$$\text{Min} \{ \text{tr}(M\Lambda); M \in \mathcal{J}^+, \text{tr}(M^{-1}) \leq 1 \} = [\text{tr}(\Lambda^{1/2})]^2,$$

le minimum n'étant réalisé que par $M = \text{tr}(\Lambda^{1/2}) \Lambda^{-1/2}$.

Démonstration.—Soit C_K le compact défini par

$$C_K = \{ M \in \mathcal{J}^+; \text{tr}(M^{-1}) \leq 1; \text{tr}(M) \leq K \}.$$

Pour K suffisamment grand, C_K est non vide et la fonction continue $M \longrightarrow \text{tr}(M\Lambda)$ y atteint ses extrema. Soit M_0 une valeur en laquelle est réalisé le minimum. Les conditions $M_0 \in \mathcal{J}^+$ et $\text{tr}(M_0^{-1}) \leq 1$ impliquent que toutes les valeurs propres de M_0 sont minorées par un. Il existe donc un voisinage de M_0 dans \mathcal{J} entièrement contenu dans \mathcal{J}^+ et ainsi M_0 réalise un minimum local de la fonction $\text{tr}(M\Lambda)$ considérée sur \mathcal{J} avec les contraintes $M \in \mathcal{J}^+$, $\text{tr}(M^{-1}) \leq 1$ et $\text{tr}(M) \leq K$. La matrice M_0 étant située à l'intérieur de \mathcal{J}^+ , la contrainte $M \in \mathcal{J}^+$ n'est pas active. De même la contrainte $\text{tr}(M) \leq K$ n'est plus active dès que K est suffisamment grand. En effet la valeur de $\text{tr}(M_0\Lambda)$ ne peut pas croître lorsque K augmente puisqu'il s'agit du minimum de $\text{tr}(M\Lambda)$ sur C_K et que C_K croît avec K . Or on a :

$$\text{tr}(M\Lambda) \geq \lambda_p(\Lambda) \text{tr}(M),$$

où $\lambda_p(\Lambda)$ désigne la plus petite valeur propre de Λ . Ainsi si la contrainte $\text{tr}(M) \leq K$ restait active la valeur de $\text{tr}(M_0\Lambda)$ deviendrait infinie en même temps que K .

La matrice M_0 doit satisfaire les conditions suivantes :

$$d_M \{ \text{tr}(M\Lambda) - \lambda[1 - \text{tr}(M^{-1})] \} = 0, \quad \lambda[1 - \text{tr}(M^{-1})] = 0, \quad \lambda \geq 0,$$

où d_M signifie que l'on dérive selon M . L'équivalence

$$d_M \{ \text{tr}(M\Lambda) - \lambda[1 - \text{tr}(M^{-1})] \} = 0 \Leftrightarrow \text{tr}\{dM[\Lambda - \lambda M^{-2}]\} = 0,$$

cette dernière égalité devant être réalisée pour tout choix de la direction dM , conduit aux conditions nécessaires suivantes:

$$\Lambda - \lambda M^{-2} = 0, \quad \lambda[1 - \text{tr}(M^{-1})] = 0, \quad \lambda \geq 0.$$

La seule solution est obtenue pour

$$M_0 = \text{tr}(\Lambda^{1/2})\Lambda^{-1/2}, \quad \lambda = \lambda \text{tr}(M_0^{-1}) = \text{tr}(M_0\Lambda) = [\text{tr}(\Lambda^{1/2})]^2. \blacklozenge$$

Une interprétation géométrique des contraintes permet de mieux les comparer. Pour cela considérons un système d'axes orthonormés représentant le vectoriel \mathbb{R}^p muni de la métrique identité. Soit $M=U'D^2U$ une décomposition spectrale d'une nouvelle métrique M sur \mathbb{R}^p ; les vecteurs lignes de la matrice $D^{-1}U$ constituent une base M -orthogonale de \mathbb{R}^p ; ils définissent un nouveau système d'axes orthogonaux, les longueurs des unités sur les axes étant données par les éléments diagonaux de D^{-1} . Un élément de \mathbb{R}^p de coordonnées x dans le système d'axes initial a pour coordonnées $y = DUx$ dans le nouveau et on a :

$$\|y\|_{I_p}^2 = y'y = x'U'D^2Ux = x'Mx = \|x\|_M^2$$

Le parallélépipède formé sur les vecteurs unitaires du nouveau système d'axes a pour volume $[|M^{-1}|]^{1/2}$ et la longueur de sa diagonale principale est $[\text{tr}(M^{-1})]^{1/2}$. Ainsi la conservation du volume se traduit par $|M^{-1}| = 1$ et la conservation de la longueur de la diagonale par $\text{tr}(M^{-1}) = p$.

Les propositions suivantes (Dégerine, 1982a) sont des conséquences directes des lemmes précédents.

I.4.3. Proposition.—*Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^p , d'espérance m_X et de matrice de covariance Λ_X non singulière. On a :*

$$E \{(X - m)'M(X - m)\} \geq p[|\Lambda_X|]^{1/p}, \quad \forall m \in \mathbb{R}^p, \quad \forall M \in \mathcal{S}^+, \quad |M^{-1}| \leq 1,$$

l'égalité n'étant réalisée que par $m = m_X$ et $M = [|\Lambda_X|]^{1/p} \Lambda_X^{-1}$.

I.4.4. Proposition.—Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^p , d'espérance m_X et de matrice de covariance Λ_X non singulière. On a :

$$E \{(X - m)'M(X - m)\} \geq [\text{tr}(\Lambda_X^{1/2})]^2, \quad \forall m \in \mathbb{R}^p, \quad \forall M \in \mathcal{S}^+, \quad \text{tr}(M^{-1}) \leq 1,$$

l'égalité n'étant réalisée que par $m = m_X$ et $M = \text{tr}(\Lambda_X^{1/2}) \Lambda_X^{-1/2}$.

Avec les hypothèses de la proposition ci-dessus soit $\Lambda_X = U'_X \Delta^2 U_X$ une décomposition spectrale de Λ_X . Le choix optimal de M conduit à définir un terme d'erreur ε en posant :

$$\text{tr}(\Lambda_X^{1/2})(X - m_X)' \Lambda_X^{-1/2} (X - m_X) = \varepsilon' \varepsilon,$$

avec

$$\varepsilon = [\text{tr}(\Delta_X)]^{1/2} \Delta_X^{-1/2} U_X (X - m_X).$$

Cela suggère de poser pour X le modèle suivant :

$$X = m_X + U'_X D_X^{1/2} \varepsilon, \quad D_X = [\text{tr}(\Delta_X)]^{-1} \Delta_X.$$

Dans ce modèle l'erreur ε est centrée, de matrice de covariance diagonale $\sigma_\varepsilon^2 D_X$ avec $\text{tr}(D_X) = 1$. En notant α_X^j les éléments diagonaux de D_X , U_X^j les vecteurs lignes de U_X et ε^j les composantes de ε on a :

$$X = m_X + \sum_{j=1}^p \varepsilon^j \sqrt{\alpha_X^j} U_X^j.$$

Lorsque les valeurs propres de Λ_X sont distinctes, ce modèle exprime que X est la résultante des composantes principales $\langle X, U_X^j \rangle U_X^j$, chacune traduisant l'effet sur X de la variable correspondante $\langle X, U_X^j \rangle$; cette variable mesure un effet moyen $\langle m_X, U_X^j \rangle$ avec une erreur ε^j de type multiplicatif dont la variance est proportionnelle à un paramètre d'échelle α_X^j . On peut aussi considérer que les $\sqrt{\alpha_X^j}$ constituent un système d'unités de

mesure pour les variables $\langle X, U_X^j \rangle$ et qu'alors les erreurs ϵ^j exprimées dans les unités ainsi déterminées ont même variance. Pour une valeur propre multiple les variables $\langle X, U_X^j \rangle$ qui lui sont associées mesurent un effet vectoriel dont les composantes ne peuvent être dissociées au vu de la variabilité qu'elles induisent sur X . Cette représentation de X possède un caractère optimal, y compris dans le cas où Λ_X est singulière.

I.4.5. Proposition.—(Dégerine, 1982a) Parmi les représentations de X sous la forme :

$$X = m + U'D^{1/2}\epsilon, \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \text{tr}(D) \leq 1,$$

où D est une matrice diagonale à éléments positifs ou nuls, celles obtenues à partir de m_X et d'une décomposition spectrale de Λ_X :

$$\Lambda_X = \sigma_\epsilon^2 U_X' D_X^2 U_X, \quad \text{tr}(D_X) = 1,$$

sont les seules à réaliser le minimum de $E\{\epsilon'\epsilon\}$. La valeur de ce minimum est égale à :

$$\sigma_\epsilon^2 = \left[\sum_{j=1}^p \lambda_j \right]^2,$$

où $\lambda_1, \dots, \lambda_p$ sont les racines carrées des valeurs propres de Λ_X .

Démonstration.—On a $E\{\epsilon'\epsilon\} = E\{[\epsilon - E(\epsilon)]'[\epsilon - E(\epsilon)]\} + [E(\epsilon)]'[E(\epsilon)]$; l'espérance $E\{\epsilon'\epsilon\}$ est donc minimum lorsque ϵ est centré, c'est-à-dire $m = m_X$. En effet dans le cas contraire, en posant $\tilde{\epsilon} = \epsilon - E(\epsilon)$, on aurait $E\{\tilde{\epsilon}'\tilde{\epsilon}\} < E\{\epsilon'\epsilon\}$ puisque

$$X = m + U'D^{1/2}E(\epsilon) + U'D^{1/2}\tilde{\epsilon}, \quad m + U'D^{1/2}E(\epsilon) = m_X \in \mathbb{R}^p.$$

Pour simplifier les notations on pose $m_X = m = E(\epsilon) = 0$. On se place dans le cas où Λ_X est singulière. X est presque sûrement à valeurs dans le sous-espace V de \mathbb{R}^p de dimension q , $q < p$, orthogonal dans \mathbb{R}^p du noyau de Λ_X . En notant α^j les éléments diagonaux de D , la condition $U'U = I_p$ dans la représentation

$$X = U'D^{1/2}\epsilon = \sum_{j=1}^p \epsilon^j \sqrt{\alpha^j} U^j,$$

montre que seuls q termes, parmi les $\epsilon^j \sqrt{\alpha^j}$, sont non nuls et que les U^j correspondant aux termes nuls constituent une base de l'orthogonal de V . Moyennant un changement

d'indice la représentation s'écrit :

$$X = \sum_{j=1}^q \varepsilon^j \sqrt{\alpha^j} U^j, \quad \alpha = \sum_{j=1}^q \alpha^j \leq 1.$$

On a

$$E\{\varepsilon^j \varepsilon^k\} = \sum_{j=1}^q \frac{1}{\alpha^j} E(\langle X, U^j \rangle^2) \geq \sum_{j=1}^q \frac{\alpha^j}{\alpha^j} E(\langle X, U^j \rangle^2).$$

L'espérance $E(\varepsilon^j \varepsilon^k)$ est donc minimum lorsque $\alpha = 1$. Ainsi $\alpha^j = 0, j=q+1, \dots, p$, et les vecteurs U^j correspondants forment une base du sous-espace propre associé à la valeur propre nulle de Λ_X . En se plaçant dans V on retrouve le problème initial en dimension q avec une matrice Λ_X non singulière et la contrainte $\text{tr}(D) = 1$. Pour ne pas changer les notations on suppose Λ_X non singulière. On a

$$E(\varepsilon^j \varepsilon^k) = \text{tr}(M \Lambda_X), \quad M = U'D^{-1}U, \quad \text{tr}(M^{-1}) = 1.$$

La solution est alors donnée par le Lemme I.4.2. ♦

On peut considérer une représentation de X basée sur le résultat de la Proposition I.4.3. On est alors conduit au modèle suivant :

$$X = m_X + U'_X D_X^{1/2} \varepsilon = m_X + \sum_{j=1}^p \varepsilon^j \sqrt{\alpha_X^j} U_X^j, \quad |D_X| = \prod_{j=1}^p \alpha_X^j = 1.$$

Dans ce cas l'erreur ε est centrée, de matrice de covariance $\sigma_\varepsilon^2 I_p$. On a une interprétation analogue mais avec des erreurs ε^j de type multiplicatif qui ont même variance ou encore dont la variance exprimée dans les unités définies par les $\sqrt{\alpha_X^j}$ est d'autant plus faible que l'unité est grande. De plus on doit se restreindre au cas où Λ_X est non singulière.

I.4.6. Proposition.—(Dégerine, 1982a) *Lorsque Λ_X est non singulière, parmi les représentations de X sous la forme*

$$X = m + U'D^{1/2}\varepsilon, \quad m \in \mathbb{R}^p, \quad U'U = I_p, \quad |D| \leq 1,$$

où D est une matrice diagonale à éléments positifs, celles obtenues à partir de m_X et d'une décomposition spectrale de Λ_X :

$$\Lambda_X = \sigma_\varepsilon^2 U' X D_X^2 U_X, |D_X| = 1,$$

sont les seules à réaliser le minimum de $E\{\varepsilon'\varepsilon\}$. La valeur de ce minimum est égale à :

$$p \sigma_\varepsilon^2 = p \left[\prod_{j=1}^p \lambda_j^2 \right]^{1/p},$$

où $\lambda_1, \dots, \lambda_p$ sont les racines carrées des valeurs propres de Λ_X .

Dans les représentations $X = m + U'D^{1/2}\varepsilon$, le critère $E\{\varepsilon'\varepsilon\}$ fait abstraction des éventuelles corrélations entre les composantes de ε mais il est minimisé précisément dans le cas où ces composantes sont non corrélées.

Pour l'estimation des paramètres m_X et Λ_X à partir de l'échantillon $\{X_1, \dots, X_n\}$, la Proposition I.4.5 montre qu'il est naturel de considérer le couple (m, Λ) qui minimise la somme des carrés des normes des écarts ε_i définis par

$$X_i = m + U'D^{1/2}\varepsilon_i, \quad i = 1, \dots, n, \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \text{tr}(D) \leq 1,$$

où D est une matrice diagonale à éléments positifs ou nuls, Λ étant alors donnée par :

$$\Lambda = \sigma_\varepsilon^2 U'D^2U, \quad \sigma_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2.$$

Par analogie avec le modèle linéaire on peut considérer qu'il s'agit d'une méthode de moindres carrés pour un modèle dans lequel les erreurs ont un effet multiplicatif.

I.4.7. Proposition.—(Dégerine, 1982a) Soit X un vecteur aléatoire à valeurs dans \mathbb{R}^p , d'espérance m_X et de matrice de covariance Λ_X . Soit $\{X_1, \dots, X_n\}$ un échantillon de taille $n, n > p$, issu de X . Les éléments empiriques

$$\hat{m}_X = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\Lambda}_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})',$$

sont les estimateurs des moindres carrés de m_X et de Λ_X au sens où $(\hat{m}_X, \hat{\Lambda}_X)$ est l'unique solution du problème de minimisation de la somme des carrés des normes des erreurs ε_i dans la représentation

$$X_i = m + U'D^{1/2}\varepsilon_i, \quad i = 1, \dots, n, \quad m \in \mathbb{R}^p, \quad UU' = I_p, \quad \text{tr}(D) \leq 1,$$

où D est une matrice diagonale à éléments positifs ou nuls associée à Λ par :

$$\Lambda = \sigma_{\varepsilon}^2 U' D^2 U, \quad \sigma_{\varepsilon}^2 = \frac{1}{n} \sum_{i=1}^n \|\varepsilon_i\|^2.$$

Ce résultat est une simple application de la Proposition I.4.5 au vecteur aléatoire Z , à valeurs dans \mathbb{R}^p , muni de la loi empirique définie par l'observation de $\{X_1, \dots, X_n\}$.

Il va de soi que l'on peut énoncer un résultat analogue basé sur la Proposition I.4.6. Toutefois dans ce cas on doit se restreindre à une matrice Λ_X non singulière alors que, dans le cas gaussien, le couple $(\hat{m}_X, \hat{\Lambda}_X)$ est de maximum de vraisemblance y compris lorsque Λ_X est singulière (la vraisemblance étant nulle pour tout Λ dont le sous-espace propre associé à la valeur propre nulle ne coïncide pas avec celui de Λ_X).

CHAPITRE II

ESTIMATION DE MAXIMUM DE VRAISEMBLANCE

D'UNE MATRICE DE TŒPLITZ

On considère désormais le cas stationnaire. On retrouve le formalisme du chapitre précédent avec une matrice aléatoire X constituée par un échantillon de taille n d'une séquence stationnaire de longueur m , gaussienne, centrée et de matrice de covariance Λ . Aucune autre hypothèse n'est faite sur la structure de la séquence. L'hypothèse \mathfrak{H} est ainsi égale au cône positif \mathfrak{C}^+ de l'ensemble des matrices de Tœplitz symétriques \mathfrak{C} d'ordre m .

On reprend, de manière spécifique, le problème de l'existence d'une solution pour la méthode du maximum de vraisemblance dans le premier paragraphe. On introduit, dans le suivant, une notion de décomposition spectrale propre aux matrices de Tœplitz. Celle-ci nous permet d'obtenir, au troisième paragraphe, une méthode de relaxation pour calculer l'estimation. Le quatrième paragraphe est consacré à la présentation de l'algorithme de calcul que nécessite la méthode et à quelques résultats de simulation. Enfin nous montrons, dans le dernier paragraphe, l'existence possible de maxima locaux pour la fonction de vraisemblance.

§ II.1. Le problème d'existence

L'ensemble \mathfrak{C} est un sous-espace vectoriel de dimension m de l'espace des matrices symétriques \mathcal{S} ($\dim \mathcal{S} = m(m+1)/2$). Plus précisément il est inclus dans l'espace des matrices doublement symétriques \mathfrak{E} dont la dimension est égale à $m(m+2)/4$ lorsque m est pair et à $(m+1)^2/4$ lorsqu'il est impair. L'espace \mathfrak{C} étant

stable par inversion on peut projeter la matrice de covariance empirique $X = (xx')/n$ sur \mathcal{C} , c'est-à-dire qu'on la remplace par la matrice doublement symétrique $(X+JXJ)/2$ encore notée X .

On note C^+ l'ensemble des matrices d'autocorrélation Γ associé à \mathcal{C}^+ par la relation $\Lambda = \sigma^2 \Gamma$ où σ^2 est l'élément diagonal de Λ . La minimisation de

$$L(\Lambda; X) = m \ln \sigma^2 + \ln |\Gamma| + \text{tr}(\Gamma^{-1}X)/\sigma^2$$

par rapport à σ^2 se traduit par la contrainte

$$\sigma^2 = \text{tr}(\Gamma^{-1}X)/m.$$

D'où

$$\inf_{\Lambda \in \mathcal{C}^+} L(\Lambda; X) = \inf_{\Gamma \in C^+} m \ln [|\Gamma|^{-1/m} \text{tr}(\Gamma^{-1}X)/m] + m. \quad (1)$$

Le cône C^+ peut être paramétré par l'ensemble des autocorrélations partielles $\{(\beta_1, \dots, \beta_{m-1}); -1 < \beta_j < 1, j=1, \dots, m-1\}$ et (1) peut se formuler comme la minimisation de

$$g(\beta_1, \dots, \beta_{m-1}; x) = |\Gamma|^{1/m} \text{tr}(x \Gamma^{-1} x)/n$$

où la matrice x est l'observation initiale.

Nous considérons tout d'abord le cas d'une seule séquence $x = (x_1, \dots, x_m)'$. Afin d'obtenir une formule plus explicite de g nous introduisons les innovations partielles progressives et rétrogrades

$$\varepsilon(t; r) = \sum_{j=0}^r b(r, j) x_{t-j} \quad (r < t \leq m; r = 0, \dots, m-1),$$

$$\varepsilon^*(t; r) = \sum_{j=0}^r b(r, j) x_{t+j} \quad (0 < t \leq m-r; r = 0, \dots, m-1),$$

dans lesquelles les coefficients $b(r, j)$ sont ceux des polynômes de Szegő

$$\phi_r(z) = \sum_{j=0}^r b(r, j) z^{r-j} \quad (r = 0, \dots, m-1)$$

construits, en fonction de $\beta_1, \dots, \beta_{m-1}$, par la récurrence

$$\phi_0(z) = \phi_0^*(z) = 1$$

$$\phi_r(z) = z \phi_{r-1}(z) - \beta_r \phi_{r-1}^*(z), \quad \phi_r^*(z) = \phi_{r-1}^*(z) - \beta_r z \phi_{r-1}(z). \quad (2)$$

Nous utilisons aussi les notations

$$\sigma_0^2 = 1, \quad \sigma_r^2 = \prod_{j=1}^r (1 - \beta_j^2) \quad (1 \leq r < m),$$

$$\eta_r(\cdot) = \varepsilon^2(\cdot; r) / \sigma_r^2, \quad \eta_r^*(\cdot) = \varepsilon^{*2}(\cdot; r) / \sigma_r^2, \quad (0 \leq r < m).$$

II.1.1. Lemme.—(Dégerine, 1987) *Soit Γ un élément de C^+ . Alors la forme quadratique associée à Γ^{-1} peut être représentée, pour tout t dans $\{0, \dots, m-1\}$, par*

$$(i) \quad x\Gamma^{-1}x = \sum_{r=0}^{m-t-1} \eta_r(t+r-1) + \sum_{r=m-t}^{m-1} \eta_r^*(m-r),$$

$$(ii) \quad x\Gamma^{-1}x = \sum_{r=0}^t \eta_r^*(t-r+1) + \sum_{r=t+1}^{m-1} \eta_r(r+1).$$
(3)

Démonstration.—Ici X désignera un vecteur aléatoire gaussien, à valeurs dans \mathbb{R}^m , centré et de matrice de covariance Γ . Sa densité de probabilité est

$$f_X(x) = (2\pi)^{-m/2} |\Gamma|^{-1/2} \exp - (x\Gamma^{-1}x)/2.$$

La transformation

$$y_{r+1} = \varepsilon(t+r+1; r) \quad (r = 0, \dots, m-t-1)$$

$$y_{r+1} = \varepsilon^*(m-r; r) \quad (r = m-t, \dots, m-1),$$

définit un nouveau vecteur aléatoire Y dont les composantes Y_1, \dots, Y_m sont des

variables gaussiennes centrées de variances respectives $\sigma_0^2, \dots, \sigma_{m-1}^2$. La densité de probabilité de Y est

$$f_Y(y) = (2\pi)^{-m/2} \left(\prod_{r=0}^{m-1} \sigma_r^2 \right)^{-1/2} \exp - \left(\sum_{r=0}^{m-1} y_{r+1}^2 / \sigma_r^2 \right).$$

Puisque le jacobien de la transformation invoquée est égal à 1 et que

$$|\Gamma| = \prod_{r=0}^{m-1} \sigma_r^2,$$

l'égalité (i) de (3) est démontrée. La relation (ii) s'obtient de façon analogue. ♦

Pour tout k , $0 \leq k < m$, soit C_k^+ le sous-ensemble de C^+ associé aux modèles $AR(j)$, $j \leq k$. Chaque élément de C_k^+ est caractérisé par des autocorrélations partielles vérifiant $\beta_j = 0$, $k < j < m$. Ainsi si Γ est dans C^+ , la matrice Γ_k désigne l'élément de C_k^+ définie par les valeurs β_1, \dots, β_k de Γ . Notons aussi

$$g_k(\beta_1, \dots, \beta_k; x) = |\Gamma_k|^{1/m} x \Gamma_k^{-1} x.$$

II.1.2. Lemme.—(Dégerine, 1987) *Pour Γ dans C^+ , les fonctions g_k , $1 \leq k < m$, vérifient les inégalités suivantes*

$$g_k(\beta_1, \dots, \beta_k; x) \geq \prod_{j=1}^k (1 - \beta_j^2)^{-j/m} \sum_{j=k}^{m-1} \varepsilon^2(j+1; k), \quad (4)$$

$$g_k(\beta_1, \dots, \beta_k; x) \geq (1 - \beta_k^2)^{(m-k)/m} g_{k-1}(\beta_1, \dots, \beta_{k-1}; x) / m. \quad (5)$$

Démonstration.—La relation (4) provient de l'inégalité (i) de (3) avec $t=0$ et $\Gamma = \Gamma_k$, compte tenu de l'égalité

$$|\Gamma_k|^{1/m} / \sigma_k^2 = \prod_{j=1}^k (1 - \beta_j^2)^{-j/m}.$$

A partir de (3) on obtient

$$x \Gamma^{-1} x = \frac{1}{2} \sum_{r=0}^{m-1} [\eta_r(r+1) + \eta_r^*(m-r)],$$

$$x \Gamma_k^{-1} x = \frac{1}{2m} \sum_{r=0}^{m-1} \left(\sum_{j=r}^{m-1} [\eta_r(j+1) + \eta_r^*(j-r+1)] + r [\eta_r(r+1) + \eta_r^*(m-r)] \right). \quad (6)$$

Cette dernière expression permet de minorer les formes quadratiques Γ_k^{-1} , $0 < k < m$:

$$x \Gamma_k^{-1} x \geq \frac{1}{2m} \sum_{r=0}^{k-1} \sum_{j=r}^{m-1} [\eta_r(j+1) + \eta_r^*(j-r+1)]$$

$$\geq \frac{1}{2m} \sum_{r=0}^{k-2} [\eta_r(r+1) + \eta_r^*(m-r)] + \frac{1}{2m} \sum_{j=k-1}^{m-1} [\eta_{k-1}(j+1) + \eta_{k-1}^*(m-j)].$$

On reconnaît ci-dessus la représentation (6) de Γ_{k-1}^{-1} et on a donc obtenu l'inégalité

$$x^T_k^{-1} x \geq (x^T_{k-1}^{-1} x)/m$$

de laquelle (5) découle immédiatement. ♦

Soit $d(x)$ le plus petit entier k , $k < m$, pour lequel x satisfait l'équation aux différences

$$\sum_{j=0}^k b(j)x_{t-j} = 0 \quad (t = k+1, \dots, m) \quad (7)$$

où le polynôme

$$\phi_k(z) = \sum_{j=0}^k b(j)z^{k-j} \quad (8)$$

est défini par la récurrence (2) dans laquelle les coefficients satisfont $|\beta_j| \leq 1$ pour $j < k$ et $|\beta_k| = 1$. Lorsqu'un tel k n'existe pas on pose $d(x) = m$.

II.1.3. Proposition.—(Dégerine, 1987) *Pour $k < d(x)$, la limite de $g_k(\beta_1, \dots, \beta_k; x)$ au bord de son domaine est infinie et ainsi le minimum*

$$\inf_{\Gamma \in C^+} |\Gamma|^{1/m} x^T^{-1} x$$

est atteint dans C_k^+ .

Démonstration.—Puisque $d(x) = 0$ équivaut à $x = 0$ on peut supposer $d(x) \geq 1$. D'après (4) une valeur frontière de $g_1(\beta_1; x)$ ne peut être finie que si, pour la valeur correspondante $\hat{\beta}_1$, on a

$$\sum_{j=1}^{m-1} \hat{\varepsilon}^2(j+1; 1) = \sum_{j=1}^{m-1} (x_{j+1} - \hat{\beta}_1 x_j)^2 = 0,$$

et donc $d(x) = 1$. Supposons maintenant $d(x) \geq k > 1$ et que les valeurs frontières de g_{k-1} sont infinies. D'après (5) une valeur frontière de g_k ne peut être finie que si $|\beta_k| = 1$ et alors de (4) nous obtenons

$$0 = \hat{\varepsilon}(t; k) = \sum_{j=0}^k \hat{b}(k, j) x_{t-j} \quad (t = k+1, \dots, m).$$

Par conséquent x satisfait (7) et $d(x) = k$. ♦

Pour une séquence $x = (x_1, \dots, x_m)'$ donnée, cette proposition indique qu'il existe une estimation de maximum de vraisemblance de Λ , sous l'hypothèse d'un modèle AR(k), lorsque k est inférieur à $d(x)$. Cependant on peut remarquer que, dès que m est supérieur à deux, $d(x)$ est plus petit que m avec une probabilité positive. L'équation aux différences (7) forme un système de $m-k$ équations dans lequel les inconnues $b(j)$, $j = 0, \dots, k$, satisfont $b(j) = \beta_k b(k-j)$ avec $|\beta_k| = 1$. On peut donc énoncer :

II.1.4. Corollaire.—(Dégerine, 1987) *Soit x_1, \dots, x_m une séquence gaussienne d'un modèle AR(k). Pour $m > 3k/2$ l'estimation de maximum de vraisemblance existe presque sûrement.*

Considérons maintenant le cas d'un échantillon; X désigne alors la matrice de covariance empirique ou sa projection sur \mathcal{E} . Soit B l'ensemble des vecteurs $b = (b(0), \dots, b(m-1))'$ de \mathbb{R}^m tels que, comme dans (8), le polynôme ϕ_{m-1} est défini par (2) avec des coefficients satisfaisant $|\beta_j| \leq 1$ pour $j < m-1$ et $|\beta_{m-1}| = 1$. $\ker(X)$ désigne le noyau de X et nous utilisons la norme $\|\Lambda\| = [\text{tr}(\Lambda^2)]^{1/2}$.

II.1.5. Proposition.—(Dégerine, 1987) *Sous la condition $\ker(X) \cap B = \emptyset$ on a $\lim L(\Lambda; X) = +\infty$ lorsque $\Lambda \rightarrow \Lambda_0$, $\Lambda_0 \in \overline{\mathcal{C}}^+ \setminus \mathcal{C}^+$, ou lorsque $\|\Lambda\| \rightarrow +\infty$.*

Démonstration.—On peut utiliser les résultats du chapitre précédent mais aussi montrer de manière récursive, comme dans le cas d'une seule séquence, que les valeurs frontières de

$$|\Pi|^{-1/m} \text{tr}(\Gamma^{-1} X) = \left(\sum_{j=1}^n |\Pi|^{-1/m} x'_j \Gamma^{-1} x_j \right) / n,$$

où x_j , $j = 1, \dots, n$, sont les colonnes de la matrice observée x , sont infinies lorsqu'il n'y a pas d'élément b dans B vérifiant $b'x_j = 0$, $j = 1, \dots, n$. Cette condition est

précisément $\ker(X) \cap B = \emptyset$. La première partie de la proposition est démontrée puisque l'on a vu, dans la preuve de (1), que

$$L(\Lambda; X) \geq m \ln [|\Gamma|^{1/m} \text{tr}(\Gamma^{-1}X)/m] + m.$$

Soit $\hat{\Gamma}$ un point d'accumulation d'une suite $\{\Gamma_k = \Lambda_k / \sigma_m^2; k \in \mathbb{N}\}$ telle que la norme $\|\Lambda_k\|$ tend vers $+\infty$. Soit $\hat{\Gamma}$ est non singulier et la démonstration de $\lim L(\Lambda; X) = +\infty$ est directe, soit $\hat{\Gamma}$ est singulier et les arguments ci-dessus conduisent au résultat. ♦

De la même façon que l'on a obtenu le Corollaire II.1.4 on obtient ici:

II.1.6. Corollaire.—(Dégerine, 1987) *Soit x_1, \dots, x_n n observations indépendantes d'une séquence de longueur m d'un processus gaussien stationnaire centré de matrice d'autocovariance Λ . Pour $n \geq m/2$ l'estimation de maximum de vraisemblance de Λ existe presque sûrement.*

En fait remplacer la matrice de covariance empirique X par sa projection sur \mathcal{C} "équivalent" à doubler la taille de l'échantillon, pour $2n \geq m$ la projection est non singulière et la vraisemblance tend vers zéro au bord de son domaine.

§ II.2. Décompositions spectrales des matrices de Tœplitz

La décomposition spectrale, en termes de valeurs propres et vecteurs propres, d'une matrice de \mathcal{C}^+ , ne laisse pas transparaître, de façon claire, le caractère Tœplitz de la matrice. La décomposition que nous proposons est par contre directement issue de cette propriété. M^* désigne ici la matrice transposée conjuguée de M .

II.2.1. Définition.—(Dégerine, 1983) *On appelle décomposition spectrale d'une matrice de Tœplitz Λ , symétrique positive d'ordre m , toute écriture de Λ sous la forme MDM^* , où D est une matrice diagonale à éléments positifs ou nuls et où M est une matrice carrée d'ordre m dont la k -ième colonne, $k = 1, \dots, m$ est de la forme $(1, e^{i\lambda_k}, \dots, e^{i(m-1)\lambda_k})'$, les fréquences $\lambda_1, \dots, \lambda_m$ vérifiant $-\pi < \lambda_1 < \dots < \lambda_m \leq \pi$.*

Soit F la mesure spectrale associée au couple (M, D) par $F(\{\lambda_k\}) = d_k$, $k = 1, \dots, m$, où d_k sont les éléments diagonaux de D . La relation $\Lambda = MDM^*$ peut se formuler par

$$\Lambda(1, 1+k) = \int_{-\pi}^{+\pi} e^{ik\lambda} F(d\lambda), \quad k = 0, \dots, m-1.$$

Lorsque Λ est non singulière, il est facile d'établir qu'elle admet exactement deux décompositions spectrales distinctes ${}_+M{}_+D{}_+M^*$ et ${}_-_M{}_-_D{}_-_M^*$. Les mesures spectrales correspondantes ${}_+F$ et ${}_-_F$ sont celles des processus linéairement singuliers d'ordre m définis par $\{\sigma^2; \beta_1, \dots, \beta_{m-1}\}$, qui caractérise Λ , et le choix de ± 1 pour β_m . Par la suite M, D, F , etc., désigneront indifféremment les éléments de l'une de ces deux décompositions. Dans le cas où Λ est de rang k , $k < m$, elle constitue la matrice d'autocovariance d'ordre m d'un processus singulier d'ordre k et seules les fréquences et les masses correspondantes de la mesure spectrale associée sont imposées dans MDM^* ; les autres fréquences sont quelconques et affectées de masses nulles. Il existe ainsi une infinité de décompositions spectrales. L'unicité peut être obtenue en imposant que les éléments diagonaux de D soient strictement positifs et M sera alors rectangulaire d'ordre $m \times k$.

La décomposition MDM^* n'est évidemment pas la décomposition spectrale classique de Λ en termes de valeurs propres et vecteurs propres. Cependant les éléments diagonaux d_k de D sont les valeurs propres de l'opérateur, sur l'espace des fonctions presque-périodiques, défini par la covariance du processus singulier sous-jacent. Les fonctions $\{e^{i\lambda_k}, \lambda \in \mathbb{Z}\}$ associées à M sont les fonctions propres correspondantes. D'autre part en posant $\tilde{\Gamma} = MM^*/m$ on a les relations classiques

$$M^*\tilde{\Gamma}^{-1}M/m = I_m, \quad \Lambda\tilde{\Gamma}^{-1}M/m = MD,$$

qui signifient que les colonnes de M sont les vecteurs propres normalisés associés aux valeurs propres constituées par les éléments diagonaux de D pour la matrice Λ et relativement au produit scalaire $\tilde{\Gamma}^{-1}/m$. Nous verrons que ce produit scalaire joue un

rôle important dans la mise en œuvre de la méthode de relaxation décrite au § II.3.

Lorsque la matrice $\Lambda = \sigma^2 \Gamma$ est non singulière, le couple $(+M, -M)$ caractérise la matrice d'autocorrélation Γ . En effet Γ est définie par les autocorrélations partielles $\beta_1, \dots, \beta_{m-1}$, ou, de façon équivalente, par le polynôme ϕ_{m-1} . Celui-ci s'écrit

$$\phi_{m-1}(z) = [{}_+\phi_m(z) + {}_-\phi_m(z)]/2z, \quad (9)$$

où ${}_+\phi_m$ et ${}_-\phi_m$ sont ceux correspondant à $\beta_m = \pm 1$. Les fréquences $+\lambda$ et $-\lambda$ associées à $(+M, -M)$ sont données par les racines de ces polynômes, elles sont distinctes et alternent sur le cercle unité.

II.2.2. Proposition.—(Dégerine, 1983) *L'ensemble des matrices d'autocorrélation non singulières Γ d'ordre m est en correspondance biunivoque avec l'ensemble des $(m-1)$ -uplets $(\theta_1, \dots, \theta_{m-1})$ vérifiant $0 < \theta_1 < \dots < \theta_{m-1} < \pi$. Cette correspondance est assurée par l'intermédiaire du polynôme ϕ_{m-1} défini par (9) et par les relations*

$${}_+\phi_m(z) = \prod_{k=0}^{m-1} (z - e^{i\theta_{2k}}), \quad {}_-\phi_m(z) = \prod_{k=0}^{m-1} (z - e^{i\theta_{2k+1}}),$$

dans lesquelles on a posé : $\theta_0 = 0, \theta_m = \pi, \theta_{m+k} = -\theta_{m-k}, k = 1, \dots, m-1$.

Démonstration.—Le polynôme ϕ_{m-1} se caractérise comme un polynôme réel de la forme

$$\phi_{m-1}(z) = \sum_{k=0}^{m-1} b(m-1, k) z^{m-1-k}, \quad b(m-1, 0) = 1,$$

dont les racines sont à l'intérieur du cercle unité (polynôme stable). D'autre part un polynôme réel $D(z)$ est stable si et seulement si dans la décomposition

$$D(z) = \sum_{k=0}^m d_k z^k = {}_+D(z) + {}_-D(z), \quad {}_+D(z) = \frac{1}{2} \sum_{k=0}^m (d_k + d_{m-k}) z^k, \quad {}_-D(z) = \frac{1}{2} \sum_{k=0}^m (d_k - d_{m-k}) z^k,$$

les racines des polynômes ${}_+D$ et ${}_D$ sont simples, sur le cercle unité et se présentent alternativement sur $[0, 2\pi[$ (cf. Schussler, 1976). En posant $D(z) = z\phi_{m-1}(z)$ on a

$${}_+D(z) = {}_+\phi_m(z)/2, \quad {}_-D(z) = {}_-\phi_m(z)/2.$$

De plus 1 est racine de ${}_+\phi_m$ et -1 est racine de ${}_+\phi_m$ ou de ${}_-\phi_m$ selon que m est pair ou impair. Ainsi si ϕ_{m-1} est associé à une matrice Γ , $z\phi_{m-1}(z)$ est stable et satisfait les

relations de la proposition. Réciproquement, compte tenu des valeurs θ_k , $k = 0, \dots, 2m-1$, ces relations définissent un polynôme stable de la forme

$$z\phi_{m-1}(z) = \sum_{k=0}^{m-1} b(m-1,k)z^{m-k}, \quad b(m-1,0) = 1.$$

ϕ_{m-1} est donc stable et de la forme requise. ♦

Soit $Y = (Y_0, \dots, Y_{m-1})'$ un vecteur aléatoire centré de matrice de covariance $\Lambda(Y)$. On notera $M(Y)$, $D(Y)$, etc., les éléments correspondants. Si M et D sont ceux d'une décomposition spectrale d'une matrice Λ , la représentation de Y sous la forme $Y=MDZ$ équivaut à poser

$$Y_t = \sum_{k=1}^m Z_k d_k e^{it\lambda_k}, \quad t = 0, \dots, m-1,$$

où, pour $k = 1, \dots, m$, Z_k est une variable aléatoire complexe reliée uniquement à la phase et à l'amplitude correspondant à la fréquence λ_k .

II.2.3. Proposition.—*Parmi les représentations de Y sous la forme :*

$$Y = MDZ, \quad \Lambda = MDM^*, \quad |\Lambda| = 1,$$

celles définies par $M = M(Y)$, $D = |\Lambda(Y)|^{-1/m} D(Y)$, sont les seules réalisant le minimum de $E(ZZ^)/m$ alors égal à*

$$\min_{\Lambda} \frac{1}{m} E(ZZ^*) = |\Lambda(Y)|^{1/m} = \sigma^2(Y) \prod_{k=1}^{m-1} [1 - \beta_k^2(Y)]^{(m-k)/m}.$$

Démonstration.—De $Z = D^{-1}M^{-1}Y$ on obtient $E(ZZ^*) = \text{tr}\{\Lambda^{-1}\Lambda(Y)\}$. L'égalité

$$\text{Min}\{\text{tr}[Q\Lambda(Y)], Q \in \mathcal{S}^+, |Q| = 1\} = \text{Min}\{\Lambda^{-1}\Lambda(Y), \Lambda = MDM^*, |\Lambda| = 1\}$$

provient de ce que la solution du premier problème, donnée par le Lemme I.4.1, satisfait les contraintes du second bien qu'elles soient plus restrictives. ♦

En se restreignant, dans la proposition ci-dessus, aux représentations de Y données par les éléments issus de matrices Λ associées à des modèles AR(j), $j \leq k < m$, la solution optimale correspond au modèle AR sous-jacent à $\Lambda(Y)$ (cf. Proposition

A.III.3.2). Notons aussi que les matrices $M(Y)$ des décompositions spectrales de $\Lambda(Y)$ se caractérisent comme les seules matrices du type M pour lesquelles la matrice de covariance de $M^{-1}Y$ est diagonale.

Pour une matrice M du type ${}_+M$ ou ${}_-_M$ on note $H(M)$ le sous-espace vectoriel de \mathcal{C} constitué par les matrices de la forme MDM^* . La matrice D est réelle diagonale et satisfait les contraintes de symétrie garantissant le caractère réel de la matrice MDM^* . Il s'en suit que les dimensions des sous-espaces $H({}_+M)$ et $H({}_-_M)$ sont respectivement $[m/2]+1$ et $m-[m/2]$ où $[x]$ désigne la partie entière de x .

II.2.3. Lemme.-(Dégerine, 1987) Soit $({}_+M, {}_-M)$ le couple caractéristique d'une matrice d'autocorrélation non singulière Γ . Les sous-espaces $H({}_+M)$ et $H({}_-_M)$ constituent une décomposition de \mathcal{C} , $\mathcal{C} = H({}_+M) + H({}_-_M)$, satisfaisant

$$H({}_+M) \cap H({}_-_M) = \{ \lambda \Gamma, \lambda \in \mathbb{R} \}.$$

Démonstration.— Compte tenu des dimensions des sous-espaces, il suffit de montrer que leur intersection n'est constituée que des matrices colinéaires à Γ . Il est clair que cette intersection contient les matrices $\lambda \Gamma$, $\lambda \in \mathbb{R}$, puisque l'on a :

$$\Gamma = {}_+M {}_+\Delta {}_+M^* = {}_-M {}_-\Delta {}_-M^*,$$

où les éléments diagonaux des matrices ${}_+\Delta$ et ${}_- \Delta$ sont strictement positifs. Supposons alors :

$$\Lambda = {}_+M {}_+D {}_+M^* = {}_-M {}_-\Delta {}_-M^*.$$

Si $\Lambda \in \mathcal{C}^+$, (ou $-\Lambda \in \mathcal{C}^+$) la matrice d'autocorrélation Λ/σ^2 (ou $-\Lambda/\sigma^2$) est paramétrée par $({}_+M, {}_-M)$ et donc égale à Γ . Sinon il existe $\rho > 0$ tel que $(\Lambda + \rho \Gamma) \in \mathcal{C}^+$ et par suite

$${}_+D + \rho {}_+\Delta = \lambda {}_+\Delta, \quad {}_-D + \rho {}_-\Delta = \lambda {}_-\Delta, \quad {}_+D = (\lambda - \rho) {}_+\Delta, \quad {}_-D = (\lambda - \rho) {}_-\Delta,$$

d'où $\Lambda = (\lambda - \rho)\Gamma$ et $\Lambda = 0$ (i.e. $\lambda = \rho$) car ni Λ ni son opposé n'appartiennent à \mathcal{C}^+ . ♦

§ II.3. Une méthode de relaxation

X désignant la matrice de covariance empirique, ou sa projection sur \mathcal{E} , on suppose que la condition $\ker(X) \cap B = \emptyset$ de la Proposition II.1.5 est satisfaite. Ainsi la méthode du maximum de vraisemblance (sous hypothèse gaussienne), qui équivaut à minimiser, par rapport à Λ dans \mathcal{E}^+ ,

$$L(\Lambda; X) = \ln |\Lambda| + \text{tr}(\Lambda^{-1}X),$$

admet une solution satisfaisant l'équation de vraisemblance

$$\text{grad } L(\Lambda; X) = 0. \quad (10)$$

Sans l'hypothèse gaussienne cette solution peut se caractériser comme la valeur $\hat{\Lambda} = \hat{\alpha} \hat{M} \hat{D} \hat{M}^*$ où \hat{M} et \hat{D} réalisent le minimum, alors égal à $\hat{\alpha}$, du problème de minimisation sous contrainte (moindres carrés)

$$\text{Min } z^*z/nm, \quad x = MDz, \quad MDM^* = |\Gamma|^{1/m} \Gamma, \quad \Gamma \in C^+$$

où la matrice x est l'observation initiale.

Pour une matrice carrée A , $\text{diag}(A)$ désigne la matrice diagonale définie par les éléments diagonaux de A .

II.3.1. Proposition.—(Dégerine, 1987) *Pour M fixé, l'estimateur de maximum de vraisemblance de $\Lambda = MDM^*$ est la matrice $\hat{\Lambda} = M\hat{D}M^*$ où $\hat{D} = \text{diag}(M^{-1}XM^{*-1})$.*

Démonstration. $L(\Lambda; X)$ est de la forme :

$$L(\Lambda; X) = \ln |MM^*| + \ln |D| + \text{tr}(D^{-1}\hat{D}).$$

Son minimum, par rapport à D , est atteint en \hat{D} lorsque les éléments diagonaux de \hat{D} sont strictement positifs. Cette condition, qui correspond à $\ker(X) \cap B = \emptyset$, est réalisée presque sûrement, même si $n = 1$. En effet $\ker(X) \cap B \neq \emptyset$ signifie que les colonnes de X (ou de x) sont des combinaisons linéaires d'au plus $(m-1)$ colonnes de M . ♦

Notons que la dérivée seconde, en $\hat{\Lambda}$, de $L(\cdot; X)$ restreinte à $H(M)$ est égale à

$$(dA, dB) \in H(M)^2 \longrightarrow \text{tr}(\hat{\Lambda}^{-1}dA\hat{\Lambda}^{-1}dB)/m = \langle dA, dB \rangle_{\hat{\Lambda}}.$$

Cette proposition et le Lemme II.2.3 montrent qu'une solution de l'équation de vraisemblance (10) se caractérise ainsi :

II.3.2. Corollaire.—(Dégerine, 1987) *La matrice $\Lambda = {}_+M_+D_+M^* = {}_M_-D_-M^*$ est solution de l'équation de vraisemblance $\text{grad } L(\Lambda; X) = 0$ si et seulement si ${}_+D$ et ${}_D_-$ satisfont :*

$${}_+D = \text{diag}({}_+M^{-1}X_+M^{*-1}), \quad {}_D_- = \text{diag}({}_M^{-1}X_-M^{*-1}).$$

En un tel Λ la dérivée seconde ci-dessus montre que la restriction de $L(\cdot; X)$ à $H({}_+M)$ ou à $H({}_M_-)$ est minimum.

Pour toute matrice $\Lambda = {}_+M_+D_+M^* = {}_M_-D_-M^*$ dans \mathcal{C}^+ on note ${}_+\hat{\Lambda}$ et ${}_{-}\hat{\Lambda}$ les matrices associées à $(\Lambda; X)$ par la proposition ci-dessus :

$${}_+\hat{\Lambda} = {}_+M_+{}_{+}\hat{D}_+M^*, \quad {}_{-}\hat{\Lambda} = {}_M_-{}_{-}\hat{D}_-M^*.$$

II.3.3. Proposition.—(Dégerine, 1987) *Soit Λ_0 quelconque dans \mathcal{C}^+ . Alors le procédé itératif suivant :*

- (i) $\Lambda_1 = {}_+\hat{\Lambda}_0$, ou $\Lambda_1 = {}_{-}\hat{\Lambda}_0$,
- (ii) $\Lambda_{k+1} = {}_+\hat{\Lambda}_k$ si $\Lambda_k = {}_{-}\hat{\Lambda}_k$, sinon $\Lambda_{k+1} = {}_{-}\hat{\Lambda}_k$, $k \geq 1$,

construit une suite $\{\Lambda_k; k \in \mathbb{N}\}$ telle que soit

$$L(\Lambda_{k+1}; X) < L(\Lambda_k; X), \quad k \in \mathbb{N}, \quad (11)$$

soit $\Lambda_{k+1} = \Lambda_k$ dès que Λ_k est solution de l'équation de vraisemblance.

Démonstration.—Le corollaire précédent montre que ${}_+\hat{\Lambda}$ (ou ${}_{-}\hat{\Lambda}$) est solution de (10) si et seulement si ${}_+\hat{\Lambda} = {}_{-}\hat{\Lambda}$. D'autre part l'unicité de la solution dans la Proposition II.3.1 prouve que, pour tout Λ dans \mathcal{C}^+ , on a

$$L({}_+\hat{\Lambda}; X) \leq L(\Lambda; X), \quad L({}_{-}\hat{\Lambda}; X) \leq L(\Lambda; X),$$

avec égalité si et seulement si ${}_+\hat{\Lambda} = \Lambda$ ou ${}_{-}\hat{\Lambda} = \Lambda$. Partant d'un Λ_0 quelconque dans \mathcal{C}^+ , (i) donne $\Lambda_1 \neq \Lambda_0$ avec $L(\Lambda_1; X) < L(\Lambda_0; X)$ sinon $\Lambda_1 = \Lambda_0 = {}_+\hat{\Lambda}_0 = {}_{-}\hat{\Lambda}_0$ et est solution de (10). A l'étape k , par construction, on a $\Lambda_k = {}_{-}\hat{\Lambda}_k$ ou $\Lambda_k = {}_+\hat{\Lambda}_k$; ainsi Λ_{k+1} est défini par (ii) et satisfait (11) sinon $\Lambda_{k+1} = \Lambda_k = {}_+\hat{\Lambda}_k = {}_{-}\hat{\Lambda}_k$ et est solution de (10). ♦

En dehors du cas $m = 2$ où la solution, lorsqu'elle existe, est obtenue dès la première étape (il existe une seule matrice $+M$), pour tout choix de Λ_0 dans \mathcal{E}^+ , ce procédé construit presque sûrement une suite infinie satisfaisant (11). On étudie ci-dessous les propriétés de convergence d'une telle suite $\{\Lambda_k; k \in \mathbb{N}\}$. Les notations sont celles introduites au chapitre précédent : $\|\cdot\|_\Lambda$ est la norme associée au produit scalaire $\langle A, B \rangle_\Lambda = \text{tr}(\Lambda^{-1}A\Lambda^{-1}B)/m$ et l'indice Λ est omis lorsqu'il est égal à I .

La suite $\{\alpha_k = |\Lambda_k|; k \in \mathbb{N}\}$ est strictement décroissante puisque $L(\Lambda_k; X)$ est égal à $\ln \alpha_k + m$. Sa limite α est supposée strictement positive ($\ker(X) \cap B = \emptyset$) et la suite satisfait :

$$\alpha_k = \text{tr}(\bar{\Lambda}_k^{-1}X)/m = \langle \Lambda_k, X \rangle_{\Lambda_k} / \alpha_k, \quad \bar{\Lambda}_k = \Lambda_k / \alpha_k.$$

II.3.4. Lemme.—(Dégerine, 1987) *Avec les notations ci dessus, la suite $\{\Lambda_k; k \in \mathbb{N}\}$ est telle que :*

$$\|\Lambda_{k+1} - \Lambda_k\|_{\Lambda_k} \sim 2\alpha(\alpha_k - \alpha_{k+1}), \quad \lim \|\Lambda_{k+1} - \Lambda_k\| = 0.$$

Démonstration.—On peut supposer $\Lambda_k = +\hat{\Lambda}_k$ et donc $\Lambda_{k+1} = -\hat{\Lambda}_k$ car la démonstration serait analogue dans la situation alternative. On peut voir que , par rapport au produit scalaire associé à $\bar{\Lambda}_k$, Λ_k et Λ_{k+1} sont les projections orthogonales de X sur respectivement $H_k = H(+M_k)$ et $H_{k+1} = H(-M_k)$. Alors, puisque Λ_k est dans l'intersection $H_k \cap H_{k+1}$, on a :

$$\|\Lambda_{k+1} - \Lambda_k\|_{\bar{\Lambda}_k}^2 = \|\Lambda_{k+1}\|_{\bar{\Lambda}_k}^2 - \|\Lambda_k\|_{\bar{\Lambda}_k}^2 = \|\Lambda_{k+1}\|_{\bar{\Lambda}_k}^2 - \alpha_k^2.$$

Maintenant, avec $M = -M_k$ et posant $\bar{\Lambda}_k = MDM^*$, $\bar{\Lambda}_{k+1} = M\hat{D}M^*$, on obtient :

$$\|\Lambda_{k+1}\|_{\bar{\Lambda}_k}^2 = \alpha_{k+1}^2 \text{tr}\{(D^{-1}\hat{D})^2\}/m, \quad \alpha_k^2 = \langle \Lambda_{k+1}, \Lambda_k \rangle_{\bar{\Lambda}_k} = \alpha_{k+1}\alpha_k \text{tr}(D^{-1}\hat{D})/m.$$

Ainsi

$$\begin{aligned} \|\Lambda_{k+1} - \Lambda_k\|_{\bar{\Lambda}_k}^2 &= \alpha_{k+1}^2 \{ \text{tr}[(D^{-1}\hat{D})^2]/m - [\text{tr}(D^{-1}\hat{D})/m]^2 \} \\ &= \alpha_{k+1}^2 \left\{ \left(\sum_{j=1}^m \varepsilon_j \right)^2 / m - \left[\left(\sum_{j=1}^m \varepsilon_j \right) / m \right]^2 \right\}, \end{aligned} \quad (12)$$

où $1+\varepsilon_k, k = 1, \dots, m$ sont les éléments diagonaux de $D^{-1}\hat{D}$. La contrainte

$$1 = |\bar{\Lambda}_k^{-1} \bar{\Lambda}_{k+1}| = |D^{-1}\hat{D}| = \prod_{j=1}^m (1+\varepsilon_j),$$

et le fait que la limite de

$$\alpha_k / \alpha_{k+1} = [\sum_{j=1}^m (1+\varepsilon_j)]/m$$

soit égale à 1 impliquent que chaque ε_j tend vers zéro. Dans ce cas la contrainte ci-dessus conduit à l'équivalence $\sum \varepsilon_j^2 \sim 2\sum \varepsilon_j$. En négligeant le second terme dans (12) nous obtenons l'équivalence de l'énoncé. D'après la Proposition II.1.5 la suite $\{\Lambda_k; k \in \mathbb{N}\}$ demeure dans un compact de \mathcal{C}^+ . C'est également le cas de $\{\bar{\Lambda}_k; k \in \mathbb{N}\}$ puisque α est strictement positif. Alors il existe deux constantes c et C telles que

$$c\|\cdot\|_{\bar{\Lambda}_k} \leq \|\cdot\| \leq C\|\cdot\|_{\bar{\Lambda}_k}, \quad k \in \mathbb{N}. \quad (13)$$

Cette équivalence et le résultat précédent donnent $\lim \|\Lambda_{k+1} - \Lambda_k\| = 0$. ♦

II.3.5. Lemme.—(Dégerine, 1987) *La suite des gradients de $L(\cdot; X)$, $\{\text{grad } L(\Lambda_k; X); k \in \mathbb{N}\}$, tend vers zéro et tout point d'accumulation de la suite $\{\Lambda_k; k \in \mathbb{N}\}$ est solution de l'équation de vraisemblance.*

Démonstration.—La dérivée de $L(\cdot; X)$ en Λ est

$$dA \longrightarrow \text{tr}(\Lambda^{-1}dA) - \text{tr}(\Lambda^{-1}X\Lambda^{-1}dA).$$

Aussi, dans la métrique donnée par le produit scalaire associé à Λ_k ,

$$\text{grad } L(\Lambda_k; X) = p(\Lambda_k - \hat{\Lambda}_k)$$

où $\hat{\Lambda}_k$ est la projection orthogonale de X sur \mathcal{C} . On peut considérer la situation présentée au début de la démonstration du lemme précédent. Avec les mêmes notations $\hat{\Lambda}_k$ et $\hat{\Lambda}_{k+1}$ restent les projections orthogonales de X sur H_k et sur H_{k+1} . D'où

$$\|\Lambda_k - \hat{\Lambda}_k\|_{\Lambda_k} = \|\Lambda_{k+1} - \Lambda_k\|_{\Lambda_k} / \sin \phi_k$$

où ϕ_k est l'angle entre $\hat{\Lambda}_k - \Lambda_k$ et $\hat{\Lambda}_k - \Lambda_{k+1}$ qui sont orthogonaux respectivement à H_k et à H_{k+1} . Rappelons que la suite $\{\Lambda_k; k \in \mathbb{N}\}$ demeure dans un compact de \mathcal{C}^+ . Alors d'une part il existe un angle ϕ tel que $\sin \phi_k \geq \sin \phi > 0$ et le lemme ci-dessus

conduit à

$$\lim \|\Lambda_k - \widehat{\Lambda}_k\|_{\Lambda_k} = 0.$$

D'autre part on a l'équivalence uniforme entre les normes analogue à (13) et ainsi

$$\lim \|\text{grad } L(\Lambda_k; X)\| = 0. \spadesuit$$

Les deux lemmes précédents et le Théorème 26.1 d'Ostrowski (1973) permettent d'énoncer :

II.2.6. Proposition.—(Dégerine, 1987) *Soit X satisfaisant $\ker(X) \cap B = \emptyset$. Alors toutes les suites $\{\Lambda_k; k \in \mathbb{N}\}$ obtenues par le procédé de la Proposition II.3.3 convergent sinon l'ensemble des solutions de l'équation de vraisemblance a la puissance du continu.*

En pratique le critère d'arrêt porte sur la proximité entre deux itérés successifs et ne peut différencier les deux situations de cette proposition.

§ II.4. Algorithme et simulations

Il est important de remarquer que la mise en œuvre de la méthode de relaxation présentée au paragraphe précédent ne nécessite pas le calcul des matrices M . Par contre on donne ici le moyen de calculer de manière exacte la "structure d'autocovariance" d'un prolongement presque-périodique d'une séquence.

Soit MDM^* une décomposition spectrale d'une matrice d'autocorrélation Γ non singulière d'ordre m . La matrice M s'identifie au polynôme

$$\phi_m(z) = \sum_{k=0}^m b(k)z^{m-k} = \prod_{j=1}^m (z - e^{i\lambda_j}) \quad (14)$$

où $\lambda_j, j = 1, \dots, m$ sont les fréquences qui composent M .

Soit $y = (y(1), \dots, y(m))'$ une séquence donnée. On appelle **prolongement presque-périodique** de y basé sur M la fonction presque-périodique (cf. § III.1)

$$y(t) = \sum_{j=1}^m \gamma_j e^{i t \lambda_j}, \quad t \in \mathbb{Z}, \quad \gamma = (\gamma_1, \dots, \gamma_m)' = M^{-1}y.$$

Le prolongement $y(\cdot)$ se calcule en fonction des valeurs initiales que constitue y à l'aide du polynôme ϕ_m selon la récurrence

$$\sum_{k=0}^m b(k)y(t-k) = 0, \quad t \in \mathbb{Z}. \quad (15)$$

La "matrice d'autocovariance" d'ordre m de $y(\cdot)$ est

$$\hat{\Lambda} = M\hat{D}M^*, \quad \hat{D} = \text{diag}(\gamma\gamma^*) = \text{diag}(M^{-1}yy'M^{*-1}).$$

Elle satisfait

$$\hat{\Lambda}(1, 1+k) = (y(1+k), \dots, y(m+k))\tilde{\Gamma}^{-1}y, \quad k = 0, \dots, m-1, \quad (16)$$

où $\tilde{\Gamma} = MM^*/m$. Les éléments de $\tilde{\Gamma}$ s'expriment comme la somme des puissances k èmes des racines du polynôme ϕ_m

$$\tilde{\Gamma}(1, 1+k) = \tilde{\rho}(k) = \left(\sum_{j=1}^m e^{i k \lambda_j} \right) / m, \quad k = 0, \dots, m-1.$$

Ils se calculent donc par récurrence en fonction des coefficients de ϕ_m

$$\tilde{\rho}(1) = -b(1)/m, \quad \tilde{\rho}(k) = -(k/m)b(k) - \sum_{j=1}^{k-1} b(j)\tilde{\rho}(k-j), \quad k = 2, \dots, m-1. \quad (17)$$

Nous sommes maintenant en mesure de décrire l'algorithme de passage de Λ_k à Λ_{k+1} . Si $\Lambda_k = MDM^*$, la matrice Λ_{k+1} est donnée par

$$\Lambda_{k+1} = M\hat{D}M^*, \quad \hat{D} = \text{diag}(M^{-1}XM^{*-1}).$$

Mais Λ_{k+1} est aussi la moyenne

$$\Lambda_{k+1} = \left(\sum_{j=1}^n M\hat{D}_jM^* \right) / n, \quad \hat{D}_j = \text{diag}(M^{-1}x_jx_j'M^{*-1}), \quad (18)$$

où $x_j, j=1, \dots, n$, sont les colonnes de l'observation initiale x .

Partant des valeurs $\rho(1), \dots, \rho(m-1)$ de Λ_k , l'algorithme de Levinson-Durbin fournit $\beta(1), \dots, \beta(m-1)$ puis $b(0), \dots, b(m)$ de (14) avec $\beta(m) = \pm 1$ selon que M dans Λ_k est du type $+M$ ou $-M$. Ensuite nous calculons $\tilde{\Gamma} = (MM^*)/m$ par la

réurrence (17) ainsi que la décomposition de Cholesky de $\tilde{\Gamma}^{-1}$, utilisant encore l'algorithme de Levinson-Durbin. Finalement, avec les prolongements $x_{m+k,j}$, $k = 0, \dots, m-1$, donnés par (15), chaque $M\hat{D}_jM^*$ dans (18) provient de (16).

Dans le cas où n est grand devant m , utilisant l'exhaustivité de $X = (xx')/n$, nous pouvons remplacer x dans (18) par toute matrice \tilde{x} d'ordre $m \times m$ satisfaisant $X = (\tilde{x}\tilde{x}')/m$.

Les essais sont effectués à partir d'un fichier de 30000 nombres pseudo-aléatoires, de loi normale centrée réduite, construit par la méthode de Box-Muller à l'aide du générateur

$$y(i+1) = \text{partie fractionnaire de } \{[y(i) + \pi]^5\}.$$

La méthode est initialisée avec la matrice d'autocovariance empirique Λ_0

$$\Lambda_0 = \frac{1}{nm} \sum_{j=1}^n \sum_{i=1}^{m-k} x_{i,j} x_{i+k,j}, \quad k = 0, \dots, m-1. \quad (19)$$

Le critère d'arrêt utilisé ici est, avec des notations évidentes, donné par

$$\max\{|\beta_{k+1}(j) - \beta_k(j)|, j = 1, \dots, m-1\} \leq \varepsilon. \quad (20)$$

Pour $m = 10$, $n = 15$ et la vraie valeur de Λ donnée dans la Table II, la méthode est appliquée, avec $\varepsilon = 10^{-3}$, sur 100 répétitions. Dans chaque cas nous calculons les éléments de la matrice empirique Λ_0 et ceux de la dernière valeur Λ_K . Le nombre d'itérations K ainsi que les premières valeurs $L(\Lambda_k; X)$, $k = 0, \dots, 10$, sont également retenus. La Table I donne la valeur moyenne et l'écart-type de K ainsi que des pourcentages suivants :

$$G(k) = 100 \times [L(\Lambda_k; X) - L(\Lambda_0; X)] / [L(\Lambda_K; X) - L(\Lambda_0; X)], \quad k = 1, \dots, 10.$$

Nous remarquons que la minimisation de $L(\cdot; X)$ est réalisée dès les premières itérations (cf. variable $G(\cdot)$) alors que la stabilisation de Λ_k est plus longue à obtenir (cf. variable K). Cela peut être dû à la méthode : à chaque itération la minimisation est réalisée dans un demi-espace dont la "direction" n'a aucun caractère optimal. Il est

vraisemblable aussi que cela provienne de l'aplatissement de la fonction $L(.;X)$ au voisinage de son minimum : il faudrait observer les variations de la valeur finale Λ_K entre deux valeurs de ϵ dans le critère d'arrêt (20). Dans ce cas on peut envisager d'utiliser $\epsilon = 10^{-2}$ dans (20) sans détériorer les qualités de l'estimateur mais il est aussi possible que ceci introduise un biais dû au choix de la valeur initiale Λ_0 . Nous avons également utilisé un critère d'arrêt portant sur la fonction $L(.;X)$ mais cela conduit parfois à des résultats erronés (cf. Dégerine, 1986) car il est des situations où la fonction $L(.;X)$ est très irrégulière. Lors de l'étude des minima locaux de $L(.;X)$ (cf. § 5), nous avons utilisé un critère d'arrêt quadratique analogue à (20) et nous avons constaté que la méthode pouvait s'arrêter en des valeurs Λ_K où le gradient était très nettement différent de zéro, y compris pour $\epsilon = 10^{-4}$. Nous suggérons donc d'adjoindre un critère mesurant la nullité de la norme du gradient à celui portant sur $\beta(.)$.

TABLE I
VITESSE DE CONVERGENCE

Variable	K	$G(1)$	$G(2)$	$G(3)$	$G(4)$	$G(5)$	$G(6)$	$G(7)$	$G(8)$	$G(9)$	$G(10)$
Moyenne	66	81.6	87.2	90.6	93.1	94.8	96.2	97.1	97.8	98.3	98.6
Ecart-type	51	6.4	5.1	4.2	3.6	3.1	2.6	2.2	1.8	1.5	1.3

La Table II illustre la supériorité de l'estimation de maximum de vraisemblance Λ_K par rapport à la matrice d'autocovariance empirique Λ_0 aussi bien au niveau du biais que de l'écart quadratique moyen. Il est clair que la plus grande part de $(EQM)^{1/2}$ dans Λ_0 est due au biais. Ce biais est bien connu et provient du facteur $(1/mn)$ utilisé dans (19) pour assurer la positivité de Λ_0 . Cet effet peut être éliminé en remplaçant $(1/mn)$ par $[1/(m-k)n]$. Cependant Λ_0 ne sera plus nécessairement positive alors que Λ_K l'est toujours, tout en fournissant de bonnes valeurs moyennes. Il existe évidemment d'autres méthodes d'estimation à confronter avec Λ_K (cf. Ch. IV).

TABLE II
MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

Paramètres	Vraie Valeur Λ	Valeur Estimée Λ_K		Valeur Initiale Λ_0	
		Moyenne	(EQM) ^{1/2}	Moyenne	(EQM) ^{1/2}
$\Lambda(0)$	1.000	1.015	0.18	1.029	0.21
$\Lambda(1)$	0.500	0.502	0.15	0.470	0.16
$\Lambda(2)$	0.400	0.407	0.09	0.336	0.10
$\Lambda(3)$	-0.316	-0.323	0.08	-0.227	0.11
$\Lambda(4)$	-0.291	-0.289	0.12	-0.185	0.14
$\Lambda(5)$	-0.680	-0.684	0.15	-0.356	0.34
$\Lambda(6)$	-0.311	-0.300	0.18	-0.130	0.20
$\Lambda(7)$	-0.526	-0.525	0.16	-0.161	0.37
$\Lambda(8)$	-0.068	-0.064	0.14	-0.015	0.07
$\Lambda(9)$	-0.178	-0.194	0.17	-0.019	0.16

§ II.5. Maxima locaux de la fonction de vraisemblance

A l'exception du cas $m = 2$ la fonction de vraisemblance présente, avec une probabilité positive, des maxima locaux. Nous établissons ce résultat de façon analytique pour $m = 3$ et par simulation pour $m = 3, 4$ et 5 .

Nous considérons tout d'abord une autre paramétrisation des matrices de Toeplitz associée à tout élément Λ_0 de \mathcal{C}^+ . Soit ${}_+M_0{}_+D_0^2{}_+M_0^* = {}_-M_0{}__-D_0^2{}__-M_0^*$ les décompositions spectrales de Λ_0 . Le Lemme II.2.3 montre que tout élément Λ de \mathcal{C} s'écrit de façon unique sous la forme

$$\Lambda = {}_+M_0{}_+D_0\{aI + {}_+D + P^*{}_-DP\}{}_+D_0{}_+M_0^*$$

où a est un réel, ${}_+D$ et ${}_-D$ sont des matrices réelles diagonales satisfaisant les symétries nécessaires au caractère réel de Λ ainsi que $\text{tr}({}_+D) = \text{tr}({}_-D) = 0$ et où $P = {}_-D_0{}__-M_0^*{}_+M_0^*{}^{-1}{}_+D_0^{-1}$. Les colonnes de ${}_+M_0{}_+D_0$, comme celles de ${}__-M_0{}__-D_0$, forment une base orthonormée pour le produit scalaire Λ_0^{-1} et P est la matrice de passage

$$P = ({}_+M_0 {}_+D_0)^* \Lambda_0^{-1} {}_+M_0 {}_+D_0, \quad P^* P = I.$$

On établit ainsi une isométrie entre \mathcal{C} , muni du produit scalaire $\langle, \rangle_{\Lambda_0}$, et l'espace $\tilde{\mathcal{C}}$ des matrices de la forme $\tilde{\Lambda} = aI + {}_+D + P^* {}_+D P$ muni de \langle, \rangle_I . On a alors

$$L(\Lambda; X) = \ln |\Lambda_0| + \ln |\tilde{\Lambda}| + \text{tr}\{\tilde{\Lambda}^{-1} \tilde{X}\}, \quad \tilde{X} = {}_+D_0^{-1} {}_+M_0^{-1} X {}_+M_0^* {}_+D_0^{-1}.$$

D'après le Corollaire II.3.2, Λ_0 est solution de l'équation de vraisemblance si et seulement si

$$\text{diag } \tilde{X} = \text{diag } P \tilde{X} P^* = I.$$

D'autre part la forme quadratique définie par la dérivée seconde de $L(\Lambda; X)$ en Λ_0 et exprimée sur $\tilde{\mathcal{C}}$ est donnée par

$$(A, B) \in \tilde{\mathcal{C}}^2 \longrightarrow 2\text{tr}\{A \tilde{X} B\} - \text{tr}\{AB\}.$$

Elle est donc définie positive dès que les valeurs propres de $(A, B) \longrightarrow \text{tr}\{A \tilde{X} B\}$ sont supérieures à 1/2.

Nous nous plaçons désormais dans le cas $m = 3$. Nous écrivons les matrices ${}_+M_0$ et ${}_+M_0$ sous la forme centro-symétrique

$${}_+M_0 = \begin{bmatrix} e^{i\lambda} & 1 & e^{-i\lambda} \\ 1 & 1 & 1 \\ e^{-i\lambda} & 1 & e^{i\lambda} \end{bmatrix}, \quad {}_+M_0 = \begin{bmatrix} e^{i\tilde{\lambda}} & -1 & e^{-i\tilde{\lambda}} \\ 1 & 1 & 1 \\ e^{-i\tilde{\lambda}} & -1 & e^{i\tilde{\lambda}} \end{bmatrix}.$$

La matrice de passage P est alors réelle et doublement symétrique

$$P = \begin{bmatrix} (1+\cos\theta)/2 & (\sin\theta)/\sqrt{2} & -(1-\cos\theta)/2 \\ (\sin\theta)/\sqrt{2} & -\cos\theta & (\sin\theta)/\sqrt{2} \\ -(1-\cos\theta)/2 & (\sin\theta)/\sqrt{2} & (1+\cos\theta)/2 \end{bmatrix}, \quad \cos\theta = \frac{\tilde{\lambda}}{\lambda} \frac{\text{tg } \frac{\tilde{\lambda}}{2}}{\text{tg } \frac{\lambda}{2}}.$$

On a $0 < \theta < \pi/2$ car λ et $\tilde{\lambda}$ satisfont $0 < \tilde{\lambda} < \lambda < \pi$. On peut supposer X doublement symétrique, alors \tilde{X} l'est aussi et est réelle. On se place dans la situation où Λ_0 est une solution de l'équation de vraisemblance associée à X et la condition $\text{diag } X = I$ permet d'écrire

$$\tilde{X} = \begin{bmatrix} 1 & y & x \\ y & 1 & y \\ x & y & 1 \end{bmatrix}.$$

La contrainte $\text{diag } P\tilde{X}P = I$ équivaut à $x\sin\theta = 2\sqrt{2}y\cos\theta$ et on constate l'égalité $P\tilde{X}P = \tilde{X}$. La matrice X étant définie positive, le domaine de variation de x est égal à l'intervalle

$$\max\{-1, -2\cos\theta / (1+\cos\theta)\} < x < \min\{1, 2\cos\theta / (1-\cos\theta)\}. \quad (21)$$

Les matrices de l'espace $\tilde{\mathcal{E}}$ sont de la forme $aI + bD + cPDP$ où a, b et c sont des coefficients réels et D la matrice diagonale

$$D = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}.$$

Les matrices PDP et $PDP\tilde{X}$ sont centro-symétriques et de trace nulle, leurs diagonales sont donc proportionnelles à D et on obtient :

$$\text{diag } PDP = (3\cos^2\theta - 1)/2 D, \quad \text{diag } PDP\tilde{X} = \{3\cos^2\theta - 1 - (3/2)x\sin^2\theta\}/2 D.$$

Les vecteurs propres de $(A, B) \longrightarrow \text{tr}\{A\tilde{X}B\}$ sont les matrices I et $D \pm PDP$ et les valeurs propres correspondantes sont alors

$$1, \quad 1 - (3/2)x\sin^2\theta / (1 + 3\cos^2\theta), \quad 1 + x/2.$$

Λ_0 représente un minimum de $L(., X)$ si et seulement si, en plus de (21), x satisfait

$$x < -1 + 4/3\sin^2\theta. \quad (22)$$

Cette condition est automatiquement vérifiée si $\cos\theta < 2/\sqrt{3} - 1$. Dans le cas contraire, la variation de x dans l'intervalle (21) fait passer Λ_0 de la position de minimum à celle de point-selle. Ceci assure déjà la non-unicité des solutions de l'équation de vraisemblance.

Nous considérons maintenant les éléments de la fonction $\ln |\tilde{\Lambda}| + \text{tr}\{\tilde{\Lambda}^{-1}\tilde{X}\}$ sous l'hypothèse, toujours possible, que Λ_0 soit une solution de l'équation de vraisemblance.

Le déterminant de $\tilde{\Lambda} = aI + bD + cPDP$ et $\text{tr}\{\tilde{\Lambda}^{-1}\tilde{X}\}$ sont donnés par

$$|\tilde{\Lambda}| = \alpha^3 \{1 - 2\beta + \beta^2\sin^2\theta - \gamma^2\sin^2\theta\},$$

$$\text{tr}\{\tilde{\Lambda}^{-1}\tilde{X}\} = \alpha^2 \{3 - 4\beta + \beta^2(1-x)\sin^2\theta - \gamma^2(1-x)\sin^2\theta\}/|\tilde{\Lambda}|,$$

$$\alpha = a + (b+c)/2, \quad \beta = 3(b+c)/4\alpha, \quad \gamma = 3(b-c)/4\alpha.$$

Le domaine de variation de (α, β, γ) garantissant la positivité de $\tilde{\Lambda}$ est défini par les inégalités :

$$\alpha > 0, \quad \beta < 1/(1 + \cos\theta), \quad \gamma^2 \sin^2\theta < 1 - 2\beta + \beta^2 \sin^2\theta.$$

Posons $h(\alpha, \beta, \gamma) = \ln |\tilde{\Lambda}| + \text{tr}\{\tilde{\Lambda}^{-1}\tilde{X}\}$. Le gradient de h ne s'annule que pour

$$\gamma = 0, \quad \alpha = \{2 - (1-x)\beta \sin^2\theta\} / 2\{1 - \beta \sin^2\theta\},$$

et β solution de

$$\beta\{(1-x)\beta^2 \sin^4\theta + 2(2x-1)\beta \sin^2\theta - 3(1+x)\sin^2\theta + 4\} = 0. \quad (23)$$

$\beta = \gamma = 0$ et $\alpha = 1$ correspond à $\tilde{\Lambda} = I$ ou encore $\Lambda = \Lambda_0$. On montre qu'il s'agit de la seule solution lorsque x vérifie

$$x < \sqrt{\frac{3\cos^2\theta}{1 + 3\cos^2\theta}}. \quad (24)$$

Dans le cas contraire, et si x satisfait (21), les trois solutions que donne (23) déterminent deux minima locaux et un point-selle car, au voisinage de chacun d'eux, l'expression dans (23) a le même signe que la dérivée de la fonction

$$g(\beta) = \min_{\gamma} \{ \min_{\alpha} h(\alpha, \beta, \gamma) \}.$$

On peut vérifier que la borne de x dans (24) est inférieure à celles de (22) et (21).

Nous nous proposons ci-dessous de constater par simulation l'existence de minima locaux pour la fonction $L(\cdot; X)$. Pour cela nous nous plaçons volontairement dans une situation très défavorable (modèle proche de la singularité, petit échantillon). Pour $m = 5$, $n = 3$ et le modèle donné dans la Table III, la méthode est appliquée sur 1000 répétitions à partir de deux valeurs initiales: la matrice d'autocovariance empirique Λ_0 et la matrice identité I . Le critère d'arrêt est quadratique et porte sur deux itérés successifs de même nature (de type $+\hat{\Lambda}$):

$$\left\{ \frac{1}{m-1} \sum_{j=1}^{m-1} [\beta_{k+2}(j) - \beta_k(j)]^2 \right\}^{1/2} \leq \varepsilon. \quad (25)$$

Nous mesurons l'écart entre les deux solutions obtenues selon ce même critère. Pour $\varepsilon = 10^{-3}$ nous avons observé 78 cas pour lesquels cet écart était supérieur à 10^{-1} et 43 pour $\varepsilon = 10^{-4}$. Nous calculons, pour chacun d'eux, la valeur du gradient et la matrice des dérivées secondes selon la méthode présentée dans Dégerine et Pham (1987). On peut

ainsi constater que le critère (25) conduit à des solutions en lesquelles le gradient reste fortement différent de zéro, d'où la nécessité de lui adjoindre un critère portant sur la nullité de la norme du gradient. En diminuant ϵ ($\epsilon = 10^{-8}$) on obtient finalement 31 cas d'existence de deux minima. L'écart entre les deux solutions est souvent dû à une forte différence sur les valeurs de $\beta(4)$ dont les signes sont en général opposés. D'autre part $|\beta(2)|$ étant proche de 1 cet écart est beaucoup moins important sur les valeurs de Λ . La Table III ci-dessous donne les caractéristiques de l'un d'eux.

TABLE III
EXEMPLE DE MINIMUM LOCAL $m = 5$

Paramètres	Vraie Valeur Λ	Valeur Initiale Λ_0	Valeur Initiale I
		Valeurs Estimées Λ_K	
$\beta(1)$	0.716	0.702	0.742
$\beta(2)$	-0.982	-0.979	-0.972
$\beta(3)$	0.704	0.730	0.893
$\beta(4)$	-0.924	-0.872	0.757
$\Lambda(0)$	7.617	10.635	13.786
$\Lambda(1)$	5.454	7.463	10.234
$\Lambda(2)$	0.260	-0.045	1.580
$\Lambda(3)$	-4.894	-7.203	-7.331
$\Lambda(4)$	-7.065	-9.702	-11.508
$L(\Lambda_K; X)$		-0.671	-0.645

Dans cet exemple la statistique X est obtenue en tronquant les valeurs de l'observation simulée initialement. La matrice X , doublement symétrique, est définie par une partie triangulaire :

$$X = \begin{pmatrix} 11.518 & & & & \\ 6.881 & 9.102 & & & \\ -1.099 & 5.965 & 9.033 & & \\ -7.818 & -1.040 & & & \\ -9.799 & & & & \end{pmatrix}$$

Les exemples en dimensions 3 et 4 sont obtenus de la même façon.

$$\begin{array}{ccc}
 & 0.335 & \\
 X = & -0.367 & 0.410 \\
 & 0.323 & \\
 & & 0.497678 \\
 X = & 0.294071 & 0.595421 \\
 & -0.212032 & 0.335699 \\
 & & -0.488536
 \end{array}$$

TABLE IV
EXEMPLE DE MINIMUM LOCAL $m = 3, 4$

Paramètres	Vraie Valeur Λ	Valeur Initiale Λ_0	Valeur Initiale I
		Valeurs Estimées Λ_K	
$\beta(1)$	-0.950	-0.900	-0.980
$\beta(2)$	-0.980	-0.976	-0.331
$\Lambda(0)$	1.000	0.220	0.303
$\Lambda(1)$	-0.950	-0.198	-0.297
$\Lambda(2)$	0.807	0.137	0.288
$L(\Lambda_K; X)$		-3.636	-3.398
$\beta(1)$	0.500	0.484	0.531
$\beta(2)$	-0.980	-0.987	-0.996
$\beta(3)$	0.990	1.000	-0.126
$\Lambda(0)$	1.000	0.495	0.546
$\Lambda(1)$	0.500	0.239	0.290
$\Lambda(2)$	-0.980	-0.258	-0.236
$\Lambda(3)$	0.990	-0.475	-0.540
$L(\Lambda_K; X)$		-4.570	-3.305

CHAPITRE III

ESTIMATION DE LA STRUCTURE DE COVARIANCE
D'UNE SEQUENCE STATIONNAIRE

Soit x_1, \dots, x_m l'observation d'une séquence de longueur m d'un processus $X(\cdot)$, réel, stationnaire au second ordre et centré. Les statistiques

$$\frac{1}{m-k} \sum_{t=1}^{m-k} x_t x_{t+k}, \quad k = 0, \dots, m-1,$$

constituent des estimateurs sans biais des autocovariances $\Lambda(k)$ correspondantes. Cependant il est bien connu que la fonction d'autocovariance ainsi estimée n'est pas nécessairement de type positif. C'est pourquoi les autocovariances empiriques usuelles sont celles définies par

$$\Lambda_e(k) = \frac{1}{m} \sum_{t=1}^{m-k} x_t x_{t+k}, \quad k = 0, \dots, m-1. \quad (1)$$

Les estimateurs sont alors biaisés mais la positivité est assurée. Ces estimations sont souvent utilisées comme base dans l'élaboration de méthodes plus spécifiques : ajustement de modèles, estimation de paramètres, etc. De même, sur le plan spectral, les méthodes sont construites, dans de nombreux cas, à partir du périodogramme

$$I_n(\lambda) = \frac{1}{2\pi m} \left| \sum_{t=1}^m x_t e^{-it\lambda} \right|^2 = \frac{1}{2\pi} \sum_{k=-(m-1)}^{m-1} \Lambda_e(k) \cos k\lambda, \quad -\pi \leq \lambda \leq \pi.$$

Ces deux statistiques sont équivalentes mais ne sont pas exhaustives dans le cas gaussien. Nous considérons ici d'autres estimations possibles de la structure de covariance $\Lambda(\cdot)$. Notons que l'autocovariance empirique $\Lambda_e(\cdot)$ est celle d'une moyenne mobile d'ordre m .

L'étude des prolongements presque-périodiques de la séquence x_1, \dots, x_m , présentée au premier paragraphe, permet de comparer, dans le paragraphe suivant, l'estimation empirique (1) avec celle du maximum de vraisemblance dans le cas gaussien. L'estimation empirique naturelle de la fonction d'autocorrélation partielle diffère de celle qui résulte de (1), dans la résolution des relations de Yule-Walker, et fait l'objet du troisième paragraphe.

§ III.1. Prolongements presque-périodiques d'une séquence x_1, \dots, x_m

Une fonction presque-périodique (p.p.) discrète peut être caractérisée comme une suite de nombres complexes $\{x(t), t \in \mathbb{Z}\}$ admettant une représentation spectrale de la forme :

$$x(t) = \sum_{k=-\infty}^{+\infty} \gamma_k e^{it\lambda_k}, \quad t \in \mathbb{Z}, \quad (2)$$

où $\{\lambda_k, k \in \mathbb{Z}\}$ est une suite de fréquences et $\{\gamma_k, k \in \mathbb{Z}\}$ une suite de nombres complexes vérifiant :

$$-\pi < \dots < \lambda_{-k} < \dots < \lambda_0 = 0 < \dots < \lambda_k < \dots \leq \pi, \quad \sum_{k=-\infty}^{+\infty} |\gamma_k|^2 < +\infty.$$

L'ensemble des fonctions p.p. discrètes, muni du produit scalaire

$$\langle x(\cdot), y(\cdot) \rangle = \lim_{T \rightarrow +\infty} \frac{1}{2T+1} \sum_{t=-T}^{+T} x(t) \overline{y(t)},$$

forme un espace de Hilbert pour lequel les suites $\{e^{it\lambda}, t \in \mathbb{Z}\}$, $-\pi < \lambda \leq \pi$, constituent une base orthonormale

$$\lim_{T \rightarrow +\infty} \frac{1}{2T+1} \sum_{t=-T}^{+T} x(t) e^{-it\lambda} = \gamma_k, \quad \text{si } \lambda = \lambda_k, \quad k \in \mathbb{Z}, \quad 0 \text{ sinon.}$$

Les fonctions p.p. discrètes à valeurs réelles forment un sous-espace de Hilbert réel et sont caractérisées par les représentations spectrales (2) vérifiant :

$$\lambda_{-k} = -\lambda_k, \quad \gamma_{-k} = \overline{\gamma_k}, \quad k \in \mathbb{Z}, \quad (\lambda_k = \pi \Rightarrow \gamma_k \in \mathbb{R}).$$

Dans le cas réel la "fonction d'autocovariance" de $x(\cdot)$ est définie par

$$\Lambda(\tau) = \lim_{T \rightarrow +\infty} \frac{1}{2T+1} \sum_{t=-T}^{+T} x(t)x(t+\tau), \quad \tau \in \mathbb{Z}.$$

$\Lambda(\cdot)$ est elle-même une fonction p.p. discrète réelle, symétrique ($\Lambda(-\tau) = \Lambda(\tau)$, $\tau \in \mathbb{Z}$), qui admet la représentation spectrale

$$\Lambda(\tau) = \sum_{k=-\infty}^{+\infty} |\gamma_k|^2 e^{i\tau\lambda_k}, \quad \tau \in \mathbb{Z}.$$

C'est une fonction de type positif. La mesure spectrale correspondante est celle qui place les masses $|\gamma_k|^2$ en les fréquences λ_k , $k \in \mathbb{Z}$, et est appelée spectre de $x(\cdot)$. Nous retrouvons, ainsi associés à la fonction $x(\cdot)$, les éléments habituels décrivant la structure du second ordre d'un processus stationnaire $X(\cdot)$: $\rho(\cdot)$, $\beta(\cdot)$, etc.

Nous appelons **dimension** de $x(\cdot)$ le cardinal de l'ensemble des fréquences λ_k , $k \in \mathbb{Z}$, effectivement présentes dans le spectre, c'est-à-dire celles pour lesquelles $|\gamma_k|^2$ est non nul. C'est aussi la dimension du sous-espace de Hilbert réel engendré par les translatés $x_\tau(\cdot)$ de $x(\cdot)$, $\tau \in \mathbb{Z}$: $x_\tau(t) = x(t+\tau)$, $t \in \mathbb{Z}$. Par la suite nous considérons essentiellement les fonctions p.p. discrètes réelles de dimension finie et nous écrivons leur représentation spectrale sous la forme

$$x(t) = \sum_{k=1}^r \gamma_k e^{i t \lambda_k}, \quad t \in \mathbb{Z}, \quad (3)$$

où les fréquences λ_k , $k = 1, \dots, r$ sont rangées par ordre croissant, $-\pi < \lambda_1 < \dots < \lambda_r \leq \pi$ et vérifient avec les nombres complexes γ_k , $k = 1, \dots, r$ les conditions qui assurent le caractère réel de $x(\cdot)$.

III.1.1. Définition.—(Dégerine, 1983) *Nous appelons prolongement presque-périodique d'une séquence de nombres réels x_1, \dots, x_m , toute fonction p.p. discrète réelle $x(\cdot)$ vérifiant $x(t) = x_t$, $t = 1, \dots, m$.*

En considérant le temps continu dans la représentation (3) de ce prolongement, celui-ci peut être interprété comme une superposition de sinusoides :

$$x(t) = \sum \alpha_k \sin(\lambda_k t + \phi_k), \quad t \in \mathbb{R}, \quad (4)$$

interpolant les valeurs x_t , $t = 1, \dots, m$.

III.1.2. Proposition.—(Dégerine, 1983) *Quelle que soit la séquence x_1, \dots, x_m , et pour tout choix des fréquences $-\pi < \lambda_1 < \dots < \lambda_m \leq \pi$, il existe un prolongement p.p. unique de x_1, \dots, x_m basé sur $\lambda_1, \dots, \lambda_m$:*

$$x(t) = \sum_{k=1}^m \gamma_k e^{it\lambda_k}, \quad t \in \mathbb{Z}.$$

Le vecteur $\gamma = (\gamma_1, \dots, \gamma_m)'$ s'exprime, en fonction de $x = (x_1, \dots, x_m)'$ et de la matrice M associée à $\lambda_1, \dots, \lambda_m$, par $\gamma = M^{-1}x$. Le spectre de $x(\cdot)$ est précisé par la matrice diagonale $D = \text{diag}(\gamma\gamma^*)$.

La dimension du prolongement est au plus égale à m ; elle est strictement inférieure à m lorsque certaines des valeurs γ_k sont nulles. Notons qu'un autre choix de l'origine des temps, $x(t+k) = x_t$, $k = 1, \dots, m$, ou encore l'inversion du temps, $x(t-k) = x_t$, $k = 1, \dots, m$, ne modifie pas le spectre de $x(\cdot)$: dans (4) les amplitudes α_k sont invariantes, seuls les déphasages ϕ_k sont modifiés. En fait la suite des valeurs x_k , $k \in \mathbb{Z}$, prises par $x(\cdot)$ et indexée à partir de la séquence donnée x_1, \dots, x_m ($x_k = x(k)$ selon la Définition III.1.1) est invariante et peut être calculée par la récurrence

$$\sum_{k=0}^m b(m,k)x_{t-k} = 0, \quad t \in \mathbb{Z},$$

où les coefficients $b(m,k)$ sont ceux du polynôme

$$\phi_m(z) = \sum_{k=0}^m b(m,k)z^{m-k} = \prod_{k=1}^m (z - e^{i\lambda_k}).$$

Notant $\tilde{\Gamma}$ la matrice de corrélation MM^*/m , le vecteur γ s'écrit $M^*\tilde{\Gamma}^{-1}x/m$ et la fonction d'autocovariance du prolongement est donnée par

$$\Lambda(\tau) = \sum_{k=1}^m \gamma_k \bar{\gamma}_k e^{it\lambda_k} = (x_{\tau+1}, \dots, x_{\tau+m}) \tilde{\Gamma}^{-1} x, \quad \tau \in \mathbb{Z}.$$

Ainsi tout choix de m fréquences, $\lambda_1, \dots, \lambda_m$, ou, de manière équivalente, toute matrice M définit un prolongement de la séquence x_1, \dots, x_m de dimension au

plus m . Soit Γ une matrice d'autocorrélation non singulière d'ordre m . Elle définit, par le couple $({}_+M, {}_-M)$ qui la caractérise, deux prolongements ${}_+x(\cdot)$ et ${}_-_x(\cdot)$ de x . Toute combinaison convexe de ces prolongements

$${}_r x(t) = r {}_+x(t) + (1-r) {}_-x(t), \quad t \in \mathbb{Z},$$

constitue un prolongement basé sur les $2m$ fréquences de $({}_+M, {}_-M)$. Avec des notations évidentes on a

$${}_r \Lambda(\tau) = r^2 {}_- \Lambda(\tau) + (1-r)^2 {}_+ \Lambda(\tau), \quad \tau \in \mathbb{Z},$$

et le prolongement de variance (énergie) minimum est réalisé lorsque les deux prolongements ${}_r {}_+x(t)$ et $(1-r) {}_-x(t)$ ont même variance, il est obtenu pour $r = {}_- \Lambda(0) / [{}_- \Lambda(0) + {}_+ \Lambda(0)]$. Par la suite nous omettrons l'indice r pour le désigner et $\Gamma(x)$ sera sa matrice d'autocorrélation d'ordre m . De cette manière une séquence x définit une application $\Gamma \longrightarrow \Gamma(x)$, notée P_x , de $C^+(m)$ dans $\bar{C}^+(m)$. Il va de soi que P_x peut être regardé comme opérant sur tout ensemble qui caractérise $C^+(m)$.

III.1.3. Définition.—(Dégerine, 1983) *Une matrice d'autocorrélation non singulière d'ordre m , Γ , ou toute paramétrisation équivalente, est dite x -stable lorsqu'elle est un point fixe de l'opérateur P_x .*

III.1.4. Proposition.—(Dégerine, 1983) *La matrice Γ est x -stable si et seulement si les matrices ${}_+ \Gamma(x)$ et ${}_-_ \Gamma(x)$ des prolongements ${}_+x(\cdot)$ et ${}_-_x(\cdot)$ sont égales. Elles sont alors égales à Γ et*

$$x(\cdot) = \{ {}_+x(\cdot) + {}_-x(\cdot) \} / 2, \quad {}_+ \Lambda(0) = {}_- \Lambda(0) = 2\Lambda(0).$$

Démonstration.— Notons les relations :

$${}_+ \Lambda(0) = \sum_{k=1}^m | \gamma_k |^2 = \frac{1}{m} x' {}_+ \Gamma(x)^{-1} x, \quad {}_+ \gamma = {}_+ M^{-1} x,$$

$${}_+ \Gamma(x) = {}_+ M {}_+ D(x) {}_+ M^*, \quad {}_+ \Lambda(0) {}_+ D(x) = \text{diag}({}_+ \gamma, {}_+ \gamma^*).$$

Il en est de même pour les éléments associés à ${}_-_x(\cdot)$. Si ${}_+ \Gamma(x)$ et ${}_-_ \Gamma(x)$ sont égales, elles coïncident avec $\Gamma(x)$, par combinaison convexe, et avec Γ selon la caractérisation

$({}_+M, {}_-M)$. Réciproquement si $\Gamma = {}_+M {}_+D {}_+M^* = \Gamma(x)$ on a

$${}_+M {}_+D {}_+M^* = [{}_-\Lambda(0) {}_+M {}_+D(x) {}_+M^* + {}_+\Lambda(0) {}_-M {}_-D(x) {}_-M^*] / [{}_+\Lambda(0) + {}_-\Lambda(0)],$$

et par suite

$${}_+M \{ [{}_+\Lambda(0) + {}_-\Lambda(0)] {}_+D - {}_-\Lambda(0) {}_+D(x) \} {}_+M^* = {}_+\Lambda(0) {}_-M {}_-D(x) {}_-M^* = {}_+\Lambda(0) {}_- \Gamma(x).$$

Ainsi ${}_-\Gamma(x)$ est égale à Γ , selon la caractérisation $({}_+M, {}_-M)$, et donc à ${}_+\Gamma(x)$ car Γ est combinaison convexe des deux. Les autres relations sont immédiates. ♦

Lorsque $x(\cdot)$ est un prolongement basé sur les $2m$ fréquences du couple $({}_+M, {}_-M)$ d'une matrice Γ x -stable, sa mesure spectrale normalisée est la demi-somme de celles de ${}_+x(\cdot)$ et de ${}_-\Gamma(x)$. La structure d'une telle mesure est précisée dans la proposition suivante :

III.1.5. Proposition.—(Dégerine, 1983) Soit ${}_+F$ (resp. ${}_-\Gamma$) la mesure spectrale définie par les autocorrélations partielles $\beta_1, \dots, \beta_{m-1}$, $|\beta_k| < 1$, et $\beta_m = +1$ (resp. -1). Alors la mesure $F = ({}_+F + {}_-F)/2$ se caractérise par $\beta_{2m} = 1$ et la séquence antisymétrique :

$$\beta_1, \dots, \beta_{m-1}, \quad \beta_m = 0, \quad \beta_{m+k} = -\beta_{m-k}, \quad k = 1, \dots, m-1.$$

La démonstration de ce résultat repose sur le lemme suivant dans lequel le polynôme ϕ_{m-1} et ceux intervenant par la suite sont les polynômes de Szegö habituellement liés aux autocorrélations partielles.

III.1.6. Lemme.—Soit F une mesure spectrale (de masse 1) définie par $\beta_1, \dots, \beta_{m-1}$, $|\beta_k| < 1$, $\beta_m = 1$ (resp. $\beta_m = -1$). La mesure G , de densité $|\phi_{m-1}(\cdot)|^2$ par rapport à F , se caractérise par la suite $\tilde{\beta}_1, \dots, \tilde{\beta}_{m-1}$ donnée par

$$\tilde{\beta}_k = -\beta_{m-k}, \quad (\text{resp. } \tilde{\beta}_k = \beta_{m-k}), \quad k = 1, \dots, m-1, \quad \tilde{\beta}_m = \beta_m,$$

et la masse totale

$$\sigma_{m-1}^2 = \prod_{k=1}^{m-1} (1 - \beta_k^2).$$

Démonstration.— Nous considérons le cas $\beta_m = 1$, la preuve étant analogue pour

$\beta_m = -1$. L'égalité $\tilde{\beta}_m = 1$ provient de ce que la fréquence $\lambda = 0$ fait partie du support de F , et donc de celui de G , car le polynôme ϕ_{m-1} ne s'annule pas sur le cercle unité.

On a

$$\int_{-\pi}^{+\pi} G(d\lambda) = \int_{-\pi}^{+\pi} |\phi_{m-1}(\lambda)|^2 F(d\lambda) = \|\phi_{m-1}\|_F^2 = \sigma_{m-1}^2.$$

L'application $J(e^{ik\lambda}) = e^{i(k+1)\lambda}\phi_{m-1}(\lambda)$, $k \in \mathbb{Z}$, se prolonge, de façon naturelle, en une isométrie, encore notée J , entre l'espace vectoriel réel engendré par les fonctions $e^{ik\lambda}$, $k \in \mathbb{Z}$, muni du produit scalaire défini par G et celui engendré par les fonctions $e^{i(k+1)\lambda}\phi_{m-1}(\lambda)$, $k \in \mathbb{Z}$, muni du produit scalaire défini par F . Notons $\tilde{\phi}_k$ et $\tilde{\phi}_k^*$, $k = 0, \dots, m$, les polynômes associés à G , et donc caractérisés par $\tilde{\beta}_1, \dots, \tilde{\beta}_m$.

Nous avons

$$\tilde{\phi}_0(\lambda) = \tilde{\phi}_0^*(\lambda) \equiv 1, \quad J\{\tilde{\phi}_0(\lambda)\} = e^{i\lambda}\phi_{m-1}(\lambda) = \phi_{m-1}^*(\lambda);$$

la dernière égalité est due à $\beta_m = 1$; elle s'entend au sens $L^2(F)$ et donc F (ou G) presque partout. Il en sera de même pour les égalités du même type qui suivront sans que cela soit précisé. Calculons $\tilde{\beta}_1$.

$$\tilde{\beta}_1^2 \sigma_{m-1}^2 = \langle e^{i\lambda}, 1 \rangle_G = \langle e^{i\lambda}\phi_{m-1}, \phi_{m-1} \rangle_F = \langle \phi_{m-1}^*, \phi_{m-1} \rangle_F = -\beta_{m-1} \sigma_{m-1}^2.$$

L'égalité ci-dessus est un cas particulier de

$$\langle \phi_k, \phi_k^* \rangle = -\beta_k \sigma_k^2, \quad \sigma_k^2 = \|\phi_k\|^2, \quad k \geq 1,$$

qui se vérifie aisément. Nous avons ainsi $\tilde{\beta}_1 = -\beta_{m-1}$ et par suite :

$$J(\tilde{\phi}_1) = e^{i\lambda}(\phi_{m-1}^* + \beta_{m-1}\phi_{m-1}) = (1 - \beta_{m-1}^2)e^{i\lambda}\phi_{m-2}^* = \tilde{\sigma}_1^2 e^{i\lambda}\phi_{m-2}^*, \quad \tilde{\sigma}_1^2 = 1 - \beta_{m-1}^2.$$

Nous obtenons de même :

$$J(\tilde{\phi}_1^*) = \tilde{\sigma}_1^2 e^{2i\lambda}\phi_{m-2}.$$

Faisons l'hypothèse que ces relations sont satisfaites jusqu'à l'ordre k , $k < m-1$, et calculons $\tilde{\beta}_{k+1}$.

$$\tilde{\beta}_{k+1} \tilde{\sigma}_k^2 \sigma_{m-1}^2 = \langle e^{i\lambda}\tilde{\phi}_k, \tilde{\phi}_k^* \rangle_G = \tilde{\sigma}_k^4 \langle \phi_{m-k-1}^*, \phi_{m-k-1} \rangle_F = -\beta_{m-k-1} \tilde{\sigma}_k^2 \sigma_{m-1}^2.$$

D'où :

$$J(\tilde{\phi}_{k+1}) = \tilde{\sigma}_k^2 e^{i(k+1)\lambda} (\phi_{m-k-1}^* + \beta_{m-k-1} \phi_{m-k-1}) = \tilde{\sigma}_{k+1}^2 e^{i(k+1)\lambda} \phi_{m-k-2}^*$$

L'expression de $J(\tilde{\phi}_{k+1}^*)$ s'obtient de la même façon et la récurrence est terminée. Notons que le calcul ci-dessus, avec $k = m-1$, conduit à $\tilde{\beta}_m = 1$ et $J(\tilde{\phi}_m) = J(\tilde{\phi}_m^*) = 0$. Dans le cas $\beta_m = -1$ les relations sont, pour $k = 1, \dots, m-1$:

$$J(\tilde{\phi}_k) = -\tilde{\sigma}_k^2 e^{ik\lambda} \phi_{m-k-1}^*, \quad J(\tilde{\phi}_k^*) = \tilde{\sigma}_k^2 e^{i(k+1)\lambda} \phi_{m-k-1} \cdot \diamond$$

Dans la démonstration ci-dessus, l'isométrie appliquée en $m-1$ montre que le polynôme normalisé $\Phi_{m-1} = \phi_{m-1}/\sigma_{m-1}$, associé à une suite $\beta_1, \dots, \beta_{m-1}$, et l'analogue Ψ_{m-1} , associé à la suite $-\beta_{m-1}, \dots, -\beta_1$, (resp. $\beta_{m-1}, \dots, \beta_1$) satisfont

$$\Phi_{m-1}(e^{i\lambda_k}) \Psi_{m-1}(e^{i\lambda_k}) = e^{i(m-2)\lambda_k}, \quad (\text{resp. } -e^{i(m-2)\lambda_k}),$$

en les fréquences $\lambda_k, k = 1, \dots, m$ solutions de $\phi_m(e^{i\lambda}) = 0$ (support de F) où ϕ_m correspond à la valeur supplémentaire $\beta_m = 1$ (resp. -1).

Démonstration de la Proposition III.1.5.—Pour la partie directe de la démonstration, nous construisons les polynômes ϕ_k et $\phi_k^*, k \geq 0$, associés à la mesure F . Pour $k \leq m-1$, il est clair que ce sont ceux définis par $\beta_1, \dots, \beta_{m-1}$. Calculons β_m .

$$\beta_m \sigma_{m-1}^2 = \langle e^{i\lambda} \phi_{m-1}, \phi_{m-1}^* \rangle_F = \{ \langle e^{i\lambda} \phi_{m-1}, \phi_{m-1}^* \rangle_{+F} + \langle e^{i\lambda} \phi_{m-1}, \phi_{m-1}^* \rangle_{-F} \} / 2 = 0,$$

et par conséquent

$$\phi_m = e^{i\lambda} \phi_{m-1}, \quad \phi_m^* = \phi_{m-1}^*, \quad \sigma_m^2 = \sigma_{m-1}^2.$$

Nous utilisons ci-après les notations et résultats de la démonstration du Lemme III.1.6, en précisant par les indices $+$ ou $-$ les références à ${}_+F$ ou $_-F$; de plus $F, {}_+F$ et $_-F$ désignent les indicatrices des supports des mesures correspondantes. Établissons, pour une récurrence, les relations à l'ordre $m+1$:

$$\begin{aligned} \beta_{m+1} \sigma_m^2 &= \langle e^{i\lambda} \phi_m, \phi_m^* \rangle_F = \{ \langle e^{2i\lambda} \phi_{m-1}, e^{i\lambda} \phi_{m-1} \rangle_{+F} - \langle e^{2i\lambda} \phi_{m-1}, e^{i\lambda} \phi_{m-1} \rangle_{-F} \} / 2 \\ &= \{ \langle e^{i\lambda}, 1 \rangle_{+G} - \langle e^{i\lambda}, 1 \rangle_{-G} \} / 2 = -\beta_{m-1} \sigma_{m-1}^2. \end{aligned}$$

$$\phi_{m+1} F = {}_+\tilde{\phi}_1 \phi_m {}_+F + {}_-\tilde{\phi}_1 \phi_m {}_-F = \tilde{\sigma}_1^2 e^{i\lambda} \phi_{m-2}^* ({}_+F - {}_-F),$$

$$\begin{aligned}\phi_{m+1}^* F &= \tilde{\phi}_{+1}^* \phi_{m+} F + \tilde{\phi}_{-1}^* \phi_{m-} F = \tilde{\sigma}_1^2 e^{2i\lambda} \phi_{m-2} \{ {}_+F - {}_-F \}, \\ \|\phi_{m+1}\|^2 &= \|\phi_{m+1}^*\|^2 = \tilde{\sigma}_1^4 \sigma_{m-2}^2 = \tilde{\sigma}_1^2 \sigma_{m-1}^2 = \sigma_{m+1}^2.\end{aligned}$$

Avec ces relations à l'ordre $m+k$ on obtient :

$$\begin{aligned}\beta_{m+k+1} \sigma_{m+k}^2 &= \langle e^{i\lambda} \phi_{m+k}, \phi_{m+k}^* \rangle_F \\ &= \tilde{\sigma}_k^4 \{ \langle \phi_{m-k-1}^*, \phi_{m-k-1} \rangle_{+F} + \langle \phi_{m-k-1}^*, \phi_{m-k-1} \rangle_{-F} \} / 2 \\ &= \tilde{\sigma}_k^4 \langle \phi_{m-k-1}^*, \phi_{m-k-1} \rangle_F = -\beta_{m-k-1} \tilde{\sigma}_k^4 \sigma_{m-k-1}^2 = -\beta_{m-k-1} \sigma_{m+k}^2, \\ \phi_{m+k+1} F &= \tilde{\sigma}_k^2 e^{i(k+1)\lambda} (\phi_{m-k-1}^* + \beta_{m-k-1} \phi_{m-k-1}) \{ {}_+F - {}_-F \} \\ &= \tilde{\sigma}_k^2 e^{i(k+1)\lambda} \phi_{m-k-2}^* \{ {}_+F - {}_-F \},\end{aligned}$$

et de même

$$\phi_{m+k+1}^* F = \tilde{\sigma}_{k+1}^2 e^{i(k+2)\lambda} \phi_{m-k-2} \{ {}_+F - {}_-F \}.$$

Enfin nous obtenons $\beta_{2m} = 1$ en terminant cette récurrence ou plus simplement en constatant que la fréquence $\lambda = 0$ est dans le support de F .

Réciproquement nous construisons, par un calcul récursif, une factorisation de ϕ_{2m} :

$$\begin{aligned}\phi_{2m} &= e^{i\lambda} \phi_{2m-1} - \phi_{2m-1}^* = e^{i\lambda} \phi_1 \phi_{2m-2} - \phi_1^* \phi_{2m-2}^* = \dots \\ &= e^{i\lambda} \phi_{m-1} \phi_m - \phi_{m-1}^* \phi_m^*.\end{aligned}$$

D'où :

$$\phi_{2m} = [e^{i\lambda} \phi_{m-1}]^2 - [\phi_{m-1}^*]^2 = [e^{i\lambda} \phi_{m-1} - \phi_{m-1}^*][e^{i\lambda} \phi_{m-1} + \phi_{m-1}^*].$$

Cette factorisation montre que le support de F (solutions de $\phi_{2m}(z) = 0$) est constitué par les fréquences du couple $({}_+M, {}_-M)$ associé à $\beta_1, \dots, \beta_{m-1}$. D'après la partie directe, et l'unicité de la mesure F liée à une suite $\beta_1, \dots, \beta_{2m}$ donnée, la démonstration est achevée. ♦

§ III.2. Le maximum de vraisemblance et ses singularités

Soit $x = (x_1, \dots, x_{m-1})'$ une séquence de longueur m issue d'un processus gaussien stationnaire centré de fonction de covariance $\Lambda(\cdot)$. La loi de x est donc précisée

par la matrice d'autocovariance Λ d'ordre m . Maximiser la vraisemblance par rapport à Λ équivaut à minimiser la fonction (cf. Chapitre I)

$$L(\Lambda; X) = \ln |\Lambda| + \text{tr}(\Lambda^{-1}X), \quad X = xx'.$$

La Proposition I.2.1 indique le comportement non tangentiel de $L(\cdot; X)$ au bord de son domaine : Λ_0 et Λ désignant respectivement une matrice singulière et une matrice définie positive, pour $\varepsilon > 0$ on pose $\Lambda_\varepsilon = \Lambda_0 + \varepsilon(\Lambda - \Lambda_0)$ et on a

$$\lim_{\varepsilon \rightarrow 0} L(\Lambda_\varepsilon; X) = +\infty \text{ si } \ker(\Lambda_0) \not\subset \ker(X), \quad -\infty \text{ sinon.}$$

La condition $\ker(\Lambda_0) \not\subset \ker(X)$ équivaut à $x \notin \text{Im}(\Lambda_0)$, où $\text{Im}(\Lambda_0)$ désigne l'image de Λ_0 . Si r est le rang de Λ_0 ($r < m$), la matrice Λ_0 admet une décomposition spectrale unique $\Lambda_0 = MDM^*$ dans laquelle D est une matrice diagonale carrée d'ordre r dont les éléments diagonaux sont strictement positifs et M est la matrice $m \times r$ associée aux r fréquences $-\pi \leq \lambda_1 < \dots < \lambda_r < \pi$ composant le support de la mesure spectrale singulière définie par Λ_0 . L'appartenance de x à l'image de Λ_0 s'exprime par sa décomposition en fonction des colonnes $M_{\cdot k}$ de M :

$$x = \sum_{k=1}^r \gamma_k M_{\cdot k} \Leftrightarrow x_t = \sum_{k=1}^r \gamma_k e^{i\lambda_k t}, \quad t = 1, \dots, m.$$

Cette condition équivaut aussi à ce que x satisfasse l'équation récurrente

$$\sum_{j=0}^r b_j x_{t-j} = 0, \quad t = r+1, \dots, m, \quad \phi_r(z) = \sum_{j=0}^r b_j z^{r-j} = \prod_{k=1}^r (z - e^{i\lambda_k}). \quad (5)$$

Ainsi si x ne satisfait pas d'équation du type (5), avec $r < m$, la limite non tangentielle de $L(\cdot; X)$ au bord de son domaine est égale à $+\infty$. Ceci ne suffit pas cependant à assurer, en toute généralité, que la vraisemblance soit bornée, mais nous verrons que cette condition est nécessaire et suffisante pour presque tout x .

Remarquons que les racines du polynôme ϕ_r de (5) sont simples et sur le cercle unité. La condition suffisante $\ker(X) \cap B = \emptyset$ de la Proposition II.1.5 équivaut à autoriser la présence de racines multiples. L'ensemble B de \mathbb{R}^m est décrit par les

coefficients des polynômes ϕ_{m-1} associés, par la récurrence habituelle, aux suites $(\beta_1, \dots, \beta_{m-1})$ satisfaisant $|\beta_k| \leq 1, k < m-1$, et $|\beta_{m-1}| = 1$. Cet ensemble coïncide avec celui des coefficients b_0, \dots, b_{m-1} des polynômes réels d'ordre $m-1$, normalisés par la condition $b_0 = 1$,

$$\phi(z) = \sum_{j=0}^{m-1} b_j z^{m-1-j} = \prod_{k=1}^{m-1} (z - e^{i\lambda_k}), \quad (6)$$

dont les racines sont sur le cercle unité. Nous obtenons ce résultat en donnant de B une description canonique en fonction des β_k .

Rappelons la récurrence :

$$\phi_0 = \phi^*_0 \equiv 1, \quad \phi_k = z\phi_{k-1} - \beta_k \phi^*_{k-1}, \quad \phi^*_k = \phi^*_{k-1} - \beta_k z\phi_{k-1}, \quad k \geq 1. \quad (7)$$

Quelle que soit la suite β_1, β_2, \dots , l'égalité $|\beta_k| = 1$ équivaut à $\phi_k + \beta_k \phi^*_k = 0$ et la récurrence se poursuit avec des polynômes de la forme

$$\phi_{k+r} = \phi_k \psi_r, \quad \phi^*_{k+r} = \phi^*_k \psi^*_r, \quad r \geq 1,$$

où ψ_r et ψ^*_r sont construits selon la même récurrence avec les paramètres :

$$-\beta_k \beta_{k+j}, \quad j = 1, \dots, r.$$

D'autre part tout polynôme ϕ dans (6) se factorise à l'aide des polynômes élémentaires :

$$z - 1, \Leftrightarrow \beta_1 = 1; \quad z + 1, \Leftrightarrow \beta_1 = -1; \quad z^2 - 2z \cos \lambda + 1, \Leftrightarrow \beta_1 = \cos \lambda, \beta_2 = -1,$$

le dernier ayant $e^{\pm i\lambda}$ comme racines. Nous pouvons donc identifier les polynômes réels dont les racines sont sur le cercle unité et ceux construits par la récurrence (7) avec la paramétrisation canonique :

$$\beta_{2k-1} = \cos \lambda_k, \quad 0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p \leq \pi, \quad \beta_{2k} = -1, \quad k = 1, \dots, p,$$

correspondant à $2p$ racines situées en $e^{\pm i\lambda_k}, k = 1, \dots, p$, et, pour un polynôme d'ordre impair, $\beta_{2p+1} = \cos \lambda, \lambda \in \{0, \pi\}$. Dans le cas pair la séquence se termine par $\beta_{2p+1} = \beta_{2p+2} = 1$ lorsque l'ordre de multiplicité des racines ± 1 est impair.

La condition $\ker(X) \cap B \neq \emptyset$ signifie qu'il existe b dans B tel que $b'x = 0$ et cela équivaut à ce que la séquence x_1, \dots, x_m satisfasse une équation récurrente du type (5) d'ordre $r, r < m$, mais dont le polynôme présente éventuellement des racines multiples. Nous noterons RM le domaine correspondant de \mathbb{R}^m et RS celui qui exclut

la possibilité de racines multiples (*RMCRS*), ceci évidemment dans le cas d'équation récurrente d'ordre minimal (cf. Proposition II.1.3).

Le complémentaire de *RM* dans \mathbb{R}^m est un domaine sur lequel la vraisemblance est bornée. L'estimation de maximum de vraisemblance est alors réalisée par une matrice Λ non singulière dont nous précisons les propriétés par la suite. Dans *RS* la vraisemblance n'est pas bornée et x admet un prolongement presque-périodique de dimension minimale inférieure à m . La matrice de covariance Λ d'ordre m correspondante est singulière. La solution n'est pas nécessairement unique mais leur nombre est fini et les matrices singulières associées sont isolées dans \mathcal{E}^+ . Dans *RM* la vraisemblance peut être bornée (la condition $\ker(X) \cap B = \emptyset$ n'est pas nécessaire) mais de tels points forment un ensemble de mesure nulle dans \mathbb{R}^m . Ce n'est pas le cas du complémentaire de *RS* dans *RM* et, à l'intérieur de ce domaine, chaque x admet une infinité non dénombrable de prolongements presque-périodiques de dimension minimale. Nous illustrons ces propos pour $m = 5$.

Compte tenu de la description canonique de B , la condition $b'x = 0$ se traduit par l'une des deux égalités suivantes :

$$(x_1 + x_5) - 2(\beta_1 + \beta_3)(x_2 + x_4) + 2(1 + 2\beta_1\beta_3)x_3 = 0,$$

$$(x_1 - x_5) - 2\beta_1(x_2 - x_4) = 0.$$

Notons que $b'x = 0$ se décompose en deux contraintes indépendantes portant séparément sur la partie symétrique de x et sur sa partie antisymétrique. L'intérieur de *RS* est caractérisé ici par l'ensemble des x satisfaisant la deuxième égalité, avec $|\beta_1| < 1$, sans satisfaire la première et donne lieu à une seule matrice singulière Λ . Il est non vide car il contient, par exemple, un voisinage de $x = (2, -1, 1, 1, 2)'$. La vraisemblance est bornée là où aucune des égalités n'est satisfaite; ceci se produit donc avec une probabilité positive strictement inférieure à 1; le point $x = (0, 0, 0, 1, 5)'$ est à l'intérieur de ce domaine. Enfin un voisinage de $x = (0, 1, 1, -1, 1)'$ est inclus dans *RMRS*; il conduit à une matrice singulière isolée, par sa partie antisymétrique, et à une famille continue de

solutions par sa partie symétrique.

La séquence $x = (0,0,0,1,4)'$ admet, comme unique solution, $\beta_1 = \beta_3 = 1$. Elle correspond à une racine d'ordre 4 en 1. Dans ce cas la vraisemblance reste bornée et tend vers zéro au bord de son domaine, prouvant ainsi que la condition $\ker(X) \cap B = \emptyset$ n'est pas nécessaire. Pour le vérifier posons

$$\Lambda = \sigma_{\varepsilon}^2 R, \quad \sigma_{\varepsilon}^2 = \sigma^2 \prod_{k=1}^{m-1} (1 - \beta_k^2), \quad \sigma^2 = \Lambda(0).$$

La matrice R n'est fonction que des autocorrélations partielles. La minimisation de $L(\Lambda; X)$ par rapport à σ_{ε}^2 a pour solution

$$\sigma_{\varepsilon}^2 = \frac{1}{m} \operatorname{tr}(R^{-1}X),$$

et la vraisemblance est alors proportionnelle à

$$|\Lambda|^{-1/2} = \left\{ \frac{1}{m} \operatorname{tr}(R^{-1}X) \right\}^{-n/2} |R|^{-1/2}.$$

Il suffit donc de montrer qu'au voisinage du point

$$\beta_1 = 1 - \alpha, \quad \beta_2 = -1 + \beta, \quad \beta_3 = 1 - \gamma, \quad \beta_4 = -1 + \delta,$$

la partie déterminant l'emporte sur la partie trace, lorsque α, β, γ et δ tendent simultanément vers zéro en restant positifs. Ceci peut se constater sur les expressions suivantes :

$$|R|^{-1} = (2 - \alpha)\alpha(2 - \beta)^2\beta^2(2 - \gamma)^3\gamma^3(2 - \delta)^4\delta^4,$$

$$\operatorname{tr}(R^{-1}X) = \delta(2 - \delta) + \{2(\alpha + \beta + \gamma) + \delta - \alpha\beta - \gamma\beta - \gamma\delta\}^2.$$

Une séquence x , satisfaisant une équation récurrente minimale avec racines multiples ($x \in RM$), peut toujours être approchée par d'autres séquences obtenues en séparant chaque racine double par deux racines simples voisines. Ainsi la partie de RM dans laquelle x ne satisfait pas (5), avec $r < m$, est de mesure nulle dans \mathbb{R}^m . La condition $\ker(X) \cap B = \emptyset$ conserve cependant son importance dans le cas des échantillons.

Il est intéressant de comparer l'estimation empirique usuelle Λ_{ε} , donnée par (1), et

celle, lorsqu'elle existe, de maximum de vraisemblance Λ_{mv} .

Soit $\{ {}_+M_i, {}_-M_i \}$ la paramétrisation de la matrice identité d'ordre m (cf. Proposition II.2.2). La matrice ${}_+M_i$ (resp. ${}_ -M_i$) est associée aux fréquences λ pour lesquelles $e^{i\lambda}$ est racine $m^{\text{ième}}$ de $+1$ (resp. -1). Les éléments de l'espace des matrices de Toeplitz réelles symétriques $\mathcal{C}(m)$ se décomposent, de façon unique, sous la forme

$$\Lambda = {}_+M_i {}_+D {}_+M_i^* + {}_-M_i {}_-D {}_-M_i^* = {}_+\Lambda_i + {}_-\Lambda_i, \quad \text{tr}({}_+D) = \text{tr}({}_-D).$$

Les matrices Λ pour lesquelles ${}_+M = {}_+M_i$ (resp. ${}_ -M = {}_-M_i$) se caractérisent par la symétrie $\Lambda(k) = \Lambda(m-k)$ (resp. l'antisymétrie $\Lambda(k) = -\Lambda(m-k)$), $k = 1, \dots, m-1$. La décomposition ci-dessus est assurée par les relations

$${}_+\Lambda_i(k) = \{\Lambda(k) + \Lambda(m-k)\}/2, \quad {}_-\Lambda_i(k) = \{\Lambda(k) - \Lambda(m-k)\}/2, \quad k=0, \dots, m,$$

où l'on a posé $\Lambda(m) = 0$.

La fonction presque-périodique, caractérisée par les éléments spectraux $\{ {}_+M_i, {}_+D, {}_-M_i, {}_-D \}$, est périodique de période $2m$, symétrique ($\Lambda(k) = \Lambda(2m-k)$, $k = 0, \dots, 2m$), satisfait $\Lambda(m) = 0$ et prolonge les éléments $\Lambda(0), \dots, \Lambda(m-1)$. Réciproquement, un tel prolongement caractérise la décomposition de Λ . Les prolongements ainsi associés aux matrices définies positives ($\Lambda \in \mathcal{C}^+$) ne sont pas nécessairement de type positif; nous noterons \mathcal{C}_e^+ le sous-ensemble des matrices de \mathcal{C}^+ pour lesquelles ils le sont. Ce sont aussi les matrices dont les termes ${}_+\Lambda_i$ et ${}_-\Lambda_i$ sont dans \mathcal{C}^+ .

L'estimation empirique Λ_e appartient à \mathcal{C}_e^+ . Le prolongement de Λ_e est le double de la fonction d'autocovariance de la fonction périodique, de période $2m$, définie par $x_1, \dots, x_m, x_{m+k} = 0, k = 1, \dots, m$. La matrice d'autocorrélation Γ_e est l'image de l'identité par l'opérateur P_x considéré au premier paragraphe : ${}_+x(\cdot)$ est le prolongement périodique de x , de période m , et ${}_ -x(\cdot)$ est périodique de période $2m$ avec $x(2m-k) = -x(k)$, $k = 1, \dots, m$; ils ont même énergie et leur demi-somme est la fonction périodique évoquée plus haut. Les éléments ${}_+\Lambda_i$ et ${}_-\Lambda_i$ de Λ_e , multipliés par deux, sont les matrices d'autocovariances d'ordres m de ${}_+x(\cdot)$ et de ${}_ -x(\cdot)$; ce sont donc les estimations de maximum de vraisemblance de Λ sous l'hypothèse a priori de

symétrie ou d'antisymétrie (cf. Proposition II.3.1) et Λ_e apparaît comme un compromis (la demi-somme) entre les deux. Cependant Λ_e n'est pas l'estimateur de maximum de vraisemblance de Λ dans \mathcal{C}_e^+ .

L'estimation de maximum de vraisemblance Λ_{mv} , lorsqu'elle existe, est solution de l'équation de vraisemblance, laquelle se caractérise par une matrice d'autocorrélation Γ_{mv} x -stable (cf. Corollaire II.3.2). Ici le prolongement de x est presque-périodique de dimension $2m$, il est basé sur les fréquences du couple $\{+M_{mv}, -M_{mv}\}$ associé à Γ_{mv} et Λ_{mv} est sa matrice d'autocovariance d'ordre m . La fonction d'autocorrélation partielle de ce prolongement est antisymétrique (Proposition III.1.5) :

$$\beta_{mv}(2m) = 1, \quad \beta_{mv}(m) = 0,$$

$$\beta_{mv}(1), \dots, \beta_{mv}(m-1), \quad \beta_{mv}(m+k) = -\beta_{mv}(m-k), \quad k = 1, \dots, m-1.$$

D'autre part, la forme des polynômes ϕ_{m+k} et ϕ_{m+k}^* , $k = 0, \dots, m$, dans la démonstration de cette proposition, montre que le prolongement est tel que les erreurs de prédiction progressive et rétrograde non observées sont égales à zéro :

$$\varepsilon(m+k+1; m+k) = \varepsilon^*(-k; m+k) = 0, \quad k \geq 0,$$

ce qui est évidemment plus naturel que de prolonger la séquence x_1, \dots, x_m par des zéros. Le calcul de Λ_{mv} peut être réalisé par des méthodes de relaxation (cf. Ch II, IV).

Il reste que, la vraisemblance n'étant pas presque sûrement bornée, la méthode du maximum de vraisemblance est compromise. On peut alors envisager diverses contraintes sur Λ afin de garantir une solution : l'estimation autorégressive en est un exemple (cf. Ch. IV). Miller et Snyder (1987) imposent que la séquence $\Lambda(0), \dots, \Lambda(m-1)$ admette un prolongement périodique, de type positif, de période N . Pour $N = 2m$ cela équivaut à l'hypothèse $\Lambda \in \overline{\mathcal{C}}_e^+$. La solution est calculée selon le principe de l'EM-algorithme de Dempster et coll. (1977). L'existence du maximum provient de ce que les bords du domaine ainsi restreint ne sont pas nécessairement constitués de matrices singulières. Lorsque la solution est atteinte à l'intérieur du domaine, elle correspond au maximum, éventuellement local, du problème sans contrainte ($N \geq 2m$).

Dans la région RM il est vraisemblable qu'elle sera atteinte au bord. La méthode de relaxation, avec un nombre d'itérations limité, devrait avoir les mêmes effets bien que le principe soit différent car le domaine est sans contrainte.

§ III.3. Autocorrélations partielles empiriques

L'estimateur sans biais de $\Lambda(k)$,

$$\frac{1}{m-k} \sum_{t=1}^{m-k} x_t x_{t+k}$$

s'interprète comme la covariance empirique, entre $X(t)$ et $X(t+k)$, calculée sur les $m-k$ observations $\{x_t, x_{t+k}\}$, $t = 1, \dots, m-k$. Il est donc tout aussi naturel de définir l'autocorrélation partielle empirique $\beta_e(k)$, entre $X(t)$ et $X(t+k)$, comme la corrélation partielle, calculée dans \mathbb{R}^{m-k} , entre $\{x_1, \dots, x_{m-k}\}$ et $\{x_{k+1}, \dots, x_m\}$ dans l'ensemble $\{x_t, \dots, x_{t+m-k-1}\}$, $t = 1, \dots, k+1$. L'intérêt est que $\beta_e(\cdot)$ définit toujours, à travers les relations de Yule-Walker, une fonction $\Gamma(\cdot)$ de type positif; celle-ci ne correspond évidemment pas à l'estimation sans biais ci-dessus, ni à la covariance empirique $\Lambda_e(\cdot)$.

Introduisons les notations suivantes :

$$\mathbf{x}_k(t) = \{x_t, \dots, x_{t+m-k-1}\}', \quad t = 1, \dots, k+1, \quad k = 0, \dots, m-1,$$

$$\mathbf{e}_k(t;j) = \mathbf{x}_k(t) - \mathbf{x}_k(t;j), \quad 1 \leq j < t, \quad \mathbf{e}_k(t;0) = \mathbf{x}_k(t),$$

où $\mathbf{x}_k(t;j)$ est la projection de $\mathbf{x}_k(t)$ sur le sous-espace engendré par $\mathbf{x}_k(t-1), \dots, \mathbf{x}_k(1)$, le produit scalaire utilisé dans \mathbb{R}^{m-k} étant

$$\langle U, V \rangle = \frac{1}{m-k} \sum_{t=1}^{m-k} u_t v_t.$$

Le vecteur $\mathbf{e}_k(t;j)$ représente "l'erreur de prédiction" de $\mathbf{x}_k(t)$ par son passé immédiat de longueur j . De même les erreurs de prédiction rétrograde sont notées :

$$\mathbf{e}_k^*(t;j) = \mathbf{x}_k(t) - \mathbf{x}_k^*(t;j), \quad 1 \leq j < k-t, \quad \mathbf{e}_k^*(t;0) = \mathbf{x}_k(t).$$

L'autocorrélation partielle empirique d'ordre k est alors définie par

$$\beta_e(k) = \langle \mathbf{e}_k(k+1;k-1), \mathbf{e}_k^*(1;k-1) \rangle / (\|\mathbf{e}_k(k+1;k-1)\| \cdot \|\mathbf{e}_k^*(1;k-1)\|), \quad (8)$$

pour $k < m/2$. Ainsi défini $\beta_e(\cdot)$ est presque sûrement à valeurs dans $]-1,+1[$. A notre connaissance cet estimateur ne figure pas dans la littérature statistique. Mais c'est également le cas de certaines techniques d'estimation utilisées en traitement du signal. En particulier nous verrons que "l'estimation autorégressive basée sur des rapports d'énergies résiduelles", proposée par Dickinson (1978), constitue une approximation de $\beta_e(\cdot)$.

Notons que $\beta_e(\cdot)$ est réversible au sens où il reste invariant par l'inversion du temps: $\{x_1, \dots, x_m\} \rightarrow \{x_m, \dots, x_1\}$. Cependant la structure, dans \mathbb{R}^{m-k} , de la séquence $\{\underline{x}_k(1), \dots, \underline{x}_k(k+1)\}$ ne l'est pas. C'est pourquoi $\beta_e(k)$ n'est pas le coefficient de $\underline{x}_k(1)$ dans la régression de $\underline{x}_k(k+1)$ sur $\{\underline{x}_k(1), \dots, \underline{x}_k(k)\}$,

$$\underline{x}_k(k+1) = -\sum b_k(j)\underline{x}_k(k+1-j),$$

qui, selon (8), est égal à

$$-b_k(k) = \beta_e(k)\|\underline{\varepsilon}_k(k+1;k-1)\| / \|\underline{\varepsilon}_k^*(1;k-1)\|,$$

bien que la relation

$$\|\underline{\varepsilon}_k(k+1;k)\|^2 = \{1-\beta_e^2(k)\}\|\underline{\varepsilon}_k(k+1;k-1)\|^2$$

soit conservée. Cela équivaut à remarquer que la matrice de covariance empirique associée à $\{\underline{x}_k(1), \dots, \underline{x}_k(k+1)\}$,

$$X_k = \underline{X}_k \underline{X}_k' / (m-k), \quad \underline{X}_k = [\underline{x}_k(1), \dots, \underline{x}_k(k+1)]',$$

n'est pas doublement symétrique, car nous verrons que $\beta_e(k)$ est aussi lié à l'identification des paramètres $\alpha_k(\cdot)$ ou $\gamma_k(\cdot)$ de X_k (cf. § A.II.3). Remplacer X_k par $X_{ks} = \{X_k + JX_k J\}/2$ conduit à une estimation différente qui apparaît comme une corrélation partielle calculée dans $\mathbb{R}^{2(m-k)}$. Pour cela considérons la matrice \underline{Z}_k d'ordre $(k+1) \times 2(m-k)$ dont les colonnes regroupent l'ensemble des séquences de longueur $(k+1)$ qui figurent dans $\{x_1, \dots, x_m\}$, indépendamment du sens du temps: $\underline{Z}_k = [\underline{X}_k, J\underline{X}_k J]$. Les lignes $\underline{z}_k(j) = [\underline{x}_k(j)', \underline{x}_k(k-j+2)'J]'$, $j = 1, \dots, k+1$, forment, dans $\mathbb{R}^{2(m-k)}$, une séquence $\{\underline{z}_k(1), \dots, \underline{z}_k(k+1)\}$ qui a la même structure que la séquence inverse $\{\underline{z}_k(k+1), \dots, \underline{z}_k(1)\}$. Avec les notations introduites pour (8), remplaçant ε_k par ε_{ks} pour préciser qu'il s'agit des vecteurs $\underline{z}_k(\cdot)$, on définit l'autocorrélation partielle empirique symétrique d'ordre k par

$$\beta_{es}(k) = \langle \underline{\varepsilon}_{k_s}(k+1;k-1), \underline{\varepsilon}_{k_s}^*(1;k-1) \rangle / (\|\underline{\varepsilon}_{k_s}(k+1;k-1)\| \|\underline{\varepsilon}_{k_s}^*(1;k-1)\|), \quad (9)$$

pour $k < 2m/3$. La réversibilité montre que $\|\underline{\varepsilon}_{k_s}(k+1;k-1)\| = \|\underline{\varepsilon}_{k_s}^*(1;k-1)\|$ et ainsi $\beta_{es}(k)$ est le coefficient de $z_k(1)$ dans la régression de $z_k(k+1)$ sur $\{z_k(1), \dots, z_k(k)\}$. Plus précisément on a :

$$\underline{\varepsilon}_{k_s}(k+1;k) = \underline{\varepsilon}_{k_s}(k+1;k-1) - \beta_{es}(k) \underline{\varepsilon}_{k_s}^*(1;k-1),$$

$$\underline{\varepsilon}_{k_s}^*(1;k) = \underline{\varepsilon}_{k_s}^*(1;k-1) - \beta_{es}(k) \underline{\varepsilon}_{k_s}(k+1;k-1).$$

Cette récurrence s'arrête là, mais elle montre aussi que les "filtres" donnant $\underline{\varepsilon}_{k_s}$ et $\underline{\varepsilon}_{k_s}^*$ sont identiques :

$$\underline{\varepsilon}_{k_s}(k+1;k) = \sum b_{k_s}(j) z_k(k+1-j), \quad \underline{\varepsilon}_{k_s}^*(1;k) = \sum b_{k_s}(j) z_k(j), \quad b_{k_s}(k) = -\beta_{es}(k).$$

Ce filtre n'est pas nécessairement stable; il est la solution du problème des moindres carrés

$$\text{Min}_{b(\cdot)} \sum_{t=k+1}^m \{\varepsilon^2(t;k) + \varepsilon^{*2}(t-k;k)\}, \quad \varepsilon(t;k) = \sum_{j=0}^k b(j) x_{t-j}, \quad \varepsilon^*(t;k) = \sum_{j=0}^k b(j) x_{t+j}. \quad (10)$$

L'estimation $\beta_{es}(\cdot)$ peut donc se caractériser à l'aide des minimisations successives ci-dessus pour $k < 2m/3$. En ce sens elle s'apparente à l'estimation proposée par Burg (cf. Ch. IV). Ici, la seule contrainte est que, à chaque minimisation, les filtres progressifs et rétrogrades sont identiques. Dans la méthode de Burg on impose en plus que le filtre d'ordre $(k+1)$ soit obtenu, à partir de celui d'ordre k , par la récurrence habituelle, ce qui implique sa stabilité. Résoudre (10) en imposant la stabilité est une situation intermédiaire qui n'a pas été envisagée.

Le calcul de $\beta_{es}(k)$ se fait à l'aide de la décomposition de Cholesky de la matrice $PX_{k_s}P'$ où X_{k_s} est la covariance empirique de $\{z_k(1), \dots, z_k(k+1)\}$ et P l'opérateur de permutation circulaire défini en A.II.3.e. La matrice $PX_{k_s}P'$ est la covariance empirique de $\{z_k(k+1), z_k(1), \dots, z_k(k)\}$. La matrice triangulaire inférieure de sa décomposition de Cholesky a pour dernière ligne les coefficients $\gamma_k(1), \dots, \gamma_k(k)$, $\sigma_k(k)$ de la décomposition orthogonale (cf. Proposition A.II.3.3)

$$z_k(k+1) = \sum \gamma_k(j) \eta_{k_s}^*(k+1-j; j-1) + \sigma_k(k) \eta_{k_s}(k+1;k),$$

où $\eta_{k_s}(\cdot; \cdot)$ et $\eta_{k_s}^*(\cdot; \cdot)$ désignent les erreurs de prédiction $\underline{\varepsilon}_{k_s}(\cdot; \cdot)$ et $\underline{\varepsilon}_{k_s}^*(\cdot; \cdot)$

après normalisation. On a alors

$$\beta_{es}(k) = \gamma_k(k) \{ \gamma_k^2(k) + \sigma_k^2(k) \}^{-1/2}.$$

Lorsque l'on désire estimer que les p premières valeurs de $\beta(\cdot)$, et que p est petit devant m , on considère, comme approximation de $\beta_{es}(k)$ pour $k < p$, la "corrélation partielle" entre $z_p(p+1)$ et $z_p(p+1-k)$ dans l'ensemble $\{z_p(p+1), \dots, z_p(1)\}$. La décomposition de Cholesky de PA_pP' permet donc, par l'identification du paramètre $\gamma_p(\cdot)$, d'obtenir cette estimation. On sait que l'identification de $\alpha_p(\cdot)$ (Proposition A.II.3.1) conduit au même résultat lorsque la matrice de covariance est doublement symétrique. L'algorithme proposé par Dickinson (1978) équivaut à identifier $\alpha_p(\cdot)$ à partir de la matrice X_p : le coefficient $\beta(k)$ est ainsi la corrélation partielle entre $x_p(k+1)$ et $x_p(1)$ dans l'ensemble $\{x_p(k+1), \dots, x_p(1)\}$. Le nom de la méthode ("rapports d'énergies résiduelles") provient de ce que $\beta(k)$ est défini à partir de la relation

$$\|\varepsilon_p(k+1;k)\|^2 = [1-\beta^2(k)] \|\varepsilon_p(k+1;k-1)\|^2,$$

le signe étant l'opposé de celui du coefficient de $x_p(1)$ dans l'expression de $\varepsilon_p(k+1;k)$, où les erreurs $\varepsilon_p(k+1;k)$ représentent les solutions aux problèmes de moindres carrés analogues à (10) considérés uniquement dans le sens progressif :

$$\text{Min}_{b(\cdot)} \sum_{t=p+1}^m \varepsilon^2(t;k), \quad \varepsilon(t;k) = \sum_{j=0}^k b(j)x_{t-j}, \quad k=1, \dots, p.$$

Dickinson (1979) étend cette méthode au cas vectoriel (voir aussi Morf et coll., 1978) et constate alors l'analogie, dans le cas scalaire, avec la présentation d'Atal (1977) où $\beta(k)$ est défini comme étant la corrélation partielle entre $x_p(k+1)$ et $x_p(1)$. Il mentionne également la version symétrisée, approximation de $\beta_{es}(k)$, obtenue en effectuant la minimisation simultanément dans le sens progressif et rétrograde.

Quelques simulations nous ont permis de constater que $\beta_{es}(\cdot)$ est préférable à $\beta_e(\cdot)$. Cela peut s'expliquer par le fait de doubler le nombre de séquences. Le point de vue suivant est peut-être plus significatif. Associons à la séquence $\underline{x} = (x_1, \dots, x_m)'$,

considérée dans \mathbb{R}^m , les séquences symétrique et antisymétrique

$${}_s\mathbf{x} = (\mathbf{x} + J\mathbf{x})/2, \quad {}_a\mathbf{x} = (\mathbf{x} - J\mathbf{x})/2.$$

Ces deux statistiques sont non corrélées et l'estimation $\gamma_p(\cdot)$, calculée sur X_{ps} , est la somme de celles obtenues de la même façon à partir des séquences ${}_s\mathbf{x}$ et ${}_a\mathbf{x}$. Chacune est l'estimation de son homologue dans une décomposition analogue du paramètre $\gamma(\cdot)$. Le même constat s'observe sur le paramètre $\alpha(\cdot)$.

D'autres méthodes peuvent être opposées à l'estimation empirique $\Lambda_e(\cdot)$ et seront présentées dans le chapitre suivant. On s'attend à ce qu'elles soient plus performantes lorsque la longueur de la séquence m est faible car le biais de $\Lambda_e(\cdot)$ est alors très important. Nous illustrons ici un autre inconvénient de $\Lambda_e(\cdot)$. Lorsque l'on estime la structure jusqu'à l'ordre p de la séquence et que m est grand devant p , toutes les méthodes donnent des résultats sensiblement identiques pour l'estimation de la matrice d'autocovariance Λ d'ordre $(p+1)$. Cependant, si Λ est proche de la singularité du fait que, pour $k < p$, la valeur de $|\beta(k)|$ est proche de 1, elle est peu sensible aux valeurs de $\beta(j)$, $j > k$. Mais ces valeurs sont très mal estimées par les relations de Yule-Walker appliquées à $\Lambda_e(\cdot)$ alors que les autres méthodes donnent toutes de bons résultats. Ces différences se retrouvent sur tous les paramètres sensibles à $\beta(\cdot)$, en particulier sur les coefficients $b(\cdot)$.

La Table I ci-dessous compare la méthode de Yule-Walker appliquée à $\Lambda_e(\cdot)$ avec l'autocorrélation partielle empirique $\beta_{es}(\cdot)$ (ACPE). Le modèle (et le fichier de données) est celui déjà cité au chapitre précédent, il est rappelé dans la table. Les estimations de $\beta(\cdot)$, $\rho(\cdot)$ (autocorrélations) et $b(\cdot)$ sont obtenues sur 25 répétitions d'une séquence de longueur 1000. Dans cet exemple, les autres méthodes (Burg, maximum de vraisemblance exact ou récursif, moindres carrés symétriques) diffèrent des résultats d'ACPE, Moyenne ou $(EQM)^{1/2}$, d'au plus $1/1000$.

TABLE I
MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

Paramètres	Vraie Valeur	YULE-WALKER		ACPE	
		Moyenne	(EQM) ^{1/2}	Moyenne	(EQM) ^{1/2}
$\beta(1)$	0.716	0.717	0.010	0.718	0.010
$\beta(2)$	-0.982	-0.978	0.005	-0.982	0.001
$\beta(3)$	0.704	0.475	0.261	0.700	0.010
$\beta(4)$	-0.924	-0.333	0.636	-0.924	0.012
$\rho(1)$	0.716	0.717	0.010	0.718	0.010
$\rho(2)$	0.034	0.040	0.028	0.040	0.028
$\rho(3)$	-0.643	-0.634	0.029	-0.636	0.029
$\rho(4)$	-0.928	-0.924	0.006	-0.928	0.004
$b(1)$	-2.761	-2.070	0.751	-2.759	0.012
$b(2)$	3.812	2.224	1.695	3.809	0.026
$b(3)$	-2.654	-1.131	1.643	-2.652	0.026
$b(4)$	0.924	0.333	0.636	0.924	0.012



CHAPITRE IV

ESTIMATION AUTOREGRESSIVE DANS LE DOMAINE TEMPOREL

L'étude de la structure de covariance du processus $X(\cdot)$ par l'ajustement d'un modèle autorégressif est proposée par J.P. Burg (1967) et souvent référencée sous le nom de méthode d'entropie maximale. Son intérêt dans le domaine spectral est souligné par E. Parzen (1974). De nombreux auteurs ont alors comparé les mérites de cette méthode par rapport aux techniques spectrales classiques (utilisation de fenêtres) ou encore ceux des différentes estimations possibles des paramètres du modèle autorégressif sous-jacent. La comparaison porte essentiellement sur le pouvoir de résolution et sur la précision de l'énergie estimée. Nous les comparons ici dans le domaine temporel sur la base des autocorrélations $\rho(\cdot)$, des autocorrélations partielles $\beta(\cdot)$ et des paramètres $b(\cdot)$ du modèle autorégressif. La variance $\sigma^2(0)$ du processus et celle, $\sigma^2(p)$, de l'innovation du modèle $AR(p)$ sont liées par la relation

$$\sigma^2(p) = \sigma^2(0) \prod_{k=1}^p [1 - \beta^2(k)]. \quad (1)$$

Le plus souvent la méthode conduit à estimer $\sigma^2(p)$ et la correspondance (1) peut donner des résultats médiocres pour $\sigma^2(0)$, comme dans le cas des moindres carrés symétriques.

Nous présentons les méthodes dans le cadre d'une seule séquence au cours du premier paragraphe et les comparons par simulation dans le suivant. L'extension au cas d'un échantillon fait l'objet du dernier paragraphe.

§ IV.1. Présentation des méthodes

Il s'agit d'estimer les paramètres du modèle $AR(p)$ sous-jacent à $X(\cdot)$ (cf. Ch. A.III) à partir de l'observation d'une séquence x_1, \dots, x_m , de longueur m ($m > 3p/2$),

sans hypothèse a priori sur la structure réelle de $X(\cdot)$. Afin de ne pas augmenter démesurément le nombre de méthodes, nous ne considérons que celles qui restent invariantes par inversion du temps. On peut les classer en deux groupes. Les méthodes récursives ajustent un modèle $AR(k+1)$ à partir du modèle $AR(k)$ précédemment estimé, le coefficient supplémentaire pouvant dépendre explicitement de ce modèle (Burg, Maximum de Vraisemblance Récursif) ou non (Yule-Walker, Autocorrélation Partielle Empirique). Les méthodes globales, dans lesquelles le modèle $AR(p-1)$ sous-jacent à la solution d'ordre p n'est donc pas celui obtenu à l'ordre $p-1$, déterminent la solution de manière itérative (Maximum de Vraisemblance Exact, Pseudo Maximum de Vraisemblance) ou non (Moindres Carrés Symétriques, Rapports d'Energies Résiduelles).

A) Méthodes récursives

a) YULE-WALKER (Y-W)

La plus couramment utilisée, elle consiste à retenir le modèle $AR(p)$ associé aux autocovariances empiriques

$$\Lambda_e(k) = \frac{1}{m} \sum_{i=1}^{m-k} x_i x_{i+k}, \quad k=0, \dots, p.$$

Les autocorrélations partielles $\beta(\cdot)$ ainsi que les coefficients $b(\cdot)$ du modèle sont obtenus en résolvant les équations de Yule-Walker par l'algorithme de Levinson-Durbin et $\sigma^2(0)$ est estimé par $\Lambda_e(0)$.

b) AUTOCORRELATION PARTIELLE EMPIRIQUE (ACPE)

Bien que $\beta_e(\cdot)$, défini par III (8) au chapitre précédent, soit invariant par inversion du temps, nous retenons ici la version symétrisée $\beta_{es}(\cdot)$ donnée en III (9). Il n'est pas nécessaire de conserver le facteur $1/2(m-k)$ dans le calcul de X_{ks} . On utilise la matrice A_k ,

$$A_k(i,j) = \sum_{i=1}^{m-k} [x_{i+i} x_{i+j} + x_{i+k-i} x_{i+k-j}], \quad i=0, \dots, k, \quad j=0, \dots, k, \quad (2)$$

qui est caractérisée par ses valeurs en (i, j) avec $j = 0, \dots, [k/2]$, $j \leq i \leq k-j$. Elle se construit par une récurrence sur k qui prend en compte la double symétrie : les valeurs de $A_k(k, 0)$ sont données par l'expression ci-dessus et sont égales à $2m\Lambda_e(k)$, les autres sont obtenues par les relations

$$A_k(k-j, j) = A_{k-1}(k-1-j, j-1) - x_j x_{k-j} - x_{m-j+1} x_{m+j-k+1}, \quad j = 1, \dots, [k/2],$$

$$A_k(i, j) = A_{k-1}(i, j) - x_{k-i} x_{k-j} - x_{m-k+1-i} x_{m-k+1-j}, \quad j=0, \dots, [k/2], j \leq i \leq k-j-1.$$

Le calcul de $\beta_{es}(\cdot)$ nécessite donc de réaliser les décompositions de Cholesky des matrices $PA_k P'$ pour $k = 1, \dots, p$. La variance du processus $\sigma^2(0)$ est estimée par $\Lambda_e(0)$. On peut aussi envisager d'estimer la variance $\sigma^2(p)$ par le minimum (divisé par $2(m-p)$) atteint par III (10) à l'ordre p et qui apparaît dans la décomposition de $PA_p P'$.

c) BURG

J.P. Burg (1968, 1975) propose d'estimer les coefficients $\beta(k)$, $k = 1, \dots, p$, en minimisant, de manière récursive, la somme des carrés des erreurs de prédiction progressive et rétrograde. Pour $k = 1, \dots, p$, la valeur de $\beta_b(k)$ est celle qui réalise le minimum dans

$$\text{Min}_{\beta(k)} \sum_{t=k+1}^m \{ [\varepsilon_b(t; k-1) - \beta(k)\varepsilon_b^*(t-k; k-1)]^2 + [\varepsilon_b^*(t-k; k-1) - \beta(k)\varepsilon_b(t; k-1)]^2 \}, \quad (3)$$

où les erreurs $\varepsilon_b(t; k-1)$, $\varepsilon_b^*(t-k; k-1)$, $t = k+1, \dots, m$, sont calculées en fonction des coefficients $\beta_b(j)$, $j = 1, \dots, k-1$, déjà estimés. Ainsi $\beta_b(k)$ est donné par :

$$\beta_b(k) = 2 \sum_{t=k+1}^m \varepsilon_b(t; k-1)\varepsilon_b^*(t-k; k-1) / \left\{ \sum_{t=k+1}^m [\varepsilon_b^*(t-k; k-1) + \varepsilon_b(t; k-1)] \right\}.$$

Notons $\underline{b}_k = [b(k-1, 0), \dots, b(k-1, k-1)]'$ le vecteur de \mathbb{R}^k formé par les coefficients de régression à l'ordre $(k-1)$. Le problème (3) s'écrit, en utilisant la matrice A_k introduite en (2),

$$\text{Min}_{\beta(k)} [\tilde{b}_k - \beta(k)\tilde{J}\tilde{b}_k]' A_k [\tilde{b}_k - \beta(k)\tilde{J}\tilde{b}_k],$$

où \tilde{b}_k est le vecteur de \mathbb{R}^{k+1} défini par $\tilde{b}_k = [\underline{b}_k', 0]'$. Par suite $\beta_b(k)$ s'exprime sous la forme

$$\beta_b(k) = \tilde{b}_k' A_k \tilde{J} \tilde{b}_k / \tilde{b}_k' A_k \tilde{b}_k. \quad (4)$$

Cela conduit à une nouvelle mise en œuvre de la méthode de Burg. Les matrices A_k sont construites selon la récurrence présentée à la suite de (2). Le numérateur de (4) se calcule en fonction de A_k et \underline{b}_k , le dénominateur satisfait la récurrence notée par Andersen (1974) :

$$\tilde{b}_k' A_k \tilde{b}_k = [1 - \beta_b^2(k-1)] \tilde{b}_{k-1}' A_{k-1} \tilde{b}_{k-1} - \varepsilon_b^2(k; k-1) - \varepsilon_b^{*2}(m-k+1; k-1),$$

et, pour l'étape suivante, on a $\underline{b}_{k+1} = \tilde{b}_k - \beta_b(k) J \tilde{b}_k$. Cette formulation améliore, par l'usage de la double symétrie, celle proposée par Makhoul (1977). Il est aussi intéressant de noter qu'en remplaçant les matrices A_k par les matrices d'autocovariance d'ordre $(k+1)$, $\Lambda_{k+1} = E\{A_k/2(m-k)\}$, $k = 1, \dots, p$, on construit de cette façon la suite des autocorrélations partielles théoriques. Comme dans ACPE, la variance du processus $\sigma^2(0)$ est estimée par $\Lambda_e(0)$. Mais ici c'est aussi l'estimation de maximum de vraisemblance sous l'hypothèse AR(p) définie par $\beta_b(k)$, $k = 1, \dots, p$. En effet, si Γ_{pb} désigne la matrice d'autocorrélation d'ordre m associée à cette hypothèse, on a

$$\frac{1}{m} x' \Gamma_{pb}^{-1} x = \frac{1}{m} x' x = \Lambda_e(0), \quad x = (x_1, \dots, x_m)'$$

Ce dernier point est une conséquence des relations suivantes : pour $k = 1, \dots, m-1$,

$$x' \Gamma_k^{-1} x = x' \Gamma_{k-1}^{-1} x + \frac{\beta(k)}{2 \sigma(k)} \left\{ \beta(k) \sum_{j=k}^{m-1} [\varepsilon^2(j; k-1) + \varepsilon^{*2}(j-k; k-1)] - 2 \sum_{j=k}^{m-1} \varepsilon(j; k-1) \varepsilon^*(j-k; k-1) \right\},$$

où tous les éléments sont définis par la suite $\beta(1), \dots, \beta(m-1)$ d'une matrice d'autocorrélation Γ d'ordre m non singulière quelconque. Ainsi, pour $k = 1, \dots, m-1$, l'égalité

$$x' \Gamma_k^{-1} x = x' \Gamma_{k-1}^{-1} x, \quad (5)$$

n'est réalisée que par $\beta(k)$ nul ou obtenu, en fonction de x et de Γ_{k-1} , par l'expression de $\beta_b(k)$. Cette deuxième solution donne une vraisemblance plus grande que la première. En ce sens la méthode de Burg apparaît comme une maximisation récursive de la vraisemblance par rapport aux valeurs successives de $\beta(\cdot)$ avec, à chaque étape, la contrainte supplémentaire (5). Là encore on reconnaît la version empirique de la propriété suivante. La matrice d'autocovariance $\Lambda(X)$ d'une séquence aléatoire stationnaire $X = (X_1, \dots, X_m)'$ est l'unique matrice $\Lambda = \sigma^2(0)\Gamma$ pour laquelle les matrices Γ_k , $k = 0$,

..., $m-1$, réalisent successivement le minimum de (cf. Proposition A.III.3.2)

$$|\Gamma_k|^{1/m} E\left\{\frac{1}{m} X' \Gamma_k^{-1} X\right\},$$

et la contrainte

$$E\left\{\frac{1}{m} X' \Gamma_k^{-1} X\right\} = E\left\{\frac{1}{m} X' \Gamma_{k-1}^{-1} X\right\} = \sigma^2(0),$$

est alors satisfaite. Dans la version empirique, la minimisation sans la contrainte (5) conduit à l'estimation de maximum de vraisemblance, lorsqu'elle existe, mais les matrices Γ_k , $k = 1, \dots, p$, ne correspondent pas à une même matrice Γ . En d'autres termes, l'estimation de maximum de vraisemblance des modèles AR(k), $k \leq p$, n'est pas récursive. Notons enfin que l'estimation de $\sigma^2(p)$ par

$$\frac{m}{m-p} \Lambda_e(0) \prod_{k=1}^p [1 - \beta_b^2(k)],$$

où l'on tient compte de l'estimation de Γ_p dans celle de $\sigma^2(0)$ par le facteur $m/(m-p)$, est plus grande que celle qui résulte du minimum atteint en (3). Cela se vérifie à partir de la récurrence satisfaite par le dénominateur de (4).

d) MAXIMUM DE VRAISEMBLANCE RECURSIF (MVR)

S.M. Kay (1983) propose de maximiser la vraisemblance du modèle AR(p) de manière récursive par rapport aux autocorrélations partielles. La matrice d'autocovariance d'ordre m étant paramétrée par $\Lambda = \sigma^2(p)R_p$, la maximisation relative à $\sigma^2(p)$ est réalisée par

$$\sigma^2(p) = \frac{1}{m} x' R_p^{-1} x, \quad x = (x_1, \dots, x_m)'. \quad (6)$$

Maximiser la vraisemblance équivaut alors à minimiser, par rapport à R_p , la fonction

$$L(x; R_p) = |R_p|^{1/m} \frac{1}{m} x' R_p^{-1} x.$$

Cette fonction ne dépend que de $\beta(k)$, $k = 1, \dots, p$, et Kay propose d'approcher la solution en effectuant une minimisation récursive sur l'ordre k du paramètre $\beta(\cdot)$. Pour cela on dispose des récurrences suivantes :

$$|R_k| = |R_{k-1}|[1-\beta^2(k)]^{-k},$$

$$x'R_k^{-1}x = [1-\beta^2(k)]x'R_{k-1}^{-1}x + e(k-1)\beta^2(k) - 2f(k-1)\beta(k),$$

où

$$e(k-1) = \sum_{t=k+1}^m [\varepsilon^2(t;k-1) + \varepsilon^{*2}(t-k;k-1)], \quad f(k-1) = \sum_{t=k+1}^m \varepsilon(t;k-1)\varepsilon^*(t-k;k-1).$$

la valeur de $\beta(k)$ est alors solution de l'équation du troisième degré

$$\beta^3(k) + p\beta^2(k) + q\beta(k) + r = 0, \quad (7)$$

avec

$$p = -\frac{m-2k}{m-k} \frac{f(k-1)}{d(k-1)}, \quad d(k-1) = e(k-1) - x'R_{k-1}^{-1}x,$$

$$q = -\left\{1 + \frac{k}{m-k} \frac{e(k-1)}{d(k-1)}\right\}, \quad r = \frac{m}{m-k} \frac{f(k-1)}{d(k-1)}.$$

Lorsque $d(k-1)$ est positif, l'équation (7) a trois racines réelles dont une seule appartient à $] -1, 1[$. Elle est donnée par

$$\beta(k) = 2a^{1/2} \cos(\theta + 4\pi/3) - p/3,$$

$$\text{où } 3a = p^2/3 - q, \quad 2b = -r + pq/3 - 2p^3/27 \text{ et } \cos 3\theta = b/a^{3/2}.$$

Lorsque $d(k-1)$ est négatif, Kay conjecture que (7) admet une seule racine réelle,

$$\beta(k) = \{b + [b^2 - a^3]^{1/2}\}^{1/3} + \{b - [b^2 - a^3]^{1/2}\}^{1/3} - p/3.$$

Nous avons cependant observé quelques cas (très rares) ne vérifiant pas cette conjecture.

Il est donc nécessaire, en principe, de rechercher la racine qui convient lorsque cette situation se produit (cf. MVE).

L'estimation de la variance $\sigma^2(p)$ est donnée par (6) et celle de $\sigma^2(0)$ qui s'en déduit selon (1) équivaut à utiliser la relation

$$\sigma^2(0) = \frac{1}{m} x' \Gamma_p^{-1} x,$$

puisque $\Gamma_p = \sigma^2(p)R_p / \sigma^2(0)$.

B) Méthodes globales

a) MOINDRES CARRÉS SYMÉTRIQUES (MCS)

La méthode des moindres carrés, sous sa forme symétrisée, est proposée par A.H. Nutall (1976) et indépendamment par T.J. Ulrych and R.W. Clayton (1976).

Marple (1980) indique un nouvel algorithme, analogue à celui donné par Morf et al. (1977) dans le cas des moindres carrés habituels, qui construit la solution de l'équation normale de manière récursive, évitant ainsi la décomposition de Cholesky de la matrice A_p . Notons que c'est la seule méthode, parmi celles présentées ici, ne garantissant pas la stabilité du filtre autorégressif. Elle conserve cependant tout son intérêt dans le domaine spectral. Rappelons que la méthode consiste à minimiser la somme des carrés des erreurs de prédiction progressive et rétrograde d'ordre p , c'est-à-dire le problème III (10) avec $k = p$. Soit TD^2T' la décomposition de Cholesky de PA_pP' , dans laquelle T est triangulaire inférieure unitaire ($\text{diag}(T) = I$) et D^2 diagonale. La variance $\sigma^2(p)$ est estimée par le minimum, divisé par $2(m-p)$, atteint par III (10). Elle est donnée par le dernier élément de D^2 . Les coefficients de régression, ou plus exactement ceux du filtre d'ordre p , $\{b(1), \dots, b(p), 1\}$, forment la dernière ligne de T^{-1} .

b) RAPPORTS D'ENERGIES RESIDUELLES (RER)

La méthode de Dickinson (1978, 1979), sous sa forme symétrisée, est une approximation d'ACPE consistant à remplacer $\beta_{es}(k)$, calculée dans $\mathbb{R}^{2(m-k)}$, par la corrélation partielle, calculée dans $\mathbb{R}^{2(m-p)}$, entre $z_p(1)$ et $z_p(k+1)$ dans l'ensemble $\{z_p(1), \dots, z_p(k+1)\}$ (cf. § III.3). Elle équivaut à identifier le paramètre $\gamma(\cdot)$ avec les coefficients de la régression orthogonale de $z_p(p+1)$ sur $z_p(p), \dots, z_p(1)$ donnés par la dernière ligne de TD ci-dessus. Elle s'apparente ainsi à la méthode précédente, mais l'orthogonalisation préalable des "variables explicatives" garantit la stabilité du polynôme de prédiction. La variance $\sigma^2(p)$ est estimée par la même valeur et $\beta(\cdot)$ est obtenu en fonction de $\gamma(\cdot)$ (cf. Proposition A.II.3.3). Notons que l'estimation de $\sigma^2(0)$ qui résulte de (1) est sans biais puisqu'elle est égale à $\|z_p(p+1)\|^2$, soit

$$\frac{1}{2(m-p)} \sum_{t=1}^{m-p} [x_t^2 + x_{m-p-t+1}^2],$$

alors que ce n'est pas le cas dans MCS si l'on procède de la même façon.

c) PSEUDO MAXIMUM DE VRAISEMBLANCE (PMV)

Les deux méthodes précédentes ne sont fonction que de la matrice de covariance empirique X_{ps} . Elles constituent, en ce sens, deux techniques d'approximation d'une matrice de covariance par une matrice de Toeplitz. Bien que X_{ps} ne provienne pas d'un échantillon, le critère

$$\underset{\Lambda}{\text{Min}} \{ \ln |\Lambda| + \text{tr}(\Lambda^{-1} X_{ps}) \},$$

conduit à une troisième possibilité. La solution est obtenue par la méthode de relaxation du chapitre II (voir aussi §3 ci-après).

d) MAXIMUM DE VRAISEMBLANCE EXACT (MVE)

Nous désignerons ainsi l'estimation de maximum de vraisemblance d'un modèle AR(p) dans le cas gaussien. D.T.Pham (1987) donne un algorithme de calcul de cet estimateur dans le cas plus général des modèles ARMA et souligne la simplicité de sa mise en œuvre dans le cas des modèles autorégressifs. Il s'agit d'une méthode de gradient dans laquelle le gradient est calculé de façon exacte et transformé, afin d'accélérer la convergence (Newton-Rhapson), à l'aide d'une approximation de la matrice des dérivées secondes. Le paramètre utilisé est constitué des coefficients $b(1), \dots, b(p)$ du modèle et de la variance $\sigma^2(p)$ de l'innovation. Afin de garantir la stabilité du polynôme nous subdivisons le pas obtenu autant de fois que cela est nécessaire. Cependant, dans les situations proches de l'instabilité, la transformation du gradient donne parfois une direction contraire à l'effet recherché. De plus l'utilisation dans ce cas du paramètre $b(\cdot)$ est mal adaptée aux méthodes de gradient. Nous proposons, avec Pham, un autre algorithme basé sur le paramètre $\beta(\cdot)$. Avec les notations introduites pour MVR, $\sigma^2(p)$ est estimé selon (6) où R_p réalise le minimum de

$$\frac{1}{m} \ln |R_p| + \ln \left[\frac{1}{m} x' R_p^{-1} x \right].$$

Le gradient de $\ln |R_p|$ est donné par

$$\frac{\partial \ln |R_p|}{\partial \beta(k)} = 2k\beta(k)[1-\beta^2(k)]^{-1}, \quad k=1, \dots, p.$$

La forme quadratique $x'R_p^{-1}x$ se compose de deux parties

$$x'R_p^{-1}x = \sum_{t=0}^{p-1} \prod_{k=t+1}^p [1-\beta^2(k)] \varepsilon^2(t+1;t) + \sum_{t=p+1}^m \varepsilon^2(t;p) = E(p) + b'Ab, \quad (8)$$

où $b = [b(p,0), \dots, b(p,p)]'$ et où la matrice A définie par

$$A(i,j) = \sum_{t=p+1}^m x_{t-i} x_{t-j}, \quad i,j = 0, \dots, p,$$

peut être obtenue selon une récurrence analogue à celle donnant A_p . On peut aussi utiliser A_p en symétrisant $E(p)$ par rapport au temps, mais cela ne modifie pas les résultats.

Le gradient de $b'Ab$ par rapport à $b(\cdot)$ est immédiat puisque $b(k) = b(p,k)$, $k = 1, \dots, p$. Celui de $E(p)$ est construit par récurrence. Avec la convention $\varepsilon(p+1;p)=0$ on a

$$E(0) = \varepsilon^2(1;0), \quad E(k) = [1-\beta^2(k)]E(k-1) + \varepsilon^2(k+1;k), \quad k = 1, \dots, p$$

et les formules de dérivation pour $k = 1, \dots, p$:

$$\frac{\partial E(k)}{\partial b(k,k)} = 2\beta(k)E(k-1) + 2\varepsilon(k+1;k)x_1 - \sum_{j=1}^{k-1} [b(k;k-j) + 2\beta(k)b(k-1,j)] \frac{\partial E(k-1)}{\partial b(k-1,j)},$$

$$\frac{\partial E(k)}{\partial b(k,j)} = \frac{\partial E(k-1)}{\partial b(k-1,j)} + \beta(k) \frac{\partial E(k-1)}{\partial b(k-1,k-j)} + 2\varepsilon(k+1;k)x_{k-j+1}, \quad j = 1, \dots, k-1.$$

Le gradient en $b(\cdot)$ permet d'obtenir celui en $\beta(\cdot)$ par changements de variables successifs. En effet le passage de $\{b(k,1), \dots, b(k,k), \beta(k+1), \dots, \beta(p)\}$ à $\{b(k-1,1), \dots, b(k-1,k-1), \beta(k), \dots, \beta(p)\}$ se traduit par

$$\frac{\partial}{\partial \beta(k)} = - \sum_{j=1}^k b(k-1,k-j) \frac{\partial}{\partial b(k,j)},$$

$$\frac{\partial}{\partial b(k-1,j)} = \frac{\partial}{\partial b(k,j)} - \beta(k) \frac{\partial}{\partial b(k,k-j)}, \quad j = 1, \dots, k-1.$$

Enfin les problèmes de bord pour les variables $\beta(k)$, $k = 1, \dots, p$ sont résolus par la transformation de Fisher :

$$z(\cdot) = \frac{1}{2} \ln \frac{1+\beta(\cdot)}{1-\beta(\cdot)}, \quad \frac{\partial}{\partial z(\cdot)} = [1-\beta^2(\cdot)] \frac{\partial}{\partial \beta(\cdot)}.$$

A chaque itération le pas est subdivisé jusqu'à ce que la fonction à minimiser diminue. Notons que l'on se déplace uniquement dans la direction du gradient. L'utilisation de méthodes plus rapides (Newton-Rhapson, gradient conjugué) se heurte, non pas à la difficulté du calcul des dérivées secondes, mais à celle de la non convexité de la fonction de log-vraisemblance (cf. Dégerine et Pham, 1987).

Une nouvelle approche est proposée par D.T.Pham (1988). Il s'agit d'une méthode de relaxation portant sur les autocorrélations partielles $\beta(k)$, $k = 1, \dots, p$, qui consiste à itérer la méthode de Kay (cf. MVR). L'expression (8) de la forme quadratique $x'R_p^{-1}x$ montre que celle-ci reste quadratique par rapport à $\beta(k)$ lorsque les $\beta(j)$, $j \neq k$, sont fixés. Ainsi $\beta(k)$ est obtenu comme l'une des solutions d'une équation semblable à (7). Pham précise le comportement des racines de cette équation en fonction des paramètres de la forme quadratique en $\beta(k)$ dont les coefficients sont eux-mêmes calculés en montrant que $x'R_p^{-1}x$ est une forme quadratique par rapport au paramètre $b(\cdot)$ du modèle AR(p).

§ IV.2. Comparaison des méthodes par simulation

Des simulations, effectuées sur plusieurs modèles et pour des séquences de longueur variable, sont présentées dans Dégerine (1986). Elles ne constituent pas une étude exhaustive de ces méthodes mais conduisent déjà à quelques conclusions générales que nous reportons ici en illustrant uniquement les points qui nous ont parus remarquables.

Les essais sont maintenant réalisés à l'aide d'un fichier de 60000 nombres pseudo-aléatoires construit à partir du générateur

$$y(i+1) = \{69,069 y(i) + 1\} \bmod(2^{32}), y(0) = 3579.$$

Aucune de ces méthodes ne prévaut en toute circonstance. Les facteurs déterminants sont la proximité du modèle à la singularité (on le dira "singulier"), la longueur de la séquence (par rapport à l'ordre du modèle estimé) et le type de paramètre étudié. Ainsi pour un modèle "non singulier" et une séquence de longueur suffisante ($m > 10p$), les méthodes sont très semblables sur l'ensemble des paramètres. Cela justifie l'utilisation de Yule-Walker dont la mise en œuvre est la plus simple.

L'exemple "singulier" MS1 que nous avons retenu est celui du modèle AR(4) cité au chapitre précédent. Ce modèle sert de référence à de nombreux auteurs dans la comparaison de méthodes en traitement du signal (Nuttal, 1976; Marple, 1980; Kay, 1983; Dégerine et Pham, 1987; Pham, 1988). La densité spectrale est composée de deux pics voisins. Ceci apparaît dans les valeurs de $\beta(\cdot)$ lorsqu'on les compare à celles de la description canonique des polynômes dont les racines sont sur le cercle unité (cf. § III.2). Les autres essais que nous avons effectués en ce sens ont été construits de la même façon. En particulier nous n'abordons pas ici le cas des modèles comportant une composante singulière pour lesquels l'estimation autorégressive n'est peut-être pas la mieux adaptée. La proximité à la singularité exclut les modèles MA d'ordre faible. Les modèles ARMA, "singuliers" par leur composante AR, devraient se comporter comme le modèle AR correspondant ou comme un modèle à composante singulière selon la forme de la partie MA. Notons que dans le cas "singulier", et pour une séquence longue ($m > 100p$), les méthodes restent équivalentes à l'exception de Yule-Walker (cf. Table I, Ch. III). Dans ce cas nous préconisons d'utiliser la méthode de Burg ou RER, qui sont comparables pour leur mise en œuvre, car elles conservent de bonnes propriétés pour des séries plus courtes (cf. Tables I, II et III ci-après).

1) YULE-WALKER

Le défaut dû au biais, pour une série courte, est bien connu et, lorsque l'écart quadratique est comparable à celui des autres méthodes, le biais est souvent plus important. Le point le plus remarquable est celui que nous avons déjà souligné à la fin du chapitre

précédent. Pour une série longue d'un modèle "singulier", la suite des autocorrélations est bien estimée alors que les autocorrélations partielles et les coefficients de régression le sont très mal.

2) AUTOCORRELATION PARTIELLE EMPIRIQUE

L'avantage d'ACPE sur son approximation RER n'est sensible que sur des séries très courtes. Notons qu'ACPE est basé sur la statistique exhaustive du modèle $AR(p)$:

$$T_x = \{(x_1, \dots, x_p), (x_m, \dots, x_{m-p+1}), \Lambda_e(0), \dots, \Lambda_e(p)\},$$

alors que RER n'est fonction que de la matrice de covariance empirique X_{ps} . Celle-ci est issue de T_x d'après la construction récursive de A_p . Elle n'en diffère que par une partie de l'information contenue dans les bords (x_1, \dots, x_p) et (x_m, \dots, x_{m-p+1}) de la séquence bien qu'elle ne soit pas équivalente à $\Lambda_e(0), \dots, \Lambda_e(p)$. Les résultats obtenus par ACPE sont très satisfaisants sur l'ensemble des paramètres et confirment l'intérêt des méthodes construites à partir des autocorrélations partielles. La variance $\sigma^2(p)$ est sous-estimée par les deux procédés. Mais le biais est plus important lorsqu'on utilise la valeur du minimum et conduit à une sous-estimation de $\sigma^2(0)$ alors que $\Lambda_e(0)$ est sans biais.

3) BURG

La formulation que nous en avons proposée, en fonction des matrices A_k , montre qu'elle s'exprime aussi en fonction de la statistique exhaustive T_x . Nous avons constaté les bonnes performances de cette méthode et la possibilité de la comparer ici avec celle du maximum de vraisemblance permet de confirmer ses qualités. En effet, basée sur une idée simple et facile à mettre en œuvre, elle réalise des résultats voisins de ceux de MVE sur l'ensemble des paramètres lorsque la série devient courte et le modèle "singulier" (cf. Tables I, II et III). Le point faible cependant est une surestimation excessive de $\sigma^2(p)$, qui s'accompagne d'un écart quadratique démesuré, lorsque la série est très courte (cf. Tables I et III). L'estimation de $\sigma^2(p)$ à partir du minimum ne change rien à ce constat. Ce phénomène est dû au fait que, dans le calcul récursif, $\beta_b(k)$ dépend des valeurs précédentes de $\beta_b(\cdot)$, ce qui n'est pas le cas dans ACPE. En effet ces deux méthodes, par

la relation (1), associent à la même estimation $\Lambda_e(0)$ de $\sigma^2(0)$ deux estimations de $\sigma^2(p)$ complètement différentes, alors que les propriétés statistiques des estimations de $\beta(\cdot)$ sont analogues. Ceci est d'ailleurs confirmé par MVR qui présente le même défaut.

4) MAXIMUM DE VRAISEMBLANCE RECURSIF

La méthode semble particulièrement bien adaptée à l'estimation des coefficients $b(\cdot)$ du modèle AR(4) "singulier" pour une série courte. Pour ce paramètre, Kay (1983) la compare avec Y-W et MCS, dans le cas du modèle MS1, pour $m = 11$ et $m = 50$. Nous confirmons la supériorité de la méthode pour l'estimation de $b(\cdot)$ ou de $\beta(\cdot)$, y compris par rapport au maximum de vraisemblance exact. Par contre MVE et MCS sont très nettement plus performantes pour l'estimation de $\rho(\cdot)$ et c'est également le cas, dans une moindre mesure, de Burg ou d'ACPE (cf. Tables I et II). Nous avons éliminé, dans les résultats, les cas pour lesquels la conjecture de Kay n'était pas satisfaite. Leur nombre n'excède pas 5/1000 et leur occurrence n'a lieu que pour des séries très courtes ($m = 10, 15$) mais reste possible pour un modèle "non singulier" (3/1000 pour un bruit blanc avec $m=10$). Nous avons constaté des performances semblables sur des modèles analogues : modifications de $\beta(1)$ et $\beta(3)$ dans l'exemple de Kay. Mais ces performances s'estompent lorsque l'on ajuste un modèle AR(p) avec $p > 4$. Par contre elles se conservent dans l'ajustement d'un AR(4) sur le modèle MS2 de la Table III. Ce modèle "singulier" est obtenu en "ajoutant" deux pics voisins supplémentaires à ceux déjà existant dans la densité spectrale de celui de Kay (cf. $\beta(\cdot)$). Il illustre bien le danger de conclusions hâtives. On retrouve ici la suprématie de MVE et MCS pour l'estimation de $\rho(\cdot)$. Par contre MVR et Burg, assez semblables, sont nettement moins performantes que les autres méthodes dans l'estimation de $\beta(\cdot)$ et de $b(\cdot)$.

5) MOINDRES CARRES SYMETRIQUES

Le nombre de polynômes instables, lorsque la série est courte et le modèle "singulier", est important : 701/2000, 138/1000 et 1601/2000 pour les exemples des Tables I, II et III. La stabilité n'est pas nécessaire pour une comparaison dans le domaine spectral. Elle l'est ici, tout au moins pour $\beta(\cdot)$ et $\rho(\cdot)$. Nous avons donc transformé

chaque polynôme instable en le polynôme stable de même module (sur le cercle unité) selon la procédure de Wilson (1969). La normalisation $b(0) = 1$ conduit à diminuer le minimum atteint par III (10) pour l'estimation de $\sigma^2(p)$ qui était déjà sous-estimé. Malgré cela l'estimation de $\sigma^2(0)$ qui résulte de (1) est médiocre. Ceci provient de l'absence de contrainte sur $b(\cdot)$, ainsi les valeurs correspondantes de $\beta(\cdot)$ peuvent être très proches de ± 1 . On retrouve le même phénomène (dans le sens inverse) d'incohérence entre les estimations des variances que celui constaté plus haut pour Burg et MVR. En effet par ailleurs la méthode réalise de bonnes performances, avec un point fort dans l'estimation de $\rho(\cdot)$, bien qu'elle ne soit fonction que de la covariance empirique X_{ps} . La transformation des polynômes instables ne semble pas modifier les propriétés statistiques des estimations car les résultats obtenus ici pour MS1 sont analogues à ceux que nous avons eus sans elle (cf. Dégerine, 1986).

6) RAPPORTS D'ÉNERGIES RESIDUELLES

Elle constitue une bonne solution de remplacement d' ACPE car elle en conserve les propriétés et nécessite une seule décomposition de Cholesky au lieu de p . L'estimation de $\sigma^2(p)$ est, par définition, celle donnée par ACPE selon le deuxième procédé. La différence entre les deux, pour $m = 20$, est de 2/1000 en moyenne sur l'ensemble des valeurs de la Table II à l'exception de $\sigma^2(0)$ (7.577, 6.285). Elle est à peine plus importante pour $m = 10$ sur $\beta(\cdot)$ et $b(\cdot)$ et atteint au plus 35/1000 sur $\rho(\cdot)$; les valeurs de $\sigma^2(0)$ sont (7.699, 7.528). Cette différence s'accroît dans l'exemple de la Table III car le rapport $m/(p+1)$ diminue encore.

7) PSEUDO MAXIMUM DE VRAISEMBLANCE

Nous ne retiendrons pas cette méthode car, pour un coût de calcul identique à celui du maximum de vraisemblance exact, elle aboutit à des résultats inférieurs. Notons, pour information, qu'elle a une qualité comparable à RER avec une amélioration dans l'estimation de $\rho(\cdot)$.

7) MAXIMUM DE VRAISEMBLANCE EXACT

Nous avons utilisé la méthode de relaxation de Pham nettement plus rapide que celle du gradient : 7 (resp. 5) itérations en moyenne pour MS1 et $m = 10$ (resp. 20) contre 18 pour la méthode du gradient. Pour MS2 et $m = 15$ ce nombre d'itérations, en relaxation, est passé à 25. Les résultats statistiques obtenus par les deux méthodes dans le cas de MS1 sont sensiblement identiques. De même l'adjonction d'un critère d'arrêt portant sur la norme du gradient (cf. § II.3) n'est pas sensible de ce point de vue. Il conserve tout son intérêt dans l'utilisation de la méthode mais il augmente de façon importante le nombre d'itérations. Les performances de MVE sont bonnes sur l'ensemble des paramètres sans toutefois se distinguer particulièrement de celles des autres méthodes excepté peut-être dans le cas de $\rho(\cdot)$. Elle est la seule méthode, avec ACPE (RER, PMV), à ne pas présenter d'incohérence dans l'estimation des deux variances. Bien que les propriétés statistiques des autres paramètres soient semblables, cette incohérence traduit des anomalies structurelles de méthode (cf. Burg, MVR, MCS) qui peuvent être importantes lors de son utilisation. Les avantages de MVR (MS1) ou d'ACPE (MS2) par rapport à MVE sont dûs à une meilleure estimation des dernières valeurs de $\beta(\cdot)$ au détriment des premières.

Nous avons mentionné dans les tables, par une ligne "moyennes", une mesure globale du biais et de $(eqm)^{1/2}$ pour chaque paramètre vectoriel. Le biais "moyen" est la distance dans \mathbb{R}^p entre les moyennes des estimations observées (valeurs de la table) et les valeurs théoriques du modèle selon le produit scalaire $(1/p)\sum x_k y_k$. La valeur "moyenne" de $(eqm)^{1/2}$ est la racine carrée de la moyenne des p écarts quadratiques moyens observés (carré des valeurs de la table). Nous avons également indiqué, dans une ligne supplémentaire, les estimations des variances $\sigma^2(0)$ et $\sigma^2(p)$ pour Burg et ACPE, lorsque $\sigma^2(p)$ est estimé à partir du minimum atteint par les erreurs de prédiction progressive et rétrograde.

TABLE I

MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

MS1 $m = 10$ $p = 4$ $r = 2000$

MODELE	BURG	ACPE	MVR	MVE	MCS						
$\beta(1)$	0.716	0.706	0.077	0.706	0.077	0.715	0.085	0.715	0.046	0.716	0.047
$\beta(2)$	-0.982	-0.960	0.062	-0.970	0.051	-0.958	0.051	-0.981	0.022	-0.978	0.025
$\beta(3)$	0.704	0.642	0.171	0.651	0.186	0.695	0.142	0.668	0.181	0.661	0.163
$\beta(4)$	-0.924	-0.786	0.245	-0.771	0.307	-0.826	0.188	-0.782	0.326	-0.737	0.326
Moyennes	0.077	0.157	0.081	0.185	0.050	0.128	0.073	0.182	0.096	0.184	
$\rho(1)$	0.716	0.706	0.077	0.706	0.077	0.715	0.085	0.715	0.046	0.716	0.047
$\rho(2)$	0.034	0.028	0.201	0.023	0.204	0.056	0.239	0.034	0.132	0.040	0.134
$\rho(3)$	-0.643	-0.588	0.230	-0.605	0.227	-0.545	0.300	-0.628	0.153	-0.617	0.161
$\rho(4)$	-0.928	-0.792	0.259	-0.815	0.247	-0.742	0.296	-0.892	0.113	-0.877	0.132
Moyennes	0.073	0.204	0.060	0.200	0.105	0.246	0.019	0.118	0.028	0.126	
$b(1)$	-2.761	-2.510	0.455	-2.526	0.424	-2.640	0.365	-2.592	0.364	-2.531	0.397
$b(2)$	3.812	3.317	0.802	3.312	0.829	3.537	0.606	3.423	0.758	3.282	0.854
$b(3)$	-2.654	-2.222	0.696	-2.206	0.766	-2.400	0.512	-2.286	0.735	-2.146	0.822
$b(4)$	0.924	0.786	0.245	0.771	0.307	0.826	0.188	0.782	0.316	0.725	0.342
Moyennes	0.358	0.590	0.364	0.622	0.203	0.447	0.289	0.580	0.397	0.648	
$\sigma^2(0)$	7.617	7.698	7.358	7.698	7.358	7.655	6.990	8.051	6.740	14.89	131.17
$\sigma^2(4)^*$	9.77	20.79	40.56	6.29	5.55	17.21	21.51	6.10	5.55	5.39	5.79
$\sigma^2(0)$	7.625	7.375	6.905	6.589							
$\sigma^2(4)^*$	20.15	38.73	5.61	5.67							

* L'unité de $\sigma^2(4)$ est égale à 10^{-3}

TABLE II

MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

MS1 $m = 20$ $p = 4$ $r = 1000$

MODELE	BURG	ACPE	MVR	MVE	MCS
$\beta(1)$	0.716	0.711 0.051	0.718 0.055	0.715 0.036	0.713 0.040
$\beta(2)$	-0.982	-0.977 0.022	-0.978 0.018	-0.984 0.012	-0.985 0.012
$\beta(3)$	0.704	0.685 0.085	0.704 0.073	0.693 0.075	0.692 0.077
$\beta(4)$	-0.924	-0.834 0.155	-0.856 0.135	-0.845 0.163	-0.826 0.180
Moyennes	0.046	0.093 0.047 0.101	0.034 0.082	0.040 0.092	0.050 0.100
$\rho(1)$	0.716	0.711 0.051	0.718 0.055	0.715 0.036	0.713 0.040
$\rho(2)$	0.034	0.028 0.140	0.048 0.161	0.032 0.101	0.028 0.113
$\rho(3)$	-0.643	-0.626 0.149	-0.603 0.197	-0.639 0.110	-0.644 0.122
$\rho(4)$	-0.928	-0.874 0.112	-0.862 0.135	-0.915 0.050	-0.915 0.049
Moyennes	0.028	0.119 0.023 0.118	0.039 0.147	0.007 0.081	0.007 0.089
$b(1)$	-2.761	-2.648 0.227	-2.711 0.186	-2.685 0.164	-2.668 0.176
$b(2)$	3.812	3.559 0.423	3.670 0.342	3.622 0.368	3.580 0.406
$b(3)$	-2.654	-2.418 0.391	-2.509 0.319	-2.467 0.367	-2.423 0.407
$b(4)$	0.924	0.834 0.155	0.856 0.135	0.845 0.163	0.826 0.180
Moyennes	0.188	0.319 0.184 0.321	0.110 0.261	0.144 0.285	0.177 0.314
$\sigma^2(0)$	7.617	7.587 6.005	7.947 5.812	7.680 5.249	15.13 48.74
$\sigma^2(4)^*$	9.77	11.83 10.71	10.66 5.51	7.99 3.47	7.82 3.53
$\sigma^2(0)$	7.396	5.904 7.284 5.764			
$\sigma^2(4)^*$	11.19	8.97 7.88 3.51			

* L'unité de $\sigma^2(4)$ est égale à 10^{-3}

TABLE III

MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

MS2 $m = 15$ $p = 8$ $r = 2000$

MODELE	BURG	ACPE	MVR	MVE	MCS						
$\beta(1)$	0.716	0.709	0.062	0.709	0.062	0.719	0.068	0.716	0.037	0.710	0.051
$\beta(2)$	-0.982	-0.974	0.034	-0.979	0.027	-0.975	0.028	-0.985	0.009	-0.970	0.041
$\beta(3)$	0.704	0.639	0.144	0.623	0.167	0.667	0.121	0.641	0.141	0.476	0.290
$\beta(4)$	-0.924	-0.829	0.150	-0.844	0.149	-0.852	0.134	-0.862	0.123	-0.765	0.188
$\beta(5)$	0.012	0.270	0.314	0.128	0.240	0.268	0.334	0.142	0.229	0.308	0.361
$\beta(6)$	-0.980	-0.770	0.268	-0.964	0.063	-0.763	0.282	-0.940	0.131	-0.956	0.049
$\beta(7)$	0.000	0.421	0.483	0.041	0.242	0.444	0.503	0.036	0.336	0.035	0.203
$\beta(8)$	-0.950	-0.684	0.348	-0.837	0.267	-0.704	0.313	-0.824	0.376	-0.801	0.290
Moyennes	0.216	0.268	0.072	0.175	0.217	0.269	0.074	0.212	0.154	0.218	
$\rho(1)$	0.716	0.709	0.062	0.709	0.062	0.719	0.068	0.716	0.037	0.710	0.051
$\rho(2)$	0.034	0.026	0.171	0.023	0.172	0.055	0.199	0.037	0.106	0.029	0.107
$\rho(3)$	-0.643	-0.618	0.194	-0.628	0.192	-0.583	0.251	-0.637	0.120	-0.627	0.114
$\rho(4)$	-0.928	-0.851	0.164	-0.866	0.151	-0.830	0.197	-0.921	0.039	-0.905	0.094
$\rho(5)$	-0.683	-0.589	0.288	-0.601	0.293	-0.593	0.273	-0.680	0.158	-0.686	0.149
$\rho(6)$	-0.088	-0.054	0.397	-0.056	0.411	-0.081	0.418	-0.091	0.290	-0.116	0.269
$\rho(7)$	0.488	0.405	0.392	0.412	0.414	0.365	0.449	0.481	0.269	0.466	0.244
$\rho(8)$	0.724	0.546	0.385	0.557	0.422	0.502	0.410	0.717	0.084	0.720	0.059
Moyennes	0.083	0.283	0.076	0.295	0.104	0.309	0.005	0.165	0.016	0.156	
$b(1)$	-2.784	-3.588	0.978	-2.685	0.447	-3.697	1.049	-2.825	0.490	-2.694	0.314
$b(2)$	5.712	7.580	2.535	5.305	1.223	7.881	2.701	5.681	1.339	5.302	0.978
$b(3)$	-7.938	-10.763	4.232	-7.077	2.218	-11.276	4.470	-7.726	2.377	-7.063	1.855
$b(4)$	9.211	11.810	4.589	8.016	2.711	12.406	4.775	8.746	2.902	7.973	2.407
$b(5)$	-7.807	-9.855	3.915	-6.652	2.573	-10.398	4.038	-7.266	2.805	-6.553	2.371
$b(6)$	5.522	6.313	2.233	4.691	1.793	6.659	2.213	5.012	2.057	4.558	1.769
$b(7)$	-2.645	-2.689	0.907	-2.234	0.890	-2.836	0.829	-2.336	1.077	-2.147	0.913
$b(8)$	0.950	0.684	0.348	0.837	0.267	0.704	0.313	0.824	0.376	0.801	0.290
Moyennes	1.723	2.907	0.754	1.756	2.091	3.024	0.341	1.920	0.810	1.580	
$\sigma^2(0)$	7.617	7.449	6.523	7.449	6.523	7.564	6.216	7.961	6.095	1.232	11.009
$\sigma^2(8)^*$	3.77	215.4	583.6	2.03	2.43	116.6	193.4	2.47	4.69	1.37	2.64
$\sigma^2(0)$		7.371	6.425	5.498	5.499						
$\sigma^2(8)^*$		220.8	634.6	1.43	2.60						

* L'unité de $\sigma^2(8)$ est égale à 10^{-5}

§ IV.3. Extension des méthodes à un échantillon

Soit $x = [x(i,j)]$ la matrice $m \times n$ associée à n observations indépendantes d'une séquence stationnaire de longueur m , gaussienne, centrée et de matrice de covariance Λ . La covariance empirique "symétrisée"

$$X = (xx' + Jxx'J)/2$$

est une statistique exhaustive, estimateur sans biais de Λ , en fonction de laquelle s'exprime chacune des méthodes. En ce sens toute méthode s'interprète comme un opérateur T de l'ensemble des matrices doublement symétriques positives $\bar{\mathcal{C}}^+$ dans le sous-ensemble $\bar{\mathcal{C}}^+$ formé des matrices de Toeplitz.

Les autocovariances empiriques sont définies par

$$\Lambda_e(k) = \frac{1}{m} \sum_{t=1}^{m-k} X(t+k,t), \quad k=0, \dots, m-1,$$

et l'opérateur correspondant est le seul à ne pas être un projecteur ($T(\Lambda_e) \neq \Lambda_e$).

Considérons, pour $k = 0, \dots, m-1$, la matrice A_k égale à la moyenne, divisée par deux, des matrices $A_{k,r}$ associées selon (2) à chaque séquence :

$$A_k = \frac{1}{2n} \sum_{r=1}^n A_{k,r}.$$

Elle s'exprime en fonction de X par

$$A_k(i,j) = \sum_{t=1}^{m-k} X(i+t,j+t), \quad i=0, \dots, k, \quad j=0, \dots, k.$$

La matrice $A_k/(m-k)$ est un estimateur sans biais de la matrice d'autocovariance d'ordre $(k+1)$, obtenue comme moyenne des blocs diagonaux du même ordre contenus dans X .

Les matrices A_k sont doublement symétriques et se construisent selon la récurrence :

$$A_k(k,0) = \sum_{t=1}^{m-k} X(k+t,t) = m\Lambda_e(k),$$

$$A_k(k-j,j) = A_{k-1}(k-1-j,j-1) - X(k-j,j), \quad j=1, \dots, [k/2],$$

$$A_k(i,j) = A_{k-1}(i,j) - X(k-i,k-j), \quad j=0, \dots, [k/2], \quad j \leq i < k-j.$$

Cette récurrence équivaut à $A_k(k,0) = m\Lambda_e(k)$ joint à la relation matricielle

$$(A_k)_k = A_{k-1} - J(X)_k J,$$

où $(M)_k$ désigne la matrice d'ordre k constituée par le premier bloc diagonal du même ordre de la matrice M . A partir de la version "symétrisée" de (8) on peut vérifier l'égalité

$$\text{tr}(\Lambda_p^{-1} X) = \text{tr}[(\Lambda_p)^{-1}(X)_p] + \frac{1}{2} \frac{b'A_p b}{\sigma(p)},$$

dans laquelle Λ_p est la matrice d'ordre m du modèle AR(p) sous-jacent à Λ et $b = [1, b(1), \dots, b(p)]'$ est le paramètre correspondant. Cela montre en particulier que chacune des statistiques $\{J(X)_p J; \Lambda_e(k), k = 0, \dots, p\}$, $\{J(X)_p J; A_p\}$ et $\{A_p; A_{p-1}\}$ est exhaustive pour le modèle AR(p).

L'estimation de la matrice de Tœplitz Λ de la séquence correspond au cas $p = m-1$. On supposera $p < 2nm/(2n+1)$ afin d'assurer une solution à chacune des méthodes.

1) AUTOCORRELATION PARTIELLE EMPIRIQUE

L'autocorrélation partielle empirique $\beta_{es}(k)$, $k = 1, \dots, p$, est une corrélation partielle exprimée dans $\mathbb{R}^{2n(m-k)}$. Elle est égale au coefficient de régression $-b(k)$ dans la solution du problème de minimisation

$$\text{Min}_{b(\cdot)} \sum_{r=1}^n \sum_{t=k+1}^m (\epsilon_r^2(t;k) + \epsilon_r^{*2}(t-k;k)), \quad \epsilon_r(t;k) = \sum_{j=0}^k b(j)x(t-j,r), \quad \epsilon_r^*(t;k) = \sum_{j=0}^k b(j)x(t+j,r).$$

Elle est donnée par

$$\beta_{es}(k) = \gamma_k(k) \{\gamma_k^2(k) + s_k^2(k)\}^{-1/2},$$

où $[\gamma_k(1), \dots, \gamma_k(k), s_k(k)]$ est la dernière ligne de la matrice triangulaire inférieure de la décomposition de Cholesky de $PA_k P'$. Les variances sont estimées en utilisant $\Lambda_e(0)$ pour $\sigma^2(0)$ ou $s_p^2(p)/(m-p)$ pour $\sigma^2(p)$, mais dans ce dernier cas la méthode n'est plus récursive selon l'ordre p du modèle. Dans les deux cas l'opérateur T reste un projecteur en vertu des propriétés des matrices de Tœplitz (cf. § A.II.3).

L'approximation RER s'obtient, comme pour une seule séquence, en fonction de la seule matrice A_p . La méthode n'est plus récursive mais demeure un projecteur.

2) BURG

Les estimations $\beta_b(k)$, $k = 1, \dots, p$, sont définies récursivement à partir des problèmes de minimisation

$$\text{Min}_{\beta(k)} \sum_{l=1}^n \sum_{t=k+1}^m \{[\varepsilon_{b,r}(t;k-1) - \beta(k)\varepsilon_{b,r}^*(t-k;k-1)]^2 + [\varepsilon_{b,r}^*(t-k;k-1) - \beta(k)\varepsilon_{b,r}(t;k-1)]^2\},$$

où les erreurs $\varepsilon_{b,r}(t;k-1)$, $\varepsilon_{b,r}^*(t-k;k-1)$, $t = k+1, \dots, m$, sont calculées en fonction du filtre $\underline{b}_k = [b(k-1,0), \dots, b(k-1,k-1)]'$ associé aux $\beta_b(j)$, $j = 1, \dots, k-1$, déjà estimés. Ainsi $\beta_b(k)$ est donné par :

$$\beta_b(k) = \tilde{b}_k' A_k \tilde{J} \tilde{b}_k / \tilde{b}_k' A_k \tilde{b}_k, \quad \tilde{b}_k' = [\underline{b}_k', 0]'$$

Comme pour ACPE la méthode est récursive lorsque $\sigma^2(0)$ est estimé par $\Lambda_e(0)$, elle ne l'est plus si $\sigma^2(p)$ est estimé par le minimum obtenu à l'ordre p :

$$\frac{1}{(m-p)} [1 - \beta_b^2(p)] \underline{b}_p' A_p \underline{b}_p.$$

Pour une matrice de Toeplitz Λ d'ordre $(k+1)$, définie positive, le minimum de $\underline{b}' \Lambda \underline{b}$, où $\underline{b} = [b(0), b(1), \dots, b(k)]'$ avec la seule contrainte $b(0) = 1$, est réalisé par le filtre autorégressif d'ordre k associé à Λ . Ainsi la méthode correspond à un projecteur indépendamment du choix de l'estimation des variances.

3) MOINDRES CARRES SYMETRIQUES

Le problème de minimisation est celui considéré pour ACPE à l'ordre p que l'on peut aussi formuler par

$$\text{Min} \{ \underline{b}' A_p \underline{b}; \underline{b} = [1, b(1), \dots, b(p)]' \}.$$

La solution est donnée par la dernière ligne $[b(1), \dots, b(p), 1]$ de la matrice triangulaire inférieure unitaire T^{-1} de la décomposition de Cholesky TD^2T' de PA_pP' . Les variances sont estimées en utilisant $s_p^2(p)/(m-p)$ pour $\sigma^2(p)$ où $s_p^2(p)$ est le dernier élément de D^2 . La méthode n'est pas récursive mais demeure un projecteur pour la même raison que ci-dessus.

4) MAXIMUM DE VRAISEMBLANCE EXACT

Posant $\Lambda_p = \sigma^2(p)R_p$, la solution du maximum de vraisemblance du modèle AR(p) gaussien est donnée par $\hat{\sigma}^2(p)\hat{R}_p$ où \hat{R}_p réalise le minimum de

$$\frac{1}{m} \ln |R_p| + \ln \left[\frac{1}{m} \text{tr} \{ R_p^{-1} X \} \right]$$

et $\hat{\sigma}^2(p) = \text{tr} \{ \hat{R}_p^{-1} X \} / m$. Il est possible de transposer la méthode du gradient utilisée pour une séquence mais la méthode de relaxation (Pham, 1988) est plus simple. D'autre part le résultat obtenu après une seule itération constitue l'extension à un échantillon du maximum de vraisemblance récursif.

Par construction MVR est récursif et par suite MVE ne l'est pas car les deux méthodes coïncident pour $p = 1$. Il est clair que MVE définit un projecteur, c'est-à-dire que le minimum ci-dessus est donné par R_p lorsque $X = \Lambda_p$. Mais c'est également le cas de MVR car, si X est une matrice de Toeplitz définie positive d'ordre m , X_k est la solution du maximum de vraisemblance pour tout k inférieur à m (cf. § A.III.3).

La différence essentielle ici est que la séquence peut être très courte par rapport à l'ordre du modèle estimé : $m = p-1$ dans le cadre de l'estimation d'une matrice de Toeplitz. C'est la situation que nous avons retenue pour le modèle MS1 présenté dans les Tables IV ($n = 3$) et V ($n = 10$). Nous retrouvons la plupart des observations faites au paragraphe précédent. Notons cependant les pertes de qualité de MVR pour l'estimation de $\beta(\cdot)$ ou de $b(\cdot)$, de MCS pour $\rho(\cdot)$, et même de la méthode de Burg pour $\beta(\cdot)$ et surtout $b(\cdot)$ pour lequel elle est celle qui réalise les plus mauvais résultats. Là encore seules MVE et ACPE sont cohérentes sur l'ensemble des paramètres tout en conservant de bonnes qualités pour chacun d'eux. Le point faible de MVE, dans l'estimation de $\beta(4)$ lorsque l'échantillon est de taille très faible (cf. Table IV), est dû à la variance de l'estimateur car le biais est le même que celui d'ACPE. Notons que c'était déjà le cas entre MVE et MVR pour une seule séquence (cf. Table I). De plus s'ajoute ici une mauvaise estimation de $\sigma^2(0)$ qui se répercute sur celle de $\Lambda(\cdot)$ et qui n'existait pas pour une séquence des modèles MS1 et MS2. Cela tient à un nombre d'observations trop faible et

n'a pas véritablement de sens statistique. En effet nous avons déjà constaté ceci sur un autre essai du modèle MS2, dans les mêmes conditions que celles de la Table III, alors que les résultats concernant les autres paramètres restaient analogues. De plus lorsque l'on donne, dans ce type de situation, une importance plus grande au critère d'arrêt portant sur le gradient, le nombre d'itérations augmente très fortement et conduit à une très légère amélioration de l'estimation de $\sigma^2(0)$. Un examen plus approfondi montre que la différence porte sur un nombre de cas restreints pour lesquels la convergence est très lente. Pour $n = 10$ (cf. Table V) MVE réalise les meilleures performances sur chacun des paramètres avec 4 itérations en moyenne contre 13 dans celui de la Table IV avec le même critère d'arrêt. A l'exception de MVE, l'augmentation de n ne modifie pas les qualités relatives des différentes méthodes. Notons la particularité de MCS qui voit l'estimation de $\sigma^2(0)$ se détériorer encore plus alors que le nombre de polynômes instables passe de 824/2000 à 121/1000. Dans le premier cas la proximité des $\beta(k)$ à ± 1 est due au hasard, alors que dans le second elle est due au modèle. Enfin signalons que l'approximation d'ACPE par RER n'est pas acceptable; elle conduit à une mauvaise estimation de $\beta(1)$, avec ses conséquences sur les autres paramètres, qui, toutes proportions gardées, ne s'améliore pas lorsque n augmente.

TABLE IV

MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

MS1 $m = 5$ $n = 3$ $p = 4$ $r = 2000$

MODELE	BURG	ACPE	MVR	MVE	MCS	
$\beta(1)$	0.716	0.712 0.043	0.712 0.043	0.715 0.057	0.716 0.033	0.723 0.053
$\beta(2)$	-0.982	-0.976 0.027	-0.979 0.024	-0.969 0.028	-0.980 0.019	-0.968 0.039
$\beta(3)$	0.704	0.553 0.287	0.636 0.238	0.654 0.191	0.675 0.166	0.678 0.159
$\beta(4)$	-0.924	-0.784 0.306	-0.894 0.223	-0.796 0.311	-0.870 0.320	-0.817 0.284
Moyennes	0.103	0.212 0.037	0.165 0.165	0.069 0.185	0.031 0.181	0.055 0.166
$\rho(1)$	0.716	0.712 0.043	0.712 0.043	0.715 0.057	0.716 0.033	0.723 0.053
$\rho(2)$	0.034	0.029 0.121	0.028 0.121	0.044 0.158	0.036 0.096	0.064 0.160
$\rho(3)$	-0.643	-0.632 0.144	-0.634 0.142	-0.594 0.178	-0.630 0.117	-0.576 0.223
$\rho(4)$	-0.928	-0.892 0.121	-0.891 0.119	-0.834 0.159	-0.902 0.091	-0.836 0.219
Moyennes	0.019	0.114 0.019	0.113 0.113	0.053 0.146	0.014 0.090	0.058 0.178
$\Lambda(1)$	5.454	5.491 3.231	5.491 3.231	5.435 3.212	5.887 4.344	6.862 32.849
$\Lambda(2)$	0.260	0.228 0.792	0.222 0.791	0.367 1.287	0.282 0.868	0.613 9.970
$\Lambda(3)$	-4.894	-4.955 3.154	-4.967 3.152	-4.545 3.004	-5.271 3.908	-5.615 29.236
$\Lambda(4)$	-7.065	-7.099 4.498	-7.095 4.495	-6.511 4.141	-7.593 5.849	-8.270 42.916
Moyennes	0.042	3.211 0.047	3.210 0.332	3.088 0.390	4.156 1.010	31.125
$b(1)$	-2.761	-2.355 0.594	-2.589 0.469	-2.545 0.427	-2.653 0.370	-2.635 0.372
$b(2)$	3.812	3.100 0.987	3.535 0.739	3.375 0.793	3.602 0.747	3.522 0.726
$b(3)$	-2.654	-2.055 0.853	-2.440 0.614	-2.269 0.728	-2.472 0.723	-2.381 0.681
$b(4)$	0.924	0.784 0.306	0.894 0.223	0.796 0.311	0.870 0.320	0.817 0.284
Moyennes	0.512	0.733 0.195	0.546 0.317	0.600 0.151	0.575 0.216	0.550
$\sigma^2(0)$	7.617	7.691 4.485	7.691 4.485	7.569 4.372	8.223 5.964	9.459 44.524
$\sigma^2(4)^*$	9.77	18.27 19.91	4.98 8.62	20.54 27.05	4.93 7.55	2.77 7.94
$\sigma^2(0)$	8.506	5.424 5.932	4.968			
$\sigma^2(4)^*$	20.77	25.66 3.01	7.84			

* L'unité de $\sigma^2(4)$ est égale à 10^{-3}

TABLE V

MOYENNE ET RACINE CARREE DE L' ECART QUADRATIQUE MOYEN DES ESTIMATIONS

MS1 $m = 5$ $n = 10$ $p = 4$ $r = 2000$

MODELE	BURG	ACPE	MVR	MVE	MCS						
$\beta(1)$	0.716	0.714	0.022	0.714	0.022	0.714	0.033	0.715	0.015	0.716	0.022
$\beta(2)$	-0.982	-0.981	0.010	-0.981	0.024	-0.978	0.011	-0.982	0.007	-0.982	0.012
$\beta(3)$	0.704	0.653	0.134	0.682	0.009	0.686	0.101	0.700	0.054	0.702	0.052
$\beta(4)$	-0.924	-0.878	0.098	-0.915	0.123	-0.889	0.084	-0.922	0.060	-0.907	0.077
Moyennes	0.034	0.084	0.012	0.068	0.020	0.068	0.003	0.041	0.009	0.048	
$\rho(1)$	0.716	0.714	0.022	0.714	0.022	0.714	0.033	0.715	0.015	0.716	0.022
$\rho(2)$	0.034	0.030	0.063	0.029	0.063	0.033	0.092	0.033	0.043	0.035	0.063
$\rho(3)$	-0.643	-0.643	0.071	-0.644	0.071	-0.631	0.102	-0.642	0.050	-0.639	0.077
$\rho(4)$	-0.928	-0.919	0.046	-0.918	0.046	-0.896	0.063	-0.923	0.032	-0.922	0.051
Moyennes	0.005	0.054	0.005	0.054	0.017	0.077	0.002	0.038	0.004	0.057	
$\Lambda(1)$	5.454	5.433	1.755	5.433	1.755	5.411	1.760	5.471	1.609	25.94	221.66
$\Lambda(2)$	0.260	0.227	0.448	0.225	0.448	0.272	0.715	0.245	0.326	0.13	8.63
$\Lambda(3)$	-4.894	-4.914	1.718	-4.916	1.718	-4.770	1.702	-4.932	1.154	-25.50	223.49
$\Lambda(4)$	-7.065	-7.048	2.443	-7.042	2.443	-6.830	2.356	-7.097	2.213	-35.96	313.66
Moyennes	0.023	1.747	0.026	1.746	0.135	1.736	0.027	1.582	20.49	222.23	
$b(1)$	-2.761	-2.625	0.260	-2.704	0.223	-2.691	0.197	-2.748	0.102	-2.746	0.110
$b(2)$	3.812	3.572	0.403	3.720	0.313	3.675	0.305	3.790	0.173	3.774	0.204
$b(3)$	-2.654	-2.455	0.322	-2.585	0.229	-2.537	0.246	-2.638	0.152	-2.616	0.188
$b(4)$	0.924	0.878	0.098	0.915	0.068	0.889	0.084	0.922	0.060	0.907	0.077
Moyennes	0.171	0.293	0.064	0.226	0.098	0.223	0.015	0.130	0.029	0.154	
$\sigma^2(0)$	7.617	7.605	2.436	7.605	2.436	7.562	2.394	7.648	2.244	36.65	314.13
$\sigma^2(4)^*$	9.77	14.01	8.48	8.46	4.54	13.81	8.79	8.41	4.26	7.67	4.15
$\sigma^2(0)$	7.923	2.797	7.167	2.679							
$\sigma^2(4)^*$	14.71	9.74	7.74	4.15							

* L'unité de $\sigma^2(4)$ est égale à 10^{-3}

REFERENCES

- A, N. Q. (1984) On the uniqueness of the maximum likelihood estimate of structured covariance matrices. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-32, n° 6, 1249-1251.
- ANDERSEN, N. (1974) On the calculation of filter coefficients for maximum entropy spectral analysis. *Geophysics*, Vol. 39, n° 1, 69-72.
- ATAL, B. S. (1977) On determining partial correlation coefficients by the covariance method of linear prediction. Presented at the 94th Meeting, Acoust. Soc. Amer.
- AZENCOTT, R. et DACUNHA-CASTELLE, D. (1984) *Séries d'observations irrégulières, modélisation et prévision*. Masson, Paris.
- BARNDORFF-NIELSEN, O. and SCHOU, G. (1973) On the parametrization of autoregressive models by partial autocorrelations. *J. Multivariate Analysis*, Vol. 3, 408-419.
- BHANSALI, R. J. (1983 a) The inverse partial correlation function of a time series and its applications. *J. Multivariate Analysis*, 13.
- BHANSALI, R. J. (1983 b) A simulation study of autoregressive and window estimators of the inverse correlation function. *Appl. Statist.* Vol. 32, n° 2, 141-149.
- BOX, G. E. P. and JENKINS, G. M. (1970) *Time series analysis, forecasting and control*, Holden-Day.
- BURG, J. P. (1967) Maximum entropy spectral analysis. Presented at the 37th Annual International S.E.G. Meeting, Oklahoma City, Oklahoma.
- BURG, J. P. (1968) A new analysis technique for time series data. Presented at NATO Advanced Study Institute on Signal Processing, Enschede, Netherlands.
- BURG, J. P. (1975) Maximum entropy spectral analysis. Ph. D. dissertation, Dep. Geophys. Stanford Univ., Stanford, CA.
- BURG, J. P., LUENBERGER, D. G. and WENGER, D. L. (1982) Estimation of structured covariance matrices. *Proc. IEEE*, 70, 963-974.
- CLEVELAND, W. S. (1972) The inverse autocorrelations of a time series and their applications. *Technometrics*, 14, 277-293.
- DEGERINE, S. (1982 a) Propriétés de moindres carrés de la matrice de covariance d'un vecteur aléatoire et applications statistiques. *Ann. Inst. Henri Poincaré, Section B*, Vol. 23, n° 3, 237-247.
- DEGERINE, S. (1982 b) Partial autocorrelation function for a scalar stationary discrete-time process. *Proc. 3rd. Franco-Belgian Meeting of Statisticians*. Bruxelles: Publications des Facultés Universitaires Saint-Louis, 79-94.
- DEGERINE, S. (1983) Décompositions spectrales d'une matrice d'autocovariance et applications. *C. R. Acad. Sci. Paris*, t. 296, Serie I, 565-568.
- DEGERINE, S. (1985) Comportement au bord et caractérisation d'un maximum pour la vraisemblance d'un vecteur aléatoire gaussien centré avec contrainte sur sa structure de covariance. *C. R. Acad. Sci. Paris*, t. 301, Série I, n° 5, 233-236.

- DEGERINE, S. (1986) Autocorrélations partielles empiriques et applications. Rap. Rech. n° 627-M- Lab. TIM3, Univ. J. Fourier, Grenoble.
- DEGERINE, S. (1987) Maximum likelihood estimation of autocovariance matrices from replicated short time series. *J. Time Series Anal.* Vol. 8, n° 2, 135-146.
- DEGERINE, S. (1988) Fonction d'autocorrélation partielle canonique d'un processus vectoriel stationnaire. *C. R. Acad. Sci. Paris*, t. 306, Serie I, 287-290.
- DEGERINE, S. (1988) Canonical matrix reflection coefficients for multichannel stationary time series. Présenté à EUSIPCO-88, Grenoble.
- DEGERINE, S. and PHAM, D.T. (1987). Une nouvelle approche du calcul de l'estimateur de maximum de vraisemblance exact d'un modèle autorégressif. Présenté au 11-ième Colloque GRETSI sur le traitement du signal et des images. Nice.
- DEMPSTER, A. D., LAIRD, N. M. and RUBIN, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, Vol. B39, 1-37.
- DICKINSON, B. W. (1978) Autoregressive estimation using residual energies ratios. *IEEE Trans. on Information Theory*, Vol. IT-24, n° 2, 503-506.
- DICKINSON, B. W. (1979) Estimation of partial correlation matrices using Cholesky decomposition. *IEEE Trans. on Automatic Control*, Vol. AC-24, n° 2, 302-305.
- DURBIN, J. (1960) The fitting of time-series models. *Rev. Inst. Internat. Statist.*, 28, 233-244.
- GERONIMUS, Y. L. (1960) Polynomials orthogonal on a circle and interval. Pergamon Press.
- HASZA, D. P. (1980) A note on maximum likelihood estimation for the first-order autoregressive process. *Commun. Statist.-Theor. Meth.*, A9 (13), 1411-1415.
- HELSON, H. and SZEGÖ, G. (1960) A problem in prediction theory. *Ann. Mat. Pura. Appl.* 51, 107-138.
- HOFFMAN, K. (1962) Banach spaces of analytic functions. Prentice-Hall.
- JEWELL, N. P. and BLOOMFIELD, P. (1983) Canonical correlations of past and future for time series: definitions and theory. *The Annals of Statistics*, Vol. 11, n° 3, 837-847.
- KAILATH, T. (1974) A view of three decades of linear filtering theory. *IEEE Trans. on Information Theory*, Vol. IT-20, n° 2, 106-181.
- KAILATH, T., VIEIRA, A. AND MORF, M. (1978) Inverses of Toeplitz operators, innovations, and orthogonal polynomials. *SIAM Review*, 20 (1), 106-119.
- KAY, S. M. (1983) Recursive maximum likelihood estimation of autoregressive processes. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-31, n° 1, 56-65.
- KENDALL, M. G. and STUART, A. (1966) Advanced theory of statistics, Vol. 3, Griffin.
- LE CADRE, J. P. et LOPEZ, P. (1984) Estimation d'une matrice interspectrale de structure imposée. Applications. *Traitement du signal*, 1, 3-17.
- LE BRETON, A. (1987) A note on maximum likelihood estimation for the complex-valued first-order autoregressive process. A paraître dans *Stat. Prob. Letters*.

- LE ROUX, J. and GUEGUEN, C. (1977) A fixed point computation of partial correlation coefficients. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, 257-259.
- LEVINSON, N. (1947) The Wiener RMS (root-mean-square) error criterion in filter design and prediction. *J. Math. Phys.*, 25, 261-278.
- MAKHOUL, J. (1977) Stable and efficient lattice methods for linear prediction. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, 423-427.
- MARPLE, L. (1980) A new autoregressive spectrum analysis algorithm. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-28, 441-454.
- MILLER, M. I. and SNYDER, D. L. (1987) The role of likelihood and entropy in incomplete-data problems: applications to estimating point process intensities and Toeplitz constrained covariances. *Proc. IEEE*, Vol. 75, n° 7, 892-907.
- MORF, M., DICKINSON, B., KAILATH, T. and VIERA, A. (1977) Efficient solution of covariance equations for linear prediction. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-25, 429-433.
- MORF, M., VIEIRA, A. and KAILATH, T. (1978 a) Covariance characterization by partial autocorrelation matrices. *Ann. Statist.* Vol. 6, n° 3, 643-648.
- MORF, M., VIEIRA, A., LEE, D. and KAILATH, T. (1978 b) Recursive multichannel maximum entropy spectral estimation. *IEEE Trans. on Geoscience Electronics*, Vol. GE-16, n° 2, 85-94.
- NUTTAL, A. H. (1976) Spectral analysis of a univariate process with bad data points, via maximum entropy and linear predictive techniques. Naval Underwater Systems Center, New London, CT, Tech. Rep. 5303.
- OSTROWSKI, A. M. (1973) Solution of equations in Euclidean and Banach spaces (3rd edn). Academic Press, New York.
- PARZEN, E. (1974) Some recent advances in time series modeling. *IEEE Trans. on Automatic Control*, Vol. AC-19, 723-730.
- PHAM, D. T. (1987) Exact maximum likelihood and Lagrange multiplier statistic for ARMA models. *J. Time Series Anal.* 18, Vol.1, p.61-78.
- PHAM, D. T. (1988) Maximum likelihood estimation of autoregressive model by relaxation on the reflection coefficients. A paraître dans *IEEE, Trans. on Acoustics, Speech, and Signal Processing*.
- PHAM, D. T. et LE BRETON, A. (1988) Levinson-Durbin type algorithms for continuous time autoregressive models and applications. Soumis pour publication.
- PHAM, D. T., LE BRETON, A. et TONG, D. Q. (1988) Maximum likelihood estimation for complex autoregressive model and Toeplitz interspectral matrices. Présenté à EUSIPCO-88, Grenoble.
- PRIESTLEY, M. B. (1981) Spectral analysis and time series. Vol.1. Academic-Press.
- QUENOUILLE, M. H. (1949) Approximate tests of correlation in time series. *J. Roy. Statist. Soc. Ser. B*, 11, 68-84.
- RAMSEY, F. (1974) Characterization of the partial autocorrelation function. *Ann. Statist.* n° 2, 1296-1301.

- RAO, C. R. (1965) *Linear statistical inference and its applications*. John Wiley & Sons, Inc., New York.
- ROMEDER, J. M. (1973) *Méthodes et programmes d'analyse discriminante*. Dunod, Paris.
- ROZANOV, Yu. A. (1960) On stationary sequences forming a basis. *Soviet. Math. Doklady* 1, 155-158.
- ROZANOV, Yu. A. (1967) *Stationary random processes*. Holden-Day.
- SAKAI, H. (1983) Covariance matrices characterization by a set of scalar partial autocorrelation coefficients. *Ann. Statist.* Vol. 11, n° 1, 337-340.
- SCHUSSLER, H. W. (1976) A stability theorem for discrete systems. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-24, 87-89.
- ULRYCH, T. J. and CLAYTON, R. W. (1976) Time series modeling and maximum entropy. *Phys. Earth and Plan. Int.*, Vol. 12, 188-200.
- WHITTLE, P. (1963) On the fitting of multivariable autoregressions and the approximate canonical factorization of a spectral density matrix. *Biometrika* 50, 129-134.
- WIENER, N. and MASANI, P. (1957) The prediction theory of multivariate stochastic processes, I. *Acta Mathematica*, 98, 111-150.
- WIENER, N. and MASANI, P. (1958) The prediction theory of multivariate stochastic processes, II. *Acta Mathematica*, 99, 93-137.
- WILSON, G. (1969) Factorization of the covariance generating function of a pure moving average process. *SIAM J. Numer. Anal.* 6, 1-7.

AUTORISATION DE SOUTENANCE

DOCTORAT D'ETAT

Vu les dispositions de l'Article 5 de l'Arrêté du 16 avril 1974,

Vu les rapports de M.A. LE BRETON et D.T. PHAM.....

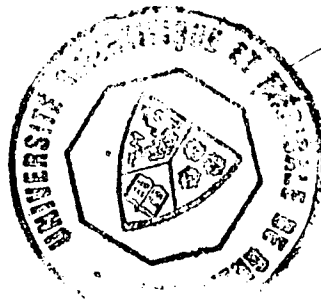
M.C. GUEGUEN.....

M.M. HALLIN.....

M. Serge DEGERINE..... est autorisé à
présenter une thèse en vue de l'obtention du grade de DOCTEUR D'ETAT
ES SCIENCES.

Fait à Grenoble, le 28 JUIN 1988

Le Président de l'U.S.T.M.G.



J.J. Payan
J.J. PAYAN

