HAL
open science

Sur l'identification des systèmes linéaires déterministes instables

Paul Caspi

To cite this version:

Paul Caspi. Sur l'identification des systèmes linéaires déterministes instables. Automatique / Robotique. Institut National Polytechnique de Grenoble - INPG; Université Joseph-Fourier - Grenoble I, 1978. Français. NNT : . tel-00011711

HAL Id: tel-00011711

https://theses.hal.science/tel-00011711

Submitted on 2 Mar 2006

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

Université Scientifique et Médicale de Grenoble
Institut National Polytechnique de Grenoble

pour obtenir le grade de
Docteur d'Etat ès - Sciences Physiques

par

Paul CASPI

SUR L'IDENTIFICATION DES SYSTEMES LINEAIRES DETERMINISTES INSTABLES

Thèse soutenue le 25 mai 1978 devant la Commission d'examen

Président	: R. PERRET
	J. AGUILAR - MARTIN
	C. FOULARD
Examinateurs	J.F. LEMAITRE
	J. RICHALET
Madame	G. SAUCIER

UNIVERSITE SCIENTIFIQUE
ET MEDICALE DE GRENOBLE

Monsieur Gabriel CAU : Président
Monsieur Pierre JULLIEN : Vice Président

MEMBRES DU CORPS ENSEIGNant : DE L'U.S.M.G.

PROFESSEURS TITULAIRES

MM	AMBLARD Pierre	Clinique de dermatologie
	ARNAUD Paul	Chimie
	ARVIEU Robert	I.S.N
	AUBERT Guy	Physique
	AYANT Yves	Physique approfondie
Mme	BARBIER Marie-Jeanne	Electrochimie
MM.	BARBIER Jean-Claude	Physique Expérimentale
	BARBIER Reynold	Géologie appliquée
	BARJON Robert	Physique nucléaire
	BARNOUD Fernand	Biosynthèse de la cellulose
	BARRA Jean-René	Statistiques
	BARRIE Joseph	Clinique chirurgicale
	BEAUDOING André	Clinique de Pédiatrie et Puériculture
	BELORIZKY Elie	Physique
	BERNARD Alain	Mathématiques Pures
Mme	BERTRANDIAS Françoise	Mathématiques Pures
MM.	BERTRANDIAS Jean-Paul	Mathématiques Pures
	BEZEZ Henri	Pathologie chirurgicale
	BLAMBERT Maurice	Mathématiques Pures
	BOLLIET Louis	Informatique (IUT B)
	BONNET Jean-Louis	Clinique ophtalmologique
	BONNET-EYMARD Joseph	Clinique gastro-entérologique
Mme	BONNIER Marie-Jeanne	Chimie générale
MM.	BOUCHERLE André	Chimie et toxicologie
	BOUCHEZ Robert	Physique nucléaire
	BOUSSARD Jean-Claude	Mathématiques appliquées
	BOUTET DE MONTVEL Louis	Mathématiques Pures
	BRAVARD Yves	Géographie
	CABANEL Guy	Clinique rhumatologique et hydrologique
	CALAS François	Anatomie
	CARLIER Georges	Biologie végétale
	CARRAZ Gilbert	Biologie animale et pharmacodynamie
	CAU Gabriel	Médecine légale et toxicologie
	CAUQUIS Georges	Chimie organique
	CHABAUTY Claude	Mathématiques Pures
	CHARACHON Robert	Clinique Oto-rhino-laryngologique
	CHATEAU Robert	Clinique de neurologie
	CHIBON Pierre	Biologie animale
	COEUR André	Pharmacie chimique et chimie analytique
	CONTAMIN Robert	Clinique gynécologique
	COUDERC Pierre	Anatomie pathologique
Mme	DEBELMAS Anne-Marie	Matière médicale
MM.	DEBELMAS Jacques	Géologie générale
	DEGRANGE Charles	Zoologie
	DELORMAS Pierre	Pneumophtisiologie

MM.	DEPORTES Charles	Chimie minérale
	DESRE Pierre	Métallurgie
	DESSAUX Georges	Physiologie animale
	DODU Jacques	Mécanique appliquée (IUT I)
	DOLIQUE Jean-Michel	Physique des plasmas
	DREYFUS Bernard	Thermodynamique
	DUCROS Pierre	Cristallographie
	GAGNAIRE Didier	Chimie Physique
	GALVANI Octave	Mathématiques Pures
	GASTINEL Noël	Analyse numérique
	GAVEND Michel	Pharmacologie
	GEINDRE Michel	Electroradiologie
	GERBER Robert	Mathématiques Pures
	GERMAIN Jean-Pierre	Mécanique
	GIRAUD Pierre	Géologie
	JANIN Bernard	Géographie
	KAHANE André	Physique généralé
	KLEIN Joseph	Mathématiques Pures
	KOSZUL Jean-Louis	Mathématiques Pures
	KRAVTCHENKO Julien	Mécanique
	KUNTZMANN Jean	Mathématiques Appliquées
	LACAZE Albert	Thermodynamique
	LACHARME Jean	Biologie végétale
Mme	LAJZEROWICZ Janine	Physique
MM.	LAJZEROWICZ Joseph	Physique
	LATREILLE René	Chirurgie générale
	LATURAZE Jean	Biochimie Pharmaceutique
	LAURENT Pierre	Mathématiques Appliquées
	LEDRU Jean	Clinique médicale B
	LE ROY Philippe	Mécanique (IUT I)
	LLIBOUTRY Louis	Géophysique
	LOISEAUX Pierre	Sciences Nucléaires
	LONGEQUEUE Jean-Pierre	Physique Nucléaire
	LOUP Jean	Géographie
Melle	LUTZ Elisabeth	Mathématiques Pures
MM.	MALINAS Yves	Clinique Obstétricale
	MARTIN-NOEL Pierre	Clinique Cardiologique
	MAZARE Yves	Clinique Médicale A
	MICHEL Robert	Minéralogie et Pétrographie
	MICOUD Max	Clinique Maladies infectieuses
	MOURIQUAND Claude	Histologie
	MOUSSA André	Chimie Nucléaire
	NOZIERES Philippe	Spectrometrie Physique
	OZENDA Paul	Botanique
	PAYAN Jean-Jacques	Mathématiques Pures
	PEBAY-PEYROULA Jean-Claude	Physique
	PERRET Jean	Semeiologie Médicale (Neurologie)
	RASSAT André	Chimie systématique
	RENARD Michel	Thermodynamique
	REVOL Michel	Urologie
	RINALDI Renaud	Physique
	DE ROUGEMONT Jacques	Neuro-Chirurgie
	SEIGNEURIN Raymond	Microbiologie et Hygiène
	SENGEL Philippe	Zoologie
	SIBILLE Robert	Construction mécanique (IUT I)
	SOUTIF Michel	Physique générale
	TANCHE Maurice	Physiologie
	TRAYNARD Philippe	Chimie générale

MM.	VAILLANT François	Zoologie
	VALENTIN Jacques	Physique Nucléaire
	VAUQUOIS Bernard	Calcul électronique
Mme	VERAIN Alice	Pharmacie galénique
MM.	VERAIN André	Physique
	VEYRET Paul	Géographie
	VIGNAIS Pierre	Biochimie médicale

PROFESSEURS ASSOCIES

MM.	CRABBE Pierre	CERMO
	DEMBICKI Eugéniuz	Mécanique
	JOHNSON Thomas	Mathématiques appliquées
	PENNEY Thomas	Physique

PROFESSEURS SANS CHAIRE

Mle	AGNIUS-DELORD Claudine	Physique pharmaceutique
	ALARY Josette	Chimie analytique
MM.	AMBROISE-THOMAS Pierre	Parasitologie
	ARMAND Gilbert	Géographie
	BENZAKEN Claude	Mathématiques appliquées
	BIAREZ Jean-Pierre	Mécanique
	BILLET Jean	Géographie
	BOUCHET Yves	Anatomie
	BRUGEL Lucien	Energétique (IUT I)
	BUISSON René	Physique (IUT I)
	BUTEL Jean	Orthopédie
	COHEN ADDAD Pierre	Spectrométrie physique
	COLOMB Maurice	Biochimie
	CONTE René	Physique (IUT I)
	DELOBEL Claude	M.I.A.G.
	DEPASSEL Roger	Mécanique des fluides
	FONTAINE Jean-Marc	Mathématiques Pures
	GAUTRON René	Chimie
	GIDON Paul	Géologie et Minéralogie
	GLENAT René	Chimie organique
	GROULADE Joseph	Biochimie médicale
	HACQUES Gérard	Calcul numérique
	HOLLARD Daniel	Hématologie
	HUGONOT Robert	Hygiène et Médecine préventive
	IDELMAN Simon	Physiologie animale
	JOLY Jean-René	Mathématiques Pures
	JULLIEN Pierre	Mathématiques Appliquées
Mme	KAHANE Josette	Physique
MM.	KRAKOWIACK Sacha	Mathématiques Appliquées
	KUHN Gérard	Physique (IUT I)
	LUU DUC Cuong	Chimie organique
	MAYNARD Roger	Physique du solide
Mme	MINIER Colette	Physique (IUT I)
MM.	PELMONT Jean	Biochimie
	PERRIAUX Jean-Jacques	Géologie et Minéralogie
	PFISTER Jean-Claude	Physique du solide
Mle	PIERY Yvette	Physiologie animale

MM.	RAYNAUD Hervé	M.I.A.G.
	REBECQ Jacques	Biologie (CUS)
	REYMOND Jean-Charles	Chirurgie générale
	RICHARD Lucien	Biologie végétale
Mme	RINAUDO Marguerite	Chimie macromoléculaire
MM.	ROBERT André	Chimie papetière
	SARRAZIN Roger	Anatomie et chirurgie
	SARROT-REYNAULD Jean	Géologie
	SIROT Louis	Chirurgie générale
Mme	SOUTIF Jeanne	Physique générale
MM.	STIEGLITZ Paul	Anesthésiologie
	VIALON Pierre	Géologie
	VAN CUTSEM Bernard	Mathématiques Appliquées

MAITRES DE CONFERENCES ET MAITRES DE CONFERENCES AGREGES

MM.	ARMAND Yves	Chimie (IUT I)
	BACHELOT Yvan	Endocrinologie
	BARGE Michel	Neuro chirurgie
	BEGUIN Claude	Chimie organique
Mme	BERIEL Hélène	Pharmacodynamie
MM.	BOST Michel	Pédiatrie
	BOUCHARLAT Jacques	Psychiatrie adultes
Mme	BOUCHE Liane	Mathématiques (CUS)
MM.	BRODEAU François	Mathématiques (IUT B) (Personne étrangère habilitée à être directeur de thèse)
	CHAMBAZ Edmond	Biochimie médicale
	CHAMPETIER Jean	Anatomie et organogénèse
	CHARDON Michel	Géographie
	CHERADAME Hervé	Chimie papetière
	CHIAVERINA Jean	Biologie appliquée (EFP)
	CONTAMIN Charles	Chirurgie thoracique et cardio-vasculaire
	CORDONNIER Daniel	Néphrologie
	COULOMB Max	Radiologie
	CROUZET Guy	Radiologie
	CYROT Michel	Physique du solide
	DENIS Bernard	Cardiologie
	DOUCE Roland	Physiologie végétale
	DUSSAUD René	Mathématiques (CUS)
Mme	ETERRADOSSI Jacqueline	Physiologie
MM.	FAURE Jacques	Médecine légale
	FAURE Gilbert	Urologie
	GAUTIER Robert	Chirurgie générale
	GIDON Maurice	Géologie
	GROS Yves	Physique (IUT I)
	GUIGNIER Michel	Thérapeutique
	GUITTON Jacques	Chimie
	HICTER Pierre	Chimie
	JALBERT Pierre	Histologie
	JUNIEN-LAVILLAVROY Claude	O.R.L.
	KOLODIE Lucien	Hématologie
	LE NOC Pierre	Bactériologie-virologie
	MACHE Régis	Physiologie végétale
	MAGNIN Robert	Hygiène et médecine préventive
	MALLION Jean-Michel	Médecine du travail

MM.	MARECHAL Jean	Mécanique (IUT I)
	MARTIN-BOUYER Michel	Chimie (CUS)
	MICHOULIER Jean	Physique (IUT I)
	NEGRE Robert	Mécanique (IUT I)
	NEMOZ Alain	Thermodynamique
	NOUGARET Marcel	Automatique (IUT I)
	PARAMELLE Bernard	Pneumologie
	PECCOUD François	Analyse (IUT B) (Personnalité étrangère habilitée à être directeur de thèse)
	PEFFEN René	Métallurgie (IUT I)
	PERRIER Guy	Géophysique-Glaciologie
	PHELIP Xavier	Rhumatologie
	RACHAIL Michel	Médecine Interne
	RACINET Claude	Gynécologie et Obstétrique
	RAMBAUD André	Hygiène et Hydrologie (Pharmacie)
	RAMBAUD Pierre	Pédiatrie
	RAPHAEL Bernard	Stomatologie
Mme	RENAUDET JAcqueline	Bactériologie (Pharmacie)
MM	ROBERT Jean-Bernard	Chimie Physique
	ROMIER Guy	Mathématiques (IUT B) (Personnalité étrangère habilité à être directeur de thèse)
	SCHAERER René	Cancérologie
	SHOM Jean-Claude	Chimie Générale
	STOEBNER Pierre	Anatomie Pathologie
	VROUSOS Constantin	Radiologie

MAITRESDE CONFERENCESASSOCIES

MM.	DEVINE Roderick	Spectro Physique
	HODGES Christopher	Transition de Phases

Fait à SAINT MARTIN D'HERES, NOVEMBRE 1976

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE ANNEE 1976-1977

Président : M. Philippe TRAYNARD

Vice-Présidents : M. Pierre-Jean LAURENT
M. René PAUTHENET

PROFESSEURS TITULAIRES

MM. BENOIT Jean	Radioélectricité
BESSON Jean	Electrochimie
BLOCH Daniel	Physique du solide
BONNETAIN Lucien	Chimie minérale
BONNIER Etienne	Electrochimie et Electrométallurgie
BRISSONNEAU Pierre	Physique du solide
BUYLE-BODIN Maurice	Electronique
COUMES André	Radioélectricité
DURAND Francis	Métallurgie
FELICI Noël	Electrostastique
FOULARD Claude	Automatique
LESPINARD Georges	Mécanique
MOREAU René	Mécanique
PARIAUD Jean-Charles	Chimie-Physique
PAUTHENET René	Physique du solide
PERRET René	Servomécanismes
POLOUJADOFF Michel	Electrotechnique
VEILLON Gérard	Informatique fondamentale et appliquée

PROFESSEURS SANS CHAIRE

MM. BLIMAN Samuel	Electronique
BOUVARD Maurice	Génie Mécanique
COHEN Joseph	Electrotechnique
LACOUME Jean-Louis	Géophysique
LANCIA Roland	Electronique
* ROBERT François	Analyse numérique
ZADWORNY François	Electronique
* ROBERT André	Chimie papetière

MAITRES DE CONFERENCES

MM. ANCEAU François	Mathématiques Appliquées
CHARTIER Germain	Electronique
GUYOT Pierre	Chimie Minérale
IVANES Marcel	Electrotechnique
JOUBERT Jean-Claude	Physique du solide
LESIEUR Marcel	Mécanique
MORET Roger	Electrotechnique Nucléaire
PIAU Jean-Michel	Mécanique
PIERRARD Jean-Marie	Mécanique
SABONNADIERE Jean-Claude	Informatique Fondamentale et Appliquée
MMe. SAUCIER Gabrièle	Informatique Fondamentale et Appliquée

CHERCHEURS DU C.N.R.S. (Directeur et Maîtres de Recherche).

M. FRUCHART Robert	Directeur de Recherche
MM. ANSARA Ibrahim	Maître de Recherche
CARRE René	Maître de Recherche
DRIOLE Jean	Maître de Recherche
LANDAU Ioan Doré	Maître de Recherche
MATHIEU Jean-Claude	Maître de Recherche
MUNIER Jacques	Maître de Recherche

Je voudrais exprimer ici mes remerciements

A Monsieur R. PERRET, Directeur du Laboratoire d'Automatique de Grenoble, qui m'a fait l'honneur de présider ce jury de thèse,

A Monsieur J. RICHALET, Directeur scientifique de l'ADERSA-GERBIOS, qui a été pendant longtemps mon "patron" de recherche. Son activité inlassable, la largeur et la diversité de ses vues, ont été pour moi un stimulant et un enrichissement permanents,

A Monsieur C. FOULARD, Professeur à l'INPG, qui a bien voulu être Directeur de ma thèse, et
A Madame G. SAUCIER, Maître de Conférences à l'INPG, qui m'a accepté dans son équipe. Tous deux m'ont cordialement reçu à mon arrivée à Grenoble et vivement encouragé à présenter cette thèse,

A Monsieur J. AGUILAR-MARTIN, Maître de Recherches au CNRS et
A Monsieur J.F. LEMAITRE, Directeur du CERT-DERA, qui ont bien voulu participer au jury,

A mes camarades du GERBIOS, notamment A. RAULT, P. LEROUX, A. MALLET, G. PREVOST, qui par leurs conseils et au cours de nombreuses discussions ont contribué à ce travail,

A mes camarades de l'ENSIMAG, en particulier C. BELLON et J. PULOU, pour notre sympathique collaboration,

Enfin, à Madame G. DUFFOURD et Madame B. HERNANDEZ, qui ont aimablement assuré la frappe de ce travail.

TABLE DES MATIERES

INTRODUCTION

I - IDENTIFICATION DES SYSTEMES ET ESTIMATION STATISTIQUE

I - 1 - Généralités sur l'identification
I - 2 - Introduction de l'aspect statistique
I - 3 - Cas du modèle instrument de mesure

II - UNE METHODE D'IDENTIFICATION DES SYSTEMES LINEAIRES INSTABLES

II - 1 - Position du problème
II - 2 - Caractérisation adaptée aux systèmes déterministes instables
II - 3 - Critères de moindre carré et de maximum de vraisemblance
II - 4 - Optimisation du critère
II - 5 - Cas des systèmes linéaires à équation d'état
II - 6 - Propriétés statistiques du critère et tests

III - PARALLELE AVEC LA METHODE DE MEHRA

III - 1 - Modèles déterministes et stochastiques
III - 2 - Méthode du maximum de vraisemblance de Mehra
III - 3 - Un exemple de modèle déterministe

IV - ETUDE DE LA CONVERGENCE DE L'ESTIMATEUR

IV - 1 - Introduction
IV - 2 - Convergence de la regression non linéaire
IV - 3 - Convergence de l'estimateur proposé
IV - 4 - Conclusion

V - ETUDE DE L'ESTIMATION DES COVARIANCES

V - 1 - Etude théorique
V - 2 - Cas de la régression non linéaire
V - 3 - Etude du modèle dynamique linéaire

VI - EXEMPLE D'APPLICATION

VI - 1 - Choix de l'exemple
VI - 2 - Rapport des variances connu
VI - 3 - Rapport des variances inconnu
VI - 4 - Cas du bruit agissant sur la boucle de retour

CONCLUSION

BIBLIOGRAPHIE

INTRODUCTION

Ce travail traite de l'identification des systèmes déterministes linéaires instables. Le problème nous a été posé de façon fortuite au terme d'une étude expérimentale sur un prototype de la SNIAS l'hydroptère H 890. Il s'agissait d'estimer les paramètres de cet engin afin de vérifier les calculs faits en bureau d'étude et d'améliorer le pilote automatique de ce système fortement instable. Cette étude qui s'est conclue sans aboutir à des conclusions complètes, nous a permis cependant de prendre conscience qu'il n'existait pas à cette époque (1974) de bonne solution à ce problème :

. la méthode du modèle est inopérante pour ce genre de problème étant donné les difficultés de simulation des systèmes instables, et le fait qu'on ne peut négliger le bruit sur les entrées ;

. la méthode des moindres carrés peut être appliquée mais dans des conditions très restrictives : il faut que le système soit linéaire stationnaire, que les paramètres à estimer soient les coefficients mêmes de la fonction de transfert et que toutes les variables d'état soient mesurées. Enfin, elle donne des résultats biaisés qui ne peuvent être corrigés que par la connaissance de covariances des bruits.

Or, dans le cas de l'hydroptère, si le système est bien linéaire stationnaire, les coefficients de la fonction de transfert sont des fonctions complexes non bijectives des paramètres de la mécanique du vol qu'il fallait estimer et toutes les variables d'état ne sont pas mesurées..

. Enfin une méthode générale de maximum de vraisemblance très séduisante proposée par R.K. MEHRA semblait pouvoir être appliquée : d'une part, elle traite de systèmes assez généraux : linéaires mais non forcément stationnaires et paramètres quelconques. D'autre part, elle conduit à remplacer la simulation du modèle par le filtrage de Kalman des données issues de l'objet à identifier. Or ce filtrage jouit de propriétés de stabilité numérique sous des conditions assez peu restrictives.

Malheureusement, cette méthode suppose que l'objet est un processus stochastique, excité par une entrée aléatoire inconnue appelée bruit de processus. Or, dans le cas de l'hydroptère il semblait que la source d'incertitude venait plus de la mauvaise qualité des mesures sur les entrées que d'un éventuel bruit de processus du type houle aléatoire, rafale de vent, etc.

C'est en discutant de cette question avec P. LEROUX que l'idée nous est venue conjointement qu'en remplaçant les bruits de processus par des bruits de mesure sur les entrées, on devait parvenir à une formulation assez voisine de celle de MEHRA.

Après avoir présenté au chapitre I les généralités sur l'identification, c'est cette approche qui est développée au chapitre II. Deux idées directrices ont été suivies dans ce développement :

Premièrement il nous a semblé important de ne pas se restreindre à présenter une simple méthode d'estimation de paramètre, mais plutôt une méthode complète d'identification de systèmes. On entend par là une méthode permettant de formuler un modèle (étape de caractérisation), d'en estimer les paramètres de même que leur domaine d'incertitude (estimation) et d'éprouver la validité du modèle en fonction de connaissances sur les propriétés statistiques des bruits (test de caractérisation). C'est pourquoi, dans cette étude nous avons étendu à cette méthode, les tests statistiques de caractérisation et de définition du domaine d'incertitude des paramètres tests classiques dans le cas de la méthode du modèle, et sur lesquels nous avions travaillé précedemment.

D'autre part, et contrairement à une tendance existant en automatique, nous avons essayé afin de donner la portée la plus générale à cette méthode, de contraindre au minimum la nature du modèle envisagé. La seule contrainte nécessaire à l'application du filtrage de Kalman est que le système soit linéaire à équation d'état. Il peut ne pas être stationnaire et les paramètres sont de type général.

S'agissant d'une méthode voisine de celle de Mehra, il nous a semblé intéressant de proposer au chapitre III un parallèle

entre ces deux méthodes. Les différences que l'on peut noter sont les suivantes :

- nous traitons de systèmes déterministes alors que Mehra traite de processus stochastiques.

- le filtrage de Kalman nous sert à résoudre un problème de commande optimal par programmation dynamique, alors que pour Mehra il s'agit d'expliciter la fonction de vraisemblance des données.

- enfin, les méthodes de test sont différentes.

De plus, face à la tendance générale en automatique de proposer des modèles de type stochastique, il nous a paru important de montrer que notre méthode qui traite de systèmes déterministes pouvait trouver des terrains d'application. Nous décrivons sommairement une étude de modèlisation du cheptel bovin français qui nous semble être le type d'étude où la méthode peut s'appliquer.

La question étudiée au chapitre IV est celle de la qualité de la méthode d'estimation proposée. Cette qualité ne pouvant se juger qu'asymptotiquement nous avons étudié la convergence presque sûre de la méthode proposée. Cette question méritait d'être posée car contrairement au cas habituel, le nombre d'inconnues de la fonction de vraisemblance croît dans notre cas vers l'infini en même temps que le nombre de données.

Cette question, d'autre part, se traite habituellement en démontrant la convergence uniforme presque sûre du critère d'identification lorsque le nombre de données tends vers l'infini, et ce genre de démonstration peut être délicat à établir. Par souci d'établir un outil général qui resservira au chapitre V, nous montrons que l'on peut se contenter de convergences simples en utilisant le théorème d'Egorof classique en analyse.

Une difficulté qui se présente dans l'application de la méthode proposée est qu'elle nécessite la connaissance des covariances des bruits d'entrée et de sortie, du moins à un coefficient proportionnel près. S'agissant de bruits de mesure, on peut penser que dans la

plupart des cas, une analyse de la méthode de mesure et des expériences appropriées peuvent fournir des estimations de ces covariances. Néanmoins des cas peuvent se présenter où celà n'est pas possible. Un problème analogue de régression non-linéaire nous a été posé par des biologistes cherchant à estimer des paramètres de cinétique enzimatique par identification de données expérimentales à l'aide de la loi de Michaelis-Menten. Lorsqu'on leur posait la question des bruits, ils nous avouaient que sans doute les deux variables mesurées étaient entachées d'erreurs sans que l'on puisse préciser davantage le rapport des variances de ces deux bruits. Ne connaissant pas de solution à ce problème, nous leur proposions alors des identifications obtenues en ne considérant de bruit que sur une seule des deux variables. Cela nous a conduit à étudier ce problème d'estimation des paramètres et des variances des bruits pour les deux modèles : régression non linéaire et système dynamique linéaire, au chapitre V. Une méthode semble-t-il originale de maximum de vraisemblance enrichi est proposée pour les résoudre et ses conditions de convergence asymptotique établies.

Enfin, nous avons complété ce travail par une étude en simulation des méthodes proposées. Cette étude fait l'objet du chapitre VI. L'objet choisi est un système linéaire stationnaire mono variable d'ordre 2, instable. Son identification est menée à bien par la méthode du chapitre II lorsque le rapport des variances des bruits est connu, par celle du chapitre V dans le cas contraire.

I - IDENTIFICATION DES SYSTEMES ET ESTIMATION STATISTIQUE

I - 1 - Généralités sur l'identification

Il ne s'agit pas ici de développer in extenso, l'identification qui est un vaste sujet ayant fait l'objet de quelques ouvrages de référence, par exemple Richalet, Rault et Pouliquen [1], mais d'en esquisser simplement la trame.

On considère un objet inconnu O dont on mesure un certain nombre N de couples d'entrées, sorties (u_n, s_n) $n = 1, N$ (figure 1.1).

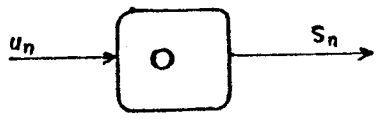


Figure 1.1

Les notations sont les suivantes :

(x_n) $n = 1, N$ ou plus simplement (x_n) lorsque les bornes sont sous-entendues, désigne la séquence $x_1, x_2, \ldots x_n$. On représentera aussi cette séquence par le symbole X_1^N.

Les systèmes considérés sont en général temporels (Sj est observé postérieurement à Si si j > i) et de type causal. Enfin les entrées u_n et sorties s_n à l'instant courant n sont des variables vectorielles réelles. Ne désignera dans la suite la dimension du vecteur d'entrée et Ns celle du vecteur de sortie.

Identifier ce système, c'est trouver un système M dit modèle de O qui, excité par la séquence d'entrées U_1^N fournisse une séquence de sorties $S'^N_1 = (s'_n)$ suffisamment voisine de S_1^N. Les étapes de cette identification sont les suivantes.

I - 1 - 1 - Caractérisation

Il s'agit de trouver, à partir d'une analyse physique biologique, économique, etc. du système O une formulation mathématique qui paraisse convenir à la représentation de ce système. Cette représentation sera du type :

$$s'_n = f_n\left(U_1^n, P\right)$$

Dans cette formulation, U_1^n représente la séquence d'entrées tronquée à l'instant n (hypothèse de causalité).

P est un vecteur en général réel de dimension p qui représente les inconnues du problème. Ces inconnues peuvent être de deux types :

- les paramètres proprement dits : ces paramètres ont pour effet de ne pas figer à priori la représentation choisie et donc de pouvoir l'ajuster, en les modifiant à la realité ;

- l'état initial du système modèle, lorsqu'il s'agit de systèmes dynamiques. p étant fini on s'interdit par là de considérer des modèles ayant une infinité de variables d'état.

Au cours de ces généralités, on ne distinguera pas paramètres et états. Lorsque cette distinction sera nécessaire (chapitre II), P désignera les paramètres (dimension p) et X_0 l'état initial (dimension No). Le modèle s'écrira alors :

$$s'_n = f_n\left(U_1^n, X_0, P\right)$$

I - 1 - 2 - Estimation des paramètres

Le modèle s'écrit donc $s'_n = f_n\left(U_1^n, P\right)$ et n'est donc pas entièrement défini puisque P est inconnu. Pour utiliser ce modèle, il convient de déterminer une valeur $\hat{P}$ convenable, et, comme il s'agit d'obtenir une analogie de comportement entre les systèmes O et M, $\hat{P}$ s'obtiendra en minimisant un critère (appelé improprement distance objet-modèle) mesurant l'écart entre S_1^N et S'^N_1. C'est donc une fonction D (P) qui pourra être du type :

$$D(P) = \max_n \| s_n - f_n(U_1^n, P) \| \quad \text{(distance de Tchebicheff)}$$

$$D(P) = \sum_1^N \| s_n - f_n(U_1^n, P) \|$$

$$D(P) = \sum_1^N \| s_n - f_n(U_1^N, P) \|^2$$

où le symbole $\| \quad \|$ représente une norme dans l'espace vectoriel $\mathbb{R}^N$ des sorties.

Afin de pouvoir parler, en toute généralité de point minimum $\hat{P}$ et de distance minimum $D(\hat{P})$, on suppose fixée à priori un domaine $\mathcal{C}$ de l'espace paramétrique dans lequel on tolère les variations de P et sur lequel se fera la minimisation de D(P). Ce domaine sera choisi fermé, borné donc compact dans l'espace paramètrique $\mathbb{R}^p$. Si de plus on suppose D(P) continue en P sur $\mathcal{C}$, l'existence et l'unicité du minimum $D(\hat{P})$ sera établie.

En revanche, le point $\hat{P}$ peut n'être pas unique. Deux raisons peuvent expliquer cela :

a) Le modèle n'est pas identifiable : c'est le cas lorsque deux points de l'espace paramétrique fournissent le même modèle :

$$\exists P, P' \in \mathcal{C}, \forall U_1^N : \forall n \quad f_n(U_1^n, P) = f_n(U_1^n, P')$$

Il faut en général éviter de proposer des modèles non identifiables. C'est la raison des recherches de représentations minimales des systèmes, par exemple pour les systèmes linéaires stationnaires multivariables, les formes canoniques observables [3].

b) Même dans le cas de modèles identifiables, il peut n'y avoir pas de minimum unique, si la séquence d'entrée U_1^N ne sensibilise pas suffisamment les paramètres. D où les recherches sur la génération d'entrées sensibilisantes sur un horizon fini [1], ou lorsque l'horizon tend vers l'infini : entrées excitantes en permanence d'Aström et Bohlin (4).

Ces questions concernent le problème d'estimation ponctuelle. Une façon complémentaire de procéder consiste à fournir tout un domaine autour de $\hat{P}$ contenant des modèles voisins qui peuvent aussi convenir à la représentation de l'objet. On parlera alors d'estimation d'un domaine, par exemple les domaines isodistance de Richalet Rault et Pouliquen [1] définis par :

$$\left\{ P \ : \quad D(P) - D(\hat{P}) \leq \text{constante} \right\}$$

Une fois ce travail fait le problème se pose de savoir si la caractérisation adoptée et les paramètres $\hat{P}$ qui en découlent satisfont le désir que l'on a de représenter le système objet : c'est l'étape de validation.

I - 1 - 3 - Validation du modèle

C'est un sujet complexe qui dépend du degré d'analogie que l'on souhaite entre l'objet et le modèle.

Un premier degré de validation consiste à examiner la distance minimum $D(\hat{P})$ obtenue, ou bien directement la séquence $\hat{S}_1^N$ en la comparant à la séquence objet.

Un second degré consiste à stimuler l'objet et le modèle par une nouvelle séquence d'entrées U_1^N , de façon à apprécier les écarts de comportement qui en découlent.

D'autres types de vérification peuvent porter sur la vraissemblance physique ou autre des valeurs de paramètres obtenus.

Dans tous les cas, une insatisfaction de ces conditions conduit à revenir sur l'étape de caractérisation et ainsi réitérer la procédure.

I - 2 - Introduction de l'aspect statistique

Dans ce cadre très général, il n'est pas fait d'hypothèses sur la structure réelle du système objet et en particulier, il n'est jamais dit que l'objet est régi par les équations du modèle. Au modèle choisi dit modèle de représentation on ne demande que de ressembler dans son comportement à l'objet.

Il en va tout autrement dans le cas particulier, mais fréquent où l'on affirme et l'on veut vérifier que l'objet est régi par des lois (physiques par exemple) de telle sorte que l'on puisse poser $s_n = f_n(U_1^n, P_o)$ où Po est un point bien particulier de l'espace paramétrique. Si tel est le cas, l'affirmation ne sera prouvée que lorsque le modèle aura fourni une séquence S'^N_1 identique à S^N_1. Le fait que cette identité ne se produise jamais revient à mettre en doute la possibilité de trouver les lois mathématiques des phénomènes observables. En réalité, dans ce problème dit du modèle de connaissance un certain nombre de facteurs ont été oubliés qui rendent l'identité effectivement impossible à atteindre. Ce sont les bruits et autres causes influant sur l'incertitude des mesures, ainsi que les phénomènes secondaires indépendants du phénomène principal que l'on cherche à cerner et qui le perturbent. Ces facteurs qui font que la répétition de deux expériences identiques ne fournit pas les mêmes résultats conduisent à poser l'hypothèse que l'on veut vérifier sous la forme :

S^N_1 est une réalisation de la séquence aléatoire S'^N_1 telle que :

$$s'_n = f_n(U_1^n, P_o)$$

où Po appelé point nominal est un point déterminé mais inconnu de l'espace paramétrique.

et où f_n est un vecteur aléatoire fonction de U_1^n et de Po.

I - 2 - 1 - Caractérisation

Ainsi, l'étape de caractérisation conduit à formuler mathématiquement la loi de probabilité de la séquence S'^N_1.

I - 2 - 2 - Estimation

L'estimation des paramètres relève de méthodes statistiques : ces méthodes peuvent être classées en deux catégories :

a) Les méthodes particulières qui ne peuvent être formulées que pour un type de problème bien particulier. Un grand nombre de méthodes ont été trouvées dans le cas des modèles linéaires stationnaires lorsque les paramètres sont les coefficients de la fonction de transfert en Z du système. Ces méthodes peuvent être souvent classées sous le nom de méthodes d'équation d'erreur dans la terminologie de Mehra [5]. On peut citer parmi ces méthodes, la méthode des moindres carrés (qu'il ne faut pas confondre avec la méthode des moindres carrés de type général formulée en b) la méthode de variables instrumentales [6], la méthode de LEVIN [7] généralisée par GROSJEAN [8] etc.

Dans cette classification les méthodes d'estimation des paramètres en temps réel ne sont pas des méthodes particulières. En effet, il simble que l'on puisse toujours les ramener à une minimisation en temps réel de critères en temps différé, à l'aide d'approximations.

- au premier ordre dans le cas des méthodes d'approximation stochastique [9] ;

- au deuxième ordre dans le cas du filtre de Kalman étendu [10].

b) Les méthodes de formulation générale

- Méthode des moindres carrés (Gauss) ou d'erreur de sortie d'après Mehra [5].

Il s'agit de minimiser la distance

$$D_{MC}(P) = \sum_1^n \left\{ s_n - E\left[f_n(U_1^n, P)\right]\right\}^T C_n^{-1} \left\{ s_n - E\left[f_n(U_1^n, P)\right]\right\}$$

où E représente l'espérance mathématique et

et C_n est une matrice carrée Ns x Ns symétrique défini positive, covariance de S'n

$$C_n = E\left\{ \left[f_n(U_1^n, P) - E\left[f_n(U_1^n, P)\right]\right]\left[f_n(U_1^n, P) - E\left[f_n(U_1^n, P)\right]\right]^T \right\}$$

On peut remarquer que la méthode des moindres carrés n'exige pas pour sa mise en oeuvre la connaissance de la loi de probabilité de S'n mais uniquement celle de ses deux premiers moments.

- Méthode de maximum de vraisemblance (Fisher)

Soit $L(S'^N_1 ; U^N_1, P)$ la densité de probabilité de la suite aléatoire S'^N_1 (si cette densité existe). La méthode du maximum de vraisemblance consiste à choisir P de telle sorte que la densité de probabilité pour la réalisation S^N_1 soit maximum. Cela revient à minimiser la "distance"

$$D_{MV}(P) = - \log\left[L(S^N_1 ; U^N_1, P)\right]$$

- Méthode d'estimation Bayesienne

Controversée dans les milieux statistiques, cette méthode suppose que le vecteur paramètre P est aléatoire et qu'une réalisation Po a été choisie pour fournir les résultats de l'expérience considérée. On est donc conduit à introduire la densité de probabilité de P p(P), et la "distance" devient :

$$D_B(P) = - \log\left[L(S^N_1 ; U^N_1, P) \cdot p(P)\right]$$

Si P est complètement incertain (loi uniforme surtout l'espace paramétrique) on retrouve la méthode précédente.

Domaine de confiance des paramètres

Nous avons vu que l'estimation pouvait aboutir non pas à un point $\hat{P}$ mais à tout un domaine (isodistance par exemple). Dans le cas statistique cette notion se précise. En effet, quelle que soit la méthode d'estimation choisie, l'optimum trouvé $\hat{P}$est fonction de la réa-

lisation S_1^N. C'est donc un vecteur aléatoire, dit estimateur dont on peut étudier des propriétés telles que :

Moyenne $M = E(\hat{P})$

Covariance $C = E\{[\hat{P} - E(\hat{P})][\hat{P} - E(\hat{P})]^T\}$

Ellipsoïde d'inertie :

$\{X : X^T C^{-1} X \leq 1\}$

Ces propriétés peuvent s'étudier par rapport au point Po supposé connu : on introduira alors le biais Po - M.

Cette question peut aussi s'appréhender par la méthode des tests statistiques. Elle conduira à interroger chaque point P de l'espace nominal Po, et à opposer à cette hypothèse simple P = Po, l'hypothèse complexe $P \neq Po$. Pour un seuil de signigication α donné (en général $\alpha = 0.05$) l'ensemble des points P tels que l'hypothèse simple P = Po ne peut être rejetée, forme le domaine de confiance à $(1 - \alpha)$ de Po.

Il faut remarquer que lorsque la méthode de test utilisée est celle du rapport de vraisemblance (Neyman-Pearson), le domaine de confiance s'exprime de façon très simple :

En effet, l'hypothèse simple P = Po ne pourra être rejetée au seuil α si

$$\frac{L(S_1^N ; P)}{\max_P L(S_1^N ; P)} \geqslant k(\alpha) \qquad 0 \leqslant k(\alpha) \leqslant 1$$

Soit $\hat{P}$ l'estimée maximum de vraisemblance . On aura aussi

$$- \mathrm{Log}\left[L(S_1^N ; P)\right] \leqslant - \log\left[L(S_1^N ; \hat{P})\right] - \mathrm{Log}\left[k(\alpha)\right]$$

Le domaine de confiance s'exprime donc comme un domaine iso distance :

$$\left\{ P : \quad D(P) \leqslant D(\hat{P}) - \mathrm{Log}\left[k(\alpha)\right] \right\}$$

En revanche la détermination de k (α) est souvent complexe.

I - 2 - 3 - Validation du modèle

Enfin, l'étape de validation du modèle se pose aussi en termes de tests statistiques. Il conviendra d'éprouver l'hypothèse complexe: S_1^N est une réalisation de $F_1^N(U_1^N, P_0)$ (Po inconnu) contre toute autre hypothèse également complexe qui la nie. Dand cette optique un modèle n'est jamais définitivement accepté, ce qui est heureux. La validation n'est que provisoire jusqu'à ce qu'un test vienne le rejeter.

I - 3 - Cas du modèle instrument de mesure

Même dans le cas d'un modèle de représentation, l'aspect statistique s'introduit du fait des bruits qui rendent S_1^N aléatoire. Il en résulte alors que quelque soient les distances utilisées, $\hat{P}$ et $D(\hat{P})$ sont aléatoires puisque fonction de S_1^N. L'étude de leurs propriétés statistiques est importante lorsque :

- Le modèle sert d'instrument de mesure des paramètres, par exemple en vue d'un diagnostic de systèmes. Un exemple d'une telle étude peut être trouvé en [2].

- On veut prévoir quel sera l'écart de comportement objet modèle au cours d'une utilisation ultérieure.

Cette étude est en principe difficile puisque la loi de probabilité de S'^N_1 n'est pas précisée. Elle devient possible lorsque cette loi est partiellement connue, par exemple lorsque l'on connait la loi de $S'^N_1 - E\,(S'^N_1)$.

II - UNE METHODE D'IDENTIFICATION DES SYSTEMES LINEAIRES INSTABLES

II - 1 - Position du problème

II - 1 - 1 - Identification d'un système déterministe par une méthode statistique

Un problème fréquent d'identification porte sur les systèmes déterministes. Un système objet O est considéré comme déterministe si l'on peut supposer que la séquence d'entrées U_1^N détermine parfaitement la séquence de sortie S_1^N.

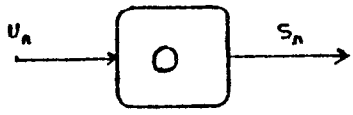


figure 2.1

Cela revient à supposer que les entrées du système prises en compte représentent complètement les excitations auxquelles est soumis le système.

Ce système peut être identifié par une méthode purement déterministe (I - 1) : On proposera un modèle:

$$s'_n = f_n(U_1^n, P)$$

et une distance:

$$D_D(P) = \sum_1^N \| s_n - f_n(U_1^n, P) \|^2 \qquad \text{par exemple.}$$

Néanmoins, nous avons vu en I.2 qu'il y a pour des raisons à la fois théoriques et pratiques intérêt à proposer un modèle aléatoire

menant à une méthode d'identification statistique.

La façon la plus simple de se placer dans un cadre aléatoire est de supposer que la sortie Sn de l'objet n'est pas mesurée parfaitement, mais entachée d'erreurs de mesure :

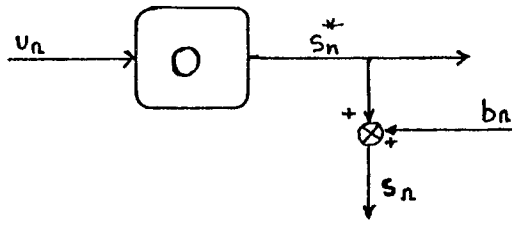


figure 2.2

Le modèle aléatoire que l'on peut proposer du système s'écrit alors

$$s'_n = f_n(U_1^n, P) + b_n$$

où b_n est un bruit de mesure. Le modèle doit être complété par la donnée de la loi de probabilité de la séquence $B_1^N = (b_n)$. L'hypothèse la plus simple consiste à définir B_1^N comme une séquence de vecteurs indépendants, stationnaires, de moyenne nulle et de covariance finie :

$$\forall n \ , \quad E(b_n) = 0$$

$$\forall n, m \ , \quad E(b_n b_m^T) = Q\,\delta_{n,m}$$

où Q est une matrice Ns x Ns symétrique définie positive (Ns est la dimension de Sn) et $\delta n,m$ est le symbole de Kronecker.

La méthode des moindres carrés fournit alors la distance

$$D_{MC}(P) = \sum_1^N \left(s_n - f_n(U_1^n, P)\right)^T Q^{-1} \left(s_n - f_n(U_1^n, P)\right)$$

Si l'on suppose de plus que le bruit est gaussien et que Q est connu, $D_{mc}(P)$ est aussi le critère résultant de la méthode de maximum de vraisemblance . Dans la méthode M.C. Q doit être connu à un facteur multiplicatif près.

Q peut aussi être inconnue. la vraisemblance des données s'écrit alors :

$$L(S_1^N ; P, Q) = \frac{e^{-\sum_1^N (s_n - f_n(U_1^n, P))^T \frac{Q^{-1}}{2} (s_n - f_n(U_1^n, P))}}{(2\pi)^{\frac{N.N_s}{2}} . (\det(Q))^{\frac{N}{2}}}$$

D où la "distance" maximum de vraissemblance:

$$D_{MV}(P,Q) = - \operatorname{Log}\left[L(S_1^N ; P, Q) \right]$$

qui peut s'écrire en multipliant par $\frac{2}{N}$ et en supprimant le terme constant

$$D_{MV}(P,Q) = \frac{1}{N} \sum_1^N \left[s_n - f_n(U_1^n, P) \right]^T Q^{-1} \left[s_n - f_n(U_1^n, P) \right] + \operatorname{Log}\left[\det(Q)\right]$$

La minimisation de D_{mv} peut se faire en deux temps : minimisation par rapport à Q puis à P :

$$\min_{P,Q} \left[D_{MV}(P,Q) \right] = \min_P \left[\min_Q \left[D_{MV}(P,Q) \right] \right] = \min_P \left[D'_{MV}(P) \right]$$

La minimisation en Q s'obtient en écrivant:

$$\frac{\partial D_{MV}(P,Q)}{\partial Q} = 0$$

qui donne:

$$\hat{Q} = \frac{1}{N} \sum_1^N \left[s_n - f_n(U_1^n, P) \right] \left[s_n - f_n(U_1^n, P) \right]^T$$

D'où après élimination des termes constants

$$D'_{MV}(P) = \operatorname{Log}\left[\det\left[\hat{Q}(P) \right] \right]$$

<u>Remarques</u> : Il faut noter que la distance moindre carré est un cas particulier de la distance déterministe (norme particulière). On peut remarquer aussi l'intérêt de la méthode M.V. qui permet de trouver Q.

Minimisation des distances

Quelle que soit la distance choisie, la minimisation en P ne peut se faire, dans le cas général, que par programmation non linéaire (P.N.L.) selon l'organigramme suivant :

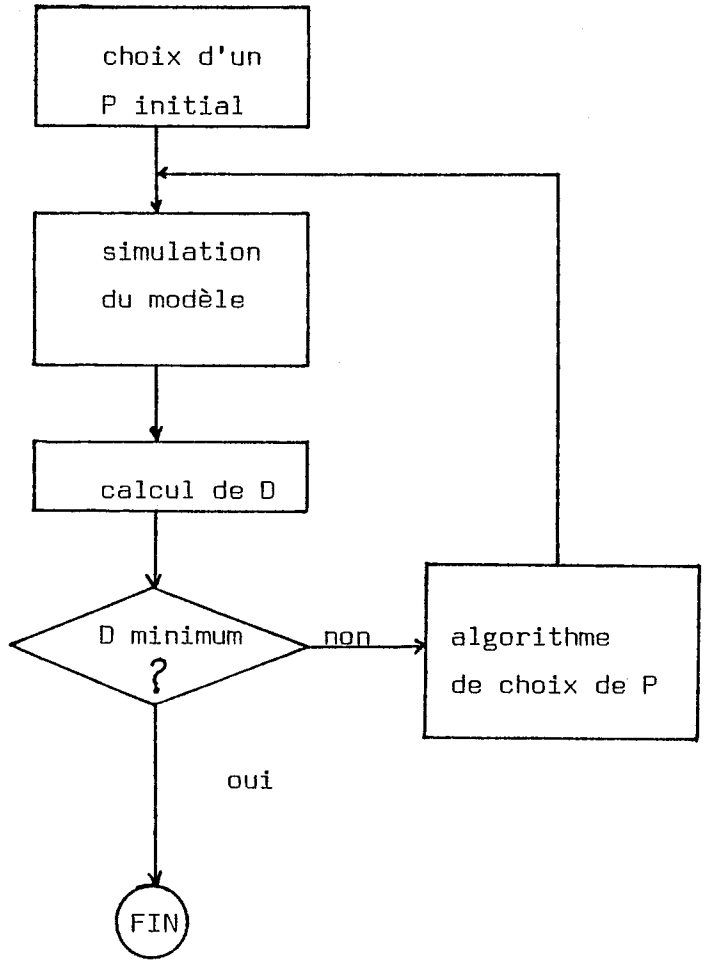


Figure 2.3

Dans certains cas : système représentés par une fonction de transfert en z des méthodes deminimisation de distance du type D ou MC ont été proposées ne faisant pas appel à la programmation non linéaire : méthode de Steigletz [11], moindres carrés généralisés de Clarke [12]. Enfin, des méthodes particulières [6], [7], [8] peuvent aussi s'appliquer.

II - 1 - 2 - Difficultés des systèmes instables

Lorsque le modèle est instable, la simulation du système au cours de l'algorithme de P.N.L. pose de délicats problèmes d'analyse numérique : ainsi les erreurs d'arrondi, loin de s'atténuer, perturbent de façon constante la sortie du modèle ou vont même en s'amplifiant. Il en est de même de la représentation digitale des entrées (celles ci devraient être connues de façon très précise). D'autre part les méthodes qui évitent la P.N.L. font explicitement référence à la stabilité asymptotique du système.

Comme cela est fréquent lorsque des problèmes d'analyse numérique se posent, la solution des difficultés ne doit pas être cherchée dans l'utilisation de méthodes d'analyse numérique plus perfectionnées mais plutôt dans une reformulation du problème qui bien souvent lève la difficulté. Dans le cas des systèmes instables une caractérisation tenant compte du fait que les entrées ne sont pas parfaitement mesurées permet de résoudre le problème [13].

II - 2 - Caractérisation adaptée aux systèmes déterministes instables

L'idée consiste donc à supposer que les entrées sont de la même façon que les sorties observées entachées de bruits de mesure, le système lui restant strictement déterministe :

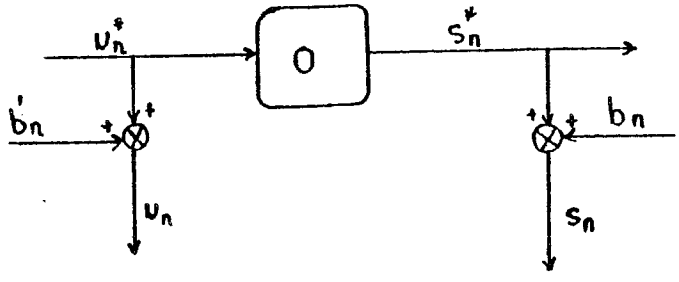


Figure 2.4

Le modèle sera donc :

$$s_n^* = f_n\left(U_1^{*n}, P\right)$$

$$s'_n = s_n^* + b_n$$

$$u'_n = u_n^* + b'_n$$

(b_n) et (b'_n) étant des séquences de bruit à moyenne nulle stationnaires indépendants et indépendants entre eux, de covariance finie.

- $E(b_n) = E(b'_n) = 0$
- $E(b_n\, b_m^T) = Q\,\delta_{n,m}$
- $E(b'_n\, b'^T_m) = R\,\delta_{n,m}$
- $E(b_n\, b'^T_m) = 0$

où Q est une matrice Ns x Ns et R une matrice Ne x Ne toutes deux symétriques définies positives.

II - 3 - Critères de moindre carré et de maximum de vrai - semblance

II - 3 - 1 - Critère de moindres carrés

Le critère ou distance de moindres carrés s'écrit :

$$D_{MC}(P) = \sum_1^N \left(u_n - u_n^*\right)^T R^{-1}\left(u_n - u_n^*\right) + \left(s_n - s_n^*\right)^T Q^{-1}\left(s_n - s_n^*\right)$$

Dans cette expression u_n^* et s_n^* sont liés par la relation:

$$s_n^* = f_n\left(U_1^{*n}, P\right)$$

- La distance ne peut être calculée que si l'on connait U_1^{*N} qui est inconnue. Il convient donc de l'introduire dans la liste des inconnues de la distance : $D_{MC}\left(U_1^{*N}, P\right)$

- La distance exige pour être calculée la connaissance de Q et R au moins à un facteur multiplicatif près. Dans le cas monovariable il s'agit de scalaires q et r dont il suffit de connaître le rapport.

II - 3 - 2 - Critère de maximum de vraisemblance

Si les bruits sont gaussiens et Q, R connus, la méthode de maximum de vraisemblance se ramène à celle des moindres carrés. De la même façon qu'au paragraphe II - 1 - 1 on peut se poser la question de savoir si la méthode de maximum de vraisemblance permet d'estimer les covariances Q et R. Nous allons voir que ce n'est pas le cas :

Il est facile de vérifier que la vraisemblance des données U_1^N, S_1^N (U_1^N est ici aléatoire) s'écrit :

$$L\left(U_1^N, S_1^N ; U_1^{*N}, P, Q, R\right) = \frac{e^{-D(U_1^{*N}, P)/2}}{(2\pi)^{N'} (\det Q)^{N/2} (\det R)^{N/2}}$$

où $N' = N\left(\frac{N_e + N_s}{2}\right)$

La maximisation de L en Q et R donne

$$\hat{Q} = \frac{1}{N} \sum_1^N \left(s_n - s_n^*\right)\left(s_n - s_n^*\right)^T$$

$$\hat{R} = \frac{1}{N} \sum_1^N \left(u_n - u_n^*\right)\left(u_n - u_n^*\right)^T$$

En reportant ces valeurs dans L il vient

$$L = \frac{e^{-N'}}{(2\pi)^{N'} (\det \hat{Q})^{N/2} (\det \hat{R})^{N/2}}$$

Deux solutions triviales se présentent alors qui rendent la vraisemblance infinie :

- $\hat{Q} = 0$ soit $S_1^{*N} = S_1^N$ (pas de bruit de sortie)
- $\hat{R} = 0$ soit $U_1^{*N} = U_1^N$ (pas de bruit d'entrée)

II - 3 - 3 - Conclusion

Il apparait donc que la méthode de maximum de vraisemblance n'apporte rien de plus par rapport à celle des moindres carrés. La connaissance des covariances des bruits de mesure parait donc nécessaire au moins à un facteur multiplicatif près. Etant donné qu'il s'agit de bruits de mesure, il n'y a pas là d'obstacle pratique à la mise en oeuvre de cette méthode : il devrait être toujours possible, avant toute expérience de mesurer l'incertitude due à la chaîne de mesures, bien que cette évaluation puisse parfois être difficile notamment en économie. Néanmoins, nous verrons au chapitre 5 une façon de modifier la méthode de maximum de vraisemblance afin de pouvoir estimer, au moins dans le cas mono entrée mono sortie ces covariances.

D'autre part, on peut remarquer que la distance trouvée peut être aussi interprétée dans un cadre purement déterministe : au lieu de chercher uniquement à rapprocher la sortie du modèle de la sortie objet, on cherche ici des paramètres et une entrée tels que l'entrée et la sortie du modèle soient le plus analogue à celles de l'objet.

II - 4 - Optimisation du critère

Le problème de la minimisation du critère se divise naturellement en deux parties :

- minimisation en U_1^{*N} : elle consiste à résoudre un problème de commande optimale à critère quadratique pour le système modèle de paramètre P fixé. On obtient ainsi le critère optimal

$$D'(P) = \min_{U_1^{*N}} \left[D(U_1^{*N}, P) \right]$$

- minimisation de la "distance objet-modèle" D'(P) qui est du domaine de la programmation non linéaire.

Dans le cas général le problème de commande optimale est difficile à résoudre. De la même façon que MEHRA en [5] on peut sans doute étendre la méthode que nous proposons en 2-5 pour les systèmes linéaires au cas non linéaire par filtrage de Kalman étendu. Le cas qui fournit les résultats les plus riches est le cas classique des systèmes linéaires à équation d'état que nous allons développer in extenso.

II - 5 - Cas des systèmes linéaires à équation d'état

II - 5 - 1 - Formulation du modèle et de la distance

La partie déterministe du modèle s'écrit alors

$$x_{n+1} = F_n(P)\, x_n + G_n(P)\, u_n^*$$

$$s_n^* = H_n(P)\, x_n$$

où x_n est le vecteur d'état du système, de dimension No

$F_n(P)$ est une matrice $N_o \times N_o$

$G_n(P)$ est une matrice $N_o \times N_e$

$H_n(P)$ est une matrice $N_s \times N_o$

Outre la connaissance de P et U_1^{*N}, la séquence de sortie du système dépend d'une inconnue de plus : l'état initial Xo. Xo sera distingué des paramètres de structure P du système pour plusieurs raisons : d'une part, la minimisation en Xo du critère sera intégrée

à la résolution du problème de commande optimale. D'autre part, Xo peut être :

- soit partiellement connu : la connaissance que l'on a de Xo peut s'exprimer dans une optique bayesienne par une distribution de probabilité à priori, gaussienne:

$$p(x_o) = \frac{e^{-(x_o - \bar{x}_o)^T \frac{C_o^{-1}}{2} (x_o - \bar{x}_o)}}{(2\pi)^{N_o/2} (\det C_o)^{1/2}}$$

où Co est une matrice symétrique définie positive No x No
$\overline{Xo}$ un vecteur No x 1

- soit parfaitement connu. Cela peut s'exprimer à la limite en posant Co = 0. La valeur de Xo est alors fixée à $\overline{Xo}$ et Xo doit être retiré de la liste des inconnues du critère. Le terme en Xo du critère est supprimé ;

- soit inconnu. Co est alors considéré comme infini. Xo est une inconnue du critère mais le terme en Xo du critère doit être supprimé.

Ces deux derniers cas sont donc des cas particuliers du premier pour lequel le critère s' écrit :

$$D(U_1^{*N}, P, x_o) = (x_o - \bar{x}_o)^T C_o^{-1} (x_o - \bar{x}_o)$$
$$+ \sum_1^N (s_n - s_n^*)^T Q^{-1} (s_n - s_n^*) + (u_n - u_n^*)^T R^{-1} (u_n - u_n^*)$$

Le problème de commande optimale revient donc à chercher le critère optimal:

$$D'(P) = \min_{U_1^{*N}, x_o} \left[D(U_1^{*N}, x_o, P) \right]$$

II - 5 - 2 - <u>Optimisation quadratique par filtrage de Kalman</u>

La façon la plus convenable de procéder pour résoudre ce pro-

blème de commande optimale consiste à utiliser la programmation dynamique. En effet, en chaque point P, il n'est pas utile d'exhiber la commande $\hat{U}_1^{*N}$ et l'état initial $\hat{X}o$ optimaux car il suffit de connaître le critère optimal D'(P) pour le transmettre au programme de recherche non linéaire. Ce n'est que lorsque $\hat{P}$ a été trouvé que la connaissance de $\hat{U}_1^{*N}$ et $\hat{X}o$ peut être utile.

Les travaux de FAGEN [14] dans le cas discret et COX [15] dans le continu, dont on trouvera l'exposé dans JAZWINSKI [16] ont établi l'équivalence entre la programmation dynamique et la théorie du filtrage optimal exposée par KALMAN en [17]. Il faut noter que dans cet article Kalman avait déjà envisagé l'application de sa théorie du filtrage aux problèmes de commande optimale sous forme d'un théorème de dualité.

Description du filtre de Kalman

Le filtre de Kalman associé à un système tel que celui formulé en II- 5 - 1 est un filtre linéaire déterminé par la donnée d'un sextuplet: FK (Fn, Gn, Hn, Q, R, Co). Les entrées sont des couples (Sn, Un). La sortie principale est la séquence d'états estimés $(\hat{X}n/n)$. D'autres sorties possibles sont la séquence des sorties des sorties prédites $(\hat{S}n/n - 1)$, la séquence d'innovations ([18]) $(Sn - \hat{S}n/n - 1)$, ou la séquence de résidus (εn) qui en est dérivée. Enfin, ces sorties dépendent d'un état initial $\overline{Xo}$.

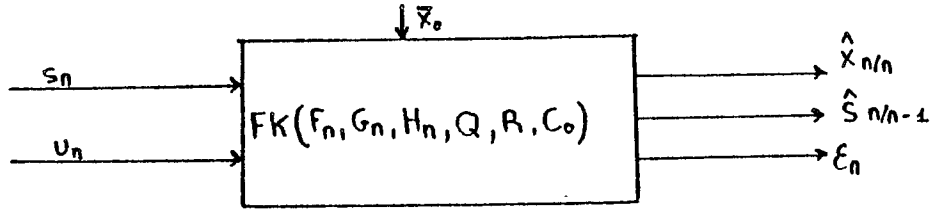


Figure 2.5

Les équations du filtre sont les suivantes :

Prédiction

$$\hat{X}_{n+1/n} = F_n \hat{X}_{n/n} + G_n U_n$$

$$C_{n+1/n} = F_n C_{n/n} F_n^T + G_n R G_n^T$$

Filtrage

$$K_n = C_{n/n-1} H_n^T \left[Q + H_n C_{n/n-1} H_n^T \right]^{-1}$$

$$\hat{X}_{n/n} = \hat{X}_{n/n-1} + K_n \left[S_n - H_n \hat{X}_{n/n-1} \right]$$

$$C_{n/n} = C_{n/n-1} \left[I - H_n^T K_n^T \right]$$

Dans ces expressions

- $\hat{X}_{n/n-1}$ et $\hat{X}_{n/n}$ sont des vecteurs de dimension No

- $C_{n/n}$ et $C_{n/n-1}$ sont des matrice symétriques $N_o \times N_o$ et I est la matrice identité d'ordre No

- K_n est une matrice (Gain de KALMAN) No x Ns

Conditions initiales

$$C_{o/o} = C_o$$

$$\hat{X}_{o/o} = \bar{X}_o$$

Lorsque l'on suppose Co infini, la valeur initiale de $\hat{X}_{o/o}$ est indiférente. Une façon de procéder consiste à poser $C_o = \alpha I$ α étant grand. Une autre façon de faire est la méthode de ALBERT [19] qui consiste à claculer une pseudo inverse exacte tant que le nombre d'inconnues dépasse le nombre de données.

Résidus

Les autres sorties du filtre sont :

- la sortie prédite $\hat{S}_{n/n-1} = H_n \hat{X}_{n/n-1}$

- les innovations $S_n - \hat{S}_{n/n-1}$

- les résidus ε_n qui se déduisent des innovations par transformation linéaire :

$$\varepsilon_n = T_n \left[S_n - \hat{S}_{n/m-1} \right]$$

où T est une matrice Ns x Ns obtenue par factorisation de la matrice symétrique positive $\left[Q + H_n C_{m/n-1} H_n^T \right]^{-1}$:

$$T_n^T T_n = \left[Q + H_n C_{n/n-1} H_n^T \right]^{-1}$$

<u>Remarque</u> : Une autre façon de procéder, qui évite au cours du filtrage l'inversion de la matrice $Q + H_n C_{n/n-1} H_n^T$ est rendue possible lorsque Q est une matrice diagonale ce qui est fréquemment le cas pour des bruits de mesure : bruits indépendants sur chaque voie de mesure des sorties. Cette méthode consiste au lieu de corriger $\hat{X}n/ n - 1$ à l'aide de la donnée Sn en un seul temps (filtrage), à répéter cette opération Ns fois en introduisant successivement les diverses composantes de Sn et en corrigeant $\hat{X}_{n/n-1}, S_n^1, S_n^2 \dots S_n^i$ à l'aide de la composante S_n^{i+1} de façon à obtenir $\hat{X}_{n/n-1}, S_n^1, \dots S_n^{i+1}$. Au terme de ce processus on obtiendra :

$$\hat{X}_{n/n-1}, S_n^1, \dots S_n^{N_s} = \hat{X}_{n/m}$$

L'intérêt de cette méthode est que chaque correction est scalaire et se fait sans inversion de matrice. Les résidus obtenus sont différents mais jouissent des mêmes propriétés.

<u>Equivalence entre programmation dynamique et filtrage de Kalman</u>

Celle-ci peut s'énnoncer sous la forme du théorème suivant dont nous ne donnerons pas la démonstration classique.

<u>Théorème 2- 5 - 2</u>

. Le critère optimal correspondant à la minimisation de

$D(U_1^{*N}, x_0, P)$ par rapport à U_1^{*N}, x_0 s'écrit:

$$D'(P) = \sum_1^N \left(S_n - \hat{S}_{n/n-1} \right)^T \left[Q + H_n C_{n/n-1} H_n^T \right]^{-1} \left(S_n - \hat{S}_{n/n-1} \right)$$

$$= \sum_1^N \varepsilon_n^T(P) . \varepsilon_n(P)$$

Ce critère peut donc se calculer par filtrage de Kalman des données $[S_1^N, U_1^N]$ à l'aide du F.K. $\left[F_n(P), G_n(P), H_n(P), Q, R, C_o\right]$ initialisé à l'état $\overline{X_o}$.

. La séquence d'entrée optimale $\hat{U}_1^{*N}$ et l'état initial $\hat{X}_o$ optimal correspondant au critère optimal D'(P) s'obtiennent en procédant au lissage inverse d'équations :

$$\hat{X}_{n/N} = \hat{X}_{n/n} + C_{n/n} F_n^T C_{n+1/n}^{-1} \left(\hat{X}_{n+1/N} - \hat{X}_{n+1/n}\right)$$

$$\hat{U}_n^* = U_n + R\, G_n^T C_{n+1/n}^{-1} \left(\hat{X}_{n+1/N} - \hat{X}_{n+1/n}\right)$$

n variant de N à 0.

$$\hat{X}_o = \hat{X}_{o/N}$$

II - 5 - 3 - <u>Mise en oeuvre et optimisation non linéaire</u>

La minimisation de D'(P) s'obtient par programmation non linéaire selon l'organigramme suivant :

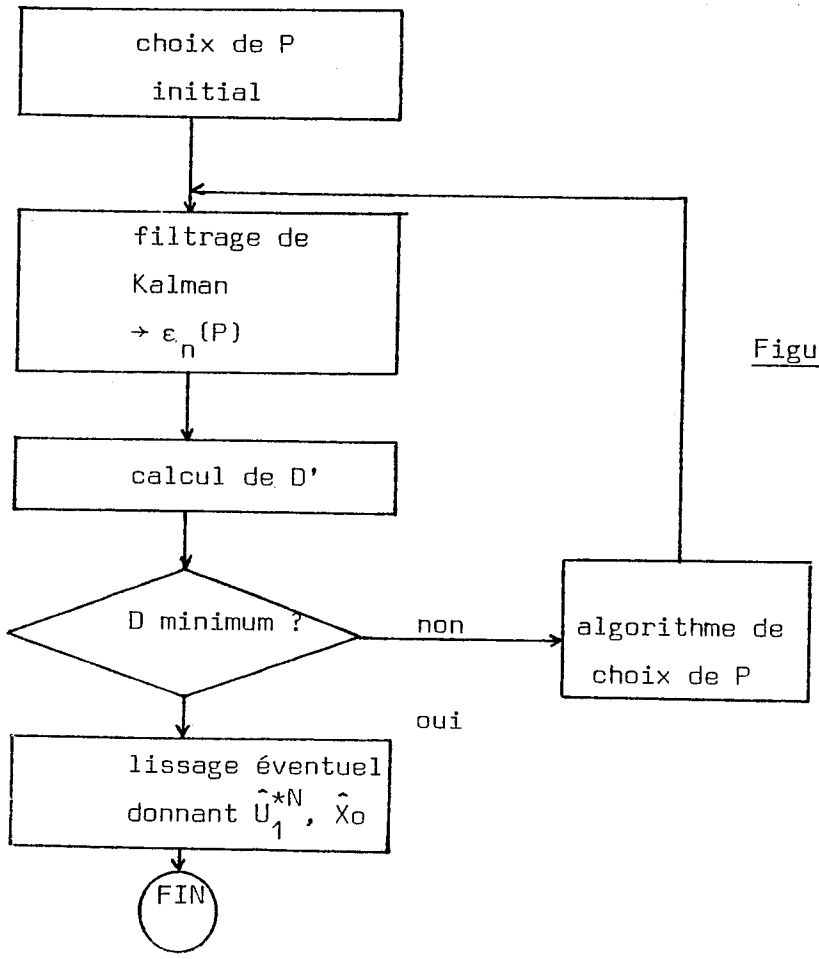


Figure 2.6

La comparaison de cet organigramme avec celui donné en II.1.1. montre que l'on a simplement remplacé la simulation du modèle par le filtrage des données à l'aide du filtre construit sur le modèle.

Le choix de l'algorithme non linéaire est guidé par le fait que D'(P) peut s'écrire sous la forme d'une somme de carrés :

$$D'(P) = \sum_{1}^{N} \varepsilon_n(P)^T \varepsilon_n(P) = \sum_{n=1}^{N} \sum_{\ell=1}^{N_s} \varepsilon_{n,\ell}^2 (P)$$

où $\varepsilon_{n,e}$ représente la ℓ ième composante de ε_n

Le Gradient s'écrit :

$$G(P) = 2 \sum_{n=1}^{N} \sum_{\ell=1}^{N_s} \varepsilon_{n,\ell}(P) \frac{\partial \varepsilon_{n,\ell}(P)}{\partial P}$$

et le Hessien modifié de Gausse Newton :

$$H(P) = 2 \sum_{n=1}^{N} \sum_{\ell=1}^{N_s} \frac{\partial \varepsilon_{n,\ell}(P)}{\partial P} \cdot \frac{\partial \varepsilon_{n,\ell}(P)}{\partial P}^T$$

Ces quantités peuvent être calculées de façon rigoureuse en adjoignant au filtrage de Kalman les équations de sensibilité du filtre. On peut aussi les calculer par différences finies.

On aura donc :

- <u>La méthode du Gradient</u> :

$$P_{k+1} = P_k - \lambda_k G(P_k)$$ avec λk tel que $D'(P_{k+1}) < D(P_k)$

- <u>La méthode de Gauss-Newton</u> :

$$P_{k+1} = P_k - \lambda_k H^{-1}(P_k) G(P_k)$$

λk étant là aussi choisi de façon algorithmique de sorte que la distance diminue.

- <u>Ou mieux, la méthode de Gauss Marquardt</u> [20]

$$P_{k+1} = P_k - \lambda_k M(P_k)^{-1} G(P_k)$$

Où la matrice $M = (m_{i,j})$ régularise la matrice $H = (h_{ij})$ par :

$$m_{i,j} = h_{i,j} + h_{i,j} \cdot \mu_k \, \delta_{i,j}$$

où $\delta_{i,j}$ est le symbole de Kronecker,

et μ_k est un coefficient tendant vers o lorsque k croît.

II - 5 - 4 - Stabilité du filtre

C'est la stabilité du filtre de Kalman, démontrée par les travaux de KALMAN [21] et de DEYST et PRICE [22] qui donne tout son sens à la méthode que nous proposons en II - 3. La propriété de stabilité peut être résumée dans le théorème suivant :

Théorème II - 5 - 4

Si le système (Fn, Gn, Hn) est uniformément complètement observable et commandable pour les covariances Q et R et pour un horizon supérieur ou égal à Noc, et si Co est positive, alors le filtre de Kalman F.K. (Fn, Gn, Hn, Q, R, Co) est uniformément asymptotiquement stable (U.A.S.).

La démonstration de ce théorème utilise la fonction de Lyapunov : $V(x,n) = x^T C_{n/n}^{-1} x$, dont on peut montrer la décroissance stricte avec n, et le lemme suivant :

Lemme II - 5 - 4

Sous les conditions du théorème II-5-4, il existe deux constantes α et β positives telles que :

$$\forall n \geqslant N_{oc} \ , \ \forall x \in \mathbb{R}^{N_o},$$

$$\frac{\alpha}{1+\alpha\beta} x^T x \leqslant x^T C_{n/n}^{-1} x \leqslant \frac{1+\alpha\beta}{\alpha} x^T x$$

On peut déduire de ces propriétés un corollaire qui sera utile au chapitre V :

Corollaire II - 5 - 4

Pour un système mono entrée, mono sortie, et en négligeant

l'influence de la condition initiale, la sortie ε_n du filtre de Kalman peut s'écrire sous forme de séquence de pondération

$$\varepsilon_n = \sum_{i=0}^{n} a_{n,i}\, s_{n-i} + b_{n,i}\, v_{n-i}$$

Sous les hypothèses du théorème II - 4, il existe deux constantes positives k_1 et k_2 telles que au dela d'un certain n' on ait

$$\forall n \geqslant n' \qquad \forall i \in [0,n] \qquad \|a_{n,i}\| \leqslant k_1\, e^{-k_2(n-i)}$$

$$\|b_{n,i}\| \leqslant k_1\, e^{-k_2(n-i)}$$

Cette propriété découle de la stabilité asymptotique uniforme du filtre, et du fait que pour assurer la commandabilité et l'observabilité complète uniforme du système il faut supposer que les variations de Gn et Hn avec n sont bornées.

II - 6 - Propriétés statistiques du critère et tests

Cette partie s'appuie, en les utilisant de façon approchée, compte tenu de l'aspect non linéaire de l'identification, sur des méthodes qui sont classiques en estimation linéaire gaussienne [23]. D'autre part, les questions de choix du modèle et de domaines d'incertitude peuvent être envisagées, on l'a vu en I, comme des problèmes de décision statistique. Il convient donc tout d'abord d'étudier les propriétés statistiques des éléments de décision dont on dispose. Ces éléments seront essentiellement la valeur du critère et la séquence des résidus .

II - 6 - 1 - Propriétés statistiques

Les propriétés statistiques ne peuvent être précisées que si l'on suppose exactes les hypothèses statistiques qui fondent le modèle. En effet, dans le cas contraire les lois de probabilité ne sont pas entièrement définies. On suppose donc que le modèle est exacte et que Po est la valeur vraie des paramètres ayant engendré les données.

Les propriétés suivantes seront données en considérant le cas général : Co définie positive non infinie. Les cas particuliers Co nulle ou infinie s'en déduisent aisément.

Propriété 1 :

$D(Po, U_1^{*N}, Xo)$ est distribué comme un χ^2 à No + N (Ne + Ns) degrés de liberté, en tant que somme d'un tel nombre de variables gaussiennes indépendantes, équidistribuées, au carré. (U_1^{*N} représente ici la séquence des entrées vraies supposées connues sans bruit).

Propriété 2 :

$D'(Po)$ est distribué comme un χ^2 à $N.N_s$ degrés de liberté. Cette distribution est exacte si le modèle est linéaire.

Propriété 3 :

$D'(\hat{P})$ est distribué de façon approchée comme un χ^2 à N.Ns - p degrés de liberté.

Remarques

. La distribution serait exacte dans le cas d'un modèle linéaire par rapport aux paramètres, ce qui est rarement le cas.

. Il faut supposer que D'(P) est deux fois continûment dérivable en P, ce qui revient à supposer qu'il en est de même pour les matrices Fn, Gn, Hn définissant le modèle.

. Les travaux de Beale [24] ont montré que cette distribution était d'autant meilleure que les bruits de mesure sont faibles et que le modèle se rapporchait d'un modèle linéaire par rapport aux paramètres. Une mesure de la non linéarité par rapport aux paramètres d'un modèle est d'ailleurs proposée en [24] ainsi que des moyens de corriger la distribution.

. Le terme -p correspond à l'optimisation de la distance par rapport à p paramètres. En toute rigueur il conviendrait de remplacer

p par p' ≤ p nombre de paramètres sensibilisés du modèle :

$$\left(\text{rang de la matrice } E\left[\frac{\partial^2 D'(P_0)}{\partial P^2}\right]\right)$$

En général, on néglige cette correction ce qui n'a pour effet que de rendre les tests plus sévères.

Propriété 4 :

$D'(Po) - D'(\hat{P})$ est distribué, sur les mêmes hypothèses de façon approchée comme un χ^2 à p (p') degrés de liberté. De plus cette quantité est indépendante de $D'(\hat{P})$.

Propriété 5 :

Soient deux modèles M_1 et M_2 d'espaces paramétriques $\{P_1\}$ et $\{P_2\}$, qui sont dits emboités, c'est à dire tels que

- $p_1 < p_2$

- M_1 se déduit de M_2 en fixant la valeur de $p_2 - p_1$ paramètres (en général on les met à 0). M_1 est donc une simplification de M_2.

Supposons M_1 juste.

Alors $D_1'(\hat{P}_1) - D_2'(\hat{P}_2)$ est distribué approximativement comme un χ^2 à $p_2 - p_1$ $(p_2' - p_1')$ degrés de libertés et cette quantité est indépendante de $D_2'(\hat{P}_2)$.

Propriété 6 :

La suite des résidus ε_n (Po) est une suite de vecteurs gaussiens, à moyenne nulle, covariance unité, indépendants entre eux.

Remarque

Cette propriété n'est pas la propriété classique d'indépendance des résidus dans le filtrage de KALMAN due à KAILATH [18]. En effet, nous ne nous situons pas dans le cas classique d'application du

filtre dans le cadre d'un problème d'estimation d'état, mais dans le cas d'application à la commande optimale.

Le cas classique peut se représenter par la figure suivante :

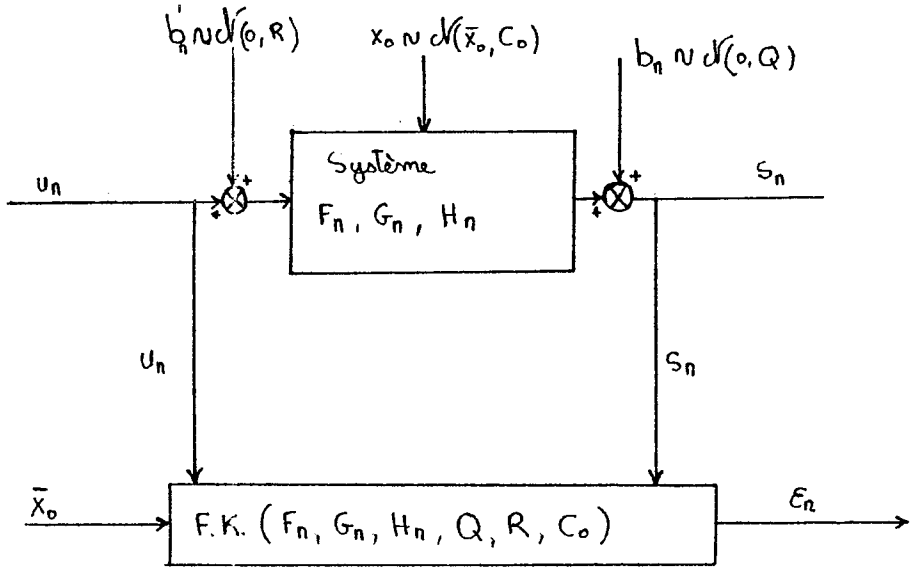


Figure 2.7

DAns cette figure le symbole $\sim \mathcal{N}$ (m, v) signifie distribué selon une loi normale de moyenne m et de variance v. Au cas classique correspond la propriété 6 bis : la suite ε_n est une suite de vecteurs gaussiens, à moyenne nulle, covariance unite, indépendants entre eux.

Le cas qui nous occupe peut se représenter par la figure 2.8.

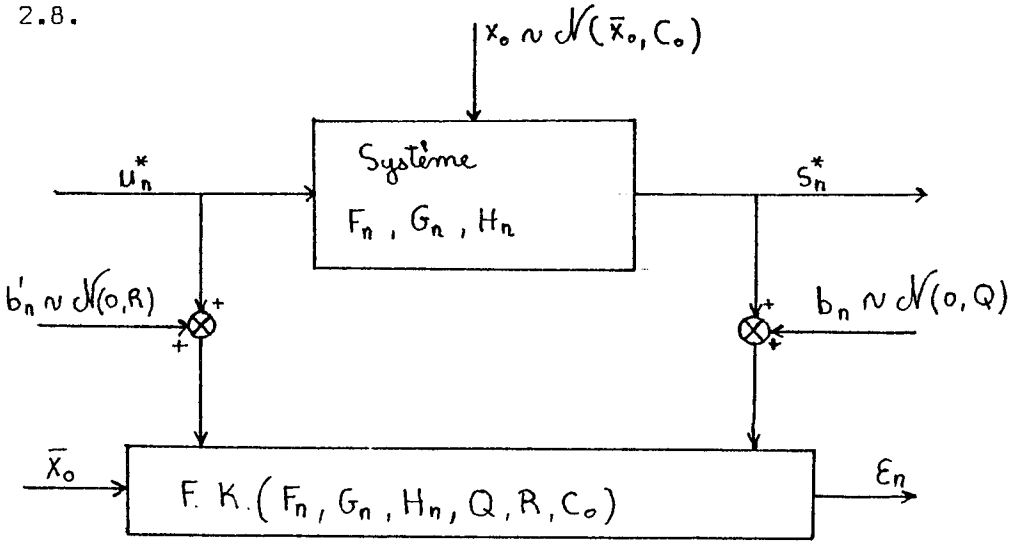


Figure 2.8

Demonstration de la propriété 6 :

Etant donné la linéarité du système (Fn, Gn, Hn) et du filtre de Kalman F.K. associé, on peut décomposer par la pensée les résidus en deux termes : (figure 2.9).

$$\varepsilon_n = v_n + \omega_n$$

. v_n provenant du filtrage de la partie déterministe des entrées (u_n^* , s_n^*) le filtre étant initialisé par Xo.

. ω_n provenant du filtrage des bruits (b'n, bn), le filtre étant initialisé par $\overline{Xo}$ - Xo.

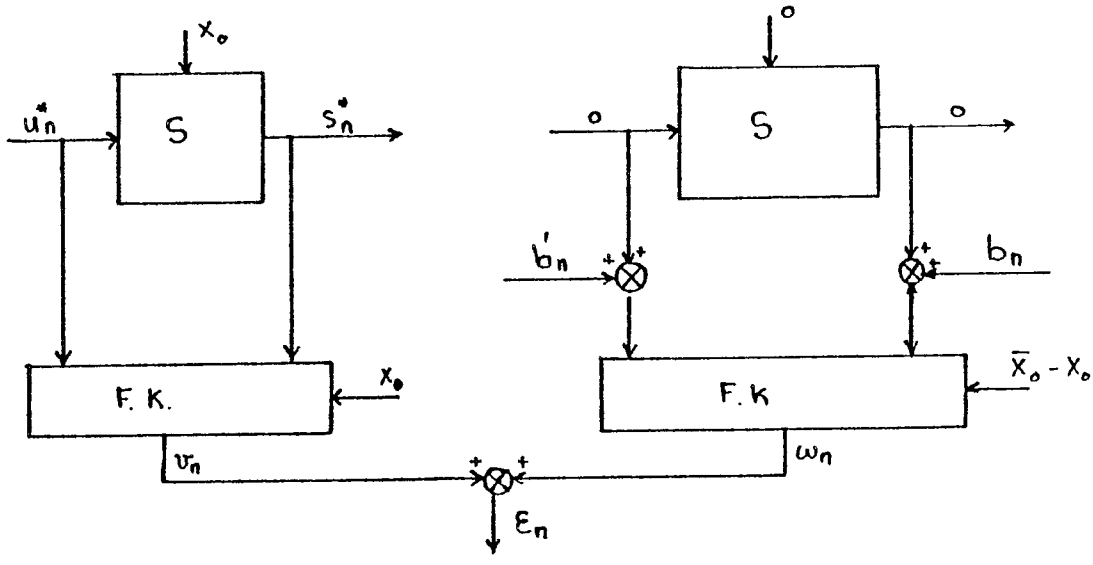


Figure 2.9

a) vn = 0 pour tout n. Cela résulte immédiatement de l'application du théorème II-5-2.: (vn) est la séquence de résidus obtenus après minimisation par filtrage de Kalman d'un critère de commande optimal trivialement nul. Donc

$$\sum_1^N v_n^T v_n = 0 \qquad \text{et} \qquad \forall n \in [1, N] \ , \ v_n = 0$$

b) Il reste à démontrer que (w_n) est une séquence de vecteurs indépendants distribués selon une loi $(0, I)$ où I est la matrice identité $N_s \times N_s$. Cette propriété découle immédiatement de la propriété classique d'indépendance des innovations de KAILATH [18]. En effet, plaçons nous dans le cas classique de la figure 2.7 avec $u_n^* = 0$ pour tout n, et en changeant le signe de b'_n :

Le système s'écrit :

$$x_{n+1} = F_n x_n - G_n b'_n$$

$$s_n = H_n x_n + b_n \quad \text{avec la condition initiale } x_0$$

Le filtre d'écrit :

$$\hat{x}_{n+1/n} = F_n \hat{x}_{n/n}$$

$$\hat{x}_{n/n} = \hat{x}_{n/n-1} + K_n [s_n - H_n \hat{x}_{n/n-1}]$$

avec la condition initiale $\hat{x}_{o/o} = \overline{x}_o$

Dans ces conditions, la séquence des résidus w_n est blanche équidistribuée. On obtient la même séquence de résidus par translation

$$(\hat{x}_{n+1/n} - x_{n+1}) = F_n(\hat{x}_{n/n} - x_n) + G_n b'_n$$

$$\hat{x}_{n/n} - x_n) = (\hat{x}_{n/n-1} - x_n) + K_n [b_n - H_n(\hat{x}_{n/n-1} - x_n)]$$

Il s'agit bien du filtrage de Kalman du couple $(+b'_n, b_n)$ avec la condition initiale $\overline{x}_o - x_o$.

Remarque

1° Dans tous ce qui précède, les hypothèses de normalité de b_n, b'_n, Xo ne sont pas nécessaires puisqu'il s'agit de propriétés d'ordre 2. Si on n'en tient pas compte, il suffira de remplacer dans ce qui précède indépendant par décorélé.

2° On obtient la propriété paradoxale suivante :

Corollaire II - 6 - 1

Soit une famille de filtres de Kalman

$FK\left[F_n(P), G_n(P), H_n(P), Q, R, C_o\right]$ dépendant d'un vecteur de paramètres P,

Soient (b_n) et (b'_n) deux séquences de vecteurs indépendants

entre eux distribués respectivement selon des lois N(o,Q), N(o,R) et soit Xo une variable indépendante des précédentes distribuée en N(o,Co).

Alors la séquence de résidus (Wn(P)) du filtrage de (b'_n, b_n) avec l'initialisation Xo est une séquence de vecteurs indépendants distribués en N(o,I), et ceci pour toute valeur de P.

II - 6 - 2 - Tests de caractérisation fondés sur la valeur du critère optimal

Ces tests sont des applications classiques de théorie du test au problème de la caractérisation. Leur utilisation en identification a fait l'objet de nombreus travaux [4], [25], [26]. Nous les appliquerons à la méthode d'identification proposée dans deux cas : le test absolu lorsque les covariances Q et R sont connues, et le test par comparaison de modèles lorsque celles-ci ne sont connues qu'à un coefficient proportionnel près.

1° Test absolu

Le test absolu consiste à étudier si le critère minimum $D'(\hat{P})$ est compatible avec la puissance des bruits (Q, et R) connues par ailleurs, ou bien s'il faut supposer des bruits supérieurs, donc soit des perturbations négligées, soit une erreur de caractérisation traduisant des écarts systématiques entre l'objet et le modèle.

Dans notre cas, cette alternative peut être posée comme un problème de choix entre deux hypothèses statistiques :

Hypothèse 1

Il existe Po tel que les données ont été engendrées par un système régi par les équations données en II-2.

Hypothèse 2

Quel que soit Po si les données sont engendrées par les équations II-2, alors il faut supposer que

$$E(b_n b_n^T) = \lambda Q \ , \ E(b'_n b'^T_n) = \lambda R \ , \quad \lambda > 1$$

En incluant λ dans la caractérisation donnée en II-2, le choix peut se résumer de la façon suivante :

Hypothèse 1 (simple) $\lambda = 1$

Hypothèse 2 (complexe) $\lambda > 1$

La fonction de vraisemblance peut s'écrire alors

$L(U_1^{*N}, x_0, P, \lambda)$ et l'hypothèse 1 sera rejetée si

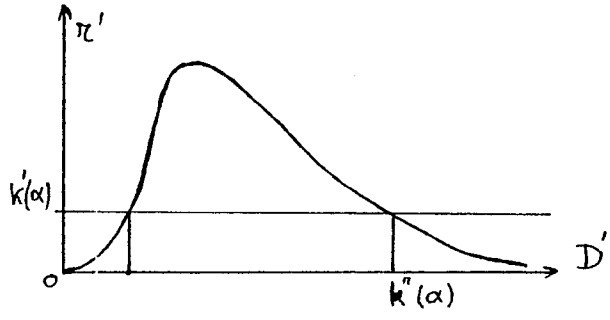
$$\tau = \frac{\max\limits_{U_1^{*N}, x_0, P} \left[L(U_1^{*N}, x_0, P, 1) \right]}{\max\limits_{U_1^{*N}, x_0, P, \lambda} \left[L(U_1^{*N}, x_0, P, \lambda) \right]} < k(\alpha)$$

où α est une probabilité fixée (en général 0.05) appelée seuil de signification du test et k (α) une fonction à déterminer connaissant la distribution de τ pour l'hypothèse 1.

Tout calcul fait, cela revient à écrire

$$\tau' = D'(\hat{P})^{N/2} e^{-D'(\hat{P})} < k'(\alpha)$$

Il y a deux seuils possibles (figure) mais seul le second



est intéressant car il correspond à l'hypothèse $\lambda > 1$. L'hypothèse 1 sera rejetée si

$$D'(\hat{P}) > k''(\alpha)$$

En général on se fixe une probabilité α (par exemple 5 %) dite seuil de signification du test, telle que l'on ait une probabilité α de rejeter H1 de façon erronée (c'est à dire si H1 est vraie). Cette définition permet de trouver la valeur de k"(α). En effet on doit avoir

$$\alpha = \mathrm{Prob}\left[D'(\hat{P}) > k''(\alpha) \,/\, H_1 \right]$$

D'après le paragraphe II-6-1 propriété 3, k"(α) est donnée dans la table ^{2} à $N.N_s - p$ degrés de liberté pour le seuil α

$$k''(\alpha) = \chi_0^2 \left(N.N_s - p , \alpha \right)$$

En revanche, on ne connait pas la probabilité de rejeter l'hypothèse 2 de façon erronée. En effet, la distribution de $D'(\hat{P})$ n'est pas définie lorsque H2 est vrai.

En conclusion, ce test n'apporte de réponse précise que si le modèle est rejeté. Dand ce cas on connait la probabilité d'erreur de cette décision. Dans le cas contraire (modèle accepté), il s'agit d'une décision provisoire qui demande d'autres vérifications. En effet, d'une part on ne connait pas la probabilité d'erreur de cette décision. D'autre part, l'hypothèse n'a été prouvée que contre un type d'hypothèse alterne et il en existe d'autes possibles.

2° <u>Tests relatifs - comparaison de modèles emboités</u>

On peut aussi opposer l'hypothèse H1 en comparant le modèle M à un modèle M2 englobant M (modèles emboités au sens de II-6-1, 5ème propriété). Le modèle M2 dépend d'un vecteur paramètre P2 à p2 dimensions et M se déduit de M2 par exemple en fixant à 0 les p2 - p paramètres supplémentaires. La comparaison conduit à opposer les deux hypothèses :

H1/ les p2 - p paramètres supplémentaires sont nuls

H2/ ces paramètres ne sont pas nuls.

La méthode de tests par rapport de vraisemblance peut alors être mise en oeuvre dans deux cas.

a) λ est fixé à 1. H1 est alors rejeté au seuil α si

$$D'(\hat{P}) - D'_2(\hat{P}_2) > \chi^2_0 (P_2 - P, \alpha)$$

Mais cette méthode qui suppose les covariances des bruits parfaitement connues n'est pas très adroite car dans ce cas il vaut mieux utiliser le test absolu.

b) λ est indéterminé. H1 est alors rejeté au seuil α si

$$\frac{N.N_S - P_2}{P_2 - P} \cdot \frac{D'(\hat{P}) - D'_2(\hat{P}_2)}{D'_2(\hat{P}_2)} > F\left(P_2 - P, N.N_S - P_2, \alpha\right)$$

où F est le seuil correspondant à la distribution de Fisher-Snedecor.

II - 6 - 3 - *Test fondé sur l'indépendance des résidus. - Condition d'application*

Ce test a été proposé dans divers cas de maximum de vrai - semblance ([25], [26]). Par exemple dans la méthode exposée en III-2 c'est le seul à pouvoir être mis en oeuvre.

Dans notre cas, il se fonde sur la propriété 6 de II-6-1. *Supposant pour simplifier un résidu monodimensionnel, il consiste à* calculer quelques coefficients de correlation.

$$r_k = \frac{\frac{1}{N-k} \sum \varepsilon_n \varepsilon_{n+k} - \left(\frac{1}{N-k} \sum \varepsilon_n\right)\left(\frac{1}{N-k} \sum \varepsilon_{n+k}\right)}{\sqrt{\left(\frac{1}{N-k} \sum \varepsilon_n^2 - \left(\frac{1}{N-k} \sum \varepsilon_n\right)^2\right)\left(\frac{1}{N-k} \sum \varepsilon_{n+k}^2 - \left(\frac{1}{N-k} \sum \varepsilon_{n+k}\right)^2\right)}}$$

et de les comparer au seuil correspondant de la distribution du coefficient de correlation dans l'hypothèse nulle.

Néanmoins cette méthode de test ne doit être utilisée qu'avec

certaines précautions, qui ont été jusqu'à présent insuffisamment explicitées. En effet, l'analyse qui précède serait exacte si l'on disposait des résidus au point Po. Or, les résidus dont on dispose ne sont connus qu'au point $\hat{P}$, qui minimise la distance D'(P). On peut montrer que même dans le cas le plus favorable, lorsque les résidus $\varepsilon_n(P)$ sont des fonctions affines de P (ce qui n'a pas lieu en général et en particulier dans notre cas) les résidus en $\hat{P}$ ne sont pas rigoureusement indépendants.

La fonction à minimiser $D'(P) = \sum_1 \varepsilon_n^2(P)$ peut s'écrire en notation vectorielle $D'(P) = \varepsilon(P)^T \varepsilon(P)$

En introduisant $\Delta P = P - Po$ et les dérivées partielles $\frac{\partial \varepsilon}{\partial P}$ (ou fonction de sensibilité) qui sont constantes dans le cas affine, il vient $D'(P) = \left(\varepsilon(P_0) + \frac{\partial \varepsilon}{\partial P}\Delta P\right)^T \left(\varepsilon(P_0) + \frac{\partial \varepsilon}{\partial P}\Delta P\right)$

La minimisation en ΔP conduit à la formule classique des moindres carrés

$$\hat{\Delta P} = \left[\frac{\partial \varepsilon}{\partial P}^T \frac{\partial \varepsilon}{\partial P}\right]^{-1} \frac{\partial \varepsilon}{\partial P}^T \varepsilon(P_0)$$

$$\varepsilon(\hat{P}) = \left\{ I - \frac{\partial \varepsilon}{\partial P}\left[\frac{\partial \varepsilon}{\partial P}^T \frac{\partial \varepsilon}{\partial P}\right]^{-1} \frac{\partial \varepsilon}{\partial P}^T \right\} \varepsilon(P_0)$$

Dans l'hypothèse ou le modèle est exact, on aura :

$$E\left[\varepsilon(P_0)\,\varepsilon(P_0)^T\right] = I$$

et $$E\left[\varepsilon(\hat{P})\,\varepsilon(\hat{P})^T\right] = I - \frac{\partial \varepsilon}{\partial P}\left[\frac{\partial \varepsilon}{\partial P}^T \frac{\partial \varepsilon}{\partial P}\right]^{-1} \frac{\partial \varepsilon}{\partial P}^T$$

La matrice $M = \frac{\partial \varepsilon}{\partial P}\left[\frac{\partial \varepsilon}{\partial P}^T \frac{\partial \varepsilon}{\partial P}^T\right]^{-1} \frac{\partial \varepsilon}{\partial P}^T$

a p valeurs propres égales à 1 correspondant au sous espace propre engendré par les fonctions de sensibilité (colonnes de $\frac{\partial \varepsilon}{\partial P}$), les N - p autres valeurs propres étant nulles. On peut donc écrire :

$M = F F^T$ ou F est la matrice N x p dont les p colonnes forment une base orthonormée du sous espace des fonctions de sensibilité. Si f_{ij} sont les éléments de F on pourra écrire

$$E\left(\frac{1}{N-k} \sum_1 \varepsilon_n(\hat{P})\, \varepsilon_{n+k}(\hat{P})\right) = \frac{1}{N-k} \sum_{i=1}^{p} \sum_{n=1}^{N-k} f_{i,n}\, f_{i,n+k}$$

Les fonctions de sensibilité étant en général à basse fréquence, pour les faibles valeurs de k (1,2...) sur lesquelles portent en général le test, leur coefficient de correlation $\sum_{n=1}^{N-k} f_{i,n} f_{i,n-k}$ sera voisin de 1.

L'espérance de la correlation $C_k = \frac{1}{N-k} \sum_{1}^{N-k} \varepsilon_n(\hat{P}) \varepsilon_{n+k}$ qui devrait être nulle en cas d'indépendance vaut environ $-\frac{p}{N-k}$

D'autre part, la variance de Ck pour un bruit gaussien blanc vaut

$$\sigma^2_{C_k} = \frac{1}{N-k}$$

On a donc pour la correlation Ck un rapport biais sur écart-type de l'ordre de $\frac{p}{\sqrt{N-k}}$. Un rapport du même ordre se retrouve sur le coefficient de correlation rk. Si l'on admet que pour que le test ne soit pas faussé, il convient que ce rapport n'exède par $\frac{1}{10}$ la condition d'applicabilité du test est que l'on dispose d'un horizon de mesures N suffisamment grand, tel que $N \geq 100\ p$. où p est le nombre de paramètres à estimer.

II - 6 - 4 - <u>Domaine d'incertitude des paramètres</u>

Il s'agit là encore d'application de méthodes classiques de test statistique à la détermination de l'incertitude sur les paramètres. Ces méthodes sont d'usage courant en identification depuis le travail d'Aström en [27].

Dans le cas de la méthode proposée le problème peut se formuler de la façon suivante :

Considérant un point P de l'espace paramétrique on se pose la question de savoir si ce point peut être le point réel Po, ou non, sous la forme de deux hypothèses :

H1 / Po = P

H2 / Po $\neq$ P

Deux cas peuvent se présenter.

a) Les covariances des bruits sont supposées bien connues. La méthode du rapport de vraisemblance est appliquée en fixant $\lambda = 1$ et conduit au rejet de H1 au seuil α si :

$$D'(P) - D'(\hat{P}) > k(\alpha)$$

La propriété 4 du II - 6 - 1 permet de fixer

$$k(\alpha) = \chi^2_0(p, \alpha)$$

b) λ est supposé mal connu. Il est alors laissé indéterminé et la même méthode conduit à la condition de rejet

$$\frac{N.N_s - p}{p} \; \frac{D'(P) - D'(\hat{P})}{D'(\hat{P})} > F(p, N.N_s - p, \alpha)$$

Dans tous les cas, la condition s'écrit D'(P) > constante. En général on considère plutôt les points qui ne sont pas rejetés, qui obéissent à la condition D'(P) ≤ constante et qui forment le domaine d'incertitude de Po. Il faut noter que la forme de ce domaine (dit aussi domaine isodistance) est rigoureuse [24]. Seule la valeur de la constante dépend des approximations qui ont été faites dans l'étude des distributions (II-6-1).

Enfin, il est souvent commede d'utiliser une approximation de ces domaines sous forme de quadrique. Celle-ci s'obtient en considérant le hessien H (défini en II-5-2 et en approximant $D'(P) - D'(\hat{P})$ par $\Delta P^T H \Delta P$
où $\Delta P = P - \hat{P}$

La quadrique d'incertitude est alors centrée en $\hat{P}$ et définie par $\Delta P^T H \Delta P \leq$ constante.

On voit là l'intérêt d'utiliser pour l'optimisation non linéaire une méthode du 2ème ordre du type Gauss Newton ou Gauss Marquardt qui fournit à la fois $\hat{P}$ et H.

III - PARALLELE AVEC LA METHODE DE MEHRA

III - 1 - Modèles déterministes et stochastiques

Nous avons vu au chapitre 1 que pour pouvoir "identifier" un objet à un modèle, il convient d'introduire dans le modèle un aspect aléatoire. La première tendance en ce sens à consisté à considérer un modèle déterministe dont les sorties sont mesurées avec des erreurs ou bruits de mesure. La méthode que nous proposons au chapitre 2 se situe dans ce cadre puisque nous montrons l'intérêt qu'il y a a introduire des bruits de mesure sur les entrées notamment pour les systèmes instables.

L'autre tendance qui est de plus en plus utilisée en automatique depuis les travaux d'Aström et Bohlin [4] et Clarke [12] consiste à considérer des modèles stochastiques. La façon la plus courante d'introduire ces modèles consiste à prendre en compte à coté des entrées connues du système des entrées aléatoires dites bruit de processus, inconnues, dont on suppose qu'elles ne sont pas corrélées avec les entrées connues. Ces bruits de processus peuvent être attribuées :

- Aux perturbations réellement aléatoires et inconnues. L'exemple type est celui des rafales de vent (Mehra [5]) dand l'identification des paramètres de vol d'un avion.

- Aux phénomènes secondaires qui pourraient être modèlisés éventuellement mais qui ne sont pas pris en compte dans le cadre d'un modèle de représentation, par exemple en vue de la commande du rpocessus.

- Enfin toujours dans le cas d'un modèle de représentation, aux erreurs de caractérisation introduites volontairement pour des raisons de simplification du modèle : par exemple modèle linéarisé d'un système non linéaire autour d'un point de fonctionnement.

Il nous a semblé intéressant de présenter en III-2 la méthode de maximum de vraisemblance adaptée au problème des modèles stochastiques. En effet, cette méthode présente de grandes analogies avec la méthode proposée au chapitre 2, notamment par l'usage du filtre

de Kalman pour calculer la distance objet-modèle. Nous reviendrons en III-3 sur le problème du choix : modèle déterministe-modèle stochastique en montrant sur un exemple l'intérêt d'un modèle déterministe.

III - 2 - Méthode du maximum de vraisemblance de Mehra

Cette méthode présentée par Mehra en [5] et [28] est une extension des travaux de Aström et Wenmark [29] et Kashyap [30].

III - 2 - 1 - Cas général

Le modèle est un processus stochastique décrit en équation d'état :

$$x_{n+1} = f_n(x_n, u_n, b'_n, P)$$

$$s_n = h_n(x_n, b_n, P)$$

où
- Xn est un vecteur d'état, de dimension No
- Un est un vecteur d'entrées connues parfaitement (dimension Ne)
- Sn est le vecteur de sorties (dimension Ns)
- b_n et b'_n sont des vecteurs aléatoires gaussiens (Ns, Ne) blancs indépendants entre eux.

$$E(b_n) = 0 \qquad E(b'_n) = 0$$

$$E(b'_n b'^T_m) = R(P)\,\delta_{n,m} \quad , \quad E(b_n b^T_m) = Q(P)\,\delta_{n,m}$$

$$E(b_n b'^T_m) = 0$$

- P est le vecteur des paramètres inconnus.

La mise en oeuvre d'une méthode de maximum de vraisemblance en vue d'estimer P nécessite le calcul de la densité de probabilité des données S_1^N (et non pas aussi U_1^N qui est déterministe) pour une valeur P des paramètres : $L(S_1^N ; P)$

C'est là un problème difficile car la transformation qui fait passer de b_n, b'_n dont les densités sont connues à S_1^N est non linéaire. En particulier, les termes s_n ne sont pas gaussiens.

III - 2 - 2 - Cas linéaire

Le cas linéaire est plus facile à traiter. Formulons le, pour se rapprocher au mieux des équations II-5-1, de la façon suivante :

$$x_{n+1} = F_n(P)\, x_n + G_n(P)\left(u_n + b'_n\right)$$

$$s_n = H_n(P)\, x_n + b_n$$

On peut décomposer la fonction de vraisemblance par application répétée de la règle de Bayes :

$$L\left(S_1^N ; P\right) = p\left(s_1 / P\right) . p\left(s_2 / s_1, P\right) \ldots p\left(s_n / S_1^{n-1}, P\right) \ldots p\left(s_N / S_1^{N-1}, P\right)$$

La théorie du filtrage de Kalman permet de montrer que le vecteur aléatoire s_n / S_1^{n-1} est distribué de façon gaussienne avec pour moyenne et covariance

$$E\left(s_n / S_1^{n-1}\right) = H_n\, \hat{x}_{n/n-1}$$

$$Cov\left[s_n / S_1^{n-1}\right] = H_n\, C_{n/n-1}\, H_n^T + Q$$

où $\hat{x}_{n/n-1}$ et $C_{n/n-1}$ sont données par le filtrage de Kalman dont les équations ont été données en II-5-2.

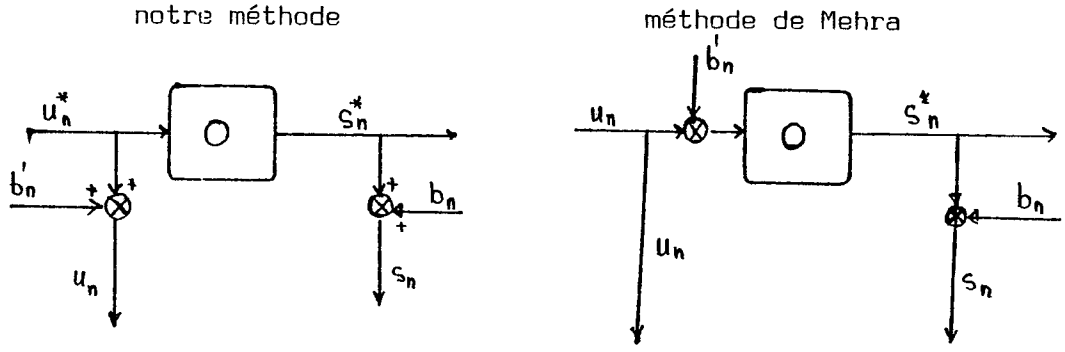
L'opposé du logarithme de L fournit donc le critère à minimiser qui s'écrit :

$$D(P) = \sum_{n=1}^{N} \Big\{ \left(s_n - H_n\, \hat{x}_{n/n-1}\right)^T \left[H_n\, C_{n/n-1}\, H_n^T + Q\right]^{-1} \left(s_n - H_n\, \hat{x}_{n/n-1}\right)$$

$$+ \operatorname{Log}\left[\det\left(H_n C_{n/n-1} H_n^T + Q\right)\right] \Big\}$$

III - 2 - 3 - Comparaison avec notre méthode

On voit ainsi apparaître un critère D(P) qui au logarithme près, est analogue au critère D'(P) donné en II-5-1. Cette différence tient à ce que l'objet des deux méthodes est différent : la figure suivante montre le xchéma bloc des modèles correspondants :



Un exemple simple permet de mieux saisir la différence.

Considérons une entrée nulle.

Dans le premier système les variables mesurées u et s seront deux bruits blancs, tandis que dans le second u sera nul et s un bruit corrélé. Si l'on applique la deuxième méthode dans le premier cas, on concluera que S est un système sans dynamique, alors qu'il est évident et notre méthode le montrerait, que le système est entièrement indéterminé.

Des différences de mise en oeuvre peuvent être aussi remarquées.

- Le critère n'est plus quadratique et son optimisation nécessite des méthodes de programmation non linéaire plus générales.

- Le problème du test de caractérisation est plus délicat.

On ne voit guère que la méthode d'indépendance des résidus qui puisse lui être appliquée.

- De même, le problème du domaine d'incertitude des paramètres est délicat. Une façon approximative de l'aborder consisterait à utiliser sous forme d'égalité l'inégalité de RAO CRAMER :

$$E\left[\Delta P \cdot \Delta P^T\right] \simeq E\left\{ \frac{\partial D(P_0)}{\partial P} \cdot \frac{\partial D^T(P_0)}{\partial P} \right\} = - E\left[\frac{\partial^2 D(P_0)}{\partial P^2} \right]$$

mais cette quantité s'avère difficile à calculer dans le cas général.

II - 2 - 4 - Cas du filtre stationnaire

Un cas particulier de mise en oeuvre plus aisée et qui se présente fréquemment est le cas où le filtre de Kalman permettant de calculer la fonction de vraisemblance est stationnaire.

Il faut pour cela supposer :

- que le modèle est linéaire stationnaire,

- que l'état initial Xo est distribué de façon gaussienne avec une moyenne connue Xo et une covariance égale à C_∞ où C_∞ est la covariance limite de l'état estimé par le filtre de Kalman : C_∞ est la solution positive de l'équation :

$$C_\infty = F\left\{ C_\infty - C_\infty H^T \left[Q + H C_\infty H^T \right]^{-1} H C_\infty \right\} F^T + G R G^T$$

Il est facile de voir que si le filtre de Kalman est initialisé à partir de $\overline{Xo}$, C_∞ , il est alors stationnaire.

Le critère s'écrit :

$$D(P) = \sum_{n=1}^{N} \left\{ \left[S_n - H \hat{X}_{n/n-1} \right]^T \left[Q + H C_\infty H^T \right]^{-1} \left[S_n - H \hat{X}_{n/n-1} \right] + \log\left[\det\left(Q + H C_\infty H^T \right) \right] \right\}$$

On suppose que l'on peut procéder au changement de paramètre

$$P \text{ (dimension p)} \rightarrow \begin{cases} P'(\text{dimension } p' = p - 1) \\ \lambda(\text{dimension } 1) \end{cases}$$

tel que :

$$\det\left[Q + H C_\infty H^T\right] = \lambda^{N_s} \quad \text{est indépendant de } P'$$

$$\lambda \sum_1^N \left(s_n - H \hat{x}_{n/n-1}\right)^T \left[Q + H C_\infty H^T\right]^{-1} \left(s_n - H \hat{x}_{n/n-1}\right) = D'(P') \quad \text{est indépendant de } \lambda.$$

Le critère s'écrit alors :

$$D(P', \lambda) = \frac{D'(P')}{\lambda} + N \operatorname{Log}(\lambda^{N_s})$$

- La minimisation se fait alors en deux temps :

On cherche $\hat{P}'$ qui minimise D'. Cette minimisation est non linéaire mais porte sur un critère quadratique.

Puis on obtient $\hat{\lambda} = \frac{1}{N.N_s - p'} D'(\hat{P}')$

- le test de caractérisation peut être mené à bien par comparaison de modèles emboités de paramètres P'1 (p'_1) et P'2 (p'_2) avec $p'_1 < p'_2$. Le modèle 1 sera rejeté au seuil α si

$$\frac{N.N_s - P'_2}{P'_2 - P'_1} \; \frac{D'_1(\hat{P}'_1) - D'_2(\hat{P}'_2)}{D'_2(\hat{P}'_2)} > F_0\left(P'_2 - P'_1, N N_s - P'_2, \alpha\right)$$

- Un domaine d'incertitude peut être calculé. Un point P' est hors de ce domaine au seuil α si

$$\frac{N.N_s - P'}{P'} \; \frac{D'(P') - D'(\hat{P}')}{D'(\hat{P}')} > F_0\left(P', N.N_s - P', \alpha\right)$$

III - 2 - 5 - <u>Difficultés du cas stationnaire</u>

Pour pouvoir en toute rigueur appliquer le filtre stationnaire il convient de connaitre l'état initial avec l'incertitude C∞ (P). Une façon de bien connaitre l'état initial consiste à laisser le système

atteindre son régime permanent avec une entrée u nulle, le système étant soumis à la seule perturbation b'. A la fin de cette période, l'observation du système peut commencer avec une entrée u variable. L'état initial ainsi obtenu est à moyenne nulle avec une covariance Cx(P) qui est uniquement fonction des paramètres du système.

Cx(P) vérifie l'équation

$$C_x = F C_x F^T + G R G^T$$

Malheureusement on peut démontrer que Cx est "supérieur" à C∞ au sens que

$$\forall x \qquad x^T C_x x \geqslant x^T C_\infty x$$

L'application de la méthode du filtre stationnaire dans ce cas n'est qu'une approximation.

Une façon de tourner le problème consiste à écrire le modèle sous la forme du système S'.

$$x'_{n+1} = F x'_n + G u_n + F K_\infty b_n$$

$$s'_n = H x'_n + b_n$$

$$\text{où} \quad E(b_n) = 0 \qquad E(b_n b_m^T) = \left[Q + H C_\infty H^T \right] \delta_{n,m}$$

et K∞ est le gain de Kalman stationnaire associé à S.

On peut démontrer que en régime stationnaire S et S' fournissent des suites S_1^N et S'^N_1 ayant les mêmes propriétés statistiques.

Le filtre de Kalman associé à S' a un gain K'∞ égal à K∞ et une covariance stationnaire C'∞ = 0. Pour que le filtre reste stationnaire il convient de connaitre l'état initial X'o de façon certaine. Par exemple, l'observation peut débuter au moment où ni l'entrée, ni la perturbation n'ont commencé à agir sur le système.

Mais alors S' n'est pas en régime stationnaire, régime dont il est facile de vérifier que la covariance Cx' obéit à la relation

$$C_{x'} + C_{\infty} = C_x$$

Donc, S' n'est pas équivalent à S. La méthode n'est rigoureuse que lorsque la caractérisation s'écrit d'emblée sous la forme S'.

La méthode du filtre stationnaire proposée par Mehra en [28] est donc bien approchée.

III - 2 - 6 - Méthode d'Aström

Un modèle particulièrement simple du type S' est le modèle d'Aström qui s'écrit sous forme de fonction de transfert rationnelle en z : dans le cas mono entrée mono sortie on a :

$$s_n = -\sum_{i=1}^{N_0} A_i(P') s_{n-i} + \sum_{i=0}^{N_0} B_i(P') u_{n-i} + \lambda \left[b_n + \sum_{i=1}^{N_0} C_i(P') b_{n-i} \right]$$

Si l'on suppose vérifiées les conditions initiales

$$b_{-i} = 0 \quad , \quad i \in [0, N_0 - 1]$$

le critère et maximum de vraisemblance s'écrira

$$D(P) = \frac{D'(P')}{\lambda} + N \operatorname{Log} \lambda$$

avec $D'(P') = \sum_1^N \varepsilon_n^2(P')$

les résidus étant donnés par l'équation de filtrage :

$$\varepsilon_n = -\sum_{i=1}^{N_0} C_i \varepsilon_{n-i} + s_n + \sum_{i=1}^{N_0} A_i s_{n-i} - \sum_{i=0}^{N_0} B_i u_{n-i}$$

III - 3 - Un exemple de modèle déterministe

Etant donnés la perfection atteinte par les méthodes d'identification de processus stochastiques présentées en III-2, et leur intérêt en automatique, la question peut se poser de l'intérêt de méthodes fondées sur des modèles déterministes telle que celle que nous proposons.

Il nous semble cependant, que ces méthodes devraient trouver un lieu d'application dans des problèmes d'économie ou d'écologie, notamment ceux qui traitent de la dynamique de populations. En effet, les variables d'état de ces systèmes sont des grands nombres ou des distributions de probabilité évoluant en suivant des équations de type diffusion ou transfert (équations de Kolmogorov par exemple). Ces équations sont donc de type déterministe puisqu'il s'agit d'équations régissant l'évolution de la loi de probabilité d'un phénomène stochastique. En revanche, les mesures que l'on peut faire sur ces systèmes sont plus souvent du type sondage que du type recensement général et donc entâchées de bruits de mesure importants et que l'on peut chiffrer. De plus, ces systèmes sont souvent instables.

Nous allons présenter sommairement un exemple de tels systèmes issu d'une étude [31] à laquelle nous avons participé.

Il s'agit de modéliser la dynamique de la population des bovins en France : on appelle $n_i(t)$ le nombre de bovins dans la tranche d'âge i en mois, par exemple. En prenant un pas de temps de un mois sur t, on aura le système d'équations :

$$n_i(t+1) = n_{i-1}(t) - m_i(t) - b_i(t) + e_i(t)$$

$$n_1(t) = \sum_{i=1}^{\infty} \lambda_i(t)\, n_i(t)$$

Dans ce système :

$b_i(t)$ représente l'abattage de bovins d'âge i à la date t,

$e_i(t)$ représente le bilan importation-exportation de bovins,

$m_i(t)$ représente la mortalité naturelle

$\lambda_i(t)$ représente la fécondité d'un bovin d'âge i à la date t (variations saisonnières. Les $\lambda_i(t)$ sont des paramètres du système. En revanche, b_i et e_i sont des entrées, m_i peut être considéré comme une entrée ou modélisée à l'aide de coefficients de mortalité μ_i:

$$m_i(t) = \mu_i\, n_i(t).$$

En réalité, le modèle est plus complexe puisqu'il faut distinguer les mâles et les femelles, les génisses non fécondables et les vaches.

L'examen de ce modèle montre que les équations régissant la population sont simples et réalistes. Il ne semble pas qu'il y ait lieu de considérer de bruit de processus (celà serait le cas si l'on étudiait un petit troupeau, mais à l'échelle d'un pays il s'agit de grands nombres dans lesquels les différences locales s'attenuent). D'autre part, ni l'état initial, pyramide d'âge des bovins à un instant donné, ni les entrées ne peuvent être connus avec une grande précision. Il s'agit de grandeurs déterministes mal mesurées. Enfin, le système est instable : une vache fournit normalement un veau tous les ans. La stabilisation du système est l'effet du processus de bouclage engendrant les entrées. Ce processus est complexe : il tient compte de facteurs économiques, écologiques (limitation des ressources) politiques etc... mais il n'entre pas dans le modèle considéré.

Ce genre de problème que l'automatique commence à envisager nous semble être un lieu d'application privilégié pour la méthode que nous proposons.

IV - ETUDE DE LA CONVERGENCE DE L'ESTIMATEUR

IV - 1. Introduction

Parmi les propriétés classiques d'un extimateur il convient de distinguer celles qui concernent un horizon d'observation N fini de celles qui apparaissent lorsque l'horizon tend vers l'infini et qui sont dites asymptotiques.

Il est clair que l'estimation proposée au chapitre 2 ne jouit d'aucune propriété particulière (non-biais, covariance minimum, distribution) sur un horizon fini. Seules paraissent donc accessibles les propriétés asymptotiques. Parmi celles-ci nous nous sommes intéressés à la question de la convergence. En supposant le modèle exact il s'agit de savoir si l'on peut approcher aussi près que l'on veut du point réel P_0 en augmentant suffisamment l'horizon d'observation, et ceci pour presque toutes les expériences, c'est-à-dire avec une probabilité unité. Cette question n'est pas totalement triviale. En effet, en général, le nombre d'inconnues est fixé et il est logique de penser que, en augmentant le nombre de données apportant de l'information sur ces inconnues, celles-ci soient de mieux en mieux connues. Dans notre cas, en revanche, le nombre d'inconnues auxiliaires à estimer U_1^{*N} croit avec l'horizon, ce qui rend le problème plus difficile. Avant de l'étudier nous exposerons la démonstration de la convergence de la régression non linéaire qui sert de point de départ à notre étude. Cette démonstration a été donnée par JENNRICH [32]. Le même type de démonstration est aussi utilisé pour démontrer la convergence d'autres méthodes de maximum de vraisemblance : par exemple SODERSTRÖM [33]. En effet, en identification les systèmes, même linéaires, donnent des critères non linéaires par rapport aux paramètres. On conçoit donc bien qu'ils soient justiciables des mêmes méthodes d'analyse qu'en régression non linéaire.

IV-2. Convergence de la régression non linéaire [32]

On considère le modèle suivant (où s_n est mono-dimensionnel pour simplifier l'exposé).

$$s_n = f_n(P) + b_n$$

modèle qui est supposé exact, dont Po est le point vrai inconnu avec les hypothèses :

a) $E(b_n) = 0$, $E(b_n^2) = \sigma^2$, les b_n sont mutuellement indépendants

b) Po appartient à un domaine compact $\mathcal{C}$ de l'ensemble {P}

c) les fonctions f_n sont continues, donc uniformément continues sur $\mathcal{C}$

L'estimation de P est obtenue par la méthode des moindres carrés.
On considère la distance :

$$D_N(P) = \frac{1}{N\sigma^2} \sum_{n=1}^{N} \left(s_n - f_n(P) \right)^2$$

et l'estimée $\hat{P}_N$ minimise $D_N(P)$ dans

Ajoutons l'hypothèse :

d) La fonction $F_N(P_1,P_2) = \frac{1}{N\sigma^2} \sum_{n=1}^{N} f_n(P_1)\, f_n(P_2)$ converge uniformément vers une limite $F(P_1,P_2)$, qui est donc uniformément continue sur

Il s'en suit que la distance sans bruit

$D_N^S(P) = \frac{1}{N\sigma^2} \sum_{n=1}^{N} \left(f_n(P) - f_n(P_o) \right)^2$ converge uniformément vers une limite $D^S(P)$

Enfin, la dernière hypothèse est :

e) $D^S(P)$ a sur $\mathcal{C}$ un minimum unique en Po : $D^S(Po) = 0$.

On a alors le théorème :

Théorème (Jennrich) 4-2-1

Sous les hypothèses a, b, c, d, e la suite $\hat{P}_N$ converge presque sûrement vers Po.

La démonstration fait appel aux lemmes suivants :

Lemme 4-2-2 $\frac{1}{N\sigma^2} \sum_{n=1}^{N} b_n^2$ tend presque sûrement vers 1.

Cela est dû à l'indépendance des b_n et à l'existence de $E(b_n^2)$.

Lemme 4-2-3

$\Delta F B_N(P) = \frac{1}{N\sigma^2} \sum_{n=1}^{N} \left(f_n(P_0) - f_n(P) \right) b_n$ tend presque sûrement vers 0.

En effet $\frac{1}{N} \sum \left(f_n(P_0) - f_n(P) \right)^2$ converge vers $\sigma^2 D^S(P)$ qui est bornée comme fonction continue sur un compact, et cette convergence entraîne celle de $\sum \frac{\left(f_n(P_0) - f_n(P) \right)^2}{n^2}$

Lemme 4-2-4

La convergence du lemme 2 est uniforme.

En effet

$$|\Delta F B_N(P)| \leq |\Delta F B_N(P')| + \sqrt{\frac{1}{N\sigma^2} \sum \left(f_n(P') - f_n(P) \right)^2} \cdot \sqrt{\frac{1}{N\sigma^2} \sum b_n^2}$$

Il existe No tel qu'au delà $\sqrt{\frac{1}{N\sigma^2} \sum b_n^2} \leq K$

Pour tout N et pour tout ε il existe η tel que presque sûrement

$$\| P' - P \| < \eta \implies \sqrt{\frac{1}{N\sigma^2} \sum \left(f_n(P') - f_n(P) \right)^2} \leq \frac{\varepsilon}{2K}$$

On peut trouver un nombre fini de boucles de diamètre η centrées en P'_i recouvrant $\mathcal{C}$.

En chaque P'_i il existe N_i tel qu'au delà, presque sûrement

$$|\Delta F B_N(P'_i)| \leq \frac{\varepsilon}{2}$$

Le maximum des N_0, N_i est un nombre tel qu'au delà et pour tout $P \in \mathcal{C}$ $|\Delta F B_N(P)| \leq \varepsilon$. L'ensemble des évènements pour lesquels cela ne se produit pas a une probabilité nulle, comme réunion finie d'ensembles de probabilité nulle.

La démonstration du théorème est alors la suivante :

Pour presque tout ensemble $\{b_n\}$:

$D_N(P)$ converge uniformément vers $D^S(P) + 1$

De la suite $\hat{P}_N$ on peut extraire une sous suite $\hat{P}_{NR}$ convergente. Soit P' sa limite

$D_{NR}(\hat{P}_{NR})$ converge vers $D^S(P') + 1$

Or $D_{NR}(\hat{P}_{NR}) \leq D_{NR}(P_0)$

Donc $D^S(P') + 1 \leq 1$ ce qui entraîne $P' = P_0$.

La suite $\hat{P}_N$ a un seul point d'accumulation P_0. C'est donc sa limite.

IV - 3. Convergence de l'estimation proposée

Nous limiterons cette étude au cas des systèmes linéaires, donc à la méthode d'estimation présentée en § 2-5-1. De plus, pour simplifier l'exposé, la sortie s_n sera considérée comme mono-dimensionnelle. Le modèle est supposé exact avec pour point vrai Po.

La méthode d'estimation conduit à la minimisation de la distance $D_N(P) = \frac{1}{N}\sum_1^N \varepsilon_n^2(P)$.

Nous avons vu en § 2-6-1 6° propriété, que les résidus $_n(P)$ peuvent se décomposer sous la forme

$$_n(P) = V_n(P) + W_n(P)$$

où $V_n(P)$ résulte du filtrage des couples (u_n^*, s_n^*) est déterministe, avec

$V_n(P_o) = 0$

et $W_n(P)$ résulte du filtrage des bruits (b_n, b'_n).

On peut donc introduire la distance déterministe

avec $D_N^s(P) = \frac{1}{N}\sum_1^N V_n^2(P)$ $\qquad D_N^s(P_o) = 0$

$$D_N(P) = D_N^s(P) + \frac{2}{N}\sum_1^N V_n(P)\,W_n(P) + \frac{1}{N}\sum_1^N W_n^2(P)$$

L'application d'une démonstration du type de celle exposée en § 4-2 exige la convergence uniforme presque sûre de chacun des 3 termes de cette égalité.

Examinons tout d'abord le cas de $\frac{1}{N}\sum_1^N W_n^2(P)$. En introduisant l'hypothèse :

a) Les bruits b_n, b'_n sont gaussiens, indépendants, à moyenne nulle et de covariances R et Q connues exactement.

On peut démontrer le lemme suivant :

Lemme 4-3-1

Pour tout P $\frac{1}{N}\sum_1^N W_n^2(P)$ converge sûrement vers 1.

Démonstration

On a vu en 2-4-1 6° propriété que quel que soit P les $W_n(P)$ ont une moyenne nulle, une covariance unité, et sont décorréllés entre eux. Si les bruits b_n, b'_n sont gaussiens, les $W_n(P)$ seront aussi gaussiens comme combinaison linéaire de termes gaussiens. Ils seront donc indépendants entre eux (gaussiens, décorrellés).

Fixons-nous P et supposons qu'il existe un ensemble de suites (b_n, b'_n) de probabilité non nulle telle que pour chacune $\frac{1}{N}\sum_1^N W_n^2(P)$ ne converge pas vers 1. Il y aura donc un ensemble de suite W_n de probabilité égale tel que cette convergence ne se produise pas. On pourrait donc trouver une suite de variables aléatoires gaussiènnes indépendantes entre elles telle que $\frac{1}{N}\sum_1^N W_n^2$ ne converge pas presque sûrement.

Le lemme 1 de § 4-2. montre que cela n'est pas possible , ce qui achève la démonstration.

L'hypothèse

b) $D_N^S(P)$ converge pour tout P vers $D^S(P)$ permet de démontrer le lemme 2.

Lemme 4-3-2. Pour tout P $\frac{1}{N}\sum_1^N V_n(P)\, W_n(P)$ converge presque sûrement vers 0.

Ce lemme se démontre de la même façon que le lemme 1.

Il faut maintenant remarquer que ces convergences qui ont lieu en tout point P pris séparément, restent vraies pour un ensemble dénombrable de points, mais non pour un ensemble ayant la puissance du continu, ce qui est en général le cas pour l'ensemble {P}. La difficulté peut être levée grâce au lemme suivant qui est dérivé du Théorème d'Egorov.

Lemme 4-3-3. (Egorov)

Soit une suite de variables aléatoires $X_n(P)$ qui sont des fonctions continues de P, P appartenant à un compact $\mathcal{P}$. L'ensemble {P} est supposé muni d'une mesure M réelle avec $M(\mathcal{P})<\infty$. Si pour tout P, $X_n(P)$ converge presque sûrement vers une fonction continu x(P), alors pour tout $\eta>0$ il existe un compact $\mathcal{P}_\eta$ inclu dans $\mathcal{P}$ avec $M(\mathcal{P}_\eta) \geq M(\mathcal{P}) - \eta$ tel que sur $\mathcal{P}$, $X_n(P)$ converge presque sûrement uniformément.

L'utilisation de ce lemme pour notre problème exige les hypothèses suivantes :

c) P_o appartient à un compact connu $\mathcal{C}$. L'espace $\{P\}$ est muni d'une mesure M. $\hat{P}_N$ minimise $D_N(P)$ sur $\mathcal{C}$.

d) Les fonctions du modèle $F_n(P)$, $G_n(P)$, $H_n(P)$ sont continues en P, ce qui entraîne la continuité du filtre de Kalman correspondant, donc des fonctions $V_n(P)$ et $W_n(P)$.

e) $D^S(P)$ a un minimum unique en P_o.

On a alors le théorème de convergence :

Théorème 4-3-4.

Moyennant les hypothèses a, b, c, d, e, pour presque tout P_o de $\mathcal{C}$, il existe une suite $\hat{P}_N$ de minimum de $D_N(P)$ qui converge presque sûrement vers P_o.

Démonstration

On peut aussi appliquer le théorème d'Egorov à $D_N^S(P)$ (ce qui permet de ne supposer que sa convergence simple), et donc trouver pour tout $\eta > 0$, $\mathcal{C}_\eta$ tel que presque sûrement uniformément

$$D_N^S(P) \to D(P)$$

$$\frac{1}{N} \sum_1^N V_n(P)\, W_n(P) \to 0$$

$$\frac{1}{N} \sum_1^N W_n^2(P) \to 1$$

avec $M(\mathcal{C}_\eta) \geqslant M(\mathcal{C}) - \eta$

Pour presque tout point P_o de $\mathcal{C}$ on peut trouver η et $\mathcal{C}_\eta$ tels que P_o est un point intérieur à $\mathcal{C}_\eta$. Soit $\hat{P}'_N$ la suite des minimum absolus de $D_N(P)$ sur $\mathcal{C}_\eta$. Une démonstration analogue à celle du théorème permet de montrer que $\hat{P}'_N$ converge presque sûrement vers P_o. P_o étant un point intérieur à $\mathcal{C}_\eta$, à partir d'un certain rang tous les $\hat{P}'_N$ sont intérieurs à $\mathcal{C}_\eta$ et sont donc des minimas relatifs de $D_N(P)$ sur $\mathcal{C}$ ce qui achève la démonstration.

IV - 4. Conclusion

Le théorème 4-3-4 est plus faible que celui démontré pour la regression non linéaire; ceci est dû au fait que le terme de bruit W_n est fonction de P et aux difficultés que cela entraîne. Il est néanmoins satisfaisant et permet de fonder la validité de la méthode d'estimation proposée.

Il est intéressant d'examiner les hypothèses faites et leur implication sur le plan d'un protocole expérimental d'étude d'un système.

Les hypothèses a et d ne posent guère de problème, et l'hypothèse c n'a pas une importance pratique très grande.

L'hypothèse b, (convergence de $D_N^S(P) = \frac{1}{N} \sum V_n^2(P)$) est d'un abord assez compliqué. Sa réalisation n'est pas liée au fait que V_n soit bornée, mais plutôt à des aspects de stationnarité. Un exemple de suite V_n ne donnant pas lieu à convergence est le suivant :

- V_n vaut 0 ou 1.
- Les dates de transition sont telles que $\begin{cases} n_{i+1} = 2_{ni} \\ n_1 = 1 \\ n_2 = 3 \end{cases}$

Aux dates de transition $\frac{1}{N} \sum V_n^2$ vaut alors alternativement $\frac{1}{3}$ ou $\frac{2}{3}$ et donc ne converge pas.

Enfin l'hypothèse a) pose deux problèmes :

- Un problème d'identificabilité : il convient qu'à deux valeurs de P distinctes ne corresponde pas le même système modèle (condition d'injectivité).
- Un problème de sensibilité : il convient que les entrées sensibilisent en permanence les paramètres du modèle. Un exemple classique en régression non linéaire, où cela n'a pas lieu est le modèle de système du premier ordre, sollicité par une entrée impulsionnelle.

La réponse du modèle est $s(t) = x_0 p_1 e^{-t/P_2}$ et la distance sans bruit

$$D_T^S(P) = \frac{1}{T} \int_0^T x_0^2 \left(p_1 e^{-t/p_2} - p_{10} e^{-t/p_{20}} \right)^2 dt$$

où P est le vecteur (p_1, p_2).

Lorsque T tend vers l'infini $D_T^S(P)$ tend vers 0 pour tout P (si $p_2 \geq 0$). L'augmentation de l'horizon T est inutile puisqu'elle apporte de moins en moins d'information sur les paramètres.

V - ETUDE DE L'ESTIMATION DES COVARIANCES

V - 1. Etude théorique

Nous avons vu en II-3.2 que pour l'estimateur proposé au Chapitre II on ne peut estimer les covariances Q et R des bruits de sortie et d'entrée, du fait de deux solutions triviales

$\hat{Q} = 0$ ou $\hat{R} = 0$ qui toutes deux rendent la vraisemblance infinie.

Avant de chercher une méthode d'estimation de ces covariances, nous allons étudier l'existence, dans les données, d'une quantité d'information suffisante pour estimer ces covariances. Nous montrerons en V-1.1 que cette quantité d'information existe. En V-1.2 une méthode sera proposée pour compléter le maximum de vraisemblance à l'aide de propriétés particulières, et les conditions de sa convergence établie. Deux cas seront présentés alors où de telles propriétés particulières peuvent être mises en évidence :

- Le cas de la régression non linéaire "linéarisée" avec bruits d'entrée et de sortie
- Le cas des systèmes dynamiques linéaires avec bruit d'entrée et de sortie.

Dans chaque cas les conditions de convergence seront précisées.

V - 1 - 1. Existence d'un information suffisante

Cette étude se fonde sur l'information de Fisher et l'inégalité de Rao-Cramer dont on peut trouver un exposé dans les ouvrages de statistique classique, par exemple DUGUE [34]:

Inégalité de Rao-Cramer

Soit X_1 un vecteur aléatoire dont la loi dépend d'un vecteur de paramètres, P et P_o la valeur vraie de ces paramètres.

Soit $L(X_1^N; P)$ la vraisemblance des données et T un estimateur de P_o : $\tilde{P} = T(X_1^N)$.

Soit C la matrice mesurant l'écart moyen entre $\tilde{P}$ et P_o :

$$C = E\,[\tilde{P} - P_o][\tilde{P} - P_o]^T$$

Alors sous des conditions de régularité de la fonction de vraisemblance C est "supérieure ou égale" à $I^{-1}(P_o)$ où I est la matrice d'information de Fisher :

$$I\left[P_o\right] = E\left(\frac{\partial}{\partial P} \text{Log } L(X_1^N;P) \Big|_{P_o} \cdot \frac{\partial \text{ Log } L(X_1^N,P)^T}{\partial P}\Big|_{P_o}\right)$$

$$= - E\left(\frac{\partial^2}{\partial P^2} \text{Log } L(X_1^N, P)\Big|_{P_o}\right)$$

<u>Remarque</u> : On entend par supérieure ou égal pour des matrices symétriques le fait que pour tout Y

$$Y^T C Y \geqslant Y^T I^{-1} Y$$

Application au problème

Il est clair que si I n'est pas régulière (existence de valeurs propres nulles) il est inutile de chercher un estimateur des paramètres puisque l'estimation résultante se trouverait infiniment dispersée par rapport à la bonne valeur P_o. Il convient donc d'étudier tout d'abord cette matrice.

Cette étude sera faite dans le cas mono entrée, mono sortie ($N_e = 1$, $N_s = 1$) en reprenant les notations définies en II-3 et en remplaçant les matrices Q et R par les scalaires q et r.

$- \text{Log } L\left(U_1^{*n}, P, q, r\right)$ s'écrit en supprimant les termes constants

$$\sum_n \left\{ \frac{(U_n - U_n^*)^2}{r} + \frac{\left(s_n - f_n(U_1^{*n}, P)\right)^2}{q} + \text{Log}(q.r) \right\}$$

Regroupons les inconnues U_1^{*n} et P en les désignant par le vecteur d'inconnues X, dont la valeur vraie est X_o et soient q_o et r_o les valeurs vraies de q et r. On peut écrire :

$$E\left(- \text{Log } L(X,q,r)\right) = \frac{A(x) + Nr_o}{r} + \frac{B(X)+Nq_o}{q} + N \text{ Log}(q.r)$$

$$= \bar{D}(X,q,r)$$

avec $A(X_o) = 0$ $\quad \frac{\partial A}{\partial X}\Big|_{X_o} = 0$

$B(x_o) = 0$ $\quad \frac{\partial B}{\partial X}\Big|_{X_o} = 0$

Sous les conditions de régularité telles que l'on peut dériver deux fois sous le signe somme, la matrice d'information I(X,q,r) sera donnée par

$$I(X,q,r) = \begin{bmatrix} \frac{\partial^2 \bar{D}}{\partial X^2} & \frac{\partial^2 \bar{D}}{\partial X \partial q} & \frac{\partial^2 \bar{D}}{\partial X \partial r} \\ \frac{\partial^2 \bar{D}^T}{\partial X \partial q} & \frac{\partial^2 \bar{D}}{\partial q^2} & \frac{\partial^2 \bar{D}}{\partial q \partial r} \\ \frac{\partial^2 \bar{D}^T}{\partial X \partial r} & \frac{\partial^2 \bar{D}}{\partial q \partial r} & \frac{\partial^2 \bar{D}}{\partial r^2} \end{bmatrix}$$

Tous calculs faits, il vient au point nominal X_0, q_0, r_0

$$I(X_0,q_0,r_0) = \begin{bmatrix} \frac{1}{r_0} \frac{\partial^2 A}{\partial X^2}\Big|_{X_0} + \frac{1}{q_0} \frac{\partial^2 B}{\partial X^2}\Big|_{X_0} & 0 & 0 \\ 0 & \frac{N}{q_0^2} & 0 \\ 0 & 0 & \frac{N}{r_0^2} \end{bmatrix}$$

I est inversible dès que

$$I'(X_0) = \frac{1}{r_0} \frac{\partial^2 A}{\partial X^2}\Big|_{X_0} + \frac{1}{q_0} \frac{\partial^2 B}{\partial X^2}\Big|_{X_0}$$

est elle-même inversible. Or $I'(X_0)$ n'est autre que la matrice d'information lorsque les variances q_0 et r_0 sont connues. Il est facile de montrer qu'elle est régulière dès que le système modèle est identifiable et que les entrées sont suffisamment sensibilisantes, du moins dans le cas des modèles linéaires stationnaires.

Ces calculs montrent donc qu'il y a en théorie suffisamment d'information pour estimer les variances q et r. De plus cette information croît en N avec la taille de l'horizon : termes en $\frac{N}{q_0^2}$ et $\frac{N}{r_0^2}$. La question de savoir pourquoi la méthode de maximum de vraisemblance ne permet pas de les estimer est un problème de statistique théorique dont la solution ne nous est pas connue. Néanmoins, l'existence de cette information nous incite à rechercher un estimateur de ces variances.

Remarque : Le cas de la régression linéaire, avec bruits de mesure sur l'entrée et la sortie, est développée dans KENDALL et STUART [40]. Lorsque les variances sont inconnues, une solution, pour estimer les paramètres et les variances, due à WALD y est présentée. Cette solution qui est du type variable instrumentale peut s'appliquer à notre problème dans certains cas particuliers : système linéaire stationnaire dont les paramètres inconnus sont les coefficients de la transformée rationnelle en 2 du système. Sa mise en oeuvre reste cependant conditionnée par la possibilité de trouver des variables instrumentales qui soient fortement corrélées aux entrées et sorties déterministes inconnues du système et décorrélées des bruits de mesure correspondants.

V - 1 - 2. Maximum de vraisemblance enrichi

La solution que nous préconisons consiste à adjoindre à la fonction de vraisemblance d'autres fonctions jouissant de propriétés intéressantes pour l'estimation. Par propriétés intéressantes pour l'estimation nous entendons des propriétés pour une fonction $\mathcal{F}_N$ du type :

$$\forall P \neq P_o \quad E\left\{\mathcal{F}_N(X_1^N, P_o)\right\} > E\left\{\mathcal{F}_N(X_1^N, P)\right\}$$

ou $$E\left\{\frac{\partial \mathcal{F}_N}{\partial P}(X_1^N, P_o)\right\} = 0 \quad , \quad E\left\{\frac{\partial \mathcal{F}_N}{\partial P}(X_1^N, P)\right\} \neq 0 \quad \text{pour} \quad P \neq P_o$$

Dans ces expressions X_1^N est une réalisation d'un vecteur aléatoire dont la loi dépend d'un vecteur paramètre P dont la valeur vraie est P_o.

En présence de telles propriétés on est conduit à penser que l'on obtiendra de bons estimateurs en prenant :

Soit $\tilde{P}$ maximise $\mathcal{F}_N(X_1^N, P)$ sur un compact $\mathcal{C}$

Soit $\frac{\partial}{\partial P} \mathcal{F}_N(X_1^N, \tilde{P}) = 0$

Considérons maintenant le problème de l'estimation des variances.

Les paramètres de la fonction de vraisemblance se décomposent en deux vecteurs P et P', où P' représente les paramètres qui ne sont pas estimables par maximum de vraisemblance (par exemple q et r) :

$\frac{\partial}{\partial P} \operatorname{Log}(L(X_1^N, \tilde{P}, P'_o)) = 0$ fournit une bonne estimation de P_o, P'_o étant connu.

En revanche $\frac{\partial}{\partial P} \operatorname{Log}(L(X_1^N, \tilde{P}, \tilde{P}')) = 0$

$$\frac{\partial}{\partial P'} \operatorname{Log}\ (L(X_1^N, \tilde{P}, \tilde{P}')) = 0$$

ne fournit pas une bonne estimation de P_o et P'_o du fait de l'existence de solutions triviales.

Supposons que l'on puisse trouver une fonction $\mathcal{F}_N$ telle que :

$$E\left\{\mathcal{F}_N(X_1^N, P_o, P'_o)\right\} = 0 \text{ et}$$

$$E\left\{\mathcal{F}_N(X_1^N, P_o, P')\right\} \neq 0 \text{ pour } P' \neq P'_o$$

Un premier estimateur que l'on peut proposer est du type :

$$\begin{cases} \frac{\partial}{\partial P} \operatorname{Log} L(X_1^N, \tilde{P}, \tilde{P}') = 0 \\ \mathcal{F}_N(X_1^N, \tilde{P}, \tilde{P}') = 0 \end{cases}$$

Il se peut cependant qu'il n'y ait pas de points $\tilde{P}$, $\tilde{P}'$ vérifiant ce système d'équations. C'est pourquoi on préfèrera construire la fonction :

$$\mathcal{G}_N(X_1^N, P, P') = \left\| \frac{\partial}{\partial P} \operatorname{Log}\ L(X_1^N, P, P') \right\| + \left| \mathcal{F}_N(X_1^N, P, P') \right|$$

l'estimation étant donnée par un point $\tilde{P}_N$, $\tilde{P}'_N$ qui minimise $\mathcal{G}_N$ sur un compact auquel appartient (P_o, P'_o).

Deux types de convergence presque sûre peuvent être envisagées pour cet estimateur en reprenant les démonstrations des théorèmes IV-2.1 et IV-3.4.

Théorème V-1.1.

Si $\mathcal{G}_N$ est continu en P, P', si $(P_o, P'_o) \in \mathcal{C}$ compact si $\mathcal{G}_N$ converge presque sûrement uniformément vers $\mathcal{G}(P, P')$ et si $\mathcal{G}(P, P')$ a un minimum unique en (P_o, P'_o) alors la suite $(\tilde{P}_N, \tilde{P}'_N)$ converge presque sûrement vers (P_o, P'_o).

Théorème V-1.2.

Si $\mathcal{G}_N$ est continue en (P,P'), si $(P_o,P'_o) \in \mathcal{E}$ compact, si $\mathcal{E}$ est muni d'une mesure M, si $\mathcal{G}_N$ converge presque sûrement simplement vers $\mathcal{G}$(P,P') continue, si $\mathcal{G}$(P,P') a un minimum unique en (P_o,P'_o), alors pour presque tout (P_o,P'_o) il existe une suite $(\tilde{P}_N,\tilde{P}'_N)$ de minimas relatifs de $\mathcal{G}_N$ sur $\mathcal{E}$ qui converge presque sûrement vers (P_o,P'_o).

Nous allons voir deux cas où cette méthode s'applique :

Le cas de la régression non linéaire et celui du modèle linéaire dynamique.

V - 2. Cas de la régression non linéaire

Nous présentons ce cas en tant qu'illustration de notre méthode. Son intérêt peut paraître limité pour l'automaticien : ne s'agit-il pas simplement d'identifier une caractéristique non linéaire statique à partir de mesures expérimentales bruitées ? Néanmoins, l'automatique débordant de plus en plus vers des problèmes économiques, biologiques etc... où la question peut se poser, ce cas pourra y trouver son utilité.

V - 2.1. Méthode d'estimation

On considère le modèle non linéaire statique à une entrée, une sortie :

$s_n^* = f(u_n^*, P)$

Les données sont supposées engendrées par le modèle avec $P = P_o$ et bruitées par des bruits blancs indépendants et indépendants entre eux :

$$s_n = s_n^* + b_n \qquad E(b_n) = 0 \qquad E(b_n\, b_m) = q\, \delta_{n,m}$$

$$u_n = u_n^* + b'_n \qquad E(b'_n) = 0 \qquad E(b'_n\, b'_m) = r\, \delta_{n,m}$$

et $E(b_n\, b'_m) = 0$

Le problème consiste à estimer (P_o, q_o, r_o) à partir d'une réalisation S_1^N, U_1^N.

Le cas de ce modèle étant trop compliqué, on va traiter celui du modèle obtenu par linéarisation en u^* du précédent. Les résultats que l'on obtiendra s'appliqueront approximativement au modèle non linéaire, et ce d'autant mieux que les puissances des bruits seront plus faibles.

Le modèle considéré sera donc :

$$s_n^* = f_n(P)\ u_n^* + g_n(P)$$

En appliquant une méthode de maximum de vraissemblance dans l'hypothèse gaussienne, et en fixant q et r, on obtient la distance

$$D(U_1^N,P,q,r) = \frac{1}{N} \sum_{n=1}^{N} \frac{(s_n-s_n^*)^2}{q} + \frac{(u_n-u_n^*)^2}{r} = \frac{1}{N} \sum_{n=1}^{N} d_n(u_n^*,P,q,r)$$

La minimisation en u_n^* peut se mener à bien sur chaque d_n et donne la nouvelle distance.

$$D'(P,q,r) = \frac{1}{N} \sum_{n=1}^{N} d_n'(P,q,r)$$

$$d_n'(P,q,r) = \frac{(s_n-f_n(P).\ u_n-g_n(P)^2}{f_n^2(P)\ r + q}$$

En P_o l'espérance de d_n' vaut :

$$E\left\{d_n'(P_o,q,r)\right\} = \frac{f_n^2\ (P_o)\ r_o + q_o}{f_n(p_o)^2\ r + q}$$

et $\quad E\left\{d_r'(P_o,\ q_o,\ r_o)\right\} = 1$

Il s'en suit que la distance minimum en chaque point dépend de n par l'intermédiaire de f_n^2 (P) sauf :

- si $f_n^2(P)$ ne dépend pas de n (cas de la régression linéaire)
- si les covariances sont correctes.

Cette particularité est mise à profit pour cônstruire la fonction auxiliaire sous forme de corrélation expérimentale :

$$\mathcal{F}_N(P,q,r) = \frac{1}{N}\Sigma_1^N\ d_n'.f_n^2(P) - (\frac{1}{N}\Sigma_1^N\ d_n')(\frac{1}{N}\Sigma_1^N\ f_n^2(P))$$

On vérifie que l'on a bien

$$E\left\{\mathcal{F}_N(P_o,q_o,r_o)\right\} = 0$$

Pour éviter une solution triviale du type $q\ \infty$ ou $r\ \infty$ il faut procéder au changement de variable

$$q = \lambda v$$

$$r = (1-\lambda)\ v$$

et poser $d_n'(P,\lambda) = v\ d_n'(P,q,r)$

On peut alors construire la fonction $_N$

$$\mathcal{G}_N(P, \ ) = \left\| \frac{\partial}{\partial P} D'_N \ (P,\lambda) \right\| + \left| F_N(P,\lambda) \right|$$

La minimisation de $\mathcal{G}_N$ donne $\tilde{P}_N$, $\tilde{\lambda}_N$ et $\tilde{v}_N = D'_N(\tilde{P}_N, \tilde{\lambda}_N)$

V - 2.2. Conditions de convergence

Pour pouvoir appliquer le théorème V-1.1., les conditions suivantes sur le modèle sont requises :

a) P_o appartient à un compact $\mathcal{C}$, d'où (P_o, λ_o) appartient au compact $\mathcal{C} \times \{0,1\}$,

b) Les fonctions $f_n(P)$ et $g_n(P)$ sont continûment dérivables.

c) Toutes les séries impliquées dans la convergence uniforme presque sûre de $\frac{\partial}{\partial P} D'_N \ (P,\lambda)$ et $\mathcal{F}_N(P,\lambda)$ doivent converger uniformément.

Ces conditions s'obtiennent de la façon suivante :

Le développement de chacun de ces deux termes peut s'écrire :

$$\frac{1}{N}\Sigma_1^N X_{1,n}(P,\lambda) + Y_{1,n}(P,\lambda)\ b_n + Y_{2,n}(P,\lambda)\ b'_n + Y_{3,n}(P,\lambda)\ b_n\ b'_n +$$

$$+ Y_{5,n}(P,\lambda)\ b_n^2 + Y_{6,n}(P,\lambda)\ b_n'^2$$

où les fonctions $X_{1,n}$ et $Y_{i,n}$ sont déterministes et ne dépendent que du modèle et des entrées.

Les conditions pour la convergence uniforme qu'il convient de vérifier sont :

1) $\frac{1}{N} \Sigma_1^N X_{1,n}(P,\lambda)$ et pour tout i $\frac{1}{N} \Sigma_1^N Y_{i,n}^2(P,\lambda)$ convergent uniformément.

2) b_n et b'_n ont un moment d'ordre 4.

d) La dernière condition est celle de l'unicité du minimum : il faut vérifier que les fonctions déterministes

$$\lim_{N\to\infty} E\left\{\frac{\partial D'_N(P,\lambda)}{\partial P}\right\}$$

et $$\lim_{N\to\infty} E\left\{\mathcal{F}_N(P,\lambda)\right\}$$

qui existent si les conditions décrites en c sont vérifiées, ne s'annulent simultanément qu'au point (P_o, λ_o).

V - 3. Etude du modèle dynamique linéaire

V - 3.1. Méthode d'estimation

Nous considérons le modèle linéaire tel qu'il est défini en II-5.1, dans le cas d'une entrée et d'une sortie. Si les covariances sont connues, la méthode de maximum de vraisemblance conduit à minimiser la distance

$$D'_N(P) = \frac{1}{N}\sum_1^N \varepsilon^2_n(P) \text{ (théorème II-5.2),}$$

où les $\varepsilon_n(P)$ sont les résidus du filtrage de Kalman. On a vu en II-6.1, 6e propriété, que les résidus peuvent se décomposer en une partie déterministe $v_n(P)$ et une aléatoire $w_n(P)$ avec

$$\varepsilon_n(P) = v_n(P) + w_n(P)$$

De plus, quel que soit P, si les covariances sont correctes, les $w_n(P)$ sont gaussiens indépendants, à moyenne nulle, et à variance unité. En revanche si les covariances sont erronnées, les $w_n(P)$ ne sont en général plus indépendants. Cette propriété est mise à profit pour construire la fonction auxiliaire F_N de la façon suivante :

On pose $r = (1-\lambda)\, v$

$q = \lambda v$

$$D_N(X_o, P, \lambda, U_1^{*N}) = \frac{1}{N}\sum_1^N \frac{(s_n - s_n^*)^2}{\lambda} + \frac{(u_n - u_n^*)^2}{(1-\lambda)}$$

Après filtrage on obtient la nouvelle distance

$$D'_N(P,\lambda) = \frac{1}{N}\sum_1^N \varepsilon^2_n(P,\lambda) = \frac{1}{N}\sum_1^N (v_n(P,\lambda) + w_n(P,\lambda))^2$$

La fonction $\mathcal{F}_N$ est fondée sur la corrélation expérimentale des résidus :

$$\mathcal{F}_N(P,\lambda) = \frac{1}{N-1}\sum_2^N \varepsilon_n(P,\lambda)\,\varepsilon_{n-1}(P,\lambda) - \left(\frac{1}{N-1}\sum_2^n \varepsilon_n(P,\lambda)\right)\left(\frac{1}{N-1}\sum_2^N \varepsilon_{n-1}(P,\lambda)\right)$$

Cette fonction vérifie bien la condition

$$E\left\{\mathcal{F}_N(P_o,\lambda_o)\right\} = 0$$

La méthode d'estimation consiste à choisir $\tilde{P}$, $\tilde{\lambda}$ qui minimisent

$$\mathcal{G}_N(P,\lambda) = \left\| \frac{\partial D_N'(P,\lambda)}{\partial P} \right\| + |\mathcal{F}_N(P,\lambda)|$$

et $\tilde{v} = D_N(\tilde{P},\tilde{\lambda})$

V - 3.2. Conditions de convergence

Il s'agit de vérifier que les conditions du théorème V.1.2. sont remplies :

- P_0, λ_0 est un point intérieur du compact $\mathcal{E} \times [0,1]$ (cela implique en particulier que λ_0 ne vaut ni 1 ni 0. Cette restriction n'est guère gênante car sur des mesures expérimentales, le bruit n'est jamais strictement nul).
- $\mathcal{E} \times (0,1)$ est muni d'un corps de Borel et d'une mesure M
- Les fonctions $F_N(P)$, $G_n(P)$, $H_n(P)$ et celles définissant le filtre de Kalman sont continûment dérivables en P, et en λ pour les dernières.
- Les suites

  $E\left(\frac{\partial D_N'(P,\lambda)}{\partial P}\right.$ et $E(\mathcal{F}_N(P,\lambda))$ convergent vers des fonctions

  $\mathcal{D}(P,\lambda)$ et $\mathcal{F}(P,\lambda)$.
- Le seul point de $\times(0,1)$ vérifiant simultanément

  $\mathcal{D}(P,\lambda) = 0$ et $\mathcal{F}(P,\lambda) = 0$

  est le point P_0,λ_0. Si cette condition n'était pas remplie, il serait possible d'adjoindre au critère $\mathcal{G}_N(P,\lambda)$ d'autres fonctions auxiliaires du type $|\mathcal{F}_{N,k}(P,\lambda)|$ avec

$$\mathcal{F}_{N,k}(P,\lambda) = \frac{1}{N-k} \sum_{k}^{N} \varepsilon_n(P,\lambda)\,\varepsilon_{n-k}(P,\lambda) - \left(\frac{1}{N-k}\sum \varepsilon_n(P,\lambda)\right)\left(\frac{1}{N-k}\sum \varepsilon_{n-k}(P,\lambda)\right)$$

- Il reste à démontrer la convergence presque sûre en tout point P, des termes aléatoires. Ceux-ci sont de deux types :

  Les termes bilinéaires : $\frac{1}{N}\sum_{1}^{N} w_n(P,\lambda)\frac{\partial w_n}{\partial P}(P,\lambda)$, $\frac{1}{N-1}\sum_{2}^{N-1} w_n(P,\lambda)\,w_{n-1}(P,\lambda)$

  Les termes linéaires: $\frac{1}{N}\sum_{1}^{N} v_n(P,\lambda)\frac{\partial w_n}{\partial P}(P,\lambda) + w_n(P,\lambda)\frac{\partial v_n(P,\lambda)}{\partial P}$

  $\frac{1}{N-1}\sum_{2}^{N} w_n(P,\lambda)\,v_{n-1}(P,\lambda) + w_{n-1}(P,\lambda)\,v_n(P,\lambda)$

  $\frac{1}{N-1}\sum_{2}^{N} w_n(P,\lambda)$

a) Etude des termes bilinéaires

Le problème qui se pose est plus complexe que celui étudié au Chapitre IV lemme 4.3.1, du fait que pour λ quelconque il s'agit de moyennes de termes non indépendants. Le problème a été étudié dans le cas où (W_n) est la sortie d'un filtre stationnaire dont l'entrée est un bruit stationnaire blanc (b_n)

$$w_n = \sum_{i=0}^{n} a_{n-i}\, b_i \qquad E(b_i) = 0$$

On trouve, par exemple, dans [36] le résultat suivant :

si $E(b_i^4)$ existe et si $\sum_{i=0}^{n} |(n-i)\, a_{n-i}|$ converge, alors $\frac{1}{N}\sum_1^N W_n^2$ tend presque sûrement vers $\lim \frac{1}{N}\sum_1^N E(w_n^2)$.

Les études sur ce sujet ont évolué [37] vers l'étude de propriétés plus fines de telles suites, dans des conditions de plus en plus générales. Mais il ne semble pas que le cas des systèmes non stationnaires ait été abordé. Plutôt que d'essayer de généraliser ces propriétés au cas non stationnaire, nous nous sommes contentés de les vérifier dans un cadre très restrictif : bruit gaussien, système uniformément asymptotiquement stable, suffisant en ce qui concerne notre propos.

Pour ce faire on utilisera essentiellement la propriété classique suivantes :

Théorème 5.3.1.

Si la suite $E(X_N)$ converge, et s'il existe un moment pair centré $m_{2k,n}$ tel que la série $\sum_n m_{2k,n}$ est convergente, alors la suite X_N converge presque sûrement vers $\lim_N E(X_N)$.

Cette propriété va nous servir à démontrer la convergence presque sûre de $\frac{1}{N}\sum_1^N W_n^2(P,\lambda)$.

Théorème 5.3.2.

Si un système linéaire à état fini est uniformément asymptotiquement stable, si ses entrées sont des bruits gaussiens blancs indépendants entre eux et si $E(\frac{1}{N}\sum_1^N w_n^2)$ converge vers une limite (où w_n est la sortie du système) alors la puissance moyenne de sa sortie $\frac{1}{N}\sum_1^N w_n^2$ converge presque sûrement vers cette limite.

Démonstration

En posant $S_N = \frac{1}{N}\sum_n w_n^2 - E(w_n^2)$, on montre facilement que $E(S_N{}^4)$ est majoré par $\frac{K}{N^2}$. C'est donc le terme général d'une série convergente. S_N' tend donc presque sûrement vers 0. Comme par hypothèse $\frac{1}{N} E(\sum w_n^2)$ converge, $\frac{1}{N}\sum w_n^2$ converge presque sûrement vers cette limite.

Corrollaire 5.3.3.

Si deux systèmes vérifient les hypothèses du théorème 5.3.2 et fournissent pour les mêmes entrées les sorties w_n et w_n' alors $\frac{1}{N}\sum_n (w_n\ w_n' - E(w_n\ w_n')$ converge presque sûrement vers 0.

En effet, le système somme vérifie lui aussi les hypothèses assurant la convergence de

$$\frac{1}{N}\sum(w_n+w_n')^2 - E(w_n+w_n')^2$$

La convergence des termes carrés permet d'assurer celle du terme croisé du développement de l'expression précédente.

Corrollaire 5.3.4.

Si le filtre de Kalman et ses dérivées en P vérifient les hypothèses de stabilité asymptotique uniforme et si les hypothèses de convergence des espérances de $\frac{\partial D_N(P,\lambda)}{\partial P}$ et $\mathcal{F}_N(P,\lambda)$ sont vérifiées alors

$$\frac{1}{N}\sum_1^N w_n(P,\lambda)\frac{\partial w_n(P,\lambda)}{\partial P} \qquad \text{et} \qquad \frac{1}{N-1}\sum_2^N w_n(P,\lambda)\, w_{n-1}(P,\lambda)$$

convergent presque sûrement.

b) Convergence des termes linéaires

La convergence des termes linéaires peut être fondée sur le théorème suivant :

Théorème 5.3.5.

On considère l'expression $S_N = \frac{1}{N}\sum \alpha_{n,N}\, b_n$

où (b_n) est un bruit blanc à moyenne nulle.

Les conditions : 1) $E(b)^2 < \infty$

Il existe une suite β_n telle que :

2) $\sum_n \frac{\beta_n^2}{n^2}$ converge

3) $\frac{1}{N} \sum_n (\alpha_{n,N} - \beta_n)^2 \to 0$ lorsque $N \to \infty$

entrainent la convergence presque sûre de S_N vers 0.

Démonstration

La condition 1 entraîne $\frac{1}{N} \sum_n b_n^2 \xrightarrow[ps]{} E(b^2)$

1 et 2 entraînent $\frac{1}{N} \sum_n \beta_n b_n \xrightarrow[ps]{} 0$

On a alors

$$\left| S_n - \frac{1}{N} \sum_n \beta_n b_n \right| \leqslant \sqrt{\frac{1}{N} \sum_n (\alpha_{n,N} - \beta_n)^2} \sqrt{\frac{1}{N} \sum_n b_n^2}$$

La condition 3 entraîne la convergence presque sûre de S_N vers 0.

Corrollaire 5.3.6.

Sous les hypothèses de stabilité asymptotique uniforme des filtres de Kalman et de leurs dérivées, et de majoration pour tout n des termes $v_n(P,\lambda)$ et $\frac{\partial v_n}{\partial P}(P,\lambda)$, les termes linéaires convergent presque sûrement vers 0.

Démonstration

La démonstration sera donnée pour le terme $S_n = \frac{1}{N} \sum_n v_{n-1}(P,\lambda)\, w_n(P,\lambda)$

$$S_n = \frac{1}{N} \sum_{n=1}^{N} v_{n-1} \left(\sum_{0 \leqslant i \leqslant n} \left(a_{n,i}\, b_i + a'_{n,i}\, b'_i \right) \right)$$

On peut se limiter au premier terme

$$S'_n = \frac{1}{N} \sum_{n=1}^{N} v_{n-1} \sum_{i=0}^{n} a_{n,i}\, b_i = \frac{1}{N} \sum_{i=0}^{N} b_i \left(\sum_{n=i}^{N} v_{n-1}\, a_{n,i} \right) = \frac{1}{N} \sum_i \alpha_{i,N}\, b_i$$

$$\alpha_{i,N} = \sum_{n=i}^{N} v_{n-1}\, a_{i,n}$$

1) $\alpha_{i,N}$ converge. En effet, $|v_{n-1}|\,|a_{n,i}| \leqslant K K_1 e^{-K_2(n-i)}$

terme général d'une série absolument convergente.

On pose $\beta_i = \lim_{N \to \infty} \alpha_{i,N}$

2) $\sum_i \frac{\beta_i^2}{i^\ell}$ converge

En effet $\beta_i^2 \leqslant K^2 K_1^2 \sum_{n=i}^{\infty} e^{-2K_2(n-i)} \leqslant$ constante

3) $\frac{1}{N} \sum_{i=1}^{N} \left(\alpha_{i,N} - \beta_i\right)^2 \longrightarrow 0$ avec N

En effet $\left|\beta_i - \alpha_{i,N}\right| \leqslant \sum_{n=N}^{\infty} \left|v_{n-1}\, a_{n,i}\right| \leqslant K K_1 \sum_{n=N}^{\infty} e^{-K_2(n-i)}$

$$\left|\beta_i - \alpha_{i,N}\right| \leqslant K' e^{-K_2(N-i)}$$

$$\sum_{i=1}^{N} \left(\alpha_{i,N} - \beta_i\right)^2 \leqslant K'^2 \sum_{i=1}^{N} e^{-2K_2(N-i)} \leqslant K''$$

Remarque

Dans cette démonstration, l'hypothèse v_n majoré par K pourrait être ramplacée par l'hypothèse moins forte $v_n \leqslant K\, n^{\frac{1}{2+\nu}}$ avec $\nu > 0$.
En pratique cependant il n'y a guère de difficultés à supposer les entrées déterministes du système majorées.

V - 3.3. Conclusion

On a donc dégagé un ensemble d'hypothèses permettant d'assurer la convergence presque sûre de l'estimateur proposé et donc de fonder la validité de la méthode proposée. Parmi celles-ci, celles qui posent un problème pratique sont :

- L'hypothèse d'unicité du minimum, qui dépend du choix judiciaire de la séquence d'entrée U_1^{*N} stimulant l'objet, les u_i^* devant être par ailleurs bornés.
- Restent les hypothèses de stabilité asymptotique uniforme. Il est facile de montrer que la stabilité du système de sensibilité d'un système linéaire par rapport à un paramètre p_i se déduit de celle du système lui-même :

  Soit un système, par exemple un filtre de Kalman

$$x_{n+1} = F_n(P)\, x_n + G_n(P)\, u_n$$

$$s_n = H_n(P)\, x_n + I_n(P)\, u_n$$

La fonction de sensibilité $\frac{\partial s_n}{\partial p_i}$ par rapport à un paramètre p_i s'obtient en adjoignant au système les équations de sensibilité

$$\frac{\partial x_{n+1}}{\partial p_i} = F_n \frac{\partial x_n}{\partial p_i} + \frac{\partial F_n}{\partial p_i} x_n + \frac{\partial G_n}{\partial p_i} u_n$$

$$\frac{\partial s_n}{\partial p_i} = H_n \frac{\partial x_n}{\partial p_i} + \frac{\partial H_n}{\partial p_i} x_n + \frac{\partial I_n}{\partial p_i} u_n$$

D'où le système de sensibilité

$$\begin{bmatrix} x_{n+1} \\ \frac{\partial x_{n+1}}{\partial p_i} \end{bmatrix} = \begin{bmatrix} F_n & 0 \\ \frac{\partial F_n}{\partial p_i} & F_n \end{bmatrix} \begin{bmatrix} x_n \\ \frac{\partial x_n}{\partial p_i} \end{bmatrix} + \begin{bmatrix} G_n \\ \frac{\partial G_n}{\partial p_i} \end{bmatrix} u_n$$

$$\frac{\partial s_n}{\partial p_i} = \begin{bmatrix} H_n & \frac{\partial H_n}{\partial p_i} \end{bmatrix} \begin{bmatrix} x_n \\ \frac{\partial x_n}{\partial p_i} \end{bmatrix} + \frac{\partial I_n}{\partial p_i} u_n$$

Or les valeurs propres de la matrice $J_n = \begin{bmatrix} F_n & 0 \\ \frac{\partial F_n}{\partial p_i} & F_n \end{bmatrix}$

sont celles de F_n. En effet, il existe un changement de variable $Y_n = R_n x_n$ qui triangularise F_n. J_n est alors triangulaire et a donc les mêmes valeurs propres que F_n.

Il suffira donc de vérifier la stabilité asymptotique uniforme du filtre de Kalman pour tout $(P,\lambda) \in \mathcal{C} \times (0,1)$. Celle-ci est établie dans les conditions du théorème 2-5.4.

VI - EXEMPLE D'APPLICATION

VI - 1. Choix de l'exemple

VI - 1.1. Equation de l'objet

Nous avons choisi d'étudier un système simple, décrit par une équation linéaire stationnaire du deuxième ordre :

$$s_n^* = a \; s_{n-1}^* + b \; s_{n-2}^* + c \; u_n^*$$

avec un pôle stable et un pôle instable, de telle sorte que la simulation du modèle soit aussi difficile que celle du modèle adjoint : les pôles choisis sont réels, inverses l'un de l'autre :

$$p_1 = 2 \qquad p_2 = 0.5$$

La difficulté de simulation du modèle est illustrée par le calcul suivant :

Soit une erreur $\mathcal{E}_0$ sur une condition initale. Sur un horizon de 100 points on aura une erreur d'état finale $\mathcal{E}_{100}$ égale à :

$$\mathcal{E}_{100} = 2^{100} \mathcal{E}_0 \simeq (10^3)^{10} \mathcal{E}_0 = 10^{30} \mathcal{E}_0$$

Ces pôles donnent les valeurs des paramètres

$$a = 2.5 \qquad b = -1.$$

De plus c est pris égal à 1.

VI - 1.2. Simulation de l'objet

L'objet est simulé en boucle fermée avec un régulateur : figure (6.1).

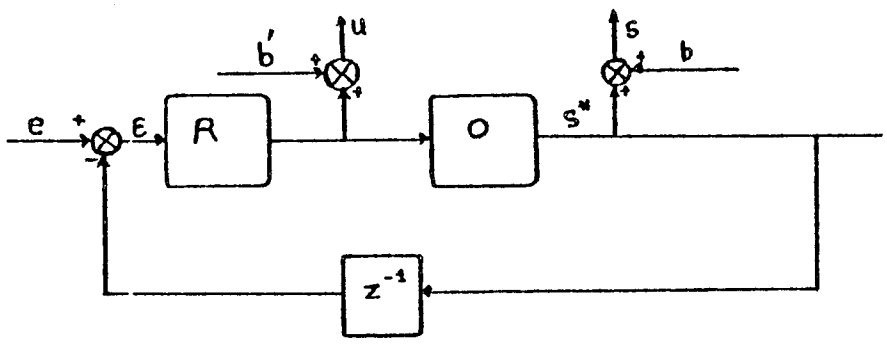


Figure 6.1.

Le régulateur choisi est à réponse plate

$$u_n^* = 2\,\varepsilon_n - \varepsilon_{n-1}$$
$$\varepsilon_n = e_n - s_{n-1}^*$$

où e_n est la valeur de consigne.

Une consigne e a été élaborée sur trois niveaux -1, 0, 1 de façon empirique. La figure 6-2 montre les signaux e, u^*, s^*. L'horizon choisi est N = 100.

Des bruits de mesure b et b' blancs, indépendants, gaussiens à moyenne nulle sont additionnés aux signaux s^*, u^*. Les variances choisies q = 1 r = 1 ont été prises élevées (pour des bruits de mesure) de façon a illustrer la méthode.

La figure 6 .3 représente les signaux u et s qui servent de données à l'identification.

VI - 1.3. Plan d'essais

Dans un premier temps :VI.2 nous avons appliqué la méthode proposées au chapitre II qui suppose le rapport des variances connu :

En VI-2.2. le modèle et le rapport des variances sont corrects.

En VI-2.3. le modèle est correct, le rapport des variances faux. Divers essais sont effectués pour étudier la sensibilité de la méthode à une erreur sur le rapport des variances.

En VI-2.4. le rapport des variances est correct mais le modèle est faux : 1er ordre.

En VI-2.5. toujours pour un rapport de variance exact, le problème plus général de la détermination du modèle par essais successif de modèles, du 1er, 2e et 3e ordre de type général, est étudié.

En VI-3. la méthode proposée au chapitre V est appliquée afin d'examiner la possibilité d'estimer les variances q et r lorsque leur rapport est inconnu.

Enfin, en VI-4. nous examinerons une hypothèse de boucle fermée plus réaliste correspondant au cas où le bruit de mesure b sur la sortie est un bruit de capteurs agissant sur la boucle de retour.

VI - 2. Rapport des variances connues

VI - 2.1. Méthode d'identification

Le critère s'écrit $D(P) = \sum_1^N \varepsilon_i^2 (P)$ où les $\varepsilon_i(P)$ sont obtenus par filtrage de Kalman. Ce critère est optimisé à l'aide du programme non linéaire VA 10 A de la bibliothèque HARWELL [35]. Ce programme met en oeuvre une méthode du 2e ordre, à l'aide de dérivées partielles du critère obtenu par différence finie. Il fournit la matrice des dérivées secondes du critère au point minimum, ce qui permet de calculer les variances des paramètres. On peut noter qu'il ne tire pas profit de la nature "somme de carrés" du critère. La raison de ce choix est qu'il permet d'utiliser le même programme d'optimisation selon que le rapport des variances est connu ou pas.

Le rapport des variances est défini par $\mu = \frac{q}{r}$. Dans cette étude μ est supposé connu donc fixé.

On a $q = r\mu$

r est estimé par $\frac{D(\hat{P})}{N-P}$ où N est l'horizon et p le nombre de paramètres

et q est estimé par $\mu \frac{D(\hat{P})}{N-p}$

Enfin, le temps de calcul en secondes C.P.U. est indiqué pour chaque essai.

VI - 2.2. Modèle et rapport des variances corrects

Deux essais ont été faits :

a) Le point de départ (initialisation de l'algorithme VAIOA) est le point exact : la figure 6.4. montre les résidus ε_n au point nominal.

b) Un point éloigné (1,1,1) est choisi.

Dans les deux cas l'algorithme converge vers deux points très voisins. Les résidus au meilleur point (cas b) sont donnés à la figure 6.5.

Le point correspondant est le suivant :

	objet	meilleur point	écart-type
a	2.5	2.614	0.21
b	-1	-1.18	0.14
c	1.	1.026	0.1
q	1.	1.066	
r	1.	1.066	

TC = 13 secondes

Il apparaît donc que la méthode donne dans ce cas de bons résultats avec un temps de calcul assez court.

VI - 2.3. Modèle correct, rapport des variances faux

Il s'agit d'étudier la sensibilité de la méthode à une erreur sur le rapport des variances.

Dans un premier essai le rapport μ est pris égal à 100 ce qui correspond à un bruit très faible sur l'entrée.

Le point de départ est le point nominal, les résidus au meilleur point sont donnés à la figure 6.6.

Le meilleur point est :

	Objet	meilleur point	écart-type
a	2.5	10.56	1.4
b	-1.	-4.06	0,52
c	1	3.78	0,77
q	1	204.	
r	1	2.04	

TC = 12 second

Un écart important apparaît sur les paramètres. Comme la limite de notre méthode lorsque la variance d'entrée tend vers 0 est la méthode classique de l'erreur quadratique de sortie, cette expérience montre que pour les systèmes instables on ne peut négliger les bruits sur l'entrée et que la méthode classique qui les néglige donne de très mauvais résultats. D'autre part, on peut noter l'aspect non blanc des résidus (figure 6.6), fait sur lequel on s'appuie au chapitre V pour estimer le rapport des variances.

Afin d'estimer la sensibilité à une erreur sur le rapport des variances, des essais successifs ont été faits en multipliant pour chaque essai le rapport μ par 2.

Pour μ = 2 les paramètres restent satisfaisants.

	Objet	modèle	écart-type
a	2.5	2.75	0,23
c	1.	1.047	0,18
b	-1.	-1.175	0.16
q	1	2.6	
r	1.	1.3	

TC = 11,4 secondes

Pour μ = 4 un écart plus important apparaît

	Objet	modèle	écart-type
a	2.5	2.91	0,1
b	-1.	-1.21	0.15
c	1.	1.12	0.12
q	1.	6	
r	1	1.5	

TC = 34 secondes

Cette étude montre donc, pour cet exemple particulier, que la méthode peut tolérer une erreur de 100% sur le rapport des variances. Il ne semble pas qu'une estimation, par l'analyse du processus de mesure, de ce rapport, avec une erreur de cet ordre de grandeur, soit la plupart du temps un objectif impossible à atteindre.

VI - 2.4. Modèle faux, rapport des variances correct

Le modèle choisi pour le système est un modèle du premier ordre :

$$s_n^* = a\, s_{n-1}^* + c\, u_n^*$$

Le résultat est le suivant : les résidus sont donnés à la figure 6.7.

	objet	meilleur point	écart-type
a	2.5	2.09	0,26
b	-1	-	-
c	1.	0,82	0,09
q	1.	1.71	
r	1.	1.71	

TC = 9,5 secondes

Appelons r_1 la valeur obtenue dans cet exemple : r_1 = 1,71 et r_2 = 1.066 celle obtenue en VI-2.2. pour le modèle correct.

Si les rapports de variances sont seuls connus, la diminution sensible de la distance optimale r par adjonction d'un paramètre conduira à préférer le modèle du deuxième ordre à celui du premier ordre (Test de F).

Si les variances sont connues, les quantités $r'_1 = 98\ r_1$ et $r'_2 = 97\ r_2$ seront comparées au seuil de χ^2 à 98 et 97 degrés de liberté respectivement, pour une probabilité α d'erreur donnée.
Pour α= 0.025, et pour un nombre de degrés de liberté important (grand par rapport à 30) le seuil est donné par

$$S = n + 2\sqrt{2n}$$

Pour n = 98 S = 125,8

Pour n = 97 S = 124, 7

Il apparaît clairement que le modèle du premier ordre sera rejeté tandis que celui du 2e ordre sera accepté, et ce en dépit de bruits de mesure relativement importants sur les données.

VI - 2.5. <u>Détermination de l'ordre du modèle</u>

Lorsque la structure du modèle est inconnue, mis à part le fait qu'il doit être linéaire, stationnaire et sans branche directe (degré du numérateur inférieur à celui du dénominateur), la méthode adoptée en général consiste à essayer successivement des modèles du type :

$$s^*_n = \sum_1^{N_0} a_i\, s^*_{n-i} + \sum_0^{N_0-1} b_i\, u^*_{n-i}$$

en accroissant l'ordre N_0 d'une unité à chaque essai.

Pour un deuxième ordre, on obtient le résultat suivant :

	objet	meilleur point	écart-type
a_1	2.5	2.56	0.20
a_2	-1.	-1.07	0.122
b_o	1.	1.0284	0.08
b_1	0.	0.0680	0.095
q	1.	1.07	
r	1.	1.07	

TC = 38 secondes

On posera r_{2g} = 1.07

Pour un troisième ordre on obtient

	objet	meilleur point	écart-type
a_1	2.5	3.04	
a_2	-1	-2.63	
a_3	0.	0.84	
b_0	1.	0.968	
b_1	0.	-0.54	
b_2	0.	0.089	
q	1.	1.08	
r	1.	1.08	

TC = 60 secondes

On posera r_3 = 1.08

Deux remarques s'imposent au sujet de cette identification :

a) Les écarts-type n'ont pas été obtenus, le hessien étant dégénéré au point obtenu : le modèle n'est pas identifiable localement au voisinage du point trouvé.

b) Les paramètres sont éloignés de ceux de l'objet. Cela s'explique d'après ce qui a été dit en a) : il y a au point trouvé une compensation entre un pôle et un zéro de la fonction de transfert du modèle. Le temps de calcul important de l'algorithme s'explique aussi par le même phénomène.

L'examen global des résultats de recherche du bon modèle, résumé dans la donnée des valeurs r_1, r_2, r_{2g}, r_3 permet d'aboutir aux conclusions suivantes :

1) Le fait que r_3 croisse par rapport à r_{2g} conduit au rejet du modèle du 3e ordre. Déjà le fait que le hessien soit dégénéré pour ce modèle est une cause de rejet.

2) Le modèle du premier ordre est rejeté par rapport au deuxième ordre général du fait de la diminution importante de r entre r_1 et r_{2g}.

3) Le deuxième ordre général n apporte aucune diminution de r par rapport au deuxième ordre particulier, et ceci bien qu'il comporte un paramètre de plus. D'autre part, ce paramètre supplémentaire b_1 a été trouvé égal à 0.068 ± 0.095. La valeur 0 est donc largement comprise dans son domaine d'incertitude. Ces raisons conduisent à adopter pour modèle le modèle exact.

Cette étude montre donc que les méthodes classiques de choix du modèle, que nous avons étendu en II.6 au cas de notre méthode d'identification donnent des résultats satisfaisants dans le cadre de cette méthode.

VI - 3. Rapport des variances inconnues

VI - 3.1. Mise en oeuvre de l'identification

En fonction de l'étude faite en V-3. on pose :

$$r = v\lambda$$

$$q = v(1-\lambda)$$

Le critère s'écrit

$$D'(P,\lambda) = \sum_{i=1}^{P} \left| \frac{\partial D(P,\lambda)}{\partial P_i} \right| + \left| Cov\left[\varepsilon_n(P,\lambda), \varepsilon_{n-1}(P,\lambda)\right] \right|$$

Les dérivées partielles de D par rapport à chaque paramètre P_i sont obtenues par différences finies. Le programme VAIOA est utilisé pour minimiser D' par rapport à P et λ. Lorsque le point optimum $\hat{P}$, $\hat{\lambda}$ est obtenu, les variances sont estimées par

$$\hat{r} = \frac{D(\hat{P},\hat{\lambda})}{N} \times \hat{\lambda}$$

$$\hat{q} = \frac{D(\hat{P},\hat{\lambda})}{N} (1-\hat{\lambda})$$

VI - 3.2. Résultats

Divers essais ont été effectués en choisissant des points d'initialisation différents.

a) Lorsque l'on part du point nominal (2.5, -1., 1., 0.5), l'algorithme converge vite vers un point très voisin.

b) Pour divers points initiaux :

(1., 1., 1., 0.9)

(1., 1., 1., 0.5)

(1.25, -2., 0.5, 0.25)

(2., -1.5, 0.75, 0.6)

obtenus en s'approchant de plus en plus du point nominal; l'algorithme converge vers des points éloignés avec des valeurs de D' plus mauvaises que celle trouvée en a).

c) En revanche pour le point initial (2.5, -1, 1., 0.9) paramètres nominaux, rapport des variances faux, l'algorithme retrouve le même point qu'en a). Les résidus en ce point sont donnés à la figure 6-8. Ce point est le suivant :

	Objet	Meilleur point
a	2.5	2.47
b	-1.	-1.096
c	1.	0.97
λ	0.5	0.525
q	1.	1.
r	1.	1.10
v	2.	2.111
$\left\lVert \frac{\partial D}{\partial P} \right\rVert$		0.17
$Cov\,\varepsilon_n \varepsilon_{n-1}$		0.078

TC = 28 secondes

Ce résultat peut être considéré comme satisfaisant en ce sens que sur 6 initialisations différentes essayées, le plus petit minimum trouvé correspond à un point très voisin du point nominal.

Néanmoins il s'avère que ce minimum sera très difficile à trouver dans les cas pratiques où l'on ne dispose pas de "bon" point d'initialisation et de ce fait de l'existence de nombreux minimes locaux. La mise en oeuvre de cette méthode peut donc s'avérer couteuse en temps calcul.

Néanmoins, en rapprochant le fait que la méthode avec rapport des variances connues semble fournir de bons résultats pour les paramètres du système, en dépit d'une erreur importante sur le rapport des variances, et que la méthode avec rapport inconnu nous a donné de bons résultats partant de bons paramètres et d'un mauvais rapport de variances, on peut suggérer la méthodologie suivante :

1) A partir d'une estimation grossière du rapport des variances issue d'une analyse du procédé de mesure, rechercher une valeur des paramètres par la méthode avec rapport connu.

2) Partant de cette valeur des paramètres, améliorer éventuellement la valeur des paramètres et celle du rapport des variances en utilisant la méthode avec rapport inconnu.

VI - 4. Cas du bruit agissant sur la boucle de retour

VI - 4.1. Conditions expérimentales

Les conditions expérimentales correspondent alors à la figure 6.9.

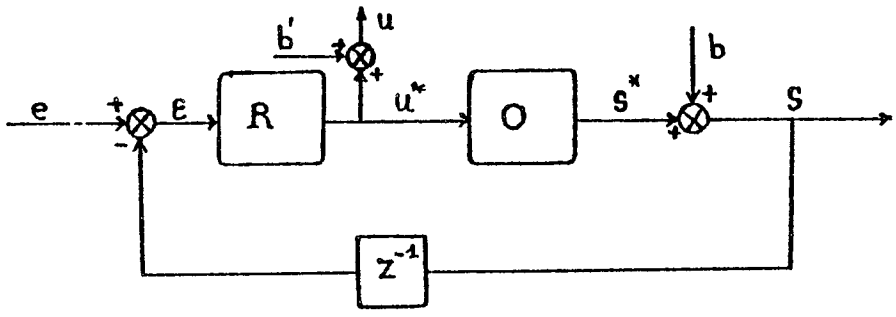


Figure 6.9

Pour les mêmes valeurs de consigne e que précédemment, et pour les mêmes variances de bruits q = 1 et r = 1 les données d'identification (u) (s) sont représentées à la figure 6.10.

VI - 4.2. Analyse du problème

Dans la littérature (38) (39), le problème est envisagé sans tenir compte du bruit de mesure b' sur l'entrée. Le régulateur est supposé linéaire stationnaire (ou stationnaire par morceaux).

L'écart d'identification $\varepsilon_i(z)$ s'écrit en transformée en z :

$$\varepsilon_i(z) = \frac{(O(z) - M(z))\, R(z)\, e(z) + (1 + z^{-1} M(z)\, R(z))\, b(z)}{1 + z^{-1}\, O(z)\, R(z)}$$

où M(z) représente la fonction de transfert du modèle.

Deux valeurs de paramètre sont à considérer :

a) Les valeurs telles que M(z) = O(z). On a alors

$$\varepsilon_i(z) = b(z)$$

b) Les valeurs telles que $M(z) = \dfrac{-1}{z^{-1}\, R(z)}$ si elles existent.

Alors $\varepsilon_i(z) = \dfrac{e(z)}{z^{-1}}$

Si ni b(z) ni e(z) ne sont identiquement nuls et s'il existe des paramètres du modèle tel que $M(z) = \frac{-1}{z^{-1}R(z)}$, l'identification conduira à des valeurs de paramètre intermédiaires entre celles-ci et celles de l'objet. Une condition nécessaire pour que l'estimateur soit asymptotiquement convergent est que le modèle ne puisse représenter l'inverse du régulateur. Il convient donc pour l'identification de choisir un régulateur adéquat (par exemple stationnaire par morceaux). Des conditions suffisantes de convergence peuvent être établies (38) (39).

Dans le cas du protocole que nous avons choisi cette condition nécessaire n'est pas remplie :

$$M(z) = \frac{c}{1+az^{-1}+bz^{-2}}$$

$$\frac{-1}{z^{-1}R(z)} = \frac{1}{-2z^{-1}+z^{-2}}$$

On obtient $M(z) = \frac{1}{z^{-1}R(z)}$ en faisant tendre c vers l'infini avec $\frac{a}{c}$ tendant vers -2 et $\frac{b}{c}$ vers 1.

On peut donc s'attendre à obtenir un résultat biaisé.

Néanmoins, on peut considérer que :

- si la puissance moyenne de la consigne est faible par rapport à celle du bruit, on obtiendra des paramètres tels que $M(z) = \frac{-1}{z^{-1}R(z)}$, le terme $\frac{1 + z^{-1} R(z) M(z)}{1 + z^{-1} O(z) R(z)}$ b(z) étant prépondérant dans le signal d'erreur,

- si au contraire la puissance moyenne de la consigne est grande par rapport à celle du bruit, les paramètres obtenus seront voisins de ceux de l'objet, le terme $\frac{(O(z) - M(z) R(z) e(z)}{1 + z^{-1} O(z) R(z)}$ devenant alors prépondérant.

Dans le cas où l'on considère le bruit b' sur les entrées et la méthode d'identification que l'on propose, l'analyse du problème est beaucoup plus complexe. Les identifications qui suivent ont pour objet de vérifier si les considérations qualitatives développées ci-dessus demeurent valables.

VI - 4.3. Identifications avec rapport des variances connu exact

Le modèle choisi est un modèle général du deuxième ordre :

$$s_n = a\, s_{n-1} + b\, s_{n-2} + c\, u_n + d\, u_{n-1}$$

Des essais successifs ont été réalisés avec des variances de bruit de sortie q décroissantes : 1, 0,5, 0,16, 0,04.

La puissance moyenne de la consigne étant voisine de 1, ces valeurs correspondent aussi au rapport $\frac{\textit{variance du bruit de sortie}}{\textit{puissance moyenne de la consigne}}$.

Des résultats satisfaisants sont obtenus pour $q \leq 0{,}16$

Par exemple pour q = 0,16 on obtient le point suivant.

	Objet	Modèle	écart-type
a	2.5	2.73	0.14
b	-1.	-1.24	0.08
c	1.	1.03	0.05
d	0.	-0.079	0.06
q	0.16	0.13	
r	1	0.83	

TC = 18 secondes

Pour q = 0.04 on obtient :

	Objet	Modèle	Ecart-type
a	2.5	2.53	0.12
b	-1.	-1.14	0.069
c	1.	.0.98	0.057
d	0.	-0.067	0.041
q	0.04	0.04	
r	1.	1.	

Ces résultats montrent donc que, même dans le cas où l'inverse du régulateur est susceptible d'être représenté par le modèle, les résultats d'identification peuvent être bons dès que le rapport $\frac{consigne}{bruit}$ est supérieur à 2.5 (en amplitude).

VI - 4.4. Identification avec rapport des variances connu faux

On considère l'objet simulé avec des variances :

$q_o = 0.04$

$r_o = 1$

soit un rapport de variances $\mu_o = \frac{q_o}{r_o} = 0.04$

L'identification est réalisée avec un rapport $\mu = 0.16$ soit un rapport 4 entre μ et μ_o. Le résultat est le suivant :

	Objet	Modèle	Ecart-type
a	2.5	2.58	0.127
b	-1.	-1.18	0.07
c	1.	0.97	0.049
d	0.	-0.085	0.05
q	0.04	0.11	
r	1	0.73	

Ce résultat montre que la méthode proposée peut tolérer une erreur sur le rapport des variances importante.

VI - 4.5. Identification avec rapport des variances inconnu

La méthode avec rapport de variances inconnu a été appliquée à l'objet correspondant à $q = 0.04$, $r = 1$.

Des essais successifs ont été menés à partir de plusieurs valeurs d'initialisation. Dans chaque cas l'algorithme converge vers un point éloigné du point nominal. En revanche, lorsque l'on initialise l'algorithme au point nominal, on obtient un point très voisin avec un critère meilleur que ceux précédemment obtenus. On retrouve donc là le problème pratique d'optimisation du critère proposé, compte tenu de la non unicité du minimum.

VI - 4.6. Conclusion

Les résultats obtenus lorsque le bruit de sortie agit sur la boucle de retour du système apparaissent donc tout à fait comparables avec ceux obtenus lorsque ce n'est pas le cas, à la condition que ce bruit soit faible par rapport à la consigne (rapport signal bruit supérieur à 2,5).

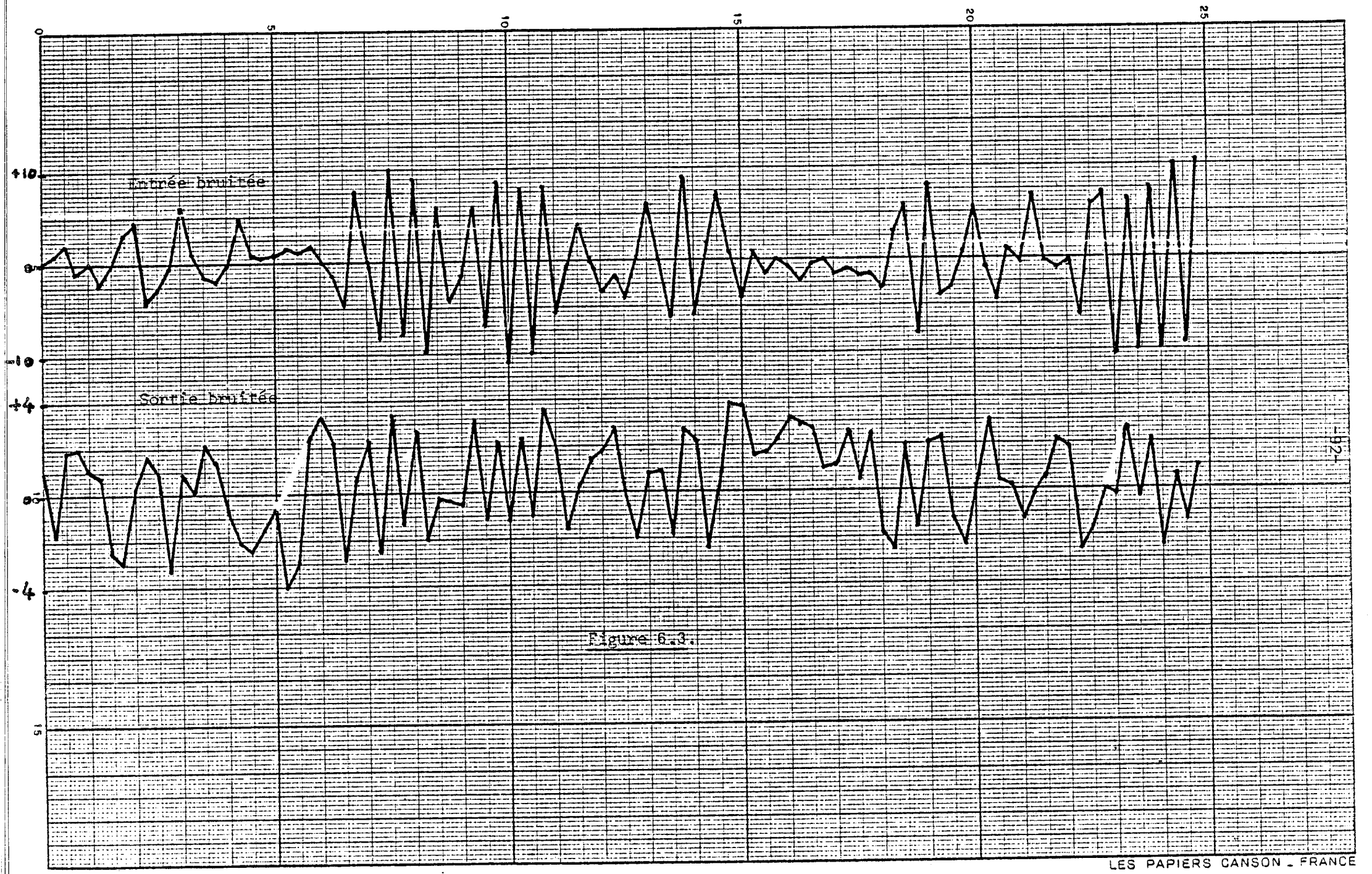


Figure 6.3.

LES PAPIERS CANSON - FRANCE

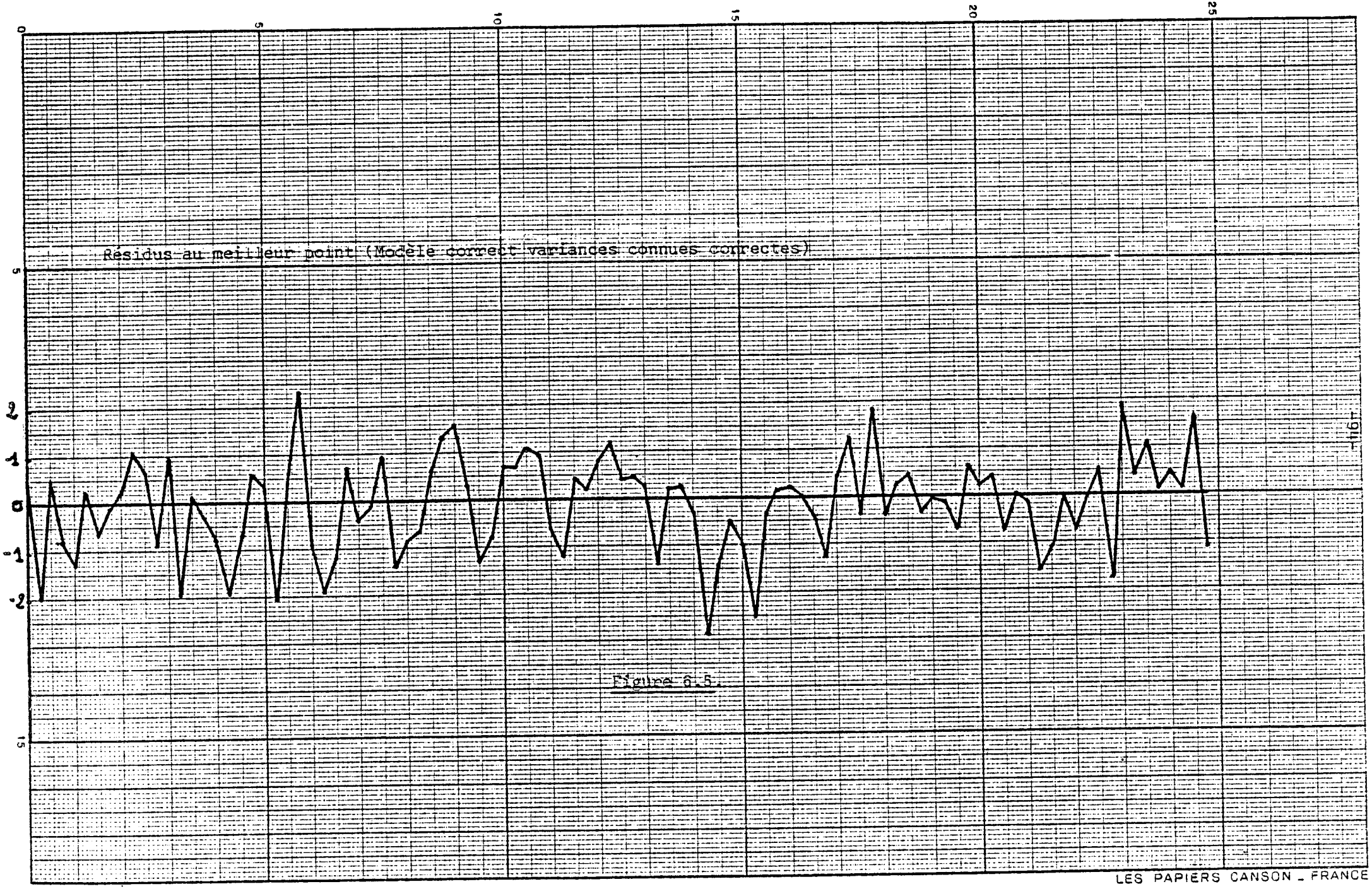


Figure 6.5.

LES PAPIERS CANSON - FRANCE

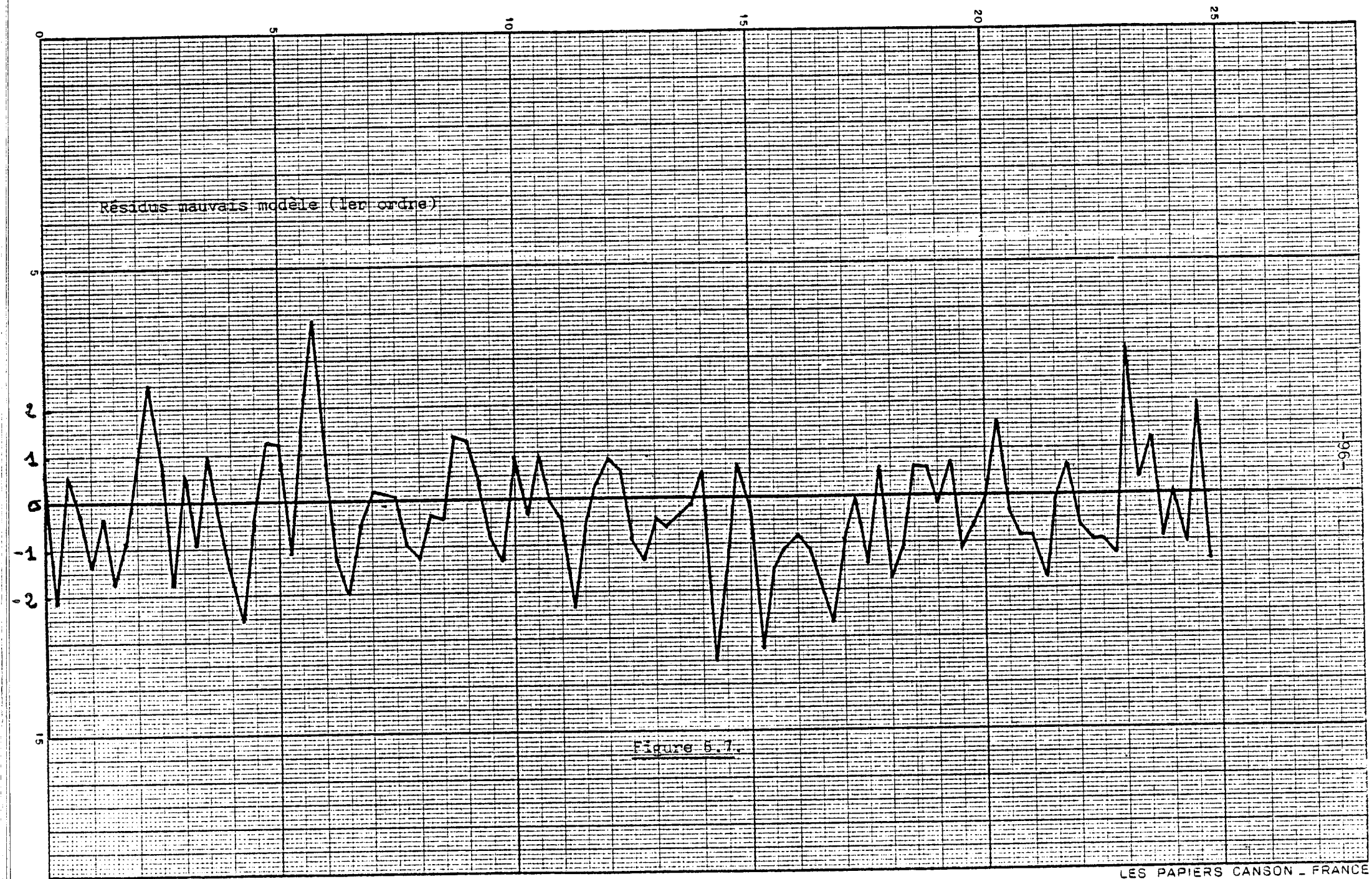


Figure 5.7.

LES PAPIERS CANSON - FRANCE

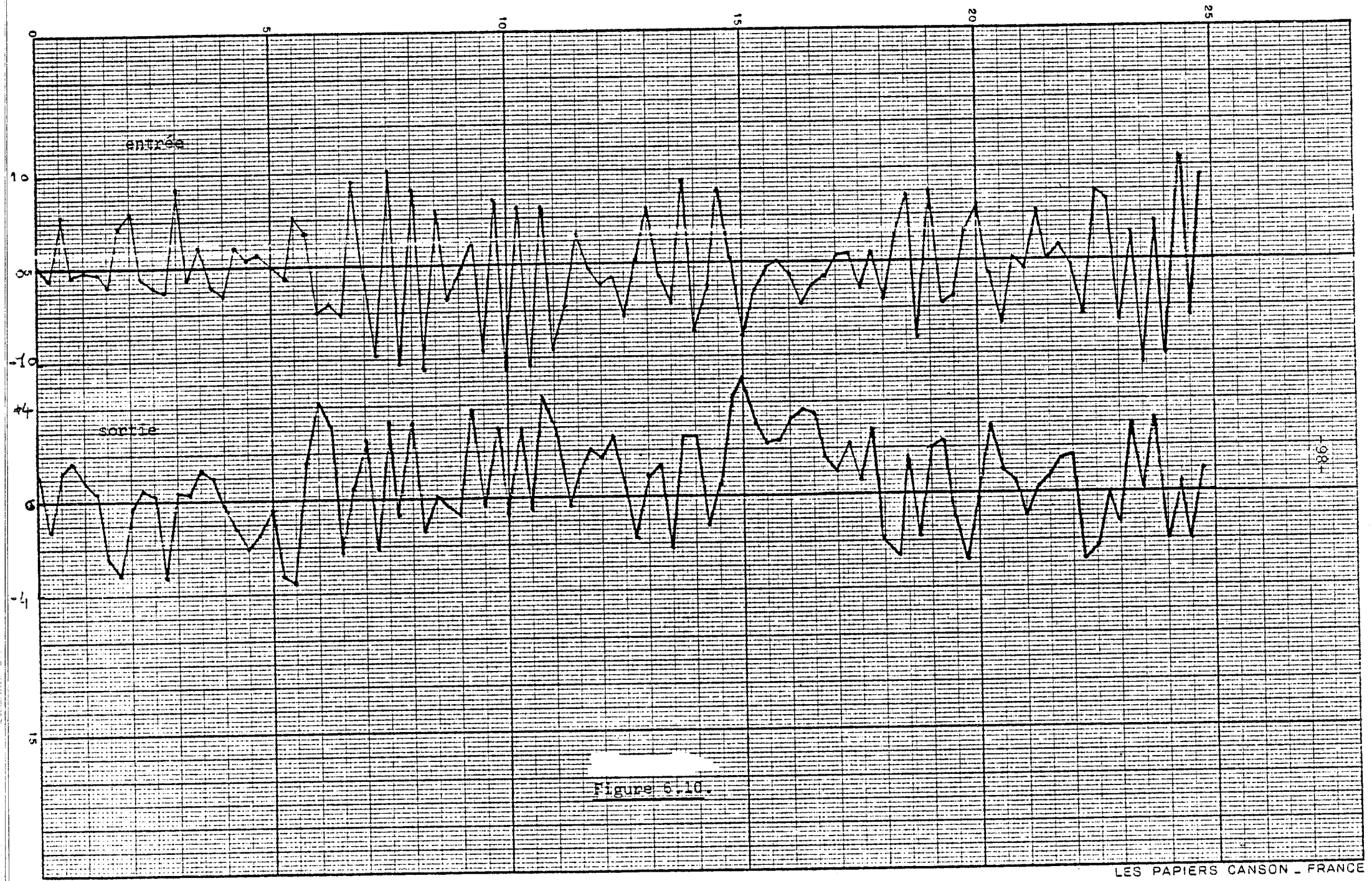

0
5
10
15
20
25
entrée
10
5
-10
sortie
4
6
-4
15
LES PAPIERS CANSON - FRANCE

Figure 6.10.

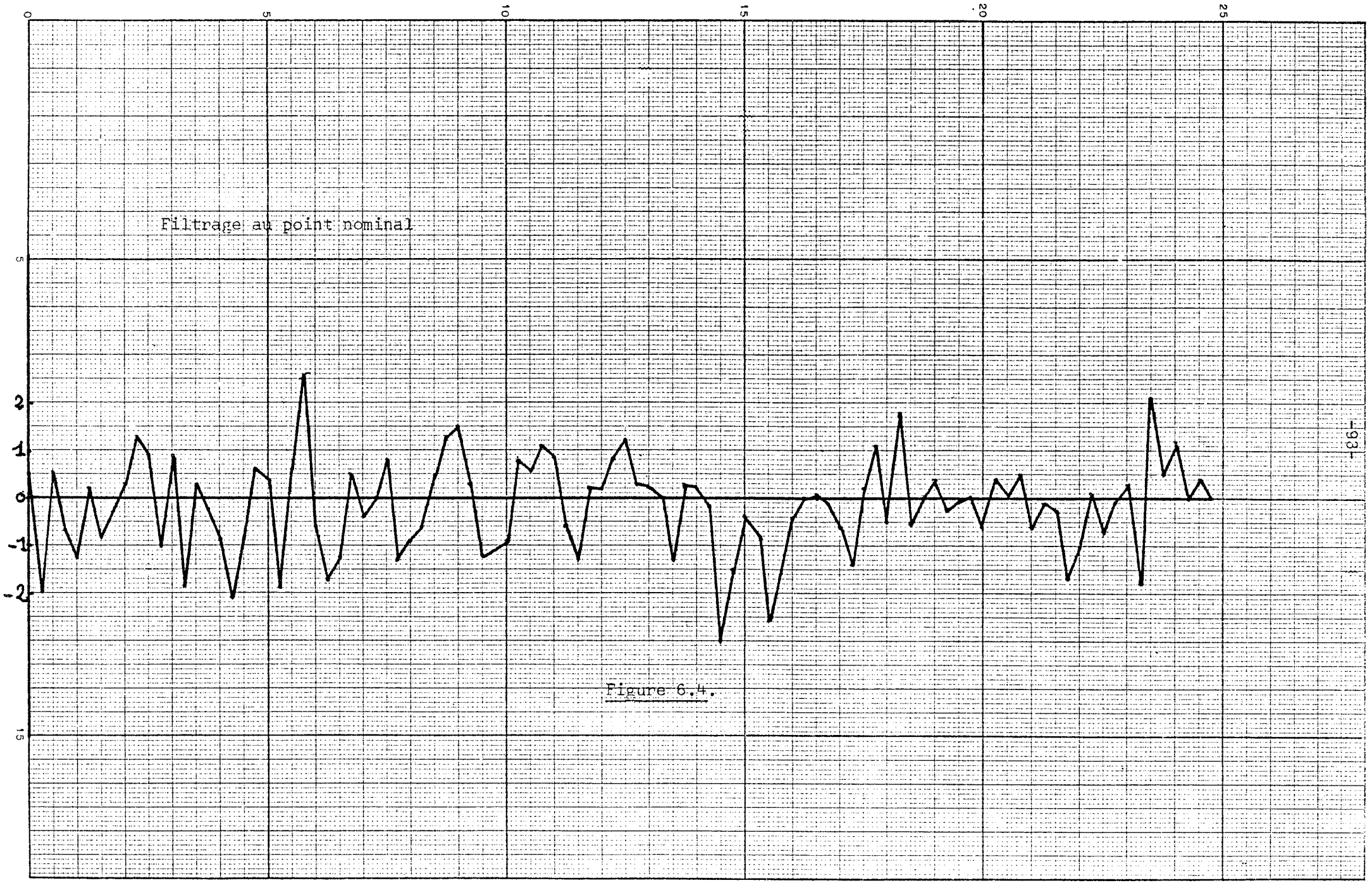


Figure 6.4.

LES PAPIERS CANSON _ FRANCE

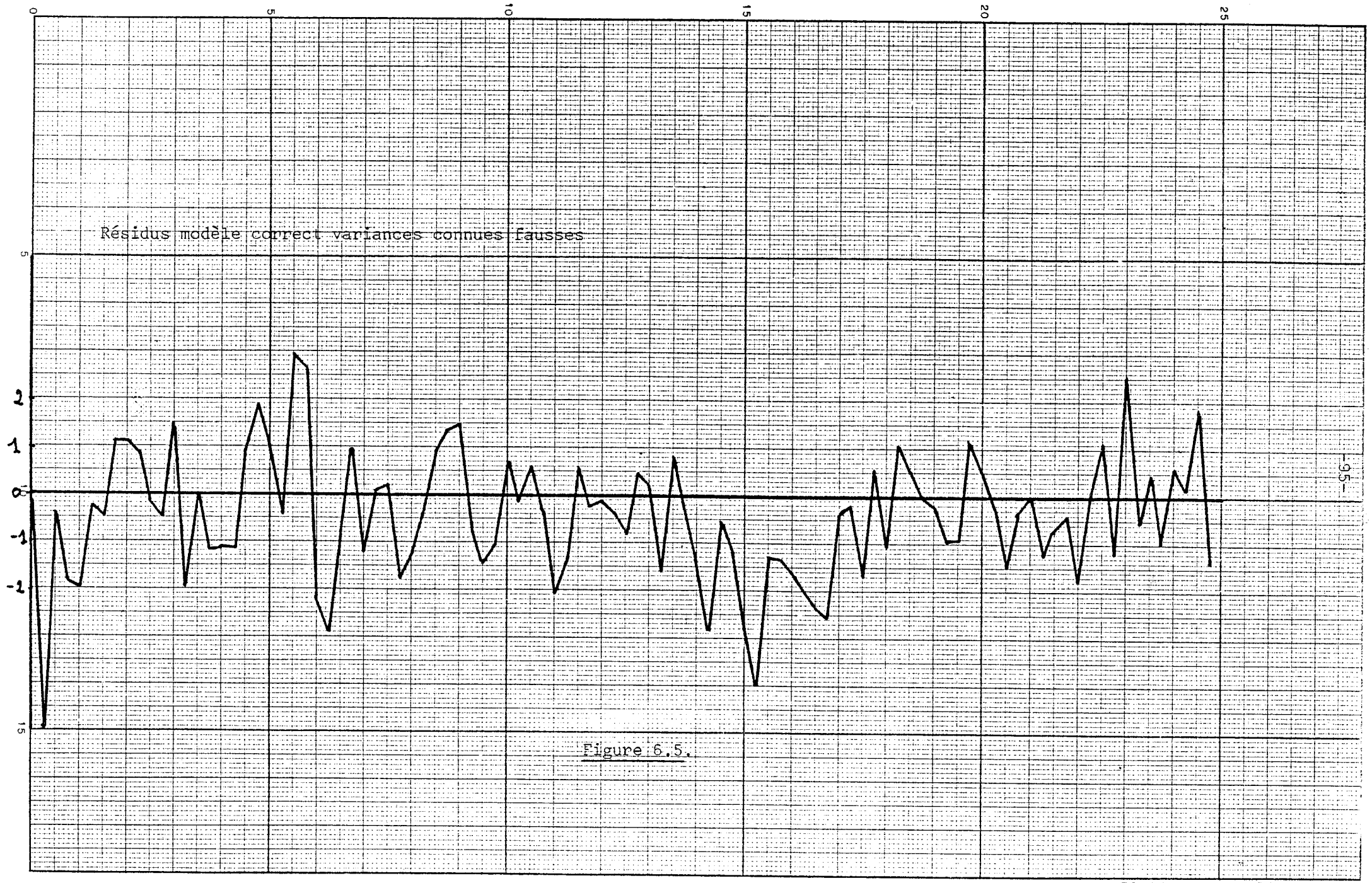


Figure 6.5.

LES PAPIERS CANSON - FRANCE

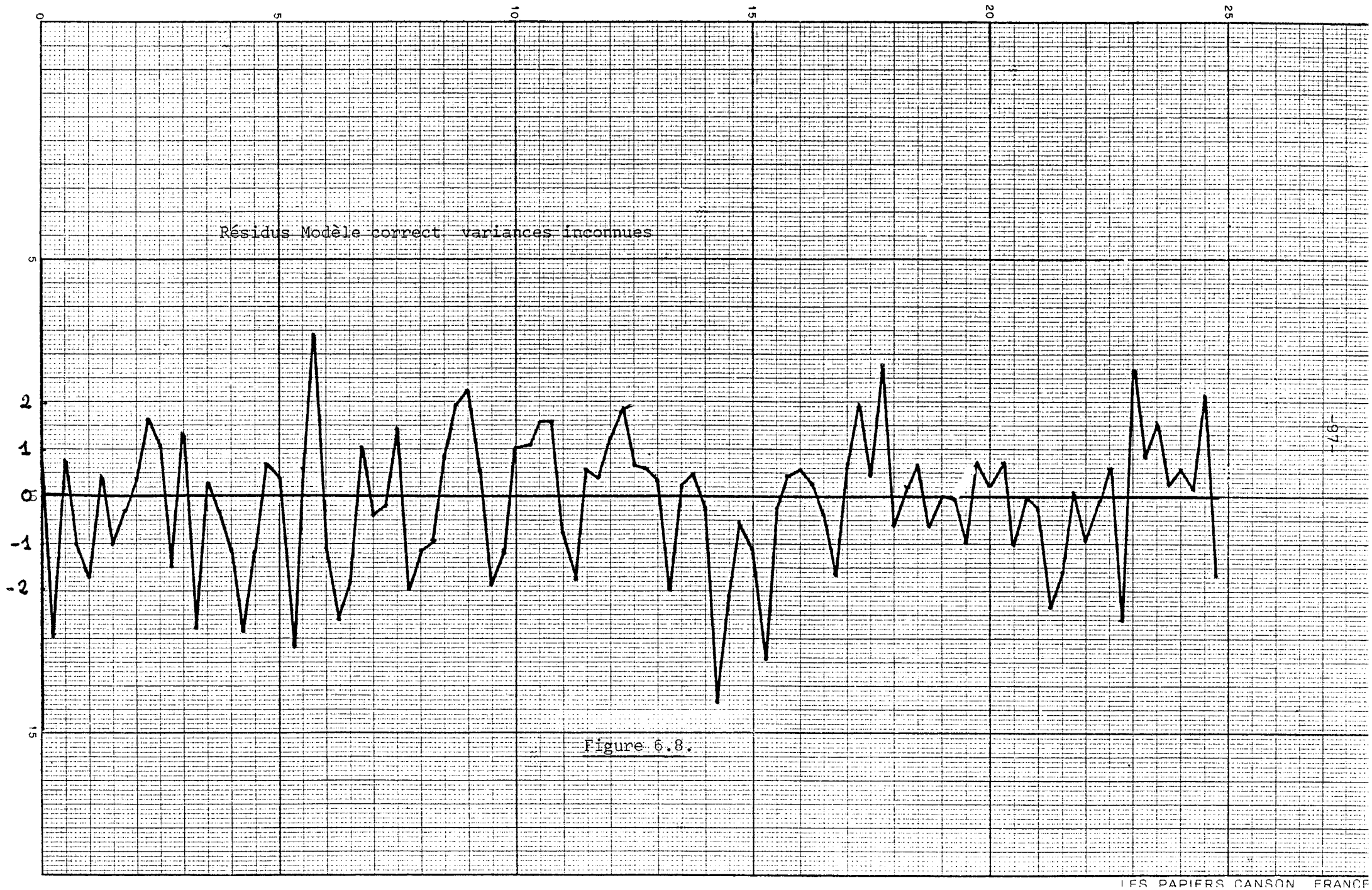


Figure 6.8.

LES PAPIERS CANSON FRANCE

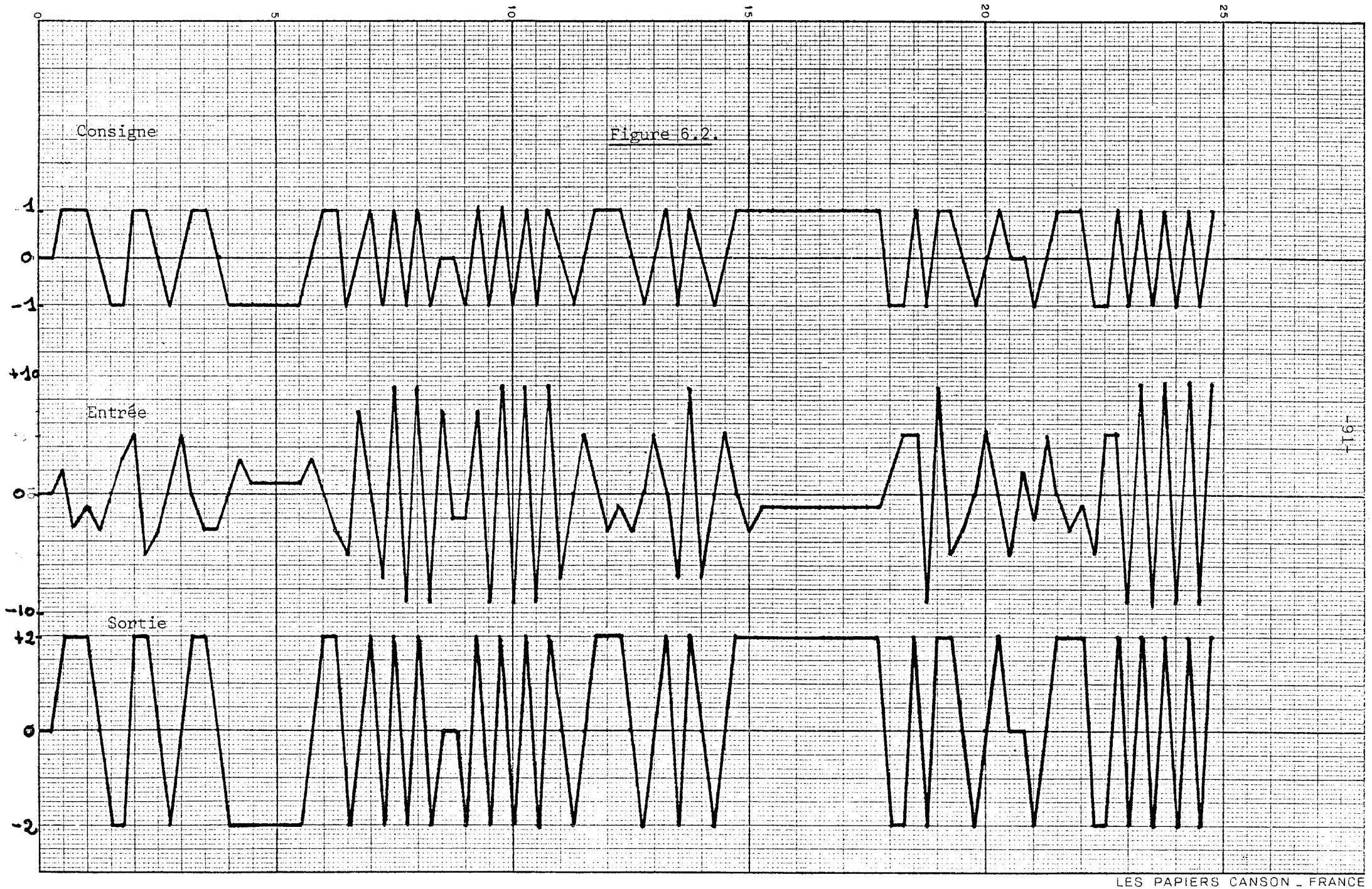

0
5
10
15
20
25
Consigne
1
0
-1
Entrée
+10
0
-10
Sortie
+2
0
-2
LES PAPIERS CANSON - FRANCE

Figure 6.2.

CONCLUSION

Dans ce travail deux méthodes ont été présentées :

Premièrement, une méthode générale d'identification des systèmes déterministes linéaires à équation d'état. Cette méthode prend en compte les bruits de mesure sur les entrées et les sorties du système bruits dont les covariances doivent être connues à un coefficient près. Elle paraît donc bien adaptée à l'identification des systèmes instables pour qui les bruits sur l'entrée ont une importance considérable. Cette méthode permet d'estimer les paramètres du système, d'en donner un intervalle de confiance, de tester le modèle dans l'absolu et de comparer entre eux divers modèles emboités, donc de pouvoir modifier éventuellement la structure du modèle. D'autre part, en ce qui concerne les systèmes instables elle présente l'intérêt de remplacer la simulation du modèle par un filtrage de Kalman uniformément asymptotiquement stable sous des conditions assez peu restrictives.

Une étude théorique a montré que cette méthode fournissait un estimateur presque sûrement convergent. D'autre part, une étude en simulation (Chapitre VI) sur un exemple simple a montré qu'elle donnait de bons résultats en ce qui concerne l'estimation des paramètres et en ce qui concerne le choix du modèle. En dépit de bruits de mesure volontairement assez importants, donc avec des mesures d'assez mauvaise qualité, nous avons pu cependant déterminer exactement la structure du modèle. Cette étude en simulation a aussi montré que la méthode tolérait une erreur assez importante dans la connaissance des covariances des bruits.

La deuxième méthode vise à compléter la première lorsque les covariances des bruits sont inconnues, du moins dans le cas monoentrée, monosortie. Il s'agit d'une méthode d'estimation à la fois des paramètres et des variances des deux bruits. Elle s'applique au cas des systèmes linéaires à equation d'état, et au problème de régression non-linéaire (identification d'un modèle monodimensionnel non linéaire statique). Sur le plan théorique nous avons pu fournir des conditions pour en assurer la convergence.

En simulation nous avons pu montrer que le minimum du critère proposé fournit une bonne estimation des paramètres du système et des variances des bruits, mais que ce minimum était difficile à atteindre du fait de la présence de nombreux minimes relatifs. Il y a donc là de sérieuses difficultés de mise en oeuvre. Néanmoins, étant donné que la première méthode tolère

des erreurs assez grandes sur les variances, il semble que la combinaison des deux puisse résoudre le problème.

A la suite de ce travail, d'autres études pourront être entreprises pour préciser la qualité de l'estimation de la première méthode. Par exemple, il devrait être possible de démontrer sa normalité asymptotique.

Sur le plan pratique, il reste à trouver à ces méthodes des débouchés afin que ce travail ne reste pas stérile. De tels débouchés peuvent naturellement être trouvés dans le domaine de l'automatique puisqu'en effet le problème qui a motivé cette recherche était de nature automatique. Néanmoins, ces méthodes qui traitent de systèmes déterministes semblent actuellement se situer à contre courant de la tendance en automatique qui consiste à modéliser les systèmes sous forme de processus stochastiques de façon à tenir compte des perturbations aléatoires, des phénomènes secondaires négligés et des erreurs de caractérisation. Il nous semble donc, comme nous avons essayé de le montrer au Chapitre III qu'un certain nombre de domaines comme la biologie, l'écologie, la démographie, voire l'économie, qui commencent à intéresser les automaticiens, étudient des phénomènes instables, déterministes en ce sens qu'ils portent sur des variables représentant des grands nombres, et mal mesurés (sondages), devraient fournir un terrain d'application privilégié pour ces méthodes.

BIBLIOGRAPHIE

1) J. RICHALET, A. RAULT, R. POULIQUEN : "Identification des Processus par la méthode du modèle", Gordon and Breach 1971.

2) P. CASPI, O.ESMENJAUD, A. MALLET, A. RAULT : "Elaboration of a system of preventive personnalized maintenance of jet engins", 4th Int. Symp. on Engine Testing and Automation, Naples 1975.

3) A. BARRAUD : "Identification de systèmes multivariables par modèle d'état sous forme canonique", RAIRO (Automatique) Vol.11 n°2, pp.161-194, 1977.

4) K.J. ASTRÖM and T. BOHLIN:"Numerical identification of linear dynamic systems from normal operating records", Theory of self Adaptation Systems (P.H. Hammond ed.) pp.94-111 Plenum Press N.Y 1966.

5) R.K. MEHRA, J. TYLER : "Case studies in aircraft parameter identification" 3nd I.F.A.C. Symp Ident and Par. Est. La Haye 1973.

6) K.Y. YOUNG, E. POLAK : "Identification of linear discrete-time systems using the instrumental variable method", I.E.E.E. Trans. on Aut. Cont. Vol. AC 12 n°6, pp.707-718, Dec. 1967.

7) M.J. LEVIN : "Estimation of a system pulse transfer function in the presence of noise", I.E.E.E. Trans. on Aut. Cont, pp.229-235, juillet 1964.

8) A. GROSJEAN, C. FOULARD : "Extensions of the Lovin's method for the identification of discrete linear multivariable stochastic dynamic systems", 4th IFAC Symp. Ident. and Syst.Par.Est, Tbilisi, Sept. 1976.

9) G.S. SARIDIS, G. STEIN : "Stochastic Approximation Algorithms for linear discrete time system identification", I.E.E.E. Trans. Aut. Cont. Vol.AC 13 n°1, pp.37-44, Feb. 1968.

10) J. AGUILAR-MARTIN, Ph. de LARMINAT : "Récent dévelopment in France of stochastic filtering and its application to identification", J. Dyn. Mes. and Cont. pp.421-425, Dec. 1974.

11) K. STEIGLITZ, M.E. Mc. BRIDGE : "A technique for the identification of linear systems", I.E.E.E. Trans. Aut. Cont. Vol. AC 10 n°4, Oct. 1965.

12) D.W. CLARKE : "Generalized least squares estimation of the parameters of a dynamic model", 1st IFAC Symp. Ident. in Aut. Cont. Systems, Prague 1967.

13) P. CASPI, P. LEROUX, Ph. TRANGOSI, A. RAULT, J.C. DUPONNOIS : "Identification of unstable linear systems by the maximum likelihood approach. Application ot the hydrofoil H 890", 4th IFAC Symp. Ident. and Syst. Par. Est, Tbilisi, Sept. 1976.

14) S.L. FAGEN : "Recursive linear Regression theory, optimal filter theory and error analysis of optimal systems", IEEE. Int. Conv. Record. 12, pp.216-240, 1964.

15) H. COX : "Estimation of state variables via dynamic programming", Proc. 1964, joint Aut. Conf. Stanford Cal., pp.376-381, 1964.

16) A.H. JAZWINSKI : "Stochastic Processes and filtering theory", Academic Press, New York 1970.

17) R.E. KALMAN : "A new approach to linear filtering and prediction", Journal of Basic Eng. Vol. 82, pp.35-45, Mars 1960.

18) T. KAILATH : "An innovations approach to least squares estimation, Part.I" IEEE Trans. on Aut. Cont. Vol.AC-13 n°6, Dec. 1968.

19) A. ALBERT : "Regression and the Moore. Penrose pseudo-inverse" Academic Press, New York 1972.

20) D.W. MARQUARDT : "An algorithm for least squares estimation of non-linear parameters", J. SOC. Indust. Appl. Math 11,pp.431-441, 1963.

21) R.E. KALMAN : "New results in Wiener Filtering Theory", Proc. Symp. Eng. Appl. Random Function Theory and Probability, Wiley New York 1963.

22) J.J. DEYST and C.F. PRICE : "Conditions for Asymptotic Stability of the Discrete Minimum Variance Linear Estimator", IEEE Trans. AC. Vol. AC 18 pp.702-705, Dec. 1968.
Corrigé dans :
J.J. DEYST : "Correction to "Conditions for asymptotic stability of the minimum variance linear estimator"", IEEE Trans. AC Vol. AC 18 n°5, pp.562-563 Oct. 1973.

23) N. DRAPPER and H. SMITH : "Applied regression analysis", Wiley, New York 1966

24) E.M.L. BEALE : "Confidence Regions in non linear estimation", J. Roy Statis. Soc. B22, pp.41-76, 1960.

25) H. UNBEHAVEN, B. GOHRING : "Tests for determining model order in parameter estimation", Automatice vol 10 pp.233-244, 1974.

26) A.J.W. VAN DEN BOOM, A.W.M. VAN DEN ENPEN : "The determination of the orders of process and noise dynamics", Automatica vol 10, pp.245-256, 1974.

27) K.J. ASTRÖM : "On the achievable accuracy in identification problems", 1st IFAC Symp. on Identification in Automatic Control Systems, Prague 1967.

28) R.K. MEHRA : "Identification of Stochastic linear dynamic systems using Kalman filter representation", AIAA Journal vol.9, n°1, Jan. 1971.

29) K.J. ASTROM ans S. WENMARK : "Numerical identification of stationnary time series", 6th Int. Instruments and Measurements Congress, Stokholm Sept. 1964.

30) R.J. KASHYAP : "Maximum likelihood identification of stochastic linear dynamic systems", IEEE Trans. Aut. Cont., Feb. 1970.

31) Rapport GERBIOS : "Modèle du cheptel bovinfrançais" 1975.

32) R.I. JENNRICH : "Asymptotic properties of non linear least squares estimators", Annals of Math. Stat. Vol. n°2, pp.633-643, 1969.

33) T. SODERSTRÖM : "Convergence properties of the generalized least squares identification method", Automatica Vol.10, pp.617-626, 1974.

34) D. DOGUE : "Traité de statistique théorique et appliquée", DUNOD, Paris 1959.

35) Librairie HARWELL : Laboratoire d'Automatique de Grenoble.

36) T.W. ANDERSON, A.M. WALKER : "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic system", Annals of Math. Stat. vol.35, pp.1296-1303, 1965.

37) E.J. HANNAN : "The asymptotic distribution of serial covariances", Annals of Math. Stat. vol.4, n°2, pp.396-399, 1976.

38) T. SODERSTRÖM, L. LJUNG, I. GUSTAVSSON : "Identificability conditions for linear multivariable systems operating under feed-back", IEEE Trans. AC, vol. AC 21 n°6, pp.837-840, 1976.

39) I. GUSTAVSSON, L. LJUNG, T. SODERSTROM : "Identification of Processes in closes loop : Identificability and accuracy aspects", AUTOMATICA vol.13, pp.59-75, 1977.

40) KENDALL, STUART : "The advanced theory of statistics" Vol.2, pp.375-415, Griffin London.

AUTORISATION DE SOUTENANCE

VU les dispositions de l'article 5 de l'arrêté du 16 Avril 1974,

VU les rapports de Messieurs :

- J. AGUILAR-MARTIN, Maître de Rcherches au C.N.R.S. - L.A.A.S. TOULOUSE -
- Cl. FOULARD, Professeur à l'Institut National Polytechnique de Grenoble
- J. RICHALET, Directeur de l'ADERSA-GERBIOS -VELISY.

Monsieur Paul C A S P I

est autorisé à présenter une thèse en soutenance pour l'obtention du grade de DOCTEUR D'ETAT ES-SCIENCES.

Grenoble, le 12 Mai 1978

Le Président de l'U.S.M.G.

Le Président de l'I.N.P.G.

Ph. TRAYNARD
Président
de l'Institut National Polytechnique