Sur quelques structures d'information Intervenant en jeux, dans les problèmes d'équipe ou de contrôle et en filtrage

Jean Lévine

THESE

DE DOCTORAT D'ETAT ES SCIENCES MATHEMATIQUES

présentée à

L'Université Paris 9 - Dauphine
UER de Mathématiques de la Décision

pour obtenir le grade de

DOCTEUR ES SCIENCES

par

Jean LEVINE

Sujet de la Thèse :

Sur Quelques structures d'Information Intervenant en Jeux, dans les Problèmes d'Equipe ou de Contrôle et en Filtrage.

Directeur de Recherche :
Alain BENSOUSSAN

Soutenue le 19 Novembre 1984

Devant la commission composée de MM A. BENSOUSSAN
P. BERNHARD
M. FLIESS
J.M. LASRY
P.L. LIONS
E. PARDOUX

A Martine et Sonia

REMERCIEMENTS

Que cette thèse soit la preuve de ma reconnaissance envers tous ceux qui, de près ou de loin, ont partagé mes préoccupations, m'aidant à assembler au fil des années les pièces de cet édifice resté, hélas, très imparfait.

Je tiens à exprimer ma profonde gratitude tout particulièrement à A. BENSOUSSAN qui m'a souvent témoigné sa confiance et m'a prodigué les conseils et encouragements sans lesquels ce travail n'aurait sans doute pas vu le jour; à P. BERNHARD qui m'a initié à la plupart des domaines abordés dans ce mémoire et qui a si bien réussi à me transmettre son enthousiasme pour la recherche appliquée Je lui dois ma présence au CAI de l'Ecole des Mines; à M. FLIESS, J.M LASRY, P.L. LIONS et E. PARDOUX qui ont souvent manifesté leur intéret pour mon travail, le prouvant par leurs nombreux conseils, et qui me font l'honneur et l'amitié de participer à ce jury.

J'adresse aussi mes sincères remerciements à tous les co-signataires des travaux ici présentés: J. THEPOT, G. PIGNIE, A. KASINSKI, F GEROMEL et P.WILLIS, ainsi qu'à tous mes collègues du CAI, G. COHEN, Y. LENOIR, L. PRALY et tous les autres qui partagent quotidiennement le dur apprentissage de la recherche méthodologique soumise aux exigences des applications.

Qu'il me soit permis aussi d'exprimer ma reconnaissance envers J. LEVY, Directeur de l'Ecole des Mines, et le souhait que les moyens de maintenir un équilibre entre théorie et applications iront s'améliorant.

Je remercie enfin J. Altimira et A. Le Gallic qui ont pris en charge le travail de frappe avec beaucoup de soin et de bonne humeur.

RESUME DE LA THESE D'ETAT DE

Jean LEVINE

Sur Quelques Structures d'Information Intervenant en Jeux, dans les Problèmes d'Equipe ou de contrôle et en Filtrage.

Ce mémoire est consacré à l'étude de certains aspects de la prise de décision ou de la commande avec information incomplète sur l'environnement déterministe ou aléatoire.

Dans la 1ère partie, on présente des structures d'information classique déterministes, comprenant la boucle ouverte, la structure de Stackelberg, la boucle fermée et la boucle fermée sur le futur On compare, sur un exemple de duopole dynamique issu de la théorie de la firme, les équilibres en boucle ouverte et fermée Puis on étudie la structure feedforward et on montre, en généralisant la méthode des caractéristiques pour les systèmes d'équations d'Hamilton-Jacobi-Bellman, une condition nécessaire d'existence locale, suggérant qu'il existe une infinité d'équilibres dans certains cas

Dans la 2ème partie, on étudie l'information non classique pour les problèmes d'équipe stochastiques dans le cas de décideurs multiples ayant des observations différentes et une mémoire limitée. On généralise la méthode de programmation dynamique en prenant la loi des trajectoires jusqu'à l'instant présent comme variable d'état. On obtient une équation d'Hamilton-Jacobi-Bellman sous des hypothèses de régularité, donnant une définition rigoureuse de la notion de "signalisation". Ces hypothèses de régularité sont vérifiées dans le cas du contrôle des diffusions avec observations partielles.

La 3ème partie est consacrée à l'étude d'une classe de systèmes nonlinéaires admettant des filtres de dimension finie Les systèmes considérés, à temps discret ou continu, sont caractérisés par le fait que les bruits n'agissent pas sur la dynamique du système, mais seulement sur les observations. On donne la condition nécessaire et suffisante d'existence d'un filtre de dimension finie ainsi que sa réalisation minimale, et on montre le lien entre dimension finie du filtre et immersion dans un système linéaire. Un exemple concret permet d'évaluer les performances de la méthode de filtrage, et de la comparer au filtre de Kalman étendu.

La 4ème partie, enfin, propose un algorithme rapide pour le calcul des commandes réalisant le découplage ou le rejet des perturbations d'un système nonlinéaire (commandes pouvant servir à définir une sous-optimalité raisonnable pour certains problèmes de contrôle stochastique) Cet algorithme nécessite la dérivation formelle (et peut être programmé dans un langage comme REDUCE ou MACSYMA) et utilise l'interprétation des nombres dits "caractéristiques" comme la longueur de chemins minimaux dans le graphe du système. Cette méthode est appliquée au calcul des commandes qui découplent la dynamique d'un bras de robot.

TABLE DES MATIERES

INTRODUCTION

INTRODUCTION

1 - LES STRUCTURES DYNAMIQUES D'INFORMATION

Lorsqu'un décideur veut élaborer rationellement sa stratégie, par exemple dans le cadre de la gestion d'une firme, il doit bien entendu tenir compte des informations qu'il a en sa possession, mais aussi d'un certain nombre d'autres facteurs dont l'oubli risquerait de faire échouer ses projets :

combien y a-t-il d'autres décideurs et quelles sont les informations en leur possession ?

. Savent-ils la nature des informations que possède notre décideur, et celui-ci connaît il la nature des informations de ses coéquipiers ou/et de ses concurrents ?

. Y a-t-il des informations dont on sait qu'elles existent mais que certains décideurs uniquement peuvent connaître sans en trahir le contenu ?

. Est-ce que le fait même de libeller ces questions est connu de tous ?

. Si chaque décideur adopte une stratégie donnée, comment le résultat sera-t-il évalué par chaque décideur et cette évaluation est-elle connue de tous ?

. Le résultat d'une telle stratégie influence-t-il les observations des décideurs et si oui comment ?

On pourrait prolonger cette liste encore longtemps, mais notre but n'est pas de tenter de décourager le décideur potentiel ! Il s'agit simplement de sensibiliser le lecteur non pas au concept vague et multiforme qu'est l'information, mais plus précisément,

à la structure dynamique d'information qui intervient dans la prise de décision ou le guidage automatique.

Notons que dans les questions précédentes ont été employés deux termes assez proches : information et observation.

C'est en fait l'ensemble des observations auxquelles les décideurs ont accès à un instant donné et la manière dont cet ensemble varie avec le temps qui servira de définition informelle à l'expression "structure dynamique d'information" Et comme précisé plus haut, on prêtera plus particulièrement attention à l'influence quantitative de certaines structures d'information sur les décisions qui en résultent ; ce qui n'interdit pas, d'ailleurs, de tirer des conclusions qualitatives.

Ainsi, les cadres théoriques que l'on s'est donné sont les Jeux Dynamiques, les Problèmes d'Equipe, et la Théorie du Contrôle, déterministes ou stochastiques, et l'influence de l'information sur les décisions se traduit ici par la notion de bouclage (voir pour certains cas particuliers [1] et [2]).

Dans tous ces problèmes, l'état est donné par la solution d'une équation d'évolution influencée par une ou plusieurs commandes, et cet état est soit observé complètement, soit observé par l'intermédiaire de mécanismes qui rendent impossible la connaissance exacte de l'état, soit encore pas observé du tout. Dans les deux premiers cas, on peut avoir une mémoire parfaite, ou se souvenir d'une partie seulement du passé des observations, ou encore n'observer qu'une fonction instantanée de l'état et oublier tout ce qui précède l'observation à chaque instant. Mais dans tous les cas, la structure dynamique d'information peut s'exprimer d'une manière simple, comme la donnée d'une famille de σ-algèbres indexées par le temps (voir Beneš [3]), contenant à chaque instant l'ensemble de toutes les observations dont on peut tenir compte dans la loi de commande, et la loi de commande est dite admissible si elle est mesurable par rapport à cette famille de σ-algèbres.

Cette définition contient naturellement les notions classiques de boucle ouverte (pas d'information autre

que le temps) ou de boucle fermée (information complète instantanée mais sans mémoire), et a un sens aussi bien dans des problèmes déterministes que stochastiques.

En contrôle ou dans les problèmes d'équipe, les performances réalisées par l'utilisation des commandes, sont évaluées à l'aide d'une fonction coût qui est une fonctionnelle des trajectoires et des commandes, et que l'on cherche à minimiser dans la classe des lois de commande admissibles.

En jeux, chaque joueur dispose d'une fonction coût et on suppose que les joueurs cherchent à réaliser un type donné d'équilibre (équilibre de Nash par exemple) à l'aide des lois de commande admissibles pour chaque joueur.

Dans tous les cas, on s'attachera à développer des techniques de calcul adaptées aux diverses structures d'information, à mettre en évidence l'influence de ces structures sur les conditions d'optimalité et, lorsqu'on le pourra, sur le minimum ou l'équilibre obtenus.

On peut trouver de nombreux exemples pratiques pour lesquels le formalisme présenté convient parfaitement. Citons, sans les détailler, les problèmes d'oligopole dynamique en économie, d'allocation des tâches dans un ordinateur multiprocesseur ou dans un atelier flexible, d'évaluation des transitoires et des capacités temps réel d'un réseau de communication, et enfin de guidage automatique en général avec observation partielle de l'état (l'un des problèmes les plus fréquents en Ingénierie !)

Cependant, à part de rares exceptions ayant trop souvent un caractère académique, le calcul effectif des stratégies optimales dépasse les possibilités des ordinateurs actuels, ce qui compromet gravement les possiblités d'application en l'état actuel de la théorie. Notamment, en contrôle stochastique avec observations partielles classique (mémoire parfaite), on est amené à calculer la loi de probabilité de l'état conditionnellement aux observations passées (le filtre), voir [4],[5],[6],[7],[8],[9],[10], loi qui dépend des commandes de manière extrêmement compliquée. Le cas

linéaire quadratique gaussien joue ici un rôle singulier puisque le filtre s'y calcule à l'aide d'un nombre fini de paramètres et que seule la moyenne conditionnelle dépend des commandes (principe de séparation de Wonham [11]). On peut donc essayer de trouver des classes de problèmes dont le filtre est de dimension finie, ce qui simplifie notablement la conduite des calculs : cette idée, popularisée par Brockett [12], a fait l'objet de tentatives encore très limitées [13],[14],[15] et nous avons cherché à la développer pour une classe de problèmes à temps discret ou continu.(Partie III.).

Une seconde approche permettant d'espérer des simplifications substantielles, consiste, à l'instar de [16], à renoncer à l'optimalité, pour l'utilisation de lois de commande induisant une structure beaucoup plus simple et même, éventuellement, permettant de se ramener à un problème déterministe. Ainsi, on proposera l'utilisation des techniques de découplage et de rejet de perturbations [17], [18], permettant en particulier de linéariser le système par bouclage, et d'appliquer, après rejet des perturbations, si besoin est, les techniques du linéaire quadratique déterministe !

Avant de passer à une revue de détail sur les points que l'on vient d'aborder, précisons que ce travail est la réunion d'une série d'articles publiés ou à publier, dont le soucis majeur est de développer des techniques de calcul lorsque celles-ci sont parcellaires (1ère partie), ou inexistantes (2ème et 3ème parties), ou déjà connues mais trop lourdes (4ème partie).

Bien entendu, les divers développements proposés n'apportent pas de solutions miracles, et d'importants efforts restent à faire, aussi bien théoriques que pratiques, particulièrement dans la seconde partie, avant de pouvoir s'attaquer à des applications réelles dont la taille est généralement colossale ! Cependant, les deux dernières parties (Filtrage nonlinéaire de dimension finie et découplage des systèmes nonlinéaires) nous semblent, du point de vue des applications, extrêmement prometteuses comme le suggèrent les exemples présentés (conduite de tir et guidage rapide d'un bras de robot, exemples émanants de secteurs industriels dont la demande d'innovation n'est plus à démontrer !).

2 - RESUME DE LA THESE

Ce travail rassemble 10 articles organisés en quatre parties :

I - Généralités sur les structures d'information, étude de quelques structures d'information en jeux différentiels déterministes non coopératifs, application au duopole dynamique.

I.1. Dynamic duopoly theory (en collaboration avec J. Thépot), publié dans l'Encyclopédia of Systems and Control. Pergamon Press. 1983.

I.2. Open-loop and closed-loop equilibria in a dynamical duopoly (en collaboration avec J. Thépot), publié dans "Optimal Control Theory and Economic Analysis, G. Feichtinger Ed., North-Holland, 1982.

I.3. On the solutions of Hamilton-Jacobi systems and applications to the dynamic duopoly. A paraître. 1983.

II - Etude des conditions d'optimalité avec information non Classique pour les problèmes de contrôle et d'équipe stochastiques.

Non classical information and optimality in continuous-time dynamic team problems. A paraître. 1984.

III - Filtrage nonlinéaire de dimension finie pour une classe de systèmes à temps discret et continu.

III 1.a. Exact finite dimensional filters for a class of nonlinear discrete-time systems. (en collaboration avec G Pignié) A paraître. 1983

III 1.b The finite dimensional filtering problem for a class of nonlinear discrete-time systems Proc of the 9th IFAC World Congress Budapest. 1984 (en collaboration avec G. Pignié).

III.2 Une classe de systèmes nonlinéaires à temps continu admettant des filtres de dimension finie. A paraître. 1984.

IV - Méthodes de graphe pour le découplage et le rejet de perturbations des systèmes nonlinéaires.

IV.1. A fast graph theoretic algorithm for the feedback decoupling problem of nonlinear systems. (en collaboration avec A. Kasinski). in Mathematical Theory of Networks and Systems. P.A. Fuhrmann. ed. Lecture Notes in Control and Information Sciences, N°58, pp. 550-562. (1983). Springer.

IV.2. A fast algorithm for systems decoupling using formal calculus (en collaboration avec F. Geromel et P. Willis). In Analysis and Optimization of Systems. A. Bensoussan, J.L. Lions ed. Lecture Notes in Control and Information Sciences, N°63, Part.2, pp. 378-390, (1983). Springer.

ANNEXE : Un aperçu élémentaire de la théorie moderme des systèmes nonlinéaires, publié dans la RAIRO - Automatique. (Dec. 83).

Partie 1 :

Dans la première partie, on donne une présentation informelle de différentes structures d'information dans le cadre de la théorie des jeux dynamiques non coopératifs à 2 joueurs (duopole dynamique). Les structures d'information sont classées en deux séries : "information complète" (qui est d'ailleurs un choix malheureux puisqu'on n'y connaît pas nécessairement tout ! mais qui veut simplement dire qu'une structure probabiliste n'est pas nécessaire) et "informatique incomplète".

L'information complète regroupe la boucle ouverte, la boucle fermée, les structures du type Stackelberg (disymétrique) et enfin la "boucle fermée sur le futur".Les techniques hamiltoniennes de calcul des stratégies optimales sont présentées Aucune structure ne donne le même résultat en général.

Cette affirmation est étayée par les deux papiers complémentaires I.2 et I.3 de ce chapitre où l'on montre (I.2) que la notion de boucle fermée n'implique pas, malgré la présence d'information complète instantanée, une concurrence plus exacerbée : au contraire, dans le cas de firmes se partageant le marché par le contrôle des prix, pour un marché de biens substituables avec une demande à élasticité constante, la boucle fermée induit une certaine coopération parce que chaque firme sait que les 2 ont intérêt à saturer les contraintes,ce qui limite les choix stratégiques au lieu de créer des menaces supplémentaires, et produit en définitive un consensus pour avoir des prix plus élevés qu'en boucle ouverte. Le second papier (I.3) présente une structure d'information originale que l'on rencontre naturellement dans le cas général de la résolution des conditions d'optimalité : l'équilibre de Nash des Hamiltoniens donne les stratégies optimales comme des fonctions de l'état, mais aussi des variables adjointes (donc contenant des informations sur le futur). Or, on montre par le calcul du système caractéristique que ces stratégies donnent lieu en général à une infinité d'équilibres possibles en tout point régulier générique. On termine en donnant un exemple linéaire quadratique où aucun des équilibres en boucle ouverte, fermée ou fermée sur le futur ne coïncide. On peut certainement en conclure que l'équilibre de Nash n'est pas une notion d'équilibre suffisamment précise pour être vraiment pertinente...

Dans la seconde série de structures d'information incomplète, on présente les structures de boucle fermée sur les observations et de boucle fermée sur la loi de probabilité de l'état. Nous reviendrons sur ces structures dans la seconde partie.

Partie II :

Cette partie est entièrement consacrée à l'information non classique, à savoir lorsque les σ-algèbres d'observation ne sont pas croissantes en fonction du temps On peut donner 2 exemples simples de structures d'information où cela a lieu : lorsque le contrôleur (décideur dans un problème de contrôle) oublie une partie du passé des observations, ou, lorsqu'il y a plusieurs joueurs, si chaque joueur a des informations différentes sur l'état du système et n'a pas accès aux informations des autres On voit que cette dernière structure est générale en théorie des jeux ou dans les problèmes d'équipe. On montre que l'on peut utiliser la méthode de programmation dynamique à condition de "grossir" l'espace d'état : au lieu de l'état du système de départ, il faut utiliser sa loi de probabilité non conditionnelle comme nouvelle variable d'état. Dans ce cas, la programmation dynamique donne la ou les stratégies optimales en fonction des observations et de la loi, ce qui oblige à considérer une structure d'information plus générale où le bouclage des commandes sur la loi est permis, et que l'on a appelée "boucle fermée". On montre alors que l'optimum en boucle fermée est égal au précédent, puis on dérive les conditions d'optimalité.

Cette étude est menée dans deux cas : lorsque les bruits sont à temps discret ou dans le cas des diffusions
On montre dans ces deux situations que la fonction valeur est concave et continue par rapport à la loi des trajectoires, et donc sur-différentiable, et, moyennant une condition de régularité sur le sur-différentiel, on peut obtenir une équation du type Hamilton-Jacobi-Bellman caractérisant la fonction valeur et la ou les stratégies optimales. L'Hamiltonien associé à cette équation comporte alors un terme supplémentaire par rapport à celui du contrôle à information complète, terme que l'on peut interpréter comme la variation du coût correspondant à une variation d'information; on donne ainsi une définition précise de la notion de "signalling", introduite heuristiquement dans [19] et [20], disant qu'à l'optimum la commande devait réaliser le meilleur compromis entre minimiser le coût et coder des informations dont la connaissance pourrait améliorer les décisions futures

Dans le cas particulier du contrôle des diffusions avec observations partielles et information classique, on montre en plus que l'équation d'Hamilton-Jacobi-Bellman peut être obtenue sans hypothèse de régularité sur la fonction valeur, donnant ainsi une condition nécesssaire et suffisante d'optimalité, généralisant les conditions nécessaires obtenues par A. Bensoussant [5]

Partie III :

Comme précédemment annoncé, c'est la pénurie de techniques de calcul efficaces en contrôle stochastique, même à information classique (à l'exception du cas linéaire-quadratique gaussien) qui montre l'importance soit des techniques de filtrage approché, soit de filtrage exact mais de dimension finie.

C'est le problème du filtrage exact de dimension finie qui est abordé ici pour une classe de systèmes nonlinéaires à temps discret ou continu , ne comportant pas de bruits de dynamique.

Du point de vue des applications, une telle modélisation peut se justifier au moins dans les deux cas suivants :

- la durée de vie ou d'observation du processus est très courte.
- les bruits de dynamique n'agissent que sur les composantes "lentes" du processus. On peut ainsi filtrer sur un court intervalle de temps la dynamique rapide non bruitée (situation précédente), puis réactualiser la loi en fonction de la dérive du processus lent et recommencer.

Les deux premiers papiers sont consacrés au temps discret, le premier exposant la théorie et le second comparant différentes méthodes de filtrage dans le cadre d'une application, et le troisisème est consacré au temps continu Les systèmes à temps discret étudiés ici sont plus généraux que ceux à temps continu puisque, pour les premiers, l'intensité des bruits d'observation peut être correlée à l'état (bruits colorés).

Dans le premier papier, on commance par prouver une formule récursive donnant la loi conditionnelle non normalisée, puis on montre qu'une orientation naturelle consiste à généraliser à la dimension infinie les techniques de réalisation des systèmes nonlinéaires à temps discret.
On montre, dans le cas des bruits gaussiens, que l'on peut construire explicitement une base canonique du filtre qui donne lieu à une condition nécessaire et suffisante d'existence de filtre de dimension finie.Cette condition est particulièrement simple et accessible au calcul, et permet de décrire explicitement la réalisation minimale du filtre dont la dimension est égale à la dimension de l'espace engendré par la base canonique. Bien entendu, on vérifie que cette réalisation minimale est bien localement faiblement observable et localement faiblement accessible, au sens de la théorie des systèmes nonlinéaires. De plus, on montre qu'un système nonlinéaire admettant un filtre de dimension finie peut être transformé en un système linéaire si et seulement si l'intensité des bruits n'est pas correlée à l'état. Enfin, on tente d'évaluer le nombre des systèmes admettant un filtre de dimension minimale donnée r, et pour une équation d'observation donnée. On montre que,
sous certaines hypothèses de régularité sur la base canonique, on peut effectivement construire au moins autant de systèmes satisfaisant aux conditions ci-dessus que d'éléments d'un sous-groupe du groupe linéaire de dimension r. En plus d'exemples académiques, on présente une application réelle à un problème de conduite de tir, donnant des résultats probants, alors qu'aucune méthode linéaire ou approchée ne donne de bons résultats. Ce point est particulièrement développé dans le second papier où l'on montre, toujours pour le problème de conduite de tir, que le filtre de Kalman étendu diverge presque systématiquement, que le filtre de Kalman sur un système linéaire obtenu en dérivant deux fois le système de départ est complètement inefficace puisque l'état n'y est plus observable, alors que le filtre nonlinéaire obtenu par les techniques précédentes donne, pour une erreur initiale de l'ordre de 40 %, une estimée en moins de 15 observations (2 secondes réelles) dont l'erreur est inférieure à 5 %. Notons enfin que pour des temps de calcul aussi courts l'utilisation du filtre nonlinéaire général (de dimension infinie) était rigoureusement impossible.

Dans le troisième papier, on montre que la plupart des résultats précédents se généralisent au temps continu. Ainsi, après avoir calculé explicitement la solution de l'équation de Zakaï pour le cas de la dynamique non bruitée, on fait apparaître comme précédemment la base canonique du filtre donnant ainsi la condition nécessaire et suffisante d'existence d'un filtre de dimension finie, ainsi que la réalisation minimale du filtre. La condition obtenue est équivalente à la dimension finie de l'algèbre de Lie associée à l'équation, de Zakaï, généralisant ainsi des résultats heuristiques [21] obtenus précédemment dans le cas où l'algèbre de Lie est nilpotente. On donne enfin un exemple d'observation polynômiale de degré quelconque d'un système linéaire non bruité où le filtre est toujours de dimension finie, alors que lorsque la dynamique est bruitée et l'observation cubique, il n'y a pas de filtre de dimension finie (voir [14]).

Partie IV :

La motivation de cette dernière partie, qui n'est pas donnée dans les papiers présentés, comportant en soi un intérêt plus général, peut être vue comme le développement de méthodes permettant de transformer un problème stochastique nonlinéaire en un problème éventuellement découplé et linéaire, mais surtout déterministe (rejet des perturbations). La classe naturelle des lois de commande assurant une telle propriété est donc la classe dans laquelle on peut chercher la "sous-optimalité".

Le premier papier IV.1, après avoir rappelé les conditions nécessaires et suffisantes de rejet de perturbation et de découplage, prouve que le calcul des lois de commande assurant le rejet de perturbations et le découplage peut être très largement simplifié à l'aide de l'interprétation, en terme de graphe, des nombres caractéristiques. Ces nombres s'interprètent comme le nombre minimal d'intégrations qu'il faut à une commande pour être "visible" dans une sortie donnée. On donne l'algorithme de calcul, utilisant des méthodes de calcul formel (Reduce ou Macsyma).

Le second papier IV.2 donne un résumé du papier IV 1 et montre comment est organisé le programme de calcul formel. L'intérêt de la méthode de graphe est chiffré sur l'exemple du découplage de la dynamique d'un bras de robot. Cet exemple montre le gain que l'on retire des méthodes de calcul formel, sans lesquelles le découplage de tels systèmes nécessiteraient des efforts extrêment lourds.

Annexes :

On donne un exposé élémentaire des résultats les plus modernes en théorie des systèmes nonlinéaires qui pourra servir à éclaircir un certain nombre de définitions et propriétés utilisées dans les deux derniers chapitres.

Conclusion :

Ce travail comportant essentiellement des méthodes de calcul, il est clair qu'un travail de comparaison et d'approfondissement sur chaque structure d'information est nécessaire. Ce travail semble cependant très difficile dans le cas de l'information non classique où de gros efforts théoriques restent à faire, surtout concernant les méthodes numériques.

D'autre part, la généralisation des méthodes développées en filtrage, au cas comportant des bruits de dynamique semble être une question très importante aussi bien théoriquement que pour les applications.

Finalement, il serait intéressant de savoir s'il est possible de trouver des algorithmes performants pour le découplage et le rejet de perturbations par retour de sortie puisqu'ici les méthodes proposées nécessitent la connaissance exacte de l'état.

Références de l'Introduction

[1] P. BERNHARD, G. COHEN, J-P QUADRAT : Le feedback en théorie de la commande. Quelques remarques. A paraître

[2] Y.C. HO, I. BLAU, T. BASAR : A tale of four information structures A paraître.

[3] V. BENES : Existence of optimal strategies based on specified information, SIAM J. Cont. Vol.8, 2 p.179-188 (1970).

[4] R. ANDERSON, A. FRIEDMAN : Multi-dimensional quality control. Parts I and II. TAMS, Vol.246, p.31-94 (1978).

[5] A. BENSOUSSAN : Maximum principle and dynamic programming approaches of the optimal control of partially observed diffusions. Stochastics, 9,3, (1983), p169-222.

[6] J.M BISMUT : Sur un problème de contrôle stochastique avec observation partielle. Z.f.W, 49, p.63-95 (1979).

[7] M.H.A. DAVIS : Nonlinear semigroups in the control of partially observed stochastic systems. Lecture Notes in Math. (1979).

[8] W.H. FLEMING : Nonlinear semigroup for controlled partially observed diffusions. To appear.

[9] W.H. FLEMING, E. PARDOUX : Existence of optimal controls for partially observed diffusions. SIAM J. Cont Vol 20 p.261-288 (1982)

[10] R.E. MORTENSEN : Stochastic optimal control with noisy observations. Int. J. Cont. 4, p.455-465 (1966).

[11] W.M WONHAM : On the separation theorem of stochastic control SIAM J. Cont Vol.6, N°2, (1968)

[12] R BROCKETT : Remarks on finite dimensional estimation Asterisque 75, 76 (1980)

[13] V BENES : Exact finite dimensional filters for certain diffusions with nonlinear drift. Stochastics 5, p.65-92 (1981).

[14] M HAZEWINKEL, S.I. MARCUS, H.J. SUSSMAN : Non existence of exact finite dimensional filters for the cubic sensor. Preprint. Université Erasmus. Amsterdam.

[15] M. CHALEYAT-MAUREL, D. MICHEL : Un théorème de non-existence de filtre de dimension finie. CRAS, t 296 (1983). Serie I. 933-936.

[16] J.P. QUADRAT : Thèse Paris 9. 1981.

[17] A. ISIDORI, A. KRENER, C. GORI-GIORGI, S. MONACO : Nonlinear decoupling via feedback. IEEE. Trans. AC. 26, 2, p.331-345 (1981).

[18] D. CLAUDE : Decoupling of nonlinear systems. Syst. Cont. Letters. 1, 4 (1982).

[19] H.S. WITSENHAUSEN : A counterexample in stochastic optimum control. SIAM J. Cont. 6,1, (1968), p. 131-147.

[20] Y.C. HO, M. KASTNER, E. WONG : Teams, market signalling, and information theory. IEEE-AC, 68,6, (1980), p. 644-654.

[21] Z.S. ROTH, K.A. LOPARO : Optimal filter realization for a class of nonlinear systems with finite dimensional estimation algebra. Syst. Cont. Letters, 4,1, (1984), p.23-26

PARTIE I

Généralités sur les structures d'information.

Etude de quelques structures d'information en jeux différentiels déterministes non coopératifs,

application au duopole dynamique

RESUME DE LA Ière PARTIE

Généralités sur les structures d'information. Duopole dynamique

Cette partie sert à introduire les diverses structures d'Information qui ont été étudiées jusqu'à présent dans le cadre des jeux dynamiques Le premier article, publié dans l'Encyclopedia of Systems and Control, en collaboration avec J. Thépot, sert en quelque sorte de fil directeur pour les 2 premières parties: on y présente, sans les démontrer, les principaux résultats sur les conditions d'équilibre pour chaque structure d'information, et les résultats originaux sont développés et démontrés dans les autres articles des parties I et II.

Les structures d'information présentées sont classées en deux groupes: l'information de nature déterministe et l'information de nature probabiliste.

a) Dans le premier groupe, on trouve la **boucle ouverte** (la seule information pour les 2 joueurs est le temps et le point de départ du jeu), la **boucle fermée** (les joueurs ont une information complète sur l'état du jeu mais purement instantanée), les **structures disymétriques du type Stackelberg** où l'un des 2 joueurs est le meneur et l'autre le suiveur. Le meneur joue en boucle ouverte alors que le suiveur joue en boucle fermée et connaissant la stratégie du meneur; et enfin, la **boucle fermée sur le futur** (information complète instantanée des 2 joueurs et en plus, observation exacte de leur revenu marginal).

Deux contributions originales y sont annoncées, et développées dans les deux articles qui suivent. Il s'agit d'une part de la comparaison entre équilibres en boucle ouverte et en boucle fermée, dans le cas d'un duopole où 2 firmes se partagent le marché par le contrôle des prix et de l'investissement, les biens produits par les 2 firmes étant substituables, et la demande

étant supposée à élasticité constante. On montre que, contrairement à ce que l'on attend, la boucle fermée induit une certaine coopération entre les firmes car chaque joueur sait que chacun a intérêt à saturer les contraintes sur les investissements, ce qui limite leurs choix stratégiques et produit, en définitive, des prix plus élevés qu'en boucle ouverte en régime permament.

Il s'agit d'autre part du calcul des équilibres en boucle fermée sur le futur. On montre d'abord que c'est cette structure d'information qui apparaît naturellement lorsque l'on cherche un équilibre de Nash des Hamiltoniens, puisqu'alors on obtient les stratégies optimales comme des fonctions du temps, de l'état, et des variables adjointes (revenus marginaux des 2 joueurs), et lorsque les stratégies optimales ne saturent pas les contraintes, on ne peut éliminer les variables adjointes. On montre alors, en généralisant à ce cas la théorie des caractéristiques de Cauchy de l'équation d'Hamilton-Jacobi, qu'il existe en tout point générique une infinité d'équilibres possibles. Enfin, on donne un exemple élémentaire où aucun des équilibres (boucle ouverte, fermée et fermée sur le futur) ne coïncide.

b) Dans le deuxième groupe, on présente des structures d'information incomplète: boucle fermée sur les observations instantanées où, d'une part, les joueurs observant l'état par des procédés différents, ils ne peuvent comparer leurs informations, et, d'autre part, les observations étant instantanées et sans mémoire, ils ne peuvent utiliser ce qu'ils auraient pu apprendre dans le passé. Ce type d'information **non classique** ne vérifie pas les conditions "habituelles" sur les σ-algèbres d'observation que l'on suppose en contrôle avec observation partielle. On présente alors une équation de programmation dynamique qui sera largement développée dans la partie II, consacrée exclusivement à l'étude de l'information non classique.

DYNAMIC DUOPOLY THEORY

J. LEVINE * J. THEPOT **

Since the prominent contribution of von Neumann and Morgenstern (1944), oligopoly theory is widely recognized as part of Game Theory. Static formulations of the oligopoly game have been developed to explain how the competitive interdependencies determine the price, quantity or advising decisions of the firms. However, it is clear that Time plays a determinant part in the definition of the strategies of the competitors. Differential Games techniques have therefore been used to extend the static traditional models to dynamic situations. By emphasizing here duopoly situations, we are going to outline the main issues arising in this Theory and to present illustrative and recent models.

I - General Statement and Informational Structures of a Dynamic Duopoly

Let us consider two firms (firm 1 and firm 2) competing on the same market over a horizon $[0,T]$. At time t, the state of firm i is represented by a vector $x_i(t)$ of $\mathbb{R}^{n_i}$ (ex : production capacity, inventory levels, balance-sheet accounts, etc...), and its decision by a vector function of its observations (to be defined later) with values $u_i(t)$ in $\mathbb{R}^{p_i}$. The result $u_i(t)$ of firm i's decision at time t knowing its observations is called a control (ex : price, quantity to be sold by unit of time, etc...), and a sequence of decisions over the horizon is called a strategy.

At any time, each firm has to take its decision according to given constraints :

$$\phi_j(t,x(t),u_1(t),u_2(t)) \leq 0 \ , \ j = 1,\ldots,m \qquad (1)$$

where $x(t) = (x_1(t),x_2(t))' \in \mathbb{R}^n$ (prime denoting transpose),

---

*Centre d'Automatique et d'Informatique de l'Ecole Nationale Supérieure des Mi...

denotes the state of the duopoly. If firm i must satisfy the set of constraints $\phi^i = \{\phi_{j_o}, \ldots, \phi_{j_i}\}$ independently of its opponent's decision, we say that this set of constraints is under firm i's responsibility.

The dynamics of the duopoly are described by the following differential system :

$$\dot{x}(t) = f(t,x(t), u_1(t), u_2(t)) \qquad (2)$$

in which the initial state $x(0) = \xi$ is given.

During the interval $[t,t+dt]$, firm i's profit is defined by $g_i(t,x,u_1,u_2)dt$, so that the net present value J_i over the horizon (with a discount rate α_i) can be written as :

$$J_i(u_1,u_2) = \int_o^T g_i(t,x(t),u_1(t),u_2(t))e^{-\alpha_i t}dt + M_i(x(T)), \quad i = 1,2, \qquad (3)$$

where M_i describes the evaluation of firm i at time T.

Before to discuss the various structures of information that can be met in these game problems, we suppose that an information structure S is given, and that $U_i(S)$ is firm i's set of strategies adapted to S, and satisfying the constraints under i's responsibility, i = 1,2. Thus, we assume that the two competitors try to realize a non-cooperative Nash equilibrium; namely, if S is a complete information structure (see below), they want to find a pair of strategies (u_1^*, u_2^*) in $U_1(S) \times U_2(S)$ such that :

$$\begin{aligned} &J_1(u_1^*,u_2^*) \leq J_1(u_1,u_2^*) \quad \forall u_1 \in U_1(S) \\ &J_2(u_1^*,u_2^*) \leq J_2(u_1^*,u_2) \quad \forall u_2 \in U_2(S) \end{aligned} \qquad (4)$$

If $\underset{\sim}{S}$ is an incomplete information structure, (4) must be adapted in replacing J_i by $E(J_i)$ the mathematical expectation of J_i, i = 1,2.

Finally, let us present a brief survey of the informational structures that have been studied or, at least, pointed out, in the literature until now : they are classified into complete and incomplete information structures.

1. Complete information.

In all this paragraph, both firms are supposed to have at least a perfect knowledge of the set of data :

$$\{\{\phi_j\}, f, J_1, J_2, \xi, t, T\}. \tag{5}$$

For all the structures introduced below, simple counterexamples prove that they yield different solutions to the Nash game.

1.1. Open-loop structure : Both firms have only the knowledge of (5). This structure is called "static" since there is no change of information during the game. The strategies of $U_i(S)$ are thus measurable functions of t and ξ, and, when ξ is fixed, reduce to controls. This class of games is by far the most studied and the reader will find a complete bibliography in (Feichtinger, Jorgensen 1983).

1.2. Feedback structure : Both firms observe exactly the sta state x at any time t. $U_i(S)$ is thus made of measurable functions $u_i(t,x)$. A careful definition of the solution of (2) must be provided in order to allow strategies that are discontinuous with respect to x.

As in (Basar,Olsder 1982); we distinguish between "Feedback" and "Closed-Loop" structures, where the initial condition ξ is also remembered. Thus strategies take the form $u_i(t,x,\xi)$. When, furthermore, the competitors perfectly remember the past of the state, we say that we are in a "Full memory structure". Whether these three structures coincide or not is not clear at all.

1.3. Stackelberg structure : The leader, say firm 1, plays open-loop and gives its control at every time to the follower which, in addition, perfectly observes the state. Thus $U_1(S)$ is made of controls $u_1(t)$; whereas $U_2(S)$ is made of measurable functions of the form $u_2(t,x,u_1(t))$. Details can be found in (Basar 1977).

1.4. Feedforward structure : Each firm takes decisions of the form $u_i(t,x,p,q)$ where p (resp. q) is the optimal marginal revenue for firm 1 (resp 2) of the game starting at (t,x) over the horizon $[t,T]$. This information structure is naturally adapted to Dynamic Programming methods (Levine 1983). As a result, the structures 1.3. and 1.4. coincide in the zero-sum situation.

2. Incomplete Information.

This is the case where at least one firm does not observe perfectly the state because of disturbances and/or of the non injectivity of the observation function. Precisely, suppose that the observations equations are given by :

$$y_i(t) = h_i(x(t),\nu_i(t)) \quad , \ i = 1,2 \qquad (6)$$

where ν_1 and ν_2 are exogeneous disturbances.
Conceptually, nothing would change if, in place of (6), the observations were described by a stochastic differential system. Following (Harsanyi 1968), the firms must agree on an a priori probability measure on the initial state ξ and on the disturbances Let $P(\xi,\nu_1,\nu_2)$ be this a priori probability measure. Then J_1 and J_2 must be replaced by their mathematical expectation with respect to P, namely :

$$\tilde{J}_i(u_1,u_2) = \int J_i(u_1,u_2)dP \quad , \ i=1,2. \qquad (7)$$

We shall assume that the constraints (1) are of the form :

$$\phi^i(t,y_i,u_1,u_2) \quad , \ i = 1,2.$$

2.1. Output Feedback Structure : Each firm perfectly knows the set of data : $\{\{\phi_j^i\}, f, h_1, h_2, \tilde{J}_1, \tilde{J}_2, P, t, T\}$, and observes y_i at every time t (and possibly all or part of the past y_i). Decisions take the form $u_i(t, y_i)$ or $u_i(t, \{y_i(s) \mid s \leq t\})$.

2.2. Closed-Loop Structure : In addition to the preceding one, the decisions take into account the actual probability measure P_t, image of P by (2), which plays the role of the state of the game with incomplete information. Thus firm i's decision is of the form $u_i(t, y_i, P_t)$. For details see (Levine 1981).2.1. and 2.2. are referred to as non-classical information structures (Witsenhausen 1968) since firms 1 and 2 have different observation and the associated sigma-fields are not included one in another.

II - Characterizations of Nash Equilibria.

We shall review the existence results and the characterizations of the solutions for the preceding information structures. We shall use the same numbering as in paragraph I.

1.1. Open-Loop structure : Existence results of an open-loop Nash solution can be proved for linear-quadratic games (Starr, Ho 1969). For the characterization of open-loop solutions, it can be proved that a two-sided minimum principle holds (Starr, Ho 1969):

Theorem 1 : Let f, g_1, g_2, M_1, M_2 be C^2 functions and ϕ_j^i depend only on (u_1, u_2), $\forall$ i,j. Then a necessary condition for (u_1^*, u_2^*) to be an open-loop Nash solution is that there exist two continuous functions p_1 and p_2 satisfying :

$$\dot{x} = f(t, x, u_1^*, u_2^*) \qquad x(0) = \xi$$

$$\dot{p}_i = -\frac{\partial H_i}{\partial x} + \alpha_i p_i \qquad p_i(T) = \frac{\partial M_i}{\partial x}(x(T)) \qquad i = 1,2.$$

with $H_1 = p_1.f(t,x,u_1^*,u_2^*) + g_1(t,x,u_1^*,u_2^*)$

$\leq p_1.f(t,x,u_1,u_2^*) + g_1(t,x,u_1,u_2^*) \quad \forall\, u_1 \text{ s.t. } \phi^1(u_1,u_2^*) \leq 0$

and $H_2 = p_2.f(t,x,u_1^*,u_2^*) + g_2(t,x,u_1^*,u_2^*)$

$\leq p_2.f(t,x,u_1^*,u_2) + g_2(t,x,u_1^*,u_2) \quad \forall\, u_2 \text{ s.t. } \phi^2(u_1^*,u_2) \leq 0.$

1.2. Feedback structure : Existence results over a small horizon can be derived for linear-quadratic games (Lukes 1971), (Bensoussan 1974). Also characterizations can be obtained under regularity assumptions on the optimal value functions, by means of the Dynamic Programming method, and under the assumption that the "local" Nash equilibrium of the Hamiltonians at every point can be obtained as functions of (t,x). Namely (Case 1969) :

Theorem 2 : f, g_1, g_2, M_1, M_2 are chosen as in theorem 1. Let :

$$e^{-\alpha_i t}V_i(t,x) = \int_t^T g_i(s,x(s),u_1^*,u_2^*)e^{-\alpha_i s}ds + M_i(x(T))$$

$$\stackrel{\text{def}}{=} J_i(t,x,u_1^*,u_2^*) \quad , \; i = 1,2, \quad \text{where } (u_1^*,u_2^*) \text{ are}$$

supposed to realize a Feedback Nash equilibrium over the horizon $[t,T]$, from the initial point x. Suppose furthermore that V_1 and V_2 are piecewise continuously differentiable. Then V_1, V_2, u_1^* and u_2^* must solve the following system of Hamilton-Jacobi equations at every regular point :

$$\frac{\partial V_1}{\partial t} - \alpha_1 V_1 + \min_{\phi^1(t,x,u_1,u_2^*)\leq 0} \left(\frac{\partial V_1}{\partial x}.f(t,x,u_1,u_2^*) + g_1(t,x,u_1,u_2^*)\right) = 0$$

$$\frac{\partial V_2}{\partial t} - \alpha_2 V_2 + \min_{\phi^2(t,x,u_1^*,u_2)\leq 0} \left(\frac{\partial V_2}{\partial x}.f(t,x,u_1^*,u_2) + g_2(t,x,u_1^*,u_2)\right) = 0 \quad \blacksquare \tag{8}$$

Corollary (Case 1969) : Under the same assumption and if (u_1^*,u_2^*) obtained by (8) are of the form $u_1^*(t,x), u_2^*(t,x)$, then

$p_1 = \frac{\partial V_1}{\partial x}$ and $p_2 = \frac{\partial V_2}{\partial x}$ solve, at every regular point, the adjoint equations :

$$\dot{p}_i = -\frac{\partial \tilde{H}_i}{\partial x} + \alpha_i p_i \quad , \quad p_i(T) = \frac{\partial M_i}{\partial x}(x(T)) \quad , \quad i = 1,2 \qquad (9)$$

with $\tilde{H}_i = p_i . f(t,x,u_1^*(t,x),u_2^*(t,x)) + g_i(t,x,u_1^*(t,x),u_2^*(t,x))$. ■

Remark : in (9) appear the derivatives of u_i^*, $i = 1,2$, with respect to x, so that its solution is generally different from the open-loop adjoints.

Let us also point out that the optimization problem of (8) determines u_1^* and u_2^* as functions of (t,x,p_1,p_2). Thus, a method to obtain u_i^* as functions of (t,x) consists in making the change of variables :

$$p_i = P_i(t,x) \quad , \quad i = 1,2. \qquad (10)$$

Thus P_i must satisfy the system :

$$\frac{\partial P_i^j}{\partial t} + \frac{\partial P_i^j}{\partial x} . f^* = - P_i . \frac{\partial f^*}{\partial x_j} - \frac{\partial g_i}{\partial x_j} - \alpha_i P_i^j - (P_i . \frac{\partial f^*}{\partial u} + \frac{\partial g_i^*}{\partial u}) \frac{\partial u^*}{\partial p} \frac{\partial P}{\partial x_j}$$

$$i = 1,2; \; j,k = 1,\ldots,n. \qquad (11)$$

$$\frac{\partial P_i^j}{\partial x_k} = \frac{\partial P_i^k}{\partial x_j}$$

where f^* denotes f evaluated at $u_i^*(t,x,P_1(t,x),P_2(t,x))$, and the same for g_i^*.

For linear f and quadratic g_i, and if we look for u_1^* and u_2^* as linear feedback functions of x, (11) becomes the well known system of two coupled Riccati equations. Nevertheless, there is no proof of the fact that, in the linear quadratic case, the linear solution of (11) is unique, and the author conjectures the contrary.

On the other hand, one can find verification theorems in (Stalford, Leitmenn 1973), (Mehlmann 1982), but an open problem remains the derivation of necessary conditions on singular surfaces

1.3. Stackelberg structure : Since the leader plays open-loop and the follower closed-loop knowing the leader's control,

the characterization of the Stackelberg equilibrium can be obtained by crossing the two preceding methods. It can be seen in (Basar 1977) that there exist infinitely many equilibria even in the simplest linear-quadratic case with strictly convex cost functions. This result illustrates the sensitivity of the Nash equilibrium to the information structures.

1.4. Feedforward : It was seen in 1.2. that one generally obtains the optimal strategies in (8) in the form :

$u_1^*(t,x,p_1,p_2)$, $u_2^*(t,x,p_1,p_2)$.

Thus, since the information structure allows the competitors to use their optimal strategies as such, without introducing the a priori change of variables (10), it remains to find the adjoint system for p_1,p_2 in order to compute the optimal trajectories. Thus, if we note $f^*(t,x,p_1,p_2)=f(t,x,u_1^*(t,x,p_1,p_2),u_2^*(t,x,p_1,p_2))$ and the same for g_1^*,g_2^*, and if we set :

$$\bar{H}_i = p_i f^*(t,x,p_1,p_2) + g_i^*(t,x,p_1,p_2) \quad , \quad i = 1,2,$$

$$\frac{\partial V_i}{\partial t} = \pi_i \qquad i = 1,2,$$

the following theorem (Levine 1983) holds true :

Theorem 3 : in the feedforward structure the adjoint equations are given, in addition to $\dot{x} = f^*(t,x,p_1,p_2)$, by :

$$\dot{p}_i^j = \sum_{k=1}^{n} a_{jk}^i(t,x,p_1,p_2) f_k^*(t,x,p_1,p_2) \quad , \quad a_{jk}^i = a_{kj}^i$$

$$i = 1,2; \; j,k = 1,\ldots,n,$$

$$\sum_{j=1}^{2} \sum_{k=1}^{n} \frac{\partial \bar{H}_i}{\partial p_k^j}(t,x,p_1,p_2) a_{kl}^j(t,x,p_1,p_2) = \alpha_i p_l^i - \frac{\partial \bar{H}_i}{\partial x_l}(t,x,p_1,p_2) \qquad (12)$$

$$i = 1,2; \; l = 1,\ldots,n,$$

with terminal conditions : $p_i(T) = \frac{\partial M_i}{\partial x}(x(T))$.

Note that , in (12), we must determine the n(n+1) functions

a^i_{jk} with 2n equations, and suitable transversality conditions. This suggests that non-uniqueness of Nash equilibria is a generic property. ∎

Remark : The non-uniqueness of (12) disappears when u_1^* and u_2^* are independent of p_1, p_2, in which case the informations on the future contained in p_1, p_2 are worthless, and the adjoint system reduces to (9). However, it can be proved that the solutions obtained by (12) are generally different from the open and closed-loop solutions.

2. Incomplete Information. Feedback structure : We shall just sketch the dynamic programming methods, for example when the observation equations are given by (6), with ν_i a piecewise constant process on prescribed intervals $[t_j, t_{j+1}[$ forming a partition of $[0,T]$. We note ν_i^j the projection of ν_i on the interval $[t_j, t_{j+1}[$ and we suppose that ν_i^j is independent of x and of ν_i^k, $k \neq j$, and we note $\rho(\nu)$ the probability measure of (ν_1, ν_2). Let us denote :

$$e^{-\alpha_i t} V_i(t,P_i) = \int\int \left(\int_t^T g_i(s, X_s^*(t,x), u_1^*, u_2^*) e^{-\alpha_i s} ds + M_i(X_T^*(t,x))\right) dP_t(x) d\rho(\nu) \quad , \quad i=1,2 \qquad (13)$$

where u_1^*, u_2^* are supposed to be a Nash pair in the Feedback structure (precisely, for every $t, y_i(t)$ and P_t, they are given by $u_1^*(t, y_1(t), P_t)$, $u_2^*(t, y_2(t), P_t)$), and where $X_s(t,x)$ is the solution of (2) at time s starting from (t,x) and generated by u_1^*, u_2^*. Finally, let us recall that the Lie derivative of P_t in the direction of u_1, u_2 is the limit when it exists :

$$L_{u_1,u_2}(P_t) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}(X_\varepsilon^{u_1,u_2}(t,P_t) - P_t) \qquad (14)$$

where $X_\varepsilon^{u_1,u_2}(t,P_t)$ is the image of P_t by the flow of trajectories solutions of :

$$\frac{d}{ds} X_s^{u_1,u_2}(t,x) = f(s, X_s^{u_1,u_2}(t,x), u_1, u_2)$$

with $X_t^{u_1,u_2}(t,x) = x$.

The following results hold true (Levine 1981) :

Proposition : V_i has the integral representation :

$$V_i(t,P_t) = \int w_i(t,x;t,P_t)dP_t(x) \quad , \quad i = 1,2,$$

with $w_i(t,x;s,P_s) = w_i(t,x;t,X_t^*(s,P_s)) \quad \forall\ (t,s,P_s)$, a.e.x.

Theorem 4 : If (u_1^*,u_2^*) is a Nash point and if w_1,w_2 are C^1 functions of all their arguments, then we have :

$$\begin{aligned}
&\int\int(\frac{\partial w_1}{\partial t}-\alpha_1 w_1)dP_t d\rho(\nu)+\int\{\min_{u_1}\int[\frac{\partial w_1}{\partial x}.f(t,x,u_1,u_2^*)+g_1(t,x,u_1,u_2^*)+\\
&\quad+<\frac{\partial w_1}{\partial P},L_{u_1,u_2^*}(P_t)-L_{u_1^*,u_2^*}(P_t)>]\,d(P_t\otimes\rho)(x,\nu|y_1)\}dQ_t^1(y_1) = 0\\
&\int\int(\frac{\partial w_2}{\partial t}-\alpha_2 w_2)dP_t d\rho(\nu)+\int\{\min_{u_2}\int[\frac{\partial w_2}{\partial x}.f(t,x,u_1^*,u_2)+g_2(t,x,u_1^*,u_2)+\\
&\quad+<\frac{\partial w_2}{\partial P},L_{u_1^*,u_2}(P_t)-L_{u_1^*,u_2^*}(P_t)>]\,d(P_t\otimes\rho)(x,\nu|y_2)\}dQ_t^2(y_2) = 0
\end{aligned} \tag{15}$$

with the boundary conditions :

$$w_i(T,x;t,P) = M_i(x) \quad , \quad \forall\ x,t,P;\ i = 1,2,$$

where $Q_t^i = y_i(t,P_t)$, $i = 1,2$, and where the brackets $<,>$ denote the duality between C^1 functions and first order distributions.

Remark : Very little is known about the solutions of (15) which constitutes a non-linear integro-differential system. It is interesting to interpret the coupled minimization problem of (15) as a trade-off between cost and information, since the Lie derivative term describes the variation of probability induced by a variation of control.

To conclude this survey of theoretic methods for non-zero sum differential games, let us just mention the analysis in (Dockner, Feichtinger, Jorgensen 1983) of classes of games showing simplifications on the open-loop Hamiltonians, so that

the optimal controls can be directly obtained by a system of differential equations :

$$\dot{u}_i = \Psi_i(u_1, u_2, t) \quad , \quad i = 1,2.$$

This situation occurs for example when $\frac{\partial H_i}{\partial u_i}$ and $\frac{\partial H_i}{\partial x_i}$ do not contain those adjoint components p_i^j corresponding to x_j , $j \neq i$.

III - An Illustrative Example : Growth Strategies in a Price-Setting Duopoly. (Lévine , Thépot 1982)

Let us consider a price setting duopoly aver an infinite horizon when the outputs of the competitors are substituable. At time t, firm i charges the price $p_i(t)$; its demand x_i by unit of time is supposed to depend on both prices :

$$x_i(t) = x_i(p_i(t), p_j(t), t). \tag{16}$$

Without a great loss of generality, we will assume henceforth that the demand functions are time independent and constant elasticities functions in the form $x_i = B_i p_i^{-\varepsilon_i} p_j^{\eta_i}$, where B_i is a constant depending on the variable units, ε_i the elasticity with respect to i's own price, η_i the crosselasticity with respect to competitor j's price, satisfying the following inequalities :

$$\varepsilon_i > 1, \ \eta_i \geq 0; \quad D = \varepsilon_1\varepsilon_2 - \eta_1\eta_2 > 0 \ , \tag{17}$$

which merely express classical assumptions on the demand functions

1. Definition of the differential game.

Each firm is supposed to maximise its net present value; then the problem can be stated as the following differential game :

$$J_i = \int_0^\infty ((p_i - c_i)x_i - v_i I_i) \exp(-d_i t)dt \ ; \tag{18}$$

$$\dot{y}_i = I_i - \omega_i y_i, \quad y_i(0) = \xi_i \ ; \tag{19}$$

$$0 \leq I_i \leq (1/v_i)(p_i - c_i)x_i \ ; \tag{20}$$

$$x_i \leq y_i \ . \tag{21}$$

where y_i is the output capacity of firm i, I_i the rate of investment in volume of capacity, c_i the production cost by unit of output, ω_i the rate of depreciation of capacity, v_i the price of unit volume of investment, ξ_i the level of capacity at time 0. Rels. (20) express that the investment is irreversible and that firm i is not allowed at any time to lose money; all the parameters v_i, c_i, ω_i are supposed to be constant throughout the horizon. Hence it is a differential game with two state variables y_1, y_2 and two control variables p_i, I_i at the disposal of each competitor.

2. Open loop strategies in the duopoly

By using the classical results (see sect. II.1.1.), we define the current value dualized Hamiltonian H_i of firm i as follows :

$$H_i = (p_i - c_i)x_i - v_i I_i + q_i(I_i - \omega_i y_i) + \psi_i(I_j - \omega_j y_j) + \alpha_i(y_i - x_i). \quad (22)$$

with q_i, ψ_i, α_i being respectively the costate variables associated to capacity y_i, y_j are the Kuhn and Tucker multiplier associated to constraint (21). The classical necessary conditions yield :

$$\dot{q}_i = (\omega_i + d_i)q_i - \alpha_i \ , \ \dot{\psi}_i = (\omega_j + d_i)\psi_i \ ; \quad (23)$$

$$(x_i + (p_i - c_i)\frac{\partial x_i}{\partial p_i})(1 + \frac{1}{v_i})(q_i - v_i)^+ - \alpha_i \frac{\partial x_i}{\partial p_i} = 0 \ ; \quad (24)$$

$$\lim_{t \to \infty} q_i(t)\exp(-d_i t) = \lim_{t \to \infty} \psi_i(t)\exp(-d_i t) = 0 \ ; \quad (25)$$

$$q_i < v_i \Rightarrow I_i = 0 \ , \ q_i = v_i \Rightarrow I_i > 0 \text{ undetermined},$$
$$q_i > v_i \Rightarrow I_i = \frac{1}{v_i}(p_i - c_i)x_i \ ; \quad (26)$$

$$\alpha_i \geq 0 \ , \ (y_i - x_i) \geq 0 \ , \ \alpha_i(y_i - x_i) = 0 \ . \quad (27)$$

For the sake of simplicity we do not consider situations where excess capacity may occur. Accordingly (26) determine the three policies likely to be chosen by each firm along the equilibrium path : [policy 1 ($q_i < v_i$): $I_i = 0$; policy 2 (permanent policy, $q_i = v_i$); policy 3 ($q_i > v_i$): $I_i = \frac{1}{v}(p_i - c_i)x_i$] . A combination

(k-s) of policies where firm i and firm j use respectively polic k and s is called a duopoly regime.

It is easy to show that regime (2-2) is the final regime of the duopoly to be held from a time t^* ($t^* < +\infty$); this regime coincides with the long term classical static equilibrium of the duopoly with constant prices $p_i^*=\varepsilon_i(c_i+(\omega_i+d_i)v_i)/(\varepsilon_i-1)$.

To emphasize growth strategies of the firm, we suppose that the initial production capacities ξ_i are lower than the long term production levels $x_i^* = x_i(p_i^*,p_j^*)$; as a result the firms are both incitated to invest and to grow from the beginning. Three types of equilibrium paths can be found according to the values of the initial capacities ξ_1 and ξ_2 :

For initial capacities of same magnitude, the equilibrium path is in the form (3-3) → (2-3) → (2-2) : at the beginning, the competitors use their maximum investment policies 3 while decreasing the prices and increasing the production until time t_i^* when the price reaches the value p_i^*. Then, firm i adopts the permanent policy 2 with its price being kept constant; firm j's production is still increasing but firm i's is decreasing. At time t_j^* price p_j becomes equal to p_J^* and the duopoly adopts its permanent regime with production and prices being constant up to infinity.

For a high initial capacity ξ_i and a low initial capacity ξ_j the equilibrium path is either in the form (3-3) → (1-3) → (1-2) → (2-2) or (3-3) → (1-3) → (2-3) → (2-2).Initially, the firms use their maximum investment policies as previously. Howeve at a time $\hat{t}_i$, firm i stops its investment although price p_i has not yet reached the value p_i^*. As a result, firm i goes through a stop in investment period while its production decreases.

It turns out that the growth of the firms is not created through monotonically increasing productions for both competitor Moreover, in some cases, one of the firms has to stop investment during a transitory period, as the decrease of the competitor's price causes too much of a decline in demand.

3. Feedback strategies

In the closed Loop formulation, the prices and the rates of investment are to be sought in the feedback form :

$$p_i = p_i(y_i,y_j,t),\ I_i = I_i(y_i,y_j,t) \qquad (28)$$

The feedbacks are determined by the Nash equilibrium of the Hamiltonians H_1 and H_2 at any point in time t and for any production capacities y_1 and y_2. As a result, the rates of investment I_i are given by (26), as in the open loop case, and the capacities constraints (21) are saturated : $x_i(p_i,p_j)=y_i$, from which we deduce the feedback laws of the prices :

$p_i = y_i^{(-\varepsilon_j/D)}\ y_j^{(-\eta_i/D)}$, and consequently those of the investments The characteristic equations (9) actually take the form :

$$\dot{q}_i = -(p_i-c_i)-\frac{\partial p_i}{\partial y_i}y_i-(q_i-v_i)\frac{\partial I_i}{\partial y_i}+(\omega_j+d_i)q_i-\psi_i\frac{\partial I_j}{\partial y_i}\ ; \qquad (29)$$

$$\dot{\psi}_i = -\frac{\partial p_i}{\partial y_j}y_i-(q_i-v_i)\frac{\partial I_i}{\partial y_j}+((\omega_j+d_i)-\frac{\partial I_j}{\partial y_j})\psi_i\ ; \qquad (30)$$

+ (25).

Clearly, the feedback strategies are sequences of the three policies defined above in the open loop case. However, some differences have to be pointed out :

a) The feedback final regime (2-2) holds with constant prices $p_i^{**} = (c_i+(\omega_i+d_i)v_i)/(1-\varepsilon_j/D)$ which are higher than permanent open loop prices p_i^{*}. This means that, in the long run,

and contrarily to what is intuitively expected, feedback strategies imply more cooperative behaviour than the open loop strategies do.

b) In the growing phase of the duopoly, regime (2-3) does not hold in feedback with firm i keeping its price constant. In this regime, the prices evolve according to a non linear differential equations system (see Lévine,Thépot 1982) which indicates that both prices are decreasing. It turns then out that the feedback strategies express a tendency towards some mimetism and synchronization of the competitor's decisions.

IV - Generalized Competition Dynamic Models

Price (or quantity) manipulation is basically considered by the managers as a two-edged sword which jeopardizes the profitability of the firm rather than really affects the rival's position. Accordingly, the firms are more and more involved in using other competitive weapons. Some dynamic duopoly models therefore have emphasized more accurate types of competition on advertising, quality of the products or R & D projects for instance. Let us outline some typical and recent contributions in this field.

1. An advertising model (Deal 1979)

Deal has developed extension of the classical monopolistic sales response model of (Vidale, Wolfe 1957) :
Let $x_i(t)$ and $a_i(t)$ the sales and the advertising expenditures per unit of time at date t of firm i. The evolution of the sales are given by the following differential equations :

$$\dot{x}_i = -\delta_i x_i + \beta_i a_i [M - x_1 - x_2]/M \; ; \tag{31}$$

$$x_i(0) = \xi_i \; , \; x_1 + x_2 \leqslant M \; . \tag{32}$$

where δ_i = the sales decay parameter, β_i = the sales response parameter and M = the total potential market size ($\beta_i, \delta_i > 0$) Eqn. (31) indicates that advertising expenditures increase the sales; however, such an increase is more efficient when the market is saturated (namely when $x_1 + x_2$ is close to M). As a result, advertising expenditures of a firm have a direct effect on its own sales and an indirect one on the competitor's as they contribute to saturate the total market.

Deal defines the objective function J_i of firm i as a weighted sum of the market share at time T and the sum of the profits earned over the horizon.

$$J_i = \omega_i x_i(T)/[x_1(T) + x_2(T)] + \int_0^T [p_i x_i(t) - a_i^2(t)]\,dt \quad . \quad (33)$$

with p_i being the net revenue coefficient and ω_i the weighting factor for the performance index. The problem is then stated as a differential game which is numerically solved in Open Loop. The obtained results for a wide range of values of the parameters give interesting insights on the relative importance over the horizon of the direct and indirect effects of advertising.

2. <u>A marketing mix model</u> (Thépot 1983)

This model is related to the price setting model presented above in Section 3. The demand of firm i is assumed to be in the form :

$$x_i(t) = X_i[p_i(t), p_j(t), A_i(t), A_j(t)] \exp(\gamma_i t) \quad . \quad (34)$$

where $A_i(t)$ denotes the goodwill of firm i, defined by the differential equation $\dot{A}_i = a_i - r_i A_i$ with a_i representing the advertising expenditures per unit of time, r_i the depreciation rat of the goodwill and γ_i the growth rate of the demand. Then the problem is stated as the differential game :

$$J_i = \int_0^{\infty}[(p_i-c_i)x_i-v_iI_i-a_i]\exp(-d_it)dt \; ; \qquad (35)$$

$$\dot{y}_i = I_i-\omega_iy_i, \dot{A}_i = a_i-r_iA_i, y_i(0) = \xi_i, A_i(0) = A_i^o \; ; \qquad (36)$$

$$I_i \geq 0 \; , \; a_i \geq 0 \; , \; y_i - x_i \geq 0 \; . \qquad (37)$$

In this model each firm has three control variables at its disposal : the price p_i, the investment I_i and the advertising expenditures a_i.

By emphasizing the Open Loop equilibrium and the case where the demand functions are constant elasticities functions, it is shown how Competition and Growth interact in the investment and marketing strategies of the firms. It turns out that the cross-elasticities of the demand with respect to the goodwill play an important role : they determine whether Competition holds through pricing or advertising decisions. This is due to the fact that pricing and advertising decisions quite differently affect the profits : the first ones have an instantaneous effect while the impact of the second ones are displayed over time through goodwill variations.

Two situations may occur : either the competitors behave in a close way to the monopoly case by cohabiting in the industry while increasing their sales and both benefiting of the growth, or one of them is self eliminated of the market. In some case, this elimination process leads this firm to manipulate its price in order to avoid excess capacity.

3. A model of R & D competition (Reinganum 1982)

J.F. Reinganum addresses the problem of resource allocation to Research and Development in a competitive context by developin a dynamic duopoly (in fact oligopoly) model which incorporates the main aspects of this type of competition over a non already existing market. Each firm is assumed to accumulate knowledge

relevant to the innovation by expending resources on research activity or knowledge acquisition. The knowledge acquisition process is assumed to be deterministic whereas the date of successful completion of the project is a random variable. Then the problem can be stated as the differential game :

$$J_i = \int_0^T [P_L \lambda \mu_i + P_F \lambda \mu_j - c_i(\mu_i)][\exp - (z_1+z_2)]\,dt;$$

$$\dot{z}_i = \mu_i \quad , \quad z_i(0) = 0 \; ; \quad 0 \leq \mu_i \leq B \; ,$$

where $\mu_i(t)$ is firm i's rate of knowledge acquisition, $c_i(\mu_i)$ the discounted cost of additional knowledge acquired at time t, B is an upper bound of knowledge acquisition; P_L is the present value of firm i's reward if it is the first to succeed in the completio of the project, P_F if it is the second ($P_F \leq P_L$). Let t_i be the time at which firm i succeeds; it is supposed that $\text{Prob}\{t_i \leq t\} = 1 - \exp[-z_i(t)]$ and that the conditional probability that firm i will succeed in the next instant, given that it hos not already done so, is $\text{Prob}\{t_i \varepsilon (t,t+dt)/t_i > t\} = \lambda \mu_i(t), (\lambda > 0)$

Consequently, J_i is the expected net present value of the gain of firm i according to the fact that imitation is costless and immediate.

Due to the specific features of the exponential distribution, it turns out that Open Loop and Closed Loop strategies coincide. Analytical solutions are obtained for interior solutions $\{0 < \mu_i < B\}$.

4. Conclusion

Differential games techniques are an appropriate conceptual framework to analyse the competitive strategies of firms in a dynamic context, although a very limited number of models can be completely analytically solved. However, they provide a unified language which makes comparisons and economic interpretations meaningful. It turns out also that Competition does not exhaust

the game situations in which firms are involved : relations between industrial firms and banks, industrial firms and unions may be studied with similar tools. Many applications of Differential Games in Economics are therefore expected.

References

Basar T. 1977 Informationally non unique equilibrium solutions in differential games SIAM J. Control 15 p 636-660.

Basar T., Olsder G. 1982 Dynamic non-cooperative game theory Academic Press New-York.

Bensoussan A. 1974 Points de Nash dans le cas de fonctionnelles quadratiques et jeux différentiels linéaires à N personnes SIAM J. Control 12 p 460-499.

Case J. 1969 Towards a theory of many player differential games SIAM J. Control 7 p 179-197.

Deal K. 1979 Optimizing advertising expenditures in a dynamic duopoly Oper. Res. 27 p 682-692.

Dockner E., Feichtinger G., Jorgensen S. 1983 Tractable classes of non-zero sum open-loop Nash differential games. Theory and examples Working paper.

Harsanyi J. 1968 Games with incomplete information played by Bayesian players. Parts I, II, III Management Science 14 N° 3, 5, 7.

Lesourne J., Leban R. 1982 Control theory and the dynamics of the firm : a survey O. R. Spektrum.

Levine J. 1981 Incomplete information in differential games and team problems : necessary and sufficient conditions 8th IFAC World Congress Kyoto.

Levine J.,Thepot J. 1982 Open-loop and closed-loop equilibria in a dynamic duopoly. In Optimal ControlTheory and Economic Analysis, G. Feichtinger Ed. North Holland, p 143-156.

Levine J. 1983 On the dynamic programming equations of Nash equilibria and their associated information structures, Working paper.

Lukes D. 1871 Equilibrium feedback control in linear games with quadratic costs SIAM J. Control 9 p 234-252.

Mehlmann A. 1982 On relations between open-loop and closed-loop N Nash solutions in deterministic differential games. In Optimal Control Theory and Economic Analysis, G. Feichtinger Ed. North Holland, p 399-413.

Owen G. 1968 Game theory Saunders Company.

Reinganum J.F. 1982 A dynamic game of R & D : patent protection and competitive behaviour. Econometrica 50 p 671-688.

Simaan M., Takayama T. 1978 Game theory applied to dynamic duopol problems with production constraints. Automatica 14 p 161-166.

Stalford H.,Leitmann G. 1973 Sufficiency conditions for Nash equilibrium in N-person differential games. In Topics in differential games A. Blaquière Ed. North Holland p 345-376.

Starr A.W., Ho Y.C. 1969 Non-zero sum differential games J. Optimiz. Theory and Appl. 3 N° 3 p 184-206 and N° 4 p 207-219.

Tapiero C.S. 1979 A generalization of the Nerlove)Arrow model to multifirms advertising under uncertainty, Management Science 25 p 907-915.

Thepot H. 1983 Marketing and investment strategies of duopolists in a growing industry, J. of Econ. Dyn. and Control.

Vidale M.L., Wolfe H.B. 1957 An operations research study of sales response to advertising. Oper. Res. 27.

Von Neumann J., Morgenstern O. 1944 Theory of games and economic behaviour Princeton University Press.

Witsenhausen H. 1968.A counterexample in stochastic optimum control SIAM J. Control 6 p 130-147.

Optimal Control Theory and Economic Analysis
G. Feichtinger (ed.)

OPEN LOOP AND CLOSED LOOP EQUILIBRIA IN A DYNAMICAL DUOPOLY

Jean Lévine[*] & Jacques Thépot[***]

[*]Centre d'Automatique et d'Informatique de l'Ecole des Mines
Fontainebleau, France

[***]European Institute for Advanced Studies in Management
Brussels, Belgium

This paper analyses a dynamical duopoly when each firm is able to decide at any point in time the price and the investment. The problem is stated as a Differential Game. One studies the strategies of the firms in Open loop and in Closed loop formulation. Analytical results are obtained in both cases.

INTRODUCTION

The aim of this paper is to compare the open loop and the closed loop solution in a price-setting dynamical duopoly and to give some economic interpretation of the differences observed between them. As far as we know, there is no closed loop analytical Nash equilibrium solution except in the classical linear-quadratic case without constraints. By emphasizing a particular case of our model where the demand functions are constant elasticities functions, analytical results are obtained. That allows us to give economic interpretations. The paper is divided into 4 sections The first one is devoted to the statement of the model ; in the second one we recall the results previously obtained by one of us in the open loop case ; in the third one we study the closed loop solution. The last section deals with the economic interpretations of the results.

1. THE MODEL

We present there the main features of a model of dynamical duopoly which has been introduced previously by one of the authors (see [5])

1.1. The demand functions

Let us consider two firms (firm 1 and firm 2) involved in a price-setting duopoly throughout the horizon $[o, +\infty[$ At time t, firm i sells its output at the price $p_i(t)$; its demand x_i by unit of time is assumed to depend on both prices $x_i = x_i\,(p_i(t), p_j(t), t)$ [1]

1 For convenience, we will note in the following i, j for $i = 1,2$; $i \neq j$

without a great loss of generality we will assume henceforth that the demand functions x_i are time independent constant elasticities functions (no trends) of the form

(1) $x_i = a_i p_i^{-\varepsilon_i} p_j^{\eta_i}$

where a_i is a constant depending on the variables units, ε_i the elasticity of the demand with respect to i's own price, η_i the elasticity of the demand with respect to the competitor price. Let us introduce now the conditions

(2) $\varepsilon_i > 1$, $\eta_i > 0$; (3) $D = \varepsilon_1 . \varepsilon_2 - \eta_1\eta_2 > 0$.

The inequalities (2) express merely classical assumptions on the demand functions The inequality (3) expresses that the determinant of the coupling matrix of the demands M defined by

$$M = \begin{bmatrix} \frac{\partial x_1}{\partial p_1} & \frac{\partial x_1}{\partial p_2} \\ \frac{\partial x_2}{\partial p_1} & \frac{\partial x_2}{\partial p_2} \end{bmatrix}$$

is always strictly positive. It means that the direct effects of a prices variation are globally stronger than the undirect ones.

1.2. Definition of the differential game

Each firm is supposed to maximize its net present value. Then the problem can be stated as a non-zero sum differential game:

(4)
$$V_1 = \int_0^{\infty} [(p_1 - c_1)\, x_1(p_1,p_2) - v_1 I_1]\, e^{-d_1 t}\, dt,$$
$$V_2 = \int_0^{\infty} [(p_2 - c_2)\, x_2(p_1,p_2) - v_2 I_2]\, e^{-d_2 t}\, dt.$$

(5) $\dot{y}_1 = I_1 - \omega_1 y_1$, $\dot{y}_2 = I_2 - \omega_2 y_2$;

(6) $I_1 \geq 0$, $I_2 \geq 0$;

(7) $I_1 \leq (1/v_1)(p_1 - c_1)\, x_1(p_1,p_2)$, $I_2 \leq (1/v_2)(p_2 - c_2)\, x_2(p_1, p_2)$;

(8) $p_1 \geq 0$, $p_2 \geq 0$;

(9) $x_1(p_1,p_2) \leq y_1$ $x_2(p_1,p_2) \leq y_2$;

(10) $y_1(0) = \xi_1$, $y_2(0) = \xi_2$.

where $y_i(t)$ is the output capacity of firm i at time t, $I_i(t)$ is the rate of investment in volume of output capacity at time t, c_i is the production cost by

unit of output, ω_i is the rate of depreciation of capacity, v_i is the price of unit volume of investment, d_i is the discount rate of firm i which equals the rate of interest of its shareholders, ξ_i is the level of output capacity at time t = 0.

Differential equations (5) state that a change in capacity equals the gross increase in capacity less depreciation for each firm; Constraints (6) that gross investment must be nonnegative (investment constraints); Constraints (7) that the cash flow must be greater or equal to the investment (financing constraints) according to the fact that no borrowing opportunities are allowed to finance the activity of the firms; Constraints (9) that the productions of the firms are limited by the capacities.

We suppose that the parameters c_i, ω_i, d_i, v_i are constant over the horizon. On the other hand, the firms are assumed to have a perfect knowledge of all the elements determining the dynamical system set up. In particular, each firm is supposed to know the demand functions x_1 and x_2 and all the parameters like v_j, c_j, d_j, ω_j describing the competitor's activity.

Relations (4) - (10) define a non-zero sum differential game with complete information, with two state variables y_1 and y_2 and two control variables p_i and I_i at the disposal of each player. As it is well known, there are two types of non cooperative equilibrium solutions of differential games the Open Loop equilibrium, the Closed Loop equilibrium.

2. OPEN LOOP SOLUTION

In this section we will recall the main results in open loop formulation. For an extensive presentation of them, the reader is invited to refer to [5].

2.1. Characteristic equations of open loop equilibrium

Open loop solution consists in considering the problem (4) - (10) as a classical optimal control problem for firm 1 (resp. firm 2) when the control variables $p_2(.)$ and $I_2(.)$ (resp. $p_1(.)$ and $I_1(.)$) are fixed over the horizon. The Current Value dualized Hamiltonian H_i^d may be written as

$$H_i^d = (p_i - c_i)\, x_i - v_i I_i + q_i(I_i - \omega_i y_i) + \Psi_i(I_j - \omega_j y_j) + \alpha_i(y_i - x_i)$$

where q_i and Ψ_i are the costate variables associated to the output, capacities y_i and y_j respectively, α_i is the Kuhn and Tucker multiplier associated to the capacity constraint.

The necessary conditions may be written as

$$(11)\quad \dot{q}_i = (\omega_i + d_i)\, q_i - \alpha_i$$

(12) $\Psi_i = (\omega_j + d_i)\ \Psi_i$

(13) $(x_1 + (p_i - c_i)\ \frac{\partial x_i}{\partial p_i})\ (1 + (1/v_i)\ (q_i - v_i)^+) - \alpha_i\ \frac{\partial x_i}{\partial p_i} = 0$

(14)
$$\begin{cases} q_i - v_i > 0 \iff I_i = (1/v_i)(p_i - c_i).x_i, \\ q_i = v_i \iff I_i > 0 \text{ undetermined}, \\ q_i - v_i < 0 \iff I_i = 0. \end{cases}$$

(15) $\alpha_i \geq 0\ ,\ y_i - x_i > 0\ ,\ \alpha_i(y_i - x_i) = 0$

N.B. We suppose here that the capacity constraint and the financial constraint of each firm is under its responsibility; consequently none of the firms has to take into account in its strategy the fact that the competitor's constraints are satisfied, namely $x_j \leq y_j$ and $I_j \leq (1/v_j)(p_j - c_j).x_j$

2.2. Policies of the firms; regimes of the duopoly

Relations (14) - (15) determine the policies that firm i will adopt in open loop equilibrium. Hence, six policies are a priori possible according to the various combinations of positive or zero α_i and $(q_i - v_i)$. But it is easy to prove that the firms do not invest when they have excess capacities (see [5]) Only four policies therefore are liable to be part of an equilibrium trajectory ; they have the following characteristics

Policy 0 The firm has excess capacity, it does not invest and distributes all the cash flow to shareholders. The output capacity is decreasing $\alpha_i=0$, $q_i-v_i < 0$.

Policy 1 The firm has no excess capacity. It does not invest and distributes all the cash flow. The output capacity and the production are decreasing $\alpha_i > 0$, $q_i-v_i < 0$.

Policy 2 The firm has no excess capacity. It invests to adapt the production to the demand; it distributes the remaining cash flow to the shareholders. This policy is called "balanced policy" $\alpha_i > 0$, $q_i-v_i = 0$.

Policy 3 The firm has no excess capacity. It invests at the maximum level without distributing anything to the shareholders $\alpha_i > 0$, $q_i-v_i > 0$.

Let us specify the terminology by the following definitions

Definition 1 A combination (k,s) k = 0,... 3; s = 0,...3, of policies where firm 1 uses policy k and firm 2 policy s is called a duopoly regime.

Definition 2 : A sequence of regimes is called a path ; the corresponding sequence of policies of firm i is called Strategy of firm i.

In the case where the demand functions are constant elasticities functions, it is possible to compute the investment rates I_1 and I_2, the variation rates $\pi_1=\dot{p}_1/p_1$ and $\pi_2=\dot{p}_1/p_2$ of the prices p_1 and p_2 for each duopoly regime. Without a great loss of generality, we do not consider the policies 0 with excess capacity since they may occur only in very specific situations [2] which are not economically relevant with regard to our problem. Let $p_i^{\ddot{}}$ be the balanced price held by firm i in balanced policy $p_i^{\ddot{}} = [\varepsilon_i/(\varepsilon_i-1)].(\omega_i + d_i)\ v_i + c_i)$
Let p_i^{++} be the price level defined by $(\omega_i + (\varepsilon_i/\eta_j)\omega_j)\ v_i + c_i$ and $\bar{p}_i=\omega_i v_i+c_i$

A regime (k-s) is said to be <u>feasible</u> if policies k and s are feasible respectively for firms 1 and 2 according to the investment and financing constraints.

2.3. <u>Open loop strategies over the horizon</u>

2.3.1. The analysis of the open loop equilibrium will be done here under the following assumptions :
a) At time t = 0 the firms have no excess capacity, they are selling their outputs at prices p_1^o and p_2^o such that (16) $x_1(p_1^o, p_2^o) = \xi_1$; $x_2(p_1^o, p_2^o) = \xi_2$.
b) The initial prices p_1^o and p_2^o are higher than the balanced prices $p_1^{\ddot{}}$ and $p_2^{\ddot{}}$.
(17) $p_1^o \geq p_1^{\ddot{}}$; $p_2^o \geq p_2^{\ddot{}}$.
c) The costate variables $q_1(t)$ and $q_2(t)$ are continuous along the equilibrium path.

Under these circumstances, the open loop equilibrium path has to be chosen among the following types of trajectoires
s_1 (3-3) → (2-3) → (2-2)
s_2 (3-3) → (1-3) → (2-3) → (2-2)
s_3 (3-3) → (1-3) → (1-2) → (2-2)

2.3.2. To simplify the analysis, we will give the results on the basis of Figure 1 where the situations are represented in the plane $\{p_1, p_2\}$ (i) the initial conditions correspond to a point of coordinates (p_1^o, p_2^o) belonging to the area s_1Ss_2; (ii) whatever are the initial conditions, the duopoly reaches the balanced regime in a finite time; all the paths are thus ending at the point S of coordinates $\{p_1^{\ddot{}}, p_2^{\ddot{}}\}$; (iii) the curve USz represents a specific situation for any initial conditions lying on it, the firms will both adopt the policy 3 of maximum investment and decrease their prices until the date $t^{\ddot{}}$ when the prices are <u>together</u> equal to the balanced prices. This curve can be numerically obtained as the solution of the differential equation

2 Due to the fact that in general each firm can avoid excess capacity by manipulating the price in order to fit the production and the demand at every moment.

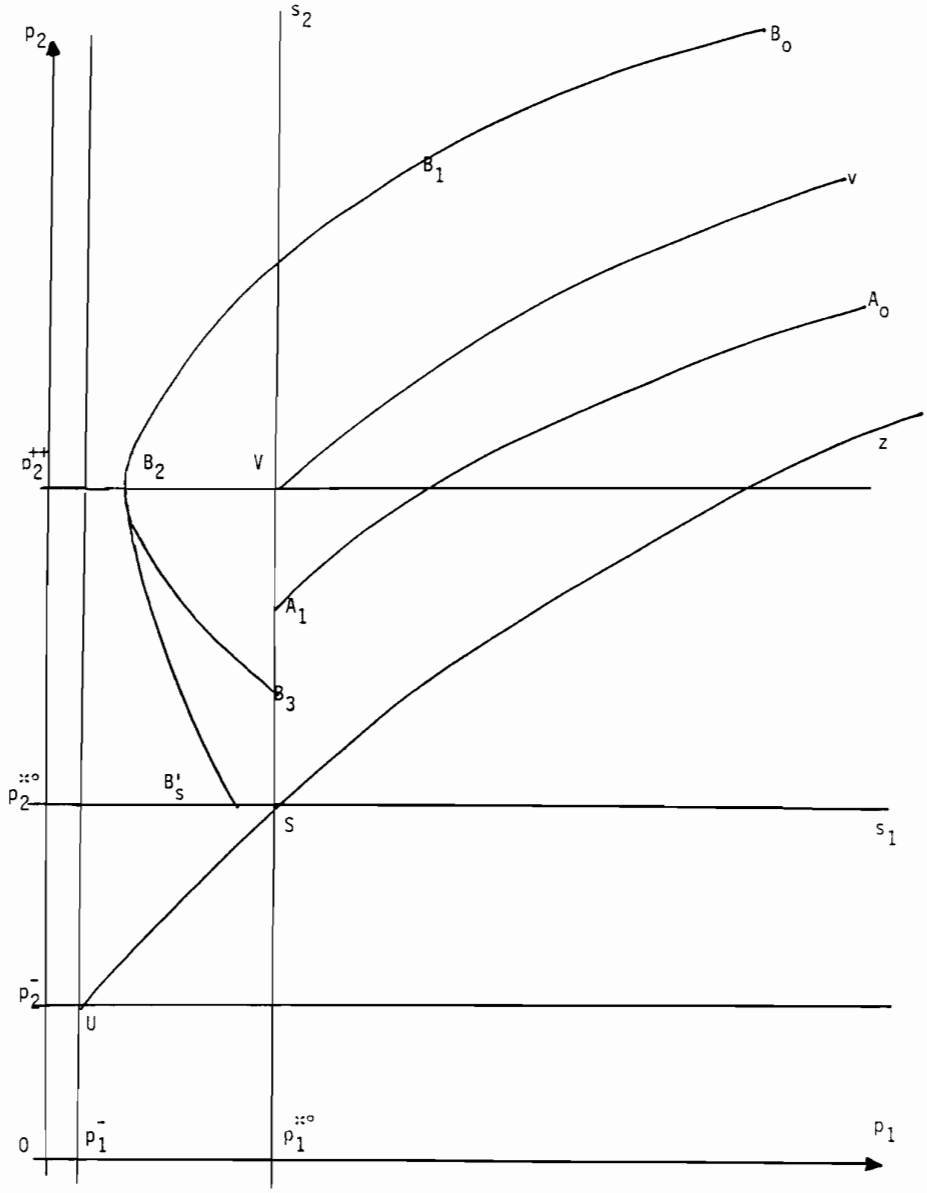


Figure 1 Open loop trajectories of the duopoly.

$$(18)\quad \frac{dp_2}{dp_1} = \frac{p_2}{p_1}\,\frac{(\varepsilon_1/v_2)(\bar{p_2} - p_2) + (n_2/v_1)(\bar{p_1} - p_1)}{(\varepsilon_2/v_1)(\bar{p_1} - p_1) + (n_1/v_2)(\bar{p_2} - p_2)}\ ;\quad p_2(p_1^{\ddot{}}) = p_2^{\ddot{}}.$$

The curve USz divided the plan into two main areas (i) the area s_2Sz where firm 1 has the possibility to reach the balanced price by using policy 3 before firm 2 can do it ; (ii) the area s_1Sz where firm 2 has the possibility to reach the balanced price by using policy 3 before firm 1 can do it. We will emphasize therefore the area s_2Sz. The results corresponding to the area s_1Sz are easy to obtain by using symetrical arguments. Two situations have to be considered

1) When the top price p_2^{++} is higher than the initial price p_2^o of firm 2, the regime (2-3) is feasible because firm 1 can keep a positive flow of investment although its production is decreasing because of the price decrease of firm 2. Consequently the equilibrium path is of the form s_1 At the beginning (represented by the point A_o), the firms use their maximum investment policy 3, they decrease their prices and increase their productions until the date $t_1^{\ddot{}}$ where the price p_1 reaches the value $p_1^{\ddot{}}$ (point A_1). Then, firm 1 adopts the balanced policy by keeping its price constant and equal to $p_1^{\ddot{}}$, its production is decreasing while the production of firm 2 is increasing again (and its price p_2 decreasing) until the date $t^{\ddot{}}$ when price p_2 becomes equal to $p_2^{\ddot{}}$ (point S).

2) When the top price p_2^{++} is lower than the initial prices p_2^o, the regime (2-3) is non feasible as long as $p_2 \geq p_2^{++}$ Accordingly, two cases may occur : (a) in the area Z S V v, nothing is changed with respect to the above situation. Firm 1 is able to adopt the balanced policy as soon as its price reaches the value $p_1^{\ddot{}}$ (initial point A'_o); (b) in the area v V s_2 the situation becomes more complicated. At the beginning the firms use their maximum investment policies, as previously. At a time t_1, firm 1 will stop its investment although its price p_1 has not reached yet the balanced value p_1 (point B_1) After this moment, the price and the production of firm 1 are decreasing while the price of firm 2 is always decreasing and its production increasing. Clearly price p_1 becomes lower than the balanced $p_1^{\ddot{}}$. When the price p_2 reaches the value p_2^{++} (point B_2), firm 1 begins to increase its price in order to find again the level $p_1^{\ddot{}}$. Two situations may occur, depending whether firm 1 is the first of the two to be able to put its price at the balanced price level (point B_3) or not (point B'_3). Accordingly the open loop equilibrium path is on the type s_2 or s_3.

2.4. *Comments*

(i) In the open loop formulation, the selling prices of the firms are the key variables which determine the strategies held by the competitors throughout the horizon. The productions are just a consequence of the prices levels

(ii) As a result, it turns out that the growth of the firms does not hold ever with monotonically increasing productions. Moreover, in some cases, one of the firms has an interest in stopping the investment during a certain period because the decrease of the competitor's price makes its demand decline too much.

3. CLOSED LOOP SOLUTION

In closed loop formulation, at any time t the firms are able to observe the production capacities $y_1(t)$ and $y_2(t)$ Such an information structure induces different strategies since the prices and the rate of investment p_i and I_i are now to be sought on the feedback form, say

$p_i = p_i(t, y_i, y_j)$, $I_i = I_i(t, y_i, y_j)$

3.1. Feedback laws of the dynamical duopoly

The feedback laws are determined by applying the Dynamic Programming approach It amounts to say that there is Nash equilibrium of the no dualized hamiltonians H_1 and H_2 at any point in time and for any values of the production capacities y_1 and y_2 :

Max $H_i = (p_i - c_i)x_i - v_iI_i + q_i(I_i - \omega_i y_i) + \Psi_i(I_j - \omega_j y_j)$

(19) $0 \le I_i \le (1/v_i)(p_i - c_i)\, x_i\,(p_i, p_j)$,

(20) $x_i(p_i, p_j) \le y_i$, for p_j, I_j, y_i, y_j, q_i, Ψ_i, t, fixed.

Since the hamiltonian H_i is a linear function of I_i, the maximization of H_i with respect to I_i gives

$$(21) \quad I_i = \begin{cases} 0 \text{ if } q_i < v_i \\ \text{undetermined if } q_i = v_i \\ (1/v_i)(p_i - c_i)\, x_i\,(p_i, p_j) \text{ if } q_i > v_i \end{cases}$$

By maximizing with respect to the price p_i, we get that the constraint (20) is saturated except in the case when $x_i + (p_i - c_i)\frac{\partial x_i}{\partial p_i} = 0$. As mentioned above, that corresponds to a very specific situation which can be avoided by supposing that the average production costs c_i are small As a result, the capacity constraints are saturated :

(22) $x_1(p_1, p_2) = y_1$, $x_2(p_1, p_2) = y_2$,

from which we deduce the feedback laws of the prices which can be written for the constant elasticities functions

(23) $p_1 = y_1^{(-\varepsilon_2/D)} \cdot y_2^{(-\eta_1/D)}$, $\quad p_2 = y_2^{(-\varepsilon_1/D)} \cdot y_1^{(-\eta_2/D)}$

Thus, the feedback laws of the investment are given by (21) where the prices p_1 and p_2 are replaced by the feedback laws (23)

3.2. Characteristic equations of closed loop equilibrium

By applying sufficient conditions (see [1]), we get the relations which determine the closed loop equilibrium of the dynamical duopoly

$$(24)\quad q_i = -(p_i - c_i) - \frac{\partial p_i}{\partial y_i}\cdot y_i - (q_i - v_i)\frac{\partial I_i}{\partial y_i} + (\omega_i + d_i)q_i - \psi_i \frac{\partial I_j}{\partial y_i}$$

$$(25)\quad \psi_i = -\frac{\partial p_i}{\partial y_j}\cdot y_i - (q_i - v_i)\frac{\partial I_i}{\partial y_j} + \left((\omega_j + d_i) - \frac{\partial I_j}{\partial y_j}\right)\cdot \psi_i$$

$$(26)\quad \lim_{t\to\infty} q_i = \lim_{t\to\infty} \psi_i = \lim_{t\to\infty} x_i = 0,$$

where $p_i(y_i, y_j)$ and $I_i(y_i, y_j)$ are given by the feedback laws.

Clearly, rules (21) define the three policies which candidate to be part of the strategies of the firms. Although these policies have the same economic interpretation as they have in open loop solution, they differ in terms of prices and production levels. In the following, we will specify all the elements of the model by a superscript "o" or "c" according to whether they are defined in open loop or in closed loop formulation.

3.3. Closed loop regimes of the duopoly

3.3.1. Final regime of the duopoly

Thanks to conditions (26) the final regime the duopoly will hold is the regime $(2\text{-}2)^c$ with $q_i = v_i$ and $\dot{y}_j = 0$. Consequently, relation (26) may be written as

$$(27)\quad (\omega_1 + d_1)v_1 = (p_1 - c_1) + \frac{\partial p_1}{\partial y_1}\cdot y_1\ ,\quad (\omega_2 + d_2)v_2 = (p_2 - c_2) + \frac{\partial p_2}{\partial y_2}\ y_2.$$

We deduce that the prices p_1 and p_2 are constant in the regime $(2\text{-}2)^c$ and equal to the balanced prices in closed loop $p_1^{\ddot{c}c}$ and $p_2^{\ddot{c}c}$ defined by

$$(28)\quad p_1^{\ddot{c}c} = \frac{c_1 + (\omega_1 + d_1)v_1}{1 - \varepsilon_2/D}\ ,\quad p_2^{\ddot{c}c} = \frac{c_2 + (\omega_2 + d_2)\ v_2}{1 - \varepsilon_1/D}$$

As in the open loop equilibrium, the final regime of the duopoly is characterized by constant prices. But the open loop balanced prices differ from the closed loop ones. We have the inequalities $p_1^{\ddot{o}} \le p_2^{\ddot{c}}$, $p_2^{\ddot{o}} \le p_2^{\ddot{c}}$

Clearly they coincide when one of the cross elasticities n_1 or n_2 vanishes.

Let us study now the transitory regime the duopoly is allowed to hold in order to reach this balanced regime. Clearly, regimes (1-3), (2-1), (3-1) and (3-3) have the same characteristics as in the open loop formulation. However, specific differ-

ences will occur with the semi-balanced regime (2-3) (symmetrically with regime (3-2))

3.3.2. Semi-balanced regime $(2\text{-}3)^C$

In this situation, we have $q_1 = v_1$, $I_2 = \frac{1}{v_2}(p_2 - c_2) \cdot y_2$.
We deduce from relations (24) and (25) that

$$(29)\quad \Psi_1 = \frac{(1 - \varepsilon_2/D)(p_1^{::} - p_1)}{\frac{\partial I_2}{\partial y_1}}$$

$$(30)\quad \dot{\Psi}_1 = -\frac{\partial p_1}{\partial y_2} \cdot y_1 + ((\omega_2 + d_1) - \frac{\partial I_2}{\partial y_2}) \cdot \Psi_1$$

By differentiating Ψ_1 with respect to time t in relation (23) and by putting the result in the differential equation (30), we get after some computations

$$(31)\quad \pi_1 (p_1 + (\varepsilon_1 + \eta_2)(p_1^{::} - p_1)) + \pi_2(1 - \eta_1 - \varepsilon_2)(p_1^{::} - p_1) =$$

$$[(\tilde{p}_2 - p_2)(p_1^{::} - p_1)(1 - \varepsilon_1/D)(1 - \varepsilon_2/D) - \frac{\eta_1\eta_2}{D^2} p_1 p_2] / v_2(1 - \varepsilon_2/D)$$

And for firm 2 which holds the maximum investment policy, we have

$$(32)\quad -\varepsilon_2\pi_2 + \eta_2\pi_1 = \frac{1}{v_2}(p_2 - \bar{p}_2)$$

Where $p_1^{::} = p_1^{::C}$ and $\tilde{p}_2 = \frac{c_2 + (\omega_2 + d_1)v_2}{1 - \varepsilon_1/D}$; $(\tilde{p}_2 = p_2^{::C}$ if $d_1 = d_2)$

Consequently the evolution of the prices p_1 and p_2 is determined by the system of differential equations (31)-(32). It differs considerably from the results obtained in open loop formulation. It must be pointed out that the constant price $p_1 = p_1^{::}$ is solution of (31) only in the case where η_1 or η_2 equals zero. In this particular situation open loop and closed loop regimes coincide since one of the firms is decoupled from its competitor. In regime $(2\text{-}3)^C$, neither the production nor the price of firm 1 remains constant.

3.4. Closed loop strategies over the horizon

We are going to study the strategies that the firms will adopt along the closed loop equilibrium path. The analysis will be done under conditions which are close to those introduced in the open loop case

a) At time t = 0, the firms have no excess capacity; they are selling their outputs at price p_1^o and p_2^o such that

$$(33)\quad x_1(p_1^o, p_2^o) = \xi_1 \ ; \ x_2(p_2^o, p_1^o) = \xi_2.$$

b) The initial prices p_1^o and p_2^o are higher than the closed loop balanced prices $p_1^{::C}$ and $p_2^{::C}$ $\quad p_1^o \geq p_1^{::C}$; $p_2^o \geq p_2^{::C}$

c) The costate variables $q_1(t)$, $q_2(t)$, $\Psi_1(t)$, $\Psi_2(t)$ are continuous along the equilibrium path.
d) The crosselasticities η_1 and η_2 have the same magnitude. This condition will be discussed later.
Under these circumstances, the closed loop equilibrium path has to be chosen among the following types of trajectories
s_1^c $(3\text{-}3) \rightarrow (2\text{-}3)^c \rightarrow (2\text{-}2)^c$,

s_2^c $(1\text{-}3) \rightarrow (2\text{-}3)^c \rightarrow (2\text{-}2)^c$

The results will be illustrated and discussed in the plane $\{p_1, p_2\}$ (Figure 2). As in the open loop case, the plane is divided into two main areas by the curve $US_c z$ of equation (18)

The curve $S_c W_1$ (and symetrically $S_c W_2$) represents the evolution of the prices in regime $(2\text{-}3)^c$ (resp. $(3\text{-}2)^c$) The equation of this curve is given by relations (31) and (32) with the boundary conditions meaning that for some value t^{**}, $p_1(t^{**}) = p_1^{**c}$, $p_2(t^{**}) = p_2^{**c}$ Although this system of differential equations has no evident solution, it can be easily implemented. The numerical experiments we made indicate clearly that, for values of the crosselasticities of the same magnitude, the prices p_1 and p_2 are monotonically decreasing while the productions y_1 and y_2 are increasing along this curve; in addition, the investment rate of firm 1 remains positive. Consequently, it can be argued that, under this condition, the regime $(2\text{-}3)^c$ is feasible. But for high values of the crosselasticity η_2, the situation becomes more complicated and appears to be very different from that intuitive guessing. It requires a more systematic use of numerical experiments that we have made by now. As in the open loop case, we consider just the case where the initial values of the prices correspond to a point in the area $s^2 S_c z$. Accordingly, two situations may occur (i) In the area $w_1 S_c z$ (point A_0), the firms begin to invest at the maximum level up to a time t_1^{**} (point A_1 on figure 2) where firm 1 is able to use the balanced policy 2^c It will use it until the time t^{**} when the two prices p_1 and p_2 are equal to the closed loop balanced prices p_1^{**} and p_2^{**}. It is important to notice that, along this path, the productions of both firms increase monotonically while the prices are strictly decreasing. None of the firms will reach its balanced price before the t^{**} when they are able to reach it together; (ii) In the area $s_2 S_c W_1$ (point B_0), firm 2 begins to invest at the maximum level while firm 1 does not. Price p_2 is decreasing and price p_1 is increasing up to the time t_1^{**} when firm 1 is able to use the balanced policy (point B_1) After this time, the duopoly will be in regime $(2\text{-}3)^c$ up to the time t^{**} when both prices are equalling the balanced prices, as previously. It is interesting to observe that firm 1 will adopt a stop in investment policy at the beginning even in the parti-

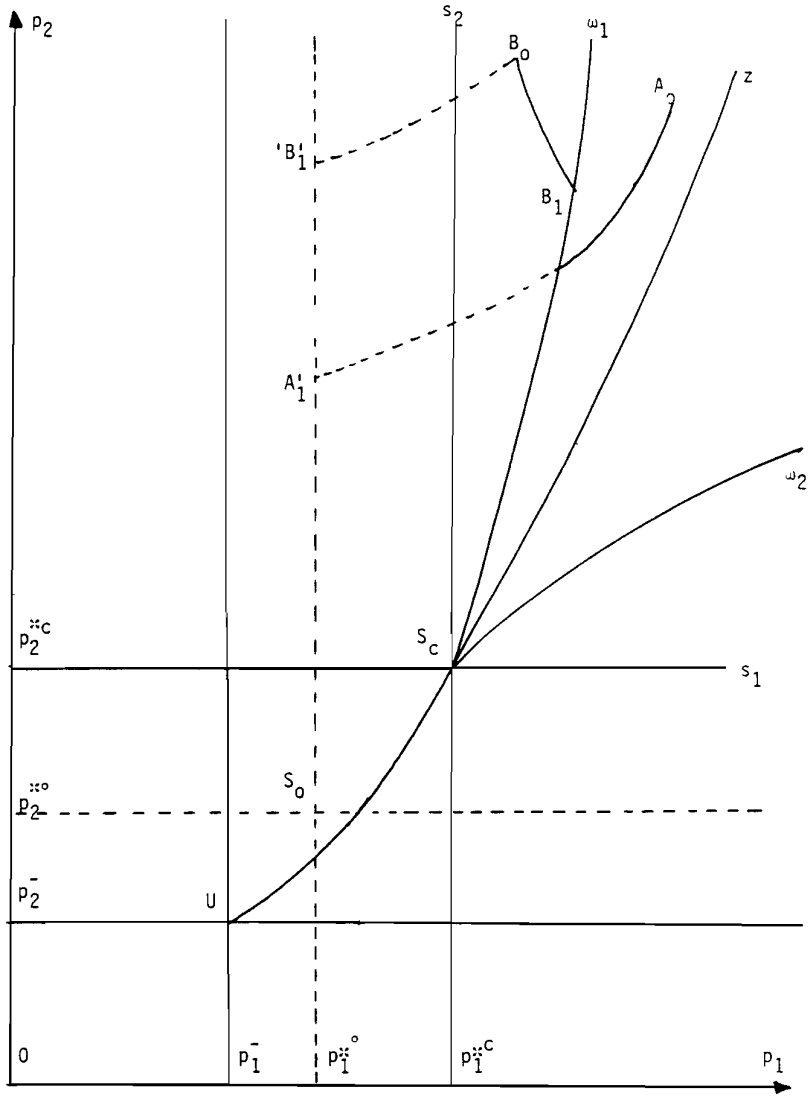


Figure 2 The dotted lines correspond to the open loop trajectories and the continuous lines to the closed loop trajectories

cular situation where the initial production level ξ_1 is lower than the balanced production $y_1^{\ddagger} = x_1^{\ddagger}(p_1^{\ddagger C}, p_2^{\ddagger C})$ (since the production y_1 is increasing along the curve $\nu_1 S_c$)

4. OPEN LOOP VERSUS CLOSED LOOP FORMULATION - WHAT ABOUT THE DIFFERENCE ?

4.1. Economic interpretation of the difference between open loop and closed loop equilibria.

In this paper we did not complete the study of the closed loop formulation it remains to investigate the cases where one of the crosselasticities is high while the other remains low. Such cases deal clearly with specific situations where the duopoly is unbalanced. In these cases, some technicalities appear in the differential equation (31) which require suitable treatments. Nevertheless some hints can be given about the differences occuring in the strategies of the firms when they operate in open loop or closed loop context.

(i) As mentioned above, the final state of the duopoly holds with constant prices in both cases. But the prices are lower in open loop than they are in closed loop. If the crosselasticities η_1 and η_2 are small enough, the productions are higher in open loop than they are in closed loop. Such a difference can be explained as follows : in closed loop, at any time, each firm knows that the competitor has no excess capacity; this information modifies the evaluation of the marginal revenues and consequently the prices. More precisely : in open loop, each firm computes its marginal revenue as if the competitor will not change its price ; in closed loop, each firm computes its marginal revenue taking into account that the competitor will react to any price variation by a suitable variation of its price in order to equal always capacity and production.

(ii) We have seen that regime (2-3) is differently defined in open loop or in closed loop. It holds with constant price for firm 1 equal to the balanced price in the first case and with a price changing over time in the second. As regime (2-3) is a transitory regime in the equilibrium paths, such a difference implies qualitatively as well as quantitatively different strategies.

Let us illustrate it on the examples given in figure 2 For initial values of the prices (and the capacities) represented by the point A_0, the open loop path corresponds to the curve $A_0 A'_1 S_0$. Disregarding the differences in balanced prices, the closed loop and the open loop strategies are qualitatively close enough : in both situations, the firms begin to invest, then firm 1 holds its balanced policy until the time when the competitor is able to do so. But it appears that the productions and the prices evolutions are more smoothed in closed loop than they are in open loop; For initial conditions defined by the point B_0, the strategies are in addition

qualitatively different in closed loop, firm 1 begins by stopping the investment while in open loop it begins by investing at the maximum level (curve $B_o B'_1 S_o$)

4.2. Concluding remarks

The open loop and closed loop equilibria of differential games are theoretically well recognized to be different. They deal with different information structures and their characteristic equations are not identical Unfortunately, closed loop equilibrium is defined as a solution of a partial differential system which is highly untractable in most cases, and economic interpretations,comparisons and empirical investigations generally remain out of touch. Under these circumstances our paper may be seen as a contribution to the difficult tool of modelling the competitive behaviour by comparing the influence of two specific information structures on the strategies of the firms.

REFERENCES

[1] Case, J.H., "Towards a Theory of Many Players Differential Games", SIAM J. Control, Vol 7, No 2, 179-197, 1969.

[2] Friedman, A., "Differential Games", J. Wiley Intersciences, New York, 1971.

[3] Lesourne, J., Modèles de Croissance des Entreprises, Dunod, Paris, 1973.

[4] Lévine, J., "Two persons zero-sum differential games with incomplete information : a bayesian model", in Differential Games and Control Theory, III, P.T. Liu and E. Resien (ed.), M. Dekker, 119-151, New York, 1979.

[5] Thépot, J., "Politiques de prix et d'investissement d'un duopole en croissance", EIASM WP 79-42, 1979.

ON THE SOLUTIONS OF HAMILTON-JACOBI SYSTEMS AND APPLICATIONS TO THE DYNAMIC DUOPOLY

J LEVINE*

* Centre d'Automatique et d'Informatique de l'Ecole Nationale Supérieure des Mines de Paris, Fontainebleau - France

ABSTRACT :

We derive the general characteristic equations solving the Hamilton-Jacobi systems of partial differential equations of a closed-loop dynamic duopoly. An equilibrium solution can be seen to belong to an information structure including some knowledge of the future through the observation of the marginal costs-to-go and we call such an equilibrium a feedforward equilibrium.

We prove that the nonuniqueness of this equilibrium surface is generic.

I - INTRODUCTION

In regard to the number of papers on Optimal control, the litterature on Nash equilibria in dynamic non zero sum oligopolies appears to be very poor, particularly concerning questions like existence, uniqueness, qualitative behaviour, etc

The most studied point of view until now remains the open-loop one, since a generalized version of the helpful minimum principle applies The open-loop solution may be interpreted as an equilibrium without a posteriori information and it is often noted ([1],[2],[4],[5]) that the information structure plays a basic role, yielding,for example,very different behaviours in open-loop or in closed-loop form.

On the other hand, closed-loop solutions still deserve a great deal of work if we want to answer questions like : "can we be sure to have all the possible solutions and how can we compute them, how does the information structure interfere into their computation, are there intermediate information structures between open-loop and closed-loop that can be obtained without introducing probabilistic models, are there games without solution of a given type, and so on"...

To specialize our subject, let us first state a fundamental remark : the local or global characterizations of the closed-loop solutions ([3],[5],[6]) always assume that we directly have a candidate solution in closed-loop form, whereas such a candidate solution is computed as a Nash equilibrium of the Hamiltonians and is thus generally a function of the state x and of the whole adjoint vector λ Therefore, we have to complete the solution of the game to be able to express λ as a function of x and finally to put the Nash Strategies into their closed-loop form, but we are just in the case where no theory exists to solve the game !

The aim of this paper is thus to give a generalized theory of characteristics to solve the coupled Hamilton-Jacobi systems of partial differential equations obtained by the dynamic programming method and to draw some conclusions about the number of possible equilibria.

II - STATEMENT OF THE PROBLEM

II.1 Open-loop and closed-loop equilibria :

For obvious notational reasons, we shall restrict ourselves to the duopoly case. We shall suppose that the state equation is given by :

$$(1) \qquad \begin{cases} \dot{x}(t) = f(t,x(t),u_1(t,x(t)),u_2(t,x(t))) \\ x(o) = x_o \end{cases}$$

in which x is an n-vector, and u_1 and u_2 are m_1 and m_2 vectors representing respectively player 1's and 2's strategy. u_1 and u_2 are subject to constraints $u_1(t,x) \in U_1$, $u_2(t,x) \in U_2$ and are such that (1) has at least one solution. The cost functions for players 1 and 2 are given by :

$$(2) \begin{cases} J_1(u_1,u_2) = \int_o^T g_1(t,x(t),u_1(t,x(t)),u_2(t,x(t)))e^{-\alpha t}\, dt \\ J_2(u_1,u_2) = \int_o^T g_2(t,x(t),u_1(t,x(t)),u_2(t,x(t)))e^{-\beta t}\, dt \end{cases}$$

where α and β are positive actualization coefficients. The functions f, g_1 and g_2 are supposed regular in all their arguments.

We recall that an open-loop Nash equilibrium (u_1^o,u_2^o) is a pair of functions $(u_1^o(.),u_2^o(\ ))$ depending on t and the initial condition x_o only, such that $(u_1^o(t),u_2^o(t)) \in U_1 \times U_2$ $\forall t$ and :

$$(3)\begin{cases} J_1(u_1^o,u_2^o) \leqslant J_1(u_1,u_2^o) & \forall u_1 \text{ function of t only, } u_1(t) \in U_1 \\ J_2(u_1^o,u_2^o) \leqslant J_2(u_1^o,u_2) & \forall u_2 \text{ function of t only, } u_2(t) \in U_2 \end{cases}$$

A closed-loop Nash equilibrium (u_1^*,u_2^*) is a pair of functions of t and x satisfying $(u_1^*(t,x),u_2^*(t,x)) \in U_1 \times U_2$ $\forall(t,x)$ such that :

$$(4)\begin{cases} J_1(u_1^*,u_2^*) \leqslant J_1(u_1 u_2^*) & \forall u_1 \text{ closed-loop function, } u_1(t,x) \in U_1 \\ J_2(u_1^*,u_2^*) \leqslant J_2(u_1^*,u_2) & \forall u_2 \text{ closed-loop function, } u_2(t,x) \in U_2 \end{cases}$$

An open-loop equilibrium is not generally a closed-loop one and we generally have $J_1(u_1^o,u_2^o) \neq J_1(u_1^*,u_2^*)$ and $J_2(u_1^o,u_2^o) \neq J_2(u_1^*,u_2^*)$ if both exist.

II.2. <u>The Hamilton-Jacobi system for closed-loop strategies</u> :

If we set :

$$(5)\begin{cases} V_1(t,x) = e^{\alpha t}\int_t^T g_1(s,x(s),u_1^*(s,x(s)),u_2^*(s,x(s)))e^{-\alpha s}\,ds \\ V_2(t,x) = e^{\beta t}\int_t^T g_2(s,x(s),u_1^*(s,x(s)),u_2^*(s,x(s)))e^{-\beta s}\,ds \end{cases}$$

and if V_1 and V_2 are regular enough, it is a well-known result that V_1 and V_2 solve the following Hamilton-Jacobi system of first-order partial differential equation :

$$(6)\begin{cases} \dfrac{\partial V_1}{\partial t} - \alpha V_1 + \underset{u_1 \in U_1}{\text{Min}}\left[\sum_{i=1}^n \dfrac{\partial V_1}{\partial x_i} f_i(t,x,u_1,u_2^*) + g_1(t,x,u_1,u_2^*)\right] = 0 \\ \dfrac{\partial V_2}{\partial t} - \beta V_2 + \underset{u_2 \in U_2}{\text{Min}}\left[\sum_{i=1}^n \dfrac{\partial V_2}{\partial x_i} f_i(t,x,u_1^*,u_2) + g_2(t,x,u_1^*,u_2)\right] = 0 \end{cases}$$

We shall use the classical notations :

$$\frac{\partial V_1}{\partial x_i} = p_i \ , \ \frac{\partial V_1}{\partial t} = p_{n+1} \ , \ \frac{\partial V_2}{\partial x_i} = q_i \ , \ \frac{\partial V_2}{\partial t} = q_{n+1} \ , \ i=1, \quad , n.$$

and (6) can be rewritten as follows :

$$(7) \quad \begin{cases} dV_1 = p_1 dx_1 + \ .. \ + p_n dx_n + p_{n+1} dt \\ dV_2 = q_1 dx_1 + \quad + q_n dx_n + q_{n+1} dt \\ p_{n+1} - \alpha V_1 + \underset{u_1 \in U_1}{\text{Min}} \left[p_i f_i(u_1, u_2^*) + g_1(u_1, u_2^*) \right] = 0 \\ q_{n+1} - \beta V_2 + \underset{u_2 \in U_2}{\text{Min}} \left[q_i f_i(u_1^*, u_2) + g_2(u_1^*, u_2) \right] = 0 \end{cases}$$

Thus, if we perform the two minimization problems of (7) it appears that, in general, u_1^* and u_2^* are functions of (t,x,p,q), that is :

$$(8) \qquad u_1^*(t,x,p,q) \ , \ u_2^*(t,x,p,q).$$

Useful exceptions appear when, for example u_1^* and u_2^* are "bang-bang", or, more generally, satisfy :

$$(9) \quad \begin{cases} u_1^*(t,x,p,q) = \tilde{u}_1(t,x,k) \\ u_2^*(t,x,p,q) = \tilde{u}_2(t,x,k) \end{cases} \quad \text{if } (p,q) \in P_k \times Q_k \quad \forall k \in N$$

that is u_1^* , u_2^* are constant with respect to (p,q) in given open sets $P_k \times Q_k$

Situation (9) generally holds when the constraints are saturated and allow the elimination of (p,q) (for an example see [4]) Then, the classical method of [3],[6] give a local solution of (7)

On the other hand, when the constraints are degenerated, or when the equilibrium (7) is attained in an interrior point of $U_1 \times U_2$, the pair (u_1^*, u_2^*) takes the general form (8). We shall now investigate a method to compute a local solution of (7) in this case. We shall also assume that u_1^* and u_2^* are uniquely defined by (7) and piecewise regular in all their arguments. In the sequel, it will be implicitly assumed that we restrict our analysis to regular points for (u_1^*, u_2^*).

It must be remarked that the informational structure induced by strategies of the form (8) is <u>not a priori included in the closed-loop structure</u>. Furthermore, since p and q summarize the future in the point of view of players 1 and 2 respectively, the strategies (8) will be called <u>feedforward strategies</u> and the cooresponding information structure, the feedforward structure.

III - THE CHARACTERISTIC EQUATIONS

Let us denote :

$$(10)\begin{cases} x_{n+1}=t,\ x=(x_1,x_2,\ldots,x_n,x_{n+1}),\ \hat{p}=(p_1,\ldots,p_n),\ \hat{q},\ (q_1,\ldots,q_n), \\ \tilde{f}_i(x,\hat{p},\hat{q})=f_i(x,u_1^*(x,\hat{p},\hat{q}),u_2^*(x,\hat{p},\hat{q})),\ i=1,\ldots,n, \\ \tilde{g}_i(x,\hat{p},\hat{q})=g_i(x,u_1^*(x,\hat{p},\hat{q}),u_2^*(x,\hat{p},\hat{q})),\ i=1,2. \end{cases}$$

It is remarkable that u_1^* and u_2^* do not depend on $p_{n+1}=\frac{\partial V_1}{\partial t}$ and $q_{n-1}=\frac{\partial V_2}{\partial t}$, hence the notation $u_i^*(x,\hat{p},\hat{q})$.

Then (6) becomes :

$$(11)\begin{cases} -\alpha V_1 + \sum_{i=1}^{n} \frac{\partial V_1}{\partial x_i} \tilde{f}_i(x,\frac{\partial V_1}{\partial x},\frac{\partial V_2}{\partial x}) + \frac{\partial V_1}{\partial x_{n+1}} + \tilde{g}_1(x,\frac{\partial V_1}{\partial x},\frac{\partial V_2}{\partial x}) = 0 \\ -\beta V_2 + \sum_{i=1}^{n} \frac{\partial V_2}{\partial x_i} \tilde{f}_i(x,\frac{\partial V_1}{\partial x}\frac{\partial V_2}{\partial x}) + \frac{\partial V_2}{\partial x_{n+1}} + \tilde{g}_2(x,\frac{\partial V_1}{\partial x},\frac{\partial V_2}{\partial x}) = 0 \end{cases}$$

Theorem : If the pair (u_1^*, u_2^*) is uniquely defined by (7) and sufficiently differentiable in a neighborhood of a regular point $(\bar{x}, \bar{p}, \bar{q})$, then one can find a local solution to (11) satisfying :

$$\frac{\partial V_1}{\partial x_i}(\bar{x}) = \bar{p}_i \ , \quad \frac{\partial V_2}{\partial x_i}(\bar{x}) = \bar{q}_i \ , \quad i=1,\dots,n+1 .$$

Furthermore, this local solution depends on $n(n+1)$ arbitrary functions, and its characteristic system is given by (12), (13):

$$(12) \quad \begin{cases} \sum_{k=1}^{n+1} \left(\frac{\partial H_1}{\partial p_k} \varphi_{i,k} + \frac{\partial H_1}{\partial q_k} \psi_{i,k} \right) = \alpha p_i - \frac{\partial H_1}{\partial x_i} \ , \quad i=1,\dots,n+1 \\ \sum_{k=1}^{n+1} \left(\frac{\partial H_2}{\partial p_k} \varphi_{i,k} + \frac{\partial H_2}{\partial q_k} \psi_{i,k} \right) = \beta q_i - \frac{\partial H_2}{\partial x_i} \ , \quad i=1,\dots,n+1 \end{cases}$$

with $$H_1 = \sum_{i=1}^{n} p_i \tilde{f}_i + \tilde{g}_1 + p_{n+1}, \quad H_2 = \sum_{i=1}^{n} q_i \tilde{f}_i + \tilde{g}_2 + q_{n+1},$$

and :

$$(13) \quad \begin{cases} \dot{p}_i = \sum_{j=1}^{n} \varphi_{i,j} \tilde{f}_j + \varphi_{i,n+1} \ , \quad i=1,\dots,n+1 \\ \dot{q}_i = \sum_{j=1}^{n} \psi_{i,j} \tilde{f}_j + \psi_{i,n+1} \ , \quad i=1,\dots,n+1 \ , \\ \dot{x}_i = \tilde{f}_i(x, \hat{p}, \hat{q}) \quad i=1,\dots,n \\ \dot{x}_{n+1} = 1 \end{cases}$$

Proof : Deriving (11) with respect to each x_i, $i=1,\dots,n+1$, we obtain :

$$(14)\quad \begin{cases} -\alpha \dfrac{\partial V_1}{\partial x_i} + \displaystyle\sum_{j=1}^{n} \dfrac{\partial^2 V_1}{\partial x_i \partial x_j} \tilde{f}_j + \dfrac{\partial^2 V_1}{\partial x_i \partial x_{n+1}} + \sum_{j,k=1}^{n} \dfrac{\partial V_1}{\partial x_j}\left(\dfrac{\partial \tilde{f}_j}{\partial x_i} + \dfrac{\partial \tilde{f}_j}{\partial p_k}\dfrac{\partial^2 V_1}{\partial x_k \partial x_i} + \dfrac{\partial \tilde{f}_j}{\partial q_k}\dfrac{\partial^2 V_2}{\partial x_k \partial x_i}\right) \\ \qquad + \dfrac{\partial \tilde{g}_1}{\partial x_i} + \displaystyle\sum_{k=1}^{n}\left(\dfrac{\partial \tilde{g}_1}{\partial p_k}\dfrac{\partial^2 V_1}{\partial x_k \partial x_i} + \dfrac{\partial \tilde{g}_1}{\partial q_k}\dfrac{\partial^2 V_2}{\partial x_k \partial x_i}\right) = 0, \quad i=1,\ldots,n+1 \\ -\beta \dfrac{\partial V_2}{\partial x_i} + \displaystyle\sum_{j=1}^{n} \dfrac{\partial^2 V_2}{\partial x_i \partial x_j} \tilde{f}_j + \dfrac{\partial^2 V_2}{\partial x_i \partial x_{n+1}} + \sum_{j,k=1}^{n} \dfrac{\partial V_2}{\partial x_j}\left(\dfrac{\partial \tilde{f}_j}{\partial x_i} + \dfrac{\partial \tilde{f}_j}{\partial p_k}\dfrac{\partial^2 V_1}{\partial x_k \partial x_i} + \dfrac{\partial \tilde{f}_j}{\partial q_k}\dfrac{\partial^2 V_2}{\partial x_k \partial x_i}\right) \\ \qquad + \dfrac{\partial \tilde{g}_2}{\partial x_i} + \displaystyle\sum_{k=1}^{n}\left(\dfrac{\partial \tilde{g}_2}{\partial p_k}\dfrac{\partial^2 V_1}{\partial x_k \partial x_i} + \dfrac{\partial \tilde{g}_2}{\partial q_k}\dfrac{\partial^2 V_2}{\partial x_k \partial x_i}\right) = 0, \quad i=1,\ldots,n+1. \end{cases}$$

and, with :

$$H_1 = \sum_{i=1}^{n} p_i \tilde{f}_i + \tilde{g}_1 + p_{n+1}\ , \quad H_2 = \sum_{i=1}^{n} q_i \tilde{f}_i + \tilde{g}_2 + q_{n+1}$$

(14) becomes :

$$(15)\quad \begin{cases} \displaystyle\sum_{k=1}^{n+1}\left(\dfrac{\partial H_1}{\partial p_k}\dfrac{\partial^2 V_1}{\partial x_k \partial x_i} + \dfrac{\partial H_1}{\partial q_k}\dfrac{\partial^2 V_2}{\partial x_k \partial x_i}\right) = \alpha p_i - \dfrac{\partial H_1}{\partial x_i}\ , \quad i=1,\ldots,n+1 \\ \displaystyle\sum_{k=1}^{n+1}\left(\dfrac{\partial H_2}{\partial p_k}\dfrac{\partial^2 V_1}{\partial x_k \partial x_i} + \dfrac{\partial H_2}{\partial q_k}\dfrac{\partial^2 V_2}{\partial x_k \partial x_i}\right) = \beta q_i - \dfrac{\partial H_2}{\partial x_i}\ , \quad i=1,\ldots,n+1 \end{cases}$$

which is precisely (12) with $\dfrac{\partial^2 V_1}{\partial x_i \partial x_j} = \varphi_{i,j}$, $\dfrac{\partial^2 V_2}{\partial x_i \partial x_j} = \psi_{i,j}$.

But (15) is a linear system of rank $2(n+1)$ (see Remark 2) in the $(n+1)(n+2)$ variables : $\{\varphi_{i,j}\ ,\ \psi_{i,j}\ ,\ 1 \leqslant i \leqslant j \leqslant n+1\}$. Thus $(n+1)(n+2) - 2(n+1) = n(n+1)$ functions among the $\{\varphi_{i,j}\ ,\ \psi_{i,j}\}$ remain arbitrary. Finally, using :

$$(16)\quad \frac{d}{dt}\left(\frac{\partial V_i}{\partial x_j}\right) = \sum_{k=1}^{n} \frac{\partial^2 V_i}{\partial x_j \partial x_k} \tilde{f}_k + \frac{\partial^2 V_i}{\partial x_j \partial x_{n+1}}\ , \quad i=1,2,\ j=1,\ldots,n+1,$$

we obtain (13).

Conversely, suppose that $\{\varphi_{i,j}, \psi_{i,j},\ 1 \leq i \leq j \leq n+1\}$ are smooth functions with respect to the variables (x,p,q) in a neighborhood of a given point $(\bar{x},\bar{p},\bar{q})$, and satisfy (12). Then, p_i and q_i, $i=1,\ldots,n+1$, can be obtained by solving (13). Furthermore, using the classical method of Cauchy, p_i and q_i can be obtained as functions of $(x_1,\ldots,x_{n+1})$ only, $i=1,\ldots,n+1$, and one can denote:

$$(17)\quad \begin{cases} \varphi_{i,j}(x,p(x),q(x)) = \dfrac{\partial p_i}{\partial x_j}(x), \quad \psi_{i,j}(x,p(x),q(x)) = \dfrac{\partial q_i}{\partial x_j}(x), \quad i,j=1,\ldots,n+1, \\ \bar{f}_i(x) = \tilde{f}_i(x,\hat{p}(x),\hat{q}(x)), \quad i=1,\ldots,n, \\ \bar{g}_i(x) = \tilde{g}_i(x,\hat{p}(x),\hat{q}(x)), \quad i=1,2. \end{cases}$$

Consequently, we have :

$$(18)\quad \frac{\partial \bar{f}_i}{\partial x_j} = \frac{\partial \tilde{f}_i}{\partial x_j} + \sum_{k=1}^{n} \left(\frac{\partial \tilde{f}_i}{\partial p_k}\frac{\partial p_k}{\partial x_j} + \frac{\partial \tilde{f}_i}{\partial q_k}\frac{\partial q_k}{\partial x_j}\right), \quad i=1,\ldots,n\ ;\ j=1,\ldots,n+1,$$

and similarly for $\dfrac{\partial \bar{g}_1}{\partial x_j}$, $\dfrac{\partial \bar{g}_2}{\partial x_j}$, $j=1,\ldots,n+1$.

Thus (12), (13) become :

$$(19)\quad \begin{cases} \left(\sum_{k=1}^{n} \tilde{f}_k \varphi_{i,k} + \varphi_{i,n+1}\right) - \alpha p_i + \sum_{j=1}^{n} p_j \left(\dfrac{\partial \tilde{f}_j}{\partial x_i} + \sum_{k=1}^{n} \left(\dfrac{\partial \tilde{f}_j}{\partial p_k}\varphi_{i,k} + \dfrac{\partial \tilde{f}_j}{\partial q_k}\psi_{i,k}\right)\right) \\ \qquad + \dfrac{\partial \tilde{g}_1}{\partial x_i} + \sum_{k=1}^{n} \left(\dfrac{\partial \tilde{g}_1}{\partial p_k}\varphi_{i,k} + \dfrac{\partial \tilde{g}_1}{\partial q_k}\psi_{i,k}\right) = 0\ , \quad i=1,\ldots,n+1 \\ \left(\sum_{k=1}^{n} \tilde{f}_k \psi_{i,k} + \psi_{i,n+1}\right) - \beta q_i + \sum_{j=1}^{n} q_j \left(\dfrac{\partial \tilde{f}_j}{\partial x_i} + \sum_{k=1}^{n} \left(\dfrac{\partial \tilde{f}_j}{\partial p_k}\varphi_{i,k} + \dfrac{\partial \tilde{f}_j}{\partial q_k}\psi_{i,k}\right)\right) \\ \qquad + \dfrac{\partial \tilde{g}_2}{\partial x_i} + \sum_{k=1}^{n} \left(\dfrac{\partial \tilde{g}_2}{\partial p_k}\varphi_{i,k} + \dfrac{\partial \tilde{g}_2}{\partial q_k}\psi_{i,k}\right) = 0\ , \quad i=1,\ldots,n+1 \end{cases}$$

or, with (17) and (18) :

$$(20)\qquad \begin{cases} \dot{p}_i - \alpha p_i + \sum_{j=1}^{n} p_j \dfrac{\partial \bar{f}_j}{\partial x_i} + \dfrac{\partial \bar{g}_1}{\partial x_i} = 0 \ , \ i=1,\dots,n+1 \\ \dot{q}_i - \beta q_i + \sum_{j=1}^{n} q_j \dfrac{\partial \bar{f}_j}{\partial x_i} + \dfrac{\partial \bar{g}_2}{\partial x_i} = 0 \ , \ i=1,\dots,n+1 \end{cases}$$

which is the characteristic system of the pair of equations :

$$(21)\qquad \begin{cases} -\alpha V_1 + \sum_{i=1}^{n} \dfrac{\partial V_1}{\partial x_i} \bar{f}_i + \dfrac{\partial V_1}{\partial t} + \bar{g}_1 = 0 \\ -\beta V_2 + \sum_{i=1}^{n} \dfrac{\partial V_2}{\partial x_i} \bar{f}_i + \dfrac{\partial V_2}{\partial t} + \bar{g}_2 = 0 \ . \end{cases}$$

Thus at a regular point, $p_i = \dfrac{\partial V_1}{\partial x_i}$, $q_i = \dfrac{\partial V_2}{\partial x_i}$, and with (17), (11) is solved, which achieves to proves that (12), (13) is a characteristic system of (11). ■

Remark 1 : when u_i^* do not depend on p and q, the general system (12), (13) gives :

$$\begin{cases} \sum_{k=1}^{n} \tilde{f}_k \varphi_{i,k} + \varphi_{i,n+1} = \alpha p_i - \dfrac{\partial H_1}{\partial x_i} \ , \quad \dot{p}_i = \sum_{k=1}^{n} \tilde{f}_k \varphi_{i,k} + \varphi_{i,n+1} \\ \sum_{k=1}^{n} \tilde{f}_k \psi_{i,k} + \psi_{i,n+1} = \beta q_i - \dfrac{\partial H_2}{\partial x_i} \ , \quad \dot{q}_i = \sum_{k=1}^{n} \tilde{f}_k \psi_{i,k} + \psi_{i,n+1} \end{cases}$$

or :

$$(22)\qquad \dot{p}_i = \alpha p_i - \frac{\partial H_1}{\partial x_i} \ , \quad \dot{q}_i = \beta q_i - \frac{\partial H_2}{\partial x_i} \ , \quad i=1,\dots,n+1$$

which is precisely the classical characteristic system of the closed-loop equilibrium.

As a result, (22) defines a unique local solution for suitable initial conditions, whereas the presence of $\dfrac{\partial u_i^*}{\partial p_k}$, $\dfrac{\partial u_i^*}{\partial q_k}$ in (12), (13) yield an indetermination of $n(n+1)$ functions among the $\dfrac{\partial^2 V_i}{\partial x_j \partial x_k}$. ■

Remark 2 : the system (15) has always rank $2(n+1)$ since it can be written :

$$\begin{cases} \varphi_{i,n+1} + \sum_{k=1}^{n} \left(\left(f_k + \sum_{j=1}^{n} p_j \frac{\partial \tilde{f}_j}{\partial p_k} + \frac{\partial \tilde{g}_1}{\partial p_k} \right) \varphi_{i,k} + \left(\sum_{j=1}^{n} p_j \frac{\partial \tilde{f}_j}{\partial q_k} + \frac{\partial \tilde{g}_1}{\partial q_k} \right) \psi_{i,k} \right) \\ \qquad = \alpha p_i - \frac{\partial H_1}{\partial x_i} \quad , \quad i=1,\ldots,n+1. \\ \psi_{i,n+1} + \sum_{k=1}^{n} \left(\left(\sum_{j=1}^{n} q_j \frac{\partial \tilde{f}_j}{\partial p_k} + \frac{\partial \tilde{g}_2}{\partial p_k} \right) \varphi_{i,k} + \left(f_k + \sum_{j=1}^{n} q_j \frac{\partial \tilde{f}_j}{\partial q_k} + \frac{\partial \tilde{g}_2}{\partial q_k} \right) \psi_{i,k} \right) \\ \qquad = \beta q_i - \frac{\partial H_2}{\partial x_i} \quad , \quad i=1,\ldots,n+1 \end{cases}$$

and thus, the submatrix corresponding to the vector :

$(\varphi_{1,n+1},\ldots, \varphi_{n+1,n+1}, \psi_{1,n+1},\ldots, \psi_{n+1,n+1})'$ (prime denoting transposition), is the identity of $R^{2(n+1)\times 2(n+1)}$, which proves the assertion. ■

III - EXAMPLE

We shall try to compare the open-loop, closed-loop and feedforward solutions on a very simple example :

$$\begin{cases} \dot{x} = u + v \ , \quad J_1(u,v) = x^2(1) + \frac{1}{2}\int_0^1 (x^2 + u^2)dt, \\ J_2(u,v) = x^2(1) + \frac{1}{2}\int_0^1 (\alpha x^2 + \beta u^2 + v^2)dt, \quad \alpha > 0, \quad \beta > 0. \end{cases}$$

1. The open-loop equilibrium

The Hamiltonians are :

(1.1) $\qquad H_1 = p(u+v) + \frac{1}{2}(x^2+u^2), \quad H_2 = q(u+v) + \frac{1}{2}(\alpha x^2+\beta u^2+\sigma^2)$

and the unique Nash point of (F_1, F_2) is :

$$u^o = -p, \quad v^o = -q \tag{1.2}$$

The adjoint equations for open-loop equilibrium are :

$$\begin{cases} \dot{x} = -(p+q) & x(o) = x_o \\ \dot{p} = -x & p(1) = q(1) = 2x(1) \\ \dot{q} = -\alpha x \end{cases} \tag{1.3}$$

Thus, the Nash equilibrium is unique and given by :

$$\begin{cases} x(t) = A\, e^{t\sqrt{1+\alpha}} + B\, e^{-t\sqrt{1+\alpha}} \\ p(t) = \frac{1}{\sqrt{1+\alpha}} \left(C - A\, e^{t\sqrt{1+\alpha}} + B\, e^{-t\sqrt{1+\alpha}}\right) = -u^o(t) \\ q(t) = \frac{1}{\sqrt{1+\alpha}} \left(-C - \alpha A\, e^{t\sqrt{1+\alpha}} + \alpha B\, e^{-t\sqrt{1+\alpha}}\right) = -v^o(t) \end{cases} \tag{1.4}$$

with :

$$\begin{cases} A = -x_o \dfrac{e^{-\sqrt{1+\alpha}}\,(4-\sqrt{1+\alpha})}{(4+\sqrt{1+\alpha})e^{\sqrt{1+\alpha}} - (4-\sqrt{1+\alpha})e^{-\sqrt{1+\alpha}}}, \quad B = x_o - A \\ C = \dfrac{4x_o(\alpha-1)}{(4+\sqrt{1+\alpha})e^{\sqrt{1+\alpha}} - (4-\sqrt{1+\alpha})e^{-\sqrt{1+\alpha}}} \end{cases} \tag{1.5}$$

2. <u>The closed-loop equilibrium</u> :

Now, (1.1) and (1.2) still hold, but we look for u* and v* in the form :

$$u^* = -Px - R \quad , \quad v^* = -Qx - S \tag{2.1}$$

with $p = Px + R$, $q = Qx + S$.

Thus, replacing (2.1) in (1.1) :

$$H_1^* = -p((P+Q)x+R+S) + \frac{1}{2}(x^2+P^2x^2+2PRx+R^2)$$

(2.2)

$$H_2^* = -q((P+Q)x+R+S) + \frac{1}{2}(\alpha x^2+\beta(P^2x^2+2PRx+R^2) + Q^2x^2 + 2QSx + S^2)$$

and, since $\dot{p} = -\frac{\partial H_1^*}{\partial x}$, $\dot{q} = -\frac{\partial H_2^*}{\partial x}$, we obtain :

$$\dot{p} = p(P+Q) - (1+P^2)x - PR$$

(2.3)

$$\dot{q} = q(P+Q) - (\alpha+\beta P^2+Q^2)x - \beta PR - QS$$

Thus, with $p = Px + R$, $q = Qx + S$, it results that :

(2.4)
$$\begin{cases} \dot{P} = Q^2 + 2PQ - 1, & P(1) = 2 \\ \dot{Q} = Q^2 + 2PQ - \beta P^2 - \alpha, & Q(1) = 2 \\ \dot{R} = (P+Q)R + PS, & R(1) = 0 \\ \dot{S} = (P+Q)S + (Q-\beta P)R, & S(1) = 0 \end{cases}$$

Obiously, $R = S \equiv 0$ and the closed-loop equilibrium is finally given by :

(2.5)
$$\begin{cases} \dot{x} = -(P+Q)x, & x(0) = x_o \\ \dot{P} = P^2 + 2PQ - 1, & P(1) = 2 \\ \dot{Q} = Q^2 + 2PQ - \beta P^2 - \alpha, & Q(1) = 2 \end{cases}$$

But it is easy to check that (p,q) solution of (1.4) satisfy :

$p(t) = P^o(t)x(t)$, $q(t) = Q^o(t)x(t)$ with P^o and Q^o solution of :

$$\begin{cases} \dot{P}^o = (P^o)^2 + 2P^oQ^o - 1, & P^o(1) = 2 \\ \dot{Q}^o = (Q^o)^2 + 2P^oQ^o - \alpha, & Q^o(1) = 2 \end{cases}$$

and thus (P^o, Q^o) cannot be solution of (2.5) if $\beta \neq 0$, which proves that (since β is assumed > 0) the closed-loop equilibrium does not coincide with the open-loop one.

3. The feedforward equilibrium :

Going back to (1.1), (1.2) and noting $\hat{u}_i$ and $\hat{H}_i$ the optimal feedforward strategies and Hamiltonians for i=1,2, we have :

$$(3.1) \qquad \begin{cases} \hat{u}_1 = -p \quad , \quad \hat{u}_2 = -q \\ \hat{H}_1 = -p(p+q) + \frac{1}{2} x^2 + \frac{1}{2} p^2 \\ \hat{H}_2 = -s(p+q) + \frac{\alpha}{2} x^2 + \frac{\beta}{2} p^2 + \frac{1}{2} q^2 \end{cases}$$

Thus :

$$\begin{cases} \frac{\partial \hat{H}_1}{\partial p} = -(p+q), \ \frac{\partial \hat{H}_1}{\partial q} = -p, \ \frac{\partial \hat{H}_1}{\partial x} = x \\ \frac{\partial \hat{H}_2}{\partial p} = \beta p - q \ , \ \frac{\partial \hat{H}_2}{\partial q} = -(p+q), \ \frac{\partial \hat{H}_2}{\partial x} = \alpha x \end{cases}$$

Now, looking for the solution for which $\frac{\partial^2 V_1}{\partial x \partial t} = \frac{\partial^2 V_2}{\partial x \partial t} \equiv 0$ (2 functions being arbitrary since n = 1), we have :

$$\begin{cases} \frac{\partial^2 V_1}{\partial t^2} = \frac{\partial^2 V_2}{\partial t^2} = 0 \\ \begin{pmatrix} -(p+q) & -p \\ \beta p - q & -(p+q) \end{pmatrix} \begin{pmatrix} \frac{\partial^2 V_1}{\partial x^2} \\ \frac{\partial^2 V_2}{\partial x^2} \end{pmatrix} = \begin{pmatrix} -x \\ -\alpha x \end{pmatrix} \end{cases}$$

or :

$$(3.2) \qquad \begin{pmatrix} \dot{p} \\ \dot{q} \end{pmatrix} = \begin{pmatrix} -(p+q) & -p \\ \beta p - q & -(p+q) \end{pmatrix}^{-1} \begin{pmatrix} (p+q)x \\ (p+q)\alpha x \end{pmatrix}$$

or also :

$$(3.3)\begin{cases} \dot{p} = -\dfrac{(p+q)((1-\alpha)p+q)}{(p+q)^2+p(\beta p-q)}\,x\ ,\ \dot{q} = -\dfrac{(p+q)((\alpha+\beta)p-(1-\alpha)q)}{(p+q)^2+p(\beta p-q)}\,x\ ,\ \dot{x} = -(p+q) \\ p(1) = q(1) = 2x(1),\ x(o) = x_o \end{cases}$$

In order to compare the solution of (3.3) to the closed-loop equilibrium or to the open-loop one, let us try to solve (3.3) under the forme $p = \hat{P}x$, $q = \hat{Q}x$, which is necessary to have the coincidence. $\hat{P}$ and $\hat{Q}$ must be solution of :

$$(3.4)\begin{cases} \dot{\hat{P}} - \hat{P}(\hat{P}+\hat{Q}) = -\dfrac{(\hat{P}+\hat{Q})^2-\alpha\hat{P}(\hat{P}+\hat{Q})}{(\hat{P}+\hat{Q})^2+\hat{P}(\beta\hat{P}-\hat{Q})} = -1+\dfrac{\hat{P}((\beta-\alpha)\hat{P}-(1+\alpha)\hat{Q})}{(\hat{P}+\hat{Q})^2+\hat{P}(\beta\hat{P}-\hat{Q})} \\ \dot{\hat{Q}} - \hat{Q}(\hat{P}+\hat{Q}) = -\dfrac{\alpha(\hat{P}+\hat{Q})^2+(\beta\hat{P}-\hat{Q})(\hat{P}+\hat{Q})}{(\hat{P}+\hat{Q})^2+\hat{P}(\beta\hat{P}-\hat{Q})} = -\dfrac{(\beta\hat{P}-\hat{Q})((1-\alpha)\hat{P}+\hat{Q})}{(\hat{P}+\hat{Q})^2+\hat{P}(\beta\hat{P}-\hat{Q})} - \alpha \end{cases}$$

But if $(\hat{P},\hat{Q})$ solves also (2.5), we must have :

$$(3.5)\quad \dot{\hat{P}} - \hat{P}(\hat{P}+\hat{Q}) = \hat{P}\hat{Q} - 1\ ,\quad \dot{\hat{Q}} - \hat{Q}(\hat{P}+\hat{Q}) = \hat{P}\hat{Q} - \beta\hat{P}^2 - \alpha$$

Thus :

$$(3.6)\quad \hat{P}\hat{Q} = \frac{\hat{P}((\beta-\alpha)\hat{P}-(1+\alpha)\hat{Q})}{(\hat{P}+\hat{Q})^2+\hat{P}(\beta\hat{P}-\hat{Q})} = \frac{\hat{Q}((1-\alpha)\hat{P}+\hat{Q})}{(\hat{P}+\hat{Q})^2+\hat{P}(\beta\hat{P}-\hat{Q})}$$

But also : $(\beta-\alpha)\hat{P}^2 - 2\hat{P}\hat{Q} - \hat{Q}^2 = 0$, or : $\dfrac{\hat{P}}{\hat{Q}} = \dfrac{1 \pm \sqrt{1+\beta-\alpha}}{\beta-\alpha} \overset{\text{def}}{=} \gamma$ if $1+\beta > \alpha$, and we see that for $1+\beta < \alpha$ the feedforward and closed-loop solutions are different. But if $1+\beta > \alpha$, we find that $\dfrac{\hat{P}}{\hat{Q}} = \gamma$ and, using (3.6) we find that $\dot{\hat{P}} = \dot{\hat{Q}} = 0$ and that α and β must satisfy a set of 4 independent relations (if we take into account that $\hat{P}(t) \equiv 2$ and $\hat{Q}(t) \equiv 2$) which is impossible, and thus the <u>feedforward and closed-loop solutions never coincide</u> ($\forall\alpha,\beta > 0$). Using the same method, one can check that also the <u>open-loop solution never coincide with the last two solutions</u> ($\forall\alpha,\beta > 0$).

IV - CONCLUDING REMARKS

We have derived the adjoint equations, for an arbitrary dimension n, solving the Nash equilibrium problem when the optimal stategies u_1^* and u_2^* are obtained as functions of (t,x,p,q). It appears that $\forall n$, these equations depend on (n^2+n) arbitrary functions and this proves that, generically, there are infinitely many surfaces that locally solve the necessary conditions. This remark is complementary to T. Basar's [1] and A. Mehlmann's [5] claim of non uniqueness depending on the information structure. It is finally worth noting that the non-uniqueness disappears when u_1^* and u_2^* are independant of p and q.

REFERENCES

[1] T. BASAR. Informationally Nonunique Equilibrium Solutions in Differential Games. SIAM J Control. 15 (1977) p. 636-660.

[2] A. BENSOUSSAN. Points de Nash dans le Cas de Fonctionnelles Quadratiques et Jeux Differentiels Linéaires à N personnes. SIAM J Control. 12 (1974) p. 460-499.

[3] J. CASE. Towards a Theory of Many-Player Differential Games. SIAM J Control. 7 (1969) p. 179-197.

[4] J. LEVINE - J. THEPOT. Open-Loop and Closed-Loop Equilibria in a Dynamic Duopoly. In : Optimal Control Theory and Economic Analysis. G. Feichtinger ed. p. 143-156. North-Holland. 1982.

[5] A. MEHLMANN. On relations between Open-Loop and Closed-Loop Nash Solutions in Deterministic Differential Games,in Optimal Control Theory and Economic Analysis, G. Feichtinger ed. p. 399-413. North-Holland. 1982.

[6] H. STALFORD - G. LEITMANN. Sufficiency Conditions for Nash Equilibria in N-Person Differential Games. In : Topics in Differential Games. A. Blaquière ed. p. 345-376. North-Holland. 1973.

PARTIE II

Etude des conditions d'optimalité avec information non classique pour les problèmes de contrôle et d'équipe stochastiques.

RESUME DE LA IIème PARTIE

Cette partie est consacrée aux problèmes d'équipe à information non classique, à savoir lorsque les σ-algèbres d'observations ne sont pas croissantes en fonction du temps. Cette situation se présente en particulier lorsque plusieurs décideurs effectuent indépendamment des observations de l'état sans échanger ces observations et/ou lorsque des décideurs ont des capacités de mémorisation limitée, et doivent minimiser une fonction coût commune à l'aide de stratégies n'utilisant que leurs observations.

Ce problème est abordé dans deux cas extrêmes : le premier, lorsque la structure probabiliste des bruits est discrète et le second dans le cas des diffusions.

On y développe une méthode de programmation dynamique où la loi de probabilité des trajectoires du processus joue le rôle d'état. On montre, sous des hypothèses de régularité du coût optimal, que les stratégies optimales sont obtenues par intégration d'une équation d'Hamilton-Jacobi-Bellman généralisée, comportant un terme supplémentaire par rapport à la théorie classique. le terme de "signalling" qui mesure l'utilité marginale d'une variation d'information. Ces résultats sont justifiés sans hypothèse de régularité dans le cas du contrôle des diffusions avec observations partielles et information classique.

NON CLASSICAL INFORMATION AND OPTIMALITY IN CONTINUOUS-TIME DYNAMIC TEAM PROBLEMS

J. LEVINE

Centre d'Automatique et Informatique
de l'Ecole Nationale Supérieure des Mines de Paris
35, Rue Saint-Honoré
F-77305 Fontainebleau - FRANCE

Abstract : This paper is devoted to the study of optimality criteria in dynamic team problems where the decision makers have noisy observations of the state and given maximal memory capacities. Thus the information structure is non classical for two reasons : on the one hand because each decision maker has no access to the observations of the other decision makers, and on the other hand because his memory can be limited.

We develop in the cases with discrete-time noises and for diffusions, a generalized dynamic programming method giving rise to efficient optimality characterizations via generalized Hamiltonians minimizations, Hamiltonians containing a "signalling term". This theory is also applied to the control of partially observed diffusions with classical information and to the famous Witsenhausen's counterexample.

I - INTRODUCTION

I.1. The team problems with non classical information

Non classical information control or team problems have been firstly introduced by H.S. Witsenhausen ([31]) and, since they appear naturally in many examples in communications ([3],[18],[31]), in mathematical economics ([5],[18],[19]), in large scale systems ([3],[4],[5],[10],[17],[18],[19],[25],[30], [31],[32],[33], and more briefly, in decision problems with both estimation and control, they have motivated an important literature.

However, besides particular examples, almost no general results are known concerning existence and optimality characterizations (see [5],[12],[21] and [31] for existence results in particular cases, and [21],[22] for optimality criteria in discrete time problems). The aim of this paper is to obtain optimality criteria for two kinds of team problems with non classical information, the only difference between them lying in the probabilistic structure of the noises : in the first model, the noises are discrete-time processes, whereas in the second they are assumed to be Wiener processes.

In both cases, the team is made of N stations observing independently, with an individual noisy observation function, the state of a system with noise perturbations, and governed by control variables. The informational structure is supposed to be the following : at each time t, the decision maker k has no access to the other decision makers observations and cannot remember his past observations before $\mathcal{K}_k(t)$, $\mathcal{K}_k$ being a given function $\forall k=1,\dots,N$, known by every decision maker, satisfying :

$$(1.1) \qquad 0 \leq \mathcal{K}_k(t) \leq t \quad \forall t \in [0,T] \ , \quad \forall k=1,\dots,N$$

This function $\mathcal{K}_k$ may be interpreted as the maximal memory capacity allocated to decision maker k at time t.

The admissible decisions for each station are thus assumed to be based on these informations, and satisfying eventually additional constraints on the controls. Finally, the N team players wish to minimize a common cost function with respect to their admissible strategies.

In the first case, the system is given by a controlled differential equation, perturbed by piecewise constant noises occuring at the given deterministic dates : $t_o, t_1, \ldots, t_M$, with :

(1.2) $o = t_o < t_1 < \ldots < t_M = T$ (the horizon T is fixed),

the noise realizations in two different intervals being uncorrelated. The observations are also perturbed by piecewise constant noises occuring at the same dates $t_o, \ldots, t_M$, having the same time-uncorrelation property.

In the second case, the system is given by a controlled stochastic differential equation driven by white noise, and the observations are also given by a stochastic differential equation driven by an independent white noise.

This problem generalizes the control problem for partially observed diffusions (see [1],[2],[5],[6],[7],[9],[11],[12],[14],[15],[24],[34],[35]) where the information structure is classical : $N = 1$ and $\mathcal{H}(t) \equiv 0$, thus the controller has perfect memory and the σ-fields of his observations are increasing with time.

For $N > 1$, even if $\mathcal{H}_k(t) \equiv 0 \quad \forall k = 1, \ldots, N$, and except for some rare situations with partially nested information structures (see [10],[17],[18], [19],[30],[33]), the increasing property of the σ-fields of observations vanishes, and the computation of conditional expectations with respect to the past observations becomes untractable. This explains why we shall avoid as much as possible the use of conditional expectations.

On the other hand, our approach generalizes the ones of [21],[22],[31], [32], essentially based on discrete-time properties.

Finally, the reasons to restrict ourselves to the cases of discrete-time incorrelated noises or to Wiener processes are essentially technical : these types of noises are extreme cases since the first is everywhere degenerated (in terms of infinitesimal generator), whereas the second is nowhere degenerated. Thus we hope to convince the reader on the flexibility of our Dynamic Programming approach. On the other hand, it is also clear that the complexity of the analysis increases with the degree of generality of the noise processes,

and for clarity's sake, we have assumed the simplest probabilistic frameworks.

I.2. Orientations of the paper

In order to explain and motivate the orientations of the paper, let us briefly recall the methods used in the precedently cited control for partially observed diffusions, with classical information structure, to obtain optimality conditions via Dynamic Programming techniques. Most of the results in this area are based on the fact that the infimum of the cost function with respect to u, namely the Value function, depends on a conditional probability measure (eventually unnormalized) of the state with respect to the past observations, and this conditional measure satisfies an evolution equation where u appears as a control variable. It results that the conditional measure can be used as the genuine state variable, and the minimization problem can be stated in an equivalent way in terms of this new state variable (see for example [1],[2],[6],[7], [11],[14],[15],[24],[25],[34],[35], The advantage, as clained above, is that the Dynamic Programming equations can be derived in this state space.

In our non classical informational structure, such a method does not work because of the non nestedness of the σ-fields of observations, as was firstly remarked in [31]. However, the same type of approach can be applied in our framework if one replaces the conditional measure by an unconditional probability measure on the space of trajectories, playing here the role of the controlled state variable. This approach has been already developed in particular cases in [21],[22],[31],[32].

Thus, the paper is organized as follows :
In section II, we shall deal with the team problem with discrete-time noises, and in section III, with the team problem for diffusions.
These two sections follow almost the same lines :
Section II : The team problem with discrete-time noises.
II.1. : the model and the assumptions
II.2. : reformulation of the problem in the space of bounded measures, with :
II.2.1. : first reformulation into a purely final cost,
II.2.2. : the value function and the optimality principle,
II.2.3. : the closed-loop formulation

II.3. : some regularity properties of the value function : concavity, subdifferentiability and integral representations. Directional derivative formula .

II.4. : the Hamilton-Jacobi-Bellman equations and the Signalling.

Section III : The team problem for diffusions

III.1. : Model and assumptions

III.2. : The dynamic programming method in the space of bounded measures

III.2.1. Reformulation into a purely final cost

III.2.2. The value function and the optimality principle

III.2.3. The closed-loop formulation

III.3. : Some regularity properties of the value function (concavity, subdifferentiability and integral representation ; directional derivative formula)

III.4. : The Hamilton-Jacobi-Bellman equations and the signalling

III.5. : Application to the control of partially observed diffusions with classical information.

III.5.1. The Hamilton-Jacobi-Bellman theory and the maximum principle.

III.5.2. The link with Mortensen's equation.

Section IV is devoted to examples, the last one being the well-known counterexample of H.S. Witsenhausen [31].

II - THE TEAM PROBLEM WITH DISCRETE-TIME NOISES

II.1. Model and assumptions

Since the measurability properties play a central role to describe the information structure, we shall have to give the precise definitions of the σ-fields of observations and of progressive measurability of a process with respect to a family of σ-fields which is not increasing. These notions will be used to define the admissible strategies. Furthermore, the concept of a solution, or trajectory, of a differential equation with such admissible strategies is not completely standard and deserves a careful introduction. We want to point out that, though heavy, these developments are essential for the sequel.

Finally, the cost function and the minimization problem will be defined with some comments on the public and private informations.

II.1.1. The dynamics and observations equations

The interval $[0,T]$ is subdivided into $t_0,\dots,t_M$ as in (1.2). The system is given by :

$$(2.1)\quad \begin{cases} \dot{x}(t) = f(t,x(t),u(t),v_j) \\ y_k(t) = h_k(t,x(t),v_j^k),\ k=1,\dots,N \end{cases} \qquad \forall t \in [t_j,t_{j+1}[,\ \forall j = 0,\dots,M-1$$

where :

- $f,h_1,\dots,h_N$ are smooth functions of all their arguments, f satisfying the classical linear growth condition and convexity condition :

$$(2.2)\quad \begin{cases} \operatorname{Sup}_{t,u,v} |x.f(t,x,u,v)| \leq C(1+\|x\|^2) \quad \forall x \in R^n \\ f(t,x,U,v) \text{ is convex } \forall (t,x,v),\ U=U_1\times\dots\times U_N \text{ closed subset of } \mathbb{R}^{p_1}\times\dots\times\mathbb{R}^{p_N} \end{cases}$$

- we denote $\omega_j = (v_j,v_j^1,\dots,v_j^N) \in R^r$, and $\omega = \{\omega_j \mid j=0,\dots,M-1\}$ a realization of a noise sequence. We set ρ_j = probability measure on $(R^r,\mathcal{A}_j)$, $j=0,\dots,M-1$. ω_j and ω_k being supposed independent $\forall j \neq k$, the probability measure ρ of ω is given by :

$$(2.3)\qquad \rho = \bigotimes_{j=0}^{M-1} \rho_j \quad \text{on} \quad (\Omega,\mathcal{A}) \quad \text{with} \quad \Omega = R^{rM} \quad \text{and} \quad \mathcal{A} = \prod_{j=0}^{M-1} \mathcal{A}_j .$$

We shall also denote $\mathcal{A}_t = \prod_{j=0}^{j(t)} \mathcal{A}_j$ with $j(t)$ defined by $t \in [t_{j(t)}, t_{j(t)+1}[$.

- The initial state $x(o)$ is supposed to be a random vector independent on ω, with probability measure μ_o on R^n and σ-field $\mathcal{F}_o$. Thus, the natural measurable space of the problem is : $(R^n \times \Omega, \mathcal{F}_o \times \mathcal{A}, \mu_o \otimes \rho)$.
A stochastic process ξ on $[o,T] \times R^n \times \Omega$ is thus said progressively measurable with res ect to $\{\mathcal{F}_o \times \mathcal{A}_t ; t \in [o,T]\}$ if the restriction of ξ to $[o,t]$ is $\mathcal{B}_{[o,t]} \times \mathcal{F}_o \times \mathcal{A}_t$ measurable $\forall t \in [o,T]$, where $\mathcal{B}_{[o,t]}$ is the Borel σ-field of $[o,t]$.

- Let $\mathcal{Y}_k(t)$ be the sub σ-field of $\mathcal{F}_o \times \mathcal{A}_t$ generated by the observation paths :

(2.4) $Y_k(t) = \{y_k(s) = h_k(s,x(s),v_j^k) | K_k(t) \leq s \leq t$, $x : \{\mathcal{F}_o \times \mathcal{A}_t\}$ progressively measurable, and $v_j^k = pr_j^k(\omega)$, $\omega \in \Omega\}$, $k=1,\ldots,N$,

with $pr_j^k(\omega) = (k+1)^{st}$ component of ω_j.
Denote $\mathcal{Y}_k(s,t) = \bigcup_{s \leq \sigma \leq t} \mathcal{Y}_k(\sigma)$, and $Y_k(o,T)$ the space of observation path on $[o,T]$.
Clearly, the family $\{\mathcal{Y}_k(t) ; t \in [o,T]\}$ is not increasing in general.

Definition 2.1 : we shall say that a process u_k on $[o,T] \times Y_k(o,T)$ is progressively measurable with respect to $\{\mathcal{Y}_k(t) ; t \in [o,T]$ if : $u_k(t)$ is $\mathcal{Y}_k(t)$-measurable $\forall t \in [o,T]$, and u_k restricted to $[o,t]$ is $\mathcal{B}_{[o,t]} \times \mathcal{Y}_k(o,t)$-measurable $\forall t \in [o,T]$. ■

Thus, we shall say that u_k is an admissible strategy for decision maker k if u_k is progressively-measurable with respect to $\{\mathcal{Y}_k(t) ; t \in [o,T]\}$ and if :

(2.5) $u_k(t,Y_k(t)) \in U_k$ $\forall t \in [o,T]$, $\forall Y_k(t) \in Y_k(o,T)$, $\forall k=1,\ldots,N$,

The set $\mathcal{U}_k$ of admissible strategies for station k is thus :

(2.6) $\mathcal{U}_k = \{u_k$: $\{\mathcal{Y}_k(t)\}$-progressively measurable process satisfying (2.5)$\}$

We denote $\mathcal{U} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_N$.

. Finally, a solution of (2.1) starting from (o,x) and generated by $u = (u_1,\dots,u_N) \in \mathcal{U}$ and $\omega \in \Omega$, may be defined in the sense of Krassovski (see [20] and the Appendix 1). However, such solutions need not be unique and we must introduce the set $\mathcal{X}^{u,\omega}(o,x)$ of all such solutions (in $C^o([o,T];R^n)$: the space of continuous functions from $[o,T]$ to R^n), issued from (o,x) and generated by (u,ω). $X_t^{u,\omega}(o,x)$ will denote the value at time t of an element of $\mathcal{X}^{u,\omega}(o,x)$ (namely a solution of (2.1) for (o,x) and (u,ω) evaluated at time t). Also $\mathcal{X}^u(o)$ will denote the set of flows of trajectories generated by u from time o, that are $\{\mathcal{F}_o \times \mathcal{A}_t\}$-progressively measurable, namely :

(2.7) $\mathcal{X}^u(o) = \{X^{u,\cdot}(o,.) \,|\, X^{u,\omega}(o,x) \in \mathcal{X}^{u,\omega}(o,x) \quad \forall (x,\omega) \in R^n \times \Omega$, and the process $t \to X_t^{u,\omega}(o,x)$ is $\{\mathcal{F}_o \times \mathcal{A}_t\}$ progressively measurable$\}$

Since $\mathcal{X}^{u,\omega}(o,x)$ is a compact subset of $C^o([o,T];R^n)$ depending upper-semicontinuously on (x,ω) (see [20]), one can prove (see for example [8]) that $\mathcal{X}^u(o) \neq \emptyset \quad \forall u \in \mathcal{U}$.

To specify the dimensionalities, let us set :

$x \in R^n$, $y_k \in R^{q_k}$ $(\sum_{k=1}^N q_k = q)$,

$u_k(Y_k(t)) \in R^{p_k}$ $(\sum_{k=1}^N p_k = p)$, with $Y_k(t)$ piecewise continuous path on $[\mathcal{K}_k(t),t] \times R^{p_k} \quad \forall t, \ \forall k$.

II.1.2. The cost functional

$$(2.8)\quad J_{o,\mu_o}(u_1,\dots,u_N) = \sup_{X^u \in \mathcal{X}^u(o)} E\Big[\int_o^T g(t,X_t^{u,\omega}(o,x),u(t,Y_t^{u,\omega}(o,x)))dt + G(X_T^{u,\omega}(o,x))\Big]$$

The expectation being taken with respect to $\mu_o \otimes \rho$, with :

- g and G smooth and bounded functions,
- the notation $u(t,Y_t^{u,\omega}(o,x))$ meaning precisely : $u(t,Y_t^{u,\omega}(o,x)) = (u_1(t,Y_{1,t}^{u,\omega}(o,x)),\dots,u_N(t,Y_{N,t}^{u,\omega}(o,x)))$, and :

$$(2.9)\quad Y_{k,t}^{u,\omega}(o,x) = \{(s,y_k(s)) \,|\, y_k(s) = h_k(s,X_s^{u,\omega}(o,x),\nu_j^k) \text{ with } \nu_j^k = pr_j^k(\omega), \quad \forall s \in [t_j,t_{j+1}[\cap [\mathcal{K}_k(t),t], \ \forall j=o,\dots,M-1\}$$

for $k = 1,\ldots,N$, $pr_j^k(\omega)$ being the $(k+1)^{st}$ component of ω ;

II.1.3. Statement of the problem

(2.10) Minimize J_{o,μ_o} (given by (2.8)) with respect to $u = (u_1,\ldots,u_N)$ in $\mathcal{U} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_N$ (given by (2.6)), and for trajectories given by (2.1)

Remark that, because of the non-uniqueness of the trajectories of (2.1), the team is interested in a minimum guaranteed cost obtained via the supremum for every possible flow generated from time o by each strategy $u \in \mathcal{U}$. Of course, if for a given $u \in \mathcal{U}$, (2.1) admits a unique solution, then we recover the usual definition of $J_{o,\mu_o}(u)$.

II.1.4. The public and private informations

To summarize, the public informations are :
the sequence $\{t_j \mid j=o,\ldots,M\}$, the functions h_k, H_k $(k=1,\ldots,N)$, f,g,G, the σ-fields $\mathcal{Y}_k(t)$ $\forall t \in [o,T]$, $\forall k=1,\ldots,N$, the probability space $(R^n \times \Omega, \mathcal{F}_o \times \mathcal{A}, \mu_o \otimes \rho)$ and the family of increasing sub σ-fields $\{\mathcal{F}_o \times \mathcal{A}_t \; ; \; t \in [o,T]\}$, the system (2.1), $\mathcal{U} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_N$ (and thus the form of the controls), $J_{o,\mu_o}(u)$ $\forall u \in \mathcal{U}$.
Each team player has the same clock and agrees on the concept of solution for (2.1).

On the other hand, player k's private information is, at every time t, a path of the form (2.9).

Finally, with this formalism, the fact that the information structure is non classical may be written :

(2.11) neither $(\forall j, k \in \{1,\ldots,N\}\ \mathcal{Y}_j(t) \subset \mathcal{Y}_k(t)\ \forall t \in [o,T])$,
nor $(\mathcal{Y}_k(t) \subset \mathcal{Y}_k(s)\ \forall t \leq s,\ \forall k = 1,\ldots,N)$, hold true in general.

II.2. Reformulation in the space of bounded measures

The aim of this paragraph is to prove that the team problem (2.10) can be reformulated in terms of bounded measures on the space of trajectories of (2.1), and that the infimum of J_{o,μ_o}, the Value function, satisfies the well-known transition (or optimality) principle of Dynamic Programming with respect to these measures. This reformulation will be done in three steps :

- a first reformulation into a purely final cost where the dependence of the cost function on the measure of the final state is linear,

- the definition of the Value function for any initial time $t \in [o,T]$ where we prove the desired transition principles (on the value function and on the measures),

- the final reformulation for closed-loop strategies (depending on the measures) and the comparison result between the optimal values with respect to closed-loop strategies and the current strategies defined in (2.6).

II.2.1. A first reformulation into a purely final cost

The most natural way to show how measures appear as new state variables, consists in making the classical change of variables to have a purely final cost :

Let us introduce a new scalar variable x_{n+1} and the differential equation :

$$(2.12)\quad \begin{cases} \dot{x}_{n+1}(t) = g(t, X_t^{u,\omega}(o,x), u(Y_t^{u,\omega}(o,x))) & \forall t \in [o,T] \\ x_{n+1}(o) = o \end{cases}$$

with $X^{u,\omega}(o,x)$ satisfying (2.1) (in the generalized sense of Krassovski), and let us denote :

$$(2.13)\quad \begin{cases} z = (x, x_{n+1})' \\ F(t,z,u,v) = (f(t,x,u,v)' \ , \ g(t,x,u))' \ , \ \forall (t,z,u,v) \\ H_k(t,z,\nu^k) = h_k(t,x,\nu^k) \ , \ \forall (t,z,\nu^k) \ , \ \forall k = 1,\dots,N \ , \\ \tilde{G}(z) = G(x) + x_{n+1} \ , \ \forall z \end{cases}$$

Thus, the system (2.1), (2.12) becomes :

$$(2.14)\quad \begin{cases} \frac{d}{dt} Z_t^{u,\omega}(o,z) = F(t, Z_t^{u,\omega}(o,z), u(t, Y_t^{u,\omega}(o,z)), v_j) \quad \forall t \in [t_j, t_{j+1}[\\ Z_o^{u,\omega}(o,z) = z = (x,o)' \end{cases}$$

with $u(t, Y_t^{u,\omega}(o,z)) = (u_1(t, Y_{1,t}^{u,\omega}(o,z)), \dots, u_N(t, Y_{N,\tau}^{u,\omega}(o,z)))$ and :

$$Y_{k,t}^{u,\omega}(o,z) = \{(s, y_k(s)) | y_k(s) = H_k(s, Z_s^{u,\omega}(o,z), \nu_j^k), \ \nu_j^k = pr_j^k(\omega), \\ \forall s \in [t_j, t_{j+1}[\cap [\mathcal{H}_k(t), t]\} \quad \forall k = 1,\dots,N.$$

Thus, since $x_{n+1}(t) = \int_o^t g(s, X_s^{u,\omega}(o,x), u(s, Y_s^{u,\omega}(o,x)))ds$

and since $\tilde{G}(z) = G(x) + x_{n+1}$, we obtain the :

Proposition 2.1

$$(2.15)\quad J_{o,\mu_o}(u) = \sup_{Z^u \in \mathcal{Z}^u(o)} \int_{\Omega \times R^{n+1}} \tilde{G}(Z_T^{u,\omega}(o,z)) dP_o(z) d\rho(\omega) \stackrel{def}{=} J_{o,P_o}(u)$$

with $\mathcal{Z}^u(o)$ defined exactly as in (2.7) with Z^u and z in place of X^u and x, and with $P_o = \mu_o \otimes \delta_o$, namely :

$$(2.16)\quad \int_{R^{n+1}} \varphi(z) dP_o(z) = \int_{R^{n+1}} \varphi(x, x_{n+1}) d\mu_o(x) d\delta_o(x_{n+1}) = \int_{R^n} \varphi(x,o) d\mu_o(x)$$

$\forall \varphi \in C_b^o(R^{n+1})$ (space of continuous bounded functions on R^{n+1})

Proof : It suffices to prove that $\mathcal{Z}^u(o) = \{(X^u, x_{n+1}(X^u)) | X^u \in \mathcal{X}^u(o)\}$ with $x_{n+1}(X^u)$ the solution of (2.12) associated to X^u ; and this is straightforward since $x_{n+1}(X^u)$ is simply the integral between 0 and T

of a function independent on x_{n+1}, and thus, uniquely defined. Hence, if $Z^u \in Z^u(o)$, it is necessarily of the form $(X^u, x_{n+1}(X^u))$ for $X^u \in \mathcal{X}^u(o)$, and the result follows. ■

It turns out from (2.15) that J is purely final but also it can be expressed as a linear functional of the measure P_T^u, image of P_o by the application Z_T^u : precisely, let us denote :

$$(2.17) \qquad P_T^u \overset{def}{=} Z_T^u(o, P_o) \overset{def}{=} \int_\Omega Z_T^{u,\omega}(o, P_o) d\rho(\omega)$$

or, equivalently :

$$(2.18) \qquad \int_{R^{n+1}} \varphi(z) dP_T^u(z) = \int_{R^{n+1} \times \Omega} \varphi(Z_T^{u,\omega}(o,z)) dP_o(z) d\rho(\omega) \quad \forall \varphi \in C_b^o(R^{n+1}).$$

We shall also denote $\int \varphi dP_T^u = \langle \varphi, P_T^u \rangle$ to recall that P_T^u is a linear continuous functional on $C_b^o(R^{n+1})$. On the other hand, since we shall need to extend the problem to P_o : general bounded measure on $C_b^o(R^{n+1})$, we shall always suppose that the measures considered are bounded with arbitrary total mass and sign.

Corollary 2.1 :

$$(2.19) \qquad J_{o,P_o}(u) = \sup_{P_T^u \in \wp_T^u(o,P_o)} \langle \tilde{G}, P_T^u \rangle$$

with :

$$(2.20) \qquad \wp_T^u(o,P_o) = \{P_T^u = Z_T^u(o,P_o) \,|\, Z^u \in Z^u(o)\}. \quad ■$$

Remark 2.1 : Clearly, $\wp_T^u(o,P_o)$ can be interpreted as the set of probability measures of z at time T knowing that the strategy $u = (u_1, \ldots, u_N)$ has been used between o and T. Formula (2.19) shows that $\wp_T^u(o,P_o)$ sums up all the dynaminal informations needed to compute the cost, and thus P^u can be seen as the controlled state variable of the problem ; this is the point of view that we shall develop in the sequel. ■

Remark 2.2 : It can be proved that since g and G are bounded on $[o,T] \times R^n \times U$ and on R^n respectively, the supremum in (2.19) is attained.

We shall not prove this result since we shall not use it in the sequel, but it is an easy consequence of the equicontinuity of the solutions of (2.1) with (2.2)-(see also [20]). ■

Remark 2.3 : If we denote $\psi^*_{\mathcal{P}^u_T(o,P_o)}$ the support function ([25]) of $\mathcal{P}^u_T(o,P_o)$ in the paired spaces $\langle C^o_b(R^{n+1}),\ \mathcal{M}(R^{n+1})\rangle$ (bounded measures on R^{n+1} with the weak* - topology), then by definition :

$$(2.21) \qquad J_{o,P_o}(u) = \psi^*_{\mathcal{P}^u_T(o,P_o)}(\tilde{G}) \qquad \forall u \in \mathcal{U}.$$

A result of the same nature can be found in [32] in a simpler context.

From the properties of support functions, we easily obtain :

$$(2.22) \qquad J_{o,P_o}(u) = \psi^*_{\overline{co}\ \mathcal{P}^u_T(o,P_o)}(\tilde{G}) \qquad \forall u \in \mathcal{U},$$

where $\overline{co}\ \mathcal{P}^u_T(o,P_o)$ is the weak* - closed convex hull of $\mathcal{P}^u_T(o,P_o)$. It is interesting to note that $\overline{co}\ \mathcal{P}^u_T(o,P_o)$ corresponds to Filippov's definition of the trajectories of (2.1) (see [13] and the proof in the Annex). Moreover, if a concept of solution for (2.1) is such that the resulting $\mathcal{Q}^u_T(o,P_o)$ satisfies : $\mathcal{P}^u_T(o,P_o) \subset \mathcal{Q}^u_T(o,P_o) \subset \overline{co}\ \mathcal{P}^u_T(o,P_o)$ $\forall u \in \mathcal{U}$, then, with (2.21) and (2.22), the cost function will not be affected. In particular, this property holds with the solution concepts of Krassovski [20], Sentis [28], and Filippov [13] that, consequently, yield the same cost function. ■

II.2.2. The value function and the optimality principle

Let us define :

$$(2.23) \qquad V(o,P_o) = \inf_{u\in\mathcal{U}} J_{o,P_o}(u) = \inf_{u\in\mathcal{U}} \sup_{P^u_T \in \mathcal{P}^u_T(o,P_o)} \langle \tilde{G}, P^u_T \rangle .$$

In the light of the preceding paragraph, we shall define the value function V for any initial time $t \in [o,T]$ and initial measure. However, since (2.14) is a delayed - differential equation by the contribution of the controls, one must be careful on the nature of such an initial measure : if (2.14) begins at time t, to define completely the solutions, one needs a function $Z_{o,t}$, continuous on $[o,t]$, as an initial condition. Accordingly, the initial measure turns out to be a measure on $C^o([o,t];\ R^{n+1})$, i.e. on the set of continuous trajectories $Z_{o,t}$ (an elementary theory of measures on C^o can be found in [29]). Such a measure will be noted $P_{o,t}$. Precisely, we define $Z^{u,\omega}(t,Z_{o,t})$ as a generalized solution of:

$$(2.24)\quad \begin{cases} \dfrac{d}{ds} Z_s^{u,\omega}(t,Z_{o,t}) = F(s,Z_s^{u,\omega}(t,Z_{o,t}),u(s,Y_s^{u,\omega}(t,Z_{o,t})),v_j) \\ \qquad\qquad \forall s \in [t_j,t_{j+1}[\cap [t,T] \\ Z_s^{u,\omega}(t,Z_{o,t}) = Z_{o,t} \quad \forall s \in [o,t] \ , \end{cases}$$

with $u(s,Y_s^{u,\omega}(t,Z_{o,t}))$ defined as in (2.14) with $Z_s^{u,\omega}(t,Z_{o,t})$ in place of $Z_s^{u,\omega}(o,z)$, and $Z_{o,t} \in C^o([o,t] ; R^{n+1})$.

Accordingly, $Z^u(t)$ is the set of all solutions $Z^{u,\bullet}(t,.)$ of (2.14) such that $s \to Z_s^{u,\omega}(t,Z_{o,t})$ is $\{\mathcal{A}_s \times \mathcal{F}_{o,t} \ ; \ s \in [t,T]\}$ progressively measurable, $\mathcal{F}_{o,t}$ being a given σ-field on $C^o([o,t] ; R^{n+1})$.

Finally, if $P_{o,t}$ is a bounded measure on $(C^o([o,t] ; R^{n+1}), \mathcal{F}_{o,t})$, one can define :

$$(2.25)\qquad \mathcal{P}_T^u(t,P_{o,t}) = \{P_T^u = Z_T^u(t,P_{o,t}) \,|\, Z^u \in Z^u(t)\}$$

with similar formulas as (2.17), (2.20) :

$$(2.26)\qquad P_T^u \overset{\text{def}}{=} Z_T^u(t,P_{o,t}) \overset{\text{def}}{=} \int_\Omega Z_T^{u,\omega}(t,P_{o,t})d\rho(\omega)$$

or, equivalently, noting $C_{o,t}$ for $C^o([0,t] ; R^{n+1})$:

$$(2.27)\qquad \int_{R^{n+1}} \varphi(z)dP_T^u(z) = \int_{\Omega\times C_{o,t}} \varphi(Z_T^{u,\omega}(t,Z_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega) \qquad \forall \varphi \in C_b^o(R^{n+1}).$$

Finally, we shall define $J_{t,P_{o,t}}(u)$ by :

$$(2.28)\qquad J_{t,P_{o,t}}(u) = \sup_{P_t^u \in \mathcal{P}_T^u(t,P_{o,t})} \langle \tilde{G}, P_T^u \rangle \qquad \forall u \in \mathcal{U}$$

and :

$$(2.29)\qquad V(t,P_{o,t}) = \inf_{u\in\mathcal{U}} J_{t,P_{o,t}}(u).$$

We have the following transition properties :

Proposition 2.2 :

$$(2.30)\qquad \wp_T^u(t,P_{o,t}) = \bigcup_{Z^u\in\mathcal{Z}^u(t)} \wp_T^u(s,Z^u_{t,s}(P_{o,t})) \quad \forall s\in[t,T],\quad \forall u\in\mathcal{U}$$

where $Z^u_{t,s}(P_{o,t})$ is the bounded measure on $C^o([o,s]\,;\,R^{n+1})$ defined by :

$$(2.31)\qquad \int_{C_{o,s}} \varphi(Z_{o,s})d(Z^u_{t,s}(P_{o,t}))(Z_{o,s}) = \int_{C_{o,t}\times\Omega} \varphi(Z^{u,\omega}(t,Z_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega)$$

$$\forall\varphi\in C_b^o(C_{o,s})$$

Consequently :

$$(2.32)\qquad J_{t,P_{o,t}}(u) = \sup_{Z^u\in\mathcal{Z}^u(t)} J_{s,Z^u_{t,s}(P_{o,t})}(u) \quad \forall u\in\mathcal{U},\quad \forall s\in[t,T]$$

and (optimality principle) :

$$(2.33)\qquad V(t,P_{o,t}) = \inf_{u\in\mathcal{U}} \sup_{Z^u\in\mathcal{Z}^u(t)} V(s,Z^u_{t,s}(P_{o,t})) \quad \forall s\in[t,T].$$

Proof : (2.30) follows from the fact that $P^u_T\in\wp^u_T(t,P_{o,t})$ is equivalent to $\exists\, Z^u\in\mathcal{Z}^u(t)$ such that $P^u_T = Z^u_T(t,P_{o,t})$, which in turn is equivalent to $P^u_T = Z^u_T(s,Z^u_{t,s}(P_{o,t}))\in\bigcup_{Z^u\in\mathcal{Z}^u(t)}\wp^u_T(s,Z^u_{t,s}(P_{o,t}))$.

To prove (2.32), it suffices to rewrite the definition (2.28) with (2.30) :

$$J_{t,P_{o,t}}(u) = \sup_{P^u_T\in\wp^u_T(t,P_{o,t})} \langle\tilde{G},P^u_T\rangle = \sup_{P^u_T\in\bigcup_{Z^u\in\mathcal{Z}^u(t)}\wp^u_T(s,Z^u_{t,s}(P_{o,t}))} \langle\tilde{G},P^u_T\rangle$$

$$= \sup_{Z^u\in\mathcal{Z}^u(t)} \sup_{P^u_T\in\wp^u_T(s,Z^u_{t,s}(P_{o,t}))} \langle\tilde{G},P^u_T\rangle$$

$$= \sup_{Z^u\in\mathcal{Z}^u(t)} J_{s,Z^u_{t,s}(P_{o,t})}(u).$$

Let us now prove (2.33) : from (2.32), we have :

$$(2.34)\quad V(t,P_{o,t}) \leq \sup_{Z^u \in \mathcal{Z}^u(t)} J_{s,Z^u_{t,s}(P_{o,t})}(u) \qquad \forall u \in \mathcal{U}$$

Thus, since $\mathcal{U}$ can be seen as the product $\mathcal{U}_{t,s} \times \mathcal{U}_{s,T}$ of concatenated $u_{t,s}$ and $u_{s,T}$ defined respectively on $[t,s[$ and $[s,T]$, $\{\mathcal{Y}(t)\}$-progressively measurable on each subinterval (the concatenation $(u_{t,s}|u_{s,T})$ being trivially $\{\mathcal{Y}(t)\}$-progressively measurable on $[t,T]$), and satisfying (2.5) on $[t,s[$ and $[s,T]$, and since $J_{s,Z^u_{t,s}(P_{o,t})}(u)$ is only defined on $\mathcal{U}_{s,T}$, taking the Inf for $u \in \mathcal{U}_{s,T}$ in the right-hand side of (2.34), we obtain, noting $Z^u(t)_{|[t,s[}$ the restriction to $[t,s[$ of the trajectories of $Z^u(t)$:

$$V(t,P_{o,t}) \leq \sup_{Z^u_{t,s} \in Z^u(t)_{|[t,s[}} V(s,Z^u_{t,s}(P_{o,t})) \qquad \forall u \in \mathcal{U}_{t,s},$$

thus :

$$V(t,P_{o,t}) \leq \inf_{u \in \mathcal{U}_{t,s}} \sup_{Z^u_{t,s} \in Z^u(t)_{|[t,s[}} V(s,Z^u_{t,s}(P_{o,t}))$$

and finally, using once more the fact that $V(s,Z^u_{t,s}(P_{o,t}))$ depends only on the past of Z^u so that Z^u can be chosen aritrarily on $[s,T]$, we have :

$$(2.35)\qquad V(t,P_{o,t}) \leq \inf_{u \in \mathcal{U}} \sup_{Z^u \in \mathcal{Z}^u(t)} V(s,Z^u_{t,s}(P_{o,t})).$$

Since, by assumption, g and G are bounded, we have : $|V(t,P_{o,t})| < +\infty$, and consequently there exists a sequence $\{u_m\}_{m \in \mathbb{N}}$ in $\mathcal{U}$ such that :

$$(2.36)\qquad \lim_{m \to \infty} J_{t,P_{o,t}}(u_m) = V(t,P_{o,t})$$

Thus, $\forall m \in \mathbb{N}$, we have :

$$J_{t,P_{o,t}}(u_m) = \sup_{Z^{u_m} \in Z^{u_m}(t)} J_{s,Z^{u_m}_{t,s}(P_{o,t})}(u_m)$$

$$\geq \sup_{Z^{u_m}_{t,s} \in Z^{u_m}(t)|[t,s[} V(s, Z^{u_m}_{t,s}(P_{o,t}))$$

$$\geq \inf_{u \in \mathcal{U}} \sup_{Z^u \in Z^u(t)} V(s, Z^u_{t,s}(P_{o,t}))$$

for the same reason as in (2.35). But the right-hand side is independent on m and we can take the limit as $m \to \infty$:

$$V(t, P_{o,t}) \geq \inf_{u \in \mathcal{U}} \sup_{Z^u \in Z^u(t)} V(s, Z^u_{t,s}(P_{o,t}))$$

and the result is proved. ∎

To have a complete picture of the new state space of probability measures here introduced, we need to describe the state variable $P^u_{o,t}$ as the result of an evolution equation controlled by u for every $t \in [o,T]$. This is the purpose of the :

Proposition 2.3 :

$\forall Z^u \in Z^u(t)$, $\forall P_{o,t} \in \mathcal{M}^b_{o,t}$, the space of bounded measures on $C_{o,t} = C^o([c,t] ; R^{n+1})$, defined on $\mathcal{F}_{o,t}$, σ-field on $C_{o,t}$, there exists a unique measure P^{Z^u} defined on $\mathcal{F}^{Z^u}_{o,T}$ = σ-field on $C_{o,T}$ generated by Z^u, satisfying $P^{Z^u}(Z_{o,t} \in B) = P_{o,t}(B)$ $\forall B \in \mathcal{F}_{o,t}$ and determined by the family of transition probability functions :

(2.37) $P^{Z^u}(t, Z_{o,t} ; s, B) = \rho(\{\omega \in \Omega | Z^{u,\omega}_\sigma(t, Z_{o,t}) = Z_{o,t}(\sigma) \quad \forall \sigma \in [o,t], \text{ and}$

$$\forall \sigma \in [t,s] \quad : \quad Z^{u,\omega}_\sigma(t, Z_{o,t}) \in B\}) \quad \forall B \in \mathcal{F}^{Z^u}_{o,s}, \quad \forall s \in [t,T].$$

satisfying the Chapman - Kolmogorow equation ;

$$P^{Z^u}(t,Z_{o,t}\ ;\ s,B) = \int_{C_{o,\sigma}} P^{Z^u}(\sigma,Z_{o,\sigma}\ ;\ s,B)\ P^{Z^u}(t,Z_{o,t}\ ;\ \sigma,dZ_{o,\sigma})$$

(2.38)

$$P^{Z^u}(t,Z_{o,t}\ ;\ t,B) = 1_B(Z_{o,t}) \quad \forall B \in \mathfrak{F}_{o,t}.$$

Furthermore, we have : $P^{Z^u} = Z^u_{t,T}(P_{o,t})$ defined by (2.31), and since Z^u satisfies :

$$\frac{d}{ds} Z_s^{u,\omega}(t,Z_{o,t}) = \tilde{F}_{u,\omega}(Z_s^{u,\omega}(t,Z_{o,t})) \quad \text{p.p. } s \in [t,T]$$

(2.39)

$$Z_\sigma^{u,\omega}(t,Z_{o,t}) = Z_{o,t}(\sigma) \quad \forall \sigma \in [o,t]$$

with $\tilde{F}_{u,\omega}(Z_s^{u,\omega}(t,Z_{o,t})) \in \overline{F}(s,Z_s^{u,\omega}(t,Z_{o,t}),u(s,Y_s^{u,\omega}(t,Z_{o,t})),v_j)$

(defined in the Annex : $\overline{F}$ is the set of accumulation points of $F(s,z,u,v_j)$ when u is piecewise constant on intervals whose magnitude tend to 0), it follows that :

$$E_{P^{Z^u}}(\varphi(z_s) - \int_t^s \frac{\partial\varphi}{\partial z}(z_\sigma).\tilde{F}_{u,\omega}(Z_{o,\sigma})d\sigma | Z_{o,t}) = \varphi(z_t) \tag{2.40}$$

$$\forall \varphi \in C_b^1(R^{n+1}),\ \forall Z_{o,t} \in C_{o,t} \quad \text{with} \quad z_\sigma = Z_{o,\sigma}(\sigma) \quad \forall \sigma \in [t,s],$$

or equivalently :

$$\frac{d}{ds}\int_{C_{o,s}} \varphi(z_s)P^{Z^u}(t,Z_{o,t}\ ;\ dZ_{o,s}) = \int_{C_{o,s}} \frac{\partial\varphi}{\partial z}(z_s).F_{u,\omega}(Z_{o,s})P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s} \tag{2.41}$$

$$\forall \varphi \in C_b^1(R^{n+1}), \quad \text{and for almost every} \quad s \in [t,T].$$

Proof : Let us first remark that (2.37) is equivalent to :

$$\int_{C_{o,s}} \varphi(Z_{o,s})P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s}) = \int_\Omega \varphi(Z_{o,s}^{u,\omega}(t,Z_{o,t}))d\rho(\omega) \quad \forall \varphi \in C_b^o(C_{o,s}) \tag{2.42}$$

and thus, 2.38) flowllows trivially from the semi -group property of $Z^{u,\omega}(t,Z_{o,t})$ with respect to the concatenation.

On the other hand, from (2.37) we have, $\forall \varphi \in C_b^o(C_{o,s})$:

$$\int_{C_{o,s}\times C_{o,t}} \varphi(Z_{o,s}) P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s})dP_{o,t}(Z_{o,t}) =$$

$$\int_{\Omega\times C_{o,t}} \varphi(Z^{u,\omega}_{o,s}(t,Z_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega)$$

$$= \int_{C_{o,s}} \varphi(Z_{o,s})d(Z^u_{t,s}(P_{o,t}))(Z_{o,s}) \quad \text{(defined by (2.31))}.$$

Finally, (2.38) ensures that the family $\{Z^u_{b,s}(P_{o,t}) | s \in [t,T]\}$ is consistent (see [29]) and thus, if we define :

$Z^u_{t,s}(P_{o,t}) = P^{Z^u}_{o,s}$, it follows from Kolmogorov's extension theorem that there exists a unique P^{Z^u} whose projection on $[o,s]$ is $P^{Z^u}_{o,s}$ $\forall s \in [t,T]$, and consequently, $Z^u_{t,T}(P_{o,t}) = P^{Z^u}$.

Let us now prove (2.40) and (2.41).

Let $\varphi \in C^1_b(R^{n+1})$, and let $\varepsilon > o$.

$$(2.43)\quad D_\varepsilon \overset{\text{def}}{=} \frac{1}{\varepsilon}\Big(\int_{C_{o,s+\varepsilon}} \varphi(z_{s+\varepsilon})P^{Z^u}(t,Z_{o,t}\ ;\ s+\varepsilon,dZ_{o,s+\varepsilon}) -$$

$$\int_{C_{o,s}} \varphi(z_s)P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s})$$

$$= \frac{1}{\varepsilon}\Big(\int_{C_{o,s+\varepsilon}\times C_{o,s}} \varphi(z_{s+\varepsilon})P^{Z^u}(s,Z_{o,s}\ ;\ s+\varepsilon,dZ_{o,s+\varepsilon}) -$$

$$\varphi(z_s)]P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s})\Big)$$

(from (2.38)). Also, using (2.42) :

$$(2.44)\quad D_\varepsilon = \frac{1}{\varepsilon}\int_{C_{o,s}\times\Omega} (\varphi(Z^{u,\omega}_{s+\varepsilon}(s,Z_{o,s})) - \varphi(Z_{o,s}(s)))P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s})d\rho(\omega).$$

Finally, since $\varphi \in C^1_b(R^{n+1})$, it follows that $\lim_{\varepsilon \to o} D_\varepsilon$ exists and, from (2.43), we have :

$$\lim_{\varepsilon \to o} D_\varepsilon = \frac{d}{ds}\int_{C_{o,s}} \varphi(z_s)P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s})$$

and, from (2.44) and (2.39) :

$$\lim_{\varepsilon \to o} D_\varepsilon = \int_{C_{o,s}} \frac{\partial\varphi}{\partial z}(z_s).\tilde{F}_{u,\omega}(Z_{o,s})P^{Z^u}(t,Z_{o,t}\ ;\ s,dZ_{o,s})$$

and (2.4') is proved. To prove (2.40), it suffices to integrate (2.41) between t and s. Conversely, if (2.40) holds, it suffices to write it in the form :

$$\int_t^s \frac{d}{d\sigma} E_{P^{Z^u}}(\varphi(z_\sigma)|Z_{o,t})d\sigma = \int_t^s E_{P^{Z^u}}(\frac{\partial \varphi}{\partial z}(z_{o,\sigma}).\tilde{F}_{u,\omega}(Z_{o,\sigma})|Z_{o,t})d\sigma$$

and thus (2.40) is equivalent to (2.41) which achieves to prove the Proposition. ∎

Proposition 2.4 :

Suppose that $(P,\mathfrak{F})$ are given on $C_{o,T}$, satisfying (2.40) $\forall \varphi \in C_b^1(R^{n+1})$, $\forall Z_{o,t} \in C_{o,t}$, for a given $\tilde{F}_{u,.}$ and $u \in \mathcal{U}$, with :

(2.45) $$P(Z_{o,t} \in B) = P_{o,t}(B) \quad \forall B \in \mathfrak{F}_{o,t} .$$

Then there exists $Z^u \in \mathcal{Z}^u(t)$ such that $P = Z^u_{t,T}(P_{o,t})$.

Proof : Let us denote $P_{o,s}$ the restriction of P to $[o,s]$ $\forall s \in [o,T]$. (2.40) and (2.45) imply that :

$$(2.46)\quad E_{P_{o,t}}(E_{P_{o,s}}(\varphi(z_s) - \varphi(z_t)|Z_{o,t})) = E_{P_{o,s}}(\varphi(z_s) - \varphi(z_t)) = E_{P_{o,s}}(\int_t^s \frac{\partial \varphi}{\partial z}(z_\sigma).\tilde{F}_{u,\omega}(Z_{o,\sigma})d\sigma)$$

But it is not difficult to check that :

$$(2.47)\quad E_{P_{o,s}}(\int_t^s \frac{\partial \varphi}{\partial z}(z_\sigma).\tilde{F}_{u,\omega}(Z_{o,\sigma})d\sigma) = \int_t^s \frac{d}{d\sigma}(\int_{C_{o,t}\times\Omega} \varphi(Z_\sigma^{u,\omega}(t,Z_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega))d\sigma$$

where $Z^{u,\omega}(t,Z_{o,t})$ satisfies :

$$(2.48)\quad \begin{cases} \frac{d}{d\sigma}Z_\sigma^{u,\omega}(t,Z_{o,t}) = \tilde{F}_{u,\omega}(Z_{o,\sigma}^{u,\omega}(t,Z_{o,t})) \\ Z_\sigma^{u,\omega}(t,Z_{o,t}) = Z_{o,t}(\sigma) \quad \forall \sigma \in [0,t] \end{cases}$$

with $\tilde{F}$ as in (2.39). Thus Z^u obviously satisfies $Z^u \in \mathcal{Z}^u(t)$, and (2.46) yields :

(2.49) $$\int_t^s \frac{d}{d\sigma}\left(\int \varphi(z_\sigma)dP_{o,\sigma}(Z_{o,\sigma})\right)d\sigma = \int_t^s \frac{d}{d\sigma}\left(\int \varphi(Z_\sigma^{u,\omega}(t,Z_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega)\right)d\sigma.$$

Hence, $P_{o,\sigma} = Z^u_{o,\sigma}(P_{o,t})$ $\forall \sigma \in [t,T]$, and the result is proved. ■

Remark 2.4 : Proposition 2.3 and its converse 2.4 give the solution to a problem which is the analogue, in the context of discrete-time noises, of the problem of martingales of Stroock and Varadhan ([29]), for the infinitesimal generator $L_\sigma^{Z^u,\omega}\varphi(z) = \frac{\partial \varphi}{\partial z}(z_\sigma).\tilde{F}_{u,\omega}(Z_{o,\sigma})$. The fact that $L_\sigma^{Z^u,\omega}$ is of order 1 confirms that the diffusion is degenerated, that we knew from the beginning ! ■

Remark 2.5 : $P_{o,t}$ is never assumed to be positive with total mass 1. This generality will be useful in the sequel since the function $V(t,P_{o,t})$ can in fact be defined on the whole vector space $\mathcal{M}^b_{o,t}$ of general bounded measures on $C^o([o,t] ; R^{n+1})$. ■

Remark 2.6 : For simplicity's sake, we have introduced measures $P_{o,t}$ on the whole past, whereas the observations of station k are only defined on $[\mathcal{K}_k(t),t]$. However, the result is the same since we take expectations of functions of the observations between $\mathcal{K}_k(t)$ and t, $k=1,\dots,N$, the observations paths before $\mathcal{K}_k(t)$ being arbitrary, and giving no contribution to the integrals. ■

Remark 2.7 : Going back to (2.33), we see that this equation is simply the principle of transition of Dynamic Programming, and therefore constitutes a backward equation : from the final condition : $V(T,P_{o,T}) = \langle \tilde{G}, P_T \rangle$ $\forall P_{o,T}$, going backward in time, one progressively computes $V(t,P_{o,t})$ $\forall t \leq T$, $\forall P_{o,t}$. A "dual" approach can also be developed, as in [32] or, in the classical informational structure, with the nonlinear semigroup theory, which gives rise to an onward equation (see [11][14]) :

Let us define the family of nonlinear operators $\{\mathcal{J}_s\}_{s\geq o}$ by :

(2.50) $$\mathcal{J}_s(V(t,P_{o,t})) = \operatorname*{Inf}_{u\in\mathcal{U}} \operatorname*{Sup}_{Z^u\in Z^u(t)} V(t+s,Z^u_{t,t+s}(P_{o,t})) \quad \forall s \in [o,T-t].$$

Thus, the reader can easily check that (2.33) becomes :

$$(2.51)\quad \begin{cases} \mathcal{J}_{s+\sigma}(V(t,P_{o,t})) = \mathcal{J}_s(\mathcal{J}_\sigma(V(t,P_{o,t}))) = \mathcal{J}_\sigma(\mathcal{J}_s(V(t,P_{o,t}))) \\ \qquad\qquad \forall s,\sigma \geqslant o \ \text{ such that } \ s+\sigma \leqslant T-t. \\ \mathcal{J}_o(V(t,P_{o,t})) = V(t,P_{o,t}), \ \text{ or equivalently : } \ \mathcal{J}_o = I. \end{cases}$$

Hence, the family $\{\mathcal{J}_s\}_{s\geqslant o}$ forms a semigroup of nonlinear operators. However, it is clear that this method is equivalent to the Dynamic Programming and the translation of the results in the sequel in terms of nonlinear semigroup is left to the reader. ■

Remark 2.8 : If we look at (2.33) as an optimization problem, and if we try to compute a solution $u^* \in \mathcal{U}$, supposed to exist, then obviously u^* will be a function of $P_{o,t}$, whereas this is not allowed in the definition of $\mathcal{U}$. By analogy to the deterministic control theory, we shall say that the current definition of $\mathcal{U}$ is in open-loop (the controls do not depend on the state $P_{o,t}$), and that u^* is obtained in closed-loop form. Thus, if we want to solve (2.10) or (2.29) by means of (2.33), we need a comparison result between open-loop and closed-loop solutions. This point will be analyzed in the next section. ■

II.2.3. The closed-loop formulation

Let $\mathcal{M}^b_{o,T}$ be the space of bounded measures on $C^o([o,T]\ ;\ R^{n+1})$ endowed with the weak* topology, let $\mathcal{B}_{\mathcal{M}}$ be its Borel σ-field, and $\mathcal{B}^t_{\mathcal{M}}$ its projection on $C^o([o,t]\ ;\ R^{n+1})$, $\forall t \in [o,T]$. Let us also denote $\mathcal{B}_{U_k}$ the Borel σ-field of U_k, $k=1,\dots,N$, and $Y_k(o,T)$ the space of piecewise continuous observation paths between 0 and T satisfying (2.4), with the family $\{\mathcal{Y}_k(t)\ ;\ t \in [o,T]\}$ of sub σ-fields of $\mathcal{F}_o \times \mathcal{A}$. We recall that this family $\{\mathcal{Y}_k(t)\}$ is not supposed to be increasing with respect to t.

Définition 2.2 : An admissible closed-loop strategy $\tilde{u}_k$ for decision maker k is an application from $[o,T] \times Y_k(o,T) \times \mathcal{M}^b_{o,T}$ to U_k which is progressively measurable with respect to $\{\mathcal{Y}_k(t) \times \mathcal{B}^t_{\mathcal{M}}\ ;\ t \in [o,T]\}$, namely :

$u_k(t)$ is $(\mathcal{Y}_k(t) \times \mathcal{B}^t_{\mathcal{M}}\ ,\ \mathcal{B}_{U_k})$-measurable $\forall t \in [o,T]$, and :
u_k restricted to $[o,t]$ is $(\mathcal{B}_{[o,t]} \times \mathcal{Y}_k(o,t) \times \mathcal{B}^t_{\mathcal{M}}\ ,\ \mathcal{B}_{U_k})$-measurable,
where we recall that $\mathcal{Y}_k(o,t) = \bigcup_{o\leqslant s\leqslant t} \mathcal{Y}_k(s)$. ■

We denote $\tilde{\mathcal{U}}_k$ the set of every admissible closed-loop strategies for station k, and $\tilde{\mathcal{U}} = \tilde{\mathcal{U}}_1 \times \ldots \times \tilde{\mathcal{U}}_N$.

Clearly, we have $\mathcal{U} \subset \tilde{\mathcal{U}}$ and every $u_k \in \tilde{\mathcal{U}}_k$ is, by definition, a nonanticipative application of $(t, Y_k(t), P_{o,t})$, with $t \to u_k(t, Y_k(t), P_{o,t}) = \mathcal{B}_{[o,T]}$-measurable, $\forall k=1,\ldots,N$.

In order to compare open-loop and closed-loop types of optimality, we need to extend the definitions of $Z^u(t)$, $\mathcal{P}_T^u(t, P_{o,t})$, $J_{t,P_{o,t}}(u)$ for $u \in \tilde{\mathcal{U}}$, and, consequently, of $V(t, P_{o,t})$.

In fact all these extensions follow directly from the extension of the concept of trajectories for closed-loop strategies. Thus, adapting Krassovski's definition [20], we set :

$\tilde{Z}^{u,\omega}(t, Z_{o,t}, P_{o,t})$ is the set of all the uniform limits in $C_{o,T}$ of trajectories $Z^{m,\omega}(t, \zeta^m_{o,t}, \pi^m_{o,t})$ of (2.24) generated by the piecewise constant approximations of u of the form :

$$(2.52) \qquad u(\tau^m_j, Y^{m,\omega}_{\tau^m_j}(t, \zeta^m_{o,t}), \pi^{Z^m}_{o,\tau^m_j})$$

$$(2.53) \quad \left\{ \begin{array}{l} \text{for every deterministic sequence } \{\tau^m_j\} \text{ of subdivisions of } [t,T] \\ \text{such that } \lim_{m \to \infty} \sup_j (\tau^m_{j+1} - \tau^m_j) = o, \text{ for every } \{\zeta^m_{o,t}\} \text{ such that} \\ \lim_{m \to \infty} \| \zeta^m_{o,t} - Z_{o,t} \|_{C_{o,t}} = o, \\ \text{and for every } \{\pi^m_{o,t}\} \text{ in } \mathcal{M}^b_{o,t} \text{ such that } \lim_{m \to \infty} \pi^m_{o,t} = P_{o,t} \\ \text{in the weak* - topology of } \mathcal{M}^b_{o,t}. \end{array} \right.$$

A precise definition of $\tilde{Z}^{u,\omega}(t, Z_{o,t}, P_{o,t})$ is given in the Appendix, as well as some of its properties.

Also, we define $\tilde{Z}^u(t)$ as the set of all the solutions $\tilde{Z}^{u,\circ}(t,\circ,\circ)$ defined here above such that the process :

$$(2.54) \quad s \to Z^{u,\omega}_s(t, Z_{o,t}, P_{o,t}) \text{ is } \{\mathcal{A}_s \times \mathcal{F}_{o,t} \times \mathcal{B}^s_{\mathcal{M}} ; s \in [t,T]\}\text{-progressively measurable.}$$

Consequently, we set $\tilde{\wp}_T^u(t,P_{o,t}) = \{\tilde{P}_T^u = \tilde{Z}_T^u(t,P_{o,t}) | \tilde{Z}^u \in \tilde{Z}^u(t)\}$ with the notation : $\tilde{Z}_T^u(t,P_{o,t})$ for the image by $\tilde{Z}_T^u$ of $P_{o,t}$, namely :

$$(2.55)\quad \int_{R^{n+1}} \varphi(z)d(\tilde{Z}_T^u(t,P_{o,t}))(z) = \int_{\Omega\times C_{o,t}} \varphi(\tilde{Z}_T^{u,\omega}(t,Z_{o,t},P_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega)$$

$$\forall \varphi \in C_b^o(R^{n+1}).$$

Finally :

$$(2.56)\quad J_{t,P_{o,t}}(u) = \sup_{\tilde{P}_t^u \in \tilde{\wp}_T^u(t,P_{o,t})} \langle \tilde{G}, \tilde{P}_T^u \rangle \quad \forall u \in \tilde{\mathcal{U}}, \quad \text{and} \quad \tilde{V}(t,P_{o,t}) = \inf_{u\in\mathcal{U}} J_{t,P_{o,t}}(u).$$

Proposition 2.5. : The conclusions of the proposition 2.2, 2.3 and 2.4 hold for $\tilde{\mathcal{U}}$, $\tilde{Z}$, $\tilde{\wp}$, $\tilde{J}$ and $\tilde{V}$ in place of $\mathcal{U}$, Z, $\wp$, J and V. ■

The proof follows exactly the same lines as those of propositions 2.2 to 2.4 and is left to the reader.

In order to compare V and $\tilde{V}$, the value functions for open-loop and closed-loop strategies, we want to point out that the situation is not as clear as in a deterministic control problem : in this case, it suffices that a closed-loop strategy $u(t,x)$ generates a unique solution $x(.)$ of the differential equation, to obtain an open-loop strategy, namely $u(t,x(t))$, giving the same cost. Here though we use the same kind of idea, we have to be sure that the open-loop representation of u does not generate other trajectories yielding a higher cost (remember that we take the supremum of every cost induced by every possible trajectories generated by a given u !).

We have the following lemma :

Lemma 2.1 : $\forall \varepsilon > o$ and $\forall u \in \tilde{\mathcal{U}}$, one can find $u_\varepsilon \in \mathcal{U}$ satisfying :

(i) $Z^{u_\varepsilon}(t)$ contains exactly one element.

(ii) $|\tilde{J}_{t,P_{o,t}}(u) - J_{t,P_{o,t}}(u_\varepsilon)| \leq \varepsilon$.

Proof : Let $u \in \tilde{U}$. Clearly, $|\tilde{J}_{t,P_{o,t}}(u)| < +\infty$.

By definition of $\tilde{J}$, one can find $\tilde{Z}^u \in Z^u(t)$, and $\tilde{P}^u_T = \tilde{Z}^u_T(t,P_{o,t})$ such that :

$$(2.57) \qquad |\tilde{J}_{t,P_{o,t}}(u) - \langle \tilde{G}, \tilde{P}^u_T \rangle| \leq \frac{\varepsilon}{3}.$$

Also, by definition of $\tilde{Z}^u$, one can find a subdivision $\tau_o < \tau_1 < \ldots < \tau_L$ of $[t,T]$, an initial condition $\zeta_{o,t}$ in a neighborhood of $Z_{o,t}$, and an initial probability measure $\pi_{o,t}$ in a neighborhood of $P_{o,t}$ such that the trajectory $Z^{L,u,\omega}(t,\zeta_{o,t},\pi_{o,t})$ obtained by the approximations scheme (2.52), (2.53), which is uniquely defined for each $L,\omega,\zeta_{o,t}$ and $\pi_{o,t}$, satisfies, by the uniform convergence :

$$(2.58) \qquad \left|\int_{\Omega\times C_{o,t}} \tilde{G}(Z^{L,u,\omega}_T(t,\zeta_{o,t},\pi_{o,t}))d\pi_{o,t}(\zeta_{o,t})d\rho(\omega) - \langle \tilde{G}, \tilde{P}^u_T \rangle\right| \leq \frac{\varepsilon}{3},$$

with $\pi_{o,t}$ satisfying

$$(2.59) \qquad \left|\int_\Omega \left(\int_{C_{o,t}} \tilde{G}(Z^{L,u,\omega}_T(t,\zeta_{o,t},\pi_{o,t}))d\pi_{o,t}(z_{o,t}) - \int_{C_{o,t}} \tilde{G}(Z^{L,u,\omega}_T(t,\zeta_{o,t},\pi_{o,t}))dP_{o,t}(\zeta_{o,t})\right)d\rho(\omega)\right| \leq \frac{\varepsilon}{3},$$

(2.59) being a consequence of the uniform convergence of Z^L_T on compect subsets of $C_{o,t}$.

Thus, putting (2.57) to (2.59) together :

$$(2.60) \qquad \left|\int_{\Omega\times C_{o,t}} \tilde{G}(Z^{L,u,\omega}_T(t,Z_{o,t},\pi_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega) - \tilde{J}_{t,P_{o,t}}(u)\right| \leq \varepsilon.$$

Finally, it suffices to define :

$$(2.61) \qquad u_\varepsilon(s, Y^{u_\varepsilon,\omega}_s(t,Z_{o,t})) = u(\tau_j, Y^{L,u,\omega}_{\tau_j}(t,Z_{o,t}), \pi^{Z^L}_{o,\tau_j}) \in U$$

$$\forall s \in [\tau_j,\tau_{j+1}[\cap [t,T],\ \forall j=o,\ldots,L-1,\quad \forall \omega \in \Omega,\quad \forall Z_{o,t} \in C_{o,t}.$$

It is straightforward to check that u_ε is $\{\mathcal{Y}(s)\ ;\ s \in [t,T]\}$-progressively measurable, for example by Benes lemma [5], and thus $u_\varepsilon \in \mathcal{U}$. On the other hand, u_ε generates only one trajectory for each ω and $Z_{o,t}$, by construction, and thus $Z^{u_\varepsilon}(t) = \{Z^{L,u,\cdot}(t,.,\tau_{o,t})\}$ has only one element. Finally :

$$J_{t,P_{o,t}}(u_\varepsilon) = \sup_{\substack{u_\varepsilon \\ P_T^{u_\varepsilon} \in \mathcal{P}_T^{u_\varepsilon}(t,P_{o,t})}} \langle \tilde{G}, P_T^{u_\varepsilon} \rangle = \int_{\Omega \times C_{o,t}} \tilde{G}(Z_T^{L,u,\omega}(t,Z_{o,t},\pi_{o,t}))dP_{o,t}(Z_{o,t})d\rho(\omega)$$

and (ii) follows from (2.60). ∎

Theorem 2.1 :

$$(2.62) \qquad \tilde{V}(t,P_{o,t}) = V(t,P_{o,t}) \quad \forall t \in [0,T], \quad \forall P_{o,t} \in \mathcal{P}^b_{o,t}\ .$$

In other words the open-loop and closed-loop team problems are equivalent.

Proof : Since $\mathcal{U} \subset \tilde{\mathcal{U}}$, $\tilde{V}(t,P_{o,t}) = \operatorname{Inf}_{u \in \tilde{\mathcal{U}}} \tilde{J}_{t,P_{o,t}}(u) \leq \operatorname{Inf}_{u \in \mathcal{U}} \tilde{J}_{t,P_{o,t}}(u)$.

But it is easy to check that if $u \in \mathcal{U}$, $\tilde{J}_{t,P_{o,t}}(u) = J_{t,P_{o,t}}(u)$, and thus :

$$(2.63) \qquad \tilde{V}(t,P_{o,t}) \leq V(t,P_{o,t}).$$

On the other hand, since $\tilde{V}$ is the infimum over $\tilde{\mathcal{U}}$, $\forall \varepsilon > 0$, one can find $\tilde{u} \in \tilde{\mathcal{U}}$ such that :

$$(2.64) \qquad \tilde{J}_{t,P_{o,t}}(\tilde{u}) \leq \tilde{V}(t,P_{o,t}) + \frac{\varepsilon}{2}\ .$$

Thus, by lemma 2.1, there exists a $u_\varepsilon \in \mathcal{U}$ such that :

$$(2.65) \qquad |J_{t,P_{o,t}}(u_\varepsilon) - \tilde{J}_{t,P_{o,t}}(\tilde{u})| \leq \frac{\varepsilon}{2}\ .$$

Finally, from (2.64) and (2.65), we obtain :

$$(2.66) \quad V(t,P_{o,t}) \leq J_{t,P_{o,t}}(u_\varepsilon) \leq \tilde{J}_{t,P_{o,t}}(\tilde{u}) + \frac{\varepsilon}{2} \leq \tilde{V}(t,P_{o,t}) + \varepsilon$$

and, since ε is arbitrary, the result is proved. ∎

To conclude, theorem 2.1 answers to the question raised in remark 2.8, and asserts that the solution to (2.10) or (2.29) can be found via the Dynamic Programming equation (2.33) for V or $\tilde{V}$ indifferently.

Remark 2.9 : The assumption that g and G are bounded is not valid in general for linear quadratic problems : However, one can restrict the analysis to a neighborhood of the optimum where this theory applies. ■

II.3. Some regularity properties of the value function

This paragraph is devoted to the proof of three regularity properties of V that will be very useful for the sequel.

. The first property states that V is not only the minimum guaranteed cost, but also the absolute minimum, giving thus the same result as if the trajectories were uniquely defined for every $u \in \tilde{\mathcal{U}}$. Furthermore, if the minimum is attained at $u^* \in \tilde{\mathcal{U}}$, every trajectories generated by u^* give the same minimum cost.

. The second one is an integral representation of V that will be useful to interpret the directional derivatives obtained in the sequel in the optimality criteria.

. Finally, the third one states that V is concave and continuous with respect to $P_{o,t}$, and thus everywhere subdifferentiable (in the sense of concave functions). The subdifferentiability will be used to obtain the derivative of V along the trajectories of P^{Z^u}.

Proposition 2.6 : V satisfies :

$$(2.67)\quad V(t,P_{o,t}) = \inf_{u\in\mathcal{U}} \inf_{\tilde{P}_T^u \in \tilde{\wp}_T^u(t,P_{o,t})} \langle \tilde{G}, \tilde{P}_T^u \rangle, \quad \forall t \in [o,T], \quad \forall P_{o,t} \in \mathcal{M}_{o,t}^b.$$

Furthermore, if $u^* \in \tilde{\mathcal{U}}$ is such that $\tilde{J}_{t,P_{o,t}}(u^*) = \tilde{V}(t,P_{o,t})$, then

$$(2.68)\quad V(t,P_{o,t}) = \langle \tilde{G}, \tilde{P}_T^{u^*} \rangle, \quad \forall \tilde{P}_T^{u^*} \in \tilde{\wp}_T^{u^*}(t,P_{o,t}), \quad \forall t \in [o,T], \quad \forall P_{o,t} \in \mathcal{M}_{o,t}^b.$$

Proof : By definition of V and by theorem 2.1, we have :

$$V(t,P_{o,t}) \leq \sup_{\tilde{P}_T^u \in \tilde{\wp}_T^u(t,P_{o,t})} \langle \tilde{G}, \tilde{P}_T^u \rangle \quad \forall u \in \tilde{\mathcal{U}}.$$

Let $u \in \tilde{\mathcal{U}}$ and $\tilde{Q}_T^u \in \tilde{\wp}_T^u(t,P_{o,t})$ be such that :

$$(2.69) \qquad \begin{aligned} &\langle G, Q_T^u \rangle < \sup_{\tilde{P}_T^u \in \tilde{\wp}_T^u(t,P_{o,t})} \langle \tilde{G}, \tilde{P}_T^u \rangle, \quad \text{and} \\ &\langle \tilde{G}, \tilde{Q}_T^u \rangle < V(t,P_{o,t}) - \varepsilon \quad \text{with} \quad \varepsilon > o \ , \end{aligned}$$

the last inequality being feasible since $V > -\infty$

But, by lemma 2.1, there exists $u_\varepsilon \in \mathcal{U}$ such that :

$$(2.70) \quad \wp_T^{u_\varepsilon}(t,P_{o,t}) = \{P_T^{u_\varepsilon}\} \quad \text{and} \quad J_{t,P_{o,t}}(u_\varepsilon) = \langle \tilde{G}, P_T^{u_\varepsilon} \rangle \leq \langle \tilde{G}, \tilde{Q}_T^u \rangle + \varepsilon,$$

and, with (2.69) :

$$(2.71) \quad V(t,P_{o,t}) \leq J_{t,P_{o,t}}(u_\varepsilon) \leq \langle \tilde{G}, \tilde{Q}_T^u \rangle + \varepsilon < V(t,P_{o,t}) + \varepsilon - \varepsilon$$

which is a contradiction, and (2.67) is proved. (2.68) follows immediately by changing u in u^*. ∎

Proposition 2.7 :

There exists a $\mathcal{B}_{[0,T]} \times \mathcal{B}_{R^{n+1}} \times \mathcal{B}_{\mathcal{M}}$ - measurable function w, integrable with respect to every measure on R^{n+1} of the form $P_t = pr_t(P_{o,t})$, pr_t being the projection operator defined by $pr_t(Z_{o,t}) = Z_{o,t}(t)$, and satisfying :

$$(2.72) \qquad V(t,P_{o,t}) = \int_{R^{n+1}} w(t,z\ ;\ P_{o,t})dP_t(z) \quad \forall t \in [0,T], \quad \forall P_{o,t} \in \mathcal{M}_{o,t}^b.$$

Remark 2.10 : w can be interpreted as the "density of value" of V with respect to the probability measure of $z(t)$, or, more precisely, as the density of cost-to-go from time t and position $z(t)$, knowing that $z(t)$ is the end point of any $Z_{o,t}$ with probability $P_{o,t}$. ∎

Proof of the Proposition 2.7 : Let $\{u_m\}_{m\geq o}$ be a minimizing sequence in $\tilde{u}$, namely a sequence satisfying :

$$\lim_{m\to\infty} \tilde{J}_{t,P_{o,t}}(u_m) = V(t,P_{o,t}) \text{ , with } u_m \in \tilde{u} \quad \forall m \geq o. \tag{2.73}$$

such a sequence exists since V is finite everywhere.

Let us denote :

$$w_m(t,z\ ;\ P_{o,t}) = \int_{C_{o,t}\times\Omega} \tilde{G}(\tilde{Z}_T^{u_m},\omega)(t,Z_{o,t},P_{o,t}))dP_{o,t}(Z_{o,t}|Z_{o,t}(t)=z)d\rho(\omega). \tag{2.74}$$

$\tilde{Z}^{u_m}$ being chosen arbitrarily in $\tilde{Z}^{u_m}(t)$. Clearly, since :

$$\int_{R^{n+1}} w_m(t,z\ ;\ P_{o,t})dP_t(z) = \langle \tilde{G}, \tilde{P}_T^{u_m}\rangle \text{ with } \tilde{P}_T^{u_m} = \tilde{Z}_T^{u_m}(t,P_{o,t}), \tag{2.75}$$

w_m is P_t - integrable for $P_t = pr_t(P_{o,t})$ $\forall P_{o,t} \in \mathcal{M}^b_{o,t}$, and $\mathcal{B}_{[0,T]} \times \mathcal{B}_{R^{n+1}} \times \mathcal{B}_{\mathcal{M}}$- measurable. On the other hand, the w_m are uniformly bounded below by a constant (the constants being P_t integrable) and :

$$\lim_{m\to\infty} \int_{R^{n+1}} w_m(t,z\ ;\ P_{o,t})dP_t(z) = V(t,P_{o,t}) \text{ ,} \tag{2.76}$$

thus, by Faton's lemma :

$$\int_{R^{n+1}} (\lim_{m\to\infty} w_m(t,z\ ;\ P_{o,t}))dP_t(z) \leq V(t,P_{o,t}). \tag{2.77}$$

Let us denote :

$$w(t,z\ ;\ P_{o,t}) = \lim_{m\to\infty} w_m(t,z\ ;\ P_{o,t}) \quad \forall t,z,P_{o,t} \text{ .} \tag{2.78}$$

then, clearly, w is $\mathcal{B}_{[0,T]} \times \mathcal{B}_{R^{n+1}} \times \mathcal{B}_{\mathcal{M}}$-measurable as limit inf of a sequence of measurable functions, and, by (2.77) w is P_t integrable. Finally, by Proposition (2.6), we have :

$$V(t,P_{o,t}) \leq \int w_m(t,z\ ;\ P_{o,t})dP_t(z) \quad \forall m, \text{ and consequently.}$$

$$V(t,P_{o,t}) \leq \int w(t,z\ ;\ P_{o,t})dP_t(z), \tag{2.79}$$

and, together with (2.77), we obtain :

$V(t,P_{o,t}) = \int w(t,z\ ;\ P_{o,t})dP_t(z)$, and (2.72) is proved if we remark that the definition of w does not depend on the choise of the sequence $\{u_m\}$.
But if $\{u_{m'}\}$ is another minimizing sequence, let us denote

$$w'(t,z;P_{o,t}) = \lim_{m' \to \infty} w_{m'}(t,z;P_{o,t}).$$

Also, noting $\{u_{m''}\} = \{u_m\} \cup \{u_{m'}\}, \{u_{m''}\}$ is also a minimizing sequence and $w''(t,z;P_{o,t}) = \lim_{m'' \to \infty} w_{m''}(t,z;P_{o,t})$ satisfies :

$$\begin{cases} w''(t,z;P_{o,t}) \leq w(t,z;P_{o,t}) \\ \\ w''(t,z;P_{o,t}) \leq w'(t,z;P_{o,t}) \end{cases} \qquad \forall t,z,P_{o,t}$$

But $E_{P_t}(w) = E_{P_t}(w') = E_{P_t}(w'') = V(t,P_{o,t})$ and thus :

$$w = w' = w'' \quad P_t \text{ - almost everywhere.} \blacksquare$$

Proposition 2.8 : V satisfies the following properties :

(i) $\forall t \in [o,T]$, the application $P \to V(t,P)$ from $\mathcal{M}^b_{o,t}$ to R, is concave.

(ii) $V(t,P)$ is everywhere finite and $P \to V(t,P)$ is continuous on $\mathcal{M}^b_{o,t}$.

(iii) $P \to V(t,P)$ is subdifferentiable on $\mathcal{M}^b_{o,t}$ and the subdifferential $\partial_P V(t,P)$ (in the sense of concave functions) defined by :

$$(2.80)\quad \partial_P V(t,P) = \{\Lambda \in C^o_b(C_{o,t}) \,|\, V(t,Q)-V(t,P) \leq \langle \Lambda, Q-P \rangle \quad \forall Q \in \mathcal{M}^b_{o,t}\}$$

is a convex compact subset of $C^o_b(C_{o,t})$ (with the uniform topology), $\forall P \in \mathcal{M}^b_{o,t}$, $\forall t \in [o,T]$.

Proof : Let $P, Q \in \mathcal{M}^b_{o,t}$ and $\alpha \in [o,1]$. From Proposition 2.6, we have : $\forall \varepsilon > o$, $\exists\, u_\varepsilon \in \mathcal{U}$ such that :

$$(2.81)\quad V(t,\alpha P+(1-\alpha)Q) \geq \int_{C_{o,t}\times\Omega} \tilde{G}(Z_T^{u_\varepsilon,\omega}(t,Z_{o,t}))d(\alpha P+(1-\alpha)Q)(Z_{o,t})d\rho(\omega) - \varepsilon$$

$$= \alpha \int_{C_{o,t}\times\Omega} \tilde{G}(Z_T^{u_\varepsilon,\omega}(t,Z_{o,t}))dP(Z_{o,t})d\rho(\omega)$$

$$+ (1-\alpha) \int_{C_{o,t}\times\Omega} \tilde{G}(Z_T^{u_\varepsilon,\omega}(t,Z_{o,t}))dQ(Z_{o,t})d\rho(\omega) - \varepsilon$$

by the linearity of the integral. Also, once more applying Proposition 2.6 :

$$(2.82)\quad \begin{cases} \int_{C_{o,t}\times\Omega} \tilde{G}(Z_T^{u_\varepsilon,\omega}(t,Z_{o,t}))dP(Z_{o,t})d\rho(\omega) \geq V(t,P) \\ \int_{C_{o,t}\times\Omega} \tilde{G}(Z_T^{u_\varepsilon,\omega}(t,Z_{o,t}))dQ(Z_{o,t})d\rho(\omega) \geq V(t,Q) \end{cases}$$

and (2.81) becomes : $V(t,\alpha P+(1-\alpha)Q) \geq \alpha V(t,P)+(1-\alpha)V(t,Q)$ which proves the concavity of $P \to V(t,P)$, $\forall t \in [o,T]$.

Then (ii) and (iii) follow from standard results of convex analysis in infinite dimensional vector spaces (see [26]), the only thing to prove being that $\mathcal{M}_{o,t}^b$ and $C_b^o(C_{o,t})$ are paired spaces when $\mathcal{M}_{o,t}^b$ is endowed with the weak* - topology and $C_b^o(C_{o,t})$ with the uniform topology, and this result can be found in [29]. ■

Remark that the continuity property of V cannot be obtained in $\mathcal{M}_{o,T}^{1,+}$ with this method since this subspace of $\mathcal{M}_{o,T}^b$ has an empty interior. ■

We shall apply the subdifferentiability property of V for a special type of increments, namely : $\frac{1}{\varepsilon}(P_{o,t+\varepsilon}-P_{o,t})$, to obtain a formula for $\frac{\partial V}{\partial P} \cdot \frac{dP_{o,t}}{dt}$ that will be of special interest in the next section.

In fact, since $P_{o,t+\varepsilon}$ and $P_{o,t}$ are not in the same space, we must define $V(t,P_{o,t+\varepsilon})$ $\forall \varepsilon > o$.

Definition 2.2 : Denote $\theta_{t'-t}$ the translation operator from $C_{o,t'}$ to $C_{o,t}$, with $t' \geq t$, defined by : $\theta_{t'-t}(Z_{o,t'})(s) = Z_{o,t'}(s+(t'-t))$.

Then, we define $\mathcal{P}_T^u(t,P_{o,t'}) = \{P = Z_T^u(t,\theta_{t'-t}(P_{o,t'}))\,|\,Z^u \in Z^u(t)\}$, and

$J_{t,P_{o,t'}}(u) = \sup_{P\in\mathcal{P}_T^u(t,P_{o,t'})} \langle \tilde{G},P\rangle$ $\forall u \in \mathcal{U}$, $V(t,P_{o,t'}) = \inf_{u\in\mathcal{U}} J_{t,P_{o,t'}}(u)$. ■

Proposition 2.9 :

If P is a measure on $(C_{o,T}, \mathcal{F}_{o,T})$ satisfying (2.40) or (2.41), and if one of the two following assumptions hold :

(i) $\frac{d}{dt} P_{o,t} \in \mathcal{M}^b_{o,t}$

(ii) $\partial_P V(t,P_{o,t}) \cap C^1_b(R^{n+1}) \neq \emptyset$

Then, the directional derivative of $V(t,P_{o,t})$ in the direction $\frac{d}{dt} P_{o,t}$ exists and is given by :

$$(2.83) \quad \lim_{\varepsilon \to o_+} \frac{1}{\varepsilon}(V(t,P_{o,t+\varepsilon}) - V(t,P_{o,t})) = \min_{\Lambda \in \partial_P V(t,P_{o,t})} \langle \Lambda, \frac{dP_{o,t}}{dt} \rangle .$$

Proof : Remark firstly that $\frac{d}{dt} P_{o,t}$ is not in general a measure on $C_{o,t}$ since, by (2.41), $\frac{d}{dt} P_{o,t}$ is a linear functional on $C^1_b(R^{n+1})$, and thus a distribution of order 1 (see [27]) on R^{n+1}, and is defined on a smaller space than $C^o_b(C_{o,t})$. Consequently, it is clear that (2.83) does not make sense for $\Lambda \in C^o_b(C_{o,t})$ in general. But if $\frac{dP_{o,t}}{dt} \in \mathcal{M}^b_{o,t}$, (2.83) is the classical formula of the directional derivative of a concave subdifferentiable function (see [26]). On the other hand, if $\tilde{\Lambda} \in \partial_P V(t,P_{o,t}) \cap C^1_b(R^{n+1})$, we have, by definition of a subgradient :

$$(2.84) \quad \frac{1}{\varepsilon}(V(t,P_{o,t+\varepsilon}) - V(t,P_{o,t})) \leq \langle \tilde{\Lambda}, \frac{1}{\varepsilon}(P_{o,t+\varepsilon} - P_{o,t}) \rangle .$$

and since $\langle \tilde{\Lambda}, \frac{dP_{o,t}}{dt} \rangle$ is finite by assumption, the right-hand side of (2.84) can be majorized by a finite constant for ε sufficiently small.

Conversely, since V is concave and finite, one has :

$$(2.85) \quad -\infty < V(t,P_{o,t+\varepsilon} + P_{o,t}) - V(t,P_{o,t}) \leq \frac{1}{\varepsilon}(V(t,P_{o,t+\varepsilon}) - V(t,P_{o,t})), \ \forall \varepsilon > o,$$

thus :

$1 = \lim_{\varepsilon \to o_+} \frac{1}{\varepsilon}(V(t,P_{o,t+\varepsilon}) - V(t,P_{o,t}))$ exists and is finite, and, by (2.84), $1 \leq \langle \tilde{\Lambda}, \frac{dP_{o,t}}{dt} \rangle$. Thus (2.83) follows by using once more the formula of the directional derivative for concave subdifferentiable functions. ■

Corollary 2.2 :

Suppose that the function w of Proposition 2.7 is differentiable with respect to z on R^{n+1}, and with respect to P in the following sense : $\frac{\partial w}{\partial P}(t,z\ ;\ P_{o,t})$ is supposed to be given by a kernel $\{\frac{\partial w}{\partial P}\}$ defined by :

$$(2.86)\quad \langle \frac{\partial w}{\partial P}(t,z\ ;\ P_{o,t}),Q\rangle = \int_{R^{n+1}} \{\frac{\partial w}{\partial P}\}(t,z,\zeta)dQ_t(\zeta) \quad \forall Q \in \mathcal{M}^b_{o,t}, \text{ with } Q_t = pr_t(Q),$$

and with $\{\frac{\partial w}{\partial P}\}$ differentiable with respect to ζ.
Then (2.83) becomes :

$$(2.87)\quad \lim_{\varepsilon \to o_+} \frac{1}{\varepsilon}\,(V(t,P_{o,t+\varepsilon}) - V(t,P_{o,t})) =$$
$$\int_{C_{o,t}\times\Omega} \frac{\partial}{\partial z}(w(t,z;P_{o,t}) + \int_{R^{n+1}} \{\frac{\partial w}{\partial P}\}(t,\zeta,z)dP_t(\zeta))\cdot$$
$$\cdot \tilde{F}_\omega(Z_{o,t})dP_{o,t}(Z_{o,t})\cdot d\rho(\omega)$$

and thus :

$$(2.88)\quad \{w(t,.\ ;\ P_{o,t}) + \int_{R^{n+1}} \{\frac{\partial w}{\partial P}\}(t,\zeta,.)dP_t(\zeta)\} = \partial_P V(t,P_{o,t}).$$

Proof : Combining (2.88) and (2.72), one obtains :

$$1 = \lim_{\varepsilon \to o_+} \frac{1}{\varepsilon}[V(t,P_{o,t+\varepsilon}) - V(t,P_{o,t})] =$$
$$\lim_{\varepsilon \to o_+} \frac{1}{\varepsilon} \int_{R^{n+1}\times\Omega} [w(t,Z^\omega_{t+\varepsilon}(t,Z_{o,t});P_{o,t+\varepsilon}) - w(t,Z_{o,t}(t);P_{o,t})]\cdot$$
$$\cdot dP_t(Z_{o,t}(t)d\rho(\omega)$$

$$(2.89)\quad 1 = \int_{C_{o,t}\times\Omega} \frac{\partial w}{\partial z}(t,z;P_{o,t})\cdot\tilde{F}_\omega(Z_{o,t})dP_{o,t}(Z_{o,t})d\rho(\omega)$$
$$+ \int_{R^{n+1}} \langle \frac{\partial w}{\partial P}(t,z;P_{o,t}), \frac{dP_{o,t}}{dt}\rangle dP_t(z)$$

by (2.40) and by the assumption on w's differentiability.

It remains to evaluate the last term of the right-hand side of (2.89).

By assumption :

$$(2.90)\quad \int_{R^{n+1}} \left\langle \frac{\partial w}{\partial P}(t,z;P_{o,t}),\ \frac{dP_{o,t}}{dt} \right\rangle dP_t(z) =$$

$$\int_{R^{n+1}\times R^{n+1}} \left\{\frac{\partial w}{\partial P}\right\}(t,z,\zeta)\,dP_t(z)\,d\left(\frac{dP_{o,t}}{dt}\right)(\zeta)$$

$$= \int_{C_{o,t}\times C} \frac{\partial}{\partial \zeta}\left(\int_{R^{n+1}} \left\{\frac{\partial w}{\partial P}\right\}(t,z,\zeta)\,dP_t(z)\right)\circ \tilde{F}_\omega(\zeta_{o,t})\,dP_{o,t}(\zeta_{o,t})\,d\rho(\omega),$$

applying once more (2.40). Thus (2.87) follows after the exchange of z and ζ in the right-hand side of (2.90). Finally, since V is differentiable (since w is), there is exactly one subgradient, thus given by (2.88). ■

Remark 2.11 : Since $\frac{d}{dt} P_{o,t}$ is not in general a measure, the existence of subgradients of V is not sufficient to compute the variation of V along a trajectory P^{Z^u}, and we finally have to assume regularity on the subgradients or on $\frac{d}{dt} P_{o,t}$. Remark that if $P_{o,t}$ is such that $P_t = pr_t(P_{o,t})$ is absolutely continuous with respect to the lebesgue measure of R^{n+1} with a smooth density and if $\tilde{F}_\omega(Z_{o,t})$ is regular enough, one can prove, by integration by parts in the right-hand side of (2.41), that $\frac{d}{dt} P_{o,t} \in \mathcal{P}^b_{o,t}$ - Namely, if $p_t(z)$ is the density of P_t with respect to dz :

$$\left\langle \varphi, \frac{d}{dt}(P_{o,t}) \right\rangle = \int_{C_{o,t}} \frac{\partial \varphi}{\partial z}(z)\circ \tilde{F}_\omega(Z_{o,t})P(s,Z_{o,s}\ ;\ t,dZ_{o,t})$$

$$= \int_{R^{n+1}} \frac{\partial \varphi}{\partial z}(z).\left(\int_{C_{o,t}} \tilde{F}_\omega(Z_{o,t})P(s,Z_{o,s}\ ;\ t,dZ_{o,t}\,|\,Z_{o,t}(t) = z)\right)p_t(z)dz$$

$$= -\int_{R^{n+1}} \varphi(z)\,\mathrm{div}_z\left(\left(\int_{C_{o,t}} \tilde{F}_\omega(Z_{o,t})P(s,Z_{o,s}\ ;\ t,dZ_{o,t}\,|\,Z_{o,t}(t) = z)\right)p_t(z)\right)dz$$

and the right-hand side can be defined $\forall \varphi \in C^o_K(R^{n+1})$ (space of continuous functions with compact support), where we have noted : $\mathrm{div}_z \psi = \sum_{i=1}^{n+1} \frac{\partial \psi_i}{\partial z_i}$ for $\psi = (\psi_1,\ldots,\psi_{n+1})$ of class C^1. Thus, if p_t has compact support in R^{n+1} the equality can also be written $\forall \varphi \in C^o_b(R^{n+1})$ and this proves that $\frac{d}{dt}(P_{o,t}) \in \mathcal{P}^b_{o,t}$. In this case, $\frac{d}{dt}(P_{o,t})$ coincides with the Lie derivative of $P_{o,t}$ in the direction $\tilde{F}_\omega$.

On the other hand, the assumption (ii) of Proposition 2.9, although strange, is natural in the case $\mathcal{K}_k(t) \equiv t \quad \forall t, \forall k$, since in this memoryless case the measures $P_{o,t}$ on $C^o([o,t]; R^{n+1})$, degenerate to measures P_t on R^{n-1}. Thus $\partial_P V(t,P_t) \subset C_b^o(R^{n+1})$ and (ii) means that we assume that there is at least one subgradient which is not only continuous and bounded on R^{n+1}, but also differentiable. In the general case, (ii) means that firstly there exists a subgradient in $C_b^o(C^o([o,t]; R^{n+1}))$ that depends only on the point $Z_{o,t}(t) = z$, and secondly that this dependence is differentiable with bounded and continuous partial derivatives with respect to z.

Another case of interest is when $N = 1$, $\mathcal{K}(t) \equiv o \quad \forall t$ (control with partial observations and classical information) where $P_{o,t}$ is replaced by the filter, i.e. the conditional probability measure of $z = Z_{o,t}(t)$, knowing all the past observations $Y_t(o,z_o)$. In this case also the conditional measure is a measure on R^{n+1} (parametrized by $Y_t(o,z_o)$) and $\partial_P V(t,P_t(\cdot|Y_t(o,z_o)))$ is a subset of $C_b^o(R^{n+1})$, as in the case $\mathcal{K}_k(t) \equiv t$. ■

II.4. The Hamilton-Jacobi-Bellman equations and the Signalling

Let $\mathcal{M}_{o,t}^{1,+}$ denote the probability measures on $C_{o,t}$.

Theorem 2.2 : Assume that V is differentiable with respect to t and that $\frac{\partial V}{\partial t}(t,P)$ is finite $\forall t \in [o,T]$, $\forall P \in \mathcal{M}_{o,t}^{1,+}$, and assume finally that :

(F) $\quad \forall t \in [o,T]$, $\forall P_{o,t} \in \mathcal{M}_{o,t}^{1,+}$, $\exists u \in \tilde{\mathcal{U}}$, $\exists \tilde{Z}^u \in \tilde{Z}^u(t)$, $\exists \Lambda_o \in \partial_P V(t,P_{o,t})$ such that $< \Lambda_o, \frac{d}{dt} P_{o,t}^{\tilde{Z}^u} >$ is well-defined and finite.

Then V satisfies the Hamilton-Jacobi-Bellman equation :

$$(2.91) \qquad \frac{\partial V}{\partial t}(t,P_{o,t}) + \inf_{\substack{u \in \mathcal{U}_1 \times \ldots \times \mathcal{U}_N \\ u_k : \mathcal{Y}_k(t)\text{-measurable} \\ \forall k=1,\ldots,N}} \min_{\Lambda \in \partial_P V(t,P_{o,t})} < \Lambda, \frac{d}{dt} P_{o,t}^u > = o$$

$$(2.92) \qquad V(T,P_{o,t}) = <\tilde{G},P_T> \quad \forall P_{o,T} \in \mathcal{M}_{o,T}^{1,+}, \quad P_T = pr_T(P_{o,T}).$$

Conversely (here (H) is not supposed to hold), if V is differentiable with respect to t with $\frac{\partial V}{\partial t}(t,P_{o,t})$ finite $\forall t$, $P_{o,t}$, concave and subdifferentiable in P, and satisfies (2.91), (2.92), then $V(t,P_{o,t}) = \operatorname{Inf}_{u\in\mathcal{U}} \tilde{J}_{t,P_{o,t}}(u)$ $\forall t \in [0,T]$, $\forall P_{o,t} \in \mathcal{M}^{1,+}_{o,t}$.

Proof : from (2.33) and (2.67), we have :

$$(2.93)\qquad \operatorname*{Inf}_{\substack{u\in\mathcal{U}\\ \tilde{Z}^u\in\tilde{Z}^u(t)}} \frac{1}{\varepsilon}\left[V(t+\varepsilon,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t})) - V(t,P_{o,t})\right] = 0 \quad \forall \varepsilon > 0.$$

or :

$$(2.94)\qquad \operatorname*{Inf}_{\substack{u\in\mathcal{U}\\ \tilde{Z}^u\in\tilde{Z}^u(t)}} \frac{1}{\varepsilon}\left[\left(V(t+\varepsilon,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t})) - V(t,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t}))\right) + \left(V(t,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t})) - V(t,P_{o,t})\right)\right] = 0$$

By assumption, we have :

$$(2.95)\qquad \lim_{\varepsilon\to 0_+} \frac{1}{\varepsilon}\left(V(t+\varepsilon,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t})) - V(t,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t}))\right) = \lim_{\varepsilon\to 0_+} \frac{1}{\varepsilon}\left[V(t+\varepsilon,P_{o,t}) - V(t,P_{o,t}) + o(\varepsilon)\right]$$

$= \frac{\partial V}{\partial t}(t,P_{o,t})$, since V is uniformly continuous on the compact $[t,t+\varepsilon_o] \times \{P^{\tilde{Z}^u}_{o,t+\varepsilon} \mid 0 \leq \varepsilon \leq \varepsilon_o\}$.

On the other hand, by (2.83) and (H), with u as in (H), we have :

$$\lim_{\varepsilon\to 0_+} \frac{1}{\varepsilon}\left[V(t,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t})) - V(t,P_{o,t}) = \operatorname*{Min}_{\Lambda\in\partial_P V(t,P_{o,t})} \left\langle \Lambda, \frac{d}{dt} P^{\tilde{Z}^u}_{o,t}\right\rangle$$

$$\geq \lim_{\varepsilon\to 0_+} \operatorname*{Inf}_{\substack{u\in\mathcal{U}\\ \tilde{Z}^u\tilde{Z}^u(t)}} \frac{1}{\varepsilon}\left[V(t,\tilde{Z}^u_{t,t+\varepsilon}(P_{o,t})) - V(t,P_{o,t})\right] = \operatorname*{Inf}_{\substack{u\in\mathcal{U}\\ u=\mathcal{Y}(t)\text{-measurable}\\ \Lambda\in\partial_P V(t,P_{o,t})}} \left\langle \Lambda, \frac{d}{dt} P^{\tilde{Z}^u}_{o,t}\right\rangle$$

$$= -\frac{\partial V}{\partial t}(t,P_{o,t}) \quad \text{by (2.94) and (2.95)},$$

and this proves (2.91), (2.92) following from the definition (2.29) at time T.

Conversely, if V is such that $\frac{\partial V}{\partial t}$ is finite everywhere and such that $\partial_P V(t,P_{o,t}) \neq \emptyset$ everywhere, and if V satisfies (2.91), (2.92), then clearly :

$$\inf_{\substack{u\in\mathcal{U} \\ \tilde{Z}^u\in\tilde{\mathcal{Z}}^u(t)}} \frac{d}{ds} V(s,P^{\tilde{Z}^u}_{o,s})\Big|_{s=t} = o \qquad \forall t \in [o,T]$$

Thus, integrating between s and T, with $o \leqslant s \leqslant T$:

$$V(T,P^{\tilde{Z}^u}_{o,T}) \geqslant V(s,P_{o,s}) \ , \quad \forall u \in \tilde{\mathcal{U}}, \quad \forall\, \tilde{Z}^u \in \tilde{\mathcal{Z}}^u(s), \quad \forall P_{o,s} \in \mathcal{M}^{1,+}_{o,s} \ , \quad \text{and}$$

$$V(T,P^{\tilde{Z}^u}_{o,T}) = \langle \tilde{G}, P^{\tilde{Z}^u}_T \rangle \ , \quad \text{with} \quad P^{\tilde{Z}^u}_T = \tilde{Z}^u_T(P_{o,s}), \quad \text{and} \quad P^{\tilde{Z}^u}_{o,T}\big|[o,s] = P_{o,s}.$$

Also : $\inf_{\substack{u\in\mathcal{U} \\ \tilde{Z}^u\in\tilde{\mathcal{Z}}^u(t)}} V(T,P^{\tilde{Z}^u}_{o,T}) = V(s,P_{o,s}) \ ,$

so that, finally : $V(s,P_{o,s}) = \inf_{u\in\mathcal{U}} \inf_{\tilde{Z}^u\in\tilde{\mathcal{Z}}^u(s)} \langle \tilde{G}, P^{\tilde{Z}^u}_T \rangle$

and thus : $V(s,P_{o,s}) = \inf_{u\in\mathcal{U}} \tilde{J}_{s,P_{o,s}}(u)$, by proposition 2.6, and the theorem is proved. ■

Remark 2.12 : The converse part of the theorem is in fact a uniqueness result of the solution of (2.91), (2.92) in the class of functions which are differentiable in t and concave and subdifferentiable in P. ■

Corollary 2.3 :

Let w, defined by Proposition 2.7, satisfy :

$$\frac{\partial w}{\partial t}(.,.,P), \ \frac{\partial w}{\partial z_i}(.,.,P) \in L^\infty([o,T] \times R^{n+1}) \quad \forall i = 1,\ldots,n+1, \quad \forall P \in \mathcal{M}^{1,+}_{o,T} \ ,$$

and with $\frac{\partial w}{\partial P}$ satisfying (2.86). Then w satisfies the generalized Hamilton - Jacobi - Bellman equations :

$$(2.96)\qquad \inf_{\substack{u_k\in\mathcal{U}_k\\ u_k=\mathcal{Y}_k(t)\text{-measurable}\\ \forall k=1,\dots,N}} \int_{C_{o,t}\times\Omega} \Big[\frac{\partial w}{\partial t}(t,z;P_{o,t})+\frac{\partial \lambda}{\partial z}(t,z;P_{o,t}).F(t,z,u(t,\mathcal{H}_\nu(Z_{o,t})),v_j)\Big]\, dP_{o,t}(Z_{o,t})d\rho(\omega) = o$$

$$\forall t\in[o,T],\quad \forall P_{o,t}\in \mathcal{M}^{1,+}_{o,T},$$

$$(2.97)\qquad w(T,z\ ;\ P_{o,T}) = \tilde{G}(z)\quad \forall z\in R^{n+1},\quad \forall P_{o,T}\in \mathcal{M}^{1,+}_{o,T}.$$

with :

$$(2.98)\qquad \lambda(t,z\ ;\ P_{o,t}) = w(t,z\ ;\ P_{o,t}) + \int_{R^{n+1}} \{\frac{\partial w}{\partial P}\}(t,\zeta\ ;\ z)dP_t(\zeta)$$

and $u(t,\mathcal{H}_\nu(Z_{o,t})) = (u_1(t,\mathcal{H}^1_\nu(Z_{o,t})),\dots,u_N(t,\mathcal{H}^N_\nu(Z_{o,t})))$,

$$\mathcal{H}^k_\nu(Z_{o,t}) = \{H_k(s,Z_{o,t}(s),\nu^k_j)\,|\,s\in[\mathcal{H}_k(t),t]\cap[t_j,t_{j+1}[,\ \forall j=o,\dots,M-1\}[$$

$$\forall \nu^k_j = pr^k_j(\omega),\qquad \forall k=1,\dots,N.$$

Consersely, if w satisfies (2.96), (2.97), (2.98), under the preceding assumptions, then $V(t,P_{o,t}) = \int_{R^{n+1}} w(t,z\ ;\ P_{o,t})dP_t(z) = \inf_{u\in\mathcal{U}} \tilde{J}_{t,P_{o,t}}(u)$, $\forall P_{o,t}\in \mathcal{M}^{1,+}_{p;t}$, $\forall t\in[o,T]$.

<u>Proof</u> : We have, by assumption : $\frac{\partial V}{\partial t}(t,P_{o,t}) = \int_{R^{n+1}} \frac{\partial w}{\partial t}(t,z;P_{o,t})dP_t(z)$, and by (2.87), (2.88) and (2.91) (the assumptions of theorem 2.2. being trivially satisfied) :

$$(2.99)\qquad \inf_{\substack{u_k\in\mathcal{U}_k\\ u_k=\mathcal{Y}_k(t)\text{-measurable}\\ \forall k=1,\dots,N}} \int_{C_{o,t}\times\Omega} \Big[\frac{\partial w}{\partial t}(t,z;P_{o,t})+\inf_{\tilde{F}_{u,\omega}(Z_{o,t})\in co\Phi_{u,\omega}(Z_{o,t})} \frac{\partial \lambda}{\partial z}(t,z;P_{o,t}).$$
$$.\tilde{F}_{u,\omega}(Z_{o,t})\Big]dP_{o,t}(Z_{o,t})d\rho(\omega) = o$$

with $\Phi_{u,\omega}(Z_{o,t}) = \bigcap_{\mathcal{V}(Z_{o,t})} F(t,Z_{o,\tau}(t),u(t,\mathcal{H}_\nu(\mathcal{V}(Z_{o,t}))),v_j)$ with $t\in[t_j,t_{j+1}[$ and $v = pr^o_j(\omega)$, $\nu^k_i = pr^k_i(\omega)$, and $\mathcal{V}(Z_{o,\tau})$ being any neighborhood of $Z_{o,t}$ in $C_{o,t}$ (see Appendix).

But, since the Inf on $\Phi_{u,\omega}(Z_{o,t})$ is the same as the one on $\text{co}\,\Phi_{u,\omega}(Z_{o,t})$, which, on turn, is also the same as the one on :

$$\{F(t,Z_{c,t}(t)u,v_j) \mid u_k \in \mathcal{U}_k \ , \ u_k = \mathcal{Y}_k(t)\text{-measurable}, \quad \forall k = 1,\dots,N\} \ ,$$

then (2.96) is proved. (2.97) follows trivially from (2.92) and (2.72). The proof of the converse is exactly the same as in the theorem 2.2 and is left to the reader. ■

Corollary 2.4 :

Under the proceding assumption, the optimal strategy u_k^* for station k is determined at each time t, for each observation path $Y_k(t)$, and each P $P_{o,t} \in \mathcal{M}_{o,t}^{1,+}$ by :

$$(2.100) \quad \underset{u_k \in U_k}{\text{Min}} \int_{C_{o,t}\times\Omega} \frac{\partial\lambda}{\partial z}(t,z;P_{o,t}) \circ F(t,z,u_k^*(t,\mathcal{H}_\nu(Z_{o,t}),P_{o,t}),u_k,v_j)\, d(P_{o,t}\otimes\rho)(Z_{o,t},\omega \mid Y_k(t))$$

with : $u_k^*(t,\mathcal{H}_\nu(Z_{o,t}),P_{o,t}) = (u_1^*(t,\mathcal{H}_\nu^1(Z_{o,t}),P_{o,t}),\dots$

$$,\dots,u_{k-1}^*(t,\mathcal{H}_\nu^{k-1}(Z_{o,t}),P_{o,t}),u_{k+1}^*(t,\mathcal{H}_\nu^{k+1}(Z_{o,t}),P_{o,t}),\dots$$

$$,\dots,u_N^*(t,\mathcal{H}_\nu^N(Z_{o,t}),P_{o,t})),$$

u_k^* being the optimal strategy of all the stations except k, and with λ given by (2.98).

Proof : trivial from (2.96). ■

Remark 2.13 : (2.100) shows that this problem can be solved by Hamiltonian techniques. Nevertheless, this Hamiltonian is far to be the classical one since the adjoint $\frac{\partial\lambda}{\partial z}$ is given by (2.98). One can see that $\frac{\partial\lambda}{\partial z} = \frac{\partial w}{\partial z} + \mu$, $\frac{\partial w}{\partial z}$ being the classical interpretation of the adjoint (the sensitivity of the cost for a variation of control, the probability measure being fixed) and $\mu = \int \frac{\partial}{\partial z}\left\{\frac{\partial w}{\partial P}\right\}(t,\zeta;z)\,dP_t(\zeta)$ can be interpreted as the variation of cost induced by a variation of information (and thus of probability measure), the controls being fixed. To summarize, the interpretation of the adjoint is :

(2.101)

adjoint = sensitivity of cost w.r. to control + sensitivity of cost w.r. to information

This phenomenon has been intuitively noted by H.S. Witsenhausen [31] in 1968, and has received the name of signalling (see also [10],[17],[18],[19],[21],[22],[30], [32],[33]). For the first time in [22], this notion has received a rigorous definition in the discrete-time case. Finally, one can see the optimization problem (2.100) as a trade-off between cheap controls, but containing poor informations, and expensive controls containing more information to recover the state $Z_{o,t}$ as precisely as possible. In this point of view, μ can also be interpreted as a normal vector to the direction where the learning is the "steepest". ■

Remark 2.14 : The only time when we need to take conditional expectation with respect to observations, is to compute the optimal u* by (2.100). Once $u_k^*(t,Y_k(t),P_{o,t})$ is computed $\forall t$, $\forall Y_k(t)$, $\forall P_{o,t}$, one can compute w by (2.96), (2.97), by replacing $Y_k(t)$ in (2.96) by $H_\nu(Z_{o,t})$, and $P_{o,t}$ by $P^{Z^{u^*}}$ obtained through (2.39), (2.41). We shall see more precisely how to handle all these quantities in the examples of section IV. ■

Remark 2.15 : A similar equation to (2.91) was first derived by Mortensen [23], in the classical information case, giving rise to a partial differential equation with derivatives with respect to the conditional density. (see also [1], [7],[11],[14],[35]) of the same nature as $\frac{\partial V}{\partial P}$. However, the reason why we obtain here more precise conditions lies on the representation (2.72) of the value function. ■

Remark 2.16 : A partial statement of the Corollaries 2.3, 2.4 can be found in Quadrat [25] in the case $H_k(t) \equiv t$ $\forall t \in [o,T]$, $\forall k = 1,\ldots,N$, with an adjoint computed via an algorithm of Howard's type but without explicit formula for the adjoint, and without convergence proof of the algorithm. ■

III - THE TEAM PROBLEM FOR DIFFUSIONS

As announced in the introduction, we shall follow almost the same lines as in section II for discrete-time noises. In fact, in spite of appearences, this case will appear easier, essentially because of the nice and simple formalism of the problem of martingale and of Girsanow's absolutely continuous change of measures, and this whatever the complexity of the information structure.

We recall that the five paragraphs of this section are :

III.1. Model and assumptions.

III.2. The dynamic programming method in the space of bounded measures.
- III.2.1. Reformulation into a purely final cost.
- III.2.2. The value function and the optimality principle.
- III.2.3. The closed-loop formulation.

III.3. Some regularity properties of the value function.

III.4. The Hamilton-Jacobi-Bellman equations and the signalling.

III.5. Application to the control of partially observed diffusions with classical information
- III.5.1. The Hamilton-Jacobi-Bellman theory and the maximum principle.
- III.5.2. The link with Mortensen's equation.

III.1. Model and assumptions

III.1.1. The dynamics and observations equations

Let us consider the following stochastic idfferential system on $[0,T]$:

$$(3.1)\qquad \begin{cases} dx_t = f_o(t,x,u_t)dt + f_1(t,x_t)dv_t^o \\ dy_t^k = h_o^k(t,x,y_t^k)dt + h_1^k(t,y_t^k)dv_t^k, \quad k=1,\dots,N. \end{cases}$$

where the state x_t lies in R^n and station k's observation y_t^k at time t, lies in R^{q_k} $\forall k=1,\dots,N$, with $\sum_{k=1}^{N} q_k = q$.

As in the preceding section, we shall focus attention on the probability measure generated by the process (x_t,y_t) for each strategy N-tuple $u = (u_1,\dots,u_N)$ adapted to a suitable observation σ-field, and the approach via the problem of martingales and Girsanow's transformations is particularly appropriate. For this purpose, let us give a precise statement of the problem and of the assumptions. The basic probability space $(\Omega, \mathcal{F}, \rho)$ is defined by :

$$(3.2)\qquad \begin{cases} \Omega = C^o([o,T]; R^{n+q}) \overset{def}{=} C_{o,T} \\ \mathcal{F}_t = \mathcal{B}\{(X_s,Y_s) \mid s \in [o,t]\} : \text{the borel σ-field generated by the family} \\ \text{of coordinate functions } \xi_s = (X_s,Y_s) = (X_s,Y_s^1,\dots,Y_s^N), \text{ with} \\ \omega(s) = (X_s(\omega), Y_s^1(\omega),\dots,Y_s^N(\omega)) = \xi_s(\omega), \ \forall s \in [o,T], \ \forall \omega \in \Omega, \\ \mathcal{F} = \mathcal{F}_T \text{ and } \rho : \text{the Wiener measure on } (\Omega,\mathcal{F}). \end{cases}$$

The station k's observation σ-field $\{\mathcal{Y}_t^k \mid o \leq t \leq T\}$ is defined by :

$$(3.3)\qquad \mathcal{Y}_t^k = \mathcal{B}\{Y_s^k \mid \nu_k(t) \leq s \leq t\}, \quad k=1,\dots,N, \quad t \in [o,T].$$

with ν_k defined as before.

Let us denote $\xi = (x,y)' \in R^{n+q}$, and :

$$(3.4)\quad \eta_1(t,\xi) = \begin{pmatrix} f_1(t,x) & & & 0 \\ & h_1^1(t,y^1) & & \\ & & \ddots & \\ 0 & & & h_1^N(t,y^N) \end{pmatrix} \in R^{(n+q)\times(n+q)}$$

$$(3.5)\qquad a(t,\xi) = \eta_1(t,\xi)\eta_1'(t,\xi)$$

We assume that $\eta_1 : [o,T] \times R^{n+q} \to R^{(n+q)\times(n+q)}$ is continuous, everywhere invertible and :

$$(3.6)\quad \begin{cases} |\eta_1^{i,j}(t,\xi) - \eta_1^{i,j}(t,\zeta)| \leq A_1\|\xi-\zeta\|, \ \forall\, \xi,\ \zeta \in R^{n+q}\ , \ \forall t \in [o,T], \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall i,j=1,\ldots,n+q \\ \eta_1 \ \text{ and } \ \eta_1^{-1} \ \text{ are uniformly bounded.} \end{cases}$$

Let us also denote :

$$\eta_o(t,\xi,u) = (f_o(t,x,u)', h_o^1(t,x,y^1)', \ldots, h_o^N(t,x,y^N)')'$$

with $\eta_o : [o,T] \times C_{o,T} \times U \to R^{n+q}$ continuous, and :

$$(3.7)\quad \begin{cases} (t,\xi) \to \eta_o(t,\xi,u) \ \ \mathcal{F}_t \text{ adapted } \ \forall u \in U,\ \eta_o(t,\xi,U) \ \text{ is convex } \ \forall t,\xi. \\ \underset{(t,u)\in[o,T]\times U}{\text{Sup}} \|\eta_o(t,\xi,u)\| \leq A_o(1 + \underset{s\in[o,t]}{\text{Sup}} \|\xi(s)\|) \ \ \forall \xi \in C_{o,T}. \end{cases}$$

where $U = U_1 \times \ldots \times U_N$, as in the preceding section, U_k being a given closed subset of R^{p_k} $\forall k=1,\ldots,N$.

Definition 3.1 : $\forall k=1,\ldots,N$, we say that $u^k : [o,T] \times \Omega \to U_k$ is admissible iff :

(i) u^k is $\mathcal{Y}^k$ adapted, namely : u_t^k is $\mathcal{Y}_t^k$ measurable $\forall \tau \in [o,T]$, and $u^k_{|[s,t]}$ is $\mathcal{B}_{[s,t]} \times \mathcal{Y}^k(s,t)$ measurable $\forall o \leq s \leq t \leq T$, with $\mathcal{Y}^k(s,t) = \underset{s\leq\sigma\leq t}{\cup} \mathcal{Y}_\sigma^k$,

(ii) $u_t^k \in U^k$ $\forall t \in [o,T]$.

We denote $\mathcal{U}_k$ the set of all admissible strategies for station k, and $\mathcal{U} = \mathcal{U}_1 \times \ldots \times \mathcal{U}_N$. ∎

In order to define a solution to :

$$(3.8)\quad \begin{cases} d\xi_t = \eta_o(t,\xi,u_t(Y))dt + \eta_1(t,\xi_t)dv_t & \forall t \geq s \\ \xi(\sigma) = \xi_{o,s}(\sigma) \quad \forall \sigma \in [o,s] \end{cases}$$

with $\xi_{o,s}$ random initial function in $C_{o,s}$, $u \in \mathcal{U}$, and $s \in [o,T]$, and, more precisely, to give a sense to the measure generated by ξ and u in (3.8) from μ_s a given bounded measure on $C_{o,s}$, let us introduce the following :

Definition 3.2 : Let $u \in \mathcal{U}$, $s \in [o,T]$, and $\mu_s \in \mathcal{M}^b_{o,s}$ (the space of bounded measures on $C_{o,s}$ endowed with the weak*-topology). We say that the bounded measure ν^u_{s,μ_s} solves the problem of martingale relatively to $(\eta_o,\eta_1,u,s,\mu_s)$ iff :

(i) $\nu^u_{s,\mu_s}(\xi(\sigma) \in B, \ o \leqslant \sigma \leqslant s) = \mu_s(B)$, $\forall B \in \mathcal{F}^s_o = \bigcup_{o \leqslant \sigma \leqslant s} \mathcal{F}_\sigma$.

(ii) $\forall t \geq s$, $\forall \varphi \in C^2_b(\Omega)$:

$$\varphi(\xi(t)) - \varphi(\xi(s)) - \int_s^t [\eta_o'(\sigma,\xi,u_\sigma(Y))\frac{\partial \varphi}{\partial \xi}(\xi(\sigma)) + \frac{1}{2}\mathrm{tr}(a(\sigma,\xi(\sigma))\frac{\partial^2 \varphi}{\partial \xi^2}(\xi(\sigma)))]d\sigma$$

is a $(\mathcal{F}_t, \nu^u_{s,\mu_s})$-martingale. ∎

Remark 3.1 : if ν^u_{s,μ_s} solves the martingale problem for $(\eta_o,\eta_1,u,s,\mu_s)$, let us show how a solution to (3.8) is deduced :

Let us denote $M^u_t = \xi_t - \xi_s - \int_s^t \eta_o(\sigma,\xi,u_\sigma(Y))d\sigma$.

Taking $\varphi_i(\xi) = \xi_i$, $i=1,\dots,n+q$, in (ii), we see that M^u_t is a local ν^u_{s,μ_s}-martingale after s. Also :

$\exp[\int_o^t \theta'(\sigma)dM^u_\sigma - \frac{1}{2}\int_s^t \theta'(\sigma)a(\sigma,\xi(\sigma))\theta(\sigma)d\sigma]$ is a local ν^u_{s,μ_s}-martingale after s, for every $\mathcal{F}$-adapted process θ.

Thus, taking $\theta(\sigma) = \eta_1'^{-1}(\sigma,\xi(\sigma))\overline{\theta}$ with $\overline{\theta} \in R^{n+q}$, we obtain that :

$\exp[\overline{\theta}'\int_s^t \eta_1^{-1}(\sigma,\xi(\sigma))dM^u_\sigma - \frac{\|\overline{\theta}\|^2}{2}(t-s)]$ is a ν^u_{s,μ_s}-martingale which proves that

$v^u_t = \int_s^t \eta_1^{-1}(\sigma,\xi(\sigma))dM^u_\sigma$ is a $(\mathcal{F}, \nu^u_{s,\mu_s})$- Wiener process. Finally we have :

$\eta_1(t,\xi(t))dv^u_t = dM^u_t = d\xi_t - \eta_o(t,\xi,u_t(Y))dt$, or :

$$d\xi_t = \eta_o(t,\xi,u_t(Y))dt + \eta_1(t,\xi_t)dv_t^u$$

with $\xi(\sigma) = \xi_{o,s}(\sigma) \quad \forall \sigma \in [o,s]$, $\xi_{o,s}$ = random function with law μ_s.
One can also easily check that ν^u_{s,μ_s} is the image of $\rho \otimes \mu_s$ by ξ since ν^u_{s,μ_s}'s evolution is given by (ii) of definition 3.2. ∎

Remark 3.2 : We have introduced general bounded measures μ_s and ν^u_{s,μ_s} rather than probability measures since this extension is needed in the sequel. ∎

Lemma 3.1 : Under the preceding assumptions, $\forall u \in \mathcal{U}$, $\forall s \in [o,T]$, $\forall \mu_s \in \mathcal{M}^b_{o,s}$, there exists a unique solution ν^u_{s,μ_s} to the problem of martingale associated to $(\eta_o,\eta_1,u,\ s,\ \mu_s)$.

Proof : Existence : Let $\nu_{s,\mu_s} = \int \rho(\cdot\,|\xi_{o,s})d\mu_s(\xi_{o,s})$, be defined, for any $B \in \mathcal{F}$, by:

$$(3.9) \qquad \nu_{s,\mu_s}(B) = \int 1_B(\xi(\omega))d\rho(\omega\,|\,\xi_\sigma(\omega) = \xi_{o,s}(\sigma) \quad \forall \sigma \in [o,s])d\mu_s(\xi_{o,s}).$$

where $\xi(\omega)$ is the unique strong solution of the stochastic differential equation ;

$$(3.10) \qquad d\xi(t) = \eta_1(t,\xi(t))dv_s(t), \ \forall t \geq s, \quad \xi(\sigma) = \xi_{o,s}(\sigma) \quad \forall \sigma \in [o,s],$$

with v_s a $\mathcal{F}_s^T$-Wiener process, ρ its associated measure, $v_s(\sigma) = o \quad \forall \sigma \in [o,s]$.
Let us denote :

$$(3.11) \qquad R^u_s(\omega) = \exp\Big\{\int_s^T (\eta_1^{-1}(\sigma,\xi(\sigma))\eta_o(\sigma,\xi,u_\sigma(Y)))\cdot dv_s(\sigma) - \int_s^T \|\eta_1^{-1}(\sigma,\xi(\sigma))\eta_o(\sigma,\xi,u_\sigma(Y))\|^2 d\sigma\Big\},$$

we have that $\nu^u_{s,\mu_s} \overset{\text{def}}{=} R^u_s \cdot \nu_{s,\mu_s}$ is such that $\forall t \geq s$:

$$v^u_s(t) \overset{\text{def}}{=} v_s(t) - \int_s^t \eta_1^{-1}(\sigma,\xi(\sigma))\eta_o(\sigma,\xi,u_\sigma(Y))d\sigma$$

is a $(\mathcal{F},\ \nu^u_{s,\mu_s})$ Wiener process (Girsanov theorem) and since :

$$\eta_1(t,\xi(t))dv^u_s(t) = d\xi(t) - \eta_o(t,\xi,u_t(Y))dt \qquad \forall t \geq s,$$

(ii) of definition 3.2 is an easy consequence of Itô's formula, and ν^u_{s,μ_s} solves the problem of martingale for $(\eta_o,\eta_1,u,s,\mu_s)$.

Uniqueness : let ν^u_{s,μ_s} be a solution to the problem of martingale for $(\eta_o,\eta_1,u,s,\mu_s)$. As in Remark 3.1, there exists a $\mathcal{F}^T_s$ - Wiener process v^u_s such that :

$$d\xi(t) = \eta_o(t,\xi,u_t(Y))dt + \eta_1(t,\xi(t))dv^u_s(t).$$

Thus, noting $\tilde{\nu}^u_{s,\mu_s}$ the bounded measure defined by :

$$\frac{d\tilde{\nu}^u_{s,\mu_s}}{d\nu^u_{s,\mu_s}} = (R^u_s)^{-1} , \tag{3.12}$$

we have that $\tilde{\nu}^u_{s,\mu_s}$ solves the problem of martingales for (o,η_1,s,μ_s) (independant of u) which is uniquely defined. Consequently $\tilde{\nu}^u_{s,\mu_s} = \nu_{s,\mu_s}$, and the uniqueness of ν^u_{s,μ_s} follows immediately. ∎

Let us now introduce the following family of bounded measures $\{\pi^u(t;s,\mu_s) \mid t \in [s,T],\ \mu_s \in \mathcal{M}^b_{o,s}\}$ defined by :

$$\begin{cases} \pi^u(t;s,\mu_s) \in \mathcal{M}^b_{o,t} \\ \pi^u(t;s,\mu_s)(B) = \nu^u_{s,\mu_s}(B) \quad \forall B \in \mathcal{F}^t_o = \bigcup_{o\leq\sigma\leq t} \mathcal{F}_\sigma \\ \forall t \in [s,T],\ \forall \mu_s \in \mathcal{M}^b_{o,s} . \end{cases} \tag{3.13}$$

Thus, in particular :

$$\pi^u(s;s,\mu_s) = \mu_s , \quad \pi^u(T;s,\mu_s) = \nu^u_{s,\mu_s} . \tag{3.14}$$

Lemma 3.2 : the family π^u forms a weakly * continuous semi-group on $\mathcal{M}^b_{o,T}$ and :

$$\nu^u_{s,\mu_\sigma} = \nu^u_{t,\pi^u(t;s,\mu_s)} \quad \forall t \in [s,T],\ \forall u \in \mathcal{U},\ \forall \mu_s \in \mathcal{M}^b_{o,s} . \tag{3.15}$$

Proof : let $\xi(.;s,\xi_{o,s},\omega^\bullet_s)$ denote the solution of

$$\begin{cases} d\xi_t = \eta_1(t,\xi_t)dv_s(t) \\ \xi(\sigma) = \xi_{o,s}(\sigma) \quad \forall \sigma \in [o,s] \end{cases} \tag{3.16}$$

where $\omega_s^\cdot$ is the application $t \to \omega_s^t$ for $t \geq s$, with :

$$(3.17) \qquad \omega_s^t = \{v_s(\sigma) - v_s(s) \mid s \leq \sigma \leq t\}.$$

Recall that, since ρ is the Wiener measure on $(\Omega,\mathfrak{F})$, we have :

$$(3.18) \qquad \rho(\omega_s^t \in B) = \rho(\omega_\sigma^t \in B \cap C_{\sigma,t})\ \rho(\omega_s^t \in B \cap C_{s,\sigma})$$

$$\forall B \in \mathfrak{F}_t\ ,\quad \forall s \leq \sigma \leq t.$$

The following formula holds :

$$(3.19) \int \varphi(\xi) d\pi^u(t;s,\mu_s)(\xi) = \int_{C_{o,s}\times\Omega} \varphi(\xi(\cdot;s,\xi_{o,s},\omega_s^\cdot)) R^u_{s,t}(\xi(\cdot;s,\xi_{o,s},\omega_s^\cdot)) d\mu_s(\xi_{o,s}) d\rho(\omega)$$

$$\forall \varphi \in C_b^o(C_{o,t}),\ \text{with :}$$

$$R^u_{s,t}(\xi) = \exp\Big[\int_s^t (\eta_1^{-1}(\sigma,\xi_\sigma)\eta_o(\sigma,\xi,u_\sigma(Y)))\cdot dv_s(\sigma) - \frac{1}{2}\int_s^t \|\eta_1^{-1}(\sigma,\xi_\sigma)\eta_o(\sigma,\xi,u_\sigma(Y))\|^2 d\sigma\Big]$$

$$\forall t \in [s,T].$$

since $$\int \varphi(\xi) d\pi^u(t;s,\mu_s)(\xi) = \int_{C_{o,t}} \varphi(\xi) d\nu^u_{s,\mu_s}(\xi)$$

$$= \int_{C_{o,t}} \varphi(\xi) R^u_{s,t}(\xi) d\nu_{s,\mu_s}(\xi)\ ,\quad \text{and since :}$$

$$\int \psi(\xi) d\nu_{s,\mu_s}(\xi) = \int_{C_{o,s}\times\Omega} \psi(\xi(\cdot;s,\xi_{o,s},\omega_s^\cdot)) d\mu_s(\xi_{o,s}) d\rho(\omega) \quad \forall\ \psi \in C_b^o(C_{o,T})$$

The formula (3.19) is proved.
But $\xi(\cdot;s,\xi_{o,s},\omega_s^\cdot)$ satisfies the following semi-group property :

$$(3.20) \qquad \xi(t;s,\xi_{o,s},\omega_s^t) = \xi(t;\sigma,\xi(\sigma;s,\xi_{o,s},\omega_s^\sigma),\omega_\sigma^t)$$

for almost every $(\omega,\xi_{o,s}) \in \Omega\times C_{o,s}$, $\forall s \leq \sigma \leq t$, and thus (3.19) becomes :

$$E_{\pi^u(t;s,\mu_s)}(\varphi) =$$

$$= \int_{C_{o,s}\times\Omega} \varphi(\xi(\circ;\sigma,\xi(\sigma;s,\xi_{o,s},\omega_s^\sigma),\omega_\sigma^\circ))\ R^u_{\sigma,t}(\xi(\circ;\sigma,\xi_{o,s},\omega_s^\sigma),\omega_\sigma^\circ) \times$$

$$\times R^u_{s,\sigma}(\xi(\circ;s,\xi_{o,s},\omega_s^\circ))d\mu_s(\xi_{o,s})d\rho(\omega)$$

$$= \int_{C_{o,\sigma}\times\Omega} \varphi(\xi(\circ;\sigma,\xi_{o,\sigma},\omega_\sigma^\circ))\ R^u_{\sigma,t}(\xi(\cdot;\sigma,\xi_{o,\sigma},\omega_\sigma^\circ))d\pi^u(\sigma;s,\mu_s)(\xi_{o,\sigma})d\rho(\omega)$$

(by (3.19) and (3.18)

$$= \int_{C_{o,t}} \varphi(\xi)d\pi^u(t;\sigma,\pi^u(\sigma;s,\mu_s))(\xi) \qquad \forall\varphi \in C_b^o(C_{o,t})$$

and thus :

$$(3.21)\quad \pi^u(t;s,\mu_s) = \pi^u(t;\sigma,\pi^u(\sigma;s,\mu_s)) \qquad \forall s \leqslant \sigma \leqslant t,\ \forall\mu_s\ ,\ \forall u.$$

And, together with (3.14), we have proved that π^u has the semi-group property. Also, since :

$$\pi^u(T;s,\mu_s) = \pi^u(T;t,\pi^u(t;s,\mu_s))$$

we obtain :

$$\nu^u_{s,\mu_s} = \nu^u_{t,\pi^u(t;s,\mu_s)}$$

and (3.15) is proved.
It remains to prove that π^u is weakly * continuous but this is obvious since $\forall\varphi \in C_b^o(\Omega)$:

$$\left|E_{\pi^u(t+\varepsilon;s,\mu_s)}(\varphi) - E_{\pi^u(t;s,\mu_s)}(\varphi)\right| = \left| E_{\pi^u(t;s,\mu_s)}(R^u_{t,t+\varepsilon}\varphi - \varphi)\right| \underset{\varepsilon\to o}{\to} 0$$

and the result is proved. ∎

III.1.2. The cost functional

Let g be a continuous function from $[o,T] \times C^n_{o,T} \times U$ to R with $C^n_{o,T} = C^o([o,T]; R^n)$ satisfying :

$$(3.22)\quad \begin{cases} (t,x) \to g(t,x,u) \text{ is } \mathcal{F}_t \text{ - progressively measurable } \forall u \in U \\ g \text{ : uniformly bounded on } [o,T] \times C^n_{o,T} \times U. \end{cases}$$

Let also G be a continuous function from R^n to R with :

$$(3.23)\qquad G \text{ : uniformly bounded on } R^n.$$

The cost functional is thus defined by :

$$(3.24)\qquad J_{s,\mu_s}(u) = E^{u}_{\nu^{u}_{s,\mu_s}} \left[\int_s^T g(t,X,u_t(Y))dt + G(X_T)\right]$$

for $u = (u_1,\ldots,u_N) \in \mathcal{U}$ and $\mu_s \in \mathcal{M}^b_{o,s}$.

Remark that, contrarily to what we saw in section II, a sophisticated definition of the cost function is not needed because of the uniqueness of the measure ν^{u}_{s,μ_s}.

III.1.3. Statement of the problem

$$(3.25)\quad \begin{cases} \text{given } s \in [o,T] \text{ and } \mu_s \in \mathcal{M}^b_{o,s}, \text{ evaluate } \operatorname{Inf}_{u \in \mathcal{U}} J_{s,\mu_s}(u) \text{ and} \\ \\ \text{characterize the optimal } u^* = (u_1^*,\ldots,u_N^*) \text{ if it exists.} \end{cases}$$

III.2. The dynamic programming method in the space of bounded measures

III.2.1. Reformulation into a purely final cost.

As in II.2.1, we introduce the auxiliary variable ζ by :

$$(3.26)\qquad \begin{cases} d\zeta_t = g(t,X,u_t(Y))dt & \forall t \in [s,T] \\ \zeta_s \text{ given} \end{cases}$$

Let us now introduce some notations :

$$(3.27)\qquad \begin{cases} z = (\xi',\zeta)' = (x',y',\zeta)' \in R^{n+q+1} \\ \tilde{G}(z) = \zeta + G(x) \\ F_o(t,z,u) = (\eta_o(t,\xi,u)',\ g(t,x,u))' \\ F_1(t,z) = \begin{pmatrix} \eta_1(t,\xi) & 0 \\ 0 & 0 \end{pmatrix}, \quad A(t,z) = F_1(t,z)\, F_1'(t,z) \end{cases}$$

clearly, the $(n+q+1)$ - dimensional process $Z = (\xi,\zeta)'$ given by (3.8) and (3.26) must satisfy (in the sense of the problem of martingales defined hereafter) :

$$(3.28)\qquad \begin{cases} dZ_t = F_o(t,Z,u_t(Y))dt + F_1(t,Z_t)d\tilde{v}_t & \forall t \in [s,T] \\ Z_\sigma = (\xi'_{o,s}(\sigma),\ \zeta_{o,s}(\sigma))' & \forall \sigma \in [o,s] \end{cases}$$

with $\tilde{v}_t = (v_t',o)'$, v_t being a suitably defined $(n+q)$ - Wiener process.

We shall also denote $C^{n+q+1}_{o,t}$ the space $C^o([o,t];\ R^{n+q+1})$, $\mathcal{F}_t(n+q+1)$ its Borel σ-field, $\tilde{\Omega} = C^{n+q+1}_{o,T}$, $\mathcal{F}^{n+q+1} = \mathcal{F}_T(n+q+1)$, and $\mathcal{M}^b_{o,t}(n+q+1)$ the space of bounded Radon measures on $C^{n+q+1}_{o,t}$ with the weak * - topology. If no confusion is possible, we shall write $C_{o,t}$ (resp $\mathcal{M}^b_{o,t}$) in place of $C^{n+q+1}_{o,t}$ (resp. $\mathcal{M}^b_{o,t}(n+q+1)$).

Remark that, by (3.6), (3.7) and (3.22), (3.23), we have :

$$(3.29)\quad \begin{cases} \sup_{(t,u)\in[o,T]\times u} \|F_o(t,Z,u)\| \leq \tilde{A}_1(1 + \sup_{s\in[o,t]} \|Z(s)\|), \quad \forall Z \in C^{n+q+1}_{o,T} \\ \tilde{G} : \text{uniformly bounded on } R^{n+q+1} \\ F_1 : \text{uniformly bounded, uniformly Lipschitz} \end{cases}$$

Definition 3.3 : Given $P_s \in \mathcal{M}^b_{o,s}(n+q+1)$, $u \in \mathcal{U}$ and $s \in [o,T]$, we say that the bounded measure P^u_{s,P_s} solves the problem of martingale relatively to (F_o,F_1,u,s,P_s) iff :

(i) $P^u_{s,P_s}(Z(\sigma) \in B,\ o \leq \sigma \leq s) = P_s(B) \quad \forall B \in \mathcal{F}^s_o(n+q+1)$

(ii) $\forall t \geq s, \quad \forall \varphi \in C^2_b(\tilde{\Omega})$:

$$\varphi(Z(t)) - \varphi(Z(s)) - \int_s^t [F'_o(\sigma,Z,u_\sigma(Y))\frac{\partial \varphi}{\partial z}(Z(\sigma)) + \frac{1}{2}\operatorname{tr}(A(\sigma,Z(\sigma))\frac{\partial^2 \varphi}{\partial z^2}(Z(\sigma)))]d\sigma$$

is a $(\mathcal{F}_t(n+q+1), P^u_{s,P_s})$ - martingale. ∎

Proposition 3.1 : Under the preceding assumptions, $\forall u \in \mathcal{U}$, $\forall s \in [o,T]$, $\forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$, there exists a unique solution P^u_{s,P_s} to the problem of martingale relatively to (F_o,F_1,u,s,P_s), and :

$$(3.30)\qquad J_{s,P_s}(u) = E_{P^u_{s,P_s}}(\tilde{G}(Z_T)) = \langle \tilde{G}, P^u_{s,P_s}(T) \rangle$$

where $P^u_{s,P_s}(T) = pr_T P^u_{s,P_s}$, namely : $\int_{\tilde{\Omega}} \varphi(Z_T)dP^u_{s,P_s}(Z) = \int_{R^{n+q+1}} \varphi(z)dP^u_{s,P_s}(T)(z)$

$\forall \varphi \in C^o_b(R^{n+q+1})$, and where $\langle \varphi,P \rangle = E_p(\varphi)$ denotes the duality product between $C^o_b(R^{n+q+1})$ and $\mathcal{M}^b(R^{n+q+1})$.

Proof : Let us denote μ_s the projection of P_s on $C^{n+q}_{o,s}$, namely $P_s = \int P_s(\cdot|\xi)d\mu_s(\xi)$. Clearly $\mu_s \in \mathcal{M}^b_{o,s}$. Let ν^u_{s,μ_s} be the solution of the problem of martingale relatively to $(\eta_o,\eta_1,u,s,\mu_s)$, and define P^u_{s,P_s} by :

$$(3.31)\quad \int_{\tilde{\Omega}} \varphi(Z)dP^u_{s,P_s}(Z) = \int_\Omega E_{P_s}(\varphi(\xi,\zeta + \int_s^\cdot g(\sigma,X,u_\sigma(Y))d\sigma))d\nu^u_{s,\mu_s}(\xi) \quad \forall \varphi \in C^o_b(\tilde{\Omega}).$$

with the notation $\int_s^\cdot g\,d\sigma$ for the process $t \to \int_s^{t\vee s} g d\sigma$.

Clearly, $P^u_{s,P_s} \in \mathcal{M}^b_{0,T}(n+q+1)$, $P^u_{s,P_s}(Z(\sigma) \in B,\ o \leq \sigma \leq s) = P_s(B) \quad \forall B \in \mathcal{F}^s_0(n+q+1)$.
Furthermore, since ν^u_{s,μ_s} solves the problem of martingale relatively to $(\eta_0,\eta_1,u,s,\mu_s)$, as in Remark 3.1, there exists a $\mathcal{F}$ - adapted $(n+q)$ dimensional Wiener process v^u_s after time s, such that :

$$d\xi_t = \eta_0(t,\xi,u_t(Y))dt + \eta_1(t,\xi_t)dv^u_s(t).$$

Thus, since $\zeta_t = \zeta_{0,s} + \int_s^t g(\sigma,X,u_\sigma(Y))d\sigma$, it is clear, by Itô's formula that :

$$\varphi(\xi(t),\zeta(t)) - \varphi(\xi(s),\zeta(s)) - \int_s^t [\eta_0'(\sigma,\xi,u_\sigma(Y))\ \frac{\partial\varphi}{\partial\xi}\ (\xi(\sigma),\zeta(\sigma)) +$$

$$+ g(\sigma,X,u_\sigma(Y))\ \frac{\partial\varphi}{\partial\zeta}\ (\xi(\sigma),\zeta(\sigma)) + \frac{1}{2}\ tr(a(\sigma,\xi(\sigma))\ \frac{\partial^2\varphi}{\partial\xi^2}\ (\xi(\sigma),\zeta(\sigma)))]d\sigma$$

$$= \int_s^t (\eta_1'(\sigma,\xi(\sigma))\ \frac{\partial\varphi}{\partial\xi}\ (\xi(\sigma),\zeta(\sigma)))'dv^u_s(\sigma) \quad \forall\varphi \in C^2_b(\tilde{\Omega})$$

and thus :

$$(3.32)\ \varphi(Z(t) - \varphi(Z(s)) - \int_s^t [F_0'(\sigma,Z,u_\sigma(Y))\ \frac{\partial\varphi}{\partial z}\ (Z(\sigma)) + \frac{1}{2}\ tr(A(\sigma,Z(\sigma))\ \frac{\partial^2\varphi}{\partial z^2}\ (Z(\sigma)))]d\sigma$$

is clearly a P^u_{s,P_s} - martingale, $\forall\varphi \in C^2_b(\tilde{\Omega})$.
The uniqueness of P^u_{s,P_s} is deduced from the one of ν^u_{s,μ_s} since by (ii) definition 3.3 with $\varphi \in C^2_b(\Omega)$ independent of ζ, we obtain (ii) of definition 3.2 for which ν^u_{s,μ_s} is uniquely defined with μ_s as precedently : $P_s = \int P_s(.|\xi)d\mu_s(\xi)$.
Consequently P^u_{s,P_s} is the product of $P_s(.|\xi)$ by the image ν^u_{s,μ_s} by the application :

$$\xi \to \xi,\zeta + \int_s^{\circ} g(\sigma,X,u_\sigma(Y))d\sigma, \quad \text{and is thus unique.}$$

Finally, since g and G are bounded, we obtain from (3.31) that :

$$J_{s,P_s}(u) = E_{\nu^u_{s,\mu_s}}(E_{P_s}(\int_s^T g(t,X,u_t(Y))dt + G(X(T))|\xi_{o,s})) = E_{P^u_{s,P_s}}(\tilde{G}(Z(T)))$$

$$= <\tilde{G},P^u_{s,P_s}(T)>. \quad \blacksquare$$

Let us now introduce the family of bounded measures :

$\{\Pi^u(t;s,P_s) | t \in [s,T],\ P_s \in \mathcal{M}^b_{o,s}(n+q+1)\}$ defined by

$$(3.33)\quad \begin{cases} \Pi^u(t;s,P_s) \in \mathcal{M}^b_{o,t} \\ \Pi^u(t;s,P_s)(B) = P^u_{s,P_s}(B) \quad \forall\, B \in \mathcal{F}^t_o(n+q+1) \\ \forall t \in [s,T],\ \forall\, P_s \in \mathcal{M}^b_{o,s}(n+q+1). \end{cases}$$

Proposition 3.2 : the family Π^u forms a weakly * continuous semi-group on $\mathcal{M}^o_{o,T}(n+q+1)$, and :

$$(3.34)\qquad P^u_{s,P_s} = P^u_{t,\Pi^u(t;s,P_s)} \quad \forall t \in [s,T],\ \forall u \in \mathcal{U},\ \forall\, P_s \in \mathcal{M}^b_{o,s}(n+q+1).$$

Furthermore, Π^u satisfies the following evolution equation :

$$(3.35)\quad \begin{cases} \dfrac{d}{dt}\int_{\Omega} \varphi(Z(t))d\Pi^u(t;s,P_s) = \int_{\Omega} L^u_t\varphi(Z)d\Pi^u(t,s,P_s) \quad \forall \varphi \in C^2_b(R^{n+q+1}) \\ \Pi^u(s;s,P_s) = P_s \end{cases}$$

with :

$$(3.36)\qquad L^u_t\varphi(Z) = F'_o(t,Z,u_t(Y))\,\frac{\partial \varphi}{\partial z}(Z(t)) + \frac{1}{2}\mathrm{tr}\left(A(t,Z(t))\,\frac{\partial^2 \varphi}{\partial z^2}(Z(t))\right)$$

Proof : by (3.31) and (3.33), one has :

$$(3.37)\qquad E_{\Pi^u(t;s,P_s)}(\varphi) = E_{\pi^u(t;s,\mu_s)}\left(E_{P_s}(\varphi|\xi)\right)$$
$$= \int_{C^{n+q}_{o,t}} \int_{C^1_{o,t}} \varphi\left(\xi,\zeta + \int_s^{\cdot} g(\sigma,X,u_\sigma(Y))d\sigma\right)dP_s(\zeta|\xi)\,d\pi^u(t;s,\mu_s)(\xi)$$

and, using the same argument as in lemma 3.2, and since :

$$\zeta_{o,s} + \int_s^t g\,d\sigma = \zeta_{o,s} + \int_s^\sigma g(\tau,X,u_\tau(Y))d\tau + \int_\sigma^t g(\tau,X,u_\tau(Y))d\tau$$

$$= \zeta_{o,\sigma} + \int_\sigma^t g\,d\tau \quad \text{with} \quad \zeta_{o,\sigma} = \zeta_{o,s} + \int_s^\sigma g\,d\tau \ , \quad \text{we obtain :}$$

$$E_{\Pi^u(t;s,P_s)}(\varphi) = \int_{C_{o,t}^{n+q}} E_{\Pi^u(\sigma;s,P_s)}(\varphi|\xi)\,d\pi^u(t;\sigma,\pi^u(\sigma;s,\mu_s))(\xi)$$

$$= E_{\Pi^u(t;\sigma,\Pi^u(\sigma;s,P_s))}(\varphi) \ , \quad \forall \varphi \in C_b^o(\tilde{\Omega})$$

since $\Pi^u(\sigma;s,P_s)$'s projection on $C_{o,t}^{n+q}$ is $\pi^u(\sigma;s,\mu_s)$ (see proposition 3.1). Also, (3.37) for $t = s$ yield $\Pi^u(s;s,P_s) = P_s$, and the simi-group property of Π^u is proved.

(3.34) follows from the fact that $\Pi^u(T;s,P_s) = P^u_{s,P_s}$ (easy consequence of (3.37) with $t = T$).

On the other hand, it suffices to prove the weak * continuity at $t = s$:

$$E_{\Pi^u(s+\varepsilon;s,P_s)}(\varphi) - E_{P_s}(\varphi) = E_{\pi^u(s+\varepsilon;s,\mu_s)}(E_{P_s}(\varphi|\xi)) - E_{\mu_s}(E_{P_s}(\varphi|\xi))$$

$$= E_{\mu_s}((R^u_{s,s+\varepsilon}(\xi) - 1)\, E_{P_s}(\varphi|\xi)) \underset{\varepsilon \to o}{\longrightarrow} 0 \qquad \forall \varphi \in C_b^o(\tilde{\Omega})$$

Finally, (3.35) is obtained by integration of (3.3?) with respect to $\Pi^u(t;s,P_s)$, and the Proposition is proved. ■

Remark 3.3 : (3.35) plays the role of a state equation controlled by u. ■

III.2.2. The Value function and the optimality principle

Now, problem (3.25) can be reformulated as follows :

(3.38)
given $s \in [o,T]$ and $P_s \in \mathcal{M}^b_{o,s}(n+q+1)$, evaluate :

$$\operatorname*{Inf}_{u \in \mathcal{U}} J_{s,P_s}(u) = \operatorname*{Inf}_{u \in \mathcal{U}} \langle \tilde{G}, P^u_{s,P_s}(T) \rangle$$

and characterize the optimal u* if it exists.

For this purpose, let us introduce the Value function, representing the optimal cost-to-go from time s with initial measure P_s :

$$(3.39)\qquad V(s,P_s) = \inf_{u \in \mathcal{U}} J_{s,P_s}(u) \quad \forall s \in [o,T], \quad \forall P_s \in \mathcal{M}^b_{o,s}(n+q+1).$$

Theorem 3.1 : The following transition properties hold true

$$\forall s \in [o,T], \quad \forall t \in [s,T], \quad \forall P_s \in \mathcal{M}^b_{o,s}(n+q+1) :$$

$$(3.40)\qquad J_{s,P_s}(u) = J_{t,\Pi^u(t;s,P_s)}(u) \qquad \forall u \in \mathcal{U}$$

$$(3.41)\qquad V(s,P_s) = \inf_{u \in \mathcal{U}} V(t,\Pi^u(t;s,P_s))$$

(3.41) is often called the optimality principle of dynamic programming.

Proof : $J_{s,P_s}(u) = \langle \tilde{G}, P^u_{s,P_s}(T) \rangle$ and, by (3.34) :

$$J_{s,P_s}(u) = \langle \tilde{G}, P^u_{t,\Pi^u(t;s,P_s)}(T) \rangle = J_{t,\Pi^u(t;s,P_s)}(u).$$

Finally, (3.41) follows immediately by taking the infimum with respect to u in both sides of (3.40). ∎

Remark 3.4 : As in remark 2.7, one can define a "dual" equation to the backward one (3.41), namely the onward equation of the nonlinear semi-group of operators.

$$(3.42)\qquad \begin{cases} \mathfrak{T}_t(V(s,P_s)) = \inf_{u \in \mathcal{U}} V(s+t,\Pi^u(s+t;s,P_s)) \\ \mathfrak{T}_o = I \end{cases}$$

One can easily check that $\mathfrak{T}_{t_1} \circ \mathfrak{T}_{t_2} = \mathfrak{T}_{t_2} \circ \mathfrak{T}_{t_1} = \mathfrak{T}_{t_1+t_2} \qquad \forall t_1, t_2$. ∎

As in remark 2.8, (3.41) defines the optimal N-tuple u*, if it exists, as a function of P_s. Such a u* need not belong to $\mathcal{U}$, and we must extend the admissibility of strategies in order to take into account this closed-loop dependence.

III.2.3. The closed-loop formulation

Let us introduce, as in definition 2.2 of II.2.3, the filtration $\{\mathcal{B}^t_{\mathcal{M}} \mid t \in [o,T]\}$ where $\mathcal{B}^t_{\mathcal{M}}$ is the Borel σ-field of $\mathcal{M}^b_{o,t}(n+q+1)$, and the Borel σ-field $\mathcal{B}_{U_k}$ of U_k, $k=1,\ldots,N$.

Definition 3.4 : an admissible closed-loop strategy u_k for decision maker $k(k=1,\dots,N)$ is a $(\mathcal{B}_{[o,t]} \times \mathcal{Y}_k^t \times \mathcal{B}_{\mathcal{M}}^t)$ - adapted process with values in U_k. The set of admissible closed-loop strategies for station k is noted $\tilde{u}_k$, and $\tilde{u} = \tilde{u}_1 \times \dots \times \tilde{u}_N$. ■

An admissible N-tuple $\tilde{u} \in \tilde{u}$ is thus allowed to depend in a non-anticipative way on the family of measures $\{\Pi^{\tilde{u}}(t;,s,P_s) \mid o \leq s \leq t \leq T\}$. It results that the definition of the problem of martingales must be extended. We shall proceed as follows :

Let us introduce the family of subdivisions of $[o,T]$:

$$(3.43) \qquad 0 = t_o^L < t_1^L \dots < t_L^L = T \quad \forall L \in N, \text{ with :}$$

$$\lim_{L \to \infty} \max_{o \leq i \leq L-1} |t_{i+1}^L - t_i^L| = 0.$$

Starting from $P_o \in \mathcal{M}^b(R^{n+q+1})$, we build $\Pi_{L,\beta}^{\tilde{u}}$ as follows :
Suppose that $\Pi_{L,\beta}^{\tilde{u}}(t_j;o,P_o)$ is given. Then on $[t_j,t_{j+1}[$, $\Pi_{L,\beta}^{\tilde{u}}(t_{j+1};o,P_o) = \Pi_{L,\beta}^{\tilde{u}}(t_{j+1};t_j,\Pi_{L,\beta}^{\tilde{u}}(t_j;o,P_o))$ is the unique solution of the problem of martingale relatively to $(F_o,F_1,\tilde{u}(\cdot,\Pi_{j,\beta}),t_j,\Pi_{L,\beta}^{\tilde{u}}(t_j;o,P_o))$, restricted to the interval $[t_j,t_{j+1}[$, where $\Pi_{j,\beta}$ is an arbitrary bounded measure on C_{o,t_j}^{n+q+1} belonging to a given weak * neighborhood of $\Pi_{L,\beta}^{\tilde{u}}(t_j;o,P_o)$, noted $\beta(\Pi_{L,\beta}^{\tilde{u}}(t_j;o,P_o))$

Thus, on each $[t_j,t_{j+1}[$, we have :

$$(3.44) \quad dZ_t = F_o(t,Z,\tilde{u}_t(Y,\Pi_{j,\beta}))dt + F_1(t,Z_t)dv_{j,\beta}(t), \quad \text{with} \quad v_{j,\beta}(t) = \begin{pmatrix} \tilde{v}_{j,\beta}(t) \\ o \end{pmatrix}.$$

Consequently, on $[o,T]$, for each L and for each denumerable basis of neighborhoods β of the measures $\Pi_{L,\beta}^{\tilde{u}}$ (with respect to the weak * topology) there exists a unique semi-group $\Pi_{L,\beta}^{\tilde{u}}$.

Definition 3.5 : We say that $\Pi^{\tilde{u}}(T;o,P_o) = P_{o,P_o}^{\tilde{u}}$ solves the problem of martingales relatively to $(F_o,F_1,\tilde{u},o,P_o)$ with $\tilde{u} \in \tilde{u}$, if there exists a family of subdivisions $\{t_j^L\}$ of $[o,T]$ and a denumerable basis of neighborhoods β in $\mathcal{M}_{o,T}^b(n+q+1)$ such that $\Pi^{\tilde{u}}(T;o,P_o)$ is the weak * limit of the sequence : $\{\Pi_{L,\beta}^{\tilde{u}}(T;o,P_o)\}$, defined as above.

The set of such solutions $\Pi^{\tilde{u}}(t;s,P_s)$ is denoted $\wp^{\tilde{u}}(t;s,P_s)$, $\forall \tilde{u} \in \tilde{\mathcal{U}}$, $\forall t \in [s,T]$, $\forall s \in [o,T]$, $\forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$. ■

Remark 3.5 : the denumerable basis β of weak * neighborhoods of each $\Pi^{\tilde{u}}_{L,\beta}(t_j;o,P_o)$, in which is chosen $\Pi_{j,\beta}$, is introduced in order to obtain a property of concatenation of $\Pi^{\tilde{u}}(t;s,\Pi^{\tilde{u}}(s;o,P_o))$ with $\Pi^{\tilde{u}}(s;o,P_o)$, since the approximations $\Pi^{\tilde{u}}_{L,\beta}(s;o,P_o)$ belong to a weak * neighborhood of $\Pi^{\tilde{u}}(s,o,P_o)$ but are different from the starting point of $\Pi^{\tilde{u}}(t;s,\Pi^{\tilde{u}}(s;o,P_o))$ (see Appendix A.III). This definition is in the spirit of Krassovski's concept of solution for a deterministic differential equation with measurable right-hand side [20]. ■

Proposition 3.3 : $\wp^{\tilde{u}}(t;s,P_s) \neq \emptyset \quad \forall \tilde{u} \in \tilde{\mathcal{U}}$, $\forall t \geq s$, $\forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$. Furthermore, for each $P^{\tilde{u}}_{s,P_s} \in \wp^{\tilde{u}}(T;s,P_s)$, there exists a $\mathcal{B}_{[o,t]} \times \mathcal{F}_t \times \mathcal{Y}_t \times \mathcal{B}^t_{\mathcal{M}}$ measurable selection $\tilde{F}_o$ of the multifunction :

$$(3.45)\quad \Phi(t,Z,Y,\tilde{u},P^{\tilde{u}}_{s,P_s}) = \overline{co} \bigcap_{\varepsilon > o} \bigcap_{\Pi \in \beta(\Pi^{\tilde{u}}(t;s,P_s))} F_o(t,Z,\tilde{u}_t(B_\varepsilon(Y),\Pi)),$$

where $B_\varepsilon(Y)$ is the ε-ball of $C^q_{o,t}$ centered at Y, namely $\tilde{F}_o$ satisfying :

$$(3.46)\quad \tilde{F}_o(t,Z,Y,\Pi^{\tilde{u}}(t;s,P_s)) \in \Phi(t,Z,Y,\tilde{u},P^{\tilde{u}}_{s,P_s}), \quad P^{\tilde{u}}_{s,P_s} \text{ a.s.},$$

such that $P^{\tilde{u}}_{s,P_s}$ solves the problem of martingale relatively to $(\tilde{F}_o,F_1,\tilde{u},s,P_s)$.

Proof : see Appendix A.III. ■

Indeed, $\wp^{\tilde{u}}(T;s,P_s)$ contains generally more than 1 element, and the cost functional (3.30) must be replaced, in the closed-loop case, by the minimum guaranteed cost functional :

$$(3.47)\quad \tilde{J}_{s,P_s}(\tilde{u}) = \sup_{P^{\tilde{u}}_{s,P_s} \in \wp^{\tilde{u}}(T;s,P_s)} \langle \tilde{G}, P^{\tilde{u}}_{s,P_s}(T) \rangle, \quad \forall \tilde{u} \in \tilde{\mathcal{U}},$$

and the Value function (3.39) becomes :

$$(3.48)\quad \tilde{V}(s,P_s) = \inf_{\tilde{u} \in \tilde{\mathcal{U}}} \tilde{J}_{s,P_s}(\tilde{u}), \quad \forall s \in [o,T], \quad \forall P_s \in \mathcal{M}^b_{o,s}(n+q+1).$$

Theorem 3.2 : $\forall s \in [o,T]$, $\forall t \in [s,T]$, $\forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$, the following transition properties hold true.

$$(3.49) \qquad \wp^u(T;s,P_s) = \bigcup_{\Pi^u_t \in \wp^u(t;s,P_s)} \wp^u(T;t,\Pi^u_t), \quad \forall u \in \tilde{u}$$

$$(3.50) \qquad \tilde{J}_{s,P_s}(u) = \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} J_{t,\Pi^u_t}(u), \quad \forall u \in \tilde{u}$$

$$(3.51) \qquad \tilde{V}(s,P_s) = \inf_{u \in \tilde{u}} \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} V(t,\Pi^u_t)$$

Proof : (3.49) is proved in Appendix A.III.
By definition of $\tilde{J}$, we have :

$$\tilde{J}_{s,P_s}(u) = \sup_{P^u_{s,P_s} \in \wp^u(T;s,P_s)} \langle \tilde{G}, P^u_{s,P_s}(T) \rangle$$

$$= \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} \sup_{P^u_{t,\Pi^u_t} \in \wp^u(T;t,\Pi^u_t)} \langle \tilde{G}, P^u_{t,\Pi^u_t}(T) \rangle$$

$$= \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} J_{t,\Pi^u_t}(u) \quad , \text{ which proves (3.50).}$$

To prove (3.51), we begin by taking the infimum of the left-hand side of (3.50) :

$$\tilde{V}(s,P_s) \leq \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} \tilde{J}_{t,\Pi^u_t}(u) \qquad \forall u \in \tilde{u}$$

and thus

$$\tilde{V}(s,P_s) \leq \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} \tilde{V}(t,\Pi^u_t) \qquad \forall u \in \tilde{u}$$

or :

$$(3.52) \qquad \tilde{V}(s,P_s) \leq \inf_{u \in \tilde{U}} \sup_{\Pi^u_t \in \wp^u(t;s,P_s)} \tilde{V}(t,\Pi^u_t).$$

Conversely, since $\tilde{G}$ is uniformly bounded, $\tilde{V}(s,P_s)$ is finite and there exists a sequence $\{u_m\}$ in $\tilde{\mathcal{U}}$ such that :

$$\lim_{m \to \infty} \tilde{J}_{s,P_s}(u_m) = \tilde{V}(s,P_s). \tag{3.53}$$

but $\forall m \geq 0$ we have :

$$\tilde{J}_{s,P_s}(u_m) = \sup_{\Pi_t^{u_m} \in \wp^{u_m}(t;s,P_s)} \tilde{J}_{t,\Pi_t^{u_m}}(u_m)$$

$$\geq \sup_{\Pi_t^{u_m} \in \wp^{u_m}(t;s,P_s)} \tilde{V}(t,\Pi_t^{u_m})$$

or :

$$\tilde{J}_{s,P_s}(u_m) \geq \inf_{u \in \tilde{\mathcal{U}}} \sup_{\Pi_t^u \in \wp^u(t;s,P_s)} \tilde{V}(t,\Pi_t^u)$$

and, as $m \to \infty$, using (3.53) :

$$\tilde{V}(s,P_s) \geq \inf_{u \in \tilde{\mathcal{U}}} \sup_{\Pi_t^u \in \wp^u(t;s,P_s)} \tilde{V}(t,\Pi_t^u) \tag{3.54}$$

and we have proved (3.51). ∎

Theorem 3.3 : The closed-loop and open-loop value functions coincide, namely : $\tilde{V}(s,P_s) = V(s,P_s)$ $\forall s \in [o,T]$, $\forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$.

Proof : Clearly $\tilde{V}(s,P_s) \leq V(s,P_s)$ $\forall P_s$, $\forall s$, since $\mathcal{U} \subset \tilde{\mathcal{U}}$.
On the other hand, by definition of the infimum, $\forall \varepsilon > o$ one can find $u_\varepsilon \in \tilde{\mathcal{U}}$ such that :

$$\tilde{J}_{s,P_s}(u_\varepsilon) \leq V(s,P_s) + \varepsilon/2 . \tag{3.55}$$

Furthermore, by lemma 3.3 below, one can find $\overline{u_\varepsilon} \in \mathcal{U}$ such that :

$$| J_{s,P_s}(\overline{u_\varepsilon}) - \tilde{J}_{s,P_s}(u_\varepsilon) | \leq \varepsilon/2 \tag{3.56}$$

Thus $J_{s,P_s}(\overline{u_\varepsilon}) \leq V(s,P_s) + \varepsilon$, and a fortiori :

(3.57) $$V(s,P_s) \leqslant \tilde{V}(s,P_s) + \varepsilon \ .$$

Finally, ε being arbitrary, the result is proved. ∎

Lemma 3.3 : $\forall \varepsilon > 0$, $\forall u \in \tilde{\mathcal{U}}$, one can find $u_\varepsilon \in \mathcal{U}$ such that :

$$\left| J_{s,P_s}(u_\varepsilon) - \tilde{J}_{s,P_s}(u) \right| \leqslant \varepsilon$$

Proof : Let $\varepsilon > 0$ be given. By definition of $\tilde{J}$, one can find $P^u_{s,P_s} \in \mathcal{P}^u(T;s,P_s)$ such that :

(3.58) $$\left| \tilde{J}_{s,P_s}(u) - < \tilde{G}, P^u_{s,P_s}(T) > \right| \leqslant \varepsilon/2$$

Also, by construction of $\mathcal{P}^u(T;s,P_s)$, there exists a sequence $\{\Pi^u_{L,\beta}(T;s,P_s)\}$ weakly * converging to P^u_{s,P_s}.

Now, since $\tilde{G}$ is continuous and bounded, we have :

$$\lim_{L,\beta} < \tilde{G}, \Pi^u_{L,\beta}(T;s,P_s)_{|T} > = < \tilde{G}, P^u_{s,P_s}(T) > \ .$$

Thus, choosing L and β such that :

$$\left| < \tilde{G}, \Pi^u_{L,\beta}(T;s,P_s)_{|T} > - < \tilde{G}, P^u_{s,P_s}(T) > \right| \leqslant \varepsilon/2 \ ,$$

and calling :

(3.59) $$u_{L,\beta}(t,Y) = u(t,Y,\Pi^u_{L,\beta}(t;s,P_s)) \quad \forall t \in [s,T], \quad \forall Y,$$

we have : $u_{L,\beta} \in \mathcal{U}$ and $J_{s,P_s}(u_{L,\beta}) = < \tilde{G}, \Pi^u_{L,\beta}(T;s,P_s)_{|T} >$, since $\Pi^u_{L,\beta}$ is uniquely defined for each L,β. It results that :

$$\left| \tilde{J}_{s,P_s}(u) - J_{s,P_s}(u_{L,\beta}) \right| = \left| \tilde{J}_{s,P_s}(u) - < \tilde{G}, \Pi^u_{L,\beta}(T;s,P_s)_{|T} > \right|$$

$$\leqslant \left| \tilde{J}_{s,P_s}(u) - < \tilde{G}, P^u_{s,P_s}(T) > \right| + \left| < \tilde{G}, P^u_{s,P_s}(T) > - < \tilde{G}, \Pi^u_{L,\beta}(T;s,P_s)_{|T} > \right|$$

$\leqslant \varepsilon$ and the result is proved. ∎

III.3. Some regularity properties of the value function

Proposition 3.4 : The value function satisfies :

$$(3.60)\qquad V(s,P_s) = \operatorname*{Inf}_{u \in \tilde{\mathcal{U}}} \operatorname*{Inf}_{\Pi \in \mathcal{P}^u(T;s,P_s)} \langle \tilde{G}, \Pi(T) \rangle$$

$$= \operatorname*{Inf}_{u \in \tilde{\mathcal{U}}} \operatorname*{Sup}_{\Pi \in \mathcal{P}^u(T;s,P_s)} \langle \tilde{G}, \Pi(T) \rangle .$$

Furthermore, if $u^* \in \tilde{\mathcal{U}}$ is such that $\tilde{J}_{s,P_s}(u^*) = V(s,P_s)$, we have :

$$(3.61)\qquad V(s,P_s) = \langle G, \Pi^*(T) \rangle \quad \forall\, \Pi^* \in \mathcal{P}^{u^*}(T;s,P_s).$$

Proof : same as proposition 2.6. ∎

Proposition 3.5 : There exists a $\mathcal{B}_{[o,T]} \times \mathcal{B}_{R^{n+q+1}} \times \mathcal{B}_{\mathcal{M}}$ measurable function w, uniformly bounded on $[o,T] \times R^{n+q+1} \times \mathcal{M}^b_{o,T}(n+q+1)$, and satisfying :

$$(3.62)\quad V(t,P_t) = \int_{R^{n+q+1}} w(t,z;P_t)\,d\Pi_t(z) \quad \forall\, P_t \in \mathcal{M}^b_{o,t}(n+q+1), \quad \forall t \in [o,T],$$

with $\Pi_t = pr_t P_t$, namely : $\int_{R^{n+q+1}} \varphi(z)\,d\Pi_t(z) = \int_{C^{n+q+1}_{o,t}} \varphi(Z(t))\,dP_t(Z)$

$\forall \varphi \in C^o_b(C^{n+q+1}_{o,t})$.

Proof : Let $\{u_m\}_{m \geq o}$ be a minimizing sequence in $\tilde{\mathcal{U}}$, namely :

$$(3.63)\qquad \lim_{m \to \infty} \tilde{J}_{t,P_t}(u_m) = V(t,P_t).$$

Such a sequence exists since V is finite. Let us denote :

$$(3.64)\quad w_m(t,z;P_t) = \int_{\tilde{\Omega}} \tilde{G}(Z(T))\,d\Pi^{u_m}(T;t,P_t)(Z \mid Z(t)=z) \quad \forall z \in R^{n+q+1},$$

with $\Pi^{u_m}(T;t,P_t)$ arbitrarily chosen in $\mathcal{P}^{u_m}(T;t,P_t)$.

Clearly, we have : $\int w_m(t,z;P_t)\,d\Pi_t(z) = \langle \tilde{G}, \Pi^{u_m}(T;t,P_t)_{|T} \rangle$, since

$\Pi_t = pr_t P_t = pr_t \Pi^{u_m}(T;t,P_t)$. Furthermore $\{w_m\}$ are uniformly bounded on $[o,T] \times R^{n+q+1} \times \mathcal{M}^b_{o,T}(n+q+1)$, and :

$$(3.65) \qquad \lim_{m \to \infty} E_{\Pi_t}(w_m) = V(t,P_t) ,$$

by (3.63) and proposition 3.4.
Applying Fatou's lemma to (3.65), we obtain :

$$(3.66) \qquad E_{\Pi_t}\left(\lim_{m \to \infty} w_m(t;.,P_t)\right) \leq V(t,P_t).$$

Let us denote :

$$(3.67) \qquad w(t,z;P_t) = \lim_{m \to \infty} w_m(t,z;P_t) \qquad \forall t,z,P_t .$$

Then w is $\mathcal{B}_{[o,T]} \times \mathcal{B}_{R^{n+q+1}} \times \mathcal{B}_{\mathcal{M}}$ measurable and bounded on $[o,T] \times R^{n+q+1} \times \mathcal{M}^b_{o,T}(n+q+1)$, and, by (3.66) :

$$E_{\Pi_t}(w(t,.;P_t)) \leq V(t,P_t).$$

On the other hand, since, by proposition 3.4 and (3.64) :

$$V(t,P_t) \leq E_{\Pi_t}(w_m(t,.;P_t)) \quad \forall m \geq o$$

we have that $E_{\Pi_t}(w) = V(t,P_t)$.
We have to check that w is independent of the sequence $\{u_m\}$:
for this purpose, if $\{u_{m'}\}$ is another minimizing sequence, let us denote
$w'(t,z;P_t) = \lim_{m' \to \infty} w_{m'}(t,z;P_t)$

But, noting $u_{m''} = \{u_m\} \cup \{u_{m'}\}$, $u_{m''}$ is also a minimizing sequence and, if
$w''(t,z;P_t) = \lim_{m'' \to \infty} w_{m''}(t,z;P_t)$.

But clearly, since $\{u_m\} \subset \{u_{m''}\}$ and $\{u_{m'}\} \subset \{u_{m''}\}$:

$$(3.68) \qquad \begin{cases} w''(t,z;P_t) \leq w(t,z;P_t) \\ \\ w''(t,z;P_t) \leq w'(t,z;P_t) \end{cases} \qquad \forall t,z,P_t$$

and $E_{\Pi_t}(w'') = E_{\Pi_t}(w') = E_{\Pi_t}(w) = V(t,P_t)$

which imply that $w'' = w' = w$ Π_t - almost surely. ∎

Proposition 3.6 :

(i) $\forall t \in [o,T]$, the application $P \to V(t,P)$ from $\mathcal{M}^b_{o,t}$ to R is concave, everywhere finite, and thus continuous on $\mathcal{M}^b_{o,t}$.

(ii) $P \to V(t,P)$ is subdifferentiable on $\mathcal{M}^b_{o,t}$ (in the sense of concave functions) and its subdifferential $\partial_P V(t,P)$ is a nonempty convex compact subset of $C^o_b(C^{n+q+1}_{o,t})$ endowed with the uniform topology.

Proof : same as proposition 2.8. ∎

In order to give a formula for the directional derivatives of V in the direction of $\frac{d}{dt}\Pi^u(t;s,P_s)$, we must introduce the following translation operator :

$\theta_{t'-t} : C^{n+q+1}_{o,t'} \to C^{n+q+1}_{o,t}$ for $t' \geq t$ is defined by :

$$(3.69) \qquad \theta_{t'-t}(Z)(\sigma) = Z(\sigma+(t'-t)) \quad \forall \sigma \in [o,t] \quad \forall Z \in C^{n+q+1}_{o,t'} .$$

$\theta_{t'-t}$ can also be extended by duality as an application $\theta^*_{t'-t}$ from $\mathcal{M}^b_{o,t'}$ to $\mathcal{M}^b_{o,t}$ by :

$$(3.70) \int_{C^{n+q+1}_{o,t}} \varphi(Z)\, d(\theta^*_{t'-t}P_{t'})(Z) = \int_{C^{n+q+1}_{o,t}} \varphi(\theta_{t'-t}(Z))\, dP_{t'}(Z) \quad \forall \varphi \in C^o_b(C^{n+q+1}_{o,t}).$$

We define for $s' \geq s$ and $P_{s'} \in \mathcal{M}^b_{o,s'}$, $\forall t \geq s$:

$$(3.71) \quad \mathcal{P}^u(t;s,P_{s'}) \stackrel{def}{=} \mathcal{P}^u(t;s,\theta^*_{s'-s}P_{s'}), \quad V(s,P_{s'}) \stackrel{def}{=} V(s,\theta^*_{s'-s}P_{s'}).$$

Proposition 3.7 : Assume that $\Pi(T;s,P_s) \in \mathcal{P}^u(T;s,P_s)$ and that there exists $\Lambda_o \in \partial_P V(t,\Pi(t;s,P_s))$ such that $\langle \Lambda_o, \frac{d}{dt}\Pi(t;s,P_s) \rangle$ is finite. Then the directional derivative of $V(t,\Pi(t;s,P_s))$ in the direction $\frac{d}{dt}\Pi(t;s,P_s)$ exists and is given by :

$$(3.72) \quad \lim_{\varepsilon \to o_+} \frac{1}{\varepsilon}\left(V(t,\Pi(t+\varepsilon;s,P_s)) - V(\tau,\Pi(t;s,P_s))\right) = \min_{\Lambda \in \partial_P V(t,\Pi(t;s,P_s))} \langle \Lambda, \frac{d}{dt}\Pi(t;s,P_s) \rangle$$

Furthermore, $< \Lambda_o, \frac{d\Pi_t}{dt} >$ is well defined if one of these two conditions hold :

(i) $\frac{d}{dt}\Pi(t;s,P_s) \in \mathcal{M}^b_{o,t}$

(ii) $\partial_p V(t,\Pi_t) \cap C^2_b(R^{n+q+1}) \neq \emptyset$

(Π_t denotes $\Pi(t;s,P_s)$).

Proof : follows the same lines as proposition 2.9, the only difference being that $< \varphi, \frac{d\Pi_t}{dt} > = E_{\Pi_t}(L^u_t \varphi)$ is finite for $\varphi \in C^2_b(R^{n+q+1})$ whereas in proposition 2.9 L^u_t was a first order operator and thus the same formula was well-defined for $\varphi \in C^1_b$. Here,

$$L^u_t\varphi = \frac{1}{2} tr(A(t,Z) \frac{\partial^2 \varphi}{\partial z^2}(Z(t))) + \overline{F^u_o}(t,Z,Y,\Pi_t)' \frac{\partial \varphi}{\partial z}(Z(t))$$

with $\overline{F^u_o}$ measurable selection of Φ (see proposition 3.3). ∎

Corollary 3.1 : Assume that w, defined by proposition 3.5, is twice differentiable with respect to z on R^{n+q+1} with bounded partial derivatives, and satisfies :

$$(3.73)\quad \begin{cases} < \frac{\partial w}{\partial P}(t,z;P_t), Q > = \int_{R^{n+q+1}} \{ \frac{\partial w}{\partial P} \}(t,z,\zeta) dQ(t)(\zeta) \quad \forall Q \in \mathcal{M}^b_{o,t} \\ \text{with } Q(t) = pr_t(Q), \text{ and with :} \\ \zeta \to \{ \frac{\partial w}{\partial P} \}(t,z,\zeta) \text{ is in } C^2_b(R^{n+q+1}) \quad \forall (t,z) \in [o,T] \times R^{n+q+1} \end{cases}$$

then (3.72) becomes :

$$(3.74)\quad \lim_{\varepsilon \to o_+} \frac{1}{\varepsilon}(V(t,\Pi_{t+\varepsilon}) - V(t,\Pi_t)) = \int_{C^{n+q+1}_{o,t}} (L^u_t \lambda)(t,Z,\Pi_t) d\Pi_t(Z)$$

with $\Pi_t = \Pi(t;s,P_s) \in \wp^u(t;s,P_s)$, $\Pi_t(t) \overset{def}{=} pr_t(\Pi_t)$,

$$(3.75)\quad (L^u_t \lambda)(t,Z,\Pi_t) = \frac{1}{2} tr(A(t,Z(t)) \frac{\partial^2 \lambda}{\partial z^2}(Z(t)) + \overline{F^u_o}(t,Z,Y,\Pi_t)' \frac{\partial \lambda}{\partial z}(Z(t),$$

$\overline{F^u_o}$ measurable selection of Φ, and :

$$(3.76)\qquad \lambda(t,z,\Pi_t) = w(t,z;\Pi_t) + \int_{R^{n+q+1}} \left\{ \frac{\partial w}{\partial P} \right\} (t,\zeta,z)\, d\Pi_t(t)(\zeta).$$

Consequently $\partial_I V(t,\Pi_t) = \{\lambda(t,.,\Pi_t)\}$ is reduced to a single gradient.

Proof : follows the same lines as in the corollary 2.2, taking once more into account the fact that, here L^u_t is a second - order operator. ∎

Remark 3.6 : The condition $\frac{\partial \Pi_t}{\partial t} \in \mathcal{M}^b_{0,t}$ is satisfied for example when Π_t is absolutely continuous with respect to the lebesgue measure of R^{n+q+1} with C^2_K density, and if F^u_o and A in (3.75) are C^2 functions with respect to z. This is easily seen, as in remark 2.11, by integrating by parts : $< \varphi, \frac{d\Pi_t}{dt} > = \int \varphi(z)(L^u_t)^*(p_t)(z)dz$, where p_t is the density of Π_t and where $(L^u_t)^*$ is the adjoint of L^u_t . ∎

III.4. The Hamilton - Jacobi - Bellman equations and the Signalling

We note $\mathcal{M}^{1,+}_{0,t}$ the space of probability measures on $C^{n+q+1}_{0,t}$.

Theorem 3.4 : Assume that V is differentiable with respect to t, with $\frac{\partial V}{\partial t}$ everywhere finite, and assume that :

$$(H)\quad \begin{cases} \forall t \in [o,T],\ \forall P_t \in \mathcal{M}^{1,+}_{0,t}\ ,\ \exists u \in \tilde{u}\ ,\ \exists \Pi^u \in \mathcal{P}^u(T;t,P_t)\ , \\ \exists \wedge_o \in \partial_p V(t,P_t) \text{ such that } < \wedge_o, \frac{d}{dt}\Pi^u_t > \text{ is finite :} \end{cases}$$

with $\frac{d}{dt}\Pi^u_t$ short notation for $\frac{d}{d\sigma}\Pi^u(\sigma;t,P_t)\big|_{\sigma = t}$.

Then V satisfies the Hamilton - Jacobi - Bellman equation :

$$(3.77)\quad \begin{cases} \frac{\partial V}{\partial t}(t,P_t) + \underset{u \in \tilde{u}}{\mathrm{Inf}}\ \underset{\Pi^u \in \mathcal{P}^u(T,t,P_t)}{\mathrm{Inf}}\ \underset{\wedge \in \partial_p V(t,P_t)}{\mathrm{Min}} < \wedge, \frac{d\Pi^u_t}{dt} > = 0 \\ V(T,P_T) = < G, P_T(T) > \quad \forall P_T \in \mathcal{M}^{1,+}_{0,T}\ ,\ P_T(\omega) = pr_T(P_T). \end{cases}$$

Conversely, if V is differentiable in t, with $\frac{\partial V}{\partial t}$ everywhere finite, concave and subdifferentiable in P, and satisfies (3.77), then :

$$(3.78)\qquad V(t,P_t) = \underset{u \in \tilde{u}}{\mathrm{Inf}}\ \tilde{J}_{t,P_t}(u) \qquad \forall t \in [o,T],\ \forall P_t \in \mathcal{M}^{1,+}_{0,t}\ .$$

Proof : (3.77) follows by combining (3.5.1) of theorem 3.2 with proposition 3.7, which is valid under (H).

Conversely, since (3.77) means that $\inf_{u \in \mathcal{U}} \inf_{\Pi^u \in \mathcal{P}^u(T;s,P_s)} \frac{d}{dt} V(t,\Pi^u_t) = 0$, (3.7 8) follows by integration with respect to t on $[s,T]$. ∎

Corollary 3.2 : Assume that w, defined in proposition 3.5, satisfies :

$$(3.79)\quad \begin{cases} \frac{\partial w}{\partial \tau}(.,.,P), \frac{\partial w}{\partial z_i}(.,.,P) \in L^\infty([o,T] \times R^{n+q+1}), \ \forall i=1,\dots,n+q+1, \\ \qquad\qquad \forall P \in \mathcal{M}^{1,+}_{o,T}, \\ \text{and } \frac{\partial w}{\partial P} \text{ satisfies (3.73).} \end{cases}$$

Then w satisfies the following Hamilton - Jacobi - Bellman equation :

$$(3.80)\quad \begin{cases} \inf_{u \in \mathcal{U}} \int_{C^{n+q+1}_{o,t}} [\frac{\partial w}{\partial t}(t,z;P_t) + (L^u_t \lambda)(t,Z;P_t)] dP_t(Z) = 0 \quad \forall t \in [o,T] \\ \qquad\qquad \forall P_t \in \mathcal{M}^{1,+}_{o,t} \\ w(T,z;P_T) = \tilde{G}(z) \quad \forall P_T \in \mathcal{M}^{1,+}_{o,T}, \quad \forall z \in R^{n+q+1} \end{cases}$$

with :

$$(3.81)\quad \lambda(t,Z;P_t) = w(t,z;P_t) + \int_{R^{n+q+1}} \{\frac{\partial w}{\partial P}\}(t,\zeta,z) dP_t(t)(\zeta),$$

$$P_t(t) = pr_t P_t \quad , \quad \text{and :}$$

$$(3.82)\quad (L^u_t \varphi)(Z) = \frac{1}{2} \operatorname{tr}(A(t,Z(t)) \frac{\partial^2 \varphi}{\partial z^2}(Z(t)) + F_o(t,Z,u_t(Y)) \frac{\partial \varphi}{\partial z}(Z(t))$$

$$\forall \varphi \in C^2_b(R^{n+q+1}).$$

Conversely, if w satisfies (3.79), (3.80) with (3.81), (3.82), then :

$$V(t,P_t) = \int_{R^{n+q+1}} w(t,z;P_t) dP_t(t)(z) = \inf_{u \in \mathcal{U}} J_{t,P_t}(u), \quad \forall t \in [o,T],$$

$$\forall P_t \in \mathcal{M}^{1,+}_{o,t}.$$

Proof : This is just a translation of theorem 3.4 in the language of corollary 3.1, the only thing to prove being the fact that $F_o(t,Z,u_t(Y))$ appears in L_t^u rather than any measurable selection of the corresponding $\Phi(t,Z,Y,u,\Pi)$. But this is a consequence of the fact that :

$$(3.83)\qquad \inf_{u\in\mathcal{U}} \inf_{F_o^u \in \Phi(t,Z,Y,u,P)} E_P(\lambda . F_o^u) = \inf_{u\in\mathcal{U}} \inf_{v\in\mathcal{U}} E_P(\lambda(t,Z).F_o(t,Z,v(Y)))$$

since $\Phi(t,Z,Y,u,P) = \overline{co} \bigcap_{\varepsilon>o} \bigcap_{\Pi\in\beta(P)} F_o(t,Z,u_t(B_\varepsilon(Y),\Pi))$

and $F_o^u \in \Phi(t,Z,Y,u,P)$ is equivalent to the existence of $\alpha_1,\dots,\alpha_r \geq o$, $\sum_{i=1}^r \alpha_i = 1$, and $v_1,\dots,v_r \in \tilde{\mathcal{U}}$ such that :

$$\sum_{i=1}^r \alpha_i F_o(t,Z,v_i(Y,P)) = F_o^u(t,Z,Y,u,P),$$

and, using the convexity of $F_o(t,Z,U)$ by (3.7), we have :

$\exists\, v \in \tilde{\mathcal{U}}$ such that $F_o(t,Z,v(Y,P)) = F_o^u(t,Z,Y,u,P)$, which proves (3.83).

Finally, the right-hand side of (3.83) being independent of u, the proof is complete. ∎

Corollary 3.3 : the statement of corollary 2.4 is true with $F_o(t,Z,u_k^*(t,Y,P_t))$ in place of F in (2.100), and with :

$$u_k^*(t,Y,P_t) = (u_1^*(t,Y^1,P_t),\dots,u_{k-1}^*(t,Y^{k-1},P_t),\ u_{k+1}^*(t,Y^{k+1},P_t),\dots$$
$$\dots,u_N^*(t,Y^N,P_t))$$

with $Z = (X,Y^1,\dots,Y^N,\zeta)$.

Furthermore, with λ given by (3.81), the adjoint is $\frac{\partial\lambda}{\partial z}$, and can be written in two parts as in (2.101) :

$$(3.84)\qquad \frac{\partial\lambda}{\partial z} = \frac{\partial w}{\partial z} + \frac{\partial}{\partial z}\int \left\{\frac{\partial w}{\partial P}\right\}(t,\zeta,z)\,dP_t(\zeta),$$

the second term being called the "signalling term". ∎

Remark 3.7 : In practice, to compute the optimal $u_1^*,\dots,u_N^*$ and V, one has to look at the Hamiltonians :

$$H_k = \int \frac{\partial \lambda}{\partial z}(t,z;P_t)F_o(t,Z,\hat{u_k^*}(t,Y,P_t),u_k)dP_t(Z|Y_\sigma^k,\nu_k(t) \leq \sigma \leq t)$$

(Corollary (3.3)), $\forall k=1,\dots,N$, and to minimize H_k with respect to u_k for every t, $\{Y_\sigma^k,\ \nu_k(t) \leq \sigma \leq t\}$, λ and P_t, all the stations, except k, using their optimal strategies : thus we are looking for a Nash equilibrium of the Hamiltonians. After having obtained the optimal N-tuple $u_1^*,\dots,u_N^*$ for every (t,Y,P_t,λ), we compute V and w by :

$$(3.85)\quad \begin{cases} E_{P_t}\left(\frac{\partial w}{\partial t} + L_t^{u^*}(\lambda)\right) = o, \quad \lambda(t,.,P_t) = w(t,.;P_t) + \int_{R^{n+q+1}} \{\frac{\partial w}{\partial P}\}(t,\zeta,.)dP_t(t)(\zeta) \\ w(T,z,P) = \tilde{G}(z) \quad \forall z\ ,\ \forall P. \end{cases}$$

together with $P_{o,t}^{u^*}$ obtained through :

$$(3.86)\quad \begin{cases} \frac{d}{dt}\int \varphi dP_t^* = E_{P_t^*}(L_t^{u^*}\varphi) \quad \forall \varphi \in C_b^2(R^{n+q+1}). \\ pr_o(P_t^*) = P_o \end{cases}$$

remark that in (3.85) and (3.86), u* is considered as a stochastic process whereas in the Nash equilibrium it is computed as a function depending on the observations' realizations.

On the other hand, if P_t^* and w are regular enough, one can solve numerically the system of PDE's.

$$(3.87)\quad \begin{cases} \frac{\partial w}{\partial t} + \frac{\partial \lambda'}{\partial z} F_o^* + \frac{1}{2}\,\mathrm{tr}\, A \frac{\partial^2 \lambda}{\partial z^2} = o, \quad \lambda = w + \int \{\frac{\partial w}{\partial P}\}\,\pi_t\, d\zeta \\ \frac{\partial \pi_t}{\partial t} - \mathrm{div}(\pi_t F_o^*) + \frac{1}{2}\sum_{i,j=1}^{n+q+1} \frac{\partial^2}{\partial z_i \partial z_j}(A_{ij}\pi_t) = o \\ pr_o(\pi_t) = \pi_o\ ,\ w(T,z,\pi_T) = \tilde{G}(z) \end{cases}$$

where π_t is supposed to be the density of P_t (supposed to exist), the only unusual problem being that we need an approximate expression for $\int \{\frac{\partial w}{\partial P}\}\pi_t d\zeta$ that can be obtained by :

$$(3.88)\qquad \frac{[w(t,z;P^{\Delta}_{t+\varepsilon}) - w(t,z;P^{\Delta}_{t})]}{P_{t+\varepsilon}(\Delta) - P_t(\Delta)} \sim \left\{ \frac{\partial w}{\partial P} \right\}(t,z;\zeta)$$

with Δ a neighborhood of the point ζ, P^{Δ} the restriction of P to Δ. More details can be found in Section IV.2.
It remains after to integrate (3.88) with respect to P_t and to exchange z and ζ to obtain the desired approximation. ∎

Remark 3.8 : in the case $\varkappa_k(t) \equiv t \ \forall t, \ \forall k$, one can introduce a slightly different function $\overline{w}$ defined by : $V(t,P_t) = \int \overline{w}(t,z;\theta,P_\theta)dP_t(z)$ where $\overline{w}(t,z;.,.)$ is invariant along the optimal trajectory (θ,P_θ) passing through (t,P_t) (see for example [22]). This approach needs more regularity but leads to a different equation than (3.80), containing a difference of Lie derivatives of P, in place of the actual signalling term $\frac{\partial \lambda}{\partial z}$. ∎

III.5. Application to the control of partially observed diffusions with classical information

III.5.1. The Hamilton-Jacobi-Bellman theory and the maximum principle

In this section, we assume that $N = 1$, $\varkappa_1(t) \equiv 0 \ \forall t$, namely there is only one decision maker having a perfect memory.

Our purpose is to prove that in this case, the Hamilton-Jacobi-Bellman equation (3.77) can be obtained without the regularity assumptions (H) or (3.79).
A candidate state variable is here the unnormalized conditional density but some slight technical differences must be introduced in our general approach : this density $\pi_t^{u,Y}$ is random and its evolution is given by a stochastic partial differential equation (Zakai equation (3.95)). In order to avoid these technicalities, we use, as in [15], an exponential transformation to change Zakaï's equation in an ordinary partial differential equation whose solution is noted $p_t^{u,Y}$ (equation (3.98)).

After having proved that $p_t^{u,Y}$ can be used as the state variable, (Proposition 3.8), nemely that $p_t^{u,Y}$ has the suitable semi-group property and that the corresponding Value function $V(t,p,Y)$ satisfies the optimality principle, we apply our general dynamic programming method and prove that, because of $p_t^{u,Y}$'s regularity, the assumption (H) is satisfied (proposition 3.10), implying immadiately that the Value function satisfies the Hamilton-Jacobi-Bellman equation (3.109), analog of (3.77). Finally, inverting the exponential transformation, we prove that the value function $\tilde{V}(t,\pi_t^{u,Y},Y)$, with state variable $\pi_t^{u,Y}$, satisfies a stochastic partial differential equation, dual in some sense to Zakaï's equation (corollary 3.4).
To finish this paragraph we prove that the adjoint variable obtained by this method satisfies the maximum principle of A. Bensoussan [6].
The second part of this section establishes the links with Mortensen's equation [24].

The state and observation equations are given by :

$$\begin{aligned} dx_t &= f_o(x_t,u_t)dt + f_1(x_t)dv_t \\ dy_t &= h(x_t)dt + d\tilde{v}_t \quad , \quad y_o = o \end{aligned} \tag{3.89}$$

with :

$$\begin{aligned} &f_o \in C_b^\infty(R^n \times R^m;R^n) \ , \ f_1 \in C_b^\infty(R^m;R^{n\times n}), \ h \in C_b^\infty(R^n;R^p) \\ &a(x) = f_1(x)f_1'(x) \ , \ \xi' a(x)\xi \geqslant \alpha\|\xi\|^2 \ \ \forall \xi \in R^n, \ \forall x \in R^n \ , \end{aligned} \tag{3.90}$$

and v, $\tilde{v}$ independent Wiener processes in R^n and R^p respectively.

We note as before $\mathcal{Y}_t = \sigma\{y_s | o \leq s \leq t\}$, and we have $\mathcal{Y}_{t_1} \subset \mathcal{Y}_{t_2} \quad \forall t_1 \leq t_2$ (classical information structure).

The set $\mathcal{U}$ of admissible controls is, by definition, the set of every $\mathcal{Y}_t$ - measurable processes u_t with values in U, a given closed and convex subset of R^m. We also assume that $f_o(x,U)$ is convex in $R^n \quad \forall x \in R^n$.

Let $\mu \in L^1(R^n) \cap L^2(R^n)$, and $u \in \mathcal{U}$. We note $P^u_{s,\mu}$ the solution of the problem of martingale associated to (3.89) on [s,T].

Let also :

$$(3.91) \qquad z_{s,t} = \exp\left(\int_s^t h'(x_\sigma)dy_\sigma - \frac{1}{2}\int_s^t \|h(x_\sigma)\|^2 d\sigma\right)$$

Then, the measure $Q^u_{s,\mu}$ defined by :

$$(3.92) \qquad \frac{dQ^u_{s,\mu}}{dP^u_{s,\mu}} \Big| \mathcal{F}_t = z^{-1}_{s,t} \qquad \forall t \in [s,T],$$

has the property that, under $Q^u_{s,\mu}$, y is a Wiener process after s, independent of x and v, and ;

$$(3.93) \qquad E_{P^u_{s,\mu}}(\varphi | \mathcal{Y}_t) = E_{Q^u_{s,\mu}}(\varphi z_{s,t} | \mathcal{Y}_t) \qquad \forall \varphi \in C^o_b(R^n),$$

$$(3.93)' \qquad E_{P^u_{s,\mu}}(\varphi) = E_\rho(E_{Q^u_{s,\mu}}(\varphi z_{s,t} | \mathcal{Y}_t)) \qquad \forall \varphi \in C^o_b(R^n),$$

where ρ is Wiener measure on $C^o([o,T]);R^p)$.

Let us also denote $\pi^{u,Y}_t$ the unnormalized conditional density given by :

$$(3.94) \qquad E_{Q^u_{s,\mu}}(\varphi z_{s,t} | \mathcal{Y}_t) = \int_{R^n} \varphi(x)\pi^{u,Y}_t(x)dx.$$

Applying the results of Fleming-Pardoux [15], this density exists in $L^2(\Omega;C^o([s,T];L^2(R^n))) \cap M^2(s,T;H^1(R^n))$ and satisfies Zakaï's equation :

$$(3.95) \qquad \begin{cases} d\pi^{u,Y}_t = ((L^{u,Y}_t)^* \pi^{u,Y}_t dt + \pi^{u,Y}_t h' dy_t , & \forall t \in [s,T] \\ \pi^{u,Y}_s = \mu \end{cases}$$

with :

$$(3.96)\quad L_t^{u,Y}\varphi(x) = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x)\frac{\partial^2\varphi}{\partial x_i \partial x_j}(x) + \sum_{i=1}^{n} f_o^i(x,u_t)\frac{\partial\varphi}{\partial x_i} \qquad \forall\varphi \in C_b^2(R^n).$$

$$= \frac{1}{2}\,\mathrm{tr}\; a(x)\frac{\partial^2\varphi}{\partial x^2}(x) + f_o'(x,u_t)\frac{\partial\varphi}{\partial x}(x)$$

Furthermore, if we make the change of variables :

$$(3.97)\quad p_t^{u,Y}(x) = \pi_t^{u,Y}(x)\exp(-h'(x)y_t), \quad \text{for each fixed } Y \in C^o([o,T];R^p),$$

then $p_t^{u,Y}$ satisfies the non-stochastic partial differential equation :

$$(3.98)\quad \begin{cases} \dfrac{d}{dt}\,p_t^{u,Y} = ((\overset{\vee}{L}{}_t^{u,Y})^* + e_t^{u,Y})p_t^{u,Y} \quad , \quad \forall t \in [s,T] \\ p_s^{u,Y} = \mu \end{cases}$$

with :

$$(3.99)\quad \begin{cases} \overset{\vee}{L}{}_t^{u,Y}\varphi(x) = L_t^{u,Y}\varphi(x) - \left(\dfrac{\partial}{\partial x}\,(h'(x)y_t)\right)'a(x)\dfrac{\partial\varphi}{\partial x}\,(x) \qquad \forall\varphi \in C_b^2(R^n), \\ e_t^{u,Y}(x) = \dfrac{1}{2}\left(\dfrac{\partial}{\partial x}\,(h'(x)y_t)\right)'a(x)\left(\dfrac{\partial}{\partial x}(h'(x)y_t)\right) - L_t^{u,Y}(h'(x)y_t) - \dfrac{1}{2}\,\|h(x)\|^2 \end{cases}$$

Precisely, $p^{u,Y} \in L^2(s,T;H^1(R^n)) \cap C^o(s,T;L^2(R^n))$, and, with (3.90),

$$(3.100)\quad \frac{d}{dt}\,p_t^{u,Y} \in L^2(R^n) \quad \forall t \in [s+\varepsilon,T], \quad \forall u \in U, \quad \forall\, Y \in C^o([o,T];R^p), \quad \forall\varepsilon > o.$$

the cost functional is defined by :

$$(3.101)\qquad J_{s,\mu}(u) = E_{P_{s,\mu}^u}\left(\int_s^T g(x_t,u_t)dt + G(x_T)\right)$$

with :

$$(3.102)\quad g \in C_b^2(R^n\times R^m;R), \quad g(\cdot,u) \in L^2(R^n) \quad \forall u \in U, \quad G \in C_b^2(R^n;R).$$

We also introduce :

$$(3.103)\qquad J_{s,\mu}(u,Y) = \int_s^T < g_t^{u,Y}, \pi_t^{u,Y} > dt + < G, \pi_T^{u,Y} >$$

where $g_t^{u,Y}(x) = g(x,u_t(Y_t))$, and $< g,\pi > = \int_{R^n} g(x)\pi(x)dx$.

Noting :

$$(3.104)\quad \tilde{g}_t^{u,Y}(x) = g_t^{u,Y}(x)\exp(h'(x)y_t),\quad \tilde{G}^Y(x) = G(x)\exp(h'(x)y_T),$$

we have, by (3.97) ;

$$(3.105)\qquad J_{s,\mu}(u,Y) = \int_s^T < \tilde{g}_t^{u,Y},\ p_t^{u,Y} > dt + < \tilde{G}^Y, p_T^{u,Y} >.$$

Of course : $E_\rho(J_{s,\mu}(u,Y)) = J_{s,\mu}(u)$.

Finally, let us introduce :

$$(3.106)\qquad V(s,\mu,Y) = \inf_{u \in \mathcal{U}} J_{s,\mu}(u,Y)$$

Proposition 3.8 :

(i) $p_t^{u,Y} = \Pi^{u,Y}(t,s)\mu$, where $\Pi^{u,Y}$ is the strongly continuous semi-group generated by $((\overset{\vee}{L}{}_t^{u,Y})^* + e_t^{u,Y})$ in $C^0(o,T;L^2(R^n))$.

(ii) $J_{s,\mu}(u,Y) = \int_s^t < \tilde{g}_\sigma^{u,Y}, \Pi^{u,Y}(\sigma,s)\mu > d\sigma + J_{t,\Pi^{u,Y}(t,s)\mu}(u,Y)$,
$\forall t \in [s,T]$, $\forall u \in \mathcal{U}$, $\forall Y \in C^0(o,T;R^p)$, $\forall \mu$, and :

(iii) $V(s,\mu,Y) = \inf_{u \in \mathcal{U}}[[V(t,\Pi^{u,Y}(t,s)\mu,Y) + \int_o^t < \tilde{g}_\sigma^{u,Y}, \Pi^{u,Y}(\sigma,s)\mu > d\sigma]$

Proof : (i) is classical and (ii), (iii) follow exactly the same lines as in the preceding sections. ∎

As before, we need to extend the value function $V(s,\mu,Y)$ for any $\mu \in \mathcal{M}^b(R^n)$. This can be obtained by considering $p_t^{u,Y}$ as the distributional solution of :

$$(3.107)\quad \begin{cases} < \varphi, \frac{d}{dt} p_t^{u,Y} > = < (\overset{\vee}{L}{}_t^{u,Y} + e_t^{u,Y})\varphi,\ p_t^{u,Y} > \quad \forall \varphi \in C_b^2(R^n) \\ p_s^{u,Y} = \mu \end{cases}$$

and $\Pi^{u,Y}(t,s)$ can be extended by density to $\mathcal{L}(\mathcal{M}^b(R^n),\mathcal{M}^b(R^n))$.

Proposition 3.9 : $V(t,\mu,Y)$ is finite on $[o,T] \times \mathcal{M}^b(R^n)$ $\forall Y \in C^0(o,T;R^p)$ and $\mu \to V(t,\mu,Y)$ is concave, continuous and subdifferentiable.

Furthermore the subdifferential $\partial_\mu V(t,\mu,Y)$ is compact in $C_b^o(R^n)$.

Proof : The finiteness of V is obvious by (3.9.) and (3.102).
Let $\alpha, \beta > o$, $\alpha + \beta = 1$, and $\mu,\nu \in \mathcal{M}^b(R^n)$.
We have :

$$\alpha V(t,\mu,Y) + \beta V(t,\nu,Y) \leq \alpha\, J_{t,\mu}(u,Y) + \beta\, J_{t,\nu}(u,Y) = J_{t,\alpha\mu+\beta\nu}(u,Y)$$

$$\forall u \in U, \quad \forall\, Y \in C^o(o,T;R^p), \quad \forall t \in [o,T].$$

and thus : $\alpha\, V(t,\mu,Y) + \beta\, V(t,\nu,Y) \leq V(t,\alpha\mu + \beta\nu,Y)$

which proves the concavity, and the proposition follows by classical convex analysis results (see [26]). ∎

As before, we have to extend the cost functional and therefore the value function to closed-loop strategies $u \in \tilde{U}$, namely the set of $\mathcal{B}_{\mathcal{M}^p(R^n)} \times \mathcal{Y}_t$ adapted processes with values in U. This extension is made exactly as in section III.2.3 and is left to the reader. One can prove exactly as in the general case that the extented value function for closed loop controls coïncides with the value function for $\mathcal{Y}_t$ adapted controls (open-loop).

Proposition 3.10 : Let $\mu \in L^1(R^n) \cap L^2(R^n)$ and $t > s$. Then the following directional derivative formula holds true :

$$(3.108) \quad \lim_{\varepsilon \to o} \frac{1}{\varepsilon} \left(V(t,\Pi^{u,Y}(t+\varepsilon,s)\mu,Y) - V(t,\Pi^{u,Y}(t,s)\mu,Y)\right)$$

$$= \min_{\lambda \in \partial_\mu V(t,\Pi^{u,Y}(t,s)\mu,Y)} < (\overset{\vee}{L}{}_t^{u,Y} + e_t^{u,Y})\lambda,\ \Pi^{u,Y}(t,s)\mu >,$$

this limit being finite $\forall u \in U$, $\forall\, Y \in C^o(o,T;R^p)$.

Proof : The only thing to prove is that the directional derivative formula can be applied, namely that :

$$\lim_{\varepsilon \to o} \frac{1}{\varepsilon} \left(\Pi^{u,Y}(t+\varepsilon,s)\mu - \Pi^{u,Y}(t,s)\mu\right) = \frac{d}{dt}\, \Pi^{u,Y}(t,s)\mu \in \mathcal{M}^b(R^n).$$

But this is a consequence of the regularity property (3.100) of the solution of (3.98) since $\frac{d}{dt}\, p_t^{u,Y} = \frac{d}{dt}\, \Pi^{u,Y}(t,s)\mu \in L^2(R^n) \subset \mathcal{M}^b(R^n)$. Finally, since :

$$< \lambda, \frac{d}{dt}\, \Pi^{u,Y}(t,s)\mu > = < \lambda, ((\overset{\vee}{L}{}_t^{u,Y})^* + e_t^{u,Y})\Pi^{u,Y}(t,s)\mu >$$

$$= < (\overset{\vee}{L}{}_t^{u,Y} + e_t^{u,Y})\lambda,\ \Pi^{u,Y}(t,s)\mu >, \quad \text{the proposition is proved.} \ ∎$$

Theorem 3.5 : Let $\mu_o \in L^2(R^n) \cap L^1(R^n)$. A necessary and sufficient condition for $\bar{u}$ to be optimal is that there exists a function V defined on $[o,T] \times \mathcal{M}^b(R^n) \times C^o([o,T];R^p)$, continuous and concave on $\mathcal{M}^b(R^n)$ for every fixed t,Y, right-differentiable on $[o,T]$ $\forall \mu,Y$, and measurable on $C^o([o,T];R^p)$ $\forall t,\mu$, satisfying the following Hamilton-Jacobi-Bellman equation :

$$
\begin{aligned}
&\frac{\partial V}{\partial t_+}(t,p_t,Y) + \min_{u \in U} \min_{\lambda_t \in \partial_\mu V(t,p_t,Y)} < (\overset{\vee}{L}{}_t^{u,Y} + e_t^{u,Y})\lambda_t + \tilde{g}_t^{u,Y}, p_t > = o \\
&= \frac{\partial V}{\partial t_+}(t,p_t,Y) + \min_{\lambda_t \in \partial_\mu V(t,p_t,Y)} < (\overset{\vee}{L}{}_t^{\bar{u},Y} + e_t^{\bar{u},Y})\lambda_t + \tilde{g}_t^{\bar{u},Y}, p_t > \qquad (3.109) \\
&\forall t \in]o,T], \quad \forall Y \in C^o([o,T];R^p), \quad \forall p_t \in \{\Pi^{u,Y}(t,o)\mu_o \mid u \in \mathcal{U}\}, \\
&V(T,p,Y) = < \tilde{G}^Y, p > \quad \forall p \in \mathcal{M}^b(R^n), \quad \forall Y \in C^o([o,T];R^p)
\end{aligned}
$$

Proof : From (iii), proposition 3.8, we have, $\forall p_t \in \{\Pi^{u,Y}(t,o)\mu_o \mid u \in \mathcal{U}\}$:

$$
\begin{aligned}
&\frac{1}{\varepsilon} \inf_{u \in \mathcal{U}} [V(t+\varepsilon, \Pi^{u,Y}(t+\varepsilon,t)p_t, Y) - V(t,p_t,Y) \\
&\qquad + \int_t^{t+\varepsilon} < \tilde{g}_\sigma^{u,Y}, \Pi^{u,Y}(\sigma,t)p_t > d\sigma] = o \\
&= \inf_{u \in \mathcal{U}} [\frac{1}{\varepsilon}(V(t+\varepsilon, \Pi^{u,Y}(t+\varepsilon,t)p_t, Y) - V(t, \Pi^{u,Y}(t+\varepsilon,t)p_t, Y)) \\
&\qquad + \frac{1}{\varepsilon}(V(t, \Pi^{u,Y}(t+\varepsilon,t)p_t, Y) - V(t,p_t,Y)) \\
&\qquad + \frac{1}{\varepsilon}\int_t^{t+\varepsilon} < \tilde{g}_\sigma^{u,Y}, \Pi^{u,Y}(\sigma,t)p_t > d\sigma]
\end{aligned}
$$

Taking the limit as $\varepsilon \to o$, the last term converges to :

$$< \tilde{g}_t^{u,Y}, p_t >$$

The second term, by (3.108), converges to :

$$\min_{\lambda \in \partial_\mu V(t,p_t,Y)} < (\overset{\vee}{L}{}_t^{u,Y} + e_t^{u,Y})\lambda, p_t > ,$$

and thus the limit of the first term also exists, equal to $\frac{\partial V}{\partial t_+}(t,p_t,Y)$, and (3.109) is proved with $V(T,p,Y) = \langle \tilde{G}^Y, p \rangle$.
Conversely, (3.109) can be rewritten as :

$$\frac{d}{dt}\left(V(t,\Pi^{\bar{u},Y}(t,o)\mu_o,Y) + \int_o^t \langle \tilde{g}_\sigma^{\bar{u},Y}, \Pi^{\bar{u},Y}(\sigma,o)\mu_o \rangle \, d\sigma\right) = 0$$

$$\leqslant \frac{d}{dt}\left(V(t,\Pi^{u,Y}(t,o)\mu_o,Y) + \int_o^t \langle \tilde{g}_\sigma^{u,Y}, \Pi^{u,Y}(\sigma,o)\mu_o \rangle \, d\sigma\right) \quad \forall u \in \mathcal{U}$$

and, integrating between o and T, one obtains :

$$\int_o^T \langle \tilde{g}_t^{\bar{u},Y}, \Pi^{\bar{u},Y}(t,o)\mu_o \rangle \, dt + \langle \tilde{G}^Y, \Pi^{\bar{u},Y}(T,o)\mu_o \rangle = V(o,\mu_o,Y)$$

$$\leqslant \int_o^T \langle \tilde{g}_t^{u,Y}, \Pi^{u,Y}(t,o)\mu_o \rangle \, dt + \langle \tilde{G}^Y, \Pi^{u,Y}(T,o)\mu_o \rangle \quad \forall u \in \mathcal{U}$$

or, by (3.105) : $J_{o,\mu_o}(\bar{u},Y) \leqslant J_{o,\mu_o}(u,Y) \quad \forall u \in \mathcal{U}$, and finally :

$$J_{o,\mu_o}(\bar{u}) = E_\rho(J_{o,\mu_o}(\bar{u},Y)) \leqslant E_\rho(J_{o,\mu_o}(u,Y)) = J_{o,\mu_o}(u)$$

$\forall u \in \mathcal{U}$, which achieves to prove the theorem. ∎

Remark 3.9 : since in (3.109) $\lambda_t^Y \in \partial_\mu V(t,p_t,Y)$, it results that $\lambda_t^Y \in C_b^o(R^n) \quad \forall t,Y$. ∎

Corollary 3.4 : The equation (3.109) can be transformed, by the change of variable :

$$\tilde{V}(t,\pi_t^{u,Y},Y) = V(t,p_t^{u,Y},Y), \quad \pi_t^{u,Y} = p_t^{u,Y} \exp(h' y_t) \tag{3.110}$$

into the stochastic partial differential equation :

$$d\tilde{V}_t(t,\pi_t^Y,Y) + \min_{u \in \mathcal{U}} \min_{\tilde{\lambda}_t \in \partial_\mu \tilde{V}(t,\pi_t^Y,Y)} \left(\langle L_t^{u,Y}\tilde{\lambda}_t + g_t^{u,Y}, \pi_t^Y \rangle \, dt + \langle \tilde{\lambda}_t h', \pi_t^Y \rangle \, dy_t\right) = 0 \tag{3.111}$$

$$\tilde{V}(T,\pi_T^Y,Y) = \langle G, \pi_T^Y \rangle$$

and $\tilde{\lambda}_t = \lambda_t \exp(h'(y_t))$.

Proof : Requires tedious but not difficult calculations. ∎

Theorem 3.6 : let $\mu_o \in L^1(R^n) \cap L^2(R^n)$, and $\bar{u} \in \mathcal{U}$ be an optimal control. We denote $\pi_t^{\bar{u},Y}$ the corresponding unnormalized conditional density starting form μ_o, and $\tilde{\lambda}_t^Y \in \partial_\mu \tilde{V}(t,\pi_t^{\bar{u},Y},Y)$ satisfying, with the notations of (3.111) :

$$(3.112) \qquad \min_{\lambda \in \partial_\mu \tilde{V}(t,\pi_t^{\bar{u},Y},Y)} < \lambda, d\pi_t^{\bar{u},Y} > = < \tilde{\lambda}_t^Y, d\pi_t^{\bar{u},Y} >$$

Then $\tilde{\lambda}_t^Y$ is the unique solution of the adjoint equation :

$$- d\tilde{\lambda}_t^Y + (L_t^{\bar{u},Y} \tilde{\lambda}_t^Y + \tilde{\lambda}_t^Y \|h\|^2)dt = (\tilde{\lambda}_t^Y h' - \exp(h'y_t) r_t')dy_t + (g_t^{\bar{u},Y} + r_t' h \exp -(h'y_t))dt$$

$$(3.113) \qquad \tilde{\lambda}_T^Y(x) = G(x)$$

$$\tilde{\lambda} \in L^2_Y(o,T;H^1(R^n)) \cap L^\infty(o,T;L^4(\Omega;H^1(R^n)))$$

$$\tilde{\lambda}_t^Y \exp h'y_t \in L^2(\Omega;C^o(o,T;H^{-1}(R^n))).$$

Proof : It suffices to remark that (3.108) implies (2.19) of [6] with :

$$\lim_{\theta \to o} \frac{1}{\theta} (\tilde{V}(t,\Pi^{\bar{u}+\theta u,Y}(t,s)\mu,Y) - \tilde{V}(t,\Pi^{\bar{u},Y}(t,s)\mu,Y)) = \min_{\lambda \in \partial_\mu \tilde{V}(t,\Pi^{\bar{u},Y}(t,s)\mu,Y)} < \lambda, \frac{d}{d\theta} \Pi^{\theta,Y}{}_o u >$$

and the result is a restatement of theorem 2.2. of [6] . ∎

Remark 3.10 : When V is differentiable with respect to p, using the representation of proposition 3.5 together with the assumption (3.73), one obtains :

$$(3.114) \qquad \tilde{\lambda}_t^Y(x) = (w(t,x;p_t^Y) + \int_{R^n} \{ \frac{\partial w}{\partial p} \} (t,\xi,x) dp_t^Y(\xi)) \exp(h'(x)y_t)$$

and $$V(t,p_t^Y,Y) = \int_{R^n} w(t,x;p_t^Y) dp_t^Y(x).$$

From Corollary 3.4 to the end, $< \tilde{\lambda}_t, d\pi_t^{u,Y} >$ is written formally since it contains the stochastic integral:

$$\int_o^t < \tilde{\lambda}_s h', \pi_s^{u,Y} > dy_s$$

which is not a priori an Itô integral:

$\tilde{\lambda}_s$ is $\mathcal{Y}_s^T$ - measurable whereas $\pi_s^{u,Y}$ is $\mathcal{Y}_o^s$ - measurable.

This integral must thus be defined through the preceding transformation :

$\tilde{\lambda}_t = \lambda_t \exp(h'y_t)$, $\pi_t^{u,Y} = p_t^{u,Y} \exp(h'y_t)$.

The equations (3.118) and (3.119) are of course written in the same formal way. ■

To conclude this paragraph, we have proved that the machinery developed in the preceding sections can be completely justified in the case of the control of partially observed diffusions with classical information. Furthermore, it provides a necessary and sufficient condition of optimality, improving the results of Bensoussan [6].

III.5.2. The link with Mortensen's equation

In this classical information structure, it appears that there are two fundamental diffusion processes: the state and the filter. In III.5.1., both are used. However, the early approach of Mortensen [24] is based on filtering properties only. Namely, one considers the semi-group $\Pi^u(t;s,\mu_s)$ defined as in (3.33) by :

$$(3.115) \qquad \Pi^u(t;s,\mu_s)(B) = Q^u_{s,\mu_s}(B) \qquad \forall B \in \mathcal{F}^n_{o,t}$$

with Q^u_{s,μ_s} defined by (3.92),
and :

$$(3.116) \qquad V(t,\Pi^u(t;s,\mu_s)) = \inf_{u \in \mathcal{U}} J_{s,\mu_s}(u).$$

We clearly have :

$$(3.117) \qquad V(t,\Pi^u(t;s,\mu_s)) = E_\rho(\tilde{V}(t,\pi_t^{u,Y})).$$

Assuming that $\tilde{V}$ is smooth enough, following [35], [24], or [7] in a different context, the following Itô's formula can be proved :

$$(3.118)\quad \tilde{V}(t,\pi_t^{u,Y}) - \tilde{V}(s,\pi_s^{u,Y}) = \int_s^t \Big(\frac{\partial \tilde{V}}{\partial \sigma}(\sigma,\pi_\sigma^{u,Y}) + \Big\langle L_\sigma^u \frac{\partial \tilde{V}}{\partial \pi}(\sigma,\pi_\sigma^{u,Y}),\pi_\sigma^{u,Y}\Big\rangle + \frac{1}{2}\Big\langle h\,\pi_\sigma^{u,Y} \frac{\partial^2 \tilde{V}}{\partial \pi^2}(\sigma,\pi_\sigma^{u,Y}),h'\pi_\sigma^{u,Y}\Big\rangle\Big)d\sigma + \int_s^t \Big\langle h' \frac{\partial \tilde{V}}{\partial \pi}(\sigma,\pi_\sigma^{u,Y}),\pi_\sigma^{u,Y}\Big),\pi_\sigma^{u,Y}\Big\rangle dy_\sigma$$

with $L_t^u\varphi = \frac{1}{2}\mathrm{tr}\, a \frac{\partial^2\varphi}{\partial x^2} + f_o' \frac{\partial \varphi}{\partial x}$, $\frac{\partial \tilde{V}}{\partial \pi}$ being the Frechet derivative of $\tilde{V}$ with respect to the second variable, and $\frac{\partial^2 \tilde{V}}{\partial \pi^2}$ being the second Frechet derivative of $\tilde{V}$ with respect to π.

Thus, from (3.118), it is easy to obtain the following Hamilton - Jacobi - Bellman equation. :

$$(3.119)\quad E_\rho\Big(\frac{\partial \tilde{V}}{\partial t} + \min_{u \in U} \Big(\Big\langle L_t^u \frac{\partial \tilde{V}}{\partial \pi}, \pi_t^Y\Big\rangle\Big) + \frac{1}{2}\Big\langle h\pi_t^Y \frac{\partial^2 \tilde{V}}{\partial \pi^2}, h'\pi_t^Y\Big\rangle\Big) = 0$$

which is known as Mortensen's equation (see [24]).
When $\tilde{V}$ is smooth enough, the adjoint in (3.119) is $\frac{\partial \tilde{V}}{\partial \pi}$, and it can be easily checked that it coincides with the one obtained in (3.114). Nevertheless, this approach is formal whereas the one of III.5.1 is rigorous. ■

IV - EXAMPLES

IV.1. A nonlinear quadratic team problem

Let x be a scalar stochastic process observed by 2 stations :

$$(4.1)\qquad \begin{cases} dx_t = (u_1 f_1(x_t) + u_2 f_2(x_t))dt + dv_t \\ dy^1_t = h_1(x_t)dt + d\nu^1_t \\ dy^2_t = h_2(x_t)dt + d\nu^2_t \end{cases}$$

u_1 and u_2 are scalars without constraints; f_1, f_2, h_1, h_2 are smooth functions of x ; $\nu_1(t) = \nu_2(t) = o \quad \forall t$ (perfect memory for each station).
The cost function is given by :

$$(4.2)\qquad J(u_1,u_2) = E\Big(\int_o^T (g_o(x_t) + u_1 g_1(x_t) + u_2 g_2(x_t) + \gamma_1 u_1^2 + \gamma_2 u_2^2 + 2\gamma u_1 u_2)dt\Big)$$

with g_o, g_1, g_2 smooth functions and $\gamma_1, \gamma_2 > 0$.
To apply the preceding theory, we introduce :

$$(4.3)\qquad d\zeta_t = g_o(x_t) + u_1 g_1(x_t) + u_2 g_2(x_t) + \gamma_1 u_1^2 + \gamma_2 u_2^2 + 2\gamma u_1 u_2 \ , \quad \zeta_o = o$$

and $\quad z = (x, y^1, y^2, \zeta)'$.
Thus : $J(u_1,u_2) = E(\zeta_T)$.
The Hamiltonians are :

$$(4.4)\qquad H_1 = E\Big(\frac{\partial\lambda}{\partial x}(u_1 f_1 + u_2^* f_2) + \frac{\partial\lambda}{\partial y^1} h_1 + \frac{\partial\lambda}{\partial y^2} h_2 + \frac{\partial\lambda}{\partial\zeta}(g_o + u_1 g_1 + u_2^* g_2 + \gamma_1 u_1^2 + \gamma_2 (u_2^*)^2 + 2\gamma u_1 u_2^*) \,|\, Y^1_t\Big)$$

$$(4.5)\qquad H_2 = E\Big(\frac{\partial\lambda}{\partial x}(u_1^* f_1 + u_2 f_2) + \frac{\partial\lambda}{\partial y^1} h_1 + \frac{\partial\lambda}{\partial y^2} h_2 + \frac{\partial\lambda}{\partial\zeta}(g_o + u_1^* g_1 + u_2 g_2 + \gamma_1 (x_1^*)^2 + \gamma_2 u_2^2 + 2\gamma u_1^* u_2) \,|\, Y^2_t\Big)$$

with $\quad Y^k_t = \{y^k_\sigma \ , \ o \leq \sigma \leq t\}$, $k = 1,2$; (u_1^*, u_2^*) is obtained by :

$$(4.6)\quad \begin{cases} \dfrac{\partial}{\partial u_1}\left[E\left(\dfrac{\partial\lambda}{\partial x} f_1 + \dfrac{\partial\lambda}{\partial\zeta}(g_1 + 2\gamma u_2^*)\,\middle|\, Y_t^1\right)u_1 + \gamma_1 u_1^2\, E\left(\dfrac{\partial\lambda}{\partial\zeta}\,\middle|\, Y_t^1\right)\right] = 0 \\ \dfrac{\partial}{\partial u_1}\left[E\left(\dfrac{\partial\lambda}{\partial x} f_2 + \dfrac{\partial\lambda}{\partial\zeta}(g_2 + 2\gamma u_1^*)\,\middle|\, Y_t^2\right)u_2 + \gamma_2 u_2^2\, E\left(\dfrac{\partial\lambda}{\partial\zeta}\,\middle|\, Y_t^2\right)\right] = 0 \end{cases}$$

Remark that (4.6) corresponds to the minimization of two coupled strictly convex functions of (u_1,u_2) and the minimum is <u>unique</u>. Thus :

$$(4.7)\quad \begin{cases} u_1^*(t, Y_t^1, P_t) = -\dfrac{1}{2\gamma_1 E_{P_t}\left(\dfrac{\partial\lambda}{\partial\zeta}\middle| Y_t^1\right)}\; E_{P_t}\left(\dfrac{\partial\lambda}{\partial x} f_1 + \dfrac{\partial\lambda}{\partial\zeta}(g_1 + 2\gamma u_2^*)\middle| Y_t^1\right) \\ u_2^*(t, Y_t^2, P_t) = -\dfrac{1}{2\gamma_2 E_{P_t}\left(\dfrac{\partial\lambda}{\partial\zeta}\middle| Y_t^2\right)}\; E_{P_t}\left(\dfrac{\partial\lambda}{\partial\zeta} f_2 + \dfrac{\partial\lambda}{\partial\zeta}(g_2 + 2\gamma u_2^*)\middle| Y_t^2\right) \end{cases}$$

It must be noted that (4.7) is a system of <u>2 integral equations</u> in u_1^*,u_2^* since, if we denote $\alpha_i = \dfrac{1}{2\gamma_i E_{P_t}\left(\dfrac{\partial\lambda}{\partial\zeta}\middle| Y_t^i\right)}$, $i=1,2$, we obtain :

$$(4.8)\quad \begin{cases} u_1^*(t, Y_t^1, P_t) = 2\gamma\alpha_1 \displaystyle\int_{C^4_{o,t}} \frac{\partial\lambda}{\partial\zeta}\, u_2^*(t,Y^2,P_t)\, dP_t(X,Y^1,Y^2,\zeta\,|\,Y_t^1) \\ \qquad\qquad + \alpha_1 \displaystyle\int_{C^4_{o,t}} \left(\frac{\partial\lambda}{\partial x} f_1 + \frac{\partial\lambda}{\partial\zeta} g_1\right) dP_t(X,Y^1,Y^2,\zeta\,|\,Y_t^1) \\ u_2^*(t, Y_t^2, P_t) = 2\gamma\alpha_2 \displaystyle\int_{C^4_{o,t}} \frac{\partial\lambda}{\partial\zeta}\, u_1^*(t,Y^1,P_t)\, dP_t(X,Y^1,Y^2,\zeta\,|\,Y_t^2) \\ \qquad\qquad + \alpha_2 \displaystyle\int_{C^4_{o,t}} \left(\frac{\partial\lambda}{\partial x} f_2 + \frac{\partial\lambda}{\partial\zeta} g_2\right) dP_t(X,Y^1,Y^2,\zeta\,|\,Y_t^2) \end{cases}$$

Nevertheless, when there is <u>no instantaneous coupling in the cost between the two controls</u>, namely when $\gamma=0$, (4.8) can easily be solved :

$$(4.9)\quad u_1^* = \alpha_1 E\left(\frac{\partial\lambda}{\partial x} f_1 + \frac{\partial\lambda}{\partial\zeta} g_1\,\middle|\, Y_t^1\right),\quad u_2^* = \alpha_2 E\left(\frac{\partial\lambda}{\partial x} f_2 + \frac{\partial\lambda}{\partial\zeta} g_2\,\middle|\, Y_t^2\right)$$

and the value function satisfies :

$$(4.10)\quad \begin{cases} E_{P^*_t}\Big[\dfrac{\partial w}{\partial t} + \dfrac{1}{2}\,\Delta_{x,y}\lambda + \dfrac{\alpha_1}{2}\,E_{P^*_t}\Big(\dfrac{\partial \lambda}{\partial x}\,f_1 + \dfrac{\partial \lambda}{\partial \zeta}\,g_1\,|\gamma^1_t\Big)^2 \\ \qquad\qquad + \dfrac{\alpha_2}{2}\,E_{P^*_t}\Big(\dfrac{\partial \lambda}{\partial x}\,f_2 + \dfrac{\partial \lambda}{\partial \zeta}\,g_2\,|\gamma^2_t\Big)^2 \\ \qquad\qquad + \dfrac{\partial \lambda}{\partial y^1}\,h_1 + \dfrac{\partial \lambda}{\partial y^2}\,h_2 + \dfrac{\partial \lambda}{\partial \zeta}\,g_o\Big] = 0 \\ w(T,z,P) = \zeta \quad \forall z = (x,y^1,y^2,\zeta),\ \forall P, \end{cases}$$

with $\Delta_{x,y}\lambda = \dfrac{\partial^2\lambda}{\partial x^2} + \dfrac{\partial^2\lambda}{\partial (y^1)^2} + \dfrac{\partial^2\lambda}{\partial (y^2)^2}$,

$$\lambda(t,x,y^1,y^2,\zeta;P^*_t) = w(t,x,y^1,y^2,\zeta;P^*_t) + \int \Big\{ \frac{\partial w}{\partial P} \Big\}(t,\xi;z)\,dP^*_t(t)(\xi),$$

and $$V(t,P^*_t) = \int w(t,x,y^1,y^2,\zeta;P^*_t)\,dP^*_t(t)(x,y^1,y^2,\zeta)\ ,$$

P^* standing for the probability measure generated by u^*_1 , u^*_2 between 0 and t, given by :

$$(4.11)\quad \begin{cases} \dfrac{d}{dt}\displaystyle\int_{C_{o,t}} \varphi\,dP^*_t = \int_{C_{o,t}} L_t^{u^*_1,u^*_2}(\varphi)\,dP^*_t \quad \forall \varphi \in C^2_b(R^4) \\ P^*_t\big|_{t=o} = P_o \end{cases}$$

with :

$$\begin{aligned}(4.12)\quad L_t^{u^*_1,u^*_2}(\varphi)(Z) &= \frac{\partial \varphi}{\partial x}(Z_t)\big[f_1(X_t)u^*_1(Y^1,P^*_t) + f_2(X_t)u^*_2(Y^2,P^*_t)\big] \\ &+ \frac{\partial \varphi}{\partial y^1}(Z_t)h_1(X_t) + \frac{\partial \varphi}{\partial y^2}(Z_t)h_2(X_t) + \frac{\partial \varphi}{\partial \zeta}(Z_t)\big[g_o(X_t) + g_1(X_t)u^*_1(Y^1,P^*_t) \\ &+ g_2(X_t)u^*_2(Y^2,P^*_t) + \gamma_1(u^*_1(Y^1,P^*_t))^2 + \gamma_2(u^*_2(Y^2,P^*_t))^2\big] \\ &+ \frac{1}{2}\Big[\frac{\partial^2\varphi}{\partial x^2}(Z_t) + \frac{\partial^2\varphi}{\partial (y^1)^2}(Z_t) + \frac{\partial^2\varphi}{\partial (y^2)^2}(Z_t)\Big].\end{aligned}$$

A first remark, to simplify a little (4.10), is that :

$$(4.13)\quad \frac{\partial \lambda}{\partial \zeta} \equiv 1 \quad \text{if} \quad w, \ \text{solution of} \ (4.10), \ \text{is regular enough.}$$

This can be seen from the fact that $\frac{\partial w}{\partial \zeta}(T,z,P) \equiv 1$,

and $\{ \frac{\partial w}{\partial P} \}(T,z;\xi) \equiv 0$, since $w(t,z,P) = \zeta \quad \forall P$.

Thus : $\frac{\partial \lambda}{\partial \zeta}(T,z,P) = 1 \quad \forall z, \ \forall P$.

But, going back to (4.10), one can evaluate $w(T-\varepsilon,z,P)$ by a Taylor's expansion in ε, from the knowledge of w and λ at $t = T$, and since $\frac{\partial \lambda}{\partial x}\Big|_{t=T} = \frac{\partial \lambda}{\partial y^i}\Big|_{t=T} = 0$, $i=1,2$, and g_o, g_1 and g_2 are functions of x only, one can see that $w(T-\varepsilon,z,P) = \zeta - \Phi_\varepsilon(x,Y^1,Y^2)$ with $\frac{\partial \Phi_\varepsilon}{\partial \zeta} \equiv 0$. Thus $\frac{\partial w}{\partial \zeta}\Big|_{t=T-\varepsilon} = 1$ and $\{ \frac{\partial w}{\partial P} \}_{|t=T-\varepsilon} = \{ \frac{\partial \Phi_\varepsilon}{\partial P} \}_{|t=T-\varepsilon}$ but $\frac{\partial}{\partial \zeta} \{ \frac{\partial \Phi_\varepsilon}{\partial P} \}_{|t=T-\varepsilon} \equiv 0$,

thus $\frac{\partial \lambda}{\partial \zeta}_{|t=T-\varepsilon} \equiv 1$ for ε sufficiently small. But, using the same argument on $[T-2\varepsilon, T-\varepsilon]$ and the fact that $f_1, f_2, g_o, g_1, g_2, h_1, h_2$ are functions of x only, we find that $\frac{\partial \lambda}{\partial \zeta} \equiv 1 \quad \forall t \in [o,T] \ \forall z, \ \forall P$.

Thus u_1^* , u_2^* and (4.10) can be rewritten :

$$(4.9)' \quad \begin{cases} u_1^*(Y_t^1 , P_t) = -\frac{1}{2\gamma_1} E_{P_t}(\frac{\partial \lambda}{\partial x} f_1 + g_1 | Y_t^1), \\ u_2^*(Y_t^2 , P_t) = -\frac{1}{2\gamma_2} E_{P_t}(\frac{\partial \lambda}{\partial x} f_2 + g_2 | Y_t^2) \end{cases}$$

and :

$$(4.10)' \quad \begin{cases} E_{P_t^*}(\frac{\partial w}{\partial t} + \frac{1}{2} \Delta_{x,y} \lambda - \frac{1}{4\gamma_1} E_{P_t^*}(\frac{\partial \lambda}{\partial x} f_1 + g_1 | \mathcal{Y}_t^1)^2 - \frac{1}{4\gamma_2} E_{P_t^*}(\frac{\partial \lambda}{\partial x} f_2 + g_2 | \mathcal{Y}_t^2)^2 \\ \qquad + \frac{\partial \lambda}{\partial y^1} h_1 + \frac{\partial \lambda}{\partial y^2} h_2 + g_o) = 0 \\ w(T,z,P) = \zeta \quad \forall x, y^1, y^2, \ \forall P. \end{cases}$$

Remark also that in (4.10) or (4.10)', u_i^* is a progressively measurable function of the process Y^i, which can be considered as a current variable for the integration with respect to P_t^* and their expressions in (4.10), (4.10)' make sense.

Now, if we try to give more insight into qualitative properties of the solution of (4.9)', (4.10)', (4.11), (4.12), some comments may be necessary :

. Firstly, (4.9)' states that the optimal strategy for each decision maker is obtained through an optimal estimate, namely the conditional expectation of $(\frac{\partial \lambda}{\partial x} f_i + g_i)$ knowing Y_t^i and thus, it confirms the intuitive rule that "a good control may be obtained through a good estimate".

. Secondly, this optimal estimate is obtained for each decision maker as a linear functional of his filter : namely, if we note :

$$(4.14) \qquad \pi_t^i(z|Y_t^i) = P_t^*(Z|Z_t = z, Y^i = Y_t^i), \qquad i=1,2,$$

the conditional measure of the state z knowing Y_t^i (the filter of player i), then $u_i^*(Y_t^i, P_t^*) = < \frac{\partial \lambda}{\partial x} f_i + g_i, \pi_t^i(.|Y_t^i) >$. π_t^i may be obtained via the solution of Zakaï's equation for its unnormalized version.
Remark also that the filter depends on u_1^* and u_2^*, as can be seen in (4.11), (4.12), (4.14) and this is generally called the "dual effect" : the controller minimizes the cost on the one hand and, on the other hand, tries to improve the quality of his information by choosing a good observation path.

It must be remarked that the filters of players 1 and 2 are not in general state variables since one cannot recover P_t^* from π_t^1, π_t^2, whereas P_t^* is needed to compute λ (compare to III.4).
This displays the gap between classical and non classical information structures. Furthermore, remark that, even in the classical case, the well-known Separation Principle does not hold in general. Thus, a fortiori, in our non classical information case (see [10],[17],[21],[30],[31],[33]).

. Thirdly, the only difference between the minimization of our Hamiltonians and the classical information's ones, at least when there is no instantaneous coupling in the cost function between the controls ($\gamma = 0$), lies in the ajoint's formula, and more precisely, in the signalling term : $\int \{ \frac{\partial W}{\partial P} \}(t,\xi,z) dP_t(t)(\xi)$.
A numerical expression for this signalling term can be obtained from (3.92).
Remark that $\frac{\partial \lambda}{\partial z} = 0$ can happen when the trade-off between cost and information is such that :

$$(4.15) \qquad \frac{\partial W}{\partial z} = - \frac{\partial}{\partial z} \int \{ \frac{\partial W}{\partial P} \} dP_t$$

and, in such a case, and if there is no cost on the controls ($g_1 = g_2 = \gamma_1 = \gamma_2 = 0$), both Hamiltonians are singular simultaneously. This situation seems to be qualitatively very specific to non classical information structures.

Let us now deal with the same problem when $f_1, f_2, g_0, g_1, g_2, h_1$ and h_2 are linear.

IV.2. A linear quadratic team problem

IV.2.1. The optimality conditions

The system is given by :

$$(4.16)\qquad \begin{cases} dx_t = (Fx + u_1 + u_2)dt + dv_t \\ dy_t^1 = H_1 x_t dt + dv_t^1 \\ dy_t^2 = H_2 x_t dt + dv_t^2 \end{cases}$$

and the information structure is the same as in IV.1.

The cost function is quadratic :

$$(4.17)\qquad J(u_1,u_2) = E\left(\int_0^T (x_t^2 + g_1 u_1 x_t + g_2 u_2 x_t + \gamma_1 u_1^2 + \gamma_2 u_2^2 + 2\gamma u_1 u_2)dt\right)$$

Applying the preceding method, we find that the Hamiltonians are :

$$(4.18)\qquad H_1 = E\Big(\frac{\partial\lambda}{\partial x}(Fx + u_1 + u_2^*) + \frac{\partial\lambda}{\partial y^1} H_1 x + \frac{\partial\lambda}{\partial y^2} H_2 x + \frac{\partial\lambda}{\partial\zeta}(x^2 + g_1 u_1 x + g_2 u_2^* x + \gamma_1 u_1^2 + \gamma_2 (u_2^*)^2 + 2\gamma u_1 u_2^*) \mid Y_t^1\Big)$$

$$(4.19)\qquad H_2 = E\Big(\frac{\partial\lambda}{\partial x}(Fx + u_1^* + u_2) + \frac{\partial\lambda}{\partial y^1} H_1 x + \frac{\partial\lambda}{\partial y^2} H_2 x + \frac{\partial\lambda}{\partial\zeta}(x^2 + g_1 u_1^* x + g_2 u_2 x + \gamma_1 (u_1^*)^2 + \gamma_2 u_2^2 + 2\gamma u_1^* u_2) \mid Y_\tau^2\Big)$$

Thus, taking as before $\frac{\partial\lambda}{\partial\zeta} \equiv 1$ for simplicity's sake (this can be proved, if w is regular enough, as in IV.1), we find u_1^* and u_2^* by :

$$(4.20)\qquad \begin{cases} u_1^*(t, Y_t^1, P_t) = -\frac{1}{2\gamma_1} E_{P_t}\left(\frac{\partial\lambda}{\partial x} + g_1 x + 2\gamma u_2^*(t, Y^2, P_t) \mid Y_t^1\right) \\ u_2^*(t, Y_t^2, P_t) = -\frac{1}{2\gamma_2} E_{P_t}\left(\frac{\partial\lambda}{\partial x} + g_2 x + 2\gamma u_1^*(t, Y^1, P_t) \mid Y_t^2\right) \end{cases}$$

As in the preceding example, u_1^* and u_2^* are given by a system of two integral equations, that can be interpreted, in view of the discussion of IV.1, by the fact that each player estimates (or filters) the other player's strategy through his own observations, and takes into account that the other player does the same.

This can be seen more easily by the equivalent setting :

$$(4.21)\quad \begin{cases} u_1^*(t, Y_t^1, P_t) = -\frac{1}{2\gamma_1} E\left(\frac{\partial\lambda}{\partial x} + g_1 x \,|\, Y_t^1\right) + \frac{\gamma}{2\gamma_1\gamma_2} E\left(E\left(\frac{\partial\lambda}{\partial x} + g_2 x \,|\, \mathcal{Y}_t^2\right) | Y_t^1\right) \\ \qquad\qquad + \frac{\gamma^2}{\gamma_1\gamma_2} E\left(E(u_1^*(t, Y^1, P_t) | \mathcal{Y}_t^2) | Y_t^1\right) \\ u_2^*(t, Y_t^2, P_t) = -\frac{1}{2\gamma_2} E\left(\frac{\partial\lambda}{\partial x} + g_2 x \,|\, Y_t^2\right) + \frac{\gamma}{2\gamma_1\gamma_2} E\left(E\left(\frac{\partial\lambda}{\partial x} + g_1 x \,|\, \mathcal{Y}_t^1\right) | Y_t^2\right) \\ \qquad\qquad + \frac{\gamma^2}{\gamma_1\gamma_2} E\left(E(u_2^*(t, Y^2, P_t) | \mathcal{Y}_t^1) | Y_t^2\right) \end{cases}$$

Of course, the chained conditional expectations cannot be simplified because of the non nestedness of $\mathcal{Y}_t^1$ and $\mathcal{Y}_t^2$, and u_1^*, for instance, must be computed knowing that the second player knows the estimator : $E(u_1^*(Y^1) | Y_t^2)$, and thus player one must also estimate this last estimator by : $E(E(u_1^*(Y^1) \,|\, \mathcal{Y}_t^2) | Y_t^1)$. The main qualitative result that can be extracted from (4.21) is that <u>there is no</u> need for a second guessing ; namely, the players do not have to estimate $E(\ldots E(u_1^* | \mathcal{Y}_t^2) | \mathcal{Y}_t^1) \ldots | Y_t^1)$, the estimator of the estimator of ..... of the estimator of u_1^*, after step 1 (guess what the other player has observed), since, firstly, every estimates are obtained iteratively through (4.20), and secondly since the adjoint $\frac{\partial\lambda}{\partial z}$ summarizes the complementary informations about the other player's strategy.

IV.2.2. <u>The case $\gamma = 0$</u>

Finally, let us conclude by a sharper analysis in the case $\gamma = 0$, when there is no instantaneous coupling in the cost function between the two team players. The optimal strategies are here <u>uniquely defined</u> by :

$$(4.22)\quad \begin{cases} u_1^*(t, Y_t^1, P_t) = -\frac{1}{2\gamma_1} E_{P_t}\left(\frac{\partial\lambda}{\partial x} + g_1 x \,|\, Y_t^1\right), \\ u_2^*(t, Y_t^2, P_t) = -\frac{1}{2\gamma_2} E_{P_t}\left(\frac{\partial\lambda}{\partial x} + g_2 x \,|\, Y_t^2\right), \end{cases}$$

and the value function is given by :

$$(4.23)\quad \begin{cases} E_{P_t^*}\left(\frac{\partial w}{\partial t} + \frac{1}{2}\Delta_{x,y}\lambda + \frac{\partial\lambda}{\partial x} Fx + \frac{\partial\lambda}{\partial y^1} H_1 x + \frac{\partial\lambda}{\partial y^2} H_2 x + x^2 - \frac{1}{4\gamma_1} E_{P_t^*}\left(\frac{\partial\lambda}{\partial x} + g_1 x \,|\, \mathcal{Y}_t^1\right)^2\right. \\ \qquad\qquad \left. - \frac{1}{4\gamma_2} E_{P_t^*}\left(\frac{\partial\lambda}{\partial x} + g_2 x \,|\, \mathcal{Y}_t^2\right)^2\right) = 0 \\ w(T, x, y^1, y^2, \zeta; P) = \zeta \quad \forall x, y^1, y^2, \ \forall P. \end{cases}$$

Also, P_t^* is given by :

$$\frac{d}{dt} < \varphi, P_t^* > = < L_t^{u_1^*,u_2^*} \varphi, P_t^* > \quad \forall \varphi \in C_b^2(R^4)$$

(4.24)

$$P_t^*|_{t=0} = P_o$$

with :

$$(4.25) \quad L_t^{u_1^*,u_2^*}(\varphi)(Z) = \frac{1}{2}\left(\frac{\partial^2 \varphi}{\partial x^2} + \frac{\partial^2 \varphi}{\partial (y^1)^2} + \frac{\partial^2 \varphi}{\partial (y^2)^2} \right)$$
$$+ \frac{\partial \varphi}{\partial x}\left[Fx - \frac{1}{2\gamma_1} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_1 x | \gamma_t^1 \right) - \frac{1}{2\gamma_2} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_2 x | \gamma_t^2 \right)\right]$$
$$+ \frac{\partial \varphi}{\partial y^1} H_1 x + \frac{\partial \varphi}{\partial y^2} H_2 x + \frac{\partial \varphi}{\partial \zeta}\left[x^2 - \frac{g_1 x}{2\gamma_1} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_1 x | \gamma_t^1 \right)\right.$$
$$- \frac{g_2 x}{2\gamma_2} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_2 x | \gamma_t^2 \right) + \frac{1}{4\gamma_1} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_1 x | \gamma_t^1 \right)^2$$
$$\left. + \frac{1}{4\gamma_2} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_2 x | \gamma_t^2 \right)^2 \right]$$

In fact, if we set $Y = (Y^1, Y^2)$, and :

$$(4.26)_1 \quad L_{x,y,t}^{u^*}(\varphi)(x,Y) = \frac{1}{2} \Delta_{x,y} \varphi(x, Y_t) + \frac{\partial \varphi}{\partial x}(x, Y_t)\left[Fx - \frac{1}{2\gamma_1} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_1 x | Y^1 \right)\right.$$
$$\left. - \frac{1}{2\gamma_2} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_2 x | Y^2 \right)\right] + \frac{\partial \varphi}{\partial y^1}(x, Y_t) H_1 x + \frac{\partial \varphi}{\partial y^2}(x, Y_t) H_2 x$$
$$\forall \varphi \in C_b^2(R^3),$$

$$(4.26)_2 \quad L_{\zeta,t}^{u^*}(\varphi)(Z) = \frac{\partial \varphi}{\partial \zeta}(Z_t)\left[x^2 - \frac{g_1 x}{2\gamma_1} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_1 x | Y^1 \right) - \frac{g_2 x}{2\gamma_2} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_2 x | Y^2 \right)\right.$$
$$\left. + \frac{1}{4\gamma_1} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_1 x | Y^1 \right)^2 + \frac{1}{4\gamma_2} E_{P_t^*}\left(\frac{\partial \lambda}{\partial x} + g_2 x | Y^2 \right)^2\right]$$
$$\forall \varphi \in C_b^2(R^4),$$

one can see that P_t^* can be obtained in two steps :

- Firstly, compute $P_t^1(X, Y^1, Y^2)$ by the operator $(4.26)_1$ in (4.24),
- and then take the image of P_t^1 by the integral part of the cost to obtain $P_t^*(Z) = P_t^*(X, Y^1, Y^2, \zeta)$.

But, because of the presence of $u_1^*(Y^1,P_t^*)$, $u_2^*(Y^2,P_t^*)$, and of $\frac{\partial\lambda}{\partial x}(x,y^1,y^2,P_t^*)$ in (4.24) with $(4.26)_1$, one cannot separate anymore the computation of $P_t^1(X,Y^1,Y^2)$ into a computation in X only and then in Y^1 and Y^2. An interesting but open problem would be to characterize when do we have weak interactions between the state and the observations to be able to compute P_t^1 "almost only function of X" ? Remark that if, in place of $H_i x$, i=1,2, we put $\varepsilon H_i x$ with $\varepsilon \to o$, and if the corresponding P_t^ε converges to some $\overline{P_t}$ weakly *, then, by proposition 3.6 the value function converges to $V(t,\overline{P_t})$, and the limit problem ($\varepsilon = o$) is a linear quadratic problem without information that is much easier (Pontryaguine's necessary conditions apply). More precisely, for $\varepsilon = o$, we easily obtain that $\frac{\partial w}{\partial P} \equiv o$ (no observation), and λ becomes the classical adjoint satisfying $\lambda(t) = Q(t)\,\overline{x}(t)$, where $\overline{x}(t) = E_{P_o \otimes p}(x(t))$ is at the same time the mean value of x(t) and the best estimate at time t knowing nothing ! Thus the solution can be completed via the separation principle.

Conversely, with $\varepsilon = 1$, there is no reason why, in (4.22), (u_1^*,u_2^*) would be linear in the observations, and thus, if we compute π_t^1 and π_t^2 by (4.14), these conditional measures are generally infinite dimensional filters, whereas for linear u_i, i=1,2, π_t^1 and π_t^2 are given by two Kalman filters which are, of course, finite dimensional. A more detailed discussion on the nonlinearity of (u_1^*,u_2^*) in a simpler context can be found in the next example.

VI.2.3. Sketch of a numerical method

To conclude this example, we shall sketch a numerical method giving the approximate solution of (4.22) to (4.25).

To this aim, let us first insist on the fact that P_t^* is a measure on the continuous paths (X,Y^1,Y^2,ζ) of $C_{o,t}^4$; thus the fact that u_1^* depends on Y^1 and u_2^* on Y^2, does not create additional difficulties in (4.23) and (4.24), and u_i^* is a well-defined P_t^* measurable function, i=1,2. Consequently of course, one cannot expect the operator (4.25) to be local. Nevertheless, the "non-local" element of (4.23), (4.24) can be easily isolated as we shall demonstrate.

Let us introduce the discretizations of $[o,T]$ and R^4 by a subdivision : $o < \theta < 2\theta < \ldots < M\theta = T$ of mesh $\theta = T/M$, and by a covering of R^4 by closed balls $B_\eta(z_i) = \prod_{j=1}^{4} [z_i^j - \eta^j, z_i^j + \eta^j]$ with $\eta = ((\eta^1)^2 + \ldots + (\eta^4)^2)^{1/2}$ being given, and $\{z_i\}$ a denumerable subset of R^4 made of points satisfying $\|z_{i_1} - z_{i_2}\| \geq 2\eta \quad \forall i_1 \neq i_2$, and $\forall i_1$, $\exists i_2$, $\exists j \in \{1,\ldots,4\}$ such that :

$$|z^j_{i_1} - z^j_{i_2}| = 2\eta^j .$$

Thus we have : $R^4 = \bigcup_{i \in N} B_\eta(z_i)$.

For clarity's sake, we shall omit the numerical expressions, with the finite differences method, of $\frac{\partial w}{\partial t}$, $\frac{\partial w}{\partial x}$, $\frac{\partial w}{\partial y^1}$, $\frac{\partial w}{\partial y^2}$, $\frac{\partial w}{\partial \zeta}$, $\frac{\partial^2 w}{\partial x^2}$, $\frac{\partial^2 w}{\partial (y^1)^2}$, $\frac{\partial^2 w}{\partial (y^2)^2}$, that can be found in classical textbooks of numerical analysis. But we shall focus attention on the non local tarms including $\frac{\partial \lambda}{\partial x}$ and conditional expectations.

Let us first derive an approximation of λ. We have seen that :

$$(4.27) \qquad \lambda(t,z;P) = w(t,z;P) + \int \{ \frac{\partial w}{\partial P} \}(t,\xi,z) dP_t(\xi)$$

with $\{ \frac{\partial w}{\partial P} \}$ defined by :

$$(4.28) \qquad < \frac{\partial w}{\partial P} , Q > = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} (w(t,z;P+\varepsilon Q) - w(t,z;P))$$

$$= \int \{ \frac{\partial w}{\partial P} \}(t,z,\xi) dQ_t(\xi) \qquad \forall Q \in \mathcal{M}^b_{0,t}(4)$$

Thus, if we define :

$$(4.29) \qquad A_i = \{ Z \in C^4_{0,t} \mid Z_t \in B_\eta(z_i) \} , \quad \forall i \in N,$$

And if we choose an arbitrary measure $Q \in \mathcal{M}^b_{0,t}$ (for instance the Brownian measure, or, more naturally, the incremental measure $\frac{P^*_{t+\theta} - P^*_t}{\theta}$) , noting Q^{A_i} the restriction of Q to A_i , and supposing that Q is suitably chosen to have : $Q_t^{A_i}(B_\eta(z_i)) \neq 0 \quad \forall i \in N$, then (4.28) can be approximated as follows :

$$\{ \frac{\partial w}{\partial P} \}(t,z,\xi_i) \times Q_t^{A_i}(B_\eta(\xi_i)) \sim \frac{w(t,z;P+\theta Q^{A_i}) - w(t,z;P)}{\theta} ,$$

or

$$(4.30) \qquad \{ \frac{\partial w}{\partial P} \}(t,z,\xi_i) \sim \frac{w(t,z;P+\theta Q^{A_i}) - w(t,z;P)}{\theta Q_t^{A_i}(B_\eta(\xi_i))} \qquad \forall i \in N ,$$

θ being the mesh of the time discretization.

Thus also, one can approximate :

$$(4.31)\qquad \int \{ \frac{\partial w}{\partial P} \}(t,\xi,z)dP_t(\xi) \sim \sum_{j \in N} \left(\frac{w(t,\xi_j;P+\theta Q_t^{A_i}) - w(t,\xi_j;P)}{\theta Q_t^{A_i}(B_{r_i}(z_i))}\right) P_t(B_\eta(\xi_j))$$

$$\text{for} \quad z \in B_\eta(z_i).$$

Accordingly, to derive (4.31) in x, y^1 and y^2, we denote :

$$(4.32)\qquad B_\eta^k(z_i) = \bigcup_{\{j \in N \mid\ |z_j^k - z_i^k| = 2\eta^k\}} B_\eta(z_j), \quad k=1,2,3,4, \quad i \in N,$$

and :

$$(4.33)\qquad A_i^k = \{Z \in C_{o,t}^4 \mid Z_t \in B_\eta^k(z_i)\}, \quad k \in \{1,\dots,4\}, \quad i \in N.$$

Thus we approximate :

$$(4.34)\qquad \frac{\partial}{\partial z^k} \int \{ \frac{\partial w}{\partial P} \}(t,\xi,z)dP_t(z) \sim \sum_{j \in N} \left[\frac{w(t,\xi_j;P+\theta Q_t^{A_i^k}) - w(t,\xi_j;P)}{\theta Q_t^{A_i^k}(B_{r_i}^k(z_i))} - \right.$$

$$\left. - \frac{w(t,\xi_j;P+\theta Q_t^{A_i}) - w(t;\xi_j,P)}{\theta Q_t^{A_i}(B_\eta(z_i))}\right] \times P_t(B_\eta(\xi_j)), \quad \forall k \in \{1,\dots,4\},$$

$$\text{for} \quad z \in B_\eta(z_i), \quad \text{with} \quad z^1 = x, \ z^2 = y^1, \ z^3 = y^2, \ z^4 = \zeta, \quad \forall i \in N.$$

Finally :

$$(4.35)\qquad \lambda(t,z_i;P) \sim w(t,z_i;P) + \sum_{j \in N} \left(\frac{w(t,\xi_j;P+\theta Q_t^{A_i}) - w(t,\xi_j;P)}{\theta Q_t^{A_i}(B_\eta(z_i))}\right) P_t(B_\eta(\xi_j))$$

and :

$$(4.36)\qquad \frac{\partial \lambda}{\partial z^k}(t,z_i;P) \sim \frac{w(t,z_i+\eta 1_k P) - w(t,z_i;P)}{\eta} + (4.34)$$

$$\forall k \in \{1,\dots,4\}, \quad \forall i \in N,$$

with $z_i + \eta 1_k$ being the point of coordinates z_i^j for $j \neq k$ and $z_i^k + \eta^k$ for the k^{th} coordinate, $k=1,\ldots,4$.

Also :

$$(4.37)\quad E_{P^*}\left(\frac{\partial\lambda}{\partial x} + g_1 x \middle| Y_t^1\right) = \sum_{r\in N}\left[\frac{\partial\lambda}{\partial z^1}(t,z_r;P) + g_1 z_r^1\right]P^*(A_r|Y_t^1)$$

and the same for $E\left(\frac{\partial\lambda}{\partial x} + g_2 x \middle| Y_t^2\right)$.

Thus, one can check that, knowing $w(t,z_i,P)$ for every $z_i, i \in N$, and $P = P^* + \theta Q^{A_i}$, $\forall A_i$, one can deduce $w(t-\theta,z_i,P)$ (backward in time since the boundary condition is for $t=T$). Thus, it remains to find $P^*_{t-\theta}$ such that with the optimal u_1^*, u_2^*, one obtains P_t^* at time t, and this is done as follows : find $P^*_{t-\theta}$ such that, $\forall \alpha \in N$,

$$(4.38)\quad \langle \varphi^\alpha, P_t^* \rangle - \langle \varphi^\alpha, P^*_{t-\theta} \rangle = \theta \langle L_{t-\theta}^{u_1^*,u_2^*} \varphi^\alpha, P^*_{t-\theta} \rangle,\ P_t^* \text{ given},$$

where $\{\varphi^\alpha\}_{\alpha\in N}$ is a Galerkin's basis of the space $C_U^2(R^4)$, the space of twice continuously differentiable with uniformly continuous bounded partial derivatives, and where $L_{t-\theta}^{u_1^*,u_2^*}$ is obtained by (4.25) and the approximation formulas (4.36), (4.37). The equation (4.38) is therefore transformed into a system of polynomial equations of degree ≤ 3 in the $\{P^*_{t-\theta}(A_i)\}_{i\in N}$ that canbe solved by algebraic methods.

Finally, when $w(o,z,P)$ and P_o^* are obtained, one has to choose P_T^* in order to have $P_o^* = P_o$ (two point boundary value problem).

This heuristic numerical method is justified if w and P^* are regular enough, but the main problem is the very large number of operations needed to solve these equations. Consequently, such a method might be very difficult to apply for dimensions higher than this example without a computer machine with very large memory capacities.

IV.3. Witsenhausen's counterexample

IV.3.1. Formulation in discrete-time

This well-known counterexample to the separation principle of linear-quadratic stochastic control, shows that when the observation σ-fields are not nested, the optimal control is not generally given by an affine function of the observations. Though a non-linear control, giving a lower cost than linear controls, can be exhibited, the optimum is not known until now. We shall show that the method of section II.1. applies to this problem.

We recall that the two-stage system is :

$$(4.39) \quad x_1 = x_o + u_1(x_o), \quad x_2 = x_1 - u_2(y),$$

x_o is a random variable with probability P_o and $y = x_1 + v$ where v is a random variable independent of x_o and with probability ρ. It is crucial to assume that u_2 depends on y only (non classical information), unless the problem becomes trivial.
The cost function is :

$$(4.40) \quad J(u_1,u_2) = E(k^2 u_1^2 + x_2^2), \quad k \neq o.$$

In [30] is used a discrete-time dynamic programming equation that leads to a minimization problem which has many local minima, that make the numerical methods fail.

IV.3.2. <u>The continuous-time equivalent problem</u>

We shall introduce a continuous time analogue to (4.39), (4.40) : let $t \in [o,1]$, $[o,1]$ representing the first period, and :

$$(4.41) \quad \dot{x}(t) = u_1(t,x(t)), \quad o \leqslant t \leqslant 1, \quad x(o) = x_o$$

We denote $x_1 = x(1)$, and $x_2 = x_1 - u_2(y)$.
The cost function is :

$$(4.42) \quad \tilde{J}(u_1,u_2) = E(\int_o^1 k^2 u_1^2(t,x(t))dt + x_2^2)$$

$E(x_2^2)$ is thus the final cost.

<u>Lemma</u> : If $(\tilde{u}_1,\tilde{u}_2)$ realize the minimum of $\tilde{J}$, then

$u_1^*(x_o) = \int_o^1 \tilde{u}_1(t,x(t))dt$, and $u_2^*(y) = \tilde{u}_2(y)$ realize the minimum of J,

and $J(u_1^*,u_2^*) = \tilde{J}(\tilde{u}_1,\tilde{u}_2)$.

<u>Proof</u> : We clearly have Min $\tilde{J}$ $\leqslant$ Min J since the controls in $\tilde{J}$ are richer than in J. On the other hand,

$u_1^*(x_o) = \int_o^1 \tilde{u}(t, X_{\tilde{u}}(t,x_o))dt$, with $X_{\tilde{u}}(t,x_o)$ the trajectory of (4.41) generated by $\tilde{u}$, is admissible for the two-stage problem, and satisfies :

$$J(u_1^*,u_2^*) = E(k^2 \Big| \int_o^1 \tilde{u}(t,X_{\tilde{u}}(t,x_o))dt\Big|^2 + x_2^2)$$

$$\leq E(k^2 \int_o^1 \Big| \tilde{u}(t,X_{\tilde{u}}(t,x_o))\Big|^2 dt + x_2^2) \quad \text{(by Cauchy-Schwartz)}$$

$$= \tilde{J}(\tilde{u}_1,\tilde{u}_2) \quad \text{and thus both problems are equivalent.}$$

Remark : The second part of the proof has been kindly communicated to me by Prof. T. Basar.
It was noted in [31] by a simple dynamic programming argument that :

$$(4.43) \qquad u_2^*(y,P_1^u) = E(x_1|y) = \int \frac{x\rho(y-x)}{\int \rho(y-\xi)dP_1^u(\xi)} dP_1^u(x)$$

with $P_1^u = X_{u_1}(1,P_o)$.
Thus, the final cost is :

$$E(x_2^2) = E((x_1 - u_2^*(y))^2) = \int x(x - \int (\frac{\int \xi\rho(y-\xi)dP_1^u(\xi)}{\int \rho(y-\zeta)dP_1^u(\zeta)})\rho(y-x)dy)dP_1^u(x)$$

and we call :

$$(4.41) \qquad w(1,x;1,P_1) = x(x - \int E(x_1|y)\rho(y-x)dy).$$

The problem is thus to determine $w(t,x;t,P_t) \quad \forall t < 1$.
Applying the corollary 2.4, we minimize the Hamiltonian :

$$(4.45) \qquad H = E(\frac{\partial\lambda}{\partial x} u_1 + k^2 u_1^2 | x) = \frac{\partial\lambda}{\partial x}(t,x;P_t)u_1(t,x) + k^2 u_1^2(t,x).$$

yielding :

$$(4.46) \qquad u_1^*(t,x,P_t) = -\frac{1}{2k^2} \frac{\partial\lambda}{\partial x}(t,x;P_t)$$

with :

$$(4.47) \qquad \frac{\partial\lambda}{\partial x}(t,x;P_t) = \frac{\partial}{\partial x}(w(t,x;t,P_t) + \int \{\frac{\partial w}{\partial P}\}(t,\xi;x)dP_t(\xi))$$

and (2.96) becomes :

$$(4.48) \qquad \int \left(\frac{\partial w}{\partial t} - \frac{1}{4k^2} \left(\frac{\partial \lambda}{\partial x} (t,x;P_t) \right)^2 \right) dP_t(x) = 0$$

with the boundary condition (4.44).

The probability P_t is finally given by its density :

$$(4.49) \qquad \begin{cases} \dfrac{\partial \pi}{\partial t}(t,x) - \dfrac{1}{2k^2} \dfrac{\partial}{\partial x} \left(\pi(t,x) \dfrac{\partial \lambda}{\partial x} (t,x;P_t) \right) = 0 \\ \pi(0,x) = \pi_0(x) \end{cases}$$

and the solution of (4.47), (4.48), (4.49), with boundary condition (4.44), give finally u_1^*.

This set of equations canbe solved numerically with the same techniques as in IV.2.3, and it is easy to obtain an approximation of u_1^* on a small interval $]1-\varepsilon,1]$ by replacing $\frac{\partial w}{\partial t}(1,x;1,P_1)$ by $\frac{1}{\varepsilon}(w(1,x;1,P_1) - w(1-\varepsilon,x;1,P_1))$.

Thus, (4.48), (4.47) become :

$$(4.50) \qquad w_\varepsilon(1-\varepsilon,x;1,P_1) = w(1,x;1,P_1) - \frac{\varepsilon}{4k^2} \left(\frac{\partial \lambda}{\partial x} (1,x;P_1) \right)^2$$

But with (4.44) :

$$(4.51) \qquad \frac{\partial \lambda}{\partial x}(1,x;P_1) = 2\Big(x - \int \hat{x}_1(y)\rho(y-x)dy + x \int \hat{x}_1(y) \frac{\partial \rho}{\partial y}(y-x)dy - \frac{1}{2}\int (\hat{x}_1(y))^2 \frac{\partial \rho}{\partial y}(y-x)dy\Big)$$

with $\hat{x}_1(y) = E(x_1|y)$ given by (4.43).

Finally, we have the desired approximation $u_\varepsilon^*(1-\varepsilon,x,P_1)$ on $]1-\varepsilon,1]$ by (4.46) and (4.51). An approximation on $[0,1]$ should require the use of the numerical method of IV.2.3.

If ρ satisfies $\lim_{|x| \to \infty} \rho(x) = 0$, then (4.51) becomes, after integration by parts :

$$u_\varepsilon(1-\varepsilon,x,P_{1-\varepsilon}) = \frac{1}{k^2}\Big(x - \int \hat{x}_1(x+y)\rho(y)dy - x\int \frac{\partial}{\partial y}(\hat{x}_1(x+y))\rho(y)dy + \frac{1}{2}\int \frac{\partial}{\partial y}[(\hat{x}_1(x+y))^2]\rho(y)dy\Big)$$

or :

$$(4.52)\quad u_\varepsilon(1-\varepsilon,x,P_{1-\varepsilon}) = -\frac{1}{k^2}\,(x - (\hat{x}_1 * \rho)(x) + \frac{1}{2}((\frac{\partial}{\partial y}\,(x-\hat{x}_1(y))^2) * \rho)(x)$$

where $(\varphi * \psi)(x)$ is the convolution between φ and ψ at the point x. u_ε can thus be interpreted as follows : the first part of u_ε corresponds to $-\frac{1}{k^2}\,(x - (\hat{x}_1 * \rho)(x))$ which is the distance between the real x known before time 1, and the best estimate after time 1, re-estimated knowing the observation x before time 1, but without knowing what is the observed y. Shortly, $(x - (\hat{x}_1 * \rho)(x))$ is the <u>error of estimation between x and the second guessing of x</u> knowing that $u_2^* = \hat{x}_1(y)$.
The second term : $-\frac{1}{2k^2}\,(\frac{\partial}{\partial y}\,(x - \hat{x}_1)^2 * \rho)(x)$, corresponds to the guessing before time 1 of the first order expansion of the variance of error, that can be interpreted as a <u>correction term</u> taking into account the fact that the variance of error could vary with a small variation of control or of probability.

To conclude, we see that the nonlinearity of u_ε is a consequence of the second guessing and of the estimated correction on the variance of error. This completes the picture already drawn by H.S. Witsenhausen, since in [31], no characterization in terms of second guessing was obtained. Finally, the most important consequence of our method lies in the fact that it avoids the problem of many local minima of the discrete-time version, by embedding it into a continuous-time version where more precise characterizations of the optimal strategy are available, namely : equations (4.47) to (4.49), and for which a numerical solution canbe obtained by the method of IV.2.3.

Furthermore, we know by the Corollary 2.3 that if w is obtained (in a suitable function space), then the value function is <u>uniquely</u> defined through w, and thus <u>there is no risk to find a local minimum</u> by this method.

V - CONCLUDING REMARKS

In the preceding sections, we have shown five types of results :

1) The dynamic programming method can be applied to non - classical information team and control problems. Although this fact was implicitly noted in [31] and [32], no fruitful results were obtained until now.

2) This method gives efficient characterizations of the optimal strategies when combined with the integral representation of the value function.

3) Non - classical team information problems do have an Hamiltonian structure (this fact is suggested in a particular case in [25]).
A natural question from now is whether adjoint equations can be derived or not. The answer is yes, at least in particular cases, as shown in the classical information case, where, moreover, the derivation of the Hamilton-Jacobi-Bellman equation is rigorous.

4) The method developed suits as well diffusions as non diffusions models and generalizes the techniques of control of partially observed diffusions ([1], [2],[6],[7],[11],[14],[15],[24]).

5) The optimality conditions allow a qualitative but precise description of phenomena like signalling, dual effect, second guessing, that were until now, at least in this context, introduced heuristically and imprecisely by many authors ([10],[17],[18],[19],[30],[31],[33]). Furthermore, numerical methods can be used to integrate the Hamilton - Jacobi - Bellman equations, and thus to obtain the value function and the optimal strategy.

Acknowledgement : The author is indebted to Professors P.L. Lions and A. Bensoussan for helpful suggestions on stochastic differential equations techniques, and for their kind encouragements.

REFERENCES

[1] R. ANDERSON, A. FRIEDMAN : Multi-dimensional quality control problems and quasi-variational inequalities. TAMS, Vol. 246 (1978) p. 31-76.

[2] R. ANDERSON, A. FRIEDMAN : Quality control for Markov chains and free boundary problems. TAMS, Vol. 246 (1978) p. 77-94.

[3] J.S. BARAS : Non commutative probability models in quantum communication and multi-agent stochastic control. Ricerce di Automatica, 10, 2 (1979) p. 217-264.

[4] T. BASAR : An equilibrium theory for multi-person decision making with multiple probabilistic models.I. Symmetric mode of decision making. Preprint. University of Illinois (1982).

[5] V. BENES : Existence of optimal stochastic control laws. SIAM J. Cont, Vol. 9, No 3, (1971).

[6] A. BENSOUSSAN : Maximum principle and dynamic programming approaches of the optimal control of partially observed diffusions. Stochastics, 9, 3, (1983), p. 169-222.

[7] J.M. BISMUT : Sur un problème de contrôle stochastique avec observation partielle. Z.F. Wahr, V. Geb. 49, (1979), p. 63-95.

[8] C. CASTAING, M. VALADIER : Convex analysis and measurable multifunctions. Lecture Notes in Maths n° 580, Springer (1977).

[9] N. CHRISTOPEIT : Existence of optimal stochastic controls under partial observation. Z.F. Wahr. V. Geb. 51 (1980) p. 201-213.

[10] K.C. CHU : Team decision theory and information structures in optimal control problems II. IEEE- AC-17, 1 (1972) p.22-28.

[11] M.H.A. DAVIS : Nonlinear semigroups in the control of partially observed stochastic systems. In measure theory and applications to stochastic analysis, Kallianpur - Kolzow editors, lecture Notes in Maths - Springer (1979).

[12] R.J. ELLIOTT, M. KOHLMANN : On the existence of optimal partially observed controls. Appl. Math. Optimiz. 9 (1982), p. 41-66.

[13] A.F. FILIPPOV : Differential equations with discontinuous right-hand side. Dokl. Akad. Nauk. SSSR. 151 (1963) p. 65-68.

[14] W.H. FLEMING : Nonlinear semigroup for controlled partially observed diffusions. To appear.

[15] W.H. FLEMING, E. PARDOUX : Existence of optimal controls for partially observed diffusions. SIAM J. Cont. Vol. 20 (1982) p. 261-285.

[16] I.I. GIHMAN, A.V. SKOROHOD : Stochastic differential equations. Springer 1972.

[17] Y.C. HO, K.C. CHU : Team decision theory and information structures in optimal control problems. IEEE - AC-17, 1 (1972).

[18] Y.C. HO, M.P. KASTNER, E. WONG : Teams, market signalling, and information theory. IEEE - AC-23 (1978), p.305-311.

[19] Y.C.HO : Teams decision theory and information structures. Proc. IEEE, Vol. 68, 6 (1980) p. 644-654.

[20] N.KRASSOVSKI, A. SOURBOTINE : Jeux différentiels. in French. MIR (1977)

[21] J. LEVINE : Principe d'optimalité et principe de séparation en contrôle stochastique à information incomplète non classique CRAS. t. 292, I, (1981) p. 877-880.

[22] J. LEVINE : Incimplete information in differential games and team problems. 8th IFAC World Congress, Kyoto (1981).

[23] R.S. LIPTSER, A.N. SHIRYAYEV : Statistics of random processes, I, Springer, 1977, (English translation).

[24] R.E. MORTENSEN : Stochastic optimal control with noisy observations, Int. J. Control, 4 (1966), p. 455-464.

[25] J.P. QUADRAT : Sur l'identification et le contrôle optimal des systèmes stochastiques. Thesis. Paris 9. (1981).

[26] T. ROCKAFELLAR : Conjugate duality and optimization. SIAM Regional Conf. Series in Applied Math n° 16. (1974).

[27] L. SCHWARTZ : Théorie des distributions. Hermann. Paris (1966).

[28] R. SENTIS : Equations différentielles à second membre mesurable. CRAS. Série A. 184, (1977), p. 113-116.

[29] D.W. STROOCK, S.R.S. VARADHAN : Multidimensional diffusion processes, Springer (1979).

[30] P. VARAIYA, J. WALRAND : On delayed sharing patterns. IEEE.AC - 23, n° 3 (1978), p. 443-445.

[31] H.S. WITSENHAUSEN : A counterexample in stochastic optimum control. SIAM J. Cont., 6, n°1, (1968), p.131-147.

[32] H.S. WITSENHAUSEN : A standard form for sequential stochastic control. J. Math. Syst. Theory, 7, n°1 (1973), p. 5-11.

[33] T. YOSHIKAWA, H. KOBAYASHI : Separation of estimation and control for decentralized stochastic control systems. 7th IFAC World Congress, Helsinki (1978).

[34] J.M. BISMUT : Partially observed diffusions and their control. SIAM J. Cont. & Optimiz. 20, 2 (1982), p. 302-309.

[35] V. BENES : Optimal stopping under partial observations. In advances in filtering and optimal stochastic control. Fleming - Gorostiza ed. Lecture Notes in Control and Inf. Sciences, 42, (1982), Springer.

APPENDIX

A.I. KRASSOVSKI'S CONCEPT OF SOLUTION FOR OUTPUT-FEEDBACK STRATEGIES

The system and the assumptions are the same as in section II.

Let $u \in \mathcal{U}$. The problem consists in defining solutions to the differential system with measurable right-hand-side :

$$(A.1)\quad \begin{cases} \dot{x}(t) = f(t,x(t),u(t,Y(t)),v_j) \\ \qquad\qquad\qquad\qquad \forall t \in [t_j,t_{j+1}[\cap [s,T] \quad \forall j=0,\dots,M-1. \\ y(t) = h(t,x(t),v_j) \\ x(t) = \xi_{0,s}(t) \qquad \forall t \in [0,s] \end{cases}$$

For simplicity's sake, we state the results in the case $N = 1$. Their generalization to arbitrary N is trivial.

The results presented here follow the same lines as those of [20] in a slightly more general context : in [20], (A.1) is studied for purely instantaneous state feedbacks of the form $u(t,x(t))$, whereas here, the right-hand-side of (A.1) depends on the whole past history of the observations between $x(t)$ and t through the dependence on u.

Let us be given a family of subdivisions of $[s,T]$ given by :

$$(A.2)\quad \begin{aligned} &\tau_0^m = s < \dots < \tau_{L_m}^m = T \\ &\text{with : } \lim_{m\to\infty} \Big(\sup_{j=0,\dots,L_m-1} (\tau_{j+1}^m - \tau_j^m) \Big) = 0 . \end{aligned}$$

Let also $\xi_{0,s} \in C^0([0,s]; R^n)$, and the sequence $\{\xi_{0,s}^m\}_m$ satisfy :

$$(A.3)\quad \lim_{m\to\infty} \|\xi_{0,s}^m - \xi_{0,s}\| \overset{\text{def}}{=} \lim_{m\to\infty} \Big(\sup_{0\le t\le s} \|\xi_{0,s}^m(t) - \xi_{0,s}(t)\| \Big) = 0$$

Let us also define :

$$(A.4)\quad u^m(Y(t)) = u(\tau_j^m, Y(\tau_j^m)) \quad \text{if} \quad t \in [t_j^m, \tau_{j+1}^m[\quad \forall j=0,\dots,L_m-1,$$

and :

$$(A.5)\quad \begin{cases} \dot{x}^m(t) = f(t,x^m(t),\ u^m(Y^m(t)),v_j) \\ y^m(t) = h(t,x^m(t),\nu_j) \\ x^m(t) = \xi^m_{o,s}(t) \quad \forall t \in [o,s] \end{cases} \qquad \begin{array}{l} \forall t \in [\tau^m_r,\tau^m_{r+1}[\cap [t_j,t_{j+1}[, \\ \forall r = o,\dots,L_m-1, \quad \forall j = o,\dots,M-1. \end{array}$$

we shall say "K-solution" for "solution in the sense of Krassovski".

Definition A.1 : We say that x is a K-solution of (A.1) on $[s,T]$ if there exists a sequence of subdivisions of $[s,T]$ satisfying (A.2), a sequence $\{\xi^m_{o,s}\}$ satisfying (A.3) and a subsequence $\{x^{m_r}\}$ of solutions of (A.5), with u^{m_r} defined by (A.4), such that :

$$(A.6)\qquad \lim_{r\to\infty} \|x^{m_r} - x\|_{C_{s,T}} = o \qquad \blacksquare$$

Theorem A.1 : If the assumptions of II.1.1 hold, there exists at least one K-solution to (A.1), and every K-solution is an absolutely continous solution on $[s,T]$ of :

$$(A.7)\quad \begin{cases} \dot{x}(t) \in \overline{co} \bigcap_{V\in\mathcal{V}(x_{o,t})} f(t,x(t),u(t,\mathcal{H}_\nu(V)),v_j) & \forall t \in [t_j,t_{j+1}[, \ \forall j = o,\dots,M-1. \\ \mathcal{H}_\nu(x_{o,t}) = \{y(\sigma) = h(\sigma,x(\sigma),\nu_r) \mid \sigma \in [t_r,t_{r+1}[\cap [\mu(t),t] \cap [s,T], \ \forall r = o,\dots,M-1\} \\ x(\sigma) = \xi_{o,s}(\sigma) \quad \forall \sigma \in [o,s] \end{cases}$$

where $\mathcal{V}(x_{o,t})$ is a basis of neighborhoods of $x_{o,t}$ for the uniform topology of $C^o([o,t];\ R^n)$.

Proof : Let $\{\tau^m_r\}$, $\{\xi^m_{o,s}\}$ be given as in (A.2), (A.3). Then, by classical results on differential equations, since the right-hand-side is smooth in x and measurable in t, x^m is uniquely defined $\forall m \geq o$, and by (2.2), x^m is defined on $[s,T]$.

Furthermore, using once more (2.2) and classical inequalities, it is straightforward to prove that the sequence $\{x^m\}$ is equicontinuous and equibounded for each $\xi_{o,s}$ satisfying $\|\xi_{o,s}\|_{C_{o,s}} < +\infty$. Thus, by Ascoli's

theorem, one can extract a subsequence converging uniformly to $x \in C^o([o,T]; R^n)$. Also, if we note $\{x^{m_r}\}$ the subsequence, we have :

$\{\dot{x}^{m_r}\}$ converges to Z in $L^2([s,T]; R^n)$ weakly, and
$\dot{x} = Z$ in the sense of distributions.

Furthermore, by Mazur's theorem one can extract another subsequence, still noted $\{x^{m_r}\}$ such that $\{\sum_{m_r \geq k} a_{m_r k} \dot{x}^{m_r}\}$ converges to Z in $L^2([s,T]; R^n)$ strongly.
Thus $\dot{x} = Z \in L^2([s,T]; R^n)$ and x is absolutely continuous on [s,T], and using the upper semi continuity of $x_{o,t} \to \Phi(x_{o,t}) = \overline{co} \bigcap_{V \in \mathcal{V}(x_{o,t})} f(t,x(t),u(t,\mathcal{H}_\nu(V),v_j)$ (see [8]), it is straightforward that $x(t) \in \Phi(x_{o,t})$ and (A.7) is established. ∎

Remark : If x is a K-solution, x does not generally satisfy (A.1), even for almost every $t \in [s,T]$ (see [20]). However, K-solutions are the most intuitive notion of a solution that satisfy the basic property needed for control theory, namely semi-group property.

On the other hand, every solution of (A.7) is not generally a K-solution (see [20]). ∎

Definition A.2 : We say that x is an F-solution (or solution in the sense of Filippov (see [13])) of the system (A.1) if x is absolutely continuous on [s,T] and satisfies (A.7). ∎

Proposition A.1 : if $K^u(s,\xi_{o,s})$ denotes the set of every K-solution on [s,T], starting from $(s,\xi_{o,s})$ and generated by u, then $K^u(s,\xi_{o,s})$ is compact in $C_{o,T} = C^o([o,T]; R^n)$ for the uniform topology, and the multivalued mapping $\xi_{o,s} \to K^u(s,\xi_{o,s})$ from $C_{o,s}$ to the compact subsets of $C_{o,T}$, is upper semi-continuous. Furthermore, the two following properties hold $\forall \xi_{o,s} \in C_{o,s}$:

(i) if $x^u(s,\xi_{o,s}) \in K^u(s,\xi_{o,s})$ then $x^u_{s,\tau}(s,\xi_{o,s})$, restriction of x^u on $[s,\tau]$, is a K-solution on $[s,\tau]$, $\forall \tau \in [s,T]$.

(ii) if $x^u_{s,\tau}(s,\xi_{o,s}) \in K^u_{s,\tau}(s,\xi_{o,s})$ and :

$$x^u_{\tau,t}(\tau, x^u_{s,\tau}(s,\xi_{o,s})) \in K^u_{\tau,t}(\tau, x^u_{s,\tau}(s,\xi_{o,s}))$$

for $s \leq \tau \leq t \leq T$, with $K^u_{s,\tau}$ the set of K-solutions restricted to $[s,\tau]$, then :

$$\exists\, x^u_{s,t}(s,\xi_{o,s}) \in K^u_{s,t}(s,\xi_{o,s}) \text{ such that :}$$

$$x^u_{s,t}(s,\xi_{o,s}) = x^u_{\tau,t}(\tau,x^u_{s,\tau}(s,\xi_{o,s})).$$

Proof : compactness and upper semicontinuity of $K^u(s,\xi_{o,s})$ follow the same argument as in [20], lemmas 7.2 and 7.3 page 33, and their adaptation to our formalism is straightforward.

For property (i), if $\{u^m\}$ is a sequence converging to $x^u(s,\xi_{o,s}) \in K^u(s,\xi_{o,s})$ and satisfying (A.5), it suffices to restrict $\{x^m|_{[s,\tau]}\}$ to obtain that $x^u|_{[s,\tau]}$ is also a K-solution on $[s,\tau]$.

Finally, for property (ii), if $x^u_{s,\tau} \in K^u_{s,\tau}(s,\xi_{o,s})$, and if $\{x^m_{s,\tau}\}$ is the associated sequence, if $x^u_{\tau,t}(\tau,x^u_{s,\tau}(s,\xi_{o,s})) \in K^u_{\tau,t}(\tau,x^u_{s,\tau}(s,\xi_{o,s}))$ and if $\{x^{m'}_{\tau,t}(\tau,x^u_{s,\tau}(s,\xi_{o,s}))\}$ is the associated converging sequence, it is not difficult to prove, using the upper semi continuity of K^u, that the sequence $\{x^{m'}_{\tau,t}(\tau,x^m_{s,\tau}(s,\xi_{o,s}))\}$ contains a subsequence converging to $x^u_{\tau,t}(\tau,x^u_{s,\tau}(s,\xi_{o,s}))$ and thus the sequence $\{x^{m'}_{\tau,t}(\tau,x^m_{s,\tau}(s,\xi_{o,s}))\}$ contains a converging subsequence to an element of $K^u_{s,t}(s,\xi_{o,s})$ which achieves the proof. ∎

Proposition A.2 : Let us denote $F^{u,\omega}(s,\xi_{o,s})$ the set of every F-solution of (A.1) starting from $(s,\xi_{o,s})$ and generated by (u,ω), and let us introduce $F^u(s)$, the set of progressively measurable flows which are F-solutions of (A.1), namely, satisfying (i) and (ii) :

(i) $(\omega,\xi_{o,s}) \to x^{u,\omega}(s,\xi_{o,s})$ is $\{\mathcal{F}_{o,s} \times \mathcal{A}_t | t \in [s,T[\}$ progressively measurable

(ii) $x^{u,\omega}(s,\xi_{o,s}) \in F^{u,\omega}(s,\xi_{o,s}) \quad \forall(\omega,\xi_{o,s}) \in \Omega \times C_{o,s}$.

Then the measures image of $P_{o,s}$ by Filippov's flows are contained in $\overline{co}\ \wp^u_T(s,P_{o,s})$, namely :

$$(A.8) \qquad \{P^u_T = x^u_T(s,P_{o,s}) | x^u \in F^u(s)\} \subset \overline{co}\ \wp^u_T(s,P_{o,s})$$

with the notations of II.2.1 : $\int \varphi d(x^u_T(s,P_{o,s})) = \int \varphi(x^{u,\omega}_T(s,\xi_{o,s}))dP_{o,s}(\xi_{o,s})d\rho(\omega)$ $\forall \varphi \in C^o_b(R^n)$.

Proof : Let $x^u \in F^u(s)$. Then, by definition :

$$(A.9) \qquad \dot{x}^{u,\omega}_t(s,\xi_{o,s}) \in \overline{co} \bigcap_{V \in \mathcal{V}(x^{u,\omega}_{s,t}(s,\xi_{o,s}))} f(t,x^{u,\omega}_t(s,\xi_{o,s}),u(\mathcal{H}_\nu(V)),v_j).$$

Applying Caratheodory's lemma, there exists n+1 positive numbers :

$$(A.10) \qquad \alpha_i(t,x^{u,\omega}_{s,t}(s,\xi_{o,s})) \geq o \ , \quad \sum_{i=1}^{n+1} \alpha_i(t,x^{u,\omega}_{s,t}(s,\xi_{o,s})) = 1$$

and n+1 elements of $\bigcap_{V\in\mathcal{V}(x^{u,\omega}_{s,t}(s,\xi_{o,s}))} f(t,x^{u,\omega}_t(s,\xi_{o,s}),u(\mathcal{H}_\nu(V)),v_j)$

denoted $f(t,x^{u,\omega}_t(s,\xi_{o,s}),\ u_i(\mathcal{H}_\nu(x^{u,\omega}_{s,t}(s,\xi_{o,s})),v_j),\quad i=1,\dots,n+1,$

with $\quad u_i(\mathcal{H}_\nu(x^{u,\omega}_{s,t}(s,\xi_{o,s}))) \in \bigcap_{V\in\mathcal{V}(x^{u,\omega}_{s,t}(s,\xi_{o,s}))} u(\mathcal{H}_\nu(V)),$

since f is continuous with respect to all his arguments. u_i is thus an accumulation point of u at $\mathcal{H}_\nu(x^{u,\omega}_{s,t}(s,\xi_{o,s}))$, i=1,...,n+1.

Furthermore, the α_i's can be chosen progressively measurable (see [8]).
Now, if we call $P^u_t = x^u_t(s,P_{o,s})$, we have :

$$(A.11) \quad \frac{d}{dt}\int\varphi(x)dP^u_t(x) = \frac{d}{dt}\int\varphi(x^{u,\omega}_t(s,\xi_{o,s}))dP_{o,s}(\xi_{o,s})d\rho(\omega)$$

$$= \sum_{i=1}^{n+1}\int \frac{\partial\varphi}{\partial x}(x^{u,\omega}_t(s,\xi_{o,s}))(\alpha_i(t,x^{u,\omega}_{s,t}(s,\xi_{o,s}))f(t,x^{u,\omega}_t(s,\xi_{o,s}),$$

$$u_i(\mathcal{H}_\nu(x^{u,\omega}_{s,t}(s,\xi_{o,s}))),v_j))dP_{o,s}(\xi_{o,s})d\rho(\omega).$$

But if we denote $dP^{i,u}_{o,t} = \alpha_i dP^u_{o,t}$, i=1,...,n+1, we have :

$$(A.12) \quad \frac{d}{dt}\int\varphi(x)dP^u_t(x) = \sum_{i=1}^{n+1}\int \frac{\partial\varphi}{\partial x}(x)\cdot f(t,x,u_i(\mathcal{H}_\nu(x_{o,t})),v_j)dP^{i,u}_{o,t}(x_{o,t})d\rho(\omega),$$

$$\forall\varphi \in C^1_b(R^n),$$

and, defining $x^{i,u,\omega}(s,\xi_{o,s}) \in K^{u,\omega}(s,\xi_{o,s})$ by :

$$\dot{x}^{i,u,\omega}_t(x,\xi_{o,s}) = f(t,x^{i,u,\omega}_t(s,\xi_{o,s}),u_i(\mathcal{H}_\nu(x^{i,u,\omega}_{s,t}(s,\xi_{o,s}))),v_j)$$

with initial condition $\xi_{o,s}$, for i=1,...,n+1, it follows that :

$$(A.13) \quad \frac{d}{dt}\int\varphi(x)dP^u_t(x) = \sum_{i=1}^{n+1}\frac{d}{d\tau}\int\varphi(x)d(x^{i,u}_\tau(t,P^{i,u}_{o,t}))(x)\Big|_{\tau=t} \qquad \forall t\in[s,T].$$

In particular for $t = s$,

$$(A.14) \qquad \frac{d}{dt}\int\varphi(x)dP_t^u(x)\Big|_{t=s} = \sum_{i=1}^{n+1}\frac{d}{dt}\int\varphi(x)d(x_t^{i,u}(s,\alpha_i P_{o,s}))(x)\Big|_{t=s}$$

But, this means that on a small interval $[s,s+\varepsilon[$, we have :

$$(A.15) \qquad P_t^u = \sum_{i=1}^{n+1}\alpha_{i,t}x_t^{i,u}(s,P_{o,s})$$

and, integrating (A.13), we find that (A.15) holds $\forall t \in [s,T]$ and thus $P_T^u = \sum_{i=1}^{n+1}\alpha_{i,T}x_T^{i,u}(s,P_{o,s})$. Since $x^{i,u} \in K^u(s)$, then $P_T^u \in \overline{co}\ \mathcal{P}_T^u(s,P_{o,s})$ and the result is proved. ∎

A.II. Krassovski's concept of solution for closed-loop strategies

As in section II.2.3, we have to define solutions to (A.1) where u depends on the measure $x_{s,t}^u(s,P_{o,s})$ at time t, and on Y(t) as before. For this purpose, let us take a subdivision as in (A.2), a sequence $\{\xi_{o,s}^m\}$ as in (A.3), and a sequence $\{\pi_{o,s}^m\}$ satisfying :

$$(A.16) \qquad \lim_{m\to\infty}\pi_{o,s}^m = P_{o,s} \quad \text{in the weak * topology of } \mathcal{M}_{o,s}^b.$$

We define :

$$(A.17) \quad u^m(Y(t),P_{o,t}) = u(\tau_j^m,Y(\tau_j^m),P_{o,\tau_j^m}) \quad \text{if} \quad t \in [\tau_j^m,\tau_{j+1}^m[, \quad \forall j=o,\dots,L_m-1,$$

for every Y and $P_{o,t}$, and we build the sequence $\{x^m\}$ as follows :

$$(A.18) \quad \begin{cases} \dot{x}_t^{m,u,\omega}(s,\xi_{o,s}^m) = f(t,x_t^{m,u,\omega}(s,\xi_{o,s}^m),u^m(\mathcal{H}_\nu(x_{s,t}^{m,u,\omega}(s,\xi_{o,s}^m)),P_{o,t}^m),v_j) \\ \qquad\qquad \forall t \in [\tau_r^m,\tau_{r+1}^m[\cap [t_j,t_{j+1}[, \quad \forall r=o,\dots,L_m-1, \quad \forall j=o,\dots,M-1, \\ \mathcal{H}_\nu(x_{s,t}^{m,u,\omega}(s,\xi_{o,s}^m)) = \{y(\sigma) = h(\sigma,x_{s,t}^{m,u,\omega}(s,\xi_{o,s}^m),v_j)\Big| \\ \qquad\qquad \forall\sigma \in [t_j,t_{j+1}[\cap [u(t),t] \cap [s,T], \quad \forall j=o,\dots,M-1\} \\ P_{o,t}^m = x_{s,t}^{m,u}(s,\pi_{o,s}^m) \end{cases}$$

Definition A.3 : We say that (x^u,P^u) is a K-solution of (A.1) for the closed-loop strategy $u \in \tilde{\mathcal{U}}$, if there exist a sequence of subdivisions of $[s,T]$ satisfying (A.2), a sequence $\{\xi^m_{o,s}\}$ satisfying (A.3), a sequence $\{\pi^m_{o,s}\}$ satisfying (A.16), and a subsequence $\{x^{m_r,u},P^{m_r}\}$ satisfying (A.18) with u^{m_r} defined by (A.17), such that :

$$\text{(A.19)} \qquad \lim_{r\to\infty} \|x^{m_r,u} - x^u\|_{C_{s,T}} = 0 \quad \text{and} \quad \lim_{r\to\infty} P^{m_r} = P^u \quad \text{(weakly * in } \mathcal{M}^b_{o,T}). \qquad \blacksquare$$

Theorem A.2 : under the assumptions of II.2.3., there exists at least one K-solution (x^u,P^u), and every K-solution (x^u,P^u) satisfies :

(i) $x^{u,\omega}(s,\xi_{o,s},P_{o,s})$ is an absolutely continuous solution on $[s,T]$ of :

$$\text{(A.20)} \quad \dot{x}^{u,\omega}_t(s,\xi_{o,s},P_{o,s}) \in \overline{\text{co}} \bigcap_{V\in\mathcal{V}(x^{u,\omega}_{s,t}(s,\xi_{o,s},P_{o,s}))} \; \bigcap_{\beta\in\mathcal{B}(P^u_{o,t})} f(t,x^{u,\omega}_t(s,\xi_{o,s}P_{o,s}), \, ,u(\mathcal{H}_\nu(V),\beta),v_j)$$

$\forall t \in [t_j,t_{j+1}[$, $\forall j=o,\dots,M-1$, with $\mathcal{H}_\nu$ defined as in (A.18), and with $\beta(P^u_{o,t})$ a denumerable filter converging to $P^u_{o,t}$ in $\mathcal{M}^b_{o,t}$ weak *.

(ii) $P^u = x^u(s,P_{o,s})$

Proof : Let $\{\tau^m_r\}$, $\{\xi^m_{o,s}\}$ and $\{\pi^m_{o,s}\}$ be given as in the definition A.3. We shall proceed by induction on r with fixed m to prove that the sequence $\{x^m,P^m\}$ is uniquely defined :

on $[\tau^m_o,\tau^m_1[$, one can define $x^{m,u,\omega}(s,\xi^m_{o,s},\pi^m_{o,s})$ in a unique way by classical arguments on differential equations, and the application : $(\omega,\xi_{o,s}) \to x^{m,u,\omega}(s,\xi_{o,s},\pi^m_{o,s})$ is measurable. Thus there is no difficulty to define in a unique way the measure $P^m_{o,\tau^m_1} = x^{m,u}(s,\pi^m_{o,s})$. Suppose that this property holds up to the order r, then for r+1, the same argument as at the order o proves that $(x^{m,u},P^m)$ is uniquely defined $\forall m \geq o$.

By the same argument as in theorem A.1, the sequence $\{x^{m,u}\}$ is equicontinuous in $C^o([s,T];R^n)$, equibounded for every initial $\xi_{o,s}$ lying in a bounded subset of $C_{o,s}$, and satisfying :

$$\text{(A.21)} \qquad \sup_m \; E\left(\left\| x^{m,u,\omega}_{s,T}(s,\xi^m_{o,s},\pi^m_{o,s})\right\|^2_{C_{o,t}}\right) < +\infty$$

(by Gronwall's inequality).

Thus, if we take a family of bounded subsets of $C_{o,s}$ whose union is $C_{o,s}$ (for example $\{\|\xi_{o,s}\|_{C_{o,s}} \leq r\}_{r\in\mathbb{N}}$), we have that on each bounded subset B_r there exists a converging subsequence $\{x^{m_r,u}\}$ to $x^u_{B_r}$, and the corresponding subsequence P^{m_r}, with the inequality (A.21) converges to P^u in the weak * topology, P^u satisfying :

$$x^u_{B_r}(s,P_{o,s}) = P^u\big|_{B^u_r} \quad \text{with} \quad B^u_r = x^u(s,B_r).$$

Finally, if $B_r \subset B_{r'}$, we have $x^u_{B_{r'}}|_{B_r} = x^u_{B_r}$ and thus by a projective limit argument x^u is uniquely defined on $C_{o,T}$ and $P^u = x^u(s,P_{o,s})$, which proves the existence of a K-solution. The proof of (ii) follows exactly the same lines as the preceding argument. Finally (i) is proved exactly as in theorem A.1. ∎

Finally, the proposition A.1, adapted in the closed-loop context, holds true and its proof, following the same lines as proposition A1's proof, is left to the reader. ∎

A.III. Existence of solutions to the problem of martingale for closed-loop strategies.

Theorem A.3 : $\mathcal{P}^u(t;s,P_s) \neq \emptyset \quad \forall u \in \tilde{\mathcal{U}}, \quad \forall t \geq s, \quad \forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$.

Furthermore, for each $P^u_{s,P_s} \in \mathcal{P}^u(T;s,P_s)$, there exists a $\mathcal{B}_{[o,t]} \times \mathcal{F}_t \times \mathcal{Y}_t \times \mathcal{B}^t_{\mathcal{M}}$ measurable selection $\tilde{F}_o$ of the multifunction Φ such that P^u_{s,P_s} solves the problem of martingale relatively to $(\tilde{F}_o,F_1,u,s,P_s)$.

Proof : Since $\|\Pi^u_{L,\beta}(T;s,P_s)\| = \|P_s\| \quad \forall u,L,\beta$, the sequence $\{\Pi^u_{L,\beta}\}$ lies in a weak * compact subset of $\mathcal{M}^b_{o,T}(n+q+1)$, and one can extract a subsequence, still noted $\{\Pi^u_{L,\beta}\}$, converging to $\Pi^u \in \mathcal{M}^b_{o,T}(n+q+1)$.

Let us firstly analyze the convergence of the projections of $\Pi^u_{L,\beta}$ on $C^{n+q}_{o,t}$ namely of $\pi^u_{L,\beta}$:

If $\varphi \in C_b^o(C_{o,t}^{n+q})$, we have :

(A.22) $E_{\Pi^u_{L,\beta}(t;s,P_s)}(\varphi) = E_{\pi^u_{L,\beta}(t;s,\mu_s)}(\varphi)$ with μ_s as in the proof of Proposition 3.1. Thus since $\pi^u_{L,\beta}(t;s,\mu_s) = R^{u,L,\beta}_{s,t}(.)\nu_{s,\mu_s}$ with

$$(A.23)\quad R^{u,L,\beta}_{s,t}(\xi) = \exp\Big[\int_s^t (\eta_1^{-1}(\sigma,\xi(\sigma))\eta_o(\sigma,\xi,u_\sigma(Y,\Pi_{j(\sigma),\beta})))'dv_s(\sigma) - \frac{1}{2}\int_s^t \|\eta_1^{-1}(\sigma,\xi(\sigma))\eta(\sigma,\xi,u_\sigma(Y,\Pi_{j(\sigma),\beta}))\|^2 d\sigma\Big]$$

where we have noted $j(\sigma) = k\ \forall\sigma \in [t_k,t_{k+1}[\cap [s,T]$, and η_o having linear growth (see (3.7)), we know (see for example [12]) that $\{R^{u,L,\beta}_{s,t}\}$ remains in a weakly compact subset of $L^1(\Omega,\nu_{s,\mu_s})$, and that there exists a subsequence, still noted $\{R^{L,\beta}_{s,t}\}$, weakly converging to $\bar{R}^u_{s,t}$ in $L^1(\Omega,\nu_{s,\mu_s})$ such that :

$$(A.24)\quad pr_{n+q}\,\bar{\Pi}^u(T;s,P_s) \overset{def}{=} \bar{\pi}^u(T;s,\mu_s) = \bar{R}^u_{s,T}\,\nu_{s,\mu_s}.$$

Consequently there exists $\bar{\eta}_o$, $\mathcal{B}_{[o,t]} \times \mathcal{F}_t \times \mathcal{Y}_t \times \mathcal{B}^t_{\mathcal{M}}$ measurable such that :

$$(A.25)\quad \bar{R}^u_{s,t}(\xi) = \exp\Big[\int_s^t (\eta_1^{-1}(\sigma,\xi(\sigma))\bar{\eta}_o(\sigma,\xi,Y,\bar{\pi}^u(\sigma;s,\mu_s)))'dv_s(\sigma) - \frac{1}{2}\int_s^t \|\eta_1^{-1}(\sigma,\xi(\sigma))\bar{\eta}_o(\sigma,\xi,Y,\bar{\pi}^u(\sigma;s,\mu_s))\|^2 d\sigma\Big]$$

But, noting $1_R(Z)$ the function equal to 1 if $\|Z\| \leq R$ and o otherwise, it is clear by (3.29) that for each $R > o$, the sequence : $\{F^{u,L,\beta}_o.1_R\}$ is bounded in $L^1(\tilde{\Omega},\tilde{\nu}_{s,P_s})$ where :

$$(A.26)\quad F^{u,L,\beta}_o(t,Z,Y,\Pi^u_{L,\beta}) = F_o(t,Z,u_t(Y,\Pi_{j(t),\beta}))$$

and $\tilde{\nu}_{s,P_s}$ is defined by $E_{\tilde{\nu}_{s,P_s}}(\varphi) = \int\varphi(\xi,\zeta)dP_s(\zeta|\xi)d\nu_{s,\mu_s}(\xi)$, $\forall\varphi \in C_b^o(\tilde{\Omega})$.

Thus, one can extract another subsequence such that $F^{u,L,\beta}_o.1_R$ converges weakly to $\bar{F}_o.1_R$ in $L^1(\tilde{\Omega},\tilde{\nu}_{s,P_s})$, with :

$$(A.27)\quad \bar{F}_o = (\bar{\eta}_o', \bar{g})'.$$

Furthermore, applying Mazur's theorem, one can find a sequence of finite convex combinations of the $\{F_o^{u,L,\beta}.1_R\}$'s, strongly converging to $\overline{F}_o.1_R$ in $L^1(\tilde{\Omega},\tilde{\nu}_{s,P_s})$. But this means that for every (t,Z,Y) such that $\|Z\| \leq R$, we have :

$$(A.28)\quad \overline{F}_o(t,Z,Y,\overline{\Pi}^u(t;s,P_s)) \in \overline{co}\Big(\bigcap_{\varepsilon>0}\bigcap_{\Pi\in\beta(\overline{\Pi}^u(t;s,P_s))} F_o(t,Z,u_t(B_\varepsilon(Y)\Pi))\Big),$$

and, letting $R \to \infty$, we obtain :

$$(A.29)\quad \overline{F}_o(t,Z,Y,\overline{\Pi}^u(t;s,P_s)) \in \Phi(t,Z,Y,u,\overline{P}^u_{s,P_s}),\quad \tilde{\nu}_{s,P_s}\ \text{a.s.}$$

Now, going back to (A.22), we have :

$$(A.30)\quad \lim_{L,\beta} E_{\Pi^u_{L,\beta}(T;s,P_s)}(\varphi) = \int \varphi(\xi)\overline{R}^u_{s,T}(\xi)d\nu_{s,\mu_s}(\xi)\quad \forall\varphi \in C_b^o(C^{n+q}_{o,T})$$

and :

$$(A.31)\quad \lim_{L,\beta} E_{\Pi^u_{L,\beta}(T;s,P_s)}(\varphi) = \int_{C^{n+q}_{o,T}}\int_{C^1_{o,s}} \varphi(\xi,\zeta + \int_s \overline{g}(\sigma,\xi,Y,\overline{\Pi}^u(\sigma;s,P_s))d\sigma)$$
$$dP_s(\zeta|\xi)\overline{R}^u_{s,T}(\xi)d\nu_{s,\mu_s}(\xi)$$

$$= E_{\overline{\Pi}^u(T;s,P_s)}(\varphi)\quad \forall\varphi \in C_b^o(\tilde{\Omega}).$$

Finally, since, by (A.30), $\overline{\pi}^u(T;s,\mu_s) = \overline{R}^u_{s,T}\nu_{s,\mu_s}$, it follows that $v_s(t) - \int_s^t \eta_1^{-1}(\sigma)\overline{\eta}_o(\sigma)d\sigma$ is a $(\mathcal{F}_t, \overline{\pi}^u(T;s,\mu_s))$ Wiener process, and concluding as in Proposition 3.1, $\overline{\Pi}^u$ solves the problem of martingale relatively to $(\overline{F}_o, F_1, u, s, P_s)$, and the result is proved. ∎

Proposition A.1 : $\forall u \in \tilde{U}$, $\forall t \in [s,T]$, $\forall s \in [o,T]$, $\forall P_s \in \mathcal{M}^b_{o,s}(n+q+1)$, we have :

$$(A.32)\qquad \mathcal{P}^u(T;s,P_s) = \bigcup_{\Pi\in\mathcal{P}^u(t;s,P_s)} \mathcal{P}^u(T;t,\Pi)$$

Proof : Clearly, we have the inclusion :

$$\mathcal{P}^u(T;s,P_s) \subset \bigcup_{\Pi \in \mathcal{P}^u(t;s,P_s)} \mathcal{P}^u(T;t,\Pi).$$

Conversely, let $\Pi^u_{t,s} \in \mathcal{P}^u(t;s,P_s)$ and $\Pi^u(T;t,\Pi^u_{t,s}) \in \mathcal{P}^u(T;t,\Pi^u_{t,s})$.
By definition, $\Pi^u_{t,s}$ is the weak * limit of a sequence $\Pi^u_{L,\beta}$.
Now, choosing L,β large enough, we have :

$$\Pi^u_{L,\beta}(t;s,P_s) \in \beta(\Pi^u_{t,s}).$$

Thus it remains to build the sequence associated to $\Pi^u(T;t,\Pi^u_{t,s})$ by letting u_t depend on $\Pi^u_{L,\beta}(t;s,P_s)$, and to choose the other $\Pi_{j,\beta}$'s to have the desired convergence to $\Pi^u(T;t,\Pi^u_{t,s})$. Thus, for every fixed L,β, we have :

$$(A.33) \qquad \Pi^u_{L,\beta}(T;t,\Pi^u_{L,\beta}(t;s,P_s)) = \Pi^u_{L,\beta}(T;s,P_s)$$

and, eventually extracting a subsequence :

$$(A.34) \qquad \Pi^u(T;t,\Pi^u_{t,s}) = \text{weak } * \lim_{L,\beta} \Pi^u_{L,\beta}(T;s,P_s),$$

which means, by definition, that $\Pi^u(T;t,\Pi^u_{t,s}) \in \mathcal{P}^u(T;s,P_s)$, and the result is proved. ∎

PARTIE III

Filtrage nonlinéaire de dimension finie pour une classe de systèmes à temps discret et continu.

RESUME DE LA IIIème PARTIE

Filtrage nonlinéaire de dimension finie pour une classe de systèmes à temps discret ou continu.

On étudie dans cette partie, le problème de l'existence et de la réalisation minimale de filtre de dimension finie pour des systèmes de la forme:

$$\begin{cases} x_{k+1} = f(x_k) \\ y_k = h(x_k) + \eta(x_k)v_k \end{cases} \quad \text{ou} \quad \begin{cases} dx_t = f(x_t)dt \\ dy_t = h(x_t)dt + dv_t \end{cases}$$

sans bruits de dynamique, mais avec, dans le cas du temps discret, des bruits d'observation dont l'intensité est correlée à l'état.

Dans le premier papier (temps discret) le système évolue sur une variété différentiable de dimension n et f est un diffeomorphisme. L'observation est dans R^p , et η est telle que $\{x \mid \det \eta(x) = 0\}$ est une sous-variété de dimension au plus n-1.

Les bruits v_k sont blancs, gaussiens, stationnaires. On commence par le calcul de l'équation d'évolution (récurrente) de la densité conditionnelle non normalisée de x_k sachant toutes les observations passées y_1 , ..., y_k , puis on montre que cette densité peut s'interprêter comme la sortie d'un système où les entrées sont les observations. C'est cette application entrée-sortie que l'on va réaliser en dimension finie. Pour cela, on montre que l'état de ce système (de dimension infinie) s'exprime de manière naturelle dans une "base canonique" obtenue simplement à partir des fonctions f, h et η, et on montre que la condition nécessaire et suffisante d'existence de filtre de dimension finie est précisément que la base canonique soit finie.

De plus, la dimension de la base canonique est égale à la dimension minimale du filtre, et on donne les équations explicites du filtre minimal. On montre ensuite des propriétés de théorie des systèmes pour le filtre minimal et pour le système (f,h,η), et on essaie d'évaluer la "grosseur" de l'ensemble des systèmes ayant un filtre de dimension finie. On donne enfin un exemple d'application à un problème de poursuite de cible mobile (conduite de tir).

Dans le second papier, on développe l'application au problème de conduite de tir en appliquant les méthodes précédentes (le filtre minimal obtenu est de dimension 4) et en les comparant aux techniques de filtrage de Kalman étendu (qui diverge presque systèmatiquement) ainsi qu'à un filtre de Kalman obtenu sur le système linéaire 2 fois dérivé du système de départ (qui s'avère inobservable).

Dans le troisième papier, on généralise les techniques précédentes au temps continu. On donne une solution explicite de l'équation de Zakaï et on déduit la base canonique comme précédemment, ainsi que la CNS d'existence d'un filtre de dimension finie et la réalisation minimale de ce filtre. La CNS est de plus équivalente à la condition de dimension finie de l'algèbre d'estimation (algèbre de Lie de l'équation de Zakaï).

EXACT FINITE DIMENSIONAL FILTERS FOR A CLASS OF NONLINEAR DISCRETE-TIME SYSTEMS[+]

J. LEVINE[*], G. PIGNIE[*]

ABSTRACT

We obtain a necessary and sufficient condition for the nonlinear discrete-time system:

$$(\Sigma) \qquad \begin{cases} x_{k+1} = f(x_k) \\ y_k = h(x_k) + \eta(x_k)v_k \end{cases}$$

to have a finite dimensional filter.

This condition is used to obtain an explicit formula of the minimal filter, as well as various systems theoretic properties of (Σ) or of the minimal filter. These results are applied to a tracking problem for a moving target.

+ This work is supported by the French Army under contract DRET, 83.34.177.

* Centre d'Automatique et d'Informatique de l'Ecole Nationale Supérieure des Mines de Paris.
35 rue St Honoré, 77305 Fontainebleau Cedex - FRANCE.

INTRODUCTION

The filtering problem, recursive estimation of a partially observed Markov process ([14],[22],[42]), has been studied by many authors. Among these papers, the Kalman filter plays a central role since it can be computed through a finite number of sufficient statistics. This property has received the name of "finite dimensionality" and the very serious difficulties met when computing nonlinear filters, have motivated to study the systems having an exact finite dimensional filter ([1],[2],[5], [6],[18],[19],[29],[30],[32],[40]), or whose filter can be approximated by a sequence of finite dimensional filters ([9],[17],[24],[25],[26], [31],[38]).

Unfortunately, in the case of exact filters, most of the results prove to be negative and establish non-existence of finite dimensional filters ([5],[19],[40]), at the exception of very rare examples ([1],[5], [6],[29]), whose applicability is questionable.

Concerning the approximation by finite dimensional filters, the situation is only clear for processes of "small" dimension since the dimensionality of the approximation increases exponentially with the degree of accuracy ([9],[17],[24],[25],[26],[31],[38]).

Starting from a different viewpoint, some authors have tried to characterize the stochastic properties of the processes that can be "realized" by a partially observed Markov chain ([3],[11],[20],[35],[41]), their finite discrete-state structure inducing naturally a finite dimensional filter.

Thus, the finite dimensionality has been mostly studied for continuous-time systems or for discrete-state systems, but very few work has been done in the case of discrete-time, continuous-state, nonlinear systems.

The aim of this paper is to study the finite dimensionality of the exact filter for a class of discrete-time (continuous-state) nonlinear systems of the form:

$$(\Sigma) \qquad \begin{cases} x_{k+1} = f(x_k) \\ y_k = h(x_k) + \eta(x_k) v_k \end{cases}$$

These systems are particularized by the fact that there is no noise on the dynamics, but the observations are perturbed by gaussian noises v_k whose intensity is correlated with the state x_k. (One sometimes says that $\eta(x).v$ is a coloured noise, but in a different sense of the one in [34]).

This correlation with the intensity of the noise can be useful in filtering applications as in communications networks with analog/digital conversions, or for systems with saturations in the observations: a simple example can be found in the observation of the state of a valve in a water reservoir by imperfectly measuring the level of water. When the water overflows, the noise cannot change the level, and we must have $\eta(\bar{x}) = 0$ where $\bar{x}$ is the height of the reservoir.

We begin the paper by a precise definition of the finite dimensional filtering problem. To this aim, we derive the evolution equation of the unnormalized conditional density in section I.2, and then define the notions of realization of this conditional density (I.3); we give the orientations of the paper in I.4.

The section II gives a complete caracterization of those systems (Σ) that admit a finite dimensional filter. The fundamental result is that the functional space generated by the functions:

$$((\eta \circ f^k(.))(\eta \circ f^k(.))')^{-1}_{ij} \quad , \quad \sum_{j=1}^{p} (\eta \circ f^k(.))(\eta \circ f^k(.))')^{-1}_{ij} \; h_j . \circ f^k(.),$$

for $i,j = 1,\ldots,p$, $k \geq 0$, is finite dimensional if and only if there exists a finite dimensional filter. Furthermore, the dimension of this space is equal to the minimal dimension of the filter. (Here $'$ denotes transposition, and f^k is the k^{th} iterate of f). The preceding space is called the canonical space (II.1). In II.2 we give an explicit formula for the minimal filter and bounds on its dimension, as well as observability and local weak reachability properties, providing thus another proof of the minimality of the filter. Three elementary examples are given. In II.3, we characterize the systems (Σ) having the same minimal filter by the notion of subordination, and deduce a necessary and sufficient condition for a system (Σ) to be immersed into a linear system.

The section III is devoted to the application of this theory to a tracking problem for a moving target.

Finally, in section IV, we evaluate the number of systems that admit a finite dimensional filter of given minimal dimension.

I - STATEMENT OF THE PROBLEM

Once the assumptions stated, we shall prove a recursion formula for the unnormalized conditional density, and then state the finite dimensional filtering problem in terms of systems realization.

I.1. The basic assumptions

We consider the following class of discrete-time nonlinear systems:

$$(1.1)\qquad \begin{cases} x_{k+1} = f(x_k) \\ y_k = h(x_k) + \eta(x_k)v_k \end{cases} \qquad k = 0, 1, \ldots$$

where:

. the state variable x_k at time k, k = 0, 1,... belongs to a pure paracompact connected C^1 manifold X of dimension n;

. the initial state x_o is a random vector in X with probability density $P_o \in C^\circ(X;R_+)$ with respect to μ_o, a given Lebesguian measure on X;

. the observation vector y_k and the noise realization v_k at time k, k = 0, 1,..., are in R^p;

and:

(H1) f is a C^1-diffeomorphism from X to X. □

(H2) $h \in C^\circ(X;R^p)$, $\eta \in C^1(X;R^{pxp})$ and:

$(\det\eta)^{-1}(0) \overset{def}{=} \{x \in X \mid \det\eta(x) = 0\}$ is a C^1 manifold of dimension at most n-1. □

(H3) the noise sequences $\{v_k\}_{k\geq o}$ are stationary and time-uncorrelated, namely: $E(v_k v_j') = E(v_k)E(v_j')$ $\forall k \neq j$ (prime denoting transpose) and $E(\phi(v_k)) = E(\phi(v_j))$ $\forall k, j$, $\forall \phi$ continuous and bounded on R^p; furthermore, each v_k has a density $V \in C^1(R^p;R_+)$ with respect to the Lebesgue measure of R^p, $\forall k = 0,1,\ldots$. □

We shall talk about the analytic, or shortly C^ω, case when X, P_o, f, h, η and V are C^ω.

Remark 1.1: The system (1.1) is a partially observed Markov process with *deterministic* transitions:

$$(1.2)\quad E(\phi(x_{k+1})/x_k) = \phi(f(x_k)) \quad \forall\, \phi \in C_b^o(X;R),\ \forall k \in N\ ,$$

where $C_b^o(X;R)$ stands for the space of continuous and bounded functions from X to R.

This quite restrictive assumption is balanced by the fact that *the intensity of the observation's noise depends on the state*. Note that, up to the authors knowledge, the filtering problem for (1.1), or for its continuous-time counterpart, has never been studied. □

Remark 2.1: (H1) holds true for example when (1.1) is obtained by "exact discretization" of an ordinary differential equation. Namely, if F is a complete vector field on X, having the form:

$$F(x) = \sum_{i=1}^{n} F^i(x)\, \frac{\partial}{\partial x^i}$$

with respect to a system of local coordinates $(x^1,\ldots,x^n)$ in a neighborhood of $x \in X$, and if $\xi_t(x) = (\xi_t^1(x),\ldots,\xi_t^n(x))$ denotes the flow defined by:

$$(1.3)\quad \begin{cases} \dfrac{d}{dt}\,\xi_t^i(x) = F^i(\xi_t(x))\ ,\ \forall t > 0 \quad ,\quad i = 1,\ldots,n, \\ \xi_o(x) = x \end{cases}$$

then, if $\theta > 0$ is given, and if we denote:

$$(1.4)\quad t_k = k\theta\ ,\ x_k = \xi_{t_k}(x)\ ,\ f(.) = \xi_\theta(.)\ ,\quad \forall k = 0,1,\ldots,$$

we obtain that:

$$x_{k+1} = f(x_k) \quad ,\quad k = 0, 1,\ldots,$$

and, if θ is small enough, it is a well-known result that $f = \xi_\theta$ is a C^1 diffeomorphism of X.

This remark is very important for the applications. □

1.2. The evolution of the unnormalized conditional density

We shall denote $Y_k = (y_1,\ldots,y_k)$, $\forall k \geq 1$, the history of observations up to time k, and Y_o the empty sequence.

Theorem 1.1: The unnormalized conditional law of x_k knowing Y_k has a density $P_k(.|Y_k) \in C^0(\tilde{X};R_+)$ with respect to μ_0, with $\tilde{X} = X - \bigcup_{j\geq o}(\det\eta \circ f^{-j})^{-1}(0)$; and $P_k(.|Y_k)$ is given by:

$$(1.5)\quad P_k(x|Y_k) = |\det\eta(x)|^{-1} V(\eta^{-1}(x)(y_k - h(x)))J_f^{-1}(f^{-1}(x))P_{k-1}(f^{-1}(x)|Y_{k-1})$$

$$\forall k > 1,\ \forall x \in X \qquad \text{(and thus almost everywhere)},$$

with: $$P_o(x|Y_o) = P_o(x),$$

and where $J_f(x)$ denotes the Jacobian of f relatively to X, evaluated at the point $x \in X$.

The proof can be found in the Appendix 1. □

Theorem 1.2: $P_k(.|Y_k)$ is the (infinite dimensional) output of the system.

$$(1.6)\quad \pi_k(x|Y_k) = V(\eta^{-1}(x)(y_k - h(x)))\pi_{k-1}(f^{-1}(x)|Y_{k-1}) \ ,\ \pi_o(x|Y_o) = P_o(x)$$

$$(1.7)\quad P_k(x|Y_k) = \prod_{j=o}^{k-1} [|\det\eta(f^{-j}(x))|\ J_f(f^{-j-1}(x))]^{-1} .\ \pi_k(x|Y_k)$$

whose state is $\pi_k(.|Y_k)$ in $C^0(\tilde{X};R_+)$, and whose inputs are the sequences $Y_k \in R^{pk}$, $\forall k \geq 1$.
In (1.7), $f^{-j}(x)$ denotes the j^{th} iterate of f^{-1}, namely:

$$f^{-j}(x) = f^{-1}(f^{-1}(\ldots(f^{-1}(x))\ldots)).$$

Proof: Expressing the outputs of (1.5) and of (1.6), (1.7) respectively, as functions of P_o, it is easy to check that they coincide $\forall k \in N$. □

Remark 1.3: Though apparently more complicated, the system (1.6), (1.7) has two advantages: on the one hand, we have separated $P_k(.|Y_k)$ into a part, (1.6), where the inputs Y_k play a dynamic role, and into a second one, (1.7), without dynamic contribution of the inputs. On the other hand, it becomes natural, with this language, to try to find a *minimal realization* (finite dimensional or not) of the system (1.6), (1.7). This will motivate the definition of a finite dimensional filter in the next paragraph. □

Remark 1.4: One can summarize the relations between inputs and outputs of (1.1) and (1.6), (1.7) by the functional diagram of Figure 1, where the noises are the inputs of the first subsystem whose outputs are the observations, which are, in turn, the inputs of the second subsystem having the conditional density as output. We shall call indifferently the second subsystem or its outputs the *filter*.

In subsystem 2 of Figure 1, we have noted:

$\tau_{f^{-1}}$ for the translation operator defined by $\tau_{f^{-1}}(\pi) = \pi(f^{-1}(.))$,
and $(D_k J_k)^{-1} = \prod_{j=o}^{k-1} [|\det \eta o f^{-j}(.)| . J_f o f^{-j-1}(.)]^{-1}$.

Clearly, our aim is to realize the second subsystem by a finite dimensional "box", having thus the same inputs and outputs, when the subsystem 1 has a suitable structure. □

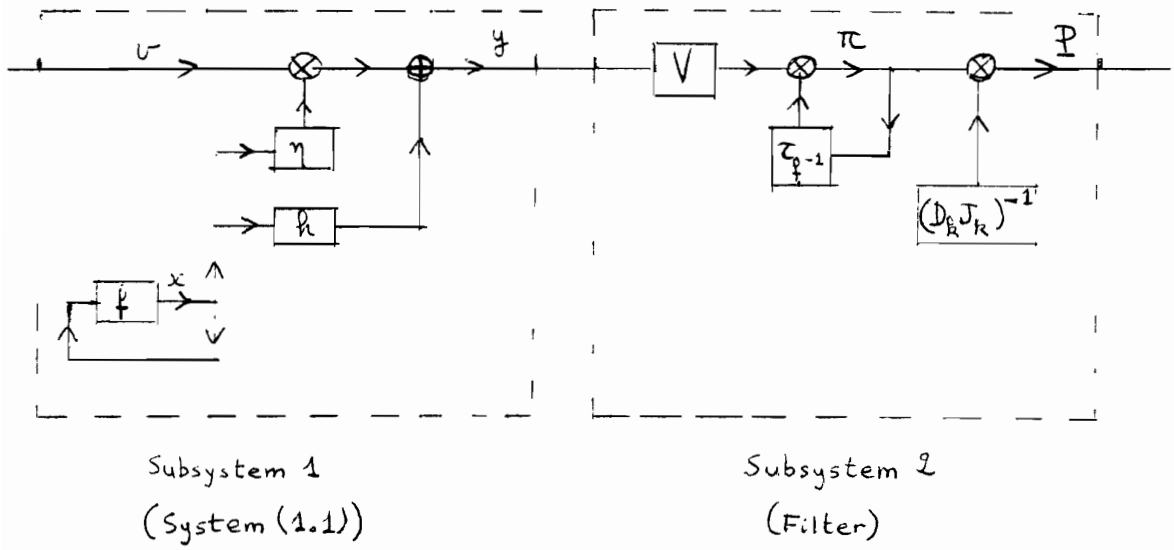


Figure 1

I.3. Definition of a finite dimensional filter through systems realization

By realization, we mean the general definition of Automata Theory (see Kalman-Falb-Arbib [22] page 11).

Definition 1.1: A *filter* (resp. local filter) for (1.1) is a realization (resp. local realization), with unspecified dimension, of (1.6), (1.7) of the form:

$$(1.8)\quad \begin{cases} a_{k+1} = \Phi_k(a_k, y_{k+1}) & \forall k \geq 0 \\ P_k(x|Y_k) = \sigma_k(a_k, x) \quad \forall x \in \tilde{X} & (\text{resp. } \forall x \in U_k = f^k(U)), \end{cases}$$

where U is a given open subset of $\tilde{X}$ and where f^k is the k^{th} iterate of f, $\sigma_k(a,.)$ being continuous on $\tilde{X}$ (resp. U_k) $\forall a$, and a_o being such that $\sigma_o(a_o,x) = P_o(x)$ $\forall x \in \tilde{X}$ (resp. $\forall x \in U$). □

Definition 1.2: 1) a *finite dimensional filter* (resp. finite dimensional local filter) is a realization of the form (1.8) where $a_k \in A$ $\forall k \geq 1$, A being a finite dimensional C^1 manifold, called the parameters space, with

$$(1.9)\quad \begin{cases} \Phi_k \in C^1(A \times R^p; A) \\ \\ \sigma_k \in C^1(A; C^\circ(\tilde{X}; R_+)) \quad (\text{resp. } C^1(A; C^\circ(U_k; R_+))). \end{cases}$$

If dim A=r, we say that the dimension of the corresponding filter is r.

2) a filter is said to be *stationary* if $\Phi_k \equiv \Phi$, $\sigma_k \equiv \sigma$, $\forall k$, and *pseudo-stationary* if $\Phi_k \equiv \Phi$ and $\pi_k(.|Y_k) = \rho(a_k,.)$ $\forall k$ (ρ independent of k).

3) a finite dimensional filter is called *analytic* if A is C^ω, if $\Phi_k \in C^\omega(A \times R^p; A)$, and $\sigma_k \in C^\omega(A; C^\circ(\tilde{X}; R_+))$ (resp. $C^\omega(A; C^\circ(U_k; R_+))$.).

4) a filter is said to be *invertible* if $\Phi_k(.,y)$ is a C^1-diffeomorphism from A to A, $\forall y \in R^p$, $\forall k \in N$, (and C^ω in the analytic case). □

Remark 1.5: In the case of a local filter, Φ_k and σ_k depend on U. Clearly, this notion is interesting in the case of a finite partitioning of $\tilde{X}$ into open submanifolds $\{U^1,\ldots,U^q\}$ such that on each U^i a finite dimensional filter exists of dimension r_i, i=1,...,q; we thus obtain a finite collection of finite dimensional local filters which generalize the global case. But apart from this situation, the interest of a finite dimensional local filter may be questionable for the applications. □

Remark 1.6: It must be noted that definitions 1.1 and 1.2 cover all the classical definitions of filters. As a result, if (Φ_k,σ_k) is a finite dimensional filter, then, firstly, P_k can be obtained as a function of a *finite number of parameters* $(a_k^1,\ldots,a_k^r)$ ($\forall k \geq 1$) whose evolution is described by $a_{k+1} = \Phi_k(a_k,y_{k+1})$, and secondly, for every ϕ continuous and bounded on R^n, $E(\phi|Y_k) = \hat{\phi}_k$ can also be obtained as the *infinite dimensional output* of the *finite dimensional system*:

$$(1.10)\quad \begin{cases} a_{k+1} = \Phi_k(a_k,y_{k+1}) \\ \\ \hat{\phi}_k = \int \phi(x)\sigma_k(a_k,x)\,dx. \end{cases}$$

In the continuous-time context, such a filter is called a "universal finite dimensional filter" (see [5]), to distinguish from the less restrictive notion of finite dimensional filter for one or a finite number of functions $\{\phi_\lambda\}$, that may be defined by (1.10) with the given functions $\{\phi_\lambda\}$.

In the language of statistics, the parameters $(a^1,\ldots,a^r)$ are called *sufficient statistics* since they sum up all the informations needed about the observations to compute the conditional density, and reject part of, or all, the redundancies of the observations $\{Y_k\}_{k\geq o}$ (see for example [27]).

For example, in the linear gaussian case, one can choose $A = R^n \times R^{nxn}$ and $a_k = (\hat{x}_k, \Sigma_k)$ with $\hat{x}_k$ the conditional mean of x_k and Σ_k the conditional variance of $(x_k-\hat{x}_k)$, ϕ_k being thus defined by the Kalman filter equation, and the output $P_k(.|Y_k)$ being the gaussian density with parameters $(\hat{x}_k,\Sigma_k) = a_k$.

Another example is the heuristic definition of remark 1.4 as a compensator (subsystem 2 of Figure 1). The next proposition proves that definition 1.1 fits with this heuristic definition. □

Proposition 1.1: The system (1.6), (1.7) is a filter in the sense of definition 1.1. Consequently, a filter of the form (1.8) always exists (but generally infinite dimensional) and can be chosen pseudo-stationary.

Proof: Let us set: $A = C^o(\tilde{X};R_+)$, $a_k = \pi_k(.|Y_k)$, $a_o = P_o(.)$,

$$(1.11)\qquad \Phi_k(a,y) = V(\eta^{-1}(.)(y-h(.))).\tau_{f^{-1}}(a) \text{ , with } \tau_{f^{-1}}(a) = a(f^{-1}(.)),$$

$$(1.12)\qquad \sigma_k(a,.) = \prod_{j=o}^{k-1} [|\det \eta o f^{-j}(.)|.J_f o f^{-j-1}(.)]^{-1}.a \text{ , } \forall k \geq 1 \text{ , } \sigma_o(a,.) = a;$$

$\sigma_k(a,.)$ being continuous on $\tilde{X}$ $\forall a$, (1.6), (1.7) takes the form (1.8), and thus (1.6), (1.7) is a filter (or equivalently the subsystem 2 of Figure 1 is a filter), and since the realization (1.6), (1.7) always exists and, since in (1.11), (1.12), $\Phi_k \equiv \Phi$ and $\rho(a,.) = a$, $\pi_k(.|Y_k) = \rho(a_k,.)$, this realization can be chosen pseudo-stationary, and the result is proved. □

Remark 1.7: In terms of functional diagram, (1.8) factorizes the subsystem 2 as shown in Figure 2:

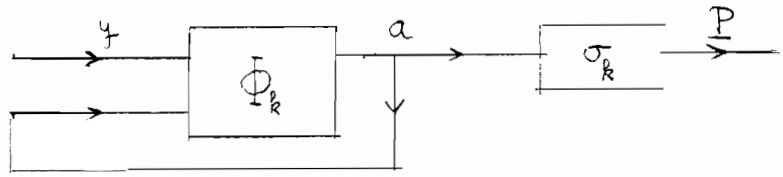


Figure 2

This diagram suggests another simplification : since only Φ_k and A are concerned with finite dimensionality, then the system made of equation (1.6) and output function $\pi_k(.|Y_k)$ (the state itself) must have the same realization structure than (1.6),(1.7). We prove this result in the following proposition. □

Proposition 1.2: To each filter (resp. local filter) of (1.1), there corresponds a realization (resp. local realization) of (1.6) with output function $\pi_k(.|Y_k)$, and conversely. Furthermore, if (1.1) admits a finite dimensional (resp. local finite dimensional, resp. analytic, resp. invertible) filter, then (1.6) with output $\pi_k(.|Y_k)$ has a realization with the same property. Finally, (1.1) has a pseudo-stationary filter if and only if (1.6) with output $\pi_k(.|Y_k)$ has a stationary realization.

Proof: Let (1.8) be a filter of (1.1). Then, denoting:

$$(1.13) \qquad \rho_k(a_k,.) = \prod_{j=o}^{k-1} \left[|\det \eta \circ f^{-j}(.)| . J_f \circ f^{-j-1}(.) \right] \sigma_k(a_k,.),$$

it follows that, for a_o such that $\rho_o(a_o,.) = \sigma_o(a_o,.) = P_o(.)$:

$$(1.14) \quad \begin{cases} a_{k+1} = \Phi_k(a_k, y_{k+1}) \\ \pi_k(.|Y_k) = \rho_k(a_k,.) \end{cases}$$

is a realization of (1.6) with output $\pi_k(.|Y_k)$, and the converse is trivial. In the local case, the proof os exactly the same with $\pi_k(.|Y_k)|_{U_k} = \rho_k(a_k,.)|_{U_k}$. For the finite dimensionality and invertibility properties, they depend only on the equation $a_{k+1} = \Phi_k(a_k, y_{k+1})$ which is the same in both systems. Finally, the analyticity property follows from the fact that ρ_k is analytic with respect to a if and only if σ_k is, and the stationarity property is an easy consequence from the definitions and (1.13), (1.14).

1.4. Statement of the existence problem for finite dimensional filters and orientations

We want to characterize the triples (f,h,η) such that, given X, V and P_o, a finite dimensional (resp. local finite dimensional) filter of the form (1.8) exists.

Furthermore, we want to know if there is a relation between the dimension of the minimal (resp. locally minimal) realization of the filter and the dimensions of X and of the observations space.

Finally, can we obtain the equations of the minimal (resp. locally minimal) filter ?

In fact, it is not surprising that we cannot answer these questions in such a generality. However, when the noises are *gaussian*, we shall give a complete solution to the three questions hereabove by algebraic and geometric methods (section II). We shall also characterize the triples (f,h,η) that admit a finite dimensional filter in terms of subordination to a system having a linear dynamics and nonlinear observations (systems of the filtering class).

A non academic example of tracking for a moving target is given in section III to illustrate these results.

A final section IV is devoted to the evaluation of the number of functions f such that (X,f,h,η,R^p) admits a finite dimensional filter of a given minimal dimension, h,η and the canonical basis being given. A partial result is obtained via algebraic topological methods.

The concluding remarks will be partly devoted to a discussion on possible methods to deal with the general case, namely the case where the noises are not gaussian, and the case of noises in the dynamical equation.

II - FINITE DIMENSIONAL FILTERS FOR GAUSSIAN NOISES

In this section, we shall suppose that:

$$(2.1) \qquad V(v) = e^{-1/2||v||^2} \quad , \qquad \text{with } ||v||^2 = v'.v \ .$$

Remark that the normalizing coefficient $(2\pi)^{-1/2}$ is useless since we consider only unnormalized densities.

Also, there is no loss of generality to assume that the variance of the noise is the identity of R^p since one can easily deduce the same results for a variance matrix Σ by changing η into $\overset{\circ}{\eta}(x) = \eta(x)\ \Sigma^{1/2}$. The same remark applies if v is not centered.

Taking into account the algebraic properties of the exponential function, and generalizing to this context the methods of exponential families in statistics (see [27]), we shall easily obtain a functional basis of the filter that will be called *canonical basis* by analogy with the classical realization theory. This canonical basis will provide the *necessary and sufficient condition* of existence of a finite dimensional filter, the *equations of the minimal realization of the filter* and, of course, the minimal dimension of the filter, as well as system theoretic caracterizations of the triples (f,h,η) by means of the concept of *immersion* (or subordination) (see [12]).

II.1. The canonical basis and the fundamental theorem

With (2.1), one can rewrite (1.6) as follows:

$$\pi_k(f(x)|Y_k) = e^{-1/2[(y_k-h(f(x)))'(\eta(f(x))\eta'(f(x)))^{-1}(y_k-h(f(x)))]} \cdot \pi_{k-1}(x|Y_{k-1}) \tag{2.2}$$

and, by induction:

$$\pi_k(f^k(x)|Y_k) = \prod_{j=1}^{k} e^{-1/2[(y_j-h(f^j(x)))'(\eta(f^j(x))\eta'(f^j(x)))^{-1}(y_j-h(f^j(x)))]} \cdot P_o(x) \tag{2.3}$$

$\forall x \in \tilde{X}$ (resp., in the local case, $\forall x \in U$).

Let us now denote:

$$\begin{cases} H_{ij}(x) = (\eta(x)\eta'(x))^{-1}_{ij} & ((i,j)^{th} \text{ component of the matrix } (\eta(x)\eta'(x))^{-1}) \\ H_i(x) = \sum_{j=1}^{p} H_{ij}(x)h_j(x) & i = 1,\dots,p \\ H_o(x) = \sum_{i=1}^{p} H_i(x)h_i(x) = \sum_{i,j=1}^{p} H_{ij}(x)h_i(x)h_j(x). & \end{cases} \tag{2.4}$$

Thus, (2.3) becomes:

$$\pi_k(f^k(x)|Y_k) = e^{-1/2 \sum_{j=1}^{k}(\sum_{\alpha,\beta=1}^{p} H_{\alpha\beta}(f^j(x))y_j^\alpha y_j^\beta - 2\sum_{\alpha=1}^{p} H_\alpha(f^j(x))y_j^\alpha + H_o(f^j(x)))} \cdot P_o(x) \tag{2.5}$$

where y_j^α is the α^{th} component of y_j at time j, $\alpha = 1,\dots,p$, $j \geq 1$.

Thus, it appears that for gaussian noises, the conditional density is the exponential of a polynomial in the $\{y_j^\alpha,\ y_j^\alpha y_j^\beta \mid \alpha,\beta=1,\dots,p,\ j \geq 1\}$ with functional coefficients $H_{\alpha\beta}of^j$ and $H_\alpha of^j$, $\alpha,\beta=1,\dots,p$, $j \geq 1$, the major fact being that these coefficients are independent of Y_j $\forall j \geq 1$

Also, one can see that the set of functions:

$$H = \{ H_{\alpha\beta}of^j,\ H_\alpha of^j \mid \alpha,\beta=1,\dots,p,\ j \geq 1 \} \tag{2.6}$$

plays a central role in the filtering problem, since $\pi_k(.|Y_k)$ and $P_k(.|Y_k)$ are completely determined by H and the observations. We shall prove in the sequel that H has intrinsic properties, as suggested by the word "canonical":

Definition 2.1: Span H is called the *canonical space* and a basis of Span H is called *canonical basis*. □

Recall that Span H is the vector space obtained by linear combinations *with constant coefficients* of the functions of H.

Remark 2.1: We have not included $\{H_o \text{of}^j,\ j \geq 1\}$ in the canonical space since this set is not affected by the observations. Furthermore, one can also reject it in the output function by considering the equivalent system:

$$(2.7)\quad \begin{cases} \tilde{\pi}_k(x|Y_k) = e^{-1/2(\sum_{\alpha,\beta=1}^{p} H_{\alpha\beta}(x)y_k^\alpha y_k^\beta - 2\sum_{\alpha=1}^{p} H_\alpha(x)y_k^\alpha)} .\tilde{\pi}_{k-1}(f^{-1}(x)|Y_{k-1}) \\ \tilde{\pi}_o(x|Y_o) \equiv 1 \\ P_k(x|Y_k) = \prod_{j=o}^{k-1} [|\det \eta \text{of}^{-j}(x)| . J_f \text{of}^{-j-1}(x)]^{-1} . e^{-1/2 \sum_{j=o}^{k-1} H_o(f^{-j}(x))} . \\ \qquad .\tilde{\pi}_k(x|Y_k). \end{cases}$$ □

Remark 2.2: Span H does not depend on P_o. We shall see that this implies that the finite dimensionality of the filter does not depend on P_o. □

Remark 2.3: Formula (2.5) can be seen as a discrete-time analogue of the *Kallianpur-Striebel's formula* (see [14]), giving the conditional measure by its density with respect to $dP_o \otimes \bigotimes_{k\geq 1} dv_k$, and then conditioned by the observations. In the same way, the exponential defining $\tilde{\pi}_k(.|Y_k)$ in (2.7) can be seen as a *likelihood ratio*.

On the other hand, the functions $H_{\alpha,\beta}\text{of}^k$, $H_\alpha \text{of}^k$ summarize the contribution of the dynamics f in the observations. Consequently, Span H has the system theoretic interpretation of an *observation space* for (1.1) (see [33]). □

Theorem 2.1: (Fundamental theorem): the following properties are equivalent.

(i) (1.1) admits a finite dimensional filter (resp. finite dimensional local filter).

(ii) Span H is finite dimensional (resp. $\exists$ U open subset of X such that Span $H|_U$ is finite dimensional).

(iii) $\exists r \in N^*$, $\exists \theta_1,\ldots,\theta_r \in$ Span H, $\exists$ R matrix (r,r) with constant coefficients such that det $R \neq 0$, and $\exists \mu_{\alpha,i}, \mu_{\alpha,\beta,i}, \alpha,\beta=1,\ldots,o$, $i=1,\ldots,r$ such that:

$$(2.8)\quad \begin{cases} \begin{pmatrix} \theta_1 \text{of}(x) \\ \vdots \\ \theta_r \text{of}(x) \end{pmatrix} = R \begin{pmatrix} \theta_1(x) \\ \vdots \\ \theta_r(x) \end{pmatrix} \qquad \forall x \in \tilde{X} \\ H_\alpha(x) = \sum_{i=1}^{r} \mu_{\alpha,i}\theta_i(x) \ , \ H_{\alpha,\beta}(x) = \sum_{i=1}^{r} \mu_{\alpha,\beta,i}\theta_i(x) \\ \forall \alpha,\beta = 1,\ldots p \ , \ \forall x \in \tilde{X}. \end{cases}$$

(resp. $\exists U$ open subset of $\tilde{X}$ such that (2.8) holds $\forall x \in U$).

Proof: See Appendix 2. □

Remark 2.4: With (2.8), we recover the intuitive idea that the filter is finite dimensional if and only if the likelihood ratio's evolution remains in a finite dimensional space. □

Remark 2.5: As previously announced, the conditions (ii) and (iii) do not depend on P_o. □

We shall now state various consequences of the fundamental theorem or of its proof.

II.2. The minimal realization of the filter

II.2.1. The equations of the minimal filter

A particularly important consequence of theorem 2.1 is the following:

Theorem 2.2: If dim Span $H = r < +\infty$, there exists an invertible analytic and pseudo-stationary minimal realization of the filter of dimension r, explicitly given, in matrix form, by: $a_k = (a_k^1,\ldots,a_k^r)'$, and:

$$(2.9)\quad \begin{cases} a_{k+1} = R'^{-1}\, a_k + M y'_{k+1} - \frac{1}{2}\, y'_{k+1}\, \Lambda\, y_{k+1}\,, \quad a_o = 0 \\ P_k(x|Y_k) = \{ \prod_{j=o}^{k-1} [|\det\eta(f^{-j}(x))|.J_f(f^{-j-1}(x))]^{-1}\, e^{-\frac{1}{2}\sum_{j=o}^{k-1} H_o(f^{-j}(x))} \\ \qquad .P_o(f^{-k}(x))\}\ \rho(a_k,x) \end{cases}.$$

with $\rho(a_k,x) = e^{\sum_{i=1}^{r} a_k^i \theta_i(x)}$, where $\{\theta_1,\ldots,\theta_r\}$ is a canonical basis with R as in (2.8), where M is the (p,r) matrix whose element μ_{ij} is the coordinate of H_i with respect to θ_j, namely:

$$(2.10)\qquad H_i(x) = \sum_{j=1}^{r} \mu_{ij}\, \theta_j(x)\,, \quad i = 1,\ldots,p\,,$$

and where Λ is the tensor of order 3 defined by:

$$(2.11)\qquad y' \Lambda\, y = \begin{pmatrix} \sum_{\alpha,\beta=1}^{p} \lambda^{1}_{\alpha,\beta}\, y^{\alpha}\, y^{\beta} \\ \vdots \\ \sum_{\alpha,\beta=1}^{p} \lambda^{r}_{\alpha,\beta}\, y^{\alpha}\, y^{\beta} \end{pmatrix} \qquad \forall y = (y^1,\ldots,y^p)' \in \mathbb{R}^p,$$

with $\lambda^{i}_{\alpha,\beta}$ the coordinate of $H_{\alpha,\beta}$ with respect to θ_i, namely:

$$(2.12)\qquad H_{\alpha,\beta}(x) = \sum_{i=1}^{r} \lambda^{i}_{\alpha,\beta}\, \theta_i(x)\ ,\ \lambda^{i}_{\alpha,\beta} = \lambda^{i}_{\beta,\alpha} \qquad \forall \alpha,\beta = 1,\ldots,p\ ,\ i=1,\ldots,r.$$

Furthermore, any other minimal invertible analytic and pseudo-stationary filter can be obtained from (2.9) by C^{ω}-diffeomorphism. Finally, if the collection $\{H_{\alpha,\beta}, H_{\alpha} \mid 1 \le \alpha \le \beta \le p\}$ is linearly independent, there exists an integer $k_o \ge 1$ such that $p(p+3)/2 \le r \le k_o p(p+3)/2$ and $r = k_o p(p+3)/2$ if the collection $\{H_{\alpha,\beta}of^{j}, H_{\alpha}of^{j} \mid 1 \le \alpha \le \beta \le p,\ 0 \le j \le k_o-1\}$ is a canonical basis.

Proof: (2.9) has been proved to be a minimal filter in Appendix 2, proof of (iii) => (i) and remark A.2.

The equivalence of every minimal invertible C^{ω} and pseudo-stationary realizations up to a C^{ω} diffeomorphism is proved in Jakubczyk [21].

Finally, the bound on r follows from (iii) of theorem 2.1 since there exists at least an integer k_o such that $\{H_{\alpha,\beta}of^{j}, H_{\alpha}of^{j} \mid 1 \le \alpha \le \beta \le p,\ 0 \le j \le k_o-1\}$ contains a canonical basis. Since the maximal number of independent functions in this collection is $k_o(p(p+1)/2 + p) = k_o p(p+3)/2$, the result is proved. □

Remark 2.6: Theorem 2.2 has been stated in the global case. The local case can be easily obtained by restricting Span H, o and P_k to U, open subset of $\hat{X}$, and these straightforward modifications are left to the reader. □

Remark 2.7: Until now, we have used the simplest definition of a minimal realization, namely, the realization having the smallest dimension of the state space (here called the parameters space A), and we have proved that dim A_{min} = dim Span H = r, with similar methods as [21].

Another approach to obtain minimal realizations consists in characterizing them through observability and reachability properties (see for example [13],[21],[33],[37],[39]). We shall prove that the minimal realization (2.9) is observable and locally weakly reachable after recalling these basic definitions. □

II.2.2. Observability and reachability

Definition 2.2: The system (2.9) is *observable* iff for every sequence $\{y_1,\ldots,y_k,\ldots\}$, $a \neq b$, $a \in A$, $b \in A$, implies $\rho(a_k,.) \neq \rho(b_k,.)$ $\forall k \geq 1$ where a_k and b_k are the solutions of the induction (2.9) generated by $\{y_1,\ldots,y_k,\ldots\}$ and initial $a_o = a$ and $b_o = b$ respectively. □

Definition 2.3: The *reachable set* $R(a_o)$ from $a_o \in A$, for (2.9), is: $R(a_o) = \{a \in A \mid \exists k \in N, \exists Y_k = (y_1,\ldots,y_k) \in R^{pk}$ such that $a_k(a_o,Y_k) = a\}$, where $a_k(a_o,Y_k)$ is the solution of (2.9) generated by Y_k from a_o. □

Definition 2.4: A *backward transition* of length $k \in N$, generated by Y_k from a_k, for (2.9), is the transition noted $a_k^{-1}(a_k,Y_k)$, defined by $a_k = a_k(a_o,Y_k) \iff a_o = a_k^{-1}(a_k,Y_k)$. a_k^{-1} is well defined since R is invertible. □

Definition 2.5: The *weakly reachable set* $R_w(a_o)$ from $a_o \in A$, for (2.9), is:

$$R_w(a_o) = \{a \in A \mid \exists\, s \in N, \exists k_1,\ldots,k_s \in N, \exists Y_{k_1},\ldots,Y_{k_s} \text{ such that}$$

$$a = a_{k_s}(a^{-1}_{k_{s-1}}(\ldots(a_{k_1}(a_o,Y_{k_1}),\ldots), Y_{k_{s-1}}), Y_{k_s}) \}$$

(namely a is obtained by k_1 onward transitions, followed by k_2 backward transitions, followed by..., followed by k_{s-1} backward transitions, followed by k_s onward transitions). □

Definition 2.6: (2.9) is *locally weakly reachable* at a_o iff, given a neighborhood V_{a_o} of a_o in R^r, $R_w(a_o) \cap V_{a_o}$ is a neighborhood of a_o. □

Remark 2.8: These definitions are borrowed from [21] and [33] and generalize the classical concepts of continuous time systems theory. It has been proved in [21] that $R_w(a_o)$ is a manifold. If we denote $T_{a_o}(R_w(a_o))$ its tangent space at the point a_o, it is not difficult to prove that:

$$\dim T_{a_o}(R_w(a_o)) = \dim A$$

is equivalent to the local weak reachability of (2.9). This idea is also borrowed from [33] and will be used to prove the local weak reachability of (2.9). □

Theorem 2.3: The minimal filter (2.9) is observable and locally weakly reachable.

Proof: see Appendix 3. □

II.2.3. Elementary examples

The examples presented in this paragraph are purely academic and our aim is just to prove the simplicity and efficiency of our method. A "real" applied problem will be developed in Section IV.

Example 1.

Let $X =]0,1[$. The system is given by:

$$(2.13)\quad \begin{cases} x_{k+1} = \dfrac{1-x_k}{1+x_k} \\ \\ y_k = \eta(x_k)v_k \end{cases} \qquad \forall k \in N$$

where $\eta(.)$ is an arbitrary scalar function from $]0,1[$ to $]0+\infty[$, and where x_o is a random variable in $]0,1[$ whose probability density is P_o on $]0,1[$.

Here, we have:

$$(2.14)\quad H_1(x) = 0 \ , \quad H_{1,1}(x) = \frac{1}{(\eta(x))^2} \qquad \forall x \in]0,1[.$$

Thus:

$$(2.15)\quad H = \{\frac{1}{(\eta(.))^2}, \frac{1}{(\eta(f(.)))^2}, \frac{1}{(\eta(f(f(.))))^2} \ ,\ldots\} \quad \text{with } f(x) = \frac{1-x}{1+x} \ .$$

But it is easy to check that

$$(2.16)\quad f^{2k}(x) = x \ , \quad f^{2k+1}(x) = \frac{1-x}{1+x} \ , \quad \forall x \in]0,1[\ , \ \forall k \geq 0,$$

and that:

$$(2.17)\quad \theta_1(x) = (\eta(x))^{-2} \ , \quad \theta_2(x) = (\eta(\frac{1-x}{1+x}))^{-2}$$

is a canonical basis of H, with dim Span $H = 2$.

Furthermore, $\theta_1(f(x)) = \theta_2(x)$, $\theta_2(f(x)) = \theta_1(x)$, so that:

$$(2.18)\quad \begin{pmatrix} \theta_1 \text{of} \\ \\ \theta_2 \text{of} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \\ \theta_2 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 0 & 1 \\ \\ 1 & 0 \end{pmatrix}.$$

On the other hand, since $H_1(x) = 0$ and $H_{1.1}(x) = \theta_1(x)$, we have:

$$(2.19)\quad M = (0,0) \quad , \quad \lambda_{11}^1 = 1 \quad , \quad \lambda_{11}^2 = 0,$$

and the minimal filter, using (2.9), is:

$$(2.20)\qquad \begin{pmatrix} a^1_k \\ a^2_k \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a^1_{k-1} \\ a^2_{k-1} \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} y^2_k \quad , \quad a^1_o = a^2_o = 0$$

with, $\forall k \geq 1$:

$$(2.21)\qquad \begin{cases} P_{2k}(x|Y_{2k}) = \dfrac{P_o(x)}{(\eta(x)\,\eta(\frac{1-x}{1+x}))^k} \cdot e^{\left(\frac{a^1_{2k}}{(\eta(x))^2} + \frac{a^2_{2k}}{(\eta(\frac{1-x}{1+x}))^2}\right)} \\ P_{2k+1}(x|Y_{2k+1}) = \dfrac{2\,P_o(\frac{1-x}{1+x})}{(\eta(x))^{k+1}\,(\eta(\frac{1-x}{1+x}))^k (1+x)^2} \cdot e^{\left(\frac{a^1_{2k+1}}{(\eta(x))^2} + \frac{a^2_{2k+1}}{(\eta(\frac{1-x}{1+x}))^2}\right)} \end{cases}$$

Remark that the local weak reachability of (2.20) is equivalent to the controllability of the *linear* system:

$$\xi_{k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \xi_k + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_k$$

since the backward transitions have the same result than $(-y^2)$ and, taking $u = \pm y^2$, the assertion is proved. Of course, Kalman's criterion gives: $\operatorname{rank} \left[\begin{array}{c|cc} 1 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right] = \operatorname{rank} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = 2$ which is the desired result.

To conclude, the same kind of example can be obtained with any function f such that: $\exists\, k_o \in N$ for which $f^{k_o}(x) = x \ \forall x \in X$, or, also, when η satisfies $\eta(\alpha x) = \phi(\alpha)\eta(x) \quad \forall x \in X, \forall \alpha \in R$, and when $\exists\, k_o \in N$ such that $f^{k_o}(x) = \alpha x \quad \forall x \in X$. All these models correspond to the filtering problem for periodic or oscillating state variable. □

Example 2 (linear gaussian system)

$$(2.22)\qquad \begin{cases} x_{k+1} = Fx_k \\ y_k = Hx_k + Jv_k \end{cases}$$

where J is an invertible (p,p) matrix, H:a(p,n) matrix, and F an invertible (n,n) matrix. $X = R^n$.

We have:

$$(2.23) \qquad H = \{(JJ')^{-1}_{\alpha,\beta},\ \sum_{\beta=1}^{p}\ \sum_{\gamma,\delta=1}^{n} (JJ')^{-1}_{\alpha\beta}\ H_{\beta\gamma}(F)^k_{\gamma\delta}x^{\delta},\ \alpha,\beta = 1,\ldots,p,\ k \geq 0\ ;$$

with $(JJ')^{-1}_{\alpha\beta}$ the (α,β) component of the matrix $(JJ')^{-1}$ and $(F)^k_{\gamma\delta}$ the (γ,δ) component of F^k. Thus:

$$(2.24) \qquad \text{Span } H \subset \text{Span } \{1,x^1,\ldots,x^n\}\ ,\ \dim \text{Span } H \leq n+1\ ,$$

where x^i denotes the i^{th} component of x.

More precisely, if:

$$(2.25) \qquad \text{rank} \begin{pmatrix} H \\ HF \\ \vdots \\ HF^{n-1} \end{pmatrix} = m \leq n\ ,$$

it is easily seen that dim Span $H = m+1$ and, taking a basis of R^m as $\{x^1,\ldots,x^m\}$, we have:

$$(2.26) \qquad \text{Span } H = \text{Span } \{1,\ x^1,\ \ldots,\ x^m\}$$

It must be noted that the classical Kalman filtering theory asserts that, under the assumption (2.25), the Kalman filter of (2.22) is the minimal *stochastic* realization of (2.22), or, equivalently, the minimal realization of the input-output system with inputs the observations and output function the *normalized* conditional density, *with dimension m* (see for example [10]) *and not m+1* as (2.26). But the normalization imposes one more relation, $\int P_k dx = 1$, between the coefficients of the unnormalized filter and thus makes the dimension decrease of 1 unit. (see also a discussion about this problem in [2]). This example shows that the unnormalized minimal realization can have a higher dimension than the normalized minimal one; nevertheless, with our local methods, it seems to be very difficult to take into account the normalizing factor which is defined by a *global* relation.

Let us now suppose that m = n for simplicity's sake. From (2.26), the canonical basis is given by:

$$(2.27) \qquad \theta_o(x) = 1\ ,\ \theta_i(x) = x^i\ ,\ i = 1,\ldots,n.$$

Thus, since $\theta(Fx) = R\theta(x)$ with:

$$(2.28) \qquad R = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & F \end{array}\right) \begin{matrix} \updownarrow 1 \\ \updownarrow n \end{matrix}$$

$$\qquad\qquad\quad \underset{1}{\longleftrightarrow}\ \underset{n}{\longleftrightarrow}$$

and: $H_\alpha(x) = \sum_{i=1}^{n} \sum_{\beta=1}^{p} (JJ')^{-1}_{\alpha\beta} H_{\beta i}\, \theta_i(x)$, $H_{\alpha\beta}(x) = (JJ')^{-1}_{\alpha\beta}\, \theta_o(x)$,

$\alpha, \beta = 1, \ldots, p,$

we have:

$$(2.29) \qquad M = \overset{\overset{1}{\longleftrightarrow}\overset{n}{\longrightarrow}}{\left(\begin{array}{c|c} 0 & (JJ')^{-1}H \end{array}\right)} \updownarrow p$$

and

$$(2.30) \qquad y'\Lambda\, y = \begin{pmatrix} y'(JJ')^{-1}y & \updownarrow 1 \\ 0 & \times \\ \vdots & \updownarrow n \\ 0 & \end{pmatrix} .$$

Thus, noting $(a, b^1, \ldots, b^n)$ the filter's variables, the minimal filter is given by:

$$(2.31) \qquad \begin{cases} a_k = a_{k-1} - \frac{1}{2}\, y_k'(JJ')^{-1}y_k \quad , & a_o = 0 \\ b_k = F'^{-1}b_{k-1} + H'(JJ')^{-1}y_k \, , & b_o = 0 \end{cases}$$

and:

$$(2.32) \qquad P_k(x|Y_k) = |J|^{-k}|F|^{-k}\, e^{-\frac{1}{2}\sum_{j=o}^{k-1} \| J^{-1}HF^{-j}x\|^2} \cdot P_o(F^{-k}x) \cdot e^{(a_k + \sum_{i=1}^{n} b_k^i x^i)}$$

It can be seen, when P_o is gaussian, that a_k and b_k can be obtained as functions of σ_k and $\hat{x}_k$, the conditional variance and mean, respectively, of the Kalman filter at time k (see [10],[22],[23]). Namely, $b_k = \sigma_k^{-1}\hat{x}_k - F'^{-k}\sigma_o^{-1}\hat{x}_o$, and a_k is the parameter left for the normalization:

$$\exp a_k = (2\pi)^{n/2}|\sigma_k|^{1/2}\, |J|^k|F|^k . \exp\left(\tfrac{1}{2}(\hat{x}_k'\sigma_k^{-1}\hat{x}_k - \hat{x}_o'\sigma_o^{-1}\hat{x}_o)\right) . \int P_k(x|Y_k)dx.$$

These tedious calculations are left to the reader.

(2.31) is known as the "information filter" (see [23]) and is numerically well adapted to degenerated initial data. Also, (2.31) is valid even if P_o is *not gaussian* whereas in this case the Kalman filter doesn't work. □

Example 3 (polynomial observations). Let $X =]0, \infty[$ and:

$$(2.33) \qquad \begin{cases} x_{k+1} = \dfrac{F}{x_k} \\ y_k = h\, x_k^p + \eta x_k^q\, v_k \end{cases}$$

where F, h and η are non zero real numbers, $F > 0$, and where p, $q \in N - \{0\}$ are arbitrary. As before, v is supposed gaussian and P_o is a given density on $]0, \infty[$.

It is easily checked that:

$$\text{(2.34)} \qquad \text{Span } H = \text{Span}\{\, x^{-2q} \,,\, x^{p-2q} \,,\, x^{2q-p} \,,\, x^{2q} \}$$

and dim Span $H = 4$ if $p \neq 2q$ (if $p = 2q$, dim Span $H = 3$).

Let us denote:

$$\text{(2.35)} \qquad \theta_1(x) = x^{-2q} \,,\ \theta_2(x) = x^{2q} \,,\ \theta_3(x) = x^{p-2q} \,,\ \theta_4(x) = x^{2q-p} \,.$$

It is easy to check that:

$$\text{(2.36)} \qquad R = \begin{pmatrix} 0 & F^{-2q} & 0 & 0 \\ F^{2q} & 0 & 0 & 0 \\ 0 & 0 & 0 & F^{p-2q} \\ 0 & 0 & F^{2q-p} & 0 \end{pmatrix} \,,\ M = (0,0,\frac{h}{\eta^2},0) \,,\ \Lambda y^2 = \begin{pmatrix} \frac{1}{\eta^2}\, y^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and thus, since $R'^{-1} = R'$:

$$\text{(2.37)} \qquad \begin{cases} a_k^1 = F^{2q}\, a_{k-1}^2 - \frac{1}{2\eta^2}\, y_k^2 \\ a_k^2 = F^{-2q}\, a_{k-1}^1 \\ a_k^3 = F^{2q-p}\, a_{k-1}^4 + \frac{h}{\eta^2}\, y_k \\ a_k^4 = F^{p-2q}\, a_{k-1}^3 \end{cases}$$

and:

$$\text{(2.38)} \qquad \begin{aligned} P_{2k}(x|Y_{2k}) &= (\eta^2 F^q)^{-k} . e^{-\frac{hk}{2\eta^2}(x^{p-2q} + F^{p-2q} x^{2q-p})} . P_o(x) . e^{\sum_{i=1}^{4} a_{2k}^i \theta_i(x)} \\ P_{2k-1}(x|Y_{2k-1}) &= (\eta^2 F^q)^{-(k-1)} |\eta F^{-1} x^{q+2}|^{-1} e^{-\frac{h}{2\eta^2}(kx^{p-2q} + (k-1)F^{p-2q} x^{2q-p})} \\ &\quad . P_o(\tfrac{F}{x}) . e^{\sum_{i=1}^{4} a_{2k-1}^i \theta_i(x)} \qquad \forall k \geq 1 . \end{aligned}$$

Remark that the same kind of result can also be obtained with the system

$$\text{(2.39)} \qquad \begin{cases} x_{k+1} = F x_k \\ y_k = h x_k^p + \eta x_k^q v_k \end{cases}$$

without any restriction on p and q, whereas for continuous-time models, (see for example the cubic sensor problem [5],[19],[40]) finite dimensional filters for linear systems with polynomial observations generically don't exist. This assertion must be however cooled down by the fact that we have no noise in the dynamics, which makes quite a difference. □

Remark 2.9: The systems (2.13) and (2.33) of examples 1 and 3 cannot be obtained, by a nonlinear transformation, from a linear system (see corollary 2.1 below). But there exist transformations that change the original systems (2.13) and (2.33) into systems having linear dynamics but nonlinear observations: for (2.13), it has been seen that

$\theta_1(x) = \frac{1}{(\eta(x))^2}$, $\theta_2(x) = \frac{1}{(\eta(\frac{1-x}{1+x}))^2}$ satisfy, with $\theta_{1,k} = \theta_1(x_k), \theta_{2,k} = \theta_2(x_k)$:

$$(2.13)' \begin{cases} \begin{pmatrix} \theta_{1,k} \\ \theta_{2,k} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \theta_{1,k-1} \\ \theta_{2,k-1} \end{pmatrix}, \begin{pmatrix} \theta_{1,k} \\ \theta_{2,k} \end{pmatrix} \in]0,\infty[\times]0,\infty[= \Theta, \quad \forall k. \\ y_k = \theta_{1,k}^{-1/2} v_k \end{cases}$$

which is equivalent to (2.13) in the filtering point of view.
For (2.33), we had:

$$\theta_1(x) = x^{-2q} , \theta_2(x) = x^{2q} , \theta_3(x) = x^{p-2q} , \theta_4(x) = x^{2q-p}$$

and, noting $\theta_{i,k} = \theta_i(x_k)$, $i = 1,\ldots,4$, we easily see that (2.33) is equivalent in the filtering point of view to:

$$(2.33)' \begin{cases} \begin{pmatrix} \theta_{1,k} \\ \theta_{2,k} \\ \theta_{3,k} \\ \theta_{4,k} \end{pmatrix} = \begin{pmatrix} 0 & F^{-2q} & 0 & 0 \\ F^{2q} & 0 & 0 & 0 \\ 0 & 0 & 0 & F^{p-2q} \\ 0 & 0 & F^{2q-p} & 0 \end{pmatrix} \begin{pmatrix} \theta_{1,k-1} \\ \theta_{2,k-1} \\ \theta_{3,k-1} \\ \theta_{4,k-1} \end{pmatrix}, \theta_k \in]0,\infty[^4 = \Theta, \ \forall k \\ y_k = h\, \theta_{2,k}\, \theta_{3,k} + \eta \theta_{2,k}^{1/2} v_k \ . \end{cases}$$

In both examples, the functions of the canonical basis play the role of the nonlinear transformation (immersion, see next section). This transformation does not preserve the dimension of the original system since (2.13) and (2.33) have dimension 1 whereas (2.13)' has dimension 2 and (2.33)' has dimension 4. This result is in fact general and will be discussed in the next section where the concept of subordination or immersion, introduced in [12] and [33], is used. Here (2.13) (resp.(2.33)) is subordinated to (2.13)' (resp. (2.33)'). □

Remark 2.10: The relation (2.8) gives a systematic technique to build systems having finite dimensional filters: fix (h(.) and η(.), compute $H_{\alpha,\beta}(.)$, $H_\alpha(.)$, $\alpha,\beta = 1,\ldots,p$, and then find f(.) such that for a given k_o, all the $H_{\alpha,\beta}\text{of}^{k_o}$, $H_\alpha\text{of}^{k_o}$ are linear combinations of the preceding $H_{\alpha,\beta}\text{of}^{j}$, $H_\alpha\text{of}^{j}$, $j \leq k_o - 1$. It is particularly simple for $k_o = 1$ where we just have to solve:

$$\begin{cases} H_{\alpha,\beta}\circ f = \sum_{\gamma,\delta=1}^{p} (r_{\alpha\beta}^{\gamma} H_{\gamma} + s_{\alpha\beta}^{\gamma\delta} H_{\gamma\delta}) & \alpha,\beta = 1,\dots,p. \\ H_{\alpha}\circ f = \sum_{\gamma,\delta=1}^{p} (r_{\alpha}^{\gamma} H_{\gamma} + s_{\alpha}^{\gamma\delta} H_{\gamma\delta}) & \alpha = 1,\dots,p. \end{cases}$$

It follows that in fact "many" nonlinear discrete-time systems of the form (1.1) have a finite dimensional filter, contrarily to the appearances. The epithet "many" will be precised in section IV. □

II.3. Subordination of the systems (X,f,h,η) having a finite dimensional filter

Let us consider two systems:

$\Sigma = (X,f,h,\eta,R^p,x_o)$ defined by:

$$(\Sigma)\begin{cases} x_{k+1} = f(x_k) & , \quad x_o \in X \\ y_k = h(x_k) + \eta(x_k)v_k \end{cases}$$

with assumptions (H1), (H2) and $v_k \in R^p$ arbitrary, (X may be replaced by an open subset UcX in the local case. The local statements are straightforward and left to the reader),

and $\Sigma' = (Z,g,\psi,\chi,R^p,z_o)$ defined by:

$$(\Sigma')\begin{cases} z_{k+1} = g(z_k) & , \quad z_o \in Z \\ \zeta_k = \psi(z_k) + \chi(z_k)v_k \end{cases}$$

where $z_k \in Z\ \forall k$, Z: C^1 r-dimensional manifold, $g \in C^1(X;X)$, $\psi \in C^1(\tilde{Z};R^p)$, $\chi \in C^1(\tilde{Z};R^{p^2})$, where $\tilde{Z}$ is an open dense subset of Z, $\zeta_k \in R^p\ \forall k$, $v_k \in R^p\ \forall k$.

Definition 2.7: We say that Σ' *is subordinated to* Σ if there exists an open dense subset $\tilde{X}$ of X and an application $\theta \in C^1(\tilde{X};Z)$ such that if $\theta(x_o) = z_o$, we have:

$$(2.40)\qquad h(x_k) + \eta(x_k)v_k = \psi(z_k) + \chi(z_k)v_k \qquad \forall v_k \in R^p\ ,\ \forall k \in N,$$

$\forall x_o \in \tilde{X}$, with $x_k = f^k(x_o)$, $z_k = g^k(z_o)$.

We say that θ is an *immersion* from Σ to Σ'. □

Remark 2.11: This definition is slightly adapted from [33] to make it fit with our problem: the system Σ' can be less regular than Σ and the immersion θ is only of class C^1 on a dense subset of X, whereas in the classical definition, everything is C^ω.

We shall now specialize to systems Σ' "of the filtering class".

Definition 2.8: A system Σ' belongs to the *filtering class* iff:

(i) $Z \subset R^r$

(ii) $g(z) = Rz$ with R: invertible (r,r) matrix.

(iii) there exists $\frac{p(p+1)}{2}$ vectors $\lambda_{\alpha,\beta} \in R^r$, $1 \leq \alpha \leq \beta \leq p$, such that the symmetric $(p,p)^2$ matrix:

$$L(z) = \begin{pmatrix} \lambda'_{11}z & \cdots & \lambda'_{1p}z \\ \vdots & & \vdots \\ \lambda'_{1p}z & \cdots & \lambda'_{pp}z \end{pmatrix} \tag{2.41}$$

is positive definite $\forall z \in Z$, and a (p,r) matrix M such that: $\psi(z) = L^{-1}(z)Mz$ and $\chi(z) = N(z)$, $\forall z \in Z$, with N(z) satisfying $N(z)N'(z) = L^{-1}(z)$, N(.): C^1 on an open dense subset $\mathring{Z}$ of Z. □

Theorem 2.4: Σ admits a finite dimensional filter if and only if there exists a system Σ' of the filtering class subordinated to Σ.

Proof: Suppose that Σ admits a finite dimensional filter. Then, by theorem 2.2, the canonical basis $\{\theta_1,\ldots,\theta_r\}$ of Span H satisfies: $\theta o f = R\theta$ with R invertible (r,r) matrix. Let us prove that θ is an immersion of Σ into:

$$(\Sigma') \begin{cases} z_{k+1} = Rz_k \quad , \quad z_0 \in Z \subset R^r \\ y_k = L^{-1}(z_k)Mz_k + N(z_k)v_k \end{cases}$$

with L(z) defined by (2.41) with $\lambda_{\alpha,\beta}$ given by (2.12), the matrix M given by (2.10), and N(.) suitably chosen.
But this is straightforward since (2.12) implies that:

$$(\eta(x)\eta'(x))^{-1}_{\alpha,\beta} = \lambda'_{\alpha,\beta}\theta(x) \qquad \forall \alpha,\beta = 1,\ldots,p \quad \text{with} \quad \lambda_{\alpha\beta} = \lambda_{\beta\alpha},$$

or:

$$(\eta(x)\eta'(x))^{-1} = L(\theta(x)). \tag{2.42}$$

Also, by (2.10):

$$(2.43) \qquad (\eta(x)\eta'(x))^{-1}h(x) = M(x)$$

or:

$$(2.44) \qquad h(x) = L^{-1}(\theta(x))M\theta(x).$$

But, from (2.42), one can find a "square root" of $L^{-1}(\theta(x))$, noted $N(\theta(x))$, namely such that: $N(z)N'(z) = L^{-1}(z)$, and:

$$(2.45) \qquad \eta(x) = N(\theta(x)) \qquad \forall x \in \tilde{X},$$

$N(.)$ being of class C^1 on an open dense subset of R^r (see for example [15]). Thus, finally:

$$y_k = h(x_k) + \eta(x_k)v_k = L^{-1}(\theta(x_k))M\theta(x_k) + N(\theta(x_k))v_k$$

and the result is proved since $\Theta(x_k) = R^k\Theta(x_o) = R^k z_o = z_k$.

Conservely, let Σ' of the filtering class be subordinated to Σ, and let $\theta = (\theta_1,\ldots,\theta_r)$ be the associated immersion from Σ to Σ'.

Since $\forall x_o, z_o$ such that $z_o = \Theta(x_o)$:

$$(2.46) \quad h(x_k) + \eta(x_k)v_k = L^{-1}(z_k)Mz_k + N(z_k)v_k \quad \forall v_k \in R^p, \quad \forall k,$$

we have, taking $v_k = o$:

$$(2.47) \quad h(x_k) = L^{-1}(z_k)Mz_k, \quad \eta(x_k) = N(z_k).$$

Since $N(z)N'(z) = L^{-1}(z)$, we obtain :

$$(2.48) \quad (\eta(x_k)\eta'(x_k))^{-1} = L(z_k), \quad (\eta(x_k)\eta'(x_k))^{-1}h(x_k) = Mz_k$$

or, $\forall x_o \in \tilde{X}$:

$$(2.49) \quad (\eta(f^k(x_o))\eta'(f^k(x_o)))^{-1} = L(R^k\Theta(x_o)),$$

$$(\eta(f^k(x_o))\eta'(f^k(x_o)))^{-1}h(f^k(x_o)) = MR^k\Theta(x_o)$$

Hence : $H_{\alpha,\beta}of^k(x_o) = L_{\alpha,\beta}(R^k\Theta(x_o)) = \lambda'_{\alpha,\beta}\, R^k\Theta(x_o)$

$H_\alpha of^k(x_o) = M'_\alpha R^k\Theta(x_o), \quad \forall x_o \in \tilde{X}, \quad \forall\alpha,\beta, \quad \forall k,$

and thus:

$H = \{H_{\alpha,\beta}of^k, H_\alpha of^k, \alpha,\beta=1,\ldots,p,\ k\geqslant o\}$ is spanned by $\{\Theta_1,\ldots,\Theta_r\}$
or : dim Span $H \leqslant r$, which is the desired result. □

Corollary 2.1: A necessary and sufficient condition for the existence of a linear system subordinated to Σ is that Σ admits a finite dimensional filter and $\eta(.) \equiv$ constant.

Proof: It suffices to prove that Σ' is linear if and only if η = constant. Suppose that Σ' is linear and subordinated to Σ. Then, noting Σ':

$$(\Sigma')\begin{cases} z_{k+1} = Rz_k \\ y_k = Hz_k + Jv_k \end{cases}$$

we have: $h(x_k) + \eta(x_k)v_k = H\theta(x_k) + Jv_k \quad \forall v_k, \forall k, \forall x_o$, θ being the corresponding immersion from Σ to Σ'.

Thus, we immediately obtain:

$h(x) = H\theta(x)$, $\eta(x) = J$, which is the desired result.

Conversely, if Σ admits a finite dimensional filter and if η is constant, by the same method as in the first part of the proof of theorem 2.4, (2.42) holds and thus: $(\eta\eta')^{-1} = L(\theta(x)) \quad \forall x$, or: $L \equiv$ constant on $\theta(X)$, and by (2.44), $h(x) = L^{-1}M\theta(x)$ on $\theta(X)$. Thus, Σ is immersed in Σ' defined by:

$$(\Sigma')\begin{cases} z_{k+1} = Rz_k \\ y_k = L^{-1}Mz_k + \eta v_k \end{cases}$$

and the result is proved. □

Remark 2.12: Theorem 2.4 can be used to classify the systems having a finite dimensional filter : the relation $\Sigma \sim \Sigma'$ defined by: "Σ is subordinated to Σ' or Σ' is subordinated to Σ", is an equivalence relation and it can be easily checked that if $\Sigma \sim \Sigma'$, Σ and Σ' have the same parameters equation of their minimal filter (the output functions can differ). Conversely, if Σ and Σ' have the same parameters equation for their minimal filters, the associated R, M and Λ are the same and, using the same method as before, one can find a system Σ'' such that Σ'' is subordinated to both Σ and Σ', if and only if the same choice of $N(\theta)$ fit to Σ and Σ'. If this is the case, we have: $\Sigma \sim \Sigma' \sim \Sigma''$. Consequently, a *complete description of the systems of the form (1.1) with the assumptions (H1), (H2), and with gaussian noises, admitting a finite dimensional filter can be made in terms of the quadruple* (R, M, Λ, N).

To illustrate this assertion, it is not difficult to see that the system

$$(\Sigma)\begin{cases} x_{k+1} = \dfrac{1-x_k}{1+x_k} \\ y_k = \eta(x_k)v_k \end{cases}$$

of Example II.2.1, is equivalent to:

$$(\Sigma')\begin{cases} x_{k+1} = x_k + \dfrac{\pi}{2} \\ y_k = \phi(\sin x_k)v_k \end{cases} \quad \text{with } \phi \text{ : arbitrary function satisfying}$$

and also to:

$$(\Sigma'') \quad \begin{cases} \begin{pmatrix} z^1_{k+1} \\ z^2_{k+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z^1_k \\ z^2_k \end{pmatrix} \\ y_k = (z^1_k)^{-1/2} v_k \end{cases}$$

since these three systems have in common the parameters equation:

$$\begin{pmatrix} a^1_{k+1} \\ a^2_{k+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a^1_k \\ a^2_k \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} y^2_{k+1}$$

with $R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $M = (0,0)$, $\Lambda = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $N(z) = (z^1)^{-1/2}$.

Note also that, by corollary 2.1, this equivalence class contains no linear systems since N depends on z. □

Remark 2.13: It may be noted that there is *no remarkable simplification* concerning the *practical* computations of a filter for a system that can be immersed in a *linear system* Σ':

Firstly, if θ is the corresponding immersion from Σ to Σ', and if $dP^\theta_o = \theta(P_o\ dx)$ is the image by θ of P_o, there is no reason why dP^θ_o would have a density on R^r, and a fortiori a gaussian density. Thus, the Kalman filter formulas cannot be used.

Secondly, θ is generally not a diffeomorphism and it is thus quite difficult, at least numerically, to recover $P_k(.|Y_k)$ from $dP^\theta_k(.|Y_k)$ the image of $P_k(.|Y_k)dx$ by θ. □

III - APPLICATION : TRACKING FOR A MOVING TARGET

A moving target is observed through an optical system giving a noisy measurement of the inverse of the angular velocity of the target The target's linear velocity V is supposed to be constant and known. We want to estimate the *initial distance* L_o from the system to the target and the *nodal distance* d (shortest distance from the system to the target's trajectory). The distances L_o and d are displayed in the following figure:

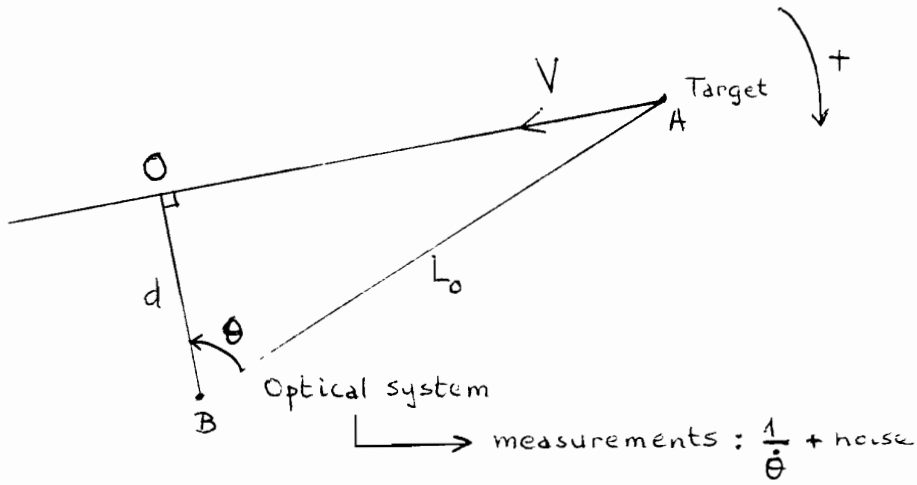


Figure 3

An appropriate choice of coordinates consists in taking the origin at the point O, the x^1-axis supported by OA, and the x^2-axis supported by OB. We shall denote $\|OA\| = x^1$, $\|OB\| = x^2 = d$.

In these coordinates, the evolution of (x^1, x^2) is: $\dot{x}^1 = -V$, $\dot{x}^2 = 0$. The observation equation being $y(t) = 1/\theta(t) + \text{noise}$, since $\operatorname{tg}\theta = x^1/x^2$, we have: $-\dot{\theta}(t)(1+\operatorname{tg}^2\theta(t)) = -V/x^2$, or $y(t) = ((x^1)^2 + (x^2)^2)/Vx^2 + \text{noise}$. The noises are supposed to be gaussian, stationary, uncorrelated, and the initial vector $(x^1(o), x^2(o))'$ is a random vector whose density P_o is the uniform density on the given rectangle $[\underline{x}^1, \bar{x}^1] \times [\underline{x}^2, \bar{x}^2]$.

Finally, since the measurements are digital, they are obtained in discrete-time with the given time mesh Δt. Thus the problem is the following:

$$(3.1)\quad \begin{cases} x^1_{k+1} = x^1_k - V\,\Delta t \\ x^2_{k+1} = x^2_k \\ y_k = \dfrac{(x^1_k)^2 + (x^2_k)^2}{V\,x^2_k} + \eta v_k \end{cases}$$

$$(3.2)\quad P_o(x^1_o, x^2_o) = \frac{1}{(\bar{x}^1 - \underline{x}^1)(\bar{x}^2 - \underline{x}^2)}\, 1_{[\underline{x}^1, \bar{x}^1]\times[\underline{x}^2, \bar{x}^2]}(x^1_o, x^2_o)\ ,$$

and the filter may be used to compute:

$$(3.3)\quad E(d|Y_k) = E(x^2|Y_k) \quad \text{and} \quad E(L_o|Y_k) = E((x^1_o)^2 + (x^2)^2)^{1/2}|Y_k).$$

It can be easily checked that dim Span $H = 4 < +\infty$.

A numerically well conditioned choice of the canonical basis is:

$$(3.4)\qquad \theta_1 \equiv 1,\ \theta_2 = \frac{1}{V}\left(\frac{(x^1)^2}{x^2} + x^2\right),\ \theta_3 = \Delta t\,\frac{x^1}{x^2},\ \theta_4 = V(\Delta t)^2\,\frac{1}{x^2}.$$

We immediately obtain from (2.8) to (2.12) that:

$$(3.5)\qquad \theta of = R\theta \quad \text{with} \quad R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix},\quad R'^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

and:

$$(3.6)\qquad M = \left(0,\ \frac{1}{\eta^2},\ 0,\ 0\right),\qquad \Lambda = \begin{pmatrix} \frac{1}{\eta^2} \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus, the minimal filter is:

$$(3.7)\qquad \begin{pmatrix} a^1_{k+1} \\ a^2_{k+1} \\ a^3_{k+1} \\ a^4_{k+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} a^1_k \\ a^2_k \\ a^3_k \\ a^4_k \end{pmatrix} + \begin{pmatrix} -\frac{1}{2\eta^2}\,y^2_{k+1} \\ \frac{1}{\eta^2}\,y_{k+1} \\ 0 \\ 0 \end{pmatrix} \qquad \begin{aligned} a^1_o &= 0 \\ a^2_o &= 0 \\ a^3_o &= 0 \\ a^4_o &= 0 \end{aligned}$$

and one can check that:

$$(3.8)\qquad P_k(x^1,x^2|Y_k) = \frac{1}{\eta k}\, e^{-\frac{k}{2\eta^2}\left[\theta_2^2+2(k-1)\theta_2\theta_3+(k-1)(2k-1)(\theta_2\theta_4-(\Delta t)^2)\right]} \cdot e^{\left[k(k-1)^2\theta_3\theta_4 + \frac{(k-1)(6k^3-9k^2+k+1)}{30}\theta_4^2\right]} \cdot e^{(a^1_k\theta_1+\ldots+a^4_k\theta_4)} \cdot P_o(x^1+kV\Delta t, x^2).$$

Thus, the filtering numerical procedure consists in computing (3.7) in real time and to tabulate $E(x^2|Y_k)$ and $E(L_o|Y_k)$ as functions of (a^1,a^2,a^3,a^4,k). At each step k, one observes the new y_k and, by (3.7), we deduce the new value of (a^1,a^2,a^3,a^4) which, in turn, is used to find in a memorized table the new values of $E(d|Y_k)$, $E(L_o|Y_k)$. This algorithm is very fast and can easily be implemented on a microcomputer.

The efficiency of the filter is shown in Figure 4. However, this filter is very sensitive to the initial measure P_o and can be improved by adapting the support of P_o from time to time in order to obtain a more precise estimate.

It msut be noted that this example is in the class of systems that can be immersed into a linear system (η = constant). However, the reader can check that it is more difficult to take this property into account by computing a linear filter for the immersed system and then pull it back to recover $E(d|Y_k)$ and $E(L_o|Y_k)$, than to use a nonlinear filter. A complete discussion on the modelling of this example and comparisons between the various corresponding filters can be found in [28].

Remark finally that if, in place of $y = 1/\dot{\theta}$ + noise, we had: $y = \dot{\theta}$ + noise, the problem would be infinite dimensional. □

IV - ON THE NUMBER OF SYSTEMS HAVING A FINITE DIMENSIONAL FILTER

We shall deal, in this section, with the following problem:

(4.1) Given X and the observation equation $y = h(x)+\eta(x)v$, with v gaussian, given the canonical space Span H, with dim Span $H = r < +\infty$ and with a basis $\{\theta_1,\ldots,\theta_r\}$, can we build the set of dynamics f that are local C^1-diffeomorphisms of X, such that the resulting system (X,f,h,η,R^p) has an r-dimensional minimal filter, and what is the cardinality of this set of dynamics ?

Roughly speaking, given the observation equation and a canonical basis, are there many f that satisfy the condition (2.8) that ensure that the minimal filter is r-dimensional? We shall give a partial answer under the following additional assumptions:

(H2)' $(\eta(x)\eta'(x))^{-1}$ is everywhere defined in X and of class C^1 from X to R^{pxp}, and $h \in C^1(X;R^p)$. □

(H4) with the notations (2.4), $H_{ij},H_j \in$ Span $\{\theta_1,\ldots,\theta_r\}$, $\forall 1 \leq i \leq j \leq p$. □

(H5) Noting $\theta(x) = (\theta_1(x),\ldots,\theta_r(x))'$, $\theta \in C^1(X;R^r)$ and: $(X,\theta(X),\theta)$ is a covering of $\theta(X)$ (see for ex. [7],[16]), $\theta(X)$ is connected and simply connected. □

Theorem 4.1: If (H2)', (H4), (H5) hold, for each $R \in GL(r,R)$ such that $R\theta(X) = \theta(X)$, there exists at least a function f_R: $X \to X$ which is a local C^1-diffeomorphism such that:

(4.2) $\theta \circ f_R = R\theta$ on X.

Furthermore, since $\gamma = \text{Card}\ \theta^{-1}(\theta(x))$ is constant $\forall x \in X$, there are γ possible choices of f_R for each R. Consequently, the set of f solving (4.1) has at least the cardinality $\gamma\, \text{Card}\ \mathcal{R}_\theta$, with:

$$\mathcal{R}_\theta = \{ R \in GL(r,\mathcal{R}) \mid R\theta(X) = \theta(X) \}. \tag{4.3}$$

Proof: Using the monodromy principle ([7], [16]), and assuming that $R \in \mathcal{R}_\theta$, there exists a unique C^1-lifting $g_R : \theta(X) \to X$ of the linear mapping $y \to Ry$, satisfying $g_R(\xi) = x_o$, for an arbitrary $\xi \in \theta(X)$ and with $x_o \in \theta^{-1}(R\xi)$. Otherwise speaking, there exists a unique C^1 function g_R such that:

$$g_R(\xi) = x_o \quad , \quad \theta(g_R(y)) = Ry \qquad \forall y \in \theta(X) \tag{4.3}$$

Thus, denoting:

$$f_R(x) = g_R(\theta(x)) \qquad \forall x \in X, \tag{4.4}$$

we obtain that $f \in C^1(X;X)$, and, with (4.3):

$$\theta(f_R(x)) = R\theta(x) \qquad \forall x \in X. \tag{4.5}$$

Thus, to prove that f_R solves (4.1), it suffices to show that f_R is a local C^1-diffeomorphism of X, since, by the theorem 2.1, f_R is such that $(X, f_R, h, \eta, \mathcal{R}^p)$ has a minimal filter of dimension r.

Since R^{-1} exists and satisfies $R^{-1}\theta(X) = \theta(X)$ for $R \in \mathcal{R}_\theta$, let us denote $g_{R^{-1}}$ the unique C^1-lifting of $y \to R^{-1}y$ satisfying:

$$g_{R^{-1}}(R^2\xi) = x_o \quad , \quad \theta(g_{R^{-1}}(y)) = R^{-1}y \qquad \forall y \in \theta(X), \tag{4.6}$$

x_o and ξ being defined as in (4.3).

Also, noting:

$$f_{R^{-1}} = g_{R^{-1}} \circ \theta \tag{4.7}$$

we have:

$$\theta(f_{R^{-1}}(x)) = R^{-1}\theta(x) \qquad \forall x \in X. \tag{4.8}$$

Furthermore, using (4.4) and (4.7), it is easy to check that:

$$\theta \circ (f_{R^{-1}} \circ g_R)(y) = (\theta \circ g_{R^{-1}}) \circ (\theta \circ g_R)(y) = R^{-1}R\, y = y \qquad \forall y \in \theta(X) \tag{4.9}$$

$$\theta \circ (f_R \circ g_{R^{-1}})(y) = (\theta \circ g_R) \circ (\theta \circ g_{R^{-1}})(y) = RR^{-1}\, y = y \qquad \forall y \in \theta(X) \tag{4.10}$$

and, by (4.3) and (4.6):

$$(4.11)\qquad (f_{R^{-1}} \circ g_R)(R\xi) = g_{R-1} \circ (\theta \circ g_R)(R\xi) = g_{R-1}(R^2\xi) = x_o$$

$$(4.12)\qquad (f_R \circ g_{R-1})(R\xi) = g_R \circ (\theta \circ g_{R-1})(R\xi) = g_R(\xi) = x_o.$$

Consequently, (4.9) and (4.10) mean that $f_{R^{-1}} \circ g_R$ and $f_R \circ g_{R-1}$ are two sections of θ, and, by (4.11), (4.12), they coincide at the point $R\xi$. Thus, using (H5), it is a well-known result (see [7], [16]) that two sections that coincide at one point are everywhere equal; thus:

$$(4.13)\qquad f_{R-1} \circ g_R = f_R \circ g_{R-1} \quad \text{on} \quad X.$$

But using once more the covering property, there is a neighborhood V of $R\xi$ in $\theta(X)$ and a neighborhood U of x_o in X such that:

$$(4.14)\qquad x = (f_{R^{-1}} \circ g_R)(\theta(x)) = (f_{R-1} \circ f_R)(x) = (f_R \circ g_{R-1})(\theta(x)) = (f_R \circ f_{R-1})(x)$$

$\forall x \in U$, and it results that

$$(4.15)\qquad (f_{R^{-1}} \circ f_R)\big|_U = (f_R \circ f_{R^{-1}})\big|_U = id_U$$

and, by a classical homotopy argument, f_R is a local C^1-homeomorphism of X with local inverse $f_{R-1} = (f_R)^{-1}$. Thus, since f_{R-1} is also C^1, we have proved that f_R is a local C^1-diffeomorphism satisfying (4.5) and thus, solving the problem (4.1).

Finally, let $\gamma(x) = \text{Card}\ \theta^{-1}(\theta(x))$. By (H5), we have that $\gamma(x) \equiv \gamma \geq 1 \quad \forall x \in X$ (see [7],[16]), and, since there are γ possible choices of $x_o \in \theta^{-1}(R\xi)$, using once more the monodromy principle, there exists exactly γ possible choices of g_R and thus of f_R, which achieves the proof. □

V - CONCLUDING REMARKS

The work presented here solves in great details the finite dimensional filtering problem for a class of discrete-time nonlinear systems without noise in the dynamics and with gaussian observations' noise. We give a complete characterization of the systems having a finite dimensional filter as well in terms of the canonical space as in terms of subordination of a system having linear dynamics (system of the filtering class) to our original system. This efficient characterization allows to obtain the general equations of the minimal filter, and of course, its dimension.

On the other hand, our analysis is entirely based on the properties of exponentials of polynomials and seems to be difficult to generalize to other kinds of noises. Nevertheless, in the particular case of bounded noises with uniform density, the same kind of method can be extended because of the elementary properties of the indicator functions, and will be developed in a forthcoming paper.

But, in the general noises case, another idea can be explored. Coming back to Proposition 1.1, it is both necessary and sufficient for a finite dimensional filter $a_{k+1} = \Phi(a_k, y_{k+1})$ to exist, that f and Φ satisfy the equation:

$$(5.1) \qquad \rho(\Phi(a_k, y_{k+1}), f(x)) = V(\eta^{-1}(f(x))(y_{k+1} - h(f(x)))) \; \rho(a_k, x)$$

that can be considered as an implicit function problem for f and Φ. But f and Φ are submitted to the constraints:

$$(5.2) \qquad \begin{cases} f \text{ must be independent of } a, y \\ \Phi \text{ must be independent of } x \end{cases}$$

and the only method known by the authors to carry on such contraints consists in differentiating (5.1), (5.2) with respect to all the variable (x,y,a) and to solve the system of partial differential equations obtained this way, applying the geometric methods of Cartan [4] (for more modern methods see Pommaret [36]). However the computations are very huge and general existence results seem to be out of touch.

Another extension of interest is in the case of noises in the dynamics. But in this case, the equation (1.5), which is a local equation (in the sense that the operators are local), becomes an integral equation, and thus global, because of the noise convolution. And so, the preceding methods fail. This problem is currently widely open.

Acknowledgement

The authors are mostly indebted to Mr Kocher and Huyhn of Sopelem, for introducing them to the tracking problem and for communicating recordings of real experiments. Particular thanks are also due to Prof. I. Kupka of the Fourier Institute who has kindly indicated the assumption (H5) and the algebraic topological methods to solve the problem (3.1) of Section IV.

REFERENCES

[1] V.E. BENES : Exact finite dimensional filters for certain diffusions with nonlinear drifts. Stochastics 5, 1981, p 65-92.

[2] R.W. BROCKETT : Remarks on finite dimensional nonlinear estimation. C. Lobry ed."Analyse des Systèmes". Astérisque 75-76 (SMF) p. 47-56.

[3] J.W. CARLYLE : On the external probability structure of finite-state channels. Int. J. Control, 7, 1964, p. 385-397.

[4] E. CARTAN : Les systèmes différentiels extérieurs et leurs applications géométriques. Hermann. Paris.

[5] M. CHALEYAT-MAUREL, D. MICHEL : Un théorème de non-existence de filtre de dimension finie. CRAS, t. 296, p 933-936, 1983.

[6] S. CHIKTE, J. T. H. LO : Optimal filters for bilinear systems with nilpotent lie algebras. IEEE Trans. AC 24, 6, 1980, 948-953.

[7] J. DIEUDONNE : Eléments d'analyse. T 3. Gauthier-Villars - Paris - 1974.

[8] G.B. DI MASI, W.J. RUNGGALDIER : On measure transformations for combined filtering and parameter estimation in discrete-time. Syst. Cont. Letters 2, 1, 1982, 57-62.

[9] G.B. DI MASI, W.J. RUNGGALDIER : Approximation and bounds for discrete-time nonlinear filtering. Lecture Notes in Control and Information Sciences n° 44. Bensoussan Lions ed. Springer 1982.

[10] P. FAURRE, M. CLERGET, F. GERMAIN: Operateurs rationnels positifs. Dunod Paris - 1978.

[11] M. FLIESS : Series rationnelles positives et processus stochastiques. Ann. Inst. H. Poincaré B, XI-1, 1975, pp. 1-21.

[12] M. FLIESS, I. KUPKA : A finiteness criterion for nonlinear input-output differential systems. SIAM J. Cont. Optimiz. 21, 5, 1983, 721-728.

[13] M. FLIESS : Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices non commutatives. Inventiones Mathematicae 71, 1983, pp. 521-537.

[14] M. FUJISAKI, G. KALLIANPUR, H. KUNITA : Stochastic differential equations for the nonlinear filtering problem. Osaka Journal of Math, 9, 19-40 (1972).

[15] GANTMACHER : Théorie des matrices. t.1, théorie générale. (French translation) Dunod, Paris, 1966.

[16] C. GODBILLON : Eléments de topologie algébrique. Hermann. Paris. 1971.

[17] M. HAZEWINKEL : On deformations, approximations and nonlinear filtering. Systems and Control Letters 1 N°1, 1981, pp. 32-36.

[18] M. HAZEWINKEL, S.I. MARCUS : On lie algebras and finite dimensional filtering, Stochastics 7, 1982, pp. 29-62.

[19] M. HAZEWINKEL, S.I. MARCUS, H.J. SUSSMANN : Non existence of exact finite dimensional filters for conditional statistics of the cubic sensor problem. Syst. Cont. Letters, 3, 331-340, 1983.

[20] A. HELLER : On stochastic processes derived from markov chains. Ann. Math. Stat, 36, 1286-1291, 1965.

[21] B. JAKUBCZYK : Invertible realizations of nonlinear discrete-time systems. Proc. of 1980 Princeton Conf. on Information Sciences and Systems.

[22] R.E. KALMAN, P.L. FALB, M.A. ARBIB : Topics in mathematical systems theory. Mc. Graw Hill. 1969.

[23] P. G. KAMINSKI, A.E. BRYSON Jr., S. F. SCHMIDT : Discrete square root filtering. A survey of current techniques. IEEE, AC, 16, 6, 727-735, 1971.

[24] H. J. KUSHNER : A robust discrete state approximation of the optimal non linear filter for a diffusion, 1979 Stochastics 3, pp. 75-83.

[25] J. L. LAMBLA : Filtrage non-linéaire des processus de diffusion à composantes périodiques. Application à deux exemples de démodulation de phase et d'amplitude. Thèse de Docteur-Ingénieur Université-Paris VI, 1977.

[26] F. LEGLAND : Application de l'équation du filtrage non-linéaire à un problème d'estimation paramétrique. Journées sur le Filtrage non-linéaire (RCP) Toulouse Mars 1981 LAAS-CNRS.

[27] E. L. LEHMANN : Testing statiscal hypothèsis. Wiley, 1959.

[28] J. LEVINE, G. PIGNIE : The finite dimensional filtering problem for a class of nonlinear discrete-time systems. 9th IFAC World Congress. Budapest. 1984.

[29] C. H. LIU, S. I. MARCUS : The Lie algebra structure of a class of finite dimensional nonlinear filters. Hazewinkel ed., "Filterday, Rotterdam 1980", Econometric Institute, Erasmus University, Rotterdam.

[30] S. I. MARCUS, A. S. WILLSKY : Algebraic structure and finite dimensional nonlinear estimation. SIAM J. Math. Anal. 9, 1978, pp. 312-327.

[31] S. I. MARCUS, C. H. LIU, G. L. BLANKENSHIP : Lie algebras and asymptotic expansions for some nonlinear filtering problems to appear.

[32] S. K. MITTER : On the analogy between mathematical problems of nonlinear filtering and quantum physics. Ricerche di Automatica 10 N° 2, pp. 163-216.

[33] D. NORMAND-CYROT : Théorie et pratique des systemes non-linéaires en temps discret. Thèse de Doctorat d'Etat, Université Paris XI Orsay, 1983.

[34] E. PARDOUX : Analyse asymptotique du problème du filtrage non-linéaire avec bruit d'observation à large bande. A. Bensoussan J. L. Lions ed, Proc. 5th Int. Conf. on Analysis and Optimization of Systems, INRIA, Versaille Dec. 82.

[35] G. PICCI : On the Internal Structure of Finite State Stochastic Processes. Proceedings. Conf. Taormina. Italy. Sept 1977.

[36] J. F. POMMARET : Systems of Partial Differential Equations and Lie Pseudogroups. Gordon and Breach, 1978.

[37] E. D. SONTAG : Polynomial response maps. Lecture Notes on Control and Inf. Sciences. 13. Springer. 1979.

[38] H. J. SUSSMANN : Approximate finite dimensional filters for some nonlinear problems. Stochastics, 7, 1982, 183-203.

[39] H. J. SUSSMANN : Existence and uniqueness of minimal realizations of nonlinear systems. Math. Syst. Theory, 10, 1977, 263-284.

[40] H. J. SUSSMANN : Rigorous results on the cubic sensor problem. In stochastic systems. Proc. Intern. Conf. Univ. Oxford 1978, Academic Press, London 1980.

[41] A. S. WILLSKY : On the albegraic structure of certain partially observable finite-state markov processes. Information and Control, 38, 1978, 179-212.

[42] M. ZAKAI : On the optimal filtering of diffusion processes. Z. Walr. Verw. geB., 11, 1969, 230-243.

APPENDIX 1

Proof of Theorem 1.1

Theorem 1.1: The unnormalized conditional law of x_k knowing y_k has a density $P_k(.|Y_k) \in C^\circ(\tilde{X};R_+)$ with respect to μ_o, with $\tilde{X} = X - \bigcup_{j\geq o} (\det\eta \circ f^{-j})^{-j}(0)$, and $P_k(.|Y_k)$ is given by:

$$(1.5) \qquad P_k(x|Y_k) = |\det\eta(x)|^{-1} \, V(\eta^{-1}(x)(y_k - h(x))) \, J_f^{-1}(f^{-1}(x)) \cdot P_{k-1}(f^{-1}(x)|Y_{k-1})$$

$$\forall k \geq 1 \ , \ \forall x \in \tilde{X}$$

$$P_o(x|Y_o) = P_o(x),$$

where $J_f(x)$ denotes the Jacobian of f relatively to X, evaluated at the point $x \in X$.

Proof: P_o is supposed to be density, $P_o \in C^\circ(X;R_+)$.
Let us suppose that $P_{k-1}(.|Y_{k-1})$ is a $C^\circ(X;R_+)$ density with respect to μ_o, with $\tilde{X} = X - \bigcup_{j\geq o}(\det\eta \circ f^{-j})^{-1}(0)$, and let us prove the same property for $P_k(.|Y_k)$.

Let us denote $\tilde{P}_k(.|Y_{k-1})$ the conditional measure of x_k knowing Y_{k-1} and let us compute its density with respect to μ_o.
If $\phi \in C^\circ_K(X)$ (the space of continuous functions with compact support in X), we have, by definition of the image of the measure $\tilde{P}_{k-1}(.|Y_{k-1}) = P_{k-1}(.|Y_{k-1})d\mu_o$ by f:

$$(A1.1) \qquad E(\psi(x_k)|Y_{k-1}) = \int \psi(f(\xi))P_{k-1}(\xi|Y_{k-1})d\mu_o(\xi).$$

Considering an atlas of X with local charts, noted (U_α, Φ_α), $\alpha \in N$, such that each U_α is relatively compact, the collection $\{U_\alpha\}$ being a locally finite covering, and considering $\{F_\alpha\}_{\alpha\in N}$ a partition of unity associated to U_α, each F_α having its support in U_α, one can define μ_o as follows:

$$(A1.2) \qquad \int \theta(x)\, d\mu_o(x) = \sum_\alpha \int_{\Phi_\alpha(U_\alpha)} \theta(\Phi_\alpha^{-1}(\xi)) \, F_\alpha(\Phi_\alpha^{-1}(\xi)) \, d\lambda(\xi)$$

$\forall \theta \in C^\circ_K(X)$, where λ is the Lebesgue measure of R^n, $\Phi_\alpha(U_\alpha)$ being an open subset of R^n.

Then, putting together (A1.1) and (A1.2):

$$(A1.3) \quad E(\psi(x_k)|Y_{k-1}) = \sum_{\alpha\in N} \int_{\Phi_\alpha(U_\alpha)} \psi(f(\Phi_\alpha^{-1}(\xi)))P_{k-1}(\Phi_\alpha^{-1}(\xi)|Y_{k-1}) \cdot F_\alpha(\Phi_\alpha^{-1}(\xi))d\lambda(\xi).$$

Let us now make the change of variables:

$$(A1.4) \qquad z = \Phi_\beta(f(\Phi_\alpha^{-1}(\xi))) \quad \text{on } \Phi_\alpha(U_\alpha \cap f^{-1}(U_\beta)) \quad , \quad \forall\, \alpha,\beta \in N$$

The application $\Phi_\beta \circ f \circ \Phi_\alpha^{-1}$ being a C^1-diffeomorphism from $\Phi_\alpha(U_\alpha \cap f^{-1}(U_\beta))$ to $\Phi_\beta(U_\beta \cap f(U_\alpha))$, we obtain with (A1.3):

$$E(\psi(x_k)|Y_{k-1}) = \sum_\beta \sum_\alpha \int_{\Phi_\beta(U_\beta \cap f(U_\alpha))} \psi(\Phi_\beta^{-1}(z)) P_{k-1}(f^{-1}(\Phi_\beta^{-1}(z))|Y_{k-1}) \cdot$$

$$\cdot F_\alpha(f^{-1}(\Phi_\beta^{-1}(z))) \, \tilde{J}^{-1}_{\Phi_\beta \circ f \circ \Phi_\alpha^{-1}}(\Phi_\alpha(f^{-1}(\Phi_\beta^{-1}(z)))) d\lambda(z)$$

where $\tilde{J}_{\Phi_\beta \circ f \circ \Phi_\alpha^{-1}}$ is the usual Jacobian on R^n of $\Phi_\beta \circ f \circ \Phi_\alpha^{-1}$, and summing in α:

$$(A1.5) \qquad E(\psi(x_k)|Y_{k-1}) = \sum_\beta \int_{\Phi_\beta(U_\beta)} \psi(\Phi_\beta^{-1}(z)) P_{k-1}(f^{-1}(\Phi_\beta^{-1}(z))|Y_{k-1}) \cdot$$

$$F_\beta(\Phi_\beta^{-1}(z)) \cdot J_f^{-1}(f^{-1}(\Phi_\beta^{-1}(z))) d\lambda(z)$$

where $J_f^{-1}(x)$ is the Jacobian of f relatively to X defined by:

$$(A1.6) \qquad \sum_\alpha F_\alpha(f^{-1}(x)) \, \tilde{J}^{-1}_{\Phi_\beta \circ f \circ \Phi_\alpha^{-1}}(\Phi_\alpha(f^{-1}(x))) = F_\beta(x) J_f^{-1}(f^{-1}(x)) \ ,$$

$$\forall x \in \text{Supp } F_\beta \subset U_\beta \ , \ \forall_\beta \ .$$

Thus, by definition of μ_o (see (A1.2)), (A1.5) becomes:

$$(A1.7) \qquad E(\psi(x_k)|Y_{k-1}) = \int_X \psi(x) P_{k-1}(f^{-1}(x)|Y_{k-1}) \, J_f^{-1}(f^{-1}(x)) d\mu_o(x)$$

which proves that $\tilde{P}_k(.|Y_{k-1})$ has a C° density with respect to μ_o:

$$(A1.8) \qquad \tilde{P}_k(.|Y_{k-1}) = (P_{k-1}(f^{-1}(.)|Y_{k-1}) \, J_f^{-1}(f^{-1}(.)))\mu_o \ ,$$

or, noting $P_k(.|Y_{k-1})$ the density of $\tilde{P}_k(.|Y_{k-1})$:

$$(A1.9) \qquad P_k(x|Y_{k-1}) = J_f^{-1}(f^{-1}(x)) \, P_{k-1}(f^{-1}(x)|Y_{k-1}).$$

Now, it remains to compute $\tilde{P}_k(.|Y_k)$ the conditional measure of x_k knowing Y_k, and to prove that it has a C° density on $\tilde{X}$. For this purpose, if Q_k is the conditional measure of (x_k,y_k) knowing Y_{k-1}, and if $\psi \in C_K^\circ(X \times R^p; R)$, we have:

$$(A1.10) \qquad \int \psi(x,y) dQ_k(x,y) = \int_{X \times R^p} \psi(x, h(x)+\eta(x)v) V(v) dv d\tilde{P}_k(x|Y_{k-1})$$

v_k and x_k being independent by construction.

Also, since $(\det\eta)^{-1}(0)$ has μ_o measure 0, $\tilde{P}_k((\det\eta)^{-1}(0)|Y_{k-1}) = 0$ by (A1.9), and, by Fubini:

$$\text{(A1.11)} \quad \int \psi(x,y)dQ_k(x,y) = \int_{X-(\det\eta)^{-1}(0)} \left(\int_{R^p} \psi(x,h(x)+\eta(x)v)V(v)dv\right)d\tilde{P}_k(x|Y_{k-1}$$

$$= \int_{X-(\det\eta)^{-1}(0)} \left(\int_{R^p} \psi(x,z)V(\eta^{-1}(x)(z-h(x)))|\det\eta(x)|^{-1}dz\right)d\tilde{P}_k(x|Y_{k-1})$$

by the change of variables in R^p: $z = h(x)+\eta(x)v$, and thus:

$$\text{(A1.12)} \quad dQ_k(x,y) = V(\eta^{-1}(x)(y-h(x)))|\det\eta(x)|^{-1}P_k(x|Y_{k-1})d\mu_o(x)dy.$$

Finally, the unnormalized conditional measure $\tilde{P}_k(.|Y_k)$ is simply:

$$\text{(A1.13)} \quad \tilde{P}_k(.|Y_k) = V(\eta^{-1}(x)(y_k-h(x)))|\det\eta(x)|^{-1} P_k(x|Y_{k-1})d\mu_o(x),$$

and its density with respect to μ_o is C^∞ on $\tilde{X}$, given by:

$$\text{(A1.14)} \quad P_k(x|Y_k) = |\det\eta(x)|^{-1}V(\eta^{-1}(x)(y_k-h(x))) J_f^{-1}(f^{-1}(x))P_{k-1}(f^{-1}(x)|Y_{k-1})$$

which is the desired result. □

APPENDIX 2

Proof of Theorem 2.1 (Fundamental theorem)

Theorem 2.1: The following properties are equivalent:

(i) the system (1.1) admits a finite dimensional filter (resp. local)

(ii) Span H is finite dimensional (resp. $\exists$ U open subset of $\tilde{X}$ such that Span $H_{|U}$ is finite dimensional)

(iii) $\exists\, r \in N$, $\exists \theta_1,\ldots,\theta_r \in$ Span H, $\exists$ R invertible$_\sim(r,r)$ matrix with constant real coefficients such that, on X (resp. on U):

$$(2.8)\quad \begin{cases} \theta o f = R\theta \quad \text{with} \quad \theta = (\theta_1,\ldots,\theta_r)', \\ H_\alpha = \sum_{i=1}^{r} \mu_{\alpha,i}\, \theta_i \ , \quad H_{\alpha,\beta} = \sum_{i=1}^{r} \mu_{\alpha,\beta,i}\, \theta_i \ , \ \forall\, \alpha,\beta = 1,\ldots,p \end{cases}$$

Proof: 1°)(i) => (ii)

Let us suppose that (1.1) admits a finite dimensional filter of dimension r given by:

$$(A2.1)\quad \begin{cases} a_{k+1} = \Phi_k(a_k, y_{k+1}) \\ \tilde{\pi}_k(x|Y_k) = \rho_k(a_k, x) \end{cases}$$

with A r-dimensional C^1 manifold, $\Phi_k \in C^1(A \times R^p; A)$, $\sigma_k \in C^1(A; C^\circ(\tilde{X}; R_+))$ (in the local case, the only change consists in replacing $C^1(A; C^\circ(\tilde{X}; R_+))$ by $C^1(A; C^\circ(U_k; R_+))$ with $U_k = f^k(U)$).

Since we have:

$$(A2.2)\quad \tilde{\pi}_k(x|Y_k) = \rho_k(\Phi_{k-1}(\Phi_{k-2}(\ldots(\Phi_o(a_o, y_1), y_2), \ldots, y_{k-1}), y_k), x)$$

and since ρ_k is differentiable with respect to a and Φ_α is differentiable with respect to $y_{\alpha+1}$ $\forall\, \alpha = 0,\ldots,k$, $\forall k$, then $\tilde{\pi}_k$ is differentiable with respect to $y_1,\ldots,y_k$, and:

$$(A2.3)\quad \frac{\partial \tilde{\pi}_k}{\partial y_j^i}(x|Y_k) = \sum_{i_o, i_1, \ldots, i_{k-j}=1}^{r} \frac{\partial \rho_k}{\partial a^{i_o}}(a_k, x)\, \frac{\partial \Phi_{k-1}^{i_o}}{\partial a^{i_1}}(a_{k-1}, y_k) \ \ldots$$

$$\ldots\, \frac{\partial \Phi_{j-1}^{i_{k-j}}}{\partial y_j^i}(a_{j-1}, y_j) = \sum_{i_o=1}^{r} \lambda_{i,j,k}^{i_o}(a_{j-1}, y_j, \ldots, y_k)\, \frac{\partial \rho_k}{\partial a^{i_o}}(a_k, x)$$

with:

$$(A2.4)\qquad \lambda^{i_o}_{i,j,k}(a_{j-1},y_j,\ldots,y_k) = \sum_{i_1,\ldots,i_{k-j}=1}^{r} \frac{\partial \Phi^{i_o}_{k-1}}{\partial a^{i_1}}(a_{k-1},y_k)\ldots \cdots \frac{\partial \Phi^{i_{k-j}}_{j-1}}{\partial y^{i}_{j}}(a_{j-1},y_j).$$

It follows that the Span of $\{ \frac{\partial \tilde{\pi}_k}{\partial y^i_j}(.|Y_k) \mid i=1,\ldots,p,\ j=1,\ldots,k \}$ is generated by the r functions:

$$(A2.5)\qquad \frac{\partial \rho_k}{\partial a^1}(a_k,.),\ldots,\frac{\partial \rho_k}{\partial a^r}(a_k,o)\ ,$$

and thus:

$$(A2.6)\qquad \operatorname*{Sup}_{\substack{k\geq 1\\ Y_k}} \dim \operatorname{Span}\{\frac{\partial \tilde{\pi}_k}{\partial y^i_j}(.|Y_k) \mid i=1,\ldots,p,\ \ j=1,\ldots,k\} \leq r.$$

On the other hand, using (2.7), we have:

$$(A2.7)\qquad \tilde{\pi}_k(x|Y_k) = \exp -\frac{1}{2}\left[\sum_{\alpha,\beta=1}^{p}\sum_{\gamma=1}^{k} H_{\alpha,\beta}(f^{\gamma-k}(x))y^{\alpha}_{\gamma}y^{\beta}_{\gamma} - 2\sum_{\alpha=1}^{p}\sum_{\gamma=1}^{k} H_{\alpha}(f^{\gamma-k}(x))y^{\alpha}_{\gamma}\right]$$

and:

$$(A2.8)\qquad \frac{\partial \tilde{\pi}_k}{\partial y^i_j}(x|Y_k) = \left[-\sum_{\alpha=1}^{p} H_{\alpha i}(f^{j-k}(x))y^{\alpha}_j + H_i(f^{j-k}(x))\right]\tilde{\pi}_k(x|Y_k)$$

Then, with (A2.6), it follows that:

$$(A2.9)\qquad \operatorname*{Sup}_{\substack{k\geq 1\\ Y_k}} \dim \operatorname{Span}\left\{-\sum_{\alpha=1}^{p} H_{\alpha i}(f^{j-k}(.))y^{\alpha}_j + H_i(f^{j-k}(.)) \;\middle|\; \begin{array}{l} i=1,\ldots,p,\\ j=1,\ldots,k\end{array}\right\} \leq r$$

Now, let us choose p+1 observation vectors $Y_{1,k},\ldots,Y_{p+1,k}$ with $Y_{m,k}=(y_{m,1},\ldots,y_{m,k})$, $m=1,\ldots,p+1$, $y_{m,j}$ being the observation vector $(y^{\alpha}_{m,j},\ \alpha=1,\ldots,p)$ at time j on the m^{th} trajectory, such that the $(p+1,p+1)$ matrix:

$$(A2.10)\qquad \begin{pmatrix} 1 & -y^1_{1,j} & \cdots\cdots & -y^p_{1,j} \\ \vdots & & & \\ 1 & -y^1_{p+1,j} & \cdots\cdots & -y^p_{p+1,j} \end{pmatrix} \text{ is invertible } \forall\, j = 1,\ldots,k$$

Such a choice is always possible since $\{ h(x)+\eta(x)v \,|\, v \in R^p\} = R^p$, $\forall x \in X$.

For the m^{th} observation trajectory $Y_{m,k}$, noting $a_{m,j-1},\ldots,a_{m,k}$ the corresponding parameters of the filter, we have by (A2.3), (A2.8):

$$(A2.11)\qquad (1\ -y^1_{m,j}\ \ldots\ -y^p_{m,j}) \begin{pmatrix} H_i \text{of}^{j-k} \\ H_{1i} \text{of}^{j-k} \\ \vdots \\ H_{pi} \text{of}^{j-k} \end{pmatrix} =$$

$$= \frac{1}{\pi_k(\cdot \,|\, Y_{m,k})} \sum_{\alpha=1}^{r} \lambda^{\alpha}_{i,j,k}(a_{m,j-1}, Y_{m,k}) \frac{\partial \rho_k}{\partial a^{\alpha}}(a_{m,k},\cdot)$$

$$= \sum_{\alpha=1}^{r} \lambda^{\alpha}_{i,j,k}(a_{m,j-1}, Y_{m,k}) \left(\frac{\partial \rho_k}{\partial a^{\alpha}}(a_{m,k},\cdot) \cdot \frac{1}{\rho_k(a_{m,k},\cdot)} \right)$$

$$\forall i = 1,\ldots,p\ ,\quad \forall j = 1,\ldots,k\ ,\quad \forall m = 1,\ldots,p+1$$

Let $$\tilde{\rho}^{\alpha}_{m,k}(\cdot) = \frac{\partial \rho_k}{\partial a^{\alpha}}(a_{m,k},\cdot) \frac{1}{\rho_k(a_{m,k},\cdot)}$$

Thus:

$$(A2.12)\qquad \begin{pmatrix} 1 & -y^1_{1,j} & \cdots\cdots & -y^p_{1,j} \\ \vdots & & & \\ 1 & -y^1_{p+1,j} & \cdots & -y^p_{p+1,j} \end{pmatrix} \begin{pmatrix} H_i \text{of}^{j-k} \\ H_{1i} \text{of}^{j-k} \\ \vdots \\ H_{pi} \text{of}^{j-k} \end{pmatrix} = \begin{pmatrix} \sum_{\alpha=1}^{r} \lambda^{\alpha}_{i,j,k}(a_{1,j-1}, Y_{1,k})\, \tilde{\rho}^{\alpha}_{1,k}(\cdot) \\ \\ \sum_{\alpha=1}^{r} \lambda^{\alpha}_{i,j,k}(a_{p+1,j-1}, Y_{p+1,k})\, \tilde{\rho}^{\alpha}_{p+1,k}(\cdot) \end{pmatrix}$$

$$\forall i = 1,\ldots,p\ ,\quad \forall j = 1,\ldots,k.$$

But this means that, with (A2.10),

$$(A2.13)\qquad \sup_k \dim \text{Span}\{ H_i \text{of}^{j-k}, H_{\beta i} \text{of}^{j-k} \,|\, i=1,\ldots,p,\ \beta=1,\ldots,p,\ j=1,\ldots,k\}$$
$$\leq \dim \text{Span}\{ \tilde{\rho}^{\alpha}_{m,k}(\cdot) \,|\, \alpha=1,\ldots,r,\ m=1,\ldots,p+1\} \leq (p+1)r$$

Finally, since f is a C^1 diffeomorphism from X to X, it is easy to deduce from (A2.13) that:

(A2.14) $\quad \dim \text{Span}\{H_i \circ f^k,\ H_{ij} \circ f^k \mid i,j = 1,\ldots p,\ k \geq 0\} = \dim \text{Span } H \leq (p+1)r$

wich proves (ii).

2°) <u>(ii) => (iii)</u>

Suppose that dim Span H = r and that $\{\theta_1,\ldots,\theta_r\}$ is a basis of Span H. If (iii) does not hold, this means that there exists at least one $\theta_i \circ f$ that cannot be expressed as a linear combination of $\{\theta_1,\ldots,\theta_r\}$ which is a contradiction since $\theta_i \circ f$ is a linear combination of elements of the form $H_\alpha \circ f^k$, $H_{\alpha,\beta} \circ f^k$, and thus $\theta_i \circ f \in$ Span H .

Thus, we necessarily have $\theta \circ f = R\theta$. With the same argument and since f is a diffeomorphism, one also has $\theta \circ f^{-1} = S\theta$.

Therefore $\quad \theta = \theta \circ f \circ f^{-1} = R\theta \circ f^{-1} = RS\theta$

and $\quad \theta = \theta \circ f^{-1} \circ f = S\theta \circ f = SR\theta$

which proves that $R^{-1} = S$ and thus R is invertible.

Finally, since $\{\theta_1,\ldots,\theta_r\}$ is a basis of Span H, the relations $H_\alpha = \sum_{i=1}^{r} \mu_{\alpha i}\theta_i$, $H_{\alpha\beta} = \sum_{i=1}^{r} \mu_{\alpha\beta i}\theta_i$ are trivial, and (III) is proved.

3°) <u>(iii) => (i)</u>

If $\theta_1,\ldots,\theta_r$ satisfy (2.8), let M be the (p,r) matrix made of the coefficients $\mu_{\alpha i}$, $\alpha = 1,\ldots,p$, $i = 1,\ldots,r$ with:

(A2.15) $\quad H_\alpha = \Sigma\, \mu_{\alpha i}\theta_i$,

and let Λ be the tensor of order 3 defined by:

(A2.16) $\quad y' \Lambda\ y = \begin{pmatrix} \sum_{\alpha,\beta=1}^{p} \lambda^1_{\alpha\beta} y^\alpha y^\beta \\ \vdots \\ \sum_{\alpha,\beta=1}^{p} \lambda^r_{\alpha\beta} y^\alpha y^\beta \end{pmatrix}$

(A2.17) $\quad$ with $H_{\alpha,\beta} = \sum_{i=1}^{r} \lambda^i_{\alpha,\beta}\theta_i$.

Let us also denote $S = R^{-1}$ and S_{ij} the entries of S.

If we set : $\tilde{\pi}_k(x|Y_k) = e^{\sum_{i=1}^{r} a_k^i \theta_i(x)}$, with (2.7), we have:

$$(A2.18) \qquad e^{\sum_{i=1}^{r} a_k^i \theta_i(x) - \frac{1}{2} \sum_{i=1}^{r} \sum_{\alpha,\beta=1}^{p} (\lambda^i_{\alpha\beta} y_k^\alpha y_k^\beta - 2\mu_{\alpha i} y_k^\alpha)\theta_i(x)} = e^{\sum_{i,j=1}^{r} a_{k-1}^i S_{ij} \theta_j(x)} \cdot e^{\sum_{i=1}^{r} a_k^i \theta_i(x)}.$$

or:

$$(A2.19) \qquad a_k^i = \sum_{j=1}^{r} S_{ji} a_{k-1}^j + \sum_{\alpha=1}^{p} \mu_{\alpha i} y_k^\alpha - \frac{1}{2} \sum_{\alpha,\beta=1}^{p} \lambda^i_{\alpha,\beta} y_k^\alpha y_k^\beta \quad , \quad i=1,\ldots,r,$$

or also:

$$(A2.20) \qquad a_k = R'^{-1} a_{k-1} + M' y_k - \frac{1}{2} y_k' \Lambda y_k.$$

Now, denoting $\Phi(a_{k-1}, y_k) = R'^{-1} a_{k-1} + M' y_k - \frac{1}{2} y_k' \Lambda y_k$,

and since $\tilde{\pi}_k(x|Y_k) = \exp \sum_{i=1}^{r} a_k^i \theta_i(x)$, it follows that there exists a finite dimensional filter of dimension r.
Furthermore, this filter is C^ω, invertible and pseudo-stationary. □

Remark A2: For (i) ⟹ (ii), one can prove a sharper result if the filter (A2.1) is supposed invertible: in this case if dimA = r, then dim Span $H \leq r$. Let us prove this fact. The only difference with the preceding proof lies in the choice of $Y_{1,k},\ldots,Y_{p+1,k}$. The idea consists in making profit of the backward transitions (see definitions 2.4 and 2.5) in order to finish always at the same point a_k (and not at $a_{1,k},\ldots,a_{p+1,k}$ as in (A2.12)) so that Span H can be expressed as linear combinations of $\partial\rho_k/\partial a^1(a_k,.)$, i = 1,...,r, implying that dim Span $H \leq r$. Thus to achieve the proof, it only remains to give a suitable choice of $Y_{1,k},\ldots,Y_{p+1,k}$ and to build the corresponding sequences of $a_{m,j}$, m = 1,...,p+1, j = 0,...,k-1.

Let us choose $Y_{1,k},\ldots,Y_{p+1,k}$ so that (A2.10) holds true, fix $a_k \in A$, and set:

$$(A2.21) \qquad a_{m,o} = \Phi_o^{-1}(\Phi_1^{-1}(\ldots(\Phi_{k-1}^{-1}(a_k, y_{m,k}),\ldots,y_{m,2}),y_{m,1}) \quad \forall m = 1,\ldots,p+1.$$

Thus:

$$(A2.22) \qquad \rho_k(a_k,x) = \rho_k(\Phi_{k-1}(\ldots(\Phi_o(a_{m,o},y_{m,1}),\ldots,y_{m,k}),x) \quad \forall m = 1,\ldots,p+1,$$

and (A2.12) becomes, with :

$$\tilde{\rho}_k^\alpha(.) = \frac{\partial \rho_k}{\partial a^\alpha}(a_k,.) \frac{1}{\rho_k(a_k,.)}$$

$$(A.2.23)\quad \begin{pmatrix} 1 & -y^1_{1,j} & \cdots & -y^p_{1,j} \\ & & & \\ 1 & -y^1_{p+1,j} & & -y^p_{p+1,j} \end{pmatrix} \begin{pmatrix} H_i of^{j-k} \\ H_{1i} of^{j-k} \\ \vdots \\ H_{pi} of^{j-k} \end{pmatrix} =$$

$$= \begin{pmatrix} \tilde{\lambda}^1_{i,j,k}(Y_{1,k}) & \cdots & \tilde{\lambda}^r_{i,j,k}(Y_{1,k}) \\ & & \\ \tilde{\lambda}^1_{i,j,k}(Y_{p+1,k}) & \cdots & \tilde{\lambda}^r_{i,j,k}(Y_{p+1,k}) \end{pmatrix} \begin{pmatrix} \tilde{\rho}^1_k(.) \\ \vdots \\ \tilde{\rho}^r_k(.) \end{pmatrix}$$

and thus concluding as before:

Sup dim Span $H \leq r$, which proves the assertion. □
$k>0$
Y_k

As a consequence of this result, one can see that the system (A2.20) is minimal since it is invertible with dimension r, and since dim Span H = r $\leq$ dim A = r.

Also, dim (minimal realization of the filter) = dim Span H . □

APPENDIX 3

Proof of Theorem 2.3

Theorem 2.3: The minimal filter (2.9) is observable and locally weakly reachable.

Proof: 1°) Observability

We have $\rho(a_k,x) = e^{\sum_{i=1}^{r} a_k^i \theta_i(x)}$, $\theta_1,\ldots,\theta_r$ being a basis of Span H. Thus, if for some k, a_k and b_k generated by $Y_k = (y_1,\ldots,y_k)$ from a_o and b_o with $a_o \neq b_o$, we have $\rho(a_k,x) = \rho(b_k,x)$ $\forall x \in X$ (or in U_k), then

$\sum_{i=1}^{r} (a_k^i - b_k^i)\theta_i(x) = 0$. But since $\theta_1,\ldots,\theta_r$ is a basis, this implies that $a_k^i = b_k^i$ $\forall i = 1,\ldots,r$.

But, since the system (2.9) is invertible, this immediately implies that $a_o = b_o$ which is contrary to the assumption. Thus $a_o \neq b_o$ implies $\rho(a_k,.) \neq \rho(b_k,.)$ $\forall k$, and (2.9) is observable.

2°) Weak local reachability

As announced in Remark 2.8, we shall compute dim $T_{a_o}(R_w(a_o))$. For this purpose, we shall proceed as follows: Let us introduce p+1 observation trajectories of length k (k is given) depending on a vector of perturbation $\varepsilon = \{(\varepsilon_{1,1},\ldots,\varepsilon_{1,k}),\ldots,(\varepsilon_{p+1,1},\ldots,\varepsilon_{p+1,k})\}$ with $\varepsilon_{m,j} \in R^p$ $\forall m = 1,\ldots,p+1$, $\forall j = 1,\ldots,k$, and:

$Y_{m,k}(\varepsilon) = (y_{m,1}+\varepsilon_{m,1},\ldots,y_{m,k}+\varepsilon_{m,k})$ $\forall m = 1,\ldots,p+1$.

We also suppose that the vectors $y_{m,j}$ satisfy (A2.10) $\forall j = 1,\ldots,k$, and we denote $Y_{m,k}(0)$ the sequence where $\varepsilon_{m,j} = 0$ $\forall j = 1,\ldots,k$.

To generate $T_{a_o}(R_w(a_o))$, we shall make:

- an onward transition of length k from a_o with $Y_{1,k}(\varepsilon)$. The endpoint will be noted $a_{1,k}(\varepsilon)$.
- a backward transition of length k from $a_{1,k}(\varepsilon)$ with $Y_{1,k}(0)$. The endpoint will be noted $a_{1,0}(\varepsilon)$.
- an onward transition of length k from $a_{1,0}(\varepsilon)$ with $Y_{2,k}(\varepsilon)$. The endpoint will be noted $a_{2,k}(\varepsilon)$.
- a backward transition of length k from $a_{2,k}(\varepsilon)$ with $Y_{2,k}(0)$. The endpoint will be noted $a_{2,0}(\varepsilon)$.

... and so on, up to $a_{p+1,0}(\varepsilon)$.

Since $a_{k+1} = R'^{-1} a_k + M' y_k - \frac{1}{2} y_k' \Lambda y_k$, it is easy to prove that:

$$(A3.1) \qquad a_{1,k}(\ ) = R'^{-k} a_o + \sum_{j=o}^{k-1} R'^{-j}(M' - \tfrac{1}{2}(y'_{1,k-j} + \varepsilon'_{1,k-j})\Lambda)(y_{1,k-j} + \varepsilon_{1,k-j})$$

and that:

$$(A3.2) \qquad a_{1,0}(\varepsilon) = a_o + \sum_{j=o}^{k-1} R'^{k-j}(M' - y'_{1,k-j}\ \Lambda)\ \varepsilon_{1,k-j} + o(\|\varepsilon_1\|^2)$$

Also, the reader can check that:

$$(A3.3) \qquad a_{p+1,o}(\varepsilon) - a_o = \sum_{m=1}^{p+1} \sum_{j=o}^{k-1} R'^{k-j}\ (M' - y'_{m,k-j}\ \Lambda)\varepsilon_{m,k-j} + o(\sum_{m=1}^{p+1} \|\varepsilon_m\|^2)$$

Consequently:

$$(A3.4) \qquad \frac{\partial a_o^i(\ )}{\partial \varepsilon^{\alpha}_{m,k-j}}\Big|_{\varepsilon=o} = (R'^{k-j}M')_{i,\alpha} - \sum_{u=1}^{r} \sum_{\beta=1}^{p} (R'^{k-j})_{iu} \Lambda^{u}_{\alpha\beta}\, y^{\beta}_{m,k-j}\ ,$$

$\forall i = 1,\dots,r,\ \forall \alpha = 1,\dots,p,\ \forall m = 1,\dots,p+1,\ \forall j = o,\dots,k-1,\ \forall k \geq 1,$
with the notation $A_{i,\alpha}$ for the coefficient of the line i and column α of the matrix A.

Thus, denoting:

$$(A3.5) \qquad B(a_o) = \text{Span}\ \{ \frac{\partial a_o(\varepsilon)}{\partial \varepsilon^{\alpha}_{m,k-j}}\Big|_{\varepsilon=o}\ ,$$

$$\alpha = 1,\dots,p\ ;\ m = 1,\dots,p+1\ ;\ j = 0,\dots,k-1\ ;\ k \geq 1\} ,$$

we have $B(a_o) \subset T_{a_o}(\mathcal{R}_w(a_o))$ by construction, and, using (A3.4):

$$(A3.6) \qquad \begin{pmatrix} \frac{\partial a_o^i(\varepsilon)}{\partial \varepsilon^{\alpha}_{1,k-j}}\Big|_{\varepsilon=o} \\ \vdots \\ \frac{\partial a_o^i(\varepsilon)}{\partial \varepsilon^{\alpha}_{p+1,k-j}}\Big|_{\varepsilon=o} \end{pmatrix} = \begin{pmatrix} 1 & -y^1_{1,k-j} & \cdots & -y^p_{1,k-j} \\ & & & \\ 1 & -y^1_{p+1,k-j} & \cdots & -y^p_{p+1,k-j} \end{pmatrix} \begin{pmatrix} (R'^{k-j}M')_{i\alpha} \\ \sum_{u=1}^{r} (R'^{k-j})_{iu} \Lambda^{u}_{1,\alpha} \\ \vdots \\ \sum_{u=1}^{r} (R'^{k-j})_{iu} \Lambda^{u}_{p,\alpha} \end{pmatrix}$$

$\forall i = 1,\dots,r,\ \forall \alpha = 1,\dots,p,\ \forall j = 0,\dots,k-1,\ \forall k \geq 1,$

and since (A2.10) holds, we have:

$$(A3.7) \qquad B(a_o) = \text{Span}\ \{ (R'^{k-j}M')_{\cdot,\alpha},\ \sum_{u=1}^{r} (R'^{k-j})_{\cdot,u}\, \Lambda^{u}_{\alpha,\beta} \ \Big|$$

$$\alpha = 1,\ldots,p\ ;\ j = 0,\ldots,k-1\ ;\ k \geq 1\ \}$$

But, since by definition:

$H = M\theta$ and $H_{\alpha\beta} = \Lambda_{\alpha\beta}\theta \quad \alpha,\beta = 1,\ldots,p$, taking into account the fact that $\theta o f = R\theta$, we have:

$$(A3.8) \qquad \sum_{i=1}^{r} \theta_i (R'^{k-j}M')_{i,\alpha} = (H_\alpha o f^{k-j})', \quad \sum_{i=1}^{r} \theta_i (R'^{k-j})_{iu}\, \Lambda^{u}_{\alpha,\beta} = (H_{\alpha,\beta} o f^{k-j})',$$

$$\forall\ \alpha,\beta = 1,\ldots,p, \quad \forall j = 0,\ldots,k-1, \quad \forall k.$$

Thus, immediately, this yields: $\dim B(a_o) = \dim \text{Span}\ H$, and:

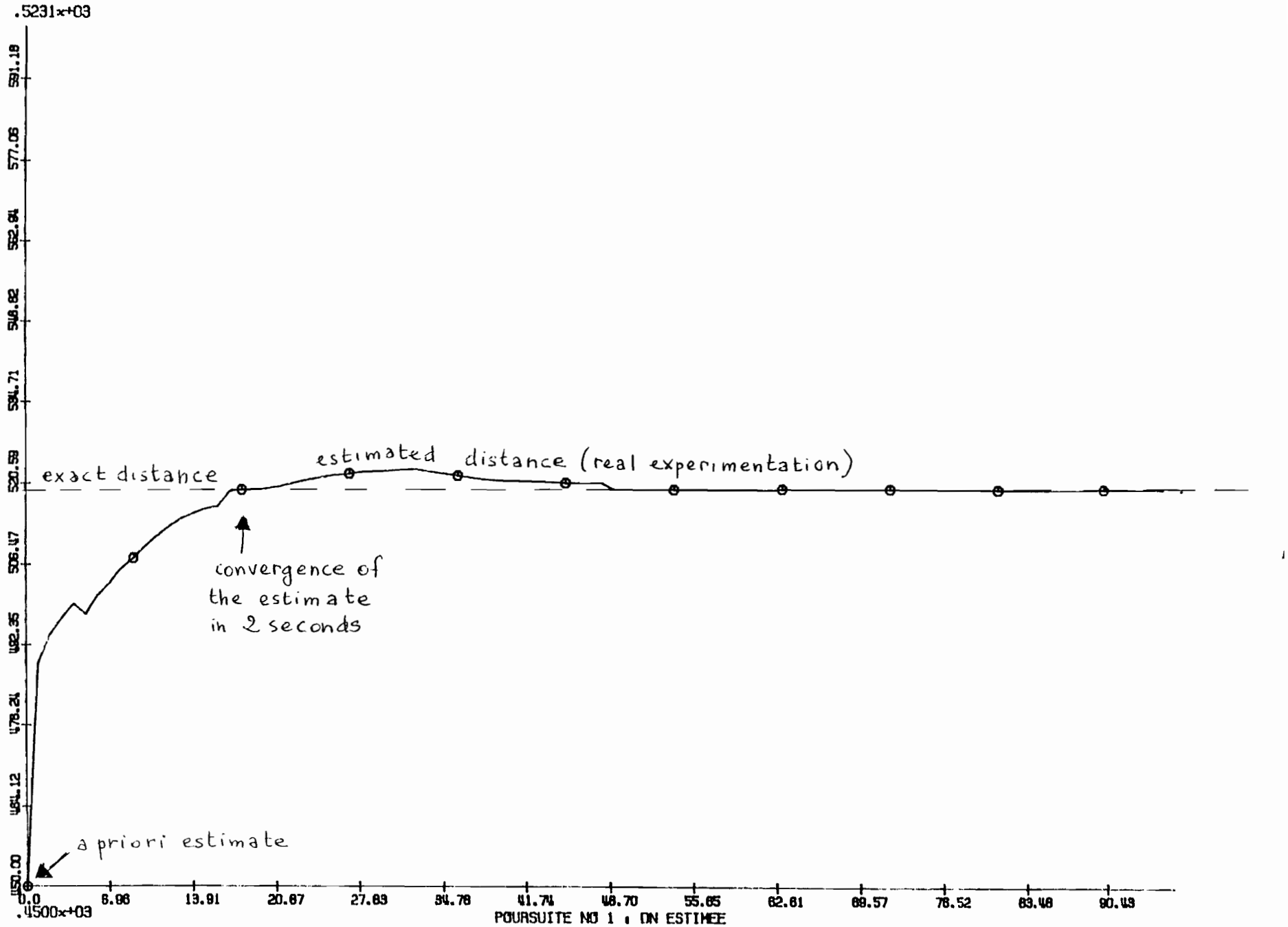
$$\dim B(a_o) = r \leq \dim T_{a_o}(\mathcal{R}_w(a_o)) \leq \dim A = r,$$

so that $\dim T_{a_o}(\mathcal{R}_w(a_o)) = r$ and the result is proved. □

Remark: Clearly the number p+1 of onward + backward transitions is the minimum number such that the matrix in (A2.10) is square, and thus can be assumed to be invertible.

This proof is very similar to the ideas of [33]. However, we have kept our elementary proof since, as a byproduct, we obtain a criterion of local weak reachability very similar to the Kalman criterion of linear discrete-time systems:

$$r = \dim B(a_o) = \sup_{y_1,\ldots,y_r \in R^p} \text{rank}\left((M' - y_1'\Lambda) \,\vdots\, R'(M' - y_2'\Lambda) \,\vdots\, \ldots \,\vdots\, R'^{r-1}(M' - y_r'\Lambda) \right) \ \square$$


.5231x+03
591.18
577.06
562.94
548.82
534.71
520.59
506.47
492.35
478.24
464.12
450.00
exact distance
estimated distance (real experimentation)
convergence of the estimate in 2 seconds
a priori estimate
0.0
6.96
13.91
20.87
27.83
34.78
41.74
48.70
55.65
62.61
69.57
76.52
83.48
90.43
.4500x+03
POURSUITE NO 1 : DN ESTIMEE

PREPRINTS

of the

9th WORLD CONGRESS

of the

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

BUDAPEST, HUNGARY

JULY 2-6, 1984

Series Editors

J. Gertler

L. Keviczky

VOLUME VII

COLLOQUIUM 14.4

Volume Editor

K.J Astrom

Published for the
INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

THE FINITE DIMENSIONAL FILTERING PROBLEM FOR A CLASS OF NONLINEAR DISCRETE-TIME SYSTEMS

J. LEVINE, J. PIGNIE

Centre d'Automatique et Informatique de l'Ecole
Nationale Supérieure des Mines de Paris
35, rue Saint-Honoré
77305 FONTAINEBLEAU (France)

This work is supported by French Army under
contract, DRET n0 83.34.177

Abstract. We obtain a necessary and sufficient condition for the nonlinear discrete-time system :

$$\begin{cases} x_{k+1} = f(x_k) \\ Y_k = h(x_k) + \eta(x_k)v_k \end{cases}$$

to have finite dimensional filters.
This condition is used to obtain an explicit formula of the minimal filter. The result is applied to a tracking problem for a moving target.

Keywords. Nonlinear filtering, Discrete-time nonlinear systems, Realization Theory.

INTRODUCTION

Among the numerous papers, on filtering (see for example [1], [2], [3], [4], [7]), the Kalman filter plays a central role since it can be computed through a finite number of sufficient statistics. This property has recieved the name of "finite dimensionality", and the very serious difficulties met when computing nonlinear filters have motivated to study the systems having an exact finite dimensional filter.

A second approach consists in approximating exact filters by finite dimensional ones (see [5], [6], [11], [13], [16]).

One can also obtain finite dimensional filters from Markovian realization for discrete-state processes (see [8], [15], [17]).

Thus, the finite dimensionality has mostly been studied for continuous-time systems, or discrete-state systems, but very few work has been done in the case of discrete-time (continuous-state) nonlinear systems.

The aim of this paper is to study the finite dimensionality of the exact filter for a class of nonlinear systems of the form :

$$(\Sigma) \quad \begin{cases} x_{k+1} = f(x_k) \\ y_k = h(x_k) + \eta(x_k)v_k \end{cases}$$

These systems are particularized by the fact that there is no noise on the dynamics, but the observations are perturbated by Gaussian noises correlated to the state x_k.

We first derive a recursion formula for the evolution of the unnormalized conditional density, analog in discrete-time to the Zakai equation [18].

In section II, we state, in the Gaussian case the fundamental results and deduce the minimal realization of the filter.

Section III is devoted to an application : a tracking problem for a moving target. We give the nonlinear filter and compare its performances to filters obtained by linear methods.

For complete proofs, the reader may refer to [12].

STATEMENT OF THE PROBLEM

I.1. The basic assumptions

We consider the systems of the folling form :

$$\begin{cases} x_{k+1} = f(x_k) & k = 0, 1, 2, .. \\ y_k = h(x_k) + \eta(x_k)v_k & k = 1, 2, \ldots \end{cases} \qquad (1.1)$$

where :

x_k is supposed to belong to a pure paracompact connected C^1 manifold X of dimension n, and the initial state x_0 to be a random variable with value in X and probability density $P_0 \in C^0(X; R_+)$ with respect to μ_0, a Lebesguian measure on X.

. The observation vector y_k and the noise realization v_k at time $k = 0, 1, \ldots$ are supposed to belong to R^p.

. Furthermore, we suppose that :

H1 : f is a C^1-diffeomorphism from X to X.

H2 : $h \in C^0(X; R^p)$, $\eta \in C^1(X; R^{p \times p})$ and : $(\det \eta)^{-1}(0) \stackrel{\text{def}}{=} \{x \in X \mid \det \eta(x) = 0\}$ is a C^1 submanifold of X, of dimension at most $n-1$.

H3 : The noises v_k $k=1,2,\ldots$ are stationnary, time-uncorrelated and have a density $V \in C^1(R^p; R_+)$ with respect to the Lebesgue measure of R^p.

We shall talk about the analytic, or shortly C^ω case, when X, f, h, η, P_0 and V are C^ω.

I.2. The evolution of the unnormalized conditional density.

Let us denote $Y_k=(y_1,\dots,y_k)$, $k\geq 1$ the history of observations up to time k and Y_0 the empty sequence, f^j and f^{-j} the j^{th} iterate of f and f^{-1} respectively and set :

$$\tilde{X} = X - \bigcup_{j\geq 0} (\det \circ f^{-j})^{-1}(0) \qquad (1.2)$$

It must be noted that $\tilde{X}$ differ from X by a null set with respect to μ_o.

Theorem 1.1. : Under H1-H3, the conditional law of x_k, knowing Y_k has an unnormalized density on $\tilde{X}$ $P_k^*(.|Y_k)\in C^0(\tilde{X}, R_+)$, with respect to μ_o, given by:

$$P_k(x|Y_k)=|\det\eta(x)|^{-1}\gamma(\eta(x)^{-1}(y_k-h(x)))J_f^{-1}(f^{-1}(x)) \cdot P_{k-1}(f^{-1}(x)|Y_{k-1}) \qquad (1.3)$$

$\forall k \geq 1$ $\forall x \in \tilde{X}$ (and thus almost everywhere) here $J_f(x)$ denote the Jacobian of f (with respect to μ_o) evaluated at x.

Equivalently, $P_k(.|Y_k)$ is given as the output of the infinite dimensional system :

$$\begin{cases} \pi_k(x|Y_k)=\gamma(\eta^{-1}(x)(y_k-h(x)))\pi_{k-1}(f^{-1}(x)|Y_{k-1}) \\ \qquad , \ \pi_0(.|Y_0) = P_0 \qquad (1.4) \\ P_k(x|Y_k)= \prod_{j=0}^{k-1} [|\det\eta(f^{-j}(x))|J_f(f^{-j-1}(x))]^{-1} \\ \qquad \cdot \pi_k(x|Y_k) \qquad (1.5) \end{cases}$$

whose state π_k is in $C^0(\tilde{X},R_+)$ and whose inputs are the finite sequences $Y_k\in R^{pk}$ $\forall k \geq 1$. ■

The results in the sequel are stated for global filters on $\tilde{X}$. But one can have the same results with local filters, up-to a straightforward transformation on the hypothesis. For more details see [12].

I.3. Systems realization and finite dimensional filtering.

We shall use the general definition of realizations given in Kalman-Falb-Arbib [10].

Definition 1.1. A filter of (1.1) is a realization with unspecified dimensions of (1.4)-(1.5), of the form :

$$\begin{cases} a_{k+1} = \Phi_k (a_k, y_k) \\ P_k(x|Y_k) = \sigma_k(a_k,x) \qquad \forall x \in \tilde{X} \end{cases} \qquad (1.6)$$

$\sigma_k(a,.)$ being continuous on $\tilde{X}$ for each a and a_o being such that $\sigma_o(a_o,x)=P_o(x)$ $\forall x \in \tilde{X}$.

It is said to be stationary if $\Phi_k \equiv \Phi$ and $\sigma_k \equiv \sigma$ $\forall k \geq 1$, to be pseudostationary if $\Phi_k \equiv \Phi$ and $\pi_k^*(.|Y_k)=\rho(.,a_k)$ with ρ independant of k.

It is said to be finite dimensional if $a_k \in A$ $\forall k \geq 1$, A being a finite dimensional C^1 manifold, called the carameter space, $\Phi_k \in C^1(A\times R^p;A)$ and $\sigma_k \in C^1(A;C^0(\tilde{X},R_+))$.

A finite dimensional filter is said to be analytic if A, Φ_k, and σ_k are analytic ; and invertible if $\Phi_k(.,y)$ is a C^1 diffeomorphism from A to A (C^ω diffeomorphism, in the analytic case), $\forall y \in R^p$, $\forall k \geq 1$ ■

Remark 1.1 : A filter of the form (1.6) always exists : for example (1.4)-(1.5) is such a filter.

Also one can see that to each filter for (1.1) corresponds a realization for (1.4) sharing the same properties (except the stationarity) and conversely. Moreover if a realization for (1.4) is stationary, the corresponding filter for (1.1) is pseudostationary.

We can now state the problem we are interested in

Problem : How to characterize the triples (f,h,r) such that, for given $\tilde{X}$, P_o, and for γ a gaussian density, the corresponding system (1.1) has finite dimensional filters of the form (1.6). And, if such a filter exists, find the minimal realization of the filter and characterize its dimension.

FINITE DIMENSIONAL FILTERING IN THE GAUSSIAN CASE.

Let us suppose that, up to the normalization coefficient :

$$\gamma(v) = e^{-\frac{1}{2}\|v\|^2} \qquad (2.1)$$

for general gaussian densities, it suffices to make a straightforward change on h and η

II.1. The canonical basis.

In the gaussian case, (1.4) may be rewritten :

$$\pi_k(f(x)|Y_k)= \exp\{-\tfrac{1}{2}(y_k-h\circ f(x))'(\eta\circ f(x)\eta'\circ f(x))^{-1}(y_k-h\circ f(x))\} \cdot \pi_{k-1}(x|Y_{k-1}) \ , \ \forall x\in\tilde{X} \qquad (2.2)$$

Let us denote :

$$\begin{cases} H_{ij}(x)=[\eta(x)\eta'(x)]^{-1}_{ij} \\ \qquad i,j=1,\dots,p \ \ ((i,j) \text{ coefficient of the matrix}) \\ H_i(x) = \sum_{j=1}^{p} H_{ij}(x)h_j(x) \quad i = 1, \dots,p \qquad (2.3) \\ H_o(x) = \sum_{i=1}^{p} H_i(x)h_i(x) \end{cases}$$

Then using (2.2) and (2.3), we obtain :

$$\pi_k(f^k(x)|Y_k) = \exp\{-\frac{1}{2}\sum_{j=1}^{k}(\sum_{\alpha,\beta=1}^{p} H_{\alpha\beta}\circ f^j(x)y_j^\alpha y_j^\beta - 2\sum_{\alpha=1}^{p} H_\alpha\circ f^j(x)y_j^\alpha + H_o\circ f^j(x))\}.P_o(x) \qquad (2.4)$$

where y^α denotes the α^{th} component of the vector y.

So, up to the terms $H_o\circ f^j$, independant from the inputs Y_k, we see that π_k is the exponential of a linear combination of the functions in the set :

$$\mathcal{H}=\{H_{\alpha\beta}\circ f^j, H_\alpha\circ f^j \ , \ \alpha,\beta = 1, \dots,p \ ; \ j \geq 1\} \qquad (2.5)$$

Definition 2.1. The vector space spanned by $\mathcal{H}$, denoted span $\mathcal{H}$ is called the canonical space and a basis of span $\mathcal{H}$ is called a canonical basis.

Theorem 2.1. : The following properties are equivalent :

(i) (1.1) admits a finite dimensional filter.

(ii) Span $\mathcal{H}$ is finite dimensional.

(iii) $\exists r \in N$ and $\theta_1 \dots \theta_r \in$ span $\mathcal{H}$, $\exists R \in R^{r\times r}$ with constant coefficients such that :

$$\begin{cases} \begin{pmatrix} \theta_1\circ f(x) \\ \vdots \\ \theta_r\circ f(x) \end{pmatrix} = R \begin{pmatrix} \theta_1(x) \\ \vdots \\ \theta_r(x) \end{pmatrix} \qquad (2.6) \\ H_\alpha(x)= \sum_{i=1}^{r} u_\alpha^i \theta_i(x) \ , \ H_{\alpha\beta}(x) = \sum_{i=1}^{r} \lambda_{\alpha\beta}^i \theta_i(x) \\ \qquad \forall \alpha,\beta=1,\dots,p \end{cases}$$

$\forall x \in \bar{X}$.

II 2. The minimal realization of the filter.

If dim span $H = r < +\infty$, let us choose a canonical basis $\{\theta_1, \ldots, \theta_r\}$, and call M the $\gamma \times r$ matrix $[u_\alpha^\gamma]$, Λ the tensor of order 3 $[\lambda_{\alpha\beta}^\gamma]$ with $u_\alpha^\gamma, \lambda_{\alpha\beta}^\gamma$ defined as in (2.6)

Theorem 2. : If dim span $H = r < +\infty$, there exists an invertible analytic and pseudo-stationary realization of the filter of dimension r. The minimal realization is given, in matrix form by $a_k = (a_k^1, \ldots, a_k^r)'$ and by :

$$\begin{aligned} a_{k+1}^k &= R'^{-1} a_k + M' y_k - \frac{1}{2} y_k' \Lambda y_k \quad , \quad a_o = 0 \\ P_k(x|Y_k) &= \{ \prod_{j=0}^{k-1} [|\det \mathrm{of}^{-j}(x)| J_f(f^{-j-1}(x))]^{-1} \\ &\exp\{-\frac{1}{2} \sum_{j=0}^{k-1} H_o \mathrm{of}^{-j}(x)\} P_o(f^{-k}(x))\} \end{aligned} \tag{2.7}$$

$$\rho(a_k, x)$$

with

$$\rho(a,x) = \exp\Big(\sum_{i=1}^{r} a^i \theta_i(x)\Big) \text{ and } R \text{ as in (2.6).}$$

Furthermore this realization is observable and locally weakly reachable in the sens given in [9] or [14].

Any other minimal invertible analytic and pseudo-stationary filter can be obtained from (2.7) by a C^ω-diffeomorphism. ∎

APPLICATION TO A TRACKING PROBLEM FOR A MOVING TARGET

One can find many examples of systems having finite dimensional filters, for example, the systems with linear dynamics and polynomial observations.

When the system is linear , the method gives a finite dimensional filter, whatever the inital density may be. Furthermore, if the initial state is Gaussian, the filter is exactly the "informations filter" of the linear filtering theory, with an extra parameter left for normalization (see [12]).

In the following application, we have compared our nonlinear filter to filters obtained by linear methods.

III.1 Traking for a moving target.

A moving target is observed through an optical system, giving a noisy measurement of the inverse of the angular velocity of the target. Assuming the target velocity is constant (what is justified as the tracking phase is very short) we want to estimate the initial distance L_o and the nodal distance d as displayed on the figure 3.1 :

We shall suppose that the noises are gaussian, stationary and uncorrelated, and the initial density is a uniform density on the rectangle $[\underline{x}^1, \bar{x}^1] \times [\underline{x}^2, \bar{x}^2]$, and as the measurement are digital, treate the problem in discrete time : If we call Δt the time mesh, we get the system :

$$\begin{cases} x_{k+1}^1 = x_k^1 - V\Delta t \\ x_{k+1}^2 = x_k^2 \\ y_k = ((x_k^1)^2 + (x_k^2)^2)/N x_k^2 + \gamma v_k \end{cases} \tag{3.1}$$

We can check that dim Span $H = 4$ and a canonical basis is given by : $\theta_1 \equiv 1$, $\theta_2 = ((x^1)^2/x^2 + x^2)/N$, $\theta_3 = \Delta t\, x^1/x^2$, $\theta_4 = \gamma(\Delta t)^2/x^2$. We obtain then the minimal filter.

$$\begin{bmatrix} a_{k+1}^1 \\ a_{k+1}^2 \\ a_{k+1}^3 \\ a_{k+1}^4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 1 & 0 \\ 0 & & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_k^1 \\ a_k^2 \\ a_k^3 \\ a_k^4 \end{bmatrix}$$

$$+ \begin{bmatrix} -1/2\eta^2 \cdot y_{k+1}^2 \\ 1/\eta^2 \cdot y_{k+1} \\ 0 \\ 0 \end{bmatrix} , \begin{bmatrix} a_o^1 \\ a_o^2 \\ a_o^3 \\ a_o^4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.2}$$

$$\begin{aligned} P_k(x|Y_k) = \frac{1}{\eta^k} \exp\{&(-k/2\eta^2)(\theta_2^2 + 2(k-1)\theta_2\theta_3 \\ &+ (k-1)(2k-1)(\theta_2\theta_4 - (\Delta t)^2) + k(k-1)^2\theta_3\theta_4 \\ &+ (1/30)\cdot(k-1)(6k^3 - 9k^2 + k + 1)\theta_4^2)\}. \end{aligned} \tag{3.3}$$

$$\exp(a_k^1\theta_1 + a_k^2\theta_2 + a_k^3\theta_3 + a_k^4\theta_4)$$

$$\cdot P_o(x^1 + kV\Delta t, x^2).$$

For real time implementation, $E(d|Y_k)$ and $E(L_o|Y_k)$ may be precomputed as functions of $(a_k^1, \ldots, a_k^4, k)$. So the only calculation to be done "on line" is (3.2).

Simulation have shown that such a filter is efficient. But it is very sensitive to the noise level η and the initial measure P_o. Moreover, as in (3.2) there are two integrators, it is very sensitive to numerical errors.

To improve it's behviour. one can adapt the support of P_o from time to time and reinitialize the filter after a certain time.

III.2. Comparison with linear filters

This filter has been compared with two linear methods. The first one is an extended Kalman filter on (3.1). The second is deduce from the fact that we can replace (3.1) by the system, written in continuous time :

$$\begin{cases} \dot{x}^1 = -V & x_o^1 = (L_o^2 - d^2)^{1/2} \\ \dot{x}^2 = 0 & x_o^2 = d \\ \dot{x}^3 = x^4 & x_o^3 = (L_o^2/d) - d \\ \dot{x}^4 = x^5 & x_o^4 = -2V(L_o^2 - d^2)^{1/2}/d \\ \dot{x}^5 = 0 & x_o^5 = 2V^2/d \\ y(t) = (x^3 + x^2)/N + \text{noise} \end{cases} \tag{3.4}$$

letting $x^3 = (x^1)^2/x^2$, $x^4 = -2Vx^1/x^2$, $x^5 = 2V^2/x^2$, by approximating the initial conditions $(x_o^1, \ldots, x_o^5)$, which are not Gaussian by a Gaussian vector and then by applying a Kalman filter.

One can see that (3.4) is not completly observables. In particular x^1 is not observable

and x^2 is only partially observable. And in fact, the Kalman filter is very unefficient for estimating x^1 and x^2.

But the filter obtained with a minimal realization of (3.4) is even worse as x^1 ans x^2 are very nonlinear statistics for this filter.

The Kalman extended filter is unefficient unless the initial estimate is close to the good value. But even in this case it explodes in about 30-40 iterations.

The nonlinear filter, for an initial estimation error of 30-40 per cent gives in less than 15 iterations an estimation at 8 per cent in the worst cases. See Fig. 3.2.-3.6.

Remark, finally that if, instead of $y=1/\underline{\theta}$ + noise we had $y = \theta$ + noise, the problem would be infinite dimensional and we would need approximation techniques to use a nonlinear filter.

ACKNOWLEDGEMENT

The authors are mostly indebted to Mr. Kocher and Huyhn of Sopelem, who have kindly communicated the tracking problem as well as recordings of real experiments.

REFERENCES

[1] Benes, V.E. (1981). Exact finite dimensional filters for certain diffusions with nonlinear drifts. Stochastics, 5, 65-92.

[2] Brockett, R.W. (1976). Remarks on finite dimensional nonlinear estimation. In C. Lobry (Ed.). Analyse des Systèmes. Asterisque, 75-76,(SMF), 47-56.

[3] Chaleyat-Maurel, M. and D. Michel (1983). Un théorème de non-existence de filtres de dimension finie. CRAS, Vol. 296, 933-936.

[4] Chikte, S. and J.T.H. Lo (1980). Optimal filters for bilinear systems with nilpotent Lie algebras. IEE Trans. AC24,6, 943-953.

[5] Di Masi, G.B. and W.J. Runggaldier, (1982). Approximation and bounds for discrete-time nonlinear filtering. In Bensoussan-Lions (Ed.). Lect. Notes in Control and Inf. Sc., 44. Springer Verlag.

[6] Hazewinkel, M. (1981). On deformations, approximations and nonlinear filtering. Systems and Control Letters, Vol.1, 1, 32-36.

[7] Hazewinkel, M., S.I. Marcus and H.J. Sussmann (1983.) Non existence of exacte finite dimensional filters for the cubic censor problem. Systems and Control letters, Vol 3, 6, 331-340.

[8] Heller, A. (1965). On stochastic processes derived from Markov chains. Ann. Math. Statistics, 36, 1236-1291.

[9] Jakubczyk, B. (1980). Invertible realizations of of nonlinear discrete-time systems. Proc. Princeton Conf. on Information and Systems.

[10] Kalman, R.E., P.L. Falb, and M.A. Arbib (1969). Topics in Mathematical Systems Theory. McGraw Hill.

[11] Kushner, H.J. (1979). A robust discrete state approximation of the optimal nonlinear filter for a diffusion. Stochastics, 3, 75-83.

[12] Lévine, J. and G. Pignié (1984). Exact finite dimensional filters for a class of nonlinear discrete time systems. To appear.

[13] Marcus, S.I. C.H. Lie and G.L. Blankenship (1983). Lie. Algébras and asymptotic expansions for nonlinear filtering problems. To appear.

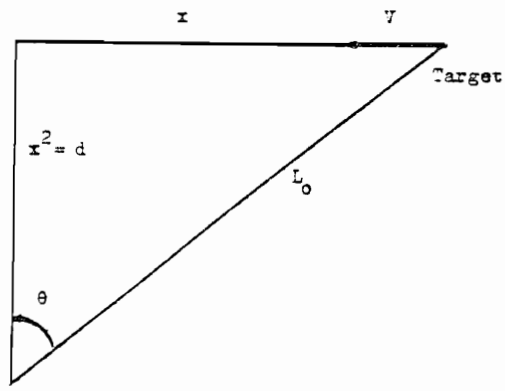
[14] Normand-Cyrot, D.(1983). Théorie et pratique des systèmes nonlinéaires en temps discret. These de Doctorat d'Etat. Université Paris XI-Orsay.

[15] Ricci, G. (1977). On the internal structure of finite state stochastic processes. Proc. Conf. Taormina, Italy.

[16] Sussman, H.J. (1982). Approximate finite dimensional filters for some nonlinear problems. Stochastics, 7, 183-203.

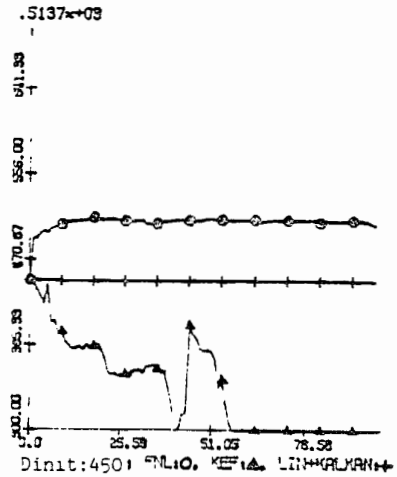
[17] Willsky, A.S. (1978). On the algebraic structure of certain partially observable finite state Markov processes. Information and Control, 38, 179-212.

[18] Zakai, L. (1969). On the optimal filtering of diffusion processes. Z. Wahr. Verw. Geb., 11, 230-243.



Optical system

Fig. 3.1



Dinit:4501 ENL=0. KEF=1. LIN=KALMAN

fig.3.2.

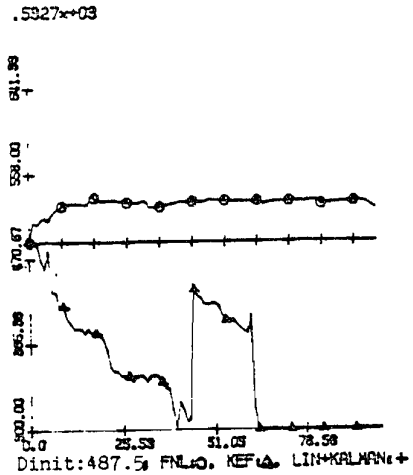


fig.3.3.

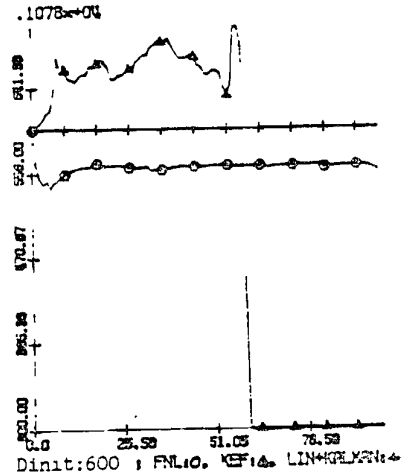


fig.3.6

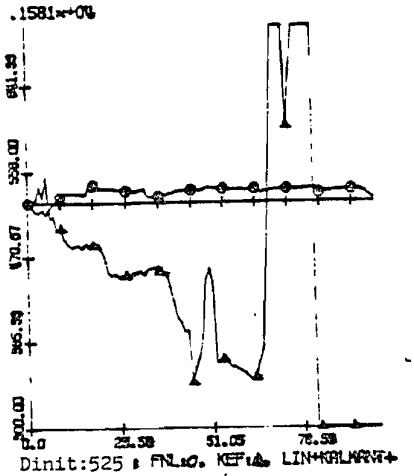


fig.3.4.

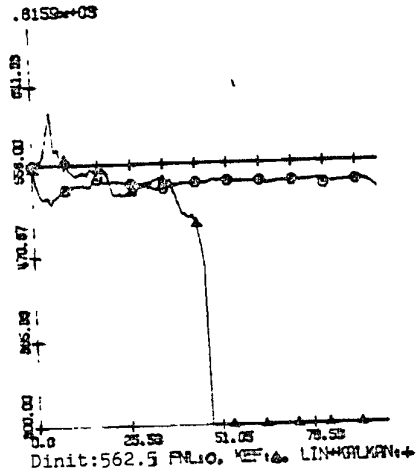


fig.3.5.

UNE CLASSE DE SYSTEMES NONLINEAIRES A TEMPS CONTINU ADMETTANT DES FILTRES DE DIMENSION FINIE.

J. LEVINE *

RESUME :

On étudie la propriété de finitude du filtre associé à un système differentiel stochastique nonlinéaire sans bruits dans la dynamique de l'état. On obtient la condition nécessaire et suffisante pour qu'un tel filtre existe, ainsi que la réalisation minimale du filtre grâce à l'obtention de la solution explicite de l'équation de Zakaï. On montre de plus que le filtre est de dimension finie si et seulement si l'algèbre de Lie de l'équation de Zakaï est de dimension finie. Enfin tous les systèmes de la classe étudiée dont le filtre est de dimension finie sont immergeables dans un système linéaire.

* CAI-ENSMP, 35 rue Saint-Honoré, 77305 FONTAINEBLEAU - FRANCE.
Cette étude a été financée par contrat DRET No 83.34.177.

I - INTRODUCTION

Le but de ce travail est de généraliser en temps continu les résultats de [6]. Précisément, on étudie les conditions sous lesquelles le système:

$$(1.1.a.)\quad \begin{cases} dx_t = f(x_t)dt\ , & t \geqslant 0 \\ dy_t = h(x_t)dt + dv_t\ , & y_o = 0 \end{cases} \qquad (1.1.b.)$$

avec $x_t \in R^n$, $y_t \in R^p$, x_o aléatoire de densité μ par rapport à la mesure de Lebesgue sur R^n , et v_t un mouvement Brownien sur $(\Omega, \mathfrak{F}_t, P)$ à valeurs dans R^p, admet un filtre de dimension finie, en abrégé FDF.
Notre définition de FDF est légèrement différente de celle de filtre universel de dimension finie de [3], mais a l'avantage d'apparaître naturellement par le calcul de la loi conditionnelle non normalisée de (1.1.a) par rapport aux observations (1.1.b), et sans faire les hypothèses d'intégrabilité de la propriété (P) (voir [3]).
Définition 1. Etant donnée la loi conditionnelle π_t^y de x_t sachant la trajectoire $Y = \{y(s) \mid o \leqslant s \leqslant t\}$, on dit que (1.1) admet un FDF ssi il existe un entier r fini, une équation différentielle stochastique sur R^r :

$$(1.2)\quad da_t = \varphi(a_t)dt + \sum_{i=1}^{p} \psi_i\ (a_t)dy_t^i$$

avec $\varphi, \psi_1, \ldots, \psi_p \in C_b^\infty(R^r;R^r)$ à dérivées de tous ordres bornées,
une application $\sigma \in C^\infty(R^r \times R^n; R_+)$, et une application $\bar{\sigma} \in C^\infty(R^n \times R_+; R_+)$ telles que :

$$(1.3)\quad \pi_t^y(x) = \sigma(a_t,x)\ \bar{\sigma}(t,x),\ \text{P-p.s.},\ \forall (t,x) \in R_+ \times R^n\ ,$$

où a_o est tel que $\mu(x) = \sigma(a_o,x)\ \bar{\sigma}(o,x)$, et où a_t est la solution de (1.2) engendrée par Y à partir de a_o. ■
Cette définition généralise les définitions (1.1) et (1.2) de [6].

L'usage en filtrage est de considérer les équations au sens de Stratonovitch pour des raisons géométriques. Cependant, comme on se place ici dans R^r et comme nous ne calculons pas de développements fonctionnels stochastiques (tout au moins pas directement) il nous a semblé que les résultats revêtaient une forme plus simple en calcul d'Ito.

Bien qu'on puisse partout raisonner localement, pour la clarté de l'exposé, nous faisons l'hypothèse simplificatrice:

$$(H)\quad \begin{cases} \text{Supposons que } f \text{ et } h \text{ sont } C^\infty \text{ et que (1.1.a) admet un groupe de } C^\infty \\ \text{difféomorphismes } \{X_t\ ,\ \forall t \in R\},\ \text{c'est-à-dire:} \\ \frac{d}{dt} X_t(x) = f(X_t(x)) \quad \forall t \in R,\ X_o(x) = x,\ X_t : C^\infty \text{ difféomorphisme de} \\ R^n \text{ sur } R^n \ \blacksquare \end{cases}$$

Notons $\mathcal{L}_f(h_i)(x)$ la dérivée de Lie de h_i suivant le champ de vecteurs f évaluée au point x :

$$(1.4)\quad \mathcal{L}_f(h_i)(x) = \sum_{j=1}^{n} f_j(x) \frac{\partial h_i}{\partial x_j}(x) \quad \text{et} \quad \mathcal{L}_f^m(h_i)(x) = \mathcal{L}_f(\mathcal{L}_f^{m-1}(h_i))(x).$$

Le principal résultat de ce papier est la caractérisation suivante des systèmes (1.1) admettant un FDF :

<u>une condition nécessaire et suffisante pour que (1.1) admette un FDF est que dim Span $\mathcal{H} < +\infty$</u>, où:

$$(1.5)\qquad \mathcal{H} = \{h_i, \mathcal{L}_f^m(h_i) \mid i=1,\dots,p,\ \forall m \geq 1\}\ .$$

Cette caractérisation permet de calculer la réalisation minimale du filtre sous la forme:

$$(1.6)\quad da_t = -R' a_t dt + \sum_{i=1}^{p} M_i dy_t^i$$

où R et $M_1,\dots,M_p$ sont respectivement une matrice et des vecteurs que l'on calcule à partir des données.

Pour faire le lien avec l'approche utilisant l'Algèbre de Lie, de l'équation de Zakaï de la densité conditionnelle non normalisée (voir par exemple [1], [2], [3], [5], [7], [8]), la condition (1.4) est équivalente à la finitude de la dimension de cette algèbre de Lie. Un cas particulier de ce résultat, dans le cas où l'Algèbre de Lie est nilpotente a été obtenu récemment par Roth et Loparo [9] par des méthodes heuristiques, utilisant en particulier la forme robuste de l'équation de Zakaï sans démontrer sa validité, et sans même démontrer l'existence d'une densité conditionnelle.

L'approche que nous suivons ici est inverse: on calcule <u>explicitement</u> la densité conditionnelle π_t^y et on calque la dérivation de la condition (1.5) sur celle de [6].

II - LA DENSITE CONDITIONNELLE π_t^Y

Théorème 1 : Si l'hypothèse (H) a lieu, π_t^Y est donnée par :

$$(2.1)\quad \pi_t^Y(x) = \exp\Big[\sum_{i=1}^p h_i(x)y_t^i - \sum_{i=1}^p \sum_{j=1}^n \int_0^t f_j(X_{s-t}(x)) \frac{\partial h_i}{\partial x_j}(X_{s-t}(x))y_s^i ds - \frac{1}{2}\sum_{i=1}^p \int_0^t h_i^2(X_{s-t}(x))ds\Big] . \Big|\det \frac{\partial X_{-t}}{\partial x}(x)\Big| . \mu(X_{-t}(x)),$$

$$\forall t \geq 0,\ \forall x \in R^n,\ P\text{-p.s.}$$

Preuve : Designons par Q la loi de (x,y), solution du problème de martingale associé à (1.1). Par le théorème de Cameron-Martin-Girsanov, la mesure Q_o définie par :

$$(2.2)\quad \frac{dQ_o}{dQ}\Big|_{\mathcal{F}_t} = \Big(\exp \sum_{i=1}^p \Big[\int_0^t h_i(X_{s-t}(x))dy_s^i - \frac{1}{2}\int_0^t h_i^2(X_{s-t}(x))ds\Big]\Big)^{-1}$$

est telle que y_t est un Q_o-Brownien, indépendant de x_t.
Comme d'autre part on vérifie facilement que la loi de x_t a pour densité par rapport à la mesure de Lebesgue de R^n :

$$(2.3)\quad \mu_t(x) = \Big|\det \frac{\partial X_{-t}}{\partial x}(x)\Big| \mu(X_{-t}(x)),$$

il vient que $Q_o|_{\mathcal{F}_t} = \mu_t dx \otimes P|_{\mathcal{F}_t}$ où P est la mesure de Wiener, et, avec (2.2):

$$(2.4)\quad Q|_{\mathcal{F}_t} = (Z_t\ \mu_t)dx \otimes P|_{\mathcal{F}_t}$$

où :

$$(2.5)\quad Z_t = \exp \sum_{i=1}^p \Big[\int_0^t h_i(X_{s-t}(x))dy_s^i - \frac{1}{2}\int_0^t h_i^2(X_{s-t}(x))ds\Big],$$

d'où immédiatement $\pi_t^Y(x) = Z_t(x,Y)\mu_t(x)$, et, intégrant par parties $\int_0^t h_i(X_{s-t}(x))dy_s^i$, ou plus précisément, appliquant la formule d'Ito à $h_i(X_t(x))y_t^i$, on trouve, pour tout $x \in R^n$:

$$(2.6)\quad h_i(X_t(x))y_t^i = \int_0^t \sum_{j=1}^n f_j(X_s(x)) \frac{\partial h_i}{\partial x_j}(X_s(x))y_s^i ds + \int_0^t h_i(X_s(x))dy_s^i ,$$

soit, changeant x en $X_{-t}(x)$:

$$(2.7)\quad \int_o^t h_i(X_{s-t}(x))dy_s^i = h_i(x)y_t^i - \int_o^t \sum_{j=1}^n f_j(X_{s-t}(x)) \frac{\partial h_i}{\partial x_j}(X_{s-t}(x))y_s^i ds$$

et, après avoir remplacé dans (2.5), on trouve (2.1). ■

Remarque 1 : l'hypothèse (H) a lieu par exemple lorsque f est à croissance linéaire et à divergence bornée. ■

Corollaire 1 : π_t^Y est solution de l'équation de Zakaï au sens d'Itô :

$$(2.8)_1\quad d\pi_t^Y = -\operatorname{div}(f\pi_t^Y)dt + \sum_{i=1}^p h_i\pi_t^Y dy_t^i \quad , \quad \pi_o^Y = \mu$$

ou, de manière equivalente, au sens de Stratonovitch :

$$(2.8)_2\quad d\pi_t^Y = (-\operatorname{div}(f\pi_t^Y) - \frac{1}{2}\sum_{i=1}^p h_i^2\pi_t^Y)dt + \sum_{i=1}^p h_i\pi_t^Y \bar{dy}_t^i, \quad \pi_o^Y = \mu,$$

où $\bar{dy}_t^i$ est la différentielle de y_t^i au sens de Stratonovitch.

Preuve : Il suffit d'appliquer la formule d'Itô à :

$$\int \varphi(x)\pi_t^Y(x)dx = \int \varphi(X_t(x))\exp\sum_{i=1}^p \left(\int_o^t h_i(X_s(x))dy_s^i - \frac{1}{2}\int_o^t (h_i(X_s(x)))^2 ds\right)\mu(x)dx. \blacksquare$$

Corollaire 2 : On a la décomposition suivante:

$$(2.9)\quad \pi_t^Y = \tilde{\pi}_t^Y \cdot \tilde{\sigma}_t \quad , \text{ avec :}$$

$$(2.10)\quad \tilde{\pi}_t^Y(x) = \exp\sum_{i=1}^p \left[h_i(x)y_t^i - \int_o^t \sum_{j=1}^n f_j(X_{s-t}(x)) \frac{\partial h_i}{\partial x_j}(X_{s-t}(x))y_s^i ds\right]$$

$$(2.11)\quad \tilde{\sigma}_t(x) = \exp -\frac{1}{2}\left[\sum_{i=1}^p \int_o^t h_i^2(X_{s-t}(x))ds\right] \left|\det \frac{\partial X_{-t}}{\partial x}(x)\right| \mu(X_{-t}(x))$$

où seul $\tilde{\pi}_t^Y$ dépend des observations Y.

Preuve : évident. ■

III - CARACTERISATION DES FDF

Grâce à (2.9), (2.10), (2.11), on a réussi à isoler la partie de π_t^Y qui ne dépend pas des observations, et, dans (2.10), à obtenir une connaissance précise de la manière dont sont pondérées les observations.
En effet, avec la notation (1.4), on a :

$$(3.1) \qquad \tilde{\pi}_t^Y(x) = \exp \sum_{i=1}^{p} [h_i(x)y_t^i - \int_0^t \mathcal{L}_f(h_i)(X_{s-t}(x))y_s^i ds]$$

On va développer la même méthode que dans [6], qui consiste à montrer que l'espace vectoriel engendré par les $\{h_i, \mathcal{L}_f^m(h_i)\}$ doit être de dimension finie pour avoir un FDF.

Théorème 2 : les 3 assertions suivantes sont équivalentes :

(i) (1.1) admet un FDF.

(ii) dim Span $\mathcal{H} < +\infty$ ($\mathcal{H}$ donné en (1.5))

(iii) $\exists r \in N$, $\exists \theta_1, \ldots, \theta_r \in$ Span $\mathcal{H}$, $\exists R \in R^{r \times r}$, $\exists M_1, \ldots, M_p \in R^r$, tels que :

$$\begin{cases} \mathcal{L}_f(\theta)(x) = R\theta(x) \\ h_i(x) = M_i' \theta(x) \quad , \quad i=1,\ldots,p \end{cases} \qquad \forall x \in R^n$$

où $\theta = (\theta_1, \ldots, \theta_r)'$

Preuve : (i) $\Rightarrow$ (ii)

Si (1.1) admet un FDF, d'après la définition 1, il existe un entier r fini, une équation différentielle stochastique sur R^r :

$$(3.2) \qquad da_t = \varphi(a_t)dt + \sum_{i=1}^{p} \psi_i(a_t)dy_t^i$$

et une fonction $\sigma : C^\infty$ de $R^n \times R^r$ dans R_+ tels que :

$$(3.3) \qquad \sigma(a_t, x) = \tilde{\pi}_t^Y(x)$$

En effet, la partie $\tilde{\sigma}_t$ peut être choisie comme en (2.11) sans restreindre la généralité.
Utilisant alors (3.1), il vient :

$$(3.4) \qquad \log \sigma(a_t, x) = \sum_{i=1}^{p} [h_i(x)y_t^i - \int_0^t \mathcal{L}_f(h_i)(X_{s-t}(x))y_s^i ds]$$

Soit alors U^i defini par $U^i_t = \int_o^t u^i(s)ds$, $u^i \in L^1(R_+;R)$, $\forall i=1,\dots,p$. On va changer, dans (3.4), la trajectoire Y en $Y+\varepsilon U^i$, et notons :

$$(3.5)\quad \begin{cases} da_t(Y+\varepsilon U^i) = \varphi(a_t(Y+\varepsilon U^i))dt + \sum_{i=1}^{p} \psi_\perp(a_t(Y+\varepsilon U^i))(dy^i_t + \varepsilon u^i_t\, d\tau) \\ a_o(Y+\varepsilon U^i) = a_o(Y) = a_o \end{cases}$$

la trajectoire du filtre correspondant à $Y+\varepsilon U^i$ qui existe et est unique pour P-presque tout Y et $\forall\, U^i$. D'après (3.4), il vient :

$$(3.6)\quad \frac{1}{\varepsilon}\left(\operatorname{Log}\sigma(a_t(Y+\varepsilon U^i),X_t(x)) - \operatorname{Log}\sigma(a_t(Y),X_t(x))\right)$$
$$= h_i(X_t(x))U^i_t - \int_o^t \mathcal{L}_f(h_i)(X_s(x))U^i_s\,ds$$

donc la limite lorsque $\varepsilon \to o$ existe, et comme σ est C^∞ en a à x fixé, le second membre de (3.6) étant linéaire et continu sur l'espace $AC(R_+)$ des fonctions absolument continues sur R_+, on peut trouver r formes linéaires continues sur $L^1(R_+)$ notées $D_Y(a^k_t)$, $k=1,\dots,r$, telles que :

$$(3.7)\quad \lim_{\varepsilon \to o} \frac{1}{\varepsilon}\left(\operatorname{Log}\sigma(a_t(Y+\varepsilon U^i),X_t(x)) - \operatorname{Log}\sigma(a_t(Y),X_t(x))\right)$$
$$= \sum_{k=1}^{r} \frac{1}{\sigma(a_t(Y),X_t(x))} \frac{\partial\sigma}{\partial a^k}(a_t(Y),X_t(x)) < D_Y(a^k_t),U^i >$$
$$= h_i(X_t(x))U^i_t - \int_o^t \mathcal{L}_f(h_i)(X_s(x))U^i_s ds \quad \forall\, U^i \in AC(R_+),\ \forall i=1,\dots,p.$$

Notons $\rho_k(a_t(Y),x) = \dfrac{1}{\sigma(a_t(Y),X_t(x))} \dfrac{\partial\sigma}{\partial a^k}(a_t(Y),X_t(x))$, et montrons que h_i, $\mathcal{L}^m_f(h_i)$ s'expriment comme des combinaisons linéaires, à coefficients indépendants de x, des r fonctions $\rho_1(a_t(Y),.),\dots,\rho_r(a_t(Y),.)$. Dans ce but, remplaçons d'abord dans (3.7) U^i par $1_t \overset{def}{=} \{U_s = 1,\ o \leqslant s \leqslant t\}$. On obtient :

$$(3.8)\quad h_i(x) = h_i(X_t(x)) - \int_o^t \mathcal{L}_f(h_i)(X_s(x))ds = \sum_{k=1}^{r} \rho_k(a_t(Y),x) < D_Y(a^k_t),1_t >$$

Remplaçant ensuite U^i dans (3.7) par $U_{\theta,\alpha}$ défini par :

$$U_{\theta,\alpha} = \{U_s = 1 \quad \forall s \in [0,\theta],\ U_s \in [0,1] \quad \forall s \in\ ]\theta,\theta+\alpha[,\ U_s = o \quad \forall s \geqslant \theta+\alpha,\ U_{\theta,\alpha} \in AC(R_+)\}$$

Pour $\theta,\alpha > o$ tels que $\theta+\alpha < t$, il vient, pour α petit :

$$\int_0^\theta \mathcal{L}_f(h_i)(X_s(x))ds + o(\alpha) = \sum_{k=1}^r \rho_k(a_t(Y),x) < D_Y(a_t^k), \Pi_{\theta,\alpha} >$$

et donc :

$$\int_0^\theta \mathcal{L}_f(h_i)(X_s(x))ds = \sum_{k=1}^r \rho_k(a_t(Y),x) \left(\lim_{\alpha \to o} < D_Y(a_t^k), U_{\theta,\alpha} > \right).$$

Aussi, comme l'application $\theta \to \int_0^\theta \mathcal{L}_f(h_i)(X_s(x))ds$ est C^∞ sur $[0,t]$, on peut trouver r fonctions $\theta \to C_k(a_t,Y,\theta)$, de classe C^∞ sur $[o,t]$ telles que :

$$\int_0^\theta \mathcal{L}_f(h_i)(X_s(x))ds = \sum_{k=1}^r \rho_k(a_t(Y),x)\, C_k(a_t,Y,\theta)$$

$$(3.9)\quad \frac{\partial^m}{\partial \theta^m} \int_0^\theta \mathcal{L}_f(h_i)(X_s(x))ds \Big|_{\theta=o} = \mathcal{L}_f^m(h_i)(x) = \sum_{k=1}^r \rho_k(a_t(Y),x)\left(\frac{\partial^m}{\partial \theta^m} C_k(a_t,Y,\theta)\right)\Big|_{\theta=o}$$

Donc, regroupant (3.8) et (3.9), on obtient :

$$\mathcal{H} = \{h_i, \mathcal{L}_f^m(h_i) \,|\, i=1,\dots,p, \quad m \geq 1\}$$

$$\subset \text{Span}\,\{\rho_1(a_t(Y),\circ),\dots,\rho_r(a_t(Y),\circ)\}$$

Ce qui prouve que dim Span $\mathcal{H} \leq r$.

<u>(ii) ⇒ (iii)</u> évident : si dim Span $\mathcal{H} = r$ et si $\{\theta_1,\dots,\theta_r\}$ est une base de Span $\mathcal{H}$, comme $\mathcal{L}_f(\mathcal{H}) \subset \mathcal{H}$, on doit avoir $\mathcal{L}_f(\theta) = R\theta$ avec R matrice (r,r). De plus, comme $h_i \in \mathcal{H}$, on a : $h_i(x) = \sum_{k=1}^r M_{i,k}\, \theta_k(x) = M_i\theta(x)$ $\forall i=1,\dots,p$, d'où (iii).

<u>(iii) ⇒ (i)</u>. Posons $\sigma(a_t,x) = \exp \sum_{k=1}^r a_t^k\, \theta_k(x) = \tilde{\pi}_t^Y(x)$, et identifions les coefficients φ et $\psi_1,\dots,\psi_p$ de la diffusion associée à a_t :

Par la formule d'Itô appliquée à $\sigma(a_t,X_t(x)) = \tilde{\pi}_t^Y(X_t(x))$, on a , $\forall x \in R^n$:

$$(3.10)\quad \Big(\Big(\sum_{k=1}^{r}(\varphi_k(a_t)\theta_k(X_t(x)) + a_t^k \mathcal{L}_f(\theta_k)(X_t(x))\Big) + \frac{1}{2}\sum_{i=1}^{p}\Big(\sum_{k=1}^{r}\psi_{i,k}(a_t)\theta_k(X_t(x))\Big)^2\Big)\sigma(a_t,X_t(x))dt$$
$$+ \sum_{k=1}^{r}\sum_{i=1}^{p}\theta_k(X_t(x))\psi_{i,k}(a_t)\sigma(a_t,X_t(x))dy_t^i =$$
$$= \sum_{i=1}^{p}\Big(\frac{1}{2}h_i^2(X_t(x))dt + h_i(X_t(x))dy_t^i\Big)\tilde{\pi}_t^Y(X_t(x))$$

et, identifiant les parties à variations bornées et martingales, avec le fait que, par définition, $\sigma(a_t,x) = \tilde{\pi}_t^Y(x)$, et que X_t est bijectif:

$$(3.11)\quad \begin{cases} \sum_{k=1}^{r}\psi_{i,k}(a_t)\theta_k(x) = h_i(x), \quad i=1,\dots,p \\ \sum_{k=1}^{r}(\varphi_k(a_t)\theta_k(x) + a_t^k \mathcal{L}_f(\theta_k)(x)) + \frac{1}{2}\sum_{i=1}^{p}\Big(\sum_{k=1}^{r}\psi_{i,k}(a_t)\theta_k(x)\Big)^2 = \frac{1}{2}\sum_{i=1}^{p}h_i^2(x) \end{cases}$$

soit, comme le second membre de la 1ère équation est égal par hypothèse à $\sum_{k=1}^{r} M_{i,k}\theta_k(x)$:

$$(3.12)\quad \psi_{i,k}(a) = M_{i,k} \quad i=1,\dots,p, \quad k=1,\dots,r$$

et la 2ème équation devient :

$$\sum_{k=1}^{r}(\varphi_k(a_t)\theta_k(x) + a_t^k \mathcal{L}_f(\theta_k)(x)) = 0$$

Utilisant enfin l'identité : $\mathcal{L}_f(\theta) = R\,\theta$, on obtient :

$$(3.13)\quad \varphi_k(a_t) = -\sum_{j=1}^{r} R_{j,k}\, a_t^j \quad , \quad k=1,\dots,r.$$

Soit finalement, comme $\sigma(a_o,x) = 1 = \exp\Big(\sum_{k=1}^{r} a_o^k\theta_k(x)\Big)$:

$$(3.14)\quad \begin{cases} \begin{cases} da_t = -R'a_t + \sum_{i=1}^{p} M_i\, dy_t^i \\ a_o = 0 \end{cases} \\ \sigma(a_t,x) = \tilde{\pi}_t^Y(x) \end{cases}$$

et $a_t \in R^r \quad \forall t \geq 0$, ce qui prouve que (1.1) admet le filtre (3.14) qui est de dimension finie, et le théorème est démontré. ■

Corollaire 3 : Considérons les opérateurs $L_o, L_1, \ldots, L_p$ définis par :

$$(3.15)\quad L_o\varphi = -\mathrm{div}(f\varphi) - \frac{1}{2}\varphi \sum_{i=1}^{p} h_i^2,\ L_i\varphi = h_i\varphi,\quad i=1,\ldots,p\ ,\quad \forall\, \varphi \in C^\infty(R^n).$$

L'algèbre de Lie $\{L_o, L_1, \ldots, L_p\}_{LA}$ engendrée par $L_o, L_1, \ldots, L_p$ est de dimension finie si et seulement si dim Span $\mathcal{H} < +\infty$.

Preuve : On a : $[L_o, L_i]\varphi = \mathcal{L}_f(h_i)\varphi$, $[L_i, L_j]\varphi = 0 \quad \forall i \neq j$,

$[L_o,[L_o,L_i]]\varphi = -\mathcal{L}_f^2(h_i)\varphi$, etc.... $\forall\, \varphi \in C^\infty(R^n)$.

Donc $\{L_o, L_1, \ldots, L_p\}_{LA} = \mathrm{Span}\,\{L_o, \tilde{h}_i, \tilde{\mathcal{L}}_f^k(h_i) \mid i=1,\ldots,p,\ \forall k \geq 0\}$

où $\tilde{h}_i$ et $\tilde{\mathcal{L}}_f^k(h_i)$ désignent les opérateurs linéaires $\varphi \rightsquigarrow h_i\varphi$ et $\varphi \rightsquigarrow \mathcal{L}_f^k(h_i)\varphi$, $\forall\, \varphi \in C^\infty(R^n)$. Posons $\tilde{\mathcal{H}} = \{\tilde{h}_i, \tilde{\mathcal{L}}_f^k(h_i) \mid i=1,\ldots,p,\ \forall k \geq 1\}$. Or a évidemment dim Span $\tilde{\mathcal{H}}$ = dim Span $\mathcal{H}$ et :

$\{L_o, L_1, \ldots, L_p\}_{LA} = \mathrm{Span}\,(\{L_o\} \oplus \tilde{\mathcal{H}})$ et le résultat est démontré. ■

Corollaire 4 : Si dim Span $\mathcal{H} = r$, la réalisation minimale du filtre est de dimension r, et indistinguable de la réalisation linéaire suivante :

$$(3.16)\quad \begin{cases} da_t = -R'a_t dt + \sum_{i=1}^{p} M_i\, dy_t^i\ , & a_o = 0 \\ \pi_t^Y(x) = \exp\left(\sum_{i=1}^{r} a_t^i \theta_i(x)\right) \circ \tilde{\sigma}_t(x) \end{cases}$$

où $(\theta_1, \ldots, \theta_r)$ est une base de Span $\mathcal{H}$, R et $M_1, \ldots, M_p$ comme (iii) du théorème 2, et $\tilde{\sigma}_t$ donné par (2.11).

Preuve : en prouvant que (i) ⇒ (ii) au théorème 2, on a montré que si dim (filtre minimal) = r, alors dim Span $\mathcal{H} \leq r$.

Inversement, dans la preuve de (iii) ⇒ (i), on avait dim Span $\mathcal{H} = r$ d'où l'existence d'une réalisation de dim r donc :

dim Span $\mathcal{H} \geq$ dim (filtre minimal), d'où l'égalité.

Enfin, comme (3.16) est une réalisation de même dimension que dim Span $\mathcal{H}$ (voir preuve de (iii) ⇒ (i) du théorème 2) le résultat est établi. ■

Remarque 2 : contrairement à ce qui se passe en temps discret, R peut ne pas être inversible. C'est le cas dès qu'il existe k_o et i_o tels que $\mathcal{L}_f^{k_o}(\theta_{i_o}) \equiv 0$.

Si cette propriété à lieu $\forall i=1,\ldots,p$, à savoir :

$\forall i\ ,\ \exists\, k \in N$ tel que $\mathcal{L}_f^k(h_i) \equiv 0$, on dit que l'algèbre de Lie $\{L_o, L_1, \ldots, L_p\}_{LA}$ est nilpotente. Ce cas a été étudié dans [9].

Cependant la réalisation qui y est présentée n'est pas minimale en général puisqu'elle n'élimine pas les dépendances linéaires éventuelles entre les $\{\mathcal{L}_f^j(h_i) \quad \forall j \leq j(i), \forall i\}$, où $j(i)$ est la première puissance telle que $\mathcal{L}_f^{j(i)}(h_i) \equiv 0$. ■

Remarque 3 : Si l'on discrétise en temps, et si l'on pose :

$$\tilde{f}(x) \equiv x + \Delta t f(x) \quad , \text{ où } \Delta t \text{ est le pas de temps,}$$

les conditions de (iii), théorème 2 redonnent au premier ordre celles déjà obtenues dans [6], (iii), théorème 2.1 dans le cas $\eta(x) = I$. Cependant le filtre minimal à temps discret n'est pas le discrétisé du filtre minimal continu, alors que la densité conditionnelle à temps discret converge vaguement vers π_t^Y Y-p.s. . ■

Remarque 4 : Les méthodes développées dans [3] ne permettent pas d'obtenir que dim Span $\mathcal{H} < +\infty$. En effet, on obtiendrait que dim Span $\mathcal{H}(x) < +\infty \quad \forall x \in R^n$, ce qui est beaucoup plus faible puisqu'alors les relations de dépendance entre les $\theta_i(x)$ peuvent dépendre de x. Une telle situation est impossible dans notre cas puisque la matrice R dépendrait de x et l'équation (3.16) ne pourrait donc plus être un filtre. ■

IV - EXEMPLE : Capteur polynomial de degré p quelconque

$$(4.1) \quad \begin{cases} dx_t = Fx_t dt & x_t \in R,\ y_t \in R,\ v_t \in R,\ y_o = 0. \\ dy_t = \left(\sum_{i=0}^{p} \alpha_i x_t^i\right)dt + dv_t & p > o \text{ quelconque.} \end{cases}$$

On vérifie que : dim Span $\mathcal{H} = p+1$, une base de Span $\mathcal{H}$ étant :

$$(4.2) \quad \theta_o(x) = 1,\ \theta_1(x) = x, \quad \theta_2(x) = x^2, \ldots, \theta_p(x) = x^p .$$

En effet, il suffit de calculer $Fx \frac{d}{dx}\left(\sum_{i=0}^{p} \alpha_i x^i\right) = \sum_{i=1}^{p} \alpha_i i F x^i$ et d'itérer cette opération. Le degré du polynôme obtenu est toujours inférieur ou égal à p, d'où le résultat.

De plus, on a dans cette base :

$$h = \sum_{i=0}^{p} \alpha_i \theta_i = (\alpha_0,\ldots,\alpha_p)\theta, \quad \text{soit :} \quad M = \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_p \end{pmatrix}, \text{ avec } \theta = \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_p \end{pmatrix}$$

Enfin : $Fx \frac{d}{dx}(\theta_i) = iFx^i = iF\theta_i$, $i=0,\ldots,p$, donc R est la matrice diagonale.

$$R = \begin{pmatrix} {}^{0}F_{2F} & & 0 \\ & \ddots & \\ 0 & & {}^{p}F \end{pmatrix} \qquad \text{et le filtre minimal est donné par :}$$

$$(4.3) \quad \begin{pmatrix} da_t^0 \\ \vdots \\ da_t^p \end{pmatrix} = - \begin{pmatrix} {}^{0}F_{2F} & & 0 \\ & \ddots & \\ 0 & & {}^{p}F \end{pmatrix} \begin{pmatrix} a_t^0 \\ \vdots \\ a_t^p \end{pmatrix} dt + \begin{pmatrix} \alpha_0 \\ \vdots \\ \alpha_p \end{pmatrix} dy_t \ , \quad \begin{pmatrix} a_0^0 \\ \vdots \\ a_0^p \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} ,$$

$$(4.4) \quad \pi_t^Y(x) = \exp\Big(\sum_{i=0}^{p} a_t^i x^i\Big).\tilde{\sigma}_t(x),$$

$$(4.5) \quad \tilde{\sigma}_t(x) = \exp\Big[-\frac{1}{2}\int_0^t \Big(\sum_{k=0}^{p} \alpha_k e^{-kF(t-s)} x^k\Big)^2 ds\Big] e^{-Ft}\mu(e^{-Ft}x). \quad \blacksquare$$

V - IMMERSION DANS UN SYSTEME LINEAIRE

On adapte ici les définitions et résultats de [4].

Définition 2 : on dit que (1.1) est immergé dans le système

$$(5.1) \quad \begin{cases} dz_t = g(z_t)dt \ , \quad z_0 \in R^m \\ dy_t = \eta(z_t)dt + dv_t \end{cases}$$

avec $z_t \in R^m$ $\forall t$ et g, η C^∞ , s'il existe une application θ de R^n dans R^m de classe C^∞ telle que si $\theta(x_0) = z_0$, alors :

$$(5.2) \qquad h(X_t(x_0)) = \eta(Z_t(z_0)) \quad \forall t \geq 0$$

où $Z_t(z_0)$ est le flot solution de $\frac{d}{dt} Z_t(z_0) = g(Z_t(z_0))$, $Z_0(z_0) = z_0$. $\blacksquare$

Théorème 3 : Une condition nécessaire et suffisante pour que (1.1) admette un FDF est qu'il soit immergeable dans un système linéaire de la forme :

$$\begin{cases} dz_t = Gz_t dt \\ dy_t = Kz_t dt + dv_t \end{cases}$$

Preuve : si (1.1) admet un FDF alors, par (iii) du théorème 2, on pose $\theta(x_o) = z_o$, $z_t = \theta(X_t(x_o))$, $h(X_t(x_o)) = M'\theta(X_t(x_o))$, et $\frac{d}{dt} z_t = \mathcal{L}_f(\theta)(X_t(x_o)) = R\ \theta(X_t(x_o)) = Rz_t$ donc $G = R$, $K = M'$, $z_t = \theta(X_t(x_o))$ et $h(X_t(x_o)) = Kz_t = Ke^{Rt}z_o$, l'immersion est donc réalisée.
Inversement, si $h(X_t(x_o)) = KZ_t(z_o) = Ke^{Gt}z_o \quad \forall t \geq o$, on a : $h(x_o) = Kz_o = K\ \theta(x_o)$. Puis, en dérivant par rapport à t :

$$\mathcal{L}_f(h)(X_t(x_p)) = KGe^{Gt}z_o$$

ou, pour $t=o$: $\mathcal{L}_f(h)(x_o) = KG\theta(x_o)$. De même, on voit facilement que $\mathcal{L}_f^k(h)(x_o) = KG^k\theta(x_o) \quad \forall k \geq 1$, et ceci pour tout $x_o \in R^n$ donc $\theta = (\theta_1,\ldots,\theta_m)$ engendre Span $\mathcal{H}$ et dim Span $\mathcal{H} \leq m$, CQFD. ■

Remarque 5 : De même qu'en [6] dans le cas $\eta(x) = I$, tous les systèmes (1.1) qui admettent un filtre de dimension finie peuvent être transformés par changement de coordonnées nonlinéaires en un système linéaire. Cependant pratiquement, aussi bien pour des questions de dimension que de stabilité numérique, il est souvent plus intéressant de travailler en nonlinéaire. ■

Conclusion : La classe des systèmes (1.1) étudiée ici, ou son équivalent à temps discret (étudiée en [6]), a pour spécificité de ne pas avoir de bruits entrant dans la dynamique. C'est cette propriété qui, aussi bien en continu qu'en discret, permet d'obtenir des FDF sous des conditions moins restrictives que dans le cas des bruits de dynamique. Notons que la condition de finitude de l'algèbre de Lie de l'équation de Zakaï est obtenue ici, comme dans les quelques exemples connus avec bruits de dynamique ([1],[2],[3],[5],[7],[8]).

REFERENCES

[1] V. BENES : Exact finite dimensional filters for certain diffusions with nonlinear drifts. Stochastics 5 (1981) p. 65-92.

[2] R.W. BROCKETT : Remarks on finite dimensional nonlinear estimation. C. Lobry ed. "Analyse des Systèmes". Astérique 75-76 (SMF) p. 47-56.

[3] M. CHALEYAT-MAUREL, D. MICHEL : Des résultats de non existence de filtre de dimension finie. A paraître.

[4] M. FLIESS, I. KUPKA : A finiteness criterion for nonlinear input-output differential systems. SIAM J. Cont. Optimiz. 21,5, (1983) p. 721-728.

[5] M. HAZEWINKEL, S.I. MARCUS : On Lie algebras and finite dimensional filtering. Stochastics 7, (1982) p. 29-62.

[6] J. LEVINE, G. PIGNIE : Exact finite dimensional filters for a class of nonlinear discrete-time systems. A paraître.

[7] S.I. MARCUS, S.K. MITTER, D. OCONE : Finite dimensional nonlinear estimation for a class of systems in continuous and discrete time, in Analysis and optimization of stochastic systems. p. 387-406, Acadamic Press (1980)

[8] S.K. MITTER : Geometric theory of nonlinear filtering. Outils et modèles mathématiques pour l'automatique, l'analyse des systèmes et le traitement du signal. Vol. 3, p. 37-60. Editions du CNRS (1983).

[9] Z.S. ROTH, K.A. LOPARO : Optimal filter realization for a class of nonlinear systems with finite dimensional estimation algebra. Syst. Cont. Letters, 4, 1, (1984) p. 23-26.

PARTIE IV

Méthodes de graphe pour le découplage et le rejet de perturbations des systèmes nonlinéaires

RESUME DE LA IVème PARTIE

Méthodes de graphe pour le découplage et le rejet de perturbations des systèmes nonlinéaires

Cette partie est consacrée à l'étude du découplage et du rejet des perturbations pour les systèmes nonlinéaires de la forme:

$$(\Sigma) \qquad \begin{cases} \dot{x} = f_o(x) + \sum_{i=1}^{N} u^i f_i(x) + \sum_{i=1}^{M} w^i g_i(x) \\ y_1 = h_1(x) \\ \vdots \\ y_p = h_p(x) \end{cases}$$

où x évolue sur une variété analytique, où f_o , f_1 ,..., f_N et g_1 ,..., g_M sont des champs de vecteurs analytiques, et où les fonctions $h_1,\ldots,h_p$ sont analytiques, chaque y_i , i=1,...,p étant scalaire.

On montre que les algorithmes de découplage existants peuvent être améliorés de manière sensible en évitant un grand nombre de calculs de dérivées de Lie des sorties h_i le long des champs de vecteurs f_o , f_j , g_i, car celles-ci peuvent s'interprêter de manière élémentaire sur le graphe du système (Σ). Plus précisément, on dérive les sorties le long de f_o , f_j et g_j pour connaître les retards minimaux du système (Σ). Or ces retards se lisent directement, tout au moins génériquement, comme la longueur du plus court chemin reliant l'une des entrées à la ième sortie.

L'algorithme est présenté dans le 1er article.

Le second article est consacré à l'application de cet algorithme au découplage d'un bras de robot.

Les commandes ainsi obtenues dépendent de paramètres que l'on peut optimiser pour obtenir un comportement entrée-sortie donné, en particulier linéaire stable.

Lecture Notes in Control and Information Sciences

Edited by A.V. Balakrishnan and M. Thoma

58

Mathematical Theory of Networks and Systems

Proceedings of the MTNS-83 International Symposium
Beer Sheva, Israel, June 20–24, 1983

Edited by P.A. Fuhrmann

Springer-Verlag
Berlin Heidelberg New York Tokyo 1984

A FAST GRAPH THEORETIC ALGORITHM FOR THE FEEDBACK DECOUPLING PROBLEM OF NONLINEAR SYSTEMS [+]

A. KASINSKI [*] and J. LEVINE [**]

ABSTRACT : The feedback decoupling problem of nonlinear systems is actually well understood in a theoretic point of view. However, to compute the decoupling feedbacks, the only method known by the authors, consists in using a formal derivation program to check if differential expressions are null [3]. We give a generic interpretation of these expressions in terms of the graph of the system in the sense of [6], and deduce a faster algorithm using the minimal length of the paths joining one of the inputs to the i^{th} output.

(*) Institut Automatyki Politechnika Poznanska
60965 POZNAN, ul. Piotrowo, 3A, POLAND.

(**) Centre d'Automatique et Informatique
Ecole Nationale Supérieure des Mines de Paris
35, rue St Honoré
77305 FONTAINEBLEAU - FRANCE.

(+) The first author was supported by a scholarship of the Ministère des Relations Extérieures de la République Française.

I - The feedback decoupling problem

We consider a linear-analytic system, given in local coordinates, by :

$$(\Sigma)\quad \begin{cases} \dot{x} = f_o(x) + \sum_{i=1}^{N} u^i f_i(x) + \sum_{i=1}^{M} w^i g_i(x) \\ y_i = h_i(x)\ ,\ i = 1, \ldots, p \\ z_i = k_i(x)\ ,\ i = 1, \ldots, q \end{cases}$$

where x belongs to a connected n-dimensional analytic manifold X, $u = (u^1, \ldots, u^N)^T$ are the input functions, $h_1, \ldots, h_p$ and $k_1, \ldots, k_q$ are the output functions, analytic on X, and where :

$$F_i(x) = \sum_{j=1}^{n} f_i^j(x) \frac{\partial}{\partial x_j}\ ,\ i = 0, \ldots, N$$
$$G_i(x) = \sum_{j=1}^{n} g_i^j(x) \frac{\partial}{\partial x_j}\ ,\ i = 1, \ldots, M \qquad (1)$$

are analytic vector fields on X.

The feedback decoupling problem consists in finding analytic functions (α^i, β_i^j), $i=1,\ldots,N$, $j=1,\ldots,N$, eventually defined on an open subset $\mathcal{O}$ of X such that

the feedback control :

$$u^i(x) = \alpha^i(x) + \sum_{j=1}^{N} \beta_i^j(x) v_j \quad , \quad i = 1, \ldots, N \tag{2}$$

makes the p outputs $y_1, \ldots, y_p$ locally independent of w^i, $i = 1, \ldots, M$. We shall denote $\hat{F}_i$, $i = 0, \ldots, N$, the vector fields obtained by the feedback (2) :

$$\begin{aligned} \hat{f}_o(x) &= f_o(x) + \sum_{i=1}^{N} \alpha^i(x) f_i(x) \quad , \quad \hat{F}_o(x) = \sum_{j=1}^{n} \hat{f}_o^j(x) \frac{\partial}{\partial x_j} \\ \hat{f}_i(x) &= \sum_{j=1}^{N} \beta_i^j(x) f_j(x) \quad , \quad \hat{F}_i(x) = \sum_{j=1}^{n} \hat{f}_i^j(x) \frac{\partial}{\partial x_j} \quad , \quad i=1,\ldots, N. \end{aligned} \tag{3}$$

The problem is actually well understood and the differential geometric methods [4] together with the algebraic ones [1] draw an almost complete picture of the theoretic solution. In the geometric approach of the "structural" decoupling, we introduce the maximal involutive distribution $\mathcal{B}$ of constant rank, which is $(F_o, F_1, \ldots, F_N)$-invariant. Isidori, Krener, Gori-Giorgi and Monaco [4] have proved the following :

Theorem 1 : The structural decoupling problem has a local solution if and only if :

$$\text{span} \{ G_1, \ldots, G_M \} \subset \mathcal{B} \subset \bigcap_{i=1}^{p} \ker dh_i \tag{4}$$

Furthermore, $\mathcal{B}$ can be obtained by the following induction (see [5]) :

$$\mathcal{B}_o = \text{span} \{dh_1, \ldots, dh_p\} \tag{5}$$

$$\mathcal{B}_k = \sum_{i=0}^{N} L_{F_i} \mathcal{B}_{k-1} + \mathcal{B}_{k-1} \tag{6}$$

where L_{F_i} is the Lie derivative with respect to the vector field F_i, and

$$\mathcal{B} = \Big(\bigcup_{k \geq 0} \mathcal{B}_k \Big)^{\perp} \quad \blacksquare \tag{7}$$

The algebraic methods, using Fliess' input-output map representation, give a "functional" point of view : in place of a distribution, one looks for a module $\mathcal{M}$ of vector fields, playing the same role as the distribution $\mathcal{B}$ but eventually with a non constant rank (see [1]). Claude [1] has proved the following :

Theorem 2 : The outputs $y_1, \ldots, y_p$ are decoupled with respect to $w^1, \ldots, w^M$, if and only if there exists an analytic module $\mathcal{M}$ which is also a Lie subalgebra of vector fields on X such that :

$$\forall i : [\hat{F}_i, \mathcal{M}] \subset \mathcal{M}, \text{ and } \text{span}\{G_1, \ldots, G_M\} \subset \mathcal{M} \subset \bigcap_{i=1}^{p} \ker dh_i \tag{8}$$

with $\hat{F}_i$ defined by (3). $\blacksquare$

Furthermore, α and β can be computed in a purely algebraic way (that is to say without solving differential or partial differential equations) by the procedure described hereafter.

For this purpose, we need the :

Definition 1 : The characteristic number ρ_i of order i is the unique integer satisfying :

$$\exists j \in \{1,\dots,N\} : F_j F_o^{\rho_i} h_i \not\equiv 0, \text{ and :} \tag{9}$$

$$\forall j \in \{1,\dots,N\}, \forall m \in \{0,\dots,\rho_i-1\}, \ F_j F_o^{m} h_i \equiv 0. \tag{10}$$

If $F_j F_o^m h_i \equiv 0 \ \forall j, \forall m$, we set $\rho_i = +\infty$, and if $\exists j : F_j h_i \not\equiv 0$, $\rho_i = 0$.

Remark that $F_o^m h_i = F_o(F_o^{m-1} h_i)$ is a polynomial of differentials of h_i up to the order m, and that $F^o h_i \equiv h_i$. ρ_i can be interpreted as the minimal number of integrations such that y_i is affected by one of the u_j.

To compute α and β, we introduce the following quantities :

$$\Delta_i^j(x) = F_j(x) F_o^{\rho_i}(x) h_i(x) \ , \ i = 1, \dots, p \ , \ j = 1, \dots, N \tag{11}$$

$$\varphi_i(x) = \tilde{\varphi}_i(h_i(x), F_o(x)h_i(x), \dots, F_o^{\rho_i}(x)h_i(x)) - F_o^{\rho_i+1}(x), i=1,\dots,p \tag{12}$$

$$\psi_i^j(x) = \tilde{\psi}_i^j(h_i(x), F_o(x)h_i(x), \dots, F_o^{\rho_i}(x)h_i(x)) \ , \ i = 1, \dots, p \ , \ j = 1, \dots, N \tag{13}$$

with $\tilde{\varphi}_i$ and $\tilde{\psi}_i^j$ arbitrary analytic functions.

Let us call : Δ the $p \times N$ matrix-valued analytic function whose $(i,j)^{th}$ element is Δ_i^j, $\varphi = (\varphi_1, \dots, \varphi_p)^T$, and ψ the $p \times N$ matrix-valued analytic function whose $(i,j)^{th}$ element is ψ_i^j.

Theorem 3 : If $G_k F_o^m h_i \equiv 0 \ \forall k \in \{1,\dots,M\}$, $\forall m \leq \rho_i$, a necessary and sufficient condition for (α,β) to realize the local functional decoupling of (Σ), is that (α,β) locally solve the system :

$$\begin{aligned} \Delta\alpha &= \varphi \\ \Delta\beta &= \psi \end{aligned} \tag{14}$$

In this case, the change of variables :

$$X_o^i = h_i, \dots, X_{\rho_i}^i = F_o^{\rho_i} h_i \ , \ i = 1, \dots, p \ , \tag{15}$$

puts the system (Σ) locally into the form :

$$\begin{cases} \dot{X}_o^i = X_1^i \\ \quad \vdots \\ \dot{X}_{\rho_i-1}^i = X_{\rho_i}^i \\ \dot{X}_{\rho_i}^i = \varphi_i(X_o^i, \dots, X_{\rho_i}^i) + \sum_{j=1}^{N} \psi_i^j(X_o^i, \dots, X_{\rho_i}^i) v_j \\ y_i = X_o^i \\ \qquad i = 1, \dots, p. \ \blacksquare \end{cases} \tag{16}$$

Clearly, this procedure involves a huge amount of formal calculus, especially to determine the characteristic numbers ρ_i, $i = 1, \ldots, p$: one must differentiate ρ_i times the expressions h_i, $F_o h_i$, etc..., whose complexity is growing very fast, and then check if $F_J F_o^m h_i$ is null or not. A program has been developped by Claude and Dufresne [3], using the language MACSYMA, to compute these formal expressions.

The aim of this paper is to introduce a faster method to compute ρ_i with the minimal number of formal differentiations : for this purpose, we shall prove that the numbers ρ_i can generically be very easily obtained on the system's graph. We shall also give a lower bound ν_i for ρ_i in the non-generic case, still obtained from the graph, and prove that either $\nu_i \leq n-1$ or $\nu_i = \rho_i = +\infty$ These results are finally synthesized in an algorithm to compute (α,β).

II-The system's graph

As in [6], we introduce the following system's graph :

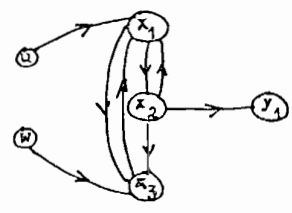
Definition 2 : We call Γ the system's graph of Σ in a given open subset $\mathcal{O}$ of X with given local coordinates, the oriented graph whose input-nodes are $(u^1,\ldots,u^N,w^1,\ldots,w^M)$, whose intermediate-nodes are the state variables $(x_1,\ldots,x_n)$, and whose output-nodes are $(y_1,\ldots,y_p)$. The oriented arcs of Γ are obtained as follows :

- There exists an oriented arc joigning u^i to x_k iff $f_i^k(x) \not\equiv 0$ in $\mathcal{O}$ $i = 1,\ldots,N$, $k=1,\ldots,n$, and joigning w^l to x_k iff $g_i^k(x) \not\equiv 0$ in $\mathcal{O}$ $i = 1, \ldots, M$, $k = 1, \ldots, n$.
- There exists an oriented arc joining x_k to x_j iff $\dfrac{\partial f_o^j}{\partial x_k}(x) \not\equiv 0$ in $\mathcal{O}$, $j, k = 1, \ldots, n$.
- There exists an oriented arc joining x_k to y_i iff $\dfrac{\partial h_i}{\partial x_k}(x) \not\equiv 0$ in $\mathcal{O}$, $i = 1, \ldots, p$, $k = 1, \ldots, n$. ■

Definition 3 : We call $d(u^j,y_i)$ the minimal number of oriented arcs of Γ forming an oriented path joining u^j to y_i, and $d_i = \underset{1\leq j\leq N}{\text{Min}}\, d(u^j,y_i)$ ■

An introductory example : $n = 3$, $N = 1$, $M = 1$, $p = 1$, $f_o^1(x) \not\equiv 0$, $\dfrac{\partial f_o^2}{\partial x_1}(x_1,x_2) \not\equiv 0$.

$$(17)\quad \begin{cases} \dot{x}_1 = f_o^1(x_1,x_2,x_3) + u\, f_1(x_1,x_2,x_3) \\ \dot{x}_2 = f_o^2(x_1,x_2) \\ \dot{x}_3 = f_o^3(x_1,x_2,x_3) + w g_1(x_1,x_2,x_3) \\ y_1 = h(x_2) \end{cases}$$



The system's graph Γ

It can be easily seen that $d_1 = d(u,y_1) = 3$, $d(w,y_1) = 4$.
We shall prove that we can predict that, generically, $\rho_1 = d_1 - 2 = 1$, and that $d(w,y_1) > d_1$ implies $G_1 h \equiv 0$ and $G_1 F_o h \equiv 0$. To check our assertion, let us go back to (9) and (10), and compute ρ_1. We first check that $(F_1 h)(x) = f_1(x) \frac{\partial h}{\partial x_3}(x_2) \equiv 0$, $(G_1 h)(x) = g_1(x) \frac{\partial h}{\partial x_3}(x_2) \equiv 0$. Then : $(F_o h)(x) = f_o^1(x) \frac{\partial h}{\partial x_1}(x_2) + f_o^2(x_1,x_2) \frac{\partial h}{\partial x_2}(x_2)$ + $f_o^3(x) \frac{\partial h}{\partial x_3}(x_2) = f_o^2(x_1,x_2) \frac{\partial h}{\partial x_2}(x_2)$, $(F_1 F_o h)(x) = f_1(x) \frac{\partial f_o^2}{\partial x_1}(x_1,x_2) \frac{\partial h}{\partial x_2}(x_2) \not\equiv 0$, and thus $\rho = 1$;
Finally, we also have $G_1 F_o h = g_1 \left(\frac{\partial f_o^2}{\partial x_3} \frac{\partial h}{\partial x_2} + f_o^2 \frac{\partial^2 h}{\partial x_3 \partial x_2} \right) \equiv 0$, as claimed above.

Thus, almost without computations, ρ_1 and the relations $G_1 h \equiv 0$ and $G_1 F_o h \equiv 0$, can be deduced from the system's graph (we only need to compute $F_1 F_o h$!). Clearly, the system's graph synthesizes the structure of the interactions of the input and output variables versus integration of the state variables. Thus, it is not surprising that, in general, but generically only, the minimal length d_i represents the minimum number of integrations for the inputs to affect y_i, namely ρ_i, up to a constant equal to 2 since the first and last arcs do not represent integrations.

Remark : in Γ, we do not take into account the fact that $f_1, \ldots, f_N, g_1, \ldots, g_M$ depend on $x_1, \ldots, x_n$ or not. For our purpose these interactions do not play any role in generic situations and, if they play a role in non-generic cases, the profit of the graph's method vanishes, as will be seen after.

III-The characteristic numbers ρ_i, their lower bounds ν_i, and the system's graph

Besides the characteristic numbers ρ_i, we shall introduce the numbers ν_i defined as follows :

Definition 4 : The number ν_i, $i = 1, \ldots, p$, is the unique integer satysfying :

$$\exists j \in \{1 \ldots, N\}, \exists k_{\nu_i}, \ldots, k_o \in \{1, \ldots, n\} \text{ such that :}$$

$$f_j^{k_{\nu_i}} \frac{\partial f_o^{k_{\nu_i - 1}}}{\partial x_{k_{\nu_i}}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_i}{\partial x_{k_o}} \not\equiv 0 \text{ in } \mathcal{O}, \tag{18}$$

and : $\forall j \in \{1, \ldots, N\}$, $\forall r < \nu_i$, $\forall k_r, \ldots, k_o \in \{1, \ldots, n\}$, we have :

$$f_j^{k_r} \frac{\partial f_o^{k_{r-1}}}{\partial x_{k_r}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_i}{\partial x_{k_o}} \equiv 0 \text{ in } \mathcal{O} \quad \blacksquare \tag{19}$$

Now we can state the main result :

Theorem 4 : . $\nu_i = d_i - 2$, $i = 1, \ldots, p$ (20)

. $\nu_i \leq \rho_i$ and $\nu_i = \rho_i$ generically $i = 1, \ldots, p$. (21)

By generically, we mean : for every system Σ whose coefficients $f_o, f_1, \ldots, f_N, g_1, \ldots, g_M, h_1 \ldots, h_p$ lie outside a closed subset, with

empty interior, of the space of analytic vector-valued functions on $\mathcal{O} \subset X$.

Proof of (20) : From the definition of ν_i, since :

$$f_j^{k_{\nu_i}} \frac{\partial f_o^{k_{\nu_i - 1}}}{\partial x_{k_{\nu_i}}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_i}{\partial x_{k_o}} \not\equiv 0 \text{ in } \mathcal{O},$$

we also have $\frac{\partial h_i}{\partial x_{k_o}} \not\equiv 0$, which means that there is an arc from x_{k_o} to y_i , $\frac{\partial f_o^{k_o}}{\partial x_{k_1}} \not\equiv 0$, which means that the arc (x_{k_1}, x_{k_o}) belongs to Γ, etc..., and $f_j^{k_{\nu_1}} \not\equiv 0$ so that $(u_j, x_{k_{\nu_1}}) \in \Gamma$. Finally $(u_j, x_{k_{\nu_i}}, x_{k_{\nu_i - 1}}, \ldots, x_{k_1}, x_{k_o}, y_i)$ is an oriented path of Γ joining u_j to y_i, of length $\nu_i + 2$. Thus, since d_i is the minimum, $\nu_i + 2 \geq d_1$

On the other hand, suppose that there exists a path of Γ of length $r < \nu_i + 2$, joining u_k to y_1 By definition of Γ, and using the same argument as before, we must have :

$$f_k^{k_r} \frac{\partial f_o^{k_{r-1}}}{\partial x_{k_r}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_i}{\partial x_{k_o}} \not\equiv 0 \quad \text{in } \mathcal{O}$$

but this is contrary to the definition of ν_i, and finally : $d_i = \nu_i + 2$ and (20) is proved. ■

To prove (21), we need the following notations :

Let us introduce the subsets Γ_i^j and $\tilde{\Gamma}_i^j$ of $\{1, \ldots, n\}$, $\forall j \geq 0$, by induction :

$$\Gamma_i^o = \{k_o \in \{1, \ldots, n\} \mid \frac{\partial h_i}{\partial x_{k_o}} \not\equiv 0 \text{ in } \mathcal{O}\} = \tilde{\Gamma}_i^o , \tag{22}$$

and , $\forall j \geq 1$:

$$\Gamma_i^j = \{k_j \in \{1, \ldots, n\} - \tilde{\Gamma}_i^{j-1} \mid \exists k_{j-1} \in \tilde{\Gamma}_1^{j-1}, \ldots, \exists k_o \in \tilde{\Gamma}_1^o \text{ s.t. } \frac{\partial f_o^{k_{j-1}}}{\partial x_{k_j}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_1}{\partial x_{k_o}} \not\equiv 0 \text{ in } \mathcal{O}\},$$

$$\tilde{\Gamma}_i^j = \Gamma_i^j \cup \tilde{\Gamma}_i^{j-1} \tag{23}$$

Let us also denote by $\pi_j(k_o, \ldots, k_j)$ the analytic function in $\mathcal{O}$ defined by :

$$\pi_i(k_o, \ldots, k_j) = \frac{\partial f_o^{k_{j-1}}}{\partial x_{k_j}} \cdot \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_i}{\partial x_{k_o}} \tag{24}$$

Γ_j^j is thus the set of state variables x_{k_j} that appear for the first time in $\pi_i(k_o, \ldots, k_j)$ (product of $j+1$ factors), and that do not appear in any of the $\pi_i(k_o, \ldots, k_r)$, $r < j$.

Lemma 1 :

$$F_o^m h_i = \sum_{\substack{k_o \in \Gamma_i^o \\ \vdots \\ k_{m-1} \in \Gamma_i^{m-1}}} f_o^{k_{m-1}} . \pi_i(k_o, \ldots, k_{m-1}) + \Phi_m(\tilde{\Gamma}_i^{m-2}) \qquad (25)$$

where $\Phi_m(\tilde{\Gamma}_i^{m-2})$ is a linear combination with analytic coefficients of all the terms of the form : $\pi_i(k_o, \ldots, k_{m-1})\ \forall k_o, \ldots, k_{m-1} \in \tilde{\Gamma}_i^{m-2}$, and :

$$\frac{\partial^{m-1-r}}{\partial x_{k_{m-1}} \ldots \partial x_{k_{r+1}}} \pi_i(k_o, \ldots, k_r) \qquad \forall r < m-1 \ , \ \forall k_o, \ldots, k_{m-1} \in \tilde{\Gamma}_i^{m-2} \ ,$$

with the convention : $\Phi_1(\tilde{\Gamma}_i^{-1}) \equiv 0$.

Proof : By induction on m. The lemma is trivial for $m = 1$, since :

$$F_o h_i = \sum_{k_o \in \Gamma_o^o} f_o^{k_o} \frac{\partial h_i}{\partial x_{k_o}} \ , \text{ and thus } \Phi_1 \equiv 0.$$

Suppose it holds up to m. For $m + 1$:

$$F_o^{m+1} h_i = F_o(F_o^m h_i) = \sum_{k_m} f_o^{k_m} \frac{\partial}{\partial x_{k_m}} (F_o^m h_i)$$

$$= \sum_{k_m} f_o^{k_m} \frac{\partial}{\partial x_{k_m}} \Big(\sum_{\substack{k_o \in \Gamma_i^o \\ \vdots \\ k_{m-1} \in \Gamma_i^{m-1}}} f_o^{k_{m-1}} . \pi_i(k_o, \ldots, k_{m-1}) + \Phi_m(\tilde{\Gamma}_i^{m-2}) \Big)$$

$$= \sum_{k_m} \Big(f_o^{k_m} \frac{\partial f_o^{k_{m-1}}}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1}) + f_o^{k_m} f_o^{k_{m-1}} \frac{\partial}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1}) + f_o^{k_m} \frac{\partial}{\partial x_{k_m}} \Phi_m(\tilde{\Gamma}_i^{m-2}) \Big)$$

$$= \sum_{k_m \in \Gamma_i^m} f_o^{k_m} \frac{\partial f_o^{k_{m-1}}}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1}) + \sum_{k_m \in \tilde{\Gamma}_i^{m-1}} f_o^{k_m} \frac{\partial f_o^{k_{m-1}}}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1})$$

$$+ \sum_{k_m \in \Gamma_i^m} f_o^{k_m} \Big(f_o^{k_{m-1}} \frac{\partial}{\partial x_{k_m}} \pi_1(k_o, \ldots, k_{m-1}) + \frac{\partial}{\partial x_{k_m}} \Phi_m(\tilde{\Gamma}_i^{m-2}) \Big)$$

$$+ \sum_{k_m \in \tilde{\Gamma}_i^{m-1}} f_o^{k_m} \Big(f_o^{k_{m-1}} \frac{\partial}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1}) + \frac{\partial}{\partial x_{k_m}} \Phi_m(\tilde{\Gamma}_i^{m-2}) \Big) \ .$$

Clearly $f_o^{k_m} \frac{\partial f_o^{k_{m-1}}}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1}) \quad f_o^{k_m} \pi_1(k_o, \ldots, k_m)$,

$\frac{\partial}{\partial x_{k_m}} \pi_i(k_o, \ldots, k_{m-1}) \equiv 0 \quad \forall k_m \in \Gamma_i^m$ by definition of Γ_i^m,

$\frac{\partial}{\partial x_{k_m}} \Phi_m(\tilde{\Gamma}_1^{m-2}) \equiv 0 \quad \forall k_m \in \Gamma_i^m$ for the same reason, and it is easy to see that :

$$\Phi_{m+1}(\tilde{\Gamma}_i^{m-1}) \stackrel{\text{def}}{=} \sum_{k_o,\ldots,k_m \in \tilde{\Gamma}_i^{m-1}} f_o^{k_m}(\pi_i(k_o,\ldots,k_m)) + f_o^{k_{m-1}} \frac{\partial}{\partial x_{k_m}} \pi_1(k_o,\ldots,k_{m-1}) + \frac{\partial}{\partial x_{k_m}} \Phi_m(\tilde{\Gamma}_i^{m-2}))$$

satisfies the desired property, and thus the Lemma is proved. ∎

Proof of (21) : a) let us prove that $\rho_i \geq \nu_1$

From Lemma 1, we have :

$$F_j F_o^{\rho_i} h_i = \sum_k f_j^k \frac{\partial}{\partial x_k} \Big(\sum_{k_o \in \Gamma_i^o} f_o^{k_{\rho_i - 1}} \pi_1(k_o, \ldots, k_{\rho_{i-1}}) + \Phi_{\rho_i}(\tilde{\Gamma}_1^{\rho_1 - 2}) \Big),$$

$$\cdot$$

$$k_{\rho_{i-1}} \in \Gamma_1^{\rho_i - 1}$$

and, using the same arguments as in Lemma 1, we find :

$$F_j F_o^{\rho_i} h_i = \sum_{k_o \in \Gamma_i^o} f_j^{k_{\rho_i}} \pi_i(k_o, \ldots, k_{\rho_1}) + \sum_{k_o, \ldots, k_{\rho_i} \in \tilde{\Gamma}_1^{\rho_i - 1}} f_j^{k_{\rho_i}}(\pi_i(k_o, \ldots, k_{\rho_i}) +$$

$$\cdot$$

$$\cdot$$

$$\cdot$$

$$k_{\rho_i} \in \Gamma_i^{\rho_i}$$

$$+ f_o^{k_{\rho_1 - 1}} \frac{\partial}{\partial x_{k_{\rho_i}}} \pi_1(k_o, \ldots, k_{\rho_{i-1}}) + \frac{\partial}{\partial x_{k_{\rho_i}}} \Phi_{\rho_i}(\tilde{\Gamma}_i^{\rho_i - 2})) \qquad (26)$$

Since $\exists j \in \{1, \ldots, N\}$ such that $F_j F_o^{\rho_i} h_i \not\equiv 0$, there is at least one term of (26) not identically 0. If the first term is non 0, namely $f_j^{k_{\rho_1}} \pi_1(k_o, \ldots, k_{\rho_i}) \not\equiv 0$, by definition of ν_1, we have $\rho_1 \geq \nu_1$.

If the first term is 0 but not the second, namely $f_j^{k_{\rho_1}} \pi_1(k_o, \ldots, k_{\rho_1}) \not\equiv 0$ for $k_o, \ldots, k_{\rho_1} \in \tilde{\Gamma}_i^{\rho_i - 1}$, using the definition of $\tilde{\Gamma}_1^{\rho_1 - 1}$, this means that there is also a subproduct of $\pi_i(k_o, \ldots, k_{\rho_1})$ which is non zero, and without loss of generality, it can be assumed that $f_j^{k_{\rho_1}} \pi_1(k_1, \ldots, k_{\rho_1}) \not\equiv 0$, and thus $\rho_i - 1 \geq \nu_1$.

Applying the same argument to each term of (26), we finally have $\rho_1 \geq \nu_1$

b) Let us now prove that $\rho_i = \nu_1$ generically.

Suppose that $\rho_i > \nu_1$ This means that :

$$F_j F_o^k h_i \equiv 0 \quad \forall j = 1, \ldots, N \quad , \quad \forall k = \nu_i, \ldots, \rho_i - 1 \qquad (27)$$

But (27) is a system of $N(\rho_i - \nu_i)$ non trivial partial differential equations in the variables f_j^k, f_o^r, h_i and their partial derivatives, system which is integrable in $\mathcal{O}$ since the coefficients of Σ give an analytic solution, by assumption. Thus, it is well-known that the set of solutions of (27) is a closed subset with empty interior of the space of analytic vector-valued functions on $\mathcal{O}$, endowed with the uniform topology on the compacts of $\mathcal{O}$. Thus, $\rho_i = \nu_i$ for almost every Σ, and the proof of Theorem 4 is achieved. ∎

Corollary : If $\nu_i > n-1$, then $\rho_i = \nu_i = +\infty$.

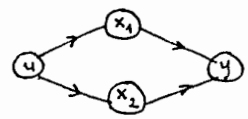
Proof : Let us first prove that $\Gamma_i^n = \emptyset$, $\forall i = 1, \ldots, p$.

Two cases can happen : either $\Gamma_i^{n-1} = \emptyset$, or $\Gamma_i^{n-1} \neq \emptyset$.
From (23), it is clear that if $\Gamma_i^{n-1} = \emptyset$, then $\Gamma_i^n = \emptyset$ also.
On the other hand, if $\Gamma_i^{n-1} \neq \emptyset$, once more from (23), we see that $\Gamma_i^{n-2} \neq \emptyset, \ldots, \Gamma_i^o \neq \emptyset$, and, since there is at least one element in each Γ_i^j, there must be n elements in :

$\tilde{\Gamma}_i^{n-1} = \Gamma_i^o \cup \Gamma_i^1 \cup \ldots \cup \Gamma_i^{n-1}$ But $\Gamma_i^n = \{1, \ldots, n\} - \tilde{\Gamma}_i^{n-1} = \emptyset$, and we have proved that $\Gamma_i^n = \emptyset$ in every case.
Now, if $\nu_i > n-1$, one has $\pi_i(k_o, \ldots, k_r) = 0 \quad \forall k_o, \ldots, k_r, \ \forall r \leq n-1$, and since $\Gamma_i^n = \emptyset$, one has also $\pi_i(k_o, \ldots, k_r) = 0 \quad \forall r \geq n$, and thus $\nu_i = +\infty$.
Finally, since $\rho_i \geq \nu_i$, the result is proved. ∎

Remark 1 : From the corollary, we conclude that ν_i is computed in at most $n-1$ steps, and, generically, the same holds for ρ_i. The result for ρ_i was proved in [2].
However, it is remarkable that one can have $\nu_i < n-1$ whereas $\rho_i = +\infty$ as the following example proves :

$$\begin{cases} \dot{x}_1 = u x_1 \\ \dot{x}_2 = -u x_2 \\ y = x_1 x_2 \end{cases}$$



- Γ -

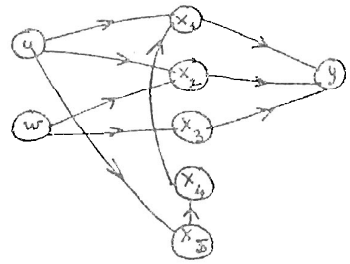
It is very easy to see that $\nu = 0$, but $\rho = +\infty$ since

$$F_1 h = x_1 \frac{\partial(x_1 x_2)}{\partial x_1} - x_2 \frac{\partial(x_1 x_2)}{\partial x_2} \equiv 0 ,$$

and $F_1 F_o h \equiv 0$ since $F_o \equiv 0$. Finally, this suffices to prove that $\rho = +\infty$ since $n = 2$. ∎

Remark 2 : It would be a nice result, if $\rho_i > \nu_i$, that there exists a (non minimal) oriented path from one of the u^j to y_i of length $\rho_i + 2$.
Unfortunately, this is not true, as the next example shows :

$$\begin{cases} \dot{x}_1 = x_4 + ux_1 \\ \dot{x}_2 = x_2 w \quad y_2 u \\ \dot{x}_3 = -x_3 w \\ \dot{x}_4 = x_5 \\ \dot{x}_5 = u \\ y = x_1 x_2 x_3 \end{cases}$$



— Γ —

We have $\nu = 0$, whereas $\rho = 1$

$$F_1 h = x_1 \frac{\partial}{\partial x_1}(x_1 x_2 x_3) - x_2 \frac{\partial}{\partial x_2}(x_1 x_2 x_3) = 0, \quad G_1 h = x_2 \frac{\partial}{\partial x_2}(x_1 x_2 x_3) - x_3 \frac{\partial}{\partial x_3}(x_1 x_2 x_3) \equiv 0,$$

$$F_o h = x_2 x_3 x_4, \quad F_1 F_o h = x_2 x_3 x_4 \not\equiv 0.$$

But it can be seen that, in Γ, there is no oriented path joining u to y with length equal to $\rho + 2 = 3$. The only path of length larger than 2 is (u, x_5, x_4, x_1, y) of length 4. Thus, if $\rho_i > \nu_i$, we see that the graph does not give anymore information on ρ_i. However, to compute $F_j F_o^r h_i$ with $r > \nu_i$, and if there is no path of length $r + 2$ in Γ, it is no need to compute the first term of (26) (with r in place of ν_i) since if there were a non zero expression in this term, there should exist a path of length $r + 2$, which contradicts our assumption. ∎

Remark 3 : the two preceding examples give a good illustration of non-generic systems : in both there were orthogonality relations between F_1 and h, so that the expressions $\pi(k_o)$ are $\not\equiv 0$, but their sum is 0. Of course, this is not generic, for if we change, for example in Remark 1, ux_1 in $(1+\varepsilon)ux_1$ we obtain : $F_1 h = (1+\varepsilon)x_1 x_2 - x_1 x_2 = \varepsilon x_1 x_2 \not\equiv 0$. ∎

Remark 4 : It is worth noting that if $r < \nu_i$, we necessarily have $F_j F_o^r h_i \equiv 0 \quad \forall j = 1 \ldots, N$. In the same way, going back to the system (17) of the introductory example, we have $d(w, y_1) = 4$, and thus $G_1 F_o^r h \equiv 0 \quad \forall r < 4-2 = 2$. Also, this remark is useful to avoid computing a number of formal expressions : if ν_i or, more precisely d_i, is obtained only for paths joining $u^{j_1}, \ldots, u^{j_r}$ to y_i, one can be sure that $F_k F_o^{\nu_i} h_i \equiv 0 \quad \forall k \neq j_1, \ldots j_r$, and one needs to check only those expressions $F_{j_1} F_o^{\nu_i} h_i, \ldots, F_{j_r} F_o^{\nu_i} h_i$ for minimal paths. ∎

IV - Description of the algorithm.

All the following computations must be done formally, for example with the languages MACSYMA or REDUCE.

1. The graph Γ

To avoid a complete construction of Γ with a number of useless nodes and arcs,

one can determine $d_i = \nu_i + 2$, and $\mathcal{U}_i$ the subset of the $(u^1,\ldots,u^N)$ corresponding to the minimal paths, directly from the data of $\sum$, and by a dynamic programming method :

. Starting from $y_i (i = 1, \ldots, p)$, we build every incident arc with: $0 \not\equiv \frac{\partial h_i}{\partial x_{k_o}}$

Then, for every x_{k_o} such that $(x_{k_o}, y_i) \in \Gamma$, we test if there is an arc (u^j, x_{k_o}) in Γ by $f_j^{k_o} \not\equiv 0$. If $(u^j, x_{k_o}) \in \Gamma$, then $d_i = 2$, $\nu_i = 0$ and $u^j \in \mathcal{U}_i$

. If $(u^j, x_{k_o}) \notin \Gamma \quad \forall j$, we change y_i into x_{k_o}, and build every incident arc to x_{k_o} by $\frac{\partial f^{k_o}_o}{\partial x_{k_1}} \not\equiv 0$; then again, for every x_{k_1} such that $(x_{k_1}, x_{k_o}) \in \Gamma$, we test if there is an arc (u^j, x_{k_1}) in Γ by $f_j^{k_1} \not\equiv 0$, and so on. If there is no arc from u^j, $\forall j$, to every path of length $\leq n-1$, then $\nu_i = +\infty$.

The same procedure can be done in parallel to determine $\underset{1 \leq j \leq M}{\text{Min}}\, d(w^j, y_i) = \mu_i$, and $\mathcal{W}_i$ the subset of the $(w^1, \ldots, w^M)$ corresponding to the $(\mu_i + 2)$ length in Γ.

2. Computation of ρ_i and the matrix Δ.

. We first compute $F_j F_o^{\nu_i} h_i \quad \forall j$ such that $u^j \in \mathcal{U}_i$.

Two cases can happen :

. either $F_j F_o^{\nu_i} h_i \not\equiv 0$ for at least one j with $u^j \in \mathcal{U}_i$.

Then $\rho_i = \nu_i$, $\Delta_i^j = F_j F_o^{\nu_i} h_i \quad \forall j$ such that $u^j \in \mathcal{U}_i$

$= 0 \quad \forall j$ such that $u^j \notin \mathcal{U}_i$.

If $\nu_i = +\infty$, then $\rho_i = +\infty$ and the i^{th} line of Δ can be deleted.

. or $F_j F_o^{\nu_i} h_i \equiv 0 \quad \forall j$ such that $u^j \in \mathcal{U}_i$

Then $\rho_i > \nu_i$ and one must compute $F_j F_o^r h_i \quad \forall r > \nu_i$,

$\forall j = 1, \ldots, N$, until the moment when one of these expressions becomes non 0 (ρ_i is then equal to the corresponding r) or until $r = n-1$ if every expression is null (then $\rho_i = +\infty$).

If ρ_i is finite, the i^{th} line of the matrix Δ is obtained by computing every expression (11) for $j = 1, \ldots, N$.

If $\rho_i = +\infty$, one can delete the i^{th} line of Δ.

3. The comparison between ρ_i and μ_i.

. If $\rho_i < \mu_i$, we have $G_j F_o^m h_i \equiv 0 \quad \forall m \leq \rho_i, \ \forall j$.

. If $\rho_i \geq \mu_i$, we have to look further if

$$G_j F_o^{\mu_i} h_i \equiv 0 \quad \forall j \text{ such that } w^j \in \mathcal{W}_i,$$

and after if $G_j F_o^m h_i \equiv 0 \quad \forall m = \mu_i + 1, \ldots, \rho_i, \ \forall j$.

Two cases can happen :

. $G_j F_o^m h_i \equiv 0 \quad \forall m \leq \rho_i, \forall j$, then the decoupling problem has a local

solution iff the system (14) has a local solution (α,β) $\exists j_o \in \{1, \ldots, M\}$ $\exists m_o < \rho_i$ such that $G_{j_o} F_o^{m_o} h_i \neq 0$, then the decoupling problem has no solution, and the system is finitely decoupled up to the order m_o, $\forall(\alpha,\beta)$. (see [2]).

4. Inversion of the system (14). Same as in [3]. ■

Remark 5 : If $\nu_i < \rho_i$, and if $\sum$ has a large dimension, it can be useful, in the evaluation of $F_j F_o^m h_i$ with $m > \nu_i$, to remark that if there is no path of length $m + 2$ joining u^j to y_i in Γ, every expression $f_j^{k_m} \pi_i(k_o, \ldots, k_m)$ is necessarily null. Thus, we eliminate this way n formal differentiations in $F_j F_o^m h_1$ ■

Remark 6 : It is clear that this method is more efficient for larger ν_i's and larger n, N, M, p. If $\nu_i = \rho_i$ and if U_1 does not contain too many elements, we need a very low number of formal derivations and the efficiency of this method is the highest. On the other hand, if $\nu_i < \rho_i$, since a minimal length in Γ is computed much faster than a formal derivation, the economy of time grows with ν_i. ■

V - Conclusion.

We have proved that the feedback decoupling method of Claude and Dufresne [3] can be significantly simplified by the introduction of the system's graph. This graph has the property that the minimal length d_i between the i^{th} output and the inputs $(u^1, \ldots, u^N)$, is generically equal to the i^{th} characteristic number ρ_i plus 2, and in general smaller or equal to $\rho_1 + 2$. This property can be used to avoid a number of formal computations and is all the more efficient as d_1 is large.

Acknowledgement. The authors are indebted to P. Willis and F. Germel, of Ecole Polytechnique, that have successfully realized the programming work.

REFERENCES

[1] D. CLAUDE. Decoupling of nonlinear systems. Syst. and Contr. Letters. Vol. 1, n°4 (1982), 242-248.

[2] D. CLAUDE. Decouplage des systèmes : du linéaire au nonlinéaire, in Developpement et utilisation d'outils et modèles mathématiques en automatique, analyse des systèmes et traitement du signal. Colloque National CNRS, Sept. 82, Belle-Ile, France.

[3] D. CLAUDE, F. DUFRESNE. An application of Macsyma to nonlinear systems decoupling. European Conference on Computer Algebra, April 82, Marseille, France .

[4] A. ISIDORI, A. KRENER, C. GORI-GIORGI, S. MONACO. Nonlinear decoupling via feedback. IEEE Trans. AC. Vol. AC26, n°2 (1981), 331-345.

[5] A. ISIDORI. The geometric approach to nonlinear feedback control : a survey. Analysis and Optimization of Systems. Lecture Notes in Control and information sciences n°44, Springer, 1982.

[6] D. SILJAK. On reachability of dynamic systems. Int. J. Syst. Sc. Vol. 8, n°3, (1977), 321-338.

Lecture Notes in Control and Information Sciences

Edited by A.V. Balakrishnan and M. Thoma

63

Analysis and Optimization of Systems

Proceedings of the Sixth International Conference on Analysis and Optimization of Systems
Nice, June 19–22, 1984

Part 2

Edited by
A. Bensoussan and J. L. Lions

Springer-Verlag
Berlin Heidelberg New York Tokyo 1984

A FAST ALGORITHM FOR SYSTEMS DECOUPLING USING FORMAL CALCULUS

F. GEROMEL*, J. LEVINE**, P. WILLIS*

ABSTRACT : The feedback decoupling problem of nonlinear systems is actually well understood in a theoretic point of view. However, to compute the decoupling feedbacks, apart of [9] the only method known by the authors, consists in using a formal derivation program to check if differential expressions are null [3]. We firstly recall the generic interpretation of these expressions in terms of the graph of the system and recall the algorithm of [9] using the minimal length of the paths joining one of the inputs to the i^{th} output. Secondly, we describe the program, and give an application to the control of robot arms.

(*) Ecole Polytechnique
91128 PALAISEAU

(**) Centre d'Automatique et d'Informatique
Ecole Nationale Supérieure des Mines de Paris
35, Rue Saint-Honoré
77305 FONTAINEBLEAU - FRANCE

A - THEORY

I - The feedback decoupling problem :

We consider a linear-analytic system, given in local coordinates, by :

$$(\Sigma) \quad \begin{cases} \dot{x} = f_o(x) + \sum_{i=1}^{N} u^i f_i(x) + \sum_{j=1}^{N} w^j g_j(x) \\ y_k = h_k(x), \quad k = 1,\dots,p \end{cases}$$

where x belongs to a connected n-dimensional analytic manifold X, $u = (u^1,\dots,u^N)^T$ are the input functions, $h_1,\dots,h_p$ are the output functions, analytic on X and where :

$$\begin{cases} F_i(x) = \sum_{j=1}^{n} f_i^j(x) \frac{\partial}{\partial x_j}, & i = o,\dots,N \\ G_j(x) = \sum_{k=1}^{n} g_j^k(x) \frac{\partial}{\partial x_k}, & j = 1,\dots,M \end{cases} \qquad (1)$$

are analytic vector fields on X.

The feedback decoupling problem consists in finding analytic functions (α^i,β^j_i), $i=1,\dots,N$, $j=1,\dots,N$, eventually defined on an open subset $\mathcal{O}$ of X such that the feedback control

$$u^i(x) = \alpha^i(x) + \sum_{j=1}^{N} \beta^j_i(x) v_j, \quad i = 1,\dots,N \tag{2}$$

makes the p outputs $y_1,\dots,y_p$ locally independent of w^i $i=1,\dots,M$. We shall denote $\hat{F}_i, i=0,\dots,N$, the vector fields obtained by the feedback (2) :

$$\begin{aligned} \hat{f}_0(x) &= f_0(x) + \sum_{i=1}^{N} \alpha^i(x) f_i(x), \quad \hat{F}_0(x) \quad \sum_{j=1}^{N} \hat{f}^j_0(x) \frac{\partial}{\partial x_j} \\ \hat{f}_i(x) &= \sum_{j=1}^{N} \beta^j_i(x) f_j(x), \quad \hat{F}_i(x) = \sum_{j=i}^{n} \hat{f}^j_i(x) \frac{\partial}{\partial x_j}, \quad i=1,\dots,N. \end{aligned} \tag{3}$$

The problem is actually well understood and the differential geometric methods [4] together with the algebraic ones [1] draw an almost complete picture of the theoretic solution. In the geometric approach of the "structural" decoupling, we introduce the maximal involutive distribution $\mathcal{B}$ of constant rank, which is $(F_0,F_1,\dots,F_N)$-invariant. Isidori, Krener, Gori-Giorgi and Monaco [4] have proved the following :

Theorem 1 The structural decoupling problem has a local solution if and only if

$$\text{span } \{G_1,\dots,G_M\} \subset \mathcal{B} \subset \bigcap_{i=1}^{P} \ker dh_i. \tag{4}$$

Furthermore, $\mathcal{B}$ can be obtained by the following induction (see [5]) :

$$\mathcal{B}_0 = \text{span } \{dh_1,\dots,dh_p\} \tag{5}$$

$$\mathcal{B}_k = \sum_{i=0}^{N} L_{F_i} \mathcal{B}_{k-1} + \mathcal{B}_{k-1} \tag{6}$$

where L_{F_i} is the Lie derivative with respect to the vector field F_i, and

$$\mathcal{B} = \Big(\bigcup_{k\geq 0} \mathcal{B}_k \Big)^{\perp} \quad \blacksquare \tag{7}$$

The algebraic methods, using Fliess' input-output map representation, give a "functional" point of view : in place of a distribution, one looks for a module $\mathcal{M}$ of vector fields, playing the same role as the distribution $\mathcal{B}$ but eventually with a non constant rank (see [1]). Claude [1] has proved the following

Theorem 2 The outputs $y_1,\dots,y_p$ are decoupled with respect to $w^1,\dots,w^M$ if and only if there exists an analytic module $\mathcal{M}$ which is also a Lie subalgebra of vector fields on X such that

$$\forall i : [\hat{F}_i,\mathcal{M}] \subset \mathcal{M} \quad \text{and} \quad \{G_1,\dots,G_M\} \subset \mathcal{M} \subset \mathcal{H} \tag{8}$$

with $\hat{F}_i$ defined by (3), and $\mathcal{H} = \{\varphi$: vector field on $X | \varphi(h_i) = 0 \quad \forall i = 1,\ldots,p\}$. ■

Furthermore, α and β can be computed in a purely algebraic way (that is to say without solving differential or partial differential equations) by the procedure described hereafter.

For this purpose, we need the :

Definition 1 : The characteristic number ρ_i of order i is the unique integer satisfying :

$$\exists j \in \{1,\ldots,N\} : F_j F_o^{\rho_i} h_i \not\equiv 0, \text{ and :} \tag{9}$$

$$\forall j \in \{1,\ldots,N\}, \forall m \in \{0,\ldots,\rho_i - 1\}, \quad F_j F_o^m h_i \equiv 0. \tag{10}$$

If $F_j F_o^m h_i \equiv 0 \quad \forall j, \forall m$, we set $\rho_i = +\infty$, and if $\exists j : F_j h_i \not\equiv 0$, $\rho_i = 0$. ■

Remark that $F_o^m h_i = F_o(F_o^{m-1} h_i)$ is a polynomial of differentials of h_i up to the order m, and that $F_o^0 h_i \equiv h_i$. ρ_i can be interpreted as the minimal number of integrations such that y_i is affected by one of the u_j.

To compute α and β, we introduce the following quantities :

$$\Delta_i^j(x) = F_j(x) F_o^{\rho_i}(x) h_i(x), \quad i = 1,\ldots,p, \quad j = 1,\ldots,N \tag{11}$$

$$\varphi_i(x) = \tilde{\varphi}_i(h_i(x), F_o(x)h_i(x), \ldots, F_o^{\rho_i}(x)h_i(x)) - F_o^{\rho_i+1}(x)h_i(x), i=1,\ldots,p \tag{12}$$

$$\psi_i^j(x) = \tilde{\psi}_i^j(h_i(x), F_o(x)h_i(x), \ldots, F_o^{\rho_i}(x)h_i(x)), \quad \begin{matrix} i=1,\ldots,p \\ j=1,\ldots,N \end{matrix} \tag{13}$$

with $\tilde{\varphi}_i$ and $\tilde{\psi}_i^j$ arbitrary analytic functions.

Let us call : Δ the $p \times N$ matrix-valued analytic function whose $(i,j)^{th}$ element is Δ_i^j, $\varphi = (\varphi_1,\ldots,\varphi_p)^T$ and ψ the $p \times N$ matrix-valued analytic function whose $(i,j)^{th}$ element is ψ_i^j.

Theorem 3 : If $G_k F_o^m h_i \equiv 0 \quad \forall k \in \{1,\ldots,M\}, \quad \forall m \leq \rho_i$, a necessary and sufficient condition for (α,β) to realize the local functional decoupling of (Σ), is that (α,β) locally solve the system :

$$\begin{aligned} \Delta\alpha &= \varphi \\ \Delta\beta &= \psi \end{aligned} \tag{14}$$

In this case, the change of variables :

$$X_o^i = h_i,\ldots, X_{\rho_i}^i = F_o^{\rho_i} h_i, \quad i = 1,\ldots,p, \tag{15}$$

Puts the system (Σ) locally into the form :

$$\begin{cases} \dot{X}_0^1 = X_1^1 \\ \vdots \\ \dot{X}_{\rho_1-1}^i = X_{\rho_1}^1 \\ \dot{X}_{\rho_i}^1 = \tilde{\varphi}_i(X_0^i,\ldots,X_{\rho_1}^i) + \sum_{j=1}^{N} \tilde{\psi}_1^j(X_0^i,\ldots,X_{\rho_i}^i)v_j \\ y_1 = X_0^i \end{cases} \qquad (16)$$

$1 = 1,\ldots,p.$ ■

Clearly, this procedure involves a huge amount of formal calculus, especially to determine the characteristic numbers ρ_1 , $i = 1,\ldots,p$: one must differentiate ρ_i times the expressions h_1, $F_0 h_1$, etc..., whose complexity is growing very fast, and then check if $F_j F_0^m h_i$ is null or not. A program has been developed by Claude and Dufresne [3], using the language MACSYMA, to compute these formal expressions.

The aim of this paper is to introduce a faster method to compute ρ_i with the minimal number of formal differentiations : for this purpose, we shall prove that the numbers ρ_i can generically be very easily obtained on the system's graph. We shall also give a lower bound ν_i for ρ_1 in the non-generic case, still obtained from the graph, and prove that either $\nu_i \leq n-1$ or $\nu_1 = \rho_i = +\infty$. These results are finally synthesized in an algorithm to compute (α,β).

II - The system's graph :

As in [6], we introduce the following system's graph :

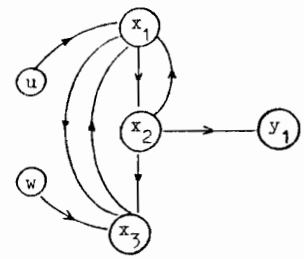
Definition 2 : We call Γ the system's graph of Σ in a given open subset $\mathcal{O}$ of X with given local coordinates, the oriented graph whose input-nodes are $(u^1,\ldots,u^N,w^1,\ldots,w^M)$, whose intermediate-nodes are the state variables $(x_1,\ldots,x_n)$, and whose output-nodes are $(y_1,\ldots,y_p)$. The oriented arcs of Γ are obtained as follows :

- There exists an oriented arc joigning u^1 to x_k iff $f_1^k(x) \not\equiv 0$ in $\mathcal{O}$, $i = 1,\ldots,N$, $k = 1,\ldots,n$, and joinging w^1 to x_k iff $g_1^k(x) \not\equiv 0$ in $\mathcal{O}$, $i = 1,\ldots,M$, $k = 1,\ldots,n$.
- There exists an oriented arc joining x_k to x_j iff $\frac{\partial f_0^j}{\partial x_k}(x) \not\equiv 0$ in $\mathcal{O}$, $j, k = 1,\ldots,n$.
- There exists an oriented arc joining x_k to y_i iff $\frac{\partial h_i}{\partial x_k}(x) \not\equiv 0$ in $\mathcal{O}$, $1 = 1,\ldots,p$, $k = 1,\ldots,n$. ■

Definition 3 : We call $d(u^j,y_i)$ the minimal number of oriented arcs of Γ forming an oriented path joining u^j to y_1, and $d_1 = \min_{1\leq j\leq N} d(u^j,y_i)$. ■

An introductory example : $n = 3$, $N = 1$, $M = 1$, $p = 1$, $f_1(x) \not\equiv 0$, $\frac{\partial f_o^2}{\partial x_1}(x_1,x_2) \not\equiv 0$.

$$(17)\quad \begin{cases} \dot{x}_1 = f_o^1(x_1,x_2,x_3) + u\, f_1(x_1,x_2,x_3) \\ \dot{x}_2 = f_o^2(x_1,x_2) \\ \dot{x}_3 = f_o^3(x_1,x_2,x_3) + w g_1(x_1,x_2,x_3) \\ y_1 = h(x_2) \end{cases}$$



The system's graph Γ

It can be easily seen that $d_1 = d(u,y_1) = 3$, $d(w,y_1) = 4$.
We shall prove that we can predict that, generically, $\rho_1 = d_1 - 2 = 1$, and that $d(w,y_1) > d_1$ implies $G_1 h \equiv 0$ and $G_1 F_o h \equiv 0$. To check our assertion, let us go back to (9) and (10), and compute ρ_1. We first check that $(F_1 h)(x) = f_1(x)\frac{\partial h}{\partial x_1}(x_2) \equiv 0$, $(G_1 h)(x) = g_1(x)\frac{\partial h}{\partial x_3}(x_2) \equiv 0$. Then : $(F_o h)(x) = f_o^1(x)\frac{\partial h}{\partial x_1}(x_2) + f_o^2(x_1,x_2)\frac{\partial h}{\partial x_2}(x_2) + f_o^3(x)\frac{\partial h}{\partial x_3}(x_2) = f_o^2(x_1,x_2)\frac{\partial h}{\partial x_2}(x_2)$, $(F_1 F_o h)(x) = f_1(x)\frac{\partial f_o}{\partial x_1}(x_1,x_2)\frac{\partial h}{\partial x_2}(x_2) \not\equiv 0$, and thus $\rho = 1$;
Finally, we also have $G_1 F_o h = g_1\left(\frac{\partial f_o^2}{\partial x_3}\frac{\partial h}{\partial x_2} + f_o^2\frac{\partial^2 h}{\partial x_3 \partial x_2}\right) \equiv 0$, as claimed above.

Thus, almost without computations, ρ_1 and the relations $G_1 h \equiv 0$ and $G_1 F_o h \equiv 0$, can be deduced from the system's graph (we only need to compute $F_1 F_o h$!). Clearly, the system's graph synthesizes the structure of the interactions of the input and output variables versus integration of the state variables. Thus, it is not surprising that, in general, but generically only, the minimal length d_i represents the minimum number of integrations for the inputs to affect y_1, namely ρ_i up to a constant equal to 2 since the first and last arcs do not represent integrations.

Remark : in Γ, we do not take into account the fact that $f_1,\ldots,f_N,g_1,\ldots,g_M$ depend on $x_1,\ldots,x_n$ or not. For our purpose these interactions do not play any role in generic situations and, if they play a role in non-generic cases, the profit of the graph's method vanishes, as will be seen after.

III - The characteristic numbers ρ_i, their lower bounds ν_i, and the system's graph

Besides the characteristic numbers ρ_i, we shall introduce the numbers ν_i defined as follows :

Definition 4 : The number ν_i, $i = 1,\ldots,p$, is the unique integer satisfying :

$$\exists j \in \{1,\ldots,N\},\ \exists k_{\nu_i},\ldots,k_o \in \{1,\ldots,n\} \quad \text{such that :}$$

$$f_j^{k_{\nu_1}} \frac{\partial f_o^{k_{\nu_1-1}}}{\partial x_{k_{\nu_i}}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_1}{\partial x_{k_o}} \not\equiv 0 \text{ in } \mathcal{O}, \tag{18}$$

and : $\forall j \in \{1,\ldots,N\}$, $\forall r < \nu_1$, $\forall k_r,\ldots,k_o \in \{1,\ldots,n\}$, we have :

$$f_j^{k_r} \frac{\partial f_o^{k_{r-1}}}{\partial x_{k_r}} \cdots \frac{\partial f_o^{k_o}}{\partial x_{k_1}} \frac{\partial h_1}{\partial x_{k_o}} \equiv 0 \text{ in } \mathcal{O} \quad \blacksquare \tag{19}$$

Now we can state the main result :

Theorem 4 : . $\nu_1 = d_1 - 2$, $1 = 1,\ldots,p$. (20)

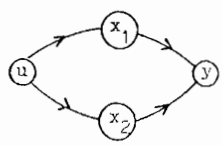
. $\nu_1 \leq \rho_i$ and $\nu_i = \rho_1$ generically, $i = 1,\ldots,p$. (21)

By generically, we mean : for every system Σ whose coefficients $f_o, f_1, \ldots, f_N, g_1, \ldots, g_M, h_1, \ldots, h_p$, lie outside a closed subset, with empty interior, of the space of analytic vector-valued functions on $\mathcal{O} \subset X$, the functions f_o, f_1, g_1, h_1, depending locally on the same variables as those of the original system

Corollary : If $\nu_i > n-1$, then $\rho_i = \nu_i = +\infty$

Remark 1 : From the corollary, we conclude that ν_1 is computed in at most $n-1$ steps, and, generically, the same holds for ρ_i The result for ρ_i was proved in [2]. However, it is remarkable that one can have $\nu_1 < n-1$ whereas $\rho_1 = +\infty$ as the following example proves :

$$\begin{cases} \dot{x}_1 = ux_1 \\ \dot{x}_2 = -ux_2 \\ y = x_1x_2 \end{cases}$$



- 1 -

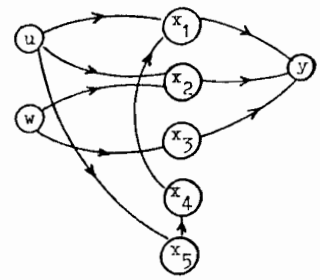
It is very easy to see that $\nu = 0$, but $\rho = +\infty$ since

$$F_1 h = x_1 \frac{\partial (x_1x_2)}{\partial x_1} - x_2 \frac{\partial (x_1x_2)}{\partial x_2} \equiv 0 ,$$

and $F_1F_o h \equiv 0$ since $F_o \equiv 0$ Finally, this suffices to prove that $\rho = +\infty$ since $n = 2$. $\blacksquare$

Remark 2 : It would be a nice result, if $\rho_i > \nu_1$, that there exists a (non minimal) oriented path from one of the u^j to y_i of length $\rho_i + 2$.
Unfortunately, this is only true for linear systems. A counterexample in the non linear case :

$$\begin{cases} \dot{x}_1 = x_4 + ux_1 \\ \dot{x}_2 = x_2 w - x_2 u \\ \dot{x}_3 = -x_3 w \\ \dot{x}_4 = x_5 \\ \dot{x}_5 = u \\ y = x_1 x_2 x_3 \end{cases}$$



- Γ -

We have $\nu = 0$, whereas $\rho = 1$:

$$F_1 h = x_1 \frac{\partial}{\partial x_1}(x_1x_2x_3) - x_2 \frac{\partial}{\partial x_2}(x_1x_2x_3) \equiv 0, \quad G_1 h = x_2 \frac{\partial}{\partial x_2}(x_1x_2x_3) - x_3 \frac{\partial}{\partial x_3}(x_1x_2x_3) \equiv 0,$$

$$F_o h = x_2x_3x_4 \ , \ F_1F_o h = -\, x_2x_3x_4 \not\equiv 0.$$

But it can be seen that, in Γ, there is no oriented path joining u to y with length equal to $\rho + 2 = 3$. The only path of length larger than 2 is (u,x_5,x_4,x_1,y) of length 4. Thus, if $\rho_i > \nu_i$, we see that the graph does not give anymore information on ρ_i. However, to compute $F_1F_o^r h_i$ with $r > \nu_i$, and if there is no path of length $r + 2$ in Γ, it is no need to compute the terms of the form (18) (with r in place of ν_i) since if there were a non zero expression in these terms, there should exist a path of length $r + 2$, which contradicts our assumption. ■

Remark 3 : the two preceding examples give a good illustration of non-generic systems: in both there were orthogonality relations between F_1 and h, so that the expressions (18) are $\not\equiv 0$, but their sum is 0. Of course, this is non generic, for if we change, for example in Remark 1, ux_1 in $(1+\varepsilon)ux_1$, we obtain : $F_1 h = (1+\varepsilon)x_1x_2 - x_1x_2 = \varepsilon x_1x_2 \not\equiv 0$. ■

Remark 4 : It is worth noting that if $r < \nu_i$, we necessarily have $F_j F_o^r h_i \equiv 0$ $\forall j = 1,\ldots,N$. In the same way, going back to the system (17) of the introductory example, we have $d(w,y_1) = 4$, and thus $G_1 F_o^r h \equiv 0$ $\forall r < 4-2 = 2$. Also, this remark is useful to avoid computing a number of formal expressions : if ν_i or, more precisely d_i, is obtained only for paths joining $u^{j_1},\ldots,u^{j_r}$ to y_i, one can be sure that $F_k F_o^{\nu_i} h_i \equiv 0$ $\forall k \neq j_1,\ldots,j_r$, and one needs to check only those expressions $F_{j_1} F_o^{\nu_i} h_i,\ldots,F_{j_r} F_o^{\nu_i} h_i$ for minimal paths. ■

IV - Description of the algorithm :

All the following computations must be done formally, for example with the languages MACSYMA or REDUCE.

1. The graph Γ

To avoid a complete construction of Γ with a number of useless nodes and arcs, one can determine $d_1 = \nu_1 + 2$, and $\mathcal{U}_1$ the subset of the $(u^1,\dots,u^N)$ corresponding to the minimal paths, directly from the data of Σ, and by a dynamic programming method :

- Starting from y_1 $(1 = 1,\dots,p)$, we build every incident arc with : $0 \not\equiv \dfrac{\partial h_1}{\partial x_{k_o}}$

  Then, for every x_{k_o} such that $(x_{k_o},y_1) \in \Gamma$, we test if there is an arc $^{k_o}(u^j,x_{k_o})$ in Γ by $f_j^{k_o} \not\equiv 0$. If $(u^j,x_{k_o}) \in \Gamma$, then $d_1 = 2$, $\nu_i = 0$ and $u^j \in \mathcal{U}_1$

- If $(u^j,x_{k_o}) \notin \Gamma$ $\forall j$, we change y_1 into x_{k_o} and build every incident arc to x_{k_o} by $\dfrac{\partial f_o^{k_o}}{\partial x_{k_1}} \not\equiv 0$; then again, for every x_{k_1} such that $(x_{k_1},x_{k_o}) \in \Gamma$, we test if there is an arc (u^j,x_{k_1}) in Γ by $f_j^{k_1} \not\equiv 0$, and so on. If there is no arc from u^j, $\forall j$, to every path of length $\leq n-1$, then $\nu_i = +\infty$.

The same procedure can be done in parallel to determine $\operatorname{Min}_{1\leq j\leq M} d(w^j,y_i) = \mu_1$, and $\mathcal{W}_i$ the subset of the $(w^1,\dots,w^M)$ corresponding to the $(\mu_i + 2)$ length in Γ.

2. Computation of ρ_i and the matrix Δ.

We first compute $F_j F_o^{\nu_1} h_1$ $\forall j$ such that $u^j \in \mathcal{U}_i$

Two cases can happen :

- either $F_j F_o^{\nu_i} h_1 \not\equiv 0$ for at least one j with $u^j \in \mathcal{U}_1$

  Then $\rho_i = \nu_i, \Delta_i^j = F_j F_o^{\nu_i} h_i$ $\forall j$ such that $u^j \in \mathcal{U}_1$

  $= 0$ $\forall j$ such that $u^j \notin \mathcal{U}_1$

  If $\nu_i = \nu_{i+1} = +\infty$, then $\rho_i = +\infty$ and the i^{th} line of Δ can be deleted.

- or $F_j F_o^{\nu_i} h_1 \equiv 0$, $\forall j$ such that $u^j \in \mathcal{U}_1$

  Then $\rho_i > \nu_i$ and one must compute $F_j F_o^{r} h_i$ $\forall r > \nu_i$,

  $\forall j = 1,\dots,N$, until the moment when one of these expressions becomes non 0 (ρ_i is then equal to the corresponding r) or until $r = n-1$ if every expression is null (then $\rho_1 = +\infty$).

  If ρ_i is finite, the 1^{th} line of the matrix Δ is obtained by computing every expression (11) for $j = 1,\dots,N$.

  If $\rho_1 = +\infty$, one can delete the i^{th} line of Δ.

3. The comparison between ρ_i and μ_1

- If $\rho_i < \mu_1$, we have $G_j F_o^{m} h_1 \equiv 0$ $\forall m \leq \rho_1$, $\forall j$.

  If $\rho_i \geq \mu_i$, we have to look further if

  $$G_j F_o^{\mu_i} h_1 \equiv 0 \quad \forall j \text{ such that } w^j \in \mathcal{W}_i ,$$

and after if $G_j F_o^m h_i \equiv 0 \quad \forall m = \mu_i + 1, \ldots, \rho_i , \quad \forall j$.

Two cases can happen :

- $G_j F_o^m h_i \equiv 0 \quad \forall m \leq \rho_i, \forall j$, then the decoupling problem has a local solution iff the system(14) has a local solution (α,β).
- $\exists j_o \in \{1,\ldots,M\}, \ \exists m_o < \rho_i$ such that $G_{j_o} F_o^{m_o} h_i \not\equiv 0$, then the decoupling problem has no solution, and the system is finitely decoupled up to the order m_o , $\forall(\alpha,\beta)$. (see [2]).

4. Inversion of the system (14). Same as in [3]. ■

Remark 5 : If $\nu_i < \rho_i$, and if Σ has a large dimension, it can be useful, in the evaluation of $F_j F_o^m h_i$ with $m > \nu_i$, to remark that if there is no path of length $m + 2$ joining u^j to y_i in Γ, every expression (18) with m in place of ν_i is necessarily null. Thus, we eliminate this way n formal differentiations in $F_j F_o^m h_i$. ■

Remark 6 : It is clear that this method is more efficient for larger ν_i's and larger n, N, M, p. If $\nu_i = \rho_i$ and if u_i does not contain too many elements, we need a very low number of formal derivations and the efficiency of this method is the highest. On the other hand, if $\nu_i < \rho_i$, since a minimal length in Γ is computed much faster than a formal derivation, the economy of time grows with ν_i. ■

Remark 7 : For linear systems, the graph's method can be significantly improved since ρ_i can be completely obtained from the graph : in place of step 2 of the algorithm, we have :

If $\nu_i < \rho_i$: delete every path of length $\nu_i + 2$, in the graph, and find the new minimal length $\nu_i' > \nu_i$. Check if $\exists j$ such that $F_j F_o^{\nu_i'} h_o \not\equiv 0$.

If yes, $\nu_i' = \rho_i$. If not, delete again every path of length $\nu_i' + 2$ and so on, until every path of length $\leq n + 2$ is deleted, then $\rho_i = +\infty$.

B - THE PROGRAM

I - Organization of the program

The programming language is MACSYMA.

The programming is made of :

1) The main program : prob ( )

It asks questions to the user and decides which subroutines to run.

2) The subroutines :

a) expli( ) : gives informations, if needed, on the program's use.

b) mit( ) : memorizes the formal equations of the system .

c) calnu( ) : computation of ν by the graph's method

d) calro( ) : computation of ρ

e) pmatcal( ) - trig( ) - reordd( ) : solves the Δ - system after reorganization of lines and columns of Δ

f) feedback( ) : gives the final result on the feedbacks.

For further informations and examples of sessions, see [8]

II - Session's display

At each step of the session, the user may choose between different tasks

1) at the beginning : the inputs can be checked and corrected, and the user can ask for further informations (expli( )).

2) during the session the user must answer the questions of recognizing null expressions. For example the program cannot check the nullity of an expression such as $x_1 \frac{\partial f}{\partial x_1} - x_2 \frac{\partial f}{\partial x_2}$ when f is not specified.

3) at the end : the user may help the program to simplify the results, for example, by giving rules of trigonometric simplifications. The user must be aware of the fact that some simplifications automatically done by MACSYMA may be sometimes worse than no simplification at all.

Remark : In order to protect the intermediate results from manipulation's errors, the main program saves them step by step in auxiliary files.

C - EXAMPLE : THE ROBOT ARM

We study the decoupling problem for a 3 degrees of freedom robot arm. It is composed of three segments of length l_1, l_2, l_3. The links have relative angles noted x_1 x_2, x_3 and x_4, x_5 and x_6 are the respective angular velocities. The cartesian coordinates of the extremity are y_1, y_2, y_3 and we wish to control its motion along the y_1 - axes - such a problem arises in automatic sizing.
The motion's equations ([7]) are

$$\dot{x}_1 = x_4$$
$$\dot{x}_2 = x_5$$
$$\dot{x}_3 = x_6$$
$$\dot{x}_4 = (u_1 - f_1(x_2,x_3,x_4,x_5,x_6)) \cdot \frac{1}{b_{11}(x_2,x_3,x_4)}$$
$$\dot{x}_5 = (b_{22}(x_3,x_5,x_6)u_2 + b_{32}(x_3,x_5,x_6\,u_3 + \\ + b_{22}(x_3,x_5,x_6)f_2(x_2,x_3,x_4,x_5)) \cdot \frac{1}{\det(x_3,x_5,x_6)}$$

$$\dot{x}_6 = (b_{23}(x_3,x_5,x_6)u_2 + b_{33}(x_3,x_5,x_6)u_3 + \\ + b_{32}(x_3,x_5,x_6)f_3(x_3,x_4,x_5,x_6)) \frac{1}{\det(x_3,x_5,x_6)}$$

with $\det(x_3,x_5,x_6) = (b_{22}b_{33} - b_{23}b_{32})(x_3,x_5,x_6)$.

The outputs are :

$$y_1 = \cos x_1 (l_3\sin(x_3+x_2)+l_2\sin x_2)$$

$$y_2 = \sin x_1 (l_3\sin(x_3+x_2)+l_2\sin x_2)$$

$$y_3 = l_3\cos(x_3+x_2)+l_2\cos x_2+l_1$$

The program finds that : $\nu_1 = \nu_2 = \nu_3 = 1$ and $\rho_1 = \rho_2 = \rho_3 = 1$ and that :

$$\det\Delta = \frac{l_2 l_3 \sin x_3}{b_{11}(x_2,x_3,x_4)\det(x_3,x_5,x_6)} (l_3\sin(x_3+x_2) + l_2\sin x_2)$$

The expressions of α_i and β_{ij}, $1\leq i, j\leq 3$, expressed as functions of $\varphi_1,\varphi_2,\varphi_3$ and $\psi_{11},\psi_{12},\psi_{33}$ as in (12),(13) are :

$$\alpha_1 = -[2l_3 b_{11}\cos(x_2+x_3)x_4(x_6+x_5)+2l_2\cos(x_2)b_{11}\, x_4x_5 - l_3f_1\sin(x_3+x_2)-l_2f_1\sin x_2 + \\ +(\varphi_1\sin(x_1) - \varphi_2\cos(x_1))b_{11}] \times \frac{1}{l_2\sin(x_3+x_2)+l_2\sin x_2}$$

$$\alpha_2 = [(l_3^2 b_{33}+(l_2l_3\cos x_3+l_3^2)b_{32})x_6^2 + (2l_3^2 b_{33}+(2l_2l_3\cos x_3+2l_3^2)b_{32})x_5x_6 + \\ + ((l_2l_3\cos(x_3)+l_3^2)b_{33}+(2l_2l_3\cos(x_3)+l_3^2+l_2^2)b_{32})x_5^2 \\ + ((l_3^2 b_{33}+l_3^2 b_{32})\sin^2(x_3+x_2)+ \\ (l_2l_3\sin(x_2)b_{33}+2l_2l_3\sin(x_2)b_{32})\sin(x_2+x_3)+l_2^2\sin^2(x_2)b_{32})x_4^2 \\ + ((\varphi_2 l_3\sin(x_1)+\varphi_1 l_3\cos(x_1))b_{33} + (\varphi_2 l_3\sin(x_1)+\varphi_1 l_3\cos(x_1))b_{32})\sin(x_3+x_2) \\ + \varphi_3 l_3(b_{33}+b_{32})\cos(x_3+x_2) \\ + ((\varphi_2 l_2\sin(x_1)+\varphi_1 l_2\cos x_1)\sin x_2+\varphi_3 l_2\cos x_2)b_{32} - l_2l_3f_2\sin x_3] \times \frac{1}{l_2l_3\sin x_3} \\ + \frac{f_3 b_{32}^2 - f_2 b_{23} b_{32}}{\det(x_3,x_5,x_6)}$$

$$\alpha_3 = [f_2 b_{22} b_{23} - f_3 b_{22} b_{32}]/\det(x_3, x_5, x_6)$$

$$- [(l_3^2 b_{23} + (l_2 l_3 \cos x_3 + l_3^2) b_{22}) x_6^2 + (2 l_3^2 b_{23} + (2 l_2 l_3 \cos x_3 + 2 l_3^2) b_{22}) x_5 x_6$$

$$+ ((l_2 l_3 \cos x_3 + l_3^2) b_{23} + (2 l_2 l_3 \cos(x_3) + l_3^2 + l_2^2) b_{22}) x_5^2$$

$$+ ((l_3^2 b_{23} + l_3^2 b_{22}) \sin^2(x_3 + x_2) + (l_2 l_3 \sin(x_2) b_{23}$$

$$+ 2 l_2 l_3 \sin(x_2) b_{22}) \sin(x_3 + x_2) + l_2^2 \sin^2 x_2 b_{22}) x_4^2$$

$$+ ((\varphi_2 l_3 \sin(x_1) + \varphi_1 l_3 \cos(x_1)) b_{23} + (\varphi_2 l_3 \sin(x_1) + \varphi_2 l_3 \cos(x_1)) b_{22}) \sin(x_3 + x_2)$$

$$+ \varphi_3 l_3 (b_{23} + b_{22}) \cos(x_3 + x_2)$$

$$+ ((\varphi_2 l_2 \sin(x_1) + \varphi_1 l_2 \cos(x_1)) \sin x_2 + (\varphi_3 l_2 \cos(x_2)) b_{22}] \times \frac{1}{l_2 l_3 \sin x_3}$$

$$\beta_{11} = - \frac{\psi_{11} \sin(x_1) b_{11}}{l_3 \sin(x_3 + x_2) + l_2 \sin(x_2)}$$

$$\beta_{12} = \frac{\psi_{22} \cos(x_1) b_{11}}{l_3 \sin(x_3 + x_2) + l_2 \sin(x_2)}$$

$$\beta_{13} = 0$$

$$\beta_{21} = \psi_{11} \cos x_1 (l_3 (b_{33} + b_{32}) \sin(x_3 + x_2) + l_2 \sin(x_2) b_{32}) \times \frac{1}{l_2 l_3 \sin x_3}$$

$$\beta_{22} = \psi_{22} [l_3 (b_{33} + b_{32}) \sin(x_3 + x_2) + l_2 \sin x_2 b_{32}] \times \frac{\sin x_1}{l_2 l_3 \sin x_3}$$

$$\beta_{23} = \psi_{33} (l_3 (b_{33} + b_{32}) \cos(x_3 + x_2) + l_2 \cos(x_2) b_{32}) \times \frac{1}{l_2 l_3 \sin x_3}$$

$$\beta_{31} = - \psi_{11} \cos x_1 (l_3 (b_{23} + b_{22}) \sin(x_3 + x_2) + l_2 \sin x_2 b_{22}) \times \frac{1}{l_2 l_3 \sin x_3}$$

$$\beta_{32} = - \psi_{22} \sin x_1 (l_3 (b_{23} + b_{22}) \sin(x_3 + x_2) + l_2 \sin x_2 b_{22}) \times \frac{1}{l_2 l_3 \sin x_3}$$

$$\beta_{33} = - \psi_{33} (l_3 (b_{23} + b_{22}) \cos(x_3 + x_2) + l_2 \cos x_2 b_{22}) \times \frac{1}{l_2 l_3 \sin x_3}$$

Remark : In this example a direct computation would need 198 partial derivations and 198 multiplications, whereas with the graph's method, only 85 partial derivations and 56 multiplications have been executed. For other examples see [8].

CONCLUSION

We have proved that the feedback decoupling method of Claude and Dufresne [3] can be significantly simplified by the introduction of the system's graph. This graph has the property that the minimal length d_i between the i^{th} output and the

inputs $(u^1,\ldots,u^N)$, is generically equal to the i^{th} characteristic number ρ_i plus 2, and in general smaller or equal to $\rho_i + 2$. This property can be used to avoid a number of formal computations and is all the more efficient as d_i is large. This method requires the use of MACSYMA because of the formal manipulations, and gives an efficient CAO tool for non linear systems decoupling.

REFERENCES

[1] D. CLAUDE. Decoupling of nonlinear systems. Syst. and Contr. Letters. Vol.1, n^o4 (1982), 242-248.

[2] D. CLAUDE. Decouplage des systèmes : du linéaire au nonlinéaire, in : Developpement et utilisation d'outils et modèles mathématiques en automatique, analyse des systèmes et traitement du signal. Vol.3, I.D. Landau ed., CNRS, Paris, 1983, 533-555.

[3] D. CLAUDE, P. DUFRESNE. An application of Macsyma to nonlinear systems decoupling Lecture Notes in Computer Sciences, Vol.144, Springer, 1982, 294-301.

[4] A. ISIDORI, A. KRENER, C. GORI-GIORGI, S. MONACO. Nonlinear decoupling via feedback. IEEE Trans. AC. Vol. AC26, n^o2 (1981), 331-345.

[5] A. ISIDORI. The geometric approach to nonlinear feedback control : a survey. Analysis and Optimization of Systems. Lecture Notes in Control and information sciences n^o44, Springer, 1982, 517-531.

[6] D. SILJAK. On reachability of dynamic systems. Int. J. Syst. Sc. Vol.8, n^o3, (1977), 321-338.

[7] S. NICOSIA, F. NICOLO, D. LENTINI. Dynamical control of industrial robots with elastic and dissipative joints. 8^{th} IFAC World Congress - Kyoto - 1981.

[8] F. GEROMEL, P. WILLIS. Algorithme de graphe pour le découplage de systèmes non-linéaires. Option Automatique. Ecole Polytechnique. Promotion 80. Juin 83.

[9] A. KASINSKY, J. LEVINE. A fast graph theoretic algorithm for the feedback decoupling problem of nonlinear systems. 8^{th} MTNS conference. Beersheva. June 1983.

ANNEXE

Un aperçu élémentaire de la Théorie Modernes des Systèmes Non linéaires.

R.A.I.R O Automatique/Systems Analysis and Control
(vol. 17, n° 4, 1983, p. 277-321)

UN APERÇU ÉLÉMENTAIRE DE LA THÉORIE MODERNE DES SYSTÈMES NON LINÉAIRES (*)

par J. LEVINE (1)

Présenté par P. BERNHARD

Résumé. — *On présente un exposé sur la théorie des systèmes non linéaires déterministes, faisant la synthèse des résultats les plus récents, dans un vocabulaire accessible aux ingénieurs, illustré de nombreux exemples, et comportant des indications d'applications déjà réalisées ou en cours. On y insiste particulièrement sur les difficultés d'étendre les méthodes du linéaire d'une part, sur le prolongement au non linéaire des techniques de descriptions externe et interne d'autre part, et enfin sur l'adaptation des méthodes d'échelles de temps multiples.*

Abstract. — *We present a survey of the deterministic nonlinear system theory, including the most recent results, in a language adapted to engineers, with many examples and applications. We focuse on the difficulties to extend the methods of linear systems on the one hand, on the generalization to the nonlinear framework of external and internal descriptions on the other hand, and finally on the adaptation of the multiple time-scales methods.*

INTRODUCTION

Ce rapport traite des développements récents de la théorie des systèmes non linéaires et répond à un besoin d'information d'une part à cause de questions de plus en plus nombreuses émanant des industriels, d'autre part à cause de la difficulté relative à accéder à ces informations, disponibles dans un langage mathématique assez peu vulgarisé. L'auteur s'est donc particulièrement attaché d'une part à présenter les résultats de la manière la moins technique possible, en motivant par des exemples simples les techniques algébriques ou géométriques qui sont, somme toute, encore très nouvelles en Automatique, la rigueur mathématique n'ayant généralement pas à en souffrir ; et d'autre part à donner un grand nombre d'exemples d'applications réelles dans de nombreux domaines allant de l'aéronautique aux réseaux de distribution d'eau, en passant par la biologie...

(*) Reçu en avril 1983. Cette étude a été financée par l'AEPA pour le compte de la DRET.
(1) Centre d'Automatique et Informatique de l'École Nationale Supérieure des Mines de Paris, 35, rue St Honoré, 77305 Fontainebleau Cedex.

R.A.I.R.O. Automatique/Systems Analysis and Control, 0399-0524/1983 277/$ 5.00

Le rapport est organisé en trois chapitres

A. Les difficultés d'extension du linéaire au non linéaire.

B. Méthodologies des systèmes non linéaires. Systèmes linéaires-analytiques (représentations externe et interne des systèmes non linéaires).

C. Perturbations singulières. Aggrégation. Cohérence.

Dans les chapitres B et C, on insiste particulièrement sur le parallèle formel complet avec la théorie linéaire, concernant les descriptions interne et externe et les échelles de temps multiples. Ainsi, les notions de « fonction de transfert » et de « réponse impulsionnelle » se généralisent aux « séries génératrices » et « série de Volterra » respectivement, la « notion d'espace d'état » devient « variété d'état », etc... On a pris soin toutefois de détailler et d'illustrer par des exemples les principaux cas pathologiques où la généralisation brutale du linéaire est en défaut.

Seuls les systèmes déterministes y sont abordés car un rapport DRET de G. Pignié et l'auteur est actuellement disponible sur les systèmes stochastiques et le filtrage non linéaire [1].

Enfin, les problèmes du non linéaire en commande optimale ne sont abordés qu'en conclusion pour éviter l'alourdissement du rapport par des développements trop hétéroclites.

SYSTÈMES DÉTERMINISTES

A. LES DIFFICULTÉS D'ÉTENDRE LES MÉTHODES DU LINÉAIRE AU NON LINÉAIRE

I. Les inconvénients de la linéarisation

Historiquement, les premières tentatives d'approche des systèmes non linéaires (qu'on notera toujours dans la suite *NL*) sont basées sur la *linéarisation autour d'une trajectoire nominale*. Cette méthode s'applique à divers domaines d'étude : stabilité asymptotique [2], commandabilité et observabilité locales [3], identification au voisinage d'un point de fonctionnement [4], filtrage de Kalman étendu [5] (bien que ce dernier point ne concerne pas les systèmes déterministes, il s'agit d'une question de théorie des systèmes qui déborde du caractère strictement probabiliste), etc...

Bien que son efficacité soit reconnue dans des situations particulières, elle comporte des défauts majeurs qui la condamnent comme outil général

• La linéarisée d'un système *NL* ne donne qu'une description très partielle du comportement entrées-sorties.

• La linéarisation de non-linéarités non bornées (par exemple polynômiales) n 'est pas robuste.

• Le filtre de Kalman étendu n'offre aucune garantie théorique d'efficacité.

• Un système linéaire ou affine par morceaux qui approxime localement un système NL peut avoir un comportement entrées-sorties qualitativement différent de celui du système NL.

Détaillons ces 4 points

• *Point* 1 : Nous allons donner un exemple de mauvaise description entrées-sorties par linéarisation, où le système NL est commandable alors que sa linéarisée ne l'est pas au voisinage de l'origine. Cet exemple, tiré de [6], représente sous forme simplifiée, la dynamique d'un satellite rigide non sphérique propulsé par un couple de tuyères.

Le système est donné par :

$$\left.\begin{aligned} \dot{x}_1 &= a_1 x_2 x_3 + b_1 u \\ \dot{x}_2 &= a_2 x_3 x_1 + b_2 u \\ \dot{x}_3 &= a_3 x_1 x_2 \end{aligned}\right\} \tag{1}$$

avec $| u | \leqslant 1$ et : $a_3 \neq 0, a_1 b_2^2 \neq a_2 b_1^2$.

($a_3 \neq 0$ s'interprète comme le fait que le satellite n'est pas sphérique.) On peut montrer (voir [6] et plus loin) que ce système est complètement commandable, c'est-à-dire que l'on peut atteindre tout R^3 à l'aide de commandes $u(t)$ continues par morceaux avec $| u(t) | \leqslant 1 \quad \forall t$.

Avant d'étudier la commandabilité du système (1) linéarisé, rappelons la condition de Lee-Markus [3]

THÉORÈME : *Soit le système* NL

$$(NL) \quad \dot{x} = f(x, u) \quad x \in R^n \quad u \in R^p$$

tel que

$$f(0, 0) = 0$$

Notons

$$F = \frac{\partial f}{\partial x}(0, 0) \qquad G = \frac{\partial f}{\partial u}(0, 0)$$

Alors si le système linéaire (linéarisée de (NL) *au voisinage de* $(0, 0)$)

$$(L) \quad \dot{x} = Fx + Gu$$

est complètement commandable, il existe un voisinage de l'origine sur lequel (NL) *est commandable.* ■

On note que (1) satisfait bien $f(0, 0) = 0$ avec

$$f(x, u) = \begin{pmatrix} f_1(x_1, x_2, x_3, u) \\ f_2(x_1, x_2, x_3, u) \\ f_3(x_1, x_2, x_3, u) \end{pmatrix} = \begin{pmatrix} a_1 x_2 x_3 + b_1 u \\ a_2 x_1 x_3 + b_2 u \\ a_3 x_1 x_2 \end{pmatrix}$$

On a

$$\frac{\partial f}{\partial x}(x, u) = \begin{pmatrix} 0 & a_1 x_3 & a_1 x_2 \\ a_2 x_3 & 0 & a_2 x_1 \\ a_3 x_2 & a_3 x_1 & 0 \end{pmatrix} \quad \frac{\partial f}{\partial u}(x, u) = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$$

donc

$$F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad G = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$$

et le système linéarisé $\dot{x} = Fx + Gu = Gu$ n'est évidemment pas commandable. Remarquons aussi que même si l'on avait un troisième coefficient b_3 dans G, le résultat serait identique puisque l'on ne dispose que d'une seule commande pour 3 dimensions.

• *Point* 2 : On va considérer un système scalaire dont la non-linéarité est quadratique (donc non bornée) :

$$\left.\begin{aligned} \dot{x}(t) &= u(t)\, x^2(t) \\ x(0) &= x_0 \neq 0 \end{aligned}\right\} \qquad (2)$$

$$u \in \mathbb{R} \qquad x \in \mathbb{R}$$

et on va considérer le problème de la robustesse de la linéarisée de (2) pour une erreur sur $| x_0 |$, lorsque $| x_0 |$ est très voisin de 0.

Notons d'abord que la solution de (2) est donnée par

$$x(t) = \frac{x_0}{1 - x_0 \int_0^t u(s)\, ds} \quad \forall t > 0 \quad \text{tel que} \quad 1 - x_0 \int_0^t u(s)\, ds \neq 0 \qquad (3)$$

Si $u(t) = -a\,(a > 0)$ et $x_0 > 0$, il est clair que le système (2) est stabilisé puisque

$$\lim_{t \to \infty} x(t) = \lim_{t \to \infty} \frac{x_0}{1 + a x_0\, t} = 0 \qquad (4)$$

et ceci pour tout $x_0 > 0$ puisque

$$\lim_{x_0 \to 0_+} \frac{x_0}{1 + ax_0\, t} = 0 \quad \forall t$$

Notons que pour $x_0 < 0$, le système (2) diverge en temps fini avec $u = -a$, puisque pour $t_1(x_0) = -\dfrac{1}{ax_0} > 0$, on a

$$\lim_{\substack{t \to t_1(x_0) \\ 0 < t < t_1(x_0)}} \frac{x_0}{1 + ax_0\, t} = +\infty \quad \text{et} \quad \lim_{x_0 \to 0} t_1(x_0) = 0_-$$

Si maintenant on linéarise autour de la trajectoire (3), on obtient

$$\dot{y}(t) = -\frac{x_0}{1 + ax_0\, t}\, y(t) + \frac{x_0^2}{(1 + ax_0\, t)^2}\, v(t) \tag{5}$$

En choisissant la commande v, en boucle fermée, comme suit

$$v(t, y, x_0) = \frac{(1 + ax_0\, t)\,((2 - \varepsilon t)\, ax_0 - \varepsilon)}{x_0^2}\, y \tag{6}$$

avec $\varepsilon > 0$, on obtient $\dot{y} = -\varepsilon y$ qui est bien un système stable. Or

$$\lim_{x_0 \to 0} \frac{v(t, y, x_0)}{y} = +\infty \quad \forall (t, y) \quad \forall \varepsilon \tag{7}$$

Ainsi, la commande (6) qui assure la stabilité de (5) autour de la trajectoire de référence $\dfrac{x_0}{1 + ax_0\, t}$, n'est pas robuste puisque son gain tend vers $+\infty$ lorsque x_0 approche de 0.

• *Point* 3 : On va considérer un système *NL* stochastique donné, au sens de Stratonovitch, par

$$(NLS) \begin{cases} dx_t = f(x_t)\, dt + g(x_t)\, dv_t \\ dy_t = h(x_t)\, dt + dw_t \end{cases}$$

ou, au sens de Itô (de manière équivalente), par

$$(NLI) \begin{cases} dx_t = \left(f(x_t) + \dfrac{1}{2} \dfrac{\partial g}{\partial x}(x_t)\, g(x_t) \right) dt + g(x_t)\, \overline{d}v_t \\ dy_t = h(x_t)\, dt + \overline{d}w_t \end{cases}$$

où $\begin{pmatrix} dv \\ dw \end{pmatrix}$ est la différentielle du brownien $\begin{pmatrix} v \\ w \end{pmatrix}$ au sens de Stratonovitch et où $\begin{pmatrix} \bar{d}v \\ \bar{d}w \end{pmatrix}$ est la différentielle de $\begin{pmatrix} v \\ w \end{pmatrix}$ au sens de Itô.

On rappelle que y est le processus d'observation de l'état x. On va montrer que la linéarisation rigoureuse (NLS) s'effectue comme pour les systèmes déterministes, mais ne correspond pas à la linéarisation utilisée pour le filtrage de Kalman étendu qui, elle, n'est pas rigoureuse.

On peut montrer [7] que la linéarisation de (NLS) ou de (NLI) a bien un sens et consiste à linéariser (NLS) au sens de Stratonovitch. Elle est donnée par

$$(NLSL)\begin{cases} dz_t = \dfrac{\partial f}{\partial x}(x_t)\, z_t\, dt + \dfrac{\partial g}{\partial x}(x_t)\, z_t\, dv_t \\ dv_t = \dfrac{\partial h}{\partial x}(x_t)\, z_t\, dt + dw_t \end{cases}$$

Insistons sur le fait important que $(NLSL)$ n'est pas un système linéaire mais *bilinéaire* puisque z_t se multiplie au bruit dv_t, et donc que le *procédé de linéarisation classique n'est pas, en stochastique, un procédé de linéarisation.*

Pour appliquer les méthodes de filtrage, il faut en plus traduire $(NLSL)$ au sens de Itô. Pour cela, il suffit de remplacer $\dfrac{\partial f}{\partial x}(x_t)\, z_t$ par

$$\frac{\partial f}{\partial x}(x_t)\, z_t + \frac{1}{2}\left(\frac{\partial}{\partial x}\left(\frac{\partial g}{\partial x}(x_t)\, z_t\right)\right)\frac{\partial g}{\partial x}(x_t)\, z_t =$$

$$= \frac{\partial f}{\partial x}(x_t)\, z_t + \frac{1}{2}\, z_t^T\, \frac{\partial^2 g}{\partial x^2}(x_t)\, \frac{\partial g}{\partial x}(x_t)\, z_t + \frac{1}{2}\, \frac{\partial g}{\partial x}(x_t)\, \frac{\partial z_t}{\partial x}\, \frac{\partial g}{\partial x}(x_t)\, z_t \qquad (8)$$

Or dans le second membre de (8) apparaît

$$\frac{\partial z}{\partial x} = \frac{\partial^2 \varphi_t}{\partial x^2}(x, \omega)$$

où $\varphi_t(x, \omega)$ est la solution de (NLS) à trajectoire $v_t(\omega)$ fixée et pour la condition initiale $\varphi_0(x, \omega) = x$. Rappelons qu'avec cette notation,

$$z_t = \frac{\partial \varphi_t}{\partial x}(x, \omega)$$

Ainsi, pour résoudre (*NLSL*) au sens de Itô, il nous faut connaître

$$\frac{\partial z_t}{\partial x} = \frac{\partial^2 \varphi_t}{\partial x^2}(x, \omega)$$

L'équation donnant $\partial z_t/\partial x$ peut être calculée comme précédemment en relinéarisant (*NLSL*) au sens de Stratonovitch, mais, lors de la traduction au sens de Itô, apparaîtra

$$\frac{\partial^2 z_t}{\partial x^2} = \frac{\partial^3 \varphi_t}{\partial x^3}(x, \omega)$$

et ainsi de suite.

Pour conclure, on a montré que le procédé de linéarisation classique ne permettait pas de linéariser en stochastique, et, pire, les équations au sens de Itô donnant $\frac{\partial \varphi_t}{\partial x}(x, \omega)$ sont en *nombre infini*. Or le procédé de linéarisation *formelle* du filtre de Kalman étendu qui consiste simplement à écrire, $\hat{x}_t$ désignant l'estimée à l'instant t

$$dz_t = \frac{\partial f}{\partial x}(\hat{x}_t)\, z_t\, dt + g(\hat{x}_t)\, dv_t$$

sans toucher au terme de bruit, ne construit pas un filtre des *variations au premier ordre autour d'une trajectoire nominale*, même dans le cas particulier où $g(x) \equiv G$, matrice constante, puisque $\partial G/\partial x \equiv 0$. En fait, le filtre de Kalman étendu correspond à un développement de f au 1er ordre et de g à l'ordre 0, et l'amélioration de la précision sur f n'a aucune raison d'être significative comparée à la mauvaise précision sur g.

• *Point* 4 : Jusqu'à présent, nous avons examiné les problèmes liés à la linéarisation « simple » mais on peut objecter qu'avec une suite de modèles linéarisés, donc un modèle linéaire ou affine par morceaux, le comportement entrées-sorties du modèle *NL* doit être mieux approché. En fait il n'en est rien comme va nous le montrer l'exemple ci-dessous (inspiré de [8]).

On suppose que l'on connaît le flot des trajectoires d'un système inconnu dans un certain voisinage, l'état x étant scalaire.

Le flot est donné par la figure 1.

Considérons la trajectoire en trait plein de la figure 1, échantillonnons-la aux instants $t_0 = 0, t_1 = 1, \ldots, t_k = k, \ldots$, et notons x_k l'ordonnée du point d'abscisse $t_k = k$.

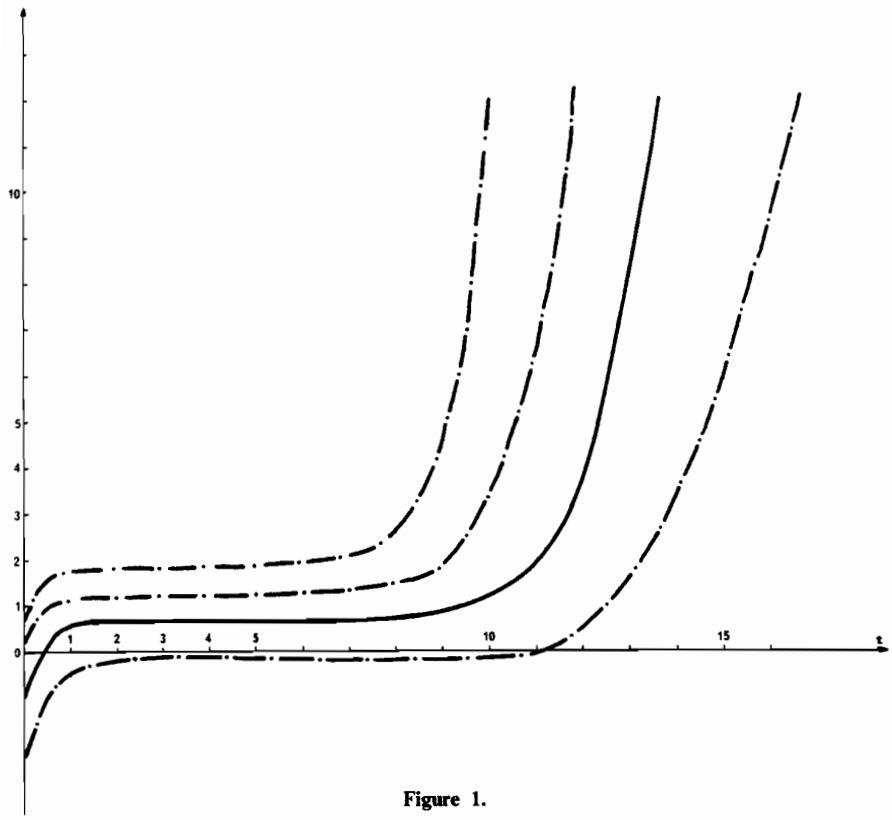


Figure 1.

On obtient

$$\begin{array}{llllllll} x_0 = -1{,}111\,2 & x_1 = 0{,}666\,8 & x_2 = 0{,}667\,0 & x_3 = 0{,}667\,5 \\ x_4 = 0{,}668\,8 & x_5 = 0{,}671\,9 & x_6 = 0{,}679\,7 & x_7 = 0{,}699\,2 \\ x_8 = 0{,}748\,0 & x_9 = 0{,}870\,1 & x_{10} = 1{,}175\,3 & x_{11} = 1{,}938\,2 \\ x_{12} = 3{,}845\,6 & x_{13} = 8{,}614\,0 & x_{14} = 20{,}534\,9 & x_{15} = 50{,}337\,2 \end{array}$$

On se pose alors le problème de trouver un modèle affine par morceaux, de la forme

$$\begin{cases} x_{k+1} = F_k x_k + g_k \overset{\text{def}}{=} f_k(x_k) \quad k = 0, \ldots, 14 \\ x_0 = -1{,}111\,2 \end{cases} \tag{9}$$

où F_k et g_k sont à déterminer.

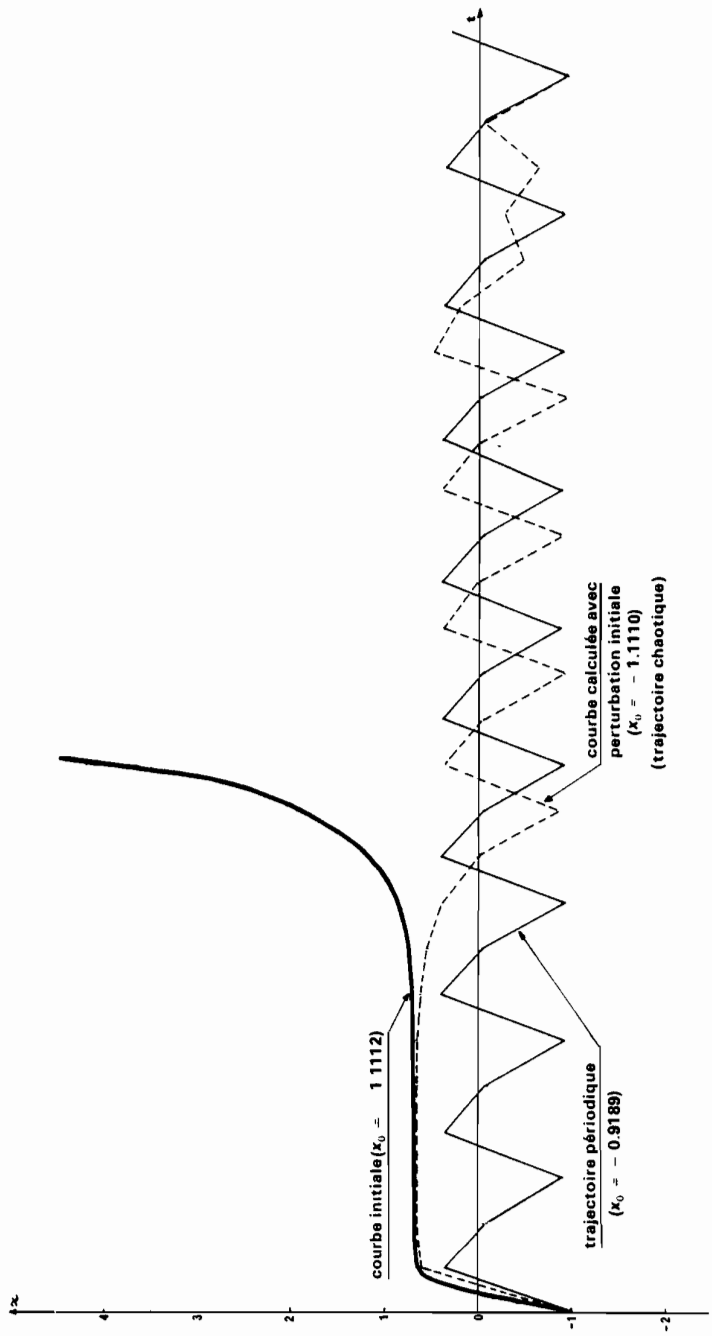

courbe initiale (x_0 = 1.1112)
courbe calculée avec perturbation initiale (x_0 = − 1.1110) (trajectoire chaotique)
trajectoire périodique (x_0 = − 0.9189)
x
t
4
3
2
1
0
-1
-2

Figure 2.

On vérifie que le modèle suivant

$$\begin{cases} \frac{5}{2}x - 1 & \text{si} \quad x \geqslant 0 \\ -\frac{3}{2}x - 1 & \text{si} \quad x < 0 \end{cases} \tag{10}$$

redonne exactement les points $x_0, ..., x_{15}$.

Or, si l'on perturbe légèrement la condition initiale $\bar{x} = -1{,}111\,0$ (soit une variation de 2.10^{-4}), on trouve avec ce modèle (9), (10), la trajectoire de la figure 2 dont on montre qu'elle est *chaotique*, x_k variant toujours entre $-\frac{10}{9}$ et $\frac{2}{3}$

Donc l'approximation (9), (10) qui est exacte pour la trajectoire partant de $x_0 = -1{,}111\,2$, et donc qui suit une trajectoire monotone non décroissante, donne un comportement qualitativement différent lorsque l'on fait varier x_0 de 2.10^{-4} puisqu'alors on obtient une trajectoire non monotone et bornée ! On peut même, en perturbant un peu plus x_0, obtenir des trajectoires périodiques de période arbitraire ; par exemple, pour la période 3

$$\begin{cases} x_0 = -\frac{34}{37} \simeq -0{,}918\,9 \quad x_1 = \frac{14}{37} \simeq 0{,}378\,3\,, \; x_2 = -\frac{2}{37} \simeq -0{,}054\,1 \\ x_3 = x_0 \quad x_4 = x_1 \quad x_5 = x_2, \text{ etc...} \end{cases}$$

En conclusion, contrairement à ce qui se passe en linéaire, la donnée du champ le long d'une trajectoire ne permet pas de décrire ce qui se passe autour, et les méthodes de linéarisation par morceaux sont ici en défaut.

II. Quelques points spécifiques aux systèmes *NL* à temps continu

Nous savons que les systèmes linéaires jouissent de propriétés très particulières comme : existence et unicité de la solution sur $]-\infty, +\infty[$ tout entier (pas d'explosion en temps fini), l'ensemble de ces solutions engendrant un espace vectoriel (l'espace d'état) lorsque l'on fait varier les lois de commande sans contraintes. De telles propriétés cessent généralement d'exister pour des systèmes *NL*, et l'espace d'état devient une variété différentiable, variété qui peut quelquefois avoir une structure topologique délicate à manier.

D'autre part, il va sans dire que de nombreux phénomènes non triviaux comme les bifurcations, peuvent apparaître en *NL*, mais nous n'aborderons pas ces problèmes ici.

Donnons quelques exemples élémentaires

a) *Non-unicité*

Le système

$$\left.\begin{aligned} \dot{x} &= u(t)\, x^m \\ x(0) &= x_0 \end{aligned}\right\}, \tag{1}$$

avec $u(t)$ fonction mesurable bornée et $0 < m < 1$, a pour solution générale

$$x(t) = \left(x_0^{1-m} + (1 - m) \int_0^t u(s)\, ds \right)^{\frac{1}{1-m}}$$

Or, pour $x_0 = 0$, il est clair que $x(t) = 0 \quad \forall t$ est une solution triviale et donc

$$x(t) = \left((1 - m) \int_0^t u(s)\, ds \right)^{\frac{1}{1-m}} \quad \text{et} \quad x(t) = 0$$

pour $x(0) = 0$ et $u \not\equiv 0$, sont deux solutions distinctes de (1).

Remarquons bien sûr que la fonction x^m n'est pas Lipschitzienne au voisinage de 0 lorsque $0 < m < 1$.

b) *Explosion*

Le système

$$\left.\begin{aligned} \dot{x} &= u(t)\, x^p \\ x(0) &= x_0 \end{aligned}\right\} \tag{2}$$

toujours avec u mesurable borné, mais, cette fois, avec $p > 1$, admet pour unique solution

$$x(t) = \frac{x_0}{\left(1 - (p - 1)\, x_0^{p-1} \int_0^t u(s)\, ds \right)^{\frac{1}{p-1}}} \tag{3}$$

Or si le dénominateur s'annule, c'est-à-dire si l'équation

$$\int_0^{t_1} u(s)\, ds = \frac{1}{(p - 1)\, x_0^{p-1}} \tag{4}$$

admet une solution $t_1 \in]0, +\infty[$, on a

$$\lim_{\substack{t \to t_1 \\ t < t_1}} | x(t) | = +\infty \text{ (explosion au temps } t_1 \text{ fini)}$$

en particulier, si $u(t) \equiv v \quad \forall t$, avec $v > 0$ et $x_0 > 0$, l'équation (4) devient

$$vt_1 = \frac{1}{(p-1)\, x_0^{p-1}}$$

et l'explosion a lieu au temps

$$t_1 = \frac{1}{v(p-1)\, x_0^{p-1}}$$

(Insistons sur le fait que t_1 dépend de la commande utilisée et de la condition initiale.)

c) *Variété d'État*

Considérons le système bilinéaire :

$$\left.\begin{aligned} \dot{x}_1 &= \quad u(t)\, x_2 \quad x_1(0) = \xi_1 \\ \dot{x}_2 &= -u(t)\, x_1 \quad x_2(0) = \xi_2 \end{aligned}\right\} \tag{5}$$

avec u mesurable et bornée.

Il est clair que l'on a $x_1\, \dot{x}_1 + x_2\, \dot{x}_2 = 0 \quad \forall u$, et donc

$$x_1^2 + x_2^2 = \xi_1^2 + \xi_2^2 \quad \forall u \tag{6}$$

(6) exprime qu'aucune commande u ne pourra faire sortir l'état du cercle C de centre l'origine et de rayon $(\xi_1^2 + \xi_2^2)^{1/2}$ qui joue donc le rôle de variété d'État. La commande u ne sert qu'à faire varier la vitesse de parcours sur C.

Remarquons enfin que la variété C est de dimension 1 alors que (5) est de dimension 2, et ne constitue donc pas la réalisation minimale. Celle-ci s'écrit simplement

$$\dot{x} = u \text{ en abscisse curviligne !}$$

Donnons un autre exemple un peu plus riche

$$\left.\begin{aligned} \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} &= \begin{pmatrix} 0 \\ x_3 \\ -x_2 \end{pmatrix} u_1 + \begin{pmatrix} -x_3 \\ 0 \\ x_1 \end{pmatrix} u_2 + \begin{pmatrix} x_2 \\ -x_1 \\ 0 \end{pmatrix} u_3 \\ x_1(0) &= \xi_1,\ x_2(0) = \xi_2,\ x_3(0) = \xi_3 \end{aligned}\right\} \tag{7}$$

On vérifie immédiatement que $x_1 \dot{x}_1 + x_2 \dot{x}_2 + x_3 \dot{x}_3 = 0 \quad \forall (u_1, u_2, u_3)$ et $\forall t$.

La variété d'état est donc la sphère S_2 de $\mathbb{R}^3$ de centre (0, 0, 0) et de rayon $(\xi_1^2 + \xi_2^2 + \xi_3^2)^{1/2} = R$, u_1, u_2 et u_3 représentant la vitesse de parcours sur S_2.

Comme précédemment, dim $S_2 = 2$ et une réalisation minimale *locale* de (7) peut être calculée, par exemple, en coordonnées en projection stéréographique

$$X = \frac{2Rx_1}{x_3 + R} \qquad Y = \frac{2Rx_2}{x_3 + R} \qquad Z = R$$

on obtient

$$\begin{cases} \begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} = \frac{u_1}{2R} \begin{pmatrix} XY \\ 2R^2 - \dfrac{X^2 - Y^2}{2} \end{pmatrix} - \frac{u_2}{2R} \begin{pmatrix} 2R^2 + \dfrac{X^2 - Y^2}{XY^2} \end{pmatrix} + u_3 \begin{pmatrix} Y \\ -X \end{pmatrix} \\ X(0) = \dfrac{2R\xi_1}{\xi_3 + R}; \quad Y(0) = \dfrac{2R\xi_2}{\xi_3 + R} \end{cases} \tag{8}$$

Précisons que (8) n'est qu'une réalisation locale puisque la transformation utilisée est singulière au point $x_3 = -R$. On laisse au lecteur le soin d'écrire la réalisation de (7) au voisinage de $x_3 = -R$ en utilisant cette fois

$$X' = \frac{-2Rx_1}{x_3 - R} \qquad Y' = \frac{-2Rx_2}{x_3 - R}$$

Notons qu'on pourrait aussi utiliser d'autres changements de coordonnées pour écrire une réalisation minimale de (7), comme les coordonnées polaires (latitude, longitude), etc..., mais tous contiennent des singularités et il n'existe pas ici de réalisation minimale globale.

d) *Singularité topologique* [9]

On va montrer que pour le système bilinéaire

$$\left.\begin{aligned} \dot{x}_1 &= x_2 & x_1(0) &= 1 \\ \dot{x}_2 &= -(1 - u)\, x_1 & x_2(0) &= 0 \end{aligned}\right\} \tag{9}$$

avec $|u| < \varepsilon \ll 1$, l'ensemble des points que l'on peut atteindre en un temps « petit » est simplement connexe, alors qu'il ne l'est plus à partir de l'instant $T = 2\pi$.

La solution de (9) pour u constant est donnée par

$$(1 - u)\, x_1^2 + x_2^2 = R^2 \tag{10}$$

Pour $u \geqslant 0$, ces courbes sont des ellipses de foyers

$$\left(\pm R \sqrt{\frac{u}{1-u}}, 0 \right) \quad \text{et de grand axe} \quad \frac{2R}{\sqrt{1-u}};$$

pour $u \leqslant 0$, ce sont les ellipses de foyers

$$\left(0, \pm R \sqrt{\frac{-u}{1-u}} \right) \quad \text{et de grand axe } 2R$$

On montre par des considérations géométriques élémentaires que l'ensemble atteignable $A(1, 0)$ à partir de $(1, 0)$ et pour $x_1 \geqslant 0$, $x_2 \leqslant 0$ est situé entre les 2 branches d'ellipse dont la branche supérieure est donnée par (voir *fig.* 3)

$$(e_1^1) \quad (1 - \varepsilon)\, x_1^2 + x_2^2 = 1 - \varepsilon \quad (u = +\varepsilon) \qquad x_1 \geqslant 0 \qquad x_2 \leqslant 0$$

et dont la branche inférieure est donnée par

$$(e_2^1) \quad (1 + \varepsilon)\, x_1^2 + x_2^2 = 1 + \varepsilon \quad (u = -\varepsilon) \qquad x_1 \geqslant 0 \qquad x_2 \leqslant 0$$

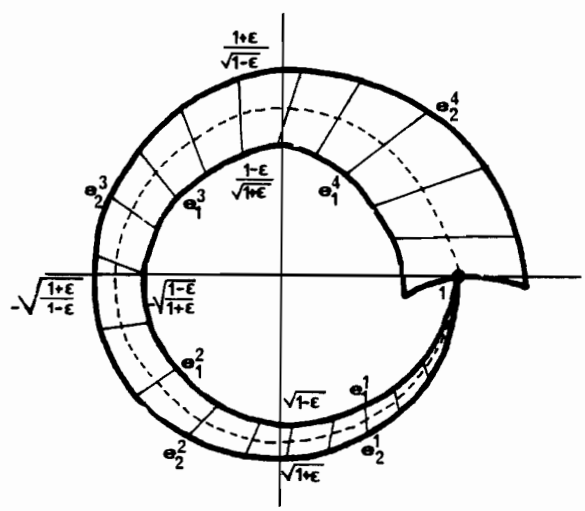


Figure 3. — $A_{2\pi}(1, 0)$.

sur (e_1^1), le temps mis à atteindre $x_1 = 0$ est $t_1 = \pi/2\sqrt{1-\varepsilon}$ et sur (e_2^1), il est $t_1' = \pi/2\sqrt{1+\varepsilon} < t_1$.

En particulier, l'ensemble atteignable avant t_1, noté $A_{t_1}(1, 0)$, est clairement simplement connexe.

Continuons à construire l'ensemble $A(1, 0)$. Il est situé entre les courbes e_1 et e_2 (voir *fig.* 3), avec

$$e_1 = \{ e_1^1, e_1^2, e_1^3, e_1^4 \} \quad \text{et} \quad e_2 = \{ e_2^1, e_2^2, e_2^3, e_2^4 \}$$

les e_i^j étant les branches d'ellipse suivantes (e_1^1 et e_1^2 étant définies plus haut)

$$e_1^2 = (1+\varepsilon)\, x_1^2 + x_2^2 = 1 - \varepsilon \qquad x_1 \leqslant 0, x_2 \leqslant 0 \quad (u = -\varepsilon)$$

$$e_1^3 = (1-\varepsilon)\, x_1^2 + x_2^2 = \frac{(1-\varepsilon)^2}{1+\varepsilon} \qquad x_1 \leqslant 0, x_2 \geqslant 0 \quad (u = +\varepsilon)$$

$$e_1^4 = (1+\varepsilon)\, x_1^2 + x_2^2 = \frac{(1-\varepsilon)^2}{1+\varepsilon} \qquad x_1 \geqslant 0, x_2 \geqslant 0 \quad (u = -\varepsilon)$$

et

$$e_2^2 = (1-\varepsilon)\, x_1^2 + x_2^2 = 1 + \varepsilon \qquad x_1 \leqslant 0, x_2 \leqslant 0 \quad (u = +\varepsilon)$$

$$e_2^3 = (1+\varepsilon)\, x_1^2 + x_2^2 = \frac{(1+\varepsilon)^2}{1-\varepsilon} \qquad x_1 \leqslant 0, x_2 \geqslant 0 \quad (u = -\varepsilon)$$

$$e_2^4 = (1-\varepsilon)\, x_1^2 + x_2^2 = \frac{(1+\varepsilon)^2}{1-\varepsilon} \qquad x_1 \geqslant 0, x_2 \geqslant 0 \quad (u = +\varepsilon)$$

On voit alors facilement que $\forall t < 2\pi$, l'ensemble $A_t(1, 0)$ des états atteignables par le système (9) à partir du point (1, 0) et dans l'intervalle de temps $[0, t]$, est simplement connexe, alors que pour $t \geqslant 2\pi$, $A_t(1, 0)$ contient des trajectoires périodiques et n'est plus simplement connexe à cause du voisinage de l'origine non contenu dans $A_t(1, 0)$. Il s'opère donc une singularité topologique de l'ensemble atteignable pour $t = 2\pi$.

B. DESCRIPTIONS EXTERNE ET INTERNE DES SYSTÈMES NON LINÉAIRES

I. Quelques exemples concrets, et non académiques, de systèmes *NL*

La liste ci-dessous n'est évidemment pas exhaustive, mais donne une indication sur la façon dont on a cherché à appliquer jusqu'à présent la théorie non linéaire.

a) *En Aéronautique*

De nombreux problèmes familiers des ingénieurs comme le contrôle de l'attitude d'un satellite rigide [10], comme la commande non interactive des avions en mouvements rapides [11], [12] ou comme la modélisation et la

commande du vol des hélicoptères [13] ont contribué à démontrer l'importance et la spécificité du non linéaire là où le linéaire n'apportait pas de réponse viable.

b) *En Robotique*

Certaines tentatives d'utilisation du NL existent déjà pour la modélisation et la commande des bras manipulateurs souples [14].

c) *En Biologie et en Chimie*

Les systèmes bilinéaires y ont très tôt été introduits pour modéliser l'évolution des recombinaisons de certaines molécules dans une solution. Leur utilité théorique est très largement reconnue dans ces domaines [15].

d) *En Pétrochimie*

La régulation et la commande non interactive des colonnes à distiller [16], [17] ont aussi apporté un exemple d'application du NL où le linéaire ne donnait que des résultats médiocres.

e) *Dans les centrales nucléaires*

Pour l'identification du comportement en charge par un modèle bilinéaire obtenu à partir des modèles linéarisés en différents points de fonctionnement [18].

f) *Dans les circuits électriques et électroniques*

Le convertisseur de courant continu par commutations entre circuits RLC [20] est l'un des premiers essais d'application des systèmes bilinéaires.

g) *Dans les réseaux de distribution d'eau et d'électricité*

D'une part se pose le problème de la classification en équilibres stables et instables des solutions du système d'équations non linéaires du régime permanent du réseau, — pour les réseaux électriques, une littérature abondante existe [21], [22] —, et d'autre part, le pilotage du réseau où commence à se développer une généralisation des méthodes par échelles de temps du linéaire [23],[24].

Remarque d'orientation

Il est facile de voir sur cette suite d'exemples que la « théorie » non linéaire est multiforme suivant que l'on s'intéresse à une modélisation fine ou à l'optimisation d'un système aggrégé, le formalisme est radicalement différent. D'une part la modélisation fine demande une structure précise (structure bilinéaire par exemple) pour pouvoir répondre à des questions aussi délicates que la commandabilité ou l'observabilité, et d'autre part l'optimisation des réseaux se contente de modèles relativement grossiers du moment que le nombre de variables d'état soit diminué et que le système réduit ainsi que son système de commande soient robustes, c'est-à-dire permettent d'obtenir du système réel les performances désirées en dépit des erreurs de modélisation. Dès lors, le lecteur pourra arguer que sans modèle précis, on ne peut avoir d'idée précise

sur le système ou, au contraire, que la précision des résultats avec un modèle fin n'est qu'apparente si les données pour construire ce modèle sont peu fiables et/ou peu nombreuses. Nous nous garderons bien de conclure et nous insisterons par contre sur l'enrichissement potentiel et réciproque des 2 approches (et des autres à venir) dans un tel débat !

Pour essayer de dresser un tableau synthétique sur les résultats représentatifs du non-linéaire, nous avons groupé toutes les théories et méthodes ayant un « degré de parenté » évident les théories géométrique et algébrique des systèmes bilinéaires ou, plus généralement, non linéaires-analytiques sont le pendant direct des méthodes interne et externe du linéaire et sont donc groupées dans le chapitre « descriptions externe et interne des systèmes non linéaires ».

Nous avons par contre séparé cette dernière approche de celle par perturbations singulières, aggrégation et cohérence, qui n'a pas les mêmes objectifs.

Pour terminer, en ce qui concerne la forme, nous nous contenterons le plus souvent de faire une bibliographie commentée, en introduisant juste ce qu'il faut des concepts mathématiques pour pouvoir garder le fil. Par contre, nous n'hésiterons pas à illustrer les résultats par des exemples lorsque ceux-ci sont suffisamment simples et... courts.

II. Comment faire un système non linéaire avec plusieurs systèmes linéaires ?

On va donner une interprétation (presque générique) d'une classe importante de NL : les bilinéaires. Pour éviter de compliquer l'exposé, considérons seulement 2 systèmes linéaires

$$\dot{x} = A_1 x \quad \text{et} \quad \dot{x} = A_2 x$$

et supposons que, entre ces 2 systèmes, un interrupteur nous serve à piloter le dispositif (voir *fig.* 4).

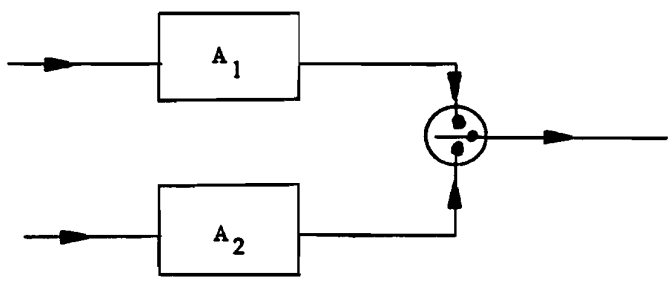


Figure 4.

Les deux seuls degrés de liberté dont nous disposons sont donc le temps passé entre deux commutations et le choix du système sur lequel on commute.

Modélisation

On introduit les variables de commande u^1 et u^2 à valeurs binaires 0 ou 1, avec la contrainte $u^1 + u^2 \leqslant 1$ puisque les 2 systèmes A_1 et A_2 ne peuvent être branchés simultanément (ce point n'est mentionné que pour la vraisemblance de la modélisation mais ne constitue en rien une limitation théorique !). Le système de la figure 4 est alors donné par

$$\left.\begin{array}{l} \dot{x}(t) = u^1(t)\, A_1\, x(t) + u^2(t)\, A_2\, x(t) = (u^1(t)\, A_1 + u^2(t)\, A_2)\, x(t) \\ u^1(t) \quad u^2(t) \in \{ 0, 1 \} \quad u^1(t) + u^2(t) \leqslant 1 \quad \forall t \end{array}\right\} \tag{1}$$

Un tel système est appelé *bilinéaire* ou régulier, la non-linéarité venant de la multiplication de l'état par la commande. On trouvera un tel dispositif par exemple dans les convertisseurs de courant continu (voir [20]).

Discussion

- Il est clair que (1) ne peut pas être représenté par un système linéaire : (1) est, par construction, linéaire par morceaux.
- Il ne s'agit pas de contrôle impulsionnel puisque les trajectoires de (1) ne comportent pas de sauts mais des commutations (sauts sur la dérivée première).
- Il existe par ailleurs d'autres techniques apparemment proches pour modéliser des systèmes à commutations comme les techniques d'équations différentielles multivoques qui s'avèrent beaucoup moins riches du point de vue qualitatif, les inéquations variationnelles ou les systèmes linéaires implicites dont le spectre, plus restreint, reste à la frontière de ce type de problème.

III. Nomenclature des systèmes *NL*

a) *Bilinéaires ou Réguliers* (voir § II)

Un système bilinéaire est un système *NL* de la forme

$$(BL) \left\{ \begin{array}{l} \dot{x} = \left(\sum_{i=1}^{N} u^i\, A_i \right) x \\ y = Cx \end{array} \right.$$

où C est la matrice d'observation et les A_i des matrices carrées (n, n), n étant la dimension du vecteur d'état x. Les matrices peuvent dépendre du temps. Dans ce cas, il s'agit d'un système bilinéaire non stationnaire.

PROPOSITION *Les systèmes linéaires sont contenus dans les systèmes bilinéaires.*

Démonstration : Considérons le système linéaire

$$\left.\begin{aligned} \dot{x} &= Fx + Gu \\ y &= Hx \end{aligned} \qquad x(0) = x_0 \right\} \tag{1}$$

avec $x \in R^k$, $u \in R^m$ et $y \in R^p$, $F \in R^{k^2}$, $G \in R^{km}$ $H \in R^{pk}$

Notons $0_{\alpha,\beta}$ la matrice nulle de $R^{\alpha\beta}$, G_i la i-ième colonne de G et

$$A_0 = \left(\begin{array}{c|c} F & 0_{k,1} \\ \hline 0_{1,k} & 0_{1,1} \end{array}\right) \qquad A_i = \left(\begin{array}{c|c} 0_{k,k} & G_i \\ \hline 0_{1,k} & 0_{1,1} \end{array}\right) \quad i = 1, \ldots, m$$

$C = (H \mid 0_{p,1})$, puis notons

$$u_0 = 1 \qquad \xi = 1 \qquad X = \begin{pmatrix} x \\ \xi \end{pmatrix} \in R^{k+1} \qquad U = \begin{pmatrix} u_0 \\ u \end{pmatrix} \in R^{m+1}$$

On vérifie alors facilement que le système bilinéaire

$$\left.\begin{aligned} \dot{X} &= \left(\sum_{i=0}^{m} U^i A_i\right) X \qquad X(0) = \begin{pmatrix} x_0 \\ 1 \end{pmatrix} \\ y &= CX \end{aligned}\right\} \tag{3}$$

est « indistinguable » du système linéaire (2), d'où le résultat. ■

Par un raisonnement analogue, on montre facilement que les bilinéaires-affines

$$(BLA) \left\{ \begin{aligned} \dot{x} &= \sum_{i=1}^{N} u^i (A_i x + b_i) \\ y &= Cx \end{aligned} \right.$$

sont eux aussi des cas particuliers des bilinéaires. On ne gagne donc pas en généralité en rajoutant des termes affines. Il n'en va pas de même lorsqu'on rajoute des termes non linéaires.

b) *Linéaires-analytiques*

Ce sont les systèmes linéaires en l'entrée et analytiques en l'état

$$(LA)\begin{cases} \dot{x} = f_0(x) + \sum_{i=1}^{N} u^i f_i(x) \\ y = h(x) \end{cases}$$

lorsque h, $f_0, \ldots, f_N$ sont analytiques. (On fera souvent l'abus de langage consistant à désigner par le même nom un système où h, $f_0, \ldots, f_N$ sont C^∞.)

En fait, en toute rigueur, et compte tenu du fait que le vecteur d'état est supposé évoluer dans une variété (C^∞ ou analytique), il faut parler de $f_0, \ldots, f_N$ en terme de *champs de vecteurs* et l'écriture de (LA) est alors dans un système de *coordonnées locales.*

Précisément, dans un système de coordonnées locales $(x^1, \ldots, x^n)$, les $f_0, \ldots, f_N$ sont les coefficients des champs de vecteurs $F_0, \ldots, F_N$ associés par la formule :

$$F_i(x) = \sum_{j=1}^{n} f_i^j(x) \frac{\partial}{\partial x^j} \quad i = 0, \ldots, N$$

Les F_i étant, eux, intrinsèques, c'est-à-dire invariants par changements de coordonnées.

Le cas où les f_i sont des polynômes a longtemps été étudié séparément, mais, étant donnée leur faible spécificité par rapport à (LA) nous n'en parlerons pas (voir [25]).

On retrouve souvent, par ailleurs, le reproche que les systèmes non linéaires étudiés ne sont généralement pas vraiment intéressants parce que trop « réguliers ». En fait, dans la pratique, il est bien rare de trouver des non-linéarités qui ne sont pas C^∞ ou analytiques. Des non-linéarités sinusoïdales ou quadratiques sont généralement considérées par les ingénieurs comme redoutables à juste titre. Ceci ne veut pas dire qu'il ne faille pas considérer de systèmes plus généraux, mais cette remarque véhicule fréquemment une confusion entre comportement non linéaire et comportement discontinu.

c) *Non linéaires généraux. Approximation bilinéaire*

$$(NLG)\begin{cases} \dot{x} = f(x, u) \\ y = h(x) \end{cases}$$

De tels systèmes sont très peu étudiés d'abord pour leur trop grande généralité, et aussi grâce au résultat d'approximation suivant dû à Fliess [26].

En fait, au lieu du système (NLG) on va considérer la relation entrées-sorties causale qu'il définit, ou, plus généralement encore, n'importe quelle relation entrées-sorties causale et continue sur $T \times C$ où T (resp. C) est un compact de $[0, \infty[$ (resp. de $C^0(T)$ espace des fonctions continues sur T muni de la topologie de la convergence uniforme).

THÉORÈME D'APPROXIMATION *Tout système non linéaire décrit par une fonctionnelle entrées-sorties causale et continue sur $T \times C$ peut être arbitrairement approché (au sens de la topologie de la convergence uniforme) par des bilinéaires de dimension finie que l'on peut choisir nilpotents.* ■

On trouvera un résultat légèrement différent dans Sussmann [27], et un résultat analogue dans le cas des systèmes à temps discret (échantillonnés) [26] où la nomenclature précédente s'adapte de manière évidente.

IV. Description externe de systèmes *NL*. Approche algébrique

Les résultats de ce paragraphe sont empruntés à (Fliess [28], Jacob [29]). On sait que les systèmes linéaires sont décrits de manière externe par leur fonction de transfert et/ou leur réponse impulsionnelle, ces 2 notions étant reliées par une relation biunivoque. Le parallèle exact existe (presque en général) pour les systèmes non linéaires réguliers ou analytiques : l'analogue de la fonction de transfert est ici la *série génératrice*, et la *série de Volterra* tient lieu de réponse impulsionnelle. Avant de préciser ces notions, rappelons la méthode de Peano-Baker pour approximer la solution d'une équation différentielle par une série (aussi appelée méthode des approximations successives).

Soit

$$(BL)\begin{cases} \dot{x} = \left(A_0 + \sum_{i=1}^{N} u^i A_i\right) x \qquad x(0) = x_0 \\ y = Cx \end{cases}$$

Le calcul est classique et consiste à écrire

$$x_1(t) = x_0 + \left(tA_0 + \sum_{i=1}^{N} \int_0^t u^i(s)\, ds A_i\right) x_0$$

puis, par récurrence

$$x_{k+1}(t) = x_0 + \int_0^t \left(A_0 + \sum_{i=1}^{N} u^i(s) A_i\right) x_k(s)\, ds \qquad \forall k > 1$$

On montre que pour t et $\{ u^i \}$ bornés, la limite lorsque $k \to \infty$ des fonctions x_k existe uniformément sur tout compact, que cette limite est x, solution de l'équation différentielle (BL), et que $x(t)$ est donnée par la série de Peano-Baker

$$x(t) = \left[I + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k=0}^{N} A_{j_1} \ldots A_{j_k} \int_0^t u^{j_1}(\tau_1) \int_0^{\tau_1} u^{j_2}(\tau_2) \cdots \times \right.$$

$$\left. \times \int_0^{\tau_{k-1}} u^{j_k}(\tau_k)\, d\tau_k \ldots d\tau_1 \right] x_0 \quad (1)$$

et $y(t) = Cx(t)$.

a) *La série génératrice de* (BL)

On commence par coder les intégrales des entrées par

$$\xi_0(t) = t, \xi_1(t) = \int_0^t u^1(s)\, ds, \ldots, \xi_N(t) = \int_0^t u^N(s)\, ds$$

de sorte que (1) devient

$$y(t) = \left(C + \sum_{k=1}^{\infty} \sum_{j_1,\ldots,j_k=0}^{N} CA_{j_1} \ldots A_{j_k} \int_0^t d\xi_{j_k} \ldots d\xi_{j_1} \right) x_0 \quad (2)$$

l'intégrale $\int_0^t d\xi_{j_k} \ldots d\xi_{j_1}$ étant une notation condensée pour la suite d'intégrales itérées de (1).

Considérons alors l'alphabet $X = \{ \zeta_0, \ldots, \zeta_N \}$, et X^* l'ensemble des mots $\zeta = \zeta_{j_k} \ldots \zeta_{j_1}$ formés à partir de symboles de X.

Posons enfin $\mu(\zeta_i) = A_i \quad i = 0, \ldots, N$, et pour

$$\zeta = \zeta_{j_k} \ldots \zeta_{j_1} \qquad \mu(\zeta) = A_{j_k} \ldots A_{j_1} = \mu(\zeta_{j_k}) \ldots \mu(\zeta_{j_1}) \qquad \forall \zeta \in X^*$$

On définit alors la série formelle rationnelle g en variables non commutatives, dite série génératrice de (BL), par la formule

$$g = \sum_{\zeta \in X^*} C\mu(\zeta)\, x_0\, \zeta \quad (3)$$

qui n'est autre qu'un codage de la formule (2).

Cette série génératrice détermine complètement le comportement entrées-sorties de (BL). On montre, voir l'exemple 2, que pour les systèmes linéaires, cette série redonne la matrice de transfert à un changement de variable près.

On mesure ainsi l'importance du concept de série génératrice qui, comme la matrice de transfert, ne dépend que de la fonctionnelle entrées-sorties et permet, en particulier, de définir correctement la notion de réalisation dans l'espace d'état.

Exemple 1 : La série génératrice de $\dot{x} = ux$ est

$$g = \left(1 + \sum_{k \geqslant 1} \zeta^k\right) x_0 = (1 - \zeta^k)^{-1} x_0 \tag{4}$$

On vérifie facilement que g est aussi la série génératrice de

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = u \begin{pmatrix} 1 & 0 \\ a_1(x_1, x_2) & a_2(x_1, x_2) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ux_1 \\ u(a_1(x_1, x_2)\, x_1 + a_2(x_1, x_2)\, x_2) \end{pmatrix}$$

$$y = (1, 0) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1$$

où a_1 et a_2 sont des fonctions analytiques arbitraires.

Ceci illustre donc bien que g ne dépend que de la fonctionnelle entrées-sorties. ■

Exemple 2 : La série génératrice du système linéaire

$$\left.\begin{aligned} \dot{x} &= Fx + \sum_{i=1}^{N} u^i G_i \quad x(0) = x_0 \\ y &= Hx \end{aligned}\right\} \tag{5}$$

est donnée pour

$$g = H(I - \zeta_0 F)^{-1} \left(x_0 + \sum_{i=1}^{N} \zeta_i G_i\right) \tag{6}$$

En effet, reprenant l'équivalent bilinéaire de III, a), et les combinant avec (2), on est amené à calculer des produits $\sum_{j_0, \ldots, j_k} A_{j_0} \ldots A_{j_k}$ où $A_0, \ldots, A_N$ sont donnés à partir de F et G_i par les formules de III, a). On vérifie alors facilement que

$$A_0 A_i = \left(\begin{array}{c|c} 0_{k,k} & FG_i \\ \hline 0_{1,k} & 0_{1,1} \end{array}\right) \quad \forall i = 1, \ldots, N$$

et $A_i A_j = 0 \quad \forall i = 1, \ldots, N, \quad \forall j = 0, \ldots, N,$ d'où (6).

Finalement, faisons $x_0 = 0$ dans (6). On obtient alors

$$g = \sum_{i=1}^{N} H(I - \zeta_0 F)^{-1} G_i \zeta_i \tag{7}$$

qui n'est autre que la matrice de transfert de (5) au changement de variable près $\zeta_0^k = z^{-k-1}$ ■

b) *La série génératrice de (LA)*

La formule de la série génératrice dans le cas linéaire analytique (en coordonnées locales)

$$(LA) \begin{cases} \dot{x}(t) = f_0(x(t)) + \sum_{i=1}^{N} u^i(t) f_i(x(t)) & x(0) = x_0 \\ y(t) = h(x(t)) \end{cases}$$

avec

$$F_i(x) = \sum_{j=1}^{n} f_i^j(x) \frac{\partial}{\partial x^j} \quad i = 0, ..., N$$

est donnée, pour t et $(\max_{0 \leqslant \tau \leqslant t} \| u(\tau) \|)$ « petits », par

$$g = h_{|x_0} + \sum_{k \geqslant 1} \sum_{j_0, ..., j_k = 0}^{N} F_{j_0} \dots F_{j_k} h_{|x_0} \zeta_{j_k} \dots \zeta_{j_0} \tag{8}$$

et généralise donc exactement (2).

c) *La série de Volterra de (LA)*

Rappelons que l'on appelle série de Volterra une fonctionnelle de la forme

$$(SV) \quad y(t, u) = w_0(t) + \int_0^t w_1(t, \tau_1) u(\tau_1) d\tau_1 +$$

$$+ \int_0^t \int_0^{\tau_1} w_2(t, \tau_1, \tau_2) u(\tau_2) u(\tau_1) d\tau_2 d\tau_1 + \cdots$$

$$+ \int_0^t \int_0^{\tau_1} \cdots \int_0^{\tau_{k-1}} w_k(t, \tau_1, ..., \tau_k) u(\tau_k) \dots u(\tau_1) d\tau_k \dots d\tau_1 + \cdots$$

où w_k est un tenseur d'ordre k, pris sous forme triangulaire pour

$$t \geqslant \tau_1 \geqslant \cdots \tau_k \geqslant 0$$

Pour un système linéaire, la série de Volterra est simplement formée des deux premiers termes. le second étant la réponse impulsionnelle en effet, la série de Volterra du système linéaire (5) s'écrit

$$y(t, u) = H\, e^{Fr}\, x_0 + \sum_{i=1}^{N} \int_0^t H\, e^{F(t-\tau_1)}\, G_i\, u^i(\tau_1)\, d\tau_1 \tag{9}$$

Ainsi, la série de Volterra décrit le comportement *temporel* alors que la série génératrice décrit les entrées-sorties dans un espace abstrait de variables non commutatives généralisant le « domaine fréquentiel ».

On a le résultat général suivant .

THÉORÈME : *Une condition nécessaire et suffisante pour que (SV) définisse une fonctionnelle causale analytique (et donc représente les entrées-sorties d'un système (LA)) est qu'elle vérifie les deux conditions suivantes .*

(i) *les w_k sont analytiques en $(t, \tau_1, ..., \tau_k)$ au voisinage de $(0, ..., 0)$,*

(ii) *les rayons de convergence des w_k sont uniformément minorés par une quantité strictement positive.* ■

Les formules explicites des w_k peuvent être trouvées dans [28].

Exemple : On va donner ces formules dans le cas bilinéaires :

$$\left.\begin{aligned} \dot{x} &= Fx + \sum_{i=1}^{N} u^i\, G_i\, x \quad x(0) = x_0 \\ y &= Hx \end{aligned}\right\} \tag{10}$$

Transformons d'abord (10) en posant : $z(t) = \mathrm{e}^{-Ft}\, x(t)$. On obtient alors

$$\left.\begin{aligned} \dot{z}(t) &= \sum_{i=1}^{N} u^i(t)\, e^{-Ft}\, G_i\, e^{Ft}\, z(t)\,, \quad z(0) = x_0 \\ y(t) &= H\, e^{Ft}\, z(t)\,, \end{aligned}\right\}$$

posant alors $\sum_{i=1}^{N} u^i(t)\, e^{-Ft}\, G_i\, e^{Ft} = \Phi(t)$, et utilisant la série de Peano-Baker

$$y(t) = H\, e^{Ft} \left(I + \sum_{k \geq 1} \int_0^t \int_0^{\tau_1} \cdots \int_0^{\tau_{k-1}} \Phi(\tau_1) \ldots \Phi(\tau_k)\, d\tau_k \ldots d\tau_1 \right) x_0$$

$$= H\, e^{Ft} \left(I + \sum_{k \geq 1} \int_0^t \int_0^{\tau_1} \cdots \int_0^{\tau_{k-1}} \sum_{i_1, \ldots, i_k = 1}^{N} \times \right.$$

$$\left. \times\, e^{-F\tau_1}\, G_{i_1}\, e^{F\tau_1} \ldots e^{-F\tau_k}\, G_{i_k}\, e^{F\tau_k}\, u^{i_1}(\tau_1) \ldots u^{i_k}(\tau_k)\, d\tau_k \ldots d\tau_1 \right) x_0$$

le k-ième noyau de Volterra est donc le tenseur d'ordre k dont le coefficient de $u^{i_1}(\tau_1) \ldots u^{i_k}(\tau_k)$ est

$$H\, e^{Ft}(e^{-F\tau_1} G_{i_1} e^{F\tau_1} \ldots e^{-F\tau_k} G_{i_k} e^{F\tau_k})\, x_0 \quad \blacksquare \tag{11}$$

d) *Fonctionnelles entrées-sorties des systèmes à temps discret*

Pour un système linéaire analytique à temps discret

$$\left.\begin{aligned} x(t+1) &= x(t) + f_0(x(t)) + \sum_{i=1}^{N} u^i(t)\, f_i(x(t)) \qquad x(0) = x_0 \\ y(t) &= h(x(t)) \end{aligned}\right\} \tag{12}$$

il est clair que les techniques employées précédemment ne s'appliquent plus, puisqu'elles utilisaient les propriétés de différentiabilité par rapport au temps, ce qui n'a plus de sens lorsque le temps est discret. En fait, on peut encore définir la notion de série génératrice pour de « petites entrées », en utilisant ces entrées comme des variables jouant un rôle analogue au temps et en différentiant d'une certaine manière par rapport à ces entrées. Le lecteur peut se référer à [30], [31] pour plus de détails.

V. Réalisation des systèmes non linéaires

a) *A temps continu*

Le problème de la réalisation consiste à se donner une fonctionnelle entrées-sorties, et à déterminer un système, c'est-à-dire une équation différentielle excitée par les entrées et une équation de sortie

$$(NL)\begin{cases} \dot{x} = f(x, u) \\ y = h(x) \end{cases}$$

que redonne la fonctionnelle entrées-sorties.

La question subsidiaire est de trouver la dimension la plus petite possible du vecteur d'état x, et un procédé permettant de construire cette « réalisation minimale ».

Dans le cas linéaire, la fonctionnelle est la matrice de transfert et on sait depuis Kalman [32], que la dimension minimale d'un système linéaire qui réalise cette matrice de transfert, est égale au rang de la matrice de Hankel associée ; on connaît divers algorithmes donnant la réalisation minimale à partir de la matrice de Hankel, par exemple l'algorithme de B. L. Ho [32].

Pour ce qui est du non linéaire, il est clair que le problème général tel qu'il vient d'être posé est beaucoup trop vague pour qu'une réponse puisse y être apportée. D'abord la notion de fonctionnelle entrées-sorties en non linéaire est une notion locale (temps et entrées petits) définie au voisinage de la condition initiale x_0. Ensuite, l'espace d'état est localement une variété différentiable mais des singularités, et donc des sauts de dimension, peuvent se produire. Donc l'énoncé est à prendre, sauf mention du contraire, au sens local, et on parlera donc de réalisation locale (voir § A, II, c)). Jusqu'à présent, seul le cas bilinéaire a reçu une description globale des réalisations locales [33], [34], [35], [36].

Dans le cas général global, des réponses partielles ont été données [37], [38], [39], [40] sous des conditions très fortes de régularité de l'équation différentielle obtenue : celle-ci doit avoir une solution régulière sur un horizon infini pour toute entrée.

Des résultats plus fins sont obtenus localement dans le cas linéaire analytique et la dimension minimale locale de la variété d'état est donnée par le rang de Lie de la série génératrice (voir [41]), qui est au plus égal au rang de la matrice de Hankel associée.

Enfin, la réalisation localement minimale est caractérisée par le fait qu'elle est localement faiblement commandable et localement faiblement observable (voir toujours [41]). Les notions de commandabilité et d'observabilité seront discutées au paragraphe suivant.

Exemple : Le système

$$(\Sigma_1)\begin{cases} \dot{x} = u \\ y = \sin x \end{cases} \qquad x(0) = 0$$

de dimension 1 est évidemment minimal. Sa série génératrice est

$$g = \sum_{k \geq 0} (-1)^k \zeta^{2k+1} = \zeta(1 - \zeta^2)^{-1}$$

Or on vérifie facilement qu'il possède un équivalent bilinéaire de dimension 2

$$(\Sigma_2)\begin{cases} \dot{x}_1 = ux_2 & x_1(0) = 0 \\ \dot{x}_2 = -ux_1 & x_2(0) = 1 \\ y = x_1 \end{cases}$$

(On le vérifie directement en posant $x_1 = \sin x$, $x_2 = \cos x$), donc de même série génératrice g. Or la matrice de Hankel de g est donnée par

$$H = \begin{pmatrix} 1 & 0 & -1 & \dots \\ 0 & -1 & 0 & \\ -1 & 0 & 1 & . \\ & & & \end{pmatrix}$$

de rang 2, ce qui montre que (Σ_2) est minimal dans la classe des bilinéaires sans l'être en général, et d'autre part que le rang de Hankel est supérieur ou égal à la dimension du système minimal. ■

b) *A temps discret*

Le problème de la réalisation minimale à temps discret peut s'énoncer comme précédemment à condition de changer « équation différentielle » par « équation récurrente » à savoir

$$(NLTD) \begin{cases} x(t+1) = f(x(t), u(t)) & x(0) = x_0 \\ y(t) = h(x(t)) & t = 0, 1, \dots, k, . \end{cases}$$

Lorsque f est bilinéaire, la réponse complète est donnée par [42], [29]. Une réponse partielle a été donnée dans [43], pour le cas où l'on cherche f polynomiale, utilisant des concepts de géométrie algébrique. Une réponse plus complète peut être apportée lorsque l'on demande à f d'être *inversible* (voir [44]). Dans ce cas, on peut caractériser les fonctionnelles entrées-sorties qui admettent une telle réalisation, et déterminer la dimension minimale sous des hypothèses de régularité. Le système minimal peut alors être construit par un algorithme donnant les transitions du système $(NLTD)$ analogue à celui de la théorie des automates, et satisfait aux conditions d'*accessibilité* et d'*observabilité* (voir paragraphe suivant).

Notons que cette théorie suffit pour l'étude d'un système obtenu par discrétisation d'un système à temps continu, puisqu'alors la discrétisée exacte est inversible.

VI. Description interne des systèmes non linéaires. Approche géométrique

On a vu au paragraphe précédent que la réalisation minimale d'un système non linéaire est caractérisée par des propriétés d'accessibilité et d'observabilité. Le but premier de la description interne des systèmes est donc de donner des critères facilement vérifiables d'accessibilité/commandabilité, d'observabilité,

etc... Comme on va le voir, cette étude se fait naturellement, depuis Lobry [6], dans le langage de la géométrie différentielle.

Avant de donner les définitions formelles et les principaux résultats, nous allons tenter de motiver, sur un calcul élémentaire, l'emploi de notions telles que « algèbre de Lie ». Une autre présentation élémentaire se trouve dans Brockett [45].

a) *Un calcul élémentaire*

Considérons le système bilinéaire

$$\left.\begin{aligned} \dot{x}(t) &= F_0\, x(t) + \sum_{i=N}^{N} u^i(t)\, F_i\, x(t) \\ x(0) &= x_0 \end{aligned}\right\} \tag{1}$$

On peut, par un changement de variables classiques, retirer le terme en F_0 $z(t) = e^{-F_0 t}\, x(t)$. En dérivant, il vient

$$\left.\begin{aligned} \dot{z}(t) &= \sum_{i=1}^{N} u^i(t)\, (e^{-F_0 t}\, F_i\, e^{F_0 t})\, z(t) \\ z(0) &= x_0 \end{aligned}\right\} \tag{2}$$

Si maintenant on veut avoir une idée des directions dans lesquelles le système peut aller pendant une durée infinitésimale, il suffit d'écrire que

$$z(t) \simeq \left(I + \sum_{i=1}^{N} \int_0^t u^i(s)\, (e^{-F_0 s}\, F_i\, e^{F_0 s})\, ds \right) x_0 \tag{3}$$

et de combiner, dans (3), le développement de Taylor

$$e^{-F_0 s}\, F_i\, e^{F_0 s} = F_i + s[F_0, F_i] + \frac{s^2}{2}\, [F_0, [F_0, F_i]] + \cdots \tag{4}$$

où $[F, G] = GF - FG$ est le « commutateur » ou « crochet de Lie » de F et G (la formule (4) est un cas particulier de la célèbre formule de Baker-Campbell-Hausdorff). Ainsi, en notant

$$\alpha^i = \int_0^t u^i(s)\, ds \qquad \beta^i = \int_0^t s u^i(s)\, ds \qquad \gamma^i = \int_0^t \frac{s^2}{2}\, u^i(s)\, ds \qquad \text{etc...},$$

on obtient

$$z(t) = \left(I + \sum_{i=1}^{N} (\alpha^i\, F_i + \beta^i [F_0, F_i] + \gamma^i [F_0, [F_0, F_i]] + \cdots) \right) x_0 \tag{5}$$

donc si l'ensemble des combinaisons linéaires des matrices F_i, $[F_0, F_i]$, $[F_0, [F_0, F_i]]$, etc..., sont de rang n, il est clair que $z(t)$, et donc $x(t)$, va pouvoir, avec des commandes bien choisies, se trouver n'importe où dans un voisinage de x_0, pour t petit.

L'algèbre de Lie engendrée par $\{ F_0, F_1, ..., F_N \}$ et notée : $\{ F_0, F_1, ..., F_N \}_{LA}$, est par définition la plus petite algèbre contenant toutes les combinaisons linéaires des crochets de Lie de $F_0, F_1, ..., F_N$. Son rang, noté $\mathrm{rg}\{ F_0, F_1, ..., F_N \}_{LA}$, est la borne supérieure des rangs des éléments de cette algèbre. On voit donc, avec (5), que ce rang joue un rôle fondamental dans l'accessibilité et donc dans la commandabilité du système (1). Pour ce qui suit, le lecteur non introduit au langage des « variétés différentiables » pourra, presque sans perte de généralité, considérer qu'il s'agit de R^n ou de surfaces régulières de dimension n.

Enfin précisons que les techniques de géométrie différentielle sont difficilement généralisables aux systèmes à temps discret. C'est pourquoi la commandabilité et l'observabilité des systèmes non linéaires à temps discret contiennent encore de nombreuses questions ouvertes et un exposé sur ces problèmes sort du cadre de ce rapport ([1]).

b) *Accessibilité. Commandabilité*

Étant donné un système linéaire-analytique

$$(LA) \begin{cases} \dot{x}(t) = f_0(x(t)) + \sum_{i=1}^{N} u^i(t)\, f_i(x(t)) \\ x(0) = x_0 \end{cases}$$

où les u^i sont choisis dans l'ensemble U des commandes admissibles (que l'on précisera plus tard), et où l'état x évolue sur une variété analytique X connexe de dimension n, on introduit les définitions suivantes .

DÉFINITION 1 . *Un point x est dit* accessible à partir de x_0 *à l'instant T s'il existe une loi de commande mesurable u de $[0, T]$ dans U, telle que la trajectoire de (LA) engendrée par u et issue de x_0, passe par x à l'instant T*

L'ensemble des états accessibles à l'instant T depuis x_0 est noté $A(x_0, T)$ (voir exemple II, c)) et l'ensemble des états accessibles : $A^+(x_0) = \bigcup_{T \geq 0} A(x_0, T)$.

On dit que (LA) est commandable *si $A^+(x) = X \quad \forall x \in X$* ■

On rajoute le qualificatif *local* lorsque l'on remplace X par un voisinage $V \subset X$, dont tous les points sont joints par des trajectoires restant entièrement dans V. On note que la locale commandabilité implique la commandabilité.

([1]) L'auteur a pris tardivement connaissance des travaux de D. Normand-Cyrot sur ce sujet. Le lecteur pourra consulter avec profit sa thèse : « Théorie et pratique des systèmes non linéaires en temps discret », Université Paris-Sud, mars 83.

DÉFINITION 2 : *L'ensemble des états faiblement accessible à partir de* x_0 *est l'ensemble* $A(x_0) = \bigcup_{T \in R} A(x_0, T)$ *(contenant les états accessibles en temps rétrograde), et on dit que* (LA) *est faiblement commandable si* $A(x) = X$ $\forall x \in X$. ■

Bien entendu un système commandable est aussi faiblement commandable, alors que le contraire n'est généralement pas vrai. Une littérature extrêmement abondante existe sur les caractérisations des états accessibles et faiblement accessibles, et la commandabilité en général. Au lieu d'exposer les subtilités des jeux d'hypothèses sous lesquelles $A^+(x_0)$ ou $A(x_0)$ sont ouverts dans X ou engendrent tout X, nous nous contenterons de renvoyer le lecteur aux deux principaux auteurs Lobry [6] et Sussmann [46], ainsi qu'à la bibliographie de l'article [6]. Nous insisterons par contre sur l'un des résultats les plus complets dû à Bonnard [10]. Il s'agit du cas où les commandes sont constantes par morceaux à valeurs dans $\{ -1 ; 0 ; +1 \}$ (le plus difficile !). Nous aurons besoin des définitions suivantes :

DÉFINITION 3 : *Soit l'équation différentielle définie sur* X

$$\dot{x}(t) = f_0(x(t)) \qquad x(0) = x_0 \tag{6}$$

obtenue à partir de (LA) *en faisant* $u^i \equiv 0 \quad \forall i = 1, ..., N$.

Si l'on note $X(t, x_0)$ *la solution à l'instant* t *de* (6), *on dit que* x_0 *est* Poisson-stable *si pour tout voisinage* V *de* x_0 *et pour tout* $T \geqslant 0$, *il existe* $t \geqslant T$ *tel que* $X(t, x_0) \in V$.

En outre, le système (6) *est dit Poisson-stable si l'ensemble des points Poisson-stables est dense dans* X. ■

DÉFINITION 4 : L'algèbre de Lie *engendrée par les* champs de vecteurs *analytiques*

$$\{ F_0, ..., F_N \} \qquad où \quad F_i(x) = \sum_{j=1}^{N} f_i^j(x) \frac{\partial}{\partial x^j} \qquad i = 0, ..., N$$

les f_i *étant ceux qui apparaissent dans* (LA) *dans un système donné de coordonnées locales, est la plus petite algèbre de champs de vecteurs contenant toutes les combinaisons linéaires à coefficients analytiques des crochets de Lie obtenus à partir de* $F_0, ..., F_N$, *le crochet de Lie de deux champs de vecteurs*

$$F = \sum_{j=1}^{n} f^j \frac{\partial}{\partial x^j} \quad et \quad G = \sum_{j=1}^{n} g^j \frac{\partial}{\partial x^j}$$

étant le champ de vecteurs défini par

$$[F, G] = GF - FG = \sum_{j,k=1}^{n} \left(g^j \frac{\partial f^k}{\partial x^j} - f^j \frac{\partial g^k}{\partial x^j} \right) \frac{\partial}{\partial x^k} \tag{7}$$

Cette algèbre est notée $\{ F_0, F_1, ..., F_N \}_{LA}$.

On définit son rang au point $x \in X$, *noté :* $\operatorname{rg} \{ F_0, F_1, ..., F_N \}_{LA}(x)$, *comme la dimension de l'espace vectoriel engendré par les vecteurs de* $\{ F_0, ..., F_N \}_{LA}$ *évalués au point* x. ■

Cette définition généralise clairement celle qui a été donnée en (B. VI, a)) pour des matrices.

On a alors le résultat [10]

THÉORÈME 1 : *Si* F_0 *est tel que le système* (6) *est Poisson-stable, alors* $\operatorname{rg} \{ F_0, ..., F_n \}_{LA}(x) = n \quad \forall x \in X$, *est une condition nécessaire et suffisante de commandabilité de* (LA) *avec des commandes constantes par morceaux prenant les valeurs*

$$u^i(t) \in \{ -1 ; 0 ; +1 \} \qquad \forall i = 1, ..., N \qquad \forall t > 0 \quad ■$$

Exemple 1 . Reprenons l'exemple de A. I. 1. à une entrée scalaire

$$\left.\begin{aligned} \dot{x}_1 &= a_1 x_2 x_3 + b_1 u \\ \dot{x}_2 &= a_2 x_3 x_1 + b_2 u \\ \dot{x}_3 &= a_3 x_1 x_2 \end{aligned}\right\} \tag{8}$$

on a ici $X = R^3$ $u \in \{ -1 ; 0 ; +1 \}$, $a_3 \neq 0$ et $a_2 b_1^2 \neq a_1 b_2^2$,

$$F_0(x_1, x_2, x_3) = a_1 x_2 x_3 \frac{\partial}{\partial x_1} + a_2 x_3 x_1 \frac{\partial}{\partial x_2} + a_3 x_1 x_2 \frac{\partial}{\partial x_3}$$

$$f_0(x) = \begin{pmatrix} a_1 x_2 x_3 \\ a_2 x_3 x_1 \\ a_3 x_1 x_2 \end{pmatrix}$$

$$F_1(x_1, x_2, x_3) = b_1 \frac{\partial}{\partial x_1} + b_2 \frac{\partial}{\partial x_2}, \quad f_1(x) = \begin{pmatrix} b_1 \\ b_2 \\ 0 \end{pmatrix}$$

On peut montrer que, si l'on fait $u = 0$, le système est Poisson-stable. Montrons que $\operatorname{rg} \{ F_0, F_1 \}_{LA}(x) = 3 \quad \forall x \in R^3$

Pour cela calculons

$$F_2 = [F_0, F_1] = a_1\, b_2\, x_3 \frac{\partial}{\partial x_1} + a_2\, b_1\, x_3 \frac{\partial}{\partial x_2} + a_3(b_1\, x_2 + b_2\, x_1) \frac{\partial}{\partial x_3}$$

$$F_3 = [[F_0, F_1], F_1] = 2\, a_3\, b_1\, b_2 \frac{\partial}{\partial x_3}$$

$$F_4 = [F_2, F_3] = 2\, a_1\, a_3\, b_1\, b_2^2 \frac{\partial}{\partial x_1} + 2\, a_2\, a_3\, b_1^2\, b_2 \frac{\partial}{\partial x_2}$$

En un point $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in R^3$, F_2, F_3 et F_4 ont pour coefficients f_2, f_3 et f_4 respectivement, définis par

$$f_2(x) = \begin{pmatrix} a_1\, b_2\, x_3 \\ a_2\, b_1\, x_3 \\ a_4(b_1\, x_2 + b_2\, x_1) \end{pmatrix} \qquad f_3(x) = \begin{pmatrix} 0 \\ 0 \\ 2\, a_3\, b_1\, b_2 \end{pmatrix}$$

$$f_4(x) = \begin{pmatrix} 2\, a_1\, a_3\, b_1\, b_2^2 \\ 2\, a_2\, a_3\, b_1^2\, b_2 \\ 0 \end{pmatrix}$$

Enfin, on remarque que f_1, f_3 et f_4 engendrent R^3 puisque

$$\operatorname{rg} \begin{pmatrix} b_1 & 0 & 2\, a_1\, a_3\, b_1\, b_2^2 \\ b_2 & 0 & 2\, a_2\, a_3\, b_1^2\, b_2 \\ 0 & 2\, a_3\, b_1\, b_2 & 0 \end{pmatrix} =$$

$$= 3 \quad \text{sauf si} \quad a_3 = 0 \quad \text{ou} \quad a_2\, b_1^2 = a_1\, b_2^2$$

ce qui est exclu par hypothèse.

Donc (8) est commandable sur R^3 ■

Exemple 2 : Le système linéaire $\dot{x} = Fx + Gu$ est commandable sur R^n avec $u \in \{ -1 : 0 : +1 \}$ scalaire si et seulement si:

(i) $\operatorname{Re} \lambda(F) = 0$ (spectre imaginaire pur)

(ii) $\operatorname{rg} \{ G, FG, \ldots, F^{n-1} G \} = n$ (critère de Kalman).

En effet, $\mathrm{Re}\,\lambda(F) = 0$ assure la Poisson-stabilité. D'autre part $\mathrm{rg}\{F, G\}_{LA}$ se calcule facilement en faisant

$$F_0(x) = \sum_{i,j=1}^{n} F_{ij}\, x_j \frac{\partial}{\partial x_i} \quad \text{et} \quad F_1(x) = \sum_{i=1}^{n} G_i \frac{\partial}{\partial x_i}$$

$$[F_0, F_1] = \sum_{i,j=1}^{n} F_{ij}\, G_j \frac{\partial}{\partial x_i} \quad \text{(le coefficient est donc } FG\text{)}$$

$$[F_0, [F_0, F_1]] = \sum_{i,j,k=1}^{n} F_{ij}\, F_{jk}\, G_k \frac{\partial}{\partial x_i} \quad \text{(le coefficient est donc } F^2 G\text{)}$$

etc..., on vérifie enfin que le $n-1$-*ième* crochet $[F_0, [F_0, \ldots, [F_0, F_1]] \ldots]$ a pour coefficient $F^{n-1} G$ donc $\mathrm{rg}\{F, G\}_{LA} = n \quad \forall x \in R^n$ équivaut à (ii).

On peut montrer (Brockett [45]), lorsque u peut prendre des valeurs arbitraires, que la condition (ii) seule (critère de Kalman), équivaut à la commandabilité, ce qui est cohérent par rapport à la théorie linéaire ! ■

c) *Observabilité*

(Pour tout ce paragraphe, le lecteur peut se référer à Isidori [47].)

DÉFINITION 5 : *Un état $x \in X$ est dit* observable *pour le système linéaire-analytique*

$$(LA) \begin{cases} \dot{x}(t) = f_0(x(t)) + \sum_{i=1}^{N} u^i(t) f_i(x(t)) & x(0) = x_0 \\ y = h(x(t)) \end{cases}$$

Si pour tout $x' \in X$, $x' \neq x$, il existe $t \geqslant 0$ et une entrée u admissible tels que $h(X_u(t, x)) \neq h(X_u(t, x'))$, où $X_u(t, x_0)$ est la solution de (LA) pour l'entrée u et la condition initiale x_0. Le système (LA) est dit observable si tout état de X est observable. ■

On note que, dans ce cas, la série génératrice correspondant à l'entrée u et à la condition initiale x ne peut être égale à celle correspondant à u et à x' (voir [29]).

Ici encore, on rajoute le qualificatif *local* lorsqu'on remplace X par un voisinage ouvert.

DÉFINITION 6 : *x et $x' \in X$ sont dits* faiblement distinguables *si pour toute fonction continue σ de $[0, 1]$ dans X avec $\sigma(0) = x$ et $\sigma(1) = x'$, il existe au moins un $s \in [0, 1]$ tel que $\sigma(s)$ et x soient distinguables, c'est-à-dire, avec les notations précédentes, $\exists t \geqslant 0$ et u admissible tels que*

$$h(X_u(t, x)) \neq h(X_u(t, \sigma(s)))$$

x est dit faiblement observable *s'il est faiblement distinguable de tout* $x' \in X$, *et le système* (LA) *est faiblement observable si tout état* $x \in X$ *est faiblement observable.* ■

On a les implications

$$\text{observabilité} \Rightarrow \text{observabilité faible}$$
$$\text{observabilité locale} \Rightarrow \text{observabilité faible}$$

mais il n'y a pas de relation simple entre observabilité et observabilité locale. Le résultat fondamental suivant est dû à Hermann et Krener [39]

THÉORÈME 2 : *Soit* Γ *la famille de fonctions donnée par*

$$\Gamma = \{ F_{j_0} \dots F_{j_k} h \quad \forall k \geqslant 0 \quad \forall j_0, \dots, j_k = 1, \dots, N \}$$

et notons : $\ker (d\Gamma)(x) = \left\{ v \in \mathbb{R}^n \,\middle|\, \frac{\partial \gamma}{\partial x}(x).v = 0 \quad \forall \gamma \in \Gamma \right\}$. *Alors si*

$$\ker (d\Gamma)(x) = \{ 0 \}$$

x est localement observable. Inversement, si (LA) *est localement observable, alors* $\ker (d\Gamma)(x) = \{ 0 \}$ *pour presque tout* $x \in X$. ■

Exemple 1 Pour un système linéaire, $\begin{cases} \dot{x} = Fx + Gu \\ y = Hx \end{cases}$ la condition $\ker (d\Gamma) = \{ 0 \}$ redonne la condition classique :

$$\operatorname{rg} \begin{pmatrix} H \\ HF \\ \\ HF^{n-1} \end{pmatrix} = n$$

En effet, par définition, Γ est engendré par $\{ H, HF, HF^2 \dots \}$. ■

Exemple 2 : Le système

$$\begin{cases} \dot{x}_1 = (1 + u)\, x_1 \\ \dot{x}_2 = u x_2 \\ y = x_1\, x_2 \end{cases}$$

n'est pas localement observable : en effet Γ est engendré par la fonction $\gamma = x_1\, x_2$ seule et donc le système est inobservable lorsque $x_1\, x_2 =$ constante, puisque $\ker (d\Gamma) = \{ (v_1, v_2) \mid x_2\, v_1 + x_1\, v_2 = 0 \} \neq \{ 0 \}$. Pour faire le lien avec le

paragraphe sur la réalisation, il est clair que l'on peut trouver un système équivalent de dimension 1

$$\begin{cases} \dot{x} = (1 + 2u)\, x \\ y = x \end{cases}$$

obtenu par le changement de variable $x = x_1\, x_2$. ■

VII. Rejet de perturbations. Commande non itéractive. Linéarisation par bouclage

C'est, pour l'heure, dans cette partie que les méthodes par descriptions interne et externe réalisent le mariage le plus harmonieux.

a) *Rejet de perturbations par bouclage. Commande non iteractive*

Considérons le système linéaire-analytique

$$\left.\begin{aligned} \dot{x}(t) &= f_0(x(t)) + \sum_{i=1}^{N} u^i(t)\, f_i(x(t)) + \sum_{i=1}^{M} w^i(t)\, g_i(x(t)) \\ y(t) &= h(x, t) \end{aligned}\right\} \qquad (1)$$

où les w^i sont des perturbations non connues (le terme « non mesurable » est souvent employé ici de manière malheureuse : il veut dire ici « que l'on ne peut pas mesurer » et il n'est pas question de théorie des fonctions mesurables !).

Le problème consiste à trouver une loi de bouclage statique :

$$u^i(x) = \alpha^i(x) + \sum_{j=1}^{N} \beta^i_j(x)\, v^j \quad i = 1, \ldots, N \qquad (2)$$

telle que la sortie y soit indépendante des w^i

Les v^j seront alors les nouvelles commandes du système bouclé, et α^i, β^i_j sont supposés analytiques $\forall i, j$.

Les conditions nécessaires et suffisantes de rejet de perturbations par bouclage ont été obtenues par Isidori-Krener-Gori-Giorgi-Monaco [48] par la méthode géométrique (description interne) généralisant les (A, B) invariants de Kalman développés ensuite par Wonham [49] ; et par Claude [50] d'autre part, par des méthodes algébriques (description externe) complétant l'approche par variable d'état et donnant des algorithmes *algébriques* de calcul des lois de bouclage.

Sans entrer dans le détail de ces résultats dont on trouve un exposé complet dans [50], il apparaît que la notion de « nombres caractéristiques » joue un rôle clé dans cette théorie. Il s'agit schématiquement du nombre d'intégrations

qu'il faut réaliser dans le système (1) pour que l'une des entrées apparaisse dans la k-ième sortie. On peut montrer [51] que ces nombres sont reliés au système (1) par les propriétés de son graphe, ce qui permet d'écrire un algorithme plus rapide. A peu de choses près, ces conclusions valent aussi pour le problème de la commande non interactive où l'on fait $w^i \equiv 0 \quad \forall i = 1, ..., M$ dans (1), et où l'on suppose que la dimension de la sortie y est égale à N.

Le problème est alors de trouver des lois de bouclage du type (2) telles que la i-ième entrée n'influence que la i-ième sortie. Remarquons que dans de nombreux cas pratiques le problème linéarisé n'est pas découplable alors que le modèle non linéaire l'est (voir [52]).

b) *Linéarisation par bouclage*

Ici encore, on se donne un système linéaire-analytique

$$\left.\begin{aligned} \dot{x}(t) &= f_0(x(t)) + \sum_{i=1}^{N} u^i(t) f_i(x(t)) \\ y(t) &= h(x) \end{aligned}\right\} \tag{3}$$

et on cherche des lois de bouclage :

$$u^i(x) = \alpha^i(x) + \sum_{j=1}^{N} \beta_j^i v^j, \quad i = 1, ..., N \tag{4}$$

qui rendent le système bouclé linéaire et commandable c'est-à-dire

$$\dot{x} = Ax + Bv \tag{5}$$

Dans (4), on suppose que $\alpha^i(0) = 0$ et $\beta = (\beta_j^i)$ est inversible en 0. La solution de ce problème (proposé par Brockett [53]) a été donnée par Jakubczyk et Respondek [54], Hunt et Su [55], Isidori et Krener [56]. Le fait interessant est que les conditions d'existence d'un tel bouclage, lorsque le nombre d'entrées est au moins égal au nombre de sorties, sont « presque toujours » vérifiées ! Il s'agit là d'un moyen (pour l'instant le seul) d'étudier la *stabilisabilité et la stabilité d'un système non linéaire.* Pour plus de détails, le lecteur peut se référer à [57].

VIII. Stabilité. Stabilisabilité

A part cette dernière idée, la stabilité des systèmes non linéaires en est à ses balbutiements, le problème principal venant de la difficulté de passer du local au global. On trouvera un exposé relativement technique dans Gauthier et Bornard [58], généralisant les techniques de fonctions de Lyapunov.

IX. Identification

Il existe deux techniques d'identification non linéaire la première [59], consiste à calculer les noyaux de la série de Volterra à partir des signaux du système. Cette méthode nécessite une énorme quantité de mesures et s'avère très instable dans la pratique. La seconde consiste à construire, à partir de modèles linéaires obtenus en différents points de fonctionnement, un modèle global non linéaire dont les linéarisées aux points de fonctionnement considérés, redonnent le plus fidèlement possible les modèles linéaires de départ. Cette technique s'avère beaucoup plus fiable. On trouvera des détails et des exemples d'application dans Normand-Cyrot [18].

C. ÉCHELLES DE TEMPS MULTIPLES. AGGRÉGATION. COHÉRENCE

L'idée qui sous-tend cette approche (Peponides-Kokotovic-Chow [23]) est que si, dans un système différentiel, certaines variables sont « faiblement couplées », leur influence se fait sentir sur le long terme et, à court terme, il suffit de résoudre un système de dimension plus petite. On dit alors qu'on *aggrège* le système par échelles de temps. Si les couplages des sous-systèmes obtenus gardent au cours du temps le « même comportement » que le comportement asymptotique, on dit qu'il y a *cohérence*.

Formellement, considérons le système

$$\frac{dx_i}{d\tau} = f_i(x_i, \varepsilon) + \varepsilon g_i(x, \varepsilon) \quad i = 1, \ldots, l, \; x_i \in R^{n_i} \tag{1}$$

où

$$\sum_{i=1}^{l} n_i = n, \; \tau = \frac{t}{\varepsilon}$$

ε est un petit paramètre, et $\varepsilon g_i(x, \varepsilon)$ représente le couplage faible entre x_i et le vecteur d'état complet x.

On suppose alors que

(i) $f_i(x, 0) = 0$ définit une variété A_i de dimension $\nu_i < n_i$ appelée *variété asymptotique*. On suppose qu'elle est définie par $n_i - \nu_i$ équations

$$\varphi_i^k(x_i) = 0 \quad i = 1, \ldots, l, \; k = 0, \ldots, n_i - \nu_i.$$

(ii) l'équation $dx_i\ d\tau = f_i(x_i, 0),\ i = 1, \ldots\ l$ admet une variété intégrale B_i de dimension $n_i - v_i$, donnée par $B_i = \{ x_i \mid \sigma_i(x_i(\tau)) = \sigma_i(x_i(0)) \}$ avec

$$\operatorname{rg} \begin{pmatrix} \dfrac{\partial \varphi_i}{\partial x_i} \\ \dfrac{\partial \sigma_i}{\partial x_i} \end{pmatrix} = n \quad \forall x \quad \text{(cohérence)}$$

Donc aussi, il existe une application différentiable γ telle que

$$x_i = \gamma_i(y_i, z_i) \quad \text{si} \quad y_i = \sigma_i(x_i) \qquad z_i = \varphi_i(x_i) \qquad i = 1, \ldots, l \tag{2}$$

Alors, si l'on fait le changement de variables dans (1)

$$y_i = \sigma_i(x_i) \qquad z_i = \varphi_i(x_i) \qquad i = 1, \ldots, l \tag{3}$$

on obtient, au premier ordre en ε

$$\left.\begin{aligned} \frac{dy_i}{dt} &= \frac{\partial \sigma_i}{\partial x_i}(\gamma_i(y_i, z_i)) \left(\frac{\partial f_i}{\partial \varepsilon}(\gamma_i(y_i, z_i), 0) + g_i(\gamma(y, z), \varepsilon) \right) \triangleq F_i(y, z, \varepsilon) \\ \varepsilon \frac{dz_i}{dt} &= \frac{\partial \varphi_i}{\partial x_i}(\gamma_i(y_i, z_i)) \left(f_i(\gamma_i(y_i, z_i), \varepsilon) + \varepsilon g_i(\gamma(y, z), \varepsilon) \right) \\ &\triangleq G_i(y_i, z_i, \varepsilon) + \varepsilon H_i(y, z, \varepsilon) \end{aligned}\right\} \tag{4}$$

soit

$$\left.\begin{aligned} \frac{dy_i}{dt} &= F_i(y, z, \varepsilon) \\ \varepsilon \frac{dz_i}{dt} &= G_i(y_i, z_i, \varepsilon) + \varepsilon H_i(y, z, \varepsilon) \quad i = 1, \ldots, l \end{aligned}\right\} \tag{5}$$

Les variables y_i s'interprètent donc comme des *modes lents* et les variables z_i comme des *modes rapides*. Heuristiquement, le système peut alors être approximé par un équivalent long terme $\bar{y}_i$ et un équivalent court terme $\bar{z}_i$

$$\left.\begin{aligned} \frac{d\bar{y}_i}{dt} &= F_i(\bar{y}, 0, 0) \quad i = 1, \ldots, l \\ \frac{d\bar{z}_i}{d\tau} &= G_i(y_i, \bar{z}_i, 0) \quad i = 1, \ldots, l \end{aligned}\right\} \tag{6}$$

On a donc *aggrégé* (1) *en* 2 *sous-systèmes lents-rapides.*

Naturellement, les hypothèses faites sont très fortes, mais le point de vue

local suffirait en fait à définir une aggrégation locale. Remarquons qu'en (6), $\bar{z}_i$ est paramétré par y_i.

On trouvera dans Peponides *et al.* [23] et dans Cohen [24] des exemples d'application et une discussion dans différents cas de figures de l'importance des concepts respectifs d'aggrégation et cohérence. Notons enfin que si (1) comportait des entrées, celles-ci pourraient détruire la séparation en variables lentes et rapides par des bouclages qui ne respectent pas cette structure. Pour déterminer les bons bouclages, on peut alors utiliser les techniques de commande non interactive et de rejet de perturbations.

Exemple [24] : On considère le réseau de distribution d'eau formé de 3 nœuds comportant chacun un réservoir et une pompe. Les équations des niveaux d'eau x_i en chaque nœud sont données par :

$$S_i \dot{x}_i = K_i(x_i) - \sum_{j \neq i} \sqrt{\frac{|x_i - x_j|}{R_{ij}}} \cdot \operatorname{sgn}(x_i - x_j) - c_i \quad i = 1, 2, 3 \qquad (7)$$

où S_i est la section du réservoir i, K_i est la caractéristique de la pompe au nœud i, c_i est la consommation en ce nœud, et R_{ij} la résistance de la canalisation de i à j.

Supposons alors que $R_{13} = R_{23} = R$ et que $R_{12} = \varepsilon^2 R$ (ε petit), avec S_1, S_2 et S_3 du même ordre de grandeur.

On vérifie alors que l'on a :

$$f_1(x_1, x_2, \varepsilon) = -\frac{1}{S_1}\sqrt{\frac{|x_1 - x_2|}{R}} \operatorname{sgn}(x_1 - x_2), f_2(x_1, x_2, \varepsilon) = \frac{S_1}{S_2} f_1(x_1, x_2, \varepsilon)$$

$f_3(x_3, \varepsilon) = 0 \quad \forall \varepsilon$,

et

$$g_i(x, \varepsilon) = \frac{1}{S_i}\left(K_i(x_i) - c_i - \sqrt{\frac{|x_i - x_3|}{R}} \operatorname{sgn}(x_i - x_3)\right) \quad i = 1, 2,$$

$$g_3(x, \varepsilon) = \frac{1}{S_3}\left(K_3(x_3) - c_3 + \sqrt{\frac{|x_1 - x_3|}{R}} \operatorname{sgn}(x_1 - x_3) + \right.$$

$$\left. + \sqrt{\frac{|x_2 - x_3|}{R}} \operatorname{sgn}(x_2 - x_3)\right), \quad \forall \varepsilon$$

La variété asymptotique est donc $x_1 - x_2 = 0$ et :

$$\begin{cases} \sigma_1(x_1, x_2) = \dfrac{S_1 x_1 + S_2 x_2}{S_1 + S_2} \\ \sigma_2(x_3) = x_3 \end{cases}$$

L'hypothèse de rang est trivialement vérifiée et on déduit le système asymptotique en posant : $x = x_1 = x_2$

$$\left.\begin{aligned}(S_1 + S_2)\frac{dx}{dt} &= (K_1(x)+K_2(x))-(C_1+C_2)-2\sqrt{\frac{|x-x_3|}{R}}\,\operatorname{sgn}(x-x_3)\\ S_3\frac{dx_3}{dt} &= K_3(x_3)-c_3+2\sqrt{\frac{|x-x_3|}{R}}\,\operatorname{sgn}(x-x_3)\end{aligned}\right\} \quad (8)$$

qui est de dimension 2.

L'interprétation de (8) est évidente : on aggrège les nœuds 1 et 2 en les remplaçant par un seul nœud où la pompe est la somme des pompes, où la consommation est la somme des consommations et où la résistance de la canalisation au nœud 3 est $R/4$. ■

CONCLUSION

En guise de conclusion, remarquons que l'achèvement de la théorie des systèmes linéaires a été réalisé en grande partie grâce au formalisme de la commande optimale linéaire-quadratique. En effet, c'est dans ce formalisme simple qu'ont pu être réglés les problèmes de stabilité des systèmes bouclés, du placement de pôles pour le contrôleur-observateur, etc...

Or la commande optimale non linéaire est encore loin de pouvoir donner des réponses aussi précises. Le problème de fond est que l'optimalité, comme les propriétés en général des systèmes non linéaires, ne se caractérise facilement que *localement* alors que c'est du *comportement global* dont on a besoin pour l'étude de la stabilité. On voit donc l'importance de développer les recherches en commande optimale. Nous allons passer en revue très brièvement les développements récents en ce domaine, à la lumière de la théorie non linéaire.

Tout d'abord, la théorie non linéaire déterministe a permis d'obtenir des résultats très fins sur les extrêmales singulières et le principe du minimum d'ordre élevé [10]. En commande stochastique, ce sont les développements récents sur le filtrage non linéaire de dimension finie qui permettent d'espérer des résultats plus calculatoires. D'autres travaux cherchent à dégager des classes de problèmes plus vastes que le linéaire-quadratique, mais suffisamment simples pour avoir des notions globales sur la solution : ainsi [60] préconisent les problèmes linéaires à coût exponentiel ; dans [61], on remarque que la solution complète est donnée lorsque l'on peut avoir une équation différentielle donnant directement la commande et on donne des classes de problèmes où l'élimination de l'état et de l'état adjoint sont possibles. Cependant, d'autres travaux tendent à montrer qu'il faut s'éloigner du linéaire-quadratique puisqu'en changeant « légèrement » les hypothèses, peuvent

apparaître des non-linéarités uniquement dues à l'optimisation le meilleur exemple dû à Witsenhausen [62] est linéaire-quadratique-gaussien, la seule différence avec le cadre classique étant le manque de mémoire. C'est ce manque de mémoire qui oblige à faire un compromis entre une commande peu coûteuse et une commande qui révèle suffisamment les actions passées pour compenser l'oubli, d'où la non-linéarité de la commande optimale par rapport à l'observation (alors qu'avec une mémoire parfaite, la commande est affine en l'observation). On trouvera la solution théorique générale des problèmes à « information non classique » dans [63], [64].

Cependant, tous ces développements mènent généralement à des calculs qui sont encore très loin des prérequis du temps réel. Ainsi l'approche par commande adaptative cherche essentiellement à éviter des calculs compliqués pour obtenir une grande précision, et à remplacer les objectifs d'optimalité par une sous-optimalité qui respecte un comportement entrées-sorties donné. Ces méthodes prolongent la théorie linéaire et peuvent être appliquées à des systèmes non linéaires qui varient « assez » lentement (voir [65]).

Citons pour terminer les méthodes qui utilisent des développements par rapport à un petit paramètre pour diminuer la complexité des calculs, dans la lignée du paragraphe C [66], et les techniques de décomposition-coordination [67].

Il est bien entendu hors de propos de donner un aperçu complet des développements actuels de la commande optimale, et nous arrêterons ici en insistant de nouveau sur l'importance de progresser aussi dans cette théorie pour pouvoir enfin parachever la stabilité non linéaire, etc...

BIBLIOGRAPHIE

Les références sont, pour des raisons de commodité, classées selon leur ordre de première apparition dans le texte et non par ordre alphabétique.

Certains ouvrages contenant des séries de publications reviennent souvent. Pour éviter d'avoir à chaque fois à réécrire les références complètes, nous donnerons les abréviations suivantes pour ces ouvrages

- *CNRS* 81 : Outils et Modèles Mathématiques pour l'Automatique, l'Analyse des Systèmes et le Traitement du Signal. I. Landau coordonnateur. Éditions du CNRS, 1981.

- *Belle-Ile* 82 Développement et Utilisation d'outils et Modèles Mathématiques en Automatique, Analyse des Systèmes et Traitement du Signal. Colloque CNRS. Belle-Ile, septembre 1982.

- *INRIA* 82 : Analysis and Optimization of Systems. A. Bensoussan, J. L. Lions Éditeurs, Lecture Notes in Control and Information, Sciences n° 44, Springer Verlag, Berlin, New York, 1982.

[1] J. LEVINE, G. PIGNIE, Rapport DRET. *Bibliographie commentée sur le filtrage non linéaire*, février 1983.

[2] V. ARNOLD. *Équations Différentielles Ordinaires* (Traduction française), Ed. MIR, 1974.

[3] E. LEE, M. MARKUS, *Foundations of Optimal Control Theory*, John Wiley, 1967

[4] E. IRVING, *Identification des systèmes*, EDF, Études et Recherches, C1-2, 1969.

[5] A. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[6] C. LOBRY, *Contrôlabilité des systèmes non linéaires*, CNRS, 1981.

[7] J. M. BISMUT, *Martingales, the Malliavin calculus and hypoellipticity under general Hormander's conditions*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 56, 1981, pp. 469-505.

[8] J. BAILLEUL, R. BROCKETT, R. WASHBURN, *Chaotic motion in non linear feedback systems*, IEEE Trans. CAS-27, nº 11, novembre 1980, pp. 990-997.

[9] R. BROCKETT, *On the reachable set for bilinear systems*, in R. Mohler, A. Ruberti Ed. Proceedings 1974 conference on bilinear systems, Springer, 1975.

[10] B. BONNARD, *Contrôle de l'attitude d'un satellite rigide*, Belle-Ile, 1982.

[11] S. SINGH, A. SCHY, *Output feedback non linear decoupled control synthesis and observer design for maneuvering aircraft*, Int. J. Control., 31, 1980, pp. 781-806.

[12] O. MERCIER, *Lois de commande multivariables non linéaires pour le pilotage en grande amplitude des avions*, Rapport ONERA, RT 5/7224 SY, décembre 1981.

[13] G. MEYER, R. SU, L. HUNT, *Applications to aeronautics of the theory of transformations of non linear systems*, Belle-Ile, 1982.

[14] G. CESAREO, F. NICOLO, S. NICOSIA, *Dynamical models of industrial robots*, 1st IASTED Int. Symposium on applied modelling and simulation, Lyon, septembre 1981.

[15] R. MOHLER, G. HSU, V. KARANAM, *Modelling and control of σ-B cell immune processes* (to appear).

[16] M. ESPANA, I. LANDAU, *Reduced order bilinear models for distillation columns*, Automatica, 1978.

[17] G. BORNARD, J. P. GAUTHIER, *Modélisation dynamique des colonnes de distillation*, CNRS, 1981.

[18] D. NORMAND-CYROT, *Identification par systèmes à état affine et applications aux centrales électriques*, CRNS, 1981.

[20] J. WOOD, *Power conversion in electrical networks*, PHD Dissertation, Harvard University, June 1974.

[21] J. BAILLEUL, C. BYRNES, *A geometric problem in electric energy systems*. Int. Symp. on Mathematical Theory of Networks and systems, Vol. 4, N. Hollywood CA, Western Periodicals Co 1981.

[22] C. TAVORA, O. SMITH, *Stability analysis of power systems*, IEEE Trans. on Power Apparatus and Systems, Vol. PAS-91, 1972, pp. 1093-1100.

[23] G. PEPONIDES, P. KOKOTOVIC, J. CHOW, *Singular perturbations and time scales in non linear models of power systems*, IEEE Trans. Circuits and Systems, Vol. CAS-29, Novembre 1983 (to appear).

[24] G. COHEN, *Commande optimale de grands systèmes. Quelques réflexions autour d'une application*, INRIA, 1982.

[25] J. BAILLEUL, *The geometry of homogeneous polynomial dynamical systems*, Non linear Analysis Theory, Methods and Applications, Pergamon Press, Vol. 4, nº 5, 1979, pp. 879-900.

[26] M. FLIESS, D. NORMAND-CYROT, *La propriété d'approximation des systèmes réguliers (ou bilinéaires)*, CNRS, 1981.

[27] H. SUSSMANN, *Semigroup representations, bilinear approximation of input, output maps and generalized inputs*, Mathematical systems Theory. G. Marchesini, S. Mitter Ed. Lecture Notes in Econ. Math. Syst., 131, Springer 1976, pp. 172, 191.

[28] M. FLIESS, *Developpements fonctionnels et calcul symbolique non commutatif*, CNRS, 1981.

[29] G. JACOB, *Réalisation des systèmes réguliers et séries formelles non commutatives*, CNRS, 1981.

[30] D. NORMAND-CYROT, *Une condition de réalisation par systèmes à état affine discrets*, INRIA, 1982.

[31] S. MONACO, D. NORMAND-CYROT, *Sur la subordination d'un système non linéaire discret à un système linéaire*, Belle-Ile, 1982.

[32] R. KALMAN, P. FALB, M. ARBIB, *Topics in Mathematical Systems Theory*, McGraw-Hill, NY, 1969.

[33] R. BROCKETT, *On the algebraic structure of bilinear systems. Theory and Applications of Variable Structure Systems*, R. Mohler, A. Ruberti Ed., Academic Press, 1972, pp. 153-168.

[34] P. D'ALESSANDRO, A. ISIDORI, A. RUBERTI, *Realization and Structure theory of bilinear systems*, SIAM J. Control, 12, 1974, pp. 517-535.

[35] M. FLIESS, *Sur la réalisation des systèmes dynamiques bilineaires*, CRAS. Série A, 277, 1973, pp. 923-926.

[36] H. SUSSMANN, *Minimal realizations and canonical forms for bilinear systems*, J. Franklin Institute, 301, 1976, pp. 593-604.

[37] H. SUSSMANN, *A generalization of the closed subgroup theorem to quotient of arbitrary manifolds*, J. Diff. Geom., 10, 1975, pp. 151-166.

[38] H. SUSSMANN, *Existence and uniqueness of minimal realizations of non linear systems*, Math. Systems Theory, 10, 1977, pp. 263-284.

[39] R. HERMANN, A. KRENER, *Non linear controllability and observability*, IEEE AC-22, 1977, pp. 728-740.

[40] B. JAKUBCZYK, *Existence and uniqueness of realization of non linear systems*, SIAM J. Control and Optimiz., 18, 1980, pp. 455-471.

[41] M. FLIESS, *Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices non commutatives*, Inventiones Math. (à paraître).

[42] M. FLIESS, *Un outil algébrique les séries formelles non commutatives*, Rapport IRIA, n° 139, octobre 1975.

[43] E. SONTAG, Y. ROUCHALEAU, *On discrete-time polynomial systems*, J. Non linear Analysis, Methods, Theory and Applications, 1, 1976, pp. 55-59.

[44] B. JAKUBCZYK, *Invertible realizations of non linear discrete time systems*, Proceedings of the 1980 Princeton Conference on Information Sciences and Systems.

[45] R. BROCKETT, *Non linear systems and differential geometry*, Proceedings of IEEE, Vol. 64, n° 1, 1976, pp. 61-71.

[46] H. SUSSMANN, V. JURDJEVIC, *Controllability of non linear systems*, J. of Diff. Equations, vol. 12, 1972, pp. 95-116.

[47] A. ISIDORI, *Observabilité et observateurs des systèmes non linéaires*, CNRS, 1981.

[48] A. ISIDORI, A. KRENER, C. GORI-GIORGI, S. MONACO, *Non linear decoupling via feedback : a differential geometric approach*, IEEE AC-26, 1981, pp. 331-345.

[49] W. WONHAM, *Linear multivariable control : a geometric approach*, 2e Ed. Springer, 1977.

[50] D CLAUDE, *Découplage des systèmes du linéaire au non linéaire*, Belle-Ile, 1982.

[51] A. KASINSKY, J. LEVINE, *A fast graph-theoretic algorithm for the feedback decoupling problem of non linear systems*, Proc. of 8th MTNS, Beersheva, juin 1983

[52] J. P. GAUTHIER, G. BORNARD, S. BACHA, M. IDIR, *Rejet de perturbations pour un modèle non linéaire de colonne à distiller*, Belle-Ile, 1982.

[53] R. BROCKETT, *Feedback invariants for non linear systems*, VII IFAC Congress, Helsinki, 1978.

[54] B. JAKUBCZYK, W. RESPONDEK, *On linearization of control systems*, Bull. Acad. Polonaise. Sci. Serie Sci. Math., 28, 1980, pp. 517-522.

[55] L. HUNT, R. SU, *Multi-input non linear systems* (à paraître).

[56] A. ISIDORI, A. KRENER, *On feedback equivalence of non linear systems*, Syst. Control Letters, 2, 1982, pp. 118-121.

[57] A. ISIDORI, *The geometric approach to nonlinear feedback control . a survey*, INRIA, 1982.

[58] J. P. GAUTHIER, G. BORNARD, *Stabilisation des systèmes non linéaires*, CNRS, 1981.

[59] P. CROUCH, *Polynomic system theory : a review*, IEE Proc., 127, 1980, p. 220-228.

[60] J. KRAINAK, F. MACHELL, S. MARCUS, J. SPEYER, *The dynamic linear exponential gaussian team problem*, IEEE AC (to appear).

[61] E. DOCKNER, G. FEICHTINGER, S. JORGENSEN, *Tractable classes of nonzero-sum open-loop Nash differential games* (à paraître).

[62] H. WITSENHAUSEN, *A counterexample in stochastic optimum control*, SIAM J Control, Vol. 6, n° 1, 1968, pp. 131-147.

[63] J. LEVINE, *Principe d'optimalité et principe de séparation en contrôle stochastique à information incomplète non classique*, CRAS Série I, 292, 1981, pp. 877-880

[64] J. LEVINE, *Incomplete information in differential games and team problems*, 8th IFAC World congress, Kyoto, 1981.

[65] K. ASTRÖM, *Theory and applications of adaptive control*, 8th IFAC. World congress, Kyoto, 1981.

[66] J. CRUZ, *Feedback systems*, McGraw-Hill, 1972.

[67] G. COHEN, *Optimization by decomposition and coordination : a unified approach*, IEEE, AC-23, 1979, pp. 222-232.

COMMENTAIRE DU RAPPORTEUR

L'article de J. Levine a le grand mérite de faire précéder l'exposé des résultats « théoriques », d'exemples simples et convaincants qui montrent pourquoi *il ne suffit pas de linéariser* pour aborder l'analyse et la régulation de certains systèmes linéaires.

Ceci indique quelles sont les limites des approches, extrêmement efficaces dans d'autres cas, du type « commande adaptative ».

Il me semble maintenant évident qu'une bonne maîtrise des deux domaines est devenue nécessaire pour aborder les problèmes très non linéaires. Il est non moins évident que des progrès importants restent à faire pour simplifier et approfondir notre compréhension de ces différents modèles non linéaires.