



**HAL**  
open science

# Estimation avancée à horizon glissant, état et paramètres, pour la conduite du coeur des réacteurs PWR

Lucas Gruss

► **To cite this version:**

Lucas Gruss. Estimation avancée à horizon glissant, état et paramètres, pour la conduite du coeur des réacteurs PWR. Automatique / Robotique. Ecole nationale supérieure Mines-Télécom Atlantique, 2024. Français. NNT : 2024IMTA0434 . tel-04960678

**HAL Id: tel-04960678**

**<https://theses.hal.science/tel-04960678v1>**

Submitted on 21 Feb 2025

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

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

# THÈSE DE DOCTORAT DE

L'ÉCOLE NATIONALE SUPÉRIEURE  
MINES-TÉLÉCOM ATLANTIQUE BRETAGNE  
PAYS DE LA LOIRE -- IMT ATLANTIQUE

ÉCOLE DOCTORALE N° 648  
*Sciences pour l'Ingénieur et le Numérique*  
Spécialité : *Génie industriel, productive, automatique et robotique*

Par

**Lucas GRUSS**

**Estimation avancée à horizon glissant, état et paramètres, pour la  
conduite du cœur des réacteurs PWR**

Thèse présentée et soutenue à IMT Atlantique - campus de Nantes, le 12 décembre 2024  
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N - UMR 6004)  
Thèse N° : 2024IMTA0434

## Rapporteurs avant soutenance :

Christina STOICA Professeure, CentraleSupélec  
Thierry POINOT Professeur ENSIP, - Université de Poitiers

## Composition du Jury :

Président :	Thierry POINOT	Professeur, ENSIP - Université de Poitiers
Examineurs :	Christina STOICA	Professeure, CentraleSupélec
	Mazen ALAMIR	Directeur de Recherche, CNRS - Université de Grenoble
Dir. de thèse :	Philippe CHEVREL	Professeur, IMT Atlantique
Encadrant :	Mohamed YAGOUBI	Maître assistant, IMT Atlantique
	Maxime THIEFFRY	Maître assistant, IMT Atlantique

## Invité :

Alain GROSSETÊTE Senior expert, Framatome



## REMERCIEMENTS

*Par une belle matinée du mois de mai, une élégante  
amazone parcourait, sur une superbe jument alezane, les  
allées fleuries du bois de Boulogne.*

— Joseph Grand dans *La Peste*  
Albert CAMUS

Quelle joie d'écrire ces quelques lignes ! Si elles marquent pour vous le début de ce manuscrit, elles symbolisent pour moi l'aboutissement de ces trois années de thèse qui m'ont tant animé. Ce travail n'aurait jamais vu le jour sans le soutien précieux de ceux qui m'ont entouré.

Je souhaite tout d'abord exprimer ma profonde gratitude à mes encadrants : Philippe Chevrel, qui m'a fait l'honneur de diriger cette thèse, ainsi que Mohamed Yagoubi et Maxime Thieffry. Philippe, c'est grâce à toi que cette thèse a pu voir le jour malgré les défis posés par la crise sanitaire de 2020. Merci pour nos nombreuses discussions et ces moments d'échange d'où je repartais toujours avec plus d'idées et de pistes de recherches que je ne pourrais explorer en une carrière. Ta bienveillance et tes encouragements m'ont porté tout au long de ce parcours. Mohamed, je te remercie pour ta rigueur scientifique et tes conseils avisés qui, je l'espère, ont contribué à rendre ce manuscrit plus clair. Maxime, merci pour ton regard calme et réfléchi sur les problématiques abordées dans mes travaux, ainsi que pour ton soutien constant. Je te souhaite le meilleur dans les nouveaux horizons que tu explores.

Je tiens également à remercier Isabelle Rivals, à qui je rends hommage, ainsi que Mazen Alamir pour leurs précieux conseils et leur écoute bienveillante au sein de mon CSI. Merci à Thierry Poinot et Cristina Stoica d'avoir accepté le rôle de rapporteurs pour mes travaux.

Mes séjours au laboratoire ont toujours été des moments agréables grâce à Fabien, Pierre, Rafael, ainsi qu'à mes codocotorants : Alexandre, François, Yuqi, Karim et Andrea. Je vous souhaite à tous le succès que vous méritez, et j'ai hâte de pouvoir vous appeler « Docteur ».

En ce qui concerne Framatome, je souhaite exprimer ma reconnaissance à Alain Grosse-tête pour la confiance qu'il m'a accordée et son expertise précieuse. Nos nombreux échanges m'ont permis d'élargir mes perspectives et de mieux comprendre les enjeux industriels. Je sais que l'OAPS est entre de bonnes mains grâce à la fine équipe qui le développe ou qui y a contribué : Guillaume Dupré mon *nakama*, qui a pavé la route de mon intérêt pour les méthodes numériques ; Perceval Beja-Battais, qui a fait un retour en force dans l'équipe ; Django le Clerre-Maraine, avec qui j'ai partagé de passionnantes discussions sur la programmation, Matlab et Modelica ; et Adam Dereszewski, dont l'aide précieuse sur le simulateur SOFIA a été déterminante. Merci à la section DTIPCE de m'avoir accueilli comme membre officieux, pour les pauses banane et les parties de Skull King : Barbara, Martin, Samy, Ozan, Théophile, Manon, Sarah, Xannath, les deux Pierre, Stéphane, Arnaud, Maxime et les deux Vincent. Un grand merci à Fady Nakache et Pierre Vanpeene pour leur aide indispensable et leur expertise sur SOFIA.

J'ai eu la chance de garder un lien fort, à Paris ou ailleurs, avec mes amis rencontrés au fil de mon parcours. Merci à mes camarades de classe préparatoire : Lucas, Louis, Marion, Victor, Thibault, Paul, Camille C., Coralie et Samuel. Une pensée également pour mes amis des Mines de Nantes et associés : Appoline, Baptiste, Jean, Ilian, Antoine, Camille B., Alexandre, Estelle, Éléonore, Lucile, Alvin, les Marly-zouz, Henri et Marvin, qui ont illuminé des moments inoubliables, du Monaco à la salle d'escalade en passant par Halloween.

Un remerciement tout particulier à mon colocataire Pierre pour sa bonne humeur et

pour avoir partagé notre « unique neurone » avec moi. Impossible de recenser tous nos fous rires, nos « blagues privatives », nos visionnages de Shrek ou d'autres chefs-d'œuvre animé ou non, la musique *ad-nauseam*, et bien sûr, tout ce qui est issu de la Catalogne. Je suis également heureux d'avoir rencontré Giomé et Teuz, les colocs officieux, avec qui j'ai passé d'excellents moments : préparez vous pour ma revanche à *Smash* !

Je mesure pleinement la chance d'avoir une famille aussi incroyable à mes côtés. Merci à ma sœur Anne-Sophie, mon meilleur public, qui me donne l'illusion d'être drôle (donc me fait bider ailleurs), ainsi qu'à sa famille : Moïse, Aaron, Noé, Naomi (et Lyxie !) pour les mardis soir, les lentilles-saucisses et les fous rires. Un immense merci à mes parents, Marie-Christine et Ronald, pour leur soutien indéfectible, même lorsque le stress a pu me rendre « moins aimable » que d'ordinaire. Je sais que je peux toujours compter sur vous, et j'en suis infiniment reconnaissant.

Je ne peux m'empêcher d'avoir une pensée pour ma petite mamie, qui a rejoint papy : je pense à toi tous les jours.

Je remercie également Dominique, Agathe, Fanny, Nathan et Aymeric pour leur accueil chaleureux. Enfin, à ma compagne Emeline : ta gentillesse, ta patience tout au long de ces années de thèse, ton soutien indéfectible, nos discussions et l'exemple que tu me donnes au quotidien sont autant de raisons de partager ma vie avec toi.

# TABLE DES MATIÈRES

<b>Avant-propos</b>	<b>ii</b>
Remerciements . . . . .	iii
Table des matières . . . . .	iv
Liste des figures . . . . .	ix
Liste des tableaux . . . . .	x
Liste des productions scientifiques . . . . .	xi
<b>Introduction</b>	<b>1</b>
Contexte . . . . .	1
Flexibilité des centrales nucléaires . . . . .	2
Contributions . . . . .	3
Plan du manuscrit . . . . .	3
<b>I Estimation à horizon glissant</b>	<b>5</b>
<b>1 État de l'art en estimation</b>	<b>7</b>
1.1 Observabilité . . . . .	8
1.1.1 Observabilité des systèmes linéaires . . . . .	8
1.1.2 Observabilité des systèmes non-linéaires . . . . .	9
1.2 Observateur d'état . . . . .	10
1.2.1 Observateur de Luenberger . . . . .	10
1.2.2 Filtre de Kalman . . . . .	11
1.2.3 Filtre de Kalman étendu . . . . .	12
1.2.4 Filtre de Kalman à « points Sigma » . . . . .	14
1.3 Estimation de paramètres . . . . .	15
1.4 Estimation à horizon glissant . . . . .	17
1.4.1 Estimation à information totale . . . . .	18
1.4.2 Introduction d'un horizon de taille fixe . . . . .	19
1.4.3 Interprétation stochastique . . . . .	21
1.4.4 Retour sur le coût d'arrivée . . . . .	21
1.4.5 Propriétés . . . . .	22
1.5 Notes et discussion . . . . .	23
1.5.1 Résolution des problèmes de contrôle optimal à temps continu . . . . .	23
1.5.2 Stratégies de réduction de la latence de calcul . . . . .	25
<b>2 Simulation numérique</b>	<b>27</b>
2.1 Modèles dynamiques . . . . .	27
2.2 Intégration numérique . . . . .	30
2.3 Variation du pas de temps, contrôle de l'erreur . . . . .	33
2.4 Sensibilité de l'intégrateur . . . . .	34
2.5 Méthode d'Euler . . . . .	35
2.6 Méthodes de Runge-Kutta . . . . .	37
2.7 Méthodes de collocation . . . . .	39
2.8 Notes et discussion . . . . .	41

<b>3</b>	<b>Méthodes de transcription</b>	<b>43</b>
3.1	Tir simple . . . . .	46
3.2	Tir multiple . . . . .	48
3.3	Transcription simultanée par collocation directe . . . . .	51
3.4	Exploitation de la structure du problème . . . . .	54
3.4.1	Remarque sur la vitesse de convergence et la complexité des itérations	56
3.4.2	Condensing . . . . .	56
3.4.3	Récursion de Riccati . . . . .	57
3.5	Notes et discussion . . . . .	57
<b>4</b>	<b>Programmation non-linéaire</b>	<b>59</b>
4.1	Conditions de KKT . . . . .	60
4.2	Méthode de Newton . . . . .	61
4.3	Recherche linéaire et région de confiance . . . . .	62
4.4	Stratégie de globalisation et mécanismes de globalisation . . . . .	63
4.4.1	Fonction mérite . . . . .	63
4.4.2	Filtres . . . . .	64
4.5	Méthode des points intérieurs . . . . .	65
4.6	Programmation quadratique séquentielle . . . . .	67
4.7	Calcul de quantités dérivées . . . . .	68
4.7.1	Différences finies . . . . .	68
4.7.2	Différentiation automatique . . . . .	69
4.8	Notes et discussion . . . . .	70
<b>II</b>	<b>Estimation pour la conduite des REP</b>	<b>73</b>
<b>5</b>	<b>Physique et modélisation des REP pour l'estimation et la commande</b>	<b>75</b>
5.1	Historique et contexte . . . . .	75
5.2	Fonctionnement d'une centrale nucléaire et pilotage d'un réacteur . . . . .	76
5.2.1	Cuve réacteur et cœur . . . . .	79
5.2.2	Générateurs de vapeurs . . . . .	79
5.3	Physique des réacteurs . . . . .	80
5.3.1	Réactions nucléaires . . . . .	80
5.3.2	Équations de la neutronique . . . . .	83
5.3.3	Modération des neutrons et effet modérateur . . . . .	84
5.3.4	Effet Doppler . . . . .	86
5.3.5	Empoisonnement par le xénon . . . . .	87
5.3.6	Action de l'acide borique . . . . .	88
5.3.7	Action des barres de contrôle . . . . .	90
<b>6</b>	<b>Modélisation pour la commande et l'estimation</b>	<b>91</b>
6.1	Modélisation des cœurs de réacteurs nucléaires . . . . .	91
6.2	SOFIA . . . . .	91
6.3	Modèle d'estimation et de commande . . . . .	92
6.3.1	Équations de la cinétique et couplage spatial . . . . .	94
6.3.2	Dynamique xénon/iode . . . . .	97
6.3.3	Modèle de température . . . . .	98
6.3.4	Synthèse des équations du modèle . . . . .	100
6.3.5	Propriétés . . . . .	101

<b>7</b>	<b>Application aux réacteurs nucléaires à eau pressurisée</b>	<b>103</b>
7.1	Conception d'un estimateur pour l'aide au pilotage . . . . .	104
7.1.1	Operator Aid Predictive System (OAPS) . . . . .	104
7.1.2	Estimation ad-hoc . . . . .	105
7.1.3	Cahier des charges de la solution d'estimation . . . . .	106
7.1.4	Choix de la méthode d'estimation . . . . .	106
7.2	Implémentation de l'estimateur . . . . .	107
7.2.1	Longueur et discrétisation de l'horizon d'estimation . . . . .	107
7.2.2	Méthode d'intégration et transcription . . . . .	108
7.2.3	Solveur d'optimisation . . . . .	109
7.2.4	Stratégie de réduction de la latence de calcul . . . . .	110
7.3	Application avec un modèle parfait de réacteur . . . . .	111
7.3.1	Estimation de l'état . . . . .	111
7.3.2	Estimation conjointe de l'état et des paramètres . . . . .	115
7.4	Validation sur les données de SOFIA . . . . .	119
7.4.1	Préliminaire : calibration du modèle de conception avec le modèle expert . . . . .	121
7.4.2	Estimation d'état . . . . .	123
7.5	Conclusion et synthèse des résultats . . . . .	124
7.6	Notes et discussion . . . . .	124
	<b>Conclusion Générale</b>	<b>127</b>
	Synthèse des travaux réalisés . . . . .	127
	Perspectives . . . . .	128
	<b>Annexes</b>	<b>129</b>
<b>A</b>	<b>Filtre de Kalman</b>	<b>131</b>
A.1	Résultat préliminaire . . . . .	131
A.2	Estimateur linéaire non biaisé . . . . .	131
A.3	Preuve du filtre de Kalman . . . . .	132
<b>B</b>	<b>Transitoire <math>PWR_{simu}</math> pour la validation sans écart modèle/process</b>	<b>135</b>
<b>C</b>	<b>Transitoire de validation SOFIA</b>	<b>139</b>
	<b>Bibliographie</b>	<b>143</b>



## TABLE DES FIGURES

1	Prix de l'électricité en Europe le 24 mai 2020. . . . .	1
1.1	Estimation d'état via un observateur. . . . .	7
1.2	Fonctionnement du filtre de Kalman. . . . .	10
1.3	Phase de prédiction dans un filtre à points sigma. . . . .	15
1.4	Filtre à information totale et estimation par horizon glissant. . . . .	18
1.5	Résolution d'un problème optimal à temps continu par une méthode directe. . . . .	25
2.1	Maillage de l'espace pour la discrétisation spatiales des EDP . . . . .	28
2.2	Discrétisation du temps . . . . .	30
2.3	Domaine de stabilité d'une méthode A-stable. . . . .	32
2.4	Méthodes d'Euler . . . . .	36
2.5	Méthode des trapèzes . . . . .	37
2.6	Méthode du point milieu . . . . .	37
2.7	Méthode de Runge-Kutta 4. . . . .	39
3.1	Différentes méthodes de transcription aboutissent à des problèmes NLP différents mais dont la structure peut être exploitée. . . . .	43
3.2	Tir simple. . . . .	46
3.3	Motif du caractère creux en tir simple . . . . .	47
3.4	Évolution de la solution pour une transcription en tir multiple. . . . .	50
3.5	Motif du caractère creux en tir multiple . . . . .	50
3.6	Comparaison de la résolution du problème pour une méthode de tir et pour la collocation directe. . . . .	51
3.7	Transcription simultanée par collocation directe. . . . .	52
3.8	Motif du caractère creux en collocation directe . . . . .	53
4.1	Filtre et critère de validation du pas . . . . .	65
5.1	Représentation simplifiée d'une centrale nucléaire. . . . .	77
5.2	Cuve du réacteur. . . . .	78
5.3	Générateur de vapeur. . . . .	78
5.4	Réactions nucléaires. . . . .	82
5.5	Domaines d'énergie des neutrons. . . . .	85
5.6	Section efficace microscopique totale de $^{235}\text{U}$ et $^{238}\text{U}$ à 0K. (ENDF/B-VIII.0) . . . . .	86
5.7	Section efficace microscopique totale de $^{135}\text{Xe}$ à 0K. (JEFF-3.3) . . . . .	87
5.8	Chaîne de désintégration du $^{135}\text{Xe}$ . . . . .	87
5.9	Section efficace microscopique totale de $^{10}\text{B}$ à 0K. (JEFF-3.3) . . . . .	89
6.1	Différences de maillage entre les modèles neutronique et thermique de SOFIA et le modèle PWR_simu. . . . .	92
6.2	Table d'interpolation pour l'effet modérateur en début de vie (DDV), milieu de vie (MDV) et fin de vie (FDV). . . . .	97
7.1	Fonctionnement de l'OAPS en mode recommandations. . . . .	105
7.2	Recalage de la prédiction de l'OAPS sur les mesures. . . . .	105
7.3	Résultats d'estimation en l'absence de terme de coût d'arrivée. . . . .	112
7.4	Impact du choix de la pondération $S$ et de l'horizon $H$ sur l'erreur d'estimation. . . . .	114

7.5	Estimation de l'état et des paramètres $K_{\text{rod}, X1}, K_{\text{rod}, X2}, K_{\text{rod}, X3}$ et $K_{\text{bore}}$ pour des mesures non bruitées. . . . .	118
7.6	Estimation de l'état et des paramètres $K_{\text{rod}, X1}, K_{\text{rod}, X2}, K_{\text{rod}, X3}$ et $K_{\text{bore}}$ pour des mesures bruitées. . . . .	118
7.7	Estimation de l'état et des paramètres $D$ et $\Sigma_X$ pour des mesures non bruitées.	119
7.8	Estimation de l'état et des paramètres $D$ et $\Sigma_X$ pour des mesures bruitées. . .	120
7.9	Résultats d'estimation de l'état avec un horizon d'une heure pour MHE et asMHE. . . . .	124
B.1	Variation de la puissance et de la concentration en bore. . . . .	135
B.2	Position des groupes de barres de contrôle au cours de la variation de charge.	136
B.3	Distribution de la puissance neutronique. . . . .	136
B.4	Températures aux frontières des mailles. . . . .	137
B.5	Évolution de la concentration en xénon (normalisée). . . . .	137
B.6	Évolution de la concentration en iode (normalisée). . . . .	138
B.7	Évolution de la population des précurseurs. . . . .	138
C.1	Baisse de charge et signaux de commande. . . . .	140
C.2	Distribution de la puissance neutronique. . . . .	140
C.3	Températures aux frontières des mailles. . . . .	141
C.4	Évolution de la concentration xénon (normalisée). . . . .	141

## LISTE DES TABLEAUX

1.1	Dualité entre la commande prédictive et l'estimation à horizon glissant . . . .	24
2.1	Tableau de Butcher. . . . .	37
2.2	Tableau de Butcher : méthode RK4. . . . .	38
2.3	Tableau de Butcher pour une méthode DIRK ( $a_{ii} \neq 0$ ). . . . .	38
2.4	Tableau de Butcher pour une méthode de collocation. . . . .	40
3.1	Variables de décision pour chaque méthode de transcription et pour le type de problème de contrôle optimal. . . . .	57
5.1	Réactions nucléaires ayant lieu au sein d'un REP [CEA15]. . . . .	81
7.1	Contraintes imposées aux variables d'états estimées. . . . .	109
7.2	Paramètres de réglage d'IPOPT. . . . .	110
7.3	Résultats de l'estimation conjointe des paramètres $K_{rod, X1}, K_{rod, X2}, K_{rod, X3}, K_{bore}$ et de l'état. . . . .	119
7.4	Résultats de l'estimation conjointe des paramètres $D, \Sigma_X$ et de l'état. . . . .	120
7.5	Calibration des paramètres du modèle. . . . .	123
7.6	Évaluation des performances de l'estimation de l'état. . . . .	124

## LISTE DES PRODUCTIONS SCIENTIFIQUES

Publications en conférence internationale à comité de relecture :

- Lucas GRUSS et al. « Moving Horizon Estimation of Xenon in Pressurized Water Nuclear Reactors Using Variable-Step Integration ». In : *ECC 2023 : European Control Conference*. Bucharest, Romania : euca, juin 2023
- Lucas GRUSS et al. « Implementation and Validation of Simultaneous State and Parameter Moving Horizon Estimation of a Pressurized Water Reactor ». In : *ECC 2024 : European Control Conference*. Stockholm, Sweden : euca, juin 2024

Brevet :

- L. GRUSS, A. GROSSETÊTE, P. CHEVREL, M. YAGOUBI, « Method for improving parameters of prediction model for reactor control », déposé le 09/01/2024

En cours de rédaction pour publication en revue :

- Adaptive Output MPC with MHE-based State and Parameter Estimation, for nuclear PWR control.
- Resilient and robust MHE for nuclear PWR State and Parameter Estimation.

# RÉSUMÉ

## Résumé en Français

La commande prédictive (MPC) comporte d'indéniables atouts pour la commande du cœur des réacteurs à eau pressurisée (REP) ; elle a vocation à enrichir l'offre commerciale de Framatome. Néanmoins, les performances de la MPC sont conditionnées par la qualité du modèle de prédiction sous-jacent et la connaissance à chaque instant de l'état courant du modèle de réacteur. Cette thèse revient sur ces deux aspects qui constituent les verrous industriels. Elle propose un modèle de cœur de complexité juste suffisante, et le moyen de procéder à la calibration adaptative de certains de ses paramètres. C'est que la dynamique du cœur évolue avec l'épuisement du combustible ; sa prise en compte par des paramètres prédéterminés nécessite un travail conséquent. A rebours, nous proposons l'utilisation d'un estimateur à horizon glissant (MHE), pour estimer conjointement l'état du modèle ainsi que certains paramètres. Raideur et non-linéarité du modèle prédictif (en lien avec les caractéristiques du process) rendent les choix d'implémentation de l'estimateur cruciaux. Exploitation directe du modèle à temps continu, intégrateur implicite, collocation, interpolation des données d'entrée sont autant de concepts participant à la performance et résilience de la MHE, là où les méthodes traditionnelles échouent (e.g. EKF). L'estimateur est validé sur la base de simulations fines réalisées sur le simulateur expert SOFIA utilisé à Framatome.

## Résumé en Anglais

The model predictive control (MPC) approach has undeniable advantages for the core control of Pressurized Water Reactors (PWR) ; it is poised to enhance Framatome's commercial offering. Nevertheless, the performance of MPC is conditioned by the quality of the underlying predictive model and the real-time knowledge of the current state of the reactor core model. This thesis revisits these two aspects that constitute industrial barriers. It proposes a model of the core with just sufficient complexity and a method to perform adaptive calibration of some of its parameters. This is because the core dynamics evolve with fuel depletion ; accounting for this through predetermined parameters requires significant effort. We propose the use of a Moving Horizon Estimator (MHE) to jointly estimate the state variables of the model. The stiffness and non-linear nature of the dynamic model (related to the characteristics of the process) make the the implementation choices of the estimator crucial. Direct exploitation of the continuous-time model, implicit integrator, collocation, interpolation of input data, etc. are key concepts resulting into a performant and resilient MHE estimator, where traditional methods fail (e.g., EKF). The estimator is validated based on fine simulations carried out on the expert simulator SOFIA used at Framatome.

# INTRODUCTION

## Contexte

Le réseau électrique permet de distribuer et d'acheminer l'électricité produite par l'ensemble des unités de production aux consommateurs. En France et en Europe, l'électricité est distribuée en courant alternatif, à une fréquence de 50Hz. Il est important que la fréquence du courant soit régulée précisément car un trop grand écart avec la référence pourrait entraîner une détérioration des équipements sur le réseau.

Il est nécessaire que les quantités d'énergie consommée et produite soient en équilibre car il n'existe pas aujourd'hui de solutions de stockage de l'électricité. Un déséquilibre important détériore la qualité du courant. En cas de déséquilibre encore plus important, on s'expose à des risques de coupures du courant voire des *blackouts*. Plusieurs mécanismes de régulation de la fréquence du courant sont mis en œuvre vis-à-vis de la production[Reb+07a] :

- le réglage primaire à l'échelle de chaque unité de production (suivant une régulation de type proportionnelle) qui vise à limiter le déséquilibre ;
- le réglage secondaire qui est centralisé et vise à rétablir la fréquence du courant à sa référence (suivant une régulation de type PI) ;
- le réglage tertiaire qui vise à rétablir les réserves primaires et secondaires.

ainsi que des mécanismes de marché[Reb+07b]. Les énergies renouvelables telles que l'éolien ou le solaire photovoltaïque sont non pilotables. Au gré des variations des conditions météorologiques, la production de ces énergies varie en temps réel et elles ont priorité sur les autres énergies. De plus, les énergies renouvelables sont massivement déployées dans le but de réduire la part des énergies carbonnées (pétrole, charbon, gaz) dans le mix énergétique. En conséquence, les unités de production pilotables telles que les centrales thermiques, hydrauliques et nucléaires doivent porter le poids de l'équilibrage de la production sur la consommation et faire preuve de toujours plus de flexibilité. Des situations de prix spot de l'électricité négatif[PD22] (figure 1) voient le jour plus en plus fréquemment.

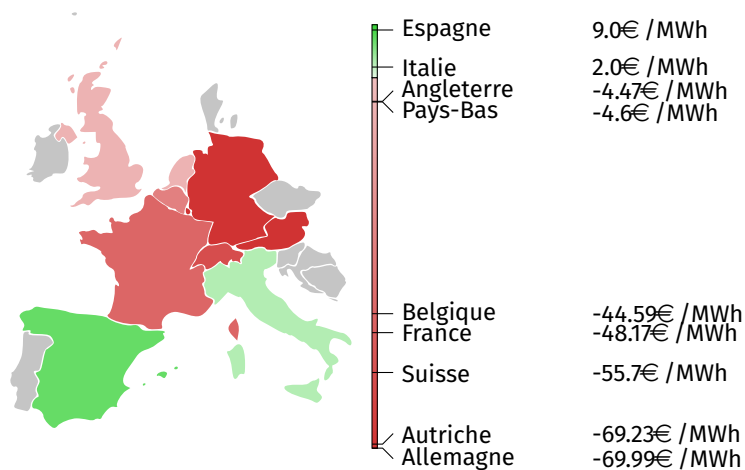


FIGURE 1 – Prix de l'électricité en Europe le 24 mai 2020.

## Flexibilité des centrales nucléaires

Les centrales nucléaires peuvent fonctionner en « base » (à 100% de leur puissance nominale) ou en « suivi de charge » où il s'agit de faire varier la puissance du réacteur pour adapter la production d'électricité de la tranche et suivre une référence de puissance. Les réacteurs sont des systèmes complexes et de nombreuses considérations entrent en jeu dans leur exploitation comme par exemple la sûreté, la réduction des coûts et la maximisation des profits, la réduction des effluents, la limitation de l'usure des actionneurs et la robustesse de la commande.

De précédents travaux à Framatome visant à l'amélioration de la flexibilité des centrales [Lem18 ; Dup23] ont mené au développement d'outils d'aide au pilotage basés sur la commande prédictive, qui possède de nombreux avantages en particulier vis-à-vis du respect de contraintes de fonctionnement difficile à garantir autrement. Ces travaux ont permis d'élaborer un modèle de cœur adapté à la commande prédictive suffisamment simple du point de vue calculatoire mais néanmoins pertinent au regard de sa capacité à prédire le comportement court terme du processus. La solution d'aide au pilotage proposée dans l'offre commerciale de Framatome, appelée OAPS (Operator Aid Predictive System), a été validée par des simulations fines et des tests en situation simulée avec des opérateurs. Néanmoins, il reste quelques verrous technologiques à lever pour faciliter le déploiement de la solution et garantir ses performances.

D'une part le comportement du cœur, à mesure que le combustible est épuisé, s'éloigne du comportement du modèle conceptuel. Les paramètres du modèle de commande peuvent être ajustés périodiquement pour assurer que le modèle reste représentatif du comportement courant du cœur. Cette modification des paramètres suppose dès lors de déterminer en amont plusieurs jeux de paramètres pour le modèle, puis d'effectuer en ligne une mise à jour de la paramétrisation adaptée. Cette approche requiert cependant une maintenance qui peut s'avérer lourde de mise en œuvre.

D'autre part, la commande prédictive nécessite de connaître à tout instant l'état du process projeté sur le modèle de prédiction utilisé par la MPC. Dans le cas du modèle de cœur employé, certaines variables d'état ne sont pas mesurées. Il est nécessaire de reconstruire les variables manquantes à partir des mesures disponibles. Un schéma d'estimation *ad-hoc* est implémenté à ce stade avec l'outil d'aide au pilotage mais présente certaines limites :

- Les commandes effectivement appliquées au système par l'opérateur peuvent différer de la suggestion de l'outil d'aide au pilotage (stratégie non suivie ou avec retard). Cependant, le schéma d'estimation actuel suppose que l'opérateur suit les recommandations à la lettre. L'estimation devrait de préférence prendre en compte les signaux de commande effectivement appliqués au système.
- Les estimations de certains états non mesurés évoluent en boucle ouverte. La stabilité de l'erreur d'estimation n'est pas garantie et la pertinence des recommandations d'aide au pilotage peut-être discutée.
- Les signaux de mesure peuvent parvenir avec un retard, en fonction des capteurs disponibles et du système d'acquisition.

Cette thèse entend répondre aux besoins soulevés précédemment en proposant un estimateur fiable pour l'état et certains paramètres d'intérêt du modèle interne de la MPC. La tâche est rendue complexe du fait de la forte raideur du process, du caractère non-

linéaire et d'ordre élevé du modèle conceptuel et des contraintes qui s'imposent au schéma d'estimation. Ces points de difficulté conditionnent lourdement le choix de la méthodologie envisagée. L'estimation par horizon glissant nous a semblé la plus à même à répondre au cahier des charges.

## Contributions

La contribution principale de cette thèse consiste en la conception et l'implémentation d'un estimateur d'état et paramètres résilient, par horizon glissant, pour un modèle de réacteur nucléaire. Le modèle d'estimation, bien que simplifié en comparaison des codes de calcul scientifique modélisant finement le système, est complexe du fait de ses non-linéarités, de sa raideur et de son ordre élevé. Il nécessite en outre d'ajuster certains paramètres de réglage.

Dans un second temps, les qualités de prédiction du modèle thermique de l'échauffement du caloporteur ont été grandement améliorées en se basant sur un calcul portant sur l'enthalpie massique et une calibration sur un simulateur fin. Le gain de représentativité du modèle contribue à la précision de l'estimateur, qui a été validée sur des données issues de simulations fines.

Enfin, nous avons réalisé et formalisé un positionnement bibliographique large et une réflexion méthodologique *ad hoc* sur la résolution des problèmes de contrôle optimal, dans un formalisme commun à la commande prédictive et l'estimation par horizon glissant. Il permet notamment de mieux cerner les choix d'implémentation nécessaires au bon fonctionnement de l'estimateur. En particulier, la discussion se concentre sur les méthodes de simulation numérique, de transcription d'un problème de contrôle optimal en programme non-linéaire et les méthodes de résolution de ces derniers. L'attention porte en particulier sur les optimisations qui rendent possibles des résolutions de ces problèmes en temps raisonnable.

## Plan du manuscrit

Le présent manuscrit est divisé en deux parties<sup>a</sup>. La **première partie** traite de l'estimation par horizon glissant en l'introduisant formellement dans le **chapitre 1** et le positionnant par rapport à d'autres méthodes populaires d'estimation. La résolution par les méthodes directes du problème d'optimisation à résoudre en MHE fait l'objet du reste de la partie. Les « briques » élémentaires de la résolution numérique sont présentées : la simulation numérique (**chapitre 2**), la transcription du problème à temps continu en un programme non-linéaire (**chapitre 3**) et enfin les méthodes de résolution des programmes non-linéaires (**chapitre 4**). La **seconde partie** présente les éléments de modélisation des cœurs de réacteurs nucléaires partant la physique du système aux modèles de conception et de validation. Enfin l'estimation par horizon glissant implémentée pour l'état et les paramètres du modèle de conception est présentée.

---

a. Ce plan ne reflète pas le cheminement qui a mené au choix de la MHE pour l'application cible. méthodologiques et applicatives. Le choix de la MHE trouve sa justification à la lumière de la deuxième partie, qui présente le système et sa modélisation.





PREMIÈRE PARTIE

**ESTIMATION À HORIZON GLISSANT**



## CHAPITRE 1

# ÉTAT DE L'ART EN ESTIMATION

Considérons un système dynamique dans une représentation d'état :

$$\begin{aligned}\dot{x}(t) &= f(x(t), u(t)) \\ y(t) &= h(x(t), u(t))\end{aligned}\tag{1.1}$$

où  $x \in \mathbb{R}^n$ ,  $u \in \mathbb{R}^{n_u}$ ,  $y \in \mathbb{R}^{n_y}$ ,  $n$  est appelé l'ordre du système, la fonction  $f$  est appelée fonction d'évolution du système, la fonction  $h$  la fonction de mesure et  $u$  désigne les signaux de commande (ou entrées) appliquées au système. Il est courant que l'état complet  $x(t)$  du système soit inaccessible<sup>a</sup>. C'est une limite pour la synthèse de lois de commande qui exploitent tout l'état mais également à des fins de diagnostic sur le système. Dans certains cas, il est possible cependant de reconstruire l'état interne du système à partir de ses sorties et de ses entrées passées. On parle de système *observable*. Pour un tel système, il est possible de concevoir un système dynamique appelé observateur dont le rôle est de reconstruire tout ou partie de variables d'état non mesurées. Un observateur est un système dynamique de la forme suivante :

$$\dot{z}(t) = F(z(t), h(x(t), u(t)))$$

où  $z \in \mathbb{R}^n$  est l'état de l'observateur. Il est possible de concevoir l'observateur tel que  $T(z) \xrightarrow[t \rightarrow \infty]{} x$  pour une certaine fonction  $T : \mathbb{R}^{n_z} \rightarrow \mathbb{R}^n$  [Lue72] (le cas trivial correspond à l'identité). On note par la suite  $\hat{x}$  l'estimée de  $x$ .

Les mesures sur le système sont utilisées pour reconstruire les états non mesurés. Ces mesures sont bruitées par nature et il est possible que la même quantité soit mesurée par plusieurs capteurs différents. L'observateur peut alors avoir un rôle de fusion des données, afin d'agréger plusieurs mesures imprécises en une estimation plus fiable. Par ailleurs, le système est soumis à des perturbations extérieures, qui ne sont pas forcément mesurées. L'observateur peut alors servir à filtrer les perturbations et bruits de mesure.

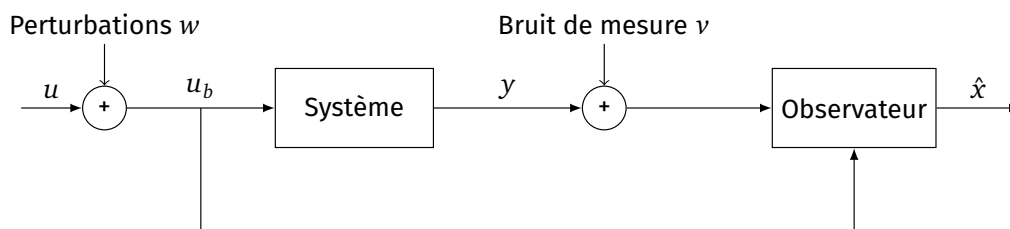


FIGURE 1.1 – Estimation d'état via un observateur.

a. Par manque d'instrumentation, soit pour des raisons économiques ou parce qu'il est physiquement impossible de réaliser une telle mesure.

## 1.1 Observabilité

Il convient en premier lieu de définir la notion d'observabilité pour un système dynamique. De manière générale, un système est dit observable s'il est possible de déterminer son état interne en « regardant » ses sorties. L'évolution des sorties, en connaissance de la dynamique du système (son comportement), permet de déterminer la valeur de l'état interne sans le mesurer. Autrement dit, un système observable est un système pour lequel il existe un observateur.

### Définition 1.1 (Système observable en temps $T$ )

On dit d'un système qu'il est **observable en temps  $T$**  s'il est possible de reconstruire l'état initial du système  $x_0$  à partir de l'observation des entrées  $U_I$  et sorties  $Y_I$  du système sur l'intervalle de temps  $I = [0, T]$ .

L'observabilité d'un système dépend de sa dynamique, ainsi que de l'instrumentation sur ce système, c'est-à-dire les mesures disponibles (certaines mesures sont cruciales pour reconstruire les états du systèmes) ainsi que la nature des bruits de mesure. Une propriété nécessaire à l'observabilité d'un système est sa détectabilité [ABS13; Ber19].

### Définition 1.2 (Système détectable)

Soit le système eq. 1.1 et soient deux trajectoires  $x_1(t)$  et  $x_2(t)$  issues du signal de commande  $u(t)$  appliqué au système à partir des conditions initiales  $x_1(t_0)$  et  $x_2(t_0)$ . Le système eq. 1.1 est dit **détectable** si :

$$\forall t > 0, h(x_1(t)) = h(x_2(t)) \Rightarrow \lim_{t \rightarrow \infty} \|x_1(t) - x_2(t)\| = 0. \quad (1.2)$$

Pour un système détectable, bien que les conditions initiales  $x_1(t_0)$  et  $x_2(t_0)$  soient indifférenciables par la mesure, les trajectoires sont asymptotiquement identiques.

### 1.1.1 Observabilité des systèmes linéaires

Dans le cas des systèmes linéaires, il existe des critères simples à mettre en œuvre pour déterminer leur observabilité. Ces critères sont intéressants car ils sont exprimés par des conditions suffisantes et nécessaires, on peut donc répondre immédiatement à la question de l'observabilité ou de la non-observabilité du système. On introduit le système linéaire invariant dans le temps :

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \quad (1.3)$$

### Définition 1.3 (Matrice d'observabilité)

La matrice d'observabilité du système (eq. 1.3) est définie comme :

$$\mathcal{O} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad (1.4)$$

**Théorème 1.4 (Critère de Kalman)**

Le système linéaire (eq. 1.3) est observable si et seulement si sa matrice d'observabilité (eq. 1.4) est de rang  $n$ .

**1.1.2 Observabilité des systèmes non-linéaires**

Nous faisons ici le choix, pour la simplicité de l'exposé, de rappeler des résultats choisis concernant la caractérisation de l'observabilité d'un système non-linéaire. Soit le système non-linéaire autonome (sans terme de commande) :

$$\begin{aligned}\dot{x}(t) &= f(x(t)) \\ y(t) &= h(x(t))\end{aligned}\tag{1.5}$$

Le critère d'observabilité de Kalman se généralise aux systèmes non-linéaires [KET73 ; Ber19], en passant par la construction de "l'observability mapping"  $H$  du système via les dérivées de Lie. Soit la fonction :

$$O_k(x) = \begin{bmatrix} h(x) \\ \vdots \\ dL_f^{k-1}(h)(x) \end{bmatrix}\tag{1.6}$$

où  $dL_f^{k-1}(h)(x) = \frac{d^{k-1}y}{dt^{k-1}}$ . S'il existe  $k$  tel que cette matrice soit de rang  $n$  en  $x$ , alors le système est observable localement en  $x$  [HK77]. Dans le cadre linéaire, le théorème de Cayley-Hamilton nous permet de déterminer qu'il n'est pas nécessaire d'aller au delà de  $k = n - 1$  (au delà, les puissances de  $A^k$  s'expriment comme combinaison linéaire des  $(I, A, \dots, A^{n-1})$  et le rang de la matrice d'observabilité ne peut plus augmenter). Un système linéaire non observable en  $n$  instants n'est pas observable. Un tel résultat n'existe pas pour les systèmes non-linéaires, donc nous ne savons pas « quand il faut s'arrêter ». Ne pas trouver de valeur de  $k$  qui puisse satisfaire la condition ne constitue pas une preuve de la non-observabilité du système. La détermination de  $O_k(x)$  pour un système non trivial s'avère difficile et sujet aux erreurs, ce qui peut nécessiter de recourir au calcul formel.

Ce critère s'étend aux systèmes avec un terme de commande  $u$ , mais dans ce cas, l'observabilité du problème dépend également du signal de commande appliqué au système [Ber19].

D'autres définitions de l'observabilité (et de la détectabilité) d'un système non-linéaire sont employées, comme la stabilité entrée/sortie-état incrémentale (i-IOSS) [SW97 ; RMD17], qui permet l'analyse de la stabilité de l'estimateur à horizon glissant.

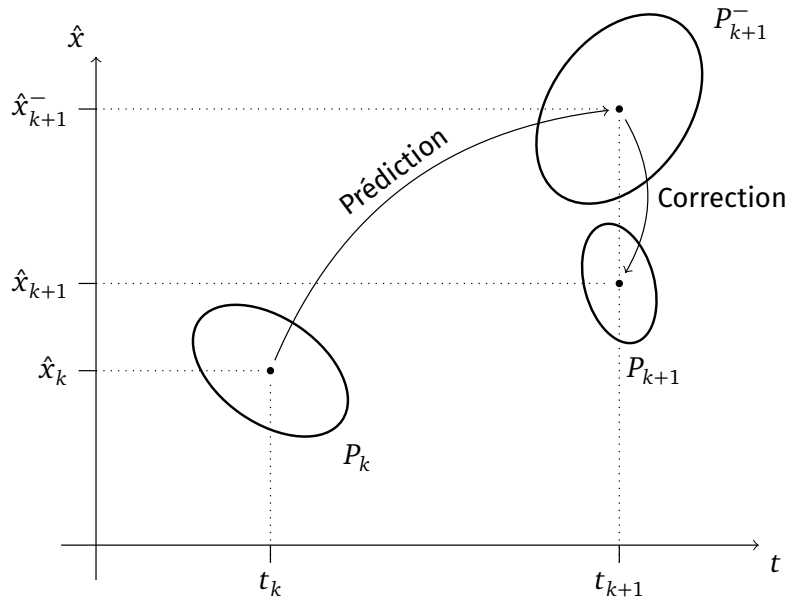


FIGURE 1.2 – Fonctionnement du filtre de Kalman. L'étape de correction permet à la fois d'intégrer les mesures pour améliorer l'estimation et de réduire l'incertitude sur l'erreur d'estimation.

## 1.2 Observateur d'état

### 1.2.1 Observateur de Luenberger

Soit le système linéaire (eq. 1.3). Si la matrice  $A$  est Hurwitz (i.e. pour toute valeur propre  $\lambda$  de  $A$  :  $\Re(\lambda) < 0$ ), alors le système :

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t)$$

Est un observateur pour le système donné par (eq. 1.3). En effet, en notant  $\varepsilon(t) = x(t) - \hat{x}(t)$ , on a :

$$\dot{\varepsilon}(t) = \dot{x}(t) - \dot{\hat{x}}(t) = A\varepsilon(t) \text{ donc } \varepsilon(t) \xrightarrow[t \rightarrow \infty]{} 0.$$

La convergence de l'estimation est la même que celle du système observé. L'observateur de Luenberger est le système dynamique suivant :

$$\begin{aligned} \dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + L(y(t) - \hat{y}(t)) = (A - LC)\hat{x}(t) + Bu(t) \\ \hat{y}(t) &= C\hat{x}(t) + Du(t) \end{aligned} \quad (1.7)$$

La synthèse de l'observateur de Luenberger est également possible à temps discret. Si le système (eq. 1.3) est observable, il est possible de placer les pôles de  $(A - LC)$  arbitrairement [Lue72] (dans la partie du plan complexe à gauche de l'axe imaginaire) pour assurer une convergence imposée à l'erreur d'estimation. Néanmoins, en cas de perturbations sur le système, le choix de pôles trop grands en valeur absolue rend l'observateur sensible aux perturbations. La connaissance *a priori* de la nature des signaux de perturbation permet de choisir au mieux le gain d'estimation  $L$ , ce que permet la synthèse du filtre de Kalman.

## 1.2.2 Filtre de Kalman

Le filtre de Kalman [Kal60] est un algorithme d'estimation d'état présenté par Rudolf Kalman dans les années 60. C'est un algorithme très populaire et simple à implémenter. Sous réserve d'observabilité du système, il permet de reconstruire asymptotiquement l'état du système à partir d'une estimation initiale erronée et de mesures bruitées. Une construction du filtre de Kalman dans le cadre des systèmes discrets (qui est le cadre classique de présentation du filtre) est proposée en annexe A. Le filtre de Kalman procède en 2 étapes (figure 1.2) :

1. Prédiction : le modèle est appliqué à partir de l'estimation courante pour obtenir une estimation *a priori*.
2. Correction : la mesure disponible sur le système est utilisée pour corriger l'estimation *a priori* et réduire la covariance de l'erreur d'estimation *a posteriori*.

Plus précisément, on estime à la fois l'état du système et l'erreur réalisée via un estimateur linéaire<sup>a</sup>. La mesure est utilisée pour corriger la prédiction via le gain de Kalman  $K_k$  (qui dépend de l'instant  $k$ ), choisi de sorte à minimiser la variance de l'erreur d'estimation commise. En ce sens, le filtre de Kalman est la solution optimale au problème d'estimation linéaire pour la variance de l'erreur d'estimation. On considère le système soumis aux perturbations gaussiennes en entrée  $w_k \sim \mathcal{N}(0, W)$  et en sortie  $v_k \sim \mathcal{N}(0, V)$ .

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = Cx_k + Du_k + v_k \end{cases} \quad (1.8)$$

L'estimation initiale de l'état est  $\hat{x}_0$  et on note  $\varepsilon_0$  l'erreur initiale qui suit une loi normale centrée en 0 de variance  $P_0$  (on note  $\varepsilon_0 \sim \mathcal{N}(0, P_0)$ ). Le filtre de Kalman est un système dynamique de la forme :

$$\begin{cases} \hat{x}_{k+1} = A\hat{x}_k + Bu_k + K_{k+1}(y_{k+1} - \hat{y}_{k+1}) \\ \hat{y}_k = C\hat{x}_k + Du_k \end{cases} \quad (1.9)$$

Soit  $\hat{x}_k$  l'estimée de l'état (*a posteriori*) à l'instant  $k$ , et l'erreur commise sur l'estimation  $x_k - \hat{x}_k = \varepsilon_k \sim \mathcal{N}(0, P_k)$ . L'estimée *a priori*  $\hat{x}_{k+1}^-$  pour l'instant  $k+1$  et l'erreur d'estimation *a priori*  $\varepsilon_{k+1}^- \sim \mathcal{N}(0, P_{k+1}^-)$  sont telles que :

$$\hat{x}_{k+1}^- = A\hat{x}_k + Bu_k \quad (1.10a)$$

$$P_{k+1}^- = AP_k A^T + W \quad (1.10b)$$

On corrige la prédiction en utilisant l'innovation  $y_{k+1} - C\hat{x}_{k+1}^-$  et le gain de Kalman  $K_{k+1}$ . L'estimation corrigée (estimation *a posteriori* de la mesure)  $\hat{x}_{k+1}$  et l'erreur d'estimation  $\varepsilon_{k+1} \sim \mathcal{N}(0, P_{k+1})$  sont telles que :

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - \hat{y}_{k+1}) \quad (1.11a)$$

$$P_{k+1} = (I - K_{k+1}C)P_{k+1}^-(I - K_{k+1}C)^T + K_{k+1}VK_{k+1}^T \quad (1.11b)$$

$$K_{k+1} = P_{k+1}^- C^T (CP_{k+1}^- C^T + V)^{-1} \quad (1.11c)$$

a. L'erreur d'estimation est un signal normal centré en 0, il suffit donc de caractériser la covariance de l'erreur pour paramétriser totalement l'erreur d'estimation.



En réunissant (eq. 1.10b) et (eq. 1.11), on obtient l'expression du filtre de Kalman :

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + K_{k+1}(y_{k+1} - \hat{y}_{k+1}) \quad (1.12a)$$

$$P_{k+1} = (I - K_{k+1}C)(AP_kA^T + W)(I - K_{k+1}C)^T + K_{k+1}VK_{k+1}^T \quad (1.12b)$$

$$K_{k+1} = (AP_kA^T + W)C^T(C(AP_kA^T + W)C^T + V)^{-1} \quad (1.12c)$$

(eq. 1.12b) est appelée équation (discrète) de Riccati. Couramment, on utilise un gain de Kalman correspondant à la solution stationnaire de l'équation de Riccati (équation algébrique discrète de Riccati), l'estimateur prenant la forme d'un estimateur de Luenberger :

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L(y_{k+1} - \hat{y}_{k+1}) \quad (1.13a)$$

$$P = (I - LC)(APA^T + W)(I - LC)^T + LVL^T \quad (1.13b)$$

$$L = (APA^T + W)C^T(C(APA^T + W)C^T + V)^{-1} \quad (1.13c)$$

Le filtre de Kalman se généralise aux systèmes linéaires variant dans le temps en remplaçant  $A$  par  $A_k$  et  $C$  par  $C_k$  dans (eq. 1.12), lorsque cela s'applique. Le filtre de Kalman se formule également pour les systèmes à temps continu [Sim06], appelé parfois filtre de Bucy-Kalman eq. 1.14 où (eq. 1.14b) est appelée équation différentielle de Riccati.

$$\dot{\hat{x}}(t) = A\dot{\hat{x}}(t) + Bu(t) + K(t)(y(t) - \hat{y}(t)) \quad (1.14a)$$

$$\dot{P}(t) = AP(t) + P(t)A^T - P(t)C^T V^{-1} CP(t) + W \quad (1.14b)$$

$$K(t) = P(t)C^T V^{-1} \quad (1.14c)$$

Dans le cas continu, on ne peut plus distinguer les étapes de prédiction et de correction, qui sont réalisées simultanément.

En outre, si  $w \sim \mathcal{N}(b, W)$  ( $w$  non centré), il est possible de se ramener au cas général en considérant  $\tilde{w} = w - b$  et le modèle d'évolution :

$$x_{k+1} = Ax_k + \begin{bmatrix} B & b \end{bmatrix} \begin{bmatrix} u_k \\ 1 \end{bmatrix} + \tilde{w}_k \quad (1.15)$$

Si le biais  $b$  est inconnu, on peut l'estimer en augmentant l'état de l'observateur de l'état  $\hat{b}$  et la dynamique de l'observateur de  $\dot{\hat{b}}_{k+1} = \hat{b}_k$  (ou  $\dot{\hat{b}}(t) = 0$ ). On peut enfin étendre le filtre de Kalman aux bruits colorés [Ala06] en augmentant le modèle de l'observateur du modèle d'évolution de la perturbation.

### 1.2.3 Filtre de Kalman étendu

Le filtre de Kalman présenté précédemment suppose que le système est linéaire. Afin de traiter les systèmes non linéaires, le filtre de Kalman étendu implémente un filtre de Kalman avec pour modèle le linéarisé tangent d'un modèle non-linéaire, autour de l'estimée courante. Soit le système :

$$\begin{cases} x_{k+1} = f(x_k, u_k, w_k) \\ y_k = h(x_k, u_k, v_k) \end{cases} \quad (1.16)$$

La prédiction *a priori* de l'état exploite directement le modèle non-linéaire ( $w_k = 0$ ) :

$$\hat{x}_{k+1}^- = f(\hat{x}_k, u_k, 0)$$

Pour calculer la propagation de la covariance de l'erreur, on considère l'expansion de Taylor de  $x_{k+1}$  à l'ordre 1 du système, autour d'une estimée  $\hat{x}_k$  :

$$x_{k+1} \approx f(\hat{x}_k, u_k, 0) + \underbrace{\frac{\partial f}{\partial x}}_{F_k} \Big|_{\hat{x}_k} (x_k - \hat{x}_k) + \underbrace{\frac{\partial f}{\partial w}}_{L_k} \Big|_{\hat{x}_k} w_k \quad (1.17)$$

On linéarise également la mesure en considérant :

$$y_k \approx h(\hat{x}_k, u_k, 0) + \underbrace{\frac{\partial h}{\partial x}}_{C_k} \Big|_{\hat{x}_k} (x_k - \hat{x}_k) + \underbrace{\frac{\partial h}{\partial v}}_{D_k} \Big|_{\hat{x}_k} v_k \quad (1.18)$$

En règle générale, la transformation non-linéaire d'une variable aléatoire gaussienne n'est pas gaussienne. La covariance de l'erreur d'estimation *a priori* après prédiction est donc propagée via l'expansion de Taylor à l'ordre 1 :

$$\varepsilon_{k+1}^- = x_{k+1} - \hat{x}_{k+1}^- \quad (1.19a)$$

$$\approx F_k(x_k - \hat{x}_k) + f(\hat{x}_k, u_k, 0) + L_k w_k - f(\hat{x}_k, u_k, 0) \quad (1.19b)$$

$$\approx F_k \varepsilon_k + L_k w_k \quad (1.19c)$$

La linéarisation du système permet d'approximer l'erreur d'estimation par une variable aléatoire gaussienne. On peut alors calculer la variance de l'erreur d'estimation *a priori* (théorème A.1) :

$$P_{k+1}^- = F_k P_k F_k^T + L_k W L_k^T \quad (1.20)$$

Du reste, on utilise la solution au filtre de Kalman linéaire usuel (eq. 1.12). Dans le cadre linéaire, le filtre de Kalman produit la solution optimale du problème d'estimation non biaisé de variance minimale. Ce n'est pas le cas de l'EKF qui étend l'utilisation du filtre de Kalman aux systèmes non-linéaire via une linéarisation du modèle, mais il n'est plus l'estimateur optimal. Cette linéarisation montre ses limites dans de nombreux cas pratiques, qui peuvent aboutir à une « divergence » de l'algorithme. Le filtre de Kalman étendu nécessite donc de disposer d'une estimation initiale suffisamment proche de l'état réel pour garantir la stabilité de l'erreur d'estimation [Unb+00]. Le filtre de Kalman étendu se généralise également pour les systèmes à temps continu [Sim06].

Il est possible d'obtenir une meilleure qualité de l'approximation en réalisant une linéarisation de Taylor à l'ordre 2, on parle alors de filtre de Kalman étendu d'ordre 2 (EKF2) [Jaz68]. Ce gain de précision se fait à un coût calculatoire accru non négligeable et il implique également que la fonction d'évolution du système soit au moins deux fois dérivable.

### 1.2.4 Filtre de Kalman à « points Sigma »

Dans les années 90 d'autres extensions au filtre de Kalman étendu ont vu le jour, afin de pallier aux limitations de ce dernier en se passant de la linéarisation du système. Les variations les plus connues sont les suivantes [JXC12] :

- l'UKF (Unscented Kalman Filter) ou « filtre de Kalman sans parfum » : il se base sur la « Transformée sans parfum » (Unscented Transform)[JUD95]<sup>a</sup>.
- le CDKF (Central Difference Kalman Filter) [IX00].
- le DDF (Divided Difference Kalman Filter) : il se base sur la formule d'interpolation de Stirling[NPR00]<sup>b</sup>.

En réalité, tous ces filtres reviennent à des implémentations légèrement différentes du « Sigma Points Kalman Filter » [Sim06].  $N$  « points sigma »  $\Sigma_k^{(i)}$  sont choisis<sup>c</sup> de manière à correspondre à la distribution de l'erreur d'estimation *a priori* (espérance et covariance) :

$$\frac{1}{N} \sum_{i=1}^N \Sigma_k^{(i)} = \hat{x}_k, \quad \frac{1}{N} \sum_{i=1}^N (\Sigma_k^{(i)} - \hat{x}_k) (\Sigma_k^{(i)} - \hat{x}_k)^T = P_k \quad (1.21)$$

On prédit la position des points via le modèle, ainsi que les mesures correspondantes :

$$\forall i \in \llbracket 1, N \rrbracket, \Sigma_{k+1}^{(i)} = f(\Sigma_k^{(i)}, u_k, 0) \quad (1.22a)$$

$$y_{k+1}^{(i)} = h(\Sigma_{k+1}^{(i)}, u_k, 0) \quad (1.22b)$$

Les covariances  $P_{x,k+1}^-$ ,  $P_{y,k+1}$  et  $P_{xy,k+1}^-$  sont calculées empiriquement :

$$P_{x,k+1}^- = \frac{1}{N} \sum_{i=1}^N (\Sigma_{k+1}^{(i)} - \hat{x}_{k+1}^-) (\Sigma_{k+1}^{(i)} - \hat{x}_{k+1}^-)^T + W \quad (1.23a)$$

$$P_{y,k+1} = \frac{1}{N} \sum_{i=1}^N (y_{k+1}^{(i)} - \hat{y}_{k+1}) (y_{k+1}^{(i)} - \hat{y}_{k+1})^T + V \quad (1.23b)$$

$$P_{xy,k+1}^- = \frac{1}{N} \sum_{i=1}^N (\Sigma_{k+1}^{(i)} - \hat{x}_{k+1}^-) (y_{k+1}^{(i)} - \hat{y}_{k+1})^T \quad (1.23c)$$

où  $\hat{x}_{k+1}^- = \frac{1}{N} \sum_{i=1}^N \Sigma_{k+1}^{(i)}$  et  $\hat{y}_{k+1} = \frac{1}{N} \sum_{i=1}^N h(\Sigma_{k+1}^{(i)}, u_k)$ . Enfin, la prédiction est corrigée avec la mesure :

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + K_{k+1} (y_{k+1} - \hat{y}_{k+1}) \quad (1.24a)$$

$$P_{k+1} = P_{k+1}^- + K_{k+1} P_{y,k+1} K_{k+1}^T \quad (1.24b)$$

$$K_{k+1} = P_{xy} (P_{y,k+1})^{-1} \quad (1.24c)$$

a. La raison pour laquelle ce filtre est appelé ainsi n'est pas tout à fait certaine. Un bruit court selon lequel le mot « unscented » serait à comprendre au sens « qui ne pue pas », en comparaison à l'EKF. La présence du mot sur un déodorant aurait marqué J. Uhlmann et le terme est resté pour son algorithme. ([https://ethw.org/First-Hand:The\\_Unscented\\_Transform](https://ethw.org/First-Hand:The_Unscented_Transform))

b. Il existe plusieurs variantes de ce filtre qui émergent de l'utilisation à l'ordre 1 ou 2 de la formule d'interpolation de Stirling.

c. Le choix des points sigma différencie les filtres UKF, DDF et CDKF.

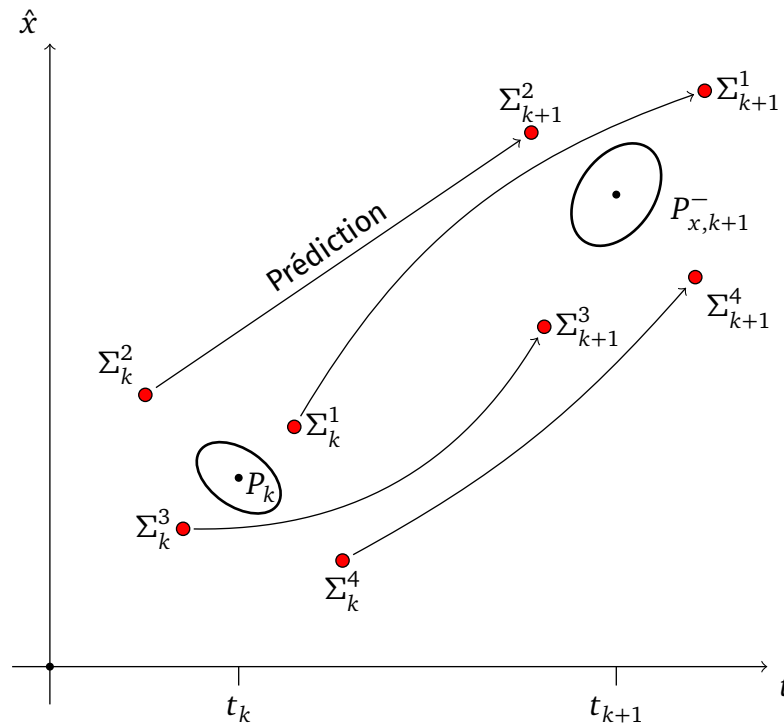


FIGURE 1.3 – Phase de prédiction dans un filtre à points sigma.

Ce type de filtre se distingue de l'EKF, en ce que la linéarisation du système –source d'erreurs sur l'estimation de la covariance– est évitée, puisque le calcul de la covariance se base sur les points sigma et non pas sur la transformation linéaire d'une variable aléatoire gaussienne. Le filtre de Kalman à point sigma est en pratique plus performant que l'EKF du fait d'une meilleure estimation du moment d'ordre 2 de l'erreur d'estimation. Il est également possible d'étendre l'approximation de la densité de probabilité de l'erreur à des ordres supérieurs, comme l'ordre 3 ou 4 [LWZ14]. On peut également formuler le filtre à points sigma dans le contexte des systèmes à temps continu [Sar07 ; KK18].

En augmentant grandement le nombre de points utilisés pour estimer la densité de probabilité de l'état du système, on aboutit au filtre particulaire [Sim06], ou filtre Monte Carlo [DGA00]. Ce type de filtre souffre néanmoins de la « malédiction de la dimension » [RMD17], qui rend le problème très calculatoire, même pour des systèmes de dimension faible (de l'ordre de la dizaine d'états). Des méthodes comme le « feedback particular filtering » [YMM13] ont été conçues afin de réduire la complexité du problème à résoudre.

### 1.3 Estimation de paramètres

Un modèle dynamique inclut un certain nombre de paramètres relatifs à la géométrie du système (longueur d'un pendule simple, masse d'un élément, etc), à la physique modélisée (cinétique d'une réaction chimique, rendement, etc) ou à des approximations. Certains paramètres impliqués dans un modèle dynamique sont entachés d'incertitude dont les causes sont multiples : variabilité des éléments constitutif du système dans un contexte industriel, variabilité dans le temps due à l'usure, erreur sur la mesure...

L'estimation des paramètres d'un modèle consiste à exploiter des données de fonctionne-

ment du système incluant à la fois les entrées et sorties du process. Le modèle du process est simulé et les paramètres sont déterminés de sorte à minimiser l'écart entre les prédictions du modèle et les sorties mesurées [Sor98]. L'estimation des paramètres peut-être réalisée hors-ligne afin d'exploiter une grande quantités de données de fonctionnement du process, ou en ligne de sorte que les paramètres du système sont continuellement ajustés pour correspondre au fonctionnement courant du système. En distinguant les paramètres dont la valeur évolue au cours de temps, et ceux dont la valeur est supposée constante, on peut choisir la méthode la plus appropriée pour garantir les performances de prédiction du modèle.

La régression linéaire aux moindres carrés [Jam+13] pour ajuster automatiquement les paramètres d'un modèle linéaire (ou affine) aux données disponibles, est couramment utilisée pour modéliser une relation linéaire entre des observables  $y \in \mathbb{R}^{n_y}$  et des entrées  $x \in \mathbb{R}^{n_x}$ , en fonction d'un paramètre  $\theta$  :

$$y = \theta x \quad (1.25)$$

En rangeant toutes les entrées et sorties disponibles sur le process dans les matrices  $X$  et  $Y$ , le problème de régression linéaire aux moindres carrés s'écrit [BK17] :

$$\min_{\theta} \|Y - X\theta^T\|_2 \quad (1.26)$$

Ce problème admet une solution analytique  $\hat{\theta} = X^\dagger Y$  où  $X^\dagger$  est la pseudo-inverse de Moore-Penrose de  $X$  [Pen55]. La méthode des moindres carrés se généralise aux modèles non-linéaires mais la solution est rarement analytique. Pour le modèle non linéaire  $y = f(x, \theta)$ , le problème d'estimation aux moindres carrés :

$$\min_{\theta} \sum_k \|y_k - f(x_k, \theta)\|_2 \quad (1.27)$$

se résout numériquement avec des méthodes comme la méthode de Newton (cf section 4.2). Pour un modèle dynamique dépendant de paramètres  $p$ , exprimé sous forme implicite ou descripteur (cf chapitre 2) :

$$\begin{aligned} 0 &= f(\dot{x}(t), x(t), u(t), p) \\ y &= h(x(t), u(t), p), \end{aligned} \quad (1.28)$$

le problème d'estimation de paramètres s'écrit :

$$\min_p \int_0^T \|y(\tau) - h(x(\tau), u(\tau), p)\|_R^2 d\tau \quad (1.29a)$$

$$\text{sous contraintes, } \forall \tau \in [0, T] \quad (1.29b)$$

$$0 = f(\dot{x}(\tau), x(\tau), u(\tau), p) \quad (1.29c)$$

$$0 = g(x(\tau), u(\tau), p) \quad (1.29d)$$

La contrainte  $0 = g(x, u, p)$  permet d'exprimer des contraintes supplémentaires sur les paramètres, telles que la connaissance de plages de valeurs admissibles sur le paramètre estimé.

La méthode des moindres carrés est adaptée à l'estimation des paramètres du système hors-ligne, qui peut exploiter une grande quantité de données. Néanmoins l'utilisation en

ligne de la méthode nécessite de l'adapter, dans la mesure où la taille du jeu de données croît à au fil de leur acquisition au cours du temps rendant le problème insoluble. Dans [Sim06], on trouve une implémentation récursive de la méthode des moindres carrés pour un modèle linéaire, permettant de mettre à jour les paramètres estimés à chaque fois que de nouvelles données sont acquises. On peut également considérer un horizon de taille finie, de sorte que seule une quantité fixe de données les plus à jour est considérée [MR95].

Une autre approche consiste à utiliser les techniques d'estimation d'état pour estimer les paramètres du modèle, en considérant un modèle où l'état du système est augmenté des paramètres à estimer. On suppose couramment qu'à l'échelle de temps de l'estimation le vecteur des paramètres à estimer  $p$  est constant ( $\dot{p} = 0$ ), mais il est également possible d'intégrer un modèle de son évolution [Lju99, chap.11] (par exemple un modèle de l'usure d'un élément, du burnup du combustible d'une centrale nucléaire, etc). On écrit :

$$\begin{aligned} \begin{pmatrix} \dot{x} \\ \dot{p} \end{pmatrix}(t) &= \begin{pmatrix} \tilde{f}(x(t), u(t), p(t)) \\ 0 \end{pmatrix} \\ y(t) &= \tilde{h}(x(t), u(t), p(t)) \end{aligned} \quad (1.30)$$

Les fonctions d'évolution  $f$  et de mesure  $h$  ont été adaptées pour considérer en entrées le vecteur des paramètres à estimer et on dit que la vecteur d'état a été **augmenté**. À partir d'une estimation *a priori*  $(x_0^T \ p_0^T)^T$  on peut alors utiliser les techniques présentées précédemment pour estimer l'état du modèle augmenté. L'estimation conjointe de l'état et de paramètres est parfois appelée estimation adaptative [Sor98] ou problème d'estimation duale [Fux+14] et mène à un problème non-linéaire.

## 1.4 Estimation à horizon glissant

*If all you have is a hammer, everything looks like a nail.*

— *The Psychology of Science*, 1966

Abraham MASLOW

L'estimation à horizon glissant (MHE pour *Moving Horizon Estimation*) est une méthode d'estimation basée sur la résolution en ligne d'un problème d'optimisation. Elle trouve ses origines dans les méthodes d'estimation aux moindres carrés [Jaz68], comme nous l'avons précédemment décrite pour l'estimation de paramètres. La MHE utilise les données disponibles sur le système (commandes et sorties mesurées) sur un horizon de temps passé et détermine la trajectoire de l'état (et éventuellement paramètres) qui « explique » le mieux les données observées. En utilisant la totalité des mesures disponibles depuis le début du fonctionnement du système, on obtient le filtre à information totale (FIE pour *Full Information Estimation*) [RMD17], appelé également *batch estimator* ou estimateur en paquet [MR95]. L'utilisation d'un horizon de taille finie permet de limiter la taille du problème d'optimisation et rendre le problème soluble en temps raisonnable (figure 1.4).

Du fait des similitudes qu'elle présente avec la commande prédictive (MPC pour *Model Predictive Control*), la MHE est une méthode répandue dans la communauté de la commande MPC et est considérée comme l'approche duale de cette dernière pour l'estimation [RMD17]. Ainsi les avancées tant méthodologiques que technologiques principalement développées

pour la commande prédictive ont également profité à la MHE.

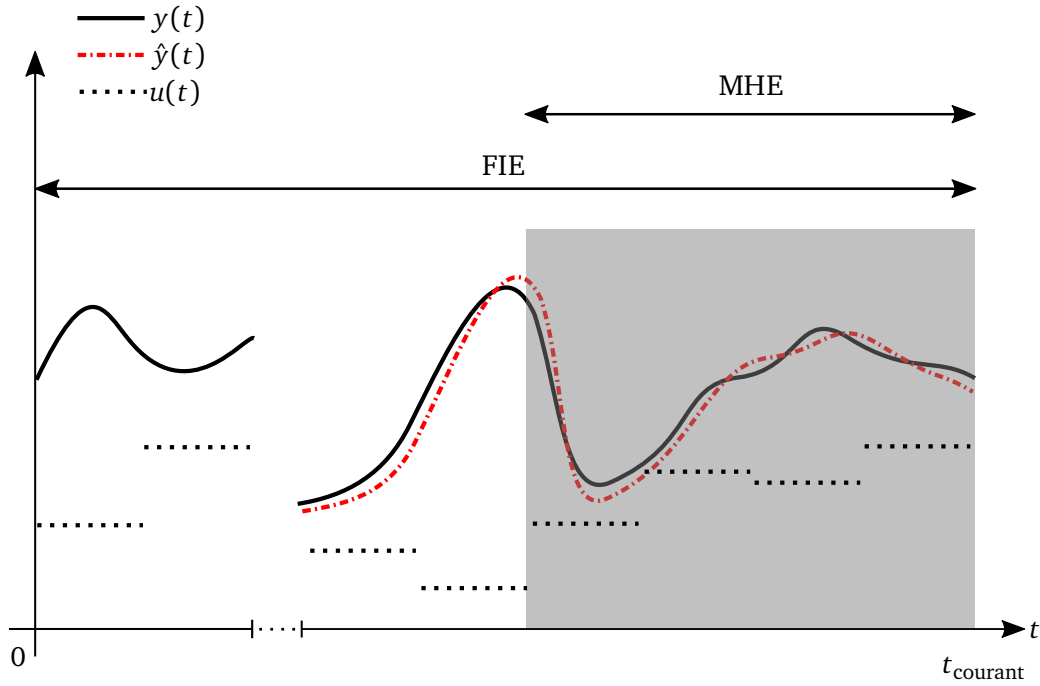


FIGURE 1.4 – Filtre à information totale et estimation par horizon glissant.

#### 1.4.1 Estimation à information totale

L'estimateur FIE utilise la totalité de l'information disponible sur le processus depuis le début de son fonctionnement. A partir de ces informations, à tout instant  $t$  le problème d'optimisation suivant est résolu :

$$\begin{aligned} \min_{x,w,v} J &= V_0(x(0), w(0), v(0)) + \int_0^t V(w(\tau), v(\tau)) d\tau \\ \text{sous contraintes } \forall \tau \in [0, t], \\ \dot{x}(\tau) &= f(x(\tau), u(\tau)) + w(\tau) \\ y(\tau) &= h(x(\tau), u(\tau)) + v(\tau) \\ 0 &\geq h_I(x(\tau), w(\tau), v(\tau)) \end{aligned} \quad (1.31)$$

ou dans sa version discrète :

$$\begin{aligned} \min_{w,v,x} J &= V_0(x_0, w_0, v_0) + \sum_{i=0}^k V(w[i], v[i]) \\ \text{sous contraintes } \forall i \in [0, k] \\ x[i+1] &= f_d(x[i], u[i]) + w[i] \\ y[i] &= h_d(x[i], u[i]) + v[i] \\ 0 &\geq h_{d,I}(x[i], w[i], v[i]) \end{aligned} \quad (1.32)$$

où  $V_0$  est la fonction coût initiale et  $V$  la fonction coût intermédiaire,  $h_I$  (resp.  $h_{d,I}$ ) exprime des contraintes d'inégalité le long de la trajectoire du système à temps continu (resp. à

temps discret). Dès qu'une nouvelle information sur le système est disponible, elle peut être exploitée pour affiner l'estimation de l'état du système depuis l'instant initial de sa mise en fonctionnement. Le problème d'estimation aux moindres carrés doit être résolu en ligne. L'implémentation de cette approche est donc impossible du fait des besoins croissants en mémoire et l'allongement du temps de calcul. Néanmoins l'estimateur FIE possède des propriétés intéressantes du point de vue de l'estimation. Ces propriétés permettent d'identifier le cas idéal ainsi que les hypothèses supplémentaires qu'il sera nécessaire d'ajouter lors de la synthèse d'estimateur à horizon glissant.

On montre [Sor70 ; RMD17], que le problème :

$$\min_{w,v,x} J = \frac{1}{2} \|x[0] - \bar{x}[0]\|_{P_0}^2 + \sum_{i=0}^k \|v[i]\|_{V^{-1}}^2 + \sum_{i=0}^{k-1} \|w[i]\|_{W^{-1}}^2 \quad (1.33)$$

$$\text{sous contraintes } \forall i \in [0, k] \quad (1.34)$$

$$x[i+1] = Ax[i] + w[i] \quad (1.35)$$

$$y[i] = Cx[i] + v[i] \quad (1.36)$$

$$(1.37)$$

c'est à dire le problème (eq. 1.32) appliqué au système linéaire (eq. 1.8) (donc sans contraintes) avec des fonctions coût intermédiaire ( $\bar{x}_0$  est une estimation *a priori* de l'état initial du système) et terminale quadratiques, a pour solution récursive le filtre de Kalman. C'est en cela que le filtre de Kalman est parfois appelé estimateur linéaire quadratique [Sor70]. Néanmoins si l'on souhaite ajouter des contraintes sur les estimations, le filtre de Kalman ne convient pas, là où le FIE s'accommode tout à fait de contraintes, à condition d'employer une méthode de résolution adaptée. Il en va de même pour les filtres à points sigma, qui ne permettent pas d'imposer *a priori* de contraintes sur les estimations de l'état. Pour les problèmes non-linéaires, l'approche par optimisation présente également l'avantage de ne pas nécessiter la linéarisation du modèle, contrairement au filtre de Kalman étendu.

Par ailleurs, sous hypothèse de détectabilité du système et que les perturbations soient bornées, l'estimateur FIE montre d'excellentes propriétés de convergence robuste, c'est à dire que l'erreur d'estimation est stable [RMD17].

#### 1.4.2 Introduction d'un horizon de taille fixe

En choisissant un horizon d'estimation de taille finie  $H$  le problème d'estimation à horizon glissant peut se formuler de la manière suivante :

$$\begin{aligned} \min_{x,w,v} J &= V_I(x(t-H), w(t-H), v(t-H)) + \int_{t-T}^t V(w(\tau), v(\tau)) d\tau \\ \text{sous contraintes } \forall \tau \in [t-T, t], & \\ \dot{x}(\tau) &= f(x(\tau), u(\tau)) + w(\tau) \\ y(\tau) &= h(x(\tau), u(\tau)) + v(\tau) \\ 0 &\geq h_I(x(\tau), w(\tau), v(\tau)) \end{aligned} \quad (1.38)$$



ou dans sa version discrète :

$$\begin{aligned}
\min_{w, v, \hat{x}[0]} J &= V_I(\hat{x}[0]) + \sum_{i=k-N+1}^k V(w[i], v[i]) \\
\text{sous contraintes } \forall i \in [k-N+1; k], & \\
\hat{x}[i+1] &= f(\hat{x}[i], u[i]) + w[i] \\
y[i] &= h(\hat{x}[i], u[i]) + v[i] \\
0 &\geq h_{d,I}(x[i], w[i], v[i])
\end{aligned} \tag{1.39}$$

Si  $V_I$  est choisi de sorte à être le reste de la fonction coût de l'estimateur FIE, c'est à dire :

$$V_i(x(t-H), w(t-H), v(t-H)) = V_0(x(0), w(0), v(0)) + \int_0^{t-H} V(w(\tau), v(\tau)) d\tau, \tag{1.40}$$

FIE et MHE sont exactement les mêmes estimateurs. Pour tout problème d'estimation à information totale, on peut donc se ramener au problème MHE décrit ci-dessus en tronquant le support de la somme et en exploitant uniquement les observations qui coïncident avec un horizon fini d'estimation. Le problème devient alors implémentable en pratique puisque les besoins en terme de mémoire sont finis et constants tout au long du fonctionnement de l'estimateur. Le terme  $V_i(x(t-H), w(t-H), v(t-H))$  est calculé en pratique en tant qu'approximation du coût restant de l'estimateur FIE.

Plusieurs interprétations et formulations de l'estimation à horizon glissant sont possibles. Il est courant de faire un choix de fonction coût intermédiaire qui rappelle le filtre de Kalman en tant qu'estimateur linéaire quadratique :

$$J = \underbrace{\frac{1}{2} \|x_{k-N+1} - \bar{x}_{k-N+1}\|_{S_k}}_{A_k} + \underbrace{\frac{1}{2} \sum_{i=k-N+1}^{k-1} \|w[i]\|_Q}_{P_k} + \underbrace{\frac{1}{2} \sum_{i=k-N+1}^k \|v[i]\|_R}_{C_k} \tag{1.41}$$

- Le **terme de correction**  $C_k$  permet de prendre en compte l'information apportée par la mesure, en associant un coût sur l'écart entre les observations et les estimées.
- Le **terme de perturbation**  $P_k$  sert essentiellement à pallier aux erreurs de modèle en estimant une perturbation sur le process. Ce terme peut être présent, ou absent si l'on considère un process sans perturbations. En l'absence de ce terme, la variable de décision du problème d'optimisation n'est plus que l'état au début de la fenêtre d'observation. Tous les états suivants sont obtenus en utilisant le modèle à partir de cet état.
- Le **coût d'arrivée**  $A_k$  permet de propager les estimations effectuées à l'itération précédente de MHE, en « complétant » l'horizon d'observation jusqu'à l'instant  $t_0$ . Ce terme est vu comme une approximation du coût restant pour atteindre l'estimateur FIE. La matrice  $S_k$  pondère l'écart entre la variable de décision  $\hat{x}[k-N+1]$  et  $\bar{x}_{k-N+1}$  une estimation *a priori* de l'état, souvent prise comme une valeur issue de la résolution précédente du problème MHE (sur un horizon différent).

Deux interprétations de la MHE se côtoient [Suw+16] : l'interprétation déterministe et l'interprétation stochastique. Dans l'interprétation déterministe de l'estimation par horizon

glissant, le choix des matrices de pondération est *ad-hoc* [RMD17] et résulte d'un ajustement par tâtonnement.

### 1.4.3 Interprétation stochastique

L'interprétation stochastique permet de positionner l'estimation par horizon glissant dans un contexte probabiliste et de faire un choix plus « informé » des matrices de pondération. On suppose que les estimées  $\hat{x}(t)$  et les perturbations sont des variables aléatoires de distribution de probabilité connue. La MHE dans le contexte stochastique peut être vue comme approximation d'un filtre bayésien [Val+17]. Le problème MHE correspond à estimer l'état au cours de la fenêtre d'estimation grâce aux mesures disponibles, c'est-à-dire à trouver la densité de probabilité  $P(X_{k-N+1}^k | Y_0^k)$ . On montre [LB12] que cette densité de probabilité s'écrit :

$$P(X_{k-N+1}^k | Y_0^k) \propto \frac{P(x_{k-N+1} | Y_0^{k-1})}{P(Y_{k-N+1}^{k-1} | x_{k-N+1})} \left( \prod_{l=k-N+1}^k P(y_l | x_l) \right) \left( \prod_{l=k-N+1}^{k-1} P(x_{l+1} | x_l) \right) \quad (1.42)$$

Il est courant de supposer que les densités de probabilités considérées sont gaussiennes, mais d'autres distributions ont été traitées [Var+22]. Dans le contexte gaussien, la maximisation de la log-vraisemblance de (eq. 1.42) mène à la formulation de la fonction coût (eq. 1.41) [Fie+20] où les matrices de pondération sont issues des densités de probabilité des perturbations affectant le système ( $R = \text{cov}(v)^{-1}$  et  $Q = \text{cov}(w)^{-1}$ ). Cette interprétation rejoint l'intuition que la contribution à la fonction coût des écarts entre les mesures prédites et acquises doit être corrélée à la confiance accordée aux mesures.

### 1.4.4 Retour sur le coût d'arrivée

Le coût d'arrivée contient deux termes : une estimation *a priori*  $\bar{x}$  de l'état au début de la fenêtre d'estimation et une pondération  $S_k$  sur l'écart à cette estimation *a priori*. L'estimation  $\bar{x}$  est en général issue des résolutions précédentes du problème d'optimisation. Au temps courant  $t$ , on produit l'estimation de l'état sur toute la fenêtre  $[t, t-H]$ , que l'on note  $\hat{x}_{[t, t-H]}^t$ . L'état à  $[t-H]$  est estimé plusieurs fois au cours du fonctionnement de l'algorithme qui produit  $\hat{x}_{[t-dt, t-dt-H]}^{t-dt}, \dots, \hat{x}_{[t-H, t-2H]}^{t-H}$ . N'importe quelle estimation précédente de  $x(t-H)$  pourrait convenir mais on distingue en général deux cas :

- $\bar{x} = \hat{x}^{t-H}(t-H)$  : coût d'arrivée dit « filtré »,
- $\bar{x} = \hat{x}^{t-dt}(t-H)$  : coût d'arrivée dit « lissé ».

Le coût filtré nécessite de stocker les solutions précédentes correspondant à la fenêtre d'estimation et mène à des performances légèrement inférieures au coût lissé [TR02].

Dans l'interprétation déterministe de la MHE, le choix  $S_k = sI_{nx}$  ( $s \in \mathbb{R}$ ) [Ale+10] garantit la stabilité de l'erreur d'estimation. Les auteurs de [Suw+16] donnent une méthode afin de déterminer  $s$  judicieusement. Dans l'approche stochastique, l'approche usuelle correspond à choisir  $S_k$  comme l'inverse de l'incertitude propagée par l'équation de Riccati d'un filtre de Kalman (étendu) [RRL01]. Plus généralement, il est envisageable d'utiliser différentes méthodes classiques d'estimation afin de déterminer la covariance en début d'horizon comme dans [Ung09] où il est proposé d'utiliser les filtres de Kalman étendu et sans parfum,

ainsi que le filtre à particules. Dans [SMG17], il est proposé une méthode récursive de calcul de la matrice de pondération. Le calcul du terme de coût d'arrivée peut se révéler délicat ou coûteux, notamment lors de l'utilisation du lisseur de Kalman, dans lequel les incertitudes sont propagées dans le sens des temps croissants sur la fenêtre d'estimation puis dans le sens des temps décroissants. Comme les calculs de propagation nécessitent des produits matriciels et inversions de matrices, la complexité numérique peut rapidement croître avec la taille du système<sup>a</sup>. Afin de pallier à cette difficulté, le calcul des incertitudes peut se faire par l'utilisation des sensibilités du problème d'optimisation de la MHE ; [ZLB08] montre que les incertitudes sont propagées implicitement par la simulation numérique et qu'elles peuvent être extraites de la Hessienne du Lagrangien du problème d'optimisation évalué en la solution du problème.

### 1.4.5 Propriétés

Il apparaît rapidement qu'en comparaison du filtre de Kalman (et variations), l'estimation à horizon glissant nécessite une grande puissance de calcul puisqu'il est nécessaire de résoudre en ligne un problème d'optimisation non-convexe (pour les systèmes non-linéaires). Heureusement, les performances des outils d'implémentation et de résolution de ces problèmes, ainsi que la puissance de calcul disponible se sont développées, de sorte que de nombreuses applications pratiques de la MHE ont pu voir le jour pour des modèles de complexité toujours croissante.

L'estimation par horizon glissant est intéressante pour les systèmes non-linéaires puisqu'à la différence des approches basées sur le filtre de Kalman, il n'est pas nécessaire de linéariser le système autour de l'estimation courante. Le filtre de Kalman étendu ou sans parfum peut ainsi échouer [HR05]. À l'opposé, les non-linéarités peuvent être exprimées explicitement dans le problème d'optimisation de la MHE.

De plus, les contraintes existant sur les états du modèle ne sont pas traitées *a priori* par le filtre de Kalman [Sim10], ce qui peut aboutir à des estimations aberrantes. Il est certes toujours possible de les vérifier *a posteriori* et de projeter les estimées dans l'ensemble admissible [SC02], mais l'information apportée par ces contraintes n'est alors pas correctement exploitée. En outre, l'expression des contraintes directement dans le problème d'optimisation permet de réduire effectivement la taille de l'espace de recherche.

Il est couramment avancé que la MHE est davantage robuste aux erreurs de modélisation ainsi qu'aux initialisations erronées que le filtre de Kalman [HR05]. Cependant, un modèle d'estimation représentatif du processus reste important pour la bonne convergence des estimées. Dans [Nej+23] il est proposé d'adapter le coût intermédiaire de la MHE en utilisant *le reinforcement learning* dans le but de palier aux erreurs du modèle. D'autres approches se penchent sur le modèle d'estimation du système et visent à en améliorer la précision par le biais des données disponibles sur le système. Dans [WLM23], des processus gaussiens sont utilisés afin d'apprendre le modèle à partir des données disponibles.

Un dernier avantage important de cet estimateur est sa capacité à prendre en compte des mesures retardées ou possédant des fréquences d'acquisition différentes. Puisque l'on

---

a. La complexité numérique pour le produit matriciel et l'inversion de matrice dépend de l'algorithme utilisé mais est typiquement  $O(n^\beta)$ ,  $2 \leq \beta \leq 3$ .

considère un horizon d'estimation qui considère un ensemble de mesures passées sur le système, il suffit de choisir un horizon plus long que les retards de mesures, quitte à ce que seul le début de la fenêtre comporte ces mesures.

D'un point de vue théorique, la MHE a fait l'objet de nombreux résultats intéressants qui cimentent l'intérêt de son utilisation en pratique. Sous hypothèse d'observabilité du problème et de perturbations bornées, la MHE est robustement globalement asymptotiquement stable (RGAS, pour *Robust Globally Asymptotically Stable*, i.e. la stabilité robuste asymptotique globale) [RMD17]. Pour un horizon d'estimation suffisamment long [Hu17] montre que pour des perturbations disparaissant, la détectabilité du système suffit pour la stabilité RGAS.

Ces résultats de stabilité sont néanmoins obtenus en supposant que la solution globale du problème d'estimation est trouvée à chaque résolution du problème. Une MHE sous-optimale est présentée dans [Ale+10], tandis que [Ala] présente une heuristique d'évitement des minima locaux basée sur la perturbation des matrices de pondération du problème.

Pour terminer, la MHE est également l'objet d'une littérature riche dans la communauté de l'assimilation de données, sous la dénomination de 4D-var (ou 3D-var dans le cas d'un horizon d'estimation d'un seul pas de temps du système discrétisé) [ABN16]. La MHE est ainsi une méthode couramment utilisée en météorologie et a prouvé ses performances sur des systèmes complexes [Cha+23].

## 1.5 Notes et discussion

### 1.5.1 Résolution des problèmes de contrôle optimal à temps continu

Le formalisme pour l'analyse de la MPC et la MHE est celui des systèmes discrétisés. Puisque la résolution numérique des problèmes d'optimisation sous-jacent n'est pas instantanée, MPC et MHE sont en pratique respectivement des contrôleur et estimateur à temps discret. Néanmoins, le formalisme à temps continu est plus naturel pour la modélisation des systèmes physiques, aboutissant généralement à un système d'équations (algébro-)différentiel. Considérons le problème de contrôle optimal (eq. 1.43). On parle également de problème de Mayer [Qui17] :

$$P(p) : \min_{v,x} V_I(x_{t_1}) + \int_{t_1}^{t_2} V(x, v, p) d\tau + V_F(x_{t_2}) \quad (1.43a)$$

$$\text{sous contraintes } \forall \tau \in [t_1, t_2] \quad (1.43b)$$

$$f(\dot{x}(\tau), x(\tau), z(\tau), u(\tau), w(\tau), p) = 0 \quad (1.43c)$$

$$x(t_1) = \bar{x} \quad (1.43d)$$

$$g(x, u) = 0 \quad (1.43e)$$

$$h(x, u) \leq 0 \quad (1.43f)$$

$V_I(x_{t_1})$  désigne un terme de coût initial (nul en MPC) et  $V_F(x_{t_2})$  désigne un terme de coût final.

L'équation (eq. 3.1d) représente la dynamique du système sous une forme implicite, où  $\dot{x}(\tau) \in \mathbb{R}^n$  est la dérivée de l'état différentiel  $x(\tau) \in \mathbb{R}^n$ ,  $u(t) \in \mathbb{R}^{n_u}$  est le vecteur des signaux

	MPC	MHE
Fonction coût	Critère de performance	Cohérence avec les mesures
Variable de décision	$x_{[t_1, t_2]}, u_{[t_1, t_2]}$	$x_{[t_1, t_2]}, w_{[t_1, t_2]}$
Signal de commande $v$	$u$ (endogène)	$w$ (exogène)
Paramètres $p$	$x(t_1)$	$y_{[t_1, t_2]}, u_{[t_1, t_2]}$

TABLE 1.1 – Dualité entre la commande prédictive et l'estimation à horizon glissant.

de commande,  $z(\tau) \in \mathbb{R}^{n_z}$  représente les variables algébriques du système,  $w(t) \in \mathbb{R}^{n_w}$  représente une perturbation (ou entrée exogène) et  $p \in \mathbb{R}^{n_p}$  représente des paramètres d'entrée du problème qui peuvent influencer le comportement de la solution.

Cette formulation couvre les problèmes de commande prédictive ainsi que les problèmes d'estimation. Pour un problème de commande, il s'agit de déterminer à partir d'un état initial  $x(t_1)$  la trajectoire du système  $x_{[t_1, t_2]}$  ainsi que les signaux de commande  $u_{[t_1, t_2]}$  correspondant, minimisant le critère du problème. Ce critère porte sur le suivi d'une trajectoire ou d'une référence ou encore des considérations d'économie d'effort de commande.

À contrario pour un problème d'estimation, les signaux de commande sont en général connus tandis que l'état  $x(t_1)$  est libre et doit être déterminé. On peut alors dans ce cas utiliser un modèle impliquant une perturbation  $w \in \mathbb{R}^{n_w}$ , qui permet de rendre compte à la fois des perturbations agissant sur le système et les erreurs de modélisation. On pourra alors employer un modèle de la forme  $\dot{x} = f(x, w, z, p)$  dans lequel  $p$  peut contenir la valeur des signaux de commande. Il s'agit de déterminer la trajectoire de l'état (en partie connue) qui correspond au mieux avec les mesures du système.

Il n'existe en règle générale pas de solution analytique à ce problème, ce qui appelle à sa résolution numérique. Les variables de décision de ce problème  $v$  et  $x$  sont des fonctions. L'espace de recherche de solutions est donc de dimension infinie. Il existe plusieurs approches pour la résolution numérique de (eq. 1.43) [Rao10] :

- La programmation dynamique produit des solutions en boucle fermée, mais nécessite de résoudre l'équation de Hamilton-Jacobi-Bellman (équation aux dérivées partielles), ce qui n'est envisageable que pour des systèmes de petite taille [AK21 ; OAT24 ; Amo+19].
- Les méthodes indirectes utilisent le principe du maximum de Pontryagin [Gam19] puis discrétisent la solution obtenue. Ces méthodes sont difficiles à mettre en œuvre en pratique. Elles ne permettent pas de prendre en compte les contraintes sur les entrée (en MPC) [TMT14].
- Les méthodes directes proposent de discrétiser les solutions du problème et de résoudre un programme non linéaire (NLP pour *Non Linear Program*), ayant un nombre fini de variables de décision.

Dans le contexte des méthodes directes, l'implémentation et la résolution numérique du problème nécessitent de transcrire le problème d'optimisation en un problème NLP. Cette étape est importante, puisque la nature du modèle de cœur de réacteur (temps continu, raideur du modèle) impose l'utilisation de méthodes adaptées, notamment pour la stabilité numérique du schéma d'intégration et la propagation de sensibilités. Les ouvrages [LY08 ; BV04] sont des références pour le traitement des algorithmes pour la résolution des problèmes NLP. Les notions de simulation numérique, de transcription en un problème NLP et la résolution de ces problèmes seront détaillées dans les chapitres suivants, mais nous

introduisons dès maintenant le contexte général, après que l'estimateur à horizon glissant ait été présenté.

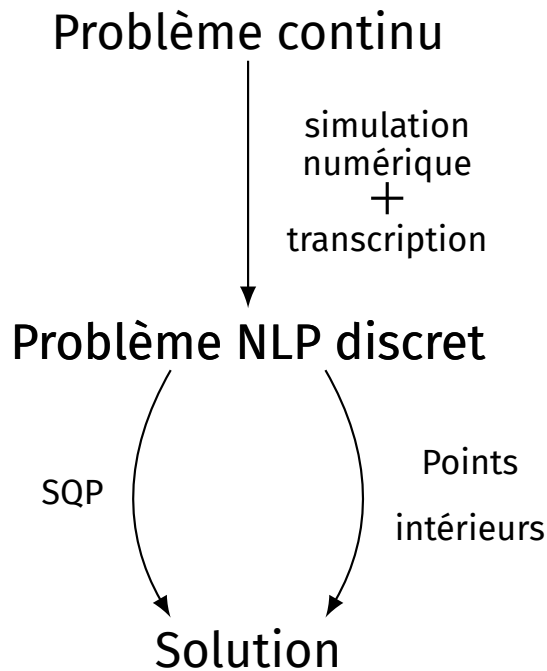


FIGURE 1.5 – Résolution d'un problème optimal à temps continu par une méthode directe.

### 1.5.2 Stratégies de réduction de la latence de calcul

La résolution du problème NLP nécessite un temps de calcul qui peut être non négligeable devant la fréquence de fonctionnement du contrôleur MPC ou de l'estimateur MHE. La solution du problème de contrôle optimal arrive avec un retard correspondant au temps de calcul. Le problème se pose pour la commande puisque ce retard impacte les performances de la régulation. D'autre part pour un estimateur impliqué dans un schéma de commande prédictive, le retard issu de l'estimation s'accumule à celui de la résolution du problème MPC.

Dans [LB12; Ale+10], une méthode de calcul à l'avance de la solution du problème d'estimation, via les sensibilités du problème d'optimisation, est proposée. Soit la solution  $w^*$  du problème NLP dépendant d'un paramètre  $p$  :

$$P(p) : \min_w F(w)$$

sous contraintes  $G(w)$   
 $H(w)$

La sensibilité du problème par rapport au paramètre  $p$  permet d'approximer la solution  $\bar{w}$  du problème  $P(\bar{p})$  en connaissant la solution  $w_0$  de  $P(p_0)$  de sorte que :

$$\bar{w} = w_0 + \frac{\partial w_0}{\partial p_0}(p - p_0) + O(|p - p_0|^2)$$

Le calcul des sensibilités du problème est implémenté dans CasADI [AR18] via le théorème des fonctions implicites appliqué aux conditions de KKT en la solution du problème.

Ce type d'approximation de la solution est particulièrement adaptée à la résolution des problèmes de contrôle optimal, dans la mesure où le problème résolu possède toujours la même structure et diffère uniquement en les données d'entrées. On peut en effet écrire un problème de commande optimale comme étant paramétrique en l'état initial de la fenêtre d'estimation (injection de la valeur initiale <sup>a</sup>) ou un problème d'estimation comme paramétrique en les signaux de commande et les mesures du système sur la fenêtre d'estimation.

Le choix de la méthode de résolution employée impacte le calcul des sensibilités. Notamment, une prédiction tangentielle de la solution déterminée par la méthode SQP est plus précise que pour la méthode des points intérieurs [DFH09]. Un algorithme de réduction de la latence de calcul par la méthodologie *advanced-step MHE* [LB12] est proposé au chapitre 7, où la solution du problème d'optimisation est obtenue par la méthode des points intérieurs mais les sensibilités de la solution aux paramètres d'entrée du problème sont obtenues par la méthode SQP.

---

a. « Initial value embedding » dans [Die01].

## CHAPITRE 2

# SIMULATION NUMÉRIQUE

La résolution des problèmes de contrôle optimal par les méthodes directes nécessite de simuler numériquement le système, en résolvant les équations qui régissent son comportement, ainsi que le calcul des sensibilités de l'intégrateur par rapport aux conditions initiales.

### 2.1 Modèles dynamiques

Modéliser un système physique mène à la formulation d'équations qui décrivent son comportement, son évolution dans le temps, sa géométrie... Le modèle résultant est un système dynamique (au sens mathématique) qui peut-être à temps continu pour une évolution décrite par des équations algébriques et différentielles (ordinaires ou aux dérivées partielles) ou à temps discret dont l'état en certains instants suit une équation aux différences. Les systèmes hybrides sont des systèmes pour lesquels certains phénomènes sont modélisés à temps continu et d'autres à temps discret. On s'intéressera spécifiquement aux modèles dynamiques à temps continu.

L'état du système  $X \in \mathbb{R}^{n_x}$  correspond à l'ensemble variables du modèle susceptibles de résumer ses évolutions passées. Ces variables d'état évolent en fonction des variables de temps, voire d'espace, mais aussi de paramètres exogènes (signaux de commande, perturbations, ...). On note  $X(t, \pi_1, \dots, \pi_n)$  pour dénoter cette dépendance, les variables et signaux exogènes étant concaténés dans le vecteur  $\pi = (t, \pi_1, \dots, \pi_n) \in \mathbb{R}^{n+1}$ . Une équation aux dérivées partielles (EDP) régissant l'évolution du modèle s'écrit alors :

$$f(\pi, X, \partial_\pi X, \dots, \partial_\pi^m X) = 0 \quad (2.1)$$

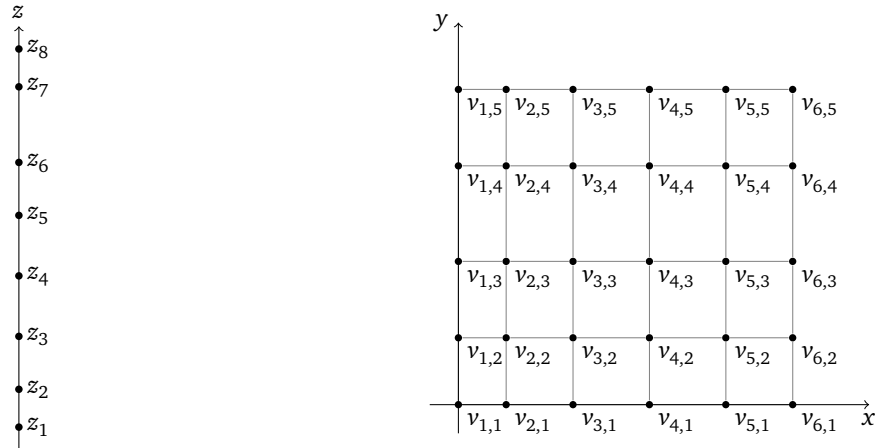
faisant apparaître les dérivées partielles de  $X$  jusqu'à l'ordre  $m > 0$ . L'équation de Boltzmann, l'équation de la diffusion ou l'équation de transport sont des équations aux dérivées partielles bien connues, qui apparaissent dans de nombreux problèmes.

La résolution analytique des EDP non triviales est difficile voire impossible. La résolution numérique de ces équations implique d'approximer la solution en certains points de l'espace et instants dans le temps. Il est courant de discrétiser la solution d'abord dans l'espace (semi-discrétisation en espace) en considérant un « maillage » de l'espace [PM96]. Le maillage peut-être adaptatif, mais considérons ici un maillage fixé (voir figure 2.1). On introduit  $N > 0$  « nœuds » dans l'espace  $\mathbb{R}^n$ . Les dérivées partielles de l'état par rapport à l'espace sont ensuite remplacées en considérant une approximation. Soit une EDP à une dimension spatiale notée  $z$  telle que :

$$f(t, z, X, \partial_z X, \partial_t X, \dots) = 0$$

Pour un maillage de l'espace  $(z_i)_{i \in \llbracket 1, N \rrbracket}$  on peut approximer (en supposant le maillage





(a) Maillage à une dimension spatiale.

(b) Maillage à deux dimensions spatiales.

FIGURE 2.1 – Maillage de l'espace pour la discrétisation spatiale des EDP. Chaque dimension est discrétisée individuellement décrivant un maillage de l'espace, qui n'est pas nécessairement uniforme.

suffisamment fin) la quantité  $\partial_z X$  par l'approximation aux différences finies suivante :

$$\forall t \in \mathbb{R}, \partial_z X(t, z_i) \approx \frac{X(t, z_{i+1}) - X(t, z_i)}{z_{i+1} - z_i} \quad (2.2)$$

De nombreuses autres approximations sont possibles, par exemple en faisant intervenir davantage de mailles adjacentes dans l'espace de la solution. Une construction systématique de méthodes d'approximation par les différences finies de la forme :

$$\partial_z^k X(t, z_i) \approx \sum_{l=-n_1}^{n_2} a_l X(t, z_{i+l})$$

est possible en considérant un développement de Taylor à un ordre plus élevé [For88]. Le cas  $n_1 = n_2$  correspond à une formule symétrique, mais l'approximation peut aussi être asymétrique (uni-latérale). Pour les mailles extérieures au domaine de l'espace discrétisées, il est nécessaire d'adapter ces formules. Pour ces mailles, on peut construire des mailles extérieures artificielles par extrapolation puis appliquer une méthode symétrique, ou alors employer une méthode asymétrique pour exploiter uniquement les mailles intérieures au domaine.

Les valeurs prises par les quantités d'intérêt aux nœuds du maillage correspondent à autant de nouveaux états liés par des équations différentielles *ordinaires*, plus simples à traiter que les équations aux dérivées partielles. Une **équation différentielle ordinaire** correspond à un cas particulier de (eq. 2.1) où seule la dérivée temporelle des variables d'état apparaît, c'est-à-dire en notant  $\partial_t X = \dot{X}$  :

$$f(t, X, \dot{X}) = 0$$

On rencontre souvent des expressions où un terme de forçage apparaît, comme par exemple un terme de commande qui permet influencer le comportement du système.

Supposons le signal de commande noté <sup>a</sup>  $u(t) \in \mathbb{R}^{n_u}$  on obtient l'équation différentielle suivante :

$$f(t, X, \dot{X}, u) = 0$$

Cette notation est dite implicite <sup>b</sup>, mais on rencontre très souvent la notation explicite  $\dot{X} = f_{\text{expl}}(t, X, u)$  lorsqu'il est possible d'exprimer directement la dérivée temporelle de l'état en fonction de ce dernier et du signal de commande. On peut alors revenir à une forme implicite par la réécriture :

$$f_{\text{impl}}(t, X, \dot{X}, u) = \dot{X} - f_{\text{expl}}(t, X, u) = 0$$

Dans le cas où la matrice Jacobienne  $\partial_{\dot{X}} f_{\text{impl}}$  est singulière, le modèle est dit algébro-différentiel (présence de relations strictement algébriques entre les variables). La totalité de l'état n'est alors pas décrit par une équation différentielle explicite et on distingue alors les parties algébriques et différentielles de l'état en les notant  $z$  et  $x$ . On aboutit à une représentation semi-explicite qui s'exprime :

$$\dot{x} = f_{\text{expl}}(t, x, z, u) \quad (2.3)$$

$$0 = g(t, x, z, u) \quad (2.4)$$

Un système linéaire est un système pour lequel l'équation  $f_{\text{impl}}$  est linéaire. Pour un système différentiel linéaire, on écrit :

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) \quad (2.5)$$

Un système pour lequel  $\forall t \in \mathbb{R}, A(t) = A, B(t) = B$ , est rang dans la classe des systèmes Linéaires à Temps Invariant (LTI).

### Définition 2.1

*En notant  $\underline{\sigma}(A)$  et  $\overline{\sigma}(A)$  respectivement la plus petite et la plus valeur grande propre de  $A$ , le système (eq. 2.5) est dit raide si  $\underline{\sigma}(A) \ll \overline{\sigma}(A)$ .*

La raideur d'un modèle le rend difficile à simuler numériquement <sup>c</sup> et implique d'employer des méthodes d'intégration adaptées afin d'assurer la stabilité numérique de la solution. Pour un modèle non linéaire, la raideur est plus délicate à appréhender. Elle est liée aux valeurs propres à partie réelle négatives de la Jacobienne  $\partial_x f$  [HNW09].

a. On doit cette notation à Lev Semenovich Pontryagin, le mot russe « управление » correspondant à « commande » [Gam19].

b. On parle également de système descripteur.

c. La grande différence entre les valeurs propres implique des dynamiques de vitesse très différentes. La simulation des dynamiques rapides avec une méthode inadaptée implique un pas de temps d'intégration excessivement petit qui peut impacter négativement les performances de l'intégrateur numérique.

## 2.2 Intégration numérique

Afin de simplifier la notation, on considère un système différentiel autonome de la forme :

$$\xi'(t) = f(t, \xi(t))$$

Cette équation différentielle a pour inconnue la fonction  $\xi : \mathbb{R} \rightarrow \mathbb{R}^n$ ,  $n \geq 1$ . Il existe une infinité de solution à ce problème [Dem06]. Le problème de la valeur initiale, ou problème de Cauchy (eq. 2.6), consiste à trouver la solution d'une équation différentielle comportant une condition initiale :

$$\begin{cases} \xi'(t) = f(t, \xi(t)) \\ \xi(t_0) = \xi_0 \end{cases} \quad (2.6)$$

La fonction d'évolution  $f$  peut être scalaire  $f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  ou vectorielle  $f : \mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $n > 1$ . En cas d'existence de solution (théorème de Cauchy-Lipchitz), la résolution analytique du problème de Cauchy pour des fonctions d'évolution non-triviales est difficile, là où les méthodes de résolution numérique permettent de trouver des solutions approchées de manière systématique. Résoudre numériquement une équation différentielle consiste à trouver la valeur prise par la solution en certains points  $(t_k)_{k \in \mathbb{N}}$ . Le maillage ainsi défini est représenté sur la figure 2.2. Ce maillage définit également des pas d'intégration, qu'on écrit  $h_k = t_{k+1} - t_k$ . Le choix du pas d'intégration est important pour la stabilité numérique du schéma d'intégration en fonction du problème à résoudre. La discrétisation du temps peut-être adaptative, c'est-à-dire que le maillage est déterminé au cours de l'intégration numérique en fonction de l'estimation de l'erreur d'intégration réalisée. La section 2.3 revient plus en détail sur les méthodes d'intégration adaptatives.

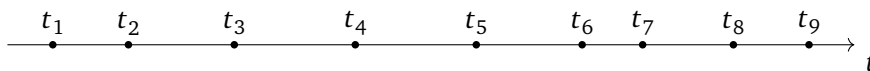


FIGURE 2.2 – Discrétisation du temps : la solution est approximée en ces instants.

La solution approchée  $(\xi_k)_{k \in \mathbb{N}}$  est ainsi une suite numérique censée coïncider avec la solution exacte en ces points, de sorte que :

$$\forall k \in \mathbb{N}, \xi_k \approx \xi(t_k)$$

Dans la suite, on notera  $\xi$  la solution exacte et  $(\xi_k)_{k \in \mathbb{N}}$  la solution approchée. La solution exacte est une fonction et la solution numérique est une suite. On appellera plus généralement « intégrateur », un algorithme mettant en place une méthode d'intégration numérique résolvant le problème de Cauchy (eq. 2.6). Pour une méthode d'intégration numérique on s'intéressera à plusieurs propriétés qui sont l'ordre de cette méthode ainsi que sa *stabilité numérique*.

On distingue pour l'intégration numérique deux classes de méthodes, dites méthodes à un pas pour l'une et méthodes multipas pour l'autre. On note  $h_k = t_{k+1} - t_k$  la longueur du pas d'intégration. Les méthodes à un pas n'utilisent que la solution numérique au pas précédent  $\xi_k$  pour déterminer  $\xi_{k+1}$  de sorte que la solution numérique à l'équation différentielle est de la forme :

$$\xi_{k+1} = \xi_k + h_k \Phi(h_k, t_k, \xi_k) \quad (2.7)$$

tandis que les méthodes d'intégration multipas réutilisent les solutions précédentes pour calculer  $\xi_{k+1}$ . Une méthode à  $n$  pas s'écrit :

$$\xi_{k+1} = \sum_{l=0}^{n-1} \alpha_l \xi_{k-l} + \sum_{l_0}^{n-1} h_l \beta_{l_0} f(t_{k+l}, \xi_{k+l}) \quad (2.8)$$

Il est nécessaire d'initialiser correctement les méthodes d'intégration multipas au début de l'intégration numérique par exemple en utilisant une méthode à un pas.

Une première question se pose pour toute méthode d'intégration numérique : est-elle une bonne approximation de la solution au problème (eq. 2.6) ?

**Définition 2.2 (Erreur de consistance)**

L'erreur de consistance  $\varepsilon_k$  d'une méthode d'intégration numérique associée au problème (eq. 2.6) est la quantité :

$$\varepsilon_k = \xi(t_{k+1}) - \xi_{k+1}.$$

où  $\xi_{k+1}$  est obtenu par intégration numérique à partir de  $\xi(t_k)$ .

L'erreur de consistance correspond à l'erreur réalisée localement en utilisant la méthode une fois à partir de la solution exacte. Lorsque l'on utilise la méthode d'intégration plusieurs fois sur tout l'intervalle d'intégration, en repartant de la solution précédente, l'erreur d'intégration s'accumule.

**Définition 2.3 (Erreur globale)**

L'intégration numérique sur l'intervalle  $[t_0, t_f]$  subdivisé en  $N$  sous-intervalles engendre l'erreur globale d'intégration :

$$\varepsilon_{globale} = \max_{0 < n < N} |\xi(t_n) - \xi_n|$$

On la définit parfois également par l'erreur d'intégration à la fin de l'intervalle :

$$\varepsilon'_{globale} = |\xi(t_N) - \xi_N|$$

Une méthode est dite consistante lorsque  $\sum_{k=0}^N |\varepsilon_k| \xrightarrow{h \rightarrow 0} 0$ . L'ordre d'une méthode d'intégration permet de quantifier la vitesse de convergence de l'erreur de consistance en fonction du pas d'intégration.

**Définition 2.4 (Ordre d'une méthode d'intégration)**

L'ordre d'une méthode d'intégration numérique est le plus grand entier  $p$  tel que  $\forall k \in \mathbb{N}, \varepsilon_k = O(h^{p+1})$ .

Il est souhaitable qu'une méthode d'intégration numérique ait un ordre élevé afin de limiter l'erreur commise sur l'approximation numérique de la solution. L'ordre d'une méthode est en réalité l'ordre maximal possible pour un problème donné.

Une méthode d'intégration numérique doit également conserver certaines propriétés du système que l'on intègre. On étudie en général la méthode sur le système test de Dahlquist  $f(t, \xi(t)) = l\xi(t)$ ,  $l \in \mathbb{C}$  de solution analytique  $\xi(t) = e^{lt}$ . Il est clair que pour  $\Re(l) < 0$ ,  $\xi(t) \xrightarrow{t \rightarrow \infty} 0$ . Supposons que l'intervalle d'intégration  $[t_0, t_f]$  soit divisé en sous-intervalles

de tailles égales de sorte que le pas d'intégration soit constant et vaille  $h$ . Une méthode à un pas prend la forme d'une suite géométrique de raison  $\phi(hl)$  telle que  $\xi_{k+1} = \phi(hl)\xi_k$ . En notant  $z = hl$ , il est clair que cette suite est convergente (comme la solution réelle) lorsque  $|\phi(hl)| < 1$ .

**Définition 2.5 (Fonction de sensibilité et domaine de convergence)**

La fonction de sensibilité d'une méthode d'intégration est la fonction :

$$\phi : \mathbb{C} \rightarrow \mathbb{C} \quad (2.9a)$$

$$z \mapsto \phi(z) \quad (2.9b)$$

Cette fonction apparaît pour les méthodes d'intégration à un pas qui prennent, sous l'hypothèse d'un pas d'intégration constant, la forme d'une suite géométrique  $\xi_{k+1} = \phi(hl)\xi_k$ . Le domaine de convergence pour une méthode d'intégration est le domaine du plan complexe pour lequel  $|\phi(z)| < 1$ .

La stabilité de la solution en fonction de  $l$  peut dépendre du choix de  $h$ . Une méthode dont la solution est stable indépendamment de la taille du pas choisi est souhaitable, c'est par exemple le cas d'une méthode A-stable.

**Définition 2.6 (A-stabilité)**

Une méthode A-stable est une méthode pour laquelle le domaine de convergence contient tout le demi-plan gauche du plan complexe. Une méthode  $A(\alpha)$ -stable est une méthode pour laquelle le domaine de convergence contient un cône d'angle  $\alpha$ .

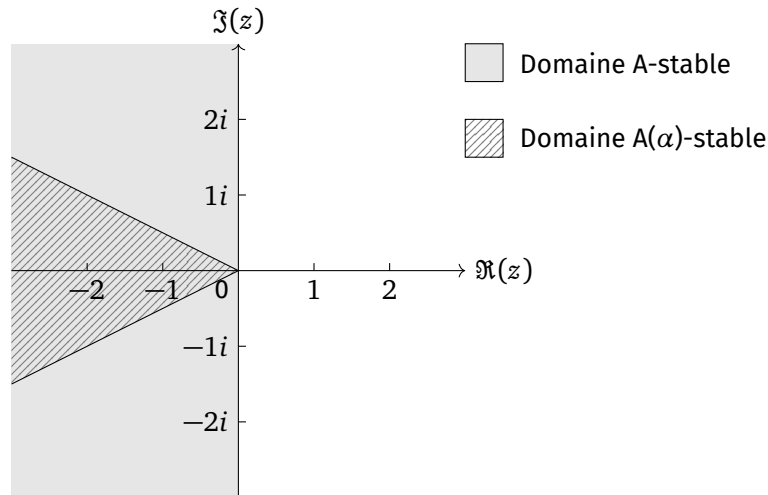


FIGURE 2.3 – Domaine de stabilité d'une méthode A-stable.

La notion d'A-stabilité n'est pas suffisante pour les systèmes raides, qui présentent des dynamiques aux temps caractéristiques très différents, de sorte que le comportement transitoire de la solution numérique peut ne pas correspondre à la solution exacte, bien que les deux convergent [HW96].

**Définition 2.7 (L-stabilité)**

Une méthode d'intégration est L-stable si elle est A-stable et si  $\lim_{|z| \rightarrow \infty} \phi(z) = 0$ .

### 2.3 Variation du pas de temps, contrôle de l'erreur

Les intégrateurs numériques implémentent une méthode de contrôle de l'erreur via un pas d'intégration variable, de sorte à limiter l'erreur d'intégration. L'enjeu est double : d'une part le choix d'un pas d'intégration fixe « trop petit » implique un grand nombre d'évaluations de la méthode d'intégration (donc un coût de calcul important), d'autant que l'on risque d'accumuler des erreurs d'arrondis (représentation numérique). D'autre part, le choix d'un pas d'intégration fixe « trop grand » implique une erreur d'intégration plus élevée. En ajustant la taille du pas d'intégration en ligne, on garantit à la fois la précision et la vitesse du schéma d'intégration.

Lors de la mise en œuvre d'un schéma d'intégration adaptatif, il s'agit d'estimer en ligne l'erreur d'intégration commise en réalisant le calcul de  $\xi_{k+1}$  de deux manières différentes. L'utilisation de la méthode de Richardson [Ric11] consiste à utiliser une méthode d'intégration deux fois avec le pas  $h$  pour obtenir  $\bar{\xi}_{k+1}$  et une fois avec le pas  $2h$  pour obtenir  $\tilde{\xi}_{k+1}$ . L'erreur commise est alors estimée avec la formule :

$$\varepsilon = \xi(t_k + 1) - \bar{\xi}_{k+1} = \frac{\bar{\xi}_{k+1} - \tilde{\xi}_{k+1}}{2^p - 1} + O(h^{p+2})$$

où  $p$  est l'ordre de la méthode d'intégration. On peut alors utiliser le résultat corrigé de l'erreur :

$$\hat{\xi}_{k+1} = \bar{\xi}_{k+1} + \frac{\bar{\xi}_{k+1} - \tilde{\xi}_{k+1}}{2^p - 1}$$

qui est une estimation d'ordre  $p+1$  (donc meilleure que  $\bar{\xi}_{k+1}$  et  $\tilde{\xi}_{k+1}$  obtenues avec des méthodes d'ordre  $p$ ).

Les méthodes de Runge-Kutta-Fehlberg (RKF) et Dormand-Prince sont des exemples de méthodes d'intégration avec contrôle de l'erreur via l'emploi de méthodes d'ordre différent. Afin de limiter la complexité de calcul, ces méthodes sont construites de sorte à ce que les mêmes évaluation de  $f$  puissent être utilisées dans les deux schémas d'intégration (on parle de méthodes emboîtées [Dem06]).

Plus précisément, l'erreur de consistance (définition théorème 2.2) est estimée en appliquant une méthode d'intégration donnée à deux ordres différents (par exemple RK aux ordres 4 et 5), donnant deux valeurs de  $\xi_{k+1}$  utilisées pour estimer l'erreur. Un algorithme de d'adaptation de la taille du pas d'intégration peut décider, en cas de rejet du pas courant, d'augmenter ou de diminuer la taille du pas afin de garantir que l'erreur d'intégration reste inférieure à un seuil spécifié par l'utilisateur. L'approche la plus simple consiste à diviser ou multiplier d'un facteur 2 le pas d'intégration, mais une approche plus élaborée peut être plus efficace<sup>a</sup>.

a. Le *contrôle* de l'erreur peut être assuré par un régulateur PI ou PID et il est possible d'estimer l'erreur d'intégration en construisant un observateur [Söd02].

## 2.4 Sensibilité de l'intégrateur

L'usage de l'intégration numérique dans le cadre de la résolution de problèmes de contrôle optimal nécessite, en plus de la simulation du système, de calculer les sensibilités (au premier ordre voire au delà) de la solution par rapport aux conditions initiales et aux paramètres (paramètres du modèle et/ou signaux de commande) du système au cours de la simulation. Ce calcul fait intervenir le calcul de quantités dérivées, qui peut être réalisé de différentes manières, ainsi que décrit plus en détail en section 4.7.

### Définition 2.8 (Sensibilités de la solution du problème de Cauchy [HNW09])

Soit  $\xi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^n$  la solution du problème de Cauchy dépendant d'un paramètre  $p$  :

$$\xi' = f(t, \xi(t, \xi_0, p)) \quad (2.10a)$$

$$\xi(t_0, \xi_0, p) = \xi_0 \quad (2.10b)$$

On note que la solution dépend de  $t$  et  $p$ , ainsi que de la condition initiale  $\xi_0$  de sorte que  $\xi(t, \xi_0, p)$ . La solution numérique, à la fin de l'intervalle d'intégration est notée  $\xi_N(\xi_0, p)$ . On appelle fonctions de sensibilité de la solution, les dérivées de cette dernière par rapport à  $\xi_0$  et  $p$  :

$$S_p = \frac{\partial \xi_N(t, \xi_0, p)}{\partial p}, \quad S_{\xi_0} = \frac{\partial \xi_N(t, \xi_0, p)}{\partial \xi_0}.$$

On s'intéresse en particulier aux valeurs prises par ces sensibilités à la fin de l'intervalle d'intégration, soit  $S_p(t_f)$  et  $S_{\xi_0}(t_f)$ .

Les sensibilités mesurent l'impact des variations des conditions initiales et des paramètres du problème sur la trajectoire du système. Ces quantités sont donc importantes pour le calcul de la direction de descente dans la méthode de Newton associée à la résolution du problème NLP résolvant le problème de contrôle optimal. Pour les méthodes directes, il faut alors calculer la sensibilité de la solution numérique par rapport à ces variables. On distingue plusieurs classes de méthodes qui permettent ce calcul. Une première approche, qualifiée de END (External Numerical Differentiation) consiste à considérer une dérivée numérique par le biais d'une différence finie, soit en définissant les directions de perturbation  $\delta p$  et  $\delta \xi_0$  (dont toutes les composantes sont nulles sauf une) et la précision machine  $\varepsilon$  :

$$S_p \approx \frac{\xi_{t_f}(\xi_0, p + \varepsilon \delta p) - \xi_{t_f}(\xi_0, p)}{\varepsilon}, \quad S_{\xi_0} \approx \frac{\xi_{t_f}(\xi_0 + \varepsilon \delta \xi, p) - \xi_{t_f}(\xi_0, p)}{\varepsilon}.$$

Cette approche est néanmoins particulièrement inefficace, en particulier pour la sensibilité par rapport à la condition initiale d'un système de grande dimension où il est nécessaire de perturber une à une chaque composante de la condition initiale et simuler le système à nouveau. Chaque simulation numérique devra alors déterminer un nouveau maillage adaptatif de l'intervalle d'intégration, au prix d'un coût calculatoire élevé. En outre, la précision de cette méthode est limitée puisque  $\varepsilon$  est nécessairement plus élevée que la précision machine. Il faut également que la trajectoire soit suffisamment perturbée pour que la différence avec la solution nominale excède l'erreur d'intégration [RMD17], ce qui n'est pas avantageux pour les systèmes raides, potentiellement sensibles à de petites variations des paramètres. Les intégrateurs capables de calculer la sensibilité de leur solution par

rapport à leurs entrées implémentent donc des techniques différentes.

En opposition, la méthode « interne » (IND Internal Numerical Differentiation) ou plus précisément du calcul « des dérivées du schéma de discrétisation sélectionné en interne » [Boc81]<sup>a</sup> consiste à calculer les sensibilités en « figeant » ou « gelant » les éléments adaptatifs du schéma d'intégration (ordre de la méthode, taille du pas d'intégration, itérations de Newton pour une méthode implicite) et en différenciant directement le schéma d'estimation utilisé (soit par les différences finies, soit par différentiation automatique).

Les méthodes que nous venons de décrire font partie des méthodes dites « discrétiser puis différentier » car la trajectoire est simulée, avant le calcul des sensibilités de la solution numérique. Une autre classe de méthodes existe, dénommée « différentier puis discrétiser », qui présentent l'intérêt de déterminer les sensibilités de la solution exacte à temps continu plutôt que pour la solution numérique [Alb11]. En différenciant d'abord le problème de Cauchy par rapport aux paramètres ou conditions initiales, on détermine une équation différentielle auxiliaire<sup>b</sup>. On calcule :

$$\partial_p \dot{\xi} = \partial_p f + \partial_{\xi} f \partial_p \xi.$$

Par ailleurs, on a :

$$\partial_p \dot{\xi} = \frac{d}{dt} \partial_p \xi.$$

Ainsi, la fonction de sensibilité par rapport au paramètre  $p$  est donc solution de l'équation différentielle suivante :

$$\dot{S}_p = \partial_p f + S \partial_{\xi} f$$

On peut ainsi obtenir la sensibilité de la solution à la fin de l'intervalle d'intégration en intégrant une équation différentielle auxiliaire dont elle est solution. Il n'est pas nécessaire d'employer la même méthode d'intégration que pour la résolution du problème de Cauchy.

## 2.5 Méthode d'Euler

La méthode d'Euler [Dem06] est une méthode d'intégration numérique bien connue. Simple à comprendre et à implémenter, elle permet de poser les bases des méthodes d'intégration numérique puisqu'elle permet de comprendre les méthodes à un pas, mais également les méthodes multi-pas<sup>c</sup>, ainsi que la distinction entre les méthodes explicites et implicites. En confondant localement la trajectoire sur  $[t_k, t_{k+1}]$  à sa tangente en  $(t_k, \xi_k)$ , on approxime la dérivée de la trajectoire par une différence finie de sorte que :  $\xi'(t_k) \approx \frac{\xi(t_{k+1}) - \xi(t_k)}{t_{k+1} - t_k}$  ce qui mène au schéma numérique de la méthode d'Euler explicite :

$$\xi_{k+1} = \xi_k + f(t, \xi_k) \times \Delta t$$

appliqué à partir de la condition initiale connue  $\xi_0$ . La méthode est dite explicite en ce que la solution numérique à  $t_{k+1}$  ne dépend que de la solution à l'instant  $t_k$ .

a. « derivatives of the internally selected discretisation scheme » dans [Boc81].

b. La différentiation automatique peut être utilisée dans ce cas également.

c. On peut imaginer la méthode d'Euler comme une méthode multi-pas... à un pas.



L'approximation de la tangente de la trajectoire en  $t_{k+1}$  à partir de la solution précédente (Euler à droite) donne la méthode d'**Euler implicite** :

$$\xi_{k+1} = \xi_k + f(t, \xi_{k+1}) \times \Delta t$$

Cette méthode est dite implicite puisque le terme  $\xi_{k+1}$  est présent dans le terme de droite et dans le terme de gauche. On ne dispose alors plus d'une formule directe pour le calcul de la solution numérique au problème de Cauchy, mais il faut à la place résoudre une équation par le biais de la méthode de Newton.

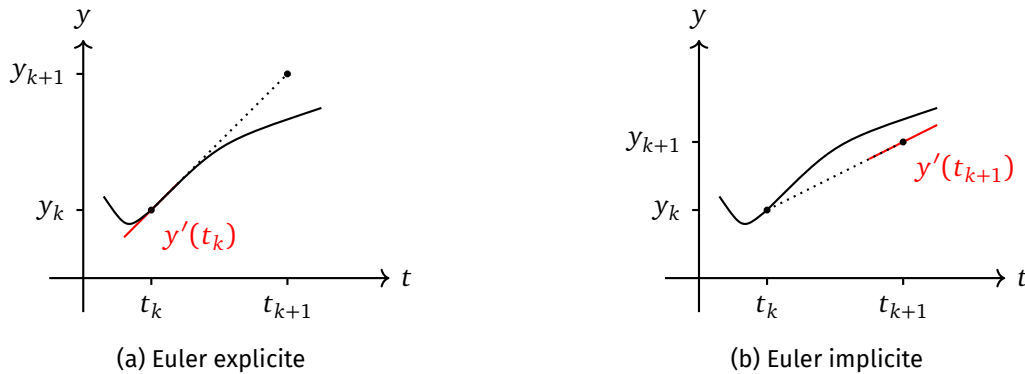


FIGURE 2.4 – Méthodes d'Euler

Étudions la stabilité numérique de la solution. On considère à nouveau le système test de Dahlquist  $\xi(t)' = l\xi(t)$  de solution  $\xi(t) = e^{lt}$ . L'utilisation de la méthode de Newton explicite donne :

$$\xi_{k+1} = \xi_k + hl\xi_k = (1 + hl)\xi_k$$

Il s'agit d'une suite géométrique de raison  $1 + hl$ , qui n'est convergente que lorsque  $|1 + hl| < 1$ . La solution du système test n'est stable que pour  $\Re(l) < 0$  mais on remarque alors que qu'il faut nécessairement  $h < 2/|l|$  pour que l'erreur converge vers 0.

La méthode d'Euler implicite mène à une suite géométrique de raison  $\frac{1}{1-hl}$  or  $\forall h > 0, \forall l \in \mathbb{C}, \Re(l) < 0, |\frac{1}{1-hl}| < 1$ , ainsi la méthode est stable pour toute valeur de  $h$  (elle est donc A-stable) et elle est même L-stable puisque  $\frac{1}{1-hl} \xrightarrow{h \rightarrow \infty} 0$ .

La méthode d'Euler (explicite ou implicite) est d'ordre 1. En d'autre terme, il est nécessaire d'utiliser un pas d'intégration fin afin de limiter l'erreur d'intégration ce qui induit de nombreuses itérations. Les méthodes d'intégration à un pas consistent à considérer une « direction moyenne » de la trajectoire pendant un temps  $h_k$  (le pas d'intégration) à partir de la solution numérique à  $t_k$ . En construisant une meilleure direction moyenne, notamment en considérant des évaluations de  $\xi'$  en  $t_k, t_{k+1}$  et/ou  $t \in ]t_k, t_{k+1}[$ , il est possible de concevoir une méthode d'intégration d'ordre plus élevé.

La  **$\theta$ -méthode** est un schéma numérique d'intégration de la forme :

$$\xi_{k+1} = \xi_k + h(1 - \theta)f(t_k, \xi_k) + h\theta f(t_{k+1}, \xi_{k+1}), \theta \in [0, 1] \quad (2.11)$$

Le cas  $\theta = 0$  correspond à la méthode d'Euler explicite tandis que le cas  $\theta = 1$  est la méthode d'Euler implicite. La méthode est implicite si  $\theta > 0$ . Le cas  $\theta = 1/2$  correspond à la **méthode des trapèzes**, qui est implicite d'ordre 2 et A-stable. Comme pour la méthode

d'Euler implicite, on résout (eq. 2.11) avec la méthode de Newton, que l'on peut initialiser à partir de la méthode d'Euler explicite avec  $\tilde{\xi}_{k+1} = \xi_k + hf(t_k, \xi_k)$ . Cela revient à faire un demi-pas dans la direction  $\xi'(t_k)$  puis un demi-pas dans la direction  $\xi'(t_{k+1})$ . Si on se contente d'évaluer (eq. 2.11) avec  $\tilde{\xi}_{k+1}$ , on arrive à la méthode **prédicteur-correcteur**, ou encore méthode de Heun [RMD17], qui correspond au schéma numérique suivant :

$$\begin{aligned} \tilde{\xi}_{k+1} &= \xi_k + hf(t_k, \xi_k) \\ \xi_{k+1} &= \xi_k + \frac{h}{2}(f(t_k, \xi_k) + f(t_{k+1}, \tilde{\xi}_{k+1})) \end{aligned}$$

Cette méthode explicite est également connue en tant que méthode de Runge Kutta d'ordre 2 (RK2). La  $\theta$ -méthode utilise uniquement des évaluations de  $\xi'$  en  $t_k$  et/ou  $t_{k+1}$  (ou leur approximation dans le cadre de la méthode prédicteur-correcteur) mais on peut également introduire des points intermédiaires sur l'horizon afin d'obtenir une estimation plus fine de la trajectoire sur  $[t_k, t_{k+1}]$ .

La méthode du **point milieu** correspond à la méthode d'Euler en considérant  $\xi'(t_k + \frac{h}{2})$ , c'est-à-dire :  $\xi_{k+1} = \xi_k + hf(t_k + \frac{h}{2}, \xi_{k+1/2})$  On peut estimer  $\xi_{k+1/2}$  avec la méthode d'Euler explicite pour un pas  $h/2$ , soit  $\xi_{k+1/2} = \xi_k + \frac{h}{2}f(t_k, \xi_k)$ . Les méthodes de Runge-Kutta présentées au paragraphe suivant permettent de construire des méthodes d'intégration d'ordre plus élevé en prenant en compte les considérations que nous venons d'énoncer concernant l'intégration à un pas.

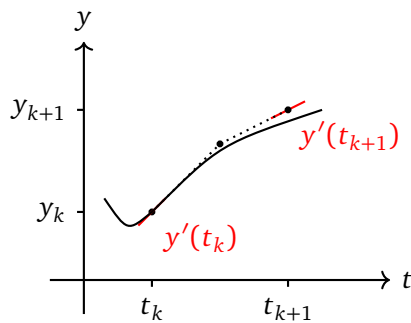


FIGURE 2.5 – Méthode des trapèzes

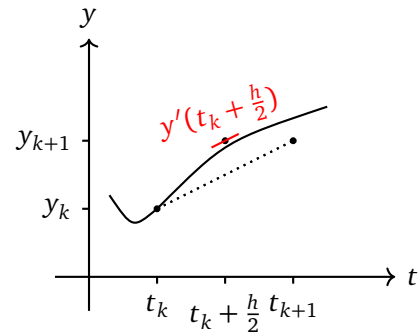


FIGURE 2.6 – Méthode du point milieu

## 2.6 Méthodes de Runge-Kutta

Les méthodes de Runge-Kutta introduisent des points intermédiaires sur l'intervalle  $[t_k, t_{k+1}]$ . Ces points, que l'on appellera nœuds de Runge-Kutta, sont définis comme  $\forall i \in \llbracket 1, s \rrbracket, t_{k,i} = t_k + c_i h$  où  $s$  est le nombre de points introduits,  $0 \leq c_1 \leq \dots \leq c_s \leq 1$ . On définit ainsi  $s$  valeurs de la dérivée de la solution aux nœuds de Runge-Kutta.

$$\begin{cases} \kappa_1 = f(t + c_1 h, \xi_k + h(a_{11}\kappa_1 + \dots + a_{1s}\kappa_s)) \\ \vdots \\ \kappa_s = f(t + c_s h, \xi_k + h(a_{s1}\kappa_1 + \dots + a_{ss}\kappa_s)) \\ \Phi(h, t_k, \xi_k) = h(b_1\kappa_1 + \dots + b_s\kappa_s) \end{cases} \quad (2.12)$$

$c_1$	$a_{11}$	$\dots$	$a_{s1}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$c_s$	$a_{s1}$	$\dots$	$a_{ss}$
	$b_1$	$\dots$	$b_s$

TABLE 2.1 – Tableau de Butcher.

On construit ainsi une approximation de la solution sur une subdivision plus fine que la suite des  $(t_k)_{k \in \mathbb{N}}$ , de sorte que :

$$\begin{aligned} \xi(t_{k,i}) &= \xi_k + \int_{t_k}^{t_{k,i}} \dot{\xi}(\tau) d\tau & \xi(t_{k+1}) &= \xi_k + \int_{t_k}^{t_{k+1}} \dot{\xi}(\tau) d\tau \\ &\approx \xi_k + h \sum_{l=1}^s a_{il} \kappa_l & & \approx \xi_k + h \sum_{i=1}^s b_i \kappa_i. \end{aligned}$$

et

Les termes intégraux apparaissant dans les formules précédentes sont approximés par des formules de quadrature qui utilisent les approximations  $k_i$ . On résume couramment les calculs présentés précédemment sous la forme d'un *tableau de Butcher* (tableau 2.1). On distingue alors deux cas de figure : lorsque la matrice  $A = (a_{ij})_{1 \leq i, j \leq s}$  est strictement triangulaire inférieure la méthode est explicite, sinon elle est implicite. Dans le cas explicite, la détermination des valeurs intermédiaires se fait par calcul direct, ce qui s'interprète comme plusieurs étapes de prédiction-correction [Dem06] (figure 2.7). Parmi ces méthodes, les méthodes de Runge Kutta 2 et 4 font partie des plus populaires et possèdent un ordre correspondant. Le tableau de Butcher de la méthode de Runge-Kutta 4 (RK4) est donné en tableau 2.2.

Lorsque la matrice  $A$  contient des termes non nuls au dessus de la diagonale, la méthode est implicite puisqu'il existe  $j \in [1, \dots, s]$  tel que  $\kappa_j$  est des deux côtés de l'équation (eq. 2.12) : on ne dispose plus d'une procédure pour obtenir  $\kappa_j$  par le calcul direct. À la place, il s'agit de résoudre le système d'équation (eq. 2.12) d'inconnues les  $\{\kappa_i, i = 1, \dots, s\}$ . Cette résolution nécessite souvent de mettre en œuvre un algorithme de recherche de zéros tel que la méthode de Newton (cf section 4.2). La recherche de zéros est plus coûteuse que le calcul direct dans les méthodes explicites, mais en contrepartie la méthode implicite est plus stable numériquement. Une méthode RK explicite à  $s$  étages aura un ordre maximal  $s^a$  tandis qu'une méthode implicite peut avoir un ordre  $2s$  [RMD17]. En outre, il existe des méthodes de Runge-Kutta implicites construites pour limiter la complexité de calcul comme la classe des méthodes DIRK (*Diagonally Implicit Runge-Kutta*) pour lesquelles certains termes diagonaux de la matrice  $A$  sont non-nuls [KC16], c'est-à-dire que tous les étages ne sont pas implicites (cf table tableau 2.3).

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	1/2	1/2	1/6

TABLE 2.2 – Tableau de Butcher : méthode RK4.

$c_1$	0			
$\vdots$	$\vdots$	$\ddots$		0
$c_i$	$a_{i1}$		$a_{ii}$	
$\vdots$	$\vdots$			$\ddots$
$c_s$	$a_{s1}$	$\cdots$	$a_{s,s-1}$	0
	$b_1$	$\cdots$		$b_s$

TABLE 2.3 – Tableau de Butcher pour une méthode DIRK ( $a_{ii} \neq 0$ ).

a. Il existe des méthodes de Runge-Kutta explicite pour n'importe quel ordre. Cependant, l'ordre  $p = 4$  est le dernier pour lequel on puisse concevoir une méthode de Runge-Kutta à  $s = p$  étages [CK63]. Afin d'atteindre un ordre plus élevé, il faudra alors impliquer davantage d'étages. Cette distinction permet néanmoins de concevoir des méthodes imbriquées ayant le même ordre pour un nombre d'étages différents. Les méthodes implicites n'ont pas ces limites.

La mise en œuvre des méthodes IRK peut passer par la méthode de Newton (section 4.2) pour résoudre les équations de (eq. 2.12), qui nécessite à chaque itération le calcul de la Jacobienne des équations de Runge-Kutta pour les conditions de collocation, et son inversion (typiquement via une factorisation LU). Ce calcul peut être coûteux, c'est pourquoi des stratégies d'approximation peuvent être mises en œuvre afin de réutiliser la même Jacobienne et sa factorisation au cours de la résolution, comme la méthode de Rosenbrock [HNW09][HI14].

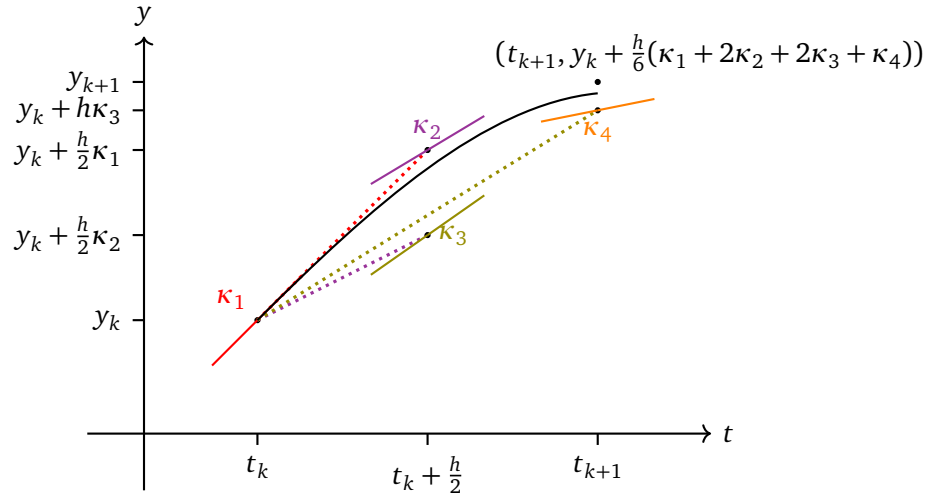


FIGURE 2.7 – Méthode de Runge-Kutta 4. Cette méthode s'appuie sur plusieurs étapes de prédiction/correction de la direction d'intégration.

## 2.7 Méthodes de collocation

On décrit à présent une classe de méthodes de Runge-Kutta implicites : les méthodes d'intégration par collocation. À partir de  $\xi_k$  connu, on approxime la solution  $\xi$  par une fonction polynomiale  $\tilde{P} : [t_k; t_{k+1}] \rightarrow \mathbb{R}^n$  associée au polynôme  $P$ . Dans la base des polynômes de Lagrange pour les points  $t_{k,i}$  définis précédemment, on écrit  $P'$  tel que  $\tilde{P}'$  interpole entre les valeurs intermédiaires de dérivées  $\{\kappa_i, i = 1, \dots, s\}$  :

$$P' = \sum_{i=1}^s \kappa_i L_i \quad \text{avec} \quad \forall i \in \llbracket 1, s \rrbracket, \quad L_i = \prod_{\substack{l=1 \\ l \neq i}}^s \frac{X - t_{k,l}}{t_{k,i} - t_{k,l}} \quad (2.13)$$

Afin de déterminer  $P$  de manière unique (par intégration de  $\tilde{P}'$ ) il faut ajouter l'équation  $\tilde{P}(t_k) = \xi_k$ . On calcule dès lors :

$$\xi(t_{k,i}) \approx \tilde{P}(t_{i,k}) = \tilde{P}(t_k) + \int_{t_k}^{t_{k,i}} \tilde{P}'(t) dt \quad (2.14)$$

$$= \xi_k + \int_{t_k}^{t_{k,i}} \left( \sum_{l=1}^s \kappa_l \tilde{L}_l(t) \right) dt \quad (2.15)$$

$$= \xi_k + \sum_{l=1}^s \left( \int_{t_k}^{t_{k,i}} \tilde{L}_l(t) dt \right) \kappa_l. \quad (2.16)$$

Un calcul similaire en intégrant  $\tilde{P}$  entre  $t_k$  et  $t_{k+1}$  donne :

$$\xi(t_{k+1}) \approx \xi_k + \sum_{l=1}^s \left( \int_{t_k}^{t_{k+1}} \tilde{L}_l(\tau) d\tau \right) \kappa_l.$$

Finalement, si l'on suppose le pas de temps d'intégration fixe et en appliquant le changement de variable  $\tau = \frac{t-t_k}{h}$  aux termes intégraux, on obtient les formules suivantes :

$$\xi(t_{k,i}) \approx \xi_k + h \underbrace{\sum_{l=1}^s \left( \int_0^{c_i} \tilde{L}_l(\tau) d\tau \right)}_{a_{il}} \kappa_l, \quad \xi(t_{k+1}) \approx \xi_k + h \sum_{l=1}^s \underbrace{\left( \int_0^1 \tilde{L}_l(\tau) d\tau \right)}_{b_l} \kappa_l.$$

La méthode d'intégration numérique par collocation s'écrit donc bien comme une méthode de Runge-Kutta implicite. Remarquons que le choix de la subdivision sur  $[t_k; t_{k+1}]$  détermine totalement la méthode de collocation. En effet, les coefficients du tableau de Butcher sont obtenus par le calcul d'intégrales sur les fonctions polynomiales  $\tilde{L}_i$ . Or ces polynômes ainsi que les bornes des intégrales ne dépendent que des  $\{c_i, i = 1, \dots, s\}$ . Le tableau de Butcher pour une méthode de collocation est donné par la table tableau 2.4.

$c_1$	$\int_0^{c_1} \tilde{L}_1(\tau) d\tau$	$\cdots$	$\int_0^{c_1} \tilde{L}_s(\tau) d\tau$
$\vdots$	$\vdots$		$\vdots$
$c_s$	$\int_0^{c_s} \tilde{L}_1(\tau) d\tau$	$\cdots$	$\int_0^{c_s} \tilde{L}_s(\tau) d\tau$
	$\int_0^1 \tilde{L}_1(\tau) d\tau$	$\cdots$	$\int_0^1 \tilde{L}_s(\tau) d\tau$

TABLE 2.4 – Tableau de Butcher pour une méthode de collocation.

En fonction des nœuds de collocation choisis, la méthode de collocation associée possède des propriétés légèrement différentes. En particulier, le choix des points de Legendre aboutit à une méthode A-stable d'ordre  $2s$  tandis que les points de Radau aboutissent à une méthode L-stable d'ordre  $2s - 1$ . Les méthodes de collocation possédant des ordres élevés sont particulièrement intéressantes en dépit de leur difficulté d'implémentation (système d'équations plutôt qu'application d'une procédure explicite). Leur stabilité numérique - en particulier pour le choix des points de Radau- en fait des méthodes privilégiées pour l'intégration des systèmes raides.

La capacité des méthodes de collocation à représenter de manière **continue** une solution approchée du problème (eq. 2.6) est une propriété intéressante pour la résolution des problèmes de contrôle optimal à temps continu. On peut distinguer deux classes (non disjointes) d'intégrateurs : ceux dont la sortie est discrète et ceux dont la sortie est *dense* (continue) [Enr+86]. Les intégrateurs discrets ne fournissent qu'un nombre fini de valeurs de la solution sur l'intervalle d'intégration  $[t_0, t_f]$  tandis que les intégrateurs continus fournissent une valeur en tout  $\tau \in [t_0, t_f]$ . En l'absence d'un traitement spécifique lors de leur conception, les méthodes de Runge-Kutta appartiennent généralement à la première catégorie tandis que les méthodes de collocation, qui emploient une paramétrisation polynomiale donc continue de la solution, font partie de la seconde. Les intégrateurs continus sont particulièrement intéressants, notamment pour le calcul la fonction coût des problèmes de contrôle optimal à temps continu. Il est néanmoins possible de mettre en œuvre une méthode d'interpolation de la solution entre les points d'une méthode dont la sortie n'est pas dense.

## 2.8 Notes et discussion

La résolution des équations aux dérivées partielles constitue un champ de recherche très large. Ici, nous avons simplement mentionné la méthode des différences finies qui permet d'approximer l'équation aux dérivées partielles par une équation différentielle ordinaire discrétisées dans l'espace. Des méthodes plus poussées peuvent être employées pour la simulation des systèmes, telles que la méthode des éléments finis [Bar16] où la solution est paramétrisée par un polynôme entre les nœuds du maillage, ou la méthode des volumes finis [EGH]. Quoiqu'il en soit, les idées utilisées pour la discrétisation temporelle d'un problème peuvent être appliquées à sa discrétisation spatiale.

Mentionnons également les méthodes *implicite-explicite* (IMEX) qui traitent différemment les dynamiques raides et non raides d'une ODE : une méthode implicite (L-stable) est utilisée pour traiter les dynamiques raides et une autre méthode pour les autres. Les méthodes de Runge-Kutta additives (ARK pour *additive Runge-Kutta*) [Rey+23] permettent d'imbriquer des méthodes (D)IRK et ERK, voire multi-pas, pour l'intégration des systèmes présentant plusieurs échelles de temps. Ce type d'approche permet de réduire la taille de l'équation à résoudre par la méthode de Newton et faire d'important gains de performance.



## CHAPITRE 3

### MÉTHODES DE TRANSCRIPTION

Le problème de contrôle optimal (eq. 1.43, sous-section 1.5.1) est difficile à résoudre analytiquement ; on le résout numériquement en pratique. Les variables de décision sont des fonctions dépendantes du temps, l'espace de recherche des solutions est donc de dimension infinie. Les méthodes directes proposent de discrétiser ces fonctions, par exemple en faisant appel à une représentation paramétrique (polynômes) puis de formuler un problème de programmation non-linéaire (NLP). L'espace de recherche est ainsi réduit à un nombre fini de variables de décision. On parlera de transcription du problème de contrôle optimal en un problème NLP. On ne résout donc pas le problème original à temps continu mais un problème discret l'approximant. On rappelle la formulation du problème de contrôle optimal :

$$P(p) : \min_{v,x} V_I(x(t_1)) + V_F(x(t_2)) + \int_{t_1}^{t_2} V(x, v, p) d\tau \quad (3.1a)$$

$$\text{sous contraintes } \forall \tau \in [t_1, t_2] \quad (3.1b)$$

$$f(\dot{x}(\tau), x(\tau), v(\tau), w(\tau), p) = 0 \quad (3.1c)$$

$$x(t_1) = \bar{x} \quad (3.1d)$$

$$g(x, v) = 0 \quad (3.1e)$$

$$h(x, v) \leq 0 \quad (3.1f)$$

où  $V_I$  désigne un terme de coût initial (nul en MPC, puisque l'état initial est imposé) et  $V_F$  désigne un coût terminal (erreur d'innovation en MHE, coût terminal approximant la fin de l'horizon en MPC) Pour un problème de commande prédictive on considère  $v = u$  (et  $w = 0$ ) et pour un problème d'estimation par horizon glissant, on considère  $v = w$ . En outre, les paramètres du problème contiennent  $x_0$  pour un problème de commande prédictive et  $y_{[t_1, t_2]}$  et  $u_{[t_1, t_2]}$  pour un problème d'estimation.  $v$  désigne donc de manière générale le signal de commande du problème de contrôle optimal.

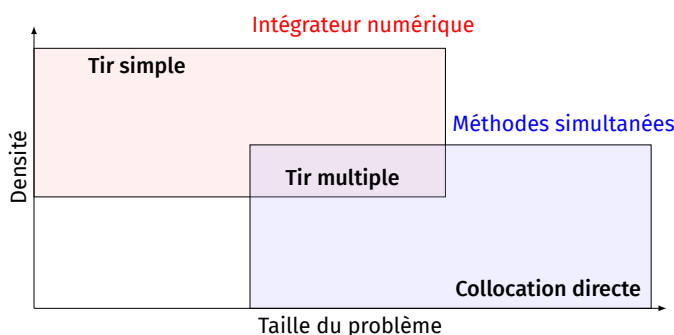


FIGURE 3.1 – Différentes méthodes de transcription aboutissent à des problèmes NLP différents mais dont la structure peut être exploitée.



### Discrétisation du temps.

L'horizon du problème (eq. 3.1) est divisé en sous-intervalles sur lesquels les signaux de commande et l'état du système sont représentés par un nombre fini de paramètres [Die01]. L'intervalle  $[t_1, t_2]$ , est découpé en  $N$  sous-intervalles définis à partir des  $N$  instants intermédiaires  $\tau_i, i \in \llbracket 0, \dots, N \rrbracket$  à partir desquels la trajectoire et les commandes sont paramétrisées.

$$t_1 = \tau_0 < \dots < \tau_N = t_2 \quad (3.2)$$

### Discrétisation de la commande.

On considère sans perte de généralité les signaux de commande constants par morceaux. En effet, pour une paramétrisation d'ordre plus élevé de la commande (polynomiale par exemple), on considère sans difficulté les coordonnées du signal dans la base paramétrique utilisée<sup>a</sup>. Par exemple, la décomposition de  $v$  pour une représentation polynomiale par morceaux dans la base  $\{P_i \in \mathbb{R}[X], i \in \llbracket 0, n-1 \rrbracket\}$  (polynômes d'ordre  $n$ ), s'écrit pour  $i \in \llbracket 1, N-1 \rrbracket$  :

$$\forall t \in [\tau_i, \tau_{i+1}], v(t) = \sum_{l=0}^{n-1} v_l^{(i)} \tilde{P}_l(t) \quad \text{où} \quad \forall l \in \llbracket 0, n-1 \rrbracket, v_l^{(i)} \in \mathbb{R}^{n_v}.$$

Le signal de commande augmenté  $V_i = (v_0^{(i)} \dots v_{n-1}^{(i)})^T$  est constant par morceaux. On peut également ajouter des contraintes de recollement pour assurer la continuité de la commande et ses dérivées aux instants  $\tau_i, \forall i \in \llbracket 1, N \rrbracket$ . En outre, certaines méthodes d'intégration peuvent prendre en compte des signaux de commande variables sur le pas d'intégration. C'est le cas des méthodes de Runge-Kutta, qui utilisent la valeur de la commande aux nœuds de Runge-Kutta. On peut mettre en œuvre une méthode d'interpolation pour déterminer la valeur des signaux de commande en ces instants.

### Discrétisation de la dynamique.

La réalisation des contraintes dynamiques (simulation du système) est supposée réalisée par un intégrateur tel que décrit au chapitre 2. Cet intégrateur réalise l'intégration numérique du modèle de sorte que pour  $k \in \mathbb{N}$  correspondant à  $t_k$  :

$$x_{k+1} = \Phi_k(x_k, v_k, w_k) \approx x_k + \int_{t_k}^{t_{k+1}} \dot{x}(\tau) d\tau$$

où  $\dot{x} = f(t, x, v, p)$ . On note que la méthode d'intégration n'est pas nécessairement la même sur chaque sous-intervalle de l'horizon du problème. Pour simplifier la notation, on omettra la dépendance en  $k$  puisque la taille du pas d'intégration ou la méthode employée n'a pas d'importance particulière dans la suite.

a. Par exemple les polynômes de Lagrange (cf eq. 2.13).

### Discrétisation de la fonction coût.

Le calcul de la fonction coût nécessite également de calculer un terme intégral dans (eq. 3.1). L'équation (eq. 3.1a) peut être réécrite, en tenant compte de la subdivision de l'horizon (eq. 3.2) :

$$\int_{t_1}^{t_2} V(x, v) d\tau = \sum_{i=0}^{N-1} \int_{\tau_i}^{\tau_{i+1}} V(x, v) d\tau.$$

Rien n'impose d'utiliser la même méthode d'intégration pour la fonction coût et pour l'équation d'évolution du modèle. On pourra ainsi rencontrer des discrétisations de type Euler (le terme intégral étant approximé par la somme des valeurs prises aux  $\tau_i$  pondérée par la largeur des intervalles), multipas ou Runge-Kutta alors que la dynamique emploie une autre méthode d'intégration adaptée aux caractéristiques du modèle telle que la collocation pour un modèle raide.

Dans le cas de schémas de Runge-Kutta pour l'intégration de la fonction coût, il est nécessaire d'évaluer la fonction coût entre les  $\tau_i$ , ce qui peut être réalisé si l'intégrateur numérique permet d'évaluer la trajectoire en des points intermédiaires. C'est possible pour des intégrateurs implémentant un schéma adaptatif de l'erreur (ou pour un appel chaîné à une méthode d'intégration avec un pas d'intégration plus petit que  $\tau_{i+1} - \tau_i$ ) ainsi que les représentations continues de la solution. À cette fin, il est avantageux d'intégrer la fonction coût en même temps que la dynamique du modèle au sein de l'intégrateur (qui pourra employer ou non différentes méthodes d'intégration pour la dynamique et la fonction coût). La fonction coût est alors obtenue en intégrant  $V(x, v)$  sur  $[t_1, t_2]$  à partir de la condition initiale  $V(x(t_1), v(t_1)) = 0$ .

En adoptant l'une des méthodes de discrétisation des commandes décrites précédemment, il est courant de réécrire pour  $i \in \llbracket 1, N - 1 \rrbracket$  :

$$\int_{\tau_i}^{\tau_{i+1}} V(x, v) d\tau = l(x_i, v_i).$$

Cette réécriture permet d'unifier la formulation du problème de contrôle optimal à temps continu et à temps discret.

Différentes stratégies de transcription existent comme le tir simple (*single-shooting*), le tir multiple (*multiple-shooting*) et la transcription simultanée par collocation directe (*direct collocation*) que nous allons décrire à présent.

### 3.1 Tir simple

Le **tir simple** constitue une approche intuitive et pragmatique pour la simulation du modèle. Pour une suite de commandes  $[v_0, \dots, v_{N-1}]$  et la condition initiale  $x_0$ , on intègre le modèle sur l'horizon du problème en appliquant successivement l'intégrateur sur chaque sous-intervalle défini par (eq. 3.2), de manière à obtenir les équations suivantes :

$$x_1 = \Phi(x_0, v_0) \quad (3.3a)$$

$$x_2 = \Phi(x_1, v_1) = \Phi(\Phi(x_0, v_0), v_1) \quad (3.3b)$$

$$\vdots$$

$$x_N = \Phi(x_{N-1}, v_{N-1}) = \Phi(\dots(\Phi(\Phi(x_0, v_0), v_1), \dots), v_{N-1}) \quad (3.3c)$$

À chaque évaluation des contraintes dynamiques du problème par le solveur d'optimisation, la trajectoire du système est simulée pour tout l'horizon d'un seul tenant, puis la solution du problème est ajustée par le solveur. On qualifie cette approche de **séquentielle**, pour signifier l'alternance entre les phases de simulation et d'optimisation. On peut également comprendre le caractère séquentiel de la méthode du point de vue de la simulation même, qui nécessite de chaîner les appels à l'intégrateur réalisés dans l'ordre présenté dans (eq. 3.3).

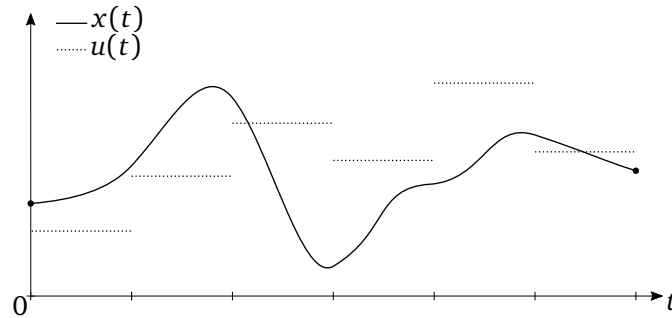


FIGURE 3.2 – Tir simple.

Pour la commande prédictive, cette transcription se traduit par un problème d'optimisation ayant pour variables de décision les commandes  $[u_0, \dots, u_{N-1}]$  (l'état initial étant supposé connu). Pour un problème d'estimation, seul le premier état  $x_0$  et les éventuelles perturbations sont variables de décision tandis que les commandes sont en général connues.

Pour un problème de commande prédictive, en ordonnant les variables de décision  $[u_0^T, \dots, u_{N-1}^T]$ , le problème NLP s'écrit :

$$\min_{u_0, \dots, u_{N-1}} V_F(x_F) + \sum_{k=0}^{N-1} l(x_k, u_k) \quad (3.4a)$$

$$\text{sous contraintes } \forall i \in \llbracket 0, N-1 \rrbracket \quad (3.4b)$$

$$x_{i+1} = \Phi(x_i, u_i) \quad (3.4c)$$

$$g(x_i, u_i) \leq 0 \quad (3.4d)$$

$$h(x_i, u_i) = 0 \quad (3.4e)$$

La mise en œuvre et la compréhension de cette méthode sont simples, ce qui explique sa popularité dans le milieu industriel. Les limites de cette méthode sont néanmoins bien connues [BP84]. Les erreurs introduites par le schéma d'intégration ainsi que celles induites par la précision numérique des quantités manipulées croissent au cours de la simulation<sup>a</sup>. L'erreur accumulée peut impacter la faisabilité du problème en violant les autres contraintes du problème.

Par ailleurs les appels imbriqués à l'intégrateur numérique accentuent les non-linéarités du problème d'optimisation. Dans cette approche séquentielle, le calcul des quantités intervenant dans la recherche de solution (évaluation de la Jacobienne des contraintes, gradient de la fonction coût, etc) nécessitent de faire appel de nombreuses fois à la règle de la chaîne. On a en effet  $\forall (i, k) \in \llbracket 0, N - 1 \rrbracket^2$  :

$$\nabla_{u_k} l(x_i, u_i) = \begin{cases} \partial_{x_i} l(x_i, u_i) \left( \prod_{j=k+1}^{i-1} \partial_{x_j} x_{j+1} \right) \partial_{u_k} x_{k+1} & \text{si } k \leq i. \\ 0 & \text{si } k > i. \end{cases} \quad (3.5)$$

Ce calcul impacte négativement le conditionnement du problème. Pour des problèmes plus complexes comme lors de la présence de raideurs ou d'instabilités du modèle, la méthode du tir simple converge rarement.

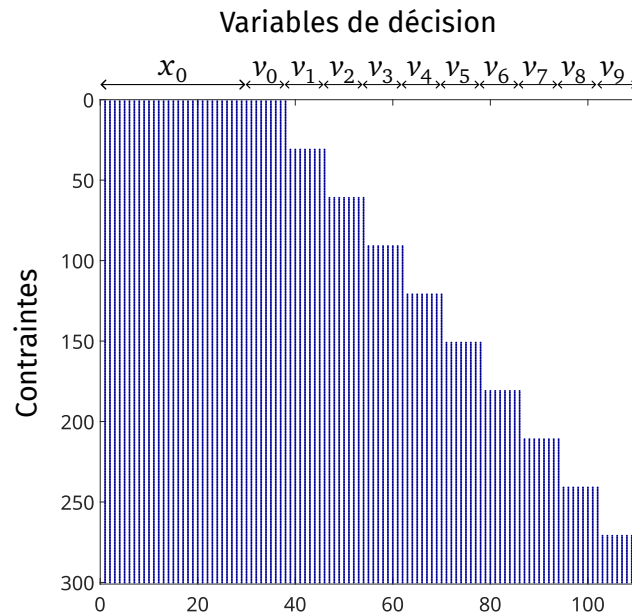


FIGURE 3.3 – Motif du caractère creux de la Jacobienne des contraintes du problème d'optimisation transcrit par tir simple, obtenu pour 10 sous-intervalles de discrétisation de l'horizon pour le problème d'estimation d'état décrit au chapitre 7. À mesure que l'on progresse dans l'horizon, la trajectoire dépend d'un nombre croissant de termes  $v_i$ .

a. On serait tenté de considérer la question de la précision numérique comme résolue, à l'heure où les ordinateurs embarquent des unités de calcul de plusieurs milliards de transistor et une précision machine très grande. Même l'IBM 3081 utilisé dans [BP84] en 1984 était capable d'opérations sur des mots de 32 voire 64 bits. Le déploiement d'algorithmes de contrôle optimal sur des équipements embarqués (micro-contrôleur, FPGA, etc) remet cette question au devant de la scène.

### 3.2 Tir multiple

La méthode du **tir multiple** pour le contrôle optimal direct a été introduite dans [BP84], mais trouve ses origines dans la résolution de problèmes aux limites [Osb69] sur lesquels le tir simple ne convenait pas. Comme nous l'avons évoqué en section 3.1, le tir simple n'est pas adapté pour les systèmes instables et non linéaires. La solution proposée en tir multiple est d'introduire les variables  $\forall k \in [0, N - 1], s_k$  de décision supplémentaires aux nœuds de tir (qui coïncident avec la discrétisation de l'horizon (eq. 3.2)) et d'imposer les contraintes de recollement suivantes :

$$\forall k \in [0, N - 1], s_{k+1} = \Phi(x_k, u_k) \quad \text{et} \quad x_{k+1} = s_{k+1} \quad (3.6)$$

On aboutit ainsi au problème NLP suivant

$$\min_{u_0, s_0, \dots, u_{N-1}, s_{N-1}, s_N} V_F(x_F) + \sum_{k=0}^{N-1} l(x_k, u_k) \quad (3.7a)$$

$$\text{sous contraintes } \forall i \in \llbracket 0, N - 1 \rrbracket \quad (3.7b)$$

$$s_{i+1} = \Phi(x_i, u_i) \quad (3.7c)$$

$$x_{i+1} = s_{i+1} \quad (3.7d)$$

$$g(x_i, u_i) \leq 0 \quad (3.7e)$$

$$h(x_i, u_i) = 0 \quad (3.7f)$$

Ainsi, au lieu de réaliser l'intégration numérique séquentielle du système sur  $[t_1, t_2]$ , on réalise  $N$  intégrations sur des intervalles plus petits. La simulation pour l'ensemble de la trajectoire n'est réalisée qu'en la solution du problème d'optimisation, on parle ainsi de méthode **simultanée** puisque l'optimisation et la simulation sur tout l'horizon sont réalisées simultanément

Les intégrations numériques sont indépendantes les unes des autres, il est donc possible (et souhaitable) de les réaliser en parallèle<sup>a</sup>. Il en va de même pour l'évaluation de la fonction coût sur chaque intervalle et des autres contraintes du problème, de sorte que l'évaluation des quantités nécessaires à la résolution du problème peut être réalisée plus rapidement en tir multiple que pour le tir simple.

En traitant la trajectoire de la sorte, des discontinuités au nœuds de tir apparaissent. L'intérêt de cette discontinuité est que les instabilités potentielles sont contenues aux sous-intervalles d'intégration, plus courts que l'horizon total, ce qui limite leur propagation. La continuité de la trajectoire est réalisée implicitement par le solveur, via le respect des contraintes (eq. 3.6) et non pas explicitement en imposant de repartir du résultat d'intégration du sous-intervalle précédent. La figure 3.4 schématise en l'exagérant l'allure de la trajectoire simulée du système au cours des itérations de l'algorithme de résolution. Il est possible d'initialiser le problème à une solution infaisable et d'assurer la faisabilité au

a. Bien qu'il soit attrayant de paralléliser les intégrations numériques sur chaque intervalle, il faut se pencher sur le temps passé dans le solveur de résolution du NLP, qui peut constituer la majorité du temps de calcul et que l'intégration numérique sur les sous-intervalles est peu coûteuse en comparaison. Les gains possibles en parallélisant peuvent alors être marginaux.

cours des itérations successives du solveur. L'introduction des variables de décision supplémentaires (non libres mais indépendantes) permet un calcul des dérivées nécessaires au problème plus aisé qu'en tir simple puisque les recours à la règle de la chaîne ne sont plus aussi longs, ce qui avait pour effet d'amplifier les non-linéarités et la sensibilité numérique. On a en effet par indépendance des variables de décision  $\forall(i, k) \in \llbracket 0, N - 1 \rrbracket^2$  :

$$\nabla_{u_k} l(x_i, u_i) = \begin{cases} \partial_{x_i} l(x_i, u_i) \partial_{u_{i-1}} \Phi(x_{i-1}, u_{i-1}) & \text{si } k = i - 1 \\ 0 & \text{sinon.} \end{cases} \quad (3.8a)$$

$$\nabla_{x_k} l(x_i, u_i) = \begin{cases} \partial_{x_i} l(x_i, u_i) \partial_{x_{i-1}} \Phi(x_{i-1}, u_{i-1}) & \text{si } k = i - 1 \\ 0 & \text{sinon.} \end{cases} \quad (3.8b)$$

En outre, le tir multiple possède des propriétés de convergence plus favorables que le tir simple. L'introduction des variables de décision supplémentaires aux nœuds de tir mène à la formulation d'un problème de dimension plus élevée mais les contraintes de recollement donnent de la structure au problème. Les auteurs de [AD10] montrent que l'introduction de variables supplémentaires non-libres dans le problème permet une convergence plus rapide des algorithmes d'optimisation implémentant une méthode de Newton.

Les avantages du tir multiple ne s'arrêtent pas là, puisqu'il est possible de réutiliser la solution d'un précédent problème pour initialiser la solution du problème courant, ce qui permet de résoudre plus rapidement le problème. En effet, les solutions déterminées pour la résolution de problèmes consécutifs sont proches. Deux stratégies se distinguent alors [Die01] :

- le **démarrage à chaud** (« warm start ») consiste à initialiser la solution du problème avec la solution précédente, sous l'hypothèse qu'entre deux résolutions les solutions sont proches.
- le **décalage** (« shifting ») qui utilise la solution du problème précédent décalée d'un pas de temps dans le futur. Il est alors nécessaire de déterminer la valeur avec laquelle initialiser le dernier sous-intervalle : il est courant d'utiliser la même valeur que pour l'avant-dernier ou de la connaissance a priori comme pour un système périodique .

En cas de distribution non uniforme des nœuds de tir (sous-intervalles de tailles différentes), le shifting est plus délicat à mettre en œuvre et on lui préfère le démarrage à chaud .

Si le tir multiple présente de nombreux avantages par rapport au tir simple, quelques désavantages sont à prendre en compte. Tout d'abord du point de vue des besoins en mémoire, il faut introduire autant de nouvelles variables de décision que de sous-intervalles, ainsi que toutes les contraintes de recollement, de sorte qu'un problème transcrit par tir multiple est de plus grande taille que le même problème transcrit par tir simple et occupe davantage de place en mémoire. Pour des applications destinées à fonctionner sur des équipements peu puissants, cela peut poser un frein. Néanmoins, le problème transcrit par tir multiple présente une structure diagonale par bloc et est creux : un solveur adapté ainsi qu'une librairie d'algèbre linéaire « creuse » (sparse linear algebra) permettent d'exploiter la connaissance des termes non nuls afin de limiter le nombre d'opérations. En outre, il existe différentes approches de réduction de la dimension du problème (voir section 3.4) qui permettent de limiter davantage les besoins en mémoire.

En outre, rien ne nous impose d'introduire des nœuds de tir multiple en chaque instant de discrétisation temporelle (eq. 3.2). Certaines méthodes mélangent les approches séquen-

tielles et simultanées, ce qui permet de réduire la taille du problème d'optimisation. Les travaux [Ret+22] [Axe15] montrent qu'il est possible d'obtenir de meilleurs résultats que le tir multiple et le tir simple en associant ces deux méthodes et en adaptant la structure du problème pour être plus ou moins creuse.

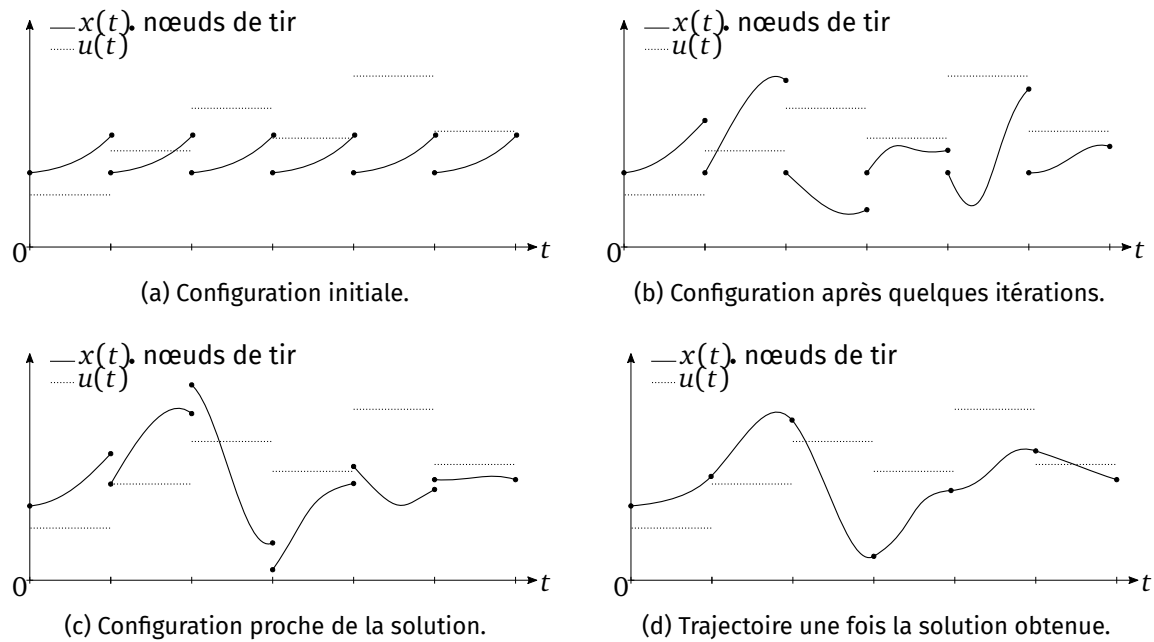


FIGURE 3.4 – Évolution de la solution pour une transcription en tir multiple.

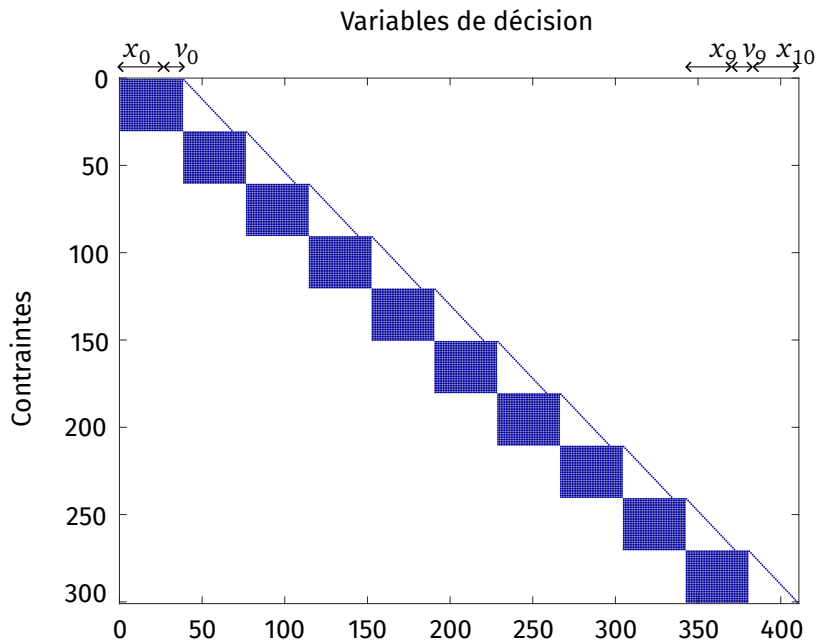
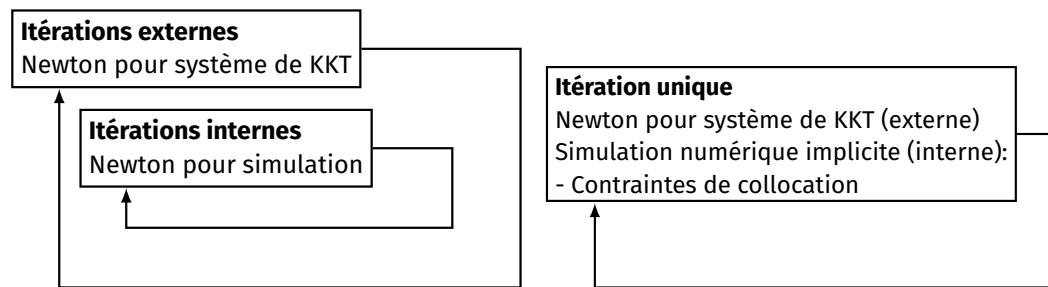


FIGURE 3.5 – Motif du caractère creux de la Jacobienne des contraintes du problème d'optimisation transcrit par tir multiple, obtenu pour 10 sous-intervalles de discrétisation de l'horizon pour le problème d'estimation d'état décrit au (chapitre 7). L'introduction de variables de décision supplémentaires, associées à l'état, et des conditions de continuité de la solution entre les sous-intervalles réduit la dépendance des contraintes du problème à un nombre réduit de variables de décision.

### 3.3 Transcription simultanée par collocation directe

La méthode de transcription par **collocation directe** peut d’une certaine manière être vue comme une extension du tir multiple. Supposons donc une transcription par tir multiple du problème d’optimisation (eq. 3.1) pour laquelle l’intégration numérique à partir de chaque nœud de tir est réalisée par une méthode de collocation (section 2.6). Dans ce cas, l’intégration nécessite de résoudre les équations de collocation (eq. 2.12), ce qui est en général réalisé via les itérations de Newton. On applique ainsi le même algorithme pour la résolution du problème d’optimisation (itérations **externes**) et pour l’intégration numérique au sein d’une itération du solveur d’optimisation (itérations **internes**).

En collocation directe, on propose de réunir les itérations internes et externes en augmentant le problème figure 3.6. Les variables de décision supplémentaires sont les variables de collocation sur chaque sous-intervalle et les conditions du problèmes sont augmentées des conditions de collocation.



(a) Itérations avec une méthode de tir (simple ou multiple) où la méthode de Runge-Kutta est utilisée. Deux itérations de Newton sont imbriquées.

(b) Itérations en collocation directe. Les itérations de Newton sont réunies.

FIGURE 3.6 – Comparaison de la résolution du problème pour une méthode de tir et pour la collocation directe.

Pour l’implémentation de la transcription par collocation directe, il est courant de modifier légèrement la formulation de la collocation présentée en section 2.6, qui est une forme **différentielle** [RMD17]. On utilisera davantage une formulation **intégrale**, dans laquelle le système d’équations à résoudre a pour inconnues les états de collocation  $\xi_{k,i}$  plutôt que les dérivées  $\kappa_i$ . L’avantage est qu’en manipulant directement les états de collocation, la quadrature de la fonction coût est plus simple à exprimer. De plus, le problème résultant est plus creux que pour la formulation différentielle de la collocation directe.

Abandonnons la notation que nous avons adoptée pour formuler le problème (eq. 2.6) et reprenons le formalisme de la représentation d’état. On veut transcrire le problème (eq. 1.38) en un problème NLP. On divise la fenêtre d’estimation en  $N \geq 1$  intervalles de taille  $h = H/N$  que l’on note  $[t_k; t_{k+1}]$ ,  $k = 0, \dots, N^a$ . Sur chaque intervalle, on approxime la trajectoire du système par une fonction polynômiale  $\tilde{P} : [t_k; t_{k+1}] \rightarrow \mathbb{R}^n$  qui interpole entre les états de collocation  $x_{k,i}^c$  qui sont variables de décision du problème NLP,  $i \in \llbracket 1, s \rrbracket$  où  $s$  est le nombre d’étages de la méthode. On approxime la fonction coût sur  $[t_k; t_{k+1}]$  par une

a. Le cas  $N = 1$  correspond aux méthodes dites pseudospectrales ou de collocation orthogonale. Dans ce type d’approche, on choisit un unique polynôme interpolateur de degré élevé, typiquement exprimé dans la base des polynômes de Chebychev [HR07].



fonction polynômiale  $\tilde{T} : [t_k; t_{k+1}] \rightarrow \mathbb{R}$ .

$$P = \sum_{l=1}^s L_l x_{k,l}^c \quad \text{et} \quad T = \sum_{l=1}^s L_l V(x_{k,l}^c)$$

$\tilde{P}'$  est une approximation de  $\dot{x}$  sur  $[t_k; t_{k+1}]$  que l'on fait coïncider avec la solution aux nœuds de collocation. Cela correspond à imposer la contrainte  $\tilde{P}'(t_{k,i}) = f(t_{k,i}, u(t_{k,i}), x_{k,i}^c, p)$  au nœud de collocation  $i$ . Ainsi les contraintes suivantes sont ajoutées au problème NLP :

$$\forall i = 1, \dots, s, \quad \sum_{l=1}^s \tilde{L}'_l(t_{k,i}) x_{k,l}^c = f(t_{k,i}, u(t_{k,i}), x_{k,i}^c, p)$$

À la fin de l'intervalle, on calcule :

$$x_{k+1} = \tilde{P}(t_{k+1}) = \sum_{l=1}^s L_l(t_{k+1}) x_{k,l}^c$$

Le terme intégral de la fonction coût est approximé sur  $[t_k; t_{k+1}]$  en calculant :

$$\int_{t_k}^{t_{k+1}} V(\mathbf{x}(t)) dt \approx \int_{t_k}^{t_{k+1}} \tilde{T}(t) dt = \sum_{l=1}^s \left( \int_{t_k}^{t_{k+1}} \tilde{L}_l(t) dt \right) V(x_{k,l}^c)$$

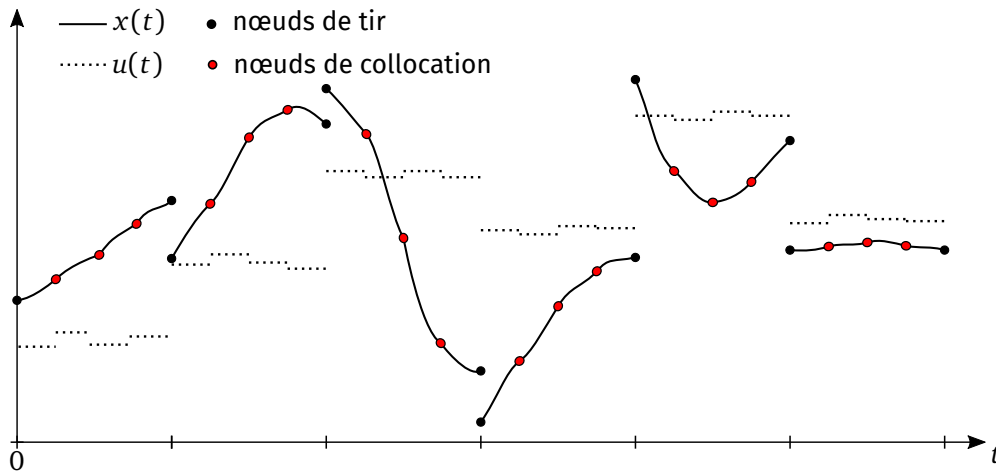


FIGURE 3.7 – Transcription simultanée par collocation directe.

Puisqu'une solution faisable du problème d'optimisation est une trajectoire faisable du système (par les contraintes de collocation), on s'affranchit de l'intégration numérique et du calcul des sensibilités. Autrement dit, l'intégration numérique du système est implicitement réalisée par le respect des contraintes de collocation. De manière similaire au tir multiple, la trajectoire est discontinue aux nœuds de tir tant que l'on n'a pas convergé vers une solution. De plus, les états intermédiaires aux nœuds de collocation peuvent ne pas respecter les contraintes de collocation, c'est-à-dire que la trajectoire simulée peut ne pas respecter la dynamique du système même au sein des sous-intervalles (eq. 3.2) de l'horizon, tant que l'on a pas convergé en une solution. Bien qu'il existe des schémas adaptatifs [LHR15], l'ordre d'intégration entre les nœuds de tir est le plus souvent fixé.

En conservant les points de collocation comme variables de décision, il est possible

d'ajouter des contraintes sur les états intermédiaires aux nœuds de collocation. On peut ainsi garantir que la trajectoire entre deux nœuds de tir reste faisable, ce qui n'est pas possible *a priori* dans le cadre du tir multiple où les contraintes ne portent que sur les nœuds de tir.

La collocation directe est une méthode simultanée au même titre que le tir multiple et présente les mêmes avantages que ce dernier par rapport au tir simple : démarrage à chaud de la solution, la propagation limitée des erreurs, instabilités et non-linéarités. La collocation directe est encore plus massivement parallélisable que le tir multiple puisque l'on peut paralléliser toutes les évaluations de la fonction d'évolution du modèle et non plus uniquement les appels à l'intégrateur. Le problème NLP résultant de la transcription par collocation directe est encore plus creux que pour une transcription par tir multiple, ce qui peut être avantageusement exploité par les solveurs NLP. Une représentation intégrale est plus creuse qu'une représentation différentielle de la méthode de collocation [RMD17]. Par ailleurs, la stratégie de globalisation (cf section 4.4) employée par le solveur d'optimisation est employée sur les nœuds de collocation, ce qui permet une meilleure convergence encore que pour le tir multiple, pour lequel cette stratégie n'est appliquée qu'aux nœuds de tir.

Néanmoins, ce type de transcription est moins adaptée pour les cibles embarquées dans la mesure où les besoins en mémoire sont bien plus grands. Il y a en effet encore plus de variables de décision que pour le tir multiple. Une approche intéressante appelée *lifted integrator* permet de concilier les avantages du tir multiple et de la collocation directe [Qui17]. Les itérations de Newton pour la résolution du problème NLP sont réalisées dans l'espace des variables de collocation, mais une étape de *condensing* permet de résoudre un problème de taille réduite équivalent à une transcription par tir multiple.

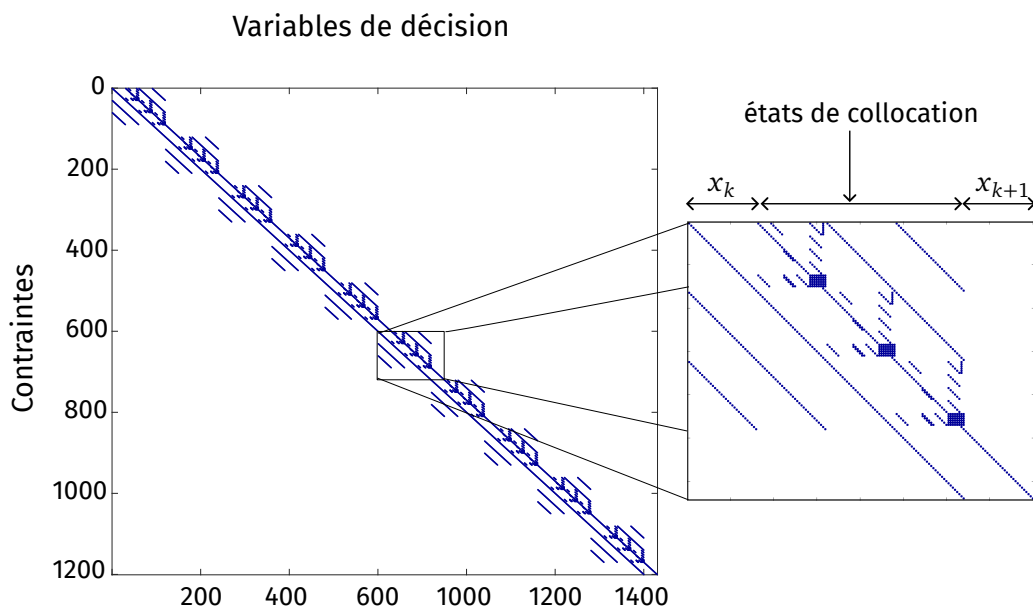


FIGURE 3.8 – Motif du caractère creux de la Jacobienne des contraintes du problème d'optimisation transcrit par collocation directe, obtenu pour 10 sous-intervalles de discrétisation de l'horizon et une méthode de collocation à l'ordre 3 pour le problème d'estimation d'état décrit au (chapitre 7). Le problème contient bien plus de variables de décision qu'en tir multiple mais est encore plus creux.

### 3.4 Exploitation de la structure du problème

Le problème NLP résultant de la transcription d'un problème de commande optimale peut présenter de nombreuses variables de décisions et contraintes. La taille du problème croît linéairement avec la taille de l'horizon et le nombre d'échantillon de la discrétisation du temps adoptée, de sorte que la taille du problème peut impacter significativement la vitesse de résolution du problème.

L'enjeu de l'exploitation de la structure du problème est d'aboutir à un problème NLP mathématiquement équivalent mais plus simple à résoudre, en éliminant certaines variables intermédiaires. Ces méthodes doivent être élaborées avec soin au risque de détériorer la vitesse de convergence de l'algorithme (voire la compromettre entièrement) [NW06].

On appelle couramment *structure* du problème l'ensemble des contraintes entre les variables de décision, comme par exemple les contraintes dynamiques entre les états et les signaux de commande, ainsi que le motif des termes non nuls dans les contraintes (*sparsity pattern*). Il convient en général d'ordonner correctement les contraintes afin de pouvoir exploiter la structure de manière systématique. Le problème résultant des transcriptions simultanées (tir multiple et collocation directe) est structuré et creux (figure 3.5 et figure 3.8) tandis qu'une transcription séquentielle mène à un problème de taille plus réduite mais dense (figure 3.3). Le caractère creux des problèmes permet de s'affranchir de nombreuses opérations calculatoires, en connaissant à l'avance l'emplacement des termes nuls dans les calculs (bibliothèques d'algèbres creuses). En outre de ce caractère creux, les contraintes du problème NLP possèdent une structure particulière (diagonale par bandes) qu'il est possible d'exploiter récursivement afin de simplifier la résolution du problème NLP. La plupart des solveurs utilisés en pratique pour le contrôle optimal implémentent une méthode d'exploitation de la structure.

Plus précisément, pour les méthodes de résolution directes des problèmes de contrôle optimal implémentant un algorithme de type itérations de Newton, il est nécessaire de résoudre un système linéaire correspondant à la linéarisation des équations de KKT. Considérons le problème NLP issu d'une transcription par tir multiple (eq. 3.7) d'un problème d'estimation sans contraintes d'inégalité. Le Lagrangien du programme non linéaire s'écrit :

$$\mathcal{L} = V_I(x_0) + V_F(x_N) + \sum_{i=0}^{N-1} l(x_i, v_i) + \lambda_{i+1}^T (x_{i+1} - \Phi(x_i, v_i))$$

Les conditions de KKT (section 4.1) s'expriment alors :

$$\begin{aligned} \nabla_{x_0} \mathcal{L} &= \nabla_{x_0} V_I(x_0) + \nabla_{x_0} l(x_0, v_0) - \nabla_{x_0} \Phi(x_0, v_0)^T \lambda_1 \\ \nabla_{x_i} \mathcal{L} &= \nabla_{x_i} l(x_i, v_i) + \lambda_i - \nabla_{x_i} \Phi(x_i, v_i)^T \lambda_{i-1}, \quad i \in \llbracket 1, N-1 \rrbracket \\ \nabla_{x_N} \mathcal{L} &= \nabla_{x_N} V_F(x_N) + \lambda_N \\ \nabla_{v_i} \mathcal{L} &= \nabla_{v_i} l(x_i, v_i) - \nabla_{v_i} \Phi(x_i, v_i)^T \lambda_{i-1}, \quad i \in \llbracket 0, N-1 \rrbracket \\ \nabla_{\lambda_{i+1}} \mathcal{L} &= x_{i+1} - \Phi(x_i, v_i), \quad \forall i \in \llbracket 0, N-1 \rrbracket \end{aligned}$$

Les variables primales et duales sont ordonnées  $w = (x_0, v_0, \lambda_1, \dots, x_{N-1}, v_{N-1}, \lambda, x_N)$  et les



### 3.4.1 Remarque sur la vitesse de convergence et la complexité des itérations

Aux section 3.2 et section 3.3 nous avançons que la projection du problème dans un espace de taille supérieur permet d'améliorer la **vitesse de convergence** puis nous venons d'énoncer que l'élimination des variables intermédiaires est souhaitable pour accélérer la résolution des itérations de Newton. Ces affirmations ne sont pas contradictoires : il faut distinguer les notions de convergence de l'algorithme, qui indique le nombre d'itérations nécessaires pour atteindre la solution, et le temps nécessaire à chaque itération. Un nombre réduit d'itérations très longues ne permet pas une résolution rapide du problème.

Afin d'assurer une convergence rapide, les itérations de Newton « de haut niveau » peuvent être réalisées dans l'espace de plus grande dimension des méthodes de transcription simultanées, tandis que l'exploitation de la structure permet de résoudre un problème linéaire plus simple. Cette approche implique l'implémentation d'une phase de restauration ou d'expansion, qui permet de retrouver les variables supplémentaires des méthodes de transcription simultanées.

Ainsi, si les problèmes résolus par tir simple et tir multiple sont exactement équivalent du point de vue mathématique, les itérations de Newton appliquées aux deux problèmes donnent en pratique des résultats différents. Pour autant, l'application d'une méthode d'exploitation de la structure du problème transcrit en tir multiple aboutit à un sous-problème d'une complexité et d'une taille exactement équivalente à celle du sous-problème en tir simple.

### 3.4.2 Condensing

Les variables intermédiaires  $\Delta x_k$ ,  $k > 0$  du problème eq. 3.10 peuvent être éliminées en considérant la dépendance linéaire des variables, de sorte que :

$$\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_N \end{bmatrix} = \underbrace{\begin{bmatrix} A_0 & B_0 & & \\ A_1 A_0 & A_1 B_0 & B_1 & \\ \vdots & \vdots & & \\ \prod_{k=N-1}^0 A_k & \prod_{k=N-1}^1 A_k B_0 & \cdots & B_{N-1} \end{bmatrix}}_M \begin{bmatrix} x_0 \\ \Delta v_0 \\ \Delta v_1 \\ \vdots \\ \Delta v_{N-1} \end{bmatrix} \quad (3.11)$$

On peut donc écrire le problème linéaire quadratique eq. 3.10 sous la forme condensée :

$$\min_U \frac{1}{2} U^T H U + h^T U$$

avec :

- $H = \begin{bmatrix} M & \mathbf{0}_{n_x} & I \end{bmatrix} \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \begin{bmatrix} M & \mathbf{0}_{n_x} & I \end{bmatrix}^T$ ,
- $Q = \text{diag}(Q_0, \dots, Q_N)$ ,
- $R = \text{diag}(R_0, \dots, R_N)$ ,
- $S = \text{diag}(S_0, \dots, S_N)$ ,
- $I = \text{diag}(I_{n_v}, \dots, I_{n_v})$ .

Les états intermédiaires introduits dans les méthodes simultanées des variables de décision du problème sont éliminés et le problème quadratique résultant est dense et de

taille réduite. On détermine alors l'itération de Newton en résolvant ce problème quadratique puis en utilisant eq. 3.11 pour retrouver les états intermédiaires éliminés (expansion).

Il existe également une approche partielle au condensing. Au lieu d'éliminer la totalité des états intermédiaires, seule une partie des états est éliminée en considérant une partition de  $\llbracket 0, N \rrbracket$ ,  $\{\llbracket N_i, N_{i+1}, i = 1, \dots, N_{\text{partition}} \rrbracket$  ( $N_1 = 0$  et  $N_{\text{partition}} = N$ ) et en éliminant tous les états  $x_j$  pour  $j \neq N_i$ . Cette approche permet de contrôler davantage la taille du problème linéaire à résoudre dans les itérations de Newton [Axe15].

### 3.4.3 Réursion de Riccati

Le problème (eq. 3.10) possède une solution analytique basée sur la réursion de Riccati. Cette méthode de résolution se base sur la programmation dynamique (la réursion de Riccati employée à la factorisation de la matrice de KKT est donc parfois appelée « programmation dynamique »). Un solveur linéaire exploitant la réursion de Riccati permet d'exploiter avantageusement la structure particulière du problème d'optimisation, et peut être vue comme une méthode sur-mesure de factorisation du système linéaire. Afin d'ajuster le caractère creux du problème, il est possible d'effectuer un *condensing* partiel du problème avant d'appliquer la réursion de Riccati [AM12]. Il est possible de mettre en œuvre la réursion de Riccati en même temps qu'une méthode des points intérieurs [RWR98].

## 3.5 Notes et discussion

Les différentes méthodes de transcription que nous avons présentées ne sont finalement que des manières différentes de contrôler *a priori* le caractère creux du problème NLP en introduisant plus ou moins de variables de décision *non libres* dans le problème non-linéaire discrétisant le problème de contrôle optimal (tableau 3.1).

Transcription	MPC	MHE
Tir simple	$[u_1, \dots, u_N]$	$[x_0, w_1, \dots, w_N]$
Tir multiple (MS)	$[u_1, x_1, \dots, u_N, x_N]$	$[x_0, w_1, x_1, \dots, w_N, x_N]$
Collocation directe	MS(MPC) + variables de collocation	MS(MHE) + variables de collocation

TABLE 3.1 – Variables de décision pour chaque méthode de transcription et pour le type de problème de contrôle optimal.

Concernant le tir simple et le tir multiple, ces deux méthodes apparaissent comme des approches extrêmes vis-à-vis de la structure du problème résultant de la discrétisation du temps (eq. 3.2), en faisant usage d'un intégrateur numérique. En tir simple il s'agit d'introduire le minimum de variables de décision associées à l'état : aucune pour la MPC pour laquelle  $x_0$  est connu et une seule en MHE —le premier état, le reste de trajectoire étant obtenu par simulation—. En tir multiple, il s'agit à l'inverse d'introduire le maximum de variables de décision associées à l'état : une pour chaque instant de discrétisation. On pourrait tout à fait envisager d'inclure un jeu quelconque d'états discrétisés le long de la trajectoire. Cette approche rejoint celle du tir multiple partiel.

Par ailleurs nous présentons le cas où une seule méthode de transcription est employée. Il est envisageable d'employer toutes ces méthodes à la fois sur la discrétisation (eq. 3.2) et

utiliser tantôt de la collocation directe sur certains intervalles et des méthodes de tir sur d'autres. Cela ne pose aucun problème au regard du programme non-linéaire qu'il s'agira de résoudre. Formellement, cette approche est équivalente à l'introduction d'« étages », chacun correspondant à un problème d'optimisation (potentiellement différent) pour lesquels on impose une continuité de la solution. En fonction des besoins<sup>a</sup>, la méthode hybride de transcription peut être mise en œuvre afin d'optimiser le fonctionnement sur un équipement cible et ses caractéristiques techniques, ou encore être adaptée entre les résolutions successives du problème de contrôle optimal, voire au sein de la même résolution du problème. Les auteurs de [Cal+24] adaptent ainsi la structure du problème d'optimisation en ligne, afin d'affiner les performances et le respect des contraintes lorsqu'elles changent.

Néanmoins, la question de la propagation des non-linéarités lors du chaînage de l'intégrateur pose toujours un problème pour les systèmes raides. Si le choix d'une méthode de transcription fixe *a priori* la structure creuse du problème, la méthodologie de *condensing* (partiel ou total) semble être particulièrement pertinente pour ajuster la structure du problème *a posteriori* de la transcription. En gardant les itérations de haut-niveau dans la transcription la plus redondante (contenant le plus de variables non libres) et en appliquant une méthode de réduction au moment de résoudre la matrice de KKT du problème, il est possible de conserver des propriétés de convergence intéressantes tout en limitant l'effort de calcul nécessaire à chaque itération et de l'adapter au problème. Ce type d'approche est également possible en collocation directe, en le réduisant à une structure de tir multiple en employant le *condensing*. Les mêmes itérations de Newton sont effectuées, mais à un coût de calcul réduit [Qui+17].

Les méthodes de transcription simultanées comme le tir multiple et la collocation directe présentent des propriétés de convergence avantageuses ainsi qu'un potentiel de parallélisation des calculs par indépendance des contraintes entre les nœuds de tir. Néanmoins, le tir multiple faisant usage d'un intégrateur numérique, il peut mettre en œuvre un contrôle de l'erreur d'intégration et ainsi assurer la précision de la trajectoire simulée, ce qui suppose une solution plus proche de l'optimalité qu'une transcription par collocation directe, dont l'ordre est figé au moment de la transcription. Par ailleurs, l'usage d'un intégrateur de type IRK permet de jouir de la même stabilité numérique que la collocation directe. Les auteurs de [PBB23] montrent ainsi que le tir multiple et la collocation directe présentent des performances similaires sur des problèmes biomécaniques.

---

a. Par exemple : évitement d'obstacle en MPC, mesures infréquentes en MHE.

## CHAPITRE 4

## PROGRAMMATION NON-LINÉAIRE

*...in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity.*

— SIAM Review, 1993

R. Tyrrell ROCKAFELLAR

**Définition 4.1 (Programme non-linéaire)**

On définit le problème de **programmation non linéaire**<sup>a</sup> ou **programme non-linéaire** comme le problème d'optimisation suivant :

$$\min_w F(w) \quad (4.1a)$$

$$\text{sous contraintes } G(w) = 0 \quad (4.1b)$$

$$H(w) \leq 0 \quad (4.1c)$$

Où :

- $F : \mathbb{R}^{n_w} \rightarrow \mathbb{R}$  de classe  $\mathcal{C}^2$  ou deux fois différentiable sur  $\mathbb{R}^{n_w}$  est appelée fonction coût.
- $G : \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_G}$  de classe  $\mathcal{C}^2$  ou deux fois différentiable sur  $\mathbb{R}^{n_w}$  formalise les contraintes d'égalité.
- $H : \mathbb{R}^{n_w} \rightarrow \mathbb{R}^{n_H}$  de classe  $\mathcal{C}^2$  ou deux fois différentiable sur  $\mathbb{R}^{n_w}$  formalise les contraintes d'inégalité.
- $w \in \mathbb{R}^{n_w}$  est le vecteur des variables de décision où  $n_w \gg 1$  (typiquement de l'ordre de plusieurs centaines à milliers de composantes).

Si le problème (eq. 4.1) est convexe, la solution est unique et est forcément globale. Dans les cas non-convexes (la non convexité peut être due à la nature de la fonction coût ou aux contraintes), les solutions que nous pouvons trouver ne sont que locales. Il est possible d'employer des **heuristiques de globalisation** telles que la recherche linéaire ou les régions de confiance qui favorisent l'éloignement des minimums locaux.

De plus, l'introduction des contraintes nécessite de réaliser un compromis entre la minimisation du critère et la faisabilité de la solution, c'est-à-dire le respect des contraintes. Des **stratégies de globalisation**, telles que l'usage de fonction mérite ou de filtres [WB06], permettent de s'assurer que la solution permet bien de faire décroître le critère au cours des itérations. Les fonctions de mérite s'assurent que la minimisation du critère et la faisabilité de la solution sont conjointement satisfaits tandis que l'approche par filtre améliore l'un ou l'autre.

De nombreuses méthodes et solveurs d'optimisation permettent de résoudre ces problèmes. Dans un premier temps, prétendons que résoudre un problème NLP sans contrainte



d'inégalité est « facile ». Une grande famille de méthodes tombant sous la terminologie de « méthodes de Newton » permet de résoudre efficacement le problème (eq. 4.1) si l'on ignore les contraintes d'inégalité. Les différentes approches pour traiter les contraintes d'inégalité (eq. 4.1c) reviennent presque toujours à trouver un moyen de résoudre une suite de problèmes plus simples, aux seules contraintes d'égalité.

Les méthodes d'ensemble actif ignorent les contraintes d'inégalité inactives et considèrent uniquement les contraintes actives. Elles peuvent être difficiles à mettre en œuvre car l'ensemble des contraintes actives est amené à évoluer au cours des itérations de l'algorithme de résolution et les sous-problèmes résultants ne présentent pas toujours la même structure. D'autre part, les méthodes du point intérieur et la programmation quadratique séquentielle (ou SQP)<sup>a</sup> résolvent des versions relaxées des conditions d'optimalité du problème (appelées conditions de KKT).

## 4.1 Conditions de KKT

### Définition 4.2 (Point faisable ou réalisable)

Un point faisable est un  $w \in \mathbb{R}^{n_w}$  tel que les conditions d'égalité et d'inégalité sont respectées :  $G(w) = 0$  et  $H(w) \leq 0$ .

### Définition 4.3 (Minimum local)

Un minimum local est un point  $w^* \in \mathbb{R}^{n_w}$  tel que  $F(w^*) \leq F(w)$  pour tout  $w$  dans un voisinage de  $w^*$ .

### Définition 4.4 (Contraintes actives)

Les contraintes d'inégalité actives au point  $w$  sont les  $H_j$  telles que  $H_j(w) = 0$ . Les contraintes d'égalité sont toujours actives. L'ensemble actif est formé de toutes les contraintes actives. On peut ranger les contraintes d'inégalité actives dans les contraintes actives :

$$\tilde{G}(w) = \begin{pmatrix} G(w) \\ H^{\text{active}}(w) \end{pmatrix}$$

On définit alors les contraintes inactives en  $w$  comme les  $H_j$  telles que  $H_j(w) < 0$ .

### Définition 4.5 (Point régulier)

Un point régulier est un point  $w \in \mathbb{R}^{n_w}$  réalisable tel que la Jacobienne des contraintes actives  $\nabla \tilde{G}(w)^T$  est de rang plein par les lignes.

### Définition 4.6 (Lagrangien)

Les multiplicateurs de Lagrange  $\lambda \in \mathbb{R}^{n_G}$  et  $\mu \in \mathbb{R}^{n_G}$  permettent de formuler la fonction Lagrangienne :

$$\mathcal{L}(w, \lambda, \mu) = F(w) - \lambda^T G(w) - \mu^T H(w).$$

a. SQP : pour Sequential Quadratic Programming.

**Théorème 4.7 (Conditions de KKT)**

Si un point régulier  $w^* \in \mathbb{R}^{n_w}$  est minimum local de  $P$ , alors il existe  $\lambda^* \in \mathbb{R}^{n_G}$  et  $\mu^* \in \mathbb{R}^{n_H}$  uniques tels que  $(w^*, \lambda^*, \mu^*)$  vérifie les conditions suivantes :

$$\nabla_w \mathcal{L}(w^*, \lambda^*, \mu^*) = 0 \quad (4.2a)$$

$$G(w^*) = 0 \quad (4.2b)$$

$$H(w^*) \leq 0 \quad (4.2c)$$

$$\mu^* \geq 0 \quad (4.2d)$$

$$\mu_j^* H_j(w^*) = 0, \quad j = 1, \dots, n_H. \quad (4.2e)$$

Les conditions de KKT (eq. 4.2) donnent les conditions d'optimalité pour une solution au problème (eq. 4.1). La condition (eq. 4.2e) est appelée condition complémentaire et implique que  $\mu_j^* = 0$  pour les contraintes inactives.

**4.2 Méthode de Newton**

La méthode de Newton [NW06] (parfois également appelée méthode de Newton-Raphson) permet de résoudre itérativement des équations non-linéaires. Partant d'une solution initiale  $x_0 \in \mathbb{R}^{n_x}$ , on construit itérativement la solution à l'équation  $E(x) = 0$  où  $E : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_E}$  en « descendant » le gradient de  $E$  en la solution courante.

Soit l'itérée courante  $x_k$ , la linéarisation de l'équation (de solution  $x^*$ ) autour de  $x_k$  s'écrit :

$$E(x^*) \approx E(x_k) + \nabla E(x_k)(x^* - x_k) = 0$$

On construit ainsi itérativement la suite des  $(x)_{k \in \mathbb{N}}$  :

$$x_{k+1} = x_k - (\nabla E(x_k))^{-1} E(x_k)$$

On détermine un critère d'arrêt pour la procédure, de sorte que la solution numérique à l'équation est « trouvée » au rang  $N \in \mathbb{N}$ ,  $x^* = x_N$ .

En particulier, la méthode de Newton est utilisée pour résoudre des problèmes sous contraintes d'égalité de la forme :

$$\min_w F(w) \quad (4.3a)$$

$$\text{sous contraintes } G(w) = 0 \quad (4.3b)$$

On transforme ce problème en un problème d'optimisation sans contraintes via le Lagrangien  $\mathcal{L}(w, \lambda) = F(w) + \lambda^T G(w)$ . Les conditions d'optimalité pour le problème (eq. 4.3) sont données par l'équation suivante :

$$\nabla_w \mathcal{L}(w, \lambda) = 0 \quad (4.4a)$$

$$G(w) = 0 \quad (4.4b)$$

La méthode de Newton appliquée à cette équation, une fois linéarisation des conditions

d'optimalité (eq. 4.4) autour de la solution courante, consiste à résoudre l'équation linéaire suivante :<sup>a</sup>

$$\begin{bmatrix} \nabla_w^2 \mathcal{L}(w, \lambda) & \nabla G(w)^T \\ \nabla G(w) & 0 \end{bmatrix} \begin{bmatrix} \Delta w_k \\ \Delta \lambda_k \end{bmatrix} = \begin{bmatrix} -\nabla_w \mathcal{L}(w_k, \lambda_k) \\ -G(w_k) \end{bmatrix} \quad (4.5)$$

### Méthodes quasi-Newton

La méthode que nous avons brièvement introduite n'est pas celle implémentée en pratique dans les solveurs d'optimisation, qui utilisent différentes stratégies qui visent à améliorer les performances de la méthode. En effet, bien que la méthode de Newton converge quadratiquement vers une solution pour un calcul exact du gradient  $\nabla E$ , les calculs à chaque itération peuvent être coûteux. Les méthodes dites **quasi-Newton** [NW06] mettent en œuvre certaines approximations pour ce calcul. Bien que leur vitesse de convergence soit inférieure à la méthode de Newton<sup>b</sup>, elles peuvent permettre des résolutions du problème plus rapides grâce à des itérations moins coûteuses.

Au sein des méthodes quasi-Newton, la question du calcul de la Hessienne est un point qui mérite notre attention. Le calcul de la Hessienne  $\nabla_w^2 \mathcal{L}(w, \lambda)$  (ou de son inverse) peut être coûteux, tandis que le calcul d'une valeur approchée peut être suffisant. La méthode de **Gauss-Newton** [DFH09] est bien connue pour les problèmes d'optimisation dont la fonction coût est une somme de carrés, ce qui est courant pour les problèmes de régression non-linéaire voire même en contrôle optimal où la fonction coût est souvent quadratique. La forme particulière de la fonction coût permet de calculer une approximation simple de la Hessienne. En outre, le calcul de la totalité de la Hessienne n'est pas nécessaire mais il peut suffire d'employer des méthodes adjointes pour résoudre (eq. 4.5).

## 4.3 Recherche linéaire et région de confiance

En tant que méthode de descente du gradient, la méthode de Newton construit la solution du problème itérativement en partant d'un point initial  $x_0$  et en « descendant le gradient », ce qui assure de faire décroître le critère du problème d'optimisation. Dans ce formalisme, la direction de descente est notée  $\Delta w_k$  et la taille du « pas » (longueur de l'incrément) est noté  $\alpha_k > \mathbb{R}$ . Les itérations de l'algorithme sont ainsi :

$$x_{i+1} = x_i + \alpha_k \Delta w_k$$

Il faut noter par ailleurs que la méthode de Newton donne la solution d'une approximation quadratique tangente au problème initial. Cette approximation n'est valable que dans un voisinage de l'itération courante. Pour pallier à cette limite, la méthode de région de confiance (*trust region*) procède en déterminant d'abord la longueur de l'incrément  $\alpha_k$ , assurant la validité de l'approximation réalisée, puis la direction de descente est déterminée.

La méthodologie de recherche linéaire (*line search*) considère plutôt que la solution

a. La résolution des équations linéaires passe rarement par l'inversion d'une matrice. En général, les solveurs de résolution résolvent le système par des méthodes de factorisation matricielles basées sur la décomposition LU, QR ou Cholesky et exploitent la structure du problème lorsque c'est possible.

b. Leur convergence est linéaire voire superlinéaire.

de (eq. 4.5) issue d'une méthode de Newton ou quasi-Newton constitue une direction de recherche de la solution. Une fois cette direction de recherche trouvée, la longueur de l'incrément  $\alpha_k > \mathbb{R}$  est déterminée comme solution du problème <sup>a</sup> :

$$\min_{\alpha_k} F(w_i + \alpha_k \Delta w_i)$$

En procédant ainsi, on souhaite s'assurer que les incréments de la solution permettent de faire décroître le critère du problème d'optimisation un maximum pour chaque itération.

La méthode de Broyden-Fletcher-Goldfarb-Shanno (BFGS) emploie une méthode d'approximation de la Hessienne en plus d'une méthode de *line-search*, tandis que la méthode de Levenberg-Marquardt est une méthode de Gauss-Newton implémentant une méthodologie *trust region* [NW06].

## 4.4 Stratégie de globalisation et mécanismes de globalisation

Si la résolution d'un problème NLP consiste à déterminer une solution aux conditions de KKT (eq. 4.2) via la méthode de Newton, la seule détermination d'un point critique ne suffit pas. D'une part, la non-convexité du problème implique que la solution est locale, ce qui pose la question de l'initialisation de la solution du problème, car toutes les initialisations n'aboutiront pas à la même solution. Il est souhaitable que l'algorithme implémenté par le solveur ne soit pas trop sensible au choix de solution initiale. D'autre part, les conditions de KKT contiennent les équations garantissant la faisabilité de la solution ainsi que la minimisation de la fonction coût du problème. Ces deux critères peuvent entrer en opposition.

Une **stratégie de globalisation** [NW06] permet d'apporter une solution aux problèmes que nous venons d'évoquer. La stratégie de globalisation a pour rôle de tester si un incrément de solution  $(\Delta w_i, \Delta \lambda_i, \Delta \mu_i)$  permet effectivement de se rapprocher d'une solution du problème où s'il faut rejeter l'incrément et en déterminer un autre. Les stratégies de globalisation les plus répandues sont les **filtres** et les **fonctions de mérite**.

En cas de rejet de l'incrément par la stratégie de globalisation, un **mécanisme de globalisation** est employé afin de calculer un nouvel incrément. Les mécanismes de globalisation les plus répandus sont le *line-search* (ou recherche linéaire) et *trust-region* (région de confiance).

Il est possible de combiner les différentes stratégies avec les mécanismes de globalisation pour obtenir les différents solveurs disponibles sur l'étagère. Par exemple, les solveurs IPOPT [WB06] et SNOPT [GMS05] exploitent la recherche linéaire, mais le premier est basé sur une stratégie de filtre tandis que le second utilise les fonctions mérite.

### 4.4.1 Fonction mérite

Une fonction mérite évalue l'incrément  $(\Delta w_i, \Delta \lambda_i, \Delta \mu_i)$  en agrégeant la minimisation de la fonction coût et le respect des contraintes dans un même indicateur de la qualité de

---

a. Une des nombreuses méthodes existant pour la résolution de ce problème est évoquée en section 4.2. On peut également employer d'autres méthodes telles que la dichotomie, l'essai/erreur ou une heuristique.

l'incrément. Plus précisément, on introduit la fonction mérite :

$$F_{\text{merit}}(w_{k+1}, \mu) = F(w) + \mu\varphi(G(w)) + \mu\varphi(\max(0, H(w))) \quad (4.6)$$

$\mu \in \mathbb{R}$  est appelé paramètre de pénalisation et  $\varphi$  est une fonction appelé fonction de pénalité. Différents choix de fonction de pénalité existent comme la norme  $l_1$  ( $\varphi(x) = \sum_i |x_i|$ ) ou la norme  $l_2$  ( $\varphi(x) = \|x\|_2$ ). Il s'agit alors d'utiliser (eq. 4.6) pour tester d'incrément  $\Delta w_k$  afin que  $F_{\text{merit}}(w_k + \Delta w_k) \leq F_{\text{merit}}(w_k)$ . En cas d'échec de ce test, le mécanisme de globalisation sert à calculer un nouvel incrément.

#### 4.4.2 Filtres

Les fonctions mérite déterminent si l'incrément de la solution  $(\Delta w_i, \Delta \lambda_i, \Delta \mu_i)$  constitue une solution qui va effectivement dans le sens d'une résolution du problème en évaluant conjointement la minimisation de la fonction coût et la faisabilité des contraintes du problème. L'autre approche, utilisant un filtre (au sens d'un ensemble partiellement ordonné), est inspirée de l'optimal de Pareto en optimisation multi-objectif [ED18] et accepte un pas qui permet d'améliorer la faisabilité du problème et/ou la minimisation de la fonction coût. La **mesure de faisabilité** du problème agrège toutes les contraintes du problème et s'écrit :

$$F_{\text{feas}}(w) = \sum_{i \leq n_G} |G_i(w)| + \sum_{i \leq n_H} |\max(0, H_i(w))|$$

$\sum_{i \leq n_G} |G_i(w)|$  correspond à la faisabilité des contraintes d'égalité du problème, nulle lorsque les contraintes d'égalité sont respectées,  $\sum_{i \leq n_H} |\max(0, H_i(w))|$  correspond à la faisabilité des contraintes d'inégalité, nulles lorsque  $\forall i \leq n_H, H_i(w) < 0$ . On peut alors considérer que le problème NLP comme un problème à deux critères :

$$\min_w \begin{cases} F(w) \\ F_{\text{feas}}(w) \end{cases}$$

Un *filtre* désigne un sous-ensemble d'un ensemble partiellement ordonné. La stratégie de globalisation construit un filtre  $\mathcal{F}$  contenant des valeurs  $(F(w_i), F_{\text{feas}}(w_i))$  au fil des itérations. Un pas est alors accepté si pour l'incrément  $(\Delta w_i, \Delta \lambda_i, \Delta \mu_i)$ ,

$$\forall l, (F(w_l), F_{\text{feas}}(w_l)) \in \mathcal{F}, \quad F(w_i + \Delta w_i) \leq F(w_l) \quad \text{ou} \quad F_{\text{feas}}(w_i + \Delta w_i) \leq F_{\text{feas}}(w_l).$$

Le nouveau point  $(F(w_i + \Delta w_i), F_{\text{feas}}(w_i + \Delta w_i))$  est ajouté au filtre et tout élément du filtre pour lequel les deux inégalités sont vraies (on dit qu'il est dominé [Jun06, def1.12]) est retiré.

Cette stratégie consiste à accepter un pas qui, s'il ne permet pas d'améliorer à la fois la faisabilité du problème et la minimisation de la fonction coût, constitue un progrès sur au moins l'un des deux critères.

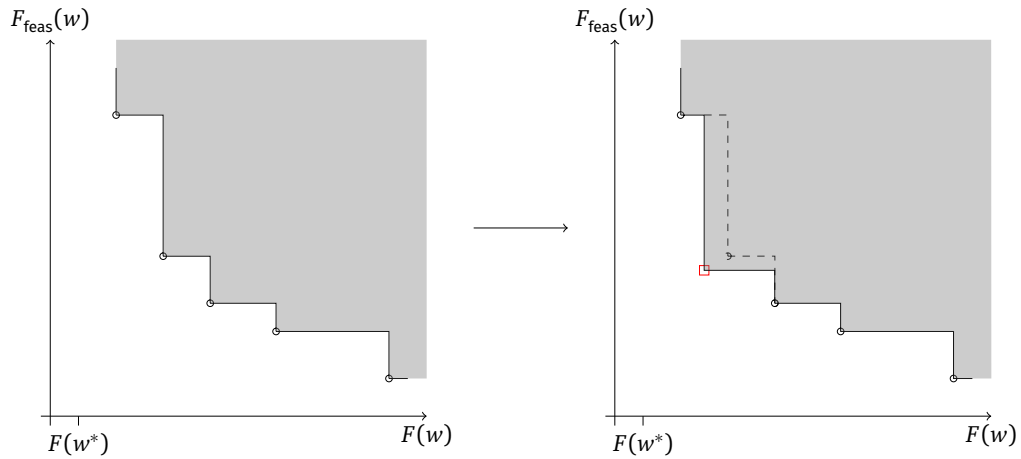


FIGURE 4.1 – Filtre et critère de validation du pas. La zone grise correspond à la zone de rejet de l'itération. L'itération  $\square$  est acceptée et le filtre est ajusté.

## 4.5 Méthode des points intérieurs

La méthode des points intérieurs a d'abord été développée dans le cadre de la résolution de problèmes convexes [FM90], qu'elle est capable de résoudre efficacement. On ce qui concerne les problèmes non linéaires, considérons le système de KKT (eq. 4.2) et en particulier la condition complémentaire (eq. 4.2e). Dans la méthode du point intérieur, cette condition est relaxée[BV04] (on dit parfois aussi qu'elle est « perturbée ») en définissant un **terme barrière**  $\nu > 0$  et on écrit :

$$\nabla_w \mathcal{L}(w^*, \lambda^*, \mu^*) = 0 \quad (4.7a)$$

$$G(w^*) = 0 \quad (4.7b)$$

$$H(w^*) < 0 \quad (4.7c)$$

$$\mu^* \geq 0 \quad (4.7d)$$

$$\mu_j^* H_j(w^*) = \nu, \quad j = 1, \dots, n_H \quad (4.7e)$$

Dans ce système modifié, on note en particulier que l'équation (eq. 4.7d) contient uniquement des inégalités **strictes** et que dans (eq. 4.7e), le membre de droite n'est plus nul. Il est clair qu'à mesure que  $t \rightarrow +\infty$ , ce problème se rapproche du problème initial. On peut également interpréter cette équation perturbée comme apparaissant dans les conditions de KKT du problème :

$$\min_w F^{\text{IP}}(w, \nu) = F(w) - \nu \sum_{i=1}^m \ln(-H_i(w)) \quad (4.8a)$$

$$\text{sous contraintes } G(w) = 0 \quad (4.8b)$$

Ce problème apparaît en remplaçant les conditions « dures » d'inégalité (eq. 4.1c) par des conditions « douces » dans la fonction coût. La fonction logarithme est utilisée comme fonction barrière : à mesure que la solution s'approche des frontières de l'ensemble de faisabilité,  $-\ln(-H_i(w)) \xrightarrow{H_i(w) \rightarrow 0^-} +\infty$  ce qui pénalise la fonction coût. Le problème résultant n'a plus

que des contraintes d'égalité. Les équations de stationnarité de ce problème sont :

$$\nabla_w \mathcal{L} = \nabla_w F(w) - \gamma \sum_{i=1}^m \frac{\nabla_w H_i(w)}{H_i(w)} - \lambda^T \nabla_w G(w) = 0 \quad (4.9a)$$

$$G(w^*) = 0 \quad (4.9b)$$

En définissant  $\mu_i = \frac{\gamma}{H_i(w)}$ , on obtient bien les équations (eq. 4.7). La méthode du point intérieur consiste ainsi à remplacer la résolution du problème (eq. 4.1) par la résolution de problèmes plus simples, ne possédant pas de conditions d'inégalité. On peut en effet montrer [Boy+94] que la suite des solutions des problèmes intermédiaires converge vers la solution du problème NLP que l'on cherche à résoudre.

Il est nécessaire de définir une stratégie de variation du terme barrière de sorte que  $\gamma \rightarrow 0$ , afin de garantir une convergence rapide vers la solution. Cette stratégie se présente sous la forme :

$$\gamma_{k+1} = \sigma_k \gamma_k, \quad 0 < \sigma_k < 1.$$

La valeur de  $\sigma_k$  peut être ajustée en fonction du progrès vers l'optimalité réalisé au cours de l'itération précédente.

Il existe beaucoup de variations autour de la méthode du point intérieur, mais les méthodes dites **primales-duales** sont les plus populaires pour la résolution de problèmes de contrôle optimal, en opposition aux méthodes **primales**. La différence entre les deux réside dans les itérations de Newton qui sont résolues, les deux problèmes étant en toute rigueur équivalents. Néanmoins, les itérations de Newton obtenue par la version primale-duale convergent plus rapidement vers une solution du fait d'un meilleur conditionnement, le terme  $\mu_i = \frac{\gamma}{H_i(w)}$  introduisant des non-linéarités.

Dans la plupart des solveurs, le problème est reformulé en introduisant des variables supplémentaires, dite variables auxiliaires (en anglais *slack variables*) qui permettent de réécrire le problème (eq. 4.1) de la manière suivante :

$$\min_{(w,s)} F(w) \quad (4.10a)$$

$$\text{sous contraintes } G(w) = 0 \quad (4.10b)$$

$$H(w) - s = 0 \quad (4.10c)$$

$$s \leq 0 \quad (4.10d)$$

La méthode des points intérieurs est équivalente à la résolution du problème sans contraintes d'inégalité suivant :

$$\min_{w,s} F(w) - \gamma \sum_{i=1}^m \ln(-s) \quad (4.11a)$$

$$\text{sous contraintes } G(w) = 0 \quad (4.11b)$$

$$H(w) - s = 0 \quad (4.11c)$$

$$s \leq 0 \quad (4.11d)$$

Cette reformulation du problème a des implications sur les solutions intermédiaires du

problème, puisqu'on peut initialiser le problème à une solution infaisable. Le corollaire est que seules les variables auxiliaires et les multiplicateurs de Lagrange sont faisables à tout instant, de sorte qu'il n'est pas possible d'interrompre les itérations de l'algorithme avant d'atteindre un point faisable [NW06]. Les méthodes SQP garantissent que les contraintes sont respectées à toute itération, ce qui les rend avantageuses pour des applications avec des contraintes d'exécution en temps réel.

Les méthodes décrites en section 4.2 telles que le *line-search* ou *trust-region* sont également applicables à la méthode des points intérieurs.

## 4.6 Programmation quadratique séquentielle

En programmation quadratique séquentielle, les conditions de KKT sont linéarisées autour de la solution courante  $w_i$ . Cette relaxation est équivalente à la résolution du problème de programmation quadratique (QP) intermédiaire (eq. 4.12).

$$\min_{\Delta w} \nabla_w F(w_i)^T \Delta w + \Delta w^T \nabla_w^2 \mathcal{L} \Delta w \quad (4.12a)$$

$$\text{sous contraintes } G(w_i) + \nabla G(w_i)^T \Delta w = 0 \quad (4.12b)$$

$$H(w_i) + \nabla H(w_i)^T \Delta w \leq 0 \quad (4.12c)$$

La solution de ce problème est notée  $(\Delta w^{QP}, \lambda^{QP}, \mu^{QP})$  où  $\lambda^{QP}$  et  $\mu^{QP}$  sont les variables duales du problème associées aux contraintes d'égalité et d'inégalité respectivement. La direction de recherche appliquée dans la méthode de Newton pour la résolution des conditions de KKT (eq. 4.2) est  $\Delta w_i = \Delta w^{QP}$  tandis que les multiplicateurs de Lagrange sont ceux de la solution du problème QP de sorte que :

$$\lambda_{i+1} = \lambda^{QP} \quad \text{et} \quad \mu_{i+1} = \mu^{QP}$$

Le problème est ensuite linéarisé autour de  $(\Delta w_i, \lambda^{QP}, \mu^{QP})$  pour donner un nouveau problème QP et ainsi de suite jusqu'à la convergence en une solution. Ainsi une suite de sous-problèmes QP est résolue pour trouver la solution des équations de KKT (eq. 4.2).

Les problèmes QP tels que (eq. 4.12) peuvent être résolus par la méthode des points intérieurs<sup>a</sup> ou par des méthodes d'ensemble actif. Si la Hessienne  $\nabla_w^2 \mathcal{L}$  est définie positive, la solution du problème est une solution globale [BV04]. Lorsque la Hessienne n'est pas définie positive, la solution n'est plus globale. Il peut être utile de mettre en œuvre une méthode de projection pour la rendre définie positive.

En fonction de la méthode de transcription utilisée pour le problème de contrôle optimal, le problème peut avoir une structure particulière. Dans le cadre du tir simple, la matrice des contraintes du problème est dense mais de taille plus petite que pour une transcription par tir multiple et encore plus que pour le tir simple. Il est cependant possible de réduire grandement la taille du problème par une étape de **condensing** et obtenir le même problème

a. Le solveur HPIPM (*High Performance Interior Point Method*) est un solveur QP utilisant la méthode des points intérieurs. Il est capable d'exploiter la structure du problème pour accélérer la résolution. Le solveur WORHP est un solveur SQP qui utilise également une méthode des points intérieurs pour la résolution des sous-problèmes QP.



que le problème obtenu par tir simple. En effet, une fois que le problème est linéarisé, les contraintes dynamiques le sont également, de sorte que :  $\Delta x_{k+1} = A_k \Delta x_k + B_k \Delta u_k$ . Il est alors possible d'éliminer toutes les variables intermédiaires introduites par tir multiple. Le problème peut alors être résolu avec un solveur générique. Cette approche peut s'avérer aussi performante que l'utilisation de solveurs spécifiques exploitant la structure du problème [Qui17]. De la même manière qu'on le décrit en section 3.2, il est possible de conserver certaines variables intermédiaires et de réduire partiellement le problème. On parle alors de **condensing partiel** [Axe15].

## 4.7 Calcul de quantités dérivées

Nous avons jusqu'à présent manipulé les quantités dérivées nécessaires à la résolution des problèmes NLP sans se soucier de leur calcul au sein des algorithmes. Ces quantités étant calculées de nombreuses fois au cours de la résolution d'un problème d'optimisation, il faut être capable de les calculer efficacement afin de garantir une résolution rapide. S'il est possible sur des fonctions simples de calculer analytiquement une expression de sa dérivée, il devient difficile et chronophage de le faire sans erreur pour des problèmes de taille importante comme on en rencontre pour la résolution des programmes non-linéaires. Des méthodes par ordinateur sont donc employées pour le calcul des dérivées [GW08].

Le calcul de dérivées intervient également dans le calcul des sensibilités d'un problème d'optimisation par rapport à un paramètre d'entrée, ainsi que dans les sensibilités des intégrateurs numériques.

### 4.7.1 Différences finies

Dans l'approche dite des **différences finies**, on approxime  $f'(x)$ , la dérivée de  $f : \mathbb{R} \rightarrow \mathbb{R}$  par un accroissement fini :

$$f'(x) \approx \Delta^h f(x) = \frac{f(x+h) - f(x)}{h} \quad \text{ou encore} \quad \Delta^h f(x) = \frac{f(x+h) - f(x-h)}{2h}$$

On choisit  $h$  « petit » de sorte que par continuité de  $f$ , l'erreur réalisée soit faible. Soit une fonction vectorielle  $f$  telle que :

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m \\ x \mapsto f(x) = [f_1(x), \dots, f_m(x)]^T, \quad \text{où } \forall i \in \llbracket 1, m \rrbracket, f_i(x) \in \mathbb{R}$$

Le calcul de  $\partial_{x_j} f_i(x)$  est réalisé par un accroissement fini, si bien que  $\nabla f$  nécessite  $n \times m$  accroissements finis :

$$\nabla f(x) \approx \left( \Delta^{h e_j} f_i(x) \right)_{1 \leq i \leq m, 1 \leq j \leq n}$$

Chacun des  $n \times m$  accroissements finis fait appel à la fonction  $f$  deux fois. Il est clair que cette approche mène à de nombreux calculs, ce qui peut rapidement impacter les performances de l'algorithme. En outre, cette méthode souffre de sensibilité numérique pour des petites valeurs de  $h$  [RMD17].

Il existe une méthode alternative des différences finies, se basant sur l'évaluation de la fonction en des valeurs complexes (lorsque c'est possible) [ST98]. Cette méthode se base

sur le développement de Taylor suivant :

$$f(x + ih) = f(x) + \sum_{k=1}^n (ih)^k \frac{f^{(k)}(x)}{k!} + o(h^n)$$

La partie réelle de ce développement ne contient que les dérivées d'ordre pair, tandis que la partie imaginaire ne contient que les dérivées d'ordre impair. On peut alors réaliser l'approximation suivante :  $f'(x) \approx \frac{\Im(f(x+ih))}{h}$ . Cette méthode est non seulement plus précise, mais elle ne nécessite qu'une seule évaluation de la fonction  $f$ .

## 4.7.2 Différentiation automatique

Une autre approche très performante pour le calcul de quantités dérivées est celle de la **différentiation automatique** [GW08]. L'avantage premier de cette méthode est qu'il est possible de calculer presque exactement (à précision machine près) les dérivées de fonctions pour un coût de calcul comparable à l'évaluation de la fonction elle-même.

Pour illustrer le fonctionnement de la différenciation automatique, considérons la procédure<sup>a</sup>  $f_{\text{info}}$  qui évalue la fonction  $f$ . L'appel à cette procédure peut nécessiter la déclaration de variables intermédiaires ou l'appel à d'autres procédures auxiliaires  $f_{\text{info},i}$ , mais il est toujours possible de les décomposer en une succession d'opérations élémentaires. En utilisant la règle de la chaîne, il est alors possible de calculer exactement la dérivée de la fonction  $f$  via les dérivées des opérations intermédiaires et des  $f_{\text{info},i}$ , ce qui permet de générer automatiquement une nouvelle procédure  $f'_{\text{info}}$  qui évalue exactement  $f'$ . Cette méthode est non seulement plus précise que les différences finies, mais le coût computationnel de la procédure  $f'_{\text{info}}$  est très souvent bien inférieur à un calcul par différences finies.

Plus précisément [And13 ; And+19], on peut représenter la fonction par un graphe orienté  $G$ , dont les nœuds sont les variables intermédiaires et opérations élémentaires. Les entrées de la fonction sont les nœuds à la « racine » de ce graphe tandis que les sorties sont les derniers nœuds. On peut ensuite déterminer un nouveau graphe  $G'$ , qui correspond à la dérivée de  $f$  par rapport à ses différentes entrées, en calculant les dérivées partielles des expressions de  $G$ . Concrètement, la dérivée de  $f(x_1, \dots, x_n)$  par rapport à une de ses entrées  $x_i$ ,  $i \in \llbracket 1, N \rrbracket$  s'écrit :

$$\frac{\partial f}{\partial x_i} = \left( \prod_{j=1}^N \frac{\partial f}{\partial c_j} \right) \frac{c_N}{x_i}$$

où les  $c_j$ ,  $j \in \llbracket 1, N \rrbracket$  sont des expressions élémentaires qui composent  $f$ .

Le calcul de  $\frac{\partial f}{\partial x_i}$  peut se faire de différentes manières puisqu'il est possible de le faire de la droite vers la gauche (on parle de *forward mode*) ou de la gauche vers la droite (on parle de *backward mode*)<sup>b</sup>.

Les langages de programmation dotés de surcharge opérateur sont adaptés pour l'implémentation de la différenciation automatique car ils permettent de réutiliser un code

a. Le terme « procédure » sert ici à distinguer les fonctions au sens informatique du terme, des fonctions au sens mathématique.

b. Pour être plus exact, il est possible d'évaluer les termes dans n'importe quel ordre. Déterminer l'ordre d'évaluation minimisant le nombre d'opérations est un problème NP-complet [Nau08].

existant en lui ajoutant des fonctionnalités lors de l'appel aux opérations élémentaires qui composent un programme informatique.

## 4.8 Notes et discussion

La programmation non-linéaire est un domaine de recherche très actif et employé dans de nombreux domaines. La brève introduction présentée ici ne saurait rendre compte de toutes les techniques, considérations, analyses et preuves nécessaires à la compréhension en profondeur de ce sujet éminemment important. De nombreuses considérations interviennent dans le choix d'une méthode de résolution, qui dépend des spécificités du problème à résoudre. Si certains choix peuvent être faits *a priori* sur les spécificités connues à l'avance du problème, il convient en général de tester différentes approches de résolution ainsi que les paramètres de réglage associés afin de garantir l'obtention de solutions satisfaisantes au problèmes (ou de solution tout court) en des temps « raisonnables »<sup>a</sup>. Les difficultés pratiques pour l'usage de différents solveurs telles que l'usage d'interfaces différentes, l'installation des dépendances logicielles ou les différences de langages de programmation pour leur implémentation sont néanmoins des freins. Des solutions logicielles telles que CasADi [And+19] (différentiation automatique), Rokit [Gil+20] (modélisation de problèmes de contrôle optimal et interfaçage avec des solveurs dédiés) ou UNO [VL24] (solveur modulaire, qui permet d'adapter la méthode de résolution au problème) qui unifient différents solveurs et méthodes de résolution sous une bannière unique sont tout aussi utiles que nécessaires.

En outre des méthodes de résolution en elles-mêmes, la question des bibliothèques de calcul utilisées est importante. Nous avons vu que la résolution des systèmes linéaires est un ingrédient essentiel pour l'optimisation, qui n'est pas traité ici par concision. De nombreux systèmes linéaires intermédiaires doivent être résolus. On distingue en général les algèbres *creuses* et les autres (on dit qu'elles sont *denses*). Dans le premier cas, elles désignent des bibliothèques d'algèbre linéaire (BLAS) capables d'exploiter la structure creuse d'un problème en utilisant des représentations adaptées des matrices (*compressed column storage*) qui permettent d'économiser la représentation en mémoire de nombreux zéros, et l'emploi de méthodes adaptées pour leur factorisation et pour le calcul matriciel. L'ouvrage [ST23] traite en profondeur les algorithmes creux pour la résolution de problèmes linéaires. En outre, certaines bibliothèques de calcul sont optimisées pour exploiter toutes les fonctionnalités des processeurs ou unités de calcul cibles : stockage des matrices dans la mémoire cache, utilisation d'instructions en parallèle (SIMD)<sup>b</sup> [Fri+17].

Nous avons présenté deux stratégies de globalisation pour l'optimisation que sont les fonctions mérite et les filtres pour le test des itérations de la méthode de Newton. Beaucoup de considérations entrent en jeu pour l'usage de ces méthode en fonction de la méthode de résolution du programme non-linéaire (SQP ou points intérieurs) ainsi que le mécanisme de globalisation employé. L'effet Maratos [NW06] décrit comment des itérations qui permettent pourtant de converger vers la solution du problème peuvent être rejetées par les stratégies

---

a. Dans le contexte de l'utilisation en ligne, comme pour la résolution de problèmes de contrôle optimal à horizon glissant, il est essentiel que la solution soit déterminée avant que celle du problème suivant ne devienne nécessaire. De plus, un temps de résolution court permet de tester un plus grand nombre de configurations des paramètres.

b. SIMD : *Single Instruction Multiple Data*.

de globalisation. Il est nécessaire de mettre en place des dispositifs supplémentaires pour s'assurer de la convergence de l'algorithme de résolution, comme des méthodes de correction au second ordre, l'emploi de pas relaxé (c'est-à-dire qui ne permettent pas d'améliorer la solution au sens de la minimisation de la fonction coût ou la faisabilité de la solution, mais qui permettent de sortir de minimums locaux) et la mise en œuvre de phases de restauration lorsque la taille du pas devient trop petite au fil des rejets de la stratégie de globalisation. L'ouvrage [NW06] est un classique présentant ces considérations.



DEUXIÈME PARTIE

**ESTIMATION POUR LA CONDUITE DES REP**



## CHAPITRE 5

# PHYSIQUE ET MODÉLISATION DES REP POUR L'ESTIMATION ET LA COMMANDE

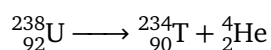
*En France, on n'a pas de pétrole, mais on a des idées.*

— Agence pour les Économies d'Énergie, 1974

### 5.1 Historique et contexte

En 1903, Henri Becquerel et Marie et Pierre Curie reçoivent le prix Nobel de physique pour leur travaux sur la radioactivité. Le premier a mis en évidence la radioactivité de l'uranium <sup>a</sup>, le couple des seconds a mis en évidence la radioactivité du radium. De nombreux physiciens entreprendront d'étudier ce nouveau phénomène, qui sera l'objet de nombreuses récompenses et prix (Nobel entres autres) en physique et en chimie.

On dit d'un noyau atomique qu'il est *radioactif*, s'il peut se transformer spontanément en un autre élément ou isotope que l'on appelle *noyau fils* [Rad10]. Cette réaction nucléaire libère un rayonnement sous forme de particule et d'énergie (photon ou énergie cinétique). On distingue différents rayonnements en fonction de la nature de la particule émise. Concernant l'uranium, la désintégration de l'uranium 238 (isotope naturel le plus abondant sur Terre) mise en évidence par Henri Becquerel correspond à la réaction nucléaire suivante :



Il s'agit d'une désintégration  $\alpha$  puisqu'elle produit un noyau d'hélium. Il existe d'autres types de rayonnement issus de réactions nucléaires tels que l'émission de protons, neutrons, neutrinos... qui caractérisent des réactions nucléaires différentes (cf sous-section 5.3.1). On s'intéresse dans ce manuscrit aux réactions de fissions et de désintégration qui interviennent dans les réacteurs nucléaires à fission. Dans ce cas, les noyaux impliqués dans les réactions sont des isotopes lourds tels que l'uranium ou le plutonium.

En 1938, le phénomène de fission induite est découvert par les allemands Lise Meitner, Fritz Strassmann et Otto Hahn [MSH38] : les auteurs montrent qu'il est possible de provoquer la fission du thorium en le bombardant de neutrons. Puis en 1939, Frédéric Joliot, Hans Halban, Lew Kowarski et Francis Perrin mettent en évidence l'émission de neutrons lors la fission de l'uranium [Jol39]. L'idée de la *réaction en chaîne* germe alors : les neutrons produits par fission servent ensuite à induire la fission d'autres noyaux, et ainsi de suite. Comme la réaction libère de l'énergie, différentes applications de la fission vont rapidement

---

a. Il a remarqué qu'une plaque photographique, pourtant vierge et recouverte de papier noir, stockée dans le même tiroir que des sels d'uranium a été marquée d'un rayonnement.



voir le jour : bombe atomique et production d'énergie<sup>a</sup>. La première met en œuvre une réaction de fission incontrôlée capable d'une puissance jusqu'alors inégalée, ainsi que le prouvent les tragiques bombardements d'Hiroshima et Nagasaki les 6 et 9 août 1945. La seconde application, bien moins funeste, nécessite de contrôler cette réaction afin de produire de l'énergie de manière plus diffuse dans le temps et plus sûre.

En France, l'intérêt croissant pour l'énergie nucléaire mène à la création du CEA (Commissariat à l'énergie atomique) dès le 18 octobre 1945 à la demande du Général de Gaulle. La première divergence d'un réacteur français est réalisée en janvier 1956. Après la construction d'une dizaine de réacteurs de puissance de ce type, la filière UNGG est rapidement abandonnée pour les nouveaux réacteurs au profit de la filière REP (Réacteur à Eau Pressurisée) sous licence Westinghouse.

En 1973, l'OPEP (Organisation des Pays Producteurs et Exportateurs de Pétrole) décrète un embargo sur Israël et ses alliés. La multiplication par quatre des prix du pétrole provoque le premier choc pétrolier. En réponse, le premier ministre Pierre Messmer annonce le 6 mars de l'année suivante le lancement du « plan Messmer ». Il est décidé de la construction de centrales nucléaires pour une puissance de 13 GWe (soit un doublement de la puissance du parc alors en exploitation par EDF) dans le but de réduire la dépendance énergétique de la France. Aujourd'hui encore, la France est de très loin le pays ayant le plus recourt au nucléaire pour sa production électrique : 56 réacteurs pour 18 centrales produisent plus de 70% de l'électricité consommée sur le territoire.

En 2021, l'AIEA (Agence Internationale de l'Énergie Atomique) recensait 441 réacteurs de puissance en fonctionnement dans le monde<sup>b</sup>.

## 5.2 Fonctionnement d'une centrale nucléaire et pilotage d'un réacteur

Toutes les centrales nucléaires fonctionnent sur le même principe : une chaudière nucléaire où se produit une réaction en chaîne chauffe un fluide caloporteur. Cette chaleur permet de vaporiser de l'eau entraînant une turbine reliée à l'arbre d'un générateur électrique<sup>c</sup>. Pour chaque chaudière nucléaire, on parle de *tranche*, une centrale pouvant en comporter plusieurs. Dans les réacteurs à eau pressurisée, 3 circuits différents sont considérés :

- le circuit primaire est celui du réacteur, où le caloporteur (eau) est chauffé. Le caloporteur transmet sa chaleur au sein d'un générateur de vapeur (GV) à l'eau du circuit secondaire. La pression de l'eau dans le cœur est d'environ 155 bars et sa température est de 300°C en entrant dans le cœur et d'environ 330°C en sortant du cœur. À ces conditions de pression et de température, l'eau du circuit primaire est principalement à l'état liquide. Des pompes assurent l'écoulement de l'eau dans le circuit primaire au débit de 4 m<sup>3</sup>/s.

---

a. Les deux ne sont pas mutuellement exclusifs. Certains sous-marins nucléaires combinent les deux : leur propulsion est assurée par un réacteur nucléaire (d'où leur appellation) et sont capables de lancer des missiles équipés de têtes nucléaires.

b. IAEA Annual Report for 2022 : <https://www.iaea.org/publications/reports/annual-report-2022>

c. Dans les réacteurs à eau bouillante, le caloporteur est vaporisé directement tandis que dans les réacteurs à eau pressurisée le caloporteur échange sa chaleur à un autre circuit où l'eau est vaporisée.

- Le circuit secondaire est celui de la turbine. La vapeur, générée dans les générateurs de vapeur à l'interface des circuits primaire et secondaire, sert à fournir un travail en entraînant l'arbre du générateur électrique. Cette vapeur est condensée dans un condenseur à l'interface des circuits secondaire et tertiaire.
- Le circuit tertiaire est le circuit de refroidissement. Il sert à évacuer la chaleur non utile produite par la centrale par le biais de tour de refroidissement (aéroréfrigérant) et/ou d'eau.

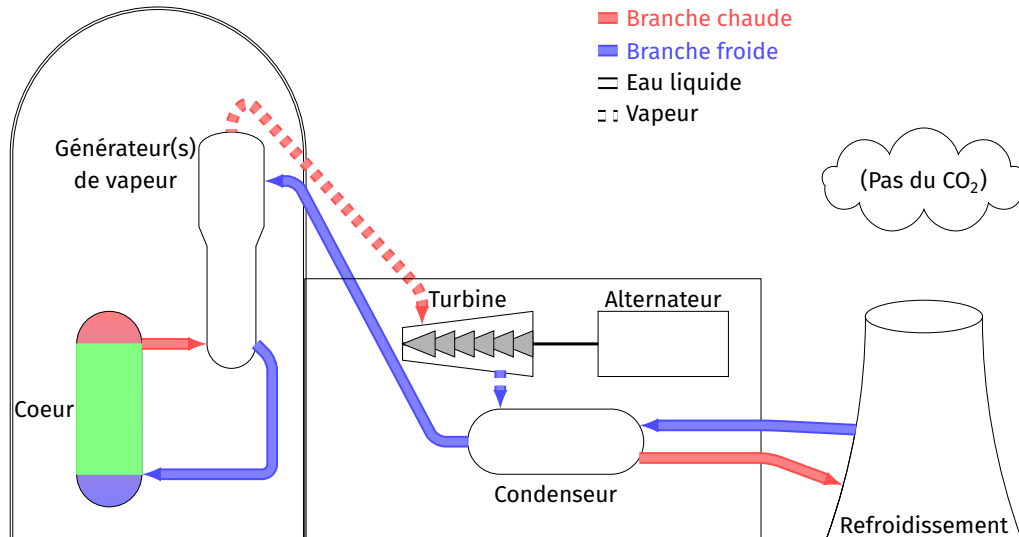


FIGURE 5.1 – Représentation simplifiée d'une centrale nucléaire.

Une centrale nucléaire est une machine thermique (circuit secondaire) dont la source chaude est le circuit primaire et la source froide est le circuit tertiaire. En fonction du réacteur considéré, la puissance thermique délivrée par le réacteur varie de 2,7 GW (Graveline) à 4,3 GW (EPR de Flamanville-3), pour des puissances électriques brutes<sup>a</sup> correspondantes de 920 à 1650 MW. L'efficacité thermodynamique d'une tranche, soit le rapport de la puissance électrique brute produite sur la puissance thermique, est ainsi de l'ordre de 35%, c'est-à-dire que seul un tiers de l'énergie thermique produite est transformée en électricité. Le reste est dissipé à l'environnement extérieur. D'autres conceptions de réacteurs visent des températures de fonctionnement plus élevée afin d'augmenter l'efficacité du cycle thermodynamique du système.

Le comportement du système est influencé par l'environnement extérieur de deux manières : d'une part la chaleur dissipable par le circuit tertiaire, d'autre part le besoin d'équilibrer la production et consommation d'électricité, assuré par les systèmes de réglage (cf Contexte).

On peut piloter le réacteur et faire varier sa puissance par le biais de deux actionneurs : du bore dissout dans l'eau du circuit primaire et l'insertion des barres de contrôle, qui bloquent les neutrons et limitent le nombre de fissions qu'ils provoquent. Les opérateurs effectuent des variations de puissance de l'ordre de quelques pourcents de puissance nominale (PN) par minute.

a. La puissance électrique *nette* correspond à la puissance électrique effectivement fournie au réseau, c'est-à-dire la puissance électrique brute moins la puissance consommée par la centrale pour son fonctionnement (pompes, systèmes auxiliaires, ...).

- ① Branche froide
- ② Branche chaude
- ③ Barres de contrôle
- ④ Assemblages de combustible

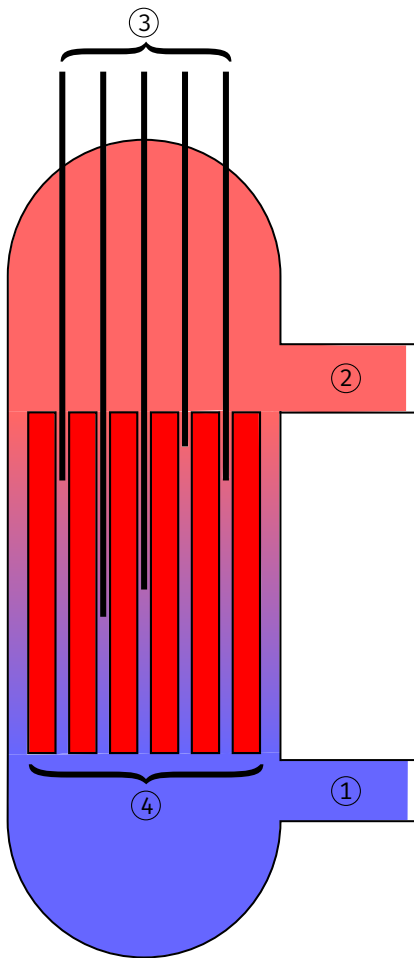


FIGURE 5.2 – Cuve du réacteur.

- ① Branche froide primaire
- ② Branche chaude primaire
- ③ Branche froide secondaire
- ④ Branche chaude secondaire
- ⑤ Faisceau tubulaire
- Vapeur

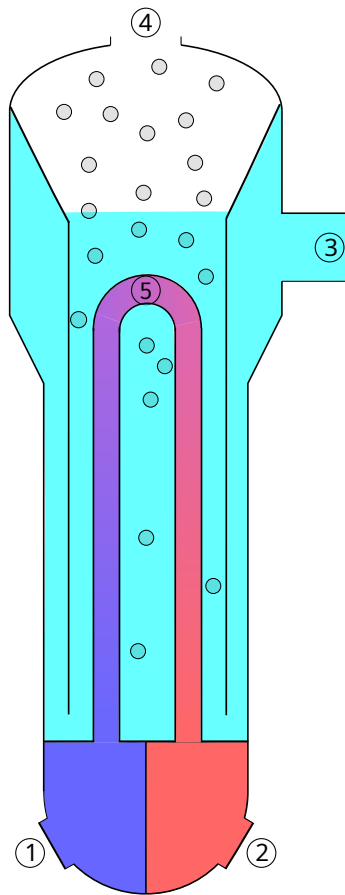


FIGURE 5.3 – Générateur de vapeur.

### 5.2.1 Cuve réacteur et cœur

Le réacteur (figure 5.2) est une cuve qui contient le cœur (qui décrit le volume où se trouve le combustible nucléaire) dans laquelle de l'eau est pompée. Le combustible est formé de pastilles en céramique contenant de l'uranium enrichi voire également du plutonium (combustible MOX). Les pastilles sont empilées en *crayons* qui sont ensuite entourés d'une gaine en zirconium protégeant les pastilles. Les crayons sont disposés en groupes appelés assemblages. Enfin, les assemblages sont disposés dans le cœur. L'eau s'écoule dans les assemblages, en contact avec la gaine des crayons.

On appelle réflecteurs des plaques en acier disposées autour du cœur afin de limiter les fuites de neutrons vers l'extérieur. Des réflecteurs lourds sont également présent en haut et en bas du cœur.

En projetant la puissance dégagée en chaque point des assemblages sur l'axe vertical, on obtient la puissance axiale qui exprime la puissance dégagée en fonction de l'altitude.

**Définition 5.1 (Déséquilibre axial de puissance : Axial Offset)**

*En notant  $P_{top}$  et  $P_{bottom}$  les puissances dégagées respectivement dans les moitiés supérieure et inférieure du cœur, on définit l'axial offset (AO) :*

$$AO = \frac{P_{top} - P_{bottom}}{P_{top} + P_{bottom}}$$

L'AO est une une quantité essentielle à la conduite de la tranche qui permet de donner une indication de la forme de la distribution de puissance. Si davantage de puissance est dégagée dans la partie inférieure, l'AO est négatif et réciproquement. En règle général, du fait que la modération des neutrons est inversement corrélée à la température (voir sous-section 5.3.3), l'AO a tendance à être négatif.

### 5.2.2 Générateurs de vapeurs

L'eau échauffée au sein du réacteur circule dans plusieurs générateurs de vapeur (GV) (figure 5.3) en parallèle. Un réacteur comporte en réalité plusieurs branches chaudes, chacune acheminant le caloporteur dans un générateur de vapeur. Dans un GV, l'eau de la branche chaude primaire s'écoule dans un faisceau tubulaire qui contient des milliers de tubes où s'écoule l'eau du circuit primaire. Ces nombreux tubes permettent d'augmenter la surface d'échange pour permettre à l'eau du circuit secondaire de se réchauffer et d'être vaporisée. La vapeur de l'ensemble des GV est ensuite collectée et envoyée dans une turbine pour y fournir un travail servant à entraîner l'arbre de la génératrice [Reu12].

L'eau ayant transmis sa chaleur au sein du GV est ensuite réinjectée dans le réacteur pour y être échauffée à nouveau. La séparation en circuit primaire et circuit secondaire est la particularité des réacteurs à eau sous pression, de sorte que l'eau dans le circuit primaire est censée être toujours à l'état liquide et ne circule que dans le réacteur et les GV, ce qui permet d'éviter la contamination de la turbine par des éléments radioactifs rencontrée dans les réacteur à eau bouillante [Hon+84].

La modélisation d'un générateur de vapeur est complexe et n'est pas l'objet des travaux présentés ici, où notre attention se porte sur le cœur du réacteur.

## 5.3 Physique des réacteurs

La modélisation d'un cœur de réacteur nécessite de faire appel à de nombreuses disciplines. Il faut considérer les phénomènes suivants :

- les réactions nucléaires ayant lieu en chaque point du cœur,
- l'échauffement du combustible du fait des réactions nucléaires,
- l'écoulement du caloporteur dans le cœur.

Un grand nombre de réactions nucléaires est à prendre en compte pour chaque isotope présent dans le cœur, nécessitant de connaître toutes les interactions possibles, leur probabilité d'occurrence (ou rendement) et ce en chaque point du domaine. Il est également important de modéliser l'évolution du flux neutronique (énergie, direction, population) afin de pouvoir prévoir le comportement de la réaction en chaîne. Une modélisation exacte des milliers de réactions nucléaires possibles et l'évolution de toutes les populations de noyaux est impossible, c'est pourquoi il est courant de faire appel à des méthodes Monte-Carlo (probabilistes) qui simplifient la modélisation [MU49]. Il est important de modéliser la diffusion des neutrons dans le milieu ainsi que leur interaction avec la matière au sein du réacteur, ce que permet de décrire l'équation de Boltzmann.

La modélisation de l'échauffement du combustible sert à concevoir le combustible et en particulier la gaine des crayons d'assemblage, qui est soumise à des conditions de pressions et de température qui varient au cours du fonctionnement du réacteur. Son échauffement impacte également les réactions nucléaires qui ont lieu au sein du combustible ainsi que les transferts thermiques avec le caloporteur. La température du caloporteur impacte également le comportement de la réaction en chaîne.

### 5.3.1 Réactions nucléaires

Les réactions nucléaires impliquent les nucléons des noyaux atomiques, à la différence des réactions chimiques qui reposent sur les interactions entre les nuages d'électrons des atomes. Les réactions nucléaires peuvent être spontanées ou provoquées par le bombardement du noyau par une particule incidente (proton, neutron, autre noyau ou énergie). La fusion nucléaire correspond à la formation d'un noyau plus lourd à partir de noyaux plus légers, elle n'a pas lieu dans les REP <sup>a</sup>.

La section efficace *totale* d'une réaction nucléaire entre une particule incidente et une particule cible caractérise la probabilité qu'une réaction du tableau 5.1 ait lieu en fonction de l'énergie de la particule incidente.

#### **Définition 5.2 (Section efficace microscopique)**

*La section efficace microscopique, ou aire équivalente d'interaction, notée  $\sigma$  (exprimée en barns =  $10^{-24} \text{cm}^2$ ) s'interprète comme la « taille » de la particule cible lors de son bombardement par la particule incidente. Plus elle est grande et plus une collision (ou interaction) est probable.*

a. Le développement de technologies industrielles exploitant la fusion nucléaire est l'objet de nombreux travaux du fait de ses avantages : abondance du combustible et sûreté. Les conditions extrêmes de température et de pression nécessaires sont cependant très difficiles à produire et à maintenir.

Réaction nucléaire	Formule	Commentaire
Désintégration $\alpha$	${}^A_Z X \longrightarrow {}^{A-4}_{Z-2} X + {}^4_2 \text{He}$	
Désintégration $\beta^+$	${}^A_Z X \longrightarrow {}^A_{Z+1} X + e^- + \nu$	$e^-$ est un électron, $\nu$ est un neutrino
Désintégration $\beta^-$	${}^A_Z X \longrightarrow {}^A_{Z-1} X + e^+ + \bar{\nu}$	$e^+$ est un positon, $\bar{\nu}$ est un antineutrino
Fission induite	${}^A_Z X + n \longrightarrow X_1 + X_2 + m n$	$X_i$ sont les noyaux fils, $m \in \mathbb{R}$
Fission spontanée	${}^A_Z X \longrightarrow X_1 + X_2 + m n$	
Capture neutronique	${}^A_Z X + n \longrightarrow {}^A_{Z-1} X + \gamma$	
Diffusion résonnante	${}^A_Z X + n \longrightarrow {}^A_Z X + n$	élastique ou inélastique

TABLE 5.1 – Réactions nucléaires ayant lieu au sein d'un REP [CEA15].

**Définition 5.3 (Section efficace macroscopique)**

La section efficace macroscopique notée  $\Sigma$  (exprimée en  $\text{cm}^{-1}$ ) résulte du produit de la section efficace microscopique par la concentration de l'isotope cible dans le milieu. Il s'agit d'une densité de probabilité.

Pour chaque nucléide, la probabilité pour chaque type de réaction est caractérisée par une section efficace qui dépend de l'énergie de la particule incidente. Pour une réaction nucléaire en particulier, on considère en général les sections efficaces *partielles* relatives à la fission  $\sigma_f$ , à la diffusion  $\sigma_s$  ( $s$  pour *scattering*) et à la capture  $\sigma_c$ , liées à la section efficace totale par la relation suivante :

$$\sigma = \sigma_f + \sigma_c + \sigma_s$$

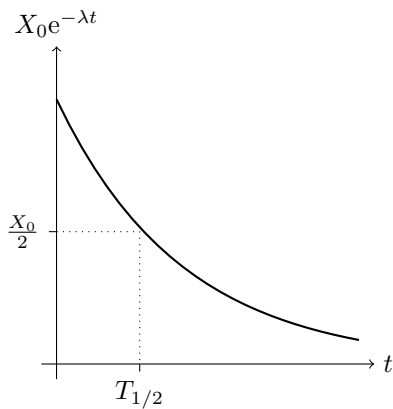
On se contente en général de la section efficace totale, dans la mesure où l'on connaît la réaction prédominante. Les données nucléaires sont disponibles dans des bases de données continuellement alimentées et entretenues par des expériences menées dans le monde entier [Plo+20; Bro+18].

**Désintégration**

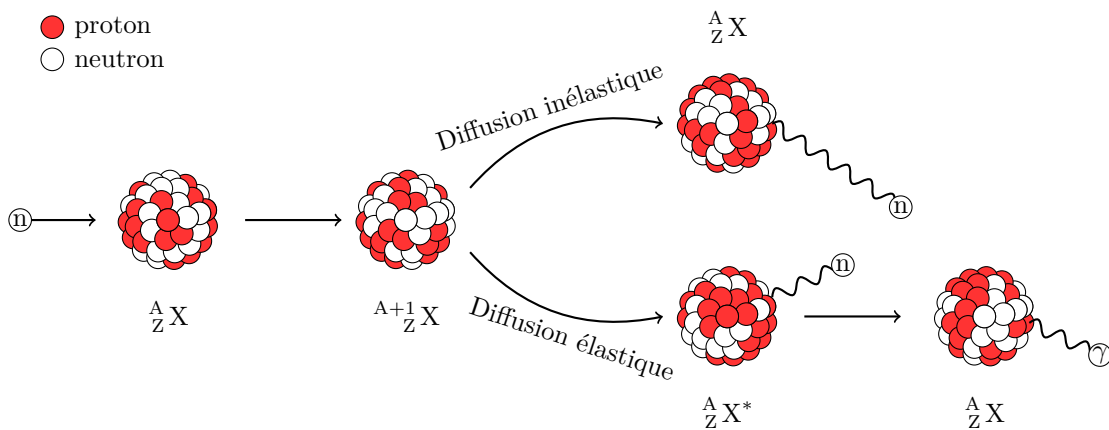
La désintégration d'un noyau se produit spontanément. La décroissance radioactive d'un élément se caractérise par sa période de demi-vie  $T_{1/2}$ , qui est le temps nécessaire pour que disparaisse la moitié d'un échantillon radioactif (figure 5.4a). La constante de désintégration  $\lambda$  permet également de caractériser la désintégration d'un isotope radioactif, dont la quantité  $x(t)$  évolue selon la loi  $\dot{x} = -\lambda x$ . La période de demi-vie et la constante de désintégration sont liées par la relation :  $\lambda = \frac{\ln(2)}{T_{1/2}}$ .

**Diffusion : résonnante élastique, inélastique et potentielle**

Dans la diffusion résonnante, le neutron incident est absorbé pour former un noyau composé puis le noyau éjecte un neutron pour revenir à l'isotope initial. Le neutron peut avoir la même énergie que le neutron incident (diffusion résonnante élastique) ou avoir une énergie plus faible (inélastique), le noyau se trouvant dans un état excité (rayonné sous la forme de photon lors de la désexcitation de ce dernier) (figure 5.4b). Lorsque le neutron « rebondit » sur le noyau sans être absorbé, on parle de diffusion potentielle



(a) Évolution de la population d'un échantillon radioactif par désintégration. Au bout du temps  $T_{1/2}$ , la moitié de la quantité initiale  $X_0$  de l'échantillon s'est désintégrée.



(b) Diffusions résonantes élastique et inélastique.

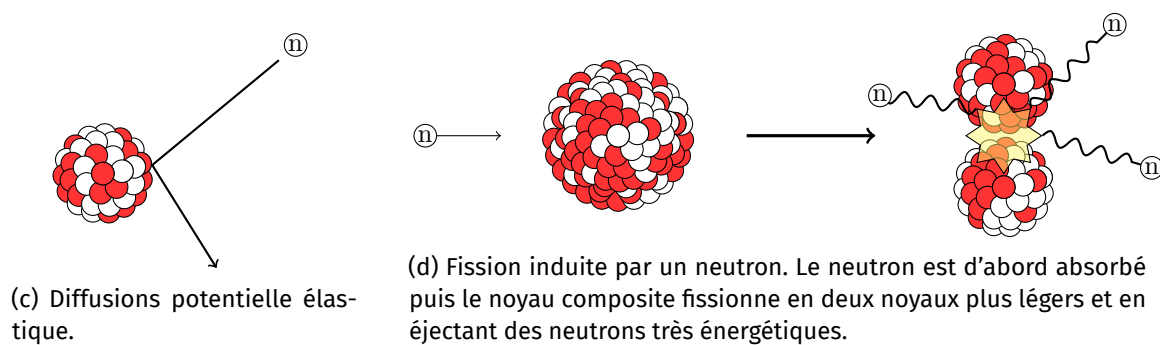


FIGURE 5.4 – Réactions nucléaires.

élastique (figure 5.4c).

### Fission nucléaire

Lors de la fission nucléaire un noyau lourd libère de l'énergie en formant deux noyaux plus légers ainsi qu'un nombre variable de neutrons (figure 5.4d). En moyenne 2,3 neutrons sont éjectés lors de la fission de l'uranium 235. La fission peut être spontanée ou induite, lorsqu'un neutron incident est absorbé par un noyau pour former un noyau composé. Le noyau est déformé et l'équilibre entre les forces coulombiennes (qui tendent à éloigner les protons entre eux) et la force de cohésion nucléaire (qui agit sur de courtes distances) peut-être rompu et la cohésion du noyau n'est plus assurée.

En notant  $p$  la probabilité qu'un neutron au sein du réacteur provoque une fission du combustible et si la fission du combustible produit en moyenne  $m$  neutrons, on définit le facteur de multiplication  $k = mp$  comme étant le nombre moyen de fissions qu'engendre une seule fission. Le nombre de fissions ayant lieu dans le cœur suit donc une loi géométrique de raison  $k$ . Le temps entre deux générations de neutrons est très court (de l'ordre de  $10^{-5}$ s). On distingue alors trois cas :

- $k > 1$  : la réaction s'emballe puisque le nombre de fissions croît rapidement. On dit que la réaction est sur-critique.
- $k < 1$  : la réaction se ralentit. On dit que la réaction est sous-critique.
- $k = 1$  : il y a toujours le même nombre de réactions de fissions, la réaction est dite critique.

Le facteur de multiplication n'est pas une grandeur très commode à manipuler, dans la mesure où elle doit être maintenue proche de 1 pour assurer la criticité du cœur, sans quoi la réaction s'emballe ou s'éteint très rapidement. La *réactivité du cœur*  $\rho$  s'écrit :

$$\rho = \frac{k - 1}{k}$$

Elle s'exprime en pcm ( $10^{-5}$ ) mais est adimensionnée. Les mêmes raisonnements que pour  $k$  s'appliquent :  $\rho = 0$  indique la criticité,  $\rho > 0$  la sur-criticité et  $\rho < 0$  la sous-criticité.

### 5.3.2 Équations de la neutronique

#### Évolution de la quantité de neutrons : équation de la cinétique

Au cours du fonctionnement du réacteur, les neutrons émis par les nombreuses réactions de fission du combustible permettent d'entretenir la réaction. En raisonnant à l'échelle du cœur tout entier, on est amené à formuler l'équation de la cinétique ponctuelle, qui caractérise l'évolution du nombre de neutrons dans le cœur. Il est nécessaire de prendre en compte les différentes voies de production des neutrons :

- voie principale : fission du combustible.
- seconde voie : neutrons retardés. Certains noyaux (appelés **précurseurs**) formés par capture neutronique subissent une désintégration  $\beta^+$ , puis le noyau fils émet des neutrons **retardés** [Reu12] (le retard est associé à la période de désintégration).



Considérons une quantité initiale de neutrons prompts  $n_0$  et de neutrons retardés  $c_0$  et notons  $l$  la durée de vie moyenne d'un neutron dans le cœur critique avant qu'il n'interagisse avec un noyau. Le nombre de neutrons dans le cœur évolue selon la loi suivante :

$$\frac{dn}{dt} = \frac{\rho - \beta}{l} n + \sum_i \lambda_i c_i \quad (5.1a)$$

$$\frac{dc_i}{dt} = \frac{\beta_i}{l} n - \lambda_i c_i \quad (5.1b)$$

$\lambda_i$  est la période de décroissance radioactive du précurseur  $i$  de la chaîne de désintégration,  $\beta_i$  la proportion des neutrons retardés du précurseur  $i$  parmi l'ensemble des neutrons issus des fissions (en pcm).

### Évolution de la distribution spatiale du flux : équation de la diffusion

L'équation de la diffusion permet de rendre compte de la distribution des neutrons dans l'espace. En chaque volume infinitésimal de l'espace, il est nécessaire de faire le bilan neutronique des neutrons qui s'y trouvent en décomptant les neutrons émis par les sources radioactives, les neutrons qui entrent dans ce volume et ceux qui s'échappent de ce volume. Ce bilan correspond à l'équation de Boltzmann.

#### Définition 5.4 (Densité neutronique)

*En chaque volume infinitésimal  $dV(x)$  de l'espace, le nombre de neutrons dans ce volume est  $n dV$ , où  $n$  est la densité neutronique au point  $x$ .*

#### Définition 5.5 (Flux neutronique)

*En notant  $v$  la vitesse des neutrons, le flux neutronique est la quantité  $\Phi = nv$ .*

#### Définition 5.6 (Spectre en énergie des neutrons, groupes d'énergie)

*Les neutrons issus de la fission sont produits à des énergies différentes. Le spectre en énergie des neutrons correspond à la distribution en énergie des neutrons. La théorie multi-groupe consiste à discrétiser le spectre en sous-intervalle  $[0, E_1, \dots, E_N]$ . La théorie à un groupe revient à considérer des neutrons monocinétiques.*

#### Définition 5.7

*L'équation de la diffusion des neutrons s'écrit [Reu12] :*

$$\frac{1}{v} \frac{\partial \Phi}{\partial t} = D \Delta \Phi - \Sigma_a \Phi + S \quad (5.2)$$

*$D$  est un terme correctif qui permet de corriger l'erreur induite par l'hypothèse d'isotropie des diffusions,  $\Delta$  décrit l'opérateur laplacien,  $\Phi$  est le flux neutronique et  $S$  le terme de sources radioactives dans le volume. Cette équation caractérise le flux neutronique et la diffusion des neutrons.*

### 5.3.3 Modération des neutrons et effet modérateur

Les neutrons issus de la fission du combustible nucléaire sont produits à des énergies de l'ordre du MeV (voire davantage) soit des vitesses de l'ordre de  $10^4$  km/s (cf figure 5.5). À

cette énergie, la section efficace de l'uranium 235 (figure 5.6) est trop faible pour qu'un nombre suffisant de fissions ait lieu pour entretenir la réaction, les neutrons ayant plutôt tendance à être absorbés par l'uranium 238 (plus abondant) sans engendrer de fission. Il est alors nécessaire<sup>a</sup> de ralentir les neutrons à une vitesse plus propice à la fission du combustible. On utilise un matériau dit *modérateur* pour ralentir les neutrons, qui en « rebondissant » sur le modérateur perdent une partie de leur énergie cinétique par interaction élastique. On parle de « thermalisation » des neutrons et à cette énergie, leur vitesse est de l'ordre du km/s.

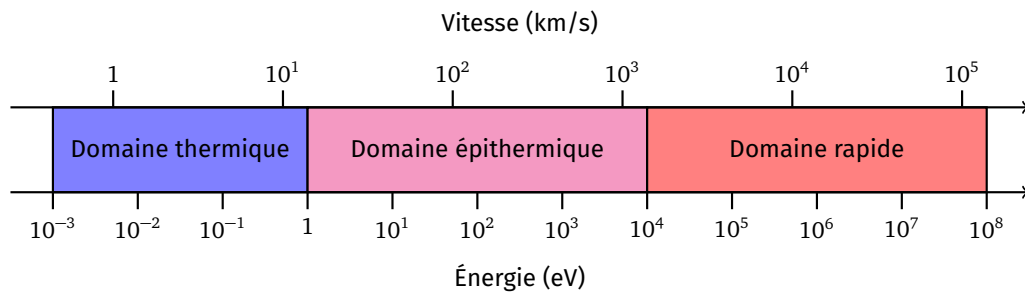


FIGURE 5.5 – Domaines d'énergie des neutrons.

Une seule collision ne suffit pas : le nombre moyen de collisions nécessaires à la thermalisation d'un seul neutron est typiquement de plusieurs dizaines [Reu12]. Plus un modérateur nécessite de collisions pour thermaliser les neutrons, moins il doit les absorber, ceci afin d'éviter que les neutrons soient absorbés avant qu'ils n'arrivent à la bonne énergie et provoquent une fission. Les modérateurs constitués d'éléments légers nécessitent le moins de collisions en moyenne. Puisqu'ils sont peu absorbants et abondants (donc peu chers), le carbone (graphite) et l'hydrogène (eau liquide) sont couramment utilisés en tant que modérateur. Dans les REP, l'eau a à la fois le rôle de caloporteur et de modérateur<sup>b</sup>. Le protium dans les molécules d'eau peut capturer des neutrons pour former du deutérium <sup>2</sup>H, il est donc nécessaire d'enrichir le combustible à 3% d'uranium 235 pour compenser ces absorptions. D'autres modérateurs moins absorbants comme le carbone et l'eau lourde permettent d'utiliser de l'uranium naturel (0.75% d'uranium 235), mais ils sont plus coûteux.

Sous l'effet de la variation de la température de l'eau et de la dilatation thermique, la densité du fluide est modifiée. L'*effet modérateur* caractérise la variation de réactivité du cœur sous l'effet de la variation de densité de l'eau. Cet effet peut-être positif (contribution à la réactivité) ou négatif (baisse de la réactivité). Les REP sont conçus de sorte à ce que l'effet modérateur soit négatif [Gas24], ce qui procure naturellement une régulation du cœur puisqu'un échauffement entraîne un ralentissement de la réaction, et inversement.

Considérons une élévation de la température du fluide, à la suite d'une accélération de la réaction en chaîne. La baisse de la densité du fluide réduit les chances de collision entre un noyau d'hydrogène et un neutron, donc une baisse du flux neutronique thermique<sup>c</sup>, si

a. Nécessaire dans les réacteurs à neutrons thermiques (choix de l'uranium comme combustible). Dans les réacteurs à neutrons rapides, on ne ralentit pas les neutrons et le combustible est choisi de sorte à subir des fissions dans le domaine rapide.

b. Cette conception permet également d'assurer la sûreté du réacteur, puisqu'en cas de perte de refroidissement (fuite), l'absence de modérateur interrompt la réaction.

c. Une légère augmentation du flux est également à noter dans la mesure où la baisse de la densité de l'eau

bien que le nombre de fissions est réduit et la réaction ralentit.

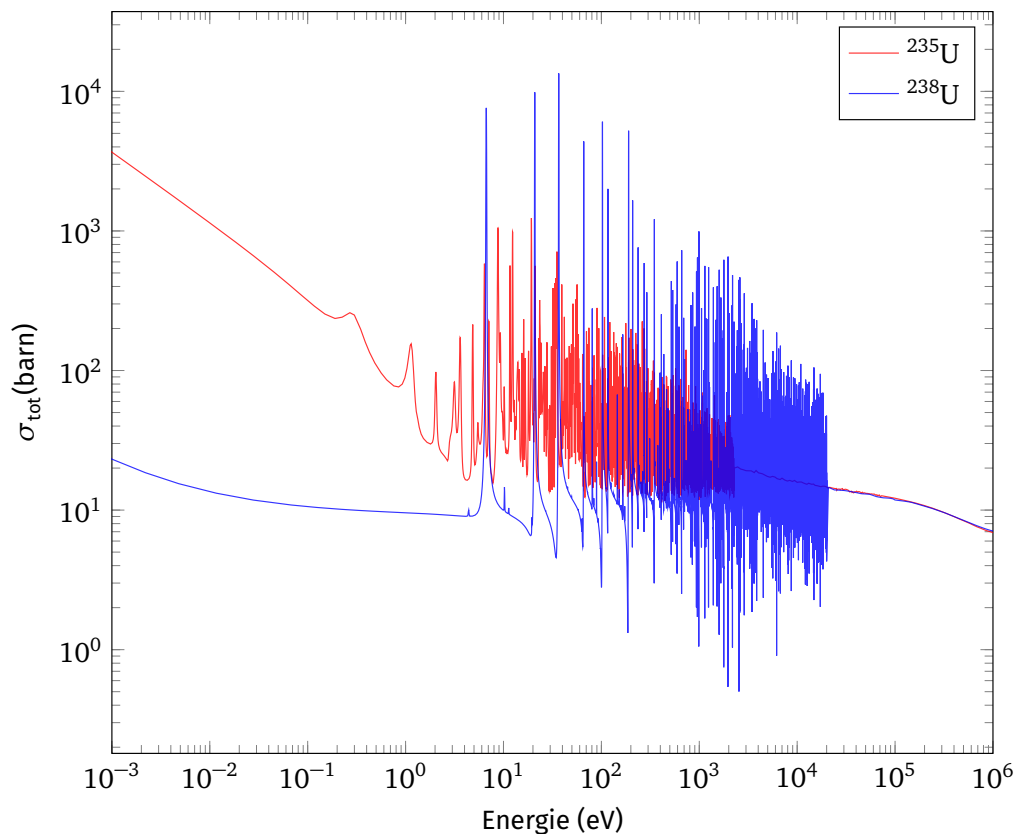


FIGURE 5.6 – Section efficace microscopique totale de  $^{235}\text{U}$  et  $^{238}\text{U}$  à 0K. (ENDF/B-VIII.0)

### 5.3.4 Effet Doppler

L'effet Doppler est provoqué par la variation de la température du combustible. La probabilité d'interaction d'un noyau avec un neutron est caractérisée par la section efficace qui dépend de la vitesse du neutron incident. En réalité, il faut considérer la vitesse relative de ce neutron par rapport au noyau. L'élévation de la température du combustible implique donc que l'énergie cinétique d'un noyau est plus élevée. En outre, la présence de pics de résonance dans la section efficace du combustible implique que de petites variations de la vitesse relative peut avoir un grand impact sur la probabilité d'absorption d'un neutron par un noyau d'uranium 238.

Le calcul de l'effet Doppler nécessite de prendre en compte l'agitation thermique du combustible et la variation de la vitesse d'oscillation du noyau autour de sa position moyenne, à la fois en direction et en norme [Reu12]. Le détail du calcul n'est pas essentiel ici, mais on notera que l'effet Doppler implique en pratique un élargissement des pics de résonance lors de l'élévation de la température du combustible, et son rétrécissement en cas de baisse de la température. L'élargissement des pics de résonance implique une capture neutronique plus importante par l'uranium 238 donc une baisse du flux neutronique. L'effet Doppler a donc un effet similaire à l'effet modérateur en ce qu'il tend à stabiliser la réaction en cas de perturbation de la température.

---

diminue également les chances d'absorption des neutrons par un noyau de bore.

### 5.3.5 Empoisonnement par le xénon

Le xénon 135 est un produit de fission de l'uranium 135. La chaîne de désintégration du xénon est représentée en figure 5.8. Il constitue le principal poison neutronique au sein du cœur du fait de sa très grande section efficace dans le domaine thermique, de l'ordre de  $10^6$  barns (figure 5.7). C'est un gaz qui se forme au sein des pastilles de combustible, il n'est donc pas possible de mesurer cette quantité.

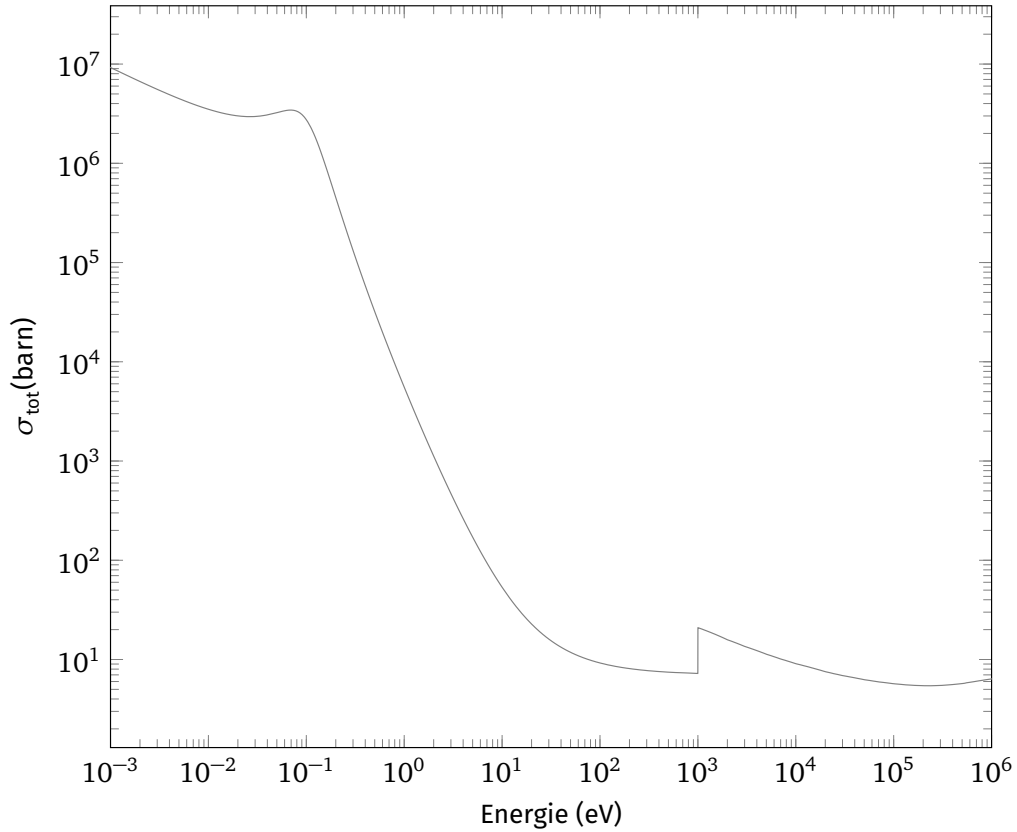


FIGURE 5.7 – Section efficace microscopique totale de  $^{135}\text{Xe}$  à 0K. (JEFF-3.3)

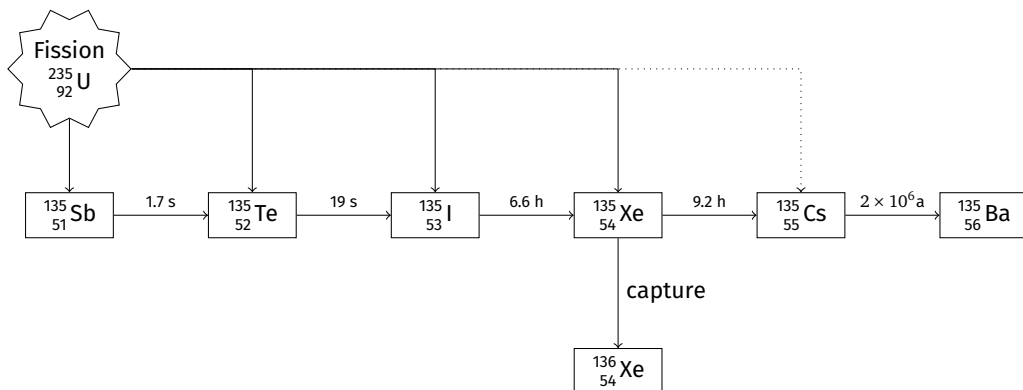


FIGURE 5.8 – Chaîne de désintégration du  $^{135}\text{Xe}$ .

Dans les éléments de la chaîne de désintégration du xénon issus de la fission du combustible, l'antimoine et le tellure ont une demi-vie si courte devant celle de l'iode qu'il est courant de considérer que l'iode est produit directement par la fission de l'uranium 235. On néglige également la très petite part d'uranium produisant directement du césium. La

dynamique de la concentration en iode comporte alors les termes de production par la fission du combustible et sa disparition par décroissance radioactive. Le xénon est produit en petite quantité avec un rendement  $\gamma_{Xe}$  et par décroissance radioactive de l'iode, et il est consommé par capture neutronique de rendement  $\sigma_{Xe}$  et décroissance radioactive. On peut alors écrire la dynamique xénon/iode décrivant la concentration en xénon et en iode en chaque point de l'espace (équation de Bateman) :

$$\dot{X}(t) = \lambda_I I(t) - \lambda_X X(t) + \gamma_X \Phi(t) - \sigma_X \Phi(t) X(t) \quad (5.3a)$$

$$\dot{I}(t) = \gamma_I \Phi(t) - \lambda_I I(t) \quad (5.3b)$$

En régime permanent, les termes de production et de disparition du xénon et de l'iode sont à l'équilibre. En cas de modification de la nappe de puissance (en réponse à une variation de charge ou des actions correctives de l'opérateur), le xénon et l'iode vont évoluer jusqu'à atteindre un nouvel équilibre. Le xénon se comporte comme un système à non minimum de phase. L'augmentation locale de la puissance aboutit à une concentration à l'équilibre de xénon plus élevée. Néanmoins pendant la phase transitoire de cette évolution on note que :

- l'iode est produit en plus grande quantité (en tant que produit de fission) pour atteindre une nouvelle valeur d'équilibre plus élevée.
- le xénon va d'abord être consommé par capture neutronique, en réponse à la hausse du flux neutronique. Puis dans un second temps (lorsque la puissance n'évolue plus), la quantité de xénon va croître à mesure que l'iode se « déverse » dans le xénon<sup>a</sup>. Puisque le xénon a une demi-vie plus rapide que l'iode, la quantité de xénon augmente davantage qu'il n'est désintégré.

Puisque le flux neutronique n'est pas le même en tout point de l'espace, la distribution de xénon n'est pas homogène dans les barres. Les variations locales de xénon ont pour effet d'impacter le flux neutronique local et peuvent induire un phénomène appelé « oscillations xénon », qui peut apparaître lorsque la distribution de puissance est perturbée. Dans ce cas, on observe que la variation de xénon dans la partie supérieure du cœur est en décalage de phase avec la variation de xénon dans la partie inférieure. Le phénomène des oscillations xénon est présenté de manière plus approfondie dans les travaux [Pon08] et [Pon+13], qui visent à estimer le xénon dans un modèle 1D en utilisant des méthodes d'assimilation de données.

### 5.3.6 Action de l'acide borique

Le bore<sup>b</sup> 10 est un poison neutronique, dont la section efficace dans le domaine thermique est de l'ordre de  $10^3$  barns. Il entre en jeu dans la composition des barres de contrôle et il est également introduit dans l'eau du circuit primaire sous forme d'acide borique. Lorsqu'un

a. Une analogie courante pour la dynamique xénon/iode est de considérer des « réservoirs » d'iode et de xénon qui fuient à la même vitesse qu'ils sont consommés et se remplissent à la même vitesse qu'ils sont produits.

b. Le bore naturel contient environ 80% de bore 11 et 20% de bore 10.

neutron interagit avec un neutron, deux réactions sont possibles <sup>a</sup> :



La concentration en bore au sein du circuit primaire peut-être contrôlée en injectant de l'eau pure pour le diluer ou en ajoutant de l'acide borique pour augmenter sa concentration. Le mélange est réalisé au sein du circuit de contrôle volumétrique et chimique, dont le contenu est injecté dans le circuit primaire. Comme la masse d'eau au sein du circuit est constante, l'injection d'eau borée ou pure nécessite d'extraire une partie de l'eau du primaire soit en l'injectant dans le RCV soit en l'extrayant du système, ce qui implique la production d'effluents qu'il est nécessaire de retraiter.

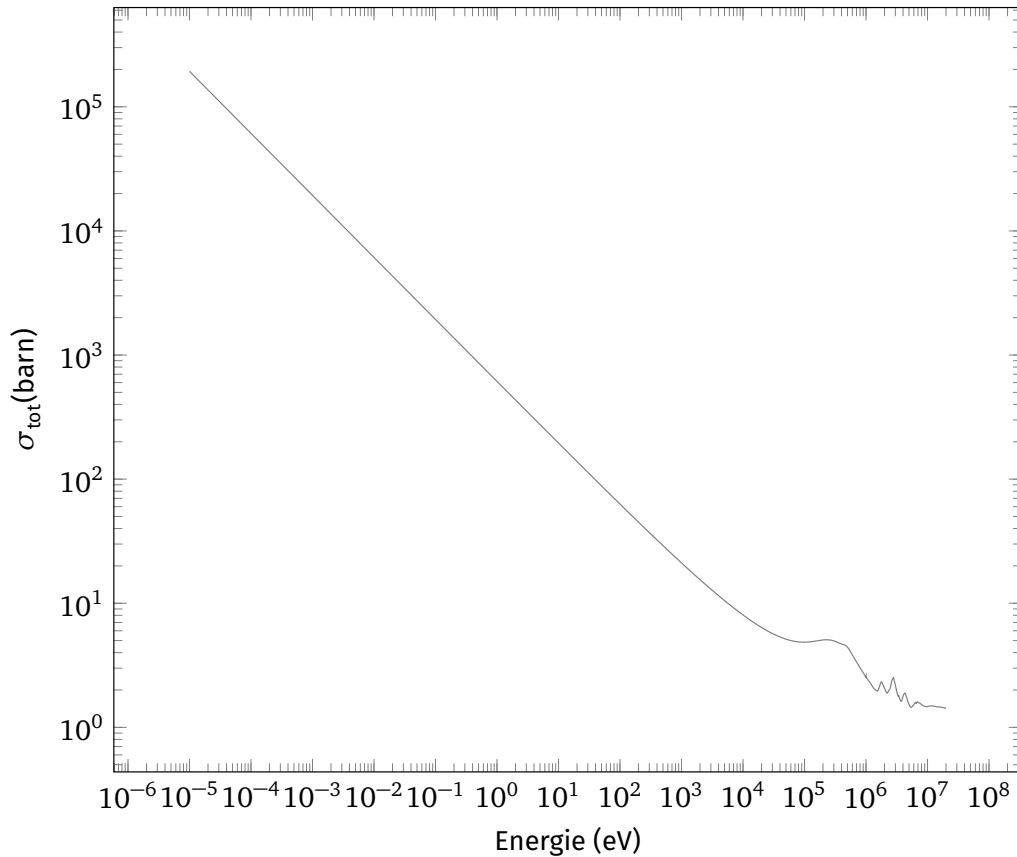


FIGURE 5.9 – Section efficace microscopique totale de <sup>10</sup>B à 0K. (JEFF-3.3)

On peut considérer que la concentration en bore dans le cœur est homogène, compte tenu du débit de l'eau parcourant le primaire. L'avantage du bore est que son action neutrophage est homogène dans le cœur, de sorte qu'il ne perturbe pas la distribution de puissance, à la différence des barres de contrôle.

a. Le lithium produit dans un état excité eq. 5.4b revient à son niveau fondamental en émettant un photon de 470keV. On peut mesurer le flux neutronique par la détection de ce photon.

### 5.3.7 Action des barres de contrôle

Les barres de contrôle sont constituées d'un matériau neutrophage (capturant les neutrons). Elles sont insérées par pas de quelques centimètres via la partie supérieure du réacteur (cf figure 5.2), en traversant le couvercle de la cuve et se glissent entre les crayons de combustible au sein des assemblages. Il est possible d'arrêter quasiment instantanément la réaction (arrêt d'urgence) en faisant tomber ces barres, ce qui explique leur insertion par le dessus de la cuve. Un mouvement des barres perturbe grandement la distribution de puissance (AO).

Il existe différentes stratégies d'insertion des grappes que l'on appelle modes de pilotage. De manière générale, les barres de contrôle sont regroupées en groupes, pouvant avoir des compositions et capacités neutrophage différentes. Les différents modes de pilotages couramment employés dans les REP sont les modes A, X, G et T et diffèrent de par le nombre et la composition des groupes de barres, leur recouvrement, ainsi que les lois de contrôle commande qui régissent leurs mouvements. La modélisation du cœur présentée au chapitre chapitre 6 est indépendante du mode de pilotage, le détail de leur fonctionnement n'est donc pas essentiel dans cette thèse, mais le lecteur pourra trouver davantage de détails dans les travaux [Gro07 ; Gas24 ; Dup23].

## CHAPITRE 6

# MODÉLISATION POUR LA COMMANDE ET L'ESTIMATION

### 6.1 Modélisation des cœurs de réacteurs nucléaires

Pour les études d'ingénierie de nombreux codes de calcul ont vu le jour pour modéliser la neutronique, la thermohydraulique ou la thermique du combustible. Ces codes sont couplés afin de modéliser le comportement à l'échelle du cœur entier et les interactions complexes entre les différents phénomènes physiques. Ces codes de calcul mettent en œuvre différentes approches pour simuler efficacement la physique du réacteur telles que les méthodes déterministes ou les méthodes Monte-Carlo [MU49]. De tels codes de calcul nécessitent des capacités de calcul conséquentes. Un usage en commande optimale en ligne, où il peut être nécessaire de réaliser un grand nombre de simulation avec des paramètres différents avec des contraintes sur les temps d'exécution, appelle à l'élaboration de modèles de conception qui sont plus simples que les modèles physiques fins précédemment décrits.

### 6.2 SOFIA

SOFIA (Simulateur d'Observation du Fonctionnement Incidentel et Accidentel) est un simulateur avancé disponible à Framatome. Il modélise la totalité des systèmes physiques de la centrale en particulier la totalité du circuit primaire et le circuit secondaire et tous les éléments qui les composent jusqu'aux vannes, pompes et capteurs individuels. Les systèmes de commande sont également modélisés ce qui permet de simuler des cas d'utilisation réels, raison pour laquelle il est peut-être employé comme outil d'entraînement à la conduite d'un réacteur et la simulation d'accident. Ce simulateur embarque différents codes de calcul pour la neutronique et la thermohydraulique.

En particulier pour la modélisation physique du cœur, les codes de calcul neutronique SMART et thermohydraulique CATHARE sont couplés. Ces modèles utilisent des résolutions spatiales différentes, illustrées par la figure 6.1. On dénombre ainsi 20 mailles pour le modèle de neutronique, les mailles 1 et 20 correspondant aux réflecteurs lourds en bas et en haut du réacteur respectivement, les mailles 2 à 19 correspondant effectivement au cœur. Le modèle thermohydraulique comprend quant à lui 36 mailles.

Des simulations de fonctionnement du cœur en condition de variation de charge ont été réalisées sur SOFIA et seront utilisées pour la validation de l'estimateur dans le chapitre 7. Il est possible de piloter le simulateur en temps réel comme en salle de commande, notamment en agissant sur l'automate de bore et la puissance de la turbine (qui sert de référence pour la puissance neutronique contrôlée). La vitesse maximale de simulation est d'environ deux fois le temps réel, ce qui limite la longueur et le nombre de transitoires réalisables sur le simulateur.



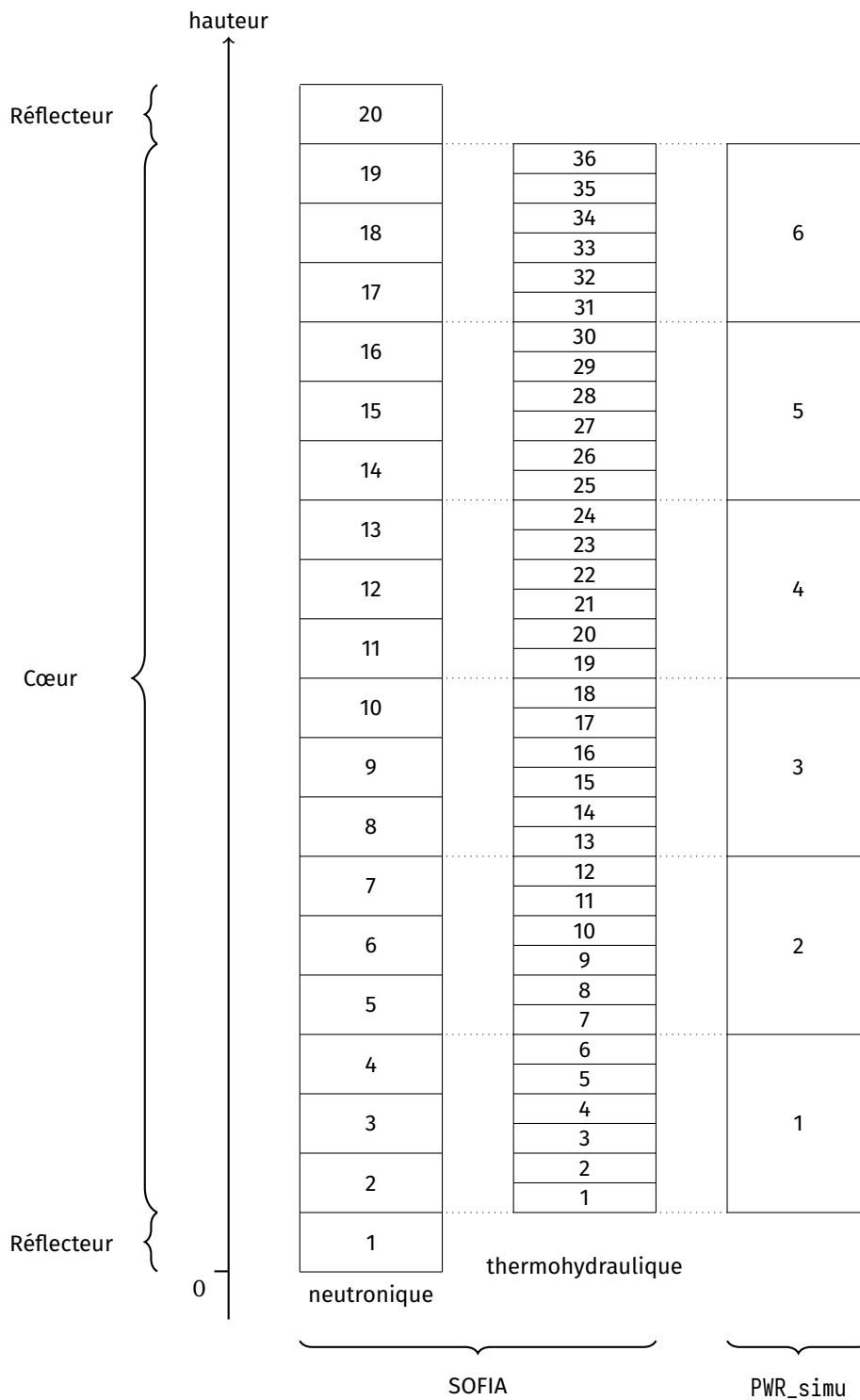


FIGURE 6.1 – Différences de maillage entre les modèles neutronique et thermique de SOFIA et le modèle PWR\_simu.

### 6.3 Modèle d'estimation et de commande

La commande prédictive a suscité un grand intérêt pour la conduite des réacteurs, en raison de ses nombreux avantages [Dup+21]. Les modèles issus de codes de calculs physiques, employés pour la conception des réacteurs et les études de sûreté, sont trop

complexes pour un usage en commande prédictive. Un modèle simplifié, permettant de prédire le comportement du système du point de vue des grandeurs importantes pour la conduite (en particulier l'AO) a été développé spécifiquement [Lem18 ; Dup23]. Nous détaillons dans ce paragraphe ce modèle que nous appelons PWR\_simu.

Dans le modèle de conception PWR\_simu, la seule dimension axiale est considérée et discrétisée en 6 mailles axiales. Pour la maille  $i \in \llbracket 1, 6 \rrbracket$ , on considère les quantités suivantes :

- concentration en xénon  $X_i$ ,
- concentration en iode  $I_i$ ,
- enthalpie massique de l'eau entrant dans la maille  $h_i$ ,
- enthalpie massique de l'eau sortant de la maille  $h_{i+1}$ ,
- population de neutrons retardés  $c_i$ ,
- puissance neutronique  $P_i$ .

Les entrées du modèle sont :

- la concentration en bore  $C_b$  (ppm) dans le cœur (supposée connue<sup>a</sup>),
- l'anti-réactivité  $\rho_{\text{rods},i}$  (ppm) introduite en chaque maille du fait de la position des barres de contrôle.
- la température  $T_{\text{in}}$  de l'eau entrant dans le cœur.

Le flux neutronique rend l'instrumentation à l'intérieur du cœur difficile (mais pas impossible). Nous supposons que nous n'avons pas d'instrumentation *in-core* disponible. Néanmoins, des chambres *ex-core* permettent de mesurer le flux neutronique s'échappant du cœur : elles sont placées autour du cœur et on dispose de 6 mesures axiales du flux neutronique. Cette résolution spatiale motive l'utilisation de 6 mailles axiales dans PWR\_simu. En outre, de nombreux capteurs de la température de l'eau en entrée et en sortie du réacteur sont disponibles. La position des grappes est connue. Avec les considérations que nous venons d'énoncer, les mesures dont on dispose sont les suivantes :

- flux neutronique en chaque maille.
- température de l'eau en sortie du réacteur  $T_{\text{out}}$  (différente de la température en sortie du cœur).
- température de l'eau entrant dans le réacteur (supposée identique en entrant dans le cœur).

Le modèle s'écrit ainsi sous la forme d'un modèle d'état :

$$\begin{aligned} \dot{x} &= f_{\text{PWR\_simu}}(x, u), & x &= [X_1, \dots, X_6, I_1, \dots, I_6, T_2, \dots, T_7, c_1, \dots, c_6, n_1, \dots, n_6]^T \\ y &= h(x), & u &= [C_b, \rho_{\text{rods},1}, \dots, \rho_{\text{rods},6}, T_1] \end{aligned}$$

Les équations du modèle, paramètres et propriétés sont détaillées dans la suite du chapitre.

---

a. Un bore-mètre est parfois présent sur site ; ses mesures prennent un certain temps, le signal de mesure est donc disponible avec un retard. Si un bore-mètre n'est pas présent, des prélèvements de l'eau du primaire sont réalisés quotidiennement pour déterminer la concentration en bore.

### 6.3.1 Équations de la cinétique et couplage spatial

On considère uniquement les neutrons thermiques et on se place dans l'hypothèse des neutrons à un groupe d'énergie. Les neutrons sont donc supposés ayant tous la même vitesse, si bien que le flux neutronique et la densité de neutrons en chaque maille sont les mêmes à un facteur multiplicatif près. On suppose que la puissance neutronique dégagée par les réactions de fission est directement proportionnelle au flux neutronique. On suppose un seul groupe de précurseurs (virtuel) qui rend compte de toutes les voies de production des neutrons retardés. Le modèle d'évolution des neutrons correspond à des équations de la cinétique (eq. 5.1) en chaque maille du modèle, couplées en s'inspirant de l'équation de la diffusion (eq. 5.2). Le Laplacien à une dimension correspond à :

$$\Delta = \frac{\partial^2}{\partial z^2}$$

Une discrétisation aux différences finies (eq. 2.2) de la dérivée première du flux par rapport à la variable d'espace donne :

$$\frac{\partial}{\partial z}\Phi(z_i) \approx \frac{\Phi(z_{i+1}) - \Phi(z_i)}{\delta z}$$

Une seconde application de la discrétisation donne :

$$\frac{\partial^2}{\partial z^2}\Phi(z_i) \approx \frac{\Phi(z_{i+1}) - 2\Phi(z_i) + \Phi(z_{i-1}))}{\delta z^2}$$

Le difficile bilan des captures neutroniques et des productions  $S - \Sigma_a \Phi$  (eq. 5.2), qui nécessiterait de suivre l'évolution de la population de nombreux noyaux (via les équations de Bateman) est substitué par l'équation de la cinétique (eq. 5.1). La dynamique du nombre de neutrons en chaque maille est alors donnée par :

$$\frac{dP_i}{dt} = \frac{1}{l}(\rho_i P_i - \beta P_i + D(P_{i+1} - 2P_i + P_{i-1})) + \lambda_c c_i \quad (6.1a)$$

$$\frac{dc_i}{dt} = \frac{\beta P_i}{l} - \lambda_c c_i \quad (6.1b)$$

On retrouve l'équation de la cinétique ponctuelle en sommant les  $\dot{P}_i$  et les  $\dot{c}_i$ , avec :

$$\rho_{\text{cœur}}(t) = \frac{\sum_i^6 \rho_i(t) P_i(t)}{\sum_i^6 P_i(t)}$$

$D$  est un paramètre d'échange de neutrons entre les mailles.  $\rho_i$  correspond à un bilan de réactivité dans la maille  $i$ . En toute rigueur, la réactivité est une donnée définie à l'échelle du cœur, mais le modèle s'apparente à la superposition de plusieurs modèles de réacteur ponctuel. Le bilan de réactivité inclut les contre-réactions neutroniques décrites à la section 5.3 de sorte que l'on écrit :

$$\rho_i = \rho_{\text{rods},i} + \rho_{X,i} + \rho_{\text{bore}} + \rho_{\text{mod},i} + \rho_{\text{dop},i}$$

- $\rho_{\text{rods},i}$  : insertion des barres dans la maille  $i$  (sous-section 5.3.7),
- $\rho_{X,i}$  : l'empoisonnement xénon dans la maille  $i$  (sous-section 5.3.5),

- $\rho_{\text{bore}}$  : bore dans l'eau du primaire, identique pour toutes les mailles (sous-section 5.3.6),
- $\rho_{\text{mod},i}$  : effet modérateur (sous-section 5.3.3),
- $\rho_{\text{dop},i}$  : effet Doppler (sous-section 5.3.4).

Les paramètres de chacun des termes du bilan de réactivité ont été obtenus en utilisant le code de neutronique SMART en considérant un état réacteur nominal critique et en faisant varier les quantités relatives aux contre-réactions. Le bilan de réactivité prend donc la forme suivante :

$$\rho_i = \rho_{\text{nom},i} + \Delta\rho_{\text{rods},i} + \Delta\rho_{X,i} + \Delta\rho_{\text{bore}} + \Delta\rho_{\text{mod},i} + \Delta\rho_{\text{dop},i} \quad (6.2)$$

où  $\rho_{\text{nom},i}$  est la réactivité nominale dans la maille  $i$  pour le cœur critique, qui s'écrit en cherchant la solution à l'équation de la dynamique de la puissance neutronique (eq. 6.1) à l'équilibre :

$$\rho_{\text{nom},1} = D \frac{P_{\text{nom},2} - P_{\text{nom},1}}{P_{\text{nom},1}} \quad (6.3a)$$

$$\rho_{\text{nom},i} = D \frac{P_{\text{nom},i+1} - 2P_{\text{nom},i} + P_{\text{nom},i-1}}{P_{\text{nom},i}}, \quad \forall i \in \llbracket 2, 5 \rrbracket \quad (6.3b)$$

$$\rho_{\text{nom},1} = D \frac{P_{\text{nom},5} - P_{\text{nom},6}}{P_{\text{nom},6}} \quad (6.3c)$$

### Insertion des barres

L'insertion des barres agit localement mais impacte le réacteur tout entier. En fonction du mode de pilotage du réacteur,  $n_{\text{group}}$  groupes de barres sont considérés. Les groupes peuvent contenir des barres de capacité neutrophage différent. Dans chaque maille du modèle, on considère l'insertion du groupe de barres  $G_i$  en pourcentage. Lorsque l'extrémité des barres du groupe descend sous l'interface de la maille  $i$  et de la maille inférieure  $i - 1$ , l'insertion de ce groupe dans  $i$  est de 100% et il commence à s'insérer dans  $i - 1$ .

À chaque groupe de barres de contrôle, on associe un coefficient d'efficacité  $k_{\text{rods},j}$ . En notant  $I = (I_{ij})$  l'insertion de la barre  $j$  dans la maille  $i$ , le coefficient d'anti-réactivité du cœur associée aux barres de contrôle s'écrit :

$$\begin{bmatrix} \rho_{\text{rods},1} \\ \vdots \\ \rho_{\text{rods},6} \end{bmatrix} = I k_{\text{rods}}$$

Le calcul de réactivité est déterministe mais non linéaire, du fait de la saturation de l'insertion des groupes de barres à 100% dans une maille. Au lieu de considérer directement la position des grappes dans les entrées du modèle, qui nécessite d'effectuer le calcul de l'insertion dans les mailles, on prend directement en entrée l'écart de réactivité des barres, calculé comme :

$$\Delta\rho_{\text{rods}} = \rho_{\text{rods},\text{nom}} - \rho_{\text{rods}}(t)$$

où  $\rho_{\text{rods}}(t)$  est calculé à partir de la position à l'instant  $t$  des barres de contrôle.

### Empoisonnement par le xénon

L'empoisonnement par le xénon est modélisé par la relation :

$$\Delta\rho_{\text{xénon},i} = K_{\text{Xe}}(X_i - X_{\text{nom},i}) \quad (6.4)$$

où  $K_{\text{Xe}}$  est identifié sur un code de neutronique et  $X_{\text{nom},i}$  correspond à la concentration de xénon dans la maille  $i$  de l'état nominal du cœur (cf eq. 6.6).

Il est courant dans les codes de calcul neutronique de ne pas considérer la distribution de xénon dans les assemblages mais une quantité en xénon totale.  $K_{\text{Xe}}$  est identifié sur la variation de réactivité du cœur en réponse à l'évolution de la quantité de xénon totale.

### Action du bore

L'absorption des neutrons par le bore est uniforme en toute maille et on calcule

$$\Delta\rho_{\text{bore}} = K_{\text{bore}}(Cb - Cb_{\text{nom}})$$

où  $K_{\text{bore}}$  est identifié sur un code de neutronique et  $Cb_{\text{nom}}$  correspond à la concentration en bore qui assure la criticité pour l'état nominal du cœur.

### Effet modérateur

L'effet modérateur en chaque maille est calculé par une table d'interpolation tabulée en température du caloporteur. Il faut noter que PWR\_simu modélise les températures à l'interface des mailles, tandis que l'effet modérateur prend en compte la température du caloporteur au sein de la maille, prise comme la moyenne de la température en entrée et en sortie de maille.

$$\rho_{\text{mod},i} = T_{T \rightarrow \rho} \left( \frac{1}{2}(T_i + T_{i+1}) \right)$$

Le calcul de l'écart de réactivité associé à l'effet modérateur, par rapport à l'état nominal est ainsi calculé :

$$\Delta\rho_{\text{mod},i} = K_{\text{mod}}(\rho_{\text{mod},i} - \rho_{\text{mod,nom},i})$$

où  $\rho_{\text{mod,nom},i} = T_{T \rightarrow \rho} \left( \frac{1}{2}(T_{\text{nom},i} + T_{\text{nom},i+1}) \right)$  Les valeurs de la table d'interpolation sont identifiées sur un code de calcul neutronique. En figure 6.2, on remarque la table d'interpolation suit toujours un profil similaire (pour le début, milieu et fin de vie). Le facteur  $k_{\text{mod}}$  est un facteur multiplicatif introduit afin d'ajuster la contribution de l'effet modérateur à la réactivité, en fonction de l'avancement dans le cycle du combustible.

### Effet Doppler

L'effet Doppler en chaque maille est calculé par une table d'interpolation tabulée en puissance. En toute rigueur, l'effet Doppler dépend de la température du combustible. Cette dernière n'est pas modélisée par PWR\_simu, mais on suppose que la température du combustible dans la maille  $i$  dépend directement de la puissance neutronique  $P_i$ , ce qui

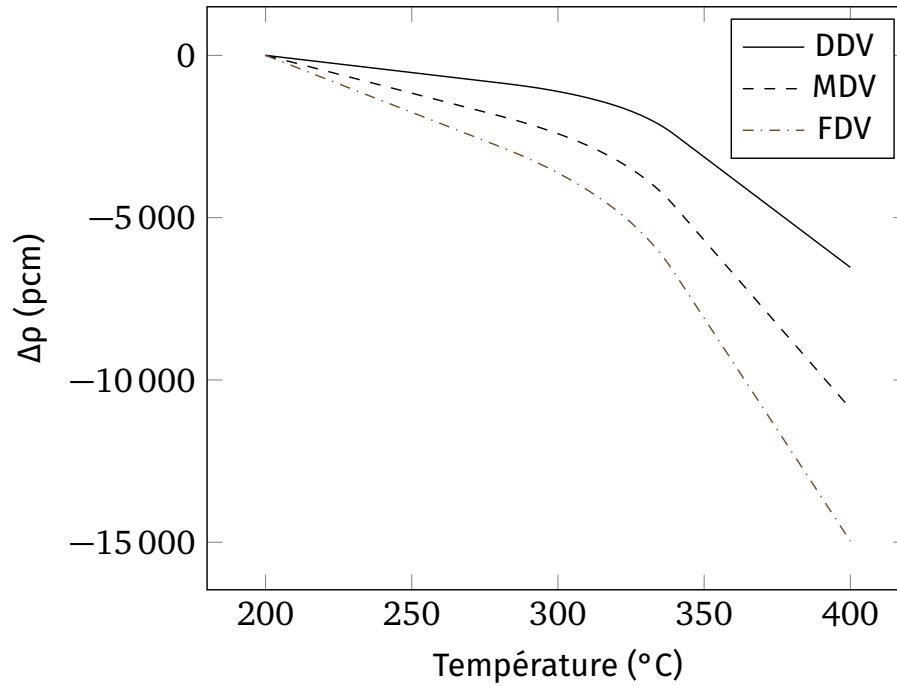


FIGURE 6.2 – Table d'interpolation pour l'effet modérateur en début de vie (DDV), milieu de vie (MDV) et fin de vie (FDV).

permet de tabuler en puissance.

$$\rho_{\text{dop},i} = T_{p \rightarrow \rho}(P_i)$$

Le calcul de l'écart de réactivité associé à l'effet Doppler par rapport à l'état nominal est ainsi calculé :

$$\Delta\rho_{\text{dop},i} = \rho_{\text{dop},i} - \rho_{\text{dop,nom},i}$$

où  $\rho_{\text{mod,nom},i} = T_{T \rightarrow \rho}(P_{\text{nom},i})$  Les valeurs de la table d'interpolation sont identifiées sur un code de calcul neutronique.

### 6.3.2 Dynamique xénon/iode

Le xénon et l'iode se forment au sein des pastilles de combustible où ils sont confinés, il n'y a donc pas de diffusion spatiale. En appliquant les équations de Bateman à la maille  $i$ , on a alors :

$$\dot{X}_i(t) = \lambda_I I_i(t) - \lambda_X X_i(t) + \Gamma_X P_i(t) - \Sigma_X P_i(t) X_i(t) \quad (6.5a)$$

$$\dot{I}_i(t) = \Gamma_I P_i(t) - \lambda_I I_i(t) \quad (6.5b)$$

En notant  $P_{\text{nom},i}$  la puissance neutronique nominale de la maille  $i$ , les concentrations nominales en xénon  $X_{\text{nom},i}$  et iode  $I_{\text{nom},i}$  s'obtiennent en résolvant (eq. 6.5) à l'équilibre ( $\dot{X}_i(t), \dot{I}_i(t) = 0$ ) et s'écrivent :

$$I_{\text{nom},i} = \frac{\Gamma_I}{\lambda_I} P_{\text{nom},i} \quad \text{et} \quad X_{\text{nom},i} = \frac{\lambda_I I_{\text{nom},i} + \Gamma_X P_{\text{nom},i}}{\lambda_X - \Sigma_X P_{\text{nom},i}} \quad (6.6)$$

### 6.3.3 Modèle de température

On suppose que la pression  $p$  de l'eau est constante dans l'ensemble du cœur (hypothèse isobare). La température de l'eau entrant dans une maille est égale à la température de l'eau sortant de la maille précédente. L'enthalpie massique de l'eau notée  $h = u + pv$  (où  $u$  est l'énergie interne massique et  $v$  le volume massique) caractérise l'énergie dans un système ouvert. On considère le volume de fluide contenu dans la maille  $i$ . Le premier principe de la thermodynamique pour ce système ouvert s'écrit :

$$\dot{q} + w = \Delta h + \Delta e_{\text{méca}} \quad (6.7)$$

- $\dot{q}$  est la puissance massique fournie au système sous forme de chaleur.
- $w$  est la puissance massique fournie au système sous forme de travail.
- $\Delta h$  est la variation d'enthalpie massique du système.
- $\Delta e_{\text{méca}}$  est la variation d'énergie mécanique (énergie potentielle de pesanteur et énergie cinétique).

Le travail reçu par le système est nul dans le circuit primaire (la turbine se trouvant dans le circuit secondaire) et la variation d'énergie mécanique du système est négligeable devant la variation de l'enthalpie. Le bilan (eq. 6.7) se réécrit :

$$\dot{q} = \Delta h \quad (6.8)$$

On suppose que la totalité de la puissance neutronique est transmise à l'eau sous-forme de chaleur. On obtient alors à l'équilibre :

$$h_{i+1} - h_i = K_{\text{ent}} P_i \quad (6.9)$$

$K_{\text{ent}}$  permet de convertir de la puissance neutronique (exprimée en %PN) à une enthalpie massique. Sa valeur est identifiée à partir de données de SOFIA.

Ce transfert n'est pas instantané : la gaine des crayons de combustible est d'abord échauffée puis transmet sa chaleur à l'eau. Afin d'éviter de modéliser la thermique des assemblages de combustible, on suppose que le transfert thermique possède un temps caractéristique  $\tau_{\text{maille}}$  et suit une dynamique du premier ordre :

$$\frac{dh_{i+1}}{dt} = \frac{1}{\tau_{\text{maille}}} (h_i - h_{i+1} + K_{\text{ent}} P_i) \quad (6.10)$$

À l'équilibre  $\frac{dh_{i+1}}{dt} = 0$  ce qui correspond bien à (eq. 6.9). D'autre part, la chaleur massique reçue par le système se relie aux températures de l'eau entrant et sortant de la maille selon la relation :

$$q = \int_{T_i}^{T_{i+1}} c_{\text{eau}}(T) dT \quad (6.11)$$

où  $c_{\text{eau}}(T)$  est la capacité thermique de l'eau ( $\text{J kg}^{-1}\text{K}^{-1}$ ) dépendant de la température.

Une approximation courante consiste à considérer la capacité thermique de l'eau constante, or ce n'est pas le cas aux températures de fonctionnement des REP. En effet,  $c_{\text{eau}}$  décroît avec la température. Les conditions de pressions et de température étant bien connues dans

le cœur, une table d'interpolation est construite<sup>a</sup> pour passer de l'enthalpie massique à la température (essentielle pour le calcul de l'effet modérateur). Ainsi la température en entrée de la maille  $i$  se détermine avec :

$$T_i = T_{\text{ent} \rightarrow \text{temp}}(h_i)$$

Et on note  $T_{\text{ent} \rightarrow \text{temp}}^{-1} = T_{\text{temp} \rightarrow \text{ent}}$ .

### Théorème 6.1

En considérant le modèle de thermique comme régi par l'enthalpie  $h_1 = T_{\text{temp} \rightarrow \text{ent}}(T_1)$  et les puissances  $P_i$ , il s'écrit comme un modèle linéaire, avec :

$$A = \begin{bmatrix} -1/\tau_{\text{maille}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/\tau_{\text{maille}} & -1/\tau_{\text{maille}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/\tau_{\text{maille}} & -1/\tau_{\text{maille}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1/\tau_{\text{maille}} & -1/\tau_{\text{maille}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/\tau_{\text{maille}} & -1/\tau_{\text{maille}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/\tau_{\text{maille}} & -1/\tau_{\text{maille}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/\tau_{\text{maille}} & -1/\tau_{\text{maille}} \end{bmatrix},$$

$$B = \begin{bmatrix} 1/\tau_{\text{maille}} & K_{\text{ent}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & K_{\text{ent}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & K_{\text{ent}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & K_{\text{ent}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & K_{\text{ent}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & K_{\text{ent}} \end{bmatrix}, \quad C = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1],$$

$$x = \begin{bmatrix} h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \end{bmatrix} \quad \text{et} \quad u = \begin{bmatrix} h_1 \\ P_2 \\ P_3 \\ P_4 \\ P_5 \\ P_6 \\ P_7 \end{bmatrix}$$

Les variables d'état associées à l'enthalpie massique sont observables.

**Preuve :** Immédiat par théorème 1.4. ■

a. L'outil TABLEDO interne à Framatome a été utilisé.



### 6.3.4 Synthèse des équations du modèle

En rassemblant les équations que nous avons présentées précédemment, le modèle complet de réacteur s'écrit comme un modèle d'état non linéaire à 30 états, 8 entrées et 7 sorties :

$$\begin{aligned}
 \dot{X}_i(t) &= \lambda_I I_i(t) - \lambda_X X_i(t) + \Gamma_X P_i(t) - \Sigma_X P_i(t) X_i(t) \\
 \dot{I}_i(t) &= \Gamma_I P_i(t) - \lambda_I I_i(t) \\
 \frac{dh_{i+1}}{dt} &= \frac{1}{\tau_{\text{maille}}} (h_i - h_{i+1} + K_{\text{ent}} P_i) \quad \forall i \in \llbracket 2, 7 \rrbracket \\
 \frac{dP_1}{dt} &= \frac{1}{l} (\rho_1 P_1 - \beta P_1 + D(P_2 - P_1)) + \lambda_c c_1 \\
 \frac{dP_i}{dt} &= \frac{1}{l} (\rho_i P_i - \beta P_i + D(P_{i+1} - 2P_i + P_{i-1})) + \lambda_c c_i \quad \forall i \in \llbracket 2, 5 \rrbracket \\
 \frac{dP_6}{dt} &= \frac{1}{l} (\rho_6 P_6 - \beta P_6 + D(P_5 - P_6)) + \lambda_c c_6 \\
 \frac{dc_i}{dt} &= \frac{\beta P_i}{l} - \lambda_c c_i \\
 h_1 &= T_{\text{temp} \rightarrow \text{ent}}(T_1) \\
 \rho_i &= \rho_{\text{nom},i} + \Delta \rho_{\text{rods},i} + \Delta \rho_{X,i} + \Delta \rho_{\text{bore}} + \Delta \rho_{\text{mod},i} + \Delta \rho_{\text{dop},i}
 \end{aligned} \tag{6.12}$$

où :

- $\lambda_I$  est la période de décroissance radioactive de l'iode 135,
- $\lambda_X$  est la période de décroissance radioactive du xénon 135,
- $\Gamma_X$  est le rendement de production du xénon 135 par voie directe de l'uranium 235,
- $\Sigma_X$  est la section efficace de capture du xénon 135,
- $\Gamma_I$  est le rendement de production de l'iode 135 par voie directe de l'uranium 235,
- $K_{\text{ent}}$  est le coefficient d'échauffement du caloporteur,
- $l$  est la durée de vie moyenne d'un neutron dans le cœur,
- $\beta$  est la proportion des neutrons retardés,
- $\lambda_c$  est la période de désintégration des précurseurs.

Les paramètres suivants dépendent de l'épuisement du combustible :

- $D$  : coefficient d'échange de neutrons entre les mailles,
- $T_{T \rightarrow \rho}$  : table d'interpolation calculant l'effet modérateur,
- $T_{P \rightarrow \rho}$  : table d'interpolation calculant l'effet Doppler,
- $K_{Xe}$  : efficacité différentielle du xénon,
- $K_{\text{bore}}$  : efficacité différentielle du bore,
- $K_{\text{rods}}$  : efficacité différentielle des barres de contrôle,
- $P_{\text{nom}}$  : distribution nominale de puissance, dont sont déduits les paramètres  $X_{\text{nom}}$  (distribution nominale de xénon) et  $T_{\text{nom}}$  (profil nominal de température),
- $C b_{\text{nom}}$  : concentration en bore assurant la criticité du bore,
- $p_{\text{rods, nom}}$  : insertion nominale des barres de contrôle.

### 6.3.5 Propriétés

Le modèle que nous venons de présenter est un modèle non-linéaire du fait de l'utilisation de tables d'interpolation ainsi que par le terme de consommation du xénon par capture neutronique (eq. 6.5a). Il est également raide du fait des différences de temps caractéristique dans ses dynamiques. D'une part, l'équation de la cinétique a un temps caractéristique correspondant au temps de parcours moyen des neutrons prompts avant d'interagir avec un noyau, qui est de l'ordre de  $10^{-5}$ s. D'autre part, les dynamiques du xénon et de l'iode sont beaucoup plus longues du fait de la demi-vie de l'iode et du xénon, qui sont respectivement de l'ordre de 5 et 9 heures. La simulation numérique du modèle d'estimation doit utiliser une méthode d'intégration numérique adaptée (L-stable cf théorème 2.7).

Le modèle d'estimation se distingue du modèle de commande (présenté dans [Dup23]) de plusieurs manières :

- le modèle thermique pour l'estimation passe par les enthalpies massiques de l'eau pour calculer la température du caloporteur. Ce modèle donne des résultats plus fidèles des températures internes au cœur et a depuis été intégré au modèle de commande.
- le modèle de commande prend en entrée non pas la concentration en bore, mais les débits d'eau diluée ou d'eau borée introduite dans l'eau du circuit primaire. On pourrait envisager d'estimer la concentration en bore mais un problème de détectabilité se pose alors si l'on essaye de déterminer à la fois la concentration en bore et le xénon.

Pour détailler le deuxième point, penchons nous à nouveau sur le bilan de réactivité (eq. 6.2) qui régit en majorité le modèle d'évolution de la puissance. On peut écrire que  $\rho_i = f(P_i, Cb, T_i, T_{i+1}, X_i, h_{rods})$  où  $h_{rods}$  décrit la position des barres de contrôle. La puissance est mesurée et est directement influencée par la réactivité dans la maille. La seule autre mesure dont nous supposons disposer est la température du caloporteur en sortie de cœur.

#### Proposition 6.2

*Pour  $i \in \llbracket 1, 6 \rrbracket$ , Il existe une infinité tuples de valeurs  $(P_i, Cb, T_i, T_{i+1}, X_i, h_{rods})$  menant à une même valeur de  $\rho_i$  et à fortiori de puissance mesurée  $P_i$ .*

Les températures  $T_{2,\dots,6}$  sont observables (théorème 6.1), on peut alors écarter les  $P_i$  et  $T_i$  des quantités « problématiques » pour la détectabilité du système, ainsi que la position des barres de contrôle supposée connue. En raisonnant sur le bore et le xénon uniquement, on est amené à affiner la théorème 6.2.

#### Proposition 6.3

*Pour  $i \in \llbracket 1, 6 \rrbracket$ , il existe une infinité de couples de valeurs  $(Cb, X_i)$  menant à une même valeur de  $\rho_i$  et à fortiori de puissance mesurée  $P_i$ .*

Si on suppose la concentration en bore connue et puisqu'elle est uniforme dans toutes les mailles, il est alors possible de déterminer le xénon en chaque maille. Cette « analyse » intuitive de la détectabilité du modèle nous mène à supposer la concentration en bore disponible. Cette hypothèse est réaliste : la concentration en bore dans le primaire est régulièrement mesurée. Son estimation par la simulation en partant d'une valeur connue de bore et en tenant compte des commandes de dilution/borication est possible.



## CHAPITRE 7

# APPLICATION AUX RÉACTEURS NUCLÉAIRES À EAU PRESSURISÉE

Dans ce chapitre, on détaille plus précisément les détails d'implémentation de l'estimateur ainsi que l'influence des paramètres de réglage de l'algorithme sur ses performances. Le problème d'estimation conjoint de paramètres et de l'état est posé et résolu pour des paramètres du modèle dans le cas où le modèle et le système sont identiques. Enfin, nous présentons des résultats d'estimation obtenus sur des données de simulation fine issues de SOFIA. Les choix sont motivés par les éléments présentés en partie I et certains sont critiques pour le bon fonctionnement de l'estimateur. L'estimateur est validé sur des données issues de simulation de SOFIA : le transitoire de validation est détaillé en annexe C. On distingue plusieurs scénarios de validation :

### Variation de l'instrumentation

**Capteurs** : les mesures disponibles sont celles détaillées en section 6.3 et correspondent aux différents capteurs et appareils de mesures installés sur site et qui fournissent un flux de données accessible à tout instant. Ces mesures correspondent à la température du caloporteur en sortie de cuve ainsi que la puissance neutronique en chacune des 6 mailles du modèle d'estimation.

**Système expert** : en plus des mesures issues de capteurs sur site, un Core Monitoring System (CMS) est un système se comportant comme un jumeau numérique du système. Embarquant des codes de calcul de la physique du réacteur, il possède une résolution numérique plus fine que le modèle de conception et modélise davantage de phénomènes physiques. Ce type de système, au regard de la complexité et du temps de calcul, n'est pas exploitable en MHE ou MPC mais les résultats de simulation peuvent être utilisés comme des mesures virtuelles. **ARGOS** est un CMS développé par Framatome avec lequel un couplage est possible. L'intérêt d'un tel couplage est étudié en ajoutant dans les mesures certains signaux simulés, en pratique inaccessibles par la mesure, mais supposés déterminés par le système expert.

### Jeux de données de validation

**PWR\_simu** : on considère un transitoire simulé avec le modèle PWR\_simu et décrit en annexe B. Il correspond à un cœur piloté en mode A, en début de cycle combustible. Ce transitoire correspond à la situation « idéale » où modèle et process sont identiques. Il permet de valider l'estimateur et d'identifier les conditions d'estimation des paramètres du modèle.

**SOFIA** : Un transitoire simulé avec SOFIA et décrit en permet de valider l'estimateur dans le contexte plus réaliste où le modèle d'estimation n'est pas identique au process. Les données issues de simulations fines, cette validation permet d'apprécier les performances plausibles de l'estimateur lors du passage au « monde réel ».

## 7.1 Conception d'un estimateur pour l'aide au pilotage

### 7.1.1 Operator Aid Predictive System (OAPS)

OAPS est un système d'aide au pilotage de l'offre commerciale de Framatome. Les travaux [Lem18 ; Dup23] ont permis d'asseoir la commande prédictive comme une méthode avantageuse pour la conduite des réacteurs nucléaires. Le cahier des charges du contrôleur est exprimé directement au sein d'un problème d'optimisation sous contraintes, ce qui facilite grandement la conception d'une loi de commande remplissant un ensemble de spécifications. L'offre a vocation à se décliner pour les différents modes de pilotage (cf sous-section 5.3.7) des centrales nucléaires et sous différents modes de fonctionnement, allant de la proposition de recommandations de commandes pour les opérateurs (borication/dilution de l'eau du primaire et/ou insertion des barres de contrôle) au pilotage entièrement automatisé du réacteur (à l'instar du mode T [Gro07]).

Les performances de la commande prédictive reposent sur la représentativité du modèle du process employé pour prédire le comportement du système en réponse à un signal de commande. Puisque l'ensemble de l'état du modèle doit être simulé, il est nécessaire de disposer de ce dernier (ou à minima d'une estimation de l'état) pour initialiser le problème d'optimisation. L'évolution du comportement du cœur, en raison de l'épuisement du combustible, induit une dérive des prédictions du modèle si ses paramètres ne sont pas convenablement ajustés pour refléter le comportement courant du système.

Les critères de pilotage du cœur portent sur un contrôle de l'*axial offset* et de la température moyenne du caloporteur lors du suivi d'un profil de charge de référence. Il est donc important que la modélisation de ces quantités soit fidèle au comportement du cœur. Le modèle de conception inclut (eq. 6.12) et modélise également le générateur de vapeur, la loi de commande d'insertion des grappes et l'automate de bore (évolution de la concentration en bore en réponse à des consignes de dilution/borication).

On considère le fonctionnement de l'outil d'aide au pilotage dans un contexte de recommandation pour un opérateur (figure 7.1). Dans ce mode de fonctionnement, l'outil produit une suite de commandes  $\tilde{u}_{[t,t+H_c]}$  et la prédiction de la trajectoire associée  $\tilde{x}_{[t,t+H_c]}$  sur l'horizon de commande  $H_c$ . Les recommandations sont produites plusieurs minutes à l'avance par rapport à l'instant où elles s'appliquent afin de laisser le temps à l'opérateur d'apprécier la pertinence de la recommandation, puis de préparer l'application de cette dernière s'il choisit de la suivre. L'opérateur conserve la main sur les commandes appliquées au process et applique le signal de commande  $\bar{u}$ . On distingue alors plusieurs cas :

- l'opérateur applique à la lettre les recommandations de commande. Cette hypothèse n'est pas réaliste puisque d'une part, elle implique que l'opérateur applique systématiquement la commande à l'instant précis où elle commence (sans en rater<sup>a</sup>), mais également qu'il ne fait aucune erreur sur la valeur à utiliser.
- l'opérateur ignore la recommandation de l'outil, en jugeant une autre action plus judicieuse.

---

a. Le système d'aide au pilotage est couramment comparé à un système de navigation tel qu'un GPS. Même en suivant rigoureusement l'itinéraire proposé, il arrive parfois que l'on rate une sortie d'autoroute ou qu'elle ne soit pas empruntable.

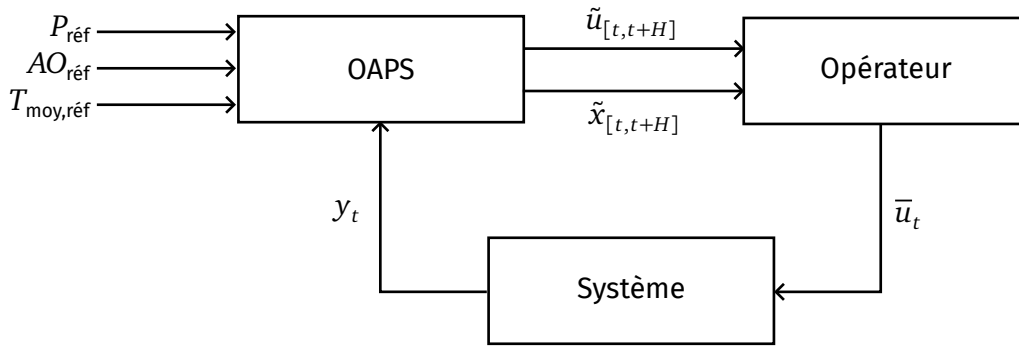


FIGURE 7.1 – Fonctionnement de l'OAPS en mode recommandations.

### 7.1.2 Estimation ad-hoc

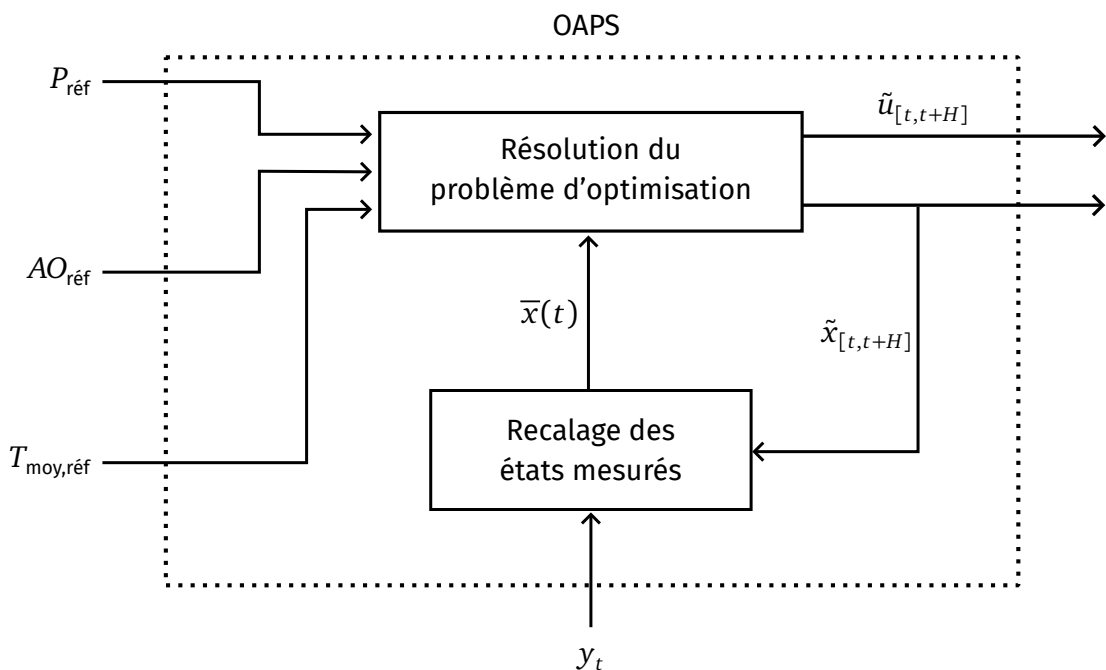


FIGURE 7.2 – Recalage de la prédiction de l'OAPS sur les mesures.

La solution du problème d'optimisation MPC est une suite de commandes  $\tilde{u}_{[t,t+H_c]}$  sur l'horizon de commande  $H_c$  ainsi qu'une trajectoire prédite  $\tilde{x}_{[t,t+H_c]}^{OAPS}$ , résultant de l'application de  $\tilde{u}_{[t,t+H_c]}$  à partir de  $\bar{x}(t)$ . La fréquence de fonctionnement du contrôleur est notée  $dt_{OAPS}$ . La solution précédente au problème d'optimisation est notée  $\tilde{x}(t)_{[t-dt_{OAPS},t-dt_{OAPS}+H_c]}$ .

Du fait de l'écart entre le modèle de prédiction de la MPC et le process, l'état prédit par le contrôleur dans le futur est erroné. On dispose en particulier des mesures de puissance neutronique en chaque maille. L'estimation de l'état  $\bar{x}(t)$  est obtenue en perturbant les puissances neutroniques supérieure et inférieure de  $\tilde{x}(t)_{[t-dt_{OAPS},t-dt_{OAPS}+H_c]}$  afin que le déséquilibre axial de puissance de  $\bar{x}(t)$  corresponde à l'AO mesuré sur le système.

Notons que la mesure est partiellement utilisée pour recalibrer l'estimation. En particulier, si la nappe de puissance prédite est déformée de sorte que l'*axial offset* prédit corresponde à l'*axial offset* mesuré, les autres variables d'état évoluent en « boucle ouverte » (ils sont simulés). Le principal problème de cette approche réside dans l'utilisation des signaux de

commande recommandé par le système, sans garantie qu'ils soient effectivement appliqués. Il y a alors un risque que les variables d'état prédites par OAPS s'écartent significativement de l'état réel du cœur et remettent en cause la pertinence des recommandations de l'outil.

Ce schéma d'estimation a été validé en même temps qu'OAPS sur des simulations fines, mais la lenteur de la dynamique xénon/iode rend difficile l'analyse de la dérive de l'état interne du contrôleur. En effet, il faudrait mener des validations sur des durées de simulation qui correspondraient à minima à plusieurs jours de fonctionnement du réacteur, ce qui est difficilement envisageable en pratique (pour des raisons économiques et pratiques de disponibilité du simulateur fin).

Pour se prémunir des limites du schéma d'estimation *ad-hoc* actuel, on propose la conception d'un estimateur de l'état du cœur exploitant les signaux réels du système pour corriger l'ensemble de l'état.

### 7.1.3 Cahier des charges de la solution d'estimation

L'estimateur MHE doit remplir plusieurs objectifs au regard des besoins de l'outil d'aide au pilotage. Dans un premier temps, l'estimateur doit remplacer le schéma d'estimation *ad hoc* actuel pour palier à ses limites et répondre aux spécifications suivantes :

- assurer la correction de l'ensemble de l'état, en particulier pour les états évoluant en boucle ouverte et dont l'estimation pourrait diverger,
- rejeter les bruits affectant les mesures site,
- dans la version « recommandations » de l'outil d'aide au pilotage, exploiter les signaux de commande effectivement appliqués au système pour réagir à l'écart entre celles-ci et les recommandations.

Pour les besoins spécifiques de la loi de commande de l'outil d'aide au pilotage, on formule les besoins suivants :

- garantir une estimation précise (et physiquement cohérentes) des quantités intervenant dans le critère de performance du contrôleur prédictif, c'est-à-dire la température moyenne du caloporteur et l'axial offset,
- ajuster des paramètres d'intérêt du modèle pour garantir sa représentativité vis-à-vis du comportement courant du cœur tout au long du cycle combustible,
- fournir aux opérateurs des informations pertinentes pour la conduite, en particulier vis-à-vis de la concentration en iode et en xénon,
- une résolution rapide du problème d'optimisation pour limiter la latence de calcul et garantir la disponibilité de l'estimation pour le contrôleur à temps.

### 7.1.4 Choix de la méthode d'estimation

L'estimation par horizon glissant est la méthode d'estimation qui a été retenue pour l'état et les paramètres du cœur. Les méthodes basées sur le filtre de Kalman sont écartées (EKF et UKF cf section 1.2). Ce choix émane des particularités du système que sont la raideur et la non-linéarité du système, qui nécessitent chacune un traitement particulier pour l'implémentation du filtre de Kalman [KK18] (cf. sous-section 1.4.5). En outre, la possibilité

d'imposer des contraintes *a priori* sur les estimations est un avantage par rapport au filtre de Kalman étendu et à points sigma, qui peuvent mener à des estimations incohérentes du point de vue de la réalité physique [HR05].

L'intérêt de l'estimation MHE réside dans le traitement explicite des spécificités que nous venons d'énoncer (raideur, non-linéarité, contraintes sur les estimées, robustesse aux erreurs initiales d'estimation) directement au sein des contraintes du problème d'optimisation, au prix d'un coût de calcul plus élevé par rapport au filtre de Kalman.

Dans [Pon+13 ; Pon08], un estimateur MHE est développé pour l'estimation à horizon glissant du xénon et de l'iode sur un modèle de cœur basé sur un code de calcul neutronique 1D. Dans cette thèse, on présente un schéma d'estimation pour l'état du cœur pour un modèle adapté à la commande (prédictive) du réacteur en présence de bruit de mesure. En outre, nous proposons un ajustement du comportement du modèle au cours du cycle combustible via l'estimation des paramètres du modèle.

## 7.2 Implémentation de l'estimateur

L'estimateur MHE utilise comme modèle de conception le modèle détaillé en section 6.3. Il s'agit d'un modèle non-linéaire raide à 30 états, 8 entrées et 7 mesures (+6 mesures virtuelles de xénon dans l'hypothèse du couplage avec ARGOS). L'implémentation est réalisée avec CasADi 3.6.5 [And+19] dans l'environnement Matlab 2021b [MAT21]. La boîte à outils CasADi permet un calcul précis et rapide des quantités dérivées nécessaires au solveur d'optimisation par la différentiation automatique.

Les mesures acquises sur le système sont exploitées en considérant une fenêtre  $[t - H, t]$  sur laquelle le problème d'optimisation est formulé. L'ensemble de la fenêtre d'estimation est estimé à chaque itération ; nous choisissons de ne considérer que le dernier instant de la fenêtre pour évaluer l'erreur d'estimation. Une fois la solution trouvée, la fenêtre est déplacée de  $dt$  vers les temps croissants, les mesures acquises et les entrées du modèle sont interpolées aux instants correspondant aux variables de collocation dans la fenêtre. Enfin, le problème est résolu numériquement.

Pour apprécier les résultats d'estimation, on introduit les quantités suivantes :

$$\varepsilon_X = \frac{\|\hat{X}_{1..6} - X_{1..6}\|_2}{\|X_{1..6}\|_2}, \quad \varepsilon_I = \frac{\|\hat{I}_{1..6} - I_{1..6}\|_2}{\|I_{1..6}\|_2}, \quad \varepsilon_P = \frac{\|\hat{P}_{1..6} - P_{1..6}\|_2}{\|P_{1..6}\|_2}$$

$$\varepsilon_c = \frac{\|\hat{c}_{1..6} - c_{1..6}\|_2}{\|c_{1..6}\|_2}, \quad \varepsilon_T = \frac{\|\hat{T}_{2..7} - T_{2..7}\|_2}{\|T_{2..7}\|_2}$$

où  $X_{1..6}$ ,  $I_{1..6}$ ,  $P_{1..6}$ ,  $c_{1..6}$  et  $T_{1..6}$  sont les valeurs exactes en xénon, iode, puissance, population de précurseurs et température pour chaque maille.

### 7.2.1 Longueur et discrétisation de l'horizon d'estimation

Les dynamiques modélisées possèdent des temps de réponse très différents :

- très rapide (10ms) : neutrons prompts.
- rapide ( $\sim 1-10s$ ) : neutrons retardés et enthalpie massique.



— lente ( $\sim 1-10h$ ) : iode et xénon.

Les dynamiques lentes nécessitent de considérer un horizon d'estimation suffisamment long pour que puissent évoluer le xénon et l'iode estimé. Dans le même temps, les entrées du modèle évoluent plus rapidement (à l'échelle de quelques dizaines de secondes), il faut donc choisir une discrétisation suffisamment fine de l'horizon, en particulier dans le cadre de l'utilisation de la collocation directe qui exclut l'usage d'un intégrateur numérique avec adaptation du pas de temps. Pour ces raisons, l'horizon d'estimation retenu est  $H = 1h$  et le pas de discrétisation est  $dt = 2 \text{ min}$ , soit une division en de l'horizon d'estimation en 30 sous-intervalles.

## 7.2.2 Méthode d'intégration et transcription

À chaque itération de l'algorithme MHE, on résout un problème de contrôle optimal par les méthodes directes. La raideur du modèle et la longueur importante de l'horizon d'estimation à utiliser exclut d'office le tir simple pour la transcription du problème d'optimisation (chapitre 3). Les méthodes de transcription simultanées sont plus indiquées. Même en employant un solveur d'intégration adapté (IDAS[Hin+05 ; Gar+22]), la résolution du problème d'optimisation dans le contexte du tir multiple est difficile. La transcription par **collocation directe** avec les **points de Radau** et un polynôme interpolateur d'**ordre 3** est la méthode ayant produit les meilleurs résultats. Cette réussite s'explique en partie par le fait que les états de collocation faisant partie des variables de décision du problème, c'est l'ensemble de la trajectoire qui profite de la stratégie de globalisation et non pas simplement les nœuds de tir. En outre, la possibilité de repartir de la solution précédente du problème permet de réduire le temps de calcul nécessaire à chaque itération.

Le problème non-linéaire à résoudre se présente sous la forme d'un problème paramétré par l'estimation *a priori* de l'état  $\bar{x}$ , les entrées du système sur l'horizon  $U$  et les sorties  $Y$ . Afin de faire correspondre les entrées et sorties de simulation aux instants de la méthode de collocation,  $U$  et  $Y$  sont obtenus par interpolation linéaire des données de simulation aux nœuds de collocation.

$$P(\bar{x}, U_{[t-H,t]}, Y_{[t-H,t]}) : \min_{v, \hat{x}} \|\bar{x} - \hat{x}_{k-N-1}\|_S^2 + \sum_{i=k-N-1}^k \sum_{l=1}^s q_l \|v_{i,l}\|_R^2$$

sous contraintes  $\forall i \in [k-N+1; k]$ ,

$$\hat{x}_{i+1} = \sum_{l=1}^s b_l \hat{x}_{i,l}^c$$

$$\forall l \in \llbracket 1 \dots s \rrbracket,$$

$$v_{i,l} = y(t_{i,l}) - h(t_{i,l}, u(t_{i,l}), \hat{x}_{i,l}^c, \hat{p})$$

$$\sum_{m=1}^s a_{l,m} \hat{x}_{i,m}^c = f(t_{i,l}, u(t_{i,l}), \hat{x}_{i,l}^c, \hat{p})$$

$$\underline{X} \leq \hat{X} \leq \bar{X}$$

$$\hat{p} \in \mathbb{P}$$

$$\text{Où } \hat{X} = \begin{bmatrix} \hat{x}_{k-N} \\ \hat{x}_{k-N,1}^c \\ \hat{x}_{k-N,2}^c \\ \hat{x}_{k-N,3}^c \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{k-1} \\ \hat{x}_{k-1,1}^c \\ \hat{x}_{k-1,2}^c \\ \hat{x}_{k-1,3}^c \\ \hat{x}_k \end{bmatrix}, U = \begin{bmatrix} u_{k-N} \\ u_{k-N,1}^c \\ u_{k-N,2}^c \\ u_{k-N,3}^c \\ u_1 \\ \vdots \\ u_{k-1} \\ u_{k-1,1}^c \\ u_{k-1,2}^c \\ u_{k-1,3}^c \\ u_k \end{bmatrix}, Y = \begin{bmatrix} y_{k-N} \\ y_{k-N,1}^c \\ y_{k-N,2}^c \\ y_{k-N,3}^c \\ y_1 \\ \vdots \\ y_{k-1} \\ y_{k-1,1}^c \\ y_{k-1,2}^c \\ y_{k-1,3}^c \\ y_k \end{bmatrix}, f \text{ est le modèle décrit en sous-section 6.3.4,}$$

$h$  la fonction de mesure,  $\underline{X}$  et  $\overline{X}$  correspondent aux bornes inférieures et supérieures imposées à l'état,  $R$  est la pondération sur l'erreur d'innovation (écart entre les mesures acquises et estimées),  $S$  est la pondération du coût d'arrivée :

$$V_F(t-H) = \|\overline{x} - \hat{x}_{k-N-1}\|_S^2$$

$\overline{x}$  est choisi comme l'estimée de l'état  $\hat{x}_{k-N}$  calculé à l'itération précédente. Les termes  $a_{l,m}$  sont les termes du tableau de Butcher de la méthode de collocation (cf tableau 2.4).

Les contraintes sur les variables d'états sont données en tableau 7.1. Elles ne sont pas très restrictives mais permettent d'assurer la cohérence des estimations du point de vue de la réalité physique, notamment la positivité (concentration, température, puissance).

Variable d'état	$X_i$	$I_i$	$h_i$	$c_i$	$P_i$
Borne inférieure	0	0	$T_{T \rightarrow ent}(250)$	0	0
Borne supérieure	10	10	$T_{T \rightarrow ent}(350)$	$1 \times 10^6$	100

TABLE 7.1 – Contraintes imposées aux variables d'états estimées.

Le temps de calcul nécessaire à la résolution du problème dépend du nombre de variables de décision. Pour un horizon divisé en  $N$  sous-intervalles et une méthode de collocation d'ordre  $d$ , un système de dimension  $n$ , le nombre de variables de décision  $N_{\text{var}}$  est donné par :

$$N_{\text{var}} = (N + 1 + dN)n$$

Pour un horizon d'une heure, une méthode de collocation d'ordre 3 et un pas de temps de discrétisation d'une minute, on dénombre ainsi 7230 variables de décision. Le temps de résolution moyen pour un problème de cette taille est de l'ordre de la seconde sur un processeur Intel Core i5-4200 sur lequel les résultats que nous présentons ont été obtenus. Le temps de calcul croît linéairement avec la taille du problème d'optimisation.

### 7.2.3 Solveur d'optimisation

Le solveur d'optimisation employé est IPOPT[BZ09], implémentant un algorithme des points intérieurs avec une stratégie de filtre. Il contient de nombreux paramètres de réglages mais les paramètres du tableau 7.2 doivent être ajustés pour le démarrage à chaud de la solution. On impose également un temps maximal de calcul de 60s afin de limiter le temps

de simulation total. La solution est en général trouvée bien avant d'atteindre cette limite temporelle. Lorsque cette limite est atteinte, les itérations de l'algorithme de résolution du problème sont interrompues et la solution courante est retournée par le programme.

Paramètre	Valeur
max_cpu_time	60s
warm_start_init_point	yes
warm_start_entire_iterate	yes
warm_start_bound_push	1e-8
warm_start_slack_bound_push	1e-8
warm_start_mult_bound_push	1e-8
mu_init	1e-8

TABLE 7.2 – Paramètres de réglage d'IPOPT.

#### 7.2.4 Stratégie de réduction de la latence de calcul

L'estimateur implémenté nécessite un effort de calcul non négligeable dans le pire des cas. Cet effort de calcul se traduit par un temps de résolution encore plus important si elle est effectuée sur un ordinateur de puissance modeste. L'approche de démarrage à chaud de la solution, simple à implémenter, permet de réduire le temps de calcul mais ne suffit pas en général. Un schéma correcteur/prédicteur [DSD11] comme l'advanced-step MHE [LB12] permet de réduire le temps de latence entre l'arrivée de nouvelles données et la disponibilité de la solution au programme non-linéaire associé. De manière contre-intuitive, ces schémas ne réduisent pas le temps de calcul nécessaire à la détermination de la solution (ils demandent davantage de calcul), mais déplacent la résolution du problème *avant* que les nouvelles données soient disponibles.

La stratégie *advanced-step MHE* (sous-section 1.5.2) est employée pour réduire la latence de calcul. Les variables duales pour un programme non-linéaire résolu par une méthode des points intérieurs ne sont pas exactes, ce qui nuit au calcul des sensibilités de la solution. Pour s'affranchir de ce problème, un second solveur basé sur la méthode SQP est utilisé pour le calcul des sensibilités de la solution. Une fois la solution déterminée par IPOPT, le solveur SQP est initialisé à partir des variables primales du problème puis la sensibilité de cette solution est utilisée.

La solution du problème  $P(\bar{x}, U_{[t-H,t]}, Y_{[t-H,t]})$  est obtenue en 3 étapes :

1. Calcul de la solution avec IPOPT :

$$X_{\text{IPOPT}}^- = \operatorname{argmin} P(\bar{x}, U'_{[t-H,t]}, Y'_{[t-H,t]}) \quad (7.1)$$

avec  $U'_{[t-H,t]} = [U_{t-H,t-dt}^T \quad \bar{U}_{[t-dt,t]}]$ ,  $\bar{U}_{[t-dt,t]}$  correspondant à un signal constant de valeur  $U_t$ .  $Y'_{[t-H,t]}$  est obtenu de manière équivalente.

2. Calcul de la solution avec le solveur SQP :

$$X_{\text{SQP}}^- = \operatorname{argmin} P(\bar{x}, U'_{[t-H,t]}, Y'_{[t-H,t]})$$

et calcul de la sensibilité de la solution par rapport aux paramètres d'entrée  $U'_{[t-H,t]}, Y'_{[t-H,t]}$  :

$$S_{X_{SQP}^-} = \frac{\partial X_{SQP}^-}{\partial U'_{[t-H,t]}, Y'_{[t-H,t]}} \quad (7.2)$$

3. Correction de la solution avec les données  $U_{[t-dt,t]}$  :

$$X = X_{IPOP}^- + S_{X_{IPOP}^-} \left( \begin{bmatrix} U_{[t-H,t]} \\ Y_{[t-H,t]} \end{bmatrix} - \begin{bmatrix} U'_{[t-H,t]} \\ Y'_{[t-H,t]} \end{bmatrix} \right) \quad (7.3)$$

On remarque que les étapes 1 et 2 ne dépendent pas de la connaissance des données sur  $[t-dt, t]$ , ce qui signifie que l'on peut les réaliser à l'instant  $t-dt$  et attendre la disponibilité des données sur le dernier sous-intervalle de l'horizon pour l'étape 3. L'algorithme asMHE fonctionne donc en 2 phases :

- calcul du *feedback* (eq. 7.3) des données disponibles immédiatement.
- préparation de la solution suivante (eq. 7.1) et calcul de la sensibilité (eq. 7.2).

Les étapes 1 et 2 sont celles qui demandent le plus de temps, tandis que l'étape 3 consiste uniquement en un calcul matriciel qui ne prend qu'une fraction du temps nécessaire à la résolution d'un programme non-linéaire et donne des garanties d'exécution en temps réel. L'effort de calcul supplémentaire représenté par l'étape 2 n'impacte pas le temps entre la disponibilité de la solution et la détermination de la solution. Un autre avantage à cette approche, par rapport à l'usage d'un filtre de Kalman ou estimateur de Luenberger sur  $[t-dt, t]$  et que l'estimation sur l'ensemble de l'horizon est corrigée.

Les résultats d'estimation qui seront présentés dans la suite font apparaître que la majorité des itérations de l'estimateur MHE<sup>a</sup> peuvent être résolues en un temps de l'ordre de la seconde à la dizaine de secondes, parfois de l'ordre de la minute dans le pire des cas. Dans la majorité des cas, la solution est donc obtenue en un temps « raisonnable » n'impactant pas significativement l'algorithme de commande prédictive. Néanmoins, la stratégie asMHE permet le déploiement de l'algorithme sur un matériel de plus faible puissance (ce qui permet de réduire le coût du matériel) ou effectuant d'autres tâches de manière concurrente (signifiant une interruption du programme et un allongement du temps de calcul effectif).

### 7.3 Application avec un modèle parfait de réacteur

On se place ici dans le contexte idéal où le modèle et le process sont identiques. On utilise les données du transitoire décrit en annexe B.

#### 7.3.1 Estimation de l'état

On traite dans un premier temps le problème d'estimation de l'état. Les mesures de flux neutronique et de température sont supposées bruitées avec un bruit blanc d'écart-

a. Une itération de l'algorithme MHE = résolution jusqu'au bout d'un problème d'estimation.

type  $\sigma = 0,2$ . On s'intéresse au comportement de l'estimateur en réponse au choix des paramètres de réglage de l'estimateur :

- le pas de temps de l'échantillonnage de la méthode d'intégration  $dt$ ,
- la taille de l'horizon d'estimation  $H$  (exprimée en nombre de pas de temps d'échantillonnage  $N = H/dt$ ),
- les matrices de pondération du problème d'estimation.

La fonction coût du problème d'optimisation résolue au sein de l'estimateur formalise la recherche d'un compromis entre :

- la confiance avec l'estimation *a priori* de l'état au début de la fenêtre : cette estimation s'interprète comme une prédiction à un pas à partir du début de la fenêtre de l'itération précédente de l'estimateur.
- la cohérence entre mesures prédites et acquises.

Pour  $S$  et  $R$  gardés constants, la variation de la longueur de l'horizon d'estimation modifie la contribution du terme d'innovation par rapport au terme de coût d'arrivée. L'allongement de l'horizon modifie la recherche du compromis en pénalisant davantage les termes d'innovations que le fonction coût. Dans toute la suite, on pose :

$$P^{-1} = 0.01^2 I_{30} \text{ et } R^{-1} = 0.2^2 I_7$$

L'estimation initiale *a priori* est initialisée à partir des mesures pour les variables d'état mesurées et les autres sont initialisées à des valeurs aléatoires centrées autour de leur valeur exacte, en introduisant une erreur moyenne de 20%.

### Rôle du coût d'arrivée dans le filtrage du bruit de mesure

On compare tout d'abord le comportement des estimées lorsque le terme de coût d'arrivée est nul. Dans ce cas, aucune information des itérations précédentes de la MHE n'est exploitée pour estimer l'état du système. La figure 7.3 donne les résultats d'estimation de l'état du système pour un horizon  $H = 1\text{min}$  et  $H = 1\text{h}$ .

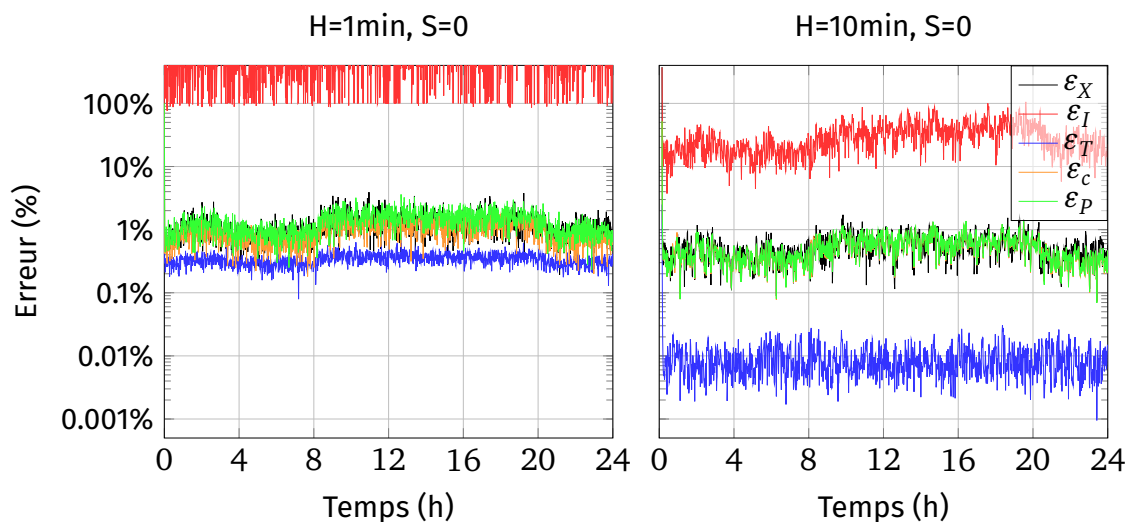


FIGURE 7.3 – Résultats d'estimation en l'absence de terme de coût d'arrivée.

Dans les deux cas, on constate que les variables d'état ayant un impact direct sur celles mesurées<sup>a</sup> sont immédiatement estimées avec une erreur de quelques pourcents. Pour ces variables d'état, il est possible de déterminer des valeurs cohérentes au regard de la dynamique et des mesures, avec des erreurs d'estimation de quelques pourcents. Ces estimations sont très sensibles au bruit de mesure ; l'allongement de l'horizon permet de réduire la sensibilité au bruit.

En revanche, les valeurs d'iode sont totalement incohérentes. L'allongement de l'horizon à 10 minutes permet d'exploiter davantage la lente dynamique du xénon et de l'iode, ce qui permet de réduire l'erreur d'estimation par rapport à l'horizon  $H = 1 \text{ min}$ , mais l'iode est toujours utilisé comme variable d'ajustement pour l'estimation des autres variables d'état, il est donc inexploitable. L'allongement de l'horizon permet de réduire l'erreur d'estimation de l'iode, sans être satisfaisante pour autant. Nous nous penchons dans la suite sur le réglage de la matrice de pondération  $S$  du terme de coût d'arrivée.

### Réglage de la matrice de pondération $S$ et longueur de l'horizon

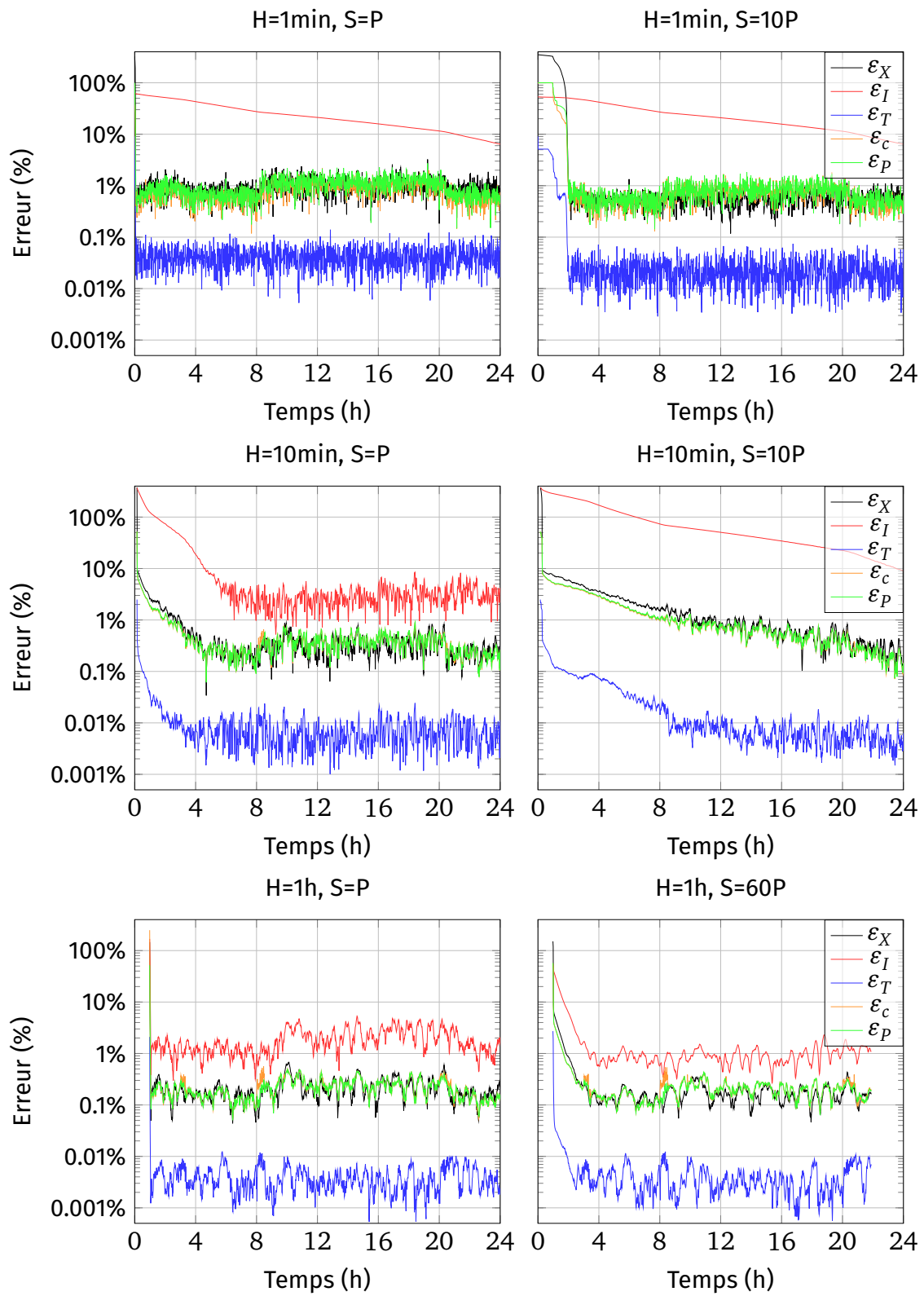
Le terme de coût d'arrivée est désormais non nul. Le choix d'un horizon d'estimation comportant un unique pas d'intégration permet d'apprécier l'impact du choix des matrices de pondération sur le compromis entre le terme de coût d'arrivée et le coût d'innovation. Les résultats de l'estimation de l'état pour un horizon d'une minute (un seul pas de temps) sont donnés en (figure 7.4).

On constate immédiatement pour  $S = P$  que l'introduction du terme de coût d'arrivée permet la convergence (bien que lente) de l'erreur d'estimation de l'iode. Le choix  $S = 10P$  pénalise davantage le terme de coût d'arrivée, ce qui introduit davantage d'information de l'itération précédente. Le bruit est mieux filtré, ce qui se caractérise par une erreur d'estimation plus faible.

L'allongement de l'horizon à  $H = 10 \text{ min}$  minutes pour le choix  $S = P$  mène à une contribution du coût d'innovation importante dans la fonction coût et une erreur d'estimation inférieure à  $H = 1 \text{ min}$ . En posant  $S = 10P$ , la convergence de l'erreur d'estimation est bien plus lente, mais le bruit de mesure est davantage filtré. En allongeant encore la taille de l'horizon à  $H = 1 \text{ h}$  et avec  $S = P$ , l'erreur d'estimation converge très rapidement, en quelques itérations.

---

a. C'est à dire les variables de xénon, de température et de puissance neutronique.

FIGURE 7.4 – Impact du choix de la pondération  $S$  et de l'horizon  $H$  sur l'erreur d'estimation.

## Conclusion

Résumons les résultats que nous venons de présenter :

- L’allongement de l’horizon d’estimation désensibilise les estimations au bruit de mesure et réduit l’erreur d’estimation. En effet, les données exploitées au cours d’un horizon plus long sont plus riches en informations.
- L’introduction du coût d’arrivée a la même effet, notamment sur l’iode qui n’est plus utilisé comme variable d’ajustement pour l’estimation des autres variables de décision. De plus, le réglage de  $S$  permet de contrôler la vitesse de convergence des estimées. Une pondération plus forte sur le terme de coût d’arrivée désensibilise les estimations au bruit de mesures, au prix d’une convergence plus lente de l’erreur d’estimation.
- La pondération  $S$  doit être ajustée en accord avec l’allongement de l’horizon. Une fois la vitesse de convergence réglée avec un choix  $S_1$  pour un horizon  $H_1$ , le choix de pondération  $S_2 = \frac{H_2}{H_1} S_1$  pour l’horizon  $H_2$  convient pour conserver une vitesse de convergence similaire. La vitesse de convergence de l’erreur d’estimation résultante n’est pas identique, car l’allongement de l’horizon contribue également à la réduction de l’erreur d’estimation.

Le réglage de  $S$  dans le contexte déterministe de la MHE est réalisé par essai/erreur et dépend du choix de la matrice  $R$ . L’emploi d’une pondération constante au cours des itérations de l’algorithme MHE est sous-optimal : dans le cadre stochastique,  $S$  s’interprète comme la confiance que l’on a dans l’estimation en début de fenêtre *a priori*  $\bar{x}$ . Or, à mesure des itérations de l’algorithme, la confiance que l’on accorde à  $\bar{x}$  évolue. Pour un pondération variable  $S_k, k \in \mathbb{N}$ , l’ajustement de  $S_k$  en fonction de la confiance accordée à  $\bar{x}$  permet de modifier en ligne le compromis réalisé entre la mesure (coût d’innovation) et la prédiction de manière automatisée.

### 7.3.2 Estimation conjointe de l’état et des paramètres

#### Paramètres d’intérêt du modèle et implémentation

On ajoute désormais des paramètres d’intérêt du modèle dans les variables à estimer. L’objectif de l’estimation des paramètres est la représentativité du modèle pour la commande prédictive. Dans l’hypothèse à un groupe, le spectre d’émission en énergie des neutrons est approximé par des neutrons monocinétiques. À mesure que le combustible s’épuise, le spectre d’émission change, ce qui se caractérise dans l’hypothèse à un groupe d’énergie par un changement de la vitesse moyenne des neutrons. Les réactions nucléaires se produisant dans le cœur dépendent de l’énergie des neutrons incident (section-efficace). Dans notre modèle, ces réactions sont la capture des neutrons par le xénon, le bore ou les barres de contrôle ainsi que la fission du combustible.

Pour la commande prédictive, le modèle de prédiction doit correctement modéliser l’effet des actionneurs sur le système. On s’intéresse ainsi aux paramètres :

- $K_{\text{rod}, X1}, K_{\text{rod}, X2}, K_{\text{rod}, X3}$  : efficacité différentielle des groupes de barres X1, X2 et X3. On ne considère pas  $K_{\text{rod}, X4}$  et  $K_{\text{rod}, X5}$  dans la mesure où les groupes X4 et X5 restent totalement extraits du cœur au cours du transitoire de validation.
- $K_{\text{bore}}$  : efficacité différentielle du bore sur la réactivité.



Ces termes d'efficacité dépendent du spectre d'émission en énergie des neutrons

Une seconde source d'incertitude porte sur les autres termes du bilan de réactivité en chaque maille. Son calcul dépend de la réactivité nominale en chaque maille  $\rho_{\text{nom},i}$  dont on rappelle la forme :

$$\rho_i = \rho_{\text{nom},i} + \Delta\rho_i(X_i, Cb, h_{\text{rods}}, P_i, \bar{h}_i)$$

où  $\bar{h}_i$  est l'enthalpie massique moyenne de la maille  $i$  et  $\rho_{\text{nom},i}$  est la réactivité du cœur nominal (100%PN) critique. On s'intéresse aux paramètres suivants :

- $D$  : coefficient d'échange de neutrons entre les mailles du modèle. Il intervient dans l'équation d'évolution de la puissance neutronique et dans la réactivité du cœur nominal critique :

$$\rho_{\text{nom},1} = D \frac{P_{\text{nom},2} - P_{\text{nom},1}}{P_{\text{nom},1}} \quad (7.4a)$$

$$\rho_{\text{nom},i} = D \frac{P_{\text{nom},i+1} - 2P_{\text{nom},i} + P_{\text{nom},i-1}}{P_{\text{nom},i}}, \quad \forall i \in \llbracket 2, 5 \rrbracket \quad (7.4b)$$

$$\rho_{\text{nom},1} = D \frac{P_{\text{nom},5} - P_{\text{nom},6}}{P_{\text{nom},6}} \quad (7.4c)$$

- $\Sigma_X$  : section efficace de capture neutronique par le xénon. Le terme  $\Delta\rho_i$  dépend du xénon, dont la contribution sur la réactivité est  $K_X(X_i - X_{\text{nom},i})$  où  $X_{\text{nom},i}$  est la concentration en xénon dans le cœur nominal, calculé en résolvant les équations du modèle d'évolution de xénon/iode à l'équilibre :

$$I_{\text{nom},i} = \frac{\Gamma_I}{\lambda_I} P_{\text{nom},i} \quad \text{et} \quad X_{\text{nom},i} = \frac{\lambda_I I_{\text{nom},i} + \Gamma_X P_{\text{nom},i}}{\lambda_X - \Sigma_X P_{\text{nom},i}}. \quad (7.5)$$

Ce paramètre intervient également dans l'équation d'évolution du xénon et caractérise la consommation du xénon par capture neutronique.

Afin d'assurer que les signaux dont on dispose sur l'horizon d'estimation sont suffisamment riches en information, on choisit un horizon d'estimation  $H = 4h$  et un pas d'intégration de  $dt = 150s$ . Les paramètres sont initialisés à des valeurs erronées (aléatoirement choisies pour correspondre à au moins 10% d'erreur) et les valeurs admissibles sont contraintes à un intervalle correspondant à  $[0.5p, 1.5p]$  où  $p$  est la valeur réelle. Ces contraintes sont essentielles pour assurer d'une part que les paramètres estimés seront cohérents avec la réalité et d'autre part pour réduire l'espace de recherche des solutions.

Le problème d'estimation conjointe état/paramètres est implémenté en considérant un modèle augmenté tel que décrit par (eq. 1.30) et on suppose que les paramètres évoluent lentement, de sorte qu'ils sont considérés constants au cours de l'horizon d'estimation. Le modèle augmenté est redonné ici :

$$\begin{cases} \begin{pmatrix} \dot{x} \\ \dot{p} \end{pmatrix}(t) = \begin{pmatrix} f(x(t), u(t), p(t)) \\ 0 \end{pmatrix} \\ y(t) = h(x(t), u(t), p(t)) \end{cases} \quad (7.6)$$

Le coût d'arrivée du problème d'optimisation est pondéré par la matrice :

$$S = N \begin{bmatrix} S_x & 0 \\ 0 & S_p \end{bmatrix} \quad (7.7)$$

où  $N$  est le nombre de pas d'intégration au cours de l'horizon,  $S_x \in \mathbb{R}^{n_x \times n_x}$  est la pondération sur les variables d'état et  $S_p \in \mathbb{R}^{n_x \times n_x}$  est la pondération sur les paramètres.  $S_x = 0.01^2 I_{30}$  et  $S_p$  prend la forme d'une matrice diagonale :

$$S_p^{-1} = \begin{bmatrix} \sigma_{p_1}^2 & \dots & 0 \\ \vdots & & \vdots \\ 0 & \dots & \sigma_{p_k}^2 \end{bmatrix}$$

où  $\sigma_{p_k}$  est la pondération sur l'écart entre le  $k$ -ième paramètre et la valeur  $\bar{p}_k$  déterminée à l'itération précédente de l'estimateur.  $S_p$  est ajustée par essai/erreur.

#### Paramètres $K_{\text{rod}, X1}, K_{\text{rod}, X2}, K_{\text{rod}, X3}, K_{\text{bore}}$

Dans un premier temps, on s'intéresse à l'estimation des paramètres  $K_{\text{rod}, X1}, K_{\text{rod}, X2}, K_{\text{rod}, X3}$  et  $K_{\text{bore}}$ . Considérons dans un premier temps que les mesures sont parfaites et non bruitées, ce qui constitue le scénario idéal. Les résultats d'estimation des paramètres sont présentés en (figure 7.5).

On constate que l'erreur d'estimation de l'état converge rapidement vers 0 (ou presque, l'erreur est d'une fraction de pourcent), il en va de même pour l'erreur d'estimation de  $K_{\text{bore}}$ . En revanche, l'erreur d'estimation de l'efficacité différentielle des barres converge bien plus lentement. Les erreurs d'estimation se compensent, du fait de l'insertion collective des groupes de barres en mode A. L'effet neutrophage des barres se superpose, si bien qu'une surestimation de l'efficacité d'un groupe de barres peut être compensée par une sous-estimation de l'effet des autres groupes. Néanmoins, les mouvements de barres au cours du transitoire permettent de différencier l'effet des différents groupes, puisqu'ils mènent à différentes configurations d'insertion des barres dans les mailles. Ainsi l'erreur d'estimation converge plus rapidement lorsque l'horizon d'estimation contient des mouvements importants des barres de contrôle. Une fois que toutes les données ont été traitées, on observe une erreur d'estimation de l'ordre du pourcent pour  $K_{\text{rods}, X1}$  et  $K_{\text{rods}, X3}$ , tandis qu'elle est un peu plus élevée pour  $K_{\text{rods}, X2}$ . Le bore intervient également dans le phénomène de compensation des estimations, mais dans la mesure où son effet est uniforme dans tout le cœur, cette compensation est moindre puisque les barres de contrôle ne s'insèrent pas jusque dans les mailles inférieures du cœur. Ainsi, il est estimé avec une faible erreur, inférieure à 0.1%.

On considère maintenant que les mesures sont bruitées d'un bruit gaussien d'écart-type  $\sigma = 0.2$ , comme pour l'estimation de l'état seul. Ces résultats sont présentés en figure 7.6. On constate dans ce cas que l'erreur d'estimation des paramètres est bien plus élevée. Les paramètres à estimer sont alors des variables d'ajustement pour faire correspondre les mesures prédites et acquises : leur estimation est sensible au bruit. En conséquence, l'estimation de l'état est détériorée par rapport au cas non bruité, tout en restant satisfaisante, avec moins d'un pourcent d'erreur. Néanmoins, on constate une amélioration des paramètres estimés, puisqu'ils étaient tous initialisés avec au moins 20% d'erreur. Les résultats sont

résumés en tableau 7.4.

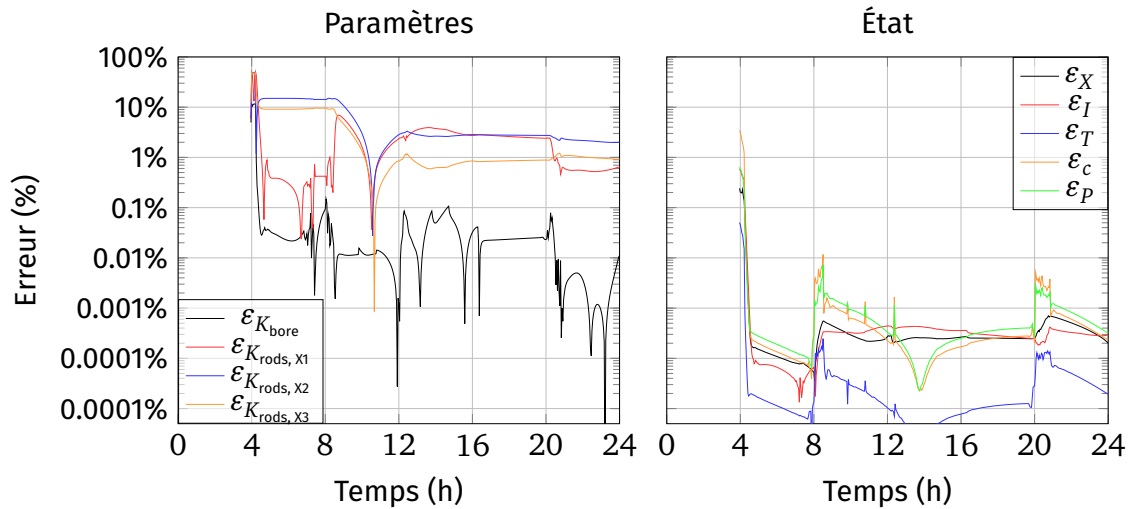


FIGURE 7.5 – Estimation de l'état et des paramètres  $K_{rod, X1}$ ,  $K_{rod, X2}$ ,  $K_{rod, X3}$  et  $K_{bore}$  pour des mesures non bruitées.

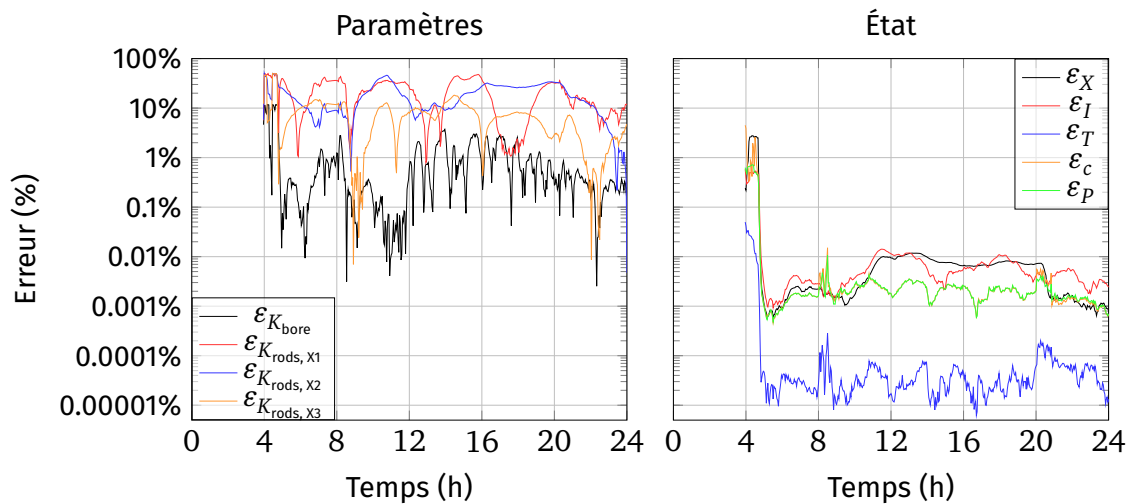


FIGURE 7.6 – Estimation de l'état et des paramètres  $K_{rod, X1}$ ,  $K_{rod, X2}$ ,  $K_{rod, X3}$  et  $K_{bore}$  pour des mesures bruitées.

### Paramètres $D$ , $\Sigma_X$

On propose à présent d'estimer les paramètres  $D$  et  $\Sigma_X$ . On considère dans un premier temps le cas où les mesures sont non bruitées (figure 7.8). Les paramètres et l'état convergent tous vers des erreurs d'estimation finales inférieures au pourcent. À l'exception du xénon, les estimations pour toutes les variables d'état convergent en quelques itérations à une erreur inférieure à 1%. Le xénon, du fait de la mauvaise estimation du paramètre  $\Sigma_X$ , prend davantage de temps à converger. On constate que  $\varepsilon_X$  et  $\varepsilon_{\Sigma_X}$  évoluent au même rythme. L'erreur d'estimation du paramètre  $D$  converge plus rapidement, puisqu'il est doublement impliqué ( $\rho_{nom}$  et échange de neutrons) dans la dynamique de la puissance neutronique, qui est mesurée. On observe des « pics » de  $\varepsilon_D$ , qui coïncident avec les mouvements importants

Paramètre	$K_{rod, X1}$	$K_{rod, X2}$	$K_{rod, X3}$	$K_{bore}$
Valeur admissible	[-1.5, -0.5]	[-3, -1]	[-4, -2]	[-10, -3]
Valeur réelle	-1	-1.8	-2.9	-6.8
Pénalisation $\sigma_p$	0.1	0.1	0.1	0.1
Initialisation	-1.25	-2.2	-2.3	-5
Erreur initiale	25%	22.2%	20.6%	26%
Valeur finale (sans bruit de mesure)	-1.006	-1.764	-2.921	-6.799
Erreur finale (sans bruit de mesure)	0.65%	2%	0.92%	0.01%
Valeur finale (bruit de mesure)	-1.115	-1.799	-2.761	-6.782
Erreur finale (bruit de mesure)	11.55%	0.004%	4.77%	0.25%

TABLE 7.3 – Résultats de l'estimation conjointe des paramètres  $K_{rod, X1}$ ,  $K_{rod, X2}$ ,  $K_{rod, X3}$ ,  $K_{bore}$  et de l'état.

des barres de contrôle dans le cœur (cf annexe B) lors des variations de la référence de puissance à  $t = 8h$  et  $t = 20h$ . Ces nouvelles données riches en information contribuent favorablement à l'estimation des paramètres.

L'introduction de bruit dans les mesures figure 7.8 détériore  $\varepsilon_D$  ainsi que sa vitesse de convergence. L'estimation du xénon est moins impactée par le bruit de mesure qui est filtré par la lente dynamique de l'iode (premier ordre) et du xénon. Néanmoins, les valeurs pour les paramètres estimés restent satisfaisantes, puisque les erreurs d'estimation à la fin de l'expérience sont inférieures à 2%.

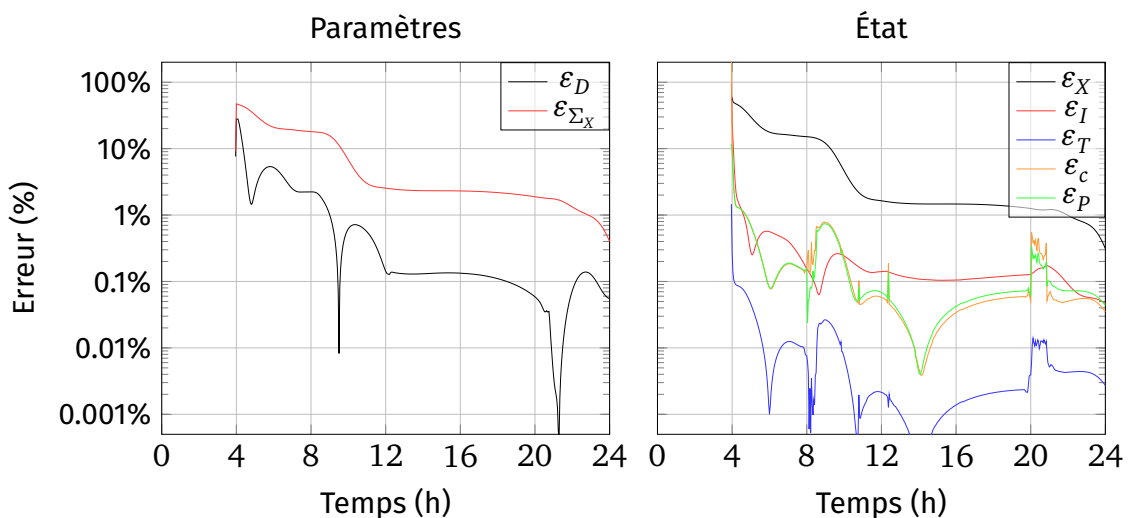
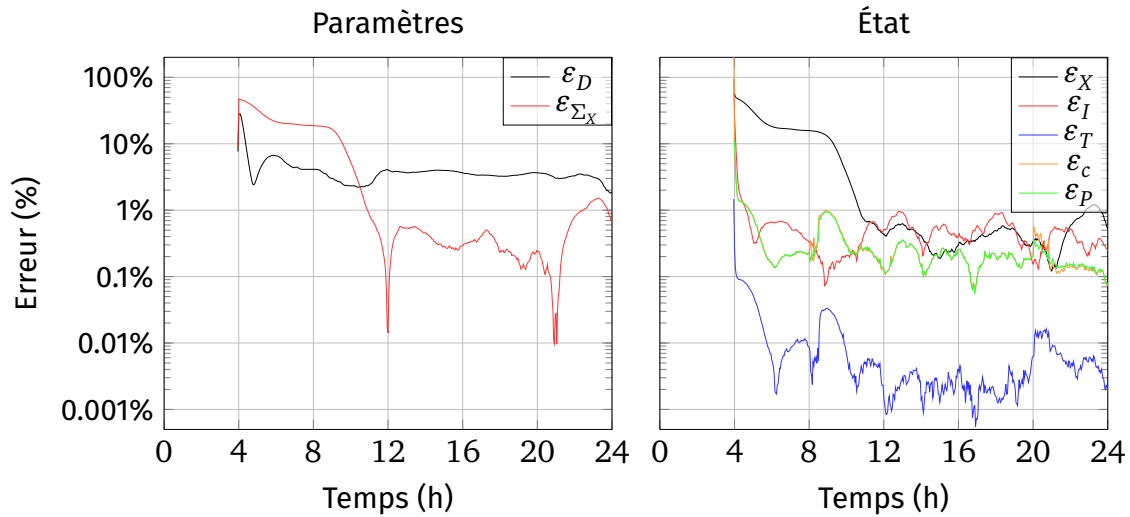


FIGURE 7.7 – Estimation de l'état et des paramètres  $D$  et  $\Sigma_X$  pour des mesures non bruitées.

## 7.4 Validation sur les données de SOFIA

On présente ici des résultats d'application de l'estimateur sur des données issues du simulateur SOFIA. Ce scénario de validation nous rapproche du comportement réel d'un réacteur. Dans ce cas, le modèle d'estimation et le process sont différents, ce qui constitue une difficulté supplémentaire pour l'estimation de l'état et des paramètres du système. Le couplage avec un modèle expert tel qu'ARGOS permet d'exploiter davantage d'informations

FIGURE 7.8 – Estimation de l'état et des paramètres  $D$  et  $\Sigma_X$  pour des mesures bruitées.

Paramètre	$D$	$\Sigma_X$
Valeur admissible	$[-400, -1300]$	$[3, 7] \times 10^{-7}$
Valeur réelle	860	$5.5 \times 10^{-5}$
Pénalisation $\sigma_p$	5	$10^3$
Initialisation	700	4
Erreur initiale	18.6%	27.3%
Valeur finale (sans bruit de mesure)	860.47	$5.478 \times 10^{-7}$
Erreur finale (sans bruit de mesure)	0.05%	0.40%
Valeur finale (bruit de mesure)	876.12	$5.45 \times 10^{-7}$
Erreur finale (bruit de mesure)	0.62%	1.87%

TABLE 7.4 – Résultats de l'estimation conjointe des paramètres  $D$ ,  $\Sigma_X$  et de l'état.

que les seules mesures acquises. Du fait de la difficulté accrue du problème du fait de l'écart entre le système réel et le modèle, seul le problème d'estimation de l'état est traité.

#### 7.4.1 Préliminaire : calibration du modèle de conception avec le modèle expert

La représentativité des prédictions du modèle dépend grandement de l'exactitude des nombreux paramètres du modèle. Les paramètres impliqués dans le bilan de réactivité du modèle, du fait de la grande sensibilité de la puissance en chaque maille aux contre-réactions, doivent impérativement être corrects. Le transitoire de référence (annexe C) est utilisé pour pré-calibrer le modèle ; il correspond à un cœur proche du début du cycle ainsi le modèle de conception utilise la paramétrisation de début de cycle (notée DDV pour début de vie). Les paramètres sont ajustés en les estimant grâce aux données du transitoire de validation qui sont plus riches que les mesures acquises sur le système.

Le modèle expert peut utiliser une discrétisation du cœur différente du modèle de conception. C'est le cas de SOFIA qui exploite un modèle de résolution spatiale supérieure (cf figure 6.1). Les données du modèle expert sont ainsi d'abord projetées sur le modèle de conception à 6 mailles avant d'être exploitées pour l'identification des paramètres du modèle. La perte en résolution spatiale dans le modèle à 6 mailles implique une erreur de modélisation, puisque chaque « macro-maille » du modèle PWR\_simu représente le comportement collectif de plusieurs mailles du modèle expert.

On estime les paramètres du modèle en résolvant un problème d'optimisation aux moindres carrés (eq. 1.29). Le problème est transcrit en un programme non-linéaire par collocation directe (avec interpolation des entrées et sorties telle que décrite en section 7.2) et résolu avec le solveur IPOPT.

Estimer l'ensemble des paramètres du modèle en une fois est souhaitable pour régler le comportement global du modèle, quitte à faire des erreurs par la compensation de paramètres. Cependant, l'absence de la totalité de l'état dans les données du transitoire de validation nous limite à considérer certaines dynamiques séparément.

On se concentre dans un premier temps sur les paramètres intervenant dans le bilan de réactivité en chaque maille. Les données de simulation contiennent toutes les quantités impliquées dans les contre-réactivités : bore, xénon, température (effet modérateur), position des groupes de barres de contrôle et puissance neutronique (effet Doppler). L'exactitude du bilan de réactivité du modèle peut être quantifiée directement par la mesure des puissances neutroniques. Le problème d'optimisation suivant est résolu pour identifier les paramètres

du modèle :

$$\min_{K_{\text{xénon}}, K_{\text{mod}}, K_{\text{bore}}} \int_0^{t_f} \sum_{i=1}^6 (P_i(\tau) - \hat{P}_i(\tau))^2 d\tau \quad (7.8a)$$

$$\text{sous contraintes, } \forall \tau \in [0, t_f], \quad (7.8b)$$

$$\frac{d\hat{P}_1}{dt} = \frac{1}{l}(\rho_1 \hat{P}_1 + D(\hat{P}_2 - \hat{P}_1)) \quad (7.8c)$$

$$\frac{d\hat{P}_i}{dt} = \frac{1}{l}(\rho_i \hat{P}_i + D(\hat{P}_{i+1} - 2\hat{P}_i + \hat{P}_{i-1})) \quad \forall i \in \llbracket 2, 5 \rrbracket \quad (7.8d)$$

$$\frac{d\hat{P}_6}{dt} = \frac{1}{l}(\rho_6 \hat{P}_6 + D(\hat{P}_5 - \hat{P}_6)) \quad (7.8e)$$

$$\rho_i = \rho_{\text{nom},i} + \Delta\rho_{\text{rods},i} + \Delta\rho_{X,i} + \Delta\rho_{\text{bore}} + \Delta\rho_{\text{mod},i} + \Delta\rho_{\text{dop},i} \quad (7.8f)$$

$$\hat{P}_i(0) = P_i(0), \quad \forall i \in \llbracket 1, 6 \rrbracket \quad (7.8g)$$

$$K_{\text{xénon}} > 0, K_{\text{mod}} > 0, K_{\text{bore}} > 0 \quad (7.8h)$$

Le paramètre  $\Gamma_I$  de la dynamique de l'iode est identifié en utilisant le transitoire SOFIA, correspondant à un même avancement dans le cycle combustible, mais qui contient les données d'iode en chaque maille, contrairement au transitoire annexe C (cette simulation modélise bien l'iode, mais n'est pas disponible dans les données de simulation). Le problème d'optimisation suivant est résolu :

$$\min_{\Gamma_I} \int_0^{t_f} \sum_{i=1}^6 (I_i(\tau) - \hat{I}_i(\tau))^2 d\tau \quad (7.9a)$$

$$\text{sous contraintes, } \forall \tau \in [0, T], \quad \forall i \in \llbracket 1, 6 \rrbracket, \quad (7.9b)$$

$$\dot{\hat{I}}_i(t) = \Gamma_I P_i(t) - \lambda_I \hat{I}_i(t) \quad (7.9c)$$

$$\hat{I}_i(0) = I_i(0) \quad (7.9d)$$

où  $P_i$  fait office d'entrée du modèle et  $I_i$  fait office de sortie mesurée, toutes issues des données. Le problème est résolu comme précédemment. La solution du problème d'identification du paramètre  $\Gamma = 2.9e - 7$  est identique à la paramétrisation début de vie.

Enfin les paramètres de la dynamique du xénon sont identifiés en résolvant le problème d'optimisation suivant :

$$\min_{\Sigma, \Gamma_X} \int_0^{t_f} \sum_{i=1}^6 (X_i(\tau) - \hat{X}_i(\tau))^2 d\tau \quad (7.10a)$$

$$\text{sous contraintes, } \forall \tau \in [0, T], \quad \forall i \in \llbracket 1, 6 \rrbracket, \quad (7.10b)$$

$$\dot{\hat{X}}_i(t) = \lambda_I I_i(t) - \lambda_X \hat{X}_i(t) + \Gamma_X P_i(t) - \Sigma_X P_i(t) \hat{X}_i(t) \quad (7.10c)$$

$$\dot{I}_i(t) = \Gamma_I P_i(t) - \lambda_I I_i(t) \quad (7.10d)$$

$$\hat{X}_i(0) = X_i(0) \quad (7.10e)$$

$$I_i(0) = \frac{\Gamma_I}{\lambda_I} P_i(t) \quad (7.10f)$$

Bien que non disponibles dans les données du transitoire, les concentrations en xénon et les puissances neutroniques en chaque maille sont à l'équilibre (cf annexe C), on peut donc en

déduire l'iode à l'équilibre, correspondant à la concentration initiale en iode. La solution obtenue correspond à la paramétrisation début de vie.

Les résultats issus de la calibration du modèle sur la base des données de SOFIA sont rassemblés en (tableau 7.5).

Paramètres	$K_{\text{xénon}}$	$K_{\text{mod}}$	$K_{\text{bore}}$	$\Gamma_I$	$\Sigma_X$	$\Gamma_X$
Paramètre DDV	-2484	1	-6.8	$2.9 \times 10^{-7}$	$5.5 \times 10^{-7}$	$5.2 \times 10^{-8}$
Valeur corrigée	-2327	1.28	-6.1	$2.9 \times 10^{-7}$	$5.3 \times 10^{-7}$	$6.8 \times 10^{-8}$

TABLE 7.5 – Calibration des paramètres du modèle.

## 7.4.2 Estimation d'état

Nous présentons ici les résultats d'estimation de l'état du cœur. Le modèle est paramétré par les valeurs qui ont été déterminées en sous-section 7.4.1. On choisit un horizon  $H = 1h$  et une discrétisation de ce dernier de 2min. Le coût d'arrivée est pondéré par  $S = 30P$ . Aucun bruit de mesure n'est considéré.

L'estimation *a priori* à l'instant  $t = 0$  des variables d'état mesurées est initialisée avec les mesures. En l'absence d'autres informations sur les états non mesurés, ils sont initialisés à leur valeur d'équilibre, donnée par la solution de l'équation  $0 = f(\bar{x}', u)$ , d'inconnue  $\bar{x}'$  et où  $u$  est issu des données de simulation. Le modèle représente le système de manière imparfaite, ce qui implique que l'état obtenu par la résolution des équations du modèle à l'équilibre ne correspond pas à l'état réel. L'erreur initiale en xénon est ainsi  $\varepsilon_X(0) = 5\%$ .

La stratégie de réduction de la latence de calcul présentée en (sous-section 7.2.4) est illustrée dans ce scénario de validation, on présente ainsi deux expériences d'estimations avec l'algorithme MHE utilisé jusqu'ici et la version modifiée asMHE, résumée en (tableau 7.6). On observe que les résultats des deux estimateurs sont très proches. L'asMHE n'introduit pas d'erreur d'estimation supplémentaire par rapport à la MHE non modifiée. Le temps de *feedback* indique la latence pour chaque variante de l'estimateur. Après l'acquisition de nouvelles données, le temps nécessaire pour déterminer la solution du problème d'optimisation est de l'ordre de la dizaine de millisecondes, soit plusieurs ordres de grandeurs inférieur au temps de calcul nécessaire à la résolution du problème dans la MHE non modifiée. Cette réduction de la latence de calcul est permise au prix d'une étape de préparation coûteuse, puisqu'elle nécessite 30s, soit la moitié du temps de calcul maximal autorisé.

Dans les deux cas, on observe une convergence  $\varepsilon_p$  et  $\varepsilon_T$  vers un niveau d'erreur inférieur au pourcent. Ce résultat n'est pas surprenant puisque nous avons déterminé que les températures entre les mailles sont facilement obtenues par la connaissance de la puissance neutronique en chaque maille, qui est mesurée et non bruitée. Ce résultat valide le modèle de température en enthalpie. Le xénon est estimé avec une valeur finale de  $\varepsilon_X = 2.64\%$ . L'erreur réalisée sur le xénon au cours du transitoire est d'environ 10% pendant la première partie du transitoire. Dans un second temps, les nouvelles informations obtenues lors de la remontée en puissance (entre 3,5 et 4,5h) permettent de passer sous la barre des 10% d'erreur.



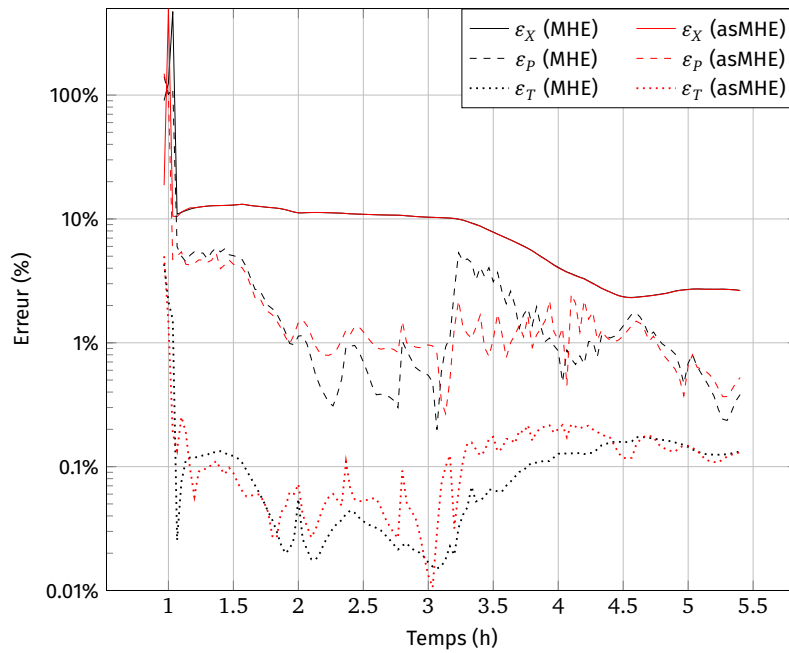


FIGURE 7.9 – Résultats d'estimation de l'état avec un horizon d'une heure pour MHE et asMHE.

Variante de l'estimateur	MHE	asMHE
Temps de calcul par résolution	~1s	~1s
Temps de <i>feedback</i>	~1s	~10ms
Temps de calcul sensibilité de la solution	-	~30s
$\varepsilon_X$ finale	2.64%	2.63%
$\varepsilon_P$ finale	0.38%	0.52%
$\varepsilon_T$ finale	0.13%	0.13%

TABLE 7.6 – Évaluation des performances de l'estimation de l'état.

## 7.5 Conclusion et synthèse des résultats

Dans ce chapitre, nous avons présenté les détails d'implémentation d'un estimateur à horizon glissant pour les réacteurs nucléaires à eau pressurisée. L'estimateur comporte un grand nombre de paramètres de réglage dont l'impact sur les performances d'estimation est présenté. En particulier, le compromis entre la vitesse de convergence et le filtrage des bruits de mesures peut-être réglé par le choix de la matrice de pondération dans le terme de coût d'arrivée, ainsi que par le choix de l'horizon d'estimation.

## 7.6 Notes et discussion

La résolution du problème d'optimisation transcrit par tir multiple s'est avérée infructueuse. La raideur du système rend l'intégration numérique très sensible aux conditions initiales pour l'intégration sur chaque sous-intervalle de l'horizon. La conséquence directe est que le solveur d'intégration peine à trouver une solution et à intégrer le système, malgré des tentatives de réglage de ses paramètres (les solveurs IDAS et CVODES de la suite SUNDIALS [Hin+05] ont été testés) ce qui dans le pire cas aboutit à une erreur dans le programme d'intégration et une interruption de ce dernier. La collocation directe, qui se passe d'un inté-

grateur, n'est pas affectée par ces problèmes et la difficulté d'intégration est reportée sur le solveur d'optimisation (contraintes de collocation). Néanmoins, le solveur d'optimisation est capable de considérer un échec de résolution du problème sans interrompre le programme. Les premières itérations de l'algorithme MHE aboutissent généralement à une erreur, telle que la non faisabilité du problème, en raison de la piètre initialisation de la solution. À mesure que de nouvelles données sont traitées et que la solution précédente est réutilisée pour initialiser la solution du problème courant, le solveur est capable d'échapper à la non faisabilité du problème et déterminer une solution. Une fois les premières itérations passées, il n'est jamais confronté à un problème infaisable.

Les résultats de validation SOFIA que nous présentons sont limités par la faible quantité de données de validation dont nous disposons. En effet, le transitoire que nous utilisons pour mener la validation ne contient que  $\sim 5.5$ h de données, ce qui limite la taille de l'horizon d'estimation que nous pouvons employer. Néanmoins, les résultats que nous présentons sont encourageants en dépit de la « courte » fenêtre d'estimation que nous considérons. Davantage de validations sur des données issues de simulations fines pourraient permettre d'affiner la qualité des résultats et apporter une réponse plus satisfaisante pour la longueur de l'horizon à utiliser. Cependant, il faut considérer que ces simulations fines sont difficiles à obtenir : il faut d'une part composer avec les créneaux de disponibilité du simulateur, mais ces simulations sont également coûteuses en budget et en temps. La vitesse de fonctionnement du simulateur étant au mieux de deux fois le temps réel, la réalisation de transitoires longs demande un temps considérable. Par ailleurs, SOFIA est un outil difficile à manipuler et à paramétrer compte tenu de la grande marge de manœuvre dont dispose l'utilisateur. Enfin, la génération et le traitement des données de simulation demandent un temps et une expertise du simulateur non négligeables pour sélectionner et instrumenter les signaux d'intérêt.



# CONCLUSION GÉNÉRALE

## Synthèse des travaux réalisés

Cette thèse a porté sur l'élaboration d'un estimateur à horizon glissant pour l'estimation de l'état et des paramètres d'un modèle de réacteur nucléaire, destiné en premier lieu à la commande prédictive d'un système d'aide au pilotage du réacteur. Ce modèle, conçu spécifiquement pour la commande, présente une complexité juste suffisante pour le contrôle du cœur du réacteur. Un schéma d'estimation *ad hoc* utilisé jusqu'alors exploitait les mesures sur le système pour corriger une partie seulement des variables d'état tandis que les autres évoluaient en boucle ouverte. Cet estimateur supposait également que les signaux de commande appliqués au système étaient exactement ceux préconisés aux opérateurs par l'outil d'aide au pilotage. Cette hypothèse n'est pas réaliste. La nouvelle solution d'estimation proposée dans ce mémoire estime l'état initial de la MPC avec davantage de confiance et de précision que l'estimateur *ad hoc*. En outre, la dynamique du réacteur nucléaire change à mesure que le combustible est consommé. Le modèle doit être ajusté pour garantir la fidélité des prédictions qu'il produit. L'estimation en ligne de paramètres d'intérêt du modèle permet de réduire les écarts de comportement entre le modèle et le système.

La solution proposée s'appuie sur le concept d'estimation à horizon glissant. Sa conception vise deux objectifs : 1) estimation de l'état et 2) estimation des paramètres d'intérêt. À cette fin, un modèle de l'évolution thermique du caloporteur basé sur les enthalpies massiques a été développé et calibré grâce à des données de simulation réalistes obtenues à l'aide d'un simulateur à haute résolution, largement utilisé pour des études de sûreté pour l'industrie nucléaire. Ce modèle remplace le modèle de l'évolution thermique mis en œuvre précédemment et permet une comparaison plus fiable avec les mesures.

Par ailleurs, les propriétés particulières du modèle (raideur et non-linéarité), en lien avec les phénomènes physiques en action au sein du réacteur, conduisent, en première instance, à une impasse avec les méthodes standards telles que l'EKF. Cependant, nous avons réussi à surmonter ce verrou grâce à une implémentation méthodologique et algorithmique sur mesure de la MHE. Ce succès s'est appuyé sur un travail bibliographique approfondi et sur la mise en relation des choix méthodologiques avec les difficultés spécifiques rencontrées.

La solution proposée a eu recours à une transcription par collocation directe du problème d'estimation, exprimé à temps continu, et pour un horizon d'estimation long. Dans le même temps, nous avons adapté et introduit la méthodologie advanced-step MHE, ce afin de réduire l'impact de la latence entre l'instant d'accès aux données sur le système et l'instant de fin de résolution du problème d'optimisation (précision accrue des estimations courantes).

Les performances de l'estimateur sont validées sur des données issues du simulateur SOFIA, supportant les études approfondies à Framatome. Le premier objectif, à savoir l'estimation d'état, est rempli par exploitation des seules entrées et sorties mesurées, en nombre limité sur le type de réacteur considéré. Le second objectif, a consisté à estimer les paramètres d'intérêt. Il a été atteint en considérant que les mesures du système sont complétées de celles disponibles sur le système expert des sites industriels de production

d'énergie nucléaire ; ce dernier fait en quelque sorte office de jumeau numérique. Les mesures « virtuelles », produites par ce système expert, sont ajoutées aux mesures réelles pour permettre l'estimation des paramètres d'intérêt.

## Perspectives

Au final, la solution proposée dans cette thèse ouvre la voie à de nouvelles pistes de développement. L'estimateur permettra la mise en œuvre de lois de commandes autocalibrées à retour de sortie, via un état reconstruit par MHE. Il pourra ainsi compléter l'outil de pilotage OAPS en cours de finalisation à Framatome. Par association avec une loi de commande MPC en l'occurrence.

En dépit de ses limites, le filtre de Kalman et ses extensions présenterait, s'il était capable de fournir l'estimation requise (ce qui n'est pas le cas ici), l'avantage de quantifier les incertitudes relatives aux estimations produites (approche Bayésienne, covariance de l'erreur d'estimation en lien avec l'équation de Riccati sous-jacente). Si la MHE ne produit pas aussi immédiatement de telles incertitudes, des perspectives existent à ce sujet que nous proposons d'étudier dans le futur. Cela participerait à l'évaluation de la qualité des estimations produites via le terme de coût d'arrivée (en lien avec l'équation de Riccati) et ouvrirait des perspectives en terme de lois de commandes robustes en ce sens qu'elles pourraient prendre en compte les incertitudes sur l'état courant. En intervenant dans le coût d'arrivée (en tant que pondération), les incertitudes pourraient faciliter le réglage automatique de l'estimateur qui pour l'instant demande un réglage *ad-hoc* par essai/erreur.

L'algorithme proposé se base sur une implémentation sur-mesure par collocation directe et emploie un solveur d'optimisation générique basé sur la méthode des points intérieurs (IPOPT). La robustesse de ce solveur permet de trouver une solution au problème d'optimisation en un temps raisonnable vis-à-vis des besoins de l'application. Néanmoins, l'usage de méthodes de résolution exploitant la structure du problème de contrôle optimal permettrait de réduire le coût de calcul. Cela ouvrirait la voie vers un déploiement sur des cibles matérielles de puissance (et de coût) moindre mais également vers une modélisation plus fine du système et donc des performances d'estimation (et de commande) accrues. La réduction du temps de résolution rend également le cycle de développement et de validation plus court.

## **ANNEXES**



## ANNEXE A : FILTRE DE KALMAN

### A.1 Résultat préliminaire

#### Théorème A.1 (Combinaison linéaire de variables aléatoires normales)

Soient  $X, Y, Z$  des variables aléatoires telles que  $Z = AX + Y$  et telles que  $X$  et  $Y$  soient indépendantes. Si  $X \sim \mathbb{N}(m_x, P_x)$  et  $Y \sim \mathbb{N}(m_y, P_y)$ , alors  $Z \sim \mathbb{N}(Am_x + m_y, AP_xA^T + P_y)$ .

**Preuve :** Par linéarité de l'espérance, il est clair que :

$$E(Z) = E(AX + Y) = AE(X) + E(Y) = Am_x + m_y.$$

En développant :

$$\begin{aligned} \text{cov}(Z) &= E((Z - E(Z))(Z - E(Z))^T) \\ &= E((AX + Y - E(AX + Y))(AX + Y - E(AX + Y))^T) \\ &= E((AX + Y - AE(X) - E(Y))(AX + Y - AE(X) - E(Y))^T) \\ &= E((AX - AE(X))(AX - AE(X))^T) + E((AX - AE(X))(Y - E(Y))^T) \\ &\quad + E((Y - E(Y))(AX - AE(X))^T) + E((Y - E(Y))(Y - E(Y))^T). \end{aligned}$$

Par indépendance de  $X$  et  $Y$ ,  $\text{cov}(X, Y) = 0$  donc les deuxième et troisième termes sont nuls. On a donc enfin :

$$\begin{aligned} \text{cov}(Z) &= E((AX - AE(X))(AX - AE(X))^T) + E((Y - E(Y))(Y - E(Y))^T) \\ &= E(A(X - E(X))(X - E(X))^T A^T) + E((Y - E(Y))(Y - E(Y))^T) \\ &= AE((X - E(X))(X - E(X))^T)A^T + E((Y - E(Y))(Y - E(Y))^T) \\ &= AP_xA^T + P_y. \end{aligned}$$

■

### A.2 Estimateur linéaire non biaisé

Considérons que l'on cherche à obtenir une estimation d'une quantité  $x^*$ . On choisit de la modéliser comme une variable aléatoire suivant une loi normale centrée sur sa valeur réelle. Considérons une estimation *a priori* de sa valeur  $\hat{x}^-$  ainsi qu'une mesure  $y$  telle que :

$$y = cx^* + v \tag{A.1}$$

On cherche à effectuer une estimation *a posteriori* de la valeur de  $x^*$ , en utilisant la connaissance que l'on a *a priori* et la mesure dont on dispose. On propose l'estimateur linéaire suivant :

$$\hat{x} = l\hat{x}^- + ky \tag{A.2}$$

On souhaite que cet estimateur soit non biaisé, c'est à dire que l'espérance de l'erreur converge asymptotiquement vers 0. C'est-à-dire, en notant  $\varepsilon = x - \hat{x}$  l'erreur d'estimation,



$\mathbb{E}(\varepsilon) = 0$ . On calcule :

$$\begin{aligned}
 \varepsilon &= x - \hat{x} \\
 &= x^* - l\hat{x}^- - ky && \text{par (eq. A.2)} \\
 &= x^* - l\hat{x}^- - k(cx^* + v) && \text{par (eq. A.1)} \\
 &= x^* - l\hat{x}^- - k(cx^* + v) + lx^* - lx^* \\
 &= (1 - l - kc)x^* + l\varepsilon - kv \\
 \Leftrightarrow (1 - l)\varepsilon &= (1 - l - kc)x^* - kv
 \end{aligned}$$

Puisque  $E(v) = 0$ ,  $kc + l = 1$  suffit pour avoir  $E(\varepsilon) = 0$ . On a donc finalement la forme :

$$\hat{x} = (1 - kc)\hat{x}^- + ky = \hat{x}^- + k(y - c\hat{x}^-)$$

Il reste un degré de liberté dans le choix de  $k$ , qui n'a pas besoin d'être un gain statique. Son choix est l'objet du filtre de Kalman.

### A.3 Preuve du filtre de Kalman

On sait que la forme obtenue au paragraphe précédent garantit que l'on aura un estimateur non biaisé. Afin de choisir  $k$ , il faut définir un critère de performance associé à un choix de  $k$  donné. On choisira alors le gain optimal au sens de ce critère, ce qui aboutira au « meilleur choix » de gain possible. Le filtre de Kalman est synthétisé comme l'estimateur de variance minimale, c'est-à-dire qu'à chaque instant il minimise le critère suivant :

$$J = E(\varepsilon_k^T \cdot \varepsilon_k) = \text{Tr}(E(\varepsilon_k \cdot \varepsilon_k^T)) \quad \text{avec} \quad \varepsilon_k = x_k - \hat{x}_k$$

$E(\varepsilon_k \cdot \varepsilon_k^T) = P_k$  est également appelé la matrice de covariance de l'erreur d'estimation.

### 1. Prédiction

L'estimation *a priori* est obtenue en appliquant le modèle à l'estimation précédente :

$$\hat{x}_{k+1}^- = A\hat{x}_k + Bu_k$$

Ainsi on a l'erreur d'estimation *a priori* :

$$\begin{aligned}\varepsilon_{k+1}^- &= x_{k+1} - \hat{x}_{k+1}^- \\ &= Ax_k + Bu_k + w_k - A\hat{x}_k - Bu_k \\ &= A\varepsilon_k + w_k\end{aligned}$$

Donc par le théorème A.1 :

$$P_{k+1}^- = AP_k^+ A^T + W$$

### 2. Correction

On utilise le gain de Kalman pour corriger l'estimation avec la mesure :

$$\begin{aligned}\hat{x}_{k+1} &= \hat{x}_{k+1}^- + K_{k+1}(y_{k+1} - C\hat{x}_{k+1}^-) \\ &= \hat{x}_{k+1}^- + K_{k+1}C(x_{k+1} - \hat{x}_{k+1}^-) + K_{k+1}v \\ &= \hat{x}_{k+1}^- + K_{k+1}C\varepsilon_{k+1}^- + K_{k+1}v\end{aligned}$$

Donc en terme de l'erreur *a posteriori* :

$$\begin{aligned}\varepsilon_{k+1} &= x_{k+1} - \hat{x}_{k+1} \\ &= x_{k+1} - \hat{x}_{k+1}^- - K_{k+1}C\varepsilon_{k+1}^- - K_{k+1}v \\ &= \varepsilon_{k+1}^- - K_{k+1}C\varepsilon_{k+1}^- - K_{k+1}v \\ &= (I - K_{k+1}C)\varepsilon_{k+1}^- - K_{k+1}v\end{aligned}$$

Donc par le lemme théorème A.1 :

$$P_{k+1} = (I - K_{k+1}C)P_{k+1}^-(I - K_{k+1}C)^T + K_{k+1}VK_{k+1}^T$$

Il reste maintenant à déterminer  $K_k$ .

### Choix du gain de Kalman

On cherche à minimiser  $J = Tr(P_k)$ ,  $\forall k \in \mathbb{Z}$  en cherchant la valeur optimale de  $K_k$  (c'est en cela que le filtre de Kalman est un estimateur linéaire non biaisé *optimal*). On calcule :

$$\frac{dJ}{dK_k} = -2(CP_k^-)^T + 2K_k(CP_k^- C^T + V)$$

Ainsi,

$$\frac{dJ}{dK_k} = 0 \Leftrightarrow K_k = P_k^- C^T (CP_k^- C^T + V)^{-1}$$



## ANNEXE B : TRANSITOIRE $PWR_{SIMU}$ POUR LA VALIDATION SANS ÉCART MODÈLE/PROCESS

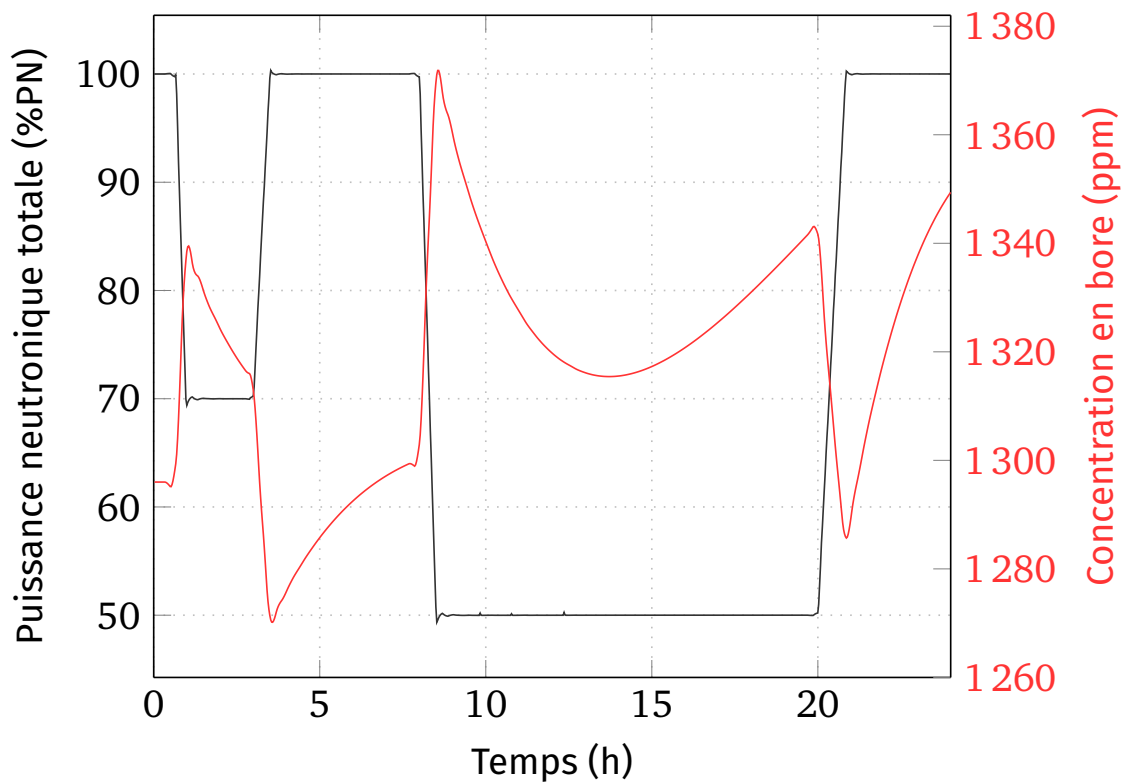


FIGURE B.1 – Variation de la puissance et de la concentration en bore.

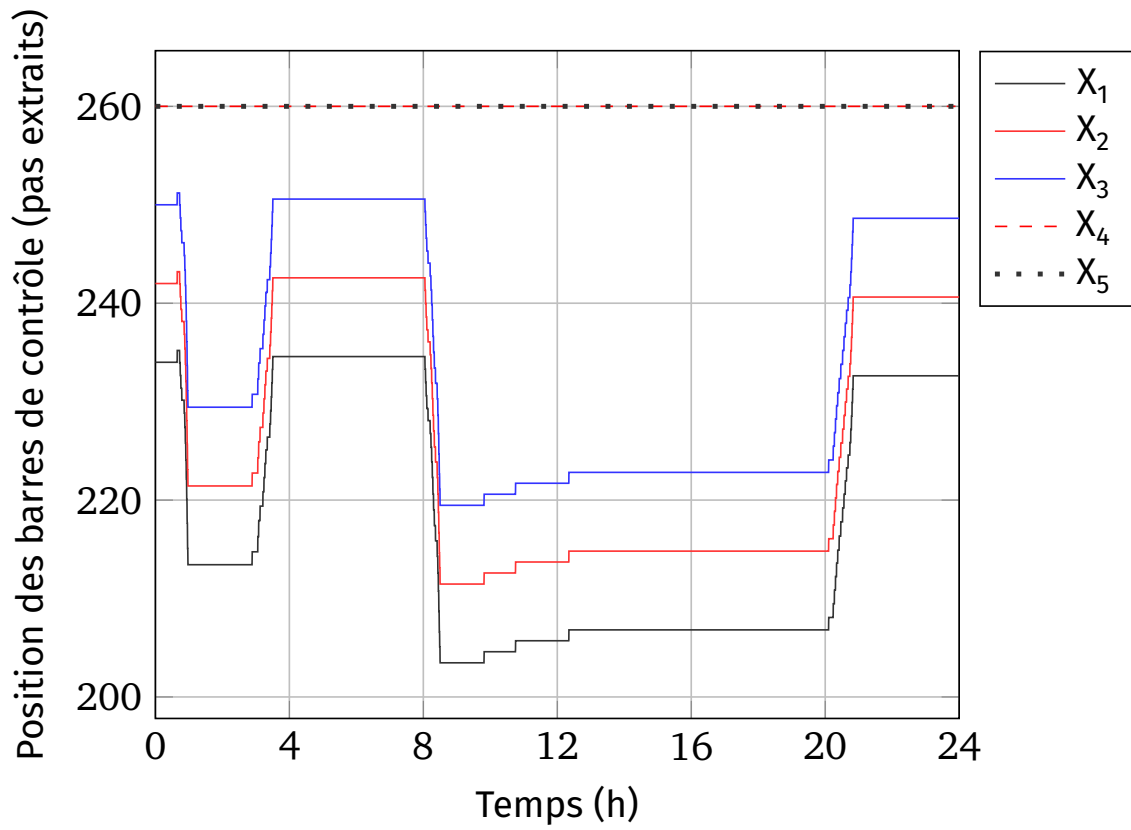


FIGURE B.2 – Position des groupes de barres de contrôle au cours de la variation de charge.

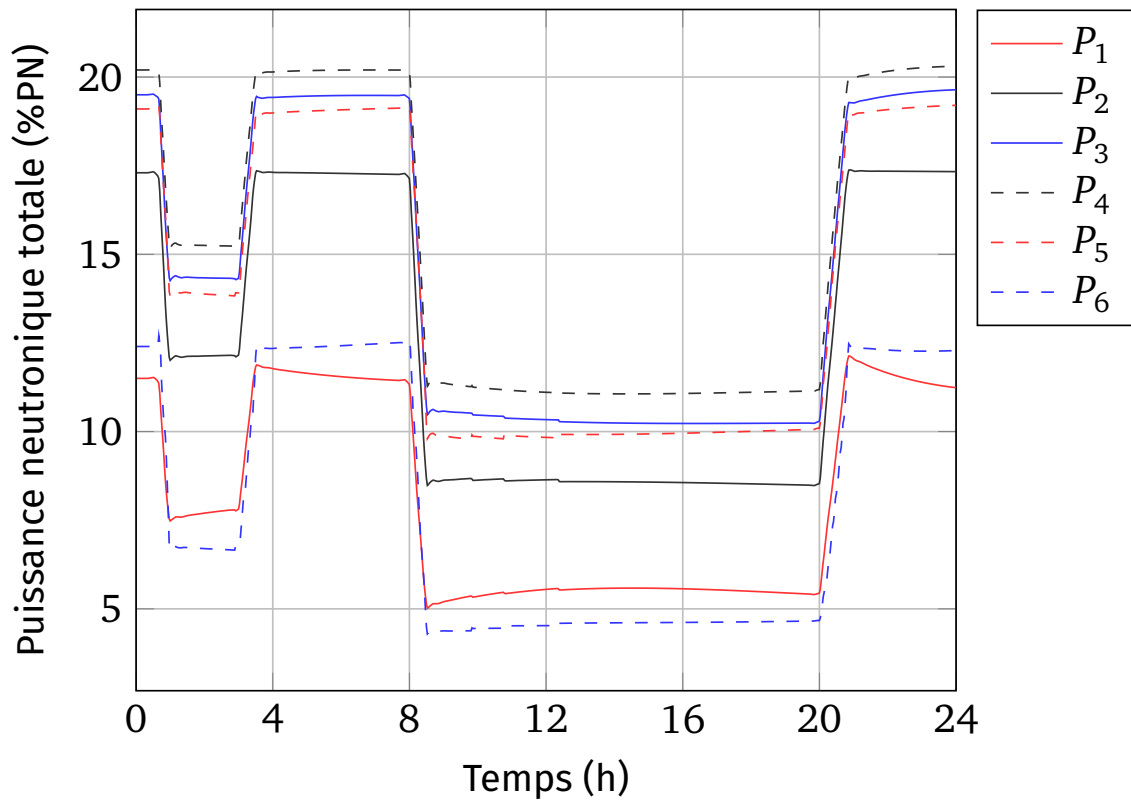


FIGURE B.3 – Distribution de la puissance neutronique.

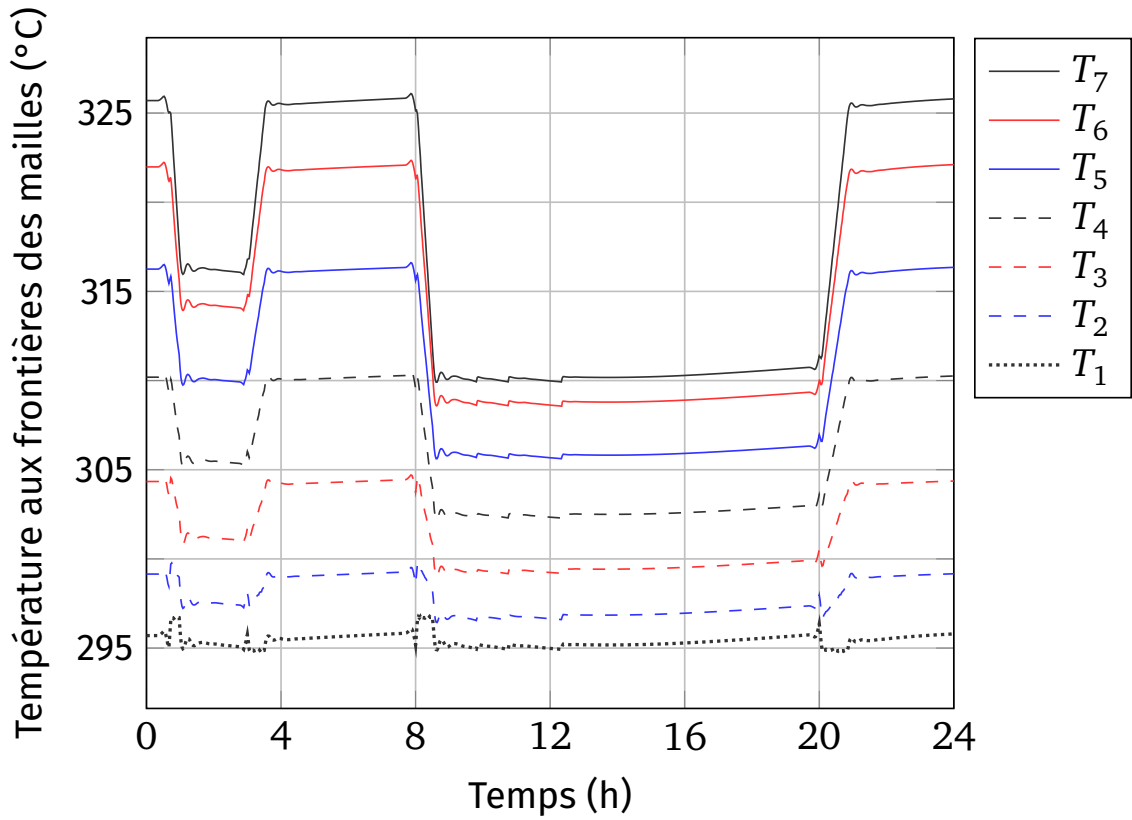


FIGURE B.4 – Températures aux frontières des mailles.

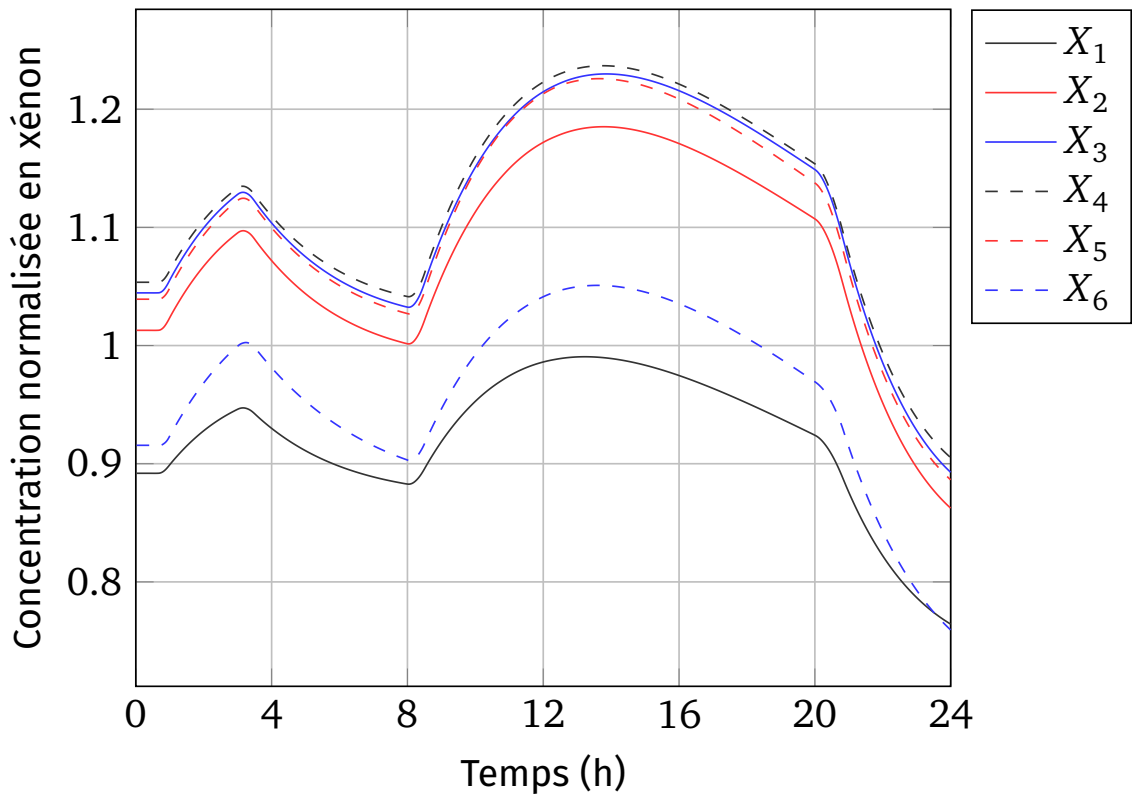


FIGURE B.5 – Évolution de la concentration en xénon (normalisée).

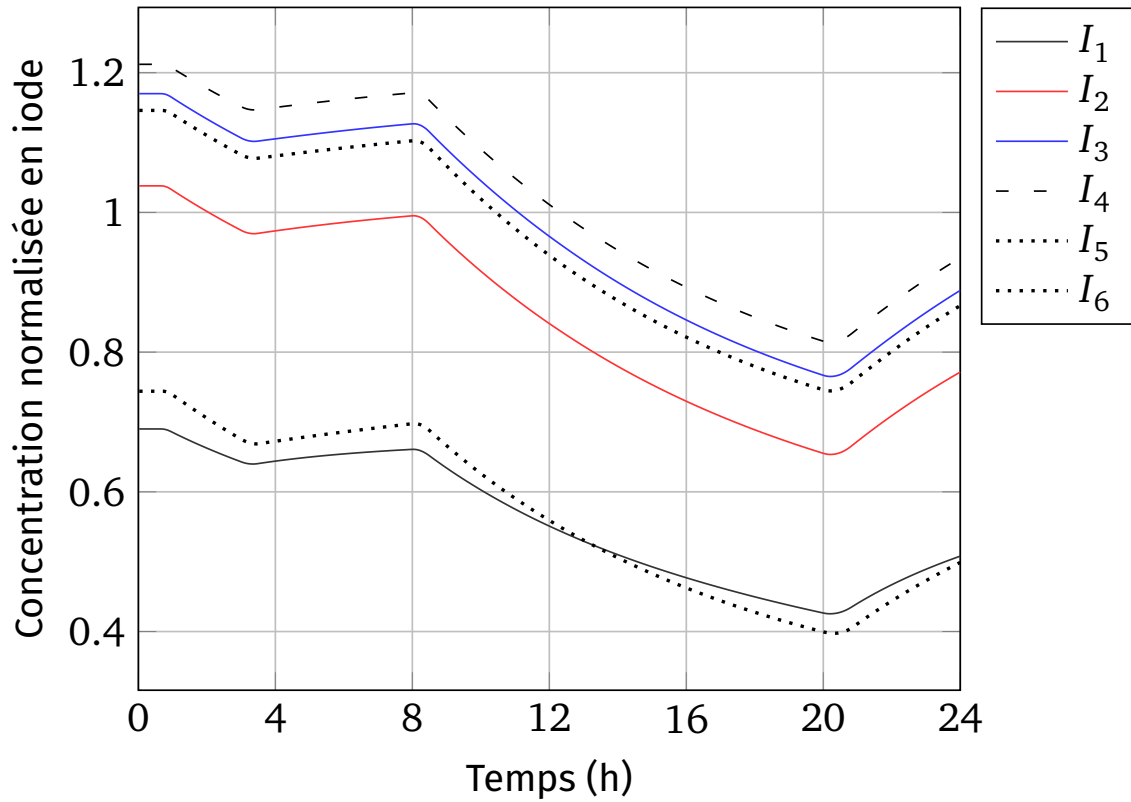


FIGURE B.6 – Évolution de la concentration en iode (normalisée).

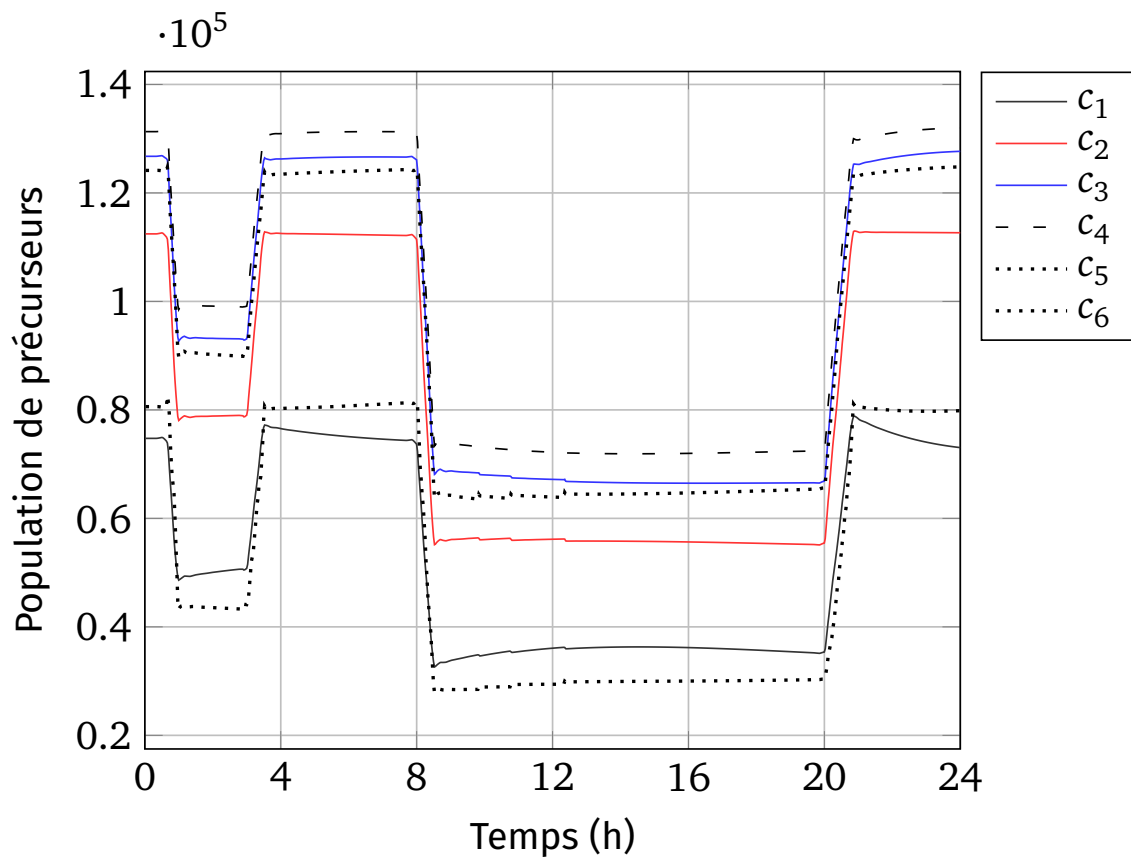


FIGURE B.7 – Évolution de la population des précurseurs.

## ANNEXE C : TRANSITOIRE DE VALIDATION SOFIA

Nous présentons ici le transitoire réalisé sur SOFIA pour la validation de l'estimateur. On considère un réacteur de type N4 piloté en mode A et on se positionne au début du cycle du combustible (faible épuisement) Le transitoire correspond à une baisse de charge en partant initialement d'une puissance de 100%PN. La position des grappes et la concentration en bore dans le circuit primaire sont ajustées afin que la puissance soit maintenue en palier bas à 37% pendant environ 2h avant de remonter à puissance quasi-maximale (98%PN) figure C.4.

Les données de simulation sont projetées sur les 6 mailles du modèle PWR\_simu afin de pouvoir les comparer à l'estimateur. Ainsi, les données du modèle neutronique de SOFIA à 20 mailles (puissance neutronique et concentration en xénon) sont projetées en les sommant par groupe de 3 mailles, de la maille 2 à la maille 19 (cf figure 6.1) puisque les mailles 1 et 20 correspondent au réflecteurs lourds non modélisés dans PWR\_simu. Le modèle de thermique de SOFIA modélise la température moyenne au sein de 36 mailles, tandis que PWR\_simu modélise les températures à l'interface des mailles. La projection sur 6 mailles est effectuée en considérant la température moyenne entre les mailles 6-7, 12-13, ..., 30-31. La température du caloporteur en entrée du réacteur est considérée égale à la température en entrée du cœur (puisque'il ne s'échauffe qu'une fois entré dans le cœur et on suppose qu'il ne refroidit pas avant d'entrer dans le cœur). La température de l'eau en sortie du cœur est prise comme la température de l'eau dans la maille 36. Bien que l'eau s'échauffe effectivement dans cette maille, l'erreur réalisée est minime en considérant que d'une part, le flux neutronique aux bords du cœur est faible et que d'autre part, l'élévation de la température de l'eau par unité de chaleur reçue est inversement corrélée à la température initiale.



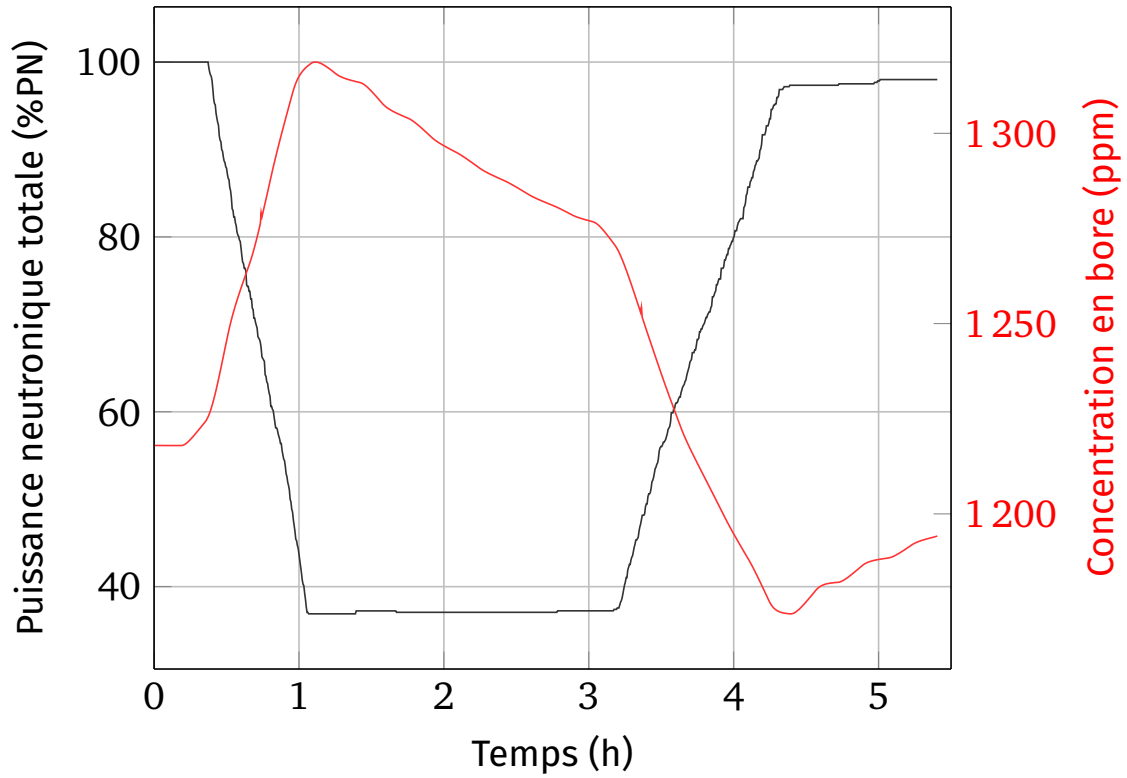


FIGURE C.1 – Baisse de charge et signaux de commande.

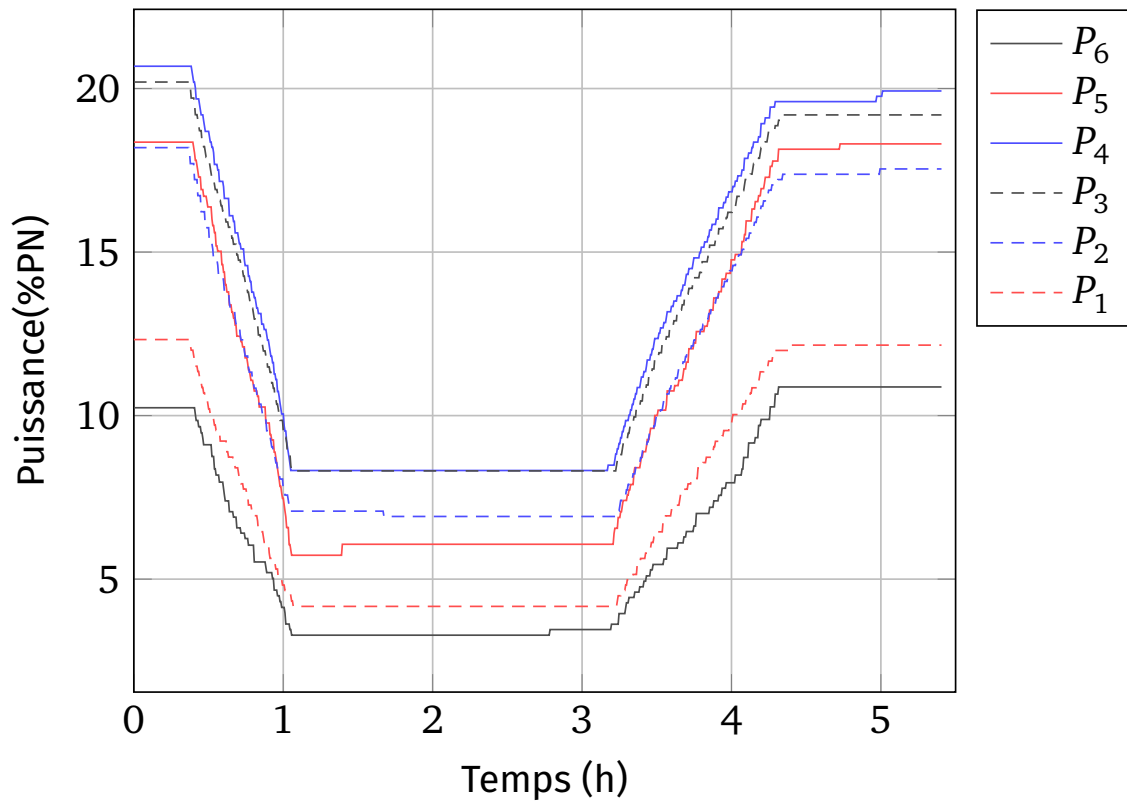


FIGURE C.2 – Distribution de la puissance neutronique.

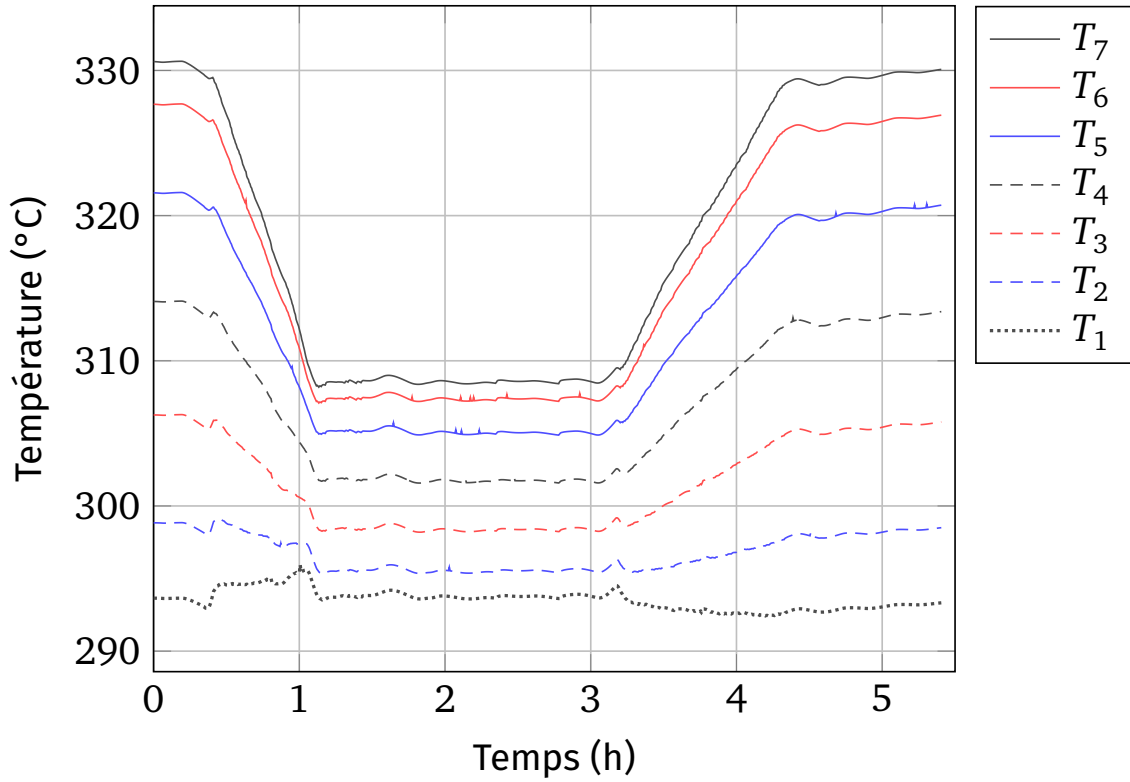


FIGURE C.3 – Températures aux frontières des mailles.

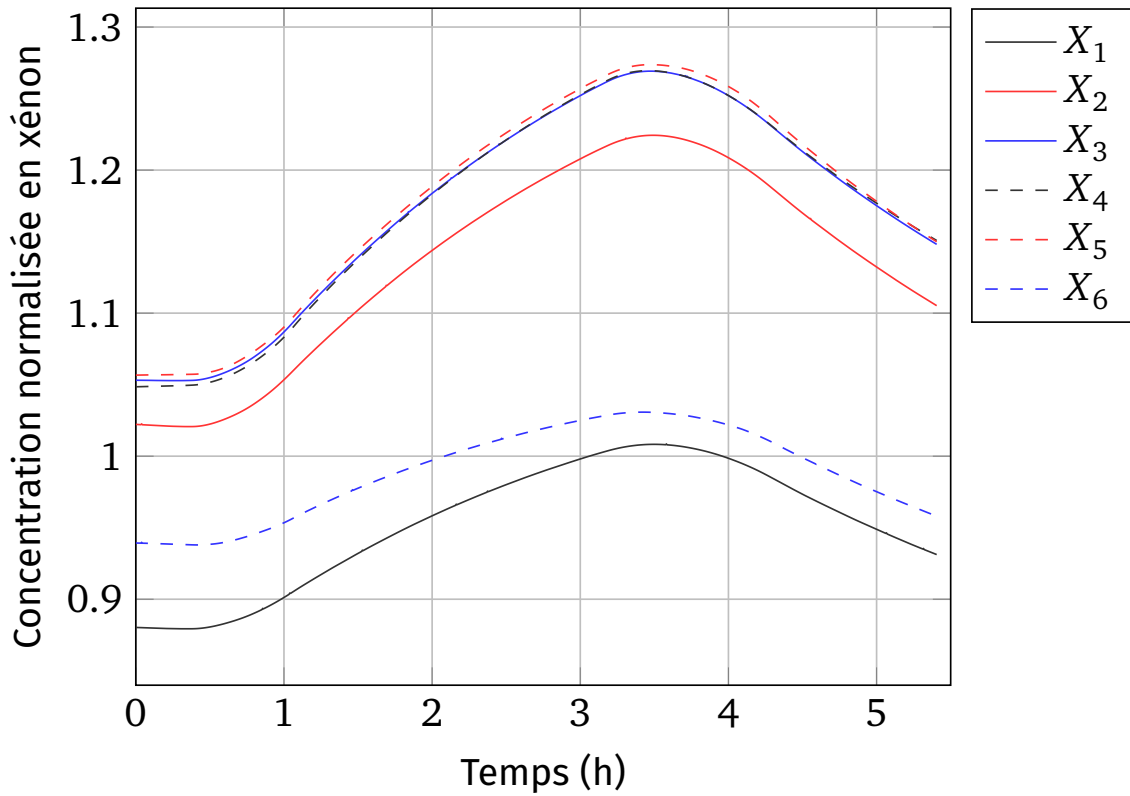


FIGURE C.4 – Évolution de la concentration xénon (normalisée).



## BIBLIOGRAPHIE

- [Ala] Mazen ALAMIR. *Nonlinear Moving Horizon Observers: Theory & Real-Time Implementation*.
- [Ala06] D. ALAZARD. « Introduction Au Filtre de Kalman ». In : 1<sup>er</sup> oct. 2006.
- [Alb11] Jan ALBERSMEYER. « Adjoint-Based Algorithms and Numerical Methods for Sensitivity Generation and Optimization of Large Scale Dynamic Systems ». Heidelberg University Library, 2011.
- [AD10] Jan ALBERSMEYER et Moritz DIEHL. « The Lifted Newton Method and Its Application in Optimization ». In : *SIAM Journal on Optimization* 20.3 (jan. 2010), p. 1655-1684.
- [Ale+10] Angelo ALESSANDRI et al. « Advances in Moving Horizon Estimation for Nonlinear Systems ». In : *49th IEEE Conference on Decision and Control (CDC)*. 49th IEEE Conference on Decision and Control (CDC). Déc. 2010, p. 5681-5688.
- [AK21] Brandon AMOS et J. Zico KOLTER. *OptNet : Differentiable Optimization as a Layer in Neural Networks*. 2 déc. 2021. URL : <http://arxiv.org/abs/1703.00443> (visité le 19/06/2024). Prépubl.
- [Amo+19] Brandon AMOS et al. *Differentiable MPC for End-to-end Planning and Control*. 14 oct. 2019. URL : <http://arxiv.org/abs/1810.13400> (visité le 19/06/2024). Prépubl.
- [And13] Joel A. E. ANDERSSON. « A General-Purpose Software Framework for Dynamic Optimization (Een Algemene Softwareomgeving Voor Dynamische Optimalisatie) ». Thèse de doct. KU Leuven, 24 oct. 2013.
- [AR18] Joel A. E. ANDERSSON et James B. RAWLINGS. « Sensitivity Analysis for Nonlinear Programming in CasADi ». In : *IFAC-PapersOnLine*. 6th IFAC Conference on Nonlinear Model Predictive Control NMPC 2018 51.20 (1<sup>er</sup> jan. 2018), p. 331-336.
- [And+19] Joel A. E. ANDERSSON et al. « CasADi : A Software Framework for Nonlinear Optimization and Optimal Control ». In : *Mathematical Programming Computation* 11.1 (mars 2019), p. 1-36.
- [ABS13] Vincent ANDRIEU, Gildas BESANÇON et Ulysse SERRES. « Observability Necessary Conditions for the Existence of Observers ». In : *52nd IEEE Conference on Decision and Control*. 52nd IEEE Conference on Decision and Control. Déc. 2013, p. 4442-4447.
- [ABN16] Mark ASCH, Marc BOCQUET et Maëlle NODET. *Data Assimilation. Fundamentals of Algorithms*. Society for Industrial and Applied Mathematics, 19 déc. 2016. 310 p.
- [Axe15] Daniel AXEHILL. « Controlling the Level of Sparsity in MPC ». In : *Systems & Control Letters* 76 (1<sup>er</sup> fév. 2015), p. 1-7.
- [AM12] Daniel AXEHILL et Manfred MORARI. « An Alternative Use of the Riccati Recursion for Efficient Optimization ». In : *Systems & Control Letters* 61.1 (1<sup>er</sup> jan. 2012), p. 37-40.
- [Bar16] Sören BARTELS. « Finite Element Method ». In : *Numerical Approximation of Partial Differential Equations*. Sous la dir. de Sören BARTELS. Cham : Springer International Publishing, 2016, p. 99-152.

- [Ber19] Pauline BERNARD. *Observer Design for Nonlinear Systems*. T. 479. Lecture Notes in Control and Information Sciences. Cham : Springer International Publishing, 2019.
- [BZ09] L.T. BIEGLER et V.M. ZAVALA. « Large-Scale Nonlinear Programming Using IPOPT : An Integrating Framework for Enterprise-Wide Dynamic Optimization ». In : *Computers & Chemical Engineering* 33.3 (mars 2009), p. 575-582.
- [Boc81] H. G. BOCK. « Numerical Treatment of Inverse Problems in Chemical Reaction Kinetics ». In : *Modelling of Chemical Reaction Systems*. Sous la dir. de Klaus H. EBERT, Peter DEUFLHARD et Willi JÄGER. Réd. par Vitalii I. GOLDANSKII et al. T. 18. Berlin, Heidelberg : Springer Berlin Heidelberg, 1981, p. 102-125.
- [BP84] H.G. BOCK et K.J. PLITT. « A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems \* ». In : *IFAC Proceedings Volumes* 17.2 (juill. 1984), p. 1603-1608.
- [Boy+94] Stephen BOYD et al. *Linear Matrix Inequalities in System and Control Theory*. Society for Industrial and Applied Mathematics, jan. 1994.
- [BV04] Stephen P. BOYD et Lieven VANDENBERGHE. *Convex Optimization*. Cambridge, UK; New York : Cambridge University Press, 2004. 716 p.
- [Bro+18] D. A. BROWN et al. « ENDF/B-VIII.0 : The 8th Major Release of the Nuclear Reaction Data Library with CIELO-project Cross Sections, New Standards and Thermal Scattering Data ». In : *Nuclear Data Sheets*. Special Issue on Nuclear Reaction Data 148 (1<sup>er</sup> fév. 2018), p. 1-142.
- [BK17] Steve BRUNTON et Nathan KUTZ. *Data Driven Science and Engineering*. 2017.
- [Cal+24] Louis CALLENS et al. « AdaptiveNLP : A Framework for Efficient Online Adaptability in NLP Structures for Optimal Control Problems ». In : *2024 European Control Conference (ECC)*. 2024 European Control Conference (ECC). Juin 2024, p. 1358-1365.
- [CEA15] CEA. *La neutronique*. CEA/Médiathèque. 20 mars 2015. URL : <https://www.cea.fr/multimedia/Pages/editions/ouvrages/monographies-nucleaire/neutronique.aspx> (visité le 27/09/2023).
- [CK63] F. CESCHINO et J. KUNTZMANN. *Problèmes Différentiels de Conditions Initiales, Par F. Ceshino [et] J. Kuntzmann*. Méthodes Numériques. 1963.
- [Cha+23] P. CHAMBON et al. « Global Observing System Experiments within the Météo-France 4D-Var Data Assimilation System ». In : *Monthly Weather Review* 151.1 (6 jan. 2023), p. 127-143.
- [Dem06] Jean-Pierre DEMAILLY. *Analyse numérique et équations différentielles*. Nouv. ed. Collection Grenoble Sciences. Les Ulis, France : EDP Sciences, 2006. 343 p.
- [Die01] Moritz DIEHL. « Real-Time Optimization for Large Scale Nonlinear Processes ». Thèse de doct. 2001.
- [DFH09] Moritz DIEHL, Hans Joachim FERREAU et Niels HAVERBEKE. « Efficient Numerical Methods for Nonlinear MPC and Moving Horizon Estimation ». In : *Nonlinear Model Predictive Control*. Sous la dir. de Lalo MAGNI, Davide Martino RAIMONDO et Frank ALLGÖWER. Réd. par Manfred MORARI et Manfred THOMA. T. 384. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009, p. 391-417.
- [DSD11] Q. Tran DINH, C. SAVORGNAN et M. DIEHL. *Adjoint-Based Predictor-Corrector Sequential Convex Programming for Parametric Nonlinear Optimization*. 13 sept. 2011. URL : <http://arxiv.org/abs/1109.2800> (visité le 24/01/2023). Pré-publ.

- [DGA00] Arnaud DOUCET, Simon GODSILL et Christophe ANDRIEU. « On Sequential Monte Carlo Sampling Methods for Bayesian Filtering ». In : *Statistics and Computing* 10.3 (1<sup>er</sup> juill. 2000), p. 197-208.
- [Dup23] Guillaume DUPRÉ. « Conception de Systèmes de Contrôle Avancé de Réacteur PWR Flexible : Vers Une Solution Industrielle ». These de doctorat. Ecole nationale supérieure Mines-Télécom Atlantique Bretagne Pays de la Loire, 6 déc. 2023.
- [Dup+21] Guillaume DUPRÉ et al. « Enhanced Flexibility of PWRs (Mode A) Using an Efficient NMPC-Based Boration/Dilution System ». In : *2021 European Control Conference (ECC)*. 2021 European Control Conference (ECC). Juin 2021, p. 1092-1098.
- [ED18] Michael T. M. EMMERICH et André H. DEUTZ. « A Tutorial on Multiobjective Optimization : Fundamentals and Evolutionary Methods ». In : *Natural Computing* 17.3 (sept. 2018), p. 585-609.
- [Enr+86] W. H. ENRIGHT et al. « Interpolants for Runge-Kutta Formulas ». In : *ACM Transactions on Mathematical Software* 12.3 (sept. 1986), p. 193-218.
- [EGH] Robert EYMARD, Thierry GALLOUËT et Raphaële HERBIN. « Finite Volume Methods ». In : ().
- [FM90] Anthony V. FIACCO et Garth P. McCORMICK. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques*. Society for Industrial and Applied Mathematics, jan. 1990.
- [Fie+20] Felix FIEDLER et al. « A Probabilistic Moving Horizon Estimation Framework Applied to the Visual-Inertial Sensor Fusion Problem ». In : *2020 European Control Conference (ECC)*. 2020 European Control Conference (ECC). Mai 2020, p. 1009-1016.
- [For88] Bengt FORNBERG. « Generation of Finite Difference Formulas on Arbitrarily Spaced Grids ». In : *Mathematics of Computation* 51.184 (1988), p. 699-706.
- [Fri+17] Gianluca FRISON et al. « BLASFEO : Basic Linear Algebra Subroutines for Embedded Optimization ». In : *ACM Transactions on Mathematical Software (TOMS)* (2017).
- [Fux+14] Samuel F. FUX et al. « EKF Based Self-Adaptive Thermal Model for a Passive House ». In : *Energy and Buildings* 68 (jan. 2014), p. 811-817.
- [Gam19] R. V. GAMKRELIDZE. « History of the Discovery of the Pontryagin Maximum Principle ». In : *Proceedings of the Steklov Institute of Mathematics* 304.1 (jan. 2019), p. 1-7.
- [Gar+22] David J GARDNER et al. « Enabling New Flexibility in the SUNDIALS Suite of Nonlinear and Differential/Algebraic Equation Solvers ». In : *ACM Transactions on Mathematical Software (TOMS)* 48.3 (2022), p. 1-24.
- [Gas24] Baptiste GASSE. « Optimisation Du Pilotage Des Réacteurs Nucléaires à Eau Sous Pression (REP) Dans Un Réseau Électrique Dans Le Contexte de La Transition Énergétique ». These de doctorat. université Paris-Saclay, 14 fév. 2024.
- [GMS05] Philip E. GILL, Walter MURRAY et Michael A. SAUNDERS. « SNOPT : An SQP Algorithm for Large-Scale Constrained Optimization ». In : *SIAM Review* 47.1 (jan. 2005), p. 99-131.
- [Gil+20] Joris GILLIS et al. « Effortless Modeling of Optimal Control Problems with Rocket ». In : *39th Benelux Meeting on Systems and Control*. T. 138. Elspeet, The Netherlands, 2020.

- [GW08] Andreas GRIEWANK et Andrea WALTHER. *Evaluating Derivatives : Principles and Techniques of Algorithmic Differentiation*. 2nd ed. Philadelphia, PA : Society for Industrial and Applied Mathematics, 2008. 438 p.
- [Gro07] Alain GROSSETETE. « Le Pilotage de l'EPR : Mode T ». In : *Revue Générale Nucléaire* 3 (mai 2007), p. 37-41.
- [Gru+23] Lucas GRUSS et al. « Moving Horizon Estimation of Xenon in Pressurized Water Nuclear Reactors Using Variable-Step Integration ». In : *ECC 2023 : European Control Conference*. Bucharest, Romania : euca, juin 2023.
- [Gru+24] Lucas GRUSS et al. « Implementation and Validation of Simultaneous State and Parameter Moving Horizon Estimation of a Pressurized Water Reactor ». In : *ECC 2024 : European Control Conference*. Stockholm, Sweden : euca, juin 2024.
- [HNW09] E. HAIRER, S. P. NØRSETT et Gerhard WANNER. *Solving Ordinary Differential Equations I : Nonstiff Problems*. 2nd rev. ed. Springer Series in Computational Mathematics 8. Heidelberg ; London : Springer, 2009. 528 p.
- [HW96] Ernst HAIRER et Gerhard WANNER. *Solving Ordinary Differential Equations II*. T. 14. Springer Series in Computational Mathematics. Berlin, Heidelberg : Springer Berlin Heidelberg, 1996.
- [HI14] Ralf HANNEMANN-TAMÁS et Lars S. IMSLAND. « Full Algorithmic Differentiation of a Rosenbrock-type Method for Direct Single Shooting ». In : *2014 European Control Conference (ECC)*. 2014 European Control Conference (ECC). Juin 2014, p. 1242-1248.
- [HR05] Eric L. HASELTINE et James B. RAWLINGS. « Critical Evaluation of Extended Kalman Filtering and Moving-Horizon Estimation ». In : *Industrial & Engineering Chemistry Research* 44.8 (1<sup>er</sup> avr. 2005), p. 2451-2460.
- [HK77] R. HERMANN et A. KRENER. « Nonlinear Controllability and Observability ». In : *IEEE Transactions on Automatic Control* 22.5 (oct. 1977), p. 728-740.
- [Hin+05] Alan C. HINDMARSH et al. « SUNDIALS : Suite of Nonlinear and Differential/Algebraic Equation Solvers ». In : *ACM Trans. Math. Softw.* 31.3 (1<sup>er</sup> sept. 2005), p. 363-396.
- [Hon+84] Takashi HONDA et al. « Radioactive Contamination of Carbon Steel in a Boiling Water Reactor ». In : *Nuclear Technology* 65.3 (1<sup>er</sup> juin 1984), p. 438-443.
- [Hu17] Wuhua HU. « Robust Stability of Optimization-based State Estimation ». 26 sept. 2017.
- [HR07] Geoffrey T. HUNTINGTON et Anil V. RAO. « A Comparison between Global and Local Orthogonal Collocation Methods for Solving Optimal Control Problems ». In : *2007 American Control Conference*. 2007 American Control Conference. Juill. 2007, p. 1950-1957.
- [IX00] K. ITO et K. XIONG. « Gaussian Filters for Nonlinear Filtering Problems ». In : *IEEE Transactions on Automatic Control* 45.5 (mai 2000), p. 910-927.
- [Jam+13] Gareth JAMES et al. *An Introduction to Statistical Learning*. T. 103. Springer Texts in Statistics. New York, NY : Springer, 2013.
- [Jaz68] A. JAZWINSKI. « Limited Memory Optimal Filtering ». In : *IEEE Transactions on Automatic Control* 13.5 (oct. 1968), p. 558-563.
- [JXC12] Bin JIA, Ming XIN et Yang CHENG. « Sparse-Grid Quadrature Nonlinear Filtering ». In : *Automatica* 48.2 (1<sup>er</sup> fév. 2012), p. 327-341.

- [Jol39] Frédéric JOLIOT. « Sur la rupture explosive des noyaux U et Th sous l'action des neutrons ». In : *Journal de Physique et le Radium* 10.3 (1<sup>er</sup> mars 1939), p. 159-160.
- [JUD95] S.J. JULIER, J.K. UHLMANN et H.F. DURRANT-WHYTE. « A New Approach for Filtering Nonlinear Systems ». In : *Proceedings of 1995 American Control Conference - ACC'95*. Proceedings of 1995 American Control Conference - ACC'95. T. 3. Juin 1995, 1628-1632 vol.3.
- [Jun06] Marc JUNGERS. « Commande robuste multicritère - une approche par la théorie des jeux ». Thèse de doct. École normale supérieure de Cachan - ENS Cachan, 13 sept. 2006.
- [Kal60] Rudolf KALMAN. « A New Approach To Linear Filtering and Prediction Problems ». In : *Journal of Basic Engineering (ASME)* 82 (1<sup>er</sup> mars 1960), p. 35-45.
- [KC16] Christopher A. KENNEDY et Mark H. CARPENTER. « Diagonally Implicit Runge-Kutta Methods for Ordinary Differential Equations. A Review ». In : (3 jan. 2016).
- [KET73] Shauying R. KOU, David L. ELLIOTT et Tzyh Jong TARN. « Observability of Nonlinear Systems ». In : *Information and Control* 22.1 (1<sup>er</sup> fév. 1973), p. 89-99.
- [KK18] Gennady Yu. KULIKOV et Maria V. KULIKOVA. « Stability Analysis of Extended, Cubature and Unscented Kalman Filters for Estimating Stiff Continuous-Discrete Stochastic Systems ». In : *Automatica* 90 (1<sup>er</sup> avr. 2018), p. 91-97.
- [Lem18] Lori LEMAZURIER. « Conception d'un système avancé de réacteur PWR flexible par les apports conjoints de l'ingénierie système et de l'automatique ». Thèse de doct. Ecole nationale supérieure Mines-Télécom Atlantique, 2 fév. 2018.
- [LHR15] Fengjin LIU, William W. HAGER et Anil V. RAO. « Adaptive Mesh Refinement Method for Optimal Control Using Nonsmoothness Detection and Mesh Size Reduction ». In : *Journal of the Franklin Institute* 352.10 (oct. 2015), p. 4081-4106.
- [LWZ14] Jiang LIU, Yujin WANG et Ju ZHANG. « A Linear Extension of Unscented Kalman Filter to Higher-Order Moment-Matching ». In : *53rd IEEE Conference on Decision and Control*. 53rd IEEE Conference on Decision and Control. Déc. 2014, p. 5021-5026.
- [Lju99] Lennart LJUNG. *System Identification : Theory for the User*. 2nd ed. Prentice Hall Information and System Sciences Series. Upper Saddle River, NJ : Prentice Hall PTR, 1999. 609 p.
- [LB12] Rodrigo LÓPEZ-NEGRETE et Lorenz T. BIEGLER. « A Moving Horizon Estimator for Processes with Multi-Rate Measurements : A Nonlinear Programming Sensitivity Approach ». In : *Journal of Process Control* 22.4 (1<sup>er</sup> avr. 2012), p. 677-688.
- [Lue72] David LUENBERGER. « An Introduction to Observers ». In : *Automatic Control, IEEE Transactions on* 16 (1<sup>er</sup> jan. 1972), p. 596-602.
- [LY08] David G. LUENBERGER et Yinyu YE. *Linear and Nonlinear Programming*. 3rd ed. International Series in Operations Research & Management Science 116. New York, NY : Springer, 2008.
- [MAT21] MATLAB. *Version 9.11.0 (R2021b)*. Natick, Massachusetts : The MathWorks Inc., 2021.



- [MSH38] Lise MEITNER, Fritz STRASSMANN et Otto HAHN. « Künstliche Umwandlungsprozesse bei Bestrahlung des Thoriums mit Neutronen ; Auftreten isomerer Reihen durch Abspaltung von  $\alpha$ -Strahlen ». In : *Zeitschrift für Physik* 109.7 (1<sup>er</sup> juill. 1938), p. 538-552.
- [MU49] Nicholas METROPOLIS et S. ULAM. « The Monte Carlo Method ». In : *Journal of the American Statistical Association* 44.247 (sept. 1949), p. 335-341.
- [MR95] Kenneth R. MUSKE et James B. RAWLINGS. « Nonlinear Moving Horizon State Estimation ». In : *Methods of Model Based Process Control*. Sous la dir. de Ridvan BERBER. Dordrecht : Springer Netherlands, 1995, p. 349-365.
- [Nau08] Uwe NAUMANN. « Optimal Jacobian Accumulation Is NP-complete ». In : *Mathematical Programming* 112.2 (1<sup>er</sup> avr. 2008), p. 427-441.
- [Nej+23] Hossein NEJATBAKHSH ESFAHANI et al. « Learning-Based State Estimation and Control Using MHE and MPC Schemes with Imperfect Models ». In : *European Journal of Control* 73 (1<sup>er</sup> sept. 2023), p. 100880.
- [NW06] Jorge NOCEDAL et Stephen WRIGHT. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2006.
- [NPR00] Magnus NØRGAARD, Niels K. POULSEN et Ole RAVN. « New Developments in State Estimation for Nonlinear Systems ». In : *Automatica* 36.11 (1<sup>er</sup> nov. 2000), p. 1627-1638.
- [Osb69] M. R OSBORNE. « On Shooting Methods for Boundary Value Problems ». In : *Journal of Mathematical Analysis and Applications* 27.2 (1<sup>er</sup> août 1969), p. 417-433.
- [OAT24] Alex OSHIN, Hassan ALMUBARAK et Evangelos A. THEODOROU. *Differentiable Robust Model Predictive Control*. 13 fév. 2024. URL : <http://arxiv.org/abs/2308.08426> (visité le 19/06/2024). Prépubl.
- [Pen55] R. PENROSE. « A Generalized Inverse for Matrices ». In : *Mathematical Proceedings of the Cambridge Philosophical Society* 51.3 (juill. 1955), p. 406-413.
- [PM96] B.-M PFEIFFER et W MARQUARDT. « Symbolic Semi-Discretization of Partial Differential Equation Systems ». In : *Mathematics and Computers in Simulation* 42.4-6 (nov. 1996), p. 617-628.
- [Plo+20] A. J. M. PLOMPEN et al. « The Joint Evaluated Fission and Fusion Nuclear Data Library, JEFF-3.3 ». In : *The European Physical Journal A* 56.7 (14 juill. 2020), p. 181.
- [Pon08] Angélique PONÇOT. « Assimilation de données pour la dynamique du xénon dans les cœurs de centrale nucléaire ». Thèse de doct. Institut National Polytechnique de Toulouse, 3 oct. 2008.
- [Pon+13] Angélique PONÇOT et al. « Variational Assimilation for Xenon Dynamical Forecasts in Neutronic Using Advanced Background Error Covariance Matrix Modelling ». In : *Annals of Nuclear Energy* 60 (1<sup>er</sup> oct. 2013), p. 39-50.
- [PD22] Oleksandr PROKHOROV et Dina DREISBACH. « The Impact of Renewables on the Incidents of Negative Prices in the Energy Spot Markets ». In : *Energy Policy* 167 (1<sup>er</sup> août 2022), p. 113073.
- [PBB23] P. PUCHAUD, F. BAILLY et M. BEGON. « Direct Multiple Shooting and Direct Collocation Perform Similarly in Biomechanical Predictive Simulations ». In : *Computer Methods in Applied Mechanics and Engineering* 414 (1<sup>er</sup> sept. 2023), p. 116162.

- [Qui17] Rien QUIRYNEN. « Numerical Simulation Methods for Embedded Optimization ». KU Leuven and University of Freiburg, 2017.
- [Qui+17] Rien QUIRYNEN et al. « Lifted Collocation Integrators for Direct Optimal Control in ACADO Toolkit ». In : *Mathematical Programming Computation* 9.4 (1<sup>er</sup> déc. 2017), p. 527-571.
- [Rad10] Pierre RADVANYI. « Ernest Rutherford : la transformation radioactive ». In : *Bibnum. Textes fondateurs de la science* (1<sup>er</sup> mai 2010).
- [Rao10] Anil RAO. « A Survey of Numerical Methods for Optimal Control ». In : *Advances in the Astronautical Sciences* 135 (1<sup>er</sup> jan. 2010).
- [RWR98] C. V. RAO, S. J. WRIGHT et J. B. RAWLINGS. « Application of Interior-Point Methods to Model Predictive Control ». In : *Journal of Optimization Theory and Applications* 99.3 (1<sup>er</sup> déc. 1998), p. 723-757.
- [RRL01] Christopher V. RAO, James B. RAWLINGS et Jay H. LEE. « Constrained Linear State Estimation—a Moving Horizon Approach ». In : *Automatica* 37.10 (1<sup>er</sup> oct. 2001), p. 1619-1628.
- [RMD17] James RAWLINGS, David MAYNE et Moritz DIEHL. *Model Predictive Control : Theory, Computation, and Design*. T. 2. Nob Hill Publishing Madison, WI, 2017.
- [Reb+07a] Yann G. REBOURS et al. « A Survey of Frequency and Voltage Control Ancillary Services—Part I : Technical Features ». In : *IEEE Transactions on Power Systems* 22.1 (fév. 2007), p. 350-357.
- [Reb+07b] Yann G. REBOURS et al. « A Survey of Frequency and Voltage Control Ancillary Services—Part II : Economic Features ». In : *IEEE Transactions on Power Systems* 22.1 (fév. 2007), p. 358-366.
- [Ret+22] András RETZLER et al. « Shooting Methods for Identification of Nonlinear State-Space Grey-Box Models ». In : *2022 IEEE 17th International Conference on Advanced Motion Control (AMC)*. 2022 IEEE 17th International Conference on Advanced Motion Control (AMC). Fév. 2022, p. 207-212.
- [Reu12] Paul REUSS. *Précis de neutronique*. EDP Sciences, 3 déc. 2012. 541 p.
- [Rey+23] Daniel R. REYNOLDS et al. « ARKODE : A Flexible IVP Solver Infrastructure for One-Step Methods ». In : *ACM Transactions on Mathematical Software* 49.2 (30 juin 2023), p. 1-26.
- [Ric11] Lewis RICHARDSON Fry. « IX. The Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations, with an Application to the Stresses in a Masonry Dam ». In : *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 210.459–470 (jan. 1911), p. 307-357.
- [SMG17] Guido SANCHEZ, Marina MURILLO et Leonardo GIOVANINI. « Adaptive Arrival Cost Update for Improving Moving Horizon Estimation Performance ». In : *ISA Transactions* 68 (1<sup>er</sup> mars 2017).
- [Sar07] Simo SARKKA. « On Unscented Kalman Filtering for State Estimation of Continuous-Time Nonlinear Systems ». In : *IEEE Transactions on Automatic Control* 52.9 (sept. 2007), p. 1631-1641.
- [ST23] Jennifer SCOTT et Miroslav TŮMA. *Algorithms for Sparse Linear Systems*. Nečas Center Series. Cham : Birkhäuser, 2023. 242 p.
- [Sim10] D. SIMON. « Kalman Filtering with State Constraints : A Survey of Linear and Nonlinear Algorithms ». In : *IET Control Theory & Applications* 4.8 (6 août 2010), p. 1303-1318.

- [SC02] D. SIMON et Tien Li CHIA. « Kalman Filtering with State Equality Constraints ». In : *IEEE Transactions on Aerospace and Electronic Systems* 38.1 (jan. 2002), p. 128-136.
- [Sim06] Dan SIMON. *Optimal State Estimation : Kalman, H [Infinity] and Nonlinear Approaches*. Hoboken, N.J : Wiley-Interscience, 2006. 526 p.
- [Söd02] Gustaf SÖDERLIND. « Automatic Control and Adaptive Time-Stepping ». In : *Numerical Algorithms* 31.1 (1<sup>er</sup> déc. 2002), p. 281-310.
- [SW97] Eduardo D. SONTAG et Yuan WANG. « Output-to-State Stability and Detectability of Nonlinear Systems ». In : *Systems & Control Letters* 29.5 (1<sup>er</sup> fév. 1997), p. 279-290.
- [Sor70] H. W. SORENSON. « Least-Squares Estimation : From Gauss to Kalman ». In : *IEEE Spectrum* 7.7 (juill. 1970), p. 63-68.
- [Sor98] Masoud SOROUSH. « State and Parameter Estimations and Their Applications in Process Control ». In : *Computers & Chemical Engineering* 23.2 (1<sup>er</sup> déc. 1998), p. 229-245.
- [ST98] William SQUIRE et George TRAPP. « Using Complex Variables to Estimate Derivatives of Real Functions ». In : *SIAM Review* 40.1 (jan. 1998), p. 110-112.
- [Suw+16] Rata SUWANTONG et al. « Stability Analysis and Robustness Assessment of Deterministic and Stochastic Nonlinear Moving Horizon Estimators ». In : *2016 IEEE 55th Conference on Decision and Control (CDC)*. 2016 IEEE 55th Conference on Decision and Control (CDC). Las Vegas, NV, USA : IEEE, déc. 2016, p. 2900-2905.
- [TMT14] Yuval TASSA, Nicolas MANSARD et Emo TODOROV. « Control-Limited Differential Dynamic Programming ». In : *2014 IEEE International Conference on Robotics and Automation (ICRA)*. 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong, China : IEEE, mai 2014, p. 1168-1175.
- [TR02] M.J. TENNY et J.B. RAWLINGS. « Efficient Moving Horizon Estimation and Nonlinear Model Predictive Control ». In : *Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301)*. Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301). T. 6. Mai 2002, 4475-4480 vol.6.
- [Unb+00] R. UNBEHAUEN et al. « Stochastic Stability of the Continuous-Time Extended Kalman Filter ». In : *IEE Proceedings - Control Theory and Applications*. T. 147. 1. 1<sup>er</sup> jan. 2000, p. 45-52.
- [Ung09] Sridhar UNGARALA. « Computing Arrival Cost Parameters in Moving Horizon Estimation Using Sampling Based Filters ». In : *Journal of Process Control* 19.9 (oct. 2009), p. 1576-1588.
- [Val+17] Jayaram VALLURU et al. « Development of Moving Window State and Parameter Estimators under Maximum Likelihood and Bayesian Frameworks ». In : *Journal of Process Control*. DYCOPS-CAB 2016 60 (1<sup>er</sup> déc. 2017), p. 48-67.
- [VL24] Charlie VANARET et Sven LEYFFER. « Unifying Nonlinearly Constrained Non-convex Optimization ». In : *Mathematical Programming Computation journal* (2024).
- [VDD24] Lander VANROYE, Joris DE SCHUTTER et Wilm DECRÉ. « A Generalization of the Riccati Recursion for Equality-constrained Linear Quadratic Optimal Control ». In : *Optimal Control Applications and Methods* 45.1 (jan. 2024), p. 436-454.

- [Var+22] Devyani VARSHNEY et al. « Moving Horizon Estimator for Nonlinear and Non-Gaussian Stochastic Disturbances ». In : *Journal of Process Control* 116 (1<sup>er</sup> août 2022), p. 234-254.
- [WB06] Andreas WÄCHTER et LORENZ T. BIEGLER. « On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming ». In : *Mathematical Programming* 106.1 (1<sup>er</sup> mars 2006), p. 25-57.
- [WLM23] Tobias M. WOLFF, Victor G. LOPEZ et Matthias A. MÜLLER. *Robust Stability of Gaussian Process Based Moving Horizon Estimation*. 14 juin 2023. URL : <http://arxiv.org/abs/2304.06530> (visité le 05/10/2023). Prépubl.
- [YMM13] Tao YANG, Prashant G. MEHTA et Sean P. MEYN. « Feedback Particle Filter ». In : *IEEE Transactions on Automatic Control* 58.10 (oct. 2013), p. 2465-2480.
- [ZLB08] Victor M. ZAVALA, Carl D. LAIRD et LORENZ T. BIEGLER. « A Fast Moving Horizon Estimation Algorithm Based on Nonlinear Programming Sensitivity ». In : *Journal of Process Control*. Selected Papers From Two Joint Conferences : 8th International Symposium on Dynamics and Control of Process Systems and the 10th Conference Applications in Biotechnology 18.9 (1<sup>er</sup> oct. 2008), p. 876-884.





---

**Titre :** Estimation avancée à horizon glissant, état et paramètres, pour la conduite du cœur des réacteurs PWR.

**Mot clés :** Estimation par horizon glissant, Identification de paramètres, Physique des réacteurs.

**Résumé :** La commande prédictive (MPC) comporte d'indéniables atouts pour la commande du coeur des réacteurs à eau pressurisée (REP); elle a vocation à enrichir l'offre commerciale de Framatome. Néanmoins, les performances de la MPC sont conditionnées par la qualité du modèle de prédiction sous-jacent et la connaissance à chaque instant de l'état courant du modèle de réacteur. Cette thèse revient sur ces deux aspects qui constituent les verrous industriels. Elle propose un modèle de coeur de complexité juste suffisante, et le moyen de procéder à la calibration adaptative de certains de ses paramètres. C'est que la dynamique du coeur évolue avec l'épuisement du combustible; sa prise en compte par des paramètres prédétermi-

nés nécessite un travail conséquent. A rebours, nous proposons un estimateur à horizon glissant (MHE), pour estimer conjointement l'état du modèle ainsi que certains paramètres. Raideur et non-linéarité du modèle prédictif (en lien avec les caractéristiques du process) rendent les choix d'implémentation de l'estimateur cruciaux. Exploitation directe du modèle à temps continu, intégrateur implicite, collocation, interpolation des données d'entrée sont autant de concepts participant à la performance et la résilience de la MHE, là où les méthodes traditionnelles échouent (e.g. EKF). L'estimateur est validé sur la base de simulations fines réalisées sur le simulateur expert SOFIA utilisé à Framatome.

---

**Title:** Moving horizon estimation of states and parameters geared towards PWR monitoring and control.

**Keywords:** Moving horizon estimation, Parameter identification, Reactor physics.

**Abstract:** The model predictive control (MPC) approach has undeniable advantages for the core control of Pressurized Water Reactors (PWR); it is poised to enhance Framatome's commercial offering. Nevertheless, the performance of MPC is conditioned by the quality of the underlying predictive model and the real-time knowledge of the current state of the reactor core model. This thesis revisits these two aspects that constitute industrial barriers. It proposes a model of the core with just sufficient complexity and a method to perform adaptive calibration of some of its parameters. This is because the core dynamics evolve with fuel depletion; accounting for this through pre-

determined parameters requires significant effort. We propose the use of a Moving Horizon Estimator (MHE) to jointly estimate the state variables of the model. The stiffness and non-linear nature of the dynamic model (related to the characteristics of the process) make the the implementation choices of the estimator crucial. Direct exploitation of the continuous-time model, implicit integrator, collocation, interpolation of input data, are key concepts resulting into a performant and resilient MHE estimator, where traditional methods fail (e.g., EKF). The estimator is validated based on fine simulations carried out on the expert simulator SOFIA used at Framatome.