Conception de systèmes de contrôle avancé de reacteur PWR flexible

Guillaume Dupre

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

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, productique, automatique et robotique*

Par

Guillaume DUPRÉ

Conception de systèmes de contrôle avancé de réacteur PWR flexible

Vers une solution industrielle

Thèse présentée et soutenue à IMT Atlantique – campus de Nantes, le 6 décembre 2023
Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N)
Thèse N° : 2023IMTA0365

Rapporteurs avant soutenance :

Gildas BESANÇON Professeur, Grenoble INP – Ense3
Guillaume SANDOU Professeur, Centrale – Supélec

Composition du Jury :

Président :	Jean-Jacques LOISEAU	Directeur de recherche CNRS, Centrale Nantes
Examinateurs :	Gildas BESANÇON	Professeur, Grenoble INP – Ense3
	Ionela PRODAN	Maître de conférences, Grenoble INP – Esisar
	Guillaume SANDOU	Professeur, Centrale – Supélec
Dir. de thèse :	Philippe CHEVREL	Professeur, IMT Atlantique
Encadrant :	Mohamed YAGOUBI	Maître assistant, IMT Atlantique

Invité(s) :

Henri BOURLÈS Professeur émérite, CNAM
Alain GROSSETÊTE Senior expert, Framatome

Remerciements

À défaut de présenter ma thèse en 180 s, la voici en 4 citations :

— Année 1 : « All models are wrong ; some models are useful. », George E. P. Box.

— Année 2 : « In theory, theory and practice are the same. In practice, they are not. », pas clair d'où vient la citation donc anonyme.

— Année 3 : « C'est plus facile d'optimiser un truc qui marche, que de faire marcher un truc optimisé. », Alain Grossetête.

— Année 4 : « Le mieux est l'ennemi du bien. », Voltaire et toute personne m'ayant côtoyé pendant l'écriture de ma thèse.

ENFIN LIBRE ! Après avoir commencé ma thèse 10 mois en retard par rapport à la date initialement proposée à M. GROSSETÊTE en 2018, puis terminé la rédaction du manuscrit 12 mois en retard par rapport à la date initialement convenue avec Alain en 2022, voici enfin venu le moment d'écrire les remerciements 3 mois après la date de soutenance (l'important c'est d'avoir PhD 2019-2023 sur Linkedin) : ça va être long donc accrochez-vous svp !

Premièrement, tout d'abord, je souhaite commencer par remercier les membres du jury d'avoir évalué mes travaux de thèse. Merci en particulier aux rapporteurs Guillaume SANDOU et Gildas BESANÇON d'avoir pris le soin de lire très attentivement l'ensemble du manuscrit qui, je trouve, est tout de même assez long. Leurs nombreux retours, dans l'ensemble très positifs, m'ont fait énormément de bien au moral, surtout après avoir passé plus d'an à écrire sur mon temps libre. Je remercie, plus sobrement mais toujours avec sincérité, Ionela PRODAN et Jean-Jacques LOISEAU d'avoir accepté d'examiner ma thèse. Merci à Henri BOURLÈS, mon grand-directeur de thèse (i.e., le directeur de thèse de mon directeur de thèse), d'avoir participé à ma soutenance : j'espère que le service de livraison de manuscrit à domicile vous aura plu. Un tonnerre de remerciements pour Philippe CHEVREL et Mohamed YAGOUBI, mes encadrants de thèse côté laboratoire et ex-professeurs aux mines de Nantes, de m'avoir supporté (au sens donner du support puisque je suis supportable j'espère) toutes ces années, y compris lorsque j'étais étudiant. Petite dédicace à Fabien CLAVEAU, qui fut également l'un de mes professeurs aux mines de Nantes en option automatique et informatique industrielle, que j'ai eu le plaisir de croiser plusieurs fois lors de mes trop rares visites au labo. Merci à Alain GROSSETÊTE, dit « le boss », de m'avoir encadré en entreprise pendant et après ma thèse (encore aujourd'hui). S'il ne fallait retenir qu'une chose de mes 3 encadrants de thèse, ce serait leur gentillesse, leur patience et leur bienveillance (ça fait 3 choses mais tant pis). Côté école doctorale, je tiens à

remercier Delphine TURLIER de m'avoir systématiquement rappelé de remplir les documents administratifs sans jamais s'énerver. De même, je souhaite remercier Yasmina BENMERIEM et Valérie HENRY de m'avoir régulièrement aidé sur le plan administratif côté entreprise.

Puis, deuxièmement, je tiens à remercier tous les collègues que j'ai pu rencontrer pendant ce périple de plus de 4 ans (mais pas sur Linkedin, 2019-2023). Un grand merci à Fady Elias NACCACHE et à Pierre VANPEENE de nous avoir aidé à valider le prototype d'aide au pilotage de centrale sur le simulateur présent à la tour AREVA. C'est toujours avec beaucoup de plaisir que je viens vous voir en salle simulateur ! Merci à Thomas RIOU, Florent SAREMBAUD et Ouiza AIT KACI ARAB, ma partenaire de quart, d'avoir participé à l'industrialisation du même prototype : Ouiza, bonne chance pour ta nouvelle vie à Toulouse, j'espère que tu n'oublieras pas comment effectuer des variations de charge en mode A pendant que je me dépêche de finir de manger pour te relayer ! Super merci à Cyril FIALA, alias « Steve » ou « agent Fialovsky », d'avoir rompu la solitude des 6 premiers mois de thèse, même s'il est finalement parti faire du conseil au bout de 3 ans : Cyril Fialovsky était donc bien un agent double ! En parlant de redemption arc, merci à Perceval BEJA-BATTAIS, dit « le p'tit Perce », de revenir faire une thèse avec nous plus d'un an après avoir terminé son stage de fin d'étude : hâte de te revoir à la pause « baneune ». Merci également aux autres stagiaires, Ilian LELARGE, Anne-Laure POINTARD, Vincent LEMERLE et Adam DERESZEWKI d'avoir travaillé sur le projet : sans vous, mes travaux de thèse n'auraient certainement pas avancés aussi vite et la startup nation d'Alain ne se serait jamais autant développée ! Elle s'est d'ailleurs tellement développée, la startup, qu'on peut maintenant donner du travail à des ingénieurs en CDI qui imputent réellement des heures, à savoir Django LE CLERRE-MARAINE et Francesco MURATORI, que je remercie tous les deux pour leur enthousiasme. Merci à Lucas GRUSS, mon successeur à qui je cède la place de senior doctorant, d'avoir relu mes remerciements afin d'en diminuer le niveau de cringe : malheureusement, ça risque d'être encore assez cringe (e.g., ganbatte pour la dernière ligne droite Lucas-kun UwU). Bonne chance à Alexandre HACHE et Adrien RISPO qui arrivent eux aussi en fin de thèse : tatakae ! ! ! J'aimerais ensuite adresser un immense merci à tous les membres de l'équipe des essais et nouveaux produits, que j'ai d'abord squattée officieusement en tant que thésard avant de l'intégrer officiellement en tant qu'ingénieur d'études, pour leur humour et leur bonne humeur : je suis très heureux de vous retrouver au boulot. Grazie mille à Arnaud WERNER, mon ex-chef d'équipe, domo arigato gozaimasu à Barbara FRIJLINK, ma cheffe d'équipe actuelle, et merci beaucoup à Martin HALLE, amicalement surnommé « Marting le chieng », d'avoir tous les 3 fait le déplacement jusqu'à Nantes pour venir me soutenir pendant ma soutenance : cela m'a énormément touché de vous avoir comme supporters. Je tiens à remercier tout particulièrement Vincent SICCARDI, mon senpai, de m'avoir soutenu moralement et imposé un rythme d'écriture quotidien lors des 2 derniers mois précédant l'envoi de mon manuscrit aux rapporteurs. Comme promis, je t'accompagnerai à la prochaine Japan-expo (ordre de

mission validé par la cheffe normalement). Merci à Samy LE GRAND, alias « Samy the Big » ou « Samig », Pierre CHEVALIER, alias « Peter Knight » ou « le loup d'Arras », Ozan GOKKAYA, alias « the real OG » ou « wide/wise Ozan » voire « coach » lors des séances de muscu, Marceau TALPIED, le « manager du savoir », Théophile LEROY et Manon DIEUAIDE pour les nombreuses parties de Skull-King jouées ensemble pendant la pause méridienne : on a bien fait d'arrêter parce que ça commençait à devenir violent. Merci à Thomas FOY, alias « Tômô », d'être tout simplement cool et sympa : on se retrouve bientôt à la boxe si je survis aux premières séances ! Merci à Vincent DESMAISON, alias « Démèze », pour ses encouragements qui m'ont toujours remonté le moral après les présentations orales ! Merci à Sarah ECARNOT d'avoir encadré mes premières études d'ingénierie réalisées en parallèle de la rédaction, sans jamais perdre son calme malgré mes erreurs d'inattention : merci également pour tes précieux conseils immobiliers post-thèse. Merci à Maxime PFEIFFER d'être devenu fan de l'émission voyage aux pays de maths diffusée sur Arte : je ne suis plus le seul ! Merci à Stephane MARECHAL de rire de temps en temps aux bêtises que je raconte avant, pendant, et après l'heure du déjeuner (cela vaut également pour les autres collègues). Merci à Pierre BARTHELET, alias « Peanut Butter » (personne l'appelle comme ça), de nous avoir aidé à connecter le prototype d'aide au pilotage de centrale sur le simulateur et d'avoir répondu à mes nombreuses questions sur les réacteurs nucléaires : vivement que tu me racontes la formation cybersécurité/hacking ! Et enfin, merci à Cyprien DEGEZ, dernier à avoir intégré l'équipe ce qui m'a permis de passer senior CDI, pour ses histoires de culture de betteraves et de tracteurs que je trouve personnellement intéressantes, n'en déplaise aux fâcheux. Il me reste encore à remercier les collègues croisés en formation, Serena COSTANZO, Pierre BRETAGNOLLES et Nore BAKKAS, les collègues croisés dans le couloir/à la cantine/en salle de pause, Alban MARTINEZ-DELCAYROU et Pierre FIRDION, et les collègues directement issus des mines de Nantes, Akram BENAZZOU (démissionné trop tôt), Julien DUFOUR (parti forger au Creusot) et Baptiste BARBIER : merci à tous de m'avoir encouragé ces 4 dernières années.

Ensuite, troisièmement, j'aimerais remercier mes amis proches de m'avoir entendu parler de ma thèse pendant bien trop longtemps. Merci à mes amis de classes préparatoires Léo POISOT #groupedekhôllesnuméro9 (coucou Marion, je viens à Londres cet été), Rémi ROSENTHAL (je viens à Bayonne cet été) et Alexandre DOLLÉ (je viens à Malmö cet été), pour les journées, soirées, et vacances passées ensemble. Il remonte à loin le temps où on calculait des « POCA » pour réduire des endomorphismes (10 ans déjà) ! Merci également à mes amis d'école parfois un peu difficiles à joindre (c'est gratuit déso), à savoir Marine DUBILLARD, Nicolas MICHEL, alias « michou », et Sarah NABI pour les journées, soirées et vacances passées ensemble sans mes potes de prépa (2 salles, 2 ambiances). Le fait d'être tous partis en thèse prouve qu'on s'était bien trouvés à l'époque. Merci aux amis d'amis devenus amis tout court par transitivité, Marie BUISSON, Othmane SAYEM, Thimothée POIREL, Pablo CABEL, que j'ai pas mal

bassiné avec le « T-word » (surtout en période de rédaction). Merci au DJ Adil CHARMOUH, ex-camarade AII et ex-directeur de prod en simulation d'entreprise, et à sa famille de m'avoir invité plusieurs fois chez eux au Maroc : je n'ai jamais aussi bien mangé de ma vie (désolé papa et maman) ! Un gigantesque merci à mon ami Az-elarabe BITANE qui, à ce stade, fait presque parti de la famille (merci Imane de l'autoriser à jouer au PC avec moi). C'est quand que tu postules à Framatome pour qu'on puisse se voir encore plus souvent ? Big mcThankies à Kévin FROHLICHER d'avoir suivi avec moi le cursus X-mines international (polytechnique *campus* Montréal et mines *campus* Nantes) : je cite « à l'aide rue Berri » ! Merci d'ailleurs de m'avoir remercié dans ta thèse, je te renvoie l'appareil. Gros merci à Martin GUILLET, alias « Martin le polytechnicien » ou « m'sieur Guillet », d'être souvent venu me rendre visite pour que je prenne l'air en période de rédaction (désolé d'avoir abandonné la natation après le covid). J'espère que tu achèteras le prototype d'aide au pilotage de centrale qu'on a développé quand tu seras PDG de Flamanville : si ça bloque niveau hiérarchie, je peux attendre que tu deviennes PDG d'EdF. Merci à Maxime PARADIS (passe le bonjour à Arzum), alias « Max le québécois » ou « golden jet », parce que « Max, c'est un mec en or » comme je disais four beers deep à la Retenue (« John Deere, 10 hivars ») : j'arrive toujours pas à croire qu'on travaille dans la même boîte alors qu'on s'était pas parlé pendant 3 ans après Montréal ! Faudrait qu'on révise notre check personnalisé, je te cale un point sur Outlook lundi pro. En parlant de Montréal, merci à Amandin PAQUET et à Sheila de m'avoir hébergé à Barneville-Carteret et de s'occuper de m'sieur Guillet depuis qu'il a déménagé là-bas. Vous êtes sûr que vous voulez pas venir me rendre visite dans ma passoire thermique de 26.4m2 loi Carrez à Paris ? Merci à Yann CORLAY, alias « le breton » (je reste poli), d'être resté en contact avec moi ces dernières années : je passerai bientôt te rendre visite dans ton bled paumé en banlieue de Brest (amicalement bien sûr). Félicitation à toi et Orlane, j'ai bien reçu la carte. Merci aux mineurs restés à Nantes (Arthur, Vincent, Guillaume V, Sofiane, Adèle, Romain, Elisa, Loïc, Tristan, Joachim) pour les soirées Among Us et Valheim passées en ligne pendant le 1er confinement.

Enfin, quatrièmement, j'aimerais remercier l'ensemble des membres de ma famille qui m'ont régulièrement croisé pendant et avant ma thèse. Merci à mon oncle Pierre et ma tante Isabelle, ainsi que mes cousins Charles et Benoît, de venir presque à chaque fois passer Noël avec nous (Adil et Azel en special guests parfois). Merci à mon oncle François, alias « tonton la pirogue », « le vieil indien », « le sorcier zoulou » ou « chairman tonton François », et ma tante Dominique de venir nous rendre visite à la maison : François, c'était trop chouette de t'avoir les weekends pendant le 1er confinement, même si c'était pas vraiment autorisé lol. (Merci)*2 François (j'en remets une couche) d'être venu assister à ma soutenance et de partager avec moi tes histoires d'ancien ingénieur de Lafarge. Merci à mon cousin Laurent de s'être souvent occupé de moi enfant et adolescent : ça me manque les parties de DBZ Budokai 3 sur PS2 (et merci de m'avoir fait découvrir MGS3 aussi, c'est un classique). Merci à ma grande sœur Aude-Isabelle et à sa

famille (Raphaël, Antonin et Quitterie) de venir passer les vacances avec nous en Bretagne, même s'ils mettent un peu le bazar dans la salle de bain : j'ai pas trop suivi tes conseils, vu que je suis une tête de mule comme papa, mais j'ai quand même réussi à finir le manuscrit ! Merci à ma petite sœur Claire d'être une super petite sœur ultra marrante, même si elle a eu rattrapage d'automatique alors que je lui avais tout expliqué (bon j'avoue, j'ai pas pu t'aider sur la synthèse des régulateurs RST en temps discret alors que c'est Philippe qui m'avait donné le cours à l'époque...) : d'ailleurs, quand est-ce que tu vas enfin payer des impôts pour financer ma retraite ? Et pour finir, je garde toujours le meilleur pour la fin, merci à mes parents adorés Jean-Luc et Catherine, que j'aime plus que tout au monde, de m'avoir toujours soutenu et éduqué avec amour, et d'avoir fait de moi la personne que je suis aujourd'hui : je leur dédie cette thèse.

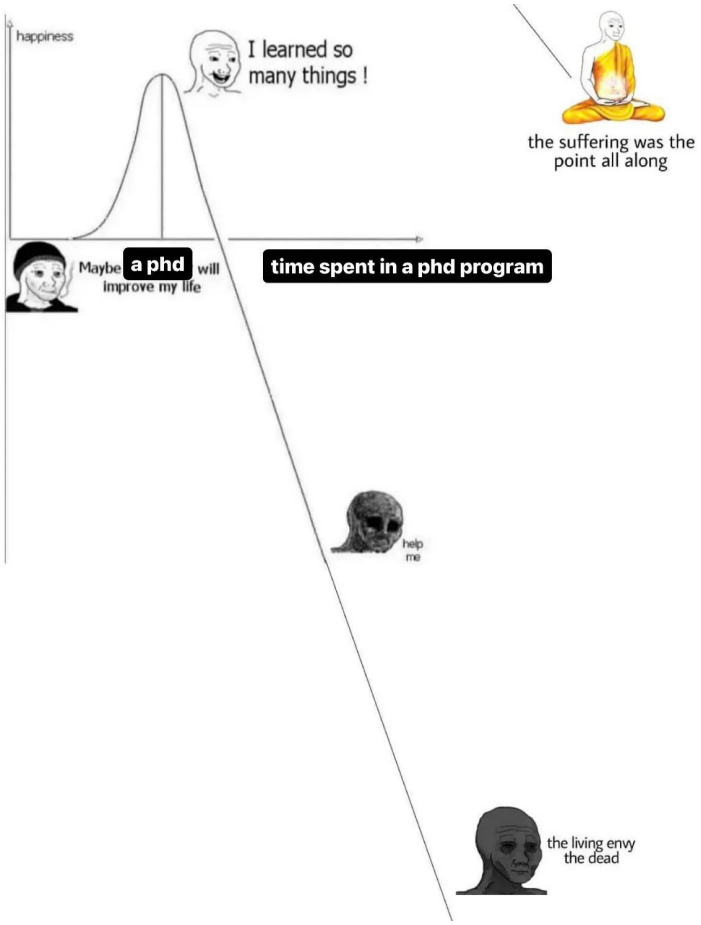

FIGURE 0.1 – Meme illustrant le bonheur d'un doctorant Wojak en fonction du temps passé en doctorat (trouvé sur le forum r/PhD de Reddit).

Bonus : je tiens à remercier mes enfants théoriques et ma femme imaginaire pour leur soutien affectif hypothétique. Et merci à moi du passé de ne pas avoir abandonné même si ça pouvait être tentant vers la fin (voir meme ci-dessus).

Résumé

La plupart des unités de production d'électricité d'origine renouvelable déployées ces dernières années sont par nature intermittentes. En l'absence de solution de stockage à grande échelle, la production et la consommation d'électricité doivent être constamment équilibrées pour garantir la stabilité du réseau. Ce rôle, traditionnellement occupé par les centrales thermiques à flamme, tend de plus en plus à être assuré par les centrales nucléaires. Ainsi, cette thèse vise à améliorer la flexibilité des réacteurs nucléaires à eau sous pression afin de répondre aux futurs besoins du réseau électrique. Pour ce faire, plusieurs systèmes de contrôle du cœur du réacteur ont été conçus en se basant sur des méthodes avancées du domaine de l'automatique, à savoir la commande prédictive et la commande à gains séquencés. Un modèle non-linéaire de réacteur multi-mailles, destiné à la synthèse de lois de commande, a notamment dû être développé. De complexité juste suffisante, il est bien adapté à des fins de prédiction court terme. La solution finalement proposée comporte deux volets : 1) un système temps réel d'aide au pilotage (brevet monde), qui fait désormais partie de l'offre commerciale de Framatome, et 2) une solution de pilotage hiérarchique compatible avec les boucles de régulation de température existantes, dont les performances sont nettement accrues en termes de flexibilité et de respect des contraintes opérationnelles, par rapport aux modes de pilotage actuels. Cette solution s'appuie sur les techniques d'implémentation de commande prédictive non-linéaire les mieux adaptées.

Abstract

Most renewable electricity generation units deployed in recent years are inherently intermittent. In the absence of large-scale storage solutions, electricity production and consumption must be constantly balanced to ensure grid stability. This role, traditionally played by fossil-fired power plants, is increasingly being filled by nuclear power plants. Hence, this thesis aims at enhancing the flexibility of pressurized water nuclear reactors to meet future grid requirements. To achieve this, several core control systems have been designed based on advanced control methods, namely model predictive control and gain-scheduling control. In particular, a non-linear multi-mesh reactor model, dedicated to the design of control laws, had to be developed. Its complexity is well-suited to short-term predictions. The solution ultimately proposed is twofold : 1) a real-time operator assistance system (world patent), which is now part of Framatome's commercial offer, and 2) a hierarchical control solution compatible with existing temperature control loops, whose performance is significantly enhanced in terms of flexibility and compliance with operational constraints, compared with current core control systems. This solution relies on the most relevant non-linear model predictive control implementation techniques.

TABLE DES MATIÈRES

LISTE DES FIGURES

Liste des tableaux

ACRONYMES

80%FDV	80 % d'avancement du cycle(*) en fin de vie
ACT	Average Coolant Temperature
AO	Axial Offset
DVX	Début de vie du cycle(*) à l'équilibre xénon
EPR	Evolutionary Power Reactor
FA3	Troisième réacteur de la centrale nucléaire de Flamanville
GCP	Groupes de compensation de puissance
IL	Insertions Limites
MPC	Model Predictive Control
NMPC	Nonlinear Model Predictive Control
PN	Puissance nominale
PVR	Power Variation Rate
PWR	Pressurized Water Reactor
REA	Circuit d'appoint en eau et en bore
RCP	Circuit primaire
RCV	Circuit volumétrique et chimique
REP	Réacteur à Eau sous Pression
TGE	Toutes Grappes Extraites.

(*) un cycle, sous-entendu d'irradiation, correspond à la durée entre deux rechargement du combustible nucléaire : en France, la durée d'un cycle est de 12 à 18 mois.

Liste des productions scientifiques

Articles de revues avec comité de lecture

[1] G. Dupré, P. Chevrel, M. Yagoubi et A. Grossetête, « Design and comparison of two advanced core control systems for flexible operation of pressurized water reactors », *Control Engineering Practice*, t. 123, p. 105 170, 2022.

[2] G. Dupré et A. Grossetête, « Système temps réel d'aide au pilotage : OAPS », *La parenthèse technique DTI*, *52*, 2023, revue interne de la Direction Technique et Ingénierie (DTI) de Framatome.

Communication avec actes et comité de lecture

[3] G. Dupré, A. Grossetête, P. Chevrel et M. Yagoubi, « Enhanced Flexibility of PWRs (mode A) Using an Efficient NMPC-Based Boration/Dilution System », in *2021 European Control Conference (ECC)*, IEEE, 2021, p. 1092-1098.

Brevet enregistré

[4] A. Grossetête, G. Dupre, C. Fiala, P. Chevrel et M. Yagoubi, « Procédé et ensemble de pilotage d'un réacteur nucléaire, réacteur nucléaire équipé d'un tel ensemble », WO2022219117, 2022.

Présentations orales

[5] A. Grossetête et G. Dupré, *Démonstrateur d'aide à la conduite et manœuvrabilité*, Présentation à l'Institut de recherche Tripartite (CEA/EDF/Framatome) sur les moyens de conduite du futur, 2021.

[6] A. Grossetête, G. Dupré et L. Gruss, *Predictive systems for flexible operation*, Présentation au conseil scientifique de Framatome, 2022.

[7] G. Dupré, *Conception de systèmes de contrôle avancé de réacteur PWR flexible : vers une nouvelle solution industrielle*, Présentation à la réunion d'équipe CODEx du LS2N, 2022.

[8] A. Grossetête, G. Dupre, C. Fiala et L. Gruss, *Real-time predictive core control for optimal flexibility*, Présentation au département conception cœur et études de transitoires de la Direction Technique et Ingénierie (DTI) de Framatome, 2022.

INTRODUCTION

Contexte et problématique

En réponse au premier choc pétrolier de 1973, la France a fait le choix de réduire sa dépendance énergétique aux énergies fossiles en produisant une électricité d'origine majoritairement nucléaire. Un total de 56 réacteurs à eau sous pression, répartis dans 18 centrales nucléaires, génèrent quotidiennement plus de 70 % de l'électricité en France [1].

La vitesse de rotation des machines tournantes synchrones connectées au réseau électrique (alternateurs, turbines, moteurs) dépend de la fréquence de l'onde électrique qui les parcourt : la vitesse de rotation augmente lorsque la fréquence augmente, et inversement. Comme il n'existe encore aucun moyen de stocker l'énergie électrique à grande l'échelle [1], [2], la seule façon de maintenir la fréquence du réseau autour de sa valeur de consigne (50 ± 0.05 Hz en Europe) est de veiller à ce que la production et la consommation d'électricité soient constamment équilibrées : à production constante, la fréquence augmente lorsque la consommation diminue et diminue lorsque la consommation augmente. Si le déséquilibre de fréquence devient trop important, les centrales électriques finiront par être automatiquement déconnectées du réseau afin de préserver l'intégrité de leurs équipements, ce qui pourrait entraîner une panne de courant généralisée [3], [4]. Pour éviter que cela se produise, un nombre minimum de centrales électriques doivent être flexibles, c'est-à-dire capables d'ajuster leur production sur demande, de sorte à pouvoir contrôler la fréquence du réseau.

Dans la plupart des pays du monde, l'équilibre entre la production et la consommation d'électricité est assuré par des centrales thermiques à flamme, faciles à piloter mais alimentées avec du combustible fossile (charbon, gaz, pétrole). Les centrales nucléaires, quand elles existent, fournissent alors la base de la production électrique, en restant à puissance maximale sans tenir compte de la demande. Ce mode d'exploitation est en effet le plus rentable sur le plan économique, car, contrairement aux centrales thermiques à flamme, les coûts marginaux liés à l'achat du combustible nucléaire sont négligeables par rapport aux coûts fixes liés à la construction, la maintenance et la sûreté de l'installation [5]. En revanche, lorsque la part du nucléaire est prépondérante dans le mix énergétique d'un pays, l'équilibre offre-demande doit forcément être compensé par tout ou partie des centrales nucléaires. En France, par exemple, la production électrique des centrales nucléaires est ajustée plusieurs fois par jours depuis les années 1980 [1].

Aujourd'hui, de nombreux pays cherchent à remplacer leurs centrales électriques à énergie fossile par des sources d'énergie renouvelables afin de lutter contre le changement climatique et

la hausse des prix de l'énergie. Or, la majorité des unités de productions d'électricité d'origine renouvelable déployées ces dernières années, telles que les éoliennes ou les panneaux photovoltaïques, sont par nature intermittentes, donc difficilement pilotables [6]. De ce fait, les centrales nucléaires sont de plus en plus amenées à participer au réglage de fréquence du réseau, notamment lorsque les baisses de puissance des centrales à énergie fossile ne suffisent pas à compenser le surplus d'électricité généré par les sources d'énergie renouvelables. Cela peut arriver, par exemple, à la suite de mauvaises prévisions météorologiques, lors de journées excessivement venteuses et/ou ensoleillées. De telles journées se traduisent généralement par l'apparition de prix bas, voire négatifs, sur le marché de l'électricité, qui incitent les exploitants à diminuer rapidement la puissance de leurs centrales. Malheureusement, cela n'est pas toujours possible, car la flexibilité de la centrale nucléaire, autrement dit de ses réacteurs, peut être limitée par le mode de pilotage du cœur.

L'objectif de cette thèse est donc d'améliorer la flexibilité des réacteurs nucléaires à eau sous pression afin de répondre aux futurs besoin du réseau électrique. Pour ce faire, plusieurs systèmes de commande du cœur ont été conçus en se basant sur des méthodes avancées du domaine de l'automatique, à savoir la commande prédictive et la commande à gains séquencés. La solution proposée devra être intelligible et de complexité limitée pour rester viable sur le plan industriel. De ce fait, certains composants des modes de pilotage existants ne pourront pas être remis en cause. Les performances de la solution proposée seront comparées à celle du mode T, le dernier mode de pilotage conçu par Framatome à destination de l'EPR.

Contributions

Une première contribution, d'ordre méthodologique, est d'avoir identifié et présenté en détails l'ensemble des étapes à suivre et des choix à effectuer pour concevoir un algorithme de commande prédictive non-linéaire. La démarche a été de montrer qu'il n'existe pas de recette miracle et que chaque brique utilisée pour mettre en place l'algorithme vient avec son lot de difficultés. En particulier, il est important de retenir que la polyvalence de la commande prédictive permet certes de résoudre des problèmes habituellement compliqués à traiter en automatique, notamment le contrôle des systèmes multivariables, non-linéaires et/ou à retard, mais nécessite en contrepartie de maîtriser de nombreuses notions théoriques et outils numériques issus de disciplines différentes (automatique, modélisation physique, analyse numérique, optimisation) pour que l'algorithme obtenu fonctionne efficacement en pratique.

Une deuxième contribution, d'ordre technique, est d'avoir repris et amélioré les travaux réalisés dans la thèse précédente :

1) Le modèle de réacteur point utilisé pour concevoir, régler et tester les systèmes de commande a été transformé en modèle de réacteur multi-mailles afin de représenter plus fi-

dèlement le comportement du déséquilibre axial de puissance du cœur. En conséquence, l'ordre du modèle a été triplé, celui-ci étant passé de 12 à 37 variables et équations d'état, ce qui complexifie forcément la conception des systèmes de commande.

2) Une seconde version du contrôleur à gains séquencés de type proportionnel-intégral a été mise au point à partir du nouveau modèle de réacteur multi-mailles. La formulation mathématique du problème H_2/H_∞ utilisé pour régler les gains a été affinée et clarifiée, et les phases d'interpolation et de réglage des gains ont été fusionnées.
3) L'algorithme de commande prédictive non-linéaire a été entièrement remanié en s'appuyant sur la méthodologie décrite dans la thèse. Le principal défi a été de diminuer le temps de résolution du problème d'optimisation. Pour ce faire, l'ordre du modèle de réacteur multi-mailles embarqué dans l'algorithme a été réduit en négligeant le comportement transitoire des variables d'état rapides par rapport aux grandeurs d'intérêts. De même, le problème de commande optimale a été transcrit en utilisant une méthode simultanée basée sur un schéma d'intégration numérique de type collocation. Enfin le problème d'optimisation obtenu a été résolu en faisant appel au solveur `IPOPT` inclus dans la boîte outils `CasADi`. Cette dernière permet de calculer efficacement les matrices Jacobienne et Hessienne du problème par différentiation algorithmique et de renvoyer leur structure au solveur d'optimisation grâce à des techniques de coloration de graphe. Un schéma de compensation du délai a néanmoins dû être ajouté à l'algorithme de commande prédictive pour tenir compte du temps de calcul de la commande.

La troisième contribution, d'ordre technique, est de proposer un système de commande du cœur crédible sur le plan industriel, dont les performances sont supérieures à celle du mode T, le dernier mode de pilotage conçu par Framatome. Le problème, en apparence similaire à celui de la thèse précédente, est en réalité beaucoup difficile à résoudre, car les exigences spécifiées par Framatome ont depuis lors été durcies :

- la régulation de température du mode T doit obligatoirement être incluse dans l'architecture finale pour ne pas complètement remettre en cause l'existant.
- le programme d'insertion du groupe $\mathrm{P_{bank}}$, qui permet de contrôler indirectement la capacité de retour en puissance de la turbine, ne doit plus être utilisé.
- les limites d'insertion des groupes de grappes doivent être respectées en début et en fin de vie du cycle d'irradiation du combustible.
- le système de commande doit permettre à la turbine de remonter rapidement, à une vitesse de $5\,\%\mathrm{PN\,min^{-1}}$, à sa puissance nominale en début et en fin de vie du cycle d'irradiation.
- le système de commande doit être capable de fonctionner en temps réel.

Le système de commande obtenu s'inspire grandement de la solution hiérarchisée proposée dans la thèse précédente et a été élaboré en combinant la boucle de régulation de température du mode T avec un algorithme de commande prédictive non-linéaire. Le rôle de la boucle de ré-

gulation de température est de garantir que la grandeur d'intérêt la plus critique en termes de sûreté, à savoir la température moyenne du circuit primaire, reste à l'intérieur du domaine de fonctionnement autorisé en dépit des incertitudes de modèle et des perturbations non mesurables. Cette partie du système de commande remplace le régulateur à gains séquencés de type proportionnel-intégral de la thèse précédente. Le rôle de l'algorithme de commande prédictive non-linéaire est d'assister la régulation de température moyenne, en fournissant une action anticipatrice « feedforward » au système, et de contrôler le déséquilibre axial de puissance du cœur, le tout en tenant compte les contraintes du cahier des charges.

Plan du manuscrit

Le manuscrit est composé de 3 chapitres, introduction et conclusion exclues :

— Le **chapitre 1** présente de façon détaillée l'ensemble des concepts théoriques et pratiques à maîtriser pour concevoir et régler un algorithme de commande prédictive. Les différentes étapes méthodologiques à suivre pour mettre en place l'algorithme apparaissent dans l'ordre chronologique.

— Le **chapitre 2** commence par rappeler les notions physiques essentielles à la compréhension du modèle de réacteur multi-mailles développé pour concevoir les systèmes de commande du cœur. Les équations différentielles du modèle sont ensuite déterminées unes par unes, avant d'être mises sous forme d'état non-linéaire.

— Le **chapitre 3** est consacré à l'élaboration des systèmes de commande du cœur. Les travaux préliminaires, ayant permis d'aboutir à la solution proposée, sont d'abord résumés. Puis, le cahier des charges à respecter et la solution proposée sont ensuite présentés. Enfin, les résultats obtenus en simulation sont comparées à ceux du mode T.

Chapitre I

Méthodologie de conception de lois de commande prédictive

1 Introduction du chapitre

L'engouement que suscite la commande prédictive dans le domaine de l'automatique depuis le début des années 2000 s'explique facilement par ses nombreux atouts. Tout d'abord, le réglage d'un algorithme de commande prédictive requiert principalement de se concentrer sur la modélisation du système à réguler et sur la formulation du problème d'optimisation associé, contrairement aux méthodes conventionnelles qui nécessitent de s'attarder longuement sur la définition et le réglage de chaque bloc fonctionnel du contrôleur (filtres, gains, hystérésis, saturations, etc.). L'accent est donc directement porté sur le besoin réel (le système et le problème à résoudre) plutôt que sur la solution (l'architecture du contrôleur), ce qui permet à la méthode de rester très intuitive, donc facilement paramétrable pour un ingénieur non spécialisé. La polyvalence de la commande prédictive permet également de traiter de nombreuses situations habituellement compliquées à gérer en automatique. En effet, l'avantage de calculer la commande en résolvant un problème d'optimisation est de pouvoir facilement prendre en compte les contraintes et les retards qui s'appliquent sur le système à réguler, ce dernier pouvant être non-linéaire et/ou multivariable (c'est-à-dire comportant plusieurs entrées et plusieurs sorties). Cette méthode est donc particulièrement adaptée au pilotage des centrales nucléaires, celles-ci étant intrinsèquement non-linéaires, multivariables, soumises à des contraintes et à divers retards. Enfin, le caractère optimal de la commande prédictive lui permet de surpasser les systèmes de commande classiques sur le plan de la performance.

Néanmoins, bien que la commande prédictive soit simple à appréhender d'un point de vue théorique, elle reste difficile à déployer en pratique. Un premier obstacle est que le calcul de la loi de commande prédictive nécessite de résoudre périodiquement et en temps réel un problème d'optimisation sous contraintes. En comparaison, le calcul d'une loi de commande conventionnelle requiert simplement d'évaluer les blocs fonctionnels du contrôleur la définissant. Un second obstacle est que la méthode se situe au croisement de plusieurs disciplines (automatique, modélisation physique, analyse numérique, optimisation) chacune étant complexe à maîtriser. Par

ailleurs, certains verrous théoriques persistent et peuvent limiter son application pratique. En particulier, la question de la robustesse de l'algorithme face aux perturbations non prévisibles et aux erreurs de modélisation reste, à l'heure actuelle, encore ouverte. De plus, l'algorithme a besoin d'avoir accès à l'ensemble des variables d'état du modèle pour fonctionner correctement. Or, comme ces dernières ne sont pas toujours toutes mesurables, sa mise en place doit souvent s'accompagner de celle d'un estimateur d'état, ce qui soulève davantage de questions quant à la robustesse du schéma de régulation global.

2 Généralités en commande prédictive

2.1 Modélisation des systèmes dynamiques contrôlés

Afin de déterminer les futures actions à envoyer au système réel, l'algorithme de commande prédictive doit, par définition, être capable de prédire son évolution dans le temps. Pour ce faire, la première étape est de décrire le comportement du système à l'aide d'un modèle mathématique. Ce dernier peut être obtenu soit de manière théorique, par raisonnement physique, soit de manière expérimentale, à partir des mesures et observations collectées sur site, soit par une combinaison des deux approches. En commande prédictive, le modèle est fréquemment représenté sous forme d'état non-linéaire à temps discret. Le choix retenu dans cette thèse est de partir d'une représentation d'état à temps continu, plus proche du monde physique, et de la convertir ensuite en temps discret. En général, ce type de modèles peut s'écrire de façon implicite [7]-[9] :

$$\mathbf{F}\Big(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t)\Big) = \mathbf{0}\,, \tag{1.1}$$

où $\mathbf{F} : \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}} \to \mathbb{R}^{n_\mathbf{x}}$ est une fonction supposée ici au moins de classe $\mathcal{C}^2$ sur son ensemble de définition, qui dépend du vecteur d'état $\mathbf{x}(t) \in \mathbb{R}^{n_\mathbf{x}}$, de sa dérivée temporelle $\dot{\mathbf{x}}(t) \in \mathbb{R}^{n_\mathbf{x}}$, et du vecteur de commande $\mathbf{u}(t) \in \mathbb{R}^{n_\mathbf{u}}$. En outre, il arrive parfois que la fonction $\mathbf{F}(t, \dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t), \mathbf{p})$ dépende explicitement du temps $t \in \mathbb{R}_{\geq 0}$ et/ou d'un vecteur de paramètres constants $\mathbf{p} \in \mathbb{R}^{n_\mathbf{p}}$. Dans ce cas, il est toujours possible de revenir à l'écriture implicite (1.4) en augmentant la taille du vecteur d'état :

$$\forall\, t_0 \geq 0,\ \forall\, t \geq t_0,\ \dot{\mathbf{x}}_{\text{aug}}(t) = \begin{bmatrix} \dot{\mathbf{x}}(t) \\ 1 \\ \mathbf{0} \end{bmatrix},\ \text{et}\ \mathbf{x}_{\text{aug}}(t_0) = \begin{bmatrix} \mathbf{x}(t_0) \\ t_0 \\ \mathbf{p} \end{bmatrix} \Rightarrow\ \mathbf{x}_{\text{aug}}(t) = \begin{bmatrix} \mathbf{x}(t) \\ t \\ \mathbf{p} \end{bmatrix}. \tag{1.2}$$

À noter toutefois que cette astuce est surtout employée pour conserver une notation uniforme dans tout le manuscrit, et qu'il sera généralement plus efficace, sur le plan numérique, de réserver un traitement particulier à ces paramètres additionnels. De même, si le modèle admet une position d'équilibre $(\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}}) \in \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}}$ vérifiant, par définition, $\mathbf{F}(\mathbf{0}, \mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}}) = \mathbf{0}$, alors

celle-ci peut être systématiquement ramenée à l'origine $(\mathbf{0}, \mathbf{0})$ via un changement de variable :

$$\widetilde{\mathbf{F}}\Big(\dot{\widetilde{\mathbf{x}}}(t), \widetilde{\mathbf{x}}(t), \widetilde{\mathbf{u}}(t)\Big) \coloneqq \mathbf{F}\Big(\dot{\widetilde{\mathbf{x}}}(t), \widetilde{\mathbf{x}}(t) + \mathbf{x}_{\text{eq}}, \widetilde{\mathbf{u}}(t) + \mathbf{u}_{\text{eq}}\Big), \text{ avec : } \begin{cases} \widetilde{\mathbf{x}}(t) \coloneqq \mathbf{x}(t) - \mathbf{x}_{\text{eq}} \\ \widetilde{\mathbf{u}}(t) \coloneqq \mathbf{u}(t) - \mathbf{u}_{\text{eq}} \,. \end{cases} \tag{1.3}$$

Pour un signal de commande $\mathbf{u} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_{\mathbf{u}}}$ donné, il est possible de calculer de façon déterministe la trajectoire de l'état $\mathbf{x} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_{\mathbf{x}}}$ du modèle à n'importe quel instant $t \geq t_0$ en résolvant le problème de Cauchy :

$$\mathbf{F}\Big(\dot{\mathbf{x}}(t), \mathbf{x}(t), \mathbf{u}(t)\Big) = \mathbf{0}, \ \mathbf{x}(t_0) = \mathbf{x}_0 \,, \tag{1.4}$$

où $\mathbf{x}_0 \in \mathbb{R}^{n_{\mathbf{x}}}$ est l'état du modèle à l'instant initial $t_0 \geq 0$. En un sens, l'état constitue donc la mémoire du passé du modèle, puisqu'il suffit de connaître sa valeur à un instant fixé pour déterminer son évolution future à partir du signal de commande.

Lorsque la matrice Jacobienne $\partial \mathbf{F} / \partial \dot{\mathbf{x}}$ est inversible, le modèle est composé uniquement d'équations différentielles ordinaires. Dans cette situation, il est beaucoup plus courant d'utiliser l'écriture explicite :

$$\dot{\mathbf{x}}(t) = \mathbf{f}\Big(\mathbf{x}(t), \mathbf{u}(t)\Big), \ \mathbf{x}(t_0) = \mathbf{x}_0 \,, \tag{1.5}$$

où $\mathbf{f} : \mathbb{R}^{n_{\mathbf{x}}} \times \mathbb{R}^{n_{\mathbf{u}}} \to \mathbb{R}^{n_{\mathbf{x}}}$ est la fonction d'évolution du modèle. En revanche, si la matrice Jacobienne $\partial \mathbf{F} / \partial \dot{\mathbf{x}}$ n'est pas inversible, alors le modèle contient à la fois des équations différentielles ordinaires et des équations algébriques. Ces modèles, dits descripteur, sont une extension de ceux représentés sous forme d'état auxquels ont été ajoutées des lois de conservations, des relations empiriques ou des contraintes géométriques. Dans ce contexte, le vecteur $\mathbf{x}(t) \in \mathbb{R}^{n_{\mathbf{x}}}$ est renommé vecteur de description, car certaines de ses composantes ne sont plus des variables d'état mais des variables algébriques. Quand la distinction entre les deux types de variables est évidente, il est souvent préférable d'exprimer le modèle descripteur sous forme semi-explicite :

$$\begin{cases} \dot{\mathbf{x}}_{\mathbf{d}}(t) = \mathbf{f}_{\mathbf{d}}\Big(\mathbf{x}_{\mathbf{d}}(t), \mathbf{x}_{\mathbf{a}}(t), \mathbf{u}(t)\Big) \\ \qquad \mathbf{0} = \mathbf{f}_{\mathbf{a}}\Big(\mathbf{x}_{\mathbf{d}}(t), \mathbf{x}_{\mathbf{a}}(t), \mathbf{u}(t)\Big) \\ \mathbf{x}_{\mathbf{d}}(t_0) = \mathbf{x}_{\mathbf{d},0} \,, \end{cases} \tag{1.6}$$

où la fonction $\mathbf{f}_{\mathbf{d}} : \mathbb{R}^{n_{\mathbf{d}}} \times \mathbb{R}^{n_{\mathbf{a}}} \times \mathbb{R}^{n_{\mathbf{u}}} \to \mathbb{R}^{n_{\mathbf{d}}}$ représente les équations qui décrivent explicitement l'évolution de l'état $\mathbf{x}_{\mathbf{d}}(t) \in \mathbb{R}^{n_{\mathbf{d}}}$ du vecteur de description $\mathbf{x}(t) = [\mathbf{x}_{\mathbf{d}}(t)^\top \ \mathbf{x}_{\mathbf{a}}(t)^\top]^\top$, et où la fonction $\mathbf{f}_{\mathbf{a}} : \mathbb{R}^{n_{\mathbf{d}}} \times \mathbb{R}^{n_{\mathbf{a}}} \times \mathbb{R}^{n_{\mathbf{u}}} \to \mathbb{R}^{n_{\mathbf{a}}}$ représente les équations qui permettent de déterminer implicitement sa partie algébrique $\mathbf{x}_{\mathbf{a}}(t) \in \mathbb{R}^{n_{\mathbf{a}}}$. N'importe quel modèle descripteur écrit sous

forme implicite (1.4) peut se mettre sous forme semi-explicite (1.6) en posant :

$$\begin{cases} \dot{\mathbf{x}}_{\mathbf{d}}(t) = \mathbf{x}_{\mathbf{a}}(t) \\ \qquad \mathbf{0} = \mathbf{F}\Big(\mathbf{x}_{\mathbf{a}}(t), \mathbf{x}_{\mathbf{d}}(t), \mathbf{u}(t)\Big) \ . \\ \mathbf{x}_{\mathbf{d}}(t_0) = \mathbf{x}_0 \, . \end{cases} \tag{1.7}$$

Néanmoins, ce changement d'écriture est rarement recommandé, car il pourrait complexifier inutilement les équations du modèle.

Une notion importante pour caractériser les modèles descripteur est celle d'indice différentiel [7]-[9]. L'indice, sous-entendu différentiel, d'un modèle descripteur est le nombre minimum $\nu_d \in \mathbb{N}$ de dérivations temporelles successives qu'il faudrait appliquer à ses équations algébriques pour les transformer en équations différentielles ordinaires. Par exemple, le fait de dériver une fois par rapport au temps les équations algébriques d'un modèle descripteur écrit sous forme semi-explicite (1.6) mène à l'expression :

$$\mathbf{0} = \frac{\partial \mathbf{f}_{\mathbf{a}}}{\partial \mathbf{x}_{\mathbf{d}}} \dot{\mathbf{x}}_{\mathbf{d}}(t) + \frac{\partial \mathbf{f}_{\mathbf{a}}}{\partial \mathbf{x}_{\mathbf{a}}} \dot{\mathbf{x}}_{\mathbf{a}}(t) + \frac{\partial \mathbf{f}_{\mathbf{a}}}{\partial \mathbf{u}} \dot{\mathbf{u}}(t) \, . \tag{1.8}$$

De ce fait, si la matrice Jacobienne $\partial \mathbf{f}_{\mathbf{a}} / \partial \mathbf{x}_{\mathbf{a}}$ est inversible, alors le modèle descripteur (1.6) est d'indice $\nu_d = 1$, car l'équation différentielle ordinaire qui régit le comportement de ses variables algébriques est simplement donnée par :

$$\dot{\mathbf{x}}_{\mathbf{a}}(t) = -\left[\frac{\partial \mathbf{f}_{\mathbf{a}}}{\partial \mathbf{x}_{\mathbf{a}}}\right]^{-1} \left(\frac{\partial \mathbf{f}_{\mathbf{a}}}{\partial \mathbf{x}_{\mathbf{d}}} \dot{\mathbf{x}}_{\mathbf{d}}(t) + \frac{\partial \mathbf{f}_{\mathbf{a}}}{\partial \mathbf{u}} \dot{\mathbf{u}}(t) \right) \coloneqq \boldsymbol{\zeta}_1 \Big(\mathbf{x}_{\mathbf{d}}(t), \mathbf{x}_{\mathbf{a}}(t), \mathbf{u}(t), \dot{\mathbf{u}}(t) \Big) \, . \tag{1.9}$$

Les modèles descripteur d'indice 1 sont donc très proches des modèles d'état qui, par définition, sont d'indice nul. En ce sens, l'indice donne une idée de la distance entre un modèle descripteur (1.6) et sa représentation d'état équivalente :

$$\begin{cases} \dot{\mathbf{x}}_{\mathbf{d}}(t) = \mathbf{f}_{\mathbf{d}}\Big(\mathbf{x}_{\mathbf{d}}(t), \mathbf{x}_{\mathbf{a}}(t), \mathbf{u}(t)\Big) \\ \dot{\mathbf{x}}_{\mathbf{a}}(t) = \boldsymbol{\zeta}_{\nu_d}\Big(\mathbf{x}_{\mathbf{d}}(t), \mathbf{x}_{\mathbf{a}}(t), \mathbf{u}(t), \ldots, \mathbf{u}^{(\nu_d)}(t)\Big) \\ \mathbf{x}_{\mathbf{d}}(t_0) = \mathbf{x}_{\mathbf{d},0}, \ \mathbf{x}_{\mathbf{a}}(t_0) = \mathbf{x}_{\mathbf{a},0} \, , \end{cases} \tag{1.10}$$

ainsi qu'une mesure de la singularité de ses équations algébriques. Par conséquent, plus l'indice d'un modèle est élevé, plus ce dernier sera compliqué à manipuler. En particulier, pour les modèles d'indice $\nu_d > 1$, la différentiation répétée des équations algébriques impose des contraintes

supplémentaires :

$$\begin{cases} \mathbf{0} = \dfrac{\mathrm{d}}{\mathrm{d}t}\mathbf{f_a}\Big(\mathbf{x_d}(t), \mathbf{x_a}(t), \mathbf{u}(t)\Big) \\ \quad \vdots \\ \mathbf{0} = \dfrac{\mathrm{d}^{\nu_d-1}}{\mathrm{d}t^{\nu_d-1}}\mathbf{f_a}\Big(\mathbf{x_d}(t), \mathbf{x_a}(t), \mathbf{u}(t)\Big)\,, \end{cases} \tag{1.11}$$

aux variables de leur représentation d'état équivalente (1.10). Ces contraintes, bien qu'invisibles dans le modèle d'origine (1.6), doivent malgré tout être respectées à chaque instant $t \geq t_0$ au même titre que les équations algébriques $\mathbf{0} = \mathbf{f_a}(\mathbf{x_d}(t), \mathbf{x_a}(t), \mathbf{u}(t))$. Cela implique notamment de trouver des conditions initiales $\mathbf{x_{d},0}$ et $\mathbf{x_{a},0}$ cohérentes, c'est-à-dire qui vérifient l'ensemble des contraintes du modèle à $t = t_0$. Or, quand $\nu_d = 1$, il suffit de déduire $\mathbf{x_{a},0}$ des équations algébriques $\mathbf{0} = \mathbf{f_a}(\mathbf{x_{d},0}, \mathbf{x_{a},0}, \mathbf{u}(t_0))$ pour obtenir des conditions initiales cohérentes, puisque l'état initial $\mathbf{x_{d},0}$ peut être choisi arbitrairement. Voilà pourquoi, en pratique, la majorité des schémas d'intégration numérique utilisés pour simuler les modèles descripteur se limitent à ceux dont l'indice vaut 1 [10], [11].

2.2 Propriétés des modèles singulièrement perturbés

Le fait de négliger certaines dynamiques d'un modèle d'état est une façon courante, en automatique, d'obtenir un modèle descripteur. Cette simplification, habituellement justifiée par une connaissance fine du comportement physique du système réel, peut être formalisée mathématiquement en s'appuyant sur la théorie des perturbations singulières [12]-[17]. Un modèle singulièrement perturbé est un modèle d'état de la forme :

$$\begin{cases} \dot{\mathbf{x}}_\mathbf{d}(t) = \mathbf{f_d}\Big(\mathbf{x_d}(t), \mathbf{x_a}(t), \mathbf{u}(t), \varepsilon\Big) \\ \varepsilon\dot{\mathbf{x}}_\mathbf{a}(t) = \mathbf{f_a}\Big(\mathbf{x_d}(t), \mathbf{x_a}(t), \mathbf{u}(t), \varepsilon\Big) \\ \mathbf{x_d}(t_0) = \mathbf{x_{d},0},\ \mathbf{x_a}(t_0) = \mathbf{x_{a},0}\,, \end{cases} \tag{1.12}$$

où $0 < \varepsilon \ll 1$ est un petit paramètre scalaire, de préférence sans dimension, pouvant être négligé. Étant donné que la dérivée temporelle des variables d'état $\mathbf{x_a}(t) \in \mathbb{R}^{n_\mathbf{a}}$ est proportionnelle à $1/\varepsilon \gg 1$, ces dernières ont tendance à évoluer bien plus rapidement que $\mathbf{x_d}(t) \in \mathbb{R}^{n_\mathbf{d}}$. Ainsi, $\mathbf{x_a}(t)$ correspond aux états rapides du modèle, alors que $\mathbf{x_d}(t)$ correspond à ses états lents. Lorsque $\varepsilon \to 0$, les équations différentielles ordinaires qui décrivent l'évolution des variables d'état rapides se transforment progressivement en équations algébriques :

$$\mathbf{0} = \mathbf{f_a}\Big(\mathbf{x_d^\infty}(t), \mathbf{x_a^\infty}(t), \mathbf{u}(t), 0\Big)\,, \tag{1.13}$$

où $\mathbf{x_d^\infty}(t)$ et $\mathbf{x_a^\infty}(t)$ sont les variables du modèle descripteur obtenu en posant $\varepsilon = 0$. D'après le théorème des fonctions implicites, si les équations algébriques (1.13) admettent au moins une

racine simple et que le modèle descripteur associé est d'indice 1, alors il est possible d'exprimer $\mathbf{x}_{\mathbf{a}}^{\infty}(t)$ en fonction des autres variables :

$$\mathbf{x}_{\mathbf{a}}^{\infty}(t) \coloneqq \boldsymbol{\varphi}\Big(\mathbf{x}_{\mathbf{d}}^{\infty}(t), \mathbf{u}(t)\Big)\,. \tag{1.14}$$

Le modèle d'état réduit qui en découle :

$$\dot{\mathbf{x}}_{\mathbf{d}}^{\infty}(t) = \mathbf{f}_{\mathbf{d}}\Big(\mathbf{x}_{\mathbf{d}}^{\infty}(t), \boldsymbol{\varphi}\Big(\mathbf{x}_{\mathbf{d}}^{\infty}(t), \mathbf{u}(t)\Big), \mathbf{u}(t), 0\Big),\ \mathbf{x}_{\mathbf{d}}^{\infty}(t_0) = \mathbf{x}_{\mathbf{d},0}\,, \tag{1.15}$$

est appelé modèle lent, ou « modèle quasi-statique », car il est obtenu en considérant que les états rapides du modèle singulièrement perturbé (1.12) atteignent instantanément leur régime permanent (1.14). Cependant, le fait de remplacer les états rapides par des variables algébriques signifie qu'il n'est plus possible de choisir librement leur valeur initiale. Par conséquent, l'approximation quasi-statique $\mathbf{x}_{\mathbf{a}}(t) \approx \mathbf{x}_{\mathbf{a}}^{\infty}(t)$ n'est pas valable à l'instant initial t_0, car il pourrait y avoir un écart important entre $\mathbf{x}_{\mathbf{a}}(t_0) = \mathbf{x}_{\mathbf{a},0}$ et $\mathbf{x}_{\mathbf{a}}^{\infty}(t_0) = \boldsymbol{\varphi}(\mathbf{x}_{\mathbf{d},0}, \mathbf{u}(t_0))$. Au mieux, $\mathbf{x}_{\mathbf{a}}(t)$ devrait commencer par se rapprocher de $\mathbf{x}_{\mathbf{a}}^{\infty}(t)$ jusqu'à ce que la relation $\mathbf{x}_{\mathbf{a}}(t) \approx \mathbf{x}_{\mathbf{a}}^{\infty}(t)$ devienne valide après un certain instant limite $t_\varepsilon > t_0$. En revanche, il est raisonnable de s'attendre à ce que $\mathbf{x}_{\mathbf{d}}^{\infty}(t)$ soit toujours une bonne approximation de $\mathbf{x}_{\mathbf{d}}(t)$, puisque leurs valeurs initiales $\mathbf{x}_{\mathbf{d}}(t_0) = \mathbf{x}_{\mathbf{d}}^{\infty}(t_0) = \mathbf{x}_{\mathbf{d},0}$ coïncident parfaitement. Afin de pouvoir analyser plus en détail le comportement transitoire des variables d'état au voisinage de t_0, le temps est étiré grâce au changement d'échelle $\tau \coloneqq (t - t_0)/\varepsilon$. En admettant que les variables d'état puissent être approchées uniformément à l'ordre 0 par un développement asymptotique composé de deux échelles de temps :

$$\mathbf{x}_{\mathbf{d}}(t) \coloneqq \mathbf{x}_{\mathbf{d}}^{0}(\tau) + \mathbf{x}_{\mathbf{d}}^{\infty}(t) + \underset{\varepsilon\to 0}{\mathcal{O}}(\varepsilon)\,, \quad \mathbf{x}_{\mathbf{a}}(t) \coloneqq \mathbf{x}_{\mathbf{a}}^{0}(\tau) + \mathbf{x}_{\mathbf{a}}^{\infty}(t) + \underset{\varepsilon\to 0}{\mathcal{O}}(\varepsilon)\,, \tag{1.16}$$

il est possible, à l'aide de la règle de dérivation en chaîne :

$$\frac{\mathrm{d}\mathbf{x}(\tau)}{\mathrm{d}t} = \frac{\mathrm{d}\mathbf{x}(\tau)}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}t} \ \text{ ou } \ \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}\tau} = \frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau},\ \text{ avec } \ \mathrm{d}t = \varepsilon\,\mathrm{d}\tau\,, \tag{1.17}$$

de réécrire le modèle singulièrement perturbé (1.12) comme suit :

$$\begin{cases} \dfrac{\mathrm{d}\mathbf{x}_{\mathbf{d}}^{0}(\tau)}{\mathrm{d}\tau} + \varepsilon\dot{\mathbf{x}}_{\mathbf{d}}^{\infty}(t) = \varepsilon\mathbf{f}_{\mathbf{d}}\Big(\mathbf{x}_{\mathbf{d}}^{0}(\tau) + \mathbf{x}_{\mathbf{d}}^{\infty}(t), \mathbf{x}_{\mathbf{a}}^{0}(\tau) + \mathbf{x}_{\mathbf{a}}^{\infty}(t), \mathbf{u}(t), \varepsilon\Big) \\ \dfrac{\mathrm{d}\mathbf{x}_{\mathbf{a}}^{0}(\tau)}{\mathrm{d}\tau} + \varepsilon\dot{\mathbf{x}}_{\mathbf{a}}^{\infty}(t) = \mathbf{f}_{\mathbf{a}}\Big(\mathbf{x}_{\mathbf{d}}^{0}(\tau) + \mathbf{x}_{\mathbf{d}}^{\infty}(t), \mathbf{x}_{\mathbf{a}}^{0}(\tau) + \mathbf{x}_{\mathbf{a}}^{\infty}(t), \mathbf{u}(t), \varepsilon\Big) \\ \mathbf{x}_{\mathbf{d}}^{0}(0) + \mathbf{x}_{\mathbf{d}}^{\infty}(t_0) = \mathbf{x}_{\mathbf{d},0},\ \mathbf{x}_{\mathbf{a}}^{0}(0) + \mathbf{x}_{\mathbf{a}}^{\infty}(t_0) = \mathbf{x}_{\mathbf{a},0}\,. \end{cases} \tag{1.18}$$

où $t \coloneqq t_0 + \varepsilon\tau$. Cette fois, lorsque $\varepsilon \to 0$, la partie rapide $\mathbf{x}_\mathbf{d}^0(\tau)$ des états lents $\mathbf{x}_\mathbf{d}(t)$ devient nulle :

$$\frac{\mathrm{d}\mathbf{x}_\mathbf{d}^0(\tau)}{\mathrm{d}\tau} = \mathbf{0},\ \mathbf{x}_\mathbf{d}^0(0) = \mathbf{0} \ \Rightarrow\ \mathbf{x}_\mathbf{d}^0(\tau) = \mathbf{0} \text{ pour tout } \tau \geq 0\,, \tag{1.19}$$

ce qui est parfaitement logique, puisque la réponse temporelle de $\mathbf{x}_\mathbf{d}(t)$ dépend principalement de celle de sa partie lente $\mathbf{x}_\mathbf{d}^\infty(t)$. De ce fait, l'évolution de la partie rapide $\mathbf{x}_\mathbf{a}^0(\tau)$ des états rapides $\mathbf{x}_\mathbf{a}(t)$ est décrite par le modèle d'état réduit :

$$\frac{\mathrm{d}\mathbf{x}_\mathbf{a}^0(\tau)}{\mathrm{d}\tau} = \mathbf{f}_\mathbf{a}\Big(\mathbf{x}_{\mathbf{d},0}, \mathbf{x}_\mathbf{a}^0(\tau) + \boldsymbol{\varphi}\Big(\mathbf{x}_{\mathbf{d},0}, \mathbf{u}(t_0)\Big), \mathbf{u}(t_0), 0\Big),\ \mathbf{x}_\mathbf{a}^0(0) = \mathbf{x}_{\mathbf{a},0} - \boldsymbol{\varphi}\Big(\mathbf{x}_{\mathbf{d},0}, \mathbf{u}(t_0)\Big)\,. \tag{1.20}$$

Ce modèle est communément appelé « modèle en couche limite », car il n'est valable que sur un petit intervalle de temps $[t_0, t_\varepsilon]$ vérifiant $t_\varepsilon - t_0 = \mathcal{O}_{\varepsilon\to 0}(\varepsilon)$. En effet, comme $\varepsilon \to 0$, le temps étiré τ aura tendance à diverger très rapidement vers l'infini si t s'écarte trop de t_0. Ainsi, les termes $\mathbf{x}_\mathbf{d}^\infty(t)$ et $\mathbf{x}_\mathbf{a}^\infty(t)$, qui évoluent déjà lentement dans l'échelle de temps standard, ont l'air presque immobiles dans la nouvelle échelle de temps étirée. En particulier, ces derniers restent figés à leurs valeurs initiales quand $\varepsilon = 0$. Sous certaines hypothèses, il peut être démontré que $\mathbf{x}_\mathbf{a}^0(\tau)$ décroît rapidement vers zéro au cours de l'intervalle de temps $[t_0, t_\varepsilon]$ et que, passée cette période, la réponse temporelle des états rapides est similaire à celle de leur partie lente $\mathbf{x}_\mathbf{a}^\infty(t)$. Dans ce cas, si ε est suffisamment petit, le comportement du modèle singulièrement perturbé (1.12) peut être d'abord approché par le modèle en couche limite (1.20) en régime transitoire, puis par le modèle quasi-statique (1.15) en régime permanent.

2.3 Formulation du problème de commande optimale

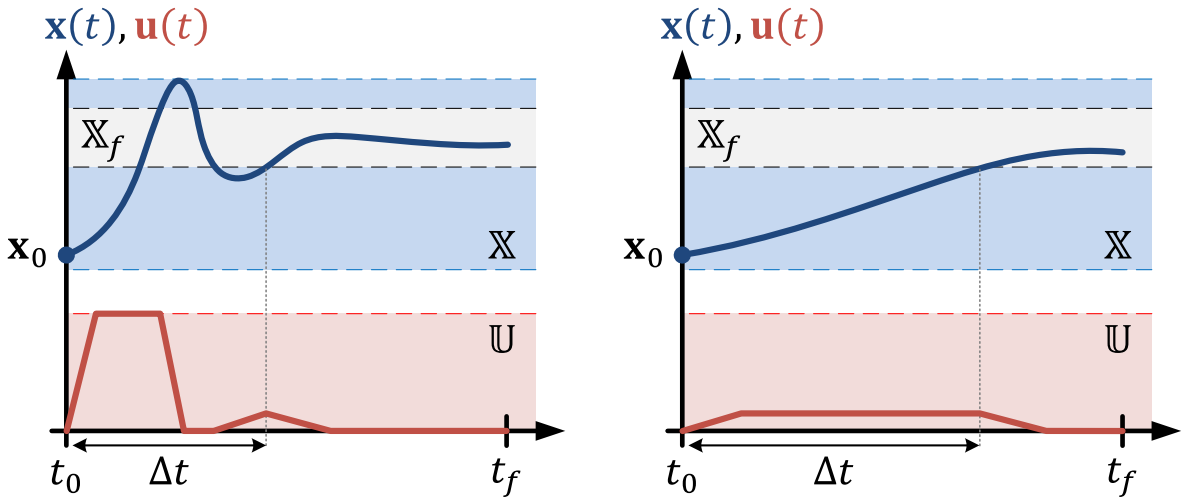

FIGURE 1.1 – Deux stratégies de commande optimales : l'une en temps, l'autre en énergie.

La prochaine étape, après avoir obtenu un modèle mathématique du système à réguler, est de formuler le problème de commande optimale. L'objectif d'un problème de commande optimale est de trouver une loi de commande qui, sur un intervalle de temps donné, dirige l'état du modèle le long d'une trajectoire minimisant un critère de performance :

$$\begin{aligned} \min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \quad & J_{[t_0,t_f]}\Big(\mathbf{x}(\cdot),\mathbf{u}(\cdot)\Big) \coloneqq V_f\Big(\mathbf{x}(t_f)\Big) + \int_{t_0}^{t_f} L\Big(\mathbf{x}(\tau),\mathbf{u}(\tau)\Big)\,\mathrm{d}\tau \\ \text{sous contraintes} : & \forall\,\tau \in [t_0,t_f]\,,\ \mathbf{F}\Big(\dot{\mathbf{x}}(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)\Big) = \mathbf{0},\ \mathbf{x}(t_0) = \mathbf{x}_0\,. \end{aligned} \tag{1.21}$$

Le premier terme $V_f : \mathbb{R}^{n_\mathbf{x}} \to \mathbb{R}_{\geq 0}$, appelé coût terminal, pénalise uniquement l'état à l'instant final t_f, tandis que le second terme $L : \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}} \to \mathbb{R}_{\geq 0}$, appelé coût de fonctionnement, pénalise la commande et l'état entre t_0 et t_f. Le coût de fonctionnement représente l'effort nécessaire (écart par rapport à une référence, énergie dépensée, temps écoulé, etc.) pour passer de l'état initial à l'état final. Cela suppose, bien évidemment, qu'il existe un signal de commande capable de déplacer l'état du modèle de $\mathbf{x}(t_0)$ à $\mathbf{x}(t_f)$ dans le temps imparti. Le coût terminal, quant à lui, sert généralement à guider l'état final vers une région particulière de l'espace d'état. Ce terme disparaît de la fonction coût lorsque le problème de commande optimale est à horizon infini :

$$\begin{aligned} \min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \quad & J_\infty\Big(\mathbf{x}(\cdot),\mathbf{u}(\cdot)\Big) \coloneqq \int_{t_0}^{+\infty} L\Big(\mathbf{x}(\tau),\mathbf{u}(\tau)\Big)\,\mathrm{d}\tau \\ \text{sous contraintes} : & \forall\,\tau \in [t_0,+\infty)\,,\ \mathbf{F}\Big(\dot{\mathbf{x}}(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)\Big) = \mathbf{0},\ \mathbf{x}(t_0) = \mathbf{x}_0\,. \end{aligned} \tag{1.22}$$

L'intérêt de choisir une fonction coût avec un horizon de temps infini est que, si le modèle admet une position d'équilibre $(\mathbf{x}_{\text{eq}},\mathbf{u}_{\text{eq}}) \in \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}}$ et que le coût de fonctionnement vérifie $L(\mathbf{x}_{\text{eq}},\mathbf{u}_{\text{eq}}) = 0$ et $L(\mathbf{x},\mathbf{u}) > 0$ pour tout $(\mathbf{x},\mathbf{u}) \neq (\mathbf{x}_{\text{eq}},\mathbf{u}_{\text{eq}})$, alors la loi de commande optimale qui en découle stabilise asymptotiquement le modèle en boucle fermée. En effet, dans ces conditions, le seul moyen de faire converger la fonction coût est que le modèle atteigne sa position d'équilibre et y reste par la suite. Toutefois, encore faut-il réussir à trouver une solution au problème à horizon infini (1.22), ce qui devient rapidement impossible lorsque le modèle est non-linéaire et/ou soumis à de nombreuses contraintes.

En plus des contraintes de stabilité et d'évolution du modèle, d'autres peuvent être ajoutées au problème de commande optimale pour tenir compte, par exemple, du fait que le signal de commande est limité par les caractéristiques physiques des actionneurs ou que certaines variables d'état doivent rester dans un domaine de fonctionnement précis. Quand les contraintes sur l'état et sur la commande sont bien séparées, l'ensemble $\mathbb{X} \subseteq \mathbb{R}^{n_\mathbf{x}}$ représente celles qui s'appliquent sur l'état et l'ensemble $\mathbb{U} \subset \mathbb{R}^{n_\mathbf{u}}$ représente celles qui s'appliquent sur la commande. En revanche, quand elles sont liées, les contraintes qui s'appliquent sur

l'état et sur la commande doivent être représentées d'un seul tenant par l'ensemble mixte $\mathbb{Y} \coloneqq \{(\mathbf{x},\mathbf{u}) \in \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}} \mid \mathbf{x} \in \mathbb{X} \text{ et } \mathbf{u} \in \mathbb{U}(\mathbf{x})\}$, avec $\mathbb{X} \coloneqq \{\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}} \mid \mathbb{U}(\mathbf{x}) \neq \emptyset\}$ et $\mathbb{U}(\mathbf{x}) \coloneqq \{\mathbf{u} \in \mathbb{R}^{n_\mathbf{u}} \mid (\mathbf{x},\mathbf{u}) \in \mathbb{Y}\}$. L'ensemble mixte se réduit simplement au produit cartésien $\mathbb{Y} = \mathbb{X} \times \mathbb{U}$ quand les contraintes sont séparées. Enfin, lorsque cela a du sens, les contraintes qui s'appliquent sur l'état final sont représentées par l'ensemble $\mathbb{X}_f \subseteq \mathbb{X}$. Avec toutes ces contraintes supplémentaires, le problème de commande optimale (1.21) devient :

$$\begin{aligned} \min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot)} \quad & J_{[t_0,t_f]}\Big(\mathbf{x}(\cdot),\mathbf{u}(\cdot)\Big) \coloneqq V_f\Big(\mathbf{x}(t_f)\Big) + \int_{t_0}^{t_f} L\Big(\mathbf{x}(\tau),\mathbf{u}(\tau)\Big)\,\mathrm{d}\tau \\ \text{sous contraintes : } & \begin{cases} \forall\,\tau \in [t_0,t_f]\,,\ \mathbf{F}\Big(\dot{\mathbf{x}}(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)\Big) = \mathbf{0},\ \mathbf{x}(t_0) = \mathbf{x}_0 \\ \forall\,\tau \in [t_0,t_f)\,,\ \Big(\mathbf{x}(\tau),\mathbf{u}(\tau)\Big) \in \mathbb{Y} \\ \mathbf{x}(t_f) \in \mathbb{X}_f\,. \end{cases} \end{aligned} \tag{1.23}$$

Étant donné que la trajectoire d'état $\mathbf{x} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_\mathbf{x}}$ est entièrement déterminée par le signal de commande $\mathbf{u} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_\mathbf{u}}$ et par les contraintes d'évolution du modèle, le problème de commande optimale (1.23) peut également s'écrire :

$$\begin{aligned} \min_{\mathbf{u}(\cdot)} \quad & J_{[t_0,t_f]}\Big(\mathbf{x}_0,\mathbf{u}(\cdot)\Big) \coloneqq V_f\Big(\mathbf{x}_\mathbf{u}(t_f;\mathbf{x}_0)\Big) + \int_{t_0}^{t_f} L\Big(\mathbf{x}_\mathbf{u}(\tau;\mathbf{x}_0),\mathbf{u}(\tau)\Big)\,\mathrm{d}\tau \\ & \text{sous contraintes : } \forall\,\tau \in [t_0,t_f)\,,\ \Big(\mathbf{x}_\mathbf{u}(\tau;\mathbf{x}_0),\mathbf{u}(\tau)\Big) \in \mathbb{Y} \text{ et } \mathbf{x}_\mathbf{u}(t_f;\mathbf{x}_0) \in \mathbb{X}_f \\ & \text{où, pour tout } \tau \in [t_0,t_f]\,,\ \mathbf{x}_\mathbf{u}(\tau;\mathbf{x}_0) \text{ est la solution de : } \begin{cases} \mathbf{F}\Big(\dot{\mathbf{x}}(\tau),\mathbf{x}(\tau),\mathbf{u}(\tau)\Big) = \mathbf{0} \\ \mathbf{x}(t_0) = \mathbf{x}_0\,. \end{cases} \end{aligned} \tag{1.24}$$

La seule différence entre les formulations (1.23) et (1.24) vient de la manière dont la trajectoire d'état et les contraintes d'évolution sont intégrées au problème. Dans la première formulation, la trajectoire d'état est vue comme une variable de décision qui doit respecter les contraintes d'évolution du modèle contrôlé par le signal de commande. Cette formulation est plutôt privilégiée pour traiter la question de la résolution numérique du problème, car les contraintes d'évolution y apparaissent explicitement. Dans la seconde formulation, la trajectoire d'état n'est plus une variable de décision du problème, mais résulte implicitement de l'action de la commande sur le modèle. Cette formulation est mieux adaptée pour analyser les propriétés de la loi de commande, car les contraintes d'évolution sont prises en compte en dehors du problème. À noter que le problème de commande optimale est paramétrique en la valeur initiale de l'état, puisque les contraintes d'évolution de (1.23) et la fonction coût de (1.24) dépendent toutes les deux du

paramètre $\mathbf{x}_0$. Il en est donc de même pour la fonction coût optimale :

$$J^*_{[t_0,t_f]}(\mathbf{x}_0) \coloneqq \inf_{\mathbf{u}(\cdot)} J_{[t_0,t_f]}\Big(\mathbf{x}_0, \mathbf{u}(\cdot)\Big)\,. \tag{1.25}$$

et pour la loi de commande optimale associée :

$$\mathbf{u}^*_{[t_0,t_f]}(\tau\,;\mathbf{x}_0) \coloneqq \operatorname*{argmin}_{\mathbf{u}(\cdot)} J_{[t_0,t_f]}\Big(\mathbf{x}_0, \mathbf{u}(\cdot)\Big),\ \text{où}\ \tau \in [t_0, t_f)\ . \tag{1.26}$$

Cependant, la nature de cette dernière (boucle ouverte ou boucle fermée, temps discret ou temps continu) change en fonction de la méthode employée pour résoudre le problème. Pour rappel, une loi de commande optimale en boucle ouverte ne dépend que d'un état initial spécifique, alors qu'une loi de commande optimale en boucle fermée dépend d'un ensemble d'états initiaux.

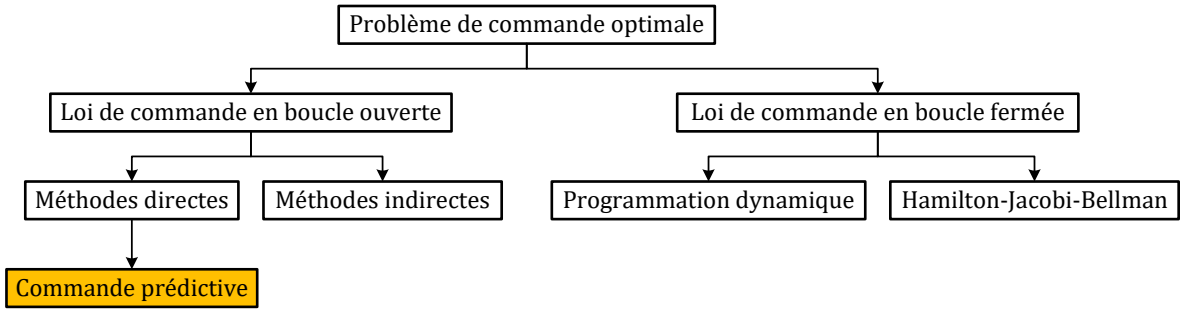

FIGURE 1.2 – Méthodes utilisées pour résoudre un problème de commande optimale.

Actuellement, trois types de méthodes existent pour résoudre un problème de commande optimale [8], [9], [18], [19] : 1) la programmation dynamique, qui s'appuie sur le principe d'optimalité de Bellman pour résoudre récursivement le problème en le décomposant en plusieurs sous-problèmes ; 2) les méthodes indirectes, qui se basent sur le principe du maximum de Pontryagin pour trouver une solution potentielle au problème à partir des conditions nécessaires d'optimalité ; 3) les méthodes directes, qui discrétisent le signal de commande et la trajectoire d'état pour transcrire le problème de commande optimale en un problème d'optimisation numérique. La programmation dynamique est la seule méthode qui permette d'obtenir une loi de commande en boucle fermée minimisant globalement la fonction coût. Néanmoins, cette méthode est rarement employée en pratique, car elle nécessite soit de résoudre une équation aux dérivées partielles potentiellement non-linéaire, soit de stocker et de tabuler le résultat de chaque sous-problème en fonction de toutes les valeurs possibles du vecteur d'état. Les méthodes indirectes, d'autre part, sont plus faciles à mettre en place, puisque les conditions nécessaires d'optimalité se ramènent à un problème aux limites constitué d'équations différentielles ordinaires. En revanche, la loi de commande qui en résulte est exprimée en boucle ouverte, et son optimalité doit

être vérifiée a posteriori. De plus, le problème aux limites a tendance à être très mal conditionné, et devient difficile à résoudre en présence de contraintes sur l'état. Enfin, les conditions nécessaires d'optimalité doivent être redéterminées analytiquement chaque fois que le problème de commande optimale est modifié. Face à tous ces inconvénients, les méthodes directes choisissent de renoncer à résoudre les conditions d'optimalité pour s'attaquer de front au problème de commande optimale. Ce dernier est alors transformé en un problème d'optimisation numérique, qui pourra ensuite être résolu par un solveur dédié. Pour ce faire, les variables de décision, la fonction coût et les contraintes du problème de commande optimale sont discrétisées à l'aide d'un nombre fini de paramètres. La perte de précision liée à la discrétisation est largement compensée par le fait que cette approche est beaucoup plus flexible et intuitive que les deux autres. Cela explique en partie pourquoi les méthodes directes sont les plus répandues de nos jours.

2.4 Définition de l'algorithme de commande prédictive

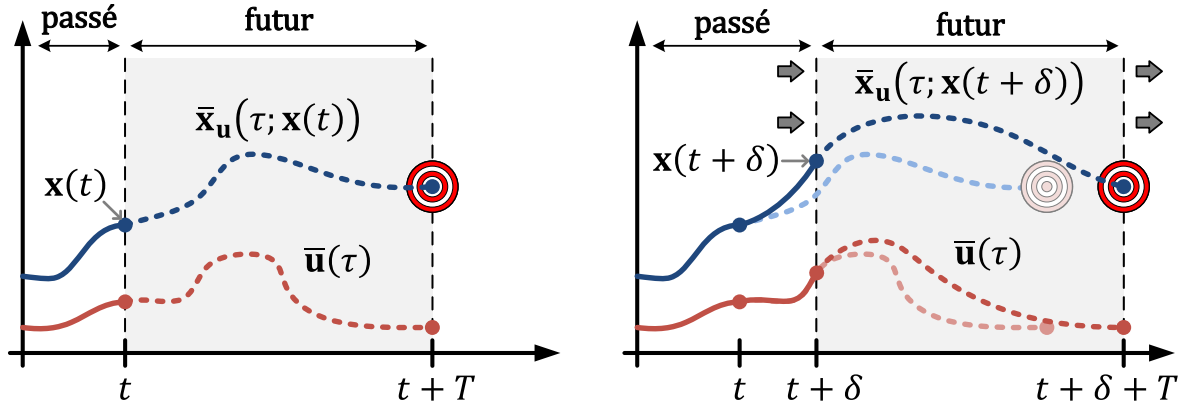

FIGURE 1.3 – Principe de base de la commande prédictive.

Le principal inconvénient des méthodes directes est que la loi de commande optimale renvoyée par le problème d'optimisation s'exprime en boucle ouverte. Celle-ci ne peut donc pas être utilisée trop longtemps pour contrôler le système réel, puisque cela reviendrait à supposer qu'il se comporte exactement comme le modèle. En revanche, de nombreux outils numériques sont disponibles pour résoudre efficacement le problème d'optimisation [9], [18], [20]-[24]. Par conséquent, l'idée de la commande prédictive [25]-[27] est de générer une loi de commande en boucle fermée en recalculant régulièrement, à différents instants, la loi de commande optimale en boucle ouverte. Plus précisément, il s'agit de résoudre périodiquement et en temps réel le problème de commande optimale à horizon fini (1.24) par une méthode directe, en prenant soin de recaler à chaque fois l'état initial du modèle sur celui du système réel. La loi de commande prédictive profite alors d'autant plus fréquemment de l'information fournie par les mesures que l'algorithme

est relancé rapidement. Ainsi, pour éviter que le système reste en boucle ouverte pendant une période prolongée, seule une partie de la loi de commande optimale est véritablement envoyée entre deux itérations de l'algorithme. Le système ne peut toutefois pas être constamment en boucle fermée, car la vitesse d'exécution de l'algorithme est limitée par le temps de résolution du problème d'optimisation. Dans sa version standard, l'algorithme de commande prédictive consiste à :

Algorithme de commande prédictive standard

1) Mesurer l'état $\mathbf{x}(t)$ du système réel à l'instant courant $t \geq t_0$.
2) Calculer la loi de commande optimale en boucle ouverte $\overline{\mathbf{u}}^*_{\mathcal{T}}(\cdot\,;\mathbf{x}(t))$ en résolvant, par une méthode directe, le problème de commande optimale à horizon fini :

$$\begin{aligned}
&\min_{\overline{\mathbf{u}}(\cdot)} \quad J_{\mathcal{T}}\Big(\mathbf{x}(t),\overline{\mathbf{u}}(\cdot)\Big) \coloneqq V_f\Big(\overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T};\mathbf{x}(t))\Big) + \int_0^{\mathcal{T}} L\Big(\overline{\mathbf{x}}_{\mathbf{u}}(\tau;\mathbf{x}(t)),\overline{\mathbf{u}}(\tau)\Big)\,\mathrm{d}\tau \\
&\text{sous contraintes : } \forall\,\tau\in[0,\mathcal{T})\,,\ \Big(\overline{\mathbf{x}}_{\mathbf{u}}(\tau;\mathbf{x}(t)),\overline{\mathbf{u}}(\tau)\Big)\in\mathbb{Y} \text{ et } \overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T};\mathbf{x}(t))\in\mathbb{X}_f \\
&\text{où, pour tout } \tau\in[0,\mathcal{T}]\,,\ \overline{\mathbf{x}}_{\mathbf{u}}(\tau;\mathbf{x}(t)) \text{ est la solution de : } \begin{cases} \mathbf{F}\Big(\dot{\overline{\mathbf{x}}}(\tau),\overline{\mathbf{x}}(\tau),\overline{\mathbf{u}}(\tau)\Big)=\mathbf{0} \\ \overline{\mathbf{x}}(0)=\mathbf{x}(t)\,. \end{cases}
\end{aligned} \tag{1.27}$$

   Le trait suscrit permet ici de distinguer les variables utilisées par le modèle interne du contrôleur de celles mesurées sur le système réel.
3) Définir la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(t+\tau;\mathbf{x}(t)) \coloneqq \overline{\mathbf{u}}^*_{\mathcal{T}}(\tau;\mathbf{x}(t))$ pour tout $\tau\in[0,\delta)$, où $0<\delta\leq\mathcal{T}$ est la période au bout de laquelle est relancé l'algorithme.
4) Appliquer la loi de commande prédictive au système réel de t à $t+\delta$.
5) Relancer l'algorithme à l'instant $t_{\text{next}} \coloneqq t+\delta$.

2.4.1 Réalisabilité récursive du problème de commande optimale

Un des atouts de la commande prédictive est de pouvoir générer une loi de commande en boucle fermée prenant directement en compte des contraintes sur l'état et sur la commande. Ces contraintes doivent néanmoins être ajoutées avec précaution et parcimonie, car elles peuvent rendre le problème de commande optimale difficile à résoudre, voire irréalisable. Il est donc important de bien poser le problème pour s'assurer que l'algorithme ne se retrouve pas sans solution au moment d'envoyer la commande au système réel. Pour un état initial $\mathbf{x}_0\in\mathbb{X}$ donné,

l'ensemble des lois de commande admissibles est défini par :

$$\mathcal{U}_{\mathcal{T}}(\mathbf{x}_0) \coloneqq \left\{ \overline{\mathbf{u}} : \mathbb{R}_{\geq 0} \to \mathbb{U}(\mathbf{x}_0) \mid \forall\, \tau \in [0, \mathcal{T}) \,, \left(\overline{\mathbf{x}}_{\mathbf{u}}(\tau; \mathbf{x}_0), \overline{\mathbf{u}}(\tau) \right) \in \mathbb{Y} \text{ et } \overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T}; \mathbf{x}_0) \in \mathbb{X}_f \right\} . \quad (1.28)$$

Les loi de commandes admissibles sont celles qui permettent de respecter toutes les contraintes. Pour que le problème de commande optimale admette potentiellement une solution en partant d'un état initial $\mathbf{x}_0 \in \mathbb{X}$, il est nécessaire que l'ensemble d'admissibilité $\mathcal{U}_{\mathcal{T}}(\mathbf{x}_0)$ soit non vide. De ce fait, l'ensemble de réalisabilité du problème est donné par :

$$\mathcal{X}_{\mathcal{T}} \coloneqq \{\, \mathbf{x}_0 \in \mathbb{X} \mid \mathcal{U}_{\mathcal{T}}(\mathbf{x}_0) \neq \emptyset \,\} \,. \quad (1.29)$$

Cet ensemble correspond aux états initiaux pour lesquels il existe au moins une loi de commande admissible. Autrement dit, il s'agit des états pouvant être dirigés sur l'intervalle de temps $[0, \mathcal{T}]$ de façon à respecter toutes les contraintes. Cependant, même si $\mathbf{x}_0 \in \mathcal{X}_{\mathcal{T}}$, rien ne garantit encore que le minimum de la fonction coût puisse être atteint avec une loi de commande admissible. Pour que cela soit possible, les hypothèses suivantes doivent être satisfaites :

Hypothèse 2.1. (Existence d'une solution au problème de commande optimale).

H1) Le modèle admet une position d'équilibre $(\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}})$ telle que $\mathbf{F}(\mathbf{0}, \mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}}) = \mathbf{0}$.

H2) Les points $(\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}})$ et $\mathbf{x}_{\text{eq}}$ appartiennent respectivement à l'intérieur des ensembles $\mathbb{Y}$ et $\mathbb{X}_f$. De plus, $\mathbb{Y}$ est fermé et $\mathbb{X}_f$ est compact. Enfin, $\mathbb{U}(\mathbf{x})$ est compact pour tout $\mathbf{x} \in \mathbb{X}$.

H3) Le coût de fonctionnement $L(\cdot)$ est continu sur $\mathbb{Y}$ et vérifie $L(\mathbf{x}, \mathbf{u}) > 0$ pour tout $(\mathbf{x}, \mathbf{u}) \neq (\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}})$ et $L(\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}}) = 0$ sinon. De même, le coût terminal $V_f(\cdot)$ est continûment différentiable sur $\mathbb{X}_f$ et vérifie $V_f(\mathbf{x}) > 0$ pour tout $\mathbf{x} \neq \mathbf{x}_{\text{eq}}$ et $V_f(\mathbf{x}_{\text{eq}}) = 0$ sinon.

En effet, dans ces conditions, une solution au problème de commande optimale existe pour tout $\mathbf{x}_0 \in \mathcal{X}_{\mathcal{T}}$, car la fonction coût $J_{\mathcal{T}}(\mathbf{x}_0, \cdot)$ est alors continue sur l'ensemble compact $\mathcal{U}_{\mathcal{T}}(\mathbf{x}_0)$. Il faudrait donc idéalement que les états mesurés à chaque itération appartiennent tous à l'ensemble $\mathcal{X}_{\mathcal{T}}$ pour être sûr que l'algorithme fonctionne toujours correctement. Pour cela, la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot)$ doit être capable de contrôler système réel de façon que son état ne sorte pas de l'ensemble de réalisabilité du problème. En d'autres termes, il faut que $\mathcal{X}_{\mathcal{T}}$ soit un ensemble positivement invariant [28], [29] pour le système en boucle fermée :

$$\forall\, \tau \geq 0,\ \mathbf{x}_0 \in \mathcal{X}_{\mathcal{T}} \ \Rightarrow\ \mathbf{x}_{\boldsymbol{\mu}}(\tau; \mathbf{x}_0) \in \mathcal{X}_{\mathcal{T}} \,, \quad (1.30)$$

où $\mathbf{x}_{\boldsymbol{\mu}}(\cdot; \mathbf{x}_0)$ est l'état obtenu en contrôlant le système réel avec la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot; \mathbf{x}_0)$ calculée en partant de $\mathbf{x}_0$. Dans ce cas, le problème de commande optimale est dit réalisable récursivement, car il suffit que celui-ci admette une solution au démarrage de l'algorithme pour que cela reste vrai aux itérations suivantes.

2.4.2 Stabilité en boucle fermée de loi de commande prédictive

Une autre point important est de veiller à ce que la loi de commande prédictive ne déstabilise pas le système réel. Rigoureusement parlant, il est impossible de vérifier strictement que le système réel est stable en boucle fermée, puisqu'il n'existe aucun modèle capable de le décrire parfaitement. La loi de commande devra donc impérativement être testée dans un environnement réaliste, comme un simulateur pleine échelle ou un banc d'essai, avant d'être installée sur site. Cette phase de validation finale, habituellement très coûteuse en temps et en ressources, n'arrive qu'après avoir conçu et réglé la loi de commande en bureau d'étude. Pour limiter au maximum le coût et la complexité du processus de validation, il est essentiel de se préoccuper de la stabilité du système bouclé dès les premières étapes de conception. L'enjeu, pendant la phase de conception, est de régler le contrôleur de sorte à obtenir des garanties de stabilité avec un modèle du système. Ce modèle de conception doit rester assez simple pour réussir à démontrer mathématiquement sa stabilité en boucle fermée, tout en étant suffisamment représentatif du comportement du système réel pour que cette preuve soit pertinente. Le besoin de trouver un modèle bien équilibré, remplissant simultanément ces deux conditions contradictoires, est encore plus prononcé en commande prédictive, car celui-ci doit retourner des prédictions fiables sans allonger excessivement le temps calcul de la loi de commande. Deux approches différentes peuvent être employées pour assurer la stabilité du système bouclé en l'absence d'erreur de modélisation et de perturbation [30]-[32]. Leur objectif commun est de chercher à montrer que la fonction coût optimale :

$$J^*_{\mathcal{T}}(\mathbf{x}_0) \coloneqq \inf_{\mathbf{u}(\cdot)\,\in\,\mathcal{U}_{\mathcal{T}}(\mathbf{x}_0)} J_{\mathcal{T}}\Big(\mathbf{x}_0, \mathbf{u}(\cdot)\Big),\ \text{avec}\ \mathbf{x}_0 \in \mathcal{X}_{\mathcal{T}}\,, \tag{1.31}$$

est une fonction de Lyapunov pour le système nominal en boucle fermée. Cela implique notamment de prouver qu'elle décroît strictement d'une itération à l'autre de l'algorithme :

$$\forall\, \delta \in (0, \mathcal{T}],\ \forall\, \mathbf{x}_0 \in \mathcal{X}_{\mathcal{T}},\ J^*_{\mathcal{T}}\Big(\mathbf{x}_{\boldsymbol{\mu}}(\delta; \mathbf{x}_0)\Big) < J^*_{\mathcal{T}}(\mathbf{x}_0)\,. \tag{1.32}$$

La première approche y parvient en se servant du coût terminal $V_f(\cdot)$ et/ou de l'ensemble des contraintes finales $\mathbb{X}_f$ [33]-[36]. Dans ce contexte, le moyen le plus simple de stabiliser le système nominal est de le forcer à atteindre sa position d'équilibre $(\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}})$ à la fin de l'horizon de prédiction en ajoutant une contrainte d'égalité finale [37]-[39] :

$$\overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T}; \mathbf{x}_0) = \mathbf{x}_{\text{eq}}\,, \tag{1.33}$$

au problème de commande optimale (1.27). La loi de commande prédictive stabilise alors asymptotiquement le système nominal en boucle fermée et le problème de commande optimale devient réalisable récursivement à partir du moment où les hypothèses (2.1) sont respectées. Toutefois, cette contrainte d'égalité finale peut également rendre le problème irréalisable si jamais

l'horizon de prédiction choisi est plus court que le temps requis pour atteindre l'équilibre. Une façon d'assouplir cette contrainte d'égalité est de la remplacer par une contrainte d'inégalité contractante [40], [41] :

$$\forall\, \mathbf{x}_0 \in \mathcal{X}_{\mathcal{T}},\ W\Big(\overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T};\mathbf{x}_0) - \mathbf{x}_{\text{eq}}\Big) \leq \gamma W(\mathbf{x}_0 - \mathbf{x}_{\text{eq}}),\ \text{avec } \gamma \in (0,1)\,, \tag{1.34}$$

où $W(\cdot)$ est une fonction définie positive sur $\mathbb{R}^{n_{\mathbf{x}}}$. Néanmoins, ce type de contrainte nécessite de modifier significativement l'algorithme de commande prédictive pour stabiliser le système nominal en boucle fermée, puisque la contraction de la fonction $W(\cdot)$ ne survient qu'à la fin de l'horizon de prédiction. En effet, même si la contrainte contractante (1.34) est vérifiée, rien n'empêche d'avoir $W(\overline{\mathbf{x}}_{\mathbf{u}}(\delta;\mathbf{x}_0) - \mathbf{x}_{\text{eq}}) > \gamma W(\mathbf{x}_0 - \mathbf{x}_{\text{eq}})$. De ce fait, relancer naïvement l'algorithme toutes les δ secondes ne permet pas de garantir que la fonction $W(\cdot)$ se contractera véritablement T secondes plus tard. Pour que cela soit le cas, il faudrait soit appliquer la totalité de la loi de commande optimale en boucle ouverte au système nominal, soit mémoriser le niveau de contraction à atteindre $\gamma W(\mathbf{x}_0 - \mathbf{x}_{\text{eq}})$ d'une itération à l'autre de l'algorithme jusqu'à ce que la fonction $W(\cdot)$ se contracte réellement, soit considérer l'horizon de prédiction T comme une variable de décision du problème. Étant donné qu'aucune de ces solutions n'est vraiment compatible avec l'algorithme standard de commande prédictive, la contrainte d'égalité finale (1.33) est plutôt remplacée par un ensemble $\mathbb{X}_f$ de contraintes finales auquel est associé un coût terminal $V_f(\cdot)$. Ces deux paramètres doivent être choisis de manière à respecter les hypothèses de base (2.1) ainsi que les hypothèses suivantes :

Hypothèse 2.2. (Stabilité nominale avec coût terminal et contraintes finales).

H1) Pour tout $\mathbf{x}_f \in \mathbb{X}_f$, il existe une loi de commande $\mathbf{u}_f : [0,\delta] \to \mathbb{U}(\mathbf{x}_f)$ telle que :

$$\forall\, \tau \in [0,\delta],\ \begin{cases} \mathbf{x}_{\mathbf{u}_f}(\tau;\mathbf{x}_f) \in \mathbb{X}_f \\ \dfrac{\mathrm{d}V_f}{\mathrm{d}t}\Big(\mathbf{x}_{\mathbf{u}_f}(\tau;\mathbf{x}_f)\Big) + L\Big(\mathbf{x}_{\mathbf{u}_f}(\tau;\mathbf{x}_f), \mathbf{u}_f(\tau)\Big) \leq 0 \end{cases} \tag{1.35}$$

H2) Il existe deux fonctions $\alpha_f : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ et $\alpha_L : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ de classe $\mathcal{K}_\infty$ telles que :

$$\begin{aligned} &\forall\, \mathbf{x}_f \in \mathbb{X}_f,\ V_f(\mathbf{x}_f) \leq \alpha_f(\|\mathbf{x} - \mathbf{x}_{\text{eq}}\|) \\ &\forall\, \mathbf{x} \in \mathbb{X},\ \forall\, \mathbf{u} \in \mathbb{U}(\mathbf{x}),\ L(\mathbf{x},\mathbf{u}) \geq \alpha_L(\|\mathbf{x} - \mathbf{x}_{\text{eq}}\|) \end{aligned} \tag{1.36}$$

Autrement dit, le coût terminal $V_f(\cdot)$ doit être une fonction de Lyapunov contrôlée, définie sur l'ensemble positivement invariant contrôlé $\mathbb{X}_f$, et les termes $V_f(\cdot)$ et $L(\cdot)$ de la fonction coût doivent être respectivement majoré et minoré par une fonction de classe $\mathcal{K}_\infty$. Ces hypothèses permettent à la fois de rendre le problème de commande optimale réalisable récursivement et de stabiliser asymptotiquement le système nominal en boucle fermée au point $(\mathbf{x}_{\text{eq}}, \mathbf{u}_{\text{eq}})$ avec $\mathcal{X}_{\mathcal{T}}$

comme région d'attraction.

Cependant, il n'est pas toujours évident de construire un coût terminal $V_f(\cdot)$ et un ensemble de contraintes finales $\mathbb{X}_f$ respectant les hypothèses (2.2). Pour contourner cette difficulté, la seconde approche se contente de trouver un horizon de prédiction suffisamment long qui garantisse la réalisabilité récursive du problème et la stabilité du système nominal en boucle fermée [42]-[45]. En revanche, les hypothèses utilisées sont bien plus compliquées à appréhender :

Hypothèse 2.3. (Stabilité nominale sans coût terminal ni contraintes finales).

H1) Il existe deux fonctions $\alpha_1 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ et $\alpha_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ de classe $\mathcal{K}_\infty$ telles que :

$$\forall\, \mathbf{x} \in \mathbb{X},\ \alpha_1(\|\mathbf{x} - \mathbf{x}_{\text{eq}}\|) \leq L^*(\mathbf{x}) \leq \alpha_2(\|\mathbf{x} - \mathbf{x}_{\text{eq}}\|),\ \text{où}\ L^*(\mathbf{x}) \coloneqq \inf_{\mathbf{u} \in \mathbb{U}(\mathbf{x})} L(\mathbf{x}, \mathbf{u})\,. \quad (1.37)$$

H2) Il existe une constante $\gamma > 0$ et un rayon $r > 0$ tels que :

$$\forall\, \mathbf{x} \in \mathcal{B}(\mathbf{x}_{\text{eq}}, r) \cap \mathbb{X},\ J^*_\infty(\mathbf{x}) \leq \gamma L^*(\mathbf{x}),\ \text{où}\ \mathcal{B}(\mathbf{x}_{\text{eq}}, r) \coloneqq \{\mathbf{x} \in \mathbb{R}^{n_\mathbf{x}} \mid \|\mathbf{x} - \mathbf{x}_{\text{eq}}\| < r\}\ . \quad (1.38)$$

H3) Pour tout horizon $T > 0$ et tout niveau $C > 0$, il existe une constante $K > 0$ telle que :

$$\forall\, \delta \in (0, \mathcal{T}],\ \forall\, \mathbf{x} \in \text{lev}_C\Big(J^*_\mathcal{T}(\mathbf{x})\Big),\ \delta L^*(\mathbf{x}) \leq K J^*_\delta(\mathbf{x}),\ \text{où}\ \text{lev}_C\Big(J^*_\mathcal{T}(\mathbf{x})\Big) \coloneqq \{\mathbf{x} \in \mathbb{X} \mid J^*_\mathcal{T}(\mathbf{x}) \leq C\}\ . \quad (1.39)$$

Le premier point de (2.3) est similaire au deuxième point de (2.2) et permet de borner la fonction coût optimale. Le deuxième point implique que le système doit pouvoir être dirigé assez rapidement vers sa position d'équilibre, sans quoi la fonction coût optimale à du problème à horizon infini risquerait de dépasser le seuil dépendant de γ. Le troisième point traduit le fait que les premiers termes de la fonction coût optimale doivent être minorés par un multiple du coût instantané.

2.5 Compensation du délai de transmission de la commande

Comme mentionné au début de la section précédente, il s'écoulera toujours un certain délai entre l'acquisition de l'état du système réel et la transmission de la loi de commande prédictive associée. Bien que souvent ignoré en théorie, ce délai :

$$\Delta(t) \coloneqq \Delta_s(t) + \Delta_c(t) + \Delta_a(t)\,, \quad (1.40)$$

principalement lié au temps de résolution $\Delta_c(t) > 0$ du problème de commande optimale [46]-[48], mais aussi aux retards $\Delta_s(t) \geq 0$ des capteurs et $\Delta_a(t) \geq 0$ des actionneurs [49], [50], est rarement négligeable en pratique, et peut sérieusement détériorer les performances et la stabilité du système en boucle fermée s'il n'est pas pris en compte. En effet, à cause du délai

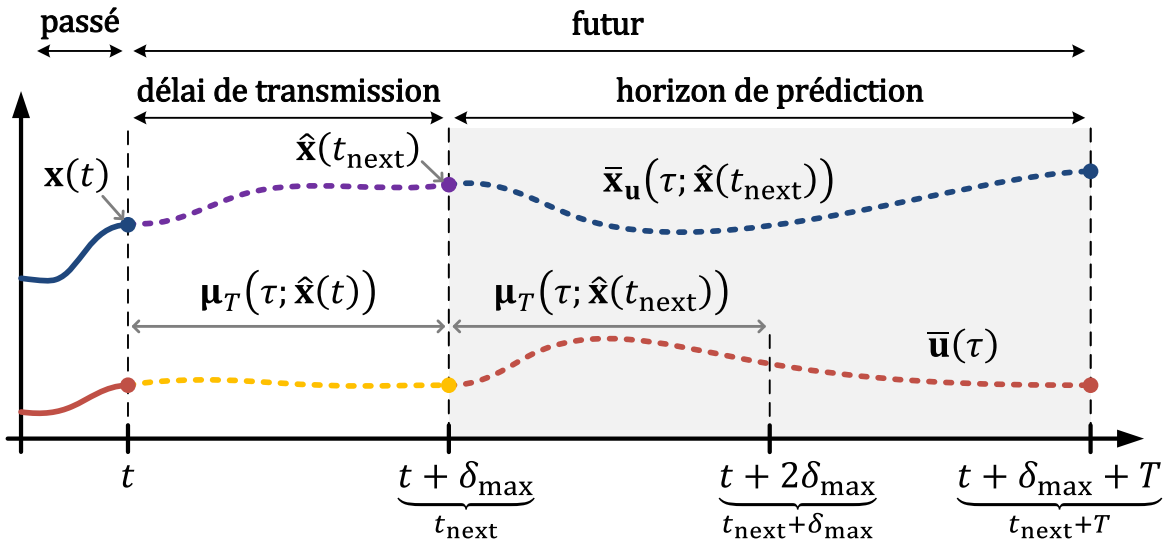

FIGURE 1.4 – Principe du schéma de compensation du délai de transmission de la commande.

de transmission, la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot;\mathbf{x}(t))$, calculée en partant du dernier état mesuré $\mathbf{x}(t)$, ne pourra être envoyée au système réel qu'à l'instant $t+\Delta(t)$ au lieu de l'instant t initialement prévu. Il est donc plus que probable que le comportement réel du système bouclé soit très éloigné de celui attendu, puisque la commande envoyée de t à $t+\Delta(t)$ ne coïncide pas avec celle calculée par l'algorithme. Ce décalage sera d'autant plus accentué que la différence entre les états $\mathbf{x}(t)$ et $\mathbf{x}(t+\Delta(t))$ sera importante.

Une manière évidente de compenser le délai de transmission serait de démarrer la résolution du problème de commande optimale $\Delta(t)$ secondes à l'avance, en partant non pas du dernier état mesuré $\mathbf{x}(t)$, mais du futur état $\mathbf{x}(t+\Delta(t))$ qu'aura le système réel au moment de recevoir la commande. Il suffirait alors d'envoyer la commande associée $\boldsymbol{\mu}_{\mathcal{T}}(\cdot;\mathbf{x}(t+\Delta(t)))$ au moment $t+\Delta(t)$ en question pour que les prédictions effectuées plus tôt restent cohérentes avec la réalité. Or, comme il est impossible d'obtenir à l'avance le futur état $\mathbf{x}(t+\Delta(t))$ du système réel, celui-ci est généralement estimé $\hat{\mathbf{x}}(t+\Delta(t))$ à partir du dernier état mesuré $\mathbf{x}(t)$, en simulant le modèle du problème de commande optimale (1.27) en boucle fermée avec la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot;\hat{\mathbf{x}}(t))$ calculée à l'itération précédente. En d'autres termes, $\hat{\mathbf{x}}(t+\Delta(t))$ est la solution à l'instant $\tau=\Delta(t)$ du problème de Cauchy :

$$\forall\,\tau\in[0,\Delta(t)],\ \mathbf{F}\Big(\dot{\bar{\mathbf{x}}}(\tau),\bar{\mathbf{x}}(\tau),\boldsymbol{\mu}_{\mathcal{T}}(t+\tau;\hat{\mathbf{x}}(t))\Big)=\mathbf{0},\ \bar{\mathbf{x}}(0)=\mathbf{x}(t)\,. \tag{1.41}$$

Néanmoins, il est essentiel de débuter la résolution du problème de commande optimale (1.27) en prenant une marge de retard supplémentaire, car le délai de transmission varie de façon aléatoire d'une itération à l'autre de l'algorithme. La stratégie la plus simple [49] consiste à surestimer le

délai de transmission $\Delta(t) \leq \delta_{\max}$ de sorte que la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot; \widehat{\mathbf{x}}(t+\delta_{\max}))$, calculée entre t et $t+\Delta(t)$, soit toujours disponible avant l'instant attendu $t+\delta_{\max}$. La mise en place de ce schéma de compensation du délai transmission conduit à l'algorithme de commande prédictive suivant :

Algorithme de commande prédictive avec compensation du délai de transmission

1) Mesurer l'état $\mathbf{x}(t)$ du système réel à l'instant courant $t \geq t_0$.
2) Envoyer la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot; \widehat{\mathbf{x}}(t))$, calculée à l'itération précédente entre $t-\delta_{\max}$ et $t-\delta_{\max}+\Delta(t-\delta_{\max})$, au système réel à l'instant courant $t \geq t_0$.
3) Estimer le futur état $\widehat{\mathbf{x}}(t_{\text{next}})$ qu'aura le système réel à l'instant $t_{\text{next}} \coloneqq t+\delta_{\max} \geq t+\Delta(t)$ en simulant le modèle (1.4) à partir du dernier état mesuré $\mathbf{x}(t)$ et en utilisant la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot; \widehat{\mathbf{x}}(t))$ calculée à l'itération précédente :

$$\widehat{\mathbf{x}}(t_{\text{next}}) \coloneqq \overline{\mathbf{x}}(\delta_{\max}) \text{ où : } \forall\, \tau \in [0, \delta_{\max}]\,, \begin{cases} \mathbf{F}\Big(\dot{\overline{\mathbf{x}}}(\tau), \overline{\mathbf{x}}(\tau), \boldsymbol{\mu}_{\mathcal{T}}(t+\tau; \widehat{\mathbf{x}}(t))\Big) = \mathbf{0} \\ \overline{\mathbf{x}}(0) = \mathbf{x}(t)\,. \end{cases} \tag{1.42}$$

4) Calculer la loi de commande optimale en boucle ouverte $\overline{\mathbf{u}}^*_{\mathcal{T}}(\cdot; \widehat{\mathbf{x}}(t_{\text{next}}))$ en résolvant, par une méthode directe, le problème de commande optimale à horizon fini :

$$\min_{\overline{\mathbf{u}}(\cdot)} \quad J_{\mathcal{T}}\Big(\widehat{\mathbf{x}}(t_{\text{next}}), \overline{\mathbf{u}}(\cdot)\Big) \coloneqq V_f\Big(\overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T}; \widehat{\mathbf{x}}(t_{\text{next}}))\Big) + \int_0^{\mathcal{T}} L\Big(\overline{\mathbf{x}}_{\mathbf{u}}(\tau; \widehat{\mathbf{x}}(t_{\text{next}})), \overline{\mathbf{u}}(\tau)\Big)\,\mathrm{d}\tau$$

$$\text{sous contraintes : } \forall\, \tau \in [0, \mathcal{T})\,,\ \Big(\overline{\mathbf{x}}_{\mathbf{u}}(\tau; \widehat{\mathbf{x}}(t_{\text{next}})), \overline{\mathbf{u}}(\tau)\Big) \in \mathbb{Y} \text{ et } \overline{\mathbf{x}}_{\mathbf{u}}(\mathcal{T}; \widehat{\mathbf{x}}(t_{\text{next}})) \in \mathbb{X}_f$$

$$\text{où, pour tout } \tau \in [0, \mathcal{T}]\,,\ \overline{\mathbf{x}}_{\mathbf{u}}(\tau; \widehat{\mathbf{x}}(t_{\text{next}})) \text{ est la solution de : } \begin{cases} \mathbf{F}\Big(\dot{\overline{\mathbf{x}}}(\tau), \overline{\mathbf{x}}(\tau), \overline{\mathbf{u}}(\tau)\Big) = \mathbf{0} \\ \overline{\mathbf{x}}(0) = \widehat{\mathbf{x}}(t_{\text{next}})\,. \end{cases} \tag{1.43}$$

Le trait suscrit permet ici de distinguer les variables utilisées par le modèle interne du contrôleur de celles mesurées sur le système réel.

5) Définir la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(t_{\text{next}}+\tau; \widehat{\mathbf{x}}(t_{\text{next}})) \coloneqq \overline{\mathbf{u}}^*_{\mathcal{T}}(\tau; \widehat{\mathbf{x}}(t_{\text{next}}))$ pour tout $\tau \in [0, \delta_{\max})$, où $\Delta(t) \leq \delta_{\max} \leq \mathcal{T}$ est la période au bout de laquelle est relancé l'algorithme.
6) Stocker la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot; \widehat{\mathbf{x}}(t_{\text{next}}))$ afin de l'envoyer au système réel à l'itération suivante, de t_{next} à $t_{\text{next}}+\delta_{\max}$.
7) Relancer l'algorithme à l'instant $t_{\text{next}} \coloneqq t+\delta_{\max}$.

Étant donné que le futur état $\mathbf{x}(t+\delta_{\max}))$ du système réel est estimé en boucle ouverte, celui-

ci sera inévitablement différent de $\widehat{\mathbf{x}}(t+\delta_{\max})$) en raison des erreurs de modélisation et autres perturbations imprévues. Toutefois, cela ne devrait pas remettre en cause la stabilité du système bouclé, puisqu'un certain niveau de robustesse est attendu face aux erreurs d'estimation [51]. Par ailleurs, le schéma de compensation du délai peut être encore amélioré [52]-[57] en ajoutant un terme correctif à la loi de commande prédictive $\boldsymbol{\mu}_{\mathcal{T}}(\cdot\,;\widehat{\mathbf{x}}(t+\delta_{\max}))$, qui dépendra de l'erreur d'estimation $\mathbf{x}(t+\delta_{\max})) - \widehat{\mathbf{x}}(t+\delta_{\max}))$ observée juste avant de l'envoyer au système réel.

3 Implémentation de l'algorithme NMPC

3.1 Simulation numérique des systèmes dynamiques

Jusqu'à présent, l'algorithme de commande prédictive a été défini en considérant que le modèle s'exprimait en temps continu, à l'aide d'équations différentielles ordinaires et d'équations algébriques. Par conséquent, l'évolution de son état ne pourra être prédite qu'en résolvant un problème de Cauchy. Or, comme il est généralement impossible de déterminer analytiquement la solution d'un problème de Cauchy, celle-ci doit être approchée par un schéma d'intégration numérique [58]-[62]. Pour éviter tout équivoque, il convient de préciser que l'expression « schéma d'intégration numérique » désigne ici les méthodes permettant de résoudre numériquement des problèmes de Cauchy, à ne pas confondre avec les formules de quadrature qui permettent de calculer numériquement la valeur d'une intégrale. Il est d'ailleurs possible d'établir des liens entre les deux problèmes, puisque le calcul d'une intégrale peut se ramener à la résolution d'un problème de Cauchy :

$$I(t) = \int_{t_0}^{t} f_I(\tau)\,\mathrm{d}\tau \;\Leftrightarrow\; \dot{I}(t) = f_I(t),\; I(t_0) = 0\,, \tag{1.44}$$

et que la résolution d'un problème de Cauchy peut se ramener au calcul d'une intégrale :

$$\dot{x}(t) = f_{\mathrm{ode}}\Big(t, x(t)\Big),\; x(t_0) = x_0 \;\Leftrightarrow\; x(t) = x_0 + \int_{t_0}^{t} f_{\mathrm{ode}}\Big(\tau, x(\tau)\Big)\,\mathrm{d}\tau\,. \tag{1.45}$$

En essence, un schéma d'intégration numérique tente d'extrapoler l'allure de la solution $x(\cdot)$ à partir de sa valeur courante et de la fonction $f_{\mathrm{ode}}(\cdot)$, alors qu'une formule de quadrature cherche à approcher l'intégrale $I(\cdot)$ en remplaçant $f_I(\cdot)$ par un polynôme d'interpolation dont la primitive est connue. Plus spécifiquement, le calcul d'une intégrale est un cas particulier de résolution de problème de Cauchy dans lequel la fonction d'évolution ne dépend pas de l'inconnu à déterminer. Cela signifie donc que n'importe quelle intégrale peut être calculée avec un schéma d'intégration numérique, tandis que seulement certains problèmes de Cauchy peuvent être résolus avec une formule de quadrature.

Pour respecter les conventions adoptées en analyse numérique, le modèle considéré dans cette

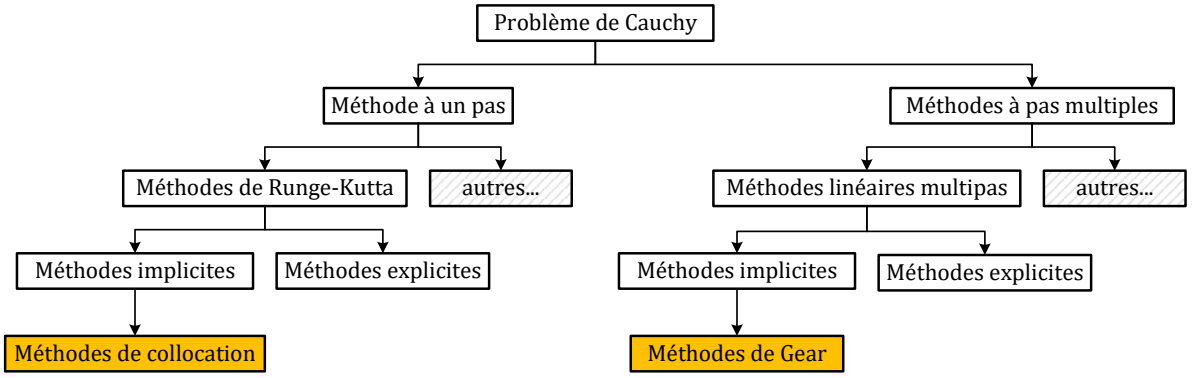

FIGURE 1.5 – Schémas d'intégration numériques utilisés pour résoudre un problème de Cauchy lorsque le modèle est raide.

sous-section est composé uniquement d'équations différentielles ordinaires qui ne dépendent pas explicitement du signal de commande $u(\cdot)$, celui-ci étant fourni au préalable par l'algorithme de commande prédictive :

$$f_{\text{ode}}\Big(t, x(t)\Big) \coloneqq f\Big(x(t), u(t)\Big)\,. \tag{1.46}$$

Le but d'un schéma d'intégration numérique est d'obtenir une valeur approchée de la solution exacte du problème de Cauchy :

$$\dot{x}(t) = f_{\text{ode}}\Big(t, x(t)\Big),\ x(t_0) = x_0\,, \tag{1.47}$$

sur un intervalle de temps $[t_0, t_f]$. Pour ce faire, la méthode part de l'instant initial $t_0 = \tau_0$ et procède itérativement, par incrément de temps $\tau_0 < \tau_1 < \cdots < \tau_n$, jusqu'à arriver à l'instant final $\tau_n = t_f$. La solution numérique du problème est alors générée par une suite d'itérés $(x_k)_{k=0}^{n}$ censés représenter la solution exacte aux instants $(\tau_0, \ldots, \tau_n)$. Pour tout $k \in [\![0, \mathcal{N}-1]\!]$, la valeur approchée $x_{k+1} \approx x(\tau_{k+1})$ de la solution exacte évaluée à l'instant τ_{k+1} est donnée par une équation aux différences :

$$x_{k+1} = \phi(\tau_{k+1-i}, x_{k+1-i}, h_k; i = i_0, \ldots, q),\ \text{où}\ h_k \coloneqq \tau_{k+1} - \tau_k\,, \tag{1.48}$$

qui découle, le plus souvent, d'un développement en séries de Taylor ou d'une approximation polynomiale. Si $q = 1$, la méthode est dite à un pas, car le calcul de x_{k+1} requiert uniquement de connaître l'itéré x_k, obtenu à l'instant précédent, pour être mené. En revanche, la méthode est dite à pas multiples si $q \in [\![2, b]\!]$, puisque le calcul de x_{k+1} ne dépend plus seulement de x_k, mais d'un historique d'itérés $(x_{k+1-q}, \ldots, x_k)$. En outre, ces deux types de méthodes peuvent être explicites ou implicites selon la valeur de l'indice $i_0 \in \{0, 1\}$. Lorsque $i_0 = 0$, la méthode est explicite, car le terme x_{k+1} peut se déduire directement de l'expression de $\phi(\cdot)$. À l'inverse, la

méthode devient implicite lorsque $i_0 = 1$, car l'expression de $\phi(\cdot)$ dépend elle-même de x_{k+1}. Les méthodes implicites nécessitent donc davantage de calculs que les méthodes explicites, puisque x_{k+1} doit être déterminé en résolvant une équation algébrique :

$$x_{k+1} - \phi(\tau_{k+1-i}, x_{k+1-i}, h_k; i = 0, \ldots, q) = 0\,. \tag{1.49}$$

Cependant, les méthodes implicites sont beaucoup plus stables que les méthodes explicites, ce qui permet d'augmenter grandement la taille du pas de temps effectué entre deux itérations.

3.1.1 Précision, zéro-stabilité et ordre de convergence

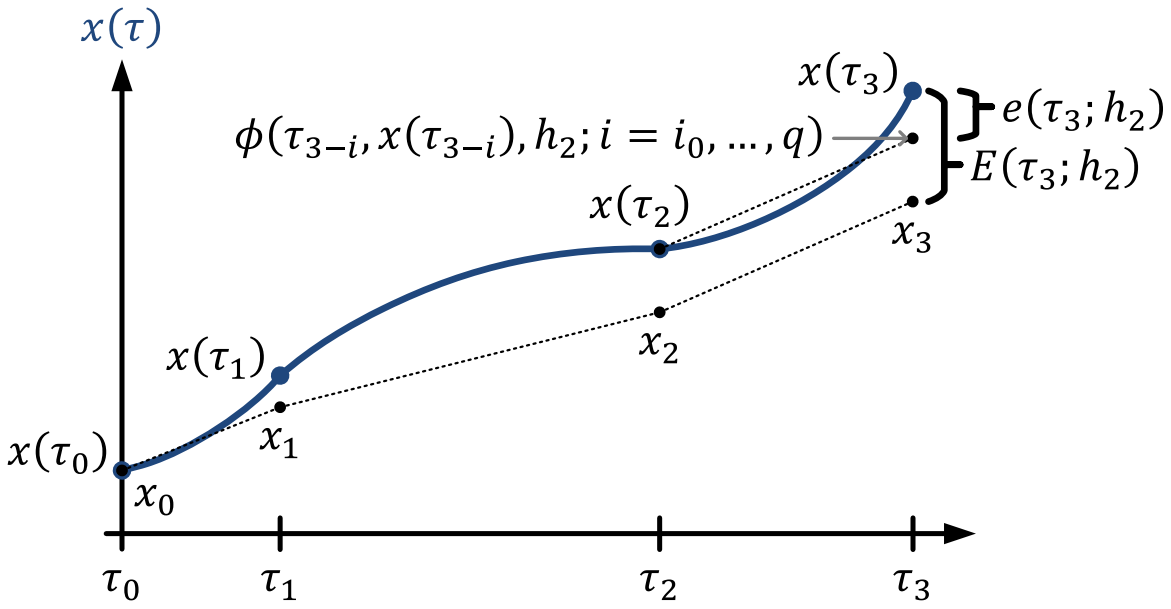

FIGURE 1.6 – Erreurs de troncature locale et globale entre la solution exacte d'un problème de Cauchy et sa solution approchée par un schéma d'intégration numérique.

En effet, la longueur du pas de temps est avant tout limitée par le domaine de stabilité de la méthode et par le niveau de précision recherché. La précision d'un schéma d'intégration numérique est quantifiée par deux indicateurs, à savoir l'erreur de troncature locale :

$$e(\tau_{k+1}; h_k) \coloneqq x(\tau_{k+1}) - \phi(\tau_{k+1-i}, x(\tau_{k+1-i}), h_k; i = i_0, \ldots, q)\,, \tag{1.50}$$

qui correspond à l'erreur induite par une itération de la méthode en partant d'un historique de solutions exactes $(x(\tau_{k+1-q}), \ldots, x(\tau_k))$, et l'erreur de troncature globale :

$$E(\tau_{k+1}; h_k) \coloneqq x(\tau_{k+1}) - x_{k+1}\,, \tag{1.51}$$

qui correspond à l'erreur accumulée à chaque itération en raison de la propagation de l'erreur de troncature locale. Pour qu'une méthode soit valide, il est essentiel que la solution numérique qu'elle génère converge vers la solution exacte du problème à mesure que le pas de temps diminue. Autrement dit, une méthode est convergente d'ordre $p \in \mathbb{N}_{>0}$ si l'erreur de troncature globale vérifie :

$$E(\tau_{k+1}; h_k) = \underset{h_{\max}\to 0}{\mathcal{O}}(h_{\max}^p), \text{ où } h_{\max} \coloneqq \max_{k\in[\![0,\mathcal{N}-1]\!]} h_k \,. \tag{1.52}$$

L'intérêt d'utiliser une méthode d'ordre élevé est de pouvoir effectuer des pas de temps plus importants qu'avec une méthode d'ordre faible, tout en gardant un niveau de précision similaire. La façon la plus simple de montrer la convergence (d'ordre p) d'une méthode est de s'assurer que celle-ci est consistante (d'ordre p) et zéro-stable. Tout d'abord, une méthode est consistante si l'erreur de troncature locale est négligeable devant le pas de temps :

$$e(\tau_{k+1}; h_k) = \underset{h_{\max}\to 0}{o}(h_{\max}) \,. \tag{1.53}$$

Plus précisément, une méthode est consistante d'ordre $p \in \mathbb{N}_{>0}$ si :

$$e(\tau_{k+1}; h_k) = \underset{h_{\max}\to 0}{\mathcal{O}}(h_{\max}^{p+1}) \,. \tag{1.54}$$

Ensuite, une méthode est zéro-stable s'il existe une constante $C_0 \in \mathbb{R}_{>0}$ et un pas de temps limite $h_{\lim} \in \mathbb{R}_{>0}$ tels que la relation suivante :

$$\forall\, h \in (0, h_{\lim}], \max_{k\in[\![0,\mathcal{N}-1]\!]} \left\|x_{k+1} - x'_{k+1}\right\| \leq C_0 \left(\left\|x_0 - x'_0\right\| + \max_{k\in[\![0,\mathcal{N}-1]\!]} \|\varepsilon_k\| \right), \tag{1.55}$$

soit satisfaite pour toutes suites d'itérés $(x_k)_{k=0}^n$ et $(x'_k)_{k=0}^n$ données par :

$$\forall\, k \in [\![0, n-1]\!], \begin{cases} x_{k+1} = \phi(\tau_{k+1-i}, x_{k+1-i}, h_k; i = i_0, \ldots, q) \\ x'_{k+1} = \phi(\tau_{k+1-i}, x'_{k+1-i}, h_k; i = i_0, \ldots, q) + \varepsilon_k \,. \end{cases} \tag{1.56}$$

La zéro-stabilité traduit le fait que l'erreur causée par la propagation de petites perturbations additives, typiquement des erreurs d'arrondis liées à l'arithmétique de l'ordinateur, reste faible du moment que le pas de temps utilisé est suffisamment petit. Néanmoins, la zéro-stabilité n'a pas vraiment d'intérêt pratique, car elle ne donne aucune indication concrète sur la taille de pas de temps à sélectionner.

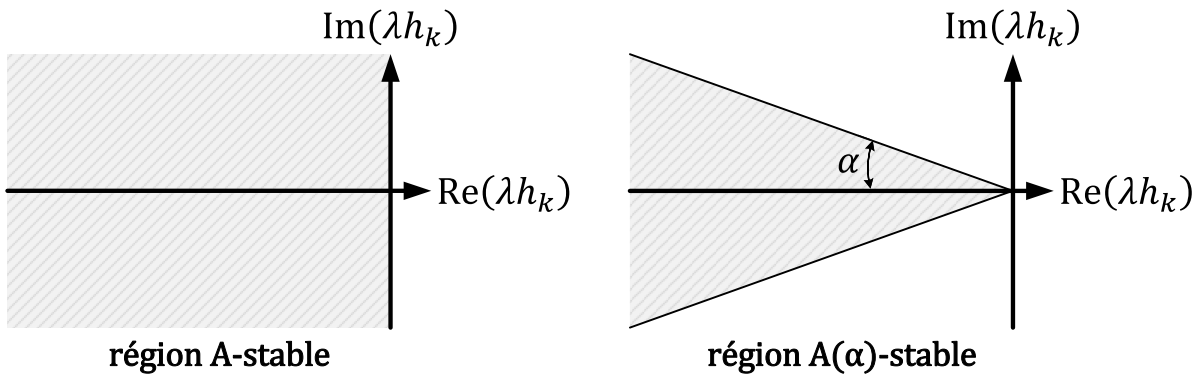

FIGURE 1.7 – Régions du plan complexe incluses respectivement dans le domaine de stabilité absolue d'une méthode A-stable et d'une méthode A(α)-stable.

3.1.2 Domaine de stabilité absolue

Un moyen simple d'estimer le domaine de stabilité de la méthode est de faire appel au problème test de Dahlquist :

$$\dot{x}(t) = \lambda x(t),\ x(t_0) = x_0,\ \text{où } \lambda \in \mathbb{C}_{<0} \coloneqq \{z \in \mathbb{C} \mid \mathrm{Re}(z) < 0\}\,. \tag{1.57}$$

Comme son nom l'indique, ce problème permet de tester la stabilité d'un schéma d'intégration numérique en comparant la réponse temporelle de sa solution exacte $x(t) = x_0 \exp(\lambda(t - t_0))$ à celle de sa solution numérique. Cette dernière est générée par une suite d'itérés dont l'équation aux différences dépend systématiquement du terme $\lambda h_k \in \mathbb{C}$, quelle que soit la méthode employée. Étant donné que $\mathrm{Re}(\lambda) < 0$, la solution exacte du problème de Dahlquist converge vers 0 quand $t \to +\infty$. Bien évidemment, il serait souhaitable que sa solution numérique présente le même comportement. De ce fait, la région de stabilité absolue d'un schéma d'intégration numérique est définie comme l'ensemble des produits $\lambda h_k \in \mathbb{C}$ pour lesquels la solution numérique du problème test (1.57) converge vers 0 quand $k \to +\infty$. En admettant que le modèle du problème d'origine (1.47) soit stable, une règle empirique est de choisir le pas de temps $h_k \in \mathbb{R}_{>0}$ de manière que le produit λh_k appartienne à la région de stabilité absolue de la méthode pour tout $\lambda \in \mathrm{Sp}(\partial f_{\mathrm{ode}}/\partial x) \subseteq \mathbb{C}_{<0}$, où $\mathrm{Sp}(\partial f_{\mathrm{ode}}/\partial x)$ désigne le spectre de la matrice jacobienne $\partial f_{\mathrm{ode}}/\partial x$. La longueur du pas temps est donc susceptible d'être limitée par la valeur propre de $\partial f_{\mathrm{ode}}/\partial x$ située le plus à gauche de l'axe des imaginaires, qui correspond, sans surprise, à l'état le plus rapide du modèle. L'enjeu est alors de trouver une méthode qui soit à la fois précise, peu coûteuse en ressources de calcul, et dont la région de stabilité absolue soit suffisamment grande pour ne pas restreindre excessivement la taille du pas de temps. En pratique, ce compromis n'est pas toujours évident à réaliser, surtout lorsque le modèle comprend des dynamiques aux échelles de temps très disparates. L'écart entre les échelles de temps d'un modèle peut être caractérisée

par son taux de raideur :

$$\frac{\max\limits_{\lambda \in \mathrm{Sp}(\,\partial f_{\mathrm{ode}}/\partial x\,)} |\mathrm{Re}(\lambda)|}{\min\limits_{\lambda \in \mathrm{Sp}(\,\partial f_{\mathrm{ode}}/\partial x\,)} |\mathrm{Re}(\lambda)|} \geq 1\,. \tag{1.58}$$

Les modèles singulièrement perturbés et les modèles descripteurs, par exemple, sont naturellement très raides. Plus la raideur d'un modèle est élevée, plus la longueur du pas de temps risque d'être limitée par la stabilité de la méthode. Ainsi, au-delà d'un certain taux de raideur, il est souvent préférable d'avoir recours à une méthode A-stable, c'est-à-dire dont la région de stabilité absolue englobe tout le demi-plan complexe gauche $\mathbb{C}_{<0}$. L'avantage d'une méthode A-stable est que le choix de son pas de temps n'est plus conditionné par sa région de stabilité absolue, puisque le produit $\lambda h_k \in \mathbb{C}$ vérifiera toujours $\mathrm{Re}(\lambda h_k) < 0$ quelles que soient les valeurs de $h_k \in \mathbb{R}_{>0}$ et de $\lambda \in \mathbb{C}_{<0}$. À noter qu'aucune méthode explicite n'est A-stable.

3.1.3 Les méthodes de Runge-Kutta

Les méthodes de Runge-Kutta sont la famille la plus répandue de méthodes à un pas. L'idée d'une méthode de Runge-Kutta est de déterminer le prochain itéré x_{k+1} à partir de l'itéré courant x_k et d'une combinaison de points additionnels, notés $x_{k,i}$, construits en évaluant la fonction $f_{\mathrm{ode}}(\cdot)$ à des instants intermédiaires, notés $\tau_{k,i}$, compris dans l'intervalle de temps $[\tau_k, \tau_{k+1}]$. L'équation aux différences d'une méthode de Runge-Kutta de rang $s \in \mathbb{N}_{>0}$ peut être écrite soit sous forme intégrale :

$$x_{k+1} = x_k + h_k \sum_{i=1}^{s} b_i \underbrace{f_{\mathrm{ode}}\Big(\tau_{k,i}, x_{k,i}\Big)}_{\approx \dot{x}(\tau_{k,i})}, \text{ avec : } \begin{cases} \tau_{k,i} \coloneqq \tau_k + c_i h_k \\ x_{k,i} \coloneqq x_k + h_k \sum\limits_{j=1}^{s} a_{ij} f_{\mathrm{ode}}\Big(\tau_{k,j}, x_{k,j}\Big)\,, \end{cases} \tag{1.59}$$

soit sous forme différentielle :

$$x_{k+1} = x_k + h_k \sum_{i=1}^{s} b_i \mathring{x}_{k,i}, \text{ avec : } \begin{cases} \tau_{k,i} \coloneqq \tau_k + c_i h_k \\ \mathring{x}_{k,i} \coloneqq f_{\mathrm{ode}}\Big(\tau_{k,i}, x_k + h_k \sum\limits_{j=1}^{s} a_{ij} \mathring{x}_{k,j}\Big)\,. \end{cases} \tag{1.60}$$

les deux formulations étant équivalentes sur le plan théorique, mais pas nécessairement sur le plan pratique. La méthode est explicite si la matrice $\mathbf{A} \coloneqq [a_{ij}]_{1 \leq i,j \leq s}$ est triangulaire inférieure stricte, et implicite sinon. Pour que la méthode soit convergente (d'ordre 1), les coefficients internes $a_{ij} \in \mathbb{R}$, les poids $b_i \in \mathbb{R}$ et les nœuds $c_i \in [0, 1]$ doivent au minimum respecter les conditions suivantes :

$$\sum_{i=1}^{s} b_i = 1 \quad \text{et} \quad \sum_{j=1}^{s} a_{ij} = c_i \text{ pour tout } i \in [\![1, s]\!]\,. \tag{1.61}$$

TABLE 1.1 – Valeurs numériques [24] des nœuds des formules de quadrature de Gauss sur $[0, 1]$.

Degré s	**nœuds c_i de Gauss-Legendre**	**nœuds c_i de Gauss-Radau**
1	0.5000	1.000
2	0.2113, 0.7887	0.3333, 1.0000
3	0.1127, 0.5000, 0.8873	0.1551, 0.6449, 0.8873
4	0.0694, 0.3300, 0.6700, 0.9306	0.0886, 0.4095, 0.7877, 1.0000
5	0.0469, 0.2308, 0.5000, 0.7692, 0.9531	0.0571, 0.2768, 0.5836, 0.8602, 1.000

Élaborer des méthodes de Runge-Kutta d'ordre élevé devient vite très complexe, car le nombre de conditions à respecter augmente exponentiellement avec l'ordre de la méthode. Dans les faits, l'ordre de convergence maximal que peut atteindre une méthode de Runge-Kutta de rang $s \in \mathbb{N}_{>0}$ vaut $p_{\max} = 2s$. L'équation aux différences d'une méthode de Runge-Kutta appliquée au problème test de Dahlquist (1.57) se réduit à :

$$x_{k+1} = R(h_k\lambda)x_k\,, \tag{1.62}$$

où $R(z) = 1 + z\mathbf{b}^\top(\mathbf{I}_s - z\mathbf{A})^{-1}\mathbf{1}_s$ est une fonction rationnelle, appelée fonction de stabilité de la méthode, avec $\mathbf{b} \coloneqq [b_1 \ \cdots \ b_s]^\top$ et $\mathbf{1}_s \coloneqq [1 \ \cdots \ 1]^\top \in \mathbb{R}^s$. Par conséquent, sa région de stabilité absolue est simplement donnée par :

$$\mathcal{S} = \{z \in \mathbb{C} \mid |R(z)| < 1\}\,. \tag{1.63}$$

Un point fort des méthodes de Runge-Kutta vient de leur capacité à produire des méthodes A-stables d'ordre quelconque. Toutefois, certaines méthodes de Runge-Kutta A-stables parviennent mieux à gérer les modèles extrêmement raides que d'autres. Par exemple, si $\lim_{z\to-\infty} |R(z)| = 1$, alors il est probable que les composantes rapides de la solution numérique convergeront bien plus lentement vers leurs valeurs d'équilibre qu'elles ne le font en réalité (comparer la vitesse de convergence de $x(t) = x_0 \exp(\lambda(t - t_0))$ et $x_{k+1} = R(\lambda h_k)^{k+1}x_0$ lorsque $\mathrm{Re}(\lambda) \to -\infty$). À l'inverse, une méthode de Runge-Kutta est dite L-stable si elle est A-stable et que sa fonction de stabilité vérifie $\lim_{z\to-\infty} |R(z)| = 0$. Les méthodes de Runge-Kutta L-stable garantissent donc que les composantes les plus rapides de la solution exacte seront atténuées correctement par sa solution numérique quelle que soit la taille du pas temps effectué. Pour qu'une méthode de Runge-Kutta A-stable devienne L-stable, il suffit juste que la matrice $\mathbf{A}$ soit inversible et que les coefficients de sa dernière ligne soient égaux aux poids de la méthode :

$$\forall\, j \in [\![1, s]\!],\ a_{sj} = b_j \ \Rightarrow\ x_{k,s} = x_{k+1},\ \text{et}\ \tau_{k,s} = \tau_{k+1}\,. \tag{1.64}$$

Les méthodes de collocation sont une catégorie très populaire de méthodes de Runge-Kutta implicites A-stables. Le principe d'une méthode de collocation [63] est d'approcher la solution exacte du problème de Cauchy (1.47) sur chaque intervalle de temps $[\tau_k, \tau_{k+1}]$ par un polynôme de degré $s \in \mathbb{N}_{>0}$, noté $\chi_k(\cdot)$, qui respecte les équations d'évolution du modèle :

$$\chi_k(\tau_k) = x_k \quad \text{et} \quad \dot{\chi}_k(\tau_{k,i}) = f_{\text{ode}}\Big(\tau_{k,i}, \chi_k(\tau_{k,i})\Big) \text{ pour tout } i \in [\![1, s]\!]\,, \tag{1.65}$$

aux instants $\tau_k < \tau_{k,1} < \cdots < \tau_{k,s} \leq \tau_{k+1}$. Pour ce faire, la dérivée temporelle du polynôme de collocation $\chi_k(\cdot)$ est généralement projetée dans une base de polynômes de Lagrange :

$$\forall\, \tau \in [\tau_k, \tau_{k+1}],\ \dot{\chi}_k(\tau) = \sum_{j=1}^{s} \widetilde{\ell}_j(\tau) \mathring{x}_{k,j},\ \text{avec } \widetilde{\ell}_j(\tau) \coloneqq \prod_{\substack{r=1 \\ r \neq j}}^{s} \left(\frac{\tau - \tau_{k,r}}{\tau_{k,j} - \tau_{k,r}} \right), \tag{1.66}$$

de sorte que ses coefficients $(\mathring{x}_{k,1}, \ldots, \mathring{x}_{k,s})$ vérifient :

$$\forall\, i \in [\![1, s]\!],\ \forall\, j \in [\![1, s]\!],\ \widetilde{\ell}_j(\tau_{k,i}) = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j\,, \end{cases} \ \Rightarrow\ \mathring{x}_{k,i} = \dot{\chi}_k(\tau_{k,i})\,. \tag{1.67}$$

L'expression du polynôme de collocation est ensuite simplement retrouvée en intégrant celle de sa dérivée temporelle (1.66) comme suit :

$$\forall\, \tau \in [\tau_k, \tau_{k+1}],\ \chi_k(\tau) = \chi_k(\tau_k) + \int_{\tau_k}^{\tau} \dot{\chi}_k(t)\, \mathrm{d}t = \chi_k(\tau_k) + h_k \sum_{j=1}^{s} \left(\int_0^{\frac{\tau - \tau_k}{h_k}} \widetilde{\ell}_j(\tau_k + c h_k)\, \mathrm{d}c \right) \dot{\chi}_k(\tau_{k,j})\,. \tag{1.68}$$

Il est alors possible, en combinant les conditions (1.65) et l'équation (1.68), de parvenir à l'équation aux différences de la méthode de collocation :

$$\begin{aligned} x_{k+1} = \chi_k(\tau_{k+1}) &= x_k + h_k \sum_{i=1}^{s} \left(\int_0^1 \widetilde{\ell}_i(\tau_k + c h_k)\, \mathrm{d}c \right) f_{\text{ode}}\Big(\tau_{k,i}, \chi_k(\tau_{k,i})\Big) \\ \text{avec : } \chi_k(\tau_{k,i}) &= x_k + h_k \sum_{j=1}^{s} \left(\int_0^{c_i} \widetilde{\ell}_j(\tau_k + c h_k)\, \mathrm{d}c \right) f_{\text{ode}}\Big(\tau_{k,j}, \chi_k(\tau_{k,j})\Big)\,, \end{aligned} \tag{1.69}$$

qui correspond bien à celle d'une méthode de Runge-Kutta de rang $s \in \mathbb{N}_{>0}$ (1.59) satisfaisant :

$$\forall\, i \in [\![1, s]\!],\ \forall\, j \in [\![1, s]\!],\ a_{ij} = \int_0^{c_i} \widetilde{\ell}_j(\tau_k + c h_k)\, \mathrm{d}c,\quad b_i = \int_0^1 \widetilde{\ell}_i(\tau_k + c h_k)\, \mathrm{d}c,\quad c_i = \left(\frac{\tau_{k,i} - \tau_k}{h_k} \right). \tag{1.70}$$

La principale différence entre deux types de méthodes de collocation réside dans le choix de leurs nœuds $(c_1, \ldots, c_s) \in [0, 1]^s$. Ceux-ci sont sélectionnés de façon que le polynôme de collocation approche le plus précisément possible la solution exacte du problème de Cauchy (1.47). Certaines

méthodes de collocation choisissent leurs nœuds en considérant la somme pondérée de l'équation aux différences (1.69) comme une formule de quadrature de Gauss :

$$\int_{\tau_k}^{\tau_{k+1}} f_{\text{ode}}\Big(\tau, \chi_k(\tau)\Big)\,\mathrm{d}\tau \approx h_k \sum_{i=1}^{s} b_i f_{\text{ode}}\Big(\tau_k + c_i h_k, \chi_k(\tau_k + c_i h_k)\Big)\,. \tag{1.71}$$

L'intérêt de cette approche est que l'approximation (1.71) devient exacte si $f_{\text{ode}}(\tau, \chi_k(\tau)) = P(\tau)$, où $P(\cdot)$ est un polynôme de degré $\deg(P) \leq 2s-1$, et si les nœuds $(c_1, \ldots, c_s) \in [0,1]^s$ sont les racines du s$^{\text{ième}}$ polynôme de Legendre :

$$P_{\text{Legendre}}(c) \coloneqq \frac{\mathrm{d}^s}{\mathrm{d}c^s}\Big(c^s(1-c)^s\Big) \tag{1.72}$$

ramené sur le domaine $[0,1]$. La méthode de collocation qui en résulte, dite de Gauss-Legendre, est la seule méthode de Runge-Kutta dont l'ordre de convergence $p_{\max} = 2s$ est maximal. Une autre méthode de collocation, obtenue en prenant les nœuds de la formule de quadrature de Gauss-Radau :

$$P_{\text{Radau}}(c) \coloneqq \frac{\mathrm{d}^{s-1}}{\mathrm{d}c^{s-1}}\Big(c^{s-1}(1-c)^s\Big) \tag{1.73}$$

est privilégiée pour les modèles extrêmement raides. La particularité d'une méthode de collocation de type Gauss-Radau est que son dernier nœud $c_s = 1$ est fixé au bout de l'intervalle $[0,1]$ afin de rendre la méthode L-stable (1.64). En effet :

$$c_s = 1 \;\Rightarrow\; a_{sj} = \int_0^{c_s} \widetilde{\ell}_j(\tau_k + c h_k)\,\mathrm{d}c = b_j \text{ pour tout } j \in [\![1, s]\!]\,. \tag{1.74}$$

En contrepartie, l'ordre de convergence $p = 2s-1$ d'une méthode de collocation de type Gauss-Radau est un cran inférieur à celui d'une méthode de type Gauss-Legendre, car figer la valeur d'un de ses nœuds revient à retirer un degré de liberté à la formule de quadrature Gauss (1.71).

3.1.4 Les méthodes linéaires multipas

La plupart des méthodes à pas multiples, pour ne pas dire toutes, appartiennent à la famille des méthodes linéaires multipas. L'équation aux différences d'une méthode linéaire multipas de rang $q \in \mathbb{N}_{>0}$ est obtenue en combinant linéairement les itérés $(x_{k+1}, \ldots, x_{k+1-q})$ et la fonction $f_{\text{ode}}(\cdot)$ évaluée aux instants $(\tau_{k+1}, \ldots, \tau_{k+1-q})$ associés :

$$\sum_{i=0}^{q} a_{q-i} x_{k+1-i} = h \sum_{i=0}^{q} b_{q-i} \underbrace{f_{\text{ode}}\Big(\tau_{k+1-i}, x_{k+1-i}\Big)}_{\approx \dot{x}(\tau_{k+1-i})}\,, \tag{1.75}$$

où $(a_0, \cdots, a_q) \in \mathbb{R}^{q+1}$ et $(b_0, \cdots, b_q) \in \mathbb{R}^{q+1}$ sont les coefficients de la méthode, avec $a_q \neq 0$ et $|a_0| + |b_0| > 0$. La méthode est explicite si $b_q = 0$, et implicite sinon. Les polynômes :

$$P_a(r) \coloneqq \sum_{i=0}^{q} a_i r^i \quad \text{et} \quad P_b(r) \coloneqq \sum_{i=0}^{q} b_i r^i \,, \tag{1.76}$$

sont appelés les polynômes générateurs de la méthode. Les méthodes linéaires multipas sont moins coûteuses en ressources de calcul que les méthodes de Runge-Kutta, car les valeurs obtenues en évaluant la fonction $f_{\text{ode}}(\cdot)$ peuvent être partiellement réutilisées d'une itération à l'autre. En revanche, la méthode ne peut pas se lancer à partir d'une seule condition initiale $x(t_0) = x_0$, ce qui implique que les premiers itérés $(x_1, \ldots, x_q)$ doivent être calculés autrement, avec une méthode à un pas. De plus, il est difficile de faire varier la taille du pas de temps $h \in \mathbb{R}_{>0}$, supposée ici constante, sans complexifier énormément la méthode. Pour qu'une méthode linéaire multipas soit convergente (d'ordre p), ses coefficients doivent respecter les conditions (consistance) :

$$P_a(1) = \sum_{i=0}^{q} a_i = 0 \quad \text{et} \quad \sum_{i=0}^{q} i^j a_i = j \sum_{i=0}^{q} i^{j-1} b_i \text{ pour tout } j \in [\![1, p]\!] \,, \tag{1.77}$$

et les racines du polynôme générateur $P_a(r)$ doivent être à l'intérieur du cercle unité, à l'exception de la racine simple $r = 1$ (zéro-stabilité). L'ordre de convergence maximal que peut atteindre une méthode linéaire multipas de rang q vaut $p_{1,\max} = q + 1$ si q est impair, et $p_{2,\max} = q + 2$ si q est pair. L'équation aux différences d'une méthode linéaire multipas appliquée au problème test de Dahlquist (1.57) peut s'écrire en fonction de ses polynômes générateurs :

$$\sum_{i=0}^{q} (a_{q-i} - \lambda h b_{q-i}) x_{k+1-i} = 0 \Leftrightarrow P_a(r) - \lambda h P_b(r) = 0, \text{ en posant } r^{q-i} = x_{k+1-i} \,. \tag{1.78}$$

Sa région de stabilité absolue dépend donc des racines, notées $r_i(z)$, du polynôme caractéristique $P_{a,b}(r, z) \coloneqq P_a(r) - z P_b(r)$:

$$\mathcal{S} = \{ z \in \mathbb{C} \mid \forall\, i \in [\![0, q]\!],\ |r_i(z)| < 1 \} \,. \tag{1.79}$$

À la différence des méthodes de Runge-Kutta, l'ordre d'une méthode linéaire multipas A-stable ne peut pas être supérieur à 2. Pourtant, certaines méthodes linéaires multipas d'ordre plus élevé semblent, en pratique, bien adaptées aux modèles raides. Ces méthodes, dites A(α)-stables, sont en fait proches d'être A-stables, puisque leur région de stabilité absolue englobe le secteur de demi-angle $\alpha \in (0, \pi/2)$ défini par l'ensemble $\mathbb{C}_{<0}(\alpha) \coloneqq \{ z \in \mathbb{C} \mid |\arg(-z)| < \alpha,\ z \neq 0 \}$. Ce secteur, inclus dans le demi-plan complexe gauche, contient notamment la droite $\mathbb{R}_{<0} \coloneqq (-\infty, 0)$, ce qui est suffisant pour ne pas limiter la longueur du pas temps lorsque le modèle est stable et que les valeurs propres de sa matrice jacobienne sont toutes réelles.

Les méthodes de Gear, aussi appelées formules de différentiation rétrograde, sont une catégorie de méthodes linéaires multipas implicites A(α)-stables couramment utilisées pour simuler les modèles raides. Le principe d'une méthode de Gear est de remplacer la solution exacte du problème de Cauchy (1.47) par l'unique polynôme d'interpolation de degré $q \in [\![2, 6]\!]$ (écrit de préférence sous forme de Newton plutôt que sous forme de Lagrange) passant par les nœuds $(x_{k+1}, \ldots, x_{k+1-q})$ aux instants $(\tau_{k+1}, \ldots, \tau_{k+1-q})$. Le prochain itéré x_{k+1} est ensuite calculé en supposant que le polynôme d'interpolation, noté $\chi(\cdot)$, vérifie l'équation d'évolution du modèle à l'instant τ_{k+1}, soit :

$$\dot{\chi}(\tau_{k+1}) = f_{\text{ode}}\Big(\tau_{k+1}, \chi(\tau_{k+1})\Big)\,. \tag{1.80}$$

Quand la taille du pas de temps $h \in \mathbb{R}_{>0}$ est constante, l'équation aux différences d'une méthode de Gear d'ordre $q \in [\![2, 6]\!]$ peut s'écrire comme une somme de différences rétrogrades :

$$\sum_{j=1}^{q} \frac{1}{j} \Delta_{\ominus}^{j} x_{k+1} = h f_{\text{ode}}\Big(\tau_{k+1}, x_{k+1}\Big), \text{ avec : } \begin{cases} \Delta_{\ominus}^{j} x_{k+1} \coloneqq \Delta_{\ominus}^{j-1} x_{k+1} - \Delta_{\ominus}^{j-1} x_k \\ \Delta_{\ominus}^{0} x_{k+1} \coloneqq x_{k+1}\,. \end{cases} \tag{1.81}$$

Contrairement aux méthodes de Runge-Kutta d'ordre $p \geq 2$, l'équation aux différences des méthodes de Gear ne nécessite qu'une seule évaluation de la fonction $f_{\text{ode}}(\cdot)$ pour être définie. De ce fait, les méthodes de Gear permettent, en moyenne, de résoudre plus rapidement le problème de Cauchy (1.47) que les méthodes de Runge-Kutta implicites, mais ce, à condition que les dynamiques du modèle, supposé stable, soient suffisamment amorties pour que les valeurs propres de la matrice jacobienne $\partial f_{\text{ode}}/\partial x$ restent à l'intérieur de la région de stabilité absolue de la méthode. En revanche, les méthodes de Gear ne sont A-stables que si $q = 2$, et seront A(α)-stables autrement, avec un demi-angle $\alpha \in (0, \pi/2)$ de plus en plus proche de 0 à mesure que l'ordre $q \in [\![3, 6]\!]$ de la méthode augmente. Par ailleurs, les méthodes de Gear ne sont plus zéro-stable au-delà de l'ordre $q = 6$, ce qui n'est pas cas des méthodes de Runge-Kutta implicites.

3.1.5 Adaptation automatique de la taille du pas de temps

En pratique, les schémas d'intégration numérique les plus sophistiqués sont capables d'adapter automatiquement la taille du pas de temps en fonction du niveau de précision recherché. Cette fonctionnalité rend la méthode beaucoup plus efficace en termes de coût de calcul, car elle permet de réduire ou d'augmenter ponctuellement le nombre de pas de temps effectués lorsque la solution se met à varier lentement ou rapidement au cours du temps. Par ailleurs, certains schémas d'intégration numérique sont également capables d'adapter l'ordre de la méthode pour économiser encore davantage de ressources de calcul. Pour contrôler la précision de la méthode, son erreur de troncature locale est généralement estimée en remplaçant la solution exacte $x(\tau_{k+1})$ du problème de Cauchy (1.47), toujours inconnue, par la solution numérique $\widetilde{x}_{k+1}$ d'une seconde

méthode plus précise :

$$x(\tau_{k+1}) \approx \widetilde{x}_{k+1} \;\Rightarrow\; e(\tau_{k+1}; h_k) \approx \mathrm{err}(\tau_{k+1}; h_k), \text{ avec } \mathrm{err}(\tau_{k+1}; h_k) \coloneqq \widetilde{x}_{k+1} - x_{k+1}\,, \tag{1.82}$$

et où les suites d'itérés $(x_k)_{k=0}^n$ et $(\widetilde{x}_k)_{k=0}^n$ sont données par deux méthodes d'ordre différent :

$$\begin{aligned} x_{k+1} &= \phi(\tau_{k+1-i}, x_{k+1-i}, h_k; i = i_0, \dots, q), \quad \text{d'ordre } p \in \mathbb{N}_{>0} \\ \widetilde{x}_{k+1} &= \widetilde{\phi}(\tau_{k+1-i}, \widetilde{x}_{k+1-i}, h_k; i = i_0, \dots, q), \quad \text{d'ordre } \widetilde{p} = p+1\,. \end{aligned} \tag{1.83}$$

En admettant que les solutions numériques $(x_{k+1-q}, \dots, x_k)$ et $(\widetilde{x}_{k+1-q}, \dots, \widetilde{x}_k)$ calculées aux itérations précédentes soient exactes, l'erreur de troncature locale de chaque méthode vérifie :

$$\begin{aligned} e(\tau_{k+1}; h_k) &= x(\tau_{k+1}) - x_{k+1} = \underset{h_k \to 0}{\mathcal{O}}(h_k^{p+1}) \\ \widetilde{e}(\tau_{k+1}; h_k) &= x(\tau_{k+1}) - \widetilde{x}_{k+1} = \underset{h_k \to 0}{\mathcal{O}}(h_k^{p+2})\,. \end{aligned} \tag{1.84}$$

De plus, s'il existe une constante $C \in \mathbb{R}_{>0}$ telle que :

$$e(\tau_{k+1}; h_k) = x(\tau_{k+1}) - x_{k+1} = C h_k^{p+1} + \underset{h_k \to 0}{\mathcal{O}}(h_k^{p+2})\,, \tag{1.85}$$

alors l'erreur de troncature locale estimée satisfait également :

$$\mathrm{err}(\tau_{k+1}; h_k) = e(\tau_{k+1}; h_k) - \widetilde{e}(\tau_{k+1}; h_k) = C h_k^{p+1} + \underset{h_k \to 0}{\mathcal{O}}(h_k^{p+2})\,. \tag{1.86}$$

Par conséquent, sa norme peut être approchée par :

$$\|\mathrm{err}(\tau_{k+1}; h_k)\| \approx C h_k^{p+1}\,, \tag{1.87}$$

en supposant que la taille du pas de temps utilisé à l'itération courante est suffisamment faible. Le pas de temps suivant sera ensuite modifié de sorte que la valeur approchée (1.87) de la norme de l'erreur de troncature locale estimée (1.86) reste inférieure à une certaine tolérance :

$$\|\mathrm{err}(\tau_{k+2}; h_{k+1})\| \approx C h_{k+1}^{p+1} \leq \mathrm{tol}(\tau_{k+1}; h_k) \;\Rightarrow\; h_{k+1} \leq h_k \left(\frac{\mathrm{tol}(\tau_{k+1}; h_k)}{\|\mathrm{err}(\tau_{k+1}; h_k)\|} \right)^{1/(p+1)}. \tag{1.88}$$

Étant donné que la détermination de cette inégalité repose sur de nombreuses hypothèses, une marge de sécurité supplémentaire est ajoutée à la formule finale, par exemple :

$$h_{k+1} = 0.9\, h_k \left(\frac{\mathrm{tol}(\tau_{k+1}; h_k)}{\|\mathrm{err}(\tau_{k+1}; h_k)\|} \right)^{1/(p+1)}. \tag{1.89}$$

La tolérance utilisée pour modifier le pas de temps est habituellement de la forme :

$$\mathrm{tol}(\tau_{k+1}; h_k) \coloneqq \mathrm{Atol} + \max\left(\|\widetilde{x}_{k+1}\|, \|x_{k+1}\|\right) \mathrm{Rtol}\,, \tag{1.90}$$

où $0 < \mathrm{Atol} \ll 1$ et $0 < \mathrm{Rtol} \ll 1$ sont des paramètres constants choisis par l'utilisateur.

3.2 Transcription du problème de commande optimale

Réussir à simuler efficacement le modèle en choisissant le bon schéma d'intégration numérique ne constitue qu'une première étape dans la mise en place de l'algorithme de commande prédictive. Une deuxième étape indispensable est de transcrire [19], [22], [24], [32], [64] le problème de commande optimale en un problème d'optimisation numérique. En effet, comme les variables de décision $\overline{\mathbf{x}} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_{\mathbf{x}}}$ et $\overline{\mathbf{u}} : \mathbb{R}_{\geq 0} \to \mathbb{R}^{n_{\mathbf{u}}}$ du problème de commande optimale sont des fonctions, celles-ci peuvent prendre une infinité de valeurs sur l'intervalle de temps $[0, \mathcal{T}]$. Or, les méthodes d'optimisation habituellement utilisées en commande prédictive ne permettent pas de manipuler un nombre illimité de variables de décision. L'un des objectifs d'une méthode de transcription est donc de représenter le plus fidèlement possible les fonctions $\overline{\mathbf{x}}(\cdot)$ et $\overline{\mathbf{u}}(\cdot)$ du problème d'origine, mais en utilisant seulement un nombre fini de paramètres réels.

3.2.1 Paramétrisation du signal de commande

Le signal de commande, par exemple, pourrait être représenté sur tout l'horizon de temps $[0, \mathcal{T}]$ à l'aide d'un polynôme de degré quelconque. L'inconvénient de cette paramétrisation globale est qu'elle ne permet pas de mettre en évidence la structure triangulaire, induite par la relation de causalité entre la trajectoire d'état du modèle et le signal de commande, du problème de Cauchy (1.4) (l'état courant ne dépend que des commandes courantes et/ou passées). De ce fait, les méthodes de transcription employées en commande prédictive proposent plutôt de représenter localement les variables $\overline{\mathbf{x}}(\cdot)$ et $\overline{\mathbf{u}}(\cdot)$ sous forme de fonctions continues par morceaux. L'intervalle de temps $[0, \mathcal{T}]$ est alors divisé en $\mathcal{N} \in \mathbb{N}_{>0}$ sous-intervalles $[t_k, t_{k+1}]$, avec $0 = t_0 < t_1 < \cdots < t_{\mathcal{N}} = \mathcal{T}$, sur lesquels le signal de commande $\overline{\mathbf{u}}(\cdot)$ vérifie :

$$\forall\, k \in [\![0, \mathcal{N}-1]\!],\ \forall\, \tau \in [t_k, t_{k+1}),\ \overline{\mathbf{u}}(\tau) = \overline{\mathbf{\Psi}}_k\Big(\tau, \mathbf{v}_k\Big)\,. \tag{1.91}$$

Pour tout $k \in [\![0, \mathcal{N}-1]\!]$, la fonction de base $\overline{\mathbf{\Psi}}_k(\cdot)$ est continue sur le sous-intervalle $[t_k, t_{k+1})$ et est paramétrée par un vecteur de dimension finie $\mathbf{v}_k \in \mathbb{R}^{n_{\mathbf{v}}}$. La fonction de base est, en règle générale, un polynôme de degré $n_{\mathbf{\Psi}} \in \mathbb{N}_{>0}$ dont les coefficients sont les composantes du vecteur de paramètres $\mathbf{v}_k \coloneqq [\mathbf{v}_{k,0}^\top \cdots \mathbf{v}_{k,n_{\mathbf{\Psi}}}^\top]^\top$, soit :

$$\forall\, k \in [\![0, \mathcal{N}-1]\!],\ \forall\, \tau \in [t_k, t_{k+1}),\ \overline{\mathbf{u}}(\tau) = \sum_{j=0}^{n_{\mathbf{\Psi}}} \mathbf{v}_{k,j} (\tau - t_k)^j\,. \tag{1.92}$$

Cependant, en pratique, le degré du polynôme $\overline{\mathbf{\Psi}}_k(\cdot)$ est la plupart du temps pris égal à $n_{\mathbf{\Psi}} = 0$ pour éviter de trop augmenter la dimension $n_{\mathbf{v}} = (n_{\mathbf{\Psi}} + 1)n_{\mathbf{u}}$ du vecteur de paramètres $\mathbf{v}_k$. Ainsi, le nombre total de variables de décision utilisées pour représenter le signal de commande est limité à $\mathcal{N}n_{\mathbf{v}} = \mathcal{N}n_{\mathbf{u}}$ variables, la fonction $\overline{\mathbf{u}}(\cdot)$ étant dorénavant constante par morceaux :

$$\forall\, k \in [\![0, \mathcal{N}-1]\!],\ \forall\, \tau \in [t_k, t_{k+1}),\ \overline{\mathbf{u}}(\tau) = \mathbf{v}_{k,0} = \mathbf{v}_k\,. \tag{1.93}$$

Toutefois, la différence entre les méthodes de transcription ne vient pas tant de la paramétrisation du signal de commande que de celle de la trajectoire d'état. Comme mentionné dans la section précédente, l'état du modèle est en réalité approché, ou discrétisé, par un schéma d'intégration numérique. Dès lors, plusieurs façons de paramétrer la trajectoire d'état peuvent être envisagées selon la manière dont le schéma intégration numérique est incorporé au problème d'optimisation.

3.2.2 Méthode de tir simple

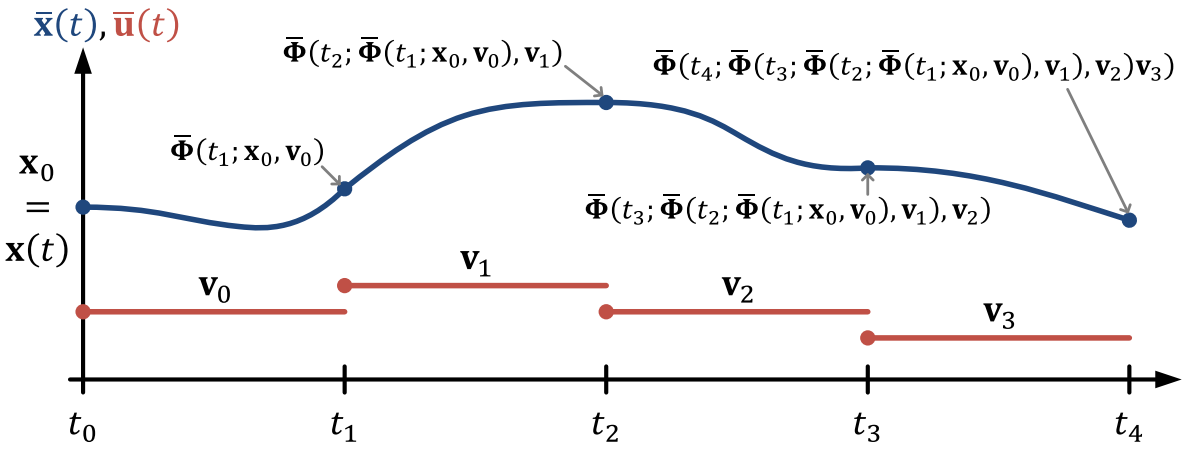

FIGURE 1.8 – Exemple de trajectoire d'état et de signal de commande obtenus à la suite de la transcription du problème de commande optimale par une méthode de tir simple.

Le tir simple, pour commencer, est la méthode de transcription la plus intuitive et la plus facile à mettre en place. En tir simple, la trajectoire d'état du modèle est calculée d'un seul tenant, sur tout l'intervalle de temps $[0, \mathcal{T}]$, en dehors du problème d'optimisation, à partir de l'état initial $\overline{\mathbf{x}}(0) = \mathbf{x}(t)$ mesuré périodiquement sur le système réel et de la loi de commande $\overline{\mathbf{u}}(\cdot)$ renvoyée par le solveur d'optimisation exécuté en arrière-plan :

$$\forall\, \tau \in [0, \mathcal{T}],\ \overline{\mathbf{x}}(\tau) = \overline{\mathbf{\Phi}}\Big(\tau; \mathbf{x}(t), \overline{\mathbf{u}}(\tau)\Big) \text{ est la solution de : } \begin{cases} \mathbf{F}\Big(\dot{\overline{\mathbf{x}}}(\tau), \overline{\mathbf{x}}(\tau), \overline{\mathbf{u}}(\tau)\Big) = \mathbf{0} \\ \overline{\mathbf{x}}(0) = \mathbf{x}(t)\,. \end{cases} \tag{1.94}$$

La fonction $\overline{\mathbf{\Phi}}(\cdot)$ est introduite ici afin de masquer la complexité du schéma d'intégration numérique retenu pour simuler le modèle. Étant donné que le signal de commande $\overline{\mathbf{u}}(\cdot)$ n'est pas continu sur l'intervalle de temps $[0, \mathcal{T}]$, le schéma d'intégration numérique doit être relancé sur chaque sous-intervalle $[t_k, t_{k+1}]$ en repartant des derniers états calculés $(\overline{\mathbf{x}}(t_1), \ldots, \overline{\mathbf{x}}(t_{\mathcal{N}-1}))$ aux instants $(t_1, \ldots, t_{\mathcal{N}-1})$ comme suit :

$$\forall\, \tau \in [t_k, t_{k+1}],\ \overline{\mathbf{x}}(\tau) = \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{x}}(t_k), \overline{\mathbf{u}}(\tau)\Big) \text{ est la solution de : } \begin{cases} \mathbf{F}\Big(\dot{\overline{\mathbf{x}}}(\tau), \overline{\mathbf{x}}(\tau), \overline{\mathbf{u}}(\tau)\Big) = \mathbf{0} \\ \overline{\mathbf{x}}(0) = \overline{\mathbf{x}}(t_k)\,. \end{cases} \tag{1.95}$$

La trajectoire d'état peut ensuite être reconstruite sur l'intervalle de temps d'origine $[0, \mathcal{T}]$ en éliminant de façon récursive les états intermédiaires $(\overline{\mathbf{x}}(t_1), \ldots, \overline{\mathbf{x}}(t_{\mathcal{N}-1}))$ qui, par continuité, relient les sous-intervalles entre eux :

$$\begin{aligned}
&\forall\, \tau \in [t_0, t_1],\ \overline{\mathbf{x}}(\tau) = \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{x}}(t_0), \overline{\mathbf{u}}(\tau)\Big) = \overline{\mathbf{\Phi}}\Big(\tau; \mathbf{x}(t), \overline{\mathbf{u}}(\tau)\Big) \\
&\forall\, \tau \in [t_1, t_2],\ \overline{\mathbf{x}}(\tau) = \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{x}}(t_1), \overline{\mathbf{u}}(\tau)\Big) = \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{\Phi}}\Big(t_1; \mathbf{x}(t), \overline{\mathbf{u}}(t_1)\Big), \overline{\mathbf{u}}(\tau)\Big) \\
&\forall\, \tau \in [t_2, t_3],\ \overline{\mathbf{x}}(\tau) = \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{x}}(t_2), \overline{\mathbf{u}}(\tau)\Big) = \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{\Phi}}\Big(t_2; \overline{\mathbf{\Phi}}\Big(t_1; \mathbf{x}(t), \overline{\mathbf{u}}(t_1)\Big), \overline{\mathbf{u}}(t_2)\Big), \overline{\mathbf{u}}(\tau)\Big) \\
&\quad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots
\end{aligned} \tag{1.96}$$

ou, de manière plus concise :

$$\begin{aligned}
&\forall\, \tau \in [t_k, t_{k+1}],\ \overline{\mathbf{x}}(\tau) = \overline{\mathbf{\Phi}}_{\mathbf{u}}\Big(\tau; \mathbf{x}(t)\Big), \text{ avec} \\
&\overline{\mathbf{\Phi}}_{\mathbf{u}}\Big(\tau; \mathbf{x}(t)\Big) \coloneqq \overline{\mathbf{\Phi}}\Big(\tau; \overline{\mathbf{\Phi}}\Big(t_k;\ \ldots; \overline{\mathbf{\Phi}}\Big(t_2; \overline{\mathbf{\Phi}}\Big(t_1; \mathbf{x}(t), \overline{\mathbf{u}}(t_1)\Big), \overline{\mathbf{u}}(t_2)\Big), \ldots, \overline{\mathbf{u}}(t_k)\Big), \overline{\mathbf{u}}(\tau)\Big)\,.
\end{aligned} \tag{1.97}$$

En pratique, les sous-intervalles $[t_k, t_{k+1}]$ sont à leur tour subdivisés en une grille d'instants $t_k = \tau_{k,0} < \cdots < \tau_{k,n} \leq t_{k+1}$ à laquelle est rattachée une suite de vecteurs $(\mathbf{x}_{k,i})_{i=0}^{n}$ qui approche numériquement l'état du modèle. Typiquement, pour une grille d'instants uniforme :

$$\forall\, k \in [\![0, \mathcal{N}-1]\!],\ \forall\, i \in [\![0, n]\!],\ \tau_{k,i} = t_k + \frac{i}{n}(t_{k+1} - t_k),\ \text{avec } \mathbf{x}_{k,i} \approx \overline{\mathbf{x}}(\tau_{k,i})\,. \tag{1.98}$$

Par conséquent, la fonction coût et les contraintes du problème de commande optimale (1.27) peuvent être approchées soit de manière fine :

$$\begin{aligned}
&J_{\mathcal{T}}\Big(\mathbf{x}(t), \overline{\mathbf{u}}(\cdot)\Big) \approx V_f\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(t_{\mathcal{N}}; \mathbf{x}(t))\Big) + \sum_{k=0}^{\mathcal{N}-1} \sum_{i=0}^{n-1} L\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(\tau_{k,i}; \mathbf{x}(t)), \overline{\mathbf{u}}(\tau_{k,i})\Big)(\tau_{k,i+1} - \tau_{k,i}) \\
&\text{s.c} : \forall\, k \in [\![0, \mathcal{N}-1]\!],\ \forall\, i \in [\![0, n]\!],\ \Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(\tau_{k,i}; \mathbf{x}(t)), \overline{\mathbf{u}}(\tau_{k,i})\Big) \in \mathbb{Y} \text{ et } \overline{\mathbf{\Phi}}_{\mathbf{u}}(t_{\mathcal{N}}; \mathbf{x}(t)) \in \mathbb{X}_f\,,
\end{aligned} \tag{1.99}$$

soit de manière plus grossière :

$$J_{\mathcal{T}}\Big(\mathbf{x}(t),\overline{\mathbf{u}}(\cdot)\Big) \approx V_f\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(t_{\mathcal{N}};\mathbf{x}(t))\Big) + \sum_{k=0}^{\mathcal{N}-1} L\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(t_k;\mathbf{x}(t)),\overline{\mathbf{u}}(t_k)\Big)(t_{k+1}-t_k)$$
$$\text{sous contraintes : } \forall\, k\in[\![0,\mathcal{N}-1]\!],\ \Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(t_k;\mathbf{x}(t)),\overline{\mathbf{u}}(t_k)\Big)\in\mathbb{Y} \text{ et } \overline{\mathbf{\Phi}}_{\mathbf{u}}(t_{\mathcal{N}};\mathbf{x}(t))\in\mathbb{X}_f\,, \quad (1.100)$$

selon le type de grille sélectionnée. Dans les faits, les deux niveaux de grilles sont utilisées conjointement, la fonction coût étant définie sur celle ayant le maillage le plus fin et les contraintes sur celle ayant le maillage le plus grossier. Ainsi, transcrire le problème de commande optimale (1.27) avec une méthode de tir simple mène au problème d'optimisation suivant :

$$\min_{\mathbf{w_v}} \quad V_f\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(t_{\mathcal{N}};\mathbf{x}(t))\Big) + \sum_{k=0}^{\mathcal{N}-1}\sum_{i=0}^{n-1} L\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(\tau_{k,i};\mathbf{x}(t)),\mathbf{v}_k\Big)(\tau_{k,i+1}-\tau_{k,i})$$
$$\text{sous contraintes : } \forall\, k\in[\![0,\mathcal{N}-1]\!],\ \Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(t_k;\mathbf{x}(t)),\mathbf{v}_k\Big)\in\mathbb{Y} \text{ et } \overline{\mathbf{\Phi}}_{\mathbf{u}}(t_{\mathcal{N}};\mathbf{x}(t))\in\mathbb{X}_f\,, \quad (1.101)$$

où $\mathbf{w_v} \coloneqq [\mathbf{v}_0^\top \cdots \mathbf{v}_{\mathcal{N}-1}^\top]^\top \in \mathbb{R}^{\mathcal{N}n_{\mathbf{u}}}$. Par ailleurs, il est intéressant de souligner ici que le terme intégral :

$$V_L\Big(\mathcal{T};\overline{\mathbf{x}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big) \coloneqq \int_0^{\mathcal{T}} L\Big(\overline{\mathbf{x}}(\tau),\overline{\mathbf{u}}(\tau)\Big)\,\mathrm{d}\tau\,, \quad (1.102)$$

de la fonction coût d'origine $J_{\mathcal{T}}(\cdot)$ peut être approché autrement qu'avec une somme de Riemann :

$$V_L\Big(\mathcal{T};\overline{\mathbf{\Phi}}_{\mathbf{u}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big) \approx \sum_{k=0}^{\mathcal{N}-1}\sum_{i=0}^{n-1} L\Big(\overline{\mathbf{\Phi}}_{\mathbf{u}}(\tau_{k,i};\mathbf{x}(t)),\overline{\mathbf{u}}(\tau_{k,i})\Big)(\tau_{k,i+1}-\tau_{k,i})\,. \quad (1.103)$$

Une solution élégante, qui fait écho au premier paragraphe de la section précédente, est d'augmenter l'état du modèle en lui ajoutant l'équation différentielle :

$$\dot{V}_L\Big(\tau;\overline{\mathbf{x}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big) = L\Big(\overline{\mathbf{x}}(\tau),\overline{\mathbf{u}}(\tau)\Big),\ V_L\Big(0;\overline{\mathbf{x}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big) = 0\,, \quad (1.104)$$

pour obtenir $V_L(\cdot)$ à l'aide du schéma d'intégration numérique ayant servi à calculer la trajectoire d'état du modèle :

$$\forall\,\tau\in[0,\mathcal{T}],\ \Big[\overline{\mathbf{x}}(\tau)^\top\ V_L\Big(\tau;\overline{\mathbf{x}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big)\Big]^\top = \overline{\mathbf{\Phi}}_{\mathbf{u}}\Big(\tau;[\mathbf{x}(t)^\top\ 0]^\top\Big) \text{ est la solution de :}$$
$$\begin{bmatrix} \mathbf{F}\Big(\dot{\overline{\mathbf{x}}}(\tau),\overline{\mathbf{x}}(\tau),\overline{\mathbf{u}}(\tau)\Big) \\ \dot{V}_L\Big(\tau;\overline{\mathbf{x}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big) - L\Big(\overline{\mathbf{x}}(\tau),\overline{\mathbf{u}}(\tau)\Big)\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\0\end{bmatrix},\ \text{avec}\ \begin{bmatrix}\overline{\mathbf{x}}(0)\\ V_L\Big(0;\overline{\mathbf{x}}(\cdot),\overline{\mathbf{u}}(\cdot)\Big)\end{bmatrix} = \begin{bmatrix}\mathbf{x}(t)\\0\end{bmatrix} \quad (1.105)$$

L'avantage du tir simple est que le problème d'optimisation (1.101) ne contient pas les contraintes d'évolution du modèle, puisque celles-ci sont éliminées récursivement par le schéma

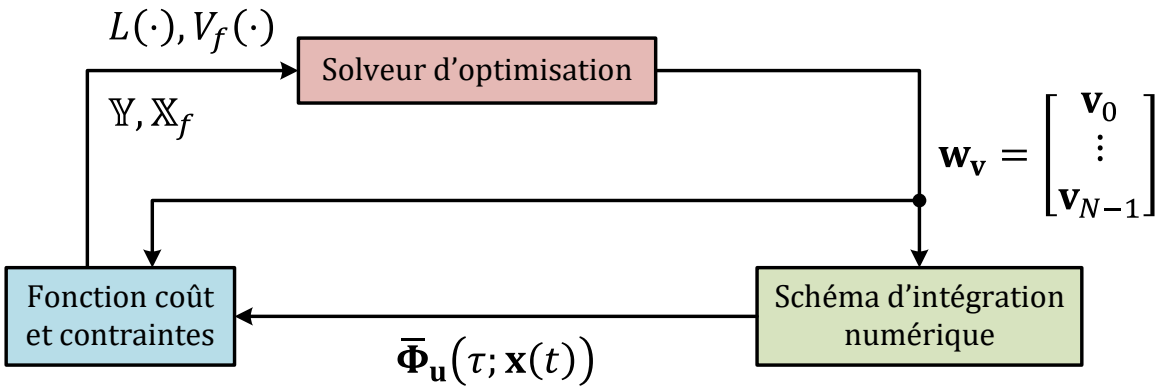

FIGURE 1.9 – Échanges de données entre les différents modules de l'algorithme de commande prédictive lorsque le problème est transcrit à l'aide d'une méthode de tir simple.

d'intégration numérique (1.97). De même, les variables d'état du modèle sont absentes du problème final, car le schéma d'intégration numérique (1.97) est perçu comme une boîte noire par le solveur d'optimisation. De ce fait, un autre avantage du tir simple est que les variables de décision du problème d'optimisation se réduisent à celles utilisées pour représenter le signal de commande. De plus, contrairement aux autres méthodes de transcription, la trajectoire d'état renvoyée par le solveur d'optimisation sera toujours cohérente avec les équations d'évolution du modèle, pourvu que le schéma d'intégration numérique fonctionne correctement. Néanmoins, l'évaluation répétée des équations du modèle par le schéma d'intégration numérique rend la fonction $\overline{\mathbf{\Phi}}_{\mathbf{u}}(\cdot)$ de plus en plus complexe, et de plus en plus sensible aux conditions initiales, à mesure que le temps de simulation avance. L'inconvénient, en tir simple, est que ces caractéristiques indésirables (complexité et sensibilité aux conditions initiales) vont forcément se retrouver dans la fonction coût et dans les contraintes du problème d'optimisation (1.101). Les effets négatifs qui en découlent sont, de surcroît, sensiblement amplifiés lorsque l'horizon de prédiction s'allonge, et lorsque le modèle est non-linéaire et/ou instable. Par conséquent, les contraintes qui s'appliquent sur l'état du modèle, notamment l'état final, sont particulièrement difficiles à traiter en tir simple. Toutes ces raisons font que la vitesse de convergence du solveur d'optimisation reste relativement lente en tir simple, bien que les contraintes d'évolution du modèle aient été éliminées du problème.

3.2.3 Méthode de tir multiple

Pour pallier les inconvénients du tir simple, une deuxième méthode de transcription, appelée tir multiple, a été mise au point. Au lieu de simuler d'un seul coup le modèle sur tout l'intervalle de temps $[0, \mathcal{T}]$, l'idée du tir multiple est de calculer séparément sa trajectoire d'état sur chaque

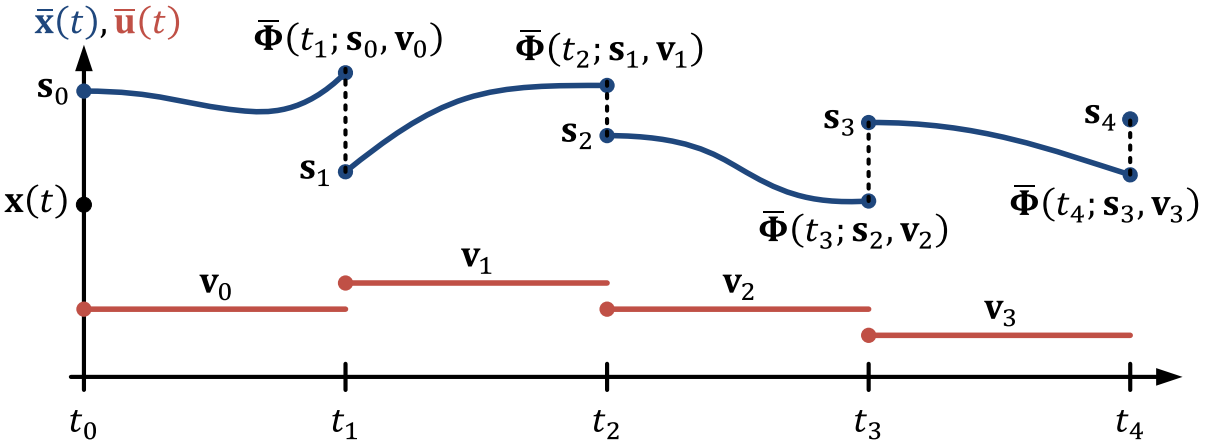

FIGURE 1.10 – Exemple de trajectoire d'état et de signal de commande obtenus à la suite de la transcription du problème de commande optimale par une méthode de tir multiple.

sous-intervalle $[t_k, t_{k+1}]$ en partant de conditions initiales $\mathbf{s}_k \in \mathbb{R}^{n_\mathbf{x}}$ artificielles :

$$\forall\, \tau \in [t_k, t_{k+1}],\ \bar{\mathbf{x}}(\tau) = \bar{\mathbf{\Phi}}\Big(\tau; \mathbf{s}_k, \bar{\mathbf{u}}(\tau)\Big) \text{ est la solution de : } \begin{cases} \mathbf{F}\Big(\dot{\bar{\mathbf{x}}}(\tau), \bar{\mathbf{x}}(\tau), \bar{\mathbf{u}}(\tau)\Big) = \mathbf{0} \\ \bar{\mathbf{x}}(0) = \mathbf{s}_k\,. \end{cases} \tag{1.106}$$

Ces états initiaux factices appartiennent à un vecteur de nœuds $\mathbf{w_s} \coloneqq [\mathbf{s}_0^\top \cdots \mathbf{s}_\mathcal{N}^\top]^\top \in \mathbb{R}^{(N+1)n_\mathbf{x}}$ qui sera, par la suite, ajouté aux variables de décision du problème d'optimisation afin de recoller les différents morceaux de la trajectoire d'état. Sa continuité, de même que sa cohérence initiale vis-à-vis de l'état mesuré sur le système réel, sont alors assurées grâce aux contraintes d'égalité :

$$\forall\, k \in [\![0, \mathcal{N}-1]\!],\ \mathbf{s}_{k+1} - \bar{\mathbf{\Phi}}\Big(t_{k+1}; \mathbf{s}_k, \mathbf{v}_k\Big) = \mathbf{0} \quad \text{et} \quad \mathbf{s}_0 - \mathbf{x}(t) = \mathbf{0}\,. \tag{1.107}$$

De ce fait, transcrire le problème de commande optimale (1.27) avec une méthode de tir multiple permet, cette fois-ci, d'aboutir au problème d'optimisation suivant :

$$\begin{aligned} \min_{\mathbf{w_s}, \mathbf{w_v}} \quad & V_f(\mathbf{s}_\mathcal{N}) + \sum_{k=0}^{\mathcal{N}-1} \sum_{i=0}^{n-1} L\Big(\bar{\mathbf{\Phi}}(\tau_{k,i}; \mathbf{s}_k, \mathbf{v}_k), \mathbf{v}_k\Big)(\tau_{k,i+1} - \tau_{k,i}) \\ \text{sous contraintes : } & \begin{cases} \forall\, k \in [\![0, \mathcal{N}-1]\!],\ (\mathbf{s}_k, \mathbf{v}_k) \in \mathbb{Y} \text{ et } \mathbf{s}_\mathcal{N} \in \mathbb{X}_f \\ \forall\, k \in [\![0, \mathcal{N}-1]\!],\ \mathbf{s}_{k+1} - \bar{\mathbf{\Phi}}\Big(t_{k+1}; \mathbf{s}_k, \mathbf{v}_k\Big) = \mathbf{0} \\ \mathbf{s}_0 - \mathbf{x}(t) = \mathbf{0}\,. \end{cases} \end{aligned} \tag{1.108}$$

L'avantage du tir multiple est que la complexité, ainsi que la sensibilité aux conditions initiales, de la fonction coût et des contraintes du problème d'optimisation (1.108) sont fortement réduites par rapport au tir simple, puisque le schéma d'intégration numérique n'est appelé que

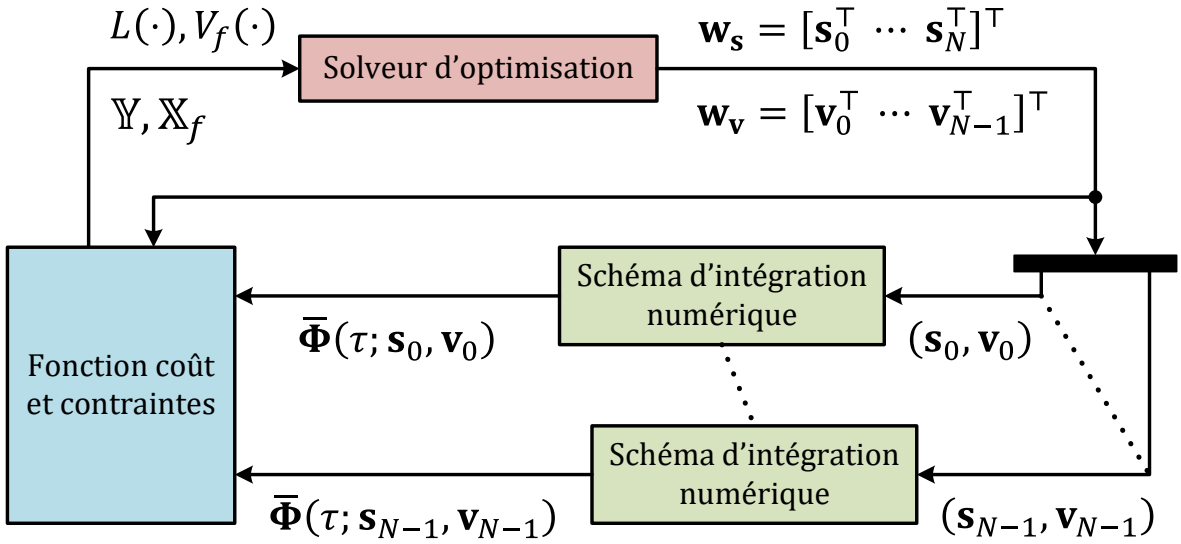

FIGURE 1.11 – Échanges de données entre les différents modules de l'algorithme de commande prédictive lorsque le problème est transcrit à l'aide d'une méthode de tir multiple.

sur les sous-intervalles $[t_k, t_{k+1}]$. Autrement dit, les potentielles non-linéarités de la fonction $\overline{\mathbf{\Phi}}(\cdot)$, qui peuvent apparaître lorsque le modèle est simulé trop longtemps, sont non seulement atténuées, mais également réparties de manière uniforme le long de l'horizon prédiction $[0, \mathcal{T}]$. En revanche, la dimension du problème d'optimisation augmente significativement, étant donné que le nombre total de variables de décision passe de $\mathcal{N} n_{\mathbf{u}}$ variables en tir simple à $(\mathcal{N}+1)n_{\mathbf{x}} + \mathcal{N} n_{\mathbf{u}}$ variables en tir multiple. De plus, $(\mathcal{N}+1)n_{\mathbf{x}}$ contraintes d'évolution sont ajoutées au problème d'optimisation pour assurer la cohérence de la trajectoire d'état sur tout l'horizon de prédiction. Toutefois, ces deux inconvénients peuvent être facilement contournés en agençant les variables de décision et les contraintes du problème de façon intelligente. En effet, comme le modèle est simulé séparément sur chaque sous-intervalle $[t_k, t_{k+1}]$, les contraintes qui s'appliquent sur le nœud $\mathbf{s}_{k+1}$ ne dépendront que du nœud précédent $\mathbf{s}_k$ et de la variable d'entrée $\mathbf{v}_k$. Par conséquent, si les variables de décision et les contraintes sont déclarées dans le même ordre que les composantes du vecteur $\mathbf{w_{sv}} \coloneqq [\mathbf{s}_0^\top \ \mathbf{v}_0^\top \cdots \mathbf{s}_{\mathcal{N}-1}^\top \ \mathbf{v}_{\mathcal{N}-1}^\top \ \mathbf{s}_{\mathcal{N}}^\top]^\top$, alors les matrices aux dérivées partielles (jacobiennes et hessiennes) du problème d'optimisation (1.108) seront diagonales par blocs. Cette structure creuse, c'est-à-dire constituée essentiellement de zéros, peut être efficacement exploitée par certains solveurs d'optimisation afin d'améliorer la vitesse de convergence de leurs algorithmes. De plus, le fait d'inclure les nœuds $\mathbf{s}_k$ dans les variables de décision du problème permet de mieux initialiser ce dernier, puisqu'il est désormais possible de lui fournir une bonne première estimation de la trajectoire d'état optimale en réutilisant celle calculée à l'itération précédente. Enfin, étant donné que la simulation du modèle est découplée d'un sous-

intervalle à l'autre, les appels à la fonction $\overline{\mathbf{\Phi}}(\cdot)$ peuvent être parallélisés pour accélérer, encore davantage, la convergence du solveur d'optimisation. Ainsi, malgré l'augmentation du nombre de variables de décision et de contraintes qui en résulte, l'emploi du tir multiple est souvent préférable à celui du tir simple, en particulier lorsque le solveur d'optimisation peut détecter et exploiter la structure creuse des matrices du problème (1.108). À noter cependant que, même en tir multiple, le solveur d'optimisation reste tributaire du bon fonctionnement du schéma d'intégration numérique pour calculer la trajectoire d'état du modèle.

3.2.4 Méthode simultanée (par collocation)

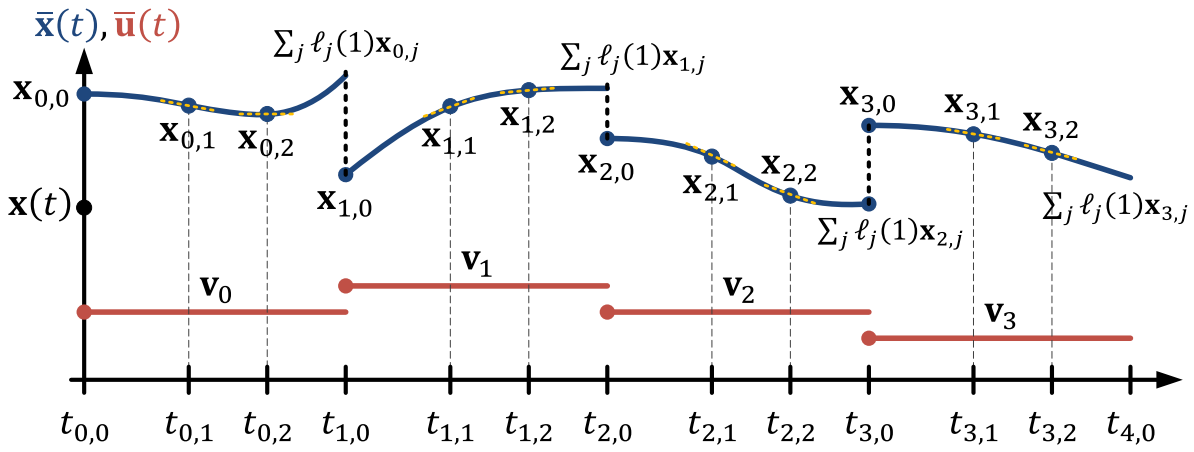

FIGURE 1.12 – Exemple de trajectoire d'état et de signal de commande obtenus à la suite de la transcription du problème de commande optimale par une méthode simultanée.

Plutôt que de considérer le schéma d'intégration numérique comme une boîte noire appelée en marge du solveur d'optimisation, une troisième méthode de transcription, dite simultanée, propose de fusionner les deux modules. Le principe d'une méthode de transcription simultanée est d'inclure l'équation aux différences et les variables internes du schéma d'intégration numérique, auparavant dissimulées au sein de la fonction $\overline{\mathbf{\Phi}}(\cdot)$, dans les contraintes et les variables de décision du problème d'optimisation. De cette manière, le calcul de la suite de commandes optimale et celui de sa trajectoire d'état associée peuvent être réalisés simultanément par le solveur d'optimisation, sans faire appel au schéma d'intégration numérique. Pour y parvenir, ce dernier est généralement remplacé par une méthode de collocation de type Gauss-Legendre ou Gauss-Radau (1.69), dont la formulation a été légèrement modifiée afin de pouvoir être mieux incorporée aux contraintes du problème. Cette fois-ci, le polynôme de collocation, et non plus sa dérivée

temporelle, est projeté dans une base de polynômes de Lagrange :

$$\forall\, \tau \in [t_k, t_{k+1}],\ \overline{\boldsymbol{\chi}}_k(\tau) = \sum_{j=0}^{s} \ell_j\Big(\frac{\tau - t_k}{t_{k+1} - t_k}\Big)\mathbf{x}_{k,j},\ \text{avec}\ \ell_j(c) \coloneqq \prod_{\substack{r=0 \\ r\neq j}}^{s}\Big(\frac{c - c_r}{c_j - c_r}\Big)\,, \tag{1.109}$$

qui, à la différence de (1.66), dépendent explicitement des $s \in \mathbb{N}_{>0}$ nœuds $0 < c_1 < \cdots < c_s \leq 1$ de la formule de quadrature de Gauss, ainsi que d'un nœud supplémentaire $c_0 = 0$. De ce fait, sachant que :

$$\forall\, i \in [\![0, s]\!],\ \forall\, j \in [\![0, s]\!],\ \ell_j(c_i) = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases} \Rightarrow\ \mathbf{x}_{k,i} = \overline{\boldsymbol{\chi}}_k(\tau_{k,i}) \approx \overline{\mathbf{x}}(\tau_{k,i})\,, \tag{1.110}$$

les coefficients $(\mathbf{x}_{k,0}, \ldots, \mathbf{x}_{k,s}) \in \mathbb{R}^{(s+1)n_{\mathbf{x}}}$ du polynôme de collocation approchent numériquement l'état du modèle aux instants intermédiaires $t_k = \tau_{k,0} < \cdots < \tau_{k,s} \leq t_{k+1}$ définis par les nœuds de la méthode :

$$\forall\, i \in [\![0, s]\!],\ \tau_{k,i} = t_k + c_i(t_{k+1} - t_k)\,. \tag{1.111}$$

Sur chaque sous-intervalle $[t_k, t_{k+1}]$, le polynôme $\overline{\boldsymbol{\chi}}_k(\cdot)$ doit respecter les conditions suivantes :

$$\overline{\boldsymbol{\chi}}_k(t_k) = \mathbf{x}_{k,0} \quad \text{et} \quad \mathbf{F}\Big(\dot{\overline{\boldsymbol{\chi}}}_k(\tau_{k,i}), \overline{\boldsymbol{\chi}}_k(\tau_{k,i}), \overline{\mathbf{u}}(\tau_{k,i})\Big) = \mathbf{0} \text{ pour tout } i \in [\![1, s]\!]\,. \tag{1.112}$$

De plus, comme en tir multiple, des contraintes de recollement sont ajoutées au problème d'optimisation pour garantir la continuité de la trajectoire d'état sur l'intervalle de temps $[0, \mathcal{T}]$, et s'assurer que celle-ci part bien de l'état initial mesuré sur le système réel :

$$\forall\, k \in [\![0, \mathcal{N} - 2]\!],\ \overline{\boldsymbol{\chi}}_{k+1}(t_{k+1}) - \overline{\boldsymbol{\chi}}_k(t_{k+1}) = \mathbf{0} \quad \text{et} \quad \overline{\boldsymbol{\chi}}_0(t_0) - \mathbf{x}(t) = \mathbf{0}\,. \tag{1.113}$$

En outre, le terme intégral (1.102) de la fonction coût d'origine peut être approché en réutilisant la formule de quadrature de Gauss (1.71) ayant permis de déterminer l'équation aux différences de la méthode de collocation :

$$\int_{t_k}^{t_{k+1}} L\Big(\overline{\mathbf{x}}(\tau), \overline{\mathbf{u}}(\tau)\Big)\,\mathrm{d}\tau \approx (t_{k+1} - t_k)\sum_{i=1}^{s}\left(\int_0^1 \ell_i(c)\,\mathrm{d}c\right) L\Big(\overline{\mathbf{x}}(\tau_{k,i}), \overline{\mathbf{u}}(\tau_{k,i})\Big)\,. \tag{1.114}$$

Enfin, étant donné que :

$$\forall\, \tau \in [t_k, t_{k+1}],\ \dot{\overline{\boldsymbol{\chi}}}_k(\tau) = \frac{1}{t_{k+1} - t_k}\sum_{j=0}^{s}\frac{\mathrm{d}\ell_j}{\mathrm{d}\tau}\Big(\frac{\tau - t_k}{t_{k+1} - t_k}\Big)\mathbf{x}_{k,j} \quad \text{et} \quad \overline{\boldsymbol{\chi}}_k(t_{k+1}) = \sum_{j=0}^{s}\ell_j(1)\mathbf{x}_{k,j}\,, \tag{1.115}$$

la transcription simultanée, par collocation, du problème de commande optimale (1.27) conduit

au problème d'optimisation suivant :

$$\min_{\mathbf{w_x},\mathbf{w_v}} \quad V_f\left(\sum_{j=0}^{s}\ell_j(1)\mathbf{x}_{s,j}\right)+\sum_{k=0}^{\mathcal{N}-1}(t_{k+1}-t_k)\sum_{i=1}^{s}\left(\int_0^1\ell_i(c)\,\mathrm{d}c\right)L\left(\mathbf{x}_{k,i},\mathbf{v}_k\right)$$

$$\text{sous contraintes :}\begin{cases}\forall\, k\in[\![0,\mathcal{N}-1]\!],\ \forall\, i\in[\![0,s]\!],\ (\mathbf{x}_{k,i},\mathbf{v}_k)\in\mathbb{Y}\text{ et }\displaystyle\sum_{j=0}^{s}\ell_j(1)\mathbf{x}_{s,j}\in\mathbb{X}_f\\ \forall\, k\in[\![0,\mathcal{N}-1]\!],\ \forall\, i\in[\![1,s]\!],\ \mathbf{F}\left(\dfrac{1}{t_{k+1}-t_k}\displaystyle\sum_{j=0}^{s}\dfrac{\mathrm{d}\ell_j(c_i)}{\mathrm{d}\tau}\mathbf{x}_{k,j},\mathbf{x}_{k,i},\mathbf{v}_k\right)=\mathbf{0}\\ \forall\, k\in[\![0,\mathcal{N}-2]\!],\ \mathbf{x}_{k+1,0}-\displaystyle\sum_{j=0}^{s}\ell_j(1)\mathbf{x}_{k,j}=\mathbf{0}\\ \mathbf{x}_{0,0}-\mathbf{x}(t)=\mathbf{0}\,.\end{cases}\tag{1.116}$$

où $\mathbf{w_x}\coloneqq[\mathbf{x}_{0,0}^\top\cdots\mathbf{x}_{0,s}^\top\cdots\mathbf{x}_{\mathcal{N}-1,0}^\top\cdots\mathbf{x}_{\mathcal{N}-1,s}^\top]^\top\in\mathbb{R}^{\mathcal{N}(s+1)n_\mathbf{x}}$.

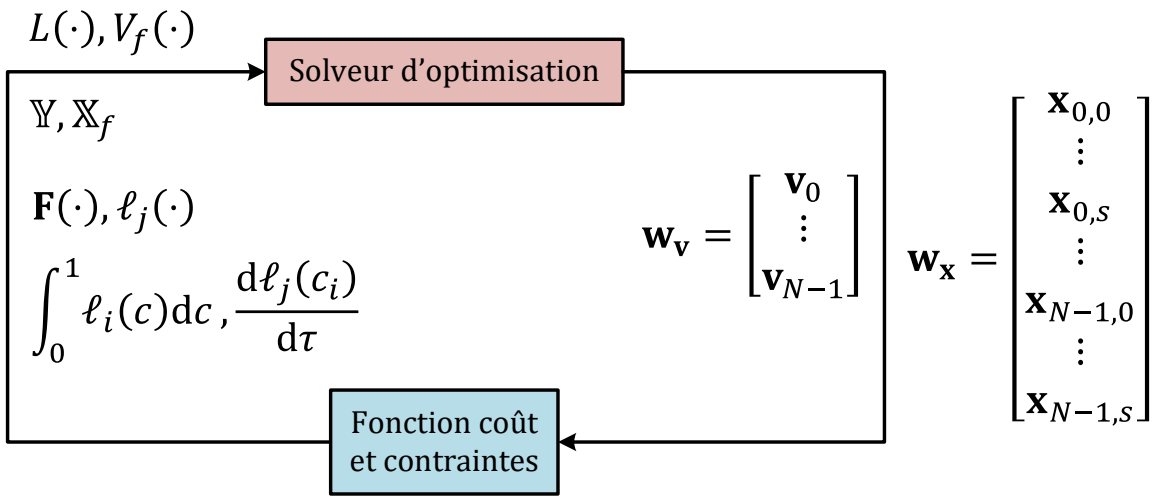

FIGURE 1.13 – Échanges de données entre les différents modules de l'algorithme de commande prédictive lorsque le problème est transcrit à l'aide d'une méthode simultanée (ici collocation).

Comme mentionné précédemment, l'avantage des méthodes simultanées est de pouvoir calculer la trajectoire d'état du modèle en même temps que la suite de commande optimale, et ce, en faisant uniquement appel au solveur d'optimisation. En revanche, lorsque le schéma d'intégration numérique est remplacé par une méthode de collocation de type Gauss-Legendre ou Gauss-Radau de rang $s\in\mathbb{N}_{>0}$, le nombre total de variables de décision et de contraintes d'évolution du problème d'optimisation passe de $(\mathcal{N}+1)n_\mathbf{x}+\mathcal{N}n_\mathbf{u}$ variables et $(\mathcal{N}+1)n_\mathbf{x}$ contraintes en tir multiple, à $\mathcal{N}(s+1)n_\mathbf{x}+\mathcal{N}n_\mathbf{u}$ variables et $\mathcal{N}(s+1)n_\mathbf{x}$ contraintes en trans-

cription simultanée. Heureusement, cet inconvénient peut de nouveau être surmonté en déclarant les variables de décision et les contraintes du problème suivant l'ordre des composantes du vecteur $\mathbf{w_{xu}} \coloneqq [\mathbf{x}_{0,0}^\top \cdots \mathbf{x}_{0,s}^\top \ \mathbf{v}_0^\top \cdots \mathbf{x}_{\mathcal{N}-1,0}^\top \cdots \mathbf{x}_{\mathcal{N}-1,s}^\top \ \mathbf{v}_{\mathcal{N}-1}^\top]^\top$. En effet, puisque les variables internes du schéma d'intégration numérique apparaissent directement dans le problème d'optimisation (1.116), la structure de ses matrices aux dérivées partielles sera beaucoup plus creuse et clairsemée qu'en tir multiple. De même, les potentielles non-linéarités qui peuvent découler de la simulation du modèle seront davantage atténuées en utilisant une méthode de transcription simultanée plutôt qu'une méthode de tir multiple.

3.2.5 Résumé : caractéristiques des différentes méthodes de transcription

Le problème de commande optimale de l'algorithme de commande prédictive est généralement transformé en un problème d'optimisation numérique en choisissant un signal de commande constant par morceaux et en utilisant l'une des trois méthodes de transcription suivantes :

- **méthode de tir simple** : la trajectoire d'état du modèle est calculée d'un seul coup, en dehors du problème d'optimisation, du début à la fin de l'horizon de prédiction.
  - \+ méthode facile à comprendre et rapide à mettre en place.
  - \+ satisfaction implicite des contraintes d'évolution du modèle.
  - \+ variables de décision limitées à celles utilisées pour représenter le signal de commande.
  - − fonction coût et contraintes du problème d'optimisation de plus en plus complexes et sensibles aux conditions initiales à mesure que l'horizon de prédiction s'allonge.
  - − difficulté à traiter les modèles non-linéaires et/ou instables, ainsi que les contraintes s'appliquant sur les variables d'état.
  - − solveur d'optimisation tributaire du bon fonctionnement du schéma d'intégration numérique appelé en aval.
- **méthode de tir multiple** : la trajectoire d'état du modèle est calculée séparément, en dehors du problème d'optimisation, sur chaque sous-intervalle utilisé pour définir le signal de commande.
  - \+ fonction coût et contraintes du problème d'optimisation beaucoup moins complexes et sensibles aux conditions initiales qu'en tir simple.
  - \+ possibilité d'initialiser le problème d'optimisation avec une partie de la trajectoire d'état calculée précédemment.
  - \+ possibilité de paralléliser les appels au schéma d'intégration numérique.
  - − augmentation du nombre de contraintes et de variables de décision du problème d'optimisation.

  - − nécessite de mettre en évidence et de pouvoir exploiter la structure creuse des matrices aux dérivées partielles (jacobiennes et hessiennes) du problème d'optimisation.
  - − solveur d'optimisation toujours tributaire du bon fonctionnement du schéma d'intégration numérique appelé en aval.
- **méthode simultanée** : la trajectoire d'état du modèle est calculée en même temps que le signal de commande, en incorporant l'équation aux différences et les variables internes du schéma d'intégration numérique aux contraintes et aux variables de décision du problème d'optimisation.
  - + trajectoire d'état directement calculée par le solveur d'optimisation.
  - + structure des matrices aux dérivées partielles du problème d'optimisation beaucoup plus creuse et clairsemée qu'en tir multiple.
  - + possibilité d'initialiser le problème d'optimisation avec une partie de la trajectoire d'état calculée précédemment.
  - − augmentation encore plus importante du nombre de contraintes et de variables de décision du problème d'optimisation.
  - − nécessite toujours de mettre en évidence et de pouvoir exploiter la structure creuse des matrices aux dérivées partielles du problème d'optimisation.

3.3 Résolution du problème d'optimisation

La dernière étape pour mettre en place l'algorithme de commande prédictive est de résoudre le problème d'optimisation paramétrique [20], [24], [65], [66] obtenu à la suite de la transcription directe du problème de commande optimale (1.27). La forme générale d'un problème d'optimisation paramétrique est donnée par :

$$\min_{\mathbf{w}} \; J(\mathbf{w};\mathbf{p}) \;\; \text{sous contraintes} : \begin{cases} \mathbf{g}(\mathbf{w};\mathbf{p}) \leq \mathbf{0} \\ \mathbf{h}(\mathbf{w};\mathbf{p}) = \mathbf{0}\,, \end{cases} \tag{1.117}$$

où $\mathbf{w} \in \mathbb{R}^{n_\mathbf{w}}$ est le vecteur des variables de décision, $\mathbf{p} \in \mathbb{R}^{n_\mathbf{p}}$ est un vecteur de paramètres défini en amont du problème, et $J : \mathbb{R}^{n_\mathbf{w}} \times \mathbb{R}^{n_\mathbf{p}} \to \mathbb{R}$ est la fonction coût à minimiser. Lorsqu'elles existent, les fonctions $\mathbf{g} : \mathbb{R}^{n_\mathbf{w}} \times \mathbb{R}^{n_\mathbf{p}} \to \mathbb{R}^{n_\mathbf{g}}$ et $\mathbf{h} : \mathbb{R}^{n_\mathbf{w}} \times \mathbb{R}^{n_\mathbf{p}} \to \mathbb{R}^{n_\mathbf{h}}$ représentent les contraintes d'inégalité et les contraintes d'égalité du problème. Il est admis, dans ce qui suit, que toutes les fonctions du problème d'optimisation (1.117) sont au moins de classe $\mathcal{C}^2$ sur leur ensemble de définition respectif. De plus, il est également admis que l'ensemble fermé $\mathbb{Y} \subseteq \mathbb{R}^{n_\mathbf{x}} \times \mathbb{R}^{n_\mathbf{u}}$ et l'ensemble compact $\mathbb{X}_f \subseteq \mathbb{R}^{n_\mathbf{x}}$ peuvent toujours être représentés par un nombre fini d'inégalités et d'égalités, ce qui est le cas, par exemple, si ces derniers sont des polyèdres ou des produits cartésiens finis d'intervalles fermés (bornés ou non). Enfin, étant donné que le problème d'optimisation (1.117) est issu de la transcription du problème de commande optimale (1.27), seules

les contraintes d'égalité $\mathbf{h}(\cdot)$ dépendent véritablement du vecteur de paramètres $\mathbf{p}$, et ce, de surcroît, de façon linéaire :

$$\min_{\mathbf{w}} \; J(\mathbf{w}) \;\; \text{sous contraintes} : \begin{cases} \mathbf{g}(\mathbf{w}) \leq \mathbf{0} \\ \mathbf{h}(\mathbf{w}; \mathbf{p}) = \mathbf{0} \, . \end{cases} \tag{1.118}$$

En effet, dans le cadre d'un algorithme de commande prédictive, le vecteur de paramètres $\mathbf{p}$ correspond à l'état initial $\overline{\mathbf{x}}(0) = \mathbf{x}(t)$ mesuré périodiquement sur le système réel. Or, celui-ci n'apparaît dans le problème d'optimisation qu'à travers les contraintes d'égalité $\mathbf{s}_0 - \mathbf{x}(t) = \mathbf{0}$ en tir multiple (1.108) et $\mathbf{x}_{0,0} - \mathbf{x}(t) = \mathbf{0}$ en transcription simultanée (1.116). De plus, même lorsque la fonction coût et/ou les contraintes d'inégalité dépendent de l'état mesuré, comme en tir simple (1.101), il est toujours possible de revenir au cas précédent en procédant de manière analogue au tir multiple, c'est-à-dire en ajoutant une variable de décision artificielle $\mathbf{s}_0 \in \mathbb{R}^{n_\mathbf{x}}$ au problème d'optimisation, qui sera liée à l'état mesuré par la contrainte $\mathbf{s}_0 - \mathbf{x}(t) = \mathbf{0}$. Cette astuce, dite « d'incorporation de l'état initial » [21], [23], est souvent employée dans les systèmes temps réel afin de générer rapidement une nouvelle commande linéarisée à partir de celle calculée précédemment [52], [55], [56], par exemple :

$$\mathbf{v}_0'(\mathbf{s}_0') \approx \mathbf{v}_0^*(\mathbf{s}_0^*) + \frac{\partial \mathbf{v}_0^*(\mathbf{s}_0^*)}{\partial \mathbf{s}_0^*}\Big(\mathbf{s}_0' - \mathbf{s}_0^*\Big), \text{ où typiquement } \mathbf{s}_0' \approx \mathbf{x}(t+\delta), \text{ et } \mathbf{s}_0^* \approx \mathbf{x}(t) \, . \tag{1.119}$$

En outre, puisque le problème d'optimisation (1.118) est paramétrique en le vecteur $\mathbf{p}$, l'ensemble des variables de décision admissibles l'est également :

$$\mathcal{W}(\mathbf{p}) \coloneqq \{ \, \mathbf{w} \in \mathbb{R}^{n_\mathbf{w}} \mid \mathbf{g}(\mathbf{w}) \leq \mathbf{0} \text{ et } \mathbf{h}(\mathbf{w}; \mathbf{p}) = \mathbf{0} \, \} \, . \tag{1.120}$$

Cela signifie donc que l'ensemble de réalisabilité du problème est donné par :

$$\mathcal{P} \coloneqq \{ \, \mathbf{p} \in \mathbb{R}^{n_\mathbf{w}} \mid \mathcal{W}(\mathbf{p}) \neq \emptyset \, \} \, . \tag{1.121}$$

La valeur optimale du problème d'optimisation (1.118) est l'infimum (le plus grand minorant) de l'image directe $\{J(\mathbf{w}) \mid \mathbf{w} \in \mathcal{W}(\mathbf{p})\}$ de l'ensemble d'admissibilité par la fonction coût :

$$J^*(\mathbf{p}) \coloneqq \inf_{\mathbf{w} \, \in \, \mathcal{W}(\mathbf{p})} J(\mathbf{w}) \, . \tag{1.122}$$

La fonction coût optimale $J^* : \mathbb{R}^{n_\mathbf{p}} \to \mathbb{R}$ est la fonction qui, à chaque vecteur de paramètres $\mathbf{p}$, associe la valeur optimale du problème d'optimisation correspondant. Par convention, la valeur optimale vaut $+\infty$ lorsque le problème est irréalisable (l'infimum de l'ensemble vide vaut $+\infty$), et le problème est dit « non borné » lorsque celle-ci vaut $-\infty$. Un vecteur de variables de décision admissibles $\mathbf{w}^* \in \mathcal{W}(\mathbf{p})$ est optimal si et seulement si la valeur optimale du problème est atteinte

par la fonction coût en ce point :

$$J(\mathbf{w}^*) = \inf_{\mathbf{w} \in \mathcal{W}(\mathbf{p})} J(\mathbf{w}) = J^*(\mathbf{p}) \;\Leftrightarrow\; \forall\, \mathbf{w} \in \mathcal{W}(\mathbf{p}),\; J(\mathbf{w}^*) \leq J(\mathbf{w})\,. \tag{1.123}$$

Comme la valeur optimale du problème peut être atteinte plusieurs fois, l'ensemble des vecteurs optimaux s'écrit :

$$\operatorname*{argmin}_{\mathbf{w} \in \mathcal{W}(\mathbf{p})} J(\mathbf{w}) \coloneqq \{\, \mathbf{w} \in \mathcal{W}(\mathbf{p}) \mid J(\mathbf{w}) = J^*(\mathbf{p}) \,\}\,. \tag{1.124}$$

Pour tout $\mathbf{p} \in \mathcal{P}$, la contrainte d'inégalité d'indice $i \in [\![1, n_{\mathbf{g}}]\!]$ est active (ou saturée) en $\mathbf{w} \in \mathcal{W}(\mathbf{p})$ si et seulement si $g_i(\mathbf{w}) = 0$. Par définition, les contraintes d'égalité sont actives en n'importe quel point admissible. L'ensemble actif d'un vecteur de variables admissibles correspond aux indices des contraintes actives en ce point :

$$\forall\, \mathbf{p} \in \mathcal{P},\; \forall\, \mathbf{w} \in \mathcal{W}(\mathbf{p}),\; \mathcal{A}(\mathbf{w}) \coloneqq [\![1, n_{\mathbf{h}}]\!] \cup \{\, i \in [\![1, n_{\mathbf{g}}]\!] \mid g_i(\mathbf{w}) = 0 \,\}\,. \tag{1.125}$$

Enfin, pour tout $\mathbf{p} \in \mathcal{P}$, la condition de qualification dite « d'indépendance linéaire » est satisfaite en $\mathbf{w} \in \mathcal{W}(\mathbf{p})$ si et seulement si les gradients des contraintes actives en ce point forment une famille de vecteurs linéairement indépendants.

3.3.1 Dualité Lagrangienne et conditions nécessaires d'optimalité

Pour tout problème d'optimisation sous contraintes (1.118), il est possible de définir une fonction de pénalisation particulière, connue sous le nom de « Lagrangien » :

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) \coloneqq J(\mathbf{w}) + \sum_{i=1}^{n_{\mathbf{g}}} \lambda_i g_i(\mathbf{w}) + \sum_{j=1}^{n_{\mathbf{h}}} \mu_j h_j(\mathbf{w}; \mathbf{p}) = J(\mathbf{w}) + \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{w}) + \boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}; \mathbf{p})\,. \tag{1.126}$$

Les composantes des vecteurs $\boldsymbol{\lambda} \in \mathbb{R}^{n_{\mathbf{g}}}$ et $\boldsymbol{\mu} \in \mathbb{R}^{n_{\mathbf{h}}}$ sont les facteurs de pénalisation, ou « multiplicateurs de Lagrange », associés respectivement aux contraintes d'inégalité et d'égalité du problème. Lorsque les multiplicateurs de Lagrange liés aux contraintes d'inégalité sont positifs, le Lagrangien permet d'exprimer le problème d'optimisation sous contraintes (1.118) comme un problème d'optimisation sans contrainte :

$$\inf_{\mathbf{w}} \sup_{\boldsymbol{\lambda} \geq \mathbf{0},\, \boldsymbol{\mu}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) = \inf_{\mathbf{w} \in \mathcal{W}(\mathbf{p})} J(\mathbf{w}) = J^*(\mathbf{p})\,, \tag{1.127}$$

car :

$$\sup_{\boldsymbol{\lambda} \geq \mathbf{0},\, \boldsymbol{\mu}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) = \begin{cases} J(\mathbf{w}), & \text{si } \mathbf{w} \in \mathcal{W}(\mathbf{p}) \\ +\infty, & \text{sinon.} \end{cases} \tag{1.128}$$

D'après l'inégalité max-min :

$$\sup_{\boldsymbol{\lambda}\geq\mathbf{0},\,\boldsymbol{\mu}}\inf_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\leq\inf_{\mathbf{w}}\sup_{\boldsymbol{\lambda}\geq\mathbf{0},\,\boldsymbol{\mu}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\,, \tag{1.129}$$

la valeur optimale $J^*(\mathbf{p})$ du problème d'optimisation (1.118) est systématiquement minorée par :

$$\mathcal{Q}^*(\mathbf{p})\coloneqq\sup_{\boldsymbol{\lambda}\geq\mathbf{0},\,\boldsymbol{\mu}}\inf_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\,. \tag{1.130}$$

Cette grandeur peut être considérée comme la valeur optimale d'un second problème d'optimisation sous contraintes, appelé problème dual :

$$\max_{\boldsymbol{\lambda},\boldsymbol{\mu}}\ \mathcal{Q}(\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\ \text{ sous contraintes : }\boldsymbol{\lambda}\geq\mathbf{0},\quad\text{où }\mathcal{Q}(\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\coloneqq\inf_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\,. \tag{1.131}$$

Le problème d'optimisation d'origine (1.118), quant à lui, est renommé problème primal. Étant donné que le Lagrangien (1.126) dépend linéairement des vecteurs $\boldsymbol{\lambda}$ et $\boldsymbol{\mu}$, la fonction coût du problème dual est concave sur l'ensemble convexe :

$$\mathcal{M}\coloneqq\{\,(\boldsymbol{\lambda},\boldsymbol{\mu})\in\mathbb{R}^{n_\mathbf{g}}\times\mathbb{R}^{n_\mathbf{h}}\mid\boldsymbol{\lambda}\geq\mathbf{0}\,\}\,. \tag{1.132}$$

Or, maximiser une fonction concave revient à minimiser une fonction convexe. Par conséquent, le problème dual (1.131) est toujours convexe quelle soit la nature du problème primal. De ce fait, l'un des avantages du problème dual est qu'il sera généralement plus simple à résoudre que le problème primal. En revanche, lorsque les variables de décision duales $(\boldsymbol{\lambda},\boldsymbol{\mu})$ sont admissibles, la fonction coût du problème dual ne renvoie qu'une borne inférieure de celle du problème primal :

$$\forall\,\mathbf{w}\in\mathbb{R}^{n_\mathbf{w}},\ \forall\,(\boldsymbol{\lambda},\boldsymbol{\mu})\in\mathcal{M},\ \mathcal{Q}(\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\leq\mathcal{L}(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu};\mathbf{p})\leq J(\mathbf{w})\,. \tag{1.133}$$

L'idée du problème dual est alors de maximiser cette borne inférieure pour se rapprocher le plus possible de la valeur optimale du problème primal. Le saut de dualité correspond à la différence $J^*(\mathbf{p})-\mathcal{Q}^*(\mathbf{p})\geq 0$ entre les valeurs optimales des problèmes primal et dual. La relation de dualité est qualifiée de forte si et seulement si le saut de dualité est nul. Dans ce cas, les formulations primale (1.118) et duale (1.131) sont équivalentes, les deux conduisant à la même valeur optimale $J^*(\mathbf{p})=\mathcal{Q}^*(\mathbf{p})$. Lorsque le problème primal est convexe, il suffit que les conditions de qualification de Slater :

$$\exists\,\mathbf{w}\in\mathcal{W}(\mathbf{p}),\ \forall\,i\in[\![1,n_\mathbf{g}]\!],\ g_i(\mathbf{w})<0\,. \tag{1.134}$$

soient respectées pour que le saut de dualité devienne nul. En d'autres termes, la relation de dualité est forte si le problème primal est convexe et que son ensemble d'admissibilité admet un

point intérieur. Plus globalement, la relation de dualité est forte si et seulement si le Lagrangien admet un point-selle dual-admissible, c'est-à-dire un point $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \mathbb{R}^{n_\mathbf{w}} \times \mathcal{M}$ tel que :

$$\forall\, \mathbf{w} \in \mathbb{R}^{n_\mathbf{w}},\ \forall\, (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathcal{M},\ \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) \leq \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) \leq \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p})\,. \tag{1.135}$$

En particulier :

$$\sup_{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathcal{M}} \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) \leq \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) \leq \inf_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p})\,, \tag{1.136}$$

ce qui implique bien que l'inégalité max-min (1.129) se transforme en égalité au point $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. De plus, il est possible de démontrer qu'un point $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \mathbb{R}^{n_\mathbf{w}} \times \mathcal{M}$ est un point-selle du Lagrangien si et seulement $\mathbf{w}^*$ est une solution optimale du problème primal et $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est une solution optimale du problème dual :

$$\mathbf{w}^* \in \underset{\mathbf{w} \in \mathcal{W}(\mathbf{p})}{\operatorname{argmin}}\ J(\mathbf{w}) \quad \text{et} \quad (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in \underset{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathcal{M}}{\operatorname{argmax}}\ \mathcal{Q}(\boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p})\,. \tag{1.137}$$

La relation de dualité forte induit donc l'existence d'un couple de solutions optimales aux problèmes primal et dual. Ce résultat peut être exploité afin de déduire plusieurs conditions nécessaires qui permettront de trouver l'ensemble des solutions potentielles aux deux problèmes. En effet, soit $\mathbf{w}^*$ et $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ un couple de variables de décision primales-duales optimales pour lequel le saut de dualité est nul. La première condition évidente que doit vérifier le point $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est de respecter les contraintes des problèmes primal et dual :

$$\begin{cases} \forall\, i \in [\![1, n_\mathbf{g}]\!],\ g_i(\mathbf{w}^*) \leq 0 \\ \forall\, j \in [\![1, n_\mathbf{h}]\!],\ h_j(\mathbf{w}^*; \mathbf{p}) = 0 \\ \forall\, i \in [\![1, n_\mathbf{g}]\!],\ \lambda_i^* \geq 0\,. \end{cases} \tag{1.138}$$

Puis, étant donné que le saut de dualité est nul et que les variables primales $\mathbf{w}^*$ et duales $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ sont optimales :

$$J(\mathbf{w}^*) = \mathcal{Q}(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) = \inf_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) \leq \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) \leq J(\mathbf{w}^*)\,, \tag{1.139}$$

ce qui signifie que les deux inégalités ci-dessus sont en fait des égalités :

$$J(\mathbf{w}^*) = \inf_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) = \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p})\,. \tag{1.140}$$

Le vecteur $\mathbf{w}^*$ minimise donc la fonction $\mathcal{L}(\cdot, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p})$, ce qui conduit à la condition, dite de « stationnarité », suivante :

$$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \nabla_{\mathbf{w}} J(\mathbf{w}^*) + \sum_{i=1}^{n_{\mathbf{g}}} \lambda_i^* \nabla_{\mathbf{w}} g_i(\mathbf{w}^*) + \sum_{j=1}^{n_{\mathbf{h}}} \mu_j^* \nabla_{\mathbf{w}} h_j(\mathbf{w}^*) = \mathbf{0} \tag{1.141}$$

À noter que le gradient (1.141) du Lagrangien ne dépend pas du vecteur de paramètres $\mathbf{p}$, car ce dernier rentre linéairement dans les contraintes d'égalité. Enfin, puisque les variables primales $\mathbf{w}^*$ et duales $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ sont admissibles :

$$J(\mathbf{w}^*) = \mathcal{L}(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*; \mathbf{p}) = J(\mathbf{w}^*) + \sum_{i=1}^{n_{\mathbf{g}}} \underbrace{\lambda_i^* g_i(\mathbf{w}^*)}_{\leq 0} + \sum_{j=1}^{n_{\mathbf{h}}} \mu_j^* \underbrace{h_j(\mathbf{w}^*; \mathbf{p})}_{=0} \;\Rightarrow\; {\boldsymbol{\lambda}^*}^\top \mathbf{g}(\mathbf{w}^*) = \mathbf{0}\,. \tag{1.142}$$

Cette dernière condition, dite de « complémentarité », implique que les multiplicateurs de Lagrange optimaux liés aux contraintes d'inégalités inactives en $\mathbf{w}^* \in \mathcal{W}(\mathbf{p})$ sont forcément nuls :

$$\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\; \lambda_i^* g_i(\mathbf{w}^*) = 0 \quad \text{et} \quad g_i(\mathbf{w}^*) < 0 \;\Rightarrow\; \lambda_i^* = 0\,. \tag{1.143}$$

En revanche, les multiplicateurs de Lagrange optimaux liés aux contraintes d'inégalités actives en $\mathbf{w}^* \in \mathcal{W}(\mathbf{p})$ peuvent être soit nuls, soit strictement positifs selon le cas :

$$\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\; \lambda_i^* g_i(\mathbf{w}^*) = 0 \quad \text{et} \quad g_i(\mathbf{w}^*) = 0 \;\Rightarrow\; \lambda_i^* \geq 0\,. \tag{1.144}$$

Une contrainte d'inégalité d'indice $i \in \mathcal{A}(\mathbf{w}^*)$ est alors faiblement active quand $\lambda_i^* = 0$ et strictement active quand $\lambda_i^* > 0$. Par ailleurs, la condition de complémentarité (1.142) devient « stricte » au point $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ si et seulement si :

$$\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\; \lambda_i^* g_i(\mathbf{w}^*) = 0 \quad \text{et} \quad g_i(\mathbf{w}^*) = 0 \;\Rightarrow\; \lambda_i^* > 0\,, \tag{1.145}$$

c'est-à-dire si et seulement si les contraintes d'inégalités actives en $\mathbf{w}^* \in \mathcal{W}(\mathbf{p})$ le sont de façon stricte. En résumé, tout couple $\mathbf{w}^*$ et $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ de variables de décision primales-duales optimales pour lequel le saut de dualité est nul doit satisfaire les conditions nécessaires d'optimalité, dites de Karush-Kuhn-Tucker (KKT), suivantes :

$$\begin{aligned}
&\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\; g_i(\mathbf{w}^*) \leq 0 \\
&\forall\, j \in [\![1, n_{\mathbf{h}}]\!],\; h_j(\mathbf{w}^*; \mathbf{p}) = 0 \\
&\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\; \lambda_i^* \geq 0 \\
&\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\; \lambda_i^* g_i(\mathbf{w}^*) = 0
\end{aligned} \tag{1.146}$$

$$\nabla_{\mathbf{w}} J(\mathbf{w}^*) + \sum_{i=1}^{n_{\mathbf{g}}} \lambda_i^* \nabla_{\mathbf{w}} g_i(\mathbf{w}^*) + \sum_{j=1}^{n_{\mathbf{h}}} \mu_j^* \nabla_{\mathbf{w}} h_j(\mathbf{w}^*) = \mathbf{0}\,.$$

Un point $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ qui vérifie les conditions de KKT (1.146) est appelé un point stationnaire. Dans les faits, l'hypothèse de dualité forte est souvent remplacée par une condition de qualification des contraintes. Plus précisément, si $\mathbf{w}^*$ est un vecteur de variables de décision primales optimales pour lequel la condition de qualification d'indépendance linéaire est satisfaite, alors il existe une unique paire $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ de multiplicateurs de Lagrange telle que $(\mathbf{w}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ est un point stationnaire.

3.3.2 Principe des méthodes de points intérieurs

Les deux types de méthodes les plus employées en commande prédictive pour résoudre le problème d'optimisation paramétrique (1.117) résultant de la transcription du problème de commande optimale (1.27) sont les méthodes de programmation quadratique successive et les méthodes de points intérieurs. Ces méthodes cherchent à trouver une solution potentielle aux problèmes primal (1.118) et dual (1.131) en résolvant les conditions nécessaires d'optimalité (1.146). Seules les méthodes de points intérieurs seront présentées ici.

Les conditions de KKT ne peuvent être résolues directement par une méthode de recherche zéro conventionnelle à cause des contraintes d'inégalité des problèmes primal et dual. La première étape d'une méthode de points intérieurs consiste donc à éliminer les contraintes d'inégalité de ces deux problèmes. Tout d'abord, les contraintes d'inégalité du problème primal sont éliminées en introduisant des variables d'écart :

$$\min_{\mathbf{w}} \; J(\mathbf{w}) \;\; \text{s.c.} : \begin{cases} \mathbf{g}(\mathbf{w}) \leq \mathbf{0} \\ \mathbf{h}(\mathbf{w};\mathbf{p}) = \mathbf{0} \end{cases} \Leftrightarrow \min_{\mathbf{w},\mathbf{s}} \; J(\mathbf{w}) \;\; \text{s.c.} : \begin{cases} \mathbf{g}(\mathbf{w}) + \mathbf{s} = \mathbf{0} \\ \mathbf{h}(\mathbf{w};\mathbf{p}) = \mathbf{0} \\ \mathbf{s} \geq 0\,. \end{cases} \tag{1.147}$$

Puis, les contraintes d'inégalité liées aux variables d'écart sont éliminées en ajoutant des fonctions barrières logarithmiques à la fonction coût du nouveau problème primal :

$$\min_{\mathbf{w},\mathbf{s}} \; J(\mathbf{w}) - \nu \sum_{i=1}^{n_{\mathbf{g}}} \ln(s_i) \;\; \text{sous contraintes} : \begin{cases} \mathbf{g}(\mathbf{w}) + \mathbf{s} = \mathbf{0} \\ \mathbf{h}(\mathbf{w};\mathbf{p}) = \mathbf{0}\,. \end{cases} \tag{1.148}$$

où $\nu \in \mathbb{R}_{>0}$ est un petit paramètre scalaire, souvent appelé « paramètre barrière ». L'avantage de ces fonctions barrières est qu'elles permettent de prendre implicitement en compte les contraintes d'inégalité associées aux variables d'écart, puisque le problème d'optimisation (1.148) devient

irréalisable dès que l'une d'entre elles se rapproche de zéro :

$$\forall\, i \in [\![1, n_{\mathbf{g}}]\!],\ \lim_{s_i \to 0}\ -\nu \sum_{i=1}^{n_{\mathbf{g}}} \ln(s_i) = +\infty\,. \tag{1.149}$$

Cependant, la fonction coût du problème d'optimisation (1.148) n'est plus égale à celle du problème primal (1.118). Par conséquent, la seconde étape d'une méthode de points intérieurs est de résoudre à plusieurs reprise le problème d'optimisation (1.148) en diminuant à chaque fois la valeur du paramètre barrière $\nu \to 0$, jusqu'à ce que la fonction coût modifiée :

$$J_\nu(\mathbf{w}, \mathbf{s}) \coloneqq J(\mathbf{w}) - \nu \sum_{i=1}^{n_{\mathbf{g}}} \ln(s_i)\,, \tag{1.150}$$

soit égale à celle du problème d'origine. Comme le gradient du Lagrangien :

$$\mathcal{L}_\nu(\mathbf{w}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) \coloneqq J_\nu(\mathbf{w}, \mathbf{s}) + \sum_{i=1}^{n_{\mathbf{g}}} \lambda_i \Big(g_i(\mathbf{w}) + s_i \Big) + \sum_{j=1}^{n_{\mathbf{h}}} \mu_j h_j(\mathbf{w}; \mathbf{p})\,, \tag{1.151}$$

est donné par :

$$\begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}_\nu(\mathbf{w}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) \\ \nabla_{\mathbf{s}} \mathcal{L}_\nu(\mathbf{w}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \mathbf{p}) \end{bmatrix} = \begin{bmatrix} \nabla_{\mathbf{w}} J(\mathbf{w}) + \sum_{i=1}^{n_{\mathbf{g}}} \lambda_i \nabla_{\mathbf{w}} g_i(\mathbf{w}) + \sum_{j=1}^{n_{\mathbf{h}}} \mu_j \nabla_{\mathbf{w}} h_j(\mathbf{w}) \\ \sum_{i=1}^{n_{\mathbf{g}}} (-\nu \nabla_{\mathbf{s}} \ln(s_i) + \lambda_i \nabla_{\mathbf{s}} s_i) \end{bmatrix}. \tag{1.152}$$

les conditions de KKT du problème d'optimisation (1.148) s'écrivent sous forme vectorielle :

$$\mathbf{V}_{\mathrm{KKT}}(\mathbf{w}, \mathbf{s}, \boldsymbol{\lambda}, \boldsymbol{\mu}; \nu) \coloneqq \begin{bmatrix} \mathbf{g}(\mathbf{w}) + \mathbf{s} \\ \mathbf{h}(\mathbf{w}; \mathbf{p}) \\ \nabla_{\mathbf{w}} J(\mathbf{w}) + \frac{\partial \mathbf{g}(\mathbf{w})}{\partial \mathbf{w}}^{\top} \boldsymbol{\lambda} + \frac{\partial \mathbf{h}(\mathbf{w})}{\partial \mathbf{w}}^{\top} \boldsymbol{\mu} \\ \mathrm{diag}(\mathbf{s}) \boldsymbol{\lambda} - \nu \mathbf{1}_{n_{\mathbf{g}}} \end{bmatrix} = \mathbf{0}\,, \tag{1.153}$$

où $\partial \mathbf{g}(\mathbf{w}) / \partial \mathbf{w} \coloneqq [\nabla_{\mathbf{w}} g_1(\mathbf{w}) \cdots \nabla_{\mathbf{w}} g_{n_{\mathbf{g}}}(\mathbf{w})]^\top$ et $\partial \mathbf{h}(\mathbf{w}) / \partial \mathbf{w} \coloneqq [\nabla_{\mathbf{w}} h_1(\mathbf{w}) \cdots \nabla_{\mathbf{w}} h_{n_{\mathbf{h}}}(\mathbf{w})]^\top$ sont les matrices jacobiennes des fonctions $\mathbf{g}(\mathbf{w})$ et $\mathbf{h}(\mathbf{w};\mathbf{p})$, $\mathrm{diag}(\mathbf{s}) \in \mathbb{R}^{n_{\mathbf{g}} \times n_{\mathbf{g}}}$ est la matrice diagonale construite à partir des éléments du vecteur $\mathbf{s} \in \mathbb{R}^{n_{\mathbf{g}}}$ et $\mathbf{1}_{n_{\mathbf{g}}} \coloneqq [1 \cdots 1]^\top \in \mathbb{R}^{n_{\mathbf{g}}}$. Contrairement aux conditions de KKT (1.146) du problème primal (1.118), les conditions de KKT (1.152) du problème barrière (1.148) peuvent être facilement résolues en faisant appel à une méthode de recherche de zéro de type Newton ou quasi-Newton.

4 Résumé du chapitre

L'objectif de ce chapitre était de mettre en évidence les différentes étapes nécessaires à la conception d'un algorithme de commande prédictive :

Étape 1 – modéliser le système à contrôler.

choix : phénomènes à représenter, dynamiques à négliger, paramètres à identifier, actionneurs à manipuler, etc.

Étape 2 – formuler le problème de commande optimale.

choix : traduction du cahier des charges, horizon de temps, coût de fonctionnement, coût terminal, contraintes, etc.

Étape 3 – définir l'algorithme de commande prédictive.

choix : stabilité en boucle fermée, réalisabilité récursive du problème, fréquence de calcul de la loi de commande, compensation du délai de transmission, etc.

Étape 4 – simuler le modèle du système à contrôler.

choix : précision et stabilité du schéma d'intégration numérique, taille du pas de temps, contrôle de l'erreur de troncature, etc.

Étape 5 – transcrire le problème de commande optimale en problème d'optimisation numérique

choix : subdivision de l'horizon de prédiction, paramétrisation du signal de commande, méthode de transcription, calcul numérique de l'intégrale.

Étape 6 – résoudre le problème d'optimisation numérique

choix : calcul numérique des dérivées, méthode d'optimisation et de recherche de zéro associée, normalisation du problème d'optimisation, etc.

Chapitre II

Modélisation pour la commande des réacteurs à eau sous pression

1 Physique des réacteurs pour l'automaticien

L'objectif de cette section est d'introduire l'ensemble des concepts physiques nécessaires à la compréhension du modèle de réacteur nucléaire à eau sous-pression élaboré dans cette thèse. Les notions présentées ici sont déjà bien établies dans la littérature scientifique et sont tirées des ouvrages [67]-[69], des thèses [6], [70], des articles [5], [71]-[75], du document technique [76], ainsi que du site internet [77].

1.1 Fonctionnement des réacteurs à eau sous pression

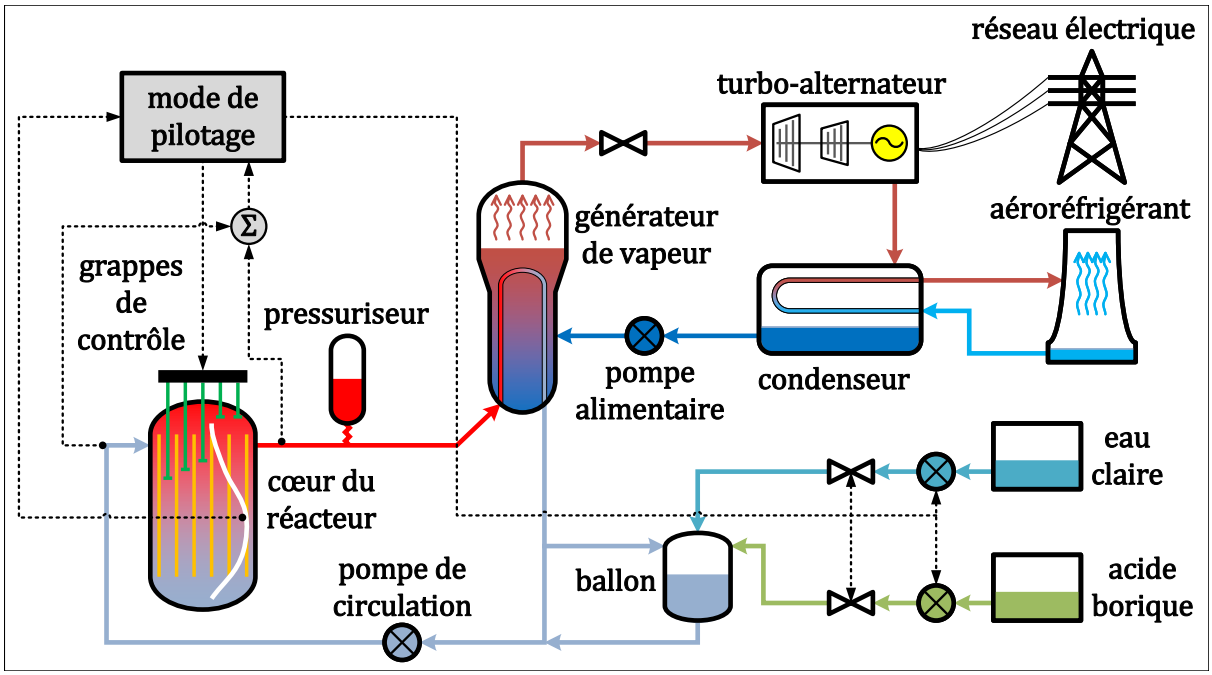

Figure 2.1 – schéma simplifié d'une des boucles d'un réacteur à eau sous pression.

Dans les grandes lignes, le principe de fonctionnement d'une centrale nucléaire est très similaire à celui des autres centrales thermiques, l'objectif étant toujours de convertir l'énergie mécanique de rotation d'une turbine en énergie électrique. Pour y parvenir, une source de chaleur fournit de l'énergie thermique à un fluide caloporteur afin de générer la vapeur entraînant la turbine. Cette turbine est couplée à un alternateur qui, en tournant, génère de l'électricité. La vapeur résiduelle est ensuite repassée à l'état liquide avant d'être de nouveau vaporisée par le fluide caloporteur. La réaction suit ainsi un cycle thermodynamique, dont le rendement maximal est donné par :

$$\eta_{\max} = 1 - \frac{T_f}{T_c}\,, \tag{2.1}$$

où $T_c \simeq 325\,^\circ\mathrm{C}$ et $T_f \simeq 215\,^\circ\mathrm{C}$ sont les températures de la source chaude et de la source froide lors de l'échange de chaleur. La particularité des centrales nucléaires est que l'énergie produite par la source de chaleur provient non pas de la combustion chimique d'un combustible fossile (charbon, gaz, fuel, etc.) avec de l'oxygène, mais de la fission de noyaux atomiques lourds (généralement des isotopes de l'uranium ou du plutonium) par des neutrons.

Le cœur du réacteur est l'endroit de la centrale où ont lieu ces réactions de fissions. Très schématiquement, il s'agit d'une cuve cylindrique composée, selon la puissance délivrée, de 157 (réacteurs 900 MWe) à 241 (réacteurs EPR 1650 MWe) assemblages, contenant chacun 264 ou 265 crayons d'environ 4 mètres de haut. Ces crayons sont constitués de tubes, ou gaines, en alliage de zirconium dans lesquels sont empilés des pastilles de matériaux fissiles, communément appelées combustible nucléaire par abus de langage. Les réacteurs à eau sous pression utilisent de l'uranium faiblement enrichi comme combustible nucléaire, et comportent deux circuits d'eau fermés : le circuit primaire, qui extrait et transmet l'énergie du cœur par l'intermédiaire des générateurs de vapeur, et le circuit secondaire, qui la récupère et la transforme en électricité grâce au groupe turbo-alternateur. Les deux circuits d'eau sont totalement séparés pour empêcher que les éléments radioactifs présents dans le circuit primaire ne se retrouvent dans le circuit secondaire.

En traversant le cœur du réacteur, l'eau du circuit primaire rentre en contact avec la gaine des crayons de combustible, où sa température passe d'environ 290 °C à 325 °C. Pour que l'eau puisse rester liquide à ces températures, la pression du circuit primaire est maintenue autour de 155 bar par un pressuriseur, de sorte que sa température d'ébullition atteigne 345 °C. L'eau chauffée par le cœur est ensuite répartie dans 3 ou 4 boucles, comprenant chacune un générateur de vapeur et une pompe de circulation.

Les générateurs de vapeur sont à l'interface entre le circuit primaire et le circuit secondaire. L'eau du circuit primaire qui les alimente circule à l'intérieur d'un échangeur de chaleur à faisceau tubulaire, le long duquel est vaporisée l'eau, plus froide d'une centaine de degrés, du circuit secondaire. La vapeur ainsi générée permet, en se détendant, de faire tourner les roues des différents étages (haute puis basse pression) de la turbine, reliés à l'arbre de transmission de

l'alternateur. Un troisième circuit d'eau ouvert sur l'environnement (fleuve ou mer) est utilisé pour refroidir la vapeur restante lors de son passage dans le condenseur. L'eau issue de la liquéfaction de la vapeur est alors renvoyée au générateur de vapeur par les pompes alimentaires du circuit secondaire.

Le rendement d'une centrale nucléaire, d'environ 33 %, est inférieur à celui d'une centrale thermique classique, d'environ 50 %, car la vapeur y est produite à plus basse température. En revanche, l'énergie libérée par une réaction de fission est considérablement supérieure à celle d'une combustion : autour de 200 MeV par atome pour une fission contre seulement quelques électronvolts par atome pour une combustion. En outre, la réaction de fission peut être pilotée pour générer de l'énergie sur demande et ne produit aucun gaz à effet de serre, contrairement à la combustion qui rejette, entres autres, du dioxyde de carbone.

1.2 Prérequis en neutronique

1.2.1 Notions de base et réaction en chaîne

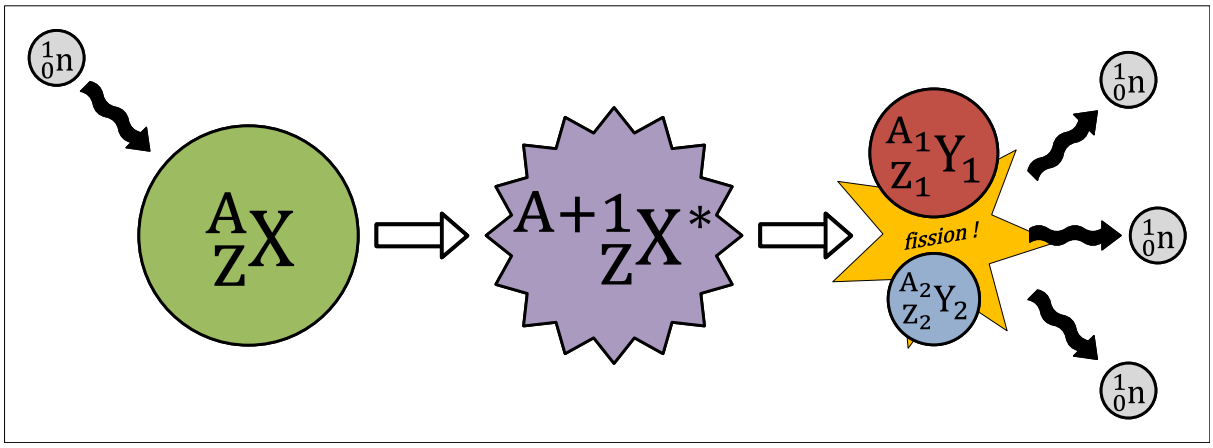

FIGURE 2.2 – Schéma simplifié d'une réaction de fission émettant $\kappa = 3$ neutrons et deux produits de fission.

Lorsqu'un neutron $^1_0$n est absorbé par le noyau atomique d'un élément chimique ^A_ZX, son nombre de masse A (nombre de protons et de neutrons) augmente d'un cran, mais son numéro atomique Z (nombre de protons) reste identique. Le noyau atomique de l'isotope $^{A+1}_{Z}$X ainsi constitué est ensuite susceptible, ou non, de se scinder en deux autres noyaux atomiques $^{A_1}_{Z_1}$Y$_1$ et $^{A_2}_{Z_2}$Y$_2$ plus légers ($A_1 + A_2 < A$ et $Z_1 + Z_2 = Z$), appelés produits de fissions. Cette réaction nucléaire $^1_0\text{n} + {}^A_Z\text{X} \Longrightarrow {}^{A+1}_{Z}\text{X}^* \Longrightarrow {}^{A_1}_{Z_1}\text{Y}_1 + {}^{A_2}_{Z_2}\text{Y}_2 + \kappa\, {}^1_0\text{n}$, dite de fission, libère une quantité d'énergie phénoménale et peut s'accompagner d'une émission de quelques neutrons ($0 \leq \kappa \leq 7$). Ces neutrons, émis par fission, pourront alors :

- soit s'échapper du cœur du réacteur (**fuite neutronique**),

- soit entrer en collision avec un noyau atomique, sans être absorbés, puis rebondir en changeant de direction (**diffusion potentielle élastique**),
- soit être absorbés par un noyau atomique. Le noyau composé, isotope du noyau cible, gagne de l'énergie d'excitation, qu'il va restituer par l'une des réactions nucléaires suivantes :

  — **diffusion résonnante élastique** : l'énergie d'excitation est dissipée par l'émission d'un nouveau neutron.

  — **diffusion résonnante inélastique** : une partie de l'énergie d'excitation est dissipée par l'émission d'un nouveau neutron, et le reste par l'émission d'un rayonnement gamma (transition isomérique).

  — **capture radiative** : l'énergie d'excitation est dissipée sous forme de photons, par l'émission d'un rayonnement gamma (transition isomérique).

  — **fission** : l'énergie d'excitation est dissipée en fragmentant le noyau atomique cible en deux (très rarement trois) produits de fission, avec émission de nouveaux neutrons.

  — **émission de particules chargées** : l'énergie d'excitation est dissipée par l'émission d'une particule chargée de masse légère ou nulle (désintégration radioactive de type alpha, bêta plus ou bêta moins, voire émission de proton).

  — **émission de neutrons** : l'énergie d'excitation est dissipée par l'émission d'un ou de plusieurs neutrons (extrêmement rare).

Par la suite, les termes « diffusion » et « capture » seront respectivement utilisés pour désigner toute réaction réémettant au moins un neutron, fission exclue, et toute réaction ne réémettant pas de neutron. Étant donné que les neutrons émis par fission ont une chance, plus ou moins grande, de provoquer de nouvelles fission, il est possible de déclencher une **réaction en chaîne**. En effet, certains noyaux atomiques lourds, comme celui de l'uranium 235 ou du plutonium 239, produisent en moyenne plus de neutrons par fission qu'ils n'en capturent. À l'inverse, d'autres noyaux atomiques, comme celui du xénon 135 ou du bore 10, capturent en moyenne plus de neutrons qu'il n'en produisent par fission. Le **facteur de multiplication effectif**, noté k_{eff}, représente le nombre moyen de fissions induites par les neutrons issus d'une même réaction de fission ayant lieu dans un milieu combustible fini :

- si $k_{\text{eff}} = 1$, la réaction en chaîne est stable et s'auto-entretient : le cœur est critique.
- si $k_{\text{eff}} < 1$, la réaction en chaîne s'étouffe : le cœur est sous-critique.
- si $k_{\text{eff}} > 1$, la réaction en chaîne s'emballe à un rythme exponentiel : le cœur est sur-critique.

En d'autre termes, le facteur de multiplication effectif donne l'évolution moyenne, d'une géné-

ration de fissions à l'autre, de la population de neutrons dans le cœur :

$$k_{\text{eff}} = \frac{\text{nombre de neutrons de la génération courante}}{\text{nombre de neutrons de la génération précédente}} \\ \Leftrightarrow \frac{\text{nombre de neutrons émis par fission à la génération courante}}{\text{nombre de neutrons perdus par fuite et capture à la génération précédente}} . \tag{2.2}$$

Comme mentionné précédemment, l'énergie thermique générée par le cœur du réacteur augmente directement avec le nombre de fissions se produisant en son sein. Or, puisque la réaction en chaîne suit une loi exponentielle, une variation, même infime, du facteur de multiplication effectif aura une influence non négligeable sur la température moyenne du cœur. Par conséquent, il est essentiel de contrôler précisément l'évolution du facteur de multiplication effectif autour de sa valeur critique $k_{\text{eff}} = 1$ afin d'éviter que le cœur du réacteur ne surchauffe, le risque étant que l'eau du circuit primaire se mette à bouillir puis se vaporise (crise d'ébullition), voire, pire encore, que les crayons de combustible commencent à fondre (fusion partielle du cœur du réacteur). La **réactivité** :

$$\rho := \frac{k_{\text{eff}} - 1}{k_{\text{eff}}} , \tag{2.3}$$

permet de quantifier l'écart relatif du cœur par rapport à la configuration critique. Cette grandeur, sans dimension, s'exprime en pour cent mille ($1\,\text{pcm} = 10^{-3}\ \%$), car sa valeur doit toujours être maintenue très proche de zéro afin garder le contrôle de la réaction en chaîne.

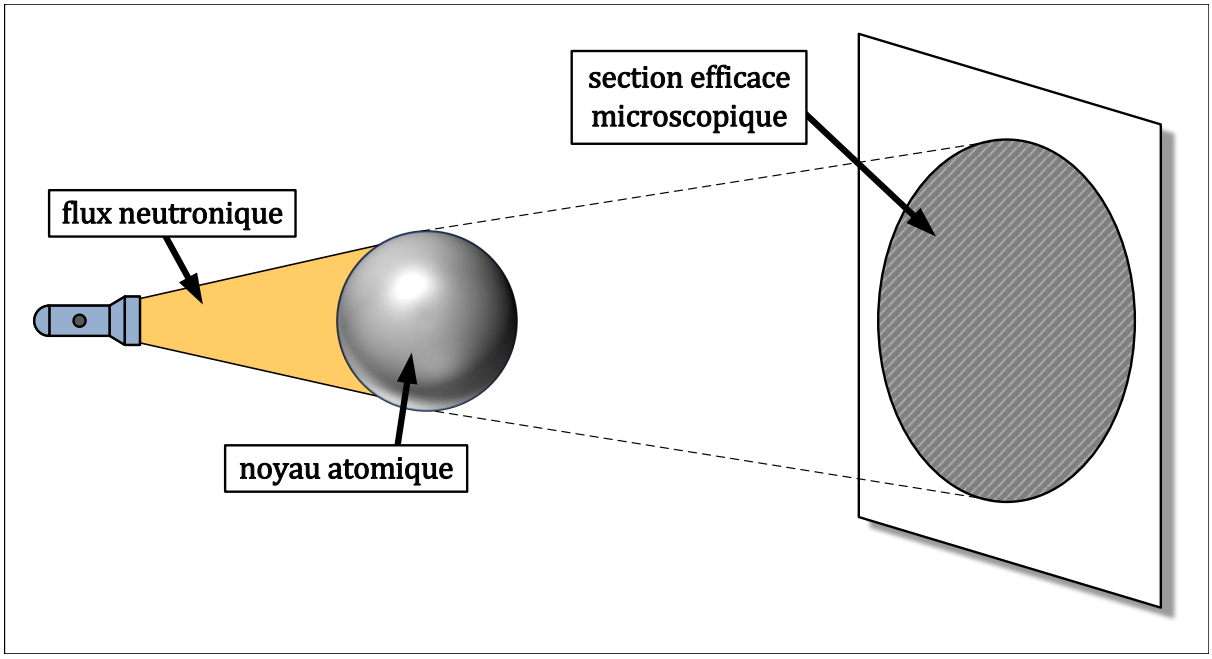

FIGURE 2.3 – Illustration du concept de section efficace microscopique.

D'autres grandeurs physiques doivent être définies pour étudier comment les neutrons in-

teragissent avec les noyaux atomiques du milieu dans lequel ils cheminent. Tout d'abord, la **densité neutronique** $n(t)$ représente le nombre de neutrons observés par unité de volume à un instant donné. De même, la **densité particulaire** $\mathrm{X}^A(t)$ représente le nombre de noyaux atomiques de l'élément chimique ${}^A_Z\mathrm{X}$ observés par unité de volume à un instant donné. Ces deux grandeurs s'expriment respectivement en neutrons par centimètre cube ($\mathrm{n\,cm^{-3}}$) et en atomes par centimètre cube ($\mathrm{at\,cm^{-3}}$). Le **taux de réaction** $R_{\mathrm{tot}}(t, E)$ correspond ensuite au nombre de réactions nucléaires (diffusion, fission, et capture) ayant lieu par unité de temps et de volume dans le milieu considéré. Il s'agit du produit de deux autres grandeurs :

$$R_{\mathrm{tot}}(t, E) \coloneqq \Sigma_{\mathrm{tot}}(t, E)\phi(t)\,, \tag{2.4}$$

à savoir la **section efficace macroscopique** $\Sigma_{\mathrm{tot}}(t, E)$, exprimée en $\mathrm{cm^{-1}}$, et le **flux neutronique** $\phi(t)$, exprimé en $\mathrm{n\,cm^{-2}\,s^{-1}}$. Le flux neutronique est défini par la relation :

$$\phi(t) \coloneqq V_n n(t)\,, \tag{2.5}$$

où V_n est le module de la vitesse des neutrons décrits par la densité neutronique $n(t)$. Contrairement à ce que son nom indique, le flux neutronique n'est pas véritablement un flux au sens physique du terme, mais s'apparente plutôt à un débit de neutrons multidirectionnel. Le flux neutronique permet de caractériser le niveau d'interaction entre une population de neutrons se déplaçant à la vitesse V_n et les noyaux atomiques qui les entourent. Une population contenant $n(t)$ neutrons par centimètre cube interagira donc autant avec les noyaux atomiques du milieu qu'une autre population deux fois moins dense mais se déplaçant deux fois plus vite. La section efficace macroscopique représente la probabilité par unité de longueur qu'une réaction nucléaire se produise lorsqu'un certain matériau est soumis à un flux neutronique. Cette probabilité :

$$\Sigma_{\mathrm{tot}}(t, E) \coloneqq \sigma_{\mathrm{X}}(E)\mathrm{X}^A(t)\,, \tag{2.6}$$

dépend évidemment de la densité particulaire $\mathrm{X}^A(t)$ de l'élément chimique ${}^A_Z\mathrm{X}$ dont est constitué le matériau. Si le matériau est composé de plusieurs éléments chimiques, par exemple ${}^{A_1}_{Z_1}\mathrm{X}_1$ et ${}^{A_2}_{Z_2}\mathrm{X}_2$, sa section efficace macroscopique sera la somme pondérée des densités particulaires de chaque élément :

$$\Sigma_{\mathrm{tot}}(t, E) \coloneqq \sigma_{\mathrm{X}_1}(E)\mathrm{X}_1^{A_1}(t) + \sigma_{\mathrm{X}_2}(E)\mathrm{X}_2^{A_2}(t)\,. \tag{2.7}$$

Le coefficient de proportionnalité $\sigma_{\mathrm{X}}(E)$, exprimé en barns ($1\,\mathrm{b} = 10^{-24}\ \mathrm{cm^2}$), est la **section efficace microscopique** de l'élément chimique ${}^A_Z\mathrm{X}$. Pour faire simple, il s'agit de l'ombre fictive que projetterait le noyau atomique de l'élément chimique ${}^A_Z\mathrm{X}$ si sa taille était équivalente à sa probabilité d'interaction avec un neutron incident, et que le flux neutronique auquel il est exposé s'apparentait à un faisceau lumineux. La section efficace microscopique dépend à la fois de

l'énergie cinétique du neutron incident et, dans une moindre mesure, de celle du noyau atomique cible, la somme des deux énergies étant simplement notée E. En outre, il est possible de définir différentes sections efficaces microscopiques partielles selon le type de réaction nucléaire (fission, diffusion ou capture) étudiée :

$$\sigma_{\mathrm{X}}(E) \coloneqq \sigma_{\mathrm{X}}^{\mathrm{fis}}(E) + \sigma_{\mathrm{X}}^{\mathrm{dif}}(E) + \sigma_{\mathrm{X}}^{\mathrm{cap}}(E)\,. \tag{2.8}$$

La même règle s'applique aux sections efficaces macroscopiques :

$$\Sigma_{\mathrm{tot}}(t) = \underbrace{\sigma_{\mathrm{X}}^{\mathrm{fis}}(E)\mathrm{X}^{A}(t)}_{\Sigma_{\mathrm{fis}}(t,E)} + \underbrace{\sigma_{\mathrm{X}}^{\mathrm{dif}}(E)\mathrm{X}^{A}(t)}_{\Sigma_{\mathrm{dif}}(t,E)} + \underbrace{\sigma_{\mathrm{X}}^{\mathrm{cap}}(E)\mathrm{X}^{A}(t)}_{\Sigma_{\mathrm{cap}}(t,E)}\,, \tag{2.9}$$

et, par extension, aux taux de réaction :

$$R_{\mathrm{tot}}(t,E) = \underbrace{\Sigma_{\mathrm{fis}}(t,E)\phi(t)}_{R_{\mathrm{fis}}(t,E)} + \underbrace{\Sigma_{\mathrm{dif}}(t,E)\phi(t)}_{R_{\mathrm{dif}}(t,E)} + \underbrace{\Sigma_{\mathrm{cap}}(t,E)\phi(t)}_{R_{\mathrm{cap}}(t,E)}\,. \tag{2.10}$$

Il est admis, dans ce qui suit, que la section efficace macroscopique de fission Σ_{fis} ne dépend pas explicitement du temps t ni de l'énergie E, car l'évolution des noyaux atomiques fissiles est lente par rapport à celle des neutrons et des produits de fission d'intérêt, et les sections efficaces microscopiques $\sigma_{\mathrm{X}}^{\mathrm{fis}}$, $\sigma_{\mathrm{X}}^{\mathrm{dif}}$, et $\sigma_{\mathrm{X}}^{\mathrm{cap}}$ sont évaluées en certaines valeurs d'énergies bien précises.

1.2.2 Évolution isotopique des produits de fission

Au cours du **cycle d'irradiation**, les réactions de fission et de capture qui se produisent dans le cœur du réacteur modifient la composition chimique du combustible nucléaire. Les noyaux atomiques fissiles, en particulier, disparaissent progressivement en donnant naissance à une multitude de produits de fission. Parmi tous les produits de fission générés, seule une vingtaine auront une influence significative sur la section efficace macroscopique, et donc le taux de réaction, du combustible nucléaire. Les équations de Bateman, de la forme :

$$\frac{\mathrm{d}\mathrm{X}_{\mathrm{i}}^{A_i}(t)}{\mathrm{d}t} = \gamma_{\mathrm{X_i}}\Sigma_{\mathrm{fis}}\phi(t) - \Big(\lambda_{\mathrm{X_i}} + \sigma_{\mathrm{X_i}}^{\mathrm{cap}}\phi(t)\Big)\mathrm{X}_{\mathrm{i}}^{A_i}(t) + \sum_{\text{isotope } \mathrm{X_j}} \Big(\lambda_{\mathrm{X_j}} + \sigma_{\mathrm{X_j}}^{\mathrm{cap}}\phi(t)\Big)\mathrm{X}_{\mathrm{j}}^{A_j}(t)\,, \tag{2.11}$$

permettent de décrire l'évolution de la densité particulaire de ces différents produits de fission. Le premier terme de l'équation (2.11) correspond au taux de production de noyaux atomiques par fission du combustible nucléaire. Cette production sera plus ou moins importante en fonction de la valeur, exprimée en %, du **rendement de fission** $\gamma_{\mathrm{X_i}}$ de l'élément chimique ${}^{A_i}_{Z_i}\mathrm{X_i}$. Tout comme les sections efficaces microscopiques, le rendement de fission ne dépend pas explicitement de la variable E, car celui-ci est évalué en certaines valeurs d'énergies bien précises. À noter que

la somme des rendements de fission ainsi évalués vaut 200 %, puisque chaque réaction fission génère deux produits de fission. Le deuxième terme correspond au taux de disparition de noyaux atomiques par désintégration radioactive et capture neutronique. À la différence des réactions de capture, la disparition par désintégration radioactive peut survenir même en l'absence de flux neutronique, mais ne concerne que certains noyaux atomiques instables, dit radioactifs. Cette désintégration naturelle suit une loi exponentielle qui dépend uniquement de la **constante de décroissance radioactive** $\lambda_{\mathrm{X_i}}$, exprimée en s^{-1}, de l'élément chimique ${}^{A_i}_{Z_i}\mathrm{X_i}$. La **période radioactive** $T^{1/2}_{\mathrm{X_i}}$ est la durée au bout de laquelle la moitié de la quantité initiale $\mathrm{X_i}^{A_i}(t_0)$ de noyaux atomiques de l'élément chimique ${}^{A_i}_{Z_i}\mathrm{X_i}$ aura disparu par désintégration radioactive :

$$\mathrm{X_i}^{A_i}(t) = \mathrm{X_i}^{A_i}(t_0)\exp\left(-\lambda_{\mathrm{X_i}}(t-t_0)\right) \quad \text{et} \quad \mathrm{X_i}^{A_i}(t_0+T^{1/2}_{\mathrm{X_i}}) = \frac{\mathrm{X_i}^{A_i}(t_0)}{2} \;\Rightarrow\; T^{1/2}_{\mathrm{X_i}} := \frac{\ln(2)}{\lambda_{\mathrm{X_i}}}. \tag{2.12}$$

Les périodes radioactives diffèrent drastiquement d'un élément chimique à l'autre, celles-ci pouvant aller de quelques milliardièmes de seconde à plusieurs milliards d'années. Le troisième et dernier terme correspond au taux de production de noyaux atomiques par désintégration radioactive et capture neutronique d'autres noyaux atomiques présents dans le combustible nucléaire. La chaîne d'évolution du produit de fission ${}^{A_i}_{Z_i}\mathrm{X_i}$ est habituellement tracée afin de retrouver quels éléments chimiques ${}^{A_j}_{Z_j}\mathrm{X_j}$ contribuent, en se transmutant, à la production de noyaux atomiques.

1.2.3 Évolution de la population de neutrons

Les réactions de fission libèrent deux types de neutrons : les neutrons prompts, qui sont émis dès que le noyau atomique cible se scinde en deux, et les neutrons retardés, qui sont émis un moment plus tard, par une petite fraction de produits de fission. En effet, la plupart des produits de fissions sont instables, car ils contiennent beaucoup plus de neutrons que de protons. Pour perdre leur excédent de neutrons, ces produits de fissions vont subir une ou plusieurs désintégrations radioactives de type β^- (où un des neutrons devient un proton en éjectant un électron et un antineutrinos ${}^1_0\mathrm{n} \Longrightarrow {}^1_1\mathrm{p} + {}^0_{-1}\mathrm{e} + {}^0_0\overline{\nu}$) jusqu'à se transformer en un noyau atomique stable. Cependant, il arrive qu'une des désintégrations de la chaîne laisse le noyau atomique résiduel dans un état suffisamment excité ${}^A_Z\mathrm{X} \Longrightarrow {}^A_{Z+1}\mathrm{Y}^* + {}^0_{-1}\mathrm{e} + {}^0_0\overline{\nu}$ pour que celui-ci émette spontanément un neutron ${}^A_{Z+1}\mathrm{Y}^* \Longrightarrow {}^{A-1}_{Z+1}\mathrm{Y} + {}^1_0\mathrm{n}$. La durée qui s'écoule entre la réaction de fission et l'émission de ce neutron retardé dépend principalement de la période radioactive du noyau, dit « précurseur », obtenu juste avant le noyau résiduel. Ce délai, compris entre la dixième de seconde et la minute, est très supérieur au temps de vie moyen des neutrons dans le cœur, plutôt de l'ordre de quelques microsecondes. Ainsi, bien que les neutrons retardés ne représentent qu'une infime partie des neutrons émis (environ $0.7\,\% = 700\,\mathrm{pcm}$ de la population de neutrons), leur influence sur la cinétique de la réaction en chaîne est considérable.

L'approche la plus simple pour décrire l'évolution de la population de neutrons est de réduire

TABLE 2.1 – Caractéristiques des six groupes de précurseurs de l'uranium 235 (fission thermique)

Groupe g	**Noyaux précurseurs**	**Période** $T_g^{1/2}$ (s)	**Fraction** β_g (pcm)
1	^{87}Br, ^{142}Cs	55,72	24
2	^{137}I, ^{88}Br	22,72	123
3	^{138}I, ^{89}Br, ^{93}Rb, ^{94}Rb	6,22	117
4	^{139}I, ^{93}Kr, ^{94}Kr, ^{143}Xe, ^{90}Br, ^{92}Br	2,3	262
5	^{140}I, ^{145}Cs	0,61	108
6	Br, Rb, As, etc.	0,23	45
Moyen	Total	8,157	679

le cœur du réacteur à un point en négligeant ses dimensions axiale et radiale. Étant donné que la notion de densité n'a plus sens, la variable $n(t)$ représente cette fois-ci le nombre de neutrons observés à un instant donné. De plus, comme les produits de fission capables d'émettre des neutrons se comptent par centaines, les noyaux **précurseurs de neutrons retardés** ne sont pas considérés individuellement, mais sont rassemblés en typiquement 6 ou 8 groupes de noyaux atomiques ayant des périodes radioactives similaires. Chaque groupe $g \in \mathbb{N}$ de précurseurs est alors caractérisé par une constante de décroissance radioactive $\lambda_g \coloneqq \ln(2)/T_g^{1/2}$ et une **fraction de neutrons retardés** β_g, exprimée en pcm. Dans ce contexte, la variable $c_g(t)$ représente le nombre de noyaux précurseurs appartenant au groupe $g \in \mathbb{N}$ observés à un instant donné. Les équations de la cinétique ponctuelle se déduisent ensuite en évaluant le nombre de neutrons et de précurseurs qui apparaissent et disparaissent du cœur du réacteur à chaque instant :

$$\begin{aligned} \frac{\mathrm{d}n(t)}{\mathrm{d}t} &= \frac{(1-\beta)k_{\text{eff}}(t)-1}{\Lambda}n(t) + \sum_g \lambda_g c_g(t) \\ \frac{\mathrm{d}c_g(t)}{\mathrm{d}t} &= \frac{\beta_g k_{\text{eff}}(t)}{\Lambda}n(t) - \lambda_g c_g(t)\,, \end{aligned} \tag{2.13}$$

où Λ est le **temps de vie moyen** d'un neutron, de son apparition à sa disparition par fuite, capture ou fission, et $\beta \coloneqq \sum_g \beta_g$ est la **fraction totale de neutrons retardés**. Le premier terme de l'équation des neutrons correspond à la différence entre le taux de production et le taux de disparition d'une génération de neutrons prompts, tandis que le second terme correspond au taux de production de neutrons retardés par désintégration radioactive des groupes de précurseurs. Bien entendu, l'effet de cette désintégration radioactive se retrouve dans le taux de disparition des équations des précurseurs. Leur taux de production, quant à lui, est égal à la fraction de neutrons retardés qu'ils génèrent. En fait, il apparaît clairement, à l'aide du changement de variable $\widetilde{c}_g(t) \coloneqq \lambda_g c_g(t)$, que chaque groupe de précurseurs se comporte comme une

source de neutrons :

$$\begin{aligned} \frac{\mathrm{d}n(t)}{\mathrm{d}t} &= \frac{(1-\beta)k_{\text{eff}}(t)-1}{\Lambda}n(t) + \sum_g \widetilde{c}_g(t) \\ \tau_g \frac{\mathrm{d}\widetilde{c}_g(t)}{\mathrm{d}t} &= -\widetilde{c}_g(t) + K_g(t)n(t)\,, \end{aligned} \tag{2.14}$$

dont la réponse temporelle est retardée par un système linéaire du 1er ordre à gain variant $K_g(t) \coloneqq \beta_g k_{\text{eff}}(t)/\Lambda$ ayant pour constante de temps $\tau_g \coloneqq 1/\lambda_g$. Par ailleurs, une méthode souvent employée pour simplifier les équations de la cinétique ponctuelle (2.13) consiste à fusionner les différents groupes de précurseurs en un seul groupe moyen :

$$\frac{\mathrm{d}c(t)}{\mathrm{d}t} = \frac{\beta k_{\text{eff}}(t)}{\Lambda}n(t) - \lambda c(t)\,, \tag{2.15}$$

caractérisé par une constante de décroissante radioactive λ, obtenue en pondérant les constantes de temps τ_g de chaque groupe par leurs fractions de neutrons retardés β_g respectives :

$$\frac{1}{\lambda} \coloneqq \frac{\sum_g \beta_g \tau_g}{\sum_g \beta_g} = \frac{1}{\beta}\sum_g \frac{\beta_g}{\lambda_g}\,. \tag{2.16}$$

Pour simplifier encore davantage l'écriture, les équations de la cinétique ponctuelle (2.13) peuvent également s'exprimer en fonction de la réactivité :

$$\begin{aligned} \frac{\mathrm{d}n(t)}{\mathrm{d}t} &= \frac{\rho(t)-\beta}{\Lambda^*(t)}n(t) + \lambda c(t) \\ \frac{\mathrm{d}c(t)}{\mathrm{d}t} &= \frac{\beta}{\Lambda^*(t)}n(t) - \lambda c(t)\,, \end{aligned} \tag{2.17}$$

et du **temps de génération moyen** $\Lambda^*(t) \coloneqq \Lambda/k_{\text{eff}}(t)$ que met un neutron pour en produire un autre par fission. À noter que le temps de génération moyen est habituellement supposé constant, car la réactivité reste toujours très proche de zéro en fonctionnement normal :

$$\forall\, t \in \mathbb{R},\ \Lambda^*(t) = \frac{\Lambda}{k_{\text{eff}}(t)} = \Lambda\Big(1-\rho(t)\Big) \approx \Lambda,\ \text{quand}\ \ \|\rho(t)\| \approx 0\,. \tag{2.18}$$

En effet, cette dernière doit impérativement être inférieure à la fraction totale des neutrons retardés $\beta \simeq 650\,\text{pcm}$ pour éviter que le cœur ne devienne sur-critique par neutrons prompts. Si cela était le cas, le taux de production de neutrons prompts permettrait à lui seul d'augmenter la taille de la population, ce qui rendrait la réaction en chaîne totalement incontrôlable, puisque le temps de génération moyen $\Lambda^* \simeq 20\,\mu\text{s}$ est extrêmement court dans les réacteurs à eau légère. Dans les faits, les paramètres du cœur du réacteur (taille et répartition des assemblages combustibles, quantité et densité de matériaux fissiles, etc.) sont dimensionnés de façon à rendre la réaction en chaîne intrinsèquement stable. En particulier, il est crucial de s'assurer que le

coefficient de température du facteur de multiplication effectif est négatif, c'est-à-dire qu'une élévation de la température du cœur n'entraîne pas une augmentation de la réactivité. Les deux principaux phénomènes physiques qui affectent le coefficient de température sont l'effet modérateur et l'effet Doppler.

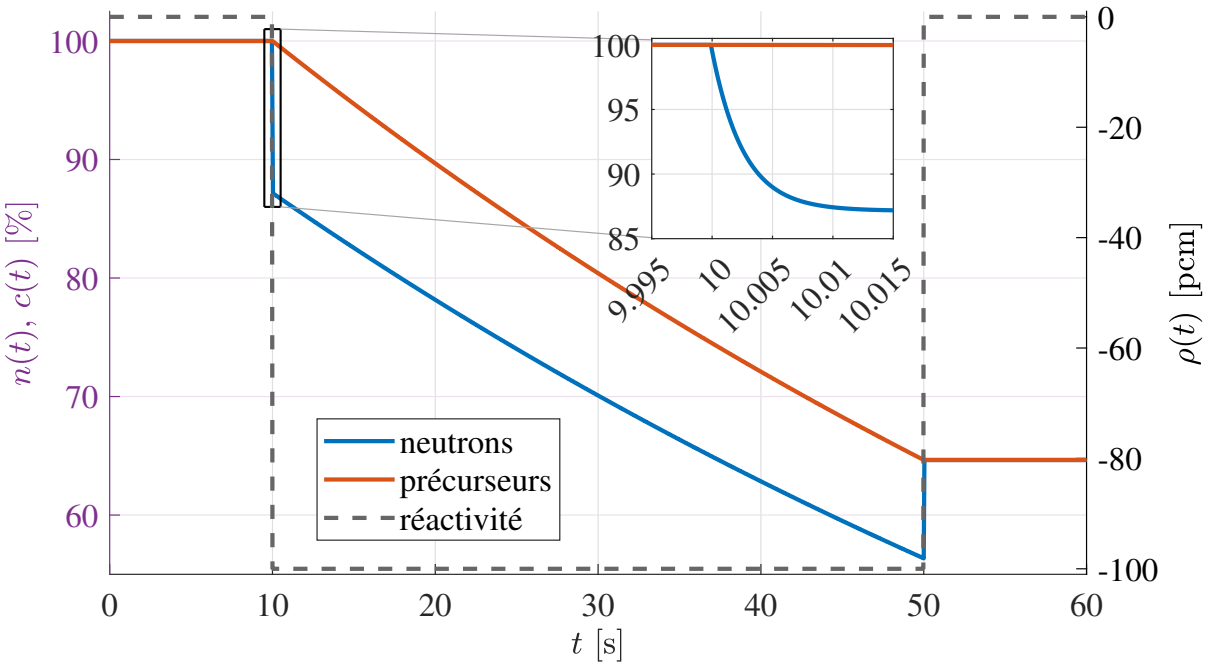

FIGURE 2.4 – Réponses temporelles des neutrons et des noyaux précurseurs données par les équations de la cinétique ponctuelle (2.17) à un créneau de réactivité de -100 pcm.

1.2.4 Effet modérateur

Lorsqu'un neutron est émis à la suite d'une fission, son énergie cinétique avoisine les 2 MeV, ce qui correspond à une vitesse d'environ 20 000 km s^{-1}. À cette vitesse, la probabilité que le neutron interagisse avec l'un des noyaux atomiques fissiles du cœur (uranium 235 ou plutonium 239) est extrêmement faible. Fort heureusement, celle-ci augmente à mesure que le neutron ralentit. Le rôle du modérateur est donc de freiner les neutrons jusqu'à ce que leur probabilité de rencontrer un noyau atomique fissile devienne suffisamment grande pour entretenir la réaction en chaîne. Pour ce faire, les neutrons doivent céder la majeure partie de leur énergie cinétique aux noyaux atomiques du modérateur par le biais de diffusions successives. Une fois complètement ralentis, les neutrons atteignent une vitesse d'environ 2 km s^{-1}, soit une énergie cinétique proche de 1/40 eV, et sont qualifiés de lents, ou plutôt de thermiques, car l'agitation thermique du milieu dans lequel ils se trouvent leur apporte désormais autant d'énergie cinétique qu'ils n'en perdent par diffusion. À noter, par ailleurs, que le nombre de diffusions nécessaires pour arri-

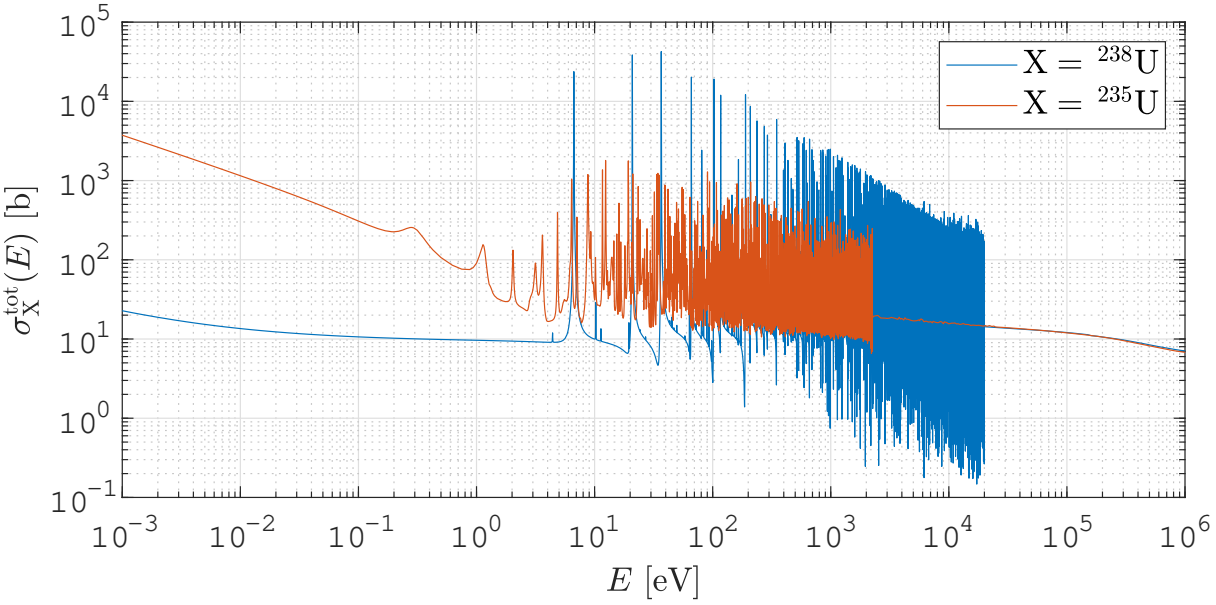

FIGURE 2.5 – Sections efficaces microscopiques totales de l'uranium 235 et de l'uranium 238.

ver à l'équilibre thermique avec la matière sera d'autant plus faible que la masse des noyaux atomiques du modérateur est proche de celle du neutron incident.

Dans les réacteurs à eau légère [1], notamment les réacteurs à eau sous-pression, les atomes d'hydrogène du fluide caloporteur font office de modérateur, ce qui contribue à stabiliser la réaction en chaîne. En effet, lorsque la puissance du cœur augmente, la température du fluide caloporteur, à savoir l'eau du circuit primaire, augmente également. Or, à cause de la dilatation thermique, la distance entre les molécules d'eau du circuit primaire s'accroît (autrement dit, la densité de l'eau devient plus faible). De ce fait, les neutrons ont moins de chance d'interagir avec les noyaux atomiques d'hydrogène, et mettent donc en moyenne plus de temps à être ralentis par le modérateur. Par conséquent, la puissance du cœur du réacteur finit par diminuer, car les réactions de fission sont moins fréquentes qu'auparavant. À l'inverse, lorsque la puissance du cœur diminue, la température de l'eau du circuit primaire diminue elle aussi, si bien que sa densité augmente. Dès lors, comme les neutrons sont à présent mieux ralentis par le modérateur, la puissance du cœur finira tôt ou tard par réaugmenter avant de se stabiliser de nouveau.

L'effet modérateur désigne l'ensemble des variations de réactivité induites par le changement de densité du milieu utilisé pour ralentir les neutrons. Une conséquence notable de l'effet modérateur est, qu'à l'équilibre, la puissance générée dans la partie basse du cœur sera légèrement supérieure à celle générée dans sa partie haute, puisque la température de l'eau du circuit

1. L'eau légère ($^1_1H_2O$), par opposition à l'eau lourde ($^2_1H_2O$), a tendance à trop capturer les neutrons, ce qui explique pourquoi le combustible nucléaire doit être enrichi à hauteur d'environ 3-4 % d'uranium 235 fissile, ce dernier ne représentant que 0.72 % de l'uranium naturel (minerai d'uranium présent naturellement dans le sol).

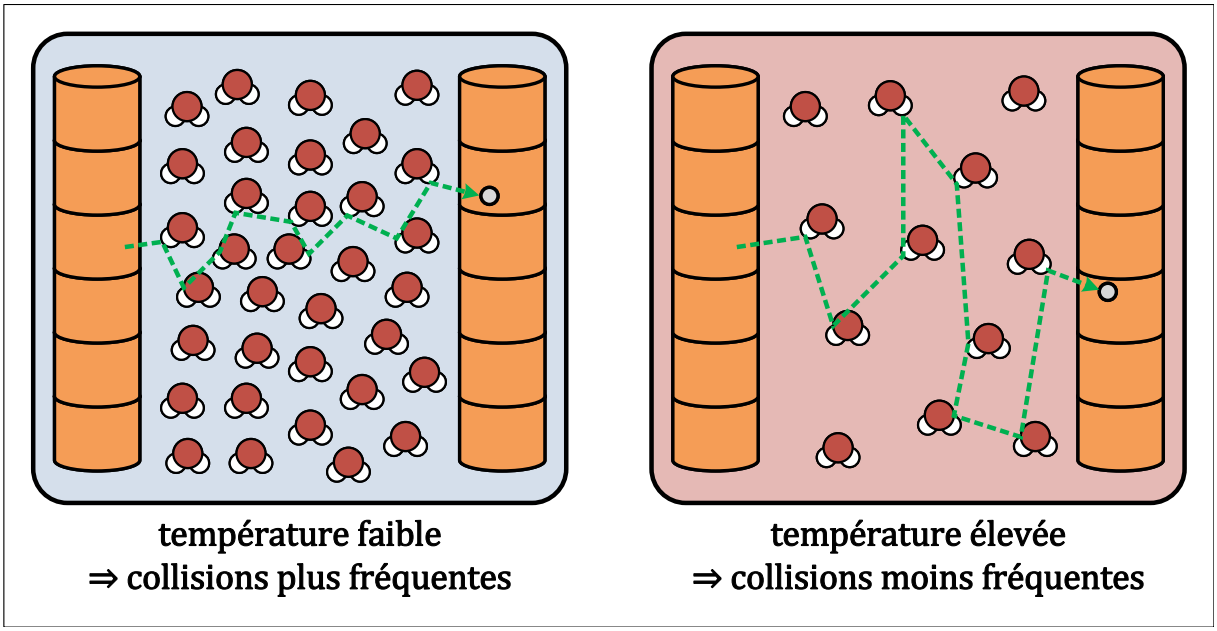

FIGURE 2.6 – Illustration de l'effet du changement de densité du milieu modérateur sur le temps de ralentissement des neutrons.

primaire est toujours moins élevée en entrée qu'en sortie du cœur.

1.2.5 Effet Doppler

Les deux isotopes primordiaux de l'uranium sont l'uranium 235 et l'uranium 238. Le noyau atomique de l'uranium 235 est qualifié de fissile, car il est susceptible de subir une fission quelle que soit l'énergie cinétique du neutron absorbé. Le noyau atomique de l'uranium 238, en revanche, n'est pas fissile, car celui-ci ne pourra subir de fission que si l'énergie cinétique du neutron absorbé dépasse un certain seuil, supérieur à 1 MeV. Dans le cas contraire, le neutron incident sera simplement capturé par le noyau atomique de l'uranium 238, sans qu'aucun nouveau neutron ne soit émis : la capture est alors dite fertile si la désintégration radioactive du noyau composé, l'uranium 239, conduit plus tard à la formation d'un noyau fissile (plutonium 239 ou plutonium 241), et stérile sinon.

Lorsque les neutrons sont ralentis par le modérateur, leur énergie cinétique tombe rapidement en-dessous du seuil de fission de l'uranium 238. Une partie d'entre eux sera donc capturée de manière stérile par l'uranium 238, qui représente plus de 95 % des noyaux atomiques du combustible nucléaire, et ne participeront plus à l'entretien de la réaction en chaîne. La probabilité qu'un neutron se fasse capturer par les noyaux atomiques d'uranium 238 n'est pas uniforme, mais comporte une multitude de pics de résonances très élevés, dont l'apparence change en fonction de la température du combustible nucléaire. Très schématiquement, ces pics de résonances

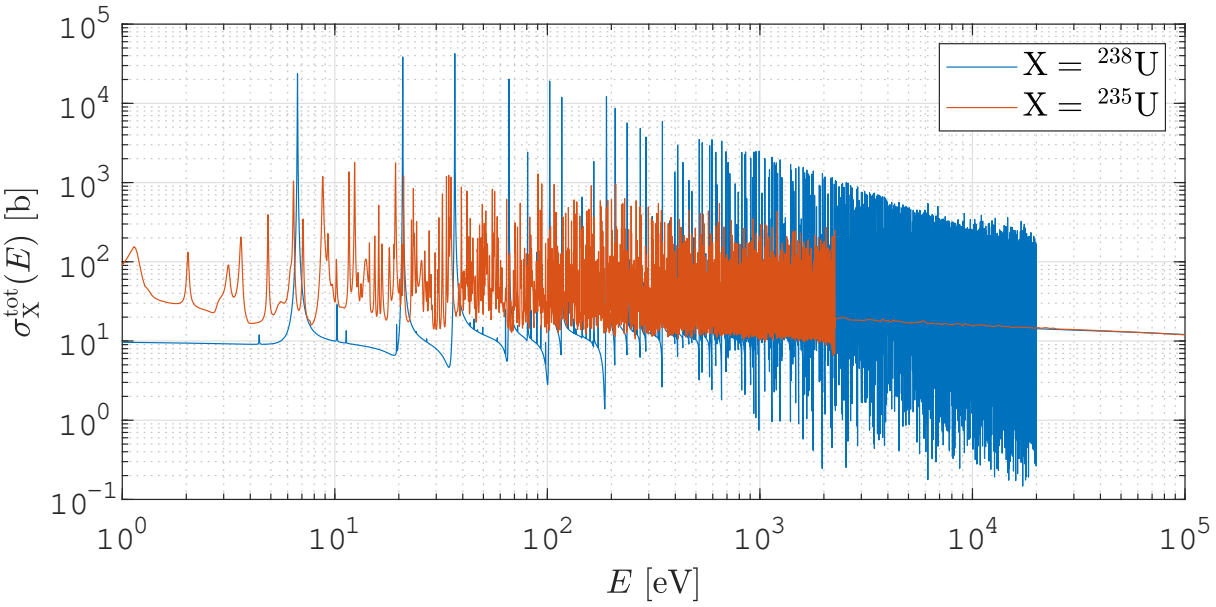

FIGURE 2.7 – Sections efficaces microscopiques totales de l'uranium 235 et de l'uranium 238 dans le domaine des neutrons épithermiques.

s'aplatissent et s'élargissent à mesure que la température du combustible croît, ce qui augmente la probabilité que les neutrons soient capturés par les noyaux atomiques d'uranium 238. À l'inverse, les résonances de capture de l'uranium 238 s'étirent et s'amincissent à mesure que la température du combustible décroît, ce qui diminue la probabilité que les neutrons soient capturés par les noyaux atomiques d'uranium 238. Ce phénomène d'élargissement ou d'amincissement des résonances est appelé effet Doppler, car il découle du changement de vitesse relative entre le neutron incident et le noyaux atomique cible, induit par les variations de température du milieu combustible.

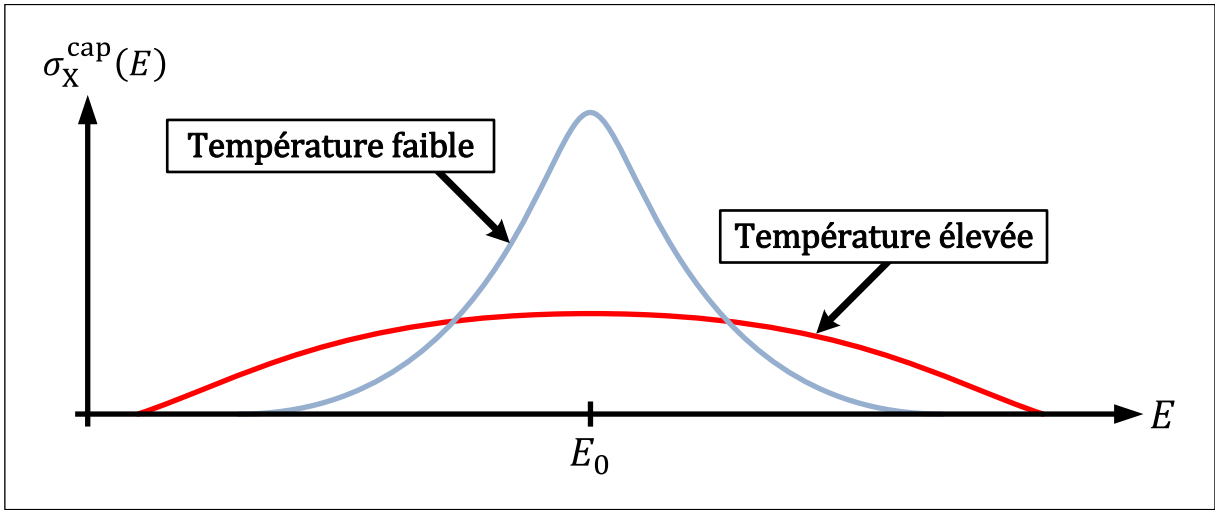

FIGURE 2.8 – Illustration du phénomène d'élargissement et d'amincissement des résonances de capture de l'uranium 238 en fonction de la température du combustible nucléaire.

À l'instar de l'effet modérateur, l'effet Doppler stabilise la réaction en chaîne, puisqu'il s'oppose aux variations de température du combustible nucléaire. Toutefois, l'effet Doppler se manifeste beaucoup plus rapidement que l'effet modérateur, étant donné que la température du combustible varie presque instantanément quand la puissance du cœur du réacteur est modifiée.

1.2.6 Empoisonnement au xénon

L'uranium 238 n'est pas le seul élément chimique qui puisse capturer un nombre important de neutrons de manière stérile. Les éléments chimiques ayant le plus de chance de capturer des neutrons sont appelés des poisons neutroniques. Le xénon 135, en particulier, est le poison neutronique qui possède la plus grande capacité de capture de neutrons thermiques : la probabilité qu'un neutron thermique soit capturé par un noyau de xénon 135 est environ 4500 fois plus élevée que celle de provoquer une fission avec un noyau d'uranium 235, toutes choses égales par ailleurs. De ce fait, l'évolution de la densité particulaire de xénon 135 dans le combustible nucléaire aura une influence considérable sur la réaction en chaîne, l'anti-réactivité apportée par ce poison neutronique étant de l'ordre de $-3000\,\text{pcm}$ pour un réacteur à eau sous pression en fonctionnement normal.

Bien que le xénon 135 soit en partie produit par fission ($\gamma_{Xe} = 0.1\,\%$), la majorité de celui-ci provient de la désintégration radioactive de l'iode 135, lui-même obtenu par fission et par désintégration radioactive du tellure 135. Néanmoins, comme la période radioactive du tellure 135 est négligeable devant celles de l'iode 135 et du xénon 135 ($T_{Te}^{1/2} = 19.2\,\text{s}$ contre $T_I^{1/2} = 6.53\,\text{h}$ et $T_{Xe}^{1/2} = 9.17\,\text{h}$ respectivement), il est souvent admis que l'iode 135 est directement produit par fission, avec un rendement cumulé de $\gamma_I = 6.4\,\%$ (somme du rendement de fission du tellure 135

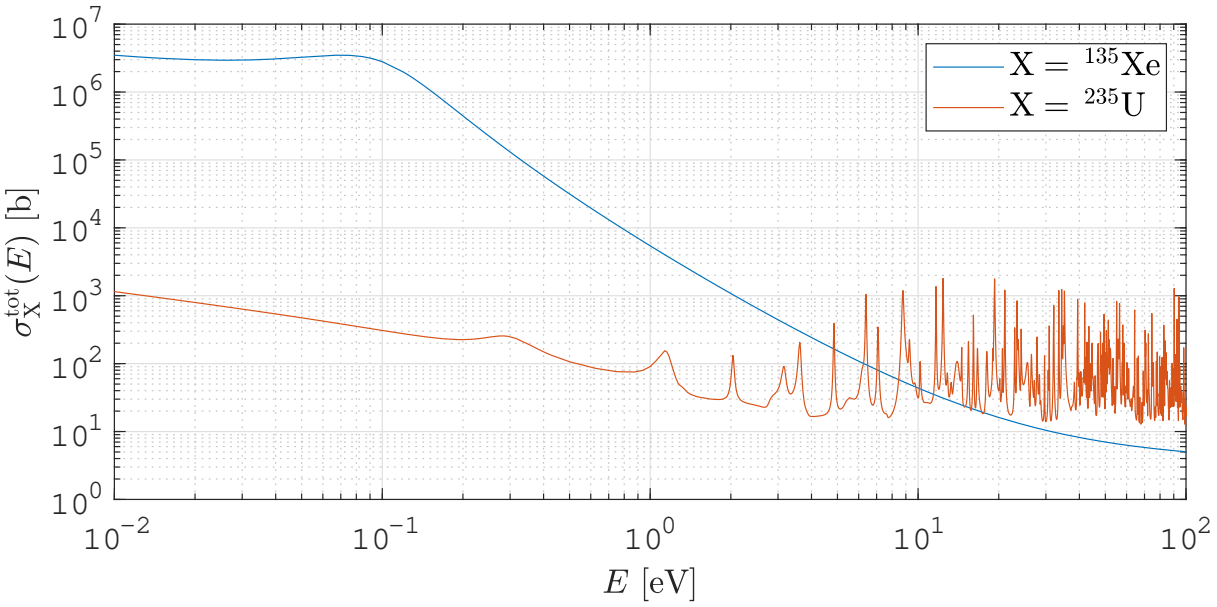

FIGURE 2.9 – Sections efficaces microscopiques totales de l'uranium 235 et du xénon 135 autour du domaine des neutrons thermiques.

et de l'iode 135). Le xénon 135 disparaît ensuite soit par désintégration radioactive, en se transformant en césium 135, soit par capture neutronique stérile, en se transformant en xénon 136. Les équations de Bateman qui décrivent l'évolution des densités particulaires d'iode 135 et de xénon 135 sont données par :

$$\begin{aligned} \frac{\mathrm{dI}^{135}(t)}{\mathrm{d}t} &= \gamma_{\mathrm{I}}\Sigma_{\mathrm{fis}}\phi(t) - \lambda_{\mathrm{I}}\mathrm{I}^{135}(t) \\ \frac{\mathrm{dXe}^{135}(t)}{\mathrm{d}t} &= \gamma_{\mathrm{Xe}}\Sigma_{\mathrm{fis}}\phi(t) + \lambda_{\mathrm{I}}\mathrm{I}^{135}(t) - \Big(\lambda_{\mathrm{Xe}} + \sigma_{\mathrm{Xe}}^{\mathrm{cap}}\phi(t)\Big)\mathrm{Xe}^{135}(t)\,. \end{aligned} \tag{2.19}$$

Le couplage entre l'iode 135 et le xénon 135 peut être facilement comparé à celui d'un système à deux réservoirs. Dans cette analogie, les densités particulaires d'iode $\mathrm{I}^{135}(t)$ et de xénon $\mathrm{Xe}^{135}(t)$ sont représentées par le niveau d'eau de leurs réservoirs respectifs, tandis que le flux neutronique $\phi(t)$ correspond au taux d'ouverture de deux vannes de régulation de débit connectées entre elles. Ces vannes agissent simultanément sur le débit d'entrée des deux réservoirs, par le biais des réactions de fission, et sur le débit de sortie du réservoir de xénon, par le biais des réactions de capture. Le réservoir d'iode, situé en amont, se vide naturellement par décroissance radioactive dans le réservoir de xénon, situé en aval, à la manière de l'eau s'écoulant par gravité. Comme le tuyau qui relie la première vanne de régulation de débit au réservoir de xénon est beaucoup moins large que celui en sortie du réservoir d'iode, le débit d'entrée du réservoir de xénon dépend surtout du débit de fuite du réservoir d'iode. Le réservoir de xénon se vide

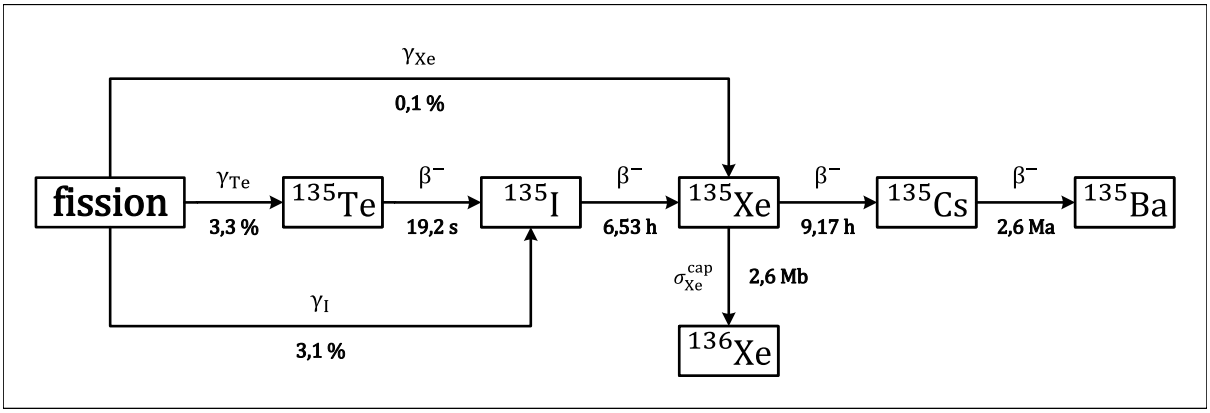

FIGURE 2.10 – Chaîne d'évolution simplifiée de l'iode 135 et du xénon 135, issue de la fission de l'uranium 235 dans le domaine thermique.

également de façon naturelle par décroissance radioactive, mais à une vitesse sensiblement plus lente que le réservoir d'iode. De ce fait, le débit de sortie du réservoir de xénon dépend avant tout du taux d'ouverture de la seconde vanne de régulation de débit.

Lorsque le flux neutronique est constant, le niveau des deux réservoirs reste stable, car leurs débits d'entrée et de sortie s'équilibrent au bout d'un certain temps. Si le flux neutronique diminue en partant de cet état d'équilibre, le niveau du réservoir d'iode va baisser jusqu'à ce que son débit d'entrée compense de nouveau les fuites liées à sa décroissance radioactive. Le niveau du réservoir de xénon, en revanche, va d'abord augmenter en passant par un pic, car les fuites liées aux captures neutroniques diminuent avant même que le débit de sortie du réservoir d'iode ne commence à baisser. L'amplitude et la durée du pic xénon seront alors d'autant plus élevées que la différence entre l'ancienne et la nouvelle valeur de flux neutronique est importante. Une fois le pic terminé, le réservoir de xénon commence à se vider jusqu'à atteindre un nouveau niveau d'équilibre, plus faible que celui duquel il est initialement parti. Les mêmes variations d'iode et de xénon se produiront en sens inverse si le flux neutronique réaugmente par la suite.

En résumé, la production de xénon est principalement dictée par l'historique du flux neutronique, symbolisé par le réservoir d'iode, tandis que sa disparition est principalement dictée par la valeur instantanée du flux neutronique, symbolisée par la seconde vanne de régulation de débit. Ce décalage temporel entre la production et la disparition de xénon peut, à long terme, faire osciller de manière instable la puissance du cœur si la distribution spatiale du flux neutronique n'est pas uniforme.

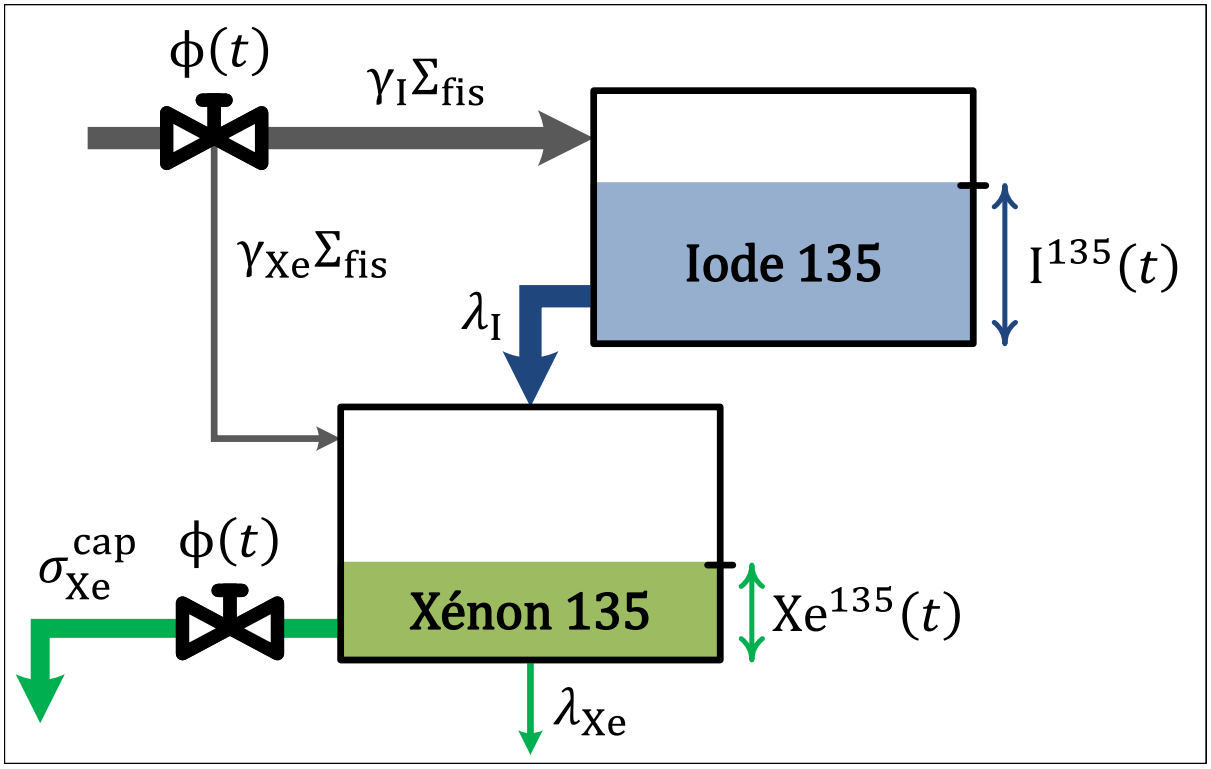

FIGURE 2.11 – Illustration du couplage entre l'iode et le xénon à l'aide d'un système à deux réservoirs.

1.3 Pilotage des réacteurs à eau sous pression

1.3.1 Description et enjeux d'une variation de charge

En dehors des transitoires d'arrêt et de redémarrage, réalisés spécifiquement pour remplacer le combustible usé ou mener des opérations de maintenance, les centrales nucléaires peuvent être amenées à modifier la puissance électrique délivrée par leur turbine à la demande du gestionnaire du réseau de transport d'électricité (RTE en France). Ces variations de puissance, indispensables pour maintenir l'équilibre entre la production et la demande d'électricité, sont habituellement classées en trois catégories, qui se distinguent par l'amplitude, la vitesse, le niveau d'occurrence, et le caractère aléatoire ou planifié du transitoire réalisé :

1) Le **suivi de charge** permet de compenser la majeure partie des fluctuations journalière de la demande d'électricité. Les variations de puissance réalisées en suivi de charge peuvent être effectuées à une vitesse maximale de 5 %PN min^{-1}, pour les centrales les plus flexibles, et couvrent une plage de fonctionnement comprise entre 30 %PN et 100 %PN. Le profil de charge que la turbine doit suivre quotidiennement est envoyé plusieurs heures à l'avance aux opérateurs de la centrale, mais celui-ci peut être modifié en urgence par le répartiteur du réseau, jusqu'à une dizaine de minutes avant de débuter la prochaine variation de

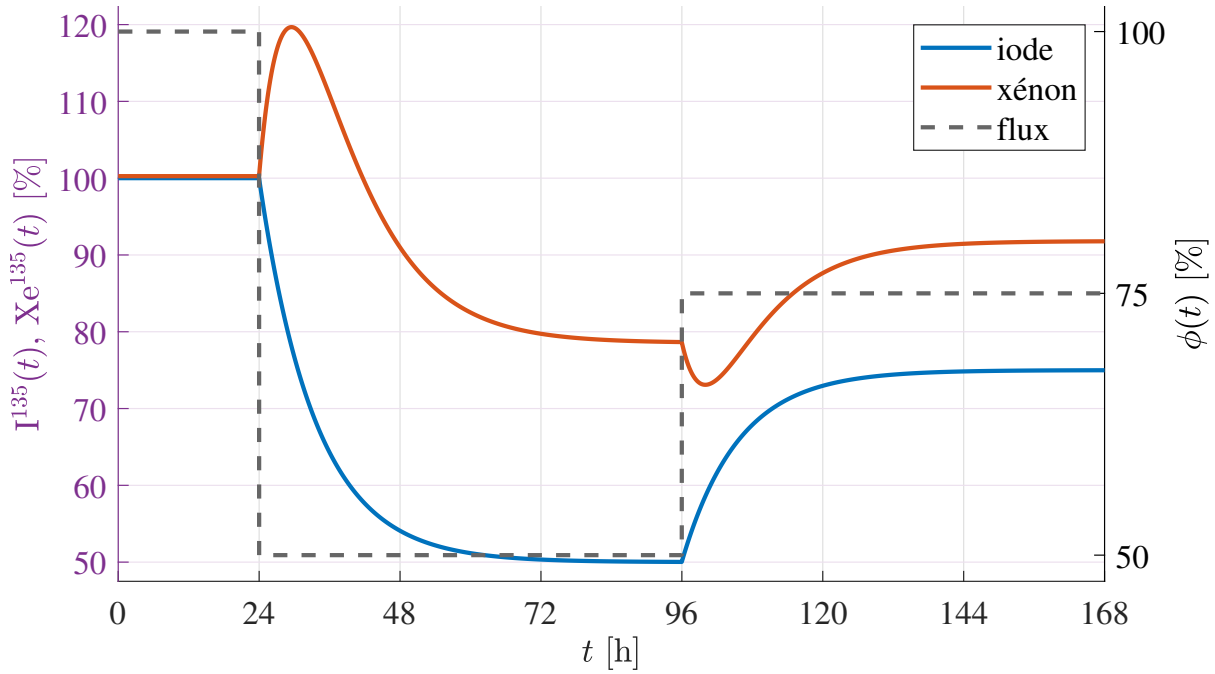

FIGURE 2.12 – Réponses temporelles des densités d'iode et de xénon données par les équations de Bateman (2.19) à des échelons de flux neutronique d'amplitudes différentes ($-50\,\%$ et $+25\,\%$).

puissance. En général, le profil de charge comprend deux variations de puissance espacées de quelques heures (par exemple une baisse de charge le soir et une reprise de charge le matin) durant lesquelles la turbine reste à puissance constante (par exemple 50 %PN en palier bas et 100 %PN en palier haut).

2) Le **réglage de fréquence primaire** permet de corriger automatiquement une partie du déséquilibre de fréquence du réseau lié aux fluctuations aléatoires de la demande d'électricité à l'échelle locale. Le signal de réglage de fréquence primaire envoyé à la centrale est proportionnel à l'écart entre la fréquence mesurée et sa valeur de référence (50 Hz en Europe), et change donc de façon imprévisible, en temps réel, à une vitesse d'environ $1\,\%\mathrm{PN\,s^{-1}}$. En France, l'amplitude des variations de puissances engendrées par le réglage de fréquence primaire est comprise entre $\pm 2.5\,\%$PN, ce qui correspond à un écart de fréquence de $\pm 50\,\mathrm{mHz}$.
3) Le **réglage de fréquence secondaire** permet de résorber automatiquement les écarts de fréquence et de puissance qui apparaissent entre les différentes zones géographiques du réseau, en raison du statisme du réglage de fréquence primaire et des échanges d'électricité avec les autres pays européens. Le signal de réglage de fréquence secondaire envoyé à la centrale est déterminé à l'échelle nationale par le gestionnaire du réseau de transport d'électricité. Il s'agit d'un signal de type proportionnel-intégral qui évolue beaucoup plus lentement, à une vitesse d'environ $1\,\%\mathrm{PN\,min^{-1}}$, que le signal de réglage de fréquence

primaire. Du fait de sa conception, l'amplitude des variation de puissance engendrées par le réglage de fréquence secondaire est comprise entre ±5 %PN en France.

Le type et le nombre de transitoires que pourra réaliser une centrale nucléaire est établi conjointement avec l'exploitant (EDF en France) en fonction, d'une part, des contraintes thermomécaniques que peuvent supporter ses éléments constitutifs et, d'autre part, de sa capacité à rejeter les perturbations causées par les variations de charge de la turbine. Ces perturbations vont affecter deux grandeurs fondamentales vis-à-vis du pilotage de la réaction en chaîne, à savoir :

1) La **température moyenne du circuit primaire** $T_{\text{moy}}(t)$, calculée à partir de la température de l'eau mesurée en entrée $T_{\text{in}}(t)$ et en sortie $T_{\text{out}}(t)$ du cœur du réacteur :

$$T_{\text{moy}}(t) \coloneqq \frac{T_{\text{in}}(t) + T_{\text{out}}(t)}{2} \,. \tag{2.20}$$

2) Le **déséquilibre axial de puissance du cœur** $AO(t)$, pour *Axial Offset*, calculé à partir de la puissance $P_H(t)$ mesurée dans sa partie haute et $P_B(t)$ mesurée dans sa partie basse :

$$AO(t) \coloneqq \frac{P_H(t) - P_B(t)}{P_H(t) + P_B(t)} \,. \tag{2.21}$$

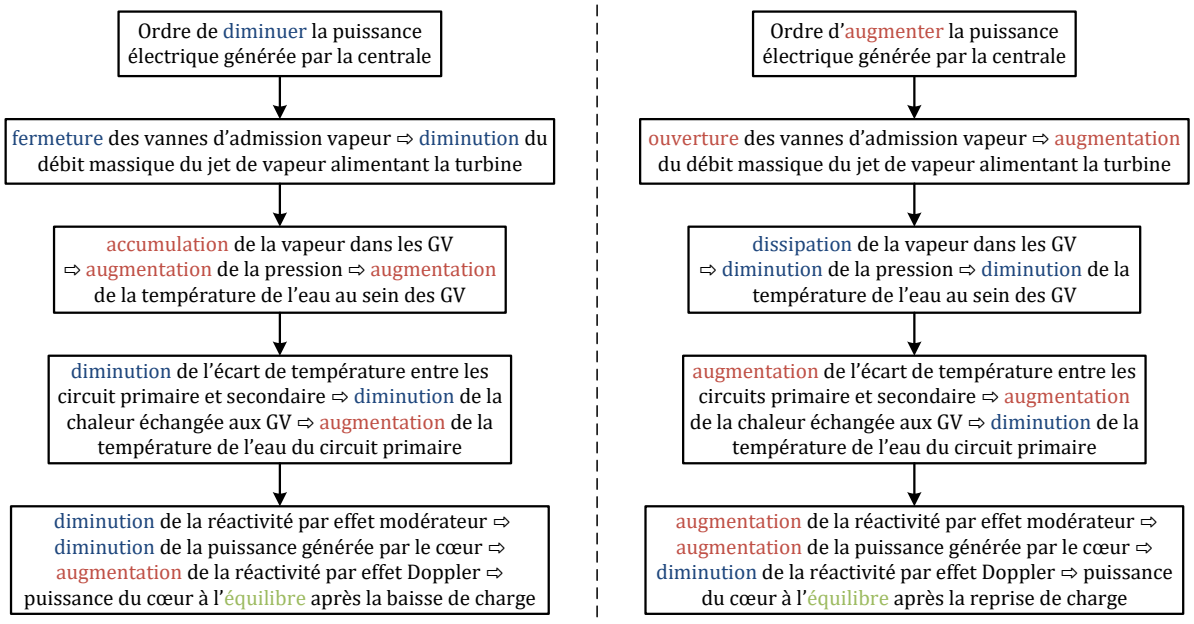

FIGURE 2.13 – Résumé du déroulement d'une variation de charge.

En effet, pendant une baisse de charge (resp. reprise de charge), la puissance électrique générée par la centrale est progressivement réduite (resp. amplifiée) en diminuant (resp. augmentant) le débit du jet de vapeur qui alimente le groupe turbo-alternateur. Cette diminution (resp. aug-

mentation) du débit vapeur augmente (resp. diminue) la pression, et donc la température, de l'eau du circuit secondaire contenue à l'intérieur des générateurs de vapeur. De ce fait, la quantité de chaleur extraite du circuit primaire vers le circuit secondaire devient moins (resp. plus) importante, ce qui provoque une augmentation (resp. diminution) de la température de l'eau du circuit primaire. Or, étant donné que le cœur est conçu de telle sorte que le coefficient de température du facteur de multiplication effectif soit négatif, cette augmentation (resp. diminution) de température fera diminuer (resp. augmenter) la réactivité jusqu'à ce qu'elle atteigne un nouvel état d'équilibre une fois la baisse (resp. la reprise) de charge terminée. La réaction en chaîne est donc capable de s'auto-réguler face aux variations de charge de la turbine du fait de la nature stabilisante des contre-réactions neutroniques (effets Doppler et modérateur). La puissance relative du cœur du réacteur, en particulier, suit naturellement celle de la turbine, avec un léger retard dû à l'inertie thermique du générateur de vapeur. Cependant, en l'absence de moyen de contrôle de la réactivité, la température moyenne de l'eau du circuit primaire et la distribution axiale de puissance du cœur risquent de s'éloigner grandement des conditions limites d'exploitation. Ces limites, définies en bureau d'étude, permettent d'obtenir des performances opérationnelles optimales au regard des exigences de sûreté de la centrale :

- Concernant la température moyenne du circuit primaire, il s'agit de suivre, à une certaine tolérance près, une température de référence qui évolue en fonction de la puissance relative de la turbine. Ce programme de température, qui diffère d'un type de réacteurs à l'autre, vise à maximiser le rendement du cycle thermodynamique de la centrale sans faire bouillir l'eau du circuit primaire.
- Concernant le déséquilibre axial de puissance du cœur, il s'agit de rester, à une certaine tolérance près, autour d'une valeur de référence fixe afin d'éviter de créer des pics de puissances locaux et de déclencher des **oscillations xénon**. Ces oscillations de puissance apparaissent typiquement lorsque la distribution de flux neutronique est soudainement perturbée, et que le déséquilibre qui s'ensuit n'est pas corrigé. Par exemple, entre le début et la fin d'une baisse de charge, la température de l'eau du circuit primaire augmentera plus fortement en entrée qu'en sortie du cœur, ce qui signifie que la réactivité, et donc le flux neutronique, diminuera davantage dans sa partie basse que dans sa partie haute. Cette perte de symétrie initiale de la distribution de puissance va s'accentuer à court terme, car le xénon disparaîtra plus rapidement dans la région où le flux neutronique est le plus élevé, c'est-à-dire en haut du cœur actuellement. La perturbation va ensuite s'inverser à long terme, car le surplus d'iode généré entre-temps dans la région à haut flux neutronique va progressivement se désintégrer en xénon. L'anti-réactivité apportée par ce dernier sera alors plus importante dans la partie haute que dans la partie basse du cœur, ce qui aura pour conséquence de faire basculer la distribution de puissance dans l'autre sens. Si cette situation perdure, les oscillations de puissance continueront de s'amplifier,

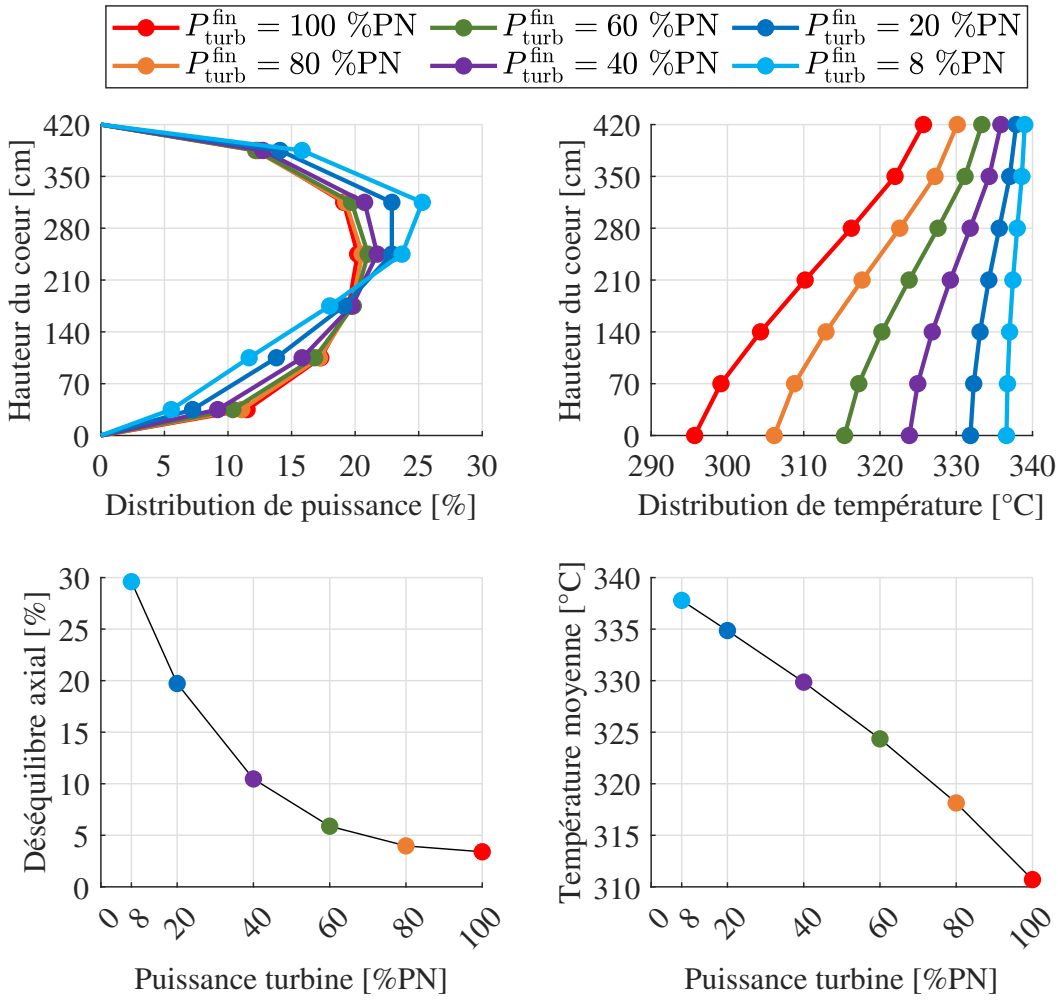

FIGURE 2.14 – Distributions de puissance et de température du cœur obtenues après avoir réalisé différentes baisses de charge sans contrôle de la réactivité en partant de 100 %PN.

jusqu'à ce que l'évolution du xénon soit parfaitement en opposition de phase avec celle du flux neutronique. Le risque, à ce stade, serait que la totalité de la puissance libérée finisse par se retrouver dans une seule moitié du cœur, ce qui conduirait à l'arrêt automatique du réacteur une fois les seuils de puissance linéique autorisés dépassés.

1.3.2 Caractéristiques des actionneurs du cœur

Deux mécanismes de contrôle de la réactivité ont été mis au point pour compenser l'impact des contre-réactions neutroniques et du xénon sur la température moyenne du circuit primaire et le déséquilibre axial de puissance du cœur :

1) Les **grappes de commande**, souvent appelées barres de contrôle par abus de langage,

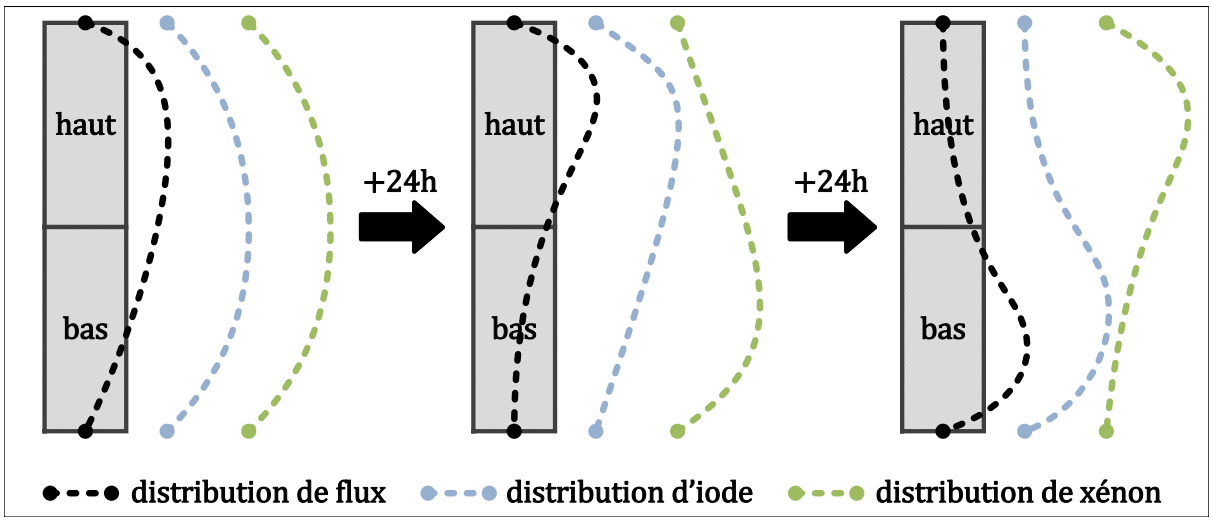

FIGURE 2.15 – Illustration du phénomène de basculement de la distribution de puissance du cœur lié aux oscillations xénon.

TABLE 2.2 – Composition des grappes de commande des réacteurs nucléaires du parc français.

Réacteur	Grappes noires	Grappes grises
900 MWe	24 crayons en AIC	8 crayons en AIC 16 crayons en acier
1300 MWe	24 crayons en AIC (partie basse) et B_4C (partie haute)	8 crayons en AIC 16 crayons en acier
1450 MWe	24 crayons en AIC (partie basse) et B_4C (partie haute)	12 crayons en AIC 12 crayons en acier
1650 MWe	24 crayons en AIC (partie basse) et B_4C (partie haute)	8 crayons en AIC 16 crayons en acier

sont des ensembles solidaires et mobiles de crayons recouverts d'acier, fabriqués à partir de matériaux capturant les neutrons thermiques. Elles permettent de modifier la réactivité du cœur en s'insérant ou s'extrayant par le haut de tubes guides, situés à l'intérieur de certains assemblages combustibles. Les **grappes noires**, très absorbantes, sont constituées uniquement de crayons neutrophages, contenant un alliage argent-indium-calcium (AIC) et/ou du carbure de bore (B_4C). Les **grappes grises**, moins absorbantes, sont constituées à la fois de crayons neutrophages en AIC et de crayons inertes en acier. Les cœurs des réacteurs du parc français sont équipés de 32 à 36 grappes de commandes réparties, par quadruplets de grappes grises ou de grappes noires, en 4 ou 5 groupes différents. L'ordre dans lequel se déplacent ces groupes est préétabli en bureau d'étude, via une contrainte de recouvrement (ou de chevauchement/écartement) fixe, si bien que seuls un ou deux

actionneurs sont utilisés, en pratique, pour manipuler toutes les grappes.
Les grappes de commandes peuvent répondre rapidement aux changements de réactivité du cœur, ce qui les rend adaptées au contrôle de la température moyenne lors des variations de charges. Néanmoins, les mouvements effectués par celles-ci perturbent la distribution axiale de puissance, puisque la réactivité est surtout modifiée dans la région du cœur où elles se trouvent. Par exemple, quand une grappe très neutrophage s'insère dans la partie haute du cœur, le déséquilibre axial de puissance et la température moyenne commencent, sans surprise, par diminuer. En revanche, dès que la grappe arrive dans la partie basse du cœur, le déséquilibre axial de puissance réaugmente tandis que la température moyenne continue de diminuer. Le déséquilibre axial de puissance atteint en fait sa valeur minimale lorsque la grappe est à moitié insérée dans le cœur, la différence de puissance entre ses deux parties étant alors maximale. Les contraintes de recouvrements auxquels sont soumis les groupes permettent notamment d'éviter que les grappes ne déforment la distribution de puissance du cœur de façon irrégulière et imprévisible. Dans l'exemple précédent, le déséquilibre axial de puissance pourrait continuer de diminuer avec la température moyenne si une deuxième grappe, plus neutrophage, commençait à s'insérer dans la partie haute du cœur au moment où la première atteignait sa partie basse. En termes de contraintes, cela reviendrait à imposer que la deuxième grappe soit entraînée par la première lorsque celle-ci arrive à la moitié du cœur.
Par ailleurs, les groupes de grappes ne doivent pas dépasser certaines **limites d'insertion**, déterminées en bureau d'étude, afin de garder une réserve suffisante d'anti-réactivité en cas d'arrêt automatique du réacteur, et de limiter l'afflux de réactivité dans le cœur en cas d'éjection accidentelle d'une des grappes.

2) Le **bore 10 soluble** est un puissant poison neutronique présent en petite quantité, sous forme d'acide borique (H_3BO_3), dans l'eau du circuit primaire. La concentration en bore, exprimée en partie par million ($1\,\text{ppm} = 1\,\text{mg}\,\text{kg}^{-1}$), peut être augmentée ou diminuée en injectant un certain volume d'eau borée (concentration supérieure à 7000 ppm) ou d'eau claire (concentration nulle) dans le circuit primaire. Un volume d'eau équivalent à celui injecté sera alors automatiquement retiré du circuit primaire, pour que sa masse, de l'ordre de 300 t, reste toujours constante. Comme l'eau du circuit primaire circule à un débit avoisinant les $22\,\text{m}^3\,\text{s}^{-1}$, le bore soluble est mélangé de façon presque homogène dans tout le cœur. Ainsi, contrairement aux mouvements de grappes, les variations de concentration en bore permettent de modifier la réactivité du cœur sans trop perturber sa distribution axiale de puissance. Néanmoins, ces variations de concentration prennent plusieurs minutes à se réaliser, car les solutions d'eau borée et d'eau claire injectées doivent d'abord traverser le circuit volumétrique et chimique avant de parvenir au circuit primaire. L'actionneur de bore est donc surtout employé pour contrer l'évolution du xénon après les

variations de charge.
Par ailleurs, le bore soluble est également utilisé pour contrebalancer la perte de réactivité du cœur, due à l'épuisement de la matière fissile du combustible nucléaire. Pour ce faire, la concentration en bore du circuit primaire est lentement réduite, du début à la fin du cycle d'irradiation, jusqu'à ce que la réaction en chaîne ne puisse plus être entrenue. La concentration en bore du circuit primaire se situe habituellement autour de 1200 ppm en début de cycle, contre seulement 200 ppm à 80 % d'avancement du cycle et 10 ppm en toute fin de cycle. Cette baisse de concentration en bore diminue l'efficacité des injections d'eau claire, car la quantité de bore soluble contenue dans l'eau évacuée du circuit primaire sera moins importante en fin de cycle qu'en début de cycle.
Cependant, le véritable inconvénient du bore soluble est de dégrader le coefficient de température du facteur de multiplication effectif. En effet, étant donné que la dilatation thermique de l'eau du circuit primaire augmente la distance entre les molécules d'acide borique, une élévation de la température du cœur entraînera une diminution de la fréquence des captures neutroniques du bore, autrement dit, une augmentation de la réactivité. Par conséquent, il existe une **concentration en bore limite** au-delà de laquelle le coefficient de température du facteur de multiplication effectif est susceptible de devenir positif.

1.3.3 Présentation des modes de pilotage de Framatome

Plusieurs systèmes de commande du cœur, communément appelées modes de pilotage dans l'industrie nucléaire, ont été développés, validés, voire installés sur site par Framatome. La principale mission d'un mode de pilotage est de veiller à ce que la température moyenne du circuit primaire et le déséquilibre axial de puissance du cœur restent constamment à l'intérieur du domaine spécifié par les conditions limites d'exploitation, et ce, pour l'ensemble des transitoires indiqués dans le cahier des charges de l'exploitant. En complément de cette exigence, les modes de pilotage les plus avancés permettent à la turbine de revenir rapidement et sans préavis à un niveau de puissance souhaité, choisi après la baisse de charge par l'opérateur, sans que les variables d'intérêt ne sortent du domaine de fonctionnement autorisé. Trois modes de pilotage sont actuellement utilisés en France ou à l'étranger :

1) Le **mode A** est le mode de pilotage le plus simple et le plus répandu à travers le monde. Il s'agit également du mode de pilotage le moins flexible, celui-ci ayant été conçu à l'origine par Westinghouse pour les centrales fonctionnant en base, c'est-à-dire pour les centrales dont la turbine reste à puissance nominale afin d'assurer la base de la production d'électricité. En France, le mode A est utilisé sur les 4 réacteurs 900 MWe du palier CP0 et sur les 4 réacteurs 1450 MWe du palier N4.
Le cœur d'un réacteur piloté en mode A est équipé de 4 groupes de grappes lourds, constitués chacun de deux quadruplets de grappes noires, reliés entre eux par une contrainte de

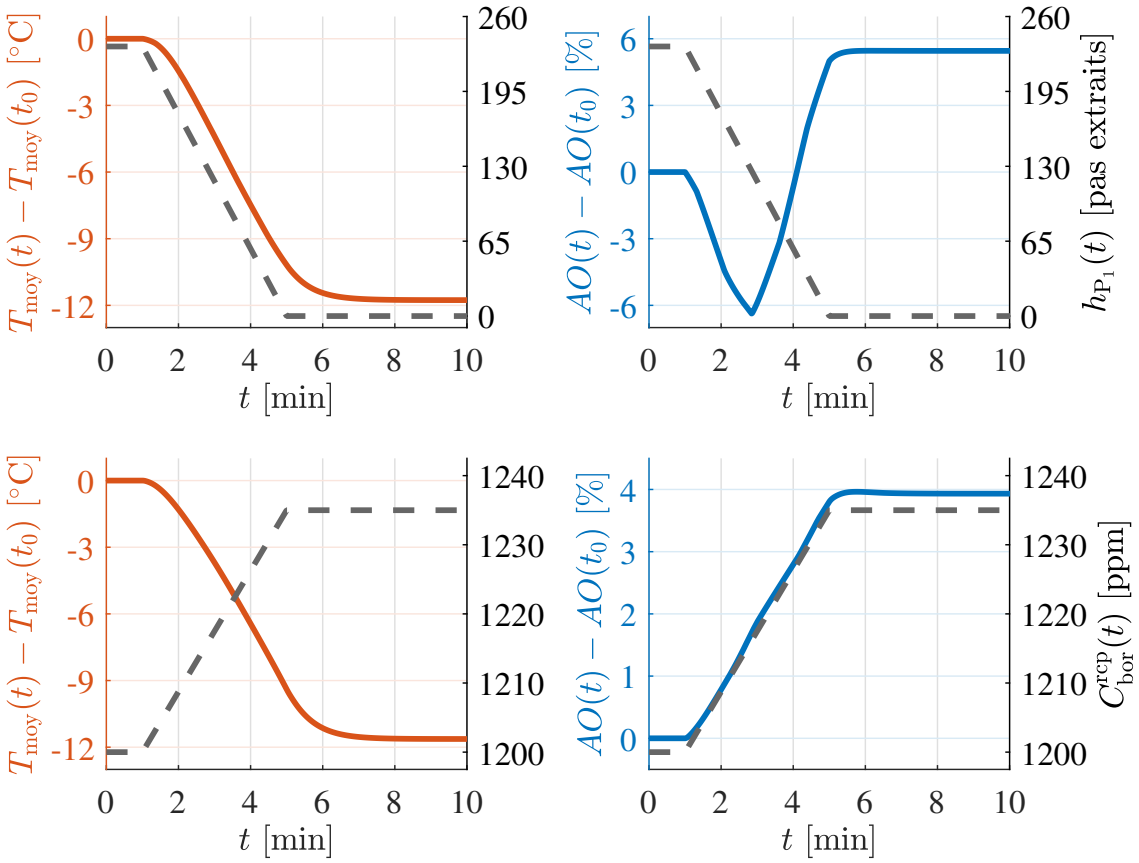

FIGURE 2.16 – Réponses temporelles de la température moyenne et du déséquilibre axial de puissance à une insertion en rampe du groupe de grappes P_1, sans recouvrement avec les autres groupes, et à une augmentation en rampe de la concentration en bore (l'anti-réactivité apportée par les deux actionneurs étant presque équivalente).

recouvrement fixe. Ainsi, contrairement aux modes de pilotages plus flexibles, le mode A ne possède qu'un seul actionneur de grappes, dédié exclusivement à la régulation de température moyenne. Pendant les variations de charge, les changements de température moyenne induits par les contre-réactions neutroniques déclenchent automatiquement l'insertion ou l'extraction des grappes de commandes dans le cœur. Celles-ci continuent de bouger après les variations de charge, en raison des changements de température moyenne induits par l'évolution du xénon. Ces mouvements de grappes perturbent très fortement la distribution axiale de puissance du cœur, et doivent être atténués manuellement par l'opérateur en ajustant la concentration en bore du circuit primaire. Plus précisément, l'opérateur doit compenser les variations de température moyenne par des injections bien planifiées d'eau borée (des « borications ») ou d'eau claire (des « dilutions ») pour éviter, autant

que possible, que les grappes ne bougent. L'obligation de recourir aux dilutions et aux boricationspour réguler le déséquilibre axial de puissance limite la vitesse des variations de charge à environ $2\,\%\mathrm{PN\,min^{-1}}$ en début de cycle, car les débits utilisés pour injecter les solutions d'eau borée et d'eau claire ne sont pas illimités. De plus, comme la concentration en bore du circuit primaire diminue progressivement au cours du cycle d'irradiation, les dilutions deviennent de moins en moins efficaces au fil du temps, ce qui limite la vitesse des reprises de charge à seulement $0.2\,\%\mathrm{PN\,min^{-1}}$ à partir de 80 % d'avancement du cycle. Les vitesses des variations de charge sont encore bien plus faibles en pratique, car il peut être délicat pour l'opérateur de réguler correctement le déséquilibre axial de puissance du fait de l'inertie de l'actionneur de bore. Par ailleurs, l'opérateur doit être suffisamment expérimenté pour contrer l'évolution du xénon après chaque variation de charge.

2) Le **mode G** a été conçu afin de permettre aux centrales nucléaires françaises de fonctionner efficacement en suivi de charge. De ce fait, une nouvelle exigence est de garantir que la turbine soit toujours en mesure de revenir rapidement à son niveau de puissance nominal après une baisse de charge. En France, le mode G est utilisé sur les 28 réacteurs 900 MWe du palier CPY, ainsi que sur les 20 réacteurs 1300 MWe des paliers P4 et P′4.
   La principale différence entre le mode A et le mode G est que les groupes de grappes sont dorénavant séparés en deux blocs fonctionnels distincts, capables de se déplacer indépendamment l'un de l'autre. Le premier bloc fonctionnel est composé d'un seul groupe lourd, constitué de deux quadruplets de grappes noires, dont la position est calculée par la régulation de température moyenne. Le second bloc fonctionnel est composé de deux groupes légers, constitués respectivement d'un et de deux quadruplets de grappes grises, et de deux groupes lourds, constitués chacun de deux quadruplets de grappes noires. La position de ces 4 groupes peut être déterminée de façon univoque à partir d'une unique position cumulée, ceux-ci étant liés par une contrainte de recouvrement fixe minimisant l'influence des mouvements grappes sur la distribution axiale de puissance. La position cumulée des groupes du second bloc fonctionnel suit alors automatiquement un programme d'insertion de référence, qui évolue en fonction de la puissance relative de la turbine. Ce programme d'insertion de référence est calibré périodiquement sur site, de telle sorte que l'extraction des groupes, dits de compensation de puissance (GCP), permette à elle seule de compenser l'impact des contre-réactions neutroniques sur la température moyenne lorsque la turbine retourne à son niveau de puissance nominal. La flexibilité des réacteurs s'est donc nettement améliorée en mode G, la vitesse des variations de charge pouvant désormais atteindre $5\,\%\mathrm{PN\,min^{-1}}$ du début à la fin du cycle. Toutefois, étant donné que le calibrage du programme d'insertion n'est jamais parfait, le groupe lourd du premier bloc fonctionnel peut parfois être sollicité par la régulation de température moyenne lors des reprises de charge. Dans ce cas, l'opérateur doit ajuster manuellement la concentration en

bore du circuit primaire, comme en mode A, pour éviter que les grappes ne perturbent la distribution axiale de puissance. L'opérateur doit également contrer l'évolution du xénon après les variations de charge, par le biais de dilutions ou de borications appropriées.

3) Le **mode T** est le dernier mode de pilotage ayant été conçu par Framatome, à destination des réacteurs 1650 MWe de type EPR. À la différence du mode A et du mode G, il s'agit d'un mode de pilotage entièrement automatisé, capable de réguler simultanément la température moyenne et le déséquilibre axial de puissance sans aucune intervention de l'opérateur. Les vitesses de variation de charge atteignables en mode T sont identiques à celles du mode G, à savoir $5\,\%\mathrm{PN\,min^{-1}}$ en début et en fin de cycle, bien que la puissance nominale de la turbine soit plus importante qu'auparavant. Pour y parvenir, les groupes de grappes sont de nouveau séparés en deux blocs fonctionnels indépendants, notés $\mathrm{P_{bank}}$ (P comme *Power*) et $\mathrm{H_{bank}}$ (H comme *Heavy*), le premier étant assigné à la régulation de température moyenne et le second à celle du déséquilibre axial de puissance. Une des particularités du mode T est que la composition de ces blocs fonctionnels change à mesure que les groupes s'insèrent dans le cœur.
   En effet, le cœur d'un réacteur piloté en mode T est équipé de 5 groupes de grappes $\mathrm{P_1}$, $\mathrm{P_2}$, $\mathrm{P_3}$, $\mathrm{P_4}$, $\mathrm{P_5}$, numérotés par ordre d'efficacité croissante. Pour le réacteur EPR de la centrale de Flamanville (FA3), les groupes $\mathrm{P_1}$ et $\mathrm{P_2}$ sont constitués respectivement d'un quadruplet de grappes grises et d'un quadruplet de grappes noires, le groupe $\mathrm{P_3}$ d'un quadruplet de grappes noires et d'un quadruplet de grappes grises, le groupe $\mathrm{P_4}$ de deux quadruplets de grappes noires, et le groupe $\mathrm{P_5}$ de trois quadruplets de grappes noires. L'affectation des groupes $\mathrm{P_1}$ à $\mathrm{P_5}$ aux blocs fonctionnels $\mathrm{P_{bank}}$ et $\mathrm{H_{bank}}$ obéit aux règles suivantes :

   a) Tout d'abord, lorsque les grappes se trouvent dans la partie haute du cœur, $\mathrm{P_{bank}}$ est composé uniquement du groupe $\mathrm{P_1}$, et $\mathrm{H_{bank}}$ des autres groupes $\mathrm{P_2}$ à $\mathrm{P_5}$ restant.
   b) Puis, lorsque l'écartement maximal entre les groupes $\mathrm{P_1}$ et $\mathrm{P_2}$ est atteint, le groupe $\mathrm{P_2}$ se désolidarise de $\mathrm{H_{bank}}$ pour rejoindre $\mathrm{P_{bank}}$.
   c) De même, les groupes $\mathrm{P_3}$ et $\mathrm{P_4}$ passeront à leur tour de $\mathrm{H_{bank}}$ à $\mathrm{P_{bank}}$ lorsque l'écartement maximal entre les groupes $\mathrm{P_2}$ et $\mathrm{P_3}$, puis $\mathrm{P_3}$ et $\mathrm{P_4}$, sera atteint.
   d) Enfin, lorsque les grappes se trouvent dans la partie basse du cœur, $\mathrm{P_{bank}}$ est composé des groupes $\mathrm{P_1}$ à $\mathrm{P_4}$, et $\mathrm{H_{bank}}$ du seul groupe $\mathrm{P_5}$ restant.

   L'écartement maximal autorisé entre deux groupes de grappes consécutifs $\mathrm{P_j}$ et $\mathrm{P_{j+1}}$ est légèrement inférieur à la moitié de la hauteur du cœur (205 sur 420 pas), ce qui permet de réduire efficacement les perturbations causées par le bloc $\mathrm{P_{bank}}$ sur la distribution axiale de puissance, tout en conservant une efficacité suffisante pour réguler la température moyenne. Après une baisse de charge, le mode T permet cette fois-ci à la turbine de revenir rapidement et sans préavis à n'importe quel niveau de puissance intermédiaire, compris entre son niveau de puissance actuel et son niveau de puissance nominal. Pour ce faire,

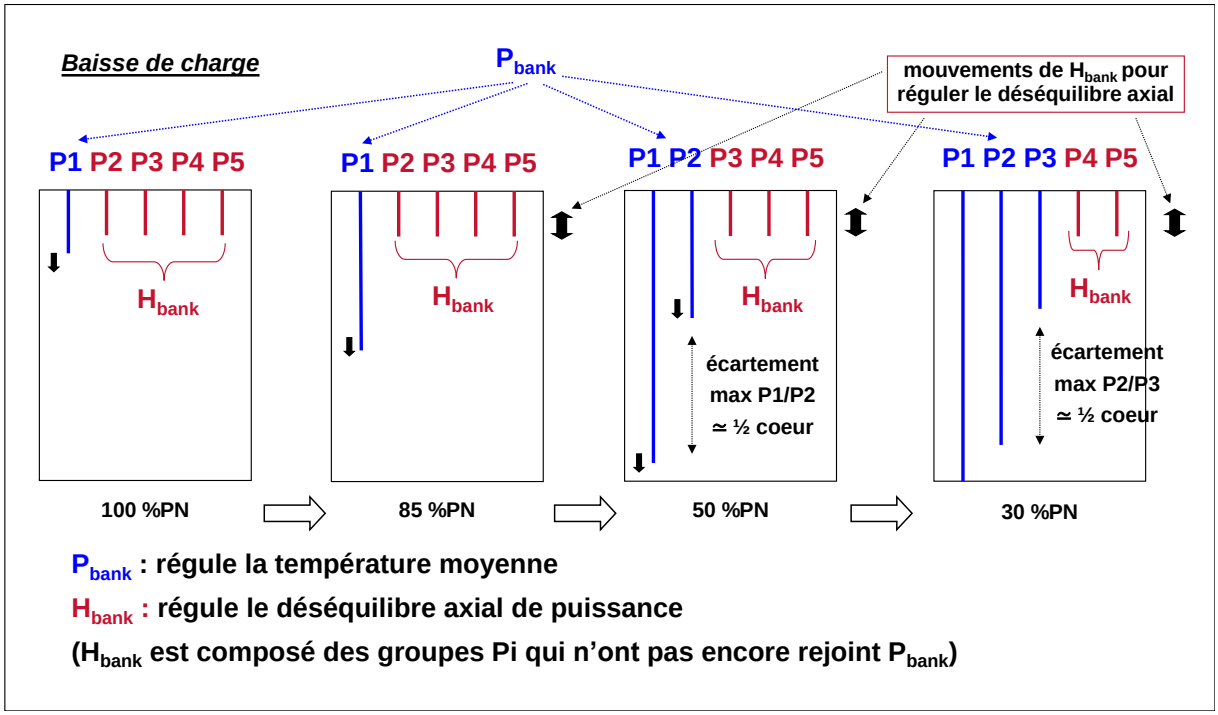

FIGURE 2.17 – Illustration du changement de composition des blocs fonctionnels P_{bank} et H_{bank} lors d'une baisse de charge (figure adaptée de [74]).

la position cumulée des groupes du bloc P_{bank} suit, à une certaine tolérance près, un programme d'insertion de référence semblable à celui des GCP du mode G, qui évolue en fonction de l'écart entre la puissance relative du cœur et la puissance de consigne à laquelle l'opérateur souhaite pouvoir retourner. Ce programme d'insertion de référence est régulièrement mis à jour en se basant sur des données déterminées en bureau d'étude (et non plus expérimentalement, comme en mode G) pour tenir compte de l'épuisement du combustible nucléaire. Si la puissance de consigne sélectionnée par l'opérateur est égale à la puissance nominale de la turbine, l'évolution du xénon sera contrée par des dilutions ou des borications automatiques afin de maintenir les groupes du bloc P_{bank} insérés dans le cœur. Cette première stratégie est optimale en termes de flexibilité, car la turbine pourra toujours retourner à son niveau de puissance nominal au moment attendu. En revanche, si la puissance de consigne est égale à la puissance courante de la turbine, alors l'évolution du xénon sera contrée par l'insertion ou l'extraction automatique des groupes du bloc P_{bank}. Cette seconde stratégie est optimale en termes d'économie d'effluents, car les volumes d'eau borée ou d'eau claire injectés dans le circuit primaire pendant le palier bas seront nuls. Une stratégie intermédiaire sera adoptée pour contrer l'évolution du xénon si la puissance de consigne est située entre la puissance courante et la puissance nominale de la turbine.

TABLE 2.3 – Résumé des caractéristiques des modes de pilotage de Framatome.

	Mode A	**Mode G**	**Mode T**
Groupes de grappes	4 groupes nommés D, C, B, A tous constitués de 2x4 grappes noires.	5 groupes nommés R, G1, G2, N1, N2 chacun constitués de : 2x4 grappes noires, 1x4 grappes grises, 2x4 grappes grises, 2x4 grappes noires, 2x4 grappes noires.	5 groupes nommés P1, P2, P3, P4, P5 chacun constitués de (FA3) : 1x4 grappes grises, 1x4 grappes noires, 1x4 grappes grises et 1x4 grappes noires, 2x4 grappes noires, 3x4 grappes noires.
Blocs fonctionnels	1 unique bloc composé des 4 groupes D, C, B, A.	1 bloc composé du seul groupe R. 1 bloc composé des 4 groupes G1, G2, N1, N2 de compensation de puissance.	2 blocs P_{bank} et H_{bank} dont la composition change avec l'insertion des groupes : P_{bank} est composé du groupe P1 et de ceux n'ayant pas encore rejoint H_{bank}. H_{bank} est composé du groupe P5 et de ceux ayant quitté P_{bank}.
Contrôle de l'ACT	Insertion/extraction automatique des groupes D, C, B, A.	Insertion/extraction automatique du groupe R.	Insertion/extraction automatique des groupes du bloc P_{bank}, ou des groupes du bloc H_{bank} quand ceux du bloc P_{bank} doivent rester insérés pour pouvoir remonter sans préavis en puissance.

Contrôle de l'AO	Dilution/borication manuelle afin de modifier indirectement la position des groupes en faisant varier l'ACT.	Dilution/borication manuelle afin de modifier indirectement la position du groupe R en faisant varier l'ACT.	Insertion/extraction automatique des groupes du bloc H_{bank}, ou dilution/borication automatique quand les groupes du bloc P_{bank} sont en haut du cœur ou que ceux du bloc H_{bank} doivent réguler l'ACT.
Retour instantané en puissance	Impossible, car les mouvements des grappes noires perturbent trop l'AO.	Possibilité de retourner sans préavis au niveau de puissance nominal grâce à l'extraction automatique des GCP.	Possibilité de retourner sans préavis au niveau de puissance de consigne en contrôlant le niveau d'insertion des groupes du bloc P_{bank}.
Vitesse des variations de charge	Lentes, car dilutions et borications obligatoires pour contrôler l'AO lors des variations de charge.	Rapides, car l'AO peut être contrôlé par des mouvements compensés du groupe R et des GCP.	Rapides, car l'AO peut être contrôlé par des mouvements compensés des groupes des blocs P_{bank} et H_{bank}.

2 Modélisation du réacteur à eau sous pression

Les systèmes de commande du cœur présentés dans cette thèse ont été conçus puis réglés en s'appuyant sur un modèle de réacteur multi-mailles non-linéaire. L'avantage de ce nouveau modèle, comparé au modèle de réacteur point non-linéaire élaboré dans la thèse précédente [70], est qu'il permet de représenter le comportement de la distribution axiale de puissance du cœur de façon beaucoup plus directe et intuitive qu'auparavant. En effet, puisque le cœur du réacteur de l'ancien modèle se réduisait à un point, son déséquilibre axial de puissance devait être reconstruit de façon empirique à partir d'autres variables d'état :

$$\begin{aligned}\frac{\mathrm{d}AO(t)}{\mathrm{d}t} = \frac{1}{\tau_{\mathrm{AO}}}\Big(&-AO(t) + K_{\mathrm{AO}}^{\mathrm{in}}\Big(T_{\mathrm{in}}(t)\Big)T_{\mathrm{in}}(t) + K_{\mathrm{AO}}^{\mathrm{out}}\Big(T_{\mathrm{out}}(t)\Big)T_{\mathrm{out}}(t) \\ &+ K_{\mathrm{AO}}^{\mathrm{P}}\Big(P_{\mathrm{bank}}(t)\Big)P_{\mathrm{bank}}(t) + K_{\mathrm{AO}}^{\mathrm{H}}\Big(H_{\mathrm{bank}}(t)\Big)H_{\mathrm{bank}}(t)\Big)\,,\end{aligned} \tag{2.22}$$

Conclusion et perspectives

L'objectif de cette thèse était d'améliorer la flexibilité des réacteurs nucléaires à eau sous pression afin de répondre aux futurs besoin du réseau électrique. La principale contribution a été de concevoir un nouveau système de commande du cœur en combinant un algorithme de commande prédictive non-linéaire avec la régulation de température déjà existante sur le mode T, le dernier mode de pilotage élaboré par Framatome. Pour ce faire, un modèle de réacteur multi-mailles de complexité juste et nécessaire a été développé pour servir de support à la synthèse des lois de commande. Ce modèle, d'ordre relativement élevé, reste néanmoins compliqué à contrôler, celui-ci étant composé de nombreuses équations différentielles non-linéaires, aux dynamiques très disparates et sujettes aux phénomènes de retard. De ce fait, la méthodologie de conception des algorithmes de commande prédictive a dû être étudiée en profondeur pour lever les verrous techniques rencontrés. Les différentes étapes de cette méthodologie ont été clairement identifiées et présentées en détails dans le manuscrit. Plusieurs système de commande ont par ailleurs été conçus avant d'arriver à la solution proposée. Ces travaux intermédiaires ont permis de progressivement prendre en main et de mettre en pratique les outils et méthodes de conception avancées utilisés, et permettent d'envisager différents niveaux d'automatisation. La solution ainsi élaborée a ensuite été comparée en simulation au mode T. Les résultats obtenus montrent que le nouveau système de commande est capable de contrôler le cœur du réacteur dans des situations très complexes, qui mettent le mode T en difficulté. Celui-ci permet notamment de conserver une flexibilité importante en fin de cycle, lorsque la marge de manœuvre du réacteur est fortement limitée, ce qui n'est pas le cas du mode T. La démarche conceptuelle adoptée est suffisamment lisible et systématique pour que le nouveau système de commande puisse être configuré rapidement en cas de modification du cahier des charges.

Cependant, plusieurs aspects peuvent être encore approfondis avant de parvenir à une solution réellement industrialisable. Tout d'abord, un observateur d'état pourrait être conçu pour reconstruire les variables d'état non mesurables du modèle. Des premiers travaux sur l'estimation à horizon mouvant ont été obtenus en 2022 dans une autre thèse en cours de réalisation à Framatome [79]. Puis, une action corrective pourrait compléter le schéma de compensation du délai de transmission incorporé à l'algorithme de commande prédictive afin d'améliorer sa robustesse vis-à-vis du délai de calcul. Il pourrait également être judicieux de mettre au point une stratégie de repli si jamais le solveur d'optimisation ne réussissait pas à trouver de solution pertinente dans le temps imparti. Enfin, un étage prédictif long terme pourrait être ajouté au contrôleur hiérarchisé afin de s'assurer que le transitoire de suivi de charge demandé à la turbine

peut être réalisé à l'avance sans sortir des conditions limites d'exploitation. Ce dernier aspect est particulièrement important en fin de cycle, puisque la marge de manœuvre dont dispose le réacteur est infime.

BIBLIOGRAPHIE

[1] P. Morilhat, S. Feutry, C. Lemaitre et J. M. Favennec, « Nuclear power plant flexibility at EDF », *atw International Journal for Nuclear Power*, t. 64, *3*, p. 131-140, 2019.

[2] K. Kosowski et F. Diercks, « Quo vadis, grid stability ? Challenges increase as generation portfolio changes », *atw International Journal for Nuclear Power*, t. 66, *2*, p. 16-26, 2021.

[3] IAEA, *Non-baseload Operation in Nuclear Power Plants : Load Following and Frequency Control Modes of Flexible Operation* (Nuclear Energy Series NP-T-3.23). IAEA Viena, 2018.

[4] G. Dupré, « Contrôle d'une unité de production d'énergie décentralisée raccordée à un réseau de distribution déséquilibré », mém. de mast., Ecole Polytechnique, Montreal (Canada), 2019.

[5] A. Lokhov, « Technical and economic aspects of load following with nuclear power plants », *NEA, OECD, Paris, France*, t. 2, 2011.

[6] V. Drouet, « Optimisation multi-objectifs du pilotage des réacteurs nucléaires à eau sous pression en suivi de charge dans le contexte de la transition énergétique à l'aide d'algorithmes évolutionnaires », thèse de doct., Université Paris-Saclay, 2020.

[7] A. Kumar et P. Daoutidis, *Control of Nonlinear Differential Algebraic Equation Systems with Applications to Chemical Processes*. CRC Press, 1999, t. 397.

[8] J. Sjöberg, « Optimal control and model reduction of nonlinear DAE models », thèse de doct., Institutionen för systemteknik, 2008.

[9] R. Quirynen, « Numerical simulation methods for embedded optimization », thèse de doct., 2017.

[10] L. F. Shampine, M. W. Reichelt et J. A. Kierzenka, « Solving index-1 DAEs in MATLAB and Simulink », *SIAM review*, t. 41, *3*, p. 538-552, 1999.

[11] A. C. Hindmarsh, P. N. Brown, K. E. Grant et al., « SUNDIALS : Suite of nonlinear and differential/algebraic equation solvers », *ACM Transactions on Mathematical Software (TOMS)*, t. 31, *3*, p. 363-396, 2005.

[12] F. Hoppensteadt, « On systems of ordinary differential equations with several parameters multiplying the derivatives », *Journal of Differential Equations*, t. 5, *1*, p. 106-116, 1969.

[13] F. Hoppensteadt, « Properties of solutions of ordinary differential equations with small parameters », *Communications on Pure and Applied Mathematics*, t. 24, *6*, p. 807-840, 1971.

[14] G. Peponides, P. Kokotovic et J. Chow, « Singular perturbations and time scales in nonlinear models of power systems », *IEEE Transactions on Circuits and systems*, t. 29, *11*, p. 758-767, 1982.

[15] P. V. Kokotović, « Applications of singular perturbation techniques to control problems », *SIAM review*, t. 26, *4*, p. 501-550, 1984.

[16] P. Kokotović, H. K. Khalil et J. O'reilly, *Singular Perturbation Methods in Control : Analysis and Design*. SIAM, 1999.

[17] H. K. Khalil, *Nonlinear systems*, 3e éd. Prentice-Hall, 2002.

[18] J. T. Betts, *Practical Methods for Optimal Control and Estimation Using Nonlinear Programming*. SIAM, 2010.

[19] J. B. Rawlings, D. Q. Mayne et M. Diehl, *Model Predictive Control : Theory, Computation, and Design*, 2e éd. Nob Hill Publishing Madison, WI, 2019.

[20] J. Nocedal et S. J. Wright, *Numerical optimization*, 2e éd. Springer, 2006.

[21] M. Diehl, H. J. Ferreau et N. Haverbeke, « Efficient numerical methods for nonlinear MPC and moving horizon estimation », in *Nonlinear model predictive control*, Springer, 2009, p. 391-417.

[22] A. V. Rao, « A survey of numerical methods for optimal control », *Advances in the Astronautical Sciences*, t. 135, *1*, p. 497-528, 2009.

[23] C. Kirches, L. Wirsching, S. Sager et H. G. Bock, « Efficient numerics for nonlinear model predictive control », in *Recent Advances in Optimization and its Applications in Engineering*, Springer, 2010, p. 339-357.

[24] L. T. Biegler, *Nonlinear Programming : Concepts, Algorithms, and Applications to Chemical Processes*. SIAM, 2010.

[25] D. Q. Mayne, J. B. Rawlings, C. V. Rao et P. O. Scokaert, « Constrained model predictive control : Stability and optimality », *Automatica*, t. 36, *6*, p. 789-814, 2000.

[26] D. Q. Mayne, « Model predictive control : Recent developments and future promise », *Automatica*, t. 50, *12*, p. 2967-2986, 2014.

[27] M. Schwenzer, M. Ay, T. Bergs et D. Abel, « Review on model predictive control : An engineering perspective », *The International Journal of Advanced Manufacturing Technology*, t. 117, *5*, p. 1327-1349, 2021.

[28] F. Blanchini, « Set invariance in control », *Automatica*, t. 35, *11*, p. 1747-1767, 1999.

[29] E. C. Kerrigan et J. M. Maciejowski, « Invariant sets for constrained nonlinear discrete-time systems with application to feasibility in model predictive control », in *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No. 00CH37187)*, IEEE, t. 5, 2000, p. 4951-4956.

[30] M. Reble, « Model predictive control for nonlinear continuous-time systems with and without time-delays », thèse de doct., 2013.

[31] M. S. Darup et M. Cannon, « A missing link between nonlinear MPC schemes with guaranteed stability », in *2015 54th IEEE Conference on Decision and Control (CDC)*, IEEE, 2015, p. 4977-4983.

[32] L. Grüne et J. Pannek, *Nonlinear Model Predictive Control : Theory and Algorithms*, 2e éd. Springer, 2017.

[33] H. Chen et F. Allgöwer, « A quasi-infinite horizon nonlinear model predictive control scheme with guaranteed stability », *Automatica*, t. 34, *10*, p. 1205-1217, 1998.

[34] F. A. Fontes, « A general framework to design stabilizing nonlinear model predictive controllers », *Systems & Control Letters*, t. 42, *2*, p. 127-143, 2001.

[35] R. Findeisen, T. Raff et F. Allgöwer, « Sampled-data nonlinear model predictive control for constrained continuous time systems », in *Advanced strategies in control systems with input and output constraints*, Springer, 2007, p. 207-235.

[36] D. Mayne, « An apologia for stabilising terminal conditions in model predictive control », *International Journal of Control*, t. 86, *11*, p. 2090-2095, 2013.

[37] C. Chen et L. Shaw, « On receding horizon feedback control », *Automatica*, t. 18, *3*, p. 349-352, 1982.

[38] S. S. Keerthi et E. G. Gilbert, « Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems : Stability and moving-horizon approximations », *Journal of optimization theory and applications*, t. 57, *2*, p. 265-293, 1988.

[39] H. Michalska et D. Q. Mayne, « Receding horizon control of nonlinear systems without differentiability of the optimal value function », *Systems & control letters*, t. 16, *2*, p. 123-130, 1991.

[40] S. L. de Oliveira Kothare et M. Morari, « Contractive model predictive control for constrained nonlinear systems », *IEEE Transactions on Automatic Control*, t. 45, *6*, p. 1053-1071, 2000.

[41] M. Alamir, « Contraction-based nonlinear model predictive control formulation without stability-related terminal constraints », *Automatica*, t. 75, p. 288-292, 2017.

[42] L. Grüne, « NMPC without terminal constraints », *IFAC Proceedings Volumes*, t. 45, *17*, p. 1-13, 2012.

[43] K. Worthmann, M. Reble, L. Grüne et F. Allgöwer, « The role of sampling for stability and performance in unconstrained nonlinear model predictive control », *SIAM Journal on Control and Optimization*, t. 52, *1*, p. 581-605, 2014.

[44] A. Boccia, L. Grüne et K. Worthmann, « Stability and feasibility of state constrained MPC without stabilizing terminal constraints », *Systems & control letters*, t. 72, p. 14-21, 2014.

[45] W. Esterhuizen, K. Worthmann et S. Streif, « Recursive feasibility of continuous-time model predictive control without stabilising constraints », *IEEE Control Systems Letters*, t. 5, *1*, p. 265-270, 2020.

[46] W.-H. Chen, D. J. Ballance et J. O'Reilly, « Model predictive control of nonlinear systems : Computational burden and stability », *IEE Proceedings-Control Theory and Applications*, t. 147, *4*, p. 387-394, 2000.

[47] R. Findeisen et F. Allgöwer, « Computational delay in nonlinear model predictive control », *IFAC Proceedings Volumes*, t. 37, *1*, p. 427-432, 2004.

[48] R. Findeisen, « Nonlinear model predictive control : a sampled data feedback perspective », thèse de doct., 2005.

[49] P. Varutti et R. Findeisen, « Compensating network delays and information loss by predictive control methods », in *2009 European Control Conference (ECC)*, IEEE, 2009, p. 1722-1727.

[50] L. Grüne, J. Pannek et K. Worthmann, « A prediction based control scheme for networked systems with delays and packet dropouts », in *Proceedings of the 48th IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, IEEE, 2009, p. 537-542.

[51] R. Findeisen, L. Grüne, J. Pannek et P. Varutti, « Robustness of prediction based delay compensation for nonlinear systems », *IFAC Proceedings Volumes*, t. 44, *1*, p. 203-208, 2011.

[52] V. M. Zavala et L. T. Biegler, « The advanced-step NMPC controller : Optimality, stability and robustness », *Automatica*, t. 45, *1*, p. 86-93, 2009.

[53] X. Yang et L. T. Biegler, « Advanced-multi-step nonlinear model predictive control », *Journal of process control*, t. 23, *8*, p. 1116-1128, 2013.

[54] L. Biegler, X. Yang et G. Fischer, « Advances in sensitivity-based nonlinear model predictive control and dynamic real-time optimization », *Journal of Process Control*, t. 30, p. 104-116, 2015.

[55] M. Diehl, H. G. Bock, J. P. Schlöder, R. Findeisen, Z. Nagy et F. Allgöwer, « Real-time optimization and nonlinear model predictive control of processes governed by differential-algebraic equations », *Journal of Process Control*, t. 12, *4*, p. 577-585, 2002.

[56] A. Nurkanović, A. Zanelli, S. Albrecht et M. Diehl, « The advanced step real time iteration for NMPC », in *2019 IEEE 58th Conference on Decision and Control (CDC)*, IEEE, 2019, p. 5298-5305.

[57] S. Gros, M. Zanon, R. Quirynen, A. Bemporad et M. Diehl, « From linear to nonlinear MPC : bridging the gap via the real-time iteration », *International Journal of Control*, t. 93, *1*, p. 62-80, 2020.

[58] E. Hairer, G. Wanner et S. P. Nørsett, *Solving Ordinary Differential Equations I : Nonstiff Problems*, 2e éd. Springer Berlin Heidelberg New York, 1996.

[59] E. Hairer et G. Wanner, *Solving Ordinary Differential Equations II : Stiff and Differential-Algebraic Problems*, 2e éd. Springer Berlin Heidelberg New York, 1993.

[60] C. Brezinski, « Méthodes numériques de base : Algèbre numérique », *Techniques de l'Ingénieur*, 2006.

[61] E. Hairer et G. Wanner, « Intégration numérique des équations différentielles raides », *Techniques de l'Ingénieur*, 2007.

[62] W. Marquis-Favre, « Simulation des mécanismes : Résolution des équations dans les logiciels », *Techniques de l'Ingénieur*, 2007.

[63] L. C. Young, « Orthogonal collocation revisited », *Computer Methods in Applied Mechanics and Engineering*, t. 345, p. 1033-1076, 2019.

[64] G. Sánchez, M. Murillo, L. Genzelis, N. Deniz et L. Giovanini, « MPC for nonlinear systems : A comparative review of discretization methods », in *2017 XVII Workshop on Information Processing and Control (RPIC)*, IEEE, 2017, p. 1-6.

[65] S. Boyd, S. P. Boyd et L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[66] M. S. Bazaraa, H. D. Sherali et C. M. Shetty, *Nonlinear programming : theory and algorithms*, 3e éd. John Wiley & Sons, 2013.

[67] P. Coppolani, N. Hassenboelher, J. Joseph, J.-F. Petetrot, J.-P. Py et J.-S. Zampa, *La chaudière des réacteurs à eau sous pression*. EDP Sciences, 2004.

[68] P. Reuss, *Précis de neutronique*. EDP sciences, 2012.

[69] M. Coste-Delclaux, C. Diop, A. Nicolas et B. Bonin, *Neutronique* (E-den, Une monographie de la Direction de l'énergie nucléaire). CEA Saclay ; Groupe Moniteur, 2013.

[70] L. 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., Institut Mines-Télécom Atlantique, 2018.

[71] P. Reuss, « Bases de neutronique : Migration des neutrons », *Techniques de l'Ingénieur*, 2005.

[72] P. Reuss, « Bases de neutronique : Physique et calcul des réacteurs », *Techniques de l'Ingénieur*, 2006.

[73] J.-L. Mourlevat, « Évolution des modes de pilotage », *Revue générale nucléaire*, *3*, p. 29-36, 2007.

[74] A. Grossetete, « Le pilotage de l'EPR : mode T », *Revue générale nucléaire*, *3*, p. 37-41, 2007.

[75] C. Herer, « Thermohydraulique des réacteurs à eau sous pression », *Techniques de l'Ingénieur*, 2021.

[76] RTE, *Documentation technique de référence*, 2009.

[77] *Nuclear Power for Everybody*, `https://www.nuclear-power.com/`, Accessed : 2023-05-23.

[78] P. Girieud, L. Daudin, C. Garat, P. Marotte et S. Tarlé, « Science Version 2 : the most recent capabilities of the Framatome 3-D nuclear code package », in *Proceedings of the 9th International Conference on Nuclear Engineering (ICONE)*, 2001.

[79] L. Gruss, P. Chevrel, M. Yagoubi, M. Thieffry et A. Grossetête, « Moving horizon estimation of xenon in pressurized water nuclear reactors using variable-step integration », in *2023 European Control Conference (ECC)*, IEEE, 2023, p. 1-6.

[80] F. Zhang et M. Yeddanapudi, « Modeling and simulation of time-varying delays », in *Proceedings of the 2012 symposium on theory of modeling and simulation-DEVS Integrative M&S Symposium*, 2012, p. 1-8.

[81] G. Dupré, A. Grossetête, P. Chevrel et M. Yagoubi, « Enhanced Flexibility of PWRs (mode A) Using an Efficient NMPC-Based Boration/Dilution System », in *2021 European Control Conference (ECC)*, IEEE, 2021, p. 1092-1098.

[82] R. Cagienard, P. Grieder, E. C. Kerrigan et M. Morari, « Move blocking strategies in receding horizon control », *Journal of Process Control*, t. 17, *6*, p. 563-570, 2007.

[83] R. Gondhalekar et J.-i. Imura, « Least-restrictive move-blocking model predictive control », *Automatica*, t. 46, *7*, p. 1234-1240, 2010.

[84] R. C. Shekhar et C. Manzie, « Optimal move blocking strategies for model predictive control », *Automatica*, t. 61, p. 27-34, 2015.

[85] S. H. Son, T. H. Oh, J. W. Kim et J. M. Lee, « Move blocked model predictive control with improved optimality using semi-explicit approach for applying time-varying blocking structure », *Journal of Process Control*, t. 92, p. 50-61, 2020.

[86] R. H. Byrd, J. C. Gilbert et J. Nocedal, « A trust region method based on interior point techniques for nonlinear programming », *Mathematical programming*, t. 89, p. 149-185, 2000.

[87] H. Cappon et B. d. Braquilanges, « SOFIA engineering simulator », in *Annual meeting on nuclear technology*, 2013.

[88] L. Lemazurier, M. Yagoubi, P. Chevrel et A. Grossetête, « Multi-Objective H_2/H_∞ Gain-Scheduled Nuclear Core Control Design », *IFAC-PapersOnLine*, t. 50, *1*, p. 3256-3262, 2017.

[89] L. Lemazurier, P. Chevrel, M. Yagoubi et A. Grossetête, « A Multi-Objective Nuclear Core Control Performing Hot and Cold Leg Temperature Control », in *2018 European Control Conference (ECC)*, IEEE, 2018, p. 3050-3056.

[90] L. Lemazurier, P. Chevrel, A. Grossetête et M. Yagoubi, « An alternative to standard nuclear core control using a multi-objective approach », *Control Engineering Practice*, t. 83, p. 98-107, 2019.

[91] S. J. Qin et T. A. Badgwell, « A survey of industrial model predictive control technology », *Control engineering practice*, t. 11, *7*, p. 733-764, 2003.

[92] J. H. Lee, « Model predictive control : Review of the three decades of development », *International Journal of Control, Automation and Systems*, t. 9, p. 415-424, 2011.

[93] D. J. Leith et W. E. Leithead, « Survey of gain-scheduling analysis and design », *International journal of control*, t. 73, *11*, p. 1001-1025, 2000.

[94] W. J. Rugh et J. S. Shamma, « Research on gain scheduling », *Automatica*, t. 36, *10*, p. 1401-1425, 2000.

[95] G. Dupré, P. Chevrel, M. Yagoubi et A. Grossetête, « Design and comparison of two advanced core control systems for flexible operation of pressurized water reactors », *Control Engineering Practice*, t. 123, p. 105 170, 2022.

[96] B. Friedland et S. W. Director, *Control Systems Design : An Introduction to State-Space Methods*. McGraw-Hill Higher Education, 1985.

[97] S. P. Boyd et C. H. Barratt, *Linear Controller Design : Limits of Performance*. Prentice-Hall, Inc., 1991.

[98] M. Green et D. J. N. Limebeer, *Linear Robust Control*. Prentice-Hall, Inc., 1994.

[99] K. Zhou, J. C. Doyle et K. Glover, *Robust and Optimal Control*. Prentice-Hall, Inc., 1996.

[100] P. Apkarian et D. Noll, « Nonsmooth H_∞ synthesis », *IEEE Transactions on Automatic Control*, t. 51, *1*, p. 71-86, 2006.

[101] D. Arzelier, G. Deaconu, S. Gumussoy et D. Henrion, « H2 for HIFOO », in *International Conference on Control and Optimization With Industrial Applications (COIA 2011)*, 2011.

[102] M. S. Sadabadi et D. Peaucelle, « From static output feedback to structured robust static output feedback : A survey », *Annual reviews in control*, t. 42, p. 11-26, 2016.

[103] P. Apkarian, D. Noll et A. Rondepierre, « Mixed H_2/H_∞ Control via Nonsmooth Optimization », *SIAM Journal on Control and Optimization*, t. 47, *3*, p. 1516-1546, 2008.

[104] S. Gumussoy, D. Henrion, M. Millstone et M. L. Overton, « Multiobjective robust control with HIFOO 2.0 », *IFAC Proceedings Volumes*, t. 42, *6*, p. 144-149, 2009.

[105] E. Davison et A. Goldenberg, « Robust control of a general servomechanism problem : The servo compensator », *Automatica*, t. 11, *5*, p. 461-471, 1975.

[106] B. A. Francis et W. M. Wonham, « The internal model principle of control theory », *Automatica*, t. 12, *5*, p. 457-465, 1976.

[107] P. Gahinet et P. Apkarian, « Automated tuning of gain-scheduled control systems », in *52nd IEEE Conference on Decision and Control*, IEEE, 2013, p. 2740-2745.

[108] X. Chen et A. Ray, « On Singular Perturbation of Neutron Point Kinetics in the Dynamic Model of a PWR Nuclear Power Plant », *Sci*, t. 2, *2*, p. 30, 2020.

[109] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings et M. Diehl, « CasADi : a software framework for nonlinear optimization and optimal control », *Mathematical Programming Computation*, t. 11, *1*, p. 1-36, 2019.

[110] A. Wächter et L. T. Biegler, « On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming », *Mathematical programming*, t. 106, *1*, p. 25-57, 2006.

[111] P. O. Scokaert, D. Q. Mayne et J. B. Rawlings, « Suboptimal model predictive control (feasibility implies stability) », *IEEE Transactions on Automatic Control*, t. 44, *3*, p. 648-654, 1999.

[112] K. Graichen et A. Kugi, « Stability and incremental improvement of suboptimal MPC without terminal constraints », *IEEE Transactions on Automatic Control*, t. 55, *11*, p. 2576-2580, 2010.

[113] L. Grüne et J. Pannek, « Analysis of unconstrained NMPC schemes with incomplete optimization », *IFAC Proceedings Volumes*, t. 43, *14*, p. 238-243, 2010.

[114] A. E. Bryson et Y.-C. Ho, *Applied optimal control : optimization, estimation, and control*. Routledge, 1975.

[115] S. Galeani, S. Tarbouriech, M. Turner et L. Zaccarian, « A tutorial on modern anti-windup design », *European Journal of Control*, t. 15, *3-4*, p. 418-440, 2009.

[116] S. Tarbouriech et M. Turner, « Anti-windup design : an overview of some recent advances and open problems », *IET control theory & applications*, t. 3, *1*, p. 1-19, 2009.

[117] J.-M. Biannic et P. Apkarian, « Anti-windup design via nonsmooth multi-objective H_∞ optimization », in *Proceedings of the 2011 American Control Conference*, IEEE, 2011, p. 4457-4462.

[118] S. Tarbouriech et I. Queinnec, « Un tour d'horizon sur les techniques anti-windup pour les systèmes saturés », *Techniques de l'Ingénieur*, 2020.

COLLEGES SCIENCES
BRETAGNE POUR L'INGENIEUR
LOIRE ET LE NUMERIQUE

IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

Titre : Conception de systèmes de contrôle avancé de réacteur PWR flexible

Mot clés : Commande prédictive, Commande à gains séquencés, Physique des réacteurs

Résumé : La plupart des unités de production d'électricité d'origine renouvelable déployées ces dernières années sont par nature intermittentes. En l'absence de solution de stockage à grande échelle, la production et la consommation d'électricité doivent être constamment équilibrées pour garantir la stabilité du réseau. Ce rôle, traditionnellement occupé par les centrales thermiques à flamme, tend de plus en plus à être assuré par les centrales nucléaires. Ainsi, cette thèse vise à améliorer la flexibilité des réacteurs nucléaires à eau sous pression afin de répondre aux futurs besoins du réseau électrique. Pour ce faire, plusieurs systèmes de contrôle du cœur du réacteur ont été conçus en se basant sur des méthodes avancées du domaine de l'automatique, à savoir la commande prédictive et la commande à gains séquencés. Un modèle non-linéaire de réacteur multi-mailles, destiné à la synthèse de lois de commande, a notamment dû être développé. De complexité juste suffisante, il est bien adapté à des fins de prédiction court terme. La solution finalement proposée comporte deux volets : 1) un système temps réel d'aide au pilotage (brevet monde), qui fait désormais partie de l'offre commerciale de Framatome, et 2) une solution de pilotage hiérarchique compatible avec les boucles de régulation de température existantes, dont les performances sont nettement accrues en termes de flexibilité et de respect des contraintes opérationnelles, par rapport aux modes de pilotage actuels. Cette solution s'appuie sur les techniques d'implémentation de commande prédictive non-linéaire les mieux adaptées.

Title: Design of advanced control systems for flexible PWR reactors

Keywords: Model predictive control, Gain-scheduling control, Reactor physics

Abstract: Most renewable electricity generation units deployed in recent years are inherently intermittent. In the absence of large-scale storage solutions, electricity production and consumption must be constantly balanced to ensure grid stability. This role, traditionally played by fossil-fired power plants, is increasingly being filled by nuclear power plants. Hence, this thesis aims at enhancing the flexibility of pressurized water nuclear reactors to meet future grid requirements. To achieve this, several core control systems have been designed based on advanced control methods, namely model predictive control and gain-scheduling control. In particular, a non-linear multi-mesh reactor model, dedicated to the design of control laws, had to be developed. Its complexity is well-suited to short-term predictions. The solution ultimately proposed is twofold: 1) a real-time operator assistance system (world patent), which is now part of Framatome's commercial offer, and 2) a hierarchical control solution compatible with existing temperature control loops, whose performance is significantly enhanced in terms of flexibility and compliance with operational constraints, compared with current core control systems. This solution relies on the most relevant non-linear model predictive control implementation techniques.