



**HAL**  
open science

# Construction de modèles réduits pour le calcul des performances des avions

Manon Bondouy

► **To cite this version:**

Manon Bondouy. Construction de modèles réduits pour le calcul des performances des avions. Ingénierie, finance et science [cs.CE]. Institut de Mathématiques de Toulouse & Institut Universitaire de France, UMR 5219, Université de Toulouse, CNRS, UPS IMT; Airbus, 2016. Français. <NNT : >. <tel-01963127>

**HAL Id: tel-01963127**

**<https://hal.science/tel-01963127v1>**

Submitted on 21 Dec 2018

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

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



HAL Authorization



# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

---

Présentée et soutenue le *08/02/2016* par :

**MANON BONDOUY**

**Construction de modèles réduits pour le calcul des  
performances des avions**

---

---

### JURY

VINCENT BARRA	Professeur	Rapporteur
PIERRE VILLON	Professeur	Rapporteur
RODOLPHE LE RICHE	Directeur de Recherche	Membre du Jury
CHRISTIAN BES	Professeur	Directeur de thèse
SOPHIE JAN	Maître de conférence	Co-directrice de thèse
SERGE LAPORTE	Ingénieur	Membre du Jury
MATTIA PADULO	Ingénieur	Invité

---

**École doctorale et spécialité :**

*MITT : Domaine Mathématiques : Mathématiques appliquées*

**Unité de Recherche :**

*Institut de Mathématiques de Toulouse (UMR 5219 )*

**Directeur(s) de Thèse :**

*Christiane BES et Sophie JAN*

**Rapporteurs :**

*Vincent BARRA et Pierre VILLON*



*À la mémoire de mes grands parents maternels*



# Remerciements

C'est avec grand plaisir que je tiens à remercier tous ceux qui m'ont aidée dans l'aboutissement de ma thèse.

Je souhaite tout d'abord remercier mes trois encadrants de thèse Sophie Jan, Christian Bès et Serge Laporte qui, par leurs compétences professionnelles et leurs qualités humaines, ont permis de réaliser mon travail de recherche dans un cadre idéal. Vous m'avez fait confiance et je garderai de bons souvenirs de tous ces moments d'échanges scientifiques ponctués par les rires. Merci également pour votre grand soutien dans les moments de doute.

Je tiens également à remercier Jérôme Chaptal qui m'a encadrée durant mes deux premières années de thèse, pour son aide précieuse dans la formalisation du contexte industriel. Je le remercie pour tous les précieux conseils qu'il a continué à me donner après son départ.

Mes remerciements vont également aux rapporteurs, Vincent Barra et Pierre Villon ; la version finale de ce mémoire a bénéficié de leur lecture très attentive et de leurs remarques pertinentes. Je remercie également Rodolphe Le Riche pour l'honneur qu'il me fait d'être dans mon jury de thèse. Je tiens finalement à remercier Mattia Padulo pour l'intérêt qu'il a manifesté en participant en qualité de membre invité à ce jury, ainsi que pour le temps qu'il a consacré à ma recherche durant ces trois années.

Je remercie par ailleurs toute l'équipe ex-EGVTO pour leur accueil, tout d'abord en tant que stagiaire puis en tant que doctorante. Je remercie particulièrement Jean-Pierre Demortier pour son intérêt permanent, Mauro Marinelli, pour le temps qu'il m'a accordé et pour son aide précieuse dans l'utilisation de modèles destinés à être embarqués ainsi que Jérémy De-Sousa pour son travail conséquent sur l'étude du modèle de qualité de vol.

Une reconnaissance particulière à l'équipe EGVX pour m'avoir accueillie suite à la réorganisation du département, plus particulièrement à Céline Dejoux qui m'a aidée à rédiger mon manuscrit en toute sérénité. Je remercie aussi Mathieu Reguerre et Julien Delbove de m'avoir fait part de leurs expériences en mécanique du vol. Un special thanks à Sebastiao Golding pour le temps qu'il a consacré à la relecture de mon article international.

Je remercie mes collègues matinaux pour les pauses café, mes collègues de midi pour les bons moments partagés à la table EGV, mes collègues sportifs pour les sorties piscine, et aussi mes collègues afterwork toujours prêts à partager un verre ou un bon repas.

Je tiens à remercier tous mes amis toulousains, plus particulièrement toute la "bande du tripode" qui s'est encore agrandie ces dernières années. Je remercie mes amis d'enfance, d'adolescence et maintenant d'adulte de Castres, ainsi que mes amis d'études supérieures de Clermont-Ferrand et de Berlin. Merci de faire partie de ma vie ! Merci plus particulièrement à Charlotte la lyonnaise, Veljko le toulousain, Clément l'écossais, Emilie l'anglaise, et Eva la polonaise ; votre présence discrète et attentionnée dans les moments difficiles

m'a beaucoup touchée et m'a aidée à avancer. Je remercie aussi mes supers colocataires, Aurore, Paul, Julie, Aurélien, Pumba, Franzi, Bruno, Manu, Aurélie et Angela. Partager tous ces bons moments a été précieux.

Enfin, je remercie ma famille, plus particulièrement mes parents qui m'ont toujours soutenue et m'ont donné la force de me surpasser dans tout ce que j'entreprends. Je remercie finalement ma grande soeur Léa, qui est mon pilier, mon soutien et mon exemple dans la vie.

# Table des matières

Table des figures	VII
Notations	XI
Glossaire	XV
Introduction générale	1
<b>1 Contexte industriel</b>	<b>5</b>
1.1 Introduction . . . . .	5
1.2 Modèles actuels de qualités de vol et de performances des avions . . . . .	6
1.2.1 Descriptions générales . . . . .	6
1.2.1.1 Modèle de performances des avions . . . . .	7
1.2.1.2 Modèle de qualités de vol . . . . .	8
1.2.1.3 Caractéristiques générales de ces deux modélisations . . . . .	8
1.2.2 Équations de la mécanique du vol . . . . .	10
1.2.2.1 Présentation des différents repères . . . . .	10
1.2.2.2 Changements de repères . . . . .	11
1.2.2.3 Modélisation des forces et des moments . . . . .	14
1.2.2.4 Variables dont dépendent les forces et les moments aérodynamiques . . . . .	20
1.2.3 Équations générales du modèle de qualités de vol . . . . .	25
1.2.4 Simplifications utilisées pour le modèle actuel de performances des avions . . . . .	29
1.2.4.1 Hypothèses physiques . . . . .	29
1.2.4.2 Simplification 1 : pré-équilibrage . . . . .	30
1.2.4.3 Simplification 2 : équilibrage moyen . . . . .	31
1.2.5 Description du modèle actuel de performances des avions . . . . .	32
1.2.5.1 Modèle aérodynamique actuel . . . . .	32
1.2.5.2 Analyse critique de la modélisation des performances des avions actuelle . . . . .	33
1.3 Vers un nouveau modèle de performance des avions . . . . .	35
1.3.1 Simplifications en vue de l'obtention du futur modèle aérodynamique de performances des avions . . . . .	35
1.3.2 Futur modèle de performances des avions . . . . .	38
1.4 Besoins et enjeux industriels de la réduction du modèle de performances des avions . . . . .	39

1.4.1	Structure et composition des modèles de qualités de vol et de performances des avions . . . . .	39
1.4.2	Construction actuelle des modèles de qualités de vol et de performances des avions . . . . .	40
1.4.3	Améliorations potentielles . . . . .	42
<b>2</b>	<b>État de l'art critique : modèles réduits, méthodes d'échantillonnage et de validation</b>	<b>45</b>
2.1	Introduction . . . . .	45
2.2	Principales étapes de construction d'un modèle réduit . . . . .	46
2.3	Différents types de modèles réduits . . . . .	52
2.3.1	Surfaces de réponses . . . . .	52
2.3.2	Réseaux de neurones . . . . .	54
2.3.3	Krigeage . . . . .	57
2.3.4	High Dimensional Model Representation . . . . .	61
2.4	Méthodes d'échantillonnage . . . . .	64
2.4.1	Grilles régulières . . . . .	65
2.4.2	Méthode de Monte Carlo directe . . . . .	66
2.4.3	Amélioration de la méthode de Monte Carlo directe par réduction de la variance . . . . .	67
2.4.3.1	Variables antithétiques . . . . .	67
2.4.3.2	Variables de contrôle . . . . .	68
2.4.3.3	Échantillonnage stratifié . . . . .	68
2.4.3.4	Méthode <i>Latin Hypercube Sampling</i> . . . . .	69
2.4.4	Méthode de quasi-Monte Carlo . . . . .	70
2.4.4.1	Notions de discrédance d'une suite . . . . .	70
2.4.4.2	Suites à faible discrédance . . . . .	73
2.4.5	Critères géométriques pour l'amélioration du remplissage de l'espace	78
2.5	Méthodes de validation . . . . .	79
2.5.1	Validation croisée . . . . .	81
2.5.2	Bootstrap . . . . .	83
2.5.3	Estimation <i>a posteriori</i> . . . . .	84
2.6	Choix des méthodes pour cette étude . . . . .	84
2.6.1	Choix des modèles réduits . . . . .	84
2.6.2	Choix des méthodes d'échantillonnage . . . . .	85
2.6.3	Choix des méthodes de validation . . . . .	85
<b>3</b>	<b>Adaptation des méthodes de réduction aux modèles de qualités de vol et de performances des avions</b>	<b>87</b>
3.1	Introduction . . . . .	87
3.2	Réduction des fonctions tabulées . . . . .	88
3.2.1	Random Sampling - High Dimensional Model Representation, cas d'un domaine de définition hypercubique . . . . .	88
3.2.1.1	Calcul analytique . . . . .	90
3.2.1.2	Application sur un exemple . . . . .	91
3.2.2	Interprétation probabiliste du Random Sampling - High Dimensional Model Representation . . . . .	95

3.2.2.1	Variables d'entrée indépendantes . . . . .	95
3.2.2.2	Variables d'entrée dépendantes . . . . .	97
3.2.2.3	Proposition d'une solution dans le cas non hypercubique . . . . .	101
3.2.3	Utilisation particulière des réseaux de neurones . . . . .	106
3.3	Réduction globale des modèles de qualité de vol et de performances des avions . . . . .	107
3.3.1	Fléau de la dimension . . . . .	108
3.3.2	Construction du Random Sampling - High Dimensional Model Representation par génération quasi-aléatoire d'un unique échantillon . . . . .	109
3.4	Conclusion . . . . .	110
<b>4</b>	<b>Méthodologie générale de réduction d'un modèle de qualités de vol et de performances des avions</b>	<b>113</b>
4.1	Introduction . . . . .	113
4.2	Formulation mathématique . . . . .	114
4.2.1	Architecture multi-niveaux hiérarchique . . . . .	114
4.2.2	Programmation Linéaire en Nombres Entiers . . . . .	116
4.2.3	Borne supérieure pour l'erreur globale . . . . .	118
4.3	Expériences numériques . . . . .	122
4.3.1	Application à un cas réel . . . . .	122
4.3.2	Caractéristiques des sous-modèles de référence et de leurs modèles réduits associés . . . . .	124
4.3.3	Résultats de l'optimisation . . . . .	127
4.3.3.1	Variation de la taille mémoire en fonction de l'erreur . . . . .	127
4.3.3.2	Variation de la taille mémoire en fonction du temps de calcul . . . . .	129
4.3.3.3	Bilan des expériences numériques . . . . .	131
4.4	Conclusion . . . . .	131
	<b>Conclusion générale</b>	<b>133</b>



# Table des figures

1.1	Construction d'un modèle numérique à partir d'un modèle physique. . . . .	6
1.2	Gouvernes d'un avion. . . . .	9
1.3	Position du repère avion par rapport au repère terrestre. . . . .	11
1.4	Position du repère aérodynamique par rapport au repère terrestre. . . . .	12
1.5	Position du repère aérodynamique par rapport au repère avion. . . . .	13
1.6	Poids à l'ordre 0. . . . .	15
1.7	Illustration du calage du moteur. . . . .	17
1.8	Illustration du pincement du moteur. . . . .	18
1.9	Atmosphère tempérée dite standard. . . . .	22
1.10	Calage du plan horizontal. . . . .	24
1.11	Schéma général simplifié du Modèle Perfo complet. . . . .	30
1.12	Schéma général simplifié du Modèle Perfo avec aérodynamique pré-équilibrée. . . . .	31
1.13	Schéma général simplifié du Modèle Perfo avec aérodynamique équilibrée. . . . .	32
1.14	Entrées/sorties du modèle aérodynamique Perfo actuel. . . . .	34
1.15	Entrées/sorties du modèle aérodynamique QdV après simplification. . . . .	38
1.16	Interpolation multi-linéaire. . . . .	39
1.17	Structure hiérarchique multi-niveaux. . . . .	41
1.18	Processus actuel de construction des modèles Perfo. . . . .	43
1.19	Processus futur de construction des modèles Perfo. . . . .	44
2.1	Erreur absolue locale au point $\mathbf{x}$ . . . . .	50
2.2	Réseau de neurones unidirectionnel et entièrement connecté. . . . .	55
2.3	Neurone de la couche d'entrée. . . . .	55
2.4	Neurone de la couche cachée $k$ . . . . .	56
2.5	Neurone de la couche de sortie. . . . .	57
2.6	Simulation de trajectoires d'un processus gaussien. . . . .	59
2.7	Simulation de trajectoires d'un processus gaussien conditionné à passer par des points donnés. . . . .	60
2.8	Grille régulière à 5 niveaux pour $[0, 1]^2$ . . . . .	65
2.9	Exemple d'échantillonnage stratifié avec $m = 4$ et $N = 20$ . . . . .	69
2.10	LHS pas uniformément réparti dans $\mathbb{H}^2$ . . . . .	70
2.11	Échantillon de 10 points équirépartis dans $[0, 1]^d$ avec $d = 1$ . . . . .	71
2.12	Discrépance centrée ( $d = 2$ ). . . . .	72
2.13	Discrépance extrême ( $d = 2$ ). . . . .	72
2.14	Projections 2D d'une suite de Halton ( $N = 50$ , $d = 15$ ) de la dimension 1 à la dimension 5. . . . .	76

2.15	Projections 2D d'une suite de Halton ( $N = 50, d = 15$ ) de la dimension 6 à la dimension 10. . . . .	76
2.16	Projections 2D d'une suite de Halton ( $N = 50, d = 15$ ) de la dimension 11 à la dimension 15. . . . .	77
2.17	Exemples de LHS optimisés. . . . .	79
2.18	Erreurs en fonction de la complexité du modèle. . . . .	80
3.1	Illustration des valeurs de sorties de $Ftab$ . . . . .	92
3.2	Illustration de la réduction de la taille mémoire par RS-HDMR. . . . .	92
3.3	Comparaison de l'erreur maximale exacte et estimée par la méthode de quasi-Monte Carlo. . . . .	94
3.4	Comparaison du temps de calcul de la méthode exacte et de la méthode de quasi-Monte Carlo. . . . .	94
3.5	Domaine non hypercubique $\Omega_1$ . . . . .	102
3.6	Fonction prolongée $\widetilde{Fp1}_{\gamma_1}$ sur $\mathbb{H}^2$ . . . . .	103
3.7	Fonction approchée $\widetilde{Ftab1}_{\gamma_1}$ sur $\mathbb{H}^2$ . . . . .	104
3.8	Fonction prolongée $\widetilde{Fp1}_{\gamma_2}$ sur $\mathbb{H}^2$ . . . . .	104
3.9	Fonction approchée $\widetilde{Ftab1}_{\gamma_2}$ sur $\mathbb{H}^2$ . . . . .	105
3.10	Fonction prolongée $\widetilde{Fp2}_{\gamma_3}$ sur $\mathbb{H}^2$ . . . . .	105
3.11	Fonction approchée $\widetilde{Ftab2}_{\gamma_3}$ sur $\mathbb{H}^2$ . . . . .	106
3.12	Réseau de neurones appelé sur une grille 1D. . . . .	107
3.13	Convergence de la méthode de quasi-Monte Carlo pour l'estimation d'une intégrale (parmi les 10 651 intégrales) en fonction du nombre de points. . .	109
3.14	Convergence de la méthode de quasi-Monte Carlo pour l'estimation de l'ensemble des coefficients des polynômes. . . . .	111
4.1	Structure hiérarchique multi-niveaux. . . . .	115
4.2	Structure hiérarchique multi-niveaux du FCM. . . . .	123
4.3	Visualisation de la taille mémoire de la solution optimale en fonction de la contrainte de précision. . . . .	127
4.4	Sous-modèles optimaux en fonction de la contrainte de précision. . . . .	128
4.5	Visualisation de la taille mémoire des solutions optimales en fonction de la contrainte de temps. . . . .	129
4.6	Sous-modèles optimaux en fonction de la contrainte de temps. . . . .	130

# Notations

Symbole	Unité	Description
$\Psi$	deg	angle d'azimut (cf. figure 1.3)
$\theta$	deg	angle d'assiette longitudinale (cf. figure 1.3)
$\phi$	deg	angle de gîte (cf. figure 1.3)
$\chi$	deg	angle d'azimut aérodynamique (cf. figure 1.4)
$\gamma$	deg	angle de pente aérodynamique (cf. figure 1.4)
$\mu$	deg	angle de gîte (cf. figure 1.4)
$\alpha$	deg	angle d'incidence (cf. figure 1.5)
$\beta$	deg	angle de dérapage (cf. figure 1.5)
$\dot{\alpha}$	deg · s <sup>-1</sup>	vitesse angulaire d'incidence de l'avion par rapport à l'air
$\dot{\beta}$	deg · s <sup>-1</sup>	vitesse angulaire de dérapage de l'avion par rapport à l'air
$\dot{\alpha}_{\text{Vent}}$	deg · s <sup>-1</sup>	vitesse angulaire d'incidence due au vent
$\dot{\beta}_{\text{Vent}}$	deg · s <sup>-1</sup>	vitesse angulaire de dérapage due au vent
$\vec{\Omega}_{R_a/R_0}$	deg · s <sup>-1</sup>	vecteur rotation du repère aérodynamique par rapport au repère terrestre
$\vec{\Omega}_{R_a/R_A}$	deg · s <sup>-1</sup>	vecteur rotation du repère aérodynamique par rapport au repère avion
$\vec{\Omega}_{R_A/R_0}$	deg · s <sup>-1</sup>	vecteur rotation du repère avion par rapport au repère terrestre
$m$	kg	masse de l'avion
$l$	m	longueur de référence
$MAC$	m	corde aérodynamique moyenne

$XCG$	%	position du centre de gravité sur l'avion par rapport à la corde aérodynamique moyenne
$S$	$m^2$	surface de référence
$Z_g$	m	altitude géopotentielle
$Z$	m	altitude géométrique
$Z_p$	m	altitude-pression fournie par un baromètre gradué
$h$	m	hauteur du centre de gravité par rapport à la piste
$P$	Pa	pression statique
$T$	K	température statique
$\Delta ISA$	deg	écart à la température standard
$conf$	–	configuration de l'aile
$PRF$	–	taux de compression du moteur
$g_0$	$m \cdot s^{-2}$	accélération de la pesanteur supposée constante
$\rho$	$kg \cdot m^{-3}$	masse volumique de l'air
$Re$	–	nombre de Reynolds
$\mu_1$	$kg \cdot m^{-1} \cdot s^{-1}$	viscosité dynamique du fluide
$\nu$	$m^2 \cdot s^{-1}$	viscosité cinématique du fluide
$\gamma_1$	–	rapport $\frac{C_p}{C_v}$ des chaleurs spécifiques (=1.4 pour l'air)
$r$	$J \cdot kg^{-1} \cdot K$	constante des gaz parfaits pour l'air (287, 053 $J \cdot kg^{-1} \cdot K$ ) : différence $C_p - C_v$
$a$	$m \cdot s^{-1}$	vitesse du son
$V_c$	$m \cdot s^{-1}$	module de la vitesse conventionnelle (Calibrated Air Speed)
$V_k$	$m \cdot s^{-1}$	module de la vitesse sol
$V_a$	$m \cdot s^{-1}$	module de la vitesse aérodynamique
$Mach$	–	Nombre de Mach
$F_g$	N	poussée brute du moteur

$F_N$	N	poussée nette du moteur
$n_z$	—	facteur de charge longitudinal
$r_s$	—	état de la piste
$\delta_s$	deg	position des spoilers
$\delta_{\text{becs}}$	deg	braquage des becs
$\delta_{\text{volets}}$	deg	braquage des volets
$\delta_{\text{train}}$	—	position des trains d'atterrissage
$\delta_{\text{trappes}}$	—	position des trappes de train
$\delta_{AF}$	deg	position des aérofreins
$\delta_p$	deg	braquage des commandes de vol en roulis
$\delta_q$	deg	braquage des commandes de vol en tangage
$\delta_r$	deg	braquage des commandes de vol en lacet
$iH$	deg	braquage du plan horizontal
$C_x$	—	coefficient de la force de traînée
$C_y$	—	coefficient de la force de portance latérale
$C_z$	—	coefficient de la force de portance verticale
$C_l$	—	coefficient du moment de roulis
$C_m$	—	coefficient du moment de tangage
$C_n$	—	coefficient du moment de lacet



# Glossaire

Nom	Définition
Ailerons	Surfaces mobiles situées à l'extrémité du bord de fuite des ailes. Ils permettent de piloter l'avion autour de son axe longitudinal.
Becs	Éléments mobiles du bord d'attaque des ailes utilisés à basse vitesse pour augmenter la portance de l'aile.
Dérive	Surface mobile située à la queue de l'avion. Elle permet de piloter l'avion autour de son axe vertical.
Domaine de vol	Ensemble des points du plan Mach-Altitude dans lequel le vol stabilisé est possible.
Empennage horizontal	Surfaces mobiles situées à la queue de l'avion qui permettent de piloter l'avion autour de son axe transversal. Il est composé du plan horizontal et de la gouverne de profondeur.
Lacet	Inclinaison de l'avion autour de son axe vertical.
Nombre de Mach	C'est le rapport entre la vitesse aérodynamique et la vitesse du son.
Pincement et calage du moteur	Angles représentant l'orientation de l'axe du moteur (qui est aussi la direction du vecteur poussée brute $F_g$ ) dans le repère avion.
Portance	Force aérodynamique perpendiculaire à la direction de la vitesse air et dirigée vers le haut de l'avion.
Poussée brute	Force de propulsion générée par le moteur.
Poussée nette	Différence des modules de la poussée brute et de la traînée de captation. Elle représente la poussée effectivement exploitable du moteur.

Roulis	Inclinaison de l'avion autour de son axe longitudinal.
Spoilers	Éléments mobiles situés sur l'extrados des ailes utilisés en particulier pendant l'atterrissage pour diminuer la portance de l'aile et donc améliorer l'efficacité du freinage.
Tangage	Inclinaison de l'avion autour de son axe transversal.
Trainée	Force aérodynamique colinéaire à la vitesse air et orientée dans le sens opposé à celle-ci.
Trainée de captation	Force de trainée aérodynamique due au moteur.
Vitesse aérodynamique	C'est la vitesse de l'avion par rapport au repère air. La différence par rapport à la vitesse sol provient de la présence éventuelle du vent.
Vitesse conventionnelle	Elle correspond à une différence de pressions convertie en vitesse dans les conditions de l'atmosphère standard à altitude nulle.
Vitesse sol	C'est la vitesse de l'avion par rapport au repère terrestre.
Volets	Surfaces mobiles situées à l'emplanture du bord de fuite des ailes utilisés à basse vitesse pour augmenter la portance de l'aile.

# Introduction générale

Dans de nombreux domaines des sciences et de l'industrie, que ce soit pour les prévisions météorologiques, économiques ou encore la conception d'un avion, la maîtrise des technologies de modélisation et de simulation numérique est devenue indispensable. Du fait de l'accroissement de la puissance des ordinateurs, une forte décroissance du coût des calculs a rendu compétitif le traitement de grands volumes de données ainsi que la simulation numérique qui complète, voire remplace, la démarche expérimentale classique. L'aéronautique fait partie de ces nombreux secteurs qui trouvent un intérêt à remplacer les essais réels par des études numériques.

Les modèles sont définis dans [83] comme des constructions matérielles ou formelles servant à représenter "quelque chose" de réel ou de fictif. Dans cette thèse, les modèles que nous allons étudier sont numériques et représentent des caractéristiques de l'avion, par exemple les forces auxquelles il est soumis. Ils sont destinés à être agrégés dans un simulateur grâce aux équations qui décrivent les mouvements de l'avion au cours du temps, notamment les lois du mouvement énoncées par Newton. La simulation numérique rend ainsi possible l'étude de plus en plus fine de l'influence de certains paramètres sur le comportement général de l'avion.

Ces simulations font appel à plusieurs modèles qui sont construits à partir de données issues de différentes sources, telles que des résultats d'essais (soufflerie, banc moteur, essais en vol), des calculs théoriques (intégration numérique des équations de Navier Stokes ou de la thermodynamique) ainsi que par similitude avec des modèles déjà existants. On parle alors de "modèles physiques", en ce sens que les relations mathématiques qui les constituent, visent à représenter la physique des phénomènes mis en jeu.

Une autre classe de modèles que nous appelons "modèles de représentation" [10] et dont la construction ne se fait plus à partir des lois physiques va être longuement abordée dans cette thèse. Ces modèles possèdent une représentativité élevée du phénomène global qu'ils modélisent au moyen d'un volume souvent plus réduit de données.

Nous venons de voir que plusieurs types de modèles existent, mais est-il toutefois possible d'en donner une définition générique ? Marvin Minsky [66] a proposé la définition suivante : "Pour un observateur B, un objet  $A^*$  est un modèle d'un objet A dans la mesure où B peut utiliser  $A^*$  pour répondre à des questions qui l'intéressent au sujet de A." . Un modèle n'est donc pas indépendant de l'utilisation qu'on en fait et sert toujours à

répondre à une question sur l'objet considéré.

Deux branches de la mécanique du vol se sont ainsi mises en place pour simuler le mouvement de l'avion faisant appel à des modèles différents. La première a pour but de représenter les qualités de vol de l'avion et la seconde ses performances au sol et en vol. Ces deux disciplines s'appuient sur des modèles présentant de nombreuses similarités, ce qui motive un besoin d'harmonisation des processus de construction des modèles en question. Comme nous le verrons par la suite, ce besoin d'harmonisation rend nécessaire la conception de nouvelles méthodes de réduction. Ces dernières sont à l'origine des travaux réalisés dans cette thèse. Ce mémoire est organisé en quatre chapitres.

Dans le chapitre 1, nous analysons les objectifs et les caractéristiques des modèles de performances des avions (dit modèle Perfo) et de qualités de vol (dit modèle QdV). Nous mettons ainsi en évidence les objectifs souvent contradictoires de précision et de temps de calcul de ces deux modèles. Le modèle Perfo, centre de cette étude, est alors étudié comme un modèle réduit du modèle QdV qui est plus complexe.

Par la suite, le besoin de stocker de plus en plus de données dans le cockpit de l'avion est souligné. Ceci mène à une nécessité de réduire la taille mémoire des modèles embarqués (en l'occurrence le modèle Perfo) tout en garantissant un temps de calcul compatible avec les exigences du temps réel. De plus, des contraintes sont identifiées en termes de précision et de préservation de la structure interne des modèles concernés.

Dans ce but, nous décrivons dans le chapitre 2 les différentes étapes d'une méthodologie générale de réduction de modèles. Puis, nous présentons un état de l'art critique des modèles réduits ainsi que des méthodes d'échantillonnage et de validation susceptibles d'être appliqués à la problématique présentée ci-dessus. Nous justifions finalement le choix de certaines méthodes sur lesquelles nous nous appuyons pour en développer de nouvelles.

Dans le chapitre 3, nous exploitons le Random Sampling - High Dimensional Model Representation (RS-HDMR) et les réseaux de neurones (RN) pour construire de nouvelles fonctions tabulées de dimension, de taille mémoire et/ou de temps de calcul inférieurs.

Dans ce contexte de fonctions tabulées, nous avons créé et développé une méthode de calcul analytique des différents termes de la décomposition RS-HDMR, ainsi qu'une extension du calcul de cette décomposition au cas des fonctions définies sur un domaine non hypercubique.

Par ailleurs, nous proposons une utilisation particulière des RN, visant à approcher au mieux les points de la fonction tabulée.

Ces méthodes donnent des résultats concluants en petite dimension. Par contre, nous mettons en évidence leurs limites d'applicabilité à la grande dimension. Ce fléau de la dimension nous conduit à exploiter dans le chapitre suivant la structure interne de ces

---

modèles hiérarchiques multi-niveaux de grande dimension composés d'un assemblage de fonctions tabulées de faible dimension.

Nous présentons dans le chapitre 4 une méthode originale de réduction de modèles basée sur la sélection optimale de sous-modèles réduits de petite dimension qui permet de préserver la structure initiale du modèle à réduire. La méthode proposée permet de formaliser cette sélection comme un problème d'optimisation linéaire en nombres entiers de telle sorte que le modèle global réduit obtenu satisfasse les exigences de rapidité, de précision et, pour certaines applications, de taille mémoire. Cette méthode est finalement appliquée et validée sur un cas test industriel.



# Chapitre 1

## Contexte industriel

### 1.1 Introduction

En aéronautique, plusieurs modélisations de l'avion existent visant chacune des utilisations différentes. Par exemple, ces modèles sont utilisés pour représenter le comportement général de l'avion (simulateur de vol), pour optimiser sa trajectoire ou encore pour prédire une consommation de carburant. Ces différentes utilisations ont amené Airbus à séparer la construction des modèles selon leurs domaines d'application et à construire plusieurs modèles répondant aux différents besoins. Cette construction parallèle selon le domaine d'application des modèles n'est plus satisfaisante aujourd'hui, car elle empêche le partage de données entre des disciplines pouvant être très proches.

Dans cette thèse, nous traitons de l'harmonisation du modèle de qualités de vol, dit QdV par la suite, et du modèle de performances des avions, dit Perfo. Plutôt que de les construire de façon complètement indépendante, le projet de recherche dans lequel s'insère cette thèse vise à dériver le modèle Perfo à partir du modèle QdV et ainsi d'améliorer la représentativité du modèle Perfo, ainsi que sa cohérence avec le modèle QdV et à harmoniser le processus global de construction.

Dans ce chapitre, nous détaillons ces deux grandes branches de modélisation, permettant d'une part d'étudier les qualités de vol d'un avion, d'autre part d'évaluer et d'optimiser ses performances. Après avoir montré les limites du modèle de performances des avions utilisé aujourd'hui, nous décrivons une nouvelle modélisation dérivée des qualités de vol et les enjeux industriels associés. Ce chapitre est composé de trois sections. Dans la section 1.2, une description générale qualitative de ces deux modélisations est donnée. Puis, en appliquant le principe fondamental de la dynamique, les équations générales du mouvement de l'avion définissant le modèle QdV sont introduites. Nous décrivons également le modèle Perfo utilisé actuellement en précisant ses limitations. Dans la section 1.3, les grandes lignes du projet de recherche dans lequel s'insèrent nos travaux sont décrites. Dans ce projet, il est envisagé de dériver le modèle Perfo à partir du modèle QdV afin d'en augmenter la représentativité et d'assurer la cohérence entre ces deux disciplines.

Par ailleurs, l'utilisation du modèle Perfo, autant sur les ordinateurs au sol que dans le cockpit de l'avion, impose des contraintes en termes de temps de calcul, de précision et de stockage mémoire qui sont distinctes des contraintes qu'impose l'utilisation du modèle QdV. À partir de ces contraintes, les besoins en terme de réduction mathématique de modèles sont décrits. Finalement, la section 1.4 est consacrée aux enjeux industriels associés à cette nouvelle modélisation.

## 1.2 Modèles actuels de qualités de vol et de performances des avions

### 1.2.1 Descriptions générales

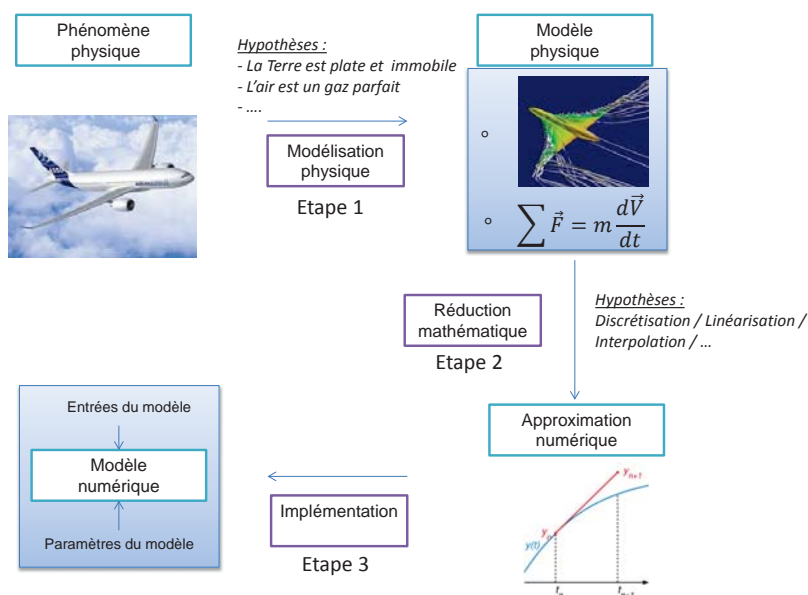


FIGURE 1.1 – Construction d'un modèle numérique à partir d'un modèle physique.

Un modèle physique est une formulation mathématique représentant un phénomène physique (cf. figure 1.1, étape 1). Une difficulté intrinsèque du concept de modèle est la définition de sa granularité, c'est-à-dire la taille et/ou le nombre de sous-systèmes permettant de décrire ce phénomène physique. Par exemple, un avion peut être vu comme composé de plusieurs sous-systèmes (propulsion, aérodynamique...), chacun d'eux étant décrit par un modèle spécifique (modèle de propulsion, modèle aérodynamique...). Mais on peut aussi voir l'avion à plus grande échelle. Ainsi, on appelle encore modèle le code qui permet de calculer la trajectoire de l'avion et comprend l'ensemble des sous-systèmes décrivant l'avion. Selon l'objectif du modèle, les caractéristiques décrites par ce dernier seront plus ou moins détaillées.

Une fois les modèles physiques et leurs niveaux de granularité déterminés, de nouvelles hypothèses dites de réduction mathématique peuvent être prises en compte (cf. figure 1.1, étapes 2 et 3).

En aéronautique, la simulation numérique est de plus en plus utilisée avec des modèles de plus en plus précis. Elle permet la mise en œuvre de méthodes mathématiques comme l’optimisation et ainsi par exemple la prédiction de trajectoires optimales. Elle permet aussi de réduire le nombre d’essais en vol. Elle facilite également l’exploration du domaine de vol, car elle permet la réalisation d’expériences à moindre frais (temps, argent...).

Dans cette section, une description générale des modèles QdV et Perfo est présentée. La connaissance des performances de l’avion ainsi que des qualités de vol est nécessaire à toutes les étapes de la conception d’un avion, c’est-à-dire autant en amont lorsque l’avion est en phase avant-projet, qu’après son entrée en service. Ces deux modélisations ont cependant des enjeux différents. De ce fait, les modèles QdV et Perfo ont une granularité différente.

### 1.2.1.1 Modèle de performances des avions

Dans le département dédié au calcul des performances des avions, l’avion est vu comme un point matériel soumis à des forces extérieures. Ce modèle est le plus simple que l’on puisse envisager pour un système mécanique. En effet, toute la masse est supposée concentrée en un point. Ce modèle est adapté dans le cas où l’on ne s’intéresse qu’au mouvement du centre de gravité, représenté par trois degrés de liberté. Dans un calcul de performances, la position de l’avion autour de son centre de gravité n’est qu’un paramètre permettant de piloter le module et la direction des forces auxquelles il est soumis (aérodynamique, propulsion, etc...).

Les modèles Perfo sont utilisés dans toutes les phases de développement d’un avion. En phase dite avant projet, ils permettent de concevoir un avion (dimension des moteurs, des surfaces des ailes...). Une fois la conception de l’avion terminée, les modèles Perfo sont utilisés dans le bureau d’étude pour calculer des trajectoires (décollage, montée, croisière, descente, atterrissage) et en déduire les performances des avions en fixant des scénarios de vol *a priori* pour établir notamment le manuel de vol, qui a pour but de déterminer des données pour le pilote, comme les distances d’atterrissage ou les vitesses de décollage selon les conditions du jour (masse, conditions atmosphériques,...).

Connaître plus précisément les performances des avions permet aussi de se placer convenablement sur le marché concurrentiel des constructeurs aéronautiques. En effet, les compagnies aériennes orientent leurs choix d’achat selon les caractéristiques de performances de l’avion et leurs besoins.

Des fonctions utilisant des modèles Perfo simplifiés sont également intégrées dans le cockpit de l’avion. Elles permettent d’améliorer en temps réel les performances, en prenant

en compte les conditions réelles du vol.

Certaines performances calculées sont réglementaires parce qu'elles sont critiques pour la sécurité du vol. Les distances de décollage et d'atterrissage ainsi que les pentes de montée citées plus haut en font partie. D'autres performances sont dites opérationnelles, comme la distance parcourue par unité de carburant consommé ou encore la trajectoire permettant de réduire les nuisances sonores à proximité des aéroports.

### 1.2.1.2 Modèle de qualités de vol

Dans le département dédié à l'étude des qualités de vol, l'avion est vu comme un solide indéformable soumis à des forces extérieures. Cette hypothèse permet de ne considérer que 6 degrés de liberté. Par rapport au modèle du point matériel, cette modélisation a pour objectif de représenter la dynamique de l'avion autour de son centre d'inertie. L'objectif est de connaître la réponse de l'avion aux sollicitations du pilote ou de l'environnement. Cette modélisation permet

- d'une part l'étude théorique de l'avion
  - stabilité : capacité de l'avion à revenir à une position d'équilibre prédéterminée après une perturbation externe ;
  - contrôlabilité : possibilité de déterminer un état d'équilibre dans tout le domaine de vol, c'est-à-dire pour toutes les conditions de vitesse, d'altitude possibles ;
  - manœuvrabilité : capacité de l'avion à atteindre un nouvel état d'équilibre ;
  - etc...
- et d'autre part des applications pratiques pour la mise en place de simulateurs de vol ou encore l'analyse d'accidents.

### 1.2.1.3 Caractéristiques générales de ces deux modélisations

Dans ces deux modélisations, malgré l'hypothèse de solide indéformable, les changements de géométrie de l'avion dus aux braquages des gouvernes ou à la souplesse de l'avion sont quand même pris en compte par les variations qu'ils induisent sur les forces extérieures (en particulier sur l'aérodynamique). Les variations de masse et de centrage au cours du vol sont quant à elles prises en compte dans le poids. Lorsqu'on connaît les modèles décrivant l'aérodynamique, la géométrie et les caractéristiques de poussée des moteurs ainsi que les conditions atmosphériques de vol, on est capable d'intégrer les équations de la mécanique du vol et ainsi de connaître le mouvement de l'avion dans l'espace.

Les forces s'appliquant sur l'avion pour le calcul des performances des avions et pour l'étude des qualités de vol sont :

- les forces aérodynamiques,
- le poids,

- les forces propulsives,
- la réaction du sol.

Pour l'étude des qualités de vol et des performances des avions, les forces aérodynamiques dépendent des conditions de vol et de la géométrie de l'avion.

Seule l'aérodynamique du modèle QdV dépend des vitesses angulaires de rotation autour du centre de masse. De plus les commandes sur lesquelles peut agir le pilote sont des entrées du modèle QdV. Ces commandes sont :

- la commande de gauchissement qui permet de contrôler les ailerons situés au bord de fuite et à l'extrémité de la voilure (notés A sur la figure 1.2). Le braquage de ces gouvernes crée essentiellement un moment de roulis, tendant à incliner l'avion à droite ou à gauche autour de l'axe  $X$  de la figure 1.2 ;
- la commande de profondeur qui permet au pilote de déplacer la gouverne de profondeur située généralement à l'arrière de l'avion (notée B sur la figure 1.2). Le braquage de cette gouverne crée un moment de tangage, qui fait cabrer ou piquer l'avion autour de l'axe  $Y$  de la figure 1.2 ;
- la commande de direction (notée C sur la figure 1.2) déplace la gouverne de direction placée au bord de fuite de la dérive. Le braquage de cette gouverne crée principalement un moment de lacet déplaçant le nez de l'avion vers la droite ou vers la gauche, c'est-à-dire autour de l'axe  $Z$  de la figure 1.2.



FIGURE 1.2 – Gouvernes d'un avion.

Dans le modèle Perfo, la trajectoire de l'avion est calculée en imposant certains paramètres (ex : pente, puissance moteur...) (cf. section 1.2.4) et ne dépend pas directement des commandes de vol qui en sont une conséquence.

Nous allons maintenant introduire les équations qui régissent le mouvement général de l'avion [8].

## 1.2.2 Équations de la mécanique du vol

### 1.2.2.1 Présentation des différents repères

Trois repères directs de référence sont généralement utilisés pour étudier le mouvement de l'avion.

#### Repère normal terrestre $R_0$

Le repère terrestre  $(0X_0Y_0Z_0)$  est lié à la terre :

- l'origine  $O$  est fixe par rapport à la terre ;
- l'axe  $OZ_0$  suit la verticale descendante de l'attraction gravitationnelle au point  $O$  ;
- $OX_0$  est arbitrairement choisi dans le plan horizontal normal à  $OZ_0$  ;
- $OY_0$  complète le repère direct.

De plus, nous faisons les hypothèses ici que :

- le repère terrestre est galiléen,
- la terre est plate. Ainsi les trajectoires à altitude constante sont rectilignes.

Il est en outre commode de définir un repère translaté du repère terrestre avec son origine liée à l'avion à un instant donné  $t$  (généralement son centre de gravité  $G$ ). C'est ce repère que l'on appelle repère normal terrestre.

#### Repère lié à l'avion $R_A$

Le repère  $(Gx_Ay_Az_A)$  est lié à l'avion :

- l'origine  $G$  est un point de référence de l'avion, en général son centre de masse. Sous l'hypothèse d'un champ de gravité constant,  $G$  est aussi le centre de gravité ;
- $Gx_A$  est orienté positivement de l'arrière à l'avant de l'avion, suivant la ligne de référence longitudinale ;
- $Gz_A$  est normal à  $Gx_A$  et appartient au plan de symétrie de l'avion. Il est orienté positivement vers le ventre de l'avion ;
- $Gy_A$  complète le repère direct (il est donc orienté positivement vers la droite du pilote).

#### Repère aérodynamique $R_a$

Le repère  $(Gx_a y_a z_a)$  est lié à la vitesse et à la géométrie de l'avion :

- l'origine  $G$  est le même point de référence que dans le repère  $R_A$  ;
- $Gx_a$  est orienté et porté par la vitesse-air, c'est-à-dire positivement de l'arrière à l'avant de l'avion ;
- $Gz_a$  est normal à  $Gx_a$ , situé dans le plan de symétrie de l'avion et orienté positivement vers le ventre de l'avion ;
- $Gy_a$  complète le repère direct.

### 1.2.2.2 Changements de repères

Nous définissons ici les matrices de rotation :  $R$  pour passer du repère avion au repère terrestre,  $U$  pour le passage du repère aérodynamique au repère terrestre et enfin  $T$  pour passer du repère aérodynamique au repère avion.

**Rappel :** En algèbre linéaire, une matrice de rotation  $P$  est une matrice orthogonale. Elle vérifie les propriétés suivantes :  $P^T P = Id = P P^T$  et  $\det(P) = 1$ , où  $P^T$  est la matrice transposée de  $P$  et  $Id$  est la matrice identité. On a donc  $P^T = P^{-1}$ .

#### Matrice de passage du repère avion au repère terrestre

L'ordre des rotations est le suivant (cf. figure 1.3) :

1. autour de l'axe  $GZ_0$  avec un angle  $\Psi$  (angle azimut) : amène  $GX_0$  en  $Gx_h$  et  $GY_0$  en  $Gy_h$  ;
2. autour de l'axe  $Gy_h$  avec un angle  $\theta$  (angle d'assiette longitudinale) : amène  $Gx_h$  en  $Gx_A$  et  $GZ_0$  en  $Gz_h$  ;
3. autour de l'axe  $Gx_A$  avec un angle  $\phi$  (angle de gîte) : amène  $Gy_h$  en  $Gy_A$  et  $Gz_h$  en  $Gz_A$ .

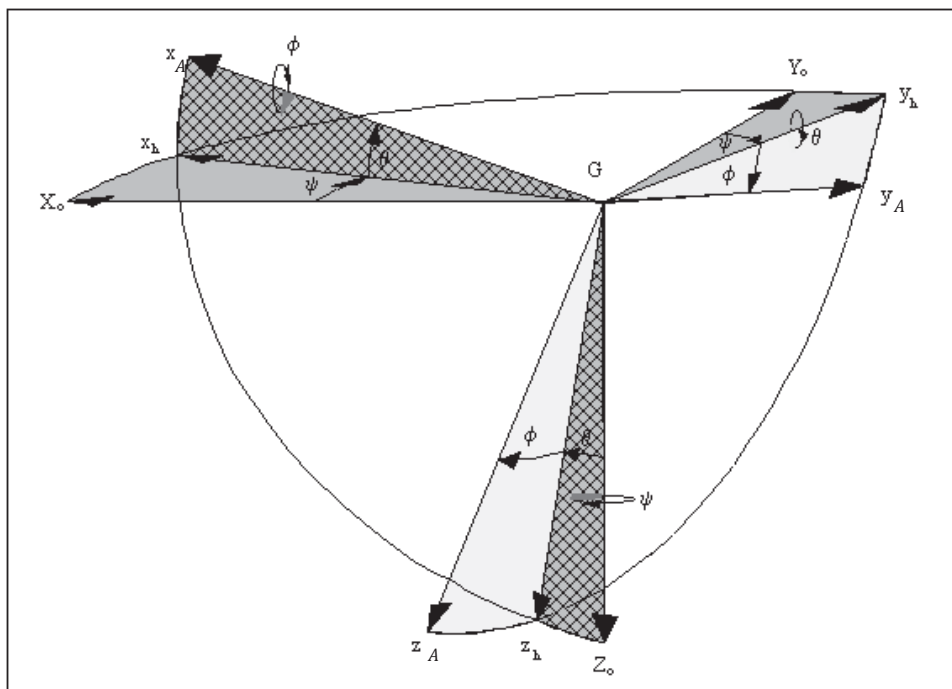


FIGURE 1.3 – Position du repère avion par rapport au repère terrestre.

La matrice de passage du repère avion au repère terrestre s'écrit donc :

$$[R] = \begin{pmatrix} \cos \Psi \cos \theta & \sin \Psi \cos \theta & -\sin \theta \\ -\sin \Psi \cos \phi + \cos \Psi \sin \theta \sin \phi & \cos \Psi \cos \phi + \sin \Psi \sin \theta \sin \phi & \cos \theta \sin \phi \\ \sin \Psi \sin \phi + \cos \Psi \sin \theta \cos \phi & -\cos \Psi \sin \phi + \sin \Psi \sin \theta \cos \phi & \cos \theta \cos \phi \end{pmatrix}$$

et on a :

$$\begin{pmatrix} x_A \\ y_A \\ z_A \end{pmatrix} = [R] \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix}.$$

### Matrice de passage du repère aérodynamique au repère terrestre

L'ordre des rotations est le suivant (cf. figure 1.4) :

1. autour de l'axe  $GZ_0$  avec un angle  $\chi$  (angle azimut aérodynamique) : amène  $GX_0$  en  $Gx_u$  et  $GY_0$  en  $Gy_u$  ;
2. autour de l'axe  $Gy_u$  avec un angle  $\gamma$  (angle de pente aérodynamique) : amène  $GZ_0$  en  $Gz_u$  et  $Gx_u$  en  $Gx_a$  ;
3. autour de l'axe  $Gx_a$  avec un angle  $\mu$  (angle de gîte) : amène  $Gy_u$  en  $Gy_a$  et  $Gz_u$  en  $Gz_a$ .

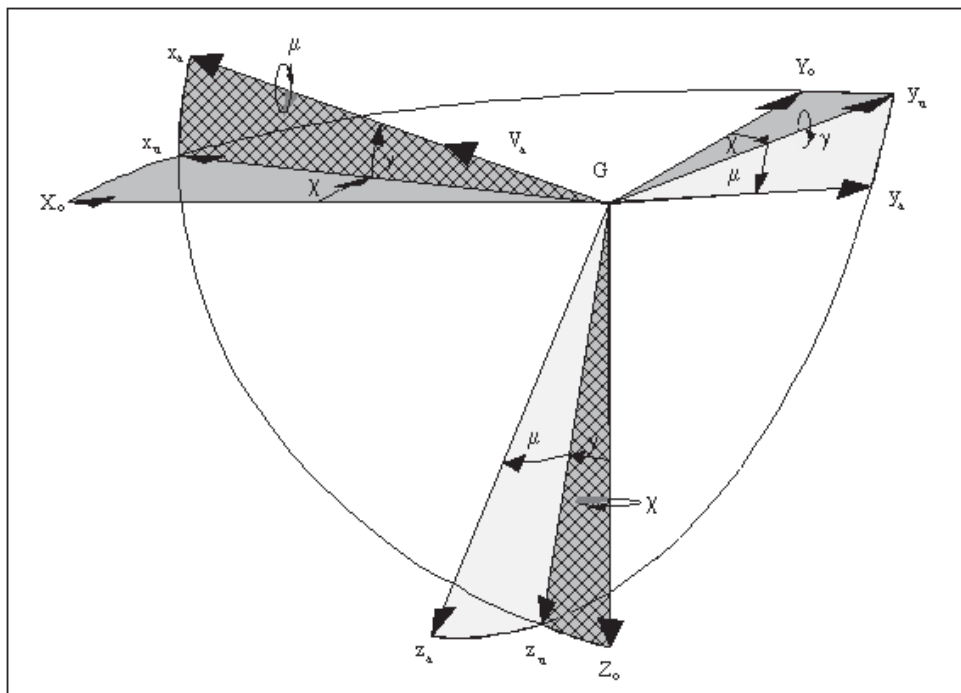


FIGURE 1.4 – Position du repère aérodynamique par rapport au repère terrestre.

La matrice de passage du repère aérodynamique au repère terrestre s'écrit :

$$[U] = \begin{pmatrix} \cos \chi \cos \gamma & \sin \chi \cos \gamma & -\sin \gamma \\ -\sin \chi \cos \mu + \cos \chi \sin \gamma \sin \mu & \cos \chi \cos \mu + \sin \chi \sin \gamma \sin \mu & \cos \gamma \sin \mu \\ \sin \chi \sin \mu + \cos \chi \sin \gamma \cos \mu & -\cos \chi \sin \mu + \sin \chi \sin \gamma \cos \mu & \cos \gamma \cos \mu \end{pmatrix}$$

et on a :

$$\begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} = [U] \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix}.$$

### Matrice de passage du repère avion au repère aérodynamique

Pour définir ce troisième changement de repère, deux angles suffisent puisque l'axe  $Gz_a$  est situé par définition dans le plan  $Gx_Az_A$  qui est le plan de symétrie de l'avion (cf. figure 1.5) :

1. l'angle d'incidence  $\alpha$  : angle de l'axe  $Gx_A$  avec le plan  $Gx_a y_a$  du repère aérodynamique ;
2. l'angle de dérapage  $\beta$  : angle de l'axe  $Gx_a$  avec le plan de symétrie  $Gx_A z_A$  de l'avion.

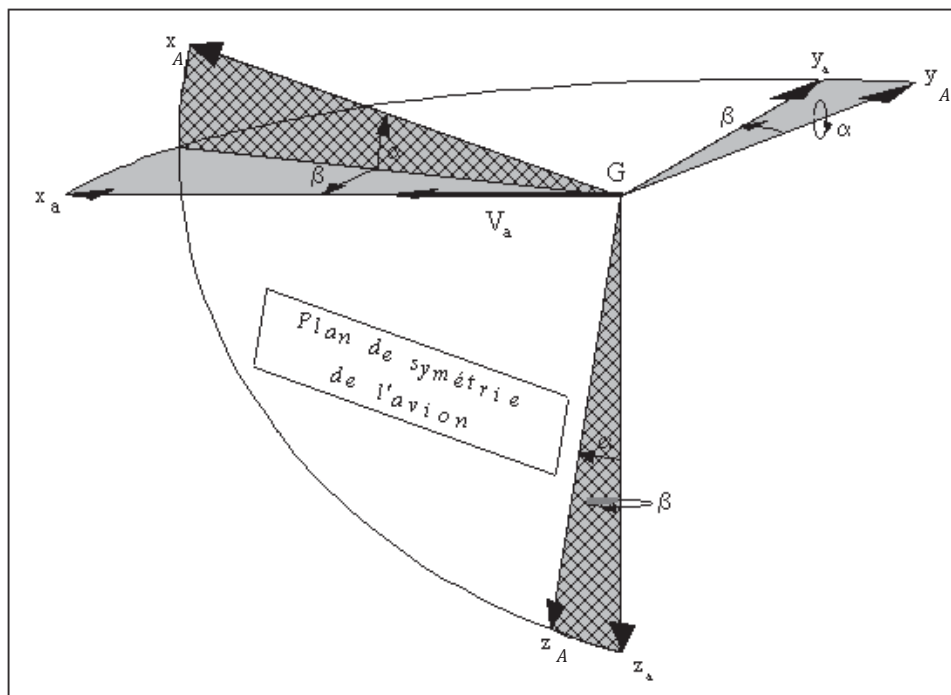


FIGURE 1.5 – Position du repère aérodynamique par rapport au repère avion.

La matrice de passage du repère avion au repère aérodynamique s'écrit :

$$[T] = \begin{pmatrix} \cos \alpha \cos \beta & -\cos \alpha \sin \beta & -\sin \alpha \\ \sin \beta & \cos \beta & 0 \\ \sin \alpha \cos \beta & -\sin \alpha \sin \beta & \cos \alpha \end{pmatrix}$$

et on a :

$$\begin{pmatrix} x_A \\ y_A \\ z_A \end{pmatrix} = [T] \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix}.$$

### 1.2.2.3 Modélisation des forces et des moments

On considère que les forces qui s'exercent sur l'avion se regroupent en trois catégories lorsque l'avion ne touche plus le sol :

- la force de gravité dont la résultante est  $\vec{P}$ ,
- les forces aérodynamiques dont la résultante est  $\vec{F}_{aero}$ , somme des contributions de chaque partie de l'avion (ailes, empennage horizontal,...),
- les forces de propulsion dont la résultante est  $\vec{F}_{prop}$ , somme des contributions de chacun des moteurs.

Au décollage et à l'atterrissage, lorsque l'avion est sur la piste, une quatrième force intervient :

- la réaction du sol dont la résultante est  $\vec{F}_{sol}$ .

### Poids

En première approximation, la force d'attraction de la Terre sur un corps de masse  $m$  à une altitude  $Z_g$  s'écrit :

$$\vec{P} = \frac{\mathcal{G}m_T m}{(R_T + Z_g)^2} \vec{e}_{z_0} = m \vec{g},$$

où  $\mathcal{G}$  est la constante universelle de la gravitation ( $\mathcal{G} = 6.67 \cdot 10^{-11} \text{ m}^3 \cdot \text{kg}^{-1} \cdot \text{s}^{-2}$ ),  $m_T$  la masse de la Terre ( $m_T = 5.97 \cdot 10^{24} \text{ kg}$ ),  $R_T$  le rayon moyen de la Terre ( $R_T = 6371 \text{ km} = 6,71 \cdot 10^6 \text{ m}$ ),  $\vec{e}_{z_0}$  le rayon vecteur unitaire dirigé vers le centre de la terre et  $\vec{g}$  l'accélération gravitationnelle.

Le module de la force de gravité vaut donc  $g = \frac{\mathcal{G}m_T}{(R_T + Z_g)^2}$  à l'ordre 0, c'est-à-dire lorsqu'on suppose que le potentiel gravitationnel de la Terre est celui d'une masse ponctuelle située en son centre (cf. figure 1.6).

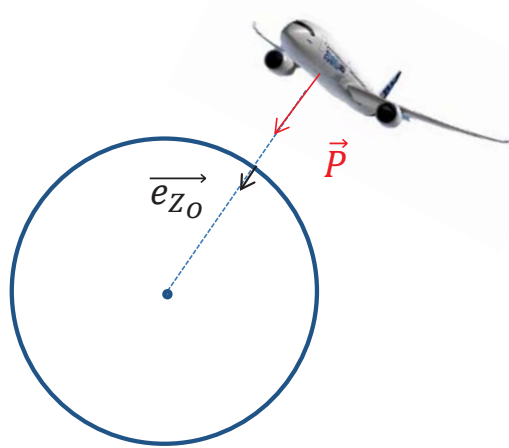


FIGURE 1.6 – Poids à l'ordre 0.

Cette hypothèse revient en fait à supposer que la symétrie sphérique de la Terre est parfaite. Or, des mesures précises montrent que des effets "non sphériques" sont détectables. De plus, la Terre tourne sur elle-même, ce qui crée une force centrifuge qui a tendance à repousser l'avion de la surface de la Terre. Cette force est d'autant plus forte que nous sommes proches de l'équateur. On introduit donc une correction à l'ordre 1 pour prendre en compte l'aplatissement de la Terre aux pôles et rendre le potentiel dépendant de la latitude.

Par souci de simplification de cet exposé, nous supposons ici que la gravité est constante quelle que soit l'altitude et quelle que soit la latitude. Ainsi, la valeur du terme de gravité utilisée est une approximation moyenne de  $g$  et vaut  $g_0 = 9.80665 \text{ m.s}^{-2}$ , qui correspond à  $g$  à l'altitude 0 et à la latitude  $45^\circ$ .

La force de gravité s'exprime simplement dans le repère terrestre  $R_0$  :

$$\vec{P} = \begin{pmatrix} 0 \\ 0 \\ mg_0 \end{pmatrix}.$$

En utilisant la matrice de passage du repère terrestre au repère aérodynamique  $[U]$ , elle s'exprime donc comme suit dans le repère aérodynamique  $R_a$  :

$$\vec{P} = \begin{pmatrix} -mg_0 \sin \gamma \\ mg_0 \cos \gamma \sin \mu \\ mg_0 \cos \gamma \cos \mu \end{pmatrix}.$$

Le moment associé au poids est nul car cette force s'applique par définition au centre de gravité.

La masse  $m$  dépend de la consommation de carburant et est donc une fonction du temps. Notons qu'elle peut aussi varier pour d'autres raisons (largages, ravitaillement en vol, accréation de glace, pressurisation de la cabine) que nous ne traiterons pas ici. La diminution de la masse totale rapportée au temps ne représente en pratique que 0.02 % de la masse totale de l'avion par minute pour un avion de ligne. Pour étudier une performance assez brève, par exemple une montée ou une distance d'atterrissage, on fait l'hypothèse que la masse est constante. En revanche, lors d'un vol en croisière de plusieurs heures, cette variation de masse intervient de façon significative, par exemple dans le calcul de la distance franchissable ou endurance (formule de Bréguet-Leduc [51]) : il n'est donc plus question de la négliger pour le calcul des performances de l'avion.

### Aérodynamique

Les efforts aérodynamiques, forces et moments, prennent la forme suivante :

$$\begin{aligned}\vec{F}_a &= \frac{1}{2}\rho V_a^2 S \vec{C}_F, \\ \vec{M}_a &= \frac{1}{2}\rho V_a^2 S l \vec{C}_M.\end{aligned}$$

Le symbole  $S$  représente la surface de référence, en général la surface des ailes. Le symbole  $l$  est la longueur de référence, en général la corde aérodynamique moyenne ( $MAC$ ). Les symboles  $\rho$  et  $V_a$  représentent la masse volumique de l'air et le module de la vitesse aérodynamique respectivement. Les vecteurs coefficients  $\vec{C}_F$  et  $\vec{C}_M$ , dont les composantes sont sans dimension, sont définis par leurs projections dans les différents repères. Projetés dans le repère aérodynamique  $R_a$ , ils ont comme expressions :

$$\begin{aligned}\vec{C}_F &= \begin{pmatrix} -C_x \\ C_y \\ -C_z \end{pmatrix}, \\ \vec{C}_M &= \begin{pmatrix} C_l \\ C_m \\ C_n \end{pmatrix}.\end{aligned}$$

Les coefficients  $C_x$ ,  $C_y$ ,  $C_z$  sont appelés respectivement coefficients de traînée, de force latérale aérodynamique et de portance. Les coefficients  $C_l$ ,  $C_m$ ,  $C_n$  sont appelés respectivement coefficients de moment de roulis, de tangage et de lacet.

On a finalement dans le repère aérodynamique  $R_a$ ,

$$\vec{F}_a = \begin{pmatrix} F_{xaero} \\ F_{yaero} \\ F_{zaero} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2}\rho V_a^2 S C_x \\ \frac{1}{2}\rho V_a^2 S C_y \\ -\frac{1}{2}\rho V_a^2 S C_z \end{pmatrix},$$

$$\vec{M}_a = \begin{pmatrix} M_{xaero} \\ M_{yaero} \\ M_{zaero} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\rho V_a^2 S l C_l \\ \frac{1}{2}\rho V_a^2 S l C_m \\ \frac{1}{2}\rho V_a^2 S l C_n \end{pmatrix}.$$

### Poussée des moteurs

La force de propulsion totale vaut dans le repère avion  $R_A$  :

$$\vec{F}_{prop} = \begin{pmatrix} F_{xprop} \\ F_{yprop} \\ F_{zprop} \end{pmatrix}.$$

Son moment associé, appliqué au centre de gravité G de l'avion, est noté  $\vec{M}_{F_{prop}}$ .

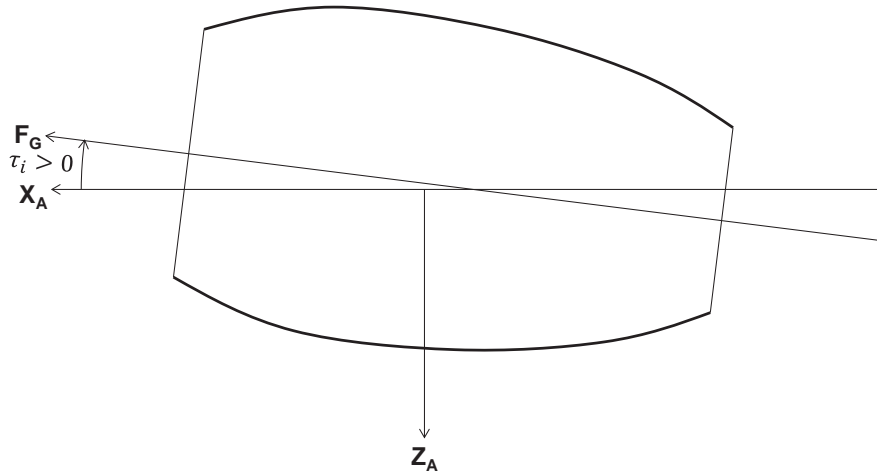


FIGURE 1.7 – Illustration du calage du moteur.

Chacun des  $N$  moteurs exerce une poussée brute notée  $\vec{F}_{G_i}$  dont la direction, qui dépend du calage  $\tau_i$  du moteur et de son pincement  $\pi_i$ , est fixe par construction dans le repère avion (cf. figures 1.7 et 1.8). La force de propulsion d'un moteur  $i$  s'exprime donc

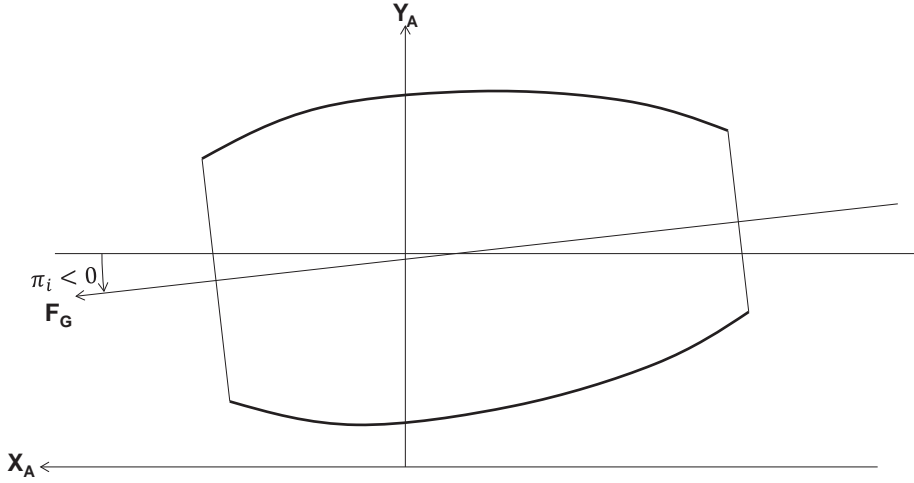


FIGURE 1.8 – Illustration du pincement du moteur.

dans le repère avion  $R_A$  avec les coordonnées suivantes :

$$\vec{F}_{G_i} = \begin{pmatrix} F_{xG_i} \\ F_{yG_i} \\ F_{zG_i} \end{pmatrix} = \begin{pmatrix} F_{G_i} \cos(\tau_i) \cos(\pi_i) \\ F_{G_i} \cos(\tau_i) \sin(\pi_i) \\ -F_{G_i} \sin(\tau_i) \end{pmatrix}.$$

Elle s'exprime donc dans le repère aérodynamique  $R_a$  de la manière suivante :

$$\vec{F}_{G_i} = \begin{pmatrix} F_{xG_i} \cos \alpha \cos \beta + F_{yG_i} \sin \beta + F_{zG_i} \sin \alpha \cos \beta \\ -F_{xG_i} \cos \alpha \sin \beta + F_{yG_i} \cos \beta - F_{zG_i} \sin \alpha \sin \beta \\ -F_{xG_i} \sin \alpha + F_{zG_i} \cos \alpha \end{pmatrix}.$$

Son module, aussi appelé poussée brute  $F_{G_i}$ , dépend du paramètre de conduite noté  $TSP$  ainsi que des conditions extérieures (Mach, température...). Le paramètre de conduite  $TSP$  dépend lui-même de la position de la manette des gaz  $\delta_{x_i}$ . La poussée brute constitue une première relation du modèle propulsif  $F_{G_i}(\delta_{x_i}, Mach, T, \dots)$ .

Chaque moteur est par ailleurs générateur d'une force de traînée aérodynamique. Cette force dépend de la position de la manette des gaz  $\delta_{x_i}$ . Elle est donc logiquement intégrée au modèle de propulsion et non au modèle aérodynamique. Cette force de traînée est notée  $\vec{F}_{RAM_i}$  et est orientée dans le sens opposé à la vitesse  $\vec{V}_a$ . Elle s'exprime donc

naturellement dans le repère aérodynamique  $R_a$  :

$$\vec{F}_{RAM_i} = \begin{pmatrix} -F_{RAM_i} \\ 0 \\ 0 \end{pmatrix}.$$

Comme pour la poussée brute, son module  $F_{RAM_i}$  dépend de  $\delta_{x_i}$  ainsi que des conditions extérieures ( $Mach$ ,  $T...$ ). Cette dépendance constitue la deuxième relation du modèle propulsif, dite trainée de captation :

$$F_{RAM_i}(\delta_{x_i}, Mach, T, \dots).$$

La force totale de propulsion s'écrit donc dans le repère aérodynamique  $R_a$  :

$$\vec{F}_{prop} = \sum_{i=1}^N (\vec{F}_{G_i} + \vec{F}_{RAM_i}).$$

Enfin, on note  $F_{N_i}$  la poussée nette, scalaire égal à la différence des modules de  $\vec{F}_{G_i}$  et  $\vec{F}_{RAM_i}$ .

$$F_{N_i} = F_{G_i} - F_{RAM_i}.$$

Comme  $\alpha$ ,  $\beta$ ,  $\tau_i$  et  $\pi_i$  sont généralement petits,  $\vec{F}_{G_i}$  et  $\vec{F}_{RAM_i}$  sont quasiment colinéaires si bien que  $F_{N_i}$  constitue une bonne approximation du module de la poussée effectivement disponible  $\vec{F}_{prop_i}$ .

Une autre équation d'évolution est constitutive du modèle de propulsion, reliant le débit de carburant à la masse :

$$WFE(\delta_{x_i}, Mach, T, \dots) = -\frac{dm}{dt}.$$

On remarque par ailleurs que la présence des moteurs modifie les caractéristiques aérodynamiques de l'avion, en raison d'une part de la perturbation de l'écoulement autour de la voilure et d'autre part de la déformation de celle-ci induite par les forces de propulsion. Ces phénomènes d'interaction entre les forces de poussée et les forces et moments aérodynamiques ne dépendent, du point de vue du moteur, que du taux de compression  $PRF$  et sont intégrés au modèle aérodynamique.

Enfin, le moment de chacune des forces de poussée au centre de gravité est déterminé par la connaissance de leurs bras de leviers respectifs  $\vec{GM}_i$  donnés par le modèle géométrique de l'avion.

## Réaction du sol

La réaction du sol sur l'avion s'effectue au niveau du contact des trains avec la piste. Sur

chacune des  $N_r$  roues  $i$ , une force  $\vec{F}_{sol_i}$  qui soutient l'avion est génératrice de frottements lorsque l'avion est en mouvement. On projette cette force dans le repère aérodynamique  $R_a$  qui est confondu avec le repère lié à la piste :

$$\vec{F}_{sol_i} = \begin{pmatrix} F_{x_{sol_i}} \\ F_{y_{sol_i}} \\ -F_{z_{sol_i}} \end{pmatrix}.$$

La composante normale  $F_{z_{sol_i}}$  dirigée selon l'axe  $GZ_a$  du repère aérodynamique et la composante tangentielle  $\sqrt{F_{x_{sol_i}}^2 + F_{y_{sol_i}}^2}$  située dans le plan  $(GX_a, OY_a)$  de ce même repère sont liées par une relation de frottement  $F_{z_{sol_i}} = f_{frott}(F_{x_{sol_i}}, F_{y_{sol_i}})$  dépendant entre autres de l'état de la piste ainsi que des commandes de freinage  $\delta f$  et de la position de la roulette de nez (qui est fonction du braquage de la gouverne de direction  $\delta r$ ). La relation  $f_{frott}$  constitue la partie principale du modèle de réaction du sol. Elle peut être assez complexe et nous ne la détaillerons pas ici [61].

La force totale de réaction du sol s'écrit donc dans le repère aérodynamique  $R_a$  :

$$\vec{F}_{sol} = \sum_{i=1}^{N_r} \vec{F}_{sol_i}.$$

Enfin, le moment de chacune des forces de réaction du sol est déterminé par la connaissance de leurs bras de leviers respectifs  $\vec{GP}_i$  donnés par le modèle géométrique de l'avion.

Les caractéristiques du modèle aérodynamique QdV sont incompatibles avec les exigences Perfo en termes de taille mémoire et de temps de calcul. Nos efforts se concentrent donc sur la construction d'un modèle aérodynamique Perfo par réduction du modèle aérodynamique QdV dont nous détaillons les variables d'entrée dans la section suivante.

#### 1.2.2.4 Variables dont dépendent les forces et les moments aérodynamiques

##### Atmosphère

À un instant donné et à une altitude déterminée au-dessus du niveau de la mer, les paramètres atmosphériques peuvent être très différents selon les conditions météorologiques. On définit donc un modèle d'atmosphère appelé atmosphère standard (ISA). Pour obtenir ce modèle simplifié, certaines hypothèses sont faites. Nous les décrivons ci dessous :

- Hypothèse 1 : L'air est assimilé à un gaz parfait. Il satisfait par conséquent l'équation d'état :  $P = \rho r T$ , avec  $P$  la pression statique,  $r$  la constante des gaz parfaits pour l'air et  $T$  la température statique.
- Hypothèse 2 : l'air est supposé en équilibre hydrostatique. Il vérifie donc la loi de Laplace,  $dP = -\rho g dZ_g$ .
- Hypothèse 3 : comme nous l'avons vu précédemment, les variations de l'accélération gravitationnelle  $g$  avec l'altitude et la latitude sont négligées ( $g = g_0$ ). On confond

donc altitude géopotentielle  $Z_g$  (liée aux équipotentielles du potentiel de gravité) et altitude géométrique  $Z$  ( $Z_g = Z$ ).

- Hypothèse 4 : en chaque point de l’atmosphère, la température ( $T$ ) varie en fonction de l’altitude géopotentielle  $Z_g$ . Une modélisation approchée des conditions atmosphériques consiste à établir un profil moyen de la variation de la température en fonction de l’altitude. La loi d’évolution de la température dans les régions tempérées qui a été retenue pour définir l’atmosphère standard (ISA) est celle caractérisée par les données suivantes.

- Lois d’évolution dans l’atmosphère standard :

- loi d’évolution de la température : la loi d’évolution de la température dans l’atmosphère standard est présentée dans la figure 1.9.

$$T = \begin{cases} T_0 = 288.15\text{K} & \text{si } Z_g = 0 \text{ m,} \\ T_0 - 0.0065 Z_g & \text{si } 0 \leq Z_g \leq 11000 \text{ m,} \\ 216.65\text{K} & \text{si } Z_g \geq 11000 \text{ m.} \end{cases}$$

Dans ces conditions, il existe une loi qui relie la masse volumique de l’air  $\rho$ , la température  $T$  et l’altitude  $Z_g$ .

- loi d’évolution de la masse volumique de l’air :

$$\rho = \begin{cases} \rho_0 = 1.225 \text{ kg/m}^3 & \text{si } Z_g = 0 \text{ m,} \\ \rho_0 \left( \frac{T}{T_0} \right)^{\left( \frac{-g_0}{-0.0065 r} + 1 \right)} & \text{si } 0 \leq Z_g \leq 11000 \text{ m,} \\ \rho_{11000} \exp \left( - \frac{g_0 (Z_g - 11000)}{r T_{11000}} \right) & \text{si } Z_g \geq 11000 \text{ m.} \end{cases}$$

- Lois d’évolution dans l’atmosphère réelle :

- loi d’évolution de la pression : un altimètre, instrument utilisé en aéronautique, est un baromètre gradué en mètres qui permet de mesurer une altitude-pression  $Z_p$ . Cette altitude-pression indiquée ne dépend que de la pression statique mesurée, laquelle est liée à l’altitude réelle par l’intermédiaire des conditions météorologiques du jour, qui sont la plupart du temps différentes de l’atmosphère standard. L’altitude mesurée par l’altimètre n’est donc égale à l’altitude géopotentielle  $Z_g$  que dans le cas où l’atmosphère réelle correspond effectivement à l’atmosphère standard. Moyennant la connaissance de la loi d’évolution de la température standard

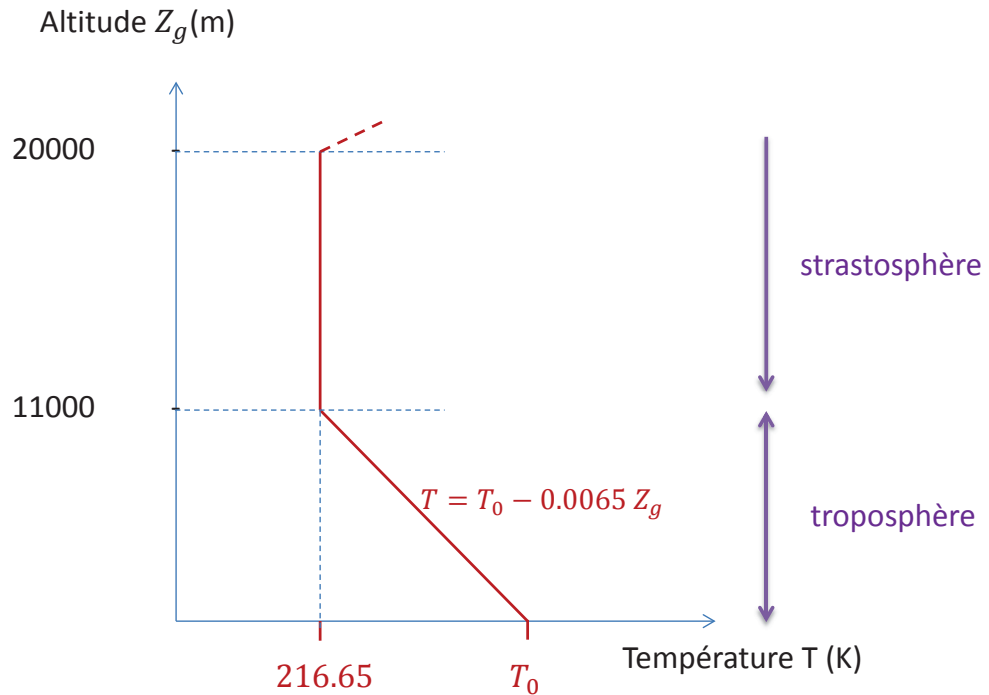


FIGURE 1.9 – Atmosphère tempérée dite standard.

(cf. figure 1.9), on est capable d'intégrer l'équation de Laplace (cf. hypothèse 2) :

$$P = \begin{cases} P_0 = 101325 \text{ Pa} & \text{si } Z_p = 0 \text{ m,} \\ P_0 \exp\left(-\frac{g_0}{0.0065 r} Z_p\right) \log\left(1 + \frac{0.0065 Z_p}{T_0}\right) & \text{si } 0 \leq Z_p \leq 11000 \text{ m,} \\ P_{11000} \exp\left(-g_0 \frac{Z_p - 11000}{r T_{11000}}\right) & \text{si } Z_p \geq 11000 \text{ m.} \end{cases}$$

En inversant ces relations, on obtient les lois de graduation de l'altimètre.

- loi d'évolution de la température réelle  $+\Delta\text{ISA}$  : pour tenir compte des variations atmosphériques, les calculs faits en atmosphère standard sont repris en supposant que pour chaque altitude-pression donnée, la température diffère de la température standard d'une valeur  $\Delta\text{ISA}$  constante. Ainsi, pour exprimer l'évolution réelle de la température, on translate la courbe de la figure 1.9 de cette valeur  $\Delta\text{ISA}$  le long de l'axe des températures. Par exemple si à 10000 m, la température observée est de 232.15 K, comme la température standard à 10000 m est  $\text{ISA} = 288.15 - 0.0065 \times 10000 = 223.15$  K, on en déduit que le vol s'effectue à 9 K au dessus de la température standard ( $\Delta\text{ISA} = 9$ ). On parle donc de vol opéré à température  $\text{ISA} + 9$ .

$$T = \begin{cases} T_0 = 288.15 + \Delta ISA \text{ K} & \text{si } Z_p = 0 \text{ m,} \\ T_0 + \Delta ISA - 0.0065 Z_p & \text{si } 0 \leq Z_p \leq 11000 \text{ m,} \\ 216.65 + \Delta ISA \text{ K} & \text{si } Z_p \geq 11000 \text{ m.} \end{cases}$$

### Vitesses

- Vitesse sol  $\vec{V}_k$  : c'est la vitesse absolue de l'avion par rapport à un repère lié à la terre  $R_0$ . Elle peut être obtenue à partir de mesures d'accélération à bord (centrale à inertie) ou par le système GPS. Du point de vue du pilotage et de l'étude du comportement de l'avion, cette vitesse a peu d'intérêt. Elle est essentiellement utilisée pour la navigation.
- Vitesse aérodynamique  $\vec{V}_a$  (TAS : True Air Speed) : c'est la vitesse de l'avion dans la masse d'air. La différence entre la vitesse sol et la vitesse aérodynamique provient de la présence du vent  $\vec{V}_w$  :  $\vec{V}_k = \vec{V}_a + \vec{V}_w$ .
- Nombre de Mach : c'est le rapport entre la vitesse aérodynamique et la vitesse du son  $a$  :  $Mach = \frac{V_a}{a}$ . La vitesse du son dans l'air est donnée par la relation :  $a = \sqrt{\gamma_1 r T}$  où  $\gamma_1$  est le rapport  $\frac{C_p}{C_v}$  (=1.4 pour l'air).
- Vitesse conventionnelle  $\vec{V}_c$  (CAS : Calibrated Air Speed) : elle correspond à une différence de pressions convertie en vitesse dans les conditions de l'atmosphère standard à  $Z_g = 0$ .

### Variables d'état

Le modèle aérodynamique a pour objet de représenter l'évolution des six composantes adimensionnées ( $C_x, C_y, C_z, C_l, C_m, C_n$ ) en fonction de la position, de la géométrie et du mouvement de l'avion, ainsi que de certaines caractéristiques de l'air dans lequel il se déplace.

- Position de l'avion : les coefficients aérodynamiques de l'avion dépendent de sa position dans l'espace par l'intermédiaire des angles d'incidence  $\alpha$  et de dérapage  $\beta$ . Lorsque l'avion est à proximité immédiate du sol, ces coefficients sont en outre influencés par les effets de sol qui dépendent de la hauteur  $h$  du centre de gravité par rapport à la piste, de l'angle de gîte  $\phi$  et de la pente  $\gamma$  de la trajectoire par rapport à la masse d'air.
- Géométrie de l'avion : les coefficients aérodynamiques de l'avion dépendent évidemment de sa géométrie. Celle-ci est susceptible d'évoluer au cours du vol en raison :
  - (a) des mouvements de certains éléments mobiles commandés par le pilote, à savoir :
    - les braquages des becs  $\delta_{\text{becs}}$  et volets  $\delta_{\text{volets}}$  correspondant aux différentes configurations *conf* de l'aile. Ainsi le pilote peut par exemple augmenter la portance de

- l'aile en Basse Vitesse ou adapter la géométrie de l'aile aux conditions de vol,
- la position des trains d'atterrissage et des trappes de trains, notée respectivement  $\delta_{\text{train}}$  et  $\delta_{\text{trappes}}$ ,
- les braquages des aérofreins, notés  $\delta_{AF}$ , qui permettent de réduire la portance et donc d'augmenter la pente de descente ou d'améliorer les capacités de freinage en augmentant les charges sur les trains d'atterrissage,
- les braquages des commandes de vol en roulis ( $\delta_p$ ), en tangage ( $\delta_q$ ) et en lacet ( $\delta_r$ ),
- la position du plan horizontal ( $iH$ ) (cf. figure 1.10) qui permet de maintenir la gouverne de profondeur au neutre ( $\delta_q = 0$ ) tout en maintenant l'avion en équilibre,

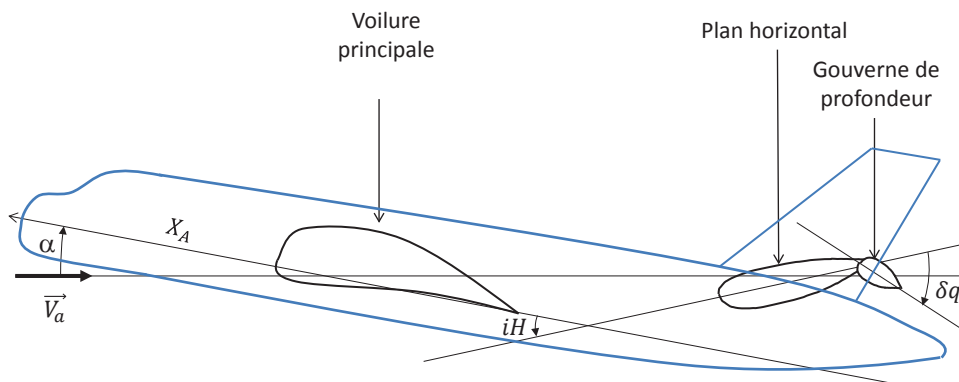


FIGURE 1.10 – Calage du plan horizontal.

(b) des déformations de l'avion dues à sa souplesse : ces déformations dépendent de l'accélération (gravité, facteur de charge ( $n_z$ )). Le facteur de charge est le rapport entre le poids apparent (gravité et forces d'inerties) et le poids réel de l'avion (gravité).

On peut aussi citer dans cette rubrique les modifications involontaires de la géométrie de l'avion (braquages intempestifs, perte de certains éléments...) qui font l'objet de traitements spécifiques (cas de pannes).

- Mouvement de l'avion : les coefficients aérodynamiques de l'avion dépendent du mouvement de l'avion par l'intermédiaire :
  - du nombre de Mach et du module  $V_a$  de la vitesse du centre de gravité par rapport à l'air. Ces deux paramètres sont liés par la température  $T$  et donc par l'altitude  $Z_p$  de l'avion ;
  - de la vitesse angulaire de l'avion par rapport à l'air, caractérisée par les variations d'incidence  $\dot{\alpha}$  et de dérapage  $\dot{\beta}$  ;
  - de la vitesse angulaire  $\vec{\Omega}_{R_A/R_0}$  du repère avion  $R_A$  par rapport au repère sol  $R_0$  ;

- des variations d'incidence  $\dot{\alpha}_{\text{Vent}}$  et de dérapage  $\dot{\beta}_{\text{Vent}}$  dues au vent.
- Poussée brute de l'avion : comme nous l'avons vu, la modification de l'écoulement autour de l'avion ainsi que la déformation de celui-ci sous l'effet des forces de propulsion entraînent une variation des coefficients aérodynamiques, qui dépend de la poussée brute  $F_{Gi}$  ou plus directement du taux de compression  $PRF$ .
- Caractéristique de l'air : les coefficients aérodynamiques de l'avion dépendent de la viscosité de l'air à travers le nombre de Reynolds  $Re = \frac{\rho V_a l}{\mu_1}$ , car on ne fait pas l'hypothèse que l'air est un fluide parfait ; ainsi même si la viscosité de l'air  $\nu = \frac{\mu_1}{\rho}$  est très faible, elle n'est pas nulle. Le nombre de Reynolds  $Re$  est également fonction de la température  $T$  et de la pression statique  $P$  de l'atmosphère via  $\rho$  et donc de l'altitude  $Z_p$  de l'avion.
- État de la piste : le coefficient de traînée aérodynamique de l'avion dépend de l'état de la piste  $r_s$ . Par exemple, lorsque la piste est mouillée à cause des projections d'eau, on observe une augmentation de la traînée.

### 1.2.3 Équations générales du modèle de qualités de vol

On étudie le mouvement du centre de gravité d'un solide indéformable, ainsi que son mouvement autour du centre de gravité par le principe fondamental de la dynamique et le théorème du moment cinétique. Dans le repère normal terrestre considéré comme galiléen, en supposant la masse constante, les équations générales du mouvement s'écrivent :

$$m \frac{d\vec{V}_a}{dt} = \sum \vec{F}_{ext}, \quad (1.1)$$

$$\frac{d\vec{\mathcal{C}}}{dt} = \sum \vec{\mathcal{M}}_{\vec{F}_{ext}}, \quad (1.2)$$

où  $\frac{d\vec{\mathcal{C}}}{dt}$  représente la variation par rapport au temps du moment cinétique de l'avion,  $\sum \vec{F}_{ext}$  la résultante des forces extérieures agissant sur l'avion (force de pesanteur, forces aérodynamiques, forces de propulsion) et  $\sum \vec{\mathcal{M}}_{\vec{F}_{ext}}$  le moment résultant des différentes forces agissant sur l'avion (moment des forces aérodynamiques et moment des forces de propulsion).

On suppose la vitesse du vent nulle  $\vec{V}_w = \vec{0}$ .

Nous avons

$$\vec{\Omega}_{R_a/R_0} = \vec{\Omega}_{R_a/R_A} + \vec{\Omega}_{R_A/R_0},$$

où

- $\vec{\Omega}_{R_a/R_0}$  est le vecteur rotation du repère aérodynamique par rapport au repère terrestre,

- $\vec{\Omega}_{R_a/R_A}$  est le vecteur rotation du repère aérodynamique par rapport au repère avion,
- $\vec{\Omega}_{R_A/R_0}$  est le vecteur rotation du repère avion par rapport au repère terrestre.

Dans le repère aérodynamique  $R_a$ , ces vecteurs ont pour expressions :

$$\vec{\Omega}_{R_a/R_0} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix},$$

$$\vec{\Omega}_{R_a/R_A} = \begin{pmatrix} -\dot{\alpha} \sin \beta \\ -\dot{\alpha} \cos \beta \\ \dot{\beta} \end{pmatrix},$$

$$\vec{\Omega}_{R_A/R_0} = \begin{pmatrix} p_a \\ q_a \\ r_a \end{pmatrix} = \begin{pmatrix} \dot{\mu} - \dot{\chi} \sin \gamma \\ \dot{\chi} \cos \gamma \sin \mu + \dot{\gamma} \cos \mu \\ \dot{\chi} \cos \gamma \cos \mu - \dot{\gamma} \sin \mu \end{pmatrix},$$

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} = \begin{pmatrix} -\dot{\alpha} \sin \beta \\ -\dot{\alpha} \cos \beta \\ \dot{\beta} \end{pmatrix} + \begin{pmatrix} \dot{\mu} - \dot{\chi} \sin \gamma \\ \dot{\chi} \cos \gamma \sin \mu + \dot{\gamma} \cos \mu \\ \dot{\chi} \cos \gamma \cos \mu - \dot{\gamma} \sin \mu \end{pmatrix}.$$

Par définition, la vitesse dans le repère aérodynamique  $R_a$  a pour coordonnées :

$$\vec{V}_a = \begin{pmatrix} V_a \\ 0 \\ 0 \end{pmatrix}$$

Par ailleurs, on note  $A, B, C$  les moments d'inertie par rapport aux axes  $Gx_A, Gy_A, Gz_A$  du repère avion  $R_A$ . Les termes hors diagonales  $D, E, F$  sont les produits d'inertie. La matrice d'inertie s'écrit donc comme suit dans le repère avion  $R_A$  :

$$\begin{pmatrix} A & -F & -E \\ -F & B & -D \\ -E & -D & C \end{pmatrix}.$$

Pour exprimer le moment cinétique dans le repère aérodynamique  $R_a$ , on a la relation suivante :

$$\begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{12} & J_{22} & J_{23} \\ J_{13} & J_{23} & J_{33} \end{pmatrix} = [T^{-1}] \begin{pmatrix} A & -F & -E \\ -F & B & -D \\ -E & -D & C \end{pmatrix} [T],$$

où  $[T^{-1}]$  est en fait la matrice transposée de  $[T]$ , comme il a été rappelé précédemment. Le moment cinétique  $\vec{C}$  dans le repère aérodynamique  $R_a$  est donc donné par :

$$\vec{C} = \begin{pmatrix} J_{11} & J_{12} & J_{13} \\ J_{12} & J_{22} & J_{23} \\ J_{13} & J_{23} & J_{33} \end{pmatrix} \cdot \vec{\Omega}_{R_a/R_0}.$$

Les composantes du moment cinétique dans le repère aérodynamique  $R_a$  sont donc :

$$\vec{C} = \begin{pmatrix} J_{11}p_a + J_{12}q_a + J_{13}r_a \\ J_{12}p_a + J_{22}q_a + J_{23}r_a \\ J_{13}p_a + J_{23}q_a + J_{33}r_a \end{pmatrix}.$$

On cherche maintenant à dériver dans le repère aérodynamique  $R_a$  qui est mobile par rapport au repère terrestre galiléen la vitesse aérodynamique  $\vec{V}_a = (V_a, 0, 0)^T$ .

Soit  $\vec{i} = (1, 0, 0)^T$  le vecteur unitaire de l'axe  $Gx_a$ ; on a donc  $\vec{V}_a = V_a \vec{i}$ .

Ainsi dans le repère aérodynamique  $R_a$ ,

$$\begin{aligned} \left. \frac{d\vec{V}_a}{dt} \right|_{R_0} &= \frac{dV_a}{dt} \vec{i} + V_a \frac{d\vec{i}}{dt} \\ &= \frac{dV_a}{dt} \vec{i} + V_a \vec{\Omega}_{R_a/R_0} \wedge \vec{i} \\ &= \begin{pmatrix} \frac{dV_a}{dt} \\ V_a \omega_3 \\ -V_a \omega_2 \end{pmatrix}. \end{aligned}$$

On retrouve une relation fondamentale de la cinématique :

$$\left. \frac{d\vec{V}_a}{dt} \right|_{R_0} = \left. \frac{d\vec{V}_a}{dt} \right|_{R_a} + \vec{\Omega}_{R_a/R_0} \wedge \vec{V}_a.$$

De la même façon, pour le moment cinétique, on a :

$$\left. \frac{d\vec{\mathcal{C}}}{dt} \right|_{R_0} = \left. \frac{d\vec{\mathcal{C}}}{dt} \right|_{R_a} + \vec{\Omega}_{R_a/R_0} \wedge \vec{\mathcal{C}}.$$

En supposant la répartition de masse constante, on obtient donc dans le repère aérodynamique  $R_a$ , :

$$\frac{d\vec{\mathcal{C}}}{dt} = \begin{pmatrix} J_{11}\dot{p}_a + J_{12}\dot{q}_a + J_{13}\dot{r}_a + J_{33}\omega_2 r_a - J_{22}\omega_3 q_a + J_{13}\omega_2 p_a - J_{12}\omega_3 p_a + J_{23}(\omega_2 q_a - \omega_3 r_a) \\ J_{12}\dot{p}_a + J_{22}\dot{q}_a + J_{23}\dot{r}_a + J_{11}\omega_3 p_a - J_{33}\omega_1 r_a + J_{12}\omega_3 q_a - J_{23}\omega_1 q_a + J_{13}(\omega_3 r_a - \omega_1 p_a) \\ J_{13}\dot{p}_a + J_{23}\dot{q}_a + J_{33}\dot{r}_a + J_{22}\omega_1 q_a - J_{11}\omega_2 p_a + J_{23}\omega_1 r_a - J_{13}\omega_2 r_a + J_{12}(\omega_1 p_a - \omega_2 q_a) \end{pmatrix}.$$

### Équilibre des forces dans le repère aérodynamique $R_a$

$$\text{Selon } Gx_a : \quad m \frac{dV_a}{dt} = -\frac{1}{2}\rho V_a^2 S C_x + F_{xprop} + F_{xsol} - mg_0 \sin \gamma. \quad (1.3a)$$

$$\text{Selon } Gy_a : \quad mV_a \omega_3 = \frac{1}{2}\rho V_a^2 S C_y + F_{yprop} + F_{ysol} + mg_0 \cos \gamma \sin \mu. \quad (1.3b)$$

$$\text{Selon } Gz_a : \quad -mV_a \omega_2 = -\frac{1}{2}\rho V_a^2 S C_z + F_{zprop} - F_{zsol} + mg_0 \cos \gamma \cos \mu. \quad (1.3c)$$

### Équilibre des moments dans le repère aérodynamique $R_a$ au point G

Autour de  $Gx_a$  :

$$\begin{aligned} J_{11}\dot{p}_a + J_{12}\dot{q}_a + J_{13}\dot{r}_a + J_{33}\omega_2 r_a - J_{22}\omega_3 q_a + J_{13}\omega_2 p_a - J_{12}\omega_3 p_a + J_{23}(\omega_2 q_a - \omega_3 r_a) \\ = \frac{1}{2}\rho V_a^2 S l C_l + \mathcal{M}_{F_{xprop}} + \mathcal{M}_{F_{xsol}}. \end{aligned} \quad (1.4a)$$

Autour de  $Gy_a$  :

$$\begin{aligned} J_{12}\dot{p}_a + J_{22}\dot{q}_a + J_{23}\dot{r}_a + J_{11}\omega_3 p_a - J_{33}\omega_1 r_a + J_{12}\omega_3 q_a - J_{23}\omega_1 q_a + J_{13}(\omega_3 r_a - \omega_1 p_a) \\ = \frac{1}{2}\rho V_a^2 S l C_m + \mathcal{M}_{F_{yprop}} + \mathcal{M}_{F_{ysol}}. \end{aligned} \quad (1.4b)$$

Autour de  $Gz_a$  :

$$\begin{aligned} J_{13}\dot{p}_a + J_{23}\dot{q}_a + J_{33}\dot{r}_a + J_{22}\omega_1 q_a - J_{11}\omega_2 p_a + J_{23}\omega_1 r_a - J_{13}\omega_2 r_a + J_{12}(\omega_1 p_a - \omega_2 q_a) \\ = \frac{1}{2}\rho V_a^2 S l C_n + \mathcal{M}_{F_{zprop}} + \mathcal{M}_{F_{zsol}}. \end{aligned} \quad (1.4c)$$

Le modèle développé par le département des qualités de vol permet d'étudier le mouvement général de l'avion en réponse aux commandes du pilote. Dans la section 1.2.4, les hypothèses associées au modèle Perfo actuel sont explicitées. Dans la section 1.3, un nouveau modèle Perfo est proposé permettant de s'affranchir de certaines d'entre elles et ainsi d'augmenter sa représentativité.

## 1.2.4 Simplifications utilisées pour le modèle actuel de performances des avions

### 1.2.4.1 Hypothèses physiques

La trajectoire de l'avion (mouvement de son centre de gravité) et son attitude (évolution de sa position autour de son centre de gravité) sont les conséquences des forces et des moments qui lui sont appliqués via les équations (1.1) et (1.2).

Ces forces et moments dépendent de l'action du pilote sur les commandes, c'est-à-dire essentiellement :

- de la commande en poussée  $\delta_x$ ,
- de la commande en roulis  $\delta_p$ ,
- des commandes en tangage  $\delta_q$  et  $iH$ ,
- de la commande en lacet  $\delta_r$ ,
- de la commande de freinage  $\delta_f$ .

Pour modéliser l'avion, on distingue trois entités : l'aérodynamique, les moteurs et la réaction du sol à l'atterrissage ou au décollage. Les commandes actionnées par le pilote ( $\delta_x, \delta_p, \delta_q, iH, \delta_r, \delta_f$ ) ont une influence sur une ou plusieurs de ces entités. Les commandes en roulis, en tangage et en lacet influent sur l'aérodynamique de l'avion. La commande en poussée agit sur la propulsion. Enfin, les commandes en roulis, en tangage et de freinage influent sur la force de réaction du sol.

### Hypothèses

- Hypothèse 1 : on suppose l'équilibre des moments vérifié tout au long de la trajectoire (vitesse angulaire constante par rapport au centre de gravité) :

$$\sum \vec{M}_{\vec{F}_{ext}} = \vec{0}. \quad (1.5)$$

Lorsque l'avion est en vol, cette hypothèse permet la détermination de la valeur du moment aérodynamique à partir de la connaissance des autres moments en présence et ainsi d'en déduire les valeurs des braquages de gouvernes à l'équilibrage de l'équation (1.5). Elle revient à considérer que la position des commandes de vol en roulis, en tangage et en lacet annule la somme des moments autour des trois axes.

- Hypothèse 2 : pour le calcul des performances des avions, les forces aérodynamiques ne dépendent pas des vitesses angulaires de rotation autour du centre de masse, l'avion étant modélisé comme un point matériel. Ces dernières sont supposées nulles en entrée du modèle aérodynamique. On parle de modèle statique.

Le schéma général très simplifié (les interdépendances entre les boîtes ne sont notamment pas représentées) de l'évolution de l'avion en vol (la force de réaction sol n'est donc pas représentée) en réponse aux commandes de pilotage est donné dans la figure 1.11. Les commandes aérodynamiques sont ajustées pour équilibrer l'avion (c'est-à-dire pour que

l'équation (1.5) soit vérifiée) et les forces aérodynamiques qui en résultent, associées aux forces de propulsion et de réaction du sol, permettent de calculer la trajectoire (1.1).

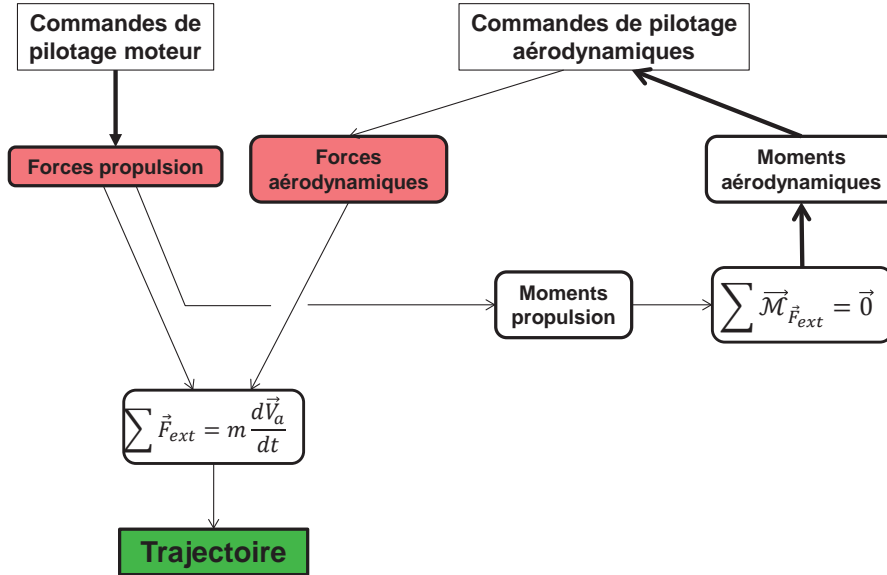


FIGURE 1.11 – Schéma général simplifié du Modèle Perfo complet.

Ces hypothèses physiques permettent de modéliser l'avion avec la granularité nécessaire pour les études Perfo. Nous allons maintenant introduire des simplifications permettant de réduire notamment le temps de calcul d'une trajectoire Perfo.

#### 1.2.4.2 Simplification 1 : pré-équilibrage

Les commandes de vol  $\delta_p$ ,  $\delta_q$ ,  $iH$  et  $\delta_r$  sont des variables d'entrée du modèle QdV complet qui ne sont pas directement utilisées pour les calculs de performances des avions. Ces commandes sont toutefois nécessaires au calcul des forces aérodynamiques. La première simplification consiste à donner à ces commandes une valeur pré-établie, représentative des conditions de vol équilibré (basses vitesses, croisière,...). On s'affranchit donc de la boucle d'équilibrage (1.5) en précalculant les braquages des gouvernes en fonction du moment de propulsion. Ces fonctions sont ensuite directement intégrées au modèle aérodynamique lui-même. De cette façon, ce dernier possède trois entrées supplémentaires (les 3 composantes du moment de propulsion) et fournit directement les coefficients  $C_x$ ,  $C_y$  et  $C_z$  correspondant aux valeurs de  $\delta_p$ ,  $\delta_q$ ,  $iH$  et  $\delta_r$  qui équilibrent les moments de propulsion en entrée. Les braquages des gouvernes ne sont alors plus que des intermédiaires de calcul et n'ont pas besoin d'apparaître explicitement comme des entrées du modèle aérodynamique, ni du modèle Perfo dont il fait partie. Le modèle aérodynamique ainsi obtenu est

représenté en figure 1.12.

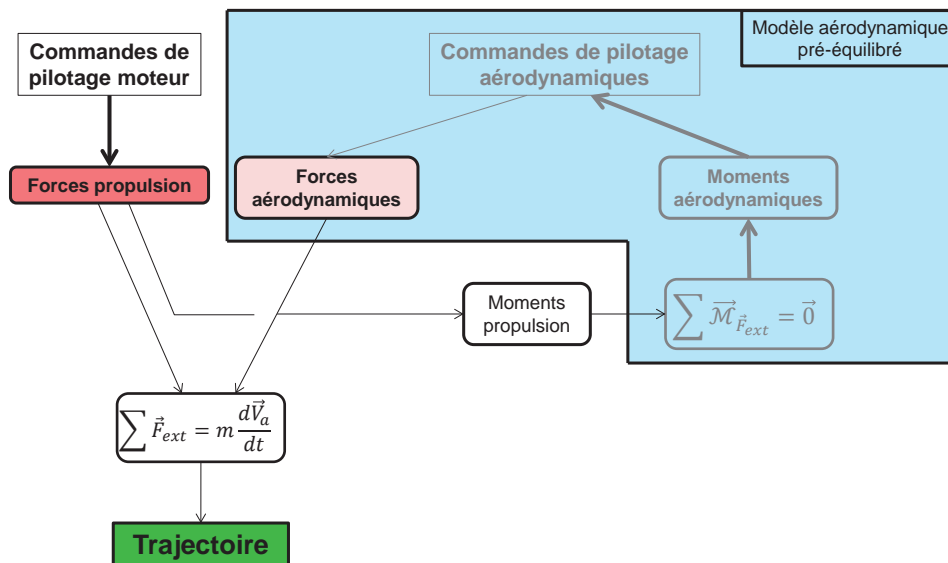


FIGURE 1.12 – Schéma général simplifié du Modèle Perfo avec aérodynamique pré-équilibrée.

- Avantages de ce pré-équilibrage : ce modèle est aussi plus rapide puisqu'il ne nécessite pas de boucle d'équilibrage du moment global. Il n'est appelé qu'une seule fois pour fournir les trois forces aérodynamiques.
- Inconvénient de ce pré-équilibrage : ce modèle reste complexe, car il rend nécessaire la modélisation des forces aérodynamiques en fonction des moments de propulsion. Comme nous allons le voir dans la section suivante, il est possible de simplifier encore ce modèle.

### 1.2.4.3 Simplification 2 : équilibrage moyen

La seconde simplification consiste à établir un équilibrage moyen forfaitaire (cf. figure 1.13). Les coefficients de forces aérodynamiques produits par ce modèle correspondent à un équilibrage moyen établi dans des conditions représentatives du vol, ou tout au moins des trajectoires avec lesquelles on va déterminer les performances de l'avion qui nous intéressent. Cette solution revient à imposer des valeurs constantes pour les commandes de vol  $\delta_p$ ,  $\delta_q$ ,  $iH$  et  $\delta_r$ .

- Avantages de cet équilibrage moyen : comme c'était déjà le cas pour le modèle Perfo pré-équilibré, on n'a pas besoin de connaître les coefficients du moment (aérody-

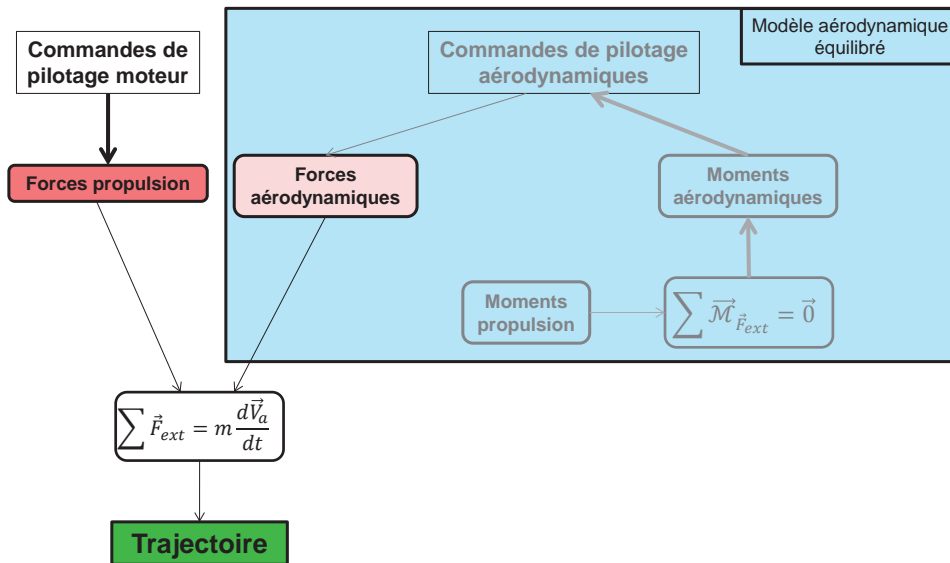


FIGURE 1.13 – Schéma général simplifié du Modèle Perfo avec aérodynamique équilibrée.

namique, propulsion) ni de boucler pour équilibrer en tangage, lacet ou roulis. De plus, ce modèle comprend moins de variables d'entrée.

- Inconvénients de cet équilibrage moyen : si le modèle aérodynamique ainsi construit est utilisé ailleurs qu'au point de vol où il a été équilibré, les coefficients aérodynamiques qu'il fournit sont entachés d'erreur. On peut limiter cette erreur en construisant plusieurs modèles équilibrés pour différents points de vol (par exemple Basse Vitesse (BV) et Grande Vitesse (GV) que nous introduisons dans la section 1.2.5).

## 1.2.5 Description du modèle actuel de performances des avions

### 1.2.5.1 Modèle aérodynamique actuel

Aujourd'hui, le modèle aérodynamique Perfo n'est pas dérivé du modèle aérodynamique QdV mais d'un modèle commun que l'on appellera modèle de la mécanique du vol. Il est construit en tirant parti des avantages des trois simplifications présentées à la section 1.2.4 et il est constitué de deux parties distinctes : une partie dite Basse Vitesse (BV) et une partie dite Grande Vitesse (GV). La partie BV s'intéresse au décollage, puis à la phase d'approche jusqu'à l'atterrissage et prend en compte les différentes configurations (*conf*) hypersustentées (position des bords et volets). Le reste du vol est modélisé par la partie GV qui ne considère que la configuration lisse (bords et volets rentrés). Ces deux modélisations ont des caractéristiques et des hypothèses simplificatrices différentes, détaillées ci-après :

- Simplification A : modèle équilibré en tangage en un point de vol représentatif de la partie BV et en un point de vol représentatif de la partie GV.
- Simplification B : modèle pré-équilibré en lacet et en roulis afin de traiter le cas de la panne moteur pour les parties BV et GV. Pour cela, il admet en entrée les poussées nettes  $F_{Ni}$ , ce qui est équivalent à donner les moments des forces de propulsion à un coefficient constant près. Si une panne moteur intervient, les équilibrages des moments ainsi que de la force latérale sont implicitement intégrés au modèle aérodynamique Perfo. L'évolution de l'avion sur sa trajectoire n'est plus représentée que par les deux équations de portance et de traînée correspondant aux projections du principe fondamental de la dynamique (1.1) sur les axes  $Gz_a$  et  $Gx_a$  du repère aérodynamique  $R_a$ . Dans la pratique, on détermine le coefficient de portance latérale de la dérive qui équilibre le moment dû à la dissymétrie de propulsion. De ce coefficient, on déduit par un modèle simplifié l'incrément de traînée résultant de l'équilibrage complet de l'avion en lacet et roulis.

Les équations (1.3a) et (1.3c) régissent le mouvement de l'avion :

$$\begin{aligned} \text{Selon } Gx_a : \quad m \frac{dV_a}{dt} &= -\frac{1}{2} \rho V_a^2 S C_x + F_{xprop} + F_{xsol} - mg_0 \sin \gamma. \\ \text{Selon } Gz_a : \quad 0 &= -\frac{1}{2} \rho V_a^2 S C_z + F_{zprop} + F_{zsol} + mg_0 \cos \gamma \cos \mu. \end{aligned}$$

Finalement le modèle aérodynamique actuel pour le calcul des performances des avions dépend des variables suivantes :

- pour les parties BV et GV : altitude-pression  $Z_p$ , vitesse aérodynamique  $V_a$ , position du centre de gravité sur l'avion  $XCG$ , température  $\Delta ISA$ , angle d'incidence  $\alpha$ , position sortie ou rentrée des trappes  $\delta_{trappes}$  et des trains d'atterrissage  $\delta_{trains}$ , position des aérofreins  $\delta_{AF}$ , poussées nettes des  $i$  moteurs  $F_{Ni}$ , configuration  $conf$  de l'aile ;
- uniquement pour la partie BV : état de la piste  $r_s$ , hauteur du centre de gravité par rapport à la piste  $h$ , taux de compression  $PRF$ , pente par rapport à la masse d'air  $\gamma$  ;
- uniquement pour la partie GV : masse  $m$ , nombre de Reynolds  $Re$ , spoilers  $\delta_s$ .

Les entrées et sorties du modèle Perfo actuel sont présentées dans la figure 1.14.

### 1.2.5.2 Analyse critique de la modélisation des performances des avions actuelle

L'avion est guidé sur sa trajectoire, c'est-à-dire qu'on impose la valeur de certaines variables d'état afin de résoudre les équations (1.3a) et (1.3c).

- Avantages de cette modélisation : l'utilisation du modèle QdV complet qui dépend directement de la position des commandes de vol  $\delta_x, \delta_p, \delta_q, iH, \delta_r$  et  $\delta_f$  est peu

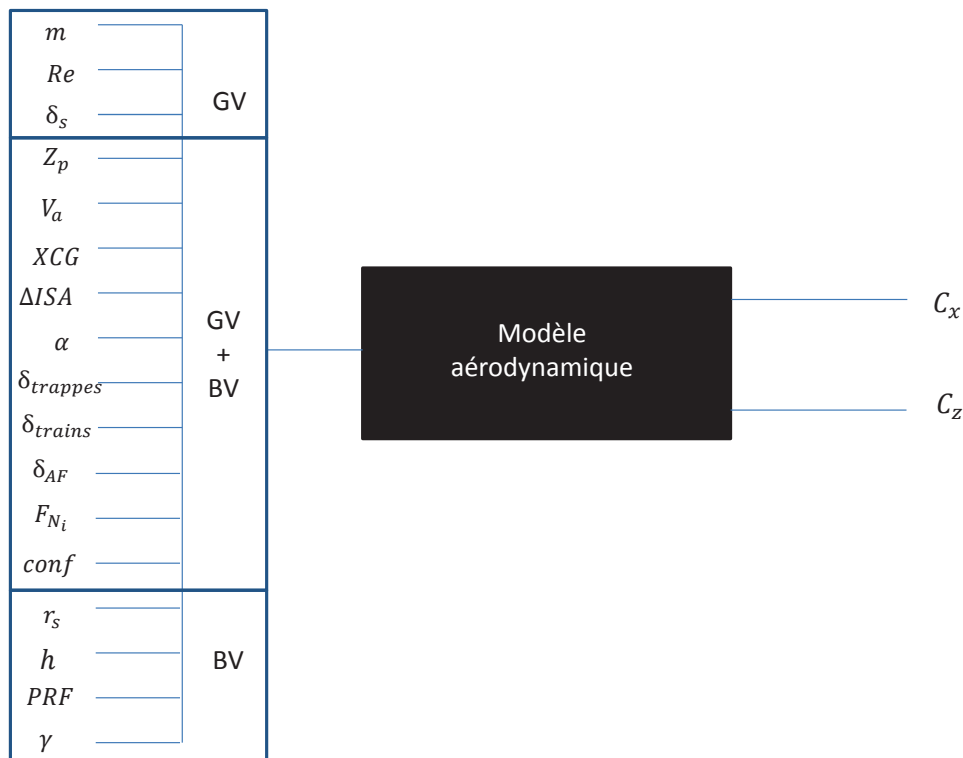


FIGURE 1.14 – Entrées/sorties du modèle aérodynamique Perfo actuel.

judicieuse pour le calcul de performances des avions pour plusieurs raisons :

- elle nécessite la mise en œuvre de modèles complexes qui dépendent de nombreuses variables telles que les différents moments (aérodynamique ou moteur), la réaction du sol, rendant difficile son recalage. La complexité de ces modèles se justifie dans le cadre d'une représentation complète du mouvement de l'avion (et en particulier de ses mouvements autour de son centre de gravité) mais elle ne permet pas forcément d'augmenter la précision en ce qui concerne le calcul de la trajectoire.
- le calcul de l'équilibrage du moment cinétique (1.2) implique de nombreuses itérations et donc plusieurs appels au modèle qui sont susceptibles d'augmenter significativement le temps de calcul.
- nous sommes amenés à calculer de nombreuses grandeurs intermédiaires inutiles pour des calculs de performances (position des commandes, vitesses et accélérations angulaires...).

Au contraire, les calculs de performances sont réalisés en utilisant des modèles spécifiques (aérodynamique, propulsion, moteur) qui sont des modèles plus simples que ceux utilisés par le modèle QdV (moins coûteux en temps de calcul et moins lourds en taille mémoire).

- Inconvénients de cette modélisation :
  - les sorties du modèle aérodynamique sont la traînée et la portance équilibrées en ce sens que le moment aérodynamique équilibre le moment de propulsion. Cette hypothèse n'est pas valide au sol (l'équilibrage en tangage et roulis est assuré par la réaction du sol et l'équilibrage en lacet est assuré par l'aérodynamique), ainsi que lors des phases transitoires telles que la rotation au décollage ou à l'atterrissage. Dans ces phases de vol, la représentativité du modèle actuel Perfo n'est donc pas garantie.
  - en vol, la traînée et la portance ont été calculées en supposant que le moment aérodynamique équilibre le moment des moteurs en un point moyen de la croisière. Lorsqu'on s'éloigne de ce point, des termes correctifs sont ajoutés (effet de centrage par exemple). Dans la section 1.3.1, on montre comment éviter ce type de correction en apportant des améliorations à cette modélisation.
  - chaque spécialité (QdV et Perfo) s'occupe séparément de recalibrer son modèle au moyen d'essais en vol spécifiques, ce qui entraîne des divergences entre ces deux disciplines.

Dans cette thèse, nos efforts se concentrent sur la construction d'un modèle aérodynamique Perfo par réduction du modèle aérodynamique des QdV après simplification physique. Actuellement, les modèles aérodynamiques Perfo et QdV sont dérivés manuellement du modèle de mécanique du vol. La construction manuelle (par connaissance métier) du modèle aérodynamique Perfo est distincte du processus manuel de dérivation du modèle aérodynamique de QdV, ce qui implique qu'aucune donnée n'est partagée entre ces deux disciplines. La construction automatique à partir du modèle QdV proposée dans cette thèse permet de garder un socle commun entre les deux domaines et de simplifier les processus de construction et de recalage par essais en vol qui pourront être communs aux deux disciplines.

## 1.3 Vers un nouveau modèle de performance des avions

### 1.3.1 Simplifications en vue de l'obtention du futur modèle aérodynamique de performances des avions

Plusieurs cas sont à considérer suivant les forces qui entrent en jeu.

1. Le cas le plus simple est la croisière tous moteurs en fonctionnement. Les forces et moments qui interviennent sont dus aux moteurs, au poids et à l'aérodynamique. Le moment associé au poids appliqué au centre de gravité est nul comme nous l'avons déjà vu. Les poussées des moteurs sont symétriques donc leurs moments autour des axes  $Gx_a$  et  $Gz_a$  sont nuls. Par conséquent, les équations (1.4a) et (1.4c) impliquent que les moments aérodynamiques associés sont nuls. De plus, le vol est supposé sy-

métrique et rectiligne (dérapiage  $\beta$  et roulis  $\mu$  nuls). Il en résulte que les braquages des gouvernes de gauchissement  $\delta_p$  et de direction  $\delta_r$  sont nuls. Il ne reste donc plus qu'à équilibrer le moment de tangage dû aux poussées moteur. Cet équilibrage est réalisé sur l'avion grâce au calage du plan horizontal  $iH$  en fixant la gouverne de profondeur  $\delta_q$  à 0. Cet équilibrage en tangage n'interagit ni sur les équilibrages en roulis et lacet (1.4a) et (1.4c), ni sur l'équilibre des forces latérales (1.3b). Il ne reste plus que les relations d'équilibre suivant  $Gx_a$  et  $Gz_a$ .

En résumé, après avoir supposé ce qui suit :

– poussées symétriques :  $\mathcal{M}_{F_{xprop}} = \mathcal{M}_{F_{zprop}} = 0$ ,

– vol symétrique et rectiligne :  $\beta = \mu = 0$ ,

le système simplifié d'équations est le suivant :

$$\Rightarrow \begin{cases} m \frac{dV_a}{dt} &= -\frac{1}{2} \rho V_a^2 S C_x + F_{xprop} - mg_0 \sin \gamma, \\ -m V_a \omega_2 &= -\frac{1}{2} \rho V_a^2 S C_z + F_{zprop} + mg_0 \cos \gamma, \\ 0 &= \frac{1}{2} \rho V_a^2 S l C_l, \\ 0 &= \frac{1}{2} \rho V_a^2 S l C_m + \mathcal{M}_{F_{yprop}}, \\ 0 &= \frac{1}{2} \rho V_a^2 S l C_n. \end{cases}$$

- Un autre cas important est celui de la panne moteur en vol. Les forces et moments en jeu sont toujours dus aux poussées moteurs, au poids et à l'aérodynamique, mais les poussées moteurs sont cette fois dissymétriques. L'équilibre des moments autour de  $Gz_a$  conduit donc à introduire un moment aérodynamique de lacet au moyen de la gouverne de direction  $\delta_r$ . Ce faisant, on génère une force aérodynamique latérale et un moment aérodynamique de roulis. L'équilibre des moments autour de  $Gx_a$  est alors assuré par une mise en dérapage ( $\beta$  non nul) qui grâce à l'effet de dièdre autour de la voilure, induit un moment de roulis  $\frac{1}{2} \rho V_a^2 S l C_l$ . Ce dernier vient s'opposer au moment dû au braquage de la direction sans qu'il soit nécessaire de faire appel au braquage des ailerons  $\delta_p$  qui doit rester au neutre ( $\delta_p = 0$ ) afin de conserver une manœuvrabilité maximale. Enfin, les forces aérodynamiques et de propulsion latérale induites par le braquage de la direction sont équilibrées par une composante latérale du poids obtenue au moyen d'une prise de roulis ( $\mu$  non nul). Le moment de tangage autour de  $Gy_a$  dû à la poussée des moteurs est quant à lui équilibré par un moment de tangage aérodynamique, généré par le braquage du plan horizontal  $iH$ , la commande de profondeur restant là encore au neutre ( $\delta_q = 0$ ).

Sous les hypothèses suivantes :

– poussées dissymétriques  $\delta_{x\text{ gauche}} \neq \delta_{x\text{ droite}}$ ,

$$- \delta p = \delta q = 0,$$

le système simplifié à résoudre est alors :

$$\Rightarrow \begin{cases} 0 = -\frac{1}{2}\rho V_a^2 S C_x + F_{xprop} - mg_0 \sin \gamma, \\ 0 = \frac{1}{2}\rho V_a^2 S C_y + F_{yprop} + mg_0 \cos \gamma \sin \mu, \\ 0 = -\frac{1}{2}\rho V_a^2 S C_z + F_{zprop} + mg_0 \cos \gamma \cos \mu, \\ 0 = \frac{1}{2}\rho V_a^2 S l C_l + \mathcal{M}_{F_{xprop}}, \\ 0 = \frac{1}{2}\rho V_a^2 S l C_m + \mathcal{M}_{F_{yprop}}, \\ 0 = \frac{1}{2}\rho V_a^2 S l C_n + \mathcal{M}_{F_{zprop}}. \end{cases}$$

3. Le troisième cas important est celui du roulement, dans lequel les forces et moments dus à la réaction du sol doivent être pris en compte. Les poussées sont symétriques sauf si une panne moteur a lieu. Concernant le tangage, les positions des gouvernes  $\delta q$  et  $iH$  sont imposées. L'équilibre est donc obtenu par la réaction des trains sur l'axe  $Gy_a$ . On notera que cette dernière est liée aux réactions des trains dans le plan horizontal (cf. section 1.2.2.3) par un modèle spécifique  $f_{frott}$ .

Avec les hypothèses qui suivent :

- poussées symétriques  $\mathcal{M}_{F_{xprop}} = \mathcal{M}_{F_{zprop}} = 0$ ,
- $\delta q, iH$  imposées,
- $\chi = 0, \mu = 0$ ,

le système simplifié d'équations est :

$$\Rightarrow \begin{cases} 0 & = \frac{1}{2}\rho V_a^2 S l C_m + \mathcal{M}_{F_{yprop}} + \mathcal{M}_{F_{y_{sol}}}, \\ m \frac{dV_a}{dt} & = -\frac{1}{2}\rho V_a^2 S C_x + F_{xprop} + F_{x_{sol}} - mg_0 \sin \gamma, \\ 0 & = -\frac{1}{2}\rho V_a^2 S C_z + F_{zprop} - F_{z_{sol}} + mg_0 \cos \gamma. \end{cases}$$

Ces calculs d'équilibrage des moments doivent être réalisés en chacun des points de la trajectoire mais restent malgré tout plus rapides que le calcul complet réalisé avec le modèle QdV puisqu'ils ne prennent pas en compte les accélérations angulaires. En dehors du troisième cas, on note que les commandes de pilotage  $\delta p, \delta q, iH$  et  $\delta r$  ne sont plus que des intermédiaires de calcul. Le schéma de principe du modèle correspondant est décrit dans la figure 1.11.

Aucune hypothèse n'a été faite concernant le modèle aérodynamique. Ce dernier peut donc être strictement identique à celui utilisé dans le modèle QdV.

### 1.3.2 Futur modèle de performances des avions

Les trois équations suivantes permettent le calcul de la trajectoire d'un avion :

$$\text{Selon } Gx_a : \quad m \frac{dV_a}{dt} = -\frac{1}{2} \rho V_a^2 S C_x + F_{xprop} + F_{xsol} - mg_0 \sin \gamma.$$

$$\text{Selon } Gz_a : \quad 0 = -\frac{1}{2} \rho V_a^2 S C_z + F_{zprop} - F_{zsol} + mg_0 \cos \gamma.$$

$$\text{Autour de } Gy_a : \quad 0 = \frac{1}{2} \rho V_a^2 S l C_m + \mathcal{M}_{F_{yprop}} + \mathcal{M}_{F_{ysol}}.$$

Le futur modèle aérodynamique est plus complexe que celui du modèle actuel Perfo, afin de représenter l'avion avec une granularité plus fine et ainsi être plus précis dans les calculs de performances de l'avion. Ce modèle est non équilibré en tangage. Le moment aérodynamique doit donc être équilibré à chaque instant du calcul d'une trajectoire. Concernant les mouvements en roulis et lacet, ce modèle est pré-équilibré pour la partie Basse Vitesse (BV) (cf. section 1.2.4.2) et équilibré en moyenne en un point représentatif de la partie Grande Vitesse (GV) (cf. section 1.2.4.3). Les modèles aérodynamiques du moment de tangage, de la force de portance et de la force de traînée sont donc nécessaires pour calculer les performances de l'avion. Les variables d'entrée de ces modèles aérodynamiques utilisées pour le futur calcul des performances sont résumées dans la figure 1.15.

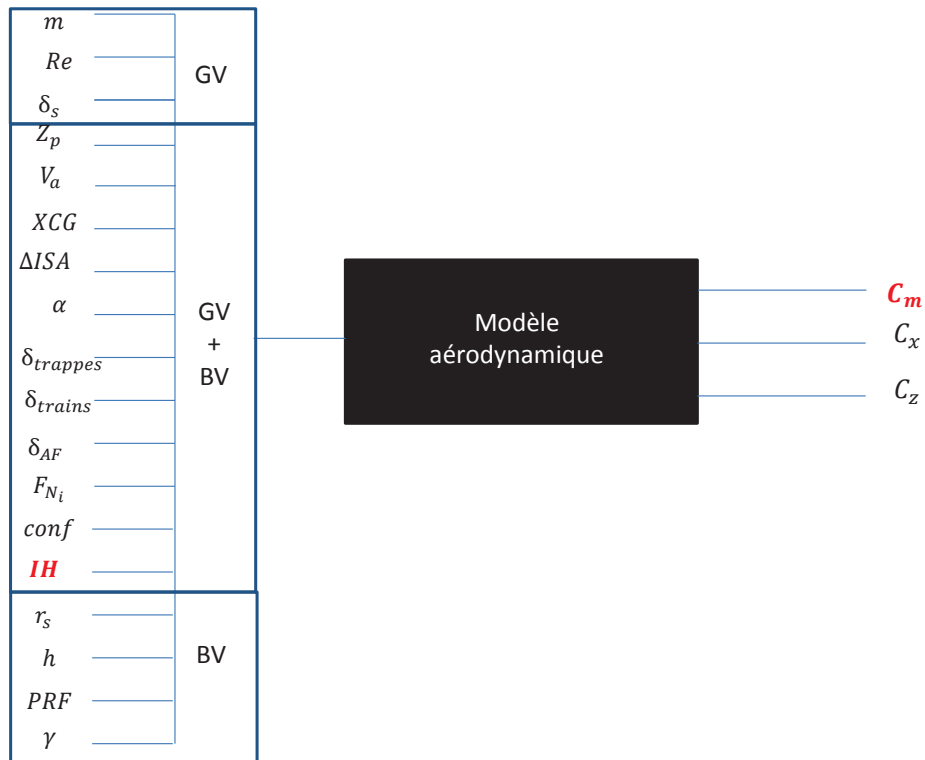


FIGURE 1.15 – Entrées/sorties du modèle aérodynamique QdV après simplification.

Le modèle aérodynamique QdV ainsi établi après simplification physique (en fixant à

zéro les variables d'entrée non pertinentes et en éliminant certains degrés de liberté) reste trop complexe pour être utilisé tel quel pour les calculs de performances, le modèle Perfo correspondant présentant l'inconvénient de nécessiter une boucle de recherche de zéro pour résoudre l'équilibre  $\sum \vec{M}_{\vec{F}_{ext}} = \vec{0}$  et donc plusieurs appels au modèle aérodynamique. Ce modèle aérodynamique devra donc être réduit pour diminuer le temps de calcul et satisfaire les contraintes Perfo qui sont introduites dans la section suivante.

## 1.4 Besoins et enjeux industriels de la réduction du modèle de performances des avions

### 1.4.1 Structure et composition des modèles de qualités de vol et de performances des avions

Les modèles Perfo et QdV permettent de décrire le mouvement de l'avion avec des niveaux de granularité différents. Ils sont tous deux construits comme un assemblage de plusieurs sous-modèles physiques (modèle aérodynamique, modèle de propulsion, modèle de l'atmosphère) qui sont couplés aux équations de la mécanique du vol. Cette construction se fait par connaissance physique mais aussi par similitude avec les avions déjà existants, puis par recalage grâce aux données réelles fournies par les essais en vol.

Ces sous-modèles physiques peuvent être eux-même vus comme un assemblage de fonctions numériques comprenant  $d$  variables d'entrée et une sortie. La méthode actuelle d'Airbus pour modéliser une telle fonction de plusieurs variables consiste à définir sa valeur en chaque point d'une grille couvrant le domaine de définition et à l'évaluer ailleurs au moyen d'une méthode d'interpolation multi-linéaire (cf. figure 1.16).

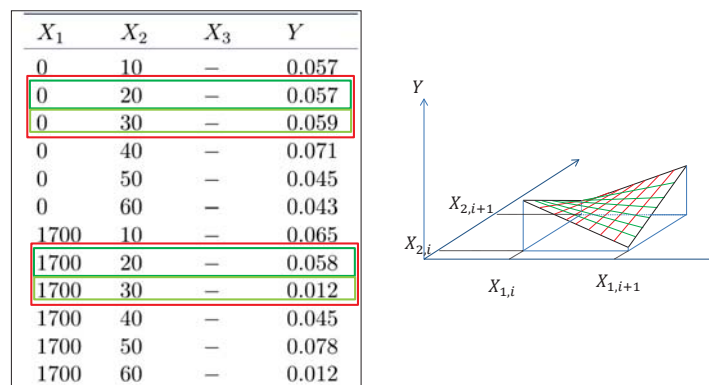


FIGURE 1.16 – Interpolation multi-linéaire.

Cette méthode s'avère rapidement inapplicable opérationnellement dès lors que le nombre de variables dépasse quelques unités (typiquement 3 ou 4). En effet, la taille mémoire requise et le temps de calcul pour interpoler deviennent vite prohibitifs, ces deux paramètres augmentant exponentiellement avec la dimension. Pour un vecteur d'entrée

de dimension  $d$  avec  $k_i$  valeurs assignées à la variable d'entrée  $x_i$ , on se retrouve avec un ensemble de  $\prod_{i=1}^d k_i$  valeurs à stocker en mémoire. Chaque variable  $x_i$  est en effet découpée en un ensemble de  $k_i$  valeurs discrètes :  $\{x_{1,1}, \dots, x_{1,k_1}\}, \{x_{2,1}, \dots, x_{2,k_2}\}, \dots, \{x_{d,1}, \dots, x_{d,k_d}\}$ . La fonction est définie sur une réunion de pavés élémentaires :

$$[x_{1,j_1}, x_{1,j_1+1}] \times [x_{2,j_2}, x_{2,j_2+1}] \times \dots \times [x_{d,j_d}, x_{d,j_d+1}]$$

avec  $j_1 \in \{1, \dots, k_1 - 1\}, j_2 \in \{1, \dots, k_2 - 1\}, \dots, j_d \in \{1, \dots, k_d - 1\}$ . Pour calculer la valeur de la fonction en un point qui n'appartient pas à la table numérique, la connaissance des valeurs en chacun des  $2^d$  sommets correspondant au pavé auquel ce point appartient est requise. Pour trouver ces sommets, on fait appel à un algorithme dichotomique par dimension  $i$ . Le nombre d'opérations de cette méthode est en  $\mathcal{O}(\log_2(k_i))$  pour la dimension  $i$ . Une fois les  $2^d$  sommets déterminées, la valeur du nouveau point est calculée par interpolation multi-linéaire. Une interpolation multi-linéaire en dimension  $d$  revient à effectuer  $2^d - 1$  interpolations linéaires uni-dimensionnelles. La complexité du calcul est exponentielle en  $\mathcal{O}(2^d)$ .

En pratique, pour gérer cette complexité, les modèles décrivant les caractéristiques aérodynamiques, propulsives, ou géométriques de l'avion sont constitués d'un assemblage de fonctions tabulées à une, deux, trois ou plus rarement quatre variables d'entrée. Leur structure peut être décrite comme une structure hiérarchique multi-niveaux. Un modèle est défini comme hiérarchique multi-niveaux quand les sous-modèles sont exécutés de façon séquentielle et les entrées du sous-modèle effectuant les calculs plus tard dans le processus séquentiel sont les sorties du sous-modèle qui le précèdent immédiatement dans cette suite. Par exemple, les différentes configurations de l'avion peuvent être comprises comme la position de toutes ses parties mobiles (position des trains, braquages des becs et volets,...). Le modèle Perfo aérodynamique peut être séparé en une série de sous-modèles représentant les effets de chacun de ces composants. Le calcul de l'effet des volets pourrait aussi par exemple être calculé avant l'effet des trains d'atterrissage, lequel prendrait donc en entrée le résultat du modèle de l'effet des volets. Sur la figure 1.17, le modèle global  $F$  décrivant par exemple un coefficient aérodynamique, est décomposé en sous-modèles notés  $F_{i,j}$ . Chacune de ces boîtes représente un sous-système physique du système global (par exemple l'effet de sortie des trains d'atterrissage sur le coefficient de traînée ou sur le coefficient de portance).

## 1.4.2 Construction actuelle des modèles de qualités de vol et de performances des avions

Les modèles QdV et Perfo sont construits à partir de données issues de différentes sources.

- Modèle aérodynamique :

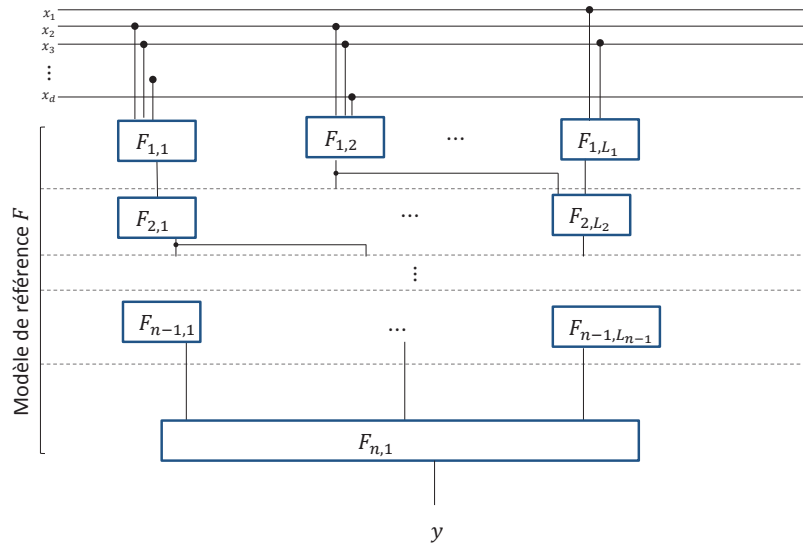


FIGURE 1.17 – Structure hiérarchique multi-niveaux.

- essais en soufflerie : de ces expériences sont tirées des informations liant les caractéristiques de l'air et de l'avion qui permettent notamment de mesurer les coefficients de trainée et portance,
- mécanique des fluides numérique (Computational Fluid Dynamic (CFD)) : on intègre les équations de Navier Stokes ou d'Euler pour modéliser l'écoulement autour d'un profil d'aile ou même de l'avion complet.
- Modèle moteur :
  - modèle fourni par les industriels qui fabriquent les moteurs,
  - bancs d'essai moteur : ces bancs permettent de mesurer les performances des moteurs dans différentes conditions.
- Modèles analytiques issus des lois de la physique (modèle de gravité, modèle atmosphérique).
- Similitude avec les modèles existants des avions déjà en service.
- Essais en vol spécifiques.

Aujourd'hui, à partir des données précédentes, les ingénieurs des départements de qualités de vol et de performances des avions construisent en parallèle de façon non automatisée un modèle spécifique adapté à leurs besoins différents. La représentativité du modèle QdV est plus large que celle du modèle Perfo (cf. sections 1.2.3 et 1.2.4), à l'exception de la trainée de spray sur piste mouillée qui n'est pas prise en compte dans le modèle QdV. Par ailleurs, les constructions de ces deux modèles ne sont pas rigoureusement indépendantes : en particulier le coefficient de trainée du modèle QdV est issu du modèle Perfo.

Le processus de construction de ces deux modèles décrit ci-dessus s'avère non optimum pour les raisons suivantes :

- on répète le même type d'opération à deux reprises pour obtenir les modèles QdV et Perfo, comme par exemple les essais en vol qui ne sont pas partagés ;
- les deux modèles obtenus ne sont pas cohérents entre eux ;
- le modèle Perfo est destiné à être représentatif de l'avion dans les phases de vol stabilisées. Or, on a de plus en plus besoin d'étendre sa validité au delà de ces phases-là pour le développement de nouvelles fonctions embarquées destinées à améliorer les performances de l'avion (trajectoires d'évitement de relief en cas de panne moteur, gestion du freinage,...).

À partir du modèle Perfo ainsi obtenu, un nouveau modèle est dérivé manuellement afin d'être embarqué dans le cockpit de l'avion. Il doit respecter des contraintes plus fortes en termes de temps de calcul et de taille mémoire.

Dans la section 1.4.3, nous proposons des améliorations potentielles de ces deux processus : dérivation du modèle Perfo à partir du modèle QdV et réduction du modèle Perfo pour l'embarquer dans le cockpit de l'avion.

### 1.4.3 Améliorations potentielles

Afin de réduire les coûts de calcul et les délais de production des modèles, nos efforts se concentrent sur la dérivation directe et automatique du modèle aérodynamique Perfo à partir du modèle aérodynamique QdV de référence. Pour cela, un premier effort est nécessaire afin que le modèle QdV englobe totalement le modèle Perfo. Cette étape ne sera pas abordée dans le cadre de cette thèse, mais est indispensable pour que les méthodes développées par la suite soient applicables.

Cette méthode de dérivation du modèle aérodynamique Perfo à partir du modèle aérodynamique QdV permettra de prendre en compte plus de variables d'entrée et des sous-modèles plus complexes et de mieux représenter ainsi l'avion réel (cf. section 1.3.2). L'objectif du processus de dérivation à mettre en place est donc de maximiser le périmètre du socle commun entre ces deux modèles tout en conservant un niveau de complexité du modèle Perfo compatible avec les exigences liées à ses applications, en particulier en terme de temps de calcul. En effet, le modèle Perfo est utilisé intensivement au quotidien par les ingénieurs du bureau d'étude pour, d'une part l'établissement du manuel de vol et, d'autre part, l'optimisation de l'utilisation opérationnelle de l'avion. Cette utilisation intensive impose d'être en mesure de générer en quelques secondes une trajectoire complète de l'avion, ce qui induit un besoin de réduction en temps de calcul du modèle QdV tout en garantissant une précision satisfaisante. Une première étape de simplification physique a été décrite dans ce chapitre qui a permis d'obtenir le modèle aérodynamique QdV de référence (cf. figure 1.15). La dérivation d'un modèle Perfo opérationnel de ce modèle

QdV nécessitera donc une réduction du modèle QdV en temps de calcul.

Une fois le modèle Perfo opérationnel obtenu, une nouvelle réduction est nécessaire en vue de l'embarquer dans le cockpit de l'avion. En effet, le besoin des compagnies aériennes de réduire leurs coûts (consommation de carburant, maintenance, etc) [67], l'intérêt fort de limiter l'impact environnemental du transport aérien (nuisance sonore, qualité de l'air, réchauffement climatique) [80] et l'augmentation continue du trafic aérien ont généré de nouveaux défis et l'utilisation croissante de fonctions dans le cockpit de l'avion pour aider le pilote en temps réel. Par exemple, trouver les meilleures trajectoires [79] permet de réduire la consommation de carburant et de réduire aussi la nuisance sonore proche des aéroports. Ces modèles embarqués doivent améliorer en temps réel (rapidité du modèle Perfo) les performances de l'avion tout en garantissant la sécurité du vol (précision). Ce besoin d'embarquer de plus en plus de données dans le cockpit de l'avion mène également à une diminution de la taille mémoire disponible.

Les figures 1.18 et 1.19 présentent de façon synthétique le processus actuel et le processus futur de construction des modèles Perfo.

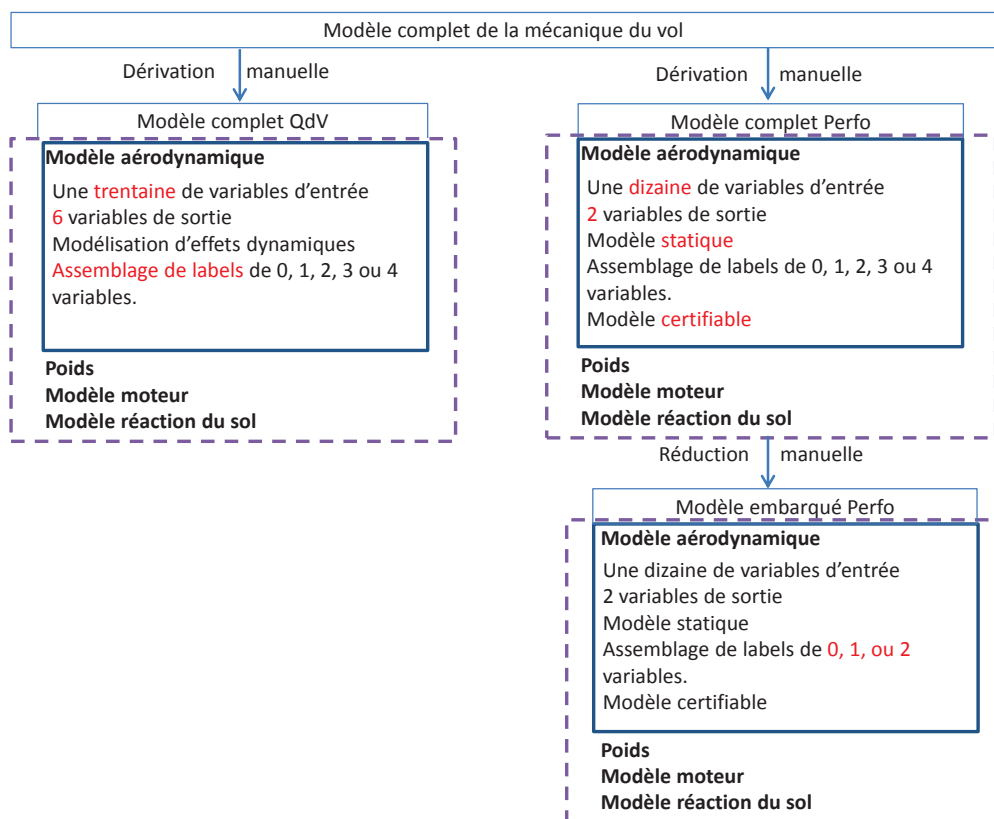


FIGURE 1.18 – Processus actuel de construction des modèles Perfo.

1.4. Besoins et enjeux industriels de la réduction du modèle de performances des avions

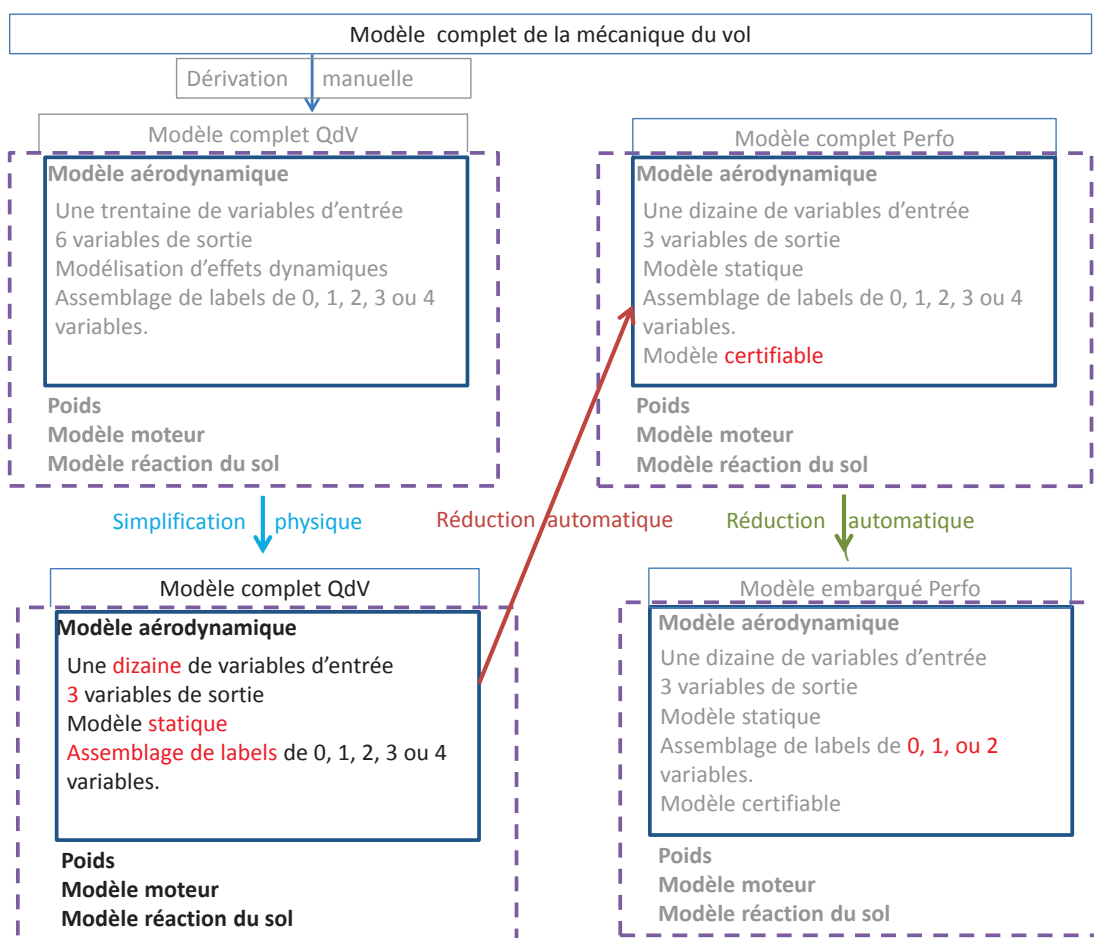


FIGURE 1.19 – Processus futur de construction des modèles Perfo.

# Chapitre 2

## État de l'art critique : modèles réduits, méthodes d'échantillonnage et de validation

### 2.1 Introduction

On introduit tout d'abord quelques notations que nous utilisons tout au long de ce chapitre.

À partir d'un modèle numérique de référence  $F$ ,

$$\begin{aligned} F : \quad \Omega \subseteq \mathbb{R}^d &\longrightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2, \dots, x_d) &\longmapsto F(\mathbf{x}), \end{aligned}$$

nous avons mis en évidence dans le chapitre 1 deux besoins de réduction de modèles.

Afin de répondre à ces exigences, un modèle réduit, noté  $\tilde{F}$  est construit,

$$\begin{aligned} \tilde{F} : \Omega \subseteq \mathbb{R}^d &\longrightarrow \mathbb{R} \\ \mathbf{x} &\longmapsto \tilde{F}(\mathbf{x}). \end{aligned}$$

Les plans d'expérience sont des moyens simples et efficaces pour connaître les valeurs de la fonction sur son domaine de définition. On note  $[\bar{x}]$  la matrice contenant les  $N$  valeurs que prend le vecteur d'entrées  $\mathbf{x}$  :

$$[\bar{x}] = (x_{ij})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq d}} = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \vdots & \vdots \\ x_{N1} & \cdots & x_{Nd} \end{pmatrix},$$

que l'on peut représenter également comme suit :

$$[\bar{x}] = (\mathbf{x}_i)_{1 \leq i \leq N} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix} \text{ avec } \forall i \in \{1, \dots, N\}, \mathbf{x}_i = (x_{i1}, \dots, x_{id}).$$

On lui associe le vecteur de sortie  $\mathbf{Y}$  par la fonction  $F$  :

$$\mathbf{Y} = \begin{pmatrix} F(\mathbf{x}_1) \\ \vdots \\ F(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix},$$

et le vecteur de sortie  $\tilde{\mathbf{Y}}$  par la fonction  $\tilde{F}$  :

$$\tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{F}(\mathbf{x}_1) \\ \vdots \\ \tilde{F}(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_N \end{pmatrix}.$$

Ce chapitre a pour but de donner une vue d'ensemble des méthodes de réduction de modèles depuis la création du plan d'apprentissage jusqu'à la validation du modèle réduit. Nous proposons tout d'abord une méthodologie générale détaillant les principales étapes de cette réduction dans la section 2.2. Cette méthodologie est directement appliquée à nos deux cas particuliers. Les différents modèles réduits qui ont été envisagés sont décrits dans la section 2.3. Puis, les différents plans d'expérience, ainsi que plusieurs mesures mathématiques permettant de caractériser certaines de leurs propriétés en terme de remplissage de l'espace en particulier, sont décrits dans la section 2.4. Enfin, différentes mesures de l'erreur d'approximation et des méthodes de validation sont décrites dans la section 2.5. Dans la section 2.6, les choix retenus et appliqués dans les chapitres 3 et 4 sont justifiés.

## 2.2 Principales étapes de construction d'un modèle réduit

Le problème de réduction revient à approcher au mieux le modèle de référence sur son domaine de définition tout en respectant certaines contraintes. Mathématiquement, ce problème se formalise généralement comme un problème d'optimisation. Il se découpe en plusieurs étapes énumérées dans la table 2.1.

Étape 1	Identification du modèle de référence à réduire
Étape 2	Spécification des caractéristiques du modèle réduit en fonction de son application envisagée
Étape 3	Choix des critères d'ajustement et de validation du modèle réduit
Étape 4	Choix de la structure du modèle réduit
Étape 5	Choix de la méthode d'échantillonnage
Étape 6	Choix de la méthode de résolution et construction du modèle réduit
Étape 7	Validation du modèle réduit

TABLE 2.1 – Description du processus de réduction.

Deux besoins de réduction ont été introduits à la section 1.4.3. Nous les rappelons brièvement dans la table 2.2.

Besoin 1	Besoin 2
Réduction du modèle aérodynamique QdV pour répondre aux exigences Perfo	Réduction du modèle Perfo utilisé au sol afin de l'embarquer dans le cockpit

TABLE 2.2 – Deux besoins de réduction.

### Étape 1 : Identification du modèle de référence à réduire

Concernant la dérivation du modèle aérodynamique Perfo à partir du modèle aérodynamique QdV (besoin 1), le modèle de référence à été introduit dans le chapitre 1 (cf. figure 1.15).

En ce qui concerne la dérivation d'un modèle Perfo embarqué (besoin 2), nous n'avons pas encore défini le modèle de référence que nous cherchons à réduire. Un modèle embarqué dans le cockpit de l'avion permet de calculer en temps réel une caractéristique Perfo, comme la consommation de carburant, la distance d'atterrissage/de décollage... afin d'aider le pilote à adapter sa vitesse, son altitude, etc... selon les conditions réelles du

vol. Nous nous concentrons dans cette thèse sur le "Fuel Consumption Model (FCM)" qui permet de prédire la consommation de carburant à bord de l'avion. Ce modèle est composé de plusieurs sous-modèles, comme le modèle aérodynamique et le modèle moteur. Il contient 6 entrées et 1 sortie et possède une structure hiérarchique multi-niveaux (cf. figure 1.17). Cet exemple est représentatif des modèles Perfo embarqués dans le cockpit et constitue le modèle de référence dans toute la suite. Les caractéristiques des modèles de référence sont présentées dans la table 2.3.

	Besoin 1	Besoin 2
Nombre de variables d'entrée	GV <sup>1</sup> : 14 / BV <sup>2</sup> : 15	6
Nombre de variables de sorties	3	1

TABLE 2.3 – Caractéristiques du modèle de référence.

**Étape 2** : Spécification des caractéristiques du modèle réduit en fonction de son application envisagée

Besoin 1 : Afin d'estimer le besoin de réduction en temps de calcul, deux trajectoires similaires ont été simulées en faisant appel d'une part au modèle aérodynamique Perfo actuel et d'autre part au modèle aérodynamique QdV. Un pseudo-équilibre en lacet et roulis a été au préalable établi pour le modèle aérodynamique QdV. L'équilibre du moment aérodynamique de tangage est quant à lui effectué en ligne. Le calcul réalisé avec le modèle Perfo actuel nécessite 10 fois moins de temps CPU que le calcul réalisé avec ce modèle aérodynamique QdV. Le temps d'appel au modèle aérodynamique QdV doit donc être divisé par 10 environ pour retrouver approximativement le même temps d'exécution pour un calcul de trajectoire. La réduction se focalise donc principalement sur le temps de calcul d'un appel au modèle aérodynamique. De plus, le modèle Perfo qui fait appel au modèle aérodynamique doit être validé; cette étape de validation s'appuie sur la structure elle-même du modèle de performance qui est un assemblage de fonctions tabulées et physiques (cf. section 1.4.1) .

Besoin 2 : Selon la criticité du modèle embarqué, les résultats fournis aux pilotes doivent être compatibles avec les exigences du temps réel. Par ailleurs, de plus en plus de modèles étant embarqués dans le cockpit de l'avion, certains de ces modèles doivent être de taille minimale. Enfin ces modèles doivent être suffisamment précis pour être validés.

---

1. Grande Vitesse  
2. Basse Vitesse

Les contraintes des modèles réduits à construire sont résumées dans la table 2.4.

	Besoin 1	Besoin 2
<b>Contraintes liées au calculateur</b>		
Taille mémoire limitée	NS <sup>3</sup>	✓
Capacité de calculs des processeurs réduite	NS <sup>3</sup>	✓
Nombre de variables d'entrée des sous-modèles	1 à 4	1 à 2
<b>Contraintes liées à l'utilisation</b>		
Réduction du temps de calcul	✓	✓
<b>Contraintes liées à la validation</b>		
Modèle Certifiable	✓	✓
<b>Contraintes liées à l'industrialisation de la méthode</b>		
Robustesse de la méthode de réduction	✓	✓
Méthode de construction automatique	✓	✓
Simplicité de mise en oeuvre de la méthode	✓	✓

TABLE 2.4 – Contraintes des modèles réduits.

### Étape 3 : Choix des critères d'ajustement et de validation du modèle réduit

Lorsqu'on construit un modèle réduit, on souhaite contrôler l'erreur commise sur tout le domaine de définition  $\Omega$  de la fonction de référence. Le critère d'ajustement du modèle permet de mesurer l'écart entre le modèle réduit et le modèle de référence. Nous détaillons dans la section 2.5, les différents échantillons utilisés pour ajuster et valider le modèle réduit.

L'erreur absolue locale est la différence entre la valeur en un point donné du modèle référence  $F$  et du modèle réduit  $\tilde{F}$  suivant le schéma de la figure 2.1.

Elle s'écrit donc  $\forall \mathbf{x} \in \Omega$ ,

$$\epsilon_{\tilde{F}}(\mathbf{x}) = |F(\mathbf{x}) - \tilde{F}(\mathbf{x})|.$$

- Critère d'ajustement de la précision

Dans notre étude, le modèle de référence est connu de façon discrète sur un échan-

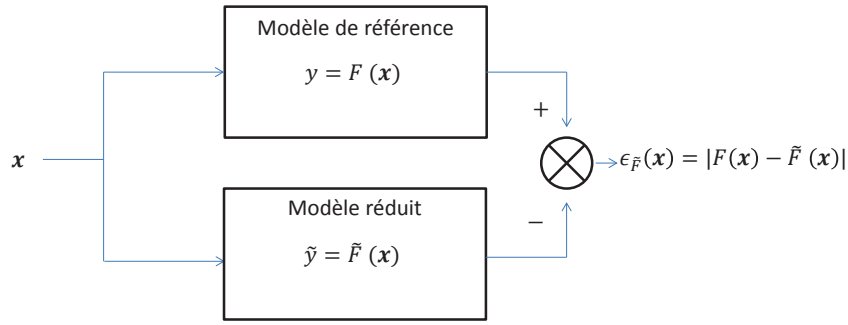


FIGURE 2.1 – Erreur absolue locale au point  $\mathbf{x}$ .

tillon  $[\bar{x}]$  et l'erreur globale du modèle réduit est estimée à partir des erreurs locales aux points  $\mathbf{x}_i$  du plan d'expérience ( $\forall 1 \leq i \leq N, \epsilon_i = |F(\mathbf{x}_i) - \tilde{F}(\mathbf{x}_i)|$ ) :

– erreur absolue moyenne :  $Err_{moy} = \frac{1}{N} \sum_{i=1}^N \epsilon_i$

– erreur quadratique :  $Err_{quad} = \sqrt{\frac{1}{N} \sum_{i=1}^N \epsilon_i^2}$

– erreur maximale :  $Err_{max} = \max_{i=1, \dots, N} (\epsilon_i)$

- Critères de validation

- Estimation de la précision : l'estimation d'une ou plusieurs de ces trois erreurs (moyenne, quadratique, maximale) est envisageable pour accepter ou non le modèle réduit construit. Ce choix se fait selon les besoins de l'utilisateur [17]. Dans cette étude, l'erreur maximale est préférée aux deux autres erreurs. En effet, lorsque l'erreur quadratique ou l'erreur absolue moyenne sont presque nulles, il n'est pas garanti que ce soit effectivement le cas sur tous les points du domaine. Dans notre cas, on ne se satisfait pas d'un modèle bon partout sauf en certains points. L'erreur maximale permet ainsi de s'assurer que le point le plus éloigné respecte la contrainte de précision.
- Comparaison de la taille mémoire du modèle réduit avec la taille mémoire du modèle de référence pour satisfaire le besoin 2.
- Comparaison du temps de calcul du modèle réduit avec le temps de calcul du modèle de référence pour satisfaire les besoins 1 et 2.

Les valeurs limites des critères de validation sont présentées dans la table 2.5.

Les méthodes de validation ainsi que l'estimation de l'erreur générée par le modèle réduit sur un plan d'expérience  $[\bar{x}]$  sont détaillées dans la section 2.5.

	Besoin 1	Besoin 2
Précision	Expertise	Expertise
Temps de calcul	10 fois plus rapide que le modèle de référence	Instantané (quelques secondes) à quasi-instantané (quelques minutes)
Taille mémoire	Pas de limitation	La plus petite possible

TABLE 2.5 – Valeurs limites des critères de validation.

#### Étape 4 : Choix de la structure du modèle réduit

La structure du modèle réduit peut être spécifiée *a priori* ou déterminée à partir des données d'apprentissage. On parle respectivement de modèles paramétriques ou non paramétriques. Un modèle non paramétrique peut contenir des paramètres et une structure adaptatives. Ils dépendent des informations disponibles sur le modèle de référence. Par exemple, un polynôme dont le degré et les coefficients sont ajustés en cours de construction du modèle réduit est un modèle non-paramétrique. Ainsi, lorsque le modèle de référence est peu connu, les modèles non-paramétriques permettent d'explorer sa structure. Les modèles paramétriques possèdent quant à eux un nombre fixe pré-déterminé de paramètres et reposent sur des hypothèses fortes posées *a priori* sur la structure du modèle réduit final. Dans le cas du polynôme, seuls les coefficients sont des paramètres ajustables, le degré est quant à lui fixé *a priori*.

#### Étape 5 : Choix de la méthode d'échantillonnage

Il existe différentes méthodes de génération du plan d'apprentissage ( $[\bar{x}_{app}], \mathbf{Y}_{app}$ ) qui garantissent notamment un bon remplissage de l'espace. Nous présentons dans la section 2.4 un état de l'art de générations de plan d'expérience optimaux selon différents critères que nous définissons.

#### Étape 6 : Choix de la méthode de résolution et construction du modèle réduit

On choisit la méthode de résolution selon les caractéristiques de la fonction de référence et du modèle réduit retenu ainsi que selon les critères d'ajustement.

Le problème est formulé comme un problème d'optimisation où l'on cherche à minimiser l'erreur commise par le modèle réduit ; l'algorithme le mieux adapté pour minimiser cette erreur dépend de la régularité de celle-ci [56]. Par exemple, si le modèle réduit choisi est linéaire par rapport à ses variables d'entrée, la méthode des moindres

carrés permet d'envisager une résolution basée sur les techniques d'algèbre linéaire. De manière plus générale, si la fonction à minimiser est régulière et convexe, l'algorithme BFGS [13, 30, 35, 76], les algorithmes à région de confiance [19], ainsi que l'algorithme de Levenberg-Marquardt [75] sont de bons candidats. Lorsque la dérivée de la fonction n'est pas accessible, les algorithmes de Nelder Mead [69] ou de Pattern Search [45] qui sont de bonnes alternatives sont des méthodes dites d'ordre zéro.

### Étape 7 : Validation du modèle réduit

Le modèle réduit construit peut ne pas être approprié, par exemple lorsque la structure du modèle  $\tilde{F}$  n'est pas adaptée ou bien lorsque la taille de l'échantillon d'apprentissage est trop faible.

Ainsi, si nous ne sommes pas en mesure d'atteindre la tolérance requise pour le critère choisi à l'étape 3, deux alternatives sont envisageables :

1. on recommence le processus à partir de l'étape 4 en choisissant une architecture de modèle réduit plus appropriée au vu des résultats obtenus ;
2. on ajoute des points dans le plan d'apprentissage et on boucle à l'étape 6.

Dans les paragraphes suivants, les modèles réduits à notre disposition, les plans d'expérience et leurs caractéristiques ainsi que les différentes méthodes de validation sont respectivement détaillés dans les sections 2.3, 2.4, 2.5.

## 2.3 Différents types de modèles réduits

### 2.3.1 Surfaces de réponses

#### Modèles polynomiaux d'ordre 1 ou 2

L'utilisation de modèles polynomiaux d'ordre 1 ou 2 remonte au début du XX<sup>ème</sup> siècle alors que les scientifiques avaient besoin d'outils pour modéliser des données empiriques dans les domaines de la biologie et de l'agriculture. En pratique, pour des expériences comportant des erreurs de mesures, et dont la génération de données est longue et coûteuse (en temps de calcul, financièrement...), les modèles linéaires ou quadratiques sont très utilisés [68, 20]. Les méthodes de surface de réponse (Response Surface Model (RSM)) faisant usuellement appel aux modèles polynômiaux d'ordre 1 ou 2 comprennent des techniques de planification d'expérience, des techniques d'analyse de régression et d'analyse de variance. Depuis une vingtaine d'années, les méthodes RSM ont été de plus en plus appliquées pour construire des approximations de modèles numériques.

Un modèle polynomial d'ordre 2 s'écrit :

$$\tilde{y} = c_0 + \sum_{1 \leq j \leq d} c_j x_j + \sum_{1 \leq j \leq k \leq d} c_{jk} x_j x_k, \quad (2.1)$$

où  $c_0$ ,  $c_j$  et  $c_{jk}$  sont les coefficients indéterminés du modèle. Si tous les coefficients  $c_{jk}$  sont nuls, le modèle polynomial est d'ordre 1. Soit  $\mathbf{c} = (c_0, c_1, \dots, c_{m-1})^T \in \mathbb{R}^m$  le vecteur des paramètres inconnus du modèle de surface de réponse d'ordre 1 ou 2 avec  $m = \frac{(d+1)(d+2)}{2}$  pour l'ordre 2 et  $m = d+1$  pour l'ordre 1. Le modèle polynomial (2.1) peut être écrit de façon synthétique :

$$\tilde{\mathbf{y}} = \mathbf{c}^T x_{\text{base}},$$

où  $x_{\text{base}} = (1, x_1, x_2, \dots, x_1^2, x_1x_2, x_1x_3, \dots, x_d^2)^T$  peut être interprété comme un vecteur de fonctions de base.

Les données observées sont substituées dans le modèle pour déterminer les paramètres inconnus. On obtient le système suivant :

$$[\mathbf{A}]\mathbf{c} = \tilde{\mathbf{Y}} \text{ où } \tilde{\mathbf{Y}} = \begin{pmatrix} \tilde{y}_1 \\ \vdots \\ \tilde{y}_N \end{pmatrix} \in \mathbb{R}^N \text{ et } [\mathbf{A}] = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d}^2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{N1} & x_{N2} & \cdots & x_{Nd}^2 \end{pmatrix} \in \mathbb{R}^{N \times m}.$$

Lorsque le nombre d'échantillons de la fonction  $F$  à approcher est égal au nombre de coefficients du modèle, on obtient un modèle interpolant. Mais il est également possible de choisir un nombre plus important d'échantillons afin de construire un modèle de régression. Les paramètres sont alors choisis dans ce cas-là par le critère des moindres carrés, c'est-à-dire qu'ils minimisent la fonction d'erreur quadratique.

### Avantages

- L'intérêt des modèles polynomiaux réside dans leur simplicité de construction et d'utilisation [34].
- Selon l'ordre du modèle, certains points caractéristiques, comme l'extremum d'une fonction quadratique, peuvent être obtenus analytiquement.

### Inconvénient

- Cette approximation polynomiale ne permet pas de modéliser des relations entrées/sorties complexes. Par exemple, les surfaces de réponses linéaires (ordre 1) ou quadratiques (ordre 2) sont incapables de capturer d'éventuels multiples extrema locaux de la fonction de référence.

On va donc s'intéresser aux modèles polynomiaux d'ordre supérieur afin de modéliser des phénomènes plus complexes.

## Modèles polynomiaux d'ordre $k$

Une approximation polynomiale multivariée à  $d$  variables et d'ordre  $k$  prend la forme

suivante :

$$\tilde{y} = \sum_{\alpha/\|\alpha\|_1 \leq k} c_\alpha \mathbf{x}^\alpha \text{ avec } \begin{cases} \mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d, \\ \alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbb{N}^d, \\ \|\alpha\|_1 = \sum_{i=1}^d \alpha_i, \\ c_\alpha \in \mathbb{R}, \\ \mathbf{x}^\alpha = \prod_{i=1}^d x_i^{\alpha_i}. \end{cases} \quad (2.2)$$

La base polynomiale standard est composée des monômes  $\mathbf{x}^\alpha$ . Les coefficients  $c_\alpha$  sont les paramètres à déterminer. Toutes les bases polynomiales (bases d'Hermite, de Legendre...) sont uniquement composées de puissance de  $\mathbf{x}$  et les modèles polynomiaux peuvent donc tous être écrits sous la forme générale présentée par (2.2).

### Avantages

- La fonction  $\tilde{y}$  est linéaire par rapport aux paramètres inconnus  $c_\alpha$ .
- Comme pour les modèles polynomiaux d'ordre 1 ou 2, ces modèles ont l'avantage principal d'être faciles à mettre en œuvre et à interpréter.
- Ils permettent par ailleurs de réaliser de bonnes approximations locales d'une fonction.

### Inconvénients

- Afin d'établir l'approximation polynomiale d'une fonction,  $\frac{(k+d)!}{k! d!}$  coefficients  $c_l$  sont à déterminer. Le nombre de coefficients explose avec la dimension  $d$  des entrées. Pour diminuer le nombre de coefficients à déterminer lorsque l'ordre du modèle augmente, on peut aussi utiliser des polynômes creux, c'est-à-dire qu'on fixe le plus judicieusement possible certains coefficients à 0.
- Augmenter l'ordre du polynôme pour coller au mieux aux données échantillonnées n'est pas forcément le meilleur choix, car des oscillations artificielles peuvent apparaître (phénomène de Runge [24]). Pour éviter ce phénomène d'oscillations, les splines sont une alternative possible (cf. [32] pour plus de détails).

## 2.3.2 Réseaux de neurones

Les réseaux de neurones sont basés sur des modèles idéalisés de la structure du cerveau humain. La notion d'apprentissage est intuitive pour les humains ou les animaux : c'est une procédure cognitive qui doit faire en sorte que l'individu réalise de manière autonome une tâche donnée. L'apprentissage supervisé numérique poursuit le même objectif : c'est un algorithme qui permet d'ajuster les paramètres d'une fonction afin que celle-ci approche au mieux la fonction de référence représentée par un échantillon de valeurs sur son domaine de

définition. Ce modèle peut être considéré comme une extension des méthodes de régression classique [7].

Un réseau de neurones unidirectionnel et entièrement connecté comprenant une couche d'entrée de  $d$  neurones,  $K_{max}$  couches cachées de  $(N_k)_{1 \leq k \leq K_{max}}$  neurones et une couche de sortie de 1 neurone est considéré en figure 2.2.

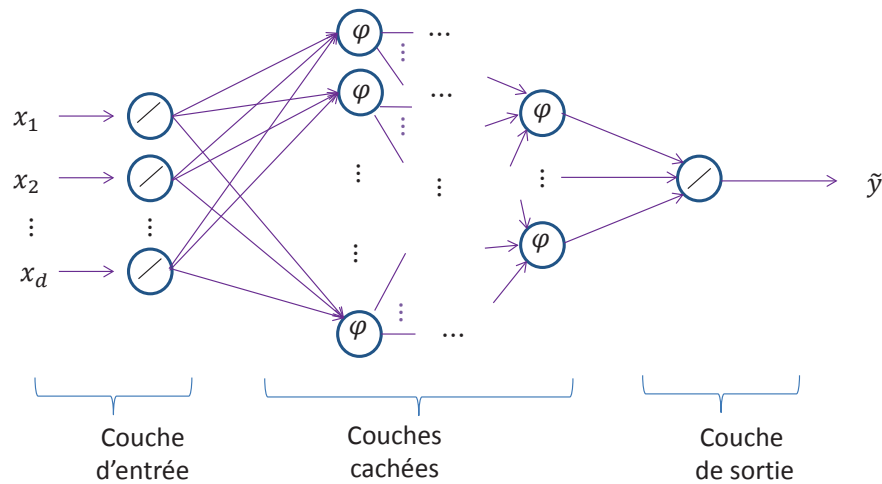


FIGURE 2.2 – Réseau de neurones unidirectionnel et entièrement connecté.

Chaque neurone  $i$ ,  $1 \leq i \leq d$ , de la couche d'entrée alloue la valeur de l'entrée à sa sortie  $s_{0,i} : s_{0,i} = x_i$  (cf. figure 2.3).

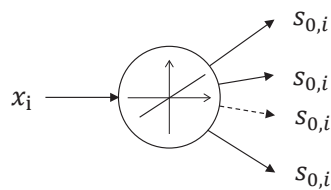


FIGURE 2.3 – Neurone de la couche d'entrée.

Pour les couches cachées (la couche d'entrée peut être vue comme la 0<sup>ème</sup> couche cachée), le calcul de la réponse d'un neurone  $j$  peut être décomposé en deux étapes (cf. figure 2.4).

- La première étape consiste en une combinaison linéaire des variables.

Pour un neurone  $j$ ,  $1 \leq j \leq N_k$ , d'une couche cachée  $k$ ,  $1 \leq k \leq K_{max}$  :

$$b_{k,j} + \sum_{i=1}^{N_{k-1}} a_{k,j,i} s_{(k-1),i}.$$

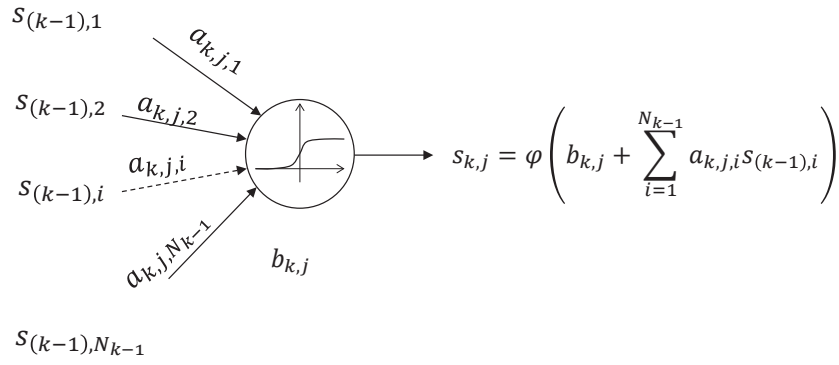


FIGURE 2.4 – Neurone de la couche cachée  $k$ .

Le coefficient  $a_{k,j,i}$  est le poids du lien entre le neurone  $i$  de la couche cachée  $k-1$  et le neurone  $j$  de la couche cachée  $k$ , le coefficient  $b_{k,j}$  est le biais associé au neurone  $j$  de la couche  $k$ . Les poids et le biais de chaque neurone des couches cachées sont des paramètres ajustables.

- Afin de conférer au réseau de neurones la capacité d’approcher des fonctions non-linéaires, la seconde étape introduit une non-linéarité grâce à la fonction d’activation  $\varphi$  des neurones des couches cachées.

Sortie d’un neurone  $j$ ,  $1 \leq j \leq N_k$ , de la couche  $k$ ,  $1 \leq k \leq K_{max}$  :

$$s_{k,j} = \varphi \left( b_{k,j} + \sum_{i=1}^{N_{k-1}} a_{k,j,i} s_{(k-1),i} \right).$$

Les deux fonctions d’activation les plus utilisées sont les suivantes [43] :

$$\forall x \in \mathbb{R}, \varphi(x) = \begin{cases} \frac{1}{1 + e^{-x}} & \text{(fonction sigmoïde),} \\ \tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} & \text{(tangente hyperbolique).} \end{cases}$$

Finalement, la couche de sortie est composée d’un neurone qui calcule la somme pondérée de ses entrées, qui sont les sorties  $s_{K_{max}j}$ ,  $1 \leq j \leq N_{K_{max}}$ , de la dernière couche cachée (cf. figure 2.5) :

$$\tilde{y} = \sum_{j=1}^{N_{K_{max}}} c_{K_{max}j} s_{K_{max}j},$$

où les  $c_{K_{max}j}$  sont des paramètres ajustables.

Une fois que la structure est choisie (principalement le nombre  $K_{max}$  de couches cachées et le nombre  $N_k$  de neurones par couche cachée), la seconde étape de la construction d’un réseau de neurones consiste à ajuster les paramètres. La fonction  $\tilde{y}$  est une fonction non-linéaire des paramètres inconnus. Cette étape, appelée le processus d’apprentissage, est un problème d’optimisation dans lequel l’objectif est de minimiser l’écart entre la

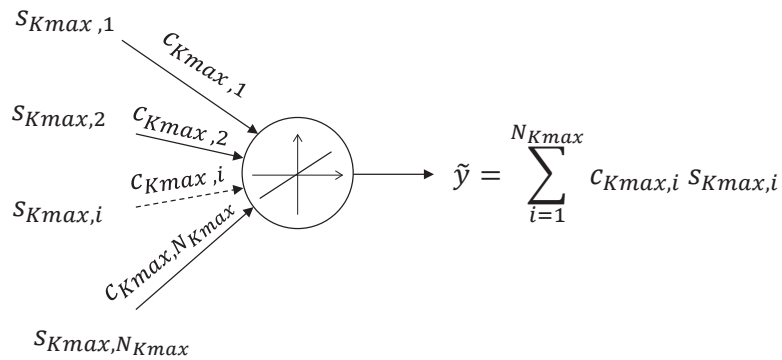


FIGURE 2.5 – Neurone de la couche de sortie.

fonction de référence et le réseau de neurones, construit sur le domaine de la fonction. La boîte à outils "Neural Network" de MATLAB permet de minimiser la fonction d'erreur quadratique en utilisant l'algorithme de Levenberg Marquadt.

#### Avantage

- Les réseaux de neurones multi-couches unidirectionnels et entièrement connectés ont comme attribut général d'être approximateurs universels parcimonieux [47, 22] :
  - les réseaux de neurones, en tant qu'approximateurs universels, ont la capacité d'approcher n'importe quelle fonction continue avec une précision arbitraire ;
  - du fait de leur caractère parcimonieux, le nombre de paramètres nécessaires est souvent plus faible que pour d'autres méthodes de réduction [4, 48].

#### Inconvénients

- Il est difficile de déterminer la structure la mieux adaptée à un problème donné.
- On ne peut généralement pas obtenir de majorant de l'erreur autre que statistique, ce qui peut poser des problèmes pour la validation des systèmes dans lesquels ils sont utilisés.

### 2.3.3 Krigeage

Le krigeage est une technique issue de l'analyse minière et de la géostatistique [21]. Les travaux de l'ingénieur minier sud-africain Krige [54] sont précurseurs de la méthode. Cependant, le terme krigeage et son formalisme sont dus au français Matheron [62, 36], qui en a aussi assuré le développement au Centre de Géostatistique de l'École des Mines de Paris. En fait, les fondements de la méthode ont été développés parallèlement par d'autres chercheurs, notamment le météorologue Gandin [33] de l'ex-URSS, mais c'est aujourd'hui sous la terminologie proposée par Matheron qu'elle est la plus connue.

Le krigeage est une méthode stochastique d'interpolation spatiale qui prévoit la valeur

d'une réponse par une combinaison linéaire sans biais et à variance minimale des données du plan d'expérience. On suppose que les points proches dans l'espace tendent à posséder des caractéristiques similaires.

L'hypothèse centrale du krigeage est que l'on cherche à approcher les valeurs de la fonction de référence  $F$  par la réalisation d'un processus stochastique gaussien  $Z$  composé d'une partie déterministe  $\mu$  et d'une partie aléatoire  $W$ . Étant donné  $\mathbf{x} \in \mathbb{R}^d$ , on décompose le processus stochastique  $Z$  :

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + W(\mathbf{x}),$$

où :

- $\mu$  représente la tendance moyenne du processus,
- $W$  est un processus gaussien stationnaire d'espérance nulle  $\mathbb{E}[W(\mathbf{x})] = 0$ , de variance  $\sigma^2$  et de fonction de corrélation  $R_\theta$ .

Différentes fonctions de corrélation  $R_\theta$  peuvent être utilisées et permettent d'introduire certaines hypothèses de régularité sur la fonction à modéliser. On suppose que  $R_\theta$  modélise le fait que plus les points  $\mathbf{x}$  et  $\mathbf{x}'$  sont proches, plus il y a de chances que leurs images respectives par  $F$  soient proches elles aussi. De plus, on suppose qu'il s'agit d'une fonction noyau continue symétrique strictement positive telle que pour tout  $\mathbf{x}$ ,  $R_\theta(\mathbf{x}, \mathbf{x}) = 1$ . La covariance du krigeage est une fonction proportionnelle à la fonction de corrélation  $R_\theta$  comme montré dans l'équation suivante :

$$\text{cov}(W(\mathbf{x}), W(\mathbf{x}')) = \sigma^2 R_\theta(\mathbf{x}, \mathbf{x}').$$

Plusieurs types de krigeage existent selon la forme donnée à  $\mu$ . Les trois principaux sont présentés ci-après allant du plus simple au plus général :

- krigeage simple :  $\mu$  est une fonction constante connue
- krigeage ordinaire :  $\mu$  est une fonction constante inconnue :  $\forall \mathbf{x} \in \mathbb{R}^d, \mu(\mathbf{x}) = \beta_0 \in \mathbb{R}$ .
- krigeage universel :  $\mu$  est une combinaison linéaire de  $n$  fonctions  $h_i$  connues, pondérées par des coefficients  $\beta_i \in \mathbb{R}$  à déterminer :  $\forall \mathbf{x} \in \mathbb{R}^d, \mu(\mathbf{x}) = \sum_{i=1}^n \beta_i h_i(\mathbf{x})$ . On choisit généralement une base de fonctions  $h_i$  polynomiales.

Nous visualisons sur la figure 2.6 plusieurs réalisations d'un processus stochastique gaussien  $Z$  de moyenne et de covariance connues.

Nous décrivons maintenant les différentes étapes de construction d'un modèle de krigeage.

- Étape 1 : choix du type de krigeage utilisé (simple, ordinaire ou universel) et génération d'un plan d'expérience ( $[\bar{\mathbf{x}}_{app}], \mathbf{Y}_{app}$ ).
- Étape 2 : choix de la fonction de corrélation  $R_\theta$ . Par exemple, une fonction de corrélation utilisée de manière classique est [1] :

$$R_\theta(\mathbf{x}, \mathbf{x}') = \exp\left(-\sum_{i=1}^d \theta_i \|\mathbf{x}_i - \mathbf{x}'_i\|^{p_i}\right),$$

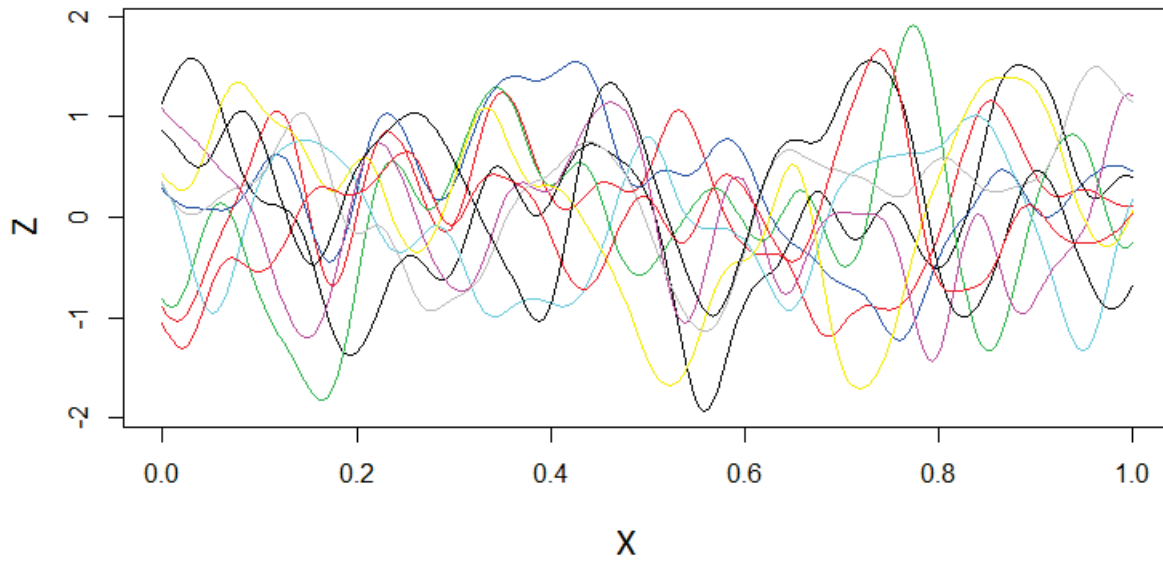


FIGURE 2.6 – Simulation de trajectoires d'un processus gaussien.

où  $0 < p_i \leq 2$  reflète la régularité de l'interpolation. Plus précisément, une valeur élevée imposera un comportement plutôt lisse du modèle alors qu'une valeur plus faible provoquera des variations plus importantes.

- Étape 3 : estimation des paramètres  $\beta$ ,  $\theta$  et  $\sigma$  par régression et/ou maximum de vraisemblance [18].
- Étape 4 : pour chaque valeur d'entrée, le krigeage fournit un prédicteur non biaisé de variance optimale  $\tilde{F}$  de la fonction de référence  $F$  comme nous allons le voir par la suite.

On conditionne le processus gaussien à passer par les  $N$  points du plan d'expérience, c'est-à-dire  $Z(\mathbf{x}_j)_{1 \leq j \leq N} = (y_j)_{1 \leq j \leq N} \in \mathbb{R}^N$  (cf. figure 2.7). Le prédicteur  $\tilde{F}$  est donné par l'espérance de ce processus conditionné :

$$\tilde{F}(\mathbf{x}) = \mathbb{E}(Z(\mathbf{x}) | Z(\mathbf{x}_1) = F(\mathbf{x}_1), \dots, Z(\mathbf{x}_N) = F(\mathbf{x}_N)).$$

On obtient donc directement la fonction de prédiction  $\tilde{F}$  du modèle de krigeage [23] :

$$\tilde{F}(\mathbf{x}) = \mu(x) + k_{\bar{x}}(\mathbf{x}) \Sigma_{\bar{x}}^{-1} y_{\bar{x}},$$

où

$$k_{\bar{x}}(\mathbf{x}) = \begin{pmatrix} R_{\theta}(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ R_{\theta}(\mathbf{x}, \mathbf{x}_N) \end{pmatrix}$$

$$y_{\bar{x}} = \begin{pmatrix} F(\mathbf{x}_1) - \mu(\mathbf{x}_1) \\ \vdots \\ F(\mathbf{x}_N) - \mu(\mathbf{x}_N) \end{pmatrix}.$$

et  $\Sigma_{\bar{x}}$  est une matrice symétrique définie positive de taille  $N \times N$  dont le coefficient  $(i, j)$  est donné par :

$$(\Sigma_{\bar{x}})_{i,j} = R_{\theta}(\mathbf{x}_i, \mathbf{x}_j).$$

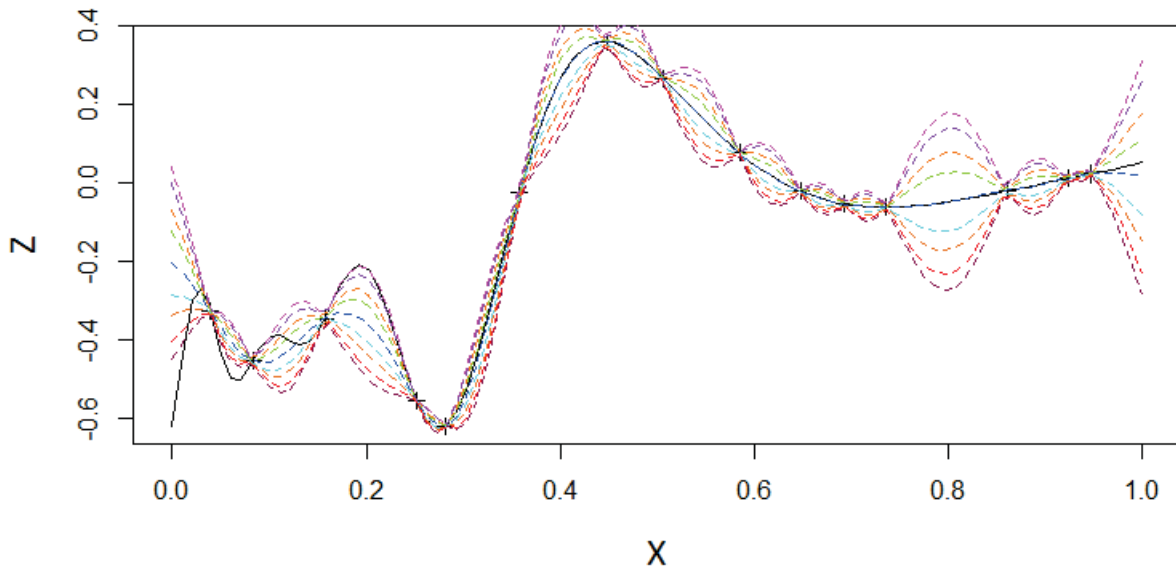


FIGURE 2.7 – Simulation de trajectoires d'un processus gaussien conditionné à passer par des points donnés.

Tout comme l'espérance du processus stochastique conditionné  $Z$  a été déterminée, on est capable d'évaluer l'incertitude de prédiction à partir de l'estimation de la variance de ce même processus  $Z$  conditionné [38, 5] :

$$\tilde{\sigma}^2(\mathbf{x}) = \text{var}(Z(\mathbf{x}) | Z(\mathbf{x}_1) = F(\mathbf{x}_1), \dots, Z(\mathbf{x}_N) = F(\mathbf{x}_N)).$$

On obtient alors l'intervalle de confiance à 95 % du processus conditionné (noté  $IC$ ) :

$$IC = [IC_{\text{inf}}, IC_{\text{sup}}] \simeq [\tilde{F}(\mathbf{x}) - 1.96\sqrt{\tilde{\sigma}^2(\mathbf{x})}, \tilde{F}(\mathbf{x}) + 1.96\sqrt{\tilde{\sigma}^2(\mathbf{x})}]. \quad (2.3)$$

**Avantage**

- Une estimation de l'incertitude associée à la prédiction de la valeur de la fonction  $F$  est fournie en tout point. On connaît donc l'écart type de l'erreur en tout point du modèle réduit. C'est un indicateur de la précision de l'estimation réalisée qui quantifie la dispersion possible de la valeur estimée autour de la vraie valeur.

### Inconvénient

- L'évaluation du modèle de krigeage nécessite d'inverser la matrice de covariance  $\Sigma$  des échantillons d'apprentissage.
  - des erreurs numériques liées à son conditionnement peuvent apparaître,
  - l'inversion de la matrice de covariance limite l'utilisation de ce modèle. Typiquement le temps d'inversion de la matrice est en  $\mathcal{O}(N^3)$ ,  $N$  étant le nombre de points contenus dans l'ensemble d'apprentissage [2, 3].

## 2.3.4 High Dimensional Model Representation

Le High Dimensional Model Representation (HDMR) permet d'exprimer une fonction  $F$  quelconque en une somme de fonctions. Soit  $F$  une fonction quelconque de  $d$  variables d'entrée  $\mathbf{x}=(x_1, \dots, x_d)$  dont l'ensemble de définition est un produit cartésien d'intervalles  $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_d, b_d]$ . La décomposition unique proposée par Hoeffding [44, 26] permet de représenter  $F$  sous la forme d'une somme de fonctions de 0, 1, 2,  $\dots$ ,  $d$  variables :

$$F(x_1, \dots, x_d) = Z_0 + \sum_{i_1=1}^d Z_{i_1}(x_{i_1}) + \sum_{1 \leq i_1 < i_2 \leq d} Z_{i_1, i_2}(x_{i_1}, x_{i_2}) + \dots \quad (2.4)$$

$$\dots + Z_{1,2,\dots,d}(x_1, x_2, \dots, x_d).$$

Le terme  $Z_0$  est constant et représente l'ordre 0 de la fonction, c'est-à-dire sa réponse moyenne. Les  $d$  fonctions  $Z_i$  sont les termes d'ordre 1 représentant l'effet de chaque variable  $x_i$  agissant seule sur la sortie  $F$ . Les  $\frac{d(d-1)}{2}$  fonctions  $Z_{i,j}$  sont des termes d'ordre 2 qui décrivent les effets coopératifs des variables  $x_i$  et  $x_j$ . Les termes d'ordre supérieur donnent les effets coopératifs d'un nombre croissant de variables agissant ensemble sur la sortie  $F$ . En pratique, les modèles réduits  $\tilde{F}$  peuvent souvent être obtenus en tronquant (2.4) à l'ordre 0, à l'ordre 1, ou à l'ordre 2. La validité de la troncature est basée sur le principe fondamental que les corrélations entre les variables d'entrée affectant le modèle décroissent généralement rapidement [78]. Cette méthode construit ainsi une carte de relations entre les entrées et la sortie grâce aux corrélations d'ordre faible parmi les variables d'entrée qui ont un impact sur la sortie.

Afin de simplifier les écritures, dans toute la suite de cette section, les variables d'entrée sont supposées normalisées sur l'hypercube unité  $\mathbb{H}^d = [0, 1]^d$ .

Un composant particulier  $Z_{i_1, i_2, \dots, i_l}(x_{i_1}, x_{i_2}, \dots, x_{i_l})$ ,  $l = 0, 1, \dots, d-1$  est obtenu comme la solution d'un problème d'optimisation qui minimise l'erreur quadratique entre

la fonction  $F$  et sa décomposition (2.4) tronquée à l'ordre  $l$  [58]. Le cut-HDMR et le RS-HDMR sont deux types de décomposition communément utilisés que nous rappelons ici [60].

### Décomposition cut-HDMR

La décomposition cut-HDMR est la solution du problème d'optimisation énoncé ci-dessus au voisinage d'un point de référence  $\mathbf{c}$ . Afin de garantir que chacun des  $2^d$  termes de la décomposition (2.4) soient définis de façon unique, on impose la *vanishing condition*,  $\forall 1 \leq k \leq d, \forall 1 \leq i_1 < \dots < i_k \leq d, \forall s \in \{i_1, \dots, i_k\}$ ,

$$\int_{[0,1]^d} Z_{i_1, i_2, \dots, i_k}(u_{i_1}, \dots, u_{i_k}) \delta(u_s - c_s) du_s = 0. \quad (2.5)$$

Un composant quelconque de cette décomposition disparaît quand une de ses variables prend la valeur de l'élément correspondant dans  $\mathbf{c}$ .

Ces conditions (2.5) sont équivalentes à une relation d'orthogonalité entre deux composants distincts de la décomposition cut-HDMR :

$$\int_{[0,1]^d} Z_{i_1, i_2, \dots, i_k}(u_{i_1}, \dots, u_{i_k}) Z_{j_1, j_2, \dots, j_l}(u_{j_1}, \dots, u_{j_l}) \prod_{i=1}^d \delta(u_i - c_i) d\mathbf{u} = 0,$$

si  $\{i_1, \dots, i_k\} \neq \{j_1, \dots, j_l\}$  avec  $1 \leq k, l \leq d$ .

Les termes de la décomposition sont calculés séquentiellement en utilisant la *vanishing condition* (2.5).

– Terme d'ordre 0 :

$$\begin{aligned} Z_0 &= \int_{[0,1]^d} F(\mathbf{u}) \prod_{i=1}^d \delta(u_i - c_i) d\mathbf{u} \\ &= F(\mathbf{c}). \end{aligned}$$

– Termes d'ordre 1 :  $\forall i_1 \in \{1, \dots, d\}$ ,

$$\begin{aligned} Z_{i_1}(x_{i_1}) &= \int_{[0,1]^{d-1}} F(\mathbf{u}) \prod_{\substack{1 \leq i \leq d \\ i \neq i_1}} \delta(u_i - c_i) du_i - Z_0 \\ &= F(x_{i_1}, (\mathbf{c} \setminus c_{i_1})) - Z_0. \end{aligned}$$

– Termes d'ordre 2 :  $\forall i_1 \in \{1, \dots, d\}, \forall i_2 \in \{1, \dots, d\}$  et  $i_1 < i_2$ ,

$$\begin{aligned} Z_{i_1, i_2}(x_{i_1}, x_{i_2}) &= \int_{[0,1]^{d-2}} F(\mathbf{u}) \prod_{\substack{1 \leq i \leq d \\ i \neq i_1 \\ i \neq i_2}} \delta(u_i - c_i) du_i - Z_0 - Z_{i_1}(x_{i_1}) - Z_{i_2}(x_{i_2}) \\ &= F(x_{i_1}, x_{i_2}, (\mathbf{c} \setminus (c_{i_1}, c_{i_2}))) - Z_0 - Z_{i_1}(x_{i_1}) - Z_{i_2}(x_{i_2}). \end{aligned}$$

- On continue la procédure jusqu'à ce que toutes les sommes de dimension  $d - 1$  soient définies. Le dernier terme est calculé grâce à (2.4).

La détermination des composants de la décomposition cut-HDMR comme solutions du problème d'optimisation sous les conditions (2.5) revient à substituer  $(x_{i_1}, \dots, x_{i_k})$  par  $(c_{i_1}, \dots, c_{i_k})$  avec des ensembles différents  $\{i_1, i_2, \dots, i_l\} \subset \{1, 2, \dots, N\}$  dans (2.4) [74].

Si chaque variable  $x_i$  est découpée en un ensemble de  $J_i$  valeurs discrètes :

$$\{x_{1,1}, \dots, x_{1,J_1}\}, \{x_{2,1}, \dots, x_{2,J_2}\}, \dots, \{x_{d,1}, \dots, x_{d,J_d}\},$$

avec  $s = \max_{i=1, \dots, d} J_i$ , alors pour un ordre donné  $N_{max}$ , le nombre maximal d'appels à la fonction de référence est égal à :

$$\begin{aligned} \sum_{i=1}^{N_{max}} \frac{d!}{(n-i)!i!} (s-1)^i &= 1 + d(s-1) + \frac{d(d-1)(s-1)^2}{2} + \dots \\ &\dots + \frac{n!}{(n-N_{max})!N_{max}!} (s-1)^{N_{max}}. \end{aligned}$$

### Avantages

- Ce modèle est rapide à obtenir.
- Le cut-HDMR est une représentation exacte de la fonction  $F$  le long des coupures (lignes, plans,...) passant par le point  $\mathbf{c}$  et permet ainsi une étude locale de la fonction  $F$ .

### Inconvénients

- Les variables d'entrée doivent être définies sur un domaine hypercubique.
- La décomposition cut-HDMR est la solution de la minimisation de l'erreur quadratique dans le voisinage du point de référence. Il ne permet donc pas une approximation globale de la fonction.

### Décomposition Random Sampling - HDMR (RS-HDMR)

La décomposition RS-HDMR est la solution de la minimisation de l'erreur quadratique entre la fonction  $F$  et sa décomposition (2.4) sur tout le domaine de définition de  $F$ .

De la même manière que pour le cut-HDMR, les composants de la décomposition RS-HDMR solutions du problème d'optimisation sont déterminés de manière unique sous la *vanishing condition*,  $\forall 1 \leq k \leq d, \forall 1 \leq i_1 < \dots < i_k \leq d, \forall s \in \{i_1, \dots, i_k\}$ ,

$$\int_{[0,1]} Z_{i_1, i_2, \dots, i_k}(u_{i_1}, u_{i_2}, \dots, u_{i_k}) du_s = 0. \quad (2.6)$$

Ces conditions permettent d'obtenir l'orthogonalité de deux composants distincts de

la décomposition HDMR :

$$\int_{[0,1]^d} Z_{i_1, i_2, \dots, i_k}(u_{i_1}, \dots, u_{i_k}) Z_{j_1, j_2, \dots, j_l}(u_{j_1}, \dots, u_{j_l}) d\mathbf{u} = 0,$$

si  $\{i_1, \dots, i_k\} \neq \{j_1, \dots, j_l\}$  avec  $1 \leq k, l \leq d$ .

En utilisant la *vanishing condition* (2.6), les termes de la décomposition s'écrivent de la façon suivante.

– Terme d'ordre 0 :

$$Z_0 = \int_{\prod_{k=1}^d [0,1]} F(\mathbf{u}) \prod_{k=1}^d du_k.$$

– Termes d'ordre 1 :  $\forall i_1 \in \{1, \dots, d\}$ ,

$$Z_{i_1}(x_{i_1}) = \int_{[0,1]^{d-1}} F(\mathbf{u}) \prod_{\substack{1 \leq k \leq d \\ k \neq i_1}} du_k - Z_0.$$

– Termes d'ordre 2 :  $\forall i_1 \in \{1, \dots, d\}, \forall i_2 \in \{1, \dots, d\}$  et  $i_1 < i_2$ ,

$$Z_{i_1, i_2}(x_{i_1}, x_{i_2}) = \int_{[0,1]^{d-2}} F(\mathbf{u}) \prod_{\substack{1 \leq k \leq d \\ k \neq i_1 \\ k \neq i_2}} du_k - Z_0 - Z_{i_1}(x_{i_1}) - Z_{i_2}(x_{i_2}).$$

– On continue la procédure jusqu'à ce que toutes les sommes de dimension  $d-1$  soient définies. Le dernier terme est calculé grâce à (2.4).

### Avantage

– Le RS-HDMR permet de décomposer de façon automatique la fonction de référence en une somme de fonctions de dimensions inférieures qui peuvent être tabulées.

### Inconvénients

- Les variables d'entrée doivent être définies sur un domaine hypercubique.
- Le nombre d'intégrales augmente rapidement lorsque le nombre de variables d'entrée augmente. Nous quantifierons cette augmentation dans le chapitre 3.
- Les intégrales multiples à calculer sont de grandes dimensions.

## 2.4 Méthodes d'échantillonnage

Les premières méthodes rigoureuses d'expérimentation basées sur les plans d'expérience sont liées à l'agronomie. La première formalisation mathématique est due à Fisher [29] en 1935. L'applicabilité de ces méthodes s'est depuis étendue à différents secteurs

industriels, grâce notamment à l'augmentation du calcul numérique depuis les années 90. On s'intéresse dans cette section aux plans d'expérience numériques sur lesquels on évalue le modèle de référence et qui servent de base pour construire le modèle réduit associé. L'objectif est d'être capable d'appeler le modèle de référence de façon optimale afin d'estimer les paramètres du modèle réduit le plus précisément possible. Nous présentons tout d'abord la construction déterministe de grilles régulières dans la section 2.4.1. Nous présentons ensuite la méthode de Monte Carlo directe dans la section 2.4.2, ainsi que des améliorations de celle-ci dans les sections 2.4.3, 2.4.4 et 2.4.5.

Le nombre de points générés doit être assez grand afin d'obtenir une bonne représentation du modèle de référence mais il est limité par le temps nécessaire à l'appel du modèle de référence et par les possibilités de stockage en mémoire. On estime le temps d'appel au modèle de référence pré-équilibré en lacet et roulis (cf. figure 1.15) à  $10^{-2}$  secondes sur un PC classique. On est donc capable de générer 360 000 points du plan d'expérience en une heure et 4 320 000 en une nuit. Du fait des contraintes industrielles, on veut être capable de générer le modèle réduit en 48 heures maximum.

### 2.4.1 Grilles régulières

La méthode la plus simple pour remplir l'espace consiste à choisir les points sur une grille régulière dans le domaine expérimental [31]. La figure 2.8 représente une grille régulière en dimension  $d = 2$  avec  $n = 5$  niveaux par variable d'entrée, soit  $N = 5^2 = 25$  points.

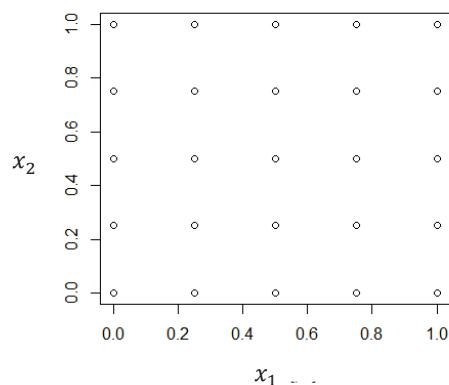


FIGURE 2.8 – Grille régulière à 5 niveaux pour  $[0, 1]^2$ .

#### Avantage

- Cette méthode permet d'assurer un bon remplissage de l'espace.

#### Inconvénients

- Le nombre de points  $N = n^d$  de la grille augmente exponentiellement avec la dimension  $d$ .

- La grille régulière devient moins intéressante en grande dimension, car ce plan est très mauvais en projection. Lorsque le modèle de référence dépend fortement de quelques unes de ses variables d'entrée, cette méthode d'exploration de l'espace conduit à des réplifications des données inutiles. Par exemple, si la réponse du modèle de référence est telle que pour  $\forall x_1 = x_1^0$  fixée  $\forall (x_2, x_2') \in [0, 1]^2$ ,  $F(x_1^0, x_2) = F(x_1^0, x_2')$  alors la grille régulière de la figure 2.8 n'est pas satisfaisante puisque l'information donnée par celle-ci se réduit à 5 points au lieu de 25.

Il peut être intéressant dans certains cas de générer une grille régulière que l'on complète par des points générés aléatoirement par les méthodes décrites dans la section 2.4.2.

## 2.4.2 Méthode de Monte Carlo directe

Soit  $G : \mathbb{R}^d \rightarrow \mathbb{R}$  et  $\mathbf{X} = (X_1, X_2, \dots, X_d)$ , un vecteur aléatoire de densité de probabilité  $\omega$  sur  $[0, 1]^d$ . Notons  $\mu$  l'intégrale sur l'hypercube unité en dimension  $d$  noté  $\mathbb{H}^d = [0, 1]^d$  :

$$\mu = \mathbb{E}(G(\mathbf{X})) = \int_{\mathbb{H}^d} G(\mathbf{x})\omega(\mathbf{x})d\mathbf{x}. \quad (2.7)$$

La méthode de Monte Carlo a été introduite pour estimer numériquement  $\mu$ . Les premiers résultats [65, 64] sont apparus au milieu du  $XX^{\text{ème}}$  siècle.

Nous prenons comme densité de probabilité  $\omega$  une distribution uniforme sur  $[0, 1]^d$ .

On introduit l'estimateur  $\mu_N$  avec les points  $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)^T$  générés selon la densité de probabilité  $\omega$  (échantillon de la loi  $\mathbf{X}$ ) :

$$\mu_N = \frac{1}{N} \sum_{i=1}^N G(\mathbf{x}_i).$$

C'est un estimateur sans biais de  $\mu$ . D'après la loi des grands nombres, la moyenne empirique  $\mu_N$  de réalisations indépendantes et identiquement distribuées converge presque sûrement vers l'espérance mathématique  $\mu$ , c'est-à-dire :

$$P \left( \lim_{N \rightarrow +\infty} |\mu_N - \mu| = 0 \right) = 1.$$

Posons  $\epsilon_N = \mu - \mu_N$ , l'erreur commise par l'estimateur et supposons  $\text{var}(G(\mathbf{X})) = \sigma^2$ . Alors en appliquant le théorème de la limite centrale  $\epsilon_N \frac{\sqrt{N}}{\sigma}$  converge en loi vers une distribution normale centrée et réduite lorsque la taille  $N$  de l'échantillon tend vers l'infini, ce qui permet la détermination d'un intervalle de confiance.

### Avantage

- La vitesse de convergence qui est de l'ordre de  $\frac{\sigma}{\sqrt{N}}$  présente l'avantage d'être insensible à la dimension et de ne pas dépendre de la régularité de la fonction  $G$  à intégrer.

### Inconvénients

- La vitesse de convergence  $\left(\frac{\sigma}{\sqrt{N}}\right)$  reste lente : si l'on veut diviser par 10 l'intervalle de confiance, c'est-à-dire augmenter la précision de l'estimateur, il faut multiplier par  $10^2$  le nombre de points compris dans l'échantillon.
- Cet intervalle de confiance est défini en fonction de l'écart-type  $\sigma$  qui n'est pas connu *a priori*. On peut faire appel à la loi de Student [25] qui s'appuie sur l'estimation de la variance, donnée par la relation suivante :  $\sigma_N^2 = \frac{1}{N-1} \sum_{i=1}^N (G(\mathbf{x}_i) - \mu_N)^2$ .

Les générateurs permettant de tirer les points selon la loi uniforme utilisés dans la méthode de Monte Carlo directe sont basés sur une relation de récurrence [84]. La suite de points générée possède les caractéristiques d'un processus aléatoire, c'est-à-dire notamment qu'elle a une période suffisamment grande avant de recommencer la même suite. Nous allons explorer par la suite les méthodes qui existent pour améliorer la convergence de la méthode de Monte Carlo directe, en générant les points du plan d'apprentissage non plus avec des générateurs pseudo-aléatoires, mais avec des générateurs capables soit de réduire la variance  $\sigma$  de l'estimateur (cf. section 2.4.3) soit d'assurer une bonne répartition des points (cf. section 2.4.4).

## 2.4.3 Amélioration de la méthode de Monte Carlo directe par réduction de la variance

Nous venons de voir qu'une manière d'améliorer l'erreur de la méthode de Monte Carlo consiste à essayer de réduire la variance des estimateurs de Monte Carlo. Ces méthodes sont appelées méthodes de réduction de la variance. Ce sujet a été largement étudié (cf. [12, 57, 41] pour plus de détails).

### 2.4.3.1 Variables antithétiques

On cherche à calculer l'intégrale  $\mu = \mathbb{E}(G(\mathbf{X})) = \mathbb{E}(\mathbf{Y})$ . L'idée des variables antithétiques est d'utiliser deux variables aléatoires  $\mathbf{Y}_1$  et  $\mathbf{Y}_2$  qui ont la même loi que  $\mathbf{Y}$ . Les égalités suivantes sont vérifiées :

$$\begin{aligned} - \mathbb{E}(\mathbf{Y}) &= \frac{1}{2} \left( \mathbb{E}(\mathbf{Y}_1) + \mathbb{E}(\mathbf{Y}_2) \right) = \mathbb{E} \left( \frac{\mathbf{Y}_1 + \mathbf{Y}_2}{2} \right), \\ - \text{var} \left( \frac{\mathbf{Y}_1 + \mathbf{Y}_2}{2} \right) &= \frac{\text{var}(\mathbf{Y}_1) + \text{var}(\mathbf{Y}_2) + 2\text{cov}(\mathbf{Y}_1, \mathbf{Y}_2)}{4}. \end{aligned}$$

Si les variables  $\mathbf{Y}_1$  et  $\mathbf{Y}_2$  sont négativement corrélées, c'est-à-dire  $\text{cov}(\mathbf{Y}_1, \mathbf{Y}_2) < 0$ , alors  $\text{var} \left( \frac{\mathbf{Y}_1 + \mathbf{Y}_2}{2} \right) < \frac{\text{var}(\mathbf{Y})}{2}$ . Il faut donc être en mesure de générer  $\mathbf{Y}_1$  et  $\mathbf{Y}_2$  négativement corrélées pour diminuer la variance [40].

### 2.4.3.2 Variables de contrôle

La méthode des variables de contrôle [52] est également basée sur l'idée d'utiliser des corrélations dans le but de réduire la variance de la méthode de Monte Carlo directe. Supposons que l'on dispose d'un modèle approché de  $G$  noté  $G_{\text{app}}$  et que l'on puisse calculer explicitement  $\mu_{\text{app}} = \mathbb{E}(G_{\text{app}}(\mathbf{X}))$ . En posant  $W_\beta = G + \beta(G_{\text{app}} - \mu_{\text{app}})$ , on a les résultats suivants :

- $\mu = \mathbb{E}(G(\mathbf{X})) = \mathbb{E}(W_\beta(\mathbf{X}))$ ,
- $\text{var}(W_\beta(\mathbf{X})) = \text{var}(G(\mathbf{X})) + \beta^2 \text{var}(G_{\text{app}}(\mathbf{X})) + 2\beta \text{cov}(G(\mathbf{X}), G_{\text{app}}(\mathbf{X}))$ .

La variance de  $W_\beta$  est minimale lorsque  $\beta^* = -\frac{\text{cov}(G(\mathbf{X}), G_{\text{app}}(\mathbf{X}))}{\text{var}(G_{\text{app}}(\mathbf{X}))}$ .

On a alors :

$$\text{var}(W_{\beta^*}(\mathbf{X})) = \text{var}(G(\mathbf{X})) - \frac{\text{cov}(G(\mathbf{X}), G_{\text{app}}(\mathbf{X}))^2}{\text{var}(G_{\text{app}}(\mathbf{X}))}.$$

Le modèle approché  $G_{\text{app}}$  doit donc préserver les principales caractéristiques du modèle d'origine afin que  $G(\mathbf{X})$  et  $G_{\text{app}}(\mathbf{X})$  soient suffisamment positivement corrélées pour réduire la variance.

### 2.4.3.3 Échantillonnage stratifié

Cette méthode utilise une partition  $\{D_i, 1 \leq i \leq m\}$  de l'espace :

$$\mathbb{H}^d = [0, 1]^d = \bigcup_{i=1}^m D_i, \quad D_i \cap D_j = \emptyset, \quad \forall i \neq j.$$

Ces régions  $D_i$  sont aussi appelées strates.

On intègre alors  $G$  sur chacune des strates avec des pondérations différentes selon les strates. Cette méthode permet de faire varier le nombre de tirages en fonction des zones de l'espace. Soit  $p_i = P(\mathbf{X} \in D_i) > 0$  la probabilité que le vecteur aléatoire  $\mathbf{X}$  appartienne au sous-domaine  $D_i$ . Les probabilités  $p_i$  sont supposées connues.

L'intégrale s'écrit alors :

$$\mu = \sum_{i=1}^m P(\mathbf{X} \in D_i) \mathbb{E}(G(\mathbf{X}) | \mathbf{X} \in D_i) = \sum_{i=1}^m p_i \mu_i.$$

Chaque intégrale  $\mu_i = \mathbb{E}(G(\mathbf{X}) | \mathbf{X} \in D_i)$  peut être estimée par une méthode de Monte Carlo directe avec  $n_i$  valeurs échantillonnées dans  $D_i$  ( $\sum_{i=1}^m n_i = N$ ). Les échantillons servant à estimer les  $\mu_i$  sont supposés indépendants et on obtient alors :

$$\text{var}(\mu) = \sum_{i=1}^m p_i^2 \frac{\text{var}(G(\mathbf{X}) | \mathbf{X} \in D_i)}{n_i}.$$

La réduction en variance n'est pas assurée par cette méthode. Elle dépend du choix des strates et du nombre de tirages affectés à chacune d'elle. Une stratégie d'affectation

consiste à choisir  $n_i$  de sorte qu'il soit proportionnel à sa probabilité  $p_i$ . Ainsi il a été démontré que la variance de l'estimateur est dans ce cas toujours diminuée [16]. Par exemple, lorsque le domaine d'intégration est un carré  $\mathbb{H}^2$ , on peut découper le domaine d'intégration en  $m$  carrés équiprobables et tirer autant de points dans chaque carré. Ainsi, pour le carré  $i$ ,  $p_i = \frac{1}{m}$  et  $n_i = \frac{N}{m}$ , la figure 2.9 illustre cet exemple classique.

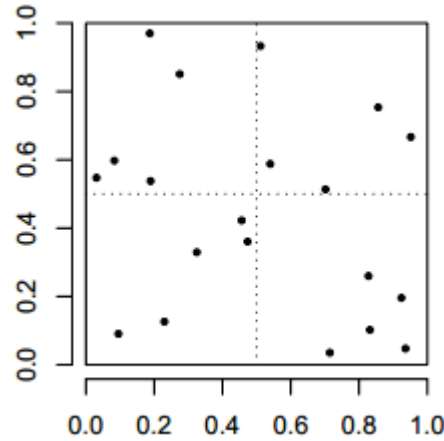


FIGURE 2.9 – Exemple d'échantillonnage stratifié avec  $m = 4$  et  $N = 20$ .

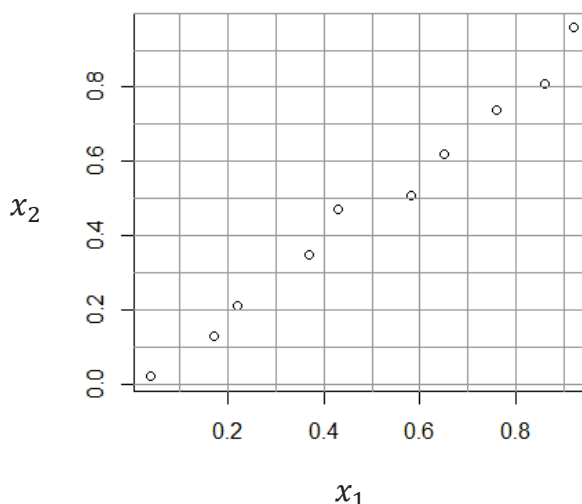
#### 2.4.3.4 Méthode *Latin Hypercube Sampling*

La méthode *Latin Hypercube Sampling* (LHS), introduite par McKay [63], reprend le principe de l'échantillonnage stratifié. Chaque axe  $[0, 1]$  du domaine de définition  $\mathbb{H}^d$  est découpé en  $N$  segments de même longueur  $\left[\frac{i}{N}, \frac{i+1}{N}\right]$  pour  $i = 0, \dots, N-1$ . En faisant le produit cartésien de ces intervalles, on obtient un découpage en  $N^d$  strates de même taille du domaine  $\mathbb{H}^d$ . Ensuite,  $N$  cellules sont choisies parmi les  $N^d$  possibles de telle manière que la projection sur un axe comporte un seul point dans l'intervalle  $\left[\frac{i}{N}, \frac{i+1}{N}\right]$  pour  $i = 0, \dots, N-1$ . Enfin, on tire au hasard dans chaque cellule sélectionnée. Cette méthode permet dans certains cas de réduire la variance, notamment lorsque  $G$  est une fonction presque additive [16].

Cette méthode garantit une couverture uniforme de l'étendue de chaque entrée et permet ainsi d'avoir des expérimentations non concentrées sur les bords du domaine expérimental, mais elle ne garantit pas l'uniformité des points dans tout le domaine expérimental (cf. figure 2.10).

##### Avantages

- Il est facile à construire et garantit une couverture uniforme de chaque entrée.
- Il permet d'avoir des expérimentations non concentrées sur les bords du domaine expérimental.

FIGURE 2.10 – LHS pas uniformément réparti dans  $\mathbb{H}^2$ .

- Si une ou plusieurs dimensions doivent être enlevées (suite par exemple à une analyse de sensibilité), le plan obtenu est toujours un hypercube latin.

### Inconvénients

- Une fois construit, il est difficile d'ajouter de nouveaux points tout en gardant les propriétés de l'hypercube latin.
- Pour  $N$  fixé, il existe  $N!$  permutations possibles pour chacune des  $d$  colonnes. Il existe donc  $(N!)^{d-1}$  hypercubes latins disponibles. Trouver le meilleur à l'aide d'un critère de sélection n'est souvent pas envisageable. D'autres méthodes existent que nous développons dans la section 2.4.5.

## 2.4.4 Méthode de quasi-Monte Carlo

### 2.4.4.1 Notions de discrédance d'une suite

L'uniformité d'une suite est caractérisée mathématiquement par sa discrédance. Le terme discrédance, d'origine latine *discrepantia* signifiant divergence, est utilisé en mathématique pour mesurer l'écart entre une distribution uniforme et une distribution donnée. En d'autres termes elle permet de mesurer l'irrégularité de la distribution. En effet, il est difficile d'appréhender l'uniformité d'une suite lorsqu'on est dans un espace à grande dimension.

Considérons le problème de construire dans l'intervalle  $[0, 1]$  une suite  $S$  de  $N$  points  $(\mathbf{x}_i)_{1 \leq i \leq N}$  uniformément répartis, ou encore équirépartis. Le plus simple est de constituer une suite de  $N$  points équidistants de la manière suivante :  $x_i = \frac{i-1}{N}$  pour  $i = 1, \dots, N$ . Une représentation de cette configuration avec 10 points est présentée dans la figure 2.11.

On considère maintenant le même problème en se plaçant cette fois dans un carré unité

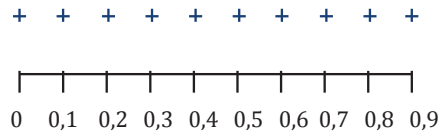


FIGURE 2.11 – Échantillon de 10 points équirépartis dans  $[0, 1]^d$  avec  $d = 1$ .

( $d = 2$ ). Plusieurs définitions de l'uniformité sont maintenant possibles. Une première définition de l'uniformité dans ce carré unité peut s'énoncer de la façon suivante : "une suite  $S = (\mathbf{x}_i)_{1 \leq i \leq N}$  de  $N$  points est dite optimalement distribuée si elle minimise la borne supérieure sur les rectangles  $R \subset [0, 1]^2$  de la déviation définie par :

$$|\text{Card}(R, S) - N\mathcal{A}(R)|,$$

où  $\text{Card}(R, S)$  est le nombre d'indices,  $1 \leq i \leq N$ , tels que  $x_i \in R$  et  $\mathcal{A}(R)$  l'aire du rectangle." Cette définition permet de juger peu uniforme une suite pour laquelle il est possible d'exhiber un rectangle avec une forte déviation. Une forte déviation correspond à un nombre de points dans le rectangle, considéré fortement différent du nombre de points totaux contenus dans le carré unité multiplié par l'aire du rectangle.

Une première mesure de la discrédance, dite discrédance centrée, consiste alors à considérer l'ensemble des hyperrectangles ayant l'origine pour sommet. Plus précisément, une suite  $S = (\mathbf{x}_i)_{1 \leq i \leq N}$  est considérée uniformément répartie dans  $[0, 1]^d$  si pour tout  $\mathbf{a} = (a_1, a_2, \dots, a_d)$  de  $[0, 1]^d$  :

$$\lim_{N \rightarrow +\infty} \frac{1}{N} \sum_{i=1}^n \mathbf{1}_{\mathbf{x}_i \in P} = \lambda(P),$$

où  $P = \prod_{j=1}^d [0, a_j]$  et  $\lambda$  sa mesure de Lebesgue ( $\lambda(P) = \prod_{i=1}^d a_i$ ).

La discrédance centrée  $D^*$  est alors définie par :

$$D^*(S) = \sup_{P \in I^d} \left| \frac{\text{Card}(P, S)}{N} - \lambda(P) \right|. \quad (2.8)$$

La détermination de cette quantité revient à chercher l'hyperrectangle ancré en zéro, dont la proportion de points contenus dans cet hyperrectangle est la plus éloignée de sa mesure de Lebesgue  $\lambda$  (cf. figure 2.12).

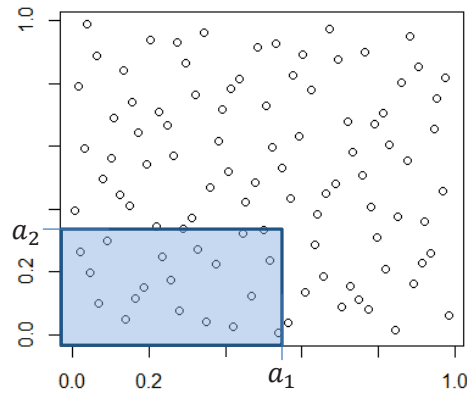


FIGURE 2.12 – Discrédance centrée ( $d = 2$ ).

Cette définition peut être élargie à tous les hyperrectangles de l'hypercube ne contenant pas nécessairement l'origine (cf. figure 2.13). On note  $R^d$  cet ensemble. Cette discrédance dite extrême est définie par :

$$D(S) = \sup_{P \in R^d} \left| \frac{\text{Card}(P, S)}{N} - \lambda(P) \right|. \quad (2.9)$$

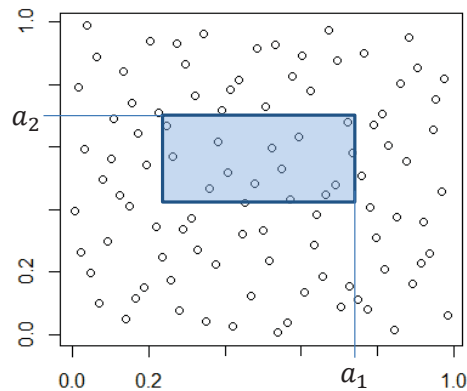


FIGURE 2.13 – Discrédance extrême ( $d = 2$ ).

Les assertions suivantes sont équivalentes [55] :

- $S_N = (\mathbf{x}_i)_{1 \leq i \leq N}$  est une suite équirépartie dans  $[0, 1]^d$ ,
- $\lim_{N \rightarrow +\infty} D^*(S_N) = 0$ ,
- $\lim_{N \rightarrow +\infty} D(S_N) = 0$ .

Ces deux discrédances sont des critères difficilement calculables, car la borne supérieure (2.8) et (2.9) est difficilement atteignable.

Cette mesure reste cependant très utile et elle est la plupart du temps utilisée pour comprendre le comportement de différentes constructions. Par exemple, le concept de discrédance est utilisé pour trouver une borne supérieure à l'erreur d'intégration via la

formule de Koksma-Hlawka. Ce résultat est largement cité [11]. Nous le mentionnons ci-dessous.

Soit  $G$  une fonction à variation  $V(G)$  finie au sens de Hardy and Krause [71], alors  $\forall N \geq 1$ , on a la formule de Koksma-Hlawka :

$$\left| \int_{[0,1]^d} G(\mathbf{x}) d\mathbf{x} - \frac{1}{N} \sum_{i=1}^N G(x_i) \right| \leq V(G) D^*(S_N).$$

Ainsi, l'erreur d'approximation est dans le pire des cas égale au produit de la variation  $V(G)$  de  $G$  (une grandeur qui ne reflète que l'irrégularité de la fonction  $G$ ) et de la discrédance  $D^*$  (qui mesure uniquement la qualité de la répartition de la suite). Contrairement aux suites aléatoires qui fournissent des intervalles de confiance pour une probabilité donnée, cette majoration est déterministe [27].

#### 2.4.4.2 Suites à faible discrédance

Des suites à faible discrédance centrée  $D^*$  ont été construites par Van der Corput [55], Halton [37], Hammersley [39], Sobol [77]. La construction de ces différentes suites utilisées en planification d'expérience est présentée ici.

##### Suite de Van der Corput

Cette suite est uni-dimensionnelle et sert de base pour construire la suite de Halton en plusieurs dimensions. On se place sur l'intervalle  $[0, 1]$  et on cherche à construire une suite équirépartie comprenant au total  $N = 16$  points. Comme nous l'avons vu (cf. figure 2.11), un choix idéal en terme d'équirépartition est de construire l'échantillon  $S = \left\{ x_i = \frac{i}{N}, 0 \leq i < N \right\}$ . Cet échantillon est optimal si les  $N = 16$  points sont effectivement tirés. En effet, si finalement seulement les 8 premiers points de cette suite croissante sont tirés, seule la première moitié de l'intervalle est explorée par l'échantillon :  $\forall i = 0, \dots, 7, x_i < 0.5$ .

Van der Cortput a proposé une méthode permettant de tirer des points uniformément répartis quel que soit le nombre de points final de l'échantillon. L'idée de Van der Corput a été d'écrire ces chiffres dans une base quelconque  $b$ , puis d'inverser l'ordre des bits. Le chiffre le plus significatif en écriture inversée correspond au chiffre le moins significatif en écriture non inversée. Ceci a pour conséquence que les termes successifs de la suite alternent entre la première et la seconde partie de l'intervalle  $[0, 1]$ . Ce comportement est illustré dans la table 2.6 pour la base 2.

La formalisation de l'écriture générale inversée d'un nombre en base  $b$  est donnée ci-après,  $\forall j = 0, \dots, N - 1$ ,

$$VP_{j,b} = \sum_{i=0}^{\infty} \frac{a_i(j)}{b^{i+1}},$$

$i$	Suite croissante	Écriture en binaire	Écriture en binaire inversé	Suite de Van der Corput en base 2
0	0	0.0000	0.0000	0
1	1/16	0.0001	0.1000	1/2
2	1/8	0.0010	0.0100	1/4
3	3/16	0.0011	0.1100	3/4
4	1/4	0.0100	0.0010	1/8
5	5/16	0.0101	0.1010	5/8
6	3/8	0.0110	0.0110	3/8
7	7/16	0.0111	0.1110	7/8
8	1/2	0.1000	0.0001	1/16
9	9/16	0.1001	0.1001	9/16
10	5/8	0.1010	0.0101	5/16
11	11/16	0.1011	0.1101	13/16
12	3/4	0.1100	0.0011	3/16
13	13/16	0.1101	0.1011	11/16
14	7/8	0.1110	0.0111	7/16
15	15/16	0.1111	0.1111	15/16

TABLE 2.6 – Construction de la suite de Van der Corput en base  $b = 2$ .

avec  $0 < a_i(j) < b$  le  $i^{\text{ème}}$  digit dans le développement en base  $b$  inversé de  $j$ . Cette suite est finie car il existe un  $m$  pour lequel tous les  $a_i(j) = 0, \forall j > m$ . La suite de Van der Corput en base  $b$  est  $(VP_{j,b})_{j=0, \dots, N-1}$ . Cette suite a l'avantage d'être facile à implémenter et d'avoir un temps d'exécution très faible. Pour passer de  $x_j = VP_{j,b}$  à  $x_{j+1} = VP_{j+1,b}$ , il suffit d'effectuer une addition de  $\frac{1}{b}$  en base  $b$ . Par exemple, on ajoute 0.1 en base 2 pour obtenir le nombre suivant de la suite (cf. table 2.6).

Le choix de base le plus commun est  $b = 2$ . La discrédance centrée  $D^*$  de cette suite en base 2 est de l'ordre de  $\mathcal{O}\left(\frac{\log N}{N}\right)$ .

### Suite de Halton

La suite de Halton (HN) est une généralisation multi-dimensionnelle de la suite de Van der Corput. Elle consiste à prendre une base différente pour chaque dimension. Soient  $b_1, \dots, b_d$  les  $d$  premiers nombres premiers supérieurs ou égaux à deux, alors si  $\forall j = 0, \dots, N-1$ ,

$$HN_j = (VP_{j,b_1}, \dots, VP_{j,b_d}),$$

la suite HN est une suite de Halton de taille  $N$ .

Lorsque les bases  $(b_1, b_2, \dots, b_d)$  sont prises dans les nombres premiers, la discrédance est diminuée [70]; elle est de l'ordre dans ce cas de  $\mathcal{O}\left(\frac{\log(N)^d}{N}\right)$ .

En grande dimension, lorsque les points de cette suite sont projetés dans le carré unité, certains comportements pathologiques peuvent être observés tels que des recouvrements de points ou l'apparition de zones lacunaires. Nous avons généré  $N = 50$  points dans l'hypercube  $\mathbb{H}^{15}$  puis nous avons projeté ces points dans des plans. Plus la dimension des variables d'entrée augmente, plus ces pathologies sont visibles (cf. figures 2.14, 2.15, 2.16). Ceci résulte de la régularité des sous-suites de Van der Corput. Des méthodes dites de *scrambling* [81] permettent d'améliorer la répartition sur les projections 2D. Toutefois, en augmentant suffisamment le nombre de points  $N$  lorsque cela est possible, l'espace est suffisamment rempli pour ne plus observer ces pathologies.

### Suite de Hammersley

La suite de Hammersley est construite à partir de la suite de Halton :  $\forall j = 0 \dots, N-1$ ,

$$HY_j = \left(\frac{j}{N}, VP_{j,b_1}, \dots, VP_{j,b_{d-1}}\right).$$

L'avantage de cette suite est que le taux de convergence de la borne pour la discrédance centrée est un peu plus rapide (elle dépend de  $d-1$  au lieu de  $d$ ). Cependant, il faut connaître le nombre de points à générer à l'avance.

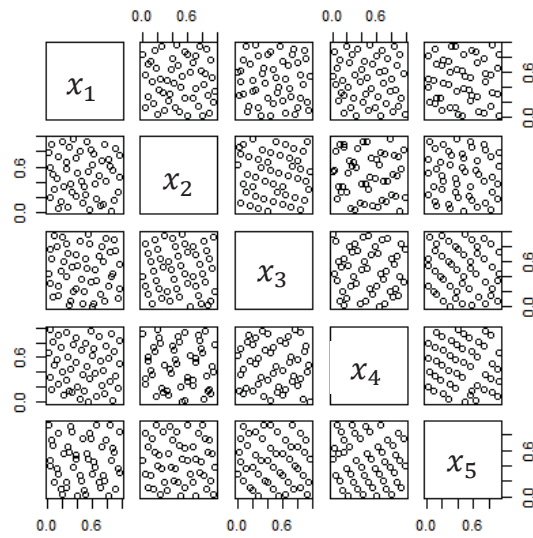


FIGURE 2.14 – Projections 2D d'une suite de Halton ( $N = 50$ ,  $d = 15$ ) de la dimension 1 à la dimension 5.

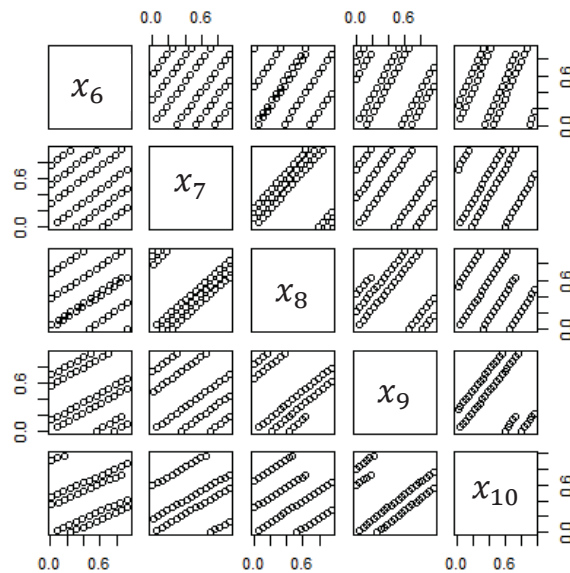


FIGURE 2.15 – Projections 2D d'une suite de Halton ( $N = 50$ ,  $d = 15$ ) de la dimension 6 à la dimension 10.

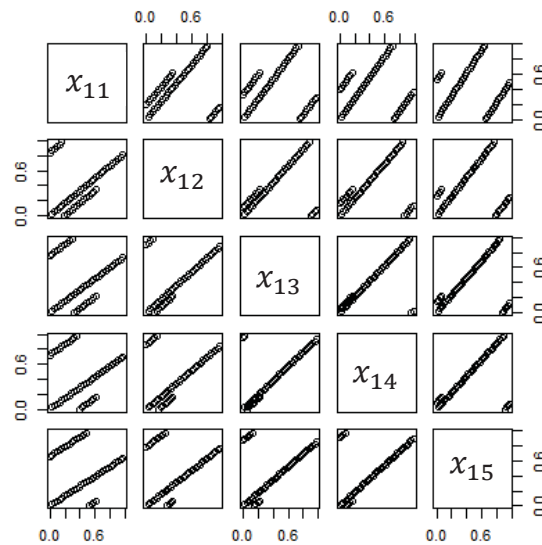


FIGURE 2.16 – Projections 2D d'une suite de Halton ( $N = 50$ ,  $d = 15$ ) de la dimension 11 à la dimension 15.

### Suite de Sobol

Nous ne détaillerons pas ici la construction de cette suite. Le lecteur intéressé pourra consulter [31]. Elle présente les mêmes problèmes en projection que la suite de Halton lorsque la dimension  $d$  augmente. Les techniques de *scrambling* peuvent alors également être mises en oeuvre [72]. La discrédance d'une suite de Sobol notée (SL) est de l'ordre de  $\mathcal{O}\left(\frac{(\log N)^d}{N}\right)$ .

### Conclusions sur les suites à faible discrédance

Grâce à ces suites, l'efficacité de la méthode de Monte Carlo peut être améliorée par réduction de la discrédance. On parle de méthode de quasi-Monte Carlo.

#### Avantages

- Les suites présentées dans cette section ont été construites de sorte qu'elles minimisent la discrédance.
- Pour les suites de Van der Corput en dimension 1, de Halton et de Sobol en dimension  $d$ , il n'est pas nécessaire de connaître *a priori* le nombre de points  $N$  nécessaires (contrairement à la grille régulière et au LHS).
- Lorsque le domaine n'est pas hypercubique, ces méthodes permettent également d'assurer une bonne répartition. Ces plans sont construits par une méthode dite de rejet, c'est-à-dire que la suite est générée dans l'hypercube puis les points hors domaine sont supprimés du plan d'expérience.

**Inconvénients**

- En grande dimension, si le nombre de points n'est pas suffisant, on peut ne pas remplir correctement l'espace et des singularités sont observées en projection.
- Une fois le plan d'expérience construit avec la suite de Hammersley, il n'est pas facile de rajouter de nouveaux points.

**2.4.5 Critères géométriques pour l'amélioration du remplissage de l'espace**

Comme nous l'avons vu, un bon remplissage de l'espace d'un plan d'expérience permet de garantir une bonne exploration du modèle. Afin de juger de ce bon remplissage, Johnson et al [50] ont introduit d'autres critères qui sont basés sur des distances entre points. On note  $\text{dist}(\mathbf{x}_i, \mathbf{x}_k)$ , la distance euclidienne entre deux points  $\mathbf{x}_i$  et  $\mathbf{x}_k$  telle que :

$$\text{dist}(\mathbf{x}_i, \mathbf{x}_k) = \left[ \sum_{j=1}^d (x_{ij} - x_{kj})^2 \right]^{1/2}.$$

**Plan Maximin**

On peut vouloir maximiser la plus petite distance  $\text{dist}(\mathbf{x}_i, \mathbf{x}_k)$  entre deux points quelconques du plan d'expérience. Le critère de distance minimale entre deux points du plan d'expérience, nommé critère *mindist* et noté  $\phi_{Mm}$  vaut :

$$\phi_{Mm}([\bar{\mathbf{x}}]) = \min_{\substack{i,k=1,\dots,N \\ i \neq k}} \text{dist}(\mathbf{x}_i, \mathbf{x}_k).$$

Le plan d'expérience dit maximin est celui qui maximise  $\phi_{Mm}$ . Plus la valeur du critère *mindist* est élevée, plus les points sont espacés, ceci permettant une meilleure occupation de l'espace. En grande dimension, ce critère a tendance à ne pas placer de points au centre du domaine mais essentiellement sur les bords.

**Plan Minimax**

On veut cette fois minimiser les zones vides de l'espace. Ce critère est basé sur la distance maximale entre un point quelconque du domaine et un point du plan d'expérience. Il est nommé critère *maxdist* et noté  $\phi_{mM}$  vaut :

$$\phi_{mM}([\bar{\mathbf{x}}]) = \max_{\mathbf{x} \in \mathbb{H}^d} \min_{\mathbf{x}_k \in [\bar{\mathbf{x}}]} \text{dist}(\mathbf{x}, \mathbf{x}_k).$$

Le plan d'expérience dit minimax est celui qui minimise  $\phi_{mM}$ , il est plus difficile à construire qu'un plan maximin et pas forcément de meilleure qualité [31].

### Plans optimisés pour remplir au mieux l'espace

On peut par exemple construire des plans d'expérience LHS qui optimisent les critères de discrédance ou de maximin et ainsi optimiser le remplissage de l'espace (cf. figure 2.17) [73].

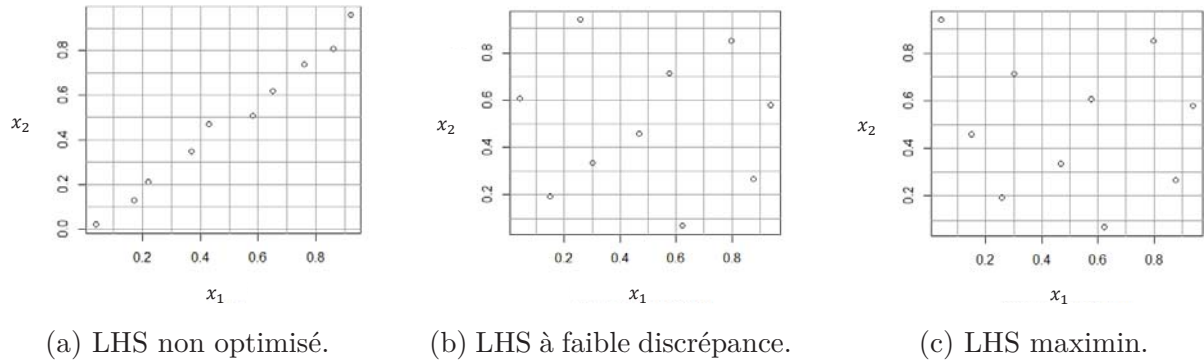


FIGURE 2.17 – Exemples de LHS optimisés.

#### Avantage

- Ces plans optimisés peuvent être utilisés pour s'assurer du bon remplissage de l'espace dans le cas non hypercubique.

#### Inconvénient

- La génération peut être très coûteuse lorsque le nombre de points est important car elle nécessite l'utilisation d'algorithmes d'optimisation.

## 2.5 Méthodes de validation

Un modèle réduit est construit ici à partir d'un échantillon d'apprentissage. Le modèle de référence est donc connu de façon discrète. Ainsi, il est impossible de connaître la véritable erreur d'un modèle en tout point, mais on peut par contre en faire des estimations à partir des données disponibles.

Si nous sommes capables de générer un nombre suffisant de points de la fonction de référence, il est intéressant de fractionner cet ensemble en trois échantillons indépendants : un échantillon d'apprentissage noté  $([\bar{x}_{\text{app}}], \mathbf{Y}_{\text{app}})$  contenant  $N_a$  valeurs de la sortie  $F$ , un échantillon de validation noté  $([\bar{x}_{\text{val}}], \mathbf{Y}_{\text{val}})$  contenant  $N_v$  valeurs et un échantillon de test noté  $([\bar{x}_{\text{test}}], \mathbf{Y}_{\text{test}})$  contenant  $N_t$  valeurs ( $N = N_a + N_t + N_v$ ). L'échantillon d'apprentissage est utilisé pour construire le modèle réduit et calculer l'erreur d'apprentissage. L'échantillon de validation est également utilisé en cours de construction du modèle réduit et permet d'estimer l'erreur de validation. L'échantillon de test est quant à lui utilisé une fois le modèle réduit construit, pour estimer l'erreur de représentativité du modèle.

Il est difficile de donner *a priori* la proportion de points dans chacun de ces échantillons. Un partage typique se fait ainsi : 50 % des points dans l'échantillon d'apprentissage, 25 % dans l'échantillon de validation et 25 % dans l'échantillon de test [42]. Lorsqu'on ne peut disposer que d'un nombre réduit de points, la répartition en trois lots fait encore diminuer le nombre de points de l'échantillon d'apprentissage. Dans les deux sections suivantes, nous allons présenter des méthodes qui réutilisent de façon efficace un même échantillon pour obtenir une estimation de l'erreur de généralisation (validation croisée et bootstrap).

Les paramètres d'un modèle réduit sont optimisés dans le but d'approcher au mieux le modèle de référence qui n'est connu que sur les échantillons. L'erreur estimée sur l'échantillon d'apprentissage est évidemment nulle pour les modèles réduits qui interpolent ces données. Pour les modèles dits de régression, cette erreur d'apprentissage diminue systématiquement avec la complexification du modèle réduit, pouvant atteindre zéro si le modèle réduit est suffisamment complexe. Or, lorsqu'un modèle trop complexe est construit, l'erreur de validation augmente. Un modèle avec une erreur d'apprentissage nulle peut être sur-ajusté sur les données d'apprentissage et donc avoir une faible capacité de généralisation. Ceci correspond à l'obtention d'un modèle qui approche très bien les échantillons d'apprentissage mais oscille entre ces points, alors que la fonction de référence est elle-même lisse. Ce phénomène est observé lorsqu'on dépasse un seuil de complexité  $\theta$  sur la figure 2.18. Un compromis doit donc être trouvé entre complexité du modèle réduit et

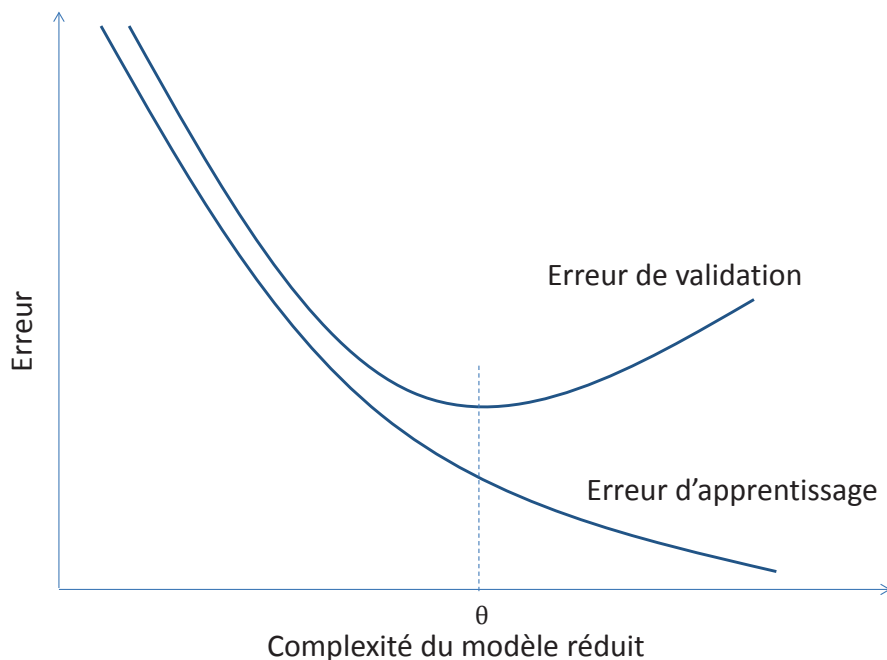


FIGURE 2.18 – Erreurs en fonction de la complexité du modèle.

capacité de généralisation. On parle de compromis biais-variance [49]. Un modèle ayant trop peu de paramètres commet des erreurs trop importantes en raison de son biais im-

portant (pas assez de souplesse), on parle de sous-apprentissage. Un modèle ayant trop de paramètres commet des erreurs importantes en raison de sa variance importante (trop sensible aux détails de l'échantillon d'apprentissage), on parle de sur-apprentissage.

La considération de l'erreur d'apprentissage seule n'est pas suffisante pour obtenir un modèle régressant de qualité. Dans le but d'obtenir un modèle réduit assurant le meilleur compromis biais-variance, les calculs de l'erreur d'apprentissage et de validation sont nécessaires [53, 82]. Dans le cas où le nombre de points n'est pas suffisant, des méthodes de validation croisée ou de bootstrap sont conseillées.

### 2.5.1 Validation croisée

#### Validation croisée classique

L'erreur de généralisation peut être estimée lors de l'apprentissage grâce à un échantillon dit de validation comme introduit précédemment. On parle alors de validation croisée. Durant l'optimisation des paramètres du modèle avec l'échantillon d'apprentissage, on évalue l'erreur sur l'échantillon de validation en comparant les valeurs prédites aux valeurs du modèle de référence en ces points. Lorsque cette erreur commence à augmenter, c'est signe que l'apprentissage peut être arrêté (cf. figure 2.18).

#### Avantages

- On augmente la capacité de généralisation du modèle réduit.
- Cette méthode peut être appliquée à tout type de modèle.

#### Inconvénient

- On réduit le nombre de points disponibles dans l'ensemble d'apprentissage.

#### Validation croisée à $k$ sous-ensembles

Lorsque le nombre d'échantillons disponibles pour la construction du modèle est limité, il peut être intéressant de tous les garder pour la construction du modèle réduit. On peut alors utiliser une stratégie dite de validation croisée à  $k$  sous-ensembles. Cette méthode repose sur la validation croisée classique présentée ci-dessus, mais elle est appliquée de manière répétée avec différents ensembles de points appartenant tous à l'échantillon d'apprentissage. L'échantillon d'apprentissage est ainsi séparé en  $k$  différents groupes de tailles égales. Par exemple pour  $k = 5$ , la première étape consiste à séparer l'échantillon d'apprentissage comme indiqué dans la table 2.7.

	1	2	3	4	5
$i = 1$	<b>Validation</b>	Apprentissage	Apprentissage	Apprentissage	Apprentissage

TABLE 2.7 – Validation croisée à 5 sous-ensembles.

On construit le modèle sur les  $k - 1$  parties des données d'apprentissage (sans la partie **Validation** de la table 2.7) et on calcule l'erreur de généralisation du modèle réduit en utilisant les données de la 1<sup>ère</sup> partie. On répète ce processus pour  $i = 2, \dots, k$  comme dans la table 2.8 (avec  $k = 5$ ).

	1	2	3	4	5
$i = 2$	Apprentissage	<b>Validation</b>	Apprentissage	Apprentissage	Apprentissage
$i = 3$	Apprentissage	Apprentissage	<b>Validation</b>	Apprentissage	Apprentissage
$i = 4$	Apprentissage	Apprentissage	Apprentissage	<b>Validation</b>	Apprentissage
$i = 5$	Apprentissage	Apprentissage	Apprentissage	Apprentissage	<b>Validation</b>

TABLE 2.8 – Validation croisée à 5 sous-ensembles.

On moyenne les  $k$  estimations obtenues de l'erreur de prédiction. Ces  $k$  modèles construits peuvent ensuite être agrégés afin de rassembler toutes les informations apprises en utilisant pour la prédiction d'un point la valeur moyenne des sorties de ces  $k$  modèles.

La validation croisée *leave-one-out* correspond au cas particulier où  $k = N$ . Le modèle réduit est construit sur  $N - 1$  observations et validé sur l'observation non utilisée pour l'apprentissage. Ce procédé est ensuite répété  $N$  fois. Cette méthode, qui nécessite de construire  $N$  modèles réduits, atteint vite ses limites du fait du temps de calcul qui peut vite devenir prohibitif dès lors que  $N$  est grand et/ou la durée d'apprentissage importante.

En pratique,  $k = 5$  ou  $k = 10$  sont les valeurs les plus souvent utilisées. La valeur finalement retenue dépend de la convergence de l'erreur d'apprentissage, en fonction de la taille de l'échantillon d'apprentissage comme montré dans [42].

### Avantage

- On utilise toutes les données disponibles pour construire le modèle réduit.

### Inconvénient

- Le modèle réduit est plus long à construire.

## 2.5.2 Bootstrap

La méthode de bootstrap est un outil général qui permet d'évaluer le biais de l'estimation de l'erreur [5]. Comme pour la validation croisée, on cherche à estimer l'erreur de généralisation. Supposons que nous disposons d'un échantillon d'apprentissage noté  $([x_{app}], \mathbf{Y}_{app})$  de taille  $N$ . L'idée est de créer  $N_b$  échantillons de données de taille  $N$  à partir de l'échantillon d'apprentissage par tirage aléatoire avec remise. On construit alors  $N_b$  modèles réduits de la fonction de référence.

En un point  $\mathbf{x}$  donné, on a donc un ensemble de valeurs  $\tilde{F}_i(\mathbf{x})$  qui forme une distribution empirique des valeurs possibles de  $F(\mathbf{x})$ . Plus  $N_b$  est grand, plus cette distribution est connue avec précision.

La fonction  $\tilde{F}$  peut être vue comme un processus stochastique. On peut alors estimer en tout point  $\mathbf{x}$  la moyenne de ce processus qui vaut :

$$\mu_{N_b}(\mathbf{x}) = \frac{1}{N_b} \sum_{i=1}^{N_b} \tilde{F}_i(\mathbf{x}),$$

et sa variance qui vaut :

$$\sigma_{N_b}^2(\mathbf{x}) = \frac{1}{N_b - 1} \sum_{i=1}^{N_b} \left( \tilde{F}_i(\mathbf{x}) - \mu_{N_b} \right)^2.$$

En prenant la moyenne des modèles bootstrap, on obtient une réalisation du modèle réduit et donc une estimation de la fonction de référence, ainsi qu'une indication sur la précision du calcul grâce à la variance. Plus la variance est grande, moins le modèle obtenu est précis.

On détermine alors un intervalle de confiance  $IC = [IC_{inf}, IC_{sup}]$  pour la prédiction du modèle réduit  $\tilde{F}$ . Cet intervalle est déterminé soit à partir de la moyenne et de la variance en faisant une approximation gaussienne comme dans le cas du krigeage (2.3), soit en calculant directement cet intervalle à partir de la distribution empirique des échantillons bootstrap.

### Avantages

- Le bootstrap permet d'estimer des intervalles de confiance pour les prédictions de n'importe quel type de modèle réduit.
- La connaissance de l'erreur des  $n_b$  modèles réduits permet de déterminer les zones de l'espace qui sont moins bien décrites et ainsi de densifier ces zones du plan d'expérience.

### Inconvénient

- L'inconvénient de cette méthode est son coût calculatoire élevé, car elle nécessite de construire un nombre important de modèles pour avoir des estimations fiables.

### 2.5.3 Estimation *a posteriori*

L'erreur globale d'un modèle réduit peut être estimée à partir de l'erreur commise sur les échantillons d'apprentissage ou de validation comme nous venons de le voir dans la section 2.5.1. Mais cette estimation de l'erreur reste biaisée car tous ces échantillons ont participé à l'établissement du modèle. Pour estimer *a posteriori* l'erreur de généralisation, il est intéressant de disposer d'un échantillon de test  $([\bar{x}_{\text{test}}], \mathbf{Y}_{\text{test}})$  indépendant des deux échantillons précédents. Plus le nombre  $N_t$  d'échantillons test est grand, plus cette estimation est précise. À iso-nombre de points, le problème est alors de savoir s'il est plus judicieux d'utiliser les points pour améliorer la construction du modèle en les affectant à l'échantillon d'apprentissage ou pour améliorer la précision de l'estimation de l'erreur en les affectant à l'échantillon de validation. De plus, l'utilisation d'un échantillon de test permet de déterminer l'erreur globale d'un modèle, mais pas son erreur locale de prédiction en un point donné (en dehors des points de test). Pour cela, la méthode de bootstrap, détaillée dans la section précédente, peut être utilisée.

## 2.6 Choix des méthodes pour cette étude

Les modèles de référence QdV et Perfo à réduire comprennent 6 à 15 variables d'entrée réelles. Ils sont composés d'un assemblage de fonctions tabulées de 1 à 4 variables d'entrée. Ces modèles sont établis sur un domaine non hypercubique des variables d'entrée appelé "domaine de vol", limité par des contraintes physiques (plafond de propulsion, décrochage, résistance structurale, etc...).

Les fonctions tabulées constituant ces modèles de référence sont, quant à elles, stockées soit sur une grille complète, auquel cas leur domaine de définition est hypercubique, soit sur une grille tronquée, ce qui entraîne la non-hypercubicité du domaine de définition. Dans cette thèse, nous avons été amenés à réduire d'une part les modèles globaux QdV et Perfo et d'autre part les fonctions tabulées de 1 à 4 variables d'entrée qui les composent.

### 2.6.1 Choix des modèles réduits

#### Random Sampling - High Dimensional Model Representation

Nous avons finalement retenu comme méthode de réduction automatique le RS-HDMR qui permet de décomposer automatiquement le modèle de référence en une somme de fonctions dépendant d'un nombre réduit de variables. On retrouve naturellement la structure actuelle de nos modèles. On peut alors choisir de tabuler les fonctions de la décomposition RS-HDMR et ainsi utiliser les méthodes d'interpolation multi-linéaire, comme cela est fait actuellement, pour calculer la sortie d'une fonction en un point non tabulé. Cette méthode a été retenue même si le calcul de chacune de ces valeurs tabulées nécessite d'intégrer numériquement une intégrale multi-dimensionnelle : ceci peut vite devenir très

coûteux en temps de calcul, comme nous le verrons par la suite. De plus, cette méthode ne s'applique pas directement aux variables d'entrée dépendantes. Le chapitre 3 propose des solutions permettant d'adapter au mieux cette méthode à notre problématique.

### Réseaux de neurones

Nous avons également retenu les RN pour leur propriété d'approximateur universel parcimonieux [47, 22]. Ils nécessitent moins de paramètres que d'autres modèles réduits pour approcher une fonction continue avec une précision arbitraire [4, 48]. Par ailleurs, nous sommes particulièrement intéressés par les propriétés des fonctions tabulées. L'erreur d'un réseau de neurones, est connue en général de manière statistique. Les RN sont donc utilisés dans cette thèse pour régénérer les points d'une grille et des interpolations muti-linéaires sont réalisées entre les points. Cette méthode sera décrite plus en détail dans le chapitre 3.

### 2.6.2 Choix des méthodes d'échantillonnage

Pour les modèles de référence QdV et Perfo, nous avons utilisé indifféremment les suites de Halton ou de Sobol pour échantillonner la fonction de référence, car elles permettent d'obtenir des résultats similaires. Le domaine étant non hypercubique et de grande dimension, il n'est pas possible d'estimer la qualité de l'échantillon généré. Toutefois, l'utilisation de ces suites quasi-aléatoires permet d'espérer un meilleur remplissage de l'espace que l'utilisation de suites pseudo-aléatoires. Cet échantillon est ensuite partitionné en un échantillon d'apprentissage et un échantillon de test. Concernant les fonctions tabulées, leurs points de définition sont utilisés pour un échantillonnage déterministe de la fonction ce qui nous assure un bon remplissage de l'espace. Ces échantillons peuvent être complétés par des points issus de suites quasi-aléatoires (suites de Halton et de Sobol).

### 2.6.3 Choix des méthodes de validation

Le modèle réduit RS-HDMR est validé par une estimation *a posteriori* de l'erreur (cf. section 2.5.3). L'échantillon généré est donc divisé en un ensemble d'apprentissage et un ensemble de test.

Concernant les cas de petite dimension, nous verrons que dans nos applications, seul l'échantillon d'apprentissage formé par les points tabulés de la fonction de référence est nécessaire pour valider le modèle réduit.



# Chapitre 3

## Adaptation des méthodes de réduction aux modèles de qualités de vol et de performances des avions

### 3.1 Introduction

Le modèle aérodynamique QdV décrit au chapitre 2 possède entre 14 (calculs Basse Vitesse, dit BV) et 15 (calculs Grande Vitesse, dit GV) variables d'entrée. Le modèle Perfo, selon la fonction embarquée considérée, a entre 6 et 15 variables d'entrée. Leurs caractéristiques communes sont rappelées ci-après.

- Leurs domaines de définition ne sont pas hypercubiques.
- Ils sont composés essentiellement de fonctions tabulées dans lesquelles des interpolations multi-linéaires sont réalisées (cf. figure 1.16).
- Ils ont une structure dite hiérarchique multi-niveaux (cf. figure 1.17).

Les spécificités des modèles réduits en fonction de leur application respective ont été décrites dans l'étape 2 du processus de construction d'un modèle réduit (cf. section 2.2).

Nous avons retenu dans le chapitre précédent deux méthodes de réduction de modèles : le Random Sampling - High Dimensional Model Representation (RS-HDMR) et les réseaux de neurones (RN) dont on peut s'inspirer pour construire de nouvelles méthodes de réduction.

Dans la section 3.2, nous développons d'une part une méthode de calcul analytique des différents termes de la décomposition RS-HDMR et d'autre part une extension du calcul de cette décomposition au cas des fonctions définies sur un domaine non hypercubique. Finalement, un emploi original des RN est proposé pour régénérer uniquement les points de la fonction tabulée.

Dans la section 3.3, nous étudions l'applicabilité des méthodes retenues aux modèles de référence complets QdV et Perfo.

## 3.2 Réduction des fonctions tabulées

Les fonctions tabulées qui composent les modèles de référence représentent l'essentiel du temps de calcul et du coût de stockage. Elles ont la structure suivante :

- elles peuvent avoir de 1 à 4 variables d'entrée ;
- selon la fonction, le domaine peut être hypercubique ou non ;
- une interpolation multi-linéaire est réalisée entre les points tabulés (cf. figure 1.16).

On cherche ici à réduire le stockage et le temps de calcul de ces fonctions tabulées. L'erreur commise en faisant cette réduction est estimée sur l'ensemble des points de définition de la fonction tabulée originale.

### 3.2.1 Random Sampling - High Dimensional Model Representation, cas d'un domaine de définition hypercubique

Nous avons introduit la décomposition RS-HDMR dans la section 2.3.4 sur le domaine  $\mathbb{H}^d = [0, 1]^d$ .

Nous allons tout d'abord donner quelques notations utilisées dans cette section. À un ordre de troncature donné  $m$ , le nombre de termes est égal au nombre de combinaisons de  $m$  variables parmi  $d$ , noté  $\binom{d}{m}$ . Chacune de ces  $\binom{d}{m}$  combinaisons est indexée par  $p$ ,  $1 \leq p \leq \binom{d}{m}$  et est représentée par l'ensemble  $V_m^p$  des  $m$  indices correspondants. On note  $U_m$  l'ensemble des ensembles de combinaisons possibles pour un ordre  $m$ . Par exemple, pour  $d = 3$  et  $m = 2$ , l'ensemble  $U_2$  est constitué des trois paires  $V_2^1 = \{1, 2\}$ ,  $V_2^2 = \{1, 3\}$ ,  $V_2^3 = \{2, 3\}$  :

$$U_2 = \left\{ \{1, 2\}, \{1, 3\}, \{2, 3\} \right\}.$$

On note aussi  $\bar{V}_m^p$  le complémentaire de  $V_m^p$  dans l'ensemble  $\{1, \dots, d\}$  et  $\mathbf{x}_{V_m^p} = (x_i)_{i \in V_m^p}$ . La décomposition peut donc être réécrite :

$$\begin{aligned} F(x_1, \dots, x_d) &= Z_0 + \sum_{V_1^p \in U_1} Z_{V_1^p}(\mathbf{x}_{V_1^p}) + \sum_{V_2^p \in U_2} Z_{V_2^p}(\mathbf{x}_{V_2^p}) + \dots \\ &+ \dots + \sum_{V_d^p \in U_d} Z_{V_d^p}(\mathbf{x}_{V_d^p}) \\ &= Z_0 + \sum_{m=1}^d \sum_{V_m^p \in U_m} Z_{V_m^p}(\mathbf{x}_{V_m^p}). \end{aligned} \quad (3.1)$$

La *vanishing condition* (2.6) peut également être réécrite :  $\forall m \in \{1, \dots, d\}, \forall s \in V_m^p$ ,

$$\int_{[0,1]} Z_{V_m^p}(\mathbf{u}_{V_m^p}) du_s = 0. \quad (3.2)$$

*CHAPITRE 3. ADAPTATION DES MÉTHODES DE RÉDUCTION AUX  
MODÈLES DE QUALITÉS DE VOL ET DE PERFORMANCES DES AVIONS*

Les composants de la décomposition RS-HDMR sont ainsi déterminés de manière unique en imposant la *vanishing condition* (3.2). On note  $I_m^p$  l'intégrale permettant de calculer le  $p^{\text{ème}}$  terme d'ordre  $m$  de cette décomposition.

Pour  $m = 0$ , on compte une combinaison ( $p = 1$ ), c'est-à-dire une intégrale;  $V_0^1 = \emptyset$  et  $\bar{V}_0^1 = \{1, \dots, d\}$  :

$$I_0^1 = Z_0 = \int \prod_{k \in \bar{V}_0^1} [0, 1] F(\mathbf{u}) \prod_{k \in \bar{V}_0^1} du_k.$$

Une fois le terme d'ordre 0 évalué, on obtient un terme  $p$  d'ordre 1 :

$$Z_{V_1^p}(\mathbf{x}_{V_1^p}) = I_1^p(\mathbf{x}_{V_1^p}) - Z_0.$$

Pour cela on évalue  $I_1^p$ , l'intégrale d'ordre 1. Pour  $m = 1$ , on compte  $\binom{d}{1}$  combinaisons, c'est-à-dire que  $p$  varie de 1 à  $d$ . On doit donc déterminer  $d$  intégrales qui dépendent chacune d'une variable :

$$I_1^p(\mathbf{x}_{V_1^p}) = \int \prod_{k \in \bar{V}_1^p} [0, 1] F(\mathbf{x}_{V_1^p}, \mathbf{u}_{\bar{V}_1^p}) \prod_{k \in \bar{V}_1^p} du_k.$$

Une fois tous les termes d'ordre 1 obtenus, on peut calculer un terme  $p$  d'ordre 2,

$$Z_{V_2^p}(\mathbf{x}_{V_2^p}) = I_2^p(\mathbf{x}_{V_2^p}) - Z_0 - \sum_{V_1^{p'} \subset V_2^p} Z_{V_1^{p'}}(\mathbf{x}_{V_1^{p'}}).$$

Pour cela on évalue l'intégrale  $I_2^p$ . Pour  $m = 2$ , on compte  $\binom{d}{2}$  combinaisons, c'est-à-dire que  $p$  varie de 1 à  $\frac{d(d-1)}{2}$ . On doit donc déterminer  $\frac{d(d-1)}{2}$  intégrales qui dépendent chacune de deux variables :

$$I_2^p(\mathbf{x}_{V_2^p}) = \int \prod_{k \in \bar{V}_2^p} [0, 1] F(\mathbf{x}_{V_2^p}, \mathbf{u}_{\bar{V}_2^p}) \prod_{k \in \bar{V}_2^p} du_k.$$

On détermine un terme  $p$  d'ordre  $m$ , en soustrayant les termes d'ordres inférieurs qui ont été déterminés au préalable :

$$Z_{V_m^p}(\mathbf{x}_{V_m^p}) = I_m^p(\mathbf{x}_{V_m^p}) - Z_0 - \sum_{j=1}^{m-1} \sum_{V_j^{p'} \subset V_m^p} Z_{V_j^{p'}}(\mathbf{x}_{V_j^{p'}}).$$

De façon générale, pour un ordre  $m \in \{1, \dots, d\}$  donné, on calcule l'intégrale suivante :

$$I_m^p(\mathbf{x}_{V_m^p}) = \int \prod_{k \in \bar{V}_m^p} [0, 1] F(\mathbf{x}_{V_m^p}, \mathbf{u}_{\bar{V}_m^p}) \prod_{k \in \bar{V}_m^p} du_k, \quad (3.3)$$

Pour alléger l'écriture, on n'écrira plus par la suite les variables dont dépendent les différentes intégrales, c'est-à-dire :  $I_m^p(\mathbf{x}_{V_m^p}) = I_m^p$ . On note que les termes de la décomposition sont obtenus de façon séquentielle.

### 3.2.1.1 Calcul analytique

Les intégrales dans la décomposition RS-HDMR peuvent être évaluées en faisant appel à des méthodes de type Monte Carlo [28] détaillées dans la section 2.4.2. Cette manière de procéder est bien adaptée dans le cas où les intégrandes ne sont pas connus analytiquement (comme c'est le cas pour les fonctions de type boîte noire). Mais on s'intéresse ici aux fonctions tabulées notées  $Ftab$  de dimension  $d$  et nous allons montrer qu'un calcul exact peut être effectué.

Pour chaque variable  $x_i$ , on considère un ensemble fini de  $J_i$  valeurs discrètes :

$$\{x_{1,1}, \dots, x_{1,J_1}\}, \{x_{2,1}, \dots, x_{2,J_2}\}, \dots, \{x_{d,1}, \dots, x_{d,J_d}\}$$

avec  $x_{1,1} = a_1, x_{1,J_1} = b_1, x_{2,1} = a_2, x_{2,J_2} = b_2, \dots, x_{d,1} = a_d, x_{d,J_d} = b_d,$

c'est-à-dire que le domaine de définition de  $Ftab$  est ainsi constitué d'une réunion de pavés élémentaires :

$$\Omega_q = [x_{1,j_1}, x_{1,j_1+1}] \times [x_{2,j_2}, x_{2,j_2+1}] \times \dots \times [x_{d,j_d}, x_{d,j_d+1}]$$

avec  $j_1 \in \{1, \dots, J_1 - 1\}, j_2 \in \{1, \dots, J_2 - 1\}, \dots,$  et  $j_d \in \{1, \dots, J_d - 1\}.$

On compte  $N_Q$  pavés avec  $N_Q = \prod_{i=1}^d \max(1, J_i - 1)$ . Pour simplifier les écritures et sans perte de généralité, on suppose que les entrées sont normalisées, c'est-à-dire,  $\forall i \in \{1, \dots, d\}, a_i = 0$  et  $b_i = 1$ . Les variables d'entrée sont ainsi définies sur l'hypercube unité  $\mathbb{H}^d = [0, 1]^d = \bigcup_{q=1}^{N_Q} \Omega_q$ .

Sur chacun de ces pavés  $\Omega_q$ , la fonction  $Ftab$  est connue. C'est une forme multilinéaire complètement définie par sa valeur en chacun des  $2^d$  sommets de chacun des  $N_Q$  pavés. Elle s'écrit analytiquement comme suit :

$$Ftab_q(\mathbf{x}) = \sum_{k=0}^{2^d-1} \alpha_{k,q} \prod_{i=1}^d x_i^{\delta_{ik}},$$

où  $\delta_{ik}$  est le  $i^{\text{ème}}$  chiffre de  $k$  écrit en base 2. Les  $2^d$  coefficients  $\alpha_{k,q}$  sont calculés grâce à la résolution du système linéaire obtenu par la connaissance des valeurs tabulées en chacun des  $2^d$  sommets du pavé  $\Omega_q$ .

De façon générale, pour un ordre  $m \in \{1, \dots, d\}$  donné, l'intégrale d'indice  $p$  s'écrit

$$I_m^p = \sum_{q=1}^{N_Q} I_{m,q}^p,$$

où  $I_{m,q}^p$  est l'intégrale sur  $\Omega_q$  :

$$I_{m,q}^p = \int \prod_{l \in \bar{V}_m^p} [x_{l,j_l}, x_{l,j_l+1}] \sum_{k=0}^{2^d-1} \alpha_{k,q} \prod_{l_1 \in V_m^p} x_{l_1}^{\delta_{l_1 k}} \prod_{l_2 \in \bar{V}_m^p} u_{l_2}^{\delta_{l_2 k}} du_{l_2}.$$

Nous détaillons maintenant le calcul des intégrales  $I_{m,q}^p$  :

$$I_{m,q}^p = \sum_{k=0}^{2^d-1} \alpha_{k,q} \prod_{l_1 \in V_m^p} x_{l_1}^{\delta_{l_1 k}} \prod_{l_2 \in \bar{V}_m^p} \int_{[x_{l_2,j_{l_2}}, x_{l_2,j_{l_2}+1}]} u_{l_2}^{\delta_{l_2 k}} du_{l_2}.$$

Deux cas se présentent alors :

– si  $\delta_{l_2 k} = 0$ , on a :

$$\int_{[x_{l_2,j_{l_2}}, x_{l_2,j_{l_2}+1}]} du_{l_2} = x_{l_2,j_{l_2}+1} - x_{l_2,j_{l_2}};$$

– si  $\delta_{l_2 k} = 1$ , on a :

$$\int_{[x_{l_2,j_{l_2}}, x_{l_2,j_{l_2}+1}]} u_{l_2} du_{l_2} = \frac{x_{l_2,j_{l_2}+1}^2 - x_{l_2,j_{l_2}}^2}{2} = (x_{l_2,j_{l_2}+1} - x_{l_2,j_{l_2}}) \left( \frac{x_{l_2,j_{l_2}+1} + x_{l_2,j_{l_2}}}{2} \right).$$

En notant  $\forall l_2 \in \bar{V}_m^p$

$$\Delta_{l_2,j_{l_2}} = x_{l_2,j_{l_2}+1} - x_{l_2,j_{l_2}} \text{ et } \mu_{l_2,j_{l_2}} = \frac{x_{l_2,j_{l_2}+1} + x_{l_2,j_{l_2}}}{2},$$

on obtient l'écriture synthétique suivante :

$$I_{m,q}^p = \sum_{k=0}^{2^d-1} \alpha_{k,q} \prod_{l_1 \in V_m^p} x_{l_1}^{\delta_{l_1 k}} \prod_{l_2 \in \bar{V}_m^p} \Delta_{l_2,j_{l_2}} \mu_{l_2,j_{l_2}}^{\delta_{l_2 k}}.$$

Finalement, on a :

$$I_m^p = \sum_{q=1}^{N_Q} \sum_{k=0}^{2^d-1} \alpha_{k,q} \prod_{l_1 \in V_m^p} x_{l_1}^{\delta_{l_1 k}} \prod_{l_2 \in \bar{V}_m^p} \Delta_{l_2,j_{l_2}} \mu_{l_2,j_{l_2}}^{\delta_{l_2 k}}. \quad (3.4)$$

### 3.2.1.2 Application sur un exemple

On prend comme exemple une fonction tabulée  $Ftab$  de dimension 3. On a besoin de stocker 1000 valeurs ( $J_1 = J_2 = J_3 = 10$ ) pour représenter  $Ftab$  (cf. figure 3.1).

La décomposition RS-HDMR tronquée à l'ordre 2 permet de décomposer cette fonction tabulée de dimension 3 en une somme de fonctions tabulées de dimensions inférieures ou égales à 2 et ainsi de diminuer le nombre de points tabulés, comme le montre la figure 3.2.

La méthode proposée en 3.2.1.1 permet de calculer de manière exacte les intégrales de la décomposition RS-HDMR dans le cas de fonctions multi-linéaires tabulées définies sur

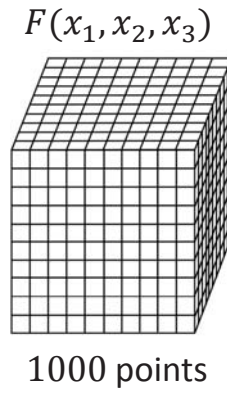


FIGURE 3.1 – Illustration des valeurs de sorties de  $F_{tab}$ .

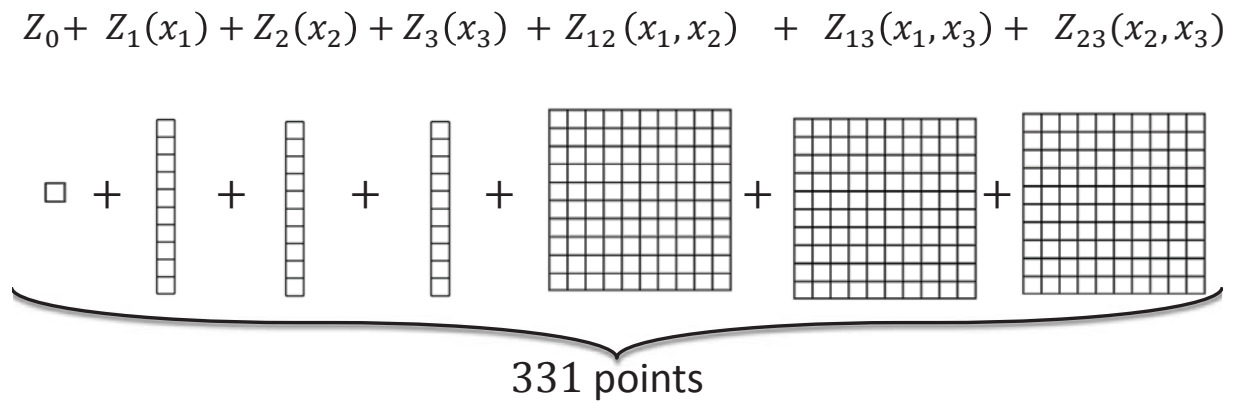


FIGURE 3.2 – Illustration de la réduction de la taille mémoire par RS-HDMR.

un hypercube. On élimine ainsi les erreurs engendrées par l'estimation des intégrales par Monte Carlo. On limite aussi le nombre d'appels à la fonction de référence, ce qui permet d'accélérer le temps de calcul de ces intégrales.

Prenons par exemple la fonction  $F$  définie,  $\forall (x_1, x_2, x_3) \in [0, 1]^3$ , par :

$$F(x_1, x_2, x_3) = x_1x_2 + x_2 + x_1x_2x_3.$$

La fonction est tabulée sur une grille régulière comprenant 10 niveaux équirépartis par variable d'entrée. La décomposition RS-HDMR tronquée à l'ordre 2 est ensuite construite. Les intégrales ont été d'une part estimées en faisant appel à une méthode de quasi-Monte Carlo et d'autre part calculées de manière exacte, grâce au résultat établi dans la section 3.2.1.1.

Le calcul exact nécessite l'inversion de 1000 systèmes linéaires de dimension 8. Une fois ces systèmes résolus de manière séquentielle, les intégrales sont calculées immédiatement avec (3.4).

La méthode de Monte Carlo nécessite l'estimation d'une intégrale triple pour calculer le terme constant  $Z_0$ , 30 intégrales doubles pour les termes  $Z_i$  et enfin 300 intégrales simples pour calculer les termes  $Z_{ij}$ . Ainsi, lorsque l'estimation par Monte Carlo de chaque intégrale nécessite  $N$  points,  $331N$  appels à la fonction  $F_{tab}$  sont nécessaires. Chacun des  $N$  points tirés aléatoirement nécessitent de réaliser une recherche pour trouver le pavé  $\Omega_q$  auquel il appartient puis de réaliser une interpolation multi-linéaire dans ce pavé. Ce nombre important d'appels à la fonction de référence du fait du grand nombre d'intégrales à estimer rend moins efficace cette méthode en terme de temps de calcul.

Comme la fonction d'origine et son approximation sont toutes les deux des fonctions multi-linéaires par morceaux définies sur la même grille, l'écart maximal entre les deux est nécessairement obtenu sur un des points de la grille. Il est donc connu exactement. L'erreur en fonction du nombre de points utilisés dans la méthode de Monte Carlo pour générer le modèle réduit est représentée sur la figure 3.3. La méthode proposée dans la section 3.2.1.1 donne directement l'erreur du modèle réduit due à la troncature de la décomposition du RS-HDMR. On observe sur la figure 3.3 la convergence de la méthode de Monte Carlo (points bleus) vers le résultat obtenu par la méthode exacte (ligne verte) lorsque le nombre total de points nécessaires pour l'estimation de l'ensemble des 331 intégrales augmente. Comme le montre la figure 3.4, alors que le résultat proposé par la méthode exacte est quasi instantané (ligne verte), plus le nombre de points nécessaires au calcul par Monte Carlo augmente, plus le temps augmente. En conséquence, plus le temps de calcul de la fonction de référence est grand, plus le calcul exact devient intéressant.

Le calcul exact permet donc d'obtenir de façon immédiate et exacte les intégrales de la décomposition RS-HDMR. Nous verrons que ce calcul nous est très utile lorsque le domaine des variables d'entrée est non hypercubique (cf. section 3.2.2.3).

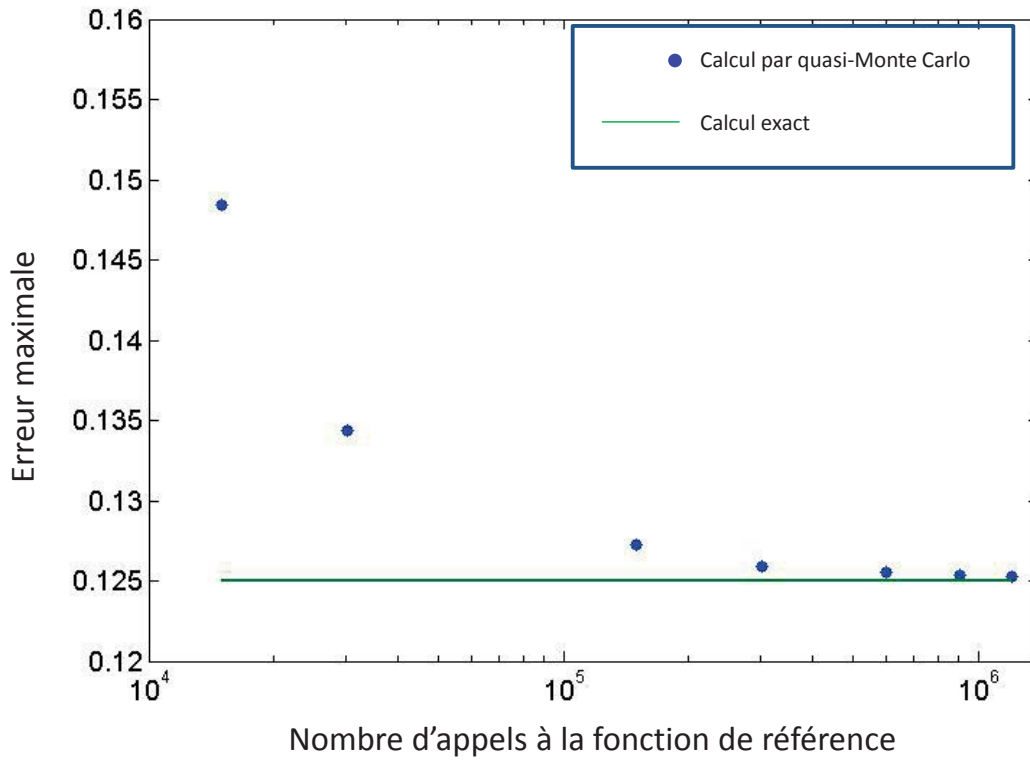


FIGURE 3.3 – Comparaison de l’erreur maximale exacte et estimée par la méthode de quasi-Monte Carlo.

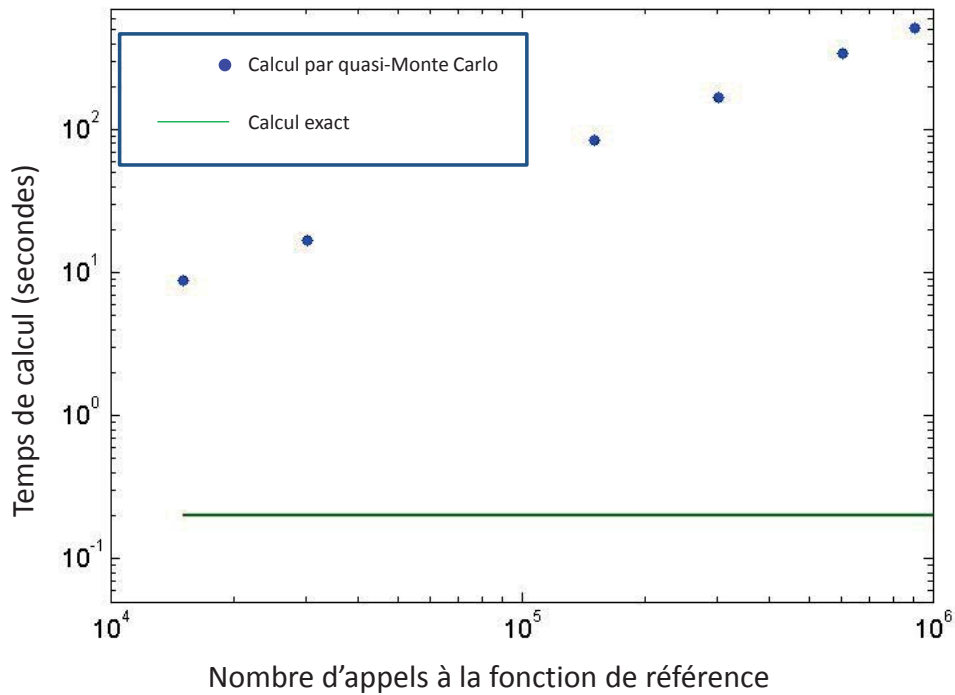


FIGURE 3.4 – Comparaison du temps de calcul de la méthode exacte et de la méthode de quasi-Monte Carlo.

### 3.2.2 Interprétation probabiliste du Random Sampling - High Dimensional Model Representation

L'unicité de la décomposition RS-HDMR proposée par Hoeffding a été établie pour des variables indépendantes et, dans la littérature, elle est essentiellement appliquée dans ce cadre-là. L'indépendance des variables d'entrée implique que leur domaine de définition soit hypercubique. Toutefois, les fonctions de référence de cette thèse sont définies sur des domaines non hypercubiques, hormis certaines fonctions tabulées comme nous venons de le voir ; cela entraîne forcément, en adoptant une interprétation probabiliste, la dépendance des variables d'entrée. Nous allons nous concentrer sur ce cas-là dans cette section.

Soit une fonction  $F$  définie pour tout vecteur aléatoire  $\mathbf{X} = (X_1, X_2, \dots, X_d)$  de densité de probabilité  $\omega$  sur un domaine non hypercubique  $\Omega$ .

#### 3.2.2.1 Variables d'entrée indépendantes

La décomposition RS-HDMR obtenue lorsque la densité de probabilité est uniforme est généralisée à une densité de probabilité quelconque  $\omega$ . Pour cela, on se place dans un premier temps dans le cas où les variables sont indépendantes et donc dans un domaine hypercubique. Pour des variables indépendantes, la densité de probabilité de  $\mathbf{X}$  est égale au produit pour  $k = 1, \dots, d$  des densités de probabilité de  $X_k$ , c'est-à-dire :

$$\omega(\mathbf{x}) = \prod_{k=1}^d \omega_k(x_k). \quad (3.5)$$

Les fonctions  $\omega_{\bar{V}_m^p}$  sont les marginales de la densité de probabilité  $\omega$  obtenues par intégration de la distribution par rapport à  $\mathbf{x}_{V_m^p}$  :

$$\omega_{\bar{V}_m^p}(\mathbf{x}_{\bar{V}_m^p}) = \int \prod_{k \in V_m^p} [0, 1] \omega(\mathbf{x}_{\bar{V}_m^p}, \mathbf{u}_{V_m^p}) d\mathbf{u}_{V_m^p}. \quad (3.6)$$

On a aussi le résultat suivant :

$$\omega_{\bar{V}_m^p}(\mathbf{x}_{\bar{V}_m^p}) = \prod_{k \in \bar{V}_m^p} \omega_k(x_k). \quad (3.7)$$

Par ailleurs, la densité de probabilité conditionnelle est égale à la densité marginale. Par exemple, la densité conditionnelle de  $\mathbf{X}_{\bar{V}_m^p}$  sachant que  $\mathbf{X}_{V_m^p} = \mathbf{x}_{V_m^p}$  est :

$$\omega_{\mathbf{X}_{\bar{V}_m^p} | \mathbf{x}_{V_m^p}}(\mathbf{x}_{\bar{V}_m^p}) = \frac{\omega(\mathbf{x}_{V_m^p}, \mathbf{x}_{\bar{V}_m^p})}{\omega_{V_m^p}(\mathbf{x}_{V_m^p})}.$$

D'après (3.5) et (3.7), on a alors :

$$\omega_{\mathbf{X}_{\bar{V}_m^p} | \mathbf{x}_{V_m^p}}(\mathbf{x}_{\bar{V}_m^p}) = \frac{\prod_{k=1}^d \omega_k(x_k)}{\prod_{k \in V_m^p} \omega_k(x_k)} = \prod_{k \in \bar{V}_m^p} \omega_k(x_k) = \omega_{\bar{V}_m^p}(\mathbf{x}_{\bar{V}_m^p}). \quad (3.8)$$

On suppose que la *vanishing condition* est vérifiée pour assurer l'unicité de la décomposition. Cette condition donnée par (3.2) s'étend pour une densité de probabilité quelconque,  $\forall m \in \{1, \dots, d\}, \forall s \in V_m^p$ ,

$$\int_{[0,1]} Z_{V_m^p}(\mathbf{u}_{V_m^p}) \omega_s(\mathbf{u}_s) du_s = 0. \quad (3.9)$$

La *vanishing condition* (3.9) permet de déterminer séquentiellement les termes de la décomposition RS-HDMR pour la densité  $\omega$  des variables d'entrée. Les termes de la décomposition RS-HDMR sont établis ici pour une densité de probabilité quelconque.

Le terme d'ordre 0 est égal à l'intégrale  $I_0^1$  déterminée comme suit :

$$\begin{aligned} I_0^1 &= \int \prod_{k \in \bar{V}_0^1} [0, 1] F(\mathbf{u}) \omega(\mathbf{u}) \prod_{k \in \bar{V}_0^1} du_k, \\ &= \mathbb{E}(F(\mathbf{X})). \end{aligned}$$

Une fois le terme d'ordre 0 déterminé, les  $p$  termes d'ordre 1 sont déterminés par :

$$Z_{V_1^p}(\mathbf{x}_{V_1^p}) = I_1^p(\mathbf{x}_{V_1^p}) - Z_0,$$

avec :

$$\begin{aligned} I_1^p(\mathbf{x}_{V_1^p}) &= \int \prod_{k \in \bar{V}_1^p} [0, 1] F(\mathbf{x}_{V_1^p}, \mathbf{u}_{\bar{V}_1^p}) \omega_{\bar{V}_1^p}(\mathbf{u}_{\bar{V}_1^p}) \prod_{k \in \bar{V}_1^p} du_k, \\ &= \int \prod_{k \in \bar{V}_1^p} [0, 1] F(\mathbf{x}_{V_1^p}, \mathbf{u}_{\bar{V}_1^p}) \prod_{k \in \bar{V}_1^p} \omega_k(u_k) du_k, \\ &= \int \prod_{k \in \bar{V}_1^p} [0, 1] F(\mathbf{x}_{V_1^p}, \mathbf{u}_{\bar{V}_1^p}) \omega_{\mathbf{X}_{\bar{V}_1^p} | \mathbf{x}_{V_1^p}}(\mathbf{u}_{\bar{V}_1^p}) du_k, \\ &= \mathbb{E}(F(\mathbf{X}) | \mathbf{X}_{V_1^p} = \mathbf{x}_{V_1^p}). \end{aligned}$$

Une fois tous les termes d'ordre 1 déterminés, on peut calculer les termes d'ordre 2 :

$$Z_{V_2^p}(\mathbf{x}_{V_2^p}) = I_2^p(\mathbf{x}_{V_2^p}) - Z_0 - \sum_{V_1^{p'} \subset V_2^p} Z_{V_1^{p'}}(\mathbf{x}_{V_1^{p'}})$$

avec :

$$\begin{aligned} I_2^p(\mathbf{x}_{V_2^p}) &= \int \prod_{k \in \bar{V}_2^p} [0, 1] F(\mathbf{x}_{V_2^p}, \mathbf{u}_{\bar{V}_2^p}) \omega(\mathbf{u}_{\bar{V}_2^p}) \prod_{k \in \bar{V}_2^p} du_k, \\ &= \int \prod_{k \in \bar{V}_2^p} [0, 1] F(\mathbf{x}_{V_2^p}, \mathbf{u}_{\bar{V}_2^p}) \prod_{k \in \bar{V}_2^p} \omega_k(u_k) du_k, \\ &= \int \prod_{k \in \bar{V}_2^p} [0, 1] F(\mathbf{x}_{V_2^p}, \mathbf{u}_{\bar{V}_2^p}) \omega_{\mathbf{X}_{\bar{V}_2^p} | \mathbf{x}_{V_2^p}}(\mathbf{u}_{\bar{V}_2^p}) du_k, \\ &= \mathbb{E}(F(\mathbf{X}) | \mathbf{X}_{V_2^p} = \mathbf{x}_{V_2^p}). \end{aligned}$$

De façon générale, une fois tous les termes d'ordre  $m - 1$  estimés, les termes d'ordre  $m \in \{1, \dots, d\}$  sont déterminés par :

$$Z_{V_m^p}(\mathbf{x}_{V_m^p}) = I_m^p(\mathbf{x}_{V_m^p}) - Z_0 - \sum_{j=1}^{m-1} \sum_{V_j^{p'} \subset V_m^p} Z_{V_j^{p'}}(\mathbf{x}_{V_j^{p'}})$$

avec :

$$\begin{aligned} I_m^p(\mathbf{x}_{V_m^p}) &= \int \prod_{k \in \bar{V}_m^p} [0, 1] F(\mathbf{x}_{V_m^p}, \mathbf{u}_{\bar{V}_m^p}) \prod_{k \in \bar{V}_m^p} \omega_k(u_k) du_k, \\ &= \mathbb{E}(F(\mathbf{X}) | \mathbf{X}_{V_m^p} = \mathbf{x}_{V_m^p}). \end{aligned} \quad (3.10)$$

### 3.2.2.2 Variables d'entrée dépendantes

Afin de simplifier les notations, le domaine d'intégration n'est plus représenté dans toute la suite.

Lorsque les variables sont dépendantes, l'équation (3.5) n'est plus vérifiée et l'écriture générale (3.10) sous la *vanishing condition* (3.9) n'est plus vraie. En effet, l'espérance conditionnelle de  $F$  pour un jeu de variables fixées peut aussi dépendre d'autres variables qui leurs sont corrélées. Une méthode a été proposée dans [46] pour traiter le cas des variables dépendantes lorsque le domaine est *grid closed*. Dans [58], un exemple est donné pour rendre explicite le concept de *grid closed* que nous rappelons ici. Un domaine  $\Omega \subset \mathbb{R}^2$  sur lequel sont définies les variables d'entrée  $(x_1, x_2)$  est dit *grid closed* si pour un point  $(x_1^0, x_2^0)$  il existe toujours un autre point avec une des deux variables fixée à la même valeur. De ce fait, la ligne  $x_1 = x_2$  n'est pas *grid closed*; dans ce cas restrictif, il n'existe pas d'autre point de  $\Omega$  en fixant une des deux variables à une valeur donnée.

On considère alors une *vanishing condition* dite relâchée qui prend la forme suivante,  $\forall m \in \{1, \dots, d\}, \forall s \in V_m^p$ ,

$$\int Z_{V_m^p}(\mathbf{u}_{V_m^p}) \omega(\mathbf{u}) du_s d\mathbf{u}_{\{1, \dots, d\} \setminus V_m^p} = 0. \quad (3.11)$$

Cette équation peut être simplifiée en réalisant implicitement l'intégration en fonction de  $\mathbf{u}_{\bar{V}_m^p}$ , c'est-à-dire en remplaçant  $\omega$  par la marginale (3.6). On obtient alors l'écriture suivante équivalente,  $\forall m \in \{1, \dots, d\}, \forall s \in V_m^p$ ,

$$\int Z_{V_m^p}(\mathbf{u}_{V_m^p}) \omega_{V_m^p}(\mathbf{u}_{V_m^p}) du_s = 0. \quad (3.12)$$

On utilise indifféremment (3.11) et (3.12) pour déterminer les termes de la décomposition RS-HDMR et assurer son unicité.

### Illustration sur un exemple pour une fonction de dimension $d = 3$

Soit une fonction  $F$  définie sur un domaine non hypercubique  $\Omega$  de dimension  $d = 3$  avec comme variables d'entrée le vecteur aléatoire  $\mathbf{X} = (X_1, X_2, X_3)$  ayant une densité de probabilité  $\omega$  sur  $\Omega$ . Les termes de la décomposition RS-HDMR de  $F$  sont définis de manière unique sous la *vanishing condition* relâchée (3.11). Les différents termes de la décomposition RS-HDMR ne peuvent plus être obtenus séquentiellement. Le système à résoudre est défini ci-dessous.

Pour le terme d'ordre 0, l'intégration sur  $\Omega$  terme à terme de la décomposition (3.1) conduit à :

$$\begin{aligned}
 & \int Z_0 \omega(\mathbf{u}) du_1 du_2 du_3 + \int Z_1(u_1) \omega(\mathbf{u}) du_1 du_2 du_3 \\
 & + \int Z_2(u_2) \omega(\mathbf{u}) du_1 du_2 du_3 + \int Z_3(u_3) \omega(\mathbf{u}) du_1 du_2 du_3 \\
 & + \int Z_{1,2}(u_1, u_2) \omega(\mathbf{u}) du_1 du_2 du_3 + \int Z_{1,3}(u_1, u_3) \omega(\mathbf{u}) du_1 du_2 du_3 \\
 & + \int Z_{2,3}(u_2, u_3) \omega(\mathbf{u}) du_1 du_2 du_3 + \int Z_{1,2,3}(\mathbf{u}) \omega(\mathbf{u}) du_1 du_2 du_3 \\
 & = \int F(\mathbf{u}) \omega(\mathbf{u}) du_1 du_2 du_3.
 \end{aligned}$$

En remarquant que  $\int \omega(\mathbf{u}) d\mathbf{u} = 1$  et d'après la *vanishing condition* relâchée (3.11), on obtient le même résultat que pour les variables indépendantes :

$$\boxed{Z_0 = \int F(\mathbf{u}) \omega(\mathbf{u}) d\mathbf{u}.} \quad (3.13)$$

On s'intéresse maintenant au calcul des termes d'ordre 1.

Pour calculer  $Z_1$ , on intègre terme à terme la décomposition (3.1) par rapport à  $x_2$  et  $x_3$  en utilisant la densité de probabilité marginale  $\omega_{2,3}$  :

$$\begin{aligned}
 & \int Z_0 \omega_{2,3}(u_2, u_3) du_2 du_3 + \int Z_1(x_1) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & + \int Z_2(u_2) \omega_{2,3}(u_2, u_3) du_2 du_3 + \int Z_3(u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & + \int Z_{1,2}(x_1, u_2) \omega_{2,3}(u_2, u_3) du_2 du_3 + \int Z_{1,3}(x_1, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & + \int Z_{2,3}(u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 + \int Z_{1,2,3}(x_1, u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & = \int F(x_1, u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3.
 \end{aligned}$$

D'après (3.6), l'égalité précédente se réécrit :

$$\begin{aligned}
 & \int Z_0 \omega_{2,3}(u_2, u_3) du_2 du_3 + \int Z_1(x_1) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & + \int Z_2(u_2) \omega_2(u_2) du_2 + \int Z_3(u_3) \omega_3(u_3) du_3 \\
 & + \int Z_{1,2}(x_1, u_2) \omega_2(u_2) du_2 + \int Z_{1,3}(x_1, u_3) \omega_3(u_3) du_3 \\
 & + \int Z_{2,3}(u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 + \int Z_{1,2,3}(x_1, u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & = \int F(x_1, u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3.
 \end{aligned}$$

En remarquant que  $\int \omega_{2,3}(u_2, u_3) du_2 du_3 = 1$  et d'après la *vanishing condition* relâchée (3.11), on obtient pour  $Z_1$  :

$$\begin{aligned}
 & Z_0 + Z_1(x_1) \\
 & + \int Z_{1,2}(x_1, u_2) \omega_2(u_2) du_2 + \int Z_{1,3}(x_1, u_3) \omega_3(u_3) du_3 \\
 & + \int Z_{1,2,3}(x_1, u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3 \\
 & = \int F(x_1, u_2, u_3) \omega_{2,3}(u_2, u_3) du_2 du_3.
 \end{aligned}$$

(3.14)

De la même manière, on obtient pour  $Z_2$  et pour  $Z_3$  :

$$\begin{aligned}
 & Z_0 + Z_2(x_2) \\
 & + \int Z_{1,2}(u_1, x_2) \omega_1(u_1) du_1 + \int Z_{2,3}(x_2, u_3) \omega_3(u_3) du_3 \\
 & + \int Z_{1,2,3}(u_1, x_2, u_3) \omega_{1,3}(u_1, u_3) du_1 du_3 \\
 & = \int F(u_1, x_2, u_3) \omega_{1,3}(u_1, u_3) du_1 du_3.
 \end{aligned}$$

(3.15)

$$\begin{aligned}
 & Z_0 + Z_3(x_3) \\
 & + \int Z_{1,3}(u_1, x_3) \omega_1(u_1) du_1 + \int Z_{2,3}(u_2, x_3) \omega_2(u_2) du_2 \\
 & + \int Z_{1,2,3}(u_1, u_2, x_3) \omega_{1,2}(u_1, u_2) du_1 du_2 \\
 & = \int F(u_1, u_2, x_3) \omega_{1,2}(u_1, u_2) du_1 du_2.
 \end{aligned}$$

(3.16)

Concernant les termes d'ordre 2, les calculs sont détaillés ci-après.

Pour calculer  $Z_{1,2}$ , on intègre terme à terme la décomposition (3.1) par rapport à  $x_3$  en

utilisant la densité de probabilité marginale  $\omega_3$  :

$$\begin{aligned}
 & \int Z_0 \omega_3(u_3) du_3 + \int Z_1(x_1) \omega_3(u_3) du_3 \\
 + & \int Z_2(x_2) \omega_3(u_3) du_3 + \int Z_3(u_3) \omega_3(u_3) du_3 \\
 + & \int Z_{1,2}(x_1, x_2) \omega_3(u_3) du_3 + \int Z_{1,3}(x_1, u_3) \omega_3(u_3) du_3 \\
 + & \int Z_{2,3}(x_2, u_3) \omega_3(u_3) du_3 + \int Z_{1,2,3}(x_1, x_2, u_3) \omega_3(u_3) du_3 \\
 = & \int F(x_1, x_2, u_3) \omega_3(u_3) du_3.
 \end{aligned}$$

En remarquant que  $\int \omega_3(u_3) du_3 = 1$  et d'après la *vanishing condition* relâchée (3.11), on obtient pour  $Z_{1,2}$  :

$$\begin{aligned}
 & Z_0 + Z_1(x_1) + Z_2(x_2) + Z_{1,2}(x_1, x_2) \\
 + & \int Z_{1,3}(x_1, u_3) \omega_3(u_3) du_3 + \int Z_{2,3}(x_2, u_3) \omega_3(u_3) du_3 \\
 + & \int Z_{1,2,3}(x_1, x_2, u_3) \omega_3(u_3) du_3 \\
 = & \int F(x_1, x_2, u_3) \omega_3(u_3) du_3.
 \end{aligned} \tag{3.17}$$

De la même manière, on obtient pour  $Z_{1,3}$  et  $Z_{2,3}$  :

$$\begin{aligned}
 & Z_0 + Z_1(x_1) + Z_3(x_3) + Z_{1,3}(x_1, x_3) \\
 + & \int Z_{1,2}(x_1, u_2) \omega_2(u_2) du_2 + \int Z_{2,3}(u_2, x_3) \omega_2(u_2) du_2 \\
 + & \int Z_{1,2,3}(x_1, u_2, x_3) \omega_2(u_2) du_2 \\
 = & \int F(x_1, u_2, x_3) \omega_2(u_2) du_2.
 \end{aligned} \tag{3.18}$$

$$\begin{aligned}
 & Z_0 + Z_2(x_2) + Z_3(x_3) + Z_{2,3}(x_2, x_3) \\
 + & \int Z_{1,2}(u_1, x_2) \omega_1(u_1) du_1 + \int Z_{1,3}(u_1, x_3) \omega_1(u_1) du_1 \\
 + & \int Z_{1,2,3}(u_1, x_2, x_3) \omega_1(u_1) du_1 \\
 = & \int F(u_1, x_2, x_3) \omega_1(u_1) du_1.
 \end{aligned} \tag{3.19}$$

Pour déterminer tous les termes de la décomposition RS-HDMR tronquée à l'ordre 1 dans le cas d'une fonction de dimension 3, on obtient un système à résoudre composé de 8 équations ((3.13), (3.14), (3.15), (3.16), (3.17), (3.18), (3.19), (3.1)) et de 8 inconnues ( $Z_0, Z_1, Z_2, Z_3, Z_{1,2}, Z_{1,3}, Z_{2,3}, Z_{1,2,3}$ ). On remarque que, contrairement au cas où les variables

sont indépendantes, pour calculer un terme d'ordre  $m$ , il est nécessaire de connaître tous les autres termes d'ordre supérieur à  $m$ .

L'estimation des termes de la décomposition RS-HDMR ne peut plus être établie séquentiellement comme cela était le cas pour les variables indépendantes [58]. Une résolution de ce système est proposée dans [46] et [58], mais via des approximations des intégrales sur une base de polynômes orthogonaux (cf. section 3.3.2), ce qui n'est pas compatible avec la précision attendue dans cette étude.

### 3.2.2.3 Proposition d'une solution dans le cas non hypercubique

Une autre approche a été proposée par Hooker dans [46]. Elle consiste à prolonger le domaine de définition de la fonction de référence  $F$  sur  $\mathbb{H}^d$ , au moyen d'une fonction  $f$  uniquement définie dans le domaine  $\mathbb{H}^d \setminus \Omega$ . Hooker a mis en avant l'influence du prolongement  $f$  sur la décomposition RS-HDMR tronquée de  $F$ .

En s'inspirant de l'idée de Hooker, nous proposons une nouvelle méthode. Celle-ci est appliquée sur une fonction  $F_{tab}$  définie sur  $\Omega$  non hypercubique que nous prolongeons sur le domaine  $\mathbb{H}^d \setminus \Omega$  par une fonction  $f_\gamma$ , dont on se donne arbitrairement la forme (linéaire, polynomiale, multi-linéaire par morceaux...), définie par un vecteur de paramètres  $\gamma \in \mathbb{R}^k$ . Ces paramètres sont alors optimisés de manière à minimiser l'erreur de la décomposition RS-HDMR à l'ordre  $m$  à l'intérieur du domaine  $\Omega$ . Le modèle réduit RS-HDMR tronqué à l'ordre  $m$ , qui dépend du prolongement par  $f_\gamma$  (et donc de  $\gamma$ ), est noté  $\widetilde{F}_{tab_\gamma}$ .

$$\widetilde{F}_{tab_\gamma}(\mathbf{x}) = Z_0^\gamma + \sum_{j=1}^m \sum_{V_j^p \in U_j} Z_{V_j^p}^\gamma(\mathbf{x}_{V_j^p}).$$

Pour un prolongement  $f_\gamma$  donné, l'erreur quadratique engendrée par le modèle RS-HDMR tronqué s'écrit

$$Err_{quad}(\gamma) = \sqrt{\frac{1}{N} \sum_{i=1}^N \epsilon_i(\gamma)^2},$$

où pour chacun des  $N$  points tabulés contenus dans  $\Omega$ , on a  $\forall 1 \leq i \leq N$ ,

$$\epsilon_i(\gamma) = \left| \widetilde{F}_{tab_\gamma}(\mathbf{x}_i) - F_{tab}(\mathbf{x}_i) \right|.$$

On cherche à minimiser cette grandeur sur  $\Omega$  en fonction de  $\gamma$  :

$$\min_{\gamma \in \mathbb{R}^k} Err_{quad}(\gamma).$$

Nous illustrons ci-après le comportement de cette méthode d'ajustement du prolongement sur deux exemples.

### Cas test 1

Soit la fonction  $Ftab1$ ,

$$\begin{aligned} Ftab1 : \quad \Omega_1 &\rightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2) &\mapsto x_1 + x_2, \end{aligned} \tag{3.20}$$

où  $\Omega_1$  est le domaine représenté sur la figure 3.5 :

$$\Omega_1 = \{(x_1, x_2) \in \mathbb{R}^2, \text{ tels que } x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}.$$

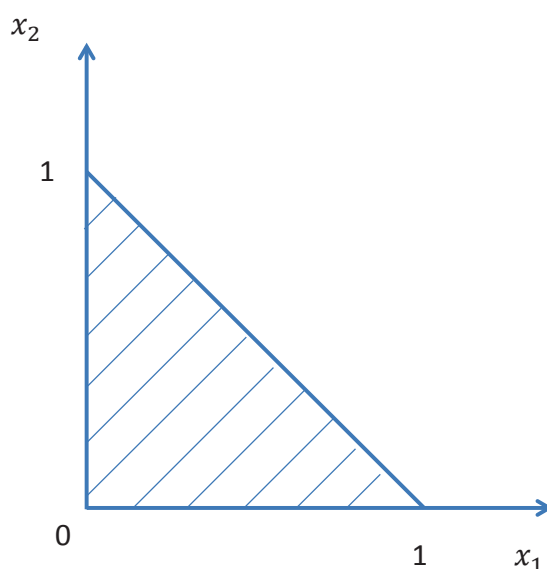


FIGURE 3.5 – Domaine non hypercubique  $\Omega_1$ .

La fonction est tabulée sur les points de la grille équirépartie  $10 \times 10$  qui sont dans  $\Omega_1$ .

Nous avons choisi la fonction  $Ftab1$  pour deux raisons :

- la décomposition RS-HDMR à l'ordre 1 sur l'hypercube unité est exacte ;
- une solution évidente à son prolongement est  $\forall (x_1, x_2) \in \mathbb{H}^d \setminus \Omega_1, f(x_1, x_2) = x_1 + x_2$ .

### Prolongement 1

Sur le domaine  $\mathbb{H}^d \setminus \Omega_1$ , nous proposons une fonction paramétrique de prolongement :

$$\begin{aligned} f_{\gamma^1} : \quad \mathbb{H}^d \setminus \Omega_1 &\rightarrow \mathbb{R} \\ \mathbf{x} = (x_1, x_2) &\mapsto \gamma_1^1 + \gamma_2^1 x_1 + \gamma_3^1 x_2. \end{aligned}$$

Les intégrales de la décomposition RS-HDMR sont obtenues par le calcul exact présenté à la section 3.2.1.1.

On cherche à minimiser  $Err_{quad}$  en fonction de  $\gamma^1$ .

Les coefficients  $\gamma^1$  ont été optimisés par la fonction *lsqnonlin* de MATLAB. L'optimum obtenu est la solution triviale  $\gamma^1 \simeq (0, 1, 1)$  qui permet de construire la décomposition RS-HDMR tronquée à l'ordre 1 suivante :

$$\begin{aligned} Z_0^{\gamma^1} &\simeq 1, \\ \forall x_1 \in [0, 1], Z_1^{\gamma^1}(x_1) &\simeq x_1 - \frac{1}{2}, \\ \forall x_2 \in [0, 1], Z_2^{\gamma^1}(x_2) &\simeq x_2 - \frac{1}{2}. \end{aligned}$$

La fonction prolongée,  $\forall (x_1, x_2) \in \mathbb{H}^2$ ,

$$Fp1_{\gamma^1}(x_1, x_2) = Ftab1(x_1, x_2)\mathbb{1}_{\Omega_1}(x_1, x_2) + f_{\gamma^1}(x_1, x_2)\mathbb{1}_{\mathbb{H}^2 \setminus \Omega_1}(x_1, x_2),$$

et son approximation RS-HDMR tronquée l'ordre 1,  $\forall (x_1, x_2) \in \mathbb{H}^2$ ,

$$\widetilde{Ftab1}_{\gamma^1}(x_1, x_2) \simeq Z_0^{\gamma^1} + Z_1^{\gamma^1}(x_1) + Z_2^{\gamma^1}(x_2),$$

sont représentées sur les figures 3.6 et 3.7. L'erreur maximale est de  $2.37 \cdot 10^{-8}$  sur  $\Omega_1$ .

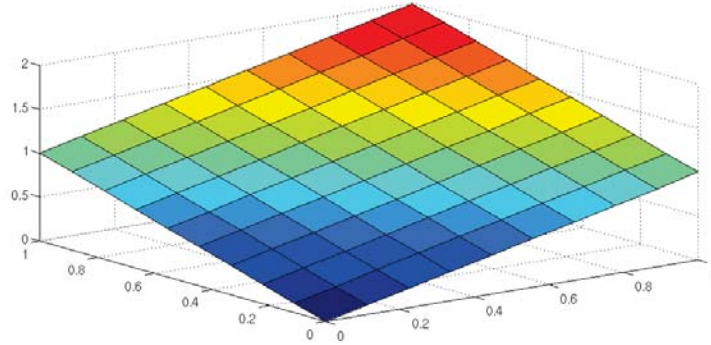


FIGURE 3.6 – Fonction prolongée  $Fp1_{\gamma^1}$  sur  $\mathbb{H}^2$ .

### Prolongement 2

Dans les cas pratiques qui nous intéressent, la fonction  $F$  à réduire est tabulée sur un domaine non hypercubique et nous n'en connaissons bien évidemment pas de représentation analytique. Nous avons donc choisi un prolongement de la fonction tabulée sur  $\mathbb{H}^d \setminus \Omega_1$  paramétré naturellement par ses valeurs en les points de la grille étendue à  $\mathbb{H}^d \setminus \Omega_1$ . Cela augmente considérablement la taille du vecteur  $\gamma$  et donc le nombre de degrés de liberté. Par exemple, pour la fonction (3.20)  $Ftab1$ , si chaque variable  $x_1$  et  $x_2$  est discrétisée en 10 valeurs distinctes, 55 points définissent la fonction dans  $\Omega_1$  et 45 points la définissent à l'extérieur. Des interpolations multi-linéaires sont réalisées entre les points tabulés tant dans  $\Omega_1$  que dans  $\mathbb{H}^d \setminus \Omega_1$ .

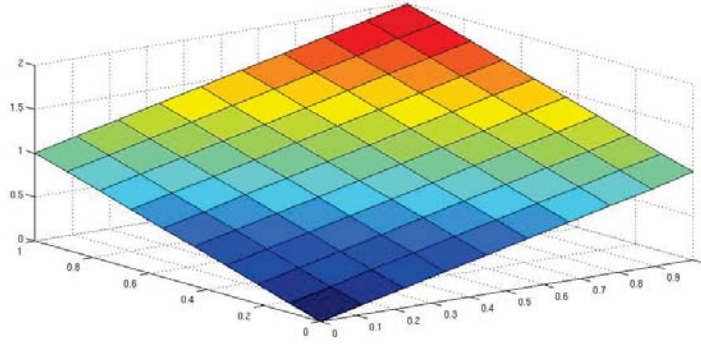


FIGURE 3.7 – Fonction approchée  $\widetilde{F}_{tab1_{\gamma_1}}$  sur  $\mathbb{H}^2$ .

La fonction prolongée après optimisation,  $\forall (x_1, x_2) \in \mathbb{H}^2$ ,

$$Fp1_{\gamma_2}(x_1, x_2) = F_{tab1}(x_1, x_2)\mathbb{1}_{\Omega_1}(x_1, x_2) + f_{\gamma^2}(x_1, x_2)\mathbb{1}_{\mathbb{H}^2 \setminus \Omega_1}(x_1, x_2)$$

est représentée figure 3.8. Le prolongement de  $F$  sur  $\mathbb{H}^d \setminus \Omega_1$  n'est pas régulier.

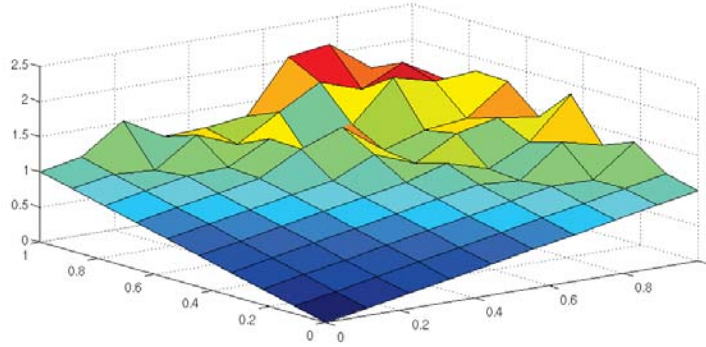


FIGURE 3.8 – Fonction prolongée  $Fp1_{\gamma_2}$  sur  $\mathbb{H}^2$ .

Néanmoins, son approximation RS-HDMR tronquée à l'ordre 1,  $\forall (x_1, x_2) \in \mathbb{H}^2$ ,

$$\widetilde{F}_{tab1_{\gamma_2}}(x_1, x_2) \simeq Z_0^{\gamma^2} + Z_1^{\gamma^2}(x_1) + Z_2^{\gamma^2}(x_2)$$

tracée sur la figure 3.9 conduit à une erreur maximale de  $4.23 \cdot 10^{-6}$  sur  $\Omega_1$  jugée tout à fait satisfaisante.

### Cas test 2

Nous considérons maintenant une fonction dont la décomposition RS-HDMR tronquée à l'ordre 1 n'est pas exacte et dont le prolongement sur  $\mathbb{H}^2$  n'est pas trivial. Nous avons pour cela choisi la fonction suivante :

$$F_{tab2} : \quad \Omega_1 \quad \rightarrow \quad \mathbb{R}$$

$$\mathbf{x} = (x_1, x_2) \quad \mapsto \quad \sqrt{1 - (x_1 + x_2)}.$$

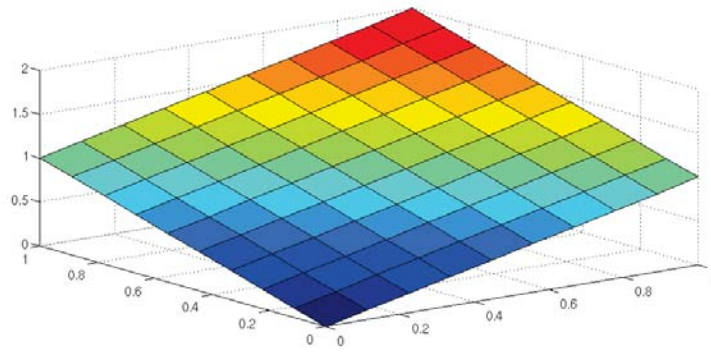


FIGURE 3.9 – Fonction approchée  $\widetilde{Ftab1}_{\gamma_2}$  sur  $\mathbb{H}^2$ .

Son prolongement  $f_{\gamma^3}$  est défini par ses valeurs sur les 45 points de la grille qui permettent de compléter le domaine  $\Omega_1$  en un domaine hypercubique. Là encore, des interpolations multi-linéaires sont réalisées entre les points tabulés.

La fonction prolongée résultant de l'optimisation  $\forall (x_1, x_2) \in \mathbb{H}^2$ ,

$$Fp2_{\gamma^3}(x_1, x_2) = Ftab2(x_1, x_2)\mathbb{1}_{\Omega_1}(x_1, x_2) + f_{\gamma^3}(x_1, x_2)\mathbb{1}_{\mathbb{H}^2 \setminus \Omega_1}(x_1, x_2),$$

est présentée sur la figure 3.10.

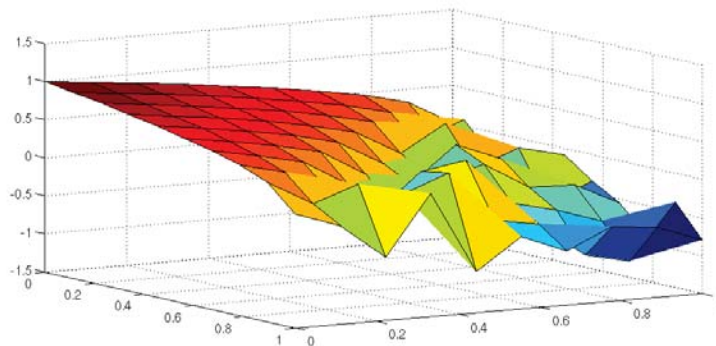


FIGURE 3.10 – Fonction prolongée  $Fp2_{\gamma^3}$  sur  $\mathbb{H}^2$ .

Comme dans le cas précédent le prolongement obtenu n'est pas régulier. Il conduit à la décomposition RS-HDMR tronquée à l'ordre 1,  $\forall (x_1, x_2) \in \mathbb{H}^2$ ,

$$\widetilde{Ftab2}_{\gamma^3}(x_1, x_2) \simeq Z_0^{\gamma^3} + Z_1^{\gamma^3}(x_1) + Z_2^{\gamma^3}(x_2),$$

représentée sur la figure 3.11.

L'erreur maximale est 0.2 sur  $\Omega_1$ . Cette valeur non négligeable s'explique par le fait que la fonction de référence ne peut pas être approchée exactement par une décomposition RS-HDMR à l'ordre 1.

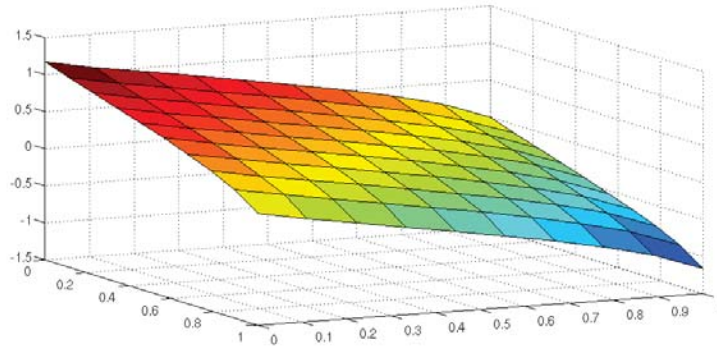


FIGURE 3.11 – Fonction approchée  $\widetilde{Ftab}_{2,3}$  sur  $\mathbb{H}^2$ .

### Conclusion

Nous avons montré ici le bon fonctionnement de la décomposition RS-HDMR d'une fonction tabulée quelconque définie sur un domaine non hypercubique, en prolongeant la fonction en dehors de son domaine de définition pour le rendre hypercubique. Les prolongements obtenus ne sont pas lisses (cf. figures 3.8 et 3.10), mais permettent d'obtenir des décompositions RS-HDMR tronquées satisfaisantes sur le domaine de définition de la fonction.

### 3.2.3 Utilisation particulière des réseaux de neurones

L'idée ici est de ne pas stocker les valeurs de la grille en mémoire mais de la faire apprendre par un réseau de neurones dont le nombre de coefficients est plus faible. Nous proposons donc une utilisation particulière des réseaux de neurones, visant à approcher au mieux les points de la fonction tabulée, sans s'intéresser à sa capacité de généralisation. L'échantillon d'apprentissage est donc constitué des points de la grille de la fonction tabulée.

Par ailleurs, puisque la fonction d'origine et son approximation sont encore une fois toutes les deux des fonctions multi-linéaires par morceaux définies sur la même grille, l'écart maximal entre les deux est nécessairement obtenu sur un des points de la grille.

L'échantillon de test, qui est le même que l'échantillon d'apprentissage, contrôle donc l'erreur commise sur les points de la fonction tabulée. Sur la figure 3.12, les croix bleues sont les points tabulés de la fonction de référence et les croix rouges sont estimées par le réseau de neurones. Cette utilisation des RN a un intérêt dès qu'on est capable d'estimer les points de la table avec un RN suffisamment précis contenant moins de coefficients que le nombre de points définissant la fonction tabulée de référence. Notons qu'outre les  $N_1 \times (2+d)$  coefficients du réseau de neurones à une couche cachée de  $N_1$  neurones, il faut également stocker pour  $i = 1, \dots, d$ , les  $J_i$  différentes valeurs prises par chaque compo-

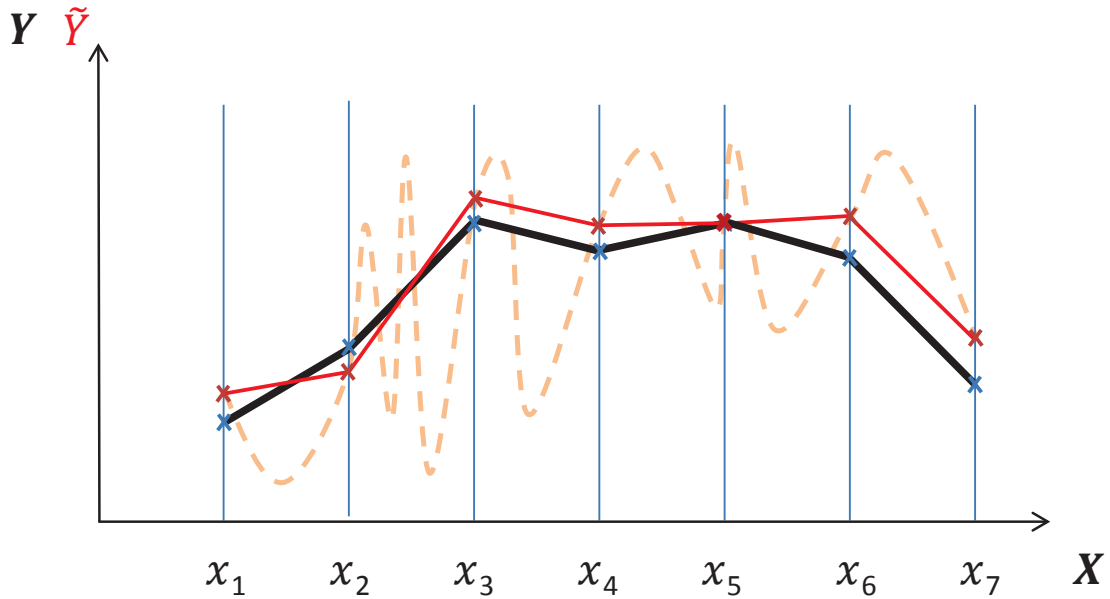


FIGURE 3.12 – Réseau de neurones appelé sur une grille 1D.

sante des points où la fonction est tabulée. Au total, on stocke donc  $N_1 \times (2 + d) + \sum_{i=1}^d J_i$  paramètres. De son côté, une fonction tabulée comprend  $\prod_{i=1}^d J_i + \sum_{i=1}^d J_i$  paramètres. La réduction est donc effective lorsque  $N_1 \times (2 + d) < \prod_{i=1}^d J_i$ . La propriété de parcimonie des RN permet de vérifier facilement cette condition pour les cas qui nous intéressent (typiquement  $J_i = 10$  et  $d = 3$ ).

### 3.3 Réduction globale des modèles de qualité de vol et de performances des avions

Le problème est maintenant de réduire en taille et en temps de calcul une fonction boîte noire comprenant un nombre important de variables d'entrée continues, typiquement 15. Pour ce faire, on peut utiliser différents types de métamodèles, tels que ceux décrits à la section 2.3. On utilise ici le RS-HDMR qui permet de construire automatiquement des fonctions tabulées de dimension faible (typiquement une ou deux variables d'entrée) et qui résument le comportement global de la fonction.

L'objet de cette section est de présenter les limites de l'application du RS-HDMR à la réduction des fonctions tabulées de grande dimension.

### 3.3.1 Fléau de la dimension

La représentation tabulée d'un modèle de référence à 15 variables d'entrée sur une grille à 10 valeurs par variable nécessite le stockage de  $10^{15}$  points, ce qui est inenvisageable. Ceci justifie l'utilisation du RS-HDMR pour construire automatiquement des fonctions de dimensions inférieures qui, elles, peuvent être tabulées.

La construction de ces fonctions tabulées nécessite toutefois un nombre d'appels important à la fonction de référence.

Le nombre de termes  $N_T$  de la décomposition RS-HDMR augmente avec la dimension  $d$  et l'ordre  $m$  de troncature de la décomposition :

$$N_T = \sum_{i=0}^m \binom{d}{i} = \sum_{i=0}^m \frac{d!}{(d-i)!i!}.$$

Chaque terme de la décomposition RS-HDMR est stocké dans une fonction tabulée. Chaque point tabulé correspond à une intégrale à estimer. Pour une fonction de dimension 15 avec  $\forall i = 1, \dots, 15, J_i = 10$ , le nombre total d'intégrales à estimer pour construire la décomposition RS-HDMR tronquée à l'ordre 2 est :

$$N_i = 1 + 10 \times 15 + 100 \times 105 = 10\,651.$$

Ce nombre  $N_i$  correspond également aux nombres de points des grilles générées par la décomposition RS-HDMR tronquée à l'ordre 2. La précision de l'estimation des intégrales par des méthodes de quasi-Monte Carlo augmente avec le nombre de points générés (cf. section 2.4.4). La précision des différents composants de la décomposition RS-HDMR dépend donc du nombre  $N$  de points générés quasi-aléatoirement. Il est important de noter que le nombre de points nécessaires aux estimations des intégrales par la méthode de quasi-Monte Carlo ne dépend pas significativement de la dimension  $d$  du problème. Cette propriété est intéressante pour le calcul d'intégrales en grande dimension comme c'est le cas ici. Toutefois, le nombre d'intégrales à estimer augmente significativement le nombre d'appels à la fonction de référence. Par exemple pour la décomposition RS-HDMR d'une fonction test de dimension 15 tronquée à l'ordre 2, nous avons estimé le nombre d'intégrales à 10 651. Si pour chaque intégrale, il est nécessaire de prendre  $10^4$  points pour que la méthode d'estimation des intégrales converge, on doit faire 106 510 000 appels à la fonction. L'appel au modèle aérodynamique QdV a été estimé à 0.01 seconde, il faudrait donc quatre mois pour évaluer toutes les intégrales.

Considérons la fonction  $F_{15}$  définie comme suit :

$$\begin{aligned} F_{15} : \mathbb{H}^{15} &\rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto \sum_{i=1}^{15} x_i^2 + x_{14}x_{15} + x_1x_3 + \exp(x_2^2x_8). \end{aligned}$$

Pour implémenter ce calcul, il a été nécessaire de faire appel à la vectorisation pour obtenir le temps de calcul nécessaire. Dans une utilisation normale des modèles QdV ou Perfo, cela n'est pas possible.

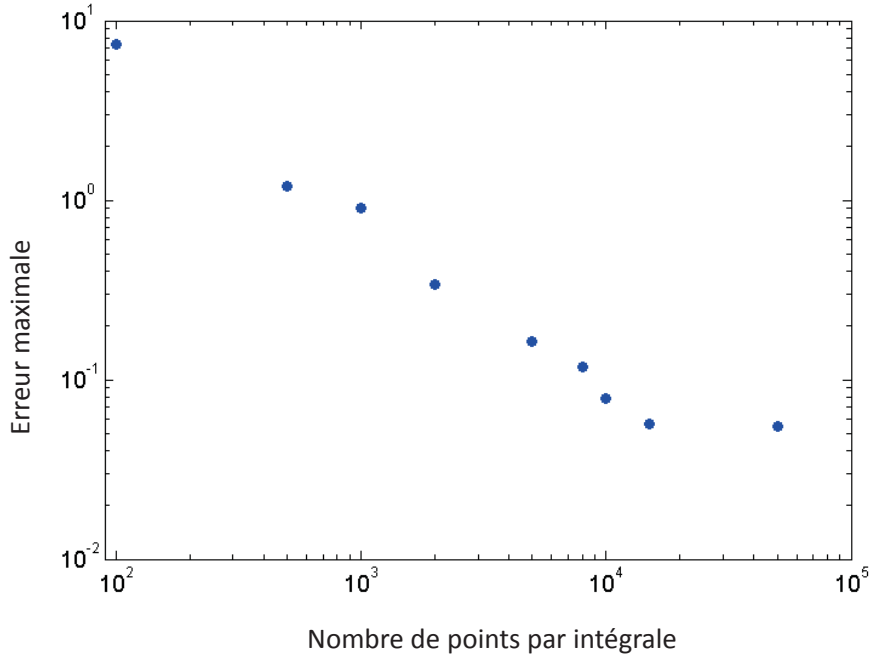


FIGURE 3.13 – Convergence de la méthode de quasi-Monte Carlo pour l'estimation d'une intégrale (parmi les 10 651 intégrales) en fonction du nombre de points.

On présente sur la figure 3.13 la courbe de convergence de l'erreur de la décomposition RS-HDMR tronquée à l'ordre 2, en fonction du nombre de points nécessaires au calcul d'une intégrale. On constate que pour estimer chaque intégrale,  $10^4$  appels à la fonction sont nécessaires pour obtenir une erreur acceptable. Comme nous venons de le voir, cette méthode n'est pas envisageable pour approcher nos modèles de référence qui nécessitent environ 0.01 seconde par appel.

### 3.3.2 Construction du Random Sampling - High Dimensional Model Representation par génération quasi-aléatoire d'un unique échantillon

Le problème soulevé ici vient du fait que chaque intégrale estimée par Monte Carlo nécessite son propre échantillon et donc d'autant plus d'appels à la fonction de référence. L'idée développée ici consiste à utiliser un échantillon unique pour l'estimation de toutes les intégrales. Pour cela, Ziehn et Tomlin [86] ont proposé l'utilisation de polynômes orthonormés (polynômes de Legendre) pour représenter les fonctions à intégrer.

Le terme  $Z_0$  est estimé comme précédemment, c'est-à-dire comme la moyenne de la fonction  $F$  sur son domaine de définition.

Les intégrales d'ordre supérieur sont approchées par des polynômes.

Termes d'ordre 1 :  $\forall i_1 \in \{1, \dots, d\}$

$$Z_{i_1}(\mathbf{x}_{i_1}) \approx \sum_{r=1}^k \alpha_r^{i_1} \varphi_r^{i_1}(x_{i_1}).$$

Termes d'ordre 2 :  $\forall i_1 \in \{1, \dots, d\}, \forall i_2 \in \{1, \dots, d\}$  et  $i_1 < i_2$ ,

$$Z_{i_1, i_2}(\mathbf{x}_{i_1}, \mathbf{x}_{i_2}) = \sum_{r_1=1}^{k_1} \sum_{r_2=1}^{k_2} \beta_{r_1 r_2}^{i_1 i_2} \varphi_{r_1}^{i_1}(x_{i_1}) \varphi_{r_2}^{i_2}(x_{i_2}),$$

où  $k, k_1, k_2$  sont des entiers, qui sont les ordres maximaux des polynômes. Les coefficients  $\alpha_r^{i_1}$  et  $\beta_{r_1 r_2}^{i_1 i_2}$  sont à déterminer. Le polynôme  $\varphi_r^i$  est un polynôme d'ordre  $r$  associé à la variable  $x_i$ . Les polynômes  $\varphi_r^{i_1}, \varphi_{r_1}^{i_1}$  et  $\varphi_{r_2}^{i_2}$  doivent être construits de telle sorte qu'ils forment une base de polynômes orthonormés. En utilisant la propriété d'orthogonalité des polynômes, les constantes sont données par [59, 60, 86] :

$\forall r = 1, \dots, k, \forall p = 1, \dots, k_1, \forall q = 1, \dots, k_2, \forall i, i_1, i_2 = 1, \dots, d, i_1 < i_2$

$$\alpha_r^{i_1} = \int F(\mathbf{x}) \varphi_r^{i_1}(x_{i_1}) d\mathbf{x},$$

$$\beta_{pq}^{i_1 i_2} = \int F(\mathbf{x}) \varphi_p^{i_1}(x_{i_1}) \varphi_q^{i_2}(x_{i_2}) d\mathbf{x}.$$

De cette façon, la génération d'un seul échantillon de points est suffisante pour déterminer toutes les composantes du RS-HDMR. Nous avons choisi des polynômes de degré 10 au maximum pour approcher globalement les différentes intégrales afin d'avoir le même ordre de grandeur de coefficients que pour l'exemple donné ci-dessus, c'est-à-dire 10 651 coefficients (1 correspondant au terme  $Z_0$ , 150 coefficients correspondant aux termes d'ordre 1  $\alpha_r^i$  et 10 500 coefficients correspondant aux termes d'ordre 2  $\beta_{i_1 i_2}^{pq}$ ).

Par exemple, pour la fonction  $F_{15}$ , la méthode de quasi-Monte Carlo [28] converge vers l'erreur maximale de 2.51 avec seulement 100 000 appels à la fonction de référence (cf. figure 3.14). Pour obtenir l'approximation de la fonction de référence par cette méthode, le nombre d'appels à la fonction de référence est donc beaucoup plus faible. Toutefois, cette approximation n'est pas satisfaisante en terme de précision.

## 3.4 Conclusion

Dans le cas de l'approximation de nos fonctions de référence, les difficultés liées à la non-hypercubicité du domaine de définition des variables d'entrée s'ajoutent au problème de l'estimation d'un grand nombre d'intégrales en grande dimension. La solution proposée, qui consiste à optimiser le prolongement de la fonction en dehors de son domaine de

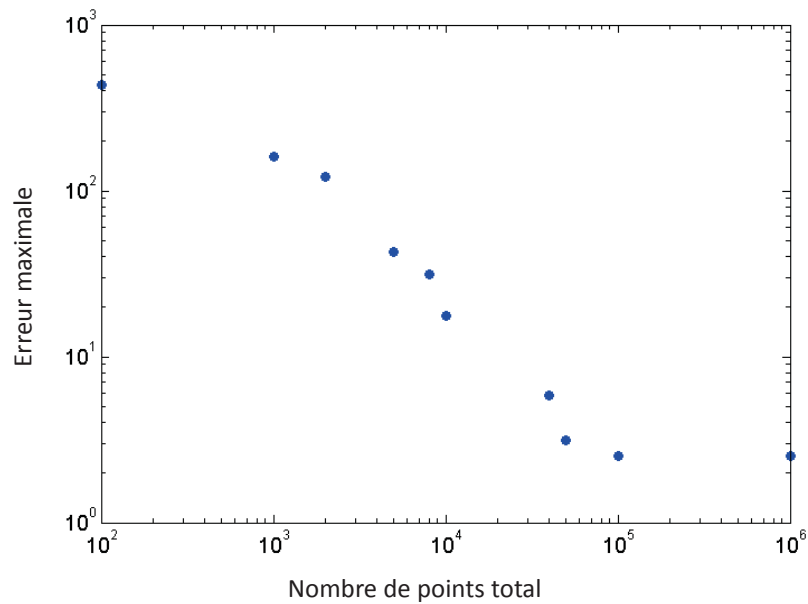


FIGURE 3.14 – Convergence de la méthode de quasi-Monte Carlo pour l’estimation de l’ensemble des coefficients des polynômes.

définition pour rendre le domaine hypercubique (cf. section 3.2.2.3) nécessite la construction de plusieurs modèles RS-HDMR dans une boucle d’optimisation, ce qui n’est pas envisageable pour des fonctions de grande dimension.

Pour résoudre ce problème de grande dimension, on propose une stratégie basée sur le constat que les modèles QdV et Perfo ont une structure hiérarchique multi-niveaux dont les éléments dimensionnants en termes de temps de calcul et de taille mémoire sont des fonctions tabulées de petites dimensions. L’idée est donc de générer des modèles réduits de ces fonctions tabulées en utilisant les techniques décrites au début de ce chapitre (cf. section 3.2). Dans le chapitre suivant, nous allons exhiber une méthode automatique d’un choix optimal de ces modèles réduits en termes de stockage, de temps de calcul et de précision.



# Chapitre 4

## Méthodologie générale de réduction d'un modèle de qualités de vol et de performances des avions

Ce chapitre a fait l'objet d'une publication dans [9].

### 4.1 Introduction

Le besoin de dériver automatiquement le modèle aérodynamique QdV pour construire un modèle aérodynamique Perfo a été introduit au chapitre 1. Comme nous l'avons vu, cette dérivation automatique a pour but de rationaliser le processus de création de ces deux modèles et ainsi de réduire les coûts de calcul et les délais de production des modèles. En effet, actuellement, ces deux modèles sont construits parallèlement et ne partagent donc aucune donnée. Nos efforts se concentrent donc sur la dérivation directe et automatique du modèle aérodynamique Perfo à partir du modèle aérodynamique QdV de référence. L'utilisation quotidienne intensive du modèle Perfo par les ingénieurs du bureau d'étude impose d'être en mesure de générer en quelques secondes une trajectoire complète de l'avion ; ce qui induit un besoin de réduction en temps de calcul du modèle aérodynamique QdV tout en garantissant une précision satisfaisante.

En ce qui concerne les modèles embarqués dans le cockpit de l'avion, nous avons mis en évidence dans le chapitre 1 des exigences critiques en termes de temps de calcul, de taille mémoire et de précision. En effet, le besoin d'embarquer de plus en plus de données dans le cockpit de l'avion mène à une diminution de la taille mémoire disponible pour chaque fonction embarquée. La réduction du temps de calcul permet, quant à elle, d'améliorer en temps réel les performances de l'avion en vol. Le modèle Perfo doit donc être à son tour réduit pour répondre aux exigences d'embarquabilité.

Les modèles actuels possèdent une structure hiérarchique multi-niveaux dont chaque élément représente un phénomène physique particulier. Leur validation repose en grande

partie sur cette architecture ; pour cette raison nous allons la conserver lors de la construction des modèles réduits. L'objectif est donc ici de proposer une méthode automatique permettant le choix optimal de modèles réduits des sous-modèles locaux qui constituent le modèle hiérarchique global de telle sorte que le modèle obtenu satisfasse les exigences en termes de taille mémoire, de temps de calcul et de précision.

Ici, pour chaque sous-modèle, nous supposons qu'un ensemble de modèles réduits est déjà disponible. Le problème n'est donc pas la définition de ces modèles réduits mais leur sélection optimale. Une méthodologie est proposée pour effectuer cette sélection, en satisfaisant les contraintes de précision et de temps de calcul liées aux modèles Perfo utilisés au bureau d'étude ou dans le cockpit de l'avion. Cette combinaison optimale est obtenue en résolvant un problème linéaire en nombres entiers [85] et permet au modèle embarqué final de garder une structure multi-niveaux. Dans la littérature, aucune publication n'a été trouvée concernant cette affectation optimale dans une structure multi-niveaux hiérarchique.

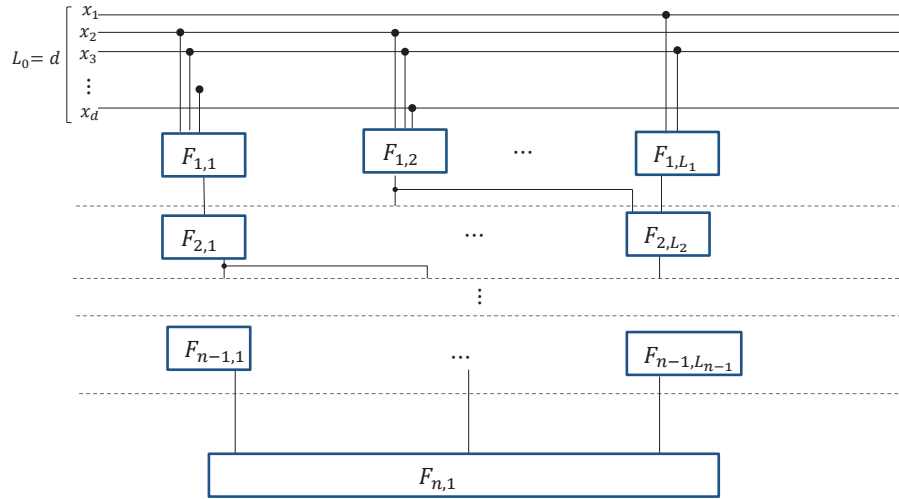
La section 4.2.1 introduit la formulation mathématique de l'architecture multi-niveaux. Ensuite, dans la section 4.2.2, les combinaisons optimales répondant aux deux besoins décrits dans le chapitre 1 sont définies comme la solution de problèmes d'optimisation non linéaires en nombres entiers. Grâce à une majoration linéaire de l'erreur propagée à travers les niveaux établie à la section 4.2.3, ces problèmes d'optimisation non linéaire en nombres entiers sont transformés en problèmes d'optimisation linéaire en nombres entiers. La méthode est validée dans la section 4.3 sur un exemple de fonction embarquée permettant de calculer la consommation de l'avion à partir d'un modèle Perfo : le Fuel Consumption Model (FCM). Une discussion et certaines améliorations sont proposées dans la section 4.4.

## 4.2 Formulation mathématique

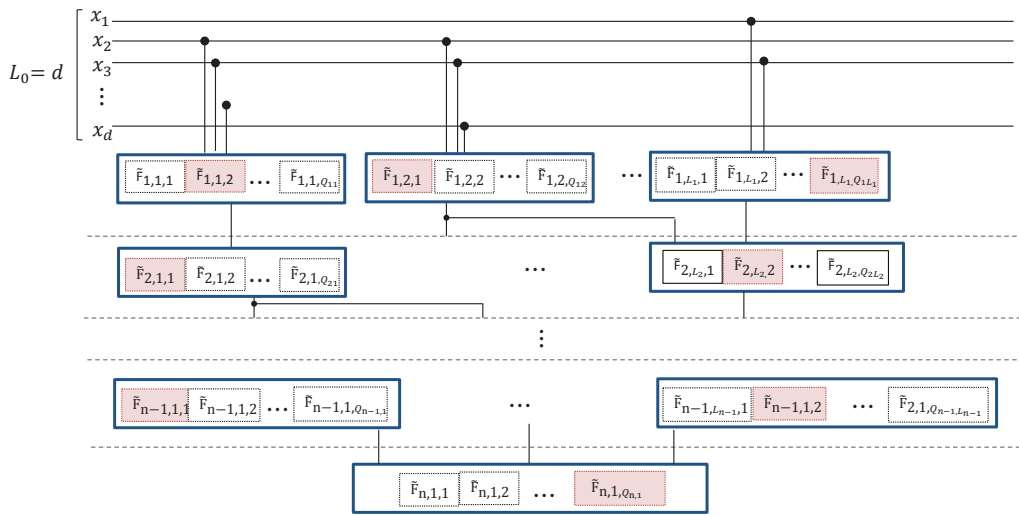
### 4.2.1 Architecture multi-niveaux hiérarchique

Le modèle global de référence  $F$  est composé d'un assemblage de sous-modèles  $F_{i,j}$ . Cette décomposition est présentée dans la figure 4.1(a). Les variables d'entrée sont représentées par le vecteur  $\mathbf{x}$  de dimension  $d$ . Le flux de données parcourt les niveaux de  $F$  de façon uni-directionnelle depuis le niveau supérieur jusqu'au niveau de sortie. Chaque niveau  $i$ ,  $1 \leq i \leq n$ , est composé de  $L_i$  sous-modèles. La sortie  $O_{i,j}$  du  $j^{\text{ème}}$  sous-modèle  $F_{i,j}$  est l'entrée d'au moins un sous-modèle au niveau  $i+1$ . Les sorties du niveau  $i$  sont regroupées dans le vecteur  $\mathbf{O}_i = (O_{i,1}, \dots, O_{i,L_i})$ . Le dernier niveau fournit la sortie unique du modèle global. Notons donc  $L_n = 1$ .

Plus précisément,  $\forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, L_i\}$ ,



(a) Le modèle de référence  $F$ .



(b) Le modèle réduit  $\tilde{F}$ .

FIGURE 4.1 – Structure hiérarchique multi-niveaux.

- le sous-modèle  $F_{i,j}$ , supposé dérivable, est défini par :

$$F_{i,j} : \mathbb{R}^{L_{i-1}} \longrightarrow \mathbb{R}$$

$$\mathbf{O}_{i-1} \longmapsto F_{i,j}(\mathbf{O}_{i-1}) ;$$

- il existe  $Q_{i,j}$  modèles réduits candidats notés  $\tilde{F}_{i,j,q}$  pour remplacer  $F_{i,j}$  et (cf. figure 4.1(b)). Ces modèles réduits incluent le sous-modèle original ( $q = 1$ ). Ils sont

définis par :

$$\begin{aligned}\tilde{F}_{i,j,q} : \mathbb{R}^{L_{i-1}} &\longrightarrow \mathbb{R} \\ \mathbf{O}_{i-1} &\longmapsto \tilde{F}_{i,j,q}(\mathbf{O}_{i-1})\end{aligned}$$

$\forall q \in \{1, \dots, Q_{i,j}\}$ ;

- le temps de calcul local  $T_{i,j,q}$ , la taille mémoire locale  $S_{i,j,q}$  et une mesure de l'erreur locale  $\epsilon_{i,j,q}$  induite par  $\tilde{F}_{i,j,q}$  sont donnés ;
- la variable de décision binaire  $z_{i,j,q}$  permet de sélectionner ou pas l'option  $\tilde{F}_{i,j,q}$  comme constituant le modèle hiérarchique réduit global. Si la  $q^{\text{ème}}$  option  $\tilde{F}_{i,j,q}$  est choisie pour le sous-modèle  $F_{i,j}$ , alors  $z_{i,j,q} = 1$ , sinon  $z_{i,j,q} = 0$ .

Notons  $\mathbf{Z}$  le vecteur composé des variables de décision élémentaires  $z_{i,j,q}$ . Puisque seulement une option est choisie pour chaque sous-modèle, les équations suivantes doivent être vérifiées :

$$\forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, L_i\}, \sum_{q=1}^{Q_{i,j}} z_{i,j,q} = 1.$$

Une fois que le vecteur  $\mathbf{Z}$  est connu, la fonction  $\tilde{F}_{\mathbf{Z}}$  est définie comme le modèle réduit global, obtenu comme l'assemblage de tous les sous-modèles sélectionnés. En conformité avec les exigences de validation,  $\tilde{F}_{\mathbf{Z}}$  garde la structure multi-niveaux de  $F$ .

Remarque : la valeur de  $L_i$  est en général différente à chaque niveau  $i$  et le nombre de modèles  $Q_{i,j}$  pour chaque sous-modèle  $F_{i,j}$  n'est pas constant, donc  $j$  dépend de  $i$  et  $q$  dépend de  $i$  et  $j$ . Néanmoins ces dépendances ne seront explicitées qu'en cas d'ambiguïté.

## 4.2.2 Programmation Linéaire en Nombres Entiers

Les besoins de réduction rappelés dans la section 4.1 se formulent naturellement comme deux problèmes d'optimisation sous contraintes (cf. table 4.1).

Concernant le besoin de réduction du modèle aérodynamique QdV, on cherche à minimiser le temps de calcul  $T$  tout en maintenant l'erreur globale  $A$  sous un certain seuil.

Le besoin de réduction du modèle Perfo en vue de l'embarquer se formule quant à lui comme un problème de minimisation de la taille mémoire globale  $S$ , en maintenant sous des bornes supérieures le temps de calcul global  $T$  et l'erreur globale  $A$  compatibles avec les exigences de temps réel des fonctions embarquées. La taille mémoire n'est pas limitative puisqu'elle reste bien en dessous des capacités du calculateur embarqué. Malgré tout, elle doit être minimisée, du fait de la volonté d'augmenter le nombre de fonctions à embarquer dans le cockpit de l'avion.

Besoin 1 : réduction QdV	Besoin 2 : réduction Perfo à embarquer
$\min_{\mathbf{Z}} T(\mathbf{Z})$ $\left\{ \begin{array}{l} A(\mathbf{Z}) \leq \bar{\epsilon}, \\ z_{i,j,q} \in \{0, 1\}, \forall i \in \{1, \dots, n\}, \\ \qquad \qquad \qquad \forall j \in \{1, \dots, L_i\}, \\ \qquad \qquad \qquad \forall q \in \{1, \dots, Q_{i,j}\}, \\ \sum_{q=1}^{Q_{i,j}} z_{i,j,q} = 1, \forall i \in \{1, \dots, n\}, \\ \qquad \qquad \qquad \forall j \in \{1, \dots, L_i\}. \end{array} \right.$	$\min_{\mathbf{Z}} S(\mathbf{Z})$ $\left\{ \begin{array}{l} T(\mathbf{Z}) \leq \bar{T}, \\ A(\mathbf{Z}) \leq \bar{\epsilon}, \\ z_{i,j,q} \in \{0, 1\}, \forall i \in \{1, \dots, n\}, \\ \qquad \qquad \qquad \forall j \in \{1, \dots, L_i\}, \\ \qquad \qquad \qquad \forall q \in \{1, \dots, Q_{i,j}\}, \\ \sum_{q=1}^{Q_{i,j}} z_{i,j,q} = 1, \forall i \in \{1, \dots, n\}, \\ \qquad \qquad \qquad \forall j \in \{1, \dots, L_i\}. \end{array} \right.$

TABLE 4.1 – Problèmes d'optimisation sous contraintes.

Dans la table 4.1, nous avons noté :

- $S(\mathbf{Z}) = \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} S_{i,j,q} z_{i,j,q}$  la taille mémoire du modèle réduit global  $\tilde{F}_{\mathbf{Z}}$ . C'est une fonction linéaire des variables de décision ;
- $T(\mathbf{Z}) = \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} T_{i,j,q} z_{i,j,q}$  le temps de calcul du modèle réduit global  $\tilde{F}_{\mathbf{Z}}$ . C'est aussi une fonction linéaire des variables de décision ;
- $A(\mathbf{Z}) = \|\epsilon_{\mathbf{Z}}\|_{\infty} = \|\tilde{F}_{\mathbf{Z}} - F\|_{\infty} = \max_{\mathbf{x}} (|\tilde{F}_{\mathbf{Z}}(\mathbf{x}) - F(\mathbf{x})|)$ , la norme infinie de la différence  $\epsilon_{\mathbf{Z}}$  entre le modèle réduit global  $\tilde{F}_{\mathbf{Z}}$  et le modèle de référence global  $F$  ;
- $\bar{T}$  le temps de calcul maximum acceptable pour des calculs en temps réels ;
- $\bar{\epsilon}$  l'erreur maximale acceptable entre le modèle réduit global  $\tilde{F}_{\mathbf{Z}}$  et le modèle original global  $F$ .

La fonction  $A$  qui dépend de  $\mathbf{Z}$  n'est pas une combinaison linéaire des variables de décision. L'existence d'une borne supérieure  $U$  linéaire par rapport à  $\mathbf{Z}$  est établie dans la section 4.2.3 :

$$U(\mathbf{Z}) = \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} a_{i,j,q} z_{i,j,q},$$

où  $a_{i,j,q}$  est obtenu en propageant l'erreur à travers les niveaux. Cette borne supérieure nous permet de reformuler les problèmes d'optimisation non linéaire de la table 4.1 sous la forme de problèmes d'optimisation linéaire en nombres entiers (cf. table 4.2) en remplaçant  $A(\mathbf{Z})$  par  $U(\mathbf{Z})$ .

Besoin 1	Besoin 2
$\min_{\mathbf{Z}} \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} T_{i,j,q} z_{i,j,q}$ $\left\{ \begin{array}{l} \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} a_{i,j,q} z_{i,j,q} \leq \bar{\epsilon}, \\ z_{i,j,q} \in \{0, 1\}, \forall i \in \{1, \dots, n\}, \\ \quad \forall j \in \{1, \dots, L_i\}, \\ \quad \forall q \in \{1, \dots, Q_{i,j}\}, \\ \sum_{q=1}^{Q_{i,j}} z_{i,j,q} = 1, \forall i \in \{1, \dots, n\}, \\ \quad \forall j \in \{1, \dots, L_i\}. \end{array} \right.$	$\min_{\mathbf{Z}} \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} S_{i,j,q} z_{i,j,q}$ $\left\{ \begin{array}{l} \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} T_{i,j,q} z_{i,j,q} \leq \bar{T}, \\ \sum_{i=1}^n \sum_{j=1}^{L_i} \sum_{q=1}^{Q_{i,j}} a_{i,j,q} z_{i,j,q} \leq \bar{\epsilon}, \\ z_{i,j,q} \in \{0, 1\}, \forall i \in \{1, \dots, n\}, \\ \quad \forall j \in \{1, \dots, L_i\}, \\ \quad \forall q \in \{1, \dots, Q_{i,j}\}, \\ \sum_{q=1}^{Q_{i,j}} z_{i,j,q} = 1, \forall i \in \{1, \dots, n\}, \\ \quad \forall j \in \{1, \dots, L_i\}. \end{array} \right.$

TABLE 4.2 – Programmation Linéaire en Nombres Entiers.

Ces problèmes d'optimisation linéaire en nombres entiers permettent d'effectuer la sélection, parmi tous les sous-modèles disponibles, de la meilleure combinaison possible. Les sous-modèles ainsi sélectionnés constituent le modèle réduit global.

### 4.2.3 Borne supérieure pour l'erreur globale

Dans cette section, l'erreur locale engendrée par l'utilisation de modèles réduits est propagée à travers les niveaux. Ceci nous mène à la formulation d'une erreur globale  $A(\cdot)$  qui est majorée par une fonction linéaire  $U(\cdot)$  des variables de décision  $\mathbf{Z}$ .

Pour le niveau 1, on a,  $\forall j_1 \in \{1, \dots, L_1\}$ ,

$$O_{1,j_1} = F_{1,j_1}(\mathbf{x}).$$

En notant  $\tilde{O}_{1,j_1}$  la sortie de l'évaluation en  $\mathbf{x}$  du modèle sélectionné pour remplacer  $F_{1,j_1}$ , comme un seul modèle réduit est sélectionné pour chaque sous-modèle, on peut écrire  $\forall j_1 \in \{1, \dots, L_1\}$ ,

$$\tilde{O}_{1,j_1} = \sum_{q=1}^{Q_{1,j_1}} \tilde{F}_{1,j_1,q}(\mathbf{x}) z_{1,j_1,q}.$$

En notant

$$\epsilon_{1,j_1,q}(\mathbf{x}) = \tilde{F}_{1,j_1,q}(\mathbf{x}) - F_{1,j_1}(\mathbf{x})$$

l'erreur locale engendrée par l'utilisation de  $\tilde{F}_{1,j_1,q}$  au lieu de  $F_{1,j_1}$  pour la variable d'entrée  $\mathbf{x}$ , on obtient facilement  $\forall j_1 \in \{1, \dots, L_1\}$ ,

$$\begin{aligned}\tilde{O}_{1,j_1} &= F_{1,j_1}(\mathbf{x}) + \sum_{q=1}^{Q_{1,j_1}} \epsilon_{1,j_1,q}(\mathbf{x}) z_{1,j_1,q} \\ &= O_{1,j_1} + \sum_{q=1}^{Q_{1,j_1}} \epsilon_{1,j_1,q}(\mathbf{x}) z_{1,j_1,q}.\end{aligned}\tag{4.1}$$

Le vecteur représentant les sorties des modèles sélectionnés au niveau 1 est noté :

$$\tilde{\mathbf{O}}_1 = (\tilde{O}_{1,1}, \dots, \tilde{O}_{1,L_1}).$$

Pour le niveau 2, on a  $\forall j_2 \in \{1, \dots, L_2\}$ ,

$$O_{2,j_2} = F_{2,j_2}(\mathbf{O}_1).\tag{4.2}$$

Comme au niveau 1, en notant  $\tilde{O}_{2,j_2}$  l'évaluation en  $\tilde{\mathbf{O}}_1$  du sous-modèle sélectionné pour remplacer le sous-modèle  $F_{2,j_2}$ , comme un seul modèle est sélectionné pour chaque sous-modèle, on obtient  $\forall j_2 \in \{1, \dots, L_2\}$ ,

$$\tilde{O}_{2,j_2} = \sum_{q=1}^{Q_{2,j_2}} \tilde{F}_{2,j_2,q}(\tilde{\mathbf{O}}_1) z_{2,j_2,q}.$$

Notons

$$\epsilon_{2,j_2,q}(\tilde{\mathbf{O}}_1) = \tilde{F}_{2,j_2,q}(\tilde{\mathbf{O}}_1) - F_{2,j_2}(\tilde{\mathbf{O}}_1)$$

l'erreur locale induite par l'utilisation de  $\tilde{F}_{2,j_2,q}$  en lieu et place de  $F_{2,j_2}$ . On obtient facilement  $\forall j_2 \in \{1, \dots, L_2\}$ ,

$$\tilde{O}_{2,j_2} = F_{2,j_2}(\tilde{\mathbf{O}}_1) + \sum_{q=1}^{Q_{2,j_2}} \epsilon_{2,j_2,q}(\tilde{\mathbf{O}}_1) z_{2,j_2,q}.\tag{4.3}$$

La notation  $\tilde{\mathbf{O}}_2 = (\tilde{O}_{2,1}, \dots, \tilde{O}_{2,L_2})$  est utilisée par la suite.

En utilisant (4.1), on a  $\forall j_2 \in \{1, \dots, L_2\}$ ,

$$F_{2,j_2}(\tilde{\mathbf{O}}_1) = F_{2,j_2}\left(O_{1,1} + \sum_{q=1}^{Q_{1,1}} \epsilon_{1,1,q}(\mathbf{x}) z_{1,1,q}; \dots; O_{1,L_1} + \sum_{q=1}^{Q_{1,L_1}} \epsilon_{1,L_1,q}(\mathbf{x}) z_{1,L_1,q}\right).$$

Comme  $\forall j_2 \in \{1, \dots, L_2\}$ ,  $F_{2,j_2}$  est dérivable en chaque point du segment  $[\mathbf{O}_1, \tilde{\mathbf{O}}_1]$ , le théorème des accroissements finis (pour plusieurs variables) garantit l'existence du point  $\zeta_{2,j_2} \in [\mathbf{O}_1, \tilde{\mathbf{O}}_1]$  tel que :

$$F_{2,j_2}(\tilde{\mathbf{O}}_1) = F_{2,j_2}(\mathbf{O}_1) + \sum_{j_1=1}^{L_1} \frac{\partial F_{2,j_2}}{\partial O_{j_1}}(\zeta_{2,j_2}) \left( \sum_{q=1}^{Q_{1,j_1}} \epsilon_{1,j_1,q}(\mathbf{x}) z_{1,j_1,q} \right).$$

Finalement, à partir de (4.2) et (4.3), on obtient  $\forall j_2 \in \{1, \dots, L_2\}$ ,

$$\begin{aligned} \tilde{O}_{2,j_2} = O_{2,j_2} &+ \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} \left( \frac{\partial F_{2,j_2}}{\partial O_{j_1}}(\boldsymbol{\zeta}_{2,j_2}) \epsilon_{1,j_1,q}(\mathbf{x}) \right) z_{1,j_1,q} \\ &+ \sum_{q=1}^{Q_{2,j_2}} \epsilon_{2,j_2,q}(\tilde{\mathbf{O}}_1) z_{2,j_2,q}. \end{aligned} \quad (4.4)$$

Pour le niveau **3**, on a  $\forall j_3 \in \{1, \dots, L_3\}$ ,

$$O_{3,j_3} = F_{3,j_3}(\mathbf{O}_2). \quad (4.5)$$

Pour les mêmes raisons et avec les mêmes notations on a  $\forall j_3 \in \{1, \dots, L_3\}$ ,

$$\tilde{O}_{3,j_3} = \sum_{q=1}^{Q_{3,j_3}} \tilde{F}_{3,j_3,q}(\tilde{\mathbf{O}}_2) z_{3,j_3,q}.$$

En notant,

$$\epsilon_{3,j_3,q}(\tilde{\mathbf{O}}_2) = \tilde{F}_{3,j_3,q}(\tilde{\mathbf{O}}_2) - F_{3,j_3}(\tilde{\mathbf{O}}_2),$$

on obtient facilement  $\forall j_3 \in \{1, \dots, L_3\}$ ,

$$\tilde{O}_{3,j_3} = F_{3,j_3}(\tilde{\mathbf{O}}_2) + \sum_{q=1}^{Q_{3,j_3}} \epsilon_{3,j_3,q}(\tilde{\mathbf{O}}_2) z_{3,j_3,q}. \quad (4.6)$$

En utilisant (4.4),  $\forall j_3 \in \{1, \dots, L_3\}$ ,

$$\begin{aligned} F_{3,j_3}(\tilde{\mathbf{O}}_2) = F_{3,j_3} &\left( \begin{array}{l} O_{2,1} \\ + \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} \left( \frac{\partial F_{2,1}}{\partial O_{j_1}}(\boldsymbol{\zeta}_{2,1}) \epsilon_{1,j_1,q}(\mathbf{x}) \right) z_{1,j_1,q} \\ + \sum_{q=1}^{Q_{2,1}} \epsilon_{2,1,q}(\tilde{\mathbf{O}}_1) z_{2,1,q} \\ , \dots , \\ O_{2,L_2} \\ + \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} \left( \frac{\partial F_{2,L_2}}{\partial O_{j_1}}(\boldsymbol{\zeta}_{2,L_2}) \epsilon_{1,j_1,q}(\mathbf{x}) \right) z_{1,j_1,q} \\ + \sum_{q=1}^{Q_{2,L_2}} \epsilon_{2,L_2,q}(\tilde{\mathbf{O}}_1) z_{2,L_2,q} \end{array} \right). \end{aligned}$$

Comme  $\forall j_3 \in \{1, \dots, L_3\}$ ,  $F_{3,j_3}$  est dérivable en chaque point du segment  $[\mathbf{O}_2, \tilde{\mathbf{O}}_2]$ , le théorème des accroissements finis (pour plusieurs variables) garantit l'existence du point

$\zeta_{3,j_3} \in [\mathbf{O}_2, \tilde{\mathbf{O}}_2]$ , tel que :

$$F_{3,j_3}(\tilde{\mathbf{O}}_2) = F_{3,j_3}(\mathbf{O}_2) + \sum_{j_2=1}^{L_2} \frac{\partial F_{3,j_3}}{\partial O_{j_2}}(\zeta_{3,j_3}) \left( \begin{aligned} & \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} \frac{\partial F_{2,j_2}}{\partial O_{j_1}}(\zeta_{2,j_2}) \epsilon_{1,j_1,q}(\mathbf{x}) z_{1,j_1,q} \\ & + \sum_{q=1}^{Q_{2,j_2}} \epsilon_{2,j_2,q}(\tilde{\mathbf{O}}_1) z_{2,j_2,q} \end{aligned} \right).$$

Finalement, à partir de (4.5) et (4.6), on obtient  $\forall j_3 \in \{1, \dots, L_3\}$ ,

$$\begin{aligned} \tilde{O}_{3,j_3} &= O_{3,j_3} \\ &+ \sum_{j_2=1}^{L_2} \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} \left( \frac{\partial F_{3,j_3}}{\partial O_{j_2}}(\zeta_{3,j_3}) \frac{\partial F_{2,j_2}}{\partial O_{j_1}}(\zeta_{2,j_2}) \epsilon_{1,j_1,q}(\mathbf{x}) \right) z_{1,j_1,q} \\ &+ \sum_{j_2=1}^{L_2} \sum_{q=1}^{Q_{2,j_2}} \frac{\partial F_{3,j_3}}{\partial O_{j_2}}(\zeta_{3,j_3}) \epsilon_{2,j_2,q}(\tilde{\mathbf{O}}_1) z_{2,j_2,q} \\ &+ \sum_{q=1}^{Q_{3,j_3}} \epsilon_{3,j_3,q}(\tilde{\mathbf{O}}_2) z_{3,j_3,q}. \end{aligned} \tag{4.7}$$

En utilisant l'inégalité triangulaire sur (4.7), on obtient  $\forall j_3 \in \{1, \dots, L_3\}$ ,

$$\begin{aligned} \|\tilde{O}_{3,j_3} - O_{3,j_3}\|_{\infty} &\leq \sum_{q=1}^{Q_{3,j_3}} a_{3,j_3,q} z_{3,j_3,q} \\ &+ \sum_{j_2=1}^{L_2} \sum_{q=1}^{Q_{2,j_2}} a_{2,j_2,q} z_{2,j_2,q} \\ &+ \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} a_{1,j_1,q} z_{1,j_1,q}, \end{aligned} \tag{4.8}$$

où  $\forall i \in \{1, 2, 3\}, \forall j_i \in \{1, \dots, L_i\}, \forall q \in \{1, \dots, Q_{i,j_i}\}$ ,

$$\begin{aligned} a_{3,j_3,q} &= \|\epsilon_{3,j_3,q}\|_{\infty} = \max_{\tilde{\mathbf{O}}_2} (|\epsilon_{3,j_3,q}(\tilde{\mathbf{O}}_2)|), \\ a_{2,j_2,q} &= \left\| \frac{\partial F_{3,j_3}}{\partial O_{j_2}} \right\|_{\infty} \|\epsilon_{2,j_2,q}\|_{\infty}, \\ &\text{avec } \left\| \frac{\partial F_{3,j_3}}{\partial O_{j_2}} \right\|_{\infty} = \max_{\zeta_{3,j_3}} \left( \left\| \frac{\partial F_{3,j_3}}{\partial O_{j_2}}(\zeta_{3,j_3}) \right\| \right) \text{ et } \|\epsilon_{2,j_2,q}\|_{\infty} = \max_{\tilde{\mathbf{O}}_1} (|\epsilon_{2,j_2,q}(\tilde{\mathbf{O}}_1)|), \\ a_{1,j_1,q} &= \sum_{j_2=1}^{L_2} \left\| \frac{\partial F_{3,j_3}}{\partial O_{j_2}} \right\|_{\infty} \left\| \frac{\partial F_{2,j_2}}{\partial O_{j_1}} \right\|_{\infty} \|\epsilon_{1,j_1,q}\|_{\infty}, \\ &\text{avec } \left\| \frac{\partial F_{2,j_2}}{\partial O_{j_1}} \right\|_{\infty} = \max_{\zeta_{2,j_2}} \left( \left\| \frac{\partial F_{2,j_2}}{\partial O_{j_1}}(\zeta_{2,j_2}) \right\| \right) \text{ et } \|\epsilon_{1,j_1,q}\|_{\infty} = \max_{\mathbf{x}} (|\epsilon_{1,j_1,q}(\mathbf{x})|). \end{aligned}$$

L'équation (4.8) peut être généralisée à  $n$  niveaux :

$$\begin{aligned}
 \left\| \tilde{O}_{n,1} - O_{n,1} \right\|_{\infty} &\leq \sum_{q=1}^{Q_{n,1}} a_{n,1,q} z_{n,1,q} \\
 &+ \sum_{j_{n-1}=1}^{L_{n-1}} \sum_{q=1}^{Q_{n-1,j_{n-1}}} a_{n-1,j_{n-1},q} z_{n-1,j_{n-1},q} \\
 &+ \dots \\
 &+ \sum_{j_3=1}^{L_3} \sum_{q=1}^{Q_{3,j_3}} a_{3,j_3,q} z_{3,j_3,q} \\
 &+ \sum_{j_2=1}^{L_2} \sum_{q=1}^{Q_{2,j_2}} a_{2,j_2,q} z_{2,j_2,q} \\
 &+ \sum_{j_1=1}^{L_1} \sum_{q=1}^{Q_{1,j_1}} a_{1,j_1,q} z_{1,j_1,q},
 \end{aligned} \tag{4.9}$$

où  $\forall i \in \{1, \dots, n-1\}, \forall j_i \in \{1, \dots, L_i\}, \forall q \in \{1, \dots, Q_{i,j_i}\}$

$$\begin{aligned}
 a_{n,1,q} &= \|\epsilon_{n,1,q}\|_{\infty}, \\
 a_{i,j_i,q} &= \sum_{j_{n-1}=1}^{L_{n-1}} \dots \sum_{j_{i+2}=1}^{L_{i+2}} \sum_{j_{i+1}=1}^{L_{i+1}} \left( \left\| \frac{\partial F_{n,1}}{\partial O_{j_{n-1}}} \right\|_{\infty} \dots \left\| \frac{\partial F_{i+3,j_{i+3}}}{\partial O_{j_{i+2}}} \right\|_{\infty} \right. \\
 &\quad \left. \left\| \frac{\partial F_{i+2,j_{i+2}}}{\partial O_{j_{i+1}}} \right\|_{\infty} \left\| \frac{\partial F_{i+1,j_{i+1}}}{\partial O_{j_i}} \right\|_{\infty} \|\epsilon_{i,j_i,q}\|_{\infty} \right).
 \end{aligned}$$

## 4.3 Expériences numériques

### 4.3.1 Application à un cas réel

Dans cette section, la méthode de réduction de modèle développée ci-dessus est appliquée au Fuel Consumption Model (FCM), qui prédit entre autres choses la consommation de carburant en vol, dans le but d'utiliser celui-ci comme une fonction embarquée dans le Flight Management System (FMS). Cette prédiction peut être recalculée durant le vol pour prendre en compte par exemple les changements de conditions météorologiques. Le FCM comprend les modèles de performances aérodynamiques et moteur ainsi qu'un calcul d'équilibre utilisant la seconde loi de Newton. Le modèle réduit FCM embarqué dans le FMS doit vérifier les exigences suivantes :

- la taille mémoire doit être aussi petite que possible,
- le temps de calcul doit être limité pour fournir des résultats en temps réel.

La sortie du modèle est un débit de carburant compris entre 200 kg/h et 9000 kg/h. Le FCM dépend de six variables : le facteur de charge longitudinal ( $Nx_{Aero}$ ), le nombre de Mach ( $Mach$ ), l'altitude ( $Z_g$ ), la température ( $\Delta ISA$ ), la masse ( $m$ ) et la position du

CHAPITRE 4. MÉTHODOLOGIE GÉNÉRALE DE RÉDUCTION D'UN MODÈLE DE QUALITÉS DE VOL ET DE PERFORMANCES DES AVIONS

centre de gravité  $G$  ( $XCG$ ). Comme le montre la figure 4.2, sa structure est hiérarchique multi-niveaux.

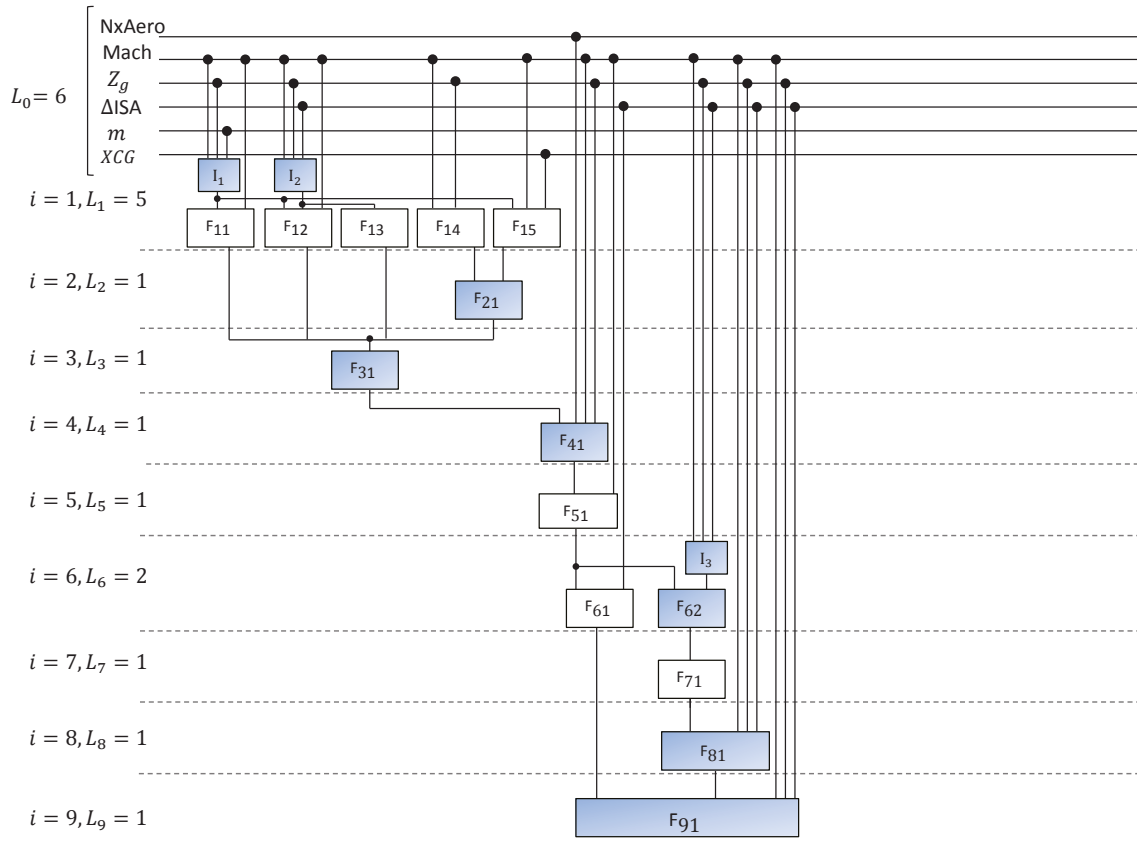


FIGURE 4.2 – Structure hiérarchique multi-niveaux du FCM.

Neuf des sous-modèles qui composent le FCM sont des fonctions analytiques. Le gain en termes de taille et de temps de calcul que l'on peut attendre de la réduction de ces fonctions analytiques n'est pas significatif; elles ne sont donc pas considérées dans le processus d'optimisation. Elles sont représentées par des boîtes colorées dans la figure 4.2. Puisque les fonctions notées  $I_1$ ,  $I_2$  et  $I_3$  dans la figure 4.2 ne sont pas réduites, elles n'introduisent pas d'erreur. De plus, puisqu'elles sont exclusivement reliées aux variables d'entrée, elles ne contribuent pas non plus à l'erreur de propagation (pas de dérivée pour le niveau 1 (4.1)). Elles peuvent donc être considérées comme faisant partie du sous-modèle qu'elles nourrissent. C'est pourquoi, le nombre de couches considérées est égal à neuf.

Chaque niveau est composé d'au plus cinq sous-modèles. Les entrées des sous-modèles intermédiaires sont aussi les sorties des sous-modèles des couches supérieures. La contribution majeure à la taille mémoire et au temps de calcul provient des huit sous-modèles représentés par des boîtes blanches dans la figure 4.2. Ils sont composés de tables numériques dans lesquelles sont réalisées des interpolations multi-linéaires. Plusieurs modèles

réduits de ces sous-modèles sont tout d'abord construits avec des caractéristiques différentes en terme de taille mémoire, de temps de calcul et de précision. Dans le but de simplifier le processus de validation, ces modèles réduits préservent la structure en table numérique. Pour l'application présentée ici, les tables numériques sont construites en utilisant le Random Sampling - High Dimensional Model Representation (RS-HDMR) et les réseaux de neurones (RN) décrits dans les sections 3.2.1.1 et 3.2.3. Le problème d'optimisation en nombres entiers est ensuite résolu (cf. Besoin 2 de la table 4.2). Finalement, le modèle réduit résultant est comparé au modèle FCM de référence au travers de sa taille mémoire, de son temps de calcul et de sa précision.

#### 4.3.2 Caractéristiques des sous-modèles de référence et de leurs modèles réduits associés

Les modèles réduits RS-HDMR et RN sont construits en utilisant les valeurs tabulées du sous-modèle original et sont eux-mêmes tabulés. Du fait de l'interpolation multilinéaire, l'erreur maximum pour chaque modèle réduit RS-HDMR et RN est atteinte aux points des tables numériques. Les caractéristiques des huit sous-modèles et de leurs modèles réduits associés (RS-HDMR 0<sup>ème</sup>, 1<sup>er</sup> et 2<sup>nd</sup>-ordre et RN) sont données dans les tables 4.3 (taille mémoire), 4.4 (temps de calcul) et 4.5 (précision).

La table 4.3 donne la taille mémoire exprimée en terme de nombre de valeurs tabulées. La table 4.4 fournit le temps de calcul d'un point arbitraire en ms sur un ordinateur portable. La table 4.5 présente l'erreur maximale relative au modèle de référence. Finalement, la table 4.6 donne la valeur de l'indice  $q$  correspondant à chaque modèle réduit disponible. Quand il n'y a pas de valeurs dans cette table, cela signifie que le modèle réduit associé n'était pas disponible ou écarté *a priori*.

Modèle	ref	ordre du RS-HDMR			RN
		2 <sup>nd</sup>	1 <sup>er</sup>	0 <sup>th</sup>	
$F_{11}$ 2D	415	—	81	—	121
$F_{12}$ 3D	504	237	49	—	125
$F_{13}$ 1D	40	—	—	1	27
$F_{14}$ 2D	27	—	19	—	18
$F_{15}$ 3D	474	234	49	—	125
$F_{51}$ 2D	2820	—	457	—	329
$F_{61}$ 2D	305	—	67	—	134
$F_{71}$ 1D	12	—	—	1	10

TABLE 4.3 – Taille mémoire  $S$  : nombre de valeurs tabulées.

Modèle	ref	ordre du RS-HDMR			RN
		2 <sup>nd</sup>	1 <sup>er</sup>	0 <sup>ème</sup>	
$F_{11}$ 2D	6.673	—	$7.117 \cdot 10^{-1}$	—	34.65
$F_{12}$ 3D	4.97	4.655	$6.655 \cdot 10^{-1}$	—	35.48
$F_{13}$ 1D	4.668	—	—	$6.7 \cdot 10^{-4}$	34.99
$F_{14}$ 2D	5.281	—	$2.884 \cdot 10^{-1}$	—	34.46
$F_{15}$ 3D	4.802	3.702	$6.863 \cdot 10^{-1}$	—	32.65
$F_{51}$ 2D	5.591	—	$5.201 \cdot 10^{-1}$	—	35.14
$F_{61}$ 2D	6.126	—	$5.098 \cdot 10^{-1}$	—	40.32
$F_{71}$ 1D	4.645	—	—	$7.27 \cdot 10^{-4}$	35.83

TABLE 4.4 – Temps de calcul  $T$  en ms.

Modèle	ref	ordre du RS-HDMR			RN
		2 <sup>nd</sup>	1 <sup>er</sup>	0 <sup>ème</sup>	
$F_{11}$ 2D	0	—	723.54	—	45.039
$F_{12}$ 3D	0	1.673	3.98	—	5.316
$F_{13}$ 1D	0	—	—	38.9	1.739
$F_{14}$ 2D	0	—	1600	—	1642
$F_{15}$ 3D	0	32.62	70.615	—	10.54
$F_{51}$ 2D	0	—	$1.271 \cdot 10^7$	—	$1.630 \cdot 10^4$
$F_{61}$ 2D	0	—	$1.282 \cdot 10^7$	—	$1.084 \cdot 10^6$
$F_{71}$ 1D	0	—	—	295	77.14

TABLE 4.5 – Erreur  $\|\epsilon\|_\infty (\cdot 10^{-6})$ .

Modèle	ref	ordre du RS-HDMR			RN
		2 <sup>nd</sup>	1 <sup>er</sup>	0 <sup>ème</sup>	
$F_{11}$ 2D	1	—	2	—	3
$F_{12}$ 3D	1	2	3	—	4
$F_{13}$ 1D	1	—	—	2	3
$F_{14}$ 2D	1	—	2	—	3
$F_{15}$ 3D	1	2	3	—	4
$F_{51}$ 2D	1	—	2	—	3
$F_{61}$ 2D	1	—	2	—	3
$F_{71}$ 1D	1	—	—	2	3

TABLE 4.6 – Valeurs de l'indice  $q$ .

### 4.3.3 Résultats de l'optimisation

Dans cette section, le problème d'optimisation linéaire en nombres entiers décrits dans la table 4.2 est résolu afin de trouver la meilleure combinaison de modèles. Pour ce faire, l'algorithme Branch and Bound [85] *bintprog* de MATLAB est utilisé. Dans cet exemple, le nombre de variables est 26 et le nombre de contraintes est 36. Pour simplifier l'interprétation des résultats, deux expériences ont été réalisées dans lesquelles nous avons étudié séparément l'impact de chaque contrainte sur la fonction objectif.

#### 4.3.3.1 Variation de la taille mémoire en fonction de l'erreur

Dans un premier temps, nous avons réalisé plusieurs optimisations pour des valeurs différentes de la contrainte de précision  $\bar{\epsilon}$ . La figure 4.3 présente la variation de la taille mémoire optimale en fonction de la tolérance  $\bar{\epsilon}$  sur l'erreur maximum globale, le temps de calcul maximum autorisé  $\bar{T}$  étant fixé à la valeur prescrite pour ce scénario spécifique ( $\bar{T} = 500$  ms). Plus précisément, les solutions optimales sont représentées en fonction de  $\bar{\beta}$  qui est un pourcentage du plus petit débit de carburant  $WFE_{min}$  :  $\bar{\beta} = 100 \frac{\bar{\epsilon}}{WFE_{min}}$ .

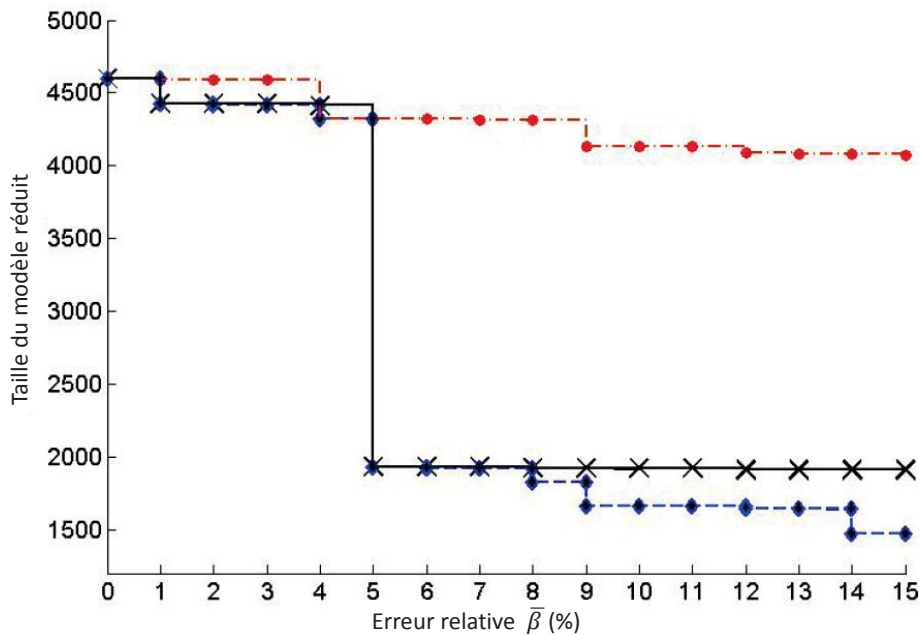


FIGURE 4.3 – Visualisation de la taille mémoire de la solution optimale en fonction de la contrainte de précision.

Les croix noires sont les résultats des optimisations lorsque seuls le RN et les modèles de référence sont proposés à l'optimiseur. Les cercles rouges sont les solutions optimales avec seulement les RS-HDMR et les modèles de référence. Enfin, les losanges bleus représentent les solutions avec les deux sortes de modèles réduits (RS-HDMR et RN) et les modèles de référence. Dans ce cas test, une réduction de 50 % en terme de taille mémoire a été obtenue

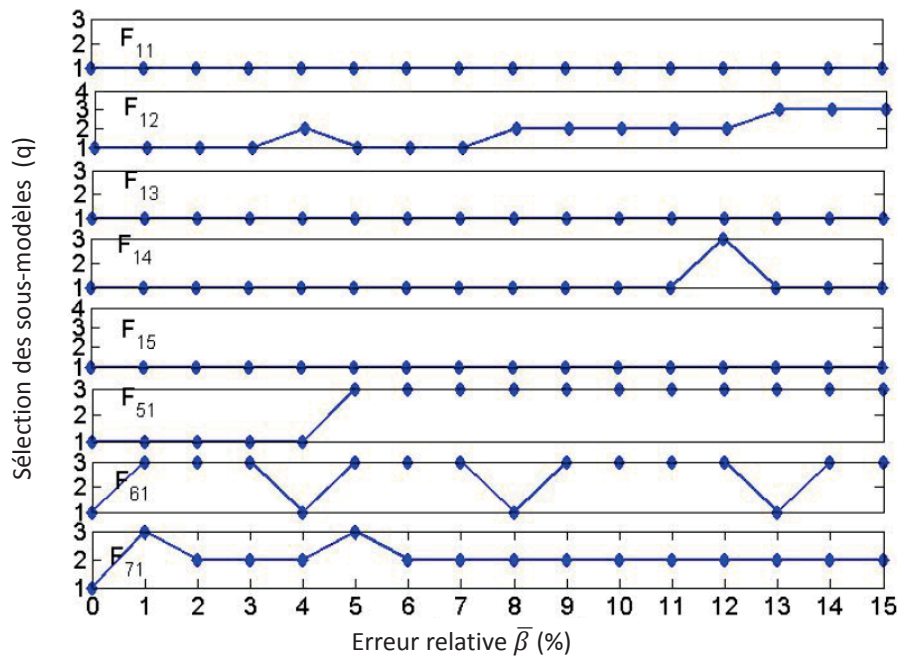


FIGURE 4.4 – Sous-modèles optimaux en fonction de la contrainte de précision.

pour une erreur relative supérieure à 5 %. L'ajout du RS-HDMR permet d'améliorer la taille du modèle réduit global seulement lorsque la tolérance sur l'erreur acceptable est portée à 8 %.

On remarque que la taille mémoire décroît lorsque la tolérance sur l'erreur augmente. La valeur de la contrainte de l'erreur  $\bar{\epsilon}$  varie entre 0 kg/h et 30 kg/h. Ceci donne dans le pire des cas, une variation de l'erreur relative  $\bar{\beta}$  entre 0 % et 15 %. Nous avons choisi de représenter la fonction constante entre les points effectivement testés (représentés par des croix, cercles et losanges). En effet, comme l'ensemble des erreurs relatives maximales globales testées est discret (tous les 1 %) seuls les croix, losanges ou cercles représentent des choix optimaux. Cette fonction constante entre deux solutions optimales présentée dans la figure 4.3 est conservative. En effet, il est possible qu'il existe une solution meilleure entre deux points représentés, mais il n'est pas possible qu'il en existe une moins bonne.

Finalement, nous soulignons le fait que plus le nombre de modèles réduits proposés est important, meilleure est la solution. Pour chacun des huit sous-modèles que l'on cherche à réduire, la variation des choix optimaux parmi la décomposition RS-HDMR, le RN et le sous-modèle de référence en fonction de la tolérance sur l'erreur est représentée dans la figure 4.4. Ces variations sont commentées ci-après.

- Pour  $\bar{\epsilon} = 0$ , tous les sous-modèles originaux ( $q = 1$ ) sont sélectionnés (pas de réduction).
- L'erreur locale due aux sous-modèles réduits  $\tilde{F}_{11}$ ,  $\tilde{F}_{13}$  and  $\tilde{F}_{15}$  a une forte influence sur la sortie car les sous-modèles de référence sont toujours gardés.

- L'erreur locale due au sous-modèle réduit  $\tilde{F}_{51}$  a une faible influence sur la sortie globale du modèle réduit : la plus grande réduction possible est choisie pour une faible erreur globale acceptée.
- $\tilde{F}_{12}$ ,  $\tilde{F}_{14}$ ,  $\tilde{F}_{61}$  and  $\tilde{F}_{71}$  sont les éléments clés de la réduction. Quand la contrainte sur l'erreur change, l'ordre de la décomposition RS-HDMR et les RN sélectionnés varient afin d'obtenir la meilleure solution en taille mémoire. Grossièrement, on peut dire que le processus d'optimisation a trouvé le meilleur équilibre entre ces quatre sous-modèles.

#### 4.3.3.2 Variation de la taille mémoire en fonction du temps de calcul

Dans notre cas industriel, la valeur requise pour  $\bar{T}$  est assez grande pour rendre la contrainte correspondante inactive. Néanmoins, il est intéressant d'étudier l'impact de la variation de  $\bar{T}$  sur la solution, car d'autres modèles avions peuvent être très sensibles à ce paramètre. On résout plusieurs fois le problème d'optimisation pour plusieurs valeurs de  $\bar{T}$ . La figure 4.5 ci-dessous représente la variation de la taille mémoire optimale en fonction de la contrainte de temps de calcul global  $\bar{T}$ , pendant que  $\bar{\beta}$  est fixé à une valeur arbitraire ( $\bar{\beta} = 10\%$ ).

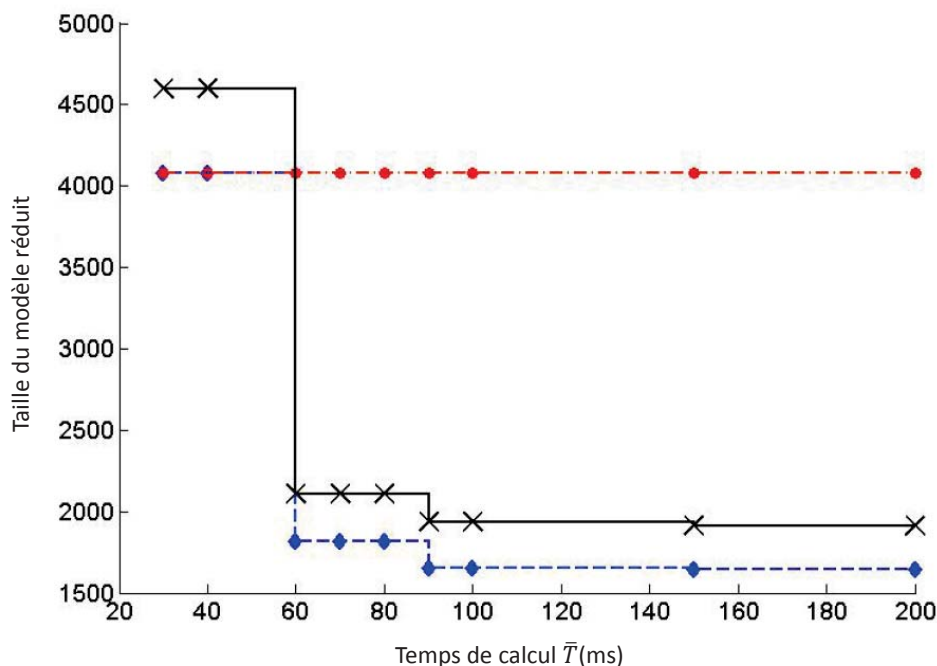


FIGURE 4.5 – Visualisation de la taille mémoire des solutions optimales en fonction de la contrainte de temps.

Comme décrit à la section 4.3.3.1, les croix noires sont les résultats des optimisations avec seulement les modèles de référence et le RN, les cercles rouges sont les solutions

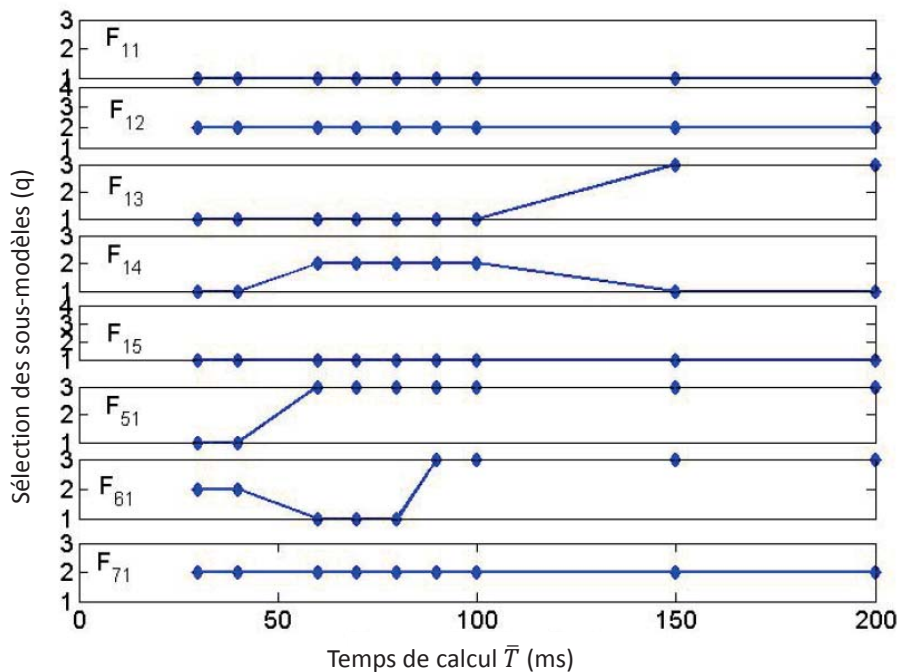


FIGURE 4.6 – Sous-modèles optimaux en fonction de la contrainte de temps.

optimales avec seulement les modèles de référence et les RS-HDMR et finalement les losanges bleus sont les solutions avec les deux modèles proposés.

Lorsque seuls les RS-HDMR et les modèles de référence sont proposés à l'optimiseur, la solution optimale est stationnaire en terme de taille mémoire quelle que soit la tolérance  $\bar{T}$  imposée sur le temps de calcul. En effet, pour un sous-modèle donné, l'optimiseur choisit parmi des modèles dont la taille et le temps de calcul augmentent en même temps que la précision (cf. tables 4.3 et 4.4). De ce fait, comme la précision est fixée à une valeur donnée, le choix parmi les RS-HDMR et les modèles de référence seuls ne permettent pas de diminuer la taille du modèle global lorsque la contrainte de temps de calcul augmente ; seule la contrainte de précision a permis de choisir parmi les RS-HDMR.

Dans le cas où les RN et les modèles de référence sont proposés, plus la contrainte en temps de calcul est relâchée ( $\bar{T}$  augmente), plus il y a de chance de choisir un RN. En effet, la taille de ces derniers est plus petite que celle des modèles de référence associés, mais leur temps d'appel est plus long (cf. tables 4.3 et 4.4). Une réduction importante est réalisée quand le temps de calcul est supérieur à 60 ms. Évidemment, quand les RS-HDMR, les RN et les modèles de référence sont proposés à l'optimiseur, la solution est encore meilleure et n'est plus triviale.

La figure 4.6 montre le changement de solution optimale entre le RS-HDMR, RN et le modèle de référence en fonction du temps de calcul. Notons que lorsque la contrainte de temps de calcul est inférieure à 30 ms, il n'y a pas de solution admissible qui peut aussi satisfaire la contrainte de précision. Pour une précision donnée, une borne inférieure

existe toujours sur le domaine admissible en terme de temps de calcul.

### 4.3.3.3 Bilan des expériences numériques

Finalement, les résultats présentés dans les figures 4.4 et 4.6 font apparaître un comportement non trivial en terme de choix de modèle réduit. En particulier, quand le RS-HDMR, le RN et le modèle de référence sont proposés à l'optimiseur, la solution optimale est composée d'un mélange de ces trois modèles qui n'est pas évident à faire *a priori*.

## 4.4 Conclusion

Nous avons développé dans ce chapitre une méthodologie permettant de trouver la meilleure combinaison de modèles dans un catalogue de modèles réduits existants. Cette méthode permet de réduire des modèles hiérarchiques multi-niveaux de grande dimension tout en conservant une architecture compatible avec nos besoins spécifiques. Sa faisabilité et son efficacité sont démontrées sur une fonction existante de calcul de performances des avions et destinée à être embarquée, le FCM. Sur ce cas test, un échantillon de sous-modèles réduits a été construit en utilisant des décompositions RS-HDMR d'ordres différents ainsi que des réseaux de neurones. Le choix optimal de modèles réduits proposé par l'optimiseur n'est pas intuitif.



# Conclusion générale

Dans le chapitre 1, nous avons introduit le contexte industriel de cette thèse. Nous avons mis en évidence un besoin d'harmonisation de création des modèles entre les départements QdV et Perfo. Actuellement, ces deux disciplines s'appuient sur des modèles très proches. Cependant, comme leurs développements sont menés indépendamment, elles ne partagent presque aucune donnée. De ce fait, certaines activités (de recalage par exemple) sont dupliquées. Le nouveau processus étudié considère le modèle QdV comme un modèle de référence. Des simplifications physiques lui sont alors apportées dans ce chapitre pour répondre aux exigences Perfo. Ce faisant, un socle commun entre ces deux disciplines est préservé et nous tirons bénéfice du modèle QdV pour améliorer la représentativité du modèle Perfo. Le temps d'appel au modèle QdV simplifié reste toutefois supérieur aux exigences Perfo, ce qui implique un besoin de réduction mathématique.

Par ailleurs, le besoin de fournir de plus en plus d'informations au pilote en temps réel sur les performances de l'avion mène à une diminution de la taille mémoire disponible dans les processeurs embarqués. Le modèle Perfo doit donc être à son tour réduit pour répondre aux exigences d'embarquabilité.

L'objectif est donc de proposer une méthode automatique de réduction qui satisfasse ces exigences en termes de temps de calcul, de précision, ainsi que de taille mémoire pour certaines applications.

Pour cela, nous avons tout d'abord proposé dans le chapitre 2 une méthodologie détaillée de réduction de modèles. Puis un état de l'art critique des méthodes de réduction, d'échantillonnage et de validation susceptibles de répondre à notre problématique a été établi. Il nous a permis de faire les choix suivants.

- Modèles réduits

Nous avons retenu comme méthode de réduction automatique le Random Sampling - High Dimensional Model Representation (RS-HDMR) qui permet de décomposer automatiquement le modèle de référence en une somme de fonctions dépendant d'un nombre réduit de variables. Nous avons également retenu les réseaux de neurones (RN) pour leur propriété de parcimonie : ils nécessitent moins de paramètres que d'autres modèles réduits pour approcher une fonction continue avec une précision arbitraire.

– Méthodes d'échantillonnage

En petite dimension, typiquement pour la réduction de taille ou de temps de calcul des fonctions tabulées, seuls les points déjà tabulés sont utilisés, puisque l'interpolation multi-linéaire garantit que c'est en ces points que l'erreur est maximale. Pour les modèles de grande dimension QdV et Perfo, nous avons choisi d'utiliser indifféremment les suites de Halton ou de Sobol, dont le but est d'assurer un bon remplissage de l'espace, pour réaliser l'échantillonnage de la fonction.

– Méthodes de validation

En petite dimension, la tabulation est possible et l'erreur est calculée exactement aux points de la grille. Pour les cas de grande dimension, le modèle réduit global est validé par une estimation *a posteriori* de l'erreur. L'échantillon généré est donc divisé en un ensemble d'apprentissage et un ensemble de test.

Dans le chapitre 3, nous avons appliqué avec succès le RS-HDMR et les RN à la réduction des fonctions tabulées de petite dimension. Pour ce faire, nous avons développé un calcul exact des termes de la décomposition RS-HDMR. De plus, nous avons étendu l'applicabilité du RS-HDMR au cas non hypercubique en optimisant un prolongement de la fonction en dehors de son domaine de définition pour rendre ce domaine hypercubique.

Par ailleurs, nous avons proposé un emploi original des RN en utilisant ces derniers pour régénérer uniquement les points de la fonction tabulée.

Toutefois, le fléau de la dimension rend impossible la construction et la validation de tels modèles réduits en un temps acceptable pour les modèles QdV et Perfo.

Pour pallier cette difficulté, nous avons exploité la structure de ces modèles. Ils sont composés d'un assemblage de sous-modèles représentant différents phénomènes physiques. L'idée est d'utiliser les techniques décrites au chapitre précédent pour construire des tables de dimensions inférieures faciles à valider. Dans ce cadre-là, nous avons donc développé dans le chapitre 4 une méthodologie permettant de trouver la meilleure combinaison de modèles réduits de ces tables choisis dans un catalogue déjà existant. Sa faisabilité et son efficacité ont été démontrées sur une fonction existante de calcul de performances des avions destinée à être embarquée.

Nous avons donc été confrontés au fléau de la dimension pour la construction du modèle RS-HDMR en grande dimension. Des outils de calcul en parallèle permettraient d'estimer simultanément les termes de la décomposition RS-HDMR et ainsi d'étendre l'utilisation de cette méthode dans le cas où le vecteur des variables d'entrée est de grande dimension et lorsque le domaine de définition de la fonction est non hypercubique.

Concernant le problème d'optimisation linéaire en nombres entiers proposé dans le cha-

pitre 4, le travail futur sera de proposer une estimation moins conservative de l'erreur. En effet, la majoration systématique par les dérivées maximales des différents sous-modèles a montré son efficacité, mais pourrait être améliorée si la propagation de l'erreur prenait en compte la dérivée effective des sous-modèles pour le point considéré. Dans ces conditions, l'erreur devient une fonction des variables d'entrées réelles et des variables de décisions entières, ce qui rend nécessaire la mise en œuvre d'un algorithme d'optimisation non-linéaire en variables mixtes, Mixed Integer Non-Linear Programming Problem (MINLP) [6, 14, 15].



# Références

- [1] M. BALESDENT, J. MORIO et J. MARZAT : Kriging-based adaptive Importance Sampling algorithms for rare event estimation. *Structural Safety*, 44:1–10, 2013.
- [2] A. BANERJEE, D.B. DUNSON et S.T. TOKDAR : Efficient Gaussian Process Regression for Large Datasets. *Biometrika*, 100:75–89, 2013.
- [3] A. BANERJEE, J. VOGELSTEIN et D.B. DUNSON : Parallel inversion of huge covariance matrices. <http://arxiv.org/abs/1312.1869>, 2013.
- [4] A. R. BARRON : Universal Approximation Bounds for Superpositions of a Sigmoidal Function. *IEEE Transaction Information Theory*, 39:930–945, 1993.
- [5] V. BAUDOUI : *Optimisation robuste multiobjectifs par modèles de substitution*. Thèse de doctorat, ISAE, 2012.
- [6] P. BELOTTI, C. KIRCHES, S. LEYFFER, J. LINDEROTH, J. LUEDTKE et A. MAHAJAN : Mixed-Integer Nonlinear Optimization. *Acta Numerica*, 22:1–131, 2013.
- [7] C.M. BISHOP : *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., 1995.
- [8] J.L BOIFFIER : *The Dynamics of Flight, The Equations*. Wiley, 1998.
- [9] M. BONDOUY, S JAN, S. LAPORTE et C BES : Optimal Surrogates Selection for Embedded, Hierarchical Multilevel Aircraft Models. *IEEE Transactions on Aerospace and Electronic Systems*, 2015.
- [10] P. BORNE, Dauphin-Tanguy G., J.-P. RICHARD et F. ROTELLA : *Modélisation et Identification des Processus*. Technip, 1992.
- [11] L. BRANDOLINI, L. COLZANI, G. GIGANTE et G. TRAVAGLINI : On the Koksma-Hlawka inequality. *Journal of Complexity*, 29:158 – 172, 2013.
- [12] P. BRATLEY, B.L FOX et L.E SCHRAGE : *A Guide to Simulation*. Springer-Verlag New York, Inc., 1986.
- [13] C. G. BROYDEN : The Convergence of a Class of Double-rank Minimization Algorithms. *Journal of the Institute of Mathematics and Its Applications*, 6:76–90, 1970.
- [14] S. CAFIERI et N. DURAND : Aircraft deconfliction with speed regulation : new models from mixed-integer optimization. *Journal of Global Optimization*, 58:613–629, 2014.
- [15] S. CAFIERI, J. LEE et L. LIBERTI : On Convex Relaxations of Quadrilinear Terms. *Journal of Global Optimization*, 47:661–685, 2010.
- [16] R.E. CALFISF : Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica*, pages 1–49, 1998.

- [17] T. CHAI et R. R. DRAXLER : Root mean square error (RMSE) or mean absolute error (MAE)? - arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, 7:1247–1250, 2014.
- [18] P. CHAUVET : *Aide-mémoire de géostatistique linéaire*. Mines Paris-les Presses ParisTech, 2008.
- [19] A.R. CONN, N.I.M. GOULD et P.L. TOINT : *Trust Region Methods*. Society for Industrial and Applied Mathematics, 2000.
- [20] J. A. CORNELL : *Response Surfaces : Designs and Analyses*. Marcel Dekker, Inc., 1987.
- [21] N.A.C CRESSIE : *Statistics for Spatial Data*. Wiley, 1993.
- [22] G. CYBENKO : Approximation by Superpositions of a Sigmoidal Function. *Mathematics of Control, Signals, and Systems*, 2:303–314, 1989.
- [23] S. DA VEIGA : Global Sensitivity Analysis with Dependence Measures. <https://hal.archives-ouvertes.fr/hal-00903283>, 2013.
- [24] J.-P. DEMAILLY : *Analyse numérique et équations différentielles*. Presses universitaires de Grenoble, 1996.
- [25] G. DREYFUS, J. M. MARTINEZ, M. SAMUELIDES, M. B GORDON, F. BADRAN, S. THIRIA et L. HÉRAULT : *Réseaux de neurones*. Eyrolles, 2002.
- [26] B. EFRON et C. STEIN : The Jackknife Estimate of Variance. *The Annals of Statistics*, 9:586–596, 1981.
- [27] R. FAIVRE, B. IOOSS, S. MAHÉVAS, D. MAKOWSKI et H. MONOD : *Analyse de sensibilité et exploration de modèles : Application aux sciences de la nature et de l’environnement*. Quae, 2013.
- [28] B. FEIL, S. KUCHERENKO et N. SHAH : Comparison of Monte Carlo and Quasi Monte Carlo Sampling Methods in High Dimensional Model Representation. *In Advances in System Simulation, 2009. SIMUL '09. First International Conference on*, pages 12–17, 2009.
- [29] R.A. FISHER : *The Design of Experiments*. Oliver and Boyd, 1935.
- [30] R. FLETCHER : A New Approach to Variable Metric Algorithms. *Computer Journal*, 13:317–322, 1970.
- [31] J. FRANCO : *Planification d’expériences numériques en phase exploratoire pour la simulation des phénomènes complexes*. Thèse de doctorat, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008.
- [32] J.H. FRIEDMAN : Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19:1–67, 1991.
- [33] L. S. GANDIN : *Objective Analysis of Meteorological Fields*. Gidrometeorologicheskoe Izdatel’stvo (GIMIZ), 1963.
- [34] A.A. GIUNTA et L.T. WATSON : A comparison of approximation modeling techniques - Polynomial versus interpolating models. *In the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 1998.
- [35] D. GOLDFARB : A Family of Variable Metric Updates Derived by Variational Means. *Mathematics of Computation*, 24:23–26, 1970.

- [36] Y. GRATTON : Le Krigeage : La Méthode Optimale d'Interpolation Spatiale. *Les articles de l'Institut d'Analyse Géographique*, 2002.
- [37] J. H. HALTON : On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2:84–90, 1960.
- [38] M. HAMDAOUI, F.-Z. OUJEBBOUR, A. HABBAL, P. BREITKOPF et P. VILLON : Kriging surrogates for evolutionary multi-objective optimization of CPU intensive sheet metal forming applications. *International Journal of Material Forming*, 1960-6206:12, 2014.
- [39] J. M. HAMMERSLEY : Monte Carlo Methods for Solving Multivariable Problems. *Annals of the New York Academy of Sciences*, 86:844–874, 1960.
- [40] J. M. HAMMERSLEY et K. W. MORTON : A new Monte Carlo technique : Antithetic Variates. *Mathematical Proceedings of the Cambridge Philosophical Society*, 52:449–475, 1956.
- [41] J.M. HAMMERSLEY et D.C. HANDSCOMB : *Monte Carlo methods*. Methuen & Co. Ltd, 1964.
- [42] T. HASTIE, R. TIBSHIRANI et J. FRIEDMAN : *The Elements of Statistical Learning : Data Mining, Inference, and Prediction*. Springer, 2003.
- [43] S. HAYKIN : *Neural Networks : A Comprehensive Foundation*. Prentice Hall, 1999.
- [44] W. HOEFFDING : A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19:293–325, 1948.
- [45] R. HOOKE et T. A. JEEVES : Direct Search Solution of Numerical and Statistical Problems. *Journal of the ACM*, 8:212–229, 1961.
- [46] G. HOOKER : Generalized Functional ANOVA Diagnostics for High-Dimensional Functions of Dependent Variables. *Journal of Computational and Graphical Statistics*, 16:709–732, 2007.
- [47] K. HORNIK, M. STINCHCOMBE et H. WHITE : Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.
- [48] K. HORNIK, M.B. STINCHCOMBE, H. WHITE et P. AUER : Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and Their Derivatives. *Neural Computation*, 6:1262–1275, 1994.
- [49] G. JAMES et T. HASTIE : Generalizations of the Bias/Variance Decomposition for Prediction Error. Rapport technique, Department of Statistics, Stanford University, Stanford, CA, 1997.
- [50] M.E JOHNSON, L.M MOORE et Ylvisajer D. : Minimax and Maximin Distance Designs. *Journal of Statistical Planning and Inference*, 26:131 – 148, 1990.
- [51] P. JOUBERT DE LA MOTTE : *Schéma à pas de temps caractéristique pour l'aérodynamique transsonique et conception aéroélastique optimale d'aile de grand allongement*. Thèse de doctorat, Arts et Métiers ParisTech, 2007.
- [52] S.L. JUN : *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.
- [53] J.P.C. KLEIJNEN et R.G. SARGENT : A Methodology for Fitting and Validating Metamodels in Simulation. *European Journal of Operational Research*, 120:14 – 29, 2000.

- [54] D. G. KRIGE : A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society*, 52:119–139, 1951.
- [55] L. KUIPERS et H. NIEDERREITER : *Uniform Distribution of Sequences*. John Wiley & Sons, 1974.
- [56] R. LE RICHE : A review of optimization methods for the identification of complex models. Lecture, 2014.
- [57] C. LEMIEUX : *Monte Carlo and Quasi-Monte Carlo Sampling*. Springer Series in Statistics. Springer, 2009.
- [58] G. LI et H. RABITZ : General formulation of HDMR component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, 50:99–130, 2012.
- [59] G. LI, S.W. WANG et H. RABITZ : Practical Approaches To Construct RS-HDMR Component Functions. *The Journal of Physical Chemistry A*, 106:8721–8733, 2002.
- [60] G. LI, S.W. WANG, H. RABITZ, S. WANG et P. JAFFE : Global uncertainty assessments by high dimensional model representations (HDMR). *Chemical Engineering Science*, 57:4445–4460, 2002.
- [61] J. LOGAN : *Modélisation des forces de contact entre le pneu d’un avion et la piste*. Thèse de doctorat, ISAE, 2012.
- [62] G. MATHERON : Principles of Geostatistics. *Economic Geology*, 58:1246–1268, 1963.
- [63] M.D. MCKAY : Latin Hypercube Sampling As a Tool in Uncertainty Analysis of Computer Models. *In the 24th Conference on Winter Simulation Conference*, 1992.
- [64] N. METROPOLIS, A.W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER et E. TELLER : Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.
- [65] N. METROPOLIS et S. M. ULAM : The Monte Carlo Method. *Journal of the American Statistical Association*, 44:335–341, 1949.
- [66] M. MINSKY : Matter, Mind and Models. *In IFIP Congress 65*, 1965.
- [67] M. MONGEAU et C. BES : Optimization of Aircraft Container Loading. *IEEE Transactions on Aerospace and Electronic Systems*, 39:140–150, 2003.
- [68] R.H. MYERS et D.C. MONTGOMERY : *Response Surface Methodology : Process and Product in Optimization Using Designed Experiments*. John Wiley & Sons, 1995.
- [69] J. A. NELDER et R. MEAD : A Simplex Method for Function Minimization. *Computer Journal*, 7:308–313, 1965.
- [70] H. NIEDERREITER : Point Sets and Sequences with Small Discrepancy. *Monatshefte für Mathematik*, 104:273–338, 1987.
- [71] H. NIEDERREITER : *Random Number Generation and quasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics, 1992.
- [72] A.B. OWEN : Randomly Permuted (t,m,s)-Nets and (t,s)-Sequences. *In Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, volume 106, pages 229–317. 1994.

- [73] L. PRONZATO et W.G. MÜLLER : Design of Computer Experiments : Space Filling and Beyond. *Statistics and Computing*, 22:681–701, 2012.
- [74] H. RABITZ, O.F ALIS, J. SHORTER et K. SHIM : Efficient input-output model representations. *Computer Physics Communication*, 117:11–20, 1999.
- [75] A. RANGANATHAN : The Levenberg-Marquardt Algorithm. [http://excelsior.cs.ucsb.edu/courses/cs290i\\_mvg/pdf/lma.pdf](http://excelsior.cs.ucsb.edu/courses/cs290i_mvg/pdf/lma.pdf), 2004.
- [76] D. F. SHANNO : Conditioning of Quasi-Newton Methods for Function Minimization. *Mathematics of Computation*, 24:647–656, 1970.
- [77] I.M. SOBOL : On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7:86 – 112, 1967.
- [78] I.M. SOBOL : Theorems and examples on high dimensional model representation. *Reliability Engineering and Systems Safety*, 79:187–193, 2003.
- [79] B. TALGORN, S. LAPORTE, C. BES et S. SEGONDS : One Engine Out Take-off Trajectory Optimization. *In the 10th AIAA Aviation Technology, Integration, and Operations Conference*, 2010.
- [80] R.M. TORRES CALDERON, J. CHAPTAL, C. BES et J.-B. HIRIART-URRUTY : Optimal, Environmentally Friendly Departure Procedures for Civil Aircraft. *Journal Of Aircraft*, 48:11–22, 2011.
- [81] B. VANDEWOESTYNE et R. COOLS : Good permutations for deterministic scrambled Halton sequences in terms of  $L_2$ -discrepancy. *In the 11th International Congress on Computational and Applied Mathematics*, 2006.
- [82] V. N. VAPNIK : *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1995.
- [83] F. VARENNE : Epistémologie des modèles et des simulations. <https://hal.archives-ouvertes.fr/hal-00674144>, 2008.
- [84] R.S. WIKRAMARATNA : The additive congruential random number generator - A special case of a multiple recursive generator. *Journal of Computational and Applied Mathematics*, 216:371 – 387, 2008.
- [85] L.A. WOLSEY : *Integer programming*. John Wiley & Sons, 1998.
- [86] T. ZIEHN et A.S. TOMLIN : GUI-HDMR - A Software Tool for Global Sensitivity Analysis of Complex Models. *Environmental Modelling and Software*, 24:775–785, 2009.

**AUTHOR :** Manon Bondouy

**TITLE:** Surrogate modeling construction for aircraft performances computation

**SUPERVISORS:** Christian Bès and Sophie Jan

**PLACE AND DATE OF THE DEFENSE:** University of Toulouse, 08/02/2016

**ABSTRACT:** The objective of this thesis is to provide a methodology and the associated tools in order to standardize the building process of performance and handling quality models. This typically leads to elaborate surrogate models in order to satisfy industrial contrasting objectives of memory size, accuracy and computation time. After listing the different steps of a construction of surrogates methodology and realizing a critical state of the art, Neural Networks and High Dimensional Model Representation methods have been selected and validated on low dimension functions. For functions of higher dimension, a reduction method based on the optimal selection of submodel surrogates has been developed which allows to satisfy the requirements on accuracy, computation time and memory size. The efficiency of this method has been demonstrated on an aircraft performance model which will be embedded into the avionic systems.

**KEYWORDS:** Handling Quality, Aircraft Performance, Surrogate model, High Dimension Model Representation, Neural Networks.

**ADMINISTRATIVE FIELD:** Applied Mathematics.

**UNIVERSITY ADDRESS:** Aircraft Performance domain. Airbus Operations S.A.S. 316 route de Bayonne, 31300 Toulouse and Institute of Mathematics of Toulouse, Paul Sabatier University, 118 Route de Narbonne, 31062 Toulouse.

**AUTEUR** : Manon Bondouy

**TITRE** : Construction de modèles réduits pour le calcul des performances des avions

**DIRECTEURS DE THÈSE** : Christian Bès et Sophie Jan

**LIEU ET DATE DE SOUTENANCE** : Université de Toulouse, 08/02/2016

**RÉSUMÉ** : L'objectif de cette thèse est de mettre en place une méthodologie et les outils associés en vue d'harmoniser le processus de construction des modèles de performances et de qualités de vol. Pour ce faire, des techniques de réduction de modèles ont été élaborées afin de satisfaire des objectifs industriels contradictoires de taille mémoire, de précision et de temps de calcul. Après avoir établi une méthodologie de construction de modèles réduits et effectué un état de l'art critique, les Réseaux de Neurones et le High Dimensional Model Representation ont été choisis, puis adaptés et validés sur des fonctions de petite dimension. Pour traiter les problèmes de dimension supérieure, une méthode de réduction basée sur la sélection optimale de sous-modèles réduits a été développée, qui permet de satisfaire les exigences de rapidité, de précision et de taille mémoire. L'efficacité de cette méthode a finalement été démontrée sur un modèle de performances des avions destiné à être embarqué.

**MOTS-CLÉS** : Qualité de Vol, Performances des avions, Modèle réduit, High Dimensional Model Representation, Réseaux de neurones.

**DISCIPLINE ADMINISTRATIVE** : Mathématiques Appliquées.

**ADRESSE DU LABORATOIRE** : Bureau des Performances des avions, Airbus Operations S.A.S., 316 route de Bayonne, 31300 Toulouse et Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse.