HAL
open science

Amélioration de la quantification des images TEP à l'yttrium 90

Maël Millardet

To cite this version:

Maël Millardet. Amélioration de la quantification des images TEP à l'yttrium 90. Algorithme et structure de données [cs.DS]. École centrale de Nantes, 2022. Français. NNT : 2022ECDN0016 . tel-03871632v2

HAL Id: tel-03871632

https://theses.hal.science/tel-03871632v2

Submitted on 25 Nov 2022

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

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

THESE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : Signal, Image, Vision

Par
Maël MILLARDET

Amélioration de la quantification des images TEP à l'yttrium 90

Thèse présentée et soutenue à l'Ecole Centrale de Nantes le 11 avril 2022
Unité de recherche : UMR 6004, Laboratoire des Sciences du Numérique de Nantes (LS2N)

Rapporteurs avant soutenance :

Claude COMTAT	Ingénieur de recherche HDR, Université Paris Saclay
Nicolas DOBIGEON	Professeur des universités, IRIT Toulouse

Composition du Jury :

Président :	Mai NGUYEN-VERGER	Professeure des universités, Cergy Paris Université
Examinatrices :	Emilie CHOUZENOUX	Chargée de recherche HDR, Université Paris Saclay
	Voichita MAXIM	Maître de conférences HDR, INSA Lyon
Dir. de thèse :	Saïd MOUSSAOUI	Professeur des universités, École Centrale de Nantes
Co-dir. de thèse :	Thomas CARLIER	Ingénieur de recherche HDR, CHU de Nantes
Co-encadrante :	Diana MATEUS LAMUS	Professeure des universités, Ecole Centrale de Nantes

Invités

Françoise KRAEBER-BODERE	Professeure, CHU de Nantes
Jérôme IDIER	Directeur de recherche CNRS, École Centrale de Nantes

Remerciements

Cette thèse n'aurait pas pu être menée à bien sans l'aide de nombreuses personnes, que je tiens à remercier.

Je tiens avant tout à remercier mes directeurs de thèse Saïd Moussaoui, Thomas Carlier et mon encadrante Diana Mateus. Je souhaite tout d'abord adresser un immense merci à Thomas Carlier pour avoir accepté d'encadrer mon stage de fin d'études, puis pour m'avoir proposé de continuer en thèse avec lui. C'est grâce à lui que l'aventure a pu commencer.

Je dois également adresser un immense merci à Saïd Moussaoui. J'ai en particulier beaucoup apprécié sa gentillesse, et ses capacités d'écoute. Tous ses conseils m'ont été très précieux.

Je dois également beaucoup à Diana Mateus, pour son aide, mais aussi pour avoir permis de financer ma thèse grâce à la chaire MILCOM.

Même s'ils ne faisaient pas partie de l'encadrement officiel, d'autres personnes m'ont apporté une aide scientifique précieuse. En particuliers Jérôme Idier, Simon Stute et Clément Bailly.

Il me faut également adresser un grand merci à toute l'équipe SIMS du LS2N pour leur accueil chaleureux. ainsi qu'à l'équipe 2 du CRCI^{2}NA, et à tout le service de médecine nucléaire.

Je remercie tous les membres du jury Claude Comtat, Nicolas Dobigeon, Emilie Chouzenoux, Mai Nguyen-Verger, Maxim Voichita Françoise Kraeber Bodéré et Jérôme Idier pour la relecture de ma thèse et leurs retours précieux.

Je remercie aussi les membres de mon comité de suivi individuel Clovis Tauber et Myriam Servières, qui m'ont soutenu et conseillé lors de mes CSI annuels.

Un grand merci également aux personnes ayant partagé mon bureau : Ramzi et Olivier pour leur accueil et leurs nombreux conseils en début de thèse, puis Dawood, Guillaume et Jitendra. Un immense merci également à mes collègues et amis Ludivine, Julian, Diane, Vanessa, Samuel, Guillaume, Valentin et Adrien. Ce groupe d'amis doctorants a vraiment été précieux tout au long de ces trois années.

Je remercie également tous mes autres très bons amis, en particulier mes colo-

cataires pendant cette thèse Stanislas, Benoît, Veronika, Louis-Marie et Matthieu, ainsi que mes parents, grands-parents, oncles et tantes, frères, sœurs, cousins et cousines.

Table des matières

Introduction générale **21**

Contexte clinique et objectifs 23

Orgaisation du manuscrit 25

Production scientifique de la thèse 26

1 État de l'art de la reconstruction TEP et problématiques spécifiques liées à l'yttrium 90 **27**

1.1 La Tomographie par émission de positons 29

1.1.1 Introduction 29

1.1.2 Principe physique 29

1.1.3 Caméra TEP 30

1.1.4 Stockage des données 31

1.1.5 Corrections 32

1.1.6 Modélisation du problème 33

1.1.7 Algorithmes de reconstruction 38

1.2 L'yttrium 90 et l'émission de positons 46

1.2.1 Introduction 46

1.2.2 Le premier état excité 0^+ du zirconium 90 47

1.2.3 Décomposition de l'yttrium 90 51

1.2.4 Problématiques spécifiques liées à la TEP à l'yttrium 90 . . . 53

2 Post-traitement **55**

2.1 Introduction 57

2.2 Formulation du problème 58

2.2.1 Principe 58

2.2.2 Prérequis 59

2.2.3 Formulation mathématique du problème 59

2.2.4 Propriétés de H 61

2.2.5 Exigence supplémentaire sur alpha 63

2.2.6 Choix de la norme 63
2.2.7 Existence et unicité de la solution 65
2.3 Formulation en termes d'optimisation linéaire et résolution à l'aide d'une version améliorée de la méthode duale du simplexe 65
2.3.1 Formulation en termes d'optimisation linéaire 65
2.3.2 La méthode duale du simplexe 67
2.3.3 Résolution à l'aide d'une version améliorée de la méthode duale du simplexe 68
2.3.4 Détails pratiques de mise en œuvre 70
2.4 Illustrations numériques 74
2.4.1 Exemple sur un fantôme 75
2.4.2 Étude d'un laplacien **H** non isotrope 78
2.4.3 Comparaison entre CPLEX et l'algorithme proposé 78
2.5 Résultats expérimentaux 81
2.5.1 Mise en œuvre expérimentale 81
2.5.2 Images 82
2.5.3 Figures de mérite 82
2.6 Discussion et conclusion 88

3 Analyse comparative 91
3.1 Algorithmes pénalisés sans contrainte 93
3.2 Cadre expérimental 94
3.2.1 Acquisitions du fantôme 94
3.2.2 Application clinique 94
3.2.3 Paramètres de la reconstruction 95
3.2.4 Métriques d'évaluation 96
3.2.5 Analyse multi-objectifs 97
3.2.6 Effet sur le calcul dosimétrique 100
3.3 Résultats 103
3.3.1 Comparaison de NEG-ML et d'AML 103
3.3.2 Comparaison d'AML et de PPGML 104
3.3.3 Évaluation globale des différents paramètres 104
3.3.4 Comparaison entre la pénalité quadratique et la pénalité aux différences relatives 107

3.3.5 Évaluation dosimétrique 108
3.3.6 Évaluation clinique 108
3.4 Discussion et conclusion 109

4 Utilisation de l'apprentissage profond pour le réglage des hyper-paramètres en reconstruction itérative 121
4.1 Introduction 123
4.2 Méthodologie générale 123
4.2.1 Présentation de l'approche 123
4.2.2 Constitution de la base de donnée pour l'apprentissage supervisé 125
4.2.3 Configuration expérimentale 129
4.2.4 Architecture 130
4.2.5 Entraînement du réseau 131
4.2.6 Méthodes « hyper-paramètres fixés » 131
4.3 Résultats 132
4.3.1 Validation croisée 132
4.3.2 Corrélation entre les variables 133
4.3.3 Analyse et discussion 136
4.4 Conclusion 142

5 Conclusion générale 143
5.1 Conclusion 145
5.2 Perspectives 145

A Annexe au chapitre 1 147
A.1 Preuve de la proposition 1 149
A.2 Preuve de la proposition 3 149
A.3 Preuve de la proposition 4 150
A.4 Preuve de la proposition 5 150
A.5 Preuve de la proposition 6 151
A.6 Coefficients de H utilisés dans cette étude 153
A.6.1 Cas des pixels carrés et des voxels cubiques 153
A.6.2 Cas des pixels rectangulaires et des voxels parallélépipédiques 154
A.6.3 Remarques conclusives 154

B Annexe au chapitre 3 **155**
B.1 Normalisation des hyper-paramètres de reconstruction 157
B.2 Hyper-paramètres parcourus lors qu parcours initial 158

Bibliographie **159**

Table des figures

1 Schéma expliquant la façon dont un patient est pris en charge dans le cadre d'une radioembolisation. D'après Arnaud Dieudonné 25

1.1 schéma d'une caméra TEP et d'un tube photomultiplicateur 30

1.2 Formation d'un sinogramme et reconstruction d'une image de cerveau 31

1.3 Description des quatre types de coïncidences 32

1.4 Illustration du biais positif introduit par le maximum de vraisemblance sous contrainte de positivité (MLEM). FBP est un algorithme analytique présenté dans la section 1.1.7.1 Image prise d'une présentation faisant suite à [Van Slambrouck et al., 2015] 36

1.5 Évolution de MLEM au cours des itérations 41

1.6 Modèle en couches 48

1.7 États excités du zirconium 90 50

1.8 Décroissance de l'yttrium 90 52

2.1 Deux stratégies de post-traitement non satisfaisantes. La première (a) peut mener à des transferts asymétriques. La deuxième (b) ne contient aucun aspect local. 60

2.2 Simulations d'un post-traitement n'imposant pas la positivité de $\boldsymbol{\alpha}$. Le fantôme utilisé est celui de la section 2.4.1. Certains pixels de la zone tiède en bordure de la zone froide « aident » les pixels négatifs proches à s'annuler en transférant leur valeur positive, ce qui érode les contours, et donne une valeur artificiellement élevée à leurs autres voisins. 64

2.3 Optimisation du vecteur $\boldsymbol{z}$ en appliquant la méthode duale du simplexe au NNEPPS (2.3a) et à un autre problème d'optimisation linéaire sans structure particulière (2.3b). La base $\mathcal{B}$ est représentée à chaque itération. Les indices entrants sont encadrés en vert et les indices sortants en rouge. 71

2.4 RMSE normalisé de l'image finale en fonction de la tolérance sur l'erreur relative de la solution des systèmes linéaires. Le RMSE est normalisé par la racine carrée de la valeur quadratique moyenne de l'image théorique. 72

2.5 RMSE normalisé de l'image issue de l'étape d'initialisation en fonction du temps de calcul. Le RMSE est normalisé par la racine carrée de la valeur quadratique moyenne de l'image théorique. 74

2.6 Différentes images du même fantôme. (a) montre l'image initiale. Les différentes parties sont remplies par un processus aléatoire suivant une distribution gaussienne de déviation standard 1, avec une moyenne de 0, 3, ou 6 selon les différentes zones. L'image non bruitée est présentée dans la figure 2.2a. (b) montre l'image après l'étape d'initialisation. (c) montre l'image après le NNEPPS. Les régions avec une activité moyenne de 0 sont nettoyées par rapport à l'image initiale, tandis que les autres zones restent inchangées. (d) montre le biais introduit par le NNEPPS, moyenné sur 1000 réalisations. 76

2.7 Représentation histogramme des deux images de la figure 2.6. L'image brute est en bleu, l'image post-traitée en rouge. Dans l'image post-traitée, 10 079 pixels appartiennent à la case contenant 0. 77

2.8 Biais et variance introduits par le NNEPPS, pour différents choix de $\boldsymbol{H}$ moyenné sur 4000 réalisations. De haut en bas, on constate clairement que le biais diminue, mais que la variance augmente. . . . 79

2.9 Comparaison du temps nécessaire au post-traitement d'un exemple d'image de taille $100 \times 100 \times n$. Les images TEP réelles ont généralement une taille de $200 \times 200 \times 109$ 80

2.10 Reconstruction du fantôme avec OSEM, AML et AML après le NNEPPS (AML+NNEPPS). Dans AML+NNEPPS, la zone froide centrale est beaucoup plus visible. 83

2.11 Reconstructions d'un patient traité à l'yttrium 90 avec OSEM, AML et AML+NNEPPS. Le NNEPPS supprime le bruit de fond introduit par AML. Les images affichées sont des fusions TEP-tomodensitométrie. Le patient fut traité pour des localisations hépatiques d'une tumeur neuroendocrine. Activité injectée : 2190 MBq (1709 MBq au moment de l'imagerie). Shunt pulmonaire : 4 %. Nombre de coïncidences : $3,1 \times 10^6$. Fraction de coïncidences aléatoires : 81 %. 84

2.12 Graphes affichant l'activité moyenne dans le cylindre froid en fonction du RMSE dans la même zone. Les tests furent effectués pour différentes valeurs du paramètre A d'AML, la borne inférieure de l'image [Van Slambrouck et al., 2015]. On obtint les différents points de la même courbe en faisant varier le nombre d'itérations après lequel on arrête AML. PP signifie NNEPPS. Les courbes pour IEC0 et IEC2 ne sont pas montrées mais sont proches de celles obtenues avec IEC1, qui est représentatif du comportement standard d'AML et du NNEPPS. IEC3 est un cas rare pour lequel AML sous-estime complètement l'activité dans le cylindre froid. 85

2.13 Graphes représentant l'activité moyenne mesurée en fonction de la distance au centre du cylindre froid. La valeur théorique et la reconstruction OSEM sont données à titre de références. Nous n'avons pas représenté les courbes pour IEC0 et IEC2, mais elles sont proches de celle d'IEC1, qui est représentatif du comportement standard d'AML et du NNEPPS. IEC3 est un cas rare pour lequel AML sous-estime complètement l'activité dans le cylindre froid. 87

3.1 ROI pour le fond 96

3.2 Activité récupérée dans le cylindre froid en fonction du RMSE dans ce même cylindre (a) et du $\mathrm{CRC}_{\mathrm{chaud}}$ en fonction du IR (b) pour AML et NEG-ML. Chaque point représente un nombre différent d'itérations. Les courbes sont pour IEC1 98

3.3 Comparaison d'AML et de NEG-ML avec et sans le NNEPPS en utilisant le fantôme IEC1 pour différentes métriques. Dans chaque graphique, les quatre fronts de Pareto 2D de deux variables (le nombre d'itérations et A ou ψ) sont représentés. L'OSEM de référence et deux autres algorithmes (choisis empiriquement pour leur bon compromis) sont également représentés. Notez que ces algorithmes se trouvent sur le front de Pareto 4D, mais pas nécessairement sur les fronts 2D-Pareto représentés ici. Notez également que bien que toutes les données soient prises en compte dans chaque graphique, les points extrêmes des fronts de Pareto peuvent être différents d'un graphique à l'autre, car un point dominé dans un front 2D peut ne pas l'être dans un autre. Par exemple, l'IR de AML+NNEPPS varie de 0,5 à 1,6 sur la figure. 3.3a et de 0,5 à 1,0 sur la figure. 3.3c. 99

3.4 Représentation du front de Pareto (à droite) et de l'ensemble de Pareto (à gauche). Image prise de [Davins-Valldaura, 2016]. 101

3.5 Comparaison d'AML (avec et sans le NNEPPS) et de PPGML (pénalité quadratique) en utilisant le fantôme IEC1 pour différentes métriques. Dans chaque graphique, les quatre fronts de Pareto 2D de 2 paramètres variables (le nombre d'itérations et A ou β) sont représentés. L'OSEM de référence et un exemple pour chaque autre algorithme (choisi empiriquement pour leur bon compromis) sont représentés. Notez que ces algorithmes se trouvent sur le front 4D-Pareto, mais pas nécessairement sur les fronts 2D-Pareto représentés ici. 102

3.6 Comparaison des quatre paramètres (intensité de la pénalité β, borne inférieure A, nombre d'itérations et inclusion ou non du NNEPPS) en utilisant le fantôme IEC1 pour différentes métriques (APPGML vs PPGML). Dans chaque graphique, les trois fronts de Pareto 2D sont représentés. Les points représentant l'OSEM de référence ainsi que deux algorithmes pertinents sont également représentés. Il est à noter que ces algorithmes se trouvent sur le front de Pareto 4D, mais pas nécessairement sur les fronts de Pareto 2D représentés ici. 105

3.7 Images de l'IEC1 provenant des trois algorithmes mis en évidence sur la figure 3.6. Bien que les caractéristiques de biais soient bonnes pour l'algorithme de la figure 3.7c, il apparaît visuellement très bruité. Le cylindre froid central est beaucoup plus visible sur la figure 3.7d que sur l'OSEM de référence. 106

3.8 Comparaison des différences relatives et des pénalités quadratiques dans le fantôme IEC1 pour différentes métriques. Dans chaque graphique, les croix représentent le front de Pareto 4D de l'ensemble des algorithmes contenant à la fois des différences relatives et des pénalités quadratiques. Les lignes représentent les ensembles de Pareto 2D (correspondant au graphique) des différences relatives et de la pénalité quadratique. Pour chaque ligne, il existe deux paramètres variables (le nombre d'itérations et la force de la pénalité β) 111

3.9 Histogramme dose-volume d'une région d'intérêt à l'intérieur du cylindre froid pour IEC0. « SIEMENS » représente l'algorithme mis en œuvre par le fabricant. 112

3.10 Histogramme dose-volume de la sphère de 22 mm pour IEC0. SIEMENS représente l'algorithme mis en œuvre par le fabricant. 112

3.11 HDV du patient pour la région d'intérêt censée être libre de toute activité. SIEMENS représente l'algorithme mis en œuvre par le fabricant. 114

3.12 HDV dans le foie en utilisant différents algorithmes. SIEMENS représente l'algorithme mis en œuvre par le fabricant. 116

3.13 HDV dans une tumeur en utilisant différents algorithmes. SIEMENS représente l'algorithme mis en œuvre par le fabricant.. 117

3.14 Image de fusion (TEP/tomodensitométrie) d'un patient avec cinq algorithmes de reconstruction différents 118

3.15 Image de fusion (TEP/tomodensitométrie) d'un patient avec cinq algorithmes de reconstruction différents. L'échelle de couleur est très saturée afin de mettre en évidence le biais positif de certains algorithmes. Certaines taches, attribuées au bruit, sont visibles dans les images OSEM et SIEMENS,et sont absentes des images APPGML + NNEPPS et PPGML 119

4.1 Récapitulatif des entrées et sorties de l'algorithme proposé. 124

4.2 Exemple d'image de la base de donnée ainsi que les reconstructions optimales et prédites par le réseau. On voit sur l'image vérité terrain une tumeur, une zone uniforme et une zone froide. 126
4.3 Courbe d'entraînement du patient 6 134
4.4 Courbe de validation du MSSIM du patient 6 135
4.5 courbe de validation de l'optimisation « métriques » du patient 6 . . . 135

Liste des tableaux

2.1 Durée nécessaire au post-traitement avec et sans l'étape d'initialisation 73
2.2 Caractéristiques des différentes zones 77
2.3 Caractéristiques des cartes de biais et de variance des images post-traitées, pour différents choix de $\boldsymbol{H}$. 80
2.4 Caractéristiques des quatre acquisitions du fantôme NEMA IEC . . . 82

3.1 Métriques dosimétriques du fantôme IEC0 110
3.2 Tableau récapitulatif des caractéristiques de chaque algorithme sur le fantôme IEC1[12] 113
3.3 Liste de tous les paramètres testés 117
3.4 Métriques dosimétriques du patient 120

4.1 Résultats de la validation croisée SSIM 133
4.2 Résultats de la validation croisée Métriques 133
4.3 Résultats de la validation croisée SSIM loss 134
4.4 Résultats de la validation croisée Métriques loss 134
4.5 Corrélation, moyenne, variance, erreur quadratique moyenne des différents hyper-paramètres. L'ensemble de validation est le patient 6. . 137

B.1 Liste de tous les paramètre testés 158

Liste des sigles et acronymes

Cette liste répertorie tous les sigles et acronymes utilisés dans cette thèse. Pour chacun, la page de la première apparition de l'acronyme, ou dans certains cas celle de sa présentation la plus complète, est indiquée.

ABEMML : algorithme de reconstruction admettant une borne inférieure A et une borne supérieure B, p. 43

Adam : adaptive moment estimation, p.131

ADMM : alternating direction method of multipliers, p.45

ADMM-Reg : ADMM régularisé, p.45

AML : algorithme de reconstruction admettant une borne inférieure A, p. 43

APPGML : PPGML avec une borne inférieure A, p.93

ART : algebraic reconstruction techniques (techniques de reconstruction algébriques), p.39

BSREM : block sequential regularized expectation maximization, p.43

CASToR : customizable and advanced software for tomographic reconstruction, p.81

CHU : centre hospitalier universitaire, p.23

CPLEX : outil informatique d'optimisation, p.78

CPU : central processing unit (unité centrale de traitement), p.58

CRC$_{chaud}$: coefficient de récupération de la concentration moyenne dans une zone chaude, p.96

CRC$_{froid}$: activité moyenne dans le cylindre froid, p.97

CRCI^{2}NA : centre de recherche en cancérologie et immunologie intégrée Nantes Angers, p.23

D70 : dose reçue par au moins 70 % du volume, p.102

D$_{avg}$: dose moyenne délivrée, p.102

DTPA : acide diéthylène triamine penta acétique, p.81

EM : expectation maximization, p.39

FBP : filtered back-projection (rétroprojection filtrée), p.38

^{18}**F-FDG** : fluorodésoxyglucose (^{18}F), p.29

Fixe$_{métriques}$, p.131
Fixe$_{MSE}$, p.131
Fixe$_{MSSIM}$, p.131
FOM : figure of merit (figure de mérite), p.95
FOV : field of view (champ de vue), p.82
FWHM : full width at half maximum (largeur à mi-hauteur), p.95
HDV : histogramme dose-volume, p.100
HypoC-PML : hypo-convergence for penalized maximum-likelihood, p.45
IEEE : institute of electrical and electronics engineers, p.58
IR : image roughness (rugosité de l'image), p.96
IRM : imagerie par résonance magnétique, p.89
KKT : conditions d'optimalité de Karush-Kuhn-Tucker, p.66
LS2N : laboratoire des sciences du numérique de Nantes, p.23
LSO : lutetium orthosilicate, p.53
MAP : maximum a posteriori, p.35
MLEM : maximum likelihood expectation maximization, p.39
MSE : mean squared error (erreur quadratique moyenne), p.78
MSSIM : mean structural similarity, p.124
NEG-ML : algorithme admettant des valeurs négatives, p.44
NEMA IEC : nom du fantôme utilisé, p.81
NNEPPS : non-negativity enforcement post-processing step, p.58
OP-OSEM : ordinary Poisson OSEM, p.100
OSEM : ordered subsets expectation maximization, p.42
PPGML : penalized preconditioned gradient maximum likelihood, p.43
PSF : point spread function (fonction d'étalement du point), p.33
RAM : random access memory (mémoire vive), p.78
ResNet-18 : réseau de neurones utilisé, p.130
RMSE : root mean squared error (racine carrée de l'erreur quadratique moyenne), p.72, p.97
ROI : region of interest (région d'intérêt), p.94
TEP : tomographie par émission de positons, p.24
TOF : time of flight (temps de vol), p.95
V50 : pourcentage du volume ayant reçu plus de 50 Gy, p.97
VMT : valeur moyenne théorique, p.75

Introduction générale

Sommaire

Contexte clinique et objectifs **23**
Orgaisation du manuscrit **25**
Production scientifique de la thèse **26**

Cette thèse a été réalisée au sein de l'équipe Signal Image et Son du LS2N en collaboration avec le laboratoire CRCI^{2}NA du CHU de Nantes. Cette collaboration entre dans le cadre d'un projet commun portant sur la reconstruction et l'analyse d'images multimodales et le développement de méthodes adaptées au contexte clinique.

Contexte clinique et objectifs

Les différents types de cancers, pris dans leur ensemble, causent chaque année environ neuf millions de décès, et les seuls cancers du foie sont responsables de huit cent mille d'entre eux. Il existe trois méthodes permettant de traiter ces tumeurs hépatiques, qu'elles soient primaires ou métastatiques : la chirurgie, la chimiothérapie et la radiothérapie. Mais lorsque le cancer atteint un stade avancé, la chirurgie est bien souvent impossible, et la chimiothérapie devient alors le traitement le plus fréquent. Cependant, dans le cas où la tumeur est entièrement comprise dans le foie, un traitement local, appelé radiothérapie interne sélective, ou radioembolisation, est possible.

Pour comprendre le fonctionnement de cette technique, il faut noter que bien que ce soit principalement la veine porte qui irrigue le foie sain en sang, les tumeurs ont tendance à se créer leur propre vascularisation, issue de l'artère hépatique. Ainsi, un radioélément injecté dans la vascularisation artérielle du foie irradiera les tumeurs en ayant un effet moindre sur le foie sain. De plus, les microsphères contenant le radioélément s'accumuleront dans les artérioles et les capillaires en les obstruant, et empêcheront de la sorte l'irrigation en sang. Voilà pourquoi l'on parle de radioembolisation.

Après l'injection des microsphères, il est bénéfique de déterminer leur distribution, et ce pour deux raisons. D'abord, une accumulation extrahépatique des microsphères, qui arrive parfois en dépit d'un bon ciblage des macro-agrégats d'albumine, peut provoquer des complications. Ensuite, la distribution intrahépatique des microsphères est un bon indicateur de l'efficacité du traitement.

Jusqu'à récemment, la tomographie monophotonique était très utilisée à cette fin, à cause de l'absence d'alternatives. Mais celle-ci souffre d'une mauvaise résolution spatiale (entre 11 et 15 mm), qui explique en partie le biais de quantification, allant de 58 % pour une distribution d'activité homogène dans une sphère de 37 mm de

rayon, à 75 % pour une sphère de 10 mm [Elschot et al., 2011]. En dépit des récents efforts pour améliorer les algorithmes de reconstruction (tels ceux concernant la correction des coïncidences atténuées ou diffusées), les performances quantitatives de la tomographie monophotonique demeurent faibles pour les lésions de moins de 30 mm, ce qui la rend peu adaptée aux analyses dose-réponse.

On sait depuis les années 1950 que la désintégration de l'yttrium 90 produit des positons, avec un rapport d'embranchement de $3{,}2 \times 10^{-5}$. En 2004, Nickles et al. ont montré la possibilité de réaliser une imagerie TEP grâce à cette production. Mais hélas ! à cause du faible nombre de positons émis, la reconstruction est difficile : on parle de « très basses statistiques ».

Ainsi, l'imagerie TEP à l'yttrium 90 est une alternative à la tomographie monophotonique permettant d'améliorer la qualité des images, mais son utilisation clinique à des fins quantitatives demeure une gageure.

Préalablement à l'injection des microsphères, on utilise des macro-agrégats d'albumine marqués par le technétium 99m comme substituts des microsphères d'yttrium 90, à des fins de simulation. On effectue ensuite une tomographie monophotonique combinée à une tomodensitométrie pour vérifier l'absence de shunt pulmonaire. Ce dispositif permet de confirmer que l'on cible bien la tumeur et d'ajuster l'activité d'yttrium 90 à injecter. Le schéma 1 décrit la façon dont les patients sont pris en charge.

Accroître la qualité et la précision des images TEP à l'yttrium 90 est un enjeu important pour améliorer le suivi des patients. Certains algorithmes ont été proposés pour tenter de répondre à cette problématique, mais une comparaison générale de ces algorithmes dans le contexte spécifique de l'yttrium 90 manquait encore à l'appel au début de cette thèse. De plus, les algorithmes en question contiennent plusieurs hyper-paramètres difficiles à choisir. Enfin, les algorithmes autorisant des valeurs négatives dans l'image ne sont pas utilisables tels quels dans les études dosimétriques.

Cette thèse a comme premier objectif d'étudier, en les combinant ces différentes méthodes, dans le cadre de la TEP à l'yttrium 90, avec un accent particulier sur la réduction du biais positif dans les zones froides. Un deuxième objectif de la thèse est d'analyser l'influence des hyper-paramètres des algorithmes de reconstruction sur l'évaluation des métriques extraites des images produites afin de proposer des stratégies d'optimisation de leurs réglage.

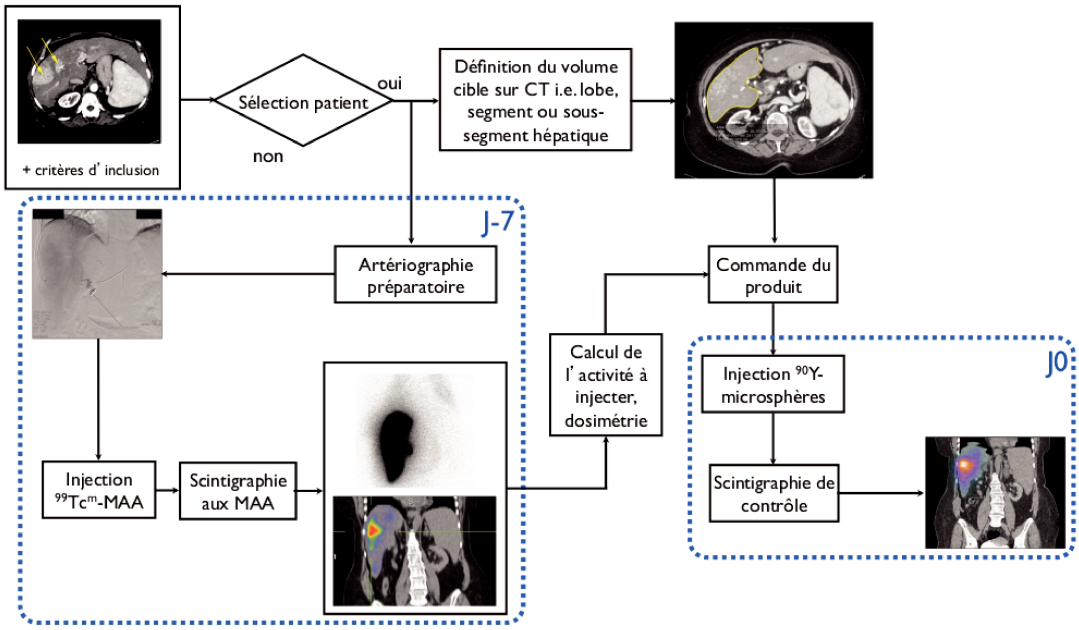


FIGURE 1 – Schéma expliquant la façon dont un patient est pris en charge dans le cadre d'une radioembolisation. D'après Arnaud Dieudonné

Contributions et organisation du manuscrit

Afin de pouvoir comparer entre elles toutes ces méthodes à partir d'études dosimétriques, nous commençons par proposer un post-traitement de l'image rendant positives les images contenant des valeurs négatives en tâchant le plus possible d'en conserver les valeurs moyennes locales (chapitre 2). Nous étudierons ensuite en détail les performances des algorithmes sans contrainte de positivité et des algorithmes pénalisés dans une étude mêlant dosimétrie, analyse de métriques sur des fantômes acquis et visualisation d'images de patients. Cette analyse (chapitre 3) étudie en les combinant l'influence de quatre hyper-paramètres : la borne inférieure des valeurs autorisées dans l'image, l'intensité de la pénalité, le nombre d'itérations des algorithmes avant de les arrêter et l'inclusion ou non du post-traitement introduit dans le chapitre 1. Enfin, nous posons dans le chapitre 4 les prémices d'une méthode destinée à choisir de manière systématique les différents hyper-paramètres.

Production scientifique de la thèse

Articles de journaux

1. M. Millardet, S. Moussaoui, D. Mateus, J. Idier and T. Carlier, « Local-Mean Preserving Post-Processing Step for Non-Negativity Enforcement in PET Imaging : Application to ^{90}Y-PET, » in IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3725-3736, Nov. 2020, doi : 10.1109/TMI.2020.3003428. **(lien)**
2. M. Millardet, S. Moussaoui, J. Idier, D. Mateus, M. Conti, C. Bailly, S. Stute and T. Carlier, « A Multi-Objective Comparative Analysis of Reconstruction Algorithms in the Context of Low-Statistics ^{90}Y-PET Imaging, » in IEEE Transactions on Radiation and Plasma Medical Sciences,
   doi : 10.1109/TRPMS.2021.3126951. **(lien)**

Communications en conférences

1. M. Millardet, S. Moussaoui, D. Mateus, J. Idier and T. Carlier, « Local-mean preserving post-processing for non negativity enforcement in ^{90}Y-PET imaging, » 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), 2018, pp. 1-3,
   doi : 10.1109/NSSMIC.2018.8824399. **(lien)**
2. M. Millardet, S. Moussaoui, D. Mateus, J. Idier, M. Conti and T. Carlier, « A comparative study of AML, NEGML and OSEM based on experimental and clinical ^{90}Y-PET data using the CASToR platform, » 2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), 2018, pp. 1-3,
   doi : 10.1109/NSSMIC.2018.8824361. **(lien)**
3. H. Carrillo, M. Millardet, T. Carlier, D. Mateus, « Low-count PET image reconstruction with Bayesian inference over a Deep Prior, » Proc. SPIE 11596, Medical Imaging 2021 : Image Processing, 115960V (15 February 2021) ; **lien**

Chapitre 1

État de l'art de la reconstruction TEP et problématiques spécifiques liées à l'yttrium 90

Sommaire

1.1 La Tomographie par émission de positons 29

1.1.1 Introduction 29

1.1.2 Principe physique 29

1.1.3 Caméra TEP 30

1.1.4 Stockage des données 31

1.1.5 Corrections 32

1.1.6 Modélisation du problème 33

1.1.7 Algorithmes de reconstruction 38

1.1.7.1 Algorithmes analytiques 38

1.1.7.2 Algorithmes itératifs 38

1.1.7.3 Pénalisation 42

1.1.7.4 Algorithmes utilisant de l'apprentissage profond 45

1.2 L'yttrium 90 et l'émission de positons 46

1.2.1 Introduction 46

1.2.2 Le premier état excité 0^+ du zirconium 90 47

1.2.2.1 Présentation du modèle en couches 47

1.2.2.2 Explication des niveaux observés du zirconium 90 49

1.2.3 Décomposition de l'yttrium 90 51

1.2.4 Problématiques spécifiques liées à la TEP à l'yttrium 90 . 53

1.1 La Tomographie par émission de positons

1.1.1 Introduction

La tomographie par émission de positons est une méthode d'imagerie médicale utilisée en médecine nucléaire. À l'instar des autres méthodes utilisées en médecine nucléaire, celle-ci fonctionne généralement à partir de l'injection d'un couple vecteur-radionucléide. Le vecteur se fixera dans les zones d'intérêt, et la désintégration du radionucléide permettra de le localiser. On choisit ce couple vecteur-radionucléide en fonction de la finalité de l'imagerie (pathologie à détecter, processus moléculaire à tracer, etc.). Les tumeurs étant avides de glucose, le couple le plus couramment utilisé pour la détection des cancers est le fluorodéoxyglucose ^{18}F-FDG, un sucre proche du glucose dont on a remplacé le groupe hydroxyle OH par du fluor 18, dont la désintégration émet des positons.

Ce chapitre a pour objectif de fournir un état de l'art succinct sur la technique d'imagerie TEP et les algorithmes de reconstruction d'images qui y sont employés. Le plan de ce chapitre est le suivant : la section 1.1.2 introduit le principe physique de la TEP, de la désintégration du radionucléide jusqu'à la création d'un signal détectable. La section 1.1.3 a pour sujet l'acquisition de ce signal. Les section 1.1.4 et 1.1.5 concernent respectivement le stockage de ces données et les corrections à y apporter. La section 1.1.6 explique comment se modélise le problème de la reconstruction d'images TEP. Enfin, la section 1.1.7 présente les divers algorithmes de reconstruction.

1.1.2 Principe physique

Les isotopes utilisés en TEP sont des émetteurs β^+, c'est-à-dire que leur désintégration inclut la transformation d'un proton en neutron accompagnée de l'émission d'un positon :

$$p \rightarrow n + e^+ + \nu_e \tag{1.1}$$

Après un parcours dans les tissus environnants de l'ordre de quelques millimètres, les positons émis rencontrent rapidement un électron du milieu ambiant, et cette paire s'annihile pour former une paire de photons :

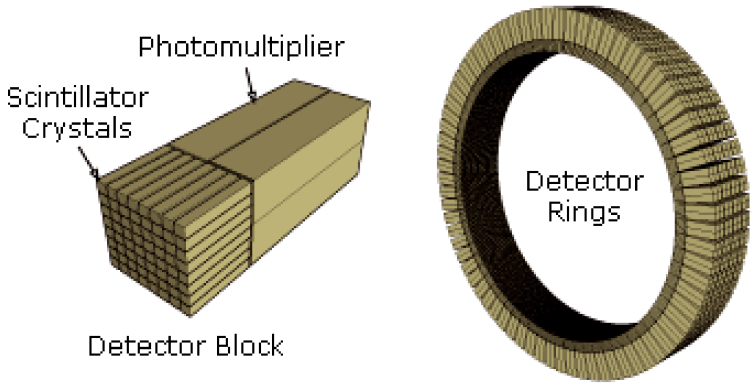


FIGURE 1.1 – schéma d'une caméra TEP et d'un tube photomultiplicateur

$$e^+ + e^- \rightarrow 2\gamma \tag{1.2}$$

Dans le référentiel lié au centre de gravité du système électron-positon, la quantité de mouvement totale est nulle et les deux photons sont émis dans des directions opposées avec une énergie égale. L'énergie cinétique du positon étant négligeable devant sa masse, l'énergie de chaque photon correspond à la masse de l'électron : 511 keV.

Dans le référentiel d'observation, les directions des deux photons ne sont pas tout à fait opposées, mais l'angle n'est généralement que de quelques degrés.

1.1.3 Caméra TEP

Comme l'indique la figure 1.1, une caméra TEP comporte une série d'anneaux de détecteurs élémentaires entourant le patient, optimisés pour la détection de photons de 511 keV. Un événement est enregistré lorsque deux photons sont détectés en coïncidence, c'est-à-dire quand la durée séparant leurs détections est inférieure à un seuil valant généralement quelques nanosecondes. On suppose alors que ces deux photons sont issus d'une même annihilation électron-positon ayant eu lieu sur la ligne, appelée ligne de réponse, reliant les deux détecteurs concernés.

Un détecteur est composé de deux éléments : un scintillateur et un tube photomultiplicateur.

Le **scintillateur** a pour fonction de convertir les photons de haute énergie reçus en photons du domaine visible — entre 1,6 eV et 3,3 eV.

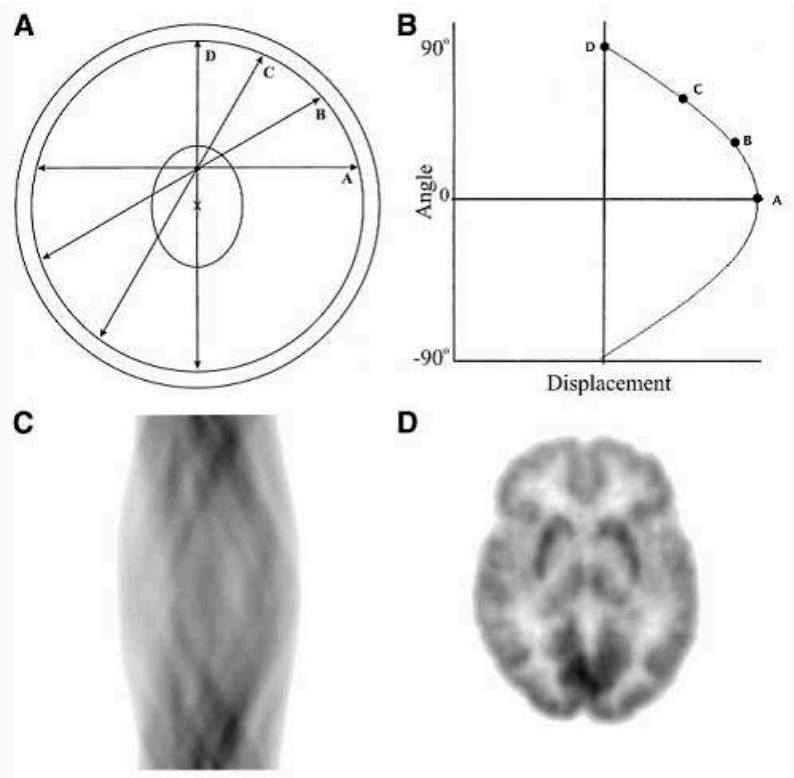


FIGURE 1.2 – Formation d'un sinogramme et reconstruction d'une image de cerveau

Le **tube photomultiplicateur** convertit ensuite les photons lumineux en courant électrique.

1.1.4 Stockage des données

Les coïncidences sont stockées dans un graphique, appelé *sinogramme*, qui, pour une paire d'anneaux donnée, représente en abscisse la distance entre la ligne de réponse et le centre du scanner, et en ordonnée l'angle que forme la ligne de réponse avec une direction donnée.

Que le lecteur se représente dans un premier temps une source d'émission ponctuelle (figure 1.2.A). Diverses lignes de réponse peuvent émerger d'une annihilation en cet unique point, selon la direction que prennent les deux photons, mais l'ensemble de ces lignes correspond à une demi-sinusoïde sur le sinogramme ((figure 1.2.B)), ce qui justifie son nom. La localisation de la source ponctuelle est possible grâce aux caractéristiques de cette sinusoïde. Le sinogramme d'un objet plus complexe est donc une superposition de sinusoïdes et la détermination de l'objet à partir de

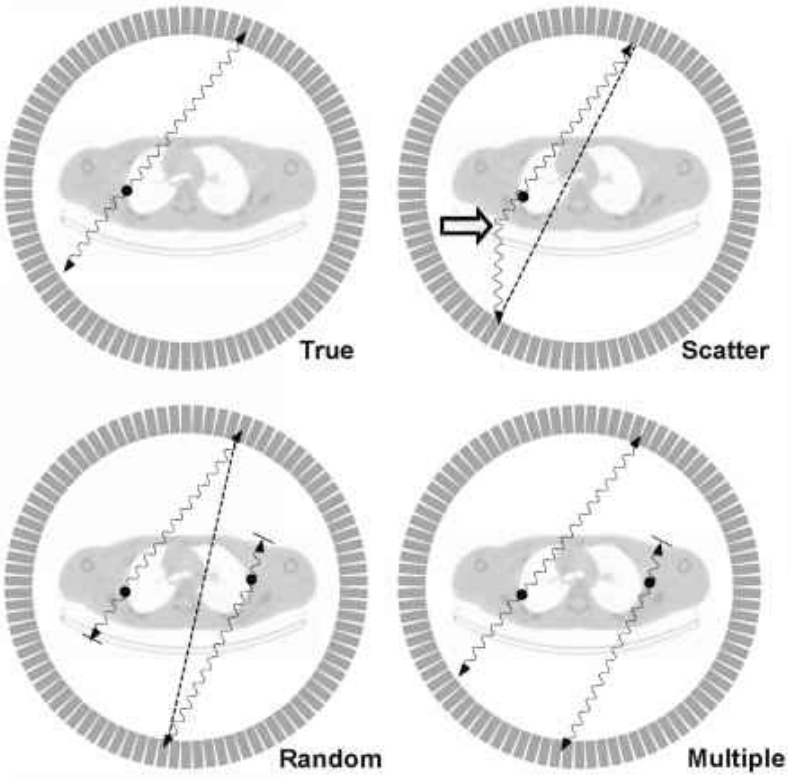


FIGURE 1.3 – Description des quatre types de coïncidences. Le point noir représente l'annihilation du positon, les lignes courbes le parcours réel des photons et la ligne pointillée la ligne de réponse enregistrée, mais erronée. Image prise de [Bailey et al., 2005]

son sinogramme est l'objet de la *reconstruction* de l'image. Les figures 1.2.C et 1.2.D représentent respectivement le sinogramme d'une coupe de cerveau et sa reconstruction.

1.1.5 Corrections

Malheureusement, le nombre de coïncidences détectées par un couple de détecteurs n'est pas proportionnel à l'activité dans la ligne de réponse correspondante et plusieurs corrections doivent être apportées avant la reconstruction.

Un premier problème est que parmi les coïncidences enregistrées, seule une partie doit être prise en compte. On peut en fait grouper les différents types de coïncidences en quatre catégories, ainsi que le montre la figure 1.3.

Un deuxième problème vient du fait que lorsqu'une paire de photons est émise,

une certaine probabilité existe que l'un au moins des deux photons interagisse avec la matière et n'atteigne pas sa cible. Cette probabilité dépend du milieu traversé et est spécifique à une ligne de réponse donnée. Elle est donc modélisée par un sinogramme d'atténuation.

Afin de réduire la quantité de données à stocker, les lignes de réponse les plus proches sont souvent groupées, aussi bien selon l'angle azimutal (mashing) que selon l'angle co-polaire (span). Mais la géométrie du scanner fait que les différents groupes ne contiennent pas le même nombre de paires de détecteurs. Le sinogramme de normalisation rend compte de ces différences, ainsi que de des variations de l'efficacité des cristaux.

Toutes ces corrections (estimation des diffusées, des coïncidences fortuites, de l'atténuation et de la normalisation) sont estimées et stockées dans des sinogrammes avant la reconstruction de l'image.

PSF La désintégration radioactive à l'origine de la détection de deux photons en coïncidence peut ne pas avoir eu exactement lieu sur la ligne joignant les deux cristaux récepteurs. Ceci peut être dû à divers phénomènes, comme le parcours du positon émis avant son annihilation, ou le fait que les directions des deux photons émis ne soient pas colinéaires. Si l'on ne prend pas en compte ce phénomène, on observe sur l'image une tache quand l'objet imagé est une source ponctuelle de radioactivité. Ce floutage de l'image est appelé la *fonction d'étalement du point* (en anglais PSF, pour *point spread function*).

Temps de vol Les scanners récents sont suffisamment précis pour pouvoir estimer la position approximative de l'annihilation à partir du temps écoulé entre la détection des deux photons reçus en coïncidence. Leur résolution temporelle va de 200 ps à 700 ps, ce qui correspond à une précision spatiale allant de 3 cm à 10,5 cm.

1.1.6 Modélisation du problème

La première étape dans la modélisation du problème de la reconstruction d'images TEP est de correctement représenter l'objet imagé que l'on cherche à reconstruire. Bien que diverses méthodes aient été proposées, la plus naturelle (et la plus utilisée) est de diviser l'espace en voxels, et d'associer à chacun de ces voxels une activité.

On représente une telle image par un vecteur $\boldsymbol{x}$ dont chaque coordonnée représente l'activité d'un voxel.

Étant donné une image $\boldsymbol{x}$, l'étape suivante consiste à modéliser combien de coïncidences le scanner TEP détecte par élément du sinogramme. On note $\boldsymbol{m}$ ce vecteur des coïncidences reçues. En TEP, on modélise assez bien le vecteur $\boldsymbol{m}$ par une loi de Poisson de moyenne $\bar{\boldsymbol{m}}(\boldsymbol{x}) = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{\rho}$, où $\boldsymbol{S}$ est la « matrice système », dont l'élément m_{ij} est proportionnel à la probabilité qu'une désintégration ayant lieu dans le voxel j soit détectée par la ligne de réponse i. Cette matrice prend en compte l'atténuation et la normalisation dont nous avons déjà parlé (section 1.1.5), l'éventuelle correction de PSF, l'information sur le temps de vol, ainsi qu'un terme purement géométrique. $\boldsymbol{\rho}$ contient la moyenne des coïncidences supplémentaires (fortuites et diffusées) reçues par la ligne de réponse. On suppose généralement que $\boldsymbol{S}$ et $\boldsymbol{\rho}$ sont parfaitement connus.

Le problème de la reconstruction d'images TEP consiste à estimer l'image $\boldsymbol{x}$ à partir des données reçues $\boldsymbol{m}$. Cependant, le nombre de lignes de réponse est généralement très supérieur au nombre de voxels, ce qui fait que l'équation $\boldsymbol{m} = \boldsymbol{S}\boldsymbol{x} + \boldsymbol{\rho}$ est généralement sans solution. Plusieurs stratégies sont alors possibles pour choisir le « meilleur » $\boldsymbol{x}$, et nous décrivons ici les principales. Il est à noter que nous ne traitons dans cette section que des objectifs à atteindre et de leurs propriétés. Les algorithmes permettant d'atteindre ces objectifs seront traités dans la section suivante. Notons également que la présentation qui suit ne tient pas compte de l'ordre dans lequel les différentes idées sont historiquement apparues.

Maximum de vraisemblance Le maximum de vraisemblance est un estimateur couramment utilisé en statistiques pour estimer les paramètres d'une distribution de probabilité. Il vise à trouver l'image $\boldsymbol{x}$ maximisant la probabilité de $\boldsymbol{m}$ sachant $\boldsymbol{x}$: $Pr(\mu = \boldsymbol{m} | \Xi = \boldsymbol{x})$, où μ et Ξ sont des variables aléatoires dont les réalisations sont les données et les images respectivement.

La loi de Poisson nous donne :

$$Pr(\mu = \boldsymbol{m} | \Xi = \boldsymbol{x}) = \prod_{i=1}^{I} \frac{\bar{m}_i(\boldsymbol{x})^{m_i} \exp(-\bar{m}_i(\boldsymbol{x}))}{m_i!} \tag{1.3}$$

Pour les besoins de l'optimisation, on n'optimise en principe pas directement cette probabilité, mais quelques modifications, menant à un problème équivalent,

lui sont d'abord appliquées :

— Le terme $m_i!$ étant indépendant de l'image à reconstruire, est ignoré.

— On optimise plutôt le logarithme de la probabilité que la probabilité elle-même.

Après ces modifications, la fonction à maximiser, appelée log-vraisemblance, est la suivante :

$$L(\boldsymbol{x}) = \sum_{i=1}^{I} (m_i \log(\bar{m}_i(\boldsymbol{x})) - \bar{m}_i(\boldsymbol{x})) \tag{1.4}$$

Dans ce formalisme, l'image projetée $\bar{\boldsymbol{m}}(\boldsymbol{x})$ doit être positive pour que la vraisemblance soit définie. Cependant, dans le cas général, l'image elle-même peut prendre des valeurs négatives, ce qui pose un problème de signification physique.

Maximum de vraisemblance sous contrainte de positivité Pour corriger ce problème, il est possible de maximiser la vraisemblance dans l'ensemble restreint des images positive. En plus de respecter une contrainte physique du problème, ceci a pour effet de réduire la variance et l'erreur quadratique moyenne de l'image estimée, mais au prix de l'introduction d'un biais positif. Ce biais positif est inhérent à la positivité de l'image. En effet, si l'image réelle contient un voxel nul, l'estimation de ce voxel, n'étant pas parfaite, aura une probabilité non nulle d'être positive, ce qui donnera une moyenne positive à l'estimateur. Ce biais positif, qu'illustre la figure 1.4, est d'autant plus important dans les zones de faible activité (appelées **zones froides**), particulièrement quand les données sont bruitées et peu nombreuses. Il est compensé par un biais négatif dans les zones d'activité élevée (les **zones chaudes**) proches des zones froides.

Maximum a posteriori (MAP) Le maximum de vraisemblance peut mener à des images extrêmement bruitées, surtout dans le domaine des basses statistiques. Cela vient notamment de ce qu'il cherche à maximiser $Pr(\mu = \boldsymbol{m}|\Xi = \boldsymbol{x})$, mais que le problème de la reconstruction tomographique consiste à trouver $\boldsymbol{x}$ sachant $\boldsymbol{m}$. La recherche du maximum a posteriori pose quant à lui le problème de la maximisation de la densité de probabilité de Ξ sachant $\mu = \boldsymbol{m}$: $g_{\Xi|\mu=\boldsymbol{m}}$. Le théorème de Bayes relie entre elles ces deux probabilités de la manière suivante :

$$g_{\Xi|\mu=\boldsymbol{m}}(\boldsymbol{x}) = \frac{Pr(\mu = \boldsymbol{m}|\Xi = \boldsymbol{x}) g_{\Xi}(\boldsymbol{x})}{Pr(\mu = \boldsymbol{m})} \tag{1.5}$$

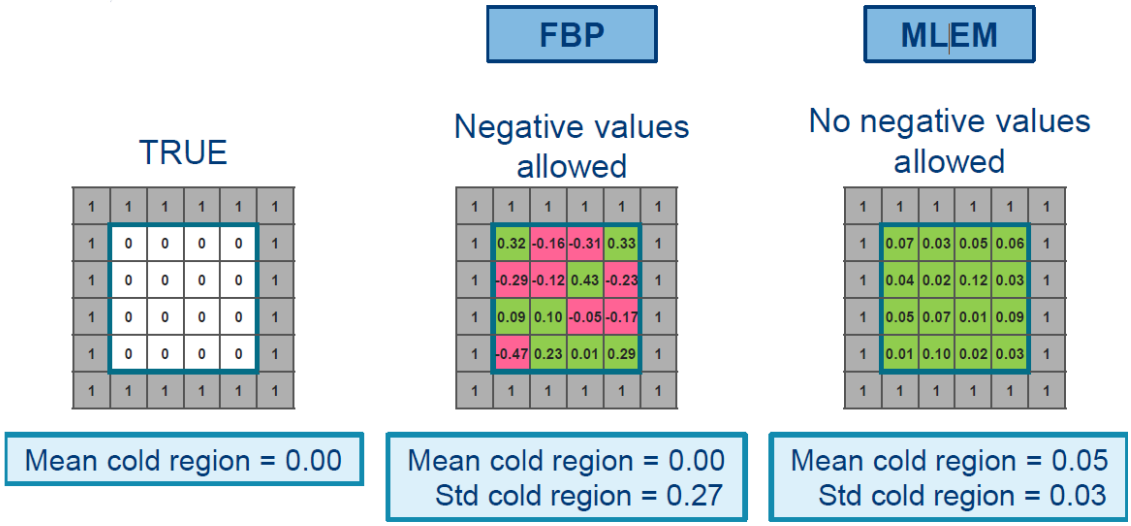


FIGURE 1.4 – Illustration du biais positif introduit par le maximum de vraisemblance sous contrainte de positivité (MLEM). FBP est un algorithme analytique présenté dans la section 1.1.7.1 Image prise d'une présentation faisant suite à [Van Slambrouck et al., 2015]

.

où $g_{\Xi}(\boldsymbol{x})$ et $Pr(\mu = \boldsymbol{m})$ représentent un a priori sur l'image $\boldsymbol{x}$ et sur les données $\boldsymbol{m}$ respectivement. En passant au logarithme, et en enlevant les termes ne dépendant pas de l'image, la fonction à maximiser devient la suivante :

$$F(\boldsymbol{x}) = L(\boldsymbol{x}) - \beta R(\boldsymbol{x}) \tag{1.6}$$

où $\beta R = -\log(g_{\Xi})$ peut aussi être vu comme une simple pénalité appliquée à la log-vraisemblance. Il n'est pas aisé de trouver un a priori correspondant réellement à la distribution réelle des images. Aussi un choix doit-il être fait, tant sur le paramètre β, donnant l'intensité de la pénalité, que sur la fonction R donnant sa forme. R doit dans l'idéal privilégier les images lisses, tout en autorisant des frontières nettes entre différentes régions de l'image. R est le plus souvent choisie convexe pour faciliter l'optimisation. Ils sont généralement de la forme suivante :

$$R(\boldsymbol{x}) = \sum_{j} \sum_{j' \in \mathcal{N}_j} \xi_{jj'} \omega_{jj'} \psi(x_j, x_{j'}) \tag{1.7}$$

où

— $\mathcal{N}_j$ est le voisinage de j.

— $\xi_{jj'}$ est le facteur de proximité, par exemple l'inverse de la distance euclidienne séparant les deux voxels.

— $\omega_{jj'}$ est l'éventuel facteur de similarité, apportant de l'information issue d'une autre modalité, par exemple une image issue d'un IRM ou d'une tomodensitométrie. Nous n'utiliserons pas ce facteur durant cette thèse.

— $\psi(x_j, x_{j'})$ est la fonction de potentiel, en principe convexe, pénalisant les valeurs trop éloignées.

La fonction de potentiel quadratique est la plus utilisée, grâce à sa forme simple : $\psi(x_j, x_{j'}) = (x_j - x_{j'})^2$

Nuyts et al. ont introduit en 2002 [Nuyts et al., 2002a] la fonction de potentiel sur les différences relatives : $\psi(x_j, x_{j'}) = \frac{(x_j - x_{j'})^2}{x_j + x_{j'} + \gamma \left| x_j - x_{j'} \right|}$, où le paramètre γ permet d'autoriser de plus ou moins grandes différences d'intensité. C'est actuellement la seule fonction de potentiel utilisée en routine clinique.

Le maximum a posteriori permet de réduire la variance de l'image à reconstruire, au prix de l'introduction d'un biais dans certaines régions de l'image, particulièrement dans les régions de petite taille, qui s'en trouvent souvent moins visibles.

Estimateur des moindres carrés Plutôt que de maximiser la vraisemblance de Poisson, on a proposé [Nuyts et al., 2002b, Van Slambrouck et al., 2015] de choisir l'image solution du problème des moindres carrés. Une telle image peut contenir des valeurs négatives dans l'espace image, mais également dans sa projection dans l'espace des données $\bar{\boldsymbol{m}}(\boldsymbol{x})$. L'avantage de cet estimateur est qu'il est connu pour être non biaisé à la seule condition que les « erreurs » — c'est-à-dire ici $\boldsymbol{m} - \bar{\boldsymbol{m}}(\boldsymbol{x}_{\mathbf{vraie}})$, où $\boldsymbol{x}_{\mathbf{vraie}}$ représente l'objet réel imagé — soient de moyenne nulle, ce qui est le cas en reconstruction TEP car elles sont issues d'un bruit de Poisson.

1.1.7 Algorithmes de reconstruction

On peut regrouper les algorithmes de reconstruction TEP en trois grandes catégories, que nous traiterons successivement : les algorithmes analytiques (rétroprojection filtrée), les algorithmes itératifs, et les algorithmes utilisant de l'apprentissage profond.

1.1.7.1 Algorithmes analytiques

Les algorithmes analytiques furent historiquement les premiers à apparaître. Ils se caractérisent par leur simplicité et leur faible temps de calcul. Ces algorithmes fonctionnent à l'aide des propriétés théoriques de la transformée de Radon. Celle-ci transforme une image bidimensionnelle en sinogramme en effectuant simplement l'intégrale curviligne de l'image le long de chaque ligne de réponse. Si on lui applique ensuite les corrections d'atténuation, de normalisation, et qu'on ajoute les diffusés et les fortuits, la transformée de Radon permet donc de modéliser simplement l'émission et la formation du sinogramme. La *rétroprojection filtrée* (en anglais filtered back projection, FBP) est simplement une inversion de la transformée de Radon permettant de retrouver l'image initiale à partir d'un sinogramme, et des diverses corrections que nous avons déjà évoquées plus haut, et qu'on calcule séparément.

Malheureusement, ces images présentent régulièrement des stries près des zones de forte activité de l'image.

1.1.7.2 Algorithmes itératifs

Pour pallier les manquements de la rétroprojection filtrée, il est possible d'utiliser un modèle plus sophistiqué pour l'émission des données et de prendre en compte

la nature statistique du bruit en utilisant explicitement le modèle décrit à la section 1.1.6. Le problème ainsi formulé ne peut cependant plus être résolu directement et une méthode itérative devient nécessaire. Le choix de l'algorithme de reconstruction commence souvent par le choix de la fonction objectif à minimiser, mais ne se limite pas à celle-ci. En effet, il convient de tenir compte également de nombreux autres paramètres comme de sa vitesse de convergence, ou de sa stabilité numérique. Ce sont de ces autres paramètres dont il est particulièrement question dans cette section.

Les méthodes itératives sont aujourd'hui les plus abouties et sont celles actuellement utilisées dans la pratique clinique. Ce sont aussi celles dont nous traiterons le plus dans cette thèse. Aussi nous détaillons dans cette section les principales d'entre elles.

Les méthodes ART Les méthodes ART, de l'anglais *algebraic reconstruction techniques*, sont les méthodes itératives les plus anciennes (1970) [Gordon et al., 1970]. Ces méthodes permettent la résolution de systèmes linéaires creux de grande dimension. Cependant, elles ne convergent pas vers une unique solution, mais entrent dans un « cycle limite ». De plus, elles ne tiennent pas compte la nature particulière du bruit dans les données TEP, qui se modélise bien par une loi de Poisson.

L'algorithme MLEM L'algorithme faisant figure de référence parmi les méthodes de reconstruction itératives est l'algorithme MLEM (*Maximum Likelihood Expectation Maximization*), proposé par [Shepp and Vardi, 1982] et indépendamment par [Lange and Carson, 1984]. C'est en fait l'application de l'algorithme EM [Dempster et al., 1977] (*Expectation Maximization*) à la reconstruction TEP. L'algorithme EM est une méthode générale permettant de maximiser la vraisemblance d'une loi de probabilité quelconque en faisant usage de données « complètes » non observées. Appliqué à la TEP, les deux étapes de l'algorithme EM se groupent et forment l'équation simple de mise à jour suivante :

$$x_j^{k+1} = x_j^k \times \frac{\sum\limits_i S_{ij} \frac{m_i}{\bar{m}_i(\boldsymbol{x}^k)}}{\sum\limits_i S_{ij}} \tag{1.8}$$

Il est démontré que MLEM converge vers l'image positive maximisant la vraisemblance (voir la section 1.1.6), ce qui en fait un algorithme particulièrement attrayant.

MLEM garantit la positivité de l'image non seulement à convergence mais aussi à chaque itération de l'algorithme. MLEM possède néanmoins quatre défauts majeurs qui ont justifié la recherche d'algorithmes alternatifs. Nous présenterons dans la section suivante les solutions que la littérature propose pour corriger ces problèmes.

1. **Lenteur de la convergence** MLEM nécessite une cinquantaine d'itérations avant d'obtenir une image acceptable et un nombre d'itérations encore bien plus important avant de converger. Chaque itération nécessitant une quinzaine de secondes [Whiteley et al., 2021], les chercheurs du domaine ont cherché des variantes plus rapides à MLEM.

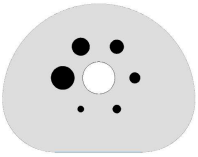
2. **L'effet damier** L'image maximisant la vraisemblance est généralement extrêmement bruitée, même avec la contrainte de positivité, et présente une apparence souvent qualifiée de damier, ou de ciel étoilé. À cause de ce phénomène, MLEM n'est jamais utilisé jusqu'à convergence, mais est arrêté prématurément. En choisissant une image initiale uniforme, on obtient alors un intermédiaire généralement acceptable entre l'image initiale uniforme et le maximum de vraisemblance contraint. Un arrêt entre 15 et 65 itérations constitue généralement le meilleur choix. La figure 1.5 illustre ce phénomène. En plus de l'arrêt prématuré, l'image issue de MLEM est généralement filtrée par une fonction Gaussienne 3D.

3. **Convergence plus lente des zones froides que des zones chaudes** L'équation 1.8 peut se reformuler sous la forme d'un algorithme du gradient dont le pas est proportionnel à l'activité du voxel mis à jour :

$$x_j^{k+1} = x_j^k + \frac{x_j^k}{\sum_i S_{ij}} \frac{\partial L}{\partial x_j}(\boldsymbol{x}^k) \tag{1.9}$$

   Ce pas est plus faible dans les zones de faible intensité et la convergence est donc plus lente dans ces zones que dans les zones chaudes. Lors d'un arrêt prématuré de l'algorithme, ce fait accroît le biais positif dans les zones froides.

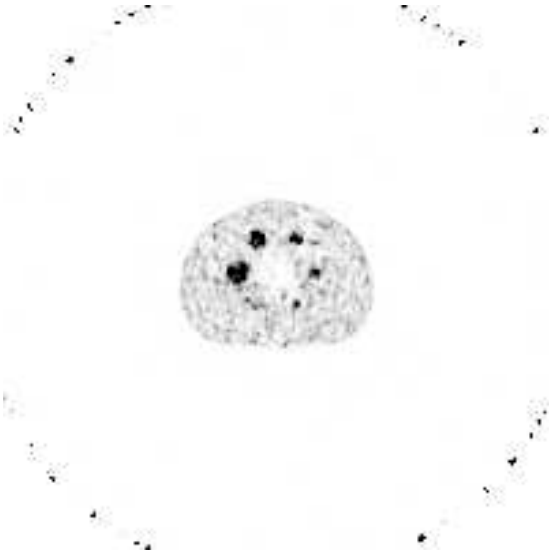
4. **Contrainte de positivité** La contrainte de positivité de l'algorithme est un des avantages de MLEM. Mais comme nous l'avons vu dans la section 1.1.6, elle induit également un biais positif dans l'image. L'utilisation d'algorithmes assouplissant cette contrainte permet de réduire, voire d'annuler, ce biais en utilisant, par exemple, l'une des autres fonctions objectif décrites dans la sec-



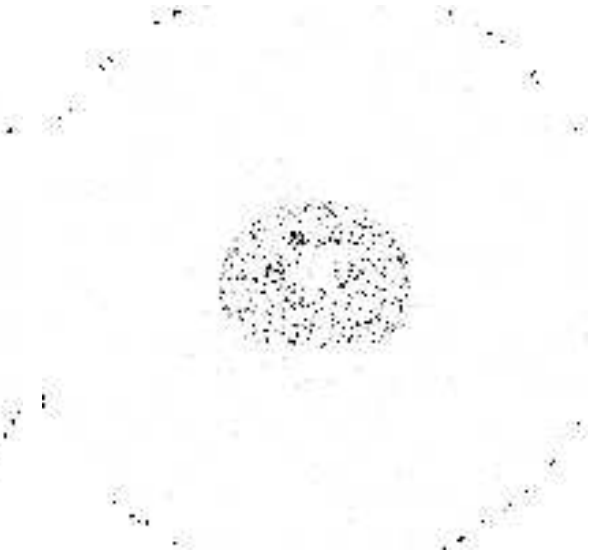
(a) image parfaite

(b) 1 itération

(c) 5 itérations



(d) 40 itérations



(e) 1000 itérations

FIGURE 1.5 – Évolution de la reconstruction d'un fantôme par MLEM au cours des itérations. 1.5a représente l'image réelle. L'image obtenue après 40 itérations est visuellement bien plus séduisante que l'image obtenue à convergence, mais la zone froide entre les sphères chaudes est loin d'être complètement blanche, comme elle devrait l'être. Toutes les images sont à la même échelle de couleur. Sauf la première itération, qui serait quasiment invisible si la même échelle que pour les autres images était appliquée.

tion 1.1.6.

Améliorations de MLEM

OSEM Afin d'accélérer MLEM, deux études ont proposé l'algorithme OSEM (de l'anglais *Ordered Subsets Expectation Maximization*) [Hudson et al., 1992, Hudson and Larkin, 1994]. Cet algorithme divise l'espace des données en S sous-ensembles. Une itération d'OSEM se compose d'm sous-itérations, chacune d'elle appliquant l'équation de MLEM en n'utilisant qu'un seul sous-ensemble, ce qui mène à l'équation suivante :

$$\begin{cases} x_j^{k,s+1} = x_j^{k,s} \dfrac{\sum\limits_{i \in E_s} S_{ij} \frac{m_i}{\bar{m}_i(\boldsymbol{x}^{k,s})}}{\sum\limits_{i \in E_s} S_{ij}} & \text{pour } s \in [\![0, S-1]\!]; \\ x_j^{k+1,0} = x_j^{k,S} (= x_j^k \text{ par définition}) \end{cases} \tag{1.10}$$

Une itération d'OSEM demande un temps de calcul à peu près égal à une itération de MLEM, mais a un résultat comparable à S itérations (du moins lors des premières itérations). Cette accélération considérable de MLEM fait d'OSEM l'algorithme le plus utilisé actuellement en routine clinique. Cependant, OSEM ne converge pas, mais finit par entrer dans un « cycle limite ». Ceci ne pose généralement pas de problème puisque MLEM est de toute façon en principe arrêté prématurément. Dans le cas où le maximum de vraisemblance contraint est réellement recherché, il est toujours possible de se mettre à utiliser MLEM au bout d'un certain nombre d'itérations.

L'idée de diviser l'espace des données en sous-ensembles n'est pas limité à MLEM, mais peut également s'appliquer à n'importe quel algorithme de reconstruction itératif. Cependant, l'usage veut que les autres algorithmes ne changent généralement pas de nom lorsqu'ils sont utilisés avec des sous-ensembles.

1.1.7.3 Pénalisation

Afin de lutter contre l'effet damier, il est possible d'opter pour la recherche du maximum a posteriori plutôt que du maximum de vraisemblance, comme décrit dans la section 1.1.6. Comme cela revient à ajouter un terme à la fonction à minimiser afin de pénaliser certaines images, on parle d'algorithmes pénalisés. Divers algorithmes convergeant vers ce maximum ont été proposés, par exemple celui de de

Pierro [Pierro, 1995]. L'algorithme BSREM (pour *block sequential regularized expectation maximization*) [Pierro and Yamagishi, 2001, Ahn and Fessler, 2003] est quant à lui le seul algorithme pénalisé actuellement utilisé en routine clinique, en utilisant la fonction de potentiel sur les différences relatives [Nuyts et al., 2002a] (voir la section 1.1.6). Nous présentons ici un algorithme que nous appellerons PPGML (pour *penalized preconditioned gradient maximum likelihood*), proposé par Johan Nuyts en 2002 [Nuyts et al., 2002a]. La convergence de cet algorithme n'a pour l'instant pas été prouvée, mais l'expérience montre qu'il est extrêmement robuste aux instabilités numériques, tout en donnant des résultats proches de ceux des autres algorithmes. Son équation de mise à jour est, pour la version n'utilisant pas de sous-ensembles, la suivante :

$$x_j^k = x_j^k + x_j^k \frac{\sum\limits_i S_{ij} \frac{m_i - \bar{m}_i(\boldsymbol{x}^k)}{\bar{m}_i(\boldsymbol{x}^k)} - \beta \partial_j R(\boldsymbol{x}^k)}{\sum\limits_i S_{ij} + x_j^k \beta \partial_j^2 R(\boldsymbol{x}^k)} \tag{1.11}$$

∂_j dénote ici la dérivée partielle par rapport à la j^{e} variable.

Algorithmes autorisant les valeurs négatives dans l'image Il est possible d'assouplir la contrainte de positivité de MLEM, ce qui a également pour effet d'accélérer la convergence de l'image dans les zones froides.

AML Byrne a proposé en 1998 un algorithme appelé ABEMML [Byrne, 1998], généralisant MLEM en contraignant l'algorithme entre deux bornes A et B, A pouvant éventuellement être négatif [Erlandsson et al., 2000]. ABEMML converge vers l'image minimisant $KL(\boldsymbol{m} - \boldsymbol{S}A, \bar{\boldsymbol{m}}(\boldsymbol{x}) - \boldsymbol{S}A) + KL(\boldsymbol{S}B - \boldsymbol{m}, \boldsymbol{S}B - \bar{m}(\boldsymbol{x}))$, où KL représente la divergence de Kullback-Leibler [Kullback and Leibler, 1951]. En 2005, Rahmim *et al.* proposèrent une simplification d'ABEMML [Rahmim et al., 2005], qui s'appellera AML par la suite, dans laquelle la borne B tend vers $+\infty$. Si l'on choisit de plus A uniforme, l'équation de mise à jour est la suivante :

$$x_j^{k+1} = A + (x_j^k - A) \times \frac{\sum\limits_i S_{ij} \frac{m_i - AS_{ij}}{\bar{m}_i(\boldsymbol{x}^k) - AS_{ij}}}{\sum\limits_i S_{ij}} \tag{1.12}$$

On retrouve bien l'équation de mise à jour de MLEM quand $A = 0$. En faisant de plus tendre A vers $-\infty$, comme proposé par [Van Slambrouck et al., 2015], l'équation

devient :

$$x_j^{k+1} = x_j^k + \frac{\sum\limits_i S_{ij} \frac{m_i - \bar{m}_i(\boldsymbol{x}^k)}{\sum_{j'} S_{ij'}}}{\sum\limits_i S_{ij}} \tag{1.13}$$

NEG-ML L'algorithme NEG-ML fut initialement proposé par Nuyts *et al.* en 2002 [Nuyts et al., 2002b], dans le but d'améliorer les images ne disposant pas de correction d'atténuation. Son équation de mise à jour était alors la suivante :

$$x_j^{k+1} = x_j^k + \max\left(\frac{x_j^k}{\sum\limits_i S_{ij}}; \frac{1}{\sum\limits_i S_{ij} \frac{\sum\limits_{j'} S_{ij'}}{\max(1; m_i)}}\right) \sum_i S_{ij} \frac{m_i - \bar{m}_i(\boldsymbol{x}^k)}{\max(1; \bar{m}_i(\boldsymbol{x}^k))} \tag{1.14}$$

Le nombre 1 dans les termes $\max(1; m_i)$ et $\max(1; \bar{m}_i(\boldsymbol{x}^k))$ a pour but d'empêcher les divisions par 0. $\frac{x_j^k}{\sum\limits_i S_{ij}}$ est le pas de MLEM, sensé mener à une convergence rapide dans les zones chaudes, quand $\frac{1}{\sum\limits_i S_{ij} \frac{\sum\limits_{j'} S_{ij'}}{\max(1; m_i)}}$ est spécifique à NEG-ML et est plus rapide dans les zones froides. L'idée de Nuyts était de prendre le plus grand des deux afin de mener à une convergence rapide dans toute l'image.

Ce pas variable avait pour conséquence que NEG-ML ne permettait pas toujours de réduire le biais positif significativement, comparé à MLEM [Van Velden et al., 2009, Verhaeghe and Reader, 2010], et pouvait de plus entraîner des problèmes de convergence [Van Slambrouck et al., 2015]. En 2015, van Slambrouck *et al.* [Van Slambrouck et al., 2015] proposèrent une nouvelle version de NEG-ML en ne conservant que le pas spécifique à NEG-ML. Ils introduisirent également un paramètre ψ en deçà duquel les données n'étaient plus modélisées par une loi de Poisson, mais par une loi Gaussienne, ce qui conduisit à la formule suivante :

$$x_j^{k+1} = x_j^k + \frac{\alpha_j \sum\limits_i S_{ij} \frac{m_i - \bar{m}_i(\boldsymbol{x}^k)}{\max(\psi; \bar{m}_i(\boldsymbol{x}^k))}}{\sum\limits_i S_{ij} \frac{\sum\limits_{j'} S_{ij'} \alpha_{j'}}{\max(\psi; \bar{m}_i^k)}} \tag{1.15}$$

dans laquelle l'ensemble des α_j étaient des paramètres additionnels. En choisissant les α_j uniformes et en faisant tendre ψ vers $+\infty$, on obtient l'équation suivante :

$$x_j^{k+1} = x_j^k + \frac{\sum_i S_{ij}(m_i - \bar{m}_i(\boldsymbol{x}^k))}{\sum_i S_{ij} \sum_{j'} S_{ij'}} \tag{1.16}$$

L'algorithme induit par cette dernière équation converge vers la solution des moindres carrés, et est par conséquent non biaisé (voir la section 1.1.6).

ADMM-Reg et HypoC-PML En 2018, Lim *et al.* [Lim et al., 2018] proposèrent un algorithme convergeant vers le maximum de vraisemblance non contraint, éventuellement pénalisé. Cet algorithme utilise la méthode ADMM (pour *alternating direction method of multipliers*) [Boyd et al., 2011], une variante de la méthode des lagrangiens augmentés permettant de résoudre des problèmes généraux d'optimisation sous contrainte [1]. Lim *et al.* ont pour cette raison appelé leur algorithme ADMM-Reg.

Bousse *et al.* présentèrent en 2020 [Bousse et al., 2020] une méthode alternative à [Lim et al., 2018] résolvant le même problème d'optimisation. Cette méthode remplace la fonction à maximiser par une suite de fonctions sans contrainte sur le domaine de faisabilité, hypo-convergeant vers la log-vraisemblance pénalisée. L'hypo-convergence est une notion particulière de convergence de suite de fonctions garantissant sous certaines conditions, réunies ici, la convergence des maxima successifs vers le maximum de la limite. Ce nouvel algorithme, appelé HypoC-PML (pour *hypo-convergence for penalized maximum-likelihood*) converge plus rapidement qu'ADMM-Reg.

1.1.7.4 Algorithmes utilisant de l'apprentissage profond

Suite aux développements récents de l'intelligence artificielle, et en particulier de l'apprentissage profond, une nouvelle classe d'algorithmes de reconstruction est apparue. Cette classe n'est cependant pas encore opérationnelle, et ne peut pas actuellement être utilisée en routine clinique. Ces méthodes sont soit des méthodes directes, permettant de reconstruire les images à partir des données enregistrées par le scanner en apprenant elles-mêmes les théories physiques et statistiques sous-jacentes, soit

1. Il faut noter que bien qu'aucune contrainte ne soit imposée aux valeurs de l'image, une contrainte de positivité est imposée dans l'espace des projections par définition de la loi de Poisson

remplacent une partie d'une méthode « conventionnelle », en servant par exemple de régularisation. Parmi les avantages souvent cités des premières méthodes, on trouve notamment la vitesse très élevée d'exécution d'un réseau de neurone une fois entraîné, qui permettrait d'ouvrir la voie à de nouveaux champs d'application de la TEP, comme la TEP interventionnelle, et la possibilité à terme de surpasser les méthodes itératives. Les deuxièmes méthodes ont pour intérêt de faire apprendre automatiquement à un réseau de neurones des éléments difficiles à modéliser physiquement. Dans tous les cas, dans le cadre d'un apprentissage supervisé à partir de données provenant de patients, ce sont des algorithmes itératifs qui fournissent l'approximation de la vérité terrain. L'amélioration de la qualité des algorithmes itératifs peut donc contribuer en retour à l'amélioration des algorithmes utilisant de l'apprentissage profond. Le lecteur intéressé pourra trouver une revue complète de ces méthodes dans la référence [Reader et al., 2021].

1.2 L'yttrium 90 et l'émission de positons

1.2.1 Introduction

L'yttrium 90 (^{90}Y) émet en se désintégrant des particules β^- d'énergie cinétique moyenne de 933,6(12) keV et maximale de 2280,1(16) keV, avec une pénétration moyenne dans les tissus de 2,5 mm et maximale de 11 mm. Ces grandeurs, ainsi que sa demi-vie (64,053(20) h), font de l'yttrium 90 l'un des radionucléides les plus utilisés pour les applications thérapeutiques en médecine nucléaire.

La désintégration la plus probable de l'yttrium 90 est une désintégration directe vers l'état fondamental du zirconium 90 (^{90}Zr). La valeur Q [2] de la réaction est de 2280,1(16) keV.

En 1955, Ford prédit que le premier état excité du zirconium 90 était un état 0^+ [Ford, 1955]. L'état prédit fut ensuite découvert par Johnson *et al.* [Johnson et al., 1955] qui observèrent également une émission de positons pendant la désintégration de l'yttrium 90, attribuée à la transition $0^+ - 0^+$. Cette émission de positons peut être utilisée pour l'imagerie TEP après un traitement à l'yttrium 90.

2. La valeur Q, aussi appelée chaleur, d'une réaction est la différence de masse entre les composants initiaux et finals. Ici, le composant initial est le noyau initial de l'yttrium 90 et les composants finals sont le noyau du zirconium 90 et l'électron nouvellement créé. La valeur Q correspond à la libération d'énergie.

La section 1.2.2 expliquera pourquoi le premier état excité du zirconium 90 a une parité de spin de 0^+ en utilisant le modèle en couches. La section 1.2.3 présentera ensuite la désintégration de l'yttrium 90. Ces deux sections sont proposées au lecteur intéressé, mais ne sont pas nécessaires à la compréhension de cette thèse. La section 1.2.4 présente quant à elle les problèmes spécifiques liés à l'utilisation de l'yttrium 90 en TEP.

1.2.2 Le premier état excité 0^+ du zirconium 90

1.2.2.1 Présentation du modèle en couches

Les niveaux d'énergie des noyaux atomiques sont bien décrits par le modèle en couches, présenté par la figure 1.6.

Fondé sur le principe d'exclusion de Pauli, ce modèle décrit les différentes énergies potentielles que peuvent prendre les nucléons à l'intérieur d'un noyau. Comme les protons et les neutrons sont des particules différentes, il faut remplir les couches séparément pour les protons et les neutrons. On nomme *couches majeures* l'ensemble des sous-couches séparées par une différence d'énergie potentielle importante. Les couches majeures sont remplies de 2, 8, 20, 28, 50, 82 et 126 nucléons respectivement pour les premières.

De manière analogue à ce qui se fait pour les couches électroniques, le modèle en couches étiquette les couches par leur nombre quantique de moment angulaire orbital l. On appelle une couche s si $l = 0$, p si $l = 1$, puis d, f, g, h, i, etc. Une couche de moment angulaire l a une dégénérescence de $2(2l + 1)$, c'est-à-dire qu'elle peut contenir $2(2l+1)$ nucléons. Le facteur $2l+1$ vient de ce que les nucléons à l'intérieur d'une couche peuvent avoir un moment orbital dans la direction z compris entre $-l\hbar$ et $l\hbar$ par pas de $\hbar$. Le facteur 2 tient du fait que la projection du spin d'un nucléon dans la direction z peut prendre deux valeurs : $-\hbar/2$ et $\hbar/2$. La première couche d'une certaine valeur de l est indexée par 1, la seconde par 2, etc. Par exemple, pour $l = 2$, nous avons les couches $1d$, $2d$, $3d$, etc.[3] [Krane, 1988]. Contrairement à la physique atomique, l'interaction spin-orbite est importante en physique nucléaire. Cela signifie que pour chaque couche, on doit distinguer si le spin et le moment angulaire orbital sont alignés ou non. Si oui, le moment angulaire *total* j prend le

3. Ceci diffère de la notation utilisée en physique atomique, où l'indice désigne le nombre quantique principal, il n'y a donc ni d'état $1d$, ni d'état $2d$.

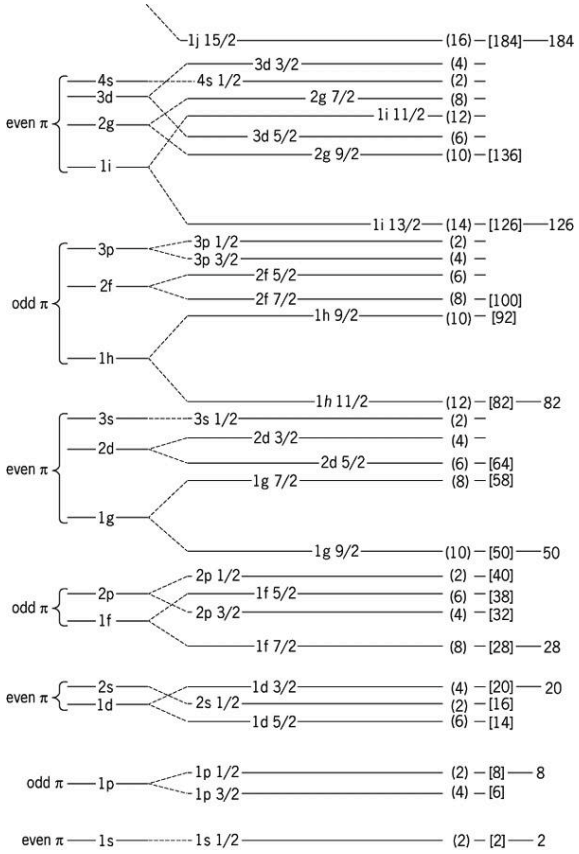


FIGURE 1.6 – Modèle en couches. Figure tirée de [D'Arienzo, 2013]. Les parenthèses rondes (2), (4), etc. indiquent les dégénérescences de niveau et les parenthèses carrées [2], [6], etc. le nombre total d'occupation.

nombre quantique $l + 1/2$, sinon $l - 1/2$. Chaque couche (excepté les couches s où $l = 0$) est alors divisée en deux couches étiquetées par la valeur j (par exemple, $1d$ est divisé en $1d_{3/2}$ et $1d_{5/2}$). Une fois encore, la projection de ce moment angulaire total le long d'une direction spécifique peut prendre une valeur allant de $-j\hbar$ à $j\hbar$ par pas de $\hbar$, de sorte que chaque couche peut contenir $2j + 1$ nucléons.[4] Le niveau d'énergie de ces couches est présenté dans la figure 1.6. La parité d'un nucléon est déterminée par $(-1)^l$, par conséquent les couches s, d, g, etc. ont une parité paire et les couches p, f, h, etc. une parité impaire.

1.2.2.2 Explication des niveaux observés du zirconium 90

$^{90}_{40}\mathrm{Zr}_{50}$ possède 40 protons et 50 neutrons. Les neutrons remplissent donc une couche majeure et il faut beaucoup d'énergie pour les exciter : les premiers états excités sont donc obtenus en excitant des protons. Dans l'état fondamental, les 40 protons remplissent des couches jusqu'à la sous-couche $2\mathrm{p}_{1/2}$.

On peut s'attendre à une augmentation de l'énergie potentielle de l'ordre de 0,9 MeV lorsque l'un des protons de la sous-couche $2\mathrm{p}_{1/2}$ est excité jusqu'à la couche $1\mathrm{g}_{9/2}$. Si les deux étaient excités, cela entraînerait une augmentation de 1,8 MeV. On s'attend à ce que ces états soient les premiers états excités. Considérons-les successivement.

À l'ordre zéro, l'état excité le plus bas devrait être trouvé lorsqu'un proton de la couche $2p_{1/2}$ est excité vers la couche $1\mathrm{g}_{9/2}$. Cela devrait normalement conduire à une augmentation de l'énergie potentielle d'environ 0,9 MeV. Dans cette configuration, seuls deux protons sont non appariés. L'un a un spin 1/2 et une parité impaire, l'autre un spin 9/2 et une parité paire. Comme le moment angulaire total est le produit des deux, la parité résultante doit être impaire. Le spin total doit se situer entre $9/2 - 1/2$ et $9/2 + 1/2$ par pas entiers. Dans ce cas, le noyau est alors soit dans un état à 4^-, soit dans un état à 5^-. Parmi les autres états excités, celui qui présente la plus faible énergie potentielle devrait être une excitation des deux protons de la couche $2\mathrm{p}_{1/2}$ vers la couche $1\mathrm{g}_{9/2}$. Cette dernière possibilité devrait entraîner une augmentation d'environ 1,8 MeV et conduire à des états de 0^+ à 9^+. Cependant, on doit considérer les deux protons $1\mathrm{g}_{9/2}$ comme des particules identiques, et donc comme une fonction d'onde correctement symétrisée. Cela réduit le spin résultant

4. Le nombre total de nucléons potentiels de $1d_{3/2}$ est donc de 4, et celui de $1d_{5/2}$ de 6. Par conséquent, $1d$ peut encore recevoir 10 nucléons.

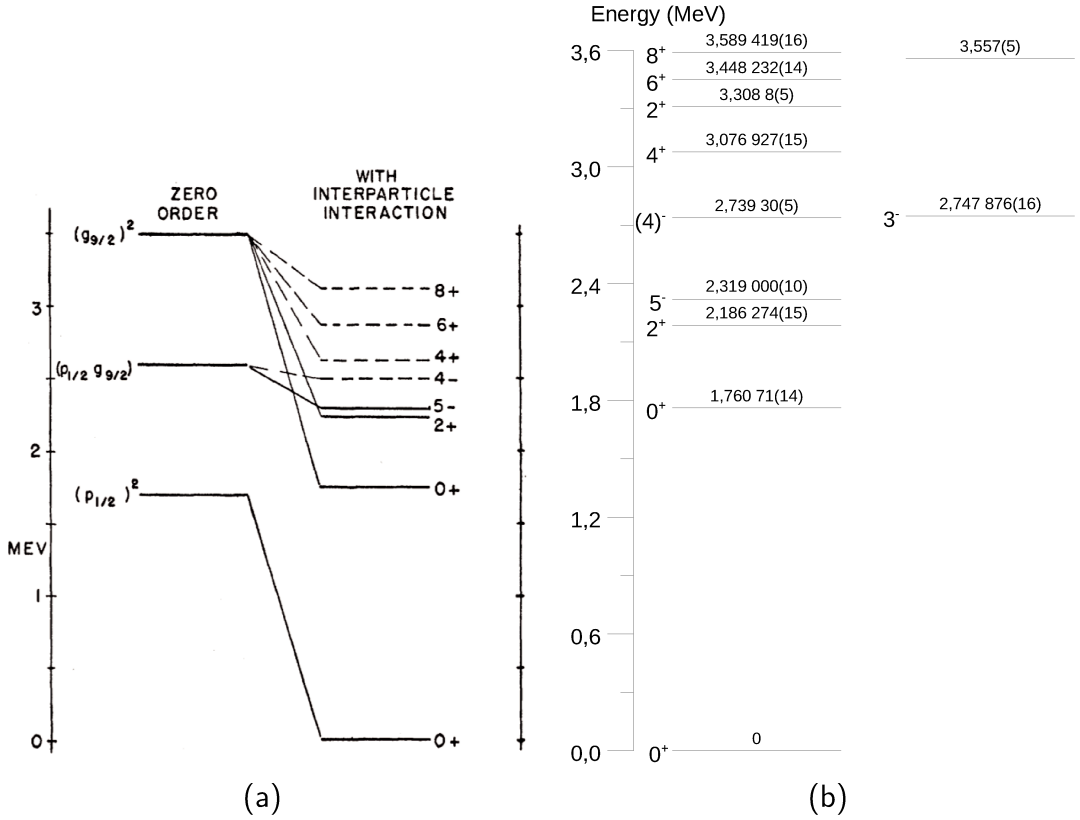


FIGURE 1.7 – [Ford, 1955].1.7b : Premiers états excités du zirconium 90. Données de [nnd,]

à des valeurs paires, et les possibilités sont donc 0^+, 2^+, 4^+ et 8^+ [Krane, 1988]. Ces deux possibilités sont présentées sur le côté gauche de la figure 1.7a. Une autre possibilité est l'excitation d'un proton de la sous-couche $1f_{5/2}$ vers la sous-couche $1g_{9/2}$, conduisant aux états 2^- à 7^-.

Afin d'obtenir un modèle plus réaliste, il faut ajouter la force d'attraction à deux corps entre les deux protons potentiellement non appariés. Dans les configurations 0^+ avec $(p_{1/2})^2$ et $(g_{9/2})^2$, l'interaction mutuelle est très forte et cela se traduit par un décalage d'environ 1,5 MeV. Au contraire, les états dans la configuration ($p_{1/2}$ $g_{9/2}$) impliquent l'interaction de deux particules non équivalentes, ce qui conduit à un décalage de seulement quelques dixièmes de MeV. L'état à 5^- est plus décalé que celui à 4^- car deux protons sont plus proches dans l'espace en alignant les moments angulaires orbitaux [D'Arienzo, 2013]. Pour la même raison, seul un petit décalage est attendu si l'on considère les autres possibilités impliquant des particules

non équivalentes, comme la configuration ($f_{5/2}$ $g_{9/2}$). Tout ceci explique pourquoi le premier état excité observé est un état 0^+.

La figure 1.7b montre les premiers états excités du zirconium 90 à partir des dernières données. Tous les niveaux présents dans l'article de Krane sont présents dans l'ordre correct.

1.2.3 Décomposition de l'yttrium 90

C'est toujours une désintégration β^- qui mène de l'état fondamental de l'yttrium 90 ($^{90}_{39}Y_{51}$) à celui (stable) du zirconium 90 ($^{90}_{40}Zr_{50}$). Cependant, même si le processus le plus probable est une désintégration directe dans l'état fondamental, un processus en deux étapes est possible : dans ce cas, l'yttrium 90 se désintègre d'abord dans un état excité du zirconium 90 par une émission β^-, avant que cet état ne se désexcite. La figure 1.7b montre les premiers états excités du zirconium 90.

La valeur Q totale de la réaction est de 2280,1(16) keV. Par conséquent, seuls les trois niveaux de plus basse énergie du zirconium 90 peuvent être atteints : le niveau fondamental stable 0^+, le premier niveau excité 0^+, et le second niveau excité 2^+. Une transition vers un niveau supérieur du zirconium 90 est interdite car elle impliquerait une création d'énergie. La transition directe vers l'état fondamental du zirconium 90 est de loin la plus probable des trois possibilités (avec un rapport d'embranchement de 99,9885(14) %), mais la transition vers le premier état excité a encore un rapport d'embranchement de 0,0115(14) %. La transition vers le second état excité ne présente quant à lui qu'un rapport d'embranchement de $1{,}4(3) \times 10^{-6}$ %. Ce rapport d'embranchement est totalement négligeable pour le processus de création de paires $e^+ - e^-$ au regard de l'imprécision que nous avons sur les autres rapports d'embranchement. La figure 1.8 montre les différentes possibilités de désintégration.

La transition de l'état excité 0^+ du zirconium 90 vers l'état fondamental implique une transition $0^+ - 0^+$, qui n'est possible que par une transition E0 [5]. L'émission d'un seul photon ne peut réaliser une telle transition, mais il existe trois possibilités :

1. **Conversion interne** : le champ du noyau interagit avec les électrons de l'atome et provoque l'émission de l'un d'entre eux.

5. E signifie électrique et 0 signifie que l'ordre multipolaire du rayonnement est $2^0 = 1$, c'est-à-dire que la transition est un monopôle électrique

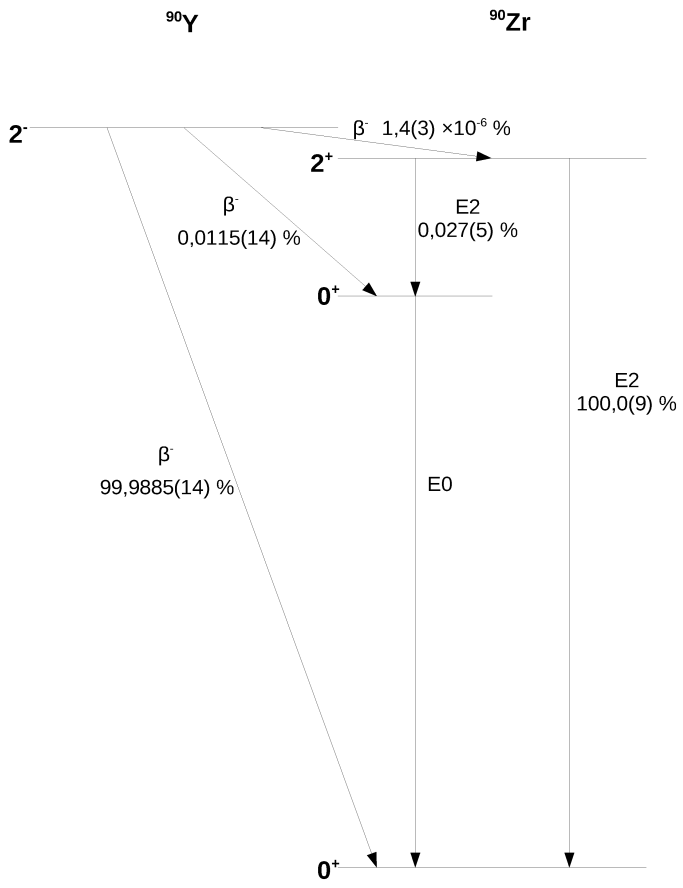


FIGURE 1.8 – Différents modes de décroissance des premiers états fondamentaux de l'yttrium 90. Les niveaux sont placés à l'échelle d'énergie correcte. Données de [nnd,]

.

2. **Création d'une paire** $e^+ - e^-$
3. **Émission de deux photons**, le plus probablement deux photons E1 ou M1.

Thomas [Greenberg and Deutsch, 1940] découvrit en 1940 l'expression générale du rapport entre la probabilité de conversion interne et la probabilité de création de paires dans les transitions $0^+ - 0^+$. Appliquée au zirconium 90, on trouve que la conversion interne est 2,6 fois plus probable que la création de paires. Ce rapport d'embranchement théorique est confirmé par les dernières expériences. La probabilité de désexcitation de l'état excité 0^+ par une émission à deux photons est d'environ 0,2 % [Langhoff and Hennies, 1961]. La probabilité de la désintégration vers l'état excité 0^+, combinée à la probabilité de création de paires explique le rapport total de $3.186(47) \times 10^{-5}$ création de paire par désintégration d'yttrium 90 tel que mesuré en 2006 par Selwyn *et al.* [Selwyn et al., 2007].

1.2.4 Problématiques spécifiques liées à la TEP à l'yttrium 90

En plus du très faible taux de positons émis, deux autres facteurs viennent limiter la qualité des images TEP à l'yttrium 90.

1. Le LSO (de l'anglais *lutetium orthosilicate*) dont est composé les détecteurs de la majorité des scanners contient environ 2,6 % de lutétium 176, qui se désintègre en émettant des photons d'une énergie comparable à ceux que produisent l'annihilation électron/positon. Ce phénomène augmente le nombre de coïncidences fortuites. Il est habituellement négligeable en TEP car le taux de positons émis est suffisamment élevé. Mais en TEP à l'yttrium 90, lorsque le niveau d'activité est faible, le LSO est responsable de la majorité des coïncidences détectées.
2. Les électrons issus de la désintégration de l'yttrium 90 produisent un Bremsstrahlung dont la détection ajoute encore au nombre des coïncidences fortuites. Cependant, le nombre de coïncidences fortuites dû au Bremsstrahlung est très inférieur à celui dû au LSO.

Ces phénomènes rendent particulièrement difficile la reconstruction d'images TEP yttrium 90, et l'étude de Strydhorst *et al.* [Strydhorst et al., 2016] suggère que c'est principalement l'algorithme de reconstruction utilisé qui pourrait améliorer la quantification de telles images.

Chapitre 2

Post-traitement des images imposant la positivité tout en préservant les moyennes locales

Sommaire

2.1 Introduction 57

2.2 Formulation du problème 58

2.2.1 Principe 58

2.2.2 Prérequis 59

2.2.3 Formulation mathématique du problème 59

2.2.4 Propriétés de H 61

2.2.5 Exigence supplémentaire sur alpha 63

2.2.6 Choix de la norme 63

2.2.7 Existence et unicité de la solution 65

2.3 Formulation en termes d'optimisation linéaire et résolution à l'aide d'une version améliorée de la méthode duale du simplexe 65

2.3.1 Formulation en termes d'optimisation linéaire 65

2.3.2 La méthode duale du simplexe 67

2.3.3 Résolution à l'aide d'une version améliorée de la méthode duale du simplexe 68

2.3.4 Détails pratiques de mise en œuvre 70

2.3.4.1 Mise en œuvre 70

2.3.4.2 Amélioration de l'algorithme (étape d'initialisation) 72

2.3.4.3 Choix de la matrice laplacienne H 73
2.4 Illustrations numériques 74
2.4.1 Exemple sur un fantôme 75
2.4.2 Étude d'un laplacien **H** non isotrope 78
2.4.3 Comparaison entre CPLEX et l'algorithme proposé 78
2.5 Résultats expérimentaux 81
2.5.1 Mise en œuvre expérimentale 81
2.5.2 Images 82
2.5.3 Figures de mérite 82
2.6 Discussion et conclusion 88

2.1 Introduction

De nombreux travaux ont été consacrés à la recherche de nouveaux algorithmes autorisant les valeurs négatives dans l'image [Byrne, 1998, Nuyts et al., 2002b, Van Slambrouck et al., 2015, Lim et al., 2018, Bousse et al., 2020]. De telles méthodes permettent de réduire, voire de supprimer le biais positif, mais les voxels négatifs ainsi introduits n'ont aucun sens physique, et, de plus, la variance des valeurs de l'image se trouve augmentée. Le travail présenté dans ce chapitre propose une stratégie de post-traitement qui peut être utilisée en plus de tout algorithme non contraint, ou partiellement contraint, en produisant une image positive.

Une stratégie de post-traitement simpliste serait d'effectuer une simple projection orthogonale de l'image reconstruite sur l'ensemble des images positives, c'est-à-dire de mettre à zéro les valeurs négatives, tout en gardant inchangées les valeurs positives. Cependant, une telle solution présenterait plusieurs inconvénients. Premièrement, l'activité moyenne s'en trouverait augmentée, ce qui conduirait à un biais positif encore plus élevé que ce qui peut être obtenu en utilisant un algorithme de reconstruction sous contrainte de positivité. Deuxièmement, cette solution ne tiendrait pas compte de la signification des valeurs négatives. En effet, les valeurs négatives ne sont pas totalement vides de sens : elles n'ont pas de signification physique au niveau du voxel mais sont importantes pour réduire l'activité moyenne de leur entourage.

Dans ce chapitre, nous proposons une approche de post-traitement non triviale qui transfère toutes les valeurs négatives aux voxels voisins afin de modifier l'image tout en conservant le plus possible les valeurs moyennes locales. Pour atteindre un tel objectif, un cadre classique serait d'utiliser le formalisme du transport optimal introduit par Gaspard Monge en 1781 [Monge, 1781], et revisité par Leonid Kantorovich en 1942 [Kantorovitch, 1942]. L'idée du transport optimal est de transférer une quantité définie de masse dans une quantité définie de trous avec un coût minimal. Il traite classiquement du cas où la quantité de masse et la capacité totale des trous sont égales, et est donc plus axé sur la manière d'arriver à la solution finale que sur la solution finale elle-même. Cependant, il a été étendu au cas où les distributions initiale et finale n'ont pas la même masse [Monge, 1781, Peyré and Cuturi, 2019]. Le principal inconvénient du transport optimal appliqué à la propagation des valeurs négatives des images TEP est que le transfert peut être asymétrique, ce qui conduit

à une propagation fallacieuse de l'information. La figure 2.1a montre un exemple simple de transfert complètement asymétrique engendré par le transport optimal. De plus, le formalisme du transport optimal ne garantit pas l'unicité de l'image finale. L'approche proposée dans ce chapitre consiste à mieux contrôler le mécanisme de transfert en prédéfinissant une fonction d'étalement des voxels symétrique décrivant la proportion de la valeur provenant d'un voxel qui sera distribuée à chacun de ses voisins. La solution est celle correspondant au transfert minimal suivant une telle règle symétrique.

La contribution de ce travail est d'abord de formaliser cette stratégie de post-traitement pour la suppression des voxels négatifs, et ensuite d'implémenter et d'éprouver la méthode sur des données réelles et simulées. Nous formalisons ce problème en termes d'optimisation linéaire [Nocedal and Wright, 1999], et nous proposons une version adaptée de la méthode duale du simplexe pour résoudre le problème d'optimisation. En outre, certaines de ses propriétés mathématiques sont discutées. Enfin, nous avons évalué les performances de l'étape de post-traitement (appelée NNEPPS, pour *non-negativity enforcement post-processing step*) sur des données yttrium 90 acquises. Ce travail a donné lieu à la publication d'un article dans IEEE *Transactions on Medical Imaging* en novembre 2020 [Millardet et al., 2020].

2.2 Formulation du problème

2.2.1 Principe

L'objectif du post-traitement que nous proposons est de rendre une image positive par un transfert minimal de ses valeurs négatives vers les voxels positifs voisins. La valeur de chaque voxel peut augmenter tandis que celle de ses voisins diminue, de telle sorte que la moyenne soit préservée. Ce post-traitement s'accompagne des garanties d'existence et d'unicité de la solution à la seule condition que la moyenne globale de l'image initiale soit positive. Cette dernière condition est nécessaire car le post-traitement préserve la moyenne globale. L'algorithme proposé dans cette étude atteint la solution en un nombre fini d'itérations, généralement autour de 15 pour une image TEP complète de 4×10^6 voxels, ce qui équivaut à un temps de calcul allant d'une à dix minutes sur les huit CPU Intel Core i7-8550U @ 1,80 GHz utilisés

dans cette étude.

2.2.2 Prérequis

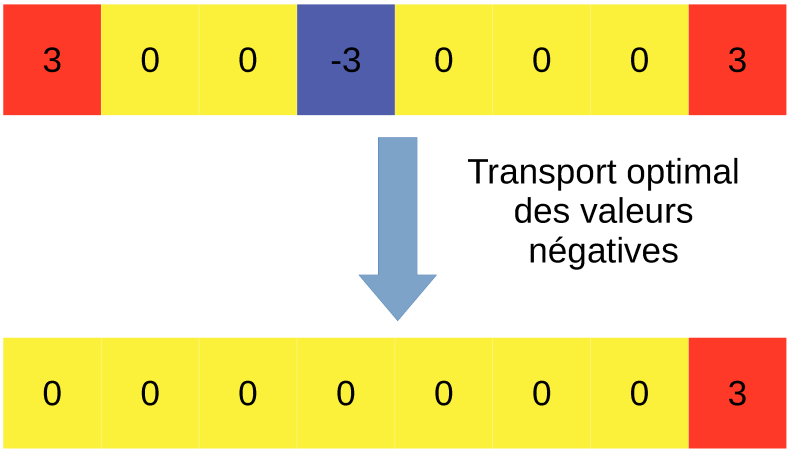
Considérons une image TEP $\boldsymbol{X}$ ayant des valeurs négatives, obtenue à partir d'un algorithme de reconstruction non contraint, et $\boldsymbol{x} \in \mathbb{R}^n$ un vecteur colonne représentant $\boldsymbol{X}$ (par exemple construit en suivant l'ordre lexicographique). Le problème abordé est celui de la suppression des valeurs négatives de $\boldsymbol{X}$ pour obtenir une image modifiée $\boldsymbol{Y}$ correspondant à un vecteur $\boldsymbol{y} \in (\mathbb{R}^+)^n$. Comme nous l'avons vu précédemment, une simple projection orthogonale de $\boldsymbol{x}$ sur $(\mathbb{R}^+)^n$ n'est pas satisfaisante. Dans ce chapitre, nous proposons une approche alternative.

Il est souhaitable que le post-traitement préserve la moyenne globale, c'est-à-dire que la moyenne de l'image post-traitée $\boldsymbol{Y}$ doit être égale à la moyenne de $\boldsymbol{X}$. Un post-traitement naïf pourrait alors être de considérer la projection orthogonale de $\boldsymbol{x}$ sur l'ensemble convexe des vecteurs positifs ayant la même moyenne que $\boldsymbol{x}$. Malheureusement, cette dernière ne respecterait pas le caractère local des transferts d'intensité. Par exemple, si $\boldsymbol{x}$ est une image avec un seul voxel négatif, ce dernier sera mis à 0 tandis que tous les autres voxels seront diminués de la même quantité. La figure 2.1b montre un exemple simple de ce phénomène.

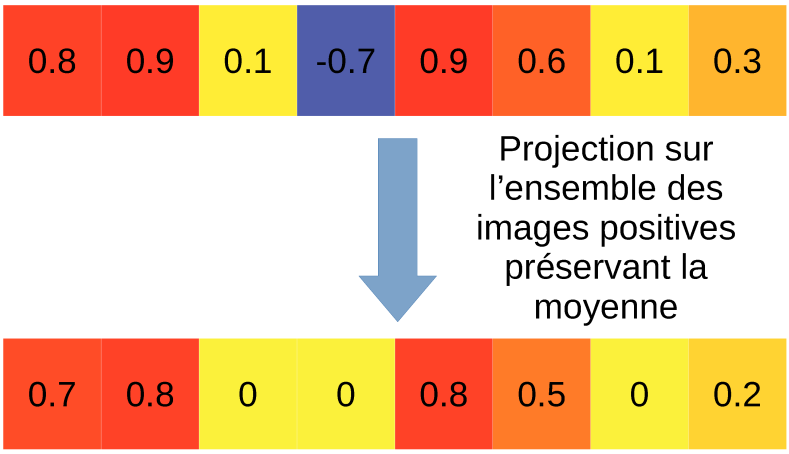
Par conséquent, nous devons introduire une notion de voisinage entre voxels. Une idée naturelle est alors de minimiser la quantité transférée entre voxels voisins, au lieu de la différence entre l'image finale et l'image initiale. On peut alors formuler le problème de la manière suivante : $\min \|\boldsymbol{\alpha}\|$, tel que $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha}$, où $\boldsymbol{\alpha}$ représente une carte des transferts élémentaires, où $\|\cdot\|$ est une norme dont le choix sera discuté dans la section 2.2.6, et où l'opérateur $\boldsymbol{H}$ décrit le mécanisme de transfert de l'intensité de chaque voxel vers les autres voxels de l'image. Remarquons que la préservation de la moyenne globale est automatiquement satisfaite dans cette formulation tant qu'elle n'implique que des *transferts*, c'est-à-dire tant que la somme de chaque ligne de $\boldsymbol{H}$ est nulle.

2.2.3 Formulation mathématique du problème

C'est la matrice $\boldsymbol{H}$ qui décrit la « proximité » des voxels entre eux, c'est-à-dire, attendu qu'un voxel transfère une quantité donnée de sa valeur, quel autre voxel reçoit quelle proportion de cette valeur. $\boldsymbol{H}$ joue donc un rôle crucial, et nous



(a) Transport optimal des valeurs négatives



(b) Projection orthogonale sur l'ensemble convexe des images positives préservant la moyenne

FIGURE 2.1 – Deux stratégies de post-traitement non satisfaisantes. La première (a) peut mener à des transferts asymétriques. La deuxième (b) ne contient aucun aspect local.

examinons dans cette partie les propriétés que cette matrice doit remplir, tout en approfondissant la formulation du problème.

2.2.4 Propriétés de $\boldsymbol{H}$

Une première possibilité serait d'autoriser tous les transferts directs entre paires de voxels voisins et de minimiser la quantité de ces transferts. Si l'on choisit de minimiser la norme L1 d'une telle carte des transferts, on entre alors dans le cadre du transport optimal, qui présente plusieurs inconvénients, comme discuté dans l'introduction de cette partie (section 2.1).

Une deuxième possibilité, qui est celle retenue ici, est d'imposer une propagation isotrope des valeurs négatives. Ainsi, le vecteur de poids $\boldsymbol{\alpha}$ a la même taille que $\boldsymbol{x}$ et $\boldsymbol{y}$, chaque coefficient de $\boldsymbol{\alpha}$ représentant la quantité que le voxel correspondant transfère à ses voisins directs de manière symétrique et prédéfinie. Cette possibilité répond à l'intuition selon laquelle l'intensité d'un voxel négatif devrait être transférée uniformément à ses voisins.

Par souci de clarté, nous avons regroupé les hypothèses que doit vérifier la matrice $\boldsymbol{H}$ en deux catégories. La première contient cinq hypothèses **O1**-**O5**. Elles sont **obligatoires** car la validité mathématique du post-traitement en dépend. Nous considérons également trois hypothèses **facultatives**. On supposera **F1** vérifiée tout au long de ce chapitre, mais son rôle est uniquement de simplifier l'écriture des propositions, et celles-ci s'étendraient facilement si elle n'était pas remplie. Les hypothèses **F2** et **F3** correspondent à des choix naturels, comme discuté ci-dessus, bien qu'elles ne soient pas des conditions nécessaires au sens strict [1]. Enfin, notons que l'énoncé de l'hypothèse **O5** est reporté à la fin de la liste, car elle joue un rôle spécifique qui mérite des commentaires additionnels.

O1 Chaque colonne de $\boldsymbol{H}$ a une somme nulle (les transferts préservent l'activité globale).

O2 La diagonale de $\boldsymbol{H}$ est uniformément positive (par convention, un α_i positif correspond à une augmentation du voxel i).

O3 $\boldsymbol{H}$ est négatif ou nul en dehors de la diagonale (Un voxel ne peut que *réduire* la valeur de ses voisins en transférant sa propre valeur).

1. Nous verrons même dans la section 2.4.2 que l'hypothèse **F3** peut être assouplie avec bénéfice

O4 $\boldsymbol{H}$ est symétrique, $H_{ij} = H_{ji}$ (les coefficients hors diagonale représentent la proximité entre les voxels : H_{ij} est la quantité donnée au voxel i par le voxel j lorsque j transfère sa valeur, qui devrait être la même que le transfert de i à j lorsque i transfère la sienne).

Remarque. *Les hypothèses **O1** à **O4** imposent que la matrice soit une matrice laplacienne, associée à un graphe non orienté particulier correspondant à l'image TEP, où les nœuds sont les voxels, et où un poids H_{ij} est attribué à chaque arête. Pour une introduction sur les graphes et les matrices laplaciennes, voir [Chung and Graham, 1997, Weisstein,].*

En conséquence de son statut de matrice laplacienne, $\boldsymbol{H}$ est une matrice semi-définie positive de rang $n-1$, chaque sous-matrice principale de $\boldsymbol{H}$ est inversible, et son inverse n'a pas de coefficients négatifs [Miekkala, 1993].

F1 L'opérateur $\boldsymbol{H}$ définit une image connexe, c'est-à-dire qu'étant donné deux voxels quelconques, il est possible de les relier par un chemin qui ne traverse que les arêtes où $H_{ij} \neq 0$.

F2 Il est naturel que les transferts élémentaires ne concernent que les voisins directs, afin que le post-traitement soit le plus local possible. La possibilité de ces transferts vers les voisins de deuxième et de troisième ordre [2] est une question de choix. Dans la présente étude, les transferts élémentaires ne sont autorisés qu'entre voisins de premier ordre.

F3 Un choix simple pour $\boldsymbol{H}$ est qu'elle résulte d'un masque élémentaire spatialement invariant et isotrope, *i.e.*, un opérateur de convolution spatialement symétrique. Le noyau de convolution peut être donné par $(-0.5, 1, -0.5)$ pour une image 1D, ou par

$$\begin{pmatrix} 0 & -0.25 & 0 \\ -0.25 & 1 & -0.25 \\ 0 & -0.25 & 0 \end{pmatrix} \tag{2.1}$$

en 2D, par exemple. C'est cette approche qui est retenue dans la majeure partie de cette étude. Des choix différents correspondant à des transferts non isotropes sont étudiés dans la section 2.4.2.

2. Ceux partageant respectivement une arête ou un sommet

2.2.5 Exigence supplémentaire sur $\boldsymbol{\alpha}$

Un des objectifs du NNEPPS est de corriger les incohérences physiques des algorithmes non contraints en redonnant à chaque voxel sa véritable signification : les valeurs négatives n'ont de sens que pour corriger les valeurs trop grandes de leurs voisins, et les voxels négatifs doivent rendre cette valeur à leurs voisins. Cependant, le problème est asymétrique, et il n'y a aucune raison pour que les voxels positifs diffusent une valeur potentiellement « trop grande » : ces valeurs positives possèdent déjà en elles-mêmes leur sens physique. Pour cette raison, nous imposons en outre l'exigence obligatoire suivante :

O5 Chaque composante d'$\boldsymbol{\alpha}$ est positive (ce qui est noté $\boldsymbol{\alpha} \geq \mathbf{0}_n$).

Des tests numériques, que présentent la figure 2.2, ont confirmé que la relaxation de cette dernière contrainte générait des images post-traitées avec de nombreux artefacts.

2.2.6 Choix de la norme

On peut formuler de la manière suivante le problème de post-traitement étudié dans ce chapitre :

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\| \quad \text{tel que} \tag{2.2a}$$

$$\begin{cases} \boldsymbol{\alpha} \geq \mathbf{0}_n \\ \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha} \geq \mathbf{0}_n \end{cases} \tag{2.2b}$$

La proposition suivante, prouvée dans l'annexe A.1, définit la notion de $\boldsymbol{\alpha}$ « minimal » indépendamment de la norme choisie[3].

Proposition 1. *Si $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ sont deux cartes des transferts satisfaisant les conditions (2.2b), le vecteur $\boldsymbol{\gamma}$ défini par $\forall i, \gamma_i = \min(\alpha_i, \beta_i)$ satisfait également (2.2b).*

La proposition 1 a la conséquence directe suivante.

Proposition 2. *(2.2a) est équivalente à $\min_{\boldsymbol{\alpha}} d(\boldsymbol{\alpha})$, pour toute fonction d de $\mathbb{R}^n$ dans $\mathbb{R}$, à condition qu'elle soit strictement croissante selon chacune de ses variables. En particulier, toutes les normes L_p, $p \in]0, \infty[$, induisent des problèmes équivalents.*

3. La proposition 1 peut être étendue à un nombre infini de cartes des transferts en remarquant que (2.2) n'implique que des sous-ensembles fermés et des opérations continues par rapport à $\boldsymbol{\alpha}$.

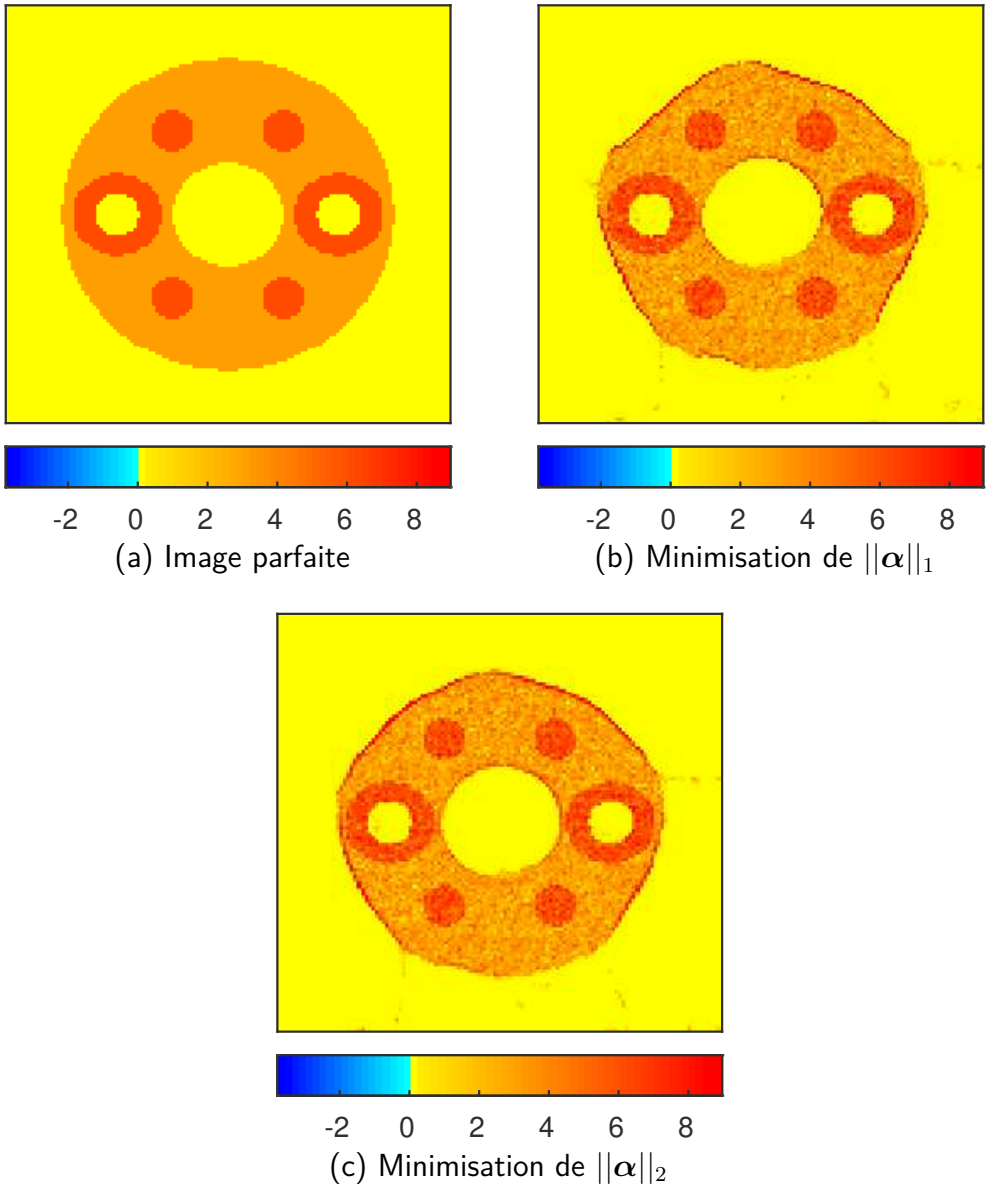


FIGURE 2.2 – Simulations d'un post-traitement n'imposant pas la positivité de $\boldsymbol{\alpha}$. Le fantôme utilisé est celui de la section 2.4.1. Certains pixels de la zone tiède en bordure de la zone froide « aident » les pixels négatifs proches à s'annuler en transférant leur valeur positive, ce qui érode les contours, et donne une valeur artificiellement élevée à leurs autres voisins.

Dans le reste de ce chapitre, on nomme la solution à (2.2) la *carte des transferts minimale.*

2.2.7 Existence et unicité de la solution

Les propriétés d'existence et d'unicité du NNEPPS sont discutées ci-dessous. Les preuves se trouvent respectivement dans les annexes A.2 et A.3.

Proposition 3. *Une solution au problème (2.2) existe si et seulement si la valeur moyenne de l'image initiale* $\boldsymbol{X}$ *est positive.*

Proposition 4. *Si une solution au problème (2.2) existe, elle est unique.*

2.3 Formulation en termes d'optimisation linéaire et résolution à l'aide d'une version améliorée de la méthode duale du simplexe

Le NNEPPS peut être reformulé comme un problème d'optimisation linéaire (section 2.3.1). La structure particulière de ce problème permet à la méthode duale du simplexe d'être particulièrement efficace. En particulier, il est possible de regrouper plusieurs itérations en une seule, et les indices entrants peuvent être choisis de telle sorte qu'ils soient effectivement dans la base d'indices de la solution.

2.3.1 Formulation en termes d'optimisation linéaire

On peut formuler le problème de post-traitement comme la minimisation de $\mathbf{1}_n^T\boldsymbol{\alpha}$ tout en satisfaisant [4] :

$$\begin{cases} \boldsymbol{\alpha} \geq \mathbf{0}_n \\ \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha} \geq \mathbf{0}_n \end{cases} \tag{2.3}$$

Ce problème peut être réécrit sous sa forme standard en optimisation linéaire de la manière suivante [Nocedal and Wright, 1999] : minimiser $\boldsymbol{c}^T\boldsymbol{z}$ sur $\boldsymbol{z}$ en respectant :

$$\begin{cases} \boldsymbol{A}\boldsymbol{z} = -\boldsymbol{x} \\ \boldsymbol{z} \geq \mathbf{0}_{2n} \end{cases} \quad \text{où } \boldsymbol{c} = \begin{pmatrix} \mathbf{0}_n \\ \mathbf{1}_n \end{pmatrix}, \boldsymbol{z} = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\alpha} \end{pmatrix} \text{ et } \boldsymbol{A} = \begin{pmatrix} -\boldsymbol{I}_n & \boldsymbol{H} \end{pmatrix}$$

4. Comme vu dans la proposition 2, la minimisation de $\boldsymbol{\alpha}$ est indépendante de la norme. Par exemple, le problème aurait pu être formulé de manière équivalente avec la norme L2. On choisit ici la norme L1 pour exprimer le problème comme un problème d'optimisation linéaire

Comme l'imposent certains auteurs pour la formulation standard en optimisation linéaire [Nocedal and Wright, 1999], notons qu'ici $\boldsymbol{A}$ a strictement plus de colonnes que de lignes (respectivement $2n$ et n) et est de rang maximal (ici n).

À chaque problème d'optimisation linéaire est associé un problème dual [Nocedal and Wright, 1999]. À partir de la forme standard du NNEPPS, on peut déduire que son problème dual consiste à maximiser $-\boldsymbol{x}^T\boldsymbol{\pi}$ tout en satisfaisant $\boldsymbol{A}^T\boldsymbol{\pi} \leq \boldsymbol{c}$. Ce qui, sous sa forme standard, donne :

$$\text{minimiser } \boldsymbol{b}^T\boldsymbol{s} \text{ tout en satisfaisant : } \begin{cases} \boldsymbol{G}\boldsymbol{s} = \mathbf{1}_n \\ \boldsymbol{s} \geq \mathbf{0}_{2n} \end{cases}$$

avec $\boldsymbol{b} = \begin{pmatrix} \boldsymbol{x} \\ \\ \mathbf{0}_n \end{pmatrix}$, $\boldsymbol{s} = \begin{pmatrix} \boldsymbol{\pi} \\ \\ \boldsymbol{t} \end{pmatrix}$, et $\boldsymbol{G} = \begin{pmatrix} \boldsymbol{H} & \boldsymbol{I}_n \end{pmatrix}$. $\boldsymbol{\pi}$, $\boldsymbol{t}$ et $\boldsymbol{s}$ sont des variables duales de dimensions respectives n, n et $2n$.

Notons qu'à nouveau, $\boldsymbol{G}$ a strictement plus de colonnes que de lignes (respectivement $2n$ et n) et est de rang maximal n.

Les conditions d'optimalité de Karush-Kuhn-Tucker (KKT) pour le problème de post-traitement sont les suivantes :

$$\begin{cases} \boldsymbol{y} = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha} & \text{(2.4a)} \\ \boldsymbol{t} = \mathbf{1}_n - \boldsymbol{H}\boldsymbol{\pi} & \text{(2.4b)} \\ \boldsymbol{s} \geq \mathbf{0}_{2n} & \text{(2.4c)} \\ \boldsymbol{z} \geq \mathbf{0}_{2n} & \text{(2.4d)} \\ \forall i \in [\![1, n]\!],\ \pi_i y_i = 0 & \text{(2.4e)} \\ \forall i \in [\![1, n]\!], t_i \alpha_i = 0 & \text{(2.4f)} \end{cases}$$

Un couple $(\boldsymbol{x}, \boldsymbol{s})$ est solution du problème si et seulement s'il remplit les conditions de Karush-Kuhn-Tucker.

La section suivante contient les grandes lignes de l'algorithme du simplexe appliqué au dual du problème de post-traitement. Nous montrerons dans la section 2.3.3 qu'appliqué au NNEPPS, cet algorithme possède des propriétés qui le rendent particulièrement intéressant.

2.3.2 La méthode duale du simplexe

Si un problème d'optimisation linéaire a des solutions, au moins une de ces solutions appartient à une catégorie appelée *point basique admissible* [Nocedal and Wright, 1999].

Un point admissible $\boldsymbol{s}$ (*i.e.*, un point satisfaisant $\boldsymbol{G}\boldsymbol{s} = \mathbf{1}_n$ et $\boldsymbol{s} \geq \mathbf{0}_{2n}$) est un point basique admissible s'il a au plus n composantes non nulles et s'il existe un sous-ensemble $\mathcal{B} \in [\![1, 2n]\!]$ — appelé la *base d'indices* — vérifiant les trois propriétés suivantes [Nocedal and Wright, 1999] :

— $\mathcal{B}$ contient exactement n indices

— $i \notin \mathcal{B} \implies s_i = 0$.

— La matrice $n \times n$ $\boldsymbol{B} = \boldsymbol{G}_{[\![1,n]\!],\mathcal{B}}$ est inversible.

Les points basiques admissibles sont en fait les sommets du polytope représentant l'ensemble des points admissibles. Nous avons déjà montré dans les propositions 3 et 4 que si l'activité moyenne est positive, —ce qui est bien sûr le seul cas intéressant !—, le problème de post-traitement a une solution unique, qui est donc un point basique admissible.

L'algorithme du simplexe [Dantzig, 1948, Dantzig, 1951, Dantzig et al., 1955] est l'algorithme de résolution le plus utilisé en optimisation linéaire [Nocedal and Wright, 1999, Koberstein, 2005]. Cet algorithme fonctionne en générant successivement une séquence de *points basiques admissibles*. L'algorithme 1 résume la façon dont fonctionne une itération du simplexe.

Algorithme 1 Une itération de l'algorithme du simplexe

1: Étant donné la base d'indices $\mathcal{B}^{(k)}$ et son complémentaire $\mathcal{N}^{(k)} = [\![1, 2n]\!] \setminus \mathcal{B}^{(k)}$.
2: Poser $\boldsymbol{s}^{(k)}_{\mathcal{N}^{(k)}} = \mathbf{0}_n$, et calculer $\boldsymbol{s}^{(k)}_{\mathcal{B}^{(k)}}$ à partir de (2.4b).
3: Poser $\boldsymbol{z}^{(k)}_{\mathcal{B}^{(k)}} = \mathbf{0}_n$, et calculer $\boldsymbol{z}^{(k)}_{\mathcal{N}^{(k)}}$ à partir de (2.4a).
4: **Si** $\boldsymbol{z} \geq \mathbf{0}_{2n}$, **alors** déclarer le succès de l'algorithme et **retourner** $\boldsymbol{z}$.
5: **Sinon**
6: choisir q dans $\mathcal{N}^{(k)}$ tel que $z_q < 0$ (l'indice entrant).
7: Augmenter s_q tout en maintenant $\boldsymbol{s}_{\mathcal{N}^{(k)}\setminus\{q\}} = \mathbf{0}_{n-1}$ et la condition (2.4b), jusqu'à ce qu'une composante p de $\boldsymbol{s}_{\mathcal{B}^{(k)}}$ soit égale à 0 (*i.e.*, l'indice sortant).
8: Poser $\mathcal{B}^{(k+1)} = \mathcal{B}^{(k)} \cup \{q\} \setminus \{p\}$
9: **Fin si**

Bien que l'algorithme du simplexe nécessite généralement au plus $3n$ itérations pour converger, son pire cas a une complexité exponentielle [Klee and Minty, 1972].

Cet inconvénient a motivé la recherche de nouveaux algorithmes qui rivalisent avec la méthode du simplexe, en particulier pour les problèmes de grande taille [Khachiyan, 1979, Karmarkar, 1984]. Dans la section suivante, une version améliorée du simplexe est proposée en exploitant les propriétés de notre problème de post-traitement (équation 2.3).

2.3.3 Résolution à l'aide d'une version améliorée de la méthode duale du simplexe

La taille d'une image TEP 3D complète est importante (généralement autour de 10^7 voxels). Cependant, l'algorithme du simplexe appliqué au problème dual [Lemke, 1954, Koberstein, 2005] (appelé méthode duale du simplexe ou simplexe dual), détient dans ce cas des propriétés très fortes, présentées dans le théorème 1, qui le rendent très efficace.

Théorème 1. *Le problème (2.2) peut être résolu efficacement par une version de la méthode duale du simplexe, en un nombre d'itérations borné par* $n - 1$.

Le théorème 1 est une conséquence des deux propositions suivantes, qui sont respectivement démontrées dans les annexes A.4 et A.5.

Proposition 5. *Si le problème de post-traitement admet une solution (voir proposition 3), la méthode duale du simplexe atteint cette solution.*

Proposition 6. *Si le point de départ choisi est* $\boldsymbol{y}^{(0)} = \boldsymbol{x}$*, à chaque itération du simplexe dual, tous les indices entrants possibles (tous les* i *pour lesquels* $y_i^{(k)} < 0$*) sont bien dans la base d'indices de la solution* i.e.*, tous les choix possibles sont des choix corrects. De plus, l'indice sortant associé à un indice entrant est connu.*

Preuve du théorème 1. La proposition 5 montre que lorsqu'une solution existe (*i.e.*, lorsque la valeur moyenne de l'image est positive, *cf.* proposition 3), la méthode duale du simplexe atteint cette solution en un nombre fini d'itérations. La proposition 6 implique qu'un changement dans la base d'indices est toujours correct. Cela signifie que même si la base d'indices de la solution est complètement différente de celle du point initial, au plus n itérations sont nécessaires. De plus, pour que n itérations fussent nécessaires, il faudrait que l'image finale fût entièrement négative, ce qui est impossible. On peut donc borner par $n - 1$ le nombre d'itérations nécessaires.

La proposition 6 implique également qu'il est possible d'effectuer simultanément nombre de changements dans la base d'indices, puisque tous les indices entrants possibles peuvent être inclus en même temps, et que leurs indices de sortie associés sont connus. □

En suivant les étapes de l'algorithme 1, et en tenant compte des propriétés précédentes, le calcul de $\boldsymbol{s}^{(k)}$ n'est pas nécessaire puisque les indices entrants et sortants sont connus grâce à z. L'algorithme peut donc être écrit comme présenté par l'algorithme 2.

Algorithme 2 Résumé de l'algorithme

1: Initialiser $\boldsymbol{y}^{(0)} = \boldsymbol{x}$, $k = 0$
2: **Tant que** $\boldsymbol{y}^{(k)}$ contient des voxels négatifs
3: $k = k + 1$
4: Créer l'ensemble $\mathcal{Z}^{(k)}$ contenant tous les voxels de $\boldsymbol{y}^{(k-1)}$ négatifs ou nuls
5: Créer $\mathcal{P}^{(k)}$, le complémentaire de $\mathcal{Z}^{(k)}$
6: $\boldsymbol{\alpha}^{(k)}_{\mathcal{Z}^{(k)}} = -(\boldsymbol{H}_{\mathcal{Z}^{(k)}\mathcal{Z}^{(k)}})^{-1}\boldsymbol{x}_{\mathcal{Z}^{(k)}}$
7: $\boldsymbol{\alpha}^{(k)}_{\mathcal{P}^{(k)}} = 0$
8: $\boldsymbol{y}^{(k)} = \boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha}^{(k)}$
9: **Fin tant que**
10: Retourner $\boldsymbol{y} = \boldsymbol{y}^{(k)}$

La borne supérieure de $n - 1$ itérations est bien la borne supérieure théorique optimale : étant donné une matrice L, il est toujours possible de construire artificiellement une image initiale x qui nécessite $n - 1$ itérations, même condensées, pour être post-traitée. Cependant, de même que l'algorithme du simplexe habituel ne demande habituellement que $2n$ à $3n$ itérations alors que son pire cas est exponentiel, l'algorithme que nous proposons ne dépasse généralement pas un nombre constant d'itérations. En effet, dans l'image initiale, les valeurs négatives sont habituellement éparpillées dans toute l'image. Une itération étend alors la zone négative au voisinage de ces valeurs, et ceci se reproduit jusqu'à ce que tous ces voisinages se rejoignent. La taille de l'image n'influençant *a priori* que bien peu la distance entre les voxels négatifs, le nombre d'itérations nécessaire s'en trouvera assez indépendant. L'expérience a montré qu'il ne dépassait généralement pas la quinzaine, même pour des images de 4×10^6 voxels.

La figure 2.3 donne une vue schématique des différences entre le simplexe dual appliqué au NNEPPS (2.3a), et à un autre problème d'optimisation linéaire sans

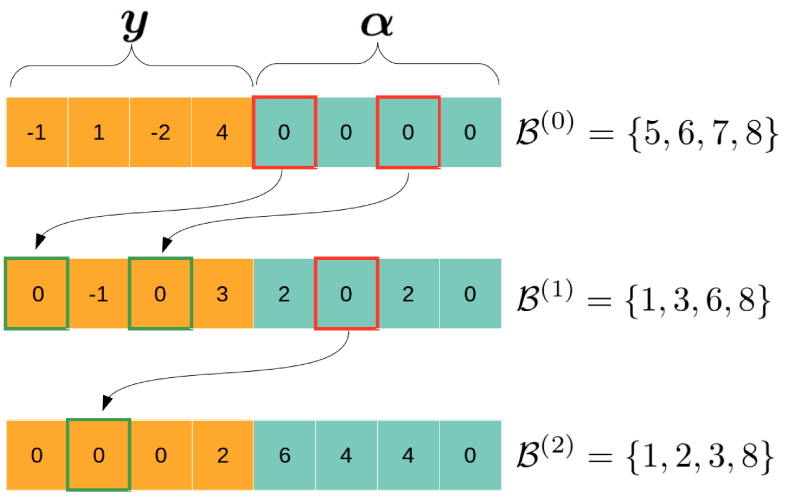
structure particulière. Dans la figure 2.3a, l'indice entrant est toujours dans la partie « $\boldsymbol{y}$ » de $\boldsymbol{z}$, et l'indice sortant est celui correspondant dans la partie $\boldsymbol{\alpha}$. Chaque indice entrant appartient réellement à la base finale. Les itérations sont condensées. Au contraire, pour une exécution normale du simplexe dual (montré dans la figure 2.3b), un seul indice entre dans la base à chaque itération. Cet indice peut être n'importe où, et il n'y a pas de règle générale pour trouver l'indice sortant correspondant. Le choix de l'indice entrant n'est pas toujours optimal : lors de la première itération, l'indice 1 entre dans la base, mais en est retiré par la suite.

Il faut également noter que même appliqué à notre problème, un simplexe dual « habituel » serait beaucoup plus lent que notre algorithme. Premièrement, les itérations n'étant pas condensées, leur nombre serait égal à la différence entre le nombre de voxels nuls dans l'image finale et le nombre de voxels négatifs dans l'image initiale, et varierait donc *a priori* linéairement avec la taille du problème. Deuxièmement, une itération de l'algorithme proposé, même condensée, est moins coûteuse qu'une itération d'un simplexe ordinaire. Celle-ci nécessite en effet de résoudre deux systèmes linéaires de taille n, quand celle-là ne demande qu'une résolution d'un système de taille valant le nombre de voxels négatifs trouvés jusqu'alors. Ce nombre est croissant avec les itérations, mais reste inférieur à n.

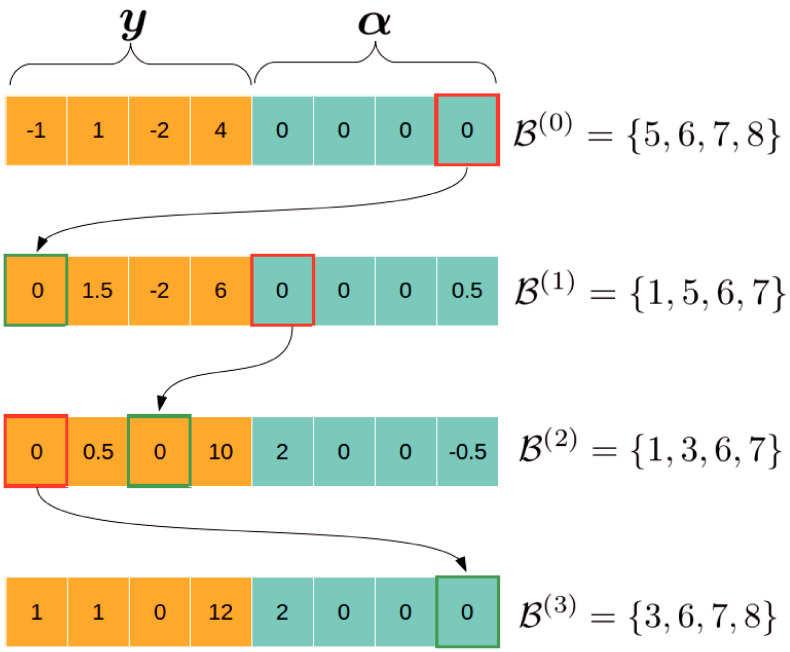
2.3.4 Détails pratiques de mise en œuvre

2.3.4.1 Mise en œuvre

Afin d'éprouver l'algorithme, nous avons développé un programme écrit en C++. Il résout les systèmes linéaires en utilisant la méthode du gradient conjugué [Hestenes and Stiefel, 1952, Atkinson, 1988, Avriel, 2003]. Cette méthode est un algorithme linéaire itératif qui est bien adapté aux matrices creuses, et en particulier aux problèmes symétriques de grande dimension comme le nôtre. Cet algorithme permet à l'utilisateur d'obtenir rapidement des solutions approximatives, avec un niveau de précision choisi, appelé tolérance. L'utilisation d'une solution approximative se justifie ici en raison du nombre limité des itérations du post-traitement, et du fait qu'une itération de notre algorithme n'utilise pas les valeurs des voxels calculées à l'itération précédente, mais seulement la liste des voxels négatifs qui en est issu. La figure 2.4 confirme que ce paramètre de tolérance correspond bien à la précision relative de l'image finale tant qu'il est supérieur à la précision numérique du format



(a) Méthode duale du simplexe appliquée au NNEPPS



(b) Méthode duale du simplexe sur un autre problème

FIGURE 2.3 – Optimisation du vecteur $\boldsymbol{z}$ en appliquant la méthode duale du simplexe au NNEPPS (2.3a) et à un autre problème d'optimisation linéaire sans structure particulière (2.3b). La base $\mathcal{B}$ est représentée à chaque itération. Les indices entrants sont encadrés en vert et les indices sortants en rouge.

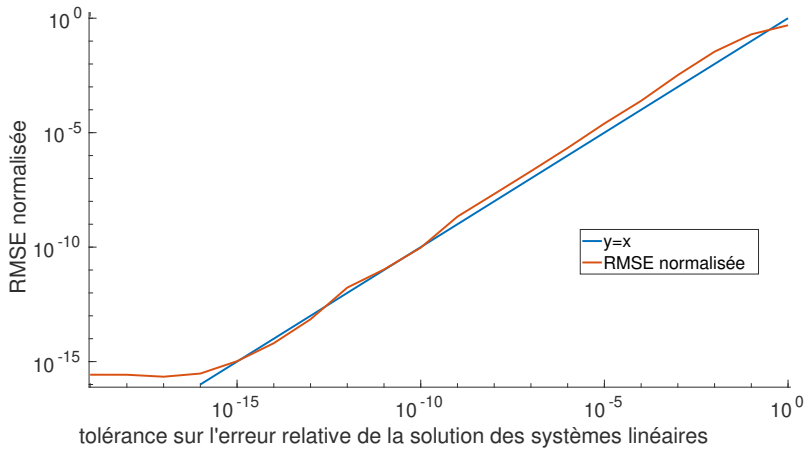


FIGURE 2.4 – RMSE normalisé de l'image finale en fonction de la tolérance sur l'erreur relative de la solution des systèmes linéaires. Le RMSE est normalisé par la racine carrée de la valeur quadratique moyenne de l'image théorique.

de nombre utilisé[5]. Il détermine le temps de calcul par itération.

2.3.4.2 Amélioration de l'algorithme (étape d'initialisation)

Nous avons vu dans la section 2.3.3 qu'en utilisant des images TEP de taille $200 \times 200 \times 109$, environ 15 itérations condensées étaient nécessaires pour que l'algorithme atteigne la solution. Mais un meilleur point de départ $\boldsymbol{y}^{(0)}$ peut être trouvé rapidement en suivant l'algorithme 3. À partir de ce nouveau point de départ, l'algorithme converge plus rapidement vers la même image finale. Le tableau 2.1 compare les temps de calcul et les nombres d'itérations (#iter) nécessaires avec et sans l'étape d'initialisation (init.), pour différents niveaux de tolérance. Nous avons optimisé le critère d'arrêt de l'étape d'initialisation pour chaque niveau de tolérance.

D'un point de vue théorique, l'étape d'initialisation seule devrait converger vers la solution de notre problème. Mais elle ne peut néanmoins pas remplacer l'algorithme que nous proposons. Outre que sa convergence n'est qu'asymptotique, et non pas réellement atteinte comme avec notre algorithme, elle est également moins robuste face aux erreurs numériques, ce qui limite fortement la précision possible, comme l'indique clairement la figure 2.5. Finalement, et toujours contrairement à l'algorithme proposé, l'image obtenue n'est pas totalement exempte de valeurs néga-

5. Cette précision numérique relative vaut 2×10^{-16} si l'on représente les réels sur 64 bits, en suivant le format double précision de la norme IEEE 754, ce qui est notre cas. Ce seuil est intrinsèquement lié au format des nombres utilisés et est donc indépassable

TABLE 2.1 – Durée nécessaire au post-traitement avec et sans l'étape d'initialisation

tolérance	**sans l'init.**		**avec l'init.**		
	durée	**#iter**	**durée de l'init.**	**durée totale**	**#iter**
10^{-3}	2'16"	14	13"	1'19"	7
10^{-6}	7'3"	15	40"	3'52"	6
2×10^{-16}	21'40"	15	1'33"	10'17"	5

tives, ainsi que le montre la figure 2.6b. Il faut donc bien la considérer uniquement comme une étape d'initialisation de l'algorithme principal.

Algorithme 3 Étape d'initialisation

1: Étant donné un nombre de voxels N choisi comme critère d'arrêt :
2: **Tant qu'** au moins N nouveaux voxels négatifs ont été trouvés lors de l'itération précédente
3: **Pour** i dans chaque voxel de l'image
4: **Si** $y_i < 0$ **alors**
5: Incrémenter le nombre de voxels négatifs trouvés
6: Poser $y_i = 0$ et diminuer y_j pour tous les voisins de i de sorte que la moyenne soit préservée.
7: **Fin si**
8: **Fin pour**
9: **Fin tant que**
10: **Retourner** $\boldsymbol{y}$.

2.3.4.3 Choix de la matrice laplacienne $\boldsymbol{H}$

Dans la section 2.2.4, nous avons formulé certaines conditions, obligatoires ou facultatives, que doit remplir $\boldsymbol{H}$. Cependant, il existe encore une certaine marge de manœuvre pour le choix du voisinage et des coefficients hors-diagonale. L'effet précis de ces choix doit encore être étudié en détail. Ici, nous nous limitons au voisinage de premier ordre (de sorte que chaque pixel ait quatre voisins en 2D, et que chaque voxel en ait six en 3D). Dans nos simulations en 2D, les pixels étaient carrés et les coefficients étaient répartis uniformément entre les voisins, de sorte que chacun se vit attribuer un coefficient de $0,25$. Pour les images TEP réelles en 3D, les voxels étaient des parallélépipèdes rectangles de taille $4\,\text{mm} \times 4\,\text{mm} \times 2\,\text{mm}$. Nous avons attribué, à partir de considérations théoriques et de simulations numériques décrites

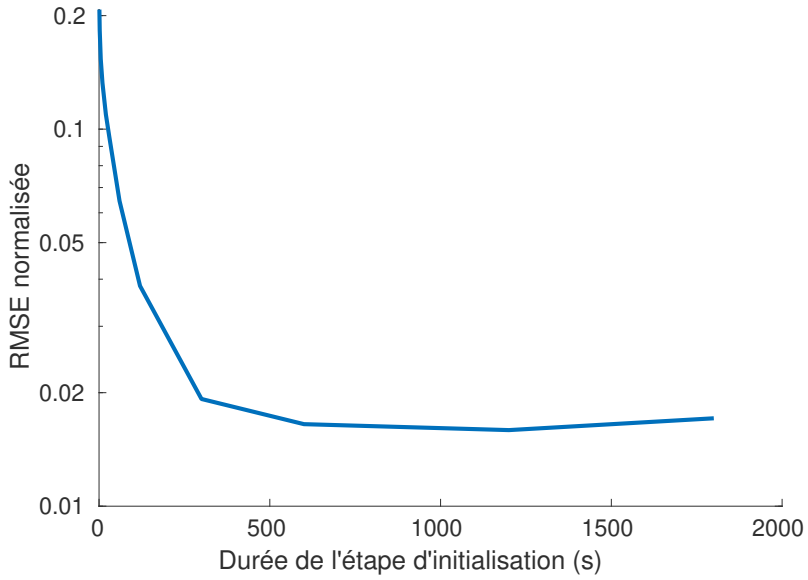


FIGURE 2.5 – RMSE normalisé de l'image issue de l'étape d'initialisation en fonction du temps de calcul. Le RMSE est normalisé par la racine carrée de la valeur quadratique moyenne de l'image théorique.

dans l'annexe A.6, des coefficients de $0,28$ aux deux voisins axiaux, et de $0,11$ aux quatre voisins transversaux. Nous avons naturellement pris en compte les limites de l'image en conservant les mêmes coefficients pour les voisins existants et en adaptant le coefficient sur la diagonale de $\boldsymbol{H}$ pour que chaque ligne eût une somme nulle. Par exemple, pour une image 1D de 4 pixels, $\boldsymbol{H}$ eût été :

$$\begin{pmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 1 & -0.5 & 0 \\ 0 & -0.5 & 1 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \end{pmatrix} \tag{2.5}$$

2.4 Illustrations numériques

Nous analyserons d'abord l'efficacité de l'approche de post-traitement proposée sur une image simulée avant de l'appliquer à un ensemble réel d'images TEP à l'yttrium 90.

2.4.1 Exemple sur un fantôme

La figure 2.6a montre une image 2D contenant des zones distinctes, prenant leur niveau d'intensité moyen parmi trois valeurs : 0, 3 et 6. L'intensité de chaque zone est perturbée par un bruit gaussien de moyenne nulle et d'écart-type unitaire, afin de simuler les erreurs de reconstruction. On remarque que les principales modifications qu'induisent le NNEPPS (figure 2.6c) concernent la zone de moyenne nulle : les pixels initialement négatifs sont mis à zéro et ceux les entourant sont soit annulés soit atténués en préservant la moyenne locale. Cet effet est plus clairement visible sur la figure 2.7 qui affiche les histogrammes des deux images (image brute et image post-traitée utilisant l'algorithme proposé). On peut remarquer que les pixels négatifs sont mis à zéro tandis que les pixels de forte intensité restent inchangés (zones de valeur moyenne théorique (VMT) 3 et 6). La figure 2.6b montre l'image initialisée[6]. Elle est proche de l'image post-traitée, mais est plus bruitée et contient encore nombre de valeurs négatives (la valeur minimale dans l'image étant de $-0,0072$). Il faut noter que l'échelle de couleurs de la figure 2.6 marque une nette différence entre les pixels positifs et les pixels négatifs. Le tableau 2.2 affiche quant à lui les caractéristiques des différentes zones. On y observe que la variance de l'image post-traitée est considérablement diminuée, tandis que les autres caractéristiques sont à peu près inchangées. Les résultats sont également comparés à une image obtenue par une simple projection de l'image brute sur l'ensemble des images positives. La figure 2.6d montre le biais systématique créé par le NNEPPS (moyenne sur 1000 réalisations). On y peut constater que le petit biais positif introduit dans la zone froide par la contrainte de positivité (environ 0,02) est compensé par un biais négatif à la limite entre les régions nulles et non nulles (environ -0,6 pour la limite extérieure). Cependant, ce biais est très localisé au niveau des pixels de la frontière. Cet effet sera examiné plus en détail à partir des données expérimentales dans la section 2.5. La prochaine section a pour but d'étudier une possibilité de correction de ce problème.

6. Nous arrêtâmes l'étape d'initialisation après que 100 parcours entiers sur l'image eurent conduit à un nombre relatif de nouveaux voxels à annuler inférieur à 10^{-6}.

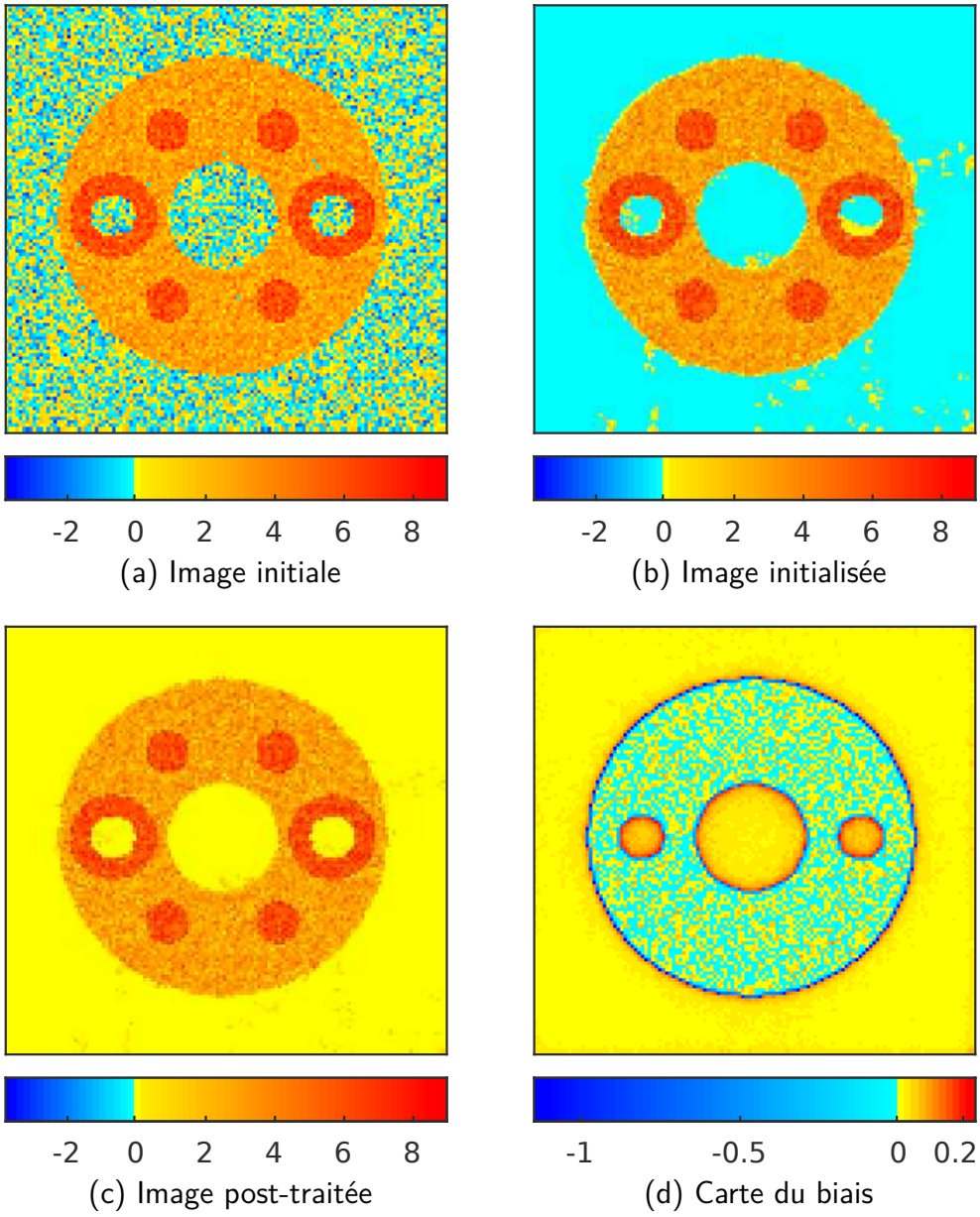


FIGURE 2.6 – Différentes images du même fantôme. (a) montre l'image initiale. Les différentes parties sont remplies par un processus aléatoire suivant une distribution gaussienne de déviation standard 1, avec une moyenne de 0, 3, ou 6 selon les différentes zones. L'image non bruitée est présentée dans la figure 2.2a. (b) montre l'image après l'étape d'initialisation. (c) montre l'image après le NNEPPS. Les régions avec une activité moyenne de 0 sont nettoyées par rapport à l'image initiale, tandis que les autres zones restent inchangées. (d) montre le biais introduit par le NNEPPS, moyenné sur 1000 réalisations.

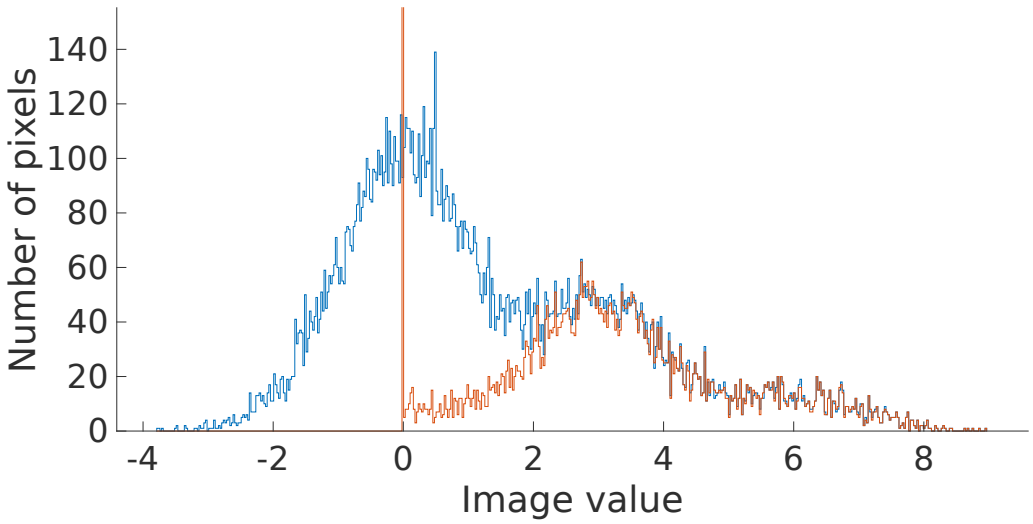


FIGURE 2.7 – Représentation histogramme des deux images de la figure 2.6. L'image brute est en bleu, l'image post-traitée en rouge. Dans l'image post-traitée, 10 079 pixels appartiennent à la case contenant 0.

TABLE 2.2 – Caractéristiques des différentes zones

VMT	**Image brute**		**Image initialisée**		**Image après le NNEPPS**		**Image projetée**	
	Moyenne	**Var.**	**Moyenne**	**Var.**	**Mean**	**Var.**	**Moyenne**	**Var.**
0	-0,018	1,00	0,0043	0,035	0,015	0,015	0,39	0,33
3	3,00	1,00	2,96	1,04	2,94	1,09	3,00	1,00
6	6,04	0,95	6,02	0,98	6,02	0,98	6,04	0,95

2.4.2 Étude d'un laplacien $\boldsymbol{H}$ non isotrope

Afin de corriger le problème de biais négatif induit par le NNEPPS à la frontière entre les régions nulles et non-nulles, il est possible de créer une matrice laplacienne $\boldsymbol{H}$ pénalisant les transferts entre deux régions d'intensités différentes. Bien sûr, on ne connaît pas *a priori* la segmentation correcte de l'image, mais il est possible de pénaliser les transferts entre voisins dont les intensités dans l'image initiale seraient trop éloignées. L'hypothèse **F3** prescrit de choisir $\boldsymbol{H}$ isotrope, mais cette hypothèse est facultative et sa transgression n'enlève rien aux propriétés mathématiques du post-traitement.

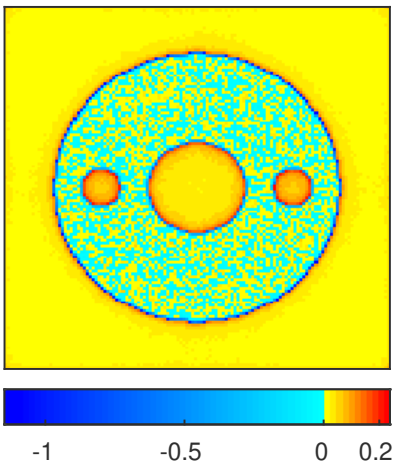
Nous étudions ici deux approches. La première consiste à choisir $H_{ij} = -1/|x_j - x_i|$ pour i et j voisins ; nous noterons cette méthode NNEPPS$_{inv}$. Dans la deuxième, $H_{ij} = -\exp(-|x_j - x_i|/\xi)$, où ξ est un paramètre positif de l'ordre de grandeur des valeurs dans l'image ; nous la noterons NNEPPS$_{exp}$. La méthode utilisée dans le reste de cette thèse, où $\boldsymbol{H}$ est isotrope, est notée NNEPPS$_{iso}$ dans cette section. La figure 2.8 montre les cartes de la variance et du biais pour chacune des trois stratégies, sur le fantôme de la section précédente. Nous avons choisi $\xi = 1$ car les valeurs dans l'image sont de l'ordre de l'unité. Le tableau 2.3 présente certaines métriques issues de ces images. On y constate que le biais, tout particulièrement le biais négatif introduit par le post-traitement à l'interface entre les zones de moyennes nulle et non nulle, est diminué par le NNEPPS$_{inv}$. Il l'est encore plus par le NNEPPS$_{exp}$. En revanche, la variance s'en trouve augmentée, surtout dans les zones que nous venons d'évoquer, et plus avec le NNEPPS$_{exp}$ qu'avec le NNEPPS$_{inv}$. Si les méthodes anisotropes laissent l'erreur quadratique moyenne (MSE) globale à peu près inchangée, elles la réduisent dans les zones difficiles, et c'est cette fois le NNEPPS$_{inv}$ qui obtient le meilleur résultat.

Cette étude mériterait d'être approfondie, mais elle confirme que le choix d'une stratégie de transferts anisotrope permettrait de réduire le biais négatif et l'erreur quadratique moyenne introduits par le NNEPPS à la bordure des zones froides.

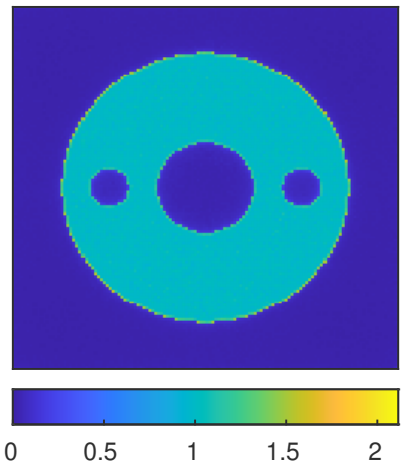
2.4.3 Comparaison entre CPLEX et l'algorithme proposé

Une comparaison de l'algorithme proposé et des algorithmes classiques d'optimisation linéaire a été effectuée et est présentée dans la figure 2.9.

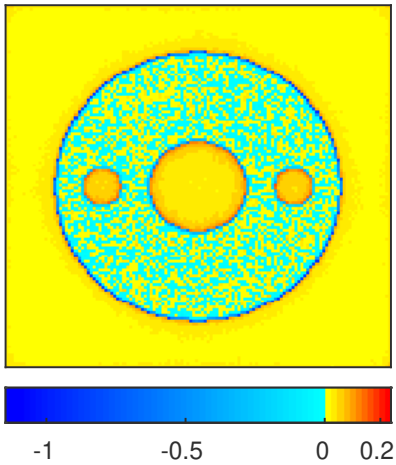
Nous avons effectué les calculs sur un Ubuntu 18.04.2 LTS 64 bits avec une RAM



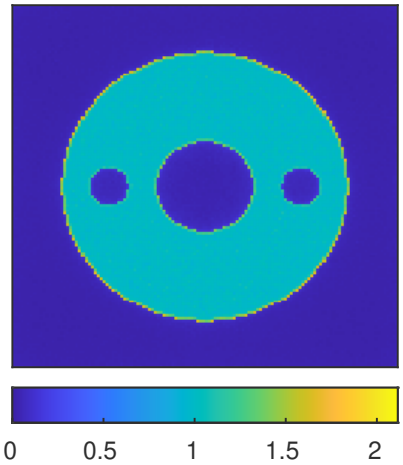
(a) Biais induit par NNEPPS$_{iso}$



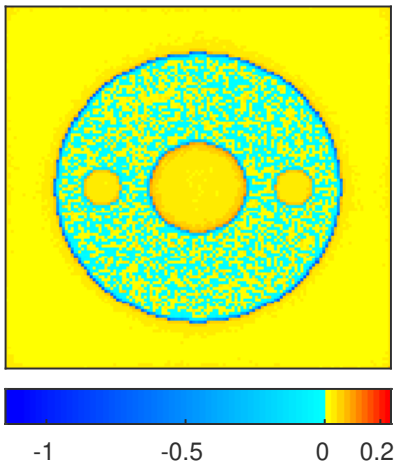
(b) Variance induite par NNEPPS$_{iso}$



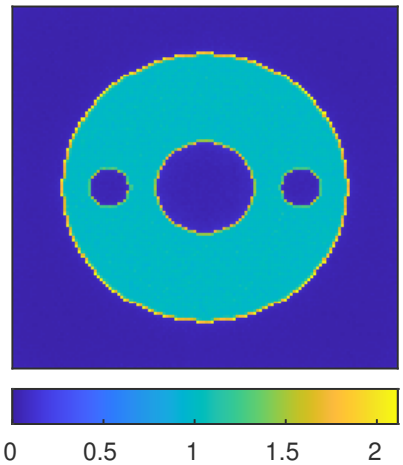
(c) Biais induit par NNEPPS$_{inv}$



(d) Variance induite par NNEPPS$_{inv}$



(e) Biais induit par NNEPPS$_{exp}$



(f) Variance induite par NNEPPS$_{exp}$

FIGURE 2.8 – Biais et variance introduits par le NNEPPS, pour différents choix de $\boldsymbol{H}$ moyenné sur 4000 réalisations. De haut en bas, on constate clairement que le biais diminue, mais que la variance augmente.

TABLE 2.3 – Caractéristiques des cartes de biais et de variance des images post-traitées, pour différents choix de $\boldsymbol{H}$.

	image brute	**NNEPPS**$_{iso}$	**NNEPPS**$_{inv}$	**NNEPPS**$_{exp}$ ($\xi = 1$)
Biais absolu moyen	0,013	0,037	0,033	0,029
Biais minimal	-0,067	-1,2	-0,96	-0,81
Biais maximal	0,062	0,23	0,16	0,12
Variance moyenne	1,0	0,39	0,40	0,41
Variance maximale	1,1	1,6	1,7	2,1
MSE moyenne	1,0	0,41	0,41	0,41
MSE maximale	1,1	2,8	2,5	2,7

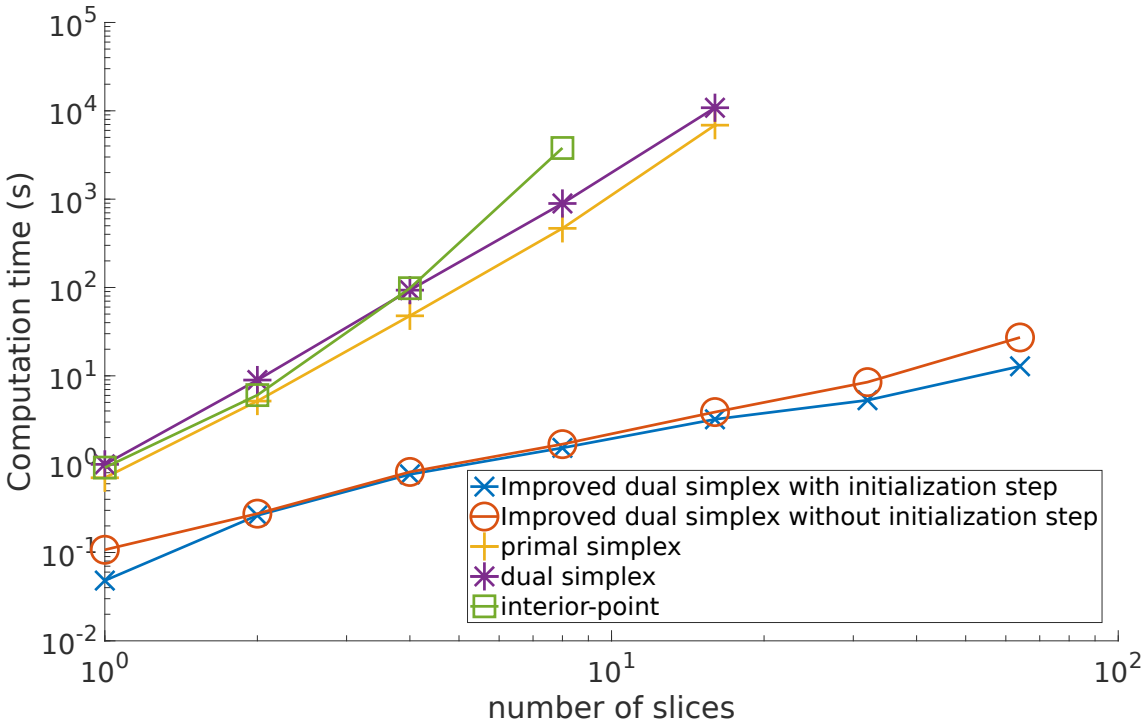


FIGURE 2.9 – Comparaison du temps nécessaire au post-traitement d'un exemple d'image de taille $100 \times 100 \times n$. Les images TEP réelles ont généralement une taille de $200 \times 200 \times 109$

de 7, 5 Go, le code était parallélisé sur 8 CPU Intel Core i7-8550U @ 1,80 GHz. Nous avons exigé un niveau de précision identique pour chaque image finale. L'algorithme proposé s'est révélé être beaucoup plus rapide que le logiciel CPLEX — qui fait figure de référence en optimisation linéaire —, et ce quel que soit l'algorithme choisi. Le temps de calcul était en effet non loin d'être linéaire par rapport au nombre de voxels. En comparaison, le meilleur algorithme CPLEX (le simplexe primal) suit une loi puissance avec un exposant compris entre 3 et 4. Étant donné une image contenant $200 \times 200 \times 109$ voxels et une précision de 10^{-3}, l'algorithme proposé nécessite environ 2 minutes, quand CPLEX est beaucoup trop lent pour être utilisé sur des images de telles dimensions : en extrapolant, on peut estimer le temps de calcul nécessaire à 80 ans.

2.5 Résultats expérimentaux

Afin d'évaluer le NNEPPS dans le contexte de l'imagerie TEP, nous avons réalisé des tests sur des images produites par l'algorithme AML [Van Slambrouck et al., 2015], qui autorise des valeurs négatives dans l'image reconstruite. Ces tests comparent certaines figures de mérite extraites de l'image après le NNEPPS (AML+NNEPPS) avec l'image originale (AML), et avec une image produite par OSEM, qui intègre une contrainte de positivité. Nous avons utilisé le logiciel CASToR [Merlin et al., 2018] pour reconstruire les images.

2.5.1 Mise en œuvre expérimentale

Dans ce travail, chaque algorithme de reconstruction utilisait 21 sous-ensembles. Le fantôme utilisé pour cette étude fut le NEMA IEC.

Quatre acquisitions du même fantôme furent réalisées pendant 7 jours à partir du Biograph mCT de Siemens [Carlier et al., 2015, Willowson et al., 2015]. Le jour 0, on remplit les six sphères du fantôme d'une solution de ^{90}Y avec une concentration d'activité de 2,9 MBq mL^{-1}, et le fond d'une solution de $^{90}YCl_3$ mélangée à du DTPA [7] — pour éviter tout collage aux parois du fantôme — avec une concentration 7,5 fois plus faible que dans les sphères. Quant au cylindre central, on le laissa libre de toute activité. On réalisa ensuite des acquisitions de 1800 s [8] les jours 0, 3, 5 et

7. acide diéthylène triamine penta acétique
8. soit de la même durée qu'en routine clinique

Table 2.4 – Caractéristiques des quatre acquisitions du fantôme NEMA IEC

Nom du fantôme	**Jour de l'acquisition**	**nombre total de coïncidences** $\times 10^6$, % of IEC0	**Fraction de coïncidences aléatoires**	**Fraction de coïncidences diffusées**	**Proportion de coïncidences vraies**
IEC0	0	5,7 (100)	76 %	32 %	16 %
IEC1	3	3,5 (62)	82 %	30 %	13 %
IEC2	5	2,8 (50)	86 %	31 %	9,4 %
IEC3	7	2,4 (43)	90 %	34 %	6,4 %

7. Nous présentons dans le tableau 2.4 le nombre de coïncidences recueillies pour chaque acquisition.

2.5.2 Images

La figure 2.10 montre qu'en utilisant OSEM, le biais positif est clairement visible dans le cylindre froid. AML introduit des valeurs négatives, de sorte que le biais positif dans le cylindre est supprimé (ou réduit), mais celui-ci devient très bruité. Après le NNEPPS, la quasi-totalité du cylindre est rigoureusement égale à 0.

La figure 2.11 montre des images cliniques acquises sur un patient traité avec des micro-sphères d'yttrium 90. Le biais positif est difficile à mettre en évidence puisque la vérité terrain n'est pas connue avec un patient. Cependant, on peut observer le nettoyage du bruit de fond effectué par le NNEPPS. L'activité injectée était de 1,71 GBq au moment de l'imagerie. L'activité du champ visuel (FOV) est de 1,65 GBq pour l'OSEM, et de 1,52 GBq pour AML et AML+NNEPPS (le NNEPPS n'affecte pas l'activité du FOV).

2.5.3 Figures de mérite

La figure 2.12 compare l'activité moyenne en fonction de la racine carrée de l'erreur quadratique moyenne (RMSE) pour AML avec et sans le NNEPPS, à différents nombres d'itérations, et différentes valeurs de A (la borne inférieure).

On en peut déduire deux observations :

1. L'approche permet de réduire le RMSE, qui est approximativement divisée par deux par rapport à l'image avant le NNEPPS. Cependant, cette diminution

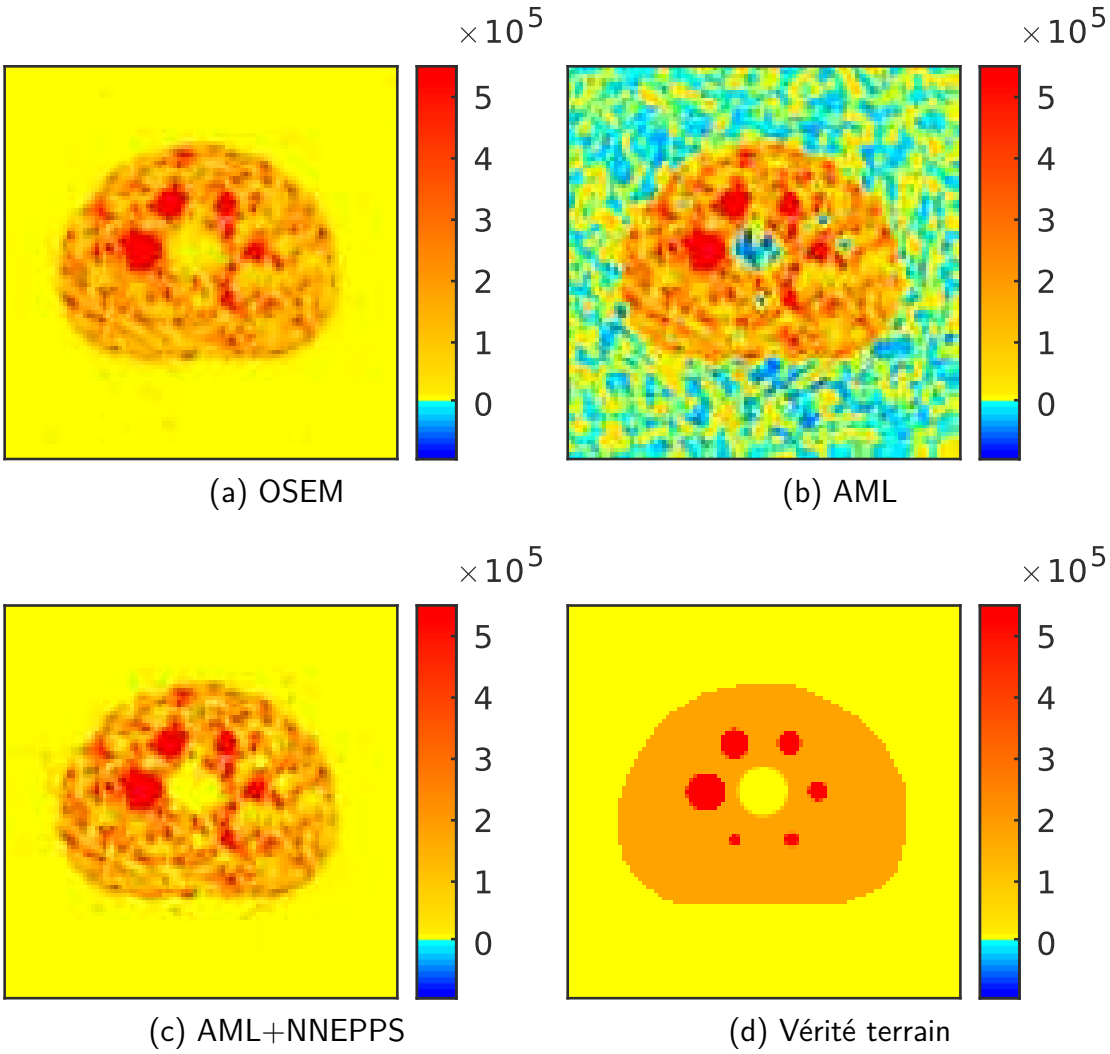
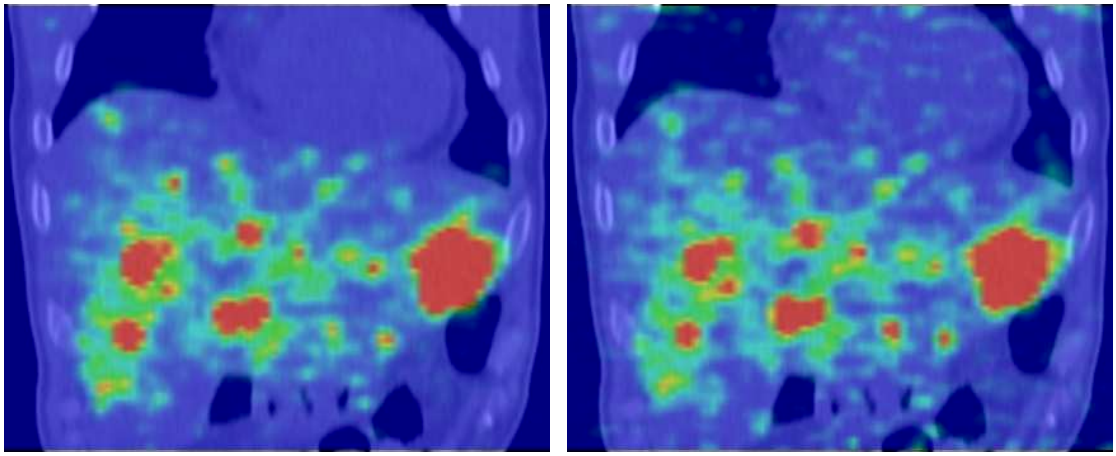
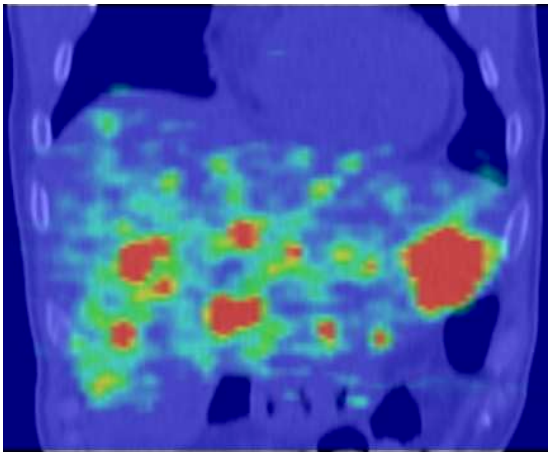


FIGURE 2.10 – Reconstruction du fantôme avec OSEM, AML et AML après le NNEPPS (AML+NNEPPS). Dans AML+NNEPPS, la zone froide centrale est beaucoup plus visible.

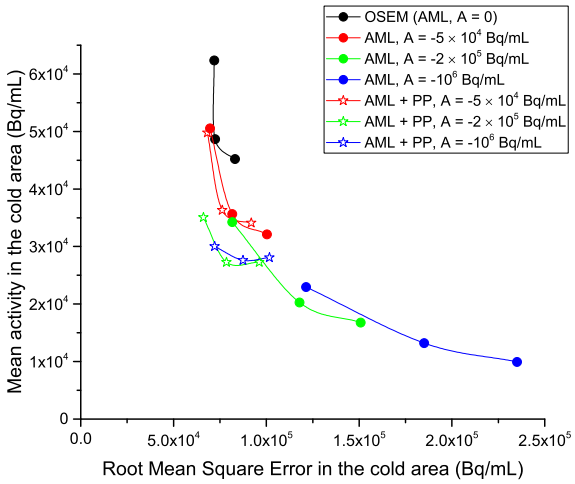


(a) OSEM (b) AML

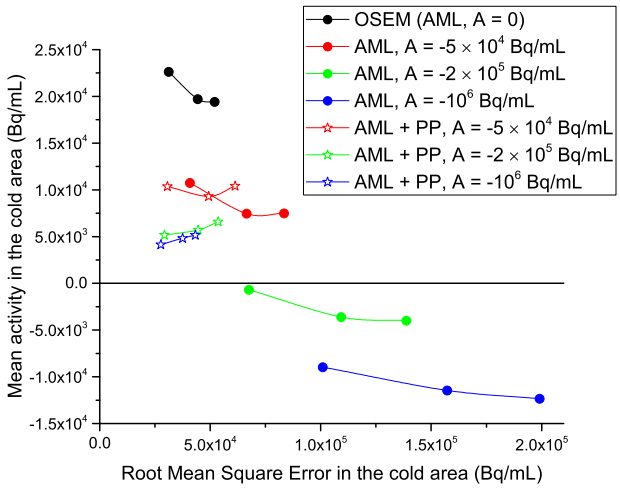


(c) AML+NNEPPS

FIGURE 2.11 – Reconstructions d'un patient traité à l'yttrium 90 avec OSEM, AML et AML+NNEPPS. Le NNEPPS supprime le bruit de fond introduit par AML. Les images affichées sont des fusions TEP-tomodensitométrie. Le patient fut traité pour des localisations hépatiques d'une tumeur neuroendocrine. Activité injectée : 2190 MBq (1709 MBq au moment de l'imagerie). Shunt pulmonaire : 4 %. Nombre de coïncidences : $3,1 \times 10^6$. Fraction de coïncidences aléatoires : 81 %.



(a) IEC1, activité moyenne réelle : 0



(b) IEC3, activité moyenne réelle : 0

FIGURE 2.12 – Graphes affichant l'activité moyenne dans le cylindre froid en fonction du RMSE dans la même zone. Les tests furent effectués pour différentes valeurs du paramètre A d'AML, la borne inférieure de l'image [Van Slambrouck et al., 2015]. On obtint les différents points de la même courbe en faisant varier le nombre d'itérations après lequel on arrête AML. PP signifie NNEPPS. Les courbes pour IEC0 et IEC2 ne sont pas montrées mais sont proches de celles obtenues avec IEC1, qui est représentatif du comportement standard d'AML et du NNEPPS. IEC3 est un cas rare pour lequel AML sous-estime complètement l'activité dans le cylindre froid.

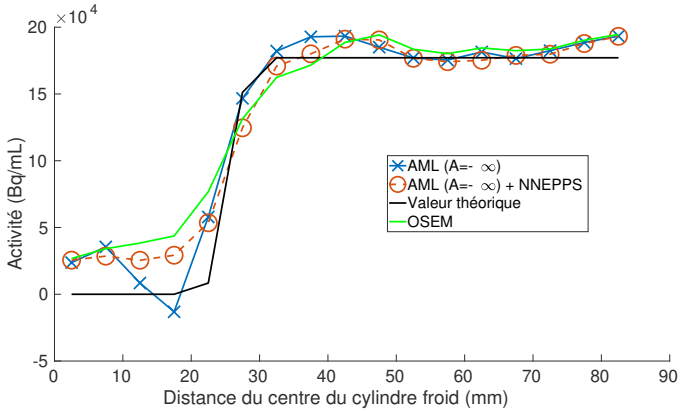
dépend de plusieurs paramètres : pour OSEM, le NNEPPS n'affecte pas l'image (et donc pas le RMSE) puisqu'elle est déjà positive. Pour IEC3, avec $A = -10^6\,\mathrm{Bq\,mL^{-1}}$, la figure 2.12b montre qu'après trois itérations, le RMSE est égal à $2{,}0 \times 10^5\,\mathrm{Bq\,mL^{-1}}$ pour AML, mais diminue à $4{,}3 \times 10^4\,\mathrm{Bq\,mL^{-1}}$ après le NNEPPS, ce qui fait une réduction d'un facteur 4,6.

2. L'activité moyenne du cylindre froid reste approximativement inchangée par le NNEPPS jusqu'à ce qu'elle atteigne une limite (d'environ $2{,}5 \times 10^4\,\mathrm{Bq\,mL^{-1}}$ pour IEC1). Par exemple, pour $A = -2 \times 10^5\,\mathrm{Bq\,mL^{-1}}$, l'activité à la première itération d'AML est d'environ $3.4 \times 10^4\,\mathrm{Bq\,mL^{-1}}$ avant le NNEPPS et reste la même après. Mais à la deuxième itération d'AML, l'activité avant le NNEPPS est inférieure à la limite (environ $2{,}0 \times 10^4\,\mathrm{Bq\,mL^{-1}}$), de sorte que l'activité après est approximativement égale à la limite.

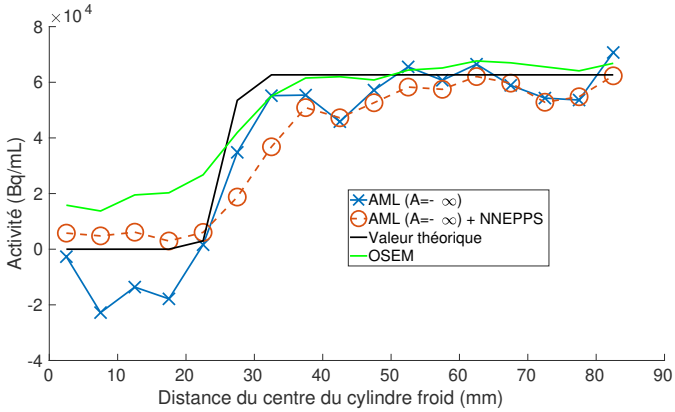
Le NNEPPS ne préserve donc pas rigoureusement l'activité moyenne dans le cylindre froid, alors même qu'il est censé préserver les moyennes locales. Une petite partie de l'activité négative s'est en réalité propagée aux régions chaudes avoisinantes, augmentant par contrecoup l'activité dans le vrai cylindre froid, qui n'est donc pas totalement fixé à 0, comme il devrait idéalement l'être. Mais ceci était inévitable, une image positive ne pouvant annuler le biais positif sans être totalement exempte de bruit.

Le biais négatif introduit par le NNEPPS à la frontière entre les zones chaudes et froides (déjà observée dans la figure 2.6d sur un exemple numérique), est illustrée sur les données du fantôme dans la figure 2.13, qui montre l'activité moyenne en fonction de la distance au centre du cylindre froid.

L'activité moyenne dans le voisinage proche du cylindre froid est effectivement réduite, mais cet effet est limité et localisé. La figure 2.13a montre que le biais négatif induit peut atteindre 15% dans le voisinage très proche du cylindre froid pour IEC1, mais il diminue rapidement et disparaît complètement à une distance comprise entre 10 mm et 15 mm de la frontière. De plus, dans cette zone frontalière, l'erreur dans l'activité estimée reste comparable à celle produite par OSEM. La figure 2.13b montre un cas rare, obtenu avec un nombre de données très faible, où AML sous-estime complètement l'activité dans le cylindre froid. Après le NNEPPS, l'image est mieux estimée dans le cylindre froid, mais la région voisine doit compenser cette augmentation d'activité. C'est une conséquence inévitable de la préservation de l'activité globale. Par rapport à OSEM, AML+NNEPPS est finalement moins



(a) IEC1



(b) IEC3

FIGURE 2.13 – Graphes représentant l'activité moyenne mesurée en fonction de la distance au centre du cylindre froid. La valeur théorique et la reconstruction OSEM sont données à titre de références. Nous n'avons pas représenté les courbes pour IEC0 et IEC2, mais elles sont proches de celle d'IEC1, qui est représentatif du comportement standard d'AML et du NNEPPS. IEC3 est un cas rare pour lequel AML sous-estime complètement l'activité dans le cylindre froid.

biaisé dans la région froide, mais l'est dans ce cas légèrement plus dans le voisinage proche de cette dernière.

2.6 Discussion et conclusion

De nombreux algorithmes de reconstruction itératifs actuellement utilisés, tels qu'OSEM, intègrent une contrainte de positivité. Cette contrainte est naturelle d'un point de vue physique, mais lorsqu'elle s'associe à du bruit, elle induit un biais positif important dans les zones froides, ce qui peut être problématique lorsqu'on envisage une analyse quantitative. Ce problème est particulièrement important dans le contexte de basses statistiques de l'yttrium 90. L'utilisation d'algorithmes non soumis à cette contrainte de positivité peut représenter une solution, mais au prix d'un niveau de bruit encore plus élevé et de l'introduction d'incohérences physiques.

À partir de considérations sur l'information portée par les valeurs négatives, nous proposons une étape de post-traitement rendant l'image positive tout en préservant son activité moyenne aussi localement que possible. Le NNEPPS introduit un petit biais positif dans les zones froides, mais qui demeure à un niveau bien inférieur à celui d'OSEM. À ce petit biais positif est associé un petit biais négatif dans les zones proches des zones froides. L'effet de ce biais négatif sur les petites lésions reste à étudier, mais là encore, ce biais négatif reste inférieur à celui d'OSEM. Une perspective intéressante pour réduire encore le nombre de transferts entre régions, et donc les biais créés, consisterait à étendre l'étape de post-traitement à un étalement anisotrope, dans l'esprit de la diffusion anisotrope [Perona and Malik, 1990]. La section 2.4.2 confirme le potentiel de telles méthodes.

Le RMSE dans les zones froides est également réduit par le NNEPPS à un niveau comparable à celui d'OSEM. Qualitativement, les images sont aussi visuellement plus faciles à interpréter après le NNEPPS. En particulier, les zones froides sont beaucoup plus visibles. La méthode a également montré des signes de robustesse en donnant de bons résultats sur un cas peu commun (voir les figures 2.12b et 2.13b).

Nous proposons un algorithme exécutant le NNEPPS. Son code source est disponible en libre accès à l'adresse codeocean.com/capsule/0893820. Il agit comme une méthode duale du simplexe condensée, ce qui est justifié par la nature particulière du problème. Le temps nécessaire pour post-traiter une image TEP 3D entière est d'environ 2 minutes à partir d'un ordinateur équipé de huit CPU Intel Core i7-8550U

@ 1,80 GHz.

Parmi plusieurs perspectives, l'une serait d'inclure le NNEPPS dans le processus de reconstruction en appliquant la méthode après chaque itération d'un algorithme itératif. Cependant, les propriétés de convergence d'un tel algorithme, ainsi que son effet sur l'image finale, doivent encore être étudiés.

Les travaux futurs pourraient également inclure une étude précise du type de voisinage à considérer et du choix des coefficients de $\boldsymbol{H}$.

Enfin, le NNEPPS n'est pas intrinsèquement limité à la TEP. Son champ application pourrait s'étendre à d'autres modalités d'imagerie, où une contrainte de positivité est imposée d'un point de vue physique. On peut penser à l'IRM, pour lequel on a également rapporté un biais positif [Gudbjartsson and Patz, 1995, Andersen and Kirsch, 1996], ou à la diffusion avec décroissance [Nagarajan and Nakshatrala, 2011, Nakshatrala and Valocchi, 2009, Nakshatrala et al., 2016]. De plus, le principe du NNEPPS n'est pas limité aux contraintes de positivité : la contrainte pourrait être un seuil quelconque, correspondant à une valeur physique extrême dans l'image.

Chapitre 3

Analyse comparative multi-objectifs de divers algorithmes de reconstruction

Sommaire

3.1 Algorithmes pénalisés sans contrainte 93

3.2 Cadre expérimental 94

3.2.1 Acquisitions du fantôme 94

3.2.2 Application clinique 94

3.2.3 Paramètres de la reconstruction 95

3.2.4 Métriques d'évaluation 96

3.2.5 Analyse multi-objectifs 97

3.2.6 Effet sur le calcul dosimétrique 100

3.3 Résultats 103

3.3.1 Comparaison de NEG-ML et d'AML 103

3.3.2 Comparaison d'AML et de PPGML 104

3.3.3 Évaluation globale des différents paramètres 104

3.3.4 Comparaison entre la pénalité quadratique et la pénalité aux différences relatives 107

3.3.5 Évaluation dosimétrique 108

3.3.6 Évaluation clinique 108

3.4 Discussion et conclusion 109

Dans ce chapitre, nous comparons les algorithmes pénalisés et non contraints (avec et sans NNEPPS) à la technique OSEM standard afin de déterminer laquelle de ces approches est la mieux adaptée à l'analyse quantitative et qualitative. Nous avons également introduit une version pénalisée d'AML pour mesurer l'influence combinée des algorithmes pénalisés et non contraints.

En utilisant des données expérimentales acquises provenant de fantômes, nous évaluons l'influence de quatre paramètres (force de la pénalité, intensité des valeurs négatives, nombre d'itérations et inclusion ou non du NNEPPS) en fonction de différentes métriques de bruit et de biais. Pour cela, nous avons suivi une approche multi-objectifs fondée sur le formalisme de Pareto [Zitzler et al., 2001]. Nous avons ensuite évalué les meilleures approches à l'aide d'études dosimétriques sur des données expérimentales. À notre connaissance, aucune étude utilisant des données réelles yttrium 90 n'avait encore été menée à bien pour évaluer la performance de ces algorithmes.

3.1 Algorithmes pénalisés sans contrainte

Nous proposons un algorithme pénalisé autorisant des valeurs négatives dans l'image, afin d'éprouver l'effet conjoint des deux méthodes. Étant donné qu'il est possible de considérer AML comme une version « décalée » de MLEM, nous avons créé une version pénalisée d'AML en utilisant la même idée sur l'algorithme PPGML, qui est très robuste aux instabilités numériques. Nous avons nommé cet algorithme APPGML [1]). L'équation de mise à jour s'obtient donc à partir de l'équation (1.11) en décalant les valeurs de l'image et du sinogramme de A et $A\sum_{j'} S_{ij'}$ respectivement :

$$x_j^{k,s+1} = x_j^{k,s} + (x_j^{k,s} - A)\frac{\sum\limits_i S_{ij}\left(\frac{y_i - A\sum_{j'} S_{ij'}}{\bar{y}_i(\boldsymbol{x}^{k,s}) - A\sum_{j'} S_{ij'}} - 1\right) - \beta\partial_j R(\boldsymbol{x}^{k,s})}{\sum\limits_i S_{ij} + \beta(x_j^{k,s} - A)\partial_j^2 R(\boldsymbol{x}^{k,s})} \tag{3.1}$$

Le lecteur notera que nous avons choisi de ne pas utiliser de décalage à l'intérieur de la fonction de pénalité, car nous pensons que celle-ci doit s'appliquer aux vraies valeurs de l'image. Cela ne fait dans tous les cas aucune différence si la fonction de

1. pour PPGML avec une borne inférieure A.

pénalité est inchangée par un décalage de l'image, *i.e.* $R(\boldsymbol{x} - A) = R(\boldsymbol{x})$, ce qui est le cas pour la pénalité quadratique. Il notera également que la fonction potentielle des différences relatives n'est pas compatible avec cet algorithme, car la première suppose implicitement des images positives et n'a pas d'extension triviale dans le cas contraire. Dans ce qui suit, nous n'utiliserons donc APPGML qu'avec la fonction potentielle quadratique.

3.2 Cadre expérimental

3.2.1 Acquisitions du fantôme

Le fantôme utilisé dans cette études, et les quatre acquisitions correspondantes, sont ceux présentés dans la section 2.5.1. Bien que nous ayons pris en compte les quatre acquisitions pour l'analyse, seuls les résultats pour IEC1 ont été tracés, sauf pour l'évaluation dosimétrique, où il s'agit d'IEC0. Cependant, sauf mention contraire, les résultats de chaque acquisition étaient similaires.

3.2.2 Application clinique

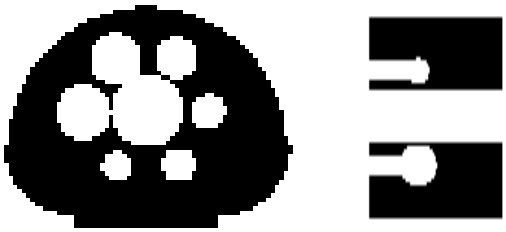
Nous avons considéré les données post-radioembolisation d'un patient traité dans notre service. La radio-embolisation fut réalisée dans un but palliatif en 4e ligne de traitement chez un patient présentant des métastases hépatiques bilobulaires d'un adénocarcinome du sigmoïde. Une dose de 1,9 GBq de microsphères de résine fut injectée dans l'artère hépatique droite. Le traitement fut effectué selon les instructions du fabricant et l'activité à injecter fut établie sur la base des informations dosimétriques, afin de délivrer une dose absorbée inférieure à 40 Gy au foie non tumoral, comme recommandé. Plusieurs régions d'intérêt (ROI) furent délimitées pour l'évaluation dosimétrique de la TEP-^{90}Y post-radioembolisation : un infiltrat tumoral du dôme hépatique, une ROI englobant tout le foie et une ROI autour du cœur, qui est considéré comme un organe exempt de toute activité, mais qui est souvent « pollué » en raison de sa proximité avec le dôme hépatique et du niveau de bruit.

3.2.3 Paramètres de la reconstruction

C'est le logiciel CASToR (de l'anglais *Customizable and Advanced Software for Tomographic Reconstruction*), dans lequel nous avons implémenté APPGML, que nous avons utilisé pour reconstruire toutes les images [Merlin et al., 2018]. Les sinogrammes d'atténuation, de normalisation, de coïncidences aléatoires et diffusées ont été calculés par le logiciel fourni par le constructeur Siemens. Les données n'ont pas été pré-corrigées. Nous choisîmes les paramètres généraux de reconstruction aussi proches que possible de ceux utilisés dans la pratique clinique actuelle pour les reconstructions ^{90}Y. La taille du voxel était de $4 \times 4 \times 2$ mm^3. Des informations sur le temps de vol (TOF) étaient disponibles. Une fonction d'étalement du point (PSF) gaussienne isotrope 3D fut utilisée dans l'image avec une largeur à mi-hauteur (FWHM) de 4,5 mm, déterminée à partir des travaux de Panin *et al.* [Panin et al., 2006]. Pour les algorithmes de reconstruction non pénalisés, un post-filtrage avec le même noyau de 4,5 mm que la PSF fut utilisé (voir section 1.1.5 pour une explication de ce qu'est la PSF). Cette technique est connue sous le nom de sieve [Grenader, 1981, Stute and Comtat, 2013]. Tous les algorithmes furent exécutés avec 21 sous-ensembles (valeur par défaut pour les reconstructions Biograph mCT).

Nous avons étudié quatre paramètres d'intérêt : la force de pénalité β pour PPGML et APPGML, l'intensité des valeurs négatives dans l'image (par ψ pour NEG-ML et A pour AML et APPGML), le nombre d'itérations et l'utilisation (ou non) de l'étape de post-traitement NNEPPS. Ces paramètres d'intérêt ont été modifiés comme détaillé dans le tableau 3.3

Pour obtenir des erreurs minimales dans la dosimétrie volumétrique à l'aide d'histogrammes dose-volume, Siman *et al.* [Siman et al., 2019] suggère d'effectuer 36 itérations équivalentes à MLEM avec un post-filtrage gaussien de 5,2 mm FWHM dans le contexte spécifique des données ^{90}Y, et pour une taille de voxels de $2,6$ mm $\times 2,6$ mm $\times 3,3$ mm. Sur la base de ces résultats, nous avons défini une « méthode de reconstruction de référence » comme étant OSEM avec 2 itérations (42 itérations équivalentes à MLEM) et un post-filtrage avec le noyau PSF. La modélisation de la PSF et le TOF ont également été utilisés conformément aux études précédentes [Carlier et al., 2015]. Cette méthode de reconstruction de référence (appelée ci-après l'« OSEM de référence ») est proche de celle actuellement utilisée dans la pratique clinique.



(a) vue de face (b) Vue XZ

FIGURE 3.1 – ROI pour le fond

3.2.4 Métriques d'évaluation

La liste suivante présente les quatre figures de mérite (FOM) utilisées dans cette étude.

— Coefficient de récupération de la concentration moyenne dans une zone chaude ($\mathrm{CRC}_{\mathrm{chaud}}$). On utilise cette métrique afin d'évaluer la récupération de l'activité dans les sphères chaudes. Elle est définie comme l'activité moyenne m_{o} dans une région divisée par l'activité réelle f_0 :

$$\mathrm{CRC}_{\mathrm{chaud}} = \frac{m_{\mathrm{o}}}{f_0} \tag{3.2}$$

La valeur optimale de cette métrique est 1. Par conséquent, dans les calculs des fronts de Pareto (définis ci-après), nous avons minimisé le paramètre $|1 - \mathrm{CRC}_{\mathrm{chaud}}|$.

Nous avons utilisé cette métrique sur la troisième sphère chaude la plus grande (diamètre de 22 mm), en raison de sa taille intermédiaire. Pour les calculs, on considère un voxel comme étant dans la ROI si son centre se trouve dans la ROI.

— Rugosité de l'image (IR). La rugosité de l'image est l'écart type des valeurs des voxels dans une ROI, normalisé par sa valeur moyenne :

$$\mathrm{IR} = \frac{\sqrt{\frac{1}{I-1} \sum\limits_{i \in \mathrm{ROI}} (x_i - m_{\mathrm{ROI}})^2}}{m_{\mathrm{ROI}}} \tag{3.3}$$

Cette FOM est étroitement liée au bruit de l'image perçu par un humain et sa valeur optimale est donc 0. Dans ce chapitre, on le calcule sur le fond du fantôme : une grande région de 200 112 voxels excluant les sphères et le cylindre central et suffisamment éloignée de toute frontière. La figure 3.1 montre les contours de cette ROI.

— Activité moyenne dans le cylindre froid ($\mathrm{CRC}_{\mathrm{froid}}$). L'un des objectifs de cette étude étant de trouver des algorithmes réduisant le biais positif, nous avons mesuré l'activité moyenne dans le cylindre froid (valeur théorique : $0\,\mathrm{MBq\,mL^{-1}}$). Nous avons minimisé $|\mathrm{CRC}_{\mathrm{froid}}|$) dans les calculs des fronts de Pareto. Nous avons inclus dans la région d'intérêt tous les voxels appartenant entièrement au cylindre froid.

— Racine carrée de l'erreur quadratique moyenne (RMSE) dans le cylindre froid. La réduction du biais positif global peut diminuer la confiance possible dans les valeurs d'un seul voxel. Par conséquent, nous avons calculé le RMSE qui quantifie l'erreur des voxels individuels. La région d'intérêt considérée pour cette métrique était identique à celle choisie pour évaluer l'activité moyenne dans le cylindre froid.

$$\mathrm{RMSE} = \sqrt{\frac{1}{I} \sum_{i \in \mathrm{ROI}} (x_i - f_0)^2} \tag{3.4}$$

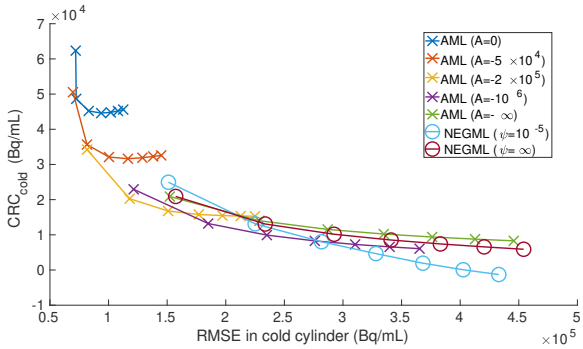
avec $f_0 = 0$. La valeur optimale de cette figure de mérite est 0.

3.2.5 Analyse multi-objectifs

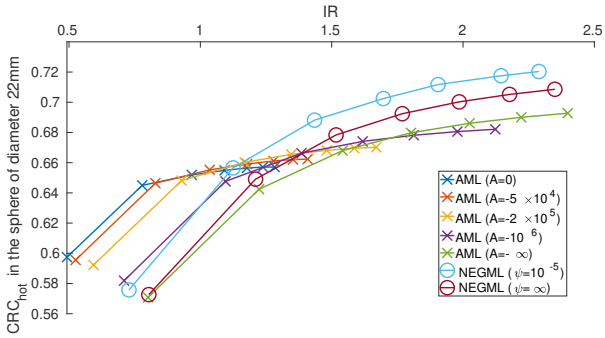
Un apport méthodologique majeur de cette étude est l'utilisation du formalisme de Pareto pour analyser les différentes métriques d'un point de vue multi-objectif. Ce formalisme, que nous présentons ci-dessous, est très utilisé dans d'autres domaines, comme l'ingénierie, mais est encore peu utilisé en TEP. À notre connaissance, et bien que de nombreuses études aient tracé des courbes représentant le contraste en fonction du bruit, seule une étude a utilisé ce formalisme, dans un contexte différent [Abouhawwash and Alessio, 2021].

Un problème d'optimisation multi-objectifs peut s'écrire de la manière suivante :

$$\min(f_1(\boldsymbol{v}), f_2(\boldsymbol{v}), \ldots, f_k(\boldsymbol{v}))\,;\ \text{tel que } \boldsymbol{v} \in V \tag{3.5}$$



(a) $\mathrm{CRC}_{\mathrm{froid}}$ vs RMSE



(b) $\mathrm{CRC}_{\mathrm{chaud}}$ vs IR

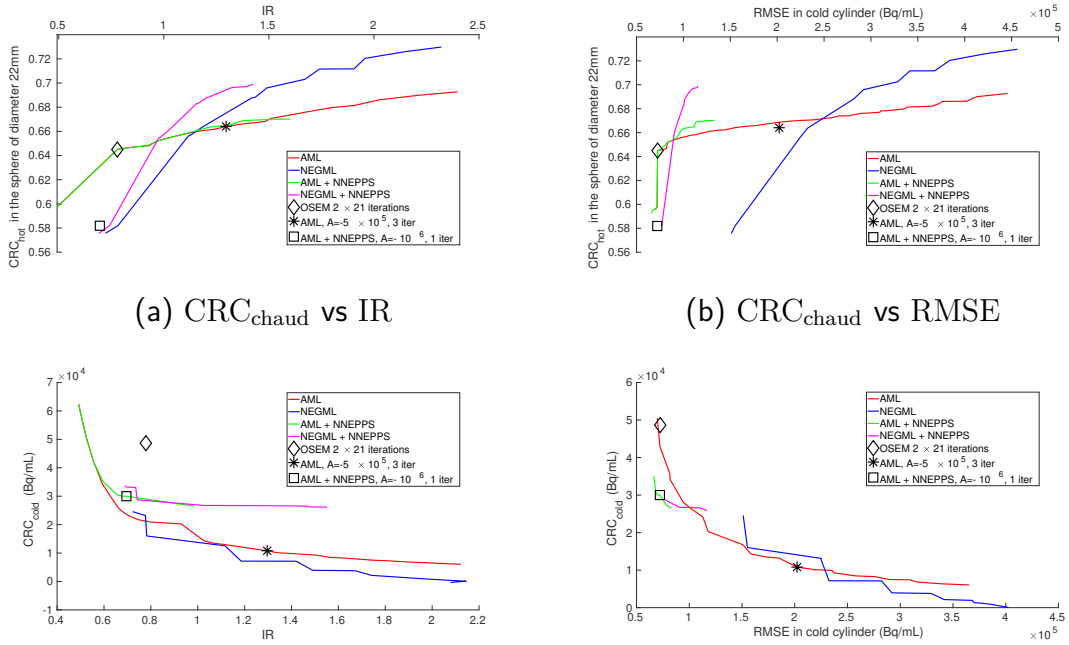
FIGURE 3.2 – Activité récupérée dans le cylindre froid en fonction du RMSE dans ce même cylindre (a) et du $\mathrm{CRC}_{\mathrm{chaud}}$ en fonction du IR (b) pour AML et NEG-ML. Chaque point représente un nombre différent d'itérations. Les courbes sont pour IEC1

où k objectifs sont à optimiser.

Dans le cas général, il n'existe pas de vecteur $\boldsymbol{v}$ minimisant tous les objectifs, et l'on s'intéresse donc à l'ensemble des solutions optimales au sens de Pareto. Une solution est optimale au sens de Pareto s'il n'est pas possible d'en améliorer un seul objectif sans en dégrader au moins un autre. Cette notion se formalise grâce aux définitions suivantes :

Définition 1. *[Collette and Siarry, 2004] Si toutes les composantes de f_a sont inférieures ou égales aux composantes de f_b, et qu'au moins une de ses composantes est strictement inférieure, on dit que f_a domine f_b.*

Remarque. *Intuitivement, cela signifie que f_a est objectivement meilleur que f_b, et ce, peu importe les objectifs les plus importants. S'il n'existe aucune solution dominant f_c, alors f_c est* non dominée.



(a) $\mathrm{CRC}_{\mathrm{chaud}}$ vs IR

(b) $\mathrm{CRC}_{\mathrm{chaud}}$ vs RMSE

(c) $\mathrm{CRC}_{\mathrm{froid}}$ vs IR

(d) $\mathrm{CRC}_{\mathrm{froid}}$ vs RMSE

FIGURE 3.3 – Comparaison d'AML et de NEG-ML avec et sans le NNEPPS en utilisant le fantôme IEC1 pour différentes métriques. Dans chaque graphique, les quatre fronts de Pareto 2D de deux variables (le nombre d'itérations et A ou ψ) sont représentés. L'OSEM de référence et deux autres algorithmes (choisis empiriquement pour leur bon compromis) sont également représentés. Notez que ces algorithmes se trouvent sur le front de Pareto 4D, mais pas nécessairement sur les fronts 2D-Pareto représentés ici. Notez également que bien que toutes les données soient prises en compte dans chaque graphique, les points extrêmes des fronts de Pareto peuvent être différents d'un graphique à l'autre, car un point dominé dans un front 2D peut ne pas l'être dans un autre. Par exemple, l'IR de AML+NNEPPS varie de 0,5 à 1,6 sur la figure. 3.3a et de 0,5 à 1,0 sur la figure. 3.3c.

Définition 2. *[Collette and Siarry, 2004] Le front de Pareto est défini, dans l'espace des objectifs, comme l'ensemble des solutions non dominées.*

Définition 3. *[Collette and Siarry, 2004] L'ensemble de Pareto est défini, dans l'espace de recherche, comme l'antécédent du front de Pareto par la fonction objectif.*

La figure 3.4 résume visuellement ces notions. Toutes les solutions que l'on peut considérer comme les meilleures font donc partie du front de Pareto. Réciproquement, une solution donnée du front de Pareto peut être considérée comme la meilleure selon une combinaison particulière des objectifs (pas nécessairement linéaire).

Dans cette étude, nous analysons quatre paramètres en fonction de quatre objectifs (*i.e.* les FOMs), en utilisant le formalisme des fronts de Pareto.

En utilisant quatre objectifs, le front de Pareto a quatre dimensions. Comme il n'est possible de visualiser facilement que des figures à deux dimensions, nous avons utilisé des fronts de Pareto en 2D pour chaque comparaison entre algorithmes.

3.2.6 Effet sur le calcul dosimétrique

L'histogramme dose-volume (HDV) [Shipley et al., 1979] est un outil majeur en dosimétrie. Il représente le pourcentage du volume qui a reçu plus de la dose en abscisse. Nous avons calculé les HDV à l'aide des meilleures méthodes de reconstruction telles qu'identifiées précédemment (section 3.2.4) [2], c'est-à-dire 1°) celle qui est la moins biaisée dans le cylindre froid, 2°) celle qui est la moins biaisée dans la sphère chaude de 22 mm, et 3°) celle qui a le plus petit biais dans le cylindre froid parmi les algorithmes dominant l'OSEM de référence (c.f. section 3.2.5). Les résultats pour l'OSEM de référence et pour la reconstruction du fabricant ont également été rapportés. Cette reconstruction du fabricant (appelée « SIEMENS » sur les figures et tableaux suivants) est fondée sur OP-OSEM en utilisant 2 itérations et 21 sous-ensembles. Une variante spatiale de la PSF est utilisée et l'image est post-filtrée avec un noyau gaussien de FWHM 5mm. Les HDV ont été calculées en utilisant l'approche des noyaux de convolution intégrée au logiciel fourni par Dosisoft (PLANET® Dose, Dosisoft SA, Cachan, France). Les ROIs et les sphères chaudes des patients ont été considérées comme des tissus mous, tandis que la densité du

2. Parmi celles incluant une contrainte de positivité puisque les valeurs négatives sont problématiques pour une étude dosimétrique.

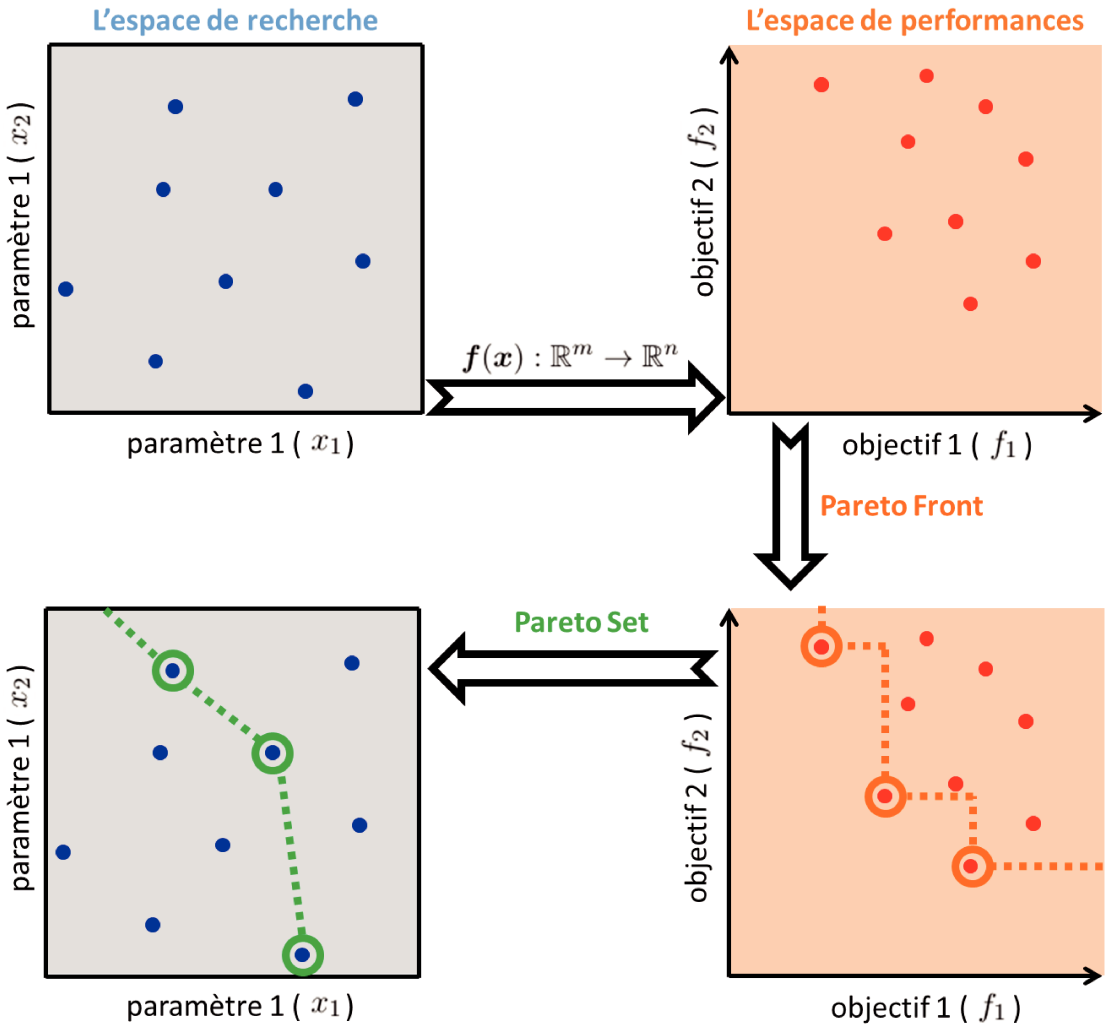
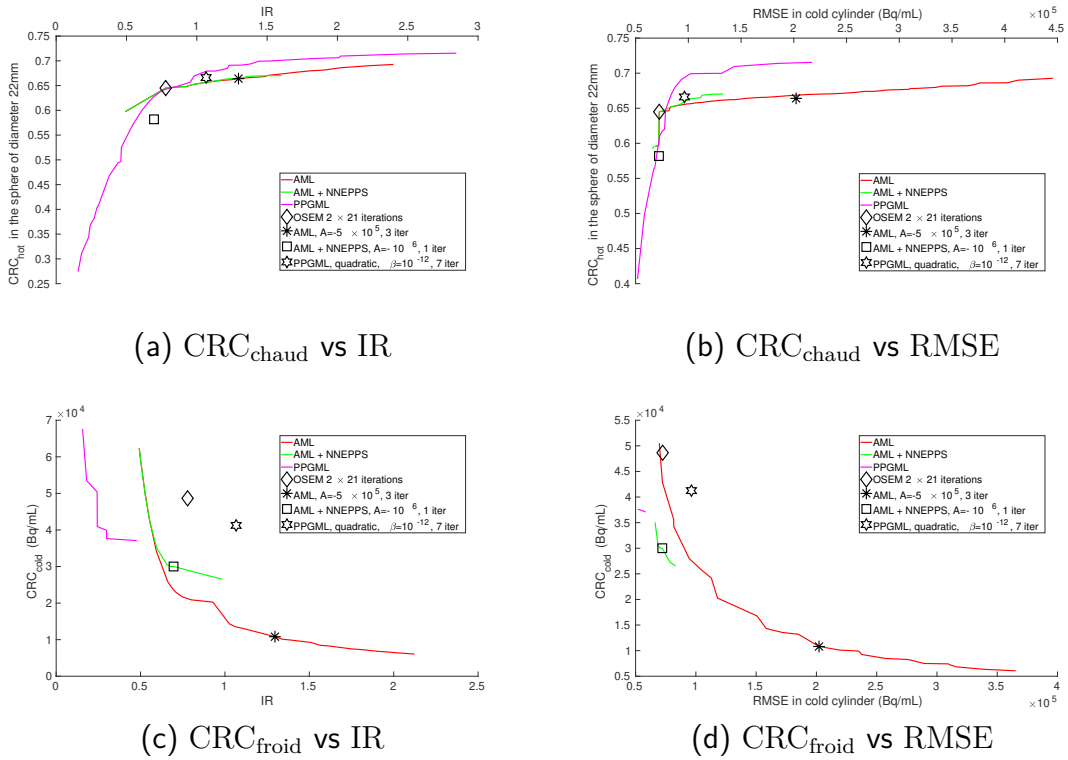


FIGURE 3.4 – Représentation du front de Pareto (à droite) et de l'ensemble de Pareto (à gauche). Image prise de [Davins-Valldaura, 2016].



(a) CRC_{chaud} vs IR

(b) CRC_{chaud} vs RMSE

(c) CRC_{froid} vs IR

(d) CRC_{froid} vs RMSE

FIGURE 3.5 – Comparaison d'AML (avec et sans le NNEPPS) et de PPGML (pénalité quadratique) en utilisant le fantôme IEC1 pour différentes métriques. Dans chaque graphique, les quatre fronts de Pareto 2D de 2 paramètres variables (le nombre d'itérations et A ou β) sont représentés. L'OSEM de référence et un exemple pour chaque autre algorithme (choisi empiriquement pour leur bon compromis) sont représentés. Notez que ces algorithmes se trouvent sur le front 4D-Pareto, mais pas nécessairement sur les fronts 2D-Pareto représentés ici.

cylindre froid a été calculée en fonction du scanner. Deux ROIs ont été considérés pour l'analyse du fantôme : la sphère chaude de 22 mm de diamètre et le cylindre froid. Afin d'éviter les effets de bord, la ROI considérée pour le cylindre froid dans les calculs dosimétriques avait un diamètre plus petit que celui présenté dans la section 3.2.4. Des HDV ont également été calculés pour le cas clinique en considérant trois régions : la tumeur, une région libre de toute activité et le foie entier (y compris la tumeur). Toutes ces régions ont été délimitées manuellement par un médecin expérimenté (plus de 8 ans de pratique clinique de la radiothérapie interne sélective). La dose moyenne délivrée (D_{avg}), la dose reçue par au moins 70 % du volume (D70), et le pourcentage du volume recevant plus de 50 Gy (V50) ont été calculés.

3.3 Résultats

Nous présentons d'abord une comparaison entre les deux méthodes non contraintes NEG-ML et AML. (section 3.3.1). Cette comparaison mène au choix d'AML plutôt que de NEG-ML. Ce choix étant fait, nous comparons AML a l'algorithme pénalisé PPGML. Nous présentons ensuite une évaluation globale de tous les paramètres, rendue possible grâce à APPGML. Nous présentons ensuite une comparaison entre la pénalité quadratique et la pénalité aux différences relatives. Ceci étant fait, nous présentons une évaluation dosimétrique, suivie d'une évaluation clinique de ces algorithmes.

3.3.1 Comparaison de NEG-ML et d'AML

La figure 3.2 montre le comportement d'AML et de NEG-ML au cours des itérations pour différentes valeurs des paramètres. Nous avons tracé les courbes de NEG-ML uniquement pour les valeurs extrêmes de ψ, mais beaucoup de valeurs intermédiaires furent testées (voir tableau 3.3 pour la liste des valeurs testées) et ont montré que ψ n'avait qu'une faible influence sur NEG-ML. En revanche, A permet à AML de passer continûment d'OSEM ($A = 0$) à un algorithme proche de NEG-ML ($A = -\infty$). Ces résultats sont cohérents avec ceux observés dans [Van Slambrouck et al., 2015].

La figure 3.3 compare AML et NEG-ML avec et sans le NNEPPS en présentant une série de fronts de Pareto pour différentes paires d'objectifs obtenus en utilisant IEC1. On notera que certains choix de paramètres attachés à AML et à NEG-ML ont donné des algorithmes presque sans biais dans le cylindre froid. Ces paramètres ont également permis une amélioration du $\mathrm{CRC}_{\mathrm{chaud}}$ d'environ 5 à 10 % par rapport à l'OSEM de référence. Ces paramètres étaient $A = -\infty$ pour AML, toutes les valeurs de ψ pour NEG-M.L, et 7 itérations (nombre maximum testé) pour les deux algorithmes. En effet, ces algorithmes conduisent à des images proches de la solution des moindres carrés de $\boldsymbol{y} = \boldsymbol{S}\boldsymbol{x} + \bar{\boldsymbol{r}}$ [Van Slambrouck et al., 2015], qui est sans biais [Lehmann and Casella, 2006]. Cependant, ces paramètres à faible biais s'accompagnent d'un niveau de bruit élevé : par rapport à l'OSEM de référence, l'IR est multipliée par un facteur de 2 à 3 et le RMSE dans le cylindre froid par environ 5.

Il est cependant possible d'identifier des paramètres offrant un bon compromis

entre le bruit et le biais. L'algorithme représenté par l'étoile sur la figure 3.3 en est un exemple. Les paramètres de biais (CRC_{chaud} et CRC_{froid}) sont toujours dans la même fourchette que les valeurs de faible biais identifiées dans le paragraphe précédent (diminution de CRC_{froid} de 78 %, et augmentation de CRC_{chaud} de 3 %), mais les caractéristiques du bruit (IR et RMSE) sont divisées par 2 environ. Ce compromis est obtenu avec AML et n'a pas pu être atteint avec NEG-ML, comme le montre la figure 3.3. Ceci est probablement dû à ce que le paramètre d'AML a plus d'influence que celui de NEG-ML et permet de trouver un compromis entre MLEM et un algorithme sans aucune contrainte relative à sa borne inférieure.

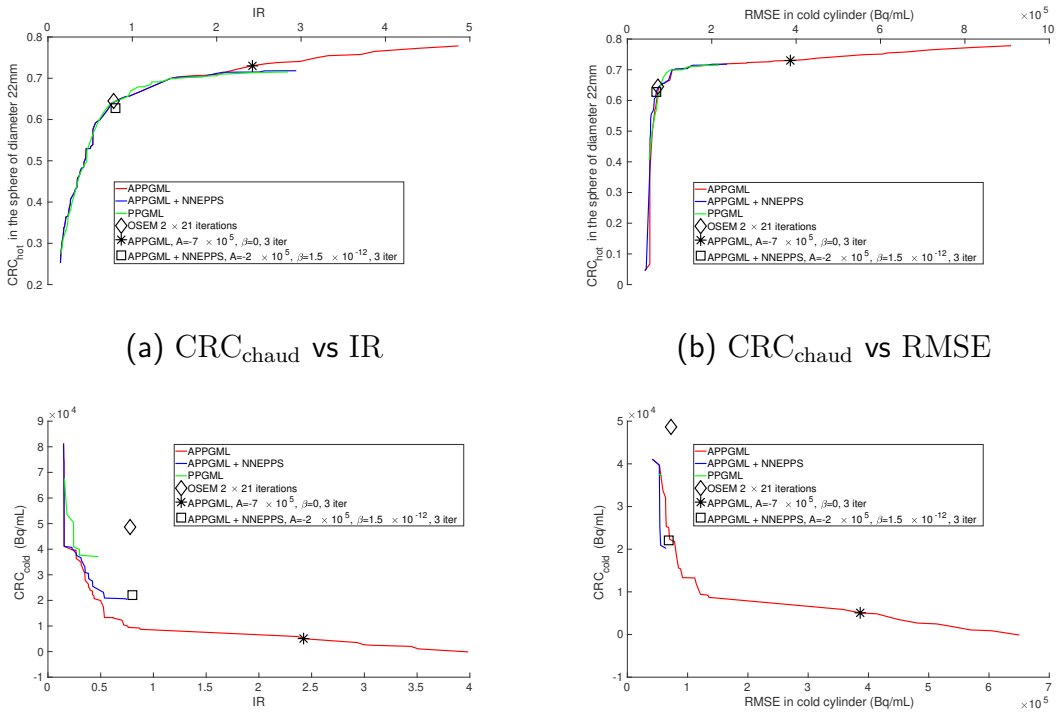
Enfin, l'utilisation d'AML combinée au NNEPPS a permis de conserver le même niveau de bruit que dans l'OSEM de référence (carrés dans la figure 3.3) et a diminué l'activité d'environ 40 % dans la région froide, au prix d'une diminution du CRC_{chaud} d'environ 10 %.

3.3.2 Comparaison d'AML et de PPGML

La figure 3.5 compare AML et PPGML pour différentes paires de métriques, en utilisant IEC1. AML et PPGML ont tous deux été capables de réduire le CRC_{froid}, mais la réduction est plus importante avec AML qu'avec PPGML. Pour un niveau de bruit similaire à l'OSEM de référence, AML + NNEPPS a offert un meilleur CRC_{froid} que PPGML. Le meilleur CRC_{chaud} a été obtenu soit en utilisant PPGML avec $\beta = 0$ (c'était notamment le cas avec IEC1 dans les sphères de 22 mm, comme le montrent les figures 3.5a et 3.5b), soit en utilisant AML, selon la taille des sphères et les acquisitions. Il est à noter que PPGML avec $\beta = 0$ revient à un OSEM sans post-fitrage. Cela montre que l'augmentation de CRC_{chaud} produite par PPGML n'est pas due à la pénalité, mais à l'absence de post-filtrage. D'autre part, pour un niveau de biais équivalent, PPGML avait généralement un niveau de bruit plus faible qu'AML. Cela suggère qu'un bon compromis entre le bruit et le biais peut être atteint en combinant pénalité et autorisation des valeurs négatives. Ces résultats nous ont donc motivés à étudier un algorithme pénalisé non contraint.

3.3.3 Évaluation globale des différents paramètres

Une évaluation simultanée de l'influence des quatre paramètres (borne inférieure d'AML, intensité de la pénalité quadratique, nombre d'itérations, utilisation ou non



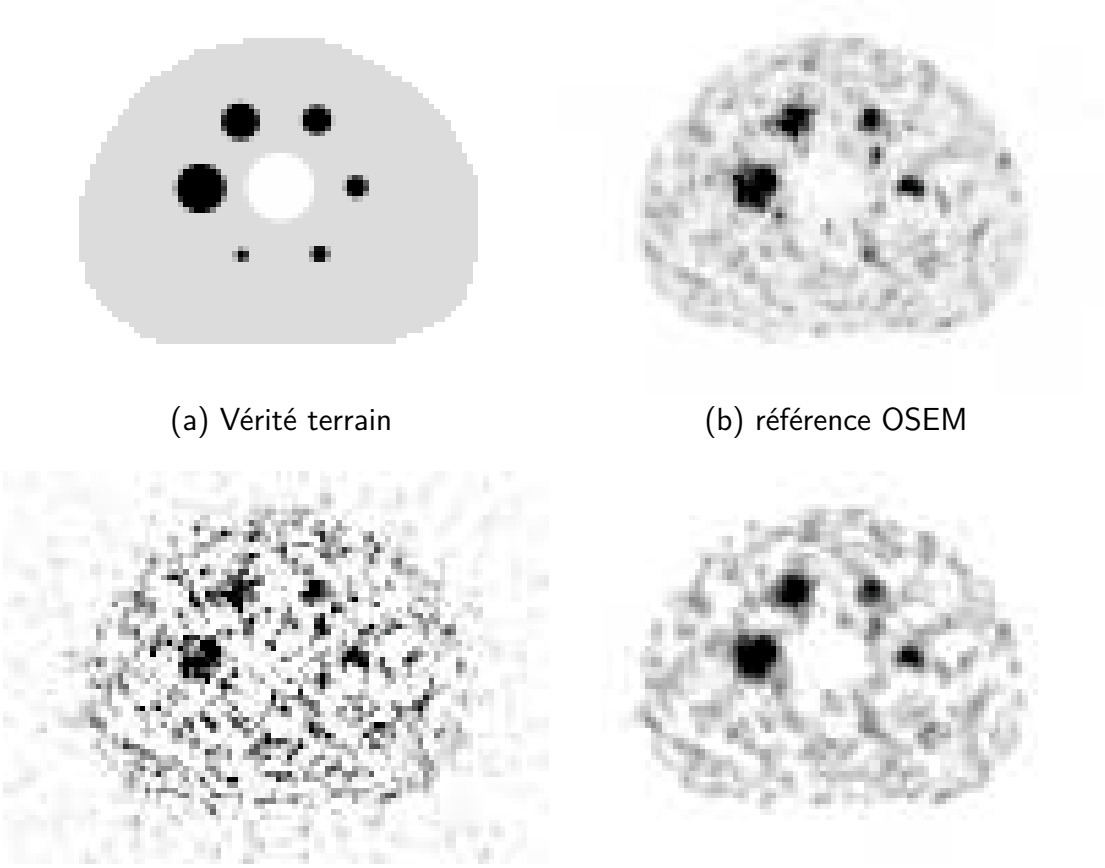
(a) CRC_{chaud} vs IR

(b) CRC_{chaud} vs RMSE

(c) CRC_{froid} vs IR

(d) CRC_{froid} vs RMSE

FIGURE 3.6 – Comparaison des quatre paramètres (intensité de la pénalité β, borne inférieure A, nombre d'itérations et inclusion ou non du NNEPPS) en utilisant le fantôme IEC1 pour différentes métriques (APPGML vs PPGML). Dans chaque graphique, les trois fronts de Pareto 2D sont représentés. Les points représentant l'OSEM de référence ainsi que deux algorithmes pertinents sont également représentés. Il est à noter que ces algorithmes se trouvent sur le front de Pareto 4D, mais pas nécessairement sur les fronts de Pareto 2D représentés ici.



(a) Vérité terrain (b) référence OSEM

(c) APPGML, $A = -7 \times 10^5$, $\beta = 0$, 3 iter (d) APPGML + NNEPPS, $A = -2 \times 10^5$, $\beta = 1.5 \times 10^{-12}$, 3 iter

FIGURE 3.7 – Images de l'IEC1 provenant des trois algorithmes mis en évidence sur la figure 3.6. Bien que les caractéristiques de biais soient bonnes pour l'algorithme de la figure 3.7c, il apparaît visuellement très bruité. Le cylindre froid central est beaucoup plus visible sur la figure 3.7d que sur l'OSEM de référence.

du NNEPPS) fut ensuite réalisée. La figure 3.6 en montre les résultats concernant IEC1.

Les meilleurs résultats en termes de biais ($\mathrm{CRC}_{\mathrm{chaud}}$ et $\mathrm{CRC}_{\mathrm{froid}}$) furent obtenus sans l'utilisation de NNEPPS et sans aucune pénalité, en utilisant une valeur négative élevée pour A et un nombre élevé d'itérations. Il est possible d'obtenir un bon compromis entre le bruit et le biais en diminuant le nombre d'itérations tout en gardant $\beta = 0$. Par exemple, APPGML utilisant $A = -7 \times 10^5\,\mathrm{Bq\,mL^{-1}}$, $\beta = 0$ (équivalent à AML sans post-filtrage), et 3 itérations (étoiles dans la figure 3.6, affiché dans la figure 3.7c) permet de réduire les métriques évaluant le bruit d'un facteur d'environ 2 par rapport aux meilleurs algorithmes en termes de biais en augmentant très peu ce dernier (figure 3.6). Enfin, pour un niveau de bruit similaire à l'OSEM de référence et un $\mathrm{CRC}_{\mathrm{chaud}}$ similaire, il est possible de diminuer significativement le biais positif (carrés dans la figure 3.6, diminuant le $\mathrm{CRC}_{\mathrm{froid}}$ de 55 %). Les images de reconstruction de ces algorithmes sont affichées sur la figure 3.7.

3.3.4 Comparaison entre la pénalité quadratique et la pénalité aux différences relatives

La figure 3.8 compare la pénalité quadratique et la pénalité aux différences relatives pour différentes combinaisons de métriques. La pénalité aux différences relatives a été la meilleure pour toutes les combinaisons de métriques, sauf pour $\mathrm{CRC}_{\mathrm{froid}}$ vs IR pour une très faible IR (*c'est-à-dire* inférieure à 64 % de celle de l'OSEM de référence : IR ≤ 0.5). Il a également été possible de trouver des paramètres dominant l'OSEM de référence en utilisant les différences relatives, par exemple en utilisant les paramètres représentés par les carrés dans la figure 3.8. Ces paramètres ont réduit le $\mathrm{CRC}_{\mathrm{froid}}$ de 34 % et le RMSE de 19 %, tout en augmentant le $\mathrm{CRC}_{\mathrm{chaud}}$ de 7 % et en préservant l'IR. La réduction possible tout en préservant toutes les autres métriques était du même ordre de grandeur pour toutes les acquisitions. Cependant, il n'a pas été possible de trouver des valeurs de paramètres qui dominaient l'OSEM de référence en utilisant la pénalité quadratique, comme cela est clairement visible sur la figure 3.8b. Pour ces raisons, la fonction potentielle aux différences relatives semble préférable à la fonction quadratique.

3.3.5 Évaluation dosimétrique

Les figures 3.9 et 3.10 montrent respectivement l'HDV du cylindre froid et de la sphère chaude de diamètre 22 mm pour l'ensemble de données IEC0. À des fins de comparaison, les HDV d'une reconstruction parfaite sont également données. Celles-ci furent obtenues en créant une image théorique haute résolution en sous-échantillonnant la matrice de l'image TEP (4 mm × 4 mm × 2 mm). L'HDV de la sphère 22 mm ne présentait pas une forme carrée parfaite en raison notamment de la taille du voxel et du noyau de convolution utilisé pour le calcul de l'HDV. Les résultats dosimétriques pour la région froide du fantôme IEC0 sont détaillés dans le tableau 3.1.

On peut en déduire qu'APPGML + NNEPPS présente les meilleurs résultats quantitatifs dans le cylindre froid, en divisant la dose moyenne par 7 par rapport à l'OSEM de référence, et D70 par 42. Dans la sphère chaude, V50 est également amélioré par rapport à l'OSEM de référence, mais la contrepartie est que D_{avg} et D70 sont réduits respectivement de 16 % et 11 % par rapport à l'OSEM de référence, qui sous-estime déjà ces métriques. PPGML avec une fonction de potentiel de différences relatives s'est révélé meilleur que l'OSEM de référence pour chaque paramètre, à l'exception de V50 dans la sphère chaude, mais la différence n'était que de 0,7 %. Comme prévu, AML + NNEPPS a obtenu la meilleure valeur de D_{avg} dans la sphère chaude, mais la variance dans cette région était trop élevée, comme le montre la figure 3.10, où l'on observe que la courbe est la plus plate. À cause de cette variance trop élevée, AML + NNEPPS sous-estime D70 et V50.

3.3.6 Évaluation clinique

Les figures 3.11, 3.12 et 3.13 montrent différents HDV pour le cas clinique. Ils montrent respectivement les HDV pour une région supposée libre de toute activité, pour le foie entier (y compris les tumeurs) et pour une tumeur. Les mêmes méthodes de reconstruction que pour le fantôme sont représentées. Comme le montre la figure 3.11, le signal lié au bruit est réduit en utilisant APPGML + NNEPPS dans la région supposée exempte de toute activité, et qualitativement réduit en considérant les autres régions. Les figures 3.11, 3.12 et 3.13 tendent à confirmer les résultats exposés dans la section 3.3.5. En comparant les figures 3.9 et 3.11 avec les figures 3.10 et 3.13, on observe que les courbes sont à peu près dans le même ordre

dans le fantôme et dans la zone correspondante du patient. La comparaison entre les tableaux 3.1 et 3.4 indique que les algorithmes présentant de bonnes performances dans la zone froide du fantôme présentent également une faible activité dans la région du patient censée être sans activité. Par exemple, APPGML + NNEPPS et PPGML sont les deux algorithmes présentant le plus petit D_{avg} dans le cylindre froid, et ce sont également les deux algorithmes présentant le plus petit D_{avg} dans la région sans activité du patient. Cela semble confirmer qu'APPGML + NNEPPS est le meilleur algorithme dans les régions à faible activité parmi celles utilisables pour la dosimétrie. De la même manière, les résultats pour la tumeur du patient confirment ceux de la sphère chaude du fantôme, où les deux D_{avg} les plus élevés sont AML + NNEPPS et PPGML pour les deux. La figure 3.14 montre une image du patient pour chacun des cinq algorithmes précédents. La différence de niveau de bruit est clairement visible sur ces images. AML + NNEPPS est trop bruité pour être interprétable. PPGML est moins bruité qu'OSEM, et APPGML + NNEPPS est encore moins bruité. La figure 3.15 présente une image du même patient avec une carte de couleurs saturée. Le biais positif du cœur présent dans OSEM est supprimé avec APPGML + NNEPPS.

3.4 Discussion et conclusion

L'imagerie TEP post-thérapeutique à l'yttrium 90 est principalement utilisée pour mesurer l'activité délivrée dans une région donnée en vue d'un calcul dosimétrique ultérieur [Pasciak et al., 2014]. Par conséquent, les images reconstruites doivent être quantitativement précises et exactes plutôt que visuellement réalistes. À cet égard, et puisqu'aucune image reconstruite ne fournit la valeur optimale pour l'ensemble des métriques, un niveau supérieur à celui trouvé à l'aide de l'OSEM de référence pourrait être toléré pour les métriques de bruit (IR et RMSE), et l'étude pourrait se concentrer préférentiellement sur les métriques de biais (CRC_{froid} et CRC_{chaud}). Il est cependant important que l'IR et le RMSE restent raisonnables à des fins de détectabilité. Ils influencent également certaines métriques dosimétriques, telles que D70 et V50.

Une comparaison entre deux algorithmes non contraints, AML et NEG-ML, a montré que si le paramètre A d'AML permettait une correspondance continue entre OSEM (un cas particulier d'AML lorsque $A = 0$) et AML avec $A = -\infty$, NEG-ML

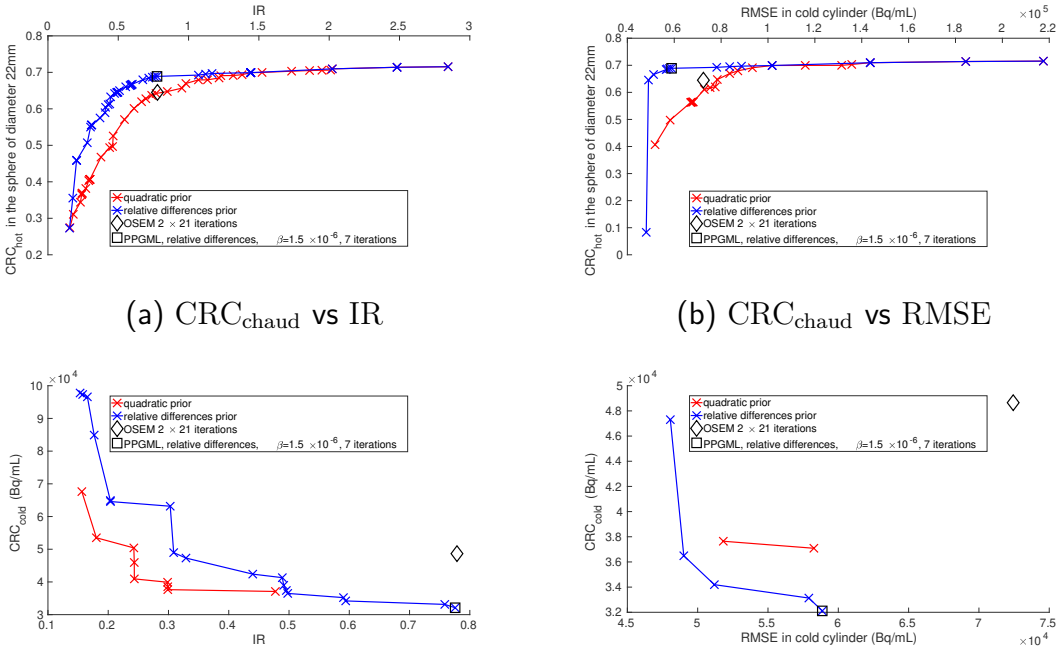
Table 3.1 – Métriques dosimétriques du fantôme IEC0

	Cylindre froid		**Sphère chaude** (22 mm)		
	Dose moyenne (Gy)	**Dose à 70 % (Gy)**	**Dose moyenne (Gy)**	**Dose à 70% (Gy)**	**Volume supérieur à 50 Gy (%)**
SIEMENS	9.5	4.8	99.4	74.1	94.0
OSEM	7.6	4.2	94.6	75.3	83.6
AML + NNEPPS	5.3	0.2	106.9	66.0	82.4
APPGML + NNEPPS	1.1	0.1	79.7	66.9	90.9
PPGML	3.3	0.8	100.3	78.4	83.0
Perfect image	0	0	105.4	100.1	97.6

est très proche de la version d'AML avec $A = -\infty$, quelle que soit la valeur de ψ, et est donc beaucoup moins flexible. Par conséquent, NEG-ML a présenté des IR et des RMSE plus grands que l'OSEM de référence de la première itération. Avec AML, le biais positif dans la région froide a été légèrement augmenté par l'utilisation du NNEPPS, mais le bruit a été réduit. Cela permet d'utiliser un $|A|$ plus grand, ce qui entraîne une plus grande réduction du biais sans augmenter le bruit.

Dans cette étude, les algorithmes pénalisés et non contraints ont permis de réduire le biais positif dans le cylindre froid, comme le montre le tableau 3.2, qui résume les caractéristiques de chaque algorithme. Les algorithmes sans contrainte ont obtenu la plus forte diminution, mais au détriment d'au moins une des autres métriques étudiées. Ils peuvent en fait être presque sans biais, mais avec un niveau de bruit élevé. Inversement, pour un même niveau de bruit, PPGML est généralement moins biaisé. Cela justifie l'utilisation d'algorithmes non contraints pénalisés (APPGML dans cette étude).

APPGML a un biais beaucoup plus faible que les autres algorithmes pour un niveau de bruit similaire. Par exemple, en utilisant le fantôme IEC1, et pour un RMSE similaire à l'OSEM de référence, AML permet une réduction du $\mathrm{CRC}_{\mathrm{froid}}$ de 12 %, PPGML avec une pénalité quadratique de 24 %, AML + NNEPPS de 38 %, APPGML de 54 %, et APPGML + NNEPPS de 59 %.



(a) CRC_{chaud} vs IR

(b) CRC_{chaud} vs RMSE

(c) CRC_{froid} vs IR

(d) CRC_{froid} vs RMSE

FIGURE 3.8 – Comparaison des différences relatives et des pénalités quadratiques dans le fantôme IEC1 pour différentes métriques. Dans chaque graphique, les croix représentent le front de Pareto 4D de l'ensemble des algorithmes contenant à la fois des différences relatives et des pénalités quadratiques. Les lignes représentent les ensembles de Pareto 2D (correspondant au graphique) des différences relatives et de la pénalité quadratique. Pour chaque ligne, il existe deux paramètres variables (le nombre d'itérations et la force de la pénalité β)

.

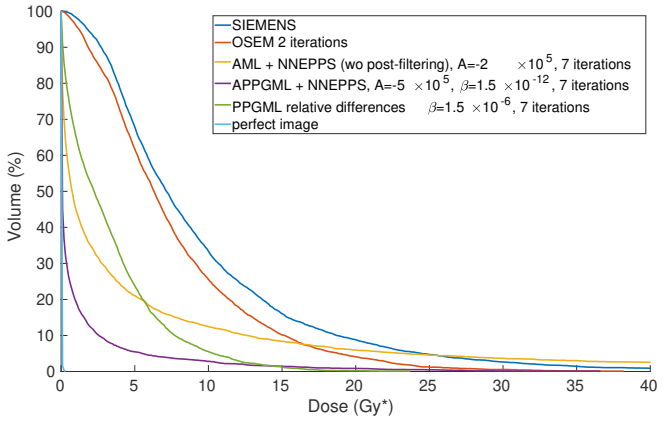


FIGURE 3.9 – Histogramme dose-volume d'une région d'intérêt à l'intérieur du cylindre froid pour IEC0. « SIEMENS » représente l'algorithme mis en œuvre par le fabricant.

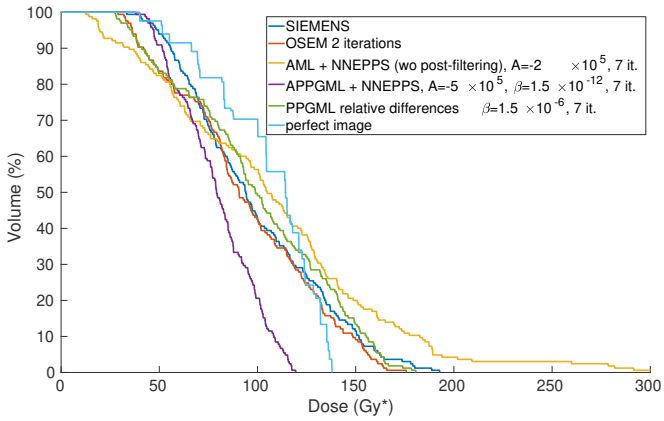


FIGURE 3.10 – Histogramme dose-volume de la sphère de 22 mm pour IEC0. SIEMENS représente l'algorithme mis en œuvre par le fabricant.

TABLE 3.2 – Tableau récapitulatif des caractéristiques de chaque algorithme sur le fantôme IEC1[12]

Algorithme	Meilleur RMSE ($\times 10^4$)	Meilleure IR	Meilleur CRC_{froid} ($\times 10^4$)	Meilleur CRC_{chaud}	Meilleur CRC_{froid} pour RMSE $\leq$ RMSE(Ref) ($\times 10^4$)	Meilleur CRC_{froid} pour IR $\leq$ IR(Ref) ($\times 10^4$)	Meilleur CRC_{froid} pour RMSE $\leq$ RMSE(Ref) and IR $\leq$ IR(Ref) ($\times 10^4$)	Meilleur CRC_{froid} pour RMSE $\leq$ RMSE(Ref), IR $\leq$ IR(Ref) et $CRC_{chaud} \geq CRC_{chaud}$(Ref) ($\times 10^4$)
OSEM de référence (post-filtré)	7.2	0.78	4.9	0.64	4.9	4.9	4.9	4.9
OSEM (post-filtré)	7.2	0.49	4.5	0.66	4.9	4.9	4.9	4.9
AML (post-filtré)	7.0	0.49	0.60	0.69	4.3	2.2	4.3	4.9
AML (post-filtré) + NNEPPS	6.6	0.49	2.7	0.67	3.0	3.0	3.0	4.9
NEG-ML (post-filtré)	11	0.76	0.56	0.71	✗	2.3	✗	✗
NEG-ML (post-filtré) + NNEPPS	8.2	0.72	2.9	0.68	✗	3.2	✗	✗
PPGML quadratique	5.2	0.16	3.7	0.72	3.7	3.7	3.7	✗
PPGML differences relatives	4.8	0.15	3.2	0.72	3.2	3.2	3.2	**3.2**
APPGML quadratique	**4.1**	**0.15**	**0.012**	**0.78**	2.2	**0.94**	2.2	✗
APPGML quadratique + NNEPPS	**4.1**	**0.15**	2.0	0.72	**2.0**	2.0	**2.0**	✗

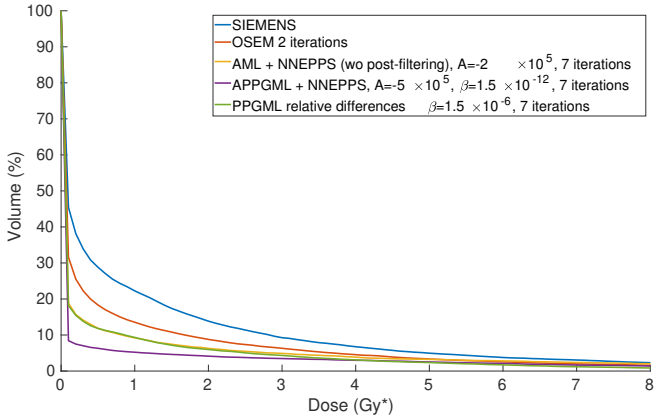


FIGURE 3.11 – HDV du patient pour la région d'intérêt censée être libre de toute activité. SIEMENS représente l'algorithme mis en œuvre par le fabricant.

APPGML + NNEPPS est en effet l'algorithme réalisant la plus forte réduction du $\mathrm{CRC}_{\mathrm{froid}}$ tout en préservant l'IR, le $\mathrm{CRC}_{\mathrm{chaud}}$ et le RMSE à peu près au même niveau que l'OSEM de référence. Le choix optimal de A et de β doit être ajusté en fonction des propriétés statistiques. Dans nos expériences, il varie de $A = -10^5\,\mathrm{Bq\,mL^{-1}}$ et $\beta = 10^{-12}$ pour IEC0 à $A = -5 \times 10^5\,\mathrm{Bq\,mL^{-1}}$ et $\beta = 2 \times 10^{-12}$ pour IEC3. C'est APPGML qui procure le biais le plus faible (à la fois en termes de $\mathrm{CRC}_{\mathrm{chaud}}$ et de $\mathrm{CRC}_{\mathrm{froid}}$), avec $A = -\infty$, $\beta = 0$, le plus grand nombre d'itérations testées, et sans le NNEPPS. Bien que les algorithmes pénalisés aient un biais beaucoup plus faible que les algorithmes non pénalisés arrêtés précocement pour un niveau de bruit comparable à l'OSEM de référence, ces derniers peuvent être aussi bons que les algorithmes pénalisés lorsqu'un compromis entre le faible niveau de bruit de l'OSEM de référence et le faible biais de l'AML convergé est recherché. Par exemple, les étoiles de la figure 3.6 ont été obtenues en utilisant un algorithme non pénalisé après 3 itérations.

En utilisant le RMSE sur des HDV pour une évaluation dosimétrique, Siman *et al.* [Siman et al., 2019] ont montré que 36 itérations équivalentes d'OSEM avec un post-filtrage de 5,2 mm était idéal. Sur la base de leurs travaux, nous avons défini une méthode de référence de 42 itérations équivalentes d'OSEM (2 itérations de 21 sous-ensembles), avec un post-filtrage de 4,5 mm, et avons comparé les résultats des algorithmes à cette référence. Les résultats dosimétriques dans le cylindre froid (figure 3.9) ont confirmé l'avantage d'utiliser des méthodes pénalisées par rapport à

OSEM. APPGML + NNEPPS a obtenu les meilleures performances dans les régions froides.

Les algorithmes pénalisés utilisant les différences relatives ont montré une diminution du biais moins impressionnante que les algorithmes non contraints, mais avec une amélioration de toutes les métriques étudiées par rapport à l'OSEM de référence. Un résultat important est que la valeur optimale de l'intensité de la pénalité lors de l'utilisation des différences relatives s'est avérée être presque indépendante du niveau de statistique (environ $1{,}5 \times 10^{-6}$). Sans l'utilisation de la fonction potentielle des différences relatives, il était impossible d'améliorer les quatre métriques en même temps par rapport à l'OSEM de référence. Par exemple, il était impossible d'améliorer à la fois l'IR et le $\mathrm{CRC}_{\mathrm{chaud}}$ avec une pénalité quadratique, même en choisissant une borne inférieure négative. Cela suggère que les différences relatives sont préférables à une fonction potentielle quadratique. Cependant, l'utilisation des différences relatives avec des algorithmes non contraints n'est pas triviale en raison des valeurs négatives impliquées. Par conséquent, APPGML a été implémenté avec une pénalité quadratique et un développement spécifique est nécessaire pour étendre la fonction potentielle des différences relatives aux algorithmes non contraints. Une autre perspective serait de tester, avec et sans NNEPPS, l'algorithme proposé par Lim *et al.*, et revisité par Bousse *et al.* [Lim et al., 2018, Bousse et al., 2020], qui impose la positivité dans l'espace des données pour maintenir la vraie distribution de Poisson tout en autorisant les valeurs négatives dans l'image.

Nos résultats sont en accord avec les études précédentes fondées sur des données réelles yttrium 90 utilisant la version commerciale de BSREM avec une pénalité aux différences relatives [Rowley et al., 2017, Seo et al., 2019, Scott and McGowan, 2019]. Ces études ont également montré une amélioration de la qualité de l'image, en termes de variabilité du fond et de rapport signal sur bruit dans les zones chaudes, par rapport à l'OSEM standard. Ces études indiquent que l'intensité de la pénalité doit être différente, selon que la finalité soit qualitative (liée à la visualisation) ou quantitative. Seo *et al.* [Seo et al., 2019] ont également signalé une augmentation de la récupération du contraste par rapport à l'OSEM, ce qui est en accord avec nos résultats.

12. Notez que même pour un algorithme donné, chaque colonne provient d'un choix de paramètres différent. Code de la carte des couleurs. Vert : plus proche de la meilleure valeur de la colonne que de la référence. Orange : meilleure (ou égale) que la référence, mais plus proche de la référence que de la meilleure valeur de la colonne. Rouge : pire que la référence. ✗ : aucun paramètre trouvé

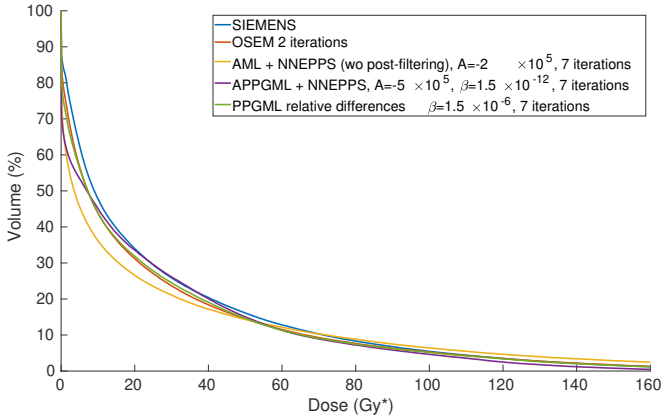


FIGURE 3.12 – HDV dans le foie en utilisant différents algorithmes. SIEMENS représente l'algorithme mis en œuvre par le fabricant.

Nos résultats suggèrent qu'APPGML + NNEPPS est idéal à la fois pour la dosimétrie et la visualisation. En effet, alors qu'une faible dose (souvent inférieure à 10 Gy) est attendue dans le foie sain, il n'est pas rare que plusieurs sessions de traitement soient nécessaires pour un patient spécifique, conduisant à une dose cumulée autour de la limite théorique pour les microsphères en résine (40 Gy [Levillain et al., 2021]). Cependant, la valeur des hyper-paramètres doit être ajustée pour correspondre aux objectifs précis de l'étude (visualisation, dosimétrie de la zone froide et/ou de la zone chaude, etc. Une solution pourrait être de reconstruire deux images, par exemple, l'une pour la visualisation, l'autre pour la quantification.

L'utilisation du RMSE sur les HDV comme métrique a permis à Siman *et al.* [Siman et al., 2019] d'optimiser une combinaison de bruit et de biais pour l'algorithme OSEM. Plutôt que d'optimiser une seule métrique, nous avons considéré simultanément quatre métriques dans une étude multi-objectifs. Cette analyse était réalisable grâce à l'utilisation des fronts de Pareto qui permettent de traiter simultanément quatre objectifs et quatre paramètres. Alors que ce formalisme est intensivement utilisé dans d'autres domaines, tels que l'ingénierie [Davins-Valldaura et al., 2017,Bompard et al., 2010], et que plusieurs études ont utilisé des métriques de quantification par rapport au bruit, à notre connaissance seule une étude a récemment appliqué ce formalisme à la TEP [Abouhawwash and Alessio, 2021] dans un contexte différent.

satisfaisant les contraintes.

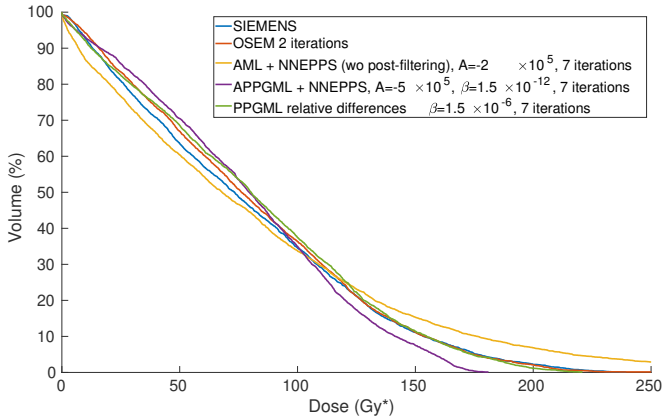
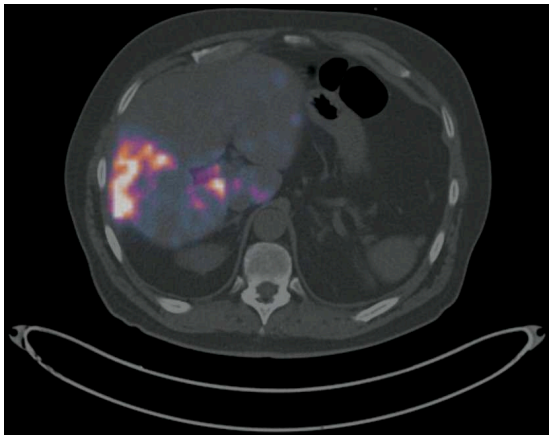


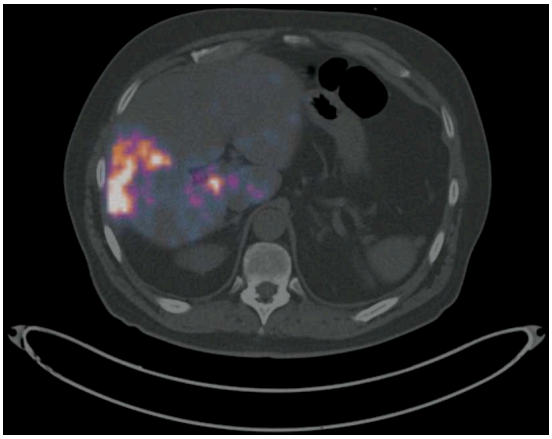
FIGURE 3.13 – HDV dans une tumeur en utilisant différents algorithmes. SIEMENS représente l'algorithme mis en œuvre par le fabricant..

TABLE 3.3 – Liste de tous les paramètres testés

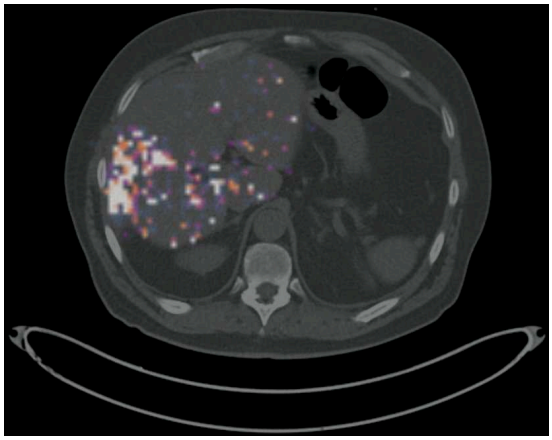
Intensité de la pénalité β (quadratique)	0, 10^{-13}, 3×10^{-13}, 5×10^{-13}, 10^{-12}, 1.5×10^{-12}, 2×10^{-12}, 4×10^{-12}, 8×10^{-12}, 2×10^{-11}, 3×10^{-11}, 5×10^{-11}, 6×10^{-11}, 8×10^{-11}, 10^{-10}
Intensité de la pénalité β (différences relatives)	0, 10^{-6}, 1.5×10^{-6}, 2×10^{-6}, 2.5×10^{-6}, 3×10^{-6}, 5×10^{-6}, 10^{-5}, 2×10^{-5}, 4×10^{-5}, 8×10^{-5}
Borne inférieure A ($\mathrm{Bq\,mL^{-1}}$)	0, -10^4, -5×10^4, -10^5, -2×10^5, -5×10^5, -7×10^5, -10^6, -2×10^6, -10^7 (seulement pour APPGML), $-\infty$ (seulement pour AML)
paramètre de NEG-ML ψ	1.4013×10^{-45}, 10^{-40}, 10^{-35}, 10^{-30}, 10^{-25}, 10^{-20}, 10^{-15}, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 0.015, 0.02, 0.025, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 1, 10, 10^2, 10^3, 10^4, 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{15}, 10^{20}, 10^{25}, 10^{30}, 10^{35}, 3.4028×10^{38}, ∞
nombre d'itérations	1, 2, 3, 4, 5, 6, 7



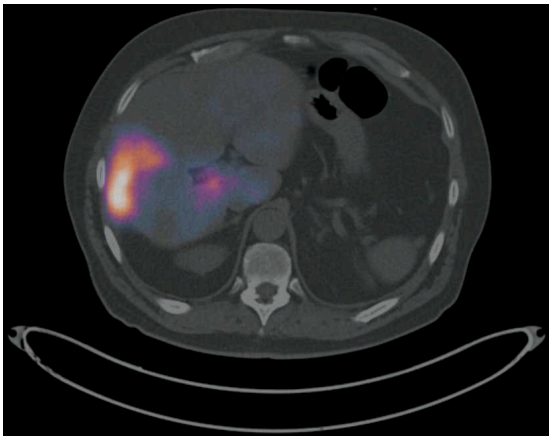
(a) SIEMENS



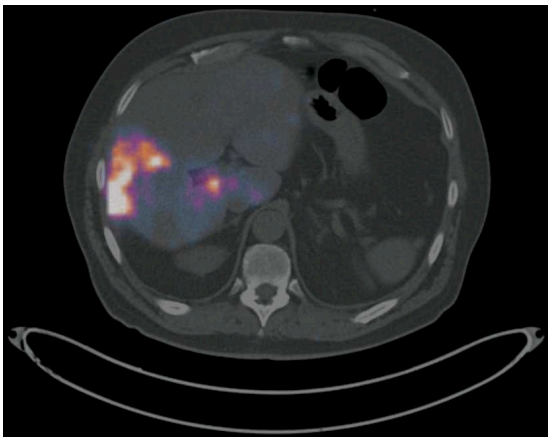
(b) OSEM 2 iterations



(c) AML + NNEPPS (sans post-filtrage), $A = -2 \times 10^5$, 7 iterations

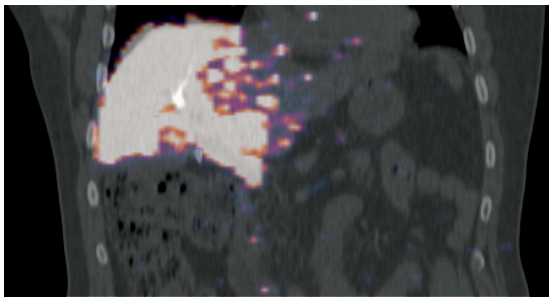


(d) APPGML + NNEPPS, $A = -5 \times 10^5$, $\beta = 1.5 \times 10^{-12}$, 7 iterations

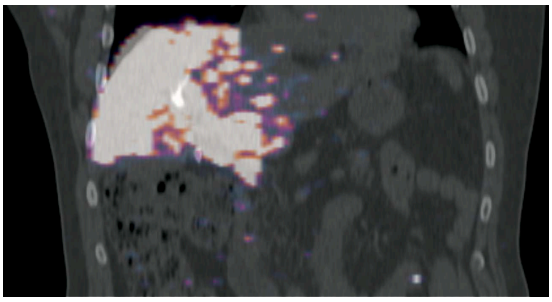


(e) PPGML differences relatives, $\beta = 1.5 \times 10^{-6}$, 7 iterations

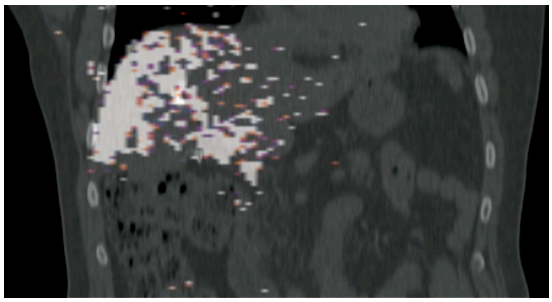
FIGURE 3.14 – Image de fusion (TEP/tomodensitométrie) d'un patient avec cinq algorithmes de reconstruction différents



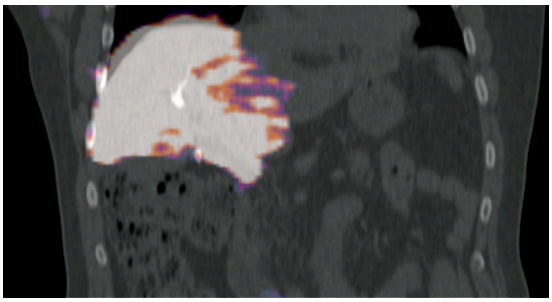
(a) SIEMENS



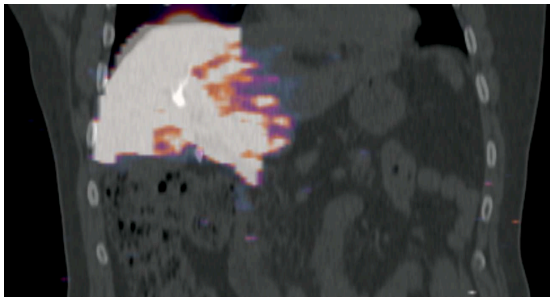
(b) OSEM 2 iterations



(c) AML + NNEPPS (sans post-filtrage), $A = -2 \times 10^5$, 7 iterations



(d) APPGML + NNEPPS, $A = -5 \times 10^5$, $\beta = 1.5 \times 10^{-12}$, 7 iterations



(e) PPGML différences relatives, $\beta = 1.5 \times 10^{-6}$, 7 iterations

FIGURE 3.15 – Image de fusion (TEP/tomodensitométrie) d'un patient avec cinq algorithmes de reconstruction différents. L'échelle de couleur est très saturée afin de mettre en évidence le biais positif de certains algorithmes. Certaines taches, attribuées au bruit, sont visibles dans les images OSEM et SIEMENS,et sont absentes des images APPGML + NNEPPS et PPGML

Table 3.4 – Métriques dosimétriques du patient

	Zone froide		**Foie entier**			**Tumeur**		
	Dose moyenne (Gy)	**Dose à 70 % (Gy)**	**Dose moyenne (Gy)**	**Dose à 70% (Gy)**	**Volume supérieur à 50 Gy (%)**	**Dose moyenne (Gy)**	**Dose à 70 % (Gy)**	**Volume supérieur à 50 Gy (%)**
SIEMENS	1.0	0.1	24.5	3.4	16.1	79.3	41.4	63.6
OSEM	0.6	0	22.6	2.1	14.5	81.7	46.2	67.0
AML + NNEPPS	0.9	0	22.8	0.4	14.3	82.6	34.3	60.4
APPGML + NNEPPS	0.3	0	21.9	0.2	15.1	79.4	50.9	70.5
PPGML	0.4	0	22.5	1.5	14.6	82.4	47.8	68.6

Chapitre 4

Utilisation de l'apprentissage profond pour le réglage des hyper-paramètres en reconstruction itérative

Sommaire

4.1 Introduction 123

4.2 Méthodologie générale 123

4.2.1 Présentation de l'approche 123

4.2.2 Constitution de la base de donnée pour l'apprentissage supervisé 125

4.2.2.1 Création des images « vérité terrain » 125

4.2.2.2 Projection de l'image pour créer un sinogramme bruité 125

4.2.2.3 Choix des objectifs 128

4.2.2.4 Optimisation des hyper-paramètres 129

4.2.3 Configuration expérimentale 129

4.2.4 Architecture 130

4.2.5 Entraînement du réseau 131

4.2.6 Méthodes « hyper-paramètres fixés » 131

4.3 Résultats 132

4.3.1 Validation croisée 132

4.3.2 Corrélation entre les variables 133

4.3.3 Analyse et discussion 136

4.3.3.1 Perspectives 136

4.3.3.2 Étude préliminaire de certaines perspectives . . 139
4.3.3.3 Formulation du problème 140
4.4 Conclusion 142

4.1 Introduction

Nous avons vu dans le chapitre précédent qu'un choix judicieux des hyper-paramètres de reconstruction pouvait mener à une amélioration qualitative et quantitative des images reconstruites. Cependant, le choix précis des hyper-paramètres dépend du problème traité et par conséquent il n'existe pas de valeur optimale unique qui s'appliquerait universellement à tous les patients.

Certains auteurs ont proposé de fixer l'intensité de la pénalité au cours de la reconstruction, par validation croisée [Reader and Ellis, 2020]. Cependant, ces méthodes sont souvent coûteuses et ne permettent pas de fixer les autres hyper-paramètres de reconstruction, telles que la borne inférieure d'AML. Dans ce chapitre, nous cherchons à savoir s'il est possible d'estimer les paramètres de reconstruction optimaux avant de commencer la reconstruction, à partir des seuls sinogrammes (données brutes) en utilisant une approche exploitant les techniques d'apprentissage statistique.

Depuis quelques années, l'apprentissage profond a une influence grandissante dans de nombreux domaines, incluant les problèmes inverses [Arridge et al., 2019]. En reconstruction TEP [Reader et al., 2021, Carrillo et al., 2021], il est utilisé à diverses fins (voir la section 1.1.7.4), mais n'a, à notre connaissance, jamais été proposé dans un but de sélection des hyper-paramètres de reconstruction.

Il s'agit pourtant là d'une approche qui pourrait être particulièrement utile, car il est difficile de modéliser théoriquement la valeur que doivent prendre les hyper-paramètres pour aboutir à une quantification optimale des images.

Ce chapitre est consacré à un travail prospectif dans cette direction. Il annonce des résultats préliminaires et propose des pistes à explorer lors d'un travail futur. Il présente également certains outils et méthodologies qui ont été développés et qui pourront être utiles pour la suite du travail, tels que la création d'une base de données présentant des couples sinogrammes-hyper-paramètres.

4.2 Méthodologie générale

4.2.1 Présentation de l'approche

L'approche proposée consiste à utiliser un réseau de neurones prenant en entrée les données brutes (cinq sinogrammes, c'est-à-dire la totalité des informations dispo-

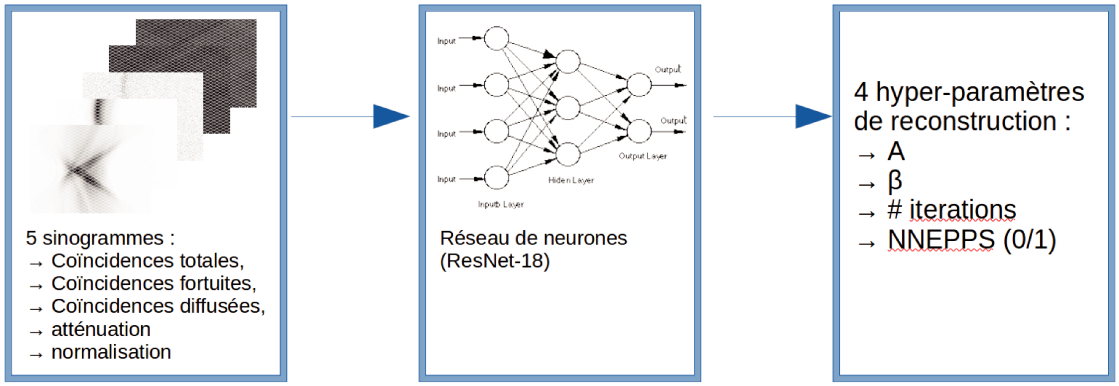


FIGURE 4.1 – Récapitulatif des entrées et sorties de l'algorithme proposé.

nibles avant la reconstruction) et renvoyant le jeu d'hyper-paramètres à utiliser lors de la reconstruction. Les cinq sinogrammes en entrée du réseau sont les sinogrammes des coïncidences totales, des coïncidences aléatoires, des coïncidences diffusées, d'atténuation et de normalisation. En cohérence avec la partie précédente, nous utiliserons l'algorithme APPGML pour nos reconstructions, et les quatre hyper-paramètres considérés sont ceux de la section précédente : la borne inférieure A, l'intensité de la pénalité β, le nombre d'itérations, et un booléen indiquant si le NNEPPS doit être appliqué ou pas. La figure 4.1 récapitule les entrées et sorties du problème traité dans cette partie.

Comme nous l'avons vu lors de la partie 3, le jeu d'hyper-paramètres à utiliser dépend de l'application considérée, et nous proposons donc d'entraîner le réseau différemment selon l'objectif à atteindre. Nous entraînons donc une architecture unique selon deux objectifs : la MSSIM (*mean structural similarity*) [Wang et al., 2004], et une combinaison des quatre métriques utilisées dans la section précédente. Ces objectifs seront détaillés dans la section 4.2.2.3.

La section 4.2.2 décrit la méthodologie utilisée pour créer la base de données. La section 4.2.3 indique la configuration expérimentale utilisée. L'architecture du réseau et les paramètres utilisés pour son entraînement sont détaillés respectivement dans les sections 4.2.4 et 4.2.5. Enfin, la section 4.2.6 décrit des méthodes simples dont les performances peuvent être comparées à celles de notre réseau.

4.2.2 Constitution de la base de donnée pour l'apprentissage supervisé

4.2.2.1 Création des images « vérité terrain »

Afin d'entraîner un réseau de neurones par apprentissage supervisé, la première étape était de constituer une base de données contenant les sinogrammes que prendrait le réseau de neurones en entrée et les hyper-paramètres de reconstruction optimaux associés à ces sinogrammes. Afin de commencer par un problème simple, nous avons fait le choix d'utiliser des images 2D associées à un sinogramme unique. Nous avons d'abord créé des images de référence (« vérité terrain »), avant d'en simuler un sinogramme. Les simulations apportent en effet cet avantage, par rapport aux acquisitions réelles de patients, que la vérité terrain est connue et peut donc servir à l'évaluation de la reconstruction.

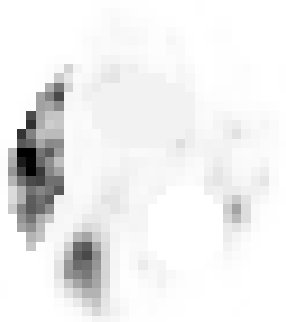
Une base de données de 112 images a été constituée, dont chaque image fut réalisée de la manière suivante :

1. Sélection d'une coupe issue d'une image TEP yttrium 90 réelle d'un patient.
2. Adjonction d'une tumeur issue d'un autre patient dans le but d'y calculer le $\mathrm{CRC}_{\mathrm{chaud}}$. La tumeur à ajouter fut sélectionnée aléatoirement et multipliée par une constante afin de régler son contraste par rapport au fond diffus. Le contraste fut choisi aléatoirement dans une distribution gaussienne correspondant à la distribution réelle du contraste sur le patient dont est extrait le fond.
3. Adjonction d'une zone d'activité nulle, dessinée à la main. Cette zone est ajoutée afin d'y pouvoir estimer le $\mathrm{CRC}_{\mathrm{froid}}$ et le RMSE.
4. Ajout d'une zone uniforme afin d'y calculer l'IR.

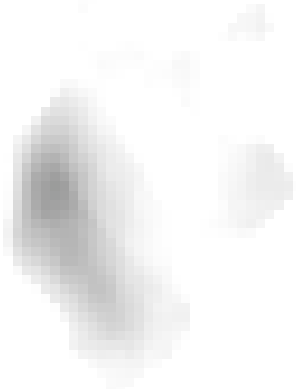
La figure 4.2a montre un exemple d'image vérité terrain ainsi créée.

4.2.2.2 Projection de l'image pour créer un sinogramme bruité

Une fois les images créées, les sinogrammes correspondants sont générés grâce à un code de simulation analytique. Les sinogrammes créés correspondent à des acquisitions du Biograph mMR de chez Siemens. La référence [Stute et al., 2015] présente ce code de façon détaillées et nous donnons ci-dessous une brève description.

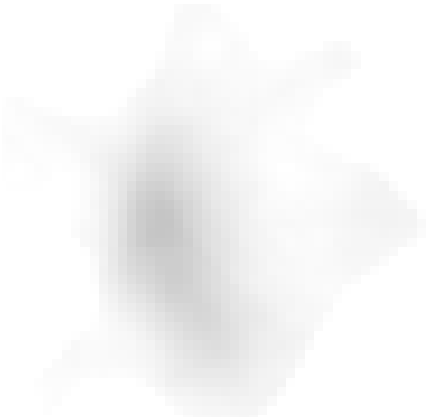


(a) Image vérité terrain



(b) Optimisation MSSIM

(c) Optimisation « métriques »



(d) Prédiction MSSIM du réseau

(e) Prédiction « métriques » du réseau

FIGURE 4.2 – Exemple d'image de la base de donnée ainsi que les reconstructions optimales et prédites par le réseau. On voit sur l'image vérité terrain une tumeur, une zone uniforme et une zone froide.

Ce code de simulation analytique prend en entrée le nombre total de coïncidences à simuler, la fraction de coïncidences diffusées, la fraction de coïncidences fortuites, ainsi que la fraction de ces coïncidences fortuites liées au fond diffus que cause le LSO (voir la section 1.2.4). Il prend également en entrée l'image vérité terrain et la carte d'atténuation associée, ainsi que la valeur de la PSF dans l'image. Voici les opérations successives que le code de simulation applique :

1. Pour chaque cristal du scanner, une efficacité est tirée au hasard entre 0,9 et 1,1 dans une distribution de probabilité uniforme.
2. Le sinogramme de normalisation est ensuite calculé à partir de la carte d'efficacité des cristaux.
3. L'image est projetée pour obtenir le sinogramme des coïncidences vraies en tenant compte de la normalisation, et de l'atténuation.
4. Les coïncidences diffusées sont estimées grâce à une convolution ligne par ligne du sinogramme des coïncidences vraies normalisées, qui sont ensuite dénormalisées.
5. Les coïncidences fortuites sont simulées par la méthode *fan-sum* à partir des coïncidences vraies et des diffusées. Les coïncidences fortuites dues au LSO sont ensuite modélisées par une distribution uniforme dans l'espace des données.
6. Enfin, le sinogramme des coïncidences totales est estimé en lui appliquant une réalisation de la loi de Poisson.

Afin de choisir les grandeurs les plus réalistes possibles, le nombre total de coïncidences, les fractions de coïncidences diffusées et aléatoires, ainsi que l'activité injectée ont été modélisées par des variables aléatoires suivant une loi de probabilité gaussienne multivariée. Les paramètres de cette loi (moyenne et covariance) ont été calculés selon les statistiques fournies par une base de 91 examens cliniques. La suite d'opérations suivante est ensuite effectuée pour chaque image afin de fournir au programme de simulation les entrées nécessaires :

1. L'activité totale de l'image en question est mesurée. Cette activité mesurée est multipliée par un facteur permettant de simuler l'activité totale injectée au patient. Ce facteur a été estimé, à partir de plusieurs patients, comme le rapport approximatif entre l'activité mesurée dans tout le champ de vue 3D et celle mesurée dans une seule coupe 2D d'intérêt.

2. À partir de cette activité injectée, la loi de probabilité conditionnelle régissant le nombre de coïncidences totales et les fractions de coïncidences fortuites et diffusées est établie à partir d'une base d'une centaine de patients. Une réalisation de cette loi de probabilité est ensuite simulée afin de générer ces trois grandeurs.
3. Le nombre de coïncidences aléatoire dues au LSO est estimé par une constante, suivant les travaux de [Strydhorst et al., 2016]. Il est tronqué à 95 % du nombre total des coïncidences aléatoires dans le cas où il dépasserait cette valeur.

4.2.2.3 Choix des objectifs

Deux jeux de paramètres optimaux ont été calculés. L'un en utilisant une métrique globale sur l'ensemble de l'image : la MSSIM, l'autre en utilisant les quatre métriques décrites précédemment. Ces jeux de paramètres optimaux constitueront la valeur cible du problème de régression que le réseau de neurones devra résoudre.

MSSIM La MSSIM (mean structural similarity) [Wang et al., 2004] est une métrique permettant d'évaluer la conformité d'une image par rapport à une image de référence telle que la percevrait l'œil humain. Elle peut par exemple servir à évaluer la qualité de la compression JPEG des images. On utilise désormais couramment cette méthode pour évaluer la qualité des images TEP lorsqu'une référence, servant de vérité terrain, est disponible [Rasch et al., 2017, Gong et al., 2017, Bland et al., 2019]. Afin d'utiliser cette métrique, les images furent d'abord tronquées à partir de la carte d'atténuation du patient ayant servi pour le fond de l'image, afin de ne conserver de l'image que sa partie intéressante. Nous les avons ensuite normalisées en les divisant par leur valeur moyenne, conformément à ce que firent les auteurs de [Gong et al., 2017].

Utilisation des quatre métriques Les quatre métriques utilisées pour évaluer les images furent les quatre mêmes que celles utilisées dans le chapitre 3 de cette thèse (voir la section 3.2.4), à la différence que l'IR fut remplacé par l'écart type, pour éviter les divisions par des grandeurs trop faibles. Afin de mettre les quatre métriques à la même échelle, elles sont d'abord normalisées de la manière suivante : Une série de reconstructions est d'abord effectuée en parcourant l'espace des hyper-paramètres de reconstruction. L'annexe B détaille les hyper-paramètres parcourus lors ce ce

premier quadrillage. On normalise ensuite chaque métrique par une transformation affine pour que sa valeur minimale sur le parcours soit 0, et sa valeur maximale 1. La transformation affine appliquée dépend donc de l'image, et est ensuite fixée : elle ne changera pas au cours de l'optimisation. Le critère à minimiser retenu est ensuite le maximum des quatre métriques. On cherche en effet à ce qu'aucune métrique ne soit trop grande.

4.2.2.4 Optimisation des hyper-paramètres

Pour optimiser les hyper-paramètres de reconstruction, pour chacune des 112 images, selon chacun des deux critères, un premier quadrillage de l'espace des hyper-paramètres est d'abord effectué. L'annexe B détaille les hyper-paramètres parcourus lors de ce premier quadrillage. Dans le cas de l'optimisation en termes de métriques, le quadrillage a déjà été effectué pour normaliser les hyper-paramètres, et aucune reconstruction supplémentaire n'est nécessaire. Le meilleur point de ce quadrillage est ensuite utilisé comme point de départ d'un algorithme de recherche directe [Audet and Dennis Jr, 2002, Kolda et al., 2003].

La figure 4.2 montre un exemple d'image obtenue après optimisation des hyper-paramètres de reconstruction. On y constate que l'optimisation MSSIM est assez floue, au contraire de l'optimisation « métriques », plutôt bruitée.

Il est à noter que pour l'optimisation MSSIM, le jeu d'hyper-paramètre optimal a toujours inclus le post-traitement NNEPPS, au contraire de l'optimisation « métriques », qui l'a inclus pour 55 images sur 112.

4.2.3 Configuration expérimentale

16 patients ayant bénéficié d'une imagerie TEP à l'yttrium 90 post-radioembolisation ont été sélectionnés. Certains de ces patients ayant été imagés plusieurs fois, nous disposions en réalité de 20 images tridimensionnelles. Parmi ces 20 images, sept, correspondant à six patients différents, avaient un fond diffus et ont pu servir de base à la création des images 2D. Les autres ont servi à ajouter des tumeurs. Pour chacune de ces sept images 3D, six à dix coupes 2D ont pu être extraites et servir de base à la constitution des images. Chacune de ces coupes ayant été utilisée deux fois, 112 images ont été créées au total.

La base de données contient donc 112 éléments, contenant chacun :

— Une image « vérité terrain », de taille 200×200. Chaque image contient une tumeur, une zone uniforme (non nulle), et une zone d'activité nulle.

— Cinq sinogrammes qui lui sont associés, chacun de taille 344×252 : les sinogrammes des coïncidences totales, diffusées, et fortuites, l'atténuation, et la normalisation.

— Un jeu d'hyper-paramètres optimisé pour maximiser la MSSIM.

— Un jeu d'hyper-paramètres optimisé pour minimiser la combinaison des quatre métriques.

Sept patients différents ont servi à constituer le fond de ces 112 images [1]. Lors de la validation croisée, chacun de ces sept patients est alternativement pris comme ensemble de validation, alors que les six autres servent à l'entraînement. Cependant, le patient 6 a toujours été mis à l'écart de tout choix d'hyper-paramètre, afin que l'évaluation sur ce dernier fût la moins biaisée possible.

4.2.4 Architecture

Le réseau utilisé est un ResNet-18 [He et al., 2016]. Le ResNet est un réseau de neurones convolutionnel qui fait figure de référence pour la vision par ordinateur et l'analyse d'images médicales [Targ et al., 2016, Ayyachamy et al., 2019]. Le choix d'utiliser un réseau de neurones convolutionnel directement à partir des sinogrammes est également effectué par certains auteurs [Häggström et al., 2019]. Bien que le ResNet ait été initialement proposé pour des tâches de classification, son utilisation pour de la régression est aisée : il suffit en effet d'enlever la fonction softmax en sortie du réseau, et de considérer les valeurs directement issues de la dernière couche entièrement connectée.

Les cinq sinogrammes de coïncidences totales, fortuites, diffusées, d'atténuation et de normalisation sont fournis en entrée du réseau comme étant cinq canaux différents. Les sinogrammes sont d'abord normalisés avant d'être fournis au réseau, afin que les valeurs des cinq sinogrammes aient un ordre de grandeur comparable. Cette normalisation se fait de sorte que chaque sinogramme ait une moyenne nulle et une variance unitaire sur l'ensemble de la base de données privée du patient

1. En réalité six. Mais l'un des patients a subi deux examens et les images qui en résultent sont suffisamment différentes pour qu'elles puissent être considérées comme provenant de deux patients différents.

6, comme expliqué dans la section 4.2.3. Les hyper-paramètres de reconstruction ont, eux aussi, été normalisés de la même manière. Nous avons utilisé l'inverse du nombre d'itérations plutôt que le nombre d'itérations afin de rendre compte qu'une différence d'une itération a moins d'importance quand le nombre d'itérations est élevé. Lors de l'évaluation des objectifs (MSSIM ou combinaison des métriques), les hyper-paramètres donnés à l'algorithme de reconstruction sont d'abord tronqués pour coïncider au domaine sur lequel les hyper-paramètres « optimaux » ont été définis. Étant donné que l'hyper-paramètre « NNEPPS » était toujours de post-traiter l'image pour l'optimisation MSSIM (voir la section 4.2.2.4), cet hyper-paramètre n'est pas demandé comme sortie du réseau « MSSIM »

4.2.5 Entraînement du réseau

Le réseau de neurones fut implémenté à partir de PyTorch [Paszke et al., 2017]. L'entraînement du réseau s'est fait avec l'algorithme Adam (*adaptive moment estimation*) [Kingma and Ba, 2014]. Adam est similaire à l'algorithme du gradient stochastique habituel, mais maintient à jour un taux d'apprentissage adaptatif différent pour chaque paramètre du réseau. Les moments retenus furent ceux proposés par les auteurs : $\beta_1 = 0,9$ et $\beta_2 = 0,999$. Le taux d'apprentissage retenu fut de 0,001.

La fonction de coût choisie fut l'erreur quadratique moyenne, à laquelle fut ajoutée une pénalité L1 sur les paramètres du réseau, afin de le régulariser. La durée de l'entraînement fut fixée à 200 époques.

4.2.6 Méthodes « hyper-paramètres fixés »

Afin d'évaluer les performances du réseau, nous avons jugé bon d'évaluer ses performances par rapport à une autre méthode qui consisterait à fixer à l'avance les hyper-paramètres de reconstruction. Afin de comparer les différentes méthodes sur un même pied d'égalité, les hyper-paramètres fixés à l'avance sont optimisés sur l'ensemble d'entraînement, avant d'être évalués sur l'ensemble de validation. Plusieurs jeux d'hyper-paramètres fixés à l'avance ont été optimisés, en utilisant différents critères : minimisation de la MSSIM sur l'ensemble d'entraînement (méthode Fixe_{MSSIM}), minimisation de la combinaison des métriques ($\text{Fixe}_{métriques}$), et minimisation de l'erreur quadratique moyenne (Fixe_{MSE}). Ce dernier jeu d'hyper-

paramètres peut être simplement déterminé en prenant la moyenne des hyper-paramètres sur l'ensemble d'entraînement. En revanche, les deux premiers nécessitent une longue optimisation [2], et ont donc seulement été évalués à partir de la base de données privée du patient 6.

4.3 Résultats

La figure 4.2 montre des exemples d'images reconstruites avec les hyper-paramètres issus du réseau. On y constate que les images sont visuellement relativement proches des images « optimales », mais avec une qualité dégradée.

4.3.1 Validation croisée

Les figures 4.3, 4.4 et 4.5 montrent respectivement des exemples de courbes d'apprentissage et de validation du réseau. On y constate que la fonction de coût sur l'ensemble d'entraînement diminue et tend vers 0, au contraire des courbes sur l'ensemble de validation. Les tableaux 4.1, 4.2, 4.3 et 4.4 montrent les résultats de la validation croisée. Les tableaux 4.1 et 4.2 montrent les performances du réseau selon l'objectif final (respectivement la MSSIM et la combinaison des métriques), patients par patients, et les comparent par rapport à la valeur optimale, ainsi que par rapport aux méthodes fixant les paramètres à l'avance. L'écart type donné dans la colonne ResNet-18 correspond à l'écart type sur trois réalisations de l'apprentissage, à partir de paramètres initiaux du réseau différents. On y constate que le ResNet-18 a de moins bons résultats que les méthodes « fixes ». Les tableaux 4.3 et 4.4 montrent les mêmes comparaisons, mais selon l'erreur quadratique moyenne par rapport aux hyper-paramètres optimaux, qui est le critère utilisé lors de l'apprentissage. On y constate que le ResNet-18 a de meilleures performances que les méthodes Fixe_{MSSIM} ou $\text{Fixe}_{Métriques}$, mais moins bonnes que Fixe_{MSE}, ce qui confirme le surapprentissage du réseau. Les écarts types sont pris sur des entraînements distincts du réseau, tous issus d'une initialisation aléatoire. Chaque entraînement a en effet été réalisé trois fois. Si l'on évalue les réseaux selon l'objectif final, les performances du ResNet-18 sont moins bonnes que celles fixant les paramètres à l'avance (figures 4.4 et 4.5). Lorsque l'on évalue ces méthodes selon l'erreur quadratique moyenne par rapport

2. l'évaluation de ces grandeurs nécessite en effet d'effectuer des reconstructions.

Table 4.1 – Résultats de la validation croisée SSIM

	Valeur optimale	Fixe_{MSSIM}	ResNet-18
Patient 1	0,4439	0,4335	$0,3802 \pm 0,0092$
Patient 2	0,4686	0,4555	$0,4092 \pm 0,0043$
Patient 3	0,2756	0,2661	$0,2489 \pm 0,0033$
Patient 4	0,4295	0,4132	$0,3687 \pm 0,0009$
Patient 6	0,4328	0,4235	$0,3844 \pm 0,0122$
Patient 9	0,4372	0,4249	$0,3812 \pm 0,0038$
Patient 13	0,4584	0,4451	$0,4034 \pm 0,0009$

Table 4.2 – Résultats de la validation croisée Métriques

	Valeur optimale	$\text{Fixe}_{Métriques}$	ResNet-18
Patient 1	0,2550	0,4158	$0,4543 \pm 0,0265$
Patient 2	0,1476	0,2352	$0,3552 \pm 0,0387$
Patient 3	0,1458	0,2932	$0,3100 \pm 0,0217$
Patient 4	0,1410	0,2193	$0,3244 \pm 0,0129$
Patient 6	0,2275	0,3435	$0,4326 \pm 0,0228$
Patient 9	0,1235	0,2159	$0,2944 \pm 0,0059$
Patient 13	0,2115	0,3249	$0,4249 \pm 0,0093$

aux paramètres optimaux, le réseau est meilleur que Fixe_{MSSIM}, mais moins bon que Fixe_{MSE}, montrant un problème de généralisation du réseau.

4.3.2 Corrélation entre les variables

Le tableau 4.5 montre la corrélation entre les différents hyper-paramètres, ainsi que la valeur moyenne, la variance et l'erreur quadratique moyenne de chacun. On y constate qu'il n'y a pas d'hyper-paramètre qui soit mieux estimé que les autres. Les corrélations entre les hyper-paramètres ont bien été comprises du réseau. En particulier, il existe une très forte corrélation entre la borne inférieure et l'intensité de la pénalité (-0,87 sur l'ensemble d'entraînement et -0,99 sur l'ensemble de validation). Le réseau a bien appris cette corrélation (-0,68). Le problème principal vient de

TABLE 4.3 – Résultats de la validation croisée SSIM loss

	Valeur optimale	Fixe_{MSSIM}	Fixe_{MSE}	ResNet-18
Patient 1	0	1,2679	0,6459	$0,6673 \pm 0,0294$
Patient 2	0	1,5136	0,9495	$1,1783 \pm 0,1415$
Patient 3	0	1,4673	0,969	$1,0960 \pm 0,0366$
Patient 4	0	1,9633	1,219	$1,2817 \pm 0,0241$
Patient 6	0	1,0996	0,6727	$0,7414 \pm 0,0514$
Patient 9	0	1,1257	0,9289	$1,0283 \pm 0,0667$
Patient 13	0	1,7963	1,25	$1,5337 \pm 0,1096$

TABLE 4.4 – Résultats de la validation croisée Métriques loss

	Valeur optimale	$\text{Fixe}_{Métriques}$	Fixe_{MSE}	ResNet-18
Patient 1	0	3,2950	2,271	$2,3787 \pm 0,2880$
Patient 2	0	1,2705	0,9522	$1,1083 \pm 0,1286$
Patient 3	0	1,7334	0,9652	$1,0997 \pm 0,0605$
Patient 4	0	1,2266	0,758	$0,7644 \pm 0,0422$
Patient 6	0	2,2672	1,608	$1,7730 \pm 0,0795$
Patient 9	0	1,0286	0,8559	$0,9951 \pm 0,0754$
Patient 13	0	1,1966	0,7177	$0,7098 \pm 0,0348$

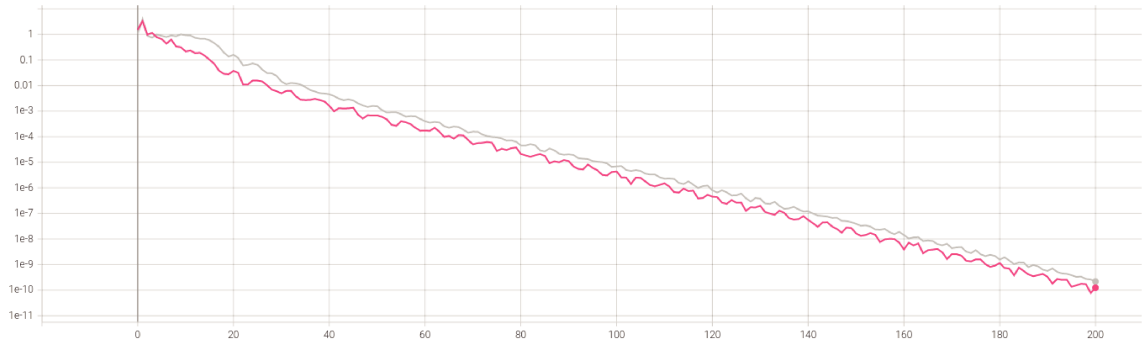


FIGURE 4.3 – Courbe d'entraînement du patient 6

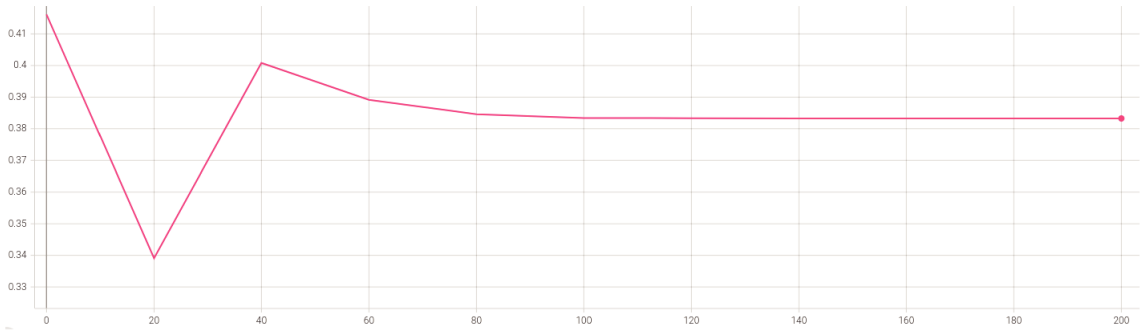


FIGURE 4.4 – Courbe de validation du MSSIM du patient 6

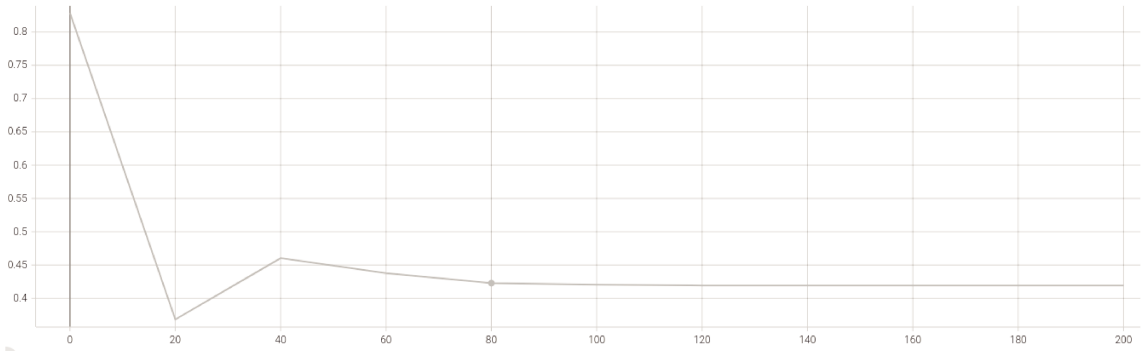


FIGURE 4.5 – courbe de validation de l'optimisation « métriques » du patient 6

ce que la variance des hyper-paramètres estimée par le réseau sur l'ensemble de validation est beaucoup plus faible que leur variance réelle. Ceci suggère que le réseau a en réalité appris une fonction proche d'une constante et a seulement appris les valeurs particulières des hyper-paramètres correspondant aux sinogrammes vus lors de l'entraînement.

4.3.3 Analyse et discussion

4.3.3.1 Perspectives

La figure 4.3 montre clairement que le réseau apprend. En effet, sur l'ensemble d'entraînement, les paramètres estimés par le réseau convergent vers les paramètres optimaux. En revanche, les performances du réseau sur l'ensemble de validation sont mauvaises : le réseau ne fait pas mieux sur cet ensemble qu'un algorithme qui se contenterait de trouver la valeur moyenne de chaque hyper-paramètre à partir de la base d'entraînement. Les solutions possibles à ce problème — que nous détaillons ci-dessous — comprennent la régularisation du réseau, le changement d'architecture, l'augmentation du nombre de données d'entraînement et le choix d'un format plus naturel pour les données d'entrée.

Régularisation du réseau Les résultats montrés dans la section 4.3 ont été obtenus en appliquant une pénalité L1 au réseau. Cependant, l'intensité de la pénalité était probablement beaucoup trop faible. Au lieu de 1, nos dernières expériences montrent qu'une pénalité de l'ordre de 10^5 serait probablement plus adaptée. Ces résultats semblent encourageants puisque l'erreur quadratique moyenne des hyper-paramètres a réellement pu être diminuée sur l'ensemble de validation. Ces résultats sont cependant à confirmer sur les autres patients. Une autre stratégie de régularisation possible consisterait à adjoindre au ResNet un décodeur qui tenterait de reconstituer les sinogrammes d'origine. Un terme pénalisant la différence entre les sinogrammes décodés et les sinogrammes de départ serait alors ajouté à la fonction de coût.

Augmentation de la taille de l'ensemble d'entraînement. C'est probablement la principale piste à privilégier. Les réseaux de neurones profonds sont en effet connus pour nécessiter un nombre particulièrement important de données d'entraî-

TABLE 4.5 – Corrélation, moyenne, variance, erreur quadratique moyenne des différents hyper-paramètres. L'ensemble de validation est le patient 6.

		ensemble d'entraînement (hyper-paramètres optimaux)	ensemble de validation (hyper-paramètres optimaux)	ensemble de validation (prédictions du resnet18)	ensemble de validation (prédictions de Fixe$_{MSE}$)
erreur quadratique moyenne	borne inférieure			1,07	1,05
	intensité de la pénalité			0,85	0,73
	1/itérations			0,22	0,24
moyenne	borne inférieure	0	-0,02	-0,10	0
	intensité de la pénalité	0	0	0,18	0
	1/itérations	0	-0,27	-0,03	0
variance	borne inférieure	1	1,14	0,04	0
	intensité de la pénalité	1	0,80	0,04	0
	1/itérations	1	0,18	0,05	0
corrélation	corrélation entre la borne inférieure et l'intensité de la pénalité	-0,87	-0,99	-0,68	
	corrélation entre la borne inférieure et 1/itérations	0,39	0,56	0,74	
	corrélation entre l'intensité de la pénalité et 1/itérations	-0,34	-0,58	-0,36	

nement [Goodfellow et al., 2016]. Une piste pourrait être de conserver la base de données créée pour constituer les ensembles de validation, de test, et éventuellement une partie de l'ensemble d'entraînement, et de compléter ce dernier ensemble par des données simulées moins réalistes en grand nombre. En effet, il a été rapporté que l'augmentation du nombre de données d'entraînement améliorait généralement les résultats du réseau sur les ensembles de validation et de test, même dans les cas où les distributions de ces différents ensembles ne sont pas les mêmes. Alternativement, les méthodes génératives récentes permettent de s'approcher de leur distribution d'origine de manière assez réaliste.

Données d'entrée plus naturelles. On pourrait penser que les grandeurs les plus importantes pour le choix des hyper-paramètres serait des grandeurs « bilan », telles que le nombre total de coïncidences collectées ou la fraction de coïncidences aléatoires. Or, ces informations, bien que présentes dans les sinogrammes, sont difficilement accessibles telles qu'elles par le réseau. On pourrait donc imaginer les donner directement en entrée du réseau. De même, certaines relations entre les sinogrammes, comme la relation multiplicative existant entre les sinogrammes d'atténuation et de normalisation, sont difficiles à apprendre, et l'on pourrait imaginer donner en entrée du réseau des sinogrammes additionnels, calculés lors d'une étape de pré-traitement. Il est également à noter que les sinogrammes ne sont pas particulièrement adaptés à l'entrée d'un réseau convolutionnel. Certains travaux utilisant des sinogrammes en entrée commencent par des couches entièrement connectées [Zhu et al., 2018], bien que d'autres utilisent aussi des couches convolutionnelles directement à partir des sinogrammes, comme [Häggström et al., 2019], qui utilise comme nous des convolutions 7×7. Une idée originale serait de fournir des histo-images en entrée du réseau [Matej et al., 2009, Whiteley et al., 2021, Ote and Hashimoto, 2021], c'est-à-dire un format où les coïncidences sont directement placés dans l'espace image, ce format est alors adapté à l'entrée d'un réseau convolutionnel. Cependant, ce format nécessite de connaître le temps de vol, ce qui n'est pas le cas dans notre étude. Alternativement, [Ye et al., 2018] ont proposé de rétro-projeter les lignes de réponse dans l'image.

Architecture du réseau L'architecture utilisée dans cette étude est une architecture connue et validée. La modification de l'architecture n'est probablement pas

la première piste à envisager. Cependant, en cas d'insuccès des autres pistes, et particulièrement si le choix des sinogrammes en entrée était maintenu, d'autres architectures pourraient être étudiées. Les résultats des tableaux 4.1, 4.2, 4.3 et 4.4 suggèrent que la minimisation de la norme L2 correspond mal à une minimisation de l'objectif fixé (MSSIM ou combinaison des métriques). Il pourrait donc être intéressant de chercher d'autres fonctions de coût plus adaptées.

4.3.3.2 Étude préliminaire de certaines perspectives

Des études préliminaires de certaines de ces perspectives ont été menées, avec deux modifications par rapport aux études présentées dans la section 4.3 : l'intensité de la pénalité L1 appliquée aux paramètres du réseau a été considérablement augmentée afin d'éviter le surapprentissage. Le nombre de sinogrammes donné en entrée du réseau a également été augmenté, créant une entrée très redondante et laissant le réseau sélectionner celles qu'il souhaite utiliser. Certaines grandeurs « bilan » ont également été proposées en entrée du réseau, en fournissant une image du même format que les autres sinogrammes, mais dans laquelle chaque pixel contient la grandeur bilan. Les vingt-deux sinogrammes fournis en entrée du réseau sont les suivants :

1. sinogramme d'atténuation
2. sinogramme de normalisation
3. sinogramme des coïncidences totales
4. sinogramme des coïncidences fortuites
5. sinogramme des coïncidences diffusées
6. sinogramme contenant le terme additif total (fortuites plus diffusées)
7. sinogramme des coïncidences vraies
8. sinogramme contenant le terme multiplicatif (atténuation multiplié par la normalisation)
9. image projetée
10. terme additif multiplié
11. coïncidences totales multipliées
12. projection d'une image uniforme

13. terme multiplicatif incluant la projection de l'image uniforme
14. nombre de coïncidences totales
15. nombre de coïncidences aléatoires
16. nombre de coïncidences diffusées
17. nombre de coïncidences vraies
18. fraction de coïncidences aléatoires
19. proportion de coïncidences diffusées (nombre de coïncidences diffusées divisé par le nombre total de coïncidences)
20. fraction de coïncidences diffusées (nombre de coïncidences diffusées divisé par le nombre de coïncidences vraies et diffusées combinées)
21. proportion de coïncidences vraies
22. estimation de l'activité totale dans l'image

Les résultats obtenus avec une pénalité de $9,15 \times 10^4$ sont encourageants, car l'erreur quadratique moyenne sur l'ensemble de validation n'était que de 0,54 après 1800 époques d'entraînement, soit assez nettement en dessous des 0,67 obtenus avec des paramètres fixes. À titre de comparaison, le même apprentissage lancé en n'autorisant en entrée que les cinq sinogrammes de départ n'a pas permis de réduire l'erreur quadratique moyenne en deçà de 0,68 (après 2800 époques). Ce résultat reste cependant à confirmer par une validation croisée.

4.3.3.3 Formulation du problème

On peut voir la constitution de l'ensemble des hyper-paramètres « optimaux » comme l'application d'une fonction prenant en entrée un sinogramme et une image théorique, et donnant en sortie un jeu d'hyper-paramètres. À ces entrées, il faut ajouter, dans le cas de l'optimisation en termes de métriques, le choix des régions d'intérêt. Il faut noter que la modification d'une de ces entrées changerait le jeu d'hyper-paramètres fourni. Le problème que nous nous sommes proposé d'étudier dans ce chapitre est l'élaboration d'un algorithme, utilisant de l'apprentissage profond, permettant de trouver une approximation des paramètres optimaux à partir des seuls sinogrammes. À ce stade, deux remarques peuvent être faites :

1. Aucun algorithme, aussi perfectionné soit-il, ne sera en mesure de trouver exactement les paramètres optimaux, à partir des sinogrammes seuls. En effet,

plusieurs images distinctes peuvent avoir généré le même sinogramme [3]. Selon l'image qui les a généré, le jeu d'hyper-paramètres optimaux ne sera pas le même.

2. Les sinogrammes permettent néanmoins à eux seuls de donner une estimation des hyper-paramètres optimaux. On sait par exemple que lorsque le nombre de coïncidences recueillies est faible, la pénalité doit être plus importante.

Le problème est donc intéressant et réaliste, à condition de le formuler clairement, et de prendre en compte que l'erreur de Bayes n'est pas nulle. Le but de cette section est de proposer une formulation claire du problème.

À un sinogramme donné est associé une distribution d'images ayant pu le générer. L'image la « plus probable » étant le maximum a posteriori. À cette distribution d'images correspond une distribution d'hyper-paramètres optimaux. Au sein de cette distribution, il faut définir un jeu d'hyper-paramètres qui sera le résultat souhaité en sortie du réseau. Dans ce chapitre, le choix implicite était de maximiser (resp. minimiser) la moyenne de la MSSIM (resp. de la combinaison des métriques) ; la moyenne étant pondérée par la densité de probabilité de chaque image [4].

La définition précédente crée une difficulté supplémentaire, généralement absente (ou minime) dans les problèmes où un apprentissage supervisé est effectué : les labels fournis au réseau pour l'apprentissage ne correspondent pas tout à fait à la sortie souhaitée du réseau. Cependant, ce problème se rapproche des problèmes de labels bruités, pour lesquels il existe une littérature abondante (voir par exemple [Natarajan et al., 2013,Sukhbaatar et al., 2014,Li et al., 2017]). Un ensemble d'entraînement suffisamment grand devrait donc permettre de compenser ce problème en présentant une bonne partie de la distribution de probabilité. Le choix de la fonction de coût est alors primordiale pour remplir l'objectif fixé.

Le problème étant posé, l'importance pratique des remarques précédentes reste à étudier. Dans le cas où la variabilité des images associées aux sinogramme s'avérerait faible, les labels pourraient être pris tels quels, comme la sortie souhaitée du réseau. Il est en tous les cas certain que plus le nombre de données collectées baisse, plus

3. Cela se comprend bien en considérant, par exemple, qu'une image est continue alors qu'un sinogramme de coïncidences est discret. Les autres sinogrammes ne le sont pas forcément mais sont issus de données discrètes.

4. Il faut noter que rien ne garantit que le jeu de paramètres ainsi défini soit dans la distribution d'hyper-paramètres optimaux. On peut cependant supposer que s'il n'est pas dedans, il ne doit au moins pas en être trop éloigné.

cette variabilité est grande.

4.4 Conclusion

Ce chapitre présente un travail préliminaire en vue de créer un algorithme capable de donner le jeu d'hyper-paramètres de reconstruction optimal dans un but précis. Ses contributions principales sont une formulation claire du problème et une discussion sur les pistes futures à explorer, ainsi que la constitution d'une base de données de 112 images, qui pourra être réutilisée. Certains résultats récents sont également encourageants, même s'ils restent à confirmer.

Chapitre 5

Conclusion générale

5.1 Conclusion

Cette thèse s'est proposé d'étudier, de proposer, et de comparer diverses techniques ayant pour objet la TEP à l'yttrium 90, avec un accent particulier sur la réduction du biais positif dans les zones froides.

Sa première partie avait pour objet la proposition d'une nouvelle technique, baptisée NNEPPS, permettant de supprimer les valeurs négatives d'une image en préservant le plus possible ses valeurs moyennes locales. Elle a été étudiée en détail et ses propriétés mathématiques ont été démontrées. Un code a également été développé et est disponible en libre accès. Cette partie a donné lieu à un article de journal [Millardet et al., 2020].

La deuxième partie de cette thèse compare différents algorithmes de reconstruction et différents choix d'hyper-paramètres, dans le cadre d'une étude multi-objectifs. Elle propose également un nouvel algorithme, APPGML. Elle conclut à la supériorité des différences relatives sur une simple pénalité quadratique, et à la plus grande souplesse d'utilisation d'AML comparée à NEG-ML. Elle conclut également qu'on peut obtenir un compromis entre le biais et le bruit dans l'image en utilisant APPGML, avec ou sans le NNEPPS suivant les cas. Cette partie a, elle aussi, donné lieu à un article de journal [Millardet et al., 2021].

Dans la troisième partie, nous avons posé les prémices de travaux futurs devant aboutir à un algorithme permettant de définir les meilleurs hyper-paramètres de reconstruction, en fonction des sinogrammes et de l'objectif visé. L'apport de cette partie est la proposition de l'idée et la formulation claire du problème et de ses difficultés, la conception d'une banque de 112 images pouvant servir de support à de futurs travaux, une première proposition de réseau de neurone et l'analyse de ses résultats, et une discussion détaillée des perspectives futures.

5.2 Perspectives

Cette thèse ouvre également de nombreuses perspectives pour de futurs travaux.

Un biais négatifs provoqué par le NNEPPS a été identifié à l'interface entre froides et chaudes. Son influences, ainsi que les solutions pour corriger ce problème restent à étudier. En particulier, son effet sur les petites lésions reste à étudier. Une solution envisagée pour corriger ce problème est d'étudier la possibilité d'un

NNEPPS non isotrope. Une première étude à ce sujet a été présentée, mais celle-ci reste à compléter. Une perspective du NNEPPS pourrait être de ne pas se limiter à une phase de post-traitement, mais de l'inclure au cours de la reconstruction. Cependant, les propriétés d'un tel algorithme resteraient à étudier. Le choix des coefficients de la matrice laplacienne $\boldsymbol{H}$ reste également une question ouverte. Enfin, l'usage du NNEPPS n'est pas limité à la TEP, mais pourrait trouver une application dans tous les problèmes d'imagerie ayant une contrainte de positivité imposée par la physique du problème, mais dont les images estimées peuvent contenir des valeurs négatives.

Parmi les perspectives importantes de l'analyse comparative multi-objectifs de divers algorithmes de reconstruction, il sera important d'évaluer les algorithmes ADMM-Reg et HypoC-PML, respectivement proposés par [Lim et al., 2018] et par [Bousse et al., 2020]. Il pourra également s'avérer utile de trouver une extension de la pénalité proposée par [Nuyts et al., 2002a] s'appliquant aux images contenant des valeurs négatives.

L'algorithme de choix des hyper-paramètres en reconstruction itérative présente également de nombreuses perspectives afin d'en améliorer les résultats. Ces perspectives peuvent se classer en quatre catégories, détaillées dans le chapitre en question. La principale perspective est l'augmentation de la taille de la base d'entraînement, mais il faut également citer la régularisation du réseau, un changement d'architecture, ainsi que l'utilisation de données d'entrée plus naturelles.

Annexe A

Annexe au chapitre 1

Sommaire

A.1 Preuve de la proposition 1 149
A.2 Preuve de la proposition 3 149
A.3 Preuve de la proposition 4 150
A.4 Preuve de la proposition 5 150
A.5 Preuve de la proposition 6 151
A.6 Coefficients de H utilisés dans cette étude 153
A.6.1 Cas des pixels carrés et des voxels cubiques 153
A.6.2 Cas des pixels rectangulaires et des voxels parallélépipédiques 154
A.6.3 Remarques conclusives 154

A.1 Preuve de la proposition 1

Proposition 1. *Si $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ sont deux cartes des transferts satisfaisant les conditions (2.2b), le vecteur $\boldsymbol{\gamma}$ défini par $\forall i, \gamma_i = \min(\alpha_i, \beta_i)$ satisfait également (2.2b).*

Démonstration. Il est évident que $\forall i, \gamma_i \geq 0$. Soit i un indice arbitraire. Supposons, sans perte de généralité, que $\alpha_i \leq \beta_i$. On a donc $\gamma_i = \alpha_i$. Puisque la diagonale de $\boldsymbol{H}$ est strictement positive, ses autres entrées étant négatives, la fonction $\begin{cases} \mathbb{R}^N \to \mathbb{R} \\ \boldsymbol{\alpha} \mapsto (\boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha})_i \end{cases}$ est croissante par rapport à la i^{e} variable, et décroissante par rapport aux autres. Puisque $\gamma_i \geq \alpha_i$ (car $\gamma_i = \alpha_i$), et $\forall j \neq i, \gamma_j \leq \alpha_j$ (car $\forall j, \gamma_j \leq \alpha_j$), on a bien $(\boldsymbol{x} + \boldsymbol{H}\boldsymbol{\gamma})_i \geq (\boldsymbol{x} + \boldsymbol{H}\boldsymbol{\alpha})_i \geq 0$. □

A.2 Preuve de la proposition 3

Proposition 3. *Une solution au problème (2.2) existe si et seulement si la valeur moyenne de l'image initiale $\boldsymbol{X}$ est positive.*

Démonstration. Si la valeur moyenne de l'image initiale est strictement négative, il est évidemment impossible de trouver une image positive sans changer sa valeur moyenne.

Montrons également la réciproque. Nous montrerons d'abord que d'après les propriétés de $\boldsymbol{H}$, il est possible de trouver un $\boldsymbol{\alpha}$ tel que $\boldsymbol{y} \geq \boldsymbol{0}_n$. Nous créerons ensuite un $\boldsymbol{\alpha}$ positif à partir du précédent. Comme mentionné dans la remarque de la section 2.2.4, la matrice $\boldsymbol{H}$ est de rang $n-1$, et chaque sous-matrice principale de $\boldsymbol{H}$ est inversible. Soit i un voxel quelconque, et $\mathcal{I} = [\![1, n]\!] \setminus i$. Puisque $\boldsymbol{H}_{\mathcal{I}\mathcal{I}}$ est inversible, on peut trouver $\tilde{\boldsymbol{\alpha}}$ tel que $\boldsymbol{x}_{\mathcal{I}} + \boldsymbol{H}_{\mathcal{I}\mathcal{I}}\tilde{\boldsymbol{\alpha}}_{\mathcal{I}} = \boldsymbol{0}$. Définissons

$$\begin{cases} \tilde{\boldsymbol{\alpha}}_{\mathcal{I}} = -\boldsymbol{H}_{\mathcal{I}\mathcal{I}}^{-1}\boldsymbol{x}_{\mathcal{I}} \\ \tilde{\alpha}_i = 0 \end{cases} \tag{A.1}$$

Cela nous donne l'image finale $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{H}\tilde{\boldsymbol{\alpha}}$. Comme $\boldsymbol{x}$ et $\boldsymbol{y}$ partagent la même moyenne (positive), et étant donné que $\boldsymbol{y}_{\mathcal{I}} = 0$, le i^{e} voxel vérifie $\boldsymbol{y}_i \geq 0$. L'image finale est donc positive. Cependant, la contrainte $\tilde{\boldsymbol{\alpha}} \geq 0$ n'est pas nécessairement remplie. Parce que les vecteurs constants appartiennent au noyau de $\boldsymbol{H}$, tout vecteur de ce type peut être ajouté à $\tilde{\boldsymbol{\alpha}}$ sans que l'image finale ne soit modifiée, et ainsi $\boldsymbol{\alpha}$ peut être rendu positif. L'existence de la solution est donc prouvée. □

A.3 Preuve de la proposition 4

Proposition 4. *Si une solution au problème (2.2) existe, elle est unique.*

Démonstration. C'est une conséquence directe de la façon dont est définie la carte des transferts $\boldsymbol{\alpha}$ minimale : sa i^{e} coordonnée est la borne inférieure des i^{e} coordonnées de chaque carte satisfaisant (2.2b). □

A.4 Preuve de la proposition 5

Proposition 5. *Si le problème de post-traitement admet une solution (voir proposition 3), le méthode duale du simplexe atteint cette solution.*

L'utilisation d'un simplexe dual se justifie par sa convergence en un nombre fini d'itérations, que nous allons prouver dans cette section. En effet, à condition que le problème d'optimisation linéaire soit non dégénéré et borné, la méthode du simplexe se termine en atteignant un point basique admissible [Nocedal and Wright, 1999]. Montrons les deux parties :

Lemme 1. *Le problème dual n'est pas dégénéré*

Démonstration. Dans un problème d'optimisation linéaire dégénéré, il existe au moins un point basique admissible qui ait moins de n composantes non nulles. Montrons que chaque point admissible du problème dual a au moins n composantes non nulles.

Si un point est admissible, $\boldsymbol{H\pi} + \boldsymbol{t} = \mathbf{1}_n$. Mais $\pi_i = 0$ implique que $(\boldsymbol{H\pi})_i \leq 0$, et donc $\boldsymbol{t}_i \neq 0$. Ainsi, pour être admissible, un point $\begin{pmatrix} \boldsymbol{\pi} \\ \boldsymbol{t} \end{pmatrix}$ doit avoir au moins n composantes non nulles.

Le problème dual n'est pas dégénéré. □

Lemme 2. *Si l'image initiale a une moyenne positive, le problème dual est borné.*

Démonstration. Ceci est une conséquence directe de la faisabilité du problème primal dans ce cas (voir proposition 3). □

La méthode duale du simplexe se termine donc si la moyenne de l'image initiale est positive.

A.5 Preuve de la proposition 6

Proposition 6. *Si le point de départ choisi est $\boldsymbol{y}^{(0)} = \boldsymbol{x}$, à chaque itération du simplexe dual, tous les indices entrants possibles (tous les i pour lesquels $y_i^{(k)} < 0$) sont bien dans la base d'indices de la solution* i.e.*, tous les choix possibles sont des choix corrects. De plus, l'indice sortant associé à un indice entrant est connu.*

L'algorithme proposé part du point naturel $\boldsymbol{y}^{(0)} = \boldsymbol{x}$, c'est-à-dire que nous choisissons $\mathcal{B}^{(0)} = [\![n+1, 2n]\!]$. Vérifions d'abord que ce point est un point basique admissible :

Ce choix correspond à $\boldsymbol{\pi} = \mathbf{0}_n$ et $\boldsymbol{t} = \mathbf{1}_n$.

- $\boldsymbol{s}$ est alors un point admissible du problème dual présenté dans la section 2.3.1, avec au plus (en réalité exactement) n composantes non nulles.
- $\mathcal{B}^{(0)}$ est un sous-ensemble de $[\![1, 2n]\!]$ tel que :
  - $\mathcal{B}^{(0)}$ contient n indices
  - $i \notin \mathcal{B}^{(0)} \implies z_i = 0$.
  - La matrice $n \times n$ $\boldsymbol{B}^{(0)} = \boldsymbol{G}_{[\![1,2n]\!],\mathcal{B}}$ (donc $\boldsymbol{B}^{(0)} = \boldsymbol{I}_n$) est inversible.

Le point initial est donc un point basique admissible du problème dual. Dans l'algorithme du simplexe (voir l'algorithme 1), l'indice entrant q doit toujours être choisi de telle sorte que $z_q = \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{\alpha} \end{pmatrix}_q < 0$.

Les deux lemmes suivants seront démontrés par récurrence.

Lemme 3. *$\boldsymbol{\alpha}$ reste toujours positif, et l'indice entrant doit donc toujours être choisi dans $[\![1, n]\!]$.*

Lemme 4. *Si l'indice entrant choisi est q, l'indice sortant est nécessairement $n+q$.*

Les lemmes 3 et 4 impliquent que les indices entrants possibles sont toujours dans $[\![1, n]\!]$, et les indices sortants dans $[\![n+1, 2n]\!]$. Par conséquent, on sait que tout indice entrant, quel que soit le choix réalisé, ne sortira jamais de la base et est effectivement dans la base de la solution. De plus, puisque les indices sortants associés aux indices entrants sont connus, il est possible de condenser efficacement les itérations.

Preuve des lemmes 3 et 4. Les deux lemmes seront démontrés en une seule démonstration par récurrence.

L'initialisation est vérifiée puisque dans l'état initial, $\boldsymbol{\alpha}^{(0)} = \mathbf{0}_n$.

Hérédité : puisque $\boldsymbol{\alpha}^{(k)} \geq 0$, l'indice entrant doit être choisi dans $[\![1, n]\!]$, et tel que $y_q^{(k)} < 0$. Une itération du simplexe consisterait à augmenter $\pi_q^{(k)}$ tout en maintenant $\boldsymbol{G}\boldsymbol{s}^{(k+1)} = \mathbf{1}_n$, et $\boldsymbol{z}_{\mathcal{N}^{(k)}\setminus q}^{(k+1)} = \mathbf{0}_{n-1}$. Si $\mathcal{Z}^{(k)} = [\![1, n]\!] \cap \mathcal{B}^{(k)}$, et $\mathcal{P}^{(k)} = [\![1, n]\!] \setminus \mathcal{B}^{(k)}$, d'après l'hypothèse de récurrence du lemme 4, $\mathcal{B}^{(k)} = \mathcal{Z}^{(k)} \cup (n + \mathcal{P}^{(k)})$, ce qui implique que $\boldsymbol{\pi}_{\mathcal{P}^{(k)}}^{(k)} = \mathbf{0}$ et $\boldsymbol{t}_{\mathcal{Z}^{(k)}}^{(k)} = \mathbf{0}$. Par conséquent :

$$\boldsymbol{H}_{\mathcal{Z}^{(k)}\mathcal{Z}^{(k)}} \boldsymbol{\pi}_{\mathcal{Z}^{(k)}}^{(k+1)} + \boldsymbol{H}_{\mathcal{Z}^{(k)}q} \pi_q^{(k+1)} = \mathbf{1} = \boldsymbol{H}_{\mathcal{Z}^{(k)}\mathcal{Z}^{(k)}} \boldsymbol{\pi}_{\mathcal{Z}^{(k)}}^{(k)}$$

$$\begin{aligned} \boldsymbol{H}_{\mathcal{P}^{(k)}\mathcal{Z}^{(k)}} \boldsymbol{\pi}_{\mathcal{Z}^{(k)}}^{(k+1)} + \boldsymbol{H}_{\mathcal{P}^{(k)}q} \pi_q^{(k+1)} + \boldsymbol{t}_{\mathcal{P}^{(k)}}^{(k+1)} &= \mathbf{1} \\ &= \boldsymbol{H}_{\mathcal{P}^{(k)}\mathcal{Z}^{(k)}} \boldsymbol{\pi}_{\mathcal{Z}^{(k)}}^{(k)} + \boldsymbol{t}_{\mathcal{P}^{(k)}}^{(k)} \end{aligned}$$

ce qui conduit à :

$$\boldsymbol{\pi}_{\mathcal{Z}^{(k)}}^{(k+1)} = \boldsymbol{\pi}_{\mathcal{Z}^{(k)}}^{(k)} - (\boldsymbol{H}_{\mathcal{Z}^{(k)}}\mathcal{Z}^{(k)})^{-1} \boldsymbol{H}_{\mathcal{Z}^{(k)}q} \pi_q^{(k+1)}$$

$$\begin{aligned} \boldsymbol{t}_{\mathcal{P}^{(k)}}^{(k+1)} = \boldsymbol{t}_{\mathcal{P}^{(k)}}^{(k)} + \Big(&\boldsymbol{H}_{\mathcal{P}^{(k)}\mathcal{Z}^{(k)}} (\boldsymbol{H}_{\mathcal{Z}^{(k)}\mathcal{Z}^{(k)}})^{-1} \\ &\boldsymbol{H}_{\mathcal{Z}^{(k)}q} - \boldsymbol{H}_{\mathcal{P}^{(k)}q} \Big) \pi_q^{(k+1)} \end{aligned}$$

De plus, d'après la remarque 2.2.4, tous les coefficients de $(\boldsymbol{H}_{\mathcal{Z}^{(k)}\mathcal{Z}^{(k)}})^{-1}$ sont positifs, et $\boldsymbol{H}_{\mathcal{Z}^{(k)}q} \leq 0$ (puisque q n'est pas dans $\mathcal{Z}^{(k)}$). Ceci implique que toutes les composantes de $\boldsymbol{\alpha}^{(k+1)}$ sont croissantes avec $\pi_q^{(k+1)}$. De la même manière, on a : $\boldsymbol{H}_{\mathcal{P}^{(k)}\mathcal{Z}^{(k)}} \leq 0$, $(\boldsymbol{H}_{\mathcal{Z}^{(k)}\mathcal{Z}^{(k)}})^{-1} \geq 0$, $\boldsymbol{H}_{\mathcal{Z}^{(k)}q} \leq 0$. Nous avons enfin $(\boldsymbol{H}_{\mathcal{P}^{(k)}q})_{i \neq q} \leq 0$ et $(\boldsymbol{H}_{\mathcal{P}^{(k)}q})_q = H_{qq} \geq 0$. Donc toutes les composantes de $\boldsymbol{t}_{\mathcal{P}^{(k)}}$ sont croissantes, sauf la q^{e} composante, sur laquelle on ne peut *a priori* rien déduire. Mais puisque le simplexe dual se termine (voir la partie 2.3.3), un indice doit être conduit à 1. Il est donc nécessairement $n + q$.

Maintenant, selon l'équation (2.4a), et puisque $\boldsymbol{y}_{\mathcal{Z}^{(k+1)}}^{(k+1)} = \mathbf{0}$, et $\boldsymbol{\alpha}_{\mathcal{P}^{(k+1)}}^{(k+1)} = \mathbf{0}$; $\mathbf{0} = \boldsymbol{x}_{\mathcal{Z}^{(k+1)}} + \boldsymbol{H}_{\mathcal{Z}^{(k+1)}\mathcal{Z}^{(k+1)}} \boldsymbol{\alpha}_{\mathcal{Z}^{(k+1)}}^{k+1}$, ce qui conduit à $\boldsymbol{\alpha}_{\mathcal{Z}^{(k+1)}}^{k+1} = -(\boldsymbol{H}_{\mathcal{Z}^{(k+1)}\mathcal{Z}^{(k+1)}})^{-1} \boldsymbol{x}_{\mathcal{Z}^{(k+1)}}$. De la même façon, et puisque $\mathcal{P}^{(k+1)} \subset \mathcal{P}^{(k)}$;

$$\boldsymbol{y}_{\mathcal{Z}^{(k+1)}}^{(k)} = x_{\mathcal{Z}^{(k+1)}} + \boldsymbol{H}_{\mathcal{Z}^{(k+1)}\mathcal{Z}^{(k+1)}} \boldsymbol{\alpha}_{\mathcal{Z}^{(k+1)}}^{k}$$

Couplé avec la forme ci-dessus de $\boldsymbol{\alpha}^{k}_{\mathcal{Z}^{(k+1)}}$, nous obtenons

$$\boldsymbol{\alpha}^{(k+1)}_{\mathcal{Z}^{(k+1)}} - \boldsymbol{\alpha}^{(k)}_{\mathcal{Z}^{(k+1)}} = -(\boldsymbol{H}_{\mathcal{Z}^{(k+1)}})^{-1}\boldsymbol{y}^{(k)}_{\mathcal{Z}^{(k+1)}}$$

De plus, $\boldsymbol{y}^{(k)}_{\mathcal{Z}^{(k+1)}} \leq \mathbf{0}$ (parce que l'indice entrant est choisi parmi les voxels négatifs). Et donc, $\boldsymbol{\alpha}^{(k+1)}_{\mathcal{Z}^{(k+1)}} - \boldsymbol{\alpha}^{(k)}_{\mathcal{Z}^{(k+1)}} \geq \mathbf{0}$. Donc, $\boldsymbol{\alpha}^{(k+1)} \geq \mathbf{0}_n$. □

A.6 Coefficients de $\boldsymbol{H}$ utilisés dans cette étude

Cette annexe détaille les considérations théoriques ayant mené au choix des coefficients de $\boldsymbol{H}$ utilisés dans cette thèse. Ces considérations suivent l'hypothèse **F3** selon laquelle les transferts doivent être isotropes, et ne tiennent donc pas compte du travail plus récent présenté dans la section 2.4.2.

Nous partons du postulat selon lequel le résultat du post-traitement ne devrait pas dépendre de la taille des voxels : l'image finale devrait idéalement être la même que si l'on avait divisé chaque voxel en un grand nombre de voxels plus petits, pour les recombiner une fois le post-traitement réalisé.

A.6.1 Cas des pixels carrés et des voxels cubiques

Dans le cas des pixels carrés et des voxels cubiques, la question qui se pose est de savoir si l'on inclut ou non les voisins de deuxième et de troisième ordre —c'est-à-dire ceux ne partageant qu'une arête ou un sommet— dans le voisinage.

Pour le savoir, imaginons une image ne possédant qu'un voxel négatif entouré de voxels positifs dont l'intensité serait suffisamment grande pour qu'ils restassent strictement positifs après le post-traitement. Divisons maintenant chacun de ses voxels en un grand nombre de petits voxels, cubiques eux aussi, en supposant une distribution uniforme de l'activité dans les pixels originaux. Appliquons le NNEPPS à l'image et recombinons ses pixels pour qu'ils retrouvent leur forme initiale. Considérons maintenant la matrice $\boldsymbol{H}$ qu'il aurait fallu pour obtenir le même résultat, directement à partir de l'image initiale. Nous observerions alors que lorsque la taille des voxels intermédiaires tendrait vers 0, la matrice $\boldsymbol{H}$ tendrait à ne considérer que les voisins de premier ordre, quel que soit le choix fait lors du post-traitement de l'image « augmentée », et ce pour deux raisons :

1. Le nombre de « petits » voxels ayant la possibilité de transférer sa valeur aux « grands » voisins de deuxième ordre serait infinitésimal en comparaison de ceux pouvant transférer leur valeur aux « grands » voisins de premier ordre.
2. Ces voxels seraient de plus éloignés du centre et auraient par conséquent un α_i plus faible.

Cela suggère que seuls les voisins de premier ordre doivent être pris en compte.

A.6.2 Cas des pixels rectangulaires et des voxels parallélépipédiques

Nous avons réalisé la même expérience en choisissant cette fois les voxels initiaux parallélépipédiques. Nous choisîmes les petits voxels cubiques afin que la matrice laplacienne de l'image augmentée fût définie sans ambiguïté, selon les conclusions de la section précédente. La simulation numérique nous montra que pour les images TEP que nous dûmes considérer durant l'étude, avec des voxels de taille 4,072 83 mm × 4,072 83 mm × 2,027 mm, une augmentation unitaire d'un voxel devait réduire ses voisins axiaux de 0,282 236 et ses voisins latéraux de 0,108 882.

A.6.3 Remarques conclusives

Le lecteur doit garder à l'esprit que même en choisissant les coefficients ci-dessus, comme nous l'avons fait, le problème original ne devient pas équivalent au problème « augmenté ». Bien que le raisonnement ci-dessus donne un premier essai de choix raisonné des coefficients de $\boldsymbol{H}$, une étude complète doit encore être réalisée à ce sujet, et est laissé à de travaux ultérieurs.

Annexe B

Annexe au chapitre 3

Sommaire

B.1 Normalisation des hyper-paramètres de reconstruction . 157

B.2 Hyper-paramètres parcourus lors qu parcours initial . . 158

Lors de l'optimisation des hyper-paramètres, un premier parcours de l'espace des hyper-paramètres est effectué, afin d'initialiser l'algorithme de recherche globale (section 4.2.2.4). Ce premier parcours est également utilisé pour normaliser les objectifs avant de les combiner (section 4.2.2.3). Cette annexe vise à détailler les hyper-paramètres évalués lors de ce premier quadrillage de l'espace des hyper-paramètres.

B.1 Normalisation des hyper-paramètres de reconstruction

Le facteur de calibration (calib) est un facteur d'échelle qui permet de convertir les comptes de l'espace des données en une image exprimée dans l'unité souhaitée. Il dépend de la durée d'acquisition, du rapport d'embranchement de l'isotope utilisé, de sa demi-vie et d'un étalonnage effectué chaque matin sur le scanner en question. Pour des acquisitions effectuées dans la même journée, de la même durée et avec le même isotope, ce paramètre est donc fixe. Ainsi, dans le cadre de cette étude, il est cohérent de s'affranchir de ce paramètres et de demander au réseau de donner des grandeurs indépendantes de ce paramètre.

Or, le programme de simulation utilisé calcule un facteur de calibration fonction de l'image d'entrée et du nombre de coïncidences demandées. Le facteur de calibration ainsi calculé varie donc d'un facteur 15 suivant les images, ce qui ne correspond pas à la réalité clinique.

Il est cependant possible de s'affranchir de ce facteur de calibration en effectuant les opérations suivantes, qui laissent le problème strictement équivalent :

— On remplace la borne inférieure A par A/calib

— On remplace l'intensité de la pénalité β par $\beta \times \text{calib}^2$

— On divise la valeur initiale de l'image par calib.

— On divise l'image théorique par calib.

— On prend calib = 1.

Une fois ces opérations effectuées, le problème est dit *non calibré*.

Table B.1 – Liste de tous les paramètre testés

Intensité de la pénalité β (quadratique)	10^{-22} $10^{-6,5}$ $10^{-4,5}$ $10^{-2,5}$ $10^{-2,2}$ 10^{-2} 10^{-1} 1 10^{1} 10^{2} 10^{4} 10^{6}
Borne inférieure A ($\mathrm{Bq\,mL^{-1}}$)	10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 1 10^{2} $10^{2,5}$ 10^{3} $10^{3,5}$ 10^{4} 10^{6}
nombre d'itérations	toutes les valeurs de 1 à 30
NNEPPS	avec et sans

B.2 Hyper-paramètres parcourus lors qu parcours initial

La table B.1 décrit tous les hyper-paramètres parcourus lors du parcours initial de l'espace des hyper-paramètres. Les hyper-paramètres indiqués sont les hyper-paramètres non calibrés, tels qu'expliqués à la section précédente.

Bibliographie

[nnd,] Interactive Chart of Nuclides.

[Abouhawwash and Alessio, 2021] Abouhawwash, M. and Alessio, A. M. (2021). Multi-Objective Evolutionary Algorithm for PET Image Reconstruction : Concept. *IEEE Transactions on Medical Imaging*, 40(8) :2142–2151.

[Ahn and Fessler, 2003] Ahn, S. and Fessler, J. A. (2003). Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Transactions on Medical Imaging*, 22(5) :613–626.

[Andersen and Kirsch, 1996] Andersen, A. and Kirsch, J. (1996). Analysis of noise in phase contrast MR imaging. *Medical Physics*, 23(6) :857–869.

[Arridge et al., 2019] Arridge, S., Maass, P., Öktem, O., and Schönlieb, C.-B. (2019). Solving inverse problems using data-driven models. *Acta Numerica*, 28 :1–174.

[Atkinson, 1988] Atkinson, K. A. (1988). *An introduction to numerical analysis*. John Wiley and Sons, second edition.

[Audet and Dennis Jr, 2002] Audet, C. and Dennis Jr, J. E. (2002). Analysis of generalized pattern searches. *SIAM Journal on optimization*, 13(3) :889–903.

[Avriel, 2003] Avriel, M. (2003). *Nonlinear Programming : Analysis and Methods*. Dover Publishing.

[Ayyachamy et al., 2019] Ayyachamy, S., Alex, V., Khened, M., and Krishnamurthi, G. (2019). Medical image retrieval using Resnet-18. In *Medical Imaging 2019 : Imaging Informatics for Healthcare, Research, and Applications*, volume 10954, page 1095410. International Society for Optics and Photonics.

[Bailey et al., 2005] Bailey, D. L., Townsend, D. W., Valk, P. E., and Maisey, M. N. (2005). *Positron emission tomography*. Springer.

[Bland et al., 2019] Bland, J., Mehranian, A., Belzunce, M. A., Ellis, S., da Costa-Luis, C., McGinnity, C. J., Hammers, A., and Reader, A. J. (2019). Intercomparison of MR-informed PET image reconstruction methods. *Medical physics*, 46(11) :5055–5074.

[Bompard et al., 2010] Bompard, M., Peter, J., and Desideri, J.-A. (2010). Surrogate models based on function and derivative values for aerodynamic global optimization. In *V European Conference on Computational Fluid Dynamics ECCOMAS CFD 2010*.

[Bousse et al., 2020] Bousse, A., Courdurier, M., Émond, E., Thielemans, K., Hutton, B. F., Irarrazaval, P., and Visvikis, D. (2020). PET Reconstruction With Non-Negativity Constraint in Projection Space : Optimization Through Hypo-Convergence. *IEEE Transactions on Medical Imaging*, 39(1) :75–86.

[Boyd et al., 2011] Boyd, S., Parikh, N., and Chu, E. (2011). *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc.

[Byrne, 1998] Byrne, C. (1998). Iterative algorithms for deblurring and deconvolution with constraints. *Inverse Problems*, 14(6) :1455–1467.

[Carlier et al., 2015] Carlier, T., Willowson, K. P., Fourkal, E., Bailey, D. L., Doss, M., and Conti, M. (2015). ^{90}Y-PET imaging : Exploring limitations and accuracy under conditions of low counts and high random fraction. *Med. Phys.*, 42(7) :4295–4309.

[Carrillo et al., 2021] Carrillo, H., Millardet, M., Carlier, T., and Mateus, D. (2021). Low-count PET image reconstruction with Bayesian inference over a Deep Prior. In *Medical Imaging 2021 : Image Processing*, volume 11596, page 115960V. International Society for Optics and Photonics.

[Chung and Graham, 1997] Chung, F. R. and Graham, F. C. (1997). *Spectral graph theory*. Number 92. American Mathematical Soc.

[Collette and Siarry, 2004] Collette, Y. and Siarry, P. (2004). *Multiobjective optimization : principles and case studies*. Springer Science & Business Media.

[Dantzig, 1948] Dantzig, G. B. (1948). Programming in a linear structure. *U. S. Air Force Comptroller, USAF, Washington, D.C.*

[Dantzig, 1951] Dantzig, G. B. (1951). Maximization of a linear function of variables subject to linear inequalities. *Activity Analysis of Production and Allocation, New York*, page 339–347.

[Dantzig et al., 1955] Dantzig, G. B., Orden, A., and Wolfe, P. (1955). The generalized simplex method for minimizing a linear form under linear inequality restraints. *Pacific Journal of Mathematics*, (5) :183–195.

[Davins-Valldaura, 2016] Davins-Valldaura, J. (2016). *Surveillance et estimation de l'état de véhicules automobiles : un champ d'applications pour l'optimisation multi-objectif.* PhD thesis, Ecole Centrale de Nantes.

[Davins-Valldaura et al., 2017] Davins-Valldaura, J., Moussaoui, S., Pita-Gil, G., and Plestan, F. (2017). ParEGO extensions for multi-objective optimization of expensive evaluation functions. *Journal of Global Optimization*, 67(1) :79–96.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1) :1–22.

[D'Arienzo, 2013] D'Arienzo, M. (2013). Emission of β^+ Particles Via Internal Pair Production in the $0^+ - 0^+$ Transition of ^{90}Zr : Historical Background and Current Applications in Nuclear Medicine Imaging. *Atoms*, (1) :2–12.

[Elschot et al., 2011] Elschot, M., Nijsen, J. F. W., Dam, A. J., and de Jong, H. W. A. M. (2011). Quantitative evaluation of scintillation camera imaging characteristics of isotopes used in liver radioembolization. *PLoS One*.

[Erlandsson et al., 2000] Erlandsson, K., Visvikis, D., Waddington, W., Cullum, I., Jarritt, P., and Polowsky, L. (2000). Low-statistics reconstruction with AB-EMML. In *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No. 00CH37149)*, volume 2, pages 15–249. IEEE.

[Ford, 1955] Ford, K. W. (1955). Predicted 0+ level in ^{90}Zr. *Physical Review*, 98 :1516–1517.

[Gong et al., 2017] Gong, K., Cheng-Liao, J., Wang, G., Chen, K. T., Catana, C., and Qi, J. (2017). Direct patlak reconstruction from dynamic PET data using the kernel method with MRI information based on structural similarity. *IEEE transactions on medical imaging*, 37(4) :955–965.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning.* MIT Press. `http://www.deeplearningbook.org`.

[Gordon et al., 1970] Gordon, R., Bender, R., and Herman, G. T. (1970). Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3) :471–481.

[Greenberg and Deutsch, 1940] Greenberg, J. S. and Deutsch, M. (1940). Internal Pair Production in Radium C'. *Physical Review*, 58 :714–715.

[Grenader, 1981] Grenader, U. (1981). *Abstract inference*. Wiley.

[Gudbjartsson and Patz, 1995] Gudbjartsson, H. and Patz, S. (1995). The Rician distribution of noisy MRI data. *Magnetic resonance in medicine*, 34(6) :910–914.

[Häggström et al., 2019] Häggström, I., Schmidtlein, C. R., Campanella, G., and Fuchs, T. J. (2019). DeepPET : A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Medical image analysis*, 54 :253–262.

[He et al., 2016] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[Hestenes and Stiefel, 1952] Hestenes, M. R. and Stiefel, E. (1952). Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49(6) :409–436.

[Hudson et al., 1992] Hudson, H. M., Hutton, B. F., and Larkin, R. (1992). Accelerated EM reconstruction using ordered subsets. *J Nucl Med*, 33 :960.

[Hudson and Larkin, 1994] Hudson, H. M. and Larkin, R. (1994). Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imag.*, 13 :601–609.

[Johnson et al., 1955] Johnson, O. E., Johnson, R. G., and Langer, L. M. (1955). Evidence for a 0+ First Excited State in Zr^{90}. *Physical Review*, 98 :1517–1518.

[Kantorovitch, 1942] Kantorovitch, L. V. (1942). On translation of mass. *Proceedings of the USSR Academy of Sciences*, (37) :199–201.

[Karmarkar, 1984] Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. *Combinatorics*, (4) :373–395.

[Khachiyan, 1979] Khachiyan, L. G. (1979). A polynomial algorithm in linear programming. *Soviet Mathematics Doklady*, (20) :191–194.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*.

[Klee and Minty, 1972] Klee, V. and Minty, G. J. (1972). How good is the simplex algorithm ? *in Inequalities, O. Shisha, ed., Academic Press, New York*, pages 159–175.

[Koberstein, 2005] Koberstein, A. (2005). *The Dual Simplex Method, Techniques for a fast and stable implementation.* PhD thesis, Fakultät für Wirtschaftswissenschaften der Universität Paderborn.

[Kolda et al., 2003] Kolda, T. G., Lewis, R. M., and Torczon, V. (2003). Optimization by direct search : New perspectives on some classical and modern methods. *SIAM review*, 45(3) :385–482.

[Krane, 1988] Krane, K. S. (1988). *Introductory nuclear physics.*

[Kullback and Leibler, 1951] Kullback, S. and Leibler, R. A. (1951). On Information and Sufficiency. *Ann. Math. Statist.*, 22(1) :79–86.

[Lange and Carson, 1984] Lange, K. and Carson, R. (1984). EM Reconstruction Algorithms for Emission and Transmission Tomography. *Journal of Computer Assisted Tomography*, 8(2) :306–316.

[Langhoff and Hennies, 1961] Langhoff, H. and Hennies, H. H. (1961). Zum experimentellen Nachweis von Zweiquantenzerfall beim $O^+ - O^+$-Übergang des Zr^{90}. *Zeitschrift für Physik*, 164 :166–173.

[Lehmann and Casella, 2006] Lehmann, E. L. and Casella, G. (2006). *Theory of point estimation.* Springer Science & Business Media.

[Lemke, 1954] Lemke, C. E. (1954). The dual method of solving the linear programming problem. *Naval Research Logistics Quarterly*, (1) :36–47.

[Levillain et al., 2021] Levillain, H., Bagni, O., Deroose, C. M., Dieudonné, A., Gnesin, S., Grosser, O. S., Kappadath, S. C., Kennedy, A., Kokabi, N., Liu, D. M., Madoff, D. C., Mahvash, A., Martinez de la Cuesta, A., Ng, D. C. E., Paprottka, P. M., Pettinato, C., Rodríguez-Fraile, M., Salem, R., Sangro, B., Strigari, L., Sze, D. Y., de Wit van der Veen, B. J., and Flamen, P. (2021). International recommendations for personalised selective internal radiation therapy of primary and metastatic liver diseases with yttrium-90 resin microspheres. *European Journal of Nuclear Medicine and Molecular Imaging*, 48(5) :1570–1584.

[Li et al., 2017] Li, Y., Yang, J., Song, Y., Cao, L., Luo, J., and Li, L.-J. (2017). Learning from noisy labels with distillation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1910–1918.

[Lim et al., 2018] Lim, H., Dewaraja, Y. K., and Fessler, J. A. (2018). A PET reconstruction formulation that enforces non-negativity in projection space for bias reduction in Y-90 imaging. *Phys. Med. Biol.*, 63(3) :35–42.

[Matej et al., 2009] Matej, S., Surti, S., Jayanthi, S., Daube-Witherspoon, M. E., Lewitt, R. M., and Karp, J. S. (2009). Efficient 3-D TOF PET reconstruction using view-grouped histo-images : DIRECT—Direct image reconstruction for TOF. *IEEE Transactions on medical imaging*, 28(5) :739–751.

[Merlin et al., 2018] Merlin, T., Stute, S., Benoit, D., Bert, J., Carlier, T., Comtat, C., Filipovic, M., Lamare, F., and Visvikis, D. (2018). CASToR : a generic data organization and processing code framework for multi-modal and multi-dimensional tomographic reconstruction. *Physics in Medicine & Biology*, 63(18) :185005.

[Miekkala, 1993] Miekkala, U. (1993). Graph properties for splitting with grounded Laplacian matrices. *BIT Numerical Mathematics*, 33(3) :485–495.

[Millardet et al., 2021] Millardet, M., Moussaoui, S., Idier, J., Mateus, D., Conti, M., Bailly, C., Stute, S., and Carlier, T. (2021). A Multi-Objective Comparative Analysis of Reconstruction Algorithms in the Context of Low-Statistics 90Y-PET Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, pages 1–1.

[Millardet et al., 2020] Millardet, M., Moussaoui, S., Mateus, D., Idier, J., and Carlier, T. (2020). Local-Mean Preserving Post-Processing Step for Non-Negativity Enforcement in PET Imaging : Application to ^{90}Y-PET. *IEEE Transactions on Medical Imaging*, 39(11) :3725–3736.

[Monge, 1781] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*, pages 666–704.

[Nagarajan and Nakshatrala, 2011] Nagarajan, H. and Nakshatrala, K. (2011). Enforcing the non-negativity constraint and maximum principles for diffusion with decay on general computational grids. *International Journal for Numerical Methods in Fluids*, 67(7) :820–847.

[Nakshatrala et al., 2016] Nakshatrala, K., Nagarajan, H., and Shabouei, M. (2016). A numerical methodology for enforcing maximum principles and the non-negative constraint for transient diffusion equations. *Communications in Computational Physics*, 19(1) :53–93.

[Nakshatrala and Valocchi, 2009] Nakshatrala, K. and Valocchi, A. J. (2009). Non-negative mixed finite element formulations for a tensorial diffusion equation. *Journal of Computational Physics*, 228(18) :6726–6752.

[Natarajan et al., 2013] Natarajan, N., Dhillon, I. S., Ravikumar, P. K., and Tewari, A. (2013). Learning with noisy labels. *Advances in neural information processing systems*, 26 :1196–1204.

[Nocedal and Wright, 1999] Nocedal, J. and Wright, S. J. (1999). *Numerical Optimization*. Springer.

[Nuyts et al., 2002a] Nuyts, J., Bequé, D., Dupont, P., and Mortelmans, L. (2002a). A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Trans. Nucl. Sci.*, pages 56–60.

[Nuyts et al., 2002b] Nuyts, J., Stroobants, S., Dupont, P., Vleugels, S., Flamen, P., and Mortelmans, L. (2002b). Reducing Loss of Image Quality Because of the Attenuation Artifact in Uncorrected PET Whole-Body Images. *J. Nucl. Med.*, 43(8) :1054–1062.

[Ote and Hashimoto, 2021] Ote, K. and Hashimoto, F. (2021). Deep Learning-Based Fast TOF-PET Image Reconstruction Using Direction Information.

[Panin et al., 2006] Panin, V. Y., Kehren, F., Rothfuss, H., Hu, D., Michel, C., and Casey, M. E. (2006). PET Reconstruction With System Matrix Derived From Point Source Measurements. *IEEE Trans. Nucl. Sci.*, 53(1) :152–159.

[Pasciak et al., 2014] Pasciak, A. S., Bourgeois, A. C., McKinney, J. M., Chang, T. T., Osborne, D. R., Acuff, S. N., and Bradley, Y. C. (2014). Radioembolization and the dynamic role of 90Y PET/CT. *Frontiers in oncology*, 4 :38.

[Paszke et al., 2017] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch.

[Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7) :629–639.

[Peyré and Cuturi, 2019] Peyré, G. and Cuturi, M. (2019). Computational Optimal Transport. *Foundations and Trends in Machine Learning*, 11(5-6) :355–607.

[Pierro, 1995] Pierro, A. R. D. (1995). A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography. *IEEE Trans. Med. Imag.*, 14(1) :132–137.

[Pierro and Yamagishi, 2001] Pierro, A. R. D. and Yamagishi, M. E. B. (2001). Fast EM-Like Methods for Maximum "*A Posteriori*" Estimates in Emission Tomography. *IEEE Trans. Med. Imag.*, 20(4) :280–288.

[Rahmim et al., 2005] Rahmim, A., Cheng, J.-C., Blinder, S., Camborde, M.-L., and Sossi, V. (2005). Statistical dynamic image reconstruction in state-of-the-art high-resolution PET. *Physics in Medicine & Biology*, 50(20) :4887.

[Rasch et al., 2017] Rasch, J., Brinkmann, E.-M., and Burger, M. (2017). Joint reconstruction via coupled Bregman iterations with applications to PET-MR imaging. *Inverse Problems*, 34(1) :014001.

[Reader et al., 2021] Reader, A. J., Corda, G., Mehranian, A., Costa-Luis, C. d., Ellis, S., and Schnabel, J. A. (2021). Deep Learning for PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1) :1–25.

[Reader and Ellis, 2020] Reader, A. J. and Ellis, S. (2020). Bootstrap-optimised regularised image reconstruction for emission tomography. *IEEE transactions on medical imaging*, 39(6) :2163–2175.

[Rowley et al., 2017] Rowley, L. M., Bradley, K. M., Boardman, P., Hallam, A., and McGowan, D. R. (2017). Optimization of image reconstruction for 90Y selective internal radiotherapy on a lutetium yttrium orthosilicate PET/CT system using a Bayesian penalized likelihood reconstruction algorithm. *Journal of Nuclear Medicine*, 58(4) :658–664.

[Scott and McGowan, 2019] Scott, N. P. and McGowan, D. R. (2019). Optimising quantitative 90 Y PET imaging : an investigation into the effects of scan length and Bayesian penalised likelihood reconstruction. *EJNMMI research*, 9(1) :40.

[Selwyn et al., 2007] Selwyn, R. G., Nickles, R. J., Thomadsen, B., DeWerd, L. A., and Micka, J. A. (2007). A new internal pair production branching ratio of 90Y : The development of a non-destructive assay for 90Y and 90Sr. *Applied Radiation and Isotopes*, 65(3) :318–327.

[Seo et al., 2019] Seo, Y., Khalighi, M. M., Wangerin, K. A., Deller, T. W., Wang, Y.-H., Jivan, S., Kohi, M. P., Aggarwal, R., Flavell, R. R., Behr, S. C., et al. (2019). Quantitative and Qualitative Improvement of Low-Count [68 Ga] Citrate and [90 Y] Microspheres PET Image Reconstructions Using Block Sequential Regularized Expectation Maximization Algorithm. *Molecular Imaging and Biology*, pages 1–9.

[Shepp and Vardi, 1982] Shepp, L. A. and Vardi, Y. (1982). Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Trans. Med. Imag.*, 1(2) :113–122.

[Shipley et al., 1979] Shipley, W. U., Tepper, J. E., Prout, George R., J., Verhey, L. J., Mendiondo, O. A., Goitein, M., Koehler, A. M., and Suit, H. D. (1979). Proton Radiation as Boost Therapy for Localized Prostatic Carcinoma. *JAMA*, 241(18) :1912–1915.

[Siman et al., 2019] Siman, W., Mikell, J. K., Mawlawi, O. R., Mourtada, F., and Kappadath, S. C. (2019). Dose volume histogram-based optimization of image reconstruction parameters for quantitative 90Y-PET imaging. *Medical Physics*, 46(1) :229–237.

[Strydhorst et al., 2016] Strydhorst, J., Carlier, T., Dieudonné, A., Conti, M., and Buvat, I. (2016). A gate evaluation of the sources of error in quantitative90Y PET. *Medical physics*, 43(10) :5320–5329.

[Stute and Comtat, 2013] Stute, S. and Comtat, C. (2013). Practical considerations for image-based PSF and blobs reconstruction in PET. *Phys. Med. Biol.*, 58 :3849–3870.

[Stute et al., 2015] Stute, S., Tauber, C., Leroy, C., Bottlaender, M., Brulon, V., and Comtat, C. (2015). Analytical simulations of dynamic PET scans with realistic count rates properties. In *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–3.

[Sukhbaatar et al., 2014] Sukhbaatar, S., Bruna, J., Paluri, M., Bourdev, L., and Fergus, R. (2014). Training convolutional networks with noisy labels. *arXiv preprint arXiv :1406.2080*.

[Targ et al., 2016] Targ, S., Almeida, D., and Lyman, K. (2016). Resnet in resnet : Generalizing residual architectures. *arXiv preprint arXiv :1603.08029*.

[Van Slambrouck et al., 2015] Van Slambrouck, K., Stute, S., Comtat, C., Sibomana, M., van Velden, F. H. P., Boellaard, R., and Nuyts, J. (2015). Bias Reduction for Low-Statistics PET : Maximum Likelihood Reconstruction With a Modified Poisson Distribution. *IEEE Transactions on Medical Imaging*, 34(1) :126–136.

[Van Velden et al., 2009] Van Velden, F., Comtat, C., Nuyts, J., Reilhac, A., Lammertsma, A., and Boellaard, R. (2009). Evaluation of a 3D iterative reconstruction

algorithm that allows for negative image values in high resolution PET studies. *Journal of cerebral blood flow and metabolism*, 29 :S3–S3.

[Verhaeghe and Reader, 2010] Verhaeghe, J. and Reader, A. J. (2010). AB-OSEM reconstruction for improved Patlak kinetic parameter estimation : a simulation study. *Physics in Medicine & Biology*, 55(22) :6739.

[Wang et al., 2004] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment : from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4) :600–612.

[Weisstein,] Weisstein, E. W. Laplacian Matrix.

[Whiteley et al., 2021] Whiteley, W., Panin, V., Zhou, C., Cabello, J., Bharkhada, D., and Gregor, J. (2021). FastPET : Near Real-Time Reconstruction of PET Histo-Image Data Using a Neural Network. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(1) :65–77.

[Willowson et al., 2015] Willowson, K. P., Tapner, M., and Bailey, D. L. (2015). A multicentre comparison of quantitative ^{90}Y PET/CT for dosimetric purposes after radioembolization with resin microspheres. *Eur J Nucl Med Mol Imaging*, (42) :1202–1222.

[Ye et al., 2018] Ye, D. H., Buzzard, G. T., Ruby, M., and Bouman, C. A. (2018). Deep back projection for sparse-view CT reconstruction. In *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1–5. IEEE.

[Zhu et al., 2018] Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R., and Rosen, M. S. (2018). Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697) :487–492.

[Zitzler et al., 2001] Zitzler, E., Thiele, L., Deb, K., Coello Coello, C. A., and Corne, D., editors (2001). *Evolutionary Multi-Criterion Optimization*. Springer Berlin Heidelberg.

CENTRALE NANTES

Titre : Amélioration de la quantification des images TEP à l'yttrium 90

Mots clés : TEP, algorithmes de reconstruction tomographique, yttrium 90, optimisation linéaire, analyse multi-objectifs

Résumé : La popularité de l'imagerie TEP à l'yttrium 90 va grandissante. Cependant, la probabilité qu'une désintégration d'un noyau d'yttrium 90 mène à l'émission d'un positon n'est que de $3{,}2 \times 10^{-5}$, et les images reconstruites sont par conséquent caractérisées par un niveau de bruit élevé, ainsi que par un biais positif dans les régions de faible activité. Pour corriger ces problèmes, les méthodes classiques consistent à utiliser des algorithmes pénalisés, ou autorisant des valeurs négatives dans l'image. Cependant, une étude comparant et combinant ces différentes méthodes dans le contexte spécifique de l'yttrium 90 manquait encore à l'appel au début de cette thèse. Cette dernière vise donc à combler ce manque. Malheureusement, les méthodes autorisant les valeurs négatives ne peuvent pas être utilisées directement dans le cadre d'une étude dosimétrique, et cette thèse commence donc par proposer une nouvelle méthode de post-traitement des images, visant à en supprimer les valeurs négatives en en conservant les valeurs moyennes le plus localement possible. Une analyse complète multi-objectifs de ces différentes méthodes est ensuite proposée. Cette thèse se termine en posant les prémices de ce qui pourra devenir un algorithme permettant de proposer un jeu d'hyper-paramètres de reconstruction adéquats, à partir des seuls sinogrammes.

Title: Improvement of the quantification of yttrium-90 PET images

Keywords: PET, tomographic reconstruction algorithms, yttrium-90, linear programmation, multi-objective analysis

Abstract: Yttrium-90 PET imaging is becoming increasingly popular. However, the probability that decay of a yttrium-90 nucleus will lead to the emission of a positron is only 3.2×10^{-5}, and the reconstructed images are therefore characterised by a high level of noise, as well as a positive bias in low activity regions. To correct these problems, classical methods use penalised algorithms or allow negative values in the image. However, a study comparing and combining these different methods in the specific context of yttrium-90 was still missing at the beginning of this thesis. This thesis, therefore, aims to fill this gap. Unfortunately, the methods allowing negative values cannot be used directly in a dosimetric study. Therefore, this thesis starts by proposing a new method of post-processing the images, aiming to remove the negative values while keeping the average values as locally as possible. A complete multi-objective analysis of these different methods is then proposed. This thesis ends by laying the foundations of what could become an algorithm providing a set of adequate reconstruction hyper-parameters from sinograms alone.