



HAL
open science

Méthodes informées de factorisation matricielle pour l'étalonnage de réseaux de capteurs mobiles et la cartographie de champs de pollution

Clément Dorffer

► **To cite this version:**

Clément Dorffer. Méthodes informées de factorisation matricielle pour l'étalonnage de réseaux de capteurs mobiles et la cartographie de champs de pollution. Traitement du signal et de l'image [eess.SP]. Université du Littoral Côte d'Opale, 2017. Français. NNT : 2017DUNK0507 . tel-02121576

HAL Id: tel-02121576

<https://theses.hal.science/tel-02121576v1>

Submitted on 6 May 2019

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

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



THÈSE

Présentée et soutenue publiquement le 13 Décembre 2017 par :
CLÉMENT DORFFER

En vue de l'obtention du doctorat délivré par :
l'Université du Littoral Côte d'Opale (ULCO)

Méthodes informées de factorisation matricielle pour l'étalonnage de réseaux de capteurs mobiles et la cartographie de champs de pollution

sous la direction de

GILLES ROUSSEL
MATTHIEU PUIGT

Professeur d'Université
Maître de Conférences

JURY

SAID MOUSSAOUI

Professeur d'Université

Président

NADÈGE THIRION-MOREAU

Professeur d'Université

Rapporteur

HERVÉ RIVANO

Professeur d'Université

Rapporteur

ANGÉLIQUE DRÉMEAU

Maître de Conférences

Examinateur

ANDA IONESCU

Maître de Conférences

Examinateur

MATTHIEU PUIGT

Maître de Conférences

Encadrant

GILLES ROUSSEL

Professeur d'Université

Dir. de thèse

GILLES DELMAIRE

Maître de Conférences

Invité

École doctorale :

Sciences pour l'ingénieur

Spécialité :

Automatique, Génie informatique, Traitement du Signal et des Images

Unité de Recherche :

Laboratoire d'Informatique Signal et Image de la Côte d'Opale, EA 4491

Remerciements

Avant tout, je souhaite remercier Messieurs Gilles Roussel et Matthieu Puigt respectivement directeur et encadrant de cette thèse. Vous m'avez fait confiance depuis notre premier entretien (le 8 octobre 2014, jour orageux de fête de la science) et avez su me pousser jusqu'à l'aboutissement de ce travail. J'adresse une mention spéciale à Matthieu pour les quelques nuits de rédaction/*brainstorming*/café+dattes passées dans le froid au labo, MERCI!

Merci au Pôle Métropolitain de la Côte d'Opale (PMCO) et à l'Université du Littoral Côte d'Opale (ULCO) qui ont rendu ce travail possible en y apportant leur soutien financier. Merci à la Région Hauts-de-France qui a financé le projet OSCAR et merci également aux différents partenaires de ce projet (ATMO Hauts-de-France, BES et l'équipe-projet Spirals de l'Inria) pour la qualité de nos échanges durant ce projet.

Je tiens à remercier Madame Nadège Thirion-Moreau et Monsieur Hervé Rivano qui ont accepté d'être mes rapporteurs et de juger ce manuscrit de thèse. Merci également à Madame Angélique Drémeau, Madame Anda Ionescu et Monsieur Saïd Moussaoui d'avoir accepté de faire parti de mon jury. L'ensemble des remarques/questions/conseils que vous avez formulé a été particulièrement apprécié.

Merci à Gilles Delmaire, pour les échanges toujours très intéressants que nous avons pu avoir et pour m'avoir fait profiter de son expertise en factorisation matricielle.

Les expériences réalisées dans ce document ont été réalisées sur la plate-forme de calcul scientifique CALCULCO, gérée par le SCoSI de l'ULCO que je remercie pour la mise en place et la maintenance (réactive) de cet outil.

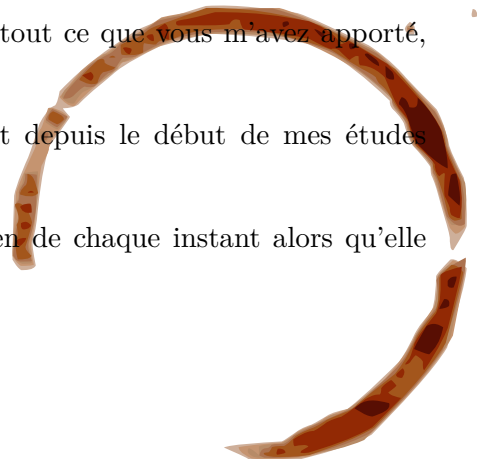
Merci aux membres du LISIC pour leur accueil, les 61 (*of course*) mais aussi les 27. Un remerciement particulier à, par ordre alphabétique : André, Christophe, Cyril, Dominique, Emilie, Eric, François, Gaëlle, Isabelle, Mahmoud, Mourad, Nicolas, Patrick, Philippe, Pierre-Alexandre, Rémi, Sébastien et Virginie.

Merci aux doctorants et en particulier à : Amaury, Ahmad, Baptiste, Emna, Florian, Florian.

Un grand MERCI à Fabien (+ Tressy) et Julien pour tout ce que vous m'avez apporté, votre amitié et votre bienveillance.

Je remercie mes parents pour tout leur investissement depuis le début de mes études jusqu'à cette fin heureuse.

Enfin je remercie ma compagne Chloé pour son soutien de chaque instant alors qu'elle



même travaillait sur sa thèse.

Table des matières

Table des figures	x
Liste des tableaux	xi
Liste des algorithmes	xiii
Table des sigles et acronymes	xv
1 Introduction générale	1
1.1 Contexte	1
1.2 Objectifs	3
1.3 Plan	4
2 État de l’art en étalonnage de capteurs	5
2.1 Introduction	5
2.2 Modèles d’étalonnage de capteurs	5
2.3 Méthodes d’étalonnage	7
2.3.1 Les approches de type régression linéaire/non-linéaire	7
2.3.2 Étalonnage par apprentissage/correction	8
2.3.3 Les approches non-supervisées ou faiblement supervisées	9
2.3.3.1 Macro-étalonnage	10
2.3.3.2 Micro-étalonnage	13
2.4 Conclusion	14
I Le problème de l’étalonnage de capteurs vu comme un problème de	

factorisation matricielle	17
3 État de l'art en factorisation matricielle	19
3.1 Introduction	19
3.2 Critères et contraintes usuelles en NMF	20
3.2.1 Critères usuels d'attache aux données	20
3.2.1.1 La norme de Frobenius	21
3.2.1.2 Divergence de Kullback-Leibler	21
3.2.1.3 Divergence d'Itakura-Saito	22
3.2.1.4 Les divergences paramétriques	22
3.2.1.5 Critère pondéré	24
3.2.1.6 Prise en compte de structures dans le critère	24
3.2.2 Les contraintes vues comme des pénalisations	26
3.2.2.1 Pénalisations d'évolution lente	26
3.2.2.2 Pénalisations de lissage	27
3.2.2.3 Pénalisations de parcimonie	27
3.2.2.4 Pénalisations douces d'égalité	28
3.2.2.5 Pénalisations de faible rang	28
3.2.3 Autres pénalisations : contraintes de bornitude	29
3.2.4 Unicité et exactitude de la NMF	30
3.3 Optimisations usuelles pour les algorithmes alternés	30
3.3.1 Méthode heuristique	31
3.3.2 Descente de gradient	32
3.3.3 Approche Majoration-Minimisation (MM)	32
3.3.4 Méthode par moindres carrés alternés (ANLS)	33
3.4 Méthodes "historiques" de NMF	34
3.5 Méthodes de NMF pour les problèmes de grandes dimensions	36

3.5.1	Méthodes de NMF accélérées	36
3.5.1.1	Méthode de NMF par mises à jour itératives alternées	36
3.5.1.2	NMF par gradient de Nesterov	36
3.5.1.3	Méthode de NMF aléatoire (<i>Randomized</i> NMF)	37
3.5.1.4	NMF distribuée	37
3.5.2	NMF pour flux de données	38
3.5.3	Co-factorisation	38
3.5.4	Factorisation tensorielle	39
3.6	Cas des données manquantes	40
3.7	Conclusion	40
4	Méthodes de NMF informées pour l'étalonnage d'un réseau de capteurs mobiles à réponse affine	43
4.1	Définitions et hypothèses	43
4.2	Mise en place de l'étalonnage par factorisation	45
4.2.1	Écriture matricielle du problème d'étalonnage	45
4.2.2	Prise en compte des données manquantes	46
4.2.3	Contrainte de non-négativité	46
4.2.4	Prise en compte de données de référence	47
4.2.5	Conditions nécessaires à l'étalonnage	48
4.3	Une première approche d'étalonnage par complétion de matrice (NMF/C-Cal)	49
4.4	Étalonnage de capteurs par factorisation matricielle informée	50
4.5	Régularisation par utilisation de données constructeur	52
4.6	Initialisation des méthodes	56
4.7	Conclusion	57
5	Méthodes de Semi-NMF informées pour l'étalonnage d'un réseau de capteurs mobiles non-linéaires	59

5.1	Hypothèses et introduction de structures dans les matrices facteurs	59
5.2	Étalonnage par complétion	61
5.3	Étalonnage par factorisation matricielle informée	61
5.3.1	Mise à jour de la matrice de Vandermonde	62
5.3.2	Mise à jour du deuxième facteur	64
5.4	Extension à la régularisation par utilisation de données constructeur	65
5.5	Initialisation et astuces algorithmiques	66
5.6	Conclusion	67
6	Étude de performances	69
6.1	Conditions de simulation	69
6.1.1	Sélection des capteurs	69
6.1.1.1	Pour le modèle affine	70
6.1.1.2	Pour le modèle non-linéaire	72
6.1.2	Simulation d'une scène et d'une matrice de données	72
6.1.3	Discussion sur les paramètres et les critères de performance	73
6.2	Performances sur le modèle affine	74
6.3	Performances sur le modèle non-linéaire	81
6.4	Conclusion	86
II	De l'étalonnage à la reconstruction de champs physiques	87
7	Raffinement de l'étalonnage et extension à la cartographie par ajout d'une régularisation spatiale	89
7.1	État de l'art en estimation de données manquantes	90
7.1.1	Les approches d'estimation de données discrètes	90
7.1.1.1	Interpolation aux K plus proches voisins (KNN)	90
7.1.1.2	Approche barycentrique d'estimation de donnée	91

7.1.1.3	Estimation de données par <i>a priori</i> de parcimonie	91
7.1.2	Les approches basées sur l'analyse de données fonctionnelles	93
7.1.2.1	Estimation de données par régression	93
7.1.2.2	Interpolation par calage de fonctions	94
7.1.3	Méthodes d'estimation de données à base de modèle	95
7.1.3.1	Méthodes basées sur la mécanique des fluides	95
7.1.3.2	Les modèles intégraux	95
7.1.3.3	Le modèle analytique gaussien	96
7.1.3.4	Modèle LUR	96
7.1.4	Les approches géostatistiques	97
7.2	Étalonnage par factorisation matricielle avec contrainte de décomposition parcimonieuse	100
7.3	Régularisation par introduction d'un modèle	103
7.3.1	Modèle polynômial	104
7.3.2	Utilisation d'un modèle géostatistique : le Krigeage	106
7.4	Conclusion	108
8	Étude de performances des approches régularisées	109
8.1	Illustration de l'apport d'une régularisation spatiale	109
8.1.1	Comparaison des approches SpIN-Cal et IN-Cal	110
8.1.2	Performances globales de SpIN-Cal et influence du paramètre de régularisation	114
8.2	Comparaison des approches régularisées sur un nouveau jeu de données	115
8.2.1	Simulations des données	115
8.2.2	Étude de performances des approches IN-Cal, SpIN-Cal, PoIN-Cal et KIN-Cal	117
8.3	Conclusion	120
9	Conclusion générale	123

9.1 Conclusion	123
9.2 Perspectives	124
A Méthode robuste d'étalonnage de capteurs fixes	127
B Méthode robuste de NMF à base de divergence paramétrique et de contrainte de somme à un	135
C Méthode accélérée de NMF avec données manquantes	143
Bibliographie	172

Table des figures

2.1	Principe d'acquisition d'images par système <i>push broom</i> . Source : [25].	12
3.1	Illustration du principe des méthodes MM.	33
4.1	D'une scène \mathcal{S} vers une matrice de données X	44
6.1	Illustration du fonctionnement d'un capteur de particules. Source [203].	70
6.2	(a) : un capteur de particules <i>Sharp GP2Y1010AU0F</i> . (b) : un projet amateur de mesure de la qualité de l'air.	71
6.3	Un exemple de simulation de paramètres de capteurs à réponses affines.	71
6.4	(a) : un capteur de particules <i>stbm 271</i> . (b) : courbe d'étalonnage théorique du <i>stbm 271</i> , source [198].	72
6.5	Scène simulée.	73
6.6	Performances globales de : (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.	75
6.7	Exemples de matrices de données X construites avec (a) 80% de données manquantes et (b) 10% de données manquantes.	75
6.8	Performances par rapport à la proportion de données manquantes ($\rho_{RV} = 0.3, \text{SNR} = \infty$) pour (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.	76
6.9	Performances par rapport à la proportion de rendez-vous ($\rho_{DM} = 0.9, \text{SNR} = \infty$) pour (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.	78
6.10	Performances par rapport au SNR d'entrée ($\rho_{DM} = 0.9, \rho_{RV} = 0.3$) pour (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.	79
6.11	Influence des paramètres de régularisation dans la méthode ACIN-Cal avec différents niveau d'erreur sur les paramètres théoriques. À gauche avec une erreur de l'ordre de $1.e - 6$ et à droite de l'ordre de 1.	79
6.12	Comparaison des performances de IN-Cal (en bleue) vs. l'étalonnage multi-sauts [182] (en rouge) dans (a) un cas sans bruit et (b) avec un SNR d'entrée d'environ 30 dB.	82
6.13	Performances globales de : (a) SC-Isn-Cal, (c) ASC-Isn-Cal.	82

6.14	Performances par rapport à la proportion de données manquantes ($\rho_{RV} = 0.3$, $SNR=\infty$) pour (a) SC-ISN-Cal, (b) ASC-ISN-Cal.	84
6.15	Performances par rapport à la proportion de rendez-vous ($\rho_{DM} = 0.9$, $SNR=\infty$) pour (a) SC-ISN-Cal, (c) ASC-ISN-Cal.	85
6.16	Performances par rapport au SNR d'entrée ($\rho_{DM} = 0.9, \rho_{RV} = 0.3$) pour (a) SC-ISN-Cal, (c) ASC-ISN-Cal.	85
7.1	Exemple de diffusion obtenue avec le modèle plume [209].	96
7.2	(d-e-f) Exemples de variogrammes obtenus pour les champs simulés (a-b-c).	99
8.1	Performances des méthodes IN-Cal (en rouge) et SpIN-Cal (en bleue) par rapport à (a) la proportion de données manquantes, (b) la proportion de rendez-vous, (c) le SNR d'entrée.	110
8.2	Illustration de l'accélération de la convergence par analyse des performances des méthodes IN-Cal (en rouge) et SpIN-Cal (en bleue) par rapport aux itérations. Figure extraite de [51].	111
8.3	Scènes reconstruites après étalonnage par les méthodes (a) IN-Cal et (b) SpIN-Cal, à partir d'une matrice de données X avec 95% de données manquantes.	112
8.4	RMSE obtenu sur la seconde colonne de la matrice G par les méthodes (a) IN-Cal et (b) SpIN-Cal, calculés par rapport à la proportion de rendez-vous, en rouge sur l'ensemble de la colonne de G et en bleue sur les éléments de G pour lesquels une mesure est disponible.	113
8.5	Performances globales de SpIN-Cal pour différentes valeurs du paramètre de régularisation : (a) $\lambda = 0.01$, (b) $\lambda = 1$, (c) $\lambda = 10$, (d) $\lambda = 100$	114
8.6	Champs physique simulé pour les tests des méthodes PolIN-Cal et KIN-Cal	116
8.7	Performances de la méthode IN-Cal en fonction du nombre de mesures par capteur, avec un pas de discrétisation de la scène fixé à 1.	118
8.8	Performances de l'approche SpIN-Cal en fonction du nombre de mesures par capteur (a) $\lambda = 10$, (b) $\lambda = 100$	118
8.9	Performances d'étalonnage de la méthode PolIN-Cal en fonction du nombre de mesures pas capteur et du paramètre de régularisation (a) $\lambda = 10$, (b) $\lambda = 100$ et (c) λ adaptatif.	119
8.10	Comparaison des performances d'étalonnage des méthode IN-Cal (courbe pointillée) et KIN-Cal (courbe pleine), en fonction du nombre de mesures pas capteur.	120

Liste des tableaux

3.1	Tableau des divergences.	23
3.2	Tableau des méthodes "historiques" de NMF.	35
6.1	Paramètres théoriques du capteur de particules <i>Sharp GP2Y1010AU0F</i>	71

Liste des algorithmes

1	Étalonnage par complétion de X et régression	50
2	Étalonnage par Factorisation Informée en Matrices Non-négatives (IN-Cal) .	52
3	Étalonnage par Factorisation Informée en Matrices Non-négatives sous contrainte douce de normalisation (ACIN-Cal)	56
4	Étalonnage par complétion de X et régression.	62
5	Étalonnage par Factorisation Informée en matrices Semi-Non-négatives avec contrainte de structure de Vandermonde (SC-ISN-Cal).	64
6	Étalonnage par Factorisation Informée en matrices Semi-Non-négatives avec contrainte de structure de Vandermonde et de moyenne (ASC-ISN-Cal). . . .	66
7	Étalonnage par Factorisation Informée en matrices Non-négatives avec contrainte de décomposition parcimonieuse (SpIN-Cal).	103
8	Étalonnage par Factorisation Informée en matrices Non-négatives avec contrainte d'approximation polynômiale (PolIN-Cal).	106
9	Étalonnage par Factorisation Informée en matrices Non-négatives avec complétion par krigeage (KIN-Cal).	108

Table des sigles et acronymes

AASQA	Association Agréée de Surveillance de la Qualité de l’Air
ACI	Analyse en Composantes Indépendantes
ACIN-Cal	(Méthode d’) Étalonnage par factorisation informée en matrices non-négatives avec contraintes de moyennes (<i>Average-Constrained Informed Nonnegative Matrix Factorization for sensor Calibration</i>)
ACPolIN-Cal	(Méthode d’) Étalonnage par factorisation informée en matrices non-négatives avec contrainte d’approximation polynômiale et de moyennes (<i>Average-Constrained PolIN-Cal</i>)
ACP	Analyse en Composantes Principales
ADM	Algorithme des directions alternées (<i>Alternating Direction Method</i>)
ANLS	Moindres carrés non-négatifs alternés (<i>Alternating Nonnegative Least Square</i>)
ASC-ISN-Cal	(Méthode d’) Étalonnage par factorisation informée en matrices semi-non-négatives avec contrainte de structure et de moyennes (<i>Average and Structure-Constrained Informed Semi-Nonnegative matrix factorization for sensor Calibration</i>)
BES	Association Bâisseurs d’Economie Solidaire
CFD	(Modèle de) mécanique des fluides numérique (<i>Computational Fluid Dynamics</i>)
FDA	Méthode d’analyse de données fonctionnelles (<i>Functional Data Analysis</i>)
GMR	Régression moyenne géométrique (<i>Geometric Mean Regression</i>)
HALS	Moindres carrés hiérarchiques alternés (<i>Hierarchical Alternating Least Square</i>)
ICA	Analyse en composantes indépendantes (<i>Independent Component Analysis</i>)
IN-Cal	(Méthode d’) Étalonnage par factorisation informée en matrices non-négatives (<i>Informed Nonnegative Matrix Factorization for sensor Calibration</i>)
IoT	Internet des objets (<i>Internet of Things</i>)
IS	Itakura-Saito
KL	Kullback-Leibler
KNN	K-Nearest Neighbor
KIN-Cal	(Méthode d’) Étalonnage par Factorisation Informée en matrices Non-négatives avec completion par krigeage (<i>Kkrigging Informed Nonnegative Matrix Factorization for sensor Calibration</i>)
LUR	Modèle de cartographie basé sur des informations de trafic (<i>Land Use Regression</i>)

MM	Majoration/Minimisation (méthode)
NeNMF	Factorisation en matrices non-négatives utilisant le gradient de Nesterov (<i>Nesterov based Nonnegative Matrix Factorization</i>)
NMF	Factorisation en matrices non-négatives (<i>Nonnegative Matrix Factorization</i>)
NMF/C	Complétion par Factorisation en matrices non-négatives (<i>Nonnegative Matrix Factorization/Completion</i>)
NMF/C-Cal	(Méthode d') Etalonnage par NMF/C (<i>NMF/C for sensor Calibration</i>)
NTF	Factorisation en tenseurs non-négatifs (<i>Nonnegative Tensor Factorization</i>)
OMP	Méthode d'approximation parcimonieuse (<i>Orthogonal Matching Pursuit</i>)
OSCAR	O bservation et S ensibilisation C itoyenne à la qualité de l' A ir en R égion
PM1	Particules de moins de 1 μm
PM10	Particules de moins de 10 μm
PM2.5	Particules de moins de 2.5 μm
PMF	Factorisation en matrices positives (<i>Positive Matrix Factorization</i>)
PolIN-Cal	(Méthode d') Etalonnage par factorisation informée en matrices non-négatives avec contrainte d'approximation polynômiale (<i>Polynomial-IN-Cal</i>)
RMSE	Critère de performances (<i>Root Mean Square Error</i>)
Semi-NMF	Factorisation en matrices semi-non-négatives (<i>Semi-Nonnegative Matrix Factorization</i>)
SC-ISN-Cal	(Méthode d') Étalonnage par factorisation informée en matrices semi-non-négatives avec contrainte de structure (<i>Structure-Constrained Informed Semi-Nonnegative matrix factorization for sensor Calibration</i>)
SNR	Rapport signal à bruit (<i>Signal to Noise Ratio</i>)
SpAIN-Cal	(Méthode d') Étalonnage par factorisation informée en matrices non-négatives avec contraintes de parcimonie et de moyenne (<i>Sparsity and Average Constrained Informed Nonnegative Matrix Factorization for sensor Calibration</i>)
SpIN-Cal	(Méthode d') Étalonnage par factorisation informée en matrices non-négatives avec contraintes de parcimonie (<i>Sparse Informed Nonnegative Matrix Factorization for sensor Calibration</i>)
SVD	Décomposition en valeurs singulières (<i>Singular Value Decomposition</i>)

Introduction générale

1.1 Contexte

La pollution environnementale est un problème majeur rencontré de par le monde. En effet, de multiples études ont montré les effets de la pollution, par exemple dans l'eau [227] ou l'air [117], sur la santé, et plus généralement sur la biodiversité [14]. L'activité humaine est aussi à l'origine du changement climatique [96], qui affecte et affectera la vie de milliards d'humains.

En conséquence, de façon à fournir des outils d'aide à la décision, en termes d'atténuation ou de compensation de l'activité humaine sur la faune et la flore, la surveillance environnementale a fait d'énormes progrès depuis plusieurs décennies. Elle ne consiste pas seulement à surveiller la qualité de l'air, de l'eau ou des sols, mais aussi à mesurer le niveau de bruit et de pollution lumineuse, qui ont des conséquences sur la santé et la biodiversité [197, 69, 133]. Aujourd'hui, il existe principalement trois méthodes d'acquisition de données.

1. La première, consiste à dépêcher une équipe de scientifiques sur le terrain. Ceux-ci effectuent alors des mesures/relevés sur la zone étudiée, pendant le temps de la campagne. Cette façon de procéder, souvent utilisée en biologie [150], permet généralement une acquisition de données précises, réparties de façon homogène sur le secteur géographique de l'étude, mais uniquement le temps de la campagne de mesure.
2. La seconde méthode d'acquisition, plus classique, consiste à placer des capteurs fixes à différents emplacements de la zone d'intérêt. Les capteurs utilisés dans ce cas sont souvent très précis, mais également volumineux et assez onéreux. Par ailleurs, ces capteurs permettent l'acquisition de données sur de grands intervalles de temps, mais avec une couverture spatiale faible.
3. La troisième, plus récente, consiste à surveiller une zone non-pas sur le terrain, à partir de mesures ou de campagnes scientifiques, mais via l'imagerie (multi- ou hyperspectrale) aérienne ou satellitaire [105].

Aujourd'hui en France, la mesure officielle de la qualité de l'air est assurée par les Associations Agréées de Surveillance de la Qualité de l'Air (AASQA), qui disposent d'un réseau de stations fixes et mobiles réparties sur le territoire. Ce réseau rencontre les mêmes problèmes qu'évoqués précédemment, à savoir un coût d'installation et d'entretien important ainsi qu'une couverture spatiale relativement pauvre. Aujourd'hui, ces mesures éparses dans l'espace servent à alimenter des modèles de la qualité de l'air qui permettent une cartographie générale mais dont les effets locaux, tels que les effets canyon dans les rues, ne peuvent

pas être pris en compte. De plus, la surveillance de la pollution est un enjeu majeur pour la santé, et malheureusement, telle qu'elle est effectuée actuellement, elle n'implique pas assez la population qui ne prend souvent pas connaissance de la qualité de l'air qu'elle respire.

Une tendance générale en surveillance environnementale consiste à compléter les réseaux de capteurs fixes par de grandes quantités de capteurs, communicant via des réseaux sans-fil, qui sont individuellement moins précis que les capteurs classiques mais couvrent de manière beaucoup plus dense une zone d'étude [84, 178]. En particulier, le développement de capteurs miniaturisés permet maintenant d'imaginer des capteurs mobiles, portés par exemple par des robots [58], des véhicules [86] ou des volontaires [175].

L'utilisation massive d'appareils électroniques portables géolocalisés, de type *smartphones* par exemple, a permis l'émergence d'un nouveau mode d'acquisition où les porteurs de ces appareils sont acteurs de l'acquisition : la *mesure par foule de capteurs mobiles* (en anglais *mobile crowdsensing*) [72]. Le *mobile crowdsensing* consiste à équiper des citoyens volontaires de capteurs (appartenant ou connectés à leurs *smartphones*), effectuant des mesures tout au long de leurs trajets quotidiens. Grâce à leurs *smartphones*, les données sont alors datées et géolocalisées avant d'être transmises à un serveur central via une connexion sans fil (par exemple GSM ou WiFi). Les campagnes de mesure par *mobile crowdsensing* exploitent donc les déplacements quotidiens des utilisateurs pour augmenter la couverture spatiale des mesures. Les capteurs qui équipent les participants à la collecte sont généralement bas-coût et nécessitent peu d'entretien, simplifiant ainsi la gestion du réseau. De plus, l'implication directe des citoyens dans la campagne de mesure permet également de sensibiliser la population au sujet de l'étude.

Le projet d'**O**bservation et **S**ensibilisation **C**itoyenne à la qualité de l'**A**ir en **R**égion (OSCAR), financé par le programme chercheurs-citoyens de la Région Hauts-de-France depuis 2014, vise à développer un boîtier de mesure de la qualité de l'air pour le mobile crowdsensing. Constitués de différents capteurs permettant de quantifier les principaux éléments impactant la qualité de l'air que nous respirons (monoxyde de carbone, ozone, dioxyde d'azote, particules fines, humidité et température), ces boîtiers d'acquisition doivent être relativement bas-coût et simples de conception afin de pouvoir être fabriqués en masse avant d'être mis à disposition de citoyens volontaires pour effectuer une campagne de mesure par *mobile crowdsensing*.

Le projet OSCAR réunit plusieurs partenaires permettant une mise en commun des domaines d'expertise. L'étape de collecte qui requiert le développement d'infrastructures permettant de récupérer les données de façon éthique, c.-à-d. en les anonymisant de façon à respecter la vie privée des participants est assurée par l'équipe-projet Spirals du centre INRIA Lille Nord-Europe grâce à la plateforme Apisense® [80]. Le développement du boîtier d'acquisition ainsi que des méthodes spécifiques de traitement du signal nécessaires pour fusionner et interpréter l'ensemble des données reviennent au Laboratoire d'Informatique Signal et Image de la Côte d'Opale (LISIC), tandis que des données de référence précises ainsi que l'expertise en surveillance de la qualité de l'air sont apportées par l'AASQA Régionale ATMO Hauts-de-France. Enfin, dans le but de maximiser la sensibilisation du public à la surveillance de la qualité de l'air, les boîtiers de mesure ont vocation à être entièrement assemblés dans des centres de formation de la région Hauts-de-France (lycées, université). La mise en relation

du LISIC avec les centres de formation est assurée par le quatrième et dernier partenaire du projet OSCAR, l'association Bâtisseurs d'Economie Solidaire (BES).

1.2 Objectifs

Même si depuis quelques années il s'est créé un réel engouement pour le *mobile crowdsensing*, notamment pour des applications environnementales [72], leur mise en place n'est pas triviale et requiert le développement d'un certain nombre de techniques pour la récupération des données et leur exploitation. En effet, dans la mouvance de l'internet des objets, on compte plusieurs problématiques en lien avec ce type de réseaux de capteurs [84, 165] :

1. Il est possible de traquer les porteurs de boîtiers, attendant ainsi grandement à leur vie privée. De nombreuses stratégies d'anonymisation ont donc été proposées dans la littérature [33].
2. L'utilisation continue de tels capteurs de mesure est coûteuse en énergie et dans le cas de réseaux de capteurs denses, de nombreuses données sont redondantes. Certains auteurs ont donc proposé des stratégies d'optimisation de l'observatoire, soit en faisant se déplacer les capteurs aux positions les plus critiques [174], soit en sélectionnant le sous-ensemble de capteurs apportant le plus d'information [16, 27, 171, 176].
3. Les données présentent des biais [100] dûs aux dérives des réponses des capteurs, nécessitant le ré-étalonnage régulier de ces derniers. Cet étalonnage ne peut généralement pas se faire *in situ* et des approches dites *aveugles* d'étalonnage de capteurs (on parle parfois aussi d'*auto-étalonnage*) ont été proposés à cette fin.

Dans cette thèse, nous nous intéressons au développement des méthodes de traitement du signal pour l'interprétation de données collectées par *mobile crowdsensing*. Les approches proposées exploitent les informations acquises par *mobile crowdsensing*, c.-à-d. les données brutes des capteurs, la date et l'heure de chaque mesure, ainsi que leur géolocalisation.

Notre premier objectif est de mettre au point une méthode d'étalonnage semi-aveugle d'un réseau de capteurs, c.-à-d. de développer une approche visant à estimer les paramètres propres à chaque capteur du réseau, permettant l'estimation d'une valeur physique, à partir d'une donnée de sortie brute. En particulier, la méthode développée doit être adaptée à une utilisation dans un projet de *mobile-crowdsensing* dans lequel les capteurs mobiles ne sont pas contraints dans leurs déplacements. Elle se doit donc d'être relativement robuste à une dispersion aléatoire des données et à une faible proportion de mesures croisées entre les différents capteurs.

Le second objectif de cette thèse est de développer une méthode permettant la construction de cartes de pollution à partir de données issues d'un réseau de capteurs mobiles. L'approche proposée doit permettre l'intégration de mesures en provenance de nombreux capteurs de précisions différentes, ainsi que des mesures ponctuelles précises fournies par les stations officielles des AASQA.

1.3 Plan

Les travaux présentés dans ce manuscrit sont organisés en 9 chapitres, dont celui-ci, et deux parties principales. Dans le chapitre 2, nous proposons un état de l’art des méthodes d’étalonnage de capteurs de la littérature, en partant des différents modèles de capteurs utilisés pour arriver aux méthodes correspondantes.

La première partie de la thèse propose de revisiter l’étalonnage aveugle de capteurs comme un problème informé de factorisation matricielle non-négative ou semi-non-négative. Le chapitre 3 de ce manuscrit dresse un état de l’art des approches de NMF de la littérature, répertoriées en fonction des différentes contraintes qu’elles permettent de prendre en compte. Dans le chapitre 4, nous reformulons le problème d’étalonnage d’un réseau de capteurs mobiles à réponse affine sous la forme d’un problème informé de factorisation matricielle tenant compte des différentes contraintes liées au modèle de réponse des capteurs. Les méthodes d’étalonnage proposées tiennent compte de contraintes de non-négativité, rentrant ainsi dans la catégorie des méthodes informées de factorisation en matrices non-négatives (NMF). Le chapitre 5 est dédié à l’extension des premières méthodes développées à des capteurs à réponses non-linéaires. Ces dernières approches tiennent compte de structures particulière des matrices facteurs et relâchent sur l’une d’elles la contrainte de non-négativité, nous parlons alors de méthodes informées de factorisation en matrices semi-non-négatives. Des tests préliminaires mettant en évidence les performances et les limites des approches proposées sont réalisés et discutés dans le chapitre 6.

La deuxième partie de la thèse propose d’incorporer dans les méthodes d’étalonnage précédentes des outils de régularisation spatiale du phénomène physique observé partiellement (permettant aussi sa cartographie). Un bref état de l’art des méthodes d’interpolation de données est présenté dans le chapitre 7. Ces types d’approches sont alors utilisés dans ce même chapitre afin de régulariser les approches d’étalonnage développées précédemment. Dans le chapitre 8, nous proposons de tester les dernières méthodes développées sur de nouveaux jeux de donnée simulés et nous mettons alors en avant l’apport de la régularisation spatiale dans les approches d’étalonnage développées.

Enfin, le chapitre 9 dresse le bilan des travaux réalisés dans cette thèse et ouvre les perspectives envisageables.

Ce manuscrit comporte également un certain nombre d’annexes dans lesquelles nous évoquons différents travaux réalisés un peu en marge de la thèse. L’annexe A présente des méthodes robustes que nous avons proposées pour un réseau de capteurs fixes. Historiquement, ces approches ont été les premières développées pendant la thèse mais elles nécessitent des hypothèses qui ne sont pas vérifiées en pratique pour le projet OSCAR. L’annexe B détaille une méthode robuste de NMF informée, proposée pour une autre application que l’étalonnage de capteurs. Enfin, nous étudions dans l’annexe C une extension d’une approche rapide de NMF au cas où des données sont manquantes. Cette approche a été développée dans un cadre général et son extension à un cadre informé pour l’étalonnage de capteurs fait partie des perspectives de cette thèse.

État de l'art en étalonnage de capteurs

2.1 Introduction

Le terme *étalonnage*, souvent remplacé par l'anglicisme *calibration*, peut avoir plusieurs définitions suivant les domaines. En traitement d'images ou de vidéos, l'étape d'étalonnage correspond généralement à estimer les distorsions induites par le dispositif d'acquisition [208, 188, 199, 236, 90] ou l'orientation des caméras [208, 88, 233] dans des applications comme par exemple la vision par ordinateur. En réseau de microphones, l'étalonnage a principalement pour objectif de déterminer la géométrie du réseau, c.-à-d. la position des microphones, par une estimation du déphasage entre les signaux reçus [202, 161, 110]. En modélisation, l'étalonnage peut consister à apprendre les paramètres d'un modèle physique, par exemple par des approches d'assimilation de données [174].

Dans l'application considérée dans cette thèse, nous nous intéressons à l'étalonnage d'un réseau de capteurs électroniques. Cet étalonnage consiste à estimer les paramètres permettant de retrouver le champ physique mesuré à partir de la grandeur électrique en sortie des capteurs. La section suivante introduit ce problème d'étalonnage auquel nous serons confrontés par la suite, et dresse un état de l'art des méthodes existantes.

2.2 Modèles d'étalonnage de capteurs

Avec le développement relativement récent de nouveaux capteurs et micro-contrôleurs bas-coût ainsi que celui de réseaux sans fils de plus en plus étendus, les applications utilisant des données en provenance de réseaux de capteurs mobiles se sont multipliées. Une problématique inhérente à l'exploitation de données issues de telles structures est l'étalonnage des capteurs. De façon générale, lors de l'utilisation d'un capteur électronique, celui-ci convertit un phénomène physique d'entrée auquel il est sensible, par exemple un taux de luminosité, une température, une concentration de gaz ou un taux d'humidité, en un signal de sortie à un instant t , cette conversion effectuée par le capteur lie sa valeur de sortie $x(t)$ (souvent une tension, exprimée en Volts) à la valeur physique d'entrée $y(t)$ par une fonction d'étalonnage \mathcal{F} non accessible à l'utilisateur et qui est propre au capteur. Le but de l'étalonnage est alors d'*inverser* la fonction \mathcal{F} pour permettre de reconstruire à partir de la tension de sortie du

capteur, la valeur physique d'entrée correspondante. Étalonner un capteur revient donc à résoudre un *problème inverse*.

Suivant le type de capteur utilisé, le modèle de réponse, c.-à-d. la fonction d'étalonnage dont il faut chercher les paramètres¹, peut avoir différentes formes.

1. Modèle linéaire/affine : Aussi connu sous le nom de modèle gain/offset, c'est le plus simple et le plus largement étudié. Les capteurs dont la réponse suit ce type de modèle convertissent la valeur physique d'entrée en une tension de sortie par une relation affine, c.-à-d.

$$x(t) \simeq \mathcal{F}(y(t)) \simeq f_0 + f_1 \cdot y(t), \quad (2.1)$$

où f_0 et f_1 sont respectivement l'offset et le gain du capteur. Certaines approches de la littérature considèrent un modèle simplifié avec l'offset f_0 connu ou nul (fonction d'étalonnage linéaire) [25, 186, 127, 21], le gain $f_1 = 1$ (estimation du biais/offset de la fonction d'étalonnage) [111, 143] tandis que d'autres s'intéressent à l'estimation à la fois du gain f_1 et de l'offset f_0 [5, 143, 128, 85].

2. Modèle non-linéaire : Le modèle non-linéaire est nettement moins étudié dans la littérature. Outre des non-linéarités très générales traitées par réseaux de neurones [101], certains auteurs ont considéré des modèles plus simples, c.-à-d. des modèles affines par morceaux [217] (dans ce cas (2.1) est localement vérifiée) ou des modèles polynômiaux [216]. Le modèle polynômial est une extension du modèle gain/offset (2.1) à un ordre supérieur. Les capteurs dont la réponse suit un modèle polynômial ont une fonction d'étalonnage qui s'exprime comme un polynôme de degré n , c.-à-d.

$$x(t) \simeq \mathcal{F}(y(t)) \simeq \sum_{j=0}^n f_j \cdot y(t)^j. \quad (2.2)$$

3. Modèle multi-linéaire : Dans certaines applications, les capteurs utilisés peuvent avoir des réponses qui dépendent de plusieurs paramètres [160] ou qui sont fortement corrélées avec d'autres phénomènes physiques [63]. Pour la surveillance de la qualité de l'air par exemple, la réponse de certains capteurs de gaz peut dépendre à la fois de la concentration du gaz mesuré mais également d'autres paramètres tels que la température et/ou l'humidité. La fonction d'étalonnage est donc une fonction de plusieurs grandeurs physiques, qui peut s'écrire, dans le cas de deux grandeurs physiques $y_1(t)$ et $y_2(t)$ influençant la réponse du capteur :

$$x(t) \simeq \mathcal{F}(y_1(t), y_2(t)) \simeq f_1 \cdot y_1(t) + f_2 \cdot y_2(t) + f_0, \quad (2.3)$$

où f_1 et f_2 sont les gains associés à chaque valeur physique et f_0 est un offset. Les méthodes d'étalonnage spécifiques pour ce type de capteurs sont peu nombreuses dans la littérature actuelle [160, 63].

1. Dans ce manuscrit, nous utiliserons de manière abusive le terme de "paramètres d'étalonnage" pour noter les paramètres qui lient la sortie de capteur $x(t)$ à la grandeur physique $y(t)$. Par exemple dans l'équation 2.1, nous appellerons gain et offset les paramètres f_1 et f_0 alors que les vrais paramètres d'étalonnage sont respectivement $1/f_1$ et $-f_0/f_1$ (en supposant $f_1 \neq 0$).

4. Modèle gain/phase : Dans des domaines plus spécifiques, la réponse d'un capteur à un signal d'entrée peut induire un déphasage de ce signal, c.-à-d.

$$x(t) \simeq \mathcal{F}(y(t)) \simeq f_1 \cdot \exp^{i\theta} \cdot y(t), \quad (2.4)$$

où f_1 et θ sont respectivement le gain et le déphasage du capteur. L'étalonnage de ces capteurs est notamment étudié dans [11].

5. Dérive (*drift* en anglais) : Certains auteurs [87, 228, 213, 222] considèrent que les capteurs qu'ils utilisent sont pré-étalonnés, c.-à-d. que les paramètres des fonctions d'étalonnage sont connus mais qu'ils évoluent au cours du temps. La dérive correspond alors à rajouter l'indice temporel t comme paramètre de la fonction d'étalonnage \mathcal{F} . Dans ce cas, le champ mesuré à l'instant t s'écrit :

$$x(t) = \mathcal{F}(x(t), t), \quad (2.5)$$

ou – par exemple pour le cas du modèle multi-linéaire – sous la forme [160]

$$x(t) \simeq \mathcal{F}(y_1(t), y_2(t), t) \quad (2.6)$$

$$\simeq f_1 \cdot y_1(t) + f_2 \cdot y_2(t) + f_0 + f_3 \cdot t. \quad (2.7)$$

Dans la section suivante, nous proposons un bref état de l'art des méthodes existantes pour étalonner un réseau de capteurs.

2.3 Méthodes d'étalonnage

Nous proposons ici une taxonomie des approches d'étalonnage existantes. Nous distinguons les approches d'étalonnage à base de régression, les approches par apprentissage, ainsi que celles dites non-supervisées ou faiblement supervisées.

2.3.1 Les approches de type régression linéaire/non-linéaire

La méthode classique d'étalonnage de capteurs consiste à exposer les capteurs à étalonner à des valeurs d'entrée contrôlées. Connaissant la valeur physique d'entrée, il est alors possible d'estimer les paramètres d'étalonnage des capteurs pour faire coïncider la tension de sortie avec la valeur physique d'entrée, par simple régression. Cette méthode d'étalonnage est notamment utilisée dans [224, 168]. Si cette façon d'étalonner les capteurs permet une grande précision dans l'estimation des paramètres, le contrôle de la valeur physique à laquelle sont exposés les capteurs est cependant assez difficile et nécessite bien souvent un équipement spécifique tel qu'un laboratoire sous atmosphère contrôlée pour l'étalonnage de capteurs de gaz, température, humidité ou particules. Ces équipements sont souvent extrêmement coûteux, de plus, la tendance actuelle est à l'utilisation de réseaux constitués de très nombreux capteurs. La taille importante de ces réseaux ne permet donc pas de manipuler manuellement tous les

capteurs pour les étalonner en laboratoire. De plus, ces capteurs sont souvent bas-coût, et de ce fait moins stables. L'étape d'étalonnage peut donc être à répéter régulièrement, entraînant alors de nombreuses manipulations. Les méthodes proposées dans [224, 168] s'appliquent donc à des réseaux de petite taille et/ou facilement accessibles, constitués de capteurs relativement stables, sous réserve de disposer des installations nécessaires.

Dans le cas de capteurs à réponse affine, une simple régression linéaire entre la mesure physique de sortie du capteur et la valeur physique d'entrée permet une estimation efficace des paramètres d'étalonnage [224, 168]. Dans le cas de modèle de réponse non-linéaire plus complexe dont l'expression n'est pas forcément connue *a priori*, les réseaux de neurones sont généralement utilisés. Ils permettent alors un apprentissage automatique des paramètres [101].

2.3.2 Étalonnage par apprentissage/correction

Les auteurs de [5, 128, 223] ont proposé des méthodes d'étalonnage à base de projection orthogonale pour un réseau de capteurs fixes à réponses affines (modèle gain/offset) et dont les mesures sont synchronisées. Celles-ci consistent à estimer les paramètres des capteurs de sorte que le signal reconstruit appartienne à un sous-espace \mathcal{S} préalablement appris. L'algorithme proposé se décompose en deux étapes : dans un premier temps, une phase d'apprentissage permet l'estimation du sous-espace \mathcal{S} dans lequel doit résider le signal observé par les capteurs. Dans un second temps, les capteurs sont étalonnés par projection sur leur espace nul, c.-à-d. l'espace \mathcal{S}^\perp orthogonal à \mathcal{S} , via $P_{\mathcal{S}^\perp}$, la matrice de projection sur \mathcal{S}^\perp . La dimension du sous-espace \mathcal{S} des données est supposée inférieure au nombre de capteurs présents, c.-à-d. les capteurs sur-échantillonnent le phénomène physique observé [5].

- *étape d'apprentissage* : l'apprentissage de la matrice $P_{\mathcal{S}^\perp}$ qui permet l'étalonnage de capteurs du réseau dans [5, 128, 223] est effectué en supposant que dans un premier temps des capteurs étalonnés fournissent directement les valeurs du phénomène physique observé. En construisant la matrice de données Y telle que sa i -ème colonne contient les mesures de tous les capteurs à un instant t_i , alors l'apprentissage de $P_{\mathcal{S}^\perp}$ peut être réalisé par exemple par réduction de l'espace via une analyse en composantes principales (ACP) ou une analyse de Fourier des données [5].
- *étape d'étalonnage* : les capteurs du réseau vieillissant, ils nécessitent d'être régulièrement ré-étalonnés. En considérant un modèle de réponse affine, alors, le vecteur $\mathbf{x}(t)$ contenant les mesures de l'ensemble du réseau à un instant t donné, est lié au vecteur $\mathbf{y}(t)$ contenant les valeurs du champ physique aux emplacements des capteurs par la relation

$$\mathbf{y}(t) = \text{diag}(\mathbf{f}_1)\mathbf{x}(t) + \mathbf{f}_0, \quad \forall t \quad (2.8)$$

où $\text{diag}(\mathbf{f}_1)$ est la matrice diagonale dont le i -ème élément diagonal est le gain f_{1i} associé au i -ème capteur et \mathbf{f}_0 est le vecteur dont l'élément f_{0i} est l'offset du i -ème capteur. En utilisant la matrice $P_{\mathcal{S}^\perp}$ de projection sur \mathcal{S}^\perp alors

$$P_{\mathcal{S}^\perp} \cdot \mathbf{y}(t) = P_{\mathcal{S}^\perp} \cdot (\text{diag}(\mathbf{f}_1)\mathbf{x}(t) + \mathbf{f}_0) = 0, \quad \forall t. \quad (2.9)$$

En considérant plusieurs échantillons, c.-à-d. plusieurs vecteurs de mesures pris à différents instants t_i et en centrant les mesures, Balzano et Nowak [5] parviennent à remplacer les offsets dans (2.9) par une expression ne dépendant plus que des gains. Les gains des capteurs doivent donc vérifier la projection nulle des données sur \mathcal{S}^\perp , et peuvent être estimés par moindres carrés [5], décomposition en valeurs singulières (SVD pour *Singular Value Decomposition*) [5], ou par moindres carrés totaux [128]. Une fois les gains estimés pour chaque capteur, en supposant connaître la valeur moyenne \bar{x} du phénomène physique au cours du temps, l'offset peut être estimé en suivant les développements ci-dessous :

$$\frac{1}{T} \sum_{i=1}^T \mathbf{f}_1 \mathbf{x}(t_i) + \mathbf{f}_0 \approx \bar{x}, \quad (2.10)$$

$$\Rightarrow \mathbf{f}_0 \approx \bar{x} - \frac{1}{T} \sum_{i=1}^T \mathbf{f}_1 \mathbf{x}(t_i), \quad (2.11)$$

où T est le nombre de échantillons disponibles pour l'étalonnage.

Cette méthode d'étalonnage par projection montre de bonnes performances dans l'application considérée par ses auteurs, cependant, elle nécessite une configuration particulière du réseau de capteurs, ainsi qu'un accès à une base de mesures étalonnées pour permettre l'apprentissage du sous-espace du signal et son orthogonal. Dans [5, 128], les auteurs proposent également une version informée de cette méthode en supposant que pendant la phase d'étalonnage certains capteurs du réseau sont étalonnés.

Plusieurs extensions de ces travaux ont été proposées, dans [223], les auteurs étendent l'approche de Balzano *et al.* pour l'estimation de la dérive des capteurs, en incluant un filtrage de Kalman et proposent également dans [221] de l'apprentissage profond (ou *deep learning*) pour l'apprentissage du sous-espace.

Dans [55], nous avons étendu [5] et [128] en proposant une version robuste aux valeurs aberrantes. Notre approche consiste en un pré-traitement à base d'ACP robuste [126] pour filtrer les données. Ces travaux ont été réalisés en début de thèse. Cependant, nous nous sommes aperçus *a posteriori* que nous ne pourrions pas bénéficier de capteurs pré-étalonnés en pratique dans le projet OSCAR, ce qui limitait l'intérêt de telles approches. Aussi, comme ces travaux ne constituent pas le cœur de la thèse, ils sont présentés en annexe A.

2.3.3 Les approches non-supervisées ou faiblement supervisées

Les méthodes d'étalonnage non-supervisées ne nécessitent pas de connaître à l'avance une grande base de données d'apprentissage. Ces méthodes travaillent directement sur les données en provenance des capteurs et sont donc particulièrement adaptées aux réseaux mobiles pour lesquels il peut être par exemple difficile d'avoir un apprentissage du signal observé, principalement à cause de la mobilité des capteurs qui n'est souvent pas contrôlée.

Ces approches non-supervisées peuvent être ordonnées en deux sous-groupes, d'un côté des méthodes dites de macro-étalonnage, de l'autre les méthodes de micro-étalonnage, dont les définitions sont données ci dessous.

Définition 2.1

Les méthodes de macro-étalonnage consistent à harmoniser en même temps les données en provenance de tous les capteurs, pour obtenir un ensemble de mesures cohérentes.

Définition 2.2

Les méthodes de micro-étalonnage travaillent capteur par capteur et font souvent intervenir des données de référence pour étalonner le réseau.

2.3.3.1 Macro-étalonnage

Étalonnage pour l'acquisition compressée

L'acquisition compressée $\mathbf{x}(t)$ (ou *compressed-sensing*) d'un signal $\mathbf{y}(t)$ consiste à rechercher un vecteur $\mathbf{z}(t)$ k -parcimonieux, c.-à-d. contenant au plus k composantes non-nulles, telle que $\mathbf{y}(t)$ peut s'écrire comme

$$\mathbf{y}(t) = D \cdot \mathbf{z}(t), \tag{2.12}$$

où D est un dictionnaire connu. L'acquisition est dite compressée car on observe non pas $\mathbf{y}(t)$ mais $\mathbf{x}(t)$ qui est de dimensions inférieures et qui est lié à $\mathbf{y}(t)$ par :

$$\mathbf{x}(t) = \Psi \cdot \mathbf{y}(t), \tag{2.13}$$

où Ψ est un opérateur de compression, supposé connu et incohérent par rapport à D [23]. La reconstruction de $\mathbf{y}(t)$ à partir de $\mathbf{x}(t)$ s'effectue donc en estimant d'abord $\mathbf{z}(t)$ par résolution du problème inverse parcimonieux

$$\begin{aligned} \tilde{\mathbf{z}}(t) = \arg \min_{\mathbf{z}(t)} \|\mathbf{x}(t) - \Psi \cdot D \cdot \mathbf{z}(t)\|_{\mathcal{F}}, \\ \text{s.c. } \mathbf{z}(t) \text{ est } k\text{-parcimonieux.} \end{aligned} \tag{2.14}$$

Le problème d'acquisition compressée a été très étudié dans la littérature [61], mais les problèmes d'étalonnage pour cette configuration n'ont été abordés que beaucoup plus récemment [11, 186, 21]. Dans ces travaux, le problème (2.14) devient

$$\begin{aligned} \tilde{\mathbf{z}}(t) = \arg \min_{\mathbf{z}(t)} \|\mathbf{x}(t) - \text{diag}(\mathbf{f}_1) \cdot \Psi \cdot D \cdot \mathbf{z}(t)\|_{\mathcal{F}}, \\ \text{s.c. } \mathbf{z}(t) \text{ est } k\text{-parcimonieux.} \end{aligned} \tag{2.15}$$

Les auteurs de [186] ont étendu l'approche de *generalized approximate message passing* [170] pour résoudre (2.15). Les auteurs de [11] ont proposé un cadre convexe à la résolution² de (2.15), en supposant connue la somme des valeurs de \mathbf{f}_1 , et utilisent l'algorithme des directions alternées (ADM) [19]. Enfin, les auteurs de [21] considèrent un cadre non-convexe faisant appel

2. Ces auteurs ont aussi considéré un modèle gain/phase d'étalonnage dans [11].

à une descente par gradient, pour une application en imagerie.

Méthodes basées sur les moments

Dans [218], les auteurs considèrent un réseau de n capteurs mobiles effectuant des mesures sur une zone géographique fixe. Ils supposent également que les capteurs suivent un modèle gain/offset et que le gain moyen \bar{f}_1 et l'offset moyen \bar{f}_0 des capteurs sont connus. En intégrant les mesures prises par un capteur sur un intervalle de temps T suffisamment grand et en supposant que le phénomène physique observé est stationnaire pendant toute la durée des acquisitions, alors la moyenne $\mu_i(T)$ des mesures prises par le capteur i pendant l'intervalle de durée T est liée à la moyenne $\bar{\mu}$ du champ physique observé par la relation

$$\mu_i(T) = f_{1i} \cdot \bar{\mu} + f_{0i} + \eta_i(T), \quad (2.16)$$

où f_{1i} , f_{0i} sont respectivement le gain et l'offset du capteur i , et $\eta_i(T)$ est l'écart à la moyenne qui tend vers zéro lorsque $T \rightarrow \infty$, c.-à-d. lorsque le nombre d'échantillons augmente. De la même façon, les auteurs de [218] montrent que la variance σ_i^2 des mesures du capteur i est liée à la variance $\bar{\sigma}^2$ du phénomène physique observé par la relation

$$\sigma_i^2 = f_{1i}^2 \cdot \bar{\sigma}^2 + \theta_i(T), \quad (2.17)$$

où $\theta_i(T)$ est l'erreur d'estimation qui tend vers zéro quand $T \rightarrow \infty$. En considérant alors les équations (2.16) et (2.17) pour tous les capteurs et en définissant ϵ_{f_1} et ϵ_{f_0} comme

$$\epsilon_{f_1} \triangleq \frac{1}{n} \cdot \sum_{i=1}^n f_{1i} - \bar{f}_1 \quad \text{et} \quad \epsilon_{f_0} \triangleq \frac{1}{n} \cdot \sum_{i=1}^n f_{0i} - \bar{f}_0, \quad (2.18)$$

les auteurs de [218] proposent d'estimer le gain et l'offset de chaque capteur en résolvant le problème de minimisation suivant

$$\min_{\mathbf{f}_1, \mathbf{f}_0, \bar{\mu}, \bar{\sigma}^2, \epsilon_{f_1}, \epsilon_{f_0}} \|\boldsymbol{\eta}(T)\|_2^2 + \|\boldsymbol{\theta}(T)\|_2^2 + \epsilon_{f_1}^2 + \epsilon_{f_0}^2, \quad (2.19)$$

où \mathbf{f}_1 et \mathbf{f}_0 sont respectivement le vecteur des gains et des offsets des capteurs et où $\boldsymbol{\eta}(T)$ et $\boldsymbol{\theta}(T)$ sont respectivement le vecteur composé des erreurs $\eta_i(T)$ et celui composé des erreurs $\theta_i(T)$. Les auteurs de [218] ont également proposé deux extensions de cette approche pour des capteurs à réponses non-linéaires [217, 216]. Ces méthodes s'affranchissent de toute connaissance même partielle du phénomène physique observé mise à part l'hypothèse de stationnarité du phénomène sur l'intervalle de temps T . Cependant, elles nécessitent une longue acquisition de données avant d'être efficaces.

Une approche similaire basée sur les moments a également été proposée dans [70] pour du *destripping* d'images satellitaires. En imagerie satellitaire, de nombreux instruments fonctionnent selon le système *push broom* présenté dans la figure 2.1. Le satellite se déplaçant dans une direction orthogonale à l'orientation du réseau de capteurs, les données observées forment une image. Dans le cas où aucun capteur n'est étalonné, l'image observée présente des bandes spatiales qui sont relativement gênantes. Les auteurs de [70] proposent une approche

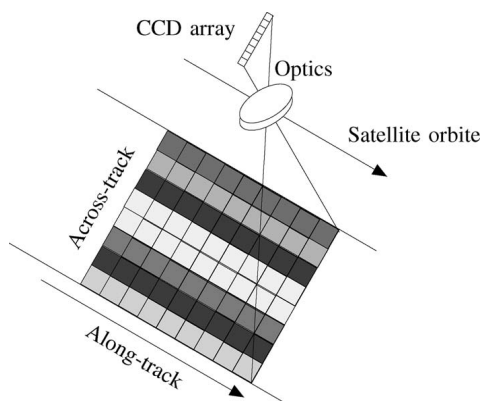


FIGURE 2.1 – Principe d’acquisition d’images par système *push broom*. Source : [25].

d’étalonnage par comparaison des histogrammes obtenus sur chaque bande. Plus récemment, une extension bayésienne de [70] a été proposée par Carfantan et Idier [25].

Étalonnage par modèle de graphe

Dans [111], les auteurs considèrent un réseau de capteurs mobiles suivant un modèle de réponse offset, c.-à-d. que la mesure $x(t)$ d’un capteur est liée à la valeur physique observée $y(t)$ par la relation

$$x(t) = y(t) + f_0, \quad (2.20)$$

où f_0 est l’offset du capteur. Les auteurs proposent d’utiliser le formalisme de *rendez-vous* qu’on trouve plus classiquement dans les approches de micro-étalonnage et dont la définition est la suivante :

Définition 2.3 ([183])

Un rendez-vous est un voisinage spatio-temporel entre deux capteurs. Lorsqu’ils sont en rendez-vous, deux capteurs sont supposés observer la même valeur physique d’entrée.

En supposant qu’au cours du temps, le déplacement des capteurs les amène à se rencontrer et donc à effectuer des mesures du même phénomène physique, les auteurs de [111] proposent de réécrire les rencontres entre les capteurs sous forme d’un graphe dans lequel chaque capteur est un nœud. Si deux capteurs ont effectué une mesure en rendez-vous, alors les nœuds correspondants sont reliés par une arête. Le graphe construit à partir des rendez-vous met en évidence les relations entre les différents capteurs. Lee *et al.* [111] déduisent alors de ces relations un système d’équations mettant en jeu la matrice laplacienne du graphe et dont la résolution mène à l’étalonnage du réseau de capteur. Cependant, comme pour les approches basées sur les moments, il faut intégrer les mesures des capteurs sur un intervalle de temps conséquent afin d’assurer un nombre important de croisements entre les différents capteurs. En effet, dans [111], les auteurs montrent que les performances sont fortement corrélées à la densité de mesures croisées entre les capteurs.

Plus récemment, les mêmes auteurs ont proposé d’inclure au réseau de capteurs mobiles,

des données issues de référence [194], réduisant ainsi la sensibilité de la méthode à la densité des mesures.

2.3.3.2 Micro-étalonnage

La majorité des méthodes de micro-étalonnage pour les réseaux de capteurs mobiles font intervenir à la fois des données de capteurs à étalonner mais également des données de capteurs de référence et exploitent la définition de rendez-vous 2.3.

Étalonnage par références directes

Dans [143], les auteurs considèrent un réseau de capteurs mobiles effectuant des mesures sur une zone où évoluent également des capteurs de référence étalonnés. Lorsqu'un capteur à étalonner effectue une mesure dans un voisinage spatio-temporel proche d'un capteur de référence, alors ils sont en rendez-vous, et sont supposés observer le même phénomène physique. Pour des capteurs suivant un modèle d'offset, lorsqu'un capteur effectue une mesure en rendez-vous avec une référence, la relation suivante doit être vérifiée :

$$x(t)_{\text{REF}} = x(t) = y(t) + f_0, \quad (2.21)$$

où $x(t)_{\text{REF}}$ est la valeur du phénomène physique fournie par la référence, $x(t)$ est la valeur du phénomène physique reconstruite à partir de la donnée de sortie $y(t)$ du capteur, et f_0 est son offset. Si f_0 n'est pas parfaitement déterminé, alors la valeur du phénomène physique reconstruite n'est pas rigoureusement égale à la valeur du capteur de référence, et l'équation précédente est biaisée. Les auteurs de [143] proposent une méthode d'étalonnage permettant de corriger l'offset d'un capteur à partir du biais de mesure, constaté lors d'un rendez-vous avec une référence. L'utilisation de mesures de références pour l'étalonnage de capteurs mobiles est une idée d'autant plus intéressante qu'en pratique, les réseaux de capteurs mobiles bas-coût qui doivent être régulièrement ré-étalonnés sont souvent déployés en renforcement d'infrastructures existantes, constituant des points de référence. Cependant, l'approche proposée dans [143] suppose que chacun des capteurs à étalonner rencontrent fréquemment des références autour desquelles ils effectuent des mesures, ce qui est une hypothèse difficilement vérifiable dans certaines applications, par exemple en *mobile crowdsensing*.

Approche multi-sauts

Dans l'application visée dans [85], des capteurs environnementaux ont été placés sur un réseau de tramways. Tout au long de la journée, les mesures brutes, c.-à-d. non étalonnées, issues des différents capteurs sont datées, géolocalisées et collectées. À ces mesures de capteurs mobiles viennent s'ajouter des mesures de référence issues de capteurs fixes précis et pré-étalonnés répartis à différents endroits le long des lignes de tramway. La méthode d'étalonnage proposée exploite alors les informations fournies par la mobilité des capteurs ainsi que les données de référence des stations fixes. En particulier, l'approche décrite exploite les rendez-vous entre les capteurs mobiles à étalonner et les capteurs de référence.

Grâce à la configuration particulière du réseau de capteurs mobiles utilisé ici, c.-à-d. grâce au passage régulier des tramways à proximité des capteurs de référence, l'accumulation de mesures en rendez-vous entre tous les capteurs (mobile/mobile et mobile/référence) est assurée. Ces rendez-vous ont alors permis aux auteurs de [85] de proposer une approche d'étalonnage dite multi-sauts étalonnant un à un l'ensemble des capteurs du réseau. Cette approche en trois étapes répétées jusqu'à étalonnage de l'ensemble du réseau consiste à :

1. Sélectionner un capteur à étalonner ayant effectué des mesures en rendez-vous avec une référence.
2. Étalonner ce capteur par régression linéaire entre ses mesures et les mesures de référence.
3. Considérer ce capteur étalonné comme étant une nouvelle référence.
4. Répéter l'opération pour un autre capteur.

Une extension de cette méthode a été proposée [182], remplaçant la régression par moindres carrés classique par une régression moyenne géométrique (GMR pour *geometric mean regression*), permettant de pallier au problème de propagation de l'erreur d'étalonnage d'un capteur à un autre. Plus récemment, la même équipe a proposé une extension de [85, 182] à des capteurs non plus à réponse gain/offset mais à réponses multi-linéaires, en tenant compte de la possible sensibilité des capteurs à différents phénomènes physiques [134].

Les très bons résultats obtenus en utilisant cet étalonnage multi-sauts en situation réelle, sont en grande partie dûs à la configuration du réseau, permettant d'assurer une grande quantité de mesures croisées entre les capteurs. La notion de rendez-vous ainsi que l'intégration de mesures de références sont les points clés développés par [85, 182] et sont particulièrement bien adaptés à une utilisation dans un réseau de capteurs mobiles. Récemment, une version distribuée de [182] a été proposée dans [109].

Dans une application de *mobile crowdsensing* classique, où les porteurs des capteurs sont humains, ces derniers se déplacent de façon non contrainte et un nombre suffisant de rendez-vous entre tous les capteurs n'est pas assuré pour permettre un étalonnage précis du réseau. Il est alors nécessaire de développer de nouvelles approches

2.4 Conclusion

Dans ce chapitre, nous avons passé en revue les méthodes d'étalonnage les plus couramment utilisées, permettant de gérer des réseaux de capteurs fixes [70, 5, 25, 128, 11], ou mobiles [143, 218, 85, 111, 182, 194, 134, 109].

Ces approches sont souvent très spécifiques et requièrent soit des connaissances précises du phénomène physique observé, soit une densité considérable de données pour permettre un bon déroulement de l'étalonnage du réseau.

En *mobile crowdsensing*, les réseaux de capteurs sont souvent déployés sur de grandes surfaces et sont mobiles. La répartition des données sur le champ physique observé est donc rela-

tivement aléatoire, et il est donc difficile de pouvoir recueillir en amont de la collecte, des informations sur le signal qui sera observé par le réseau. Les approches d'étalonnage supervisées ne sont donc clairement pas adaptées à ce type de réseaux. De plus, en *mobile crowdsensing* classique, les capteurs sont bien souvent libres dans leurs mouvements. L'acquisition de données croisées entre les capteurs mobiles et les références n'est donc pas assurée. Les approches non-supervisées nécessitant une connexion de chaque capteur à une référence ne sont donc pas non plus envisageables pour l'étalonnage des réseaux de capteurs en *mobile crowdsensing*.

Dans la partie suivante, nous proposons de revisiter le problème d'étalonnage d'un réseau de capteurs mobiles sous la forme d'un problème informé de factorisation matricielle, où la matrice observée qui collecte les données n'est que partiellement connue, c.-à-d. que certaines de ses données sont manquantes. En particulier, les approches proposées permettent de gérer l'hétérogénéité du réseau de capteurs en tenant compte de l'ensemble des données des capteurs mobiles peu précis et des capteurs de référence très précis.

Première partie

Le problème de l'étalonnage de capteurs vu comme un problème de factorisation matricielle

État de l'art en factorisation matricielle

3.1 Introduction

Dans les prochains chapitres, nous montrerons comment ré-ordonner les données acquises par un réseau de capteurs mobiles sous la forme d'une matrice réécrivable comme un produit matriciel dont l'estimation des facteurs permet l'étalonnage des capteurs. Il est donc important d'avoir connaissance de la grande richesse de la littérature sur les méthodes de factorisation matricielle dont nous proposons ici un état de l'art.

Dans beaucoup d'applications, une matrice de données X peut s'écrire sous la forme d'un produit matriciel

$$X \approx G \cdot F. \quad (3.1)$$

De très nombreuses approches permettent de décomposer cette matrice en un produit matriciel. L'analyse en composantes principales (ACP) [158] par exemple qui est généralement utilisée pour réduire la dimension des données, décompose les colonnes de X en combinaisons de vecteurs orthogonaux. L'analyse en composantes indépendantes (ACI) [24], étendant l'idée de l'ACP, sépare les colonnes de X en combinaisons de vecteurs statistiquement indépendants. L'apprentissage de dictionnaire, consiste à construire une matrice D généralement orthonormée dont la combinaison de quelques colonnes permet de reconstruire chaque colonne de X . La matrice D (le dictionnaire), apprise sur une matrice X , est ensuite utilisée sur d'autres données pour, par exemple les débruiter, les restaurer [136] ou les classifier [29]. Le filtrage collaboratif vise également à factoriser une matrice de données souvent à trous de façon à en extraire des profils types [201]. Cette technique est en grande partie utilisée dans les systèmes de recommandation et permet par exemple d'apprendre le profil de nombreux utilisateurs de façon à pouvoir leur suggérer des produits. Enfin, nous citons ici la factorisation en matrices non-négatives (NMF pour *Nonnegative Matrix Factorization*) comme dernier exemple de méthode de factorisation matricielle. En NMF, une matrice de données non-négatives X est factorisée sous la forme du produit de deux matrices G et F non-négatives fournissant la meilleure approximation de X . Par rapport aux approches sans contrainte de signe, les techniques de NMF permettent généralement une interprétabilité accrue des facteurs matriciels [113].

Les méthodes d'étalonnage que nous proposons de développer dans la suite de ce manuscrit

sont principalement basées sur des approches de type NMF. Ainsi, nous proposons dans la suite de ce chapitre un bref état de l'art des méthodes de NMF rencontrées dans la littérature.

3.2 Critères et contraintes usuelles en NMF

En considérant une matrice de données positive $X \in \mathbb{R}_+^{m \times n}$, un problème de factorisation matricielle peut généralement s'écrire sous la forme d'un problème d'optimisation visant à minimiser la fonction de coût

$$\mathcal{J}(G, F) = \mathcal{D}(X || G \cdot F) + \mathcal{C}(G, F), \quad (3.2)$$

où $\mathcal{D}(\cdot)$ est un terme d'attache aux données et $\mathcal{C}(\cdot)$ est une fonction de pénalisation associée aux contraintes et régularisations éventuelles.

Les méthodes proposées dans ce manuscrit sont principalement basées sur des approches de factorisation en matrices non-négatives (NMF) ou semi-non-négatives (Semi-NMF). En NMF, la matrice de données X ainsi que les matrices facteurs G et F sont supposées non-négatives, tandis qu'en Semi-NMF, seule l'une des matrices facteurs est supposée non-négative. Les approches les plus classiques de NMF et Semi-NMF consistent à minimiser la fonction de coût (3.2) en tenant compte des contraintes de positivité du/des facteur(s). Ces contraintes de signe des matrices ne peuvent pas être directement incluses dans le problème (3.2) par pénalisation sont donc rajoutées en marge du problème d'optimisation (généralement via le choix de la méthode d'optimisation) qui s'écrit alors

$$\begin{aligned} G, F &= \mathcal{D}(X || \tilde{G} \cdot \tilde{F}) + \mathcal{C}(\tilde{G}, \tilde{F}), \\ \text{s.c. } \tilde{G}, \tilde{F} &\geq 0. \end{aligned} \quad (3.3)$$

Dans cette section, nous présenterons les principaux critères utilisés comme terme d'attache aux données en NMF, ainsi que les contraintes et pénalisations les plus fréquentes.

3.2.1 Critères usuels d'attache aux données

Nous considérons ici le problème de NMF général le plus standard, c.-à-d. sans contraintes autre que la non-négativité des matrices, qui consiste à minimiser :

$$\begin{aligned} \mathcal{J}(G, F) &= \mathcal{D}(X || G \cdot F), \\ \text{s.c. } \tilde{G}, \tilde{F} &\geq 0. \end{aligned} \quad (3.4)$$

Pour ce problème, les termes d'attache aux données usuels sont la norme de Frobenius, la divergence de Kullback-Leibler, la divergence d'Itakura-Saito, les divergences paramétriques ou encore un terme d'attache aux données pondéré. D'autres critères, moins classiques, permettent de tenir compte d'informations complémentaires ou de structures spécifiques de fac-

teurs.

3.2.1.1 La norme de Frobenius

Le problème de factorisation le plus classique consiste à minimiser la fonction de coût (3.4), en choisissant comme terme d'attache aux données la norme de Frobenius de $(X - G \cdot F)$, c.-à-d.

$$\begin{aligned} \mathcal{D}(X||G \cdot F) &= \frac{1}{2} \cdot \|X - G \cdot F\|_{\mathcal{F}}^2 = \frac{1}{2} \cdot \sum_{i,j} (x_{i,j} - (G \cdot F)_{i,j})^2, \\ &= \frac{1}{2} \cdot \text{Tr} \left((X - G \cdot F)(X - G \cdot F)^T \right), \end{aligned} \quad (3.5)$$

où $(A)_{i,j}$ est l'élément de la matrice A situé à la i -ème ligne et j -ème colonne. Cette norme, obtenue par généralisation du produit scalaire hermitien, est la plus utilisée. En effet, la norme de Frobenius est le critère d'attache aux données le plus adapté à la factorisation d'une matrice corrompue par un bruit gaussien.

Suivant la nature de la matrice de données à factoriser, il peut être intéressant d'utiliser un terme d'attache aux données autre que la norme de Frobenius, permettant de prendre en compte certaines caractéristiques de la matrice, par exemple un bruit particulier ou la présence de valeurs aberrantes. Parmi les termes d'attache aux données les plus courants, la divergence de Kullback-Leibler [113, 112], d'Itakura-Saito [97, 67, 116] ainsi que les divergences α [2] ou β [40]. Plus généralement, ces divergences peuvent être exprimées sous la forme d'une unique formule paramétrique, la divergence $\alpha\beta$ [35].

3.2.1.2 Divergence de Kullback-Leibler

La divergence de Kullback-Leibler (divergence KL) [179] mesure généralement la dissimilarité entre deux vecteurs \mathbf{u} , \mathbf{v} et est définie comme

$$d^{\text{KL}}(\mathbf{u}||\mathbf{v}) = \sum_i \left(u_i \cdot \ln \left(\frac{u_i}{v_i} \right) - u_i + v_i \right), \quad (3.6)$$

où u_i est le i -ème élément de \mathbf{u} et v_i celui de \mathbf{v} . Cette formule se généralise naturellement sur les matrices et l'attache aux données

$$\mathcal{D}^{\text{KL}}(X||G \cdot F) = \sum_{ij} \left(x_{ij} \cdot \ln \left(\frac{x_{ij}}{(G \cdot F)_{ij}} \right) - x_{ij} + (G \cdot F)_{ij} \right). \quad (3.7)$$

Cette mesure a notamment été utilisée dans l'une des approches de référence en NMF, celle de Lee et Seung [112].

3.2.1.3 Divergence d'Itakura-Saito

La divergence d'Itakura-Saito (divergence IS) [97] qui s'exprime entre deux scalaires u et v positifs comme

$$d^{\text{IS}}(u||v) = \frac{u}{v} - \ln \frac{u}{v} - 1, \quad (3.8)$$

se généralise aux matrices U et V comme

$$\mathcal{D}^{\text{IS}}(U||V) = \sum_{ij} d^{\text{IS}}(u_{ij}||v_{ij}). \quad (3.9)$$

Cette divergence est principalement utilisée pour des applications en traitement audiophonique, du fait qu'elle est invariante à l'échelle, c.-à-d. que pour tout scalaire λ positif

$$d^{\text{IS}}(\lambda \cdot u||\lambda \cdot v) = d^{\text{IS}}(u||v). \quad (3.10)$$

Cette propriété permet d'accorder autant d'importance aux termes de grande comme de faible amplitude, ce qui permet de traiter de façon équivalente toutes les valeurs d'une matrice U , même lorsque celles-ci sont très dispersées, comme cela peut être le cas en audio lorsque U correspond par exemple à un spectrogramme [67, 116]. L'attache aux données utilisant la divergence IS s'écrit

$$\mathcal{D}^{\text{IS}}(X||G \cdot F) = \sum_{ij} \left(\frac{x_{ij}}{(G \cdot F)_{ij}} - \ln \frac{x_{ij}}{(G \cdot F)_{ij}} - 1 \right). \quad (3.11)$$

3.2.1.4 Les divergences paramétriques

Divergence α

La divergence α , proposée par Chernoff [30] puis reprise en NMF par Amari [2] est un premier cas de mesure de similarité entre deux signaux, permettant d'ajuster suivant la valeur de α le comportement de l'attache aux données vis à vis d'éventuelles valeurs aberrantes. Si le choix d'un paramètre α faible permet en général de réduire l'influence des valeurs extrêmes/aberrantes, le choix d'un paramètre fort a l'effet inverse et produit une approximation majorante des signaux de données. La divergence α s'exprime entre deux matrices U et V comme

$$\mathcal{D}^{\alpha}(U||V) = \begin{cases} \frac{-1}{\alpha(\alpha-1)} \cdot \sum_{i,j} (u_{ij}^{\alpha} v_{ij}^{\alpha-1} - \alpha u_{ij} + (\alpha-1)v_{ij}), & \alpha \neq 0 \text{ et } \alpha \neq 1, \\ \sum_{i,j} (u_{ij} \ln \frac{u_{ij}}{v_{ij}} - u_{ij} + v_{ij}), & \alpha = 1, \\ \sum_{i,j} (v_{ij} \ln \frac{v_{ij}}{u_{ij}} - v_{ij} + u_{ij}), & \alpha = 0. \end{cases} \quad (3.12)$$

Divergence β

Initialement étudiée par Basu en 1998 [6], la divergence β permet tout comme la divergence α de changer le comportement du terme d'attache aux données par l'ajustement d'un

α	β	$\alpha + \beta$	$\mathcal{D}^{\alpha\beta}(X G \cdot F)$	équivalence
1	1		eq. (3.5)	norme de Frobenius
1	0		eq. (3.7)	divergence de Kullback-Leibler (KL)
1	-1		eq. (3.11)	divergence de Itakura-Saito (IS)
		1	eq. (3.12)	divergence α
1			eq. (3.13)	divergence β

TABLE 3.1 – Tableau des divergences.

paramètre β . Elle s'exprime entre deux matrices U et V comme

$$\mathcal{D}^\beta(U||V) = \begin{cases} \frac{-1}{\beta} \cdot \sum_{i,j} (u_{ij}v_{ij}^\beta - \frac{1}{1+\beta}u_{ij}^{\beta+1} - \frac{\beta}{1+\beta}v_{ij}^{\beta+1}), & \beta \neq 0 \text{ et } \beta \neq -1, \\ \sum_{i,j} (u_{ij} \ln \frac{u_{ij}}{v_{ij}} - u_{ij} + v_{ij}), & \beta = 1, \\ \sum_{i,j} (v_{ij} \ln \frac{v_{ij}}{u_{ij}} + \frac{v_{ij}}{u_{ij}} - 1), & \beta = 0. \end{cases} \quad (3.13)$$

Divergence $\alpha\beta$

Les divergences paramétriques [35] utilisées en NMF permettent un réglage fin du comportement de l'attache aux données. En particulier, les divergences $\alpha\beta$ [35] dépendent de deux paramètres et s'expriment entre deux scalaires u et v comme

$$d^{\alpha\beta}(u||v) = \begin{cases} \frac{-1}{\alpha\beta} \cdot (u^\alpha v^\beta - \frac{\alpha}{\alpha+\beta}u^{\alpha+\beta} - \frac{\beta}{\alpha+\beta}v^{\alpha+\beta}), & (\alpha, \beta, \alpha + \beta) \neq 0, \\ \frac{1}{\alpha^2} (u^\alpha \ln \frac{u^\alpha}{v^\alpha} - u^\alpha + v^\alpha), & \alpha \neq 0, \beta = 0, \\ \frac{1}{\alpha^2} (\ln \frac{v^\alpha}{u^\alpha} + \frac{u^\alpha}{v^\alpha} - 1), & \alpha = -\beta \neq 0, \\ \frac{1}{\beta^2} (v^\beta \ln \frac{v^\beta}{u^\beta} - v^\beta + u^\beta), & \alpha = 0, \beta \neq 0, \\ \frac{1}{2} (\ln u - \ln v), & \alpha, \beta = 0. \end{cases} \quad (3.14)$$

Cette divergence peut être généralisée aux matrices et s'exprime alors comme

$$\mathcal{D}^{\alpha\beta}(U||V) = \sum_{ij} d^{\alpha\beta}(u_{ij}||v_{ij}), \quad (3.15)$$

pour toutes matrices U et V positives. En ajustant les valeurs de α et β , cette formulation proposée par Cichocki *et al.* en 2011 [35] permet de rassembler sous une unique expression les divergences évoquées précédemment mais également la norme de Frobenius. Le tableau 3.1 montre le terme d'attache aux données correspondant au choix des paramètres α et β .

Les divergences présentées dans le tableau 3.1 ont toutes fait l'objet de nombreux travaux [30, 6, 112, 144, 38, 36, 2, 34, 89, 68, 123], avant d'être généralisées par Cichocki *et al.* [35]. Les méthodes développées dans cette thèse reposent essentiellement sur de la NMF avec des fonctions de coût à base de norme de Frobenius. L'utilisation des divergences paramétriques ne sera donc pas étudiée plus en détails dans ce manuscrit. Par ailleurs, un état de l'art plus complet sur les divergences paramétriques est présenté dans la thèse de Limem [123].

Il est toutefois important de retenir que les divergences paramétriques permettent d'unifier sous la forme d'une même expression (3.14), de nombreuses variantes du terme d'attache aux données, et que le choix adéquat des paramètres α et β permet une certaine robustesse aux valeurs aberrantes et/ou de pénaliser différemment les erreurs en fonction de leur signe (sous- et sur-approximations de X par le produit $G \cdot F$) [35].

3.2.1.5 Critère pondéré

Si une mesure de confiance sur la qualité des données présentes dans X est disponible, elle peut être insérée dans le problème d'optimisation et permet alors de limiter par exemple les effets du bruit. Pour tenir compte de cette fluctuation de confiance aux données, il est possible de pondérer ces dernières en affectant un poids à chaque entrée de la matrice X . Les données les plus sûres sont alors favorisées en se voyant affectées de poids forts, au dépend des autres données auxquelles sont associés des poids plus faibles.

Les premiers travaux en NMF pondérée ont été menés en 2001 par Guillaumet *et al.* [78]. Les auteurs considèrent alors un poids constant par ligne de la matrice X en définissant une matrice diagonale $D \in \mathbb{R}_+^{m \times m}$ dont l'élément d_{ii} correspond à la pondération associée à la i -ème ligne de X . Cette pondération peut être incluse directement dans le terme d'attache aux données qui s'écrit alors

$$\mathcal{D}_D(X||G \cdot F) \triangleq \mathcal{D}(D \cdot X||D \cdot G \cdot F). \quad (3.16)$$

Depuis ces travaux, les mêmes auteurs ont proposé des extensions permettant une pondération de chaque élément de X indépendamment [79]. En définissant une matrice de pondération $W \in \mathbb{R}_+^{m \times n}$ dont un élément w_{ij} correspond au poids associé à x_{ij} , le terme d'attache aux données s'écrit alors

$$\mathcal{D}_W(X||G \cdot F) \triangleq \mathcal{D}(W \circ X||W \circ (G \cdot F)), \quad (3.17)$$

où \circ désigne le produit de Hadamard. L'utilisation d'une pondération terme à terme permet également de gérer le cas d'une matrice X à données manquantes, en associant aux entrées correspondantes un poids nul de façon à ce qu'elles n'influencent pas la factorisation. Si la matrice W est binaire, alors elle fait office d'opérateur d'échantillonnage de X . La prise en compte d'une pondération dans le terme d'attache aux données a été étudié de nombreuses fois avec la norme de Frobenius [78, 79, 234, 91, 56] ou des divergences paramétriques [31].

3.2.1.6 Prise en compte de structures dans le critère

Dans certains problèmes spécifiques, les facteurs matriciels peuvent avoir des structures spécifiques dont il peut être intéressant de tenir compte dans le critère d'attache aux données $\mathcal{D}(\cdot)$.

Contraintes strictes d'égalité

Si quelques éléments des matrices sont connus de façon précise, alors ils peuvent être forcés de façon stricte. Dans [123, 119], une paramétrisation des matrices en la somme d'une partie forcée connue et d'une partie libre à estimer permet la gestion de ces contraintes d'égalité strictes. L'approche développée consiste dans un premier temps à définir deux matrices $\Phi_F \in \mathbb{R}_+^{r \times n}$ et $\Delta_F \in \mathbb{R}_+^{r \times n}$ contenant respectivement les valeurs connues de F et les valeurs inconnues à déterminer. Ainsi, la matrice F recherchée s'écrit comme la somme d'une partie forcée et d'une partie libre à estimer, c.-à-d.

$$F = \underbrace{\Phi_F}_{\text{valeurs forcées}} + \underbrace{\Delta_F}_{\text{valeurs libres}} \quad (3.18)$$

Il est évident que l'intersection de la partie forcée et de la partie libre de F doit être nulle, une valeur ne pouvant pas être à la fois connue et inconnue. Or, l'introduction de (3.18) comme une contrainte du problème d'optimisation de NMF (3.3) ne suffit pas à assurer cette intersection nulle entre la matrice Φ_F fixée et la matrice estimée Δ_F .

Limem [123] définit alors deux masques binaires Ω_F et $\bar{\Omega}_F$ informant respectivement de la présence ou de l'absence de contraintes égalité, c.-à-d.

$$\Omega_{f_{ij}} = \begin{cases} 1 & \text{si } f_{ij} \text{ est connu} \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad \bar{\Omega}_{F_{ij}} = \begin{cases} 0 & \text{si } f_{ij} \text{ est connu} \\ 1 & \text{sinon.} \end{cases} \quad (3.19)$$

Ainsi, en associant (3.18) et (3.19) dans le problème de NMF général celui-ci se réécrit

$$\begin{aligned} \{G, F\} &= \arg \min_{\tilde{G}, \tilde{F} \geq 0} \mathcal{D}(X || G \cdot F), \\ \text{s.c.} \quad \tilde{F} &= \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F, \end{aligned} \quad (3.20)$$

où \circ désigne le produit de Hadamard. Il est à noter que cette paramétrisation a été proposée dans le cadre d'un critère pondéré de NMF et des approches ont été proposées avec la norme de Frobenius [119], des divergences KL [121] et paramétriques (β - [120, 122], α - et $\alpha\beta$ -divergences [123]).

Contrainte de structure

Il arrive que les colonnes ou les lignes des matrices soient liées, imposant alors une certaine structure aux matrices recherchées. Dans [141, 140, 47] par exemple, la matrice F contient des colonnes qui sont des combinaisons d'autres colonnes de F . La matrice est alors séparée en deux parties $F = [F^{\mathcal{M}}, F^{\mathcal{E}}]$, où $F^{\mathcal{M}}$ est la partie maître contenant les colonnes de F linéairement indépendantes, et où $F^{\mathcal{E}}$ est la partie esclave, entièrement définie par la partie maître via une relation $F^{\mathcal{E}} = R(F^{\mathcal{M}})$. Cette contrainte de structure sur F mène alors au

problème d’optimisation

$$\begin{aligned} \{G, F\} &= \arg \min_{\tilde{G}, \tilde{F} \geq 0} \frac{1}{2} \cdot \|X - \tilde{G} \cdot \tilde{F}\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{F} &= [F^{\mathcal{M}}, F^{\mathcal{E}}], \\ F^{\mathcal{E}} &= R(F^{\mathcal{M}}). \end{aligned} \tag{3.21}$$

Dans [141, 140, 47], les colonnes de la matrice esclave $F^{\mathcal{E}}$ sont des produits de Hadamard des colonnes de la matrice maître $F^{\mathcal{M}}$. En factorisation tensorielle, des approches ont été proposées afin de permettre d’estimer des facteurs avec une structure de Vandermonde [195]. Toutefois, cette structure de Vandermonde n’est pas développée en factorisation matricielle et nous proposerons une méthode spécifique de factorisation pour ce genre de structure dans les travaux présentés dans cette thèse.

3.2.2 Les contraintes vues comme des pénalisations

Le problème de NMF (3.4) est très générique et il est classique d’y ajouter une ou plusieurs contraintes et/ou régularisations afin de guider la factorisation vers le résultat souhaité. Dans la section ci-après nous présentons les contraintes et pénalisations les plus couramment rencontrées dans la littérature.

Les informations supplémentaires qui sont ajoutées à la NMF standard (3.4) peuvent être incluses dans le problème d’optimisation de deux façons différentes. Outre certaines contraintes structurant la fonction d’attache aux données, de nombreuses contraintes – dites douces – peuvent être incluses au problème par l’ajout d’une pénalisation, menant alors au problème d’optimisation (3.2). Dans cette section nous présentons les contraintes les plus classiques qui peuvent s’inclure au problème d’optimisation au moyen d’un terme de pénalisation.

3.2.2.1 Pénalisations d’évolution lente

Dans certains problèmes (audio, video), il peut être intéressant de contraindre les lignes ou colonnes des facteurs à estimer à être proches. Dans [220] pour du traitement vidéo, la différence de deux colonnes successives de la matrice F doit être faible. Les auteurs proposent alors de lisser les colonnes en pénalisant le problème d’optimisation en y ajoutant le terme

$$\mathcal{C}(G, F) = \lambda \cdot \|F \cdot R\|_{\mathcal{F}}, \tag{3.22}$$

où

$$R = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & -1 \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \quad (3.23)$$

Le produit $(F \cdot R)$ fournit la matrice des différences des colonnes successives de F . La minimisation du critère (3.22) permet alors de pénaliser les différences entre ces colonnes et donc de lisser F . Dans [62] cette régularisation a également été utilisée pour l'estimation de spectrogramme lisses en application audio.

3.2.2.2 Pénalisations de lissage

Afin d'assurer l'obtention d'une solution "lisse" (*smooth* en anglais), il est classique d'utiliser la régularisation de Tikonov [155] qui ajoute une pénalisation au problème général (3.3). Cette dernière consiste à minimiser la norme de Frobenius de la matrice à lisser, et s'écrit alors pour F

$$\mathcal{C}(G, F) = \frac{\lambda}{2} \|F\|_{\mathcal{F}}^2, \quad (3.24)$$

où λ est le poids associé à la régularisation. Cette contrainte est tout à fait applicable à la matrice G . Dans [196] des contraintes de lissage sur G et F sont combinées pour assurer la construction d'une matrice de rang faible.

3.2.2.3 Pénalisations de parcimonie

Dans certaines applications [235], l'une des matrices recherchées peut être soumise à une contrainte de parcimonie, c.-à-d. qu'elle ne doit avoir que peu d'éléments non-nuls. Le terme de pénalisation associé à une contrainte de parcimonie sur une matrice utilise souvent la norme ℓ_0 qui donne le nombre d'éléments non-nuls de la matrice. La minimisation de cette norme dans le problème d'optimisation est difficile puisqu'elle relève des mathématiques combinatoires et il lui est souvent préféré une relaxation convexe. En 2002, Hoyer [93] a proposé de choisir comme fonction de pénalisation la norme ℓ_1 de la matrice F , c.-à-d.

$$\mathcal{C}(G, F) = \|F\|_1 = \sum_{ij} |f_{ij}|, \quad (3.25)$$

qui pour une matrice non-négative n'est autre que la somme de ses éléments. Hoyer explique alors que le problème de NMF avec comme unique contrainte (3.25) mène à des solutions dégénérées dans lesquelles la matrice G tend vers l'infini tandis que la matrice F tend vers zéro pour minimiser sa norme ℓ_1 . Ce problème de dégénérescence est alors fixé par normalisation des matrices G et F .

Plus récemment, Kim et Park [103] on proposé de choisir comme terme de pénalisation

$$\mathcal{C}(G, F) = \nu \cdot \left\| \tilde{G} \right\|_{\mathcal{F}}^2 + \lambda \cdot \sum_i \|\mathbf{f}_i\|_1, \quad (3.26)$$

où \mathbf{f}_i est la i -ème colonne de F . La norme $\left\| \tilde{G} \right\|_{\mathcal{F}}^2$ permet alors de limiter la dégénérescence de la solution évoquée précédemment, tandis que $\sum_i \|\mathbf{f}_i\|_1$ impose une certaine parcimonie de F . Les paramètres ν , λ sont les poids associés à chaque partie du terme de pénalisation. Cette pénalisation est un premier exemple de mélange de contraintes. Il s'agit ici de considérer une contrainte de lissage sur la matrice G ainsi qu'une contrainte de parcimonie sur F .

Dans [142], les auteurs proposent d'utiliser la norme $\ell_{1,2}$ comme contrainte de parcimonie. Cette dernière offre un compromis entre le lissage induit par la norme ℓ_2 et la parcimonie provoquée par la norme ℓ_1 . La norme $\ell_{1,2}$ consiste pour une matrice F en l'expression

$$\|F\|_{1,2} = \sum_i \left\| \bar{f}_i \right\|_2, \quad (3.27)$$

où \bar{f}_i est la i -ème ligne de F .

L'intérêt certain de la NMF avec contraintes de parcimonie [235] a inspiré de très nombreux autres travaux [108], utilisant des divergences paramétriques comme la divergence β ou KL [131], ou des critères de pénalisation particuliers comme les normes ℓ_p [181].

3.2.2.4 Pénalisations douces d'égalité

Si les valeurs approximatives des matrice G et F sont connues et contenues dans G^* et F^* , alors G et F peuvent être régularisées par l'ajout de la pénalisation

$$\mathcal{C}(G, F) = \frac{\lambda}{2} \|G - G^*\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \|F - F^*\|_{\mathcal{F}}^2, \quad (3.28)$$

au problème (3.3). Si le produit $(G \cdot F)$ doit permettre de reconstruire les données X , les matrice G et F doivent elles rester proches des données G^* et F^* . Cette pénalisation permet une régularisation douce, en n'imposant pas une égalité stricte entre les matrice G et G^* d'une part, et F et F^* d'autre part.

3.2.2.5 Pénalisations de faible rang

Dans (3.3), le rang de la matrice reconstruite par le produit $(G \cdot F)$ est fixé par les dimensions des matrices. Cependant, il peut être souhaité de rechercher une solution faible rang, sans pour autant connaître à l'avance la valeur du rang. De façon intuitive, la contrainte de faible rang peut être ajoutée en marge du problème d'optimisation (3.3) où la solution doit

vérifier

$$\text{Rang}(G \cdot F) \leq r, \quad (3.29)$$

où r est le rang maximum souhaité pour la matrice reconstruite. Toutefois, la résolution de (3.3) avec (3.29) est difficile et une alternative – issue des travaux en complétion de matrices [22] – consiste à pénaliser le rang de la matrice reconstruite directement dans le problème d’optimisation, en incluant dans (3.3) la pénalisation

$$\mathcal{C}(G, F) = \frac{\lambda}{2} \|G \cdot F\|_{\star}^2, \quad (3.30)$$

où, pour une matrice A , $\|A\|_{\star}$ est la norme nucléaire définie comme la somme des valeurs propres de A et λ est le poids de la pénalisation dans le problème globale (3.2). Une alternative à la norme nucléaire [196] consiste à minimiser la somme des normes de Frobenius de G et F et donc à utiliser la pénalisation

$$\mathcal{C}(G, F) = \frac{\lambda}{2} \|G\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \|F\|_{\mathcal{F}}^2. \quad (3.31)$$

Cette pénalisation est également un exemple de mélange de contraintes, en l’occurrence, il s’agit ici de considérer une contrainte de lissage sur chaque matrice G et F .

3.2.3 Autres pénalisations : contraintes de bornitude

Il arrive dans certains problèmes que toutes les contraintes ne soient pas insérables dans le problème d’optimisation sous forme de pénalisations. Ces contraintes sont alors ajoutées en marge du problème d’optimisation et sont traitées de façon particulière. C’est notamment le cas des contraintes de bornitudes, que nous détaillons ci-dessous.

Afin de faciliter la résolution du problème de factorisation, il peut être intéressant de réduire le champ de recherche des matrices facteur, par exemple, en bornant l’intervalle de recherche. Ceci est réalisable dans beaucoup d’applications où les valeurs des matrice recherchées ont un sens physique et où certaines valeurs sont interdites. Cette contrainte n’est pas incluse dans le problème sous la forme d’une pénalisation mais mène à la relation

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \cdot \left\| X - \tilde{G} \cdot \tilde{F} \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \mathcal{I}_G \leq \tilde{G} \leq \mathcal{S}_G, \quad \text{et} \quad \mathcal{I}_F \leq \tilde{F} \leq \mathcal{S}_F, \end{aligned} \quad (3.32)$$

où $\mathcal{I}_G, \mathcal{I}_F$ sont respectivement la borne inférieure de G et F et où $\mathcal{S}_G, \mathcal{S}_F$ sont respectivement la borne supérieure de G et F . Ce type de contrainte a été proposé par exemple par Lin [125] ou Limem *et al.* [122]. Dans les deux cas, la contrainte de bornitude est traitée par projection des valeurs en dehors des intervalles admissibles vers la borne la plus proche.

3.2.4 Unicité et exactitude de la NMF

Il est à noter que dans son formalisme général (3.2), le problème de NMF n'est pas convexe et présente de multiples optima locaux. Les conditions garantissant l'unicité des solutions ont donc été particulièrement étudiées. Ces aspects ont été présentés de manière détaillée dans [123] et nous n'en fournissons ci-dessous qu'une synthèse.

Les premiers travaux de la littérature ont étudié la convergence de la NMF en utilisant les conditions de Karush-Kuhn Tucker (KKT) [124]. Celles-ci sont des conditions nécessaires pour obtenir un optimum local mais la suffisance n'est garantie que pour des fonctions de coûts convexes, ce qui n'est pas le cas des problèmes généraux de NMF. L'unicité de la solution a par la suite été étudiée par Donoho et Stodden [49] et Moussaoui [145]. Ces auteurs ont montré qu'à une matrice de permutation près, l'unicité de la solution de la NMF est possible si certaines conditions de parcimonie conjointe des facteurs matriciels – conditions appelées quasi-séparabilité dans [49] – est vérifiée. Ces hypothèses ont été particulièrement étudiées en démélange hyperspectral (dans le cas de mélange linéaire) où ces conditions de parcimonie sont vérifiées par les sources à séparer [12].

Un autre problème actuel en NMF concerne l'exactitude de la factorisation, dans le cas où la matrice X n'est pas bruitée. Les approches présentées dans ce chapitre et celles développées dans cette thèse sont des méthodes de factorisation approchée. Obtenir une solution exacte de NMF peut être vu comme un problème combinatoire [211]. Des approches exactes – faisant appel à des hypothèses de parcimonie proches de [49] voire plus contraignantes – ont été proposées par exemple dans [4, 13, 210].

3.3 Optimisations usuelles pour les algorithmes alternés

Bien qu'il soit clairement identifié par (3.3), la résolution du problème de NMF n'est pas triviale, celui-ci étant non-convexe par rapport aux deux matrices G et F . Les premiers travaux sur la NMF remontent aux années 90. Introduits par Paatero *et al.* [151, 152, 3], on parle alors de PMF pour factorisation en matrices positives. Dans ces premiers travaux [151], les matrices G et F sont estimées simultanément, en résolvant un problème de moindres carrés. L'utilisation de cette méthode est limitée à des méthodes de petite taille. Depuis, la communauté scientifique en traitement du signal et en apprentissage s'est fortement intéressée au problème de NMF, et des stratégies alternées permettant la résolution du problème d'optimisation ont été très étudiées. Cette stratégie consiste à considérer non pas le problème non-convexe d'optimisation général (3.3) mais les sous-problèmes en G et en F , c.-à-d.

$$G = \arg \min_{\tilde{G} \geq 0} \mathcal{J}(\tilde{G}, F), \quad (3.33)$$

$$F = \arg \min_{\tilde{F} \geq 0} \mathcal{J}(G, \tilde{F}). \quad (3.34)$$

L'intérêt des approches alternées réside dans le fait que pour la plupart des critères d'attache aux données vus précédemment, les sous-problèmes (3.33) et (3.34) sont convexes et beaucoup plus simples à gérer.

Les méthodes alternées de résolution du problème de NMF sont généralement basées sur l'une des techniques d'optimisation suivante : l'approche heuristique, la descente de gradient, la majoration-minimisation (stratégie MM) ou des méthodes proximales comme par exemple la méthode des directions alternées (ADM pour *Alternating Direction Method*).

3.3.1 Méthode heuristique

L'approche heuristique est l'approche d'optimisation qui a été proposée par Lee et Seung [112]. Cette technique peut être vue comme une descente de gradient avec un pas spécifique et consiste en une mise à jour multiplicative. Elle consiste à écrire le gradient de $\mathcal{J}(G, F)$ par rapport à un facteur – disons F par exemple – comme la différence entre deux termes de signes opposés, c.-à-d., en partant de l'équation (3.5)

$$\nabla_F \mathcal{J}(G, F) = \frac{\partial \mathcal{J}(G, F)}{\partial F} = G^T \cdot G \cdot F - G^T \cdot X, \quad (3.35)$$

$$= \nabla_F^+ \mathcal{J}(G, F) - \nabla_F^- \mathcal{J}(G, F), \quad (3.36)$$

où

$$\nabla_F^+ \mathcal{J}(G, F) \triangleq G^T \cdot G \cdot F \quad (3.37)$$

et

$$\nabla_F^- \mathcal{J}(G, F) \triangleq G^T \cdot X. \quad (3.38)$$

La formule de mise à jour de F consiste alors en

$$F = F \circ \frac{\nabla_F^- \mathcal{J}(G, F)}{\nabla_F^+ \mathcal{J}(G, F)}. \quad (3.39)$$

De manière similaire, la mise à jour de G s'effectue en calculant

$$\nabla_G^+ = G \cdot F \cdot F^T, \quad \text{et} \quad \nabla_G^- = X \cdot F^T, \quad (3.40)$$

ce qui mène à la mise à jour

$$G = G \circ \frac{\nabla_G^- \mathcal{J}(G, F)}{\nabla_G^+ \mathcal{J}(G, F)}. \quad (3.41)$$

En partant de matrices initiales positives, les mises à jour multiplicative, qu'elles soient obtenues par la méthode heuristique ou une autre approche (comme la stratégie de Majoration-Minimisation discutée ci-après) garantissent la positivité des solutions. L'idée consiste – mis à part pour les valeurs nulles – à avoir les rapports dans (3.39) et (3.41) qui tendent vers 1, ce qui conduit les gradients à tendre vers zéro.

3.3.2 Descente de gradient

La méthode classique de descente de gradient pour la recherche du minimum d'une fonction consiste, à partir d'un point de paramètres initial, à mettre à jour ce point en suivant une direction diminuant le critère. De façon plus formelle, la direction choisie, appelée *direction de descente* est définie comme

Définition 3.1

d^k est une direction de descente si, partant d'un point courant f^k

$$\nabla \mathcal{D} \left(f^k \right)^T \cdot d^k < 0. \quad (3.42)$$

Avec une direction de descente d^k , tout point construit par la relation

$$f^{k+1} = f^k + \rho^k \cdot d^k, \quad (3.43)$$

avec un pas de descente $\rho^k > 0$ suffisamment petit, fait décroître la fonction de coût $\mathcal{D}(\cdot)$. Il est très classique de choisir comme direction de descente, l'opposé du gradient, c.-à-d. $d^k = -\nabla \mathcal{D} \left(f^k \right)^T$. Ces méthodes sont cependant connues pour leur convergence lente en cas de mauvais choix du pas de descente ρ^k . Des versions accélérées comme par exemple la méthode du gradient conjugué leur sont souvent préférées. Il est à noter que les formules de mise à jour multiplicatives obtenues avec la méthode heuristique discutée ci-dessus peuvent être vues sous l'angle de la descente de gradient, avec un choix particulier du pas [112] qui permet de transformer la somme dans (3.43) en une mise à jour multiplicative qui s'écrit :

$$f^{k+1} = f^k \cdot \alpha^k \cdot d^k, \quad (3.44)$$

De nombreuses méthodes de NMF construites à partir de descente de gradient ont été proposées, par exemple [112, 93, 125, 140].

3.3.3 Approche Majoration-Minimisation (MM)

En considérant le problème d'optimisation convexe

$$f = \arg \min_{\tilde{f}} \mathcal{D} \left(\tilde{f} \right), \quad (3.45)$$

où $\mathcal{D}(\cdot)$ est la fonction de coût du problème, alors, le résoudre itérativement par majoration-minimisation (MM) consiste à

1. Construire une fonction majorante généralement convexe $\mathcal{H}(f, f^k)$ de la fonction de coût initiale $\mathcal{D}(\cdot)$, et qui soit tangente à $\mathcal{D}(\cdot)$ au point f^k . Cette fonction est appelée fonction auxiliaire de $\mathcal{D}(\cdot)$ au point f^k . La fonction \mathcal{H} a donc les propriétés

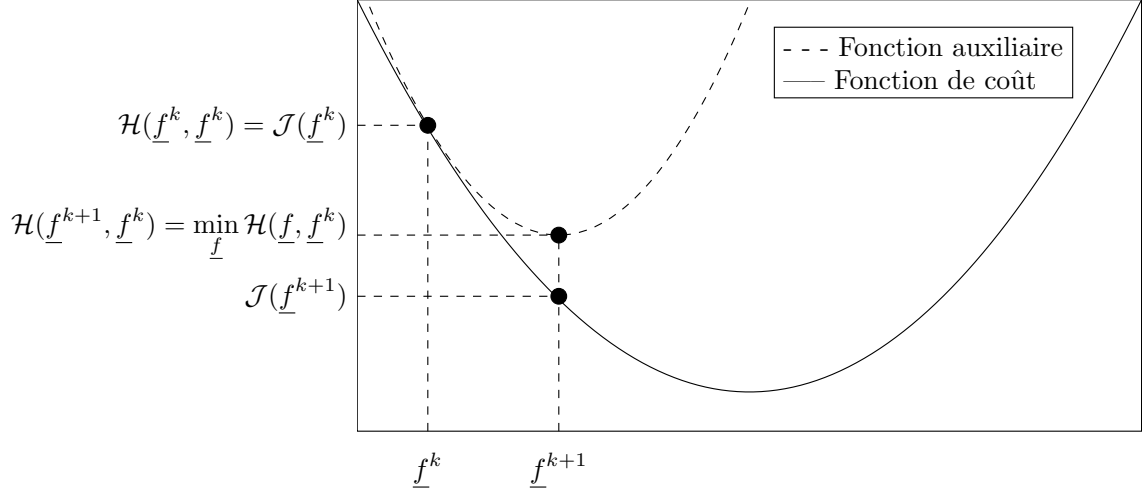


FIGURE 3.1 – Illustration du principe des méthodes MM.

suivantes

$$\mathcal{H}(f, f^k) \geq \mathcal{D}(f), \quad (3.46)$$

$$\mathcal{H}(f^k, f^k) = \mathcal{D}(f^k). \quad (3.47)$$

2. Calculer f^{k+1} par minimisation de $\mathcal{H}(f, f^k)$ au lieu de minimiser directement $\mathcal{D}(f)$.
3. Répéter jusqu'à atteindre un critère d'arrêt ou un nombre maximum d'itérations.

La mise à jour par la formule $f^{k+1} = \arg \min_{f \geq 0} \mathcal{H}(f, f^k)$ permet, grâce aux propriétés de la fonction auxiliaire (3.46) et (3.47), d'assurer la décroissance de la fonction de coût $\mathcal{D}(\cdot)$. Le principe de fonctionnement de la recherche de minimum par méthode MM est illustré dans la figure 3.1.

En utilisant cette méthode MM de façon alternée sur les sous-problèmes en G (3.33) et en F (3.34), de nombreuses méthodes de NMF ont été proposées, par exemple [112, 65, 119].

3.3.4 Méthode par moindres carrés alternés (ANLS)

Contrairement aux approches d'optimisation présentées jusqu'ici, les approches basées sur les moindres carrés alternés cherchent réellement à minimiser (3.33) et (3.34) à chaque itération de l'algorithme, une fois l'autre facteur estimé à l'itération précédente. Pour cela, ces techniques de NMF consistent, dans leur version de base, à appliquer une approche de moindres carrés non-négatifs, assez gourmande en temps de calcul. Plusieurs extensions plus rapides, par exemple [39, 102] ont proposé d'accélérer ce type de techniques en partitionnant les données de G et F , respectivement.

La méthode des moindres carrés hiérarchiques alternés [39] partitionne le problème de NMF de base en $2 \cdot r$ sous-problèmes, où r est le nombre de colonnes (respectivement de

lignes) de la matrice G (resp. F). La fonction de coût \mathcal{J} est alors reformulée comme

$$\mathcal{J}(\mathbf{g}_1, \dots, \mathbf{g}_r, \bar{f}_1, \dots, \bar{f}_r) = \frac{1}{2} \cdot \left\| X - \sum_{l=1}^r \mathbf{g}_l \cdot \bar{f}_l \right\|_{\mathcal{F}}^2, \quad (3.48)$$

où pour une matrice A , \mathbf{a}_l représente sa l -ème colonne et \bar{a}_l sa l -ème ligne. Ce problème est alors partitionnable en sous-problèmes suivant les colonnes de G et les lignes de F (différent donc de l'approche ANLS qui utilise un partitionnement dans le sens contraire). Ainsi, pour la recherche de \bar{f}_l , en définissant le résidu $R(l) = X - \sum_{i \neq l}^r \mathbf{g}_i \cdot \bar{f}_i$, alors

$$R(l) \simeq \mathbf{g}_l \cdot \bar{f}_l, \quad (3.49)$$

et le problème d'optimisation associé à (3.49) s'écrit

$$\bar{f}_l = \arg \min_{\bar{f} \geq 0} \frac{1}{2} \left\| R(l) - \mathbf{g}_l \cdot \bar{f} \right\|_{\mathcal{F}}^2. \quad (3.50)$$

Ce dernier peut être re partitionné en problèmes indépendants suivant un élément f_{lk} de F , ce qui conduit à la résolution de

$$\tilde{f}_{lj} = \arg \min_{f \geq 0} \|\mathbf{r}_j(l) - \mathbf{g}_l \cdot f\|_{\mathcal{F}}, \quad \forall j \in \{1, \dots, n\}, \quad (3.51)$$

où $\mathbf{r}_j(l)$ est la j -ème colonne de $R(l)$. La solution de ce problème pour un élément f_{lk} est obtenue de façon triviale par la formule

$$f_{lj} = \left[\frac{\mathbf{g}_l^T \cdot \mathbf{r}_j(l)}{\mathbf{g}_l^T \cdot \mathbf{g}_l} \right]^+, \quad (3.52)$$

et est facilement généralisable à l'ensemble du vecteur ligne \bar{f}_l donnant alors la formule de mise à jour de \bar{f}_l . Toute la démarche pour arriver à l'écriture de cette formule de mise à jour est réalisable pour une colonne de G .

3.4 Méthodes "historiques" de NMF

Dans le tableau 3.2, nous présentons les principales méthodes "historiques" de NMF proposées dans la littérature, ainsi que les contraintes associées. La liste des méthodes présentées dans ce tableau est non-exhaustive, l'attention étant portée sur les approches de la littérature utilisant un terme d'attache aux données basée sur la norme de Frobenius. Il est à noter que la NMF a aussi été étudiée sous l'angle d'approches bayésiennes, par exemple dans [146, 184, 185], qui ne sont pas présentées dans ce tableau.

méthode	fonction de coût	contraintes	mises à jour
Lee et Seung [113, 112]	Frobenius ou KL		multiplicatives
Lin [125]	Frobenius	bornitude (égalité stricte possible)	gradient projeté
Cichocki [39]	Frobenius	parcimonie sur G et F	moindres carrés hiérarchiques alternés
Kim et Park [102]	Frobenius		moindres carrés non-négatifs alternés
Hoyer [93]	Frobenius	parcimonie sur F	multiplicative
Guillamet [78]	Frobenius pondérée par lignes/colonnes		multiplicative
Guillamet [79]	Frobenius pondérée terme à terme		multiplicatives
Ho [91]	Frobenius pondérée terme à terme		multiplicatives

TABLE 3.2 – Tableau des méthodes "historiques" de NMF.

3.5 Méthodes de NMF pour les problèmes de grandes dimensions

3.5.1 Méthodes de NMF accélérées

Aujourd'hui, le traitement des *Big Data* monopolise beaucoup d'attention en recherche. L'acquisition de jeux de données de très grandes dimensions étant de plus en plus facile, les méthodes de traitement doivent s'adapter pour permettre une analyse/extraction des informations en un temps raisonnable. En NMF, la matrice de données X à factoriser peut être gigantesque (matrice 480000×17000 pour les données du prix Netflix¹), et les méthodes de factorisation présentées précédemment ne sont pas adaptées pour gérer une telle quantité d'information. Dans les sections suivantes, nous proposons un bref aperçu de quelques méthodes visant à accélérer la NMF standard.

3.5.1.1 Méthode de NMF par mises à jour itératives alternées

Dans [73], les auteurs étudient la complexité des formules de mise à jour multiplicatives de Lee et Seung [113]. Ils proposent alors de modifier l'algorithme de Lee et Seung en effectuant à chaque itération plusieurs mises à jour successives pour G et pour F , de façon à réduire le nombre d'évaluations des termes les plus coûteux en calculs. En suivant la même démarche, Gillis *et al.* proposent également une version accélérée de l'algorithme de NMF par HALS [39].

3.5.1.2 NMF par gradient de Nesterov

La méthode de gradient optimal de Nesterov [149], est une méthode d'optimisation rapide permettant la résolution du problème

$$F = \arg \min_{\tilde{F}} f(\tilde{F}), \quad (3.53)$$

dans le cas où la fonction $f(\cdot)$ est convexe et de gradient Lipschitz-continu, c.-à-d. s'il existe une constante L telle que pour tout X, Y du domaine de définition de $f(\cdot)$,

$$\|f(X) - f(Y)\| \leq L \cdot \|X - Y\|. \quad (3.54)$$

En calculant la valeur de la constante de Lipschitz L , la méthode de gradient optimal de Nesterov consiste en une succession de descentes de gradient avec un pas adaptatif dépendant de L .

En considérant les sous-problèmes de NMF (3.33) et (3.34), Guan *et al.* [76] montrent que la fonction de coût associée à chacun des sous-problèmes est convexe et de gradient Lipschitz-

1. Voir <http://www.netflixprize.com/>.

continu. Ils proposent alors d’alterner les mises à jour des matrices en utilisant le gradient optimal de Nesterov. Cette méthode de NMF à base de gradient de Nesterov (NeNMF pour *Nesterov-based NMF*) fait aujourd’hui partie des méthodes les plus rapides de NMF selon [193]. Son extension pondérée – permettant éventuellement de traiter des données manquantes – a été proposé dans [52] et est détaillée dans l’Annexe C.

3.5.1.3 Méthode de NMF aléatoire (*Randomized NMF*)

Lorsque le problème à traiter est de grande taille, il est probable qu’il soit de faible rang, c.-à-d. que la taille de données nécessaires à l’estimation des variables latentes soit beaucoup plus faible que la taille de la matrice X . Cette hypothèse est au cœur des approches d’ACP robuste et de complétion de matrices [241]. En particulier, l’utilisation d’opérateurs de mélange aléatoires pour réduire la taille des données et accélérer une décomposition ont fait l’objet de nombreux travaux, qui sont en partie regroupés dans [82]. Combiner ces stratégies avec la NMF a été proposé par plusieurs auteurs.

Une autre façon de traiter les matrices de très grande dimension consiste à n’en considérer qu’une partie. Dans [219], les auteurs proposent de mettre à jour les matrices à partir de sous-matrices de données de tailles réduites, extraites de la matrice complète X par projections aléatoires [212]. Wang *et al.* [219] proposent donc de résoudre de façon alternée les problèmes d’optimisation suivants

$$G = \arg \min_{\tilde{G}} \frac{1}{2} \cdot \left\| X \cdot R_G - \tilde{G} \cdot (F \cdot R_G) \right\|_{\mathcal{F}}, \quad (3.55)$$

et

$$F = \arg \min_{\tilde{F}} \frac{1}{2} \cdot \left\| R_F \cdot X - (R_F \cdot G) \cdot \tilde{F} \right\|_{\mathcal{F}}, \quad (3.56)$$

où les tailles des matrices R_G et R_F sont définies telles que les produits $X \cdot R_G$ et $R_F \cdot X$ réduisent la matrice X , c.-à-d. que si $X \in \mathbb{R}^{m \times n}$ alors $R_G \in \mathbb{R}^{m \times r_G}$ et $R_F \in \mathbb{R}^{r_F \times n}$ avec $r_G \ll n$ et $r_F \ll m$. Wang *et al.* [219] utilisent alors les formules classiques de mise à jour multiplicative de Lee et Seung [113] pour les deux sous-problèmes en G et F .

Dans [239], les auteurs décomposent la matrice X par ACP randomisée – pour stocker un produit $A \cdot S$ beaucoup moins volumineux – et proposent de nouvelles mises à jour multiplicative mettant en jeu $A \cdot S$ au lieu de X . Les auteurs de [204] ont proposé des techniques similaires aux deux approches ci-dessus mais les ont étendues au cas où la NMF est séparable.

3.5.1.4 NMF distribuée

Sur un fond plus orienté informatique, et sur le même principe que les méthodes ANLS et HALS, certains auteurs ont proposé de découper les sous-problèmes de NMF en problèmes indépendants suivant les lignes/colonnes des matrices G et F . Plusieurs stratégies sont alors envisagées pour distribuer les calculs de façon à profiter au maximum de la puissance du

calculateur à disposition (calcul GPU, cluster de calcul, etc..) [129, 106, 118, 99].

3.5.2 NMF pour flux de données

Dans le cas où la factorisation doit être effectuée en temps réel sur une matrice X de données entrantes, alors X est construite au fur et à mesure de l'arrivée des données. Des méthodes spécifiques de NMF doivent alors être utilisées, permettant l'insertion des nouvelles données dans X , et/ou l'oubli de données trop anciennes. On parle alors de NMF en ligne (*Online NMF*) ou de NMF par flux (*Streaming NMF*). De nombreuses approches pour gérer la factorisation de matrices de données entrantes ont été proposées [135, 116, 77, 237, 238], pour la plupart basées sur une construction itérative de X dans laquelle les nouvelles données reçues sont ajoutées dans une nouvelle ligne ou colonne de X , se traduisant par l'ajout d'une ligne de G ou colonne de F à rechercher.

3.5.3 Co-factorisation

Dans certaines situations, plusieurs jeux de données à factoriser peuvent avoir des similarités dans leur décomposition factorielle. Plutôt que de considérer un problème d'optimisation individuel pour chaque jeu de données, certains auteurs ont proposé de lier les problèmes en tenant compte de ces similarités entre les différents facteurs. On parle alors de co-factorisation, de factorisations groupées ou encore de factorisations jointes [114, 231, 130, 187]. Pour simplifier les explications, nous présentons ci-dessous les problèmes dans le cas où seuls deux jeux de données, notés X_1 et X_2 , sont disponibles. Les facteurs à estimer sont notés G_i et F_i pour $i = 1, 2$.

Le problème traité par la co-factorisation consiste souvent à supposer strictement égaux certains facteurs, par exemple $F_1 = F_2 = F$, et à rechercher des matrices G_1, G_2, F , telles que pour deux matrices de données X_1 et X_2

$$\{G_1, G_2, F\} = \arg \min_{\tilde{G}_1, \tilde{G}_2, \tilde{F}} \frac{1}{2} \cdot \left\| X_1 - \tilde{G}_1 \cdot \tilde{F} \right\|_{\mathcal{F}}^2 + \frac{1}{2} \cdot \left\| X_2 - \tilde{G}_2 \cdot \tilde{F} \right\|_{\mathcal{F}}^2, \quad (3.57)$$

c.-à-d. que la matrice F est commune à la factorisation des matrices de données X_1 et X_2 .

Dans [130], les auteurs ont proposé une version moins contraignante du problème (3.57), dans lequel la matrice F n'est plus rigoureusement identique pour chaque problème, mais autorise quelques variations : certains facteurs – par exemple les F_i , $i = 1, 2$ – doivent être "proches" d'une matrice F^* connue (à une indétermination de permutation et de facteur d'échelle près, que l'on suppose résolue ci-dessous) et proposent d'optimiser le problème suivant :

$$\{G_1, G_2, F_1, F_2\} = \arg \min_{\tilde{G}_1, \tilde{G}_2, \tilde{F}_1, \tilde{F}_2} \frac{1}{2} \cdot \left\| X_1 - \tilde{G}_1 \cdot \tilde{F}_1 \right\|_{\mathcal{F}}^2 + \frac{1}{2} \cdot \left\| X_2 - \tilde{G}_2 \cdot \tilde{F}_2 \right\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \sum_{i=1}^2 \left\| \tilde{F}_i - F^* \right\|_{\mathcal{F}}^2. \quad (3.58)$$

Les auteurs de [187] ont proposé une alternative, uniquement pour le cas de deux matrices à factoriser, dans lequel on contraint F_1 et F_2 à être "proches" entre elles, c.-à-d.

$$\{G_1, G_2, F_1, F_2\} = \arg \min_{\tilde{G}_1, \tilde{G}_2, \tilde{F}_1, \tilde{F}_2} \frac{1}{2} \cdot \left\| X_1 - \tilde{G}_1 \cdot \tilde{F}_1 \right\|_{\mathcal{F}}^2 + \frac{1}{2} \cdot \left\| X_2 - \tilde{G}_2 \cdot \tilde{F}_2 \right\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \cdot \left\| \tilde{F}_1 - \tilde{F}_2 \right\|_{\mathcal{F}}^2, \quad (3.59)$$

ou

$$\{G_1, G_2, F_1, F_2\} = \arg \min_{\tilde{G}_1, \tilde{G}_2, \tilde{F}_1, \tilde{F}_2} \frac{1}{2} \cdot \left\| X_1 - \tilde{G}_1 \cdot \tilde{F}_1 \right\|_{\mathcal{F}}^2 + \frac{1}{2} \cdot \left\| X_2 - \tilde{G}_2 \cdot \tilde{F}_2 \right\|_{\mathcal{F}}^2 + \lambda \cdot \left\| \tilde{F}_1 - \tilde{F}_2 \right\|_1. \quad (3.60)$$

Alors que (3.59) impose des contraintes de lissage dans la différence entre les facteurs F_1 et F_2 , le problème (3.60) contraint cette différence à être parcimonieuse.

Il est à noter que des approches de co-factorisation ont aussi été proposées lorsque les facteurs matriciels sont de taille différente, par exemple pour la factorisation conjointe de données audio et vidéo [132] ou d'images multi- et hyper-spectrales [9].

3.5.4 Factorisation tensorielle

Il peut également arriver que les données ne puissent pas être "naturellement" stockées sous forme d'une matrice X . Si une dimension supplémentaire est nécessaire pour stocker les données, alors il est possible de les organiser sous la forme d'un tenseur, qui n'est autre qu'un tableau multi-dimensions de données, et pour lequel des opérations telles que l'addition ou la multiplication sont définies de façon particulière. Les données écrites sous forme de tenseur peuvent alors être factorisées, par des méthodes spécifiques de factorisation tensorielle [104, 191] avec ou sans contrainte de signe. Dans le cas où tous les facteurs sont non-négatifs, les méthodes sont regroupées sous l'acronyme NTF pour *Nonnegative Tensor Factorization* [189, 240].

Historiquement, les tenseurs de dimension 3 ont souvent été ré-écrits sous forme de matrices avant d'être factorisés, par exemple par des approches de NMF. Cependant, ce traitement pouvait amener des pertes d'informations utiles, dues à la structure multi-voie intrinsèque aux données. Fort heureusement, les décompositions de type NTF héritent non-seulement de tous les avantages de NMF (dont l'interprétabilité des facteurs), mais bénéficient également des avantages dûs aux structures multilinéaires des données organisées en tenseurs [240]. Il est à noter que de nombreuses fonctions de coût et de régularisation présentées dans ce chapitre pour la NMF ont aussi été considérées en NTF. On trouve notamment dans la littérature des approches capables de gérer l'absence de données [177], conçues pour traiter des problèmes de grande dimension [37, 214] ou encore développées avec un cadre bayésien [94].

Les travaux réalisés dans cette thèse s'appuyant exclusivement sur des approches de factorisation matricielle, nous n'entrerons pas ici dans les détails des approches de factorisation tensorielle, qui nécessitent de redévelopper l'ensemble des opérations de base. Le lecteur in-

téressé trouvera par exemple énormément d'information sur ce type de factorisation dans [191].

3.6 Cas des données manquantes

Comme évoqué dans la section 3.2.1.5, la matrice de données X peut avoir des données manquantes. Celles-ci peuvent être prises en compte dans la factorisation au moyen d'une matrice de pondération W (3.17). Les travaux présentés dans cette thèse rentrent aussi dans cette catégorie, puisque les données à traiter ne seront que partiellement connues. Cependant, il existe une riche littérature sur l'estimation de données manquantes dans des matrices de faible rang [44], depuis les travaux pionniers de [64, 22, 225]. Le problème consiste à estimer une matrice M de faible rang qui approche la matrice X partiellement observée, c.-à-d.

$$M = \arg \min_{\tilde{M}} \|W \circ (\tilde{M} - X)\|_F + \lambda \text{Rang}(\tilde{M}), \quad (3.61)$$

où W est l'opérateur binaire indiquant la présence ou absence de données dans X , \circ est le produit de Hadamard et $\text{Rang}(\cdot)$ est un opérateur dont l'expression pénalisée dans (3.61) permet de minimiser le rang de la matrice à estimer. Minimiser le rang est généralement un problème complexe et l'idée souvent utilisée en complétion de matrices consiste à rendre convexe (3.61) en remplaçant l'opérateur de Rang par une norme nucléaire [22, 225], comme discuté dans la section 3.2.2.5.

Le problème (3.61) présente des similitudes avec les problèmes de factorisation matricielle si on ré-écrit la matrice M à estimer sous la forme d'un produit matriciel [20]. Cependant, les approches historiques, basées par exemple sur des décompositions en valeurs singulières tronquées, ont plutôt été vues comme des extensions matricielles à l'acquisition compressée [22, 44], permettant ainsi de rattacher à cette problématique tous les résultats obtenus en acquisition compressée en termes de probabilité d'estimation de données manquantes en fonction de l'échantillonnage des données.

3.7 Conclusion

La littérature traitant du problème de factorisation en matrices non-négatives (NMF) est extrêmement riche puisque ce type de problème intéresse les chercheurs depuis les années 90 et est aujourd'hui encore un sujet de recherche sur lequel se penche une grande communauté.

L'état de l'art qui a été présenté dans ce chapitre est non-exhaustif mais procure une vue générale des méthodes permettant de traiter le problème de NMF, en partant des approches historiques/classiques [113, 125, 39, 102] jusqu'aux méthodes plus modernes adaptées aux problématiques actuelles [219, 129, 135, 76, 73, 118, 106, 99, 232], tout en montrant que la NMF permet de tenir compte d'un très grand nombre de contraintes et informations supplémentaires [78, 48, 103, 241, 119, 140, 123, 108, 181, 221].

Dans les chapitres suivants, nous proposons de développer des méthodes d'étalonnage d'un réseau de capteurs mobiles, à base de factorisation pondérée et informée en matrices non-négatives et semi-non-négatives. En particulier, nous nous servirons dans le chapitre 4 de la paramétrisation développée dans la thèse de Limem [123] pour informer le problème de NMF dans le cadre d'un étalonnage de capteurs à réponse affine. Dans le chapitre 5, nous introduirons cette paramétrisation dans un problème de NMF structurée pour lequel nous développerons des formules de mise à jour à base de gradient projeté.

Méthodes de NMF informées pour l'étalonnage d'un réseau de capteurs mobiles à réponse affine

Nous présentons ici une partie des développements réalisés dans cette thèse pour l'étalonnage de capteurs et qui ont fait l'objet des publications [50, 53]. Dans un premier temps nous présentons les définitions et hypothèses nécessaires à la mise en place de nos méthodes d'étalonnage. Nous montrons ensuite comment poser ce problème sous forme d'une factorisation matricielle informée. Enfin nous proposons différentes approches permettant la résolution du problème de factorisation tout en respectant les contraintes qui auront été introduites.

4.1 Définitions et hypothèses

Les campagnes de mesure par foule de capteurs (en anglais *mobile crowdsensing*), utilisent des données datées et géolocalisées en provenance de nombreux capteurs mobiles, déployés sur une zone géographique bien définie.

Les méthodes d'étalonnage que nous proposons dans ce manuscrit sont dédiées au *mobile crowdsensing* pour la surveillance de la qualité de l'air. Nous considérerons donc dans tout ce manuscrit qu'un réseau de n capteurs mobiles effectue des mesures sur une zone géographique fixée. De façon générale, nous écrivons la mesure x_j fournie par le j -ème capteur du réseau lorsqu'il est exposé à une valeur physique y comme

$$x_j = \mathcal{F}_j(y), \quad (4.1)$$

où $\mathcal{F}_j(\cdot)$ est la fonction d'étalonnage dont la forme générale est supposée connue, associée au capteur et dont nous cherchons à estimer les paramètres, de façon à retrouver la valeur physique d'entrée y .

Dans un premier temps, nous cherchons à réécrire les données fournies par l'ensemble du réseau de capteur sous forme d'une matrice de données X . Pour cela, nous utilisons le formalisme de rendez-vous de la définition 2.3 page 12.

Un *rendez-vous* est caractérisé par une distance spatiale Δ_d et une durée Δ_t . Lorsque deux

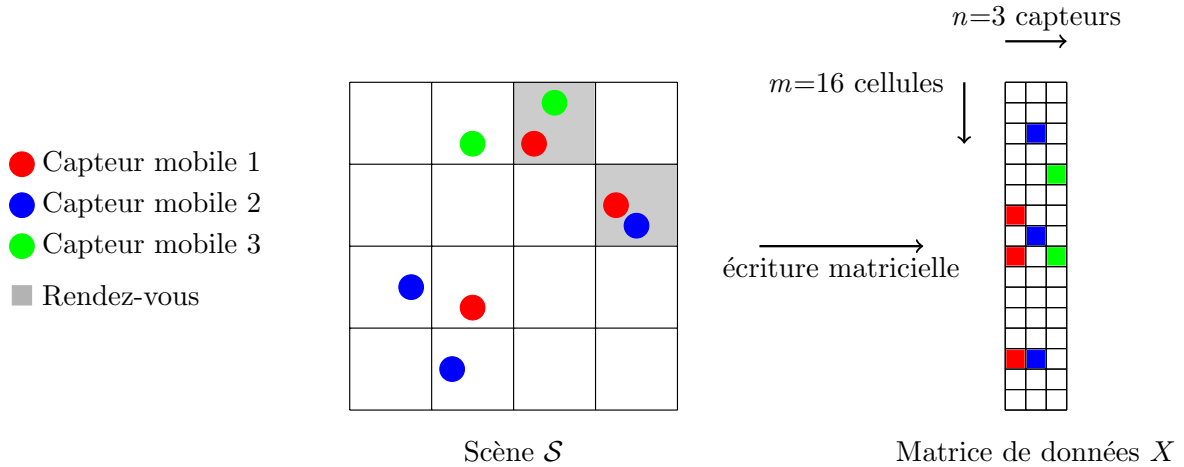


FIGURE 4.1 – D’une scène \mathcal{S} vers une matrice de données X .

capteurs sont en rendez-vous, la fluctuation du phénomène physique entre deux emplacements de mesure éloignés d’une distance inférieure à Δ_d pendant un intervalle de temps $[t, t + \Delta_t]$ est négligeable. Les capteurs en rendez-vous sont donc supposés être exposés à la même valeur physique d’entrée $y(t)$. Le choix de la distance Δ_d et de la durée Δ_t définissant un rendez-vous est très dépendant du phénomène physique observé [183]. Par exemple, le monoxyde de carbone est très fugace et, pour garantir que deux capteurs soient en rendez-vous et observent bien la même concentration de ce gaz, les paramètres Δ_d et Δ_t doivent être choisis bien plus petits que pour l’observation d’un gaz tel que l’ozone, moins volatile.

Nous définissons alors une *scène* comme :

Définition 4.1

Une scène \mathcal{S} est une zone discrétisée, observée durant un intervalle de temps $[t, t + \Delta_t]$. La taille d’une cellule est fixée de sorte que si deux capteurs sont dans une même cellule, alors ils sont l’un de l’autre à une distance inférieure à Δ_d .

En d’autres termes, une scène \mathcal{S} n’est autre qu’une discrétisation du champ physique observé pendant un intervalle de temps durant lequel il n’y a pas d’évolution du champ. Si deux capteurs se trouvent dans une même cellule de la scène, alors, par construction, ils sont en rendez-vous et observent la même valeur physique d’entrée.

En supposant que la scène \mathcal{S} est divisée en m cellules et en notant $x_{i,j}$ la mesure produite par le capteur j dans la i -ème cellule de \mathcal{S} , alors nous pouvons ordonner les mesures prises par chaque capteur dans chaque cellule, sous la forme d’une matrice $X \in \mathbb{R}^{m,n}$ dont un élément $x_{i,j}$ correspond à la mesure du j -ème capteur dans la cellule i . Ainsi, une ligne de X contient les mesures de tous les capteurs dans une même cellule, tandis qu’une colonne de X contient les mesures d’un même capteur pour chaque cellule. La construction de la matrice X à partir de la définition de la scène \mathcal{S} est illustrée au niveau de la figure 4.1. L’indice i de la cellule est obtenu en scrutant les colonnes de la scène \mathcal{S} de haut en bas, de gauche à droite. La matrice

X est donc une concaténation des colonnes de la scène pour chaque capteur.

La matrice de données X construite par mise en forme des mesures sur une scène \mathcal{S} est la base à partir de laquelle nous proposons d'élaborer nos méthodes d'étalonnage par factorisation matricielle présentées ci-après. Il est à noter que l'ensemble des matrices de données X_k issues de plusieurs scènes $\mathcal{S}_k - k$ étant un indice temporel – peuvent sans difficulté être concaténées sous la forme d'une unique matrice de données X . Dans la suite de ce chapitre, nous développons des approches pour étalonner un réseau de capteurs à partir de données issues d'une unique scène. Les méthodes proposées peuvent être trivialement étendues au cas multi-scènes.

4.2 Mise en place de l'étalonnage par factorisation

4.2.1 Écriture matricielle du problème d'étalonnage

Dans cette section, nous nous intéressons à l'étalonnage de capteurs à réponses affines, c.-à-d. aux capteurs dont la mesure de sortie est liée à la valeur physique d'entrée par une relation affine. Ainsi, la mesure $x_{i,j}$ fournie par le capteur j dans la cellule i d'une scène \mathcal{S} est liée à la valeur physique y_i de la cellule par la relation

$$x_{i,j} = f_{1,j} + y_i \cdot f_{2,j}, \quad (4.2)$$

où $f_{1,j}$ est l'offset du capteur et $f_{2,j}$ est son gain. En considérant la relation (4.2) pour toutes les cellules $i \in \llbracket 1 \cdots m \rrbracket$ et tous les capteurs $j \in \llbracket 1 \cdots n \rrbracket$, alors la matrice de données X peut s'écrire comme le produit

$$\underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix}}_X = \underbrace{\begin{bmatrix} 1 & y_1 \\ \vdots & \vdots \\ 1 & y_m \end{bmatrix}}_G \cdot \underbrace{\begin{bmatrix} f_{1,1} & \cdots & f_{1,n} \\ f_{2,1} & \cdots & f_{2,n} \end{bmatrix}}_F. \quad (4.3)$$

Les matrices G et F contiennent respectivement la structure du modèle d'étalonnage (forçant une colonne de uns dans G), et les paramètres d'étalonnage des capteurs. Étalonner le réseau de capteurs en utilisant la factorisation (4.3) consiste alors à estimer les matrices G et F qui produisent la meilleure approximation de X , tout en respectant la contrainte de structure de G , c.-à-d. en résolvant le problème d'optimisation

$$\begin{aligned} \{G, F\} &= \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \cdot \left\| X - \tilde{G} \cdot \tilde{F} \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \mathbf{g}_1 &= \mathbb{1}_{[m \times 1]}, \end{aligned} \quad (4.4)$$

où $\| \cdot \|_{\mathcal{F}}$ est la norme de Frobenius, \mathbf{g}_1 est la première colonne de G et $\mathbb{1}_{[m \times 1]}$ est un vecteur colonne de uns.

4.2.2 Prise en compte des données manquantes

En pratique, lors de campagnes de *mobile crowdsensing*, les capteurs ne sont pas contraints dans leurs mouvements, puisqu'ils effectuent des mesures au cours des déplacements de leurs porteurs. Ils ne parcourent donc que partiellement la scène. Si un capteur j effectue une mesure dans une cellule i de \mathcal{S} , alors l'équation (4.2) s'applique, sinon la mesure $x_{i,j}$ n'est pas disponible et est donc manquante dans la matrice X , comme illustré dans la figure 4.1. Ces données manquantes dans X ne permettent pas de résoudre directement le problème (4.4). Cependant, en définissant Ω_X le domaine sur lequel X est observé et $\mathcal{P}_{\Omega_X}(\cdot)$ l'opérateur projetant X sur Ω_X , alors le problème d'optimisation pour l'étalonnage des capteurs permettant de tenir compte des données manquantes dans X s'écrit

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \cdot \left\| \mathcal{P}_{\Omega_X}(X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \mathbf{g}_1 = \mathbb{1}_{[m \times 1]}. \end{aligned} \quad (4.5)$$

En omettant la contrainte qui s'applique à la première colonne de G , le problème d'optimisation (4.5) correspond au problème de factorisation matricielle pour la complétion faible rang de matrice [81]. En pratique, ce problème peut être abordé en utilisant par exemple une technique de factorisation matricielle pondérée. L'extension pondérée de (4.5) s'écrit alors

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \mathbf{g}_1 = \mathbb{1}_{[m \times 1]}, \end{aligned} \quad (4.6)$$

où \circ est le produit de Hadamard, et $W \triangleq (w_{i,j})$ est la matrice de poids dont les éléments $w_{i,j}$ sont définis tels que

$$w_{i,j} = \begin{cases} \rho_i & \text{si la donnée } x_{i,j} \text{ est disponible,} \\ 0 & \text{sinon.} \end{cases} \quad (4.7)$$

Le poids $\rho_i > 0$ peut être choisi proportionnellement à la confiance dans le capteur i . Si tous les capteurs ont une sensibilité équivalente, tous les poids ρ_i peuvent être fixés à un, rendant alors la matrice W binaire. A ce stade, le problème d'étalonnage d'un réseau de capteurs mobiles peut être réécrit sous la forme du problème de factorisation matricielle pondérée spécifique (4.6). Dans la section suivante, nous proposons l'ajout de contraintes supplémentaires dans (4.6). Ces contraintes sont liées aux capteurs utilisés pour la surveillance de la qualité de l'air, aux propriétés des phénomènes physiques observés, ainsi qu'à la configuration particulière du réseau de capteurs considéré.

4.2.3 Contrainte de non-négativité

La matrice de données X contient uniquement les valeurs de sortie des capteurs, qui dans la majorité des cas sont des tensions. De plus, pour la surveillance de la qualité de l'air, la

scène \mathcal{S} et donc le vecteur $\mathbf{y} = [y_1, \dots, y_m]^T$ correspondent à un phénomène physique non-négatif comme par exemple une concentration, une proportion ou une température (qui peut s'exprimer en degrés Kelvin). Les matrices X et G sont par conséquent non-négatives. Enfin, la matrice F contient les paramètres des capteurs qui sont positifs pour un grand nombre de capteurs à réponses affines [74, 190]. La matrice F peut donc également être supposée non-négative. A ce stade, nous avons finalement réécrit le problème d'étalonnage de capteurs comme un problème de factorisation pondéré et informé en matrices non-négatives, c.-à-d.,

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \mathbf{g}_1 = \mathbb{1}_{[m \times 1]}. \end{aligned} \quad (4.8)$$

4.2.4 Prise en compte de données de référence

Les réseaux constitués de capteurs mobiles bas-coût sont souvent utilisés pour compléter les mesures fournies par un réseau fixe existant, moins dense, plus onéreux, mais également plus précis. Dans d'autres applications environnementales, combiner la grande quantité d'informations en provenance du réseau de capteurs bas-coût avec les quelques mesures précises produites par le réseau existant permet d'augmenter la distribution spatiale des données grâce à la mobilité du réseau bas-coût, tout en conservant la précision des mesures fixes [41]. Pour la surveillance de la qualité de l'air, le réseau de capteurs mobiles est déployé en complément du réseau de mesure officiel des AASQA constitué de quelques stations fixes, régulièrement étalonnées. Ces stations fournissent directement une estimation précise de la valeur physique qu'elles mesurent. Ces données constituent une vérité de terrain valable dans les cellules de chaque station, c.-à-d. que si une station est située dans la cellule i de \mathcal{S} , alors la valeur physique réelle de cette cellule y_i est connue. L'ensemble des mesures fournies par les stations officielles fixes peuvent alors être modélisées comme les données d'un unique capteur mobile, le $(n + 1)$ -ème capteur, ayant pour paramètres un offset nul et un gain égal à un. La prise en compte de ces données de référence se traduit dans le problème d'étalonnage par l'ajout d'une colonne à la matrice de données X qui contient alors les mesures des stations fixes, ainsi que par l'ajout d'une colonne à la matrice F qui contient les paramètres des capteurs fixes. De plus, comme les mesures fournies par les capteurs fixes correspondent directement aux valeurs du champ physique, ces données peuvent être reportées dans la deuxième colonne de G . Le problème d'optimisation à résoudre pour étalonner le réseau de capteurs mobiles en tenant compte à la fois de la non-négativité des matrices, des mesures de référence et des données manquantes dans X dues à l'échantillonnage spatiale des capteurs mobiles s'écrit

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \mathbf{g}_1 = \mathbb{1}_{[m \times 1]}, \\ g_{i,2} = x_{i,n+1} \quad \forall i \in \mathcal{I}, \\ \mathbf{f}_{n+1} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \end{aligned} \quad (4.9)$$

où \mathcal{I} est l'ensemble des cellules pour lesquelles une donnée de référence est disponible, $g_{i,2}$ est le i -ème élément de la deuxième colonne de G et \mathbf{f}_{n+1} est la $(n+1)$ -ème colonne de F .

L'équation (4.9) montre que l'étalonnage d'un réseau de capteurs mobiles peut être envisagé en utilisant une méthode de NMF particulière permettant de conserver la structure spécifique des facteurs tout en tenant compte des données manquantes dans X et des contraintes d'égalité dans G et F .

4.2.5 Conditions nécessaires à l'étalonnage

Il est important de remarquer que dans le problème de NMF pondérée

$$\{G, F\} = \arg \min_{\tilde{G}, \tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \quad (4.10)$$

la solution n'est pas unique et est déterminée à un facteur d'échelle et à une permutation près [173], c.-à-d. que si G et F sont solutions de (4.10), alors en considérant une matrice diagonale D ainsi qu'une matrice de permutation P , les matrices $\tilde{G} = G \cdot D \cdot P$ et $\tilde{F} = P^{-1} \cdot D^{-1} \cdot F$ sont également solutions de (4.10). L'ambiguïté sur D mène à un étalonnage relatif des capteurs, c.-à-d. un étalonnage permettant uniquement d'assurer une certaine cohérence des mesures des capteurs les uns par rapport aux autres, mais pas un étalonnage complet permettant de revenir aux valeurs physiques d'entrée. Une façon classique de fixer ce problème d'échelle en NMF consiste à normaliser l'une des matrices G ou F [229, 107, 32]. Cependant, pour l'étalonnage de capteurs par factorisation matricielle, il est impératif de retrouver la *vraie* matrice F , contenant les *vrais* paramètres des capteurs. Une normalisation arbitraire des matrices n'est donc pas envisageable. En réalité, l'ambiguïté d'échelle est naturellement fixée dans (4.9) grâce aux données de référence qui sont introduites dans la matrice X , sous certaines hypothèses de bon conditionnement de cette matrice.

Afin de permettre un bon déroulement de l'étalonnage, c.-à-d. afin de résoudre le problème d'optimisation (4.9), il doit y avoir suffisamment de données et de diversité dans la matrice X . Il est très facile de voir dans un cas extrême que si toutes les données contenues dans X sont identiques, ses colonnes sont constantes et l'estimation des matrices G et F n'est alors pas possible. De la même façon, si un capteur ne fournit qu'une mesure, alors la colonne de F correspondante ne pourra pas être estimée sans information complémentaire. Par conséquent, et comme pour toute méthode d'analyse de variables latentes, nous avons besoin d'une certaine diversité dans les mesures pour pouvoir effectuer la factorisation.

De façon similaire, il doit y avoir suffisamment de connections entre les capteurs mobiles à étalonner et les mesures de référence, de façon à permettre un étalonnage complet de tous les capteurs. En imaginant par exemple le cas où les capteurs peuvent être divisés en deux groupes disjoints, c.-à-d. que les capteurs du premier groupe n'ont pas de mesures en rendez-vous avec les capteurs du deuxième groupe, alors sous réserve que les conditions de diversité et de quantité des mesures soient vérifiées, le groupe ayant des mesures en rendez-vous avec des mesures de référence pourra être complètement étalonné, tandis que le groupe n'en ayant

pas ne pourra être étalonné que de façon relative.

4.3 Une première approche d'étalonnage par complétion de matrice (NMF/C-Cal)

Dans cette section, nous proposons une première méthode "naturelle" d'étalonnage basée sur le faible rang de la matrice de données X . Supposons dans un premier temps que la matrice X soit complète, c.-à-d. que tous les capteurs ont effectué une mesure dans chaque cellule de la scène \mathcal{S} . D'après l'équation (4.9), la dernière colonne de X qui contient les mesures de référence doit être égale à la deuxième colonne de G qui contient les valeurs du phénomène physique observé. Par conséquent, les matrice X et G sont connues et à partir d'elles, la matrice F peut alors être estimée par exemple par moindres carrés non-négatifs. Cependant, comme un certain nombre de données peuvent être manquantes dans X , cette approche ne peut pas être appliquée directement, et une première étape d'estimation des valeurs manquantes dans X est alors nécessaire.

Comme la matrice X est de faible rang, la restauration de ses données manquantes par l'utilisation d'une méthode de complétion de matrice de faible rang permet l'obtention d'une matrice complète X^{comp} vérifiant

$$\begin{aligned} X^{\text{comp}} &= \arg \min_{\tilde{X}} \text{rang}(\tilde{X}), \\ \text{s.c. } \mathcal{P}_{\Omega_X}(X) &= \mathcal{P}_{\Omega_X}(\tilde{X}), \end{aligned} \quad (4.11)$$

où $\mathcal{P}_{\Omega_X}(\cdot)$ est l'opérateur défini dans la section 4.2.2. Ce problème étant NP -difficile, il peut être relâché et s'écrire alors comme

$$\begin{aligned} X^{\text{comp}} &= \arg \min_{\tilde{X}} \|\tilde{X}\|_{\star}, \\ \text{s.c. } \mathcal{P}_{\Omega_X}(X) &= \mathcal{P}_{\Omega_X}(\tilde{X}), \end{aligned} \quad (4.12)$$

où $\|\cdot\|_{\star}$ est la norme nucléaire. En pratique, de nombreuses approches ont été développées pour résoudre le problème (4.12), en utilisant par exemple une SVD tronquée [22, 7]. Cependant, par construction, le rang de la matrice X est connu et est égal à 2 pour le modèle affine. La matrice X peut donc s'écrire comme le produit de deux matrices non-négatives (4.3). Il est alors possible de chercher à résoudre le problème de complétion (4.11) par une approche de complétion de matrice par factorisation en matrices non-négatives (ou NMF/C pour *Nonnegative Matrix Factorization/Completion*) [234, 230, 52], qui consiste à résoudre¹

$$\{A, S\} = \arg \min_{\tilde{A}, \tilde{S}} \|\mathcal{P}_{\Omega_X}(X - \tilde{A} \cdot \tilde{S})\|_{\mathcal{F}}, \quad (4.13)$$

1. Les matrices A et S obtenues par la complétion de matrice (4.13) ne correspondent pas nécessairement aux matrices G et F recherchées puisque rien dans le problème de complétion ne permet de fixer les indéterminations et les données connues.

en utilisant une méthode de direction alternée (ADM pour *Alternating Direction Method*) [230], des mises à jour multiplicatives [234] ou une descente de gradient de Nesterov [52]. Les deux matrices A et S obtenues permettent par produit de reconstruire une matrice X^{comp} complète vérifiant la propriété de faible rang de X et sa décomposition en produit de matrices non-négatives. La matrice X^{comp} ainsi construite est alors utilisée pour étalonner le réseau de capteurs comme proposé précédemment. L'algorithme complet d'étalonnage par complétion de matrice est finalement résumé dans l'Algorithme 1. Cette méthode sera appelée NMF/C-Cal (pour *Nonnegative Matrix Factorization/Completion for Calibration*) dans la suite de ce manuscrit.

Malgré les propriétés intéressantes de cette approche qui conserve la non-négativité des facteurs estimés et le rang de la matrice X , les performances obtenues lors des tests d'étalonnage sont liées à la réussite de l'étape de complétion de matrice qui donne des résultats très variables suivant le conditionnement de la matrice de mesure X . Les performances de cette approche seront étudiées dans le chapitre 6, consacré aux tests des méthodes d'étalonnage développées.

Algorithme 1 Étalonage par complétion de X et régression

Construire la matrice X^{comp} par une approche de complétion de matrice faible rang, par exemple à base de SVD tronquée [7] ou de NMF/C [230].

Construire G comme

$$G \triangleq [\mathbf{1}_{m \times 1}, \mathbf{x}_{n+1}^{\text{comp}}] \quad (4.14)$$

Estimer F par moindres carrés non-négatifs en résolvant

$$F = \arg \min_{\tilde{F} \geq 0} \|X^{\text{comp}} - G \cdot \tilde{F}\|_{\mathcal{F}} \quad (4.15)$$

4.4 Étalonage de capteurs par factorisation matricielle informée

Afin d'inclure l'ensemble des contraintes dans une nouvelle méthode d'étalonnage par NMF informée, nous proposons de reformuler le problème d'optimisation (4.9) en utilisant la paramétrisation proposée dans [119]. Celle-ci consiste à décomposer les matrices G et F en la somme de leur partie libre à estimer et une partie forcée qui contient les valeurs connues. Cette paramétrisation s'écrit alors

$$G = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \quad (4.16)$$

pour la matrice G , et

$$F = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F, \quad (4.17)$$

pour la matrice F , où

- Ω_G et Ω_F sont les matrices binaires informant la position des valeurs connues dans G et F ,
- $\bar{\Omega}_G$ et $\bar{\Omega}_F$ sont les matrices binaires informant la position des valeurs libres dans G et F ,
- Φ_G et Φ_F sont les matrices contenant les valeurs connues de G et F ,
- Δ_G et Δ_F sont les matrices à estimer, contenant les valeurs libres de G et F .

Les masques binaires Ω_G , Ω_F , $\bar{\Omega}_G$ et $\bar{\Omega}_F$ sont définis de façon à garantir une intersection nulle des parties libres et connues de chaque matrice, c.-à-d. $\Omega_G \circ \bar{\Omega}_G = 0$ et $\Omega_F \circ \bar{\Omega}_F = 0$ [119].

En utilisant les paramétrisations (4.16) et (4.17), le problème d'optimisation (4.9) s'écrit alors

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{G} = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \\ \tilde{F} = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (4.18)$$

Le problème (4.18) étant non-convexe par rapport aux deux matrices G et F , nous proposons de l'aborder par une stratégie de mises à jour alternées en considérant les sous-problèmes en G :

$$\begin{aligned} G = \arg \min_{\tilde{G} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot F) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{G} = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \end{aligned} \quad (4.19)$$

et en F :

$$\begin{aligned} F = \arg \min_{\tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{F} = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (4.20)$$

Ces deux sous-problèmes sont ensuite résolus de façon itérative et alternée. Pour la mise à jour de F , les auteurs de [119] ont proposé la formule multiplicative suivante

$$F \leftarrow \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F \circ \left(\frac{G^T \cdot (W \circ [X - G \cdot \Omega_F \circ \Phi_F]^+)}{G^T \cdot (W \circ (G \cdot \bar{\Omega}_F \Delta_F))} \right), \quad (4.21)$$

obtenue en suivant une stratégie MM où, pour une matrice A quelconque, $[A]^+$ seuil les éléments négatifs de A à un réel ϵ donné. Il est intéressant de remarquer que si la matrice F n'a pas d'éléments connus, Ω_F est donc une matrice nulle, $\bar{\Omega}_F = \mathbb{1}_{2 \times (n+1)}$, $F = \Delta_F$, et la formule de mise à jour (4.21) est alors identique à la version pondérée [79] de la formule de mise à jour multiplicative de Lee et Seung [113]. Pour le sous-problème (4.19), nous proposons

d'adapter la mise à jour de F à la matrice G , c.-à-d.

$$G \leftarrow \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G \circ \left(\frac{(W \circ [X - \Phi_G \cdot F]^+) \cdot F^T}{(W \circ (\Delta_G \cdot F)) \cdot F^T} \right). \quad (4.22)$$

Comme évoqué précédemment, il est important de garder en tête qu'en factorisation matricielle, une ambiguïté d'échelle entre G et F est généralement observée. Ce problème est habituellement résolu en normalisant les lignes de F ou les colonnes de G . Cependant, pour l'étalonnage de capteurs, nous cherchons à estimer la vraie matrice F et nous ne pouvons donc pas normaliser de façon arbitraire les matrices estimées. En réalité, dans la méthode d'étalonnage proposée ci-dessus, le problème d'ambiguïté d'échelle est résolue grâce aux capteurs de référence qui contraignent la dernière colonne de F et permettent donc de fixer l'échelle du problème, tant qu'il y a suffisamment de mesures dans la matrice X pour lier chaque capteur du réseau avec au minimum deux mesures de référence (pour le modèle affine) distinctes, via des rendez-vous directs ou un chemin entre les différents capteurs. Si la matrice X ne contient pas suffisamment de mesures de références et/ou pas suffisamment de rendez-vous entre les capteurs mobiles à étalonner et les capteurs de références, alors la méthode d'étalonnage par NMF informée présentée ci-dessus que nous appellerons IN-Cal (pour *Informed Nmf-based Calibration*) effectue un étalonnage relatif des capteurs, c.-à-d. que les mesures des capteurs sont harmonisées de façon à être cohérentes d'un capteur à un autre, mais l'ambiguïté d'échelle reste non-résolue. L'algorithme de cette méthode d'étalonnage par NMF informée (IN-Cal) est présenté dans l'Algorithme 2.

Algorithme 2 Etalonnage par Factorisation Informée en Matrices Non-négatives (IN-Cal)

Initialisation de G et F

pour k de 1 à itération max **faire**

Mettre à jour G selon (4.22)

Mettre à jour F selon (4.21)

fin pour

4.5 Régularisation par utilisation de données constructeur

Les fabricants des capteurs fournissent très souvent dans la documentation de leurs produits des paramètres moyens d'étalonnage. Ces informations ne sont pas prises en compte dans les méthodes d'étalonnage précédentes NMF/C-Cal et IN-Cal, alors que leur utilisation permettrait de régulariser l'étalonnage des capteurs dans les cas où la matrice X ne serait pas bien conditionnée pour permettre un étalonnage efficace des capteurs par les méthodes précédentes, c.-à-d. s'il n'y a pas suffisamment de rendez-vous ou de mesures de référence dans X pour lever le problème d'ambiguïté d'échelle du problème de NMF. Par ailleurs, cette information est utilisée par plusieurs méthodes d'étalonnage présentes dans la littérature,

notamment [11, 218, 111].

En considérant des capteurs à réponses affines, alors les paramètres théoriques fournis par le fabricant sont une valeur moyenne \bar{f}_1 pour l'offset et une valeur moyenne \bar{f}_2 pour le gain. En pratique, la moyenne des offsets de tous les capteurs du réseau et la moyenne des gains de tous les capteurs doivent être proches des moyennes théoriques, ce qui s'écrit

$$\frac{1}{n} \cdot \sum_{j=1}^n f_{1,j} \approx \bar{f}_1 \quad \text{et} \quad \frac{1}{n} \cdot \sum_{j=1}^n f_{2,j} \approx \bar{f}_2. \quad (4.23)$$

Il est important de noter que la normalisation des lignes de F ou des colonnes de G est très classique en factorisation matricielle et est habituellement effectuée après mise à jour de la matrice [229], ou directement dans la mise à jour [107]. La normalisation de la matrice F dans le cas de factorisations matricielles informées a également été proposée dans [122, 32]. Cependant, ces normalisations forcent de façon brutale les normes des lignes de F à être rigoureusement égales aux valeurs choisies, ce qui peut entraîner une erreur dans la cas de l'étalonnage de capteurs où la moyenne fixée par le constructeur n'est pas rigoureusement égale à la "vraie" moyenne des paramètres des capteurs. Ce type d'erreur est une des principales limites des approches d'étalonnage n'utilisant pas de capteurs de référence pour fournir une vérité de terrain, comme par exemple [11]. En considérant la contrainte de norme comme une pénalisation du problème d'optimisation, alors l'impact de la pénalisation sur le problème peut être ajusté en faisant varier le poids associé à la régularisation.

En définissant $\bar{\mathbf{f}} \triangleq [\bar{f}_1, \bar{f}_2]$, alors les approximations (4.23) s'écrivent

$$\frac{1}{n} \cdot F \cdot \begin{bmatrix} \mathbb{1}_{n \times 1} \\ 0 \end{bmatrix} \approx \bar{\mathbf{f}}, \quad (4.24)$$

où $\begin{bmatrix} \mathbb{1}_{n \times 1} \\ 0 \end{bmatrix}$ est un vecteur colonnes de taille $[(n+1) \times 1]$ constitué d'une colonne de uns concaténée avec un zéro afin de ne pas tenir compte de la dernière colonne de F qui correspond aux capteurs de référence, dans le calcul de la moyenne des paramètres. En utilisant la paramétrisation (4.17) et en remarquant que la dernière colonne de $\bar{\Omega}_F$ est nulle, alors l'approximation (4.24) s'écrit finalement

$$\frac{1}{n} \cdot (\bar{\Omega}_F \circ F) \cdot \mathbb{1}_{(n+1) \times 1} \approx \bar{\mathbf{f}}. \quad (4.25)$$

Nous proposons de déduire de l'expression (4.25) un terme de pénalisation qui une fois intro-

duit dans le problème (4.18) mène au problème d'optimisation

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F} \geq 0} & \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2 \\ & + \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot (\bar{\Omega}_F \circ \tilde{F}) \cdot \mathbf{1}_{(n+1) \times 1} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2, \\ \text{s.c. } & \tilde{G} = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \\ & \tilde{F} = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (4.26)$$

où ν est le poids permettant de gérer la force du terme de pénalisation. Comme pour la méthode IN-Cal, le problème (4.26) est non-convexe par rapport à G et F , et peut donc être séparé en deux sous-problèmes. Il est immédiat de remarquer que le terme de pénalisation ajouté au problème n'intervient que dans le sous-problème en F . Le sous-problème en G est donc identique à celui de la méthode IN-Cal proposée précédemment et la formule de mise à jour associée (4.22) est donc toujours applicable. Le sous-problème en F quant à lui s'écrit

$$\begin{aligned} F = \arg \min_{\tilde{F} \geq 0} & \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2 \\ & + \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot (\bar{\Omega}_F \circ \tilde{F}) \cdot \mathbf{1}_{(n+1) \times 1} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2, \\ \text{s.c. } & \tilde{F} = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (4.27)$$

Pour le calcul de la formule de mise à jour associée au problème (4.27), nous proposons d'utiliser l'approche heuristique décrite dans [112]. Ainsi, nous commençons par définir \mathcal{J} la fonction de coût associée à (4.27) comme

$$\mathcal{J} = \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2 + \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot (\bar{\Omega}_F \circ \tilde{F}) \cdot \mathbf{1}_{(n+1) \times 1} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2. \quad (4.28)$$

Nous réécrivons alors \mathcal{J} comme la somme $\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2$, où

$$\mathcal{J}_1 = \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2 \quad (4.29)$$

est la fonction de coût du problème de NMF informée présenté dans (4.20) et

$$\mathcal{J}_2 = \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot (\bar{\Omega}_F \circ \tilde{F}) \cdot \mathbf{1}_{(n+1) \times 1} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2 \quad (4.30)$$

est la partie de la fonction de coût due à la pénalisation.

L'introduction de la paramétrisation (4.17) dans la fonction de coût \mathcal{J}_1 mène alors à l'expression

$$\mathcal{J}_1 = \frac{1}{2} \cdot \left\| W \circ (X - G \cdot (\Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F)) \right\|_{\mathcal{F}}^2, \quad (4.31)$$

$$= \frac{1}{2} \cdot \left\| W \circ (\tilde{X} - G \cdot (\bar{\Omega}_F \circ \Delta_F)) \right\|_{\mathcal{F}}^2, \quad (4.32)$$

où $\tilde{X} = [X - G \cdot (\Omega_F \circ \Phi_F)]^+$ avec $[\cdot]^+$ l'opérateur de projection sur \mathbb{R}^+ qui permet d'assurer la non-négativité de la matrice \tilde{X} . Par développement de la norme de Frobenius, l'équation (4.32) peut s'écrire

$$\mathcal{J}_1 = \frac{1}{2} \cdot \sum_{j=1}^m \sum_{i=1}^n \left[(W \circ (\tilde{X} - G \cdot (\bar{\Omega}_F \circ \Delta_F)))_{i,j} \right]^2, \quad (4.33)$$

$$= \frac{1}{2} \cdot \sum_{j=1}^m \sum_{i=1}^n w_{i,j}^2 \cdot \left[\tilde{x}_{i,j} - (G \cdot (\bar{\Omega}_F \circ \Delta_F))_{i,j} \right]^2, \quad (4.34)$$

$$= \frac{1}{2} \cdot \sum_{j=1}^m \sum_{i=1}^n w_{i,j}^2 \cdot \left[\tilde{x}_{i,j} - \sum_{k=1}^{N+1} g_{j,k} \cdot \bar{\Omega}_{F_{k,i}} \cdot \Delta_{F_{k,i}} \right]^2. \quad (4.35)$$

A partir de (4.35), il est facile de dériver la fonction de coût \mathcal{J}_1 par rapport à un élément $\Delta_{F_{k,j}}$ de Δ_F , ce qui donne

$$\frac{\partial \mathcal{J}_1}{\partial \Delta_{F_{k,i}}} = \sum_{j=1}^m w_{i,j}^2 \cdot \left[\tilde{x}_{i,j} - (G \cdot (\bar{\Omega}_F \circ \Delta_F))_{i,j} \right] \cdot (-g_{j,k} \cdot \bar{\Omega}_{F_{k,i}}), \quad (4.36)$$

$$= -\bar{\Omega}_{F_{k,i}} \cdot \sum_{j=1}^m g_{j,k} \cdot (W^2 \circ \tilde{X})_{i,j} - g_{j,k} \cdot (W^2 \circ (G \cdot (\bar{\Omega}_F \circ \Delta_F)))_{i,j}, \quad (4.37)$$

$$= -\bar{\Omega}_{F_{k,i}} \cdot \left[G^T \cdot (W^2 \circ \tilde{X}) - G^T \cdot (W^2 \circ (G \cdot (\bar{\Omega}_F \circ \Delta_F))) \right]_{k,i}. \quad (4.38)$$

La dérivée (4.38) se généralise alors trivialement pour l'ensemble de la matrice Δ_F , ce qui donne

$$\frac{\partial \mathcal{J}_1}{\partial \Delta_F} = \bar{\Omega}_F \circ \left(G^T \cdot (W^2 \circ (G \cdot (\bar{\Omega}_F \circ \Delta_F))) - G^T \cdot (W^2 \circ [X - G \cdot (\Omega_F \circ \Phi_F)]^+) \right). \quad (4.39)$$

De la même façon, l'introduction de la paramétrisation (4.17) dans la fonction de coût \mathcal{J}_2 mène alors à l'expression

$$\mathcal{J}_2 = \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot \left(\bar{\Omega}_F \circ (\Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F) \right) \cdot \mathbb{1}_{[(n+1) \times 1]} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2. \quad (4.40)$$

En remarquant que $\bar{\Omega}_F \circ \Omega_F \circ \Phi_F = 0$, alors nous obtenons

$$\begin{aligned} \mathcal{J}_2 &= \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot \left(\bar{\Omega}_F \circ \Delta_F \right) \cdot \mathbb{1}_{[(n+1) \times 1]} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2, \\ &= \frac{\nu}{2} \cdot \sum_{j=1}^2 \left[\frac{1}{n} \cdot \sum_{i=1}^{n+1} \left(\bar{\Omega}_F \circ \Delta_F \right)_{i,j} - \bar{f}_j \right]^2. \end{aligned} \quad (4.41)$$

La dérivée de la fonction de coût \mathcal{J}_2 par rapport à l'élément $\Delta_{F_{i,j}}$ de Δ_F s'écrit alors

$$\frac{\partial \mathcal{J}_2}{\partial \Delta_{F_{i,j}}} = \nu \cdot \left[\frac{1}{n} \cdot \left(\sum_{k=1}^{n+1} \bar{\Omega}_{F_{i,k}} \cdot \Delta_{F_{j,k}} \right) - \bar{f}_j \right] \cdot \frac{1}{n} \cdot \bar{\Omega}_{F_{i,j}}. \quad (4.42)$$

L'extension matricielle de cette dérivée s'écrit alors

$$\frac{\partial J_2}{\partial \Delta_F} = \frac{\nu}{n^2} \cdot \text{diag} \left(\left(\bar{\Omega}_F \circ \Delta_F \right) \cdot \mathbb{1}_{[(n+1) \times 1]} \right) \cdot \bar{\Omega}_F - \frac{\nu}{n} \cdot \text{diag}(\bar{\mathbf{f}}) \cdot \bar{\Omega}_F. \quad (4.43)$$

En notant finalement \mathcal{J}_1^+ et \mathcal{J}_2^+ les parties positives de (4.39) et (4.43), et \mathcal{J}_1^- et \mathcal{J}_2^- leurs parties négatives, la dérivée de la fonction de coût globale \mathcal{J} par rapport à la matrice Δ_F s'écrit

$$\frac{\partial \mathcal{J}}{\partial \Delta_F} = (\mathcal{J}_1^+ + \mathcal{J}_2^+) - (\mathcal{J}_1^- + \mathcal{J}_2^-), \quad (4.44)$$

et la formule de mise à jour multiplicative de Δ_F obtenue par la méthode heuristique s'écrit

$$\Delta_F \leftarrow \Delta_F \circ \frac{\mathcal{J}_1^- + \mathcal{J}_2^-}{\mathcal{J}_1^+ + \mathcal{J}_2^+}, \quad (4.45)$$

c.-à-d.

$$\Delta_F \leftarrow \Delta_F \circ \frac{\bar{\Omega}_F \circ (G^T(W^2 \circ [X - G \cdot (\Omega_F \circ \Phi_F)]^+ + \frac{\nu}{n} \text{diag}(\bar{\mathbf{f}}) \cdot \bar{\Omega}_F)}{\bar{\Omega}_F \circ (G^T(W^2 \circ (G(\bar{\Omega}_F \circ \Delta_F))) + \frac{\nu}{n^2} \text{diag}((\bar{\Omega}_G \circ \Delta_F) \cdot \mathbb{1}_{[(n+1) \times 1]}) \cdot \bar{\Omega}_F)}. \quad (4.46)$$

La mise à jour de la matrice F complète s'écrit finalement

$$F \leftarrow \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \quad (4.47)$$

La mise à jour (4.46) permet de rajouter à la méthode IN-Cal une régularisation douce de la valeur moyenne des lignes de F , d'où le nom que nous donnons à cette méthode dans la suite de ce manuscrit, ACIN-Cal (pour *Average Constrained IN-Cal*). L'algorithme de cette méthode est présenté dans l'Algorithme 3

Algorithme 3 Étalonnage par Factorisation Informée en Matrices Non-négatives sous contrainte douce de normalisation (ACIN-Cal)

Initialisation de G et F .

pour k de 1 à itération max **faire**

Mise à jour de G selon (4.22).

Mise à jour de F selon (4.46) et (4.47).

fin pour

4.6 Initialisation des méthodes

Chacune des approches d'étalonnage qui ont été présentées précédemment sont itératives et nécessitent une initialisation des matrices G et F . Cette étape d'initialisation est un pro-

blème bien connu en NMF et différentes stratégies sont envisageables. Une approche classique consiste à initialiser les matrices aléatoirement, tandis que certains auteurs ont proposé une initialisation par des experts [119], la sortie d'une autre approche de factorisation [8] ou encore un modèle physique [162].

Dans les premiers tests réalisés dans le cadre de cette thèse, qui ont fait l'objet de la publication [50], nous avons choisi d'initialiser les matrices par la sortie d'une première version de l'approche "naturelle" présentée dans la section 4.3. Nous avons d'abord utilisé l'approche de complétion faible rang proposée dans [7]. Malheureusement, cette initialisation était très coûteuse en temps de calcul et nous avons finalement remplacé l'approche de complétion par SVD tronquée [7] par l'approche NMF/C [230]. Cependant, comme toute méthode de NMF, la NMF/C nécessite aussi d'être initialisée. Nous effectuons finalement cette étape par tirages aléatoires autour des paramètres théoriques du constructeur des capteurs pour la matrice F , tandis que pour G , la première colonne est initialisée à 1 – pour satisfaire au modèle affine – tandis que sa deuxième colonne est initialisée aléatoirement. Les éventuelles entrées négatives des matrices G et F initialisées sont seuillées à une constante ϵ faible et non-négative pour assurer la non-négativité de toutes les matrices. L'approche "naturelle" d'étalonnage a alors montré des performances légèrement meilleures qu'avec la SVD tronquée dans la phase de complétion, mais en deçà des performances atteintes par les autres approches proposées dans ce chapitre.

Par ailleurs, dans d'autres tests complémentaires, nous avons constaté qu'initialiser nos approches *(i)* avec la sortie de la méthode naturelle (utilisant la SVD tronquée ou la NMF/C ci-dessus) ou *(ii)* de manière aléatoire comme pour la NMF/C apportait *in fine* les mêmes performances d'étalonnage. Ce sont donc des initialisations aléatoires d'IN-Cal et d'ACIN-Cal que nous utiliserons dans tous les tests de cette thèse.

4.7 Conclusion

Dans ce chapitre, nous avons proposé de reformuler le problème d'étalonnage d'un réseau de capteurs mobiles à réponses affines sous la forme de différents problèmes d'optimisation.

En exploitant des hypothèses et définitions couramment utilisées dans la littérature pour l'étalonnage de réseaux de capteurs, nous avons proposé de regrouper les mesures collectées par l'ensemble du réseau en une matrice de données X factorisable en un produit de deux matrices G et F , la première contenant la structure du modèle d'étalonnage, la seconde les valeurs des paramètres de chaque capteur. L'échantillonnage spatial irrégulier du à la mobilité des capteurs se traduit sur X par l'absence de certaines données. Nous avons alors proposé de gérer l'absence de certaines données par une pondération du problème de factorisation.

Le problème d'étalonnage proposé a alors pris la forme d'une factorisation pondérée en matrices non-négatives dans laquelle des données de capteurs de référence étalonnés sont mêlées à celles des capteurs mobiles afin de pallier au problème d'ambiguïté d'échelle rencontré classiquement en factorisation matricielle. Ces données de référence sont incluses dans

les problèmes d'optimisation sous la forme de contraintes égalité dans les matrices facteur recherchées.

Deux méthodes de résolution ont été développées. Nous avons dans un premier temps proposé une approche d'étalonnage naturelle à base de complétion de matrice faible rang. Dans un second temps, nous avons développé une méthode d'étalonnage utilisant une factorisation matricielle informée dans laquelle les matrices sont décomposées en la somme de parties libres à estimer et de parties forcées permettant de respecter les contraintes égalité liées aux modèles et aux données de référence.

Dans le cas où la matrice de données X serait mal conditionnée et ne permettrait donc pas une estimation correcte de G et F , nous proposons de régulariser les problèmes de factorisation matricielle en exploitant les paramètres théoriques fournis par le fabriquant des capteurs. Ces informations se traduisent dans les différents problèmes d'optimisation par l'ajout d'un terme de pénalisation influençant la mise à jour de la matrice des paramètres F .

Dans le chapitre suivant, nous proposons d'étendre les méthodes proposées ici pour des capteurs à réponses affines à des capteurs à réponse non-linéaire.

Méthodes de Semi-NMF informées pour l'étalonnage d'un réseau de capteurs mobiles non-linéaires

Dans ce chapitre, nous proposons de généraliser les approches d'étalonnage développées précédemment au cas de l'étalonnage d'un réseau de capteurs à réponses polynômiales de degrés et de signes quelconques. La forme polynômiale des réponses des capteurs impose alors des structures particulières aux matrices recherchées. La matrice G contenant les valeurs du phénomène physique observé prend alors la forme d'une matrice de Vandermonde. Les méthodes développées ici ont fait l'objet de la publication [54].

5.1 Hypothèses et introduction de structures dans les matrices facteurs

Les méthodes d'étalonnage qui sont proposées dans ce chapitre reposent sur les mêmes définitions que précédemment. En particulier, n capteurs mobiles sont supposés effectuer des mesures sur une scène \mathcal{S} , à partir desquelles se construit une matrice de données X comme illustré dans la figure 4.1.

Nous nous intéressons maintenant à l'étalonnage de capteurs à réponses polynômiales. La fonction d'étalonnage \mathcal{F}_j associée au j -ème capteur du réseau est alors un polynôme de degré N fixé, c.-à-d. que la mesure $x_{i,j}$ fournie par le j -ème capteur du réseau lorsqu'il est exposé à la valeur physique y_i de la cellule i s'écrit comme

$$x_{i,j} = \mathcal{F}_j(y_i) = f_{1,j} + f_{2,j} \cdot y + \cdots + f_{N+1,j} \cdot y^N, \quad (5.1)$$

où les termes $f_{k,j}$ sont les paramètres d'étalonnage associés au capteur j . Dans le cas où les capteurs suivent une réponse polynômiale de degré 1, alors nous revenons au modèle de réponse affine étudié précédemment.

En supposant que la scène \mathcal{S} soit composée de m cellules qui ont toutes été visitées par

l'ensemble des capteurs, alors la matrice de données X peut s'écrire sous la forme

$$\underbrace{\begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix}}_X = \underbrace{\begin{bmatrix} 1 & y_1 & \cdots & y_1^N \\ \vdots & \vdots & & \vdots \\ 1 & y_m & \cdots & y_m^N \end{bmatrix}}_G \cdot \underbrace{\begin{bmatrix} f_{1,1} & \cdots & f_{1,n} \\ \vdots & & \vdots \\ f_{N+1,1} & \cdots & f_{N+1,n} \end{bmatrix}}_F. \quad (5.2)$$

La matrice G a une structure de Vandermonde, notée $G \triangleq \text{VDM}(\mathbf{y}, N)$, dont les colonnes sont toutes des expansions monômiales du phénomène physique observé \mathbf{y} , alors que F contient les paramètres des capteurs. Étalonner le réseau consiste alors à estimer G et F telles que

$$\begin{aligned} \{G, F\} &= \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \left\| X - \tilde{G} \cdot \tilde{F} \right\|_{\mathcal{F}}^2 \\ \text{s.c. } &\tilde{G} = \text{VDM}(\mathbf{y}, N) \geq 0, \end{aligned} \quad (5.3)$$

En pratique, les capteurs étant toujours non-contraints dans leurs déplacements, des données sont manquantes dans la matrice X et peuvent être prises en compte dans le problème (5.3) au moyen de la matrice de pondération W définie dans (4.7) qui informe de la position des données manquantes dans X . Le problème d'optimisation à résoudre s'écrit alors

$$\begin{aligned} \{G, F\} &= \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2 \\ \text{s.c. } &\tilde{G} = \text{VDM}(\mathbf{y}, N) \geq 0, \end{aligned} \quad (5.4)$$

De plus, les matrices X et G sont supposées non-négatives puisque contenant respectivement, les valeurs de sortie des capteurs, et le phénomène physique observé. Nous considérons comme précédemment que les mesures fournies par des capteurs de référence fixes et étalonnés peuvent être vues comme les mesure d'un unique $(n + 1)$ -ème capteur mobile dont les paramètres d'étalonnage sont

$$f_{2,n+1} = 1 \quad \text{et} \quad f_{k,n+1} = 0 \quad \forall k \neq 2. \quad (5.5)$$

En effet, avec ces paramètres d'étalonnage, si une station de référence fournit une mesure dans la cellule i de la scène \mathcal{S} , alors $x_{i,n+1}$ est disponible et est égal à la valeur physique y_i présente dans la deuxième colonne de G . Grâce à la structure de Vandermonde de G , une mesure de référence informe donc une ligne complète de G . Les capteurs de référence apportent donc des contraintes sur les deux matrices facteurs recherchées qui sont pour G des lignes connues, et pour F une colonne (la $(n + 1)$ -ème) qui est connue. Nous proposons d'utiliser ici encore la paramétrisation (4.16) et (4.17) développée par Limem *et al.* [119] permettant de séparer chaque matrice en la somme de sa partie libre à estimer et de sa partie connue.

Étalonner le réseau de capteurs consiste donc à retrouver les matrices G et F respectant les contraintes égalité imposées par la paramétrisation (4.16) et (4.17), ainsi que la structure

de Vandermonde de la matrice G , c.-à-d. qu'il faut résoudre le problème d'optimisation

$$\begin{aligned} \{G, F\} = & \arg \min_{\tilde{G}, \tilde{F}} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } & \tilde{G} = \text{VDM}(\mathbf{y}, N) \geq 0, \\ & \tilde{G} \triangleq \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \\ & \tilde{F} \triangleq \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (5.6)$$

Il est important de remarquer que contrairement à l'étalonnage des capteurs à réponses affines, nous n'avons pas posé ici de contrainte de non-négativité sur la matrice F . En effet, si une grande partie des capteurs à réponse affine utilisés pour la surveillance de la qualité de l'air ont des paramètres non-négatifs, ce n'est pas le cas de la plupart des capteurs à réponses non-linéaires. La méthode d'étalonnage proposée dans ce cas doit donc permettre de gérer des paramètres de signes quelconques.

5.2 Étalonnage par complétion

Comme pour les capteurs à réponses affines, il est évident que si la matrice X est complète, alors la matrice G est également connue car entièrement définie à partir de la dernière colonne de X . La matrice F peut alors être estimée par moindres carrés en résolvant

$$F = \arg \min_{\tilde{F}} \frac{1}{2} \cdot \left\| X^{\text{comp}} - G \cdot \tilde{F} \right\|_{\mathcal{F}}^2. \quad (5.7)$$

Cette approche est très similaire à la méthode d'étalonnage par complétion proposée pour les capteurs à réponses affines, à ceci près que dans le cas présent, une fois la matrice X complétée, sa dernière colonne doit être propagée dans G de façon à respecter la structure de Vandermonde. De plus, la matrice F n'étant plus sujette à aucune contrainte de signe, elle peut être estimée par la méthode classique des moindres carrés. L'algorithme de cette méthode est présenté dans l'Algorithme 4.

5.3 Étalonnage par factorisation matricielle informée

Le problème d'optimisation (5.6) étant non-convexe par rapport aux deux matrices G et F , nous le divisons en deux sous-problèmes qui seront résolus de façon itérative et alternée. Le sous-problème par rapport à G s'écrit

$$\begin{aligned} G = & \arg \min_{\tilde{G} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot F) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } & \tilde{G} = \text{VDM}(\mathbf{y}, N), \\ & \tilde{G} = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \end{aligned} \quad (5.10)$$

Algorithme 4 Étalonage par complétion de X et régression.

Construire la matrice X^{comp} par une approche de complétion de matrice faible rang, par exemple à base de SVD tronquée [7].

Construire G comme

$$G \triangleq \left[\mathbb{1}_{m \times 1}, \mathbf{x}_{n+1}^{\text{comp}}, \dots, (\mathbf{x}_{n+1}^{\text{comp}})^N \right] \quad (5.8)$$

Estimer F par moindres carrés en résolvant

$$F = \arg \min_{\tilde{F} \geq 0} \left\| X^{\text{comp}} - G \cdot \tilde{F} \right\|_{\mathcal{F}} \quad (5.9)$$

tandis que celui par rapport à F s'écrit

$$\begin{aligned} F = \arg \min_{\tilde{F}} \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \tilde{F} \triangleq \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (5.11)$$

5.3.1 Mise à jour de la matrice de Vandermonde

Grâce à la paramétrisation (4.16) et (4.17), il est facile de voir que seule la partie libre de G doit être mise à jour dans le problème d'optimisation (5.10) et la fonction de coût associée s'écrit

$$\begin{aligned} \mathcal{J} &= \frac{1}{2} \cdot \left\| W \circ (\tilde{X} - (\bar{\Omega}_G \circ \Delta_G) \cdot F) \right\|_{\mathcal{F}}^2, \\ &= \frac{1}{2} \cdot \sum_{j=1}^m \sum_{i=1}^n (W \circ (\tilde{X} - (\bar{\Omega}_G \circ \Delta_G) \cdot F))_{i,j}^2, \\ &= \frac{1}{2} \cdot \sum_{j=1}^m \sum_{i=1}^n w_{i,j}^2 \cdot (\tilde{x}_{i,j} - ((\bar{\Omega}_G \circ \Delta_G) \cdot F)_{i,j})^2, \end{aligned} \quad (5.12)$$

où $\tilde{x}_{i,j}$ est l'élément de la j -ème ligne et i -ème colonne de la matrice $\tilde{X} \triangleq X - (\Omega_G \circ \Phi_G) \cdot F$. Comme expliqué précédemment, lorsqu'une donnée de \mathbf{y} est connue, alors toute la ligne correspondante dans G est connue, et inversement, lorsqu'une donnée de \mathbf{y} est inconnue, alors toute la ligne correspondante est à estimer. La matrice Δ_G a donc également une structure de Vandermonde, ce qui permet de reformuler la fonction de coût \mathcal{J} par rapport à la deuxième

colonne de Δ_G que nous notons $\Delta_{\mathbf{y}}$. Cette nouvelle expression de \mathcal{J} s'écrit

$$\mathcal{J} = \frac{1}{2} \cdot \sum_{j=1}^m \sum_{i=1}^n w_{i,j}^2 \cdot \left(\tilde{x}_{i,j} - \sum_{k=1}^{N+1} \bar{\Omega}_{G_{i,k}} \cdot \Delta_{\mathbf{y}}^{k-1} \cdot f_{k,j} \right)^2. \quad (5.13)$$

A partir de cette dernière expression, nous proposons une stratégie de mise à jour en deux étapes pour la matrice Δ_G qui consiste en une première mise à jour de $\Delta_{\mathbf{y}}$ par une itération de descente de gradient projetée, suivie par une étape de propagation de $\Delta_{\mathbf{y}}$ dans Δ_G de façon à conserver la structure de Vandermonde de Δ_G . Cette stratégie en deux temps de mise à jour d'une partie des données puis de leur propagation dans l'ensemble de la matrice a été utilisée dans [140, 66] dans un contexte de factorisation en matrices linéaires-quadratiques.

En dérivant (5.13) par rapport à un élément $\Delta_{\mathbf{y}_i}$ de $\Delta_{\mathbf{y}}$, nous obtenons

$$\frac{\partial \mathcal{J}}{\partial \Delta_{\mathbf{y}_i}} = \sum_{i=1}^n w_{i,j}^2 \cdot (\tilde{x}_{i,j} - ((\bar{\Omega}_G \circ \Delta_G) \cdot F)_{i,j}) \cdot \sum_{k=2}^{N+1} (k-1) \cdot \bar{\Omega}_{G_{i,k}} \cdot \Delta_{\mathbf{y}}^{k-2} \cdot f_{k,j}. \quad (5.14)$$

$$= \sum_{i=1}^n (W^2 \circ (\tilde{X} - (\bar{\Omega}_G \circ \Delta_G) \cdot F))_{i,j} \cdot \sum_{k=1}^N k \cdot \bar{\Omega}_{G_{i,k+1}} \cdot \Delta_{G_{i,k}} \cdot f_{k+1,j}, \quad (5.15)$$

$$= \sum_{i=1}^n (W^2 \circ (\tilde{X} - (\bar{\Omega}_G \circ \Delta_G) \cdot F))_{i,j} \cdot (D \cdot (\bar{\Omega}_G^S \circ \Delta_G^S) \cdot F^S)_{i,j}, \quad (5.16)$$

$$= \sum_{i=1}^n (W^2 \circ (\tilde{X} - (\bar{\Omega}_G \circ \Delta_G) \cdot F))_{i,j} \cdot ((F^S)^T \cdot (\bar{\Omega}_G^S \circ \Delta_G^S)^T \cdot D^T)_{i,j}. \quad (5.17)$$

où $\bar{\Omega}_G^S$ est la matrice composée des $N-1$ dernières colonnes de $\bar{\Omega}_G$, Δ_G^S est le matrice composée des $N-1$ premières colonnes de Δ_G , F^S est la matrice composée des $N-1$ dernières lignes de F , et D est la matrice diagonale dont les élément sont $d_{i,j} \triangleq i$. A partir de l'expression (5.17) il est facile de généraliser l'expression du gradient de \mathcal{J} par rapport à $\Delta_{\mathbf{y}}$ et de trouver

$$\frac{\partial \mathcal{J}}{\partial \Delta_{\mathbf{y}}} = \text{diag} \left[(W^2 \circ (\tilde{X} - (\bar{\Omega}_G \circ \Delta_G) \cdot F)) \cdot (F^S)^T \cdot (\bar{\Omega}_G^S \circ \Delta_G^S)^T \cdot D^T \right]. \quad (5.18)$$

Nous proposons alors de mettre à jour le vecteur colonne $\Delta_{\mathbf{y}}$ par une simple descente de gradient projeté de Lin [125]. La mise à jour s'écrit donc

$$\Delta_{\mathbf{y}} \leftarrow \left[\Delta_{\mathbf{y}} - \rho_G \cdot \frac{\partial \mathcal{J}}{\partial \Delta_{\mathbf{y}}} \right]^+, \quad (5.19)$$

où ρ_G est le pas de descente fixé, et $[\cdot]^+$ est l'opérateur de projection sur \mathbb{R}^+ . Une fois le vecteur $\Delta_{\mathbf{y}}$ mis à jour, nous pouvons alors reformer le vecteur colonne \mathbf{y} complet en réunissant sa partie libre $\Delta_{\mathbf{y}}$ mise à jour et sa partie connue $\Phi_{\mathbf{y}}$

$$\mathbf{y} \leftarrow \Phi_{\mathbf{y}} + \Delta_{\mathbf{y}}. \quad (5.20)$$

La matrice G complète peut être mise à jour à son tour en propageant \mathbf{y} de façon à conserver

la structure de Vandermonde, ce qui s'écrit alors pour chaque colonne \mathbf{g}_i de G comme

$$\mathbf{g}_i \leftarrow \mathbf{y}^{i-1}, \quad \forall i \in \llbracket 1 \cdots N + 1 \rrbracket. \quad (5.21)$$

5.3.2 Mise à jour du deuxième facteur

La mise en place de la mise à jour de la matrice F est immédiate, en remarquant que la contrainte de structure de Vandermonde de G n'intervient pas dans le sous-problème (5.11). La fonction de coût \mathcal{J} associée au problème s'écrit alors

$$\begin{aligned} \mathcal{J} &= \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{F} &= \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F, \end{aligned} \quad (5.22)$$

et correspond à la fonction de coût du problème de factorisation matricielle informée présenté dans l'équation (4.32). Sa dérivée s'écrit donc comme (4.39), à ceci près que la matrice F n'est pas sujette ici aux contraintes de non-négativité appliquées dans (4.32), et que donc l'opérateur de projection $[\cdot]^+$ utilisé dans (4.39) n'a pas à être appliqué. Nous obtenons finalement que

$$\frac{\partial \mathcal{J}}{\partial \Delta_F} = \bar{\Omega}_F \circ \left(G^T \cdot (W^2 \circ (G \cdot (\bar{\Omega}_F \circ \Delta_F))) - G^T \cdot (W^2 \circ (X - G \cdot (\Omega_F \circ \Phi_F))) \right). \quad (5.23)$$

La mise à jour par descente de gradient est ensuite immédiate et s'écrit

$$\Delta_F \leftarrow \Delta_F - \rho_F \cdot \frac{\partial \mathcal{J}}{\partial \Delta_F}, \quad (5.24)$$

où ρ_F est le pas de descente fixé. L'algorithme proposé pour l'étalonnage d'un réseau de capteurs par factorisation informée en matrices semi-non-négatives avec contrainte de structure de Vandermonde (SC-ISN-Cal pour *Structure-Constrained Informed Semi-NMF based Calibration*) est finalement résumé dans l'Algorithme 5.

Algorithme 5 Étalonnage par Factorisation Informée en matrices Semi-Non-négatives avec contrainte de structure de Vandermonde (SC-ISN-Cal).

Initialisation de G et F .

pour k de 1 à itération max **faire**

Mise à jour de \mathbf{y} selon (5.19) et (5.20).

Mise à jour de G selon (5.21).

Mise à jour de F selon (5.24).

fin pour

5.4 Extension à la régularisation par utilisation de données constructeur

Comme pour les capteurs à réponses affines, nous disposons pour les capteurs à réponses polynômiales de paramètres d'étalonnage théoriques fournis par le fabriquant. Il peut être intéressant de s'en servir afin de régulariser la factorisation en cas de mauvais conditionnement de la matrice de données, comme évoqué précédemment.

Pour rappel, ces paramètres théoriques correspondent aux valeurs moyennes des paramètres d'étalonnage observés par le fabriquant. Pour les capteurs à réponses polynômiales d'ordre N il s'agit donc de $N + 1$ valeurs théoriques $\bar{f}_1, \dots, \bar{f}_{N+1}$ dont doivent s'approcher les moyennes des paramètres d'étalonnage estimés dans la matrice F , c.-à-d.

$$\begin{cases} \frac{1}{n} \cdot \sum_{i=1}^n f_{1,j} & \approx \bar{f}_1, \\ \vdots & \vdots \\ \frac{1}{n} \cdot \sum_{i=1}^n f_{N+1,j} & \approx \bar{f}_{N+1} \end{cases} \quad (5.25)$$

Comme présenté dans la section 4.5, l'information apportée par les données constructeur peuvent être incluses dans le problème d'optimisation par rapport à F (5.11). La fonction de coût associée s'écrit alors

$$\mathcal{J} = \frac{1}{2} \cdot \|W \circ (X - G \cdot F)\|_{\mathcal{F}}^2 + \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot (\bar{\Omega}_F \circ F) \cdot \mathbb{1}_{(n+1) \times 1} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2, \quad (5.26)$$

s.c. $\tilde{F} = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F,$

et peut être séparée en la somme $\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2$ avec

$$\mathcal{J}_1 = \frac{1}{2} \cdot \|W \circ (X - G \cdot F)\|_{\mathcal{F}}^2 \quad \text{et} \quad \mathcal{J}_2 = \frac{\nu}{2} \cdot \left\| \frac{1}{n} \cdot (\bar{\Omega}_F \circ F) \cdot \mathbb{1}_{(n+1) \times 1} - \bar{\mathbf{f}} \right\|_{\mathcal{F}}^2. \quad (5.27)$$

La dérivée de \mathcal{J}_1 par rapport à Δ_F est présentée dans l'équation (5.23), tandis que celle de \mathcal{J}_2 est présentée dans l'équation (4.43). La mise à jour de Δ_F par descente de gradient est alors immédiate et s'écrit

$$\Delta_F \leftarrow \Delta_F - \rho_F \cdot \left(\frac{\partial \mathcal{J}_1}{\partial \Delta_F} + \frac{\partial \mathcal{J}_2}{\partial \Delta_F} \right). \quad (5.28)$$

Les données fournies par le fabriquant des capteurs permettent de contraindre la matrice F . La régularisation apportée n'influence donc que le problème d'optimisation sur F et le problème sur G reste tel qu'écrit dans (5.10). La formule de mise à jour de G proposée dans la méthode SC-ISON-Cal est donc toujours applicable ici. L'algorithme de cette approche est présenté dans l'Algorithme 6.

Algorithme 6 Étalonnage par Factorisation Informée en matrices Semi-Non-négatives avec contrainte de structure de Vandermonde et de moyenne (ASC-ISN-Cal).

Initialisation de G et F .

pour k de 1 à itération max **faire**

Mise à jour de \mathbf{y} selon (5.19) et (5.20).

Mise à jour de G selon (5.21).

Mise à jour de F selon (5.28) avec (5.23) et (4.43).

fin pour

5.5 Initialisation et astuces algorithmiques

Comme expliqué dans le chapitre précédent, les méthodes développées ici pour des capteurs à réponses non-linéaires doivent être initialisées. Dans les premiers tests réalisés dans [54], nous avons utilisé la même stratégie que celle utilisée dans [50], c.-à-d. que nous avons d'abord appliqué une approche de complétion [7] à la matrice X . Une fois X complétée, la matrice G peut à son tour être estimée par identification de la dernière colonne de X avec la seconde colonne de G à partir des équations (5.2) et (5.5). La matrice F peut ensuite être estimée par moindres carrés. Cependant et contrairement aux approches IN-Cal et ACIN-Cal, ce type d'initialisation amenait souvent en pratique vers des minima locaux. Nous avons alors initialisé aléatoirement F et G , comme discuté dans la section 4.6 (mis à part que G doit ici vérifier la structure de Vandermonde), ce qui a permis de meilleures performances dans des tests préliminaires. Dans les tests des méthodes non-linéaires qui seront effectués dans le chapitre suivant, nous initialiserons donc les deux matrices G et F par tirages aléatoires.

Un autre aspect non trivial des approches développées ici est le choix des pas de descente ρ_G et ρ_F introduits dans les équations (5.19) et (5.24). En effet, un choix de pas de descente trop petit risque d'entraîner une très faible vitesse de convergence de la méthode, tandis qu'avec un pas trop grand, la projection présente dans la mise à jour de G tend à être largement appliquée, menant à de médiocres performances d'étalonnage. Par ailleurs, dans nos tests préliminaires, nous avons observé que le choix des poids de descente adaptés dépend également de la proportion de données manquantes dans la matrice X . Les pas de descente choisis sont présentés dans le chapitre suivant consacré aux tests de toutes les méthodes développées jusqu'ici.

5.6 Conclusion

Dans ce chapitre, nous avons étendu les approches d'étalonnage proposées précédemment pour des capteurs à réponses affines et à paramètres non-négatifs au cas plus général de capteurs à réponses non-linéaires et à paramètres de signes quelconques.

Dans les approches proposées, les mesures effectuées par les capteurs sont disposées dans une matrice de données X non-négative, factorisable sous la forme d'un produit $G \cdot F$, où G contient la nouvelle structure du problème d'étalonnage qui prend la forme d'une matrice de Vandermonde, et F contient les valeurs des paramètres des capteurs. Les matrices X et G sont supposées non-négatives tandis que la matrice F est libre de toute contrainte de signe. Des données de référence issues de capteurs étalonnés sont également incluses aux données des capteurs mobiles, forçant alors certaines entrées des deux matrices. Par ailleurs, l'échantillonnage irrégulier produit par la mobilité des capteurs laisse place ici aussi à des données manquantes dans X , qui sont gérées dans le problème de factorisation par l'ajout d'une matrice de pondération dans le problème de factorisation.

Une première approche "naturelle" d'étalonnage par complétion de matrice a été proposée, avant de transposer le problème d'étalonnage des capteurs sous la forme d'un problème informé et pondéré de factorisation en matrice semi-non-négatives avec un facteur de Vandermonde. Dans cette dernière approche, les mises à jour multiplicatives introduites pour le modèle affine dans le chapitre précédent sont délaissées au profit de mises à jour par descente de gradient projeté dans le cas de la matrice non-négative G et standard dans le cas de la matrice F .

Une dernière méthode d'étalonnage a ensuite été décrite, permettant de tenir compte dans la factorisation de données constructeur, de façon à guider la factorisation dans le cas où la matrice de données X mal conditionnée, ne permettant pas à elle seule d'étalonner les capteurs de façon efficace.

Dans le chapitre suivant, nous proposons de mettre à l'épreuve l'ensemble des méthodes d'étalonnage proposées jusqu'à présent, pour les capteurs à réponses affines et pour ceux à réponses non-linéaires.

Étude de performances

Dans ce chapitre, nous proposons d'étudier les performances des différentes méthodes d'étalonnage présentées précédemment sur jeux de données simulés. Dans une première section, nous présentons donc la façon dont nous avons procédé pour obtenir des simulations "réalistes". Dans un second temps, nous comparons les performances obtenues par les différentes approches développées sur les jeux de données simulés.

6.1 Conditions de simulation

Afin d'obtenir des jeux de données qui soient réalistes, trois aspects de la simulation doivent être étudiés avec attention. Dans un premier temps, les paramètres des capteurs simulés doivent être proches de paramètres de capteurs réels, de façon à s'assurer que les méthodes d'étalonnage proposées sont suffisamment précises pour permettre l'estimation de paramètres réels. Dans un second temps, la scène \mathcal{S} simulée doit contenir des valeurs possibles du phénomène physique observé et être en accord avec les limites de sensibilité des capteurs simulés. En effet, les méthodes d'étalonnage développées que nous proposons de tester ici ne prennent pas en compte (i) les effets de saturation de capteurs qui arrivent lorsque ceux-ci sont exposés à des valeurs physiques dépassant les seuils de sensibilité et (ii) les effets de non-détection lorsque les valeurs physique sont en dessous du seuil de sensibilité des capteurs. Enfin, la proportion de données manquantes et de rendez-vous dans la matrice de données X sont induites par le déplacement des capteurs et doivent donc être simulées de façon à être cohérentes avec des données de *mobile crowdsensing*.

6.1.1 Sélection des capteurs

Dans les tests qui seront réalisés dans ce chapitre, nous nous intéresserons à la mesure du taux de particules fines. Ces polluants, émis par une grande partie des activités humaines (automobile, industries, chauffage, etc..) sont parmi les principaux éléments surveillés par les organismes de mesure de la qualité de l'air et représentent un enjeu majeur pour la santé. Les particules surveillées sont généralement classées suivant leur taille. Les PM10 qui correspondent aux particules de moins de 10 μm , les PM2.5 aux particules de moins de 2.5 μm et les PM1 à celles de moins de 1 μm . Les mesures du taux de particules fines sont exprimées en mg/m^3

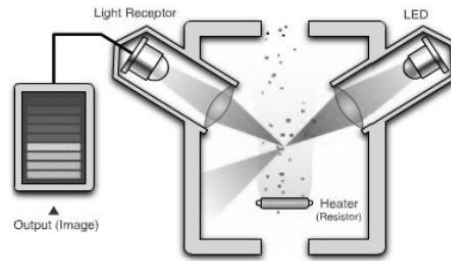


FIGURE 6.1 – Illustration du fonctionnement d’un capteur de particules. Source [203].

Les capteurs miniaturisés permettant la mesure du taux de particules fines sont pour la plupart basés sur un système optique utilisant la diffraction produite lorsqu’un faisceau lumineux entre en contact avec une particule. Une source lumineuse (laser ou diode infrarouge) et un détecteur photosensible sont placés de manière non-alignée dans le capteur. Lorsqu’une particule entre en contact avec le faisceau, la lumière est alors diffractée vers le photorécepteur qui détecte alors une particule. Suivant la taille de la particule coupant le faisceau, l’angle de diffraction varie, permettant de classer les particules suivant leur taille. Ce principe de fonctionnement est illustré dans la figure 6.1.

De très nombreuses références de capteurs de particules fines sont disponibles dans le commerce à des prix très variables suivant la qualité de l’optique utilisée dans le capteur mais également le système de circulation de l’air mis en place. Dans les capteurs les plus bas coût [74, 198], une résistance chauffante (comme présentée sur la figure 6.1) assure la circulation de l’air dans le boîtier. Ces résistances perturbent légèrement la mesure puisqu’elles détruisent les aérosols qui ne sont alors plus comptés dans les particules fines. Dans les modèles de capteur plus haut de gamme, le système chauffant est souvent remplacé par un ventilateur assurant la circulation de l’air tout en conservant intacts tous les éléments potentiellement mesurables [163]. Toutefois, les capteurs bas-coût avec circulation de l’air par chauffage sont les plus utilisés tant dans les projets de recherche [26, 153] que dans des projets/applications/bricolages amateurs (voir Fig. 6.2b)

6.1.1.1 Pour le modèle affine

Pour la simulation des capteurs à réponses affines, nous avons sélectionné le capteur de particules *Sharp GP2Y1010AU0F* [74], évoqué précédemment et présenté dans la figure 6.2a.

Dans la documentation technique [74], le fabricant fournit les valeurs moyennes pour le gain et l’offset du capteur, ainsi que des intervalles à l’intérieur desquels doivent se situer les paramètres réels des capteurs. Ces données constructeurs sont présentées dans la tableau 6.1.

À partir de ces données constructeur, nous avons effectué 40 simulations de réseaux constitués de $n = 25$ capteurs en tirant aléatoirement pour chaque simulation 25 offsets et 25 gains



FIGURE 6.2 – (a) : un capteur de particules *Sharp GP2Y1010AU0F*. (b) : un projet amateur de mesure de la qualité de l’air.

paramètre	min.	moyenne	max.
offset	0 V	0.9 V	1.5 V
gain	3.5 V/mg/m ³	5 V/mg/m ³	6.5 V/mg/m ³

TABLE 6.1 – Paramètres théoriques du capteur de particules *Sharp GP2Y1010AU0F*.

autour des moyennes théoriques. Les gains et offsets obtenus sont ensuite tronqués de façon à appartenir aux intervalles théoriques fournis par le constructeur. Ces paramètres sont stockés pour chaque simulation dans la matrice F_{theo} des paramètres théoriques, qui devra être retrouvée par les méthodes d’étalonnage. À titre d’exemple, les paramètres obtenus pour le premier réseau simulé sont présentés dans la figure 6.3. À cette matrice F_{theo} simulée, nous juxtaposons le vecteur $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ correspondant aux paramètres des capteurs de référence.

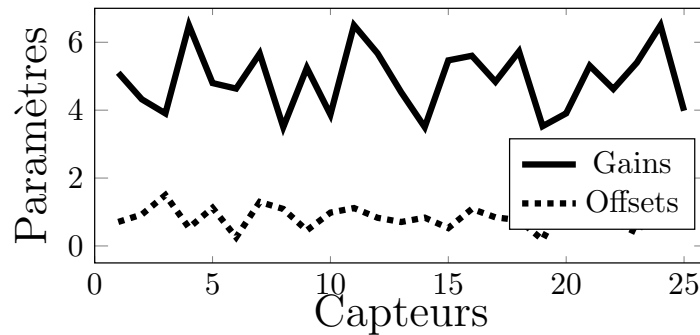


FIGURE 6.3 – Un exemple de simulation de paramètres de capteurs à réponses affines.

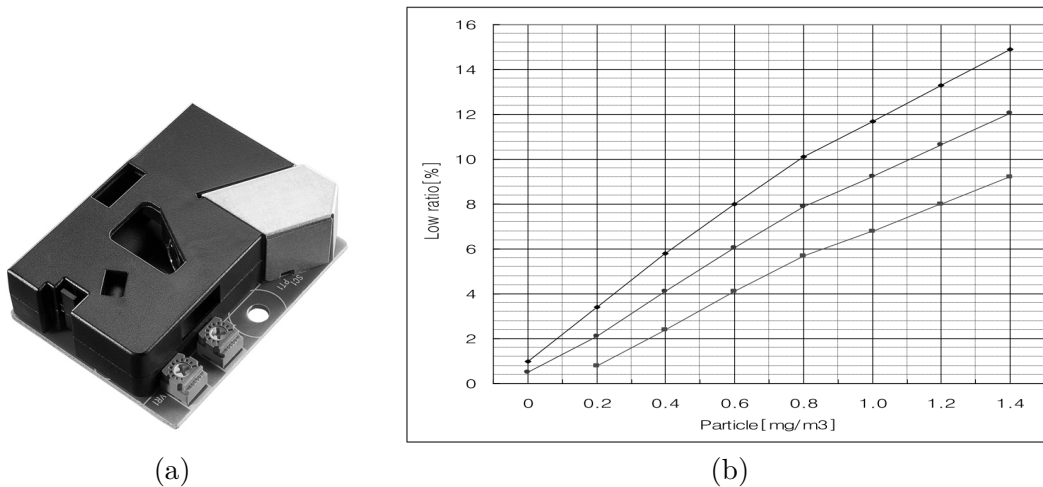


FIGURE 6.4 – (a) : un capteur de particules *stbm 271*. (b) : courbe d’étalonnage théorique du *stbm 271*, source [198].

6.1.1.2 Pour le modèle non-linéaire

Pour la simulation de capteurs à réponses non-linéaires, nous avons sélectionné le capteur de particules *STBM 2731* [198], présenté dans la figure 6.4a. Ce capteur, réputé plus précis que le précédent est fourni avec une notice technique de laquelle nous pouvons extraire la courbe d’étalonnage présentée dans la figure 6.4b. À partir de cette courbe, nous pouvons estimer les paramètres d’étalonnage moyens du capteurs, qui sont d’environ 0.9 pour l’offset, 13 pour le gain associé au terme d’ordre 1 et -2.5 pour celui associé au terme d’ordre 2. Comme précédemment, nous avons effectué 40 simulations de réseaux de 30 capteurs mobiles. Tous ces paramètres simulés constituent la matrice F_{theo} des paramètres théoriques des capteurs, à laquelle nous juxtaposons le vecteur $[0, 1, 0]^T$ des paramètres des capteurs de référence. Une fois les paramètres des capteurs simulés, nous nous intéressons à produire une scène \mathcal{S} à partir de laquelle nous pourrions construire les mesures prises par les réseaux de capteurs simulés précédemment.

6.1.2 Simulation d’une scène et d’une matrice de données

Pour la simulation d’une scène, nous avons considéré une gaussienne, discrétisée suivant un maillage de taille 10×10 cellules et normalisée de sorte que les valeurs physiques des cellules soient comprises entre 0 et 1.5 mg/m^3 , ce qui correspond à l’intervalle de sensibilité théorique des capteurs de particules qui ont été simulés précédemment. La scène produite, présentée dans la figure 6.5, peut par exemple représenter grossièrement la diffusion d’une source ponctuelle de pollution. Il est important de noter que la scène simulée ici n’a pas pour objectif d’être proche de la réalité d’une carte de pollution. Elle a principalement pour but de permettre de tester les méthodes d’étalonnage proposées précédemment sur des données simulées qui soient cohérentes avec les mesures pouvant être produites par des capteurs de

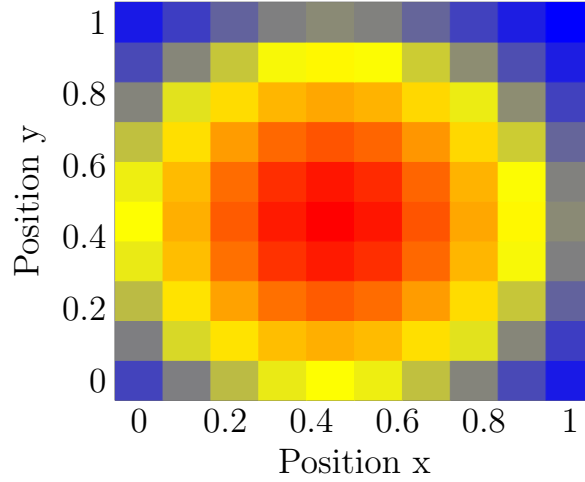


FIGURE 6.5 – Scène simulée.

pollution.

À partir de la scène simulée \mathcal{S} , il est facile de construire la matrice G_{theo} , en juxtaposant une colonne de uns et la scène vectorisée. Dans le cas des simulations pour les capteurs à réponses non-linéaires, la scène vectorisée est dupliquée dans G_{theo} , de façon à respecter la structure de Vandermonde de la matrice. Nous pouvons alors construire pour chaque jeu de paramètres la matrice de mesures théorique X_{theo} complète par le produit

$$X_{\text{theo}} = G_{\text{theo}} \cdot F_{\text{theo}}. \quad (6.1)$$

À ce stade, nous disposons de tous les éléments permettant de tester les approches d'étalonnage développées précédemment, c.-à-d. des matrices de données X_{theo} complètes qui seront partiellement vidées aléatoirement par la suite pour visualiser l'influence des données manquantes dues aux déplacements des capteurs, ainsi que les matrices de paramètres théoriques F_{theo} qui permettront d'évaluer la qualité de la reconstruction des paramètres par nos méthodes d'étalonnage.

6.1.3 Discussion sur les paramètres et les critères de performance

Paramètres de simulation

Afin de tester le comportement des méthodes d'étalonnage développées, nous proposons de faire varier trois paramètres. La proportion ρ_{DM} de données manquantes dans X liées aux déplacements des capteurs, la proportion ρ_{RV} de capteurs mobiles ayant effectué une mesure en rendez-vous avec un capteur de référence, et le bruit de mesure, c.-à-d. le SNR d'entrée sur la matrice de données X . Dans chaque simulation, 4 mesures de référence sont placées aléatoirement dans la dernière colonne de X .

Quelques précisions méthodologiques

Pour toutes les méthodes d'étalonnage proposées, les matrices G et F doivent être initialisées, nous proposons des initialisations aléatoires autour des valeurs théoriques des capteurs pour F , autour de la valeur moyenne de la scène pour G . De plus les méthodes étant itératives, nous fixons le nombre d'itérations à $5.e5$. Pour la méthode ACIN-Cal, le paramètre de régularisation est, sauf mention contraire, fixé à $\nu = 10$, tandis que les valeurs $\bar{\mathbf{f}}$ des paramètres de régularisation sont, sauf mention contraire, fixés aux valeurs exactes des moyennes des paramètres de F .

Dans le cas des approches SC-ISN-Cal et ASC-ISN-Cal, le pas de gradient doit également être défini. Suite à de nombreux tests nous avons pu effectuer deux constats, le premier étant que le pas de gradient pour la mise à jour de G doit être choisi avec plus d'attention que celui pour F . La matrice G étant contrairement à F sujette à une contrainte de positivité un pas de gradient trop important mène à un seuillage de G perturbant totalement les performances des méthodes. Par ailleurs, nous avons également pu remarquer qu'il est intéressant d'ajuster le pas de gradient en fonction de la proportion de données manquantes. Après de nombreux tests, nous avons choisi d'utiliser comme pas de gradient ρ_G pour G et ρ_F pour F

$$\rho_G = 0.001(\rho_{DM} + 0.1) \quad \text{et} \quad \rho_F = 0.01(\rho_{DM} + 0.1). \quad (6.2)$$

Il est important de noter que le choix des pas de gradient pourrait sans doute être optimisé. Cependant, comme nous l'avons brièvement évoqué dans le chapitre précédent, la paramétrisation utilisée dans nos méthodes – avec notamment la présence de produit de Hadamard – rend la recherche d'un pas optimal non triviale. Par ailleurs, nous avons proposé une approche rapide de NMF basée sur le gradient de Nesterov et adaptée à la factorisation d'une matrice à données manquantes. Cette approche – présentée dans l'annexe C – pourrait tout à fait être appliquée ici et ainsi pallier au problème de selection des pas de descente. Toutefois, nous n'avons pas pu la tester dans le cadre de cette thèse par manque de temps et cela restera dans les proches perspectives de ce travail.

Critère de performance

Le critère d'évaluation choisi pour évaluer les performances des méthodes est le RMSE (pour *root mean square error*) entre les paramètres estimés et les paramètres théoriques. Pour l'évaluation de l'estimation des offsets, le RMSE s'écrit

$$\text{RMSE}(\bar{\mathbf{f}}_1, \bar{\mathbf{f}}_{\text{theo}_1}) = \frac{\|\bar{\mathbf{f}}_1 - \bar{\mathbf{f}}_{\text{theo}_1}\|_2}{\sqrt{n}}, \quad (6.3)$$

où n est le nombre de capteurs.

6.2 Performances sur le modèle affine

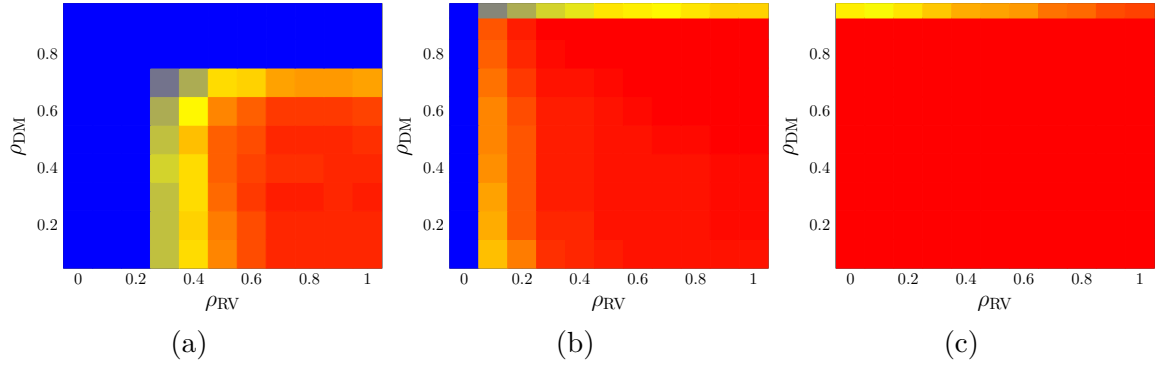


FIGURE 6.6 – Performances globales de : (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.

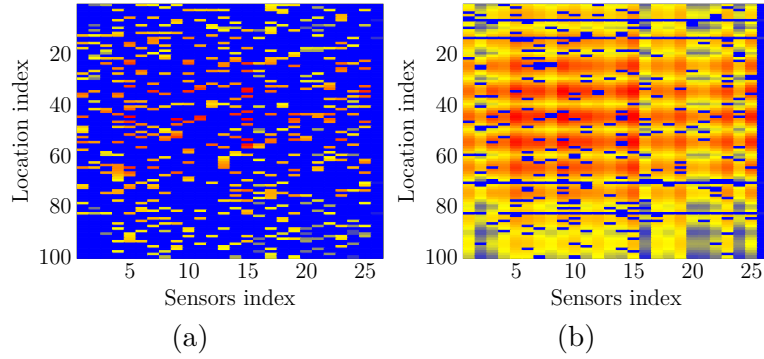


FIGURE 6.7 – Exemples de matrices de données X construites avec (a) 80% de données manquantes et (b) 10% de données manquantes.

Performances globales

Dans un premier temps, nous proposons une vue générale des performances obtenues avec les méthodes d'étalonnage NMF/C-Cal, IN-Cal et ACIN-Cal. Nous faisons donc varier la proportion de données manquantes ρ_{DM} de 10 à 95%, la proportion de rendez-vous de 0 à 100% et nous considérons le cas sans bruit ($SNR = \infty$). Pour chaque couple de paramètres (ρ_{DM}, ρ_{RV}) , nous effectuons l'étalonnage avec les trois méthodes présentées précédemment sur 40 simulations différentes. La figure 6.7 montre les matrices de données X obtenues sur une simulation pour différentes proportion de données manquantes ρ_{DM} .

La figure 6.6 montre pour chaque méthode et pour chaque couple de paramètres la proportion de simulations pour lesquelles le RMSE final est sous le seuil de $1.e - 8$. Dans cette figure, une case de couleur bleue signifie que pour le couple de paramètres correspondant, aucune des simulations n'a atteint le seuil de $1.e - 8$, tandis qu'une case rouge signifie que toutes les simulations se sont bien déroulées et ont atteint un RMSE final d'au plus $1.e - 8$.

L'approche d'étalonnage NMF/C-Cal, parvient à étalonner le réseau de capteurs lorsque celui-ci est suffisamment dense, c.-à-d. lorsque la proportion de données manquantes ρ_{DM} est inférieure à 50% et que la proportion de rendez-vous entre les capteurs de référence et ceux à

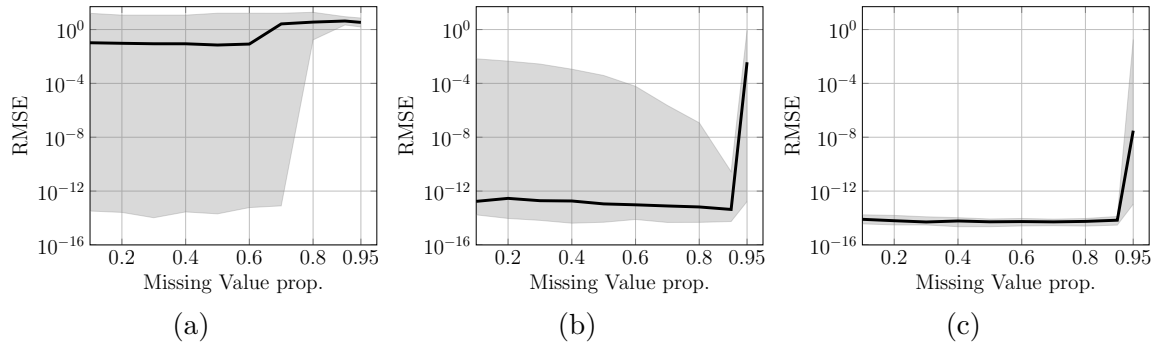


FIGURE 6.8 – Performances par rapport à la proportion de données manquantes ($\rho_{RV} = 0.3, \text{SNR} = \infty$) pour (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.

étalonner ρ_{RV} est au dessus de 60%. Lorsque ces deux conditions sont respectées, NMF/C-Cal atteint un taux de succès pour l'étalonnage entre 80 et 100%. Lorsque ρ_{DM} augmente et/ou lorsque ρ_{RV} diminue, les performances chutent brutalement. Ainsi, lorsque $\rho_{DM} \geq 80\%$ ou lorsque $\rho_{RV} \leq 20\%$, NMF/C-Cal ne parvient jamais à un RMSE final inférieur à $1.e - 8$.

Comme on peut le voir sur la figure 6.6b IN-Cal en revanche se montre bien plus robuste à la proportion de données manquantes ainsi qu'à la proportion de rendez-vous. En effet, mis à part lorsque $\rho_{DM} = 95\%$ et/ou lorsqu'il n'y a pas de rendez-vous entre les capteurs de référence et les capteurs à étalonner, cette approche permet, dans la majorité des cas, un étalonnage précis des capteurs. La chute de performances pour des valeurs extrêmes de ρ_{DM} et ρ_{RV} était attendue puisque dans ces conditions, les hypothèses nécessaires à la méthode ne sont pas vérifiées. En particulier, lorsqu'il n'y a pas de rendez-vous entre les références et les capteurs mobiles, l'ambiguïté d'échelle du problème de NMF n'est pas résolue. De même, lorsqu'il y a trop de données manquantes, certains capteurs du réseau sont isolés et ne sont plus reliés aux références. Dans ce dernier cas, le problème d'échelle n'intervient donc que sur certains capteurs et ne peut être résolu.

Les paramètres de régularisation apportés à ACIN-Cal rendent les performances de la méthode totalement indépendante de la proportion de rendez-vous entre les références et les capteurs mobiles. En effet, les rendez-vous étaient nécessaires dans IN-Cal afin de résoudre l'ambiguïté d'échelle du problème de NMF. En introduisant la régularisation avec les données constructeur, les lignes de F sont automatiquement mises à l'échelle et les rendez-vous ne sont donc plus obligatoires. Il faut toutefois souligner le fait que dans les tests réalisés ici, la moyenne exacte des paramètres des capteurs est utilisée dans ACIN-Cal. Les lignes de F sont donc parfaitement normalisées et mises à la bonne échelle. L'influence de l'utilisation de paramètres de régularisation erronés sera étudiée plus loin dans ce chapitre.

Nous proposons maintenant d'étudier plus en détail le comportement des méthodes d'étalonnage développées par rapport aux paramètres de simulation, c.-à-d. par rapport à ρ_{DM} , ρ_{RV} et le SNR d'entrée, en les faisant varier un par un.

Influence des données manquantes

Afin de mettre en évidence l'influence de la proportion de données manquantes dans la matrice X , nous fixons la proportion de rendez-vous ρ_{RV} à 30% et le SNR d'entrée à ∞ (pas de bruit de mesure). Nous faisons alors varier la proportion de données manquantes de 10 à 95%. Pour chaque valeur de ρ_{DM} , nous appliquons les méthodes d'étalonnage sur 40 simulations. La figure 6.8 montre les performances obtenues avec chaque méthode. Dans cette figure, la courbe montre le RMSE médian obtenu sur 40 simulations pour chaque méthode d'étalonnage, tandis que l'enveloppe délimite les valeurs maximale et minimale de RMSE obtenues pour chaque méthode.

Il est immédiat de constater que la méthode d'étalonnage par complétion de matrice NMF/C-Cal, ne fournit de résultats satisfaisants ni au niveau de son enveloppe (valeur maximale toujours haute), ni pour sa médiane et est surpassée par les méthodes d'étalonnage à base de NMF informée IN-Cal et ACIN-Cal quelque soit la valeur de ρ_{DM} . Il est à noter que la même méthode utilisant une SVD tronquée pour estimer X_{comp} fournit des performances médianes relativement similaires avec une partie basse de l'enveloppe beaucoup plus haute (autour de 10^{-1}), comme on peut le voir dans [50].

La méthode IN-Cal produit des résultats satisfaisant dans la majorité des simulations, avec un RMSE médian en dessus de $1.e - 12$ pour des valeurs de ρ_{DM} entre 0 et 90%. Lorsque $\rho_{DM} = 95\%$, les capteurs mobiles ne sont plus suffisamment connectés entre eux, et des groupes se retrouvent isolés des mesures de références. Ces groupes de capteurs isolés sont alors sujet à l'ambiguïté d'échelle et ne sont donc pas parfaitement étalonnés. Il est important de remarquer que la taille de l'enveloppe est essentiellement due à la faible vitesse de convergence de cette la méthode IN-Cal. L'enveloppe est considérablement réduite autour de la valeur médiane du RMSE en choisissant un nombre d'itérations plus élevé comme nous l'avons montré dans [50] où la même méthode a été testée avec $1.e6$ itérations.

Comme on peut le voir sur la figure 6.6c, ACIN-Cal propose des performances médianes légèrement meilleures qu'IN-Cal. Par contre, mis à part quand $\rho_{DM} = 95\%$, l'enveloppe obtenue avec ACIN-Cal est très réduite, ce qui montre l'apport de la moyenne dans l'approche puisque la régularisation apportée dans ACIN-Cal accélère la convergence de la méthode originale IN-Cal. Dans le cas le plus extrême où 95% des données sont manquantes dans X , ACIN-Cal possède toujours une enveloppe élevée, ce qui était attendu : l'introduction de la régularisation par les données constructeur est destinée à résoudre le problème d'étalonnage dans le cas où aucune des mesures de référence ne serait disponible.

Influence des rendez-vous

Nous proposons ici d'étudier l'influence de la proportion de rendez-vous sur les approches d'étalonnage proposées. Ainsi, nous fixons la proportion de données manquantes à 90%, et le SNR d'entrée à ∞ (pas de bruit). Nous faisons alors varier la proportion ρ_{RV} de capteurs mobiles ayant un rendez-vous avec une mesure de référence de 0 à 100%. Comme précédemment, 40 simulations d'étalonnage sont réalisées pour chaque valeur de ρ_{RV} . La figure 6.9 montre les performances obtenues avec chaque méthode d'étalonnage. La courbe y représente la valeur médiane de RMSE obtenue pour chaque méthode d'étalonnage, tandis que l'enveloppe montre

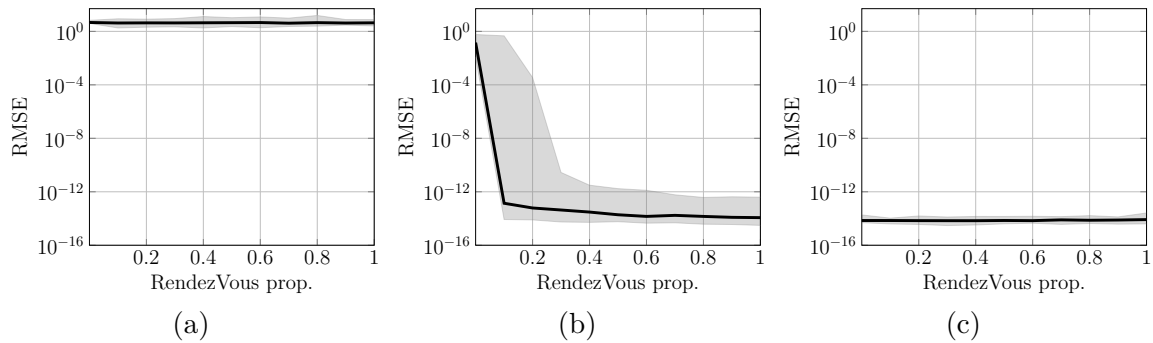


FIGURE 6.9 – Performances par rapport à la proportion de rendez-vous ($\rho_{DM} = 0.9, SNR = \infty$) pour (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.

les valeurs minimales et maximales de RMSE atteintes.

Comme précédemment, il est immédiat de constater que les méthode d'étalonnage à base de factorisation matricielle informée IN-Cal et ACIN-Cal produisent de bien meilleures performances que NMF/C-Cal. IN-Cal montre de bonnes performances tant qu'il existe des rendez-vous entre les capteurs mobiles à étalonner et les capteurs de référence. Lorsque la proportion de rendez-vous diminue jusqu'à atteindre 0%, l'ambiguïté d'échelle classique de NMF ne peut plus être fixée et les performances d'étalonnage chutent alors brutalement. Toutefois, il est bon de rappeler que lorsque l'ambiguïté d'échelle n'est pas résolue, IN-Cal permet d'étalonner le réseau de capteurs de façon relative et que donc les paramètres des capteurs sont bien estimés, à un facteur près.

Quelque soit la proportion de rendez-vous dans la matrice X , la méthode ACIN-Cal parvient à maintenir un RMSE remarquablement bas, grâce à l'utilisation de la moyenne théorique des paramètres qui permet de réduire l'erreur d'estimation dans les cas où IN-Cal n'effectue qu'un étalonnage relatif des capteurs. Toutefois, rappelons que les très bonnes performances obtenue par cette méthode sont susceptibles de se dégrader légèrement en conditions réelles où les valeurs théoriques moyennes ne correspondront pas à la moyenne réelle des paramètres des capteurs. L'influence des paramètres de régularisation et notamment de l'emploi de valeurs théoriques erronées, sera discutée plus loin dans ce chapitre.

Sensibilité au bruit

Afin de montrer la sensibilité des méthodes d'étalonnage proposées par rapport au bruit de mesure, nous fixons la proportion de données manquantes ρ_{DM} à 90%, la proportion de rendez-vous ρ_{RV} à 30%, et nous faisons varier le SNR d'entrée de ∞ (pas de bruit) à 20 dB par l'ajout d'un bruit gaussien sur la matrice de données X . Les simulations de bruit sont réalisées de façon à conserver des matrices de données bruitées qui soient non-négatives.

Dans le cas de l'approche NMF/C-Cal, les performances obtenues sont relativement mauvaises et le niveau de bruit a finalement peu d'impact sur les valeurs de RMSE atteintes. Les méthodes IN-Cal et ACIN-Cal montrent des performances qui évoluent de façon similaire

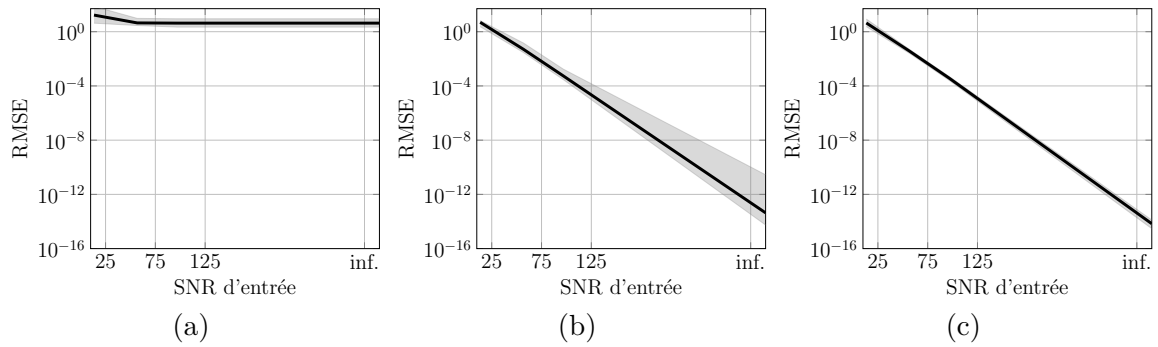


FIGURE 6.10 – Performances par rapport au SNR d'entrée ($\rho_{DM} = 0.9, \rho_{RV} = 0.3$) pour (a) NMF/C-Cal, (b) IN-Cal, (c) ACIN-Cal.

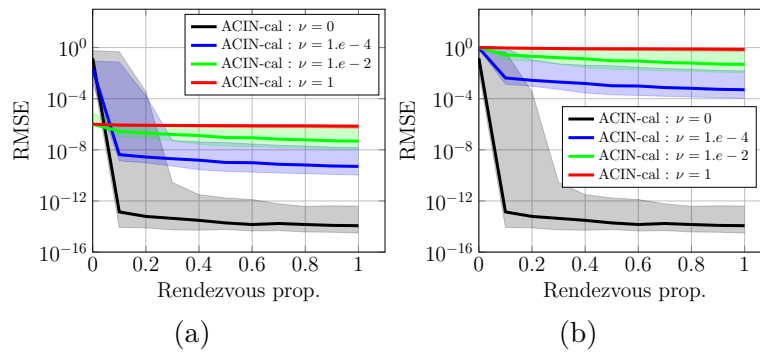


FIGURE 6.11 – Influence des paramètres de régularisation dans la méthode ACIN-Cal avec différents niveau d'erreur sur les paramètres théoriques. À gauche avec une erreur de l'ordre de $1.e - 6$ et à droite de l'ordre de 1.

par rapport au niveau de bruit. Pour ces deux méthodes, le RMSE médian évolue de façon linéaire avec l'augmentation du niveau de bruit.

Choix des paramètres de régularisation

Dans l'approche ACIN-Cal, le choix du poids ν , associé au terme de régularisation doit être attentivement choisi. Dans les simulations précédentes, nous avons fixé le poids $\nu = 10$ et les moyennes théoriques des paramètres utilisées dans la régularisation aux moyennes exactes des paramètres à estimer. Ce choix idéal de paramètres théoriques est également utilisé dans d'autres approches de la littérature, comme par exemple [11]. En situation réelle, les paramètres théoriques fournis par le constructeur ne sont pas rigoureusement égaux aux moyennes exactes des paramètres. Dans ce cas, le choix d'un poids ν trop fort contraindra la moyenne des paramètres estimés à être proche de valeurs théoriques erronées et les performances d'étalement se verront alors fortement dégradées. Inversement, si le poids ν est choisi trop faible, la régularisation aura alors moins d'influence sur la méthode.

L'influence du choix du paramètre ν est mise en évidence dans la figure 6.11 où nous

montrons les performances obtenues avec ACIN-Cal, sur des simulations sans bruit, avec 90% de données manquantes et en faisant varier la proportion de rendez-vous de 0 à 100%. Pour chaque valeur de ρ_{RV} , nous avons réalisé 40 simulations. Les courbes dans la figure 6.11 montrent la valeur médiane de RMSE atteinte par la méthode, tandis que les enveloppes délimitent l'intervalle dans lequel se situe l'ensemble des RMSE obtenus après étalonnage. Dans la figure 6.11a l'erreur sur la valeur théorique est fixée à 1 et à $1.e-6$ dans la figure 6.11b. Les courbes et enveloppes en noir correspondent aux tests réalisés avec un poids $\nu = 0$, celles en bleu avec $\nu = 1.e-4$, en vert $\nu = 1.e-2$ et en rouge $\nu = 1$.

Dans la figure 6.11a, l'erreur due au paramètres moyen d'étalonnage est importante – autour de $1.e-4$ – et affecte grandement les performance d'étalonnage pour toutes les valeurs non nulles de ν . En effet, lorsque $\nu = 1$, le RMSE median reste autour de 1 et ce même quand $\rho_{RV} = 100\%$, c.-à-d. quand tous les capteurs sont en rendez-vous. Celà montre que la régularisation a un impact majeur sur la factorisation. Lorsque la valeur de ν diminue, les RMSE sont légèrement plus bas – du au fait que l'impact de la régularisation est diminué – excepté quand $\rho_{RV} = 0$ où le RMSE obtenu avec IN-Cal et l'erreur sur la moyenne théorique sont similaires.

Dans la figure 6.11b, l'analyse est un peu différente. Dans ce cas, l'erreur due au paramètres moyen d'étalonnage est plutôt faible. Pas conséquent, lorsque $\rho_{RV} = 0$, la régularisation permet d'améliorer les performances par rapport à l'approche IN-Cal. Cependant, lorsque $\nu = 1.e-4$, l'influence de la régularisation est trop faible et les performances obtenue sont alors similaires à celle de IN-Cal. Lorsque ρ_{RV} augmente – c.-à-d. lorsque les hypothèses nécessaire à un bon étalonnage par IN-Cal sont satisfaites – nous retrouvons le même comportement que dans la figure 6.11a, c.-à-d. que les valeurs de ν les plus fortes fixent l'erreur d'étalonnage à l'erreur sur les paramètres théoriques, alors que des valeurs plus faibles de ν permettent d'obtenir un RMSE plus faible.

Ces tests montrent l'importance de l'erreur due à la régularisation par les paramètres constructeur sur les performances d'étalonnage, d'où l'intérêt de tenir compte de cette régularisation sous la forme d'un terme de pénalisation et non comme une normalisation stricte. Ces tests ont également montré que la selection du paramètre ν n'est pas triviale et pourrait sans doute être optimisée.

Comparaison avec une approche d'étalonnage multi-sauts

Nous nous intéressons ici à comparer les performances de notre approche IN-Cal avec l'approche d'étalonnage multi-sauts développée par Saukh *et al.* dans [182]. Cette méthode présentée brièvement dans le chapitre d'état de l'art de cette thèse page 13 a été conçue dans le cadre d'un projet de mesure de la qualité de l'air par *mobile crowdsensing* dans lequel les capteurs mobiles ont été placés sur des tramways. Les mouvements des capteurs sont donc réguliers et assurent un nombre de rendez-vous considérable entre les capteurs mobiles mais également avec des données de stations de référence fixes. Avec cette configuration particulière de réseau, l'approche d'étalonnage multi-sauts a montré des performances intéressantes dans [182]. Cependant, dans les tests que nous avons menés jusqu'ici, chaque capteur n'entraîne en rendez-vous qu'avec au plus une donnée de référence. Dans ce cas, les approches multi-sauts

ne sont pas applicables puisque nécessitant à chaque étape que le capteur à étalonner ait au moins deux mesures distinctes en rendez-vous avec un capteur de référence, de façon à pouvoir estimer ses deux paramètres, gain et offset.

Nous avons donc proposé de comparer les performances de notre approche IN-Cal avec celles atteintes par l’approche multi-sauts [182] sur un nouveau jeu de données simulées qui respecte les contraintes nécessaires à [182]. Ces nouvelles simulations consistent à retirer aléatoirement des entrées de la matrice de données X de façon à obtenir une proportion de données manquantes ρ_{DM} que nous faisons varier de 0 à 95%. Dans ces simulations, nous laissons libre ρ_{RV} la proportion de rendez-vous qui évoluera naturellement avec le changement de la proportion de données manquantes.

Pour chaque proportion de données manquantes, 40 simulations sont effectuées. Dans les figures 6.12a et 6.12b, nous montrons le RMSE médian ainsi que son enveloppe pour chaque valeur de ρ_{DM} avec, en rouge, l’approche multi-sauts et en bleue IN-Cal. Dans les simulations correspondant à la figure 6.12a la matrice X est non bruitée, tandis que dans celles de la figure 6.12b le SNR d’entrée est fixé à environ 30 dB.

Dans la figure 6.12a, les deux approches testées montrent des performances similaires lorsque la proportion de données manquantes dans la matrice X est faible, c.-à-d. lorsque $\rho_{DM} \leq 80\%$. Cependant, lorsque ρ_{DM} augmente, IN-Cal produit de bien meilleures performances que l’approche multi-sauts, ce qui était attendu, du au fait que lorsque $\rho_{DM} > 80\%$, le nombre minimum de rendez-vous entre un capteur mobile et une référence peut être inférieur à 2 dans certaines simulations. Enfin, lorsque $\rho_{DM} > 90\%$, alors les deux méthodes montrent des performances fortement dégradées, dues aux simulations dans lesquelles au moins un capteur effectue moins de 2 mesures distinctes.

Dans le cas bruité montré dans la figure 6.12b, le RMSE médian obtenu par IN-Cal est toujours en dessous de celui atteint par l’approche multi-sauts. Lorsque la proportion de données manquantes est grande, les performances de l’approche multi-sauts sont très fortement dégradées tandis que IN-Cal produit des résultats bien meilleurs. En réalité, lorsque $\rho_{DM} = 95\%$, les performances obtenue par IN-Cal avec et sans bruit sont similaires, pour ce niveau de bruit fixé.

6.3 Performances sur le modèle non-linéaire

Comme pour les capteurs à réponse affine, nous proposons ici d’étudier les performances des approches SC-ISN-Cal et ASC-ISN-Cal. Compte tenu des piètres performances obtenues avec l’approche d’étalonnage par complétion matricielle et regression sur le modèle affine, nous nous focaliserons donc dans ce chapitre sur les performances obtenues par nos approches à base de factorisations matricielles informées.

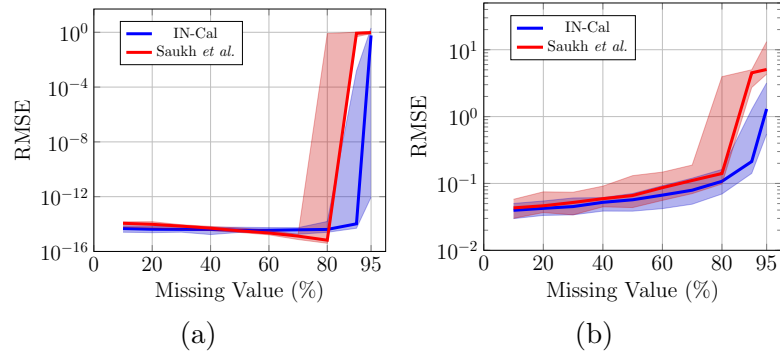


FIGURE 6.12 – Comparaison des performances de IN-Cal (en bleu) vs. l'étalonnage multi-sauts [182] (en rouge) dans (a) un cas sans bruit et (b) avec un SNR d'entrée d'environ 30 dB.

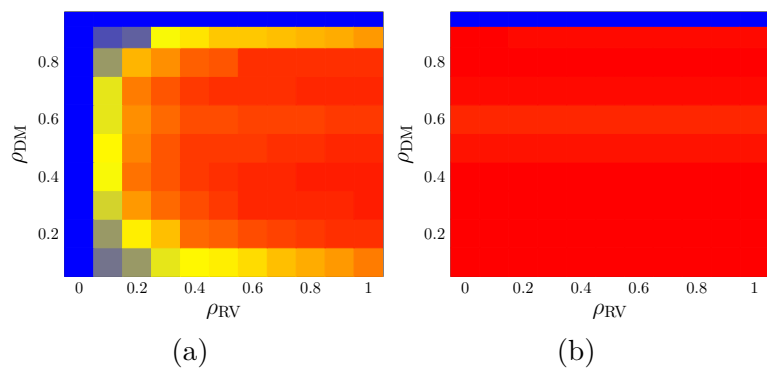


FIGURE 6.13 – Performances globales de : (a) SC-ISN-Cal, (c) ASC-ISN-Cal.

Performances globales

Dans un premier temps, nous proposons un aperçu général du comportement des deux méthodes SC-ISN-Cal et ASC-ISN-Cal, par rapport à la proportion de données manquantes ρ_{DM} et à la proportion de rendez-vous ρ_{RV} . Nous fixons donc le SNR d'entrée à ∞ (pas de bruit) et nous effectuons 40 simulations par couple de paramètres (ρ_{DM}, ρ_{RV}) . Il est important de noter que le problème à résoudre ici est bien plus complexe que celui à traiter pour l'étalonnage de capteurs à réponse affine. En effet, les paramètres à estimer avec le modèle non-linéaire sont plus nombreux pour chaque capteur, sans contraintes de signe et une structure de matrice est à respecter. Par ailleurs, bien que le problème à résoudre soit plus complexe, nous avons conservé pour toutes les simulations effectuées dans cette section le même nombre de capteur de mesures de référence que dans les tests effectués pour le modèle affine, c.-à-d. 4 mesures de référence par jeu de données. Le risque que le problème soit mal posé est donc plus important ici que dans les tests précédents.

La figure 6.13 montre pour chaque couple de paramètres, le comptage des simulations dont l'erreur finale d'étalonnage – obtenue après $5.e5$ itérations – est inférieure ou égale à $1.e - 2$. Dans cette figure, un "pixel" rouge signifie que 100% des 40 simulations ont atteint le seuil de RMSE fixé – ici $1.e - 2$ – tandis qu'un pixel bleu indique au contraire que tous les tests ont échoué à atteindre ce seuil. Le dégradé de couleur du bleu au rouge rend compte de la proportion de simulations ayant atteint le seuil de RMSE.

Il est immédiat de constater que le comportement de SC-ISN-Cal montré dans la figure 6.13a est relativement similaire à celui de l'approche IN-Cal montré dans la figure 6.6b. En effet, les meilleures performances sont obtenues lorsqu'il y a beaucoup de rendez-vous dans la matrice X et une proportion raisonnable de données manquantes, autour de 50% pour obtenir les meilleurs RMSE. Lorsque la proportion de données manquantes devient trop importante, certains capteurs ne fournissent plus suffisamment de mesures ou ne sont plus connectés à suffisamment de données de référence pour permettre un étalonnage correcte du réseau avec SC-ISN-Cal. De même lorsqu'il n'y a pas suffisamment de rendez-vous entre les capteurs mobiles et les références alors les performances de SC-ISN-Cal sont limitées.

Comme nous pouvons le voir dans la figure 6.13b, la régularisation apportée par les données constructeur des capteurs dans ASC-ISN-Cal améliore considérablement les performances d'étalonnage puisqu'elle permet une bonne estimation des paramètres même lorsqu'il y a peu voir aucun rendez-vous entre les capteurs mobiles et les données de référence. Par ailleurs, l'utilisation de la régularisation avec données constructeur accélère également la vitesse de convergence de la méthode ce qui permet d'améliorer les performances obtenues même dans les configurations où l'approche SC-ISN-Cal devrait fournir de bons résultats mais est trop lente à converger. En revanche, lorsque trop de données sont manquantes, cette régularisation ne permet pas d'améliorer les performances d'étalonnage puisqu'elle ne permet pas de connecter les capteurs mobiles les uns aux autres. Dans ce cas le problème d'ambiguïté d'échelle propre au problème de factorisation matricielle n'est pas fixé.

Le comportement des approches SC-ISN-Cal et ASC-ISN-Cal montré dans la figure 6.13 est semblable à celui des approches IN-Cal et ACIN-Cal montré dans la figure 6.6, à un facteur d'échelle près, puisque pour la construction de la figure 6.6 nous sélectionné un seuil de RMSE

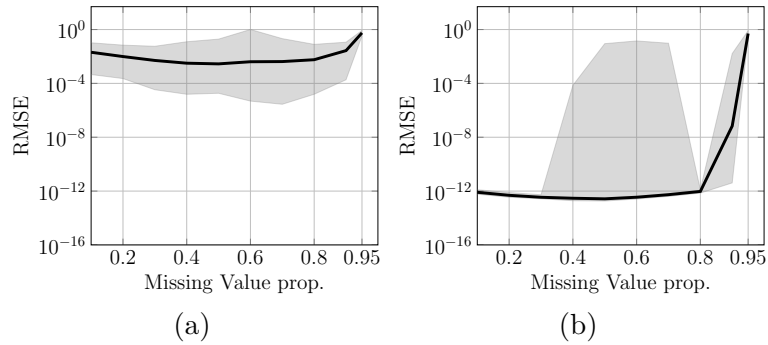


FIGURE 6.14 – Performances par rapport à la proportion de données manquantes ($\rho_{RV} = 0.3$, $SNR = \infty$) pour (a) SC-ISN-Cal, (b) ASC-ISN-Cal.

fixé à $1.e - 8$ tandis qu'ici nous avons fixé ce seuil à $1.e - 2$. Cette différence de choix de seuil s'explique par le problème à résoudre qui est nettement plus complexe dans le cas du modèle non-linéaire que dans le cas du modèle affine étudié précédemment.

Dans les paragraphes suivants, nous étudions plus en détails l'influence de chaque paramètre ρ_{DM} , ρ_{RV} ainsi que du SNR d'entrée.

Influence des données manquantes

Afin d'évaluer plus précisément l'impact de la proportion de données manquantes sur les performances de ces méthodes, nous fixons la proportion de rendez-vous ρ_{RV} à 30% et le SNR d'entrée à ∞ (pas de bruit). Nous faisons alors varier la proportion de données manquantes ρ_{DM} de 10 à 95% et réalisons comme précédemment 40 simulations pour chaque valeur de ρ_{DM} . Dans la figure 6.14, la courbe montre la valeur médiane de RMSE atteinte par chaque méthode sur les 40 simulations réalisées tandis que l'enveloppe rend compte de la variabilité des RMSE au cours de ces tests. Nous pouvons voir sur la figure 6.14a que le RMSE médian obtenu par l'approche SC-ISN-Cal est relativement stable pour toutes les valeurs de ρ_{DM} (autour de $1.e-2$), sauf lorsque 95% des données sont manquantes. Dans ce cas, la valeur médiane est proche de 1. L'ajout de la régularisation à l'aide des paramètres théoriques dans ASC-ISN-Cal permet de diminuer fortement l'erreur d'étalonnage pour atteindre un RMSE médian de $1.e - 12$ pour toute valeur de ρ_{DM} en dessous de 90%. La hauteur importante de l'enveloppe dans la figure 6.14b est probablement due à la faible vitesse de convergence de quelques simulations parmi les 40 effectuées.

Influence des rendez-vous

Pour la mise en évidence de l'impact de la proportion de rendez-vous sur l'étalonnage, nous fixons la proportion de données manquantes ρ_{DM} à 90%, le SNR d'entrée à ∞ (pas de bruit) et nous faisons varier la proportion de rendez-vous ρ_{RV} de 0 à 100%. Les figures 6.15a et 6.15b montrent les performances obtenues par les méthodes d'étalonnage sur 40 simulations par valeur de ρ_{RV} . Là encore, les courbes représentent la valeur médiane de RMSE atteinte tandis que les enveloppes délimitent les valeurs hautes et basses du RMSE. Nous constatons alors que

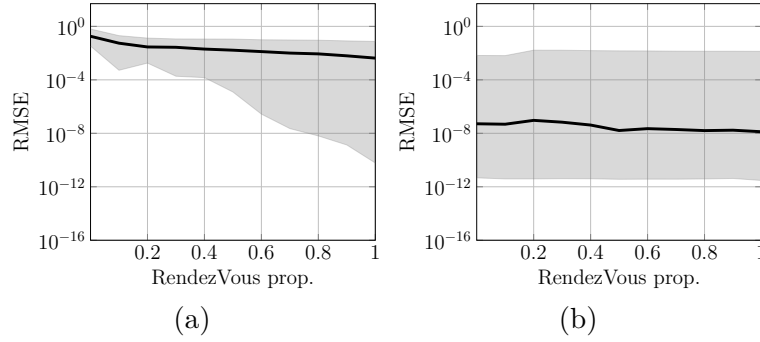


FIGURE 6.15 – Performances par rapport à la proportion de rendez-vous ($\rho_{DM} = 0.9, \text{SNR}=\infty$) pour (a) SC-ISN-Cal, (c) ASC-ISN-Cal.

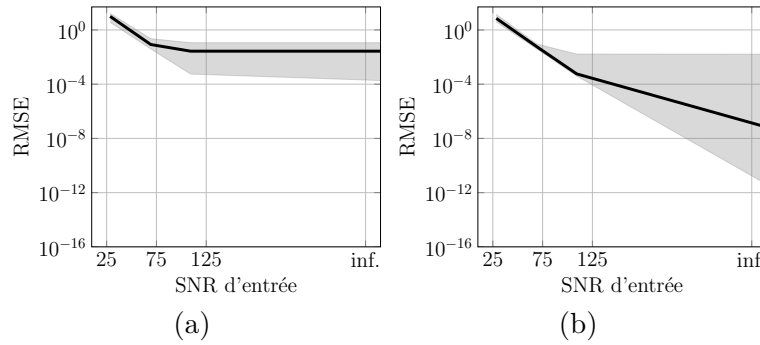


FIGURE 6.16 – Performances par rapport au SNR d'entrée ($\rho_{DM} = 0.9, \rho_{RV} = 0.3$) pour (a) SC-ISN-Cal, (c) ASC-ISN-Cal.

naturellement les performances de SC-ISN-Cal s'améliorent légèrement avec l'augmentation du nombre de rendez-vous. En effet, si la valeur médiane du RMSE est stabilisée autour de $1.e - 3$ pour $\rho_{RV} > 0$, l'enveloppe tend à se déplacer vers des valeurs plus basses, principalement pour la partie inférieure. Dans la figure 6.15b, nous constatons qu'avec ASC-ISN-Cal, la proportion de rendez-vous avec des capteurs étalonnés n'a plus d'impact sur les performances obtenues. La valeur médiane tout comme l'enveloppe sont alors bien moins élevées qu'avec SC-ISN-Cal, ce qui justifie l'intérêt de cette approche avec régularisation. Notons toutefois qu'ici encore, les paramètres théoriques choisis pour la régularisation correspondent aux valeurs exactes des moyennes des paramètres. Ces valeurs sont inaccessibles en pratique et ne sont qu'approximativement connues, une erreur est alors présente et influence l'étalonnage de la même façon que présenté précédemment pour le modèle affine. La valeur du paramètre de régularisation est donc à choisir avec attention.

Sensibilité au bruit

Pour tester la sensibilité au bruit des approches SC-ISN-Cal et ASC-ISN-Cal, nous fixons la proportion de données manquantes ρ_{DM} à 90% et la proportion de rendez-vous ρ_{RV} à 30%. Nous faisons alors varier le SNR d'entrée entre ∞ (pas de bruit) et 25 dB par ajout d'un bruit

gaussien. Pour chaque valeur de SNR, 40 tests sont effectués et les performances tracées dans la figure 6.16a pour SC-ISN-Cal et 6.16b pour ASC-ISN-Cal. Les performances obtenues sont attendues, le RMSE évolue de la même façon qu’avec les méthodes IN-Cal et ACIN-Cal, à ceci près que lorsqu’il n’y a pas de bruit ses valeurs sont plus élevées que dans le cas affine. Comme expliqué précédemment, cette chute de performances est notamment due au fait que le problème à résoudre est plus complexe mais qu’aucune connaissance, contrainte ou donnée supplémentaire n’est ajoutée pour faciliter l’étalonnage.

6.4 Conclusion

Dans ce chapitre, nous avons étudié les performances obtenues par les méthodes d’étalonnage proposées précédemment pour des capteurs à réponse affine ainsi que pour des capteurs à réponse non-linéaire. Les approches testées tiennent notamment compte des contraintes liées aux campagnes de mesures par foule de capteurs mobiles, c.-à-d. des données éparses donnant lieu à des entrées manquantes dans la matrice X , et un faible taux de croisement entre les capteurs mobiles à étalonner et les références. En particulier, elles montrent des performances intéressantes dans des conditions difficiles, où peu de mesures étalonnées (4 dans chaque simulation) sont mêlées aux mesures des capteurs mobiles. Dans ces conditions, les approches classiques d’étalonnage de la littérature telles que les approches multi-sauts [183, 182] ne sont pas applicables puisque nécessitant généralement une plus grande densité de mesures afin d’assurer une grande quantité de rendez-vous entre l’ensemble des capteurs du réseau.

Toutefois, les méthodes proposées restent sensibles lorsqu’une proportion trop importante de données sont manquantes, ne permettant alors plus de lier l’ensemble du réseau. Par ailleurs, lorsque l’étalonnage se déroule pour le mieux la valeur du phénomène physique est estimée précisément en chaque point de mesure. Une étape est ensuite nécessaire pour interpoler le champ de pollution et ainsi reconstruire une carte complète. Dans la partie suivante, nous proposons d’introduire dans les approches d’étalonnage des méthodes d’interpolation qui permettront à la fois de connecter l’ensemble du réseau de capteurs dans les conditions les moins favorables à l’étalonnage et de permettre de mettre en place une cartographie du champ partiellement mesuré par les capteurs.

Deuxième partie

De l'étalonnage à la reconstruction de champs physiques

Raffinement de l'étalonnage et extension à la cartographie par ajout d'une régularisation spatiale

Les approches présentées dans la première partie de ce manuscrit montrent des performances globales assez intéressantes dans des situations d'étalonnage relativement difficiles comptant peu de rendez-vous entre les capteurs. Cependant, toutes ces méthodes restent assez sensibles dans les cas extrêmes où la trop grande proportion de données manquantes ainsi que la trop faible proportion de rendez-vous ne permettent plus de vérifier les hypothèses nécessaires à l'étalonnage, c.-à-d. que certains capteurs sont totalement isolés ou ne sont plus liés, même indirectement, à suffisamment de références. En *mobile crowdsensing* classique où les capteurs sont portés par des humains, les déplacements sont non-contraints et ces cas extrêmes comportant beaucoup de données manquantes et peu de rendez-vous sont amenés à être assez courants. De plus, toutes ces approches sont basées sur la construction de la matrice de données X à partir d'une scène \mathcal{S} discrétisée suivant un maillage adapté au phénomène physique observé. En pratique, le choix du paramètre Δ_d est difficile puisqu'il est dépendant à la fois du phénomène physique observé et des conditions de mesure. Un mauvais choix de Δ_d entraîne alors une mauvaise discrétisation de la scène qui induit à son tour un biais dans l'étalonnage. La robustesse des méthodes d'étalonnage proposées précédemment doit donc être renforcée pour permettre d'atteindre de bonnes performances d'estimation des paramètres d'étalonnage dans des conditions extrêmes et en minimisant l'influence du paramètre Δ_d utilisé pour discrétiser la scène.

Par ailleurs, l'objectif final du projet OSCAR dans le cadre duquel ont été menés les travaux présentés dans ce manuscrit est de permettre la production d'une cartographie de la qualité de l'air. L'aspect restauration du champ de pollution observé par le réseau n'a pas encore été développé ici mais il est important de noter qu'avec chacune des méthodes présentées précédemment, lorsque l'étalonnage des capteurs est mené à bien, la matrice des paramètres F est estimée avec précision, tout comme une partie de la matrice G dont la seconde colonne contient la carte discrétisée du phénomène physique observé. Ces méthodes permettent donc une première fusion des mesures produites par chaque capteur ainsi qu'une première cartographie du champ mesuré.

Les deux objectifs à atteindre dans cette seconde partie de manuscrit sont donc d'améliorer la robustesse des méthodes dans le cas extrême de mesures très dispersées en espace, et de

permettre en sortie d'étalonnage de produire une cartographie fine du champ de pollution observé. Nous proposons ici d'atteindre ces objectifs en couplant les méthodes précédentes à des approches d'estimation de données manquantes. Notons que les régularisations proposées dans ce chapitre peuvent être appliquées à toutes les approches proposées dans la première partie de la thèse, c.-à-d. IN-Cal, ACIN-Cal et SC-Isn-Cal. Dans ce chapitre, pour des raisons de concision, nous ne présentons que les extensions de IN-Cal.

7.1 État de l'art en estimation de données manquantes

L'estimation de données manquantes est une problématique qui a rencontré un écho, tant en traitement du signal qu'en mathématiques et en physique. En effet, dans beaucoup d'applications, les données à traiter peuvent être incomplètes. Outre les approches basées sur la complétion de matrices, discutées dans le chapitre 3, de nombreuses approches ont été proposées dans la littérature pour l'estimation de données manquantes dans un signal [1, 17], une image [10, 60], des séries temporelles [159] ou plus généralement une collection de données. Nous proposons ici un bref aperçu de certaines de ces méthodes que nous divisons en plusieurs catégories, c.-à-d. :

1. les approches d'estimation de données discrètes,
2. les techniques d'estimation de données continues,
3. les méthodes d'estimation utilisées par exemple en génie de l'environnement.

7.1.1 Les approches d'estimation de données discrètes

Dans cette section, nous considérons le même formalisme que dans la partie précédente de la thèse, c.-à-d. que le phénomène observé est discrétisé sur une grille de dimension finie, générant ainsi des vecteurs finis et éventuellement très "longs". Nous considérons le cas d'un signal y observé en q points (possiblement multi-dimensionnels) $\{\mathbf{p}_1, \dots, \mathbf{p}_q\}$ parmi m disponibles et dont nous cherchons à estimer la valeur y_{q+1} en un $(q + 1)$ -ème point \mathbf{p}_{q+1} , sans ajout d'*a priori* sur la nature du signal observé. Nous définissons donc $\mathbf{y} \triangleq [y_1 \dots y_q]^T$ le vecteur des observations aux points $\{\mathbf{p}_1 \dots \mathbf{p}_q\}$.

7.1.1.1 Interpolation aux K plus proches voisins (KNN)

Les K -plus proches voisins (ou KNN pour *K-Nearest Neighbor* en anglais) est une méthode très simple utilisée en classification qui consiste à attribuer à un objet la classe la plus représentée parmi ses K plus proches voisins [115]. La méthode des KNN est également utilisée en interpolation de données [207] et consiste, pour estimer la valeur d'un signal en un point \mathbf{p} à interpoler, à lui attribuer comme valeur celle la plus largement représentée parmi les K points d'observation présents dans son voisinage. Cette méthode, très simple, ne peut fonctionner correctement que si le choix du nombre de voisins K a été correctement effectué.

Par ailleurs, le nombre de voisins autour d'une valeur manquante doit être relativement élevé pour que l'interpolation soit convenablement réalisée. Aussi, ce type d'approche n'est *a priori* pas adapté au problème considéré et ne sera pas étudié par la suite.

7.1.1.2 Approche barycentrique d'estimation de donnée

L'interpolation barycentrique – ou combinaison convexe – légèrement différente de l'approche KNN, consiste à estimer la valeur y_{q+1} en un point \mathbf{p}_{q+1} comme une combinaison linéaire des points voisins, c.-à-d.

$$y_{q+1} = \sum_{i=1}^q w_i \cdot y_i, \quad (7.1)$$

où $\{w_1, \dots, w_k\}$ sont les poids associés à chaque élément de la combinaison linéaire. En interpolation barycentrique, le poids w_i associé à point de mesure \mathbf{p}_i est souvent fonction de sa distance au point d'interpolation. La formule la plus courante pour le calcul des poids est

$$w_i = \frac{\frac{1}{d_i}^\alpha}{\sum_{j=1}^q \frac{1}{d_j}^\alpha}, \quad (7.2)$$

où d_i correspond à la distance euclidienne entre le i -ème point d'observation et le point d'interpolation, et le paramètre α est généralement fixé à 1 ou 2. Cette formule de calcul des poids assure que $\sum_i w_i = 1$. Il est à noter que lorsque la distance euclidienne entre deux points est importante, le poids w_i associé est négligeable. Les calculs peuvent être accélérés en les limitant à la zone de voisinage du point à interpoler. D'autres fonctions de calcul des poids peuvent également être utilisées [43].

L'interpolation barycentrique est une approche d'interpolation dont la mise en place est très simple, cependant ses performances sont relativement limitées, une valeur interpolée ne pourra pas par exemple dépasser la valeur maximale atteinte par les observations environnantes. Dans l'application visée ici, nous avons besoin d'une approche d'interpolation permettant une plus grande souplesse, c.-à-d. autorisant un comportement plus libre du phénomène physique reconstruit et ne limitant pas ainsi les valeurs estimées.

7.1.1.3 Estimation de données par *a priori* de parcimonie

Dans de nombreuses applications, il est possible de considérer qu'un signal \mathbf{y} peut être écrit comme une combinaison parcimonieuse des atomes d'un dictionnaire D donné. C'est à dire que si $\mathbf{y} \in \mathbb{R}^{m \times 1}$ est le signal considéré et que $D \in \mathbb{R}^{m \times d}$ est le dictionnaire, alors il existe un vecteur $\mathbf{a} \in \mathbb{R}^{d \times 1}$ dont un nombre $k \ll d$ de composantes sont non-nulles et tel que

$$\mathbf{y} \approx D \cdot \mathbf{a}. \quad (7.3)$$

Cette propriété de décomposition parcimonieuse d'un signal suivant un dictionnaire donné peut permettre par exemple de débruiter/restaurer un signal ou une image [59, 137], de faire de la classification ou de l'identification de formes [95, 226].

En considérant maintenant que le signal observé \mathbf{y}_{obs} est une version "à trous" – c.-à-d. avec des entrées manquantes – de \mathbf{y} , celui-ci peut s'écrire comme

$$\mathbf{y}_{obs} = E \cdot \mathbf{y} \quad (7.4)$$

$$\approx E \cdot D \cdot \mathbf{a}, \quad (7.5)$$

où E est la matrice d'échantillonnage diagonale dont un élément $e_{i,i}$ est égal à 0 ou 1. Ainsi, si $e_{i,i} = 0$ alors la i -ème donnée de \mathbf{y} est manquante dans \mathbf{y}_{obs} . Il est à noter que le problème (7.5) présente des similitudes avec l'acquisition compressée. En effet, définissons \mathbf{y}'_{obs} le vecteur de taille q concaténant toutes les valeurs connues dans \mathbf{y}_{obs} . Alors (7.5) s'écrit sous la forme (2.13) où Ψ est une matrice binaire $m \times q$ telle que

$$\mathbf{y}'_{obs} = \Psi \cdot \mathbf{y}. \quad (7.6)$$

Le problème d'estimation manquante peut alors être vu comme un problème d'acquisition compressé particulier où la matrice Ψ n'est plus un opérateur de mélange aléatoire mais une matrice extraite de E . Cependant, les outils de résolution de (2.14) restent valables et la plupart des approches sont basées sur des techniques de relaxation convexe de la parcimonie – comme par exemple l'approche LASSO [205] faisant appel à la norme ℓ_1 – ou des techniques gloutonnes, comme par exemple OMP (*Orthogonal Matching Pursuit* en anglais) [154, 45].

Ces techniques d'estimation de données manquantes ont été principalement utilisées en traitement d'images, où la problématique s'appelle couramment *inpainting* [10]. Dans [59], les auteurs proposent une telle technique, basée sur la parcimonie, où le dictionnaire regroupe plusieurs transformées parcimonieuses, chacune permettant la reconstruction des contours ou de la texture. Dans [17], les auteurs estiment un signal en le supposant parcimonieux dans le domaine de Fourier, pour l'observation astrophysique. Cette même idée a été considérée dans [1] pour la restauration de signaux audiophoniques.

Il est à noter que pour que ces approches soient efficaces, notamment sur des problèmes de grande dimension, il est nécessaire :

- de connaître un dictionnaire bien adapté au problème considéré,
- d'avoir des atomes relativement peu corrélés.

Dans le cas où les atomes sont corrélés, les approches classiques d'approximation parcimonieuse ne fonctionnent pas et le problème parcimonieux ne peut alors être résolu que par optimisation combinatoire exacte, connue pour être coûteuse et seulement adaptée aux problèmes de petite taille [18].

7.1.2 Les approches basées sur l'analyse de données fonctionnelles

Lorsque la taille du vecteur \mathbf{y} est importante, il peut être plus intéressant de considérer les données comme étant de longueur infinie et discrétisant une fonction à estimer. Nous supposons alors que nous observons q points $\{y_1, \dots, y_q\}$ échantillonnant de manière irrégulière, en les points $\{\mathbf{p}_1, \dots, \mathbf{p}_q\}$, une fonction \mathcal{Y} à estimer. Cette vision inclut notamment l'estimation paramétrique de fonctions (régression linéaire ou polynômiale) et la régression de données fonctionnelles, un sous-domaine de l'analyse de données fonctionnelles (FDA pour *Functional Data Analysis* en anglais) [169].

7.1.2.1 Estimation de données par régression

Nous supposons ici que la fonction \mathcal{Y} est gouvernée par un nombre réduit de paramètres à estimer en supposant que, pour tout $i = 1, \dots, q$,

$$y_i = \mathcal{Y}(\mathbf{p}_i). \quad (7.7)$$

Le modèle le plus connu de régression est la régression linéaire [71] qui cherche à lier les variables par une relation linéaire ou affine. Ainsi, dans le cas où les points d'observation sont bi-dimensionnels, c.-à-d. qu'un point $\mathbf{p}_i = (p_{i,1}, p_{i,2})$, en définissant la matrice P comme

$$P = \begin{bmatrix} 1 & p_{1,1} & p_{1,2} \\ \vdots & \vdots & \vdots \\ 1 & p_{q,1} & p_{q,2} \end{bmatrix}, \quad (7.8)$$

la régression linéaire consiste à estimer le vecteur de paramètres $\mathbf{a} = [\beta, \alpha_1, \alpha_2]^T$ tel que $(P \cdot \mathbf{a})$ soit le plus proche possible de \mathbf{y}_{obs} au sens des moindres carrés. Le vecteur \mathbf{a} est alors défini par l'équation $\mathbf{a} = (P^T P)^{-1} P^T \mathbf{y}$. Ainsi, la valeur de la fonction \mathcal{Y} en un nouveau point $\mathbf{p}_{q+1} = (p_{1,q+1}, p_{2,q+1})$ peut être estimée par l'application de la relation

$$\mathcal{Y}(\mathbf{p}_{q+1}) = \beta + \alpha_1 \cdot p_{1,q+1} + \alpha_2 \cdot p_{2,q+1}. \quad (7.9)$$

Cette approche de régression linéaire par moindres carrés est la plus classique et facile à mettre en œuvre, toutefois les profils de fonctions reconstruits sont limités à des droites dans le cas mono-dimensionnel, et à des plans en dimensions plus élevées.

La régression polynômiale consiste à rechercher le vecteur de paramètres \mathbf{a} qui permet de lier le vecteur des observations \mathbf{y} aux points \mathbf{p}_i non plus par une relation affine mais par une expression polynômiale. En mono-dimensionnel, cette approche peut être vue comme une extension immédiate de la régression linéaire qui consiste donc à rechercher les coefficients $\{a_0, \dots, a_N\}$ tels que chaque observation $\mathcal{Y}(\mathbf{p}_i)$ puisse s'écrire comme une fonction polynômiale de son point d'observation, c.-à-d. que $\mathcal{Y}(p_i) \approx \sum_{l=0}^N a_l p_i^l$. Cette régression peut être

résolue par moindres carrés en considérant la matrice

$$P = \begin{bmatrix} 1 & p_1 & \cdots & p_1^N \\ \vdots & \vdots & & \vdots \\ 1 & p_q & \cdots & p_q^N \end{bmatrix}. \quad (7.10)$$

L'inconvénient de ce modèle de régression réside dans le fait que le degré du polynôme recherché est fixé et qu'il doit être sélectionné précisément pour permettre de bien approximer les données observées. Il est important de noter que ces régressions peuvent être étendues trivialement à des dimensions supérieures.

7.1.2.2 Interpolation par calage de fonctions

En analyse de données fonctionnelles [169], une alternative à la régression paramétrique ci-dessus consiste à supposer que la fonction \mathcal{Y} à estimer peut s'écrire comme une combinaison linéaire de fonctions Φ_i issues d'une base plus simple à calculer, c.-à-d.

$$\mathcal{Y}(\mathbf{p}) = \sum_{i=1}^k c_i \Phi_i(\mathbf{p}). \quad (7.11)$$

L'intérêt de ce type de représentation est la compacité de description de \mathcal{Y} puisque une fois la base choisie, on peut représenter la fonction \mathcal{Y} par le vecteur de pondération $\mathbf{c} \triangleq [c_1, \dots, c_k]^T$, d'où son intérêt pour la classification [169] et par extension, pour certaines approches de séparation de sources [167, 166]. Les bases de fonctions les plus classiques sont les séries de Fourier – lorsque les données à estimer sont périodiques – ou les (bases de) splines pour les autres cas [169]. Dans la suite, nous ne nous intéressons qu'à ces dernières :

- Splines : la méthode des splines consiste à ajuster sur les données des courbes dont l'énergie de flexion entre deux points de mesure est minimale [57, 215]. Deux variantes des splines existent, d'un côté une approche exacte dont la courbe d'interpolation est contrainte à passer exactement par les points de mesure, d'un autre côté une approche relâchée qui permet un compromis entre l'attache aux données disponibles et le "degré de lissage" de la courbe obtenue.
- Les fonctions de base radiales : cette approche consiste à rechercher une courbe collant au mieux aux données observées et qui soit combinaison linéaire de fonctions d'une base prédéfinie. Différentes familles de bases sont considérées, la base exponentielle, multiquadratique, multiquadratique inverse, spline cubique ou spline plaque mince. Ces bases et leur utilisations sont détaillées dans la thèse d'Adriana Coman [42]. Une des difficultés qui peuvent être rencontrées avec ce type d'approches est le choix du nombre de nœuds et des degrés des polynômes constituant la base [166]. En effet, la régression ne peut être que possible que si, pour un degré de spline donnée, le nombre de points observés entre deux nœuds est suffisamment élevé. Le positionnement optimal des nœuds est donc dépendant de l'échantillonnage du phénomène observé.

7.1.3 Méthodes d'estimation de données à base de modèle

L'estimation de champs spatiaux ayant un sens physique à partir de données irrégulièrement échantillonnées a été aussi très étudié par les physiciens, qui ont proposé des modèles physiques à cet effet. Cependant, ces modèles peuvent fortement dépendre de la zone étudiée. Jerrett *et al.* ont par exemple parcouru l'ensemble des modèles utilisés pour la modélisation de la qualité de l'air en zone urbaine [98]. Ces types de modèle ont aussi été étudiés en traitement du signal et en informatique, principalement sous forme de problème inverse/déconvolution pour de la localisation de sources en 1D [46] ou 2D [148, 147], l'estimation de débit de sources [172, 139] ou le placement optimal/sélection de capteurs [15, 174, 176].

7.1.3.1 Méthodes basées sur la mécanique des fluides

Les modèles CFD (pour *Computational Fluid Dynamics* en anglais) permettent une résolution numérique des équations de la mécanique des fluides – c.-à-d. les équations d'Euler et de Navier-Stokes [164]– et sont ainsi utilisables dans un grand nombre d'applications, comme par exemple la qualité de l'air [180], l'épandage de gaz liquide [138], ou l'industrie aéronautique [164]. Ces approches sont capables de prendre en compte des sources multiples et leurs activités – par exemple, des débits variables – associées. Les méthodes utilisant les modèles CFD peuvent approcher de manière précise la réalité du terrain et tenir compte de différents facteurs intervenant dans le phénomène modélisé (écoulements fluidiques laminaires ou turbulents, effets thermiques, présence d'obstacles et de reliefs, etc). Cependant, ces modèles peuvent être assez complexes à mettre en œuvre (choix de la discrétisation spatiale, choix de la méthode de résolution) et sont généralement très coûteux en temps de calcul.

Aussi, coupler ces techniques aux approches IN-Cal et ACIN-Cal (et leurs extensions non-linéaires) – itératives et nécessitant donc potentiellement plusieurs ré-estimations du champ physique – n'est donc pas *a priori* un choix judicieux si on veut conserver un temps de calcul raisonnable.

7.1.3.2 Les modèles intégraux

Les modèles intégraux sont basés sur la résolution simplifiée des équations de la mécanique des fluides. Contrairement aux approches basées sur les modèles CFD, les modèles intégraux ne permettent pas de tenir compte des obstacles et du relief. Cependant, ils introduisent des paramètres supplémentaires – par exemple un registre d'émission des polluants – pour pallier aux phénomènes non-représentés par les équations simplifiées de la mécanique des fluides et tiennent compte de la turbulence atmosphérique via des classes de stabilité atmosphérique, évitant ainsi une modélisation lourde de la turbulence. Ce type de modèle – comme ADMS [180] par exemple – est actuellement utilisé par certaines AASQA pour modéliser la qualité de l'air à l'échelle urbaine [206].

Par contre, les modèles intégraux ont été proposés pour des réseaux de capteurs fixes

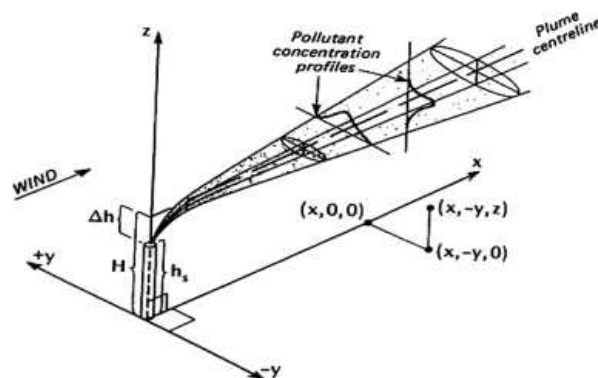


FIGURE 7.1 – Exemple de diffusion obtenue avec le modèle plume [209].

et très parcimonieux dans l'espace. Ils ne sont donc pas actuellement adaptés aux grandes quantités de données fournies par des réseaux mobiles. Par ailleurs, l'estimation du registre d'émission est une tâche fastidieuse qui doit être réalisée sur chaque site étudié. Ce type de méthodes n'est donc pas a priori adapté aux données issues du *mobile crowdsensing*.

7.1.3.3 Le modèle analytique gaussien

Le modèle analytique gaussien [209] permet de simuler, à partir de positions connues de sources ponctuelles de pollution – ayant chacune un débit constant connu – et d'un champ de vent (direction et force du vent) la dispersion d'un polluant dans l'air jusqu'à un site voulu. La dispersion est estimée en suivant un étalement gaussien dont l'orientation est influencée par le champ de vent, comme illustré dans la figure 7.1. Ce modèle, assez simple, a été notamment couplé à la NMF dans [162] pour estimer les contributions nulles de sources dans un problème de séparation de sources. Il a aussi été utilisé dans [15] dans un problème d'optimisation de l'observatoire. Cependant, dans le cadre du projet OSCAR, nous ne disposons pas de champs de vent *locaux* mais uniquement globaux. Or, la trajectoire du vent peut être assez complexe à modéliser en milieu urbain [174]. Ce type de modèle ne sera donc pas étudié dans la suite du manuscrit.

7.1.3.4 Modèle LUR

Le modèle LUR (pour *Land Use Regression* en anglais) cherche à prédire la valeur d'un phénomène physique en un point d'interpolation \mathbf{p} en se basant sur une connaissance des caractéristiques du terrain environnant ainsi que sur des informations de trafic par exemple. Le modèle LUR a été utilisé dans d'autres projets d'estimation de la qualité de l'air à partir d'un réseau mobile de capteurs [86]. Cependant, dans le cadre du projet OSCAR, nous ne disposons pas d'estimation du flot de circulation et ne pouvons donc enrichir le modèle pour obtenir une modélisation précise.

7.1.4 Les approches géostatistiques

Dans les approches d'interpolations géostatistiques, le phénomène physique étudié est modélisé par une fonction aléatoire $\mathcal{Y}(\mathbf{p})$ où \mathbf{p} est point de l'espace géographique. Les différentes variables aléatoires indexées par le processus aléatoire $\mathcal{Y}(\mathbf{p})$ sont corrélées et c'est à partir des corrélations entre les points d'observations que les propriétés de dispersion spatiale du phénomène physique sont alors déduites et utilisées pour l'interpoler.

Appréhension globale du krigeage

Le krigeage fait partie de la famille des méthodes géostatistiques pour l'interpolation de données. Meilleur outil d'interpolation linéaire [75], le krigeage consiste à apprendre à partir des observations disponibles le comportement du phénomène physique d'intérêt avant de l'étendre aux points d'interpolations désirés. Plusieurs méthodes de krigeage sont étudiées dans la littérature :

- le krigeage simple, dans lequel le phénomène physique observé est modélisé par la fonction aléatoire

$$\mathcal{Y}(\mathbf{p}) = \mu + \delta(\mathbf{p}), \quad \mathbf{p} \in D, \quad (7.12)$$

où μ est une constante connue, $\delta(\cdot)$ est une fonction aléatoire stationnaire de second ordre centrée et D est le domaine d'étude du phénomène physique.

- le krigeage ordinaire, dans lequel le phénomène physique suit le même modèle que le krigeage simple, avec une constante μ inconnue.
- le krigeage universel, dans lequel la fonction aléatoire $\mathcal{Y}(\mathbf{p})$ n'est plus stationnaire.

Quelque soit le type de krigeage choisi, la méthode employée pour arriver aux formules d'interpolation est la même et repose sur les quatre contraintes suivantes :

- C1. **Linéarité** : chaque estimation du phénomène physique en un nouveau point \mathbf{p}_{q+1} doit être combinaison linéaire de l'ensemble des q observations $\mathcal{Y}(\mathbf{p}_1), \dots, \mathcal{Y}(\mathbf{p}_q)$ disponibles, c.-à-d.

$$\tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) = a + \sum_{i=1}^q w_i \cdot \mathcal{Y}(\mathbf{p}_i), \quad (7.13)$$

avec a et w_i des paramètres à déterminer.

- C2. **Autorisation** : l'interpolation par krigeage est réalisée en minimisant l'erreur d'estimation. Cette dernière doit donc admettre un moment d'ordre 2.

- C3. **Non-biais** : le krigeage est une méthode d'interpolation sans biais. L'erreur d'estimation doit être d'espérance nulle, c.-à-d. $\mathbb{E}[\tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) - \mathcal{Y}(\mathbf{p}_{q+1})] = 0$.

- C4. **Optimalité** : la variance de l'erreur d'estimation doit être minimale, c.-à-d. que $\tilde{\mathcal{Y}}(\mathbf{p}_{q+1})$ minimise

$$\text{Var}(\tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) - \mathcal{Y}(\mathbf{p}_{q+1})). \quad (7.14)$$

Les développements nécessaires à l'écriture des contraintes ci-dessus sont détaillés dans l'article de Chauvet [28]. Suivant les hypothèses de krigeage simple, ordinaire ou universel, la vérification des contraintes C1 à C4 mène aux formules d'interpolation. Dans cette sous-section, nous présentons l'approche d'interpolation par krigeage ordinaire, qui se trouve être

la technique la plus utilisée [75]. Nous reprenons les notations de la thèse de Coman [42] dans laquelle sont présentées plus en détails les différentes approches de krigeage.

Le krigeage ordinaire

Comme évoqué précédemment, pour arriver à l'écriture de la formule d'interpolation par krigeage, les hypothèses C1 à C4 ci-dessus doivent être vérifiées une à une. Nous nous plaçons ici dans le cas du krigeage ordinaire et modélisons donc le phénomène physique étudié par la fonction aléatoire

$$\mathcal{Y}(\mathbf{p}) = \mu + \delta(\mathbf{p}), \quad \mathbf{p} \in D, \quad (7.15)$$

où μ est une constante inconnue, $\delta(\cdot)$ est une fonction aléatoire stationnaire de second ordre centrée et D est le domaine d'étude du phénomène physique.

Dans le cas du krigeage ordinaire, la vérification des contraintes mène aux résultats suivants :

(C1) $\Rightarrow \tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) = a + \sum_{i=1}^q w_i \cdot \mathcal{Y}(\mathbf{p}_i) = a + \mathbf{w}^T \cdot \hat{\mathbf{y}}$, où \mathbf{w} est le vecteur des poids w_i et $\hat{\mathbf{y}}$ est le vecteur des $\mathcal{Y}(\mathbf{p}_i)$.

(C2) $\Rightarrow \sum w_i = 1$.

(C3) $\Rightarrow a = 0$ et donc $\tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) = \mathbf{w}^T \cdot \hat{\mathbf{y}}$.

(C4) \Rightarrow Le vecteur minimisant la variance $\text{Var}(\tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) - \mathcal{Y}(\mathbf{p}_{q+1}))$ est solution de

$$\begin{bmatrix} \gamma_{1,1} & \dots & \gamma_{1,q} & 1 \\ \vdots & & & \vdots \\ \gamma_{q,1} & \dots & \gamma_{q,q} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_q \\ \nu \end{bmatrix} = \begin{bmatrix} \gamma_{q+1,1} \\ \vdots \\ \gamma_{q+1,q} \\ 1 \end{bmatrix}, \quad (7.16)$$

où $\gamma_{i,j} = \frac{1}{2}(\mathcal{Y}(p_i) - \mathcal{Y}(p_j))^2$.

Afin de pouvoir procéder à une interpolation par krigeage ordinaire en un point \mathbf{p}_{q+1} , il faut donc déterminer la valeurs des poids w_i par résolution du problème (7.16) obtenu par vérification de la contrainte C4. Or, si les valeurs $\gamma_{i,j}$ sont connues pour tout couple $(i, j) \in \{1, \dots, q\}^2$, les valeurs $\gamma_{q+1,i}$ sont elles inconnues et leur estimation passe par l'étude du variogramme qui constitue un point clé du krigeage.

Analyse de variogramme : Le variogramme, calculé à partir des observations du phénomène physique et de leur localisation permet de représenter la structure spatiale du champ observé. Il s'agit de calculer pour chaque couple d'observations $(\mathcal{Y}(\mathbf{p}_i), \mathcal{Y}(\mathbf{p}_j))$ la distance h_{ij} entre les emplacements des mesures ainsi que la différence γ_{ij} entre les valeurs observées. Le variogramme est alors le tracé de toutes les valeurs γ en fonction des distances h , son allure dépend des fluctuations du phénomène physique mesuré, comme illustré dans la figure 7.2.

Les valeurs $\gamma_{q+1,i}$ nécessaires à la résolution du problème (7.16) sont finalement estimées à partir du variogramme associé au phénomène physique observé. Une fois ces paramètres obtenus, l'estimation des poids w_i est triviale et peut être effectuée par exemple par moindres

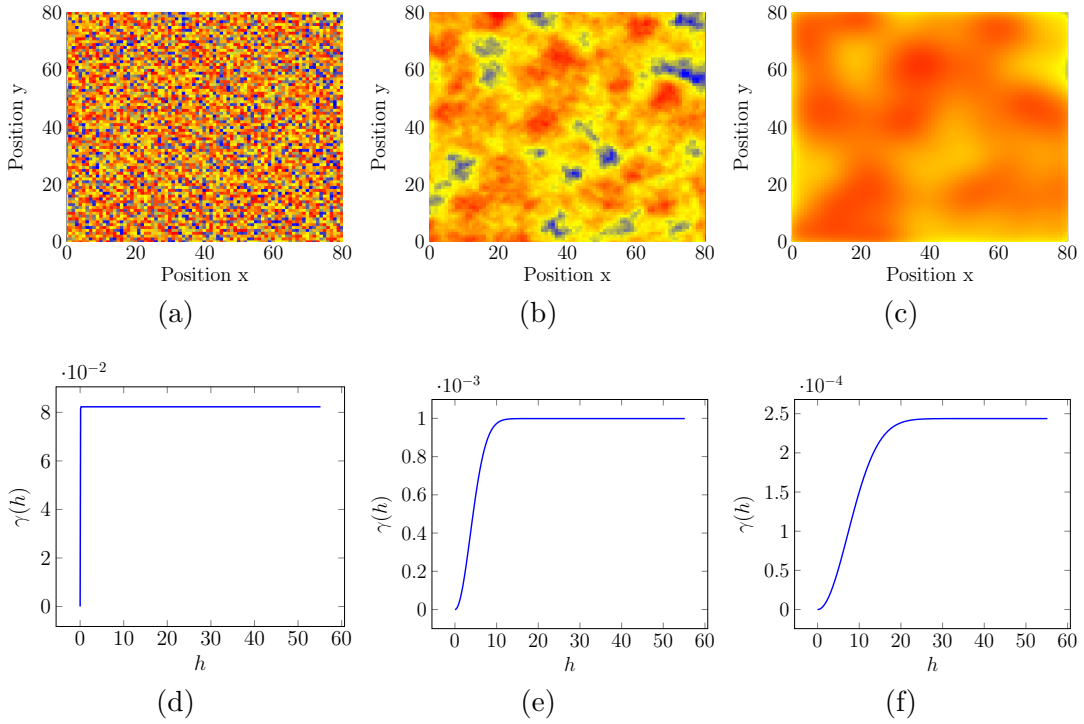


FIGURE 7.2 – (d-e-f) Exemples de variogrammes obtenus pour les champs simulés (a-b-c).

carrés. L'interpolation de la valeur au site \mathbf{p}_{q+1} est ensuite immédiate par la formule

$$\tilde{\mathcal{Y}}(\mathbf{p}_{q+1}) = \mathbf{w}^T \cdot \hat{\mathbf{y}}. \quad (7.17)$$

Par ailleurs, le krigeage permet également d'évaluer la variance autour du point interpolé, ce qui donne un indice de confiance dans l'estimation. Cette variance est calculée par la formule

$$\sigma(\mathbf{p}_{q+1}) = \sum_{i=1}^q w_i \cdot \gamma_{q+1,i}. \quad (7.18)$$

Notons que les approches d'interpolation par calage de fonctions, approximation barycentrique ou krigeage sont toutes basées sur une idée commune d'estimer la valeur en un nouveau point \mathbf{p}_{q+1} par un mélange pondéré des observation environnantes. Seule varie d'une approche à une autre la méthode d'estimation des paramètres de mélange. Dans les méthodes d'interpolation par calage de fonctions, ces paramètres sont en partie contraints par le choix de la base de fonction. Dans les méthodes barycentriques, il s'agit de fixer la valeur du paramètre α intervenant dans les différents poids du modèle, tandis qu'avec le krigeage, tous les paramètres sont automatiquement ajustés par apprentissage à partir des points d'observation et une mesure de confiance associée aux points interpolés est aussi fournie.

7.2 Étalonnage par factorisation matricielle avec contrainte de décomposition parcimonieuse

Dans cette section, nous proposons de régulariser les méthodes d'étalonnage présentées précédemment en supposant que le phénomène physique observé par le réseau de capteurs mobiles admet une décomposition parcimonieuse suivant un dictionnaire D connu. Cette nouvelle hypothèse ajoute aux problèmes de factorisation matricielle présentés précédemment une contrainte supplémentaire permettant un relâchement de la définition 2.3 de rendez-vous page 12. Par ailleurs, l'utilisation du dictionnaire permettra également d'affiner la cartographie grossière produite par les méthodes d'étalonnage précédentes. La méthode proposée ci-dessous a été publiée dans [51, 53].

Ajout d'une nouvelle hypothèse à l'étalonnage

Pour rappel, le problème de factorisation matricielle considéré dans la méthode IN-Cal consistait originalement à chercher les matrices G et F minimisant

$$\begin{aligned} \{G, F\} = \arg \min_{\tilde{G}, \tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{G} = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \\ \tilde{F} = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F. \end{aligned} \quad (7.19)$$

La seconde colonne \mathbf{g}_2 de la matrice G devant contenir les valeurs \mathbf{y} du phénomène physique observé par les capteurs. Nous supposons que ce phénomène admet une décomposition parcimonieuse suivant un dictionnaire D connu, c.-à-d. qu'il existe un vecteur \mathbf{a} parcimonieux tel que $\mathbf{y} \approx D \cdot \mathbf{a}$. Cette hypothèse semble tout à fait réaliste : en effet, dans le cas de la mesure de la qualité de l'air, le dictionnaire D peut être appris à partir de différentes réalisations d'un modèle de simulation – par exemple un modèle discuté dans la section 7.1.3.1, comme le modèle analytique gaussien [200] – permettant de générer différentes "formes" possibles du champ physique. Le vecteur \mathbf{y} est ensuite vu comme une combinaison de quelques formes caractéristiques contenues dans D . Cette hypothèse de décomposition parcimonieuse s'appliquant à \mathbf{y} , elle s'applique également à \mathbf{g}_2 , à ceci près qu'avec les données manquantes dans X , \mathbf{g}_2 est une version potentiellement à trous de \mathbf{y} . La décomposition parcimonieuse suivant le dictionnaire permet donc d'approximer les entrées de \mathbf{g}_2 correspondant aux cellules de la scène qui ont été visitées par au moins un capteur. Autrement dit,

$$\omega \circ \mathbf{g}_2 \approx \omega \circ (D \cdot \mathbf{a}), \quad (7.20)$$

où ω est le vecteur binaire de poids ω_i assurant la présence d'une mesure par un capteur de la cellule i , c.-à-d.

$$\omega_i = \begin{cases} 0 & \text{si } \sum_j W_{ij} = 0 \\ 1 & \text{sinon} \end{cases} \quad (7.21)$$

L'équation 7.20 peut alors être insérée dans le problème de factorisation (7.19) sous la forme d'une nouvelle contrainte donnant

$$\begin{aligned} \{G, F, \mathbf{a}\} &= \arg \min_{\tilde{G}, \tilde{F} \geq 0, \mathbf{a}} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \cdot \left\| \boldsymbol{\omega} \circ (\mathbf{g}_2 - D \cdot \tilde{\mathbf{a}}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{G} &= \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \\ \tilde{F} &= \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F, \\ \|\mathbf{a}\|_0 &\ll k, \end{aligned} \quad (7.22)$$

où λ est le paramètre de régularisation associé à la contrainte de décomposition parcimonieuse de \mathbf{g}_2 . Ce problème diffère des méthodes classiques de NMF parcimonieuse [93, 92, 235, 108] qui consistent à rechercher des matrices G et F creuses. Au contraire, nous cherchons ici une matrice G dense dont une colonne admet une décomposition parcimonieuse. Cette démarche est quelque peu similaire à celle présentée dans [83]. Comme pour les méthodes présentées précédemment, la fonction de coût associée au problème d'optimisation global (7.22) est non-convexe par rapport à G , F et \mathbf{a} . Par ailleurs, la contrainte de parcimonie ne faisant intervenir que le vecteur \mathbf{a} et la matrice G , elle n'influence que les mises à jour de ces variables. Les sous-problèmes que nous devons considérer ici sont donc

$$\begin{aligned} G &= \arg \min_{\tilde{G} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - \tilde{G} \cdot F) \right\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \cdot \left\| \boldsymbol{\omega} \circ (\mathbf{g}_2 - D \cdot \mathbf{a}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{G} &= \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \end{aligned} \quad (7.23)$$

et

$$\begin{aligned} \mathbf{a} &= \arg \min_{\tilde{\mathbf{a}}} \left\| \boldsymbol{\omega} \circ (\mathbf{g}_2 - D \cdot \tilde{\mathbf{a}}) \right\|_{\mathcal{F}}^2, \\ \|\mathbf{a}\|_0 &\ll k. \end{aligned} \quad (7.24)$$

Nous pouvons remarquer que le problème d'optimisation en F n'est pas affecté par la nouvelle contrainte de décomposition parcimonieuse. Aussi, la formule de mise à jour (4.21) de F développée pour la méthode IN-Cal reste applicable ici.

Mise à jour de la matrice G

Pour la résolution de (7.23), c.-à-d. du problème d'optimisation en G , en remarquant simplement que $\mathbf{g}_2 = G \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, le terme de pénalisation dans (7.23) peut s'écrire

$$\frac{\lambda}{2} \cdot \left\| \boldsymbol{\omega} \circ (\mathbf{g}_2 - D \cdot \mathbf{a}) \right\|_{\mathcal{F}}^2 = \left\| (\sqrt{\lambda} \cdot \boldsymbol{\omega}) \circ \left(D \cdot \mathbf{a} - G \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \right\|_{\mathcal{F}}^2. \quad (7.25)$$

En définissant les matrices

$$\mathcal{W} \triangleq [W, \sqrt{\lambda} \cdot \boldsymbol{\omega}], \quad \mathcal{X} \triangleq [X, D \cdot \mathbf{a}] \quad \text{et} \quad \mathcal{F} \triangleq \left[F, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right], \quad (7.26)$$

alors il est évident que

$$\frac{1}{2} \cdot \|W \circ (X - G \cdot F)\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \cdot \|\omega \circ (\mathbf{g}_2 - D \cdot \mathbf{a})\|_{\mathcal{F}}^2 = \frac{1}{2} \cdot \|W \circ (\mathcal{X} - G \cdot \mathcal{F})\|_{\mathcal{F}}^2. \quad (7.27)$$

Chercher G solution du problème d'optimisation (7.23) est donc équivalent à chercher G solution du problème

$$\begin{aligned} G &= \arg \min_{\tilde{G} \geq 0} \frac{1}{2} \cdot \left\| W \circ (\mathcal{X} - \tilde{G} \cdot \mathcal{F}) \right\|_{\mathcal{F}}^2, \\ \text{s.c. } \tilde{G} &= \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G. \end{aligned} \quad (7.28)$$

La mise à jour de G avec la contrainte de décomposition parcimonieuse selon D peut donc être effectuée en utilisant la formule (4.22) page 52 développée pour IN-Cal, en considérant dans cette dernière les matrices \mathcal{W} , \mathcal{X} et \mathcal{F} respectivement en lieu et place des matrices W , X et F .

Mise à jour du vecteur \mathbf{a}

Pour la mise à jour du vecteur \mathbf{a} , le problème d'optimisation associé (7.24) vise à estimer la meilleure décomposition parcimonieuse de \mathbf{g}_2 . Cette estimation peut être effectuée de différentes manières. Le problème (7.24) étant NP-difficile, de nombreuses alternatives ont été proposées, comme par exemple sa relaxation convexe, c.-à-d.

$$\mathbf{a} = \arg \min_{\tilde{\mathbf{a}}} \|\omega \circ (\mathbf{g}_2 - D \cdot \tilde{\mathbf{a}})\|_{\mathcal{F}}^2 + \mu \cdot \|\tilde{\mathbf{a}}\|_1, \quad (7.29)$$

où μ est le poids associé à la contrainte de parcimonie de \mathbf{a} , c.-à-d. que plus μ est élevé, plus le vecteur solution trouvé sera proche de zéro. L'équation (7.29) est l'expression d'un problème pondéré de LASSO ou de *Basis Pursuit Denoising* pour lesquels de nombreuses approches ont été développées [235]. Le problème pondéré de LASSO est bien souvent très sensible au choix du paramètre μ , dont la valeur optimale dépend de la variance du bruit additif. Une alternative à LASSO et ses dérivées consiste à utiliser des méthodes gloutonnes comme l'OMP [154] présentée précédemment. Pour rappel, cette méthode consiste à sélectionner dans le dictionnaire D l'atome ayant la meilleure corrélation avec le signal à approximer, à projeter le signal sur l'espace orthogonal à cet atome puis à réitérer ces deux étapes jusqu'à sélection du nombre souhaité d'atomes. La version pondérée d'OMP sera utilisée pour la mise à jour du vecteur \mathbf{a} dans la méthode d'étalonnage par NMF informée avec contrainte de parcimonie développée ici et que nous appellerons désormais SpIN-Cal (pour *Sparse IN-Cal*). L'ajout de cette contrainte de parcimonie permet de relâcher le besoin de nombreux rendez-vous exacts rencontrés dans la méthode IN-Cal. En effet, en utilisant le dictionnaire D , deux capteurs sont en rendez-vous virtuels lorsqu'ils effectuent chacun une mesure sur le même support atomique. La connexion entre deux capteurs est alors assurée grâce à la contrainte $\mathbf{g}_2 \approx D \cdot \mathbf{a}$ qui permet de lier les mesures à travers les atomes du dictionnaire.

L'algorithme de la méthode d'étalonnage SpIN-Cal est résumé dans l'Algorithme 7. Il est intéressant de remarquer que la contrainte de décomposition parcimonieuse de \mathbf{g}_2 n'influençant pas la mise à jour de la matrice F , il est tout à fait envisageable de rajouter au problème

(7.22) les contraintes de régularisation par moyenne qui ont été présentées dans ACIN-Cal page 54. En effet ces contraintes n'intervenaient que dans la mise à jour de la matrice F et sont donc trivialement insérables dans une nouvelle méthode d'étalonnage contenant à la fois la contrainte de parcimonie de \mathbf{g}_2 et la régularisation de F . Dans cette dernière méthode que nous appellons SpAIN-Cal (pour *Sparsity and Average Constrained IN-Cal*) la mise à jour de F est effectuée par la formule (4.46) d'ACIN-Cal, tandis que G et \mathbf{a} sont mis à jour avec les formules de SpIN-Cal.

Algorithme 7 Étalonnage par Factorisation Informée en matrices Non-négatives avec contrainte de décomposition parcimonieuse (SpIN-Cal).

Initialisation de G et F .

pour k de 1 à itération max **faire**

Mettre à jour F selon (4.21).

Mise à jour de \mathbf{a} par OMP pondéré.

Mettre à jour G selon (4.22) avec les matrices \mathcal{W} , \mathcal{X} et \mathcal{F} à la place de W , X et F .

fin pour

7.3 Régularisation par introduction d'un modèle

L'approche SpIN-Cal proposée précédemment permet de lier l'ensemble du réseau de capteurs grâce aux atomes d'un dictionnaire D connu. Cette nouvelles façon de connecter les capteurs – qui ne sont plus obligés d'effectuer de mesure dans une même cellule de la scène, voir figure 4.1 – permet de réduire la taille Δ_d du voisinage utilisé pour la construction de la matrice de données X . Cette nouvelle taille de voisinage est définie en fonction des données utilisées lors de la phase d'apprentissage du dictionnaire. Un problème majeur de l'approche par dictionnaire réside dans la sélection des atomes lorsque le dictionnaire est corrélé. En localisation/comptage de sources sonores, certains auteurs ont proposé de construire par exemple un dictionnaire de fenêtres de Blackman de moyennes [157] et écart-type [156] différents, et de chercher à décomposer le signal étudié suivant ce dictionnaire en "fusionnant" les atomes corrélés [156]. Cependant, malgré la très grande quantité de méthodes permettant la sélections des atomes, lorsque ceux-ci sont fortement corrélés, la sélection reste difficile et ne permet pas l'obtention de résultats satisfaisants, sauf à considérer le problème combinatoire exact [18].

Dans cette section, nous proposons donc de remplacer la régularisation avec dictionnaire par un modèle. La contrainte de décomposition parcimonieuses du phénomène physique y est alors remplacée par une contrainte de cohérence avec un modèle \mathcal{Y} évalué aux emplacements des mesures. Notons que la scène observée n'étant plus discrétisée, nous réécrivons la matrice observée X comme suit. Reprenons la définition 4.1 de la scène. Nous considérons maintenant

une discrétisation spatiale δ_d suffisamment fine pour être considérée comme ponctuelle. Avec cette grille, aucun capteur n'est en rendez-vous avec un autre capteur, au sens de la définition 2.3. La matrice X générée est donc de très grande dimension mais contient énormément de lignes sans aucune information. Nous proposons dans un premier temps de supprimer les lignes sans information de la matrice observée, en la multipliant par une matrice Ψ qui jouerait le même rôle que dans les approches parcimonieuses pour l'*inpainting*, comme expliqué dans la section 7.1.1.3. Par ailleurs, les lignes restantes ne contiennent qu'une seule valeur. Dans cette section, nous utiliserons malgré tout la notation X pour décrire cette matrice vidée des lignes nulles.

Deux approches sont présentées ci-après : dans un premier temps nous proposons d'utiliser un simple modèle polynômial pour régulariser le signal reconstruit dans \mathbf{g}_2 . Dans un deuxième temps, la connexion entre les capteurs sera assurée par interpolation des données manquantes par krigeage ordinaire.

7.3.1 Modèle polynômial

Nous supposons ici que le phénomène physique observé par le réseau de capteurs mobiles peut être approché par un polynôme de degré N donné. En considérant alors $\mathbf{y} \in \mathbb{R}^{q,1}$ le vecteur contenant les valeurs y_i du phénomène physique à tous les emplacements \mathbf{p}_i mesurés par les capteurs, alors

$$y_i \approx \hat{y}_i \triangleq \mathcal{Y}(\mathbf{p}_i) \quad \forall i \in \{1, \dots, q\}, \quad (7.30)$$

où $\mathcal{Y}(\mathbf{p}_i)$ est un polynôme de degré N , évalué aux coordonnées $\mathbf{p}_i = (p_{i,1}, p_{i,2})$. De la même façon qu'avec le dictionnaire, cette contrainte d'approximation polynômiale sur le phénomène physique observé peut être reportée sur la seconde colonne de la matrice G et est alors intégrée dans le problème de factorisation associé à l'étalonnage des capteurs, menant à l'optimisation de

$$\begin{aligned} \min_{G, F \geq 0, \hat{\mathbf{y}}} & \frac{1}{2} \cdot \|W \circ (X - G \cdot F)\|_{\mathcal{F}}^2 + \frac{\lambda}{2} \cdot \|\mathbf{g}_2 - \hat{\mathbf{y}}\|_{\mathcal{F}}^2, \\ \text{s.c.} & \quad G = \Omega_G \circ \Phi_G + \bar{\Omega}_G \circ \Delta_G, \\ & \quad F = \Omega_F \circ \Phi_F + \bar{\Omega}_F \circ \Delta_F, \\ & \quad \hat{\mathbf{y}} = [\mathcal{Y}(\mathbf{p}_1), \dots, \mathcal{Y}(\mathbf{p}_q)]^T. \end{aligned} \quad (7.31)$$

L'algorithme proposé pour tenir compte d'une régularisation par approximation polynômiale du phénomène physique observé est alors très simple et consiste à mettre à jour successivement les matrices G , F et le vecteur $\hat{\mathbf{y}}$. La nouvelle contrainte de régularisation n'intervenant pas dans le problème d'optimisation de la matrice F , cette dernière peut être mise à jour par la formule (4.21) développée pour la méthode IN-Cal.

Mise à jour du vecteur $\hat{\mathbf{y}}$ (contrainte de régression polynômiale)

Pour la mise à jour du vecteur $\hat{\mathbf{y}}$, le problème d'optimisation à considérer est le suivant

$$\min_{\hat{\mathbf{y}}} \|\mathbf{g}_2 - \hat{\mathbf{y}}\|_{\mathcal{F}}^2, \quad \text{s.c.} \quad \hat{y}_i = \mathcal{Y}(\mathbf{p}_i) \quad \forall i \in \{1, \dots, q\}, \quad \text{avec } \mathcal{Y} \text{ un polynôme de degré } N \text{ fixé.} \quad (7.32)$$

Dans le cas où le degré N du polynôme approximant est connu et fixé à 2, alors la contrainte qui s'applique au vecteur $\hat{\mathbf{y}}$ permet d'écrire que

$$\hat{\mathbf{y}} = P \cdot \mathbf{a}, \quad (7.33)$$

où

$$P \triangleq \begin{bmatrix} 1 & p_{1,1} & p_{1,2} & p_{1,1} \cdot p_{1,2} & p_{1,1}^2 & p_{1,2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & p_{q,1} & p_{q,2} & p_{q,1} \cdot p_{q,2} & p_{q,1}^2 & p_{q,2}^2 \end{bmatrix}, \quad (7.34)$$

est la matrices contenant les coordonnées des points d'observation et $\mathbf{a} \triangleq [a_0, \dots, a_6]^T$ est le vecteur des paramètres à déterminer du polynôme approximant. En réinjectant l'égalité (7.33) dans le problème d'optimisation (7.32), ce dernier revient à résoudre le problème

$$\mathbf{a} = \arg \min_{\tilde{\mathbf{a}}} \|\mathbf{g}_2 - P \cdot \tilde{\mathbf{a}}\|_{\mathcal{F}}^2, \quad (7.35)$$

ce qui peut être fait pas une méthode de moindres carrés classique. La mise à jour de $\hat{\mathbf{y}}$ consiste alors en ces deux étapes :

1. Une mise à jour de \mathbf{a} par moindres carrés

$$\mathbf{a} \leftarrow (P^T \cdot P)^{-1} \cdot P^T \cdot \mathbf{g}_2. \quad (7.36)$$

2. Une mise à jour de $\hat{\mathbf{y}}$ par construction

$$\hat{\mathbf{y}} \leftarrow P \cdot \mathbf{a}. \quad (7.37)$$

Mise à jour de G

La mise à jour associée à la matrice G est immédiate en constatant que le problème par rapport à G associé à l'optimisation (7.31) est identique à celui de la méthode SpIN-Cal précédente, à ceci près que l'approximation parcimonieuse $D \cdot \mathbf{a}$ est remplacée ici par le vecteur approximant $\hat{\mathbf{y}}$ et qu'il n'y a plus de donnée manquante dans la colonne \mathbf{g}_2 . De la même façon donc que pour SpIN-Cal, la mise à jour (4.22) peut être utilisée en remplaçant les matrices W , X et F par

$$\mathcal{W} \triangleq [W, \sqrt{\lambda} \cdot \mathbb{1}_{[q \times 1]}], \quad \mathcal{X} \triangleq [X, \mathbf{Y}] \quad \text{et} \quad \mathcal{F} \triangleq \left[F, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]. \quad (7.38)$$

L'algorithme de cette approche d'étalonnage à base de NMF informée et régularisée par approximation polynômiale – que nous appellons PolIN-Cal pour *Polynomial-based IN-Cal* – est présenté dans l'algorithme 8.

Algorithme 8 Étalonnage par Factorisation Informée en matrices Non-négatives avec contrainte d'approximation polynômiale (PolIN-Cal).

Initialisation de G et F .

pour k de 1 à itération max **faire**

Mettre à jour F selon (4.21) pour PolIN-Cal.

Mettre à jour de $\hat{\mathbf{y}}$ par les deux étapes (7.36) puis (7.37).

Mettre à jour G selon (4.22) avec les matrices \mathcal{W} , \mathcal{X} et \mathcal{F} .

fin pour

7.3.2 Utilisation d'un modèle géostatistique : le Krigeage

Comme il le sera montré dans le chapitre suivant, lorsque le phénomène physique peut être approché de façon précise par un polynôme de degré connu et fixé, la régularisation produite par l'ajout de cette contrainte montre de bonnes performances. Cependant, lors de l'observation d'un phénomène physique sur une zone étendue avec des phénomènes locaux assez complexes à modéliser (ou anticiper), l'approximation polynômiale du phénomène observé peut s'éloigner du phénomène physique réel et ne permet alors plus une régularisation efficace de l'étalonnage. Par ailleurs, la sélection du degré N du polynôme régularisant peut également être délicate et entraîner une forte erreur d'étalonnage en cas de mauvais choix. Nous proposons ici de régulariser la factorisation – et donc l'étalonnage – d'une façon différente, à l'aide d'une méthode d'interpolation géostatistique : le krigeage. Cette approche, plus souple que les précédentes, ne demandera pas l'ajout de connaissances supplémentaires et travaillera directement à partir des données disponibles, fournies par le réseau de capteurs.

Dans les méthodes SpIN-Cal et PolIN-Cal développées ci-dessus, nous avons régularisé la factorisation en introduisant une contrainte sur le phénomène physique, c.-à-d. sur le vecteur $\hat{\mathbf{y}}$ contenu dans la seconde colonne de G . De façon équivalente, nous aurions pu introduire cette régularisation sur les colonnes de la matrice X en cherchant la combinaison d'atomes – dans le cas de SpIN-Cal – ou le polynôme – dans le cas de PolIN-Cal – approximant chaque colonne de X . En procédant ainsi, la matrice X est alors complétée et les capteurs sont liés par les valeurs interpolées, créant ainsi des rendez-vous "virtuels" entre les capteurs. Cependant, cette approche alternative nous semble relativement similaire à approcher la seconde colonne de G : celle-ci peut en effet être vue comme une fusion des valeurs issues de tous les capteurs post-étalonnage.

Le krigeage est une méthode d'interpolation de données dite exacte dans le sens où il ne modifie pas les valeurs aux points d'observation. Cette caractéristique du krigeage ne nous permet donc pas de régulariser directement le vecteur \mathbf{y} correspondant à la deuxième colonne de G , puisque celui-ci étant complet, il ne serait pas modifié par application du krigeage. Nous

proposons donc de travailler ici sur les colonnes de la matrice de données X .

En considérant le modèle de capteurs à réponse affine, dans un cas idéal, les éléments x_{ij} de la j -ème colonne de X devrait correspondre à la mesure du j -ème capteur du réseau, à la i -ème position de mesure. Cependant, l'échantillonnage irrégulier des mesures du aux déplacements des capteurs induit un certain nombre de données manquantes dans la matrice X . Certains éléments x_{ij} ne sont donc pas disponibles. Nous pouvons voir la j -ème colonne de X , c.-à-d. \mathbf{x}_j , comme un champ physique – correspondant au champ \mathbf{y} affecté par la relation affine (4.2) – dont nous avons certaines observations – les emplacements des mesures du capteur j – et des valeurs inconnues à estimer – les entrées de \mathbf{x}_j manquantes. Nous proposons alors d'estimer les valeurs manquantes dans \mathbf{x}_j par krigeage à partir des ses entrées connues. En réalisant répétant cette complétion sur toutes les colonnes de X , la matrice obtenue – notée X^{comp} – est complète et les différents capteurs sont tous connectés par rendez-vous "virtuels". L'étalonnage peut alors être effectué par application de notre approche IN-Cal sur X^{comp} .

Il est important de noter ici que l'estimation des entrées manquantes dans X ne peut pas être réalisée en une fois pour toutes les entrées de la matrice X , mais bien colonne par colonne. En effet, nous ne pouvons pas considérer dans le krigeage les données brutes issues de différents capteurs. Ces dernières n'étant pas étalonnées, elles ne sont pas à la même échelle et les utiliser de façon simultanée perturberait fortement l'estimation du variogramme et par conséquent la reconstruction des entrées manquantes. Cependant, la qualité d'estimation des données manquantes sur une colonne qui n'aurait que peu de valeurs observées devra être prise en considération. Il est intéressant de remarquer que le krigeage est une méthode d'interpolation statistique et qu'il permet d'avoir une estimation de la fiabilité des valeurs estimées. En effet, pour chaque valeur interpolée \hat{x}_{ij} , le krigeage permet de déduire la variance d'erreur d'estimation, notée σ_{ij} , qui vaut zéro si l'estimée est parfaite – c.-à-d. si le point \hat{x}_{ij} est un point de mesure – ou est strictement supérieure à 0 sinon. Cette mesure σ_{ij} peut être prise en compte dans la pondération de la matrice W utilisée dans l'approche d'étalonnage IN-Cal. En effet, nous proposons de remplacer les valeurs nulles de w_{ij} de W par

$$\hat{w}_{ij} \triangleq \exp \left\{ \frac{\sigma_{ij}}{\hat{x}_{ij}^2 + \epsilon} \right\}, \quad (7.39)$$

où $\epsilon > 0$ est un seuil fixé par l'utilisateur pour éviter la division par zéro. La variance normalisée $\frac{\sigma_{ij}}{\hat{x}_{ij}^2 + \epsilon}$ fournit une vision relative non-biaisée de l'erreur d'estimation, qui a été montrée comme plus performante que la variance non-normalisée dans des problèmes de séparation de sources [192]. Ainsi, nous pouvons utiliser cette information sur la fiabilité des valeurs interpolées pour pondérer l'étalonnage. Cette approche d'étalonnage informée par krigeage – que nous appelons KIN-Cal pour *Krigging-based IN-Cal* en anglais – est résumée dans l'algorithme 9.

Algorithme 9 Étalonnage par Factorisation Informée en matrices Non-négatives avec completion par krigeage (KIN-Cal).

Initialisation de G et F .

Estimer X^{comp} par krigeage de chaque colonne de X .

Estimer la matrice \hat{W} dont les valeurs \hat{w}_{ij} sont définies dans (7.39).

Étalonner les capteurs par IN-Cal en remplaçant respectivement X et W par X^{comp} et \hat{W} .

7.4 Conclusion

Dans ce chapitre nous avons proposé trois régularisations permettant de relâcher la contrainte de rendez-vous nécessaire aux approches d'étalonnage développées dans la partie précédente de cette thèse. Les régularisations proposées ici consistent à ajouter artificiellement une connexion entre les mesures des différents capteurs, générant ainsi des rendez-vous "virtuels" entre les capteurs. Nous avons proposé l'emploi d'informations différentes : dictionnaire discret, régression polynômiale ou géostatistiques. Ces informations sont utilisées de manière relativement similaires dans la factorisation et il est donc possible d'utiliser un autre modèle, apportant éventuellement plus d'informations, à la place.

Il est à noter aussi que ces techniques demandent un certain degré d'expertise pour être utilisées de manière efficiente. En effet, les approches SpIN-Cal et PolIN-Cal nécessitent respectivement qu'on leur fournisse un dictionnaire et un degré de polynôme qui correspondent aux données à approcher. Au contraire, l'approche KIN-Cal utilise du krigeage et est *de facto* moins supervisée que les précédentes. Nous pouvons donc nous attendre à ce que cette approche fournisse de moins bonnes performances sur des exemples spécifiques mais soit plus versatile en pratique. Ces effets seront discutés dans le chapitre suivant, où nous testons les méthodes présentées.

Enfin, comme expliqué précédemment, les méthodes d'étalonnage de capteurs à réponse polynômiale proposées dans le chapitre 5 peuvent aussi être étendues de la même manière que les approches de NMF informées pour l'étalonnage de capteurs à réponse affine.

Étude de performances des approches régularisées

Dans ce chapitre, nous proposons d'étudier les performances des approches SpIN-Cal, PolIN-Cal et KIN-Cal proposées dans le chapitre précédent. Il est important de rappeler que les méthodes PolIN-Cal et KIN-Cal font intervenir les coordonnées des points de mesure directement dans leur processus d'interpolation de données. L'information spatiale est alors prise en compte sans avoir besoin de discrétiser le phénomène physique observé, contrairement à toutes les autres approches développées dans ce manuscrit. L'étude des performances de SpIN-Cal, PolIN-Cal et KIN-Cal sera donc menée en deux temps.

Nous comparerons d'abord les résultats de SpIN-Cal avec ceux de IN-Cal sur les jeux de données simulés dans le chapitre 6.

Dans un second temps, nous étudierons les performances de PolIN-Cal et KIN-Cal sur de nouveaux jeux de données sans discrétisation spatiale du phénomène physique observé.

8.1 Illustration de l'apport d'une régularisation spatiale

L'objectif de cette section est de montrer l'intérêt de la régularisation spatiale apportée par le dictionnaire dans SpIN-Cal, par rapport aux résultats obtenus avec notre approche IN-Cal. Nous reconsidérons donc les mêmes jeux de données que ceux du chapitre 6.

Nous reprenons donc la scène \mathcal{S} simulée page 73, permettant d'obtenir la matrice théorique G_{theo} , ainsi que la matrice des paramètres théoriques F_{theo} simulées page 71. Par le produit $G_{theo} \cdot F_{theo}$, nous construisons alors la matrice de mesures théorique X_{theo} dont toutes les entrées sont connues. Pour simuler des mesures par *mobile crowdsensing*, nous vidons ensuite cette matrice jusqu'à obtention des proportions de données manquantes ρ_{DM} et de rendez-vous ρ_{RV} souhaitées, et nous ajoutons un bruit gaussien de façon à obtenir le SNR d'entrée voulu. La matrice X obtenue après retrait des entrées constitue la matrice de données à partir de laquelle sera effectué l'étalonnage des capteurs, c.-à-d. l'estimation des matrices G et F .

Pour construire le dictionnaire D utilisé par l'approche SpIN-Cal, nous nous plaçons ici dans un cadre idéal dans lequel la scène \mathcal{S} peut être parfaitement reconstruite à partir de 2 atomes de D (c.-à-d. que le signal \mathbf{y} est 2-parcimonieux) constitué d'un total de 62 atomes

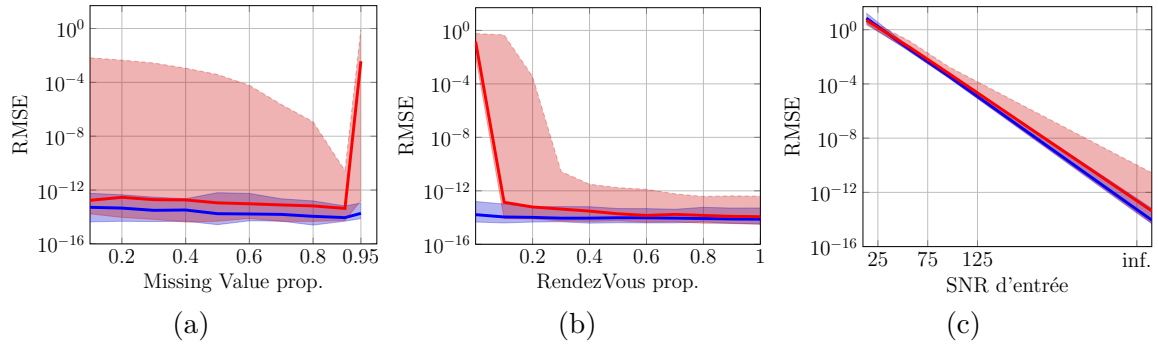


FIGURE 8.1 – Performances des méthodes IN-Cal (en rouge) et SpIN-Cal (en bleue) par rapport à (a) la proportion de données manquantes, (b) la proportion de rendez-vous, (c) le SNR d'entrée.

peu corrélés. Ainsi, le vecteur \mathbf{y} construit par vectorisation de la scène \mathcal{S} s'écrit $\mathbf{y} = D \cdot \mathbf{a}$ avec $\|\mathbf{a}\|_0 = 2$, c.-à-d. toutes les entrées du vecteur \mathbf{a} sont nulles excepté les deux entrées associées aux atomes permettant l'approximation parcimonieuse de \mathbf{y} .

8.1.1 Comparaison des approches SpIN-Cal et IN-Cal

Afin de mettre en évidence l'apport de la contrainte de décomposition parcimonieuse de SpIN-Cal par rapport à l'approche IN-Cal, nous réalisons trois séries de tests. Dans la première, nous fixons la proportion de rendez-vous ρ_{RV} à 30%, le SNR d'entrée à ∞ (pas de bruit) et nous faisons varier la proportion de données manquantes ρ_{DM} de 0 à 95%. Dans une seconde série de tests nous fixons $\rho_{DM} = 90%$, $\text{SNR} = \infty$ et nous faisons varier ρ_{RV} de 0 à 100%. Enfin dans une dernière série, nous fixons $\rho_{DM} = 90%$, $\rho_{RV} = 30%$ et nous faisons varier le SNR d'entrée de ∞ (pas de bruit) jusqu'à environ 20 dB. Tous les tests sont effectués à partir de la même scène simulée \mathcal{S} , cependant, pour chaque jeu de paramètres $(\rho_{DM}, \rho_{RV}, \text{SNR})$, 40 simulations sont réalisées, chacune avec une nouvelle matrice F à estimer.

Les deux méthodes testées ici, IN-Cal et SpIN-Cal sont toutes deux lancées avec $5.e5$ itérations. Le RMSE final est ensuite calculé sur les gains des capteurs, c.-à-d. la seconde ligne de F . Les figures 8.1a, 8.1b et 8.1c montrent les performances obtenues avec chaque méthode (SpIN-Cal en bleue et IN-Cal en rouge), par rapport à la proportion de données manquantes, la proportion de rendez-vous et le SNR d'entrée, respectivement. Comme dans le chapitre 6, les courbes pleines des figures représentent les valeurs médianes de RMSE tandis que les enveloppes délimitent les valeurs minimales et maximales obtenues sur les 40 simulations. Dans le cas de la méthode SpIN-Cal, nous fixons le paramètre de régularisation λ de (7.22) à 10 dans chaque série de tests. L'influence de ce paramètre sera étudié plus en détails dans des tests plus complets.

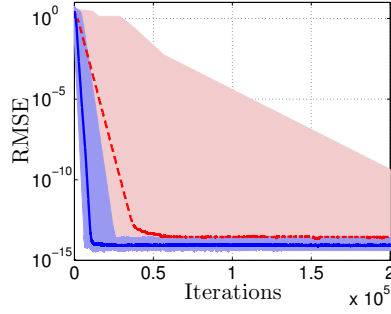


FIGURE 8.2 – Illustration de l’accélération de la convergence par analyse des performances des méthodes IN-Cal (en rouge) et SpIN-Cal (en bleu) par rapport aux itérations. Figure extraite de [51].

Influence des données manquantes

Comme expliqué précédemment, les performances obtenues avec IN-Cal sur la figure 8.1a sont celles commentées page 76. Nous rappelons en particulier que IN-Cal produit des performances intéressantes jusqu’à environ 90% de données manquantes. Au delà de ce seuil, les performances sont fortement dégradées. Avec SpIN-Cal en revanche, les performances sont relativement constantes sur tout l’intervalle testé, c.-à-d. de 0 à 95% de données manquantes. Par ailleurs, l’enveloppe obtenue avec IN-Cal reste élevée quelque soit la quantité de données manquantes dans X , ceci étant du à la faible vitesse de convergence de la méthode sur quelques simulations. La taille de l’enveloppe est fortement diminuée par l’utilisation du dictionnaire qui, en ajoutant une contrainte à la factorisation, accélère sa vitesse de convergence. Cette accélération est visible dans la figure 8.2, extraite de [51] où nous montrons l’évolution du RMSE en fonction des itérations. Pour la réalisation de cette figure, nous avons appliqué les méthodes IN-Cal et SpIN-Cal sur 25 jeux de données simulés, non bruités, comportant 90% de données manquantes et 30% de rendez-vous. La courbe et l’enveloppe rouge correspondent aux performances obtenues avec IN-Cal, tandis que celles en bleu correspondent à SpIN-Cal. Il est aussi intéressant de remarquer que SpIN-Cal produit également de meilleures performances qu’ACIN-Cal – dont les résultats sont présentés dans la figure 6.8 – dans le cas où 95% des données sont manquantes dans X . Ce gain de performance est du au fait que l’approximation parcimonieuse permet de lier l’ensemble des capteurs les uns aux autres et mène donc à un étalonnage cohérent de l’ensemble du réseau, ce que ne permet pas de faire notre approche normalisée ACIN-Cal.

Influence des rendez-vous

Ici encore les performances obtenues par IN-Cal sont identiques à celles présentées dans la figure 6.9. Nous notons en particulier que IN-Cal donne de bon résultats en médiane et en enveloppe pour les proportions de rendez-vous avec des capteurs étalonnés supérieures à 20%. Cependant, les performances sont fortement dégradées dès lors que $\rho_{RV} \leq 20\%$. L’introduction du dictionnaire permet de lier l’ensemble des capteurs du réseau grâce à ses atomes, ce qui permet de créer artificiellement les rendez-vous manquants entre les capteurs mobiles et les références, ce qui stabilise les performances de la méthode quelque soit la proportion de rendez-vous. Les performances obtenues avec SpIN-Cal sont donc totalement indépendantes

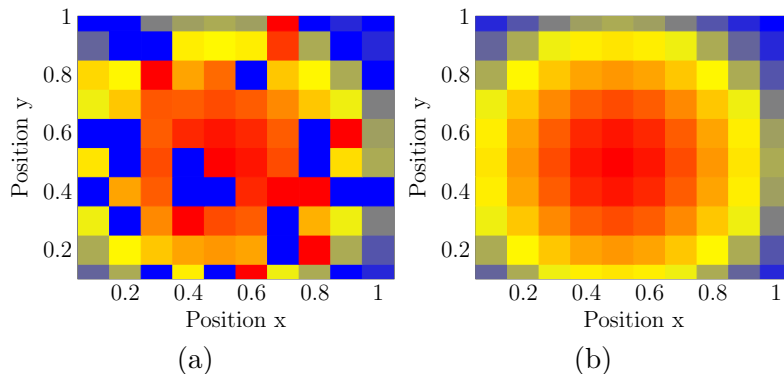


FIGURE 8.3 – Scènes reconstruites après étalonnage par les méthodes (a) IN-Cal et (b) SpIN-Cal, à partir d’une matrice de données X avec 95% de données manquantes.

de la proportion de rendez-vous et le RMSE médian obtenu est ici toujours situé en dessous de $1.e - 12$. Ceci démontre en particulier que le dictionnaire autorise l’utilisation d’un pas de discrétisation plus fin pour la méthode SpIN-Cal par rapport à IN-Cal.

Influence du bruit

Nous nous intéressons ici à l’impact du bruit de mesure sur les performances de l’approche SpIN-Cal. Comme nous pouvons le voir sur la figure 8.1c, le comportement des méthodes IN-Cal et SpIN-Cal est relativement similaire puisque les deux approches montrent un RMSE médian qui évolue linéairement avec l’augmentation du SNR d’entrée. Dans les cas peu bruités – jusqu’à environ 50 dB – SpIN-Cal produit des performances légèrement meilleures que l’approche IN-Cal. Cependant, dans les cas très bruités, l’approche avec dictionnaire peut fournir des résultats moins bons qu’avec IN-Cal. Ceci est dû au fait que dans certaines simulations très bruitées, les atomes sélectionnés lors pour la régularisation ne correspondent pas aux atomes théoriques permettant de reconstruire au mieux le phénomène physique, ce qui entraîne de médiocres performances d’étalonnage.

Au travers des trois séries de tests présentées ci-dessus, nous avons montré que l’apport d’une régularisation spatiale – ici grâce à un dictionnaire – permet de guider l’étalonnage et d’atteindre des performances assez remarquables même dans des configuration extrême comportant peu ou pas de rendez-vous entre les capteurs mobiles et les données de référence, et beaucoup de données manquantes. Dans la sous-section suivante, nous proposons d’étudier le comportement global de l’approche SpIN-Cal ainsi que l’influence du paramètre de régularisation λ qui n’a pas été considéré jusqu’à présent.

Un exemple de cartographie

Comme évoqué dans le chapitre 7, un bon déroulement des méthodes d’étalonnage par factorisation matricielles présentées dans ce manuscrit produit une bonne estimation des matrices G et F . Or, la seconde colonne de G contient une estimation du phénomène physique partiellement observé par les capteurs et permet donc de reconstruire partiellement une première

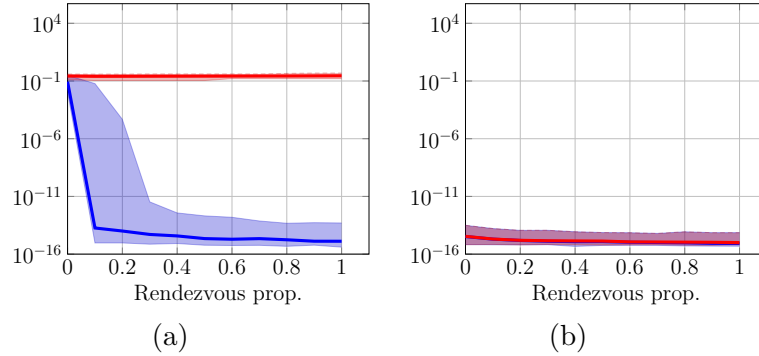


FIGURE 8.4 – RMSE obtenu sur la seconde colonne de la matrice G par les méthodes (a) IN-Cal et (b) SpIN-Cal, calculés par rapport à la proportion de rendez-vous, en rouge sur l'ensemble de la colonne de G et en bleu sur les éléments de G pour lesquels une mesure est disponible.

cartographie du champ physique. Les méthodes développées dans la première partie de ce manuscrit – dont IN-Cal – ne permettent pas d'interpoler les valeurs du champ physique aux emplacements qui n'ont pas été visités par des capteurs : la cartographie produite reste donc "à trous". Dans le cas de l'approche SpIN-Cal en revanche, la décomposition du phénomène physique via le dictionnaire permet bien de reconstruire l'ensemble du champ physique.

Ces différences de reconstruction sont mises en évidence dans les figures 8.3a et 8.3b, dans lesquelles sont présentées les scènes reconstruites avec les méthodes IN-Cal et SpIN-Cal, respectivement. Ces scènes sont obtenues après application des approches d'étalonnage sur la matrice simulée X comportant 95% de données manquantes. Sur ces figures, nous pouvons voir que la reconstruction produite par la méthode IN-Cal est de très mauvaise qualité car (i) il n'y a pas assez de rendez-vous pour assurer un étalonnage du réseau de capteurs et (ii) des données manquantes ne sont pas reconstruites. Avec SpIN-Cal en revanche, la reconstruction de la scène est bien plus précise et ne compte plus de cellules vides.

Une autre façon de mettre en évidence cette différence de performances consiste à suivre l'évolution du RMSE sur la deuxième colonne de la matrice G en fonction de la proportion de rendez-vous. Dans les figures 8.4a et 8.4b, nous montrons les RMSE obtenus avec IN-Cal et SpIN-Cal respectivement. Dans ces figures, la courbe rouge représente le RMSE médian calculé sur la seconde colonne de G complète, tandis que la courbe bleue représente le RMSE médian calculé sur la seconde colonne de G en ne considérant pas les cellules vides, pour lesquels aucune mesure n'est disponible dans X . Nous remarquons que les deux RMSE sont similaires (et bas) pour l'approche SpIN-Cal alors que la méthode IN-Cal fournit des RMSE beaucoup plus élevés, non-seulement on tient compte de toutes les données de \mathbf{g}_2 , dont certaines ne peuvent pas être directement estimées, mais aussi quand on ne tient compte que des données observées dans X .

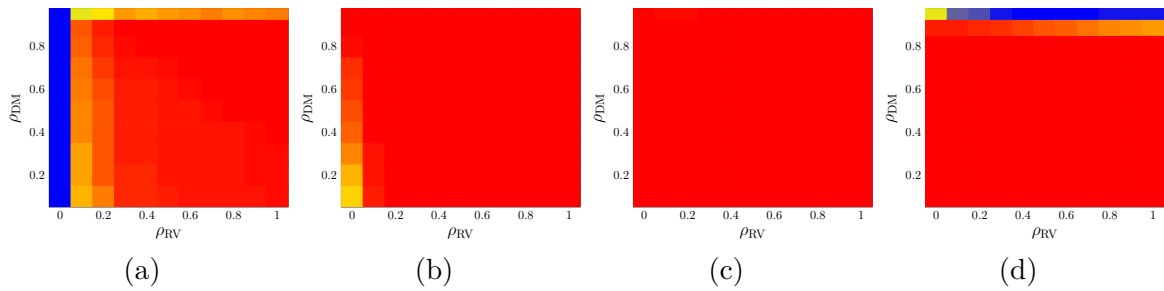


FIGURE 8.5 – Performances globales de SpIN-Cal pour différentes valeurs du paramètre de régularisation : (a) $\lambda = 0.01$, (b) $\lambda = 1$, (c) $\lambda = 10$, (d) $\lambda = 100$.

8.1.2 Performances globales de SpIN-Cal et influence du paramètre de régularisation

Nous nous plaçons ici dans le cas d'une matrice de données X construite comme précédemment, sans bruit et dans laquelle nous contrôlons la proportion de données manquantes ρ_{DM} qui varie de 10 à 95%, ainsi que la proportion de rendez-vous ρ_{RV} qui varie de 0 à 100%. Pour chaque couple de paramètres (ρ_{DM}, ρ_{RV}) , nous effectuons 40 simulations sur lesquelles nous appliquons SpIN-Cal afin de procéder à l'étalonnage. Les figures 8.5a, 8.5b, 8.5c et 8.5d montrent les performances obtenues par SpIN-Cal après $5.e5$ itérations, avec un paramètre λ respectivement égal à 0.01, 1, 10 et 100 respectivement. Ces figures mettent en évidence la proportion de simulation ayant atteint un RMSE final inférieur à $1.e - 8$. Comme pour la figure 6.6, un "pixel" rouge signifie que 100% des simulations ont atteint le seuil fixé – ici $1.e - 8$ – tandis qu'un pixel bleu signifie qu'aucune simulation n'a atteint ce seuil.

Nous constatons immédiatement que lorsque le paramètre de régularisation λ est choisi trop petit – $\lambda = 0.01$ – alors la méthode SpIN-Cal ne profite pas de la régularisation apportée par le dictionnaire, et les performances obtenues sont identiques à celles présentées dans la figure 6.6b pour la méthode IN-Cal.

Dans le cas la figure 8.5b où le paramètre λ est fixé à 1, nous constatons une nette amélioration des performances dans toutes les situations extrêmes, c.-à-d. lorsqu'il y a beaucoup de données manquantes et/ou très peu de rendez-vous. Seuls quelques simulations dans le cas où la proportion de rendez-vous est nulle et où peu de données sont manquantes n'atteignent pas le seuil de RMSE fixé. Ceci est principalement dû au fait que lorsqu'il y a peu de données manquantes et peu de rendez-vous, la convergence de la méthode d'étalonnage est lente, et le paramètre $\lambda = 1$ n'est pas encore suffisant pour profiter pleinement de l'accélération produite par la régularisation.

Dans cette série de tests, les meilleures performances sont atteintes lorsque $\lambda = 10$, comme présenté dans la figure 8.5c. Avec ce choix de paramètre, la méthode profite à la fois de l'accélération amenée par la régularisation, mais également de la connexion des mesures grâce au dictionnaire. Dans ce cas, toutes les simulations réalisées ont atteint le seuil de $1.e - 8$.

Enfin comme nous pouvons le voir sur la figure 8.5d, lorsque $\lambda = 100$, le poids est trop fort et une confiance trop importante est accordée à l'approximation parcimonieuse. Ce poids limite les performances dans les cas où l'approximation est mal estimée lors des premières itérations.

Dans cette section, nous avons montré l'intérêt de l'ajout d'une régularisation spatiale – ici en utilisant un dictionnaire – dans les méthodes d'étalonnage développées précédemment. En effet, cette régularisation permet de lier l'ensemble des capteurs les uns aux autres mais également aux mesures de référence, rendant les méthodes présentées précédemment totalement indépendantes de la proportion de données manquantes ainsi que de la proportion de rendez-vous. Cependant, nous nous sommes placés jusqu'à présent dans un cas idéal où les mesures effectuées par les capteurs étaient en rendez-vous exacts, et où le dictionnaire était parfaitement construit de façon à obtenir une reconstruction parfaite du champ physique observé. Ces deux hypothèses de rendez-vous et de reconstruction exacte ne sont pas toujours vérifiées en pratique, et des erreurs d'estimation ou de rendez-vous peuvent perturber l'étalonnage.

Dans la section suivante, nous proposons d'évaluer les performances des approches PolIN-Cal et KIN-Cal présentées dans le chapitre précédent. En particulier, contrairement à SpIN-Cal, ces dernières approches ne demandent pas de discrétisation du phénomène physique observé et ne sont donc *a priori* pas sensibles aux erreurs de rendez-vous.

8.2 Comparaison des approches régularisées sur un nouveau jeu de données

Dans cette section, nous proposons de comparer les performances des approches IN-Cal, SpIN-Cal, PolIN-Cal et KIN-Cal sur un jeu de données plus "réaliste" dans lequel le phénomène physique n'est plus discrétisé *a priori*. En effet, nous considérons le cas où nous obtenons des mesures géo-localisées et où le pas Δ_d n'est pas fourni mais doit être estimé. De fait, les approches PolIN-Cal et KIN-Cal ne sont pas affectées par l'absence de cette information puisque ces approches utilisent la position exacte des mesures dans le processus d'approximation/interpolation des données. Au contraire, l'approche IN-Cal – et dans une moindre mesure SpIN-Cal – nécessite de fournir un pas de discrétisation spatiale adéquat : en effet, si ce pas est trop grand, l'écart possible entre deux mesures dans une même cellule sera trop élevé et perturbera l'étalonnage. Au contraire, si ce pas est trop petit, le risque que des capteurs ne puissent pas être en rendez-vous n'est pas négligeable.

8.2.1 Simulations des données

Comme expliqué dans la section 7.3, sans discrétisation, la matrice de données X est construite de sorte que chaque ligne corresponde à une position \mathbf{p} à laquelle au moins un capteur a effectué une mesure. Ainsi, en supposant que les capteurs ont fourni des mesures

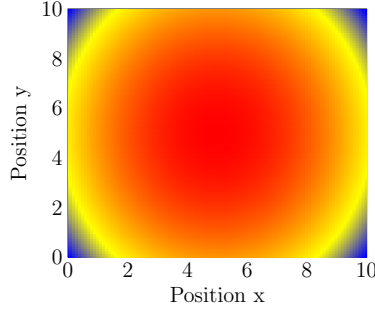


FIGURE 8.6 – Champs physique simulé pour les tests des méthodes PolIN-Cal et KIN-Cal

en m points $\{\mathbf{p}_1 \dots \mathbf{p}_m\}$ de l'espace, alors $X = (x_{ij})$, où

$$x_{ij} = \begin{cases} f_{0,j} + f_{1,j} \cdot \mathcal{Y}(\mathbf{p}_i) & \text{si le capteur } j \text{ a effectué une mesure à la position } \mathbf{p}_i, \\ 0 & \text{sinon,} \end{cases} \quad (8.1)$$

avec $\mathcal{Y}(\mathbf{p}_i)$ la valeur du champ physique à la position \mathbf{p}_i , $f_{0,j}$ l'offset du capteur et $f_{1,j}$ son gain. Il semble évident qu'en considérant directement les positions des mesures des capteurs et non plus les cellules de la scène discrétisée, les rendez-vous entre les capteurs deviennent très rares voir même inexistant et la matrice X est alors très creuse, ne comportant plus qu'une mesure par ligne. La taille de la matrice X est entièrement définie par le nombre n de capteurs et le nombre q de mesure totales. Dans les simulations ci-dessous, nous proposons de fixer le même nombre de mesures m_{cpt} pour chaque capteur, c.-à-d. $q = n \cdot m_{\text{cpt}}$. Afin de simuler la matrice de données X , nous commençons par simuler un réseau constitué de $n = 26$ capteurs (25 capteurs mobiles et 1 capteur de référence), comme expliqué dans le chapitre 6, en tirant aléatoirement 25 gains et 25 offsets autour des paramètres théoriques fournis dans la documentation du capteur de particules [74]. Les paramètres simulés sont ensuite tronqués de façon à respecter les spécifications constructeur, puis sont agencés dans une matrice F_{theo} . La matrice G_{theo} est construite en juxtaposant une colonne de uns et le vecteur contenant les valeurs du phénomène physique simulé en $q = n \cdot m_{\text{cpt}}$ points de l'espace tirés aléatoirement, où $m_{\text{cpt}} = 10, 25, 50$ ou 100 . Pour simuler le phénomène physique observé, nous avons défini la fonction

$$\mathcal{Y}(\mathbf{p}) = 0.175 + 0.08p_1 + 0.05p_2 - 0.008p_1^2 - 0.005p_2^2, \quad (8.2)$$

simulant un épisode de pollution issu d'une source ponctuelle, comme représenté dans la figure 8.6, et dont les valeurs se situent entre 0.1 et 0.5 mg/m^3 , pour satisfaire aux limites de sensibilité et aux contraintes de linéarité des capteurs de particules simulés. Une fois les matrices F_{theo} et G_{theo} simulées, nous créons la matrice de données complète X_{theo} par le produit $G_{\text{theo}} \cdot F_{\text{theo}}$. Finalement, nous pouvons créer la matrice de données X en enlevant aléatoirement des mesures dans X_{theo} de façon à ce que chaque ligne de X comporte une donnée et que chaque colonne en compte m_{cpt} . L'étalonnage peut ensuite être réalisé par PolIN-Cal ou KIN-Cal à partir de la matrice de données X et est présenté dans les sous-sections ci-dessous. Il est à noter qu'on ne peut pas étudier ici l'influence du taux de valeurs manquantes ou de la proportion de rendez-vous entre capteurs. Nous ne jouons donc que sur la densité de mesure – via la valeur de m_{cpt} – pour étudier ces effets. Enfin, nous considérons

ci-dessous un cas idéal non-bruité.

Pour étudier les performances des méthodes IN-Cal et SpIN-Cal, il faut fournir un pas de discrétisation pour chaque approche ainsi qu'un dictionnaire. Nous supposons ici que la scène définie dans la figure 8.6 représente un champ sur une zone de $10 \text{ m} \times 10 \text{ m}$. Pour l'approche IN-Cal, nous avons testé dans des expériences préliminaires des cellules carrées de côté fixé respectivement à 0.5, 1 et 2 m. Les résultats obtenus avec le pas de 1 m étaient légèrement meilleurs et nous pavons donc la scène dans les tests ci-dessous avec des cellules carrées de cette taille. Par ailleurs, si deux points de mesure sont aux bords de deux cellules voisines et distinctes, il est possible que leur distance soit inférieure à Δ_d et donc qu'ils soient en rendez-vous (au sens de la définition 2.3) sans l'être au sens de la discrétisation de la scène. Pour pallier à ce problème, nous proposons un maillage recouvrant dans lequel deux cellules horizontalement ou verticalement adjacentes sont recouvrantes de 50%.

Pour SpIN-Cal, le dictionnaire est constitué de plusieurs gaussiennes scindées de centres différents et discrétisées avec un pas $1.e - 2 \text{ m}$. La scène simulée, représentée dans la figure 8.6 est alors 2-parcimonieuse suivant le dictionnaire D , constitué de 52 atomes. Il est à noter que contrairement aux tests précédents, les atomes permettent une approximation parcimonieuse de la scène qui n'est pas exacte, même en l'absence de bruit. Par ailleurs, deux gaussiennes de centres voisins fournissent des atomes corrélés, ce qui rend plus difficile leur sélection dans la décomposition parcimonieuse, comme discuté précédemment.

8.2.2 Étude de performances des approches IN-Cal, SpIN-Cal, PolIN-Cal et KIN-Cal

Nous présentons ici les performances d'étalonnage obtenues avec les approches IN-Cal, SpIN-Cal, PolIN-Cal et KIN-Cal. Comme expliqué précédemment, la matrice de données X est entièrement définie par le nombre n de capteurs et le nombre m_{cpt} de mesures par capteur. Dans les tests suivants, nous proposons donc d'analyser les performances d'étalonnage – c.-à-d. le RMSE final obtenu après $5.e5$ itérations – en sortie des approches PolIN-Cal et KIN-Cal par rapport au nombre m_{cpt} de mesures par capteur.

Performances de l'approche IN-Cal

La figure 8.7 montre les performances de la méthode IN-Cal avec les paramètres ci-dessus pour la scène considérée, en fonction du nombre m_{compt} de points de mesures. Nous remarquons que l'étalonnage n'est pas réalisé de manière satisfaisante, avec un RMSE médian proche de 4 quelque soit la valeur de m_{compt} . Cela est dû à la discrétisation qui amène un effet similaire à du bruit de mesure sur les valeurs observées dans chaque cellule.

Performances de l'approche SpIN-Cal

La figure 8.8 présente les performances obtenues avec SpIN-Cal quand le poids λ de la contrainte de parcimonie de \mathbf{g}_2 est fixé à 10 ou 100, respectivement. Nous notons de meilleures

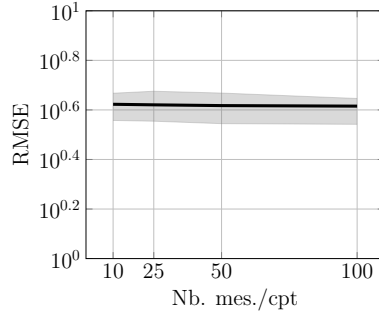


FIGURE 8.7 – Performances de la méthode IN-Cal en fonction du nombre de mesures par capteur, avec un pas de discrétisation de la scène fixé à 1.

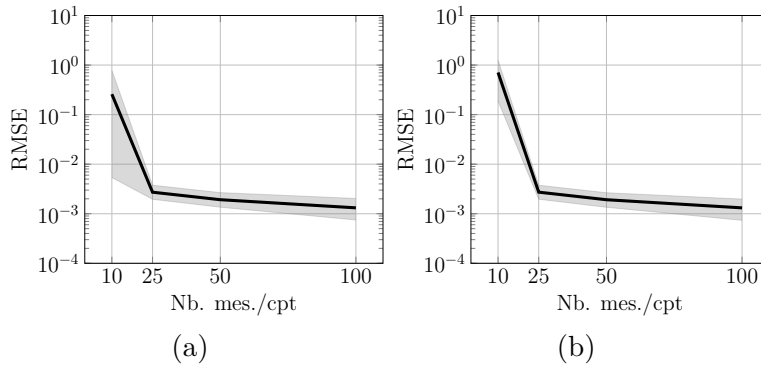


FIGURE 8.8 – Performances de l’approche SpIN-Cal en fonction du nombre de mesures par capteur (a) $\lambda = 10$, (b) $\lambda = 100$.

performance qu’avec IN-Cal, mais en deça des tests réalisés précédemment avec des RMSE médians légèrement au dessus de $1e - 3$ lorsque $m_{\text{cpt}} \geq 25$. Comme discuté ci-dessus, cela peut être dû au fait que le dictionnaire approche seulement la scène et que ses atomes sont corrélés. Lorsque le nombre de mesures est faible, c.-à-d. $m_{\text{cpt}} = 10$, l’étalonnage n’est pas bien réalisé (avec un RMSE médian respectivement proche de 0.3 et 0.8 quand $\lambda = 10$ ou 100). Cette chute de performance peut être due au poids trop fort du dictionnaire sur les premières itérations de la NMF, engendrant une erreur d’étalonnage qui ne se corrige pas au fur et à mesure des itérations. Ce comportement est aussi visible sur la figure 8.5, comme discuté dans la section 8.1.2.

Performances de l’approche PolIN-Cal

Dans les figures 8.9a, 8.9b et 8.9c, nous montrons les performances obtenue par la méthode PolIN-Cal par rapport au nombre de mesures par capteurs que nous faisons varier de 10 à 100, pour différentes valeurs du paramètre de régularisation λ . Dans la figure 8.9a le paramètre est fixé à $\lambda = 10$, dans la figure 8.9b à 100 et dans la figure 8.9c le paramètre λ évolue en fonction de nombre de mesures par capteurs avec la formule $\lambda = 10 \cdot m_{\text{cpt}}$. Pour chaque valeur de m_{cpt} , 40 simulations ont été réalisées. Le degré N du polynôme recherché est fixé

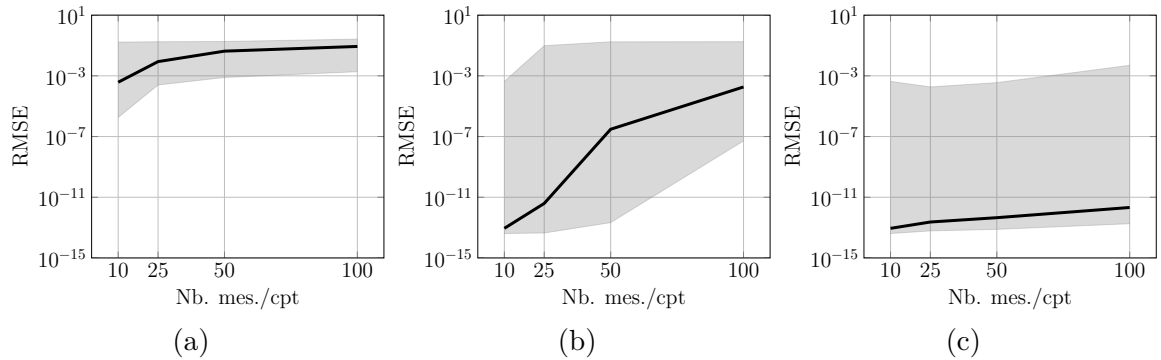


FIGURE 8.9 – Performances d'étalonnage de la méthode PolIN-Cal en fonction du nombre de mesures par capteur et du paramètre de régularisation (a) $\lambda = 10$, (b) $\lambda = 100$ et (c) λ adaptatif.

à 3 et est donc suréstimé par rapport à celui du champs simulé. Dans ces figures, la courbe bleue montre le RMSE médian obtenu en sortie d'étalonnage, sur 40 simulations, tandis que l'enveloppe délimite les valeurs minimales et maximales obtenues.

Dans la figure 8.9a, les valeurs de RMSE obtenues sont relativement stabilisées – en médiane et en enveloppe – pour m_{cpt} entre 25 et 100. Cependant, ces valeurs sont plutôt élevées (bien que plus bas qu'IN-Cal sur ces mêmes tests) : le RMSE médian se situe autour de $1.e - 2$. Pour $m_{cpt} = 10$ cependant, le RMSE médian ainsi que le partie basse de l'enveloppe diminuent, ce qui semble assez peu naturel. En effet, avec la régularisation par régression polynômiale, nous nous attendons ici à ce que les performances s'améliorent avec l'augmentation du nombre de mesures m_{cpt} . Cette amélioration des performances est probablement due au fait que la pondération de l'approximation polynômiale dans la fonction de coût (7.31) devient plus importante quand le nombre de points de mesure diminue, ce qui confirme l'expérience suivante.

Dans la figure 8.9b, nous augmentons la valeur du paramètre de régularisation λ que nous fixons alors à 100. Les performances obtenues – en comparaison de celle présentées dans la figure 8.9a – sont largement améliorées pour toutes les valeurs de m_{cpt} testées, ce qui confirme que les performances limitées obtenues avec $\lambda = 10$ étaient dues à une valeur trop faible de ce paramètre de régularisation, ne permettant pas alors à PolIN-Cal de profiter pleinement de la régularisation. Toutefois, nous remarquons également que l'amélioration de performance obtenue avec l'augmentation du paramètre de pondération est nettement plus importante pour les plus faibles valeurs testées de m_{cpt} . Le RMSE croît avec la valeur de m_{cpt} , laissant penser que le paramètre de régularisation λ devrait être sélectionné en fonction du nombre de mesures par capteurs.

Finalement, dans la figure 8.9c, nous adaptons la valeur du paramètre λ en fonction de m_{cpt} avec la formule $\lambda = 10 \cdot m_{cpt}$. Nous constatons alors une nette amélioration des performances pour toutes les valeurs de m_{cpt} , même les plus élevées. Dans cette figure, les valeurs médianes de RMSE sont toutes autour de $1.e - 12$. La partie haute est également diminuée et passe de

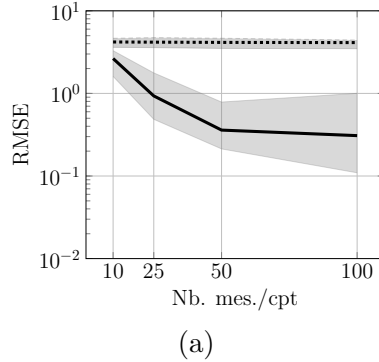


FIGURE 8.10 – Comparaison des performances d’étalonnage des méthode IN-Cal (courbe pointillée) et KIN-Cal (courbe pleine), en fonction du nombre de mesures pas capteur.

1.e-1 pour toutes les valeurs de m_{cpt} dans la figure 8.9a, à environ $1.e - 3$ dans la figure 8.9c.

Performances de l’approche KIN-Cal

Dans la figure 8.10, nous montrons les performances obtenues avec la méthode KIN-Cal (en bleu). Le RMSE médian obtenu en sortie de l’étalonnage diminue avec le nombre de mesures par capteurs. Cependant, les performances sont assez mauvaises, avec un RMSE médian autour 5 lorsque $m_{\text{cpt}} = 10$ et baissant à environ 0.5 lorsque 100 mesures par capteurs sont disponibles. Ce résultat, qui peut sembler décevant par rapport aux performances obtenues avec PolIN-Cal ou SpIN-Cal, est probablement du au fait que l’approche KIN-cal n’effectue le krigeage qu’une seule fois sur chaque colonne de X :

- la qualité d’estimation des données manquantes dans chaque colonne de X dépend du nombre et de la répartition des échantillons de la scène à reconstruire. Ainsi, plus il y a de points de mesures et meilleure est l’estimation du champ physique.
- contrairement aux approches PolIN-Cal et SpIN-Cal, la méthode KIN-Cal n’améliore pas la qualité du krigeage au fur et à mesure des itérations de la NMF informée.

Cependant, bien que les performances de KIN-Cal semblent beaucoup moins intéressantes que celles obtenues avec PolIN-Cal ou SpIN-Cal par exemple, il est important de rappeler que, comme évoqué dans le chapitre précédent, KIN-Cal est l’approche d’étalonnage régularisée la moins supervisée qui ait été développée dans ce manuscrit. De ce fait, cette approche devrait être plus facilement s’adapter aux différentes concentrations de particules fines sur une zone d’étude en pratique que les autres approches regularisées, qui nécessitent de trouver la bonne fonction de régression polynômiale ou le bon dictionnaire.

8.3 Conclusion

Dans ce chapitre, nous avons étudié les performances des extensions régularisées de la méthode IN-Cal que nous avons proposées dans le chapitre précédent. Ces approches, en ajoutant une régularisation spatiale au problème d’étalonnage, permettent de s’affranchir des rendez-

vous exacts entre les capteurs qui limitaient les performances et les possibilité d'application des approches d'étalonnage développées dans la partie précédente de ce manuscrit.

Les méthodes SpIN-Cal et PolIN-Cal ont montré des performances très intéressantes en atteignant des valeurs de RMSE particulièrement basse en fin d'étalonnage. Cependant, les tests réalisés ont également mis en évidence que l'une des principales difficultés de ces méthodes réside dans la sélection des paramètres de régularisation, qui, loin d'être triviale, devrait également être optimisée.

L'approche KIN-Cal présente des performances plus faibles qui pouvaient être attendues et qui ne remettent pas en cause l'intérêt de cette approche. Rappelons que dans les tests présentés précédemment, le phénomène physique simulé pouvait être parfaitement reconstruit par régression polynômiale, ce qui n'est pas toujours le cas en pratique lorsque le champs physique est moins régulier et comporte des phénomènes plus localisés.

Conclusion générale

9.1 Conclusion

Dans cette thèse, nous nous sommes intéressés au problème d'étalonnage d'un réseau de capteurs mobiles dans le cadre d'un projet de surveillance de la qualité de l'air par mesure citoyenne, dans le cadre du projet OSCAR. Nous avons commencé par reformuler le problème d'étalonnage d'un réseau de capteurs mobiles sous la forme d'un problème de factorisation matricielle à données manquantes dans lequel une matrice de données X contenant les mesures des capteurs mobiles est factorisée suivant un produit $G \cdot F$, où les facteurs contiennent respectivement le modèle d'étalonnage fonction du phénomène physique observé et les paramètres d'étalonnage de chaque capteur. La réécriture du problème proposée tient compte de diverses hypothèses classiques en étalonnage de capteurs mobiles et a permis d'ouvrir le problème d'étalonnage de capteurs à la littérature très riche des méthodes de factorisation matricielle. Dans le cas de capteurs à réponse affine, le problème d'étalonnage prend la forme d'un problème informé de factorisation en matrices non-négatives, tandis que pour des capteurs à réponse non-linéaire il s'agit d'un problème informé de factorisation en matrices semi-non-négatives.

La formulation proposée dans ce manuscrit, fait principalement intervenir la paramétrisation développée dans [123] pour informer la factorisation. Cette dernière est employée ici afin de tenir compte dans la factorisation de mesures issues de capteurs de référence, idéalement fournies par une AASQA, comme par exemple ATMO dans le cadre du projet OSCAR.

Nous avons dans un premier temps proposé deux méthodes d'étalonnage pour des capteurs à réponse affine, l'une à base de complétion de matrice – méthode NMF/C-Cal – l'autre utilisant des mises à jour multiplicatives spécifiques – méthode IN-Cal. Les algorithmes de ces deux approches sont résumés aux pages 50 et 52, respectivement. En comparaison aux approches de la littérature en étalonnage de capteurs mobiles, les deux méthodes proposées permettent de tenir compte de l'ensemble des connexions entre les capteurs du réseau et se sont montrées moins sensibles à l'échantillonnage irrégulier des données du au mouvement des capteurs.

À la suite de ces premiers développements, nous avons proposé d'inclure les informations provenant de la documentation constructeur au problème de factorisation, de façon à permettre sa résolution dans le cas où certaines hypothèses nécessaires à l'étalonnage ne seraient pas vérifiées. En particulier, l'introduction de cette régularisation permet de conserver de bonnes performances d'étalonnage dans le cas où aucune donnée de référence ne serait dis-

ponible. La méthode de résolution du problème de NMF régularisé – méthode ACIN-Cal – consiste alors en des mises à jour multiplicatives particulières. L’algorithme de cette approche est présenté à la page 56 de ce manuscrit.

Nous avons par la suite étendu ces premières approches à l’étalonnage de capteurs à réponse non-linéaire. Dans ce nouveau modèle, la non-linéarité de la fonction d’étalonnage se traduit dans la factorisation par l’ajout d’une contrainte de structure sur l’une des matrices facteur, qui doit alors prendre la forme d’une matrice de Vandermonde, ainsi que par une possible négativité de certains paramètres. L’extension des approches proposées précédemment à ce type de capteurs donne trois nouvelles méthodes, l’une à base de complétion de matrice et régression, une autre – SC-ISN-Cal – à base de factorisation matricielle avec mises à jour par descentes de gradient, et une dernière – ASC-ISN-Cal – tenant compte des données constructeur des capteurs.

En seconde partie de cette thèse, nous avons proposé de régulariser les approches précédentes en ajoutant au problème de factorisation une contrainte d’approximation du phénomène physique reconstruit par un modèle. Cette nouvelle contrainte permet principalement de rendre les approches proposées moins sensibles à l’échantillonnage irrégulier des mesures des capteurs, en connectant toutes les données entre elles via un dictionnaire – méthode SpIN-Cal – un modèle polynômial – méthode PolIN-Cal – ou via une estimation du phénomène observé par krigeage – méthode KIN-Cal. Les approches SpIN-Cal et PolIN-Cal ont montré de très bonnes performances sur des jeux de données simulés. Cependant, elles nécessitent en pratique de pouvoir effectuer un apprentissage de dictionnaire – pour SpIN-Cal – ou d’avoir une connaissance *a priori* des paramètres de calage du modèle physique – degré de polynôme dans le cas d’PolIN-Cal. *A contrario*, avec KIN-Cal, l’approximation régularisante est estimée directement à partir des données et ne demande pas d’ajout de connaissances particulières. Les performances obtenues sur données simulées sont moins impressionnantes que celles atteintes avec nos autres méthodes. Cependant, en application sur données réelles, cette dernière méthode plus souple doit pouvoir s’adapter plus facilement à n’importe quel jeu de données, sans nécessiter de réglages particuliers, contrairement aux autres approches.

9.2 Perspectives

Le travail présenté dans ce manuscrit ouvre de nombreuses perspectives parmi lesquelles :

Application des approches proposées sur des données réelles : Dans le cadre du projet OSCAR, le développement du boîtier de mesure citoyenne a connu des phases de retard. Cependant, les premières campagnes de fabrication ont eu lieu fin 2017. Il reste à finaliser l’application APISENSE® pour cette expérience, afin de faire remonter des données et tester nos approches.

Régularisation spatiale parcimonieuse de données continues : Les méthodes régularisées par un a priori spatial qui ont proposées dans cette thèse, notamment SpIN-Cal et PolIN-Cal, ont montré de bonnes performances sur l'ensemble des tests. Cependant, elles nécessitent le réglage d'un pas de discrétisation et d'un dictionnaire pour la première, et un degré de polynôme pour la seconde. Une idée intéressante pourrait consister à mixer les deux approches, en fournissant à l'utilisateur un dictionnaire de fonctions paramétriques (polynômes, splines, bases radiales) qui seraient optimisées pour coller aux données, sous contrainte de parcimonie du nombre de fonctions utilisées.

L'accélération des méthodes proposées : Les méthodes de factorisation utilisées dans cette thèse pour la mise en place des approches d'étalonnage ne sont pour l'instant pas des plus optimales en terme de temps de calcul et un travail reste à faire pour les accélérer et les rendre applicables en pratique, sur de grand jeux de données. Pendant la thèse, nous avons développé une variante de la méthode de factorisation matricielle avec gradient de Nesterov au cas d'une matrice à factoriser comportant des données manquantes. Cette approche – qui est présentée en détails dans l'article [52] joint dans l'annexe C – est l'une des méthodes envisagées pour accélérer les approches d'étalonnage proposées dans ce travail.

La prise en compte de l'évolution temporelle du phénomène physique : Dans cette thèse, tous les tests et développements ont été réalisés en considérant une unique scène \mathcal{S} . Si l'extension au cas multi-scènes par concaténation des matrices de données est immédiate, il peut être intéressant de considérer l'intégration des données au court du temps d'une façon qui permettrait de tenir compte de l'évolution temporelle du phénomène physique. Une approche envisagée est la remise en forme du problème d'étalonnage non plus comme un problème de factorisation matricielle, mais comme un problème de factorisation tensorielle où la dimension supplémentaire introduite par l'écriture des données sous forme de tenseur correspondrait alors au temps. Une alternative consisterait à traiter le problème sous forme de co-factorisation.

L'extension à des modèles de réponse plus complexe : Nous avons considéré ici le cas de capteurs à réponse affine ou non-linéaire, qui correspondent à la plupart des réponses de compteurs de particules du commerce. Cependant, d'autres capteurs – comme les capteurs de gaz – font intervenir des modèles d'étalonnage plus complexes, et tiennent par exemple compte d'interactions entre différents phénomènes physiques (température, humidité, etc).

Optimisation de l'observatoire : Comme expliqué précédemment, le *mobile crowdsensing* classique fait appel aux (capteurs connectés ou issues de) smartphones des utilisateurs. Ceux-ci utilisant une batterie, il est important de limiter au maximum la consommation énergétique des applications de *mobile crowdsensing*, au risque de perdre des utilisateurs volontaires sinon. Comme expliqué dans l'introduction de la section 7.1.3, les modèles physiques ont souvent été couplés à des stratégies d'optimisation de l'observatoire [15, 174, 176]. Il pourrait être

intéressant de coupler de telles stratégies aux approches présentées dans cette thèse, de façon à (i) étalonner de manière fine le réseau, (ii) obtenir une cartographie fine du champ observé et (iii) minimiser la consommation énergétique.

Méthode robuste d'étalonnage de capteurs fixes

Dans les tous premiers travaux réalisés dans le cadre de cette thèse, nous nous sommes d'abord intéressés aux méthodes d'étalonnages existantes à base de projection en sous-espace [5, 128], dans le but d'étendre ces approches destinées à des réseaux de capteurs fixes à des réseaux de capteurs mobiles. Nous sommes rapidement arrivés à la conclusion que de telles méthodes, sans ajout de contraintes ou connaissances supplémentaires n'étaient pas envisageables pour l'application considérée : en effet, les approches [5, 128] ont d'abord besoin d'une phase d'apprentissage du phénomène physique, pendant laquelle les capteurs sont étalonnés. Or, dans le cadre du projet OSCAR, cette phase n'était pas envisageable. Toutefois, nous avons proposé un pré-traitement à base d'ACP robuste permettant de filtrer les valeurs aberrantes parmi les données des capteurs et ainsi de limiter leur très fort impact sur les résultats d'étalonnage. Nous avons montré dans [55] que ce pré-traitement est particulièrement bien adapté à une utilisation en amont de l'utilisation de méthode d'étalonnage de la littérature telles que [5, 128], puisque celui-ci peut s'appliquer directement sur les matrices de données utilisées par ces méthodes, sans l'ajout d'hypothèses supplémentaires. Ces premiers travaux, qui n'ont pas été évoqués dans ce manuscrit, sont présentés ci-après.

Copyright

©2017 IEEE. Ré-imprimé avec l'autorisation des auteurs de : C. Dorffer, M. Puigt, G. Delmaire, G. Roussel, Outlier-robust calibration method for sensor networks, in Proceedings of the 2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their application to Mechatronics (ECMSM), San Sebastian, May 24-27, 2017.

Outlier-Robust Calibration Method for Sensor Networks

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel

Univ. Littoral Côte d’Opale, LISIC – EA 4491, F-62228 Calais, France

clement.dorffer@univ-littoral.fr, matthieu.puigt@univ-littoral.fr

gilles.delmaire@univ-littoral.fr, gilles.roussel@univ-littoral.fr

Abstract—In this paper, we aim to blindly calibrate the responses of a sensor network whose outputs are possibly corrupted by outliers. In particular, we extend some well-known nullspace-based blind calibration approaches, proposed for fixed sensors with affine responses—i.e., with unknown gain and offset for each sensor—to that difficult case. These state-of-the-art approaches assume that the true data lie in a known lower dimensional subspace, so that in practice sensors can be calibrated by projection of the uncalibrated observations to this subspace. A robust extension was recently proposed in order to provide less sensitivity to noise. In this paper, we show that such methods (including the robust extensions) are very sensitive to outliers and we propose new extensions able to deal with such issues. For that purpose, we assume the outliers to be rare events, which can be modeled as a sparse contribution to the low-rank observed data. Using such an assumption, we separate sparse outliers from the low-rank data, so that we can perform calibration. We show that the proposed approach is able to handle up to 10% of outliers in the data without major impact on the calibration accuracy while state-of-the-art methods are already sensitive to the presence of one unique outlier.

I. INTRODUCTION

While being increasingly cheap, sensors are now part of our daily lives. Indeed, deployed in large-scale networks, they provide massive quantities of data and allow many applications for, e.g., smart cities [1], agriculture [2], or environment monitoring [3]. In such scenarios, the high number of measurements provided by the numerous sensors compensate the poor sensitivity and quality of an individual sensor (which is due to the fact that such sensors are low-cost) [4].

However, an inherent problem to any sensor network is the sensor calibration step which is necessary to get interpretable and consistent data. Calibration usually consists of estimating the unknown parameters that link the sensors outputs to the corresponding physical inputs. Traditionally, calibration can be performed in a laboratory by determining the calibration parameters to fit a known and controlled input to which sensors are exposed to. However, when the sensors are too numerous or inaccessible, calibrating them in laboratory is no longer possible and specific signal processing methods have been proposed to remotely calibrate the sensors of the network from the collected measurements.

In the current literature, these calibration approaches can be classified into two main families, namely micro- and macro-calibration [5]. Micro-calibration consists of calibrating one single sensor of a network at a time while macro-calibration

operates on the whole set of sensors. Depending whether or not the sensors are mobile in the network, different assumptions have been proposed to solve such problem. When the sensors are mobile, they can be in *rendezvous*, i.e., they are in the same spatio-temporal neighborhood, thus sensing the same phenomenon [6]. Such an assumption was recently used¹ in both micro- [8], [9], [10] and macro-calibration² [12], [13], [14]. However, in the case of fixed sensors, other assumptions are needed. For fixed sensors, state-of-the-art calibration approaches assume to know the low-rank subspace where the sensed phenomenon lies [15], [16] or consider a compressed sensing framework [17], [18]. Other approaches use statistical properties³ of the sensed phenomenon to derive estimates of the calibration parameters [19].

Another issue with sensor networks may happen if the sensors are faulty. In such a case, the readings result in outliers which must be taken into account [20]. In this paper, we investigate the influence of outliers on blind calibration. In particular, we focus on the subspace-based approach [15] and its robust extension [16]. We show that these methods are very sensitive to the presence of even a unique outlier and we propose an outlier-robust extension, which assumes that the number of outliers is low with respect to the mass of generated data, i.e., the outliers can be modeled as a sparse contribution in the observed low-rank data.

The remainder of the paper reads as follows. We introduce the sensor calibration problem in Section II. In Section III, we briefly recall the principles of state-of-the-art methods [15] and we show how sensitive to outliers these methods are. In Section IV, we propose a novel outlier-robust calibration method which combines a preprocessing stage—which finds and removes outliers from the sensor readings—to the nullspace-based strategies. In Section V, we experimentally show the enhancement provided by our proposed method. In particular, the preprocessing stage is shown to be able to filter up to 10% of outliers in the data. Lastly, we conclude and discuss the possible extension of the work in Section VI.

For the sake of readability, the notations used in this paper are defined in Table I.

¹Some approaches, e.g., [7], do not use the rendezvous assumption.

²Please note that in that case, macro-calibration has similarities with low-rank matrix completion [11].

³Moment-based calibration methods were also proposed for mobile sensor networks, e.g., in [7].

TABLE I
NOTATIONS USED IN THIS PAPER.

Symbol	Definition
a	a scalar
\underline{a}	a column vector
A	a matrix
\mathbf{A}	a diagonal matrix
variables x , \underline{x} , or X	the sensed physical phenomenon
variables y , \underline{y} , or Y	the outlier-free contributions of the sensor readings
variables z , \underline{z} , or Z	the sensor readings
$\mathcal{P}(\cdot)$	the projection operator on the nullspace of $\underline{x}(t)$

II. PROBLEM STATEMENT

In this paper, we assume to observe the sensed signals $\underline{z}(t) \in \mathbb{R}^n$ provided by a synchronized network composed of n fixed sensors. Moreover, we also assume that the sensors are possibly defecting. Sensor faults might be due to, e.g., communication errors between sensors, signal subsampling due to desynchronization (assuming that missing samples are set to zero), sparse or continuous spurious sensor readings (see Fig. 1). In that case, the corrupted sensor readings $\underline{z}(t)$, observed at time t read

$$\underline{z}(t) = \underline{y}(t) + \underline{q}(t), \quad (1)$$

where $\underline{y}(t)$ and $\underline{q}(t)$ are the vectors of outlier-free sensor readings and outliers, respectively. In this paper, we assume the outliers to be sparse, i.e., most values in $\underline{q}(t)$ are null.

If one consider that sensor readings are outlier-free, i.e., $\underline{q}(t) = 0$ for any time t , then one can directly observe $\underline{y}(t)$. If we assume that the calibration functions for each sensor of the network to be affine, then the true input phenomenon—denoted $x_i(t)$ —reads

$$x_i(t) = y_i(t) \cdot \alpha_i + \beta_i, \quad (2)$$

where $y_i(t)$ is the signal observed by the i -th sensor at time t , and α_i and β_i are the unknown sensor gain and offset associated with the sensor i , respectively. Then, by defining $\underline{x}(t) \in \mathbb{R}^n$ as the column vector containing the input phenomenon observed by the n sensors and $\mathbf{Y}(t)$ the diagonal matrix of $\underline{y}(t)$, one can write

$$\underline{x}(t) = \mathbf{Y}(t) \cdot \underline{\alpha} + \underline{\beta}, \quad (3)$$

where $\underline{\alpha}$ and $\underline{\beta}$ are the vectors containing the sensor gains and offsets, respectively. Calibrating the sensor network then consists of estimating $\underline{\alpha}$ and $\underline{\beta}$.

For that purpose, we briefly recall the assumptions made in [15], upon which we build our proposed method. We first assume that the sensor network is oversampling the observed signal, i.e., $\underline{x}(t)$ lies in a low dimensional subspace \mathcal{S} which is assumed to be spanned by a set of r basis vectors, where $r \ll n$ is the rank of \mathcal{S} . These basis vectors are assumed to be known. Such an assumption might look strong. However, it arises in several situations, e.g., for bandlimited signals with sensor distances lower than required by the Shannon-Nyquist sampling rate [15]. It is also valid in a machine learning framework—where the sensors are calibrated in their early

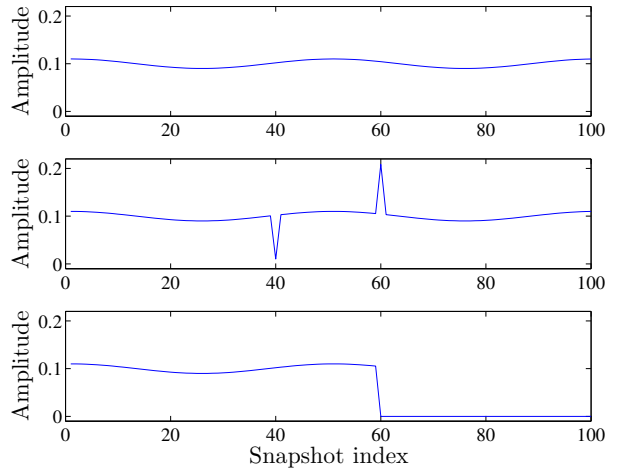


Fig. 1. Examples of sensor faults: (top) outlier-free sensor reading, (middle) presence of 2 local sensor faults (e.g., constant or bottom-noise fault), and (bottom) fault due to defective sensor or communication lost.

life—so that \mathcal{S} can be learned from early sensor readings. Then, while sensors are aging, their responses drift with time, i.e., their calibration parameters vary. This implies that at a given time t , the sensor readings follow the calibration model (3). Besides, as \mathcal{S} is known, one can compute \mathcal{P} , the linear projection on its orthogonal subspace. The key idea used in [15] (and in its robust extension [16]) consists of noticing that $\underline{x}(t)$ is in the nullspace of \mathcal{P} , i.e.,

$$\mathcal{P}(\underline{x}(t)) = \mathcal{P}(\mathbf{Y}(t) \cdot \underline{\alpha} + \underline{\beta}) = 0. \quad (4)$$

In practice, we assume to collect synchronized snapshots $\underline{y}(t_1), \underline{y}(t_2), \dots, \underline{y}(t_k)$ from the sensor network, where t_j denotes the time of the j -th snapshot. In order to perform calibration, the number k of snapshots satisfies [15]

$$k \geq \left\lceil \frac{n-1}{n-r} \right\rceil + 1. \quad (5)$$

Defining $\mathbf{Y}_j \triangleq \mathbf{Y}(t_j)$, the nullspace blind calibration problem thus aims to solve

$$\{\hat{\underline{\alpha}}, \hat{\underline{\beta}}\} = \arg \min_{\underline{\alpha}, \underline{\beta}} |\mathcal{P}(\mathbf{Y}_j \cdot \underline{\alpha} + \underline{\beta})|, \quad \forall j \in \llbracket 1, k \rrbracket. \quad (6)$$

In the next sections, we briefly introduce the principles of the nullspace-based calibration methods [15], [16]. We then show that they are very sensitive to the presence of outliers in the observed signals—i.e., when $\underline{q}(t) \neq 0$ —and we propose a novel nullspace calibration method which introduces a preprocessing stage to remove outliers.

III. NULLSPACE-BASED SENSOR CALIBRATION

We now briefly recall the principles of the nullspace-based calibration methods. As \mathcal{P} is linear and defining

$$\bar{\mathbf{Y}} \triangleq \frac{1}{k} \sum_{j=1}^k \mathbf{Y}_j, \quad (7)$$

the centered version of Eq. (4) reads

$$\mathcal{P}((\mathbf{Y}_j - \bar{\mathbf{Y}}) \cdot \underline{\alpha}) = 0, \quad \forall j \in \llbracket 1, k \rrbracket, \quad (8)$$

which yields

$$\mathcal{P}(\mathbf{Y}_j - \bar{\mathbf{Y}}) \cdot \underline{\alpha} = 0, \quad \forall j \in \llbracket 1, k \rrbracket. \quad (9)$$

The authors in [15] then defined a matrix C by stacking the projection in each snapshot, i.e.,

$$C \triangleq \begin{bmatrix} \mathcal{P}(\mathbf{Y}_1 - \bar{\mathbf{Y}}) \\ \vdots \\ \mathcal{P}(\mathbf{Y}_k - \bar{\mathbf{Y}}) \end{bmatrix}. \quad (10)$$

Combining Eqs. (9) and (10) implies that the unknown gain vector $\underline{\alpha}$ is solution of

$$\hat{\underline{\alpha}} = \arg \min_{\underline{\alpha}} \|C \cdot \underline{\alpha}\|_2, \quad (11)$$

which can be solved using ordinary Least Square (LS) or SVD—noticing that the last eigenvector is proportional to $\underline{\alpha}$ —as proposed in [15]. However, the solutions of Eq. (11) are not unique: indeed, if $\underline{\alpha}$ is a solution to Eq. (11), then $\lambda \cdot \underline{\alpha}$ is also a solution. The authors in [15], [16] removed the scale ambiguity by setting the first element⁴ of $\underline{\alpha}$ to 1.

The authors in [16] then proposed an extension of [15] which uses Total Least Square (TLS) to solve Eq. (11). TLS mainly consists of a weighted SVD where the weights allow to take into account the accuracy of each sensor.

Once the gain vector is estimated, the authors in [15] estimate the offsets⁵, assuming the observed phenomenon is zero-mean (or at least its average value is known). In that case, one easily derive

$$\underline{\beta} = -\bar{\mathbf{Y}} \cdot \underline{\alpha}. \quad (12)$$

Finally, the calibration process is summarized in Algorithm 1 and provides an interesting performance while remaining easy to implement. However, the projection step makes

Algorithm 1 Nullspace-based calibration methods [15], [16].

- Center the measurement matrices \mathbf{Y}_j (see Eq. (9)) ;
 - Construct C (see Eq. (10)) ;
 - Estimate $\underline{\alpha}$ from Eq. (11) using LS, SVD or TLS ;
 - Estimate $\underline{\beta}$ (see Eq. (12)).
-

the method very sensitive to outliers. Indeed, in a realistic framework, the measurements provided by the sensor network can be corrupted by outliers, as explained in Section II. The outlier-free vectors $\underline{y}(t_j)$ are thus inaccessible and one must deal with corrupted sensor readings $\underline{z}(t_j)$ instead. If

⁴Please note that other strategies—e.g., assuming to know the sum of the values of the unknown calibration parameters [18]—have been proposed in the literature and can be also applied with these methods instead. In the remainder of the paper, we follow the same strategy as in [15], [16].

⁵An approach for estimating $\underline{\beta}$ when $n \ll m$ offsets are known was also proposed in [15].

one directly applies the above nullspace-based calibration techniques on $\underline{z}(t_j)$, then

$$\mathcal{P}(\mathbf{Z}_j \cdot \underline{\alpha} + \underline{\beta}) = \mathcal{P}((\mathbf{Y}_j + \mathbf{O}_j) \cdot \underline{\alpha} + \underline{\beta}), \quad (13)$$

$$= \mathcal{P}(\mathbf{Y}_j \cdot \underline{\alpha} + \underline{\beta}) + \mathcal{P}(\mathbf{O}_j \cdot \underline{\alpha}), \quad (14)$$

where $\mathbf{Z}_j \triangleq \text{diag}(\underline{z}(t_j))$ and $\mathbf{O}_j \triangleq \text{diag}(\underline{o}(t_j))$. It turns out that the estimation of $\underline{\alpha}$ is perturbed by \mathbf{O}_j , which also implies a poor estimation accuracy of $\underline{\beta}$. Please note that the TLS method should be robust to outliers if the latter are present in very few sensors. Indeed, by lowering the weights associated to the relevant sensors, it should be still possible to perform accurate calibration on the outlier-free sensors. However, TLS should provide a very limited enhancement (or even no enhancement at all) if the outliers are distributed over all the sensors readings, which is more likely to happen in practice.

It should be noticed that the above methods were also extended to a partially blind calibration problem in [15], [16]. In that case, some sensor calibration parameters are assumed to be known, thus improving the calibration performance. However, even these methods are sensitive to outliers as they are also based on nullspace projection.

As a consequence, a novel approach able to handle outliers must be proposed, which we introduce in the next section.

IV. PROPOSED OUTLIER-ROBUST CALIBRATION METHOD

The outliers that may corrupt the data can be seen as sparse noisy data. In sensor networks, it usually corresponds to spurious readings or transmission faults.

At this stage, it should be noticed that $\mathbf{Z}_j \cdot \underline{\alpha} + \underline{\beta}$ in Eq. (13) can also be expressed as

$$\mathbf{Z}_j \cdot \underline{\alpha} + \underline{\beta} = \mathbf{A} \cdot \underline{z}(t_j) + \mathbf{B}, \quad (15)$$

where $\mathbf{A} \triangleq \text{diag}(\underline{\alpha})$ and $\mathbf{B} \triangleq \text{diag}(\underline{\beta})$, respectively. We now define the matrix $Z \triangleq [\underline{z}(t_1), \dots, \underline{z}(t_k)]$, i.e., whose (i, j) -th element reads

$$(Z)_{ij} \triangleq z_i(t_j). \quad (16)$$

From Eq. (1) and Eq. (16), we can derive

$$Z = Y + O, \quad (17)$$

where Y and O contain the entries of $\underline{y}(t_j)$ and $\underline{o}(t_j)$, respectively.

Using the same formalism as in Eq. (15), Eq. (3) provides

$$X = \mathbf{A} \cdot Y + \mathbf{B}, \quad (18)$$

where $X \triangleq [\underline{x}(t_1), \dots, \underline{x}(t_k)]$. Since $\underline{x}(t)$ lies in a low dimensional subspace \mathcal{S} defined by $r \ll \min(n, k)$ basis functions, it turns out that X is a low-rank matrix, and as a consequence, Y is also low-rank. As O is assumed to be sparse, it turns out that the sensor readings Z are written as the sum of a low-rank and a sparse matrix, which has been intensively studied under the name of Robust Principal

Component Analysis (RPCA) since the pioneering work in [21], [22]. RPCA aims to solve

$$\min_{Y,O} \|Y\|_*^2 + \lambda \|O\|_1, \quad \text{s.t. } Z = Y + O, \quad (19)$$

where $\|\cdot\|_*$ is the nuclear norm of a matrix—i.e., the sum of its eigenvalues— $\|\cdot\|_1$ is the ℓ_1 norm, and λ a positive weighting parameter. RPCA became a popular tool as it finds numerous applications, e.g., in video surveillance, face recognition, or ranking and collaborative filtering [22].

Numerous RPCA algorithms have been proposed in the literature and can thus be applied to Z , in order to estimate Y and O . In some preliminary tests, we found the Inexact Augmented Lagrange Multiplier (IALM) method [23] to be fast and accurate for the considered application. As most RPCA algorithms, the IALM-RPCA technique is iterative and computes partial SVDs of the size of Z to estimate O and to derive \hat{Y} , an estimation of Y .

Then, we can easily extract the vectors $\hat{y}(t_j)$ and build the diagonal matrices $\hat{Y}_j, \forall j \in \llbracket 1, k \rrbracket$. At this stage, it is straightforward to perform blind calibration by applying any of the above projection-based calibration techniques to \hat{C} , the matrix defined in Eq. (10) using the estimated \hat{Y}_j . The structure of the proposed methods is provided in Algorithm 2. Let us stress that using RPCA as a preprocessing stage of

Algorithm 2 Proposed outlier-robust calibration methods.

- Construct Z (see Eq. (16))
 - Estimate \hat{Y} using IALM-RPCA [23]
 - Center the matrices \hat{Y}_j (see Eq. (9)) ;
 - Construct \hat{C} (see Eq. (10)) ;
 - Estimate $\underline{\alpha}$ from Eq. (11) using LS, SVD or TLS ;
 - Estimate $\underline{\beta}$ (see Eq. (12)).
-

the nullspace-based calibration does not involve any extra-assumption. Moreover, in a machine learning framework, it can also be applied to early readings—where the sensors are assumed to be calibrated—as sensor faults or communication errors might happen. In that case, the estimation of \mathcal{S} might be improved as well. Lastly, such a preprocessing can also be applied to the partially blind extensions of the above nullspace-based blind calibration methods, as proposed in [15], [16].

In the next section, we investigate the enhancement provided by our proposed method with respect to the state-of-the-art ones, in both blind and partially blind scenarios.

V. EXPERIMENTAL VALIDATION AND DISCUSSION

In this section, we investigate the performance of the proposed outlier-robust calibration methods. To that end, we simulate a low-dimensional subspace \mathcal{S} of rank $r = 20$ and we derive a set of $n = 100$ sensors by randomly picking 100 gains and offsets. We then simulate $k = \lceil 3 \cdot r \cdot \log(n) \rceil$ signals from the subspace \mathcal{S} —as done in [16]—and we derive the outlier-free sensor responses associated to each signal—using Eq. (2)—thus providing the vectors $y_j, \forall j \in \llbracket 1, k \rrbracket$.

In the tests below, we explore the effects of both the subspace error and the number of outliers on the calibration performance. Indeed, as the subspace is assumed to be learned prior to the calibration, it may be not perfectly known. Its estimation error is simulated by adding random zero-mean Gaussian perturbations to the theoretical subspace with increasing variance [15]. Such an error is then estimated as

$$\epsilon_S \triangleq \frac{1}{k} \cdot \sum_{i=1}^k \frac{\|P(\underline{x}_i)\|_2}{\|\underline{x}_i\|_2}. \quad (20)$$

In the tests reported below, the values of ϵ_S range from 0 to 0.45. A few outliers are then randomly added to the observed data matrix Y . Their proportion ρ_o ranges between 0 and 20% of the number of data points in Z . For each pair of parameters (ϵ_S, ρ_o) , we apply the LS [15] and TLS [16] nullspace-based calibration methods to the sensor readings corrupted by outliers. The weights used in TLS are all set to 1, so that TLS behaves as SVD. We also apply our proposed extension, where IALM-RPCA is applied to Z before applying LS or TLS calibration.

As a performance criterion, we compute the error between the estimated parameters $\tilde{\alpha}$ and the true ones $\underline{\alpha}$ using

$$\text{Error} = \frac{\|\underline{\alpha} - \tilde{\alpha}\|_2}{\|\underline{\alpha}\|_2}. \quad (21)$$

Figure 2 shows the median performance of the tested methods obtained over 100 trials. Let us first focus on the enhancement reached by the state-of-the-art methods. When the proportion of outliers is set to 0, the TLS calibration approach outperforms the LS one, which is consistent with [16]. While the median calibration error reached using LS is above 0.9 when the subspace error ϵ_S is above 0.26, the TLS-based technique provides median calibration errors always below 0.30. However, its spread (shown by considering minimum and maximum errors) is higher than with LS. As an example, the errors reached when $\epsilon_S = 0.45$ are between 0.17 and 0.92 with TLS calibration while they are between 0.94 and 0.98 with LS calibration. When some outliers are added to the data, the calibration error reached by both methods significantly increases. However, their behaviour is not similar. On the one hand, the median error reached with LS calibration remains close to 1 for almost any tested value ϵ_S of the subspace error (it is always above 0.9 when only $\rho_o = 0.2\%$ of outliers are present in the data, for any initial subspace error). On the contrary, the median TLS calibration error remains always below 0.6 when $\rho_o \leq 0.2\%$. However, for higher values of ρ_o , the median TLS calibration error drastically increases (it is always above 10 when $\rho_o \geq 2\%$, for any tested value of ϵ_S), thus showing that such a robust approach is actually very sensitive to the presence of outliers.

On the contrary, both our proposed methods—which combine the IALM-RPCA preprocessing with the above calibration techniques—provide almost constant performance up to $\rho_o = 10\%$ of outliers. Interestingly, the preprocessing does not affect the calibration performance when no outlier is

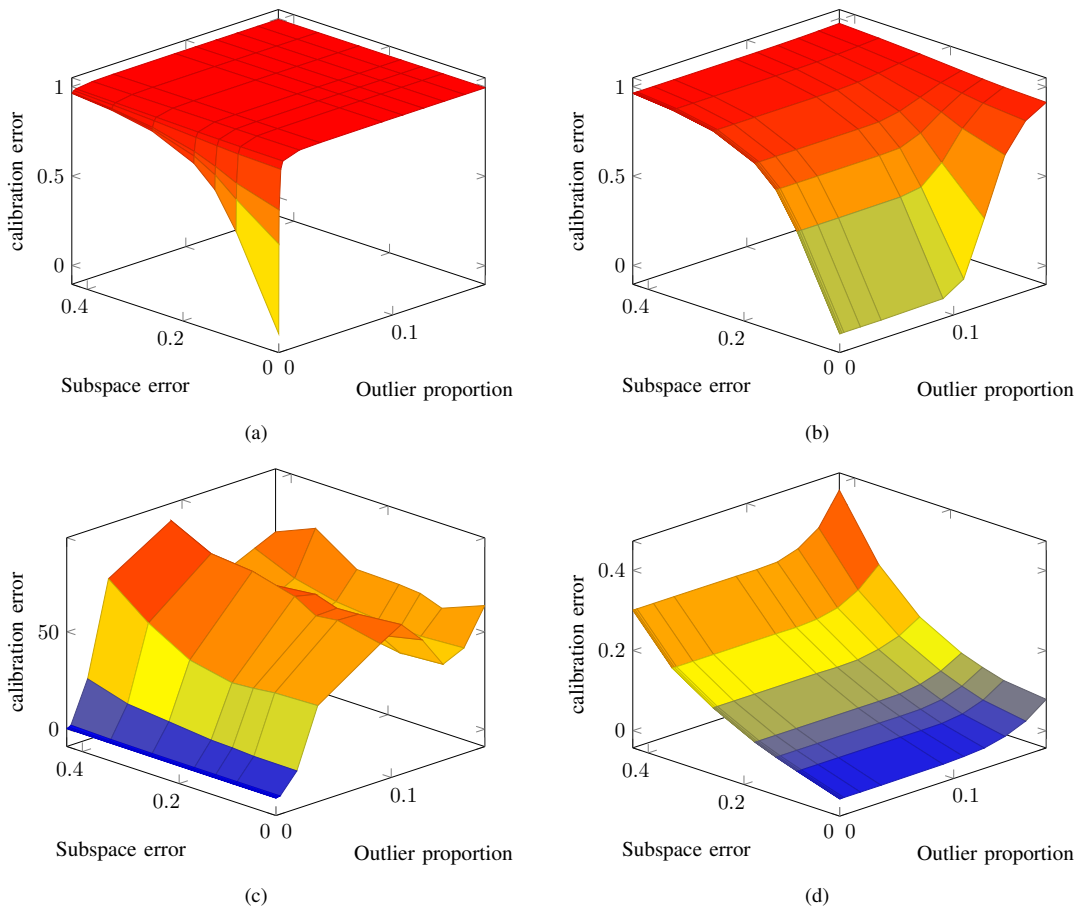


Fig. 2. Performance of the nullspace-based calibration methods with respect to the subspace estimation error ϵ_S and the outlier proportion ρ_o . The nullspace in Eq. (11) is estimated using: (a) LS [15], (b) RPCA preprocessing + LS, (c) TLS [16], and (d) RPCA preprocessing + TLS.

present in the data. Indeed, in that case, the performance reached by the state-of-the-art calibration methods and their respective proposed extensions are almost equal. When the outlier proportion is higher than 20% (the situation is not plotted on Fig. 2), the outlier matrix O is not very sparse and the IALM-RPCA method fails in accurately estimating it, thus leading to a degradation of the calibration performance.

We also investigated the case of partially blind (PB) calibration. We reproduced the same experiment as above, except that we considered a set of 5 sensors to be perfectly calibrated, i.e., their calibration parameters are assumed to be known. Due to space consideration, we cannot show the plots on the paper but the performance reached by the state-of-the-art and the proposed methods are consistent with the blind case. Indeed, the PB state-of-the-art techniques using LS and TLS show the same sensitivity to outliers than the blind LS and TLS approached tested above. The use of well calibrated sensors is then not sufficient to make these methods robust to outliers. Using RPCA as a preprocessing for PB calibration methods allows some robustness to the presence of up to 10% of outliers while it does not affect the performance of the methods in outlier-free simulations.

It should be noticed that in the nullspace-based calibration

problem, the rank of Y is assumed to be known, as it was estimated with \mathcal{S} . This knowledge is not fully taken into account in the IALM-RPCA algorithm, as it aims to find a matrix of minimal rank but it does not take into account the known rank $r + 1$ of matrix Y . Keeping the rank equality constraint in RPCA might allow to filter additive noise in the sensor readings, and thus to improve the performance reached by the proposed techniques.

VI. CONCLUSION AND PERSPECTIVES

In this paper, we proposed a robust extension of nullspace-based blind calibration methods. We reordered the collected data in a low-rank matrix that we filtered using an RPCA algorithm. While not adding any extra-assumption, such a preprocessing stage allows removing up to 10% of outliers without any loss of calibration performance. This work opens many perspectives. We recently proposed some mobile sensor calibration techniques based on informed matrix factorization [12], [13], [14]. In future work, we will investigate some outlier-robust extensions of these approaches, using a similar low-rank modeling. We will compare such a formalism to robust informed matrix factorization using parametric divergences [24], [25] or the Huber norm [26], that we recently

proposed for another application [27]. Moreover, the proposed techniques need to know the low-rank subspace where the sensed phenomenon lie. We will investigate the use of machine learning to learn S and to perform sensor calibration.

ACKNOWLEDGEMENT

This work was funded partly by the “OSCAR” project within the Région Hauts-de-France “Chercheurs Citoyens” Program, and partly by the CNRS MASTODONS project under the grant “DoMaSQ’Air 2017”.

REFERENCES

- [1] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, “Internet of things for smart cities,” *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
- [2] A. U. Rehman, A. Z. Abbasi, N. Islam, and Z. A. Shaikh, “A review of wireless sensors and networks’ applications in agriculture,” *Computer Standards & Interfaces*, vol. 36, no. 2, pp. 263–270, 2014.
- [3] M. Puigt, G. Delmaire, and G. Roussel, “Environmental signal processing: new trends and applications,” in *Proc. of ESANN*, 2017.
- [4] E. Cochran, J. Lawrence, A. Kaiser, B. Fry, A. Chung, and C. Christensen, “Comparison between low-cost and traditional MEMS accelerometers: a case study from the M7.1 Darfield, New Zealand, aftershock deployment,” *Annals of Geophysics*, vol. 54, no. 6, 2012.
- [5] K. Whitehouse and D. Culler, “Macro-calibration in sensor/actuator networks,” *Mobile Networks and Applications*, vol. 8, no. 4, pp. 463–472, Aug. 2003.
- [6] O. Saukh, D. Hasenfratz, C. Walser, and L. Thiele, “On rendezvous in mobile sensing networks,” in *Proc. of REALWSN*, 2014, vol. 281 of LNCS, pp. 29–42.
- [7] C. Wang, P. Ramanathan, and K.K. Saluja, “Moments based blind calibration in mobile sensor networks,” in *Proc. of ICC ’08*, May 2008, pp. 896–900.
- [8] O. Saukh, D. Hasenfratz, and L. Thiele, “Reducing multi-hop calibration errors in large-scale mobile sensor networks,” in *Proc. of IPSN*, 2015.
- [9] E. Miluzzo, N.D. Lane, A.T. Campbell, and R. Olfati-Saber, “CaliBree: A self-calibration system for mobile sensor networks,” in *Proc. of DCOSS*, 2008, vol. 5067 of LNCS, pp. 314–331.
- [10] B.-T. Lee, S.-C. Son, and K. Kang, “A blind calibration scheme exploiting mutual calibration relationships for a dense mobile sensor network,” *IEEE Sensors Journal*, vol. 14, no. 5, pp. 1518–1526, May 2014.
- [11] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, “Fast nonnegative matrix factorization and completion using nesterov iterations,” in *Proc. of LVA-ICA*, vol. 10179 of LNCS, pp. 26–35, 2017.
- [12] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, “Blind calibration of mobile sensors using informed nonnegative matrix factorization,” in *Proc. of LVA-ICA*, vol. 9237 of LNCS, pp. 497–505, 2015.
- [13] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, “Blind mobile sensor calibration using a nonnegative matrix factorization with a relaxed rendezvous model,” in *Proc. of ICASSP*, Mar. 2016, pp. 2941–2945.
- [14] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, “Nonlinear mobile sensor calibration using informed semi-nonnegative matrix factorization with a Vandermonde factor,” in *Proc. of SAM*, 2016.
- [15] L. Balzano and R. Nowak, “Blind calibration of sensor networks,” in *Proc. of IPSN*, 2007, pp. 79–88.
- [16] J. Lipor and L. Balzano, “Robust blind calibration via total least squares,” in *Proc. of ICASSP*, May 2014, pp. 4244–4248.
- [17] C. Schulke, F. Caltagirone, F. Krzakala, and L. Zdeborova, “Blind calibration in compressed sensing using message passing algorithms,” in *NIPS* 26, 2013, pp. 566–574.
- [18] C. Bilén, G. Puy, R. Gribonval, and L. Daudet, “Convex optimization approaches for blind sensor calibration using sparsity,” *IEEE Trans. on Signal Processing*, vol. 62, no. 18, pp. 4847–4856, Sept 2014.
- [19] H. Carfantan and J. Idier, “Statistical linear destriping of satellite-based pushbroom-type images,” *IEEE Trans. on Geoscience and Remote Sensing*, vol. 48, no. 4, pp. 1860–1871, April 2010.
- [20] Y. Zhang, N. Meratnia, and P. Havinga, “Outlier detection techniques for wireless sensor networks: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 12, no. 2, pp. 159–170, 2010.
- [21] M. Fazel, *Matrix rank minimization with applications*, Ph.D. thesis, Stanford University, 2002.
- [22] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.
- [23] Z. Lin, M. Chen, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” UIUC Technical Report UILU-ENG-09-2215, 2009.
- [24] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, “Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment,” in *Proc. of MLSP*, 2013.
- [25] R. Chreiky, G. Delmaire, C. Dorffer, M. Puigt, G. Roussel, and A. Abche, “Robust informed split gradient nmf using alpha beta-divergence for source apportionment,” in *Proc. of MLSP*, 2016.
- [26] R. Chreiky, G. Delmaire, M. Puigt, G. Roussel, and A. Abche, “Informed split gradient non-negative matrix factorization using huber cost function for source apportionment,” in *Proc. of ISSPIT*, 2016.
- [27] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, “Non-negative matrix factorization under equality constraints—a study of industrial source identification,” *Applied Numerical Mathematics*, vol. 85, pp. 1–15, November 2014.

Méthode robuste de NMF à base de divergence paramétrique et de contrainte de somme à un

Dans le cadre d'une autre application au sein de l'équipe, nous avons travaillé au développement d'une méthode de factorisation matricielle robuste, avec une fonction de coût à base de divergence paramétrique. Cette approche permet notamment d'effectuer la factorisation d'une matrice de données corrompue par des valeurs aberrantes, tout en incluant des données forcées dans les matrices facteur et une normalisation dure des matrices. Cette approche qui a donné lieu à la publication [31] présentée ci-après est une perspective intéressante pour renforcer la robustesse de nos méthodes d'étalonnage aux valeurs aberrantes.

Copyright

©2016 IEEE. Ré-imprimé avec l'autorisation des auteurs de : R. Chreiky, G. Delmaire, M. Puigt, G. Roussel, A. Abche, Outlier-robust calibration method for sensor networks, in Proceedings of the 16th IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), December 2016, 12-14th, Limassol, Cyprus.

ROBUST INFORMED SPLIT GRADIENT NMF USING $\alpha\beta$ -DIVERGENCE FOR SOURCE APPORTIONMENT

Robert Chreiky^{*†}, Gilles Delmaire^{*}, Clément Dorffer^{*}, Matthieu Puigt^{*}, Gilles Roussel^{*}, and Antoine Abche[†]

^{*} Univ. Littoral Côte d’Opale, LISIC – EA 4491, F-62228 Calais, France,

[†] University of Balamand, Balamand Al Kurah, Lebanon

ABSTRACT

Source apportionment is a very challenging topic for which non-negative source separation is well-suited. Recently, we proposed several informed Non-negative Matrix Factorization (NMF) for which some expert knowledge was introduced. These methods were all dealing with some set values of one factor together with the row sum-to-one property by either processing each constraint alternatingly or using a new parameterization which involves all of them. However, this last method was sensitive to the presence of outliers. In this paper, we thus propose a new robust informed Split Gradient NMF method which is based on a weighted $\alpha\beta$ -divergence cost function. Experiments conducted for several input SNR with and without outliers on simulated mixtures of particulate matter sources show the relevance of the new approach.

Index Terms— Informed Source Separation, Non-negative Matrix Factorization, Split Gradient, Source apportionment, $\alpha\beta$ -divergence

1. INTRODUCTION

Source apportionment consists of estimating the particulate matter sources and their contributions in the ambient air. A source is fully described by a *profile* which stands for a kind of fingerprint of the source by gathering m chemical species proportions (expressed in ng/ng). The observed particulate matter concentration obtained through chemical analysis may be written as a non-negative linear combination of p profiles. In practice, chemists collect n concentration samples which are mathematically modeled by a $n \times m$ non-negative data matrix X . Denoting G and F , the $n \times p$ contribution matrix and $p \times m$ profile matrix, respectively, we derive

$$X \approx G \cdot F, \quad (1)$$

which can be solved using Non-negative Matrix Factorization (NMF). Moreover, due to the nature of the source apportionment problem, the rows of F sum to one.

Sum-to-one constraints are often met in NMF—e.g., in remote sensing [1–3]—but those constraints are applied on

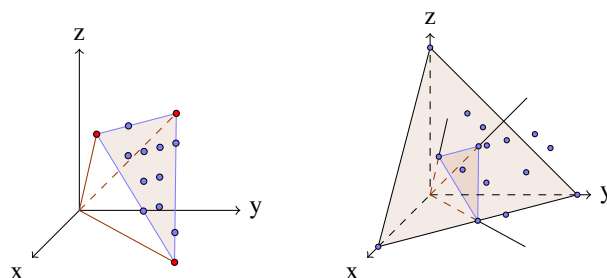


Fig. 1. Geometric point of view for (Left) hyperspectral data, (Right) Chemical data

the columns of the second factor, namely the mixing coefficients. This implies that the data points of X lie in a polytope of the non-negative orthant, as shown with 3 endmembers on the left plot of Fig. 1. In that case, searching for the mixing matrix may be split into independent sub-problems related to one column in which the non-negativity and sum-to-one apply. Then, most algorithms are based on running independent Fully Constrained Least Squares (FCLS) algorithms [4]. In source pollution, the situation is quite different. The sum-to-one constraint is applied on the rows of F , which implies that the data points belong to a convex hull of a simplicial cone in the non-negative orthant spanned by the different source profiles, as shown on the right plot of Fig. 1. As a consequence, splitting the NMF problem into independent sub-problems is impossible since the constraints apply along the rows.

Moreover, source apportionment is known to be a difficult task, as the data are noisy and possibly corrupted. Equation (1) is usually solved by an alternating Frobenius norm minimization under the non-negativity of all the entries in X , G , and F [5]. It is well-known that such a norm is well-suited to the case of additive zero-mean Gaussian noise while failing in the presence of even a single atypical observation. As a result, robust methods are designed in order to withstand a certain fraction of contamination named outliers. The latter may be defined as data with unusual statistical properties with respect to the majority of the available data. While “leverage points” and “vertical outliers” are met in robust

regression [6], only vertical outliers—corresponding to corrupted data points in the observed matrix—are met in NMF. In source apportionment, they usually come from a contamination during the measurement process but may also occur during the transmission process. The methods able to deal with outliers should be split into two main families. While some papers, e.g., [3, 7, 8], assume that X may be decomposed as the sum of a low-rank term L —which can then be factorized as in Eq. (1)—and a sparse matrix S —which contains the outliers present in the collected data—some authors propose instead modified cost functions, which provide less penalization to large residuals than the Frobenius norm, e.g., $\ell_{2,1}$ [9], $\ell_1 - \ell_2$ [10] or Huber [11] norms. An alternative consists of using some parametric divergences, e.g., α - [12], β - [13, 14], or $\alpha\beta$ -divergences [15], which are tuned by the value of α and β .

Lastly, classical NMF outputs in atmospheric pollution often lack physical meaning [16], probably because some rows of the profile matrix are geometrically close to each other. Moreover, we found in [17] that even weighted NMF using β -divergence do not provide satisfactory results. To tackle such an issue, we recently proposed some *informed* NMF methods in, e.g., [17–19]. Indeed, chemical experts are able to set some values of the profile matrix F , which can help the NMF to converge to an interpretable result. Such constraints should be applied together with the above row sum-to-one constraint. In [17, 18], these constraints were alternately satisfied within iterations and the resulting methods were shown to provide a good performance. However, we also found that some specific set-value constraints may provide poor results. In our previous work [19], we proposed a solution which always lives in the space of both constraints and which is less sensitive to the set constraint location than [17, 18]. However, the method in [19] is based on the Frobenius norm, which is not robust to outliers as explained above. Indeed, in that case, some outlying points may lie outside the real convex hull (see Fig. 1) and classical NMF may be affected by these outliers.

As a consequence we here extend [19] and we propose a new informed NMF approach using $\alpha\beta$ -divergence. The proposed method thus (i) incorporates some known values of the profile matrix while satisfying the sum-to-one constraints at each iteration, and (ii) uses a robust parametric cost function to deal with outliers.

The remainder of the paper is structured as follows. In Section 2, we introduce the NMF parameterization used in this paper. Section 3 is devoted to the expression of a weighted $\alpha\beta$ -divergence for NMF. We then introduce the updates rules of the proposed method in Section 4 while its performance is investigated in Section 5. Lastly, Appendix A provides the proof that the third KKT is satisfied by our method.

2. PROBLEM PARAMETERIZATION

As explained above, we assume in this paper that some entries of the profile matrix are known as they are provided by experts. Such a knowledge can be taken into account as a specific parameterization proposed in [18], that we briefly recall hereafter. We define Ω as a $p \times m$ binary matrix which indicates the presence or the absence of constraints on each element f_{ij} of the matrix F , i.e.,

$$\Omega_{ij} = \begin{cases} 1 & \text{if } f_{ij} \text{ is known,} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Moreover, we define $\bar{\Omega}$ as the $p \times m$ binary matrix which reads

$$\bar{\Omega} \triangleq 1_{pm} - \Omega, \quad (3)$$

where 1_{pm} is the $p \times m$ matrix of ones. F then reads

$$F = \Omega \circ \Phi + \bar{\Omega} \circ \Delta F, \quad (4)$$

where \circ denotes the elementwise matrix multiplication, Φ and ΔF are both $p \times m$ matrices of set and free values, respectively, i.e., $\Phi \triangleq F \circ \Omega$ and $\Delta F \triangleq F - \Phi \circ \Omega$. It should be noticed that the parameterization (4) can also be applied to G , as shown in, e.g., [20, 21] for another application.

However, it should be noticed that Eq. (4) does not allow to take into account the row sum-to-one constraint. To this end, we proposed in [19] a novel parameterization inspired by the work in [22], i.e.,

$$F = \Omega \circ \Phi + \frac{\bar{\Omega} \circ Z}{(\bar{\Omega} \circ Z) \cdot 1_{mm}} \cdot (1_{pm} - \Phi \cdot 1_{mm}). \quad (5)$$

where the fraction denotes elementwise division. It should be noticed that Z is a non-negative matrix with no row sum-to-one constraint. The sum-to-one property is taken into account by the denominator of the ratio in Eq. (5) which is weighted by the total value of the free part, i.e., $1_{pm} - \Phi \cdot 1_{mm}$. As an example, let F be a 1×4 profile matrix and $\Phi = [0, 0.1, 0, 0.3]$. The maximum free part value is then $1_{pm} - \Phi \cdot 1_{mm} = [0.6, 0.6, 0.6, 0.6]$ and Eq. (5) reads

$$F = \left[\frac{Z_{11} \cdot 0.6}{Z_{11} + Z_{13}}, 0.1, \frac{Z_{13} \cdot 0.6}{Z_{11} + Z_{13}}, 0.3 \right].$$

As for Eq. (4), it should be noticed that a similar parameterization may be applied to G . In that case, the sum of each row of G should be equal to the sum of each row of X . Such a strategy was proposed in [22] for non-informed NMF. However, it should be noticed that in the application considered in this paper, such a parameterization is not applicable since it is valid only in an exact factorization framework. Indeed, while the parameterization (5) is derived from the definition of the profile matrix, only an approximate row sum constraint may be applied to G . In addition, outliers may affect even more this relationship. This problem is addressed below in the general framework of the parametric $\alpha\beta$ -divergence cost function, which is well suited for managing outliers and which encompasses several classical cost functions.

3. WEIGHTED $\alpha\beta$ -DIVERGENCE

Many cost functions may be designed to cope with outliers. They can rely on modified norms of the matrix of residuals, thus leading to the same output for opposite residual entries. Another choice is to choose a parametric divergence tuned by several parameters, thus enabling to penalize in a different way a positive or negative residual. This is of particular interest in source apportionment, where the sample data may be corrupted afterwards by an additional contamination. This situation corresponds to the case of positive outliers only.

The $\alpha\beta$ -divergence function is a parametric asymmetric discrepancy measure which may be used to evaluate the gap between two matrices P and Q by evaluating the corresponding divergence of each entry, i.e., $\forall(\alpha, \beta, \alpha + \beta) \neq 0$,

$$\mathcal{D}^{\alpha, \beta}(P||Q) = \sum_{i,j} \frac{-1}{\alpha\beta} (P_{ij}^\alpha Q_{ij}^\beta - \frac{\alpha}{\alpha + \beta} P_{ij}^{\alpha+\beta} - \frac{\beta}{\alpha + \beta} Q_{ij}^{\alpha+\beta}). \quad (6)$$

The classical Frobenius norm may be viewed as a special case of the $\alpha\beta$ -divergence by choosing $\alpha = 1$ and $\beta = 1$. It underlies a real-valued Gaussian distribution whereas the data are only non-negative. This fact promotes the use of different parameters, particularly, some special parameters lead to very famous divergences measures such as α -divergences [15] or β -divergences [13, 14]. This divergence is used for example in clustering or non-negative matrix factorization [15]. It has many advantages: firstly, it may be differentiated with respect to the estimation, and secondly it remains convex for a wide range of parameters. Moreover, its asymmetry enables to give different penalizations to a negative or positive gap. Additionally, by introducing a weight matrix W , an individual weight may also be given to each divergence entry [23, 24] as

$$\mathcal{D}_W^{\alpha, \beta}(P||Q) = W \circ \mathcal{D}^{\alpha, \beta}(P||Q). \quad (7)$$

In analytic chemistry, measurements x_{ij} are provided with absolute uncertainties σ_{ij} . In the case of an $\alpha\beta$ -divergence, the entries of the weight matrix W read $w_{ij} \triangleq \sigma_{ij}^{-(\alpha+\beta)}$. Such a weighted divergence is used below to derive the update rules of the proposed informed NMF method.

4. PROPOSED INFORMED SPLIT GRADIENT NMF

In this section, we aim to derive the update rules for G and F . It should be noticed that such rules may be obtained from the differentiation of a cost function \mathcal{J} defined as

$$\begin{aligned} \mathcal{J} &\triangleq \mathcal{D}_W^{\alpha, \beta}(X||G \cdot F) \\ \text{s.t. } F &= \Omega \circ \Phi + \frac{\bar{\Omega} \circ Z}{(\bar{\Omega} \circ Z) \cdot 1_{mm}} \cdot (1_{pm} - \Phi \cdot 1_{mm}). \end{aligned} \quad (8)$$

4.1. Update rules for G

As explained above, G is unconstrained in the considered application. As a consequence, the update rules, derived Eq. (8),

should only take into account the weight matrix W . For that, we extend the work in [23, 24] from one hand and in [15] from the other hand, to a Weighted $\alpha\beta$ -NMF ($\alpha\beta$ -WNMF) which yields the following update rules for G :

$$G^{k+1} = G^k \circ \left[\frac{(W \circ (X^\alpha \circ (GF)^{\beta-1})) F^T}{(W \circ (GF)^{\alpha+\beta-1}) F^T} \right]^{(\frac{1}{\alpha})}. \quad (9)$$

These rules encompass several other methods, especially $\alpha\beta$ -NMF [15] if $W = 1_{nm}$, e.g., for any i and j , $w_{ij} = 1$, and β -NMF [13] if additionally $\alpha = 1$. This divergence is able to deal with outliers since it can accept an asymmetric gap between the data and the estimated ones [15].

4.2. Update rules for F

In this subsection, we introduce the proposed update rules for the profile matrix F . we aim to differentiate Eq. (8) with respect to F , so that we can propose heuristic rules based on the KKT conditions which satisfy the expected properties of the solutions. Because of space limitations, only the main steps of the proof are provided below. The differentiation of \mathcal{J} with respect to F may be performed by expressing \mathcal{J} in term of several matrix traces, which leads to

$$\frac{\partial \mathcal{J}}{\partial F} = \frac{1}{\alpha} G^T [W \circ (G \cdot F)^{\beta-1} \circ ((G \cdot F)^\alpha - X^\alpha)] \quad (10)$$

From the parameterization (5), F_{ij} —the (i, j) -th entry of F —only depends on the i -th row of Z —i.e., on Z_{ik} , $\forall k \in \{1 \dots m\}$ —and its differentiation may then be expressed with respect to Z_{ik} , hence providing

$$\frac{\partial F_{ij}}{\partial Z_{ik}} = (1 - \sum_{l=1}^m \Phi_{il}) \frac{\bar{\Omega}_{ij}}{\sum_{l=1}^m \bar{\Omega}_{il} Z_{il}} \left[\delta_{jk} - \frac{\bar{\Omega}_{ik} Z_{ij}}{\sum_{l=1}^m \bar{\Omega}_{il} Z_{il}} \right]. \quad (11)$$

where δ_{jk} is the Kronecker function which is equal to 1 when $j = k$. We then define B the $p \times m$ strictly negative matrix whose entries b_{ik} read

$$b_{ik} \triangleq \frac{1 - \sum_{l=1}^m \Phi_{il}}{(\sum_{l=1}^m Z_{il} \bar{\Omega}_{il})}. \quad (12)$$

Combining Eqs. (10), (11), and (12), we derive the matrix form of the differentiation of \mathcal{J} with respect to Z , i.e.,

$$\frac{\partial \mathcal{J}}{\partial Z} = B \circ \left[\frac{\partial \mathcal{J}}{\partial F} \circ \bar{\Omega} - \frac{\bar{\Omega}}{1_{pm} - \Phi \cdot 1_{mm}} \circ \left(\left(\frac{\partial \mathcal{J}}{\partial F} \circ F \circ \bar{\Omega} \right) \cdot 1_{mm} \right) \right]. \quad (13)$$

The third KKT condition allows to express a general condition to obtain a stationary point w. r. t. Z , as $Z \circ \frac{\partial \mathcal{J}}{\partial Z} = 0$. This condition may be expressed with respect to F , i.e.,

$$F \circ \bar{\Omega} \circ \frac{\partial \mathcal{J}}{\partial Z} = 0. \quad (14)$$

In the case of a stationary point ($F = F^k = F^{k+1}$), combining Eqs. (13) and (14) leads to the following general condition

$$\bar{\Omega} \circ F^k \circ U - \frac{(U \circ \bar{\Omega} \circ F^k) \cdot 1_{m \times m}}{1_{p \times m} - \Phi \cdot 1_{m \times m}} \circ F^{k+1} \circ \bar{\Omega} = 0, \quad (15)$$

where $U \triangleq -\frac{\partial \mathcal{J}}{\partial F}$. From Eq. (15), we derive the update rules for F , e.g.,

$$F^{k+1} = \Omega \circ \Phi + \bar{\Omega} \circ \frac{F^k \circ U}{[U \circ \bar{\Omega} \circ F^k] \cdot 1_{mm}} \circ (1_{pm} - \Phi \cdot 1_{mm}). \quad (16)$$

By forgetting the scalar factor in Eq. (10)—due to the presence of U in both the denominator and numerator— U may be written as

$$U = G^T(W \circ (X^\alpha \circ (GF)^{\beta-1})) - G^T(W \circ (GF)^{\alpha+\beta-1}). \quad (17)$$

The general form of the update rules given in Eq. (16) fulfills the parameterization introduced in (5). However, the non-negativity of F should be preserved which is ensured when U is positive. The authors in [22] proposed to shift U by adding $\eta \cdot 1_{pm}$, where η is computed at each iteration. As an alternative, we here propose an analytic shift equal to the sum of the negative part of $-\frac{\partial \mathcal{J}}{\partial F}$ provided in Eq. (10). Thus, from Appendix A, the shifted matrix U_s is given by

$$U_s = G^T(W \circ (X^\alpha \circ (GF)^{\beta-1})) - G^T(W \circ (GF)^{\alpha+\beta-1}) + [G^T(W \circ (GF)^{\alpha+\beta-1})]1_{mm}. \quad (18)$$

Thus, the updated rules of the proposed $\alpha\beta$ -divergence Split-Gradient Method for Constrained Weighted NMF ($\alpha\beta$ -SG-CWNMF) consists of using U_s as defined in Eq. (18) instead of U in Eq. (16). These updated rules together with U_s satisfy (i) the non-negativity of F , (ii) the sum-to-one constraints, and (iii) the third KKT condition (see Appendix A).

5. SIMULATIONS

In this section, the performance of the newly proposed method is investigated by conducting several experiments with and without outliers. The data matrix X consists of 50 samples and 7 species listed in Table 1, with a known uncertainty measure σ_{ij} provided by a chemical expert and associated with each data point x_{ij} . The collected sampled values are firstly corrupted by a contamination simulated by one or several individual multiplicative outliers located at random locations and tuned by a gain. This gain is always greater than one in order to simulate a pollutant contamination. Secondly, the offline chemical analysis is modeled by an additive uniform noise within the interval $[-\min(\lambda\sigma_{ij}; x_{ij}); \lambda\sigma_{ij}]$ while preserving their positiveness. It is to be noted that λ is related to an input Signal-to-Noise Ratio (SNR) before

Table 1. Positions and values of the constraints used in the informed NMF methods. X means no constraint.

	Fe	Ca	SO ₄	Zn	Mg	Al	Cr
Source 1	0.7	X	X	X	0.04	X	X
Source 2	0.3	X	X	X	X	0.075	X
Source 3	X	0.2	X	X	0.12	X	X

outliers are applied. The profile matrix consists of three partially correlated industrial profiles. In addition, all the tested informed methods are using 6 set values, provided in Table 1. In the tests, the initialization for F is performed with the same expert's profiles which are consistent with the chosen constraints. The initial contribution is then computed with a weighted quadratic program. The performance is measured by using an average MER [25], computed on the columns of G , expressed in dB.

Four blind methods are experimented. Two of them are based on the β -divergence and are tested with $\beta = 0.8$, namely the β -NMF [13] and the β -WNMF [24]. The two others blind approaches are using a Frobenius cost function, i.e., the robust NMF (rNMF) [3] which put the outliers into a sparse matrix and a weighted version of the Split-Gradient NMF (SG-WNMF) proposed in [22]. Moreover, two state-of-the-art informed approaches are tested, i.e., our previous informed β -NMF methods [17] denoted β -CWNMF and β -CWNMF-R (but we only show the best achieved performance in these tests, i.e., the one obtained with β -CWNMF-R). We compare all these methods with our new Split Gradient Constrained Weighted NMF (denoted $\alpha\beta$ -SGCWNMF) for which we set $\alpha = 0.6$ and $\beta = 0.6$. In addition, we show the performance obtained on $\alpha\beta$ -WNMF when F is set to the true profile matrix (the corresponding plot is denoted “Bound” in the figures). In order to evaluate all these methods, we run 400 tests for each plot in Fig. 2 with various input Signal-to-Noise Ratios (SNRs) ranging from 15 to 75 dB. However, we limit the representation of the x axis in Figure 2 to 55 dB for space considerations. The three plots in Figure 2 show the reached MERs with respect to the input SNR. In these plots, the Bound is roughly equal to the input SNR. We investigate the case without outliers (left plot), and with 5 outliers (right plot). Please note that the shown SNR in these cases stand for the SNRs computed after setting several outliers. In the case of large SNRs, this means that several small outliers coexist with very low noise. On the contrary, low SNRs may be obtained by medium size outliers with large noise. It turns out in Fig. 2 that the rNMF approach [3] is always failing. This is probably due to the fact that in source apportionment, an outlier is not necessarily of large amplitude (it is outside the range provided by the uncertainty). In that case, the estimation of the sparse matrix of outliers completely fails in rNMF, thus providing a poor MER performance. Our new method $\alpha\beta$ -SGCWNMF outperforms all the other approaches in ev-

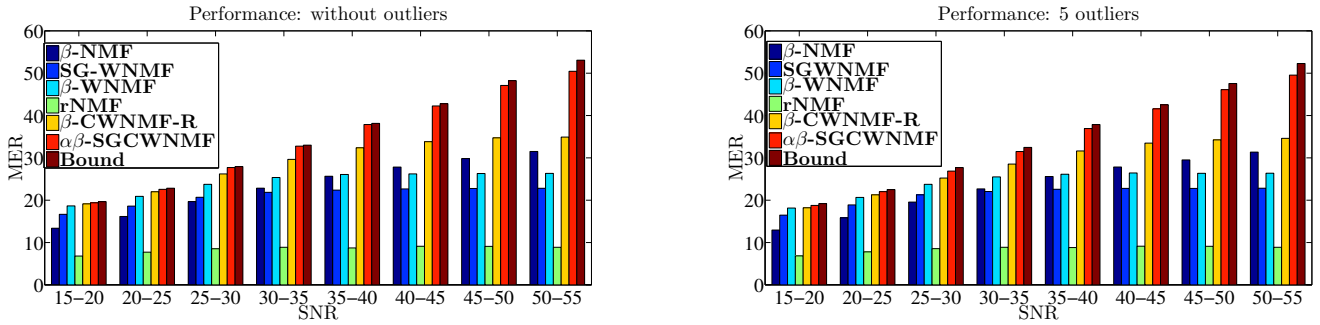


Fig. 2. MERs provided by the tested NMF methods vs. the input SNR: (left) no outlier, and (right) 5 outliers.

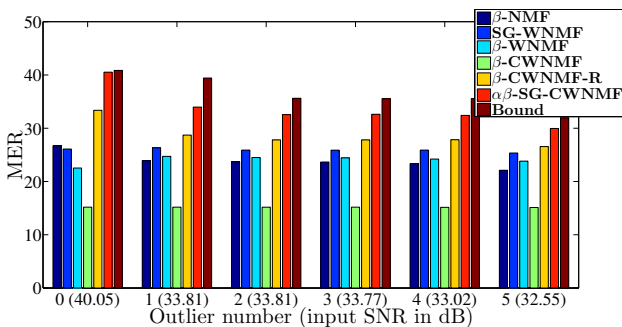


Fig. 3. Performance of the tested NMF methods vs. the number of outliers. Perf. Criterion: MER (in dB).

ery slice of SNR. Moreover, it remains very close to the computed Bound with and without outliers, thus showing its relevance. In order to confirm this behaviour, we set the input SNR before outliers to 40 dB and we investigate the influence of the number of outliers on the performance of all methods. The multiplicative outlier factor is chosen such that it enables to increase the unnoisy entry data by 30 percent. Figure 3 shows the achieved MER versus the outlier number, associated with the SNR after outliers inside brackets. We notice that all methods provide a decreasing performance when the number of outliers increases. Our $\alpha\beta$ -SGCWNMF outperforms all the others and remains very close to the Bound obtained through regression on G . To conclude, our new method seems to bring an important enhancement in the performance even in the presence of a few outliers.

6. CONCLUSION

In this paper, a new informed NMF method is proposed for the source apportionment problem. It relies on a weighted $\alpha\beta$ -divergence able to resist to medium size asymmetric outliers. It is also based on a previously introduced parameterization of the profile matrix which enables to fulfill automatically both the set values and the row sum-to-one constraint. Then, a general form of the update rules are derived heuristically by using

the differentiation of the criterion with respect to the newly introduced matrix and the third KKT condition. By proposing a new analytic shift, it yields new analytic expressions of the update rules. This method is extensively compared with state-of-the-art methods and seems to outperform all the others in every slice of input SNR and for all the tested numbers of outliers. In addition, our method remains very close to the computed bound obtained with the exact profile matrix, even in the presence of outliers.

7. REFERENCES

- [1] E. Chouzenoux, S. Moussaoui, M. Legendre, and J. Idier, "Algorithme primal-dual de points intérieurs pour l'estimation pénalisée des cartes d'abondances en imagerie hyperspectrale," *Traitement du Signal*, 2013.
- [2] W.-K. Ma, J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C.-Y. Chi, "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Proc. Mag.*, pp. 67–81, Jan. 2014.
- [3] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Processing*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.
- [4] D. C. Heinz and C. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 529–545, 2001.
- [5] D. D. Lee and H. S. Seung, "Learning the parts of objects by non negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [6] P. Filzmoser, S. Serneels, R. Maronna, and P.J. Van Espen, *Robust multivariate methods in Chemometrics*, In B. Walczak, R.T. Ferre, and S. Brown, editors, *Comprehensive Chemometrics*, 2nd edition, 2009.

- [7] B. Shen, L. Si, R. Ji, and B. Liu, “Robust nonnegative matrix factorization via l_1 norm regularization,” *arXiv preprint arXiv:1204.2311*, 2012.
- [8] J. Rapin, J. Bobin, A. Larue, and J. L. Starck, “Sparse and non-negative BSS for noisy data,” *IEEE Trans. on Signal Processing*, vol. 61, no. 22, pp. 5620–5632, 2013.
- [9] D. Kong, C. Ding, and H. Huang, “Robust non-negative matrix factorization using l_{21} -norm,” in *Proc. of CIKM*, 2011, pp. 673–682.
- [10] A. Ben Hamza and D. J. Brady, “Reconstruction of reflectance spectra using robust nonnegative matrix factorization,” *IEEE Trans. on Signal Processing*, vol. 54, no. 9, pp. 3637–3642, 2006.
- [11] L. Du, X. Li, and Y.-D. Shen, “Robust non-negative matrix factorization via half-quadratic minimization,” in *Proc. of ICDM*. 2012, pp. 201–210, Springer.
- [12] A. Cichocki, H. Lee, Y. Kim, and S. Choi, “Nonnegative matrix factorization with alpha-divergence,” *Pattern Recog. Letters*, vol. 29, no. 9, pp. 1433–1440, 2008.
- [13] C. Févotte and J. Idier, “Algorithms for nonnegative matrix factorization with the beta-divergence,” *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011.
- [14] D. L. Sun and C. Févotte, “Alternating direction method of multipliers for non-negative matrix factorization with the beta-divergence,” in *Proc. of ICASSP*, May 2014.
- [15] A. Cichocki, S. Cruces, and S. Amari, “Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization,” *Entropy*, vol. 13, pp. 134–170, 2011.
- [16] M. Viana, A. Pandolfi, M. C. Minguillo, X. Querol, A. Alastuey, E. Monfort, and I. Celades, “Inter-comparison of receptor models for PM source apportionment: Case study in an industrial area,” *Atmospheric Environment*, vol. 42, pp. 3820–3832, 2008.
- [17] A. Limem, M. Puigt, G. Delmaire, G. Roussel, and D. Courcot, “Bound constrained weighted NMF for industrial source apportionment,” in *Proc. of MLSP*, 2014.
- [18] A. Limem, G. Delmaire, M. Puigt, G. Roussel, and D. Courcot, “Non-negative matrix factorization under equality constraints—a study of industrial source identification,” *Applied Numerical Mathematics*, vol. 85, pp. 1–15, November 2014.
- [19] R. Chreiky, G. Delmaire, M. Puigt, G. Roussel, and A. Abche, “Split gradient method for informed non-negative matrix factorization,” in *Proc. of LVA-ICA*, 2015, pp. 376–383.
- [20] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, “Blind calibration of mobile sensors using informed nonnegative matrix factorization,” in *Proc. of LVA-ICA*, vol. 9237 of *LNCS*, pp. 497–505. 2015.
- [21] C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, “Blind mobile sensor calibration using a nonnegative matrix factorization with a relaxed rendezvous model,” in *Proc. of ICASSP*, Mar. 2016.
- [22] H. Lantéri, C. Theys, C. Richard, and C. Févotte, “Split gradient method for nonnegative matrix factorization,” in *Proc. of EUSIPCO*, 2010.
- [23] D. Guillaumet, J. Vitria, and B. Schiele, “Introducing a weighted non-negative matrix factorization for image classification,” *Pattern Recognition Letters*, vol. 24, no. 14, pp. 2447–2454, 2003.
- [24] N.-D. Ho, *Non negative matrix factorization algorithms and applications*, Phd thesis, Univ. Catholique de Louvain, 2008.
- [25] E. Vincent, S. Araki, and P. Bofill, “The 2008 signal separation evaluation campaign: A community-based approach to large-scale evaluation,” in *Proc. of ICA*, 2009, pp. 734–741.

A. PROOF OF THE KKT CONDITION WITH THE NEW MATRIX U_S

Let U be defined in Eq. (17).

Theorem 1. *Every U_s of the form $U_s = U + S \cdot 1_{mm}$ fulfills the third KKT condition (15).*

Provided that U satisfies the condition (15), it amounts to check the necessary condition at convergence,

$$\bar{\Omega} \circ F^k \circ (S \cdot 1_{mm}) - \frac{((S \cdot 1_{mm}) \circ \bar{\Omega} \circ F^k) \cdot 1_{mm}}{1_{pm} - \Phi \cdot 1_{mm}} \circ F^{k+1} \circ \bar{\Omega} = 0. \quad (19)$$

By using the property

$$((S \cdot 1_{mm}) \circ B) \circ 1_{mm} = (S \cdot 1_{mm}) \circ (B \cdot 1_{mm}), \quad (20)$$

it turns out that

$$\bar{\Omega} \circ F^k \circ (S \cdot 1_{mm}) - \frac{(S \cdot 1_{mm}) \circ ((\bar{\Omega} \circ F^k) \cdot 1_{mm})}{1_{pm} - \Phi \cdot 1_{mm}} \circ F^{k+1} \circ \bar{\Omega} = 0. \quad (21)$$

Knowing that $(\bar{\Omega} \circ F^k) \cdot 1_{mm} = 1_{pm} - \Phi \cdot 1_{mm}$, it leads that the necessary condition to get a stationary points reads

$$\bar{\Omega} \circ F^k \circ (S \cdot 1_{mm}) - (S \cdot 1_{mm}) \circ F^{k+1} \circ \bar{\Omega} = 0 \quad (22)$$

Given that F reaches a limit point ($F^k = F^{k+1}$), the necessary condition (22) is always fulfilled. This completes the proof.

In addition, if we choose $S \triangleq G^T (W \circ (GF)^{\alpha+\beta-1})$, then U_s given in Eq (15) is solution of the third KKT condition.

Méthode accélérée de NMF avec données manquantes

Dans cette thèse, nous avons principalement utilisé des approches de NMF multiplicatives ou à base de descente de gradient. Dans le cas des méthodes multiplicatives, celles-ci sont connues pour leur grande simplicité de mise en pratique, cependant, il est également connu que ces méthodes sont lentes à converger, en particulier si les matrices recherchées sont de grande dimension. Dans le cas des approches par descente de gradient, la vitesse de convergence dépend fortement du pas de descente choisi, et sa sélection n'est pas toujours triviale. Par ailleurs, les méthodes qui optimisent la recherche de ce pas font une recherche par ligne, qui est coûteuse, ce qui rend l'approche peu applicable aux problèmes de grande dimensions.

Aujourd'hui, la NMF à base de gradient de Nesterov (NeNMF) est connue pour être parmi les méthodes les plus rapides [193]. Cependant, dans sa version classique [76], la NeNMF ne permet pas de tenir compte de données manquantes dans la matrice à factoriser.

En perspective de l'application de nos approches d'étalonnage à des jeux de données de grandes dimensions, nous avons développé deux méthodes de factorisation matricielle accélérées basées sur l'approche de Nesterov qui permet de gérer la possible absence de certaines entrées dans la matrice de données. L'une de ces approches permet d'inclure une pondération dans la méthode de NeNMF tandis que la seconde combine une stratégie de EM (pour *Expectation-Maximization* en anglais) avec le gradient de Nesterov. Ces deux méthodes sont présentées en détails dans la publication [52] présentée ci-après.

Copyright

©Reprinted by permission from Springer Nature : C. Dorffer, M. Puigt, G. Delmaire, and G. Roussel, "*Fast nonnegative matrix factorization and completion using Nesterov iterations*", in Proceedings of the 13th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2017), Lecture Notes in Computer Sciences (LNCS), vol. 10179, pp. 25-36, Copyright Springer International Publishing AG, 2017.

Fast Nonnegative Matrix Factorization and Completion using Nesterov Iterations

Clément Dorffer, Matthieu Puigt, Gilles Delmaire, and Gilles Roussel

Univ. Littoral Côte d’Opale, LISIC – EA 4491, F-62228 Calais, France
clement.dorffer@univ-littoral.fr, matthieu.puigt@univ-littoral.fr
gilles.delmaire@univ-littoral.fr, gilles.roussel@univ-littoral.fr

Abstract. In this paper, we aim to extend Nonnegative Matrix Factorization with Nesterov iterations (Ne-NMF)—well-suited to large-scale problems—to the situation when some entries are missing in the observed matrix. In particular, we investigate the Weighted and Expectation-Maximization strategies which both provide a way to process missing data. We derive their associated extensions named W-NeNMF and EM-W-NeNMF, respectively. The proposed approaches are then tested on simulated nonnegative low-rank matrix completion problems where the EM-W-NeNMF is shown to outperform state-of-the-art methods and the W-NeNMF technique.

Keywords: Low-rank matrix completion; nonnegative matrix factorization; Nesterov iterations; gradient descent

1 Introduction

Estimating missing entries in a low-rank matrix—known as Matrix Completion (MC)—received a lot of attention since some pioneering work in [8, 2]. Indeed, such a problem finds many applications [1] including environmental monitoring [17, 26] for example. Mathematically, it consists of estimating a matrix $M \in \mathbb{R}^{m \times n}$ —partially known on a subset Ω of entries—by a low-rank matrix X , i.e.,

$$\min_X \text{rank}(X) \quad \text{s.t.} \quad \mathcal{P}_\Omega(X) = \mathcal{P}_\Omega(M), \quad (1)$$

where $\mathcal{P}_\Omega(\cdot)$ is the sampling operator of M .

Among the many approaches recently proposed to solve MC, matrix factorization techniques emerged as a powerful tool, especially when the rank of the data matrix M is known [11, 31]. Moreover in some applications—e.g., blind sensor calibration [5, 6]— M is nonnegative, thus yielding a Nonnegative MC (NMC) problem which can be tackled by Nonnegative Matrix Factorization (NMF). NMF is widely used in, e.g., signal processing or machine learning, and consists of finding two nonnegative matrices $G \in \mathbb{R}_+^{m \times r}$ and $F \in \mathbb{R}_+^{r \times n}$ such that

$$X \approx G \cdot F. \quad (2)$$

Most NMF algorithms consider two separate convex sub-problems that respectively and alternately estimate G for a fixed F and F for a fixed G . Such estimates are obtained using, e.g., Multiplicative Updates (MU) [19], Alternating Nonnegative Least Squares (ANLS) [15], or Projected Gradient (PG) [22]. Additionally, some authors incorporated some extra-information in the NMF model [30], e.g., weights [13, 10], sum-to-one constraints [18, 21, 4], sparsity assumptions [14, 6, 12], or known and/or bounded values [22, 3, 20, 21, 5, 6, 12]. With the Big Data era, computational time reduction of NMF is particularly investigated, e.g., through optimal solvers [9], distributed strategies [23], online estimation [24], or randomization [33, 29].

When applied to a data matrix M which is partially known on a subset Ω of entries, NMF consists of estimating both matrices G and F which satisfy

$$\min_{G, F \geq 0} \|\mathcal{P}_\Omega(M) - \mathcal{P}_\Omega(G \cdot F)\|_{\mathcal{F}}, \quad (3)$$

where $\|\cdot\|_{\mathcal{F}}$ is the Frobenius norm. Solving Eq. (3) is usually done either by applying Weighted NMF (WNMF) [13, 10] to M or by using an Expectation-Maximization (EM) strategy [32, 16]. However, these approaches use MU [32, 13] or ANLS iterations [16], and are limited to relatively small-sized matrices. As NMC may be applied to very large-scale problems, efficient WNMF methods must be proposed. As a consequence, in this paper we investigate NMF with Nesterov iterations (Ne-NMF) [9] that we propose to extend to the NMC problem. Indeed, NeNMF is improving PG-NMF [22] by replacing the line search of the optimal step in the update rules—known to be time consuming—by Nesterov iterations [25]. NeNMF is thus much faster than MU-NMF or PG-NMF¹. Moreover, NeNMF can be easily extended to semi-NMF, which is of particular interest in some sensor calibration problems [7]. We thus propose to extend the NeNMF algorithm to deal with missing entries in a nonnegative data matrix M . In particular, we discuss the respective advantages and drawbacks of extensions using the above weighted and EM strategies. We experimentally show that the latter outperforms both the state-of-the-art methods and the proposed weighted extension in simulations of NMC problems.

The remainder of the paper is organized as follows. We recall the principles of the NeNMF method in Section 2. We then extend such an approach to the case of an observed matrix with missing entries in Section 3. The enhancement of the proposed methods is investigated in Section 4 while we conclude and discuss about future work in Section 5.

2 Standard NeNMF

As explained above, for a fixed $m \times n$ nonnegative data matrix X , NMF consists of finding both the $m \times r$ and $r \times n$ matrices G and F which provide the best

¹ See for example the CPU-time consumptions of [27] at <https://github.com/andrewssobral/lrslibrary>.

low-rank approximation of X , as shown in Eq. (2). The NeNMF method [9] iteratively and alternately updates G and F . For that, it solves

$$F = \arg \min_{\tilde{F} \geq 0} \mathcal{J}(G, \tilde{F}) = \arg \min_{\tilde{F} \geq 0} \frac{1}{2} \cdot \left\| X - G \cdot \tilde{F} \right\|_{\mathcal{F}}^2, \quad (4)$$

and

$$G = \arg \min_{\tilde{G} \geq 0} \mathcal{J}(\tilde{G}, F) = \arg \min_{\tilde{G} \geq 0} \frac{1}{2} \cdot \left\| X - \tilde{G} \cdot F \right\|_{\mathcal{F}}^2, \quad (5)$$

respectively, by applying in an inner loop the Nesterov accelerated gradient descent [25]. In order to update a factor, say F , the latter initializes $Y_0 \triangleq F^t$ —where t is an NeNMF outer iteration—and considers a series α_k defined as

$$\alpha_0 = 1, \quad \alpha_{k+1} = \frac{1 + \sqrt{4\alpha_k^2 + 1}}{2}, \quad \forall k \in \mathbb{N}. \quad (6)$$

For each inner loop index k , the Nesterov gradient descent then computes

$$F_k = \left(Y_k - \frac{1}{L} \nabla_F \mathcal{J}(G, Y_k) \right)^+, \quad (7)$$

and

$$Y_{k+1} = F_k + \frac{\alpha_k - 1}{\alpha_{k+1}} (F_k - F_{k-1}), \quad (8)$$

where $(\cdot)^+$ is the operator which projects any negative entries to zero, $\nabla_F \mathcal{J}(G, \cdot)$ is the gradient of $\mathcal{J}(G, \cdot)$ in Eq. (4), and L is a Lipschitz constant equal to

$$L = \left\| G \cdot G^T \right\|_2 = \|G\|_2^2, \quad (9)$$

where $\|\cdot\|_2$ is the spectral norm. Nesterov gradient descent thus performs a single step of gradient to go to F_k and then slides it in the direction of F_{k-1} to derive Y_{k+1} . Using the Karush-Kuhn-Tucker (KKT) conditions, a stopping criterion—considering both a maximum number Max_{iter} of iterations and a gradient bound—is proposed in [9], thus yielding $F^{t+1} = Y_K$, where Y_K is the last iterate of the above inner iterative gradient descent. In practice, the gradient descent stops if $K = \text{Max}_{\text{iter}}$ or if

$$\left\| \nabla_F^+ \mathcal{J}(G^t, Y_K) \right\|_{\mathcal{F}} \leq \epsilon \cdot \left\| \left[\nabla_G^+ \mathcal{J}(G^1, F^1), \nabla_F^+ \mathcal{J}(G^1, F^1) \right] \right\|_{\mathcal{F}}, \quad (10)$$

where ϵ is a user-defined threshold, $\nabla_F^+ \mathcal{J}(G^t, \cdot)$ is the projected gradient of $\mathcal{J}(G^t, \cdot)$, and $\mathcal{J}(G^1, F^1)$ is the value of the cost function computed at the outer loop initialization. The same strategy is applied to G .

3 Extending NeNMF to missing entries

3.1 Weighted extension of NeNMF

As explained in Section 1, two main strategies exist to process missing data in matrix factorization, i.e., using weights or an EM framework. In the weighted

strategy, a binary matrix W is included in the NMF formalism. The (i, j) -th entry in W is defined as

$$W(i, j) \triangleq \begin{cases} 1 & \text{if } (i, j) \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \quad (11)$$

Combining Eqs. (3) and (11), WNMF for NMC considers the following problem:

$$\min_{\tilde{G}, \tilde{F} \geq 0} \left\| W \circ (X - \tilde{G} \cdot \tilde{F}) \right\|_{\mathcal{F}}^2, \quad (12)$$

which is solved using the same splitting strategy as for NMF, e.g., for F :

$$F = \arg \min_{\tilde{F} \geq 0} \mathcal{J}_W(G, \tilde{F}) = \arg \min_{\tilde{F} \geq 0} \frac{1}{2} \cdot \left\| W \circ (X - G \cdot \tilde{F}) \right\|_{\mathcal{F}}^2. \quad (13)$$

Lemma 1. *The cost function $\mathcal{J}_W(G, \cdot)$ is convex.*

Lemma 2. *In the case of binary weights—and assuming that any row and any column of W contains at least one nonzero element—the gradient of $\mathcal{J}_W(G, \cdot)$ is L -Lipschitz continuous, where L is defined in Eq. (9).*

The proofs are provided in Appendix A.

Lemmas 1 and 2 allow the use of Nesterov iterations, which alternately compute updates (7) and (8), except that the gradient $\nabla_F \mathcal{J}(G, Y_k)$ is here replaced by

$$\nabla_F \mathcal{J}_W(G, Y_k) = G^T \cdot (W^2 \circ (G \cdot Y_k - X)), \quad (14)$$

where $W^2 = W \circ W$ and \circ is the Hadamard product. The stopping criterion for the inner loop is the same as in Section 2 for the standard approach, except that $\mathcal{J}(\cdot, \cdot)$ is here replaced by $\mathcal{J}_W(\cdot, \cdot)$.

It should be noticed that weighted gradient optimization was also investigated in matrix factorization with no sign constraint [28]. In particular, the authors explain that many local minima of $\mathcal{J}_W(G, F)$ exist but can be avoided in practice by gradient descent methods with a well-chosen step size. As a consequence, the major risk of the proposed W-NeNMF might be to converge to a local minimum.

3.2 EM extension of NeNMF

As an alternative to the above weighted framework, the authors in [32] proposed an EM strategy which consists of an unweighted low-rank approximation of a weighted combination of a previously computed approximation. It turns out that for a given NMF iteration t , the EM problem aims to solve

$$\min_{G^{t+1}, F^{t+1} \geq 0} \left\| W \circ M + (\mathbb{1}_{m,n} - W) \circ (G^t \cdot F^t) - G^{t+1} \cdot F^{t+1} \right\|_{\mathcal{F}}^2, \quad (15)$$

where $\mathbb{1}_{m,n}$ is the $m \times n$ matrix of ones. Indeed, the term $(\mathbb{1}_{m,n} - W) \circ (G^t \cdot F^t)$ in Eq. (15) provides the expected missing values of M . The latter are derived

from the complete-data log-likelihood expression, with respect to the unknown data X and given the observed data matrix M and the current estimate of X , i.e., $G^t \cdot F^t$. Denoting $X^t \triangleq W \circ M + (\mathbb{1}_{m,n} - W) \circ (G^t \cdot F^t)$, Eq. (15) reads as a classical NMF problem, i.e.,

$$\min_{G^{t+1}, F^{t+1} \geq 0} \|X^t - G^{t+1} \cdot F^{t+1}\|_{\mathcal{F}}^2, \quad (16)$$

whose solution maximizes the previously computed likelihood [32]. As a consequence, the whole approach consists of:

E-step: Estimate the completed version X of M using

$$X \leftarrow W \circ M + (\mathbb{1}_{m,n} - W) \circ (G \cdot F). \quad (17)$$

M-step: Apply a standard NMF to update G and F using X instead of M .

The resulting EM extension of NeNMF then follows the same structure, except that the M-step runs Nesterov iterations instead of MU [32] or ANLS [16].

It should be noticed that the E-step is a key point in the method. Indeed, if the estimation of X is not accurate enough—which might happen in early NMF iterations or in noisy conditions—this might propagate inaccuracies in the estimation of G and F along iterations. We experimentally investigate such potential issues in Section 4.

Moreover, the authors in [32] state that the EM strategy combined with MU-NMF is less sensitive to initialization than MU-WNMF. However, they notice that MU-WNMF is faster than EM-NMF in their tests, because they run many MUs at each M-step. When applied to Nesterov iterations, such a strategy has to be rethought. As a consequence, we investigated in preliminary tests the number of Nesterov outer loops processed at each M-stage. Actually, this number is linked to the proportion of missing data in M . Indeed, when the proportion of missing data is low, the E-step is accurate enough to run many Nesterov outer loops and provide a good performance. On the contrary, when this proportion is high, we faced that running relatively few outer loops provides a better performance. As a trade-off, we run 2 outer loops per M-step.

4 Performance of the weighted NeNMF extensions

In this section, we investigate the completion performance of the proposed W-NeNMF and EM-W-NeNMF, with respect to the state-of-the-art WNMF [13] and EM-WNMF² [32] techniques³. For that purpose, we generate simulations which each consist of the generation of 500×10 and 10×500 random matrices G

² As explained above, the standard approach was shown to be slow to converge in [32], because many MUs were applied at each M-step. To prevent such an issue, at each M-stage, we here run one NMF update per matrix factor, i.e., one MU.

³ In some preliminary tests, we noticed that the EM-W-ANLS method [16] provided a performance similar to EM-WNMF [32]. We thus do not include it in these tests.

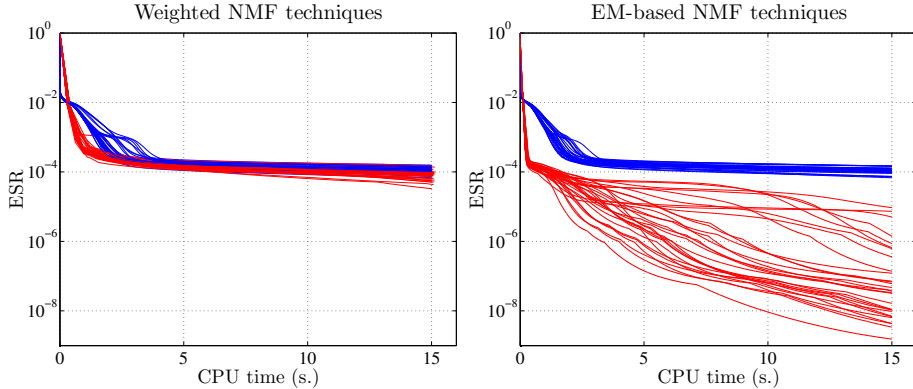


Fig. 1: Evolution of the NMC ESRs vs the CPU time, achieved with MU (in blue), and Nesterov iterations (in red).

and F , respectively. For each pair of matrices, we derive the rank-10 500×500 nonnegative matrix X and we randomly sample a percentage of data through an operator \mathcal{P}_Ω , where this percentage ranges from 10 to 90% (with a step-size of 20%). The number Max_{iter} of passes in the Nesterov inner loops is set to 100. Indeed, such a value was shown to be a good trade-off in preliminary tests where Max_{iter} was set to 1, 10, 100, and 1000. In particular, we noticed a very low performance when $\text{Max}_{\text{iter}} = 1$ —i.e., when the proposed extensions turn out to use a classical gradient descent instead of Nesterov iterations—while the achieved performance was close in the other cases, with a slightly better enhancement with 100 passes per inner loop for each extension. The user-defined threshold ϵ is initialized to 10^{-3} and is divided by 10 each time the stopping criterion (10) is reached within a very low number of inner iterations that we set to $\text{Min}_{\text{iter}} = 5$ in our tests.

All the methods are run using Matlab R2014a on a laptop with an Intel Core i7-4800MQ Quad Core processor, and 32 GB RAM memory. In particular, they are initialized with the same matrices and are stopped after 15 s. In order to compare their performance, we finally compute an Error Signal Ratio (ESR) defined as

$$\text{ESR}(X, G \cdot F) = \frac{\|X - G \cdot F\|_{\mathcal{F}}^2}{\|X\|_{\mathcal{F}}^2}. \quad (18)$$

Figure 1 shows the ESRs—reached by the four tested methods on 30 simulations made with 50% of missing entries in M —with respect to the CPU time expressed in seconds. Interestingly, the MU-based methods and the W-NeNMF method provide the same global ESR after 15 s. However, the W-NeNMF ESRs decrease faster in the early iterations. This shows the acceleration due to the Nesterov iterations. All these methods are outperformed by the EM-W-NeNMF approach whose much lower ESRs are still significantly decreasing. This implies that running the EM-W-NeNMF method more than 15 s should provide a much

better enhancement. On the contrary, the other methods provide slightly decreasing ESRs, which implies that they will need a lot of time to converge to a limit matrix.

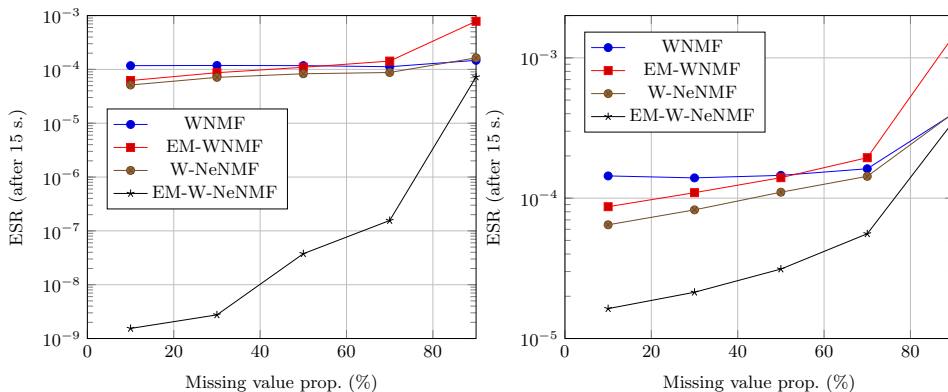


Fig. 2: Median NMC performance reached by the tested NMF methods: (left) noiseless data, (right) noisy data.

The left plot in Fig. 2 shows the median ESRs achieved by the tested methods after 15 s CPU time, with respect to the missing value proportion. Again, we notice that the EM-W-NeNMF outperforms all the other approaches. However, when only 10% of the data in M are available, the achieved performance is almost similar to the one obtained with both the WNMF and W-NeNMF methods. Actually, the weighed extensions of MU-NMF and NeNMF are not very sensitive to the sampling rate, in contrast to the EM methods which look very sensitive to this parameter. Indeed, when a lot of data is missing, the E-step does not yield an accurate estimate of X , thus providing a little enhancement in the M-step, which is consistent with [32]. This results in a slower strategy, which is particularly visible with the EM-WNMF performance and which is partly compensated by the speed of Nesterov gradient for the EM-W-NeNMF method.

The right plot in Fig. 2 shows the achieved ESRs in the same simulations when an additive noise is added to X , with an input SNR around 34 dB. We first notice that the median performance achieved by the WNMF, EM-WNMF, and W-NeNMF methods is almost not affected by the presence of noise. However, the obtained ESRs were already quite high in the noiseless case and thus remain high. Despite the fact that the achieved ESRs are higher in the noisy configuration than in the previous noiseless case, the EM-W-NeNMF still outperforms the other methods. The performance loss is probably due to the fact that the E-step is less efficient, which prevents accurate estimation of G and F in the M-step.

5 Conclusion

In this paper, we proposed to extend a fast NMF method using Nesterov gradient descent to solve large-scale low-rank nonnegative matrix completion. Indeed, contrary to classical gradient methods, Nesterov gradient descent does not need the line search of the optimal step in the update rules, which makes the approach really fast. In order to process missing data, we proposed two extensions using the weighted and the EM strategies, respectively. The latter—which consists of an unweighted optimization of a weighted problem—was found to be better suited than the weighted strategy, in noiseless and noisy NMC simulations, hence showing the relevance of the method.

In future work, we aim to extend such approaches to an informed framework [5, 6, 20]—where some entries of G and/or F are provided—and to apply such extensions to large-scale blind calibration problems [5–7].

Acknowledgments. This work was funded by the "OSCAR" project within the Région Hauts-de-France "Chercheurs Citoyens" Program.

A Proof of Lemmas 1 and 2

Following the structure of the proof in [9], the proof of Lemma 1 is shown by noticing that

$$\mathcal{J}_W(G, \lambda F_1 + (1 - \lambda)F_2) - \lambda \mathcal{J}_W(G, F_1) - (1 - \lambda) \mathcal{J}_W(G, F_2) = \frac{-\lambda(1-\lambda)}{2} \mathcal{J}_W(G, F_1 - F_2) \leq 0. \quad (19)$$

Let us now focus on the proof of Lemma 2. The Lipschitz continuity of $\nabla_F \mathcal{J}_W$ can be shown by extending the proof in [9] to the weighted situation considered in this paper. However, the key point lies in the estimation of the "best" Lipschitz constant. Indeed, let us first recall that Q is a Lipschitz constant of $\nabla_F \mathcal{J}_W(G, \cdot)$ if, for any matrices F_1 and F_2 ,

$$\|\nabla_F \mathcal{J}_W(G, F_1) - \nabla_F \mathcal{J}_W(G, F_2)\|_{\mathcal{F}} \leq Q \cdot \|F_1 - F_2\|_{\mathcal{F}}. \quad (20)$$

From Eq. (7), it is obvious that the larger is the majoring constant Q , the smaller is the associated gradient step size and thus the convergence speed.

Considering the gradient expression (14), we derive

$$\|\nabla_F \mathcal{J}_W(G, F_1) - \nabla_F \mathcal{J}_W(G, F_2)\|_{\mathcal{F}} = \|G^T \cdot (W^2 \circ (G \cdot (F_1 - F_2)))\|_{\mathcal{F}}. \quad (21)$$

The singular value decomposition of G yields

$$\|\nabla_F \mathcal{J}_W(G, F_1) - \nabla_F \mathcal{J}_W(G, F_2)\|_{\mathcal{F}} \leq \|G\|_2 \|W^2 \circ (G \cdot (F_1 - F_2))\|_{\mathcal{F}}. \quad (22)$$

At this stage, we need to put W and G out of the norm to obtain a Lipschitz constant. This can be done by assuming that any column and any row of M

contains at least one element—i.e., any column and row of W contains at least one 1—and thus noticing that

$$\|W^2 \circ (G \cdot (F_1 - F_2))\|_{\mathcal{F}} \leq \|G \cdot (F_1 - F_2)\|_{\mathcal{F}}, \quad (23)$$

which provides, using another singular value decomposition of G :

$$\|\nabla_F \mathcal{J}_W(G, F_1) - \nabla_F \mathcal{J}_W(G, F_2)\|_{\mathcal{F}} \leq \|G\|_2^2 \|F_1 - F_2\|_{\mathcal{F}}, \quad (24)$$

i.e., the Lipschitz constant for W-NeNMF is L .

References

1. Candès, E.J., Plan, Y.: Matrix completion with noise. *Proc. of the IEEE* 98(6), 925–936 (June 2010)
2. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9(6), 717–772 (2009)
3. Choo, J., Lee, C., Reddy, C.K., Park, H.: Weakly supervised nonnegative matrix factorization for user-driven clustering. *Data Mining and Knowledge Discovery* 29(6), 1598–1621 (Nov 2015)
4. Chreiky, R., Delmaire, G., Puigt, M., Roussel, G., Courcot, D., Abche, A.: Split gradient method for informed non-negative matrix factorization. In: Vincent, E., Yeredor, A., Koldovský, Z., Tichavský, P. (eds.) *Proc. of LVA/ICA*. pp. 376–383 (2015)
5. Dorffer, C., Puigt, M., Delmaire, G., Roussel, G.: Blind calibration of mobile sensors using informed nonnegative matrix factorization. In: Vincent, E., Yeredor, A., Koldovský, Z., Tichavský, P. (eds.) *Proc. of LVA/ICA*. pp. 497–505 (2015)
6. Dorffer, C., Puigt, M., Delmaire, G., Roussel, G.: Blind mobile sensor calibration using an informed nonnegative matrix factorization with a relaxed rendezvous model. In: *Proc. of ICASSP*. pp. 2941–2945 (2016)
7. Dorffer, C., Puigt, M., Delmaire, G., Roussel, G.: Nonlinear mobile sensor calibration using informed semi-nonnegative matrix factorization with a Vandermonde factor. In: *Proc. of SAM* (2016)
8. Fazel, M.: Matrix rank minimization with applications. Ph.D. thesis, Stanford University (2002)
9. Guan, N., Tao, D., Luo, Z., Yuan, B.: NeNMF: An optimal gradient method for nonnegative matrix factorization. *IEEE Trans. Sig. Proc.* 60(6), 2882–2898 (June 2012)
10. Guillaumet, D., Vitrià, J., Schiele, B.: Introducing a weighted non-negative matrix factorization for image classification. *Pattern Recognition Letters* 24(14), 2447–2454 (2003)
11. Haldar, J.P., Hernando, D.: Rank-constrained solutions to linear matrix equations using powerfactorization. *IEEE Sig. Proc. Letters* 16(7), 584–587 (July 2009)
12. Hamon, R., Emiya, V., Févotte, C.: Convex nonnegative matrix factorization with missing data. In: *Proc. of MLSP* (2016)
13. Ho, N.D.: Nonnegative matrix factorizations algorithms and applications. Ph.D. thesis, Université Catholique de Louvain (2008)
14. Hoyer, P.O.: Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.* 5, 1457–1469 (Dec 2004)

15. Kim, H., Park, H.: Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM Journal on Matrix Analysis and Applications* 30(2), 713–730 (2008)
16. Kim, Y.D., Choi, S.: Weighted nonnegative matrix factorization. In: Proc. of ICASSP. pp. 1541–1544 (April 2009)
17. Kumar, R., Da Silva, C., Akalin, O., Aravkin, A.Y., Mansour, H., Recht, B., Herrmann, F.J.: Efficient matrix completion for seismic data reconstruction. *Geophysics* 80(5), V97–V114 (2015)
18. Lantéri, H., Theys, C., Richard, C., Févotte, C.: Split gradient method for nonnegative matrix factorization. In: Proc. of EUSIPCO (2010)
19. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS, pp. 556–562 (2001)
20. Limem, A., Delmaire, G., Puigt, M., Roussel, G., Courcot, D.: Non-negative matrix factorization under equality constraints—a study of industrial source identification. *Applied Numerical Mathematics* 85, 1–15 (2014)
21. Limem, A., Puigt, M., Delmaire, G., Roussel, G., Courcot, D.: Bound constrained weighted NMF for industrial source apportionment. In: Proc. of MLSP (2014)
22. Lin, C.: Projected gradient methods for nonnegative matrix factorization. *Neural Computation* 19(10), 2756–2779 (Oct 2007)
23. Liu, C., Yang, H.C., Fan, J., He, L.W., Wang, Y.M.: Distributed nonnegative matrix factorization for web-scale dyadic data analysis on MapReduce. In: Proc. of WWW Conf. (April 2010)
24. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research* 11(January), 19–60 (2010)
25. Nesterov, Y.: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In: *Soviet Mathematics Doklady*. vol. 27, pp. 372–376 (1983)
26. Savvaki, S., Tsagkatakis, G., Panousopoulou, A., Tsakalides, P.: Application of matrix completion on water treatment data. In: Proc. of CySWater. pp. 3:1–3:6 (2015)
27. Sobral, A., Bouwmans, T., Zahzah, E.: LRSLibrary: Low-rank and sparse tools for background modeling and subtraction in videos. In: *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*. CRC Press, Taylor and Francis Group.
28. Srebro, N., Jaakkola, T.: Weighted low-rank approximations. In: Proc. of ICML (2003)
29. Tepper, M., Sapiro, G.: Compressed nonnegative matrix factorization is fast and accurate. *IEEE Trans. Sig. Proc.* 64(9), 2269–2283 (May 2016)
30. Wang, Y.X., Zhang, Y.J.: Nonnegative matrix factorization: A comprehensive review. *IEEE Trans. on Knowledge and Data Eng.* 25(6), 1336–1353 (June 2013)
31. Wen, Z., Yin, W., Zhang, Y.: Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm. *Mathematical Programming Computation* 4(4), 333–361 (2012)
32. Zhang, S., Wang, W., Ford, J., Makedon, F.: *Learning from Incomplete Ratings Using Non-negative Matrix Factorization*, chap. 58, pp. 549–553 (2006)
33. Zhou, G., Cichocki, A., Xie, S.: Fast nonnegative matrix/tensor factorization based on low-rank approximation. *IEEE Trans. Sig. Proc.* 60(6), 2928–2940 (June 2012)

Bibliographie

- [1] A. ADLER, V. EMIYA, M. G. JAFARI, M. ELAD, R. GRIBONVAL et M. D. PLUMBLEY. « Audio inpainting ». In : *IEEE Transactions on Audio, Speech, and Language Processing* 20.3 (2012), p. 922-932.
- [2] S.-I. AMARI. « alpha-Divergence Is Unique, Belonging to Both f-Divergence and Bregman Divergence Classes ». In : *IEEE Transactions on Information Theory* 55.11 (2009), p. 4925-4931.
- [3] P. ANTILA, P. PAATERO, U. TAPPER et O. JÄRVINEN. « Source identification of bulk wet deposition in Finland by positive matrix factorization ». In : *Atmospheric Environment* 29.14 (1995), p. 1705-1718.
- [4] S. ARORA, R. GE, R. KANNAN et A. MOITRA. « Computing a nonnegative matrix factorization – provably ». In : *Proceedings of the 44th symposium on Theory of Computing*. 2012, p. 145-162.
- [5] L. BALZANO et R. NOWAK. « Blind Calibration of Sensor Networks ». In : *Proc. of IPSN*. 2007, p. 79-88.
- [6] A. BASU, I. R. HARRIS, N. L. HJORT et M. JONES. « Robust and efficient estimation by minimising a density power divergence ». In : *Biometrika* 85.3 (1998), p. 549-559.
- [7] S. R. BECKER, E. J. CANDÈS et M. C. GRANT. « Templates for convex cone problems with applications to sparse signal recovery ». In : *Mathematical programming computation* 3.3 (2011), p. 165-218.
- [8] D. BENACHIR, Y. DEVILLE, S. HOSSEINI, M. S. KAROUI et A. HAMEURLAIN. « Hyperspectral image unmixing by non-negative matrix factorization initialized with modified independent component analysis ». In : *Proc. of WHISPERS*. 2013.
- [9] Y. K. BENKOUIDER, M. S. KAROUI, Y. D. VILLE et S. HOSSEINI. « A new multiplicative nonnegative matrix factorization method for unmixing hyperspectral images combined with multispectral data ». In : *Proceedings of EUSIPCO*. Août 2017, p. 483-487.
- [10] M. BERTALMIO, G. SAPIRO, V. CASELLES et C. BALLESTER. « Image inpainting ». In : *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co. 2000, p. 417-424.
- [11] C. BILEN, G. PUY, R. GRIBONVAL et L. DAUDET. « Convex Optimization Approaches for Blind Sensor Calibration Using Sparsity ». In : *IEEE Trans. on Signal Processing* 62.18 (sept. 2014), p. 4847-4856.
- [12] J. M. BIOUSCAS-DIAS, A. PLAZA, N. DOBIGEON, M. PARENTE, Q. DU, P. GADER et J. CHANUSSOT. « Hyperspectral unmixing overview : Geometrical, statistical, and sparse regression-based approaches ». In : *IEEE journal of selected topics in applied earth observations and remote sensing* 5.2 (2012), p. 354-379.

- [13] V. BITTORF, B. RECHT, C. RÉ et J. TROPP. « Factoring nonnegative matrices with linear programs ». In : *Advances in Neural Information Processing Systems*. 2012, p. 1214-1222.
- [14] R. BOBBINK, M. HORNUNG et J. G. M. ROELOFS. « The effects of air-borne nitrogen pollutants on species diversity in natural and semi-natural European vegetation ». In : *Journal of Ecology* 86.5 (1998), p. 717-738.
- [15] A. BOUBRIMA, W. BECHKIT et H. RIVANO. « Optimal WSN Deployment Models for Air Pollution Monitoring ». In : *IEEE Transactions on Wireless Communications* 16.5 (2017), p. 2723-2735.
- [16] A. BOUBRIMA, F. MATIGOT, W. BECHKIT, H. RIVANO et A. RUAS. « Optimal deployment of wireless sensor networks for air pollution monitoring ». In : *Proceedings of ICCCN*. IEEE. 2015, p. 1-7.
- [17] S. BOURGUIGNON, H. CARFANTAN et J. IDIER. « A sparsity-based method for the estimation of spectral lines from irregularly sampled data ». In : *IEEE Journal of Selected Topics in Signal Processing* 1.4 (déc. 2007), p. 575-585.
- [18] S. BOURGUIGNON, J. NININ, H. CARFANTAN et M. MONGEAU. « Exact Sparse Approximation Problems via Mixed-Integer Programming : Formulations and Computational Performance ». In : *IEEE Transactions on Signal Processing* 64.6 (mar. 2016), p. 1405-1419.
- [19] S. BOYD, N. PARIKH, E. CHU, B. PELEATO et J. ECKSTEIN. « Distributed optimization and statistical learning via the alternating direction method of multipliers ». In : *Foundations and Trends® in Machine Learning* 3.1 (2011), p. 1-122.
- [20] R. CABRAL, F. DE LA TORRE, J. P. COSTEIRA et A. BERNARDINO. « Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition ». In : *Proceedings of the IEEE International Conference on Computer Vision*. 2013, p. 2488-2495.
- [21] V. CAMBARERI et L. JACQUES. « A non-convex blind calibration method for randomised sensing strategies ». In : *Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), 2016 4th International Workshop on*. IEEE. 2016, p. 16-20.
- [22] E. J. CANDÈS et B. RECHT. « Exact matrix completion via convex optimization ». In : *Foundations of Computational mathematics* 9.6 (2009), p. 717.
- [23] E. J. CANDÈS et M. B. WAKIN. « An introduction to compressive sampling ». In : *IEEE signal processing magazine* 25.2 (2008), p. 21-30.
- [24] J.-F. CARDOSO. « Analyse en composantes indépendantes ». In : *Proc. of XXXIV Journées de Statistique, Bruxelles* (2002).
- [25] H. CARFANTAN et J. IDIER. « Statistical linear destriping of satellite-based pushbroom-type images ». In : *IEEE transactions on geoscience and remote sensing* 48.4 (2010), p. 1860-1871.

- [26] S. CHANG et K. JEONG. « A Mobile Application for Fine Dust Monitoring System ». In : *Mobile Data Management (MDM), 2017 18th IEEE International Conference on*. IEEE. 2017, p. 336-339.
- [27] C. CHAUDET, E. FLEURY, I. G. LASSOUS, H. RIVANO et M.-E. VOGÉ. « Optimal positioning of active and passive monitoring devices ». In : *Proceedings of the 2005 ACM conference on Emerging network experiment and technology*. ACM. 2005, p. 71-82.
- [28] P. CHAUVET. « Aide-mémoire de géostatistique linéaire ». In : *Ecole des Mines de Paris, Cahiers de Géostatistique 2* (1994).
- [29] Y. CHEN, N. M. NASRABADI et T. D. TRAN. « Hyperspectral image classification using dictionary-based sparse representation ». In : *IEEE Transactions on Geoscience and Remote Sensing* 49.10 (2011), p. 3973-3985.
- [30] H. CHERNOFF. « A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations ». In : *The Annals of Mathematical Statistics* (1952), p. 493-507.
- [31] R. CHREIKY, G. DELMAIRE, C. DORFFER, M. PUIGT, G. ROUSSEL et A. ABCHE. « Robust informed split gradient NMF using $\alpha\beta$ -divergence for source apportionment ». In : *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE. 2016, p. 1-6.
- [32] R. CHREIKY, G. DELMAIRE, M. PUIGT, G. ROUSSEL, D. COURCOT et A. ABCHE. « Split gradient method for informed non-negative matrix factorization ». In : *International Conference on Latent Variable Analysis and Signal Separation*. Springer. 2015, p. 376-383.
- [33] D. CHRISTIN, A. REINHARDT, S. KANHERE et M. HOLLICK. « A survey on privacy in mobile participatory sensing applications ». In : *Journal of systems and software* 84.11 (2011), p. 1928-1946.
- [34] A. CICHOCKI et S.-i. AMARI. « Families of alpha-beta-and gamma-divergences : Flexible and robust measures of similarities ». In : *Entropy* 12.6 (2010), p. 1532-1568.
- [35] A. CICHOCKI, S. CRUCES et S.-i. AMARI. « Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization ». In : *Entropy* 13.1 (2011), p. 134-170.
- [36] A. CICHOCKI, H. LEE, Y.-D. KIM et S. CHOI. « Non-negative matrix factorization with α -divergence ». In : *Pattern Recognition Letters* 29.9 (2008), p. 1433-1440.
- [37] A. CICHOCKI et A.-H. PHAN. « Fast local algorithms for large scale nonnegative matrix and tensor factorizations ». In : *IEICE transactions on fundamentals of electronics, communications and computer sciences* 92.3 (2009), p. 708-721.
- [38] A. CICHOCKI, R. ZDUNEK et S.-i. AMARI. « Csiszar's divergences for non-negative matrix factorization : Family of new algorithms ». In : *International Conference on Independent Component Analysis and Signal Separation*. Springer. 2006, p. 32-39.

- [39] A. CICHOCKI, R. ZDUNEK et S.-i. AMARI. « Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization ». In : *International Conference on Independent Component Analysis and Signal Separation*. Springer. 2007, p. 169-176.
- [40] A. CICHOCKI, R. ZDUNEK, A. H. PHAN et S.-i. AMARI. *Nonnegative matrix and tensor factorizations : applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [41] E. S. COCHRAN, J. F. LAWRENCE, A. KAISER, B. FRY, A. CHUNG et C. CHRISTENSEN. « Comparison between low-cost and traditional MEMS accelerometers : a case study from the M7. 1 Darfield, New Zealand, aftershock deployment ». In : *Annals of Geophysics* 54.6 (2012).
- [42] A. COMAN. « Modélisation spatio-temporelle de la pollution atmosphérique urbaine à partir des mesures d'un réseau de surveillance de la qualité de l'air ». Thèse de doct. Université Paris Est, sept. 2008.
- [43] G. P. CRESSMAN. « An operational objective analysis system ». In : *Mon. Wea. Rev* 87.10 (1959), p. 367-374.
- [44] M. A. DAVENPORT et J. ROMBERG. « An Overview of Low-Rank Matrix Recovery From Incomplete Observations ». In : *IEEE Journal of Selected Topics in Signal Processing* 10.4 (juin 2016), p. 608-622.
- [45] G. DAVIS, S. MALLAT et M. AVELLANEDA. « Adaptive greedy approximations ». In : *Constructive approximation* 13.1 (1997), p. 57-98.
- [46] G. DELMAIRE et G. ROUSSEL. « Joint estimation decision methods for source localization and restoration in parametric convolution processes. Application to accidental pollutant release ». In : *Digital Signal Processing* 22.1 (2012), p. 34-46.
- [47] Y. DEVILLE. « Matrix factorization for bilinear blind source separation : Methods, separability and conditioning ». In : *Signal Processing Conference (EUSIPCO), 2015 23rd European*. IEEE. 2015, p. 1900-1904.
- [48] C. DING, X. HE et H. D. SIMON. « On the equivalence of nonnegative matrix factorization and spectral clustering ». In : *Proceedings of the 2005 SIAM International Conference on Data Mining*. SIAM. 2005, p. 606-610.
- [49] D. DONOHO et V. STODDEN. « When does non-negative matrix factorization give a correct decomposition into parts? » In : *Advances in neural information processing systems*. 2004, p. 1141-1148.
- [50] C. DORFFER, M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Blind calibration of mobile sensors using informed nonnegative matrix factorization ». In : *12th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2015)*. Août 2015, p. 497-505.
- [51] C. DORFFER, M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Blind mobile sensor calibration using an informed nonnegative matrix factorization with a relaxed rendezvous model ». In : *Proceedings of the 41st IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016)*. Mar. 2016, p. 2941-2945.

- [52] C. DORFFER, M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Fast nonnegative matrix factorization and completion using nesterov iterations ». In : *International Conference on Latent Variable Analysis and Signal Separation*. Springer. 2017, p. 26-35.
- [53] C. DORFFER, M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration ». working paper or preprint. Juil. 2017.
- [54] C. DORFFER, M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Nonlinear mobile sensor calibration using informed semi-nonnegative matrix factorization with a Vandermonde factor ». In : *9th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM 2016)*. Juil. 2016.
- [55] C. DORFFER, M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Outlier-Robust Calibration Method for Sensor Networks ». In : *2017 IEEE International Workshop of Electronics, Control, Measurement, Signals and their Application to Mechatronics (ECMSM)*. Mai 2017.
- [56] R. DUBROCA, C. JUNOR et A. SOULOUMIAC. « Weighted NMF for high-resolution mass spectrometry analysis ». In : *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. IEEE. 2012, p. 1806-1810.
- [57] J. DUCHON. « Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces ». In : *ESAIM : Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique* 10.R3 (1976), p. 5-12.
- [58] M. DUNBABIN et L. MARQUES. « Robots for environmental monitoring : Significant advancements and applications ». In : *IEEE Robotics & Automation Magazine* 19.1 (2012), p. 24-39.
- [59] M. ELAD et M. AHARON. « Image denoising via sparse and redundant representations over learned dictionaries ». In : *IEEE Transactions on Image processing* 15.12 (2006), p. 3736-3745.
- [60] M. ELAD, J.-L. STARCK, P. QUERRE et D. L. DONOHO. « Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA) ». In : *Applied and Computational Harmonic Analysis* 19.3 (2005), p. 340-358.
- [61] Y. ELДАР et G. KUTYNIOK. *Compressed Sensing : Theory and Applications*. Cambridge University Press, 2012.
- [62] S. ESSID et C. FÉVOTTE. « Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring ». In : *IEEE Transactions on Multimedia* 15.2 (2013), p. 415-425.
- [63] X. FANG et I. BATE. « Using Multi-parameters for Calibration of Low-cost Sensors in Urban Environment ». In : *Proceedings of the 2017 International Conference on Embedded Wireless Systems and Networks*. EWSN ’17. Uppsala, Sweden : Junction Publishing, 2017, p. 1-11.
- [64] M. FAZEL. « Matrix Rank Minimization with Applications ». Thèse de doct. Stanford University, 2002.

- [65] C. FÉVOTTE. « Majorization-minimization algorithm for smooth Itakura-Saito nonnegative matrix factorization ». In : *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mai 2011, p. 1980-1983.
- [66] C. FÉVOTTE et N. DOBIGEON. « Nonlinear Hyperspectral Unmixing With Robust Nonnegative Matrix Factorization ». In : *IEEE Transactions on Image Processing* 24.12 (déc. 2015), p. 4810-4819.
- [67] C. FÉVOTTE, N. BERTIN et J.-L. DURRIEU. « Nonnegative matrix factorization with the Itakura-Saito divergence : With application to music analysis ». In : *Neural computation* 21.3 (2009), p. 793-830.
- [68] C. FÉVOTTE et J. IDIER. « Algorithms for nonnegative matrix factorization with the β -divergence ». In : *Neural computation* 23.9 (2011), p. 2421-2456.
- [69] C. D. FRANCIS, C. P. ORTEGA et A. CRUZ. « Noise Pollution Changes Avian Communities and Species Interactions ». In : *Current Biology* 19.16 (2009), p. 1415-1419.
- [70] F. GADALLAH, F. CSILLAG et E. SMITH. « Destriping multisensor imagery with moment matching ». In : *International journal of remote sensing* 21.12 (2000), p. 2505-2511.
- [71] F. GALTON. « Regression towards mediocrity in hereditary stature. » In : *The Journal of the Anthropological Institute of Great Britain and Ireland* 15 (1886), p. 246-263.
- [72] R. K. GANTI, F. YE et H. LEI. « Mobile crowdsensing : current state and future challenges ». In : *IEEE Communications Magazine* 49.11 (nov. 2011), p. 32-39.
- [73] N. GILLIS et F. GLINEUR. « Accelerated multiplicative updates and hierarchical ALS algorithms for nonnegative matrix factorization ». In : *Neural computation* 24.4 (2012), p. 1085-1105.
- [74] *GP2Y1010AU0F compact optical dst sensor*. Sharp Corp. 2006.
- [75] Y. GRATTON. « Le krigeage : la méthode optimale d'interpolation spatiale ». In : *Les articles de l'Institut d'Analyse Géographique* 1 (2002).
- [76] N. GUAN, D. TAO, Z. LUO et B. YUAN. « NeNMF : An optimal gradient method for nonnegative matrix factorization ». In : *IEEE Transactions on Signal Processing* 60.6 (2012), p. 2882-2898.
- [77] N. GUAN, D. TAO, Z. LUO et B. YUAN. « Online nonnegative matrix factorization with robust stochastic approximation ». In : *IEEE Transactions on Neural Networks and Learning Systems* 23.7 (2012), p. 1087-1099.
- [78] D. GUILLAMET, M. BRESSAN et J. VITRIA. « A weighted non-negative matrix factorization for local representations ». In : *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. T. 1. IEEE. 2001, p. I-I.
- [79] D. GUILLAMET, J. VITRIA et B. SCHIELE. « Introducing a weighted non-negative matrix factorization for image classification ». In : *Pattern Recognition Letters* 24.14 (2003), p. 2447-2454.

- [80] N. HADERER, R. ROUVOY, C. RIBEIRO et L. SEINTURIER. « APISENSE : Crowd-Sensing Made Easy ». In : *ERCIM News*. Special theme : Mobile Computing 93 (avr. 2013), p. 28-29.
- [81] J. P. HALDAR et D. HERNANDO. « Rank-constrained solutions to linear matrix equations using powerfactorization ». In : *IEEE Signal Processing Letters* 16.7 (2009), p. 584-587.
- [82] N. HALKO, P.-G. MARTINSSON et J. A. TROPP. « Finding structure with randomness : Probabilistic algorithms for constructing approximate matrix decompositions ». In : *SIAM review* 53.2 (2011), p. 217-288.
- [83] R. HAMON, V. EMIYA et C. FÉVOTTE. « Convex nonnegative matrix factorization with missing data ». In : *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE. 2016, p. 1-6.
- [84] J. K. HART et K. MARTINEZ. « Environmental Sensor Networks : A revolution in the earth system science ? » In : *Earth-Science Reviews* 78.3 (2006), p. 177-191.
- [85] D. HASENFRATZ, O. SAUKH et L. THIELE. « On-the-Fly Calibration of Low-Cost Gas Sensors ». In : *Proc. of EWSN*. 2012.
- [86] D. HASENFRATZ, O. SAUKH, C. WALSER, C. HUEGLIN, M. FIERZ, T. ARN, J. BEUTEL et L. THIELE. « Deriving high-resolution urban air pollution maps using mobile sensor nodes ». In : *Pervasive and Mobile Computing* 16 (2015), p. 268-285.
- [87] J.-E. HAUGEN, O. TOMIC et K. KVAAL. « A calibration method for handling the temporal drift of solid state gas-sensors ». In : *Analytica chimica acta* 407.1 (2000), p. 23-39.
- [88] J. HEIKKILA et O. SILVEN. « A four-step camera calibration procedure with implicit image correction ». In : *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE. 1997, p. 1106-1112.
- [89] R. HENNEQUIN, B. DAVID et R. BADEAU. « Beta-divergence as a subclass of Bregman divergence ». In : *IEEE Signal Processing Letters* 18.2 (2011), p. 83-86.
- [90] J. HENRIQUE BRITO, R. ANGST, K. KOSER et M. POLLEFEYS. « Radial distortion self-calibration ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, p. 1368-1375.
- [91] N.-D. HO. « Nonnegative matrix factorization algorithms and applications ». Thèse de doct. 2008.
- [92] P. O. HOYER. « Non-negative matrix factorization with sparseness constraints ». In : *Journal of machine learning research* 5.Nov (2004), p. 1457-1469.
- [93] P. O. HOYER. « Non-negative sparse coding ». In : *Neural Networks for Signal Processing, 2002. Proceedings of the 2002 12th IEEE Workshop on*. IEEE. 2002, p. 557-565.
- [94] C. HU, P. RAI, C. CHEN, M. HARDING et L. CARIN. « Scalable Bayesian non-negative tensor factorization for massive count data ». In : *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer. 2015, p. 53-70.

- [95] K. HUANG et S. AVIYENTE. « Sparse representation for signal classification ». In : *Advances in neural information processing systems*. 2007, p. 609-616.
- [96] IPCC. *Climate change 2014 : synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds)]. 2014.
- [97] F. ITAKURA. « Analysis synthesis telephony based on the maximum likelihood method ». In : *International Congress on Acoustics, 1968*. 1968, p. 17-20.
- [98] M. JERRETT, A. ARAIN, P. KANAROGLOU, B. BECKERMAN, D. POTOGLOU, T. SAHSUVAROGLU, J. MORRISON et C. GIOVIS. « A review and evaluation of intraurban air pollution exposure models ». In : *Journal of Exposure Science and Environmental Epidemiology* 15.2 (2005), p. 185-204.
- [99] R. KANNAN, G. BALLARD et H. PARK. « A high-performance parallel algorithm for nonnegative matrix factorization ». In : *Proceedings of the 21st ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming*. ACM. 2016, p. 9.
- [100] T. R. KARL, A. ARGUEZ, B. HUANG, J. H. LAWRIMORE, J. R. MCMAHON, M. J. MENNE, T. C. PETERSON, R. S. VOSE et H.-M. ZHANG. « Possible artifacts of data biases in the recent global surface warming hiatus ». In : *Science* (2015).
- [101] S. A. KHAN, T. ISLAM et G. HUSAIN. « Artificial neural network based online sensor calibration and compensation ». In : *International Journal of Computing* 6.3 (2014), p. 74-78.
- [102] H. KIM et H. PARK. « Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method ». In : *SIAM journal on matrix analysis and applications* 30.2 (2008), p. 713-730.
- [103] H. KIM et H. PARK. « Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis ». In : *Bioinformatics* 23.12 (2007), p. 1495-1502.
- [104] T. G. KOLDA et B. W. BADER. « Tensor decompositions and applications ». In : *SIAM review* 51.3 (2009), p. 455-500.
- [105] H. J. KRAMER. *Observation of the Earth and its Environment : Survey of Missions and Sensors*. Springer Science & Business Media, 2002.
- [106] V. KYSENKO, K. RUPP, O. MARCHENKO, S. SELBERHERR et A. ANISIMOV. « GPU-Accelerated non-negative matrix factorization for text mining ». In : *Natural language processing and information systems* (2012), p. 158-163.
- [107] H. LANTÉRI, C. THEYS, C. RICHARD et C. FÉVOTTE. « Split gradient method for non-negative matrix factorization ». In : *Signal Processing Conference, 2010 18th European*. IEEE. 2010, p. 1199-1203.
- [108] J. LE ROUX, F. J. WENINGER et J. R. HERSHEY. « Sparse NMF—half-baked or well done? » In : *Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA, Tech. Rep., no. TR2015-023* (2015).

- [109] L. LE TARO et H. RIVANO. « Collaborative Sensor Calibration : Convergence Study ». In : *CoRes 2017- 2ème Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication*. Quiberon, France, mai 2017, p. 4.
- [110] T. K. LE, N. ONO, T. NOWAKOWSKI, L. DAUDET et J. D. ROSNY. « Experimental validation of TOA-based methods for microphones array positions calibration ». In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Mar. 2016, p. 3216-3220.
- [111] B.-T. LEE, S.-C. SON et K. KANG. « A blind calibration scheme exploiting mutual calibration relationships for a dense mobile sensor network ». In : *IEEE Sensors Journal* 14.5 (2014), p. 1518-1526.
- [112] D. D. LEE et H. S. SEUNG. « Algorithms for non-negative matrix factorization ». In : *Advances in neural information processing systems*. 2001, p. 556-562.
- [113] D. D. LEE et H. S. SEUNG. « Learning the parts of objects by non-negative matrix factorization ». In : *Nature* 401.6755 (1999), p. 788.
- [114] H. LEE et S. CHOI. « Group nonnegative matrix factorization for EEG classification ». In : *Artificial Intelligence and Statistics*. 2009, p. 320-327.
- [115] Y. LEE. « Handwritten digit recognition using k nearest-neighbor, radial-basis function, and backpropagation neural networks ». In : *Neural computation* 3.3 (1991), p. 440-449.
- [116] A. LEFEVRE, F. BACH et C. FÉVOTTE. « Online algorithms for nonnegative matrix factorization with the Itakura-Saito divergence ». In : *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*. IEEE. 2011, p. 313-316.
- [117] J. LELIEVELD, J. EVANS, M. FNAIS, D. GIANNADAKI et A. POZZER. « The contribution of outdoor air pollution sources to premature mortality on a global scale ». In : *Nature* 525.7569 (2015), p. 367-371.
- [118] R. LIAO, Y. ZHANG, J. GUAN et S. ZHOU. « CloudNMF : a MapReduce implementation of nonnegative matrix factorization for large-scale biological datasets ». In : *Genomics, proteomics & bioinformatics* 12.1 (2014), p. 48-51.
- [119] A. LIMEM, G. DELMAIRE, M. PUIGT, G. ROUSSEL et D. COURCOT. « Non-negative Matrix Factorization under equality constraints—a study of industrial source identification ». In : *Applied Numerical Mathematics* 85 (2014), p. 1-15.
- [120] A. LIMEM, G. DELMAIRE, M. PUIGT, G. ROUSSEL et D. COURCOT. « Non-negative matrix factorization using weighted beta divergence and equality constraints for industrial source apportionment ». In : *Proceedings of the 23rd IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2013)*. 2013.
- [121] A. LIMEM, G. DELMAIRE, G. ROUSSEL et D. COURCOT. « Kullback-Leibler NMF Under Linear Equality Constraints. Application to Pollution Source Apportionment ». In : *Proceedings of the International Conference on Information Science Signal Processing and their Applications (ISSPA 2012)*. 2012.

- [122] A. LIMEM, M. PUIGT, G. DELMAIRE, G. ROUSSEL et D. COURCOT. « Bound constrained weighted NMF for industrial source apportionment ». In : *Proceedings of the 24th IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2014)*. 2014.
- [123] A. LIMEM. « Méthodes informées de factorisation matricielle non négative : Application à l'identification de sources de particules industrielles ». Thèse de doct. 2014.
- [124] C.-J. LIN. « On the Convergence of Multiplicative Update Algorithms for Nonnegative Matrix Factorization ». In : *IEEE Transactions on Neural Networks* 18.6 (nov. 2007), p. 1589-1596.
- [125] C.-J. LIN. « Projected gradient methods for nonnegative matrix factorization ». In : *Neural computation* 19.10 (2007), p. 2756-2779.
- [126] Z. LIN, M. CHEN et Y. MA. « The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices ». In : *arXiv preprint arXiv :1009.5055* (2010).
- [127] S. LING et T. STROHMER. « Self-calibration and biconvex compressive sensing ». In : *Inverse Problems* 31.11 (2015), p. 115002.
- [128] J. LIPOR et L. BALZANO. « Robust blind calibration via total least squares ». In : *Proc. of ICASSP*. Mai 2014, p. 4244-4248.
- [129] C. LIU, H.-c. YANG, J. FAN, L.-W. HE et Y.-M. WANG. « Distributed nonnegative matrix factorization for web-scale dyadic data analysis on mapreduce ». In : *Proceedings of the 19th international conference on World wide web*. ACM. 2010, p. 681-690.
- [130] J. LIU, C. WANG, J. GAO et J. HAN. « Multi-view clustering via joint nonnegative matrix factorization ». In : *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM. 2013, p. 252-260.
- [131] W. LIU, N. ZHENG et X. LU. « Non-negative matrix factorization for visual coding ». In : *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. T. 3. IEEE. 2003, p. III-293.
- [132] A. LIUTKUS, U. ŞİMŞEKLI et T. CEMGIL. « Extraction of Temporal Patterns in Multi-rate and Multi-modal Datasets ». In : *International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*. Liberec, Czech Republic, août 2015.
- [133] T. LONGCORE et C. RICH. « Ecological light pollution ». In : *Frontiers in Ecology and the Environment* 2.4 (2004), p. 191-198.
- [134] B. MAAG, Z. ZHOU, O. SAUKH et L. THIELE. « SCAN : Multi-Hop Calibration for Mobile Sensor Arrays ». In : *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1.2 (juin 2017), 19 :1-19 :21.
- [135] J. MAIRAL, F. BACH, J. PONCE et G. SAPIRO. « Online learning for matrix factorization and sparse coding ». In : *Journal of Machine Learning Research* 11.Jan (2010), p. 19-60.
- [136] J. MAIRAL, F. BACH, J. PONCE, G. SAPIRO et A. ZISSERMAN. « Non-local sparse models for image restoration ». In : *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE. 2009, p. 2272-2279.

- [137] J. MAIRAL, M. ELAD et G. SAPIRO. « Sparse representation for color image restoration ». In : *IEEE Transactions on image processing* 17.1 (2008), p. 53-69.
- [138] N. MAQUIGNON. « Vers un modèle multi-phases et multi-composants (MPMC) de type Lattice Boltzmann Method (LBM) pour la simulation dynamique d'un fluide cryogénique dans l'eau. » Thèse de doct. Université du Littoral Côte d'Opale, nov. 2014.
- [139] M. MARTINEZ-CAMARA, I. DOKMANIĆ, J. RANIERI, R. SCHEIBLER, M. VETTERLI et A. STOHL. « The Fukushima inverse problem ». In : *Proceedings of ICASSP*. Mai 2013, p. 4330-4334.
- [140] I. MEGANEM, Y. DEVILLE, S. HOSSEINI, P. DELIOT et X. BRIOTTET. « Linear-quadratic blind source separation using NMF to unmix urban hyperspectral images ». In : *IEEE Transactions on Signal Processing* 62.7 (2014), p. 1822-1833.
- [141] I. MEGANEM, Y. DEVILLE, S. HOSSEINI, P. DELIOT, X. BRIOTTET et L. T. DUARTE. « Linear-quadratic and polynomial non-negative matrix factorization; application to spectral unmixing ». In : *Signal Processing Conference, 2011 19th European*. IEEE. 2011, p. 1859-1863.
- [142] L. MEIER, S. VAN DE GEER et P. BÜHLMANN. « The group lasso for logistic regression ». In : *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* 70.1 (2008), p. 53-71.
- [143] E. MILUZZO, N. D. LANE, A. T. CAMPBELL et R. OLFIATI-SABER. « CaliBree : A self-calibration system for mobile sensor networks ». In : *International Conference on Distributed Computing in Sensor Systems*. Springer. 2008, p. 314-331.
- [144] M. N. H. MOLLAH, M. MINAMI et S. EGUCHI. « Exploring latent structure of mixture ICA models by the minimum β -divergence method ». In : *Neural Computation* 18.1 (2006), p. 166-190.
- [145] S. MOUSSAOUI. « Séparation de sources non-négatives. Application au traitement des signaux de spectroscopie ». Thèse de doct. Université Henri Poincaré-Nancy I, 2005.
- [146] S. MOUSSAOUI, D. BRIE, A. MOHAMMAD-DJAFARI et C. CARTER. « Separation of non-negative mixture of non-negative sources using a Bayesian approach and mcmc sampling ». In : *IEEE Transactions on Signal Processing* 54.11 (2006), p. 4133-4145.
- [147] J. MURRAY-BRUCE et P. L. DRAGOTTI. « A Sampling Framework for Solving Physics-driven Inverse Source Problems ». In : *IEEE Transactions on Signal Processing* (2017).
- [148] J. MURRAY-BRUCE et P. L. DRAGOTTI. « Physics-driven quantized consensus for distributed diffusion source estimation using sensor networks ». In : *EURASIP Journal on Advances in Signal Processing* (2016).
- [149] Y. NESTEROV. « A method of solving a convex programming problem with convergence rate $O(1/k^2)$ ». In : *Soviet Mathematics Doklady*. T. 27. 2. 1983, p. 372-376.
- [150] B. OCHANDO. « Méthode d'inventaire et de dénombrement d'oiseaux en milieu forestier. Application à l'Algérie. » In : *Annales de l'Institut national agronomique-El Harrach (Alger)* 12.3 (1988), p. 47-59.

- [151] P. PAATERO. « Least squares formulation of robust non-negative factor analysis ». In : *Chemometrics and intelligent laboratory systems* 37.1 (1997), p. 23-35.
- [152] P. PAATERO et U. TAPPER. « Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values ». In : *Environmetrics* 5.2 (1994), p. 111-126.
- [153] S. PATEL, J. LI, A. PANDEY, S. PERVEZ, R. K. CHAKRABARTY et P. BISWAS. « Spatio-temporal measurement of indoor particulate matter concentrations using a wireless network of low-cost sensors in households using solid fuels ». In : *Environmental Research* 152 (2017), p. 59-65.
- [154] Y. C. PATI, R. REZAIIFAR et P. S. KRISHNAPRASAD. « Orthogonal matching pursuit : Recursive function approximation with applications to wavelet decomposition ». In : *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*. IEEE. 1993, p. 40-44.
- [155] V. P. PAUCA, F. SHAHNAZ, M. W. BERRY et R. J. PLEMMONS. « Text mining using non-negative matrix factorizations ». In : *Proceedings of the 2004 SIAM International Conference on Data Mining*. SIAM. 2004, p. 452-456.
- [156] D. PAVLIDI, A. GRIFFIN, M. PUIGT et A. MOUCHTARIS. « Real-Time Multiple Sound Source Localization and Counting Using a Circular Microphone Array ». In : *IEEE Transactions on Audio, Speech, and Language Processing* 21.10 (oct. 2013), p. 2193-2206.
- [157] D. PAVLIDI, A. GRIFFIN, M. PUIGT et A. MOUCHTARIS. « Source counting in real-time sound source localization using a circular microphone array ». In : *Proceedings of SAM*. Juin 2012, p. 521-524.
- [158] K. PEARSON. « LIII. On lines and planes of closest fit to systems of points in space ». In : *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2.11 (1901), p. 559-572.
- [159] T.-T.-H. PHAN, É. P. CAILLAULT, A. LEFEBVRE et A. BIGAND. « Dynamic time warping-based imputation for univariate time series data ». In : *Pattern Recognition Letters* (2017).
- [160] R. PIEDRAHITA, Y. XIANG, N. MASSON, J. ORTEGA, A. COLLIER, Y. JIANG, K. LI, R. DICK, Q. LV, M. HANNIGAN et L. SHANG. « The next generation of low-cost personal air quality sensors for quantitative exposure monitoring ». In : *Atmospheric Measurement Techniques* 7.10 (2014), p. 3325.
- [161] A. PLINGE et G. A. FINK. « Geometry calibration of distributed microphone arrays exploiting audio-visual correspondences ». In : *Signal Processing Conference (EU-SIPCO), 2014 Proceedings of the 22nd European*. IEEE. 2014, p. 116-120.
- [162] M. PLOUVIN, A. LIMEM, M. PUIGT, G. DELMAIRE, G. ROUSSEL et D. COURCOT. « Enhanced NMF initialization using a physical model for pollution source apportionment ». In : *Proc. of ESANN*. 2014, p. 261-266.
- [163] *PM2.5 laser dust sensor SKU : SEN017*. DFRobot Corp. 2008.

- [164] G. PUIGT. « CFD dedicated techniques applied to industrial problems ». Habilitation à diriger des recherches. Université de Toulouse, déc. 2014.
- [165] M. PUIGT, G. DELMAIRE et G. ROUSSEL. « Environmental signal processing : new trends and applications ». In : *Proceedings of ESANN*. 2017, p. 205-214.
- [166] M. PUIGT, A. GRIFFIN et A. MOUCHTARIS. « Nonlinear blind mixture identification using local source sparsity and functional data clustering ». In : *Proc. of SAM*. 2012, p. 489-492.
- [167] M. PUIGT, A. GRIFFIN et A. MOUCHTARIS. « Post-nonlinear speech mixture identification using single-source temporal zones & curve clustering ». In : *Proc. of EUSIPCO*. 2011, p. 1844-1848.
- [168] N. RAMANATHAN, L. BALZANO, M. BURT, D. ESTRIN, T. HARMON, C. HARVEY, J. JAY, E. KOHLER, S. ROTHENBERG et M. SRIVASTAVA. « Rapid deployment with confidence : Calibration and fault detection in environmental sensor networks ». In : *Center for Embedded Network Sensing (2006)*.
- [169] J. O. RAMSAY et B. W. SILVERMAN. *Functional data analysis*. Springer-Verlag, 2005.
- [170] S. RANGAN. « Generalized approximate message passing for estimation with random linear mixing ». In : *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE. 2011, p. 2168-2172.
- [171] J. RANIERI, A. CHEBIRA et M. VETTERLI. « Near-optimal sensor placement for linear inverse problems ». In : *IEEE Trans. on Signal Processing* 62.5 (2014), p. 1135-1146.
- [172] J. RANIERI, I. DOKMANIĆ, A. CHEBIRA et M. VETTERLI. « Sampling and reconstruction of time-varying atmospheric emissions ». In : *Proceedings of ICASSP*. 2012, p. 3673-3676.
- [173] S. RICKARD et A. CICHOCKI. « When is non-negative matrix decomposition unique ? » In : *Information Sciences and Systems, 2008. CISS 2008. 42nd Annual Conference on*. IEEE. 2008, p. 1091-1092.
- [174] G. ROUSSEL, L. BOURGOIS, M. BENJELLOUN et G. DELMAIRE. « Estimation of a semi-physical GLBE model using dual EnKF learning algorithm coupled with a sensor network design strategy : Application to air field monitoring ». In : *Information Fusion* 14.4 (2013), p. 335-348.
- [175] H. ROY, M. POCOCK, C. PRESTON, D. ROY, J. SAVAGE, J. TWEDDLE et L. ROBINSON. « Understanding citizen science & environmental monitoring ». In : *NERC-NHM (2012)*. Final report on behalf of UK-EOF.
- [176] V. ROY, A. SIMONETTO et G. LEUS. « Spatio-temporal sensor management for environmental field estimation ». In : *Signal Processing* 128 (2016), p. 369-381.
- [177] J. P. ROYER, N. THIRION-MOREAU et P. COMON. « Nonnegative 3-way tensor factorization taking in to account possible missing data ». In : *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. Août 2012, p. 71-75.
- [178] P. W. RUNDEL, E. A. GRAHAM, M. F. ALLEN, J. C. FISHER et T. C. HARMON. « Environmental sensor networks in ecological research ». In : *New Phytologist* 182.3 (mai 2009), p. 589-607.

- [179] K. S. « Information Theory and Statistics ». In : *John Wiley & Sons*. 1959.
- [180] S. D. SABATINO, R. BUCCOLIERIA, B. PULVIRENTI et R. BRITTER. « Simulations of pollutant dispersion within idealised urban-type geometries with CFD and integral models ». In : *Atmospheric Environment* 41.5 (2007), p. 8316-8329.
- [181] Y. E. SALEHANI et S. GAZOR. « Sparse data reconstruction via adaptive l_p -norm and multilayer NMF ». In : *2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. Oct. 2016, p. 1-6.
- [182] O. SAUKH, D. HASENFRATZ et L. THIELE. « Reducing Multi-Hop Calibration Errors in Large-Scale Mobile Sensor Networks ». In : *Proc. of IPSN*. 2015.
- [183] O. SAUKH, D. HASENFRATZ, C. WALSER et L. THIELE. « On rendezvous in mobile sensing networks ». In : *Real-World Wireless Sensor Networks*. Springer, Cham, 2014, p. 29-42.
- [184] M. N. SCHMIDT et H. LAURBERG. « Non-negative matrix factorization with Gaussian process priors ». In : *Computational Intelligence and Neuroscience* (2008).
- [185] M. N. SCHMIDT, O. WINTHER et L. K. HANSEN. « Bayesian Non-negative Matrix Factorization ». In : *Independent Component Analysis and Signal Separation : 8th International Conference, ICA 2009, Paraty, Brazil, March 15-18, 2009. Proceedings*. Sous la dir. de T. ADALI, C. JUTTEN, J. M. T. ROMANO et A. K. BARROS. Berlin, Heidelberg : Springer Berlin Heidelberg, 2009, p. 540-547.
- [186] C. SCHÜLKE, F. CALTAGIRONE, F. KRZAKALA et L. ZDEBOROVÁ. « Blind Calibration in Compressed Sensing using Message Passing Algorithms ». In : *CoRR* abs/1306.4355 (2013).
- [187] N. SEICHEPINE, S. ESSID, C. FÉVOTTE et O. CAPPÉ. « Soft nonnegative matrix co-factorization ». In : *IEEE Transactions on Signal Processing* 62.22 (2014), p. 5940-5949.
- [188] S. SHAH et J. AGGARWAL. « Intrinsic parameter calibration procedure for a (high-distortion) fish-eye lens camera with distortion model and accuracy estimation ». In : *Pattern Recognition* 29.11 (1996), p. 1775-1788.
- [189] A. SHASHUA et T. HAZAN. « Non-negative tensor factorization with applications to statistics and computer vision ». In : *Proceedings of the 22nd international conference on Machine learning*. ACM. 2005, p. 792-799.
- [190] *SHT1x humidity and temperature sensor*. Sensirion Corp. 2008.
- [191] N. D. SIDIROPOULOS, L. D. LATHAUWER, X. FU, K. HUANG, E. E. PAPALEXAKIS et C. FALOUTSOS. « Tensor Decomposition for Signal Processing and Machine Learning ». In : *IEEE Transactions on Signal Processing* 65.13 (juil. 2017), p. 3551-3582.
- [192] D. SMITH, J. LUKASIAK et I. S. BURNETT. « A two channel, block-adaptive audio separation technique based upon time-frequency information ». In : *2004 12th European Signal Processing Conference*. Sept. 2004, p. 393-396.

- [193] A. SOBRAL, T. BOUWMANS et E.-h. ZAHZAH. « LRSLibrary : Low-rank and sparse tools for background modeling and subtraction in videos ». In : *Robust Low-Rank and Sparse Matrix Decomposition : Applications in Image and Video Processing*. CRC Press, Taylor and Francis Group (2016).
- [194] S.-C. SON, B.-T. LEE, S. K. KO et K. KANG. « Hybrid Sensor Calibration Scheme for Mobile Crowdsensing-Based City-Scale Environmental Measurements ». In : *ETRI Journal* 38.3 (2016), p. 551-559.
- [195] M. SØRENSEN et L. DE LATHAUWER. « Blind signal separation via tensor decomposition with Vandermonde factor : Canonical polyadic decomposition ». In : *IEEE Transactions on Signal Processing* 61.22 (2013), p. 5507-5519.
- [196] N. SREBRO, J. RENNIE et T. S. JAAKKOLA. « Maximum-margin matrix factorization ». In : *Advances in neural information processing systems*. 2005, p. 1329-1336.
- [197] S. A. STANSFELD et M. P. MATHESON. « Noise pollution : non-auditory effects on health ». In : *British medical bulletin* 68.1 (2003), p. 243-257.
- [198] *STBM 271 dust sensor module*. B+B Sensors Corp. 2014.
- [199] G. P. STEIN. « Lens distortion calibration using point correspondences ». In : *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE. 1997, p. 602-608.
- [200] J. M. STOCKIE. « The mathematics of atmospheric dispersion modeling ». In : *Siam Review* 53.2 (2011), p. 349-372.
- [201] X. SU et T. M. KHOSHGOFTAAR. « A survey of collaborative filtering techniques ». In : *Advances in artificial intelligence* 2009 (2009), p. 4.
- [202] M. J. TAGHIZADEH, R. PARHIZKAR, P. N. GARNER et H. BOURLARD. « Euclidean distance matrix completion for ad-hoc microphone array calibration ». In : *Digital Signal Processing (DSP), 2013 18th International Conference on*. IEEE. 2013, p. 1-7.
- [203] S. TECHNOLOGY. *Particle Sensor Unit-How It Works*. 2017. URL : www.shinyei.co.jp/stc/eng/optical/main_dust.html.
- [204] M. TEPPER et G. SAPIRO. « Compressed Nonnegative Matrix Factorization Is Fast and Accurate ». In : *IEEE Transactions on Signal Processing* 64.9 (mai 2016), p. 2269-2283.
- [205] R. TIBSHIRANI. « Regression shrinkage and selection via the lasso ». In : *Journal of the Royal Statistical Society. Series B (Methodological)* (1996), p. 267-288.
- [206] A. TILLOY, V. MALLET, D. POULET, C. PESIN et F. BROCHETON. « BLUE-based NO2 data assimilation at urban scale ». In : *Journal of Geophysical Research : Atmospheres* 118.4 (2013), p. 2031-2040.
- [207] O. TROYANSKAYA, M. CANTOR, G. SHERLOCK, P. BROWN, T. HASTIE, R. TIBSHIRANI, D. BOTSTEIN et R. B. ALTMAN. « Missing value estimation methods for DNA microarrays ». In : *Bioinformatics* 17.6 (2001), p. 520-525.
- [208] R. TSAI. « A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses ». In : *IEEE Journal on Robotics and Automation* 3.4 (1987), p. 323-344.

- [209] D. B. TURNER. *Workbook of atmospheric dispersion estimates : an introduction to dispersion modeling*. CRC press, 1994.
- [210] A. VANDAELE, N. GILLIS, F. GLINEUR et D. TUYTTENS. « Heuristics for exact non-negative matrix factorization ». In : *Journal of Global Optimization* 65.2 (juin 2016), p. 369-400.
- [211] S. A. VAVASIS. « On the Complexity of Nonnegative Matrix Factorization ». In : *SIAM Journal on Optimization* 20.3 (2010), p. 1364-1377.
- [212] S. S. VEMPALA. *The random projection method*. T. 65. American Mathematical Soc., 2005.
- [213] A. VERGARA, S. VEMBU, T. AYHAN, M. A. RYAN, M. L. HOMER et R. HUERTA. « Chemical gas sensor drift compensation using classifier ensembles ». In : *Sensors and Actuators B : Chemical* 166 (2012), p. 320-329.
- [214] X. T. VU, S. MAIRE, C. CHAUX et N. THIRION-MOREAU. « A New Stochastic Optimization Algorithm to Decompose Large Nonnegative Tensors ». In : *IEEE Signal Processing Letters* 22.10 (oct. 2015), p. 1713-1717.
- [215] G. WAHBA. *Spline models for observational data*. SIAM, 1990.
- [216] C. WANG, P. RAMANATHAN et K. K. SALUJA. « Blindly calibrating mobile sensors using piecewise linear functions ». In : *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON'09. 6th Annual IEEE Communications Society Conference on*. IEEE. 2009, p. 1-9.
- [217] C. WANG, P. RAMANATHAN et K. K. SALUJA. « Calibrating Nonlinear Mobile Sensors ». In : *2008 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*. Juin 2008, p. 533-541.
- [218] C. WANG, P. RAMANATHAN et K. K. SALUJA. « Moments Based Blind Calibration in Mobile Sensor Networks ». In : *Proc. of ICC '08*. Mai 2008, p. 896-900.
- [219] F. WANG et P. LI. « Efficient nonnegative matrix factorization with random projections ». In : *Proceedings of the 2010 SIAM International Conference on Data Mining*. SIAM. 2010, p. 281-292.
- [220] J. WANG, F. TIAN, C. H. LIU, H. YU, X. WANG et X. TANG. « Robust nonnegative matrix factorization with ordered structure constraints ». In : *Neural Networks (IJCNN), 2017 International Joint Conference on*. IEEE. 2017, p. 478-485.
- [221] Y. WANG, A. YANG, X. CHEN, P. WANG, Y. WANG et H. YANG. « A Deep Learning Approach for Blind Drift Calibration of Sensor Networks ». In : *IEEE Sensors Journal* (2017).
- [222] Y. WANG, A. YANG, Z. LI, X. CHEN, P. WANG et H. YANG. « Blind drift calibration of sensor networks using sparse Bayesian learning ». In : *IEEE Sensors Journal* 16.16 (2016), p. 6249-6260.
- [223] Y. WANG, A. YANG, Z. LI, P. WANG et H. YANG. « Blind drift calibration of sensor networks using signal space projection and Kalman filter ». In : *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*. IEEE. 2015, p. 1-6.

- [224] K. WHITEHOUSE et D. CULLER. « Calibration as parameter estimation in sensor networks ». In : *Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*. ACM. 2002, p. 59-67.
- [225] J. WRIGHT, A. GANESH, S. RAO, Y. PENG et Y. MA. « Robust principal component analysis : Exact recovery of corrupted low-rank matrices via convex optimization ». In : *Advances in neural information processing systems*. 2009, p. 2080-2088.
- [226] J. WRIGHT, A. Y. YANG, A. GANESH, S. S. SASTRY et Y. MA. « Robust face recognition via sparse representation ». In : *IEEE transactions on pattern analysis and machine intelligence* 31.2 (2009), p. 210-227.
- [227] S. L. WRIGHT, R. C. THOMPSON et T. S. GALLOWAY. « The physical impacts of microplastics on marine organisms : A review ». In : *Environmental Pollution* 178 (2013), p. 483-492.
- [228] Y. XIANG, B. L. S., R. PLEDRAHITA, D. R. P., L. Q., H. M. et S. L. « Collaborative calibration and sensor placement for mobile sensor networks ». In : *2012 ACM/IEEE 11th International Conference on Information Processing in Sensor Networks (IPSN)*. Avr. 2012, p. 73-83.
- [229] W. XU, X. LIU et Y. GONG. « Document clustering based on non-negative matrix factorization ». In : *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM. 2003, p. 267-273.
- [230] Y. XU, W. YIN, Z. WEN et Y. ZHANG. « An alternating direction algorithm for matrix completion with nonnegative factors ». In : *Frontiers of Mathematics in China* 7.2 (2012), p. 365-384.
- [231] J. YOO, M. KIM, K. KANG et S. CHOI. « Nonnegative matrix partial co-factorization for drum source separation ». In : *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. IEEE. 2010, p. 1942-1945.
- [232] L. ZHANG et S. ZHANG. « A Unified Joint Matrix Factorization Framework for Data Integration ». In : *arXiv preprint arXiv :1707.08183* (2017).
- [233] Q. ZHANG et R. PLESS. « Extrinsic calibration of a camera and laser range finder (improves camera calibration) ». In : *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*. T. 3. IEEE. 2004, p. 2301-2306.
- [234] S. ZHANG, W. WANG, J. FORD et F. MAKEDON. « Learning from incomplete ratings using non-negative matrix factorization ». In : *Proceedings of the 2006 SIAM International Conference on Data Mining*. SIAM. 2006, p. 549-553.
- [235] Z. ZHANG, Y. XU, J. YANG, X. LI et D. ZHANG. « A survey of sparse representation : algorithms and applications ». In : *IEEE access* 3 (2015), p. 490-530.
- [236] Z. ZHANG. « A flexible new technique for camera calibration ». In : *IEEE Transactions on pattern analysis and machine intelligence* 22.11 (2000), p. 1330-1334.
- [237] R. ZHAO et V. Y. TAN. « Online nonnegative matrix factorization with outliers ». In : *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE. 2016, p. 2662-2666.

- [238] R. ZHAO, V. Y. TAN et H. XU. « Online nonnegative matrix factorization with general divergences ». In : *arXiv preprint arXiv :1608.00075* (2016).
- [239] G. ZHOU, A. CICHOCKI et S. XIE. « Fast nonnegative matrix/tensor factorization based on low-rank approximation ». In : *IEEE Transactions on Signal Processing* 60.6 (2012), p. 2928-2940.
- [240] G. ZHOU, A. CICHOCKI, Q. ZHAO et S. XIE. « Nonnegative matrix and tensor factorizations : An algorithmic perspective ». In : *IEEE Signal Processing Magazine* 31.3 (2014), p. 54-65.
- [241] T. ZHOU et D. TAO. « Godec : Randomized low-rank & sparse matrix decomposition in noisy case ». In : *International conference on machine learning*. Omnipress. 2011.

Résumé — Le mobile crowdsensing consiste à acquérir des données géolocalisées et datées d'une foule de capteurs mobiles (issus de ou connectés à des smartphones). Dans cette thèse, nous nous intéressons au traitement de données issues du mobile crowdsensing environnemental. En particulier, nous proposons de revisiter le problème d'étalonnage aveugle de capteurs comme un problème informé de factorisation matricielle à données manquantes, où les facteurs contiennent respectivement le modèle d'étalonnage fonction du phénomène physique observé (nous proposons des approches pour des modèles affines et non-linéaires) et les paramètres d'étalonnage de chaque capteur. Par ailleurs, dans l'application de surveillance de la qualité de l'air que nous considérons, nous supposons avoir à notre disposition des mesures très précises mais distribuées de manière très parcimonieuse dans le temps et l'espace, que nous couplons aux multiples mesures issues de capteurs mobiles. Nos approches sont dites informées car (i) les facteurs matriciels sont structurés par la nature du problème, (ii) le phénomène observé peut être décomposé sous forme parcimonieuse dans un dictionnaire connu ou approché par un modèle physique/géostatistique, et (iii) nous connaissons la fonction d'étalonnage moyenne des capteurs à étalonner. Les approches proposées sont plus performantes que des méthodes basées sur la complétion de la matrice de données observées ou les techniques multi-sauts de la littérature, basées sur des régressions robustes. Enfin, le formalisme informé de factorisation matricielle nous permet aussi de reconstruire une carte fine du phénomène physique observé.

Mots clés : Étalonnage de capteurs, factorisation matricielle, estimation de données manquantes, internet des objets, mobile crowdsensing

Abstract — Mobile crowdsensing aims to acquire geolocated and timestamped data from a crowd of sensors (from or connected to smartphones). In this thesis, we focus on processing data from environmental mobile crowdsensing. In particular, we propose to revisit blind sensor calibration as an informed matrix factorization problem with missing entries, where factor matrices respectively contain the calibration model which is a function of the observed physical phenomenon (we focus on approaches for affine or nonlinear sensor responses) and the calibration parameters of each sensor. Moreover, in the considered air quality monitoring application, we assume to possess some precise measurements—which are sparsely distributed in space and time—that we melt with multiple measurements from the mobile sensors. Our approaches are "informed" because (i) factor matrices are structured by the problem nature, (ii) the physical phenomenon can be decomposed using sparse decomposition with a known dictionary or can be approximated by a physical or a geostatistical model, and (iii) we know the mean calibration function of the sensors to be calibrated. The proposed approaches demonstrate better performances than the one based on the completion of the observed data matrix or the multi-hop calibration method from the literature, based on robust regression. Finally, the informed matrix factorization formalism also provides an accurate reconstruction of the observed physical field.

Keywords : Sensor calibration, matrix factorization, missing value estimation, internet of things, mobile crowdsensing

Laboratoire d'Informatique Signal et Image de la Côte d'Opale, EA 4491
Maison de la Recherche – 50 rue F. Buisson, B.P. 719, 62228 Calais Cedex, France