



**DOSSIER DE CANDIDATURE
A UNE ALLOCATION DE RECHERCHE
POUR LA RENTREE 2019**

Dossier complété et revêtu des signatures à transmettre impérativement pour
le :

22 mars 2019 au plus tard,

A la Direction de la Recherche et Valorisation
secretariat.recherche@univ-littoral.fr

Titre de la thèse :

Apprentissage automatique profond par (co-)factorisation matricielle multi-objective évolutionnaire. Applications en science des données industrielles, environnementales et marines.

Laboratoire d'accueil ULCO : LISIC

Priorité du laboratoire, tous supports de financements confondus :

Directeur de thèse ULCO : Gilles ROUSSEL (PR)

X LIBAN - Université Libanaise

Pour ce dispositif, merci d'indiquer en plus :

- le nom du codirecteur étranger et le laboratoire partenaire

En cours d'identification

- Thématique : Autres

- (1) La qualité de l'air
- (2) Le milieu aquatique
- (3) L'obésité, la nutrition et les activités sportives
- (4) Les énergies propres et renouvelables
- (5) La gestion et le traitement des déchets
- (6) L'urbanisme

X LIBAN - CNRS Libanais

Pour ce dispositif, merci d'indiquer en plus :

- le nom du codirecteur étranger et le laboratoire partenaire

- Thématique : Autre



- (1) La qualité de l'air
- (2) Le milieu aquatique
- (3) L'obésité, la nutrition et les activités sportives
- (4) Les énergies propres et renouvelables
- (5) La gestion et le traitement des déchets
- (6) L'urbanisme

***LABORATOIRE D'ACCUEIL**

Nombre de HDR dans le laboratoire : **16 HDR**

Nombre de thèses encadrées dans le laboratoire (rentrée 2018) : **20**

Durée moyenne des thèses soutenues dans le laboratoire, sur la période 2014-2018 : 3.5 années

ENCADREMENT

Nom, Prénom du directeur de laboratoire : **Christophe RENAUD**

Nom, Prénom du directeur de thèse (si différent du directeur de laboratoire) :
Gilles ROUSSEL

Nombre de doctorats en préparation sous la direction du directeur de thèse : **1**

Avis détaillé du directeur de thèse :

Ce sujet aborde une nouvelle extension de l'outil NMF, déjà bien maîtrisé et utilisé dans l'équipe pour différentes applications. Il est ici envisagé de contribuer au concept de "deep learning" par une approche naissante et alternative au réseau de neurones profond, appelée NMF profond dans l'état de l'art. Dans ces travaux prospectifs, Matthieu Puigt propose une co-factorisation contrainte, optimisée, en conjuguant des facteurs linéaires et non linéaires, comme ceux apparaissant en réseau de neurones. Cependant, cette structure particulière de factorisation permettant une formulation plus parcimonieuse, doit démontrer un apprentissage automatique plus rapide car nécessitant moins d'échantillons d'apprentissage que le réseau de neurones profond. Ces travaux pourront s'appliquer à une palette d'applications dans les domaines industriels et environnementaux. Je soutiens bien sûr très favorablement ces travaux originaux.

Gilles ROUSSEL

Signature du directeur de thèse



Avis détaillé du directeur de laboratoire :

Le sujet présenté dans ce dossier est original et potentiellement novateur. Il s'appuie sur des compétences avérées des co-encadrants de notre laboratoire et doit permettre une amélioration des techniques d'apprentissage « profond », tout en réduisant la quantité de données à utiliser, qui reste un frein à l'extension des techniques de « deep-learning ». Les applications des études théoriques proposées sont nombreuses et couvrent potentiellement la plupart des domaines prioritaires du dispositif de co-financement ULCO/CNRS libanais. J'emets donc un avis favorable au financement de ce sujet.

Signature du directeur de laboratoire

A handwritten signature in black ink, consisting of a large, stylized loop followed by a horizontal line extending to the right.



PROJET DE THESE

Intitulé du projet de thèse : **Apprentissage automatique profond par (co-)factorisation matricielle multi-objective évolutive. Applications en science des données industrielles, environnementales et marines**

Domaine scientifique : **Apprentissage pour le traitement du signal**

Résumé (1/2 page maxi.) :

La factorisation matricielle non-négative (NMF) consiste à écrire une matrice de données non-négatives comme une combinaison linéaire de facteurs latents non-négatifs, c.-à-d. des variables expliquant les données. Cet outil est extrêmement populaire en apprentissage (machine learning), traitement du signal et traitement des images, de par la meilleure interprétabilité qu'il permet par rapport à des approches sans contrainte de signe. Notre équipe a notamment proposé plusieurs méthodes de NMF dites *informées* pour des problématiques de (i) séparation de sources de pollution et (ii) de cartographie de la qualité de l'air par des réseaux de capteurs aux réponses hétérogènes.

Dans les années 2000, les outils de NMF ont permis des avancées notables en apprentissage supervisé et non-supervisé. Aujourd'hui, les techniques de *deep learning*, utilisant un grand nombre de couches cachées dans des réseaux de neurones, ont permis d'atteindre des résultats sans précédent en apprentissage supervisé, notamment en classification d'images et en reconnaissance de la parole. Cependant, les approches de NMF restent l'état de l'art en apprentissage non-supervisé.

De manière intéressante, l'idée de l'apprentissage profond a été reprise sur des techniques de NMF, connues sous le nom de *deep NMF* ou *multi-layered NMF*, où plusieurs factorisations en couches permettent une meilleure décomposition des données observées. L'objectif de cette thèse consistera à développer de nouvelles approches de Deep NMF. L'originalité de ces travaux consistera à combiner ces techniques profondes à des approches d'optimisation multi-objectif évolutive pour gérer de manière fine l'ensemble des contraintes supplémentaires du problème à traiter.

Ces méthodes pourront être appliquées en surveillance industrielle, en lien avec l'internet des objets pour l'industrie (industrie 4.0) et à des problèmes de démixage en télédétection et vision embarquée marine ou terrestre.



Projet de thèse (5 pages maxi.) :

Apprentissage automatique profond par (co-)factorisation matricielle multi-objective évolutionnaire. Applications en science des données industrielles, environnementales et marines

Mots clés : *factorisation en matrices (semi-)non-négatives ; données de grande dimension ; optimisation multi-objective évolutionnaire ; factorisation multi-couches ; apprentissage profond ; surveillance marine ; surveillance industrielle ; environnement*

I. Contexte scientifique :

Depuis l'invention de l'ordinateur, l'homme a très tôt voulu le doter d'une capacité d'analyse pour classer ou reconnaître des événements. Ainsi, dès les années 1940, des méthodes d'apprentissage étaient basées sur des réseaux de neurones [1]. Cependant, les capacités de calcul des machines étant extrêmement limitées, le nombre de couches des réseaux l'était lui-aussi. En conséquence, des alternatives demandant moins de ressources en calcul ont été proposées par la suite. Parmi celles-ci, les techniques basées sur la factorisation matricielle non-négative (NMF) ont connu un vif intérêt de la part de la communauté dès la fin des années 1990 et les travaux de Lee et Seung [2]. Elles ont notamment été appliquées avec succès en traitement d'images [2], traitement du son [3], télédétection [4], chimie environnementale [5], analyse de réseaux sociaux [6], ou encore en étalonnage de capteurs mobiles [7].

L'apprentissage profond (ou *deep learning* en anglais) a révolutionné l'intelligence artificielle depuis une demi-douzaine d'années [8]. Historiquement, les premières approches proposées ont repris le formalisme de méthodes développées à la fin des années 1980 qui utilisaient les réseaux de neurones, mais en rajoutant un grand nombre de couches cachées. Elles permettent notamment de réaliser un apprentissage de manière supervisée avec des performances inégalées, sous réserve d'avoir une base d'apprentissage de très grande taille (*big data*). Aujourd'hui, certains travaux en intelligence artificielle consistent à reconsidérer des alternatives aux réseaux de neurones classiques et d'en proposer des extensions profondes. Comme les approches alternatives classiques étaient beaucoup plus performantes que les approches à base de réseaux de neurones historiques (c.-à-d. avec peu de couches cachées), on peut "espérer" que leurs extensions profondes seront plus performantes que les approches à base de réseaux profonds de neurones, c.-à-d. qu'ils pourraient fournir des performances similaires avec une base d'apprentissage beaucoup plus réduite, ce qui pourrait avoir des conséquences très positives en terme de besoins énergétiques (par exemple grâce aux calculs plus rapides, du fait de la base plus réduite).

Il est démontré depuis longtemps qu'une couche de réseau de neurones peut être ré-écrite comme un produit matriciel [2]. Quelques travaux ont ainsi débuté en NMF profonde¹ (sous le nom de *deep NMF* [9] ou *multi-layered NMF* [10] dans la littérature) et ont montré comment ces approches pouvaient se

¹Notons aussi l'existence de travaux sur des factorisations profondes semi-non-négatives [11].



comporter comme des classifieurs hiérarchiques qui permettraient d'apprendre, dans chaque sous-couche du réseau de factorisation, certains patterns d'intérêt. Cependant, de nombreux efforts restent à faire. En particulier, les approches de *deep learning* classiques nécessitent le réglage fin de différentes contraintes dans chaque couche cachée du réseau de neurones. Ce réglage est donc très long et fastidieux [8], même si des travaux très récents en optimisation cherchent à apprendre la structure et les hyper-paramètres optimaux du réseau profond [12]. Cet apprentissage automatique sera le principal verrou scientifique du sujet de thèse proposé, pour des techniques de NMF profonde. Ce travail s'effectuera notamment en collaboration avec l'équipe OSMOSE du LISIC, qui possède une expertise en optimisation multi-objectif [13]. Un petit nombre de travaux existe en NMF-MOE [14-16] : Les travaux dans [14] et [16] proposent de combiner NMF avec des stratégies de recuit simulé et d'optimisation évolutionnaire pour contraindre un niveau de parcimonie sur un des facteurs. Les travaux dans [15] proposent de trouver un compromis entre deux NMF différentes, l'une étant le pendant non-linéaire de l'autre. Pour cela, une extension de la NMF, consistant en une co-factorisation (CoNMF) est proposée et le front de Pareto est étudié de manière extensive (sans aucun apport de l'optimisation évolutionnaire). Nous proposons de nous appuyer sur ces travaux pour les généraliser.

Comme indiqué précédemment, les méthodes de (Co)NMF permettent un grand nombre d'applications. Dans le cadre de ce projet, d'un point de vue applicatif, nous nous focaliserons sur :

- l'analyse de données issues de mesures ubiquitaires pour l'Industrie 4.0, avec des applications possibles en analyse de la qualité de la production. Cette application s'inscrit notamment dans le cadre des activités de recherche émergentes sur le site audomarois du LISIC ;

le démixage d'images en télédétection marine (en collaboration avec le LOG avec qui des discussions sont déjà engagées). Ce travail s'inscrit dans le cadre de la SFR Mer qui se monte sur le Littoral et dans le cadre du CPER Marco.

II. Sujet de recherche :

Ce sujet de thèse sera découpé en 2 parties :

- 1.** Dans un premier temps, nous chercherons à proposer une méthode "universelle" de NMF utilisant l'optimisation multi-objective (MOE). Nos méthodes de NMF-MOE et CoNMF-MOE tiendront compte d'un ensemble fonctions de contraintes propres aux problèmes considérés et proposeront un ensemble de solutions "optimales au sens de Pareto" pour l'aide à la décision. L'objectif est de généraliser les travaux [14-16] en proposant de gérer bien plus de contraintes que ce qui est actuellement réalisé dans la littérature.

Dans un deuxième temps, nous chercherons à automatiser l'apprentissage de ces fonctions de contraintes en cherchant à éliminer automatiquement celles qui seraient inutiles pour répondre au problème d'apprentissage. Pour cela, un formalisme original consisterait à apprendre les hyper-



paramètres de la (Co)NMF-MOE comme un problème de MOE en soit. Ce premier travail est déjà ambitieux car il permettrait de sérieuses avancées à la fois en apprentissage pour le traitement du signal et des images, mais aussi en recherche opérationnelle qui est le domaine scientifique dans lequel l'optimisation MOE est la plus active.

Dans un deuxième temps, il s'agira d'étendre ces approches "universelles" au cas de la factorisation profonde. Chaque couche cachée de (Co)NMF sera potentiellement contrainte et il s'agira de trouver, pour chaque couche, les contraintes les mieux adaptées au problème. Ces techniques automatisées pourraient alors concurrencer les approches basées sur des réseaux de neurones pour l'analyse de données.



III. Travail et échéancier proposés :

Tâche	Durée	Période
Bibliographie	6 mois	De Mois-1 à Mois-34
Développement d'approches de (Co)NMF-MOE	12 mois	De Mois-3 à Mois-18
Développement d'approches de Deep (Co)NMF-MOE	9 mois	De Mois-18 à Mois-33
Validation méthodes et rédaction d'articles	6 mois	De Mois-18 à Mois-34
Rédaction thèse	3 mois	De Mois-30 à Mois-36
Total	36 mois	De Mois-1 à Mois-36

IV. Etat du sujet dans le laboratoire et l'équipe d'accueil :

La NMF fait l'objet de nombreux travaux dans l'équipe depuis plusieurs années. Nous avons notamment obtenu des contrats industriels (ArcelorMittal) et institutionnels (DREAL, Région, CNRS) autour de cette thématique.

Nous avons pu développer de nouvelles approches robustes et informées que nous appliquées à la séparation de sources de pollution atmosphérique, en collaboration avec l'UCEIV de l'ULCO. Un logiciel, nommé SPamApp a notamment été proposé en ce sens.

Nous avons aussi proposé des approches spécifiques pour conjointement étalonner des réseaux de capteurs mobiles de la qualité de l'air (internet des objets) et réaliser des cartographies de la qualité de l'air. Dans cette application, la matrice de données à factoriser est de grande taille mais partiellement connue (présence de nombreuses données manquantes). Nous avons proposé plusieurs approches structurées et informées de NMF pondérée à cet effet (certains codes Matlab de ces méthodes que nous avons proposées sont publiquement accessibles à : https://gogs.univ-littoral.fr/puigt/Informed_NMF_Mobile_Sensor_Calibration/).

Actuellement, nous travaillons à l'accélération de ces approches d'étalonnage, compte tenu de la grande masse de données fournies, en combinant des solveurs optimaux (basés sur les méthodes des moments - de manière similaire à ce qui est fait en *deep learning*) et des projections aléatoires pour réduire la taille des données à traiter. De même que précédemment, certains de nos codes sont accessibles en ligne : https://gogs.univ-littoral.fr/puigt/Faster-than-fast_NMF.



Le travail de thèse que nous proposons ici serait donc parfaitement complémentaire des travaux de l'équipe puisque nous tirerions partie de notre expertise et de l'expertise de l'équipe OSMOSE du LISIC pour mener à bien ces travaux.

V. Retombées scientifiques et économiques attendues :

Ces travaux nous permettront de nous positionner de manière originale par rapport aux communautés en traitement du signal et des images d'une part, en apprentissage automatique (*machine learning*) d'autre part, et enfin en recherche opérationnelle. Ces travaux nous permettront de renforcer les collaborations internes au LISIC, nous permettant ainsi de pouvoir répondre de manière originale à des appels à projets nationaux (ANR) et européens (FEDER, H2020). Par ailleurs, nous profiterons de cette expérience pour accompagner le développement des entreprises du territoire, en lien avec l'intelligence artificielle et l'aide à la décision.

VI. Collaborations prévues :

Comme indiqué précédemment, ces travaux seront réalisés avec le soutien d'une autre équipe du LISIC, spécialiste de l'optimisation multi-objective évolutionnaire.

Les approches proposées pourront être appliquées dans un grand nombre de domaines en lien avec les axes prioritaires de la Région et de l'ULCO :

1. surveillance industrielle/internet des objets pour l'industrie (en lien avec les activités de recherche du laboratoire sur Saint-Omer), télédétection marine (en collaboration avec le laboratoire LOG de l'ULCO/Univ. Lille).

Par ailleurs et comme indiqué précédemment, nous sommes aussi en discussion avec plusieurs entreprises de la région, pour mettre en place des collaborations autour de l'apprentissage par NMF autour de leurs applications. Ce travail de thèse nous permettrait de pouvoir prendre une avance technique et scientifique pour mieux les accompagner dans leurs développements, par exemple dans le cadre de futurs partenariats public/privé.

VII. Publications significatives de l'équipe en lien avec le sujet (membres actuels de l'équipe en gras, doctorants de l'équipe soulignés) :

1. **F. Yahaya**, **M. Puigt**, **G. Delmaire**, **G. Roussel**, *Weighted NMF using random projections and Nesterov iterations*, soumis.
1. C. Dorffer, **M. Puigt**, **G. Delmaire**, **G. Roussel**, *Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration*, IEEE Transactions on Signal and Information Processing over Networks, Volume 4, Issue 4, pp. 667-682, December 2018.
2. **F. Yahaya**, **M. Puigt**, **G. Delmaire**, **G. Roussel**, *Faster-than-fast NMF using random projections and Nesterov iterations*, in Proc. of iTWIST'18,



- Marseille, France, November 21-23, 2018.
3. **C. Dorffer, M. Puigt, G. Delmaire, G. Roussel**, *Fast nonnegative matrix factorization and completion using Nesterov iterations*, in Proc. of LVA/ICA'17, Springer International Publishing AG, vol. LNCS 10179, pp. 26-35, Grenoble, France, February 21-24, 2017.
 4. **S. Verel**, *Apport à l'analyse des paysages de fitness pour l'optimisation mono-objective et multiobjective*, Thèse d'habilitation à diriger des recherches, ULCO, Décembre 2016.
 5. **R. Chreiky, G. Delmaire, C. Dorffer, M. Puigt, G. Roussel**, A. Abche, *Robust informed split gradient NMF using Alpha Beta-divergence for source apportionment*, in Proc. of MLSP'16, Vietri Sul Mare, Salerno, Italy, September 13-16, 2016.
 6. **C. Dorffer, M. Puigt, G. Delmaire, G. Roussel**, *Nonlinear mobile sensor calibration using informed semi-nonnegative matrix factorization with a Vandermonde factor*, in Proc. of SAM'16, Rio de Janeiro, Brazil, July 10-13, 2016.
 7. **A. Limem, G. Delmaire, M. Puigt, G. Roussel**, D. Courcot, *Non-negative matrix factorization under equality constraints—a study of industrial source identification*, Applied Numerical Mathematics (APNUM), Volume 85, pp. 1-15, November 2014.

VIII. Références :

- [1] McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4), 115-133.
- [2] Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755), 788.
- [3] Virtanen, T. (2007). Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *IEEE transactions on audio, speech, and language processing*, 15(3), 1066-1074.
- [4] Ma, W. K., Bioucas-Dias, J. M., Chan, T. H., Gillis, N., Gader, P., Plaza, A. J., ... & Chi, C. Y. (2014). A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine*, 31(1), 67-81.
- [5] Limem, A., Delmaire, G., Puigt, M., Roussel, G., & Courcot, D. (2014). Non-negative matrix factorization under equality constraints—a study of industrial source identification. *Applied Numerical Mathematics*, 85, 1-15.
- [6] Yang, J., & Leskovec, J. (2013, February). Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 587-596). ACM.
- [7] Dorffer, C., Puigt, M., Delmaire, G., & Roussel, G. (2018, December). Informed Nonnegative Matrix Factorization Methods for Mobile Sensor Network Calibration. *IEEE Transactions on Signal and Information Processing over Networks*, 4(4):667-682.
- [8] Cardon, D., Cointet, J. P., & Mazières, A. (2018). La revanche des neurones. *Réseaux*, (5), 173-220.
- [9] Le Roux, J., Hershey, J. R., & Wenginger, F. (2015, April). Deep NMF for speech separation. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE*



International Conference on (pp. 66-70). IEEE.

[10] Cichocki, A., Zdunek, R., & Amari, S. I. (2007, September). Hierarchical ALS algorithms for nonnegative matrix and 3D tensor factorization. In *International Conference on Independent Component Analysis and Signal Separation* (pp. 169-176). Springer, Berlin, Heidelberg.

[11] Trigeorgis, G., Bousmalis, K., Zafeiriou, S., & Schuller, B. W. (2017). A deep matrix factorization method for learning attribute representations. *IEEE transactions on pattern analysis and machine intelligence*, 39(3), 417-429.

[12] Assunção, F., Lourenço, N., Machado, P., & Ribeiro, B. (2018). DENSER: Deep Evolutionary Network Structured Representation. *arXiv preprint arXiv:1801.01563*.

[13] Verel S. (2016, December), Apport à l'analyse des paysages de fitness pour l'optimisation mono-objective et multiobjective, *Thèse d'habilitation à diriger des recherches*, ULCO.

[14] Foley, K., Greene, D., & Cunningham, P. (2010, August). Optimizing conflicting objectives in NMF using Pareto simulated annealing. In *Paper presented at the 21st National Conference on Artificial Intelligence and Cognitive Science (AICS 2010), Galway, Ireland, 30 August-1 September, 2010*.

[15] Zhu, F., & Honeine, P. (2016). Biobjective nonnegative matrix factorization: Linear versus kernel-based models. *IEEE Transactions on Geoscience and Remote Sensing*, 54(7), 4012-4022.

[16] Gong, M., Jiang, X., Li, H., & Tan, K. C. (2018). Multiobjective Sparse Non-Negative Matrix Factorization. *IEEE Transactions on Cybernetics*.