



**HAL**  
open science

## Factorisation Matricielle Semi Non-Négative: Application à la Décomposition de Consommations Electriques

Simon Henriet, Umut Şimşekli, Sérgio F. Santos, Benoît Fuentes, Gael Richard

► **To cite this version:**

Simon Henriet, Umut Şimşekli, Sérgio F. Santos, Benoît Fuentes, Gael Richard. Factorisation Matricielle Semi Non-Négative: Application à la Décomposition de Consommations Electriques. Colloque francophone de traitement du signal et des images (GRETSI), Aug 2019, Lille, France. hal-02381367

**HAL Id: hal-02381367**

**<https://telecom-paris.hal.science/hal-02381367>**

Submitted on 26 Nov 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.

# Factorisation Matricielle Semi Non-Négative: Application à la Décomposition de Consommations Electriques

Simon HENRIET<sup>1,2</sup>, Umut SIMSEKLI<sup>1</sup>, Sergio DOS SANTOS<sup>2</sup>, Benoit FUENTES<sup>2</sup>, Gaël RICHARD<sup>1</sup>

<sup>1</sup>LTCI, Télécom ParisTech, Université Paris-Saclay  
46 Rue Barrault, Paris, France

<sup>2</sup>Smart Impulse  
157 Boulevard Macdonald, Paris, France  
simon.henriet@telecom-paristech.fr

**Résumé** – Depuis de nombreuses années, la mesure et le suivi des consommations électriques dans les bâtiments résidentiels et commerciaux comme les bureaux, les centres commerciaux ou les entrepôts ont connu un essor important. Cependant, obtenir la consommation individuelle des équipements à partir de la consommation totale (NILM), est un problème complexe. Plusieurs approches ont été proposées dans le cadre des bâtiments résidentiels. Des résultats prometteurs ont notamment été obtenus par le biais de techniques de factorisation de matrices appliquées aux mesures haute fréquence de la tension et du courant. Ces méthodes ne sont pas efficaces lorsqu'on les applique aux bâtiments commerciaux. Dans ce papier, nous proposons une nouvelle méthode de factorisation basée sur une extension de la factorisation semi non négative de matrices (SNMF) à laquelle est ajoutée une pénalisation de la variation totale (TV-SNMF). Pour résoudre ce problème d'optimisation sous contraintes, nous avons développé une stratégie d'optimisation alternée qui utilise une méthode quasi-Newton. Les expériences sur une base de données de simulations de bâtiments commerciaux montrent clairement un gain d'efficacité comparé à d'autres approches comme l'analyse en composantes indépendantes (ICA) ou la SNMF classique.

**Abstract** – In the recent years, there has been an increasing academic and industrial interest for analysing the electrical consumption of commercial buildings. One approach to enable energy efficiency is to disaggregate total energy consumptions into individual ones. This problem is also called Non Intrusive Load Monitoring (NILM). While several approaches have been studied to solve it for residential building using high frequency current and voltage measurements, none of them seems efficient applied to commercial buildings. Amongst the NILM method for residential buildings, matrix factorization approached showed promising results. In this paper, we propose a novel method as an extension of factorization techniques based on Semi Non-Negative Matrix Factorization constrained with a total variation penalization (TV-SNMF). To solve this constrained optimization problem, we rely on an alternating minimization strategy involving a quasi-newton algorithm. The experiments on a simulated commercial building dataset demonstrate clear improvements compared to other approaches such as Independent Component Analysis (ICA) and classic SNMF.

## 1 Introduction

Dans un contexte d'efficacité énergétique, la surveillance non intrusive de courbes de charges (*Non Intrusive Load Monitoring* ou NILM) consiste à utiliser un capteur électrique ainsi qu'un algorithme de désagrégation de consommations [1]. Le NILM peut être formulé comme un problème de séparation de sources dans lequel les mesures de courant sont un mélange linéaire de sources inconnues correspondant à différents appareils électriques. Le but est donc de retrouver les sources inconnues à partir de l'observation du mélange.

Dans [2], les auteurs utilisent une factorisation de matrices semi-binaire appliquée aux mesures de périodes du courant (une période de courant correspond à un échantillonnage du courant pendant une période de tension, 20 ms pour un réseau en courant alternatif à 50 Hz). La matrice de mesure (une colonne par période) est ainsi factorisée en : (i) une matrice réelle représentant la forme caractéristique du courant de chaque ap-

pareil et (ii) une matrice binaire décrivant le caractère éteint ou allumé des appareils au cours du temps.

Bien que cette méthode obtienne des résultats encourageants sur des bâtiments résidentiels, elle montre des limitations en ce qui concerne les bâtiments commerciaux. En effet, les consommations électriques dans ces bâtiments peuvent être très spécifiques [3, 4]. En particulier, la présence d'appareils ayant une consommation variant continûment au cours du temps est en contradiction avec la contrainte binaire discutée précédemment. Pour contourner ces limitations, nous proposons d'utiliser le cadre de la factorisation semi non négative de matrices (SNMF) [5], où la matrice binaire est remplacée par une matrice non négative.

Dans ce travail, nous étendons la SNMF en introduisant une pénalisation de la variation totale de la matrice non négative (TV-SNMF), motivée par la dépendance temporelle dans les consommations des appareils (section 2). Le choix de la variation totale peut être interprété comme le moyen de prendre en

compte l'hypothèse de parcimonie dans la dérivée des consommations ([4]). Pour résoudre ce problème d'optimisation sous contrainte, nous avons développé une stratégie d'optimisation alternée basée sur une méthode quasi-Newton (section 3). Finalement, nous avons étudié les performances de cette méthode sur une base de données NILM simulant des consommations de bâtiments commerciaux (section 4). Cette approche, utilisant des données haute fréquence de courant issues de bâtiments commerciaux, constitue une nouveauté.

## 2 SNMF et variation totale

SNMF est un problème classique de factorisation matricielle dans lequel une matrice d'observation est approximée par le produit de deux facteurs :  $X \approx SA$ , avec la contrainte que  $S$  soit une matrice à valeurs réelles alors que  $A$  est une matrice à valeurs non négatives. La SNMF s'écrit comme le problème d'optimisation suivant [5] :

$$\begin{aligned} & \underset{S,A}{\text{minimise}} \quad \frac{1}{2} \|X - SA\|_{Fro}^2 & (1) \\ & \text{sachant que} \quad A \geq 0 \end{aligned}$$

où  $X \in \mathbb{R}^{N \times T}$  est la matrice d'observations,  $S \in \mathbb{R}^{N \times K}$  est la matrice de *signatures* et  $A \in \mathbb{R}_+^{K \times T}$  est la matrice d'*activations*. Les dimensions du problème sont le nombre d'échantillons par période de tension  $N$ , le nombre de périodes  $T$  et le nombre de sous-composantes recherchées  $K$ .

Nous étendons (1) en introduisant la pénalisation en variation totale (TV) des lignes de la matrice d'activation ( $A$ ) :

$$\begin{aligned} & \underset{S,A}{\text{minimise}} \quad \frac{1}{2} \|X - SA\|_{Fro}^2 + \lambda \|\Delta A\|_1 & (2) \\ & \text{sachant que} \quad \forall k \in \llbracket 1, K \rrbracket, \|S_k\|_2^2 \leq 1 \quad \text{et} \quad A \geq 0 \end{aligned}$$

où  $S_k$  est la  $k^{eme}$  colonne de  $S$  et  $\Delta A \in \mathbb{R}^{K \times T-1}$  avec  $[\Delta A]_{k,t} = A_{k,t+1} - A_{k,t}$ , et où  $\lambda > 0$  est le paramètre de pénalisation. La contrainte sur les colonnes de  $S$  sert à résoudre l'ambiguïté multiplicative inhérente aux factorisations. En effet, sans cette contrainte, multiplier  $S$  par un scalaire et diviser  $A$  par la même valeur aurait pour conséquence de réduire artificiellement la pénalisation L1 sans changer le terme de fidélité aux données ni la forme des activations. On peut remarquer que même si la contrainte sur les colonnes de  $S$  est une inégalité, à l'optimum nous avons :  $\forall k \in \llbracket 1, K \rrbracket, \|S_k\|_2^2 = 1$ .

## 3 Estimation des paramètres

Ding et. al. [5] ont utilisé un algorithme d'optimisation alternée avec des mises à jour multiplicatives pour  $A$  mais n'ont pas abordé le problème de la TV-SNMF (2). Seichepine et. al [6] ont utilisé une pénalisation en variation totale mais uniquement dans le cadre de la factorisation non-négative de matrices, où les matrices  $S$  et  $A$  sont non-négatives. Ils ont utilisé un algorithme de majoration-minimisation pour mettre à jour

itérativement  $S$  et  $A$ . Nous proposons d'utiliser une stratégie d'optimisation alternée. La mise à jour de  $S$  conduit à un problème d'optimisation des moindres carrés sous contraintes quadratiques, alors que la mise à jour de  $A$  est un problème de moindres carrés avec un terme de variation totale ainsi qu'une contrainte de positivité.

### 3.1 Mise à jour de $S$

Le problème de mise à jour de  $S$  s'écrit :

$$\begin{aligned} & \underset{S}{\text{minimise}} \quad \frac{1}{2} \|X - SA\|_{Fro}^2 & (3) \\ & \text{sachant que} \quad \forall k \|S_k\|_2^2 \leq 1 \end{aligned}$$

Nous proposons de résoudre le problème dual équivalent [7], à la manière de ce qui est fait dans [8]. Le concept central de cette méthode est de calculer la formulation duale du problème ainsi que ses dérivées partielles par rapport aux variables duales dans le but de maximiser la fonction duale. Son principal avantage est que le problème dual est convexe et possède des contraintes plus simples. Nous utilisons une méthode d'optimisation de quasi newton (L-BFGS-B [9]) pour maximiser numériquement la fonction duale.

La formulation duale du problème s'écrit :

$$\begin{aligned} & \underset{\mu}{\text{maximise}} \quad \inf_S \mathcal{L}(S, \mu) \\ & \text{sachant que} \quad \mu \in \mathbb{R}_+^K \end{aligned}$$

où  $\mathcal{L}(S, \mu) = \frac{1}{2} \|X - SA\|_{Fro}^2 + \sum_k \mu_k (\|S_k\|_2^2 - 1)$  est le Lagrangien du problème et  $\mu$  sont les variables duales.

La fonction duale est donnée par :

$$\mathcal{D}(\mu) = \inf_S \mathcal{L}(S, \mu) = \mathcal{L}(S^*(\mu), \mu)$$

et  $\nabla_S \mathcal{L} = 0$  donne  $S^*(\mu) = XA^\top (AA^\top + 2M(\mu))^{-1}$

où  $M(\mu)$  est la matrice diagonale dont les éléments diagonaux sont  $\mu$ . Nous avons utilisé le fait que les dérivées partielles du Lagrangien par rapport à  $S$  sont égales à 0. Les dérivées par rapport aux variables duales sont données par :

$$\forall k, \quad \nabla_{\mu_k} \mathcal{D}(\mu) = S_k^{*\top}(\mu) S_k^*(\mu) - 1 \quad (4)$$

### 3.2 Mise à jour de $A$

L'optimisation de  $A$  en considérant  $S$  fixe est un problème de moindres carrés non négatifs avec pénalisation L1 :

$$\begin{aligned} & \underset{A}{\text{minimise}} \quad \frac{1}{2} \|X - SA\|_{Fro}^2 + \lambda \|\Delta A\|_1 \stackrel{\text{def}}{=} \mathcal{G}(A) & (5) \\ & \text{sachant que} \quad A \geq 0 \end{aligned}$$

La pénalisation TV est une fonction convexe mais non différentiable (en 0). Il existe plusieurs stratégies pour résoudre ce type de problème [10].

En particulier, on peut montrer qu'utiliser une approximation lisse de la norme L1 est efficace et rend le problème (5) convexe

et différentiable. Un candidat naturel pour cette approximation est la fonction :  $f(x) = \sqrt{x^2 + \epsilon} \approx |x|$ , où  $\epsilon > 0$  et petit. Comme le problème est maintenant convexe (sous condition que  $S^\top S$  soit semi définie positive) et différentiable, il peut être résolu à l'aide d'une méthode de quasi-Newton et du gradient :

$$\nabla_{A_{k,t}} \tilde{\mathcal{G}}(A) = B_{k,t} + \lambda(\mathbb{1}_{(t>1)} f'(\Delta A_{k,t-1}) - \mathbb{1}_{(t<T)} f'(\Delta A_{k,t}))$$

où  $B = -S^\top(X - SA) \in \mathbb{R}^{K \times T}$  et  $f'(x) = x/\sqrt{x^2 + \epsilon}$ .

### 3.3 L'Algorithme TV-SNMF

L'algorithme TV-SNMF consiste à itérer des mises à jour de  $S$  et de  $A$  de façon alternée jusqu'à convergence de la fonction de coût ou des paramètres. Il utilise l'implémentation de L-BFGS-B fourni par Scipy [11]. L'optimisation alternée garantit la réduction de la valeur de la fonction de coût à chaque itération mais comme le problème n'est pas convexe en  $S$  et  $A$  simultanément, la solution peut être un minimum local. On utilisera plusieurs initialisations des paramètres dans les expériences suivantes.

## 4 Résultats et discussion

Nous nous intéressons à l'application de techniques de factorisation matricielle au problème du NILM. Nous montrons d'abord comment transformer cet objectif en un problème de factorisation. Nous présenterons ensuite les résultats sur la base de données SHED [4].

Le problème du NILM consiste à retrouver les consommations individuelles d'appareils à partir d'une mesure de la consommation totale du bâtiment, en utilisant le fait que :

$$p_{\text{total}}(t) = \sum_k p_k(t)$$

où  $t$  est l'indice des périodes de tension,  $p_{\text{total}}$  est la puissance totale consommée par le bâtiment et  $\forall k \in \llbracket 1, K \rrbracket$ ,  $p_k$  sont les consommations individuelles d'appareils  $k$ . En considérant que l'on a accès aux mesures de périodes du courant ( $X$ ) et de tension ( $U$ ), la puissance totale s'écrit :

$$p_{\text{total}}(t) = \frac{1}{N} \sum_n X(n, t) U(n, t)$$

où  $n \in \llbracket 1, N \rrbracket$  correspond à l'indice des échantillons au sein d'une période  $t$ . Factoriser les mesures de courant implique que :  $X(n, t) \approx \sum_k S(n, k) A(k, t)$ ; et la puissance totale s'écrit donc :

$$\begin{aligned} p_{\text{total}}(t) &\approx \frac{1}{N} \sum_n \sum_k S(n, k) A(k, t) U(n, t) \\ &= \sum_k A(k, t) \underbrace{\frac{1}{N} \sum_n S(n, k) U(n, t)}_{\alpha(k, t)} = \sum_k \hat{p}_k(t) \end{aligned}$$

où  $\alpha(k, t)$  est un facteur de normalisation dépendant des signatures ( $S$ ) et de la tension ( $U$ ). On s'aperçoit qu'en utilisant

TABLE 1 – Comparaison des performances entre ICA, SNMF, TV-SNMF sur la base données SHED.

Bâtiment	ICA		SNMF		TV-SNMF		
	Moy	1 <sup>er</sup>	Moy	1 <sup>er</sup>	$\lambda^*$	Moy	1 <sup>er</sup>
1	1.00	0.96	0.61	0.44	0.30	0.58	0.17
2	1.07	0.98	0.58	0.44	0.25	0.55	0.28
3	0.99	0.98	0.63	0.43	0.35	0.80	0.57
4	1.07	1.00	0.57	0.42	0.35	0.70	0.37
5	1.03	0.99	0.58	0.43	0.30	0.42	0.27
6	1.03	0.96	0.58	0.48	0.35	0.61	0.39
7	0.99	0.96	0.59	0.38	0.25	0.47	0.27
8	1.02	0.99	0.59	0.43	0.20	0.55	0.36

Moy : moyenne des erreurs  $E_p$  sur 50 exécutions.

1<sup>er</sup> : erreur minimale  $E_p$  parmi les 50 exécutions.

$\lambda^*$  : valeur de  $\lambda$  obtenue par validation croisée.

une factorisation du courant et si  $\forall k, t \alpha(k, t) > 0$ , on peut écrire la puissance totale comme la somme de composantes non négatives ( $\hat{p}_k$ ). On observe, dans l'expérience suivante, que nous avons bien  $\alpha(k, t) > 0$ . Dans l'évaluation du NILM, ces composantes seront directement comparées aux consommations individuelles. Nous définissons l'erreur en puissance :

$$E_p = \sum_k \|p_k - \hat{p}_k\|_2^2 / \sum_k \|p_k\|_2^2 \quad (6)$$

où  $p_k \in \mathbb{R}_+^T$  est la vérité terrain et  $\hat{p}_k \in \mathbb{R}_+^T$  est l'estimation correspondante.

Dans cette expérience, nous appliquons nos algorithmes à une base de données nommé SHED [4]. SHED contient des simulations de consommations électriques pour 8 bâtiments commerciaux ainsi que les consommations individuelles des appareils présents. Il a été montré que ces simulations étaient réalistes pour le NILM [4]. Nous pouvons noter que le procédé de simulation inclut un modèle de factorisation plus complexe. Nous allons utiliser une partie de cette base, à savoir une journée de mesure pour chaque bâtiment, moyennée au pas de temps de 5 minutes. Le nombre original d'appareils (ou de catégories d'appareils) va de 5 à 10. Les dimensions du problème sont donc :  $N = 100$ ,  $K = \llbracket 5, 10 \rrbracket$  et  $T = 288$ .

Pour évaluer les algorithmes, la méthode de base que nous avons utilisée est l'analyse en composantes indépendantes (ICA). L'ICA est une technique de séparation de sources qui peut être vue comme une méthode de factorisation matricielle. Dans celle-ci, une matrice de signatures est recherchée, qui minimise la dépendance statistique entre ses composantes [12, 13].

L'expérience consiste à exécuter les algorithmes ICA, SNMF et TV-SNMF avec 50 initialisations différentes en fixant la valeur de  $K$  à 10. Pour TV-SNMF, le choix du paramètre  $\lambda$  suit un protocole de validation croisée *leave one out* : la valeur optimale de  $\lambda$  pour chaque bâtiment de l'échantillon d'apprentissage est obtenue par recherche dans un ensemble discret de valeurs possibles. La valeur de  $\lambda$  utilisée pour le bâtiment de

validation est pris égal à la moyenne des  $\lambda$  optimaux obtenus sur les autres bâtiments. Ceci est répété pour chaque bâtiment. Les résultats de l'erreur en puissance présentés dans le Tableau 1 montrent que TV-SNMF surpasse SNMF et ICA.

La Figure 1 présente 2 composantes estimées appartenant à la meilleure exécution de chaque algorithme pour le bâtiment 7 de la base de données SHED, ainsi que la vérité terrain correspondante. On remarque que les estimations de TV-SNMF sont plus lisses et plus proches de la vérité terrain. En ce qui concerne ICA, les valeurs peuvent être négatives, ce qui est en contradiction avec la contrainte de positivité des consommations. Enfin pour la composante 2, la reconstruction est imparfaite, ce qui peut s'expliquer par le fait que certains appareils, plus complexes, nécessitent plusieurs composantes pour être modélisés, voir [4].

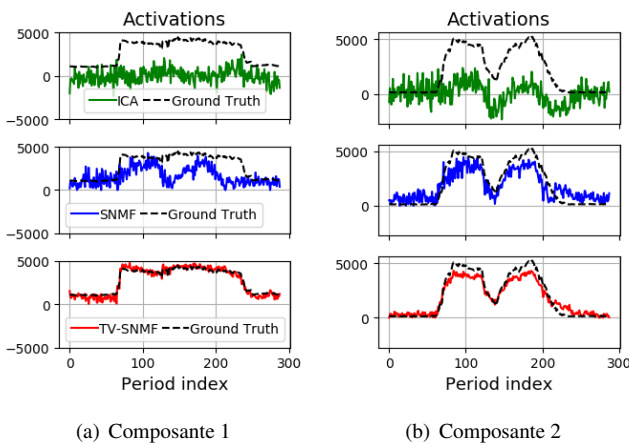


FIGURE 1 – Comparaison des activations estimées via ICA, SNMF et TV-SNMF (lignes colorées pleines) et de la vérité terrain (ligne pointillée noire) des consommations données par la base de données SHED, bâtiment 7. Les résultats présentés ici sont comparables à ceux des autres bâtiments de la base de données SHED.

## 5 Conclusion

Nous avons décrit un nouvel algorithme permettant de traiter le problème du NILM pour les bâtiments commerciaux. La méthode proposée associe la factorisation matricielle semi non négative et une pénalisation de la variation totale. Nous avons montré que notre approche est plus efficace pour l'analyse des consommations électriques de bâtiments commerciaux car elle estime mieux les facteurs non négatifs qui présentent une dépendance temporelle. Les résultats ont été obtenus sur une base de données réaliste de bâtiments simulés. L'évaluation de la performance de TV-SNMF sur données réelles est difficile du fait du manque de données disponibles combinant mesures haute fréquence et vérité terrain. Ceci, ainsi que l'étude sur des bâtiments résidentiels, est laissé pour des recherches futures.

## Références

- [1] C. Klemenjak and P. Goldsborough, "Non-intrusive load monitoring : A review and outlook," *arXiv preprint arXiv:1610.01191*, 2016.
- [2] H. Lange and M. Bergés, "Bolt : Energy disaggregation by online binary matrix factorization of current waveforms," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, pp. 11–20, ACM, 2016.
- [3] N. Batra, O. Parson, M. Berges, A. Singh, and A. Rogers, "A comparison of non-intrusive load monitoring methods for commercial and residential buildings," *arXiv preprint arXiv:1408.6595*, 2014.
- [4] S. Henriët, U. Simsekli, B. Fuentes, and G. Richard, "A generative model for non-intrusive load monitoring in commercial buildings," *Energy and Buildings*, vol. 177, pp. 268 – 278, 2018.
- [5] C. H. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 1, pp. 45–55, 2010.
- [6] N. Seichepine, S. Essid, C. Févotte, and O. Cappe, "Piecewise constant nonnegative matrix factorization," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6721–6725, IEEE, 2014.
- [7] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [8] M. Jas, T. D. La Tour, U. Simsekli, and A. Gramfort, "Learning the morphology of brain signals using alpha-stable convolutional sparse coding," in *Advances in Neural Information Processing Systems*, pp. 1099–1108, 2017.
- [9] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.
- [10] E. D. Dohmatob, A. Gramfort, B. Thirion, and G. Varoquaux, "Benchmarking solvers for tv-l1 least-squares and logistic regression in brain imaging," in *Pattern Recognition in Neuroimaging, 2014 International Workshop on*, pp. 1–4, IEEE, 2014.
- [11] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy : Open source scientific tools for Python," 2001.
- [12] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [13] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.