# Matrix Factorization for High Frequency Non Intrusive Load Monitoring: Definitions and Algorithms.

Simon Henriet
Télécom Paris, Institut Polytechnique de Paris, France
simon.henriet@telecom-paristech.fr

Benoit Fuentes
Télécom Paris, Institut Polytechnique de Paris, France
benoit.fuentes@smart-impulse.com

Umut Şimşekli
Télécom Paris, Institut Polytechnique de Paris, France
umut.simsekli@telecom-paristech.fr

Gaël Richard
Télécom Paris, Institut Polytechnique de Paris, France
gael.richard@telecom-paristech.fr

## ABSTRACT

Non Intrusive Load Monitoring has been introduced 30 years ago in order to monitor the electric consumption of specific equipments inside a building without the need of installing multiples sensors. During three decades, researchers and industrials have described the NILM problems according to the electric data available, the desired quantity to be monitored and the application it was used for. As a consequence of the multitude of choices, a lot of different formulations can be found in the literature. This diversity makes it difficult for researchers from general domains such as machine learning to tackle the NILM problem. In this paper we aim at defining the NILM problem as a Matrix Factorization task using high frequency measurements and also to review methods to solve this problem. We start by defining the general concepts driving the NILM problem and then show how to cast high frequency NILM into a Matrix Factorization problem. Once casted as a machine learning problem, we will review general purposes algorithms applicable to this problem such as Independent Component Analysis, Sparse Coding or Semi Non-negative Matrix Factorization and specific NILM methods such as BOLT and IVMF.

## CCS CONCEPTS

• **Computing methodologies** → **Factorization methods**; **Source separation**; • **Hardware** → *Energy metering*.

## KEYWORDS

NILM, Energy Disaggregation, High Frequency Data, Matrix Factorization, Source Separation.

**ACM Reference Format:**
Simon Henriet, Benoit Fuentes, Umut Şimşekli, and Gaël Richard. 2020. Matrix Factorization for High Frequency Non Intrusive Load Monitoring: Definitions and Algorithms.. In *The 5th International Workshop on Non-Intrusive Load Monitoring (NILM'20), November 18, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3427771.3427847


*NILM workshop,, November 18 2020, Online,*

ACM ISBN 978-1-4503-8191-8/20/11…$15.00
https://doi.org/10.1145/3427771.3427847

## 1 INTRODUCTION

There is one fundamental concept behind all the Non Intrusive Load Monitoring formulations: the conservation of energy. Established by Kirchhoff in 1845, the current law states that: the algebraic sum of currents in a network of conductors meeting at a point is zero: $\sum_k \mathbf{i}_k(\tau) = 0$, where $k$ is the index of conductors and $\tau$ is the time index. In other words the Kirchhoff's current law says that for any node or junction in an electrical circuit, the sum of currents flowing into that node is equal to the sum of currents flowing out of that node. In electrical circuits for modern buildings, all the devices are plugged in parallel.

The main breaker is a node where the Kirchhoff's current law applies. This results in the NILM equation on current:

$$\mathbf{i}_{main}(\tau) = \sum_{d \in \mathcal{D}} \mathbf{i}_d(\tau), \tag{1}$$

where $\mathbf{i}_{main}(\tau)$ is the current at the main breaker, $d$ is the index of an electric device and $\mathcal{D}$ is the set of all the device in the electric network.

Using Equation (1) and the fact that all the devices share the same voltage $\mathbf{u}_{main}(\tau)$, we can establish the NILM equation on power:

$$\mathbf{p}_{main}(\tau) = \sum_{d \in \mathcal{D}} \mathbf{p}_d(\tau). \tag{2}$$

The NILM problem on high frequency current data can then be defined as:

*Definition 1.1.* From the current and voltage measurements acquired at the breaker panel of a building and at a high frequency sampling rate (> 50Hz): $\mathbf{i}_{main}(\tau), \mathbf{u}_{main}(\tau)$; estimate, the real power consumptions of categories of equipments $\mathbf{P}_c(t)$, indexed by $c \in \mathcal{C}$, in the building and such that: $\mathbf{P}_{main}(t) = \sum_{c \in \mathcal{C}} \mathbf{P}_c(t)$.

We use category indexes to group all the devices with similar electric or electronic components (e.g. lights or computers).

## 2 PROBLEM FORMULATION

### 2.1 From the NILM Software Problem to Matrix Factorization

We can now show how to transform this single-channel source separation problem into a Matrix Factorization problem. Let us first explain the matrix representation. From a unidimensional time serie $\mathbf{i}(\tau) \in \mathbb{R}^{NT}$, where $N$ is the number of samples during one voltage period and $T$ is the number of voltage periods in the measurement, we cut slices of size $N$ (one voltage period) which are then set as the columns of a matrix. The beginning of a voltage is defined at the time where the voltage crosses zero from negative to positive value. Let us denote by $\mathbf{I}_{main} \in \mathbb{R}^{N\times T}$ this current matrix observation. Due to the pseudo sinusoidal shape of current it is often also referred as the current waveform matrix.

This transformation being only a reshaping of the current time serie, the current conservation equation (1) still holds. We denote by $\mathbf{I}_c$ the unobserved current matrix of a category of equipment indexed by $c$:

$$\mathbf{I}_{main}(n,t) = \sum_{c\in C} \mathbf{I}_c(n,t) \tag{3}$$

Using this matrix representation for the voltage $\mathbf{u}$, the power calculations are given by:

$$\mathbf{P}_i(t) = \frac{1}{N}\sum_n \mathbf{I}_i(n,t)\,\mathbf{U}_i(n,t), \quad \forall i \in C \cup \{main\} \tag{4}$$

In [6], the authors showed that current matrices ($\mathbf{I}_c$) of individual equipment or group of same equipment (such as lights or computers) can be accurately approximated by low rank matrices. In the following we make the assumption that a rank one matrix can approximate well the individual current matrices. We recall that a rank one matrix can be defined as the multiplication between a column vector and a row vector:

$$\mathbf{I}_c(n,t) \approx \mathbf{s}_c(n)\,\mathbf{a}_c^\top(t), \quad \forall c \in C, \tag{5}$$

where, $\mathbf{s}_c \in \mathbb{R}^N$ is called a *signature* and $\mathbf{a}_c \in \mathbb{R}_+^T$ is called an *activation*. See Figure 1 for an illustration.

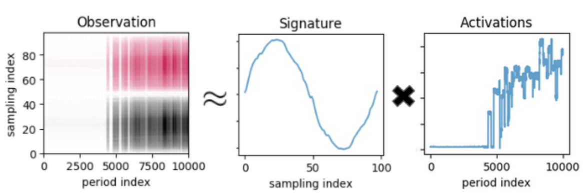


**Figure 1: Rank one approximation of current matrices.**

Merging Equations (3) with (5) results in a matrix factorization equation:

$$\mathbf{I}_{main} \approx \mathbf{S}\,\mathbf{A} \tag{6}$$

where $\mathbf{S} \in \mathbb{R}^{N\times C}$ is called the *signature* matrix and its columns contain the signatures $\mathbf{s}$ for each category. The other factor $\mathbf{A} \in \mathbb{R}^{C\times T}$ is called the *activation* matrix and its rows correspond to the activations $\mathbf{a}$ for each category.

## 2.2 Learning Strategy

The learning problem defined in this Matrix Factorization framework is then defined as follows:

*Definition 2.1.* From the total current $\mathbf{I}_{main}$ and voltage $\mathbf{U}_{main}$ measurements (in matrix shape), estimate $\hat{\mathbf{S}}$ and $\hat{\mathbf{A}}$, such that:

$$\mathbf{I}_{main} \approx \hat{\mathbf{S}}\hat{\mathbf{A}} \tag{7}$$

$$\hat{\mathbf{P}}_c(t) = \frac{1}{N}\sum_n \hat{\mathbf{S}}(n,c)\hat{\mathbf{A}}(c,t)\,\mathbf{U}(n,t) \tag{8}$$

$$\sum_c \mathcal{L}(\mathbf{P}_c(t)\|\hat{\mathbf{P}}_c(t)) \quad \text{is minimized.} \tag{9}$$

where $\{\mathbf{P}_c\}_{c\in C}$ are the true power consumptions per category (often referred as ground truth) and $\mathcal{L}$ is a divergence between the ground truth and the estimation.

A supervised learning strategy would involve an important amount of observation/ground truth couples: $(\mathbf{I}^b_{main}, \{\mathbf{P}^b_c\}_{c\in C})$ where $b$ is the index of a building. In such a setting one uses a training set of couples (observation, ground truth) to learn a mapping function from observation to the outputs by minimizing the loss function $\mathcal{L}$. In this case the factorization $\hat{\mathbf{S}}\hat{\mathbf{A}}$ is an intermediate quantity.

Due to the unavailability of such a training set in our particular application, a supervised learning approach is not possible. Oppositely, an unsupervised learning approach would try to infer $\hat{\mathbf{S}}$ and $\hat{\mathbf{A}}$ without having access to the ground truth ($\mathbf{P}_c$). Estimating the performance of an unsupervised approach can be accomplished in two ways. First, one can use a limited size testing set to quantify the estimation error using the same loss function as in the supervised case. Secondly, one can theoretically analyze the method and demonstrate conditions under which the method guarantees the recovery of the unknown sources. This includes the notion of identifiability developed in the following section.

In the rest of the paper we denote by $X$ the observation matrix instead of $\mathbf{I}_{main}$ for generality and simplicity purposes and by $k$ the index of a component in the factorization, which can be greater than the number of category $C$.

## 3 GENERAL PURPOSE MATRIX FACTORIZATION TECHNIQUES

Matrix Factorization refers to the wide ensemble of techniques that can decompose a real valued observation matrix $X \in \mathbb{R}^{N\times T}$ into the product of two matrices $\mathbf{S} \in \mathbb{R}^{N\times K}$ and $\mathbf{A} \in \mathbb{R}^{K\times T}$, called factors. Most of the time the decomposition is qualified as *approximated* since: $X \approx \mathbf{S}\,\mathbf{A}$; but in some cases the decomposition is qualified as *exact* and: $X = \mathbf{S}\,\mathbf{A}$.

The factors learning problem is traditionally cast into an optimization problem where a *fit* function of the observation $X$ and the factorization $\mathbf{S}\,\mathbf{A}$ is minimized. The fit function $\mathcal{D}$ is here a function of 2 variables, from $\mathbb{R}^{N\times K} \times \mathbb{R}^{N\times K}$ to $\mathbb{R}_+$. It equals 0 if and only if the 2 variables are equals. It can be a divergence or a distance. Classical examples are based on the euclidean distance or the Kullback-Leibler divergence for instance. Moreover, regularizer functions are usually added to enforce particular characteristics to the factors (sum of vector norms, matrix norms) or to reduce the number of solutions. Finally, the factors may be constrained to lie in a specific space such as the space of positive valued matrices, the orthogonal group or the unit ball defined by a norm. Then, the generic matrix factorization optimization takes the form of:

$$\hat{\mathbf{S}},\hat{\mathbf{A}} = \underset{\mathbf{S}\in E_\mathbf{S},\mathbf{A}\in E_\mathbf{A}}{\operatorname{argmin}} \; \mathcal{D}(X,\mathbf{S}\,\mathbf{A}) + \lambda_\mathbf{S}\mathcal{R}_\mathbf{S}(\mathbf{S}) + \lambda_\mathbf{A}\mathcal{R}_\mathbf{A}(\mathbf{A}) \tag{10}$$

where $\mathcal{D}$ is a fit function, $\mathcal{R}_\mathbf{S}$ and $\mathcal{R}_\mathbf{A}$ the regularizer functions, $\lambda_\mathbf{S}$ and $\lambda_\mathbf{A}$ are the regularizer parameters, $E_\mathbf{S}$ and $E_\mathbf{A}$ are the subspaces of $\mathbf{S}$ and $\mathbf{A}$.

Matrix factorization has a long and successful history for solving mathematical and signal processing problems (image, audio, neuroscience, recommender systems). Famous techniques such as Principal Component Analysis, Dictionary Learning [1, 13, 14], Non-negative Matrix Factorization [11], Semi Non-negative Matrix Factorization [3] or Independent Component Analysis [8, 9] lie into this framework.

## 3.1 Semi Non-negative Matrix Factorization

Semi Non-negative Matrix Factorization (SNMF) has been introduced in [3]. In SNMF, the observation matrix $X$ is approximated by the matrix product of two factors: a real valued factor $\mathbf{S} \in \mathbb{R}^{N\times K}$ and a nonnegative factor $\mathbf{A} \in \mathbb{R}_+^{K\times T}$. The divergence is chosen to be the squared Euclidean distance defined by the Frobenius matrix norm and there is no regularization function. SNMF can formally be written down as the following problem:

$$\begin{aligned} \underset{\mathbf{S},\mathbf{A}}{\text{minimize}} \quad & \frac{1}{2}\|X - \mathbf{S}\,\mathbf{A}\|_{Fro}^2 \\ \text{subject to} \quad & \mathbf{A} \geq 0. \end{aligned} \tag{11}$$

In [3], the authors use an alternating optimization to solve (11) as the problem is convex on $\mathbf{S}$ and $\mathbf{A}$ separately.

Such a structure in Matrix Factorization introduce indeterminacies, i.e. an infinite number of factorization respecting the definition (11) may exist. The two classic indeterminacies are the permutation and the scale ambiguities. Let $\hat{\mathbf{S}}$ and $\hat{\mathbf{A}}$ be a solution of (11). Let $P$ be a $K \times K$ permutation matrix (a matrix with exactly one 1 on every rows and columns), and $C$ a $K\times K$ diagonal matrix with nonnegative entries, then $\tilde{\mathbf{A}} = PC\hat{\mathbf{A}}$ and $\tilde{\mathbf{S}} = \hat{\mathbf{S}}C^{-1}P^{-1}$ are also solution of (11).

Another kind of more problematic indeterminacy exists. Let us consider an observation matrix $X$ and a solution ($\hat{\mathbf{S}} = [\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2]$, $\hat{\mathbf{A}} = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2]^\top$) of (11), one can find an infinite number of admissible solutions as:

$$\tilde{\mathbf{s}}_1 = \hat{\mathbf{s}}_1 + \alpha\hat{\mathbf{s}}_2, \quad \tilde{\mathbf{s}}_2 = \hat{\mathbf{s}}_2, \tag{12}$$

$$\tilde{\mathbf{a}}_1 = \hat{\mathbf{a}}_1, \quad \tilde{\mathbf{a}}_2 = \hat{\mathbf{a}}_2 - \alpha\hat{\mathbf{a}}_1, \tag{13}$$

$$0 \leq \alpha \leq \min_{t=1\ldots T} \frac{\hat{\mathbf{a}}_2(t)}{\hat{\mathbf{a}}_1(t)}. \tag{14}$$

We can verify that $\hat{\mathbf{S}}\hat{\mathbf{A}} = X = \tilde{\mathbf{S}}\tilde{\mathbf{A}}$. The condition on $\alpha$ ensures that $\tilde{\mathbf{a}}_2 \geq 0$. This kind of indeterminacy may cause confusion in estimation.

This is major problem of SNMF, indeed for a same observation, several *equivalent* (in terms of fit function) factorization coexist. We will see how this *confusion* problem can be resolved under further assumptions on the factors. One way to reduce the number of solution to Problem (11) is to use regularization function. The idea is to choose among the infinity of equivalent solutions the one that minimizes a certain quantity. The regularization defined by the $\ell_1$ norm of the activations is widely used.

## 3.2 Sparse Coding

Sparse Coding (SC) or Sparse Dictionary Learning is a method introduced by [14] in neuroscience. The principle is to learn basis vectors such that the observations have a sparse representation in such a basis. We recall that a vector or a matrix is said to be sparse if it has a limited number of non zero elements. The $\ell_1$ norm is widely acknowledge to induce sparsity when used as a regularizing function such as in the well studied Lasso regression [16].

The classic SC optimization formulation reads:

$$\underset{\mathbf{S},\mathbf{A}}{\text{minimize}} \quad \frac{1}{2}\|X - \mathbf{S}\,\mathbf{A}\|_{Fro}^2 + \lambda \sum_{k,t} |\,\mathbf{A}(k,t)| \tag{15}$$

$$\text{subject to} \quad \|\,\mathbf{s}_k\,\|_2^2 \leq 1, \qquad \forall k \in [\![1, K]\!], \tag{16}$$

The constraint on the columns of $\mathbf{S}$ is essential for the regularization to operate. Indeed, without this constraint, multiplying $\mathbf{S}$ by a scalar and dividing $\mathbf{A}$ by the same value would artificially decrease the penalization term without changing the *data fitting* term or the shape of the solution. [12] have proposed an efficient algorithm to solve Problem (15) using an alternating optimization strategy. They also proposed extensions to constraint $\mathbf{A}$ to be nonnegative like in SNMF.

## 3.3 Independent Component Analysis

[9] have introduced Independent Component Analysis (ICA). ICA can be viewed as a special case of matrix factorization. Its main particularity is that in contrary to previously presented techniques, ICA constraints the factorization to be exact: $X = \mathbf{S}\,\mathbf{A}$.

The fundamental principle of the ICA model is that the rows $\mathbf{a}_k$ of $\mathbf{A}$ represent realizations of statistically independent random variables called sources.

While ICA has a strong literature involving statistics and information theory, we concentrate here on its optimization formulation. ICA algorithms can be viewed either as maximizing the likelihood of a couple model/observation or maximizing an approximation of the independence of samples via entropy-like measures. [7] has developed the FastICA algorithm which reduces to an iterative procedure that finds extremal points (minimizers and maximizers) of a certain non-linear and non-convex function under orthogonality conditions:

$$\begin{aligned} \underset{\mathbf{S}}{\text{maximize}} \quad & \|\hat{\mathrm{E}}\{\mathrm{G}(\mathbf{S}^\top X)\} - \mathbb{E}\{\mathrm{G}(\nu)\}\|_2^2 \\ \text{subject to} \quad & X = \mathbf{S}\,\mathbf{A} \quad \text{and} \quad \mathbf{S}^\top\mathbf{S} = I. \end{aligned} \tag{17}$$

where $X$ has been centered and whitened, $I$ is the identity matrix, $\mathcal{G}$ is a non quadratic function, $\mathbb{E}$ is the expectation, $\nu$ is a multivariate Gaussian variable with identity covariance matrix and $\hat{\mathrm{E}}$ represents the mean (over the columns of a matrix).

We have previously seen that matrix factorization models suffers from a number of indeterminacies. An important result in [2] stipulates that, if the original sources are independent and the density of at most one source is Gaussian, then, expect from a scale and permutation indeterminacy, the model is identifiable. This fundamental theorem shows why ICA is able to recover original sources from a linear mixture.

# 4 LIMITATIONS OF EXISTING MATRIX FACTORIZATION METHODS

After having defined matrix factorization techniques let us investigates their limitations for solving the high frequency NILM problem (Def. 2.1).

*Semi Non-negative Matrix Factorization.* In practice, the activation rows estimated by SNMF exhibit high correlations which is a non expected property of individual equipment power in buildings. Another weakness is that the positivity constraints on $A$ is not sufficient to ensure positivity of the estimated power consumption. As introduced in Equation (8), the estimated power consumption given by the matrix factorization can be rewritten as $\hat{\mathbf{P}}_k(t) = \alpha_k(t)\hat{\mathbf{A}}(k,t)$ with $\alpha_k(t) = \frac{1}{N}\sum_n \hat{\mathbf{S}}(n,k)\,\mathbf{U}(n,t)$. One can see that constraining $\hat{\mathbf{A}}$ to be positive is not sufficient to ensure that $\hat{\mathbf{P}}_k(t) \geq 0$.

*Independent Component Analysis.* ICA has important advantages over other Matrix Factorization technique which mainly include its identifiability and the high rate of convergence of the developed algorithms. However, the independence hypothesis of the devices consumption is not reasonable since in big buildings many devices are more likely to consume energy during the opening hours than during the night for instance. A refined assumption is that the power variations (or sometimes called derivatives) of the devices are independent. It is then usual (see [4]) to apply ICA to a transformation of the data such that the independence assumption is fulfilled in this new domain. Another weakness of ICA in our problem resides in the positivity of estimated consumptions.

*Sparse Coding.* As we have already seen, Sparse Coding is based on $\ell_1$ norm regularization on the activation $A$ to promote sparsity. However, this sparsity hypothesis does not hold at all for our signals. It is obvious that many different equipment are ON at the same time in a big building. However, the power differences are more sparse, that is to say, only a few devices switched ON or change their consumption state at the same instant. As done for, ICA, one can try Sparse Coding on a transformation of the original data. Unfortunately, Sparse Coding will suffer from the same positivity problem as ICA and SNMF.

# 5 NILM SPECIFIC METHODS

## 5.1 BOLT

In [10], the authors designed a matrix factorization technique, called BOLT, to deal with high frequency current measurements ($\mathbf{I}$). This method is based on the same matrix representation as presented here. The main idea is to factorize the current matrix so that it can be expressed as the sum of sub-components representing *individual* current matrices. These sub-components are afterwards used to infer devices activity (which device is ON and when). Inspired by FHMM approaches, the authors chose to use 2-state sub-components (as it is done to model ON/OFF devices). This choice results in a binary constraint on the activation matrix $\mathbf{A}$, while the signature matrix $\mathbf{S}$ is left unconstrained. The optimization problem reads:

$$\underset{\mathbf{S},\mathbf{A}}{\text{minimize}} \quad \|X - \mathbf{S}\,\mathbf{A}\|_{Fro}^2 \tag{18}$$

$$\text{subject to} \quad \mathbf{A} \in \{0,1\}^{K\times T} \tag{19}$$

From an algorithmic point of view such a problem is said combinatorial and no polynomial time algorithm can solve it. Interestingly, the authors introduced an additional constraint on the factor $\mathbf{S}$ (called signature matrix) so that $S = f_\theta(X)$ where $f$ is defined as a neural network with parameters $\theta$. This constraint can be seen as a mean of casting the binary matrix factorization into a deep neural network framework and then efficiently optimize Problem (18).

This approach works well for residential buildings and especially with simple ON/OFF devices, but lacks generality for more complex cases such as commercial buildings with continuously varying consumptions.

## 5.2 IVMF

In [5] the authors extend SNMF, ICA and SC by introducing: (i) a specific regularization and a positivity constraint over the activation matrix; (ii) linear and quadratic constraints on the signature matrix. The method called IVMF, can be defined as an optimization problem:

$$\begin{aligned}
\underset{\mathbf{S},\mathbf{A}}{\text{minimize}} \quad & \frac{1}{2}\|\,\mathbf{I} - \mathbf{S}\,\mathbf{A}\,\|_{Fro}^2 + \lambda\mathcal{G}(\mathbf{A}) \\
\text{subject to} \quad & \|\,\mathbf{s}_k\,\|_2^2 \leq 1, \quad \forall k \in [\![1,K]\!], \\
& \mathbf{s}_k^\top\,\mathbf{u}_0 \geq \alpha_0, \quad \forall k \in [\![1,K]\!], \\
& \mathbf{A} \geq 0,
\end{aligned} \tag{20}$$

where $\mathcal{G}(\mathbf{A}) = \sum_{k,t} G(\mathbf{A}(k,t{+}1) - \mathbf{A}(k,t))$ and $G$ is a non-quadratic scalar function and $\lambda > 0$ is the regularization parameter. To induce sparsity on the variation, two choices for $G$ are proposed: $G_{abs}(x) = |x|$ or $G(x) = \sqrt{x^2+\epsilon} - \sqrt{\epsilon}$ where $\epsilon$ is a small positive constant.

Note first that if we take $\epsilon = 1$, the smooth absolute value is equivalent to the classic $\text{logcosh}(x) = \log(\cosh(x))$ function in ICA (both of them being equivalent to $\frac{x^2}{2}$ near 0). When $\epsilon$ decrease to 0 the limit of the *smooth absolute value* is the absolute value. A second remark is that $G_{abs}(x)$ correspond to the widely used total variation regularization [15].

The *quadratic constraint* on the columns $\mathbf{s}_k$ of $\mathbf{S}$ is used to fix the inherent scaling ambiguity in such factorization problems and for the regularization to operate (as explained for Sparse coding). The *linear constraint* on $\mathbf{s}_k$, on top of the positivity constraint on $\mathbf{A}$, ensures the positivity of the power estimation (with $\mathbf{u}_0$ being the voltage vector and $\alpha_0$ a fixed parameter).

IVMF showed in practice the ability of recovering time dependent sources whose variations exhibit independence or sparsity due to the specific regularizer used. This property enables IVMF to better estimate the consumption of continuously varying devices.

# 6 CONCLUSION AND DISCUSSION

In this paper we have recalled the fundamental concept defining Non Intrusive Load Monitoring. We have also showed how to transform this problem into a Matrix Factorization problem using high frequency data. Finally, we have reviewed the advantages and drawbacks of general purposes and NILM specific matrix factorization methods.

We have first seen that BOLT used the binary hypothesis for its activation in order to fit to the ON/OFF kind of device. We have also seen that IVMF can address the important problem of continuously varying devices present in large buildings such as commercial ones. However, as shown in [6], for certain devices, called *Varying signature*, the rank-one approximation of their current matrix does not hold. A relaxation of this hypothesis and the use of group regularization could lead to a new matrix factorization methods.

We can also remark that the presented methods treat the data in batch and achieving an online treatment of data could be an interesting feature for NILM algorithms.

## REFERENCES

[1] M Aharon, M Elad, and A Bruckstein. 2006. The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation. to appear in the IEEE Trans. On Signal Processing (2006).
[2] Pierre Comon. 1994. Independent component analysis, a new concept? Signal processing 36, 3 (1994), 287–314.
[3] Chris HQ Ding, Tao Li, and Michael I Jordan. 2010. Convex and semi-nonnegative matrix factorizations. IEEE transactions on pattern analysis and machine intelligence 32, 1 (2010), 45–55.
[4] Fangchen Feng and Matthieu Kowalski. 2018. Revisiting sparse ICA from a synthesis point of view: Blind Source Separation for over and underdetermined mixtures. Signal Processing 152 (2018), 165–177.
[5] Simon Henriet, Umut Şimşekli, Sergio Dos Santos, Benoît Fuentes, and Gael Richard. 2019. Independent-Variation Matrix Factorization With Application to Energy Disaggregation. IEEE Signal Processing Letters 26, 11 (2019), 1643–1647.
[6] Simon Henriet, Umut Simsekli, Benoit Fuentes, and Gaél Richard. 2018. A Generative Model For Non-Intrusive Load Monitoring in Commercial Buildings. Energy and Buildings 177 (2018), 268 – 278. https://doi.org/10.1016/j.enbuild.2018.07.060
[7] Aapo Hyvarinen. 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE transactions on Neural Networks 10, 3 (1999), 626–634.
[8] Aapo Hyvärinen and Erkki Oja. 2000. Independent component analysis: algorithms and applications. Neural networks 13, 4 (2000), 411–430.
[9] Christian Jutten and Jeanny Herault. 1991. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal processing 24, 1 (1991), 1–10.
[10] Henning Lange and Mario Bergés. 2016. 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. ACM, 11–20.
[11] Daniel D Lee and H Sebastian Seung. 2001. Algorithms for non-negative matrix factorization. In Advances in neural information processing systems. 556–562.
[12] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. 2007. Efficient sparse coding algorithms. In Advances in neural information processing systems. 801–808.
[13] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. 2010. Online learning for matrix factorization and sparse coding. Journal of Machine Learning Research 11, Jan (2010), 19–60.
[14] Bruno A Olshausen and David J Field. 1997. Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision research 37, 23 (1997), 3311–3325.
[15] Leonid I Rudin, Stanley Osher, and Emad Fatemi. 1992. Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena 60, 1-4 (1992), 259–268.
[16] Robert Tibshirani. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 1 (1996), 267–288.