Sparse optimization for automated energy end use disaggregation

Dario Piga[1], Andrea Cominola[2], Matteo Giuliani[2], Andrea Castelletti[2], and Andrea Emilio Rizzoli[3]
[1]IMT Institute for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy
[2]Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy
[3]Dalle Molle Institute for Artificial Intelligence Research, SUPSI, Galleria 1, 6928 Manno, Switzerland

Retrieving the household electricity consumption at individual appliance level is an essential requirement to assess the contribution of different end uses to the total household consumption, and thus to design energy saving policies and user-tailored feedback for reducing household electricity usage. This has led to the development of Nonintrusive Appliance Load Monitoring (NIALM), or energy disaggregation, algorithms, which aim to decompose the aggregate energy consumption data collected from a single measurement point into device-level consumption estimations. Existing NIALM algorithms are able to provide accurate estimate of the fraction of energy consumed by each appliance. Yet, to the authors' experience, they provide poor performance in reconstructing the power consumption trajectories over time. In this work, a new NIALM algorithm is presented, which, besides providing very accurate estimates of the aggregated consumption by appliance, also accurately characterises the appliance power consumption profiles over time. The proposed algorithm is based on the assumption that the unknown appliance power consumption profiles are piecewise constant over time (as it is typical for power use patterns of household appliances) and it exploits the information on the time-of-day probability in which a specific appliance might be used. The disaggregation problem is formulated as a least-square error minimization problem, with an additional (convex) penalty term aiming at enforcing the disaggregate signals to be piecewise constant over time. Testing on household electricity data available in the literature is reported.

***Index Terms*—Energy disaggregation, Sparse optimization.**

I. Introduction

Managing world energy demand is one of the most challenging issue that governments are facing today. Global energy consumption is expected to increase by nearly 35% by 2035 [1] and the consequent impacts on climate have already started to generate policy change on how energy is generated, stored, and managed [2].

Acting on the reduction of energy demand, more than on expanding the energy production capacity, becomes essential to secure reliable energy supply, while reducing utilities' costs and financial risks. For example, the residential sector accounts for almost 30% of electricity consumption in the EU [3], and the study reported in [4] shows that household energy consumption can be reduced by 10% to 15% through better energy demand management. In order to design efficient energy management programs, it is essential to monitor the household end use energy utilization and to provide personalized feedback to consumers, so that: (i) users are aware of how much energy each appliance is consuming, and personalized hints for reducing their energy consumption can be given; (ii) household's occupants can be informed on potential savings in deferring the use of some appliances to off-peak hours.

These challenges have motivated researchers to develop *Nonintrusive Appliance Load Monitoring* (NIALM) algorithms to decompose the aggregate power consumption, as provided by traditional smart meters, into its individual component appliances, without installing on-device monitoring equipment. The first algorithm for NIALM was proposed by Hart in [5], where the aggregate power signal is segmented into successive steps, which are matched to the appliance signatures (i.e., their typical power demand curves). However, Hart's approach is not able to detect multistate appliances and it is neither able to decompose power signals made of simultaneous on/off events on multiple appliances. Since Hart's contribution, the problem of NIALM has been extensively studied in the literature (for a review, see [6], [7], [8] and references therein). State-of-the-art NIALM algorithms can be classified into two main categories: machine learning and optimization based approaches. In the first category, we mention the methods based on *Hidden Markov Models* [9], [10], [11] and on *Artificial Neural Networks* [12], [13]; while the approaches based on integer programming optimization [14], [15] and sparse coding [16] belong to the second category. All of the aforementioned algorithms have generally shown good performance in estimating the fraction of energy consumed by each appliance, however most of them lack in skill in accurately reconstructing the power consumption trajectories over time. This represents a serious drawback, since: (i) no information on the time of use of each appliance can be derived, and so feedback on potential savings in differing the usage of some devices to peak-off hours cannot be provided; (ii) anomalous events, such as a device consuming an exceptional amount of power over an extended period, can be barely detected; (iii) it is not evident if the accuracy in the estimate of the fraction of energy consumed by each appliance is due to fortuitous balancing mechanisms.

In this paper, a novel NIALM algorithm based on sparse optimization is presented. The proposed approach exploits the assumption that the power demand profiles of each appliance are piecewise constant over time (as it is typical for energy use patterns of household appliances), and exploits the information on the time-of-day probability in which a specific appliance is likely to be used. The disaggregation problem is treated as a least-square error minimization problem, with an additional (convex) penalty term aiming at enforcing the disaggregated signals to be piecewise constant over time. Beside being able to handle situations where multiple appliances are operating simultaneously, the proposed algorithm is able to reconstruct

Corresponding author: D. Piga (email: dario.piga@imtlucca.it).

the consumption trajectories over time, thus overcoming the main drawback of the disaggregation methods available in the literature. The paper is organized as follows. In Section II, the NIALM problem is formally defined and the main assumptions behind the proposed algorithm are provided. The disaggregation algorithm is discussed in Section III, and suggestions for its practical implementation are given in Section IV. In Section V, the proposed algorithm is tested against the AMPds dataset [17], containing the energy consumption readings of a single house located in Canada, and its performance is contrasted with the one attained by a benchmark NIALM algorithm based on Hidden Markov Models. Concluding remarks are given in Section VI, together with potential directions for future works.

II. Problem description

Consider the situation where N different electric appliances $(L_1, \ldots, L_N)$ are available in a house and connected to the electric power line. Each appliance L_i has $C_i \in \mathbb{N}$ operating modes, and let $B_i^{(j)}$ be the power demand of the i-th appliance at the j-th operating mode (with $j = 1, \ldots, C_i$). The power demand $y_i(t)$ of the i-th appliance at time t is then given by:

$$y_i(t) = \begin{bmatrix} B_i^{(1)} & \cdots & B_i^{(C_i)} \end{bmatrix} \begin{bmatrix} \theta_i^{(1)}(t) \\ \vdots \\ \theta_i^{(C_i)}(t) \end{bmatrix} + e_i(t), \quad (1)$$

with $e_i(t)$ denoting an intrinsic modeling error. The variables $\theta_i^{(1)}(t), \ldots, \theta_i^{(C_i)}(t)$ can be either 0 or 1, and satisfy the constraint $\sum_{j=1}^{C_i} \theta_i^{(j)}(t) = 1$ (i.e., each appliance can only operate at a single mode at each time instant).

Let $y(t)$ be the household aggregate power reading, which is given by:

$$y(t) = \sum_{i=1}^{N} y_i(t) + e(t), \quad (2)$$

where $e(t)$ is a measurement noise. Given a sequence $\mathcal{D}_T = \{y(t)\}_{t=1}^{T}$ of observations of the aggregate power signal $y(t)$, the aim of energy disaggregation is to estimate the power demand $y_i(t)$ (with $i = 1, \ldots, N$ and $t = 1, \ldots, T$) of each appliance based on the aggregate power readings $\mathcal{D}_T$.

Remark 1: The energy disaggregation problem can be seen as a blind identification problem [18], which aims at estimating the behaviour of a system (and eventually the input signal profiles) only based on the output signal observations. In a blind identification setting, the observed output of the system is the aggregate power consumption $y(t)$, the (unmeasured) input signals are the end-use power consumption profiles $y_i(t)$ and the underlying system is a static system defined as: $y(t) = \sum_{i=1}^{N} y_i(t)$. ■

In this paper, a sparse optimization based NIALM algorithm is presented. The following conditions are assumed to hold:

C1: A training data set $\mathcal{D}_{T_t}$ is available. The training set consists of the observations of the power signatures of each appliance available in the house. An intrusive period is needed to construct $\mathcal{D}_{T_t}$. During this period, the patterns of electricity demand of each appliance are observed, and information on time-of-day probability characterizing the usage of each appliance can be also gathered.

C2: A roughly knowledge of the power demand of each appliance at each operating mode (i.e., the terms $B_i^{(j)}$ in (1)) is supposed to be available. For instance, the terms $B_i^{(j)}$ can be evaluated from the training set $\mathcal{D}_{T_t}$ through k-means clustering [19] or through a simple visual inspection.

C3: The energy consumption profiles of each appliance are piecewise constant over time.

III. Disaggregation algorithm

The main ideas behind the proposed NIALM algorithm are discussed in this section.

A. Standard Least Squares

In order to estimate the power demand $y_i(t)$ of each appliance based on the aggregate power consumption observations $\mathcal{D}_T$, the time-varying parameters $\theta_i^{(1)}(t), \ldots, \theta_i^{(C_i)}(t)$ (with $i = 1, \ldots, N$ and $t = 1, \ldots, T$) can, in principle, be computed solving the standard least-square problem:

$$\min_{\substack{\theta_i^{(1)}(t), \ldots, \theta_i^{(C_i)}(t) \\ i = 1, \ldots, N \\ t = 1, \ldots, T}} \sum_{t=1}^{T} \left(y(t) - \sum_{i=1}^{N} \hat{y}_i(t, \theta_i) \right)^2, \quad (3)$$

with

$$\hat{y}_i(t, \theta_i) = y_i(t) - e_i(t) = \begin{bmatrix} B_i^{(1)} & \cdots & B_i^{(C_i)} \end{bmatrix} \begin{bmatrix} \theta_i^{(1)}(t) \\ \vdots \\ \theta_i^{(C_i)}(t) \end{bmatrix}.$$

However, Problem (3) is overparameterized, since it involves more parameters than measurements. As a consequence, overfitting occurs, and thus no generalization property is guaranteed. One possible solution to overcome this problem is to introduce regularization terms in (3) to:

- enforce each appliance to operate at a single mode at each time instant;
- according to **C3**, enforce the energy consumption profiles $\hat{y}_i(t, \theta_i)$ to be piecewise constant signals over time.

B. Adding regularization

In order to exploit the information that: (i) the parameters $\theta_i^{(1)}(t), \theta_i^{(2)}(t), \ldots, \theta_i^{(C_i)}(t)$ can be either 0 or 1 and (ii) each appliance can only operate at a single mode at each time instant, the following regularized problem can be solved instead of (3):

$$\min_{\substack{\theta_i^{(1)}(t),\ldots,\theta_i^{(C_i)}(t)\\ i=1,\ldots,N\\ t=1,\ldots,T}} \sum_{t=1}^{T}\left(y(t)-\sum_{i=1}^{N}\hat{y}_i(t,\theta_i)\right)^2 + \tag{4a}$$

$$+\lambda_1\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\begin{bmatrix}\theta_i^{(1)}(t)\\ \vdots\\ \theta_i^{(C_i)}(t)\end{bmatrix}\right\|_0 \tag{4b}$$

$$\text{s.t.} \sum_{j=1}^{C_i}\theta_i^{(j)}(t)=1,\ \theta_i^{(j)}(t)\geq 0,\ i=1,\ldots,N;\ t=1,\ldots,T,$$

where $\|\cdot\|_0$ denotes the cardinality of a vector (i.e., number of its nonzero components). Note that, on one hand, the second term in the objective function of Problem (4) aims at enforcing sparsity in the vector $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$. On the other hand, the vector $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$ is guaranteed to have at least one element different than zero, because of the equality constraint appearing in Problem (4). The hyper-parameter $\lambda_1 > 0$ is tuned by the user (for instance, through cross-validation) for balancing the trade-off between minimizing the fitting error (by decreasing the value of λ_1) and maximising sparsity of the parameter vector $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$ (by increasing the value of λ_1).

Note that, because of the $\|\cdot\|_0$ operator, Problem (4) is not convex. According to the Lasso [20], [21], an approximate solution of Problem (4) can be obtained by replacing the cardinality of a vector (i.e., the operator $\|\cdot\|_0$) with its ℓ_1 norm. Furthermore, in order to improve the accuracy of the final estimate, the parameters $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$ can be scaled by nonnegative weights $\begin{bmatrix}w_i^{(1)}(t) & \cdots & w_i^{(C_i)}(t)\end{bmatrix}$. This leads to the following convex approximation of Problem (4):

$$\min_{\substack{\theta_i^{(1)}(t),\ldots,\theta_i^{(C_i)}(t)\\ i=1,\ldots,N\\ t=1,\ldots,T}} \sum_{t=1}^{T}\left(y(t)-\sum_{i=1}^{N}\hat{y}_i(t,\theta_i)\right)^2 + \tag{5a}$$

$$+\lambda_1\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\begin{bmatrix}w_i^{(1)}(t)\\ \vdots\\ w_i^{(C_i)}(t)\end{bmatrix}\odot\begin{bmatrix}\theta_i^{(1)}(t)\\ \vdots\\ \theta_i^{(C_i)}(t)\end{bmatrix}\right\|_1 \tag{5b}$$

$$\text{s.t.} \sum_{j=1}^{C_i}\theta_i^{(j)}(t)=1,\ \theta_i^{(j)}(t)\geq 0,\ i=1,\ldots,N;\ t=1,\ldots,T,$$

where $\odot$ denotes the element-wise multiplication. The choice of the weights $w_i^{(j)}(t)$ is discussed in Section IV-A1. Note that the ℓ_1-norm regulation promotes sparsity of the vector $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$. In fact, in the ideal case, only one component of the vector $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$ should be nonzero (i.e., the i-th appliance operates at a single mode at each time instant). The reader is referred to the works [22], [23], [24], [25] for a detailed analysis on the properties of ℓ_1-regularization in sparse estimation problems.

C. Adding regularization to enforce piece-wise constant signals power demand profiles

In order to further improve the accuracy of the estimate given by (5), we might exploit the additional information that the power demand signatures of the electric appliances are piecewise constant over time (Assumption **C3**). In order to enforce the power signals to be piecewise constant, a new regularization term aiming at penalizing the variation of the time-varying coefficients $\theta_i(t)$ is added to Problem (5), i.e.,

$$\min_{\substack{\theta_i^{(1)}(t),\ldots,\theta_i^{(C_i)}(t)\\ i=1,\ldots,N\\ t=1,\ldots,T}} \sum_{t=1}^{T}\left(y(t)-\sum_{i=1}^{N}\hat{y}_i(t,\theta_i)\right)^2 + \tag{6a}$$

$$+\lambda_1\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\begin{bmatrix}w_i^{(1)}(t)\\ \vdots\\ w_i^{(C_i)}(t)\end{bmatrix}\odot\begin{bmatrix}\theta_i^{(1)}(t)\\ \vdots\\ \theta_i^{(C_i)}(t)\end{bmatrix}\right\|_1 + \tag{6b}$$

$$+\lambda_2\sum_{i=1}^{N}\sum_{t=2}^{T}\left\|k_i\begin{bmatrix}\theta_i^{(1)}(t)-\theta_i^{(1)}(t-1)\\ \vdots\\ \theta_i^{(C_i)}(t)-\theta_i^{(C_i)}(t-1)\end{bmatrix}\right\|_\infty \tag{6c}$$

$$\text{s.t.} \sum_{j=1}^{C_i}\theta_i^{(j)}(t)=1,\ \theta_i^{(j)}(t)\geq 0,\ i=1,\ldots,N;\ t=1,\ldots,T,$$

with λ_2 being a tuning parameter playing a role similar to λ_1. The terms k_i (with $i=1,\ldots,N$) are a-priori specified nonnegative weights which can be chosen through the method described in Section IV-A2. It is worth remarking that:

- penalizing the norm of the difference between two consecutive parameters $\theta_i^{(j)}(t)$ and $\theta_i^{(j)}(t-1)$ is commonly referred in the literature as *Fused Lasso* [26] and it is used to promote sparsity in the discrete-time derivative of the signal $\theta_i^{(j)}(t)$ (thus enforcing the signal $\theta_i^{(j)}(t)$ to be piecewise constant over time).
- the term (6c) is a *group (fused) Lasso* penalty [27], [28], [29], penalizing the mixed $\ell_{1,\infty}$-norm (i.e., sum of the infinity norms) of the groups

  $$\begin{bmatrix}\theta_i^{(1)}(t)-\theta_i^{(1)}(t-1)\\ \vdots\\ \theta_i^{(C_i)}(t)-\theta_i^{(C_i)}(t-1)\end{bmatrix},$$

  with $i=1,\ldots,N$ and $t=2,\ldots,T$. The infinity norm is considered in (6c) so that, at the solution, the vector

  $$\begin{bmatrix}\theta_i^{(1)}(t)-\theta_i^{(1)}(t-1)\\ \vdots\\ \theta_i^{(C_i)}(t)-\theta_i^{(C_i)}(t-1)\end{bmatrix},$$

  is enforced to be either identically zero or full. In fact, if one of the parameters $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$ changes from time $t-1$ to t, a variation of the other parameters does not change the cost function. Specifically, only the largest time variation among the elements of the vector $\begin{bmatrix}\theta_i^{(1)}(t) & \cdots & \theta_i^{(C_i)}(t)\end{bmatrix}$ affects the objective function. Following the same rationale, the ℓ_2-norm can be alternatively used instead of the ℓ_∞-norm. The choice of the norm of the group is a problem at hand, mainly related to the numerical algorithms used to solve the formulated group Lasso problem.

D. Exploiting the information that every appliance cannot change state simultaneously

If the sampling interval $\Delta_t = t - (t-1)$ is small enough, it is also reasonable assuming that at most one appliance can change operating mode at each time instant. If this assumption holds, this additional information can be exploited by adding the following convex constraints to Problem (6):

$$\sum_{i=1}^{N} \left\| \begin{bmatrix} \theta_i^{(1)}(t) - \theta_i^{(1)}(t-1) \\ \vdots \\ \theta_i^{(C_i)}(t) - \theta_i^{(C_i)}(t-1) \end{bmatrix} \right\|_\infty \leq 1, \quad t = 2, \ldots, T \quad (7)$$

IV. Practical implementation

Some suggestions for a practical implementation of the proposed disaggregation algorithm, including the choice of the weighting parameters $w_i^{(j)}(t)$ and k_i from the training set $\mathcal{D}_{T_t}$, are given in this section.

A. Weight setting

1) On the choice of the weights $w_i^{(j)}(t)$

The main idea behind the choice of the weights $w_i^{(1)}(t), \ldots, w_i^{(C_i)}(t)$ is the following: if the i-appliance is more likely to operate at mode C_j at time t, then the parameter $\theta_i^{(j)}(t)$ is more likely to be equal to 1, while the other parameters $\theta_i^{(g)}(t)$ (with $g \neq j$) are more likely to be equal to zero. In terms of the optimization problem (6), the parameters $\theta_i^{(g)}(t)$ (with $g \neq j$) should be more penalized than $\theta_i^{(j)}(t)$, or equivalently, $w_i^{(g)}(t)$ (with $g \neq j$) should be higher than $w_i^{(j)}(t)$. The information on time-of-day probability of the usage of each appliance can be inferred from the training data set $\mathcal{D}_{T_t}$. Specifically, for given i and t, the weights $w_i^{(1)}(t), \ldots, w_i^{(C_i)}(t)$ can be chosen as follows:

1) given the training dataset $\mathcal{D}_{T_t}$, for each time sample t compute $q_i^j(t)$ as the number of times the i-th appliance is classified to be at mode c_j at the time samples $t + k24\text{h}$, with $k \in \mathbb{Z}$.
2) if $q_i^j(t) \neq 0$, the weight $w_i^j(t)$ is then given by: $w_i^j(t) = \frac{1}{q_i^j(t)}$. Otherwise set the weight $w_i^j(t)$ to a large number.

Note that the parameter $q_i^j(t)$ might also be computed considering not only the observations t, $t - 24\text{h}$, $t + 24\text{h}$, $t - 48\text{h}$, $t + 48\text{h}$, ..., but also the observations (possibly weighted) within given time intervals $[t + k24\text{h} - \Delta, \quad t + k24\text{h} + \Delta]$.

2) On the choice of the weights k_i

The weights k_i can be chosen as follows: if the i-th appliance rarely changes its operating mode over time, than the time variation of the parameters $\theta_i(t)$ should be more penalized w.r.t. the time variation of the parameters characterizing an other appliance which frequently changes its operating mode. The weight k_i can be then inversely proportional to the number of mode changes observed in the training dataset for the i-th appliance, and scaled by the length of the training dataset.

B. Reducing the computational complexity

As the number of optimization variables in Problem (6) grows linearly with the length T of the signal $y(t)$ to be disaggregated, the applicability of the proposed approach is limited to small/medium values of T. In order to overcome this problem, a sub-optimal solution of Problem (6) can be computed by splitting the dataset $\mathcal{D}_T$ into M disjoint subsets $\mathcal{D}^{(h)}$ of length T_h (with $h = 1, \ldots, \mathrm{M}$) such that $\mathcal{D}_T = \bigcup_{h=1}^{\mathrm{M}} \mathcal{D}^{(h)}$. Problem (6) is then solved for each subset $\mathcal{D}^{(h)}$.

The computational complexity of the algorithm can be further reduced as follows. If at time t the i-th appliance is guaranteed not to operate at the j-th mode, then the parameter $\theta_i^{(j)}(t)$ can be set to zero, thus reducing the number of decision variables for Problem (6). Such an information can be simply obtained by analyzing the observed aggregate power consumption $y(t)$. In fact, if $y(t) \ll B_i^{(j)}$ (i.e., the observed aggregate power consumption at time t is largely lower than the power consumption of the i-th appliance when operating at mode j), then $\theta_i^{(j)}(t)$ can be directly set to zero.

V. Application on real data

The proposed disaggregation algorithm is tested against the AMPds dataset [17], which contains the energy consumption readings of a single house located in the Vancouver region in British Columbia, Canada. Specifically, the AMPds dataset contains the power consumption profiles of 19 appliances monitored for an entire year (from April 1, 2012 to March 31, 2013) at one minute read intervals.

A. Preprocessing phase

For the sake of analysis, we consider only the aggregate power consumption given by the sum of the power consumption readings of the following four electric appliances: clothes dryer; fridge; dishwasher; heat pump. The contribution of the selected appliances is about 45% of the total energy consumption. Furthermore, in order to assess the robustness of the disaggregation algorithm w.r.t. the measurement noise, a fictitious white noise $e(t)$ with Gaussian distribution $\mathcal{N}(0, \sigma_e^2)$ and standard deviation $\sigma_e = 4$ W is added to the aggregate power consumption signal $y(t)$.

The AMPds dataset is divided as follows:

- a training set, which consists of the power readings from May 17, 2012 to May 29, 2012. The training set is used to estimate the power demand of each appliance at each operating mode (i.e., the terms $B_i^{(j)}$ in eq. (1)). Therefore, the sub-metered power consumption trajectories of each appliance are supposed to be available in the training phase. Specifically, the set of power demands of each appliance at each operating mode are chosen through a simple visual inspection of the sub-metered power consumptions in the training dataset. The chosen values of $B_i^{(j)}$ are:
  1) clothes dryer: $[0 \;\; 260 \;\; 4700]$ W,
  2) fridge: $[0 \;\; 128 \;\; 200]$ W,
  3) dish washer: $[0 \;\; 120 \;\; 800]$ W,
  4) heat pump: $[39 \;\; 1900]$ W.

  The training set is also used to estimate the weights $w_i(t)$ and k_i in eq. (6) through the procedure discussed in Section IV-A. The obtained values of the (time-invariant) weights k_i associated to each appliance are:

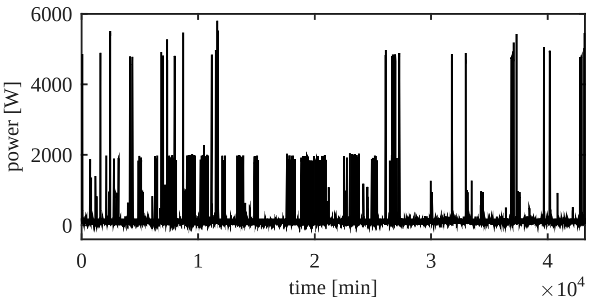

Fig. 1. Validation dataset $\mathcal{D}_T$: aggregate electric power consumption from June 1, 2012 to June 30, 2012

1) clothes dryer: $k_1 = 273$,
2) fridge: $k_2 = 11$,
3) dish washer: $k_3 = 165$,
4) heat pump: $k_4 = 444$.

- a calibration dataset, which consists of the measurements from May 30, 2012 to May 31, 2012. The calibration dataset is used to tune the hyper-parameters λ_1 and λ_2 in (6). Also in the calibration phase, the sub-metered power consumptions $y_i(t)$ are supposed to be available. The values of λ_1 and λ_2 are chosen through a cross-validation procedure, that is by minimizing (with a grid search) the *Total Relative Square Error* (TRSE) w.r.t. the calibration dataset, where the TRSE is defined as:
$$\text{TRSE} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T_c} \left(y_i(t) - \hat{y}_i(t)\right)^2}{\sum_{t=1}^{T_c} y_i^2(t)},$$
with T_c being the length of the calibration dataset. The chosen values of λ_1 and λ_2 are 10 and 750, respectively.
- a validation dataset $\mathcal{D}_T$, which consists of the data for the days 1-30 June 2012 (plotted in Fig. 1). The proposed algorithm in applied to disaggregate the data of the set $\mathcal{D}_T$.

In order to reduce the computational burden, a sub-optimal solution of Problem (6) is computed according to Section IV-B, by splitting the set of data to be disaggregated into 4320 subsets, each of equal length (i.e., 10 minutes). Finally, the aggregate power consumption observations are used to further reduce the computational complexity of Problem (6), by a-priori setting some parameters $\theta_i^{(j)}(t)$ to 0. Specifically:

- at the time instants when the aggregate power consumption is less than 3000 W, the parameters $\theta_i^{(j)}(t)$ associated to the clothes dryer and that multiply the basis $B_i^{(j)} = 4700$ W are set to 0;
- at the time instants when the aggregate power consumption is less than 1000 W, the parameters $\theta_i^{(j)}(t)$ associated to the heat pump and that multiply the basis $B_i^{(j)} = 1900$ W are set to 0;
- at the time instants when the aggregate power consumption is less than 400 W, the parameters $\theta_i^{(j)}(t)$ associated to the dish washer and that multiply the basis $B_i^{(j)} = 800$ W are set to 0.

B. Benchmark comparison: Factorial Hidden Markov Models

The performance of the optimization-based algorithm presented in this paper is compared to the performance of the disaggregation approach based on *Factorial Hidden Markov Model* (FHMM) and implemented in the open source *Non-intrusive Load Monitoring Toolkit* (NILMTK) [30]. In the FHMM-based approach, the power demand of each appliance is modelled as the observed value of a *Hidden Markov Model* (HMM). The hidden component of these HMMs are the states (i.e., the operating modes) of the appliances. The state of the whole system is modelled through a Factorial Hidden Markov Model and is given by the most probable combination of the states of each appliance, and the observed output of the whole system is the aggregate power consumption. For a fair comparison with the optimization-based algorithm presented in the paper, the FHMM algorithm is trained based on the data from May 17, 2012 to May 31, 2012 and used to disaggregate the data belonging to the validation dataset $\mathcal{D}_T$. The number of states of each HMM, or equivalently, the number of operating modes for each appliance, is the same across all appliances and it is set equal to 2.

C. Performance metrics

The following metrics are used to assess the performance of the optimization-based algorithm discussed in the paper and the performance of the FHMM-based approach:

- The *Estimated Energy Fraction Index* (EEFI), defined as:
$$\hat{h}_i = \frac{\sum_{t=1}^{T} \hat{y}_i(t)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{y}_i(t)}.$$
The index $\hat{h}_i$ provides the fraction of energy assigned to the i-th appliance, and it should be compared to the *Actual Energy Fraction Index* (AEFI), defined as
$$h_i = \frac{\sum_{t=1}^{T} y_i(t)}{\sum_{i=1}^{N} \sum_{t=1}^{T} y_i(t)},$$
which in turn provides the actual fraction of energy consumed by the i-th appliance. The EEFI $\hat{h}_i$ gives the users the information on how much energy each appliance is consuming, and so personalized hints for reducing their energy consumption can be provided.
- The *Relative Square Error* (RSE), defined as:
$$\text{RSE}_i = \frac{\sum_{t=1}^{T} \left(y_i(t) - \hat{y}_i(t)\right)^2}{\sum_{t=1}^{T} y_i^2(t)}.$$
The RSE provides a normalized measure of the difference between the actual and the estimated power consumption of the i-th appliance.
- The R^2 coefficient, defined for the i-th appliance as:
$$R_i^2 = 1 - \frac{\sum_{t=1}^{T} \left(y_i(t) - \hat{y}_i(t)\right)^2}{\sum_{t=1}^{T} \left(y_i(t) - \bar{y}_i\right)^2},$$
with $\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_i(t)$. Both the R^2 coefficient and the RSE measure how well the estimated power profiles match the actual power profiles over time. An accurate estimate of the power consumption profiles over time is essential to inform the customer on potential savings in deferring the use of some appliances to peak-off hours. Obviously, high value of the R^2 coefficients (or equivalently low values of the RSE) imply an accurate estimate of the EEFI $\hat{h}_i$.

TABLE I
FRACTION OF ENERGY ASSIGNED TO EACH APPLIANCE ($\hat{h}_i$) AND ACTUAL FRACTION OF ENERGY CONSUMED BY EACH APPLIANCE (h_i). RESULTS OBTAINED BY USING THE OPTIMIZATION-BASED ALGORITHM PRESENTED IN THE PAPER AND THE FHMM-BASED APPROACH.

	optimization-based algorithm	FHMM-based algorithm	ground truth
	$\hat{h}_i$	$\hat{h}_i$	h_i
Clothes dryer	30.7 %	31.6 %	31.3 %
Fridge	22.0 %	22.7 %	21.3 %
Dishwasher	4.0 %	7.7 %	5.1 %
Heat Pump	43.3 %	37.9 %	42.3 %

TABLE II
RELATIVE SQUARE ERRORS AND R^2 COEFFICIENTS. RESULTS OBTAINED BY USING THE OPTIMIZATION-BASED ALGORITHM PRESENTED IN THE PAPER AND THE FHMM-BASED APPROACH.

	optimization-based algorithm		FHMM-based algorithm	
	$RSE_{\rm i}$	R_i^2	$RSE_{\rm i}$	R_i^2
Clothes dryer	0.8 %	99.2 %	0.3 %	99.7 %
Fridge	24.2 %	63.3 %	20.6 %	68.7 %
Dishwasher	28.2 %	71.4 %	161.6 %	-63.9 %
Heat Pump	2.7 %	97.1 %	31.9 %	65.1 %

D. Numerical results

The performance metrics introduced in the previous section and the estimated disaggregate power profiles are computed in order to assess the performance of the two disaggregation algorithms. The obtained results are reported in Table I, Table II and in Fig. 2. It is worth remarking that the RSE and the R^2 coefficients, as well as the indexes $\hat{h}_i$ and h_i, are referred to the portion of the dataset to be disaggregated (i.e., the whole month of June), while, for the sake of visualization, only a portion of the disaggregated power profiles is plotted in Fig. 2. The obtained results show that the developed optimization-based algorithm is able to accurately estimate the fraction of energy consumed by each appliance in the household (see Table I). As a matter of fact, the EEFI $\hat{h}_i$ is very close to the AEFI h_i for each appliance. This good performance is mainly due to an accurate estimate of the disaggregated consumption trajectories over time (as shown by Table II and Figs. 2). The obtained results also reveal that:

- both the optimization-based and the FHMM-based algorithms provide an accurate estimate of the power consumption of the clothes dryer. This is mainly due to the fact that clothes dryer events can be better distinguished from the other end-use events, as they usually show the highest power consumption peaks and large durations.
- the FHMM-based algorithm slightly outperforms the optimization-based approach in the estimate of the fridge power consumption. This is mainly due to the fact that the power consumption profile of the fridge has a marked pattern, with periodic ON/OFF cycles, which is accurately captured by probabilistic models like Markov models;
- the optimization-based approach provides better performance than the FHMM-based algorithm in the estimate of the power consumptions of dishwasher and heat pump. In fact, the FHMM-based method tends to underestimate the consumption of the heat pump (see Fig. 2, bottom panels) and thus to erroneously assign the residual power to the dishwasher.

VI. CONCLUSION

In this paper, a novel algorithm for Nonintrusive Appliance Load Monitoring is presented. The disaggregation problem is treated as a least-square error minimization problem, with an additional penalty term aiming at enforcing the disaggregate power consumption signals to be piece-wise constant over time. The proposed method is able to handle situations where multiple appliances are operating simultaneously, and also to accurately estimate the appliance power consumption profiles over time. Ongoing research activities are focused on:

- extensive testing of the algorithm's generalization potential across different data sampling resolutions (i.e., 1s, 15 min, 1 h) and w.r.t. new, unseen, and continuously varying appliances;
- application to high-resolution water consumption data.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 619172 (SmartH2O: an ICT Platform to leverage on Social Computing for the efficient management of Water Consumption).

The authors would like to thank the contribution of Andrea Vescovi to parts of this study.

REFERENCES

[1] "U. S. Energy Information Administration (EIA), *Annual Energy Outlook 2012*," June 2012.
[2] J. O. Ebinger, *Climate impacts on energy systems: key issues for energy sector adaptation*. World Bank Publications, 2011.
[3] L. D. Shorrock and J. Utley, *Domestic energy fact file 2008*. BRE Bookshop, 2008.
[4] S. Darby, "The effectiveness of feedback on energy consumption: A review for DEFRA of the literature on metering, billing and direct displays," *Environmental Change Institue, Oxford, UK*, 2006.
[5] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
[6] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 1, pp. 76–84, 2011.
[7] A. Zoha, A. Gluhak, M. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.
[8] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli, "Benefits and challenges of using smart meters for advancing residential water demand modeling and management: a review," *Environmental Modelling & Software*, vol. 72, pp. 198–214, 2015.
[9] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "An unsupervised training method for non-intrusive appliance load monitoring," *Artificial Intelligence*, vol. 217, pp. 1–19, 2014.
[10] M. Johnson and A. Willsky, "Bayesian nonparametric hidden semi-markov models," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 673–701, 2013.
[11] J. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Int. Conference on Artificial Intelligence and Statistics*, 2012, pp. 1472–1482.
[12] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. O'Hare, "Real-time recognition and profiling of appliances through a single electricity sensor," in *IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks (SECON)*, 2010, pp. 1–9.
[13] D. Srinivasan, W. Ng, and A. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 398–405, 2006.
[14] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *SICE Annual Conference, 2008*, 2008, pp. 2742–2747.
[15] T. R. Camier, S. Giroux, B. Bouchard, and A. Bouzouane, "Designing a NIALM in smart homes for cognitive assistance," *Procedia Computer Science*, vol. 19, pp. 524–532, 2013.

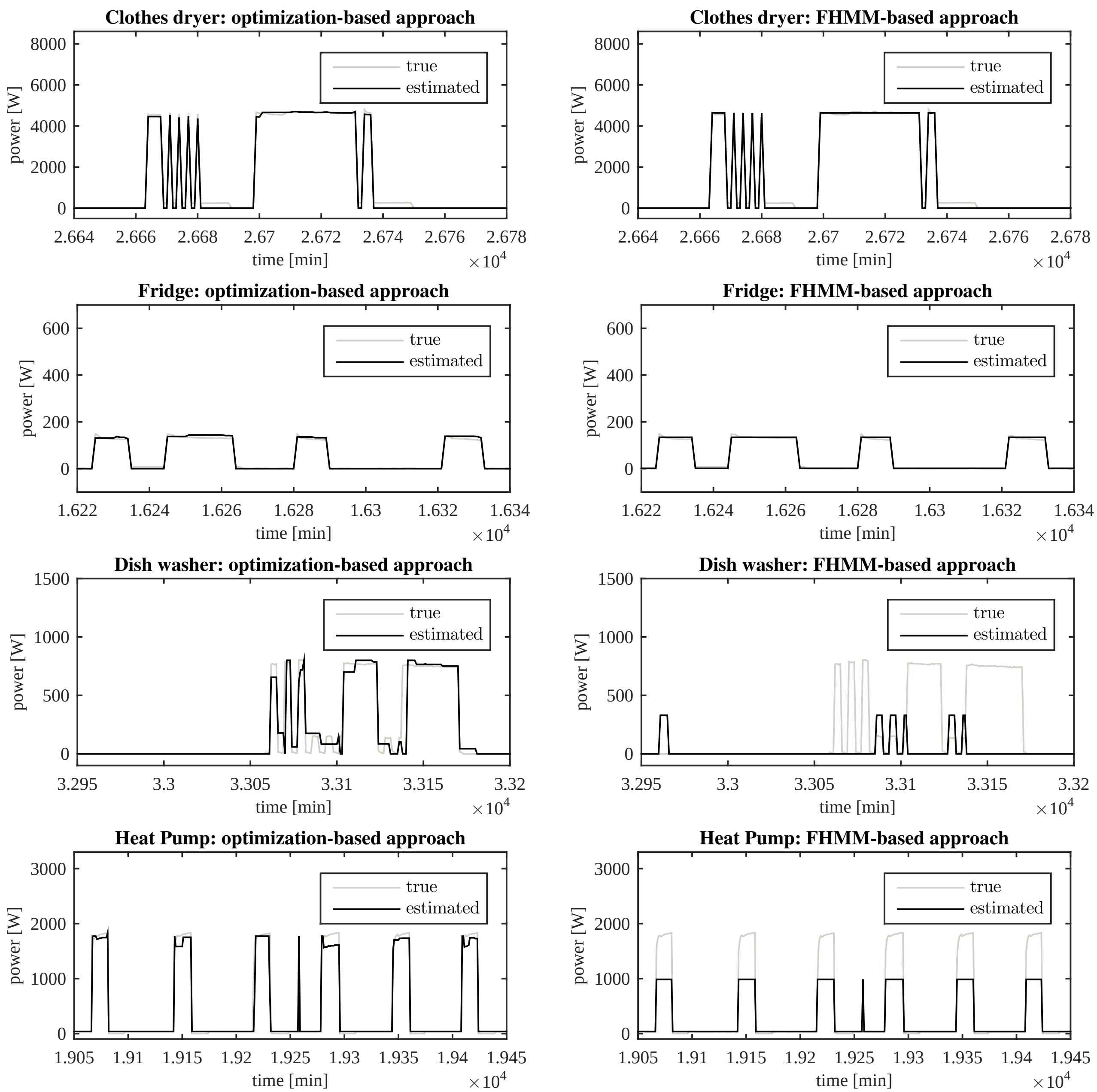

Fig. 2. Disaggregate power consumption profiles. Results obtained through the optimization-based approach presented in the paper (left panels) and through the FHMM-based approach (right panels).

[16] M. Figueiredo, B. Ribeiro, and A. de Almeida, "On the regularization parameter selection for sparse code learning in electrical source separation," in *Adaptive and Natural Computing Algorithms*. Springer, 2013, pp. 277–286.

[17] S. Makonin, F. Popowich, L. Bartram, B. Gill, and I. Bajic, "AMPds: a public dataset for load disaggregation and eco-feedback research," in *Electrical Power and Energy Conference (EPEC)*, 2013, pp. 1–6.

[18] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1310–1322, 1997.

[19] A. Likas, N. Vlassis, and J. Verbeek, "The global k-means clustering algorithm," *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[20] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society. Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[21] H. Zou, "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, vol. 101, no. 476, pp. 1418–1429, 2006.

[22] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. on Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[23] J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. on Information Theory*, vol. 50, no. 6, pp. 1341–1344, 2004.

[24] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[25] T. Vincent and C. Novara, "Mixed parametric/non-parametric identification of systems with discontinuous nonlinearities," *Automatica*, vol. 49, no. 12, pp. 3661–3669, 2013.

[26] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B*, vol. 67, no. 1, pp. 91–108, 2005.

[27] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, no. 1, pp. 49–67, 2006.

[28] P. Zhao, G. Rocha, and B. Yu, "The composite absolute penalties family for grouped and hierarchical variable selection," *The Annals of Statistics*, pp. 3468–3497, 2009.

[29] J. E. Vogt and V. Roth, "The group-lasso: $\ell_{1,\infty}$ regularization versus $\ell_{1,2}$ regularization," in *Pattern Recognition*, 2010, pp. 252–261.

[30] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proceedings of the 5th international conference on Future energy systems*, 2014, pp. 265–276.