This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Sparse Optimization for Automated Energy End Use Disaggregation

Dario Piga, Andrea Cominola, Matteo Giuliani, Andrea Castelletti, and Andrea Emilio Rizzoli

***Abstract*— 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, in the authors' experience, they provide poor performance in reconstructing the power consumption trajectories overtime. In this brief, a new NIALM algorithm is presented, which, besides providing very accurate estimates of the aggregated consumption by appliance, also accurately characterizes the appliance power consumption profiles overtime. The proposed algorithm is based on the assumption that the unknown appliance power consumption profiles are piecewise constant overtime (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 overtime. 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 European Union [3], and the study reported in [4] shows that household energy consumption can be reduced by 10% to 15% through better energy demand management. 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: 1) users are aware of how much energy each appliance is consuming, and personalized hints for reducing their energy consumption can be given and 2) 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, nor is it 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]–[8] and references therein). State-of-the-art NIALM algorithms can be classified into two main categories: 1) machine learning and 2) optimization-based approaches. In the first category, we mention the methods based on hidden Markov models (HMMs) [9]–[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 overtime. This represents a serious drawback, since: 1) 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; 2) anomalous events, such as a device consuming an exceptional amount of power over an extended period, can be barely detected; and 3) 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 brief, 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 overtime (as it is typical for energy use patterns of household appliances), and exploits the information

Manuscript received January 30, 2015; revised June 2, 2015; accepted August 22, 2015. Manuscript received in final form August 31, 2015. This work was supported by the European Union Seventh Framework Programme (FP7/2007-2013) under Grant 619172. Recommended by Associate Editor L. Fagiano. *(Corresponding author: Dario Piga.)*

D. Piga is with the IMT Institute for Advanced Studies Lucca, Lucca 55100, Italy (e-mail: dario.piga@imtlucca.it).

A. Cominola, M. Giuliani, and A. Castelletti are with the Politecnico di Milano, Milan 20133, Italy (e-mail: andrea.cominola@polimi.it; matteo.giuliani@polimi.it; andrea.castelletti@polimi.it).

A. E. Rizzoli is with the Dalle Molle Institute for Artificial Intelligence Research, USI/SUPSI, Manno 6928, Switzerland (e-mail: andrea@idsia.ch).

Digital Object Identifier 10.1109/TCST.2015.2476777

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 overtime. Beside being able to handle situations where multiple appliances are operating simultaneously, the proposed algorithm is able to reconstruct the consumption trajectories overtime, thus overcoming the main drawback of the disaggregation methods available in the literature. This brief 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 data set [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 HMMs. The conclusion is 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 ith appliance at the jth operating mode (with $j = 1, \ldots, C_i$). The power demand $y_i(t)$ of the ith appliance at time t is then given by

$$y_i(t) = \left[B_i^{(1)} \cdots B_i^{(C_i)}\right] \begin{bmatrix} \theta_i^{(1)}(t) \\ \vdots \\ \theta_i^{(C_i)}(t) \end{bmatrix} + e_i(t) \tag{1}$$

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

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

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 behavior 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 brief, 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 overtime.

III. Disaggregation Algorithm

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

A. Standard Least Squares

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

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

with

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

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 achieve the following:

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

B. Adding Regularization

To exploit the information that: 1) the parameters $\theta_i^{(1)}(t), \theta_i^{(2)}(t), \ldots, \theta_i^{(C_i)}(t)$ can be either 0 or 1 and 2) each appliance can only operate at a single mode at each time

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

instant, the following regularized problem can be solved instead of (3):

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

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

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

where $\|\cdot\|_0$ denotes the cardinality of a vector (i.e., number of its nonzero components). Note that, on the one hand, the second term in the objective function of Problem (4) aims at enforcing sparsity in the vector $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]$. On the other hand, the vector $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]$ is guaranteed to have at least one element different than zero, because of the equality constraint appearing in Problem (4). The hyperparameter $\lambda_1 > 0$ is tuned by the user (for instance, through cross validation) for balancing the tradeoff between minimizing the fitting error (by decreasing the value of λ_1) and maximizing sparsity of the parameter vector $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]$ (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, to improve the accuracy of the final estimate, the parameters $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]$ can be scaled by nonnegative weights $[w_i^{(1)}(t)\ \ldots\ w_i^{(C_i)}(t)]$. This leads to the following convex approximation of Problem (4):

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

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

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

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 $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]]$. In fact, in the ideal case, only one component of the vector $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]$ should be nonzero (i.e., the ith appliance operates at a single mode at each time instant). The reader is referred to [22]–[25] for a detailed analysis of the properties of ℓ_1-regularization in sparse estimation problems.

C. Adding Regularization to Enforce Piece-Wise Constant Signals Power Demand Profiles

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 overtime (Assumption **C3**). 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)

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

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

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

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

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 the following.

1) Penalizing the norm of the difference between two consecutive parameters $\theta_i^{(j)}(t)$ and $\theta_i^{(j)}(t-1)$ is commonly referred to 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 overtime).
2) The term (6c) is a group (fused) Lasso penalty [27]–[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 $[\theta_i^{(1)}(t)\ \ldots\ \theta_i^{(C_i)}(t)]$ changes

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

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 $[\theta_i^{(1)}(t) \ \dots \ \theta_i^{(C_i)}(t)]$ 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):

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

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 $w_i^{(1)}(t),\dots,w_i^{(C_i)}(t)$ is the following: if the ith 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 $w_i^{(1)}(t),\dots,w_i^{(C_i)}(t)$ can be chosen as follows.

1) Given the training data set $\mathcal{D}_{T_t}$, for each time sample t compute $q_i^j(t)$ as the number of times the ith appliance is classified to be at mode c_j at the time samples $t+k24$h, with $k \in \mathbb{Z}$.
2) If $q_i^j(t) \neq 0$, the weight $w_i^j(t)$ is then given by: $w_i^j(t) = (1/q_i^j(t))$. 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 $t, t-24h, t+24h, t-48h, t+48h, \dots,$ but also the observations (possibly weighted) within given time intervals $[t+k24h-\Delta, t+k24h+\Delta]$.

2) On the Choice of the Weights k_i: The weights k_i can be chosen as follows: if the ith appliance rarely changes its operating mode overtime, then the time variation of the parameters $\theta_i(t)$ should be more penalized with respect to the time variation of the parameters characterizing another 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 data set for the ith appliance, and scaled by the length of the training data set.

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. To overcome this problem, a suboptimal solution of Problem (6) can be computed by splitting the data set $\mathcal{D}_T$ into M disjoint subsets $\mathcal{D}^{(h)}$ of length T_h (with $h=1,\dots,\mathrm{M}$) 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 ith appliance is guaranteed not to operate at the jth 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 ith 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 data set [17], which contains the energy consumption readings of a single house located in the Vancouver region in British Columbia, Canada. Specifically, the AMPds data set contains the power consumption profiles of 19 appliances monitored for an entire year (from April 1, 2012 to March 31, 2013) at 1 min 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: 1) clothes dryer; 2) fridge; 3) dishwasher; and 4) heat pump. The contribution of the selected appliances is about 45% of the total energy consumption. Furthermore, to assess the robustness of the disaggregation algorithm with respect to 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 data set is divided as follows.

1) 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 (1)]. Therefore, the submetered power consumption trajectories of each

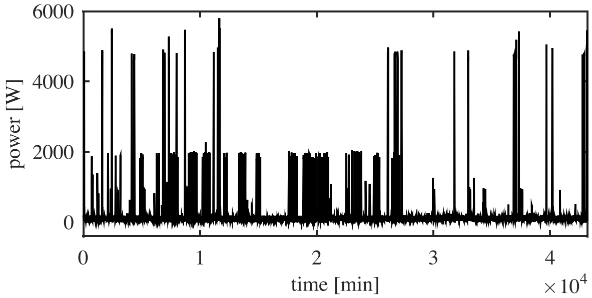

Fig. 1. Validation data set $\mathcal{D}_T$: aggregate electric power consumption from June 1, 2012 to June 30, 2012.

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 submetered power consumptions in the training data set. The chosen values of $B_i^{(j)}$ are as follows:

a) clothes dryer: [0 260 4700] W;
b) fridge: [0 128 200] W;
c) dish washer: [0 120 800] W;
d) heat pump: [39 1900] W.

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

a) clothes dryer: $k_1 = 273$;
b) fridge: $k_2 = 11$;
c) dish washer: $k_3 = 165$;
d) heat pump: $k_4 = 444$.

2) A calibration data set, which consists of the measurements from May 30, 2012 to May 31, 2012. The calibration data set is used to tune the hyperparameters λ_1 and λ_2 in (6). Also in the calibration phase, the submetered 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) with respect to the calibration data set, where the TRSE is defined as

$$\text{TRSE} = \sum_{i=1}^{N} \frac{\sum_{t=1}^{T_c} (y_i(t) - \hat{y}_i(t))^2}{\sum_{t=1}^{T_c} y_i^2(t)}$$

with T_c being the length of the calibration data set. The chosen values of λ_1 and λ_2 are 10 and 750, respectively.

3) A validation data set $\mathcal{D}_T$, which consists of the data for the days June 1–30, 2012 (as plotted in Fig. 1). The proposed algorithm is applied to disaggregate the data of the set $\mathcal{D}_T$.

To reduce the computational burden, a suboptimal 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 min). 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 the following points are to be noted.

1) 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.
2) 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.
3) 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 brief is compared with the performance of the disaggregation approach based on factorial HMM (FHMM) and implemented in the open source nonintrusive load monitoring toolkit [30]. In the FHMM-based approach, the power demand of each appliance is modeled as the observed value of an 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 modeled through an FHMM 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 brief, 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 data set $\mathcal{D}_T$. 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 brief and the performance of the FHMM-based approach.

1) 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 ith appliance, and it should be compared with 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 ith appliance. The EEFI $\hat{h}_i$ gives the users the information on how much energy each

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

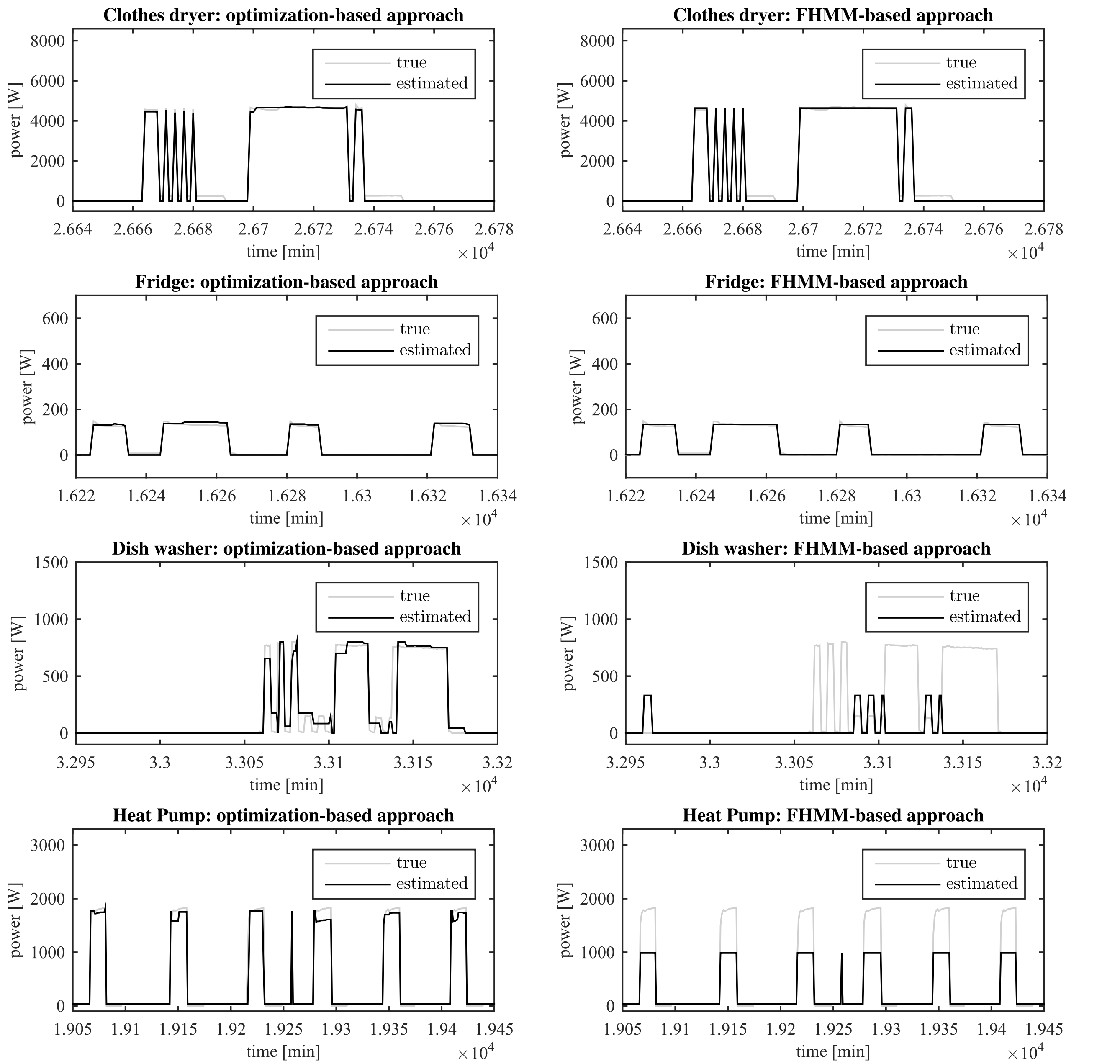

Fig. 2. Disaggregate power consumption profiles. Results obtained through the optimization-based approach presented in this brief (left panels) and through the FHMM-based approach (right panels).

appliance is consuming, and so personalized hints for reducing their energy consumption can be provided.

2) The relative square error (RSE), defined as

$$\mathrm{RSE}_i = \frac{\sum_{t=1}^{T}(y_i(t) - \hat{y}_i(t))^2}{\sum_{t=1}^{T} y_i^2(t)}.$$

The RSE provides a normalized measure of the difference between the actual and the estimated power consumptions of the ith appliance.

3) The R^2 coefficient, defined for the ith appliance as

$$R_i^2 = 1 - \frac{\sum_{t=1}^{T}(y_i(t) - \hat{y}_i(t))^2}{\sum_{t=1}^{T}(y_i(t) - \bar{y}_i)^2}$$

with $\bar{y}_i = (1/T)\sum_{t=1}^{T} y_i(t)$. Both the R^2 coefficient and the RSE measure how well the estimated power profiles match the actual power profiles overtime. An accurate estimate of the power consumption profiles overtime 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 $\hat{h}_i$.

D. Numerical Results

The performance metrics introduced in the previous section and the estimated disaggregate power profiles are computed to assess the performance of the two disaggregation algorithms. The obtained results are reported in Tables I and II and 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

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 THIS BRIEF 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

RSES AND R^2 COEFFICIENTS. RESULTS OBTAINED USING THE OPTIMIZATION-BASED ALGORITHM PRESENTED IN THIS BRIEF AND THE FHMM-BASED APPROACH

	optimization-based algorithm		FHMM-based algorithm	
	RSE_{i}	R_i^2	RSE_{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 %

of the data set 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 overtime (as shown in Table II and Fig. 2). The obtained results also reveal the following.

1) 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.
2) 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.
3) 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 brief, a novel algorithm for NIALM 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 overtime. The proposed method is able to handle situations where multiple appliances are operating simultaneously, and also to accurately estimate the appliance power consumption profiles overtime. Ongoing research activities are focused on the following:

1) extensive testing of the algorithm's generalization potential across different data sampling resolutions (i.e., 1 s, 15 min, and 1 h) and with respect to new, unseen, and continuously varying appliances;
2) application to high-resolution water consumption data.

ACKNOWLEDGMENT

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

REFERENCES

[1] *Annual Energy Outlook 2012*, U.S. Energy Inf. Administration, Washington, DC, USA, Jun. 2012.
[2] J. Ebinger and W. Vergara, *Climate Impacts on Energy Systems: Key Issues for Energy Sector Adaptation*. Washington, DC, USA: World Bank Publications, 2011.
[3] J. I. Utley and L. D. Shorrock, "Domestic energy fact file 2008," Dept. Energy Climate Change, Building Res. Establishment, Garston, U.K., Tech. Rep., 2008. [Online]. Available: http://www.bre.co.uk/filelibrary/pdf/rpts/fact_file_2008.pdf
[4] S. Darby, *The Effectiveness of Feedback on Energy Consumption: A Review for DEFRA of the Literature on Metering, Billing and Direct Displays*. Oxford, U.K.: Environmental Change Institute, 2006.
[5] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.
[6] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.
[7] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 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," *Environ. Model. Softw.*, vol. 72, pp. 198–214, Oct. 2015.
[9] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "An unsupervised training method for non-intrusive appliance load monitoring," *Artif. Intell.*, vol. 217, pp. 1–19, Dec. 2014.
[10] M. J. Johnson and A. S. Willsky, "Bayesian nonparametric hidden semi-Markov models," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 673–701, 2013.
[11] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. 15th Int. Conf. Artif. Intell. Statist.*, 2012, pp. 1472–1482.
[12] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G. M. P. O'Hare, "Real-time recognition and profiling of appliances through a single electricity sensor," in *Proc. 7th Annu. IEEE Commun. Soc. Conf. Sensor Mesh Ad Hoc Commun. Netw. (SECON)*, Jun. 2010, pp. 1–9.
[13] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Trans. Power Del.*, vol. 21, no. 1, pp. 398–405, Jan. 2006.
[14] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, "Nonintrusive appliance load monitoring based on integer programming," in *Proc. SICE Annu. Conf.*, Aug. 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 Comput. Sci.*, vol. 19, pp. 524–532, Jun. 2013.
[16] M. Figueiredo, B. Ribeiro, and A. M. de Almeida, "On the regularization parameter selection for sparse code learning in electrical source separation," in *Adaptive and Natural Computing Algorithms*. Berlin, Germany: Springer-Verlag, 2013, pp. 277–286.

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

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

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

[20] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Statist. Soc. B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[21] H. Zou, "The adaptive lasso and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

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

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

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

[25] T. L. 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," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, vol. 67, no. 1, pp. 91–108, 2005.

[27] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc. B (Statist. Methodol.)*, 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," *Ann. Statist.*, vol. 37, no. 6A, 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*. Berlin, Germany: Springer-Verlag, 2010, pp. 252–261.

[30] N. Batra *et al.*, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proc. 5th Int. Conf. Future Energy Syst.*, 2014, pp. 265–276.