Unsupervised Energy Disaggregation via Convolutional Sparse Coding

Christian Aarset\textsuperscript{D}, Andreas Habring\textsuperscript{D}, Martin Holler\textsuperscript{D}, and Mario Mitter

Abstract—In this work, a method for unsupervised energy disaggregation in private households equipped with smart meters is proposed. The method aims to classify power consumption as active or passive, granting the ability to report on the residents' activity and presence without direct interaction. This lays the foundation for applications like non-intrusive health monitoring of private homes. The proposed method is based on minimizing a suitable energy functional, for which the IPALM (inertial proximal alternating linearized minimization) algorithm is employed, demonstrating that various conditions guaranteeing convergence are satisfied. In order to confirm feasibility of the proposed method, experiments on semi-synthetic test data sets and a comparison to existing methods are provided.

Index Terms—Energy disaggregation, non-intrusive load monitoring, health monitoring, convolutional sparse coding.

I. INTRODUCTION

SMART meters – energy meters capable of digital communication – are becoming increasingly available in European domestic households. The regulation Intelligente Messgeräte-Einführungsverordnung – IME-VO \cite{1} derived from the EU directive \cite{2} dictates that 95\% of all Austrian meter points have to be equipped with smart meters by the end of 2024. As a result, regular readings of household energy consumption will be available on a large scale, potentially enabling a variety of insights into power consumption, usage patterns and human activity. For an overview of potential health care applications of smart meters see \cite{12,43}. As a lack of human activity in a household of, e.g., an elderly or care-dependent person over an extended period of time could be an indication of a potential emergency, in this work we focus on the possibility of detecting human activity, respectively a lack thereof, based on smart meter readings. Such an automatic detection would alleviate the need for the consumer to report incidents manually, and bypass the usage of intrusive surveillance devices.

Smart meters measure the aggregate energy consumption of a household rather than the consumptions of individual devices. Obtaining detailed information about the composition of the aggregate data is difficult, as it requires either directly measuring the energy consumption at many different points in the electrical circuit, or computational separation of the smart meter signal into distinct contributions, which is inherently ill-posed. For the latter problem, under the term computational energy disaggregation or non-intrusive load monitoring (NILM), there exists a variety of methods; we give an overview over existing literature in the following.

A. Related Works

Reviews of existing methods for NILM can be found in \cite{11,13}. NILM methods are frequently split into the categories supervised if they use labeled training data and unsupervised in the converse case. While the latter case is substantially more difficult, it is also relevant for practical applications, due to the cost and difficulty of collecting sufficient amounts of labeled training data.

a) Supervised methods: One of the first approaches introduced in \cite{14} is based on matching the changes in power consumption (event detection) to certain devices in the household. In \cite{16}, the authors use graph signal processing to predict changes in the state of individual devices. There are many works on different versions of hidden Markov models (HMM), and in particular factorial versions thereof \cite{8,28,45,51}. In \cite{26,40}, the authors propose sparse coding for NILM, and in \cite{47}, this approach is extended to a deep sparse coding model. Deep learning approaches have also been applied to NILM. In \cite{21}, the authors propose to use recurrent convolutional neural networks in a sequence-to-sequence fashion. In \cite{22}, the authors employ a sequence-to-sequence generative adversarial network (GAN) for energy disaggregation, and in \cite{20}, this model is improved by using a deep learning recurrent classifier to model the discriminator component of the GAN. For more work on deep learning for NILM, see also \cite{19,29,37,38,50}.

b) Unsupervised methods: By employing a priori assumptions on the device distributions, HMMs can be used in an unsupervised manner \cite{3,17,18,24,27,35,42}. In \cite{41}, the authors propose a method of tuning pre-trained general appliance models of an HMM to devices of a specific household by only using aggregate data from this household. In \cite{30}, the authors propose a method comprised of event detection, feature extraction, and classification. The method is proposed in an unsupervised and a supervised manner, using a decision tree and dynamic time wrapping for classification, respectively. Similar strategies are considered in \cite{32} and \cite{33}.
and in [49], such an approach is combined with graph signal processing methods. In [36], the authors propose to solve the NILM problem using non-negative matrix factorization, and also a shift-invariant version thereof.

**Contribution**

In this work, we propose a convolutional sparse coding (CSC) approach similar to [26], [40]. Contrary to these works, however, our method completely omits any prior dictionary learning using labeled data. We believe this poses a clear benefit, as we are able to circumvent several crucial issues: As elaborated in the introduction of [48], numerous difficulties are encountered in the acquisition, sharing, and reuse of household energy data, which leads to a rather limited amount of training data. Avoiding difficulties related to the scarcity of data by using a representative dictionary of general appliance models currently only works well for simple devices [35, Sec. II-D], especially since a sufficiently extensive dictionary is not yet available [39, Sec. 3.5.7]. Obtaining a dictionary by measuring one activity cycle of each electrical device in the household, on the other hand, poses an excessive and intrusive installation effort, and is consequently undesirable. Altogether, we therefore believe that methods not relying on training data are crucially needed for NILM. Moreover, our experiments suggest that, even when training data is available, our unsupervised approach is more robust compared to supervised ones with respect to shifts in distribution between training and testing data. (see Table II, Figure 1 and also [31], [46]).

Motivated by the aforementioned monitoring applications, in this work we will focus on experimental results for activity prediction, but provide also results for full disaggregation into individual devices. For activity prediction, we disaggregate into an active and a passive channel, where the presence of the former will be used as an indicator of human activity.

**II. MODELING AND PROBLEM FORMULATION**

The power consumption \( \mathbf{u} \in \mathbb{R}^m \) of a household, with \( m \in \mathbb{N} \) the number of evenly spaced time points, is given as the sum of the power consumptions of all individual devices \( \mathbf{u}^i \in \mathbb{R}^m \) for \( i = 1, 2, \ldots, N \) contained in this household as

\[
\mathbf{u} = \sum_{i=1}^{N} \mathbf{u}^i + \epsilon,
\]

where \( \epsilon \) denotes measurement noise. The goal of NILM is to recover all \( \mathbf{u}^i \) given only knowledge of \( \mathbf{u} \). Due to the severe ill-posedness of this problem, additional modeling assumptions have to be made in order to obtain reasonable estimates of the \( \mathbf{u}^i \). Specifically, our model builds on the observation that the main contributors to a household’s energy consumption consist of disparate device power consumption peaks, or simply device characteristics. That is, whenever a device is used for a fixed period of time, its individual power consumption over time will consistently be of the same approximate shape and magnitude. Let for instance \( p \in \mathbb{R}^{2q+1} \) be one activation cycle of length \( 2q + 1 \) of a dishwasher. Assume at each of the timestamps \( t_1, t_2, \ldots \) the dishwasher is activated for one cycle. Then, via the convolution, the dishwasher’s power consumption can be computed as

\[
\mathbf{c} * p := \left( \sum_{\Delta q = -q}^{q} c_{j+q} \mathbf{p}_{q+1-i} \right)_{j=1}^{m} \in \mathbb{R}^m
\]

where \( c = (0, \ldots, 0, 1, 0, \ldots, 0, 1, 0, \ldots, 0) \in \mathbb{R}^{m+2q} \) with the ones appearing at indices \( t_1, t_2, \ldots \). Figuratively speaking, the coefficient \( c \) places the activation \( p \) at the respective times.
Inspired by this, we model each device $u^i$ as a convolution $c^i * p^i$ – or a sum thereof for devices with multiple different characteristic activation cycles – of an atom $p^i \in \mathbb{R}^{2q+1}$ resembling one activation cycle of device $i$ and a coefficient $c^i \in \mathbb{R}^{m+2q}$ indicating the times of activation of device $i$. Explicitly, this leads to the following formulation:

$$u = \sum_{i=1}^{N} c^i * p^i + \epsilon$$

For the sake of simplifying notation, we also denote the sum-of-convolutions map for multiple devices $c * p := \sum_{i=1}^{N} c^i * p^i$.

In the application of activity detection, we are particularly interested in the case where we have several active channels, representing the characteristics of all devices that require human interaction – thus indicating human activity – and one passive channel, consisting of the characteristics of all remaining devices. Examples of devices contributing to the active channels include, but are not limited to, ovens, washing machines and hair dryers, while devices such as fridges and routers might contribute to the passive channel. Devices in the active channels are assumed to be needed only a few times in the span of a single day, in the sense that the number of times they are activated each day is relatively low. This leads to a natural sparsity of incidents of each device characteristic on the temporal axis (see Figure 1, top right). In summary, we desire the following properties for our atoms and coefficients:

- The atoms should be bounded, so that scaling is left to the coefficients. Atoms and coefficients should be positive.
- With $(c^1, p^1)$ representing the passive channel, we let all remaining pairs, that is, $(c^i, p^i)$ for $i \geq 2$, serve to explain the active channels; thus, active energy consumption is approximated by $\sum_{i=2}^{N} c^i * p^i$. The coefficients $c^i, i \geq 2$ will be penalized in the $\| \cdot \|_2$-norm, as this norm emulates the previously mentioned temporal sparsity (see [15]), while remaining analytically tractable. However, as some passive devices, such as a router, may consume power throughout most of the day, a sparsity-enforcing penalty term for $c^1$ becomes undesirable. Accordingly, we penalize $c^1$ in the $\| \cdot \|_2$-norm.

Combining the above leads to the minimization problem

$$\min_{c, p} \| u - c * p \|^2_2 + \lambda_p \| c^1 \|_2 + \lambda_a \sum_{i=2}^{N} \| c^i \|_1$$

s.t. $c \geq 0, p \geq 0, \| p^i \|_2 \leq 1$ for all $i = 1, \ldots, N$,

with $\lambda_p, \lambda_a > 0$; the positivity constraints $c \geq 0$ and $p \geq 0$ are meant in the componentwise sense. The first term in (2) ensures the solution to approximately solve the energy disaggregation problem and the remaining terms incorporate constraints and ensure well-posedness.

### III. Inertial Proximal Alternating Linearized Minimization

As (2) does not generally allow for an analytical solution, we require an appropriate algorithm to numerically solve the minimization problem. In that respect, we face two main difficulties with our problem formulation: It is neither jointly convex nor differentiable in $(c, p)$, meaning standard convex or gradient descent-based algorithms cannot be used as is. To overcome this, we employ the inertial Proximal Alternating Linearized Minimization (iPALM) algorithm as proposed in [44], which is suitable for nonconvex, nonsmooth optimization problems of a particular structure. Therefore, we first rewrite (2) into a form where the iPALM algorithm is applicable, i.e., we consider

$$\min_{c, p} F(c, p), \text{ where } F(c, p) := R(c, p) + f_1(c) + f_2(p). \quad (2)$$

Here, $R(c, p) := \| u - c * p \|^2_2$.

$$f_1(c) := \begin{cases} \lambda_p \| c^1 \|_2 + \lambda_a \sum_{i=2}^{N} \| c^i \|_1, & c \geq 0, \\ \infty, & \text{otherwise,} \end{cases}$$

$$f_2(p) := \begin{cases} 0, & p \geq 0 \text{ and max} \{ \leq i \leq N \} \| p^i \|_2 \leq 1, \\ \infty, & \text{otherwise.} \end{cases}$$

To ensure convergence of iPALM according to [44], several properties of the functionals involved in (2) need to be satisfied. It is clear that $f_1, f_2$ and $R$ are proper, lower-semicontinuous, bounded from below, and that $R$ is differentiable. Moreover, $f_1$ and $f_2$ are both convex, with $f_1$ being a linear combination of componentwise norms under a convex constraint, and $f_2$ simply being a convex constraint. We further need Lipschitz continuity of the derivative of $R$ as stated in the following proposition.

**Proposition 1:** For $R$ as above, the following holds:

1) For any fixed $p \in (\mathbb{R}^{2q+1})^N, c \mapsto \nabla_c R(c, p)$ is Lipschitz continuous. Similarly, for any fixed $c \in (\mathbb{R}^{m+2q})^N, p \mapsto \nabla_p R(p, c)$ is Lipschitz continuous.

2) $\nabla R$ is Lipschitz continuous on bounded subsets of $(\mathbb{R}^{m+2q})^N \times (\mathbb{R}^{2q+1})^N$.

**Proof:** Write $R(c, p) := \| u - \cdot \|_2^2 \circ [u - \cdot] \circ (c * p)$. As the map $c \mapsto c * p$ is linear for fixed $p$, we have $c * p = P(p)c$ with $P : (\mathbb{R}^{2q+1})^N \to \mathbb{R}^{m \times (m+2q)}$. Simultaneously, for fixed $c, c * p = C(c)p$ with $C : (\mathbb{R}^{m+2q})^N \to \mathbb{R}^{m \times N(2q+1)}$. Thus,

$$\nabla_c R(c, p) = -2P(p)^T[u - c * p],$$

$$\nabla_p R(c, p) = -2C(c)^T[u - c * p].$$

It follows that for any fixed $p \in (\mathbb{R}^{2q+1})^N$ and any $c, \tilde{c} \in (\mathbb{R}^{m+2q})^N$, one has

$$\| \nabla_c R(c, p) - \nabla_c R(\tilde{c}, p) \|_2 \leq 2 \| P(p) \|^2 \| c - \tilde{c} \|_2 =: L_1(p) \| c - \tilde{c} \|_2$$

and analogously, for all $c \in (\mathbb{R}^{m+2q})^N$ and all $p, \tilde{p} \in (\mathbb{R}^{2q+1})^N$,

$$\| \nabla_p R(c, p) - \nabla_p R(c, \tilde{p}) \|_2 \leq 2 \| C(c) \|^2 \| p - \tilde{p} \|_2 =: L_2(c) \| p - \tilde{p} \|_2,$$

showing 1).

To show 2), we note that as the maps $p \mapsto P(p), p \mapsto P(p)^T, c \mapsto C(c)$ and $c \mapsto C(c)^T$ are linear and bounded, it is clear that $\nabla_c R(c, p)$ and $\nabla_p R(c, p)$ are differentiable with respect to $(c, p)$ jointly, with bounded derivatives on bounded subsets of $(\mathbb{R}^{m+2q})^N \times (\mathbb{R}^{2q+1})^N$. The mean value inequality immediately yields Lipschitz continuity on bounded sets.

**Remark 1:** The preceding theorem ensures that [44, Assumption A] is satisfied (note that the lower bound in
Writing it out explicitly, one has

\[ c_{k+1} \in \text{prox}_{\beta_i}(y_{k+1} - \frac{1}{\tau_i} \nabla R(c_i^k, p_k)). \]

**Remark 2:** The Lipschitz constants from 1) in Proposition 1 are needed for the stepsize choice in the iPALM algorithm. Those correspond to the operator norms of \( \|P(p)\| \) resp. \( \|C(c)\| \), which can be computed explicitly via spectral radii, but this may prove impractical for numerics. Denoting the matrix representation of \( c^i \mapsto c^i * p^j \) as \( P^i(p^j) \), norm estimates on the \( P^i(p^j) \) can be obtained as follows: With \( J : \mathbb{R}^{m+2q} \rightarrow \ell^2(\mathbb{R}) \), \( Jx_i := x_i \) for \( 1 \leq i \leq m + 2q \) and \( Jx_i := 0 \) otherwise, with \( K : \mathbb{R}^{2q+1} \rightarrow \ell^2(\mathbb{R}) \), \( (Kx)_i := x_{i+2q+1} \) for \( -2q \leq i \leq 0 \) and \( (Kx)_0 := 0 \) otherwise, and with the discrete convolution between \( \ell^2(\mathbb{R}) \) and \( \ell^2(\mathbb{R}) \), Young’s inequality implies

\[
\|P^i(p^j) c_i^j\|_2 \leq \|c_i^j * p^j\|_2 \leq \|Jc_i^j\| \|Kp^j\|_2 \leq \|c_i^j\|_2 \|p^j\|_1.
\]

Taking the supremum yields \( \|P^i(p^j)\| \leq \|p^j\|_1 \). A similar argument shows \( \|C(c)\| \leq \|c\|_1 \) for \( C(c) \) the matrix representation of \( p^j \mapsto c^i * p^j \). Applying standard norm estimates, one altogether has

\[
\|C(c)\| \leq \min \left\{ \|c\|_1, \sqrt{2q+1} \max_{0 \leq \Delta \leq 2q} \sum_{j=1}^m c_j^{i+j} \right\},
\]

\[
\sqrt{m} \max_{0 \leq \Delta \leq m-1} \sum_{j=1}^{2q+1} \sum_{j=1}^m \sum_{\Delta=0}^m |c_j^{i+j+\Delta}|^2 ,
\]

where it can be seen that each of these estimates may be optimal under various circumstances. Note also that although similar estimates are available for \( P^i(p^j) \), its particular structure causes these to be all bounded below by \( \|p^j\|_1 \). Norm estimates for \( P(p) \), respectively \( C(c) \) are obtained using that, by Cauchy-Schwarz’s inequality, for \( v \in (\mathbb{R}^{m+2q})^N \), \( \|P(p) v\|_2 \leq \sqrt{\sum_{i=1}^N \|P^i(p)\|_2^2 \|v\|_2^2} \) and, analogously, \( \|C(c)\| \leq \sqrt{\sum_{i=1}^N \|C(c^i)\|^2} \).

Applied to our setting, the iPALM algorithm takes the following form: Fix a suitable non-zero initialisation \((c_{-1}, p_{-1}) = (c_0, p_0) \in (\mathbb{R}^{m+2q})^N \times (\mathbb{R}^{2q+1})^N \), all non-negative componentwise. Fix \( \varepsilon \in (0, 1) \) small, and for \( i \in [1, 2] \), choose \( \delta_i \in (0, 1 - \varepsilon) \), \( \beta_i > 0 \) arbitrary. Define

\[ \gamma_i := \frac{\delta_i + 2\beta_i}{2(1 - \varepsilon - \alpha_i)} \lambda_i^+ \]

where \( \lambda_i^+ \) resp. \( \lambda_i^- \) corresponds to the upper bound over all \( k \) on \( L_1(p_k) \) as in (3), resp. \( L_2(c_k) \) as in (4); when such a bound is not analytically apparent, a viable substitute in the \( k \)-th iteration step is \( \max_{1 \leq j \leq k} L_1(p_j) \) resp. \( \max_{1 \leq j \leq k+1} L_2(c_j) \), possibly dropping early indices to avoid excessive size. For each \( k \in \mathbb{N}_0 \), repeat the following process:

**Require:** Pick \( \alpha_i^k \in [0, \alpha_i] \), \( \beta_i^k \in [0, \beta_i] \).

\[
y_{k+1}^i := c_k^i + \alpha_i^k (c_k^i - c_{k-1}^i)
\]

\[
z_{k+1}^i := c_k^i + \beta_i^k (c_k^i - c_{k-1}^i)
\]

\[
\tau_k^i := \frac{(1 + \varepsilon)\gamma_i + (1 + \beta_i^k)\delta_i}{2 - 2\alpha_i^k}
\]

\[
c_{k+1}^i \in \text{prox}_{\beta_i^k}(y_{k+1}^i - \frac{1}{\tau_k^i} \nabla R(z_{i+1}^k, p_k)).
\]

From this, it is apparent that any solution \( \hat{c} \in \mathbb{R}^{m+2q} \) will satisfy \( d_i = 0 \) if \( c_i^1 < 0 \). As such, one can replace \( c_1 \) by \( \{c_1^1\}_+ \) in the above minimization problem without changing the solution. But with \( \{c_1^1\}_+ \) instead of \( c_1 \), it is clear that any solution of the above problem will be non-negative even without the non-negativity constraint, making it equivalent to

\[
\text{prox}_{\lambda_i^+ \| \cdot \|_2 \left( \{c_1^1\}_+ \right) := \arg \min_{\hat{c} \in \mathbb{R}^{m+2q}} \frac{1}{2} \sum_{i=1}^{m+2q} d_i^2 + \frac{1}{2} \sum_{i=1}^{m+2q} d_i - \left\{ \{c_1^1\}_+ \right\}^2.
\]
By [6, Example 6.19], the solution of the latter is of the claimed form.

\[ (\text{prox}^f_\lambda(c))^i \text{ for } i \geq 2; \]

Writing \( f_1(c) = |c| + \delta_+(c) \) for \( 1 \leq j \leq N \), a combination of [6, Example 6.8] with [9, Lemma 6.14(b)] immediately yields the claimed form.

\[ \text{prox}^f_\lambda(p) \text{: Observe that } f_2(p) = \sum_{i=1}^N \delta_{C_i}(p), \text{ where } C_i := \{ x \in \mathbb{R}^{2q+1} \mid x \geq 0, \|x\|_2 \leq 1 \}. \]

Employing [6, Ths. 6.6, 6.24 and 6.30], as well as noting that \( C_i = \{ x \in \mathbb{R}^{2q+1} \mid g_i(x) \leq 1 \} \), where \( g_i(x) := \delta_+(x) + \|x\|_2^2 \) with effective domain \( \mathbb{R}_+^{2q+1} \), one has

\[ \left(\text{prox}^f_\lambda(p)\right)^i = P_{C_i}(p^i) = \begin{cases} p^i, & g_i\left(\left[p^i\right]_+\right) \leq 1, \\ \text{prox}_{\lambda g_i}\left(p^i\right), & \text{else}, \end{cases} \]

where \( \lambda > 0 \) solves

\[ 1 = g_i(\text{prox}_{\lambda g_i}(p^i)) = g_i\left(\frac{p^i}{2\lambda + 1}\right) \]

\[ = g_i\left(\left[p^i\right]_+\right) = \frac{1}{2\lambda + 1} \|p^i\|_2^2 \]

with the last three equalities following from [6, Ths. 6.13, 6.24 and Lemma 6.26]. Simplifying leads to the desired form. \( \blacksquare \)

### IV. Experiments

In this section, we provide experimental results of the proposed method (CSC) for two different scenarios, namely activity prediction, where we try to detect whenever energy is consumed actively by a person, and multichannel disaggregation, where we aim to reconstruct all individual device signals that amount to an aggregate energy signal. In particular, Table II in Section IV-D1, and Section IV-D2 contain results for the setting that no ground truth training data is available for the household of interest, constituting a realistic use case as explained in Section I.

#### A. Data

For our experiments, we employ the SynD data set generator [25], an algorithm to generate synthetic energy signals. This enables us to work with clean, yet realistic, data. The algorithm uses real device energy signatures, which it randomly alters and places throughout the day. It became apparent, however, that the device’s energy signal generated with SynD is unrealistically regular, being identical each day; for this reason, we replaced the fridge by an energy signal taken from the UK-DALE 6 seconds data set [23], and resampled to the desired sampling rate by weighted averaging.

Explicitly, we generate two different synthetic households, referred to as syn1 and syn2. Each of the two data sets consists of four devices, syn1 of the fridge from UK-DALE’s house 1 as well as SynD’s electric space heater, washing machine, and hair dryer, and syn2 of the fridge from UK-DALE’s house 2 and SynD’s dishwasher, iron, and watercooker. We use a sampling rate of 1 minute, and the data sets are normalized to have values between 0 and 1.

#### B. Performance Metrics

1) Binary Performance Metrics: Both experiments – activity prediction and multichannel disaggregation – can be cast as binary classification tasks with multiple channels. At each time point each of the channels to predict has to be classified as either ON or OFF. In case of activity prediction, for each time point we aim to classify the active signal as ON or OFF depending on whether a person is actively consuming energy or not, and similarly classify the passive signal as ON or OFF depending on whether a passive device is consuming energy or not. In case of multichannel disaggregation, we aim to classify each individual device as ON or OFF. Note that the proposed method, as well as the comparison methods, yield non-binary output signals, namely the actively consumed energy or the individual device energy signatures, rather than a binary ON/OFF signal. Thus, before evaluating classification performance, non-binary predictions are first transformed to binary signals via pointwise thresholding by 0.01 times the maximum value of the ground truth signal. Given \( \hat{u}, u \in \mathbb{R}^m \) a true non-binary signal and its prediction, respectively, we compute \( \hat{U} \) and \( U \) as \( \hat{U} = (\hat{u} \geq 0.01\|\hat{u}\|_\infty) \) and \( U = (u \geq 0.01\|u\|_\infty) \) in Python notation. As a metric for evaluating binary classification, we use a modified Matthews correlation coefficient (MCC) and F1 score. Given \( \hat{U}, U \in \{0, 1\}^m \) a true signal and its prediction, respectively, the MCC and F1 are defined as

\[ \text{MCC}(\hat{U}, U) := \frac{(\text{TP})(\text{TN}) - (\text{FP})(\text{FN})}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \]

and

\[ \text{F1}(\hat{U}, U) := \frac{2 \text{TP}}{2 \text{TP} + \text{FP} + \text{FN}} \]

where TP \( = \hat{U} \cdot U \), TN \( = (1 - \hat{U}) \cdot (1 - U) \), FP \( = (1 - \hat{U}) \cdot U \), and FN \( = \hat{U} \cdot (1 - U) \), with \( \cdot \) the Euclidean scalar product, denote the number of true positives, true negatives, false positives, and false negatives, respectively. The MCC score is a value in \([-1, 1]\), where +1 and −1 correspond to perfect agreement or disagreement, respectively, and 0 indicates no relationship. The F1 score is a value in \([0, 1]\), 1 being the best score. In view of relevant monitoring applications, where it is sufficient to predict activity up to a reasonable inaccuracy in time, we introduce an activity prediction radius \( R \in \mathbb{N}_0 \) into the computation of the scores. We first define a modified true activity \( \hat{U} \) via \( \hat{U}_i = 1 \) if there exists \( j \in \{1, 2, \ldots, m\}, |i - j| \leq R \) such that \( \hat{U}_j = 1 \) and \( U_j = 0 \) otherwise. Hence, \( \hat{U}_i \) is equal to one whenever there was true activity within a (2R + 1) minute time window around \( i \). We then compute TP \( = \hat{U} \cdot U \), TN \( = (1 - \hat{U}) \cdot (1 - U) \), FP \( = (1 - \hat{U}) \cdot U \), and FN \( = \hat{U} \cdot (1 - U) \). Using these values, the modified MCC and F1 are obtained via (5) and (6) and denoted as MCC\(_R\) and F1\(_R\), respectively. In our experiments, we use \( R = 7 \).

While both metrics are presented going forward, we generally place additional emphasis on the MCC, as it is more robust. Whereas the F1 score may falsely assign near-perfect scores by, e.g., classifying every instance as positive in a data set containing significantly less negatives than positives, the MCC inherently adjusts for such data set imbalance, see [10].

2) Non-Binary Performance Metrics: We do not only want to quantify whether the state of a device was correctly predicted, but also how well a method predicts the magnitude of its energy consumption. Thus, we also employ two non-binary metrics for evaluating activity and energy consumption predictions, respectively. For evaluating activity prediction, we use the area under the ROC curve (AUC-ROC) and the Dice coefficient (Dice) as metrics. For evaluating energy consumption predictions, we use the mean absolute percentage error (MAPE) and the mean squared error (MSE) as metrics.
TABLE I
TESTING SCORES OF THE ACTIVE ENERGY USAGE FOR DIFFERENT METHODS APPLIED TO SYND1 AND SYND2. TRAINING AND TESTING ON DIFFERENT TIME FRAMES OF THE SAME HOUSEHOLD.  BOLD INDICATES BEST RESULT

|                  | CSC | GSP | FHMM | CO | S2S | DSC |
|------------------|-----|-----|------|----|-----|-----|
| synd1 MCC-7      | 0.78| 0.84| 0.79 | 0.91| 0.97| 0.44|
| Fl1              | 0.78| 0.84| 0.81 | 0.92| 0.97| 0.47|
| NMAE             | 0.01| 0.07| 0.12 | 0.09| 0.04| 0.36|
| synd2 MCC-7      | 0.13| 0.96| 0.38 | 0.30| 0.79| 0.95|
| Fl1              | 0.95| 0.88| 0.87 | 0.90| 0.98| 0.73|
| NMAE             | 0.002| 0.11| 0.14 | 0.11| 0.05| 0.21|
| NMAE             | 0.17| 0.63| 0.66 | 0.57| 0.53| 0.92|

metrics to evaluate performance, namely the normalized mean squared error (NMSE) and the normalized maximum absolute error (NMAE), defined as

\[ \text{NMSE}(\hat{u}, u) = \frac{\|\hat{u} - u\|^2_2}{\|u\|^2_2}, \quad \text{NMAE}(\hat{u}, u) = \frac{\max_i |\hat{u}_i - u_i|}{\max_i |u_i|}. \]

C. Experimental Setup

The CSC experiments were carried out in the following manner: Given an aggregate energy signal \( u \in \mathbb{R}^m \), initialize \((c_{-1}, p_{-1}) = (c_0, p_0) \in (\mathbb{R}^{m+2q})^N \times (\mathbb{R}^{2q+1})^N\) by, e.g., selecting random non-zero features of \( u \) of appropriate length as atoms and their location, shifted appropriately, as coefficients. Pick \( \alpha_k^k = \beta_k^0 = \beta_k^k = 0.7 \) for all \( k \in \mathbb{N}_0 \), and compute the \( \tau_1^k \) and \( \tau_2^k \) in accordance with [44, Remark 4.1]. Employ the relative change in the modified energy

\[ \Psi_k := \|c_{k} - c_{k-1}\|^2_2 + \|p_k - p_{k-1}\|^2_2 \] (see [44, (4.6)]) as the stopping rule; explicitly, we stop when \( |\Psi_k - \Psi_{k-1}| < 10^{-6} \) or when \( \Psi_k = 0 \). Across all experiments, the method terminates after around 3000-6000 iterations. One iteration takes roughly 0.55sec on an AMD Ryzen 7 3800X 8-Core Processor CPU.

D. Results

1) Activity Prediction: In our data sets, the passive consumption is comprised solely of the fridge signal, while the active consumption consists of all remaining devices. We compare to four different supervised methods from the non-intrusive load monitoring toolkit NILMTK [4], [5], namely, combinatorial optimisation (CO), a factorial hidden Markov model (FHMM), a neural network model (S2S), and discriminative sparse coding (DSC), and to the fully unsupervised graph signal processing method (GSP) [49], for which a Python implementation is available at [34]. All methods are used to predict the active and the passive energy consumption. The data is normalized before plotting and computing the performance metrics. We report the scores on the active component of the signal in Table I, and II. In Table I, we show the testing scores for the case of training and testing on the same household. That is, all trainable algorithms use four weeks of the energy signal for supervised training, and a different selection of four weeks of the same energy signal for testing. In this setting, we observe that the proposed method performs comparably to the deep learning method S2S, and that they both outperform the remaining methods. For the scores in Table II, on the other hand, we use 4 weeks of data of different households for training and testing. This corresponds to the use case, that no training data of the household of interest is available and, thus, training is conducted using accessible data of a different house, which leads to a distribution shift between training and testing data, especially since synd1 and synd2 even contain different devices. Empirically, the proposed method on average outperforms its (supervised) rivals in this setting. While CSC also uses data to determine hyperparameters, it uses significantly less information for this process than the other methods, and is thus less prone to over-fitting. That is, it is not necessary for CSC to learn, e.g., the distribution of individual devices (FHMM), the shapes of device signatures (DSC), or the parameters of a neural network etc., but solely to determine the hyperparameters \( \lambda_\alpha, \lambda_\beta, N \) and \( q \) of our energy functional. We believe that this is the reason for CSC specializing better than its supervised rivals to unseen data with a potentially slightly different distribution. In general, in Table I, and II we observe that S2S yields the best MCC/F1 scores and CSC performs superior in terms of NMSE/NMAE scores. We believe that this relates to the fact that CSC recognizes the specific shapes of energy consumption more accurately. This is also supported by Figure 1. While FHMM and CO also closely fit the ground truth passive energy signal, CSC is the only method to capture the spiked shape of the fridge signal. The deficiency in the MCC score is a result of the reconstructed signal not being close enough to zero at times where the ground truth is zero. It should also be noted that the second unsupervised method, GSP, performs very well on average, with only the reconstruction of the passive signal failing severely in Figure 1, which is most likely caused by the method misinterpreting starting and ending points of events.

2) Unsupervised Multichannel Disaggregation: In this experiment, we investigate the performance of CSC and GSP applied to multichannel disaggregation, that is, inferring all individual devices from the given aggregate energy signal. We consider again the use case of no ground truth training data being available for the household of interest and estimate the CSC hyperparameters on a different dataset than the one used for testing. We cannot perform this experiment with the remaining methods, as those learn characteristics of specific devices, however, here the devices contained in the training and testing data differ. Therefore, this use case

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
TABLE III

|       | synd1 | synd2 |
|-------|-------|-------|
|       | Frigde | Hair dryer | Space heater | Washing machine |
|       | MCC<sub>7</sub> | 0.48 | 0.18 | 0.74 | 0.5 | 0.55 | 0.69 | 0.56 | 0.8 |
|       | F1<sub>7</sub> | 0.74 | 0.3 | 0.73 | 0.88 | 0.88 | 0.57 | 0.72 | 0.87 |
|       | RMSE | 0.24 | 1.53 | 0.63 | 1.05 | 0.95 | 1.05 | 0.76 | 0.8 |
|       | NMAE | 1.53 | 1.05 | 0.99 | 1.96 | 1.95 | 1.98 | 2.06 | 0.98 |
|       | MCC<sub>7</sub> | 0.48 | 0.18 | 0.74 | 0.5 | 0.55 | 0.69 | 0.56 | 0.8 |
|       | F1<sub>7</sub> | 0.74 | 0.3 | 0.73 | 0.88 | 0.88 | 0.57 | 0.72 | 0.87 |
|       | RMSE | 0.24 | 1.53 | 0.63 | 1.05 | 0.95 | 1.05 | 0.76 | 0.8 |
|       | NMAE | 1.53 | 1.05 | 0.99 | 1.96 | 1.95 | 1.98 | 2.06 | 0.98 |
|       | MCC<sub>7</sub> | 0.48 | 0.18 | 0.74 | 0.5 | 0.55 | 0.69 | 0.56 | 0.8 |
|       | F1<sub>7</sub> | 0.74 | 0.3 | 0.73 | 0.88 | 0.88 | 0.57 | 0.72 | 0.87 |
|       | RMSE | 0.24 | 1.53 | 0.63 | 1.05 | 0.95 | 1.05 | 0.76 | 0.8 |
|       | NMAE | 1.53 | 1.05 | 0.99 | 1.96 | 1.95 | 1.98 | 2.06 | 0.98 |

constitutes a setting where only unsupervised methods are feasible. Quantitative results can be found in Table III, and qualitative ones in Figure 2. Overall, CSC slightly outperforms GSP in terms of quantitative scores. In general, the reconstruction quality of active devices with CSC is dependent on the specific device signature. In Figure 2, we observe that devices with a more distinct and/or extended signature, such as the space heater or the iron, are reconstructed properly. The reconstruction is less precise for devices like the watercooker, with a signature composed of simple peaks of a very short duration.

V. Conclusion

In this work, we developed a new unsupervised method for energy disaggregation based on convolutional sparse coding with a focus on classifying energy consumption as active or passive. Prior training is not needed, as the method relies only on the choice of hyperparameters, which can be fixed for a broad class of input data. We employ the iPALM algorithm to minimize the used objective functional, ensuring strong convergence guarantees compared to many competing methods. As the results of our numerical experiments show, our unsupervised approach performs comparable to the state of the art in activity prediction in terms of classification scores, and yields superior performance in terms of NMSE and NMAE scores (see Table I, and II). Further, it recognizes specific energy consumption signatures better (see Figure 1), and offers more flexibility due to the untrained approach. Applied to unsupervised multichannel disaggregation, our method performs slightly superior to the comparison method GSP [49].

REFERENCES

[1] “Directive (EU) 2019/944 of the European Parliament and of the council of 5 June 2019 on common rules for the internal market for electricity and amending directive 2012/27/EU.” Accessed: Jun. 23, 2022. [Online]. Available: eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32019L0944

[2] “Intelligente messgeräte-einführungsverordnung.” Accessed: May 23, 2022. [Online]. Available: www.ris.bka.gv.at

[3] M. Aiad and P. H. Lee, “Unsupervised approach for load disaggregation with devices interactions,” Energy Build., vol. 116, pp. 96–103, Mar. 2016.

[4] 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.

[5] N. Batra et al., “Towards reproducible state-of-the-art energy disaggregation,” in Proc. 6th ACM Int. Conf. Syst. Energy-Effic. Build. Cities Transp., 2019, pp. 193–202.

[6] A. Beck, First-Order Methods in Optimization. Philadelphia, PA, USA: SIAM, 2017.

[7] J. Bolte, S. Sabach, and M. Teboulle, “Proximal alternating linearized minimization for nonconvex and nonsmooth problems,” Math. Program., vol. 146, pp. 459–494, Aug. 2014.

[8] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, “Non-intrusive load monitoring by using active and reactive power in additive factorial hidden Markov models,” Appl. Energy, vol. 208, pp. 1590–1607, Dec. 2017.

[9] K. Bredies and M. Holler, “Higher-order total variation approaches and generalisations,” Inverse Probl., vol. 36, Dec. 2020, Art. no. 123001.

[10] D. Chico and G. Jurman, “The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” BMC Genom., vol. 21, p. 6, Jan. 2020.

[11] S. Dash and N. Sahoo, “Electric energy disaggregation via non-intrusive load monitoring: A state-of-the-art systematic review,” Electr. Power Syst. Res., vol. 213, Dec. 2022, Art. no. 108673.

[12] M. Fell, H. Kennard, G. Huebner, M. Nicolson, S. Elam, and D. Shipworth, “Energising health: A review of the health and care applications of smart meter data. SMART Energy GB, London, U.K., 2017.

[13] R. Gopinath, M. Kumar, C. P. C. Joshua, and K. Sriwats, “Energy management using non-intrusive load monitoring techniques—state-of-the-art and future research directions,” Sustain. Cities Soc., vol. 62, Nov. 2020, Art. no. 102411.
