Balancing services provision via optimization and management of the flexibility offered by a pool of energy resources*

Daniel Zamudio[1], Alessandro Falsone[1] *Member, IEEE*,
Federico Bianchi[2] *Member, IEEE*, Maria Prandini[1] *Fellow, IEEE*

***Abstract*— Aggregators of energy resources providing balancing services to the grid face the twofold challenge of assessing the overall amount of flexibility of the pool of energy resources, and mapping the power requests by the grid back to each single energy resource during the service window. In this paper, we propose a framework that allows to jointly optimize the power flexibility limits and determine the disaggregation policy for managing the pool of energy resources, thus avoiding an additional computational step in the operational service phase. The baseline power exchange profiles of the energy resources can also be optimized to enhance the pool flexibility, and practical constraints such as time availability of the resources within the service window or network congestion constraints can be included. Notably, the resulting optimization problem is amenable for privacy preserving and scalable distributed resolution schemes.**

***Index Terms*— Balancing services; Aggregation; Power flexibility; Disaggregation policy.**

I. Introduction

According to the European Green Deal, the European Union aims at becoming net carbon neutral by 2050, [1]. This ambitious goal calls for a substantial increase in the share of renewable energy sources (RESs). However, integrating a large amount of RESs like solar and wind into the electrical supply poses challenges, particularly in terms of balancing services due to the intermittent nature of these sources. The distributed nature of RESs makes traditional centralized electrical plants less suitable for providing balancing services, leading to a policy shift promoting the use of distributed energy resources (DERs). In response to this shift, the Ancillary Service Market (ASM) has introduced products to encourage the participation of smaller producers, with a reduction of the minimum bid requirements, [2]. Aggregators play a key role in coordinating pools of prosumers (small-scale consumers and producers) to participate in the market, [3], [4]. However, assessing the flexibility of a pool of prosumers, i.e., computing the set of all the power trajectories that the resources in the pool can jointly provide to the grid along a given time horizon, is a challenging task, also because the flexibility of the pool must be represented by a set with a concise and compact description in order to reduce the complexity of planning, trading, and control by the aggregator and grid operator, [5].

Due to the inherent complexity of precisely computing the aggregated flexibility set [6], various approximation methods have been presented in the literature. We next briefly recall methods that search for an inner approximation of the flexibility set. Interested readers are referred to [7] for a thorough review.

Several methods address the approximation of the aggregated flexibility set by first finding an approximation of the polytopic flexibility set of each individual unit. Key to these methods is the selection of a suitable geometric representation for the individual sets to facilitate their subsequent aggregation. The method in [5] employs zonotopes, a subclass of polytopes, which can be easily aggregated if they share the same set of generators, which, however, may limit the quality of the approximation. The approach in [8], using unions of homothets of hyper-rectangles, decomposes the polytopic set and iteratively covers it for an accurate approximation. However, computation time increases exponentially with the coverage level, thus causing early stopping of the covering procedure and limiting the approximation accuracy. A general strength of these methods is the preservation of privacy since each unit can compute locally its flexibility set and communicate it to the aggregator without sharing any sensitive data, an essential aspect since prosumers are willing to offer flexibility but not necessarily to disclose private information related to their own use of local resources. A significant drawback of these methods is the introduction of conservatism, as the aggregated flexibility is not directly maximized but is instead computed by approximating the flexibility at the prosumer level, which also makes it difficult to incorporate congestion constraints within the distribution network, thus possibly resulting in an overestimation of the aggregated flexibility. In contrast to the above mentioned approaches, the approaches in [9]–[12] seek to maximize the flexibility directly over the aggregated space. [9] and [10] formulate the aggregation problem as a projection operation. In particular the aggregated flexibility is expressed as the projection of a higher-dimensional set defined by the prosumers' power trajectories. Notably, the resulting problem can be expressed as a robust optimization program, which, however, requires certain assumptions over the geometry of the aggregated set to be solvable. For instance, [9] fixes

*This work was supported by the Italian Ministry of University and Research (MUR) and the European Union (EU) under the PON/REACT project, and partially funded by the Research Fund for the Italian Electrical System under the Three-Year Research Plan 2022-2024 (DM MITE n. 337, 15.09.2022), in compliance with the Decree of April 16th, 2018, and by the PRIN PNRR project P2022NB77E "A data-driven cooperative framework for the management of distributed energy and water resources" (CUP: D53D23016100001), funded by the NextGeneration EU program (Mission 4, Component 2, Investment 1.1).

[1]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Italy. Email: {name.surname}@polimi.it

[2]Ricerca sul Sistema Energetico - RSE S.p.A. Email: federico.bianchi@rse-web.it

the structure of the aggregated flexibility by using a simple polytopic set, while [10] opts for an ellipsoidal set. These reformulations, however, do not allow to account for network congestion constraints.

More recent work in [11] and [12] retains the projection-based formulation while incorporating network congestion constraints through an additional adaptive step (adaptive robust optimization). This step ensures that the solution adheres to network congestion requirements while reducing conservatism of the solution.

Although the main focus in the literature has been assessing the overall flexibility of a pool of prosumers that are coordinated by an aggregator, it is important to note that the aggregator should be able to map any (admissible) power request by the grid back into the power exchange profile of each single resource of the pool (disaggregation policy), which is a demanding task especially in the context of the Balancing Market (BM), where short response times are required. Balancing service markets consist of an auction session where participants offer to the grid their availability to modify their power request within certain upward/downward limits during a given service window, and a subsequent real-time session where accepted participants must provide the actual service when receiving a power request from the grid within their declared limits [13]. Aggregators of energy resources offering balancing services then face the challenges of assessing the upward and downward power variations that the pool can withstand in each time slot of the service window for the auction session, and mapping admissible power requests by the grid back to each single energy resource for the real-time session.

Unfortunately, methods [11], [12] do not provide a disaggregation policy, thus requiring the additional step of defining the contribution of each prosumer as soon as the grid power request becomes available. An interesting approach that avoids this two-step procedure is proposed in [14], where a finitely parameterized disaggregation policy that depends linearly on the power profile requested by the grid is assumed and, accordingly, an equivalent battery model is derived to assess the flexibility of the aggregate. By jointly determining the aggregated feasible region and the dispatching policy, method [14] allows to promptly react to a grid balancing request. Although inspiring, [14] has some limitations: it assumes an initial energy condition equal to zero and symmetric bounds for the energy resources involved, which is hardly the case in practical problems, and it does not account for network constraints caused by limits on the capacity of the transmission lines. [15] proposes an approach to impose network constraints to the set derived in [14] while reducing conservatism. However, differently from [14], it does not provide the disaggregation policy.

Building upon [14], we proposed in [16] a framework for handling time-varying power and energy constraints while including arbitrary initial energy conditions. We use a hypercube with constant upward and downward power flexibility limits over all time slots in the service window to represent the aggregated flexibility and a linearly parameterized time-invariant disaggregation policy affine in the grid power request. We then show that the policy parameters can be derived by maximizing the volume of the aggregated flexibility set while considering global constraints like minimum service levels and network congestion limits and that the resulting problem is convex with a multi-agent constraint-coupled structure, thus allowing for a decentralized solution that is computationally tractable as the population grows while preserving information privacy. In [17] a disaggregation policy with a time-varying affine term is provided, which better copes with time variations of power and energy limits while accounting for general energy limits and initial state. However, the proposed method does not account for network constraints and is tailored to Thermostatically Controlled Loads (TCLs), which limits its application domain. The aggregator collects the TCL parameters from each prosumer and computes an average TCL model, whose flexibility polytopic set serves as a template for all prosumers to compute an inner approximation of their local flexibility set via a re-scaling scalar factor and a translation vector defining their disaggregation policy. Since the resulting local flexibility sets are homothets, the inner approximation of the aggregate flexibility set is easy to compute using the local re-scaling factors and translation vectors.

In this paper, we extend our previous work [16] by adding the following important features:

i) allow the disaggregation policy to have time-varying gains and affine terms, and the upward and downward power flexibility limits to differ over different time slots in the service window;
ii) include the possibility of computing policies with temporal dependencies, leveraging the knowledge of the system operator requested power profile some time before the time window when it needs to be actually delivered, or adapt the prosumers local power profile after the requested power profile has been delivered based on the actual service provided;
iii) enabling the optimization of the baseline exchange profile of each prosumer, which provides an additional degree of freedom to increase the overall amount of flexibility, while satisfying the local demand.

The proposed framework is suited for the aggregation of energy resources that can be modeled as polyhedral set in the power domain with possibly linear coupling constraints. This includes generators as well as storage-like units, such as batteries, thermostatically controlled loads, and electric vehicles, subject to constraints on power - both at the individual and aggregate levels - and energy as well as to service market requirements, such as a minimum amount of offered downward/upward flexibility and also symmetric/asymmetric bids. Possible applications include the provision of ancillary services to the grid through the control of air conditioning systems in commercial buildings [18], the coordinated charging of electric vehicle fleets [19], and the optimal scheduling of flexibility-oriented microgrids [20].

The paper is organized as follows. In Section II we introduce the addressed problem and discuss the challenges involved in its solution. In Section III, we propose an approximate resolution strategy where the flexibility set is inner-approximated by

a box and the disaggregation policy is finitely parameterized as a linear function of the power request by the grid. In Section IV, we illustrate how several practical cases can be captured by the proposed framework. In Section V we showcase the effectiveness of the framework via numerical simulations and in Section VI we draw some concluding remarks.

Notation: We denote by $\mathbb{N}$ the set of positive integers, by $\mathbb{R}$ the set of real numbers, and by $\mathbb{R}_+$ the set of non-negative real numbers. Given two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, we denote by $\mathcal{S}_1 \setminus \mathcal{S}_2$ the relative complement of $\mathcal{S}_2$ with respect to $\mathcal{S}_1$, i.e., the set of elements in $\mathcal{S}_1$ and not in $\mathcal{S}_2$. Given a vector $v \in \mathbb{R}^n$ and an index $j \in \{1,\dots,n\}$, we denote with $v(j)$ its j-th component, while if a set of indices $\mathcal{J} \subseteq \{1,\dots,n\}$ is given, we denote with $v[\mathcal{J}]$ the subvector containing the components of v corresponding to the indices in $\mathcal{J}$. Similarly, for a matrix $V \in \mathbb{R}^{m\times n}$ and two indices $r \in \{1,\dots,m\}$ and $c \in \{1,\dots,n\}$, we denote with $V(r,c)$ the element of V in the r-th row and c-th column, while if two set of indices $\mathcal{J}_r \subseteq \{1,\dots,m\}$ and $\mathcal{J}_c \subseteq \{1,\dots,n\}$ are given, we denote with $V[\mathcal{J}_r,\mathcal{J}_c]$ the submatrix of V obtained removing from V the rows with index $r \in \{1,\dots,m\} \setminus \mathcal{J}_r$ and the columns with index $c \in \{1,\dots,n\}\setminus\mathcal{J}_c$. For a vector or matrix, superscript $\cdot^\top$ denotes its transpose and $|\cdot|$ its component-wise absolute value. For any vector $v \in \mathbb{R}^n$, we denote with $\|v\|_\infty = \max_j |v(j)|$ its infinity norm and with $\mathrm{diag}(v)$ a diagonal matrix with the components of v on the main diagonal. The identity matrix of order n is denoted by I_n, while the vector in $\mathbb{R}^n$ with all ones or zeros is denoted as 1_n or 0_n, respectively, the subscript being omitted when the size n is clear from the context. Throughout the paper (in)equalities between vectors or between a vector and a scalar have to be intended as component-wise, with the scalar interpreted as a vector with all components identical and equal to that scalar.

II. The Aggregation/Disaggregation Problem

Let us consider a set $\mathcal{I} = \{1,\dots,N\}$ of N prosumers willing to cooperate to offer balancing services to the grid operator, possibly via a coordinating unit (the aggregator). The coordination problem is posed with reference to a discrete-time horizon $\mathcal{T} = \{1,\dots,M\}$ over which the prosumers can plan their electrical energy exchange with the main grid.

For each prosumer i, $i \in \mathcal{I}$, we denote with $p_i(k) \in \mathbb{R}$ its (average) power consumption (if $p_i(k) > 0$) or production (if $p_i(k) < 0$) within time slot k, $k \in \mathcal{T}$, and with $p_i \in \mathbb{R}^M$ the corresponding power exchange profile over the horizon $\mathcal{T}$. We assume that the set of admissible power profiles of prosumer $i \in \mathcal{I}$ can be described (see, e.g., [21]) via a polyhedron

$$\mathcal{P}_i = \{p_i \in \mathbb{R}^M : \; F_i p_i \le h_i\},$$

which possibly accounts for operating constraints, e.g., related to capacity limits of a storage device, final status of charge of an electric vehicle, comfort conditions of a thermostatically controlled load, etc.[1]

[1]Equality constraints can be accounted for by means of double-sided inequalities.

In balancing service provision, a power deviation profile $\delta \in \mathbb{R}^M$ from a nominal power exchange profile $\bar{p} \in \mathbb{R}^M$ is requested by the grid to the pool within horizon $\mathcal{T}$. It is thus convenient to decompose the power profile of prosumer i as

$$p_i = \bar{p}_i + \delta_i, \tag{1}$$

where $\bar{p}_i \in \mathbb{R}^M$ is its power exchange profile when no power deviation δ is requested by the grid, whereas $\delta_i \in \mathbb{R}^M$ is its contribution to the deviation profile δ, i.e.,

$$\textstyle\sum_{i\in\mathcal{I}} \delta_i = \delta. \tag{2}$$

In turn, $\bar{p}_i$ is conveniently expressed as the sum of two components

$$\bar{p}_i = g_i + a_i, \tag{3}$$

where $g_i \in \mathbb{R}^M$ and $a_i \in \mathbb{R}^M$ respectively represent the power exchanged with the grid and with the other prosumers within the aggregate so that we have

$$\textstyle\sum_{i\in\mathcal{I}} a_i = 0, \tag{4}$$

resulting into

$$\textstyle\sum_{i\in\mathcal{I}} \bar{p}_i = \sum_{i\in\mathcal{I}} g_i = \bar{p}. \tag{5}$$

Clearly, the service requested by the grid may involve only a subset $\mathcal{S} \subseteq \mathcal{T}$ of time slots, which we refer to as the *service window*. Consequently, the requested deviation is zero outside the service window, i.e.,

$$\delta(k) = 0, \qquad k \in \mathcal{S}^{\mathsf{c}} = \mathcal{T} \setminus \mathcal{S}. \tag{6}$$

Prosumer i may be connected to the grid and the other prosumers in a subset $\mathcal{C}_i \subseteq \mathcal{T}$ of time slots, which we refer to as *connection window*, and may allow for a deviation from its nominal profile only in a subset $\mathcal{F}_i \subseteq \mathcal{C}_i$ of its connection window, which we refer to as *flexibility window*. This is coded via the following constraints:

$$g_i(k) = a_i(k) = 0, \qquad k \in \mathcal{C}_i^{\mathsf{c}} = \mathcal{T} \setminus \mathcal{C}_i, \qquad i \in \mathcal{I}, \tag{7}$$

$$\delta_i(k) = 0, \qquad k \in \mathcal{F}_i^{\mathsf{c}} = \mathcal{T} \setminus \mathcal{F}_i, \qquad i \in \mathcal{I}. \tag{8}$$

Note that even if for a time slot k outside the service window $\mathcal{S}$ the grid request $\delta(k)$ is zero (cf. (6)), the prosumers within the pool can still have a non-zero deviation $\delta_i(k)$ from their nominal profile $\bar{p}_i$ within their flexibility window $\mathcal{F}_i$ so as to better prepare for or recover from a non-zero deviation request within the service window. Such a deviation must be coordinated with the other prosumers in the pool so that the aggregate deviation $\delta(k) = \sum_{i\in\mathcal{I}} \delta_i(k)$ is zero, or $k \in \mathcal{S}^{\mathsf{c}}$. Also, better preparing for a non-zero deviation request is possible only if the whole request is known before the flexibility window starts. We shall assume that this is the case without any loss of generality, since we can eventually set the flexibility window to start when we know the requested profile will be available, or, in the worst case, at the beginning of the service window, if the request were not available in advance, so that no anticipative action can be taken anyhow.

Given the nominal power profiles $\bar{p}_i = g_i + a_i$, $i \in \mathcal{I}$, of all prosumers, we can define the corresponding set of possible

power deviations that the prosumers pool can accommodate, which we refer to as the *flexibility set*, as

$$\Delta = \big\{\delta \in \mathbb{R}^M : \ \delta = \textstyle\sum_{i\in\mathcal{I}} \delta_i \\ (g_i + a_i + \delta_i \in \mathcal{P}_i, \ i \in \mathcal{I}) \wedge (6) \wedge (8)\big\}. \tag{9}$$

Note that the flexibility set Δ associated to $\bar{p}_i = g_i + a_i$, $i \in \mathcal{I}$, is a polyhedron given by the Minkowski sum over all prosumers of the polyhedral sets collecting their admissible deviations. Unfortunately, determining an explicit representation of Δ in terms of linear inequalities or vertices enumeration is computationally hard when N is large. For this reason several approaches aiming at inner approximating Δ with a simpler set have been proposed in the literature, [7].

Even more difficult is the problem of finding the nominal power profiles that optimize a certain feature of Δ (e.g., maximize its extent), which can be formalized as the following joint optimization problem

$$\begin{aligned} \max_{\{g_i,a_i\}_{i\in\mathcal{I}}} \quad & f(\Delta, g_1, \dots, g_N, a_1, \dots, a_N), \\ \text{subject to:} \quad & (4) \wedge (7) \end{aligned} \tag{10}$$

where $f(\cdot)$ is a suitably-defined performance criterion, possibly accounting for the different energy prices for exchanges with the grid and exchanges within the aggregate. Despite its inherent complexity, problem (10) plays a central role in ancillary service provision, e.g., in vehicle-to-grid optimal operation planning, [22].

Suppose now that $g_1^\circ, \dots, g_N^\circ$, $a_1^\circ, \dots, a_N^\circ$ is an optimal solution to (10) and Δ° the corresponding flexibility set computed according to (9) with $\bar{p}_i = g_i^\circ + a_i^\circ$, $i \in \mathcal{I}$. Then, if a (feasible) service $\delta = \tilde{\delta} \in \Delta^\circ$ is requested by the grid, it has to be mapped into a (feasible) variation $\tilde{\delta}_i$ for each prosumer $i \in \mathcal{I}$. This can be done by solving a second optimization problem

$$\begin{aligned} \min_{\{\delta_i\}_{i\in\mathcal{I}}} \quad & f_d(\delta_1, \dots, \delta_N) \\ \text{subject to:} \quad & \textstyle\sum_{i\in\mathcal{I}} \delta_i = \tilde{\delta} \\ & g_i^\circ + a_i^\circ + \delta_i \in \mathcal{P}_i \quad i \in \mathcal{I} \\ & (8) \end{aligned} \tag{11}$$

where a cost criterion $f_d(\cdot)$ can be used to promote specific disaggregation outcomes over others. If $f_d(\cdot)$ is "easy" to minimize (e.g., convex), then (11) is easy to solve. However, besides requiring a coordination among the prosumers so as to match the required service $\tilde{\delta}$, problem (11) needs to be solved online within the time interval between the time when the request by the grid operator is received and the time when the deviation profile has to be implemented.

Alternatively, one can interpret (11) as a multiparametric programming problem depending on $\tilde{\delta} \in \Delta^\circ$ and precompute the disaggregation policy $(\tilde{\delta}_1, \dots, \tilde{\delta}_N) : \Delta^\circ \to \mathbb{R}^{MN}$. In the case of a convex cost criterion $f_d(\cdot)$ that is linear or quadratic, the parametric solution is known to be piecewise affine and defined over a polyhedral partition of Δ°, whose cardinality grows exponentially with the number of prosumers, [23]. This perspective has the advantage of providing a policy that is readily available for online application. However, i) it requires to solve two optimization problems, one of which is parametric, ii) it provides to the grid a flexibility set where the admissible deviations $\delta(k)$, $k \in \mathcal{S}$, are correlated, and iii) the resulting disaggregation policy is quite complex since the polyhedral partition of Δ° has a cardinality that grows exponentially with the number of prosumers, [24].

In the next section we propose a method to approximately solve (10) and jointly obtain a disaggregation policy that is linear in the service request δ, by searching for a box-shaped inner approximation of Δ, while maximizing the performance index in (10) with respect to the nominal power exchange profiles.

III. Flexibility Assessment and Policy Design

In order to comply with the requirements of the energy service market, where the offered flexibility has to be given in terms of upward and downward power variations per time slot, we inner-approximate the flexibility set Δ with a box

$$\mathcal{B}_{c,d} = \{\delta \in \mathbb{R}^M : \ |\delta - c| \le d\} \subseteq \Delta \tag{12}$$

where vector $c \in \mathbb{R}^M$ parameterizes the center of the box and $d \in \mathbb{R}^M_+$ the box side half-lengths. Clearly, $d(k) > 0$ for all $k \in \mathcal{S}$ and $d(k) = 0$ for all $k \in \mathcal{S}^c$, so that we can accommodate any deviation $\delta(k) \in [c(k) - d(k), c(k) + d(k)]$ for any time slot $k \in \mathcal{S}$.

Similarly to past works [14], [16], we propose to fix the family of disaggregation policies a-priori. To this end, we impose the deviation profile of prosumer i, $i \in \mathcal{I}$, to be a linear function

$$\delta_i = K_i \delta \tag{13}$$

of the deviation profile δ requested by the grid. Note that this entails that when the grid does not request any deviation ($\delta = 0$), then, all prosumers do not modify their exchange profile with respect to the nominal one since $\delta_i = 0$, for all $i \in \mathcal{I}$. Differently from [14], [16], the gain K_i is not a scalar, but it is rather a matrix $K_i \in \mathbb{R}^{M\times M}$, whose sparsity pattern depends on the considered scenario (see Section IV-D for some examples). Using (13) inside (1) and (3) we obtain

$$\begin{aligned} p_i &= \bar{p}_i + K_i \delta \\ &= g_i + a_i + K_i \delta. \end{aligned} \tag{14}$$

Differently from [16], we jointly optimize the policy and the nominal profile.

We can now approximate (10) as

$$\begin{aligned} \max_{c,d,\{g_i,a_i,K_i\}_{i\in\mathcal{I}}} \quad & f_a(c, d, g_1, \dots, g_N, a_1, \dots, a_N) \\ \text{subject to:} \quad & (4) \wedge (7) \\ & \mathcal{B}_{c,d} \subseteq \Delta_K, \end{aligned} \tag{15}$$

where $f(\cdot)$ is replaced by its surrogate $f_a(\cdot)$ in which the dependence on Δ is through center and size of the box $\mathcal{B}_{c,d}$ defined in (12) and the flexibility set Δ in (9) is replaced by

$$\Delta_K = \big\{\delta \in \mathbb{R}^M : \ \delta = \textstyle\sum_{i\in\mathcal{I}} \delta_i \\ (g_i + a_i + \delta_i \in \mathcal{P}_i \wedge \delta_i = K_i\delta, \ i \in \mathcal{I}) \wedge (6) \wedge (8)\big\} \tag{16}$$

accounting for the parameterized policy (13). Note that, as a byproduct of solving (15), we obtain also a disaggregation policy and we thus do not need to solve (11) online nor its multiparametric version offline.

In the next theorem, we show that, under a mild assumption on $f_a(\cdot)$, problem (15) can be reformulated as a problem that is much easier to solve.

Theorem 1: Problem (15) is equivalent to

$$\begin{aligned} \max_{c,d,\{g_i,a_i,K_i\}_{i\in\mathcal{I}}} \quad & f_a(c,d,g_1,\dots,g_N,a_1,\dots,a_N) \qquad (17)\\ \text{subject to:} \quad & (4)\wedge(7)\\ & I_{\mathcal{S}}=\textstyle\sum_{i\in\mathcal{I}} K_i\\ & c[\mathcal{S}^{\mathsf{c}}]=0,\ d[\mathcal{S}^{\mathsf{c}}]=0,\ d[\mathcal{S}]>0\\ & K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]=0, \qquad i\in\mathcal{I}\\ & F_i(g_i+a_i+K_ic)+|F_iK_i|d\le h_i, \quad i\in\mathcal{I}, \end{aligned}$$

where $I_{\mathcal{S}}=\operatorname{diag}(e_{\mathcal{S}})\in\mathbb{R}^{M\times M}$ with $e_{\mathcal{S}}\in\mathbb{R}^M$ such that $e_{\mathcal{S}}[\mathcal{S}]=1$ and $e_{\mathcal{S}}[\mathcal{S}^{\mathsf{c}}]=0$. Furthermore, if $f_a(\cdot)$ does not depend on $a_1,\dots,a_N$, then, problem (17) is equivalent to

$$\begin{aligned} \max_{c,d,\{g_i,\gamma_i,G_i\}_{i\in\mathcal{I}}} \quad & f_a(c,d,g_1,\dots,g_N) \qquad (18)\\ \text{subject to:} \quad & c=\textstyle\sum_{i\in\mathcal{I}}\gamma_i,\ \operatorname{diag}(d)=\sum_{i\in\mathcal{I}} G_i\\ & c[\mathcal{S}^{\mathsf{c}}]=0,\ d[\mathcal{S}^{\mathsf{c}}]=0,\ d[\mathcal{S}]>0\\ & g_i[\mathcal{C}_i^{\mathsf{c}}]=0,\ \gamma_i[\mathcal{C}_i^{\mathsf{c}}]=0, \qquad i\in\mathcal{I}\\ & G_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]=0, \qquad i\in\mathcal{I}\\ & F_i(g_i+\gamma_i)+|F_iG_i|1_M\le h_i, \qquad i\in\mathcal{I} \end{aligned}$$

where $\gamma_i=a_i+K_ic$ and $G_i=K_i\operatorname{diag}(d)$.
In particular, if $c^\star$, $d^\star$, and $\{g_i^\star,\gamma_i^\star,G_i^\star\}_{i\in\mathcal{I}}$ is an optimal solution to (18), then the corresponding optimizer for (17) (and thus (15)) is given by $c^\star$, $d^\star$, and $\{g_i^\star,a_i^\star,K_i^\star\}_{i\in\mathcal{I}}$ where $K_i^\star=G_i^\star\operatorname{diag}(d^\dagger)$ and $a_i^\star=\gamma_i^\star-K_i^\star c^\star$, with $d^\dagger\in\mathbb{R}_+^M$ such that $d^\dagger(k)=1/d^\star(k)$ for $k\in\mathcal{S}$ and $d^\dagger(k)=0$ for $k\in\mathcal{S}^{\mathsf{c}}$.

Proof: See Appendix A. ∎

According to Theorem 1, problem (15) thus admits a reformulation with a convex feasible set. If the cost function $f_a(\cdot)$ is concave, then (15) admits a reformulation as a convex problem, which can be easily solved with off-the-shelf solvers. Additionally, note that (18) can be regarded as a multi-agent optimization problem where the agents are the prosumers (in charge of optimizing variables g_i, γ_i, and G_i) and the aggregator (in charge of optimizing variables c and d), with decisions coupled by the constraints $c=\sum_{i\in\mathcal{I}}\gamma_i$ and $\operatorname{diag}(d)=\sum_{i\in\mathcal{I}}G_i$. Therefore, if the cost function $f_a(\cdot)$ is separable (see Section IV-B for some examples), then problem (18) can be solved by means of a distributed optimization algorithm capable of handling such a structure, e.g., [25]–[28], so as to ease the computation of the solution and preserve privacy of the local prosumers' information.

Since the change of variables involves a_i and K_i quantities only (which are mapped into γ_i and G_i), any additional constraint on c, d, and g_i can be easily accounted for by adding it to both (15) and (18) without any effect on the validity of Theorem 1. Additional constraints linking a subset of the p_i's can also be accounted for, leveraging the same idea behind the proof of Theorem 1. Examples of both kind of constraints are reported in Section IV-C.

Lastly, the reader should note that the set $\mathcal{P}_i$ may be the projection onto the p_i-subspace of a higher-dimensional polyhedron describing the prosumer internal behavior, for example when the power p_i is not the only decision variable. If the relation between p_i and other variables is still linear, then we can simply add these decision variables to all optimization problems and all derivations will carry through.

IV. Practical Cases

In this section we show how practical scenarios in ancillary service provision fit our framework so that the solution described in Section III can be applied.

A. Prosumers

The methodology developed in Section III applies to the case when the flexibility set of each prosumer is given by a polyhedron $\mathcal{P}_i$. As in [21] it has been suggested that prosumers are equipped with different types of devices that can be modeled as resource polytopes, in the following, we briefly review resource polytopes modeling devices and show how to translate them into linear inequalities to be used in the definition of $\mathcal{P}_i$.

Dispatchable generator: the power $p_i(k)\le 0$ of the generator is constrained by

$$-u_{p,i}(k)\le p_i(k)\le 0, \qquad (19)$$

where $u_{p,i}(k)>0$ represents the maximum amount of power that the generator can deliver, which can possibly be time-varying within the time horizon $\mathcal{T}$.

Ramp-rate constraints limiting the rate at which the generator power can vary are given by

$$\ell_{r,i}(k)\le p_i(k)-p_i(k-1)\le u_{r,i}(k), \qquad (20)$$

with lower and upper bounds that may also vary along $\mathcal{T}$.

Storage-like units, including batteries and thermostatically controlled loads (TCLs, see [14]): the power p_i exchanged by the storage (positive if injected in the device, negative if withdrawn from it) is bounded

$$\ell_{p,i}(k)\le p_i(k)\le u_{p,i}(k), \qquad (21)$$

and (possibly) subject to rate constraints like (20). Further constraints on p_i arise due to capacity limits. Indeed, a storage-like device is subject to constraints on the amount of energy $e_i(k)$ that can be stored at the beginning of any time slot k, which can be expressed as

$$\ell_{e,i}(k)\le e_i(k)\le u_{e,i}(k), \qquad (22)$$

and can also encompass constraints on a minimum/maximum desired energy level at specific time slots. Note that in the case of a TCL, $e_i(k)$ represents thermal energy converted into electrical energy through the TCL coefficient of performance and capacity constraints model constraints on the TCL temperature (cf. [29, Section 3.3.1]).

The evolution of $e_i(k)$ can be described by a recursive equation of the following form

$$e_i(k+1) = \zeta_i e_i(k) + b_{p,i} p_i(k) + b_{d,i} d_{x,i}(k), \tag{23}$$

where $\zeta_i \in (0,1]$ represents a self-discharge coefficient that models the dissipation of energy, while $b_{p,i}$ and $b_{d,i}$ are conversion efficiencies of the power $p_i(k)$ and a possible further exogenous input $d_{x,i}(k)$ into the energy state. This entails that $e_i(k)$ can be expressed as:

$$e_i(k) = \zeta_i^{k-k_{s,i}} e_i(k_{s,i}) + b_{p,i} \sum_{s=k_{s,i}}^{k-1} \zeta_i^{k-1-s} p_i(s) + b_{d,i} \sum_{s=k_{s,i}}^{k-1} \zeta_i^{k-1-s} d_{x,i}(s),$$

$k_{s,i}$ being the first time slot in $\mathcal{C}_i$ and $e_i(k_{s,i})$ the storage energy content at connection. We can then translate (22) into a constraint on $p_i(s)$, $s = k_{s,i}, \dots, k-1$, as follows

$$\tilde{\ell}_{e,i}(k) \le b_{p,i} \sum_{s=k_{s,i}}^{k-1} \zeta_i^{k-1-s} p_i(s) \le \tilde{u}_{e,i}(k), \tag{24}$$

with:

$$\tilde{\ell}_{e,i}(k) = \ell_{e,i}(k) - \sum_{s=k_s}^{k-1} \zeta^{k-1-s} b_d d_{x,i}(s) - \zeta_i^{k-k_{s,i}} e_i(k_{s,i})$$

$$\tilde{u}_{e,i}(k) = u_{e,i}(k) - \sum_{s=k_s}^{k-1} \zeta^{k-1-s} b_d d_{x,i}(s) - \zeta_i^{k-k_{s,i}} e_i(k_{s,i})$$

Since constraints (19), (20), (21), and (24) are all linear in p_i, they are encompassed by $\mathcal{P}_i = \{p_i \in \mathbb{R}^M : \ F_i p_i \le h_i\}$ by suitably defining matrix F_i and vector h_i based on τ, ζ_i, $\ell_{p,i}$, $u_{p,i}$, $\ell_{r,i}$, $u_{r,i}$, $\tilde{\ell}_{e,i}$, $\tilde{u}_{e,i}$, and $k_{s,i}$.

B. Cost Function

Different cost functions $f_a(\cdot)$ model different objectives and clearly lead to different flexibility sets and disaggregation policies. The expression of $f_a(\cdot)$ should obviously reflect the needs of the specific application at hand. Here, we suggest some practical choices.

A first possible objective for the aggregator is to maximize the set of profiles δ that can be accommodated. This can be encoded using as cost function $f_a(\cdot)$ the volume of the box $\mathcal{B}_{c,d}$, which is given by the product of the length of all edges in the service window

$$\mathrm{vol}(\mathcal{B}_{c,d}) = \prod_{k \in \mathcal{S}} 2\, d(k).$$

Taking the logarithm of $\mathrm{vol}(\mathcal{B}_{c,d})$

$$\log\big(\mathrm{vol}(\mathcal{B}_{c,d})\big) = \sum_{k \in \mathcal{S}} \log(2\, d(k)) \tag{25}$$

we obtain a cost that is strictly concave in the decision variable d. Alternatively, one can use the sum of the lengths of the edges

$$\sum_{k \in \mathcal{S}} 2\, d(k) \tag{26}$$

as a different measure for the extent of $\mathcal{B}_{c,d}$, to keep the problem linear. If the aggregator is interested in promoting symmetric upward and downward services, then it can encourage the center of $\mathcal{B}_{c,d}$ to be close to the origin with the term

$$-\|c\|_2^2, \tag{27}$$

which is strictly concave in c.

If the revenues of the aggregator depend on how much flexibility it provides in different time slots, then we can modify (26) as follows

$$\sum_{k \in \mathcal{S}} 2\, \rho_k d(k), \tag{28}$$

where $\rho_k \ge 0$ is the aggregator revenue per unit of flexibility offered in time slot k. If we further want to differentiate between amounts of downward ($\delta(k) > 0$) and upward ($\delta(k) < 0$) flexibility offered, then we can use

$$\sum_{k \in \mathcal{S}} \rho_k^+ \max\{0, c(k)+d(k)\} - \rho_k^- \min\{0, c(k)-d(k)\}, \tag{29}$$

where $\rho_k^+ \ge 0$ and $\rho_k^- \ge 0$ are the aggregator revenue per unit of flexibility offered downward and upward, respectively, in time slot k. Even if (29) is not necessarily concave in c and d, it simplifies to the concave expression

$$\sum_{k \in \mathcal{S}} \rho_k^+ (c(k)+d(k)) - \rho_k^- (c(k)-d(k)) \tag{30}$$

in the case when the interval $[c(k)-d(k), c(k)+d(k)]$ contains $\delta(k) = 0$, for all $k \in \mathcal{S}$, which is always the case because it is sensible to ensure that a zero-deviation request is admissible for the problem at hand (see Section IV-C).

The cost function $f_a(\cdot)$ can be set to any positive combination of the terms in (25), (26), (27), (28), (29), and (30).

C. Additional Constraints

The framework proposed in Section III can be easily extended to account for application-specific requirements. We next cover some useful constraints that can be added to problems (15) and (18) without invalidating the result of Theorem 1.

To ensure that it is always possible to stick to the nominal profile, we can enforce $0 \in \mathcal{B}_{c,d}$ by imposing that

$$c[\mathcal{S}] + d[\mathcal{S}] \ge 0 \tag{31a}$$

$$c[\mathcal{S}] - d[\mathcal{S}] \le 0. \tag{31b}$$

In many cases in order to access the balancing service market, a minimum amount of downward/upward flexibility has to be offered. This requirement can be easily enforced via the constraints

$$c[\mathcal{S}] + d[\mathcal{S}] \ge u_{\min}[\mathcal{S}], \tag{32a}$$

$$c[\mathcal{S}] - d[\mathcal{S}] \le \ell_{\min}[\mathcal{S}], \tag{32b}$$

where $u_{\min}(k) \ge 0$ and $\ell_{\min}(k) \le 0$ respectively represent the minimum level of downward and upward flexibility that must be offered within time slot $k \in \mathcal{S}$. In this case, constraints (31) are redundant.

If instead the aggregator is forced to provide a symmetric flexibility around its baseline profile $\bar{p}$, then this can be enforced imposing

$$c[\mathcal{S}] = 0. \tag{33}$$

On the contrary, if the aggregator is interested in bidding asymmetrically for either a downward or upward service only in a specific time slot k, it can impose either of the following constraint

$$c(k) - d(k) \geq 0, \tag{34a}$$
$$c(k) + d(k) \leq 0, \tag{34b}$$

respectively, for some $k \in \mathcal{S}$.

Lastly, if a subset $\mathcal{N}_j \subseteq \mathcal{I}$ of prosumers is subject to technical limits on their cumulative power exchange profile, then we must enforce

$$\ell_n^{(j)} \leq \min_{\delta \in \mathcal{B}_{c,d}} \textstyle\sum_{i \in \mathcal{N}_j} p_i \leq \max_{\delta \in \mathcal{B}_{c,d}} \textstyle\sum_{i \in \mathcal{N}_j} p_i \leq u_n^{(j)}, \tag{35}$$

where $u_n^{(j)}(k)$ and $\ell_n^{(j)}(k)$ are the maximum power that the j-th subset of prosumers $\mathcal{N}_j$ can absorb from and inject into the grid during time slot k, respectively, thus modeling congestion constraints. If we now exploit the fact that $|\sum_{i \in \mathcal{N}_j} K_i| d = |\sum_{i \in \mathcal{N}_j} K_i \operatorname{diag}(d)| 1_M$ since $d \geq 0$, by (14) and (42) in the Appendix, constraint (35) is equivalent to enforcing

$$\textstyle\sum_{i \in \mathcal{N}_j} (g_i + a_i + K_i c) + \left|\sum_{i \in \mathcal{N}_j} K_i \operatorname{diag}(d)\right| 1_M \leq u_n^{(j)}$$
$$\textstyle\sum_{i \in \mathcal{N}_j} (g_i + a_i + K_i c) - \left|\sum_{i \in \mathcal{N}_j} K_i \operatorname{diag}(d)\right| 1_M \geq \ell_n^{(j)}$$

in problem (15). Leveraging the change of variables $G_i = K_i \operatorname{diag}(d)$ and $\gamma_i = a_i + K_i c$ used in Theorem 1, the previous constraints are equivalent to

$$\textstyle\sum_{i \in \mathcal{N}_j} (g_i + \gamma_i) + \left|\sum_{i \in \mathcal{N}_j} G_i\right| 1_M \leq u_n^{(j)} \tag{36a}$$
$$\textstyle\sum_{i \in \mathcal{N}_j} (g_i + \gamma_i) - \left|\sum_{i \in \mathcal{N}_j} G_i\right| 1_M \geq \ell_n^{(j)} \tag{36b}$$

to be added to problem (18) for as many subset $\mathcal{N}_j$ as needed. Note that, if the whole aggregate power profile is subject to technical limits, then its is sufficient to enforce (36) with $\mathcal{N}_j = \mathcal{I}$, which, using $c = \sum_{i \in \mathcal{I}} \gamma_i$ and $\operatorname{diag}(d) = \sum_{i \in \mathcal{I}} G_i$ from (18), simplifies into

$$\textstyle\sum_{i \in \mathcal{I}} g_i + c + d \leq u_n \tag{37a}$$
$$\textstyle\sum_{i \in \mathcal{I}} g_i + c - d \geq \ell_n \tag{37b}$$

to be added to problem (18), where $u_n(k)$ and $\ell_n(k)$ are the maximum power that the whole aggregate can absorb from and inject into the grid during time slot k, respectively.

Clearly, any combination of constraints (32), (33), (34), (36), and (37) is allowed except for (32a) with (34b) and $u_{\min}(k) > 0$ for some $k \in \mathcal{S}$ and for (32b) with (34a) and $\ell_{\min}(k) < 0$ for some $k \in \mathcal{S}$, because they would be mutually exclusive.

D. Policy Structure

As we mentioned in Section III, the sparsity pattern of the disaggregation policy matrix K_i of prosumer i depends on the considered scenario. We shall describe next three different types of increasingly restrictive policies.

The less restrictive one is the *proactive policy*, which assumes that the actual deviation profile δ that the grid requests to the prosumers pool is known before the flexibility window $\mathcal{F}_i$ of prosumer i begins. In this case, prosumer i is allowed to deviate from its nominal profile (i.e., have $\delta_i(k) \neq 0$ for some $k \in \mathcal{F}_i \setminus \mathcal{S}$) to prepare for the upcoming request. The proactive policy is already captured by the formulation in Section III and requires no modifications.

A more restrictive case is the *reactive policy*, which assumes that the actual deviation $\delta(k)$ that the grid requests to the prosumers pool in time slot k is known to prosumer i at the beginning of time slot k only and not before. In this case, prosumer i is not allowed to deviate from its nominal profile to prepare for the upcoming request simply because it does not know the request in advance, but it is still allowed to deviate from its nominal profile to recover from the request after the service window. This scenario forces $K_i(j_r, j_c) = 0$ for all (j_r, j_c) such that $j_c \in \mathcal{S}$ and $j_r = \{1, \dots, j_c - 1\}$ and, hence, a specific sparsity pattern for $K_i[\mathcal{T}, \mathcal{S}]$.

An even more restrictive case is the *greedy policy*, which forces the deviation $\delta_i(k)$ of prosumer i to be a function of $\delta(k)$ only. This implies that prosumer i is not allowed to deviate from its nominal profile before and after the service window but has the benefit of having a simple policy to optimize. In this case, $K_i(j_r, j_c) = 0$ for all (j_r, j_c) such that $j_c \in \mathcal{S}$ and $j_r \neq j_c$ and, hence, a specific sparsity pattern for $K_i[\mathcal{T}, \mathcal{S}]$.

Note that, since according to Theorem 1, matrix K_i in (15) and matrix G_i in (18) are related by $G_i = K_i \operatorname{diag}(d)$, their corresponding submatrices $K_i[\mathcal{T}, \mathcal{S}]$ and $G_i[\mathcal{T}, \mathcal{S}]$ are related by $G_i[\mathcal{T}, \mathcal{S}] = K_i[\mathcal{T}, \mathcal{S}] \operatorname{diag}(d[\mathcal{S}])$ with $d[\mathcal{S}] > 0$. As a consequence, any constraint on the sparsity pattern of $K_i[\mathcal{T}, \mathcal{S}]$ straightforwardly translates into a constraint on the sparsity pattern of $G_i[\mathcal{T}, \mathcal{S}]$.

E. Policy Optimization

As shown in Theorem 1, for problem (15) and (18) to be equivalent, the cost function $f_a(\cdot)$ is not allowed to depend on a_i or K_i, $i \in \mathcal{I}$. Therefore, even if some possible cost functions introduced in Section IV-B are strictly concave in c and d (thus resulting in a unique optimal flexibility box $\mathcal{B}_{c^\star, d^\star}$), there might be different combinations of a_i and K_i yielding the same box $\mathcal{B}_{c^\star, d^\star}$.

To break the tie we can look for the best policy g_i, a_i, and K_i, $i \in \mathcal{I}$, according to some cost criterion $f_p(g_1, a_1, K_1, \dots, g_N, a_N, K_N)$, among those yielding the box $\mathcal{B}_{c^\star, d^\star}$. This can be achieved by solving the following optimization problem

$$\begin{aligned} \min_{\{g_i, a_i, K_i\}_{i \in \mathcal{I}}} \quad & f_p(g_1, a_1, K_1, \dots, g_N, a_N, K_N) \\ \text{subject to:} \quad & (4) \wedge (7) \\ & \mathcal{B}_{c^\star, d^\star} \subseteq \Delta_K, \end{aligned} \tag{38}$$

which is equal to (15) but with $c = c^\star$, $d = d^\star > 0$ and a different cost function $f_p(\cdot)$ in place of $f_a(\cdot)$. Therefore, by the first part of Theorem 1, (38) is equivalent to (17) with

$c = c^\star$, $d = d^\star > 0$ and $f_p(\cdot)$ in place of $f_a(\cdot)$, i.e.,

$$\begin{aligned} \min_{\{g_i,a_i,K_i\}_{i\in\mathcal{I}}} \quad & f_p(g_1, a_1, K_1, \ldots, g_N, a_N, K_N) \\ \text{subject to:} \quad & 0 = \textstyle\sum_{i\in\mathcal{I}} a_i,\ I_{\mathcal{S}} = \sum_{i\in\mathcal{I}} K_i \\ & g_i[\mathcal{C}_i^{\mathsf{c}}] = 0,\ a_i[\mathcal{C}_i^{\mathsf{c}}] = 0, && i \in \mathcal{I} \\ & K_i[\mathcal{F}_i^{\mathsf{c}}, \mathcal{S}] = 0, && i \in \mathcal{I} \\ & F_i(g_i + a_i + K_i c^\star) + |F_i K_i| d^\star \le h_i, && i \in \mathcal{I}, \end{aligned} \tag{39}$$

where we made explicit constraints (4) and (7) and we neglected constraints on c and d only, since they are already satisfied by $c^\star$ and $d^\star > 0$.

Note that, whenever $f_p(\cdot)$ is convex in the decision variables, problem (39) is convex and easy to solve with available solvers.

V. Simulation results

We next present two case studies to showcase the effectiveness of our approach via simulation. The first one refers to the aggregation of a pool of TCLs. This allows to compare our method with that in [17] which is tailored to TCL energy resources. We also make a comparison with the method in [16], which can be viewed as an instance of our approach by setting $\mathcal{C}_i = \mathcal{F}_i = \mathcal{S} = \mathcal{T}$, making G_i a diagonal matrix with equal diagonal entries, forcing vectors γ_i, c, and d to have all identical components, and fixing g_i. Lastly, we also comment on the differences between our approach and the one in [8].

The second case study is a comprehensive example featuring a heterogeneous pool of prosumers including air conditioners (ACs), refrigerators (RFs), electric vehicles (EVs), electrical batteries (EBs), and power generators (PGs), which is instrumental to show the versatility of our method and its capability to cope with various settings and requirements including time window mismatches, final energy condition requirements, and network constraints. In our experiments, we adopt populations of energy resources with parameters extracted at random from intervals that reflect a realistic setting and are taken from [30] for ACs and RFs and from [31], [32] for EVs, EBs, and PGs.

A. First case study

We consider a horizon $\mathcal{T}$ of $M = 24$ time slots of $\tau = 15$ minutes duration and a service time window $\mathcal{S} = \mathcal{T}$ for a pool $\mathcal{I}$ of $N = 100$ residential ACs, which can be modeled as energy storage systems as mentioned in Section IV-A and further detailed in [29, Section 3.3.1]. Since [17] does not consider any misalignment between service, connection, and flexibility windows, we restrict our setup to the case when $\mathcal{C}_i = \mathcal{F}_i = \mathcal{S}$, $i \in \mathcal{I}$. We assume that each AC i, $i \in \mathcal{I}$, has a constant temperature set-point $\theta_{s,i}$ and an admissible temperature deviation $\Delta_{s,i}$ with respect to $\theta_{s,i}$ that is also constant along $\mathcal{T}$. The values of $\theta_{s,i}$ and $\Delta_{s,i}$ as well as of the further TCL parameters (thermal resistance $R_{th,i}$, thermal capacitance $C_{th,i}$, coefficient of performance η_i, and rated power $P_{r,i}$) involved in the TCL model (see [29, Section 3.3.1]) belong to the ranges specified in Table II. For simplicity, we assume a constant ambient temperature $\theta_a = 33$ °C throughout the time horizon $\mathcal{T}$ for the whole pool. The pool of N ACs is then generated by random extraction from the discrete population of AC types in Table III. Note that the electrical parameters of the AC types reported in that table are derived from the thermal parameters according to the modeling in [29, Section 3.3.1]. For each AC, the value of the initial energy $e_i(0)$ is extracted uniformly at random from the interval $[\ell_{e,i}, u_{e,i}]$.

Since in the method in [17] the power exchanged with the grid by each TCL is not optimized, we fix it also for our approach and set $g_i = P_{0,i}$, $i \in \mathcal{I}$, where $P_{0,i}$ is the power to maintain the temperature at the extracted constant set-point value $\theta_{s,i}$ (see [29, Section 3.3.1] for the details).

In [17] the flexibility set is described as a polytope in halfspace representation $\mathcal{P} = \{\delta \in \mathbb{R}^M : F\delta \le h\}$. Set $\mathcal{P}$ can be inner approximated by a maximum volume box $\mathcal{B} = \{\delta \in \mathbb{R}^M : \ell \le \delta \le u\}$ defined by the upward and downward margins $\ell \in \mathbb{R}^M$ and $u \in \mathbb{R}^M$ solving the following optimization problem

$$\begin{aligned} \max_{u,\ell} \quad & \sum_{k\in\mathcal{S}} \log(u(k) - \ell(k)) \\ \text{subject to:} \quad & \max\{F, 0\}u - \max\{-F, 0\}\ell \le h \\ & \ell \le 0 \\ & u \ge 0. \end{aligned} \tag{40}$$

Note that the first constraint represents the containment condition $\mathcal{B} \subseteq \mathcal{P}$, while the remaining ones ensure that $0 \in \mathcal{B}$ so that the aggregator can maintain the baseline profile if there is no power request from the grid. As for our method, we directly determine the maximum volume box by solving the following optimization problem:

$$\begin{aligned} \max_{c,d,\{g_i,\gamma_i,G_i\}_{i\in\mathcal{I}}} \quad & \sum_{k\in\mathcal{S}} \log(2\, d(k)) \\ \text{subject to:} \quad & c = \textstyle\sum_{i\in\mathcal{I}} \gamma_i,\ \mathrm{diag}(d) = \sum_{i\in\mathcal{I}} G_i \\ & c[\mathcal{S}^{\mathsf{c}}] = 0,\ d[\mathcal{S}^{\mathsf{c}}] = 0,\ d[\mathcal{S}] > 0 \\ & g_i[\mathcal{C}_i^{\mathsf{c}}] = 0,\ \gamma_i[\mathcal{C}_i^{\mathsf{c}}] = 0, && i \in \mathcal{I} \\ & G_i[\mathcal{F}_i^{\mathsf{c}}, \mathcal{S}] = 0, && i \in \mathcal{I} \\ & F_i(g_i + \gamma_i) + |F_i G_i| 1_M \le h_i, && i \in \mathcal{I}, \\ & c[\mathcal{S}] + d[\mathcal{S}] \ge u_{\min}[\mathcal{S}], \\ & c[\mathcal{S}] - d[\mathcal{S}] \le \ell_{\min}[\mathcal{S}], \end{aligned} \tag{41}$$

with $\ell_{min} = u_{min} = 0$, so that $0 \in \mathcal{B}_{c,d}$.

In Figure 1 we report the upward and downward flexibility margins of the pool obtained with the method in [17] (yellow dashed lines) and our method with the constant policy in our earlier paper [16] (green solid lines), greedy (red dotted lines), and reactive policies (turquoise dotted lines). It can be observed that for most of the time slots, the margins obtained with the method in [17] (yellow dashed lines) are contained within those obtained with our method using the three different policy variants. In order to better compare the performance of the methods, we report in Table I the ratio between the volume of the box obtained with our method and the volume of the box obtained with [17]. Note that the volume ratio obtained is always above one and gradually increases as more degrees of freedom are used in our policy.

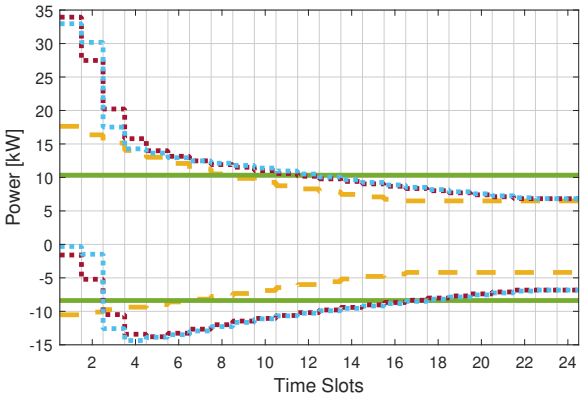


Fig. 1: Power flexibility margins of the pool along the 24 time slots using the method in [17] (yellow dashed lines), and the proposed method with: the constant policy of [16] (green solid lines), the greedy policy (red dotted lines), and the reactive policy (turquoise dotted lines).

We next examine the contribution of the a_i component of the baseline power profile $\bar{p}_i$ in (3), which provides a zero-net contribution to the overall power exchange with the grid according to the constraint (4) and can be used to increase the flexibility of the pool. To this purpose, consider the baseline power profile of one of the ACs and the corresponding energy evolution of its equivalent battery model, respectively reported in the left and right plots of Figure 2. Note that the initial energy condition is very close to the upper capacity limit, which hampers the AC in providing downward flexibility, as it is able to absorb less power (give away stored energy, upward service) but not extra power (store more energy, downward service). The a_i contribution to $\bar{p}_i$ then makes the energy content evolution (solid blue line) towards half capacity more rapid than with the g_i contribution only (dashed magenta line), so as to enable the exploitation of the full power range.

If we optimize jointly g_i and a_i, we do not get any significant improvement in terms of cost in the considered set-up. This is probably due to the fact that the distribution of the initial energy content throughout the population can support the net-zero internal exchanges a_i, $i \in \mathcal{I}$, so as to allow for the cost maximization despite g_i being fixed. If, instead, we impose that a final temperature condition $\theta_i(24) \leq \theta_{s,i}$, $i \in \mathcal{I}$, has to be reached, this limits the amount of net-zero internal exchanges because part of the initial energy content is used to enforce this new constraint, and we then obtain that optimizing the energy grid exchange g_i jointly with a_i, $i \in \mathcal{I}$, does provide a cost improvement. In fact, by optimizing g_i, the volume of the flexibility box increases of a factor 1.7 and 4 for the greedy and reactive policy, respectively.

TABLE I: Comparative analysis with respect to [17]: ratio between the obtained flexibility boxes for different policies.

Policy	Volume ratio
Constant, [16]	2.13×10^2
Greedy	15.93×10^2
Reactive	18.74×10^2

In order to understand the reason why our method is better performing than the method in [17], we now consider the case of the flexibility offered by two ACs over a two-slot time horizon. Figure 3 reports the exact flexibility set (black solid lines), the flexibility sets obtained with [17] (yellow solid lines), the flexibility sets obtained with our greedy policy (red dashed lines), and those obtained with our reactive policy (turquoise dotted lines), for the aggregator (Figure 3c) and, using the appropriate disaggregation strategy, for the two ACs (Figures 3a and 3b). Firstly, note that while [17] dilates the prototype set at each AC level as much as possible (yellow lines are touching the black lines in Figures 3a and 3b), it falls short in fully dilating the aggregated set (yellow line is *not* touching the black line in Figure 3c). The maximum volume box inscribed the aggregate flexibility set returned by [17] (reported with a green dotted line in Figure 3c) is therefore smaller than the optimal one (indeed, it is *not* touching the black line). As for our method with the greedy policy, the resulting aggregate flexibility box is also not optimal in terms of volume (red dashed line is *not* touching the black lines in Figure 3c), but it is nonetheless bigger than the one obtained using [17]. It is worth highlighting that by using the greedy policy, the resulting flexibility set of the second AC is lower dimensional and contracted in the first dimension. In practice, this means that the aggregator found it optimal to use the power flexibility of AC 2 only during the second time slot. The fact that the box in the AC 2 space has zero volume is *not* a contradiction since our approach aims at maximizing the volume directly in the aggregator space and not in each prosumer spaces. Notably, the reactive policy computes two non-axis aligned zonotopes (turquoise dotted lines in Figures 3a and 3b) that effectively maximize the aggregated flexibility box (turquoise dotted line in Figure 3c *does* touch the black line of the exact aggregate flexibility set). In fact, the maximum box inscribed in the exact polytopic flexibility set of Figure 3c corresponds to the same box found by the reactive policy, proving that more flexibility can be caught by considering a policy that accounts for past power requests from the grid.

This simple example also shows a major drawback of the approach in [8], which presents a method for approximating the aggregated flexibility set using a union of boxes. In order for the resulting aggregated set to be a box (to comply with balancing market regulations), we restrict the method in [8] to what the authors refer to as "Stage 0". In this stage, the method first computes the maximum volume box within each prosumer space and then returns their Minkowski sum, which is easy to compute for boxes, and is guaranteed to be contained within the aggregate flexibility set. However, this approach may lead to a suboptimal box, as maximizing the volume of each individual box in the prosumer space does *not* necessarily correspond to maximize the volume of the box within the aggregated space. This is very well illustrated by the results obtained by our greedy policy in Figure 3 (red dashed lines). Due to its structure, the greedy policy aims at maximizing the volume of the box in the aggregated space by fitting boxes

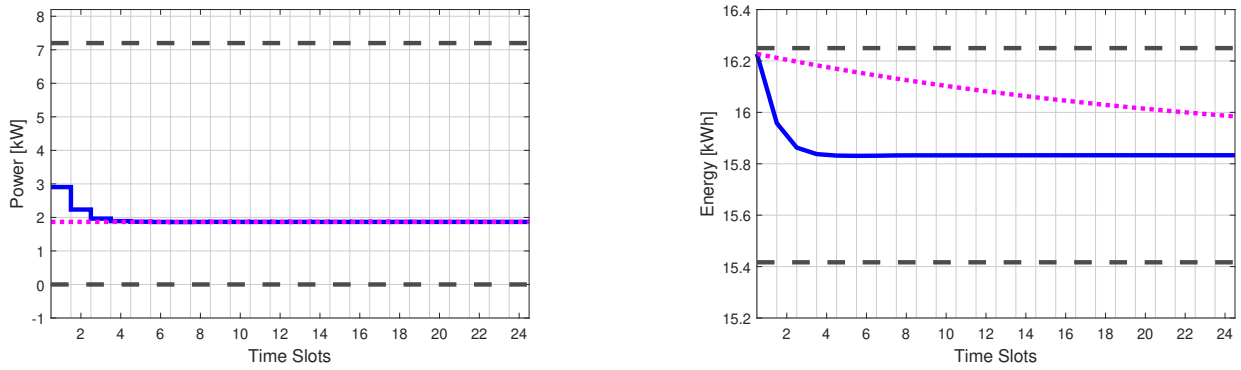
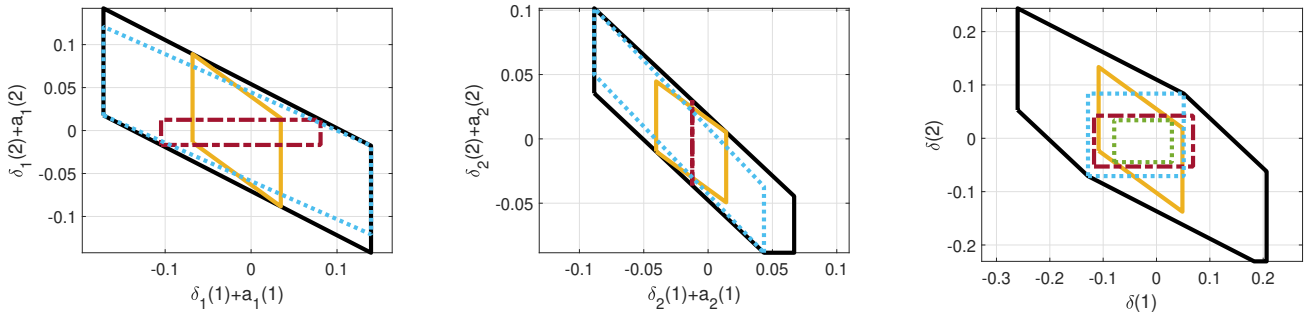


Fig. 2: Power (left plot) and energy (right plot) evolution for the battery equivalent model of one AC with: $\bar{p}_i = g_i$ (dotted magenta line) and $\bar{p}_i = g_i + a_i$ (solid blue line). Black dashed lines represents power/energy upper and lower limits.



(a) AC 1 space (b) AC 2 space (c) Aggregated space

Fig. 3: Flexibility sets for two ACs over a two-time slot horizon in each single AC space (left and center) and in the aggregated space (right). The exact flexibility sets are represented by black solid lines. The sets obtained using the method in [17] are shown as yellow solid lines, while those obtained using the method proposed in this paper with a greedy policy and a reactive policy are shown as red dashed and turquoise dotted lines, respectively. In the aggregated space, the maximum volume box contained inside the set obtained using [17] is represented with a dotted green line.

into each prosumer space. As can be seen from the picture, the (unique) maximum volume box in the aggregated space (that can be obtained with boxes in the prosumer spaces) does *not* correspond to maximum volume boxes within the prosumer spaces (in fact the box of AC 2 has zero volume, as noted above). This proves that maximizing the boxes within the prosumer space will lead to a suboptimal box in the aggregated space. Since our method encompasses also more complex policies than the greedy one, we can conclude that our method will always outperform the one in [8] when restricted to return a box-shaped aggregate flexibility set. Moreover, [8] does not readily provide a disaggregation policy.

B. Second case study

We consider a discrete time horizon $\mathcal{T} = \{1, 2, \ldots, M\}$ of $M = 24$ time slots of $\tau = 15$ minutes duration each, and a pool $\mathcal{I}$ of $N = 250$ prosumers offering balancing services in a 4-hour service window $\mathcal{S} = \{5, \ldots, 20\}$, with minimum required upward and downward flexibility of 500 kW. The pool of prosumers comprises 50 units of each type of the following energy resources: ACs, RFs, EVs, EBs, and PGs.

The 50 units of each energy resource are extracted at random from the discrete populations in Table III. As for the TCLs, we assume that the temperature set-point $\theta_{s,i}$ and admissible deviation $\Delta_{s,i}$ are both constant, and that both ACs and RFs are exposed to a constant ambient temperature θ_a equal to 33 °C and 20 °C, respectively. The values of $\theta_{s,i}$ and $\Delta_{s,i}$ and of the TCL thermal parameters $R_{th,i}$, $C_{th,i}$, η_i, and $P_{r,i}$ involved in the TCL model belong to the intervals specified in Table II. The electrical parameters of the AC and RF types in Table III are derived from the thermal parameters according to the modeling in [29, Section 3.3.1].

EVs are scheduled to arrive at the charging stations at slot $k^{ev}_{s,i}$ and depart at slot $k^{ev}_{f,i}$ with a minimum desired energy level $e^{ev}_{f,i}$. We assume that EBs are fully dedicated to providing flexibility and require their final charge $e^{eb}_{f,i}$ to be equal to half their capacity C_i, so that they are already balanced for future service provision. Finally, PGs are subject to ramp-rate constraints and are assumed to be always connected but with a flexibility window that starts at $k^{pg}_{s,i}$ and ends at $k^{pg}_{f,i}$, which are selected at random in the same way as $k^{ev}_{s,i}$ and $k^{ev}_{f,i}$.

The initial energy conditions for the battery or battery-like energy resources and the initial power value for the PGs are selected at random according to the uniform distribution over the interval specified in Table IV, where also information on the final energy amount and the connection and flexibility

windows are summarized, with $\mathcal{K}_i^{ev} = \{k_{s,i}^{ev}, \dots, k_{f,i}^{ev}\}$ and $\mathcal{K}_i^{pg} = \{k_{s,i}^{pg}, \dots, k_{f,i}^{pg}\}$.

In our tests, we consider different objective functions, i.e., i) the box volume (25), ii) the sum of the lengths of the box edges (26), and iii) the possible revenue obtained for flexibility provision (30). The downward and upward revenue coefficients ρ_k^+ and ρ_k^- appearing in (30), are real prices taken from the Italian ancillary service market, whose values per time slot is reported in Table V. For all cases, we impose the constraints (31), so that $0 \in \mathcal{B}_{c,d}$. We, however, do not consider any network congestion constraint when formulating the optimization problem.

The parameters of the causal policy and the baseline power profile are both optimized, resulting in the values of the indices reported in Table VI, which are consistent with the chosen objective function. The resulting power flexibility is shown in Figure 4.

Note that when the cost function represents the possible revenue of the aggregator (red dashed line), the downward flexibility margin (positive limit) is set to the minimal admissible value (500 kW), due to the fact that the values of ρ_k^- are larger than ρ_k^+ in every time slot k. When the objective is maximizing the box volume (turquoise dashed line), the power flexibility margins tend to be almost constant, whist maximizing the sum of the lengths of the box edges (green dashed line) leads to high margins in the first slots and a progressively decreasing trend within the service window.

Figure 6 shows the power and energy profiles of individual energy resources corresponding to the power request from the grid in Figure 5, when the disaggregation policy is the optimal reactive one maximizing the sum of the edge lengths of the flexibility box. Note that all energy resources track the baseline before the service window starts, while the reactive nature of the policy has an impact on their power profile after time slot $k = 20$, when the service windows is over, yet within their flexibility window. This deviation allows the aggregator to offer a wider power margin during the service window since the power profile can be adjusted after the service window is over based on past grid requests to meet local constraints, as is the case with the EB and EV. The net amount of power deviation when summed over all energy resources in the pool will however be zero so as to match the request in Figure 5.

It is worth observing that the baseline power profiles (Figure 6, solid turquoise lines) of the battery-like energy resources drive their energy state close to the center of its range. For the PG, which is not subject to energy constraints, the baseline power profile gets centered within its range. This is not surprising since in this way the PG can either decrease or increase its power production of the same amount, depending on the grid request, and, similarly, the battery-like energy resources are ready to offer both upward and downward services, compatibly with their final energy constraints, if any. In the EV case, the minimum required energy level at the end of the connection window can be recovered in only a few time slots.

Now, consider the time slots $k = 5$ and $k = 6$, where the grid requests around 3800 kW of downward flexibility (Figure 5). During time slot $k = 5$, the AC contributes with a high power value by significantly deviating from its baseline profile (see the green dashed line versus the solid turquoise one in the plot (a)). This however leads its energy state (plot

TABLE II: Residential TCL parameter ranges taken from [30, Table 3].

Parameter	AC	RF	Description
$\Delta_{s,i}$	$0.25 - 1$	$1 - 2$	Half dead-band width [°C]
$\theta_{s,i}$	$18 - 27$	$1.7 - 3.3$	Temperature set-point [°C]
$R_{th,i}$	$1.5 - 2.5$	$80 - 105$	Thermal resistance [°C/kW]
$C_{th,i}$	$1.5 - 2.5$	$0.4 - 0.8$	Thermal capacitance [kWh/°C]
η_i	$2 - 3$	$1.5 - 2.5$	Coefficient of performance
$P_{r,i}$	$4 - 7.2$	$0.2 - 0.5$	Power Rate [kW]

TABLE III: Electrical parameters of the discrete pool of EVs, EBs, ACs, RFs, and PGs used in the second case study.

Parameter	ζ_i	$u_{p,i}$	$\ell_{p,i}$	$u_{e,i}$	$\ell_{e,i}$	$u_{r,i}$	$\ell_{r,i}$	$b_{p,i}$	$b_{d,i}$
Unit	[-]	[kW]	[kW]	[kWh]	[kWh]	[kW]	[kW]	[-]	[-]
EV/EB-1	1	6.0	-6.0	40	0	-	-	0.25	-
EV/EB-2	1	7.5	-7.5	45	0	-	-	0.25	-
EV/EB-3	1	11.5	-11.5	75	0	-	-	0.25	-
EV/EB-4	1	22.5	-22.5	90	0	-	-	0.25	-
EV/EB-5	0.99	30	-30	90	0	-	-	0.25	-
EV/EB-6	0.99	30	-30	100	0	-	-	0.25	-
EV/EB-7	0.985	90	-90	100	0	-	-	0.25	-
EV/EB-8	0.98	100	-100	105	0	-	-	0.25	-
PG-1	-	20	-	-	-	2.4	-2.4	0.25	-
PG-2	-	20	-	-	-	4.8	-4.8	0.25	-
PG-3	-	20	-	-	-	6.0	-6.0	0.25	-
PG-4	-	20	-	-	-	9.0	-9.0	0.25	-
PG-5	-	40	-	-	-	4.8	-4.8	0.25	-
PG-6	-	40	-	-	-	9.6	-9.6	0.25	-
PG-7	-	40	-	-	-	12.0	-12.0	0.25	-
PG-8	-	40	-	-	-	18.0	-18.0	0.25	-
AC-1	0.926	4.5	0	16.31	16.09	-	-	-0.241	0.0668
AC-2	0.920	5.8	0	17.50	17.05	-	-	-0.240	0.0727
AC-3	0.939	6.5	0	17.50	17.05	-	-	-0.242	0.0551
AC-4	0.951	7.0	0	18.50	17.50	-	-	-0.244	0.0488
AC-5	0.961	7.2	0	16.25	15.42	-	-	-0.245	0.0327
RF-1	0.996	0.2	0	0.88	0.48	-	-	-0.25	0.0016
RF-2	0.992	0.2	0	0.69	0.37	-	-	-0.25	0.0012
RF-3	0.995	0.3	0	0.90	0.30	-	-	-0.25	0.0014
RF-4	0.997	0.3	0	1.20	0.40	-	-	-0.25	0.0013
RF-5	0.997	0.5	0	1.12	0.48	-	-	-0.25	0.0010

TABLE IV: Time windows, initial energy/power values, and final energy constraints for the heterogeneous population of the second case study.

Device	$\mathcal{C}_i$	$\mathcal{F}_i$	$e_{0,i}$	$p_{0,i}$	$e_{f,i}$
AC	$\mathcal{T}$	$\mathcal{T}$	$[\ell_{e,i}, u_{e,i}]$	-	-
RF	$\mathcal{T}$	$\mathcal{T}$	$[\ell_{e,i}, u_{e,i}]$	-	-
EV	$\mathcal{K}_i^{ev}$	$\mathcal{C}_i$	$[0.4, 0.6]C_i$	-	$[0.8, 1]C_i$
EB	$\mathcal{T}$	$\mathcal{T}$	$[\ell_{e,i}, u_{e,i}]$	-	$0.5C_i$
PG	$\mathcal{T}$	$\mathcal{K}_i^{pg}$	-	$[-u_{p,i}, 0]$	-

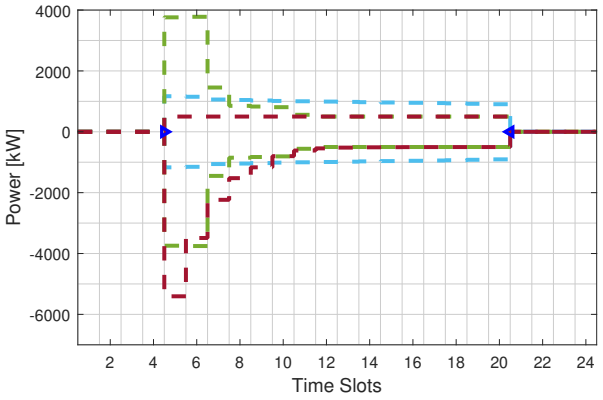


Fig. 4: Power flexibility margins obtained by aggregating a mixed population of 250 energy resources using as objective function: the box volume (25) (in turquoise), the sum of the lengths of the box edges (26) (in green), and the possible revenue (30) (in red). Blue arrows denote the beginning (pointing to the right) and the ending (pointing to the left) of the service window.

(b), green dashed line) to saturation thus causing a power reduction during time slot $k = 6$. This demonstrates the effectiveness of using a reactive policy, enabling the aggregator to exploit the AC power range to contribute to a high peak request from the grid and subsequently reduce the power in the following time slots to ensure feasibility. It is important to note that it is not just about having a smaller gain in different time slots but about being able to provide a high power peak during any time slot and then reduce power as needed. In contrast, a greedy policy cannot make such adjustments, leading to a more conservative behavior and thus reducing the flexibility available to the grid. In fact, by aggregating the pool of prosumers using a greedy policy instead of a reactive, the aggregator is not able to find a solution that ensures the minimum of 500 kW of downward and upward services. Moreover, when the limits of the constraints are relaxed to $\ell_{min} = u_{min} = 0$ kW, the power indicator obtained is only 1127 kW against 3304 kW reported in Table VI for the reactive policy, which instead enforces the minimum 500 kW constraint.

As pointed out in Section IV-C, an important feature of our method is that we can account for network constraints of the form (36). To demonstrate this feature, we randomly distribute the 250 energy resources across the nodes of an IEEE 33-bus distribution feeder with 32 transmission lines subject to congestion limits as indicated in Figure 7. We then compute the optimal reactive policy according to the objective

TABLE V: Prices per kWh along the time horizon $\mathcal{T}$ in the second case study.

k	$1 \to 4$	$5 \to 8$	$9 \to 12$	$13 \to 16$	$17 \to 20$	$21 \to 24$
ρ_k^+	0.031	0.034	0.035	0.036	0.036	0.036
ρ_k^-	0.093	0.096	0.098	0.098	0.097	0.097

TABLE VI: Performance indicators using different cost functions for a mixed population of 250 energy resources.

Objective function	Volume [kW^{16}]	Power [kW]	Revenue [Euro]
(25)	$\mathbf{6.51 \times 10^{52}}$	3207	530
(26)	8.38×10^{50}	**3304**	542
(30)	3.47×10^{50}	2784	**549**

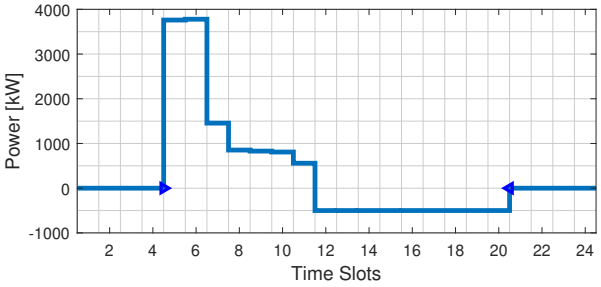


Fig. 5: Power profile request by the grid. Blue arrows denote the beginning (pointing to the right) and the ending (pointing to the left) of the service window.

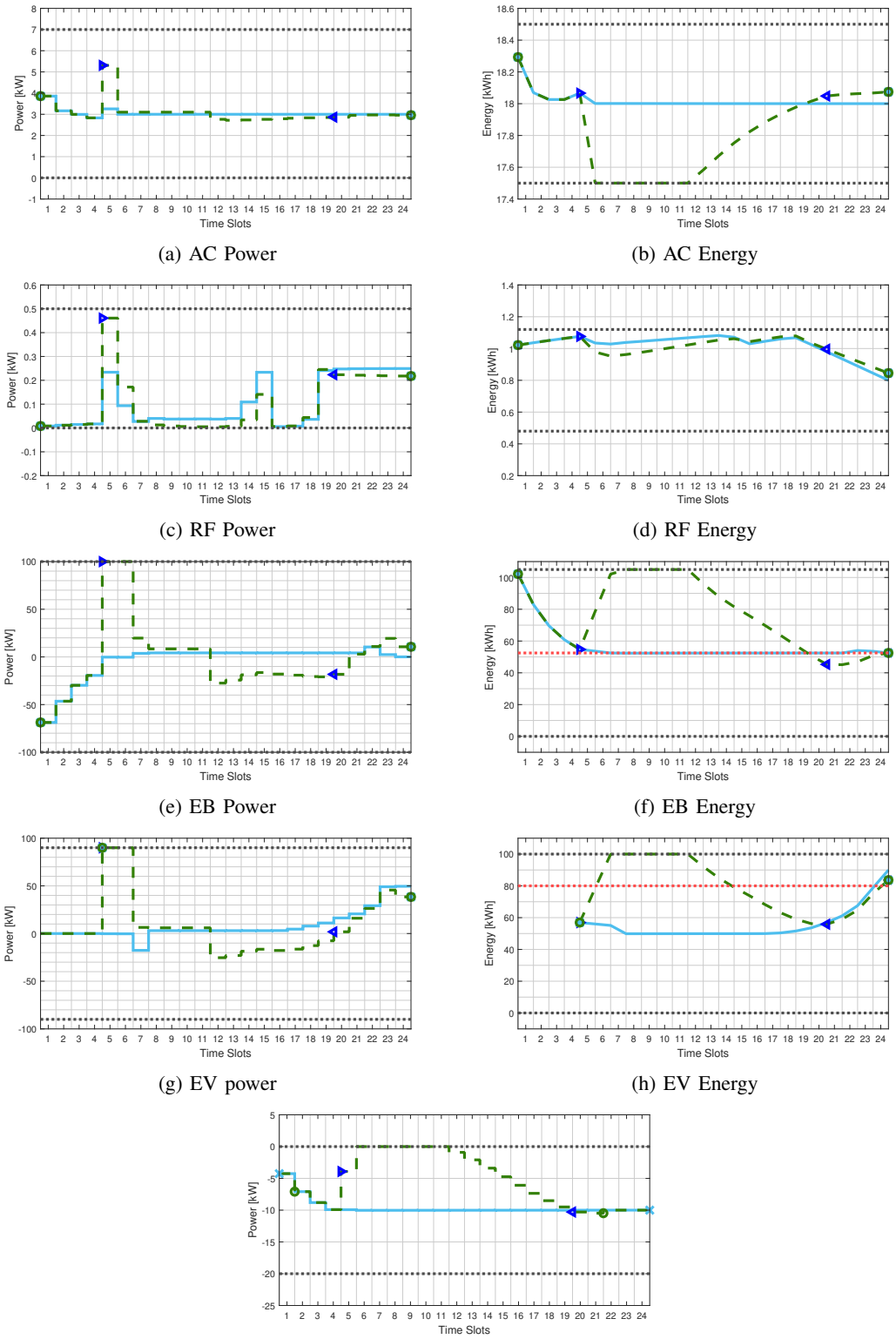
function (26) in the presence of the network constraints. The power flexibility margins of the aggregated energy resources (at node 0) obtained with and without accounting for congestion constraints are represented in Figure 8. The resulting cost function value is 3296, which is slightly less than the 3304 obtained when network constraints are not enforced. However, it can be seen from the plots that by neglecting the network constraints, an infeasible flexibility box is obtained since the power flow through ℓ_0 is allowed to exceed its limits in the time slots $k = 5$ and $k = 6$.

VI. Conclusion

In this paper, we address the problem of optimizing and managing the flexibility that a pool of distributed prosumers can offer to the grid for providing balancing services. Prosumers are equipped with energy resources whose power availability throughout the service window can be described as a polytopic set.

We introduce a linearly parameterized disaggregation policy to distribute the power grid request to the prosumers in the pool in the real-time session of the balancing market. The policy parameters are optimized jointly with the baseline power profiles so as to maximize the size of a flexibility box that inner approximates the overall flexibility set and thus make the best offer in the auction session of the balancing market. The proposed method allows to account also for practical constraints including congestion limits in the distribution network. Notably, the resulting optimization problem is well-suited for distributed implementation, enhancing computational scalability and preserving the privacy of prosumers' information.

Note that the optimization problem depends on the initial state of the energy resources. An interesting further direction



(a) AC Power

(b) AC Energy

(c) RF Power

(d) RF Energy

(e) EB Power

(f) EB Energy

(g) EV power

(h) EV Energy

(i) PG Power

Fig. 6: Power and energy profiles obtained by maximizing the sum of the half-edges of the power box shown in Figure 4 using a reactive policy. The solid turquoise lines represent the baseline profiles. The complete profiles (i.e. the sum of the baseline profile and the contribution of the device to the grid power request in Figure 5) are depicted with green dashed lines. Black dotted lines denote minimum and maximum power and energy limits. The desired final energy for the EB and the minimum desired energy for the EV are represented by red dotted lines. Blue arrows denote the beginning (pointing right) and ending (pointing left) of the service windows $\mathcal{S}$; green circles indicate the connection windows $\mathcal{C}_i$, and turquoise x indicates the flexibility windows $\mathcal{F}_i$.

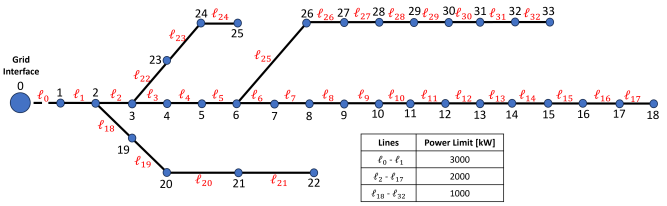


Fig. 7: IEEE 33-bus distribution system. Blue solid circles represent the nodes where energy resources are connected. The feeder lines connecting the nodes are labeled with a red ℓ_i, $i = 1, \ldots, 32$. Each of the lines is subject to symmetric power limits as reported next to the feeder.

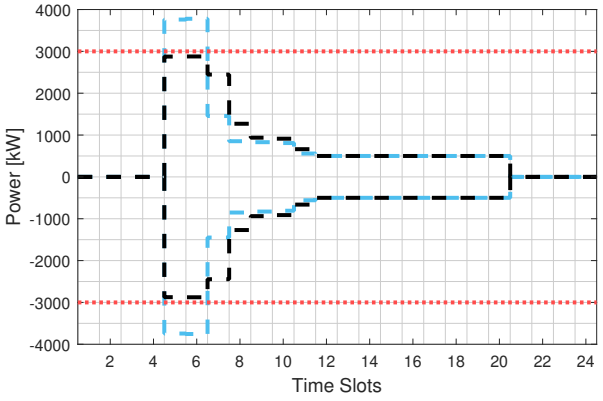


Fig. 8: Extent of power flexibility provided through line ℓ_0 when the network congestion constraints are neglected (turquoise dashed line) and accounted for (black dashed lines). The congestion power constraints of line ℓ_0 are represented by the red dotted lines.

of research would then be investigating the robustness properties of the derived policy with respect to perturbations of the initial conditions and the development of solutions that are parametric in the initial state by adopting a multiparametric programming approach.

APPENDIX

In the following, we will make extensive use of the following (well-known) relations, which we report here for ease of reference:

$$\max_{\delta \in \mathcal{B}_{c,d}} V\delta = Vc + |V|d \tag{42a}$$

$$\min_{\delta \in \mathcal{B}_{c,d}} V\delta = Vc - |V|d \tag{42b}$$

for any matrix V of appropriate dimensions, where $\mathcal{B}_{c,d}$ is the box-shaped set defined in (12) parameterized by its center c and half-side lengths d.

A. Proof of Theorem 1

Let us first focus on the set Δ_K. By definition of Δ_K in (16) and $\mathcal{P}_i$, and the fact that $\delta_i = K_i\delta$, we have that $\delta \in \Delta_K$ if and only if

$$\delta[\mathcal{S}^c] = 0, \tag{43a}$$
$$\delta = \textstyle\sum_{i\in\mathcal{I}} K_i\delta, \tag{43b}$$
$$(K_i\delta)[\mathcal{F}_i^c] = 0, \qquad i \in \mathcal{I}, \tag{43c}$$
$$F_i(g_i + a_i) + F_iK_i\delta \le h_i, \qquad i \in \mathcal{I}, \tag{43d}$$

for given nominal profiles $\bar{p}_i = g_i + a_i$ satisfying (4) and (7). Imposing the constraint $\mathcal{B}_{c,d} \subseteq \Delta_K$ in (15) is thus equivalent to enforcing the satisfaction of (43) for any $\delta \in \mathcal{B}_{c,d}$. This is analyzed next.

Let us start with (43a). Constraint (43a) holds for all $\delta \in \mathcal{B}_{c,d}$ if and only if

$$0 \le \min_{\delta\in\mathcal{B}_{c,d}} \delta[\mathcal{S}^c] \le \max_{\delta\in\mathcal{B}_{c,d}} \delta[\mathcal{S}^c] \le 0,$$

which, by (42) is equivalent to

$$0 \le c[\mathcal{S}^c] - d[\mathcal{S}^c] \le c[\mathcal{S}^c] + d[\mathcal{S}^c] \le 0,$$

thus yielding

$$\delta[\mathcal{S}^c] = 0 \quad \delta \in \mathcal{B}_{c,d} \iff c[\mathcal{S}^c] = d[\mathcal{S}^c] = 0. \tag{44}$$

Let us now define $K = \sum_{i\in\mathcal{I}} K_i$, consider (43b) and split it into

$$\delta[\mathcal{S}] = K[\mathcal{S},\mathcal{S}]\delta[\mathcal{S}] + K[\mathcal{S},\mathcal{S}^c]\delta[\mathcal{S}^c]$$
$$\delta[\mathcal{S}^c] = K[\mathcal{S}^c,\mathcal{S}]\delta[\mathcal{S}] + K[\mathcal{S}^c,\mathcal{S}^c]\delta[\mathcal{S}^c].$$

Since we know that (43a) must hold for all $\delta \in \mathcal{B}_{c,d}$, then, the previous constraints simplifies into

$$\delta[\mathcal{S}] = K[\mathcal{S},\mathcal{S}]\delta[\mathcal{S}]$$
$$0 = K[\mathcal{S}^c,\mathcal{S}]\delta[\mathcal{S}],$$

which holds for all $\delta \in \mathcal{B}_{c,d}$ if and only if

$$0 \le \min_{\delta\in\mathcal{B}_{c,d}} (I - K[\mathcal{S},\mathcal{S}])\delta[\mathcal{S}] \le \max_{\delta\in\mathcal{B}_{c,d}} (I - K[\mathcal{S},\mathcal{S}])\delta[\mathcal{S}] \le 0,$$
$$0 \le \min_{\delta\in\mathcal{B}_{c,d}} K[\mathcal{S}^c,\mathcal{S}]\delta[\mathcal{S}] \le \max_{\delta\in\mathcal{B}_{c,d}} K[\mathcal{S}^c,\mathcal{S}]\delta[\mathcal{S}] \le 0,$$

which, by (42) is equivalent to

$$\begin{aligned} 0 &\le (I - K[\mathcal{S},\mathcal{S}])c[\mathcal{S}] - |I - K[\mathcal{S},\mathcal{S}]|d[\mathcal{S}] \\ &\le (I - K[\mathcal{S},\mathcal{S}])c[\mathcal{S}] + |I - K[\mathcal{S},\mathcal{S}]|d[\mathcal{S}] \le 0, \\ 0 &\le K[\mathcal{S}^c,\mathcal{S}]c[\mathcal{S}] - |K[\mathcal{S}^c,\mathcal{S}]|d[\mathcal{S}] \\ &\le K[\mathcal{S}^c,\mathcal{S}]c[\mathcal{S}] + |K[\mathcal{S}^c,\mathcal{S}]|d[\mathcal{S}] \le 0, \end{aligned}$$

thus yielding

$$\delta = \textstyle\sum_{i\in\mathcal{I}} K_i\delta \quad \delta \in \mathcal{B}_{c,d} \iff \begin{matrix} K[\mathcal{S},\mathcal{S}] = I \\ K[\mathcal{S}^c,\mathcal{S}] = 0 \end{matrix} \tag{45}$$

since $d[\mathcal{S}] > 0$. Since K_i always appears multiplied by δ and since, by (43a), $\delta[\mathcal{S}^c] = 0$, then all entries of $K_i[\mathcal{T},\mathcal{S}^c]$ are free and we can thus set $K_i[\mathcal{T},\mathcal{S}^c] = 0$ yielding $K[\mathcal{T},\mathcal{S}^c] = 0$. This way, recalling the definition of $I_{\mathcal{S}} = \mathrm{diag}(e_{\mathcal{S}}) \in \mathbb{R}^{M\times M}$, with $e_{\mathcal{S}} \in \mathbb{R}^M$ such that $e_{\mathcal{S}}[\mathcal{S}] = 1$ and $e_{\mathcal{S}}[\mathcal{S}^c] = 0$, we have

$$\delta = \textstyle\sum_{i\in\mathcal{I}} K_i\delta \quad \delta \in \mathcal{B}_{c,d} \iff K = \sum_{i\in\mathcal{I}} K_i = I_{\mathcal{S}}. \tag{46}$$

Next, using again (43a), constraint (43c) can be simplified into

$$\begin{aligned}0 &= (K_i\delta)[\mathcal{F}_i^{\mathsf{c}}]\\ &= K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{T}]\delta\\ &= K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]\delta[\mathcal{S}] + K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}^{\mathsf{c}}]\delta[\mathcal{S}^{\mathsf{c}}]\\ &= K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]\delta[\mathcal{S}],\end{aligned}$$

which holds for all $\delta \in \mathcal{B}_{c,d}$ if and only if, for all $i \in \mathcal{I}$,

$$0 \leq \min_{\delta\in\mathcal{B}_{c,d}} K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]\delta[\mathcal{S}] \leq \max_{\delta\in\mathcal{B}_{c,d}} K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]\delta[\mathcal{S}] \leq 0,$$

which, by (42) is equivalent to

$$\begin{aligned}0 &\leq K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]c[\mathcal{S}] - |K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]|d[\mathcal{S}]\\ &\leq K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]c[\mathcal{S}] + |K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]|d[\mathcal{S}] \leq 0,\end{aligned}$$

for all $i \in \mathcal{I}$, thus yielding

$$(K_i\delta)[\mathcal{F}_i^{\mathsf{c}}] = 0 \quad \delta \in \mathcal{B}_{c,d} \iff K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}] = 0, \quad i \in \mathcal{I}, \tag{47}$$

since $d[\mathcal{S}] > 0$. Lastly, constraint (43d) holds for all $\delta \in \mathcal{B}_{c,d}$ if and only if, for all $i \in \mathcal{I}$,

$$F_i(g_i + a_i) + \max_{\delta\in\mathcal{B}_{c,d}} F_iK_i\delta \leq h_i,$$

which, by (42), yields

$$\begin{aligned}&F_i(g_i + a_i) + F_iK_i\delta \leq h_i \quad \delta \in \mathcal{B}_{c,d}\\ &\iff F_i(g_i + a_i) + F_iK_ic + |F_iK_i|d \leq h_i\\ &\iff F_i(g_i + a_i + K_ic) + |F_iK_iD|1_M \leq h_i,\ i \in \mathcal{I},\end{aligned} \tag{48}$$

where $D = \mathrm{diag}(d)$ and $d \geq 0$.

In summary, the double implications in (44), (46), (47), and (48) together with $d[\mathcal{S}] > 0$, show that problem (15) and (17) share the same feasible set. Since they also have the same cost function, the two problems are equivalent, thus concluding the first part of the proof.

While most constraints in problem (17) are linear in the decision variables c, d, g_i, a_i, and K_i, in the last constraint (cf. right hand side of (48)) we have a product between K_i and $D = \mathrm{diag}(d)$. In order to handle the bilinearity, we introduce the following change of variables

$$G_i = K_iD \quad \text{and} \quad \gamma_i = a_i + K_ic, \tag{49}$$

where G_i and γ_i are new decision variables in place of K_i and a_i. We thus have to check whether all constraints in (17) can be equivalently expressed in terms of c, d, g_i, γ_i, and G_i, without any bilinearity. From (49) the right hand side of (48) satisfies

$$\begin{aligned}&F_i(g_i + a_i + K_ic) + |F_iK_iD|1_M \leq h_i\\ &\iff F_i(g_i + \gamma_i) + |F_iG_i|1_M \leq h_i,\ i \in \mathcal{I}.\end{aligned} \tag{50}$$

Since

$$\begin{aligned}G_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}] &= K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{T}]D[\mathcal{T},\mathcal{S}]\\ &= K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]D[\mathcal{S},\mathcal{S}] + K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}^{\mathsf{c}}]\underbrace{D[\mathcal{S}^{\mathsf{c}},\mathcal{S}]}_{0}\\ &= K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}]D[\mathcal{S},\mathcal{S}],\end{aligned}$$

we immediately have that the right hand side of (47) satisfies

$$K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}] = 0 \iff G_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}] = 0, \quad i \in \mathcal{I}, \tag{51}$$

as an effect of $D[\mathcal{S},\mathcal{S}] = \mathrm{diag}(d[\mathcal{S}])$ being invertible due to $d[\mathcal{S}] > 0$. Moreover,

$$\begin{aligned}G_i[\mathcal{F}_i,\mathcal{S}] &= K_i[\mathcal{F}_i,\mathcal{T}]D[\mathcal{T},\mathcal{S}]\\ &= K_i[\mathcal{F}_i,\mathcal{S}]D[\mathcal{S},\mathcal{S}] + K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}^{\mathsf{c}}]\underbrace{D[\mathcal{S}^{\mathsf{c}},\mathcal{S}]}_{0}\\ &= K_i[\mathcal{F}_i,\mathcal{S}]D[\mathcal{S},\mathcal{S}],\end{aligned}$$

and, hence

$$K_i[\mathcal{F}_i,\mathcal{S}] = G_i[\mathcal{F}_i,\mathcal{S}]D[\mathcal{S},\mathcal{S}]^{-1},$$

as an effect of $D[\mathcal{S},\mathcal{S}] = \mathrm{diag}(d[\mathcal{S}])$ being invertible due to $d[\mathcal{S}] > 0$. Finally, since by (51) we want $K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}] = 0$ and we arbitrarily set $K_i[\mathcal{T},\mathcal{S}^{\mathsf{c}}] = 0$, we can compactly write part of the inverse coordinate change as

$$K_i = G_iD^\dagger, \tag{52}$$

with $D^\dagger = \mathrm{diag}(d^\dagger)$ and $d^\dagger \in \mathbb{R}_+^M$ such that[2] $d^\dagger(k) = 1/d(k)$ for $k \in \mathcal{S}$ and $d^\dagger(k) = 0$ for $k \in \mathcal{S}^{\mathsf{c}}$. Let us define $G = \sum_{i\in\mathcal{I}} G_i$. Let us define $G = \sum_{i\in\mathcal{I}} G_i$. By (49),

$$G = \textstyle\sum_{i\in\mathcal{I}} G_i = \sum_{i\in\mathcal{I}} K_iD = KD,$$

hence

$$\begin{aligned}G[\mathcal{S},\mathcal{S}] &= K[\mathcal{S},\mathcal{T}]D[\mathcal{T},\mathcal{S}]\\ &= K[\mathcal{S},\mathcal{S}]D[\mathcal{S},\mathcal{S}] + K[\mathcal{S},\mathcal{S}^{\mathsf{c}}]\underbrace{D[\mathcal{S}^{\mathsf{c}},\mathcal{S}]}_{0}\\ &= K[\mathcal{S},\mathcal{S}]D[\mathcal{S},\mathcal{S}]\\ G[\mathcal{S}^{\mathsf{c}},\mathcal{S}] &= K[\mathcal{S}^{\mathsf{c}},\mathcal{T}]D[\mathcal{T},\mathcal{S}]\\ &= K[\mathcal{S}^{\mathsf{c}},\mathcal{S}]D[\mathcal{S},\mathcal{S}] + K[\mathcal{S}^{\mathsf{c}},\mathcal{S}^{\mathsf{c}}]\underbrace{D[\mathcal{S}^{\mathsf{c}},\mathcal{S}]}_{0}\\ &= K[\mathcal{S}^{\mathsf{c}},\mathcal{S}]D[\mathcal{S},\mathcal{S}],\end{aligned}$$

from which we can see that the right hand side of (45) satisfies

$$\begin{aligned}K[\mathcal{S},\mathcal{S}] &= I\\ K[\mathcal{S}^{\mathsf{c}},\mathcal{S}] &= 0\end{aligned} \iff \begin{aligned}G[\mathcal{S},\mathcal{S}] &= D[\mathcal{S},\mathcal{S}]\\ G[\mathcal{S}^{\mathsf{c}},\mathcal{S}] &= 0,\end{aligned}$$

as an effect of $D[\mathcal{S},\mathcal{S}] = \mathrm{diag}(d[\mathcal{S}])$ being invertible due to $d[\mathcal{S}] > 0$. Since $G[\mathcal{T},\mathcal{S}^{\mathsf{c}}] = K[\mathcal{T},\mathcal{T}]D[\mathcal{T},\mathcal{S}^{\mathsf{c}}] = 0$, we finally have that the right hand side of (46) satisfies

$$K = \textstyle\sum_{i\in\mathcal{I}} K_i = I_{\mathcal{S}} \iff G = \sum_{i\in\mathcal{I}} G_i = D. \tag{53}$$

Constraint $c[\mathcal{S}^{\mathsf{c}}] = d[\mathcal{S}^{\mathsf{c}}] = 0$ in the right hand side of (44) depends on c and d only and is not affected by the proposed coordinate change. Since $D = \mathrm{diag}(d)$ now appears linearly in all constraints, we can keep d as a decision variable and substitute $D = \mathrm{diag}(d)$. Using (52) in the expression of γ_i in (49) and summing over all prosumers gives

$$\begin{aligned}\textstyle\sum_{i\in\mathcal{I}} \gamma_i &= \textstyle\sum_{i\in\mathcal{I}} a_i + \big(\sum_{i\in\mathcal{I}} K_i\big)c\\ &= \textstyle\sum_{i\in\mathcal{I}} a_i + \big(\sum_{i\in\mathcal{I}} G_i\big)D^\dagger c\\ &= \textstyle\sum_{i\in\mathcal{I}} a_i + DD^\dagger c\\ &= \textstyle\sum_{i\in\mathcal{I}} a_i + I_{\mathcal{S}}c,\end{aligned}$$

[2]Note that $D^\dagger$ is the unique pseudoinverse (Moore-Penrose inverse) of D and $DD^\dagger = D^\dagger D = I_{\mathcal{S}}$.

where the second-last equality holds because we are imposing $G = D$ according to (53) and the last equality is true by definition of $I_{\mathcal{S}}$. This immediately yields

$$\begin{aligned}&\textstyle\sum_{i\in\mathcal{I}}\gamma_i[\mathcal{S}] = \sum_{i\in\mathcal{I}} a_i[\mathcal{S}] + c[\mathcal{S}]\\ &\textstyle\sum_{i\in\mathcal{I}}\gamma_i[\mathcal{S}^{\mathsf{c}}] = \sum_{i\in\mathcal{I}} a_i[\mathcal{S}^{\mathsf{c}}] + \underbrace{I_{\mathcal{S}}[\mathcal{S}^{\mathsf{c}},\mathcal{T}]}_{0}\, c = \sum_{i\in\mathcal{I}} a_i[\mathcal{S}^{\mathsf{c}}],\end{aligned}$$

from which we obtain

$$\textstyle\sum_{i\in\mathcal{I}} a_i = 0 \wedge c[\mathcal{S}^{\mathsf{c}}] = 0 \iff c = \sum_{i\in\mathcal{I}}\gamma_i \wedge c[\mathcal{S}^{\mathsf{c}}] = 0, \tag{54}$$

where, similarly to d, we can keep c as decision variable as it now appears linearly in all constraints. Lastly, we have to reformulate constraint (7). Since g_i is still a decision variable, we can readily enforce $g_i[\mathcal{C}_i^{\mathsf{c}}] = 0$. As for $a_i[\mathcal{C}_i^{\mathsf{c}}] = 0$ we observe that

$$\begin{aligned}\gamma_i[\mathcal{C}_i^{\mathsf{c}}] &= a_i[\mathcal{C}_i^{\mathsf{c}}] + K_i[\mathcal{C}_i^{\mathsf{c}},\mathcal{T}]c\\ &= a_i[\mathcal{C}_i^{\mathsf{c}}] + \underbrace{K_i[\mathcal{C}_i^{\mathsf{c}},\mathcal{S}]}_{0} c[\mathcal{S}] + K_i[\mathcal{C}_i^{\mathsf{c}},\mathcal{S}^{\mathsf{c}}]\underbrace{c[\mathcal{S}^{\mathsf{c}}]}_{0}\\ &= a_i[\mathcal{C}_i^{\mathsf{c}}],\end{aligned}$$

where $K_i[\mathcal{C}_i^{\mathsf{c}},\mathcal{S}] = 0$ because $\mathcal{C}_i^{\mathsf{c}} \subseteq \mathcal{F}_i^{\mathsf{c}}$ and $K_i[\mathcal{F}_i^{\mathsf{c}},\mathcal{S}] = 0$ by (47), while $c[\mathcal{S}^{\mathsf{c}}] = 0$ by (54). By the previous discussion we immediately have

$$g_i[\mathcal{C}_i^{\mathsf{c}}] = 0 \wedge a_i[\mathcal{C}_i^{\mathsf{c}}] = 0 \iff g_i[\mathcal{C}_i^{\mathsf{c}}] = 0 \wedge \gamma_i[\mathcal{C}_i^{\mathsf{c}}] = 0. \tag{55}$$

By equivalences (50), (51), (53), (54), and (55) we proved that problem (17) and (18) have an equivalent feasible set under the change of variable (49). If we further assume that the cost function does not depend on the a_i's, then the cost function is unchanged and the two problems are equivalent (and equivalent to (15)). Moreover we have shown that we can recover K_i from G_i and D using (52), while by (49) we can recover a_i from γ_i, c, and K_i using $a_i = \gamma_i - K_i c$, which concludes the proof. ∎

References

[1] C. Fetting, "The european green deal," *ESDN report*, vol. 53, 2020.

[2] G. Rancilio, A. Rossi, D. Falabretti, A. Galliani, and M. Merlo, "Ancillary services markets in Europe: Evolution and regulatory trade-offs," *Renewable and Sustainable Energy Reviews*, vol. 154, Nov. 2022, Art. no. 111850.

[3] A. M. Carreiro, H. M. Jorge, and C. H. Antunes, "Energy management systems aggregators: A literature survey," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 1160–1172, Feb. 2017.

[4] S. Kerscher and P. Arboleya, "The key role of aggregators in the energy transition under the latest European regulatory framework," *International Journal of Electrical Power & Energy Systems*, vol. 134, Jul. 2022, Art. no. 107361.

[5] F. L. Müller, J. Szabó, O. Sundström, and J. Lygeros, "Aggregation and disaggregation of energetic flexibility from distributed energy resources," *IEEE Trans. on Smart Grid*, vol. 10, pp. 1205–1214, 2019.

[6] K. Trangbaek and J. Bendtsen, "Exact constraint aggregation with applications to smart grids and resource distribution," in *IEEE Conf. on Decision and Control*, 2012, pp. 4181–4186.

[7] E. Öztürk, K. Rheinberger, T. Faulwasser, K. Worthmann, and M. Preißinger, "Aggregation of demand-side flexibilities: A comparative study of approximation algorithms," *Energies*, vol. 15, 2022, Art. no. 2501.

[8] M. S. Nazir, I. A. Hiskens, A. Bernstein, and E. Dall'Anese, "Inner approximation of Minkowski sums: A union-based approach and applications to aggregated energy resources," in *IEEE Conf. on Decision and Control*, 2018, pp. 5708–5715.

[9] L. Zhao, H. Hao, and W. Zhang, "Extracting flexibility of heterogeneous deferrable loads via polytopic projection approximation," in *IEEE Conf. on Decision and Control*, 2016, pp. 6651–6656.

[10] S. Barot, "Aggregate load modeling for demand response via the Minkowski sum," Ph.D. dissertation, University of Toronto, 2017.

[11] B. Cui, A. Zamzam, and A. Bernstein, "Network-cognizant time-coupled aggregate flexibility of distribution systems under uncertainties," *IEEE Control Systems Letters*, vol. 5, pp. 1723–1728, 2021.

[12] X. Chen and N. Li, "Leveraging two-stage adaptive robust optimization for power flexibility aggregation," *IEEE Trans. on Smart Grid*, vol. 12, pp. 3954–3965, 2021.

[13] B. B. Belmonte, P. Mouratidis, G. Franke, and S. Rinderknecht, "Developments in the cost of grid balancing services and the design of the European balancing market," *Energy Reports*, vol. 10, pp. 910–931, Nov. 2023.

[14] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, "Aggregate flexibility of thermostatically controlled loads," *IEEE Trans. on Power Systems*, vol. 30, pp. 189–198, 2015.

[15] G. Wang, Z. Li, and F. Wang, "Enhanced sufficient battery model for aggregate flexibility of thermostatically controlled loads considering coupling constraints," *IEEE Trans. on Sustainable Energy*, vol. 12, pp. 2493–2496, 2021.

[16] D. Zamudio, A. Falsone, F. Bianchi, and M. Prandini, "Ancillary services provision via aggregation: Joint power flexibility assessment and disaggregation policy design," *IEEE Control Systems Letters*, vol. 7, pp. 3108–3113, 2023.

[17] L. Zhao, W. Zhang, H. Hao, and K. Kalsi, "A geometric approach to aggregate flexibility modeling of thermostatically controlled loads," *IEEE Trans. on Power Systems*, vol. 32, pp. 4721–4731, 2017.

[18] H. Hao, Y. Lin, A. Kowli, P. Barooah, and S. Meyn, "Ancillary service to the grid through control of fans in commercial building HVAC systems," *IEEE Trans. on Smart Grid*, vol. 5, no. 4, pp. 2066–2074, 2014.

[19] N. I. Nimalsiri, C. P. Mediwaththe, E. L. Ratnam, M. Shaw, D. B. Smith, and S. K. Halgamuge, "A survey of algorithms for distributed charging control of electric vehicles in smart grid," *IEEE Trans. on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4497–4515, 2019.

[20] A. Majzoobi and A. Khodaei, "Application of microgrids in supporting distribution grid flexibility," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3660–3669, 2016.

[21] S. Barot and J. A. Taylor, "A concise, approximate representation of a collection of loads described by polytopes," *International Journal of Electrical Power and Energy Systems*, vol. 84, pp. 55–63, 2017.

[22] R. Vignali, A. Falsone, F. Ruiz, and G. Gruosso, "Towards a comprehensive framework for V2G optimal operation in presence of uncertainty," *Sustainable Energy, Grids and Networks*, vol. 31, May. 2022, Art. no. 100740.

[23] D. Kenefake, I. Pappas, N. A. Diangelakis, S. Avraamidou, R. Oberdieck, and E. N. Pistikopoulos, *Multi-parametric Linear and Quadratic Programming*. Springer, 2020, pp. 1–5.

[24] P. Tøndel and T. A. Johansen, "Complexity reduction in explicit linear model predictive control," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 189–194, 2002.

[25] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization with coupling constraints," *Automatica*, vol. 84, pp. 149–158, 2017.

[26] A. Falsone and M. Prandini, "A distributed dual proximal minimization algorithm for constraint-coupled optimization problems," *IEEE Control Systems Letters*, vol. 5, no. 1, pp. 259–264, 2021.

[27] A. Falsone, I. Notarnicola, G. Notarstefano, and M. Prandini, "Tracking-ADMM for distributed constraint-coupled optimization," *Automatica*, vol. 117, 2020, Art. no. 108962.

[28] A. Falsone and M. Prandini, "Augmented Lagrangian Tracking for distributed optimization with equality and inequality coupling constraints," *Automatica*, vol. 157, 2023, Art. no. 111269.

[29] D. Zamudio, "Balancing service provision via aggregation of flexible prosumers," Ph.D. dissertation, Politecnico di Milano, 2024.

[30] J. L. Mathieu, M. E. Dyson, and D. S. Callaway, "Resource and revenue potential of California residential load participation in ancillary services," *Energy Policy*, vol. 80, pp. 76–87, May. 2015.

[31] A. König, L. Nicoletti, D. Schröder, S. Wolff, A. Waclaw, and M. Lienkamp, "An overview of parameter and cost for battery electric vehicles," *World Electric Vehicle Journal*, vol. 12, pp. 1–29, Feb. 2021.

[32] I. A. Khan, H. Mokhlis, N. N. Mansor, H. A. Illias, L. J. Awalin, and L. Wang, "New trends and future directions in load frequency control and flexible power system: A comprehensive review," *Alexandria Engineering Journal*, vol. 71, pp. 263–308, May. 2023.