Peak Load Energy Management by Direct Load Control Contracts

Ali Fattahi

Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095,
ali.fattahi.1@anderson.ucla.edu

Sriram Dasu

Marshall School of Business, University of Southern California, Los Angeles, Los Angeles, California 90089,
dasu@marshall.usc.edu

Reza Ahmadi

Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095,
rahmadi@anderson.ucla.edu

Energy firms use demand-response programs such as direct load control contracts (DLCCs) to curtail electricity consumption during peak load periods. We study contracts that allow utilities to directly reduce customers' consumption via remote control devices. Regulators limit the number of times (calls) in a year a customer's power consumption can be reduced, the duration of each call, and the total number of hours of power reduction over the life of the contract. For technical reasons, utilities partition customers that enroll in DLCCs into groups. To optimally implement DLCCs, utilities have to solve a stochastic dynamic optimization problem each day that determines how many groups to call and the timing and duration of each call. This is a provably difficult optimization problem. A practically appealing version of DLCCs is one in which all calls are restricted to the same duration. Under this restriction, we show that the problem has a special structure that enables us to find optimal solutions. We develop a near-optimum heuristic procedure for the general stochastic problem. There are two simplifications in our approximation scheme. First, we aggregate the groups into a single group, which allows us to reduce the state space. Second, we ignore the detailed structure of the uncertainty. We refer to the error induced by these two schemes as aggregation and information error, respectively. The asymptotic relative error, which includes the information error ($\mathcal{O}(1/\sqrt{n})$) and the aggregation error ($\mathcal{O}(1/n)$), is shown to be zero. The effectiveness of our proposed solution procedures is also verified numerically. Finally, applying our heuristic to an industrial instance from the Southern California region results in approximately 8% saving in the energy generation cost.

Subject classifications: Electric Industries. Dynamic programming. Stochastic Programming.

1. Introduction

For energy firms, matching supply and demand is challenging due to variability in demand, large fluctuations in cost among alternative power generation methods, and the inability to store power. Power consumption is often weather dependent and varies by hour of day. High demand can cause grid failure, which is estimated to cost tens of billions of dollars per year.[1] Inability to meet demand, besides being unacceptable, results in significant direct and indirect financial penalties for the utilities (Taylor and Taylor 2015).

Supply costs are highly non-linear (see Fig. 1, and Meulemeester (2014), Posner (2015), Next-Kraftwerke-Belgium (2016)). When demand rises above amounts that can be supplied by primary power generators such as renewable (Alizamir et. al 2016), nuclear, coal, and gas, secondary sources such as diesel and gasoline-powered generators are used. Peak period refers to the time interval when demand exceeds primary generators' capacity and peak units must be activated. As indicated in Fig. 1, peak units are significantly more expensive. The U.S. Government Accountability Office finds that the cost of generating electricity during hot summer days is about 10 times higher than at night. The top 100-highest priced hours in one year account for nearly 20% of the total energy cost.[2] Therefore, dynamic demand management during peak periods is central to matching supply and demand in this industry.

Utilities have introduced a number of programs that reduce or shift the peak load through a pricing mechanism (Peura and Bunn 2015, Baldick et. al 2006, Kamat and Oren 2002) or by direct load control contracts (DLCCs) (Oren and Smith 1992). In this paper, we concentrate on DLCCs, which are very popular in developed countries and, according to the Federal Energy Regulatory Commission, in the U.S., "as of 2012, more than 200 utilities across the country offered some type of direct load control program for residential customers" (Michigan Public Service Commission, 2015). DLCCs permit utility companies to directly reduce a customer's energy usage using a remote control device that is installed on site. In residential areas, DLCCs usually target central air conditioning systems, controlled electric appliances, electric water heaters, and pool pumps. By
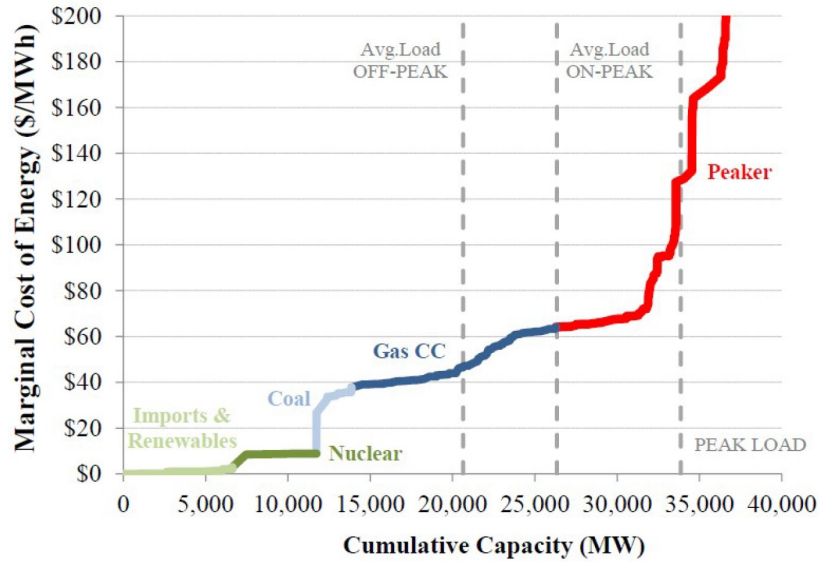
Figure 1   Power generation cost is strictly convex and increases in hourly consumption rate (Solar Energy Indus-
tries Association, 2016).

reducing some consumers' energy consumption, utility companies can avoid launching peak units
(Michigan Public Service Commission, 2015). It is also anticipated that the implementation of
DLCCs will rapidly expand due to the Internet of Things.

The general structure of a DLCC that we study in this paper is as follows. Due to technological
constraints, the DLCC participants are partitioned into identical groups to allow for load shedding
in equal increments and ensure that the load is reduced in a balanced way.[3] A utility company
informs customers in advance that their consumption will be reduced by a pre-specified amount
for a certain period of time. Regulators allow the utility companies to reduce customer groups'
consumption for a maximum of K events in a year, for a maximum cumulative period of H hours,
and each event is required to be continuous and last no longer than L hours.

The energy consumption profile (ECP), illustrated in Fig. 2, depicts the amount of power con-
sumed over a typical day. The horizontal axis denotes the time of day and the vertical axis shows
the energy consumption rate in mega watts per hour. ECPs can have various shapes—e.g., uni-
modal on a summer day and bimodal on a winter day (see Fig. 2). Our work is based on problems
faced by a large utility serving the Southern California region. This utility classifies ECPs into
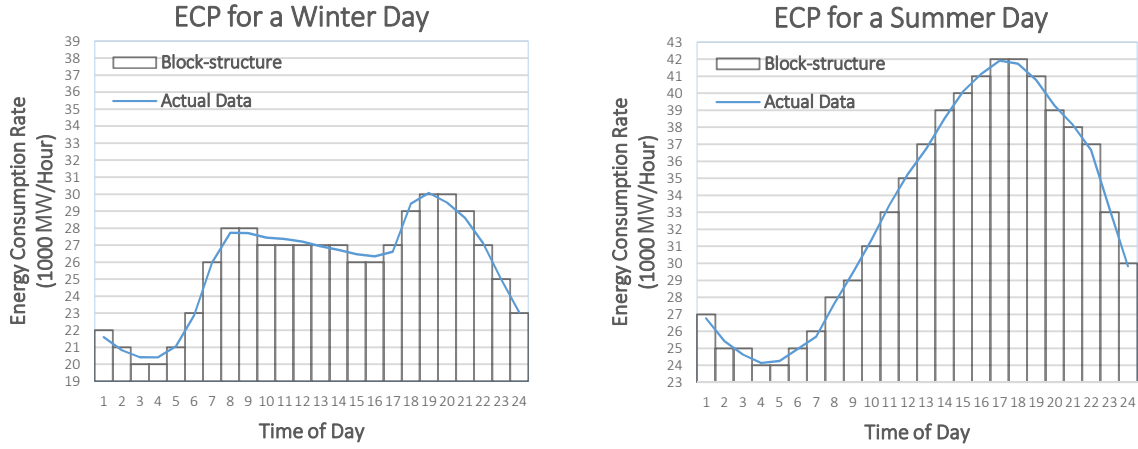
Figure 2   Actual data and block-structure for a summer day and a winter day (actual data from California Independent System Operator (2012)).

day-types. There are approximately 10 day-types. To optimize the value of DLCCs, the firm must decide each day the number of groups to "call", the starting time, and the duration for each call. This decision is typically made a day in advance and communicated to the affected consumers. In addition, the type of day is almost always known a day ahead. Although the type of day for the affected day is known, the types of days in the rest of the contract period remain uncertain. For modeling purposes, the ECP is divided into identical rectangular blocks, in which each block represents the amount of load reduction that can be achieved if one customer group is called for one time unit.

To illustrate an aspect of the optimization problem we are addressing, assume that we have decided to call three groups for 1, 1, and 2 hours, respectively (see Fig. 3). Observe that the savings will depend on the starting time for each of these groups. Assignment (a) uses starting times 1, 7, and 7, while assignment (b) uses 7, 7, and 1. Fig. 3 shows the differences in power generation costs for these two assignments.

In summary, each day, after the type of day is determined, the following questions must be answered: Which of the G groups should be called that day, for how long, and when should the call for each group begin? The objective is to minimize the total power generation cost over the
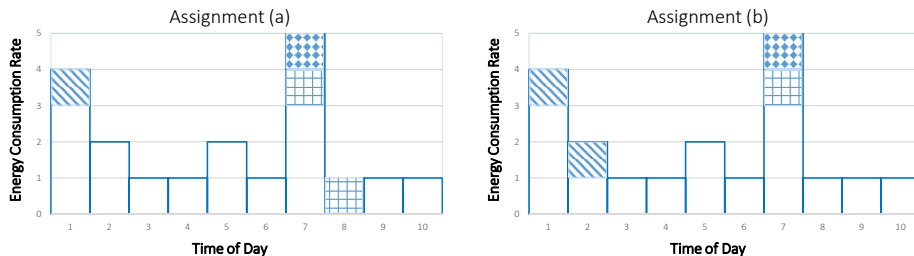
Figure 3   An example: Assigning 3 calls with lengthes 1, 1, and 2 to customer groups.

contract horizon, which is equivalent to maximizing the total saving. The constraints are (i) each group of customers is called no more than K times in a year, (ii) the duration of each call is less than L hours, (iii) the power reduction must occur over a continuous interval, and (iv) the total number of hours of load reduction for a group in a year cannot exceed H hours.

We refer to our problem as the general stochastic problem (GSP), in which the uncertainty in the type of day is modeled via multi-nomial distributions. The GSP can be viewed as a general version of the revenue management problem (Adelman 2007, Zhang and Adelman 2009, Cooper 2002). Specifically, with one customer group, the GSP reduces to the well-known revenue management problem (Jasin and Kumar 2012). The GSP with two groups is weakly NP-complete and with G groups is strongly NP-complete. The GSP is a finite-horizon stochastic dynamic program with a $2G$-dimensional state space. The state space is the number of available calls and hours for each group. We present a heuristic approach (HGSP) that is based on two approximations: (i) ignoring uncertainty which creates *information error*, and (ii) ignoring group identities which creates *aggregation error*. Our analysis of the information error adapts the approach of Jasin and Kumar (2012). We show that the worst-case relative performance of the information and aggregation errors are $\mathcal{O}(1/\sqrt{n})$ and $\mathcal{O}(1/n)$, respectively, where n denotes the problem size.

The power company that we are working with is interested in a policy in which all calls are of the same duration. The corresponding problem is referred to as the fixed-length stochastic problem (FSP). We present a polynomial time algorithm to find an optimal solution to the FSP.

The remainder of the paper is organized as follows. Section 2 presents the mathematical formulation of our problem and establishes its NP-completeness. In Section 3, we present a polynomial

time algorithm to solve the FSP. Sections 4 and 5 present our heuristic approach to solve the GSP, study its theoretical properties, and establish its worst-case error bounds. In Section 6, we test the effectiveness of the HGSP and solve an industrial instance of the problem. Finally, we conclude the paper and offer some directions for future work.

2. Problem Formulation

The GSP is a *finite-horizon stochastic dynamic program* in which stages correspond to the days in the planning horizon $d = 1, \ldots, D$. For a complete list of notations, see Appendix F. The index d refers to the beginning of a day and, hence, $d = 1$ is the beginning of the planning horizon. Uncertainty is induced by day-type $\omega \in \Omega$. The state variables are the remaining number of calls and hours for each group. As previously stated, a decision is made each day to determine which groups to call, the duration of the calls, and the starting time for each call.

There are G identical groups that are indexed as $g = 1, \ldots, G$. Group g can be called on day d for a duration of l_{dg} hours, satisfying $0 \leq l_{dg} \leq L$. Let m_{dg} be a binary variable that shows if group g is called on day d. If $l_{dg} > 0$, then $m_{dg} = 1$, and otherwise $m_{dg} = 0$.

We define vectors $\mathbf{L}_d = (l_{d1}, \ldots, l_{dG})$ and $\mathbf{M}_d = (m_{d1}, \ldots, m_{dG})$, for all d, which respectively show the duration of calls and whether each group is called on day d. We define vectors $\mathbf{H}_d = (h_{d1}, \ldots, h_{dG})$ and $\mathbf{K}_d = (k_{d1}, \ldots, k_{dG})$, for all d, which respectively show the number of calls available and the duration of time available from each group on day d. At the beginning of the planning horizon $d = 1$, we have $\mathbf{H}_1 = (H, \ldots, H)$ and $\mathbf{K}_1 = (K, \ldots, K)$, which represent identical contracts of groups.

Let the augmented vector $(\mathbf{L}_d, \mathbf{M}_d)$ denote a *decision* on day d. Let $\boldsymbol{\Theta}_d$ denote the set of *feasible decisions* on day d. We define $\boldsymbol{\Theta}_d := \{(\mathbf{L}_d, \mathbf{M}_d) \mid m_{dg} \leq k_{dg}, \ l_{dg} \leq \min\{h_{dg}, L\}, \ m_{dg} = \mathbb{I}(l_{dg}), \text{ for all } g\}$, for all d. Here, $\mathbb{I}(.)$ is the indicator function: $\mathbb{I}(l_{dg}) = 1$ if $l_{dg} > 0$, and otherwise $\mathbb{I}(l_{dg}) = 0$. The *finite-horizon stochastic dynamic program* for the GSP is as follows:

$$\mathcal{V}_d(\mathbf{H}_d, \mathbf{K}_d) = \min_{(\mathbf{L}_d, \mathbf{M}_d) \in \boldsymbol{\Theta}_d} \{\psi(\mathbf{L}_d, \omega_d) +$$

$$\mathbf{E}\left[\mathcal{V}_{d+1}(\mathbf{H}_d - \mathbf{L}_d, \mathbf{K}_d - \mathbf{M}_d)\right]\}, \ \forall d = 1, \ldots, D-1,$$

$$\mathcal{V}_D(\mathbf{H}_D, \mathbf{K}_D) = \min_{(\mathbf{L}_D, \mathbf{M}_D) \in \boldsymbol{\Theta}_D} \{\psi(\mathbf{L}_D, \omega_D)\}, \tag{1}$$

where, $\mathcal{V}_d(\mathbf{H}_d, \mathbf{K}_d)$ is the optimal cost for days $d, d+1, \ldots, D$ given $\mathbf{H}_d$ available time and $\mathbf{K}_d$ available number of calls. Finally, $\psi(\mathbf{L}_d, \omega_d)$ denotes the minimum cost that one could achieve in day d, which is found by solving a mixed integer linear programming (MILP) problem (see Section 2.1).

2.1. Single day problem

We present an MILP formulation to obtain $\psi(\mathbf{L}_d, \omega_d)$. This MILP is referred to as the *single day problem* (SDP).Each day consists of T time periods—e.g., 24 hours. Let t denote the index for time periods in a day. We assume that the ECP height is constant during time period t on day d and is given by function $r_t : \Omega \longrightarrow \mathbb{Z}_+$, and hence $r_t(\omega_d)$ is the ECP height in period t and day d. Moreover, the energy cost in period t and day d depends solely on $r_t(\omega_d)$, which is given by the function $f : \mathbb{Z}_+ \longrightarrow \mathbb{R}$. For each $z \in \mathbb{Z}, z \geq 1$, we define $\partial f(z) := f(z) - f(z-1)$ and $\partial^2 f(z) := \partial f(z) - \partial f(z-1)$. We assume that the cost function satisfies: (i) $f(0) = \partial f(0) = \partial^2 f(0) = 0$, (ii) $\partial f(z) > 0$, for all $z \in \mathbb{Z}$, $z \geq 1$, and (iii) $\partial^2 f(z) > 0$, for all $z \in \mathbb{Z}$, $z \geq 1$. Properties (ii) and (iii) respectively mean that $f(z)$ is strictly increasing in z and that the marginal increase in z is also strictly increasing in z (as indicated in Fig.1). Finally, note that f does not have index d and/or t since the energy cost depends solely on the ECP height.

Let binary variables u_{gt} and w_{gt}, for all g and t, be defined as follows: $u_{gt} = 1$ shows that group g is on call in period t of day d, and $w_{gt} = 1$ indicates that group g starts in t. The SDP is formulated as follows:

$$\psi(\mathbf{L}_d, \omega_d) = \min \ \sum_{t=1}^{T} f\left(\left[r_t(\omega_d) - \sum_{g=1}^{G} u_{gt}\right]^+\right) \tag{2}$$

$$\text{s.t.} \ \sum_{t=1}^{T} u_{gt} = l_{dg}, \ \forall g, \tag{3}$$

$$w_{gt} \geq u_{gt} - u_{g(t-1)}, \ \forall g, t \geq 2, \tag{4}$$

$$w_{g1} \geq u_{g1}, \ \forall g, \tag{5}$$

$$\sum_{t=1}^{T} w_{gt} \leq 1, \ \forall g, \tag{6}$$

$$u_{gt}, w_{gt} \in \{0,1\}, \ \forall g, t, \tag{7}$$

where, $[z]^+ := \max\{z, 0\}$. In the objective function (Eq. 2), $\sum_{g=1}^{G} u_{gt}$ is the number of groups on call in period t. The ECP height in period t decreases by $\sum_{g=1}^{G} u_{gt}$ and hence becomes $r_t(\omega_d) - \sum_{g=1}^{G} u_{gt}$ if this value is positive, and 0 otherwise. Eq. (3) ensures that group g is called for l_{dg} time units and Eqs. (4)–(6) guarantee that the reduction occurs over a continuous interval. Eqs. (4)–(5) initialize the starting time of the calls. Eq. (6) confirms that each group can be called once on day d at most.

We next show that the deterministic version of the GSP, which we refer to as the *general deterministic problem* (GDP), is very difficult. The following proposition establishes that special case of the GDP with two groups is *weakly NP-complete* and the general case of the GDP is *strongly NP-complete*.

PROPOSITION 1. *(a) the GDP with $G = 2$ is weakly NP-complete, and (b) the GDP is strongly NP-complete.*

The proof is standard and follows from a reduction from the partition problem for part (a) and a reduction from the 3-partition problem in part (b) (see Appendix A). We note that the GDP is an easier, special case of the GSP because it considers only one realization of day-types; hence, the GSP is significantly more difficult.

3. Fixed-Length Policy

Due to the ease of implementation, a policy that is used frequently is the *fixed-length policy* (FLP). According to this policy, the duration of calls are equal and fixed to L' hours, where $0 \leq L' \leq L$. Under the FLP, the energy companies always contract with their customers for $H = KL'$ hours. In fact, the constraint on the total hours of energy-consumption reduction is no longer necessary. We refer to this problem as the *fixed-length stochastic problem* (FSP), and it includes only the constraint on the total number of calls per group.

We aggregate the state variables, solve the low-dimensional problem, and disaggregate the solution. This approach is proven to find an optimal solution for the FSP. The aggregate version of the FSP is refered to as the *aggregate fixed-length stochastic problem* (AFSP). Let $\mathcal{K}_d$ denote the total number of calls that we want to assign to groups on day d, and $\mathscr{K}_d$ denote the remaining number

of calls at the beginning of day d. Obviously, $\mathscr{K}_1 = GK$. On each day d, our decision must satisfy

$\mathcal{K}_d \leq \mathscr{K}_d$ because we have at most $\mathscr{K}_d$ calls to use for the rest of the planning horizon $d, d+1, \ldots, D$.

Moreover, $\mathcal{K}_d \leq G$, for each day d, because G groups exist and each group can be called at most once

a day. We define the set of feasible decisions for the AFSP as $\boldsymbol{\Theta}_d^{\mathrm{AF}} := \{\mathcal{K}_d \in \mathbb{Z}_+ | \mathcal{K}_d \leq \min\{\mathscr{K}_d, G\}\}$,

for all d. Let $\psi^{\mathrm{AF}}(\mathcal{K}_d, \omega_d)$ be the optimal cost of using $\mathcal{K}_d$ on day d. The dynamic programming

(DP) formulation of AFSP is as follows:

$$\mathcal{V}_d^{\mathrm{AF}}(\mathscr{K}_d) = \min_{\mathcal{K}_d \in \boldsymbol{\Theta}_d^{\mathrm{AF}}} \{\psi^{\mathrm{AF}}(\mathcal{K}_d, \omega_d) + \mathbf{E}\left[\mathcal{V}_{d+1}^{\mathrm{AF}}(\mathscr{K}_d - \mathcal{K}_d)\right]\}, \ \forall d = 1, \cdots, D-1,$$

$$\mathcal{V}_D^{\mathrm{AF}}(\mathscr{K}_D) = \min_{\mathcal{K}_D \in \boldsymbol{\Theta}_D^{\mathrm{AF}}} \{\psi^{\mathrm{AF}}(\mathcal{K}_D, \omega_D)\}, \tag{8}$$

where, $\psi^{\mathrm{AF}}(\mathcal{K}_d, \omega_d)$ is the optimal cost of using $\mathcal{K}_d$ calls on day d knowing that the day-type is ω_d.

We next formulate a mathematical model to determine $\psi^{\mathrm{AF}}(\mathcal{K}_d, \omega_d)$. This problem is referred to as

the *aggregate fixed-length single day problem* (AFSDP). Let integer variables s_t and R_t, for each t,

indicate the number of groups that start reducing their energy consumption at period t, and the

ECP height at period t after customers reduce their energy consumption, respectively. The AFSDP

is formulated as follows:

$$\psi^{\mathrm{AF}}(\mathcal{K}_d, \omega_d) = \min \ \sum_{t=1}^{T} f(R_t), \tag{9}$$

$$\text{s.t.} \ \ R_t \geq r_t(\omega_d) - \sum_{t'=\max\{1,t-L'+1\}}^{\min\{T-L'+1,t\}} s_{t'}, \ \forall t \tag{10}$$

$$\sum_{t=1}^{T-L'+1} s_t = \mathcal{K}_d \tag{11}$$

$$s_t, R_t \in \mathbb{Z}_+, \ \forall t. \tag{12}$$

The summation in constraint (10) represents the number of groups that are on call at time t.

Constraint (11) ensures that the total number of groups reducing their consumption on day d is

equal to $\mathcal{K}_d$.

Algorithm 1 presents our approach for solving the FSP. We first solve the AFSDP for all $0 \leq \mathcal{K}_d \leq$

GK and $\omega_d \in \Omega$ and we then solve the AFSP. For each day d, we find a set of calls that is denoted

by $\mathcal{D}_d$. We assign the calls to the groups that have been called the fewest in days $1, \ldots, d-1$. Note that the starting times of calls are known; hence, when a call is assigned to a group, the group is called starting from that time.

Algorithm 1 Solving the FSP.

  1: Solve the AFSDP for all $(\mathcal{K}_d, \omega_d)$ s.t. $0 \leq \mathcal{K}_d \leq GK$ and $\omega_d \in \Omega$

  2: Solve AFSP which results in $\mathcal{D}_d$, for all d

  3: Define $TC_g := 0$, for all g

  4: **for** $d = 1, \cdots, D$ **do**

  5:      Sort TC_g's in a non-decreasing order: $TC_{[1]} \leq TC_{[2]} \leq \cdots \leq TC_{[G]}$

  6:      Define $\hat{g} := 1$

  7:      **for** each call in $\mathcal{D}_d$ **do**

  8:          Assign the call to group $[\hat{g}]$

  9:          $TC_{[\hat{g}]} := TC_{[\hat{g}]} + 1$

10:          $\hat{g} := \hat{g} + 1$

11:      **end for**

12: **end for**.

THEOREM 1. *Algorithm 1 finds an optimal solution for the FSP in a polynomial time.*

Next, we present some lemmas that are used in the proof of theorem 1. The following remark presents an alternative way of formulating the AFSDP.

REMARK 1. Define $\hat{f}_{d,t} : \{z \in \mathbb{Z} | 0 \leq z \leq r_t(\omega_d)\} \longrightarrow \mathbb{R}$ as $\hat{f}_{d,t}(z) = f(r_t(\omega_d)) - f(z)$. Define the AFSDP* as $\min \sum_{t=1}^{T} \sum_{g=1}^{G} \partial \hat{f}_{d,t}(g) u_{gt}$ subject to $r_t(\omega_d) - R_t = \sum_{g=1}^{G} u_{gt}$, for all t, and Eqs. (10), (11), and (12), where $u_{gt} \in \{0, 1\}$, for all g, t. The AFSDP* and the AFSDP are equivalent.

Here, $\partial \hat{f}_{d,t}(z) = \hat{f}_{d,t}(z) - \hat{f}_{d,t}(z-1)$ for $z \in \mathbb{Z}$, $z \geq 1$. We note that $\hat{f}_{d,t}$ is a concave increasing function and $\hat{f}_{d,t}(z)$ measures the saving at period t of day d given that the ECP height has reduced to z. Finally, note that the objective function of the AFSDP* is a piecewise linear function.

Consider the linear programming (LP) relaxation of the AFSDP*, which is obtained by replacing constraints (12) with $s_t, R_t \in \mathbb{R}_+$, for all t, and replacing constraints $u_{gt} \in \{0, 1\}$ with $0 \leq u_{gt} \leq 1$, for all g, t. We denote the LP relaxation by AFSDP^r.

PROPOSITION 2. *The AFSDP^r solves the AFSDP*.*

LEMMA 1. *In Algorithm 1, at each iteration d of the* for*-loop, such that $d \leq D - 1$, we have $|TC_{g'} - TC_{g''}| \leq 1$ for each pair of groups g', g''.*

See Appendix B for the proofs. Based on Proposition 2, the AFSDP* is solved in a polynomial time. Moreover, according to Lemma 1, Algorithm 1 assigns calls to groups in an almost balanced fashion, for all $d \leq D - 1$.

LEMMA 2. *The solution of Algorithm 1 satisfies $TC_g \leq K$, for all g.*

Lemma 2 states that the solution that we obtain by applying Algorithm 1 is feasible.

PROOF OF THEOREM 1. The AFSP is a relaxation of the FSP because the AFSP does not consider group identities. This means that a feasible solution for the FSP is always feasible for the AFSP; however, a feasible solution for the AFSP may not be feasible for the FSP. Therefore, to prove the optimality of the solution, it suffices to show that the obtained solution is feasible for the FSP. For this, we must prove that the use of *for*-loop for assigning the solution of each day $\mathcal{D}_d$ to groups does not violate the maximum number of calls per group constraint. This is proven in Lemma 2. Thus, Algorithm 1 finds an optimal solution for the FSP. Note that once the solution of the AFSDP for all $0 \leq \mathcal{K}_d \leq GK$ and $\omega_d \in \Omega$ is given, which is found in a polynomial time (see Proposition 2), the rest of the algorithm requires a polynomial time. Therefore, the proof of theorem 1 is complete.

4. General Stochastic Problem (GSP)

Solving the GSP is difficult because of the stochasticity and $2G$-dimensional state space. We present an effective heuristic approach, which is denoted by HGSP, to solve the GSP and show that, for a sufficiently large problem size, the error is negligible. The main steps of the HGSP are as follows: (1) aggregating state space to obtain the *aggregate general stochastic problem* (AGSP), (2) fixing the number of days of each type to their expected value, to obtain the *aggregate general deterministic problem* (AGDP), (3) solving the AGDP in a polynomial time using a DP algorithm, and, (4) disaggregating the AGDP solution.

To analyze the asymptotic behavior of our heuristic approach for sufficiently large D, K, and H, we define $n \in \mathbb{Z}$, $n \geq 1$ to represent the problem size. We consider a series of problems with nD days, nK calls, and nH hours per group for all $n \in \mathbb{Z}$, $n \geq 1$. We aim to show that our heuristic approach finds an asymptotically optimal solution to the GSP. Let $v_{\text{GSP}}^{(n)}$ denote the optimal saving for the GSP, $v_{\text{HGSP}}^{(n)}$ denote the HGSP saving, $v_{\text{AGDP}}^{(n)}$ denote the optimal saving for the AGDP, and $v_{\text{AGSP}}^{(n)}$ denote the optimal saving for the AGSP, where the problem size is n.

THEOREM 2. $\mathbf{E}[v_{GSP}^{(n)}] - v_{HGSP}^{(n)} = \mathcal{O}(\sqrt{n})$ *for all $n \in \mathbb{Z}$, $n \geq 1$.*

Thus, as $n \longrightarrow \infty$, the relative error approaches to zero. The proof of theorem 2 follows from noting that the difference between $\mathbf{E}[v_{\text{GSP}}^{(n)}]$ and $v_{\text{HGSP}}^{(n)}$ includes two types of error—*information error* and *aggregation error*—and finding upper bounds on these errors. Information error refers to the error that is created from replacing random variables with their expected values.

THEOREM 3 (**Information Error**). $v_{AGDP}^{(n)} - \mathbf{E}[v_{AGSP}^{(n)}] = \mathcal{O}(\sqrt{n})$ *for all $n \in \mathbb{Z}$, $n \geq 1$.*

The proof of theorem 3 is deferred to Section 4.1. The second type of error in HGSP is called aggregation error, which is created from aggregating the state space in step (2).

THEOREM 4 (**Aggregation Error**). *There exists a non-negative constant ρ, independent of n, such that $v_{AGDP}^{(n)} - v_{HGSP}^{(n)} \leq \rho$ and $\mathbf{E}[v_{AGSP}^{(n)}] - \mathbf{E}[v_{GSP}^{(n)}] \leq \rho$, for all $n \in \mathbb{Z}$, $n \geq 1$.*

The proof of theorem 4 is deferred to Section 5.3. Theorem 4 states that there exists a constant upper bound on the aggregation error that does not increase in n; hence, the relative aggregation error is $\mathcal{O}(\frac{1}{n})$. Finally, the proof of theorem 2 follows from theorems 3 and 4.

4.1. Information error

Information error is referred to the difference between the optimal value of the AGDP and the expected optimal value of the AGSP. Here, the AGDP considers $p_\omega D$ days of type ω for each $\omega \in \Omega$, where p_ω denotes the proportion of days of type ω (the probability of day-type ω). We assume $p_\omega D$ is an integer for all ω, which is a reasonable assumption because $p_\omega D$ is the expected number of days of type ω. We show that the absolute information error is $\mathcal{O}(\sqrt{n})$.

In the AGSP, at the beginning of each day, the day-type is revealed. We want to determine the total number of calls and hours that should be used for that day. The AGSP is an special case of finite horizon revenue management problems (see, for example, Jasin and Kumar (2012), Cooper (2002)). The day-type resembles "customer type," and the total number of calls and hours that we use in a day resembles the "offers." Let "offers" be indexed by $j = 1, \ldots, J$. Different offers are distinguished by the number of calls, number of hours, and their day-types. In total, we have $J = |\Omega| G^2 L + 1$ different "offers" from which we can choose one to "make" in a day. We also include a no-offer scenario that explains the "+1" in the total number of available offers. Each offer can be used in only one day-type; hence, we use the notation $\omega(j)$ to denote the day-type in which offer j can be used. Once offer j is made, a saving of θ_j is realized with probability 1, where θ_j is found by solving the aggregate single day problem (see Section 5.1). We use index $\gamma = 1, \ldots, \Gamma$ to denote resources. In the AGSP, we have $\Gamma = 2$, which are the total number of calls and hours in the aggregate sense. Let $a_{\gamma j} \geq 0$ show the amount of resource γ used when offer j is made, and c_γ denote the total available resource γ at the beginning of the planning horizon. Let $\mathbf{a}_\gamma := (a_{\gamma 1}, \ldots, a_{\gamma J})$ and $a_{\gamma,\max} := \max_j \{a_{\gamma j}\}$ for all γ.

To analyze the asymptotic behavior of the error, we use superscript (n) to denote scaling by factor n, where $n \in \mathbb{Z}$, $n \geq 1$; hence, the AGSP$^{(n)}$ and the AGDP$^{(n)}$ respectively denote the AGSP and the AGDP with $c_\gamma^{(n)} = nc_\gamma$, for all γ, and $D^{(n)} = nD$. The superscript (n) is dropped whenever $n = 1$.

Our analysis of the information error adapts the approach of Jasin and Kumar (2012). They consider a network revenue management problem with uncertain customer choice, where the customer arrival is modeled via Poisson processes. They show that the expected revenue loss is bounded by a constant, independent of the problem size, when the problem is re-solved according to a pre-determined schedule. Whereas in our problem, the day-type, which is analogous to the customer type is modeled via multi-nomial distributions. We establish that the absolute information error is $\mathcal{O}(\sqrt{n})$, when the problem is solved once in the beginning of the planning horizon.

The LP relaxation of the AGDP$^{(n)}$ is denoted by the RAGDP$^{(n)}$, which is formulated as follows. Here, x_j denotes the number of times that offer j is made in the planning horizon (x_j can be fractional):

$$\max \quad \sum_j \theta_j x_j \tag{13}$$

$$\text{s.t.} \quad \sum_j a_{\gamma j} x_j \leq n c_\gamma, \ \forall \gamma, \tag{14}$$

$$\sum_{j: \omega(j)=\omega} x_j \leq n p_\omega D, \ \forall \omega, \tag{15}$$

$$x_j \in \mathbb{R}_+, \ \forall j. \tag{16}$$

Let the optimal solution of the RAGDP$^{(n)}$ be denoted by $\mathbf{x}^{*(n)} = (x_1^{*(n)}, \ldots, x_J^{*(n)})$. Moreover, let $v_{\text{RAGDP}}^{(n)}$ denote the optimal value of the RAGDP$^{(n)}$. Note that $\mathbf{x}^{*(n)} = n\mathbf{x}^*$ for all $n \geq 1$. Moreover, $v_{\text{RAGDP}}^{(n)}$ provides an upper bound on the optimal value of the AGDP$^{(n)}$.

We next present a simple heuristic approach to obtain a lower bound on the optimal value of the AGSP$^{(n)}$. Let $q_j := x_j^* / (p_{\omega(j)} D)$, for all j, be the allocation probability of offer j. Note that q_j does not need superscript because $q_j^{(n)} = q_j$ for all $n \geq 1$.

Algorithm 2, referred to as the LBH$^{(n)}$, finds a lower bound on the optimal value of the AGSP$^{(n)}$. If the condition in Line 3 is satisfied in day d, the LBH$^{(n)}$ selects offer $\hat{j}$ using probability $\{q_j\}$. This will consume some of the resources and generate some profit. If the condition in Line 3 is violated, the LBH$^{(n)}$ does not generate any profit.

Algorithm 2 LOWER-BOUND HEURISTIC$^{(n)}$ (LBH$^{(n)}$).

1:   Define $v_{\text{LBH}}^{(n)} := 0$, and $\tilde{c}_\gamma := n c_\gamma$, for all γ

2:   **for** $d = 1, \cdots, nD$ **do**

3:      **if** $\tilde{c}_\gamma \leq a_{\gamma,\max}$ for all γ **then**

4:         Select offer j using allocation probability $\{q_j\}$, and denote it by $\hat{j}$

5:         Update $v_{\text{LBH}}^{(n)} := v_{\text{LBH}}^{(n)} + \theta_{\hat{j}}$, and $\tilde{c}_\gamma := \tilde{c}_\gamma - a_{\gamma\hat{j}}$, for all γ

6:      **end if**

7:   **end for**

8:   **return** $v_{\text{LBH}}^{(n)}$.

We introduce an upper-bound heuristic, given in Algorithm 3, in a similar fashion, ignoring resource constraints. In each day, the $\text{UBH}^{(n)}$ selects an offer j with probability $\{q_j\}$ and realizes the corresponding profit θ_j.

Algorithm 3 UPPER-BOUND HEURISTIC$^{(n)}$ ($\text{UBH}^{(n)}$).

1: Define $v_{\text{UBH}}^{(n)} := 0$

2: **for** $d = 1, \cdots, nD$ **do**

3:     Select offer j using allocation probability $\{q_j\}$, and denote it by $\hat{j}$

4:     Update $v_{\text{UBH}}^{(n)} := v_{\text{UBH}}^{(n)} + \theta_{\hat{j}}$

5: **end for**

6: **return** $v_{\text{UBH}}^{(n)}$.

LEMMA 3. $v_{AGDP}^{(n)} - \mathbf{E}[v_{AGSP}^{(n)}] \leq \mathbf{E}[v_{UBH}^{(n)}] - \mathbf{E}[v_{LBH}^{(n)}]$, *for all* $n \in \mathbb{Z}$, $n \geq 1$.

The proof follows from $\mathbf{E}[v_{\text{UBH}}^{(n)}] = v_{\text{RAGDP}}^{(n)} \geq v_{\text{AGDP}}^{(n)} \geq \mathbf{E}[v_{\text{AGSP}}^{(n)}] \geq \mathbf{E}[v_{\text{LBH}}^{(n)}]$, for all $n \in \mathbb{Z}$, $n \geq 1$, and is given in Appendix C. In the remainder, we find an upper bound on $\mathbf{E}[v_{\text{UBH}}^{(n)}] - \mathbf{E}[v_{\text{LBH}}^{(n)}]$ that is also valid as an upper bound for the information error.

For any realization of the day-types, the $\text{LBH}^{(n)}$ and $\text{UBH}^{(n)}$ produce equal profits on all days as long as the condition in Line 3, Algorithm 2, is satisfied. Let $d^{*(n)}$ denote the minimum of nD and the last day in which this condition is satisfied. The expected value of profit that is then produced by the $\text{LBH}^{(n)}$ and $\text{UBH}^{(n)}$ during days $1, 2, \ldots, d^{*(n)}$ is equivalent. In days $d^{*(n)} + 1, \ldots, nD$, the $\text{UBH}^{(n)}$ produces profit but the $\text{LBH}^{(n)}$ does not. Hence, $\mathbf{E}[v_{\text{UBH}}^{(n)}] - \mathbf{E}[v_{\text{LBH}}^{(n)}] \leq \theta_{\max} \mathbf{E}[nD - d^{*(n)}]$, where $\theta_{\max} := \max_j \{\theta_j\}$. We note that $\mathbf{E}[d^{*(n)}] = \sum_{d=1}^{nD} P(d^{*(n)} \geq d)$. Therefore, $\mathbf{E}[nD - d^{*(n)}] = \sum_{d=1}^{nD} P(d^{*(n)} < d)$, and we have:

$$\mathbf{E}[v_{\text{UBH}}^{(n)}] - \mathbf{E}[v_{\text{LBH}}^{(n)}] \leq \theta_{\max} \sum_{d=1}^{nD} P(d^{*(n)} < d), \quad n \in \mathbb{Z}, \ n \geq 1. \tag{17}$$

Thus, it remains to obtain an upper bound on $P(d^{*(n)} < d)$, for all n and d. On each day, after knowing the day-type, the $\text{UBH}^{(n)}$ makes an offer based on probability $\{q_j\}$ and then some of the resources are utilized. Although the expected value of the resource usage is $\frac{\mathbf{a}_\gamma \mathbf{x}^*}{D}$, for all γ and d, the

actual utilization of resource γ on day d is $\frac{\mathbf{a}_\gamma \mathbf{x}^*}{D} + \Delta_\gamma^{(n)}(d)$, where $\Delta_\gamma^{(n)}(d)$ is the deviation from the expected utilization. We note that $\Delta_\gamma^{(n)}(d)$ is a discrete random variable that follows a multi-nomial distribution with 1 trial and event probability $\{p_{\omega(j)} q_j\}$. Let $\Im_d$ denote the accumulated information up to day d. Then, $\{\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')\}_{d=1}^{nD}$ and $\{|\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')|\}_{d=1}^{nD}$ are respectively martingale and submartingale with respect to filtration $\{\Im_d\}$. Note that $\mathbf{E}[\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')|\{\Im_d\}] = 0$, for all n and d. The following lemma provides an upper bound on the second moment of $\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')$ conditional on the filtration $\{\Im_d\}$.

LEMMA 4. $\mathbf{E}[(\sum_{d'=1}^d \Delta_\gamma^{(n)}(d'))^2|\{\Im_d\}] \leq \frac{1}{4} a_{\gamma,\max}^2 d$, *for all γ and d.*

We next provide an intuition for the information error being $\mathcal{O}(\sqrt{n})$ by comparing $n = 1$ and $n = 2$ (see Fig. 4). Consider $P(|\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')| > \kappa) = \vartheta$, where ϑ is a significantly small positive number. For this equation to hold for all d, κ must be a linear function of the standard deviation of $|\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')|$. Hence, using Lemma 4, κ must be a linear function of $\sqrt{d}$. Therefore, in Fig. 4, the shaded area indicates a region where $|\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')|$ would fall with probability $1 - \vartheta$. On the other hand, consider a resource γ for which the corresponding resource constraint is binding (a similar argument can be made for the non-binding constraints). Thus, the expected value of the remaining resource γ decreases linearly in d, from nc_γ to 0. We can roughly assume that the LBH$^{(n)}$ stops offering when the expected value of the remaining resource γ hits the shaded region. Moreover, the information error is related to the days after this hitting time. The number of days after the hitting time for $n = 1$ and $n = 2$ are $\tilde{D}$ and $\sqrt{2}\tilde{D}$, respectively. Therefore, the information error is $\mathcal{O}(\sqrt{n})$. We next rigorously prove this result.

Given that the LBH$^{(n)}$ makes an offer on day d, the utilization of γ during days $1, 2, \ldots, d$ is given by $\frac{d}{D} \mathbf{a}_\gamma \mathbf{x}^* + \sum_{d'=1}^d \Delta_\gamma^{(n)}(d')$. Recall that, for all γ, the total available resource γ at the beginning of the planning horizon is nc_γ. Thus, the condition for LBH$^{(n)}$ to make an offer on day $d+1$ is as follows:

$$\text{CONDITION } (\dagger): \quad \sum_{d'=1}^d \Delta_\gamma^{(n)}(d') \leq nc_\gamma - \frac{d}{D} \mathbf{a}_\gamma \mathbf{x}^* - a_{\gamma,\max}, \text{ for all } \gamma.$$

In the following lemma, we use condition $(\dagger)$ and Doob's submartingale inequality to present an upper bound on $P(d^{*(n)} < d)$, for all n and d.
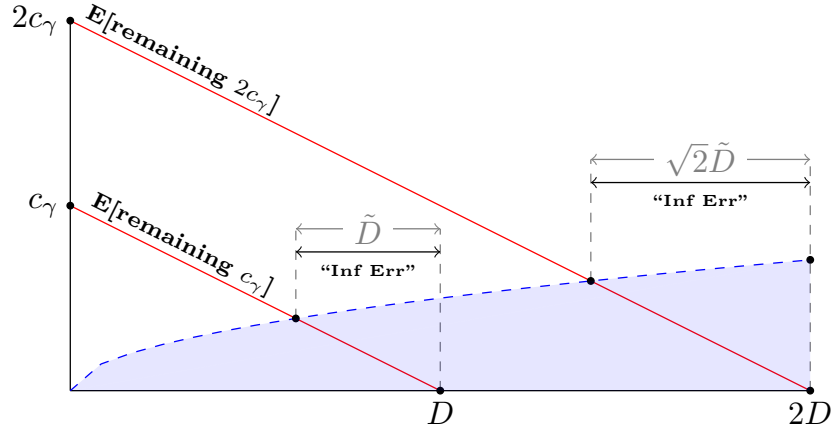
Figure 4   Intuition for the absolute information error being $\mathcal{O}(\sqrt{n})$.

LEMMA 5. $P(d^{*(n)} < d) \leq \min\left\{1, \sum_\gamma d\left(\frac{2n}{a_{\gamma,\max}}(c_\gamma - \frac{d}{nD}\mathbf{a}_\gamma \mathbf{x}^*) - 2\right)^{-2}\right\}$, *for all n and d.*

Note that $a_{1,\max} = G$, $a_{2,\max} = GL$, $c_1 = GK$, and $c_2 = GH$. Moreover, $\mathbf{a}_1 \mathbf{x}^*$ and $\mathbf{a}_2 \mathbf{x}^*$ are respectively the utilized number of calls and hours, based on the solution of RAGDP. Using Lemmas 3 and 5, and Eq. (17), we obtain the following upper bound on the information error, for all $n \in \mathbb{Z}$, $n \geq 1$.

$$v_{\mathrm{AGDP}}^{(n)} - \mathbf{E}[v_{\mathrm{AGSP}}^{(n)}] \leq \theta_{\max} \sum_{d=1}^{nD} \min\left\{1, \sum_\gamma d\left(\frac{2n}{a_{\gamma,\max}}(c_\gamma - \frac{d}{nD}\mathbf{a}_\gamma \mathbf{x}^*) - 2\right)^{-2}\right\}. \qquad (18)$$

The right-hand-side of Eq. (18) provides a tight upper bound that can be calculated for any given instance. However, it may not be obvious why the right-hand-side is $\mathcal{O}(\sqrt{n})$. Next, we simplify the upper bound in Eq. (18). (See Fig. 5 for the graphical illustration.) Let the planning horizon consist of two parts: $nD - \sqrt{n}D$ days starting from the beginning of the horizon, and the remaining $\sqrt{n}D$ days. Using Lemma 3 and Eq. (17), we have:

$$v_{\mathrm{AGDP}}^{(n)} - \mathbf{E}[v_{\mathrm{AGSP}}^{(n)}] \leq \theta_{\max} \sum_{i=1}^{\lceil n-\sqrt{n}\rceil} \sum_{d=(i-1)D+1}^{iD} P(d^{*(n)} < d) + \theta_{\max} D\sqrt{n},$$

for all $n \in \mathbb{Z}$, $n \geq 1$. Here, we use $\lceil n - \sqrt{n}\rceil$ instead of $n - \sqrt{n}$, which is valid because we increase the upper bound. Moreover, for the last $D\sqrt{n}$ days, we use 1 as the upper bound for $P(d^{*(n)} < d)$. Using Lemma 5, we have:

$$v_{\mathrm{AGDP}}^{(n)} - \mathbf{E}[v_{\mathrm{AGSP}}^{(n)}] \leq \theta_{\max} \sum_{i=1}^{\lceil n-\sqrt{n}\rceil} \sum_{d=(i-1)D+1}^{iD} \sum_\gamma \frac{\frac{1}{4}d a_{\gamma,\max}^2}{\left(nc_\gamma - \frac{d}{D}\mathbf{a}_\gamma \mathbf{x}^* - a_{\gamma,\max}\right)^2} + \theta_{\max} D\sqrt{n}$$

18

Fattahi, Dasu and Ahmadi: *Peak Load Energy Management*
Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)
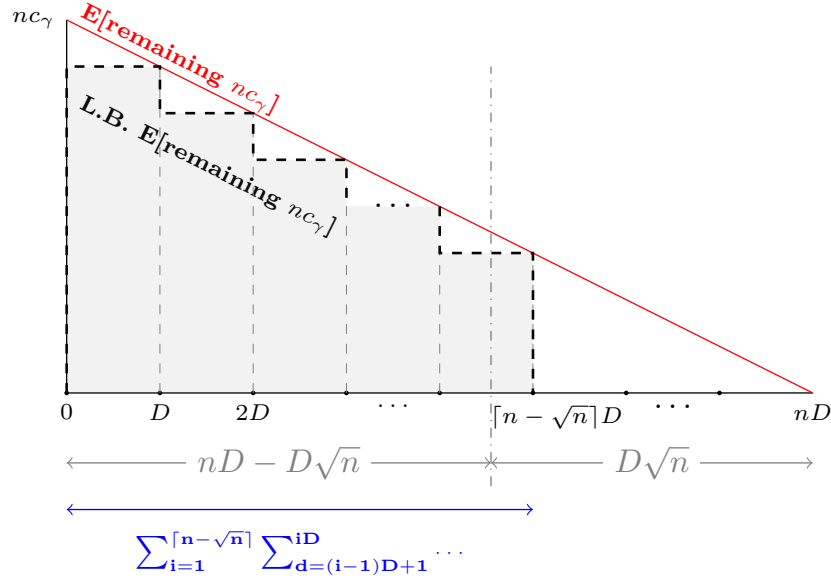
Figure 5   A graphical illustration of simplifying the upper bound given in Eq. (18).

$$\leq \theta_{\max} \sum_{i=1}^{\lceil n - \sqrt{n} \rceil} \sum_{d=(i-1)D+1}^{iD} \sum_{\gamma} \frac{\frac{1}{4} d a_{\gamma,\max}^2}{\left(n c_\gamma - \frac{iD}{D} c_\gamma - c_\gamma \right)^2} + \theta_{\max} D \sqrt{n},$$

for all $n \in \mathbb{Z}$, $n \geq 1$. We replace $\mathbf{a}_\gamma \mathbf{x}^*$ by c_γ, since $\mathbf{a}_\gamma \mathbf{x}^* \leq c_\gamma$, for all γ, due to the feasibility of $\mathbf{x}^*$. We also replace $a_{\gamma,\max}$ by c_γ, because logically, $a_{\gamma,\max} \leq c_\gamma$, for all γ. Moreover, in the denominator, we replace d by D, which corresponds to the lower bound on the expected remaining resource γ that is shown in Fig. 5. Finally, we have:

$$v_{\mathrm{AGDP}}^{(n)} - \mathbf{E}[v_{\mathrm{AGSP}}^{(n)}] \leq \frac{1}{4} \theta_{\max} \sum_{\gamma} \frac{a_{\gamma,\max}^2}{c_\gamma^2} \sum_{i=1}^{\lceil n-\sqrt{n} \rceil} \frac{1}{(n-i-1)^2} \sum_{d=(i-1)D+1}^{iD} d + \theta_{\max} D \sqrt{n}$$

$$\leq \frac{1}{4} \theta_{\max} D^2 \sum_{\gamma} \frac{a_{\gamma,\max}^2}{c_\gamma^2} \sum_{i=1}^{\lceil n-\sqrt{n} \rceil} \frac{i}{(n-i-1)^2} + \theta_{\max} D \sqrt{n} \tag{19}$$

$$\leq \frac{1}{4} \theta_{\max} D^2 \sum_{\gamma} \frac{a_{\gamma,\max}^2}{c_\gamma^2} \int_{i=1}^{n-\sqrt{n}+2} \frac{i}{(n-i-1)^2} di + \theta_{\max} D \sqrt{n} \tag{20}$$

$$= \frac{1}{4} \theta_{\max} D^2 \sum_{\gamma} \frac{a_{\gamma,\max}^2}{c_\gamma^2} \mathcal{O}(\sqrt{n}) + \theta_{\max} D \sqrt{n}. \tag{21}$$

Eq. (19) follows from $\sum_{d=(i-1)D+1}^{iD} d = iD^2 - \frac{1}{2}(D^2 - D)$ and ignoring $\frac{1}{2}(D^2 - D)$. In Eq. (20), we use an integral that is an upper bound for $\sum_{i=1}^{\lceil n-\sqrt{n} \rceil} \frac{i}{(n-i-1)^2}$. Eq. (21) represents the result of some

standard calculations and shows that the integral in Eq. (20) is $\mathcal{O}(\sqrt{n})$. Therefore, it follows that

$$v_{\text{AGDP}}^{(n)} - \mathbf{E}[v_{\text{AGSP}}^{(n)}] = \mathcal{O}(\sqrt{n}).$$

5. General Deterministic Problem (GDP)

We use a DP approach to formulate the *aggregate general deterministic problem*. The state variables are the total remaining time and the total number of available calls, which are denoted by $\mathscr{H}_d$ and $\mathscr{K}_d$, respectively. At the beginning of the planning horizon $d = 1$, we have $\mathscr{H}_1 = GH$ and $\mathscr{K}_1 = GK$. Let $\mathcal{H}_d$ and $\mathcal{K}_d$ denote the total energy consumption reduction time and the number of calls in day d, respectively. The set of feasible vectors $(\mathcal{H}_d, \mathcal{K}_d)$ is defined as $\mathbf{\Theta}_d^{\text{AD}} := \left\{ (\mathcal{H}_d, \mathcal{K}_d) \in \mathbb{Z}_+^2 \mid \mathcal{H}_d \leq \min\{\mathscr{H}_d, GL\}, \ \mathcal{K}_d \leq \min\{\mathscr{K}_d, G\} \right\}$, for all d. After using $\mathcal{H}_d$ and $\mathcal{K}_d$ in day d, the remaining time and calls become $\mathscr{H}_{d+1} = \mathscr{H}_d - \mathcal{H}_d$ and $\mathscr{K}_{d+1} = \mathscr{K}_d - \mathcal{K}_d$, respectively. The DP formulation for the AGDP is as follows:

$$
\begin{aligned}
\mathcal{V}_d^{\text{AD}}(\mathscr{H}_d, \mathscr{K}_d) &= \min_{(\mathcal{H}_d, \mathcal{K}_d) \in \mathbf{\Theta}_d^{\text{AD}}} \left\{ \psi^{\text{AD}}(\mathcal{H}_d, \mathcal{K}_d, \omega_d) + \mathcal{V}_{d+1}^{\text{AD}}(\mathscr{H}_d - \mathcal{H}_d, \mathscr{K}_d - \mathcal{K}_d) \right\}, \ \forall d = 1, \cdots, D-1, \\
\mathcal{V}_D^{\text{AD}}(\mathscr{H}_D, \mathscr{K}_D) &= \min_{(\mathcal{H}_D, \mathcal{K}_D) \in \mathbf{\Theta}_D^{\text{AD}}} \left\{ \psi^{\text{AD}}(\mathcal{H}_D, \mathcal{K}_D, \omega_D) \right\},
\end{aligned}
\tag{22}
$$

where $\mathcal{V}_d^{\text{AD}}(\mathscr{H}_d, \mathscr{K}_d)$ is the optimal cost of using $\mathscr{H}_d$ and $\mathscr{K}_d$ in days $d, d+1, \cdots, D$. The stage cost $\psi^{\text{AD}}(\mathcal{H}_d, \mathcal{K}_d, \omega_d)$ is the optimal value of the *aggregate single day problem* (ASDP; see Section 5.1), which is defined as follows. Assign $\mathcal{H}_d$ units of energy consumption reduction time to consumers by, at most, $\mathcal{K}_d$ calls to achieve the minimum energy cost on day d. The state space of the AGDP consists of 2 non-negative integer variables, $\mathscr{H}_d$ and $\mathscr{K}_d$, that are upper bounded by GH and GK, respectively. Therefore, given the solution of the ASDP, the AGDP is solved in a polynomial time.

5.1. Aggregate single day problem (ASDP)

Given $\mathcal{H}_d$ and $\mathcal{K}_d$, the ASDP minimizes the total cost in day d by using, at most, $\mathcal{K}_d$ calls, and with a total time $\leq \mathcal{H}_d$. Day-type ω_d must be also given as it determines the shape of the ECP for day d. The ASDP is aggregate in the sense that it ignores the group identities and does not assign calls to groups.

Recall that a group can be called for at most L units of time during day d. In addition, it is not logical to call a group for more than $T - t + 1$ units starting from period t. Let $\mathcal{L}(t) := \min\{L, T - t + 1\}$ denote the maximum duration of a call starting from t. We define non-negative integer variables $s_{t\ell}$, for all t, ℓ, $1 \le \ell \le \mathcal{L}(t)$, as the number of calls with duration ℓ and starting time t. The ASDP is formulated as follows:

$$\psi^{\mathrm{AD}}(\mathcal{H}_d, \mathcal{K}_d, \omega_d) = \min \ \sum_{t=1}^{T} f\left(\left[r_t(\omega_d) - \sum_{i=1}^{\min\{L,t\}} \sum_{\ell=i}^{\mathcal{L}(t)} s_{(t-i+1)\ell}\right]^+\right) \tag{23}$$

$$\text{s.t.} \ \sum_{t=1}^{T} \sum_{\ell=1}^{\mathcal{L}(t)} s_{t\ell} \le \mathcal{K}_d, \tag{24}$$

$$\sum_{t=1}^{T} \sum_{\ell=1}^{\mathcal{L}(t)} \ell s_{t\ell} \le \mathcal{H}_d, \tag{25}$$

$$s_{t\ell} \in \mathbb{Z}_+, \ \forall t, \ell, \ 1 \le \ell \le \mathcal{L}(t). \tag{26}$$

In the objective function, $\sum_{i=1}^{\min\{L,t\}} \sum_{\ell=i}^{\mathcal{L}(t)} s_{(t-i+1)\ell}$ indicates the reduction in the ECP height in period t. Eqs. (24) and (25) guarantee that the number of calls and the total reduction time are upper bounded by $\mathcal{K}_d$ and $\mathcal{H}_d$, respectively. The ASDP is non-linear; however, as we state in the following remark, it can be formulated as a MILP problem.

REMARK 2. Define the ASDP* as $\min \sum_{z=1}^{Z} \{\partial^2 f(z)(\sum_{t=1}^{T} \varepsilon_{zt})\}$ subject to:

$$r_t(\omega_d) - z + 1 - \left(\sum_{i=1}^{\min\{L,t\}} \sum_{\ell=i}^{\mathcal{L}(t)} s_{(t-i+1)\ell}\right) \le \varepsilon_{zt}, \ \forall z, t, \tag{27}$$

$$\varepsilon_{zt} \ge 0, \ \forall z, \ t, \tag{28}$$

and Eqs. (24), (25), and (26). The ASDP* and the ASDP are equivalent.

The explanation is given in Appendix D.1. Our experiments show that the Cplex MIP solver is quite efficient for solving the ASDP. Another approach for solving the ASDP is to use a DP algorithm, which also shows that the ASDP can be solved in a polynomial time.

One could initially solve the ASDP for all $(\mathcal{H}_d, \mathcal{K}_d, \omega_d)$ such that $0 \le \mathcal{H}_d \le GH$, $0 \le \mathcal{K}_d \le GK$, and $\omega_d \in \Omega$. Thus, the AGDP is solved using a DP algorithm in a polynomial time. We next show that the AGDP solution can be disaggregated and assigned to groups with a constant absolute disaggregation error.

5.2. Assigning calls to groups

Let $\mathcal{D}_d$, for all d, denote the set of calls in day d based on the AGDP solution. A call, $\eta \in \mathcal{D}_d$, is in the form of $\eta = (st(\eta), |\eta|)$ where $st(\eta) \in \mathbb{Z}_+$ is the starting time and $|\eta| \in \mathbb{Z}_+$ is the length of the call. For example, $\mathcal{D}_1 = \{(1,1),(1,4)\}$ means that two groups must be called starting from $t = 1$ on day 1: one group for 1 hour and the other group for 4 hours. However, it is not known which two groups must reduce their consumption, hence, $\mathcal{D}_1 = \{(1,1),(1,4)\}$ can be assigned to G groups in $G(G-1)$ different ways. In general, given $\mathcal{D}_d$, for all d, there are $\prod_{d=1}^{D} \frac{G!}{(G-|\mathcal{D}_d|)!}$ possible assignments. We also note that a feasible assignment of calls to groups may not exist. We next present a more efficient approach for assigning calls to groups. The following example shows our approach.

EXAMPLE 1. Consider $D = 5$, $G = 4$, $K = 4$, $L = 4$, and $H = 16$. Assume that the AGDP solution is given by: $\mathcal{D}_1 = \{(1,1),(1,4),(1,3)\}$, $\mathcal{D}_2 = \{(1,2),(1,3),(1,1),(1,4)\}$, $\mathcal{D}_3 = \{(1,1)\}$, $\mathcal{D}_4 = \{(1,4),(1,4),(1,4),(1,4)\}$, and $\mathcal{D}_5 = \{(1,1),(1,2),(1,4),(1,4)\}$. Note that the starting times of all calls are $t = 1$. The starting times of calls do not play any role in the disaggregation step, as each group is assigned one call in a day at most. Once a call is assigned to a group, they can start reducing their consumption any time during that day.

We sort the calls in a non-increasing length order as shown in Fig. 6. The length of a bar shows the duration of the corresponding call. The first number below the bar indicates the corresponding day for the call. We divide the calls into classes of size G. This results in 4 classes, which are denoted by $\varsigma = 1, 2, 3, 4$. We assign one call from each class to each group. This assignment is always possible since each class has G calls. An important constraint is that, at most, one call from a day can be assigned to a group, as without this constraint, the assignment would be straightforward. In Fig. 6, each group is assigned exactly four calls. The second number below the bar shows the corresponding group for the call. In the following, we formally state this approach and discuss its properties.

Assume the AGDP is solved and outputs $\mathcal{D}_d$, for all d, and let $\mathbf{D} := \bigcup_{d=1}^{D} \mathcal{D}_d$. Sort the members of $\mathbf{D}$ in a non-increasing length order. If the total number of calls, $|\mathbf{D}|$, is less than GK, add $\eta = (1,0)$
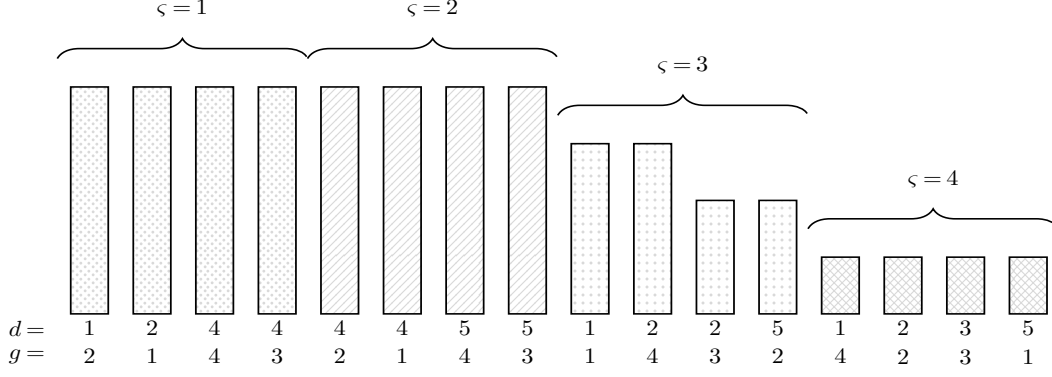
Figure 6   An example for assigning calls to groups.

to the end of the sorted list so that the number of calls become GK. We refer to the first G calls

as class 1, the second G calls as class 2, and so on. Denote the classes by $\varsigma = 1, 2, \cdots, K$. Define

binary variables $\lambda_{i\varsigma g}$, for all i, ς, g, such that $\lambda_{i\varsigma g} = 1$ if call i from class ς is assigned to group g,

and $\lambda_{i\varsigma g} = 0$ otherwise. We obtain an assignment of calls to groups by solving the following system,

which we refer to as the *disaggregation problem* (DAP):

$$\sum_{g=1}^{G} \lambda_{i\varsigma g} = 1, \ \forall i, \varsigma, \tag{29}$$

$$\sum_{i=1}^{G} \lambda_{i\varsigma g} = 1, \ \forall \varsigma, g, \tag{30}$$

$$\sum_{i\varsigma \in \mathcal{D}_d} \lambda_{i\varsigma g} \leq 1, \ \forall g, \mathcal{D}_d, \tag{31}$$

$$\lambda_{i\varsigma g} \in \{0, 1\}, \ \forall i, \varsigma, g. \tag{32}$$

Eqs. (29)–(30) ensure a 1-to-1 assignment of calls to groups for each class. Eq. (31) guarantees

that each group g can receive one call for each day d at most. The condition $i\varsigma \in \mathcal{D}_d$ means that if

the ith call from class ς belongs to day d.

Let TH_g, for all g, denote the sum of the call durations that are assigned to group g, which

are calculated using the DAP solution. Observe that, in example 1, $TC_g = 4$, for all g. Moreover,

$TH_1 = TH_4 = 12$ and $TH_2 = TH_3 = 11$. The following proposition presents a property of the DAP

solution (the proof is given in Appendix D.2).

PROPOSITION 3. *Consider TH_g and TC_g, for all g, that are calculated using the DAP solution.*

Then: (a) $|TC_{g'} - TC_{g''}| \leq 1$, for all g', g'', and (b) $|TH_{g'} - TH_{g''}| \leq L$, for all g', g''.

Let $\mathcal{G}_g$, for all g, denote the set of calls that are assigned to group g based on the DAP solution. Recall that, at most, K calls are assigned to each group because there are K classes. Thus, the solution satisfies the *maximum call per group* constraint because $|\mathcal{G}_g| \leq K$ for all g.

The total consumption reduction time for group g is TH_g. We note that the *total time per group* constraint may be violated. If $TH_g > H$, for some g, we find $\eta \in \mathcal{G}_g$ that corresponds to the minimum saving, and reduce its length. We repeat this process until $TH_g \leq H$ for all g. The main steps of our heuristic approach are summarised in Algorithm 4.

Algorithm 4 HGSP heuristic.

1: Solve the ASDP for all $(\mathcal{H}_d, \mathcal{K}_d, \omega_d)$ s.t. $0 \leq \mathcal{H}_d \leq GH$, $0 \leq \mathcal{K}_d \leq GK$, and $\omega_d \in \Omega$

2: Solve the AGDP, which results in $\mathcal{D}_d$, for all d

3: Let $\mathbf{D} := \bigcup_{d=1}^{D} \mathcal{D}_d$

4: Sort the members of $\mathbf{D}$ in a non-increasing order of their lengthes

5: Add zeros to the end of the sorted list so that the number of calls become GK

6: Denote the first G calls class 1, the second G calls class 2, and so on

7: Solve the DAP, which results in $\mathcal{G}_g$, for all g

8: **for** $g = 1, \cdots, G$ **do**

9:      **while** $TH_g > H$ **do**

10:          Reduce the length of $\eta \in \mathcal{G}_g$, which corresponds to the least amount of saving

11:      **end while**

12: **end for**

13: Return $\mathcal{G}_g$, for all g.

THEOREM 5. *The HGSP always finds a feasible solution to the GDP in a polynomial time.*

We next present some results that are used in the proof of theorem 5. We note that the effectiveness of the HGSP relies heavily on the following issues. First, the DAP must always have a feasible solution to guarantee that the HGSP can assign calls to groups. Second, the DAP must be solved in a polynomial time. Let the DAP^r denote the LP relaxation of the DAP, which is obtained by replacing Eq. (32) with $0 \leq \lambda_{i\varsigma g} \leq 1$, for all i, ς, g.

For a fixed g, let $\mathcal{B}$ denote the coefficient matrix for Eq. (30), $\mathcal{E}_1$ denote the coefficient matrix for Eq. (31) with $|\mathcal{D}_d| = G$, and $\mathcal{E}_2$ denote the coefficient matrix for Eq. (31) with $|\mathcal{D}_d| < G$. See Appendix D.3 for a detailed characterization of $\mathcal{B}$, $\mathcal{E}_1$, and $\mathcal{E}_2$.

PROPOSITION 4. *Let* $\Phi := \{\lambda \in \mathbb{R}^{KG} | \mathcal{B}\lambda = \mathbf{1}_K, \mathcal{E}_1\lambda = \mathbf{1}_{D_1}, \mathcal{E}_2\lambda \leq \mathbf{1}_{D_2}, \lambda \geq \mathbf{0}\}$. *Then,* Φ *is a non-empty and integral polyhedron.*

The proof is given in Appendix D.3. Thus, for a fixed g, one can find an integer solution in Φ in a polynomial time. In the obtained solution, one call from each class is assigned to g. By eliminating the assigned calls and group g, the number of groups decreases by 1 while the structure of the problem is preserved. Hence, by repeating this procedure G times, a feasible solution for the DAP is found in a polynomial time. Therefore, the proof of theorem 5 is complete.

5.3. Worst-case analysis for the aggregation error

Recall that if $TH_g \leq H$, for all g, then the obtained solution is optimal for the GDP; otherwise, we reduce the duration of some of the calls to achieve $TH_g \leq H$, for all g. Let ζ_{sum} denote the sum of the durations of all calls. Proposition 5 states a condition under which the HGSP is guaranteed to find an optimal solution for the GDP.

PROPOSITION 5. *If* $\zeta_{sum} \leq G(H - L + 1)$, *then the HGSP finds an optimal solution for the GDP.*

The proof is given in Appendix E.1. If the condition of Proposition 5 does not hold, the HGSP solution may be infeasible. Recall that if for any g, $TH_g > H$, we then reduce the duration of some of the calls that are assigned to group g.

Next, we prove theorem 4. According to Proposition 3, $|TH_{g'}^{(n)} - TH_{g''}^{(n)}| \leq L$, for all g', g''. Moreover, one could show, using a contradiction, there exists a group g' such that $TH_g^{(n)} \leq nH$. Hence, by some algebra, we have the following remark.

REMARK 3. $[TH_g^{(n)} - nH]^+ \leq L$, for all n and g.

Define $\delta := \max_\omega\{\psi^{\mathrm{AD}}(1,1,\omega)\}$; hence, δ is the maximum saving that one could achieve by assigning a call of length 1 to a group. Therefore, for all $n \in \mathbb{Z}$, $n \geq 1$, we have:

$$v_{\mathrm{AGDP}}^{(n)} - v_{\mathrm{HGSP}}^{(n)} \leq \delta \sum_g [TH_g^{(n)} - nH]^+$$

$$\leq \delta G \max_g \{[TH_g^{(n)} - nH]^+\}$$

$$\leq \delta GL.$$

This completes the first part of the proof. To prove the second part, note that for any realization of day-types, we have

$$v_{\mathrm{AGDP}}^{(n)} - v_{\mathrm{GDP}}^{(n)} \leq v_{\mathrm{AGDP}}^{(n)} - v_{\mathrm{HGSP}}^{(n)} \leq \delta GL,$$

for all $n \in \mathbb{Z}$, $n \geq 1$, because $v_{\mathrm{HGSP}}^{(n)} \leq v_{\mathrm{GDP}}^{(n)}$. Since this is true for any realization, then $\mathbf{E}[v_{\mathrm{AGSP}}^{(n)}] - \mathbf{E}[v_{\mathrm{GSP}}^{(n)}] \leq \delta GL$, for all $n \in \mathbb{Z}$, $n \geq 1$. Hence, the proof of theorem 4 is complete.

We next present an alternative approach for the characterization of the aggregation error. Define the *percentage of relative infeasibility* as $\%\mathrm{RI} := \frac{100}{\zeta_{\mathrm{sum}}} \sum_g [TH_g - H]^+$. Note that if $TH_g \leq H$, for all g, $\%\mathrm{RI} = 0$. Moreover, note that although we do not use cost in the definition of $\%\mathrm{RI}$, the costs are considered because the HGSP reduces the duration of calls that are associated with the least amount of saving.

We present a mathematical model to determine the worse-case $\%\mathrm{RI}$ for the HGSP. The parameters of this model are G, K, H, $L \in \mathbb{Z}$, and we assume they satisfy $G \geq 1$, $K \geq 1$, and $H \geq L \geq 2$. Note that the HGSP always finds an optimal solution for the GDP if $L = 1$. Recall that after sorting calls, a call is assigned to each group from each class; hence, after the classes are formed, one can change the order of the calls inside a class. Therefore, the DAP is equivalent to the following problem: *change the order of the calls inside the classes such that the following assignment is feasible: assign the first call from each class to group 1, the second call from each class to group 2, and so on.* Let $\zeta_{i\varsigma} \in \mathbb{Z}_+$, for all i, ς, denote the duration of the ith call in class ς. Thus, group g

is assigned $\zeta_{g1}, \zeta_{g2}, \cdots, \zeta_{gK}$ and hence $TH_g = \sum_{\varsigma=1}^{K} \zeta_{g\varsigma}$, for all g. Let $\zeta_{\min,\varsigma}, \zeta_{\max,\varsigma}$, for all ς, denote the minimum and maximum $\zeta_{i\varsigma}$ in class ς. The worst-case %RI problem (WRIP) is as follows:

$$\%\text{WRI} = \max \quad \frac{100}{\zeta_{\text{sum}}} \sum_{g=1}^{G} \left[\left(\sum_{\varsigma=1}^{K} \zeta_{g\varsigma} \right) - H \right]^{+} \tag{33}$$

$$\text{s.t.} \quad \zeta_{\min,\varsigma} \leq \zeta_{i\varsigma} \leq \zeta_{\max,\varsigma}, \ \forall i, \varsigma, \tag{34}$$

$$\zeta_{\max,1} \leq L, \tag{35}$$

$$\zeta_{\min,\varsigma} \geq \zeta_{\max,\varsigma+1}, \ \forall \varsigma \leq K-1, \tag{36}$$

$$\zeta_{\min,K} \geq 0, \tag{37}$$

$$\sum_{i=1}^{G} \sum_{\varsigma=1}^{K} \zeta_{i\varsigma} = \zeta_{\text{sum}}, \tag{38}$$

$$\zeta_{i\varsigma}, \zeta_{\min,\varsigma}, \zeta_{\max,\varsigma}, \zeta_{\text{sum}} \in \mathbb{R}, \ \forall i, \varsigma. \tag{39}$$

Eq. (34) ensures that the $\zeta_{i\varsigma}$ is lower bounded by $\zeta_{\min,\varsigma}$ and upper bounded by $\zeta_{\max,\varsigma}$. The first class includes the largest $\zeta_{i\varsigma}$, which must be less than or equal to L. This is guaranteed by Eq. (35). Eq. (36) assures that every call in class ς is at least as big as the maximum duration of the calls in class $\varsigma + 1$. Eq. (37) guarantees that the calls in class K have non-negative durations. Eq. (38) relates the values of ζ_{sum} and $\zeta_{i\varsigma}$.

In the WRIP, we assume $\zeta_{i\varsigma}$ are continuous variables while they must be integers. Hence, one could assume integral $\zeta_{i\varsigma}$ that may improve the %WRI; however, the WRIP may become more difficult to solve. A major drawback of the WRIP that makes it very difficult to solve is the nonlinearity of the objective function. The numerator of the objective function is a piecewise linear and convex function, and the denominator ζ_{sum} is a variable. For a fixed ζ_{sum}, one could use binary variables to convert the WRIP to a MILP problem; however, according to our experiment, the obtained MILP problem is significantly more difficult. The following proposition presents and alternative formulation and an effective approach for solving the WRIP in a reasonable time.

PROPOSITION 6. *The WRIP is equivalent to the following problem.*

$$\%WRI = \max_{\substack{g^* = 1, \cdots, G, \\ G(H-L+1)+1 \leq \zeta_{sum} \leq GH.}} \left\{ \begin{array}{l} \max \frac{100}{\zeta_{sum}} \sum_{g=1}^{g^*} \left(\left(\sum_{c=1}^{K} \zeta_{g\varsigma} \right) - H \right) \\ \text{s.t. Eqs. (34)–(39).} \end{array} \right\} \tag{40}$$

The proof is given in Appendix E.2. We applied Proposition 6 to a real-size problem with $G = 20$, $K = 100$, $H = 180$, and $L = 4$, and obtained that the %WRI is 0.56%. Hence, the aggregation error is significantly small for real-size problems.

6. Computational Experiment

First, we show the quality of our heuristic approach on a set of mid-size problems. Next, we present the results obtained from applying our heuristic to a real problem which is provided to us by a large utility company in the Southern California region.

6.1. Error analysis on mid-sized problems

We use the ECPs for a large region of the Southern California from 2012 to 2014. The number of day-types is set to 5, 10, and 15, and $G = 20$ and $L = 4$. We set $K = 6n$, $H = 10n$, and $D = 20n$, where $1 \leq n \leq 5$; hence, K varies between 6 and 30, H varies between 10 and 50, and D varies between 20 and 100. We use a strictly increasing cost function similar to Fig. 1.

Fig. 7 presents the relative and absolute *information errors* obtained for different problem sizes, and for different number of day-types. Each data point corresponds to the average information error for the years 2012–2014. As expected, the absolute and relative errors decrease as the number of day-types decreases, with the lower graph corresponding to the errors obtained for the smallest number of day-types. As n increases, the absolute error modestly increases, which is bounded by $\mathcal{O}(\sqrt{n})$, as indicated by theorem 3. The relative error rapidly decreases as n increases, which is expected as proven in theorem 3. Overall, Fig. 7 shows that the relative information error decreases very fast and reaches 0.06%, 0.07%, and 0.23%, for 5, 10, and 15 day-types, respectively. As n increases, the relative information error is projected to be significantly small.

Fig. 8 depicts the results obtained for the absolute and relative *aggregation error* for different problem sizes. In all of our experiments, we use the *infeasibility* (defined in Section 5.3) as a proxy for the aggregation error. Our computational experiments suggest that the effect of the number of day-types, unlike in the case of information error, is not significant. The absolute error is almost
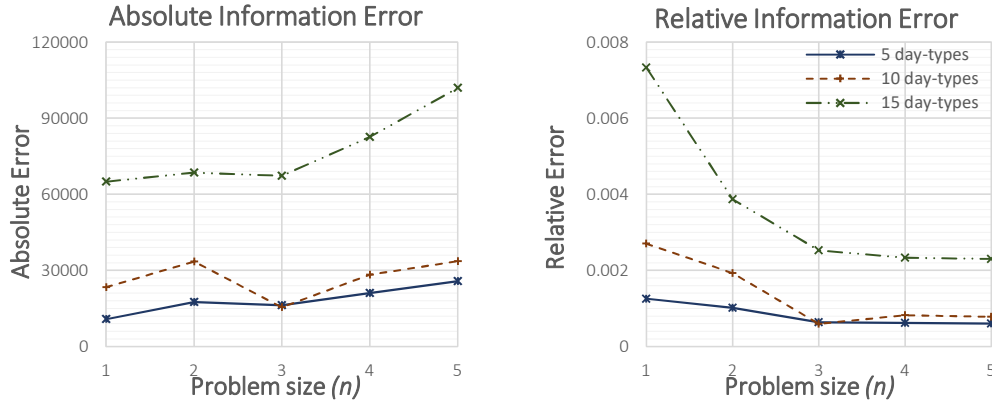
Figure 7   Information error and its sensitivity to the problem size and the number of day-types.
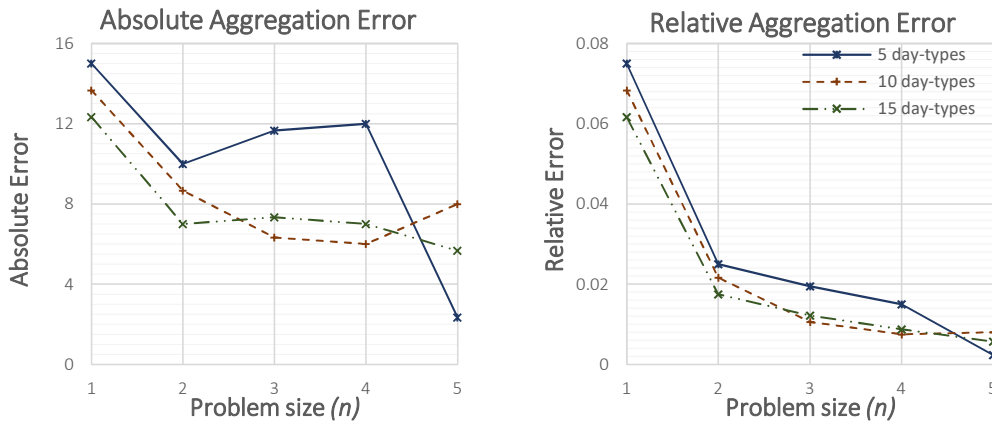

Figure 8   Aggregation error and its sensitivity on problem size and the number of day-types.

constant and the relative error reduces at a very fast rate and reaches 1% for $n = 5$. Finally, we note that the total error of our heuristic approach is the sum of the information and aggregation error. For 5, 10, and 15 day-types, the average (over three years) total relative error is 0.29%, 0.87%, and 0.79%, respectively.

6.2. An industrial instance

We present the results of applying our proposed methodology on a real problem with ECPs from 2012 to 2014. The instance is provided to us by a large utility company in the Southern California region. In these instances, $K = 100$, $H = 180$, $D = 365$, $|\Omega| = 10$, and $L = 4$. For managerial considerations, we looked at 5, 10, 15, and 20 customer groups. Fig. 9 exhibits the percentage of saving
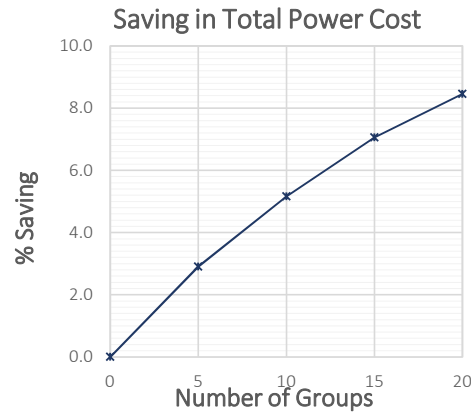
Figure 9   Percentage of saving as the number of groups increases.
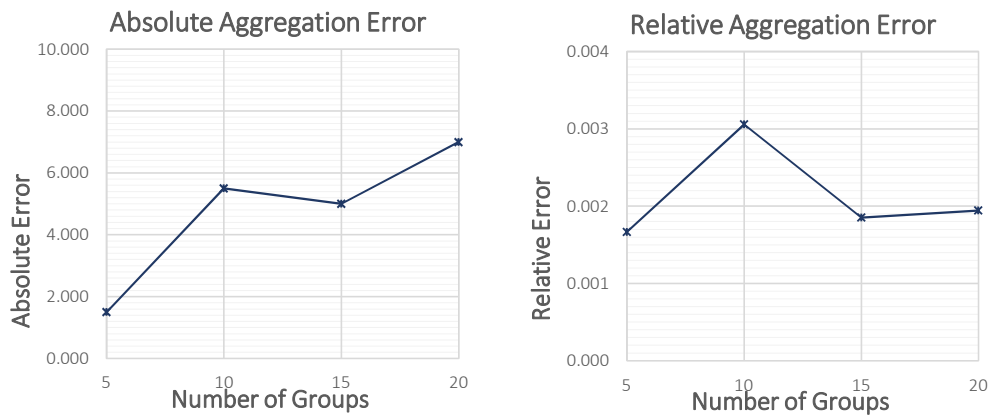

Figure 10   The effects of the number of groups on the aggregation error.

in total power cost as the number of groups increases. The total saving is a function of the number

of customer groups and increases from 2.9% to 8.45% for 5 to 20 groups, respectively.

Fig. 10 shows the absolute and relative aggregation errors for different numbers of groups. The

absolute error increases almost linearly in G, and the relative error is almost constant in G, with

an average of 0.21%. Due to the complexity and size of the industrial instances—i.e., the inability

to optimally solve the AGSP—we can only report the aggregation error. However, we note that

based on our computational experiment for mid-size problems (Section 6.1), the information error
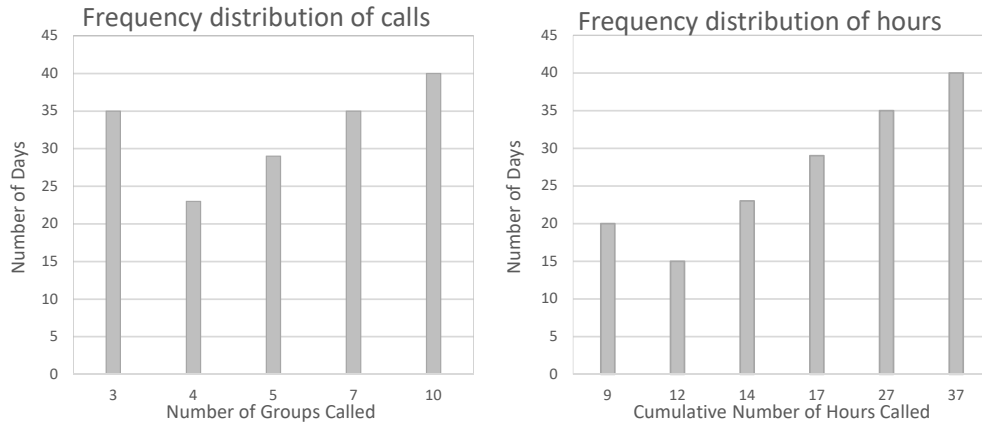
is expected to be less than 0.07%.

Figure 11   The frequency distributions of calls and hours.

Our solution scheduled calls for 162 days throughout the year and the frequency distribution of the daily cumulative calls and hours are shown in Fig. 11. For example, in 35 days, 3 calls per day are scheduled, and in 20 days, a total of 9 hours per day is assigned to groups.

Finally, we note that the impetus behind DLCCs is to reduce peak load consumption and influence customers' energy consumption patterns. Fig. 12 shows the before and after ECP patterns for the first weeks of winter and summer 2014. In winter, the energy consumption fluctuates between 19 and 29 GW/h, while in summer, it varies between 22 and 36 GW/h. According to our solution, the summer peak load reduces from 36 to 32 GW/h, which may correspond to a reduction in the energy production rate from 160 to 80 \$/MWh (see Fig. 1).

7. Concluding Remarks

We analyze the problem of efficiently implementing DLCCs used by many utility companies in developed countries. We optimally solve an special case of the problem in which the lengths of interventions are equal and fixed. A heuristic approach is proposed based on approximating the stochasticity, and aggregating the customer groups and forming a representative group. Our heuristic approach is asymptotically optimal and its worst-case relative error is $\mathcal{O}(\frac{1}{\sqrt{n}})$. The quality of our heuristic approach under different settings were experimentally verified. An industrial instance of this problem is thoroughly which results in approximately 8% in the energy generation cost.
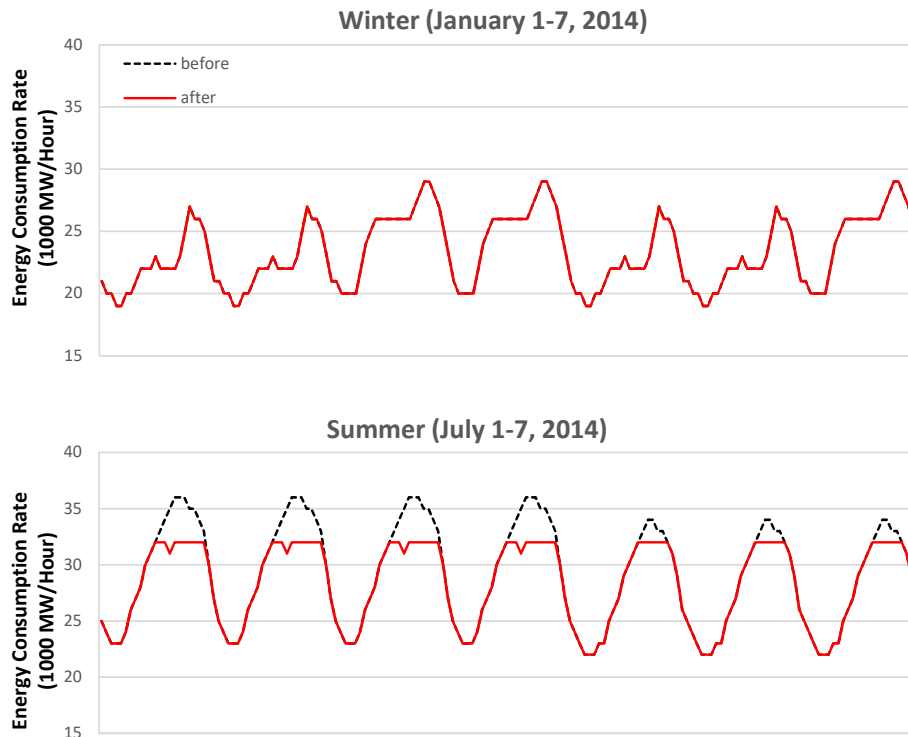
Figure 12 The peak load reduction during the first weeks of winter and summer 2014.

Finally, we emphasize that our approach can also be used to efficiently design DLCCs and set essential contract parameters.

We offer the following directions for future research. First, although we assumed the customer groups are identical in terms of their demands and contracts, this may not be true in some demand-response programs. Second, in the DLCCs that were considered in this paper, the peak demand is eliminated—not shifted—through the scheduled interventions. However, in some programs, in anticipation of being called for a certain period, a customer group may shift their energy consumption to before and after that period. This phenomenon is referred to as "the shoulder effect". Finally, this paper assumes that customers can subscribe to DLCCs only at the beginning of the planning horizon. However, we note that there are demand-response programs that allow customers subscribe any time throughout the year. Future research could look at demand-response programs with non-identical groups, shoulder effects, and/or rolling contracts.

Appendix A: Complexity Proofs

A.1. Proof of Proposition 1(a)

We show that the GDP with $G = 2$ is as hard as the 2-partition problem, which is known to be weakly NP-complete (Garey and Johnson 1979). Assume that we are given a general instance of the 2-partition problem consisting of the positive integers $\alpha_1, \alpha_2, \ldots, \alpha_{2D}$, and A such that $\sum_{d=1}^{2D} \alpha_d = 2A$. The question is: Can $\alpha_1, \alpha_2, \ldots, \alpha_{2D}$ be partitions into two subsets with equal size and equal number of elements?

Consider the following specific instance of the GDP. Let $G = 2$ and consider a planning horizon of $2D$ days where in day d the height of the ECP is equal to 1 at periods $1, \ldots, \alpha_d$ and 0 at periods $\alpha_d + 1, \ldots, T$. Thus, there are $2D$ ECPs with durations $\alpha_1, \alpha_2, \ldots, \alpha_{2D}$ (see Fig. 13). We assume $T \geq \max_{d=1,\ldots,2D}\{\alpha_d\}$ and $L \geq \max_{d=1,\ldots,2D}\{\alpha_d\}$. We assume $f(1) = 1$ meaning that if we assign α_d to a group, we achieve a saving of α_d. Let $K = D$ and $H = A$.
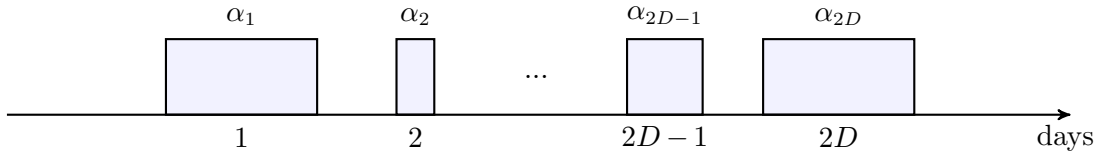

Figure 13 The ECPs for the planning horizon of $2D$ days.

Case 1: If the 2-partition has a solution, then the optimal solution of the GDP is obtained by assigning partitions to groups. The saving that is achieved by this assignment is $2A$.

Case 2: If the 2-partition does not have solution, we show that the GDP cannot have a solution with a total saving of $2A$. If the 2-partition does not have solution, then for all partitions with equal size, there exists $\Lambda > 0$ such that the size of the partitions are $A - \Lambda$ and $A + \Lambda$. Assigning these partitions to the groups, one could achieve savings of $A - \Lambda$ and A, respectively. Therefore, the total saving is $2A - \Lambda$ which is strictly less than $2A$.

To complete the proof, we note that the GDP with two groups is solved in a pseudo polynomial time, which implies that the problem is weakly NP-complete.

A.2. Proof of Proposition 1(b)

We show that the GDP is as hard as the 3-partition problem, which is known to be strongly NP-complete (Garey and Johnson 1979). Assume that we are given a general instance of the 3-partition problem consisting of positive integers $\alpha_1, \alpha_2, \ldots, \alpha_{3G}$, and A such that $\sum_{d=1}^{3G} \alpha_d$ is divisible by G and $A = \frac{1}{G}\sum_{d=1}^{3G} \alpha_d$, and $\frac{A}{4} \leq \alpha_d \leq \frac{A}{2}$, for all d. The question is: Can $\alpha_1, \alpha_2, \ldots, \alpha_{3G}$ be partitioned into G subsets with equal size and equal number of elements?

Consider the following specific instance of the GDP, with G groups. Let $D = 3G$, and assume that the ECP height is 1 in periods $1, \ldots, \alpha_d$ and 0 in periods $\alpha_d + 1, \ldots, T$, on day d, for all d. Thus, there are $3G$ energy consumption periods with durations $\alpha_1, \alpha_2, \ldots, \alpha_{3G}$. We assume $T \geq \max_{d=1,\ldots,3G}\{\alpha_d\}$ and $L \geq \max_{d=1,\ldots,3G}\{\alpha_d\}$. We assume $f(1) = 1$ meaning that if we assign α_d to a group, we achieve a saving of α_d. Let $K = 3$ and $H = A$.

Case 1: If the 3-partition has a solution, then the optimal solution of the GDP is obtained by assigning the partitions to groups. The saving that is achieved by this assignment is GA.

Case 2: The 3-partition does not have a solution. Let us partition $\alpha_1, \alpha_2, \ldots, \alpha_{3G}$ into G subsets with equal number of elements and assign each set to a group. Hence, each group receives exactly 3 elements. There exists a positive integer $\Lambda > 0$ and groups g' and g'' such that $\sum_{d:\alpha_d \text{ is assigned to } g'} \alpha_d \geq A + \Lambda$ and $\sum_{d:\alpha_d \text{ is assigned to } g''} \alpha_d \leq A - \Lambda$. Therefore, the total saving of this assignment is less than or equal to $GA - \Lambda$ and hence is strictly less than GA. Note that this holds for all such partitions with equal number of elements. Therefore, the optimal solution of the GDP is strictly less than GA.

Appendix B: FSP Proofs

B.1. Proof of Proposition 2

Let s_t^* denote the optimal solution to AFSDP^r. If AFSDP^r results in a solution in which all s_t^*'s are integer, then the obtained solution is an optimal solution to AFSDP. Consider the case where some of the s_t^*'s are fractional. We define, for each t, $S_t^* := \sum_{t'=\max\{1,t-L+1\}}^{\min\{T-L+1,t\}} s_{t'}^*$ which is the number of groups who are on call during time period t in the optimal solution. Two or more S_t^*'s must be fractional. Consider the following linear program (P1).

$$\max \ \sum_{t=1}^{T} \partial \hat{f}_{d,t}(\lfloor S_t^* \rfloor) \sum_{t'=\max\{1,t-L+1\}}^{\min\{T-L+1,t\}} s_{t'}, \tag{41}$$

$$\text{s.t. } R_t \geq r_t(\omega_d) - \sum_{t'=\max\{1,t-L+1\}}^{\min\{T-L+1,t\}} s_t, \ \forall t \tag{42}$$

$$\sum_{t=1}^{T-L+1} s_t = \mathcal{K}_d \tag{43}$$

$$\lfloor S_t^* \rfloor \leq \sum_{t'=\max\{1,t-L+1\}}^{\min\{T-L+1,t\}} s_{t'} \leq \lceil S_t^* \rceil, \ \forall t, \tag{44}$$

$$s_t \in \mathbb{R}_+, \ \forall t \leq T - L + 1. \tag{45}$$

We note that s_t^* are feasible for P1. A feasible solution to P1 is also feasible for AFSDP^r and the objective values of AFSDP^r and P1 differ by a constant that is independent of s_t's. We show next that all the extreme points of P1 are integer. We note that P1 contains $T + 1$ constraints

and $T - L + 1$ variables. We define $\aleph_t := \sum_{t'=\max\{1,t-L+1\}}^{\min\{T-L+1,t\}} s_{t'}$, for all $1 \le t \le T$. Moreover, we define $\aleph_0 := \sum_{t'=1}^{T-L+1} s_{t'}$, and $\aleph_t := 0$, for all $t > T$.

Assume that $T = \alpha L + \beta$, where α is an integer and $0 \le \beta < L$. Then, $\aleph_0 = \sum_{i=0}^{k+1} \aleph_{j+Li}$ for $j = 1, \ldots, L$. Thus constraints corresponding to $\aleph_{T-L+1}, \ldots, \aleph_T$ are redundant constraints. They are obtained through elementary row operations on the remaining $T - L + 1$ constraints. We note that $\aleph_0$ and $\aleph_1, \ldots, \aleph_T$ are independent of each other as each of them introduces one new variable. Consider the solution to the following set of equations:

$$\aleph_0 = \mathrm{RHS}_0 \tag{46}$$

$$\aleph_t = \mathrm{RHS}_t, \ \forall t : 1 \le t \le T - L, \tag{47}$$

where all the right hand side variables are integers. The solution to this set is easily found, because $\aleph_1 = s_1$, $\aleph_2 = s_1 + s_2$, and so on. Each constraint introduces exactly one new variable. All the s_t's will be integer.

Any other independent set of $T - L + 1$ constraints from the set $\aleph_0, \ldots, \aleph_T$ can be transformed to the set $\aleph_0, \ldots, \aleph_{T-L+1}$ through elementary row operations that does not involve any division, hence maintaining the integrality of the right hand side. Thus, every extreme point of the polytope corresponding to P1 is an integer. This implies that there exists a set of integer extreme points such that their convex combination is s^*, the optimal solution to AFSDP^r.

Using property (iii) of the function f, we have, for $z' \ne z''$, $\partial \hat{f}_{d,t}(z') \ne \partial \hat{f}_{d,t}(z'')$. Then, if s^* is not an extreme point, then there is a strict improvement direction. Hence P1 contains an integer solution that has a higher objective than that given by s^*. This in turn implies that s^* is not optimal to AFSDP^r. Thus we proved that AFSDP^r yields integer solutions. $\square$

B.2. Proof of Lemma 1

We prove this claim using an induction. Before starting the *for*-loop, we have $TC_g = 0$, for all g. Assume that in iteration $\hat{d}$, we have $|TC_{g'} - TC_{g''}| \le 1$, for all g', g''. Sort TC_g's in a non-decreasing order $TC_{[1]} \le TC_{[2]} \le \cdots \le TC_{[G]}$; hence, $TC_{[1]} + 1 \ge TC_{[G]}$. If $TC_{[1]} + 1 = TC_{[G]}$, then there exists $[\dot{g}]$ such that $TC_{[1]} = \cdots = TC_{[\dot{g}]-1} = TC_{[\dot{g}]} - 1 = \cdots = TC_{[G]} - 1$. If $TC_{[1]} = TC_{[G]}$, let $[\dot{g}] = [G] + 1$. Let $TC := TC_{[1]}$. Then, $TC_{[g]} = TC$, for all $[g]$ such that $[g] < [\dot{g}]$, and $TC_{[g]} = TC + 1$, for all $[g]$ such that $[g] \ge [\dot{g}]$.

There exists $[\ddot{g}]$ such that the *for*-loop assigns a call to each group $[g]$ such that $[g] < [\ddot{g}]$. Let $[g]_{\min} := \min\{[\dot{g}], [\ddot{g}]\}$ and $[g]_{\max} := \max\{[\dot{g}], [\ddot{g}]\}$. In iteration $\hat{d} + 1$, we have:

$$TC_{[g]} = \begin{cases} TC + 1 & , \ \forall [g] : [g] < [g]_{\min} \\ TC & , \ \forall [g] : [g]_{\min} \le [g] < [g]_{\max}, \text{ if } [\dot{g}] > [\ddot{g}] \\ TC + 2 & , \ \forall [g] : [g]_{\min} \le [g] < [g]_{\max}, \text{ if } [\dot{g}] < [\ddot{g}] \\ TC + 1 & , \ \forall [g] : [g] \ge [g]_{\max} \end{cases}$$

Therefore, in iteration $\hat{d} + 1$, we have $|TC_{g'} - TC_{g''}| \le 1$, for all g', g''. So, the proof is complete.

B.3. Proof of Lemma 2

According to lemma 1, for $d = D - 1$, we have $|TC_{g'} - TC_{g''}| \leq 1$, for all g', g''. Now consider day D. We assign $\min\{\mathscr{K}_D, G\}$ or less but we consider assigning $\min\{\mathscr{K}_D, G\}$ because if we can assign $\min\{\mathscr{K}_D, G\}$ to groups, then we can assign smaller values as well. One of the following cases can happen:

<u>CASE 1</u>: $\mathscr{K}_D < G$; hence, we assign $\mathscr{K}_D$. We claim that in this case each group has at most 1 call available—i.e., $K - TC_g \leq 1$, for all g. This claim is proven as follows. Note that at the beginning of iteration D we have assigned TC_g calls to group g, and group g has $K - TC_g$ unused calls. The total unused calls is $\sum_g (K - TC_g) = GK - \sum_g TC_g = \mathscr{K}_D < G$. Assume there exists $\hat{g}$ such that $K - TC_{\hat{g}} > 1$, then $TC_{\hat{g}} < K - 1$. Using Lemma 1, we must have $TC_g \leq K - 1$, for all g, with strict inequality for $\hat{g}$. Therefore, $\sum_g TC_g < GK - G$ and hence $GK - \sum_g TC_g > G$. This is a contradiction. Thus, each group has at most 1 call available. Therefore, to assign $\mathscr{K}_D$, we call each group g with an available call.

<u>CASE 2</u>: $\mathscr{K}_D \geq G$; hence, we assign G. We claim that in this case each group has at least 1 call available—i.e., $K - TC_g \geq 1$, for all g. We prove this claim by contradiction. Assume there exists $\hat{g}$ such that $K - TC_{\hat{g}} < 1$. Then, $TC_{\hat{g}} > K - 1$. Using Lemma 1, we must have $TC_g \geq K - 1$, for all g, with strict inequality for $\hat{g}$. Therefore, $\sum_g TC_g > GK - G$ and hence $\mathscr{K}_D = GK - \sum_g TC_g < G$. This is a contradiction. Thus, each group has at least 1 call available. Therefore, to assign G calls, we call all groups.

Appendix C: Information Error Proofs

C.1. Proof of Lemma 3

We prove the following relationship, for all $n \in \mathbb{Z}$, $n \geq 1$, which proves the lemma.

$$\mathbf{E}[v_{\text{UBH}}^{(n)}] = v_{\text{RAGDP}}^{(n)} \geq v_{\text{AGDP}}^{(n)} \geq \mathbf{E}[v_{\text{AGSP}}^{(n)}] \geq \mathbf{E}[v_{\text{LBH}}^{(n)}].$$

Note that $\mathbf{E}[v_{\text{UBH}}^{(n)}] = \sum_{d=1}^{nD} \sum_j q_j p_{\omega(j)} \theta_j = \sum_j \sum_{d=1}^{nD} \frac{x_j^*}{D} \theta_j = n \sum_j \theta_j x_j^* = v_{\text{RAGDP}}^{(n)}$, for all $n \in \mathbb{Z}$, $n \geq 1$. Because RAGDP$^{(n)}$ is a relaxation of AGDP$^{(n)}$, then $v_{\text{RAGDP}}^{(n)} \geq v_{\text{AGDP}}^{(n)}$, for all $n \in \mathbb{Z}$, $n \geq 1$. Recall that in AGDP$^{(n)}$ we assume that $n p_\omega D$ days are from type ω, for all ω. Also, recall that we assume $p_\omega D$ is integer for all ω. Note that the optimal value of AGDP$^{(n)}$ is concave in the number of days of type ω, for all ω. Also, using Jensen's inequality, we have $v_{\text{AGDP}}^{(n)} \geq \mathbf{E}[v_{\text{AGSP}}^{(n)}]$, for all $n \in \mathbb{Z}$, $n \geq 1$. Because LBH$^{(n)}$ provides a feasible solution to AGSP$^{(n)}$, then $\mathbf{E}[v_{\text{AGSP}}^{(n)}] \geq \mathbf{E}[v_{\text{LBH}}^{(n)}]$, for all $n \in \mathbb{Z}$, $n \geq 1$.

36

Fattahi, Dasu and Ahmadi: *Peak Load Energy Management*

Article submitted to *Operations Research*; manuscript no. (Please, provide the manuscript number!)

C.2. Proof of Lemma 4

For all γ and d, we have:

$$\mathbf{E}[(\Delta_\gamma^{(n)}(d))^2|\Im_d] = \sum_j p_{\omega(j)}q_j\left(a_{\gamma j} - \sum_{j'} p_{\omega(j')}q_{j'}a_{\gamma j'}\right)^2 \le \frac{1}{4}a_{\gamma,\max}^2.$$

This inequality follows from Popoviciu's inequality according to which $Var(x) \le \frac{1}{4}(\overline{x}-\underline{x})^2$, where $\overline{x} \le x \le \underline{x}$ with probability 1. We note that $\sum_j p_{\omega(j)}q_j(a_{\gamma j} - \sum_{j'} p_{\omega(j')}q_{j'}a_{\gamma j'})^2$ is the variance of $a_{\gamma 1},\ldots,a_{\gamma J}$. Also note that for all j, $0 \le a_{\gamma j} \le a_{\gamma,\max}$. Hence, the variance does not exceed $\frac{1}{4}a_{\gamma,\max}^2$. To complete the proof, we note that $\mathbf{E}[(\sum_{d'=1}^d \Delta_\gamma^{(n)}(d'))^2|\{\Im_d\}] = \sum_{d'=1}^d \mathbf{E}[(\Delta_\gamma^{(n)}(d'))^2|\Im_{d'}]$ because considering the filtration $\{\Im_d\}$, the error terms, $\Delta_\gamma^{(n)}(d')$, are independent.

C.3. Proof of Lemma 5

The probability $P(d^{*(n)} < d)$ is found by calculating the probability of the violation of condition (†) for some resource γ and for some days $\hat{d} < d$. Therefore, we have:

$$P(d^{*(n)} < d) = P\left(\sum_{d'=1}^{\hat{d}} \Delta_\gamma^{(n)}(d') > nc_\gamma - \frac{\hat{d}}{D}\mathbf{a}_\gamma\mathbf{x}^* - a_{\gamma,\max}, \text{ for some } \gamma, \hat{d} < d\right)$$

$$\le \sum_\gamma P\left(\sum_{d'=1}^{\hat{d}} \Delta_\gamma^{(n)}(d') > nc_\gamma - \frac{\hat{d}}{D}\mathbf{a}_\gamma\mathbf{x}^* - a_{\gamma,\max}, \text{ for some } \hat{d} \le d\right)$$

Since, $\hat{d} \le d$, we replace $\hat{d}$ with d in the right-hand-side of the condition; hence, the right-hand-side becomes a constant (independent of $\hat{d}$). Therefore, we can take the maximum of $\sum_{d'=1}^{\hat{d}} \Delta_\gamma^{(n)}(d')$ over $\hat{d} = 1,\ldots,d$, and hence we have:

$$P(d^{*(n)} < d) \le \sum_\gamma P\left(\max_{\hat{d}=1,\ldots,d} \sum_{d'=1}^{\hat{d}} \Delta_\gamma^{(n)}(d') > nc_\gamma - \frac{d}{D}\mathbf{a}_\gamma\mathbf{x}^* - a_{\gamma,\max}\right)$$

$$\le \sum_\gamma P\left(\max_{\hat{d}=1,\ldots,d} \left|\sum_{d'=1}^{\hat{d}} \Delta_\gamma^{(n)}(d')\right| \ge n\left(c_\gamma - \frac{d}{nD}\mathbf{a}_\gamma\mathbf{x}^*\right) - a_{\gamma,\max}\right)$$

$$\le \sum_\gamma \frac{\mathbf{E}\left[\left(\sum_{d'=1}^d \Delta_\gamma^{(n)}(d')\right)^2\right]}{\left(n(c_\gamma - \frac{d}{nD}\mathbf{a}_\gamma\mathbf{x}^*) - a_{\gamma,\max}\right)^2} \tag{48}$$

$$\le \sum_\gamma d\left(\frac{2n}{a_{\gamma,\max}}\left(c_\gamma - \frac{d}{nD}\mathbf{a}_\gamma\mathbf{x}^*\right) - 2\right)^{-2} \tag{49}$$

Inequality (48) follows from Doob's submartingale inequality, and inequality (49) follows from applying Lemma 4. To complete the proof, we note that $P(d^{*(n)} < d) \le 1$, for all n and d.

Appendix D: HGSP Proofs

D.1. Proof of Remak 2

We first note that the objective function can be alternatively formulated as the following piecewise linear and convex function.

$$\sum_{z=1}^{Z}\left\{\partial^2 f(z)\left(\sum_{t=1}^{T}\left[r_t(\omega_d)-z+1-\left(\sum_{i=1}^{\min\{L,t\}}\sum_{\ell=i}^{\mathcal{L}(t)}s_{(t-i+1)\ell}\right)\right]^{+}\right)\right\},\tag{50}$$

where $Z \in \mathbb{Z}$ is an upper bound on the ECP height. We linearize the objective function by defining the non-negative variables ε_{zt}, for all z, t, replacing $\left[r_t(\omega_d)-z+1-\left(\sum_{i=1}^{\min\{L,t\}}\sum_{\ell=i}^{\mathcal{L}(t)}s_{(t-i+1)\ell}\right)\right]^{+}$ with ε_{zt}, for all z, t, and adding constraints to ensure $\varepsilon_{zt} \geq r_t(\omega_d) - z + 1 - \left(\sum_{i=1}^{\min\{L,t\}}\sum_{\ell=i}^{\mathcal{L}(t)}s_{(t-i+1)\ell}\right)$, for all z, t. Since we solve a minimization problem, this constraint is always satisfied as equality unless the right-hand-side is negative in which case $\varepsilon_{zt}=0$.

D.2. Proof of Proposition 3

(a) Note that we form classes based on a sorted list of calls; hence, there exists $\hat{\varsigma}$ such that every class $1,\cdots,\hat{\varsigma}-1$ has exactly G calls with nonzero durations, and every class $\hat{\varsigma}+1,\cdots,K$ has no call with nonzero duration. Moreover, we assign exactly one call from each class to each group. Thus, each group receives exactly one call from each class $1,\cdots,\hat{\varsigma}-1$ and at most one call from class $\hat{\varsigma}$. Therefore, a group receives either $\hat{\varsigma}-1$ or $\hat{\varsigma}$ calls; hence, $|TC_{g'}-TC_{g''}|\leq 1$, for all g', g''.

(b) Let $\zeta_{i\varsigma}\in\mathbb{Z}_{+}$, for all i, ς, denote the duration of the ith call in class ς. Since we sort the calls in a non-increasing length order, we have $\zeta_{11}\leq L$ and $\zeta_{GK}\geq 0$; hence, $\zeta_{11}-\zeta_{GK}\leq L$. Moreover, we have $\zeta_{G\varsigma}\geq\zeta_{1(\varsigma+1)}$, for all $\varsigma\leq K-1$ because of the sorted list. Finally, the maximum difference between the durations of the calls in a class is $(\zeta_{1\varsigma}-\zeta_{G\varsigma})$, for all ς. Thus, we have:

$$\sum_{\varsigma=1}^{K}(\zeta_{1\varsigma}-\zeta_{G\varsigma})=\zeta_{11}-(\zeta_{G1}-\zeta_{12})-\cdots-(\zeta_{G(K-1)}-\zeta_{1K})-\zeta_{GK}\leq\zeta_{11}-\zeta_{GK}\leq L.$$

Therefore, $\sum_{\varsigma=1}^{K}(\zeta_{1\varsigma}-\zeta_{G\varsigma})\leq L$. This means that the sum of the class differences is upper bounded by the total difference which itself is upper bounded by L. The DAP assigns exactly one call from each class to each group. Consider the groups g' and g''. The difference $(TH_{g'}-TH_{g''})$ is maximized if group g' is assigned the maximum duration from each class and group g'' is assigned the minimum duration from each class.

$$\begin{aligned}
TH_{g'}-TH_{g''}&=\sum_{i\varsigma}\lambda_{i\varsigma g'}\zeta_{i\varsigma}-\sum_{i\varsigma}\lambda_{i\varsigma g''}\zeta_{i\varsigma}\\
&\leq\sum_{\varsigma}\left(\max_{i}\{\zeta_{i\varsigma}\}\right)-\sum_{\varsigma}\left(\min_{i}\{\zeta_{i\varsigma}\}\right)\\
&=\sum_{\varsigma}\zeta_{1\varsigma}-\sum_{\varsigma}\zeta_{G\varsigma}=\sum_{\varsigma}(\zeta_{1\varsigma}-\zeta_{G\varsigma})\leq L.
\end{aligned}$$

Hence, the proof is complete and $|TH_{g'}-TH_{g''}|\leq L$, for all g', g''.

D.3. Proof of Proposition 4

Adding slack variables to Eq. (31), the DAP^r is equivalent to the following system of equations.

$$\sum_{g=1}^{G} \lambda_{i\varsigma g} = 1, \ \forall i, \varsigma, \tag{51}$$

$$\sum_{i=1}^{G} \lambda_{i\varsigma g} = 1, \ \forall \varsigma, g, \tag{52}$$

$$\sum_{i\varsigma \in \mathcal{D}_d} \lambda_{i\varsigma g} + \sigma_{gd} = 1, \ \forall g, \mathcal{D}_d, \tag{53}$$

where, $\lambda_{i\varsigma g}$, for all i, ς, g, and σ_{gd}, for all g, d, are non-negative continuous variables. We note that $\lambda_{i\varsigma g}$ cannot take a value greater than 1 because the right-hand-side values are 1, all of the coefficients in the left-hand-sides are 1, and all variables are non-negative and continuous. Thus, constraint $\lambda_{i\varsigma g} \leq 1$, for all i, ς, g, is automatically ensured. Thus, the coefficient matrix of DAP^r is a $(2KG + DG) \times (KG^2 + DG)$-matrix of zeros and ones as follows.

$$\mathcal{A} = \begin{bmatrix} \mathbf{I}_{KG} & \mathbf{I}_{KG} & \cdots & \mathbf{I}_{KG} & 0 \\ \hline \mathcal{B} & 0 & \cdots & 0 & 0 \\ 0 & \mathcal{B} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & \cdots & \mathcal{B} & 0 \\ \hline \mathcal{E} & 0 & \cdots & 0 & \\ 0 & \mathcal{E} & \cdots & 0 & \mathbf{I}_{DG} \\ \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & \mathcal{E} & \end{bmatrix}, \qquad \mathcal{B}_{K \times KG} = \begin{bmatrix} \mathbf{1}'_G & 0 & \cdots & 0 \\ 0 & \mathbf{1}'_G & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{1}'_G \end{bmatrix},$$

and $\mathcal{E}$ is a $(D \times KG)$-matrix of zeros and ones. The following remark states that the matrix $\mathcal{A}$ has $(DG + 2KG - K)$ linearly independent rows and columns. The proof is standard and follows from performing some elementary operations on $\mathcal{A}$.

REMARK 4. The rank of matrix $\mathcal{A}$ is $(DG + 2KG - K)$.

It is seen that $\mathcal{B}$ is a matrix of zeros and ones. Each column of $\mathcal{B}$ has exactly one 1 and exactly $(K - 1)$ zeros. Moreover, each row of $\mathcal{B}$ has exactly G ones and exactly $(K - 1)G$ zeros. We similarly characterize the columns and rows of $\mathcal{E}$ in the following remark. The proof is standard and follows from analyzing Eq. (53).

REMARK 5. Matrix $\mathcal{E}$ has the following properties: (i) each entry is either 0 or 1, (ii) each column has exactly one 1, and (iii) each row has at most G ones. Moreover, if $D = K$, then each row of $\mathcal{E}$ has exactly G ones.

We partition $\mathcal{E}$ into two sub-matrices as follows: $\mathcal{E}_1$ consists of the rows of $\mathcal{E}$ that have exactly G ones, and $\mathcal{E}_2$ consists of the rows of $\mathcal{E}$ that have strictly less than G ones. Let $\Phi := \{\lambda \in \mathbb{R}^{KG} | \mathcal{B}\lambda =$

$\mathbf{1}_K, \mathcal{E}_1\lambda = \mathbf{1}_{D_1}, \mathcal{E}_2\lambda \leq \mathbf{1}_{D_2}, \lambda \geq \mathbf{0}\}$. In the following, we show that Φ is a nonempty and integral polyhedron.

Adding slack variables to $\mathcal{E}_2\lambda \leq \mathbf{1}_{D_2}$, the coefficient matrix becomes $\begin{bmatrix} \mathcal{B} & 0 \\ \mathcal{E}_1 & 0 \\ \mathcal{E}_2 & \mathbf{I}_{D_2} \end{bmatrix}$ and the right-hand-side vector is $\mathbf{1}_{(K+D)}$. Recall that each column of $\mathcal{B}$ and $\mathcal{E}$ has exactly one 1 (see Remark 5); hence, each column of the coefficient matrix has at most 2 ones. Additionally, if there are 2 ones in a column of the coefficient matrix, one of them is in the first K rows and the other one is in the last D rows. Therefore, the coefficient matrix is *totally unimodular* meaning that every extreme point of the feasible region is integral and thus the set Φ is an integral polyhedron.

We next prove that Φ is nonempty. Consider the following LP problem: $\max 0$ s.t. $\lambda \in \Phi$. The dual problem is as follows.

$$\min \quad \mathbf{1}_G'\chi\mathbf{1} + \mathbf{1}_{D_1}'\chi\mathbf{2} + \mathbf{1}_{D_2}'\chi\mathbf{3} \tag{54}$$

$$\text{s.t.} \quad \begin{bmatrix} \mathcal{B} \\ \mathcal{E}_1 \\ \mathcal{E}_2 \end{bmatrix}' \begin{bmatrix} \chi\mathbf{1} \\ \chi\mathbf{2} \\ \chi\mathbf{3} \end{bmatrix} \geq \mathbf{0} \tag{55}$$

$$\chi\mathbf{3} \geq \mathbf{0}, \tag{56}$$

where, $\chi\mathbf{1}$, $\chi\mathbf{2}$, $\chi\mathbf{3}$ are the dual variables. One could see that the dual problem is feasible because $\chi\mathbf{1} = \mathbf{0}$, $\chi\mathbf{2} = \mathbf{0}$, and $\chi\mathbf{3} = \mathbf{0}$ satisfy the constraints. We need to show that the optimal value for the dual problem is finite. Consider the relaxation of the dual problem that is obtained by multiplying Eq. (55) by $\mathbf{1}_{K+D}'$ from the left. This results in the following inequality.

$$\begin{bmatrix} G\mathbf{1}_K \\ G\mathbf{1}_{D_1} \\ \aleph \end{bmatrix}' \begin{bmatrix} \chi\mathbf{1} \\ \chi\mathbf{2} \\ \chi\mathbf{3} \end{bmatrix} \geq 0 \tag{57}$$

where, $\aleph \in \mathbb{R}^{D_2}$ and $\aleph \leq G\mathbf{1}_{D_2}$. Thus, the relaxation of the dual problem includes Eqs. (54), (56), and (57). Note that every feasible solution to the dual problem is also feasible for the relaxation problem but the converse may not be true. Eq. (57) can be written as follows.

$$G\mathbf{1}_K'\chi\mathbf{1} + G\mathbf{1}_{D_1}'\chi\mathbf{2} + \aleph'\chi\mathbf{3} \geq 0$$
$$\implies G\left(\mathbf{1}_K'\chi\mathbf{1} + \mathbf{1}_{D_1}'\chi\mathbf{2} + \mathbf{1}_{D_2}'\chi\mathbf{3}\right) - \left(G\mathbf{1}_{D_2}'\chi\mathbf{3} - \aleph'\chi\mathbf{3}\right) \geq 0$$
$$\implies \mathbf{1}_K'\chi\mathbf{1} + \mathbf{1}_{D_1}'\chi\mathbf{2} + \mathbf{1}_{D_2}'\chi\mathbf{3} \geq \frac{1}{G}\left(G\mathbf{1}_{D_2} - \aleph\right)'\chi\mathbf{3}$$
$$\implies \mathbf{1}_K'\chi\mathbf{1} + \mathbf{1}_{D_1}'\chi\mathbf{2} + \mathbf{1}_{D_2}'\chi\mathbf{3} \geq 0$$

Thus, the optimal value of the relaxation of the dual problem is bounded below; hence, the optimal value of the dual problem is bounded. Therefore, since the dual problem has a finite optimal solution, then the primal problem also has a solution. Thus, we proved that Φ is nonempty.

Appendix E: Aggregation Error Proofs

E.1. Proof of Proposition 5

We show that if $\zeta_{sum} \leq G(H - L + 1)$, then $TH_g \leq H$, for all g. We use a contradiction as follows. Assume $\zeta_{sum} \leq G(H - L + 1)$ and there exists g' such that $TH_{g'} \geq H + 1$. Using Proposition 3(b), we must have $TH_g \geq H - L + 1$, for all g. Note that if there exists g'' such that $TH_{g''} \leq H - L$, then $TH_{g'} - TH_{g''} \geq L + 1$ which contradicts Proposition 3(b). Hence, we showed that $TH_g \geq H - L + 1$, for all g. Then, $\zeta_{\mathrm{sum}} = \sum_g TH_g > G(H - L + 1)$ which contradicts our earlier assumption. Therefore, $TH_g \leq H$, for all g.

E.2. Proof of Proposition 6

Consider $G(H - L + 1) + 1 \leq \zeta_{sum} \leq GH$. We first show that this interval is nonempty if $L \geq 2$. We have:

$$1 \leq \zeta_{\mathrm{sum}} - G(H - L + 1) \leq G(L - 1)$$

The right-hand-side is lower bounded by G. If $G = 1$, we have at least one feasible point. Therefore, the interval for ζ_{sum} is always nonempty. Using Proposition 5, if $\zeta_{\mathrm{sum}} \leq G(H - L + 1)$, then the HGSP solution is optimal and hence %RI $= 0$. Thus, the optimal value of ζ_{sum} must satisfy $G(H - L + 1) + 1 \leq \zeta_{sum} \leq GH$. Moreover, since this interval is nonempty and %RI is always non-negative, we can limit our search to this interval.

Let us fix ζ_{sum} and focus on maximizing the numerator of the objective function. We note that, for some groups, we have $TH_g > H$, and for the remaining groups we have $TH_g \leq H$. Thus, we can assume that the groups with $TH_g > H$ are located first in the list. Therefore, there must exist an optimal solution that satisfies the following condition: there exists g^* such that $TH_g > H$ for all $g \leq g^*$ and $TH_g \leq H$ for all $g > g^*$. Thus, we can limit our search to finding the optimal g^*.

Appendix F: Notations

Abbreviations:

AFSDP	Aggregate fixed-length single day problem
AFSDP*	An alternative formulation for the AFSDP
AFSDP^r	The linear programming relaxation of the AFSDP*
AFSP	Aggregate fixed-length stochastic problem
AGDP	Aggregate general deterministic problem
AGSP	Aggregate general stochastic problem
ASDP	Aggregate single day problem
ASDP*	An alternative formulation for the ASDP
DAP	Disaggregation problem
DAP^r	The LP relaxation of the DAP
DLCC	Direct load control contract

DP	Dynamic programming
ECP	Energy consumption profile
FLP	Fixed-length policy
FSP	Fixed-length stochastic problem
GDP	General deterministic problem
GSP	General stochastic problem
HGSP	Our heuristic approach for solving the GSP
LBH	Lower-bound heuristic
LP	Linear programming
MILP	Mixed-integer linear programming
RAGDP	The LP relaxation of the AGDP
SDP	Single day problem
UBH	Upper-bound heuristic
WRIP	worst-case %RI problem

Notations:

$g = 1, \ldots, G$	Groups.
$d = 1, \ldots, D$	Days.
$t = 1, \ldots, T$	Time periods in a day.
$\omega \in \Omega$	Day-types.
ω_d	Type of day d.
l_{dg}	Duration of the call for group g in day d.
m_{dg}	1, if group g reduces their consumption in day d (if $l_{dg} > 0$), and 0 otherwise.
h_{dg}	Remaining hours that group g can reduce their consumption in days $d, \ldots, D$.
k_{dg}	Remaining number of times that group g can be called in days $d, \ldots, D$.
$\mathbf{L}_d$	Vector $\mathbf{L}_d = (l_{d1}, \ldots, l_{dG})$
$\mathbf{M}_d$	Vector $\mathbf{M}_d = (m_{d1}, \ldots, m_{dG})$
$\mathbf{H}_d$	Vector $\mathbf{H}_d = (h_{d1}, \ldots, h_{dG})$
$\mathbf{K}_d$	Vector $\mathbf{K}_d = (k_{d1}, \ldots, k_{dG})$
L	Upper bound on the duration of a call.
H	Upper bound on the total duration of the calls for a group in the entire horizon.
K	Upper bound on the number of the calls for a group in the entire horizon.
$\mathbf{\Theta}_d$	The set of feasible decisions $(\mathbf{L}_d, \mathbf{M}_d)$ on day d.
$\mathbf{\Theta}_d^{\mathrm{F}}$	The set of feasible decisions under FLP on day d.
$\mathbf{\Theta}_d^{\mathrm{AF}}$	The set of feasible decisions for the AFSP on day d.
$\mathbf{\Theta}_d^{\mathrm{AD}}$	The set of feasible decisions for the AGDP on day d.
$\mathbb{I}(.)$	The indicator function, i.e., $\mathbb{I}(z) = 1$ if $z > 0$ and otherwise $\mathbb{I}(z) = 0$.
$\mathcal{V}_d(\mathbf{H}_d, \mathbf{K}_d)$	The optimal cost for days $d, d+1, \ldots, D$ given $\mathbf{H}_d$ available time and $\mathbf{K}_d$ available number of calls to groups.
$\mathcal{V}_d^{\mathrm{F}}(\mathbf{K}_d)$	The optimal cost for days $d, d+1, \ldots, D$ under FLP, given $\mathbf{K}_d$ available calls.
$\mathcal{V}_d^{\mathrm{AF}}(\mathscr{K}_d)$	The optimal cost for days $d, d+1, \ldots, D$ for AFSP, given a total of $\mathscr{K}_d$ available calls.
$\mathcal{V}_d^{\mathrm{AD}}(\mathscr{H}_d, \mathscr{K}_d)$	The optimal cost for days $d, d+1, \ldots, D$ for the AGDP, given a total of $\mathscr{H}_d$ available hours and $\mathscr{K}_d$ available calls.
$\psi(\mathbf{L}_d, \omega_d)$	The minimum cost for day d of type ω_d using the calls $\mathbf{L}_d = (l_{d1}, \ldots, l_{dG})$.
$\psi^{\mathrm{AF}}(\mathcal{K}_d, \omega_d)$	The minimum cost for AFSDP for day d of type ω_d using $\mathcal{K}_d$ calls.
$\psi^{\mathrm{AD}}(\mathcal{H}_d, \mathcal{K}_d, \omega_d)$	The minimum cost for the ASDP on day d of type ω_d using $\mathcal{H}_d$ hours and $\mathcal{K}_d$ calls.
$r_t(\omega_d)$	The height of ECP in period t and day d.
$f(.)$	Energy cost function.

$\hat{f}_{d,t}(.)$	Energy saving function at period t of day d. Note: $\hat{f}_{d,t}(z)$ measures the saving at period t of day d given that the height of the ECP has reduced to z.
$\partial f(.)$	The first difference of function f, i.e., $\partial f(z) := f(z) - f(z-1)$, $z \in \mathbb{Z}$, $z \geq 1$.
$\partial^2 f(.)$	The second difference of function f, i.e., $\partial^2 f(z) := \partial f(z) - \partial f(z-1)$, $z \in \mathbb{Z}$, $z \geq 1$.
u_{gt}	1 if group g is on call in period t of day d, and 0 otherwise.
w_{gt}	1 if group g begins reducing their consumption in period t of day d, and 0 otherwise.
$[.]^+$	The function that is defined as $[z]^+ := \max\{z, 0\}$.
$\mathcal{K}_d$	The total number of calls to be used in day d.
$\mathcal{H}_d$	The total hours to be used in day d.
$\mathscr{K}_d$	The number of available calls that can be used in days $d, d+1, \cdots, D$.
$\mathscr{H}_d$	The available hours that can be used in days $d, d+1, \cdots, D$.
s_t	The number of groups that start reducing their energy consumption at period t in day d.
$s_{t\ell}$	The number of calls with duration ℓ and starting time t.
R_t	The height of the ECP at period t of day d after the customers reduce their energy consumption.
$\mathcal{D}_d$	The set of calls that we must use in day d based on the optimal solution of AFSP.
$\mathbf{D}$	Defined as $\mathbf{D} := \bigcup_{d=1}^{D} \mathcal{D}_d$.
TC_g	The total number of calls to group g in the entire planning horizon.
TH_g	The sum of the call durations that are assigned to group g in the entire planning horizon.
n	problem size, $n \in \mathbb{Z}$, $n \geq 1$.
$v_{\text{GSP}}^{(n)}$	The optimal saving for the GSP where the problem size is n.
$v_{\text{HGSP}}^{(n)}$	The HGSP saving where the problem size is n.
$v_{\text{AGDP}}^{(n)}$	The optimal saving for the AGDP where the problem size is n.
$v_{\text{AGSP}}^{(n)}$	The optimal saving for the AGSP where the problem size is n.
$v_{\text{RAGDP}}^{(n)}$	The optimal saving for the RAGDP where the problem size is n.
ρ	A non-negative constant, independent of n, as the upper bound on the aggregation error.
$j = 1, \ldots, J$	Offers.
q_j	The allocation probability for offer j.
x_j	The number of times that offer j is made in the planning horizon.
$\mathbf{x}^{*(n)}$	The vector $(x_1^{*(n)}, \ldots, x_J^{*(n)})$ and the optimal solution of the RAGDP$^{(n)}$.
$\gamma = 1, \ldots, \Gamma$	Resources.
$a_{\gamma j}$	The amount of resource γ used when offer j is made.
c_γ	The total available resource γ at the beginning of the planning horizon.
$\mathbf{a}_\gamma$	The vector $(a_{\gamma 1}, \ldots, a_{\gamma J})$.
$a_{\gamma, \max}$	Defined as $\max_j\{a_{\gamma j}\}$ for all γ.
θ_j	The saving from offer j.
$\theta_{\max}$	Defined as $\theta_{\max} := \max_j\{\theta_j\}$.
$\omega(j)$	The day-type in which offer j can be used.
p_ω	The proportion of days of type ω (the probability of day-type ω).
$d^{*(n)}$	The minimum of nD and the last day in which the condition (†) is satisfied.
$\Delta_\gamma^{(n)}(d)$	The deviation from the expected resource usage for γ on day d.
$\Im_d$	The accumulated information up to day d.
$\mathbf{E}[.]$	The expected value function.
$P(.)$	The probability function.

$\lvert . \rvert$	The cardinality/absolute value function.
$\lceil . \rceil$	The ceiling function.
$\mathcal{L}(t)$	Defined as $\mathcal{L}(t) := \min\{L, T-t+1\}$ which denotes the maximum duration of a call starting from t.
$s_{t\ell}$	The number of calls with duration ℓ and starting time t.
$\eta \in \mathcal{D}_d$	Calls that are scheduled on day d. A call is in the form of $\eta = (st(\eta), \lvert \eta \rvert)$ where $st(\eta) \in \mathbb{Z}_+$ is the starting time and $\lvert \eta \rvert \in \mathbb{Z}_+$ is the length of the call.
$\varsigma = 1, \ldots, K$	Classes.
$\lambda_{i\varsigma g}$	1 if call i from class ς is assigned to group g, and 0 otherwise.
$\mathcal{G}_g$	The set of calls that are assigned to group g.
$\mathcal{B}, \mathcal{E}_1, \mathcal{E}_2$	The coefficient matrices for the DAP.
Φ	Defined as $\Phi := \{\lambda \in \mathbb{R}^{KG} \mid \mathcal{B}\lambda = \mathbf{1}_K, \mathcal{E}_1\lambda = \mathbf{1}_{D_1}, \mathcal{E}_2\lambda \le \mathbf{1}_{D_2}, \lambda \ge \mathbf{0}\}$.
ζ_{sum}	The sum of the durations of all calls.
δ	Defined as $\delta := \max_\omega \{\psi^{\text{AD}}(1, 1, \omega)\}$.
$\zeta_{i\varsigma}$	The duration of the ith call in class ς.
$\zeta_{\min,\varsigma}, \zeta_{\max,\varsigma}$	The minimum and maximum $\zeta_{i\varsigma}$ in class ς.
%RI	Percentage of relative infeasibility
%WRI	Worst-case for %RI

Acknowledgement

Endnotes

1. http://www.sandc.com/en/outages#Solutions

2. http://www.popularmechanics.com/science/energy/a6013/how-a-smarter-grid-can-prevent-blackouts/

3. Creating identical groups may be somewhat difficult in some situations; however, according to the energy company we are working with, it is common to assume there are identical groups with the same contracts and power consumption.

References

Adelman D (2007) Dynamic bid prices in revenue management. *Operations Research* 55(4):647-661.

Alizamir S, de Vericourt F, Sun P (2016) Efficient feed-in-tariff policies for renewable energy technologies. *Operations Research* 64(1):52-66.

Baldick R, Kolos S, Tompaidis S (2006) Interruptible electricity contracts from an electricity retailer's point of view: valuation and optimal interruption. *Operations Research* 54(4):627-642.

California Independent System Operator (2012) `www.caiso.com`.

Cooper WL (2002) Asymptotic behavior of an allocation policy for revenue management. *Operations Research* 50(4):720-727.

Garey MR, Johnson DS (1979) Computers and Intractability: A Guide to the Theory of NP-Completeness. WH Freeman & Co., San Francisco.

Jasin S, Kumar S (2012) A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* 37(2):313-345.

Kamat R, Oren SS (2002) Exotic options for interruptible electricity supply contracts. *Operations Research* 50(5):835-850.

Meulemeester BD (2014) Capacity payments: expensive solution for a non-existing problem. `http://energypost.eu/capacity-payments-expensive-solution-non-existing-problem/`.

Michigan Public Service Commission (2015) Common Demand Response Practices and Program Designs. `https://www.michigan.gov/documents/energy/Common_Practices_Feb22_522983_7.pdf`.

Next-Kraftwerke-Belgium (2016) Merit order curve. `https://www.next-kraftwerke.be/`

Oren SS, Smith SA (1992) Design and management of curtailable electricity service to reduce annual peaks. *Operations Research* 40(2):213-228.

Peura H, Bunn DW (2015) Dynamic Pricing of Peak Production. *Operations Research* 63(6):1262-1279.

Posner B (2015) The Fundamentals of Electricity Markets. `https://www.e-education.psu.edu/ebf200wd/node/151`

Solar Energy Industries Association (2016) `www.seia.org`.

Taylor B, Taylor C (2015) Demand Response: Managing Electric Power Peak Load Shortages with Market Mechanisms. `http://www.raponline.org/wp-content/uploads/2016/05/rap-china-demandresponsemanagingpeakloadshortages-2015-mar.pdf`.

Zhang D, Adelman D (2009) An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* 43(3):381394.