Utilities' Managed Home Charging Programs for Electric Vehicles

Ali Fattahi
Johns Hopkins University, Carey Business School, Baltimore, Maryland 21202, USA, ali.fattahi@jhu.edu

Experts estimate 20 million electric vehicles will be on U.S. roads by 2030, and the majority (around 80%) of the electric vehicle drivers will use home charging. Many utilities are designing *managed home charging* programs to centrally manage charging times to reduce cost, avoid new and aggravated peaks and blackouts, and ensure grid stability. These managed home charging programs are either *active*, in which the utility continuously controls the charging while the vehicle is plugged in, or *passive*, in which the participants decide when to charge based on pre-announced low-rate episodes. We study jointly designing and executing these active and passive programs. We present a *program-design* model, which produces a menu of the charging programs, tailored for each driver type, and a *load-management* model, which dynamically manages the load supply to each individual participant. The load-management model consists of a large number of non-homogeneous participants, and it is a large-scale mixed-integer nonlinear stochastic problem. We present an effective approximation method, conduct thorough theoretical and numerical analyses of our approximation, and provide worst-case bounds for its error components. Our methodology provides detailed insights on the amount and timing of the improvements achievable in cost and demand variability by offering managed home charging programs, and by customizing the passive programs. It also offers detailed insights on the significance of the trade-off between cost and demand variability. We find promoting a culture of charging electric vehicles every night may significantly increase utilities' total cost if passive programs have high participation levels.

Key words: electric vehicles, managed charging programs, home charging, large-scale optimization, approximation, error analysis

Acknowledgement: We thank Garrett Fitzgerald, PhD, Senior Director, Electrification, *Smart Electric Power Alliance* (SEPA), for the continued support and providing feedback throughout this project.

1. Introduction

An estimated 20 million electric vehicles (EVs) will be on U.S. roads by 2030. Therefore, experts believe EVs are "the most significant new electric load since the rise of air conditioning in the 1950s" (Myers 2019, Hanvey 2019, Trabish 2019, Cooper and Schefter 2018, Blair et al. 2021). Reasonably assuming an EV on average requires 3.8 megawatt-hours (MWh) per year (GITT and ISATT 2019), 20 million EVs require 76 terawatt-hours (TWh) in a year. The National Renewable Energy Laboratory predicts electrified transportation may lead to 58-336 TWh of consumption per year by 2030, which is equivalent to an average annual consumption of 5.6 million-32.3 million U.S. homes (NREL 2018, Myers 2019). If this load is effectively integrated and managed, it represents a paramount opportunity for utilities; otherwise, it can pose a significant challenge in operating

and balancing the grid loads (Hanvey 2019, Myers 2019, Blair et al. 2021).

An EV supply equipment — such as a Level-1 or a Level-2 home charging station, or a Level-3 fast charging station — is required to charge an EV, which is also referred to as a *charging station*. Home charging, using Level-1 or Level-2 ports, is the most common practice, representing around 80% of all charging done by EV drivers (ChargeHub 2020). Cooper and Schefter (2018) estimate around 9.6 million charging ports will be needed by 2030 to satisfy the EV load requirement, of which 7.5 million are expected to be home-charging stations.

An EV driver with an installed charging station usually keeps her EV plugged in overnight, after returning home from work, until she leaves the next morning. We refer to the time interval in which an EV is left plugged in as a *charging window*. The time needed to charge an EV depends on the *battery size*, the *speed* of the charger, and whether it is an *empty-to-full* charging or a *top-up* charging. Most drivers prefer top-up charging instead of letting the battery run empty (pod-point.com); for example, an EV driver with an 80% charge may plug in to fully charge the battery by charging overnight at home. Thus, conveniently, at a full charging speed, an EV load requirement is usually satisfiable in a much shorter time than the charging window. Additionally, most EV drivers do not have a preference for when their EV is charged within their charging windows. Such flexibilities provide a significant opportunity for managing the EV load.

Managed home charging (MH) programs allow utilities to centrally manage EV drivers' home charging times to avoid new and aggravated peaks. In the absence of MH programs, EV drivers naturally start charging their EVs at their full charging speeds after returning home in the afternoon/evening, which may create new peaks and significantly aggravate the existing peaks. Reducing energy consumption during peak periods has been among the primary objectives for the utilities, and various programs and solutions have been designed over the past several decades (see, e.g., Oren and Smith 1992, Ata et al. 2018, Agrawal and Yücel 2022, Alizamir et al. 2020, Fattahi et al. 2023a,b, Keskin et al. 2020). Thus, MH programs are essential for effectively integrating and managing the EV load, to realize "the maximum benefits for consumers, the grid, and society as a whole" (Myers 2019, Blair et al. 2021).

An MH program is either an (i) *active* program (AMH), also referred to as direct load control, in which the utility continuously controls the EV charging while the vehicle is plugged in, or a (ii) *passive* program (PMH), also known as behavioral load control, in which the EV driver decides when to charge based on pre-announced *low-rate episodes* (Fitzgerald and Dougherty 2021, Blair et al. 2021). As an example of AMH, in 2019, Eversource launched a new EV load-management program, through which it has "access to customer charging data and the ability to control" charging during peak times, and, in return, participants receive a one-time bonus of "$300 either through a $300 rebate for newly purchased chargers or a combined incentive of $150 to enroll and $50 per year of

load control for 3 years" (Blair et al. 2021). As an example of PMH, Southern California Edison (SCE: www.sce.com) offers "TOU-D-PRIME Rates," in which 9 p.m. to 4 p.m. (the next day) is the low-rate episode and priced at 19 ¢/KWh, whereas the cost of charging outside of the low-rate episode is 45 ¢/KWh. In a recent survey, of the 84 utilities that responded to the survey, 53% were interested in managed charging programs (Myers 2019). Fitzgerald and Dougherty (2021) provide an up-to-date assessment of approximately 40 managed charging programs by different utilities. Blair et al. (2021) present seven case studies of such programs.

In this paper, we present a *program-design model* (PD) and a *load-management model* (LM) to jointly design and execute AMH and PMH programs over a "season" (e.g., summer, winter, or a calendar year). We provide an overview of PD and LM below.

Program-design model (PD). During the "subscription phase," which precedes the season, a utility presents a menu of its MH programs to the EV drivers. Each driver selects a program and subscribes to it after examining the menu, or she decides to subscribe to none of the offered MH programs, which we refer to as NMH. Using PD, the utility optimally designs its menu of MH programs, with the objective of minimizing the total expected cost, which consists of the total bonus paid to AMH participants, the lost revenue due to PMH participants, and the total cost of procuring electricity during the season. We construct PD as follows. First, we form driver types such that the members of a type are homogeneous in their total EV load requirement over the season, the flexibility of their charging times, preferences toward different MH programs, and their compatibility with the potential low-rate episodes. Second, we evaluate the appropriateness of each MH program for each driver type and decide which MH programs should be offered to each driver type. Third, we formulate EV drivers' choice probabilities to estimate the participation levels in MH programs as well as the utility's total cost. Last, by solving PD, we optimize the AMH bonus and PMH low rate.

Load-management model (LM). The subscribers of each program are known within the season, and the utility needs to dynamically manage the supply to each AMH participant. We present a large-scale mixed-integer nonlinear stochastic optimization model to address the utility's load-management decisions. LM considers a large number of *non-homogeneous* participants. These participants have different *arrival times*, *charging windows*, *load requirements*, and *charging speed limits*. Each AMH participant plugs in her EV at some time and indicates her required load and charging window. This information is transmitted to the utility. The utility is *obligated* to satisfy each AMH participant's load requirement within her charging window. The utility minimizes the total energy cost by dynamically managing the load supply to each individual participant. LM also determines a low-rate episode for each PMH participant on a daily basis. We solve LM on a rolling-horizon basis, at the beginning of each period (e.g., hour). After each time solving LM,

we only implement the first-period decisions and then continue to the next period. The rolling-horizon feature allows for updating all inputs on an hourly basis, incorporating the most recent information.

1.1. Summary of Main Results and Insights

LM is a large-scale mixed-integer nonlinear stochastic problem with three sources of difficulty. First, it has a significantly large number of decision variables and constraints, which are introduced to dynamically manage each AMH participant's charging rate over her charging window. Second, it has a long horizon. Third, its objective function is nonlinear. We present a scalable solution method that effectively addresses these three challenges. Our approach consists of three approximations, which we briefly outline below. First, we partition the AMH participants into groups based on their load requirement and home charging speed limit, referred to as *aggregation*, which transforms LM to a practically sound model, because its size no longer grows in the number of participants. Second, we truncate the long horizon (season) to a shorter horizon (e.g., 15 hours), which we refer to as *truncation*. Consider an AMH participant with a charging window of t' periods within the truncated horizon and t'' periods after the truncated horizon ends. To accommodate this participant in our truncation, we assume we need to supply the $\left(\frac{t'}{t'+t''}\right)$-fraction of her requested load within the truncated horizon. Third, we linearize the objective function by approximately assuming the cost function is constant over intervals of length δ. This approximation, referred to as *linearization*, transforms our model to a mixed-integer linear program (MILP). We present thorough theoretical analyses and worst-case error bounds on the components of our approximation. We confirm the effectiveness and near optimality of our approximation through extensive numerical experiments. We use real data from California Independent System Operator (CAISO) to simulate instances for LM. We find our approximation produces near-optimal solutions to the large instances of LM in a reasonable amount of time.

Our methodology can effectively be employed as a decision support tool to address various managerial questions that arise when designing and promoting MH programs. It offers valuable managerial insights into the potential improvements in cost and demand variability achievable through the implementation of AMH, PMH, and a combination of both, as well as the customization of PMH. It also provides detailed insights into the trade-off between cost and demand variability. Last, we explore the effects of charging frequency on utilities' costs. Some EV drivers may skip charging on certain nights for various reasons, whereas others may prefer charging nightly to avoid running out of power. Households with multiple EVs might alternate charging each vehicle every other night to save costs associated with installing new chargers. We find that increased charging frequency among AMH participants could benefit utilities by enhancing load management

flexibility. However, promoting nightly EV charging could substantially increase utilities' costs if NMH and/or PMH participation is high.

In summary, our contributions are as follows. First, to the best of our knowledge, we present the first models for jointly designing and executing AMH and PMH programs. Second, we present an approximation method that consists of aggregation, truncation, and linearization, which transforms our large-scale mixed-integer nonlinear program LM to a MILP with a practically reasonable size. Third, we present thorough theoretical analyses of our approximation and provide worst-case bounds for its error components. Fourth, we perform extensive numerical experiments that confirm the effectiveness of our methodology. Last, we present several managerial insights on how utilities should design and promote their MH programs. Our methodology can be used as a decision support tool to address various managerial questions regarding MH programs.

The remainder of this paper is organized as follows. We present a literature review in §2. Our model formulation is given in §3. We present our approximation method and its theoretical analyses in §4. We provide the results of our numerical experiments and offer several managerial insights in §5, followed by our concluding remarks.

2. Literature Review

Our paper contributes to the domains of EV operations management and EV managed charging programs. Lim et al. (2015) study the major barriers to mass adoption of EVs, Avci et al. (2015) study the key mechanisms driving EV adoption, and Agrawal et al. (2022) study the role of dealer demonstration in the EV adoption. Shi et al. (2022) study an optimal subsidy structure by modeling the interaction between the government and a charging supplier. He et al. (2021) integrate vehicle repositioning and charging infrastructure planning. Varma et al. (2023) study an EV fleet and charging infrastructure capacity planning problem. Zhang et al. (2021) study EV sharing systems and vehicle-to-grid operations. Schneider et al. (2018), Sun et al. (2019), and Mak et al. (2013) study battery-swap stations. Different from these papers, we study jointly designing and executing MH programs.

Optimal management of EV charging has recently been studied by Jin et al. (2013), Zhang et al. (2013), Jiang and Powell (2016), Wu et al. (2021), and Chen et al. (2023). In particular, Jin et al. (2013) consider deterministic EV arrivals, whereas in our paper, the participants' plug-in times, charging windows, and charging requirements are stochastic. Our paper is different from Zhang et al. (2013) and Jiang and Powell (2016) because they study a charging station with multiple charge points, whereas we study management of the charging supply of a large number of home charging stations. In Wu et al. (2021), participants are incentivized by a price menu to delay their charging, whereas we study jointly desiging and executing AMH and PMH, where in AMH, a utility

directly manages the charging supply for each participant, and in PMH, the utility determines a low-rate episode for each participant and announces it ahead of time. Chen et al. (2023) study the scheduling of vehicle charging for a service provider with a limited number of chargers, where non-homogeneous participants of different types (defined based on arrival/departure times and charging requirements) arrive according to a Poisson process and employ an exponential cone programming approach to solve their stochastic optimization problem. Different from their paper, in LM, a utility dynamically manages supplies and low-rate episodes for a large number of non-homogeneous participants who vary in their arrival times, charging windows, charging requirements, and charging speed limits; we do not assume the participants' arrival process is Poisson; and our solution approach is different and consists of aggregation, truncation, and linearization.

Methodologically, our aggregation is related to the classical aggregation-disaggregation methodology (see, e.g., Whitt 1978, Mendelssohn 1982, Bean et al. 1987, Bertsekas et al. 1988, Rogers et al. 1991, Van Roy 2006). In our aggregation, we define a measure that is unique to our problem, called *normalized load requirement*, to form groups. We design a disaggregation procedure, show it always produces a feasible solution, and present a worst-case bound on the error of our aggregation. Our scenario-based modeling of uncertainty is a common approach in modeling uncertainty in the context of energy management (see, e.g., Fattahi et al. 2023a,b). This scenario-based approach allows us to model our stochastic problem as a large-scale mixed-integer nonlinear program. To the best of our knowledge, our truncation approach, which considers the $\left(\frac{t'}{t'+t''}\right)$-fraction of each participant's load requirement, as well as our theoretical analyses of the truncation, are new. Our linearization is related to the classical approach of approximating a nonlinear function using a piecewise linear function (see, e.g., Bertsimas and Tsitsiklis 1997), which partitions the domain into subsets and assumes the function is linear over each subset. We also partition the domain into intervals of length δ; however, we assume the function is constant over each interval. We present thorough theoretical analyses and error bounds on the components of our approximation method.

3. Model Formulation

We introduce our preliminaries in §3.1. In §3.2, we tailor each MH program for each type of EV driver. In §3.3, we present a choice model for EV drivers. We formulate PD and LM, respectively, in §3.4 and §3.5.

3.1. Preliminaries

We introduce our preliminaries below.

Season and subscription phase. We consider a "season" (e.g., summer, winter, or a calendar year) over which a utility executes its MH programs. During the "subscription phase," which

precedes the season, the utility presents a menu of its MH programs to the EV drivers who perform home charging. Each EV driver selects a program and subscribes to it after examining the menu, or she decides to subscribe to none of the offered MH programs, which we refer to as NMH. That is, an NMH driver is an EV driver with an installed charger at home, who performs home charging, but she does not participate in MH programs. If an EV driver subscribes to a program, she remains in the same program throughout the entire season.

Periods (e.g., hours) and days. Let $t \in \{1, \dots, T\}$ denote the time periods in the season; for example, $t = 1$ and $t = T$ are the first and last periods in the season, respectively. Without loss of generality, in this paper, we assume each period is an hour. Also, we let index t refer to the beginning of hour t and assume the EV drivers' arrivals and departures (i.e., plug-in and plug-out times) occur at the beginning of periods. Let $d \in \{1, \dots, \mathcal{D}\}$ denote the days in the season. We assume each driver has $\mathcal{D}$ non-overlapping charging events within the season. We allow the required load of a driver to be zero in some charging events, accommodating the days when the driver does not charge her car. We use index d to represent the d^{th} charging event of each driver.

Charging options. We assume each EV driver has an installed charger at home. And each EV driver belongs to one of the following groups over the season.

- AMH: In each charging event, an AMH participant plugs in her EV at some time and indicates her charging requirement and plug-out time. The utility fulfills the requirement within her charging window. The driver pays the common electricity rate $\bar{\mu}$ (\$/KWh) for charging her EV, and she receives a one-time bonus β (\$) for participating in AMH.
- PMH: The utility announces the starting time (τ), duration ($\mathcal{L}$), and electricity rate (μ^{L}) for the *low-rate episode* ahead of time. The utility may customize the low-rate episode (§3.2). PMH participants enjoy a lower rate μ^{L} ($\leq \bar{\mu}$) for charging within the low-rate episode.
- NMH: An NMH driver does not participate in AMH or PMH. In each charging event, she starts charging her EV at her full charging speed after plugging in her EV until her load requirement is fulfilled. She pays the common electricity rate $\bar{\mu}$ and receives no bonus.

Charging events. In her d^{th} charging event, driver j plugs in her EV at the beginning of some period $\tilde{t}^{\text{A}}_{j,d}$ (arrival or plug-in time) and requires some load amount $\tilde{Q}_{j,d}$ to be charged until she leaves at the beginning of some period $\tilde{t}^{\text{D}}_{j,d}$ (departure or plug-out time). We use tilde ($\tilde{}$) to distinguish random variables in this paper. We have $1 \leq \tilde{t}^{\text{A}}_{j,1} < \tilde{t}^{\text{D}}_{j,1} \leq \tilde{t}^{\text{A}}_{j,2} < \tilde{t}^{\text{D}}_{j,2} \leq \cdots \leq \tilde{t}^{\text{A}}_{j,\mathcal{D}} < \tilde{t}^{\text{D}}_{j,\mathcal{D}} \leq T + 1$ for each driver j. In AMH, the utility decides how much load to supply to driver j in each period $\tilde{t}^{\text{A}}_{j,d}, \dots, \tilde{t}^{\text{D}}_{j,d} - 1$ to satisfy her requirement $\tilde{Q}_{j,d}$. By contrast, the driver decides the charging time in PMH and NMH.

Let $\tilde{\ell}_{j,d} \triangleq \tilde{t}^{\text{D}}_{j,d} - \tilde{t}^{\text{A}}_{j,d} \geq 1$ denote driver j's charging window in her d^{th} charging event; that is, her car is plugged in for $\tilde{\ell}_{j,d}$ hours. Let $q^U_j > 0$ denote her maximum charging speed, and define $\tilde{\gamma}_{j,d} \triangleq \frac{\tilde{Q}_{j,d}}{q^U_j}$

as her *normalized load requirement*, which indicates the amount of time needed to charge her EV at the full charging speed to satisfy her requirement. We assume $\tilde{\gamma}_{j,d} \le \tilde{\ell}_{j,d}$ to ensure feasibility, meaning she plugs in her EV for at least $\tilde{\gamma}_{j,d}$ hours in her d^{th} charging event.

Total load requirement and flexibility. For driver j, define

$$\mathcal{Q}_j \triangleq q_j^U \sum_{d=1}^{\mathcal{D}} \mathbf{E}[\tilde{\gamma}_{j,d}], \quad \mathcal{I}_j \triangleq \sum_{d=1}^{\mathcal{D}} \mathbf{E}[\tilde{\ell}_{j,d} - \tilde{\gamma}_{j,d}],$$

where $\mathcal{Q}_j$ is the total expected EV load requirement of driver j over the season and $\mathcal{I}_j$ is the total expected time the EV is plugged in beyond the total time that is actually needed to charge the EV. We propose $\mathcal{I}_j$ as a proxy for the amount of *flexibility* driver j could offer to the utility for managing her charging times over the season, if she participates in AMH. Utilities can estimate $\mathcal{Q}_j$ and $\mathcal{I}_j$ through analyzing the drivers' charging history and/or by asking questions regarding their driving routines. Utilities must target the drivers with large $\mathcal{I}_j$ for their AMH programs.

Driver types. We partition EV drivers into a small number of driver types, denoted by $\varphi \in \Phi$. We assume the type-φ drivers are homogeneous in terms of (H1) total load $\mathcal{Q}_j$ and flexibility $\mathcal{I}_j$, (H2) preferences toward different charging programs, and (H3) compatibility with the potential low-rate episodes. We rigorously specify these conditions in §3.4. Let $\mathcal{K}_\varphi$ denote the number of type-φ drivers. Let $\Pi \triangleq \{\text{NMH}, \text{AMH}, \text{PMH}\}$ denote the set of all three home charging options that we consider in this paper. The utility offers a subset of these programs to each driver type φ, denoted by $\Pi_\varphi \subseteq \Pi$. During the subscription phase, each type-φ driver chooses an option $\pi \in \Pi_\varphi$, subscribes to it, and remains in it throughout the season. We design Π_φ in §3.2.

With some abuse of notation, we replace index j with φ to denote a common quantity for the drivers of type φ; for example, $\mathcal{I}_\varphi$ denotes the common load flexibility of driver type φ.

Utility's cost function. Let $f_t(y_t)$ denote the utility's cost of procuring its total supply of y_t in period t. In this paper, we assume (i) the cost function is additive in hours, (ii) each hourly cost function $f_t(y)$ is non-decreasing in the hourly consumption y, for $y \ge 0$, and (iii) $f_t(y) = 0$, for $y \le 0$, for all t. These assumptions are mild. In a few places, we consider a cost function in the form of $f_t(y) = cy^3$, for $y \ge 0$, for some $c > 0$, as a special case, which we explicitly state. Fattahi et al. (2023b) show this cubic function is a good fit for the energy generation cost in CAISO.

Grid constraints. The scope of our problem includes $g \in \{1, \ldots, G\}$ service regions, which are served by several substations. Substation i serves a subset of the service regions $\{g | g \in i\}$. Driver j's service region is denoted by g_j. In each period, the total load supplied by substation i cannot exceed its limit, which is denoted by h_i.

3.2. Designing MH Programs

We show how the utility designs its offer set Π_φ for each driver type φ. We include NMH in all offer sets; that is, NMH $\in \Pi_\varphi$, for all φ, representing the *no-participation* option. The utility offers AMH and/or PMH to a driver type if the program is compatible with their driving routines and worthwhile for the utility. For example, PMH may be unsuitable for a driver type with irregular arrival and departure times; PMH participants with long and inflexible charging times are worthless for the utility; and AMH participants with high load requirements and short charging windows are worthless for the utility. We present eligibility requirements to determine which MH programs should be offered to each driver type, while ensuring the utility's interest and the participants' high satisfaction. Eligibility requirements that are defined based on the customers' electricity consumption profiles are common in the existing demand-response programs (see, e.g., PG&E 2023, SCE 2023).

Designing AMH. We introduce a threshold $\mathcal{I}^\circ$ that determines whether driver type φ qualifies for AMH; we add AMH to Π_φ if $\mathcal{I}_\varphi \geq \mathcal{I}^\circ$. By increasing $\mathcal{I}^\circ$, fewer drivers qualify for AMH, but each of them provides a higher value to the utility. We assume $\mathcal{I}^\circ$ is exogenous and given by the utility. Last, the utility optimizes the AMH bonus β by solving PD (§3.4).

Designing PMH. A *low-rate episode* is specified by the following three elements: starting time τ, duration $\mathcal{L}$, and rate μ^{L}. We propose customizing the starting time each day, while keeping the duration and rate constant throughout the season and across the drivers. Customizing μ^{L} could cause price discrimination, leading to dissatisfaction among the participants; hence, we do not customize μ^{L}, due to its practical challenges. Customizing $\mathcal{L}$ has practical challenges as well because, for example, a short low-rate episode may restrain the drivers from fulfilling their requirement within the low-rate episode, leading to dissatisfaction. On the other hand, a long low-rate episode is ineffective because the drivers usually fulfill their charging in the first few hours of the low-rate episode. We propose designing a customized PMH through the following three steps.

1. Duration $\mathcal{L}$: We assume $\mathcal{L}$ is exogenous and given by the utility. Utilities must target EV drivers with small load requirements for PMH; otherwise, the charging times would be long and inflexible. For example, the utility may target the EV drivers j such that $P(\tilde{\gamma}_{j,d} \leq 6) = 1$, for all d, meaning each of the normalized load requirements is at most six hours with probability 1. In this example, the utility may set $\mathcal{L} = 6$ hours, ensuring the PMH participants can always fulfill their requirement within the low-rate episode.

2. Potential starting times $\mathcal{T}_{\varphi,d}$: Let $\mathcal{T}_d \subseteq \{1, \ldots, T\}$ denote the set of all starting times that are worth considering by the utility on day d; for example, assuming $\mathcal{L} = 6$ hours, $\mathcal{T}_d$ could consist

of the periods corresponding to 9 p.m., 10 p.m., 11 p.m., and 12 a.m. on day d. We assume $\mathcal{T}_d$ is exogenous and given by the utility, and $|\mathcal{T}_d|$ is small. Define

$$\mathcal{T}_{j,d} \triangleq \{\tau \in \mathcal{T}_d |\ P(\tilde{t}^{\mathrm{A}}_{j,d} \le \tau,\ \tilde{t}^{\mathrm{D}}_{j,d} \ge \tau + \mathcal{L},\ \tilde{\gamma}_{j,d} \le \mathcal{L}) = 1\},$$

for all j and d; that is, $\mathcal{T}_{j,d}$ consists of all potential starting times on day d that are compatible with driver j. By properly forming the driver types, as we discuss in §3.4, we ensure the members of each type are homogeneous in $\mathcal{T}_{j,d}$. Thus, let $\mathcal{T}_{\varphi,d}$ denote the common compatible starting times for driver type φ on day d. The utility adds PMH to Π_φ if $\mathcal{T}_{\varphi,d} \neq \{\}$ for all d. During the subscription phase that the utility offers Π_φ to the drivers of type φ, the utility shows $\mathcal{T}_{\varphi,d}$ to the drivers and, as one of the contract terms, the utility guarantees it will assign one of the starting times $\mathcal{T}_{\varphi,d}$ to the drivers of type φ on each day d. Moreover, the utility guarantees it will communicate the assigned starting times to the drivers at a prespecified time $\check{t}_d$ each day d (e.g., 10 a.m.).

3. Rate μ^{L}: The utility optimizes μ^{L} by solving PD (§3.4).

Thus, $\mathcal{L}$ and $\mathcal{T}_d$ are exogenous and given by the utility, whereas μ^{L} is endogenous and optimized by solving PD. The second step ensures the following property:

$$P\left(\tilde{t}^{\mathrm{A}}_{j,d} \le \tau,\ \tilde{t}^{\mathrm{D}}_{j,d} \ge \tau + \mathcal{L},\ \tilde{\gamma}_{j,d} \le \mathcal{L}\right) = 1, \quad \forall \varphi,\ \forall j \in \varphi,\ \forall \tau \in \mathcal{T}_{\varphi,d},\ \forall d, \qquad \text{(CST)}$$

referred to as *compatible starting times* (CST). This property is particularly interesting because it guarantees high satisfaction over the season. In PMH, if the timing of the low-rate episode is set inappropriately, a driver may partially or completely miss the low-rate episode, meaning she needs to perform a portion of her charging outside of the low-rate episode, which causes dissatisfaction. CST helps utilities avoid such occurrences to ensure high satisfaction over the season, which is vital for the success of MH programs.

In short, we showed how the utility designs its MH programs while ensuring its interest and the participants' high satisfaction. We assume $\mathcal{I}^\circ$, $\mathcal{L}$, and $\mathcal{T}_d$ are exogenous and given by the utility. The utility optimizes β and μ^{L} by solving PD.

3.3. Choice Model

We present a choice model to obtain the drivers' choice probabilities for a given $(\beta, \mu^{\mathrm{L}})$. We first determine the drivers' expected costs under different programs. Recall drivers of the same type φ are homogeneous in their total expected load requirements, which we denote by $\mathcal{Q}_\varphi$. Thus, a type-φ driver's expected cost over the season under NMH, AMH, and PMH, respectively, are

$$\mathcal{E}^{\mathrm{NMH}}_\varphi \triangleq \bar{\mu}\mathcal{Q}_\varphi, \quad \mathcal{E}^{\mathrm{AMH}}_\varphi \triangleq -\beta + \bar{\mu}\mathcal{Q}_\varphi, \quad \mathcal{E}^{\mathrm{PMH}}_\varphi \triangleq \mu^{\mathrm{L}}\mathcal{Q}_\varphi.$$

We let the drivers' choice probabilities be proportional to the attractiveness of the programs, which is a common approach in the literature (see, e.g., Gallego et al. 2015, Fattahi et al. 2022). For a type-φ driver, the offered charging programs Π_φ and their expected costs ($\mathcal{E}^\pi_\varphi$) are respectively akin to a set of products offered to a customer and their prices. A driver must select exactly one of the offers, which could be NMH. We define the attractiveness of option $\pi \in \Pi_\varphi$ to driver type φ as $\mathcal{A}^\pi_\varphi \triangleq \hat{\rho}^\pi_\varphi - \rho^\pi_\varphi \mathcal{E}^\pi_\varphi > 0$, where $\hat{\rho}^\pi_\varphi > 0$ denotes the attractiveness of option π if its cost ($\mathcal{E}^\pi_\varphi$) is zero, and $\rho^\pi_\varphi \geq 0$ captures the cost elasticity of option π for driver type φ. Parameters ρ^π_φ and $\hat{\rho}^\pi_\varphi$ can be estimated for each driver type φ offline. They respectively capture the financial and environmental considerations of the EV drivers. Parameter $\hat{\rho}^\pi_\varphi$ also captures other considerations, such as the discomfort due to signing up for the MH programs, communications with the utility, and sharing information in AMH. For example, we may have $\hat{\rho}^{\mathrm{AMH}}_{\varphi'} > \hat{\rho}^{\mathrm{PMH}}_{\varphi'}$ because driver type φ' may value giving full control of their charging to the utility for the environmental benefits, whereas for another driver type φ'', we may have $\hat{\rho}^{\mathrm{AMH}}_{\varphi''} < \hat{\rho}^{\mathrm{PMH}}_{\varphi''}$ due to information privacy. Last, a type-φ driver chooses option $\pi \in \Pi_\varphi$ with probability $\mathcal{P}^\pi_\varphi \triangleq \frac{\mathcal{A}^\pi_\varphi}{\sum_{\pi' \in \Pi_\varphi} \mathcal{A}^{\pi'}_\varphi}$.[1]

3.4. Program-Design Model (PD)

The utility's expected cost due to the bonus payments to the AMH participants plus the lost revenue because of the PMH participants' charging during low-rate episodes is

$$\mathcal{H}(\beta, \mu^{\mathrm{L}}) \triangleq \beta \sum_{\varphi:\ \mathrm{AMH} \in \Pi_\varphi} \mathcal{P}^{\mathrm{AMH}}_\varphi \mathcal{K}_\varphi + (\bar{\mu} - \mu^{\mathrm{L}}) \sum_{\varphi:\ \mathrm{PMH} \in \Pi_\varphi} \mathcal{P}^{\mathrm{PMH}}_\varphi \mathcal{K}_\varphi \mathcal{Q}_\varphi.$$

In return, the participants help the utility reduce its electricity cost. Let $\mathcal{F}(\beta, \mu^{\mathrm{L}})$ denote the utility's total expected cost of purchasing/producing electricity to satisfy its demand over the season. LM in §3.5 could be used to estimate $\mathcal{F}(\beta, \mu^{\mathrm{L}})$. The utility's PD is

$$(\mathrm{PD}): \quad \min_{(\beta, \mu^{\mathrm{L}})} \quad \mathcal{H}(\beta, \mu^{\mathrm{L}}) + \mathcal{F}(\beta, \mu^{\mathrm{L}}).$$

The utility designs its MH programs by constructing and solving PD as follows. First, it learns the following input data for each driver: total load requirement $\mathcal{Q}_j$, flexibility $\mathcal{I}_j$, parameters of the attractiveness functions $\hat{\rho}^\pi_j$ and ρ^π_j for each option π, and the compatible starting times $\mathcal{T}_{j,d}$. Second, it partitions the drivers into several types $\varphi \in \Phi$, such that the members of each type are homogeneous in the following conditions: (H1) $\mathcal{Q}_j$ and $\mathcal{I}_j$; (H2) $\hat{\rho}^\pi_j$ and ρ^π_j, for all π; and (H3) $\mathcal{T}_{j,d}$, for all d. Third, it defines $\Pi_\varphi \triangleq \{\mathrm{NMH}\}$, adds AMH to Π_φ if $\mathcal{I}_\varphi \geq \mathcal{I}^\circ$, and adds PMH to Π_φ if $\mathcal{T}_{\varphi,d} \neq \{\}$ for all d. Fourth, it fixes its decision vector to some practical values $(\check{\beta}, \check{\mu}^{\mathrm{L}})$ and performs the following steps:

[1] One may assume a type-φ driver chooses option $\bar{\pi}_\varphi \triangleq \arg\max_{\pi \in \Pi_\varphi} \mathcal{A}^\pi_\varphi$ with probability 1. Our methodology readily extends. We prefer probabilistic choices because (i) they effectively reflect the real-world behavior of the EV drivers and (ii) the optimal solution of PD is robust with respect to the inputs of the attractiveness functions.

- (i) compute costs $\mathcal{E}^{\pi}_{\varphi}$, attractiveness $\mathcal{A}^{\pi}_{\varphi}$, and choice probabilities $\mathcal{P}^{\pi}_{\varphi}$, for all $\varphi,\ \pi \in \Pi_{\varphi}$; and
- (ii) estimate the utility's total cost $\mathcal{H}(\check{\beta}, \check{\mu}^{\mathrm{L}}) + \mathcal{F}(\check{\beta}, \check{\mu}^{\mathrm{L}})$.

The utility optimizes its program-design decisions by repeating steps (i)-(ii) for all practical values of $(\beta, \mu^{\mathrm{L}})$. Online Appendix A provides an illustrative example for PD. After solving PD, the utility invites the EV drivers to select an option from the offer sets. After the subscription phase concludes and the season starts, the subscribers of each program are known. Next, we present LM to jointly execute AMH and PMH within the season.

3.5. Load-Management Model (LM)

We formulate LM to jointly execute AMH and PMH within the season. We construct and solve LM on a rolling-horizon basis, at the beginning of each hour, throughout the season.

We formulate LM at the beginning of a given hour as follows. With some abuse of notation, let $t \in \{1, \ldots, T\}$ denote the *remaining* hours in the season, including the current hour, and let $d \in \{1, \ldots, \mathcal{D}\}$ denote the *remaining* days in the season, including the current day. We incorporate the stochasticity in our model by considering $s \in \{1, \ldots, S\}$ scenarios, where scenario s occurs with probability $p_s > 0$. We assume all aspects of each scenario are deterministic — our extension in Online Appendix G.1 incorporates some stochasticity beyond scenarios. The EV drivers who arrive at the beginning of period 1 (current hour) or earlier are the same in different scenarios, whereas the drivers who arrive at the beginning of the second period or later may vary across scenarios. After each time we solve LM, we implement the supply decisions that correspond to the first period. Then, we proceed to the next period and update the scenarios based on the most recent information. Online Appendix B provides an illustrative example for scenarios.

NMH load. Let j be an NMH driver under scenario s. In a charging event, she plugs in at the beginning of some hour t' and starts charging with her maximum speed q^U_j to satisfy her requirement γ'. The utility supplies the following amount to her in period t, under scenario s:

$$q_{s,j,t} = \begin{cases} q^U_j & \text{if } t' \le t < t' + \lfloor \gamma' \rfloor, \\ \gamma' - \lfloor \gamma' \rfloor & \text{if } t = t' + \lfloor \gamma' \rfloor, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Thus, one could pre-compute the supplied load to all NMH drivers. Let $\lambda^{\mathrm{NMH}}_{s,g,t}$ denote the total load supplied to all NMH drivers in period t, in service region g, and under scenario s.

PMH load. The utility must decide the starting times of the low-rate episodes and notify the PMH participants at the beginning of a pre-specified hour, denoted by $\check{t}_d$, as stipulated in the contract (e.g., $\check{t}_d = 10$ a.m. for all d). To simplify our model, we assume all potential low-rate episodes on day d start at or after the beginning of period $\check{t}_d$ and end at or before the end of period $\check{t}_{d+1} - 1$. Thus, the charging of the PMH participants in periods $\{\check{t}_d, \ldots, \check{t}_{d+1} - 1\}$ is influenced by

the decisions made at the beginning of period $\check{t}_d$. Let binary variable $\vartheta_{s,\tau,\varphi,d}$ be 1 if starting time $\tau \in \mathcal{T}_{\varphi,d}$ is assigned to the PMH participants of type φ on day d, under scenario s. Each time we solve LM, it optimizes $\vartheta_{s,\tau,\varphi,d}$ for all days $d \in \{1,\ldots,\mathcal{D}\}$. Recall $d=1$ denotes the current day (today). After solving LM at the beginning of hour $\check{t}_1$, the utility communicates today's starting times to the PMH participants. We fix the value of $\vartheta_{s,\tau,\varphi,1}$ when we solve our model at the beginning of $\check{t}_1+1$, $\check{t}_1+2$, and so forth.

Recall from Property CST that a PMH participant arrives home before her low-rate episode starts. She starts charging at her full charging speed after her low-rate episode starts. Thus, the charging behavior of the PMH participants is similar to the NMH drivers, except PMH participants start charging after their low-rate episode starts, despite the NMH drivers who start charging after they arrive home. Therefore, given that the low-rate episode of participant j starts at $\tau \in \mathcal{T}_{\varphi,d}$, one could pre-compute the supplied load to her using equation (1). For driver type φ, where PMH $\in \Pi_\varphi$, let $\lambda^{\text{PMH}}_{s,g,t,\varphi,\tau}$ denote the total load supplied to all PMH participants of type φ in period t, in service region g, and under scenario s, assuming the assigned starting time is $\tau \in \mathcal{T}_{\varphi,d}$. Thus, the total load supplied to all PMH participants in period t, in region g, and under scenario s, is

$$\sum_{\varphi:\ \text{PMH}\in\Pi_\varphi} \left(\sum_{\tau\in\mathcal{T}_{\varphi,d(t)}} \lambda^{\text{PMH}}_{s,g,t,\varphi,\tau}\vartheta_{s,\tau,\varphi,d(t)} \right),$$

where $d(t)$ denotes the day on (the morning of) which the utility decides the starting times of the low-rate episodes that influence period t — for example, $d(t)=d'$, for all $t \in \{\check{t}_{d'},\ldots,\check{t}_{d'+1}-1\}$.

AMH load. An AMH participant has $\mathcal{D}$ remaining charging events. Because these charging events do not overlap, instead of each AMH participant, one could equivalently assume $\mathcal{D}$ participants exist, each of whom has one charging event in the remainder of the season. Let $\mathbb{J}_s$ denote the set of such AMH participants who have one charging event, under scenario s. To simplify our notation, for $j \in \mathbb{J}_s$, let a_j, ℓ_j, Q_j, and γ_j respectively denote her plug-in time, charging window, requested load, and normalized load requirement. Observe we dropped index d because each $j \in \mathbb{J}_s$ has one charging event. Furthermore, under scenario s, these inputs (a_j, ℓ_j, Q_j, and γ_j) are deterministic. As we previously stated, the AMH participants who arrive at the beginning of period 1 (current hour) or earlier are the same across different scenarios; that is, all $\mathbb{J}_s$'s include such participants. The AMH participants who arrive at the beginning of the second period or later may vary across scenarios. By solving our model, we determine the amount of load to be supplied to each AMH participant in each period and under each scenario. Let decision variable $q_{s,j,t}$ denote the amount of load to be supplied to participant j in period t, under scenario s. We include a constraint in LM that enforces the same first-period decisions across different scenarios. After each time we solve LM, we implement the supply decisions that correspond to the first (current) period; then,

we proceed to the next period and update the scenarios based on the most recent information. We formulate LM as follows:

$$
\begin{aligned}
(\mathrm{LM}): \quad \min \ & \sum_{s} p_s \sum_{t} f_t(y_{s,t}) && (2)\\
\text{s.t.} \ & \sum_{t=a_j}^{a_j+\ell_j-1} q_{s,j,t} = Q_j, \quad \forall s,\ j \in \mathbb{J}_s, && (3)\\
& 0 \le q_{s,j,t} \le q_j^U, \quad \forall s,\ t,\ j \in \mathbb{J}_s, && (4)\\
& q_{s,j,1} = q_{s',j,1}, \quad \forall s \ne s',\ j \in \mathbb{J}_s \cap \mathbb{J}_{s'}, && (5)\\
& x_{s,g,t} = \lambda_{s,g,t} + \sum_{j \in \mathbb{J}_s:\ g_j = g} q_{s,j,t}, \quad \forall s,\ g,\ t, && (6)\\
& \lambda_{s,g,t} = \lambda_{s,g,t}^{\mathrm{EXI}} + \lambda_{s,g,t}^{\mathrm{NMH}} + \sum_{\varphi:\ \mathrm{PMH} \in \Pi_\varphi} \left(\sum_{\tau \in \mathcal{T}_{\varphi,d(t)}} \lambda_{s,g,t,\varphi,\tau}^{\mathrm{PMH}} \vartheta_{s,\tau,\varphi,d(t)} \right), \quad \forall s,\ g,\ t, && (7)\\
& \sum_{\tau \in \mathcal{T}_{\varphi,d}} \vartheta_{s,\tau,\varphi,d} = 1, \quad \forall s,\ \varphi:\ \mathrm{PMH} \in \Pi_\varphi,\ d, && (8)\\
& \vartheta_{s,\tau,\varphi,1} = \vartheta_{s',\tau,\varphi,1}, \quad \forall s \ne s',\ \tau \in \mathcal{T}_{\varphi,1},\ \varphi:\ \mathrm{PMH} \in \Pi_\varphi, && (9)\\
& \vartheta_{s,\tau,\varphi,d} \in \{0,1\}, \quad \forall s,\ \tau \in \mathcal{T}_{\varphi,d},\ \varphi:\ \mathrm{PMH} \in \Pi_\varphi,\ d, && (10)\\
& \sum_{g \in i} x_{s,g,t} \le h_i, \quad \forall s,\ t,\ i, && (11)\\
& y_{s,t} = \sum_{g} x_{s,g,t}, \quad \forall s,\ t. && (12)
\end{aligned}
$$

In the objective function (2), $y_{s,t}$ is the total consumption in period t, under scenario s, and hence, the objective function minimizes the expected energy cost in the rest of the season. Constraint (3) ensures the load required by each AMH participant is supplied within her charging window. The supplied load to a participant in any period must satisfy her charging speed limit q_j^U, which is ensured by constraint (4). Constraint (5) enforces the same first-period decisions across different scenarios. To model the grid constraints, we let $x_{s,g,t}$ denote the total load supplied to service region g in period t and under scenario s. The value of $x_{s,g,t}$ is determined through constraint (6). Quantity $\lambda_{s,g,t}$ is the total load excluding the EV load of the AMH participants, also referred to as the *non-AMH load.* Constraints (7)-(10) compute the non-AMH load $\lambda_{s,g,t}$, where $\lambda_{s,g,t}^{\mathrm{EXI}} \ge 0$ is the total non-EV load. Grid constraints are imposed on $x_{s,g,t}$, as we model in constraint (11); under each scenario s, the total load for the service regions that are served by substation i (i.e., $g \in i$) cannot exceed the limit of the substation h_i. Last, constraint (12) computes the values of $y_{s,t}$'s.

In short, we formulate LM to determine the charging supply to each AMH participant and the starting times of the low-rate episodes, while minimizing the total energy cost and respecting the grid-stability constraints. LM has three sources of difficulty. First, the number of decision

variables ($q_{s,j,t}$) and constraints (e.g., constraints (3) and (4)) grow linearly in the number of AMH participants. The size of this model (i.e., its variables and constraints) is $\mathcal{O}(STJ)$, where $J \triangleq \max_s |\mathbb{J}_s|$, and reasonably assuming the number of binary variables ($\vartheta_{s,\tau,\varphi,d}$) is small. Real-size instances of LM can have extremely large numbers of participants, making it impractical in its current form; hence, designing an approximation model such that its size does not grow in the number of participants is necessary. We address this challenge by presenting an *aggregation* scheme to form groups of participants. Second, T can be very large. The size of LM grows in T; hence, as an approximation, we propose a *truncation* procedure to solve LM using a short horizon. Investigating how this approximation affects the solution, both theoretically and numerically, is indispensable. Third, the objective function of LM is nonlinear. Nonlinear problems are extremely more difficult to solve than linear ones. To address the nonlinearity, we present a *linearization* of the objective function, which transforms our problem to an MILP. In the remainder, we present our approximations for LM.

4. Approximations for LM

In this section, we present a solution method that consists of three approximations to address the three challenges of LM, which we discussed at the end of §3. We briefly introduce these approximations below and defer their details to §4.1-§4.3.

Aggregation. In §4.1, we aggregate the AMH participants who belong to the same service region and have the same arrival time and charging window into groups, based on their normalized load requirement (γ_j). We also introduce a parameter ϵ that indicates which participants should be in the same group; the members of a group have γ-values within ϵ of each other. The value of ϵ controls the number of groups, the size of groups, and the approximation quality. Our aggregation reduces the size of LM from $\mathcal{O}(STJ)$ to $\mathcal{O}(\frac{1}{\epsilon}SGT^4)$, transforming it to a practically sound model because its size no longer grows in the number of participants. Let ALM denote the aggregate model. We show how to disaggregate the solution of ALM and create a feasible solution for LM. Last, we present a worst-case error bound for aggregation and discuss how it is influenced by ϵ.

Truncation. Recall T can be large. Thus, we truncate the long horizon to $\breve{T}$ periods, where $\breve{T} \leq T$. Consider an AMH participant with a charging window of t' periods within the truncated horizon and t'' periods after the truncated horizon ends. To accommodate this participant in our truncation, we assume we need to supply only the $\left(\frac{t'}{t'+t''}\right)$-fraction of her requested load within the truncated horizon. Interestingly, we show if the horizon is truncated to one period (i.e., $\breve{T} = 1$), the solution has a nice structure in which each participant's requirement is uniformly supplied over her charging window. We present a theoretical analysis of our truncation and show it affects the solution by (i) restricting the solution and (ii) discarding forecast information. Inconsistent with

intuition, we show the error of truncation does not necessarily reduce as $\breve{T}$ increases. In a special case, we show the relative error of truncation is at most 86.6%, and we establish the tightness of this bound. Our numerical experiments show $\breve{T} \geq 15$ produces near-optimal solutions. Last, we present a special case of our problem, which we obtain by imposing two assumptions that are motivated by practical cases, and show a reasonably small T^* exists such that truncating the horizon to $\breve{T} \geq T^*$ produces an optimal solution.

Linearization. In §4.3, we linearize the objective function by approximately assuming the cost function $f_t(\cdot)$ is constant over intervals of length δ. We formulate an approximation model, denoted by LLM, that is an MILP. We present a worst-case error bound for this approximation, which shows how the error is related to the value of δ.

The above three approximations transform our large-scale nonlinear program to an MILP, which has a practically reasonable size. Our numerical experiments indicate this MILP problem is solvable in a short time using commercial solvers, such as Cplex. We characterize the error of the linear-programming relaxation of this MILP problem and show it is small, supporting our conclusion that our approximation is suitable for large-scale applications. In §4.1-§4.3, we present the details of our aggregation, truncation, and linearization.

Notations for error characterization. Let $q^{\diamond} \triangleq \max_j q_j^U$ be the maximum home-charging speed, $J \triangleq \max_s |\mathbb{J}_s|$, and let $\check{J}$ be a practical upper bound on J. Define

$$\lambda_{s,t}^{\mathrm{L}} \triangleq \sum_g \left(\lambda_{s,g,t}^{\mathrm{EXI}} + \lambda_{s,g,t}^{\mathrm{NMH}} + \sum_{\varphi:\ \mathrm{PMH} \in \Pi_\varphi} \min_{\tau \in \mathcal{T}_{\varphi,d(t)}} \lambda_{s,g,t,\varphi,\tau}^{\mathrm{PMH}} \right),$$

$$\lambda_{s,t}^{\mathrm{U}} \triangleq \sum_g \left(\lambda_{s,g,t}^{\mathrm{EXI}} + \lambda_{s,g,t}^{\mathrm{NMH}} + \sum_{\varphi:\ \mathrm{PMH} \in \Pi_\varphi} \max_{\tau \in \mathcal{T}_{\varphi,d(t)}} \lambda_{s,g,t,\varphi,\tau}^{\mathrm{PMH}} \right),$$

$$\Theta(\delta) \triangleq \max_{s,t} \left\{ \max_{\lambda_{s,t}^{\mathrm{L}} \leq y \leq \lambda_{s,t}^{\mathrm{U}} + \check{J} q^{\diamond}} f_t(y+\delta) - f_t(y) \right\},$$

$$\Psi(\varepsilon) \ \triangleq \max_{\sum_t \delta_t \leq \varepsilon,\ \delta_t \geq 0,\ \forall t} \sum_t \Theta(\delta_t),$$

for some $\delta \geq 0$ and $\varepsilon \geq 0$. Quantities $\lambda_{s,t}^{\mathrm{L}}$ and $\lambda_{s,t}^{\mathrm{U}}$ are respectively lower and upper bounds on the non-AMH load, and $\Theta(\delta)$ (respectively, $\Psi(\varepsilon)$) is an upper bound on the increase in energy cost if an hour's (respectively, the horizon's) load increases by δ (respectively, ε). Let $\lambda^{\mathrm{L}} \triangleq \min_{s,t} \lambda_{s,t}^{\mathrm{L}}$ and $\lambda^{\mathrm{U}} \triangleq \check{J} q^{\diamond} + \max_{s,t} \lambda_{s,t}^{\mathrm{U}}$. Last, $[e]^+ \triangleq \max\{0, e\}$, for $e \in \mathbb{R}$.

4.1. Aggregation

We present an aggregation-disaggregation (AD) procedure that (i) constructs and solves an approximation of LM, denoted by ALM, and (ii) creates a feasible solution for LM using the solution of ALM. Steps (i) and (ii) are called *aggregation* and *disaggregation*, respectively.

Aggregation. For each scenario s, we partition the AMH participants who belong to the same service region g and have the same arrival time and charging window into multiple groups, based on the closeness of their normalized load requirements (γ_j). We offer the following intuitive interpretation. Consider two participants j' and j'' and assume $q^U_{j'} = 2q^U_{j''}$ and $Q_{j'} = 2Q_{j''}$, meaning participant j' has twice the charging speed limit and requests twice the load requested by participant j''. In this example, we may assume participant j' consists of two sub-participants who are similar to j'', and hence, we can reasonably call participants j' and j'' "similar" and assign them to the same group.

To simplify our presentation, we define a combined index as $\kappa \triangleq (s, g, t, \ell)$; that is, each κ refers to a specific combination of scenario s, service region g, arrival time t, and charging window ℓ. Moreover, with some abuse of notation, we denote the s, g, t, and ℓ that correspond to κ as $s(\kappa)$, $g(\kappa)$, $t(\kappa)$, and $\ell(\kappa)$. With some abuse of notation, let $\mathbb{J}_\kappa$ denote the set of participants in service region $g(\kappa)$ who plug in their EVs at the beginning of period $t(\kappa)$ and require a load amount to be supplied during the next $\ell(\kappa)$ periods, under scenario $s(\kappa)$.

We partition $\mathbb{J}_\kappa$ into groups based on the closeness of their γ-values. We introduce a parameter $\epsilon > 0$ that controls the closeness in the γ-values of the participants who fall into the same group. In other words, we create groups such that the difference between the γ-values of the participants in the same group is at most ϵ. The value of ϵ controls a trade-off between the number of groups (size of the approximation model) and the approximation quality. Our detailed procedure for constructing these groups is given below. For each nonempty $\mathbb{J}_\kappa$, we perform the following three steps.

1. Determine the smallest γ-value, that is, $\gamma_L \triangleq \min_{j\in\mathbb{J}_\kappa} \gamma_j$. Create a group consisting of all participants whose γ-values belong to the closed interval $[\gamma_L, \gamma_L + \epsilon]$. Denote this group using index m, and let $\mathbb{M}_m$ be the set of all participants who belong to this group.

2. Define $\gamma_U \triangleq \max_{j\in\mathbb{M}_m} \gamma_j$. For participant $j \in \mathbb{M}_m$, consider a synthetic load requirement $\hat{Q}_j \triangleq \gamma_U q^U_j$ (instead of her actual load requirement Q_j). Note we increase the required load by each participant j by at most ϵq^U_j. We assume ϵ is sufficiently small that the model remains feasible after such increases. Last, we consider group m a "super-participant" with a load requirement of $\hat{Q}_m = \sum_{j\in\mathbb{M}_m} \hat{Q}_j$ and a charging speed of $q^U_m = \sum_{j\in\mathbb{M}_m} q^U_j$. In simple terms, a group's load requirement and charging speed are equal to sums of those of its members.

3. Let $\mathbb{J}_\kappa \triangleq \mathbb{J}_\kappa \backslash \mathbb{M}_m$, and repeat steps 1-3 until $\mathbb{J}_\kappa$ becomes empty.

After forming groups, instead of individual participants, we have groups that act as super-participants. Each group has a synthetic load requirement and a synthetic charging speed limit. We construct ALM using these groups; in other words, ALM is similar to LM except that instead of individual participants, its inputs are the groups.

Size of ALM. We partition each $\mathbb{J}_\kappa$ into at most $\frac{1}{\epsilon}T$ groups, because γ-values are between 0 and T. In each scenario, the number of $\mathbb{J}_\kappa$'s is at most GT^2; hence, in each scenario, we create at most $\frac{1}{\epsilon}GT^3$ groups. Recall the size of LM is $\mathcal{O}(STJ)$. Using $\frac{1}{\epsilon}GT^3$ groups instead of J participants, the size of ALM is $\mathcal{O}(\frac{1}{\epsilon}SGT^4)$. Note the size of this approximation model is not affected by the number of participants J; that is, regardless of the number of participants, the size of ALM remains practically reasonable.

Disaggregation. So far, we have described our procedure for creating ALM. We solve this model on a rolling basis, and each time, we implement the decisions corresponding to the first period. Let $q_{s,m,1}$ denote the amount of load to be supplied to group m in period 1, based on the solution of ALM. Recall that because of constraint (5), the values of $q_{s,m,1}$'s are equal across different scenarios. ALM supplies a cumulative amount $q_{s,m,1}$ to group m, but it does not specify how to dispense this cumulative load to the members of this group. We allocate such cumulative loads to the individual participants as follows. For each participant j in group m, we supply the following load amount in the first period:

$$q_{s,j,1} \triangleq \min\left\{Q_j, \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,1}\right\}.$$

In simple terms, we allocate the group's total load to its members proportional to their synthetic load requirements. We are also careful not to supply more than a participant's actual load requirement; that is, if the proportional allocation is more than her requirement Q_j, we supply Q_j.

Analysis of AD procedure. To show the effectiveness and appropriateness of our AD procedure, we need to investigate two properties. First, we need to show AD always creates a feasible solution — for example, it does not under-supply early in a participant's charging window, resulting in an unsatisfiable remaining load closer to the end of her charging window. Second, we need to theoretically and numerically investigate its solution quality. We study these properties below.

THEOREM 1. *(a)* (**Feasibility of AD**). *Our AD procedure creates a feasible solution for LM.*

(b) (**Aggregation Error**). *If $h_i \geq \lambda^U$, for all i, the absolute error of our AD procedure is at most $\Psi(J\epsilon q^\diamond)$.*

All proofs are available in Online Appendix C. Part (a) establishes that AD always produces a feasible solution for LM. The proof of part (a) entails showing the disaggregated supply quantities ($q_{s,j,1}$'s) satisfy the constraints of LM. Part (b) relates the error of AD to the value of ϵ and J. Intuitively, because we increase each participant's load by at most $\epsilon q^\diamond$, the total EV load is increased by at most $J\epsilon q^\diamond$ over the season, which creates an error of at most $\Psi(J\epsilon q^\diamond)$. The assumption $h_i \geq \lambda^{\mathrm{U}}$, for all i, is made to simplify our proof; under this assumption, with small changes in the supply quantities, the solution of LM remains feasible. Importantly, part (b) outlines the trade-off,

controlled by the value of ϵ, between the solution quality of AD and the computational performance in solving ALM. One could reduce the error by selecting ϵ small; however, this advantage comes at the price of increasing the size of the approximation model ALM, recalling the number of variables and constraints of ALM are both in the order of $\mathcal{O}(\frac{1}{\epsilon}SGT^4)$. Our numerical study demonstrates AD produces near-optimal solutions. Also note our aggregation is different than the existing *clustering* approaches (see, e.g., Mettu and Plaxton 2004, Feldman et al. 2007), in which the *number of clusters* is fixed and a partition of some given points into clusters must be found, without controlling the *maximum distance* within each cluster. By contrast, we impose the *maximum distance* of ϵ within each cluster and let the algorithm decide the *number of clusters*. The advantage of our aggregation is that it allows us to track the total changes in the requested loads and bound the error in Theorem 1(b).

4.2. Truncation: Truncating the Long Horizon

We propose truncating the long horizon to $\breve{T}$ periods, where $\breve{T} \leq T$. Intuitively, we consider a truncated horizon (TH) that consists of periods $\{1, \ldots, \breve{T}\}$, and discard all periods succeeding $\breve{T}$. Truncation affects an AMH participant if her charging window exceeds TH. For example, assume participant j arrives t' hours before TH ends, and her charging window is $t' + t''$ hours; that is, her charging window consist of t' hours inside TH and t'' hours outside of TH. In our truncation, we accommodate this participant by assuming she arrives t' hours before TH ends and requests a $\left(\frac{t'}{t'+t''}\right)$-fraction of her actual load requirement to be supplied in the next t' hours. For example, if a participant plugs in her EV at the beginning of hour $\breve{T}-1$ and requires 20 KWh to be supplied in the next eight hours, we assume this participant requires 5 KWh to be supplied in the next two periods $\{\breve{T}-1, \breve{T}\}$. Because we solve our problem on a rolling basis and $\breve{T}$ is reasonably large, the impact of these manipulations in the latter part of TH is negligible, which our theoretical and numerical analyses verify. Truncation creates an approximate problem, denoted by TLM($\breve{T}$). We assume TLM($\breve{T}$) is feasible for all $\breve{T} \in \{1, \ldots, T\}$.

Analysis of truncation. Truncation affects the solution in the following two ways: (i) *restricting the solution*, recalling it enforces the supply of a $\left(\frac{t'}{t'+t''}\right)$-fraction of a participant's load within the truncated horizon; and (ii) *discarding forecast information*, because TLM($\breve{T}$) does not consider beyond $\breve{T}$ periods. Therefore, truncation may introduce an error. Inconsistent with intuition, we show the error does not necessarily reduce as $\breve{T}$ increases, if the forecast is highly inaccurate. Because the forecast of the near future (e.g., the next 24 hours) is highly accurate in the context of this paper, we believe the solution quality should improve in $\breve{T}$ in a practical setting, which our numerical experiments confirm. In other words, we believe that in a practical setting, truncating the horizon to $\breve{T} = 1$ produces the worst solution. Interestingly, if $\breve{T} = 1$, we show the utility

uniformly supplies each participant's requirement over her charging window. In a special case, we show truncation increases the utility's energy cost by at most 86.6%. Our numerical experiments show $\breve{T} \geq 15$ produces near-optimal solutions. The following theorem summarizes our theoretical analyses of truncation.

THEOREM 2. *(a)* (**Uniform Supply**). *TLM(1) uniformly supplies an AMH participant's load Q_j over her charging window ℓ_j; that is, she receives $\frac{Q_j}{\ell_j}$ in each period $t \in \{a_j, \ldots, a_j + \ell_j - 1\}$.*

(b) (**Violation of Monotonicity in $\breve{T}$**). *The utility's total energy cost may increase in $\breve{T}$.*

(c) (**Relaxation of the Feasible Set**). *Let TLM($\breve{T}$) and TLM($\breve{T}'$), for some $\breve{T}' \geq \breve{T} \geq 1$, be constructed at the same period. If $h_i \geq \lambda^U$, for all i, an optimal solution to TLM($\breve{T}$) is a feasible solution to TLM($\breve{T}'$).*

(d) (**Truncation Error**). *If $f_t(y) = cy^3$, for all t and $y \geq 0$, and for some $c > 0$, $S = G = 1$, and only one EV driver exists who participates in AMH and has one charging event with $\ell_j = 2$, truncating the horizon to $\breve{T} = 1$ increases the utility's energy cost by at most 86.6%. This worst-case error bound is tight.*

In part (a), by setting $\breve{T} = 1$, the utility supplies each participant's load uniformly over her charging window. This solution is nice and consistent with the main purpose of AMH programs, which is to distribute the EV loads across the participants' charging windows; however, it is done myopically and without requiring any optimization. If $\breve{T} \geq 2$, this restriction is relaxed in the following two ways. First, TLM($\breve{T}$) freely assigns the load to any periods within TH. For example, it may assign no supply to a participant in the first period, implying the supply decision to this participant is postponed until the next period. Second, if $\breve{T} \geq 2$, in the solution of TLM($\breve{T}$), we discard the supply decisions beyond the current period. In essence, we reevaluate our supply decisions on an hourly basis. TLM(1) does not have this advantage. Part (b) investigates whether the utility benefits from increasing $\breve{T}$. Using larger $\breve{T}$ implies discarding less information, and hence, one would expect the utility's energy cost to be non-increasing in $\breve{T}$. Interestingly, and inconsistent with intuition, we show in part (b) that if the forecast has low accuracy beyond, say, $\breve{T}'$ periods, increasing $\breve{T}$ beyond $\breve{T}'$ may worsen the solution. By contrast, part (c) supports the monotonicity in $\breve{T}$. Part (c) implies larger $\breve{T}$ provides more flexibility to achieve better solutions. In short, parts (b) and (c) respectively refute and support monotonicity in $\breve{T}$. In a real-world setting, because the forecast of the near future is highly accurate, instances such as that in part (b) are less likely to happen. Motivated by these arguments, in part (d), we conduct our worst-case analysis of truncation for the case of $\breve{T} = 1$. In this special case, we show truncation can increase the utility's energy cost by at most 86.6%. Other assumptions of part (d) are mild and reasonable. Our numerical experiments show truncation with $\breve{T} \geq 15$ produces near-optimal solutions.

An optimal special case for truncation. We provide a theoretical justification of why truncation is likely to produce near-optimal solutions in the context of this paper. We create a special case by imposing some assumptions that are motivated by our observations from practice. We show a reasonably small T^* exists such that truncating the horizon to $\breve{T} \geq T^*$ periods is optimal.

THEOREM 3 **(Truncation Optimality)**. *Assume $|\mathcal{T}_{\varphi,d}| = 1$, for all φ, d, $f_t(y) = f_{t'}(y)$ for all t and t', $f_t(y)$ is strictly convex over $y \geq 0$, and the horizon $\{1, \dots, T\}$ consists of non-overlapping intervals in the form of $\{\dot{t}, \dots, \ddot{t}\}$ that contain AMH participants' charging windows, and satisfy $\ddot{t} - \dot{t} + 1 \leq T_{NOI}$, for some $T_{NOI} \in \{1, 2, \dots\}$. If $\sum_{g \in i'} (\lambda_{s,g,t} - \lambda_{s,g,t+1}) \geq (J+2)q^{\diamond}$, for all s, $i' \in \{i\} \cup \{1, \dots, G\}$, and $t \in \{\dot{t}, \dots, \ddot{t} - 1\}$, truncating the horizon to $\breve{T} \geq T^*$ is optimal, where*

$$T^* \triangleq \min \left\{ T_{NOI}, \max_j \left\{ \lceil \gamma_j \rceil, \left\lceil \frac{\lceil \gamma_j \rceil + 1}{\lceil \gamma_j \rceil + 1 - \gamma_j} \right\rceil \right\} \right\}.$$

The assumption that the horizon consists of the non-overlapping intervals is motivated by the observation that most drivers arrive home in the evening, plug in their EVs, and leave for work the next morning. In other words, most charging windows are between afternoon and the morning of the next day. Thus, the horizon seems to consist of non-overlapping intervals, which contain charging windows. Furthermore, we find consumption usually sharply reduces during the time interval between evening and the next pre-morning (i.e., before people wake up). This pattern is common both in summer and winter. Motivated by this observation, we assume a sharp reduction in consumption over the non-overlapping intervals. These two assumptions enable us to characterize T^*. In the definition of T^*, the first term T_{NOI} is a consequence of the non-overlapping intervals. Because the intervals do not overlap, we show in the proof that only the current interval is relevant; hence, one needs to consider at most T_{NOI} periods. The second term in the definition of T^* is due to imposing the assumption of sharp reduction in consumption. In the proof, we note that under this assumption, supplying each participant's load requirement in the latest hours of her charging window is optimal. Therefore, we determine the smallest horizon for which the load supply is not assigned to the first period unless the first period is sufficiently close to the end of her charging window. This analysis leads to the second term in the definition of T^*. For example, if $T_{\text{NOI}} = 24$ and $\gamma_j \leq 4$, for all j, then $T^* \leq 5$, meaning the horizon could be truncated to five hours with no loss in optimality. In short, Theorem 3's assumptions are motivated by practical cases and the proposed T^* is reasonably small. In our numerical experiments, the assumptions of Theorem 3 do not hold; however, we find $\breve{T} \geq 15$ produces near-optimal solutions.

4.3. Linearization: Linearizing the Objective Function

We partition the domain of $f_t(\cdot)$, which is the total hourly load, into intervals of length δ. The z^{th} interval is $((z-1)\delta, z\delta]$, for $z \in \{1, 2, \dots\}$. Interval 0 is a singleton consisting of 0. For each $y_{s,t}$,

we define an integer variable $y^{\text{int}}_{s,t}$ that indicates the interval to which $y_{s,t}$ belongs. For example, if $y^{\text{int}}_{s,t}=1$, then $y_{s,t}$ belongs to the first interval; that is, $y_{s,t}\in(0,\delta]$; if $y^{\text{int}}_{s,t}=2$, then $y_{s,t}\in(\delta,2\delta]$; and so forth. To compute the objective value of a solution, we round up each $y_{s,t}$ to the largest value in the corresponding interval; that is, we approximate $f_t(y_{s,t})$ by $f_t(\delta y^{\text{int}}_{s,t})$. Because we increase $y_{s,t}$'s, our approximation provides an upper bound on the optimal value of LM. The following theorem outlines our linearization and bounds its error.

THEOREM 4. *(a)* (**Linearization**). *Let $\delta>0$ be a given scalar, define $Z\triangleq\left\lceil\frac{1}{\delta}\lambda^U+1\right\rceil$, and let $\sigma_{t,z}\triangleq f_t(z\delta)-2f_t((z-1)\delta)+f_t((z-2)\delta)$, for all $t\in\{1,\dots,T\}$ and $z\in\{1,\dots,Z\}$. An optimal solution of the following MILP problem (LLM) is a feasible solution to LM, and an optimal value of LLM is an upper bound on the optimal value of LM:*

$$\textit{(LLM)}: \quad \min \ \sum_s p_s\sum_{t=1}^{T}\sum_{z=1}^{Z}\sigma_{t,z}\varepsilon_{s,t,z} \tag{13}$$

$$s.t. \ \ 0\le\varepsilon_{s,t,z}-\left(y^{int}_{s,t}-z+1\right)\le Z\left(1-\upsilon_{s,t,z}\right), \quad \forall s,\ t,\ z, \tag{14}$$

$$0\le\varepsilon_{s,t,z}\le Z\upsilon_{s,t,z},\ \upsilon_{s,t,z}\in\{0,1\}, \quad \forall s,\ t,\ z, \tag{15}$$

$$y_{s,t}\le\delta y^{int}_{s,t}\le y_{s,t}+\delta, \quad y^{int}_{s,t}\in\mathbb{Z}_+, \quad \forall s,\ t, \tag{16}$$

and constraints (3)-(12).

(b) (**Linearization Error**). *The error of linearization is at most $T\Theta(\delta)$.*

In Theorem 4(a), we introduce $\sigma_{t,z}$, which enables us to evaluate the value of $f_t(\cdot)$ on $0,\ \delta,\ 2\delta$, and so forth. Then, we are able to formulate a mixed-integer program, referred to as LLM, by introducing new variables $y^{\text{int}}_{s,t}$, $\upsilon_{s,t,z}$, and $\varepsilon_{s,t,z}$, as well as new constraints (14)-(16). Because the total hourly loads $y_{s,t}$ are increased, the optimal value of LLM is a bound for LM. This feature is used in part (b) to obtain an upper bound on the error of linearization. Intuitively, as we increase the load in each period by at most δ, we introduce at most $\Theta(\delta)$ error in each period; hence, the total error is at most $T\Theta(\delta)$. Observe the error increases in δ and T, as one would expect.

Our linearization is similar to the classical approach of approximating a nonlinear function with a piecewise linear function (see, e.g., Bertsimas and Tsitsiklis 1997), which partitions the domain into subsets and assumes the function is linear over each subset. We too partition the domain into intervals of length δ. However, we assume the function is constant over each interval; that is, we are not sensitive within δ of the hourly load. This assumption enables us to model our linear approximation by introducing ST integer variables and STZ binary variables to determine the index of the interval for the total hourly load. We also introduce STZ continuous variables and $4STZ+2ST$ constraints. Obviously, we obtain a better approximation by decreasing δ. However, as δ decreases, the number of variables and constraints grow, because $Z=\left\lceil\frac{1}{\delta}\lambda^{\text{U}}+1\right\rceil$. In our numerical

experiments in Online Appendix F, we change the value of δ from 0.01% to 10% of the maximum load, and LLM is solved quickly in all cases. We find the error is small if $\delta \leq 1\%$.

In short, we formulated an approximation model LLM, which is an MILP. MILPs are known to be difficult to solve; however, as we previously stated, our numerical experiments indicate LLM is quickly solvable using commercial solvers, such as Cplex. This observation motivated us to further investigate the properties of LLM. In the following theorem, we show LLM has a desirable formulation, because its linear-programming relaxation error is conveniently small.

THEOREM 5 **(Worst-Case Error of LP Relaxation)**. *If $|\mathcal{T}_{\varphi,d}| = 1$, for all φ, d, the error of the linear programming relaxation of LLM is at most*

$$\sum_{t=1}^{T}\sum_{z=1}^{Z}\left\{[\sigma_{t,z}]^{+} + \frac{1}{2}[-\sigma_{t,z}]^{+}\left(Z - \frac{1}{\delta}\lambda^{L} + z - 2\right)\right\}.$$

The above error bound includes a term for each combination of t and z, which is defined depending on the sign of $\sigma_{t,z}$. An interesting special case happens when the hourly cost functions are convex, implying $\sigma_{t,z} \geq 0$ for all t and z, in which case, the error bound becomes

$$\sum_{t=1}^{T}\sum_{z=1}^{Z}\sigma_{t,z} = \sum_{t=1}^{T} f_t(\delta Z) - f_t(\delta(Z-1)),$$

which is similar to the error of linearization $T\Theta(\delta)$. We conclude our approximation LLM is an effective model for finding near-optimal solutions to large instances of our problem in a reasonable amount of time. In the remainder, we present our numerical study.

5. Numerical Experiments

We present our numerical analysis of PD and managerial insights in §5.1 and §5.2, respectively. Robustness and generalizations are discussed in Online Appendix E. The details of our numerical study on LM are provided in Online Appendix F, with an overview of key findings summarized as follows: (i) aggregation error increases in ϵ and in the AMH participation level, (ii) truncation error significantly reduces when we increase $\breve{T}$ from 2 hours to 14 hours, and the error is negligible beyond 14 hours, (iii) linearization error is small when $\delta \leq 1\%$, (iv) the CPU time of our approximation model is small for large instances, and (v) $\epsilon = 0.5$, $\breve{T} = 15$, and $\delta = 1\%$ are appropriate choices, leading to fast computation times and near-optimal solutions.

5.1. Numerical Analyses of PD

We conduct a numerical study using CAISO data to demonstrate the application of PD.

Daily Consumption Profiles. We let the season consist of July-August 2022. Then, $\mathcal{D} = 62$. We construct a representative daily-consumption profile based on CAISO system demand in July-August 2022 and assume this daily profile repeats every day in the season. For this purpose, we

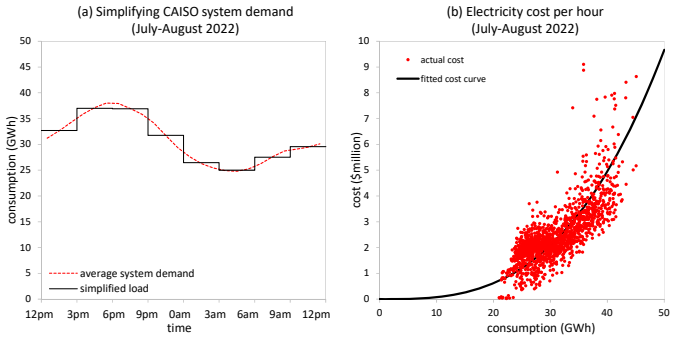

Figure 1 A simplified daily consumption and cost function using CAISO data (July-August 2022)

compute the average daily consumption during July-August 2022, which is plotted using a dotted line in Figure 1(a). We let a day consist of eight three-hour intervals: [0 a.m., 3 a.m.), [3 a.m., 6 a.m.), and so forth. During each interval, we let the consumption be constant and equal to the average demand during the interval. This procedure creates our simplified daily profile, which is shown using a solid line in Figure 1(a). We assume this profile occurs every day in the season.

Cost Function. We downloaded the CAISO real-time prices for the period of July-August 2022 from www.energyonline.com and computed hourly costs, which are shown in Figure 1(b). Each data point represents the total hourly cost of supplying electricity during some hour in July-August 2022. We have 62×24=1,488 data points. We fitted a cost curve in the form of $f(y) = cy^3$ and obtained $f(y) = 7.73 \times 10^{-8} y^3$ \$/MWh.

Charging Behavior of the EV Drivers. For simplicity, we assume only one driver type exists. All EV drivers arrive home at 6 p.m. and leave at 6 a.m. the next morning; that is, each EV driver has a 12-hour charging window starting at 6 p.m. Each driver has a charging speed limit of 3.7 KW and a normalized load requirement of three hours each day. An NMH driver charges her EV at full charging speed during the interval [6 p.m., 9 p.m.), which coincides with the existing evening peak. The utility needs to shift this EV consumption to off-peak times through offering its AMH and PMH programs. An AMH driver allows the utility to decide her charging time during her 12-hour charging window. In PMH, the utility decides the timing of the low-rate episode, and each PMH driver starts charging her EV at her full charging speed as soon as the low-rate episode starts. For simplicity, we assume the utility always starts the low-rate episode at midnight. We discuss customizing the low-rate episode in §5.2.

EV Drivers' Expected Costs. A driver's total load requirement during the season is $\mathcal{Q} = 62 \times 11.1 = 688.2$ KWh. Let the electricity rate during the season be $\bar{\mu} = 26$ ¢/KWh. An NMH driver's total expected cost is $\mathcal{E}^{\mathrm{NMH}} = 688.2 \times 0.26 =$\$178.93. Similarly, the total expected cost for an AMH driver and a PMH driver are $\mathcal{E}^{\mathrm{AMH}} =$ \$178.93 $- \beta$ and $\mathcal{E}^{\mathrm{PMH}} = 688.2\mu^{\mathrm{L}}$, respectively. For illustration, let the AMH bonus and PMH low rate be $\beta =$\$50 and μ^{L}=20 ¢/KWh (we endogenize these variables shortly). Thus, $\mathcal{E}^{\mathrm{AMH}} =$\$128.93 and $\mathcal{E}^{\mathrm{PMH}} =$\$137.64.

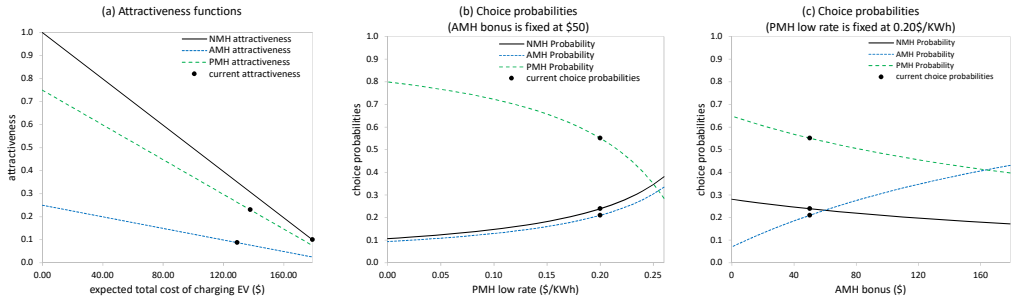

Figure 2 Attractiveness and choice probabilities

Attractiveness. Recall the utility estimates $\hat{\rho}^{\pi}$ and ρ^{π}, for each option π, offline. We normalize $\hat{\rho}^{\mathrm{NMH}}$ to 1 (see Figure 2(a)); that is, the attractiveness of NMH would be 1 if we assume its cost is zero. The attractiveness of AMH and PMH when their costs are zero may be smaller or larger than $\hat{\rho}^{\mathrm{NMH}}$; hence, we let $\hat{\rho}^{\mathrm{AMH}}, \hat{\rho}^{\mathrm{PMH}} \in [0, \infty)$. Let $\bar{\mathcal{A}} \in [0, 1]$ denote the attractiveness of NMH when its cost is $\mathcal{E}^{\mathrm{NMH}}$ =\$178.93; that is, $\bar{\mathcal{A}}$ could be interpreted as the attractiveness-reduction coefficient when the cost increases from \$0 to \$178.93. We let the attractiveness of AMH and PMH reduce using the same coefficient to $\bar{\mathcal{A}}\hat{\rho}^{\mathrm{AMH}}$ and $\bar{\mathcal{A}}\hat{\rho}^{\mathrm{PMH}}$ when their cost is \$178.93. In this example, quantities $\bar{\mathcal{A}}$, $\hat{\rho}^{\mathrm{AMH}}$, and $\hat{\rho}^{\mathrm{PMH}}$ are sufficient to fully characterize the attractiveness functions. Here, $\bar{\mathcal{A}}$ indicates the drivers' cost sensitivity, and $\hat{\rho}^{\mathrm{AMH}}$ and $\hat{\rho}^{\mathrm{PMH}}$ indicate the drivers' interest in AMH and PMH, compared with NMH, ignoring the costs. For illustration, we let $\bar{\mathcal{A}} = 0.1$, $\hat{\rho}^{\mathrm{AMH}} = 0.25$, and $\hat{\rho}^{\mathrm{PMH}} = 0.75$. The attractiveness functions are plotted in Figure 2(a). Using the expected costs $\mathcal{E}^{\mathrm{AMH}}$ =\$128.93 and $\mathcal{E}^{\mathrm{PMH}}$ =\$137.64, we obtain the attractiveness of AMH and PMH as $\mathcal{A}^{\mathrm{AMH}} = 0.09$ and $\mathcal{A}^{\mathrm{PMH}} = 0.23$. Also, $\mathcal{A}^{\mathrm{NMH}} = \bar{\mathcal{A}} = 0.1$. These attractiveness values are shown in Figure 2(a) using filled circles.

Choice Probabilities. In Figure 2(b), we plot the choice probabilities for different charging options, when the low rate μ^{L} changes, keeping AMH bonus β fixed at \$50. Similarly, in Figure 2(c), we plot the choice probabilities when the AMH bonus β changes, while the PMH low rate is kept fixed at 20 ¢/KWh. Using the current attractiveness values, we compute the current choice probabilities as $\mathcal{P}^{\mathrm{NMH}} = 0.24$, $\mathcal{P}^{\mathrm{AMH}} = 0.21$, and $\mathcal{P}^{\mathrm{PMH}} = 0.55$. Observe a driver is highly likely to choose PMH with probability 0.55, whereas AMH is less attractive and is selected with probability 0.21. These choice probabilities are shown in Figure 2(b)-(c) using filled circles.

Expected Participation. We expect $\mathcal{K}^{\pi} = \mathcal{P}^{\pi}\mathcal{K}$ drivers to sign up for program π, where $\mathcal{K}$ denotes the total number of EV drivers. For illustration, let $\mathcal{K}$ be 10 million. Thus, we expect 2.4 million, 2.1 million, and 5.5 million participants in NMH, AMH, and PMH, respectively. The expected consumption of the participants of each program is shown in Figure 3(a). The cumulative demand is plotted in Figure 3(b). The charging times of the NMH and PMH drivers are respectively

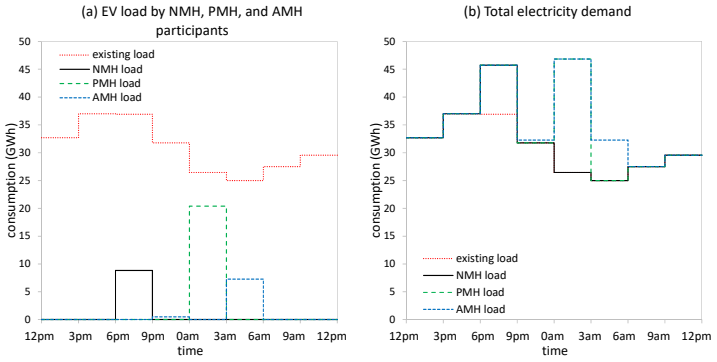

Figure 3 Expected EV consumption and the cumulative electricity demand

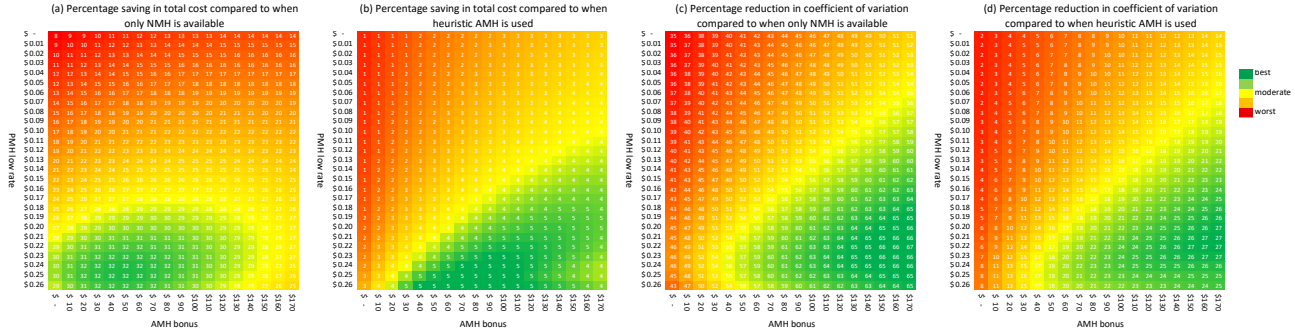

Figure 4 Minimizing cost and the variability in consumption

[6 p.m., 9 p.m.) and [0 a.m., 3 a.m.). The utility assigns the charging times of the AMH drivers to the low-consumption intervals, taking into account the existing load and the EV load by NMH and PMH drivers. Observe the utility assigns the majority of the AMH drivers to [3 a.m., 6 a.m.), and the remaining to [9 p.m., 0 a.m.).

Cost Minimization. We compute the cost of the above solution (Figure 3) and compare it with the cost if only NMH is available. Our computations indicate this solution has a 30.07% lower cost. Recall, for illustration, we let the AMH bonus and PMH low rate be β =\$50 and μ^{L}=20 ¢/KWh. One could repeat the above steps for all practical values of β and μ^{L} to identify an optimal combination. The heatmap in Figure 4(a) shows this step. Using this heatmap, the best solution is β =\$50 and μ^{L}=25 ¢/KWh. If we use this optimal solution, the expected cost of an AMH driver and a PMH driver are respectively \$128.93 and \$172.05. Recalling the choice probabilities, the utility needs to provide a significantly higher incentive to the AMH participants, because we assume (ignoring costs) PMH is three times more attractive than AMH. If we assume AMH is more attractive and set, for example, $\hat{\rho}^{\mathrm{AMH}} = 0.75$ and $\hat{\rho}^{\mathrm{PMH}} = 0.25$, the optimal solution would be β =\$20 and μ^{L}=22 ¢/KWh, with an optimal cost saving of 33.54%. One could similarly obtain an optimal solution for any set of inputs.

The cost reductions shown in Figure 4(a) result from (i) the availability of AMH and PMH and

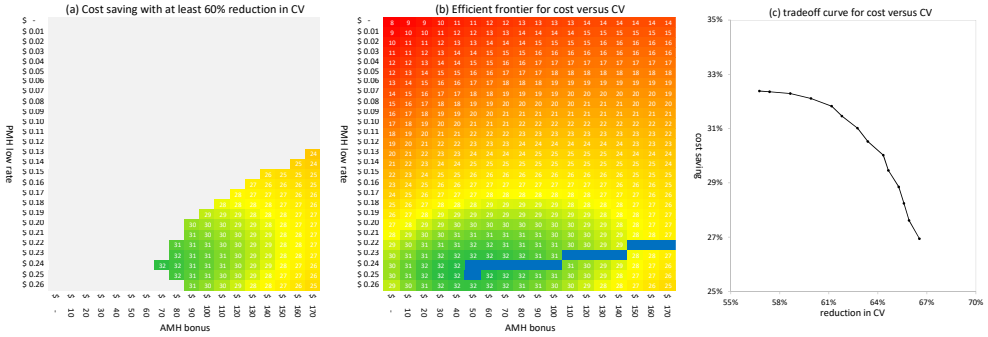

Figure 5 Trade-off between cost and demand variability

(ii) optimizing AMH. In Figure 4(b), the impact of optimizing AMH is demonstrated by comparing solutions to a heuristic implementation of AMH. In line with the primary objectives of AMH programs, an intuitive practical heuristic is to evenly distribute EV loads across participants' charging windows. This heuristic is applied myopically, without optimization. Notably, this practical heuristic is theoretically related to our optimization methodology, as detailed in Theorem 2(a). The value added by optimizing AMH beyond this practical heuristic is depicted in Figure 4(b). For instance, with β =\$50 and μ^{L}=25 ¢/KWh, optimizing AMH contributes 4.74% to the total cost reduction (32.40%), with the remainder attributed to the availability of AMH and PMH.

5.2. Managerial Insights

We investigate the trade-off between cost and demand variability, evaluate customizing the low-rate episode in PMH, and study the impact of charging frequency on utilities' total cost and demand variability. Online Appendix D presents insights on the amount and timing of the improvements achievable in cost and demand variability by MH programs.

Trade-off Between Cost and Coefficient of Variation (CV). Besides minimizing cost, utilities are concerned with reducing the variability of demand. In Figure 4(c) and (d), we compute the reduction in the CV of expected demand and the contribution of optimizing AMH, respectively. Using the cost-minimizing solution (β =\$50 and μ^{L}=25 ¢/KWh), CV reduces 57% compared with the case in which only NMH is available. Optimzing AMH contributes 18.95% of the CV reduction beyond the practical heuristic. Observe the cost-minimizing solution does not minimize CV. An explanation is as follows. Minimizing cost incorporates the bonus paid to the participants, whereas minimizing CV solely concentrates on the final consumption curve. Thus, a trade-off exists between cost and demand variability. If the utility is interested in simultaneously minimizing cost and demand variability, one could compute the efficient frontier as follows. Recall from Figure 4(a) that the cost-minimizing solution is β =\$50 and μ^{L}=25 ¢/KWh, which results in a 57% reduction in CV. Assume the utility is not convinced the demand variability is reasonably low and requires a

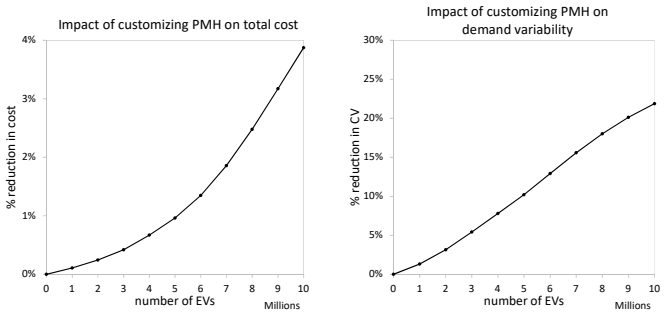

Figure 6 Effects of customizing PMH

solution that achieves at least a 60% or higher reduction in CV. Imposing this additional constraint eliminates some combinations of β and μ^{L}. The combinations that are still feasible are shown in Figure 5(a). Among the feasible combinations, the best combination is β =\$70 and μ^{L}=24 ¢/KWh, which leads to a 32.11% reduction in cost and a 60% reduction in CV. Note that if the utility decides to use this solution, it slightly increases the cost to achieve a smaller variability in demand. We repeat the above procedure for different limits on the CV reduction and create a trade-off curve, shown in Figure 5(c). The corresponding efficient solutions are shown in Figure 5(b).

Figure 5(c) provides detailed insights on the significance of the trade-off between cost and CV. For example, compared with the cost-minimizing solution (the upper-left solution in Figure 5(c)), the utility could achieve an additional 9.8% reduction in CV if it is willing to sacrifice 5.4% of the cost saving. This result is similar to findings in supply-chain management where slight flexibility in cost can significantly mitigate risks. For example, in the newsvendor problem, the optimal solution balances a trade-off between understocking and overstocking (see, e.g., Porteus 1990, Petruzzi and Dada 1999). Overstocking results in increased ordering and inventory holding costs but decreases the likelihood of shortages. Similarly, in Figure 5(c), the 9.8% reduction in CV could account for significant increases in grid reliability. In short, our methodology can effectively be used as a decision support tool to identify an appropriate trade-off between cost and demand variability.

Customizing PMH. To demonstrate the benefits that could be achieved by customizing the timing of the low-rate episode in PMH, we assume the utility assigns a starting time of 0 a.m. to half of its PMH participants and a starting time of 3 a.m. to the remaining half. Figure 6 shows the additional improvements in cost and demand variability for different numbers of EVs (β =\$50 and μ^{L}=20 ¢/KWh). Observe that in this instance, customization is always beneficial, and its benefits increase as the number of EVs grows. Thus, because the number of EVs will significantly increase in the next decades, utilities need to customize PMH.

Charging Frequency. So far, we have assumed each driver charges her EV every night. Some EV drivers may skip some nights, due to various work-/life-related circumstances and to improve the lifespan of their batteries, whereas others may prefer charging every night to ensure they do

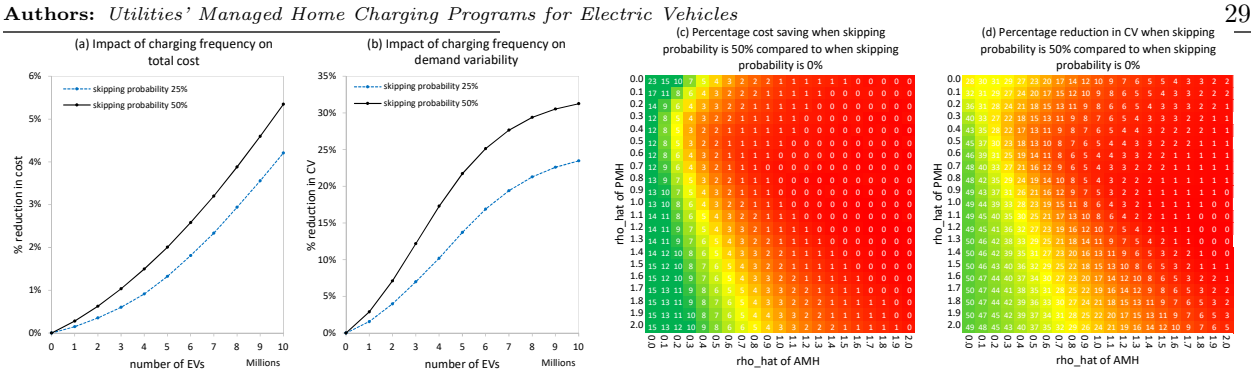

Figure 7 Effects of charging frequency

not run out of power. A household with two EVs may find it economical to alternate the use of a single charger to charge each EV every other night. Utilities may benefit if AMH participants charge more frequently, because of the increased flexibility to manage their EV loads. Thus, we investigate whether utilities would benefit from all EV drivers charging more frequently. We consider a simple stationary model as follows. We assume if a driver skips a night, she will charge her EV (with a normalized load requirement of six hours) the following night with probability 1; that is, $P(\text{“charge tonight”}\,|\,\text{“skipped last night”}) = 1$. Also, given that a driver charged her EV last night, she charges tonight with probability $\hat{p}$; that is, $\hat{p} \triangleq P(\text{“charge tonight”}\,|\,\text{“charged last night”})$. We assume an i.i.d. skipping distribution across the drivers. With some algebra, one could show $\bar{p} \triangleq P(\text{“skip tonight”}) = \frac{1-\hat{p}}{2-\hat{p}}$; for example, if $\hat{p} = 0.75$, then $\bar{p} = 0.2$, meaning, on any night, 20% of the drivers skip charging. With some algebra, one could show the skipping probability satisfies $\bar{p} \in [0, 0.5]$. Figure 7(a)-(b) shows how the skipping probability affects the total cost and demand variability. Inconsistent with intuition, we observe a higher skipping probability is more beneficial. An explanation is as follows. An NMH or a PMH driver who skips tonight has a normalized load requirement of six hours tomorrow, and hence, she naturally shifts half of her charging to off-peak times. Thus, the utility will be worse off if it promotes a culture of charging EVs every night. We further investigate the effects of driver type on this insight. The driver type influences the participation levels in MH programs through the attractiveness functions. Thus, we plot two heatmaps in Figure 7(c)-(d), considering different attractiveness parameters. According to this figure, a higher skipping probability becomes more advantageous as AMH becomes less attractive and PMH becomes more attractive. We conclude promoting a culture of charging EVs every night may significantly increase utilities' total cost if NMH and/or PMH have high participation levels.

In short, in this section, we presented our numerical analysis of PD and managerial insights. Robustness and generalizations are discussed in Online Appendix E.

6. Concluding Remarks

We present mathematical models and solution approaches to jointly design and execute MH programs. A summary of our main results and insights is provided in §1.1. We offer the following

directions for future research.

- Because MH programs are still in the design and pilot stages, real data depicting the behavior of EV drivers under large-scale implementations are currently unavailable. Therefore, our numerical study of LM relies on a limited number of simulated scenarios. Further analysis of the scenarios using real data, once such datasets become available, is an important avenue for future research.
- We scaled down the number of EV drivers in our numerical study of LM to ensure manageable experimentation times. In real-world implementations, more computational resources will be necessary for processing inputs and outputs. Subsequent research endeavors could explore innovative techniques to optimize the data-processing steps.
- Last, in Online Appendix G, we extend our methodology to cases where (i) drivers' requested loads are uncertain and (ii) grid capacity is uncertain. We offer further analyses on these extensions and study other decisions utilities face alongside EVs, such as capacity investment decisions (renewable vs. conventional), as directions for future research.

References

Agrawal V, Bellos I, Ren H (2022) The role of dealer demonstration in the adoption of electric vehicles .

Agrawal VV, Yücel Ş (2022) Design of electricity demand-response programs. *Management Science* 68(10):7441–7456.

Alizamir S, Farajbakhsh F, Wang S (2020) Electricity pricing with limited consumer response. .

Ata B, Duran AS, Islegen O (2018) An analysis of time-based pricing in retail electricity markets. *Available at SSRN 2826055* .

Avci B, Girotra K, Netessine S (2015) Electric vehicles with a battery switching station: Adoption and environmental impact. *Management Science* 61(4):772–794.

Bean JC, Birge JR, Smith RL (1987) Aggregation in dynamic programming. *Operations Research* 35(2):215–220.

Bertsekas DP, Castanon DA, et al. (1988) Adaptive aggregation methods for infinite horizon dynamic programming .

Bertsimas D, Tsitsiklis JN (1997) *Introduction to linear optimization*, volume 6 (Athena Scientific Belmont, MA).

Blair B, Fitzgerald G, Dougherty C (2021) The state of managed charging in 2021. *Smart Electric Power Alliance. https://sepapower.org/resource/the-state-of-managed-charging-in-2021/* .

ChargeHub (2020) 2020 guide on how to charge your electric car with charging stations. *https://chargehub.com/en/electric-car-charging-guide.html#homecharging* .

Chen L, He L, Zhou Y (2023) An exponential cone programming approach for managing electric vehicle charging. *Operations Research* .

Cooper A, Schefter K (2018) Electric vehicle sales forecast and the charging infrastructure required through 2030. *Institute for Electric Innovation. Report.* *https://www.edisonfoundation.net/-/media/Files/IEI/publications/IEI_EEI-EV-Forecast-Report_Nov2018.ashx* .

Fattahi A, Dasu S, Ahmadi R (2023a) Peak-load energy management by direct load control contracts. *Management Science* 69(5):2788–2813.

Fattahi A, Ghodsi S, Dasu S, Ahmadi R (2023b) Flattening energy-consumption curves by monthly constrained direct load control contracts. *Operations Research* .

Fattahi A, Li Y, Sahin O (2022) Customer-driven bundle promotion optimization at scale. *Available at SSRN 4200758* .

Feldman D, Monemizadeh M, Sohler C (2007) A ptas for k-means clustering based on weak coresets. *Proceedings of the twenty-third annual symposium on Computational geometry*, 11–18.

Fitzgerald G, Dougherty C (2021) Managed charging incentive design: Guide to utility program development. *Smart Electric Power Alliance.* *https://sepapower.org/resource/managed-charging-incentive-design/* .

Gallego G, Ratliff R, Shebalov S (2015) A general attraction model and sales-based linear program for network revenue management under customer choice. *Operations Research* 63(1):212–232.

GITT, ISATT (2019) Summary report on evs at scale and the u.s. electric power system. *Grid Integration Tech Team and Integrated Systems Analysis Tech Team. Report.* *https://www.energy.gov/sites/prod/files/2019/12/f69/GITT%20ISATT%20EVs%20at%20Scale%20Grid%20Summary%20Report%20FINAL%20Nov2019.pdf* .

Hanvey C (2019) Ev managed charging: lessons from utility pilot programs. *Smart Electric Power Alliance* .

He L, Ma G, Qi W, Wang X (2021) Charging an electric vehicle-sharing fleet. *Manufacturing & Service Operations Management* 23(2):471–487.

Jiang DR, Powell WB (2016) Practicality of nested risk measures for dynamic electric vehicle charging. *arXiv preprint arXiv:1605.02848* .

Jin C, Tang J, Ghosh P (2013) Optimizing electric vehicle charging with energy storage in the electricity market. *IEEE Transactions on Smart Grid* 4(1):311–320.

Keskin NB, Li Y, Sunar N (2020) Data-driven clustering and feature-based retail electricity pricing with smart meters. *Available at SSRN 3686518* .

Lim MK, Mak HY, Rong Y (2015) Toward mass adoption of electric vehicles: Impact of the range and resale anxieties. *Manufacturing & Service Operations Management* 17(1):101–119.

Mak HY, Rong Y, Shen ZJM (2013) Infrastructure planning for electric vehicles with battery swapping. *Management science* 59(7):1557–1575.

Mendelssohn R (1982) An iterative aggregation procedure for markov decision processes. *Operations Research* 30(1):62–73.

Mettu RR, Plaxton CG (2004) Optimal time bounds for approximate clustering. *Machine Learning* 56(1):35–60.

Myers E (2019) A comprehensive guide to electric vehicle managed charging. *Smart Electric Power Alliance.* *https://sepapower.org/resource/a-comprehensive-guide-to-electric-vehicle-managed-charging* .

NREL (2018) Electrification futures study: Scenarios of electric technology adoption and power consumption for the united states. *National Renewable Energy Laboratory.* *https://www.nrel.gov/docs/fy18osti/71500.pdf* .

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

Petruzzi NC, Dada M (1999) Pricing and the newsvendor problem: A review with extensions. *Operations research* 47(2):183–194.

PG&E (2023) Base interruptible program details. `https://www.pge.com/en_US/large-business/save-energy-and-money/energy-management-programs/demand-response-programs/base-inerruptible/base-inerruptible.page`, accessed: 11-06-2023.

Porteus EL (1990) Chapter 12 stochastic inventory theory. *Stochastic Models*, volume 2 of *Handbooks in Operations Research and Management Science*, 605–652 (Elsevier), URL `http://dx.doi.org/https://doi.org/10.1016/S0927-0507(05)80176-8`.

Rogers DF, Plante RD, Wong RT, Evans JR (1991) Aggregation and disaggregation techniques and methodology in optimization. *Operations Research* 39(4):553–582.

SCE (2023) Summer discount plan. `https://www.sce.com/residential/rebates-savings/summer-discount-plan`, accessed: 11-06-2023.

Schneider F, Thonemann UW, Klabjan D (2018) Optimization of battery charging and purchasing at electric vehicle battery swap stations. *Transportation Science* 52(5):1211–1234.

Shi L, Sethi SP, Çakanyıldırım M (2022) Promoting electric vehicles: Reducing charging inconvenience and price via station and consumer subsidies. *Production and Operations Management* 31(12):4333–4350.

Sun B, Sun X, Tsang DH, Whitt W (2019) Optimal battery purchasing and charging strategy at electric vehicle battery swap stations. *European Journal of Operational Research* 279(2):524–539.

Trabish H (2019) The rise of evs could overwhelm the grid, but pge has a better plan. *DEEP DIVE.* *https://www.utilitydive.com/news/the-rise-of-evs-could-overwhelm-the-grid-but-pge-has-a-better-plan/555984/* .

Van Roy B (2006) Performance loss bounds for approximate value iteration with state aggregation. *Mathematics of Operations Research* 31(2):234–244.

Varma SM, Castro F, Maguluri ST (2023) Electric vehicle fleet and charging infrastructure planning. *arXiv preprint arXiv:2306.10178* .

Whitt W (1978) Approximations of dynamic programs, i. *Mathematics of Operations Research* 3(3):231–243.

Wu OQ, Yücel Ş, Zhou Y (2021) Smart charging of electric vehicles: An innovative business model for utility firms. *Manufacturing & Service Operations Management* .

Zhang T, Chen W, Han Z, Cao Z (2013) Charging scheduling of electric vehicles with local renewable energy under uncertain electric vehicle arrival and grid power price. *IEEE Transactions on Vehicular Technology* 63(6):2600–2612.

Zhang Y, Lu M, Shen S (2021) On the values of vehicle-to-grid electricity selling in electric vehicle sharing. *Manufacturing & Service Operations Management* 23(2):488–507.

Utilities' Managed Home Charging Programs for Electric Vehicles

(Online Appendix)

Appendix A: An illustrative example of PD

Assume a utility is planning to execute its MH programs in July-August. Assume it has set $\mathcal{I}^\circ = 600$ hours and $\mathcal{L} = 6$ hours, and it has identified that the worthwhile starting times (in terms of shifting the EV load) are 9 p.m., 10 p.m., 11 p.m., and 12 a.m. on each day (i.e., $\mathcal{T}_d$ is the same for different d). Table 1 is an example of the input data that we need to construct and solve PD. We assume 10 EV drivers exist. The last four columns of Table 1 need to be collected for all d, and one could consider day types rather than individual days, for example, weekdays and weekends. Table 2 shows three driver types that we have formed by partitioning the 10 drivers. The members of each type are homogeneous. Observe types 1 and 2 have lower flexibility than the threshold ($\mathcal{I}^\circ = 600$ hours), and hence, they are not qualified for AMH. By contrast, type-3 drivers have high flexibility, meaning they qualify for AMH, but because none of the potential starting times are compatible with their driving routines, they are not qualified for PMH. In short, the utility offers PMH and NMH to types 1 and 2, and it offers AMH and NMH to type 3.

Let the common electricity rate be $\bar{\mu} = 26$ ¢/KWh. In Table 3, assuming $\mu^{\mathrm{L}} = 20$ ¢/KWh and $\beta =$\$50, we compute the choice probabilities and the corresponding offer sets. For example, the type-1 drivers select PMH and NMH with probabilities 0.43 and 0.57, respectively.

One could compute $\mathcal{H}(\beta, \mu^{\mathrm{L}})$ using the choice probabilities. To estimate $\mathcal{F}(\beta, \mu^{\mathrm{L}})$, one could simulate several participation scenarios, denoted by index s^{P}, using the choice probabilities; in each participation scenario s^{P}, each customer belongs to a program AMH, PMH, or NMH. For each s^{P}, one could use LM to estimate $\mathcal{F}(\beta, \mu^{\mathrm{L}}|s^{\mathrm{P}})$. Last, the expected energy cost of the solution ($\mu^{\mathrm{L}} = 20$ ¢/KWh and $\beta =$\$50) is $\mathcal{F}(\beta, \mu^{\mathrm{L}}) = \mathbf{E}_{s^{\mathrm{P}}}[\mathcal{F}(\beta, \mu^{\mathrm{L}}|s^{\mathrm{P}})]$. By repeating the above steps for all practical values of μ^{L} and β, the utility finds an optimal solution for PD.

Table 1 Input data of EV drivers

j	$\mathcal{Q}_j$ (KWh)	$\mathcal{I}_j$ (hours)	$\hat{\rho}^\pi_j$ AMH	$\hat{\rho}^\pi_j$ PMH	$\hat{\rho}^\pi_j$ NMH	ρ^π_j AMH	ρ^π_j PMH	ρ^π_j NMH	Does $\tau \in \mathcal{T}_d$ work? 9pm	10pm	11pm	12am
1	600	500	0.25	0.75	1	0.001	0.001	0.001	No	Yes	Yes	Yes
2	600	500	0.25	0.75	1	0.001	0.001	0.001	No	Yes	Yes	Yes
3	600	500	0.25	0.75	1	0.001	0.001	0.001	No	Yes	Yes	Yes
4	400	500	0.75	0.25	1	0.001	0.001	0.001	Yes	Yes	Yes	Yes
5	400	500	0.75	0.25	1	0.001	0.001	0.001	Yes	Yes	Yes	Yes
6	400	500	0.75	0.25	1	0.001	0.001	0.001	Yes	Yes	Yes	Yes
7	400	500	0.75	0.25	1	0.001	0.001	0.001	Yes	Yes	Yes	Yes
8	400	700	0.75	0.25	1	0.001	0.001	0.001	No	No	No	No
9	400	700	0.75	0.25	1	0.001	0.001	0.001	No	No	No	No
10	400	700	0.75	0.25	1	0.001	0.001	0.001	No	No	No	No

Table 2 Forming driver types

φ	$\mathcal{K}_\varphi$	members	$\mathcal{Q}_\varphi$ (KWh)	$\mathcal{I}_\varphi$ (hours)	$\hat{\rho}^\pi_\varphi$ AMH	$\hat{\rho}^\pi_\varphi$ PMH	$\hat{\rho}^\pi_\varphi$ NMH	ρ^π_φ AMH	ρ^π_φ PMH	ρ^π_φ NMH	$\mathcal{T}_{\varphi,d}$
1	3	1,2,3	600	500	0.25	0.75	1	0.001	0.001	0.001	{10pm, 11pm, 12am}
2	4	4,5,6,7	400	500	0.75	0.25	1	0.001	0.001	0.001	{9pm, 10pm, 11pm, 12am}
3	3	8,9,10	400	700	0.75	0.25	1	0.001	0.001	0.001	{}

Table 3 Choice probabilities and offer sets if $\mu^{\mathrm{L}} = 20$ ¢/KWh and $\beta =$\$50

φ	$\mathcal{E}^\pi_\varphi$ (\$) AMH	$\mathcal{E}^\pi_\varphi$ (\$) PMH	$\mathcal{E}^\pi_\varphi$ (\$) NMH	$\mathcal{A}^\pi_\varphi$ AMH	$\mathcal{A}^\pi_\varphi$ PMH	$\mathcal{A}^\pi_\varphi$ NMH	$\mathcal{P}^\pi_\varphi$ AMH	$\mathcal{P}^\pi_\varphi$ PMH	$\mathcal{P}^\pi_\varphi$ NMH	menu
1	—	120	156	—	0.63	0.84	—	0.43	0.57	Π_1 ={PMH,NMH}, $\mathcal{T}_{1,d}$ ={10pm, 11pm, 12am}
2	—	80	104	—	0.17	0.90	—	0.16	0.84	Π_2 ={PMH,NMH}, $\mathcal{T}_{2,d}$ ={9pm, 10pm, 11pm, 12am}
3	54	—	104	0.70	—	0.90	0.44	—	0.56	Π_3 ={AMH,NMH}

Appendix B: An illustrative example for scenarios

The existing data could be used to form scenarios, as we explain through the following toy example. Assume the first three days of the season are Tuesday, Wednesday, and Thursday. Assume the current time is 10 a.m. on Thursday (the third day). Assume we are using a truncated horizon of 24 hours, meaning LM's horizon has 24 periods, starting at 10 a.m. on Thursday and ending at 9 a.m. on Friday (i.e., the starting and ending periods are [10 a.m., 11 a.m.) on Thursday and [9 a.m., 10 a.m.) on Friday, respectively). One could use the two observations from Tuesday and Wednesday as two scenarios to construct LM. Assume AMH has two participants, referred to as A and B. Assume in the first observation (Tuesday), both participants plugged in at 6 p.m. and requested 10 KWh to be charged until 5 a.m. (the next morning). Assume in the second observation (Wednesday), both participants plugged in at 5 p.m. and requested 9 KWh to be charged until 5 a.m. To construct LM at 10 a.m. on Thursday, we use the following two scenarios: (s_1) both participants will plug in at 6 p.m. and request 10 KWh to be charged until 5 a.m. on Friday, and (s_2) both participants will plug in at 5 p.m. and request 9 KWh to be charged until 5 a.m. on Friday. Ignoring weather and other external factors, both scenarios are equally likely because they are observations from similar workdays; hence, one could assume each scenario occurs with probability 0.5.

We also explain how the scenarios are updated over time. In the above example, by solving LM at 10 a.m. on Thursday (assuming none of the participants show up until then), no supply is assigned to the first hour, and hence, no supply is made. We repeat the above steps to construct and solve LM at 11 a.m., 12 p.m., 1 p.m., and so forth. As long as no participant shows up, LM assigns no supply to the first period. Assume participant A plugs in at 4 p.m. on Thursday and requests 9 KWh to be charged until 5 a.m. on Friday. Because we have full information about A, we update her information in both scenarios. The updated scenarios follow: (s_1') participant A plugged in at 4 p.m. and requested 9 KWh to be charged until 5 a.m. on Friday, and B will plug in at 6 p.m. and request 10 KWh to be charged until 5 a.m. on Friday, and (s_2') participant A plugged in at 4 p.m. and requested 9 KWh to be charged until 5 a.m. on Friday, and B will plug in at 5 p.m. and request 9 KWh to be charged until 5 a.m. on Friday.

In short, the above toy example illustrates how the existing data could be used to form scenarios. In a real application, one could use more sophisticated approaches to form scenarios, which is beyond the scope of our paper. If the existing data are large, one could cluster them, in which case, the centers of clusters would be the scenarios and their sizes would be the probabilities of the scenarios. High-dimensional clustering techniques could be used in such settings (see, e.g., Mettu and Plaxton 2004, Feldman et al. 2007). To use LM in the program-design phase, if no data exist, one could simulate multiple participation scenarios using the choice probabilities (as we discuss in Online Appendix A), and for each participation scenario, one could simulate multiple scenarios based on the driving routines of the EV drivers.

Scenario-based modeling offers multiple advantages. First, it directly uses the existing (clustered) data, as we explained above. Second, the stochastic model is formulated as a standard (deterministic) optimization model using scenarios. Last, correlations among the EV drivers are automatically incorporated in the model through scenarios.

Appendix C: Proofs

Proof of Theorem 1(a). Suppose we repeat our disaggregation procedure for all $q_{s,m,t}$'s as follows: for each $j \in \mathbb{M}_m$, let $q_{s,j,t} \triangleq \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,t}$. We first analyze the feasibility of these $q_{s,j,t}$'s and then reduce the values of some of these variables and investigate the consequences of these reductions. The total load supplied to group m in period t, under scenario s, is

$$\sum_{j\in\mathbb{M}_m} q_{s,j,t} = \sum_{j\in\mathbb{M}_m} \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,t} = q_{s,m,t}.$$

Therefore, our disaggregation does not violate constraints (6)-(12). We skip constraint (3) for now. Our disaggregation satisfies constraint (4) because

$$q_{s,j,t} = \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,t} \geq 0,$$

and

$$q_{s,j,t} = \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,t} \leq \frac{\hat{Q}_j}{\hat{Q}_m} q^U_m = \frac{\gamma_{m,U} q^U_j}{\hat{Q}_m} \sum_{j'\in\mathbb{M}_m} q^U_{j'} = \frac{q^U_j}{\hat{Q}_m} \sum_{j'\in\mathbb{M}_m} \gamma_{m,U} q^U_{j'} = \frac{q^U_j}{\hat{Q}_m} \sum_{j'\in\mathbb{M}_m} \hat{Q}_{j'} = q^U_j,$$

where $\gamma_{m,U} \triangleq \max_{j\in\mathbb{M}_m} \gamma_j$ for all m. Also note $\hat{Q}_j = \gamma_{m,U} q^U_j$ for all $j \in \mathbb{M}_m$.

Constraint (5) is for the arrivals in the first period. We partition such arrivals with equal time windows into groups. Because arrivals at $t = 1$ are common in all scenarios, we create groups that are also common across scenarios. Our disaggregation procedure works identically for the members of these groups, and hence, constraint (5) is satisfied. Now, let us revisit $q_{s,j,t} \triangleq \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,t}$ and note each participant j may receive more than her actual requirement. By starting from the latest assignment to this participant, let us reduce the values of $q_{s,j,t}$'s until the total load supplied to this participant becomes equal to her actual load. Because we are reducing the values of some $q_{s,j,t}$'s, constraints (4)-(12) remain feasible. We perform such reductions to satisfy constraint (3), and hence, our disaggregation creates a feasible solution.

Proof of Theorem 1(b). Similar to part (a), let our disaggregation procedure consist of two steps. In the first step, for each s, t, m, and $j \in \mathbb{M}_m$, we let $q_{s,j,t} \triangleq \frac{\hat{Q}_j}{\hat{Q}_m} q_{s,m,t}$. In the second step, we reduce the values of some $q_{s,j,t}$'s to ensure we do not supply more than a participant's actual load requirement. These two steps create our AD solution with an objective value of v_{AD}.

Because we increase the required loads of some participants when forming groups, after the first step, we create a solution for a synthetic problem (with increased load requirements), denoted by SLM, in which each participant requests at least as much as that in LM. Note the solution is optimal for SLM because ALM is a relaxation of SLM, and the optimal value of ALM is equal to the objective value of the solution for SLM; hence, $v_{\text{SLM}} = v_{\text{ALM}}$ (note v_{P} denotes the optimal value of problem P).

In the optimal solution of SLM, by decreasing the supply to some participants (the second step), we create the AD solution, which is feasible for LM and has an objective value of v_{AD}. Thus, $v_{\text{LM}} \leq v_{\text{AD}}$. Moreover, in the generated AD solution, the total hourly loads ($y_{s,t}$) are less than or equal to the hourly loads in the optimal solution of SLM. Recall we assume $f_t(y)$ is non-decreasing in $y \geq 0$ for all t. Therefore, $v_{\text{AD}} \leq v_{\text{SLM}}$. In short,

$$v_{\text{LM}} \leq v_{\text{AD}} \leq v_{\text{SLM}} = v_{\text{ALM}}.$$

Let $q^\circ_{s,j,t}$ and $y^\circ_{s,t}$ denote an optimal solution for LM. For each s and j, we have $\sum_{t=a_j}^{a_j+\ell_j-1} q^\circ_{s,j,t} = Q_j$, because of constraint (3). We arbitrarily increase the values of some of these $q^\circ_{s,j,t}$ to $q^\triangle_{s,j,t}$, for $t \in \{a_j, \ldots, a_j + \ell_j - 1\}$, such that the following three conditions hold: (i) $q^\triangle_{s,j,t} \leq q^U_j$; (ii) for each s and j, the total increase is equal to $\hat{Q}_j - Q_j$; and (iii) $q^\triangle_{s,j,1} = q^\triangle_{s',j,1}$ for all $s \neq s'$ and $j \in \mathbb{J}_s \cap \mathbb{J}_{s'}$. Achieving properties (i) and (ii) is feasible because when we were creating groups, we increased each participant's load to $\hat{Q}_j$ as long as her synthetic load is satisfiable within her charging window. Moreover, condition (iii) is feasible because the participants who arrive in the first period exist in all $\mathbb{J}_s$'s, and hence, one needs to employ the same increases

for all such participants. Furthermore, because $h_i \geq \lambda^{\mathrm{U}} \triangleq \check{J}q^{\diamond} + \max_{s,t} \lambda^{\mathrm{U}}_{s,t}$, for all i, and $\check{J}q^{\diamond} \geq \sum_{j\in\mathbb{J}_s} q^{\mathrm{U}}_j \geq \sum_{j\in\mathbb{J}_s} q^{\triangle}_{s,j,t}$, for all s and t, we obtain

$$\begin{aligned}
h_i &\geq \sum_{j\in\mathbb{J}_s} q^{\triangle}_{s,j,t} + \max_{s,t} \lambda^{\mathrm{U}}_{s,t} \\
&\geq \sum_{j\in\mathbb{J}_s} q^{\triangle}_{s,j,t} + \lambda^{\mathrm{U}}_{s,t} \\
&= \sum_{j\in\mathbb{J}_s} q^{\triangle}_{s,j,t} + \sum_{g} \left(\lambda^{\mathrm{EXI}}_{s,g,t} + \lambda^{\mathrm{NMH}}_{s,g,t} + \sum_{\varphi:\ \mathrm{PMH}\in\Pi_\varphi} \max_{\tau\in\mathcal{T}_{\varphi,d(t)}} \lambda^{\mathrm{PMH}}_{s,g,t,\varphi,\tau} \right) \\
&\geq \sum_{j\in\mathbb{J}_s} q^{\triangle}_{s,j,t} + \sum_{g} \left(\lambda^{\mathrm{EXI}}_{s,g,t} + \lambda^{\mathrm{NMH}}_{s,g,t} + \sum_{\varphi:\ \mathrm{PMH}\in\Pi_\varphi} \sum_{\tau\in\mathcal{T}_{\varphi,d(t)}} \lambda^{\mathrm{PMH}}_{s,g,t,\varphi,\tau} \vartheta^{\triangle}_{s,\tau,\varphi,d(t)} \right) \\
&= \sum_{g} \left(\sum_{j\in\mathbb{J}_s:\ g_j=g} q^{\triangle}_{s,j,t} + \lambda^{\mathrm{EXI}}_{s,g,t} + \lambda^{\mathrm{NMH}}_{s,g,t} + \sum_{\varphi:\ \mathrm{PMH}\in\Pi_\varphi} \sum_{\tau\in\mathcal{T}_{\varphi,d(t)}} \lambda^{\mathrm{PMH}}_{s,g,t,\varphi,\tau} \vartheta^{\triangle}_{s,\tau,\varphi,d(t)} \right) \\
&= \sum_{g} x^{\triangle}_{s,g,t},
\end{aligned}$$

for all $s,\ t$, and i; hence, the grid constraints remain feasible with the increases in the participants' requirements. Note $\vartheta^{\triangle}_{s,\tau,\varphi,d} \triangleq \vartheta^{\circ}_{s,\tau,\varphi,d}$, for all $s,\ \varphi:\ \mathrm{PMH}\in\Pi_\varphi,\ \tau\in\mathcal{T}_{\varphi,d}$, and d. Also, $x^{\triangle}_{s,g,t}$'s are the corresponding values in the solution that we created for SLM. Now we have $\sum_{t=a_j}^{a_j+\ell_j-1} q^{\triangle}_{s,j,t} = \hat{Q}_j$; that is, constraint (3) is satisfied for SLM. All other constraints are also satisfied, and hence, $q^{\triangle}_{s,j,t}$ is a feasible solution for SLM. Let $y^{\triangle}_{s,t}$ denote the corresponding values for $y_{s,t}$'s in this feasible solution. Observe we have created a feasible solution for SLM. Let $v^{\triangle}$ denote the objective value of this feasible solution. We have $v^{\triangle} \geq v_{\mathrm{SLM}}$. Thus, in short,

$$v_{\mathrm{LM}} \leq v_{\mathrm{AD}} \leq v_{\mathrm{SLM}} = v_{\mathrm{ALM}} \leq v^{\triangle}.$$

Therefore, the error is upper bounded as

$$\begin{aligned}
v_{\mathrm{AD}} - v_{\mathrm{LM}} &\leq v^{\triangle} - v_{\mathrm{LM}} \\
&= \sum_s p_s \sum_t \left(f_t(y^{\triangle}_{s,t}) - f_t(y^{\circ}_{s,t}) \right) \\
&= \sum_s p_s \sum_t \left(f_t(y^{\circ}_{s,t} + \varepsilon_{s,t}) - f_t(y^{\circ}_{s,t}) \right),
\end{aligned}$$

where $\varepsilon_{s,t} \triangleq y^{\triangle}_{s,t} - y^{\circ}_{s,t} \geq 0$ is the increase in $y^{\circ}_{s,t}$ in period t, under scenario s. Recalling $y^{\circ}_{s,t}$ is optimal for LM, we have $\lambda^{\mathrm{L}}_{s,t} \leq y^{\circ}_{s,t} \leq \lambda^{\mathrm{U}}_{s,t} + \check{J}q^{\diamond}$, for all s and t, because AMH participants can increase each hourly load by at most $\check{J}q^{\diamond}$. Thus,

$$\begin{aligned}
v_{\mathrm{AD}} - v_{\mathrm{LM}} &\leq \sum_s p_s \sum_t \left\{ \max_{\lambda^{\mathrm{L}}_{s,t} \leq y \leq \lambda^{\mathrm{U}}_{s,t} + \check{J}q^{\diamond}} \left(f_t(y + \varepsilon_{s,t}) - f_t(y) \right) \right\} \\
&\leq \sum_s p_s \sum_t \max_{s,t\ [\text{fix } \varepsilon_{s,t}]} \left\{ \max_{\lambda^{\mathrm{L}}_{s,t} \leq y \leq \lambda^{\mathrm{U}}_{s,t} + \check{J}q^{\diamond}} \left(f_t(y + \varepsilon_{s,t}) - f_t(y) \right) \right\} \\
&= \sum_s p_s \sum_t \Theta(\varepsilon_{s,t}).
\end{aligned}$$

The first maximization in the second line keeps the value of $\varepsilon_{s,t}$ fixed. In the remainder, we bound the value

of $\varepsilon_{s,t}$. In each scenario, the total increase in the values of $q^{\circ}_{s,j,t}$'s is upper bounded by $J\epsilon q^{\diamond}$. Thus, the worst-case error is upper bounded as

$$\begin{aligned} v_{\mathrm{AD}} - v_{\mathrm{LM}} &\le \sum_s p_s \left\{ \max_{\substack{\sum_t \varepsilon_{s,t} \le J\epsilon q^{\diamond} \\ \varepsilon_{s,t} \ge 0,\ \forall t}} \sum_t \Theta(\varepsilon_{s,t}) \right\} \\ &= \max_{\substack{\sum_t \varepsilon_t \le J\epsilon q^{\diamond} \\ \varepsilon_t \ge 0,\ \forall t}} \sum_t \Theta(\varepsilon_t) \\ &= \Psi(J\epsilon q^{\diamond}), \end{aligned}$$

which completes the proof.

Proof of Theorem 2(a). First, receiving Q_j/ℓ_j in a period is within her charging speed limit, because otherwise, if $\frac{Q_j}{\ell_j} > q^U_j$, then $\frac{Q_j}{q^U_j} > \ell_j$, implying $\gamma_j = \frac{Q_j}{q^U_j} > \ell_j$, which is a contradiction.

Consider problem TLM(1), which is constructed and solved at the beginning of her arrival time $t = a_j$. We consider a $\left(\frac{1}{\ell_j}\right)$-fraction of her load to be supplied within the truncated horizon of one period. Observe no optimization is required, and we supply $\frac{Q_j}{\ell_j}$ to participant j in period $t = a_j$. Also recall we assume TLM(1) is feasible. At the beginning of the next period $t = a_j + 1$, she needs $\frac{(\ell_j - 1)Q_j}{\ell_j}$ to be supplied in the next $\ell_j - 1$ periods. Consider constructing TLM(1) at the beginning of period $t = a_j + 1$. We consider a $\left(\frac{1}{\ell_j - 1}\right)$-fraction of her remaining load, which is

$$\left(\frac{1}{\ell_j - 1}\right)\left(\frac{(\ell_j - 1)Q_j}{\ell_j}\right) = \frac{Q_j}{\ell_j},$$

to be supplied within the truncated horizon of one period. Again, no optimization is required, and we supply $\frac{Q_j}{\ell_j}$ to participant j in period $t = a_j + 1$. Similarly, one could show she receives $\frac{Q_j}{\ell_j}$ in each period $t \in \{a_j, \ldots, a_j + \ell_j - 1\}$, which completes the proof.

Proof of Theorem 2(b). Let STLM($\breve{T}$) denote the rolling-horizon solution (i.e., supply amounts made over the horizon) achieved by solving TLM($\breve{T}$) on an hourly basis over the horizon. We construct an example to show STLM(1) can have a strictly better objective value than STLM(2).

Let the horizon consist of two periods. Assume only one EV driver exists and she participates in AMH; she arrives at the beginning of $t = 1$ and requests $Q > 0$ to be supplied in the next two periods ($\ell = 2$). We drop index j to simplify our notations. Assume $Q \le q^U$, meaning her requirement can be supplied in one period. Assume $f_1(\cdot) = f_2(\cdot) = f(\cdot)$, and $f(y)$ is strictly convex in $y \ge 0$. Assume $G = 1$, and the region is served by one substation with capacity $h_1 = 2Q$. Assume the actual existing load in the first and second periods are Q and 0.

We compare the solutions of STLM(1) and STLM(2).

- STLM(1): The solution of TLM(1) assigns $\frac{Q}{2}$ to the first and second periods; hence, the total cost is $f\left(\frac{3}{2}Q\right) + f\left(\frac{1}{2}Q\right)$.
- STLM(2): Consider constructing TLM(2) at the beginning of the first period. Assume only one scenario exists in which the existing load is $\lambda^{\mathrm{EXI}}_{1,1,1} = Q$ and $\lambda^{\mathrm{EXI}}_{1,1,2} = 2Q$. Observe the forecast of the second period at the beginning of the first period is inaccurate. The solution of TLM(2) assigns the supply of Q to the first period, and hence, it is implemented. Thus, the total cost of this solution is $f(2Q) + f(0)$.

Before we proceed, note the solution of TLM(2) assigns the supply of Q to the first period because of the following property (also known as *increasing marginal return*): for all $y'' > y' \ge 0$ and $\alpha > 0$, we have $f(y' + \alpha) - f(y') < f(y'' + \alpha) - f(y'')$, because

$$f(y'+\alpha) = f\left(\frac{y''-y'}{y''-y'+\alpha}y' + \frac{\alpha}{y''-y'+\alpha}(y''+\alpha)\right) < \frac{y''-y'}{y''-y'+\alpha}f(y') + \frac{\alpha}{y''-y'+\alpha}f(y''+\alpha), \text{ and}$$
$$f(y'') = f\left(\frac{\alpha}{y''-y'+\alpha}y' + \frac{y''-y'}{y''-y'+\alpha}(y''+\alpha)\right) < \frac{\alpha}{y''-y'+\alpha}f(y') + \frac{y''-y'}{y''-y'+\alpha}f(y''+\alpha),$$

and hence, by combining these inequalities, we obtain

$$\begin{aligned} & f(y'+\alpha)+f(y'') < f(y')+f(y''+\alpha) \\ \Leftrightarrow\ & f(y'+\alpha)-f(y') < f(y''+\alpha)-f(y''), \quad \forall y''>y'\geq 0,\ \forall \alpha>0. \end{aligned} \tag{17}$$

Moreover, if either $\alpha=0$ or $y''=y'$, inequality (17) holds as an equality.

Next, we compare the objective values of the solutions of STLM(1) and STLM(2). Because we assume $f(0)=0$ and $f(\cdot)$ is strictly convex, we have $f(y')+f(y'') < f(y'+y'')$ for all $y',y''>0$. To see why this inequality holds, in inequality (17), let $y'=0$ and $\alpha=y'''>0$; then,

$$\begin{aligned} & f(y''')-f(0) < f(y''+y''')-f(y'') \\ \Leftrightarrow\ & f(y''')+f(y'') < f(y''+y'''), \quad \forall y'',y'''>0. \end{aligned} \tag{18}$$

Therefore, using inequality (18), we have $f\left(\frac{3}{2}Q\right)+f\left(\frac{1}{2}Q\right) < f(2Q)$, which completes the proof.

Proof of Theorem 2(c). We prove this part for $\breve{T}' = \breve{T}+1$, and it can readily be extended to any $\breve{T}' \geq \breve{T}$. Note truncation is only applied to the AMH participants, meaning $\lambda^{\text{EXI}}_{s,g,t}$, $\lambda^{\text{NMH}}_{s,g,t}$, and $\lambda^{\text{PMH}}_{s,g,t,\varphi,\tau}$ in the two problems are equal in periods $t\in\{1,\ldots,\breve{T}\}$.

Consider an optimal solution to TLM($\breve{T}$). In this optimal solution, for participant j, let $q^{\circ}_{s,j,t}$ denote the amount of load supplied to this participant in period $t\in\{1,\ldots,\breve{T}\}$, under scenario s. Consider increasing the horizon to $\breve{T}+1$ periods. In problem TLM($\breve{T}+1$), let us use the same $q^{\circ}_{s,j,t}$ for $t\in\{1,\ldots,\breve{T}\}$, and for $t=\breve{T}+1$, let $q_{s,j,t}$ be as follows: if participant j's charging window ends before $t=\breve{T}+1$, then $q_{s,j,\breve{T}+1}=0$; otherwise, let $q_{s,j,\breve{T}+1}$ be equal to the additional load to be considered when going from $\breve{T}$ to $\breve{T}+1$. This additional load can be shown to satisfy the charging-speed constraint (4) and demand-satisfaction constraint (3). The grid constraints (11) are also satisfied following the assumption $h_i \geq \lambda^{\text{U}}$, for all i (because the supply quantities in period $\breve{T}+1$ satisfy the charging speed constraints, i.e., $q_{s,j,\breve{T}+1} \leq q^{\text{U}}_j$, for all s and j, similar to the proof of Theorem 1(b), one could show the grid constraints are satisfied). Thus, an optimal solution of TLM($\breve{T}$) is feasible for TLM($\breve{T}+1$), completing the proof.

Proof of Theorem 2(d). With some abuse of notation, let $\lambda_t \geq 0$ denote the existing load in period $t\in\{1,\ldots,T\}$ (before truncation), recalling $S=G=1$. Let periods $\hat{t}$ and $\hat{t}+1$ be the two periods in the participant's charging window, and let $Q>0$ denote her requirement (note if $Q=0$, the error is zero; hence, we assume $Q>0$ in the rest of the proof). Because $\breve{T}=1$, $\frac{Q}{2}$ is assigned to each period $\hat{t}$ and $\hat{t}+1$ (following part (a)). Let $q_{\hat{t}}$ and $q_{\hat{t}+1}$ denote the optimal amounts supplied to this participant in periods $\hat{t}$ and $\hat{t}+1$, respectively. The worst-case relative error is

$$\begin{aligned} \rho_{\text{T}} &\leq \max_{Q>0,\ \lambda_t\geq 0,\ \forall t} \left\{ \frac{\sum_{t'\notin\{\hat{t},\hat{t}+1\}} \lambda^3_{t'} + \sum_{t'\in\{\hat{t},\hat{t}+1\}} \left(\frac{Q}{2}+\lambda_{t'}\right)^3}{\min\limits_{q_{\hat{t}},\ q_{\hat{t}+1}\geq 0,\ q_{\hat{t}}+q_{\hat{t}+1}=Q} \sum_{t'\notin\{\hat{t},\hat{t}+1\}} \lambda^3_{t'} + \sum_{t'\in\{\hat{t},\hat{t}+1\}} \left(q_{t'}+\lambda_{t'}\right)^3} \right\} - 1 \\ &= \max_{Q>0,\ \lambda_{\hat{t}},\ \lambda_{\hat{t}+1}\geq 0} \left\{ \frac{\sum_{t'\in\{\hat{t},\hat{t}+1\}} \left(\frac{Q}{2}+\lambda_{t'}\right)^3}{\min\limits_{q_{\hat{t}},\ q_{\hat{t}+1}\geq 0,\ q_{\hat{t}}+q_{\hat{t}+1}=Q} \sum_{t'\in\{\hat{t},\hat{t}+1\}} \left(q_{t'}+\lambda_{t'}\right)^3} \right\} - 1 \end{aligned}$$

$$= \max_{Q>0,\ \frac{\lambda_{\hat{t}}}{Q},\ \frac{\lambda_{\hat{t}+1}}{Q}\geq 0} \left\{ \frac{\sum_{t'\in\{\hat{t},\hat{t}+1\}} \left(\frac{1}{2}+\frac{\lambda_{t'}}{Q}\right)^3}{\min_{\frac{q_{\hat{t}}}{Q},\ \frac{q_{\hat{t}+1}}{Q}\geq 0,\ \frac{q_{\hat{t}}}{Q}+\frac{q_{\hat{t}+1}}{Q}=1} \sum_{t'\in\{\hat{t},\hat{t}+1\}} \left(\frac{q_{t'}}{Q}+\frac{\lambda_{t'}}{Q}\right)^3} \right\} - 1.$$

To obtain the second line, the outer maximization sets $\sum_{t'\notin\{\hat{t},\hat{t}+1\}} \lambda_{t'}^3$ to zero because both the numerator and denominator of the fraction are positive, and the denominator is less than or equal to the numerator. To obtain the third line, we divide the objective functions and constraints by Q^3 and Q, respectively (note $Q^3>0$). To simplify our notations, let $A \triangleq \frac{\lambda_{\hat{t}}}{Q}$, $B \triangleq \frac{\lambda_{\hat{t}+1}}{Q}$, $C \triangleq \frac{q_{\hat{t}}}{Q}$, and $D \triangleq \frac{q_{\hat{t}+1}}{Q}$. Then,

$$\begin{aligned}\rho_{\mathrm{T}} &\leq \max_{A,\ B\geq 0} \left\{ \frac{\left(A+\frac{1}{2}\right)^3+\left(B+\frac{1}{2}\right)^3}{\min_{C,\ D\geq 0,\ C+D=1} (A+C)^3+(B+D)^3} \right\} - 1 \\ &= \max_{A,\ B\geq 0} \left\{ \frac{\left(A+\frac{1}{2}\right)^3+\left(B+\frac{1}{2}\right)^3}{\min_{0\leq C\leq 1} (A+C)^3+(B+1-C)^3} \right\} - 1.\end{aligned}$$

We provide an intuition before we proceed. We have normalized the load requirement of the EV driver to $\check{Q}=1$. Quantities A and B are the existing loads in the first and second periods ($\hat{t}$ and $\hat{t}+1$), respectively, after normalization, and C and $(1-C)$ are the fractions of $\check{Q}$ that we assign to the first and second periods, respectively. We will consider two cases. In Case (i), either $B>A+1$ or $A>B+1$ (these sub-cases are equivalent, if one switches A and B). Consider $B>A+1$ and note that the existing load in the second period is so high that assigning $\check{Q}$ to the first period is optimal (because the objective function is strictly convex); that is, $C^*=1$. Then, the denominator becomes $(A+1)^3+B^3$. Next, we show the value of the fraction decreases in A, and hence, we set A to 0. The resulting fraction is $\frac{\left(\frac{1}{2}\right)^3+\left(B+\frac{1}{2}\right)^3}{1+B^3}$. The maximum value of this fraction is 1.866, which occurs at $B=1.366$. In Case (ii), $A\leq B+1$ and $B\leq A+1$. We assign some fraction of $\check{Q}$ to each period, such that the cumulative load is equal across the two periods (because the objective function is strictly convex). In other words, the total load in the two periods is $A+B+1$ (including $\check{Q}$), and after assigning $\check{Q}$, half of $A+B+1$ must be in the first period and the other half must be in the second period. Thus, the denominator becomes $\frac{1}{4}(A+B+1)^3$. Intuitively, the myopic solution in the numerator is worst when A and B have the maximum difference; hence, we let $A=B+1$. We obtain $\frac{\left(B+\frac{3}{2}\right)^3+\left(B+\frac{1}{2}\right)^3}{\frac{1}{4}(2B+2)^3}$. This fraction is decreasing in B, and hence we set B to 0. Thus, the maximum value of the fraction is 1.75. Combining Cases (i) and (ii), the maximum error is 86.6%.

We provide a detailed proof below. The minimization problem in the denominator is

$$\min_{0\leq C\leq 1} \mathcal{G}(C) \triangleq (A+C)^3+(B-C+1)^3.$$

Moreover,

$$\begin{aligned}\frac{\partial}{\partial C}\mathcal{G}(C) &= 3(A+C)^2-3(B-C+1)^2, \\ \frac{\partial^2}{\partial C^2}\mathcal{G}(C) &= 6(A+B+1)>0,\end{aligned}$$

meaning $\mathcal{G}(C)$ is strictly convex. Using $\frac{\partial}{\partial C}\mathcal{G}(C)=0$, we obtain $C^*=\frac{1}{2}(B-A+1)$. Consider the following two cases.

Case (i): $C^*>1$ or $C^*<0$. If $C^*>1$, then $C=1$ is optimal and the optimal value of the denominator is

$\mathcal{G}(1) = (A+1)^3 + B^3$. In other words, if $B > A+1$, assigning the customer's load to $\hat{t}$ is optimal (because the existing load in this period is much lower than that in period $\hat{t}+1$). In this case, we have

$$\rho_{\mathrm{T}} \leq \max_{\substack{A,\ B\geq 0,\\ B>A+1}} \left\{ \frac{\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3}{(A+1)^3+B^3} \right\} - 1. \tag{19}$$

On the other hand, if $C^* < 0$, then $C = 0$ is optimal and the optimal value of the denominator is $\mathcal{G}(0) = A^3 + (B+1)^3$. This scenario results in an inequality that is identical to (19) (except one needs to switch A and B). Thus, consider inequality (19) and note

$$\frac{\partial}{\partial A} \frac{\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3}{(A+1)^3+B^3} = \frac{3\left(A+\frac{1}{2}\right)^2 \left((A+1)^3+B^3\right) - 3(A+1)^2 \left(\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3\right)}{\left((A+1)^3+B^3\right)^2}.$$

Moreover, because of inequality (17) (let $y' = A+\frac{1}{2}$, $\alpha = \frac{1}{2}$, and $y'' = B$), we have

$$\begin{aligned}
& (A+1)^3 + B^3 < \left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3, \quad \text{and} \\
& 3\left(A+\frac{1}{2}\right)^2 < 3(A+1)^2 \\
\Rightarrow\ & 3\left(A+\frac{1}{2}\right)^2 \left((A+1)^3+B^3\right) < 3(A+1)^2 \left(\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3\right) \\
\Rightarrow\ & \frac{\partial}{\partial A} \frac{\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3}{(A+1)^3+B^3} < 0,
\end{aligned}$$

meaning the fraction is always decreasing in A (for any value of B). Thus, setting A to its smallest value 0 is optimal. In this case, inequality (19) reduces to

$$\rho_{\mathrm{T}} \leq \max_{B>1} \left\{ \frac{\left(\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3}{1+B^3} \right\} - 1.$$

The maximum value of the fraction is 1.866, which occurs at $B = 1.366$. This result implies $\rho_{\mathrm{T}} \leq 86.6\%$. Note this case happens if $Q = 1$ and the existing load in periods $\hat{t}$ and $\hat{t}+1$ is respectively 0 and 1.366, meaning this worst-case error bound is tight.

Case (ii): $0 \leq C^* \leq 1$. In other words, $0 \leq \frac{1}{2}(B-A+1) \leq 1$, implying $A \leq B+1$ and $B \leq A+1$. Then,

$$\rho_{\mathrm{T}} \leq \max_{\substack{A,\ B\geq 0,\\ A\leq B+1,\\ B\leq A+1}} \left\{ \frac{\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3}{\frac{1}{4}(A+B+1)^3} \right\} - 1.$$

Observe by switching A and B, the right-hand side remains identical. Thus, without loss of generality, we restrict the feasible set of the maximization to $B \leq A$. Then, the feasible set becomes $0 \leq B \leq A \leq B+1$. Keeping B fixed,

$$\begin{aligned}
\frac{\partial}{\partial A} \frac{\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3}{\frac{1}{4}(A+B+1)^3} &= \frac{\frac{3}{4}\left(A+\frac{1}{2}\right)^2 (A+B+1)^3 - \frac{3}{4}(A+B+1)^2 \left(\left(A+\frac{1}{2}\right)^3 + \left(B+\frac{1}{2}\right)^3\right)}{\frac{1}{16}(A+B+1)^6} \\
&= \frac{\frac{3}{4}\left(A+\frac{1}{2}\right)^2 - \frac{3}{4}\left(\left(A+\frac{1}{2}\right)^2 - \left(A+\frac{1}{2}\right)\left(B+\frac{1}{2}\right) + \left(B+\frac{1}{2}\right)^2\right)}{\frac{1}{16}(A+B+1)^3}
\end{aligned}$$

$$= \frac{\frac{3}{4}\left(\left(A+\frac{1}{2}\right)\left(B+\frac{1}{2}\right)-\left(B+\frac{1}{2}\right)^2\right)}{\frac{1}{16}\left(A+B+1\right)^3} \geq 0.$$

To obtain the second line, we apply $a^3+b^3=(a+b)(a^2-ab+b^2)$ to the last term in the numerator of the right-hand side of the first line — that is, $\left(A+\frac{1}{2}\right)^3+\left(B+\frac{1}{2}\right)^3$. The last fraction is non-negative because $A \geq B$. Thus, if B is fixed, the fraction is non-decreasing in A; hence, setting A to its largest value $B+1$ would be optimal. Therefore,

$$\rho_{\mathrm{T}} \leq \max_{B\geq 0}\left\{\frac{\left(B+\frac{3}{2}\right)^3+\left(B+\frac{1}{2}\right)^3}{\frac{1}{4}(2B+2)^3}\right\}-1,$$

and

$$\begin{aligned}
\frac{\partial}{\partial B}\frac{\left(B+\frac{3}{2}\right)^3+\left(B+\frac{1}{2}\right)^3}{\frac{1}{4}(2B+2)^3} &= \frac{\frac{3}{4}\left(\left(B+\frac{3}{2}\right)^2+\left(B+\frac{1}{2}\right)^2\right)(2B+2)^3-\frac{6}{4}(2B+2)^2\left(\left(B+\frac{3}{2}\right)^3+\left(B+\frac{1}{2}\right)^3\right)}{\frac{1}{16}\left(2B+2\right)^6} \\
&= \frac{\frac{3}{4}\left(\left(B+\frac{3}{2}\right)^2+\left(B+\frac{1}{2}\right)^2\right)-\frac{6}{4}\left(\left(B+\frac{3}{2}\right)^2-\left(B+\frac{3}{2}\right)\left(B+\frac{1}{2}\right)+\left(B+\frac{1}{2}\right)^2\right)}{\frac{1}{16}\left(2B+2\right)^3} \\
&= \frac{-\frac{3}{4}\left(B+\frac{3}{2}\right)^2+\frac{6}{4}\left(B+\frac{3}{2}\right)\left(B+\frac{1}{2}\right)-\frac{3}{4}\left(B+\frac{1}{2}\right)^2}{\frac{1}{16}\left(2B+2\right)^3} \\
&= \frac{-\frac{3}{4}\left(\left(B+\frac{3}{2}\right)-\left(B+\frac{1}{2}\right)\right)^2}{\frac{1}{16}\left(2B+2\right)^3} \\
&= \frac{-12}{\left(2B+2\right)^3} < 0.
\end{aligned}$$

Once again, to obtain the second line, we apply $a^3+b^3=(a+b)(a^2-ab+b^2)$ to the last term in the numerator of the right-hand side of the first line. Thus, the fraction is decreasing in B, and hence, setting B to its smallest value 0 is optimal. In summary, an optimal solution for Case (ii) is $B=0$ and $A=1$ (or $B=1$ and $A=0$), and the optimal value is 1.75. Thus, in Case (ii), $\rho_{\mathrm{T}} \leq 75\%$. In short, using Cases (i) and (ii), we have $\rho_{\mathrm{T}} \leq 86.6\%$. Furthermore, this bound is tight. Hence, the proof is complete.

Proof of Theorem 3.

Note $\{i\}$ denotes the set of all substations. The horizon $\{1,\ldots,T\}$ decomposes into mutually exclusive and collectively exhaustive intervals, where ς^{th} interval is $\{\dot{t}_\varsigma,\ldots,\ddot{t}_\varsigma\}$. We use index ς to refer to these intervals; $\varsigma=1$ refers to the first interval that contains the current period, that is, $\dot{t}_1=1$; and $\dot{t}_{\varsigma+1}=\ddot{t}_\varsigma+1$, for all ς (except the last one). Correspondingly, we decompose $\mathbb{J}_s$ into mutually exclusive and collectively exhaustive subsets $\mathbb{J}_{s,\varsigma}$. For participant $j \in \mathbb{J}_{s,\varsigma}$, we have $\dot{t}_\varsigma \leq a_j \leq a_j+\ell_j-1 \leq \ddot{t}_\varsigma$ (also because $\ell_j \geq 1$). Therefore, following such a decomposition, LM can equivalently be stated as

$$\begin{aligned}
\text{DTLM:}\quad \min\ & \sum_{\varsigma}\sum_{s} p_s \sum_{t=\dot{t}_\varsigma}^{\ddot{t}_\varsigma} f(y_{s,t}) \\
\text{s.t.}\ & \sum_{t=a_j}^{a_j+\ell_j-1} q_{s,j,t} = Q_j, \quad \forall s,\ \varsigma,\ j \in \mathbb{J}_{s,\varsigma}, \\
& 0 \leq q_{s,j,t} \leq q_j^U, \quad \forall s,\ \varsigma,\ t \in \{\dot{t}_\varsigma,\ldots,\ddot{t}_\varsigma\},\ j \in \mathbb{J}_{s,\varsigma}, \\
& q_{s,j,1} = q_{s',j,1}, \quad \forall s \neq s',\ j \in \mathbb{J}_{s,1} \cap \mathbb{J}_{s',1},
\end{aligned}$$

$$x_{s,g,t} = \lambda_{s,g,t} + \sum_{j \in \mathbb{J}_{s,\varsigma}:\ g_j = g} q_{s,j,t}, \quad \forall s,\ g,\ \varsigma,\ t \in \{\dot{t}_\varsigma, \ldots, \ddot{t}_\varsigma\},$$
$$\sum_{g \in i} x_{s,g,t} \le h_i, \quad \forall s,\ \varsigma,\ t \in \{\dot{t}_\varsigma, \ldots, \ddot{t}_\varsigma\},\ i,$$
$$y_{s,t} = \sum_{g} x_{s,g,t}, \quad \forall s,\ \varsigma,\ t \in \{\dot{t}_\varsigma, \ldots, \ddot{t}_\varsigma\}.$$

Note we drop constraints (7)-(10) because $|\mathcal{T}_{\varphi,d}| = 1$, for all $\varphi,\ d$ (the values of $\lambda_{s,g,t}$ are pre-computed and not impacted by truncation). Also, for each $j \in \mathbb{J}_{s,\varsigma}$, we discarded the decision variable $q_{s,j,t}$ if $t \notin \{\dot{t}_\varsigma, \ldots, \ddot{t}_\varsigma\}$, because its value is 0 in the optimal solution. Moreover, the third constraint is written only for the first period of the first interval $\varsigma = 1$. In this reformulation of LM, the objective function is additive in ς and all constraints are written separately for different ς's. This observation implies LM decomposes into subproblems (one subproblem for each non-overlapping interval ς) such that an optimal solution for LM is created by combining the optimal solutions of these subproblems. Consider the first subproblem (corresponding to $\varsigma = 1$). The horizon for this subproblem is $\{\dot{t}_1, \ldots, \ddot{t}_1\}$, where $\dot{t}_1 = 1$. Because of the above decomposition, the decisions corresponding to the first period (which are the ones that are implemented) are not affected by the other subproblems. Therefore, solving the first subproblem is equivalent to solving LM. Now, consider solving TLM($\breve{T}$) for some $\breve{T} \ge T_{\text{NOI}} \ge \ddot{t}_1 - \dot{t}_1 + 1$; then, TLM($\breve{T}$) decomposes into one or more subproblems such that the first-period decisions depend on only the first subproblem. Thus, we proved solving LM is equivalent to solving TLM($\breve{T}$) for $\breve{T} \ge T_{\text{NOI}}$. This proof establishes the first term in the definition of T^*, which is solely a result of assuming the non-overlapping intervals (NOI).

In the remainder, we further impose the following assumption: $\sum_{g \in i'} (\lambda_{s,g,t} - \lambda_{s,g,t+1}) \ge (J+2)q^\diamond$, for all $s,\ i' \in \{i\} \cup \{1, \ldots, G\}$, and $t \in \{\dot{t}, \ldots, \ddot{t} - 1\}$. This assumption implies a sharp reduction in consumption over the non-overlapping intervals. We refer to this assumption as SRC. First, we establish the following property.

Optimality of Delaying Supply (ODS). For each participant $j \in \mathbb{J}_{s,\varsigma}$, for arbitrary s and ς, delaying her supply to the last $\hat{T}_j \triangleq \lceil \gamma_j \rceil$ periods of her charging window is optimal.

To prove ODS, consider an arbitrary NOI interval $\{\dot{t}_\varsigma, \ldots, \ddot{t}_\varsigma\}$, and let $j \in \mathbb{J}_{s,\varsigma}$. Let her charging window $\{a_j, \ldots, a_j + \ell_j - 1\}$ consist of two parts: (i) $\mathbb{T}'_j$: the first $\ell_j - \lceil \gamma_j \rceil$ periods and (ii) $\mathbb{T}''_j$: the last $\lceil \gamma_j \rceil$ periods. Because $\ell_j \ge \lceil \gamma_j \rceil$, the first part $\mathbb{T}'_j$ may be empty, and the second part $\mathbb{T}''_j$ always consists of $\lceil \gamma_j \rceil$ periods. If $\ell_j = \lceil \gamma_j \rceil$, then $\mathbb{T}'_j$ is empty, meaning in an optimal solution, her load is assigned to $\mathbb{T}''_j$, completing our proof; thus, assume $\mathbb{T}'_j$ is nonempty. Let $q^*_{s,j,t}$, $x^*_{s,g,t}$, and $y^*_{s,t}$ denote an optimal solution. Assume to the contrary that in this optimal solution, $q^*_{s,j,t'} > 0$ for some $t' \in \mathbb{T}'_j$. Then, $t'' \in \mathbb{T}''_j$ exists whereby $q^*_{s,j,t''} < q^U_j$. Define $e \triangleq \min\{q^*_{s,j,t'}, q^U_j - q^*_{s,j,t''}\}$. We introduce a new solution as follows: $q^\circ_{s,j,t'} = q^*_{s,j,t'} - e$, $q^\circ_{s,j,t''} = q^*_{s,j,t''} + e$, and all other variables $q^\circ_{s,j,t}$ are equal to those in the optimal solution $q^*_{s,j,t}$. Then, $x^\circ_{s,g_j,t'} = x^*_{s,g_j,t'} - e$ and $x^\circ_{s,g_j,t''} = x^*_{s,g_j,t''} + e$. Moreover, $y^\circ_{s,t'} = y^*_{s,t'} - e$ and $y^\circ_{s,t''} = y^*_{s,t''} + e$. Consider problem DTLM that we introduced above. The first two constraints are feasible. The third constraint is feasible if $t' \ne 1$; otherwise, we repeat the above modification for all scenarios (in this case, $e \triangleq \min_s e_s$, where e_s is defined as we did above for scenario s). Without loss of generality, assume $t' \ne 1$. The fourth and sixth constraints are feasible too. For the fifth constraint, if $g_j \in i$, we have

$$\begin{aligned}
\sum_{g \in i} x^\circ_{s,g,t''} &= \sum_{g \in i} \lambda_{s,g,t''} + \sum_{j \in \mathbb{J}_{s,\varsigma}:\ g_j \in i} q^\circ_{s,j,t''} \\
&\le -(J+2)q^\diamond + \sum_{g \in i} \lambda_{s,g,t'} + \sum_{j \in \mathbb{J}_{s,\varsigma}:\ g_j \in i} q^*_{s,j,t''} + e \\
&= -(J+2)q^\diamond + \sum_{g \in i} \lambda_{s,g,t'} + \sum_{j \in \mathbb{J}_{s,\varsigma}:\ g_j \in i} q^*_{s,j,t''} + e \pm \sum_{j \in \mathbb{J}_{s,\varsigma}:\ g_j \in i} q^*_{s,j,t'}
\end{aligned}$$

$$
\begin{aligned}
&= -(J+2)q^\diamond + \sum_{g\in i} x^*_{s,g,t'} + \sum_{j\in\mathbb{J}_{s,\varsigma}:\ g_j\in i} q^*_{s,j,t''} + e - \sum_{j\in\mathbb{J}_{s,\varsigma}:\ g_j\in i} q^*_{s,j,t'} \\
&\le -(J+2)q^\diamond + h_i + \sum_{j\in\mathbb{J}_{s,\varsigma}:\ g_j\in i} \left(q^*_{s,j,t''} - q^*_{s,j,t'}\right) + e \\
&\le -(J+2)q^\diamond + h_i + (J+1)q^\diamond \\
&\le h_i.
\end{aligned}
$$

To obtain the second line, we use $\sum_{g\in i}\lambda_{s,g,t''} \le \sum_{g\in i}\lambda_{s,g,t'} - (J+2)q^\diamond$, because of SRC and $t''>t'$. To obtain the third line, we add and subtract the last term. To obtain the fifth line, we use $\sum_{g\in i} x^*_{s,g,t'} \le h_i$. In the fifth line, we note

$$
\sum_{j\in\mathbb{J}_{s,\varsigma}:\ g_j\in i} \left(q^*_{s,j,t''} - q^*_{s,j,t'}\right) \le \sum_{j\in\mathbb{J}_{s,\varsigma}:\ g_j\in i} q^\diamond \le |\mathbb{J}_{s,\varsigma}|q^\diamond \le Jq^\diamond,
$$

and $e\le q^\diamond$, which leads to the sixth line. Thus, our solution satisfies the grid constraints. In short, we have created a feasible solution for DTLM. In this solution,

$$
\begin{aligned}
y^*_{s,t''} + e &= \sum_g \lambda_{s,g,t''} + \sum_{j\in\mathbb{J}_{s,\varsigma}} q^*_{s,j,t''} + e \\
&\le -(J+2)q^\diamond + \sum_g \lambda_{s,g,t'} + \sum_{j\in\mathbb{J}_{s,\varsigma}} q^*_{s,j,t''} \pm \sum_{j\in\mathbb{J}_{s,\varsigma}} q^*_{s,j,t'} + e \\
&= -(J+2)q^\diamond + y^*_{s,t'} + \sum_{j\in\mathbb{J}_{s,\varsigma}} (q^*_{s,j,t''} - q^*_{s,j,t'}) + e \\
&\le -(J+2)q^\diamond + y^*_{s,t'} + (J+1)q^\diamond \\
&< y^*_{s,t'},
\end{aligned}
$$

or $y^*_{s,t''} < y^*_{s,t'} - e$. To obtain the last line, we note $q^\diamond > 0$.

Last, we compare the objective values of the two solutions.

$$
\begin{aligned}
\sum_\varsigma \sum_s p_s \sum_{t=\dot{\imath}_\varsigma}^{\ddot{\imath}_\varsigma} f(y^\circ_{s,t}) - \sum_\varsigma \sum_s p_s \sum_{t=\dot{\imath}_\varsigma}^{\ddot{\imath}_\varsigma} f(y^*_{s,t}) &= p_s \left(f(y^\circ_{s,t'}) + f(y^\circ_{s,t''}) - f(y^*_{s,t'}) - f(y^*_{s,t''})\right) \\
&= p_s \left(f(y^*_{s,t'} - e) - f(y^*_{s,t'}) + f(y^*_{s,t''} + e) - f(y^*_{s,t''})\right) < 0.
\end{aligned}
$$

To see why the second line is negative, note $p_s>0$, and in inequality (17), set $y' = y^*_{s,t''}$, $y'' = y^*_{s,t'} - e$, and $\alpha = e$. Thus, we have created a feasible solution that is strictly better than the optimal solution, which is a contradiction; hence, our proof of ODS is complete.

So far, we have shown that for each participant $j\in\mathbb{J}_{s,\varsigma}$, for arbitrary s and ς, delaying her supply to the last $\hat{T}_j \triangleq \lceil\gamma_j\rceil$ periods of her charging window is optimal. Thus, the length of the truncated horizon must be long enough that no load of participant j is assigned to the first period unless the first period is one of her last $\hat{T}_j$ periods in her charging window.

If a participant arrives at $t>1$, her load supply is not assigned to the first period. Thus, consider a participant with an arrival time $a_j = 1$. Recalling $\ell_j \ge \lceil\gamma_j\rceil$, exactly one of the following two cases happens.

Case 1: $\ell_j \ge \lceil\gamma_j\rceil + 1$. In this case, no supply must be assigned to the first period. If $T^* \ge \ell_j$, because of ODS, no supply is assigned to the first period. If $T^* < \ell_j$, we must have

$$
\frac{T^*}{\ell_j} Q_j \le (T^* - 1) q^U_j.
$$

The left-hand side is the fraction of participant j's load that must be satisfied in periods $1,\dots,T^*$, and the right-hand side is the maximum load that can be supplied in periods $2,\dots,T^*$. If this inequality holds, because of ODS, no supply is assigned to the first period. The left-hand side of this inequality grows as ℓ_j decreases; hence, the worst case for this inequality occurs if $\ell_j = \lceil \gamma_j \rceil + 1$. Therefore, we must have

$$\begin{aligned}
&\frac{T^*}{\lceil \gamma_j \rceil + 1}\gamma_j \le T^* - 1, \quad \forall j \\
\Leftrightarrow\ & T^*\left(1 - \frac{\gamma_j}{\lceil \gamma_j \rceil + 1}\right) \ge 1, \quad \forall j \\
\Leftarrow\ & T^* \ge \left\lceil \frac{\lceil \gamma_j \rceil + 1}{\lceil \gamma_j \rceil + 1 - \gamma_j} \right\rceil, \quad \forall j \\
\Leftarrow\ & T^* \ge \max_j \left\lceil \frac{\lceil \gamma_j \rceil + 1}{\lceil \gamma_j \rceil + 1 - \gamma_j} \right\rceil.
\end{aligned}$$

If the last inequality holds, no supply of participant j is assigned to any time prior to the last $\lceil \gamma_j \rceil$ periods in her charging window.

Case 2: $\ell_j = \lceil \gamma_j \rceil$. In this case, if γ_j is integral, regardless of the value of T^*, the EV will be charged at a full speed in the first period (similar to the uniform supply property that we showed earlier). If γ_j is fractional, only the fractional part of γ_j must be assigned to the first period; in this case, we show T^* must be at least $\lceil \gamma_j \rceil$. The fractional part of γ_j is $1 + \gamma_j - \lceil \gamma_j \rceil$. We must have

$$\begin{aligned}
&\frac{T^*}{\ell_j} Q_j \le (T^* - 1) q_j^U + (1 + \gamma_j - \lceil \gamma_j \rceil) q_j^U, \quad \forall j \\
\Leftrightarrow\ & \frac{T^*}{\lceil \gamma_j \rceil}\gamma_j \le (T^* - 1) + (1 + \gamma_j - \lceil \gamma_j \rceil), \quad \forall j \\
\Leftrightarrow\ & \frac{\gamma_j}{\lceil \gamma_j \rceil} T^* \le T^* + (\gamma_j - \lceil \gamma_j \rceil), \quad \forall j \\
\Leftrightarrow\ & \frac{\gamma_j - \lceil \gamma_j \rceil}{\lceil \gamma_j \rceil} T^* \le \gamma_j - \lceil \gamma_j \rceil, \quad \forall j \\
\Leftarrow\ & T^* \ge \lceil \gamma_j \rceil, \quad \forall j \\
\Leftarrow\ & T^* \ge \max_j \lceil \gamma_j \rceil.
\end{aligned}$$

Note we drop participant j if $\gamma_j = 0$, which is optimal, because the utility does not supply any load to this participant; thus, assume $\gamma_j > 0$ for all j. If γ_j is integral, the first line holds regardless of the value of T^*. Otherwise, if γ_j is not integral, the first line holds if $T^* \ge \max_j \lceil \gamma_j \rceil$. Thus, $T^* \ge \ell_j$, implying the charging window of participant j is completely inside the truncated horizon of T^* periods. Therefore, because of ODS, we assign her total requirement to the first ℓ_j periods. Because of ODS, we charge her EV at the full charging speed in periods $2,\dots,\ell_j$, and the leftover is assigned to the first period.

Therefore, if $T^* \ge \max_j \left\{ \lceil \gamma_j \rceil, \left\lceil \frac{\lceil \gamma_j \rceil + 1}{\lceil \gamma_j \rceil + 1 - \gamma_j} \right\rceil \right\}$, because of ODS, each participant's requirement is satisfied in the latest periods of her charging window. This solution is optimal, and hence, the proof is complete.

Proof of Theorem 4(a). We first show

$$f_t(\delta b) = \sum_{z=1}^{Z} \sigma_{t,z} \left[b - z + 1 \right]^+, \quad \forall b \in \{0, 1, 2, \dots, Z\}, \tag{20}$$

where $[x]^+ \triangleq \max\{0, x\}$. Fattahi et al. (2023) prove a similar equation holds for the case of integral load values; we do not impose such a restriction. Recall $f_t(y) = 0$ for $y \le 0$, and hence, we have $\sigma_{t,1} \triangleq f_t(\delta)$,

$\sigma_{t,2} \triangleq f_t(2\delta) - 2f_t(\delta)$, and $\sigma_{t,z} \triangleq f_t(z\delta) - 2f_t((z-1)\delta) + f_t((z-2)\delta)$, for all $z \geq 3$. If $b = 0$, equation (20) holds; hence, assume $b \in \{1, \ldots, Z\}$. Then,

$$\begin{aligned}
\sum_{z=1}^{Z} \sigma_{t,z} \left[b - z + 1\right]^+ &= \sum_{z=1}^{b} \sigma_{t,z} \left(b - z + 1\right) \\
&= \left(\sum_{z=1}^{b} f_t(z\delta) \left(b - z + 1\right)\right) - 2\left(\sum_{z=2}^{b} f_t((z-1)\delta) \left(b - z + 1\right)\right) + \left(\sum_{z=3}^{b} f_t((z-2)\delta) \left(b - z + 1\right)\right) \\
&= \left(\sum_{z=1}^{b} f_t(z\delta) \left(b - z + 1\right)\right) - 2\left(\sum_{z=1}^{b-1} f_t(z\delta) \left(b - z\right)\right) + \left(\sum_{z=1}^{b-2} f_t(z\delta) \left(b - z - 1\right)\right) \\
&= 2f_t((b-1)\delta) + f_t(b\delta) - 2f_t((b-1)\delta) \\
&= f_t(b\delta).
\end{aligned}$$

To obtain the fourth line, note the three summations in the third line cancel out for all $z \in \{1, \ldots, b-2\}$. Thus, the first two terms in the fourth line are the terms of the first summation in the third line corresponding to $z = b-1$ and $z = b$, respectively. The last term in the fourth line is the term corresponding to $z = b-1$ in the second summation in the third line.

Using equation (20), the objective function (2) is

$$\begin{aligned}
&\min \ \sum_s p_s \sum_{t=1}^{T} f_t(y_{s,t}) \\
&\leq \ \min \ \sum_s p_s \sum_{t=1}^{T} f_t(\delta y^{\text{int}}_{s,t}) \quad \text{s.t.} \quad y_{s,t} \leq \delta y^{\text{int}}_{s,t} \leq y_{s,t} + \delta, \ y^{\text{int}}_{s,t} \in \mathbb{Z}_+, \quad \forall s, \ t \\
&= \ \min \ \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \left[y^{\text{int}}_{s,t} - z + 1\right]^+ \quad \text{s.t.} \quad y_{s,t} \leq \delta y^{\text{int}}_{s,t} \leq y_{s,t} + \delta, \ y^{\text{int}}_{s,t} \in \mathbb{Z}_+, \quad \forall s, \ t \\
&= \ \min \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \varepsilon_{s,t,z} \\
&\quad \text{s.t.} \quad 0 \leq \varepsilon_{s,t,z} - \left(y^{\text{int}}_{s,t} - z + 1\right) \leq Z\left(1 - \upsilon_{s,t,z}\right), \quad \forall s, \ t, \ z, \\
&\qquad\quad 0 \leq \varepsilon_{s,t,z} \leq Z\upsilon_{s,t,z}, \ \upsilon_{s,t,z} \in \{0,1\}, \quad \forall s, \ t, \ z, \\
&\qquad\quad y_{s,t} \leq \delta y^{\text{int}}_{s,t} \leq y_{s,t} + \delta, \quad y^{\text{int}}_{s,t} \in \mathbb{Z}_+, \quad \forall s, \ t.
\end{aligned}$$

The second line holds because $f_t(\cdot)$ is non-decreasing in the hourly load. Note that, in the second line, $y^{\text{int}}_{s,t} \leq \frac{1}{\delta} y_{s,t} + 1 \leq \frac{1}{\delta}\lambda^{\text{U}} + 1 \leq \left\lceil \frac{1}{\delta}\lambda^{\text{U}} + 1 \right\rceil = Z$. Thus, the third line follows directly from equation (20). To obtain the last line, we introduce continuous variables $\varepsilon_{s,t,z} = \left[y^{\text{int}}_{s,t} - z + 1\right]^+$. The first two constraints ensure $\varepsilon_{s,t,z} = \left[y^{\text{int}}_{s,t} - z + 1\right]^+$. If $\left(y^{\text{int}}_{s,t} - z + 1\right) > 0$, then $\upsilon_{s,t,z} = 1$; if $\left(y^{\text{int}}_{s,t} - z + 1\right) < 0$, then $\upsilon_{s,t,z} = 0$; and if $\left(y^{\text{int}}_{s,t} - z + 1\right) = 0$, then $\upsilon_{s,t,z}$ could take either value. Therefore, LLM is an upper bound for LM. Moreover, because LLM includes constraints (3)-(12), its optimal solution is feasible to LM.

Proof of Theorem 4(b). Let $y^{\circ}_{s,t}$, $q^{\circ}_{s,j,t}$, $x^{\circ}_{s,g,t}$, and $\vartheta^{\circ}_{s,t,\varphi,d}$ denote an optimal solution to LM. We use this optimal solution to create a feasible solution for LLM as follows. We keep the values for $y^{\circ}_{s,t}$, $q^{\circ}_{s,j,t}$, $x^{\circ}_{s,g,t}$, and $\vartheta^{\circ}_{s,t,\varphi,d}$, and hence, constraints (3)-(12) will be satisfied. Let $y^{\text{int}}_{s,t}$ be the smallest integer number that satisfies $\delta y^{\text{int}}_{s,t} \geq y^{\circ}_{s,t}$ for all s and t. Define $\varepsilon_{s,t,z} \triangleq \left[y^{\text{int}}_{s,t} - z + 1\right]^+$ and $\upsilon_{s,t,z} \triangleq \mathbb{I}(\varepsilon_{s,t,z} > 0)$ for all s, t, and z. This solution is feasible to LLM, and hence, its objective value is an upper bound on the optimal value of LLM.

We have $\delta y^{\text{int}}_{s,t} \le y^{\circ}_{s,t} + \delta$ for all s and t, and hence, the absolute error of linearization is upper bounded as

$$\begin{aligned} v_{\text{LLM}} - v_{\text{LM}} &\le \sum_s p_s \sum_t f_t(y^{\circ}_{s,t} + \delta) - \sum_s p_s \sum_t f_t(y^{\circ}_{s,t}) \\ &\le \sum_s p_s \sum_t \max_{s,t} \left\{ \max_{\lambda^{\text{L}}_{s,t} \le y \le \lambda^{\text{U}}_{s,t} + \check{J} q^{\diamond}} f_t(y+\delta) - f_t(y) \right\} \\ &= \sum_s p_s \sum_t \Theta(\delta) \\ &= T\Theta(\delta), \end{aligned}$$

completing the proof.

Proof of Theorem 5. The LP relaxation of LLM is equivalently written as

$$\begin{aligned} \text{RLLM}: \quad \min \ & \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \varepsilon_{s,t,z} \\ \text{s.t.} \ & \max\{0, \left(y^{\text{int}}_{s,t} - z + 1\right)\} \le \varepsilon_{s,t,z} \\ & \qquad \le \min\{Z\upsilon_{s,t,z}, \ Z\left(1 - \upsilon_{s,t,z}\right) + \left(y^{\text{int}}_{s,t} - z + 1\right)\}, \quad \forall s, \ t, \ z, \\ & y_{s,t} \le \delta y^{\text{int}}_{s,t} \le y_{s,t} + \delta, \ \ y^{\text{int}}_{s,t} \ge 0, \quad \forall s, \ t, \\ & 0 \le \upsilon_{s,t,z} \le 1, \quad \forall s, \ t, \ z, \\ & \text{and constraints (3)-(12).} \end{aligned}$$

Let $y^{\circ}_{s,t}$ denote an optimal solution for RLLM. If we fix $y_{s,t} = y^{\circ}_{s,t}$, the resulting problem is

$$\begin{aligned} \text{RLLM}': \quad \min \ & \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \varepsilon_{s,t,z} \\ \text{s.t.} \ & \max\{0, \left(y^{\text{int}}_{s,t} - z + 1\right)\} \le \varepsilon_{s,t,z} \\ & \qquad \le \min\{Z\upsilon_{s,t,z}, \ Z\left(1 - \upsilon_{s,t,z}\right) + \left(y^{\text{int}}_{s,t} - z + 1\right)\}, \quad \forall s, \ t, \ z, \\ & y^{\circ}_{s,t} \le \delta y^{\text{int}}_{s,t} \le y^{\circ}_{s,t} + \delta, \ \ y^{\text{int}}_{s,t} \ge 0, \quad \forall s, \ t, \\ & 0 \le \upsilon_{s,t,z} \le 1, \quad \forall s, \ t, \ z. \end{aligned}$$

The optimal values of RLLM and RLLM′ are equal. The following problem provides a lower bound on the optimal value of RLLM.

$$\begin{aligned} \text{LRLLM}: \quad \min \ & \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \varepsilon_{s,t,z} \\ \text{s.t.} \ & \min_{\frac{1}{\delta} y^{\circ}_{s,t} \le y \le \frac{1}{\delta} y^{\circ}_{s,t} + 1} \{\max\{0, (y - z + 1)\}\} \le \varepsilon_{s,t,z} \\ & \qquad \le \max_{\substack{\frac{1}{\delta} y^{\circ}_{s,t} \le y \le \frac{1}{\delta} y^{\circ}_{s,t} + 1 \\ \upsilon \in \mathbb{R}}} \{\min\{Z\upsilon, \ Z\left(1 - \upsilon\right) + (y - z + 1)\}\}, \quad \forall s, \ t, \ z, \\ & y^{\circ}_{s,t} \le \delta y^{\text{int}}_{s,t} \le y^{\circ}_{s,t} + \delta, \ \ y^{\text{int}}_{s,t} \ge 0, \quad \forall s, \ t, \\ & 0 \le \upsilon_{s,t,z} \le 1, \quad \forall s, \ t, \ z. \end{aligned}$$

The second and third constraints in LRLLM are redundant; hence, we drop them in the rest of the proof. The resulting problem is

$$\text{LRLLM}': \quad \min \ \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \varepsilon_{s,t,z}$$

$$\text{s.t.} \quad \min_{\frac{1}{\delta}y^{\circ}_{s,t}\le y\le \frac{1}{\delta}y^{\circ}_{s,t}+1} \{\max\{0,(y-z+1)\}\} \le \varepsilon_{s,t,z} \le \max_{\substack{\frac{1}{\delta}y^{\circ}_{s,t}\le y\le \frac{1}{\delta}y^{\circ}_{s,t}+1 \\ \upsilon\in\mathbb{R}}} \{\min\{Z\upsilon,\ Z(1-\upsilon)+(y-z+1)\}\}, \quad \forall s,\ t,\ z.$$

The optimal value of LRLLM′ is equal to the optimal value of LRLLM. Observe LRLLM′ decomposes into some subproblems (one subproblem for each $s,\ t,\ z$). Each subproblem has one continuous variable $\varepsilon_{s,t,z}$, which is restricted to take a value within the specified limits. In an optimal solution, $\varepsilon_{s,t,z}$ will be at either its lower limit (if $\sigma_{t,z}>0$) or upper limit (if $\sigma_{t,z}<0$). If $\sigma_{t,z}=0$, $\varepsilon_{s,t,z}$ could take any value. Thus, an optimal solution for LRLLM′ is

$$\varepsilon^{\circ}_{s,t,z} \triangleq \begin{cases} \min_{\frac{1}{\delta}y^{\circ}_{s,t}\le y\le \frac{1}{\delta}y^{\circ}_{s,t}+1}[y-z+1]^{+} & \text{if } \sigma_{t,z}\ge 0, \\ \max_{\substack{\frac{1}{\delta}y^{\circ}_{s,t}\le y\le \frac{1}{\delta}y^{\circ}_{s,t}+1 \\ \upsilon\in\mathbb{R}}} \{\min\{Z\upsilon,\ Z(1-\upsilon)+(y-z+1)\}\} & \text{if } \sigma_{t,z}<0. \end{cases}$$

In the first case ($\sigma_{t,z}\ge 0$), the minimization assigns $\frac{1}{\delta}y^{\circ}_{s,t}$ to its decision variable y. In the second case ($\sigma_{t,z}<0$), considering the maximization problem, for any value of υ, it would be optimal for y to take its upper limit $\frac{1}{\delta}y^{\circ}_{s,t}+1$. Therefore,

$$\varepsilon^{\circ}_{s,t,z} \triangleq \begin{cases} [\frac{1}{\delta}y^{\circ}_{s,t}-z+1]^{+} & \text{if } \sigma_{t,z}\ge 0, \\ \max_{\upsilon\in\mathbb{R}}\left\{\min\{Z\upsilon,\ Z(1-\upsilon)+\left(\frac{1}{\delta}y^{\circ}_{s,t}-z+2\right)\}\right\} & \text{if } \sigma_{t,z}<0. \end{cases}$$

In the second case, both terms inside the minimization are linear in υ. The first term ($Z\upsilon$) increases in υ, and the second term ($Z(1-\upsilon)+\left(\frac{1}{\delta}y^{\circ}_{s,t}-z+2\right)$) decreases in υ. These two lines intersect at

$$\upsilon^{\circ}_{s,t,z} \triangleq \frac{1}{2}+\frac{1}{2Z}\left(\frac{1}{\delta}y^{\circ}_{s,t}-z+2\right).$$

Observe $\upsilon^{\circ}_{s,t,z}$ optimizes the maximization. Thus,

$$\varepsilon^{\circ}_{s,t,z} \triangleq \begin{cases} [\frac{1}{\delta}y^{\circ}_{s,t}-z+1]^{+} & \text{if } \sigma_{t,z}\ge 0, \\ \frac{1}{2}Z+\frac{1}{2}\left(\frac{1}{\delta}y^{\circ}_{s,t}-z+2\right) & \text{if } \sigma_{t,z}<0. \end{cases}$$

Therefore, we obtain a lower bound on the optimal value of the LP relaxation:

$$\begin{aligned} v_{\text{RLLM}} &\ge \sum_{s} p_s \sum_{t=1}^{T}\sum_{z=1}^{Z} \sigma_{t,z}\left\{[\frac{1}{\delta}y^{\circ}_{s,t}-z+1]^{+}\mathbb{I}(\sigma_{t,z}\ge 0)+\left(\frac{1}{2}Z+\frac{1}{2}\left(\frac{1}{\delta}y^{\circ}_{s,t}-z+2\right)\right)\mathbb{I}(\sigma_{t,z}<0)\right\} \\ &= \sum_{s} p_s \sum_{t=1}^{T}\sum_{z=1}^{Z}\left\{[\sigma_{t,z}]^{+}[\frac{1}{\delta}y^{\circ}_{s,t}-z+1]^{+}-\frac{1}{2}[-\sigma_{t,z}]^{+}\left(Z+\left(\frac{1}{\delta}y^{\circ}_{s,t}-z+2\right)\right)\right\}, \end{aligned}$$

noting $\sigma_{t,z}\mathbb{I}(\sigma_{t,z}\ge 0)=[\sigma_{t,z}]^{+}$ and $\sigma_{t,z}\mathbb{I}(\sigma_{t,z}<0)=-[-\sigma_{t,z}]^{+}$. To simplify our notations, let $B_{s,t,z}\triangleq\frac{1}{\delta}y^{\circ}_{s,t}-z+1$, for all $s,\ t,\ z$. Thus,

$$v_{\text{RLLM}} \ge \sum_{s} p_s \sum_{t=1}^{T}\sum_{z=1}^{Z}\left\{[\sigma_{t,z}]^{+}[B_{s,t,z}]^{+}-\frac{1}{2}[-\sigma_{t,z}]^{+}(Z+B_{s,t,z}+1)\right\}.$$

Next, we create an upper bound on the optimal value of LLM. Let $y^{\circ}_{s,t}$, $q^{\circ}_{s,j,t}$, $x^{\circ}_{s,g,t}$, and $\vartheta^{\circ}_{s,t,\varphi,d}$ denote an optimal solution to RLLM. We use this optimal solution to create a feasible solution for LLM as follows. We keep the values for $y^{\circ}_{s,t}$, $q^{\circ}_{s,j,t}$, $x^{\circ}_{s,g,t}$, and $\vartheta^{\circ}_{s,t,\varphi,d}$, and hence, constraints (3)-(12) will be satisfied (also because we assume $|\mathcal{T}_{\varphi,d}|=1$, for all φ,d). Let $y^{\text{int}}_{s,t}$ be the smallest integer number that satisfies $\delta y^{\text{int}}_{s,t}\ge y^{\circ}_{s,t}$

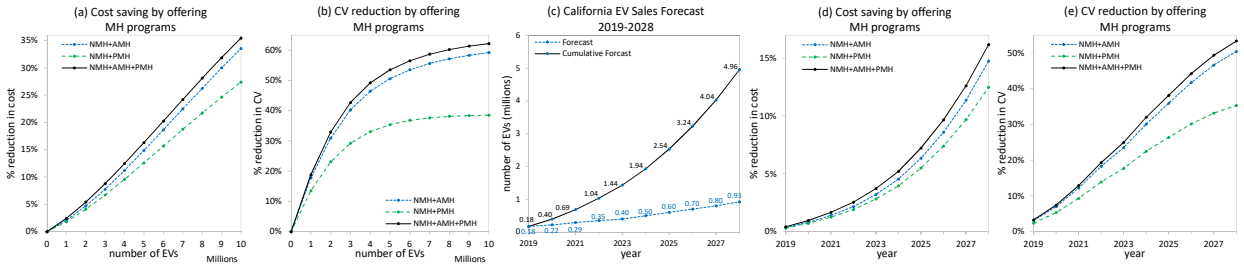

Figure 8 Potential cost savings and CV reductions for different offerings of MH programs.

for all s and t. Define $\varepsilon_{s,t,z} \triangleq \left[y^{\text{int}}_{s,t} - z + 1\right]^+$ and $v_{s,t,z} \triangleq \mathbb{I}(\varepsilon_{s,t,z} > 0)$ for all $s,\ t$, and z. Note this solution is feasible to LLM, and hence, its objective value is an upper bound on the optimal value of LLM. Moreover, because $y^{\text{int}}_{s,t} \le \frac{1}{\delta} y^{\circ}_{s,t} + 1$, we have

$$\begin{aligned} v_{\text{LLM}} &\le \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \sigma_{t,z} \left[\frac{1}{\delta} y^{\circ}_{s,t} - z + 2\right]^+ \\ &= \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \left([\sigma_{t,z}]^+ - [-\sigma_{t,z}]^+\right) \left[B_{s,t,z} + 1\right]^+, \end{aligned}$$

noting $\sigma_{t,z} = [\sigma_{t,z}]^+ - [-\sigma_{t,z}]^+$, for all t and z. Thus, by subtracting the upper bound of LLM minus the lower bound of RLLM, the absolute error of the LP relaxation of LLM is bounded as

$$\begin{aligned} v_{\text{LLM}} - v_{\text{RLLM}} &\le \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ \left([B_{s,t,z}+1]^+ - [B_{s,t,z}]^+\right) [\sigma_{t,z}]^+ - \frac{1}{2}[-\sigma_{t,z}]^+ \left(2\left[B_{s,t,z}+1\right]^+ - Z - B_{s,t,z} - 1\right) \right\} \\ &\le \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ [\sigma_{t,z}]^+ - \frac{1}{2}[-\sigma_{t,z}]^+ \left(2B_{s,t,z} + 2 - Z - B_{s,t,z} - 1\right) \right\} \\ &= \sum_s p_s \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ [\sigma_{t,z}]^+ + \frac{1}{2}[-\sigma_{t,z}]^+ \left(Z - B_{s,t,z} - 1\right) \right\} \\ &= \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ [\sigma_{t,z}]^+ + \frac{1}{2}[-\sigma_{t,z}]^+ \left(Z - \sum_s p_s B_{s,t,z} - 1\right) \right\} \\ &\le \sum_{t=1}^{T} \sum_{z=1}^{Z} \left\{ [\sigma_{t,z}]^+ + \frac{1}{2}[-\sigma_{t,z}]^+ \left(Z - \frac{1}{\delta}\lambda^{\text{L}} + z - 2\right) \right\}. \end{aligned}$$

To obtain the second line, we use $[B+1]^+ - [B]^+ \le 1$ and $B + 1 \le [B+1]^+$, for $B \in \mathbb{R}$. To obtain the last line, we note $B_{s,t,z} = \frac{1}{\delta} y^{\circ}_{s,t} - z + 1 \ge \frac{1}{\delta}\lambda^{\text{L}} - z + 1$, for all $s,\ t,\ z$, and $\sum_s p_s = 1$. Thus, the proof is complete.

Appendix D: Predicting Cost Saving and CV Reduction.

Our methodology provides detailed managerial insights on the amount and timing of the improvements achievable in cost and demand variability by offering AMH, PMH, and both. We illustrate using the following example. Figures 8(a) and (b), respectively, show the expected cost savings and CV reductions for offering AMH, PMH, and both, for different numbers of EVs in CAISO. The inputs for this figure are the same as Figure 3 and $\hat{\rho}^{\pi} = 1$, for all π. Figure 8(c) shows the predicted EV sales in California, which we use for illustration. We downloaded the EV sales forecast from EV-Adoption (2023) and computed the cumulative forecast using the sales forecast. Combining Figures 8(a) and (c), we produce Figure 8(d), which shows the

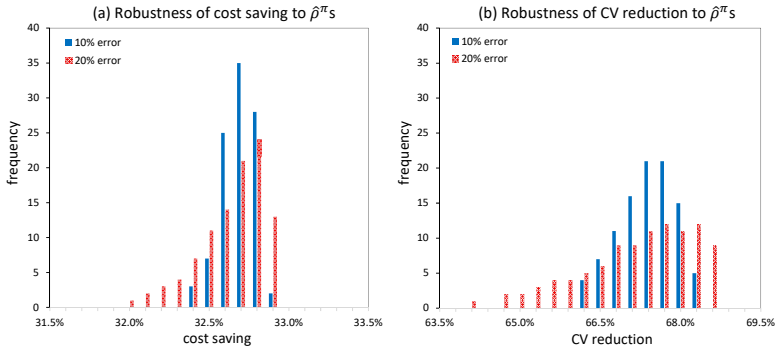

Figure 9 Robustness of cost saving and CV reduction

amount and timing of the improvements achievable in cost saving by different offerings of MH programs. Similarly, we produce Figure 8(e) by combining (b) and (c). For example, by 2028, California will have 4.96 million EVs and the utilities could achieve (i) 14.8% cost saving and 50.4% reduction in CV by offering AMH, (ii) 12.5% cost saving and 35.3% reduction in CV by offering PMH, and (iii) 16.2% cost saving and 53.3% reduction in CV by offering both AMH and PMH. In this illustrative example, offering both AMH and PMH optimizes the cost and CV. Not offering AMH leads to high participation in NMH and PMH, increasing cost and CV by, for example, 4.4% and 38.6%, respectively, in 2028. In short, our methodology produces a cost-saving and CV-reduction profile, across different years, for various offerings of MH programs. These detailed predictions enable utilities to make informed decisions on whether to launch MH programs, which ones, and when.

Similar to Figure 8, combined with a forecasting model, Figure 6 maps to a cost-saving and CV-reduction profile, across different years. For example, recalling the forecast of 4.96 million EVs in 2028, customizing PMH reduces the cost and CV by 0.95% and 10.11%, respectively, in 2028. These detailed predictions enable utilities to make informed decisions about whether and when to customize PMH.

Our predictions in Figure 8(d)-(e) depend on the accuracy of the EV forecast shown in Figure 8(c). Furthermore, for simplicity, we assume the existing load and cost function remain the same across years, and we do not account for the consumption of non-home charging EVs. Our methodology and results extend to incorporate the evolving factors such as the existing load, renewable power generation sources, energy procurement costs, and the non-home charging EVs. Additional analyses of these factors and their corresponding impacts would be of great value.

Appendix E: Robustness and Generalizations

We evaluate the robustness of cost saving and CV reduction with respect to $\hat{\rho}^{\mathrm{AMH}}$ and $\hat{\rho}^{\mathrm{PMH}}$. Figure 9 presents the histograms of cost savings and CV reductions for 100 combinations of values associated with $\hat{\rho}^{\mathrm{AMH}}$ and $\hat{\rho}^{\mathrm{PMH}}$, under two scenarios: 10% and 20% errors. We let $\beta = \$50$, $\mu^{\mathrm{L}} = 20$ ¢/KWh, and $\mathcal{K} = 10$ millions. For the case of 10% error, we let $\hat{\rho}^{\mathrm{AMH}},\ \hat{\rho}^{\mathrm{PMH}} \in \{0.91, 0.93, \ldots, 1.09\}$, which results in 100 combinations of values for $\hat{\rho}^{\mathrm{AMH}}$ and $\hat{\rho}^{\mathrm{PMH}}$. Similarly, for the case of 20% error, we let $\hat{\rho}^{\mathrm{AMH}},\ \hat{\rho}^{\mathrm{PMH}} \in \{0.82, 0.86, \ldots, 1.18\}$. As expected, the histograms for the 20% errors exhibit greater variability in the objective values. Interestingly, the cost saving's variability falls within the range of 32.0% to 32.9%, and the CV reduction's variability ranges from 64.1% to 68.6%. These ranges are quite reasonable, demonstrating the robustness of both cost saving and CV reduction concerning variations in the inputs for the attractiveness functions.

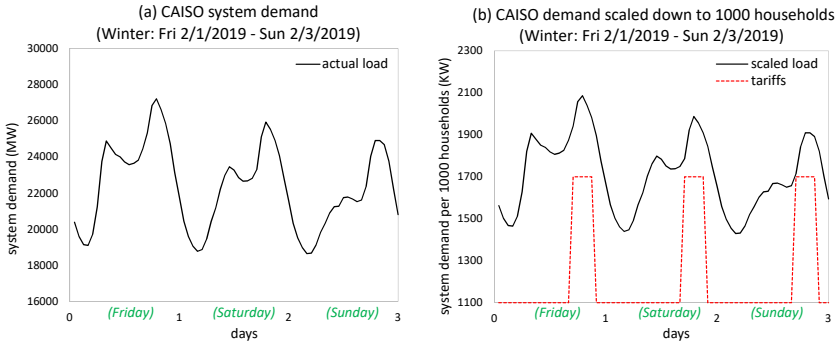

Figure 10 CAISO data scaled down to 1,000 households

Our numerical study of PD extends to multiple driver types and multiple day types. For example, one could consider three day types to represent the days in the season: (i) cold weekday, (ii) hot weekday, and (iii) weekend day. Fattahi et al. (2023) also use day types for a long-term forecast of energy-consumption profiles.

Appendix F: Numerical Experiments of LM

In this section, we investigate the effectiveness and near optimality of our approximation through extensive numerical experiments. We use real data from CAISO to simulate instances for our problem. The set-up of our numerical analyses is given in §F.1. We present the result of our experiments in §F.2. Our numerical results in this section indicate our approximation produces near-optimal solutions to the large instances of our problem in a reasonable amount of time.

F.1. Numerical-Analysis Setup

To generate an instance for our problem, we need to simulate the existing load, the low-rate episodes for PMH, and the EV drivers' plug-in times, charging requirements, and charging windows. We explain our approach in generating these inputs below.

Generating the Existing Load. We use CAISO system demand (www.energyonline.com) to generate the existing load in our instances. To create computationally manageable instances, we scale down the CAISO system demand to 1,000 households, assuming California has 13,044,266 households (see, e.g., www.census.gov). For example, Figure 10(a) is the CAISO system demand during the first three days of February 2019. In Figure 10(b), we multiply the total demand by 1,000/13,044,266, representing an instance with 1,000 households. Note the shape of the existing consumption profile and its peak and off-peak periods are preserved. In the remainder, we use Figure 10(b) as the existing load. Our analysis can be repeated using the existing load during other times.

PMH Tariffs. For the low-rate episode in PMH, we use TOU-D-PRIME Rates offered by SCE (www.sce.com), in which, in winter, 9 p.m to 4 p.m. (next day) is the low-rate episode. Figure 10(b) uses dotted lines to plot the timings of the low-rate episodes. For simplicity, we assume all PMH participants have the same low-rate episode.

EV Owners' Charging Requirements and Timings. We assume an EV driver leaves her home at $t \sim \mathcal{N}(7$ a.m., 2 hours); that is, her departure time is normally distributed around 7 a.m., with a standard deviation of two hours. We assume she returns home at $t \sim \mathcal{N}(6$ p.m., 3 hours). Moreover, by driving her EV during the

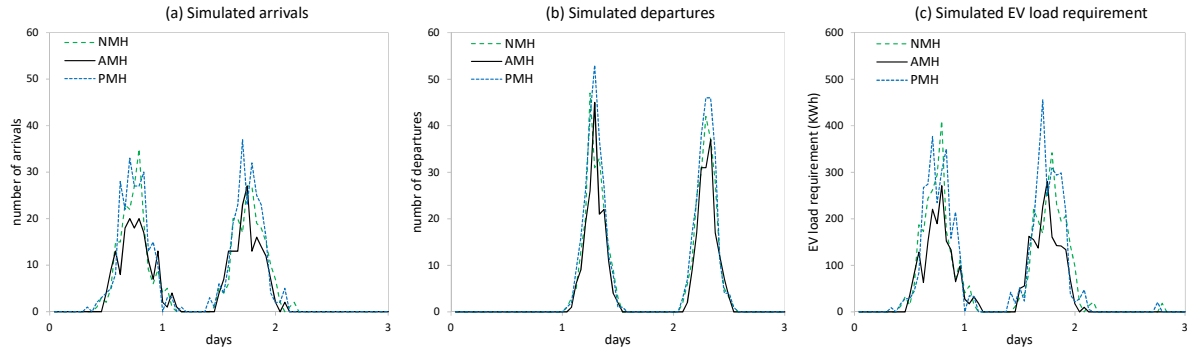

Figure 11 Simulating EV drivers' arrival and departure times and their load requirements (20% NMH, 20% AMH, and 20% PMH)

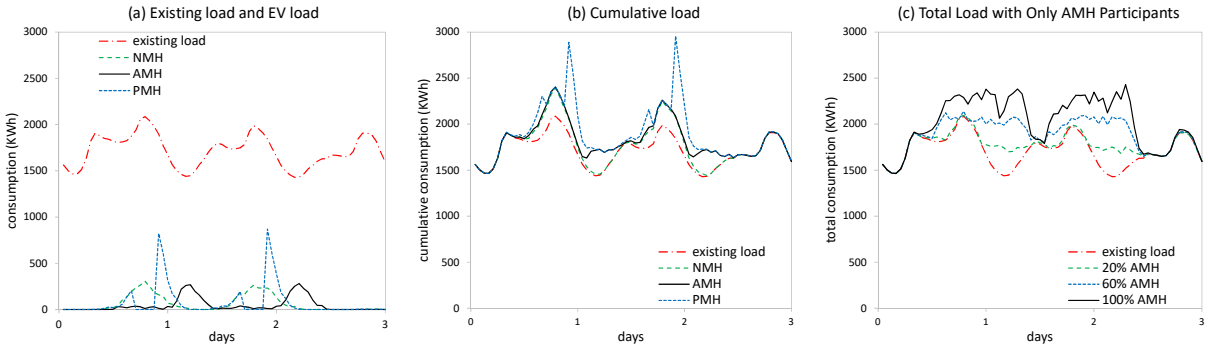

Figure 12 Total load after the addition of EV load: (a) EV load for 20% NMH, 20% AMH, and 20% PMH; (b) cumulative consumption; and (c) total consumption if all EV drivers participate in AMH.

day (e.g., to/from work), she consumes $l \sim \mathcal{U}[0, 22]$ KWh (note 22 KWh is determined based on the average daily driving distance in the U.S. and the average EV consumption per mile). Figure 11 shows the simulated arrivals, departures, and load requirements for an instance of our problem with 20% NMH, 20% AMH, and 20% PMH (the remaining 40% of households do not own EVs). We generate arrivals and departures only for days 1 and 2, which are respectively Friday and Saturday; hence, our instance consists of a weekday and a weekend day.

We assume an EV driver's charging window consists of all times that she is at home. An NMH driver starts charging her EV at the full charging speed after she arrives home in the evening; an AMH participant plugs in her EV after she returns home in the evening and leaves it plugged in until she leaves for work the next morning; and a PMH participant assigns her charging hours to the hours with the lowest tariffs within her charging window (if alternative optimal solutions exist, she minimizes her secondary objective, which is her charging completion time, to avoid uncertainty). Figure 12(a)-(b) shows integration of the total EV load to the existing load for the instance given in Figure 11. NMH drivers' load is distributed over the first few hours of their arrival time, which happens naturally because of their charging speed limit. AMH participants' load is assigned to before- and after-peak periods, which is obtained by solving LM. PMH participants start charging at their full speed after the low-tariff begins, which creates a sharp peak after 9 p.m. every day.

In Figure 12(c), we show how the total load changes if all EV drivers participate in AMH; for example,

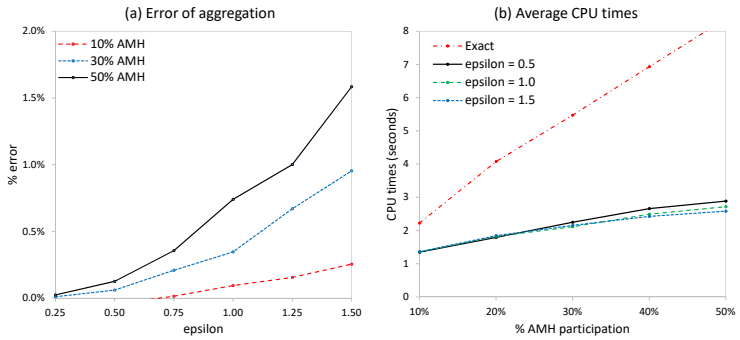

Figure 13 Numerical analysis of aggregation; impact on error and CPU times

the solid (dashed and dotted) line represents the case in which all (20% and 60%) of the households own EVs and participate in AMH. Figure 12(c) represents an ideal case in which the utility manages the EV load for all EV drivers. Observe the total load is nicely balanced for different EV adoption rates.

In our instances, we consider two scenarios to incorporate stochasticity. We also assume two substations exist, and we uniformly assign the 1,000 participants to these substations to incorporate grid constraints. Moreover, we use a cost function in the form of $f_t(y) = cy^3$, for $y \geq 0$. Because we only report relative changes in cost, the value of c is inconsequential, and hence, we set $c = 1$. We extend our numerical experiments to non-convex cost functions in F.2.3.

F.2. Numerical Analysis of Our Approximation

Recall our solution approach for LM consists of aggregation, truncation, and linearization. Next, we perform an extensive numerical study to investigate the effectiveness of these approximations.

F.2.1. Aggregation. Figure 13 summarizes our numerical results and shows our aggregation is near optimal and significantly reduces the CPU time. We compute the errors in Figure 13(a) by comparing the increase in cost when the aggregate model ALM is used with when the original model LM is used. Each data point is the average of 10 randomly generated instances. Because we solve the exact model for each instance, solving larger instances with more than 50% AMH is not time-wise manageable, which explains why we only consider up to 50% AMH participation. Figure 13(a) shows the error is always very small; the largest error is for the case of 50% AMH and $\epsilon = 1.5$, which is around 1.5%. We find the error increases in ϵ and in the AMH participation level.

Figure 13(b) shows how the CPU time grows as we increase the number of participants. The CPU time of the exact model grows rapidly, which makes it impractical for real-size instances (note because we have scaled down our instances to 1,000 customers, the 50% participation corresponds to an instance with 500 EV drivers participating in AMH). The CPU time of the aggregate model modestly increases initially but stays reasonable for large instances. The reason is that for initial increases in the number of participants, the number of groups increases, whereas after all potential groups are formed, new participants join the existing groups, and hence, the number of groups remains the same beyond some threshold.

In short, Figure 13 supports the appropriateness and effectiveness of our aggregation. This approximation is essential to ensure our solution approach applies to practical cases with many participants. Motivated by Figure 13, in the remainder, we use $\epsilon = 0.5$. Observe the error is less than 0.13% when $\epsilon = 0.5$.

F.2.2. Truncation. Figure 14 summarizes our numerical analysis of the truncation. We show in Figure 14(a) how the error changes as we increase the length of the truncated horizon, from two hours to 26 hours.

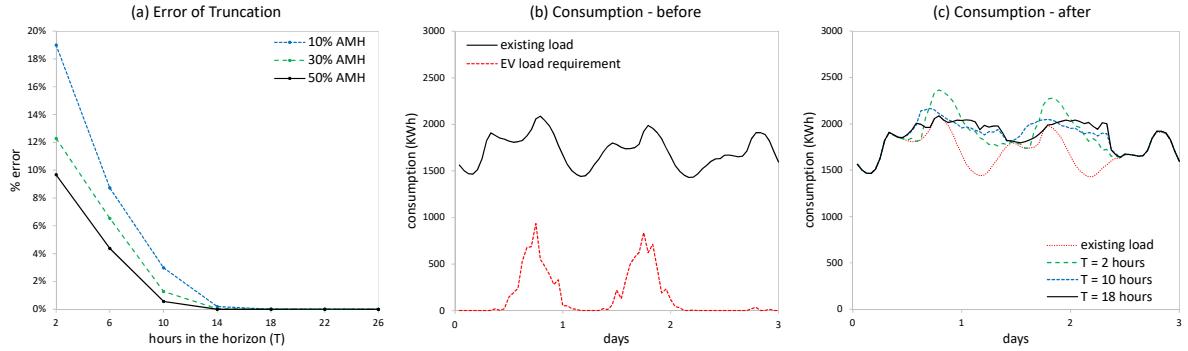

Figure 14 Numerical analysis of truncation

Each data point is an average of 10 randomly generated instances. We compute these errors by comparing them with the case in which the horizon is 26 hours (as a proxy for the optimal increase in cost); for example, for 10% AMH, the increase in cost when we truncate the horizon to two hours is 19% higher than when we truncate the horizon to 26 hours. Beyond 18 hours, the error is zero, meaning the objective value remains the same. In other words, the objective value does not improve by further extending the horizon beyond 18 hours. This observation justifies the appropriateness of using the case of 26 hours as a proxy for the optimal value. We find a significant reduction in error when we increase the length of the horizon from two to 14 hours, whereas the error is significantly small beyond 14 hours. Motivated by this numerical analysis, we use $\breve{T} = 15$ in the remainder of this appendix.

Figure 14(b)-(c) shows the impact of truncating the horizon on integrating the EV load. For example, a truncated horizon of two hours implies we have the flexibility to move the load only within a two-hour interval each time we solve the problem on a rolling basis. Thus, as the length of the truncated horizon increases, we have higher flexibility to move around the EV load; thus, we achieve a better load balance.

In short, truncation is an appropriate and effective approximation, which our theoretical and numerical analyses support. This approximation is necessary because our problem can have a long horizon, in which case, it cannot be constructed and solved in practice by known techniques/software. Truncation allows us to find near-optimal solutions for the large instances of our problem in a reasonable amount of time.

F.2.3. Linearization. Figure 15(a) shows the error of our linearization for different values of δ. Recall δ is a measure of our sensitivity to the variations in the hourly consumption when we compute the total energy cost. Each data point is the average of 10 randomly generated instances. The horizontal axis is in logarithmic scale. We find $\delta = 0.1$ (i.e., 10% of the maximum hourly load) results in large errors, whereas all δ-values less than 0.01 (meaning 1% variation in the hourly load is acceptable) provide negligible error. Figure 15(b)-(c) shows how the choice of δ affects the consumption profile. We note huge fluctuations when $\delta = 0.1$, because the model is sensitive only within 10% of the hourly load. We observe an approximately balanced consumption profile when δ is 0.01 and 0.0001. Motivated by these results, we set $\delta = 0.01$.

So far in this appendix, we have considered a convex objective function. Some utilities may have concave cost functions, for example, due to inflexible generation resources. To demonstrate the generality of our method, we also consider a non-convex objective function $f(y) = \max\{c'\sqrt{y},\ c''y^3\}$, which is concave in the interval $[0, \left(\frac{c'}{c''}\right)^{-2.5}]$ and convex in $[\left(\frac{c'}{c''}\right)^{-2.5}, +\infty]$. Considering CAISO's existing load, we let $\left(\frac{c'}{c''}\right)^{-2.5} = 25,000$ MWh, implying the cost function is concave below 25,000 MWh and convex thereafter. Figure 16 shows our numerical results. We find the CPU time and the number of integer variables are small when

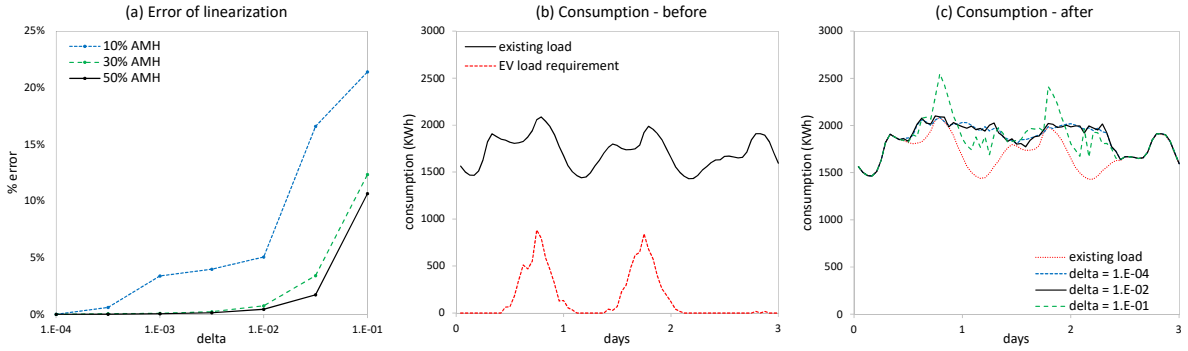

Figure 15 Numerical analysis of linearization

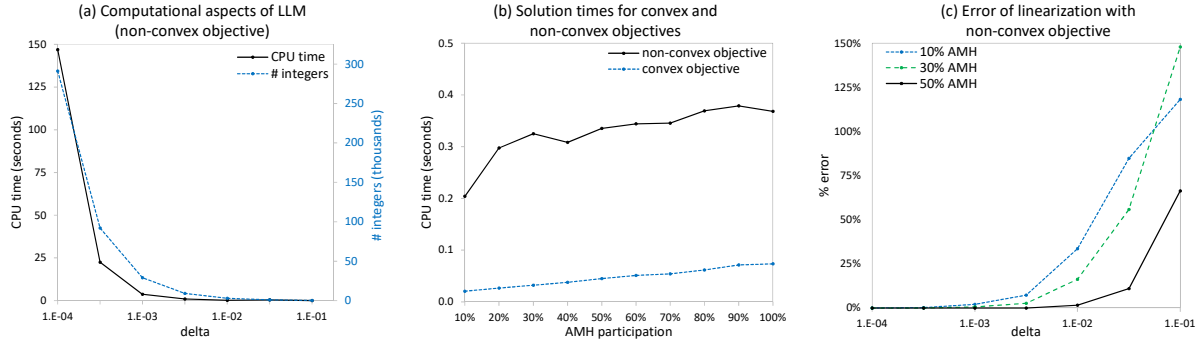

Figure 16 Numerical analysis of linearization with a non-convex objective

$\delta = 0.01$ (Figure 16(a)). According to Figure 16(b), although the CPU time is larger for the non-convex objective, it remains practically reasonable for different participation levels in AMH (we let $\delta = 0.01$ in Figure 16(b)). We also repeat Figure 15(a) in Figure 16(c). We find the relative error is larger for the non-convex objective. Nicely, the relative error decreases as the participation in AMH increases, and it is reasonably small when $\delta = 0.01$.

Thus, we find linearization with $\delta = 0.01$ provides near-optimal solutions. As we previously stated, our linearization is essential for practicality of our approach, because it converts our problem to a mixed-integer linear program solvable in a few seconds (CPU times are provided in Figure 13(b)).

In summary, in this section, we performed extensive numerical experiments to demonstrate the solution quality and computational efficacy of our approach. We find (i) aggregation error increases in ϵ and in the AMH participation level, (ii) truncation error significantly reduces when we increase $\breve{T}$ from 2 hours to 14 hours, and the error is negligible beyond 14 hours, (iii) linearization error is small when $\delta \leq 1\%$, (iv) the CPU time of our approximation model is small for large instances, and (v) $\epsilon = 0.5$, $\breve{T} = 15$, and $\delta = 1\%$ are appropriate choices, leading to fast computation times and near-optimal solutions.

Appendix G: Extensions and Generalizations

We extend and generalize our methodology to the cases where (i) the drivers' requested loads are uncertain within scenarios and (ii) the grid capacity is uncertain.

G.1. Uncertainty in the Requested Loads within Scenarios

We extend our method to the case in which the realized charging deviates from the requested load, which could happen due to, for example, uncertainty in the total load needed to fully charge the EV, inaccurate driver estimates of the required load, and so forth.

Consider an AMH participant j. In her d^{th} charging event, she plugs in her EV at the beginning of $\tilde{t}^{\text{A}}_{j,d}$ and requests $\tilde{Q}^{\text{R}}_{j,d}$ to be charged until she leaves at the beginning of $\tilde{t}^{\text{D}}_{j,d}$. Let $\tilde{Q}^{\text{N}}_{j,d}$ denote the amount of load that the utility needs to supply to the EV driver in her d^{th} charging event. We assume the conditional cumulative distribution of $\tilde{Q}^{\text{N}}_{j,d}$ given $\tilde{Q}^{\text{R}}_{j,d}$ is known, which we denote by $F_{\tilde{Q}^{\text{N}}_{j,d}|\tilde{Q}^{\text{R}}_{j,d}}(\cdot)$. We introduce a service level $\Gamma \in [0,1]$, which allows the utility to fully (respectively, partially) satisfy the load with probability Γ (respectively, $1-\Gamma$). We assume $0 \le F^{-1}_{\tilde{Q}^{\text{N}}_{j,d}|\tilde{Q}^{\text{R}}_{j,d}}(\Gamma) < \infty$. The utility prepares to supply any amount up to $\tilde{Q}^{\text{U}}_{j,d} \triangleq F^{-1}_{\tilde{Q}^{\text{N}}_{j,d}|\tilde{Q}^{\text{R}}_{j,d}}(\Gamma)$. Finally, given the requested load $\tilde{Q}^{\text{R}}_{j,d}$ and the service level Γ, the utility ends up supplying $\tilde{Q}_{j,d}$ in her d^{th} charging event. For example, assume she requests $\tilde{Q}^{\text{R}}_{j,d} = 10$ KWh, and the load that is needed to be supplied is uniformly distributed around the requested load as $\tilde{Q}^{\text{N}}_{j,d} \sim \mathcal{U}[9,11]$ KWh. Assume the utility's service level is $\Gamma = 90\%$; hence, $\tilde{Q}^{\text{U}}_{j,d} = F^{-1}_{\tilde{Q}^{\text{N}}_{j,d}|\tilde{Q}^{\text{R}}_{j,d}}(0.9) = 10.8$ KWh. That is, the utility fully satisfies the load if $\tilde{Q}^{\text{N}}_{j,d} \le 10.8$ KWh; otherwise, if $\tilde{Q}^{\text{N}}_{j,d} > 10.8$ KWh, the utility supplies $\tilde{Q}_{j,d} = 10.8$ KWh, meaning the load is partially satisfied. The expected value of $\tilde{Q}_{j,d}$ is $\mathbf{E}[\tilde{Q}_{j,d}] = \int_{9}^{10.8} \frac{1}{2} Q dQ + \int_{10.8}^{11} \frac{1}{2} 10.8 dQ = 9.99$ KWh. PD readily extends by using $\tilde{Q}_{j,d}$ to estimate her total load requirement. In LM, under a given scenario, the requested load Q^{R}_j is deterministic (recall we drop index d in LM). The utility computes Q^{U}_j based on its service level and uses Q^{U}_j in the right-hand side of constraint (3). Our extension of LM is similar to the robust optimization techniques, in which the solution is feasible with a high probability and the optimal value is a bound (see, e.g., Bertsimas and Sim 2004). As a special case, if $\Gamma = 1$, our extended LM would be similar to the robust modeling approach of Soyster (1973).

G.2. Uncertainty in the Grid Capacity

In LM, we assumed the capacity of a substation h_i is deterministic and constant over time. Renewable energy sources create randomness in the supply capacities. Let $\tilde{h}_{i,s,t}$ denote the capacity of substation i, under scenario s, and in period t. Assume the cumulative distribution of $\tilde{h}_{i,s,t}$ is known, which we denote by $F_{\tilde{h}_{i,s,t}}(\cdot)$. To extend LM, similar to §G.1, we introduce a service level Γ', define $h^{\text{U}}_{i,s,t} \triangleq F^{-1}_{\tilde{h}_{i,s,t}}(1-\Gamma')$, and use $h^{\text{U}}_{i,s,t}$ in the right-hand side of constraint (11). We offer further analyses on this extension as a direction for future research. With an increasing number of innovative business models and utility investments in renewable energy sources, such an analysis could be of interest to researchers who study renewable energy (see, e.g., Hu et al. 2015, Sunar and Birge 2019, Sunar and Swaminathan 2021).

References

Bertsimas D, Sim M (2004) The price of robustness. *Operations research* 52(1):35–53.

EV-Adoption (2023) Ev sales forecast. `https://evadoption.com/ev-sales/ev-sales-forecasts/`, accessed: 11-06-2023.

Fattahi A, Dasu S, Ahmadi R (2023) Peak-load energy management by direct load control contracts. *Management Science* 69(5):2788–2813.

Feldman D, Monemizadeh M, Sohler C (2007) A ptas for k-means clustering based on weak coresets. *Proceedings of the twenty-third annual symposium on Computational geometry*, 11–18.

Hu S, Souza GC, Ferguson ME, Wang W (2015) Capacity investment in renewable energy technology with supply intermittency: Data granularity matters! *Manufacturing & Service Operations Management* 17(4):480–494.

Mettu RR, Plaxton CG (2004) Optimal time bounds for approximate clustering. *Machine Learning* 56(1):35–60.

Soyster AL (1973) Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research* 21(5):1154–1157.

Sunar N, Birge JR (2019) Strategic commitment to a production schedule with uncertain supply and demand: Renewable energy in day-ahead electricity markets. *Management Science* 65(2):714–734.

Sunar N, Swaminathan JM (2021) Net-metered distributed renewable energy: A peril for utilities? *Management Science* 67(11):6716–6733.