Proc. of IEEE Transportation Electrification Conference (ITEC)
19-21 June 2024, Rosemont, IL, USA

A Reinforcement Learning-Augmented Lyapunov Optimization Approach to DC Fast Charging Station Management

Mohammad Hossein Abbasi
Automotive Engineering Department
Clemson University
Greenville, SC, USA
mabbasi@clemson.edu

Ziba Arjmandzadeh
Mechanical Engineering Department
University of Oklahoma
Norman, OK, USA
ziba.arjmandzadeh-1@ou.edu

Dillip Kumar Mishra
Automotive Engineering Department
Clemson University
Greenville, SC, USA
dmishra@clemson.edu

Jiangfeng Zhang
Automotive Engineering Department
Clemson University
Greenville, SC, USA
jiangfz@clemson.edu

Bin Xu
Mechanical Engineering Department
University of Oklahoma
Norman, OK, USA
binxu@ou.edu

Venkat Krovi
Automotive Engineering Department
Clemson University
Greenville, SC, USA
vkrovi@clemson.edu

***Abstract*—A Lyapunov optimization (LO) approach is proposed in this paper to minimize the operation costs of a DC fast charging station (FCS). The LO method eliminates the need for future forecasts, e.g., EV arrival time or required charging energy, and reduces computation time. The FCS is equipped with a battery energy storage system (ESS) to mitigate the costs and the station's strain on the grid during peak hours. However, using LO can lead to underutilization of the ESS. To address this issue, a reinforcement learning (RL) agent is trained to change the desired energy level in the ESS such that it is close to the optimal value. Lastly, simulation results demonstrate that the RL-augmented LO method diminishes the costs by 30%.**

***Index Terms*—DC fast charging station, Lyapunov optimization, reinforcement learning, deep deterministic policy gradient, energy storage system**

I. Introduction

Electric vehicles (EVs) powered by renewable generation have great potential to decrease the transportation sector's carbon footprint [1]. Therefore, the U.S. bipartisan infrastructure bill allocated $15 billion to develop low-emission buses and ferries and build a nationwide network of plug-in EV chargers [2]. Furthermore, in order to enable long-distance travel by EVs and address users' range anxiety, it is essential to establish fast charging infrastructure [3]. Moreover, cost optimization of fast charging stations (FCSs) is crucial to attract investors. However, the optimization depends on uncertain user behavior, such as arrival time, power and energy demand, etc. Forecasting such uncertain aspects of users' random behavior is challenging. Hence, we propose a Lyapunov optimization (LO) approach to tackle the FCS optimization problem, eliminating the need for future forecasts and drastically increasing computation speed. Further, as demonstrated in the results, the LO approach underutilizes the ESS, a crucial component to reduce costs and FCS load on the grid during peak time. Hence, the proposed LO strategy is augmented by a deep learning algorithm based on reinforcement learning (RL) to resolve the underutilization issue.

In recent years, many scholars have been conducting research on FCS optimization. In [4], adaptable charging ports are designed to maximize FCS's profit. Relying on data forecasting, the authors in [5] devise a strategy based on which EV users decide whether to charge their batteries. A novel mechanism is proposed in [6] to run an FCS under limited available power while maximizing user quality of service. In [7], FCS operation costs and EV waiting time are minimized without using ESS or renewable generation at the FCS. Notably, the works in [4]–[7] rely on forecast information, which negatively impacts the optimal solution as forecasting user behavior requires unrealistic assumptions such as a known constant traveling speed or identical energy consumption of all EVs as they travel. On the other hand, the LO approach relaxes the need for future forecasts [8].

Introduced in [9], LO is an optimization approach that decomposes the problem into many subproblems, each solved for a single time step [10]. Additionally, LO does not require future information as the underlying problem only depends on the current time step. Besides, any constraint that depends on several time steps, such as difference equations, is converted into a virtual queue backlog. An LO model is investigated in [8] to maximize FCS's long-term profit. The researchers in [11] use LO to optimize power allocation among charging EVs in an FCS. Finally, in [12], the LO algorithm is employed to control ESS utilization of a commercial building. However,

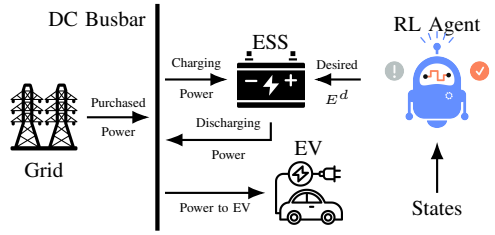

Fig. 1. The components of the FCS along with the RL agent.

in the mentioned literature, the virtual queues follow a fixed desired level. In contrast, the present work integrates an RL agent to dynamically change the desired ESS energy level to improve the LO results.

In this paper, we propose a novel optimization problem based on LO where the operation cost of a DC FCS is minimized. As illustrated in Fig. 1, the FCS is equipped with a battery ESS and can directly purchase electricity from the grid. The contributions and novelties of this work are twofold: i) The LO algorithm eliminates the need for future forecasts. In other words, the LO problem's time steps are 5 minutes long, for which we assume the forecast is not required and accurate user behavior is known. ii) An RL agent is trained through a deep deterministic policy gradient (DDPG) algorithm to change the desired ESS energy level, denoted as E^d, such that ESS energy, E_t^{ESS}, is close to the optimal value, denoted as E_t^*. RL training is performed on historical data for which perfect information is available (forecast is not required), and E_t^* can be calculated by solving the problem via a mixed-integer quadratic programming (MIQP) solver.

II. Problem Formulation

In this paper, a novel approach for optimizing the operational costs of an FCS equipped with an ESS is presented. First, the optimization problem is defined in (1) considering the cost of purchasing electricity from the grid, ESS utilization cost, and the penalty corresponding to unsatisfied EV demands. Subsequently, the proposed optimization problem in (1) is cast into (12) to enable utilizing LO, which decomposes the problem into many subproblems, one subproblem for each time step. With the help of Theorems 1 and 2, it is proved that the optimal solution of (12) is within a controllable boundary around the global optimal solution of (1). The benefit of exploiting LO is improving computation speed and removing the need for forecast. However, the solution of the LO leads to the underutilization of ESS, as is demonstrated in Section III. Consequently, a deep learning approach based on a DDPG algorithm is designed to dynamically change the desired energy level in ESS and resolve the underutilization issue. Further, it is shown in Section III that exploiting the proposed RL approach improves the results. The FCS operation costs optimization problem is as follows,

$$\min \quad \frac{1}{T}\mathbb{E}\left\{\sum_{t=1}^{T}\left(\lambda_t D_t + C_t^{\text{ESS}} + C_t^{\text{EV}}\right)\right\} dt \tag{1a}$$

subject to

$$z_{t,v} = z_{t-1,v} + \frac{\eta_{\text{EV}} P_{t,v}^{\text{EV}}}{C_v} dt, \quad \forall t, v \tag{1b}$$

$$z_v^{\min} \le z_{t,v} \le z_v^{\max}, \quad \forall t, v \tag{1c}$$

$$P_{t,v}^{\text{EV}} \le P_v^{\max} A_{t,v}^{\text{EV}} b_{t,v}, \quad \forall t, v \tag{1d}$$

$$x_t = \sum_{\mathcal{V}} b_{t,v}, \quad \forall t \tag{1e}$$

$$C_t^{\text{ESS}} = \lambda^{\text{ESS}} \left(P_t^c + P_t^d\right), \quad \forall t \tag{1f}$$

$$C_t^{\text{EV}} = \sum_{v=1}^{N_{\text{EV}}} \left(z_v^{\text{target}} - z_{t,v}\right) C_v, \quad \forall t, v \tag{1g}$$

$$D_t^{\text{D}} = P_t^c - P_t^d + \sum_{v=1}^{N_{\text{EV}}} P_{t,v}^{\text{EV}}, \quad \forall t \tag{1h}$$

$$P_t^d \le \sum_{v=1}^{N_{\text{EV}}} P_{t,v}^{\text{EV}}, \quad \forall t \tag{1i}$$

$$P_t^c \le P^{\text{ESS}} \kappa_t^{\text{ESS}}, \quad \forall t \tag{1j}$$

$$P_t^d \le P^{\text{ESS}} (1 - \kappa_t^{\text{ESS}}), \quad \forall t \tag{1k}$$

$$P_{t,v}^{\text{EV}}, P_t^c, P_t^d \ge 0, \quad \forall t \tag{1l}$$

$$E_{t+1}^{\text{ESS}} = E_t^{\text{ESS}} + \left(\eta_{\text{ESS}} P_t^c - \frac{P_t^d}{\eta_{\text{ESS}}}\right) dt, \quad \forall t \tag{1m}$$

$$E^{\min} \le E_t^{\text{ESS}} \le E^{\max}, \quad \forall t \tag{1n}$$

$$E_T^{\text{ESS}} = E_1^{\text{ESS}} = E^d, \tag{1o}$$

where t is time step, v is EV index, T shows total time steps, λ_t indicates electricity price in \$/kWh, D_t is purchased power from the grid in kW, C_t^{ESS} expresses the cost of utilizing ESS [\$/h], N_{EV} indicates number of EVs, κ_t^{ESS} is a binary variable to differentiate between ESS charge and discharge, λ^{ESS} is operation cost of ESS [\$/kWh], $P_{t,v}^{\text{EV}}$ shows EV charging power [kW], $P_v^{\max}$ is maximum EV charging power in kW, $z_{t,v}$ indicates EV state of charge (SOC), C_v shows EV battery capacity, η_{EV} denotes EV battery charging efficiency, $z_v^{\min}$ and $z_v^{\max}$ are EV SOC limits, z_v^{target} is EV target SOC, P_t^c and P_t^d represent ESS charging and discharging powers [kW], P^{ESS} presents ESS (dis)charging power limit [kW], E_t^{ESS} states ESS energy level at t in kWh, η_{ESS} is ESS power transaction efficiency, $E^{\min}$ and $E^{\max}$ are ESS energy limits [kWh]. $A_{t,v}^{\text{EV}}$ is a binary input that is zero before EV arrival at the station and is one from the time EV arrives at the station to the end of the day, $b_{t,v}$ is a binary variable that is zero if EV is not connected, and $x_t \in \{0, 1, 2, 3, 4\}$ is an integer variable that shows how many chargers are occupied at time t. $A_{t,v}^{\text{EV}}$ ensures EV charging power is zero when the EV has not arrived yet. (1e) limits the maximum number of EVs connected to the chargers to four. Denote the set of decision variables as $\Pi_t = \{P_t^c, P_t^d, z_{t,v}, P_{t,v}^{\text{EV}}, E_t^{\text{ESS}}, \kappa_t^{\text{ESS}}\} \forall v \in \mathcal{V}$.

In order to cast problem (1) into the LO framework, (1b) and (1m) should be rewritten in the form of virtual queues. In this respect, the ESS depth of discharge (DOD) is defined as queue backlog as follows,

$$Q_t^{\text{ESS}} \triangleq E^d - E_t^{\text{ESS}}. \tag{2}$$

Using (1m), (2) can be expressed as

$$Q_{t+1}^{\text{ESS}} = Q_t^{\text{ESS}} + \left(\frac{P_t^d}{\eta_{\text{ESS}}} - P_t^c \eta_{\text{ESS}}\right) dt = Q_t^{\text{ESS}} + \varrho_t, \quad (3)$$

where $\varrho_t = \left(P_t^d/\eta_{\text{ESS}} - P_t^c \eta_{\text{ESS}}\right) dt$.

Likewise, to tackle Eq. (1b), the DOD of each EV battery is defined as a queue backlog as,

$$Q_{t,v}^{\text{EV}} \triangleq z_v^{\text{target}} - z_{t,v}, \quad (4)$$

$$Q_{t+1,v}^{\text{EV}} = Q_{t,v}^{\text{EV}} + \varphi_{t,v}, \quad (5)$$

where $\varphi_{t,v} = -\frac{\eta_{\text{EV}} P_{t,v}^{\text{EV}}}{C_v} dt$.

A. Reformulation

Exploiting the virtual queues in (2) and (4), we can transform the original problem in (1) to a queue stability problem (6), where constraints (1b) and (1m) are changed to stability of virtual queues.

$$\min \quad \frac{1}{T}\mathbb{E}\left\{\sum_{t=1}^{T}\left(\lambda_t D_t + C_t^{\text{ESS}} + C_t^{\text{EV}}\right)\right\} dt \quad (6a)$$

subject to

$$\text{(1d)-(1l)},$$
$$\text{stability of virtual queues } \Theta_t, \quad \forall t \quad (6b)$$

where $\Theta_t = \left(Q_t^{\text{ESS}}, Q_{t,1}^{\text{EV}}, \cdots, Q_{t,N_{\text{EV}}}^{\text{EV}}\right)$ is the queue length vector. Subsequently, LO is applied to (6) to design an adaptive control policy. Problem (6) is decomposed into T subproblems and solved for each time slot t separately, while the system stability is guaranteed.

B. Lyapunov Optimization

Define the Lyapunov function as,

$$L[\Theta_t] \triangleq \frac{\alpha}{2}\left(Q_t^{\text{ESS}}\right)^2 + \frac{\beta}{2}\sum_{v=1}^{N_{\text{EV}}}\left(Q_{t,v}^{\text{EV}}\right)^2, \quad (7)$$

which is a positive-definite function ($\alpha, \beta > 0$) and is used to define the Lyapunov drift [8] as,

$$\Delta[\Theta_t] \triangleq \mathbb{E}\{L[\Theta_{t+1}] - L[\Theta_t]|\Theta_t\}. \quad (8)$$

Finally, the drift-plus-penalty term is defined as [11],

$$\Delta[\Theta_t] + V\mathbb{E}\{u_t|\Theta_t\}, \quad (9)$$

$$u_t = \left(\lambda_t D_t + C_t^{\text{EV}} + C_t^{\text{ESS}}\right) dt, \quad (10)$$

where V is a non-negative constant that adjusts the weight between minimizing the queue or the objective in (6a). According to the LO theory, if (8) has an upper bound for all t, then minimizing the upper bound at each time slot t solves problem (6). The following theorem derives the upper bound for the Lyapunov drift.

Theorem 1. *At any time slot t, the Lyapunov drift-plus-penalty term has the following upper bound,*

$$\Delta[\Theta_t] + V\mathbb{E}\{u_t|\Theta_t\} \le B + \alpha Q_t^{ESS}\mathbb{E}\{\varrho_t|\Theta_t\} + \beta\sum_{v=1}^{N_{EV}} Q_{t,v}^{EV}\mathbb{E}\{\varphi_{t,v}|\Theta_t\} + V\mathbb{E}\{u_t|\Theta_t\}, \quad (11)$$

where B is a positive constant as follows,

$$B = \frac{\alpha}{2}\left[\frac{P^{ESS}}{\eta_{ESS}}dt\right]^2 + \frac{\beta N_{EV}}{2}\left[Mdt\right]^2,$$
$$M = \max_v\left\{\frac{P_v^{max}\eta_{EV}}{C_v}\right\}.$$

Proof. Using (7), we can write,

$$L[\Theta_{t+1}] - L[\Theta_t] = \frac{\alpha}{2}\left(\left(Q_{t+1}^{\text{ESS}}\right)^2 - \left(Q_t^{\text{ESS}}\right)^2\right) + \frac{\beta}{2}\sum_{v=1}^{N_{\text{EV}}}\left(\left(Q_{t+1,v}^{\text{EV}}\right)^2 - \left(Q_{t,v}^{\text{EV}}\right)^2\right).$$

With the help of (2) and (4), we obtain,

$$L[\Theta_{t+1}] - L[\Theta_t] = \frac{\alpha}{2}\left(E_{t+1}^{\text{ESS}} - E_t^{\text{ESS}}\right)\left(E_{t+1}^{\text{ESS}} + E_t^{\text{ESS}} - 2E^d\right) + \frac{\beta}{2}\sum_{v=1}^{N_{\text{EV}}}\left[\left(z_{t+1,v} - z_{t,v}\right)\left(z_{t+1,v} + z_{t,v} - 2z_v^{\text{target}}\right)\right].$$

Combining (1m) and (3) as well as (1b) and (5) we have $E_{t+1}^{\text{ESS}} = E_t^{\text{ESS}} - \varrho_t$ and $z_{t+1,v} = z_{t,v} - \varphi_{t,v}$. Leveraging (2) and (4) we obtain,

$$L[\Theta_{t+1}] - L[\Theta_t] = \frac{\alpha}{2}\varrho_t\left(\varrho_t + 2Q_t^{\text{ESS}}\right) + \frac{\beta}{2}\sum_{v=1}^{N_{\text{EV}}}\varphi_{t,v}\left(\varphi_{t,v} + 2Q_{t,v}^{\text{EV}}\right) = \frac{\alpha}{2}\varrho_t^2 + \alpha Q_t^{\text{ESS}}\varrho_t + \beta\sum_{v=1}^{N_{\text{EV}}}\left(\frac{\varphi_{t,v}^2}{2} + Q_{t,v}^{\text{EV}}\varphi_{t,v}\right).$$

Utilizing (1j)-(1k) and the definition of ϱ_t, we have,

$$-P^{\text{ESS}}\eta_{\text{ESS}}dt \le \varrho_t \le \frac{P^{\text{ESS}}}{\eta_{\text{ESS}}}dt.$$

Similarly, by (1d) and the definition of $\varphi_{t,v}$, we obtain,

$$-\frac{P_v^{\max}dt}{C_v}\eta_{\text{EV}} \le \varphi_{t,v} \le 0$$

Since $0 < \eta_{\text{ESS}} \le 1$, then $\varrho_t^2 \le \left[P^{\text{ESS}}dt/\eta_{\text{ESS}}\right]^2$. On the other hand, $\varphi_{t,v}^2 \le \left[Mdt\right]^2$. Therefore,

$$L[\Theta_{t+1}] - L[\Theta_t] \le B + \alpha Q_t^{\text{ESS}}\varrho_t + \beta\sum_{v=1}^{N_{\text{EV}}} Q_{t,v}^{\text{EV}}\varphi_{t,v}.$$

By taking conditional expectations from both hand-sides with respect to Θ_t and adding $V\mathbb{E}\{u_t|\Theta_t\}$ to both hand-sides,

$$\Delta[\Theta_t] + V\mathbb{E}\{u_t|\Theta_t\} \le B + \alpha Q_t^{\text{ESS}}\mathbb{E}\{\varrho_t|\Theta_t\} + \beta\sum_{v=1}^{N_{\text{EV}}} Q_{t,v}^{\text{EV}}\mathbb{E}\{\varphi_{t,v}|\Theta_t\} + V\mathbb{E}\{u_t|\Theta_t\},$$

where $\Delta[\Theta_t]$ is substituted using (8). ■

We can minimize the upper bound of the drift-plus-penalty in (11) through the following problem,

$$\min \quad \beta\sum_{v=1}^{N_{\text{EV}}} Q_{t,v}^{\text{EV}}\varphi_{t,v} + \alpha Q_t^{\text{ESS}}\varrho_t + Vu_t, \quad (12)$$

Algorithm 1 Minimizing FCS Operation Cost

Initialize: $Q_1^{\text{ESS}} \leftarrow E^d - E_1^{\text{ESS}}$ and $Q_{1,v}^{\text{EV}} \leftarrow z_v^{\text{target}} - z_{1,v},\ \forall v$

1: **for** $t \in [1, T]$ **do**
2: solve problem (12) and find outputs for t
3: $\varrho_t = \left(P_t^d/\eta_{\text{ESS}} - P_t^c \eta_{\text{ESS}}\right) dt$
4: $\varphi_{t,v} = -P_{t,v}^{\text{EV}} \eta_{\text{EV}} dt / C_v$
5: $Q_t^{\text{ESS}} \leftarrow Q_t^{\text{ESS}} + \varrho_t$
6: $Q_{t,v}^{\text{EV}} \leftarrow Q_{t,v}^{\text{EV}} + \varphi_{t,v},\ \forall v$
7: **end for**

subject to (1d)-(1l),

In addition, we demonstrate all queues in Θ_t are bounded, and (6b) holds. By replacing u_t, $Q_{t,v}^{\text{EV}}$, Q_t^{ESS}, $\varphi_{t,v}$, and ϱ_t in (12) and rearranging, the objective function becomes,

$$\left(V\left(\lambda_t + \lambda_t^{\text{ESS}}\right) - \alpha\eta_{\text{ESS}} Q_t^{\text{ESS}}\right) dt P_t^c +$$
$$\left(V\left(-\lambda_t + \lambda_t^{\text{ESS}}\right) + \frac{\alpha}{\eta_{\text{ESS}}} Q_t^{\text{ESS}}\right) dt P_t^d +$$
$$\sum_{v=1}^{N_{\text{EV}}} \left(V(z_v^{\text{target}} - z_{t,v}) C_v + V\lambda_t P_{v,t}^{\text{EV}} - \beta \frac{Q_v^{\text{EV}} \eta_{\text{EV}}}{C_v} P_{t,v}^{\text{EV}}\right) dt.$$

Based on (1j) and (1k), either P_t^c or P_t^d is always zero. Therefore, since the problem is minimization, ESS will be charged or discharged under the following conditions, which yield Q_t^{ESS} boundaries,

$$\begin{cases} P_t^c \geq 0, P_t^d = 0, & \text{if } Q_t^{\text{ESS}} \geq V\frac{\lambda_t + \lambda^{\text{ESS}}}{\alpha\eta_{\text{ESS}}}, \\ P_t^c = 0, P_t^d \geq 0, & \text{if } Q_t^{\text{ESS}} \leq V\frac{\lambda_t - \lambda^{\text{ESS}}}{\alpha}\eta_{\text{ESS}}, \end{cases}$$

$$Q_t^{\text{ESS}} \leq \min\left\{V\frac{\lambda_t + \lambda^{\text{ESS}}}{\alpha\eta_{\text{ESS}}} + \frac{P^{\text{ESS}}}{\eta_{\text{ESS}}} dt, E^d - E^{\min}\right\}, \quad (13a)$$

$$Q_t^{\text{ESS}} \geq \max\left\{V\frac{\lambda_t - \lambda^{\text{ESS}}}{\alpha}\eta_{\text{ESS}} - P^{\text{ESS}}\eta_{\text{ESS}} dt, E^d - E^{\max}\right\}. \quad (13b)$$

Similarly, $Q_{t,v}^{\text{EV}}$ boundaries are,

$$\min\left\{\frac{V\lambda_t C_v}{\beta\eta_{\text{EV}}}, z_v^{\text{target}} - z_{t,v}\right\} \leq Q_{t,v}^{\text{EV}} \leq z_v^{\text{target}} - z_{t,v}.$$

Since (12) is independently solved for each time slot t, the solving procedure can be implemented via Algorithm 1. Subsequently, the following theorem proves the solution of Algorithm 1 converges to the global optimal solution of (1), denoted as p^*.

Theorem 2. *For all $T > 1$, the optimal solution of* (12)*, denoted by $\hat{\Pi}_t$, and the optimal solution of* (1)*, denoted by Π_t^*, satisfy the following inequality,*

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left\{\hat{u}_t\right\} \leq \frac{B_1}{V} + p^*, \quad (14)$$

where p^* is the optimal objective function value of (1)*, $\hat{u}_t$ is calculated by* (10) *using $\hat{\Pi}_t$, and B_1 is*

$$B_1 = B + \beta N_{EV} M dt + \frac{1}{T}\mathbb{E}\{L[\hat{\Theta}_1]\}$$

Proof. For Π_t satisfying the constraints of (1), define $A_t = P_t^d dt/\eta_{\text{ESS}}$ and $\xi_t = P_t^c \eta_{\text{ESS}} dt$. Consequently, based on (3) and (1m) we derive $Q_{t+1}^{\text{ESS}} = Q_t^{\text{ESS}} + A_t - \xi_t$ and $E_{t+1}^{\text{ESS}} = E_t^{\text{ESS}} - A_t + \xi_t$, respectively. By taking expectation on (1m) and summing over $\{1, 2, \cdots, T\}$, we have

$$\mathbb{E}\{E_T^{\text{ESS}}\} - \mathbb{E}\{E_1^{\text{ESS}}\} = \sum_{t=1}^{T}\left(-\mathbb{E}\{A_t\} + \mathbb{E}\{\xi_t\}\right). \quad (15)$$

Using (1o) and dividing (15) by T we obtain

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\{A_t\} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\{\xi_t\}. \quad (16)$$

This implies that all feasible solutions Π_t satisfying the constraints of (1) will satisfy the constraint (16). Now, by replacing (1m) with (16), we define a new problem as

$$\min \quad \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{u_t\right\}$$
$$\text{subject to (1d)-(1l), (16)} \quad (17)$$

Moreover, because all Π_t satisfying the constraints of (1) will also satisfy (16) and thus all constraints of (17), problem (17) can be called a relaxed version of problem (1) in the sense that the optimal solution of (1) is a feasible solution of (17), and the optimal objective of (17) is less than or equal to the optimal objective value of (1). Assume $\tilde{p}$ denotes the optimal objective function value of (17) under the optimal solution $\tilde{\Pi}_t$,

$$\tilde{p} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\tilde{u}_t\right\} \leq p^* = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{u_t^*\right\}, \quad (18)$$

where $\tilde{u}_t$ and u_t^* are calculated by the optimal solutions of (17) and (1) using (10), respectively. On the other hand, because (12) minimizes the upper bound of the drift-plus-penalty term, the solutions of (12) yield the minimum of $\beta\sum_{v=1}^{N_{\text{EV}}} Q_{t,v}^{\text{EV}} \cdot \varphi_{t,v} + \alpha Q_t^{\text{ESS}} \cdot \varrho_t + V \cdot u_t$, which equals to the following,

$$\beta\sum_{v=1}^{N_{\text{EV}}} \hat{Q}_{t,v}^{\text{EV}} \cdot \hat{\varphi}_{t,v} + \alpha\hat{Q}_t^{\text{ESS}}\left(\hat{A}_t - \hat{\xi}_t\right) + V \cdot \hat{u}_t, \quad (19)$$

where $\hat{A}_t$ and $\hat{\xi}_t$ correspond to the values of A_t and ξ_t evaluated at $\hat{\Pi}_t$, respectively. Evaluating the objective of (12) with any other values satisfying the constraints of (12) (i.e., (1d)-(1l)) results in a value that is greater than or equal to (19). Note that the solution $\tilde{\Pi}_t$ of the relaxed problem (17) also satisfies (1d)-(1l), thus

$$\beta\sum_{v=1}^{N_{\text{EV}}} \hat{Q}_{t,v}^{\text{EV}} \cdot \hat{\varphi}_{t,v} + \alpha\hat{Q}_t^{\text{ESS}}\left(\hat{A}_t - \hat{\xi}_t\right) + V \cdot \hat{u}_t \leq$$
$$\beta\sum_{v=1}^{N_{\text{EV}}} \tilde{Q}_{t,v}^{\text{EV}} \cdot \tilde{\varphi}_{t,v} + \alpha\tilde{Q}_t^{\text{ESS}}\left(\tilde{A}_t - \tilde{\xi}_t\right) + V \cdot \tilde{u}_t, \quad (20)$$

where $\tilde{A}_t$ and $\tilde{\xi}_t$ correspond to the values of A_t and ξ_t evaluated at $\tilde{\Pi}_t$, respectively. By taking expectations with respect to $\hat{\Theta}_t$ of (19) and Plugging the result in (11),

$$\Delta[\hat{\Theta}_t] + V\mathbb{E}\{\hat{u}_t|\hat{\Theta}_t\} \leq B + \alpha\hat{Q}_t^{\text{ESS}}\mathbb{E}\{\hat{A}_t - \hat{\xi}_t|\hat{\Theta}_t\} +$$

$$\beta \sum_{v=1}^{N_{\text{EV}}} \hat{Q}_{t,v}^{\text{EV}} \mathbb{E}\{\hat{\varphi}_{t,v} | \hat{\Theta}_t\} + V \mathbb{E}\{\hat{u}_t | \hat{\Theta}_t\}.$$

By taking expectations from both hand sides, we obtain

$$\mathbb{E}\{\Delta[\hat{\Theta}_t]\} + V\mathbb{E}\{\hat{u}_t\} \leq B + \alpha \mathbb{E}\{\hat{Q}_t^{\text{ESS}}\}\mathbb{E}\{\hat{A}_t - \hat{\xi}_t\} + \beta \sum_{v=1}^{N_{\text{EV}}} \mathbb{E}\{\hat{Q}_{t,v}^{\text{EV}}\}\mathbb{E}\{\hat{\varphi}_{t,v}\} + V\mathbb{E}\{\hat{u}_t\}. \quad (21)$$

Using (20), (21), and (8) and by summing both hand sides over all time slots, we obtain the following,

$$\mathbb{E}\{L[\hat{\Theta}_{T+1}]\} - \mathbb{E}\{L[\hat{\Theta}_1]\} + V \sum_{t=1}^{T} \mathbb{E}\{\hat{u}_t\} \leq V \sum_{t=1}^{T} \mathbb{E}\{\tilde{u}_t\} + TB + \alpha \sum_{t=1}^{T} \mathbb{E}\{\tilde{Q}_t^{\text{ESS}}\}\mathbb{E}\{\tilde{A}_t - \tilde{\xi}_t\} + \beta \sum_{t=1}^{T} \sum_{v=1}^{N_{\text{EV}}} \mathbb{E}\{\tilde{Q}_{t,v}^{\text{EV}}\}\mathbb{E}\{\tilde{\varphi}_{t,v}\}.$$

Since $L[\Theta_t]$ is a positive definite function, we can drop $\mathbb{E}\{L[\hat{\Theta}_{T+1}]\}$ from the left-hand side of the inequality. By rearranging, dividing by TV, and using (16) and (18),

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\{\hat{u}_t\} \leq \frac{B}{V} + p^* + \frac{\beta}{TV} \sum_{t=1}^{T} \sum_{v=1}^{N_{\text{EV}}} \mathbb{E}\{\tilde{Q}_{t,v}^{\text{EV}}\}\mathbb{E}\{\tilde{\varphi}_{t,v}\} + \frac{1}{TV}\mathbb{E}\{L[\hat{\Theta}_1]\}$$

As mentioned in the proof of Theorem 1, $\tilde{\varphi}_{t,v} \leq Mdt$. Besides, according to (4), $\tilde{Q}_{t,v}^{\text{EV}} \leq 1$. Hence,

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\{\hat{u}_t\} \leq \frac{B_1}{V} + p^*,$$

which is identical to (14). ■

C. Dynamic ESS Desired Energy Level

The Q_t^{ESS} bounds in (13) render underutilization of ESS as shown in Fig. 2b. To resolve the issue, an RL agent is trained using the DDPG algorithm to change the ESS desired energy level, E^d, per time step. The RL agent learns to change E^d based on time of day, current electricity price, energy level in ESS, and EV demand. The training is executed on historical data. To generate reward signals during training, E_t^* is found for training data via an MIQP solver.

1) Environment: The environment is the FCS. For each time step, Algorithm. 1 receives RL action and finds P_t^c and P_t^d. Then, a reward is generated based on the updated E_t^{ESS}.

2) State Space: The state space comprises the current hour, electricity price λ_t, ESS energy level E_t^{ESS}, and the charging EVs' demands. The states are,

$$\left[\frac{\text{hour}}{24} \quad \frac{\lambda_t}{100} \quad \frac{E_t^{\text{ESS}}}{E^{\max}} \quad \sum_{v=1}^{N_{\text{EV}}} \frac{P_{t,v}^{\text{EV}}}{4\times 150}\right]^T,$$

where the values are normalized by division by their maximums. EV charging power is divided by 4×150 as there are four 150 kW chargers at the station.

3) Action Space: The RL action determines the updated value for the desired ESS energy level E^d. Based on the new E^d, Q_t^{ESS} is updated.

4) Reward Signal: During training, the reward signal is calculated using E_t^*, denoted as E_t^*, obtained by solving problem (1) for June 2023. Since June 2023 is past, we assume perfect knowledge of electricity prices and user behavior. The reward function is as follows,

$$R_t = -\frac{\left\| E_t^* - E_t^{\text{ESS}} \right\|}{E^{\max}}.$$

5) Deep Deterministic Policy Gradient: DDPG generates a continuous action that maximizes Q-function $Q_\phi(s,a)$, which is obtained through gradient ascent over,

$$\max_\theta \mathbb{E}\left\{Q_\phi\left(s, \mu_\theta(s)\right)\right\}.$$

Further, during training, Q-learning in DDPG is performed by minimizing the following loss via gradient decent [13], [14],

$$L(\phi) = \mathbb{E}\left\{\left(Q_\phi(s,a) - (r + \gamma(1-d)Q_{\phi_t}(s', \mu_{\theta_t}(s'))\right)^2\right\}.$$

III. Results and Discussion

In this section, the proposed augmented LO method with RL is applied to a DC FCS with four 150 kW charging ports, which is in accordance with the Federal Highway Administration's minimum number of required charging points in an FCS [15]. Further, the FCS has a 500 kWh ESS with a maximum charging/discharging power of 200 kW. On the other hand, the electricity prices are obtained from the NYISO electricity market [16]. In addition, it is assumed that a maximum of 100 EVs visit the FCS daily, comprising fifteen different types of EVs with the highest sales in 2021 and 2022 [17]. Moreover, EV arrival time is simulated based on historical data in [18].

The RL agent is trained using June 2023 data, and the trained RL is evaluated on the last week of May 2023. Table. I represents the results of applying LO on the train and test days with or without using the trained RL to adjust E^d. The first row of Table. I is June 2023 averaged results. The following seven rows present the results for each test day during the last week of May 2023. The first column, "Optimal solution," presents global optimal solution which is the operation cost, i.e., (1a) excluding C_t^{EV}, obtained by solving the problem with an MIQP solver, assuming perfect information is available. The two middle columns titled "LO without RL" and "LO with RL" show the results of solving the problem with LO when E^d is fixed and when E^d is adjusted by RL, respectively. Finally, the last column shows the percentage improvement when RL is used compared to the global optimal solution.

Fig. 2 illustrates detailed results for test day #2. As seen in Fig. 2b, when RL is not used, the ESS is not charged when prices are low and discharged when prices are high. This behavior stems from the Q_t^{ESS} bounds in (13). In fact, based on λ_t values for the day, the smallest lower and greatest upper bounds of Q_t^{ESS} are -14.81 and 18.87 [kWh], respectively. It means that if Q_t^{ESS} becomes smaller than $E^d - 14.81 = 250 - 14.81$, ESS will be charged in the next time step. Similarly, if Q_t^{ESS} becomes greater than $250+18.87$, ESS will be discharged in the following time step. Adjusting E^d with the help of RL resolves this limitation.

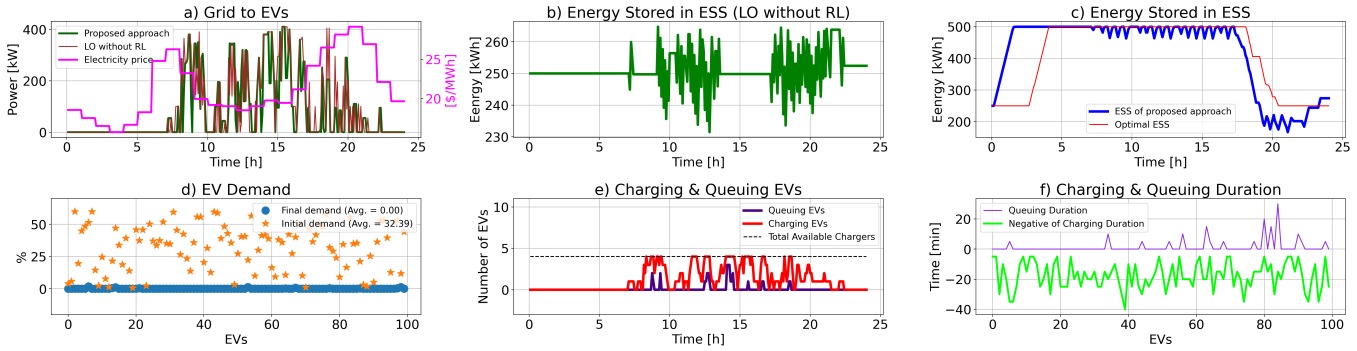

Fig. 2. Test day #2 (05/26/2023). a) Power flow from grid to EVs (averaged every 5 minutes) as well as electricity price. b) E_t^{ESS} when RL is not used to adjust E^d. c) E_t^{ESS} when E^d is adjusted by RL (the blue line). The adjustment reduces the overall cost of the day by 11.98%. The red line is E_t^* from the global optimal solution. d) EVs' initial and final demand in terms of SOC percentage. e) Charger occupation and queue length at the FCS. f) Each EV's charging and queuing duration. Note that subplots (d), (e), and (f) are the same for both cases, i.e., with or without RL adjustment.

TABLE I
RESULTS FOR THE TRAINING MONTH AND THE TEST WEEK

	Optimal solution	LO without RL	LO with RL	Improvement
June 2023 Average	58.9553	78.8494	74.7662	20.52%
Test day 1	43.0842	58.5527	53.4648	32.89%
Test day 2	51.5378	66.2411	64.4798	11.98%
Test day 3	39.2181	54.1628	50.512	24.43%
Test day 4	40.0221	56.5487	50.9714	33.74%
Test day 5	48.5916	67.7079	61.0943	34.60%
Test day 6	53.5637	78.0473	68.2638	39.96%
Test day 7	69.2387	97.4979	89.3151	28.96%
Average	**49.3223**	**68.394**	**62.5859**	**30.45%**

IV. CONCLUSION

This paper presents a novel technique based on an RL-augmented LO approach for minimizing DC FCS operation costs. The FCS under study has a battery ESS whose charging and discharging powers should be optimized. The LO algorithm handles the optimization problem without the need for future information. However, the LO framework instigates underutilization of ESS. To address this issue, an RL agent is trained through the DDPG algorithm to alter the desired energy level in ESS, which resolves the underutilization issue. Lastly, in simulation results, it is demonstrated that the proposed approach reduces the costs by 30% compared to the results of solving the problem with LO alone.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-2312196.

REFERENCES

[1] M. H. Abbasi, J. Zhang, and V. Krovi, "A lyapunov optimization approach to the quality of service for electric vehicle fast charging stations," in *2022 IEEE Vehicle Power and Propulsion Conf. (VPPC)*, pp. 1–6, IEEE, 2022.

[2] "White house us. president biden's bipartisan infrastructure law." https://www.whitehouse.gov/build/, 2021.

[3] B. Xu and Z. Arjmandzadeh, "Parametric study on thermal management system for the range of full (tesla model s)/compact-size (tesla model 3) electric vehicles," *Energy Convers. Manag.*, vol. 278, p. 116753, 2023.

[4] S. Mishra, A. Mondal, and S. Mondal, "A multi-objective optimization framework for electric vehicle charge scheduling with adaptable charging ports," *IEEE Trans. Veh. Technol.*, 2022.

[5] B. Kim, M. Paik, Y. Kim, H. Ko, and S. Pack, "Distributed electric vehicle charging mechanism: A game-theoretical approach," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8309–8317, 2022.

[6] R. Buckreus, R. Aksu, M. Kisacikoglu, M. Yavuz, and B. Balasubramanian, "Optimization of multiport dc fast charging stations operating with power cap policy," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 4, pp. 2402–2413, 2021.

[7] M. Tan, Y. Ren, R. Pan, L. Wang, and J. Chen, "Fair and efficient electric vehicle charging scheduling optimization considering the maximum individual waiting time and operating cost," *IEEE Trans. Veh. Technol.*, 2023.

[8] G. Fan, Z. Yang, H. Jin, X. Gan, and X. Wang, "Enabling optimal control under demand elasticity for electric vehicle charging systems," *IEEE Trans. Mobile Comput.*, vol. 21, no. 3, pp. 955–970, 2020.

[9] M. Neely, *Stochastic network optimization with application to communication and queueing systems*. Springer Nature, 2010.

[10] Y. Wang, F. Xu, S. Mao, S. Yang, and Y. Shen, "Adaptive online power management for more electric aircraft with hybrid energy storage systems," *IEEE Trans. Transport. Electrific.*, vol. 6, no. 4, pp. 1780–1790, 2020.

[11] M. H. Abbasi, J. Zhang, and V. Krovi, "A lyapunov optimization approach to the quality of service for electric vehicle fast charging stations," in *2022 IEEE Vehicle Power and Propulsion Conference (VPPC)*, pp. 1–6, IEEE, 2022.

[12] J. Shi, Z. Ye, H. O. Gao, and N. Yu, "Lyapunov optimization in online battery energy storage system control for commercial buildings," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 328–340, 2022.

[13] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[14] "Openai." https://spinningup.openai.com/en/latest/algorithms/ddpg.html, 2023.

[15] "National electric vehicle infrastructure standards and requirements." https://www.federalregister.gov/documents/2023/02/28/2023-03500/national-electric-vehicle-infrastructure-standards-and-requirements, 2023.

[16] "Lcg consulting." http://www.energyonline.com/Data/Default.aspx, 2023.

[17] "Electric vehicle sales report." https://www.coxautoinc.com/wp-content/uploads/2023/01/Kelley-Blue-Book-EV-Sales-and-Data-Report-for-Q4-2022.pdf, 2023.

[18] "Ev charge station use sept 2018 to aug 2019." https://data.pkc.gov.uk/datasets/ev-charge-station-use-sept-2018-to-aug-2019/about, 2023.