Load Restoration in Islanded Microgrids: Formulation and Solution Strategies

Shourya Bose and Yu Zhang, *Member, IEEE*

Abstract

Extreme weather events induced by climate change can cause significant disruptions to the normal operation of electric distribution systems (DS), including isolation of parts of the DS due to damaged transmission equipment. In this paper, we consider the problem of load restoration in a microgrid (MG) that is islanded from the upstream DS because of an extreme weather event. The MG contains sources of distributed generation such as microturbines and renewable energy sources, in addition to energy storage systems. We formulate the load restoration task as a non-convex optimization problem with complementarity constraints. We propose a convex relaxation of the problem that can be solved via model predictive control. In addition, we propose a data-driven policy-learning method called constrained policy optimization. The solutions from both methods are compared by evaluating their performance in load restoration, which is tested on a 12-bus MG.

I. Introduction

Extreme weather events such as the wildfires, hurricanes, and winter storms pose a big threat to the reliable operation of electric distribution systems (DS) [1]. Those events can disrupt the operation of DS by damaging electric transmission equipment such as overhead cables, thereby curbing the distribution of electric power. Traditionally, DS have been designed to be reliable during nominal operations and in face of predictable off-nominal operating conditions. Recently, however, a new paradigm of *resilience* is being explored by the power engineering community, which posits that a DS must be capable of rapidly recovering to a state of nominal operations post extreme weather events [2]. A cornerstones in achieving DS resilience is the concept

S. Bose and Y. Zhang are with the Department of Electrical and Computer Engineering at the University of California, Santa Cruz. Emails: {shbose,zhangy}@ucsc.edu

This work was supported in part by the Faculty Research Grant of UC Santa Cruz, Seed Fund Award from CITRIS and the Banatao Institute at the University of California, and the Hellman Fellowship.

of microgrids MGs, which are localized distribution systems, often equipped with sources of distributed power generation such as microturbines (MTs) and renewable energy sources (RES). RES may in turn contain elements such as photovoltaics and wind turbines. The intermittent nature of power production by RES necessitates the storage of generated energy, that can be dispatched as and when needed. Therefore, MGs also contain energy storage systems (ESS), which can be used store energy during periods of low power demand and inject energy into the MG during periods of high power demand.

Since MGs contain both sources of power generation and energy storage, they are an ideal candidate for restoring the demand for power by loads such as residential homes and industries, should the MG become disconnected from the upstream DS due to an extreme weather event (a phenomenon known as *islanding*), until the connection can be restored. One of the most important advantages of a MG is the control of generation as well as storage resources available to the MG controller (MGC) [3], which allows the implementation of algorithms for rapid load restoration post extreme weather event. The survey [4] covers the state-of-art in MG research.

Some of the previous works on the load restoration in DSs involve risk-limiting strategies [5], utilization of wide-area monitoring systems [6], distributed algorithms [7], and expert systems [8]. However, one of the common assumptions in conventional load restoration is the capability to withdraw power from the upstream DS through a point of common coupling (PCC), which is no longer valid in case a MG is islanded from the DS. In this paper, we formulate the load restoration problem for an islanded MG, and propose and compare two different methods.

Model predictive control (MPC) is a popular technique for the control of dynamical systems in which an optimization problem is solved on every time step to generate the optimal control over a fixed time horizon, and then the first input of the optimal control is applied to the actual system. MPC has been utilized by the power engineering community for several applications such as voltage stability assurance [9], demand response in industrial loads [10], volt/var control [11], and scheduling PV storage systems [12]. Reinforcement learning (RL) is concerned with determining which actions an agent, which interacts with its environment, should take such that the reward collected (as a function of actions taken) is maximized over a given time horizon. In the last decade, RL has been used in various power systems applications such as volt/var control [13], EV charging schedule determination [14], power management in networked MGs [15], and optimal control of ESS in MGs [16].

In this paper, we first propose an efficient convex relaxation for the load restoration problem,

which can be solved via MPC. Moreover, we leverage a RL policy learning method called constrained policy optimization (CPO) [17] to learn the optimum policy for load restoration while ensuring safety of learned policy in terms of constraint satisfaction. The derived policy is compared with the MPC solution through simulations.

Notation: The notations $\mathbb{R}$, $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}_+$ denote the sets of real numbers, natural numbers, integers, and non-negative reals, respectively. $[a,b]_{\mathbb{Z}} \triangleq [a,b] \cap \mathbb{Z}$ denotes the set collecting all integer indices between a and b. $\text{conv}(\mathcal{A})$ denotes the convex hull of the set $\mathcal{A}$. $D_{\text{KL}}\left(p_1 \| p_2\right)$ denotes the KL-divergence between probability distributions p_1 and p_2. Boldface variables represent vectors and matrices. $\mathbf{I}_n$ denotes the identity matrix of size n. $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

II. Load Restoration Problem Formulation

A. Objective Function and Constraints

Consider a MG operating on alternating current (AC) that has been islanded from the upstream DS due to an extreme weather event. Let the MG be represented as a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ represents the set of buses and $\mathcal{E}$ is the set of transmission lines. We assume that $\mathcal{G}$ is a *radial network* (i.e. $\mathcal{G}$ is a tree). $\mathcal{N}$ is the union of disjoint sets $\mathcal{N}^{\text{L}}$, $\mathcal{N}^{\text{MT}}$, $\mathcal{N}^{\text{RES}}$, and $\mathcal{N}^{\text{ESS}}$, where $\mathcal{N}^{\text{L}}$ represents the load buses, $\mathcal{N}^{\text{MT}}$ represents the buses connected to a MT (which we assume to be nonempty), $\mathcal{N}^{\text{RES}}$ represents the buses connected to a RES, and $\mathcal{N}^{\text{ESS}}$ represents the buses connected to an ESS. We consider the load restoration problem over a time horizon of $\mathcal{T} \triangleq \{1, 2, \cdots, T\}$, and time steps are indexed by the variables t or k. We let $P_{i,t}$, $Q_{i,t}$, and $v_{i,t}$ denote the real power injection, reactive power injection and squared magnitude of voltage phasor at bus $i \in \mathcal{N}$ respectively on time step t. For every line $(i,j) \in \mathcal{E}$, we let $P_{ij,t}$, $Q_{ij,t}$, and $l_{ij,t}$ denote the *sending-end* active power, reactive power and squared magnitude of the current phasor respectively. For buses i and j, the notation $i \to j$ indicates the presence of a transmission line from i to j, i.e. $(i,j) \in \mathcal{E}$. Note that the power terms $P_{i,t}, Q_{i,t}, Q_{ij,t}, Q_{ij,t}$ assume values in $\mathbb{R}$, while the squared voltage magnitude $v_{i,t}$ and current magnitude $l_{ij,t}$ take nonnegative real values. Furthermore, a positive value of any power quantity represents injection into the MG, while a negative value represents withdrawal from the MG.

Objective function: The objective is to maximize load restoration while minimizing MT fuel consumption. Assuming a linear relation between fuel consumption and power generation for the MTs, the objective function is given as

$$J = -\sum_{k\in\mathcal{T}}\left(\sum_{i\in\mathcal{N}^{\mathrm{L}}}\xi_i^{\mathrm{L}}P_{i,k}+\sum_{i\in\mathcal{N}^{\mathrm{MT}}}\xi_i^{\mathrm{MT}}P_{i,k}\right), \tag{1}$$

where coefficients ξ_i^{L} specify the priority order of load restoration to different load buses, while coefficients ξ_i^{MT} specify the per-unit costs of power generation of different MTs. Note that the negative sign accounts for the fact that load power is constrained to be withdrawn while MT power is constrained to be injected into the MG, and these constraints will be formalized in the following subsections. In general, we can assume quadratic or piecewise-linear relation between MT power and cost of generation, without introducing additional theoretical complexity into the problem.

Power flow constraints: Let r_{ij} and x_{ij} represent the resistance and reactance of line $(i,j)\in\mathcal{E}$. We consider the *DistFlow* model for quantifying the power flow in the MG, which is given as follows:

$$P_{j,t} = \sum_{k:j\to k}P_{jk,t} - \sum_{i:i\to j}(P_{ij,t}-r_{ij}l_{ij,t}),\ \forall j\in\mathcal{N} \tag{2a}$$

$$Q_{j,t} = \sum_{k:j\to k}Q_{jk,t} - \sum_{i:i\to j}(Q_{ij,t}-x_{ij}l_{ij,t}),\ \forall j\in\mathcal{N} \tag{2b}$$

$$v_{j,t} = v_{i,t} - 2(r_{ij}P_{ij,t}+x_{ij}Q_{ij,t}) + (r_{ij}^2+x_{ij}^2)l_{ij},\ \forall(i,j)\in\mathcal{E} \tag{2c}$$

$$l_{ij,t}v_{i,t} = P_{ij,t}^2+Q_{ij,t}^2,\ \forall(i,j)\in\mathcal{E}. \tag{2d}$$

The equations (2a) to (2d) precisely describe the power flows in a radial network [18]. To ensure nominal operation of the DS, the nodal voltage magnitudes are constrained within appropriate upper and lower limits:

$$\underline{v}\le v_{i,t}\le\bar{v},\ \forall i\in\mathcal{N},\ \forall t\in\mathcal{T}. \tag{3}$$

Furthermore, in order to prevent damage to transmission equipment due to current-induced heating, the squared current magnitude on a transmission lines should be less than a given threshold:

$$l_{ij,t}\le\bar{l}_{ij},\ \forall(i,j)\in\mathcal{E},\forall t\in\mathcal{T}. \tag{4}$$

MT Constraints: The power generation of each MT is subject to several constraints due to its physical characteristics. The MT real power is constrained to assume values in a certain interval:

$$\underline{P}_i^{\text{MT}} \leq P_{i,t} \leq \bar{P}_i^{\text{MT}},\ \forall i \in \mathcal{N}^{\text{MT}}, \forall t \in \mathcal{T}. \tag{5}$$

Due to inertia in the MT, the ramp-up and ramp-down of the output power is constrained as

$$\underline{P}_{\text{rd}}^{\text{MT}} \leq P_{i,t} - P_{i,t-1} \leq \bar{P}_{\text{ru}}^{\text{MT}}, \forall i \in \mathcal{N}^{\text{MT}}, t \in \mathcal{T}\backslash\{1\}, \tag{6a}$$

$$P_{i,t} \leq P_{\text{ru}}^{\text{MT}},\ \forall i \in \mathcal{N}^{\text{MT}}, t = 1. \tag{6b}$$

Each MT is assume to have a fixed amount of fuel at the beginning of $\mathcal{T}$, which constrains the total amount of active power produced over $\mathcal{T}$ as

$$\left(\sum_{t\in\mathcal{T}} P_{i,t}\right)\tau_i \leq E_i,\ \forall i \in \mathcal{N}^{\text{MT}},$$

where E_i is the total fuel initially available at MT i, and τ_i is the conversion factor between quantity of fuel expended and power produced. The total fuel constraint can be reformulated as a recursive relation in the amount of fuel left in the MT on a given time step. Denoting by $\varsigma_{i,t}$ the amount of fuel remaining in MT i at time t:

$$\varsigma_{i,t+1} = \varsigma_{i,t} - \tau_i P_{i,t},\ \forall i \in \mathcal{N}^{\text{L}}, \forall t \in \mathcal{T}\backslash\{T\} \tag{7a}$$

$$\varsigma_{i,T} \geq 0,\ \forall i \in \mathcal{N}^{\text{L}} \tag{7b}$$

with the initial condition $\varsigma_{i,1} = E_i$.

ESS Constraints: Each ESS is a large-scale energy reservoir which may discharge power into the MG when required, and otherwise charge in order to replenish its energy reserves. The energy available to ESS i at time step t is represented by its state of charge (SoC) $\mathcal{S}_{i,t}$ which takes values in $[\underline{\mathcal{S}}_i, \bar{\mathcal{S}}_i]$. Let $P_{i,t}^{\text{ch}}$ and $P_{i,t}^{\text{dis}}$ denote the charge and discharge powers of ESS i on time step t, and both these quantities are assumed to be bounded. We assume that on any given time step, the ESS may either charge or discharge but cannot do both simultaneously. Such a *complementarity constraint* (CC) can be formulated as

$$0 \leq P_{i,t}^{\text{dis}} \leq \bar{P}_i^{\text{dis}},\ \forall i \in \mathcal{N}^{\text{ESS}}, \forall t \in \mathcal{T}, \tag{8a}$$

$$0 \leq P_{i,t}^{\text{ch}} \leq \bar{P}_i^{\text{ch}},\ \forall i \in \mathcal{N}^{\text{ESS}}, \forall t \in \mathcal{T} \tag{8b}$$

$$P_{i,t}^{\text{ch}} P_{i,t}^{\text{dis}} = 0,\ \forall i \in \mathcal{N}^{\text{ESS}}, \forall t \in \mathcal{T} \tag{8c}$$

The net injection of power from an ESS into the DS is the sum of the charge and discharge powers:

$$P_{i,t} = P^{\text{dis}}_{i,t} - P^{\text{ch}}_{i,t},\ \forall i \in \mathcal{N}^{\text{ESS}}, \forall t \in \mathcal{T}. \tag{9}$$

Lastly, we consider the evolution of ESS SoC $\mathcal{S}_{i,t}$. The SoC increases proportional to the charge power and reduces proportional to the discharge power, which can be formalized as an evolution equation:

$$\mathcal{S}_{i,t+1} = \mathcal{S}_{i,t} + \eta^{\text{ch}}_i P^{\text{ch}}_{i,t}\Delta t - \frac{1}{\eta^{\text{dis}}_i} P^{\text{dis}}_{i,t}\Delta t,\ \forall i \in \mathcal{N}^{\text{ESS}}, \forall t \in \mathcal{T}\backslash\{T\}, \tag{10a}$$

$$\mathcal{S}_{i,1} = \mathcal{S}^{\text{init}}_i,\ \mathcal{S}_{i,t} \in [\underline{\mathcal{S}}, \bar{\mathcal{S}}],\ \forall i \in \mathcal{N}^{\text{ESS}}, \forall t \in \mathcal{T}, \tag{10b}$$

where $\eta^{\text{ch}}_i \in (0, 1]$ and $\eta^{\text{dis}}_i \in (0, 1]$ respectively denote the charge and discharge efficiency for ESS i, Δt denotes the time duration corresponding to each time step, and $\mathcal{S}^{\text{init}}_i$ denotes the initial SoC of ESS i.

RES Constraints: The power output of RES is stochastic in nature and cannot be predicted with certainty ahead-of-time. Therefore, we use a forecast value of $\hat{P}^{\text{RES}}_{i,t}$ rather than the actual RES power output in the load restoration problem. We also assume that each RES has the capability to curtail its active power output. Denoting by $\kappa_{i,t} \in [0, 1]$ the curtailment ratio of the injected active power, the RES active power constraint is given as

$$P_{i,t} = (1 - \kappa_{i,t})\hat{P}^{\text{RES}}_{i,t},\ \forall i \in \mathcal{N}^{\text{RES}},\ \forall t \in \mathcal{T}. \tag{11}$$

Reactive Power Constraints: We assume that the MT, RES, and ESS buses are equipped with controllable inverters, and therefore the reactive power constraint at these buses is given as

$$|Q_{i,t}| \leq \bar{Q}_i,\ \forall i \in \mathcal{N}^{\text{MT}} \cup \mathcal{N}^{\text{ESS}} \cup \mathcal{N}^{\text{RES}},\ \forall t \in \mathcal{T}, \tag{12}$$

where $\bar{Q}_i$ is the nameplate capacity of the inverter at bus i.

Load Constraints: We let forecasts of active and reactive power demands of the loads attached to the load buses, which could possibly be time-varying, be denoted by $\tilde{P}^{\text{L}}_{i,t}$ and $\tilde{Q}^{\text{L}}_{i,t}$ respectively. We let $\rho_{i,t} \in [0, 1]$ denote the ratio of satisfied load demand of load $i \in \mathcal{N}^{\text{L}}$ on time step k. $\rho_{i,t}$ is also referred to as *pickup ratio* in literature. Assuming a constant power factor at the load buses, we have the constraints

$$P_{i,t} = \rho_{i,t}\tilde{P}^{\text{L}}_{i,t},\ \ Q_{i,t} = \rho_{i,t}\tilde{Q}^{\text{L}}_{i,t},\ \forall i \in \mathcal{N}^{\text{L}}, \forall t \in \mathcal{T}. \tag{13}$$

A monotonically increasing pickup ratio is preferable over frequent dropping of loads already picked up. This is ensured by the *almost-monotonic* load restoration constraint,

$$\rho_{i,t} - \rho_{i,t-1} \geq -\epsilon,\ \forall i \in \mathcal{N}^{\text{L}}, \forall t \in \mathcal{T}\backslash\{1\} \tag{14}$$

where the parameter $\epsilon \geq 0$ allows for a small leeway in monotonicity of load pickup, and is chosen by the system designer.

Slack Bus Constraints: In our formulation, we assume the bus which has been disconnected from the upstream DS is the slack bus. In general, a slack bus can be any bus in the MG which does not contain any upstream buses. Note that a slack bus can be any one among a load bus, ESS bus, RES bus, or MT bus, and the corresponding active and reactive power constraints apply to it. However, the voltage of the slack bus is assumed to be constant at 1 p.u. Thus, for slackbus $i_0 \in \mathcal{N}$,

$$v_{i_0,t} = 1,\ \forall t \in \mathcal{T}. \tag{15}$$

The purpose of modeling a slack bus is to provide a voltage reference for all other buses.

B. Optimization Problem

We collect the relevant decision variables of the optimization problem into appropriate sets as follows.

$$\mathcal{X}_t^{\text{MT}} \triangleq \left\{P_{i,t}, Q_{i,t}, \varsigma_{i,t}\right\}_{i\in\mathcal{N}^{\text{MT}}},\ \mathcal{X}_t^{\text{RES}} \triangleq \left\{P_{i,t}, Q_{i,t}, \kappa_{i,t}\right\}_{i\in\mathcal{N}^{\text{RES}}}$$
$$\mathcal{X}_t^{\text{ESS}} \triangleq \left\{P_{i,t}, Q_{i,t}, \mathcal{S}_{i,t}, P_{i,t}^{\text{ch}}, P_{i,t}^{\text{dis}}\right\}_{i\in\mathcal{N}^{\text{ESS}}},\ \mathcal{X}_t^{\text{BV}} \triangleq \left\{v_{i,t}\right\}_{i\in\mathcal{N}}$$
$$\mathcal{X}_t^{\text{L}} \triangleq \left\{P_{i,t}, Q_{i,t}, \rho_{i,t}\right\}_{i\in\mathcal{N}^{\text{L}}},\ \mathcal{X}_t^{\text{TL}} \triangleq \left\{l_{ij,t}, P_{ij,t}, Q_{ij,t}\right\}_{(i,j)\in\mathcal{E}}.$$

In the above, the new superscripts TL and BV denote transmission line variables and bus voltages respectively. We collect all the variables on time step t as

$$\mathcal{X}_t \triangleq \left\{\mathcal{X}_t^{\text{MT}}, \mathcal{X}_t^{\text{RES}}, \mathcal{X}_t^{\text{ESS}}, \mathcal{X}_t^{\text{BV}}, \mathcal{X}_t^{\text{L}}, \mathcal{X}_t^{\text{TL}}\right\}.$$

The optimization problem to be solved by the MGC is

$$\begin{aligned} &\max_{\{\mathcal{X}_t\}_{t\in\mathcal{T}}} \quad (1) \\ &\text{s.t.} \quad (2) - (15), \\ &\left(\kappa_{i,t} \in [0,1], \forall i \in \mathcal{N}^{\text{RES}};\ \rho_{i,t} \in [0,1], \forall i \in \mathcal{N}^{\text{L}}\right), \forall t \in \mathcal{T}. \end{aligned} \tag{16}$$

Computation of an optimal solution of (16) allows the MGC to implement said solution for load restoration in the islanded MG.

III. Solution using Model Predictive Control

In this section, we consider relaxations to problem (16) which make it amenable to being solved with MPC. The MPC approach involves solving (16) over subhorizons of $\mathcal{T}$, and using the subhorizon-optimal solutions to construct a near-optimal solution over $\mathcal{T}$ [19]. However, the power flow constraint (2d) and CC (8c) are non-convex and therefore hinder MPC implementation using efficient convex solvers. Thus we provide relaxations for the same in the following subsections.

A. Convex Relaxations of Non-Convex Constraints

Convex Relaxation of Power Flow Constraints: The non-convex constraint (2d) can be relaxed into multiple linear inequalities by first approximating $v_{i,t} \approx 1$ (which is a good approximation as long as $\bar{v}$ and $\underline{v}$ in (3) are close to 1 p.u.), and then using a convex *polygon inner approximation* [20] to replace the quadratic terms as

$$l_{ij,t} \geq h_c(P_{ij,t}) + h_c(Q_{ij,t}),\ \forall (i,j) \in \mathcal{E}, \forall t \in \mathcal{T}, \forall c \in \mathcal{C} \tag{17}$$

where $h_c(x) \triangleq \gamma_c x + \psi_c$ represents each side c of the approximating polygon. A higher number of polygon sides $|\mathcal{C}|$ provides a finer approximation of the quadratic function. Note that, any constraint of the form $y = x^2$ in variables (x, y) can be approximated with the constraints $y = z$ and $z = \max_{c \in \mathcal{C}} \{\gamma_c x + \psi_c\}$ where $\mathcal{C}$ denotes each side of the inner approximating polygon. Eliminating z, the constraints can be equivalently stated as $y \geq \gamma_c x + \psi_c$ for all $c \in \mathcal{C}$.

Convex Relaxation of ESS CC: We propose a convex relaxation for the ESS CC by replacing the nonconvex feasible region of $P^{\text{ch}}_{i,t}$ and $P^{\text{dis}}_{i,t}$ described by (8) with its convex hull. The rationale behind choosing the convex hull of the feasible region in the original problem is to limit the number of feasible CC-violating solutions in the relaxed problem, since the convex hull is the smallest convex superset of any given set.

Lemma 1 (Convex hull of nonconvex ESS feasible region)**.** *Define the set $\mathcal{P}_i$ as*

$$\mathcal{P}_i \triangleq \left\{ (P^{ch}, P^{dis}) \,|\, P^{ch}, P^{dis} \text{ satisfy (8a)-(8c)} \right\}.$$

Then $\mathcal{P}_i^{conv} \triangleq conv(\mathcal{P}_i)$ is given as

$$\mathcal{P}_i^{conv} = \left\{ (P^{ch}, P^{dis}) \,\middle|\, P^{ch} \geq 0, P^{dis} \geq 0, \frac{P^{ch}}{\bar{P}_i^{ch}} + \frac{P^{dis}}{\bar{P}_i^{dis}} \leq 1 \right\}.$$

Proof. Note that the points $(0,0)$, $(0,\bar{P}_i^{\text{dis}})$, and $(\bar{P}_i^{\text{ch}},0)$ are contained in $\mathcal{P}_i$, and therefore any convex set which contains $\mathcal{P}_i$ should contain $\text{conv}\left\{(0,0),(0,\bar{P}_i^{\text{dis}}),(\bar{P}_i^{\text{ch}},0)\right\}$, which is exactly $\mathcal{P}_i^{\text{conv}}$. Thus $\mathcal{P}_i^{\text{conv}} \subseteq \text{conv}(\mathcal{P}_i)$. On the other hand, $\mathcal{P}_i^{\text{conv}}$ is convex and contains the set $\mathcal{P}_i$, and therefore $\text{conv}(\mathcal{P}_i) \subseteq \mathcal{P}_i^{\text{conv}}$. It follows that $\text{conv}(\mathcal{P}_i) = \mathcal{P}_i^{\text{conv}}$. □

Thus, the constraints (8) are replaced by

$$P_{i,t}^{\text{ch}} \geq 0,\ P_{i,t}^{\text{dis}} \geq 0,\ \frac{P_{i,t}^{\text{ch}}}{\bar{P}_i^{\text{ch}}} + \frac{P_{i,t}^{\text{dis}}}{\bar{P}_i^{\text{dis}}} \leq 1,\ \forall i \in \mathcal{N}^{\text{ESS}},\ \forall t \in \mathcal{T}. \tag{18}$$

B. MPC Problem Formulation

We now formulate the MPC problem which solves (16). We let $\hat{H} \in \mathbb{N}$, with $\hat{H} \leq T$ be the length of the *look-ahead* time horizon for the MPC problem on every time step $t \in \mathcal{T}$, with the time horizon itself denoted as $\mathcal{T}_t^{\hat{H}} \triangleq \left[t, t+\hat{H}\right]_{\mathbb{Z}}$. Since the remaining MT fuel $\varsigma_{i,t+1}$ and SoC $S_{i,t+1}$ and can be inferred from $\mathcal{X}_t$ using (7a) and (10a) respectively, we let $\varsigma_{i,t+1} \sim \mathcal{X}_t$ and $\mathcal{S}_{i,t+1} \sim \mathcal{X}_t$ denote the same. We let $J_t^{\hat{H}}$ denote the t^{th} MPC objective function, which is defined similarly to (1), but with the outer summation summing over $k \in \mathcal{T}_t^{\hat{H}}$ (instead of $k \in \mathcal{T}$). The MPC problem on time step $t+1$ is defined as

$$\mathcal{X}_t^* = \underset{\{\mathcal{X}_k\}_{k\in\mathcal{T}_t^{\hat{H}}}}{\arg\max}\ J_t^{\hat{H}} \tag{19}$$

$$\text{s.t.} \quad \text{(2a)-(2c), (3)-(7), (9)-(15) (17)-(18)},$$

$$\varsigma_{i,t} \sim \mathcal{X}_{t-1}^*,\ \forall i \in \mathcal{N}^{\text{MT}}, \quad \mathcal{S}_{i,t} \sim \mathcal{X}_{t-1}^*,\ \forall i \in \mathcal{N}^{\text{ESS}},$$

$$\left(\kappa_{i,t} \in [0,1], \forall i \in \mathcal{N}^{\text{RES}};\ \rho_{i,t} \in [0,1], \forall i \in \mathcal{N}^{\text{L}}\right), \forall t \in \mathcal{T}_{t+1}^{\hat{H}}.$$

Observe that (19) defines the MPC problem on all time steps $t \in \mathcal{T}\setminus\{1\}$. For $t = 1$, the MPC problem remains unchanged except for $\mathcal{S}_{i,1} = \mathcal{S}_i^{\text{init}}$ and $\varsigma_{i,1} = E_i$. Not only is the MPC problem (19) convex, but it is actually a linear program, and can therefore be solved efficiently using commercially available solvers.

IV. Solution using Constrained Policy Optimization

In this section, we pose problem (16) as a constrained Markov decision process (CMDP) to solve it using the RL policy-learning framework of CPO. A CMDP is defined as the 6-tuple $\{\mathcal{S}, \mathcal{A}, p, R, \mathbf{C}, \gamma\}$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $p : \mathcal{S}\times\mathcal{A}\times\mathcal{S} \mapsto [0,1]$ is the state transition probability, $R : \mathcal{S}\times\mathcal{A} \mapsto \mathbb{R}$ is the reward function, $\mathbf{C} : \mathcal{S}\times\mathcal{A} \mapsto \mathbb{R}^M$ is the

constraint function (assuming the CMDP has M constraints attached to it), and $\gamma \in (0,1]$ is the discount factor.

We define the *state* on time step t, denoted by $\mathbf{s}_t$ as

$$\mathbf{s}_t \triangleq \Bigg\{ \{\mathcal{S}_{i,k}\}_{i\in\mathcal{N}^{\text{ESS}}}, \{\nu_{i,k}\}_{i\in\mathcal{N}^{\text{MT}}}, \Bigg[\Big\{ \tilde{P}^{\text{L}}_{i,k}, \tilde{Q}^{\text{L}}_{i,k} \Big\}_{i\in\mathcal{N}^{\text{L}}}, \Big\{ \hat{P}^{\text{RES}}_{i,k} \Big\}_{i\in\mathcal{N}^{\text{RES}}} \Bigg]_{k=t}^{t+\tilde{H}} \Bigg\}, \tag{20}$$

and correspondingly, the *action* on time step t, denoted by $\mathbf{a}_t$ is defined as

$$\mathbf{a}_t \triangleq \Bigg\{ \Bigg[\{P_{i,k}, Q_{i,k}\}_{i\in\mathcal{N}^{\text{RES}}}, \{P_{i,k}, Q_{i,k}\}_{i\in\mathcal{N}^{\text{MT}}}, \Big\{ P^{\text{ch}}_{i,k}, P^{\text{dis}}_{i,k}, Q_{i,k} \Big\}_{i\in\mathcal{N}^{\text{MT}}} \Bigg]_{k=t}^{t+\tilde{H}} \Bigg\}, \tag{21}$$

where $\tilde{H} \in \mathbb{N}$ denotes the look-ahead time horizon for policy search.

Remark 1 (Choice of variables in $\mathbf{s}_t$ and $\mathbf{a}_t$)**.** *The action $\mathbf{a}_t$ contains the variables which are under control of the MGC. Correspondingly, the state variable $\mathbf{s}_t$ should be chosen such that knowledge of $\mathbf{a}_t$ and $\mathbf{s}_t$ should be sufficient information to determine $\mathbf{s}_{t+1}$ (either deterministically, or in terms of a distribution over all possible values $\mathbf{s}_{t+1}$ can assume). Our choice of state and action are such that once $\mathbf{s}_t$ and $\mathbf{a}_t$ are known, the power flow equations* (2)*, SOC evolution* (10)*, and MT fuel quantity evolution* (7) *subject to appropriate constraints over the window $[t, t+\tilde{H}]$ allow for a deterministic s_{t+1}.*

A *policy* $\pi_{\boldsymbol{\theta}} : \mathcal{S} \times \mathcal{A} \mapsto [0,1]$ parameterized by $\boldsymbol{\theta} \in \mathbb{R}^h$ and denoted as $\pi_{\boldsymbol{\theta}}(\mathbf{a}|\mathbf{s})$ gives the probability of taking action $\mathbf{a}$ on current state $\mathbf{s}$. We model the policy as a multivariate gaussian distribution whose mean vector and covariance matrix are generated by a feedforward neural network (FNN). Thus, letting d denote the dimension of the action variable,

$$\pi(\mathbf{a}|\mathbf{s}) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}_{\mathbf{s}}|(2\pi)^d}} e^{-\frac{1}{2}(\mathbf{a}-\boldsymbol{\mu}_{\mathbf{s}})^\top \boldsymbol{\Sigma}_{\mathbf{s}}^{-1}(\mathbf{a}-\boldsymbol{\mu}_{\mathbf{s}})} \tag{22}$$

where $\boldsymbol{\mu}_{\mathbf{s}} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}_{\mathbf{s}} \in \mathbb{R}^{d\times d}$ are generated by using FNNs $\hat{f}^{(1)}_{\boldsymbol{\theta}}$ and $\hat{f}^{(2)}_{\boldsymbol{\theta}}$ as

$$\boldsymbol{\mu}_{\mathbf{s}} = \hat{f}^{(1)}_{\boldsymbol{\theta}}(\mathbf{s}), \quad \boldsymbol{\Sigma}_{\mathbf{s}} = h(\hat{f}^{(2)}_{\boldsymbol{\theta}}(\mathbf{s})), \tag{23}$$

wherein the function $h(\cdot)$ converts the outputs of the FNN $\hat{f}^{(2)}_{\boldsymbol{\theta}}$ into the positive semidefinite covariance matrix $\boldsymbol{\Sigma}_s$. The vector $\boldsymbol{\theta}$ contains the weights and biases which parameterize the FNNs. We can decompose $\boldsymbol{\theta}$ as $\boldsymbol{\theta} = \Big[\boldsymbol{\theta}^{\mu}, \boldsymbol{\theta}^{\Sigma}\Big]$, wherein $\boldsymbol{\theta}^{\mu}$ contains the parameters of the FNN $\hat{f}^{(1)}_{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{\Sigma}$ contains the parameters of the FNN $\hat{f}^{(2)}_{\boldsymbol{\theta}}$. The closed form of function $h(\cdot)$ is provided in Theorem 1.

For the rest of the paper, we refer to both policy distribution $\pi_{\boldsymbol{\theta}}$ and parameter vector $\boldsymbol{\theta}$ as 'policy', with the exact meaning evident from context. The *reward* on time step t is given as

$$R(\mathbf{s}_t, \mathbf{a}_t) = -\sum_{k=t}^{t+\tilde{H}} \gamma^{(k-t)} \left(\sum_{i \in \mathcal{N}^{\mathrm{L}}} \xi_i^{\mathrm{L}} P_{i,k} + \sum_{i \in \mathcal{N}^{\mathrm{MT}}} \xi_i^{\mathrm{MT}} P_{i,k} \right),$$

where $\gamma \in (0, 1]$ is the *discount factor* used to de-emphasize the contribution of uncertain future quantities to the reward.

The constraints on the variables in action $\mathbf{a}_t$ can be denoted in vectorized notation as $\mathbf{C}(\mathbf{s}_t, \mathbf{a}_t) \leq \mathbf{0}$. We define the *reward function* and *constraint function* on time t as

$$J^R(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t) \triangleq \mathbb{E}_{\mathbf{a}_t \sim \pi_{\boldsymbol{\theta}}} \left[R(\mathbf{s}_t, \mathbf{a}_t) \,|\, \mathbf{s}_t \right],$$
$$J^{\mathbf{C}}(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t) \triangleq \mathbb{E}_{\mathbf{a}_t \sim \pi_{\boldsymbol{\theta}}} \left[\mathbf{C}(\mathbf{s}_t, \mathbf{a}_t) \,|\, \mathbf{s}_t \right].$$

According to the CPO approach, the load restoration problem can be solved by determining a policy $\boldsymbol{\theta}^*$ such that for any state s, $\pi_{\boldsymbol{\theta}^*}$ maximizes $J^R(\pi_{\boldsymbol{\theta}^*}, \mathbf{s})$ while respecting $J^{\mathbf{C}}(\pi_{\boldsymbol{\theta}^*}, \mathbf{s}) \leq \mathbf{0}$. Such a policy $\boldsymbol{\theta}^*$ can be found in an episodic fashion by determining a sequence of policies $\{\boldsymbol{\theta}_t\}$ such that

$$\boldsymbol{\theta}_{t+1} = \arg\max_{\boldsymbol{\theta}} J^R(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t) \tag{24a}$$
$$\text{s.t.} \quad J^{\mathbf{C}}(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t) \leq \mathbf{0} \tag{24b}$$
$$D_{\mathrm{KL}}\left(\pi_{\boldsymbol{\theta}}(\,.\,|\mathbf{s}_t) \| \pi_{\boldsymbol{\theta}_t}(\,.\,|\mathbf{s}_t)\right) \leq \delta \tag{24c}$$

wherein, for some $\delta > 0$ (with δ being the *trust region parameter*), (24c) ensures that successive policies do not have large variations. Directly solving (24) can be challenging due to the highly nonlinear and nonconvex nature of the FNNs. Instead, as proposed in [17], on every time step we solve a convex quadratically constrained quadratic program (QCQP) approximation of 24 which is given as

$$\boldsymbol{\theta}_{t+1} = \arg\max_{\boldsymbol{\theta}} \mathbf{a}_t^\top (\boldsymbol{\theta} - \boldsymbol{\theta}_t) \tag{25a}$$
$$\text{s.t.} \quad \mathbf{B}_t^\top (\boldsymbol{\theta} - \boldsymbol{\theta}_t) + \mathbf{c}_t \leq \mathbf{0} \tag{25b}$$
$$(\boldsymbol{\theta} - \boldsymbol{\theta}_t)^\top \mathbf{F}_t (\boldsymbol{\theta} - \boldsymbol{\theta}_t) \leq \delta, \tag{25c}$$

where $\mathbf{F}_t$ is a positive (semi)definite matrix. The closed forms of $\mathbf{a}_t$, $\mathbf{B}_t$, $\mathbf{c}_t$, and $\mathbf{F}_t$ are given in the following result.

Theorem 1 (Parameters in QCQP approximation of (24))**.** *Let the output of the FNN $\hat{f}_{\boldsymbol{\theta}}^{(2)}(\mathbf{s})$ be a lower triangular matrix of size $d \times d$, and let the function $h(\cdot)$ in* (23) *be defined as $h(\mathbf{L}) \triangleq \mathbf{L}\mathbf{L}^\top$. Furthermore, let $\hat{f}_{\boldsymbol{\theta}}^{(1)}(\mathbf{s}_t)$ and $\hat{f}_{\boldsymbol{\theta}}^{(2)}(\mathbf{s}_t)$ be denoted as $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and $\mathbf{L}_{\boldsymbol{\theta}}$ respectively. Then,*

$$\mathbf{a}_t = \left(\mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)} \left[\frac{\partial R_t}{\partial \mathbf{a}_t} \left((\boldsymbol{\epsilon}^\top \otimes \mathbf{I}_d) \frac{\partial \mathbf{v}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \right) \,\middle|\, \mathbf{s}_t \right] \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}$$

$$\mathbf{B}_t = \left(\mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)} \left[\frac{\partial \mathbf{C}_t}{\partial \mathbf{a}_t} \left((\boldsymbol{\epsilon}^\top \otimes \mathbf{I}_d) \frac{\partial \mathbf{v}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \right) \,\middle|\, \mathbf{s}_t \right] \right)_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}$$

$$\mathbf{c}_t = J^{\mathbf{C}}(\pi_{\boldsymbol{\theta}_t}, \mathbf{s}_t)$$

$$\mathbf{F}_t(i,j) = \left[\frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}^\top}{\partial \boldsymbol{\theta}(i)} \left(\mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top \right)^{-1} \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}(j)} + \frac{1}{2} \operatorname{Tr} \left(\left(\mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top \right)^{-1} \frac{\partial \mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top}{\partial \boldsymbol{\theta}(i)} \left(\mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top \right)^{-1} \frac{\partial \mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top}{\partial \boldsymbol{\theta}(j)} \right) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}$$

where $R_t \triangleq R(\mathbf{s}_t, \mathbf{a}_t)$, $\mathbf{C}_t \triangleq \mathbf{C}(\mathbf{s}_t, \mathbf{a}_t)$, $\mathbf{v}_{\boldsymbol{\theta}} \triangleq vec(\mathbf{L}_{\boldsymbol{\theta}})$, and $\mathbf{F}_t(i,j)$ and $\boldsymbol{\theta}(i)$ denote the $(i,j)^{th}$ and i^{th} element of $\mathbf{F}_t$ and $\boldsymbol{\theta}$, respectively.

Proof. The main idea of the proof is to replace the reward function, constraint function, and KL-divergence terms in (24) with their respective Taylor-series approximations around $\boldsymbol{\theta}_t$. The first-order approximation of the objective function is given as

$$J^R(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t) \approx J^R(\pi_{\boldsymbol{\theta}_t}, \mathbf{s}_t) + \nabla_{\boldsymbol{\theta}} J^R(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t)\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t} (\boldsymbol{\theta} - \boldsymbol{\theta}_t),$$

and by comparing the above with (25), we see that $\mathbf{a}_t^\top = \nabla_{\boldsymbol{\theta}} J^R(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t)\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_t}$. In order to compute $\mathbf{a}_t$, we need a closed-form of the gradient

$$\nabla_{\boldsymbol{\theta}} J^R(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t) = \nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{a}_t \sim \pi_{\boldsymbol{\theta}}} \left[R(\mathbf{s}_t, \mathbf{a}_t) \,\middle|\, \mathbf{s}_t \right],$$

which is difficult to evaluate since the gradient is with respect to $\boldsymbol{\theta}$ which parameterizes the distribution over which the expectation is being taken. In order to alleviate this difficulty, we use the *reparametrization trick*. For a standard normal vector $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$, it holds that

$$\mathbf{L}_{\boldsymbol{\theta}} \boldsymbol{\epsilon} + \boldsymbol{\mu}_{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\theta}}, \mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top).$$

Therefore, letting $\mathbf{a}_t = \mathbf{L}_{\boldsymbol{\theta}} \boldsymbol{\epsilon} + \boldsymbol{\mu}_{\boldsymbol{\theta}}$ is equivalent to defining $\mathbf{a}_t$ as a multivariate gaussian random variable with mean and variance given in (23). Thus, we have

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{a}_t \sim \pi_{\boldsymbol{\theta}}} \left[R(\mathbf{s}_t, \mathbf{a}_t) \,\middle|\, \mathbf{s}_t \right] &= \nabla_{\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)} \left[R(\mathbf{s}_t, \mathbf{L}_{\boldsymbol{\theta}} \boldsymbol{\epsilon} + \boldsymbol{\mu}_{\boldsymbol{\theta}}) \,\middle|\, \mathbf{s}_t \right] \\
&= \mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)} \left[\nabla_{\boldsymbol{\theta}} R(\mathbf{s}_t, \mathbf{L}_{\boldsymbol{\theta}} \boldsymbol{\epsilon} + \boldsymbol{\mu}_{\boldsymbol{\theta}}) \,\middle|\, \mathbf{s}_t \right] \\
&= \mathbb{E}_{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)} \left[\frac{\partial R_t}{\partial \mathbf{a}_t} \frac{\partial \mathbf{a}_t}{\partial \mathbf{L}_{\boldsymbol{\theta}}} \frac{\partial \mathbf{L}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} + \frac{\partial R_t}{\partial \mathbf{a}_t} \frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}} \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} \,\middle|\, \mathbf{s}_t \right] \qquad (26)
\end{aligned}$$

Computing $\frac{\partial \mathbf{a}_t}{\partial \mathbf{L}_{\boldsymbol{\theta}}} \frac{\partial \mathbf{L}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}$ involves a multiplication of two tensors, but we can overcome the tensor multiplication by vectorizing $\mathbf{L}_{\boldsymbol{\theta}}$, and it follows that

$$\frac{\partial \mathbf{a}_t}{\partial \mathbf{L}_{\boldsymbol{\theta}}} \frac{\partial \mathbf{L}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \frac{\partial \mathbf{a}_t}{\partial \mathbf{v}_{\boldsymbol{\theta}}} \frac{\partial \mathbf{v}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \left(\boldsymbol{\epsilon}^\top \otimes \mathbf{I}_d\right) \frac{\partial \mathbf{v}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}.$$

On the other hand, $\frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}} = \frac{\partial \mathbf{L}_{\boldsymbol{\theta}} \boldsymbol{\epsilon} + \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}} = \mathbf{I}_d$, and therefore, $\frac{\partial \mathbf{a}_t}{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}} \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}} = \frac{\partial \boldsymbol{\mu}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}$. This verifies the closed form of $\mathbf{a}_t$ given in the theorem statement. The closed form of $\mathbf{B}_t$ can be similarly derived by computing the Hessian of $J^{\mathbf{C}}(\pi_{\boldsymbol{\theta}}, \mathbf{s}_t)$ with respect to $\boldsymbol{\theta}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_t$ (with $\mathbf{c}_t$ is simply being the zeroth-order term in the Taylor expansion), carrying out the reparametrization trick and then deriving the closed form of the partial derivatives.

The first-order term in the Taylor approximation of the constraint (24c) vanishes, and the second order term is used in the relaxed constraint (25c). Matrix $\mathbf{F}_t$ is the *Fisher information matrix* (FIM) that is positive semidefinite by construction. The closed form of FIM for a Gaussian vector with mean $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and covariance matrix $\mathbf{L}_{\boldsymbol{\theta}} \mathbf{L}_{\boldsymbol{\theta}}^\top$ as provided in the theorem statement is well-known [21]. □

We now provide a result that facilitates the computation of the partial derivatives $\frac{\partial R_t}{\partial \mathbf{a}_t}$ and $\frac{\partial \mathbf{C}_t}{\partial \mathbf{a}_t}$. Recall that R_t contains real power terms of the MT and load. Let $P \in \mathbf{a}_t$ denote an arbitrary element of action vector $\mathbf{a}_t$. Calculating the partial derivative of MT real power terms is straightforward, since

$$\frac{\partial P_{i,k}}{\partial P} = \begin{cases} 1, & \text{if } P \text{ is } P_{i,k} \\ 0, & \text{otherwise,} \end{cases} \quad \forall i \in \mathcal{N}^{\text{MT}},\ \forall k \in \mathcal{T}_t^{\tilde{H}},$$

where $\mathcal{T}_t^{\tilde{H}} \triangleq \left[t, t + \tilde{H}\right]_{\mathbb{Z}}$.

The partial derivative of load real power terms with respect to the action variables is more difficult to compute since they are related implicitly through (2), (7a), (9), and (10a). However, we can calculate the partial derivatives numerically using the aforementioned relation as follows. The total differential of (2), (7a), (9), and (10a) is given as

$$dP_{i,k} = \sum_{j:i\to j} dP_{ij,k} - \sum_{h:h\to i} (dP_{hi,k} - r_{hi} dl_{hi,k}), \tag{27a}$$

$$dQ_{i,k} = \sum_{j:i\to j} dQ_{ij,k} - \sum_{h:h\to i} (dQ_{hi,k} - x_{hi} dl_{hi,k}), \tag{27b}$$

$$dv_{j,k} = dv_{i,k} - 2(r_{ij} dP_{ij,k} + x_{ij} dQ_{ij,k}) + (r_{ij}^2 + x_{ij}^2) dl_{ij,k}, \tag{27c}$$

$$dl_{ij,k} v_{i,k} + l_{ij,k} dv_{j,k} = 2P_{ij,k} dP_{ij,k} + 2Q_{ij,k} dQ_{ij,k}, \tag{27d}$$

$$dP_{i,k} = dP^{\text{dis}}_{i,k} - dP^{\text{ch}}_{i,k}, \tag{27e}$$

$$d\mathcal{S}_{i,k+1} = d\mathcal{S}_{i,k} + (\eta^{\text{ch}}_i \Delta t) dP^{\text{ch}}_{i,k} - \left(\frac{\Delta t}{\eta^{\text{dis}}_i}\right) dP^{\text{dis}}_{i,k} \tag{27f}$$

$$d\varsigma_{i,k+1} = d\varsigma_{i,k} - \tau_i dP_{i,k} \tag{27g}$$

wherein (27a) and (27b) hold for all $j \in \mathcal{N}$, (27c) and (27d) hold for all $(i,j) \in \mathcal{E}$, (27e) and (27f) hold for all $i \in \mathcal{N}^{\text{ESS}}$, (27g) holds for all $i \in \mathcal{N}^{\text{MT}}$. In terms of time steps, (27a)-(27e) hold for all $k \in \mathcal{T}_t^{\tilde{H}}$, while (27f) and (27g) hold for $k \in \mathcal{T}_{t+1}^{\tilde{H}-1}$. (27) is a homogenous system of linear equations in the differentials. This system can be used to numerically compute the required partial derivatives using the following result.

Lemma 2. *Under the assumption that the values of $\mathbf{s}_t$ and $\mathbf{a}_t$ on time step t are known, the partial derivative $\frac{\partial P_{i,k}}{\partial P}$ for $P \in \mathbf{a}_t$ and $i \in \mathcal{N}^{\text{L}}, k \in \mathcal{T}_t^{\tilde{H}}$ can be computed by setting $dP' = 0$ for all $P' \in \mathbf{a}_t$ such that $P' \neq P$ in* (27)*, and solving the resulting homogenous linear system. A solution always exists. Then, $\frac{\partial P_{i,k}}{\partial P} = \frac{dP_{i,k}}{dP}$.*

Proof. The method of calculation of the partial derivatives follows from the definition of a partial derivative. Consider variables $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{y} \in \mathbb{R}^n$ which are related through the implicit equation $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ where $\mathbf{F} : \mathbb{R}^m \times \mathbb{R}^n \mapsto \mathbb{R}^p$ is a smooth function. Let $\mathbf{x}$ be the independent variable, $\mathbf{y}$ be the dependent variable, and suppose we are interested in calculating $\frac{\partial y_i}{\partial x_j}$. The total differential of $\mathbf{F}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ is given as

$$\left[\frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}\right]^\top d\mathbf{x} + \left[\frac{\partial \mathbf{F}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}}\right]^\top d\mathbf{y} = \mathbf{0}. \tag{28}$$

In order to calculate $\frac{\partial y_i}{\partial x_j}$, we set $dx_g = 0$ for all $g \neq j$ (a partial derivative with respect to x_j means that any variable x_g with $g \neq j$ is assumed to be constant), and solve (28) for dx_j and $d\mathbf{y}$. From here, we have $\frac{\partial y_i}{\partial x_j} = \frac{dy_i}{dx_j}$.

Thus, it only remains to verify the existence of a solution. We first enumerate the number of equations in (27). (27a) and (27b) consist of $2\tilde{H}|\mathcal{N}|$ equations, (27c) and (27d) consist of $2\tilde{H}|\mathcal{E}|$ equations, (27e) consists of $\tilde{H}|\mathcal{N}^{\text{ESS}}|$ equations, (27f) consists of $(\tilde{H}-1)|\mathcal{N}^{\text{ESS}}|$ equations, and (27g) consists of $(\tilde{H}-1)|\mathcal{N}^{\text{MT}}|$ equations.

We now enumerate the number of variables, and the number of action variables which will be set to 0. The branch quantities $\{dP_{ij,k}\}$, $\{dQ_{ij,k}\}$, and $\{dl_{ij,k}\}$ consist of $2\tilde{H}|\mathcal{E}|$ variables, the bus quantities $\{dP_{i,k}\}$, $\{dQ_{i,k}\}$, and $\{dv_{i,k}\}$ consist of $2\tilde{H}|\mathcal{N}|$ variables, the charge & discharge powers $\left\{dP^{\text{ch}}_{i,k}\right\}$ and $\left\{dP^{\text{dis}}_{i,k}\right\}$ consist of $2\tilde{H}|\mathcal{N}^{\text{ESS}}|$ variables, the SoC $\{d\mathcal{S}_{i,k}\}$ consists of $(\tilde{H} -$

$1)|\mathcal{N}^{\text{ESS}}|$ variables, and remaining fuel quantity $\{d_{\varsigma_{i,k}}\}$ consists of $(\tilde{H}-1)|\mathcal{N}^{\text{MT}}|$ variables (in the last two cases, $\mathcal{S}_{i,t}$ and $\varsigma_{i,t}$ are prescribed due to knowledge of $\mathbf{s}_t$, so their differentials vanish). On the other hand, the action $\mathbf{a}_t$ contains $\tilde{H}\left(2|\mathcal{N}^{\text{MT}}|+2|\mathcal{N}^{\text{RES}}|+3|\mathcal{N}^{\text{ESS}}|\right)$ variables, and at most 1 of the differential of these variables can be nonzero. Thus the difference between the number of differentials allowed to assume nonzero values and the number of equations is

$$
\begin{aligned}
&\left\{3\tilde{H}|\mathcal{N}|+3\tilde{H}|\mathcal{E}|+2\tilde{H}|\mathcal{N}^{\text{ESS}}|+(\tilde{H}-1)|\mathcal{N}^{\text{ESS}}|+(\tilde{H}-1)|\mathcal{N}^{\text{MT}}|\right\}-\left\{2\tilde{H}|\mathcal{N}^{\text{MT}}|+3\tilde{H}|\mathcal{N}^{\text{ESS}}|\right.\\
&\left.+2\tilde{H}|\mathcal{N}^{\text{RES}}|-1\right\}-\left\{2\tilde{H}|\mathcal{N}|+2\tilde{H}|\mathcal{E}|+\tilde{H}|\mathcal{N}^{\text{ESS}}|+(\tilde{H}-1)|\mathcal{N}^{\text{ESS}}|+(\tilde{H}-1)\mathcal{N}^{\text{MT}}|\right\}\\
&=\tilde{H}\left(|\mathcal{N}|+|\mathcal{E}|-2\left(|\mathcal{N}^{\text{MT}}|+|\mathcal{N}^{\text{ESS}}|+|\mathcal{N}^{\text{RES}}|\right)\right)+1\\
&\overset{[a]}{=}\tilde{H}\left(2|\mathcal{N}|-1-2\left(|\mathcal{N}|-|\mathcal{N}^{\text{L}}|\right)\right)+1\\
&=\tilde{H}\left(2|\mathcal{N}^{\text{L}}|-1\right)+1\overset{[b]}{>}0,
\end{aligned}
$$

wherein $[a]$ follows from the fact that for any tree graph, $|\mathcal{N}|=|\mathcal{E}|+1$, while $[b]$ follows from the assumption that the MG comprises of at least one load bus. Thus, (28), after having appropriate differentials set to 0, is a homogenous system of linear equations with strictly larger number of variables than equations, and therefore, the system has a nonempty nullspace. Any vector in the nullspace of said system of linear equations is a solution of the system, which guarantees existence of a solution. □

Note that the values of the variables $v_{i,k}$, $l_{ij,k}$, $P_{ij,k}$, and $Q_{ij,k}$ are required for solving (28), which are not present in either $\mathbf{s}_t$ or $\mathbf{a}_t$. These variables are *observation variables*, as detailed in the following remark.

Remark 2 (Observation variables)**.** *Remark 1 provides us with a method to evaluate next state $\mathbf{s}_{t+1}$ deterministically, given values of current state $\mathbf{s}_t$ and current action $\mathbf{a}_t$. However, in the process of evaluating $\mathbf{s}_{t+1}$ we have to solve the DistFlow equations* (2)*, and a solution will also include the observation variables mentioned above. Since these variables do not participate in either the state or the action but are available as 'observations' during state transition, they are named appropriately.*

The partial derivative $\frac{\partial \mathbf{C}_t}{\partial \mathbf{a}_t}$ is easier to compute than $\frac{\partial \mathbf{C}_t}{\partial \mathbf{a}_t}$ since all the terms in $\mathbf{C}_t$ are either constants or action variables. Thus, all partial derivatives involved are trivial to compute.

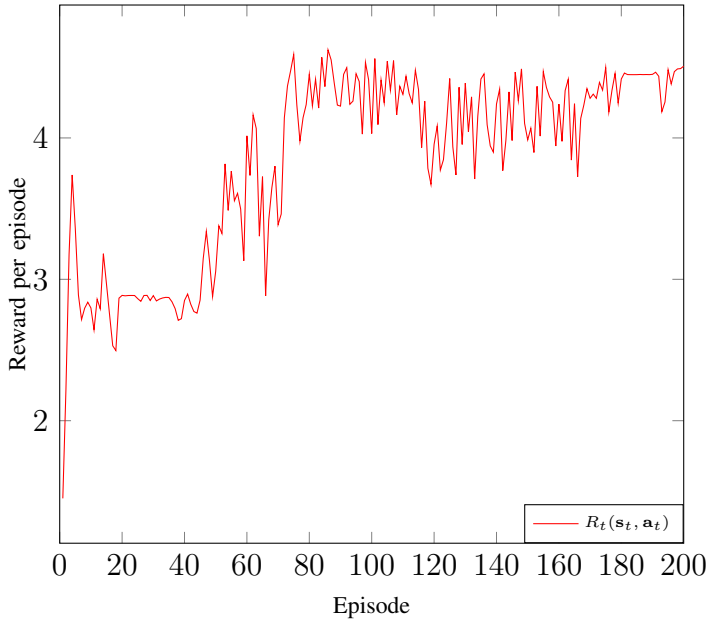


Fig. 1: Reward accumulated in CPO as a function of the number of episodes trained.

MGC Implementation: In case of CPO, the FNNs $\hat{f}_{\boldsymbol{\theta}}^{(1)}$ and $\hat{f}_{\boldsymbol{\theta}}^{(2)}$ represent the 'controller' since they take as input the current state, and their output prescribes a distribution over the action space, which is then used to determine the action to be taken. The training of the FNNs with CPO is carried out over the entire episodic time horizon $\mathcal{T}$ for multiple episodes. Each episode is started off with a slightly different value of $\mathcal{S}_i^{\text{init}}$ and E_i so as to make the controller robust with respect to initial conditions. Once training is complete, the controller can be implemented for real-time operation in the MGC.

V. Simulation Results

In this section, we compare the two methods presented, and validate their performance though simulations. We carry out all simulations on the 12-bus distribution system '`case12da`' in MATPOWER [22]. This test case prescribes the transmission line parameters, line flow limits, as well as real and reactive demands at each of the load buses. We consider the presence of a MT at bus 2, an ESS at bus 4, and a RES at bus 10. The RES forecast is assumed to have a gauussian distribution with mean μ_{RES} and variance σ_{RES}. The system parameters are presented in Table I (all power values are in MW).

All simulations were carried out in MATLAB on a PC with Intel i7-8700k CPU and 32GB RAM. Both MPC and CPO use a look-ahead window $\hat{H} = \tilde{H} = 3$ time steps. CPO parameters

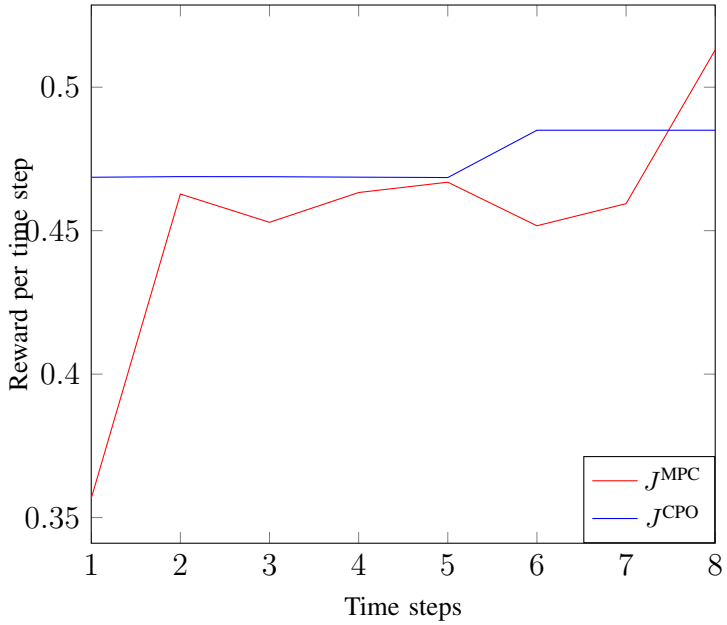
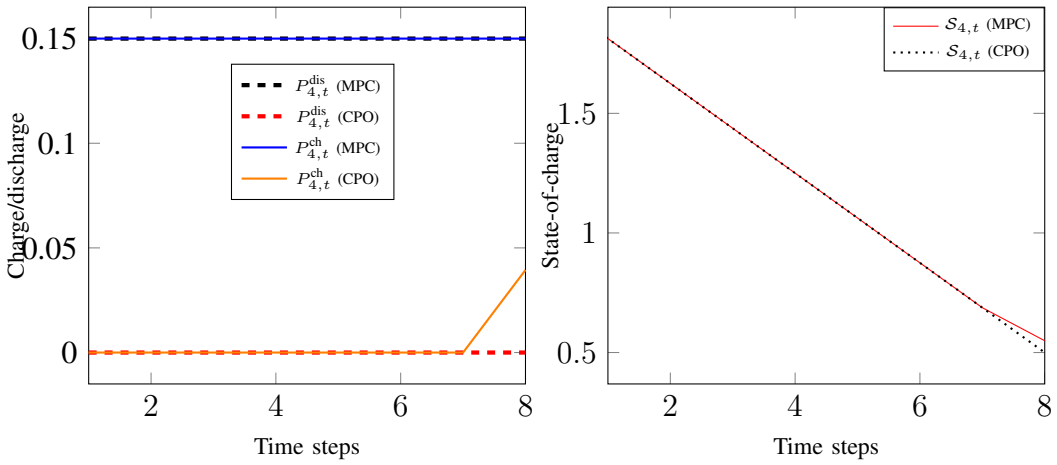


Fig. 2: Comparison of CPO reward and MPC cost over a single episode.



(a) (b)

Fig. 3: ESS charge/discharge power and SoC under MPC and CPO.

are provided in Table II. During implementation of CPO, the transition to state $\mathbf{s}_{t+1}$ from $\mathbf{s}_t$ under action $\mathbf{a}_t$ is achieved by solving equations (2a)-(2c), (7a), and (10a) while respecting the constraints (3)-(4), (15), and (17). The constraints on the variables in action $\mathbf{a}_t$ are implemented via the constraint function $\mathbf{C}(\mathbf{s}_t, \mathbf{a}_t)$. In order to implement the CC, we stipulate that $P^{\text{ch}}_{i,t} P^{\text{dis}}_{i,t} \leq 0$, and this along with the constraints $P^{\text{ch}}_{i,t} \in [0, \bar{P}^{\text{ch}}_i]$ and $P^{\text{dis}}_{i,t} \in [0, \bar{P}^{\text{dis}}_i]$ ensure that the CC always

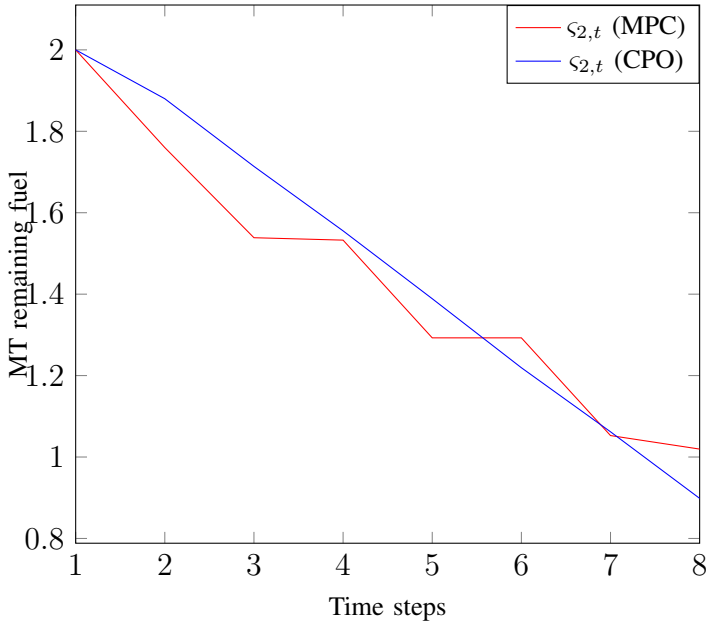


Fig. 4: Comparison of remaining MT fuel under CPO and MPC.

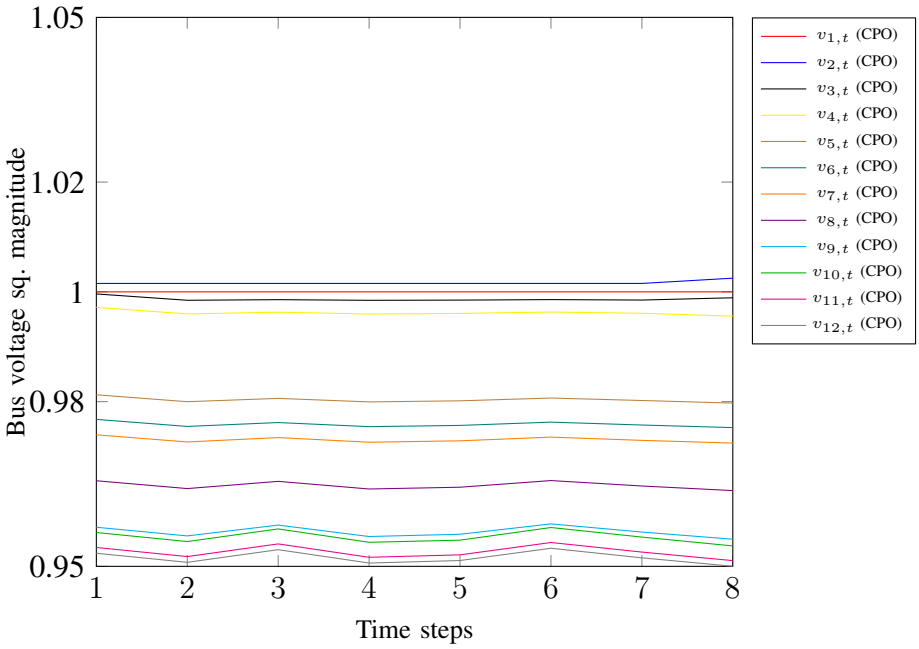


Fig. 5: Bus voltages under CPO. Here $\bar{v} = 1.05$ p.u. and $\underline{v} = 0.95$ p.u.

hold. Another important detail in the implementation of CPO is that on every time step, we evaluate the expectation terms in Theorem 1 by sampling the standard normal $\boldsymbol{\epsilon}$ multiple times as mentioned in Table II, followed by averaging the outcomes.

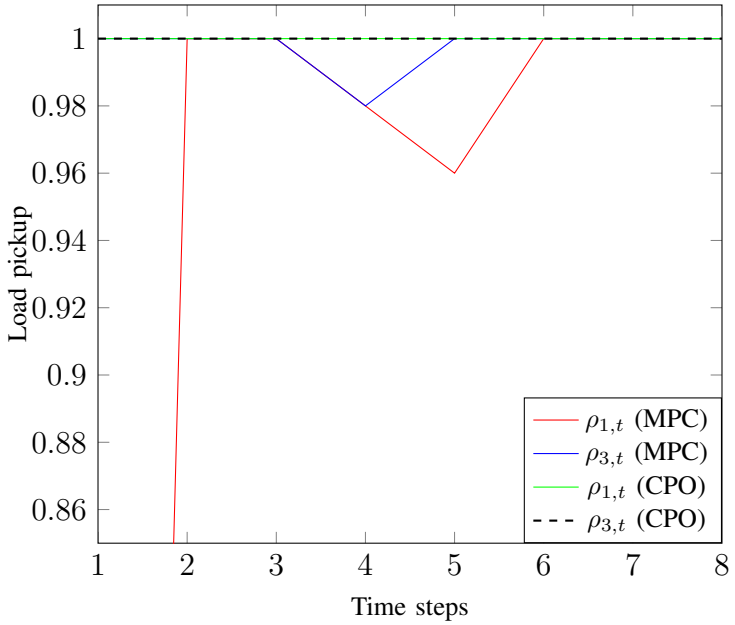


Fig. 6: Load pickup for two load buses 1 and 3 under MPC and CPO.

TABLE I: System Parameters

System Parameter	Value
T	8
$(\underline{v}, \bar{v})$	$(0.95, 1.05)$
$(\underline{P}^{\text{MT}}, \overline{P}^{\text{MT}}, \tau, E_0, \underline{P}^{\text{MT}}_{\text{rd}}, \bar{P}^{\text{MT}}_{\text{ru}})$	$(0, 0.3, 0.8, 2, 0.15, -0.1)$
$(\bar{P}^{\text{ch}}, \bar{P}^{\text{dis}}, \underline{\mathcal{S}}, \bar{\mathcal{S}}, \mathcal{S}^{\text{init}}, \eta^{\text{ch}}, \eta^{\text{dis}})$	$(0.2, 0.15, 1, 5, 2, 0.8, 0.8)$
$(\mu_{\text{RES}}, \sigma_{\text{RES}}, \epsilon, \Delta t)$	$(0.07, 0.01, 0.02, 1$ hr$)$
$(\bar{Q}^{\text{MT}}, \bar{Q}^{\text{ESS}}, \bar{Q}^{\text{RES}})$	$(0.3, 0.2, 0.1)$

TABLE II: Parameters of the CPO

CPO Parameter	Value
Discount Factor γ	0.9
Mean network $\hat{f}^{(1)}_{\boldsymbol{\theta}}(\mathbf{s})$	Two hidden layers (10 neurons each)
Var. network $\hat{f}^{(2)}_{\boldsymbol{\theta}}(\mathbf{s})$	Two hidden layers (20 neurons each)
Activation Function	$\tanh$
Trust region δ	0.1
$\boldsymbol{\epsilon}$ samples per time step	40

The simulation results are provided in Figures 1–6. Figure 1 shows the episodic reward accumulated over the training duration of 200 episodes. The episodic reward is low in the initial episodes, but as training progress, the reward accumulated increases, demonstrating the acquisition of better policies. In Figure 2, we compare the per time step reward of the trained CPO solution with the cost of the MPC solution. It can be seen that the performance of CPO is consistent and generates higher rewards over MPC in all but the terminal time step. Note that the CPO reward contains the discount factor γ as compared to the MPC cost, and we have accounted for the same while generating Figure 2. Figure 3a compares the charge/discharge powers of the ESS under CPO and MPC respectively. It is obvious that both the approaches let the ESS discharge at maximum allowable discharge power for the entirety of $\mathcal{T}$; however, MPC violates CC in the terminal time step by allowing for a nonzero ESS charge power. This CC violation is also visible in ESS SoC in Figure 3b under the two approaches. Thus Figure 3 demonstrates that CPO is better equipped to handle ESS CC than MPC.

Figure 4 shows the amount of fuel left in the MT under CPO and MPC. While the fuel remaining under CPO is higher for most of $\mathcal{T}$, said quantity is higher under MPC for the three terminal time steps. Figure 5 demonstrates the bus voltage squared magnitudes under CPO over $\mathcal{T}$. It can be seen that the resulting voltage values remain between the bound of 0.95 p.u. and $1.05p.u.$, as stipulated. Furthermore, bus 1, which is the slack bus, has its voltge held constant at 1 p.u.. Finally, we select two buses, *viz.* buses 1 and 3, to compare the load pickup under CPO and MPC in Figure 6. It can be seen that the load pickup under CPO is constant at 100% over the entire time horizon $\mathcal{T}$, while under MPC, the pickup ratio for both the buses falls below 100% on multiple time steps. Furthermore, MPC has a very low pickup ratio for bus 1 on time step 1. Thus, the simulation results demonstrate the superiority of the CPO approach over the MPC approach in solving the load restoration problem.

VI. Conclusion

In this paper we formulated the load restoration problem for an islanded MG, which contains sources of distributed power generation such as RES and MTs, as well as sources of power storage, such as ESS. We considered two methods of finding a solution to the problem which can be implemented on the MGC. We considered MPC, and proposed a convex relaxation of the load restoration problem which can be efficiently solved by the MGC. We also considered CPO which finds an optimal policy through training over multiple episodes. We compared the performance

of MPC and CPO on a 12-bus distribution system. Future work will involve coordinated load restoration using multiple MGs, and consideration of user data privacy concerns that may arise as a consequence of the cooperation.

REFERENCES

[1] H. Nazaripouya, "Power grid resilience under wildfire: A review on challenges and solutions," in *2020 IEEE Power Energy Society General Meeting*, 2020, pp. 1–5.

[2] M. Panteli and P. Mancarella, "Modeling and evaluating the resilience of critical electrical power infrastructure to extreme weather events," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1733–1742, 2017.

[3] "IEEE standard for the specification of microgrid controllers," *IEEE Std 2030.7-2017*, pp. 1–43, 2018.

[4] A. Vasilakis, I. Zafeiratou, D. T. Lagos, and N. D. Hatziargyriou, "The evolution of research in microgrids control," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 331–343, 2020.

[5] Z. Wang, C. Shen, Y. Xu, F. Liu, X. Wu, and C. Liu, "Risk-limiting load restoration for resilience enhancement with intermittent energy resources," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2507–2522, 2019.

[6] W. Liu, Z. Lin, F. Wen, and G. Ledwich, "A wide area monitoring system based load restoration method," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 2025–2034, 2013.

[7] R. Roofegari Nejad and W. Sun, "Distributed load restoration in unbalanced active distribution systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5759–5769, 2019.

[8] M. Tsai, "Development of an object-oriented service restoration expert system with load variations," *IEEE Transactions on Power Systems*, vol. 23, no. 1, pp. 219–225, 2008.

[9] L. Jin, R. Kumar, and N. Elia, "Model predictive control-based real-time power system protection schemes," *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 988–998, 2010.

[10] X. Zhang, G. Hug, J. Z. Kolter, and I. Harjunkoski, "Model predictive control of industrial loads and energy storage for demand response," in *2016 IEEE Power and Energy Society General Meeting*, 2016, pp. 1–5.

[11] S. C. Dhulipala, R. V. A. Monteiro, R. F. d. Silva Teixeira, C. Ruben, A. S. Bretas, and G. C. Guimarães, "Distributed model-predictive control strategy for distribution network volt/var control: A smart-building-based approach," *IEEE Transactions on Industry Applications*, vol. 55, no. 6, pp. 7041–7051, 2019.

[12] T. Wang, H. Kamath, and S. Willard, "Control and optimization of grid-tied photovoltaic storage systems using model predictive control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 1010–1017, 2014.

[13] W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3008–3018, 2020.

[14] H. Li, Z. Wan, and H. He, "Constrained ev charging scheduling based on safe deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427–2439, 2020.

[15] Q. Zhang, K. Dehghanpour, Z. Wang, F. Qiu, and D. Zhao, "Multi-agent safe policy learning for power management of networked microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1048–1062, 2021.

[16] J. Duan, Z. Yi, D. Shi, C. Lin, X. Lu, and Z. Wang, "Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid ac–dc microgrids," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5355–5364, 2019.

[17] J. Achiam, D. Held, A. Tamar, and P. Abbeel, "Constrained policy optimization," in *Proceedings of the 34th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, D. Precup and Y. W. Teh, Eds., vol. 70. PMLR, 06–11 Aug 2017, pp. 22–31.

[18] M. Farivar and S. H. Low, "Branch flow model: Relaxations and convexification—Part I," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2554–2564, 2013.

[19] J. Rawlings, "Tutorial overview of model predictive control," *IEEE Control Systems Magazine*, vol. 20, no. 3, pp. 38–52, 2000.

[20] V. Dvorkin, F. Fioretto, P. Van Hentenryck, P. Pinson, and J. Kazempour, "Differentially private optimal power flow for distribution grids," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2186–2196, 2021.

[21] L. Malagò and G. Pistone, "Information geometry of the gaussian distribution in view of stochastic optimization," in *Proceedings of the 2015 ACM Conference on Foundations of Genetic Algorithms XIII*, 2015, pp. 150–162.

[22] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, 2011.