Optimal Control of Microgrids with Multi-stage Mixed-integer Nonlinear Programming Guided *Q*-learning Algorithm

Yeliz Yoldas, Selcuk Goren, and Ahmet Onen

***Abstract*—This paper proposes an energy management system (EMS) for the real-time operation of a pilot stochastic and dynamic microgrid on a university campus in Malta consisting of a diesel generator, photovoltaic panels, and batteries. The objective is to minimize the total daily operation costs, which include the degradation cost of batteries, the cost of energy bought from the main grid, the fuel cost of the diesel generator, and the emission cost. The optimization problem is modeled as a finite Markov decision process (MDP) by combining network and technical constraints, and *Q*-learning algorithm is adopted to solve the sequential decision subproblems. The proposed algorithm decomposes a multi-stage mixed-integer nonlinear programming (MINLP) problem into a series of single-stage problems so that each subproblem can be solved by using Bellman's equation. To prove the effectiveness of the proposed algorithm, three case studies are taken into consideration: ① minimizing the daily energy cost; ② minimizing the emission cost; ③ minimizing the daily energy cost and emission cost simultaneously. Moreover, each case is operated under different battery operation conditions to investigate the battery lifetime. Finally, performance comparisons are carried out with a conventional *Q*-learning algorithm.**

***Index Terms*—Cost minimization, energy management system, microgrid, real-time optimization, reinforcement learning.**

I. Introduction

DISTRIBUTED energy resources (DERs) such as wind power, solar power, and energy storage system (ESS) are viewed as a solution due to the reduction in primary energy reserves and ever-increasing load demand. Thus, microgrid plays a crucial role in the integration of DERs into the future electric power grids. Despite many advantages of microgrid [1], there are several technical challenges such as stability and reliability issues due to the natural uncertainty and unpredictability of renewable energy sources (RESs) [2]. The management of power system operation is already quite complex because instability and unreliability make it very difficult to maintain a balance between supply and demand of energy in real-time operation. When integrating RESs into the power systems, the complicated systems get even more complex, rendering the management of power systems including DERs a real challenge. It is crucial to have an appropriate energy management in place for the success of such complicated power systems. A microgrid energy management system (EMS) plays a critical role in economic, sustainable and reliable operation by providing the optimal coordination between conventional energy resources, RESs, ESSs, and consumers [3].

The existing studies in the literature can be classified according to the objectives of EMSs or the optimization approaches used. Microgrid energy management has been studied for many purposes such as operation cost reduction [4]-[7], maximization of battery life and renewable energy penetration [8], environmental pollution and operation cost reduction [9]-[11], and improvement of stability and reliability of the system [12], [13]. For example, while the main objective in [5] is to minimize the total operation cost of microgrid including the fuel cost of power generators, the cost of operation and maintenance, the cost of purchasing electricity from main grid and penalties on the curtailment of renewable energy and load shedding, [8] intends to maximize the reliability and customer satisfaction.

The intermittent nature of RESs and nonlinear characteristics of other devices make it inevitable to have an optimization process in place: trivial straightforward decisions result in severely suboptimal management systems. In this regard, mixed-integer linear programming (MILP) and mixed-integer nonlinear programming (MINLP) are frequently employed [14], [15]. Unfortunately, even though the algorithms to get exact solutions to integer programming problems have improved significantly through the years, the state-of-the-art solution algorithms still rely on implicit enumeration, which has large computational burden for practical problems. Hence, several heuristic algorithms such as metaheuristic al-

Manuscript received: July 22, 2020; accepted: November 2, 2020. Date of CrossCheck: November 2, 2020. Date of online publication: November 26, 2020.

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) (No. 215E373), Malta Council for Science and Technology (MCST) (No. ENM-2016-002a), Jordan The Higher Council for Science and Technology (HCST), Cyprus Research Promotion Foundation (RPF), Greece General Secretariat for Research and Technology (GRST), Spain Ministerio de Economía, Industria y Competitividad (MINECO), Germany and Algeria through the ERANETMED Initiative of Member States, Associated Countries and Mediterranean Partner Countries (3DMgrid Project ID eranetmed_energy-11-286).

Y. Yoldas (corresponding author) and A. Onen are with the Department of Electrical and Electronics Engineering, Abdullah Gul University, Kayseri, 38080, Turkey (e-mails: yeliz.yoldas@agu.edu.tr; ahmet.onen@agu.edu.tr).

S. Goren is with the Department of Industrial Engineering, Abdullah Gul University, Kayseri, 38080, Turkey (e-mail: selcuk.goren@agu.edu.tr).

DOI: 10.35833/MPCE.2020.000506

gorithms [16], [17], model predictive control (MPC) algorithms [18]-[20], and artificial intelligence algorithms [21], [22] are proposed to find near-optimal solution. In [23], a stochastic MPC algorithm is proposed to cope with the uncertainties in both supply and demand. In [24], a scenario-based stochastic particle swarm optimization (PSO) algorithm is employed. The authors claim that the balance between the solution accuracy and the computational burden is improved using scenario modeling and scenario reduction methods. Scenario reduction methods are used to make the algorithm faster, thus the system performance may depend on the scenario reduction method used. One of the commonly used methods as a metaheuristic is the genetic algorithm (GA). In [25], the author develops a memory-based GA (MGA) that tries to share the demand among several power generation sources with minimum production cost. The main limitation of these heuristic algorithms is that they cannot guarantee optimality, nor can they provide bounds on the amount of suboptimality, i.e., the optimality gap.

The determination of the optimal operation involves a sequential decision-making process to tackle the uncertainty in weather-related generation units, demand, electricity price, and problem arising from the integration of variable power sources into the main grid. Thus, energy management for a microgrid becomes unavoidable to enable stable and reliable operation, seek optimal dispatch and maximize its performance. To solve these issues, the adaptive and intelligent methods are essential, especially for a large-scale microgrid. Reinforcement learning (RL) is a promising computational method to solve the stochastic sequential decision-making problems, in which a learning agent learns what actions to take by interacting with its environment to maximize a reward signal [26]. In this method, the agent is not told what to do in the current state, but instead needs to try the actions to find out which one gives the maximum reward. However, the RL suffers from "curse of dimensionality" as the complexity of microgrid system increases. Due to the fact that coarse-grained discretization causes information loss, fine-grained discretization is required, and that causes the "curse of dimensionality" problem. Several studies have been published in the literature with RL. In [27], an RL-based optimal control method is proposed to improve the transient performance of hybrid microgrid systems. In [28], a well-known batch RL, i.e., fitted Q-iteration, for residential demand response is suggested. In [29], the fitted Q-iteration is also used on a residential scale to minimize the amount of imported power from the main grid.

In [30], a dynamic pricing strategy using Q-learning (QL) algorithm is proposed by considering hierarchical electricity market. The aim is to find a financial balance between profits of service providers and costs of customers. However, customers, service providers and main grids constitute the whole system. In [31], a strategic bidding is proposed by using QL algorithm. In this study, customers need a bidding strategy to maximize their long-term profit. In [32], a two-step ahead RL method is proposed for a simple microgrid system to plan battery schedules without considering the detailed mathematical model of devices. In [33]-[35], a multi-agent RL method is applied to a microgrid considering the uncertainties. Moreover, operation cost reduction is targeted with RL method in [36], [37].

This paper proposes an EMS that employs an MINLP guided QL algorithm for microgrid operation in a stochastic and dynamic environment to tackle the aforementioned challenges. The main feature of the proposed algorithm is that the "curse of dimensionality" can be handled without coarse-grained discretization. The proposed algorithm decomposes the multi-time horizon optimization problem into sub-problems based on consecutive time-indexed periods. Then each sub-component at each time is solved by MINLP method. The purpose of the study is to minimize the total daily operation costs which include the degradation cost of batteries, the cost of energy bought from the main grid, the fuel cost of the diesel generator (DG), and the emission cost. Compared with prior studies, e.g., [10], [30], [38], the main contributions of the paper are as follows:

1) The proposed real-time EMS is formulated as a Markov decision process (MDP) problem, where the solar energy, DG and battery are considered. The proposed algorithm has been developed to provide efficient energy management of a real microgrid pilot of the Malta College of Arts, Science and Technology (MCAST) by considering constraints of the network model and technical model.

2) This paper tackles the problem with multiple smaller sub-problems by decomposing multiple time period operation cost optimization over a finite horizon. Thus, MINLP sub-problems can be solved effectively.

3) In order to reduce the dependency on the forecasted information, the historical data are used offline to deal with uncertainties of load demand and photovoltaic (PV).

4) The proposed algorithm enables finding optimal solutions without applying an approximation method, which enhances the performance of QL-based optimization with large state space.

The remainder of this paper is organized as follow. Section II presents the mathematical model description of the microgrid model and the devices used. Section III formulates the real-time operation problem of microgrid as an MDP. The proposed algorithm and the QL algorithm are given in Section IV. The simulation environment using real data and result analysis are shown in Section V. Finally, Section VI draws the conclusions.

II. Microgrid Model Description

The structure of the microgrid system is illustrated in Fig. 1, where PCC stands for point of common coupling and SS stands for substation. The system is comprised of solar PV arrays (63 kW in total), a DG (300 kW), lithium-ion batteries (300 kWh capacity in total) and loads. This paper assumes that the microgrid operates in grid-connected mode. A finite time horizon of the microgrid operation is considered as $t = \{0, \Delta t, 2\Delta t, \ldots, T - \Delta t, T\}$, where $\Delta t = 5$ min is the time interval and $T = 24$ hours.

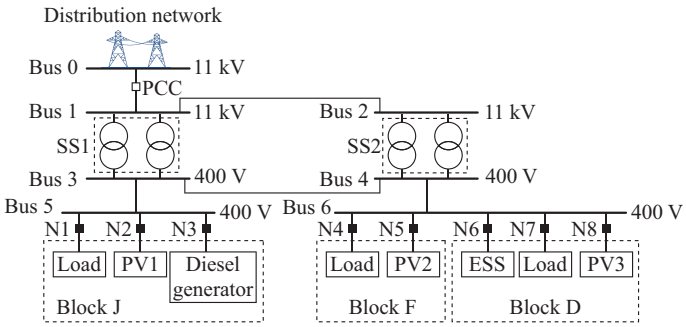

Fig. 1. Schematic diagram of microgrid.

A. Battery Model

The ESS is one of the core parts of the microgrid system, which can improve the performance of the microgrid system. Since the initial investment cost of batteries is high, it is crucial to extend the battery life. The battery cycle life is directly related to the depth of discharge (DoD). The cycle life data are given by the battery manufacturer in the form of total cycle number with respect to the DoD. The relationship between expected cycle life and DoD is exponential for the lithium-ion battery as given in (1).

$$L(D)=D^{a}\mathrm{e}^{b} \tag{1}$$

where D is the DoD in percentage at which the battery is cycled; $L(D)$ is the average cycle number at that particular D; and a and b are the battery dependent coefficients. From the logarithmic fitted curve between DoD and cycle life specified in the data sheet of the battery used, these coefficients are found as $a=-1.24$, $b=7.043$. From the fitted curve, the battery wear cost can be calculated as

$$C_{W}=\frac{C_{inv}}{2E_{\max}L(D)D\eta^{d}\eta^{c}} \tag{2}$$

where C_{inv} is the capital cost of battery; $E_{\max}$ is the total capacity of battery; and η^{c} and η^{d} are the charging and discharging efficiencies, respectively.

The operation cost of battery based on wear cost is written as

$$C_{bat,t}=C_{W}P_{bat,t}\Delta t \tag{3}$$

where $P_{bat,t}$ is the charging or discharging power of the battery at time t.

At any given time, the state of charge (SOC) of the lithium-ion battery system should be within a certain range. It can be expressed as

$$SOC_{\min}\le SOC_{t}\le SOC_{\max} \tag{4}$$

where $SOC_{\min}$ and $SOC_{\max}$ are the lower limit and upper limit of SOC, respectively. The charging and discharging states, charging and discharging power limits, and SOC formulation of the lithium-ion battery are given respectively as follows:

$$\begin{cases}u^{c}_{bat,t}+u^{d}_{bat,t}\le 1\\ u^{c}_{bat,t},u^{d}_{bat,t}\in\{0,1\}\end{cases} \tag{5}$$

$$0\le P^{d}_{bat,t}\le u^{d}_{bat,t}P^{d}_{\max} \tag{6}$$

$$0\le P^{c}_{bat,t}\le u^{c}_{bat,t}P^{c}_{\max} \tag{7}$$

$$SOC_{t}=\begin{cases}SOC_{t-\Delta t}-\dfrac{P^{d}_{bat,t}\Delta t}{\eta^{d}E_{\max}} & P^{d}_{bat,t}>0\\ SOC_{t-\Delta t}+\dfrac{\eta^{c}P^{c}_{bat,t}\Delta t}{E_{\max}} & P^{c}_{bat,t}>0\end{cases} \tag{8}$$

where $u^{c}_{bat,t}$ and $u^{d}_{bat,t}$ are the charging and discharging states of the battery, respectively; $P^{c}_{bat,t}$ and $P^{d}_{bat,t}$ are the charging and discharging power of the battery, respectively; and $P^{c}_{\max}$ and $P^{d}_{\max}$ are the maximum charging and discharging power of the battery, respectively. η^{c} and η^{d} are both assumed to be 95%, according to the practical situation of the MCAST system.

B. DG

The hourly fuel consumption FC_{t} of a DG is modeled as a linear function, which is based on data provided by the manufacturer.

$$FC_{t}=F_{1}P_{rated}+F_{2}P_{dg,t} \tag{9}$$

where F_{1} and F_{2} are the coefficients of fuel consumption function, which are set as 0.0183 and 0.22, respectively; and P_{rated} and $P_{dg,t}$ are the rated power and the actual output power of DG, respectively.

The power limits of DG are imposed as

$$kP_{rated}\le P_{dg,t}\le P_{rated} \tag{10}$$

where k is set to be 0.3 based on the suggestion of manufacturers.

The fuel cost of DG at time step t can be calculated as

$$C_{dg,t}=C_{fuel}\cdot FC_{t}\cdot\Delta t \tag{11}$$

where C_{fuel} is the fuel cost.

C. Main Grid

The power transaction between main grid and microgrid should be constrained as

$$-P^{\max}_{grid}\le P_{grid,t}\le P^{\max}_{grid} \tag{12}$$

where $P_{grid,t}$ is the active power exchange between microgrid and main grid at time t; and $P^{\max}_{grid}$ is the maximum active power that can be exported to and imported from the main grid.

The cost related to the power transaction at time step t is

$$C_{grid,t}=prc_{t}\cdot P_{grid,t}\Delta t \tag{13}$$

where prc_{t} is the real-time electricity price at time step t.

D. AC Power Flow

The power flow limits in each branch ij are considered as

$$P_{ij,t}=\frac{\left|V^{2}_{i,t}\right|\cos\left(\theta_{ij}\right)}{\left|Z_{ij}\right|}-\frac{\left|V_{i,t}\right|\left|V_{j,t}\right|\cos\left(\delta_{i,t}-\delta_{j,t}+\theta_{ij}\right)}{\left|Z_{ij}\right|} \tag{14}$$

$$Q_{ij,t}=\frac{\left|V^{2}_{i,t}\right|\sin\left(\theta_{ij}\right)}{\left|Z_{ij}\right|}-\frac{\left|V_{i,t}\right|\left|V_{j,t}\right|\sin\left(\delta_{i,t}-\delta_{j,t}+\theta_{ij}\right)}{\left|Z_{ij}\right|} \tag{15}$$

$$P^{2}_{ij,t}+Q^{2}_{ij,t}\le\left(S^{\max}_{ij}\right)^{2} \tag{16}$$

where $i,j\in\{1,2,\ldots,N_{b}\}$, and N_{b} is the total number of buses; $P_{ij,t}$ and $Q_{ij,t}$ are the active and reactive power flows of branch ij, respectively; $\left|V_{i,t}\right|$ and $\delta_{i,t}$ are the voltage ampli-

tude and angle at bus i, respectively; $|Z_{ij}|$ and θ_{ij} are the impedance magnitude and corresponding phase angle of branch ij, respectively; and $S_{ij}^{\max}$ is the maximum complex power flow of branch ij.

The transmission capacity limit of power cables is also considered as

$$P_{ij,t} \le P_{ij}^{\max} \tag{17}$$

where $P_{ij}^{\max}$ is the maximum power flow limit from bus i to bus j.

The voltage amplitude limit is bounded by

$$V_i^{\min} \le |V_{i,t}| \le V_i^{\max} \tag{18}$$

where $V_i^{\min}$ and $V_i^{\max}$ are the minimum and maximum voltage magnitudes of bus i, respectively.

The power balance equation is also considered as

$$P_{pv,t} + P_{grid,t} + P_{dg,t} + P_{bat,t}^{d} + P_{bat,t}^{c} = P_{ij,t} + P_{L,t} \tag{19}$$

where $P_{pv,t}$ is the total active power output of PV arrays; and $P_{L,t}$ is the total active load demand.

E. Emission Cost Calculation

Toxic gas externalities including CO_2, NO_x and SO_2 must be considered as cost function to reduce the greenhouse gas effect. The mass of the three gases are calculated with mathematical equation of the generated power of the DG and electricity grid as

$$C_{em,t} = \sum_{k=1}^{N_{em}} \sum_{i=1}^{N_{ps}} EC_k \cdot EF_{i,k} \cdot P_{i,t} \tag{20}$$

where N_{em} is the number of emission types (CO_2, NO_x, SO_2); N_{ps} is the number of power sources that release the toxic gases (main grid and DG); EC_k is the externality cost of emission type k; $EF_{i,k}$ is the emission factor of power source i and the emission type k; and $P_{i,t}$ is the power output of power source i.

III. MDP Model for Real-time Scheduling of Microgrid

In the MDP model, there are four components: state variables, decision (action) variables, state transitions and rewards. The state variables denote the current state of the system and the basis for making operation decisions. The decision variables identify the choices and the agent selects an action from a set of available actions, which is then sent to the environment. A time step later, the agent receives a reward which is an evaluation of taken actions, and the environment responds to these actions as a new state transition. Clearly, MDP allows us to predict the next state and reward given the current state and action. The next state depends only on the states and actions at time t instead of the previous history.

The centralized EMS collects two types of information to make optimal decisions: the first is the historical data of PV generation and demand at the annual, monthly, daily, hourly and minute levels, and the second is the real-time information from microgrid assets including SOC of the battery, electricity price and the output of the battery and DG. Based on this information, EMS decides the power outputs of DG, PV and battery, and the power exchange between main grid and microgrid to achieve the objectives.

A. State Variables and Decision (Action) Variables

The state variables S_t at time t include SOC_t, available active power outputs of PVs $P_{pv1,t}^{a}, P_{pv2,t}^{a}, P_{pv3,t}^{a}$, total active load demand $P_{L,t}$ and real-time electricity price prc_t. Hence, S_t can be given as

$$S_t = \left\{SOC_t, P_{pv1,t}^{a}, P_{pv2,t}^{a}, P_{pv3,t}^{a}, P_{L,t}, prc_t\right\} \tag{21}$$

The decision variable x_t at time t of the problem can be given as

$$x_t = \left\{P_{bat,t}^{d}, P_{bat,t}^{c}, P_{dg,t}\right\} \tag{22}$$

The transition function for the battery SOC can be formulated as

$$SOC_{t+\Delta t} = SOC_t + \left(\frac{P_{bat,t}^{d}}{\eta^{d}} - P_{bat,t}^{c}\eta^{c}\right)\Delta t \tag{23}$$

B. Objective Function

The total cost of microgrid is considered as a trade-off between power generation cost and emission cost caused by the grid and DG. In this study, three case studies are considered including individual minimization of power generation cost, individual minimization of emission cost, and simultaneous minimization of power generation cost and emission cost. Thus, the objective function can be expressed as

$$C_t(S_t, x_t) = C_{bat,t}(S_t, x_t) + C_{dg,t}(S_t, x_t) + C_{grid,t}(S_t, x_t) + C_{em,t}(S_t, x_t) \tag{24}$$

where x_t is an action variable; and (S_t, x_t) is the state-action pair.

IV. Proposed Optimization Model

A. QL Algorithm

QL algorithm is an efficient algorithm of RL to solve the MDP-based optimization problem (24) without an explicit environment model. The objective of the QL algorithm is to seek the optimal policy by maximizing the expected discounted reward of actions based on the given states. The output of the Q-value table for a state variable S_t and an action variable x_t is represented as $Q(S_t,x_t)$. In the QL algorithm, the Q-values of each action x_t performed in a state S_t can be updated recursively using Bellman's action-value function as follows:

$$Q(S_t, x_t) \leftarrow Q(S_t, x_t) + \alpha\left(R_{t+\Delta t} + \gamma \max_{x' \in A} Q(S_{t+\Delta t}, x') - Q(S_t, x_t)\right) \tag{25}$$

where $\gamma \in [0, 1)$ is a discount parameter; α is the learning parameter, which decreases over time interval Δt in the suitable way; $R_{t+\Delta t}$ is the immediate reward when the agent takes action x_t at state S_t; and A is the set of feasible actions. The immediate reward is defined as the daily cost of the microgrid system.

The basic principle behind the QL algorithm is that the agent takes an action based on the ε-greedy policy, which is a way to choose an action from a set of feasible actions. The

agent selects the best action with probability $1-\varepsilon$ or takes actions randomly with probability ε to discover new actions. The taken action gives rise to a change of the environmental state, thus the agent makes a transition to a new state and observes the immediate reward from taking action x_t in state S_t. Then, Q-value for a given state S_t and action x_t is updated. The optimal value of a state at each iteration is obtained by computing the maximum value, as shown in (26).

$$Q^*\left(S_t,x_t\right)\triangleq \max_{x\in A} Q\left(S_t,x\right) \tag{26}$$

B. Proposed Algorithm

The main advantage of the QL algorithm is that it does not need any environment model and can handle uncertainties and stochastic transitions without requiring full information of the system. However, it can be inefficient for large state-action space and cannot be applied easily to continuous state-action spaces involved in our problem. The simplest solution to continuous working space is to discretize the space. Making discretization at smaller intervals can compensate the changes of the system, but state-action pair number will increase exponentially. In this study, after making discretization with larger intervals, each subproblem at each time step is solved by MINLP method using DICOPT solver of General Algebraic Modeling System (GAMS) to get precise results. Thus, the problem can be handled with MINLP guided QL algorithm without discretization in smaller way. By this way, it can overcome the challenges and find a more precise solution instead of an approximated value.

The complete training process of the proposed algorithm using a combination of QL algorithm and MINLP optimization is presented in Fig. 2.

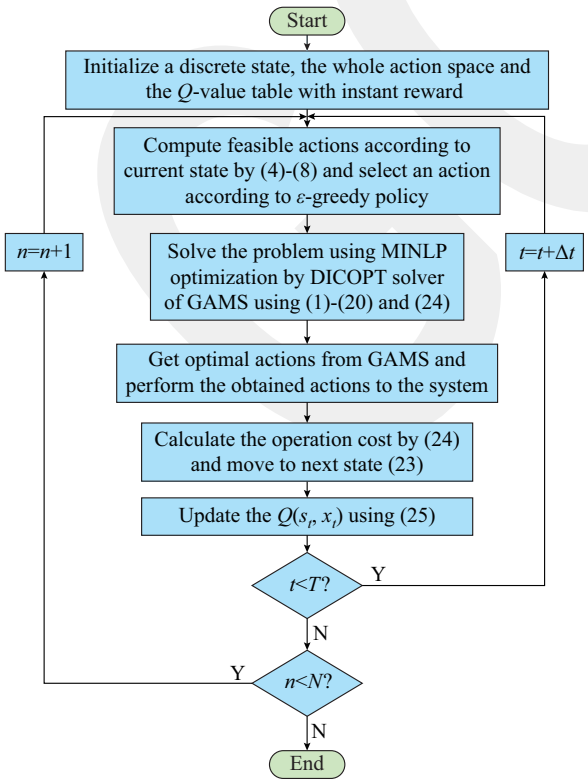

Fig. 2. Flowchart of training process.

In the flowchart, at the beginning, a discrete state, the whole action space of the system and Q-value table are initialized. Instead of storing every state-action pair of the system, iteration is started by choosing only a discrete state. Additionally, the $Q(S_t,x_t)$ values of each state-action pairs are initialized with total discounted reward r_0 with $\gamma=0$ to reduce the convergence time, which can be obtained as the instant reward at time step 0 before the learning process starts [36]. Then, the iteration starts by finding feasible actions at that state. An action is then selected from the feasible action set using ε-greedy policy. The selected actions are sent to the GAMS to solve the economic dispatching problem as MINLP. Thus, GAMS that uses the discretized actions at large intervals as inputs will give us optimal actions that minimize the cost function. The obtained optimal actions are then performed in the microgrid system. In the next step, the objective function at time t is calculated using (24). Then, $Q(S_t,x_t)$ value and time are updated, respectively. Finally, the number of episode n is updated and if $n<N$, where N is the total number of episodes, the system goes to the next episode.

V. Numerical and Result Analysis

A. Simulation Environment

The microgrid is equipped with a 300 kW/375 kVA DG, 3×21 kW solar generators, and 150 kW/300 kWh battery, as shown in Fig. 1. Moreover, the profiles of load demand and the electricity price are shown in Figs. 3 and 4, respectively. The parameters of DG and lithium-ion battery are given in Tables I and II, respectively. The parameters of distribution lines are given in Table III.

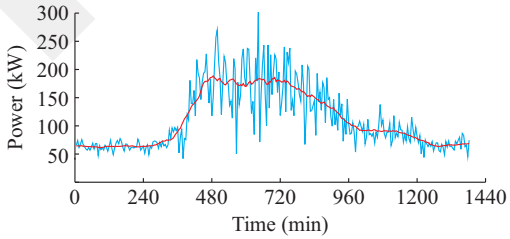

Fig. 3. Profiles of load demand.

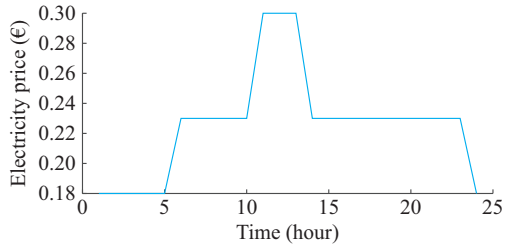

Fig. 4. Profile of electricity price.

TABLE I
Parameters of DG

Parameter	Value	Parameter	Value
P_{rated} (kW)	300	k	0.3
F_1 (L·h^{-1}·kW^{-1})	0.0183	C_{fuel} (€/L)	1.1
F_2 (L·h^{-1}·kW^{-1})	0.22		

TABLE II
PARAMETERS OF LITHIUM-ION BATTERY

Parameter	Value	Parameter	Value
E_{max} (kWh)	300	P^d_{max} (kW)	50
Cycle life	2700 (50% DoD)	P^c_{max} (kW)	40
η^d, η^c	0.95, 0.95	a	−1.24
SOC_{min} (%)	50	b	7.043
SOC_{max} (%)	100	Battery cost (€/kWh)	220

TABLE III
PARAMETERS OF DISTRIBUTION LINES

Line		Resistance (mΩ)	Reactance (mΩ)
From	To		
Bus 0	Bus 1	129.000	78.225
Bus 1	Bus 2	19.737	11.969
Bus 3	Bus 4	11.536	12.208
Bus 3	Bus 5	3.770	3.989
Bus 4	Bus 6	3.770	3.989
Bus 5	N1	3.770	3.989
Bus 5	N2	3.770	3.989
Bus 5	N3	3.770	3.989
Bus 6	N4	9.048	9.550
Bus 6	N5	9.048	9.550
Bus 6	N6	4.901	5.186
Bus 6	N7	6.786	7.181
Bus 6	N8	6.786	7.181

For all simulation cases, since the SOC changes from 40% to 60%, it is discretized into 70 to 130 states. The discharging/charging power of the battery, the output power of DG, the generated PV power and the load demand are discretized into 10/8 states, 7 states, 5 states, 60 states, respectively. Table IV demonstrates the externality costs and emission factors of main grid and DG. The optimization horizon of all simulations is set to be 24 hours, and $\Delta t = 5$ min. Although the time interval is 5 min in the operation of the algorithm, the results in all cases are drawn with a time interval of 1 hour so that the graphics can be clearly seen. All studies have been simulated using MATLAB 2020 and GAMS 24.9.2 on 64-bit Linux based computer with 250 GB of RAM and 2.10 GHz Intel® Xeon® processor.

TABLE IV
PARAMETERS OF EXTERNALITY COSTS AND EMISSION FACTORS OF DG AND MAIN GRID

Emission type	Externality cost (€/kg)	Emission factor of DG (kg/kWh)	Emission factor of main grid (kg/kWh)
CO_2	0.0308	0.7430000	0.922000
SO_2	2.1810	0.0004045	0.003583
NO_x	9.2527	0.0093600	0.002295

B. Case Studies

1) Case 1: Minimize Operation Cost Without Emission Cost

In this case, the main objective is to minimize the operation costs of battery, DG and main grid. Emission costs are not considered. When the battery is operated at SOC of 50%, the simulation results are illustrated in Fig. 5.

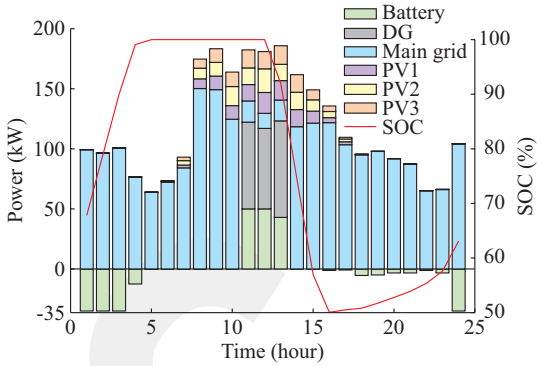

Fig. 5. Output power of all sources for Case 1.

It can be observed from Fig. 5 that the battery stores energy during 0th-4th hour. Then the power generated by PV is dispatched. When operation cost of DG is less than electricity price, DG is turned on between 11th-13th hour. Since DG is operated at minimum 90 kW, the power can be bought from the main grid in that situation. Table V shows the effect of battery SOC on the average daily operation cost.

TABLE V
SIMULATION RESULTS OF PROPOSED ALGORITHM COMPARED WITH QL ALGORITHM FOR CASE 1

SOC (%)	Algorithm	Total emission (kg/kWh)	Emission cost (€)	Battery throughput (kWh)	Daily energy cost (€)
40	Proposed	2139.0457	129.6250	150.0000	547.3606
	QL	2161.4613	130.8125	142.0833	554.4926
45	Proposed	2150.2869	130.2584	150.0000	550.6309
	QL	2161.5040	132.0833	142.0833	556.9808
50	Proposed	2153.6972	133.8653	142.9167	555.4273
	QL	2163.9308	134.1898	142.0833	558.2455
55	Proposed	2163.8568	134.1640	129.5833	555.6116
	QL	2165.2087	135.3101	128.2500	558.5268
60	Proposed	2171.6735	136.1739	115.4167	558.2421
	QL	2178.3315	137.3908	114.0000	561.9658

When the battery is operated at SOC of 40%, the average daily operation cost is €547.3606, while it goes up to €558.2421 at SOC of 60%. According to this table, the proposed algorithm performs better than the QL algorithm on the average daily operation cost. If we assume that the battery is operated at that power level as average during a year, by changing SOC level from 40% to 60%, the battery life increases from 6.71 years to 9.65 years. In this way, the capital cost of battery is deferred as 2.94 years, if we assume the lithium-ion battery life as average ten years and the total capital cost of battery as (300×220) €66000. The annual cost during the whole battery life is €6600. Thus, the net saving of battery renewal is (2.94×6600) €19404.

2) Case 2: Minimize Operation Cost with Emission Cost

DG is not used in this case, because the total cost of DG including fuel cost and emission cost is higher than that of the main grid. The simulation results are illustrated in Fig. 6. As in the Case 1, the battery charges at low electricity price intervals and discharges at peak price intervals to support the load demand. Table VI shows the results of the proposed algorithm and QL algorithm according to different SOC values. It can be seen from Table VI that the proposed algorithm works better than QL algorithm. Comparing Table VI with Table V, it can be seen that for the proposed algorithm, the emission cost decreases by 5.68% (7.97%), while the daily energy cost increases by 0.66% (0.37%) with SOC of 40% (60%).

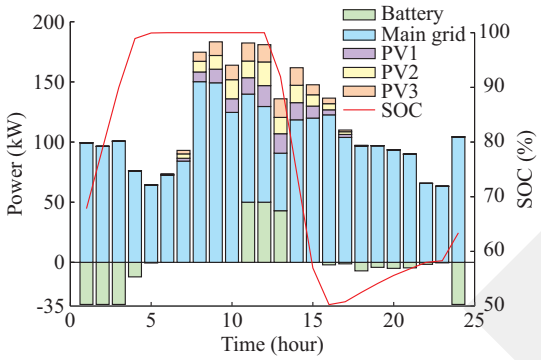

Fig. 6. Output power of all sources for Case 2.

TABLE VI
SIMULATION RESULTS OF PROPOSED ALGORITHM COMPARED WITH QL ALGORITHM FOR CASE 2

SOC (%)	Algorithm	Total emission (kg/kWh)	Emission cost (€)	Battery throughput (kWh)	Daily energy cost (€)
40	Proposed	2167.8582	122.2626	150.0000	550.9850
	QL	2192.7543	123.6544	141.6667	557.6795
45	Proposed	2179.9132	122.9418	150.0000	554.5246
	QL	2194.6874	123.7633	141.6667	558.7408
50	Proposed	2189.6389	123.4897	142.9167	556.5094
	QL	2199.1335	124.0138	141.6667	559.6482
55	Proposed	2202.1638	124.1954	129.5000	558.7096
	QL	2208.6829	124.5518	128.2500	560.8216
60	Proposed	2211.7894	125.3180	115.8333	560.3257
	QL	2220.3729	125.7182	114.0000	563.0461

3) Case 3: Minimize Emission Cost

Figure 7 shows the dispatched power by the proposed algorithm considering the goal of reducing the emission cost. According to figure, the microgrid system uses the maximum capacity of renewable sources since they have no emission. Since the emission cost of main grid is less than that of DG, the whole day demand is supplied by the main grid, and the battery also contributes to supply the demand. It can be observed clearly from Fig. 7 that the battery charging and discharging states are constantly changing, which adversely affects battery life. Based on Table VII, the battery lifetime can be calculated, which varies from 5.32 to 7.03 years as the SOC value of the battery increases. Thus, the capital cost of the battery renewal is deferred for 1.71 years and the net saving is (1.71×6600) €11286. Comparing the net savings of Case 3 with that of Case 1, the net saving decreases by 41.84%.

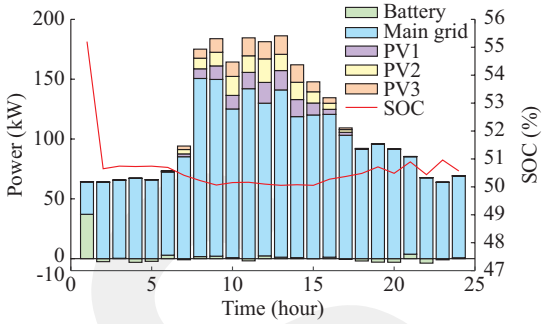

Fig. 7. Output power of all sources for Case 3.

TABLE VII
SIMULATION RESULTS OF PROPOSED ALGORITHM COMPARED WITH QL ALGORITHM FOR CASE 3

SOC (%)	Algorithm	Total emission (kg/kWh)	Emission cost (€)	Battery throughput (kWh)	Daily energy cost (€)
40	Proposed	2102.0332	118.4323	189.0191	555.8055
	QL	2108.0267	118.7715	187.2234	557.9767
45	Proposed	2120.3695	119.4620	181.2250	559.4833
	QL	2124.4356	119.6959	177.4557	560.4852
50	Proposed	2129.7326	119.9896	184.4583	561.9195
	QL	2137.0311	120.4056	172.8813	562.8523
55	Proposed	2147.7348	121.0038	169.3868	563.6109
	QL	2152.8422	121.2964	162.4661	564.7938
60	Proposed	2162.6823	121.8460	158.3448	565.7577
	QL	2165.4063	122.0151	148.6896	568.3321

Table VIII shows the emission cost and daily energy cost comparison of the three cases for with SOC of 50% the proposed algorithm.

TABLE VIII
EMISSION COST AND DAILY ENERGY COST COMPARISON FOR THREE CASES WITH SOC OF 50%

Case	Emission cost (€)	Daily energy cost (€)	Total cost (€)
Case 1	133.8633	555.4273	689.2906
Case 2	123.4897	556.5094	679.9991
Case 3	119.9896	561.9195	681.9090

It can be seen that the emission cost in Case 3 decreases by 10.364% compared with that of Case 1. However, in Case 3, the daily energy cost is higher than those of other cases since only the emission cost is taken into consideration. Case 2 gives a relatively balanced result compared with other two cases in terms of daily energy cost and emission cost. As both emission cost and energy cost are tried to be minimized, the total operation cost is the lowest in Case 2.

VI. Conclusion

This paper proposes an MINLP guided QL algorithm for the real-time energy management of the stochastic and dynamic microgrid in Malta. The AC power flow equations and constraints, the battery wear cost and constraints, the fuel cost and the emission cost are considered for the economic and environment-friendly operation of the microgrid system. Three different cases are considered with three different objective functions: ① minimization of daily operation cost regardless of emission cost; ② minimization of both daily energy cost and emission cost; ③ minimization of emission cost without considering daily energy cost. The simulation results using real pilot data of MCAST prove the cost effectiveness of the proposed algorithm compared with the traditional QL algorithm. In case studies using the proposed algorithm, there is a 1.348% reduction in the daily total operation cost in Case 2 compared with Case 1. The daily total operation cost of the proposed algorithm is up to 1.25% lower than that of the QL algorithm. From the simulation results, we can also find that the battery lifetime is affected by the adjustment of SOC value of the battery.

References

[1] A. Kaur, J. Kaushal, and P. Basak, "A review on microgrid central controller," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 338-345, Mar. 2016.

[2] Y. Yoldaş, A. Önen, S. M. Muyeen *et al*., "Enhancing smart grid with microgrids: challenges and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 205-214, May 2017.

[3] Y. Li and F. Nejabatkhah, "Overview of control, integration and energy management of microgrids," *Journal of Modern Power Systems and Clean Energy*, vol. 2, no. 3, pp. 212-222, Sept. 2014.

[4] A. Das and Z. Ni, "A computationally efficient optimization approach for battery systems in islanded microgrid," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6489-6499, Nov. 2018.

[5] H. Shuai, J. Fang, X. Ai *et al*., "Stochastic optimization of economic dispatch for microgrid based on approximate dynamic programming," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2440-2452, May 2019.

[6] H. Shuai, J. Fang, X. Ai *et al*., "Optimal real-time operation strategy for microgrid: an ADP-based stochastic nonlinear optimization approach," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 931-942, Apr. 2019.

[7] P. Zeng, H. Li, H. He *et al*., "Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 4435-4445, Jul. 2019.

[8] G. K. Venayagamoorthy, R. K. Sharma, P. K. Gautam *et al*., "Dynamic energy management system for a smart microgrid," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 8, pp. 1643-1656, Aug. 2016.

[9] H. Kanchev, F. Colas, V. Lazarov *et al*., "Emission reduction and economical optimization of an urban microgrid operation including dispatched PV-based active generators," *IEEE Transactions on Sustainable Energy*, vol. 5, no. 4, pp. 1397-1405, Oct. 2014.

[10] M. Sedighizadeh, M. Esmaili, A. Jamshidi *et al*., "Stochastic multi-objective economic-environmental energy and reserve scheduling of microgrids considering battery energy storage system," *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 1-16, Mar. 2019.

[11] V. S. Tabar, M. A. Jirdehi, and R. Hemmati, "Energy management in microgrid based on the multi objective stochastic programming incorporating portable renewable energy resource as demand response option," *Energy*, vol. 118, pp. 827-839, Jan. 2017.

[12] J. Han, S. Khushalani-Solanki, J. Solanki *et al*., "Adaptive critic design-based dynamic stochastic optimal control design for a microgrid with multiple renewable resources," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2694-2703, Nov. 2015.

[13] S. F. Contreras, C. A. Cortes, and J. M. A. Myrzik, "Optimal microgrid planning for enhancing ancillary service provision," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 4, pp. 862-875, Jul. 2019.

[14] T. Shekari, A. Gholami, and F. Aminifar, "Optimal energy management in multi-carrier microgrids: an MILP approach," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 4, pp. 876-886, Jul. 2019.

[15] D. Prudhviraj, P. B. S. Kiran, and N. M. Pindoriya, "Stochastic energy management of microgrid with nodal pricing," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 1, pp. 102-110, Mar. 2020.

[16] B. Khan and P. Singh, "Selecting a meta-heuristic technique for smart micro-grid optimization problem: a comprehensive analysis," *IEEE Access*, vol. 5, pp. 13951-13977, Jul. 2017.

[17] S. Leonori, M. Paschero, F. M. F. Mascioli *et al*., "Optimization strategies for microgrid energy management systems by genetic algorithms," *Applied Soft Computing*, vol. 86, pp. 1-8, Jan. 2020.

[18] H. Farzin, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie, "Enhancing power system resilience through hierarchical outage management in multi-microgrids," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2869-2879, Nov. 2016.

[19] X. Yang, Y. Zhang, H. He *et al*., "Real-time demand side management for a microgrid considering uncertainties," *IEEE Transactions on Smart Grid*, vol. 10, no.3, pp. 3401-3414, May 2019.

[20] U. C. Yilmaz, M. E. Sezgin, and M. Go, "A model predictive control for microgrids considering battery aging," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 2, pp. 296-304, Mar. 2020.

[21] L. Meng, E. Sanseverino, A. Luna *et al*., "Microgrid supervisory controllers and energy management systems: a literature review," *Renewable and Sustainable Energy Reviews*, vol. 60, pp. 1263-1273, Jul. 2016.

[22] M. Zia, E. Elbouchikhi, and M. Benbouzid, "Microgrids energy management systems: a critical review on methods, solutions, and prospects," *Applied Energy*, vol. 222, pp. 1033-1055, Jul. 2018.

[23] P. Kou, D. Liang, and L. Gao, "Stochastic energy scheduling in microgrids considering the uncertainties in both supply and demand," *IEEE Systems Journal*, vol. 12, no. 3, pp. 2589-2600, Sept. 2018.

[24] H. Farzin, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie, "A stochastic multi-objective framework for optimal scheduling of energy storage systems in microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 117-127, Jan. 2017.

[25] A. Askarzadeh, "A memory-based genetic algorithm for optimization of power generation in a microgrid," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1081-1089, Jul. 2018.

[26] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge: MIT Press, 2018.

[27] J. Duan, Z. Yi, D. Shi *et al*., "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, Sept. 2019.

[28] F. Ruelens, B. J. Claessens, S. Vandael *et al*., "Residential demand response of thermostatically controlled loads using batch reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2149-2159, Sept. 2017.

[29] B. Mbuwir, F. Ruelens, F. Spiessens *et al*., "Battery energy management in a microgrid using batch reinforcement learning," *Energies*, vol. 10, no. 11, p. 1846, Nov. 2017.

[30] Z. Ding, L. Xie, Y. Lu *et al*., "Emission-aware stochastic resource planning scheme for data center microgrid considering batch workload scheduling and risk management," *IEEE Transactions on Industry Applications*, vol. 54, no. 6, pp. 5599-5608, Nov. 2018.

[31] Y. Lim and H. M. Kim, "Strategic bidding using reinforcement learning for load shedding in microgrids," *Computers & Electrical Engineering*, vol. 40, no. 5, pp. 1439-1446, Jul. 2014.

[32] E. Kuznetsova, Y. Li, C. Ruiz *et al*., "Reinforcement learning for microgrid energy management," *Energy*, vol. 59, pp. 133-146, Sept. 2013.

[33] E. Foruzan, L. K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5749-5758, Sept. 2018.

[34] P. Kofinas, A. I. Dounis, and G. A. Vouros, "Fuzzy *Q*-learning for multi-agent decentralized energy management in microgrids," *Applied Energy*, vol. 219, pp. 53-67, Jun. 2018.

[35] X. Xu, Y. Jia, Y. Xu *et al*., "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3201-3211, Jul. 2020.

[36] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, Aug. 2018.
[37] Y. Liu, C. Yuen, N. U. Hassan *et al.*, "Electricity cost minimization for a microgrid with distributed energy resource under different information availability," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 4, pp. 2571-2583, Apr. 2015.
[38] L. An and T. Tuan, "Dynamic programming for optimal energy management of hybrid wind-PV-diesel-battery," *Energies*, vol. 11, no. 11, p. 3039, Nov. 2018.

Yeliz Yoldas received the B.S. degree in electrical and electronics engineering from Cumhuriyet University, Sivas, Turkey, in 2011, and the M.S. degree in electrical and electronics engineering from Cukurova University, Adana, Turkey, in 2015. She is currently pursuing the Ph.D. degree in electrical and computer engineering at Abdullah Gul University, Kayseri, Turkey. Since 2013, she is working at Electrical and Electronics Engineering Department, Abdullah Gul University as a Research Assistant. Her research interests include power system operation and optimization, smart grid and microgrid.

Selcuk Goren received his B.S., M.S., and Ph.D. degrees from the Department of Industrial Engineering, Bilkent University, Ankara, Turkey. He is currently working as an Assistant Professor in Abdullah Gul University, Kayseri, Turkey. His research interests include robust optimization, renewable energy systems, smart grids, production scheduling under uncertainty, stochastic modeling, simulation, applied probability, discrete optimization, and metaheuristics.

Ahmet Onen received the B.Sc. degree in electrical and electronics engineering from Gaziantep University, Gaziantep, Turkey, in 2005. He received the M.S. degree in electrical and computer engineering from Clemson University, Clemson, USA, in 2010, and the Ph.D. from Electrical and Computer Engineering Department, Virginia Tech, Blacksburg, USA, in 2014. He is currently working as an Associate Professor in Abdullah Gul University, Kayseri, Turkey. His research interests include transmission networks and smart grids, big data in power systems, renewables and integrations, power system optimization, cost-benefit analysis, energy storage, demand response, FACTS devices and their application, block chain application, microgrid and virtual power plants.