Electrical Power and Energy Systems 174 (2026) 111556

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

RL-driven problem decomposition for computationally efficient AC optimal power flow

Ye-Eun Jang [a], Jaepil Ban [b,*], Young-Jin Kim [c,*], Chen Chen [d]

[a] Energy ICT Research Section, Electronics and Telecommunications Research Institute (ETRI), Daejeon 34129, South Korea
[b] Department of Electrical Engineering, Kumoh National Institute of Technology, Gumi, Gyeongbuk 39177, South Korea
[c] Department of Electrical Engineering, Pohang University of Science and Technology (POSTECH), Pohang, Gyeongbuk 37673, South Korea
[d] The School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

ARTICLE INFO

Keywords:
AC optimal power flow problem
Computational efficiency
Problem decomposition
Feasible area restriction
Reinforcement learning (RL)
RL-based cutting planes

ABSTRACT

Solving the AC optimal power flow (AC OPF) problem poses significant challenges in power system operations because of its inherent nonlinearity and complexity. This paper introduces a novel strategy to solve the AC OPF problem by utilizing a new problem decomposition framework combined with reinforcement learning (RL)-based cutting planes. The problem is decomposed into two sub-problems, DC OPF and AC power flow (AC PF) calculation sub-problems. To yield the AC-feasible solution, linear inequality constraints (i.e., cuts) are obtained by an RL agent and added into the DC OPF sub-problem. Then, the AC PF calculation is performed using the solution to the DC OPF sub-problem (i.e., power generation profiles) and voltage magnitude reference values, which is the output of the RL agent. Additionally, the action selection method is employed for the RL agent's training efficiency. Case studies under various simulation scenarios are conducted to show the effectiveness of the proposed strategy compared to the conventional strategies. The simulation results indicate that the proposed strategy significantly enhances computational efficiency and solution feasibility compared to the conventional methods.

1. Introduction

Optimal power flow (OPF) calculations are fundamental for maintaining secure and cost-efficient power system operations and are regularly performed by system operators across varying time scales from minutes to months. Since its introduction in the 1960 s [1], the studies on OPF have primarily addressed the single-period problems, neglecting the intertemporal coupling effects. However, as power systems evolve towards greater sustainability, integrating various energy sources and adhering to policies against climate crisis, the need to solve the multi-period OPF problem increases. The multi-period OPF reflects the intertemporal dependencies, thereby enhancing the cost-efficiency and security of power systems while supporting environmental sustainability goals.

Although solving the AC OPF problem is essential for ensuring system security, its nonlinearity and non-convexity make it NP-hard [2]. Thus, most system operators reformulate the AC OPF problem. To manage its complexity, the system operators typically solve the DC OPF problem [2], which simplifies the nonlinear constraints in the AC OPF problem to be solved by linear programming (LP) solvers. However, it relies on several simplifying assumptions, including constant voltage magnitudes at buses and reactive power ignorance, which may lead to the system insecurities, such as voltage limit violations and line overloads. Various linearization methods, such as piecewise linear approximation and Taylor series expansion, have been utilized to address the nonlinearity of power flow constraints [1–3]. While these methods simplify the original nonlinear problem, they can also compromise the intrinsic nonlinear properties of power systems and introduce additional decision variables, including integer variables. Consequently, the computational burden escalates with system sizes.

Recent research has sought to overcome the linearization methods' limitations by leveraging heuristic methods [4–6] and machine learning (ML) algorithms [7–12]. In [4–6], various evolutionary algorithms have been utilized, but the obtained solutions are often sensitive to initial values and prone to converging to local optima. To address these challenges, deep neural networks (DNNs) have been utilized in [7–12] to approximate the OPF solution and trained by both supervised learning and reinforcement learning (RL). In the supervised learning-based approaches, fully connected neural networks (FCNs) were used in [7], and

* Corresponding authors.
E-mail addresses: jpban@kumoh.ac.kr (J. Ban), powersys@postech.ac.kr (Y.-J. Kim).

https://doi.org/10.1016/j.ijepes.2025.111556
Received 25 September 2025; Received in revised form 1 December 2025; Accepted 29 December 2025
Available online 11 January 2026

Nomenclature

Sets and Indices

b, g, l, t	Indices for bus, generator, power line, and time step
Case•	Subscripts for Case •
max, min	Superscripts for maximum and minimum limits
$S_{B,l}$, $S_{G,b}$	Sets of buses linked to line l and generators at bus b
$\bar{\cdot}$, $\hat{\cdot}$	Solution and cut of the DC OPF sub-problem

Variables

α	Weighting factor for the expected entropy
$\mathbf{a}$, a_n^t	Vector expression for an action and its n-th element at time t
C_G	Cost for power generation
$\lvert F_l^t \rvert$	Magnitude of line flow on power line l
r_e^t	Power system states on line (or bus) e
P_g^t, Q_g^t, $\lvert V_g^t \rvert$	Active/reactive generated power and voltage magnitude reference of generator g
$\hat{P}_g^{t,\max}$	Added cut (power output limit) for generator g
$\lvert V_b^t \rvert$, θ_b^t	Voltage magnitude and phase angle at bus b

Parameters

$B_{bb'}$, $G_{bb'}$	Susceptance and conductance of the power line between buses b to b'
$\mathbf{E}$	Environmental data for the system operation
$F_l^{\max}$	Maximum value of line flow on power line l
N_{ASP}, N_{ASV}	Time-step intervals for which each cut and voltage magnitude reference is held constant
N_B, N_G, N_L	Numbers of buses, generators, and loads
N_{HL1}, N_{HL2}	Numbers of hidden layers in actor and critic networks
N_T	Number of time steps of the optimization horizon
$N_{R,2\text{-}4}$, $N_{R,2\text{-}6}$	Repetition numbers of Steps 2–4 and 2–6
$P_{L,b}^t$, $Q_{L,b}^t$	Active and reactive power loads
$P_g^{\max}$, $P_g^{\min}$, $Q_g^{\max}$, $Q_g^{\min}$, $\Delta\theta_l^{\max}$, $\Delta\theta_l^{\min}$, $\lvert V_b \rvert^{\max}$, $\lvert V_b \rvert^{\min}$	Max-/minimum values of active/reactive generated power, phase angle difference, and voltage magnitude
$\mathbf{P_{NL}}$	Vector expression for net loads profile
$R_{H,g}$, $R_{L,g}$	Up-/downward ramp rates of generated power for generator g
$\mathbf{S}$, $\mathbf{S}'$	Present and subsequent states for the RL agent
$\mathbf{Z}$	System information (i.e., component and topology)
γ	Discount factor for reward calculation
$\boldsymbol{\lambda}$	Vector expression for weighting factors in objective and reward function
Δt	Unit time step

graph neural networks (GNNs) were adopted in [8,9] as surrogate models for AC OPF. In particular, GNNs were introduced to explicitly incorporate grid topology, and these studies demonstrate improved generalization to varying topologies compared to FCNs. Furthermore, in [10], multiple FCNs are utilized for the physics-informed surrogate model, decomposing the overall model into three sub-models to mitigate the learning complexity and bias. However, the supervised learning requires large datasets generated by nonlinear programming (NLP) solvers under various system conditions, which is challenging to obtain due to the inherent complexity of the AC OPF problem, especially when the solver's initial conditions are not appropriately chosen [1]. In the reinforcement learning-based approaches, a convex-constrained DRL algorithm was applied in [11] to obtain the AC OPF solution directly, aiming to ensure safe control actions, sample efficiency, and training stability. In [12], a physics-informed RL approach was introduced, where the AC power flow (AC PF) module was embedded within the actor network to satisfy the constraints of the OPF. However, the conventional RL-based methods that directly search for the solutions within the nonconvex solution space of the AC OPF problem often require extensive exploration and longer training to achieve high-quality solutions. Moreover, physics-informed RL approaches, which integrate physical models directly, can affect the efficiency of the algorithm modification and the agent training.

To reduce the original problem's complexity, decomposition approaches have been applied [13–16]. In [13], the optimization problem incorporating unit commitments was divided into smaller sub-problems over a shorter optimization horizon (i.e., temporal decomposition). In [14], the original problem was partitioned into sub-problems for partitioned sub-systems from the entire power system (i.e., spatial decomposition). Similarly, in [15], the hierarchical decomposition was proposed, dividing the AC OPF problem into a master problem and sub-problems for the transmission system and distribution systems, respectively. However, these decomposition approaches did not fully resolve the nonconvex nature of the AC OPF problem, and the decomposed problems retained their nonconvex natures. In [16], the original problem was decomposed into sub-problems by integer variable introduction and bilinear programming-based reformulation. The reformulation was based on the assumptions that voltage magnitudes were approximately one and that angle differences across power lines were close to zero. Nevertheless, these assumptions may not be practical in real power systems.

Cutting plane methods have also been utilized as an effective approach for addressing the AC OPF's complexity. [17] introduced semidefinite programming (SDP) cuts to enhance the solution accuracy of second-order cone programming relaxations. [18] relaxed the OPF problem for high voltage DC and AC grids based on SDP and incorporated additional SDP cuts. However, these approaches can lead to infeasible solutions unless specific conditions in the power system hold [2], which help in minimizing the optimality gap. [19] iteratively incorporated the supporting hyperplanes and half-spaces into the LP-based optimization problem to obtain a solution that satisfies the AC OPF problem's constraints. However, these approaches are sensitive to initial conditions since the added constraints are derived from prior iterations, potentially affecting solution quality. Recently, cutting plane methods in various domains beyond power systems have incorporated ML algorithms [20,21]. [20] obtained the minimal number of cuts for the enhanced performance. [21] employed imitation learning to train the agent that mimics expert decision-making in cut selection. Despite their effectiveness, they concentrated on cut selection procedures rather than cut generation, and the computational complexity of the original problem were not considered.

Beyond classical cutting plane methods, recent studies [22,23] have explored feasibility restoration techniques that share conceptual similarities with cut incorporation, as they refine relaxed solutions to recover AC feasibility. For example, [22] applied AC feasibility restoration to DC OPF–based market clearing by using a linearized OPF algorithm with PTDFs for generator ranking and selection. However, this method primarily addresses active power violations and lacks reactive power and voltage coordination, while requiring proper initial points. [23] trained a DNN to output AC OPF solutions and introduced a feasibility restoration module that projects the DNN output onto the AC-feasible set via an optimization problem based on quadratically constrained quadratic programs (QCQP). However, the projection step is computationally expensive, and solving a relaxed QCQP can still lead to infeasible solutions, similar to SDP-based approaches.

This paper presents problem decomposition method and an RL-based cutting plane method to address the computational complexity associated with solving the AC OPF problem. The problem decomposition

method divides the AC OPF problem into two sub-problems: DC OPF and AC PF sub-problems. To obtain an AC-feasible solution, linear inequality constraints (i.e., cuts) are added to the DC OPF sub-problem and iteratively adjusted for the sub-problem's feasible region restriction. The corresponding solution for the restricted region is converted into the final solution for the AC-feasible region after its AC PF calculation. This procedure is referred to as *problem decomposition* in this paper. The added cuts are adjusted by the RL algorithm, and it can be seen that the cut generation and selection are performed simultaneously. This is referred to as *RL-based cutting plane method*. Once the agent is adequately trained, the proposed strategy ensures convergence by solving the DC OPF sub-problem with optimal cuts, which is convex. The proposed strategy successfully mitigates the challenges incurred by the AC OPF's non-convexity without relying on specific assumptions in conventional SDP-based approaches. In addition, the efficiency of the agent training is improved through the proposed action selection method that reduces fluctuations in generator operations over time. The primary contributions of this study are as follows:

- A novel decomposition strategy for solving AC OPF problem is introduced, where the problem is divided into a DC OPF and AC PF calculation sub-problems to enhance the computational efficiency.
- This study pioneers the RL-based cutting plane method for power system operation, which concurrently generates and selects cuts for computation efficiency against conventional cutting plane procedures. In addition, the proposed method is shown to be more efficient in solving the optimization problem than conventional RL-based methods that find the solution directly.
- The DC OPF sub-problem is refined through strategically adjusted cutting planes, ensuring that the resulting AC PF solutions exhibit fewer constraint violations and are more likely to be AC-feasible.
- Comprehensive case studies validate the proposed strategy's effectiveness, highlighting its improvements in computational efficiency and convergence.

The structure of this paper is as follows: Section 2 proposes and explains a new RL-based strategy for the OPF calculation. Section 3 analyses the proposed strategy, focusing on feasible regions and computational complexities. Section 4 discusses the results of case studies under various conditions. Finally, Section 5 summarizes and concludes this study.

2. Problem decomposition and RL-Based cutting plane methods for AC OPF

2.1. Overview

To efficiently achieve an AC-feasible solution, the proposed strategy utilizes a decomposition framework that separates the AC OPF problem into the DC OPF and AC PF sub-problems. Although both sub-problems can be solved using conventional solvers within reasonable time, the DC OPF sub-problem, whose decision variables consist of active power and phase angles only, often results in AC-infeasible solutions. To address this issue, the proposed strategy introduces additional linear inequality constraints (i.e., cuts) into the DC OPF sub-problem with an RL agent optimizing these cuts, and the active generated power is constrained by the additional cuts. The cuts serve as the time-varying generation limits in the sub-problem. Then, the AC PF sub-problem reconstructs system states, including voltage magnitudes and reactive power, which are used to compute the reward function for the RL agent training.

The proposed strategy follows the seven-step procedures:

Step 1) Apply the problem decomposition method (i.e., the reformulation of the AC OPF problem), dividing the AC OPF problem into the DC OPF and AC PF sub-problems. The RL agent is then pretrained with an initial dataset for initial action.

Step 2) Obtain the action (i.e., cuts for power generations, and voltage magnitude references at PV and slack buses) from the RL agent, add the cuts into the DC OPF sub-problem, and solve the extended sub-problem.

Step 3) Compute the AC PF solution based on the solution of the DC OPF sub-problem and the voltage references (both are obtained in Step 2). This step calculates the system variables that are not determined in Step 2 (e.g., voltage magnitudes at PQ buses and reactive power generation).

Step 4) Evaluate the reward function based on generation costs and constraint violations. The corresponding action, state, and reward are stored for the next agent training.

Step 5) Train the RL agent with the data, which is stored in Step 4, for the maximized reward.

Step 6) Adjust the cuts and voltage magnitude references (i.e., action) based on the updated agent.

Step 7) Iterate Steps 2 to 6 until the iteration number exceeds the predetermined value. Once the agent training is completed, Steps 2 and 3 are executed to obtain the final OPF solution.

Fig. 1 compares the feasible regions of three OPF problems: the conventional AC OPF and DC OPF problems and the DC OPF sub-problem including the cuts. The grey and red regions represent the feasible regions of the conventional DC OPF and AC OPF problems, respectively. The yellow and pink regions indicate the converted feasible regions of the conventional DC OPF (i.e., the grey one) and the DC OPF sub-problem with the cuts, respectively. The conversion is performed through AC PF calculation. The yellow and red markers represent the solutions resulting from the conventional DC OPF problem and the DC OPF sub-problem formulated under the proposed strategy, respectively. Due to the simplifications inherent in the DC OPF formulation –particularly the omission of power losses and reactive power– the feasible region of the DC OPF problem does not intersect with that of the AC OPF one, as depicted in Fig. 1(a) [24]. The impact of the cuts on final solutions with respect to the AC-feasibility is assessed by performing the AC PF calculation. Calculating the AC PF at each point in the feasible region of the DC OPF problem makes the feasible region converted to be aligned with the AC OPF problem's feasible region. For simplicity, the converted feasible region, along with its two-dimensional projection of the DC OPF sub-problem, is used throughout the remainder of this paper as Fig. 1(b) without the additional explanations.

2.2. Original AC OPF problem

The following formulation describes the original AC OPF problem:

$\mathcal{P}_1$ (Original AC OPF problem):

$$\min_{\boldsymbol{P}_G,|\boldsymbol{V}_G|} C_G(\boldsymbol{C},\boldsymbol{P}_G), \tag{1}$$

$$\text{s.t.} \sum_i |V_b^t||V_i^t|\left(G_{ib}\cos\left(\theta_i^t-\theta_b^t\right)+B_{ib}\sin\left(\theta_i^t-\theta_b^t\right)\right)=\sum_{g\in S_{G,b}} P_g^t - P_{L,b}^t,\ \forall t,b \tag{2}$$

$$\sum_i |V_b^t||V_i^t|\left(G_{ib}\sin\left(\theta_i^t-\theta_b^t\right)-B_{ib}\cos\left(\theta_i^t-\theta_b^t\right)\right)=\sum_{g\in S_{G,b}} Q_g^t - Q_{L,b}^t,\ \forall t,b \tag{3}$$

$$F_l^t = V_i^t\left(Y_{ii}V_i^t + Y_{ij}V_j^t\right)^*,\ \forall t,\ l,\ (i,j)\in S_{B,l} \tag{4}$$

$$\left|F_l^t\right| \le F_l^{\max},\ \forall t,\ l \tag{5}$$

$$\Delta\theta_l^{\min} \le \theta_i^t - \theta_j^t \le \Delta\theta_l^{\min},\ \forall t,\ l,\ \{i,j\}\in S_{B,l} \tag{6}$$

$$P_g^{\min} \le P_g^t \le P_g^{\max},\ \forall t,\ g \tag{7}$$

$$Q_g^{\min} \le Q_g^t \le Q_g^{\max},\ \forall t,\ g \tag{8}$$

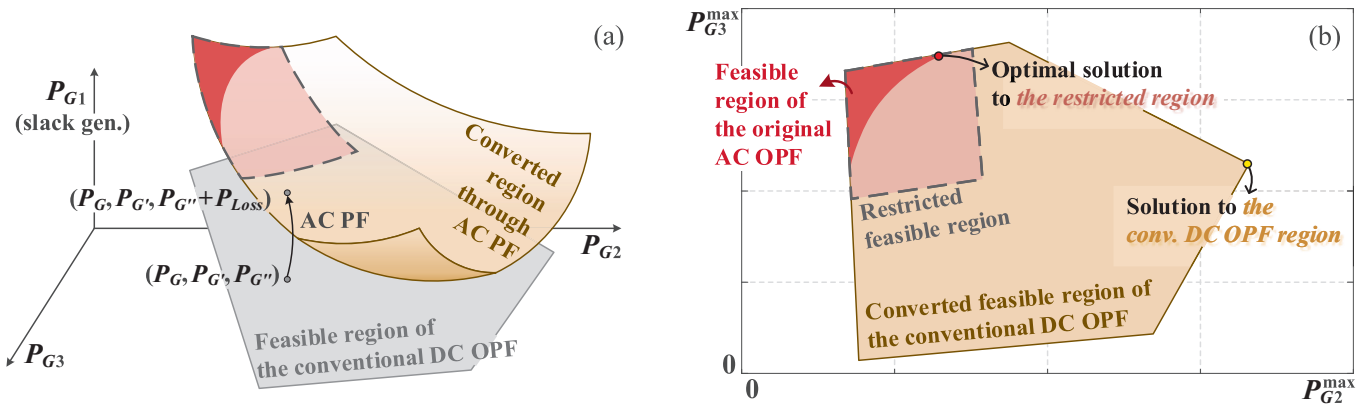

Fig. 1. A depiction of the feasible region comparison among the conventional AC and DC OPF problems, and the DC OPF sub-problem of the proposed strategy.

$$|V_b|^{\min} \le |V_b^t| \le |V_b|^{\max}, \ \forall t, b \tag{9}$$

$$R_{L,g} \le \left(P_g^t - P_g^{t-\Delta t}\right) \Big/ \Delta t \le R_{H,g}, \ \forall t, g. \tag{10}$$

The objective function (1) consists of the generation cost C_G, and it aims to be minimized. C_G is determined by the unit generation price $\mathbf{C}$ and active power generation P_G. (2) and (3) indicate active/reactive power balance constraints, respectively. P_g^t and Q_g^t are active/reactive power outputs of generator g, and $P_{L,b}^t$ and $Q_{L,b}^t$ are active/reactive load at bus b. (4) and (5) specify power flow and its limits $F_l^{\max}$ on line l that connects buses i and j, where V_i^t and Y_{ij} are complex voltage at bus i and (i, j)-th element of the admittance matrix of the power system. (6) to (10) represent the system operating constraints. They include constraints on phase angle differences across power lines (i.e., θ_i^t–θ_j^t), active/reactive power generations (i.e., P_g^t and Q_g^t), bus voltage magnitude (i.e., $|V_b^t|$), and ramp rates of power generations (i.e., $(P_g^t$–$P_g^{t-\Delta t})$ /Δt), respectively. The ramp rate constraint (10) incurs intertemporal coupling effects in power generation scheduling, and the power balance and flow constraints (2) to (4) incurs the non-convexity in $\boldsymbol{\mathcal{P}_1}$. Both characteristics make solving $\boldsymbol{\mathcal{P}_1}$ challenging.

2.3. Problem reformulation using problem decomposition method (Steps 1 – 3)

To alleviate the computational challenge, the proposed strategy reformulates the original problem $\boldsymbol{\mathcal{P}_1}$ through the problem decomposition as follows:

$\boldsymbol{\mathcal{P}_2}$ (Reformulated AC OPF problem in compact form):

$$\min_{\mathbf{P_G}, |\mathbf{V_G}|} C_G(\mathbf{C}, \mathbf{P_G}), \tag{11}$$

$$\text{s.t.} \ \ P_g^{\min} \le P_g^t \le P_g^{\max}, \ \forall t, g \tag{12}$$

$$Q_g^{\min} \le Q_g^t \le Q_g^{\max}, \ \forall t, g \tag{13}$$

$$\mathbf{A}\cdot\mathbf{r_E} \le \mathbf{r_E^M}, \ \forall \mathbf{E} \tag{14}$$

$$\text{where} \ \ [\mathbf{P_G}, \mathbf{Q_G}, \mathbf{r_E}] = f_{\text{ACPF}}(\overline{\mathbf{P}}_\mathbf{G}, |\mathbf{V_G}|; \mathbf{E}, \mathbf{Z}), \tag{15}$$

$$\overline{\mathbf{P}}_\mathbf{G} = f_{\text{DCOPF+}}\left(\widehat{\mathbf{P}}_\mathbf{G}^{\max}; \mathbf{C}, \mathbf{E}, \mathbf{Z}, \varepsilon_{DC}\right), \tag{16}$$

$$\left[\widehat{\mathbf{P}}_\mathbf{G}^{\max}, |\mathbf{V_G}|\right] = f_{\text{Agent}}(\mathbf{E}; \mathbf{C}, \mathbf{Z}, \varepsilon_{RL}), \tag{17}$$

$$P_g^{\min} \le \widehat{P}_g^{t,\max} \le P_g^{\max}, \ \forall t, g. \tag{18}$$

The objective function (11) and constraints (12) and (13) are same as (1), (7), and (8), respectively, in the original problem $\boldsymbol{\mathcal{P}_1}$. For notational simplicity, $\mathbf{r_E}$ (i.e., r_e^t) and $\mathbf{r_E^M}$ in (14) and (15) are used to indicate the power system states $[\{|V_b^t|, F_l^t, \theta_b^t\}_{t,b,l}]^{\mathrm{T}}$ and its minimum and maximum values, respectively. $\mathbf{A}$ is an incidence matrix that reformulates individual variable bounds and linear difference constraints into a unified compact inequality form. Thus, (14) corresponds to (5), (6), and (9).

(15) represents AC PF calculation function $f_{\text{ACPF}}(\cdot)$ given the active power generation schedules $\overline{P}_g^t$ obtained by DC OPF sub-problem, generators' voltage magnitudes $|V_g^t|$ obtained by the RL agent, environmental data $\mathbf{E}$ (e.g., load profiles), and system information $\mathbf{Z}$ (e.g., system components and topology). Because of the formulation for AC PF calculation, (15) corresponds to (2)–(4) in $\boldsymbol{\mathcal{P}_1}$. (16) indicates the DC OPF sub-problem solving function $f_{\text{DCOPF+}}(\cdot)$ that provides the active power generation schedules $\overline{P}_g^t$ based on additional cuts $\widehat{P}_g^{t,\max}$, the generation price $\mathbf{C}$, $\mathbf{E}$, $\mathbf{Z}$, and an initial value ε_{DC}. (16) includes additional constraints on cuts $\widehat{P}_g^{t,\max}$ to restrict its feasible area as follows:

$$P_g^t \le \widehat{P}_g^{t,\max}, \ \forall t, g. \tag{19}$$

$\widehat{P}_g^{t,\max}$ is distinct from $P_g^{\max}$. The former represents the additional generation limit introduced by the cuts, while the latter indicates the actual power generation limit. Their relationship including P_g^t is expressed as follows:

$$P_g^t \le \widehat{P}_g^{t,\max} \le P_g^{\max}, \ \forall t, g. \tag{20}$$

For clarity, the compact formulations presented in Eqs. (14)–(18) represent the functional expressions of the detailed equations introduced earlier. Specifically, each component (e.g., AC PF calculation, DC OPF sub-problem, and forward evaluation) is expressed as a single functional module to enhance readability and generalize the proposed strategy. This compact representation has been widely used in previous studies [8,12,25], as it maintains mathematical rigor while simplifying the overall optimization structure.

(17) represents forward estimation function f_{Agent} of the RL agent given $\mathbf{E}$, $\mathbf{C}$, $\mathbf{Z}$, and the probabilistic factor ε_{RL} for the RL algorithm. The outputs (i.e., action) of the agent are $|V_g^t|$ and $\widehat{P}_g^{t,\max}$, which are the input of (15) and (16), respectively. Because $f_{\text{DCOPF+}}$ in (16) does not include explicit constraints on voltage magnitudes and reactive power flow, $\boldsymbol{\mathcal{P}_2}'$s solution is induced towards the AC feasibility and C_G minimization through the RL agent's actions. (18) specifies the acceptable ranges of $\widehat{P}_g^{t,\max}$, which are same as the ranges of P_g^t to consider operation constraints of generators. To solve $\boldsymbol{\mathcal{P}_2}$, f_{ACPF}, $f_{\text{DCOPF+}}$, and f_{Agent} in (15)–(17) are evaluated sequentially instead of solving them simultaneously. Table 1 lists the comparisons among the original problem and both sub-problems using $\boldsymbol{\mathcal{P}_1}$ and $\boldsymbol{\mathcal{P}_2}$. Notably, $\boldsymbol{\mathcal{P}_1}$ and $\boldsymbol{\mathcal{P}_2}$ are closely resembled (but not equivalent). When the action of the RL agent is optimally determined, the solution of $\boldsymbol{\mathcal{P}_2}$ is determined satisfying AC feasibility. To

Table 1
Comparisons of the AC OPF problem and the decomposed sub-problems in the proposed strategy.

Elements	AC OPF	Decomposed problems of AC OPF		
		DC OPF with cuts ($f_{\text{DCOPF+}}$)	AC PF (f_{ACPF})	RL-based cutting plane (f_{Agent})
Objective	(1)	(1)	Min. violations of (2)–(4)	Min. (1) and violations of (5)–(9)
Decision var.	$\{P_g^t, Q_g^t, \lvert V_b^t\rvert, \theta_b^t\}_{b,g,t}$	$\{P_g^t, \theta_b^t\}_{b,g,t}$	$\{P_{loss}^t, Q_g^t, \lvert V_b^t\rvert, \theta_b^t\}_{b,g,t}$	$\{\widehat{P}_g^{t,\max}, \lvert V_g^t\rvert\}_{g,t}$
Const.	(2)–(10)	DC power balance, (5)–(7), (10), (19)	–	(9) for slack and PV buses, (18)
Solver	NLP solver	LP solver	Newton's method	RL

ensure the feasibility of solutions to both sub-problems, it is assumed that the system operates under normal conditions. Specifically, power loads are kept within the operating capacity of generators and power lines, and the system remains reasonably balanced [26]. In addition, the selection of the initial solution for AC PF calculation can affect its convergence. Therefore, a well-established approach (i.e., $|V| = 1$p.u. and $\theta = 0$ rad) proposed in [27] is employed in this study to maintain the feasibility and enhance the convergence of the Newton-based AC PF calculation.

2.4. Cutting plane generation and selection (Steps 4 – 6)

As an RL algorithm, soft actor-critic (SAC) algorithm [28] is employed in the RL-based cutting plane method of the proposed strategy. As illustrated in Fig. 2, the agent consists of an actor and critic networks. The actor network estimates the best action **a** corresponding to the given state **S** (i.e., the net load profile) as follows:

$$\mathbf{a} = \left[a_1^1, a_1^2, \cdots, a_{2N_G}^{N_T}\right]^{\mathrm{T}},\ a_n^t \sim N\left(\mu_n^{\pi,t}(\mathbf{S}), \sigma_n^{\pi,t}(\mathbf{S})\right),\ \forall t, n \tag{21}$$

where N_T and N_G represent the numbers of the unit time steps and the controllable generators, respectively. $\mu_n^{\pi,t}$ and $\sigma_n^{\pi,t}$ denote the mean and standard deviation of a_n^t, and these values are obtained by policy function π for **S**.

The action a is defined by $\widehat{P}_g^{t,\max}$ and $|V_g^t|$ for all generators g and $1 \le t \le N_T$, i.e., $\mathbf{a} = [\widehat{P}_1^{1,\max}, \widehat{P}_1^{2,\max}, \ldots, \widehat{P}_2^{1,\max}, \widehat{P}_2^{2,\max}, \ldots, |V_1^1|, |V_1^2|, \ldots, |V_2^1|, |V_2^2|, \ldots]^{\mathrm{T}}$. The size of the action space is $2N_TN_G$. Note that, as N_T or the system size increases, the action space expands significantly, making the agent training challenging, particularly in case that the state has high variance [29]. To enhance training efficiency and exploration performance, the action space reduction is advisable [30]. In the proposed strategy, each value of $\widehat{P}_g^{t,\max}$ and $|V_g^t|$ is held constant over multiple time steps, denoted as N_{ASP} and N_{ASV}, based on the system operators' preferences. This method, which is denoted as *action selection* throughout this study, reduces the action size to $N_G(\lceil N_T/N_{ASP}\rceil + \lceil N_T/N_{ASV}\rceil)$, improving training efficiency.

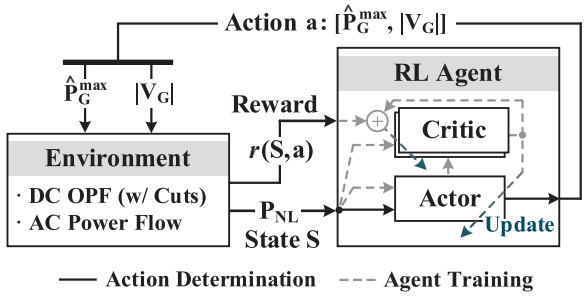

Fig. 2. A diagram for the input–output relationship between the environment and agent of the proposed RL-based cutting plane method.

The policy function π is trained by minimizing (22):

$$J_\pi = \mathbb{E}_{\mathbf{S}}[\mathbb{E}_{\mathbf{a}}[\alpha \log\pi(\mathbf{a}|\mathbf{S}) - \Omega(\mathbf{S},\mathbf{a})]]. \tag{22}$$

Similarly, the Q-function Ω is trained by minimizing (23):

$$J_\Omega = \mathbb{E}_{(\mathbf{S},\mathbf{a})}\left[\left(\Omega(\mathbf{S},\mathbf{a}) - r(\mathbf{S},\mathbf{a}) - \gamma\mathbb{E}_{\mathbf{S}'}[\Psi(\mathbf{S}')]\right)^2\right], \tag{23}$$

$$\Psi(\mathbf{S}) = \mathbb{E}_{\mathbf{a}}[\Omega(\mathbf{S},\mathbf{a}) - \alpha\log\pi(\mathbf{a}|\mathbf{S})]. \tag{24}$$

To fully capture the intertemporal coupling effect [31], the state **S** consists of multi-period data across the optimization horizon, and the next state **S′** is not explicitly considered, with a zero value for the discount factor γ, indicating that it is a single-step RL algorithm. This structure stems from the fact that the current action (i.e., $\widehat{P}_g^{t,\max}$ and $|V_g^t|$) has almost no effect on the subsequent state (i.e., net load). Thus, the multi-period AC OPF problem in this study does not need to be modeled as a Markov decision process (MDP). Instead, the RL agent is employed as a static optimizer without relying on the sequential decision-making mechanism of conventional multi-step RL algorithms.

The update of weighting factor α is conducted by minimizing (25):

$$J_\alpha = \mathbb{E}_{\mathbf{a}}[-\alpha\log\pi(\mathbf{a}|\mathbf{S}) - \alpha H], \tag{25}$$

where H represents a desired minimum expected entropy. The procedure of the RL-based cutting plane method including the agent training is outlined in **Algorithm 1**. In this study, explicit termination criteria of a training loop are not defined; instead, the numbers of iterations are set sufficiently large for the agent training loops (i.e., lines 3–13 in **Algorithm 1**) to follow the standard RL algorithms [28]. However, incorporating termination criteria could further reduce the computational cost by stopping the training process once the reward sufficiently saturates.

To formulate the reward function, $\mathcal{P}_2$ is reformulated into $\mathcal{P}_3$ by relaxing the constraints (12)–(14) as follows:

$\mathcal{P}_3$ (Reformulation of $\mathcal{P}_2$ for reward calculation):

$$\min_{\mathbf{P_G},\,|\mathbf{V_G}|} \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{y}, \tag{26}$$

$$\text{where } \mathbf{y} = \left[C_G(\mathbf{C},\mathbf{P_G}), \mathbf{D_X}^{\mathrm{T}}\right]^{\mathrm{T}}, \tag{27}$$

(15)–(18)

The objective function of $\mathcal{P}_3$ consists of the objective function, or the operating cost, (11) of $\mathcal{P}_2$ and the penalty $\mathbf{D_X}$ of (12)–(14) as expressed in (26) and (27). $\mathbf{D_X}$ is the constraint violations for P_g^t, Q_g^t and r_e^t from their acceptable ranges, and each value is computed as follows:

$$D_x := \sum_t \left[\sum_n \left\{\max\left(x_n^t - x_n^{\max}, 0\right) + \max\left(x_n^{\min} - x_n^t, 0\right)\right\}\right], \tag{28}$$

where x_n^t denotes a variable in (12)–(14) (e.g., P_g^t, Q_g^t, $|V_b^t|$), and $\mathbf{D_X}$ is a non-negative value, meaning that the constraint violation treats as a cost in the optimization. $\boldsymbol{\lambda}$ is a set of the weighting factors, which are selected empirically, considering the trade-offs among the various components of the objective function as well as the system operator's preferences. Based on $\mathcal{P}_3$, the reward function is formulated as follows:

$$r := -\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{y}. \tag{29}$$

To compute (29), the environment first provides P_g^t of the non-slack generators by solving the DC OPF sub-problem. These values, together with $|V_g^t|$ determined by the RL agent, are then used in the subsequent AC PF calculation. The AC PF results determine C_G and D_x for (29).

During the RL agent training process, the agent in (17) is updated continuously. After sufficient training iterations, the well-trained agent

can be used to yield the solution for $\mathcal{P}_2$, ensuring computational efficiency and the improved feasibility. Once the training process is conducted offline, only the *action determination* part in Fig. 2 is executed for the online application. Specifically, the RL agent receives the net load $\mathbf{P_{NL}}$ as the state $\mathbf{S}$ and produces the action $\mathbf{a}$ corresponding to (21), which is then used in the DC OPF sub-problem and AC PF calculation to ultimately derive the solution of the AC OPF.

In this study, the grid topology is assumed to remain fixed and the system is operated under normal conditions without considering contingencies or high RES variability, as the main objective is to reduce the computational complexity of the multi-period AC OPF problem. For future work, the proposed strategy can be extended to handle network reconfiguration and uncertain operating conditions by incorporating topological parameters and formally defined uncertainty sets into the RL-based framework and further integrating them into both sub-problems.

Algorithm 1: RL-based cutting plane method to solve AC OPF problem

1: Initialize policy network π, Q-networks Ω_1, Ω_2, entropy weighting factor α, and replay buffer
2: Reset environment and set state $\mathbf{S}$.
3: **For** $i_{R,2\text{-}6} = 1$ to $N_{R,2\text{-}6}$ **do**
4: **For** $i_{R,2\text{-}4} = 1$ to $N_{R,2\text{-}4}$ **do**
5: Sample action $\mathbf{a} = \pi(\mathbf{S})$ based on (21): $a_n^t \sim N(\mu_n^{\pi,t}(\mathbf{S}), \sigma_n^{\pi,t})$ for n, t
6: Observe environment feedback.
7: Compute reward $r(\mathbf{S}, \mathbf{a})$ using (29).
8: Store tuple data $\{\mathbf{S}, \mathbf{a}, r(\mathbf{S}, \mathbf{a})\}$ into the replay buffer.
9: **End**
10: Sample mini-batches from the stored data.
11: Update π, Ω_1, Ω_2, and α by minimizing (22), (23), and (25).
12: Minimize (22) with $\Omega(\mathbf{S}, \mathbf{a})$ replaced by $\min(\Omega_1(\mathbf{S}, \mathbf{a}), \Omega_2(\mathbf{S}, \mathbf{a}))$.
13: **End**
14: **Output**: optimal action $\mathbf{a}$ (i.e., $\{\hat{P}_g^{t,\max}, |V_g^t|\}_{g,t}$)

3. Analysis of The proposed OPF strategy

3.1. Feasible region comparison

To show the feasible region differences among the AC OPF problem, the conventional DC OPF problem, and the DC OPF sub-problem (16) incorporating the cuts, these problems are formulated for the IEEE 9-bus system [32] that has three generators and power loads at three PQ buses. Figs. 3 and 4 show the projected feasible regions onto P_g^t, Q_g^t, and $|F_l^t|$ of the selected generators and power lines. Specifically, Fig. 3 visualizes variations of the DC OPF sub-problem's feasible region and solution throughout the agent training procedure, while Fig. 4 shows only variations of the sub-problem's solution. In contrast to Fig. 3, Fig. 4 shows only the AC OPF problem's feasible region, as the DC OPF problem does not account for the reactive power generations. In Fig. 3, the portion of the DC OPF's feasible region that lies outside the AC OPF's region represents the solution set that are infeasible for the AC OPF problem. It highlights that the DC OPF's solution is not inherently AC-feasible, as the DC OPF formulation neglects certain state variables of the system, such as reactive power and voltage magnitudes.

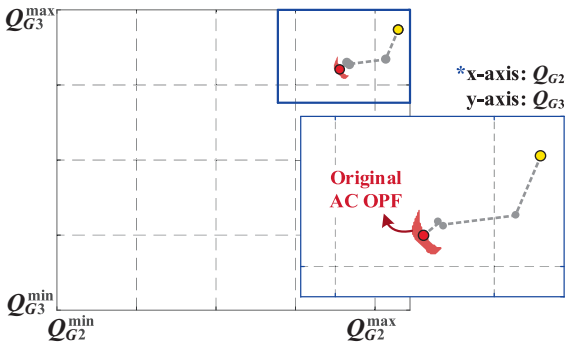

Fig. 4. Solution variations of the DC OPF sub-problem in the proposed strategy during the agent training compared to the original AC OPF problem's feasible region with the axis representing Q_g^t.

To address this concern, the RL-based cutting plane method progressively refines the feasible region of the DC OPF sub-problem (16), as shown in Fig. 3(a)–(d). Among the added cuts, the cuts for the selected two generators (i.e., $\hat{P}_{G2}^{\max}$ $\hat{P}_{G3}^{\max}$) are depicted as vertical and horizontal boundaries of the feasible region. Throughout the iterative procedures from (a) to (d), these cuts are continuously adjusted, guiding the solution toward the feasible region of the AC OPF problem. It is also observed in Fig. 3(e)–(h) and 4. The gap between the solution and the AC OPF's feasible region represents constraint violations of the decision variables, and its reduction increases the reward (29) for the RL agent. Consequently, by leveraging the proposed cutting plane method with a well-trained RL agent, an AC-feasible solution can be achieved effectively.

Because the cuts in the proposed strategy are the additional linear inequality constraints for power ratings, the resulting feasible region can be conservative, as illustrated in Fig. 3. However, it can be regarded as a trade-off for ensuring AC feasibility and achieving low computational complexity [33]. Addressing this conservativeness within the proposed strategy remains a future research area. Additionally, a key contribution of this study is obtaining high-quality solutions that satisfy AC feasibility for the OPF problem. The conventional DC OPF solutions often lead to increased operational costs due to the system constraints violations,

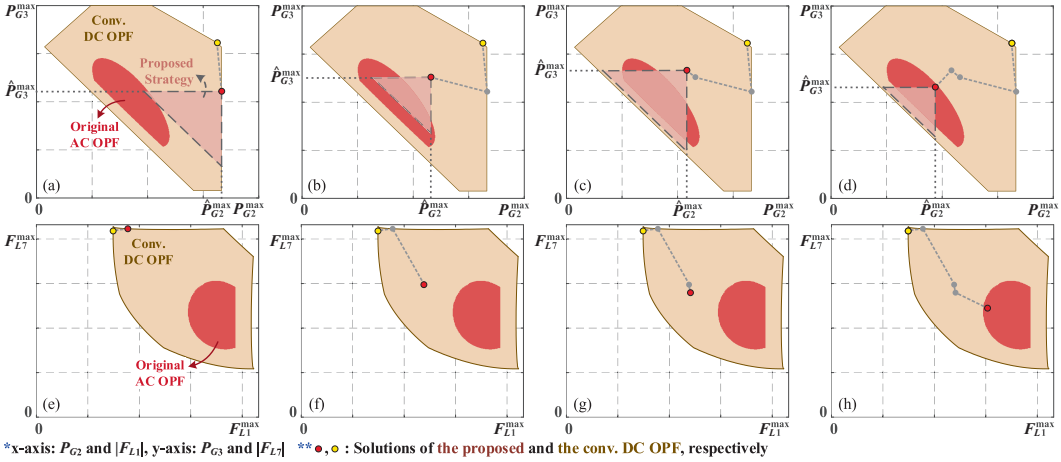

Fig. 3. Feasible region and solution variations of the DC OPF sub-problem in the proposed strategy during the agent training compared to feasible regions of the original AC OPF and conventional DC OPF problems with axes representing P_g^t and $|F_l^t|$.

requiring corrective actions [34]. In contrast, the proposed strategy focuses on minimizing the violations, thereby mitigating the need for costly adjustments in system operations.

3.2. Computational complexity analysis

The computational complexities of the three main procedures in the proposed strategy–solving the DC OPF sub-problem, estimating cuts and voltage references using the RL agent, and calculating AC PF–are determined as follows:

$$\mathrm{CC}_{\mathrm{DCOPF}} = O\left(\{N_T(N_G + N_B)\}^{3.5}\right), \tag{30}$$

$$\mathrm{CC}_{\mathrm{RL}} = O\left(\{N_T N_B\}^2\right), \tag{31}$$

$$\mathrm{CC}_{\mathrm{ACPF}} = O\left(N_T N_B^3\right). \tag{32}$$

Thus, the overall complexity is derived by summing (30)–(32), resulting in $O(\{N_T(N_G + N_B)\}^{3.5})$. This complexity is equivalent to that of solving the DC OPF problem $\mathrm{CC}_{\mathrm{DCOPF}}$ and remains notably lower than that of solving the AC OPF problem, which is known to be NP-hard. It confirms the proposed strategy achieves the same computation efficiency as the DC OPF formulation. Each complexity of (30)–(32) is derived as follows.

The computational complexity of the interior-point method for solving an LP-based problem is $O(N_{DV}{}^{3.5})$ [35] for N_{DV} decision variables. As listed in Table 1, for a multi-period DC OPF problem, N_{DV} is determined by $N_T(N_G + N_B)$. Consequently, the computational complexity of solving the DC OPF problem can be expressed as:

$$\mathrm{CC}_{\mathrm{DCOPF}} = O\left(\{N_T(N_G + N_B)\}^{3.5}\right). \tag{33}$$

The computational complexity of the cut and voltage reference estimation is same with that of the DNN's forward evaluation, and it is expressed as $O(N_{HL}M_{HN}{}^2)$ [36], where N_{HL} is the number of hidden layers in a DNN, and M_{HN} is the maximum number of hidden nodes of the layers. In this paper, the DNN is the actor network in SAC algorithm, and N_{HL} for the network is set as 3. M_{HN} is chosen to be of the same order as $N_T N_B$, because the actor network's input and output node sizes are $N_T N_L$ and $\lceil N_T N_G / N_{ASP} \rceil + \lceil N_T N_G / N_{ASV} \rceil$, respectively, and they increase in the same order with $N_T N_B$, when N_{ASP} and N_{ASV} are set as 1. Consequently, the complexity of the agent's action estimation is derived as:

$$\mathrm{CC}_{\mathrm{RL}} = O\left(\{N_T N_B\}^2\right). \tag{34}$$

Although the training is performed offline and at predetermined periods set by the system operators, this work also incorporates the computational complexity of the agent training. The agent training involves sequential operations of forward evaluation, backward propagation, and the subsequent update of model parameters. Through experimental analysis in previous studies [37,38], it has been confirmed that the computational complexity of the parameter update is negligible, whereas that of the backward propagation is approximately twice that of the forward evaluation. Nonetheless, as both forward and backward propagations have the same asymptotic complexity in big-O notation, $\mathrm{CC}_{\mathrm{RL}}$ can be represented as given in (31).

The AC PF calculation involves two operations: matrix multiplication and inversion. In case of an $n \times n$ matrix, both operations have computational complexities of $O(n^3)$ and $O(n^2)$, respectively [35]. In this study, n is set to $2N_B$ to calculation the AC PF. For multi-period AC PF calculation, $\mathrm{CC}_{\mathrm{ACPF}}$ is obtained as:

$$\mathrm{CC}_{\mathrm{ACPF}} = O\left(N_T N_B^3\right). \tag{35}$$

The original AC OPF problem is NP-hard due to its nonlinear and nonconvex power flow constraints, leading to non-polynomial computational complexity [39]. In contrast, the proposed method decomposes the original problem into tractable subproblems and employs the RL-based cutting plane method, allowing each subproblem to be solved in polynomial time. As a result, the proposed strategy achieves improved computational efficiency compared with the original AC OPF strategy.

4. Case studies and Numerical results

4.1. Test conditions

The IEEE 39-bus [40] and 9-bus [32] systems were used to evaluate the proposed strategy. In Section 4.1, both systems were tested to show the solution degradations as the system size increases. In the other sections, the 39-bus system was employed to assess the effectiveness of the proposed strategy. Both profiles of RESs' generation and power loads were reconstructed from the NYISO's measured data during June to August 2019 and June 2020 [41]. Among the reconstructed data, the RL agent was trained using the first 100 days of data, while the rest was utilized to evaluate the proposed strategy. The parameters for the case studies were listed in Table 2. In Sections 4.2, 4.4, and 4.5, the RES penetration rate was set to 0 %, whereas in Section 4.3, the rate was increased to 50 %. The types, capacities, and operating prices of generators were randomly assigned based on recorded statistics in 2020 on oil, coal, and natural gas-based power generation in the United States [42].

The case studies evaluated eight strategies with respect to the OPF results: the proposed strategy, five conventional strategies based on linearization, RL, heuristic optimization, and DC/AC OPF formulations, a variant of the proposed strategy with an alternative action selection, and another variant without the problem decomposition, referred to as Cases 1–8, respectively. Table 3 provides a detailed comparison among the eight strategies. In Case 2, nonlinear constraints (2)–(5) in the AC OPF problem are linearized by applying Taylor series [2]. In Case 3, the action of the RL agent consists of $\{P_g^t, |V_g^t|\}_{g,t}$ to find the solution of the AC OPF problem directly. In Case 4, a hybrid DE-PSO algorithm proposed recently is utilized to find the solution directly [5]. In Cases 5 and 6, the DC OPF and AC OPF problems are solved by an LP and NLP solver, respectively. In Case 7, the action space is reduced by increasing the number of the retained time steps (i.e., N_{ASP}, N_{ASV}) for each action element. Thus, the action sizes for each generator and the agent network sizes of Case 7 are smaller than that of Case 1, as listed in Table 4. In Case 8, the AC OPF problem is addressed solely using the proposed cutting plane method, without no decomposition involved. In this section, CPLEX was used for solving LP-based problems, while Matpower Interior Point solver (MIPS) [43] was for NLP-based problems. All case studies were performed on a system equipped with a 32 GB RAM and 3.8 GHz Intel Core i7 processor.

Sections 4.2, 4.3, and 4.4 evaluated the proposed strategy based on

Table 2
Parameters for Section 4.

Explanations	Parameters	Values	Units
# of time steps in optimization period	N_T	24	[h]
unit time step	Δt	1	[h]
max./min. of voltage magnitude	$\lvert V_n\rvert^{min}$, $\lvert V_n\rvert^{max}$	0.94, 1.06	[p. u.]
up-/downward ramp rates of generators	$R_{H,g}$, $R_{L,g}$	60, 80	[%]

Explanations	Parameters	Values		Units
		39-bus	9-bus	
# of repetitions of Step 2–4 and 2–6	$N_{R,2\text{-}4}$, $N_{R,2\text{-}6}$	30, 100	30, 80	–
# of epochs for agent training	N_{TR}	100	80	–
batch size in agent training	N_{BS}	8000	1000	–
# of hidden layers for agent	N_{HL1}, N_{HL2}	2, 2		–
learning rate for agent training	R_T	1×10^{-2}		–

Table 3
Comparison of features among the proposed and conventional OPF strategies.

Strategies		Solver	PR	CP	NC	AS
Proposed	Case 1	RL, LP	PD	○	○	–
LOPF-based	Case 2	LP	L	–	–	–
Conv. RL-based	Case 3	RL	–	–	○	–
Heuristics-based	Case 4	DE-PSO	–	–	○	–
DC OPF-based	Case 5	LP	S	–	–	–
AC OPF-based	Case 6	MIPS	–	–	○	–
Proposed w/ AS	Case 7	RL, LP	PD	○	○	○
Proposed w/o DaC	Case 8	RL, MIPS	–	○	○	–

*__AS__: different action selection, __CP__: cutting plane method, __DE-PSO__: hybrid DE and PSO algorithm, __L__: linearization, __NC__: nonlinear system constraints, __PD__: problem decomposition, __PR__: problem reformulation, __S__: simplification

the deviations in $|V_b^t|$ and Q_g^t, with line flow limits relaxed to allow deviations in these variables. In contrast, Section 4.5 examines the strategy based on the deviations in $|F_l^t|$ and Q_g^t, where the original line flow limits from [40] were reinstated. This is because that the proposed strategy is assessed in scenarios where the DC OPF strategy (i.e., Case 2) results in simultaneous deviations in $|F_l^t|$ and Q_g^t. In Section 4.6, the convergence analyses were performed by executing the proposed strategy multiple times under various daily net-load profiles.

To evaluate the solution quality of each case (i.e., Case •), certain properties (e.g., generation cost and constraint deviations) were compared with those of Cases 5 and 6, and four metrics were formulated as follows: cost difference (η_C), speed-up (η_T) [7], reduction ratio of maximum constraint deviation ($\zeta_{|V|}$, ζ_Q, and $\zeta_{|F|}$), and reduction ratio of total summation of constraint deviations ($\kappa_{|V|}$, κ_Q, and $\kappa_{|F|}$).

$$\eta_C = 100\left(C_{G,\text{Case}\bullet} - C_{G,\text{Case6}}\right)/C_{G,\text{Case6}}, \tag{36}$$

$$\eta_T = T_{\text{Case6}}/T_{\text{Case}\bullet}, \tag{37}$$

$$\zeta_x = 100\left(M_{x,\text{Case5}} - M_{x,\text{Case}\bullet}\right)/M_{x,\text{Case5}}, \tag{38}$$

$$\kappa_x = 100\left(D_{x,\text{Case5}} - D_{x,\text{Case}\bullet}\right)/D_{x,\text{Case5}}. \tag{39}$$

where C_G and T are the generation cost and computation time, respectively. M_x is the maximum value of constraint deviations on variable x_n^t, and it is calculated as follows:

$$M_x : \quad = \max\left(\max(\boldsymbol{x} - \boldsymbol{x}^{\max}, 0) + \max(\boldsymbol{x}^{\min} - \boldsymbol{x}, 0)\right). \tag{40}$$

D_x is the total summation of constraint deviations on variable x_n^t, which is calculated by (28). η_C calculates the relative difference in generation cost C_G between Case • and Case 6. Similarly, η_T measures the ratio of computation time in Case 6 to that in Case •. ζ_x calculates the relative reduction in the maximum constraint deviation compared to Case 5, while κ_x evaluates the overall reduction in the constraint deviations relative to Case 5. To ensure reliable performance evaluation, each case was run 100 times, and for all non-deterministic algorithms (Cases 1, 3, and 4), the 95 % confidence intervals of the metrics (i.e., η_C, ζ_x, and κ_x) for the operating cost and constraint deviations are reported in the form of "mean ± half-width of the interval". Since Case 2 is a deterministic linearization-based method, confidence intervals are not applicable and therefore omitted.

4.2. Comparisons of the OPF results

Tables 5 and 6 list the OPF results of Cases 1–4 of the 9-bus and 39-bus systems, respectively. As indicated by the values of $\zeta_{|V|}$, ζ_Q, $\kappa_{|V|}$, and κ_Q, the proposed strategy (i.e., Case 1) significantly reduced constraint deviations compared to Cases 2–4 even the system size increased. In the 39-bus system, the constraint deviations for Cases 3 and 4 are significantly larger than those of the proposed strategy, and Case 2 was infeasible for all given load profiles. In Case 4, the maximum deviation M_Q of Q_g^t was larger than the conventional DC OPF result (i.e., Case 5), while the total deviations D_Q of Q_g^t were slightly lower than Case 5, as shown in $\zeta_Q < 0$, $\kappa_Q > 0$ in Table 6. Regarding the operation costs, η_C of Cases 2–4 were higher than Case 1, indicating that the operating costs of both cases were higher than Case 1, as well as Case 6. The lower cost difference η_C and the higher maximum and total deviation reductions $\zeta_{|V|}$, ζ_Q, $\kappa_{|V|}$, and κ_Q, of Case 1 implies that the solution of Case 1 was closer to the optimal solution rather than that of Cases 2–4.

Additionally, the proposed strategy demonstrated a substantial reduction in the computation time for the OPF calculation. Specifically, it achieved the speed-up η_T of 26.65 and 30.58 times compared to Case 6 in the 9-bus and 39-bus systems, respectively. It is acknowledged that the RL-based strategy utilizes a pre-trained agent, whereas Case 6 solves each instance from scratch. This may appear to introduce a bias in computation time comparison; however, such an offline–online framework is a well-established paradigm in learning-based optimization studies [11,44,45]. The proposed strategy can require substantial computational load during the offline training stage but significantly reduces the time and resources required for online operation. The computation time of Case 1 was slightly higher than the time of Cases 2 (in 9-bus system) and 3 (in both systems). This is because the LP-based algorithm in Case 2 can be computationally efficient, and the DC OPF sub-problem was not solved in Case 3. However, as the system size increases, the solution qualities of both cases can be degraded as shown in the simulation results of the larger systems. The simulation results verified that the problem decomposition and RL-based cutting plane method are effective in solving the OPF problem, while ensuring the computational efficiency.

When comparing the OPF results of Cases 2–4 in both systems, Cases 3 and 4 outperformed Case 2 in terms of the constraint deviations and feasibility, because a RL algorithm and heuristic algorithm can be considered as kinds of NLP solvers, and the solution quality of the linearization method can be degraded by the increased model complexity. However, the conventional RL-based strategy (Case 3) and heuristic-based strategy (Case 4) finds the solution of the OPF problem directly. Thus, it requires more iterations for the agent training than

Table 5
Comparisons of the proposed strategy with the conventional linearization, RL, and heuristics-based strategies in the 9-bus system.

Metrics	Case 1 (9B)	Case 2 (9B)	Case 3 (9B)	Case 4 (9B)
η_C [%]	4.38 ± 0.01	4.56	16.91 ± 0.41	42.41 ± 0.98
$\zeta_{\|V\|}$ [%]	99.99 ± 0.00	51.24	48.93 ± 4.05	31.77 ± 3.93
$\kappa_{\|V\|}$ [%]	99.99 ± 0.00	35.57	86.94 ± 1.37	72.13 ± 1.91
η_T	×26.65	×42.87	×31.36	×0.0062
$T_{\text{Case1–4,6}}$ [sec]	0.1227, 0.0763, 0.1043, 527.3, 3.2706			

Table 4
Parameter comparisons between Cases 1 and 7.

Strategy	Action sizes per a generator		# of nodes in both networks of the RL agent					
			Actor network			Critic network		
	for $\widehat{P}_g^{t,\max}$	for $\|V_g^t\|$	Input	Hidden	Output	Input	Hidden	Output
Case 1	8	24	408	420	320	728	930	1
Case 7	2	6	408	420	80	488	520	1

Table 6
Comparisons of the proposed strategy with the conventional linearization, RL, and heuristics-based strategies in the 39-bus system.

Metrics	Case 1 (39B)	Case 2 (39B)	Case 3 (39B)	Case 4 (39B)
η_C [%]	3.97 ± 0.23	NA	13.97 ± 0.19	14.28 ± 0.17
$\zeta_{\|V\|}$ [%]	79.63 ± 2.20	NA	41.19 ± 2.94	38.31 ± 3.59
ζ_Q [%]	76.65 ± 2.26	NA	20.67 ± 3.83	−12.66 ± 6.98
$\kappa_{\|V\|}$ [%]	97.04 ± 0.65	NA	79.22 ± 1.78	72.27 ± 3.07
κ_Q [%]	93.01 ± 0.79	NA	34.58 ± 2.32	1.134 ± 6.94
η_T	×30.58	NA	×35.77	×0.0025
$T_{\text{Case1–4,6}}$ [sec]	0.1990, NA, 0.1701, 2452.6, 6.0845			

* η_C, ζ_x, and κ_x are expressed as "mean ± half-*width* of 95 % confidence interval".
* **NA**: not available (no feasible solution).

Case 1 as the system size increases.

To further examine the stability of the proposed strategy, the 95 % confidence intervals of η_C, ζ_x, and κ_x were evaluated. The results show that, for Case 1, the deviations from the mean were notably small, indicating that the proposed strategy provides highly consistent solutions. Moreover, the intervals for Case 1 were substantially narrower than those of the conventional RL-based method (Case 3) and the heuristic-based methods (Case 4). This implies that the proposed strategy not only achieves lower constraint deviations and operation costs on average but also yields solutions with considerably lower variability, demonstrating improved robustness against randomness in initialization and learning dynamics.

4.3. OPF results for different RES penetration rates

The RES penetration rate R_{RES} is defined as:

$$R_{RES} := \sum_{t=1}^{N_T}\sum_{g=1}^{N_{RES}} P^t_{RES,g} \Big/ \sum_{t=1}^{N_T}\sum_{b=1}^{N_B} P^t_{L,b}, \tag{41}$$

where $P^t_{RES,g}$ represents the generated power of RES g, and $P^t_{L,b}$ denotes the active power load at bus b. R_{RES} is defined as the total summation of $P_{RES,g}$ relative to that of $P_{L,b}$ throughout the optimization horizon. As the share of RES in the system grew, the net load decreased, making the violation of $|V^t_b|$ negligible even in Case 5. Thus, the simulation results in Table 7 focus on Q^t_g deviation. In Case 1, κ_Q remained similar at 0 % and 50 % penetration. However, at 50 % penetration, the cost difference η_C was 3.73 % higher, and the reduction ratio of the maximum constraint deviation ζ_Q was 5.12 % lower than at 0 % penetration. In addition, the confidence intervals of Case 1 at 50 % penetration widened slightly, indicating a modest increase in solution variability. This is attributed to the increased variance in the state **S** during the agent training. The RES's generations, particularly from wind power, follows no fixed pattern, leading to greater variance in net load profiles (i.e., the agent's states). Fig. 5(a) illustrates the normalized reward profiles along with their upward and downward variations of Case 1 for both penetration rates. The slower saturation of the rewards at the higher RES penetration indicates greater difficulty in identifying the optimal action.

Fig. 5(b) presents the reward profiles of Cases 1 and 7. These profiles not only provide empirical evidence for the convergence of the proposed strategy, but also show that the agent required less training time in Case 7 compared to Case 1. It indicates that reducing the action size improves the exploration efficiency and enhances the training performance within the same number of iterations. Table 7 also shows the confidence intervals were further narrowed in Case 7, demonstrating additional robustness gains. While Case 7 achieved higher feasibility (i.e., higher ζ_Q and κ_Q) than Case 1, it also incurred higher operation cost η_C than Case 1 at the same R_{RES}. A smaller action space implies reduced variability in generator operation, meaning that generators experience less operational strain. However, low operational fluctuations can make it more challenging to precisely respond to rapid load variations while preserving solution feasibility and cost efficiency. This highlights the inherent trade-off between reducing operational stress on generators and minimizing total generation costs.

Table 7
Evaluation of the proposed strategy based on RES penetration rates and action selection in the 39-bus system.

Metrics	Case 1		Case 7
	RES 0 %	RES 50 %	RES 50 %
η_C [%]	3.97 ± 0.23	7.70 ± 0.17	8.89 ± 0.20
ζ_Q [%]	76.65 ± 2.26	71.53 ± 2.41	82.68 ± 1.53
κ_Q [%]	93.01 ± 0.79	93.12 ± 1.45	96.06 ± 0.26
$T_{\text{Case}\bullet}$ [sec]	0.1990	0.2117	0.2119

* η_C, ζ_x, and κ_x are expressed as "mean ± half-width of 95 % confidence *interval*".

4.4. OPF results without problem decomposition method

To assess the effectiveness of the proposed decomposition approach, Case 1 was compared with its variant (i.e., Case 8) that does not incorporate the problem decomposition. In Case 8, the RL-driven cuts were integrated directly to the full nonlinear AC OPF problem defined by Eqs. (1)–(10). It indicates that the cuts for P^t_g were introduced into the problem, and it was solved using MIPS. Unlike in Case 1, the RL agent estimated the cuts only, as $|V^t_g|$ was determined by solving the AC OPF problem. To maintain a fair comparison, λ_i were kept identical in both cases.

The OPF results of both cases are listed in Table 8. While Case 8 obtained the AC-feasible solution and produced extremely small confidence intervals, which indicate high solution stability, it required much longer computation time than Case 1. This is because that Case 8 solves the NLP-based problem and estimates the optimal action. Additionally, the total generation cost incurred in Case 8 was 1.55 % higher than that in Case 6, indicating a decrease in economic efficiency. Fig. 6 illustrates the reward profiles during the training process for Cases 1 and 8. While the rewards of Case 8 achieved faster saturation compared to Case 1, a single execution time of Case 8 was approximately 31 times longer than that of Case 1. Thus, it is hard to show that Case 8 achieved computational efficiency compared to Case 1. These findings suggest that the problem decomposition is also necessary to enhance computational efficiency.

4.5. OPF results considering line flow limits

The effectiveness of the proposed strategy was assessed based on the violations of the line flow $|F^t_l|$ and reactive power generation Q^t_g constraints. As listed in Table 9, the OPF results indicate that the computation time of Case 1 was 31.06 times shorter than that of Case 6, and that the constraint violations were reduced sufficiently as indicated by $\zeta_{|F|}$, ζ_Q, $\kappa_{|F|}$ and κ_Q. This verifies that the proposed strategy is capable of obtaining feasible solutions under the condition that various constraints are considered, provided that the reward function incorporates the corresponding penalty terms. As shown in Section 4.2, the proposed strategy (i.e., Case 1) outperformed the other conventional strategies (i.e., Cases 2–4), in terms of the feasibility, cost difference, and constraint deviation reductions. Moreover, it yields substantially narrower confidence intervals than Cases 3 and 4, consistent with the observations in Section 4.2.

4.6. Convergence analysis

The comprehensive convergence performance of the proposed strategy was examined using 20-day test dataset. The RL training and execution were repeated 100 times for each daily dataset, with the initial weights of the neural networks being randomly selected for each execution. Fig. 7 shows the proportion of acceptable solutions among 100 executions for each day, where the acceptable solutions were

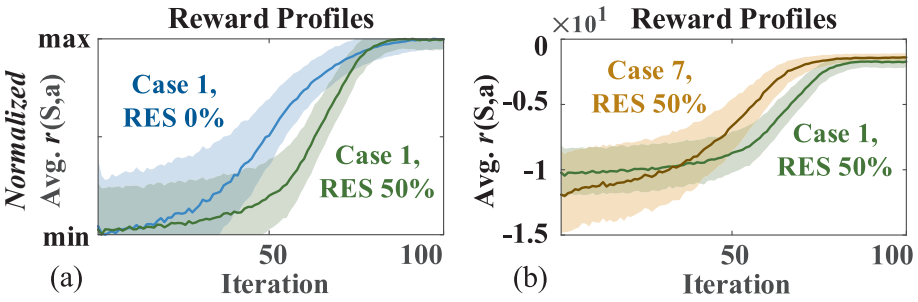

Fig. 5. Reward profile comparisons: (a) Case 1 at RES penetration rates of 0% and 50 %, and (b) Cases 1 and 7 at the rate of 50%.

Table 8
Comparison of OPF results in the 39-bus system: the proposed strategy with and without problem decomposition.

Metrics	Case 1	Case 8
η_C [%]	3.97 ± 0.23	1.55 ± 0.05
$\zeta_{\|V\|}$ [%]	79.63 ± 2.20	100.0 ± 0.00
ζ_Q [%]	76.65 ± 2.26	100.0 ± 0.00
$\kappa_{\|V\|}$ [%]	97.04 ± 0.65	100.0 ± 0.00
κ_Q [%]	93.01 ± 0.79	100.0 ± 0.00
η_T	×30.58	×0.9908
$T_{\text{Case1,6,8}}$ [sec]	0.1990, 6.0845, 6.1412	

* η_C, ζ_x, and κ_x are expressed as "mean ± half-width of 95 % confidence interval".

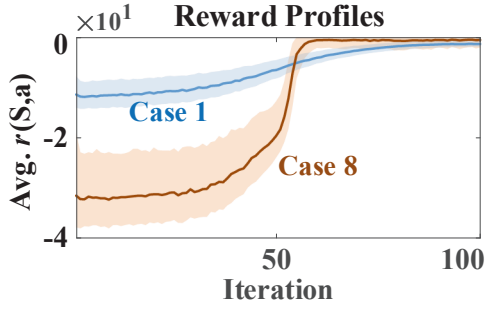

Fig. 6. Reward profiles of Cases 1 and 8.

Table 9
Comparisons of the proposed strategy with the conventional linearization and RL, and heuristics-based strategies in the 39-bus system.

Metrics	Case 1 ($\|F_l^t\|$, Q_g^t)	Case 2	Case 3	Case 4
η_C [%]	3.89 ± 0.26	NA	14.61 ± 0.21	15.38 ± 0.18
$\zeta_{\|F\|}$ [%]	99.01 ± 0.42	NA	8.36 ± 5.56	−336.4 ± 118.19
ζ_Q [%]	97.99 ± 2.57	NA	83.85 ± 11.68	35.24 ± 6.09
$\kappa_{\|F\|}$ [%]	99.94 ± 0.16	NA	48.55 ± 3.31	60.58 ± 11.71
κ_Q [%]	99.59 ± 0.61	NA	98.94 ± 0.77	71.23 ± 3.14
η_T	×31.06	NA	×40.34	×0.0034
$T_{\text{Case1–4,6}}$ [sec]	0.2351, NA, 0.1810, 2117.1, 7.3014			

* η_C, ζ_x, and κ_x are expressed as "mean ± half-*width* of 95 % confidence interval".
*NA: not available (no feasible solution).

defined as solutions that exhibited at least 85.0 % of the constraint deviation reduction ratios (i.e., $\kappa_{|V|}$ and κ_Q). The proposed strategy achieved a minimum ratio of acceptable solutions of 90.0 % and an average ratio of 98.9 % despite the inherent stochastic nature of the agent training procedure and variations in the daily load profiles, whereas none of the conventional DC OPF solutions were acceptable. Moreover, the proposed strategy reduced the maximum value of the constraint deviations (i.e., $\zeta_{|V|}$ and ζ_Q) by 80.0 % on average compared to those of the conventional DC OPF solutions. The results show the solutions of the proposed strategy converge closely to the AC-feasible area.

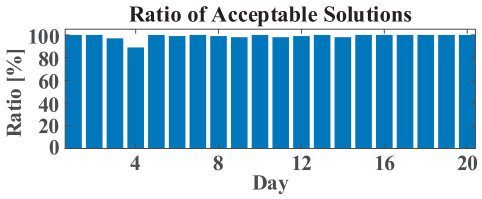

Fig. 7. The proportion of acceptable solutions obtained from multiple executions of Case 1 under various net-load profiles.

5. Conclusions

This study proposed the problem decomposition and RL-based cutting plane methods to solve the multi-period AC OPF problem with computational efficiency. The problem was decomposed into the DC OPF and AC PF sub-problems. To obtain the AC-feasible solution, additional linear inequality constraints (i.e., cuts) and voltage magnitude references were required by the DC OPF and AC PF sub-problems, respectively, and these were determined by the RL agent. Comprehensive case studies verified the proposed strategy's computational efficiency and solution feasibility compared to the conventional OPF strategies. In addition, the computational complexity analysis was conducted, and it also confirmed that the proposed strategy had lower complexity than the original AC OPF strategy. Although the proposed strategy can exhibit conservativeness in their solution, due to the characteristics of the feasible area restriction, it is the trade-off of computational efficiency. Further research will explore methods to reduce the conservativeness while maintaining the efficiency.

CRediT authorship contribution statement

Ye-Eun Jang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jaepil Ban:** Writing – review & editing, Visualization, Supervision, Investigation, Formal analysis, Data curation. **Young-Jin Kim:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis. **Chen Chen:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was partly supported by the National Research Council of

Science & Technology (NST) grant by the Korea government (MSIT) (CRC23031-000) and Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government (MCEE) (RS-2025-16063445).

Data availability

The data that has been used is confidential.

References

[1] Akbari T, Bina MT. Linear approximated formulation of AC optimal power flow using binary dis-cretization. IET Gener Transmiss Distrib 2016;10:1117–23. https://doi.org/10.1049/iet-gtd.2015.0388.
[2] Yang Z, et al. A linearized OPF model with reactive power and voltage magnitude: a pathway to improve the MW-only DC OPF. IEEE Trans Power Syst 2018;33:1734–45. https://doi.org/10.1109/TPWRS.2017.2718551.
[3] Yang Z, et al. A general formulation of linear power flow models: basic theory and error analysis. IEEE Trans Power Syst 2019;34:1315–24. https://doi.org/10.1109/TPWRS.2018.2871182.
[4] Rousis AO, et al. A planning model for a hybrid AC–DC microgrid using a novel GA/AC OPF algorithm. IEEE Trans Power Syst 2020;35:227–37. https://doi.org/10.1109/TPWRS.2019.2924137.
[5] Asija D, et al. Transmission network congestion control by DESS through interval computation and capacity optimization via hybrid DE-PSO technique. IET Gener Transmiss Distrib 2022;17:551–72. https://doi.org/10.1049/gtd2.12577.
[6] Deng Z, et al. Kriging Assisted surrogate evolutionary computation to solve optimal power flow prob-lems. IEEE Trans Power Syst 2020;35:831–9. https://doi.org/10.1109/TPWRS.2019.2936999.
[7] Huang W, et al. DeepOPF-V: solving AC-OPF problems efficiently. IEEE Trans Power Syst 2022;37:800–3. https://doi.org/10.1109/TPWRS.2021.3114092.
[8] Falconer T, Mones L. Leveraging power grid topology in machine learning assisted optimal power flow. IEEE Trans Power Syst 2023;38:2234–46. https://doi.org/10.1109/TPWRS.2022.3187218.
[9] Liu S, et al. Topology-aware graph neural networks for learning feasible and adaptive AC-OPF solutions. IEEE Trans Power Syst 2023;38:5660–70. https://doi.org/10.1109/TPWRS.2022.3230555.
[10] Lei X, et al. Data-driven optimal power flow: a physics-informed machine learning approach. IEEE Trans Power Syst 2021;36:346–54. https://doi.org/10.1109/TPWRS.2020.3001919.
[11] Sayed AR, et al. Feasibility constrained online calculation for real-time optimal power flow: a convex constrained deep reinforcement learning approach. IEEE Trans Power Syst 2023;38:5215–27. https://doi.org/10.1109/TPWRS.2022.3220799.
[12] Wu Z, et al. Physics-informed reinforcement learning for real-time optimal power flow with re-newable energy resources. IEEE Trans Sustain Energy 2025;16:216–26. https://doi.org/10.1109/TSTE.2024. 3452489.
[13] Safdarian F, et al. Temporal decomposition for security-constrained unit commitment. IEEE Trans Power Syst 2020;35:1834–45. https://doi.org/10.1109/TPWRS.2019.2947410.
[14] Chatzos M, et al. Spatial network decomposition for fast and scalable AC-OPF learning. IEEE Trans Power Syst 2022;37:2601–12. https://doi.org/10.1109/TPWRS.2021.3124726.
[15] Tu S, et al. A two-stage decomposition approach for AC optimal power flow. IEEE Trans Power Syst 2021;36:303–12. https://doi.org/10.1109/TPWRS.2020.3002189.
[16] Liu P, et al. An improved spatial branch-and-bound algorithm for non-convex optimal electricity-gas flow. IEEE Trans Power Syst 2022;37:1326–39. https://doi.org/10.1109/TPWRS.2021.3101883.
[17] Miao Z, et al. Least squares estimation based SDP cuts for SOCP relaxation of AC OPF. IEEE Trans Automat Contr 2018;63:241–8. https://doi.org/10.1109/TAC.2017.2719607.
[18] Venzke A, Chatzivasileiadis S. Convex relaxations of probabilistic AC optimal power flow for interconnected AC and HVDC grids. IEEE Trans Power Syst 2019;34:2706. https://doi.org/10.1109/TPWRS.2019.2895122 2718.
[19] Mhanna S, Mancarella P. An exact sequential linear programming algorithm for the optimal power flow problem. IEEE Trans Power Syst 2022;37:666–79. https://doi.org/10.1109/TPWRS.2021.309 7066.
[20] Tang Y, et al. Reinforcement learning for integer programming: Learning to cut. Proceedings of the 37th International Conference on Machine Learning, PMLR 2020 2020;119:9367–9376.
[21] Paulus MB, et al. Learning to cut by looking ahead: cutting plane selection via imitation learning. Proceedings of the 39th International Conference on Machine Learning, PMLR 2022 2022;162:17584-17600.
[22] Fang X, et al. AC feasibility restoration in market clearing: Problem formulation and improvement. IEEE Trans Ind Inform 2022;18:7597–607. https://doi.org/10.1109/TII.2021.3139619.
[23] Han J, et al. FRMNet: a feasibility restoration mapping deep neural network for AC optimal power flow. IEEE Trans Power Syst 2024;39:6566–77. https://doi.org/10.1109/TPWRS.2024.3354733.
[24] Wang Z and Chiang H. On the feasibility of AC and DC optimal power flow models: analytics and comparison. 2021 IEEE Power & Energy Society General Meeting, 2021 IEEE PESGM 2021:1-5.
[25] Jang YE, et al. Optimal HVAC system operation using online learning of interconnected neural networks. IEEE Trans Smart Grid 2021;12:3030–42. https://doi.org/10.1109/TSG.2021.3051564.
[26] Teng J. A direct approach for distribution system load flow solutions. IEEE Trans Power Deliv 2003;18:882–7. https://doi.org/10.1109/TPWRD.2003.813818.
[27] Li S. Power flow modeling to doubly-fed induction generators (DFIGs) under power regulation. IEEE Trans Power Syst 2013;28:3292–301. https://doi.org/10.1109/TPWRS.2013.2251914.
[28] Haarnoja T, et al. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. Proceedings of the 35th International Conference on Machine Learning, PMLR 2018 2018;80:1861-1870.
[29] Zhou D, et al. Learning to classify with incremental new class. IEEE Trans Neural Netw Learn Syst 2022;33:2429–43. https://doi.org/10.1109/TNNLS.2021.3104882.
[30] Tavakoli A, et al. Action branching architectures for deep reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, AAAI, Vol. 32, No. 1; 2018. p. 4131–8. https://doi.org/10.1609/aaai.v32i1.11798.
[31] Jabr RA, et al. Robust multi-period OPF with storage and renewables. IEEE Trans Power Syst 2015;30:2790–9. https://doi.org/10.1109/TPWRS.2014.2365835.
[32] Chow JH, et al. Time-scale modeling of dynamic networks with applications to power systems. Berlin, Germany: Springer; 1982.
[33] Pan Z, et al. Feasible region method based integrated heat and electricity dispatch considering building thermal inertia. Appl Energy 2017;192:395–407. https://doi.org/10.1016/j.apenergy.2016.09.016.
[34] Glover J, et al. Power system analysis and design. 5th ed. Boston, MA, USA: Cengage; 2011.
[35] Ploskas N, Samaras N. Linear programming using MATLAB. Cham, Switzerland: Springer; 2018.
[36] Pan X, et al. DeepOPF: a deep neural network approach for security-constrained DC optimal power flow. IEEE Trans Power Syst 2021;36:1725–35. https://doi.org/10.1109/TPWRS.2020.3026379.
[37] Narayanan D, et al. Efficient large-scale language model training on gpu clusters using megatron-LM. In: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. 2021, p. 1–15. https://doi.org/10.1145/3458817.3476209.
[38] Li S, et al. Pytorch distributed: Experiences on accelerating data parallel training. In: Proceedings of the VLDB Endowment. 2020, p. 3005–3018. https://doi.org/10.14778/3415478.3415530.
[39] Nguyen Q, et al. Loss minimization with optimal power dispatch in multi-frequency HVac power systems. IEEE Trans Power Syst 2020;35:1979–89. https://doi.org/10.1109/TPWRS.2019.2953161.
[40] Athay T, et al. A practical method for the direct analysis of transient stability. IEEE Trans Power App Syst 1979;PAS-98:573–84. https://doi.org/10.1109/TPAS.1979.319407.
[41] NYISO. Load Data – NYISO, https://www.nyiso.com/load-data/; 2022 [accessed 26 Feb 2022].
[42] EIA. Annual Energy Outlook 2021 – Table: Table 9. Electricity Generating Capacity, https://www.eia. gov/outlooks/aeo/data/browser/#/?id=9-AEO2021&region=0-; 2021 [accessed 26 Feb 2022].
[43] Zimmerman RD, et al. MATPOWER: steady-state operations, planning, and analysis tools for power systems research and education. IEEE Trans Power Syst 2011;26:12–9. https://doi.org/10.1109/TPWRS.2010.2051168.
[44] Sayed AR, et al. Deep reinforcement learning-assisted convex programming for AC unit commit-ment and its variants. IEEE Trans Power Syst 2024;39:5561–74. https://doi.org/10.1109/TPWRS. 2023.3340674.
[45] Zheng S, et al. Offline-training online-execution framework for volt-var control in distribution networks. J Mod Power Syst Clean Energy 2025;13:1726–37. https://doi.org/10.35833/MPCE.2024.000887.