Research on Optimal Strategy of Peak-shaving of Photovoltaic Grid-connected System Based on Simulated Annealing-Q Learning Algorithm

Xu Jun¹, Chen Jinhui¹ and Zhang Zhe¹
¹ College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu, 210023, China
1018051321@njupt.edu.cn

Abstract. In order to alleviate the problem of large-scale grid-connected photovoltaics and increase the pressure of grid peak regulation, this paper proposes a dynamic economic dispatch method considering the system peak regulation margin. The system peak shaving margin is proposed to deal with the flexibility problem in economic dispatch and improve the enthusiasm of thermal power unit peak shaving. The economic dispatch model is transformed into a dynamic programming model and solved by the simulated annealing-Q learning algorithm. The analysis of the results shows that the proposed dynamic programming strategy can effectively improve the operating economy of the system and increase the flexibility of peak shaving of thermal power units.

1. Introduction
Under the situation of large-scale grid connection of new energy, the intermittent, volatile, and random characteristics of new energy output have aggravated the peak-shaving pressure of the power grid. As the main peak-shaving resource of the power grid at this stage, thermal power units have to pass deep peak-shaving and start-ups. Stopping peak shaving to relieve peak shaving pressure, leading to deterioration of unit operating conditions and increased costs, severe decline in unit peak shaving enthusiasm, and local blockage of the grid caused by new energy power generation peaks, which caused the peak shaving capacity of thermal power units to not be fully utilized. Therefore, it is urgent to analyze the comprehensive benefits of new energy consumption and grid peak shaving from the perspective of the whole society, and study and formulate the operation strategy of grid peak shaving auxiliary service under new energy consumption.

For peak shaving technology, the existing research is mostly carried out from the energy side. According to the development scale of new energy and the operating status of thermal power units, the goal is to achieve the coordinated peak shaving of new energy and thermal power to solve the optimal distribution of energy. Literature [1] stimulated thermal power units to participate in peak shaving by formulating peak shaving rights trading methods, formulating wind power secondary consumption methods, and proposing a combined heat and power economic dispatch model, which reduced the system’s abandonment rate and improved the overall economic benefits. Literature [2-3] analyzes the economics of deep peak shaving of thermal power plants from the perspective of technology, thermal power companies, wind power companies, and society. Literature [4] sets the output prohibition zone of thermal power units, judges the operation status of pumped storage units through the reserve capacity rate and load rate, and realizes the synchronization and coordination of energy storage and
new energy. Literature [5] established a decision-making model for virtual energy storage and deep peak shaving to participate in the reserve to ensure the peak shaving reserve capacity of the system. Literature [6-7] provides methods for coordinated peak shaving in multiple regions to ensure the economic benefits of cross-regional network peak shaving.

Reinforcement learning, as the forefront of algorithms for dealing with complex problems, can be solved quickly and has strong convergence. Literature [8] proposed a new energy consumption strategy based on flexibility and balance theory, and the equilibrium solution was obtained by algorithm. Literature [9-10] applies reinforcement learning to microgrid dispatch in different scenarios, and achieves maximum benefit and convergence under the disturbance of photovoltaic and load. Literature [11] uses a reinforcement learning algorithm to develop a strategy for managing storage systems in the grid. This strategy is applied to predict uncertainty with parameterized modification, and the effect is good. Literature [12] considers the uncertainty of load demand, renewable energy and electricity prices, models microgrid energy management as a Markov process, and develops a deep reinforcement learning method to solve the Markov process.

This paper proposes a dynamic planning and scheduling strategy based on machine learning. The enthusiasm for peak shaving of thermal power units is mobilized by proposing the system peak shaving margin, and the cost of the power grid during peak shaving period is paid attention to. The economic dispatch model is transformed into dynamic planning, and the dispatch mode of flexible consumption of photovoltaics is proposed, which is solved by simulated annealing-Q learning algorithm. A simulation example shows the effectiveness of the proposed method.

2. Economic dispatch of photovoltaic grid-connected power system

Economic dispatch based on peak shaving margin is to improve the flexibility of system peak shaving in traditional economic dispatch. By proposing the peak shaving margin to mobilize the enthusiasm for the operation of thermal power units, the balanced allocation of peak shaving resources can be realized. In this article, the peak shaving margin is mainly applied to thermal power units.

Objective function:

\[
\min \ f = \sum_{k=0}^{K} C_1 + \sum_{k=0}^{K} C_2 + \sum_{k=0}^{K} C_3
\]

(1)

Where \( C_1 \) is the coal consumption cost of the thermal power unit; \( C_2 \) is the shutdown cost of the thermal power unit; \( C_3 \) is the compensation income of the deep peak shaving of the thermal power unit.

\[
C_1 = \sum_{i=1}^{N} \left( g_i P_{i,k}^2 + b_i P_{i,k} + c_i \right) \times C_{i,\text{coal}}
\]

(2)

\[
C_2 = \sum_{k=0}^{K} \sum_{i=1}^{N} I_{i,k} \times C_{i,k}
\]

(3)

\[
C_3 = \sum_{i=1}^{N} \sum_{j=1}^{4} \left( c_{\text{mbc},i,j,k} \times P_{\text{deep},i,j,k} \right)
\]

(4)

\( a, b, c \) are the coal consumption characteristic parameters of the i-th thermal power unit; \( N \) is the number of peak-shaving units; \( c_{\text{mbc},i,j,k} \) is the peak-shaving compensation price based on the i-th group in the j-th gear in the k period; \( P_{\text{deep},i,j,k} \) is the deep peak shaving power of the i-th unit in the j-th gear in the k period. \( C_{i,k} \) is the emergency shutdown cost of the i-th peak-shaving unit in time period k; \( I_{i,k} \) is the operating state of the i-th thermal power unit in time period k.

Constraints:

1) Output constraints of thermal power units

\[
P_{i,k,\text{min}} \leq P_{i,k} \leq P_{i,k,\text{max}}
\]

(5)

Where \( P_{i,k,\text{min}} \) and \( P_{i,k,\text{max}} \) are the minimum and maximum output limits of unit i in period k.
2) Power balance constraint
\[ \sum_{j=1}^{n} l_{ij} P_{ij,k} + P_{pv,k} + P_{L,k} = L_{k} \] (6)
Where \( P_{en,k} \) is the network loss; \( P_{pv,k} \) is the photovoltaic consumption; \( L_{k} \) is the load of the time period \( k \).

3) Unit start and stop time constraints
\[ T_{on,i,k} \geq M_{on,i}^{\min} \] (7)
\[ T_{off,i,k} \geq M_{off,i}^{\min} \] (8)
Where \( M_{on,i}^{\min} \) and \( M_{off,i}^{\min} \) are the minimum start-up time and minimum stop time of the \( i \)-th unit; \( T_{on,i,k} \) and \( T_{off,i,k} \) are respectively the running time and shutdown time of the \( i \)-th unit in period \( k \).

4) Crew gradeability constraints
\[ P_{i,k} - P_{i,k-1} \leq P_{up,i} \] (9)
\[ P_{i,k-1} - P_{i,k} \leq P_{down,i} \] (10)
Where \( P_{up,i} \) and \( P_{down,i} \) are the maximum changes when the \( i \)-th unit increases/decreases output; \( P_{i,k} \) and \( P_{i,k-1} \) are the output of the \( i \)-th thermal power unit in period \( k \) and \( k-1 \).

5) System peak shaving margin constraint
\[ 0.8 < E_{k} = P_{en,k} / P_{ed,k} < 1.4 \] (11)
Where \( E_{k} \) is the system peak shaving margin; \( P_{ed,k} \) is the peak shaving demand; \( P_{en,k} \) is the peak shaving capacity. If the peak shaving margin of the system is greater than 1, the photovoltaic can be fully consumed, and if the peak shaving margin is less than 1, the full photovoltaic can be abandoned. \( P_{ed,k} \) and \( P_{en,k} \) are obtained by (12) and (13) respectively.
\[ P_{ed,k} = \mu P_{pv,k} + |L_{k+1} - L_{k}| \] (12)
\[ P_{en,k} = \sum_{i=1}^{n} \min(P_{i,k}, P_{i,max}, P_{i,min}) \] (13)
Where \( \mu \) is the percentage coefficient of the spinning reserve capacity; \( P_{by,k} \) is the spinning reserve capacity of the thermal power unit; \( L_{k+1} \) and \( L_{k} \) are the equivalent loads in the period \( k+1 \) and period \( k \).

3. Solution method
Discrete the maximum range \([-P_{peak,\text{max}}, P_{peak,\text{max}}]\) of grid peak shaving power demand in each decision period \( k \) into \([N_{pp}, \cdots, 0, \cdots, N_{pp}]\) a total of \( 2N_{pp} + 1 \) levels, \( a'_{i} \in \{0,1\} \) is the action of the \( i \)-th thermal power unit in each decision period, where \( a'_{i} = 0 \) represents the shutdown of the thermal power unit, \( a'_{i} = 1 \) represents the unit is in operation. There are a total of \( 2N_{ag} + 1 \) levels of power adjustment actions \( a_{ag} \in [-N_{ag}, \cdots, 0, \cdots, N_{ag}] \) of thermal power units in decision period \( k \). The system action includes the operation behavior of the thermal power unit \( a_{ag} \in D_{ag} = \{0,1\} \) and the power adjustment action of the thermal power unit \( a_{ag} \in D_{ag} = \{0,1\} \). The system action vector can be expressed as \( a = \{a_{ag}, a_{ag}\} \in D \), and its action set can be expressed as \( D = D_{ag} \times D_{ag} \).

This paper proposes a limited time period learning algorithm based on the simulated annealing method to solve the established optimal scheduling model, and the specific process is as follows:

Step 1: Initialize the Q value table, learn the total number of sample tracks \( M \), single sample track decision period \( K \), initial learning rate \( \alpha_{m} \) and learning rate update coefficient \( \mu_{\alpha} \), discount factor \( \gamma \).
simulated annealing temperature $T_{\text{temp}}$, simulated annealing coefficient $\mu_{\text{temp}}$, and set the sample track $m = 0$;

Step 2: $k = 0$, Initialize the system status data randomly to determine the system status, that is, the current system load, the total real-time output of the regional photovoltaic power plants, and the actual output completion status of the regional thermal power units;

Step 3: According to Q value and greedy strategy, choose action $a_k$ with greedy strategy: current state $s_k$ corresponds to the greedy action $a_k^{\text{greedy}} = \arg\min_{a_k} Q(s_k, a_k)$ of peak shaving output of the grid thermal power unit with the best Q value, and randomly select action $a_k^{\text{rand}}$, which is randomly generated A feasible solution. If $e^{Q(s_k, a_k^{\text{greedy}}) - Q(s_k, a_k^{\text{rand}})))/T_{\text{temp}} < \text{random}(0,1)$, the system action is $a_k = a_k^{\text{greedy}}$, otherwise $a_k = a_k^{\text{rand}}$. If $k < K - 1$, skip to step 4, if $k = K - 1$, skip to step 5;

Step 4: Execute the action $a_k$ selected by the current system, observe the system state $s_{k+1}$ in the next decision period according to the built system model, and calculate the system cost $r_k$ generated during the state transition by executing action $a_k$ in the decision period $k$, update Q Value, update the strategy at the same time, let $k = k + 1$, return to step 3;

Step 5: Execute the current action $a_{k+1}$, observe the cost $r_k$ and the final state cost $r_{Q}(s_k)$ generated after a decision period, update the Q value according to the formula, and update the strategy;

$$Q'(s_k, a_k) = Q(s_k, a_k) + \alpha_m (r_k - (Q(s_k, a_k) + \gamma \min(Q(s_{k+1}, a_{k+1})))$$

(14)

Step 6: Let $m = m + 1$; if $m < M$, update the simulated annealing temperature according to the simulated annealing coefficient $\mu_{\text{temp}}$: $T_{\text{temp}} = T_{\text{temp}} \cdot \mu_{\text{temp}}$, and update the learning rate according to the learning rate update coefficient $\alpha_m$: $\alpha_m = \alpha_m \cdot \alpha_m$, return to step 2, otherwise exit the loop end procedure.

4. Case study

The application cases in this article include 6 thermal power units and 4 photovoltaic power plants. The specific parameters of the thermal power unit are shown in the table 1.

| Thermal power unit capacity/MW | a(t/MWh^2) | b(t/MWh) | c(t) | Fuel consumption during peak-shaving stage of fuel injection depth (t) |
|-------------------------------|------------|----------|------|-------------------------------|
| 2×1000 | 0.000 0744 | 0.159 37 | 8.263 70 | 0.265 |
| 2×660 | 0.000 0810 | 0.133 56 | 11.088 0 | 0.240 |
| 1×300 | 0.000 0921 | 0.145 45 | 14.139 5 | 0.200 |

Based on the randomness of photovoltaic output, this paper proposes the operation mode of photovoltaic elastic consumption. Modes 1, 2, and 3 are operating modes with full-consumption photovoltaics, with a flexible photovoltaic consumption interval of [80%, 85%] and a photovoltaic flexible consumption interval of [90%, 95%]. The operating results of each mode are shown in Table 2.

| Operating expenses (Ten thousand yuan) | Peak shaving compensation (Ten thousand yuan) | System peak shaving margin |
|----------------------------------------|---------------------------------------------|---------------------------|
| Mode 1 | 137.57 | 45.4 | 1.4347 |
| Mode 2 | 63.08 | 73.21 | 1.342 |
| Mode 3 | 30.98 | 97.81 | 1.1341 |
It can be seen from the data in Table 2 that the peak shaving margin of the system is positively correlated with the operating cost, that is, the operating cost increases with the increase of the peak shaving margin. The operating cost of mode 3 is significantly better than that of mode 1 and mode 2, and the peak shaving compensation of mode 3 is higher than that of mode 1 and mode 2. The operating cost of Mode 1 is the highest. The fundamental reason is that thermal power units have to be shut down and peak-shaving in order to fully consume photovoltaics. It can be seen from the improved Q learning algorithm that modes 2 and 3 do not stop peak shaving after flexibly absorbing photovoltaics, which effectively reduces system operating costs and ensures the economy and flexibility of system operation. The output of the thermal power unit in each mode is shown in Figure 1, 2, 3.

![Figure 1. The output of mode 1](image1)
![Figure 2. The output of mode 2](image2)
![Figure 3. The output of mode 3](image3)
![Figure 4. Load factor of thermal power unit in each mode](image4)

It can be seen from Figure 1, 2, 3 that when the simulated annealing-Q learning algorithm is used to solve the peak shaving strategy, Unit 3 in Mode 1 starts and stops peak shaving. Mode 2 and Mode 3 are affected by photovoltaic elastic consumption and load, the scheduling strategy instead of starting and stopping peak load regulation, each thermal power unit actively responds to adjust output, has a faster response speed and strong peak load regulation ability to the lack of photovoltaic power.

It can be seen from Figure 4 that the average load rate of mode 1 is not significantly improved compared to the average load rate of the unoptimized system, and the operating cost of this mode is the highest, which fails to ensure economy. Modes 2 and 3 have improved the average load rate of the system to a certain extent, and have a certain guarantee for the stability of the system. In these two modes, the simulated annealing learning algorithm not only ensures the elastic constraints of photovoltaic consumption, but also It ensures that the operating cost of the system and the peak shaving margin are balanced, that is, both economy and flexibility are taken into account.

Taking the operating result at 13:00 as an example, the peaking margin of mode 1 is 1.4506, the peaking margin of mode 1 is 1.4506, the peaking margin of mode 2 is 1.2542, and the peaking margin of mode 3 is 1.195. For mode 1, when the system operating cost of fully absorbing photovoltaics is the
lowest as the goal, the peak shaving margin of the system during peak shaving period reaches 1.4506. When the photovoltaics are fully absorbing, the system peaking margin under the operating mode of mode 1 is followed. The peak shaving margin of Broad Mode 2 and 3, at this time the system peak shaving margin is too large and the operating cost is too high, the scheduling strategy of Mode 1 is obviously unreasonable. When modes 2 and 3 aim at the lowest operating cost of the system that can absorb photovoltaics flexibly, the peak shaving margin of the system is greater than 1, indicating that part of the photovoltaic can be used as spare capacity at this time, and the peak shaving margin of mode 2 is larger. In addition, the operating cost is relatively high. The peak shaving margin of Mode 3 is slightly lower than that of Mode 2, but the operating cost is significantly better than that of Mode 2. Considering comprehensively, the operating mode of Mode 3 is more reasonable.

5. Conclusion
This paper proposes a photovoltaic grid-connected peak shaving strategy based on simulated annealing-Q learning algorithm. The system peak shaving margin is proposed, which fully taps the flexibility of thermal power units; in the model, flexible consumption of photovoltaics is considered, and the peak shaving auxiliary service resources provided by each thermal power unit are reasonably invoked. The results show that the operating cost of the system increases with the increase of the peak shaving margin of the system. If the photovoltaic output level exceeds expectations, the flexible consumption of photovoltaics can be considered. The simulated annealing-learning algorithm can better deal with the uncertainty problem, which not only ensures the elastic constraints of photovoltaic absorption, but also ensures the balance between the system peak shaving margin and the operating cost.

References
[1] YANG Lijun, LIANG Xuri, WANG Xinrui, et al. (2020) Combined Heat and Power Economic Dispatching Considering Peak Regulation Right Trading to Improve Secondary Accommodation Capability of Wind Power. Power System Technology, 44(05):1872-1880.
[2] LIN Li, ZOU Lanqing, ZHOU Peng, et al. (2017) Multi-angle Economic Analysis on Deep Peak Regulation of Thermal Power Units with Large-scale Wind Power Integration. Automation of Electric Power Systems, 41(07):21-27.
[3] LIN Li, TIAN Xinyu. (2017) Analysis of Deep Peak Regulation and Its Benefit of Thermal Units in Power System With Large Scale Wind Power Integrated. Power System Technology, 41(07):2255-2263.
[4] YANG Lijun, ZHANG Zhao, LÜ Xuejiao, et al. (2017) Economic Dispatch on Power Systems With Wind Power and Pumped-Storage Station Considering Prohibited Zones Under Influence of Reserve Capacity. Power System Technology, 41(05):1548-1554.
[5] XUE Chen, REN Jing, ZHANG Xiaodong, et al. (2019) A Reserve Decision Model for High-Proportional Renewal Energy Integrated Power Grid Based on Deep Peak-Shaving and Virtual Storage. Electric Power, 52(11):35-43.
[6] XU Zhiqiang, ZHANG Fuqiang, CHEN Shizhe, et al. (2018) Analysis of Multi-Region Coordinated Peak-Shaving and Peak-Valley Mutual Operation Mode under the Energy Internet. Electric Power, 51(08):64-69.
[7] DING Ming, SHI Jianxiong, HAN Pingping, et al. (2020) Integrated control strategy of optical storage system participating in power grid frequency and peak regulation. Electric Power: 1-7 [2020-10-05]. http://kns.cnki.net/kcms/detail/11.3265.TM.20200226.1457.009.html.
[8] LI Hongzhong, WANG Lei, LIN Dong, et al. (2019) A Nash Game Model of Multi-agent Participation in Renewable Energy Consumption and the Solving Method via Transfer Reinforcement Learning. Proceedings of the CSEE, 39(14):4135-4150.
[9] LIU Jinhua, KE Zhongming, ZHOU Wenhui. (2020) Microgrid energy dispatch strategy and optimization based on reinforcement learning. Journal of Beijing University of Posts and Telecommunications, 43(01):28-34.
[10] WANG Yadong, CUI Chenggang, QIAN Shensheng, et al. (2019) Research on Microgrid Energy Storage Dispatching Strategy Based on Deep Reinforcement Learning. Renewable Energy, 37(08):1220-1228.

[11] Vinícius de Carvalho Neiva Pinheiro, Francato A L, Powell W B. (2020) Reinforcement learning for electricity dispatch in grids with high intermittent generation and energy storage systems: A case study for the Brazilian grid. International Journal of Energy Research, (3).

[12] Ji Y, Wang J, Xu J, et al. (2019) Real-Time Energy Management of a Microgrid Using Deep Reinforcement Learning. Energies, 12(12):2291.