



THE HONG KONG
POLYTECHNIC UNIVERSITY

香港理工大學

Pao Yue-kong Library

包玉剛圖書館

Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

By reading and using the thesis, the reader understands and agrees to the following terms:

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

**MULTI-OBJECTIVE GENERATION
CONTROL AND OPTIMIZATION FOR
LARGE-SCALE POWER SYSTEM DISPATCH**

ZHOU BIN

Ph.D

The Hong Kong Polytechnic University

2013



The Hong Kong Polytechnic University

Department of Electrical Engineering

**MULTI-OBJECTIVE GENERATION
CONTROL AND OPTIMIZATION FOR
LARGE-SCALE POWER SYSTEM DISPATCH**

ZHOU BIN

A thesis submitted in partial fulfillment of the requirements
for the Degree of Doctor of Philosophy

August 2012

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written or material which has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____._____(Signed)

ZHOU Bin (Name of student)

Abstract

With the renew interest in energy-saving generation dispatch and the growing environmental concern for the power industry, developing a modernized dispatch infrastructure, associated with its optimized generation scheduling and dispatch strategies, have become a global priority to contribute towards the formation and development of the future smart grid. Under the smart and green grid paradigm, automatic generation control (AGC) and electric power dispatch will play an important role in improving the long-term control performance and sustainability of energy management system (EMS) for the efficient operation of large-scale interconnected power systems. Therefore, based on the foundation laid by the pioneering research studies already presented, this thesis strives to make further investigations for the design of the AGC strategies and power dispatch algorithm with multi-objective operation in mind.

The AGC performance in interconnected power system operation is used to measure against the control performance standards (CPS) released by North American Electric Reliability Council (NERC) in 1997. Since the introduction of this new NERC CPS, there are fundamental changes in the conventional AGC control philosophies. So far, researches on AGC strategies under CPS were mostly based on the classical proportional-integral (PI) control structure. Even with the wide adoption of the CPS nowadays, the existing AGC systems have commonly not yet been optimized to fully explore the potential of CPS standards. Dynamic response studies showed that the AGC system under CPS in fact can better be formulated as an uncertain stochastic system from the statistical and probabilistic point of view. Consequently, this thesis harnesses reinforcement learning (RL) and Markov decision process (MDP) techniques to develop a robust and adaptable AGC optimized for the NERC's CPS standards and optimal

relaxed operation.

Over the years, in order to determine the optimal steady-state operation of dispatchable generators, economic dispatch (ED) is a standard function in the EMS so that the total generation cost is minimized while satisfying a set of operational and physical constraints. However, nowadays more operating objectives, such as energy conservation and emission reduction, should also be considered to establish a multiobjective power dispatch (MOPD) optimization. Consequently, this thesis would also propose a new algorithm based on the Pareto optimality to solve this highly constrained large-scale MOPD problem with multiple contradictory and noncommensurable objectives.

First and foremost, in this thesis, a new concept referred as optimal relaxed AGC control is proposed to allow the AGC plants to maneuver in less costly manner in finding the optimal economic dispatch policy on the premise of complying with the CPS1 and CPS2 metrics. A Q-learning based AGC controller, in which the CPS control and relaxed control objectives are formulated as multi-criteria reward function via linear weighted aggregate method, is presented for interactive self-learning control rules to maximize the long-term discounted reward. In addition, for the thermal-dominated power systems, to overcome the long time-delay problem caused by the steam turbine of AGC thermal units in the secondary frequency control loop, a multi-step $Q(\lambda)$ learning based AGC is proposed to regulate the degree of CPS compliance and relaxation for the desirable relaxed control. The effectiveness and validity of the proposed RL based AGC strategies have been successfully verified on a two-area load frequency control (LFC) model and the practical-sized China Southern Power Grid (CSG) with four control areas.

The goal of average reward RL is to maximize the long-term average rewards of a generic system. This coincides with the design objective of the CPS which was established to improve the long-term performance of an AGC used for real-time control of interconnected power systems. Therefore, a novel $R(\lambda)$ imitation

learning ($R(\lambda)$ IL) method based on average reward optimality criterion is proposed to develop an optimal AGC under the CPS. This $R(\lambda)$ IL based AGC is capable of operating online in real-time with high CPS compliances and fast convergence rate in the imitation pre-learning process. Its capability to learn the control behaviors of the existing AGC by observing the system variations enable it to overcome the main difficulty in practically applying the conventional RL controllers, in which an accurate power system model is required for the offline pre-learning process, and significantly enhance the learning efficiency and control performance for power generation control in various power system operation scenarios.

On the other hand, an equilibrium-inspired multiple group search optimizer (MGSO) is developed to solve the highly constrained MOPD problem. In this algorithm, a stochastic learning automata based synergistic learning is employed to achieve the information interaction and credit assignment among groups for cooperative search, and an average linkage-based hierarchical clustering is used to provide the power dispatcher with a manageable and representative Pareto-optimal front (PF). Furthermore, the Nash equilibrium point is first introduced to identify the best compromise solution from the PF. An alternative constraint handling, which separates constraints and objectives with different search strategies, is presented to orient towards a well-distributed PF. In addition, two special implementations, space reduction strategy and chaotic sequence dispersion, have been incorporated in convergence process to facilitate the local exploitation and global exploration of PF solutions respectively. Simulation tests on the benchmark power systems, including the IEEE 30-bus system with 6 units and the IEEE 118-bus system with 54 units, have demonstrated the enhancement and superiority of the proposed MGSO algorithm, and confirmed its potential to cope with this type of large-scale multiobjective optimization problems with high-dimensional and more objective functions.

Acknowledgement

First and foremost, I would like to express my deep and sincere appreciation to my chief supervisor, Dr. Kevin K. W. Chan, for his continuous support and invaluable suggestions throughout my PhD study and research. His enthusiasm, patience, rich knowledge and creative way of thinking have been of significant value for me. I am very grateful to my external co-supervisor, Prof. T. Yu, for his kindness of offering useful resources, constructive criticism and encouragement. I learned a great deal and gained valuable inspiration through discussions with him on the optimization and control methodologies. It has been an honor to study under their guidance, and this academic experience in Hong Kong will have a profound influence on my future professional and personal development.

I reserve special gratitude to Prof. Q. H. Wu at the University of Liverpool for his insightful comments and deliberations on my research papers.

I have spent three really memorable and substantial years with the team of I heartily appreciate them all. Special thanks to my best friend, Mr. S. W. Xia, who gave me a great help on Matlab programming. Moreover, I would like to convey my thanks to all staffs of the Department of Electrical Engineering for their help and assistance provided, and also to my paper reviewers and thesis examiners for their valuable comments to this research. My sincere thanks also go to everyone who directly or indirectly offered his or her help while working on this thesis.

My genuine appreciation and love are due to my dear parents and my wife who have borne the burden of unconditionally supporting me in chasing dreams.

Last but not least, I would like to acknowledge the support from my Ph.D. studentship awarded by The Hong Kong Polytechnic University.

Table of Contents

Abstract	I
Acknowledgement	IV
Table of Contents	V
Lists of Figures, Tables and Abbreviations	IX

Chapter I

Introduction	1
1.1 Background of Research	1
1.2 Incentives of Thesis.....	5
1.3 Primary Contributions	8
1.4 Thesis Layout	12
1.5 List of Publications	15

Chapter II

Essentials for Multiobjective Generation Control and Optimization	18
2.1 Overview of Power Generation Dispatch and Control	18
2.2 Current State of the Art	20
2.2.1 Control Performance Standards	21
2.2.2 State-of-the-Art AGC Schemes	24
2.2.3 State-of-the-Art MOPD optimization	27
2.3 Reinforcement Learning & Markov Decision Processes	30
2.3.1 Markov Decision Processes	30
2.3.2 Reinforcement Learning Methodology	31
2.4 Multiobjective Optimization Algorithms	34
2.5 Summary	36

Chapter III

Stochastic Optimal CPS Control Methodology for Interconnected

Power Systems Using Q-Learning	37
3.1 Introduction	37
3.2 Q-Learning	38
3.3 Design of Q-Learning Based AGC	41
3.3.1 Controller Framework.....	41
3.3.2 State and Action Space Discretization	42
3.3.3 Control Objective of QAGC	43
3.3.4 Design of Reward Function	45
3.3.5 Q-learning Parameter Settings	46
3.3.6 Execution Steps of QAGC	47
3.4 Procedure of Semisupervisory Group Prelearning.....	48
3.5 Simulation Studies	50
3.5.1 Investigated System	50
3.5.2 Performance Results	51
3.6 Summary	55

Chapter IV

Stochastic Optimal Relaxed AGC Methodology in Non-Markov

Environment Based on Multi-step $Q(\lambda)$ Learning	56
4.1 Introduction	56
4.2 Multi-Step $Q(\lambda)$ Learning.....	57
4.3 Thermal-Dominated AGC Process under CPS	60
4.4 Design of Optimal Relaxed AGC Strategy	62
4.4.1 State and Action Space Discretization	62
4.4.2 Reward Function	63
4.4.3 $Q(\lambda)$ Learning Parameters	65
4.4.4 Implementation Procedures of $Q(\lambda)$ AGC	66

4.5	Simulation Results and Analysis.....	67
4.5.1	Study on Two-Area LFC Power System Model	68
4.5.2	Study on China Southern Power Grid.....	73
4.6	Summary	81

Chapter V

A Novel $R(\lambda)$ Imitation Learning Methodology for Online AGC

Optimization	83	
5.1	Introduction	83
5.2	Multi-Step $R(\lambda)$ Learning.....	84
5.3	Design of Optimal $R(\lambda)$ ILAGC.....	87
5.3.1	$R(\lambda)$ Learning for AGC Strategy.....	87
5.3.2	Imitation Pre-learning Process	89
5.4	Simulation Studies	93
5.4.1	Evaluation on Two-Area LFC Power System Model	93
5.4.2	Evaluation on China Southern Power Grid.....	96
5.5	Summary	98

Chapter VI

Equilibrium-Inspired Multiple Group Search Optimizer with

Synergistic Learning for Multi-objective Electric Power Dispatch

6.1	Introduction	100
6.2	Problem Formulation	102
6.2.1	MOPD Objectives	102
6.2.2	MOPD Constraints	103
6.3	Proposed Multiple Group Search Optimizer.....	105
6.3.1	Algorithm Framework.....	105
6.3.2	Initialization	106
6.3.3	Variable-size External Repository	107

6.3.4	Space Reduction-based Scanning Strategy for Producer.....	108
6.3.5	Synergistic Learning for Scroungers.....	109
6.3.6	Chaotic Sequence Dispersion for Rangers.....	111
6.3.7	Constraint Handling Strategy.....	112
6.3.8	Pruning Pareto Set.....	114
6.3.9	Nash Equilibrium-based Decision Making.....	114
6.3.10	Procedures for MGSO.....	116
6.4	EED Studies on IEEE 30-Bus Power System.....	117
6.4.1	Experimental Settings.....	117
6.4.2	Case Study 1.....	118
6.4.3	Case Study 2.....	119
6.4.4	Case Study 3.....	120
6.4.5	Solution Quality Analysis.....	121
6.4.6	Compromise Solutions.....	122
6.5	Computational EEED Studies on IEEE 118-Bus System.....	123
6.5.1	Experimental Settings.....	123
6.5.2	Comparative Results, Analysis and Discussion.....	124
6.6	Summary.....	133
 Chapter VII		
Conclusions and Future Work.....		135
7.1	Conclusion.....	135
7.2	Future Work.....	138
 Appendices.....		140
A.	Data of Two-Area LFC System and Guangdong Power Grid.....	140
B.	Data of IEEE 30-Bus 6-Generator Power System.....	141
C.	Data of IEEE 118-Bus 54-Generator Power System.....	145
 Reference.....		157

Lists of Figures, Tables and Abbreviations

List of Figures

Fig. 1.1	Illustration of overall organizational structure of this thesis	14
Fig. 2.1	Functional diagram of the studied generation dispatch framework	19
Fig. 2.2	Flow chart for the determination of CPS compliance	23
Fig. 2.3	Agent-world interactive structure for reinforcement learning	32
Fig. 3.1	Q-learning based optimized AGC control framework	41
Fig. 3.2	The schematic diagram of CPS control state space	43
Fig. 3.3	Execution steps of the proposed QAGC	47
Fig. 3.4	Framework of semi-supervisory group pre-learning method	48
Fig. 3.5	Semi-supervisory group pre-learning procedures	49
Fig. 3.6	Two-area power system LFC model	50
Fig. 3.7	Pre-learning process of the proposed QAGC in area A	52
Fig. 3.8	Simulation test of the optimal QAGCs	54
Fig. 4.1	The stream of states for backward estimation	60
Fig. 4.2	Backward estimation for $Q(\lambda)$ backtracking learning	61
Fig. 4.3	$Q(\lambda)$ learning algorithm for AGC relaxed control	67
Fig. 4.4	Two-area thermal-dominated LFC model	68
Fig. 4.5	Pre-learning process of the $Q(\lambda)$ AGC in area A	69
Fig. 4.6	Simulation experiment for AGC controllers	72
Fig. 4.7	The interconnected network of China Southern Power Grid	74
Fig. 4.8	Time characteristics of ΔF , ACE and generators output	76
Fig. 4.9	Relaxation factor effects on AGC performance	77
Fig. 5.1	Pseudo code of the $R(\lambda)L$ algorithm for AGC strategy	87

Fig. 5.2	Imitation pre-learning process of $R(\lambda)$ ILAGC	89
Fig. 5.3	Pseudo code of the proposed imitation pre-learning process	91
Fig. 5.4	Relationship between control actions of PI controller and $R(\lambda)$ ILAGC	92
Fig. 5.5	Imitation pre-learning process of $R(\lambda)$ ILAGC	93
Fig. 5.6	Step response plots of AGCs (solid line: $R(\lambda)$ ILAGC; dot line: $Q(\lambda)$ AGC; dash line: PI controller)	94
Fig. 6.1	Flowchart of the proposed algorithm for the MOPD problem	115
Fig. 6.2	Comparison of the PFs obtained for Case 1	118
Fig. 6.3	Comparison of the PFs obtained for Case 2	118
Fig. 6.4	Comparison of the PFs obtained for Case 3	119
Fig. 6.5	Reference PF surface	123
Fig. 6.6	Best PF solutions obtained with NSGA-II	124
Fig. 6.7	Best PF solutions obtained with SPEA2	125
Fig. 6.8	Best PF solutions obtained with MGSO	125
Fig. 6.9	Producer's convergence of each group objective	126
Fig. 6.10	Maximum pursuit distance and ranger percentage of each group in the convergence process	127
Fig. 7.1	Dynamic optimization process for hierarchical AGC framework	137
Fig. B.1	Single-line diagram of IEEE 30-bus test system	140
Fig. C.1	Single-line diagram of IEEE 118-bus test system	144

List of Tables

Table 4.1	Trace-decay factor effects on $Q(\lambda)$ AGC	70
Table 4.2	Discount rate effects on $Q(\lambda)$ AGC	70
Table 4.3	AGC performance test for different time-delay and GRC	73
Table 4.4	Weight factors for the $Q(\lambda)$ AGCs	75
Table 4.5	Statistical experiment results of Guangdong power grid I	80
Table 4.6	Statistical experiment results of Guangdong power grid II	80
Table 5.1	Statistical experiment results of Guangdong power grid I	96
Table 5.2	Statistical experiment results of Guangdong power grid II	96
Table 6.1	Parameter settings of MGSO for EED of IEEE 30-bus system	116
Table 6.2	Comparison of best fuel cost for Case 1	117
Table 6.3	Comparison of best emission for Case 1	117
Table 6.4	Comparison of best fuel cost for Case 2	118
Table 6.5	Comparison of best emission for Case 2	119
Table 6.6	Comparison of best fuel cost for Case 3	120
Table 6.7	Comparison of best emission for Case 3	120
Table 6.8	Normalized distance measure of different algorithms	121
Table 6.9	Compromise solutions of MGSO for different decision making	121
Table 6.10	Parameter settings of MGSO for EEED of IEEE 118-bus system	122
Table 6.11	Comparison of best solutions for cost, emission and system loss	127
Table 6.12	Performance measures of the best run of each algorithm	129
Table 6.13	Resulting statistics of convergence metrics in 50 runs	129
Table 6.14	Resulting statistics of spacing metrics in 50 runs	130

Table 6.15	Resulting statistics of normalized distance measures in 50 runs	130
Table 6.16	Run time per generation of different algorithms	131
Table 6.17	Reinforcement factor effects on performance of MGSO	132
Table A.1	System parameters for two-area LFC system model	139
Table A.2	Model parameters of AGC units in Guangdong power grid	139
Table B.1	Generator cost coefficients of IEEE 30-bus system	140
Table B.2	Generator emission coefficients of IEEE 30-bus system	141
Table B.3	Generation data of IEEE 30-bus system	141
Table B.4	Transmission line data of IEEE 30-bus system	141
Table B.5	Bus and demand data of IEEE 30-bus system	142
Table C.1	Generator cost coefficients of IEEE 118-bus system	145
Table C.2	Generator emission coefficients of IEEE 118-bus system	146
Table C.3	Generation data of IEEE 118-bus system	147
Table C.4	Transmission line data of IEEE 118-bus system	149
Table C.5	Bus and demand data of IEEE 118-bus system	153

List of Abbreviations

ACE	Area Control Error
AGC	Automatic Generation Control
AHC	Adaptive Heuristic Critic
ANN	Artificial Neural Networks
AROC	Average Reward Optimality Criterion
BES	Battery Energy Storage
CF	Compliance Factor
CNIC	Constant Net Interchange Control
CPC	Control Performance Criteria
CPS	Control Performance Standards
CSG	China Southern Power Grid
CTMDPs	Continuous Time Markov Decision Processes
DCS	Disturbance Control Standards
DGs	Distributed Generations
DM	Decision Maker
DMS	Distribution Management System
DROC	Discounted Reward Optimality Criterion
DTMDPs	Discrete-Time Markov Decision Processes
EA	Evolutionary Algorithm
ED	Economic Dispatch
EED	Environmental/Economic Dispatch
EEED	Economic/Environmental/Energy-saving Power Dispatch
EMS	Energy Management System
FCPSO	Fuzzy Clustering Based Multiobjective PSO
FFC	Flat Frequency Control
GA	Genetic Algorithm

GRC	Generation Rate Constraint
GSO	Group Search Optimizer
HAM	Hierarchical Abstract Machines
HVDC	High Voltage Direct Current
SPEA2	Improved Strength Pareto Evolutionary Algorithm
LFC	Load Frequency Control
LNG	Liquefied Natural Gas
LP	Linear Programming
LQR	Linear Quadratic Regulator
MAS	Multi-Agent System
MCDM	Multiple Criteria Decision Making
MDPs	Markov Decision Processes
MGSO	Multiple Group Search Optimizer
MOEAs	Multiobjective Evolutionary Algorithms
MOPD	Multiobjective Power Dispatch
MOPSO	Multiobjective Particle Swarm Optimization
MOSST	Multi-objective Stochastic Search Technique
NARI	Nanjing Automation Research Institute
NERC	North American Electric Reliability Council
NLP	Nonlinear Programming
NPGA	Niched Pareto Genetic Algorithm
NSGA	Nondominated Sorting Genetic Algorithm
NSGA-II	Nondominated Sorting Genetic Algorithm-II
OPF	Optimal Power Flow
PFs	Pareto-optimal Fronts
PI	Proportional-Integral
POZs	Prohibited Operating Zones
PSO	Particle Swarm Optimization
PV	Photovoltaics

QAGC	Q-learning Based AGC Controller
$Q(\lambda)$ AGC	$Q(\lambda)$ Learning Based AGC Controller
RL	Reinforcement Learning
RMS	Root-Mean-Squares
$R(\lambda)$ IL	$R(\lambda)$ Imitation Learning
$R(\lambda)$ ILAGC	$R(\lambda)$ Imitation Learning Based AGC Controller
$R(\lambda)$ L	$R(\lambda)$ Learning
SCADA	Supervisory Control And Data Acquisition
SMDPs	Semi-Markov Decision Processes
SMES	Super Conducting Magnetic Energy Storage
SPEA	Strength Pareto Evolutionary Algorithm
SQP	Sequential Quadratic Programming
TBC	Tie-line Bias Control
TD	Temporal Difference

Chapter I

Introduction

1.1 Background of Research

As modern power systems have evolved over the centuries and have become large-scale and geographically expansive electrical networks with more and more neighbouring grids and energy resources being interconnected, the development of the operation and management for interconnected power systems requires the successive updating of load frequency control (LFC) for maintaining the power system frequency and optimizing the power flows of transmission grid [1]. The primary objective of the LFC is to maintain the frequency of each control area and to keep tie-line power close to the scheduled power exchanges by real-time regulating the generation outputs of dispatchable generators to accommodate the fluctuating load demands as well as system losses. Under normal operating condition, the power system LFC scheme in energy management system (EMS) can be implemented on the following three control levels according to the control time horizons [2]:

- (1) Primary frequency control: the first LFC level is executed with the turbine governors in synchronized generating units, and it can respond rapidly and automatically to frequency deviations within a timescale of seconds for the fast and stochastic small load disturbances;
- (2) Secondary frequency control: the secondary control is implemented through automatic generation control (AGC), a closed-loop feedback control system in the centralized power system dispatch of each interconnected control area. The output of AGC system, AGC regulating command, is to modulate the

power outputs of AGC plants to re-establish the overall system frequency and tie-line interchanges to their nominal values. Besides, the AGC decision cycle can be set in the range from 4 to 16 seconds [3];

- (3) Tertiary frequency control: economic dispatch (ED), which is performed to allocate the system predictable load amongst all the dispatchable generators throughout the control area so as to minimize the total fuel costs, is always employed to determine the base loading setpoints for the AGC and primary frequency control, and its function is activated every 5 minutes to 1 hour [4].

From the perspective of power dispatch center, AGC and ED play a crucial role in enhancing the long-term control performance and sustainability for the efficient and successful operation of large-scale interconnected power grids [5]. The interconnected power systems consist of a number of control areas, which are mainly responsible for two tasks: (i) supplying power to their native load; (ii) maintaining interchange power with their neighbors to its scheduled value. First of all, there are three fundamental AGC operation modes [6]: (i) flat frequency control (FFC), (ii) constant net interchange control (CNIC), and (iii) tie-line bias control (TBC). Since the TBC operating mode supports the inter-area assistance of the overall system reserves for interconnection frequency, and hence is usually adopted for the most of existing interconnected power grids [7].

Over the years, the AGC performance of interconnected control areas under the TBC mode had been assessed and monitored by control performance criteria (CPC), A1 and A2 criteria, since the late 1960s [3]. Both the criteria are designed based on the area control error (ACE) which is a variable signal calculated by the linear weighted combination of the deviations of frequency and tie-line power exchanges. In the old CPC, A1 criterion requires a control area's ACE to cross zero at least once every ten minutes, and A2 criterion requires that the average of ACE over each six ten-minute clock interval is limited within a specific threshold. Since A1 and A2 were based on engineering judgment and had little analytical basis, previous onsite operating experiences indicated that the old criteria had

often led to unnecessary and frequent reversal adjustments in the AGC regulating commands, and this is harmful to the AGC plants due to the maneuvering wear-and-tear and operational inefficiency. Also, forcing ACE to return to zero may increase the system frequency error if the ACE and frequency have contradictory control directions. Therefore, the CPC criteria cannot satisfy the needs of the current power industry environment and have received lots of criticisms [8].

Consequently, the North American Electric Reliability Council (NERC) has replaced A1 and A2 with new control performance standards (CPS) [9], CPS1 and CPS2, which have the solid technical foundations but less straight forward to interpret directly, in February 1997. The Standards have since then been adopted by most of power utilities over the world including East China Power Grid and China Southern Power Grid (CSG) in 2001 and 2005, respectively [10, 11].

Meanwhile, with the alarming deterioration in air quality conditions, environmental protection has attracted an ever-increasing public awareness and concern. The Clean Air Act was significantly amended and passed in 1990 to control evaporative emissions from fossil-fueled power plants [12]. Therefore, the emission dispatch option has been incorporated in the operational and control strategies of power systems in order to reduce the environmental pollution emissions, such as sulphur oxides SO_x and nitrogen oxides NO_x , from the thermal power plants, and the traditional ED problem is then transformed to the environmental/economic dispatch (EED) which is a more attractive and sustainable alternative for active power generation dispatch and optimization, as it considers both the economic and atmospheric emission simultaneously [13, 14].

In recent years, the EED problem has received much attention from both the academia and industry, and several techniques to reduce emission pollutants in power dispatch problem have been reported and investigated in literatures [12-17]. In general, these techniques can be classified into the following categories:

- (1) Investment incentives should be prioritized for technological innovations to the development of clean and green energy and fuels. Firstly, the alternative

low-emission generation mix, such as wind farm, solar photovoltaics (PV), hydro and nuclear power, can be grid-connected to reduce the use of traditional fossil fuels [18]. Furthermore, cleaning equipment can be installed in various thermal power plants to capture the emission pollutants;

- (2) Compared to the techniques in above, developing the environmental dispatch strategies is more important and preferable to the power dispatch center due to its less capital investment and higher immediacy. The original EED only requires minor modification to the basic ED with an explicit environmental constraint to limit the emission gases within a permissible threshold [19-22]. However, there is a major difficulty in this approach to further study the trade-off relationship between the two dispatch objectives, emission and economy [16, 23];
- (3) Furthermore, different techniques have been proposed to combine both the economic and emission objectives using linear weighted aggregate method [24-28], and then the biobjective EED problem can be transformed into a single-objective optimization problem. Nevertheless, this approach requires many trials to obtain a desired noninferior set of solutions by varying the weights in the objective function, and it is also not effective to cope with the problems including nonconvex Pareto-optimal fronts (PFs) [29, 30];
- (4) In order to involve the process for simultaneously optimizing the fuel cost and emission objective functions, the Pareto optimality based multiobjective optimization theory [31], in which the EED objectives are considered as the conflicting and incommensurable objectives, is usually used for the pursuit of a group of Pareto-optimal solutions to express the trade-offs of multiple objectives. Consequently, based on the Pareto dominance principle, various multiobjective optimization algorithms have been proposed to handle this EED problem [13, 32-34], and a series of nondominated solutions can be generated by these computational algorithms with different search strategies in a single simulation run. Thus, a set of manageable and representative PF

solutions can be intuitively perceived and provided to the power system dispatchers to extract a suitable solution as the final dispatch decision.

1.2 Incentives of Thesis

The ongoing and deepening energy shortage over the world has significantly hard hit the global economy and society to the greatest extent [35]. With the surge in the oil prices and the shrinking stock of other traditional fuel resources, significant efforts have been made to promote and develop alternative renewable energy technologies for the sustainable electricity supply. The energy shortage concern propels the power dispatch centers to reschedule their strategies, and brings in the heatedly intensive research into the energy-saving generation dispatch. Nevertheless, the major challenge to incorporate energy conservation in grid dispatch is the highly constrained and high-dimensional optimization space with the increased problem complexity due to the multiple objectives over a great variety of dispatchable generators, and these problems are quite difficult to solve using conventional methods. Therefore, it is the aim of this thesis to develop a new generation of LFC control and optimization schemes designed for the CPS standards, optimum economic-emission and energy-saving operation over large-scale interconnected power systems.

The NERC's new CPS standards pay more attention to the medium and long-run returns of AGC control performance [36]. Furthermore, previous works have concluded that the time averages of the ACE, and so are the CPS1 and CPS2, for interconnected control areas are essentially random and unpredictable with the averaging interval beyond a specified measurement period [37]. Coupled with the renewed interest in the energy-saving dispatch and the statistical concepts for CPS1/CPS2 metrics as well as the rapid development of smart grid, there are fundamental changes in the conventional AGC control philosophies. The existing AGC systems mostly adopt the conventional proportional-integral (PI) control

structure with fixed parameters and do not adapt well to large changes in operating mode, parameters and structure of power grids [38, 39]. In particular, after the separation of the generation sectors from electricity grids in the deregulation of the Chinese power systems in 2002 [40], genuine AGC strategies in the provincial power grids would only give priority to promoting CPS compliances for the operation benefits of dispatch centers. As a result, it may lead to an “over-compliant” problem which high CPS compliance is blindly pursued without considering the maneuvering costs of plants. It has been observed from the dispatch centers of CSG that there are many tiny frequent reversal actions in AGC commands for some plants, and this is ineffective for enhancing the CPS compliances and harmful to the thermal generators because of wear-and-tear [41]. Moreover, the increments in AGC regulating intensity for thermal units would always result in the evident rise in the fuel consumption as well as the decrease in generation benefits of interconnected operation. Therefore, in addition to CPS metrics, the regulating cost of AGC plants is also a considerable issue and should be incorporated in the AGC control objective for the beneficial coordination between the electricity grids and generation plant.

The AGC system under CPS standards is a key technology and core content for the energy saving generation dispatch. So far, researches reported are mostly based on classical linear or nonlinear LFC models as recommended by the IEEE [42]. The performance of the resulted AGC schemes would largely affected by the accuracy of the system parameters as well as the load demand fluctuation characteristics. Further study on its dynamics and CPS statistical characteristics indicated that AGC process under CPS in fact is a stochastic multistage decision problem [43] and can be modeled as the Markov chain control problem which forms the base for the proposed research on AGC using reinforcement learning (RL) [44]. Moreover, the goal of RL models is to find an optimal control policy by maximizing the expected cumulative reward value in the long-run, and this coincides with the purpose of the CPS standards. Consequently, in the thesis, the

RL algorithms [45] based on Markov Decision Processes (MDPs) [46, 47] are introduced for the design of the optimal AGC controllers under CPS.

Thermal power plants have played a dominant role in most AGC systems, the CSG power system being an example. One of the challenges in thermal-dominated AGC strategies is the long time-delay problem caused by the steam turbines of thermal generators in secondary frequency control loop [48]. Field tests showed that typical time-delay ranges from 0.5 to 2 minutes and would vary with the ramping direction. Coupled with the fast random fluctuations in load demand, the problem of AGC optimization becomes even more complicated and so far no satisfactory solution method has been reported publicly. Moreover, it is an unavoidable fact that differences always exist between the system simulation model and real power system, and most of modern control algorithms suffer a severe drawback that the results optimized in the simulation environment cannot be practically implemented for real power system application. Hence, enhancing the feasibility and applicability of the proposed AGC controllers for the onsite operation has become a pressing need.

On the other hand, for solving the Pareto-based biobjective EED problem, it has to be admitted that currently numerous algorithms can be available to form the nondominated trade-off surface representing Pareto optima, [13, 49-63]. Most notably, various multiobjective evolutionary algorithms (MOEAs) [32, 64] and swarm intelligence techniques [65] have been introduced for the multiobjective power dispatch (MOPD) optimization to generate a set of comprehensive Pareto-optimal solutions. However, those previous multiobjective algorithms have been designed and implemented for a small IEEE 30-bus 6-generator power system, and limited attention has been paid to the MOPD problems for large-scale power systems with more objectives. Following the No Free Lunch theorem, “for any algorithm, any elevated performance over one class of problems is exactly paid for in performance over another class” [66]. Actually, for the nowadays large-scale and topologically complicated power grids, the multiobjective algorithms

should be designed specifically and applicable for the real-world power dispatch requirements. Consequently, coupled with the energy-saving dispatch objective, there is still significant room to explore novel Pareto optimality-based MOPD approaches.

Moreover, the ultimate goal of the multiobjective optimization is to identify a unique solution expressing the best compromise among multiple competing and conflicting objectives [67]. The compromise solution should be extracted from the obtained PF as the decision maker (DM)'s preference with utilities' interest. Previous decision-making methods have made use of the fuzzy reasoning for this bargaining solution [68, 69]. Although such a way of fuzzy-based decision-making method is efficient and readily to be implemented, the definition of fuzzy membership functions is based on the designer's experience without considering the PF's trade-off characteristics. Consequently, a multiple criteria decision making (MCDM) scheme should be integrated in the multiobjective formulation to select a more reasonable solution for the actual operating task.

1.3 Primary Contributions

The main motivation of this research is to develop new methodologies for regional power interconnections and system dispatchers to handle the new issues resulted from the CPS standard and energy-saving generation dispatch. More specifically speaking, the original contributions of this thesis can be summarized into the following aspects:

First and foremost, the NERC's CPS, which has been widely adopted around the world including power grids in the Mainland China, can provide a theoretical basis for relaxing the regulation constraints of AGC plants and enhancing the frequency support effect from neighbouring control areas [9]. Consequently, in order to take full advantage of the CPS standards, this thesis addresses the main design objectives for a synthetic optimum AGC strategy as follows: (i) to comply

with CPS metrics for interconnected control areas in various complex operation scenarios of power systems, (ii) to achieve the relaxation of AGC system which can release the regulating pressure and conserve control cost of AGC plants at the most, (iii) to develop a feasible and reasonable control structure having a high level of applicability and practicability for real power systems. Based on the design objectives above, a new concept is firstly proposed in this thesis referred as “the optimal relaxed AGC control” to allow the AGC plants to maneuver less costly in finding the optimal economic policy on the premise of complying with the assessment standards CPS1 and CPS2. To be more specific, this optimal economic policy is to pursue less control actions, less wear-and-tear, and further savings on fuel costs for the generation sectors.

Secondly, the methodology developed employs the well-known Q-learning algorithm [70] based on Discrete-time Markov Decision Process (DTMDP) [71] to design the optimal relaxed AGC strategy, with the CPS metrics regarded as the rewards from interconnected power systems. In addition, by means of the space discretization technique, the CPS state space and AGC regulating commands can be divided into a finite number of state-action pairs after in-depth investigation on the relaxed AGC control and unit regulating characteristics. Subsequently, in order to enhance stability and convergence capability of the proposed Q-learning based AGC controller (QAGC), a semi-supervisory group pre-learning method is presented using a PI/RL structure to deal with the pre-learning problem involving multiple QAGCs in multi-area interconnected power systems. With a decreasing rule for the gains of the supplementary PI controller, the optimal AGC strategy can be obtained gradually to maximize the long-term cumulative rewards in the procedure of interactive self-learning mechanism. The corresponding simulations demonstrated that, compared with the conventional PI controller, the QAGC can effectively improve the robustness and adaptability of AGC systems while the CPS metrics are ensured under the complex changing operation conditions.

Thirdly, for the thermal-dominated power systems, the AGC strategies arisen from the power dispatch centers will be implemented by the fuel supply in the turbine-boiler control system of thermal plants, and will therefore be performed slowly with large time constants. Consequently, further in-depth researches and analysis on the after-effect behaviors of the thermal-dominated LFC indicated that, the AGC decision-making problem is in fact a non-Markov chain control process, and its non-Markov property is difficult to solve using the standard Q-learning [72]. Hence, this thesis firstly applies the multi-step $Q(\lambda)$ learning with eligibility traces [73] to effectively tackle the long time-delayed feedback control problem of thermal AGC generators in non-Markov environment. Furthermore, in the $Q(\lambda)$ learning based AGC controller ($Q(\lambda)$ AGC), the moving averages of CPS1 and ACE are adopted as the state feedback input, and the AGC control pulses are integrated in the reward function using linear weighted aggregate method. This novel AGC strategy also can provide a convenient way for controlling the degree of CPS compliance and relaxation of system by online tuning relaxation factors to implement a desirable relaxed control. The statistical experiments on the CSG power system showed that the $Q(\lambda)$ AGC can significantly enhance the robustness and dynamic performance of AGC systems, and reduce the number of pulses and pulse reversals so as to avoid the CPS over-compliant problem.

However, it should be mentioned that the above QAGC and $Q(\lambda)$ AGC need to be scheduled to experience a series of trial-and-error training phase called the pre-learning process before their real application, and a high-accuracy simulation model is required to be established for the offline pre-learning process [74]. If a significant difference exists between the system simulation model and real power system, it may leads to the intolerable “trial-and-error noise” for these RL based AGCs after being put into the practical system. In fact, this problem is the main obstacle to the application of the conventional RL algorithms for the practical implementation in power systems. Therefore, this thesis continues the stochastic

optimal relaxed AGC along this direction but with full consideration to overcome the difficulties explained above. Thus, another fundamental contribution of this thesis is to develop a novel imitation pre-learning technique to facilitate the transition from the off-line pre-learning process of the RL controllers to the on-line pre-learning. Besides, a $R(\lambda)$ learning ($R(\lambda)L$) [75], studied on the basis of average reward optimality criterion (AROC) [76], is to seek for the maximization of long-term average rewards of generic systems. This truly provides the opportunity for this research work to design the optimized AGC scheme for the improvement of the CPS performance assessed over a long period of time. Thereby, the proposed $R(\lambda)$ imitation learning based AGC controller ($R(\lambda)ILAGC$) can combine the merits of the average reward RL model and imitation pre-learning process. Case studies illustrated that the $R(\lambda)ILAGC$ not only can act online in real time as an observer to learn from the control behaviors of the existing AGC controller such that the feasibility and applicability of the proposed AGCs are enhanced greatly, but also can improve the convergence efficiency and control performance of AGC system.

Last but not least, the final main contribution of this thesis is to propose a novel multiple group search optimizer (MGSO) algorithm to solve the large-scale highly constrained MOPD optimization problem with several incommensurable and contradictory objectives. Although the traditional EED problem can supply an optimum steady-state operating point for EMS and supervisory control and data acquisition (SCADA) system [15, 62], the economic/environmental/energy-saving power dispatch (EEED) for large-scale power systems does not appear to have been reported by any researchers. Therefore, this tri-objective EEED issue is specifically addressed in this thesis with the help of a newly emerged group search optimizer (GSO) [77] which is capable of solving large-scale multimodal complex optimization problems. In this thesis, the GSO is utilized and extended by combining it with RL mechanisms to form the MGSO algorithm which allows the multi-group cooperative search to establish a Pareto tradeoff for the

multiobjective optimality. In the developed algorithm, a better constraint handling strategy, which separates the constraints and objectives with different search strategies, and two enhancements, namely space reduction and chaotic sequence dispersion, have been incorporated to further enhance the convergence performance of the proposed algorithm. Moreover, this thesis also introduces the Nash equilibrium [78] for designing a more reasonable decision making method in order to identify the best compromise solution from the obtained optimal front. In comparison with other earlier multiobjective heuristic algorithms on various performance measures, the proposed algorithm possesses the greatly enhanced capacity in solving the large-scaled MOPD problems, and hence has a large potential to accommodate the highly complex problem characteristics.

1.4 Thesis Layout

The remainder of this thesis consists of six chapters and two appendices, and is organized as follows:

Chapter II discusses some essential details regarding generation dispatch and control in modern power system dispatch centers, which serve as the foundations for the research work addressed in this thesis. The relationship between the two issues studied, AGC and MOPD, is explained, and the multi-criteria optimization and control objectives for large-scale power system dispatch are investigated and designed. Furthermore, a brief literature survey on the development progress of present AGC and MOPD schemes are reviewed, and the RL and multiobjective algorithms used throughout the thesis is also introduced in this chapter.

Chapter III presents the application and design of a stochastic optimal control methodology using the Q-learning algorithm for AGC controller under NERC's CPS standards. The two-area interconnected LFC model and CSG power system model used for simulation studies are investigated and established in this chapter. In addition, a novel semi-supervisory group pre-learning approach is proposed

for the QAGC to improve its stability and convergence capability in the pre-learning process, and the corresponding simulation results are also presented.

Chapter IV focuses on applying the multi-step $Q(\lambda)$ learning to specifically design the stochastic optimal relaxed AGC strategy for thermal-dominated power systems. The long time-delay secondary frequency control loop for AGC thermal plants in non-Markov environment is thoroughly analyzed and discussed, and the eligibility traces as well as its backward estimation are introduced to effectively alleviate the non-Markovian effects in the thermal-dominated AGC process. In this chapter, a multi-criteria relaxed reward function is proposed, and studies on the learning parameters and relaxation factors for the $Q(\lambda)$ AGC are carried out in simulations for comparison and discussion.

Chapter V continues to exploit the potential of RL theory for updating the AGC to improve its practical onsite operation and long-run control performance. A novel imitation pre-learning method is proposed and devised to overcome the major obstacle for the application of RL algorithms to the onsite implementation in real power systems. Meanwhile, the AROC based $R(\lambda)$ L method is harnessed to develop an optimal $R(\lambda)$ imitation learning ($R(\lambda)$ IL) methodology for the AGC under CPS. Also, the in-depth comparisons of the $R(\lambda)$ ILAGC with the previous QAGC and $Q(\lambda)$ AGC are investigated via statistical comparative experiments.

Chapter VI proposes a new equilibrium-inspired multiobjective optimization algorithm, MGSO, to solve the highly constrained EEED problem including the fuel cost, emission and energy-saving objectives for large-scale power systems. The proposed approach integrates various search techniques, including the GSO, Nash equilibrium, synergistic learning, Boltzmann distribution based constraint handling, space reduction and chaotic sequence dispersion, into a unique MGSO algorithm. As demonstrated in this chapter, the performance of MGSO has been fully evaluated and benchmarked using different Pareto metrics on the IEEE 30-bus 6-generator system and the IEEE 118-bus 54-generator system.

Finally, the concluding remarks of the thesis are summarized in Chapter VII, and some prospective extensions and possible directions for future research work are also presented.

Furthermore, the overall organization of this thesis is illustrated in Fig. 1.1.

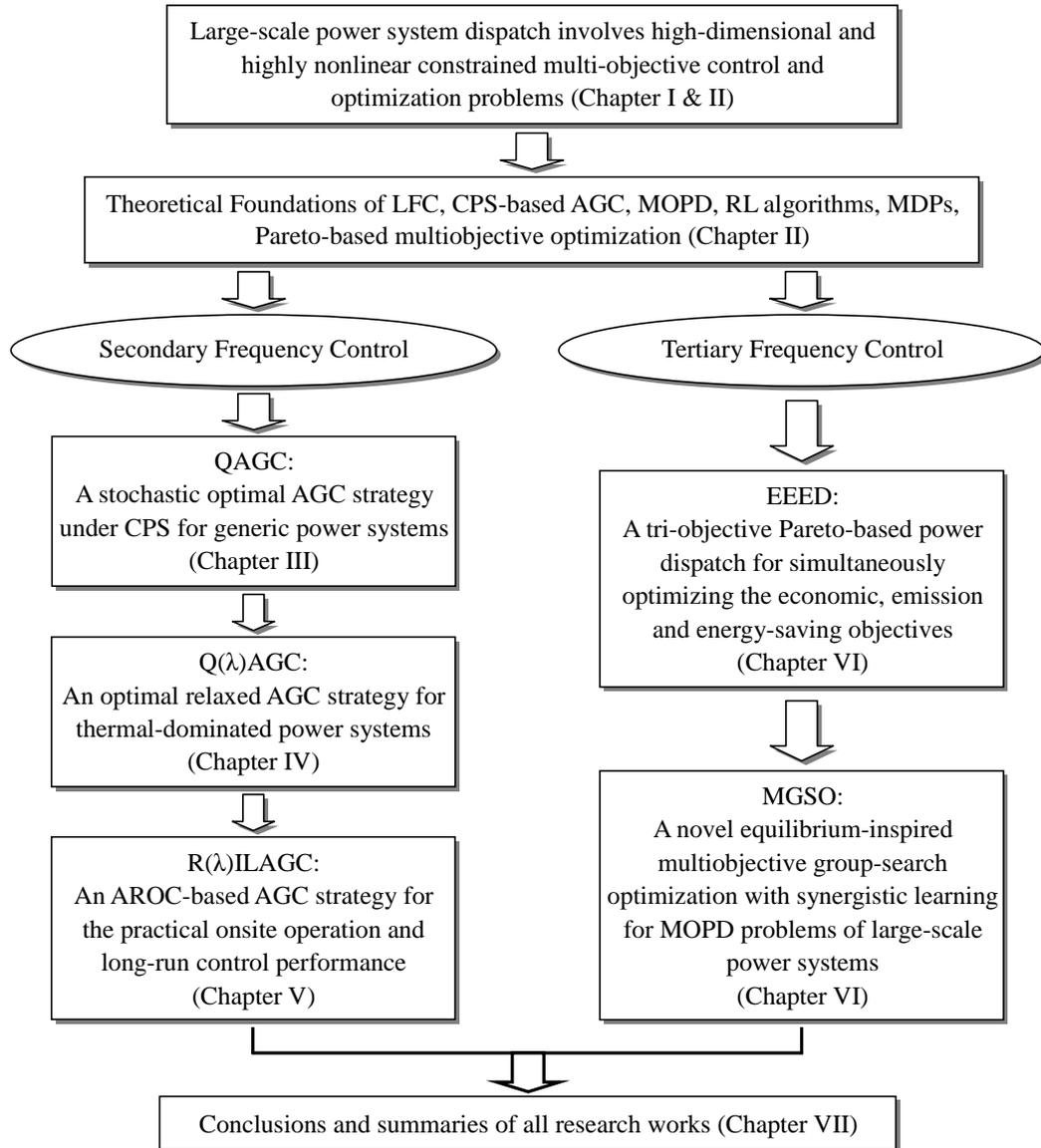


Fig. 1.1 Illustration of overall organizational structure of this thesis

1.5 List of Publications

Journal paper published:

1. B. Zhou, K. W. Chan, T. Yu, and C. Y. Chung, "Equilibrium-inspired multiple group search optimization with synergistic learning for multi-objective electric power dispatch," *IEEE Transactions on Power Systems*, Manuscript ID: TPWRS-00121-2012.R2, Accepted for publication on 15 Apr 2013.
2. T. Yu, B. Zhou, K. W. Chan, Y. Yuan, B. Yang, and Q. H. Wu, " $R(\lambda)$ imitation learning for automatic generation control of interconnected power grids," *Automatica*, vol. 48, no. 9, pp. 2130-2136, Sep. 2012.
3. T. Yu, B. Zhou, K. W. Chan, and E. Lu, "Stochastic optimal CPS relaxed control methodology for interconnected power systems using Q-learning method," *Journal of Energy Engineering*, vol. 137, no. 3, pp. 116-129, Sep. 2011.
4. T. Yu, B. Zhou, K. W. Chan, L. Chen, and B. Yang, "Stochastic optimal relaxed automatic generation control in Non-Markov environment based on multi-step $Q(\lambda)$ learning," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1272-1282, Aug. 2011.
5. T. Yu, Y. M. Wang, W. J. Ye, B. Zhou, and K. W. Chan, "Stochastic optimal generation command dispatch based on improved hierarchical reinforcement learning approach," *IET Generation, Transmission & Distribution*, vol. 5, no. 8, pp. 789-797, Aug. 2011.
6. Y. Tian, B. Zhou, Y. T. Zhang, and K. W. Chan, "Investigation on the use of GPGPU for fast sparse matrix factorization," *Journal of International Council on Electrical Engineering*, vol. 1, no. 1, pp. 116-122, Jan. 2011.
7. T. Yu, K. W. Chan, J. P. Tong, B. Zhou, and D. H. Li, "Coordinated robust nonlinear boiler-turbine-generator control systems via approximate dynamic

feedback linearization,” *Journal of Process Control*, vol. 20, no. 4, pp. 365-374, Apr. 2010.

8. T. Yu, H. H. Liang, and B. Zhou, “Smart power generation control for Microgrids islanded operation based on $R(\lambda)$ learning,” *Power System Protection and Control*, vol. 40, no. 13, pp. 7-13, Jul. 2012. (in Chinese)
9. T. Yu, B. Zhou, and K. W. Chan, “Stochastic optimal CPS control methodology for interconnected power systems using multi-step backtrack $Q(\lambda)$ learning,” *Transactions of China Electrotechnical Society*, vol. 26, no. 6, pp. 179-186, Jun. 2011. (in Chinese)

Journal paper under review or preparation:

10. B. Zhou, K. W. Chan, T. Yu, and J. Liu, “Strength Pareto multi-group search optimizer for multiobjective optimal VAR dispatch,” in preparation for submission to *IEEE Transactions on Smart Grid*.
11. B. Zhou, T. Yu, K. W. Chan, J. Liu, and J. M. Li, “Hierarchical multi-group search optimizer for distributed Pareto-optimal power flow of multi-area interconnected power systems,” in preparation for submission to *IEEE Transactions on Power Systems*.

Conference paper presented:

12. B. Zhou, K. W. Chan, and T. Yu, “Q-learning approach for hierarchical AGC scheme of interconnected power grids,” *The Proceedings of International Conference on Smart Grid and Clean Energy Technologies, ICSGCE 2011*, 27-30 Sep. 2011, UESTC Hotel, Chengdu, China.
13. Y. T. Zhang, B. Zhou, and K. W. Chan, “Investigation of impact of wind power fluctuations on AGC performance,” *International Conference on Electrical Engineering, ICEE 2010*, 11-14 July 2010, Paradise Hotel, Busan, Korea.

14. B. Zhou, K. W. Chan, and T. Yu, "Application of eigen-analysis method in auxiliary decision for black start scheme," *International Conference on Advances in Power System Control, Operation and Management, APSCOM 2009, IET, 8-11 Nov. 2009, Kowloon Shangri-la Hotel, Hong Kong.*
15. Y. H. Wang, K. W. Chan, D. S. Ying, B. Zhou, and S. W. Mei, "Optimal secondary voltage preventive control based on particle swarm optimization," *International Conference on Advances in Power System Control, Operation and Management, APSCOM 2009, IET, 8-11 Nov. 2009, Kowloon Shangri-la Hotel, Hong Kong.*

Chapter II

Essentials for Multiobjective Generation Control and Optimization

2.1 Overview of Power Generation Dispatch and Control

In modern power system LFC schemes, the generation dispatch actions and control pulses for each interconnected control area are always determined and maintained by a central grid facility, called power dispatch center [1]. Usually, the control area is an electric power utility for an individual service area, taking provincial power grids in the CSG power system as an example. In general, the obligation of the power dispatch center is to monitor system load, generation of plants and tie-line power interchanges to ensure the balance of electricity demand and supply with secondary and tertiary generation control.

Fig. 2.1 illustrates the functional diagram of the studied modern generation dispatch schemes in a specified control area including the primary control of the governors in generating plants, AGC for secondary control, and MOPD for the tertiary generation optimization. With SCADA/EMS & distribution management system (DMS), the day-ahead generation scheduling [2] is executed day-by-day on the basis of daily load forecasting for all types of generators throughout the system, including non-dispatchable units (nuclear power and PV generations), partially dispatchable units (some fossil-fuel-fired thermal plants and wind farms [79]), dispatchable units (pumped storage hydro power, hydroelectric plants, liquefied natural gas (LNG) plants, and some dispatchable thermal plants). On the other hand, the real-time information, pertaining to tie-line flow, network topologies, system frequency, load flow data, and so on, is telemetered to power

dispatchers for further processing and analysis by means of state estimation [1]. Meanwhile, the dispatch center, which is equipped with a digital integrated computer system, also delivers control signals to the dispatchable generators for MOPD (reference loading set-points) and AGC (raise/lower pulses with changing lengths) through the telemetering channels [80].

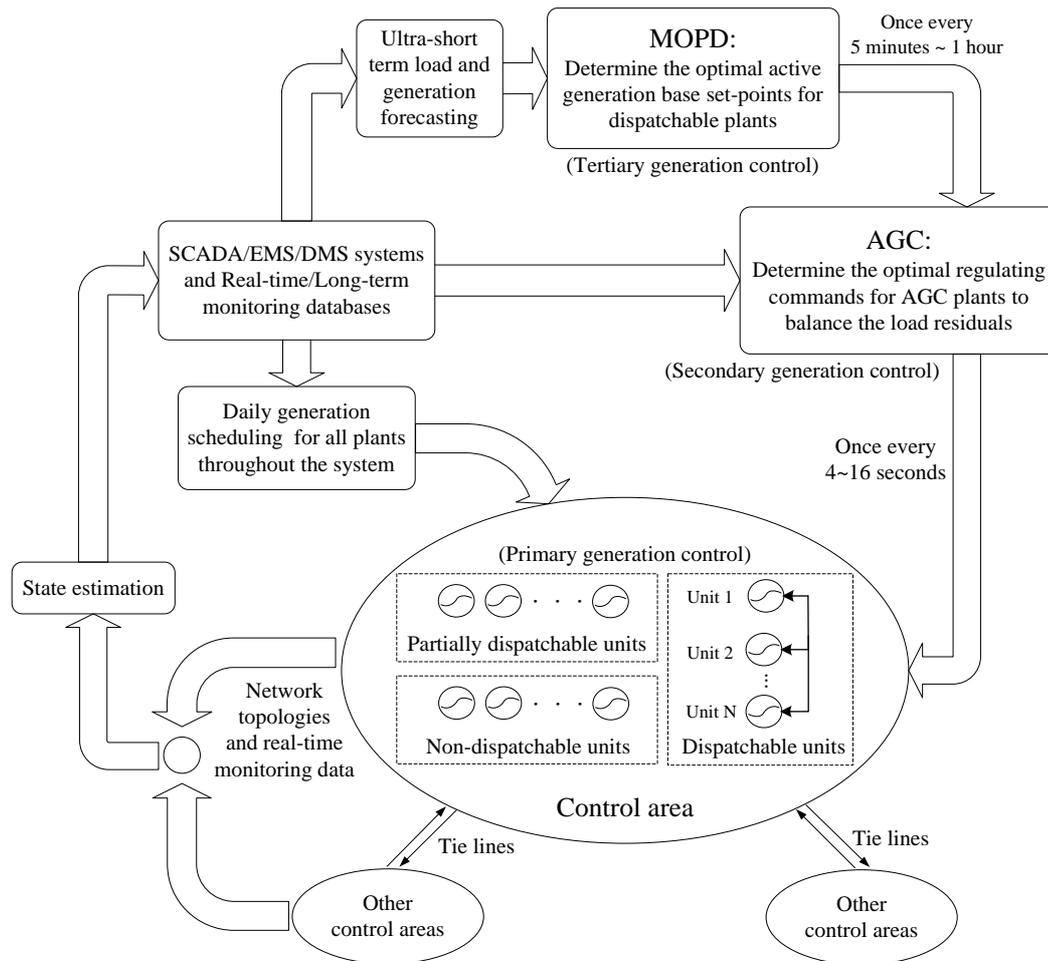


Fig. 2.1 Functional diagram of the studied generation dispatch framework

For the dispatch center of each control area in an interconnected power grid, the two essential dispatch strategies, MOPD and AGC, are usually carried out in area-wise decentralized manner, and these two issues will be intensively studied and designed in this thesis. Nevertheless, it should be pointed out that there are both connections and differences between MOPD and AGC problems. For each decision cycle, the amount of load demands, which is predicted from ultra-short

term load forecasting [5], should be optimally allocated from the central to each individual generating unit for its base MW loading. This allocation optimization problem is solved by the MOPD program to determine the optimized regulation participation factor for each dispatchable unit. Also, the optimal AGC regulating command is calculated for AGC generators to update their generation outputs in order to accommodate fluctuating load disturbances. Furthermore, the total AGC regulating generation is also assigned to each AGC unit according to its MOPD participation factor. Therefore, MOPD provides the base load set-points for AGC, and the AGC command can be regarded as correction component in combination with MOPD base set-points and regulation participation factors to determine the desired generation outputs for individual AGC generators [4].

Theoretically speaking, since MOPD and AGC have different time horizons and optimization objectives, the two problems can be investigated independently, designed to pursue their own objectives and produce their own control outputs. The MOPD is performed to optimally distribute the system load amongst all dispatchable generating sources with all the objectives minimized simultaneously; whereas, the AGC function is to modulate the power outputs of AGC plants so as to balance the load residuals resulted from load forecasting errors or of plant outputs failed to follow their prescribed trajectories and ensure the CPS compliances. In this thesis, a triple objective EEED is designed for the tertiary LFC optimization to accommodate economic, emission and energy-saving objectives. Furthermore, in addition to CPS compliances, the generation regulating cost shall also be considered in the proposed AGC strategy.

2.2 Current State of the Art

Present schemes of generation control and dispatch have been developed over half a century, and AGC assessment criteria A1 and A2 have been replaced by CPS1 and CPS2 standards, while the ED have evolved into a Pareto-based

MOPD optimization problem. In the following, a literature review on the methodologies and strategies of AGC and MOPD are presented in details.

2.2.1 Control Performance Standards

NERC's new CPS standards [81] consist of CPS1 and CPS2. Both metrics are established based on the area control error to assess the energy balance for an interconnected control area, and the ACE is expressed as follows:

$$ACE = \Delta P_T - 10B\Delta F = (P_a - P_s) - 10B(F_a - F_s) \quad (2.1)$$

where ΔF is the frequency deviation which is the instantaneous difference of the actual and base system frequency in this control area, $F_a - F_s$; ΔP_T is the tie-line power error computed from the instantaneous difference between the net actual and scheduled interchanges, $P_a - P_s$; B represents a control area's frequency bias expressed with a unit of $-MW/0.1Hz$.

CPS1 is a limit on the average of a function combining ACE and ΔF from schedules to control the frequency of interconnection, while CPS2 is to restrict the unacceptable and unpredictable tie-line power flow. The former performance standard, CPS1, is formulated based upon statistical theory. Firstly, an expression, called Compliance Factor (CF) [8], is identified which represents, quantitatively, a control area's contribution to the reliability operation of the interconnected grid to which it belongs. The CF requires a control area i to satisfy the constraint in a certain assessment period (such as one minute), as follows:

$$CF = \frac{\sum (E_{AVE-\min} \cdot \Delta F_{AVE})}{(-10B_i) \cdot n} \leq \varepsilon_1^2 \quad (2.2)$$

where $E_{AVE-\min}$ and ΔF_{AVE} are the clock-minute averages of ACE and frequency deviation, respectively; B_i represents the frequency bias of the i th control area, ε_1 denotes a bound for CPS1 control target of the interconnection, n represents the number of minutes in the assessment period. Consequently, the percentage CPS1 compliance of this assessment period can then be calculated as follows:

$$CPS1 = (2 - CF/\varepsilon_1^2) \times 100\% \quad (2.3)$$

The second performance standard, CPS2, limits the magnitude of short-term ACE. It requires the 10-min averages of a control area's ACE less than a constant (L_{10}) given as follows:

$$|\sum E_{AVE-min}|/10 \leq L_{10} \quad (2.4)$$

$$L_{10} = 1.65 \cdot \varepsilon_{10} \cdot \sqrt{(-10B_i) \cdot (-10B_s)} \quad (2.5)$$

where B_s is the summation of the frequency bias settings of all control areas, and ε_{10} is the targeted root-mean-squares (RMS) of clock-10-min average frequency deviation. Then, the percentage of CPS2 compliance can be obtained as follows:

$$CPS2 = \left[1 - \frac{\text{violations}_{\text{month}}}{\text{total periods} - \text{unavailable periods}} \right] \times 100\% \quad (2.6)$$

where $\text{violations}_{\text{month}}$ is a count of the number of periods that the clock-10-min averages of ACE are greater than L_{10} within one month. The AGC performance of each control area should be evaluated and reported with compliances CPS1 and CPS2 at the end of each month or year. In order to comply with the NERC's CPS, the control area must ensure its CPS1 compliance no less than 100% while CPS2 compliance should be greater than 90%. Detailed statistical and dynamic analysis as well as field tests of NERC's CPS were reported in [9, 37, 81-87].

In Chinese electricity industry, a new metric, CPS compliance, is introduced based on the CPS1/CPS2 metrics to measure the overall AGC performance under CPS [10, 11]. According to Grid Code of the China Southern Power Grid [41], the logical flow chart for determining the CPS compliance is shown in Fig. 2.2, and the determination of the CPS compliance for a specified assessment period (typically 10 minutes) is interpreted as follows:

- If $CPS1 \geq 200\%$, then there is no need to consider CPS2 metric and CPS compliance rating is Pass. This is to encourage the load frequency supporting efforts from the other interconnected control areas during emergency and to fully explore the benefits of interconnection. Therefore, AGC behaviors of

the control area in this case are considered advantageous to the improvement of overall interconnection frequency quality over this assessment period.

- If $100\% \leq \text{CPS1} < 200\%$ and CPS2 compliance is satisfied, then the CPS compliance rating is Pass.

If $100\% \leq \text{CPS1} < 200\%$ and CPS2 standard is in violation, then the CPS compliance rating is Fail.

- If $\text{CPS1} \leq 100\%$, then the CPS compliance rating is Fail.

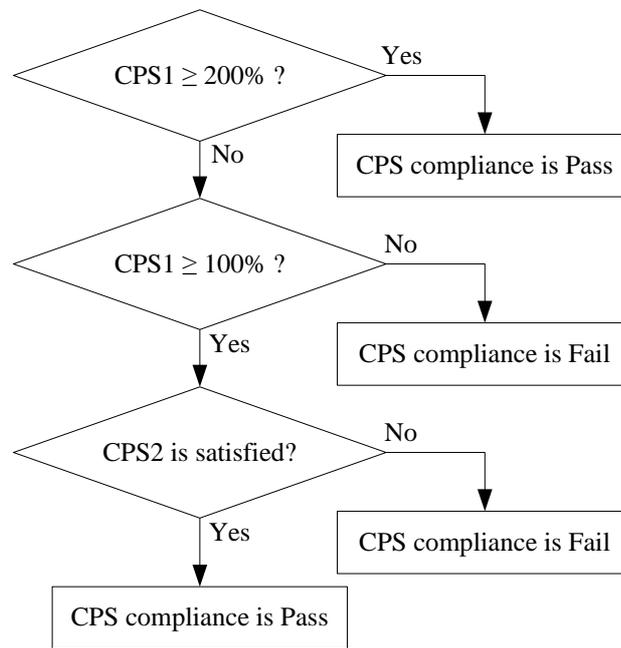


Fig. 2.2 Flow chart for the determination of CPS compliance

Consequently, the percentage metric of CPS compliance on a daily, monthly or yearly basis can be calculated as follows:

$$\text{CPS}(\%) = \left[1 - \frac{\text{violation periods}}{\text{total periods} - \text{unavailable periods}} \right] \times 100\% \quad (2.7)$$

Compared to the old CPC, CPS are technically defensible with solid technical foundations but is difficult to interpret directly. In summary, the main advantages of the CPS are listed as below:

- (1) CPS standards were formulated based on a more reasonable statistical and probabilistic theory with consideration of stochastic behaviors of fluctuating

frequency. Besides, both CPS1 and CPS2 are long-term average evaluation criteria for AGC performance and frequency quality, and thus can reduce the non-predictable maneuverings of generation.

- (2) Another significant superiority of the CPS is the elimination of the control requirement for ACE zero-crossing, and the resultant efficient operation for AGC generators could achieve the fuel efficiency improvements, unit wear-and-tear reductions and less maintenance.
- (3) The CPS can quantitatively assess AGC behaviors with various metrics on a long-term basis, and thus offer control areas with flexibility to design their own control strategies.
- (4) In CPS1 standard, a CF with positive sign means that, during this specific period, the control area acts as a LFC regulating burden to interconnections' operation. On the contrary, a negative CF indicates that the area contributes to the LFC requirement of interconnections. Therefore, the adoption of CPS would promote the mutual energy supports amongst control areas.
- (5) Finally, in CPS2 standard, the limit for clock-10-min averages of ACE, L_{10} , would be remarkably greater than the threshold value of A2 criterion. Then, a larger limit for average ACE allows looser control, lower control costs and less unit adjustments.

2.2.2 State-of-the-Art AGC Schemes

Over the years, extensive research works in the area of AGC strategies have been done. The state-of-the-art in recent AGC control strategies and philosophies, including the theoretical conclusions for the extended research and development, have been comprehensively reported in [3, 6, 88] and further investigated in [38, 39]. Power system models as well as control techniques covered, including linear and nonlinear power system models [89-91], classical and optimal control [92-94], centralized [90] and decentralized control [95-97], two-level [98] and multi-level control [99], have appeared in the literatures. Considerable research on the

control strategies involved includes digital technology [100, 101], adaptive control schemes [102-104], self-tuning methodologies [105-107], robust control [97, 108, 109], discrete control [98, 110], variable structure systems [111-113], and sliding mode control [114]. There also have been lots of AGC studies using intelligent control methodologies such as artificial neural networks (ANN) [115-117], fuzzy logic [97, 118-120], evolutionary algorithm (EA) [121], genetic algorithm (GA) [112, 122, 123], particle swarm optimization (PSO) [124-126], bacteria foraging optimization [127, 128], RL algorithms [43, 129, 130], and hybrid intelligent algorithms [131-134].

Moreover, the basic problem of AGC has also been extended to incorporate ED function [135, 136], optimal power flow (OPF) [137], the restructuring and deregulated environment of power systems [122, 138-142]. Meanwhile, with the rapid developments and advancements on distributed generations (DGs), the new power electronic technologies for wind turbine, photovoltaics, fuel cell as well as battery energy storage (BES) have enabled those DGs to be grid-connected with gradually higher penetration [143]. Thereby, there have been investigations and analyses attempted on LFC system dynamics including BES [144], redox flow batteries [145], photovoltaic systems [146], super conducting magnetic energy storage (SMES) [138, 147-150], and fuel cells [151]. Supervisory AGC schemes for wind farms and wind generation variations have also been illustrated and studied to improve the LFC performance [79, 152, 153].

However, the above AGC schemes and control strategies are designed and tailored for the previous performance assessment criteria A1 and A2 based on the classical concept of certain linear/nonlinear systems. So far, the AGC strategies designed to work under NERC's CPS standards have been reviewed in [38, 39], and statistical analysis as well as dynamic physical characteristics for the CPS were addressed in [81, 83]. An analytical framework was presented in [82] for estimation and formulation of the CPS with the uncertainties and random characteristics of the LFC variables taken account. In addition, a decomposition

model was proposed for the 1-min averages of ACE to quantify its influence of the control area's loadings on CPS1 and CPS2 metrics [85], and the corresponding field implementation of this practical decomposing method were investigated in [87]. Following which the AGC strategy was conducted based on practical tests and field data, with the refinements of ACE decomposition, so as to confirm the regulation burden attributable to varying of loadings [86].

So far, investigations on the AGC strategies under CPS were mostly based on the classical PI control structures [8, 10, 11, 154, 155]. Most notably, an AGC logic based on NERC's CPS and disturbance control standards (DCS) was specifically developed [8] for the practical operation with enhanced efficiency and effectiveness. An integral-type LFC controller manipulated by a set of fuzzy logic rules to comply with CPS metrics has been proposed in [156]. In [157], a decentralized multi-area AGC under CPS is addressed, in which the eigenstructure assignment technique is utilized in each local controller, to incorporate varieties of transactions for the competitive restructured electricity markets. A methodology and its practical implementation for redesigning the optimal LFC strategy considering regulation energy costs, power system security and reliability, and power markets was illustrated in [158]. A novel AGC scheme based on neuro-fuzzy inference system and adaptive gain scheduling has been reported in [159] to achieve frequency relaxation. Modern control methodologies, such as distributed model predictive control [160, 161], two-degree-of-freedom and internal model control [162], have also been introduced into the area of LFC strategies. Furthermore, the hierarchical structure of Spanish AGC system has been described and analyzed in [155], and an adaptive gain control strategy was also designed in [154] for AGC with a master regulator to coordinate the control areas. In China, an improved PI control based hierarchical AGC system has been developed by Nanjing Automation Research Institute (NARI) and implemented in several large interconnected power grids [10, 11].

2.2.3 State-of-the-Art MOPD optimization

With the development of MOPD optimization problem, a number of Pareto optimality-based (deterministic or stochastic) methods can be available to solve the nonlinear and multiobjective combinatorial optimization problem. The state-of-the-art multiobjective techniques based on deterministic mathematical models for the well-know biobjective EED has been summarized and reported in [13, 15, 17]. Specifically, the classical EED approaches contain the linear and nonlinear goal programming techniques [163], linear programming (LP) [164], ϵ -constraint method [165, 166], and weighted mini-max technique [167], and these methods have been tested and validated on small power systems comprising six generators or three generators. However, these deterministic optimization methods cannot effectively handle the highly constrained MOPD with high discontinuity effects due to the complex system constraints and nonconvex prohibited operating zones (POZs) in the generation patch.

In recent years, various multiobjective optimality-based heuristic stochastic search algorithms have been proposed and successfully applied to the EED issue on a small IEEE 30-bus 6-generator power system model to obtain the Pareto tradeoff between fuel cost and pollutant emissions [14, 49-63]. These algorithms can generate a series of noninferior solutions based on different stochastic search strategies in a single simulation run. A fuzzy satisfaction-maximizing decision method has been utilized for solving this EED problem [168], but the inclusion of more objective functions in the MOPD is difficult to be tackled by this method. In order to resolve this problem, an interactive fuzzy satisfying approach with fuzzy goals was updated for DMs and addressed in [169] to optimize economy, emission and transmission line security together. However, its fuzzy membership functions were defined with intuitive knowledge and experiences of the designer. In literature [53], an abductive reasoning network as well as technique for order preference by similarity to ideal solution has been adopted for the real-time EED

decision making to ascertain the best compromise solution continuously. At the same time, the goal-attainment and adaptive polynomial network were used to achieve this real-time dual-objective optimization [170]. However, the methods cannot provide the scientific framework for guiding and guaranteeing solutions toward the true PF. Furthermore, a modified non-inferior surface estimation was proposed in [55] for the tradeoff Pareto curve of generation cost versus emission in consideration of multiple pollutants SO_2 , CO_2 and NO_x , and for each type of pollutant, the emission reduction rate was also researched in this study.

In the context of Pareto optimality theoretics for multiobjective problems, a series of famous MOEAs and group search techniques have been developed and reported in this research field. In particular, an improved EA-based framework was presented in [14] to model and estimate the economic impacts of emission dispatching and fuel switching, but some dominated solutions might be mistaken as nondominated ones during the selection procedure, which would lead to loss of some Pareto-optimal solutions. In literature [171], a multi-objective stochastic search technique (MOSST) based on real coded GA and simulated annealing was presented. Nevertheless, the computational process of this technique is always time-consuming and premature convergence may be caused owing to its search bias. For the sake of alleviating the defects above, a strength Pareto evolutionary algorithm (SPEA) has been proposed in [59] for the nonlinear constrained EED, in which the diversity-preserving policy and fuzzy clustering were integrated in the algorithm. Later, two similar methods, namely nondominated sorting genetic algorithm (NSGA) [54] and niched Pareto genetic algorithm (NPGA) [52], were reported. Then, these three MOEA approaches were intensively compared and discussed in [172] with different PF quality measures.

Recently, several advanced MOEAs have further been presented on the basis of the algorithms above. A modularized framework for the EED generation mix dispatch with a solid oxide fuel cell system was developed in [49], in which a biobjective linear programming model and a Queuing multi-objective optimizer

were integrated in the stated framework. In literature [173], an improved GA was addressed to the optimal EED for a hydrothermal power system, and a multiplier updating was introduced in this algorithm as the constraint handling strategy. An improved NSGA, called the nondominated sorting genetic algorithm-II (NSGA-II), was employed to deal with the dynamic EED and compared to the previous MOEA techniques in [174]. However, the NSGA-II cannot maintain and enhance the uniformity and diversity of the resulting PF solutions. In order to overcome the drawback, the controlled elitism and dynamic crowding distance mechanisms were introduced in the developed approach to formulate a modified NSGA-II in [175], and then this meta-heuristic was applied to different benchmark power systems in the literature.

Most remarkably, a number of well-known single-objective stochastic group-search algorithms have been extended lately to establish a set of multiobjective optimization algorithms. A multiobjective PSO (MOPSO) was implemented to the MOPD in [62] by redefining the local and global best particle individuals. Besides, the PSO algorithm was combined with fuzzy set theory to formulate a fuzzified multiobjective PSO in [57] and a fuzzy clustering based multiobjective PSO (FCPSO) in [61]. Subsequently, a modified MOPSO has been proposed and performed to investigate and analyze both the stochastic and deterministic EED models in [176]. Also, [63] focuses on employing the MOPSO with local search to address a reserve-sharing based multi-area EED scheme. With the burgeoning rise of chaotic theory, a novel multiobjective chaotic ant swarm optimization was proposed in [50] by integrating the chaotic behaviors into each individual ant of ant colony algorithm, while a multiobjective chaotic particle swarm optimization with chaotic local search was also developed in [51] to study the EED problem including reliability constraint, ramp rate limits and prohibited operating zones. As presented in [56], a multiobjective differential evolution as well as crowding entropy-based diversity operator was carried out on the EED to demonstrate its effectiveness and verifiability. A hybrid Pareto-based optimization metaheuristic

was then raised in [58] on the strength of the improved MOPSO and differential evolution to solve the EED problem. The bacteria foraging optimization has been extended to produce a multiobjective bacteria foraging technique in [60] and then introduced into the domain of EED to verify its efficacy. However, the extension of these Pareto optimization algorithms to include more objectives on large-scale power systems is still a very involved query.

2.3 Reinforcement Learning & Markov Decision Processes

2.3.1 Markov Decision Processes

MDP theory is a branch of stochastic operations research to handle dynamic decision-making problems under stochastic circumstances with multiple period scales [46]. First and foremost, Markov process (or Markov chain) is defined as a specific memoryless type of stochastic process with Markov property, in which, for the process of future states, the conditional probability distribution of state transitions does not depend on the sequence of past preceding events, but instead, it only relies upon the current state. Generally speaking, there are three types of complicated stochastic processes in power systems: Markov, semi-Markov, and non-Markov environments [71]. Accordingly, the MDP is an extended research area in which an action strategy is applied to the Markov chain process as well as a reinforcement reward adopted as the motivation [47]. Typically, a MDP model is always used for the establishment of an action strategy to maximize a utility function related to expected rewards.

MDPs have firstly been formulated and investigated in 1960 [47], and have now been developed into discrete-time MDPs, continuous time Markov decision processes (CTMDPs), and semi-Markov decision processes (SMDPs). Except for the three basic MDPs above, several generalized MDPs, such as adaptive MDPs, constrained MDPs and partially observable MDPs, were also described in [71]. In addition, a stochastic system with inertia property (taking large time-delay link

for example) can be approximately depicted as a non-Markov sequence model. While the CTMDP and SMDP models can always be transformed reasonably to the DTMDPs, a typical DTMDP can be expressed as follows:

$$\{S, A(s_i), p_{ij}(a), r(s_i, a), V, s_i, s_j \in S, a \in A(s_j)\} \quad (2.8)$$

where S is the observed state space; $A(s_i)$ denotes the nonempty action set for the i th state; $p_{ij}(a)$ is the probability from state s_i to state s_j under action $a \in A(s_i)$; and $r(s_i, a)$ represents the reward of action a in the i th state; V expresses the objective or criterion in the model, called value function.

The MDP policy provides a mathematical decision-making framework for the decision maker or system operators with action selection rules in each situation, and it can be depicted as $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \dots, \pi_k, \pi_{k+1}, \dots)$, where π_k represents the action selection strategy at the k th iterative step. For $k \geq 0$, let s_k and a_k be the state and the action selected at iteration k , respectively, then it has been proved that a stochastic process $\{s_k, a_k, k \geq 0\}$ can well be modeled as a discrete time Markov chain under MDP policy $\boldsymbol{\pi}$ based on stochastic process theory [46]. This criterion lays a solid foundation for the extensive application of MDP methods in power systems. Consequently, MDPs are applicable for studying a wide range of optimization and control problems solved via various RL algorithms.

2.3.2 Reinforcement Learning Methodology

RL is an online self-learning technique in the domain of machine learning to solve MDP problems without explicit specification of the transition probabilities [45], and its basic philosophy is to regulate a closed-loop control rule mapping from operating states to control actions by means of training a series of online representative data so as to maximize the long-term rewards. Consequently, RL algorithms, which differ from statistical pattern recognition and ANN in that it does not rely on accurate historical training samples and *a priori* knowledge of

the controlled system, are computational approaches to learn experiences from interactions with environment by trial-and-error for the policy improvement.

Fig. 2.3 illustrates the generalized framework of RL systems, which consists of two interactive modules, 'world' and 'agent'. In the structure, RL assumes that the world can be described as a complex dynamic system by a finite set of states called state space S , and the agent can choose a sequence of actions from action space A to interact with the environment [177]. The agent is composed of three main submodules: state identification, reinforcement reward and action strategy. The first submodule is to identify the environment state as the input form of the agent and the reward submodule determines the agent's control objective. The final submodule, action strategy, which is the core decision-making module for the RL structure, is to update and evaluate the knowledge policy of the agent and select an action for each state based on this strategy submodule in order to effect on the environment.

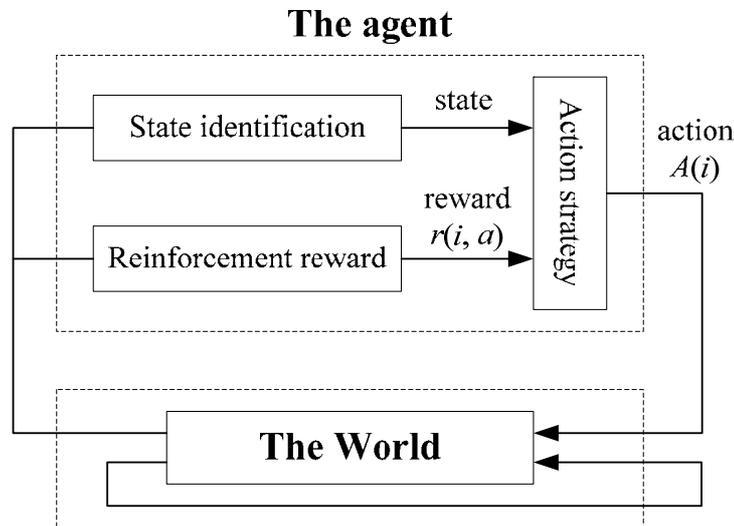


Fig. 2.3 Agent-world interactive structure for reinforcement learning

Almost all RL algorithms are developed to form the long-term optimal policy via repeatedly estimating a value function. There are main two optimality criteria for the value function model, discounted reward optimality criterion (DROC) and AROC. The RL value functions are always constituted from the viewpoint of

long-run cumulative rewards. Firstly, the value function of DROC is expressed as the sum of the discounted rewards,

$$V_{\text{DROC}} = r_k + \gamma r_{k+1} + \gamma^2 r_{k+2} + \cdots = \sum_{n=0}^{\infty} \gamma^n r_{k+n} \quad (2.9)$$

where parameter γ ($0 \leq \gamma \leq 1$) is the discount rate, and r_k denotes the immediate reward from the world at the k th iteration. Typical DROC-based RL algorithms include Q-learning, temporal difference (TD), dynamic programming, and $Q(\lambda)$ learning, etc. On the other hand, the value function of the AROC-based RL can be defined as the long-run average expected rewards, as follows,

$$V_{\text{AROC}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \text{Exp}\{r_k(s_k, a_k)\} \quad (2.10)$$

where $\text{Exp}(\bullet)$ represents the expectation function, and $r_k(s_k, a_k)$ is the immediate reward obtained at the k th iterative step after implementing action a_k in state s_k . The classical AROC-based RL algorithms include R-learning, H-learning, $R(\lambda)$ learning, and LC-learning [76].

The RL algorithms can be classified into two categories: model-free method and model-based method. The former consists of Q-learning and $Q(\lambda)$ learning, TD and $TD(\lambda)$ algorithms, R-learning and $R(\lambda)$ learning, Sarsa method, etc. And the latter include dynamic programming, Monte Carlo, adaptive heuristic critic (AHC), and so on [45]. Moreover, the emergence of multi-agent system (MAS) also gives impetus to the rapid development of distributed RL and hierarchical RL methodologies. Most hierarchical RL methods are underlain by the discrete-time SMDP formulation. Recently emerged hierarchical RL contain hierarchical abstract machines (HAM), options algorithm, MAXQ and HEXQ methods [178], while commonly used multi-agent RL involves Nash Q-learning, correlated Q-learning, and asymmetric Q-learning [179]. So far, the MDP and RL methods has been applied to various fields in power system including reactive control [180], power market [181], nonlinear turbogenerator system [182], stability control [74], and optimum maintenance policy [183], etc.

2.4 Multiobjective Optimization Algorithms

Real-world optimization problems usually need to cope with two or more conflicting and incommensurable objectives, and the multiobjective optimization aims to find a family of Pareto-optimal solutions in which none of the solutions can outperform any other with regard to the fitness values of all objectives [67]. In mathematical terms, multiobjective optimization can be formulated as,

$$\begin{aligned} \text{Min} \quad & f_i(X) \quad i = 1, 2, \dots, M_{\text{obj}} \\ \text{s.t.} \quad & \begin{cases} g_j(X) \leq 0, j = 1, 2, \dots, M_{\text{ineq}} \\ h_k(X) = 0, k = 1, 2, \dots, M_{\text{eq}} \end{cases} \end{aligned} \quad (2.11)$$

where f_i is the i th objective function; X is the decision vector that represents a solution to be optimized; M_{obj} , M_{ineq} and M_{eq} denote the numbers of objectives, inequality constraints and equality constraints, respectively.

Generally, for each solution vector in a multiobjective minimization problem, there is a dominance relationship between the solution being considered and the others. Any solution X_a covers or dominates X_b once conditions in (2.12) are satisfied [33].

$$\begin{cases} \forall i \in \{1, 2, \dots, M_{\text{obj}}\}: f_i(X_a) \leq f_i(X_b) \\ \exists m \in \{1, 2, \dots, M_{\text{obj}}\}: f_m(X_a) < f_m(X_b) \end{cases} \quad (2.12)$$

Any solution which cannot be dominated by other solutions of a given set is called the nondominated solution. The solutions which are nondominated within the entire feasible search space are known as Pareto-optimal solutions, and the set obtained by mapping these solutions to the fitness vectors in the objective space is the PF [65]. The front can be expressed as follows:

$$\text{PF} = \left\{ F = \left(f_1(X), f_2(X), \dots, f_{M_{\text{obj}}}(X) \right) \mid X \in X^* \right\} \quad (2.13)$$

where X^* is the set of all true Pareto-optimal solutions called the Pareto set. Since the complete Pareto-optimal solutions are always infinite, the determination of the Pareto set is extremely difficult and even infeasible due to memory constraint

and computational complexity caused by the presence of infinite suboptimal PFs. Consequently, the optimization goal of MGSO is to acquire a widely spread and well-distributed PF, in which the Pareto set can be diversified to cover maximum possible regions of the solution space, within a limited repository [34, 64]. The following are the three basic quality measure criteria for evaluating PFs resulted from various multiobjective optimization algorithms [184].

- (1) The distance of the resulting nondominated PF set to the true PF should be minimized.
- (2) The PF solutions shall be as uniformly distributed as possible.
- (3) The extent of the obtained nondominated PF in objective space should be maximized.

It should be pointed out that single-objective optimization algorithms require numerous simulation runs to obtain a desired set of Pareto solutions by varying the weights of objective functions while the Pareto set can be generated in one single run with the multiobjective algorithms. Therefore, it is unfair for comparisons of single-objective and multi-objective optimization solutions using above quality measure criteria. On the other hand, in the process of multiobjective optimization, the obtained nondominated set needs to be pruned within a desirable size while its trade-off characteristics shall be maintained [185]. There are already lots of cluster analysis techniques based on data mining in the literatures, such as block clustering, hierarchical clustering, and direct clustering, etc. Finally, the best compromise solution will be yielded by a MCDM approach to simulate the DM's preferences, and diverse MCDM categories have been summarized in [186], as follows:

- (1) The adoption of weight factors in which each weight expresses the relative importance degree of the corresponding objective;
- (2) The employment of aspiration levels corresponding to the desired levels of objective functions which the DM wants to reach;
- (3) The adoption of trade-offs for Pareto front surface between the objectives;

- (4) The use of reservation levels which correspond to the recognized levels of objective functions which the DM wishes to achieve;
- (5) The classification of these optimization objectives into different categories.

In this thesis, the equilibrium theory is firstly introduced to propose a novel equilibrium inspired decision making approach with the competing objectives regarded as noncooperative decision making players [78], which is belong to the third category of MCDM as mentioned above. The equilibrium theory is a branch of game theory to study the mixed-strategic decision making problem of games with many players [187]. There are various types of equilibrium model reported in the literatures for different stochastic game problems, such as Nash equilibria [188], Correlated equilibria [189], slightly altruistic equilibria [190], Stackelberg equilibria [191], and so on.

2.5 Summary

In this chapter, the fundamentals of multiobjective optimization and control for the power generation dispatch have been introduced. While a functional LFC framework is given to illustrate the essence of power system dispatch schemes, a survey is conducted to address the control performance standards as well as its modified version from a mathematical viewpoint. Then, a brief literature review on the current state of the art for AGC and MOPD methodologies is discussed in this chapter. Furthermore, the theories of MDPs and RL have been reviewed, and the Pareto-based multiobjective optimization is also presented and analyzed.

Chapter III

Stochastic Optimal CPS Control Methodology for Interconnected Power Systems Using Q-Learning

3.1 Introduction

Automatic generation control is an essential function for the daily operation of interconnected power grids to guarantee the frequency quality of each control area and to keep tie-line power close to the scheduled values by regulating the power outputs of AGC generators to accommodate fluctuating load demands [38]. In 1997, the NERC released new CPS standards to monitor and assess the TBC control performance in normal interconnected power system operation [9], and then the standards were widely adopted and carried out by most utilities over the world to replace the old CPC criteria. The purposes of CPS are to provide a solid statistical and physical foundation to loosen the requirement of ACE regulation and eliminate unnecessary AGC pulses which enforce ACE return to zero within ten minutes. Compared to the empirical A1 and A2, the pioneering studies on analysis and discussion of this new and more sophisticated CPS1 and CPS2, with the field-test experiences, have been reported in [37, 81-85].

The AGC strategy under CPS adopted in the SCADA/EMS systems delivers AGC regulating command every 4~16 seconds in a discrete manner. At present several large-scale interconnected power systems in China have implemented the NARI's hierarchical AGC system based on improved PI control structure [10, 11]. Although this control strategy is effective in maintaining frequency quality and providing energy support to interconnected control areas during emergencies, its PI control system with fixed PI gains, like most existing AGC systems, cannot

adapt well to large changes in operating mode, parameters, or structure of power grids. This has been observed in previous field experiments in Guangdong power grid. As a remedy, the model-free RL algorithms are introduced in this thesis in order to ameliorate the adaptivity and dynamic optimization performance of the existing AGC systems.

In CPS standards, the CPS1 and CPS2 metrics are considered as stochastic variables, and [37] concluded that the time averages of ACE in an interconnected control area is essentially unpredictable (even in sign) due to the randomness of the load if the averaging interval is greater than a specified measurement period. Furthermore, [82, 192] also discussed the uncertainties in the measured variables for LFC from a probabilistic and statistical point of view. Therefore, the modern AGC system under NERC's new CPS can better be formulated as an uncertain stochastic system, which can be modeled as a Gauss-Markov decision process based on the stochastic optimal control methodology [182]. Consequently, a DTMDP-based Q-learning algorithm is therefore introduced into the domain of CPS control to create a new QAGC controller. This model-free Q-learning method does not need to make any strong assumptions on the system dynamics and can be employed as an online learning control strategy, taking advantage of the CPS rewards to create a group of self-learning rules from the viewpoint of long-term optimal objective. As a result, the outputs of the proposed QAGC can be tuned rapidly and automatically.

3.2 Q-Learning

Q-learning [70] is the classical branch of model-free RL algorithms in the stochastic optimal MDP control theory, and it can be viewed as an asynchronous dynamic programming to solve the DTMDP domains with incomplete or partial information. Also, the Q-learning agent with the capability of self-learning can operate optimally by evaluating the consequences of events, without building the

maps of Markovian fields, in order to improve successively its action policy, and the goal is to find an optimal action strategy via maximizing a long-term DROC value function. Following the RL theory outlined in Section 2.3, the so-called state-action value function or Q-function can be defined as follows,

$$Q(s, a) = r(s, s', a) + \gamma \sum_{s' \in S} P(s' | s, a) \max_{a \in A} Q(s', a) \quad (3.1)$$

where s and s' represent the current and next state, respectively; $r(s, s', a)$ denotes the immediate reward after implementing action a ; $P(s'/s, a)$ is the probability of transition from state s to s' under action a ; $Q(s, a)$ represents the expected sum of rewards when starting from an initial state s_0 , taking action a , and performing optimal action a' in next searches, until the optimal value of Q-function ($Q^*(s, a)$) is achieved.

An iterative policy should be designed for Q-learning to estimate the optimal Q-function, and the policy will impact the performance of the algorithm in that it determines which state-action pairs will be visited and updated. Suppose $Q_k(s, a)$ is the estimation of the optimal Q-function Q^* at the k th iterative step, and s_k and a_k represent the state and action at the k th iterative step, respectively. Hence, the value function $Q_{k+1}(s, a)$ can make a useful update based on the action taken and reward received, and the following iterative policy is adopted for the one-step Q-learning as below,

$$Q_{k+1}(s, a) = \begin{cases} Q_k(s, a) + \alpha [r(s, s', a) + \gamma \max_{a \in A} Q_k(s', a) - Q_k(s, a)] & (s, a) = (s_k, a_k) \\ Q_k(s, a) & \forall (s, a) \neq (s_k, a_k) \end{cases} \quad (3.2)$$

where parameter α ($0 < \alpha < 1$) is a constant called learning factor to indicate the step size of learning for Q-functions, and the term within the square brackets can be considered as the estimation for the current value function error.

The algorithm is an online self-learning and dynamic optimization technique which learns to act in an optimal interactive way through experience gained by exploration and exploitation. The agent can handle one sample at a time and does

not require to explicitly storing all samples. For the action strategy, the greedy action policy a_g is defined firstly as follows,

$$a_g = \arg \max_{a \in A} Q_k(s, a) \quad (3.3)$$

where $\arg \max$ represents the action a corresponding to maximal Q-function for a given state s at the k th iteration.

The purpose of the action greedy policy is to execute the best control actions in the short term. However, this greedy policy always exploits current knowledge to maximize the immediate reward, and only implementation of the greedy action in each state mean that the agent cannot explore sufficiently rich combinations of state-action pairs. Therefore, in order to achieve better sampling and convergence to the optimal policy, an action exploration policy called pursuit method [44] can be adopted in which, for each state s , actions are chosen based on a probability distribution over the action space. Such a sequence of samples can be obtained either through a simulation model of the system or observing the practical system in operation. Initially, a uniform probability distribution is firstly used so that the QAGC acts randomly in the beginning of self-learning procedure, then the action selection probabilities can be updated as follows,

$$P_{k+1}(s, a) = \begin{cases} P_k(s, a) + \xi(1 - P_k(s, a)) & s = s_k, a = a_g \\ P_k(s, a)(1 - \xi) & s = s_k, a \neq a_g \\ P_k(s, a) & \text{otherwise} \end{cases} \quad (3.4)$$

where parameter ξ ($0 < \xi < 1$) is a constant called the action exploration factor; $P_k(s, a)$ represents the probability with which action a selected for state s at the k th iteration of learning; a_g is the greedy action in state s with respect to $Q_k(s, a)$. Thus, for all states of interest, Q-function would be close to $Q^*(s, a)$ for all state-action pairs after a sufficiently large number of iterations. A rigorous proof has been concluded in [193] that Q-learning can converge to its optimal strategy with probability 1 once all action-state pairs are represented discretely and repeatedly sampled. In addition, more general convergence characteristics of RL have been

proved later in [194]. Since both action space A and state space S are finite, the probability distribution of control actions can be stored as a finite matrix (exactly in a tabular form) and the value of Q-functions as a lookup table, respectively.

3.3 Design of Q-Learning Based AGC

3.3.1 Controller Framework

The CPS metrics are not only assessment criteria for AGC performance of interconnected operation but also can be regarded as the reward measures for the LFC control quality. Moreover, there are always some uncertain parameters and stochastic load disturbances existing in the system, especially for natural frequency response coefficient β which is a time-varying and nonlinear variable [3]. Consequently, the AGC problem under CPS is a dynamic decision-making problem, which can be transformed as a non-time-homogeneous MDP optimization process. Q-learning is an optimized way to maximize the long-term cumulative reward on the basis of stochastic control theory, and hence is very applicable for the design of stochastic optimal CPS control strategy.

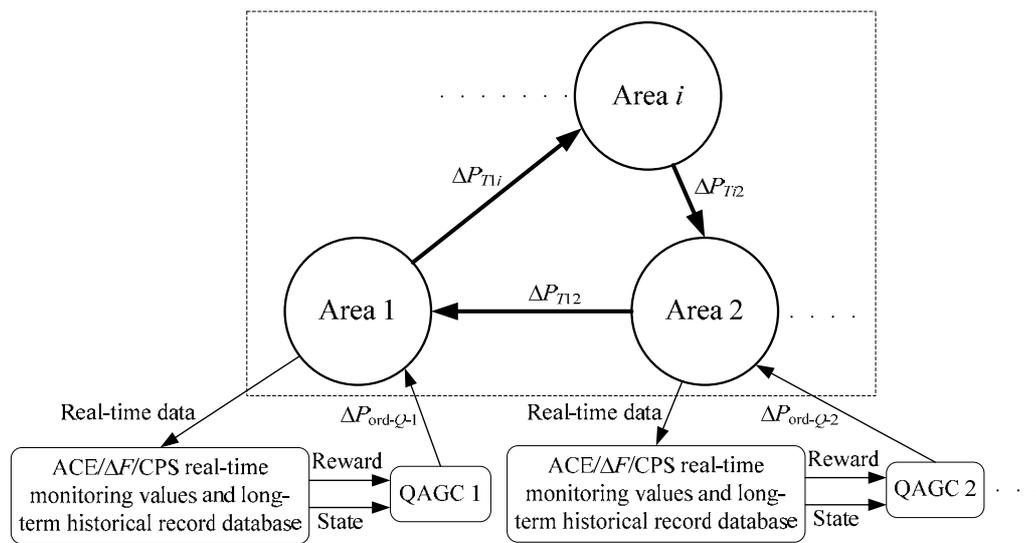


Fig. 3.1 Q-learning based optimized AGC control framework

The proposed Q-learning based dynamic optimal AGC control framework is illustrated in Fig. 3.1. In this framework, the independent QAGC governs each control area in a decentralized manner, and there are two modules in each area: the “ACE/ ΔF /CPS real-time monitoring database” is responsible for the real-time supervision and data acquisition, such as the instantaneous and average values of ACE, ΔF and CPS1 [80]. Besides, the statistical data of CPS compliances can be recorded and deposited in the “long-term historical record database” [41], and these databases can provide the “state” and “reward” from interconnected power grids as the input signals for each QAGC. Thereby, the QAGCs can be implemented for the online optimization of AGC generation regulating commands $\Delta P_{\text{ord-}Q-i}$ on the basis of the RL mechanisms [70].

3.3.2 State and Action Space Discretization

One of the major attractions of the Q-learning is the significant flexibility in designing state-action pairs and reward function for a specified control strategy. In the proposed QAGC, the state-action spaces should be reasonably discretized, and the reward function should be defined properly. The action space A involves a finite set of the optimized AGC regulating command $\Delta P_{\text{ord-}i}$ for the output of QAGC, which denotes the variation in generator setpoint within an AGC cycle. The discretization of action vector should be conducted in terms of the system spinning reserve and units’ regulating characteristics in the specified control area, and the specific action spaces will be presented in the following case studies.

Furthermore, the instantaneous values of CPS1 and ACE indices are selected as the state variables to constitute state space S (the state space of Markov chain). Since the CPS state space is infinite, as depicted to a 2-dimensional vector space in Fig. 3.2, the space discretization technique shall be used for dividing it into a finite range of regions to comply with CPS assessment and the discrete relaxed control characteristics. Therefore, according to CPS assessment and management principles in [41], the first state variable CPS1 can be quantized as the following

twenty-three levels: $(-\infty, 0)$, $[0, 100)$, $[100, 105)$, $[105, 110)$, $[110, 115)$, \dots , $[185, 190)$, $[190, 195)$, $[195, 200)$, $[200, +\infty)$. It should be pointed out that the range of CPS1 from 100 to 200 is discretized into 20 levels by 5%. This is because power utilities usually utilize these levels as the performance assessment criterion for dispatching operators [41]. The second state variable ACE is to distinguish the cause for the change of CPS1 index. It depends on whether the ACE value is negative or positive, and the state can be set as -1 or 1 level-state, respectively. Thus, the total number of the input states for QAGC is 46, and the learning step T_{step} is determined by AGC decision cycle time. Consequently, the CPS control state space is then transformed into a discrete-time stochastic state model.

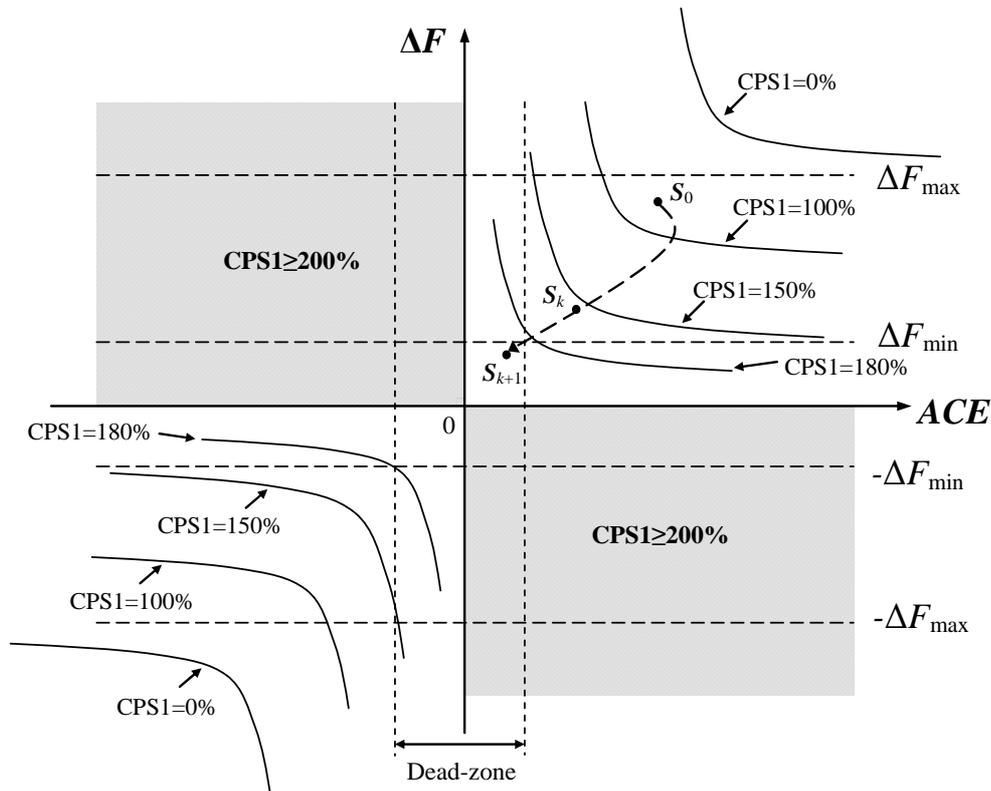


Fig. 3.2 The schematic diagram of CPS control state space

3.3.3 Control Objective of QAGC

To take full advantage of NERC's new CPS, the design objectives of a highly robust and adaptable AGC controller are (1) to comply with the CPS1 and CPS2

metrics for interconnected power systems under complicated dynamic operating conditions; (2) to relax the regulating pressure as well as the control cost of AGC plants. The existing AGC strategies always dispatch the outputs of AGC plants blindly in pursuit of high CPS compliances, and this will often result in the so-called “over-compliant” problem for CPS standards [40]. A major cause of this over-compliant problem is that the power dispatch centers adopt the overly strict AGC schemes. Hence, a new multiobjective control philosophy, namely optimal relaxed AGC control, is proposed in this thesis, in which the AGC plants would realize costly loose maneuvering to carry out the dynamic optimal AGC strategy on the premise of complying with the assessment metrics CPS1 and CPS2. The proposed AGC for the relaxation of LFC system can be implemented with high feasibility and practicability on the foundation of previous field CPS experiments on the CSG power systems [195].

Based on the above optimal relaxed control, the multiobjective function of QAGC should combine the CPS compliance and relaxed control objectives. For most of control problems, the multiple control objectives cannot be optimized simultaneously for Pareto optimality due to the real-time control requirement. Therefore, the control function can be expressed as the following quadratic form,

$$\begin{aligned}
\min \quad & J = \int_{t_0}^{t_k} X^T(t) Q_L X(t) + u^T(t) R_L u(t) \\
& = \sum_{t=1}^T q_{Li} \sum_{k=1}^t [K_{CPS^*} - K_{CPS}]^2 + \sum_{t=1}^T \sum_{i \in S_G} r_{Li} \sum_{k=1}^t \Delta P_{Gi}^2(k) \quad (3.5) \\
s.t. \quad & P_{Gi, \min} \leq P_{Gi} \leq P_{Gi, \max} \\
& P_{Tmn} \leq P_{Tmn, \max}
\end{aligned}$$

where Q_L and R_L are the weight matrices of state variables and control inputs in linear quadratic regulator (LQR) [196], respectively; $X(t)$ and $u(t)$ are state matrix and input matrix of the control system at iterative time t ; K_{CPS} and K_{CPS^*} express the degree of CPS compliance and the preset target of CPS compliance during a specified assessment period respectively; q_{Li} and r_{Li} are the cost coefficients for the i th AGC generator in weight matrices Q and R ; S_G represents the set of AGC

generators; $\Delta P_{Gi}(k)$ is the regulating power variation in the i th AGC generator at the k th iterative time; P_{Gi} represents the active power output of the i th AGC unit; $P_{Gi,\max}$ and $P_{Gi,\min}$ are upper and lower generation output limits of the i th AGC unit, respectively; P_{Tmn} and $P_{Tmn,\max}$ are the active power flow and its security limit of the transmission line between bus m and bus n .

3.3.4 Design of Reward Function

The next step of the QAGC design is to define the immediate reward function $r(s,s',a)$ for the k th AGC decision cycle. The controller observes operating state s_k of the power systems, as described by CPS1 and ACE, and chooses a control action a_k from the action space. Then a new state s_{k+1} can be obtained and an immediate reward ($S \times A \rightarrow r$) is provided to express the degree of satisfaction of AGC performances. This procedure will be repeated for a sufficient number of state-action events until no more changes in the Q-function can be achieved, and then the optimal control strategy can be gradually learned. In the reward function, the 10-min averages of CPS1 and ACE (refer to CPS2) cannot be taken as the input state variables, because the QAGC would be completely incapable of satisfying the real-time control requirement if the optimization is executed at intervals of 10 minutes. Therefore, the feedback reward function should hereby be determined using the real-time values of CPS1 and ACE, which can be calculated with the instantaneous values of ACE and frequency deviation using Eq. (2.1)~(2.3).

The control target of QAGC, as formulated in Eq. (3.5), should be embodied with its reward function. For the sake of establishing a desirable reward function, the reward function $r_i(s,s',a)$ for control area i can hence be computed using the following piecewise function:

$$\begin{cases} r_i(k) = o_i, & o_i \geq 0 & CPS1_i(k) \geq 200 \\ r_i(k) = -\{\eta_{1i}[ACE_i(k) - ACE_i^*]^2 + \mu_{1i}[a_{\text{ord-}i}(k) - a_{\text{ord-}i}^*]^2\} & CPS1_i(k) \in [100, 200) \\ r_i(k) = -\{\eta_{2i}[CPS1_i(k) - CPS1_i^*]^2 + \mu_{2i}[a_{\text{ord-}i}(k) - a_{\text{ord-}i}^*]^2\} & CPS1_i(k) < 100 \end{cases} \quad (3.6)$$

where o_i is a nonnegative constant which is set to 0 in the case study; $CPS1_i(k)$ and $ACE_i(k)$ are the instantaneous values of CPS1 and ACE at the k th iteration; $CPS1_i^*$ and ACE_i^* express the preset target values for the state inputs CPS1 and ACE, respectively. As for the $CPS1_i^*$, experiences showed that a value of 200 works well if high CPS compliance is required, and here, the daily or monthly mean of CPS1 compliance can be set to implement the relaxed control; The value of ACE_i^* , in this application, is used to specify the threshold value of the dead-zone of AGC system [11] in order to ensure CPS2 compliance, reduce inadvertent power interchanges, and prevent ACE from zero-crossing frequently; $a_{ord-i}(k)$ is the index to a selected action from the action space A at the k th iteration instead of its actual power value, and a_{ord-i}^* is an index to the null control action in the space A . The purpose of the second square terms in reward function is to decrease the mechanical wear-and-tear of AGC generators and the economic cost resulting from large fluctuations in the power regulating commands. η_{1i} , η_{2i} and μ_{1i} , μ_{2i} are the optimum weight factors for the reward function in area i , which are equivalent to the parameters of matrix Q and R in the LQR theory [196].

3.3.5 Q-learning Parameter Settings

In addition, the three parameters γ , α and ξ in Eq. (3.1)~(3.4) are important in implementing the algorithm and shall be set wisely, for example, following the generic guidelines [44, 70, 194]. For the AGC problem, parameter γ used in Eq. (3.1) is the control factor by which later rewards are discounted, and simulation studies show that any value in the range from 0.5 to 0.98 works well. Since later rewards are very important for QAGC and a value close to 1 shall be chosen, γ is set to 0.9 in this case study. The critical parameter α , as introduced in Eq. (3.2), expresses the step size for updating Q-function and the rate of self-learning. This parameter determines the extent to which a single training example modifies the policy in the learning process. Since the QAGC problem under CPS depends on previous control actions, simulation studies show the value in the wide range of

0.001~0.1 is adequate to this problem due to the randomness of load disturbances, and here it is set to 0.1. The third factor ζ , used in Eq. (3.4) for updating the probability distribution of action selection, essentially determines the extent of explorations of action strategy. Case studies indicate that a very small value will slow down the learning convergence while the controller will tend to act as the greedy policy with a large one, and the policy remains unaffected with ζ between 0.3 and 0.6. In the following, an intermediate value of 0.5 is used.

3.3.6 Execution Steps of QAGC

Following the analytical framework as described above, the QAGC learns the optimal control strategy by a set of online training tuples of form $\{s, s', a, r\}$. The execution steps of Q-learning applied to AGC under CPS is illustrated in Fig. 3.3.

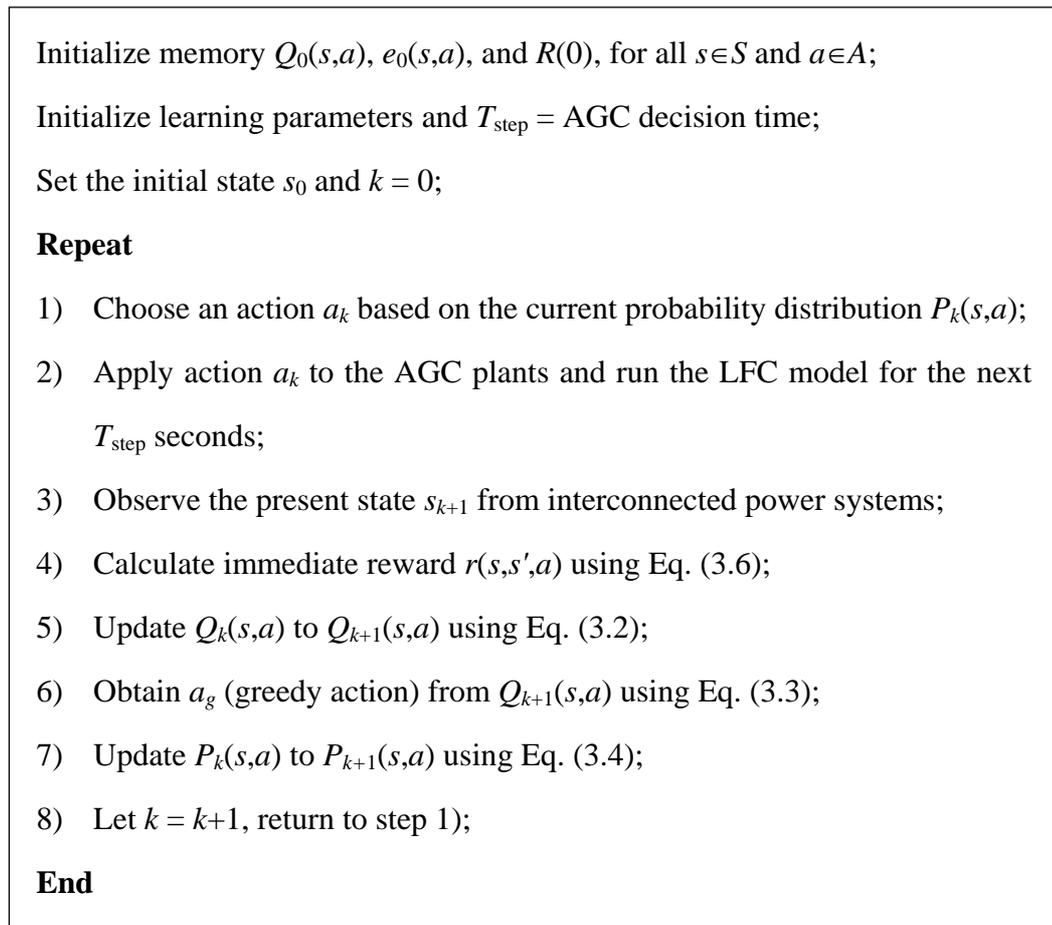


Fig. 3.3 Execution steps of the proposed QAGC

3.4 Procedure of Semisupervisory Group Prelearning

Normally, the RL controllers should be scheduled to experience a series of pre-learning procedure for the onsite application in the initial stage of online self-learning, which is a stochastic action exploration process in the CPS state space. The main drawback of this online pre-learning benefiting from the fact is that the controller may jeopardize security and stability of the real system, because at the beginning of the interaction, no experience is available to the RL controllers. One effective solution to this problem is to implement the controller in a simulation environment first [74].

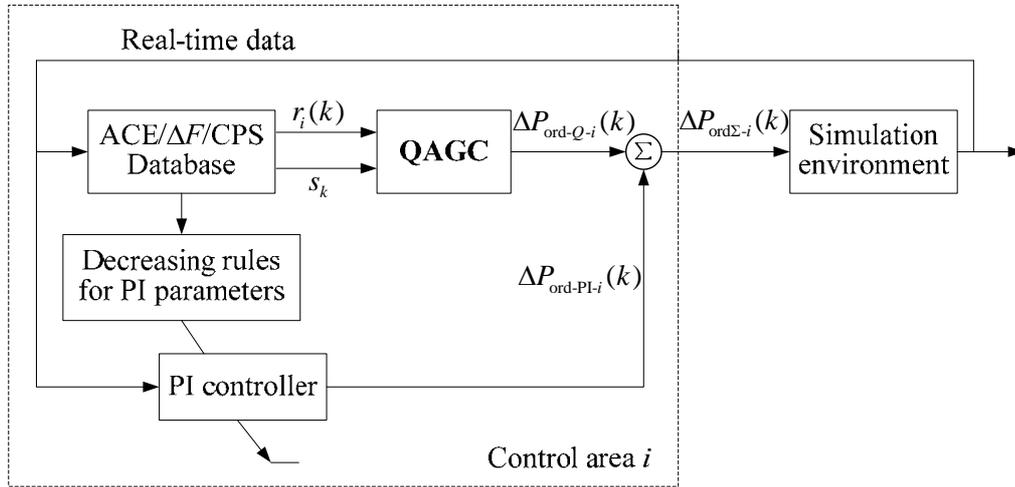


Fig. 3.4 Framework of semi-supervisory group pre-learning method

This chapter hereby presents a novel preconditioning technique, called semi-supervisory group pre-learning method, to solve the group pre-learning problem involving multiple QAGCs in multi-area interconnected power systems. This method can be implemented as auxiliary correction and stabilization for multi-controllers pre-learning by means of the PI/RL structure, as described in Fig. 3.4 and 3.5. Moreover, the proposed method can also provide a better convergence property for the QAGC, and its iterative termination criterion for the pre-learning process is determined as a matrix 2-norms $\|Q_{k+1}(s,a) - Q_k(s,a)\|_2 \leq \zeta$, ζ is a given precision factor. Fig. 3.5 illustrates the execution procedures of the proposed pre-

learning method, and this process will end once the iterative termination criterion is satisfied. The controller would make online learning control in a smooth way instead of random output fluctuations after the pre-learning process, and then the proposed QAGC can be put into onsite operation for its practical implementation thereafter [74].

- 1) Firstly, a high-accuracy digital simulation model is established for real power systems, and all AGC controllers adopt the PI control structure [1] in each control area;
- 2) The QAGC is designed as an additional corrective control module for the AGC system in area i , as shown in Fig. 3.4, and then it interacts with the simulation environment as in Fig. 3.3 until the stable convergence of the PI/RL control structure;
- 3) Reduce the gains of the PI controller based on the linear decreasing rules while the QAGC would adapt to the changing operation conditions, such that eventually an AGC framework without PI controller can be obtained once the PI gains and outputs finally decrease to zero;
- 4) Repeat the step 2 and 3 until all AGC controllers only consist of QAGCs in all interconnected areas, i.e. all PI controller gains become zero, and then the PI controller in each area can be removed;
- 5) Observe system states and control strategy of the AGC system, terminate the pre-learning process until no more changes in the Q-functions;
- 6) Store all the current action probability distributions and Q-functions after the pre-learning process. The QAGCs can then be implemented into the real power system and extract the learned AGC strategies for their online optimization with the practical system.

Fig. 3.5 Semi-supervisory group pre-learning procedures

3.5 Simulation Studies

3.5.1 Investigated System

The performance of AGC strategies are mostly evaluated based on classical linear LFC models as recommended in [1, 42], and the uncertainties in the power system models are usually simulated with the load demand fluctuations as well as parameter perturbations. In this chapter, a representative LFC model of two-area interconnected power system is adopted as the benchmark system to analyze and investigate the AGC methodologies, as shown in Fig. 3.6. This LFC model, in which active and reactive power are decoupled to quantitatively investigate the power system load-frequency characteristics, mainly contains four fundamental system models: load and generator model, governor model, prime-mover model and tie-line model. In the following case studies, the model parameters are taken from [42] and listed in Appendix A with the base capacity of 5000MW. Besides, the simulation of LFC model is executed using the software platform of Matlab /Simulink and the QAGC is programmed as the S-Function module.

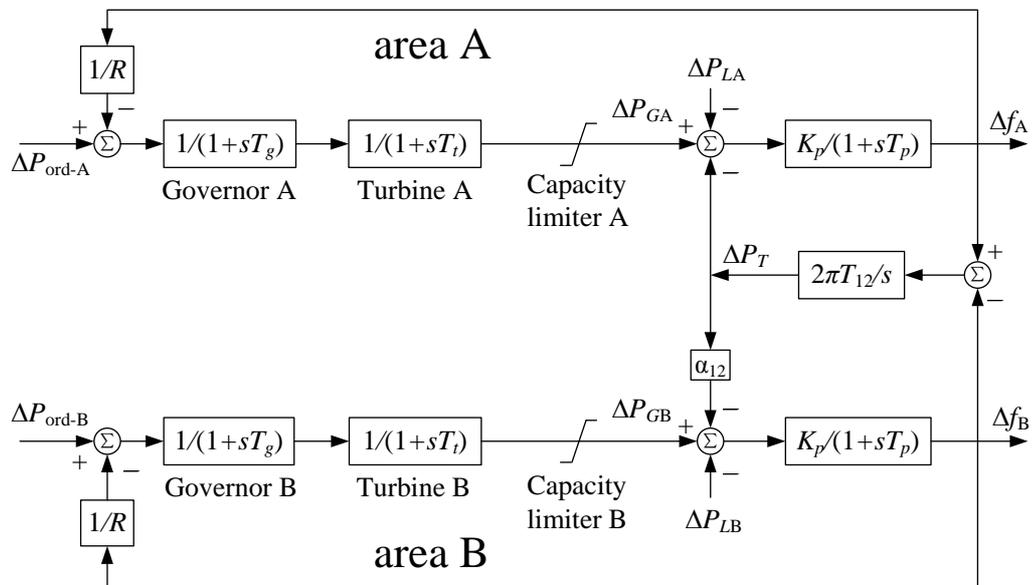
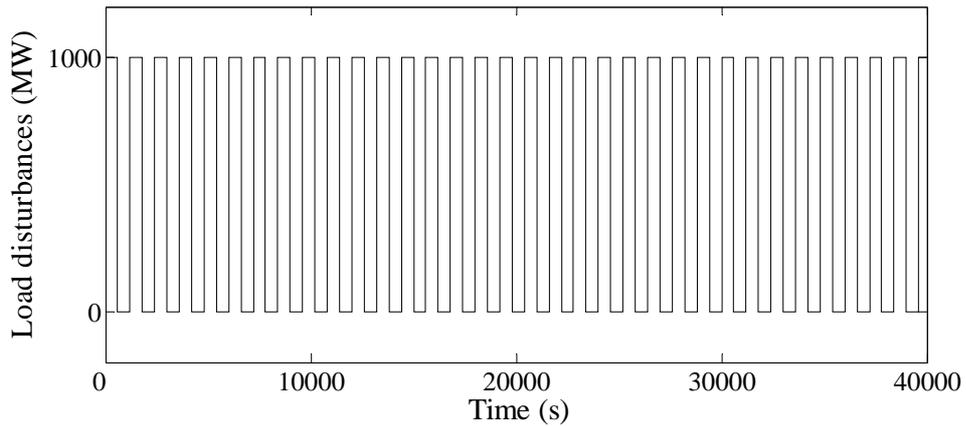


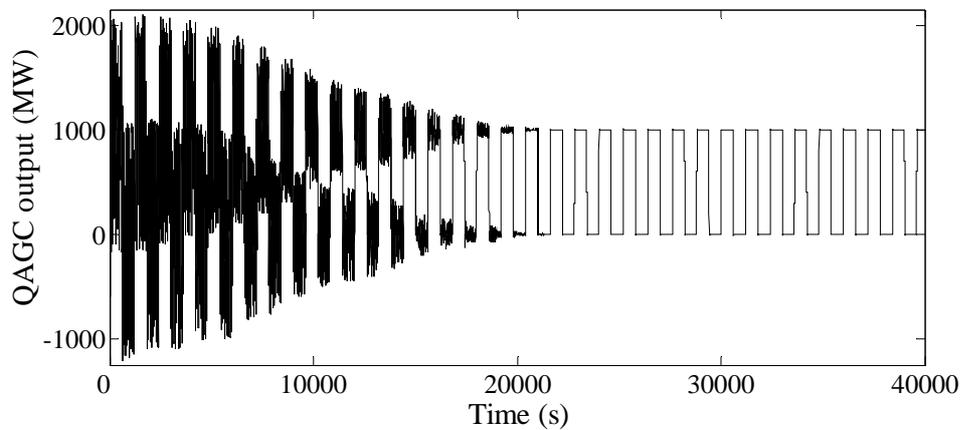
Fig. 3.6 Two-area power system LFC model

3.5.2 Performance Results

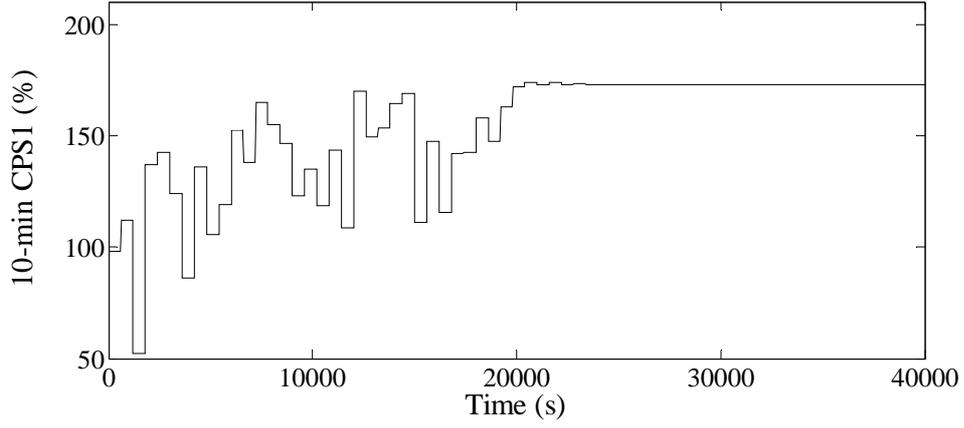
For clear illustration of the implementation steps involved and the proposed pre-learning method, the following settings were designed. As the AGC decision cycle time is set to 3 seconds and the two-dimensional variables ($CPS1$, ACE) are set as state input signals to QAGCs, the action vector A is discretized into fifteen levels as $\{-500, -300, -100, -50, -20, -10, -5, 0, 5, 10, 20, 50, 100, 300, 500\}$ MW. Then, the QAGCs are trained by carrying out the simulation for the semi-supervisory group pre-learning technique as in Fig. 3.5. During the pre-learning process, the controllers can explore the state-action spaces with a group of 10-min periodic square-wave load disturbances, and the typical pre-learning process corresponding to the disturbances in area A is shown in Fig. 3.7.



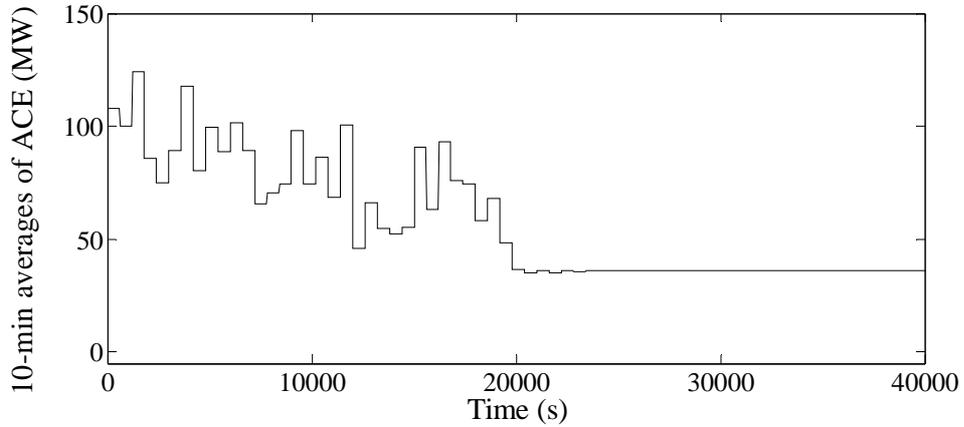
(a) Periodic square-wave load disturbances ΔP_{LA}



(b) QAGC output ΔP_{ord-A} in pre-learning process



(c) The plot of clock-10-min CPS1 index in pre-learning process

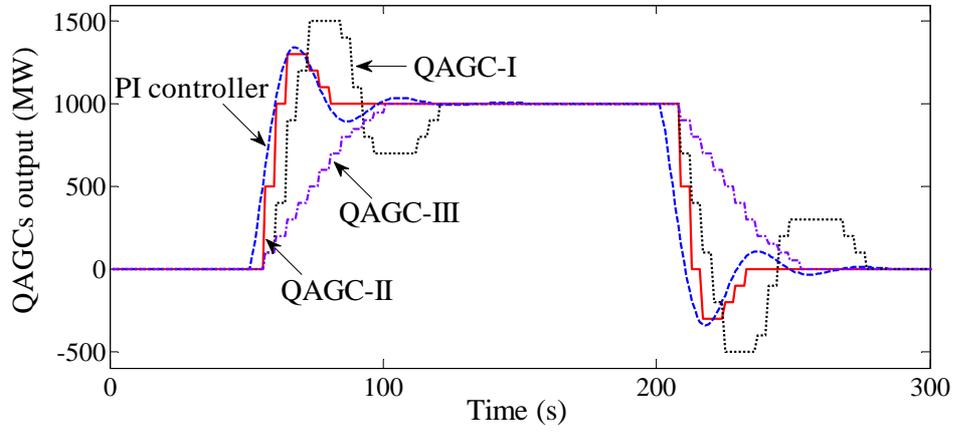


(d) The plot of clock-10-min $|E_{AVE-min}|$ in pre-learning process

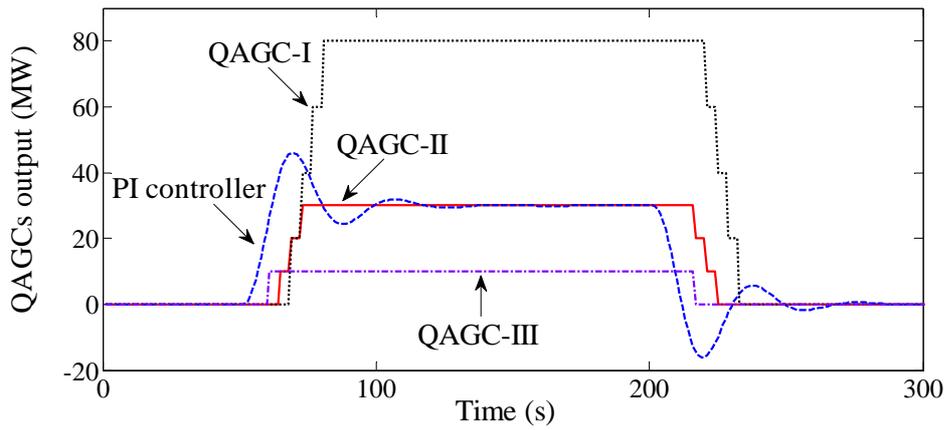
Fig. 3.7 Pre-learning process of the proposed QAGC in area A

It can be found in Fig. 3.7 that the controller's actions at various states have converged to their deterministic control policy while the clock-10-min CPS1 and $|E_{AVE-min}|$ (10-min absolute averages of ACE) tend to stabilize to constant values. In this pre-learning process, the contribution of PI controllers gradually decreases as their gains reduce, and the QAGCs can adapt well to the ever-changing external environment via the RL mechanism. Consequently, the Q-learning can obtain the optimal action strategy for the AGC system (apart from PI controller) once the PI gains reduce to 0. Then, the proposed controllers can be implemented in practical operation once the pre-learning process in each control area is completed, and the performance of QAGCs can further be improved via the interactions with the real power systems.

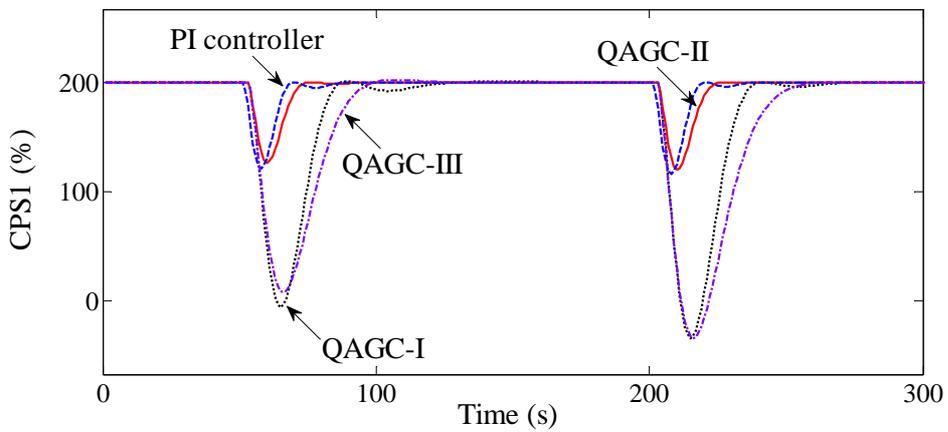
For further illustration of the important role of the reward functions for the Q-learning control performance, three typical groups of weight factors in Eq. (3.6) are adopted to the QAGCs for comparison and discussion as below,



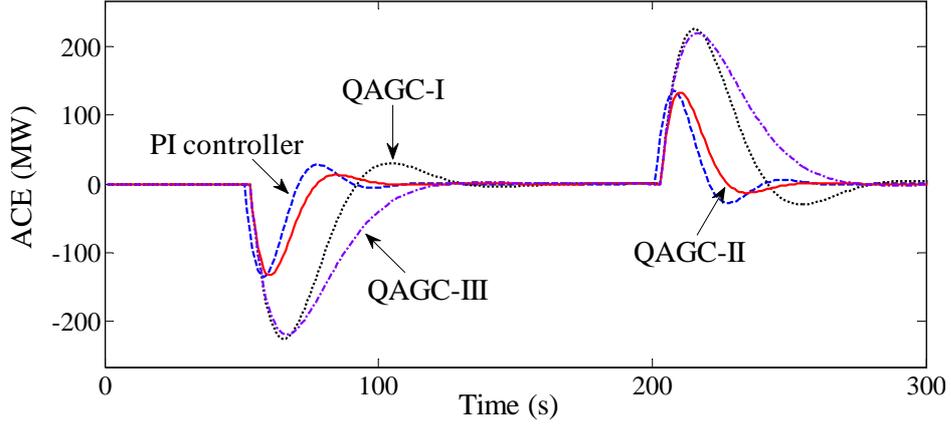
(a) The plot of QAGCs output $\Delta P_{ord-Q-A}$ in area A



(b) The plot of QAGCs output $\Delta P_{ord-Q-B}$ in area B



(c) The curve of instantaneous CPS1 index in area A



(d) The curve of instantaneous ACE in area A

Fig. 3.8 Simulation test of the optimal QAGCs

- QAGC-I: $\eta_1=1, \eta_2=50, \mu_1=1, \mu_2=1$;
- QAGC-II: $\eta_1=1, \eta_2=50, \mu_1=10, \mu_2=10$;
- QAGC-III: $\eta_1=1, \eta_2=50, \mu_1=50, \mu_2=50$.

Here, the fairness and conformance in the reward function can be ensured by maintaining a constant weight ratio η_1/η_2 , which is related to the parameter B_i and ε_1 , a value of 50 is used here from our observation, and μ_1 can also be equal to μ_2 thereby. Since the step responses are the most typical representations to evaluate the dynamic performance of control systems, a square-wave load disturbance is introduced to simulate generator shutdown and load shedding in the LFC system model. In this simulation study, the reward functions of QAGCs in two control areas are identical, and the performance results of the simulation tests are shown in Fig. 3.8. It can be found that different reward functions have the significantly influences on the optimal behaviors of the QAGC, and the control effects of the QAGC-II can outperform PI controller by choosing the appropriate weights.

On the basis of the LQR principle [196] and the simulation test results, the conclusion can be drawn that the meanings of η_1, η_2 and μ_1, μ_2 are very similar to weight matrices Q_L and R_L in the LQR approach. The power control outputs will slow down with the decrease in weight ratio η/μ , so that the regulating cost and maneuvering pressure of AGC plants would be released to achieve the loosened

control based on NERC's standards; conversely, QAGCs would also tend toward the tightened control and thus attach more relative importance to the CPS metrics. Consequently, with regard to the optimum QAGC design, weight ratio η/μ should be thoughtfully set for the coordination and balance between the CPS assessment objective and AGC relaxed objective in terms of the above mentioned laws. The selection guideline for weight ratio η/μ will be further analyzed and discussed in the following Chapter IV.

3.6 Summary

Conserving energy in grid dispatch is a key technology and core content for AGC control strategies based on CPS criterion. Interpreting the previous results, the proposed stochastic optimal CPS control methodology based on Q-learning possesses the following several advantages:

- (1) The design of the proposed QAGC does not depend on any knowledge of the electric network configuration and system dynamics, and the RL driven controller can learn the closed-loop control laws known to be well adapt to the stochastic behaviors and effects of nonlinearities;
- (2) The reward function of QAGC plays a very important role to AGC control performances, which originates from the combination of the CPS evaluation objective and AGC relaxed objective through the linear weighted aggregate approach in this chapter;
- (3) As the stochastic optimal control based Q-learning can learn continuously by interaction with its environment, the algorithm opens avenues to the self-adaptive control. Therefore, the QAGC can cope with the complex changing operation scenarios in power systems, and the robustness and adaptability of the AGC system can be enhanced and the degree of CPS compliance is evidently improved.

Chapter IV

Stochastic Optimal Relaxed AGC Methodology in Non-Markov Environment Based on Multi-step $Q(\lambda)$ Learning

4.1 Introduction

The AGC performance in interconnected power system operation has usually been measured against a number of evaluation metrics including the compliance with some reliability standards based on ACE and system frequency [84]. The CPS [37] established by NERC in 1997 have been widely adopted by most of utilities around the world including East China Power Grid and the CSG in 2001 and 2005, respectively. Coupled with the renewed interest in the energy saving in power dispatch [158] and the statistical concepts for CPS1 and CPS2 as well as the rapid development of smart grid technologies [35], a new generation of CPS oriented optimal AGC system has become a pressing need.

The state-of-the-art in AGC strategies designed to work under CPS has been comprehensively reported in [38, 39]. Even with the wide adoption of the CPS nowadays, the existing AGC systems have not yet been optimized to fully explore the potential of this more sophisticated criterion. CPS pays more attention to the medium and long-run returns of AGC performances and regards CPS1/CPS2 indices as random variables. As mentioned in [37] that the time averages of the ACE for some averaging intervals may be sufficiently random to make CPS2 a sufficient control on 10-min averages of interconnection frequency, while the CPS1 directly controls the 1-min averages of interconnection frequency. Coupled with the studies on the uncertainties in the measured variables for LFC

systems [82, 192, 197], further dynamic response studies show that AGC system under CPS in fact can better be formulated as an uncertain stochastic system from the statistical and probabilistic point of view. On the other hand, the CPS provides a theoretical foundation for the relaxation of frequency. If the frequency characteristics could be relaxed without threatening CPS compliances and system reliability, the LFC strategy is allowed to achieve for the desired relaxed control in order to economize on energy in power system dispatch.

Based on the studies in Chapter III, the design of a stochastic optimal relaxed AGC scheme using multi-step $Q(\lambda)$ learning algorithm [73], which has recently emerged and is suitable for the non-Markov environment [72], is investigated and presented in this chapter. The proposed optimal relaxed AGC means that AGC plants would suffer less from costly maneuvering to find the optimum policy with the objectives of complying with the CPS1 and CPS2 standards. The multi-step $Q(\lambda)$ backtracking learning is utilized to overcome the long time-delay problem caused by the steam turbine of AGC thermal units in the secondary frequency control loop. Meanwhile, an eligibility trace [42] in the area of RL theory is introduced to assign explicitly the credit or blame for the multi-step historic backward decision in the actor-critic learning process. The proposed $Q(\lambda)$ AGC has been successfully applied to the two-area LFC system model.

4.2 Multi-Step $Q(\lambda)$ Learning

Multi-step $Q(\lambda)$ learning is a model-free RL algorithm that extends the one-step Q-learning [70] by combining it with $TD(\lambda)$ returns for general λ [198] in an incremental way for the delayed RL control problems. The resulting hybrid RL algorithm thus combines some best characteristics of Q-learning and actor-critic learning paradigms. The $Q(\lambda)$ learning can serve as the basis of various multiple time scale learning mechanisms, which are essential for the application of RL to practical systems, for developing the discrete-time Markov chain control process

[71]. By estimating the corrections to predictions of state-action events occurring in the past, this algorithm propagates information incrementally to the important elements of Q-function. There are two different $Q(\lambda)$ algorithms proposed via the combination of eligibility traces with Q-learning, called Watkins's $Q(\lambda)$ [44] and Peng's $Q(\lambda)$ [73], respectively. Here, the latter $Q(\lambda)$ learning with a greedy policy, instead of the former with an exploratory action policy, is investigated.

The multi-step $Q(\lambda)$ learning, unlike one-step Q-learning algorithm in Section 3.2, can take advantage of the multi-step feedback information to update the Q-functions. The parameter λ in eligibility traces is used to assign the credits throughout sequences of the actions, lead to faster convergence and also alleviate the non-Markovian effects caused by the long time-delay and coarse state-action space quantization [73]. Eligibility trace [44] is a temporary record for the occurrence of taking of actions and state trajectory, and it marks the memory parameters associated with the decision-making events as eligible for interacting online learning. This trace for state s and action a at the k th iteration is indicated as $e_k(s,a)$. In each iterative step, eligibility traces for all the state-action pairs would decay exponentially by a factor of $(\gamma\lambda)^k$ while the one for the visited pair would also be incremented by 1, as follows,

$$e_{k+1}(s,a) = \begin{cases} \gamma\lambda e_k(s,a) + 1 & (s,a) = (s_k, a_k) \\ \gamma\lambda e_k(s,a) & \text{otherwise} \end{cases} \quad (4.1)$$

where parameter λ ($0 < \lambda < 1$) is referred as the trace-decay factor. Since both the state space S and action space A are finite, the values of eligibility trace and Q-function can be stored as matrices and implemented in two-dimensional lookup tabular forms indexed by the state-action pairs.

The $Q(\lambda)$ learning makes use of $TD(\lambda)$ returns as the value function estimator through a trace mechanism, and additional traces would help to bridge the gap between frequency and recency information for the heuristic events. The purpose of the $Q(\lambda)$ learner is to pursue the optimum policies by maximizing cumulative discounted rewards over the entire optimization period, and the estimation for the

current value function errors at the k th iteration can be computed in Eq. (4.2) and Eq. (4.3), as follows,

$$\rho_k = r(s_k, s_{k+1}, a_k) + \gamma Q_k(s_{k+1}, a_g) - Q_k(s_k, a_k) \quad (4.2)$$

$$\delta_k = r(s_k, s_{k+1}, a_k) + \gamma Q_k(s_{k+1}, a_g) - Q_k(s_k, a_g) \quad (4.3)$$

where ρ_k can be viewed as the one step Q-function error while δ_k as the TD(0) value function error for the Q-function at the k th iteration; $r(s_k, s_{k+1}, a_k)$ denotes the reward function of transition from state s_k to s_{k+1} under action a_k ; In the next searches, Q_{k+1} can make a useful iterative update based on the action taken and reward received as the following iterative rules,

$$Q_{k+1}(s, a) = Q_k(s, a) + \alpha \delta_k e_{k+1}(s, a) \quad (4.4)$$

$$Q_{k+1}(s_k, a_k) = Q_{k+1}(s_k, a_k) + \alpha \rho_k \quad (4.5)$$

It has been demonstrated [73] that the overall performance of $Q(\lambda)$ learning exhibits less sensitivity to the choice of the training parameters and more robust behavior than the standard Q-learning. Both Markov and non-Markov tasks have been carried out to validate the efficacy of this algorithm. Rigorous proof in [194] shows that the estimated Q-function converges to the optimal $Q^*(s, a)$ values with the probability 1 once the action values are represented discretely and all actions are sufficiently sampled in state space.

The backward estimation mechanism of eligibility traces plays an important role for the backtracking process of the multi-step $Q(\lambda)$ learning. The backward estimation with its additional memory, eligibility trace, can proportionally update the global Q-function errors to all recently visited states. For each iterative step, value function errors can be calculated and assigned backward to each previous state according to the states' eligibility traces, and the updating policies depend on the current Q-function error combined with eligibility traces of the past events. All in all, the backward estimation is an incremental mechanism to assign the credits backward to the previously visited “the stream of states”, as illustrated in Fig. 4.1 [44].

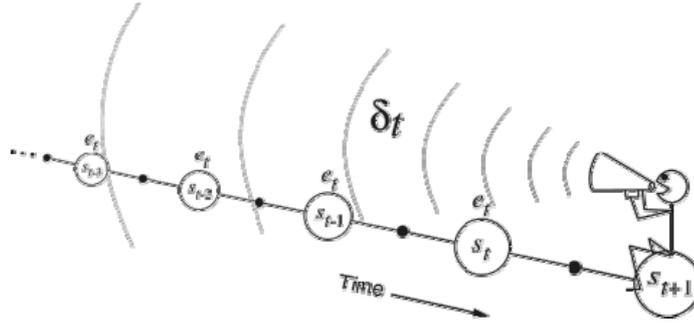


Fig. 4.1 The stream of states for backward estimation

4.3 Thermal-Dominated AGC Process under CPS

Thermal power plants have played a dominant role in most AGC systems in China. The Guangdong power grid in the CSG system is a good example. AGC strategy arisen from the power dispatch center will be implemented via the fuel supply in the turbine-boiler control system of thermal plants, and will therefore be executed slowly with large time constants [48]. The typical time-delay in the secondary frequency regulation ranges from 0.5 to 2 minutes and would vary with the ramping direction. Since a certain amount of the memory effect for the transition trajectory would be resulted in the CPS state space, the past states and decisions will have after-effect influence on the transition probability distribution of future behaviors in the thermal-dominated AGC process. Consequently, this LFC process for interconnected power systems is, actually, a non-Markov chain (non-Markov environment) [72]. Previously, the AGC decision-making problem was viewed as a Markov chain control process and was solved using the standard Q-learning [40], as stated in Chapter III. However, this approach would not be sufficient to cope with the modern AGC with CPS because of its non-Markov property. Furthermore, the long time-delay secondary frequency control link in thermal units should also be accounted in the AGC scheme.

This chapter focuses on applying the $Q(\lambda)$ learning with eligibility traces to cope with the thermal-dominated AGC strategy in the non-Markov environment.

The multi-step backtracking learning takes full account of the previously visited state-action pairs and the temporal sequence in the stochastic process. It could be effective for alleviating the non-Markovian effects of time-delay control loop and coarse state-space quantization. Eligibility trace, which is a special mechanism of temporal reward assignment for value function errors, can also offer a backward estimation approach to explicitly allocate the credits or blames backward to prior states according to the eligibility trace at that iterative moment. In the backward estimation of $Q(\lambda)$ learning, as shown in Fig. 4.2, the global value function errors will trigger proportional updates to all the recently visited “stream of state-action pairs” as signaled by their nonzero traces.

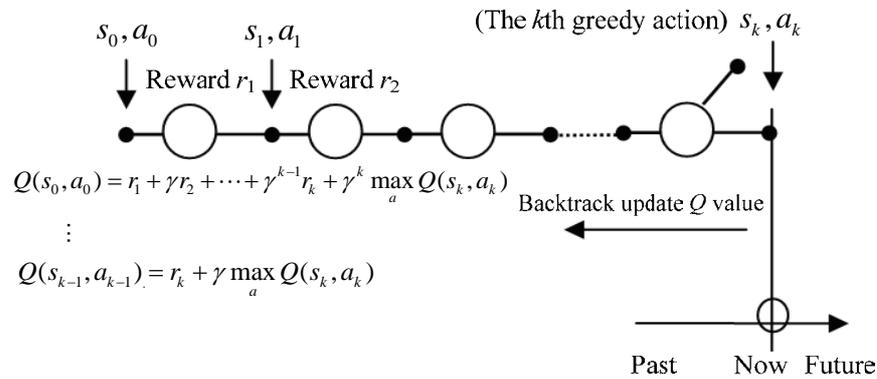


Fig. 4.2 Backward estimation for $Q(\lambda)$ backtracking learning

NARI’s control strategy, as adopted in its EMS system, is to deliver an AGC regulating command every 4 seconds in a discrete manner. After the separation of the generation sectors from electricity grids in the deregulation of the Chinese power systems in 2002 [199], genuine AGC strategies in provincial power grids would give priority to promoting CPS metrics and, as a result, may lead to an “over-compliant” problem that high CPS compliance is blindly pursued without considering the maneuvering costs of AGC plants. Measurements from dispatch centers of CSG power system show that there are frequent tiny reversal actions in AGC commands for some generators and it is harmful to the units because of the mechanical wear-and-tear [41]. Moreover, the increments in the AGC regulating

intensity for thermal plants would result in the evident rise in fuel consumption as well as the decrease in generation benefits of the interconnected operation. In addition, detailed analysis and comparison for extra energy losses and potential economic aspects of thermal plants involved in AGC implementation have been reported in [200]. Hence, a multi-step $Q(\lambda)$ learning based stochastic optimal relaxed AGC strategy is proposed here for thermal-dominated power systems.

4.4 Design of Optimal Relaxed AGC Strategy

Following the analytical framework described above, the thermal-dominated AGC problem under CPS can be transformed to a dynamic non-Markov decision process. Coupled with the fast random fluctuation in the load demand and some uncertain parameters existing in the system, the problem of optimal AGC relaxed control becomes even more complicated and heretofore no satisfactory solution method has been reported publicly. The $Q(\lambda)$ learning not only can provide AGC controller with capability of self-learning for the discounted infinite time-horizon optimization in non-Markov domains by means of undergoing the backtracking consequences of actions, but also would be flexible in accommodating various control actions and control objectives for solving this challenging multi-criteria optimization problem [43].

4.4.1 State and Action Space Discretization

In the following state and action space analysis, a two-class classification is assumed. While the state space S comprises all the acceptable operating points not violating any system constraints, action space A combines the set of discrete values of control adjustments. The degree of state and action space discretization plays an important role to the AGC performance of $Q(\lambda)$ AGCs. By means of space discretization, the state-action space can be divided into a finite number of regions, called the state-action pairs. Though a large number of state-action pairs

could give a high-resolution solution, the convergence rate of $Q(\lambda)$ learning would be lowered owing to the curse of dimensionality of Q -functions, and as a result, excessive AGC commands would occur and go against the relaxed control objective. On the other hand, poor AGC performance would result if the number of state-action pairs was too small.

Both CPS1 and CPS2 are long-term statistical evaluation criteria, which can be assessed in the form of the clock-minute averages. For the convenience and timeliness of control and input filtering, 1-min moving averages of CPS1/ACE [8], which can be obtained from “ACE/ ΔF /CPS real-time monitoring database” [40], are used as the binary state signals for the $Q(\lambda)$ AGC. CPS compliance data are recorded and deposited in “long-term historical database”. Hence, the CPS control space, as shown in Fig. 3.2, can be transformed as a discrete-time stochastic state space constituted by 1-min moving averages of CPS1 and ACE mathematically. According to CPS assessment principle [41] and relaxed control characteristics, the state space discretization can be designed following the QAGC in Section 3.3.2.

One of the main advantages of $Q(\lambda)$ learning is the flexibility in specifying an AGC control action vector. The output variable of the $Q(\lambda)$ AGC is the optimized AGC regulating command $\Delta P_{\text{ord-}i}$ which is the variation in the generator setpoint within an AGC cycle. The discretization of action space A should be carried out based on the system spinning reserve capacity and unit regulating characteristics for each control area, and the specific action vector will be presented in the case studies in Section 4.5.

4.4.2 Reward Function

The control objective of the $Q(\lambda)$ AGC can be defined by the discount factor γ and immediate reinforcement/reward function $r(k)$. Since the reward function in $Q(\lambda)$ AGC will be used to assess the AGC performance, it shall consider both the CPS performance and relaxed control objectives in order to avoid the “over-

compliant” problem existing in CPS control strategies and minimize the control pressure imposed on the AGC generators. The proposed control for relaxation of AGC system can be implemented with high feasibility based on the previous researches on CPS for the CSG [195]. Consequently, the desirable relaxed AGC control shall give higher priority to lessening the number of maneuvering and reversal actions for AGC units according to the CPS and the stochastic sequential decision emerged in the relaxed control. Here, a new relaxed reward function $r_i(k)$ for control area i is defined by modifying the multi-criteria piecewise function Eq. (3.6), as below,

$$\left\{ \begin{array}{l} r_i(k) = \sigma_i - \mu_{1i} [a_{\text{ord-}i}(k) - a_{\text{ord-}i}^*]^2 - \nu_{1i} N_R, \quad CPS1_{\text{mov-}i}(k) \geq 200 \\ r_i(k) = -\{\eta_{1i} [\max(|ACE_{\text{mov-}i}(k)| - ACE_i^*, 0)]^2 \\ \quad + \mu_{1i} [a_{\text{ord-}i}(k) - a_{\text{ord-}i}^*]^2 + \nu_{1i} N_R \}, \quad CPS1_{\text{mov-}i}(k) \in [100, 200) \\ r_i(k) = -\{\eta_{2i} [CPS1_{\text{mov-}i}(k) - CPS1_i^*]^2 \\ \quad + \mu_{2i} [a_{\text{ord-}i}(k) - a_{\text{ord-}i}^*]^2 + \nu_{2i} N_R \}, \quad CPS1_{\text{mov-}i}(k) < 100 \end{array} \right. \quad (4.6)$$

where σ_i is an arbitrary non-negative number which is set to 0 in the case study; $ACE_{\text{mov-}i}(k)$ and $CPS1_{\text{mov-}i}(k)$ are the 1-min moving averages of ACE and CPS1 at the k th iterative time in area i , respectively. As for the $CPS1_i^*$, experiences show that a value of 200 works well if high CPS compliance is required since CPS1 would be the dominated factor in the CPS compliance when it is high; vice versa, if the $CPS1_i^*$ is too small, say 100, CPS compliance would be dominated by CPS2 and become too low. As a compromise for the relaxed control, the historical daily or monthly mean of CPS1 metric in area i is adopted here as the $CPS1_i^*$ for area i . The clock-minute averages of ACE can be used to represent the control area’s CPS2 condition. The value of the ACE_i^* , in our application to the CSG power system, is used to specify the threshold value of the AGC dead-zone (5MW) so as to prevent real-time ACE from crossing zero frequently and reduce the inadvertent power exchanges. Consequently, the received reward will be kept constant for the values of $ACE_{\text{mov-}i}(k)$ within the dead-zone. $a_{\text{ord-}i}(k)$ is the index to a selected action from action vector A at the k th iteration, while $a_{\text{ord-}i}^*$ denotes

the index of the null control action in the action vector. The use of the quadratic term of action variation in Eq. (4.6) is to lower wear-and-tear of AGC generators and economic cost resulting from large fluctuation in AGC control commands. N_R represents a pulse identifier of control signals, and it is equal to 1 or 0 to indicate whether the control signal at the k th iteration is a reversal pulse or not, respectively. In addition, the weights η_{1i} , η_{2i} , μ_{1i} , μ_{2i} and v_{1i} , v_{2i} are equivalent to the parameters of weight matrices in LQR methodology [196].

Since CPS compliance will be satisfied in our application to the CSG system when “CPS1 \geq 200%”, the reward is assigned to the maximum value if “CPS1 \geq 200%”. When “100% \leq CPS1 $<$ 200%”, CPS compliance is determined by CPS2 metric, and thus the reward should be assigned based on the averages of ACE. When “CPS1 $<$ 100%”, CPS compliance fails and CPS1 index is the predominant factor for the CPS assessment. Therefore, the reward for the control area with a lower CPS1 value should be assigned with a smaller value. Furthermore, in order to ensure the conformance and fairness in the piecewise reward function Eq. (4.6), the weight ratios η_{1i}/η_{2i} , μ_{1i}/μ_{2i} and v_{1i}/v_{2i} shall be kept constant while their exact values depend on the frequency bias B_i and CPS1 control target ε_1 in control area i due to different dimensions between state variables CPS1 and ACE. Moreover, weights μ_{1i} , μ_{2i} are critical elements for the relaxed control performance of the $Q(\lambda)$ AGCs, which is named as the relaxation factor here, and the relaxation vector is $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_i, \dots, \mu_m]$ for an interconnected power systems within m control areas.

4.4.3 $Q(\lambda)$ Learning Parameters

The learning step T_{step} is determined by AGC decision cycle time. In addition to the above algorithm settings, three parameters α , γ and λ in Eq. (4.1)~(4.5) are crucial in implementing the $Q(\lambda)$ learning and shall be set with following generic guidelines in [44, 73, 193].

The learning factor α is the step size of learning which essentially determines the amount of update in the Q-function [194]. A larger α (close to 1) allows fast convergence of the algorithm, while a smaller α (close to 0) tends to enhance the stability of the Q(λ)AGC. For the RL enforced AGC strategy which depends on previous control steps, α shall be very small owing to the time-lag control loop and the randomness of load disturbance. Simulation studies show that a value in the range of 0.001~0.1 is acceptable. Here, the factor is set to 0.01.

The discount rate γ is the control factor by which later rewards are discounted in control objective. Since later rewards in the AGC process under CPS are very important, a value close to 1 shall be taken. Experiences show that a value in the range from 0.6 to 0.98 works well. Here, an intermediate value of 0.9 is used.

The trace-decay factor λ in eligibility traces is used for allocating the rewards throughout sequences of actions, leading to faster self-learning and also help to overcome the non-Markov effects in the thermal-dominated AGC problem. For larger values of λ , traces of the preceding events would decay more slowly, and more of farther backward operation information can be utilized to optimize the Q-functions. Conversely, the prior state-action pairs are given less credit for the Q-function error if λ is set to a small value. In general, λ can be considered as a time scale factor in the backtracking process, and it is a trade-off between bias and variance [73]. For the AGC problem in non-Markov environment, the value shall be chosen close to 1 and the policy remains unaffected with λ between 0.85 and 0.995 in the simulations. With respect to the proposed Q(λ)AGC, a value of 0.9 is selected for factor λ .

4.4.4 Implementation Procedures of Q(λ)AGC

The implementation steps of the proposed optimal relaxed AGC algorithm are summarized in Fig. 4.3.

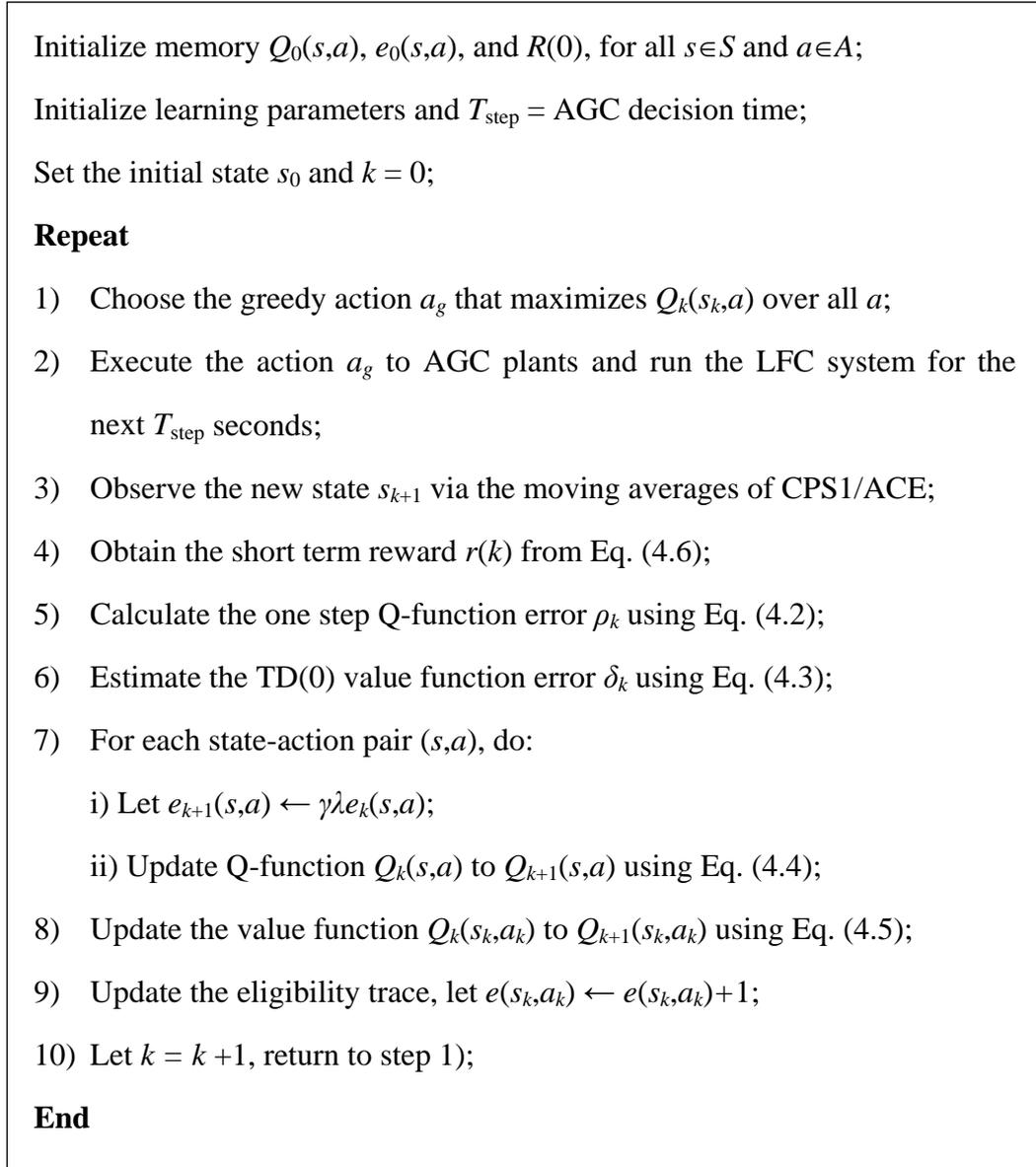


Fig. 4.3 $Q(\lambda)$ learning algorithm for AGC relaxed control

4.5 Simulation Results and Analysis

The performance of the developed optimal relaxed AGC strategy has been evaluated through the following case studies. The two test systems adopted for studying the $Q(\lambda)$ AGC are a small two-area thermal-dominated LFC power system with long secondary time delay and the practical CSG power systems with four control areas.

4.5.1 Study on Two-Area LFC Power System Model

Fig. 4.4 shows the block schematic diagram of a typical thermal-dominated LFC power system with the model parameters taken from [42]. While the dynamic simulation of the LFC system is established using Matlab/Simulink, the $Q(\lambda)$ AGC is implemented as S-Function module. For clear illustration of the execution steps involved, the following settings are adopted in this case study. As the AGC decision cycle is set to 3 seconds and the time-delay T_s of thermal unit is set to 20 seconds, action vector A is discretized into fifteen values equal to $\{-500, -300, -100, -50, -20, -10, -5, 0, 5, 10, 20, 50, 100, 300, 500\}$ MW.

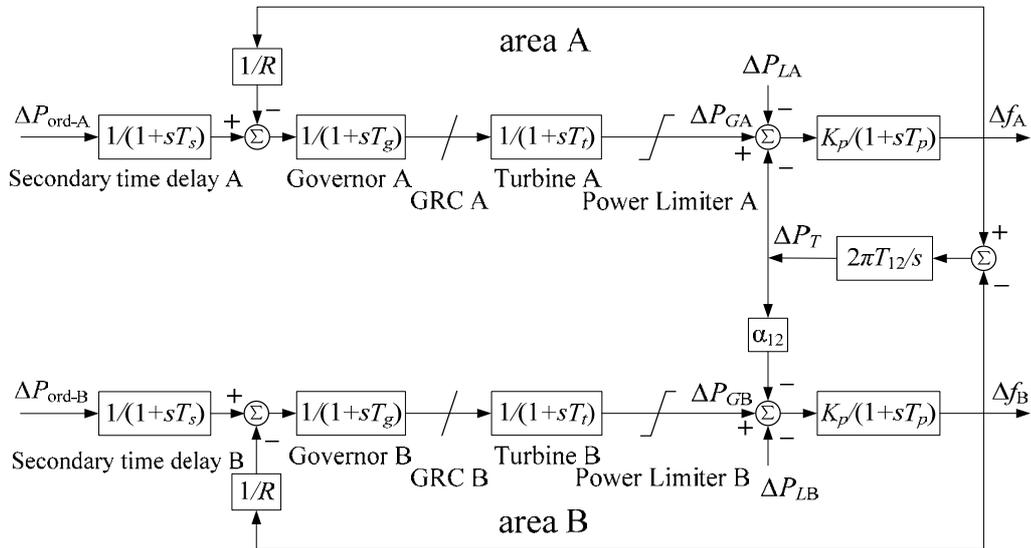
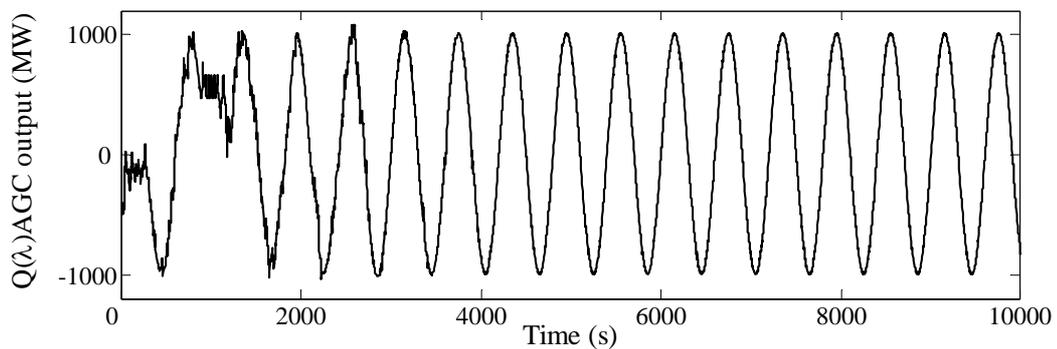


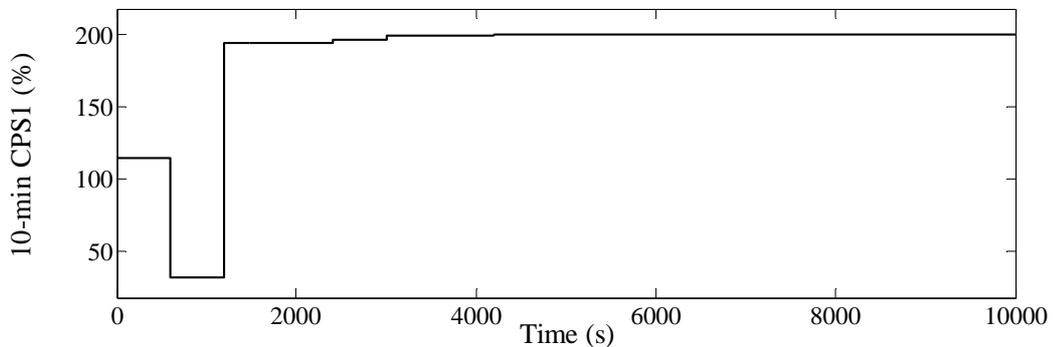
Fig. 4.4 Two-area thermal-dominated LFC model

Normally, no experiences will be available in the initial stage of interactive learning for the RL driven controller to control the AGC system adequately. As a remedy, $Q(\lambda)$ AGCs should be scheduled to experience a series of trial-and-error procedures called “pre-learning process”, which is a stochastic action exploration process in the CPS state space to consummate the optimal tabular form $Q^*(s,a)$, before its practical onsite application. There are two preconditioning techniques for RL methods discussed and compared in [74]: on-line mode and off-line mode. The on-line mode, in which the interaction occurs with the real power system, is

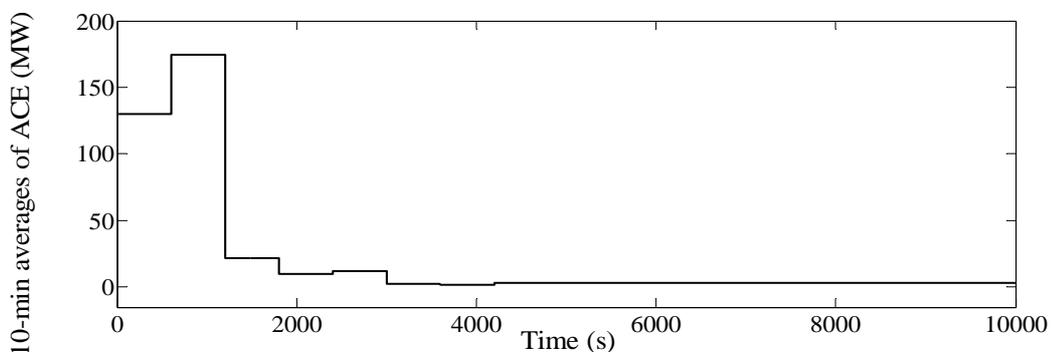
particularly interesting when it is difficult to model the system or when some scenarios are difficult to reproduce in the simulation model. However, the main drawback of this on-line learning mode is that the RL controller may jeopardize security and stability of the real system. One effective solution is to employ the off-line learning mode by using the $Q(\lambda)$ AGC in a simulation environment first. In this mode, the controller can extract the offline learned control policy from a high-accuracy simulation system and then implement it on the real system.



(a) $Q(\lambda)$ AGC output $\Delta P_{\text{ord-A}}$ in pre-learning process



(b) The plot of 10-min CPS1 metric in pre-learning process



(c) The plot of clock-10-min $|E_{\text{AVE-min}}|$ in pre-learning process

Fig. 4.5 Pre-learning process of the $Q(\lambda)$ AGC in area A

The typical pre-learning process corresponding to a group of 10-min periodic sine-wave load disturbances in area A is given in Fig. 4.5. It can be found that the action outputs at various states in Q-functions converge to their deterministic optimum control strategy after about 3000 iterations while the CPS1 (in 10-min assessment period) and $|E_{AVE-min}|$ (clock-10-min absolute averages of ACE) tend to become stable as shown in Fig. 4.5(b) and 4.5(c). This pre-learning process should be repeated continuously for a large number of operating states with different typical load disturbances until there are no more changes in the Q-functions, and the optimal AGC strategy can be gradually learned. Moreover, the iterative termination criterion for pre-learning process is determined as the matrix 2-norms $\|Q_{k+1}(s,a)-Q_k(s,a)\|_2 \leq \zeta$ where ζ is a given small precision factor. Then, all parameters and memory such as Q-function matrices and eligibility traces should be stored after the pre-learning process in each area is completed, and $Q(\lambda)$ AGCs can be put into normal operation for the practical implementation of AGC system thereafter. Furthermore, the proposed $Q(\lambda)$ AGCs will continue to learn from the on-line system and improve its control strategy and behaviors by interacting with the real power systems. Also, unlike the QAGC, the pre-learning phase of $Q(\lambda)$ AGC does not need the semisupervisory controller as in Section 3.4.

Table 4.1 Trace-decay factor effects on $Q(\lambda)$ AGC

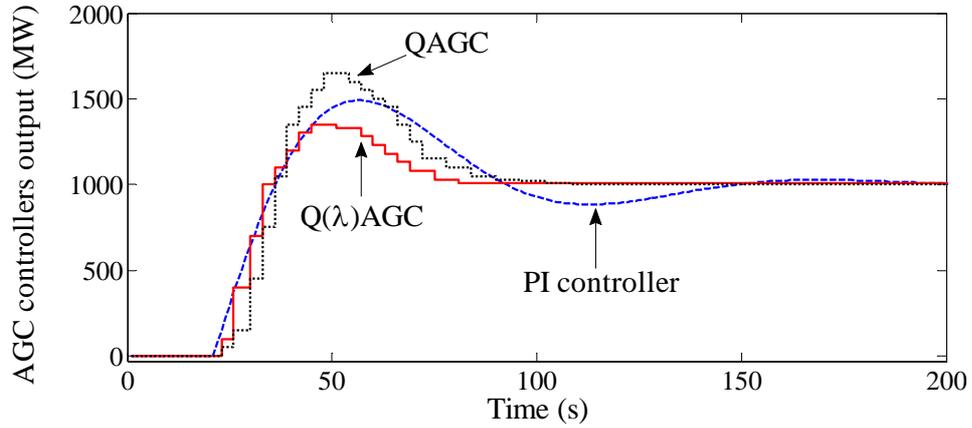
λ	0.7	0.8	0.85	0.9	0.95	0.99
T_c (s)	21629	13146	9285	8261	7927	7813

Table 4.2 Discount rate effects on $Q(\lambda)$ AGC

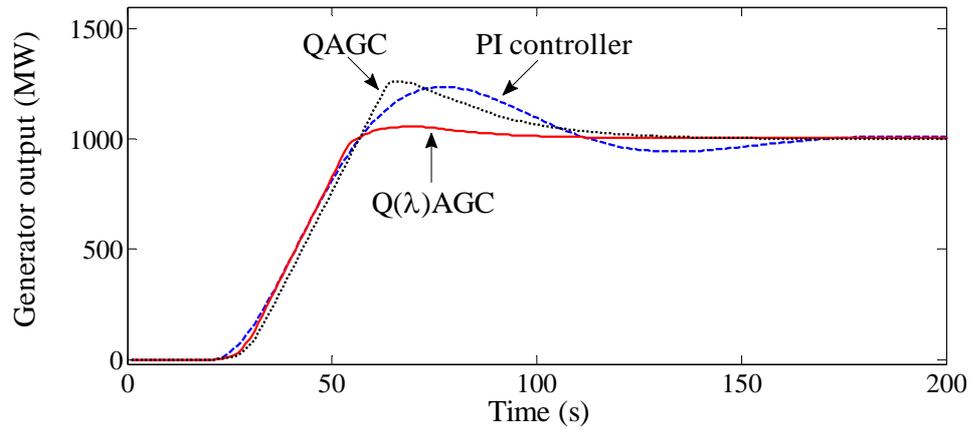
γ	0.5	0.6	0.7	0.8	0.9	0.95
T_c (s)	15758	10630	9347	8565	8261	8242

Since the trace-decay factor λ and discount rate γ determine the record and backtracking mechanism of eligibility traces, they would have crucial impacts on the pre-learning process of the non-Markov delayed control problems. Table 4.1 and 4.2 illustrate the effects of those two parameters on the convergence property

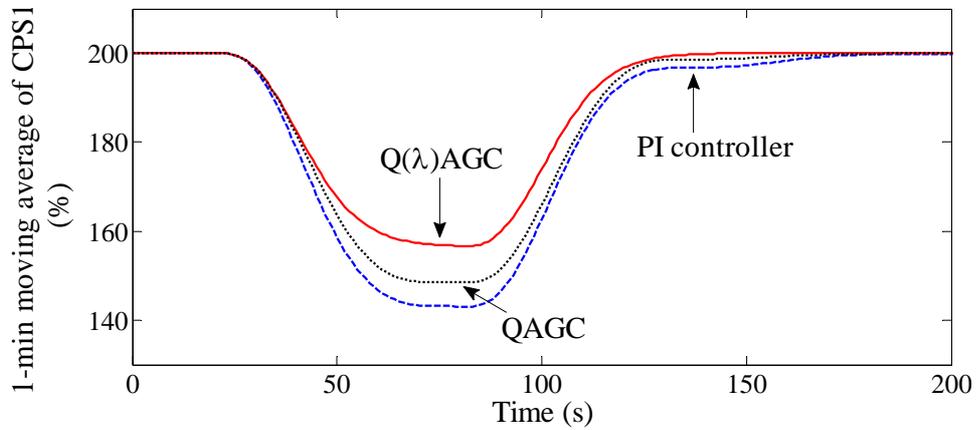
of the learning algorithm. T_c represents the average convergence time under the periodic sine-wave load fluctuations as in Fig. 4.5. Simulation studies indicate that large values of λ and γ should be set to cope with the delayed consequences of actions and rewards.



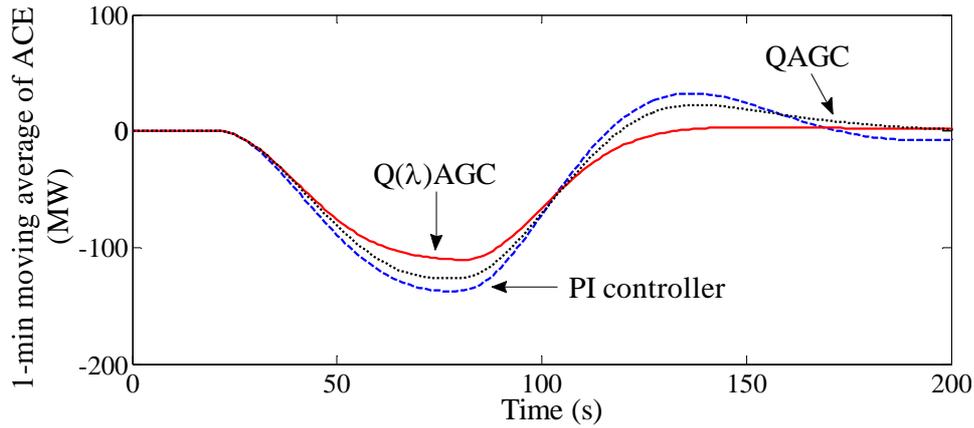
(a) The plot of AGC power regulating commands $\Delta P_{\text{ord-A}}$



(b) The plot of AGC generating unit output ΔP_{GA}



(c) The curve of 1-min moving average of CPS1_A



(d) The curve of 1-min moving average of ACE_A

Fig. 4.6 Simulation experiment for AGC controllers

The dynamic performance of $Q(\lambda)$ AGCs for a step load disturbance in area A is shown in Fig. 4.6. The simulation results are compared with the well-designed PI controller in [11] and one-step QAGC in Chapter III. This step response can provide information on the stability and dynamic performance for a closed-loop control system [201]. Here, the reward functions of AGC controllers in two areas are identical, and the weight factors in Eq. (4.6) for the $Q(\lambda)$ AGC and the QAGC are set as follows after numerous simulation studies: $\eta_1=1$, $\eta_2=50$, $\mu_1=\mu_2=10$, $\nu_1=\nu_2=0$. Fig. 4.6(a) and 4.6(b) illustrate the plots of AGC regulating commands and the dynamic response of an AGC unit. In general, the smoother the regulating commands delivered to the AGC generators, the higher the energy utilization of those generators will be. Thus, the $Q(\lambda)$ AGC can provide the better relaxation properties for AGC thermal plants. Consequently, less control actions, less wear-and-tear, and further savings on fuel costs would be expected. On the other hand, the experiment suggests that $Q(\lambda)$ learning can distribute backward information rapidly by allowing Q-function corrections to be made incrementally for the prediction of past observations, and hence the $Q(\lambda)$ AGC works notably satisfactorily on the dynamic behaviors such as settling time, overshoot, etc.

Thermal generating units, such as LNG and coal-fired units, have a wide variety of regulating characteristics for the secondary frequency response. For

further illustration of AGC behaviors of the $Q(\lambda)$ AGCs in different non-Markov environment, taking control area A in Fig. 4.4 as the studied system, Table 4.3 shows the performance results of $Q(\lambda)$ AGC for thermal units with a set of typical AGC time-delay parameters and generation rate constraint (GRC) [202]. Here, a group of band-limited white noise fluctuations is used as the load disturbances for the performance comparison after the pre-learning process. $|\Delta F|$ and CPS1 are average indices over a 24-hour period, while $|E_{AVE-min}|$ (refer to CPS2) in Table 4.3 represents the average of absolute values of $E_{AVE-min}$ over the period. As the convergence time of $Q(\lambda)$ AGC in pre-learning process goes up remarkably with the increase of AGC time-delay T_s and GRC, it demonstrates that the controller requires more training iterations to pursue optimal policy under the condition of long time-lag control loop. In conclusion, as the delay time and generation rate get worse, the AGC metrics would only deteriorate moderately, and this shows the ability and efficacy of $Q(\lambda)$ AGCs for overcoming the non-Markov effects.

Table 4.3 AGC performance test for different time-delay and GRC

Thermal Unit	T_s (s)	GRC (p.u./min)	T_c (s)	$ \Delta F $ (Hz)	$ E_{AVE-min} $ (MW)	CPS1 (%)
Unit 1	8	10%	7650	0.0113	22.47	197.69
Unit 2	20	10%	8261	0.0138	27.76	196.35
Unit 3	20	8%	9084	0.0199	34.18	194.30
Unit 4	30	8%	10553	0.0217	35.21	192.81
Unit 5	30	5%	12882	0.0231	36.95	191.63
Unit 6	45	5%	16035	0.0252	39.13	189.15
Unit 7	45	3%	21196	0.0285	43.65	187.07
Unit 8	60	3%	28269	0.0328	48.39	183.36

4.5.2 Study on China Southern Power Grid

For the in-depth analysis of the proposed optimal relaxed AGC strategy in a realistic simulation environment, a detailed CSG power system model previously developed using Matlab/Simulink for Guangdong power dispatch center projects

[41] was used as the benchmark power system in this thesis. The CSG power system has four control areas interconnected by parallel HVDC-HVAC transmission systems, as illustrated in Fig. 4.7. In this AGC simulator, the plant models [197, 202] for fossil-fuel-fired, LNG and hydro generators are included in each control area, and each plant output is determined by the governor and set-point of AGC pulse from the ED function according to its participation factor. In addition, the high voltage direct current (HVDC) system is modeled as a first-order constant-power control model [203], and the AGC decision cycle time in CSG is set to 4 seconds.

In the CSG power system model, the AGC plants in each area are classified into different unit groups while each unit group is established using an equivalent generator model. The system LFC characteristics of the study area are available in [41], as depicted in Fig. 4.7 and further tabulated in Appendix A. Besides, the parameters used in CPS standards are the practical onsite values for CSG power system in 2009. The frequency bias factors B_i are set to -225 in Guangdong area, -35 in Guangxi, -37.5 in Yunnan and -40 in Guizhou. The threshold limit for the CPS2 assessment (L_{10}) is fixed to 288 in Guangdong, 75 in Guangxi, 78 in Yunnan, and 81 in Guizhou. The ε_{10} and ε_1 are 0.042 and 0.052, respectively.

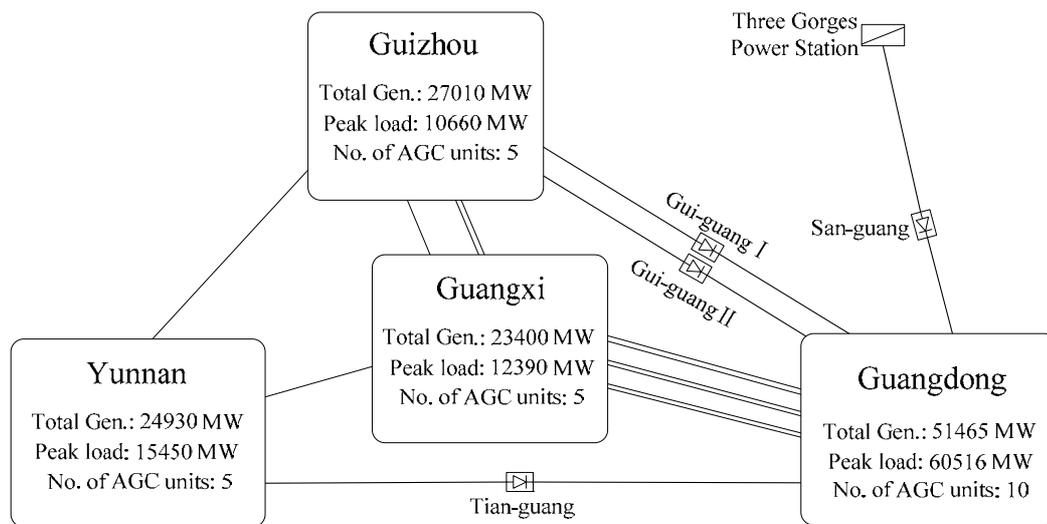


Fig. 4.7 The interconnected network of China Southern Power Grid

In this case study, since the magnitudes of CPS1 and ACE are different in different provincial control areas due to the different frequency bias factors, four different types of $Q(\lambda)$ AGCs are specifically designed for the four control areas based on their load characteristics and installed generation capacities. Table 4.4 gives typical weight factors for the reward function Eq. (4.6) of the $Q(\lambda)$ AGC in each control area. Furthermore, by analyzing the load disturbance and spinning reserve capacity in each power grid, the action discrete vector for the total AGC regulating command can be fixed as $A = \{-1000, -600, -300, -100, -50, -20, -10, -5, 0, 5, 10, 20, 50, 100, 300, 600, 1000\}$ MW. Also, the simulation results presented here correspond to the performance of AGC controllers after the pre-learning process.

Table 4.4 Weight factors for the $Q(\lambda)$ AGCs

Control area	η_1	η_2	μ_1	μ_2	ν_1	ν_2
Guangdong	1	38	30	30	20	20
Guangxi	1	3.5	7.6	7.6	0	0
Yunnan	1	5	8	8	0	0
Guizhou	1	6	8.5	8.5	0	0

Since the load disturbance in the AGC simulator is the forcing function, it is important to provide representative load data for this case study. The adaptability and robustness of the proposed AGC controller under a series of various complex load fluctuations and system parameter perturbations can be verified with white noise moderated samples to simulate synthesized disturbances and uncertainties in power systems. The white noise model is a random fluctuation with flat power spectral density. In this simulation study, a typical white-noise stochastic load [199] and white-noise parameter perturbation are chosen as a group of fixed disturbances for the comparison of the strategies, as follows:

- Applying the white-noise load model in each provincial power grid, in which the sample time is 10 minutes and the magnitude is less than 1500MW (this

value is equivalent to the value caused by the HVDC monopole block in Guangdong power grid);

- Adding the band-limited white noise perturbation, in which the sample time is 20 minutes and the magnitude is less than 10 percent of the nominal value, on natural frequency response coefficient β in each provincial power grid.

Here, Guangdong power grid is taken as the study area. Fig. 4.8 shows the time characteristics of frequency deviation, ACE, tie-line power error, and AGC generation in Guangdong area corresponding to the above system disturbances.

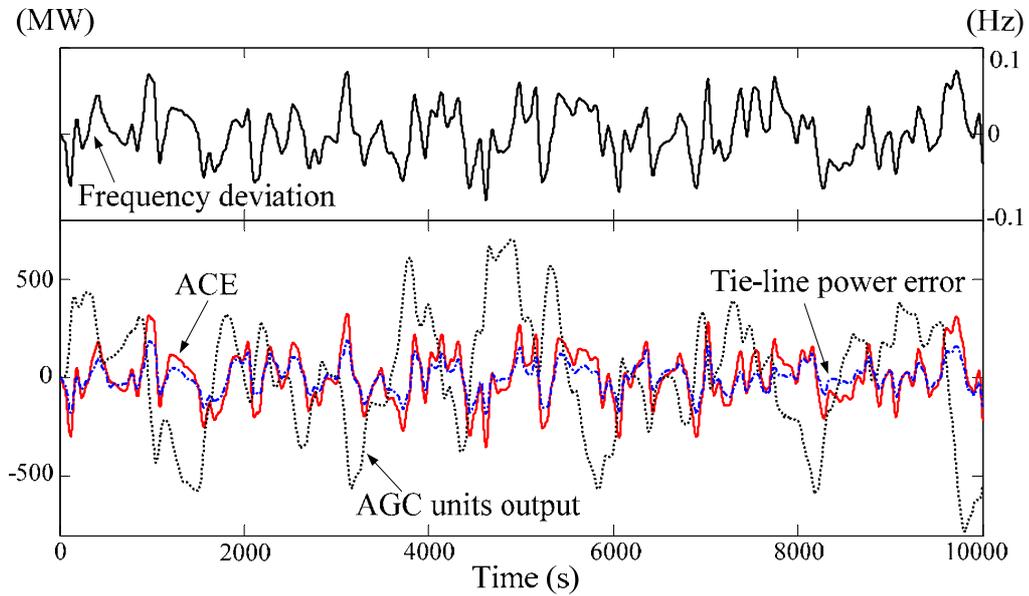
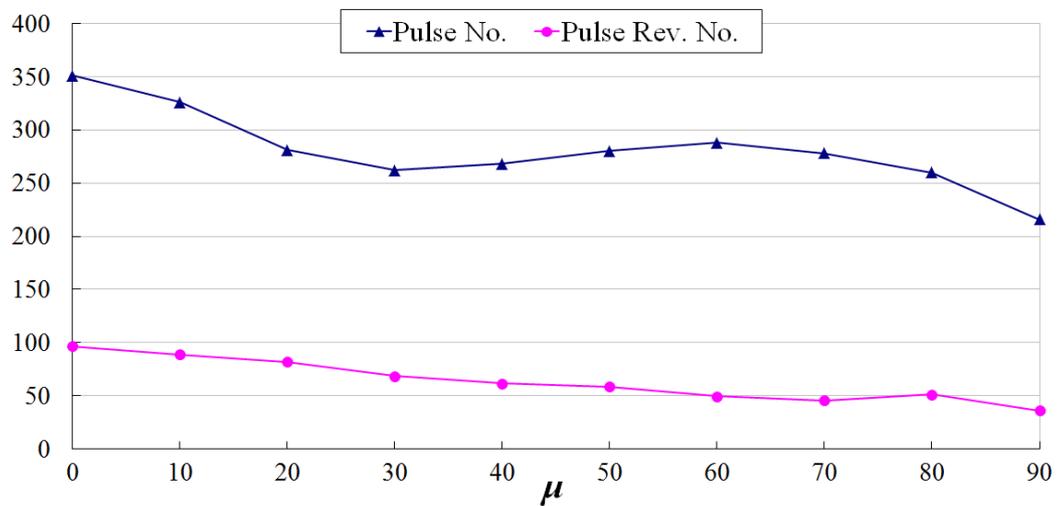


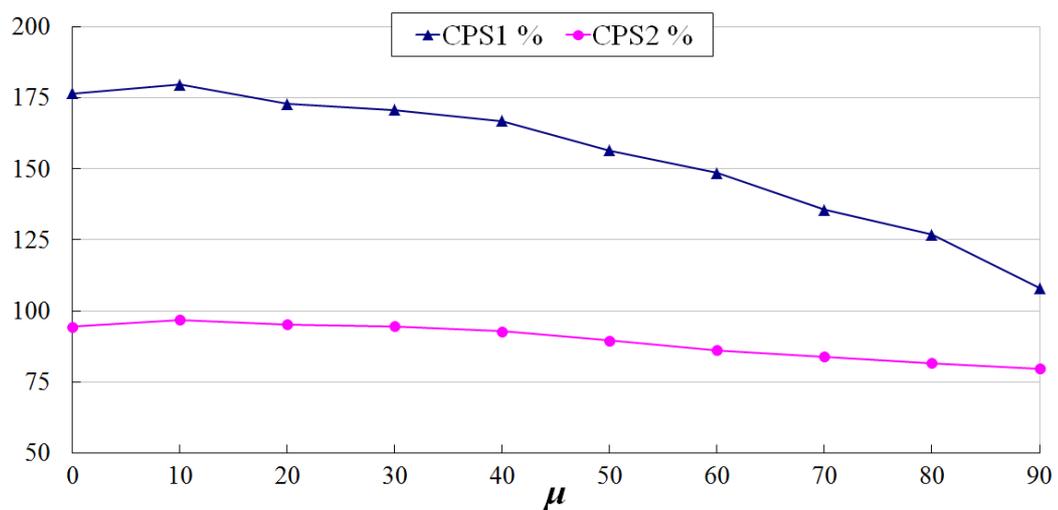
Fig. 4.8 Time characteristics of ΔF , ACE, ΔP_T and generators output

The AGC relaxed control performance can be accessed from the number of control pulses in the power regulating commands delivered to AGC units from the dispatch center. Here, Pulse No. is defined as the average number of pulses that are sent to each regulating unit per hour. Similarly, Pulse Rev. No. is defined as the average number of direction changes in pulses being sent to each AGC unit per hour. The relaxation factor in the reward function has been found to have a large influence on the performance of $Q(\lambda)$ AGC in a given control area. Fig. 4.9 illustrates the effects of relaxation factor μ on the AGC behaviors in the

Guangdong power grid under the load disturbances and parameter perturbations as described above. It can be revealed from Fig. 4.9 that as the relaxation factor μ (μ_1 is set to μ_2 in the case studies) increases, the relaxed control performances improve while the CPS1 and CPS2 compliance decrease. Furthermore, it is noted that Pulse No. begins to rise when μ increases beyond 40. In addition, there will be many small amplitude pulses in the total regulating command curve when μ is set to 60 or above. As a compromise between relaxed control and CPS metrics, the relaxation factor for this study area shall be set within the range of 10 to 40.



(a) Correlation between relaxed property and relaxation factor



(b) Correlation between CPS1, CPS2 and relaxation factor

Fig. 4.9 Relaxation factor effects on AGC performance

Based on the LQR theory and simulation test results, the proposed optimal relaxed control methodology provide a customized platform for online regulation of the control tightness so as to achieve a desired level of CPS compliance and relaxed performance. The AGC output would slow down along with the increase of relaxation factor μ , so that the regulating pressure and maneuvering cost of AGC plants would be released to achieve the “loosened control”. Conversely, the $Q(\lambda)$ AGC would tend toward the “tightened control” for the overstrict and high CPS compliance requirement. Consequently, the relaxation vector μ should be thoughtfully selected for the coordination and balance between the CPS control objective and AGC relaxed objective in terms of the above-mentioned laws and the requirements of actual operating condition. The proposed $Q(\lambda)$ AGC can also employ a performance feedback mechanism which helps to adjust the relaxation vector or control tightness based upon historical AGC performance data.

Moreover, the long-term AGC performance can be evaluated with the data statistical comparative experiments. In the experiments, the AGC simulator of CSG system was started in a quiescent state where frequency and tie-line flow were at their scheduled values. Simulations were then implemented with the preset disturbance conditions over a 30-day period. Table 4.5 and 4.6 show the AGC performance metrics for the study area Guangdong power system with the assessment period 10 minutes, in which $|\Delta F|$ and $|ACE|$ are the averages of the absolute values of frequency deviation and ACE over the entire period, and CPS1, CPS2 and CPS are the monthly compliance percentages. Furthermore, the following four types of AGC controllers were considered and implemented for the comparison and discussion.

- NARI’s improved-PI control strategy
- QAGC: $\eta_1=1, \eta_2=38, \mu_1=\mu_2=30, v_1=v_2=0$
- $Q(\lambda)$ AGC-I: $\eta_1=1, \eta_2=38, \mu_1=\mu_2=30, v_1=v_2=0$
- $Q(\lambda)$ AGC-II: $\eta_1=1, \eta_2=38, \mu_1=\mu_2=30, v_1=v_2=20$

Table 4.5 and 4.6 provide the overall comparisons among AGC strategies in Guangdong power grid with the increase of the white noise perturbations in the coefficient β from 0% (nominal parameters) to 10% of its nominal value. Then, highlights from the results are summarized as follows:

- (i) The $Q(\lambda)$ AGC performs very well and compares well with the performance of a properly tuned improved-PI controller even in the case with nominal parameters. When the parameters were perturbed, the performances of $Q(\lambda)$ AGC could still be maintained and become significant superior to the improved-PI controller on CPS compliances and relaxed metrics. Since the proposed AGC would continue to learn the optimized closed-loop control laws online, it is able to adapt well to the changing operating conditions [45] and hence has a superior adaptability and dynamic optimization capability than NARI's PI controller.
- (ii) The main advantage of the $Q(\lambda)$ learning is that its multi-step information updating mechanism as well as backward estimation approach can enhance the efficiency of online learning. As the optimal Q-function estimator takes into account the influence of the previous multi-step decision-making policy, it can make the dispatching commands more reasonable and credible with the help of its prediction capability. As shown in Table 4.5 and 4.6, $Q(\lambda)$ backtracking learning can provide better control effects than the standard Q-learning control when the system is subjected to a series of complicated system disturbances.
- (iii) The addition of the pulse identifier N_R to the reward function in $Q(\lambda)$ AGC-II would reduce the Pulse Rev. No. to the AGC plants and hence improve the relaxed property. Since any improvement on the relaxed performance would inevitably lead to deterioration in the CPS metrics, the weight factors v_1 and v_2 should not be set so large that CPS1 and CPS2 performances fall below minimum levels necessary for compliances. Besides, the two-dimensional state variables (1-min moving averages of CPS1 and ACE) are used as input

signals for $Q(\lambda)$ AGCs to ensure the timeliness and feasibility for the AGC system under CPS. Thus, information on the operation sequences within the preceding 1 minute will be embedded in the reward function for strategic decisions and contribute to the relaxation of frequency control as well as alleviate the maneuvering cost for AGC generators. This reduction in the maneuvering of plants output would give opportunities for AGC generating units to operate more efficiently with the significant decrease in unnecessary mechanical wear-and-tear and saving in fuel costs.

Table 4.5 Statistical experiment results of Guangdong power grid I

Metrics	Nominal parameters			
	NARI's PI	QAGC	$Q(\lambda)$ AGC-I	$Q(\lambda)$ AGC-II
$ \Delta F $ (Hz)	0.0282	0.0271	0.0273	0.0289
$ \text{ACE} $ (MW)	151.89	146.18	148.63	157.95
CPS1 (%)	183.33	187.35	185.67	182.24
CPS2 (%)	97.85%	98.55%	98.50%	97.65%
CPS (%)	94.51%	95.51%	95.29%	94.28%
Pulse No.	294	267	251	239
Pulse Rev. No.	78	67.8	63.3	60.9

Table 4.6 Statistical experiment results of Guangdong power grid II

Metrics	White noise parameter perturbations			
	NARI's PI	QAGC	$Q(\lambda)$ AGC-I	$Q(\lambda)$ AGC-II
$ \Delta F $ (Hz)	0.0561	0.0412	0.0365	0.0382
$ \text{ACE} $ (MW)	233.17	213.03	185.75	196.18
CPS1 (%)	142.05	163.59	176.03	170.69
CPS2 (%)	90.83%	94.11%	95.89%	94.81%
CPS (%)	85.79%	90.48%	93.76%	92.16%
Pulse No.	338	303	271	262
Pulse Rev. No.	91.5	80.4	73.5	68.3

4.6 Summary

The thermal-dominated AGC control process can be formulated as a dynamic decision-making optimization problem. In this chapter, a novel optimal relaxed AGC control in non-Markov environment based on $Q(\lambda)$ learning is developed and successfully implemented and tested on a small two-area power system and the China Southern Grid power systems. The following are the main advantages of the proposed approach:

- (1) The application of multi-step $TD(\lambda)$ return mechanism as the value function estimator in $Q(\lambda)$ AGC can effectively resolve the delayed feedback problem with non-Markov effects caused by long time-delay control loop in thermal plants. Since CPS pay more attention to the long-run statistical returns of AGC performance, it is noteworthy that the $Q(\lambda)$ learning is an incremental iterative algorithm to maximize the cumulative discounted reward from the viewpoint of long-term optimal objective as well.
- (2) The $Q(\lambda)$ AGC can cope with the partial information, nonlinear effects and stochastic behaviors in power systems [45], especially when the real system is facing with scenarios which have not been previously encountered in the simulations. As supported by the statistical comparative experiments, the robustness and adaptability of the studied AGC system are enhanced and the degree of CPS compliance is improved evidently under various complex changing operation conditions.
- (3) The CPS compliance and AGC relaxed control objectives are embodied in the multi-criteria reward function via the linear weighted aggregate method. Simulation studies showed that the proposed $Q(\lambda)$ AGC can realize significant reduction in the generator movements for thermal plants. For example, with respect to the NARI's PI controller in Table 4.6, $Q(\lambda)$ AGC-II lowered the Pulse No. and Pulse Rev. No. by 29% and 34%, respectively. Moreover, it also provides a mechanism for the dispatch center to optimize the combined

objectives of CPS compliance and reducing the control cost by regulating the relaxation factors online. This means the AGC system can operate at a desired relaxed control level with least amount of control actions while the plants would operate at the high-efficiency mode under the optimal relaxed AGC strategy, and contribute to the energy-saving generation in power dispatch thereafter.

Chapter V

A Novel $R(\lambda)$ Imitation Learning Methodology for Online AGC Optimization

5.1 Introduction

The AGC performances of interconnected power systems have always been evaluated by NERC's CPS standards released in 1997 [37]. While the state-of-the-art of AGC strategies under CPS has been comprehensively addressed and investigated in [8, 38, 39], the existing CPS based AGC controllers mostly adopt fixed-parameter PI control strategies and do not adapt well to the changes in operation modes, parameters and structures of power systems [154, 156, 159]. In order to improve the dynamic performance and adaptability of AGC controllers, RL algorithms have been introduced to the design of optimal AGCs for control of interconnected power grids in [40, 43, 130, 195]. Most notably, in Chapter IV, a DROC model based multi-step $Q(\lambda)$ AGC was proposed for the optimal relaxed AGC to effectively overcome the long-time delay problem caused by the steam turbine of thermal units in the secondary frequency control loop.

The main obstacle for the practical onsite application of the above-mentioned conventional RL controllers is that it needs to be scheduled to experience a series of pre-learning processes before its onsite operation, and an accurate simulation model of the power system is required for this offline pre-learning process [74]. It is unavoidable that differences do exist between the power system model and real power system or even some uncertain scenarios cannot be modeled in the simulation. Therefore, the intolerable "trial-and-error noise" [71] may exist in power systems when RL controllers operate in real systems. To overcome this

problem, a new imitation pre-learning technique is presented in this chapter. The imitation learning philosophy [204] has been applied in robot control, in which the RL controllers learn from the movements and behaviors of human being by imitation [205, 206]. This technique can be executed with high applicability and feasibility because it does not rely on the accurate knowledge of system model, and thus can be utilized as a solution to the online pre-learning problem.

Furthermore, RL algorithms based on the AROC model [76] are designed to maximize the long-run cumulative average rewards of the generic systems, and would therefore match well with the control objective of CPS which considers the long-term average return performances of AGCs. In this chapter, a novel $R(\lambda)$ imitation learning method is proposed to combine the $R(\lambda)L$ algorithm [75] with an imitation pre-learning process, so as to develop an optimal AGC under CPS. This $R(\lambda)ILAGC$ has been successfully tested on the two-area thermal-dominated LFC system and the practical-sized CSG system outlined in Chapter IV.

5.2 Multi-Step $R(\lambda)$ Learning

The computational goal of the AROC is to find an optimum policy to achieve the maximal expected average reward from interaction with environment by trial-and-error [207]. For each state s and policy π , π^* is called gain-optimal policy if $V_{AROC}(\pi^*) \geq V_{AROC}(\pi)$ is satisfied [76]. However, the gain-optimal policy may not be the optimal policy in that, compared to the DROC model, the average function V_{AROC} in Eq. (2.10) ignores the relative importance of the short-term reward and long-term reward. It has been demonstrated that, among these numerous gain-optimal policies, bias-optimal policy can achieve optimal control objectives with the least iterative steps. The bias-optimal policy can be measured using the bias value function denoted by the average-adjusted sum of rewards, as follows,

$$V_B(\pi) = \lim_{n \rightarrow \infty} \text{Exp} \left\{ \sum_{k=0}^{n-1} [r_k(s_k, a_k) - V_{AROC}(\pi)] \right\} \quad (5.1)$$

where $V_B(\pi)$ is called the bias value function or relative value function. For each state s and MDP policy π , if $V_B(\pi^*) \geq V_B(\pi)$, then π^* is defined as the bias-optimal policy, and thus the goal of solving the AROC model is to find a bias-optimal policy π^* .

The R(λ)L is a relative new AROC-based RL algorithm which extends one-step R-learning [44] by combining it with multi-step TD(λ) returns for general λ [198] in an incremental way for the delayed MDP problems [72]. The algorithm can operate by successively improving its evaluations of an action value function, named R-function, to approach the bias-optimal control strategy [75]. Here, the R-function represents the average-adjusted reward of implementing an action a in state s once, and then following policy π subsequently. The R-function can be defined as below:

$$R^\pi(s, a) = r(s, s', a) - V_{\text{AROC}}(\pi) + \sum_{s' \in S} P(s' | s, a) \max_{a \in A} R^\pi(s', a) \quad (5.2)$$

Furthermore, in the algorithm, eligibility trace is used, as a temporary record of the occurrence of taken actions and state trajectory, to assign explicitly the rewards for multi-step backward sequences of decisions. In each iterative time, eligibility trace $e_k(s, a)$ for all of state-action pairs would decay exponentially by a factor of λ^k while the one for the visited pair would also be incremented by 1, as follows:

$$e_{k+1}(s, a) = \lambda e_k(s, a) \quad (5.3)$$

$$e_{k+1}(s_k, a_k) = e_{k+1}(s_k, a_k) + 1 \quad (5.4)$$

The designs of value function iteration policies for the optimal R-function are important and have a critical influence on the learning stability and numerical convergence of the algorithm. Here, the iteration policies with rigorous proof on the learning stability guarantees in [177] are employed. Therefore, the R(λ)L makes use of the TD(λ) returns as the value function estimator [198] for the estimation of value function errors $\rho_R(k)$ and $\delta_R(k)$, as follows:

$$\rho_R(k) = r(k) + R_k(s_{k+1}, a_{g(k+1)}) - R_k(s_k, a_k) - V_{\text{AROC}}(k) \quad (5.5)$$

$$\delta_R(k) = r(k) + R_k(s_{k+1}, a_{g(k+1)}) - R_k(s_k, a_{g(k)}) - V_{\text{AROC}}(k) \quad (5.6)$$

where $r(k)$ is a reward function of transition from state s_k to s_{k+1} under action a_k ; $R_k(s, a)$ represents the estimation for the R-function for action a set in state s at the k th iteration; $V_{\text{AROC}}(k)$ expresses the estimation of the average function at the k th iteration of learning. In the algorithm, the action selected from the maximal R-function in a given state is referred as greedy action [45]. Here, $a_{g(k)}$ and $a_{g(k+1)}$ represent the greedy actions in state s_k and state s_{k+1} , respectively. Then, the R-function can be updated using the following iterative rules:

$$R_{k+1}(s, a) = R_k(s, a) + \alpha e_{k+1}(s, a) \delta_R(k) \quad (5.7)$$

$$R_{k+1}(s_k, a_k) = R_{k+1}(s_k, a_k) + \alpha \rho_R(k) \quad (5.8)$$

where parameter α ($0 < \alpha < 1$) is the learning factor of the R-function. In addition to the iteration of R-function, the updating iterative policy for average function $V_{\text{AROC}}(k)$ can be designed as follows:

$$V_{\text{AROC}}(k+1) = V_{\text{AROC}}(k) + \tau \delta_R(k) \quad \text{if } a_k = a_{g(k)} \quad (5.9)$$

where parameter τ ($0 < \tau < \alpha^2$) is the learning factor of the average function.

The $R(\lambda)$ L algorithm is sensitive to the action selection strategy, and thus in order to achieve the reasonable action sampling as well as enhance the control performance of $R(\lambda)$ ILAGC, an action exploration policy called pursuit method [44], in which, for each state s , the actions are selected based on the probability distribution over the action space, is adopted, as expressed in Eq. (3.4) in Section 3.2. Since both state space S and action space A are finite, the values of eligibility trace, R-function and probability distribution of action selection can be stored as finite matrices with tabular forms. As the R-function and average function iterate and update continuously, the selection possibilities of the actions with larger R-functions will tend to become higher. Therefore, the $R(\lambda)$ ILAGC will gradually converge to the bias-optimal control strategy represented by optimal R-function once the state-action space are represented discretely and sampled sufficiently [75].

5.3 Design of Optimal $R(\lambda)$ ILAGC

5.3.1 $R(\lambda)$ Learning for AGC Strategy

The $R(\lambda)$ L not only can provide AGC with capability of self-learning for the average time-horizon optimization, but also would be flexible in accommodating various state-action pairs and control objectives for the multi-criteria CPS-based AGC optimization. Here, in the following state-action space analysis, two AGC performance indices, 1-min moving averages of CPS1 and ACE, are utilized as the binary input states of $R(\lambda)$ ILAGC to pursue the optimum average returns of CPS1 and CPS2 compliances, respectively [195]. On the other hand, action space A consists of a set of discrete control pulses $\Delta P_{\text{ord-RL}}$ which is the variation in the AGC generation within an AGC decision cycle. Here, the state-action space is specified and discretized following the space discretization in Section 4.4.1. Furthermore, the control objective of $R(\lambda)$ ILAGC is determined by reward function $r(k)$, and a relaxed reward function is adopted with following the multi-criteria piecewise function in Eq. (4.6). Finally, the learning-step of $R(\lambda)$ L, T_{step} , is determined by the AGC decision cycle.

In addition to the above definitions, four factors λ , α , τ and ζ in Eq. (5.3)~(5.9) and Eq. (3.4) are important in executing the algorithm and shall be set following generic guidelines [44, 75, 76]. Firstly, trace-decay factor λ is used to accelerate the convergence and alleviate the non-Markov effects of time-delay control [73]. For larger values of λ (close to 1), traces of preceding events would decay more slowly, and more of farther backward information can be utilized to optimize the R-function. For the thermal-dominated AGC problem under CPS which depends on previous control steps, the λ should be set close to 1 and the control strategy remains unaffected with λ between 0.75 and 0.9 in the simulation study. Here, it is set to 0.8. Secondly, the action exploration factor ζ in Eq. (3.4) is used for updating the action probability distribution. In order to avoid the local optima in the learning process, a small ζ (close to 0) shall be taken. Experiences show that

a value in the range from 0.05 to 0.3 would work well to balance the trade-off between exploration and exploitation. Here, the ζ is set to 0.1.

Initialize memory $R_0(s,a)$, $e_0(s,a)$ and $P_0(s,a)$, for all $s \in S$ and $a \in A$;
Initialize learning parameters and $T_{\text{step}} = \text{AGC decision cycle}$;
Set the initial state s_0 and $k = 0$;

Repeat

- 1) Choose an action a_k from the action space A based on the current action probability distribution $P_k(s,a)$;
- 2) Apply action a_k to the AGC system for the next T_{step} seconds;
- 3) Observe new state s_{k+1} using the moving averages of CPS1/ACE;
- 4) Obtain an immediate reward $r(k)$ from the relaxed reward function;
- 5) Calculate the value function errors via Eq. (5.5) and (5.6) respectively;
- 6) Update eligibility trace $e_k(s,a)$ and R-function $R_k(s,a)$ for all state-action pairs using Eq. (5.3) and (5.7), respectively;
- 7) Update the eligibility trace $e_{k+1}(s_k, a_k)$ and R-function $R_{k+1}(s_k, a_k)$ for the current state-action pair using Eq. (5.4) and (5.8), respectively;
- 8) If a_k is the greedy action, then update average function $V_{\text{AROC}}(k)$ using Eq. (5.9);
- 9) Update action probability distribution $P_k(s,a)$ to $P_{k+1}(s,a)$ via Eq. (3.4);
- 10) Let $k = k + 1$, return to step 1);

End

Fig. 5.1 Pseudo code of the $R(\lambda)L$ algorithm for AGC strategy

Moreover, the two learning factors, α and τ , shall be well-tuned through large amount of simulation testing heuristics as improper settings may lead to potential instability of the whole closed-loop control system. The first factor α essentially controls the extent of update in the R-function. A larger α (close to 1) allows fast convergence rate, while a smaller value of α (close to 0) can effectively enhance

learning stability of the $R(\lambda)L$. In order to guarantee the stability of $R(\lambda)ILAGC$ in onsite operation, α shall be very small. Simulation studies showed that a value in the range of 0.01~0.3 is acceptable. In the simulations, α is set to 0.01. The factor τ , introduced in Eq. (5.9), controls the extent of update for average reward function V_{AROC} . Since large variations of V_{AROC} would lead to the deterioration in control performance of the proposed $R(\lambda)ILAGC$ or even instability of the whole control system [76], a very small τ (close to 0) should be selected. In this AGC application, it is set to 0.001.

To sum up, the pseudo code of the $R(\lambda)L$ -based AGC controller is depicted in Fig. 5.1.

5.3.2 Imitation Pre-learning Process

The RL pre-learning process is a preconditioning technique which involves a series of iterative exploration procedures in the CPS state space to consummate the optimal R-function. Two modes of the pre-learning process, namely online mode and offline mode, were discussed and compared in [74]. In online learning mode, interaction occurs directly with the real system and simulation model is not required. The main drawback is that security and stability of the real power system may be jeopardized. Offline mode carries out in a simulation environment in which AGC extracts the offline-learned policy from an accurate power system model. However, if there are large differences between the system model and the real power system, the conventional RL controllers will not be able to obtain the optimum control strategy for its onsite implementation in the real power system, or even would result in mal-operation of the power system. Therefore, an online imitation pre-learning method is proposed to overcome those problems.

In the imitation pre-learning process, $R(\lambda)ILAGC$ can be directly put into the real power system, and act as an observer for online imitation and identification of the existing AGC in a real-time pattern. As illustrated in Fig. 5.2, the proposed AGC strategy observes AGC generation commands ΔP_{ord-PI} from the existing PI

controller to learn and imitate its control behaviors through $R(\lambda)$ learning policy evaluation [75]. During the proposed pre-learning process, the AGC system is still governed by the PI controller while the $R(\lambda)$ ILAGC would not influence the stability and security of the power system.

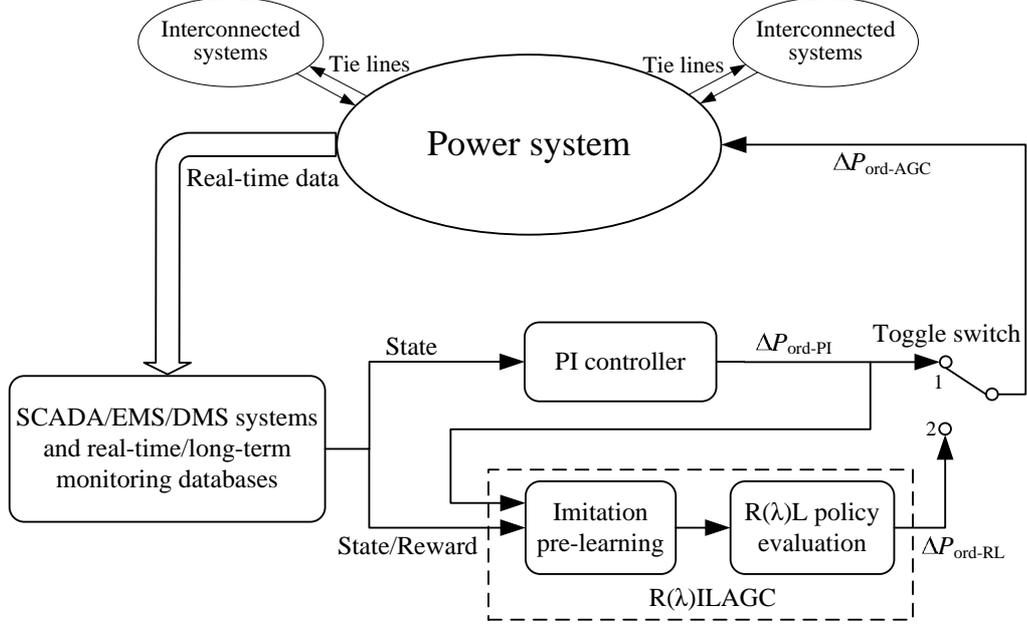


Fig. 5.2 Imitation pre-learning process of $R(\lambda)$ ILAGC

For the purpose of imitation learning, a mapping principle from continuous variable $\Delta P_{\text{ord-PI}}$ to discrete action space A shall be defined. Firstly, the variation of AGC generation commands, $dP_{\text{ord-PI}}$, can be calculated using Eq. (5.11). Then, the $dP_{\text{ord-PI}}$ is discretized into a set of intervals, and the interval vertices can be determined by the averages of two corresponding adjacent actions in action space A . Suppose space A consists of a set of monotone increasing actions as $\{a_1, a_2, \dots, a_{i-1}, a_i, \dots, a_{m-1}, a_m\}$, and the correspondence relationship between $dP_{\text{ord-PI}}$ and action a_i can be defined as follows:

$$\begin{cases} a_1 \leftarrow dP_{\text{ord-PI}}(k) & dP_{\text{ord-PI}}(k) \leq \frac{a_1 + a_2}{2} \\ a_i \leftarrow dP_{\text{ord-PI}}(k) & \frac{a_{i-1} + a_i}{2} < dP_{\text{ord-PI}}(k) \leq \frac{a_i + a_{i+1}}{2} \\ a_m \leftarrow dP_{\text{ord-PI}}(k) & dP_{\text{ord-PI}}(k) > \frac{a_{m-1} + a_m}{2} \end{cases} \quad (5.10)$$

$$dP_{\text{ord-PI}}(k) = \Delta P_{\text{ord-PI}}(k) - \Delta P_{\text{ord-PI}}(k-1) \quad (5.11)$$

After the design of mapping principle, the state-action space of $R(\lambda)$ ILAGC for the pre-learning process can thus be defined. Furthermore, the reward function, $r_{\text{pre}}(k)$, for the imitation pre-learning process is designed as follows:

$$r_{\text{pre}}(k) = \begin{cases} \sigma & a_k = a_i \\ -(a_k - a_i)^2 & a_k \neq a_i \end{cases} \quad (5.12)$$

where σ is a given non-negative number which is set to 0 in this case study. This reward function is to determine the imitation objective of $R(\lambda)$ ILAGC from the PI controller. Moreover, the iterative termination criterion for the imitation pre-learning process is determined by the following matrix 2-norm D_s :

$$D_s = \|R_k(s, a) - R_{k-1}(s, a)\|_2 \leq \zeta \quad (5.13)$$

where ζ (close to 0) is a given small termination tolerance which influences the convergence time and imitation precision for the imitation pre-learning process. The setting of tolerance factor ζ can be well-tuned by testing heuristics [208]. In general, a small tolerance factor ζ tends to increase the computational burden of the pre-learning but may enhance the imitation precision from the PI controller. It has been found from a series of trial heuristics that satisfactory imitation pre-learning could be obtained for different scenarios with ζ in the range of 0.35~0.7. Here, a value of 0.5 is selected. The pre-learning process will end once the termination criterion Eq. (5.13) is satisfied. All the parameters and memory, such as matrices $R(s, a)$, $P(s, a)$ and $e(s, a)$, should be stored, and then the toggle switch can be actuated to the contactor 2 to deactivate the PI controller and put the $R(\lambda)$ ILAGC online for live operation.

The developed online imitation pre-learning method provides a feasible way for practical application and implementation of RL controllers under a real power system environment and this would be the major original contribution in this chapter. The measurement of the real-time power system data as well as the delivery of AGC regulating commands can be carried out by the SCADA system

[80]. To sum up, the detailed pseudo code of the imitation pre-learning process is shown in Fig. 5.3.

Initialize memory $R_0(s,a)$, $e_0(s,a)$ and $P_0(s,a)$, for all $s \in S$ and $a \in A$;

Initialize learning parameters and $T_{\text{step}} = \text{AGC decision cycle}$;

Set the initial state s_0 and $k = 0$;

Repeat

- 1) Choose an action a_k from action space A based on the current action probability distribution $P_k(s,a)$;
- 2) Observe AGC generation command $\Delta P_{\text{ord-PI}}(k)$ from the PI controller;
- 3) Calculate AGC command variation $dP_{\text{ord-PI}}(k)$ using Eq. (5.11);
- 4) Determine the corresponding action a_i using Eq. (5.10)
- 5) Obtain an immediate reward $r_{\text{pre}}(k)$ using Eq. (5.12);
- 6) Apply $\Delta P_{\text{ord-PI}}(k)$ to the AGC system for the next T_{step} seconds;
- 7) Observe the next state s_{k+1} via the moving averages of CPS1/ACE;
- 8) Update R-function, eligibility trace, average function as well as action probability distribution with following Step 5) ~ Step 9) in Fig. 5.1;
- 9) Calculate iterative termination criterion D_s using Eq. (5.13);
- 10) If $D_s > \zeta$, then let $k = k+1$, and return to Step 1);
 Else, the imitation pre-learning process is completed, go to Step 11);
- 11) Actuate the toggle switch to contactor 2 in order to inactivate the PI controller, and put the R(λ)ILAGC into practical onsite operation;

End

Fig. 5.3 Pseudo code of the proposed imitation pre-learning process

5.4 Simulation Studies

5.4.1 Evaluation on Two-Area LFC Power System Model

The performance of $R(\lambda)$ ILAGC has been evaluated through a two-area LFC power system model as shown in Fig. 4.4. In this case study, control area A is taken as the study area, and the mapping principle from the $dP_{\text{ord-PI}}$ to the action space is illustrated in Fig. 5.4 through Eq. (5.10). The imitation pre-learning process should be executed continuously for a large number of operating states with different typical load disturbances until there are no more changes in R-functions. Therefore, a sequence of white noise load disturbances, which is a random pattern of square-wave fluctuations with flat power spectral density, is used as forcing function for the pre-learning process, as shown in Fig. 5.5. It can be found that the actions for various states in the R-function converge to their deterministic optimum policies, while termination criterion Eq. (5.13) is satisfied within 6537 seconds (the time boundary between the pre-learning and the onsite operation). On the other hand, since the PI controller is still functioned properly during the pre-learning process, the proposed $R(\lambda)$ IL can provide the satisfactory performance, with 10-min CPS1 indices and clock-10-min absolute averages of ACE $|E_{\text{AVE-min}}|$ (referred as CPS2), throughout the entire period.

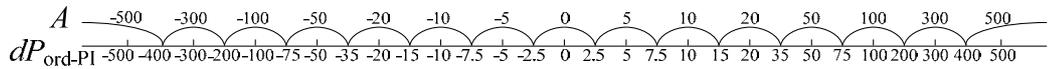


Fig. 5.4 Relationship between control actions of PI controller and $R(\lambda)$ ILAGC

It should be pointed out that the original PI controller installed in the real system is used to facilitate the transition from the offline pre-learning to online learning, provide the initial AGC strategy, and also improve the convergence rate of the $R(\lambda)$ ILAGC in the pre-learning process. The selection of PI controller is therefore important to the $R(\lambda)$ IL method. Here, an improved PI-based AGC developed by the NARI [11] is adopted. Furthermore, after the completion of the

imitation pre-learning process in each control area, the $R(\lambda)$ ILAGC, which has already benefited from the pre-learning experiences, will continue to make steady online optimization and still be able to improve its control behaviors by interaction with the real power system. Hence, as shown in Fig. 5.5, $R(\lambda)$ ILAGC in the onsite operation has exhibited the slightly better control performance than the PI controller.

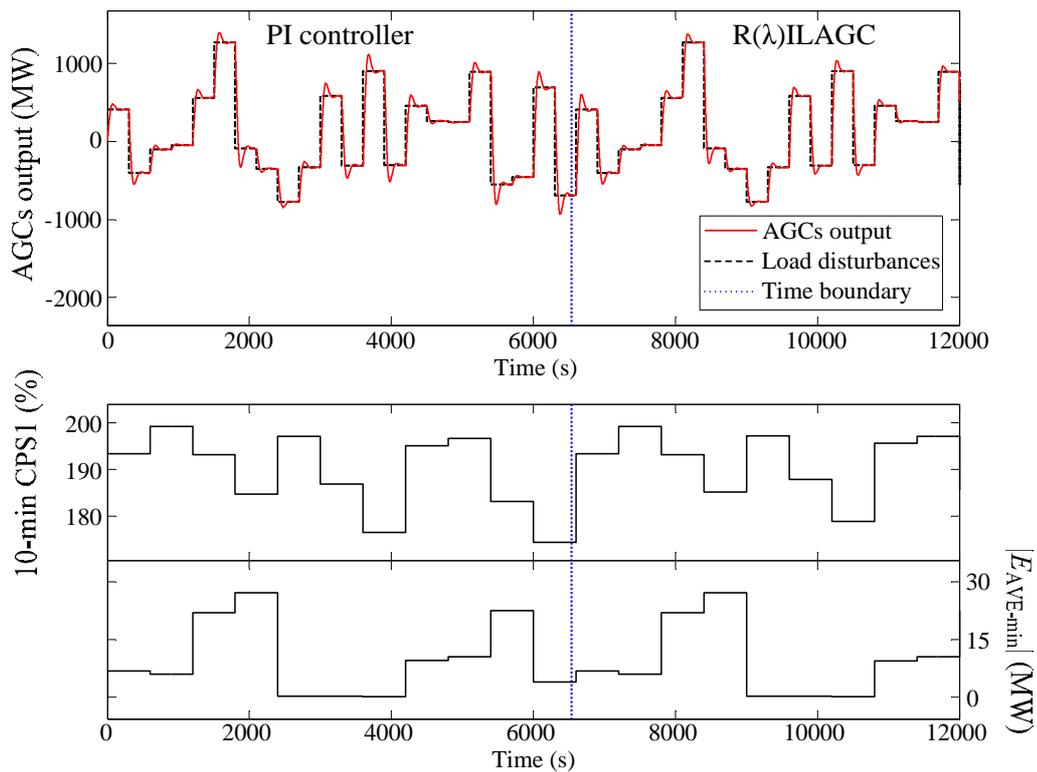


Fig. 5.5 Imitation pre-learning process of $R(\lambda)$ ILAGC

Though artificial load fluctuations shown in Fig. 5.5 will not be available in the practical application, the great variety of load disturbances found in the daily operation of a real power system can be served as the forcing function for the pre-learning process. The $R(\lambda)$ ILAGC would first be put into operation as an observer for a sufficient long period of time, say a month or even several months, to experience enough system scenarios for enhancing its control strategy. When all state-action pairs in the R-function are sampled sufficiently and termination

criterion Eq. (5.13) is satisfied, the $R(\lambda)$ ILAGC can then be activated to replace the PI controller for normal interconnected power system operation.

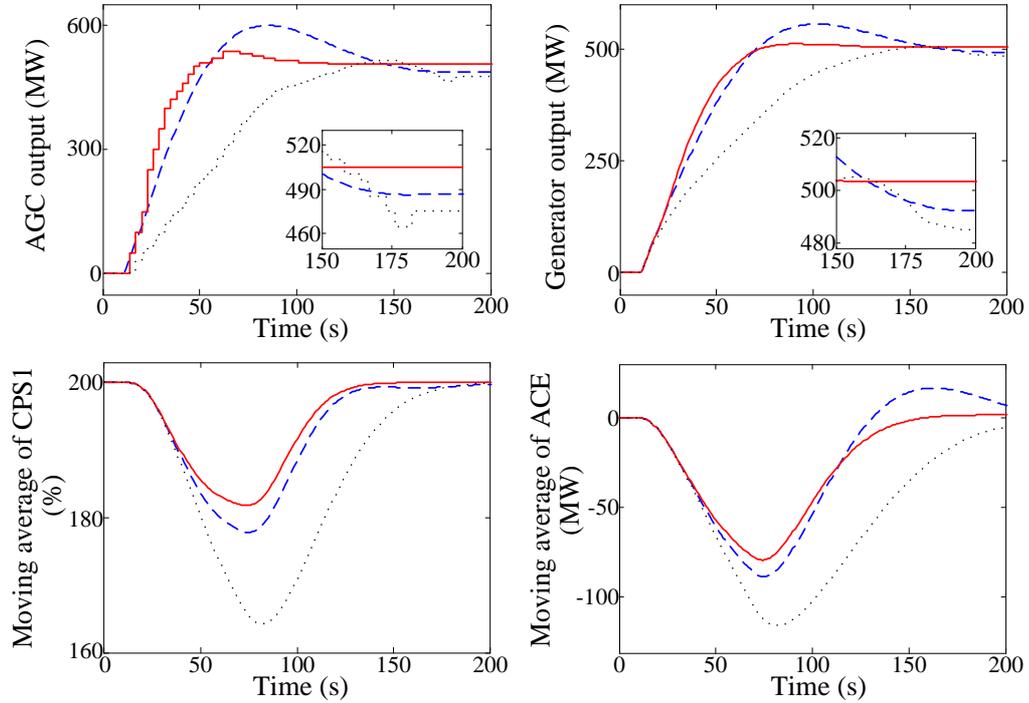


Fig. 5.6 Step response plots of AGCs (solid line: $R(\lambda)$ ILAGC; dot line: $Q(\lambda)$ AGC; dash line: PI controller)

Moreover, Fig. 5.6 shows the short-term dynamic response of the proposed AGC for a step load disturbance, and the performance results are benchmarked against NARI's PI controller and the $Q(\lambda)$ AGC in Chapter IV. In the experiment, $Q(\lambda)$ AGC employs the conventional offline pre-learning using a two-area LFC model with the wildly inaccurate parameters, and the simulations presented here correspond to the performances of the AGCs after their pre-learning processes. In general, the step response plots indicate that $R(\lambda)$ ILAGC performs satisfactorily and compares well with the performances of the two previously developed AGCs, and also exhibits different steady-state error from the other AGCs. With respect to $Q(\lambda)$ AGC, Fig. 5.6 reveals that the $R(\lambda)$ IL method can effectively avoid the deteriorated pre-learning strategy caused by the inaccurate power system model. In addition, the $R(\lambda)$ L algorithm can explicitly design its control objective using

reward function. Consequently, it is also interesting to note that, by interacting with the real power system during the onsite operation, $R(\lambda)$ ILAGC continuously optimizes its control strategy through the relaxed reward function, and thus can outperform the PI controller.

Furthermore, the transient dynamic response of the proposed $R(\lambda)$ ILAGC is mainly influenced by its reward function, and could be tuned by regulating the weights of the relaxed reward function. Taking the step response as an example, the influence of the relaxed reward function on the dynamic response of the RL driven AGC controllers, as well as the detailed tuning laws of the weights, have been thoroughly discussed and investigated in Chapter III and IV [40, 195].

5.4.2 Evaluation on China Southern Power Grid

For in-depth evaluation of the developed $R(\lambda)$ ILAGC in a realistic simulation environment, the CSG power system model, which was previously developed by utilities for Guangdong power dispatch center projects, is used as the benchmark system, as shown in Fig. 4.7. Furthermore, the long-term AGC performances are assessed with data statistical comparative experiments under the representative disturbances and uncertainties in power systems. A typical stochastic load model as well as white noise parameter perturbations [195], which have the same system disturbance settings as in Section 4.5.2, is set as a series of fixed scenarios for the comparisons of AGC strategies in Chapter III and IV.

In this experiment, the AGC simulator of the CSG power systems was started in a quiescent state in which frequency and tie-line flow were at the scheduled values, and the simulations were then implemented with the preset disturbance conditions over a 30-day period. The AGC performance metrics, including CPS compliances and relaxed indices, for the study area Guangdong power grid with assessment period 10 minutes are recorded, and the overall comparisons among four CPS-based AGCs, with the increase of white noise parameter perturbations

from 0% (nominal parameters) to 10% of its nominal value, are tabulated in Table 5.1 and 5.2.

Table 5.1 Statistical experiment results of Guangdong power grid I

Metrics	Nominal parameters			
	NARI's PI	QAGC	Q(λ)AGC	R(λ)ILAGC
$ \Delta F $ (Hz)	0.0282	0.0271	0.0273	0.0265
$ \text{ACE} $ (MW)	151.89	146.18	148.63	140.28
CPS1 (%)	183.33	187.35	185.67	190.32
CPS2 (%)	97.85%	98.55%	98.50%	98.75%
CPS (%)	94.51%	95.51%	95.29%	96.02%
Pulse No.	294	267	251	245
Pulse Rev. No.	78	67.8	63.3	61.9

Table 5.2 Statistical experiment results of Guangdong power grid II

Metrics	White noise parameter perturbations			
	NARI's PI	QAGC	Q(λ)AGC	R(λ)ILAGC
$ \Delta F $ (Hz)	0.0561	0.0412	0.0365	0.0353
$ \text{ACE} $ (MW)	233.17	213.03	185.75	175.64
CPS1 (%)	142.05	163.59	176.03	184.21
CPS2 (%)	90.83%	94.11%	95.89%	96.53%
CPS (%)	85.79%	90.48%	93.76%	94.17%
Pulse No.	338	303	271	269
Pulse Rev. No.	91.5	80.4	73.5	72.2

From Table 5.1 and 5.2, it can be found that, since RL algorithms would continue to learn the closed-loop control laws online, the three RL-based AGCs can adapt well to the changing operation scenarios and show the superior adaptability and dynamic optimization than NARI's PI controller, especially when the parameters were perturbed. On the other hand, in comparison with the QAGC in Chapter III and PI controller, R(λ)L allows the multi-step backward estimation for prediction of observations in the past, and thus can enhance the efficiency of online learning and overcome the non-Markov delayed AGC

problem in the thermal-dominated CSG power systems [195]. Moreover, the DROC-based QAGC and $Q(\lambda)$ AGC pay more attention to their immediate rewards for the iterative policy evaluation [44] while AROC would pursue for the maximization of the cumulative average returns, without discounting the forward rewards in the control objective, over the entire assessment period. Therefore, it could be seen from Table 5.1 and 5.2 that, compared with QAGC and $Q(\lambda)$ AGC in Chapter III and IV, the proposed $R(\lambda)$ ILAGC provides the best performance enhancements, and hence can effectively improve CPS percentage compliances and AGC relaxed performance when the power system is subjected to a series of complex system disturbances.

5.5 Summary

This chapter develops a novel $R(\lambda)$ IL control methodology for the optimal AGC under NERC's CPS, which possesses the following advantages:

- (1) The proposed online imitation pre-learning technique does not rely on any accurate system model for the offline pre-learning process and therefore can provide a feasible and promising means for the practical implementation of RL algorithms in real power systems.
- (2) For the first time, the AROC model is introduced to the design of AGC for interconnected power grids. The undiscounted optimality framework-based $R(\lambda)$ L algorithm is more applicable to the CPS optimization problem than the DROC-based algorithms.
- (3) $R(\lambda)$ IL can form the basis of performance-adaptive control for the nonlinear and stochastic behaviors in power systems, and can effectively cope with the time-delayed feedback control problem which is an inherent issue lying in various power system control problems.

Simulation studies have demonstrated that $R(\lambda)$ ILAGC is capable of online observing system variations, identifying and learning the existing AGC controller

in the real-time operation with fast convergence rate and high CPS compliances in the imitation pre-learning process. Moreover, as confirmed by the statistical comparative experiments, the proposed $R(\lambda)$ ILAGC outperforms previous AGC controllers, and the adaptability and control performance of the AGC system can be enhanced greatly under complex changing power system scenarios.

Chapter VI

Equilibrium-Inspired Multiple Group Search Optimizer with Synergistic Learning for Multi-objective Electric Power Dispatch

6.1 Introduction

Electric power dispatch is an essential function required in the modern EMS system to determine the optimal steady-state operation of dispatchable generating plants with multiple contradictory objectives, such as generation cost, emission and energy saving, subjected to a set of operational and physical constraints [5]. Over the years, extensive research has been done in the area of multiobjective power dispatch problems. Most notably, different techniques have been proposed to transform the dual-objective EED problem into a single-objective optimization using the linear weighted aggregate approach [24-28]. Nevertheless, many trials are required for those methods to obtain a desired set of noninferior solutions by varying the weights of optimization objective function, and they are not effective to handle the non-differential and nonconvex problems. Therefore, in recent years, various Pareto-based multiobjective stochastic optimization algorithms have been proposed to solve this MOPD problem.

So far, the state-of-the-art in Pareto-based optimization algorithms, including niched Pareto genetic algorithm [52], multiobjective stochastic search technique [171], nondominated sorting genetic algorithm [172], NSGA-II [175], strength Pareto evolutionary algorithm [59], multiobjective bacteria foraging algorithm [60], multiobjective differential evolution [56], multiobjective chaotic ant swarm optimization [50], and fuzzy clustering-based particle swarm optimization [61],

etc., have been successfully applied to the biobjective EED problem on a small IEEE 30-bus 6-generator system to obtain a Pareto tradeoff between fuel cost and atmospheric emission. Those Pareto algorithms operate on a set of nondominated solutions with different search mechanisms in a single simulation run. However, some important system constraints such as system voltage and reserve constraints have not been considered in the literatures yet, and this work is hence aimed to develop a novel multiobjective algorithm designed for the highly constrained and high-dimensional MOPD optimization problems with more objective functions.

Recently, a new optimization algorithm inspired by group living and foraging behaviors of animals, called group search optimizer, has been proposed based on the producer-scrounger model [77]. It has also been demonstrated that the overall performance of the GSO methodology exhibits superiority and high efficiency on the non-differential, high-dimensional and multimodal optimization problems.

Here, a meta-heuristic MGSO algorithm is further developed ingeniously to form a significantly improved multiobjective algorithm for large-scale MOPD applications. A novel stochastic learning automata based reinforcement scheme [209] is formulated to explicitly assign the rewards among searching individuals for synergistic learning [210] which allows parallel groups to have information interaction and resource sharing in the cooperative search process. Furthermore, a dynamic search-space reduction strategy [211] is used in scanning mechanism to obtain accurate and extreme vertex solutions in the PF surface, and chaotic sequence dispersion [212] is introduced to improve the population diversity and avoid entrapment into local optima. Meanwhile, the algorithm handles problem objectives and constraints separately based on the Boltzmann distribution [44] so as to direct infeasible members towards the sparsely populated regions of PF surface. In addition, the average linkage-based hierarchical clustering method is implemented to extract a representative elitist PF from the nondominated set for the decision maker. Lastly, a Nash equilibrium-inspired decision-making method [78] is then proposed to extract the best compromise solution from the elitist PF.

The effectiveness and validity of the proposed algorithm have been thoroughly verified on the IEEE 30-bus and IEEE 118-bus test power systems.

6.2 Problem Formulation

6.2.1 MOPD Objectives

1) *Economic Objective*: The economic objective of MOPD is to minimize the total generation cost. The fuel cost of thermal generators with high nonlinearity and non-convexity caused by valve-point effects can be modeled as ripple curve [211], and the total fuel cost $F(P_G)$ in (\$/h) can then be expressed with quadratic functions and sine components as follows:

$$F(P_G) = \sum_{i=1}^{N_G} (a_i P_{Gi}^2 + b_i P_{Gi} + c_i + |d_i \sin[e_i(P_{Gi,\min} - P_{Gi})]|) \quad (6.1)$$

where N_G is the number of dispatchable generators; a_i , b_i , c_i , d_i , and e_i are the cost coefficients of the i th generator; $P_{Gi,\min}$ is the minimum active power output of unit i ; P_G is the vector of the active power generations and defined as:

$$P_G = [P_{G1}, P_{G2}, P_{G3}, \dots, P_{GN_G}] \quad (6.2)$$

2) *Emission Objective*: The objective of emission dispatch is to minimize the atmospheric pollutants due to fossil-fueled thermal units, such as sulfur dioxides and nitrogen oxides, and so on [171]. The total environmental emission $E(P_G)$ in (ton/h) can be represented as:

$$E(P_G) = \sum_{i=1}^{N_G} 10^{-2} (\alpha_i P_{Gi}^2 + \beta_i P_{Gi} + \gamma_i + \zeta_i \exp(\lambda_i P_{Gi})) \quad (6.3)$$

where α_i , β_i , γ_i , ζ_i , and λ_i are the emission coefficients of the i th dispatchable generator characteristics.

3) *Energy-Saving Objective*: The aim of energy-saving generation dispatch is to minimize power transmission losses, and minimization of active power loss in transmission lines P_{Loss} can therefore be used as an objective of the MOPD. The solution of system loss P_{Loss} involves the calculation and analysis of load flow

[5], which includes the equality constraints on active and reactive power on each bus, as follows:

$$\begin{cases} P_{Gi} - P_{Di} - V_i \sum_{j=1}^{N_B} V_j [G_{ij} \cos(\delta_i - \delta_j) + B_{ij} \sin(\delta_i - \delta_j)] = 0 \\ Q_{Gi} - Q_{Di} - V_i \sum_{j=1}^{N_B} V_j [G_{ij} \sin(\delta_i - \delta_j) - B_{ij} \cos(\delta_i - \delta_j)] = 0 \end{cases} \quad (6.4)$$

where $i = 1, 2, \dots, N_B$; N_B is the number of system buses; P_{Di} and Q_{Di} are active and reactive load demand at the i th bus, respectively; Q_{Gi} is the reactive power generated at the i th bus; G_{ij} and B_{ij} are the transfer conductance and susceptance between bus i and j , respectively; V_i and V_j are the voltage magnitudes at bus i and j , and δ_i and δ_j denotes the voltage angles at bus i and j , respectively. The nonlinear equality constraints Eq. (6.4) can be readily solved using the Newton-Raphson method [1] for the solution of load flow problem, and then the active transmission power loss can be obtained as:

$$P_{\text{Loss}} = \sum_{k=1}^{N_L} g_k [V_i^2 + V_j^2 - 2V_i V_j \cos(\delta_i - \delta_j)] \quad (6.5)$$

where N_L is the number of transmission lines, and g_k is the conductance of the k th line connecting bus i and j .

6.2.2 MOPD Constraints

1) *Power Balance Constraint*: Since the total power generation outputs of generators must equal to the sum of total load demand P_D plus power loss P_{Loss} , after the load flow calculation the active power output of the slack generator is reassigned to satisfy the equality constraint Eq. (6.6) as below:

$$\sum_{i=1}^{N_G} P_{Gi} - P_D - P_{\text{Loss}} = 0 \quad (6.6)$$

2) *Active Power Generation Constraints*: The active power output of each unit should be restricted within its lower and upper limits, and the corresponding inequality constraint is:

$$P_{Gi, \min} \leq P_{Gi} \leq P_{Gi, \max} \quad i = 1, 2, \dots, N_G \quad (6.7)$$

where $P_{Gi,\max}$ is the maximum active generation for unit i . For unit i with NP_i prohibited operating zones, its feasible operating zones can then be described as a disjoint nonconvex set, as follows:

$$\begin{cases} P_{Gi,\min} \leq P_{Gi} \leq P_{Gi(1),lb} \\ P_{Gi(j-1),ub} \leq P_{Gi} \leq P_{Gi(j),lb} & j = 2, 3, \dots, NP_i \\ P_{Gi(j),ub} \leq P_{Gi} \leq P_{Gi,\max} & j = NP_i \end{cases} \quad (6.8)$$

where $i = 1, 2, \dots, N_G$; $P_{Gi(j),lb}$ and $P_{Gi(j),ub}$ represent lower and upper boundary of the j th POZ of generator i , respectively.

3) *System Spinning Reserve Constraint*: For reliable and secure operation, the spinning reserve demand [213] should be considered for contingency conditions as follows:

$$\sum_{i=1}^{N_G} SP_{Gi} \geq SP_R \quad (6.9)$$

where SP_R represents the spinning reserve requirement in MW. Since the POZs of the generators would severely limit their flexibility to regulate the system load, these generators cannot contribute to the system spinning reserve [213] and the spinning reserve with POZs considered can be calculated as below:

$$SP_{Gi} = \begin{cases} 0 & \forall i \in \Omega \\ P_{Gi,\max} - P_{Gi} & \text{others} \end{cases} \quad (6.10)$$

where Ω represents the set of the generators with POZs.

4) *Reactive Power Generation Constraints*: Similarly, reactive power output generated by generator i should also be within its lower limit $Q_{Gi,\min}$ and upper limit $Q_{Gi,\max}$, as follows:

$$Q_{Gi,\min} \leq Q_{Gi} \leq Q_{Gi,\max} \quad i = 1, 2, \dots, N_G \quad (6.11)$$

5) *Nodal Voltage Constraints*: While the voltage magnitude of generator bus i , V_{Gi} , should be maintained at its rating voltage $V_{Gi,\text{rating}}$, the nodal voltage magnitude of load bus j should be constrained between its minimum limit $V_{j,\min}$ and maximum limit $V_{j,\max}$, as follows:

$$V_{Gi} = V_{Gi,\text{rating}} \quad i = 1, 2, \dots, N_G \quad (6.12)$$

$$V_{j,\text{min}} \leq V_j \leq V_{j,\text{max}} \quad j = 1, 2, \dots, N_B \quad (6.13)$$

6) *Transmission Security Constraints*: The apparent power flow through the k th transmission line should not exceed its loading limit $LF_{k,\text{max}}$ as follows so as to avoid any overloading:

$$|LF_k| \leq LF_{k,\text{max}} \quad k = 1, 2, \dots, N_L \quad (6.14)$$

where LF_k is the apparent power flow of the k th branch line connecting bus i and j , and its magnitude can be calculated as the larger line flow at bus i and j , as follows:

$$|LF_k| = \max \left[|LF_{ij}|, |LF_{ji}| \right] = \max \left[|V_i \angle \delta_i \cdot I_{ij}^*|, |V_j \angle \delta_j \cdot I_{ji}^*| \right] \quad (6.15)$$

where I_{ij}^* is the complex conjugate of current flow from bus i to j , and the line flow to bus i can be computed as follows:

$$LF_{ij} = V_i \angle \delta_i \cdot [(V_i \angle \delta_i - V_j \angle \delta_j) \cdot y_{ij} + V_i \angle \delta_i \cdot j y_b / 2]^* \quad (6.16)$$

where y_{ij} is the line admittance in nodal admittance matrix; and y_b is the shunt susceptance of the transmission line.

6.3 Proposed Multiple Group Search Optimizer

6.3.1 Algorithm Framework

The proposed MGSO algorithm integrates various stochastic global searching and probability selection techniques for different types of population members. Firstly, the population of MGSO consists of multiple searching groups, and each group is designed based on producer-scrounger model [214] for each objective of MOPD problem. For each searching group, there are four categories of members for four different searching strategies as following: 1) Producer: this member is designated to the member conferring the best single-objective fitness in the group in each iterative step, and it is the group leader which has a critical impact on the overall searching direction of the group. 2) Scroungers: except for the producer,

80% of the remaining feasible members are selected randomly as the scroungers which constitute the main searching force in the algorithm. Therefore, the update policy of this swarm should take all the objectives into account through the social cooperative mechanisms among these groups. 3) Rangers: the remaining feasible members in the group are the rangers, and they should move in an unpredictable dispersion to discover new resources globally. 4) Infeasible members: the update policy of this swarm should be a constraint satisfaction process for handling the complex constraints.

Every individual in the population has a current position $X \in \mathbb{R}^{N_G}$ and a head angle $\varphi \in \mathbb{R}^{N_G-1}$. The search direction vector of the head angle [77], $D(\varphi) = (d_1, d_2, d_3, \dots, d_{N_G}) \in \mathbb{R}^{N_G}$, can be calculated by converting polar to Cartesian coordinates, as follows:

$$\begin{aligned}
 d_1 &= \prod_{j=1}^{N_G-1} \cos(\varphi_j) \\
 d_i &= \sin(\varphi_{i-1}) \cdot \prod_{j=i}^{N_G-1} \cos(\varphi_j) \quad (i = 2, \dots, N_G - 1) \\
 d_{N_G} &= \sin(\varphi_{N_G-1})
 \end{aligned} \tag{6.17}$$

6.3.2 Initialization

The population can be initialized by generating the members of each group randomly within the boundary search patch as:

$$\begin{aligned}
 X_{ij}^0 &= LB + r_1 \circ (UB - LB) \\
 i &= 1, 2, \dots, M_{\text{obj}}; \quad j = 1, 2, \dots, M_p
 \end{aligned} \tag{6.18}$$

where M_p represents the population size of each searching group; $r_1 \in \mathbb{R}^{N_G}$ is a uniform random sequence in the range (0, 1); LB and UB are the lower and upper bound for variable vector X , respectively; operator “ \circ ” calculates the entrywise product of two matrices. After the initialization of members in each group, the multiobjective fitness sequence for each member can then be calculated from Eq. (6.1)~(6.5). For each initial group, if there is no feasible solution that satisfies all

the problem constraints, the members in this seeking group will be reinitialized until there is at least one feasible member that can be used as the producer in each group. In addition, the initialized head angle for each group member is set to $(\pi/4, \pi/4, \dots, \pi/4)$ as recommended in [77].

6.3.3 Variable-size External Repository

The external repository is a bounded elite archive used for preserving the nondominated solutions found along the search process. After the initialization of each searching group, the initial repository is determined by the nondominated members obtained in the initial population. For each iterative step of the MGSO algorithm, each nondominated individual obtained in the new generation of the population is checked for dominance with the solutions in the current repository. The dominance comparison strategies for updating the archive are as follows:

- If the new solution obtained is infeasible or dominated by other members in the population, the solution will not be saved into the repository;
- In case a nondominated member in the population cannot be dominated by any solution in the current repository, then the solution will be saved into the repository;
- Any dominated solution in the current repository by this new nondominated member will be removed from the repository.

Though a large-size memory of the repository tends to represent the better characteristics of the PF, it would lead to explosion in the computational burden due to the dominance comparisons [215]. Therefore, the number of PF solutions saved in the repository, i.e. the size of the PF, should be limited. In this thesis, a variable-size elite repository, which can be resized on demand, is adopted. After each iterative step of the algorithm, the repository will be resized to cover the entire nondominated set, including the survival solutions in the repository and new nondominated members. The resized repository could further be shrunk if the clustering enhancement described in Section 6.3.8 was adopted.

6.3.4 Space Reduction-based Scanning Strategy for Producer

The producer employs the scanning strategy inspired from white crappies [77] to pursue new Pareto-optimal solutions. The scanning field can be characterized by maximum pursuit angle $\theta_{\max} \in \mathbb{R}^1$ and maximum pursuit distance $L_{\max} \in \mathbb{R}^1$, and the apex is the producer's current position X_p . In the i th group at the k th iteration, the producer will scan the N_G -dimensional hypercube field by randomly sampling three points: one point at the zero degree, two points in the left and right sides symmetrically as:

$$X_{iz} = X_{ip}^k + r_2 L_{i\max} D(\varphi_{ip}^k) \quad (6.19)$$

$$X_{ir} = X_{ip}^k + r_2 L_{i\max} D(\varphi_{ip}^k + r_3 \theta_{\max} / 2) \quad (6.20)$$

$$X_{il} = X_{ip}^k + r_2 L_{i\max} D(\varphi_{ip}^k - r_3 \theta_{\max} / 2) \quad (6.21)$$

where $r_2 \in \mathbb{R}^1$ is a random number based on normal distribution with mean 0 and standard deviation 1; $r_3 \in \mathbb{R}^{N_G-1}$ is a random sequence with uniform distribution in the range (0, 1).

After the scanning at each iterative step, three objective fitness vectors are sampled by the producer in each group, and the corresponding nondominated solution will be stored to the repository via dominance comparisons. Meanwhile, if the producer can find a better group single-objective fitness than its existing fitness value, then it will move to this point; otherwise, it will stay at its current position and update its head angle as:

$$\varphi_{ip}^{k+1} = \varphi_{ip}^k + r_3 \psi_{\max}, \quad \psi_{\max} = \theta_{\max} / 2 \quad (6.22)$$

where ψ_{\max} is the maximum turning angle. Furthermore, in case the producer cannot find a better position within A_{\max} iterations, it will turn its head angle back to zero degree, as follows:

$$\varphi_{ip}^{k+A_{\max}} = \varphi_{ip}^k, \quad A_{\max} = \text{round}(\sqrt{N_G + 1}) \quad (6.23)$$

In order to generate an accurate optimal solution for each objective as well as facilitate the convergence process, a space reduction strategy [211] is added to

adaptively adjust the maximum pursuit distance for better local exploitation in the scanning space. Initially, the maximum pursuit distance for the i th group can be determined as follows:

$$L_{i\max}^0 = \|UB - LB\|_2 = \sqrt{\sum_{j=1}^{N_G} (P_{Gj,\max} - P_{Gj,\min})^2} \quad (6.24)$$

This strategy will be activated when the performance of the producer in the i th group is not improved after a prespecified iteration period $iter_{L\max}$. In this case, its scanning space will be dynamically shrunk based on the maximum pursuit distance at the k th iteration, as follows:

$$L_{i\max}^{k+1} = L_{i\max}^k (1 - \Delta_L) \quad (6.25)$$

where Δ_L is the predetermined step-size constant.

Moreover, after each iterative cycle, the member found the best fitness value for the corresponding objective is chosen as the producer in this group. At the same time, if a better group single-objective fitness could be found from the member of external groups, the producer will move to the position of the member for the most promising resource.

6.3.5 Synergistic Learning for Scroungers

In this thesis, a synergistic learning strategy inspired from stochastic learning automata [209] is proposed for extending the single objective GSO to cope with the multiobjective problems. During the k th searching bout of the algorithm, the scroungers in the i th group use a special area copying behavior [214], which move across and learn from the promising resources found by their leader and external group members, to pursue Pareto-optimal solutions, as follows:

$$X_{ij}^{k+1} = X_{ij}^k + C_p^k r_p \circ (X_{ip}^k - X_{ij}^k) + \sum_{m=1, m \neq i}^{M_{\text{obj}}} C_m^k r_m \circ (X_{lm}^k - X_{ij}^k) \quad (6.26)$$

$$i = 1, 2, \dots, M_{\text{obj}}; \quad j = 1, 2, \dots, M_{si}^k$$

where $r_p, r_m \in \mathbb{R}^{N_G}$ are uniform random vectors in the range (0, 1); the second and third terms in Eq. (6.26) are referred as the leader component and synergistic

component, respectively; C_p^k and C_m^k are the coefficients of leader component and synergistic components at the k th iteration; X_{lm}^k represents the position of the member selected from the m th group at the k th iteration; M_{si}^k is the number of scroungers in the i th group at the k th iteration.

Since the selection of the members from external groups for the interactive cooperation is important to the performance of synergistic learning, a new linear reinforcement scheme based probability distribution selection is proposed such that all nondominated solutions found in the search process are regarded as the social achievement of the searching groups, and reinforcement rewards [45] can be assigned to each member in terms of the good solutions found by this member. Initially, a uniform probability distribution for members in each group is adopted. If T_{ni}^k represents the set of members seeking out the nondominated solutions in the i th group at the k th iteration, the selection probability of the j th member in the i th group at the k th iteration, $prob_{ij}^k$, can be updated as follows:

$$prob_{ij}^{k+1} = \begin{cases} prob_{ij}^k + \eta(1 - \sum_{t \in T_{ni}^k} prob_{it}^k) & j \in T_{ni}^k \\ prob_{ij}^k(1 - \eta N_{ni}^k) & \text{otherwise} \end{cases} \quad (6.27)$$

where N_{ni}^k is the number of elements in set T_{ni}^k ; η ($0 < \eta < 1/N_{ni}^k$) is a small constant called reinforcement factor. It can be found that the member which can find most nondominated solutions, e.g. the scanning strategy-based producers, has the high selection probability for the synergistic component in Eq. (6.26).

Simulation studies indicated that the search performance of MGSO can be significantly enhanced by fine-tuning the coefficients in Eq. (6.26). In the initial stage of the search process, since most of group members have not sought out the promising areas, scroungers should give priority to following the producer in its own group. On the other hand, along with the convergence of searching groups, scroungers can learn from other groups for the PF set, and the leader component should decrease while synergistic components increase during the search process.

After comparative studies, the coefficients at the k th iteration can be set using hyperbolic tangent functions as below:

$$C_p^k = \frac{(C_{p,\max} - C_{p,\min})}{2} \left[-\tanh\left(c_{rp}\left(k - \frac{Iter_{\max}}{2}\right)\right) + 1 \right] + C_{p,\min} \quad (6.28)$$

$$C_m^k = \frac{(C_{m,\max} - C_{m,\min})}{2} \left[\tanh\left(c_{rm}\left(k - \frac{Iter_{\max}}{2}\right)\right) + 1 \right] + C_{m,\min} \quad (6.29)$$

where $Iter_{\max}$ is the maximum iteration number for termination criterion of the proposed MGSO; $C_{p,\max}$, $C_{p,\min}$ and $C_{m,\max}$, $C_{m,\min}$ are the constants used for the upper and lower bounds of coefficients C_p and C_m , respectively; c_{rp} and c_{rm} are constants which are designed to control the variation ramp rates of the hyperbolic tangent functions.

6.3.6 Chaotic Sequence Dispersion for Rangers

Here, a special random walk dispersion is employed by rangers to improve the population diversity and the global exploration for dispersive PF resources over the entire search space as below:

$$X_{ij}^{k+1} = X_{ij}^k + r_{\text{sign}} r_{il}^k L_{i\max}^0 D(\varphi_{ij}^k + r_{i\varphi}^k 2\pi) \quad (6.30)$$

$$i = 1, 2, \dots, M_{\text{obj}}; \quad j = 1, 2, \dots, M_{ri}^k$$

where M_{ri}^k is the number of rangers in the i th group at the k th iteration; r_{sign} is a randomly generated sign equal to 1 or -1 . Previous work in [212] showed that the global searching performance can be enhanced with the use of the chaotic sequences instead of random sequences. Therefore, here, $r_{il}^k \in \mathbb{R}^1$, $r_{i\varphi}^k \in \mathbb{R}^{N_G-1}$ are time series generated by chaotic logistic functions rather than random number generators, and thus the logistic map based chaotic sequence iterator in Eq. (6.30) can be expressed as follows:

$$r_{il}^{k+1} = \mu r_{il}^k (1 - r_{il}^k) \quad (6.31)$$

$$r_{i\varphi}^{k+1} = \mu \circ r_{i\varphi}^k \circ (1 - r_{i\varphi}^k) \quad (6.32)$$

where μ is a control factor determining the time series to be constants, oscillate within limits, or behave chaotically in an unpredictable pattern. Here, sequences

Eq. (6.31) and (6.32) are deterministic and will display chaotic behaviors when $\mu=4.0$, and the initial values of the chaotic sequences should not contain any members of the following $\{0, 0.25, 0.50, 0.75, 1.0\}$ [212].

Furthermore, GSO is not sensitive to most of its parameters except for the percentage of rangers, and the recommended percentage in [77] is 20%. In order to motivate the individuals to explore the global search space when getting stuck into the local PF, an adaptive strategy for the ranger percentage in i th group, $\%_{iR}$, is adopted. This strategy will be activated once no nondominated solutions can be sought to improve the PF in a given iteration period $iter_{Rmax}$. Thereafter, the percentage of rangers at the k th iteration can be increased as follows:

$$\%_{iR}^{k+1} = \%_{iR}^k + \Delta_R \quad (6.33)$$

where Δ_R is the prespecified step-size percentage constant.

6.3.7 Constraint Handling Strategy

Firstly, in order to restrict group members to search within their generation constraints Eq. (6.7)~(6.8), the following strategy is placed to cope with the bounded search patch: when a member moves outside the search patch, it will be turned back to the search patch by setting the violated dimensional variables in the member to its previous values.

For the sake of effective constraint handling for highly complex constrained search space, infeasible members will be separated from the population so as to guide them towards the feasible space for pursuing new Pareto-optimal solutions. Here, this is achieved through the following policy:

$$\begin{aligned} X_{ij}^{k+1} &= X_{ij}^k + r_{pf} \circ (X_{pf}^k - X_{ij}^k) \\ i &= 1, 2, \dots, M_{obj}; \quad j = 1, 2, \dots, M_{ini}^k \end{aligned} \quad (6.34)$$

where M_{ini}^k represents the number of infeasible members in the i th group at the k th iteration; $r_{pf} \in \mathbb{R}^{N_G}$ is a uniform random vector in the range $(0, 1)$; X_{pf}^k denotes a nondominated solution selected from the current PF in the repository for the j th

infeasible member in the i th group at the k th iteration. Furthermore, in order to direct the infeasible swarm towards sparsely populated regions for a uniformly-distributed PF, a Boltzmann distribution based on the crowding distance [216] is used to form the probability distribution for selecting X_{pf}^k , as following:

$$prob_{Bn}^k = \frac{e^{d_{cn}^k / wL_{dn}^k}}{\sum_{n=1}^{M_{\text{rep}}^k} e^{d_{cn}^k / wL_{dn}^k}} \quad (6.35)$$

where M_{rep}^k is the number of feasible nondominated solutions to be selected for Eq. (6.34) in the current repository; w is a positive parameter called temperature; L_{dn}^k denotes the normalized Euclidean distance, in the solution space, between the corresponding nondominated solution of the n th fitness element in the elite repository and a given infeasible member at the k th iteration; d_{cn}^k represents the normalized crowding distance, in the objective space, of the n th element in the repository at the k th iteration. The crowding distance, which expresses a measure of front density with the neighborhood, requires sorting the current PF fitness values, in terms of each objective function in an ascending order. For the i th objective, the boundary solutions, with the smallest fitness f_i^{\min} and largest fitness f_i^{\max} , are assigned to 1 in this thesis. On the other hand, the normalized crowding distance of all other intermediate solutions can be computed as follows:

$$d_{cn}^k = \sum_{i=1}^{M_{\text{obj}}} \frac{|f_{i(+1)} - f_{i(-1)}|}{f_i^{\max} - f_i^{\min}} \quad (6.36)$$

where, following the sorting order by the i th objective, $f_{i(+1)}$ and $f_{i(-1)}$ represent the fitness of two adjacent solutions of the n th element in the current repository. Therefore, the infeasible members can be hauled towards the nearby preferable feasible PF regions, and thereby the border of feasible search space can also be readily located to seek the potential Pareto-optimal solutions.

6.3.8 Pruning Pareto Set

After each searching bout, new nondominated solutions will be found and saved in the variable-size repository. When the number of repository elements exceeds the prespecified size of PF, M_{pf} , an average linkage-based hierarchical clustering method [185] is adopted to prune the nondominated set to a desirable size with its trade-off characteristics preserved. This method is used to iteratively classify and join the repository solutions into the required number of clusters. Firstly, each repository element constitutes a distinct cluster, and the average linkage distance d_{al} of two clusters cl_1 and cl_2 can be calculated as follows:

$$d_{al}(cl_1, cl_2) = \frac{1}{n_1 n_2} \sum_{i_1 \in cl_1, i_2 \in cl_2} d_E(i_1, i_2) \quad (6.37)$$

where n_1 and n_2 represent the numbers of individuals in the clusters cl_1 and cl_2 , respectively; function $d_E(i_1, i_2)$ represents the Euclidean distance in the objective space between individuals i_1 and i_2 . Then, the adjacent clusters with the minimal distance d_{al} in all possible pairs of clusters will be combined into a larger cluster. This iterative procedure will be continued until the required number of clusters is satisfied, and then the nearest individual to the centroid of each cluster can be extracted as the representative to form the elitist PF [59].

6.3.9 Nash Equilibrium-based Decision Making

The best compromise solution should be identified from the resulting PF set to simulate the DM's preference. Previous MOPD algorithms generate the best compromise solution using the fuzzy logic theory [68, 69, 217] in which a simple fuzzy membership function is defined based on experiences without considering the PF's trade-off characteristics. Hence, in the MGSO, the competing objectives are considered as noncooperative decision-making players [188], and the PF's objective fitness can be modeled as the players' set of actions for Nash equilibria of the game theory [187]. Consequently, an alternative multi-criteria decision

making is proposed on the basis of Nash equilibrium to extract an individual with the best joint actions as the best compromise solution.

Based on the concept of Pareto optimality, this equilibrium selection problem with several noncooperative objectives can thus be modeled and transformed to find a Nash equilibrium point [187] of multiobjective players, which involves an optimization problem with probability and rationality constraints to yield a joint probability distribution over the PF's action space, as follows:

$$\begin{aligned}
& \text{Max } Nash(H_1, H_2, \dots, H_i, \dots, H_{M_{\text{obj}}}, v_1, \dots, v_i, \dots, v_{M_{\text{obj}}}) \\
& = \sum_{i=1}^{M_{\text{obj}}} \left(\sum_{j=1}^{M_{\text{pf}}} (-\omega_i f_{ij}) (\prod_{i=1}^{M_{\text{obj}}} h_{ij}) \right) - \sum_{i=1}^{M_{\text{obj}}} v_i \\
& \text{s.t. } \sum_{j=1}^{M_{\text{pf}}} h_{ij} = 1, \quad i = 1, 2, \dots, M_{\text{obj}} \\
& h_{ij} \geq 0, \quad i = 1, 2, \dots, M_{\text{obj}}, \quad j = 1, 2, \dots, M_{\text{pf}} \\
& \sum_{j=1}^{M_{\text{pf}}} (-\omega_i f_{ij}) h_{ij} \leq v_i, \quad i = 1, 2, \dots, M_{\text{obj}}
\end{aligned} \tag{6.38}$$

where $H_i = [h_{i1}, h_{i2}, \dots, h_{ij}, \dots, h_{iM_{\text{pf}}}]$ is equilibrium solution of the i th objective, which represents a probability distribution over the PF's fitness, and h_{ij} is the i th objective's equilibrium value of the j th element in the PF; f_{ij} is the i th objective's normalized fitness of the j th element in the PF; v_i expresses the upper limit of the expectation of fitness values for the i th objective player; M_{pf} is the prespecified size to represent the limited PF set; ω_i is the weight for the relative importance of the i th objective function, and its value could be determined by utilities based on their operation preferences. Here, the weights are set to 1 for the unbiased preference of DM in the case study. The optimization problem formulated in Eq. (6.38) is a standard constrained nonlinear programming (NLP) solved in this algorithm by sequential quadratic programming (SQP) [218], a highly effective and matured method for the NLP. As a result, a list of equilibrium values will be provided for the DM, and the best compromise PF solution can then be derived from the best joint equilibrium which represents the highest payoff outcome obtained from this joint action, as follows:

$$\max \left[\prod_{i=1}^{M_{obj}} h_{i1}, \prod_{i=1}^{M_{obj}} h_{i2}, \dots, \prod_{i=1}^{M_{obj}} h_{ij}, \dots, \prod_{i=1}^{M_{obj}} h_{iM_{pf}} \right] \quad (6.39)$$

6.3.10 Procedures for MGSO

To sum up, the flowchart of the equilibrium-inspired MGSO with synergistic learning is depicted in Fig. 6.1.

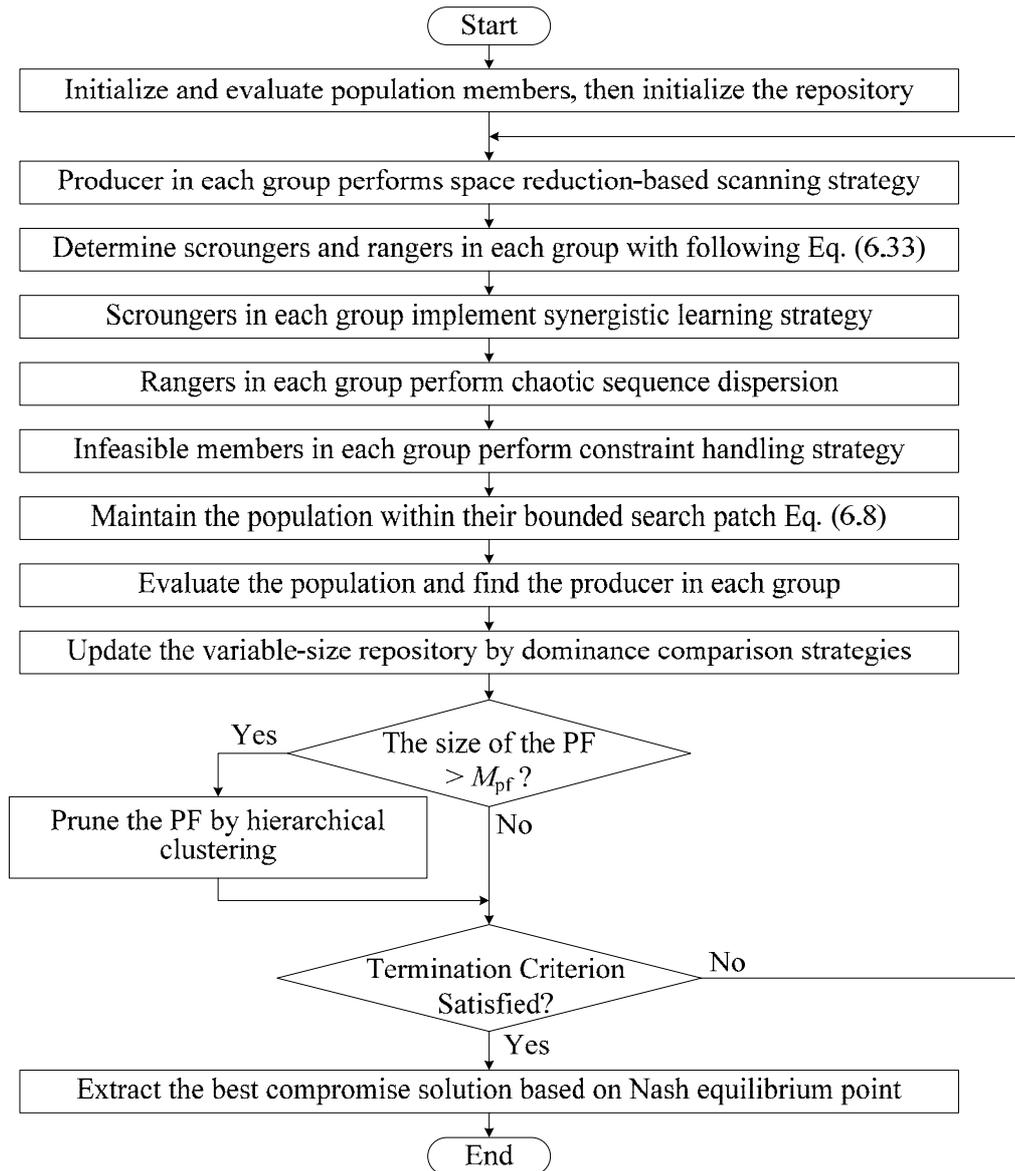


Fig. 6.1 Flowchart of the proposed algorithm for MOPD problems

6.4 EED Studies on IEEE 30-Bus Power System

6.4.1 Experimental Settings

For the purpose of the comparison with previously published algorithms and results, the proposed MGSO is tested on the standard IEEE 30-bus 6-generator power system to investigate the dual-objective EED problem. While the detailed power system data are given in [164, 165, 172] and tabulated in Appendix B, the fuel cost and emission coefficients in Eq. (6.1) and (6.3) are available in [61] and listed in Table B.1 and B.2. As the overall performance of the algorithm is not sensitive to most of its parameters, the parameter setting guidelines in [77] were adopted such that maximum pursuit angle θ_{\max} is set to $\pi/(A_{\max})^2$ and termination criterion $Iter_{\max}$ is fixed to 300. The iteration periods $iter_{L\max}$ and $iter_{R\max}$ were set to 35 and 16, respectively. As the producer in Eq. (6.19)~(6.21) requires three function evaluations, the population size of each group M_p is set to 28 such that the total number of function evaluations in a generation is 60. Besides, for all the optimization runs, the preset size of the PF, M_{pf} , is fixed to 50. Meanwhile, the settings for other parameters in this chapter are heuristically well-tuned as shown in Table 6.1 through a large amount of simulations and comparative studies [208].

Table 6.1 Parameter settings of MGSO for EED of IEEE 30-bus system

Δ_L	Δ_R	$C_{p,\max}$	$C_{p,\min}$	$C_{m,\max}$	$C_{m,\min}$	c_{rp}	c_{rm}	η	w
0.03	0.01	1.20	0.30	1.05	0	0.015	0.015	0.03	0.15

The following are the three study cases [59] being investigated for the MOPD algorithms. The benchmarking was carried out with ten independent optimization runs of MGSO on each case, and the PFs of the best optimization runs from all algorithms were used to compare the algorithm performances [172].

Case Study 1): Lossless system with constraint Eq. (6.7) considered only;

Case Study 2): Transmission losses Eq. (6.5) are also considered in Eq. (6.6);

Case Study 3): Constraints Eq. (6.6), (6.7) and (6.14) are considered.

6.4.2 Case Study 1

The problem constraints include only active generation capacity constraint as well as the power balance constraint without P_{Loss} . Table 6.2 and 6.3 detailed the best solutions for fuel cost and emission obtained by the extreme vertices in the resulting PF of MGSO and other algorithms published in [59, 61, 171, 172]. In addition, the PF obtained from MGSO was plotted in Fig. 6.2 and compared with those from the well-know NSGA-II algorithm [67]. The results demonstrate that MGSO performs well with the two better outer solutions and compares well with other EED algorithms. As indicated from the distribution of solutions, which are found by different groups of the algorithm and marked differently, the synergistic learning of multi-groups for diverse regions of the PF indeed can maintain the diversity over the tradeoff surface.

Table 6.2 Comparison of best fuel cost for Case 1

Case 1	LP [164]	MOSST [171]	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
P_{G1}	0.1500	0.1125	0.1038	0.1116	0.1062	0.1070	0.1073	0.1096
P_{G2}	0.3000	0.3020	0.3228	0.3153	0.2897	0.2897	0.3138	0.2997
P_{G3}	0.5500	0.5311	0.5123	0.5419	0.5289	0.5250	0.5289	0.5243
P_{G4}	1.0500	1.0208	1.0387	1.0415	1.0025	1.0150	0.9931	1.0162
P_{G5}	0.4600	0.5311	0.5324	0.4726	0.5402	0.5300	0.5360	0.5248
P_{G6}	0.3500	0.3625	0.3241	0.3512	0.3664	0.3673	0.3550	0.3594
$F(P_G)$	606.314	605.889	600.34	600.31	600.15	600.1315	600.1763	600.1114
$E(P_G)$	0.22330	0.2222	0.2241	0.2238	0.2215	0.2223	0.22056	0.22216

Table 6.3 Comparison of best emission for Case 1

Case 1	LP [164]	MOSST [171]	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
P_{G1}	0.4000	0.4095	0.4072	0.4146	0.4116	0.4097	0.3982	0.4069
P_{G2}	0.4500	0.4626	0.4536	0.4419	0.4532	0.4550	0.4503	0.4576
P_{G3}	0.5500	0.5426	0.4888	0.5411	0.5329	0.5363	0.5527	0.5377
P_{G4}	0.4000	0.3884	0.4302	0.4067	0.3832	0.3842	0.3817	0.3840
P_{G5}	0.5500	0.5427	0.5836	0.5318	0.5383	0.5348	0.5416	0.5373
P_{G6}	0.5000	0.5152	0.4707	0.4979	0.5148	0.5140	0.5096	0.5105
$F(P_G)$	639.600	644.112	633.83	636.04	638.51	638.3577	637.5980	638.1992
$E(P_G)$	0.19424	0.19418	0.1946	0.1943	0.1942	0.1942	0.19422	0.19420

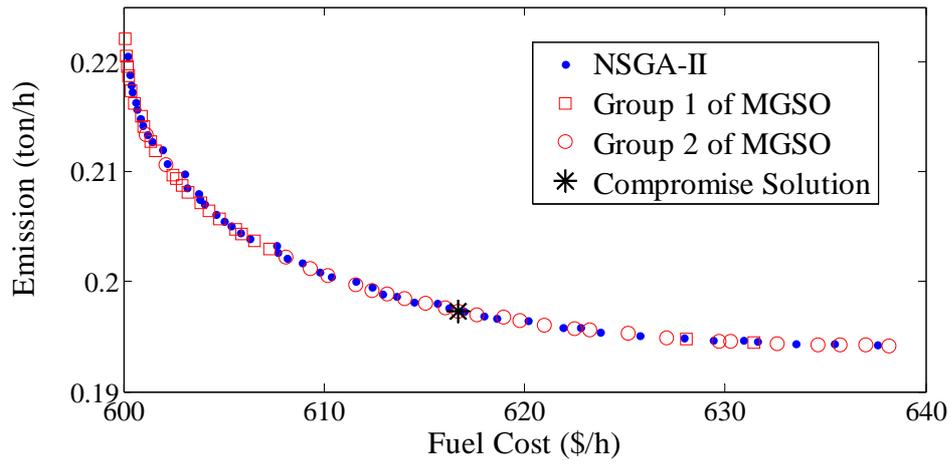


Fig. 6.2 Comparison of the PFs obtained for Case 1

6.4.3 Case Study 2

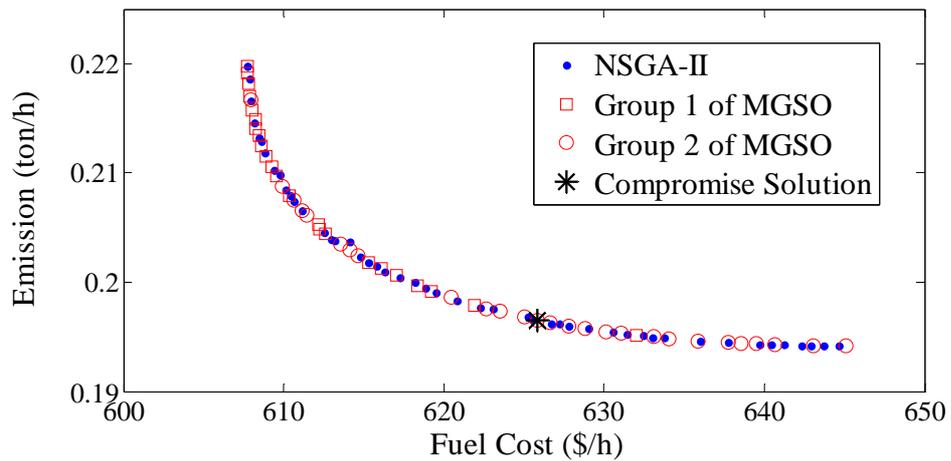


Fig. 6.3 Comparison of the PFs obtained for Case 2

Table 6.4 Comparison of best fuel cost for Case 2

Case 2	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
P_{G1}	0.1447	0.1425	0.1086	0.1130	0.1147	0.1155
P_{G2}	0.3066	0.2693	0.3056	0.3145	0.3093	0.3059
P_{G3}	0.5493	0.5908	0.5818	0.5826	0.5895	0.5974
P_{G4}	0.9894	0.9944	0.9846	0.9860	0.9820	0.9804
P_{G5}	0.5244	0.5315	0.5288	0.5264	0.5164	0.5136
P_{G6}	0.3542	0.3392	0.3584	0.3450	0.3555	0.3543
$F(P_G)$	607.98	608.06	607.807	607.7862	607.7684	607.7633
$E(P_G)$	0.2191	0.2207	0.22015	0.22010	0.21978	0.21980

Table 6.5 Comparison of best emission for Case 2

Case 2	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
P_{G1}	0.3929	0.4064	0.4043	0.4063	0.4113	0.4103
P_{G2}	0.3937	0.4876	0.4525	0.4586	0.4547	0.4628
P_{G3}	0.5818	0.5251	0.5525	0.5510	0.5471	0.5435
P_{G4}	0.4316	0.4085	0.4079	0.4084	0.3913	0.3904
P_{G5}	0.5445	0.5386	0.5468	0.5432	0.5490	0.5438
P_{G6}	0.5192	0.4992	0.5005	0.4974	0.5113	0.5141
$F(P_G)$	638.98	644.23	642.603	642.8964	644.6877	645.1108
$E(P_G)$	0.1947	0.1943	0.19422	0.19420	0.19419	0.19418

Table 6.4 and 6.5 listed the best results obtained for Case 2 with transmission losses included. As shown in the plot of MGSO and NSGA-II results in Fig. 6.3, the frontier of MGSO are well-distributed and dispersedly covered the entire PF of NSGA-II. This verifies that the Pareto solutions can effectively be solved by MGSO.

6.4.4 Case Study 3

The maximum line flow limits used in Case 3 are 115% of their rating values as given in Table B.4 [172]. From the simulation results presented in Fig. 6.4, Table 6.6 and 6.7, the performance of NSGA-II deteriorates with the increase of problem complexity while MGSO performs well with the outstanding diversity characteristics and full extent of the PF.

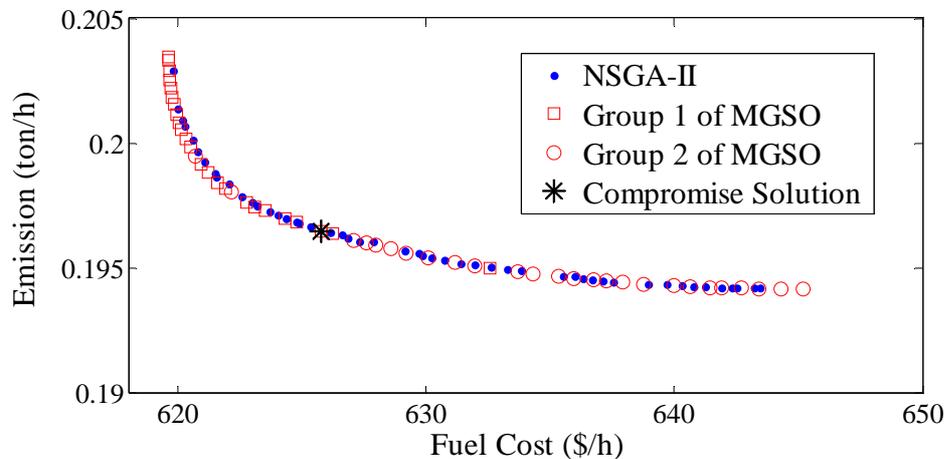


Fig. 6.4 Comparison of the PFs obtained for Case 3

Table 6.6 Comparison of best fuel cost for Case 3

Case 3	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
P_{G1}	0.1358	0.1127	0.15975	0.1596	0.1646	0.1775
P_{G2}	0.3151	0.3747	0.35339	0.3535	0.3757	0.3588
P_{G3}	0.8418	0.8057	0.79600	0.7974	0.7108	0.7448
P_{G4}	1.0431	0.9031	0.97176	0.9719	0.5861	0.5913
P_{G5}	0.0631	0.1347	0.08684	0.08624	0.6229	0.5996
P_{G6}	0.4664	0.5331	0.49709	0.49609	0.3995	0.3870
$F(P_G)$	620.87	620.46	620.165	620.18	619.8271	619.6269
$E(P_G)$	0.2368	0.2243	0.22826	0.2283	0.20289	0.20346

Table 6.7 Comparison of best emission for Case 3

Case 3	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
P_{G1}	0.4403	0.4753	0.47975	0.47969	0.4188	0.4102
P_{G2}	0.4940	0.5162	0.52868	0.5287	0.4457	0.4630
P_{G3}	0.7509	0.6513	0.67109	0.67116	0.5424	0.5436
P_{G4}	0.5060	0.4363	0.53174	0.5318	0.4069	0.3896
P_{G5}	0.1375	0.1896	0.12571	0.1257	0.5465	0.5438
P_{G6}	0.5364	0.5988	0.53010	0.5299	0.5045	0.5147
$F(P_G)$	649.24	657.59	651.633	651.62	643.4741	645.1976
$E(P_G)$	0.2048	0.2017	0.20470	0.2047	0.19423	0.19418

6.4.5 Solution Quality Analysis

From the optimization runs, it can be found that, with the help of the space reduction-based scanning, the MGSO is able to find Pareto solutions with better fitness on each objective compared to other algorithms. This confirms the Pareto optimality of the proposed algorithm and its potential to find solutions covering the entire true PF.

In order to estimate the spread of the PFs, a performance metric in [184] is used to measure the normalized distance in the objective space between the PF's two extreme solutions. The averages of this metric over ten independent runs for different algorithms are listed in Table 6.8 and further confirm that the proposed MGSO outperforms the other five multiobjective algorithms for all cases.

Moreover, as shown in Fig. 6.4, the solutions of MGSO are more widely and evenly distributed than the ones of NSGA-II. This validates that the proposed constraint handling strategy for infeasible members does improve the uniformity of the PF's distribution. Also, compared to NSGA, NPGA, SPEA and NSGA-II on function evaluation analysis [219], FCPSO and MGSO require much fewer function evaluations to form the optimal front. Coupled with its high exploratory capability, the proposed MGSO can give the best performance but with less number of fitness function evaluations.

Table 6.8 Normalized distance measure of different algorithms

	NSGA [172]	NPGA [172]	SPEA [59]	FCPSO [61]	NSGA-II [67]	MGSO
Case 1	0.93757	0.95001	0.93809	0.95012	0.94210	0.96539
Case 2	0.92211	0.93747	0.94509	0.94531	0.94829	0.95845
Case 3	0.85539	0.81312	0.85363	0.85358	0.87961	0.90133

6.4.6 Compromise Solutions

Table 6.9 Compromise solutions of MGSO for different decision making

	Case 1		Case 2		Case 3	
	Fuzzy	Equilibrium	Fuzzy	Equilibrium	Fuzzy	Equilibrium
P_{G1}	0.25895	0.31427	0.26880	0.31633	0.31496	0.31496
P_{G2}	0.37878	0.41163	0.39617	0.40981	0.41595	0.41595
P_{G3}	0.52849	0.52786	0.54228	0.56694	0.55377	0.55377
P_{G4}	0.71059	0.60474	0.70125	0.56448	0.56393	0.56393
P_{G5}	0.51670	0.52964	0.52423	0.53808	0.56120	0.56120
P_{G6}	0.44049	0.44586	0.43185	0.46775	0.45378	0.45378
$F(P_G)$	609.3465	616.6989	617.0303	625.8472	625.7582	625.7582
$E(P_G)$	0.20119	0.19732	0.20058	0.19646	0.19651	0.19651

Lastly, the best compromise solutions of MGSO solved by the proposed Nash equilibrium point were marked with * in Fig. 6.2~6.4, and Table 6.9 presents the compromising solutions obtained for the three cases with the fuzzy logic decision making [61, 69] and the proposed method. It is interesting to note that the results

in Case 1 and 2 are different while the compromising solution for Case 3 is the same. This confirms that the proposed equilibria-based decision making is able to generate a reasonable bargaining solution for power system dispatcher.

6.5 Computational EEED Studies on IEEE 118-Bus System

6.5.1 Experimental Settings

For in-depth investigation of the proposed algorithm on larger power systems, a modified IEEE 54-generator 118-bus power system is used for a tri-objective EEED optimization including fuel cost, emission and energy saving objectives in Eq. (6.1)~(6.5). The power system topologies and generator data as well as the constraint data are available from [220] and listed in Appendix C. The total system demand is 4,242MW. Table C.1~C.3 list the voltage ratings, fuel cost and emission coefficients for the 54 generators. In addition, for unit 1~24, a POZ is set in the middle of its generation capacity patch for each generator, and the POZ interval is 10% of the generator's original feasible operating zone. In this study, all system constraints in Section 6.2.2 were considered. Besides, except for the modified settings listed in Table 6.10, settings in Section 6.4.1 were also adopted for the IEEE 118-bus system.

Table 6.10 Parameter settings of MGSO for EEED of IEEE 118-bus system

$Iter_{\max}$	M_{pf}	M_p	SP_R (p.u.)	c_{rp}	c_{rm}	η	w
2000	150	98	38.0	0.025	0.025	0.01	0.10

For further comparison and discussion, the two advanced MOEA algorithms, namely NSGA-II and improved SPEA (SPEA2) [221], were also implemented and considered here. Numerous trials have been carried out to determine the optimum settings for these two evolutionary algorithms, and their population size and maximum number of generations are set to 300 and 2000, respectively, such that the number of fitness function evaluations of the algorithms is the same with

MGSO for fair comparisons. Moreover, the probabilities of crossover and mutation are selected as 0.9 and 0.01, respectively [32-34].

6.5.2 Comparative Results, Analysis and Discussion

The PFs resulted from the above three MOPD algorithms should be assessed systematically with the performance metrics derived from the three basic quality measure criteria stated in Section 2.4 and the true Pareto front of EEED problem. However, since the true PF is difficult to determine and guarantee for problems with high-dimensional and highly complex search space, a pseudo PF, named as reference PF [31], is used instead as true PF here to compare the PFs generated by multiobjective algorithms. The reference PF can be formed by 50 independent runs of all the MOPD algorithms, NSGA-II, SPEA2 and MGSO, i.e. 150 sets of nondominated solutions. All these 150 sets of PF solutions were then combined and ranked by dominance comparisons so as to generate the reference PF, which consists of 1688 nondominated solutions as illustrated in Fig. 6.5. Furthermore, the PF of the overall best run in the 50 optimization runs of each algorithm, as shown in Fig. 6.6~6.8, is selected for further analysis.

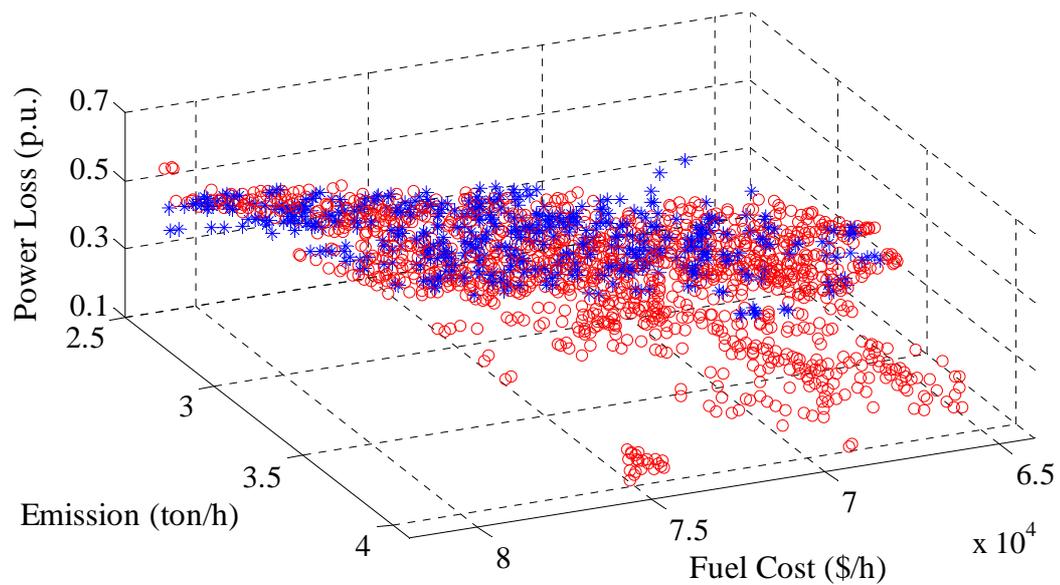


Fig. 6.5 Reference PF surface (dot mark: NSGA-II; star mark: SPEA2; circle mark: MGSO)

For the reference PF shown in Fig. 6.5, approximately 68.48%, 31.52% and 0% of the Pareto solutions are contributed by MGSO, SPEA2 and NSGA-II, respectively. It can be observed that all the solutions found with NSGA-II are covered by those with SPEA2 and MGSO, and MGSO has contributed the majority of the reference PF solutions. This also reveals that the solutions of MGSO are closer to the true Pareto set. From Fig. 6.6~6.8, it is clear that, in comparison with NSGA-II and SPEA2, MGSO can obtain wider spread of solutions and better scattered PF. In particular, the results from MGSO in Fig. 6.8 indicate that the multi-groups can effectively seek out the Pareto solutions in the separated feasible islands, and this also validates its superior efficiency of solution searching for complex nonlinear constrained problems with high-dimensional search space. In addition, for the best run of MGSO in Fig. 6.8, the simulation results in the convergence process, including the convergence of each objective function, ranger percentages and maximum pursuit distances are shown in Fig. 6.9 and 6.10. The experimental results also confirmed that different types of functional members do facilitate the developed algorithm to effectively propagate the search towards the well-spread and diverse PF.

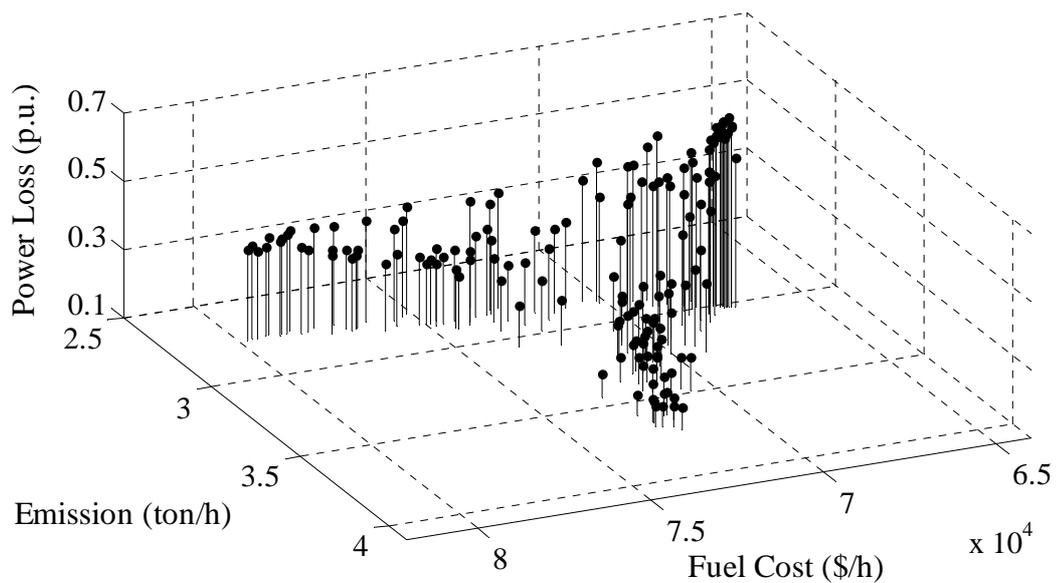


Fig. 6.6 Best PF solutions obtained with NSGA-II

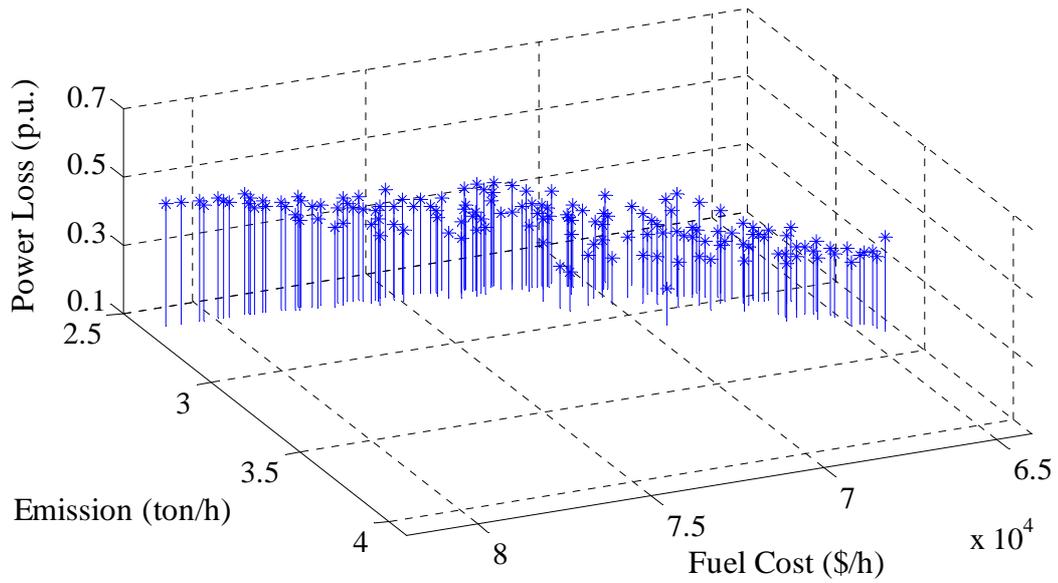


Fig. 6.7 Best PF solutions obtained with SPEA2

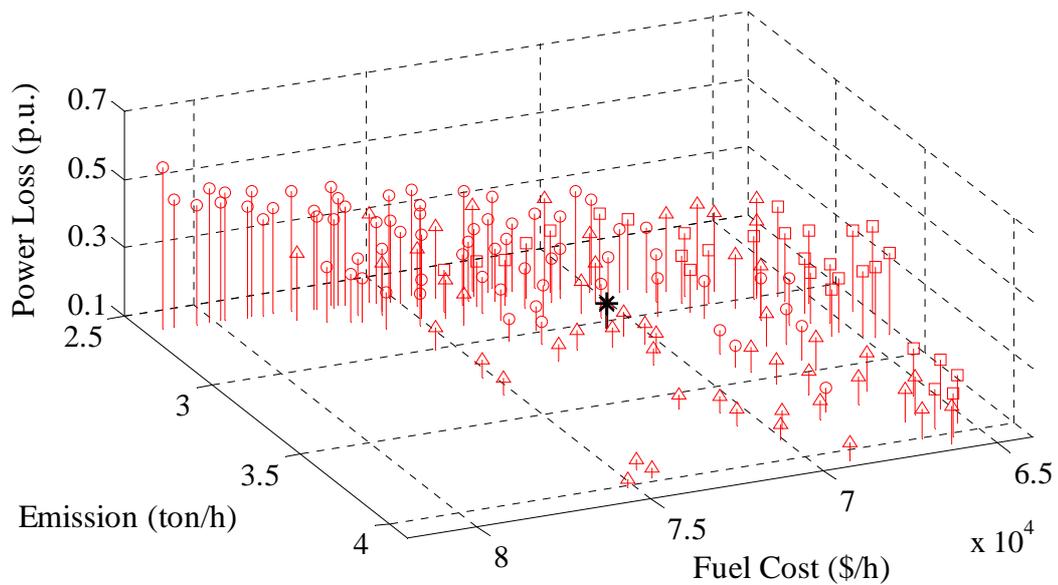


Fig. 6.8 Best PF solutions obtained with MGS0 (square mark: group 1; circle mark: group 2; triangle mark: group 3; star mark: compromise solution)

In this chapter, three typical performance metrics were used to compare and analyze the solution quality of PFs obtained from various MOPD algorithms. The first is the convergence metric [216] adopted to measure the degree of closeness between the obtained PF and the reference PF because of its simplicity of use and clarity of definition. For each PF solution obtained, the Euclidean distance from it to the nearest solution of the reference PF in the objective space is calculated

firstly, and the convergence metric can then be obtained using the average value of these distances.

Secondly, the distribution uniformity of the PF solutions can be assessed by the spacing metric [67] which is calculated as the relative crowding distance S_d between consecutive solutions in the resulting PF set as follows:

$$S_d = \sqrt{\frac{1}{M_{\text{nbs}}} \sum_{n=1}^{M_{\text{nbs}}} (d_{cn} - d_{c,\text{avg}})^2} \quad (6.40)$$

where $d_{c,\text{avg}}$ is the average of all crowding distances d_{cn} , $n = 1, 2, \dots, M_{\text{nbs}}$, and M_{nbs} is the number of the non-boundary solutions in the PF. The desired value for this metric is 0, which means the elements of PF solutions can be equidistantly spaced. Thirdly, as explained in Section 6.4.5, the extent of PF can be assessed with the normalized distance of extreme solutions for the three MOPD objectives.

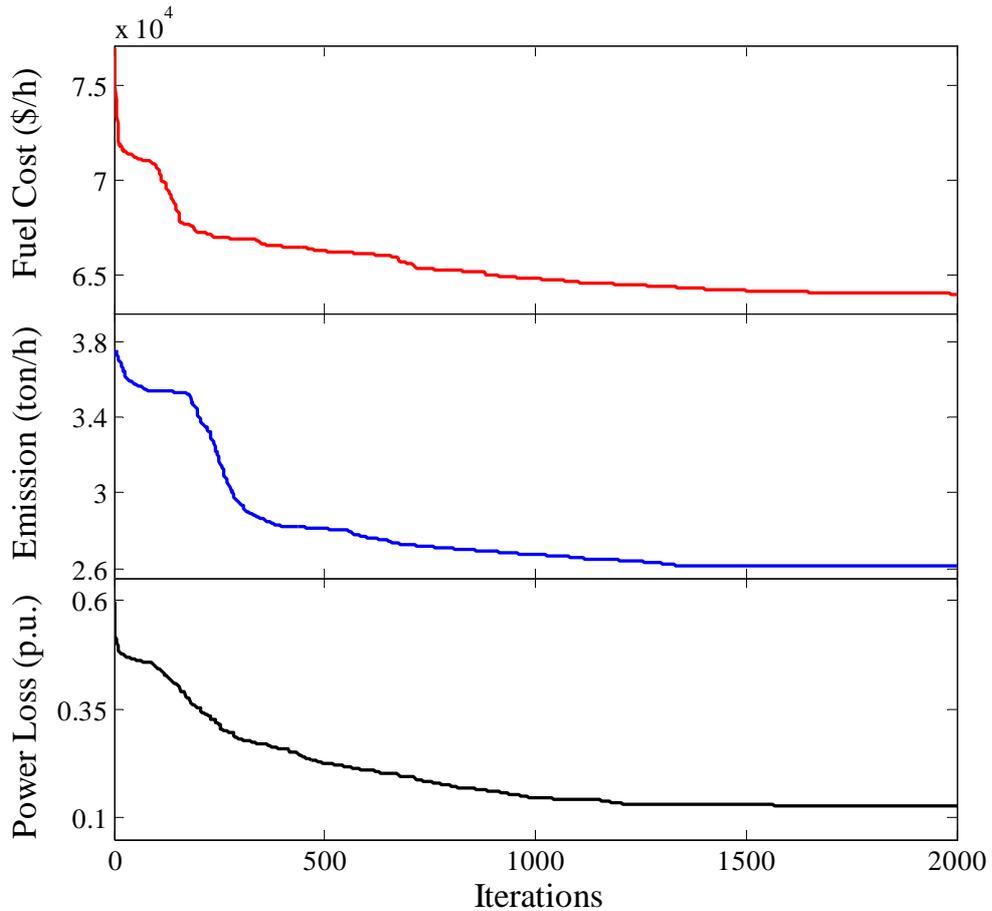


Fig. 6.9 Producer's convergence of each group objective

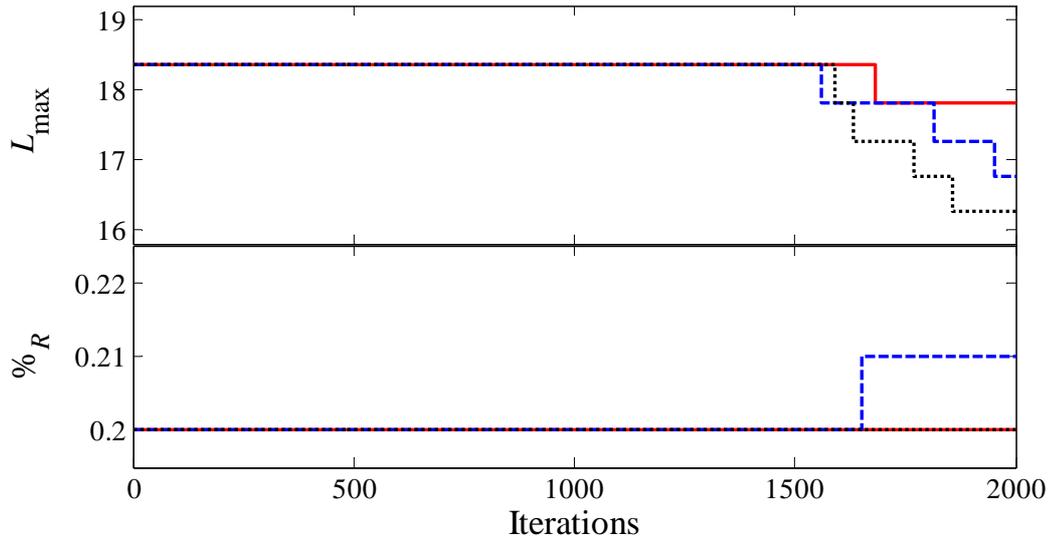


Fig. 6.10 Maximum pursuit distance and ranger percentage of each group in the convergence process (solid line: group 1; dash line: group 2; dot line: group 3)

Table 6.11 tabulates the best solutions for fuel cost, emission, and power loss obtained from the best run of each algorithm shown in Fig. 6.6~6.8. It can be found that, in terms of the best results for each objective, MGSO can find more optimal extreme solutions to maintain a widespread Pareto set over the entire true PF region. Meanwhile, the performance measures of the overall best run of each algorithm are given in Table 6.12, and the resulting statistics on the convergence, spacing and normalized distance metrics over the 50 optimization runs are listed in Table 6.13~6.15, respectively. The statistical results reveal that, with the same number of fitness function evaluations, MGSO can markedly outperform the two earlier MOEAs, and provides satisfactory performance on various Pareto metrics, especially on the convergence and normalized distance measures.

Table 6.11 Comparison of best solutions for cost, emission and system loss

	NSGA-II			SPEA2			MGSO		
	Best Cost	Best Emission	Best Loss	Best Cost	Best Emission	Best Loss	Best Cost	Best Emission	Best Loss
P_{G1}	0.80948	0.84045	0.86252	0.78758	0.78833	0.80867	0.74845	0.78085	0.72239
P_{G2}	0.75667	0.73915	0.73955	0.75233	0.76964	0.79850	0.74474	0.76074	0.75670
P_{G3}	0.42762	0.42711	0.35031	0.72894	0.75702	0.73766	0.75765	0.79534	0.67587
P_{G4}	0.64978	0.67181	0.36543	0.66449	0.78321	0.76971	0.75913	0.75657	0.45000
P_{G5}	1.03647	0.65580	0.88960	0.03740	0.39996	0.02667	1.01994	0.45531	0.09824

P_{G6}	1.10115	1.17702	1.35370	0.50369	0.48576	0.50839	0.50507	0.58498	1.81110
P_{G7}	0.65123	0.76760	0.76059	0.99784	0.77049	0.97726	0.95866	0.77333	0.97846
P_{G8}	0.44534	0.65061	0.64810	0.57781	0.75525	0.85099	0.99940	0.77101	0.67117
P_{G9}	0.82597	0.83584	0.83123	0.98992	0.77196	0.98387	0.97719	0.77586	0.76452
P_{G10}	0.85827	0.69759	0.27954	0.98739	0.73889	0.68794	0.98230	0.76992	0.22894
P_{G11}	1.31544	0.96681	0.04979	0.75034	1.36378	0.23050	0.74518	1.76001	0.09093
P_{G12}	0.92568	0.66631	0.76115	0.88504	0.63861	0.29217	0.01619	0.72401	0.13482
P_{G13}	0.44401	0.65877	0.37474	0.98925	0.81072	0.91967	0.99461	0.83285	0.95033
P_{G14}	0.89826	0.78658	0.81484	0.86904	0.60565	0.91941	0.88797	0.65236	0.67120
P_{G15}	0.92114	0.92612	0.90215	0.90312	0.84719	0.80713	0.89564	0.87747	0.78182
P_{G16}	0.80327	0.68594	0.42311	0.91610	0.96068	0.92911	0.89478	0.87238	0.86258
P_{G17}	0.87401	0.85052	0.86254	0.88890	0.93308	0.93073	0.89744	0.88128	0.81568
P_{G18}	0.90285	0.97906	0.97076	0.97473	0.99248	0.94031	0.90272	0.87583	0.99375
P_{G19}	0.90573	0.90901	0.86629	0.93515	0.98564	0.95785	0.91231	0.89898	0.99433
P_{G20}	0.81906	0.91135	0.96474	0.79751	0.66162	0.76254	0.82181	0.50198	0.76231
P_{G21}	0.93964	0.64443	2.06825	1.10515	0.56650	1.19427	1.12156	0.54504	1.96112
P_{G22}	0.65542	0.66198	1.08344	1.46625	0.50454	0.87519	1.47762	0.46665	1.41607
P_{G23}	0.80711	0.84795	0.87241	0.74993	0.88983	0.79896	0.80552	0.84775	0.77138
P_{G24}	0.85418	0.87302	0.90812	0.74251	0.91620	0.96607	0.81959	0.84326	0.94078
P_{G25}	1.12072	1.75493	2.31877	1.51642	1.47481	1.72985	1.85326	1.42155	2.53576
P_{G26}	1.00649	1.06999	1.40456	1.47125	0.95238	1.37838	1.01485	0.85412	0.89165
P_{G27}	0.75150	0.84611	0.68705	0.86575	0.87414	0.86889	0.84836	0.82999	0.92799
P_{G28}	1.44691	0.68314	1.16578	1.88697	0.69066	1.68078	1.50011	0.63309	0.52916
P_{G29}	0.80233	1.22254	0.70439	0.01624	1.47159	1.49834	0.73357	1.44356	0.60491
P_{G30}	0.60725	1.09120	0.89879	0.00188	1.00489	0.81014	0.00845	0.99471	0.81291
P_{G31}	0.19555	0.83678	0.36140	0.04808	0.92147	0.44051	0.02577	0.90859	0.91771
P_{G32}	0.21819	0.61924	0.27975	0.03761	0.88289	0.16391	0.04905	0.88797	0.11755
P_{G33}	0.24555	0.77721	0.29332	0.00104	0.85174	0.18549	0.08277	0.89347	0.08327
P_{G34}	0.55747	0.62962	0.59346	0.57901	0.71898	0.53912	0.59563	0.73946	0.59470
P_{G35}	0.67570	0.80385	0.80825	0.91894	0.89547	0.92290	0.92163	0.90497	0.92707
P_{G36}	0.83601	0.84688	0.85278	0.99486	0.95615	0.98148	0.99736	0.91526	0.99786
P_{G37}	1.32874	0.64748	2.86540	2.02229	0.67973	1.93127	1.63543	0.63388	3.36588
P_{G38}	0.77572	0.79446	0.74840	0.74881	0.82996	0.88484	0.78557	0.80768	0.89856
P_{G39}	0.48294	0.39866	0.28885	0.02025	0.42194	0.27322	0.02974	0.43542	0.13029
P_{G40}	0.94810	1.20705	1.15391	0.89806	1.02777	0.94228	0.86704	0.98974	1.19254
P_{G41}	0.81360	0.79442	0.84690	0.76395	0.73281	0.73113	0.75109	0.70823	0.99991
P_{G42}	0.74546	0.74209	0.50493	0.74902	0.70756	0.75193	0.76934	0.71840	0.44458
P_{G43}	0.50631	0.38811	0.36299	0.36906	0.47169	0.35759	0.37907	0.51631	0.35542
P_{G44}	0.74319	0.66865	0.62350	0.71749	0.71277	0.73499	0.77736	0.70893	0.40362
P_{G45}	1.71778	1.54197	1.55313	1.77699	1.42717	1.27394	0.90188	1.41896	1.47011
P_{G46}	1.00609	0.50175	0.40332	0.02469	0.33478	0.09787	0.00283	0.41330	0.24505
P_{G47}	0.01773	0.52401	0.40323	0.78567	0.82804	0.74074	0.91334	0.90805	0.34214
P_{G48}	0.78292	0.52908	0.43381	0.80445	0.77749	0.15622	0.89071	0.81450	0.62452
P_{G49}	0.93352	0.81579	0.54891	0.89996	0.67243	0.59006	0.88319	0.72618	0.55597
P_{G50}	0.88193	0.67332	0.41231	0.88543	0.69164	0.86544	0.88026	0.73960	0.42933
P_{G51}	0.57815	0.46397	0.35011	0.31974	0.38400	0.23814	0.28811	0.39721	0.03406
P_{G52}	0.85821	0.62301	0.41259	0.90070	0.52294	0.49069	0.89972	0.59980	0.64806
P_{G53}	0.88800	0.67538	0.43999	0.89139	0.61589	0.85388	0.92036	0.60533	0.19887
P_{G54}	0.85991	0.81254	0.85766	0.87863	0.64970	0.84110	0.92583	0.62374	0.96185
$F(P_G)$	67032.4	78447.6	72454.3	64296.4	81370.4	69388.8	64029.3	81443.0	75189.0
$E(P_G)$	3.03191	2.74248	3.67364	3.34088	2.62034	3.09897	3.21255	2.61602	4.01564
P_{Loss}	0.53977	0.39406	0.16111	0.37501	0.46053	0.20868	0.35717	0.57575	0.12580

It should be pointed out that the MGSO can always find the nondominated solutions in the separated feasible space so as to guarantee the diversity of the PF, and therefore the spacing metrics of MGSO are greater than those of NSGA-II and SPEA2. Besides, it can also be found that the PFs from NSGA-II are more uniformly-spaced, but perform worst on the other two metrics. Also, the variance and standard deviation values of these metrics indicate the stable performance of MGSO for the resulting Pareto set.

Table 6.12 Performance measures of the best run of each algorithm

Algorithms	Convergence	Spacing metric	Normalized distance
NSGA-II	3.217613	0.012542	0.657604
SPEA2	2.005540	0.018667	0.978645
MGSO	0.937165	0.023775	0.998119

For the investigation of contribution of the scanning-space reduction strategy for producer and the adaptive ranger percentage with chaotic sequence, statistical results were collected over 50 runs for the MGSO with fixed L_{imax} (referred as MGSO-I) and the MGSO with the fixed $\%_{iR}$ and the random number sequence [77] (referred as MGSO-II), and tabulated in Table 6.13~6.15. It is worth noting that the normalized distances of the PFs can be statistically improved with the help of the space reduction strategy. Furthermore, referring to MGSO-II, it also verified that the proposed chaotic sequence dispersion can statistically enhance the overall performance of MGSO, especially on the convergence metric.

Table 6.13 Resulting statistics of convergence metrics in 50 runs

Algorithms	Best	Worst	Average	Variance	Std. Dev.
NSGA-II	2.277862	3.636170	2.877324	0.108673	0.329656
SPEA2	1.334764	3.211466	1.956421	0.273424	0.522899
MGSO	0.810330	2.153011	1.237079	0.095313	0.308729
MGSO-I	0.857980	2.145775	1.296715	0.101716	0.318930
MGSO-II	0.913451	2.315498	1.423473	0.166899	0.408533

Table 6.14 Resulting statistics of spacing metrics in 50 runs

Algorithms	Best	Worst	Average	Variance	Std. Dev.
NSGA-II	0.008896	0.053966	0.013163	4.3963E-5	0.006631
SPEA2	0.016049	0.081216	0.033993	5.8215E-4	0.024128
MGSO	0.017113	0.070548	0.030553	2.8154E-4	0.016778
MGSO-I	0.018596	0.069413	0.031077	2.2648E-4	0.015048
MGSO-II	0.018071	0.087473	0.030929	3.0917E-4	0.017583

Table 6.15 Resulting statistics of normalized distance measures in 50 runs

Algorithms	Best	Worst	Average	Variance	Std. Dev.
NSGA-II	0.657604	0.136544	0.390349	0.015487	0.124446
SPEA2	0.978645	0.488013	0.754910	0.035914	0.189510
MGSO	0.998119	0.653782	0.826710	0.010782	0.103836
MGSO-I	0.965013	0.611423	0.793783	0.014355	0.119814
MGSO-II	0.981434	0.678453	0.811077	0.015388	0.124051

The ultimate goal of any Pareto-based multiobjective algorithms is to identify a unique solution expressing the best compromise among multiple objectives. In the proposed MGSO, the solution having the maximum joint equilibrium value will be chosen as the PF's best compromise solution as marked in Fig. 6.8. The fitness of the best compromise solution obtained with Nash equilibrium-based decision making method is (70850.4, 3.04343, 0.17382), as compared to solution (68189.8, 3.07348, 0.20682) obtained by the fuzzy method [61]. It can be seen that the two compromise solutions are quite different in this case study because the proposed method takes into account the objectives' trade-off of PF solutions and its solution model is on the basis of the compromising between the gain of one objective and the degradation in other objectives [186] with solid technical foundation based on the non-cooperative game theory [187]. In conclusion, the proposed MCDM is applicable to handle the complex trade-off of PF solutions with many objective functions.

From the investigations and analysis above, it can be found that, though the performance improvement of MGSO algorithm is moderate compared with other algorithms on the small IEEE 30-bus system, the proposed MGSO has exhibited its superior capability to provide largely improved solution for the larger 118-bus system with limited number of fitness function evaluations and can significantly enhance the searching capability, ensures the quality of Pareto solutions, and also efficiently manages the highly complex power system constraints in solving the high-dimensional MOPD problems with more objectives.

A comparative study of the average run time per generation over the 50 optimization runs for each of the three MOPD algorithms is given in Table 6.16. All the algorithms were implemented in Matlab 7.6 and executed on a personal computer with 3.2 GHz Intel Core 2 Quad CPU and 4GB RAM. It is quite evident that the computation time of MGSO is less than that of the other two techniques.

Table 6.16 Run time per generation of different algorithms

	NSGA-II	SPEA2	MGSO
Run time (s)	7.895	8.322	7.655

As the main searching force members, scroungers perform the proposed synergistic learning strategy and their behaviors are mainly determined by reinforcement factor η which is crucial to the credit assignment and information interaction in the cooperative searching process. In order to investigate the tuning rules for design parameters on the search performance of MGSO, the sensitivity of η was studied over a range from 0.001 to 0.01 in step of 0.001. Table 6.17 tabulates the statistical results of the Pareto performance measures for MGSO with various reinforcement factors over 50 optimization runs. It can be seen that a large value of η can effectively enhance the interactive cooperation between searching groups and hence can improve the average performances of MGSO algorithm. Therefore, the overall best PF results can be achieved when η was set

to 0.01. Similarly, the other algorithm parameters can also be heuristically fine-tuned, as listed in Table 6.1 and 6.10, using this cut-and-try approach [208].

However, it shall also be noted that, according to the No Free Lunch theorem, “for any search algorithm, any elevated performance over one class of problems would be exactly paid for in performance over another class” [66]. Therefore, to solve a given optimization problem requires finding a suitable algorithm which is better than other algorithms for this problem, and in this thesis, this implies that the proposed MGSO is more superior in this class of power system dispatch problems.

Table 6.17 Reinforcement factor effects on performance of MGSO

η	Convergence		Spacing metric		Normalized distance	
	Average	Std. Dev.	Average	Std. Dev.	Average	Std. Dev.
0.001	4.281376	0.239019	0.049208	0.021571	0.614183	0.293906
0.002	3.336863	0.528563	0.058155	0.031308	0.563386	0.145856
0.003	3.393176	0.391236	0.039840	0.026977	0.627215	0.191327
0.004	3.490579	0.263235	0.030686	0.019706	0.742178	0.293504
0.005	3.019370	0.429266	0.025169	0.017366	0.512825	0.168161
0.006	2.675614	0.325410	0.029576	0.018761	0.726275	0.090572
0.007	2.312698	0.309755	0.029543	0.020696	0.791573	0.121694
0.008	1.568162	0.340045	0.037298	0.019274	0.755257	0.175619
0.009	1.493507	0.319943	0.031997	0.018536	0.808960	0.119370
0.01	1.237079	0.308729	0.030553	0.016778	0.826710	0.103836

6.6 Summary

In this chapter, a novel Pareto optimization algorithm, MGSO, is developed to solve the highly nonlinear constrained and large-scale MOPD problems. The following are main advantages of the proposed approach:

- (1) In the algorithm, four categories of group members, in association with their searching strategies, are designed for effective exploration and formation of the PF and improving the extension, convergence, uniformity and diversity of the resulting Pareto solutions.

- (2) A synergistic learning mechanism based on the stochastic learning automata is firstly proposed for credit assignment and information interaction among multiple groups to achieve the cooperative search for Pareto set.
- (3) A new decision-making criterion on the basis of Nash equilibrium point is presented to identify a more reasonable compromising solution from the obtained PF with multiple contradictory objectives.

The proposed MGSO has been successfully applied to the bi-objective EED problem on the small IEEE 30-bus system and a tri-objective EEED problem on the large IEEE 118-bus test system. In-depth numerical simulation studies have confirmed the superior efficiency of the MGSO for PF solution searching and the effectiveness of the proposed new searching strategies. Compared to previously published algorithms on various performance measures, the proposed MGSO is very competitive in the small dual-objective EED problems and clearly superior in the large-scale high-dimensional MOPD problems with complex constraints and objectives.

Chapter VII

Conclusions and Future Work

7.1 Conclusion

The contemporary energy crisis impels most of countries all over the world to implement their energy saving and emission reduction policies for environmental protection and green economy. New issues have since been raised in electricity industry to accommodate the energy-saving objective and need to be addressed through updated multiobjective generation dispatch strategies. In the meanwhile, medium-size local power networks are increasingly being interconnected for the safety of operations and reliability of demand-supply balance. Heuristically, such large-scale interconnections have given rise to difficulties in the management and optimization of the high-dimensional power system dispatch. This multiobjective generation dispatch for modern power systems is a relatively new topic and has received remarkable attention from international academic researchers. Therefore, multiobjective AGC and MOPD strategies are the central themes of this thesis, with the innovative design and considerable improvements on the foundation of pioneering research studies already reported. Specifically speaking, the primary contributions of this research can be summarized as the following aspects:

- i) A novel optimal relaxed AGC control philosophy has been proposed for solving the AGC problem under NERC's CPS standards*

NERC's new CPS provides a theoretical basis for relaxation of LFC system as well as benefit coordination between dispatch center and AGC plants. Hence, this thesis explicitly defines the optimal relaxed AGC objectives as follows: (i) to

comply with CPS assessment standards in various complex operation scenarios of power systems; (ii) to allow the AGC plants to maneuver with least wear-and-tear and control actions for further cost saving in the generation sectors; (iii) to develop a reasonable control structure having a high level of practicability and applicability for real power systems.

ii) Three RL algorithms were investigated and adopted in the design of optimal AGC strategies under CPS

In this thesis, various RL methods were thoroughly examined and three stochastic optimal AGC strategies under CPS, including QAGC, $Q(\lambda)$ AGC and $R(\lambda)$ ILAGC, were designed as a result. On a long-run basis with the relaxed control objectives transformed to MDP reward functions, the optimized control strategy can be gradually obtained for the maximization of the long-term cumulative rewards in the procedure of interactive self-learning processes. Simulation studies revealed that, compared to the NARI's PI controller, the RL-based AGCs can effectively enhance the adaptability and dynamic performance of AGC systems, and reduce the number of pulses and pulse reversals while CPS metrics are assured under the complicated and dynamic operational conditions.

iii) Backward estimation technique with eligibility traces is utilized to overcome the problem of long time-delay control in thermal-dominated AGC process

For the thermal-dominated AGC control strategies, one of main challenges is the long time-delayed feedback problem caused by the steam turbine of thermal AGC plants. This research firstly applies the backward estimation with eligibility traces in $Q(\lambda)$ learning and $R(\lambda)$ learning algorithms to effectively cope with the time-delay control problem of thermal generators in non-Markov environment. Furthermore, the moving averages of CPS1 and ACE are used as the state inputs for further alleviating the over-compliant problem, and the AGC strategies also can provide a feasible mean to achieve the desirable degree of CPS compliances and relaxation of AGC system by online tuning the relaxation factors.

iv) A novel imitation pre-learning method is proposed for the RL-based AGCs to facilitate the transition from the conventional offline pre-learning process to online learning

The conventional RL-based AGCs cannot handle their offline pre-learning processes for onsite implementation if there are large differences existed between the system simulation model and real power system. Thus, another important contribution of this thesis is the proposed imitation online pre-learning technique in combination with the AROC-based $R(\lambda)$ learning, in which the $R(\lambda)$ ILAGC not only can act online in a real-time mode as a third-party observer to imitate and learn from the control behaviors of the PI controller such that the applicability and feasibility of proposed AGC methodology is enhanced significantly, but also the convergence efficiency and control performance of AGC system can be improved.

v) A new equilibrium-inspired multiobjective group search optimizer based on synergistic learning is proposed to solve the high-dimensional tri-objective EEED optimization problem

The final major contribution of this thesis is to develop a novel multiobjective meta-heuristic algorithm, MGSO, to solve the large-scale highly constrained EEED optimization problem. In the proposed algorithm, multiple searching groups are formed with RL mechanisms to achieve the synergistic learning and cooperative search for Pareto solutions. A better constraint handling strategy, which separates constraints and objectives with different search strategies, and two enhancements, namely space reduction and chaotic sequence dispersion, have been incorporated to improve the algorithm performance. Moreover, the Nash equilibrium point is introduced to extract the best compromise solution from the obtained PF, and an equilibrium-inspired clustering method is also developed to form a representative PF for power operators. Compared to other multiobjective algorithms on various performance

measures, the MGSO is a good alternative in solving large-scaled MOPD problems with highly complex problem characteristics.

7.2 Future Work

This thesis has laid a substantial foundation for the use of the multiobjective control and optimization to design two LFC dispatch problems, AGC and EEED schemes. With the further development of the multi-agent hierarchical RL theory and stochastic search optimization, following research directions may be worthy of further studying in the future:

- (1) Various multi-agent RL algorithms based on Nash equilibria [78], correlated equilibria [189], and Stackelberg equilibria [191] can be employed to design the optimal coordinated AGC strategies for multi-area interconnected power grids in order to enhance the overall control performance of interconnection operations. Most specifically, Nash equilibrium with individual rationality can be introduced for the coordinated LFC optimization in normal operating conditions, while the correlated equilibrium with collective rationality can be used to study the coordinated LFC strategy in emergency conditions.

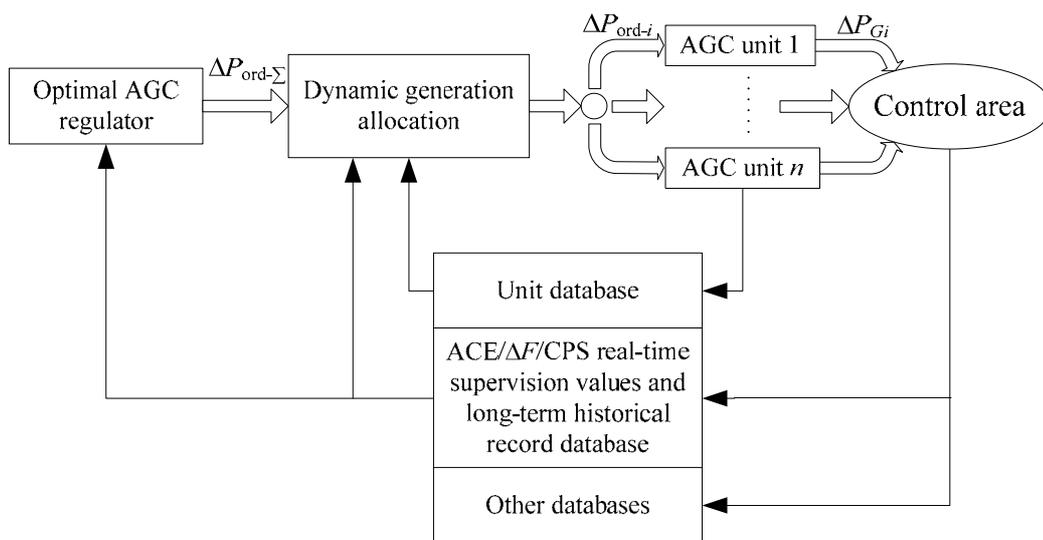


Fig. 7.1 Dynamic optimization process for hierarchical AGC framework

- (2) The existing AGC generation allocation among various types of AGC plants is based on their participation factors which are always determined with the ED or MOPD function. However, this method cannot provide the optimum performance over a wide range of dynamic operational scenarios of power grids. Therefore, further on-going research would focus on the optimization problem of dynamic AGC generation allocation so as to formulate a two-layer hierarchical AGC framework using multi-agent RL algorithms with imitation pre-learning process, as depicted in Fig. 7.1. In addition, the effects of relaxation vector μ on the maneuvering cost and efficiency loss of AGC plants shall also be investigated and analyzed quantitatively.
- (3) The tri-objective EEED optimization can be extended to accommodate more objective functions, e.g. voltage stability, and take into account more system constraints, such as dynamic dispatch, security and contingency constraints. An enhanced MGSO can also be investigated and applied to solve the multi-objective optimal VAR dispatch problem in which voltage stability index, voltage profiles and power loss are to be improved simultaneously. Study on the more powerful Pareto-based multiobjective algorithms can be continued to optimize and coordinate the active/reactive power outputs of dispatchable generators simultaneously for distributed multi-area interconnected power systems with highly constrained and high-dimensional problem complexity. Furthermore, the execution time of the proposed EEED algorithm should be improved with the help of modern advanced computer systems and parallel computation theory to satisfy the practical real-time requirements of tertiary frequency control.

Appendices

A. Data of Two-Area LFC System and Guangdong Power Grid

Note: If not specified, all data in per unit are calculated on the basis of power rating of 5000 MW.

Table A.1 System parameters for two-area LFC system model

T_g (s)	T_t (s)	T_p (s)	R (Hz/pu)	K_p (Hz/pu)	T_{12}
0.08	0.3	20	2.4	120	0.545

Table A.2 Model parameters of AGC units in Guangdong power grid

Types	Groups	T_s (s)	$\Delta P_{Gi,max}$ (MW)	$\Delta P_{Gi,min}$ (MW)	Ramp GRC (MW·min ⁻¹)
Coal-fired	Thermal 1	45	2800	-2800	140
	Thermal 2	43	1912	-1912	71.7
	Thermal 3	40	2680	-2680	134
	Thermal 4	38	1788	-1788	70.5
LNG	Thermal 5	12	1028	-1028	128.5
	Thermal 6	8	688	-688	68.82
Oil-fired	Thermal 7	25	720	-720	54
	Thermal 8	20	480	-480	28.8
Hydro power	Hydro 1	5	600	0	600
	Hydro 2	5	400	0	400

B. Data of IEEE 30-Bus 6-Generator Power System

Note: If not specified, all data in per unit are calculated on the basis of power rating of 100MVA.

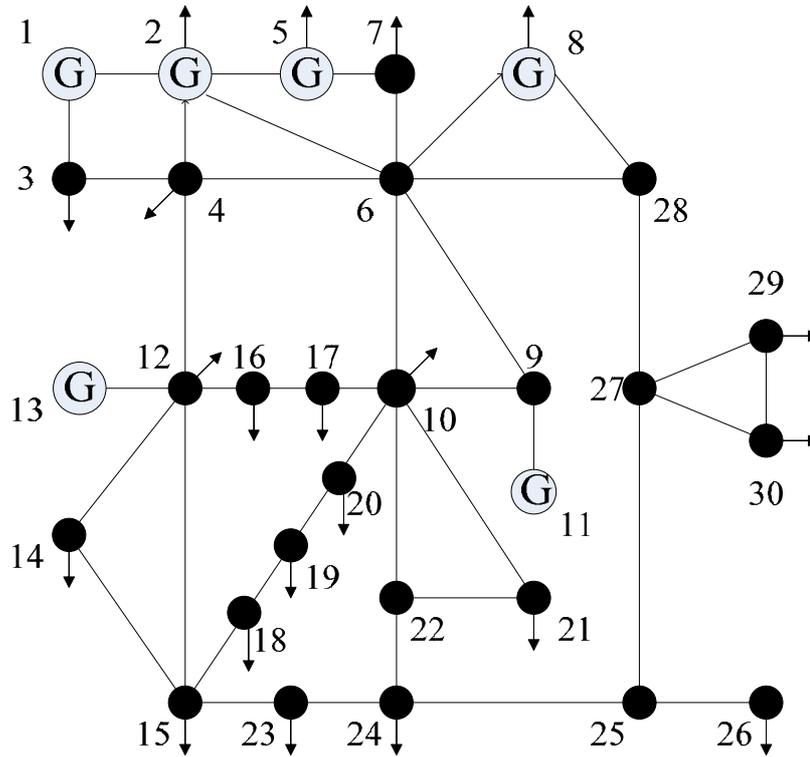


Fig. B.1 Single-line diagram of IEEE 30-bus test system

Table B.1 Generator cost coefficients of IEEE 30-bus system

Generator	a_i	b_i	c_i	d_i	e_i
G1	100	200	10	0	0
G2	120	150	10	0	0
G3	40	180	20	0	0
G4	60	100	10	0	0
G5	40	180	20	0	0
G6	100	150	10	0	0

Table B.2 Generator emission coefficients of IEEE 30-bus system

Generator	α_i	β_i	γ_i	ζ_i	λ_i
G1	6.490	-5.554	4.091	2.00E-04	2.857
G2	5.638	-6.047	2.543	5.00E-04	3.333
G3	4.586	-5.094	4.258	1.00E-06	8.000
G4	3.380	-3.55	5.326	2.00E-03	2.000
G5	4.586	-5.094	4.258	1.00E-06	8.000
G6	5.151	-5.555	6.131	1.00E-05	6.667

Table B.3 Generation data of IEEE 30-bus system

Generator	Bus #	$V_{Gi,rating}$ (pu)	$P_{Gi,max}$ (pu)	$P_{Gi,min}$ (pu)	$Q_{Gi,max}$ (pu)	$Q_{Gi,min}$ (pu)
G1 (Slack)	1	1.060	0.5	0.05	1.5	-0.2
G2	2	1.045	0.6	0.05	0.6	-0.2
G3	5	1.010	1	0.05	0.625	-0.15
G4	8	1.010	1.2	0.05	0.5	-0.15
G5	11	1.082	1	0.05	0.4	-0.1
G6	13	1.071	0.6	0.05	0.45	-0.15

Table B.4 Transmission line data of IEEE 30-bus system

Line #	From bus	To bus	Resistance (pu)	Reactance (pu)	Susceptance (pu)	Transformer ratio	Rating (MVA)
1	1	2	0.0192	0.0575	0.0264	1	1.3
2	1	3	0.0452	0.1852	0.0204	1	1.3
3	2	4	0.057	0.1737	0.0184	1	0.65
4	3	4	0.0132	0.0379	0.0042	1	1.3
5	2	5	0.0472	0.1983	0.0209	1	1.3
6	2	6	0.0581	0.1763	0.0187	1	0.65
7	4	6	0.0119	0.0414	0.0045	1	0.9
8	5	7	0.046	0.116	0.0102	1	0.7
9	6	7	0.0267	0.082	0.0085	1	1.3
10	6	8	0.012	0.042	0.0045	1	0.32
11	9	6	0	0.208	0	1	0.65
12	6	10	0	0.556	0	1	0.32
13	9	11	0	0.208	0	1	0.65
14	9	10	0	0.11	0	1	0.65
15	12	4	0	0.256	0	1	0.65
16	12	13	0	0.14	0	1	0.65
17	12	14	0.1231	0.2559	0	1	0.32

18	12	15	0.0662	0.1304	0	1	0.32
19	12	16	0.0945	0.1987	0	1	0.32
20	14	15	0.221	0.1997	0	1	0.16
21	16	17	0.0824	0.1923	0	1	0.16
22	15	18	0.107	0.2185	0	1	0.16
23	18	19	0.0639	0.1292	0	1	0.16
24	19	20	0.034	0.068	0	1	0.32
25	10	20	0.0936	0.209	0	1	0.32
26	10	17	0.0324	0.0845	0	1	0.32
27	10	21	0.0348	0.0749	0	1	0.32
28	10	22	0.0727	0.1499	0	1	0.32
29	21	22	0.0116	0.0236	0	1	0.32
30	15	23	0.1	0.202	0	1	0.16
31	22	24	0.115	0.179	0	1	0.16
32	23	24	0.132	0.27	0	1	0.16
33	24	25	0.1885	0.3292	0	1	0.16
34	25	26	0.2544	0.38	0	1	0.16
35	25	27	0.1093	0.2087	0	1	0.16
36	28	27	0	0.396	0	1	0.65
37	27	29	0.2198	0.4153	0	1	0.16
38	27	30	0.3202	0.6027	0	1	0.16
39	29	30	0.2399	0.4533	0	1	0.16
40	8	28	0.0636	0.2	0.0214	1	0.32
41	6	28	0.0169	0.0599	0.0065	1	0.32

Table B.5 Bus and demand data of IEEE 30-bus system

Bus #	P_{Di} (pu)	Q_{Di} (pu)	$V_{i,max}$ (pu)	$V_{i,min}$ (pu)
1	0	0	1.05	0.95
2	0.217	0.127	1.1	0.95
3	0.024	0.012	1.05	0.95
4	0.076	0.016	1.05	0.95
5	0.942	0.19	1.1	0.95
6	0	0	1.05	0.95
7	0.228	0.109	1.05	0.95
8	0.3	0.3	1.1	0.95
9	0	0	1.05	0.95
10	0.058	0.02	1.05	0.95
11	0	0	1.1	0.95
12	0.112	0.075	1.05	0.95
13	0	0	1.1	0.95
14	0.062	0.016	1.05	0.95
15	0.082	0.025	1.05	0.95
16	0.035	0.018	1.05	0.95
17	0.09	0.058	1.05	0.95

18	0.032	0.009	1.05	0.95
19	0.095	0.034	1.05	0.95
20	0.022	0.007	1.05	0.95
21	0.175	0.112	1.05	0.95
22	0	0	1.05	0.95
23	0.032	0.016	1.05	0.95
24	0.087	0.067	1.05	0.95
25	0	0	1.05	0.95
26	0.035	0.023	1.05	0.95
27	0	0	1.05	0.95
28	0	0	1.05	0.95
29	0.024	0.009	1.05	0.95
30	0.106	0.019	1.05	0.95

C. Data of IEEE 118-Bus 54-Generator Power System

Note: If not specified, all data in per unit are calculated on the basis of power rating of 100MVA.

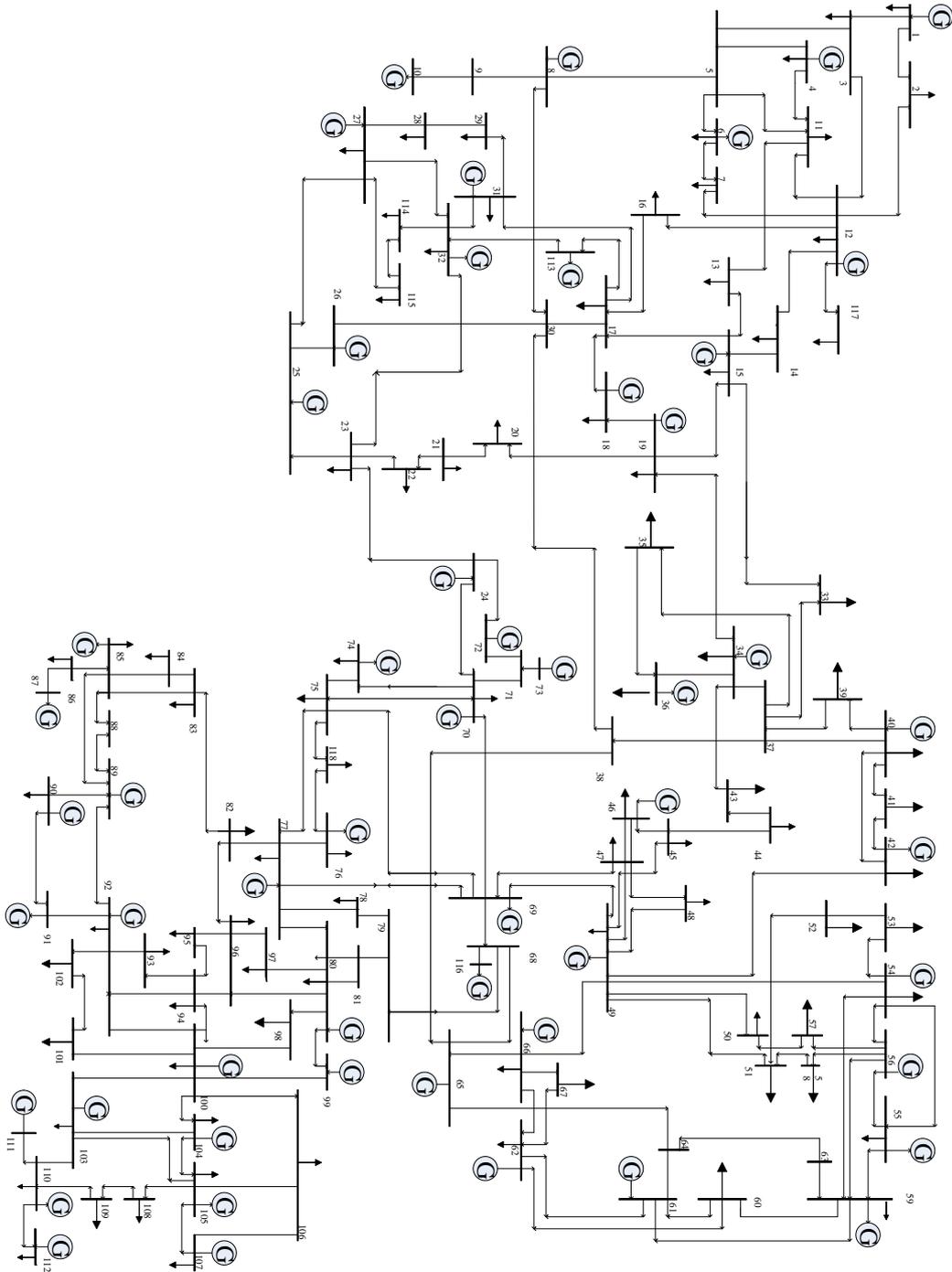


Fig. C.1 Single-line diagram of IEEE 118-bus test system

Table C.1 Generator cost coefficients of IEEE 118-bus system

Generator	a_i	b_i	c_i	d_i	e_i
G1	28.4	860	126	100	8.4
G2	28.4	860	126	100	8.4
G3	28.4	860	126	100	8.4
G4	28.4	860	126	100	8.4
G5	15.62	792	561	300	3.15
G6	48.2	797	78	150	6.3
G7	32.4	774	240	150	6.3
G8	32.4	774	240	150	6.3
G9	32.4	774	240	150	6.3
G10	32.4	774	240	150	6.3
G11	19.4	785	310	200	4.2
G12	2.8	810	550	300	3.5
G13	32.4	774	240	150	6.3
G14	31.3	797	647.83	300	3.5
G15	31.6	795	649.69	300	3.5
G16	31.6	795	649.69	300	3.5
G17	31.6	795	649.69	300	3.5
G18	31.6	795	649.69	300	3.5
G19	31.6	795	649.69	300	3.5
G20	114	535	148.89	120	7.7
G21	114.2	805	222.33	100	8.4
G22	35.7	803	287.71	200	4.2
G23	5.6	810	309	200	4.2
G24	5.6	810	307	200	4.2
G25	202.8	707	309.54	100	8.4
G26	94.2	818	369.03	150	6.3
G27	49.2	699	391.98	200	4.2
G28	57.3	660	455.76	200	4.2
G29	60.5	1290	722.82	200	4.2
G30	51.5	1290	635.2	200	4.2
G31	5212.4	333	1055.1	120	7.7
G32	5212.4	333	1055.1	120	7.7
G33	5212.4	333	1055.1	120	7.7
G34	16	643	222.92	150	6.3
G35	16	643	222.92	150	6.3
G36	16	643	222.92	150	6.3
G37	114	535	148.89	120	7.7
G38	1	895	107.87	200	4.2
G39	56.9	1280	654.69	200	4.2
G40	42.1	1250	913.4	300	3.5
G41	1	862	116.58	200	4.2
G42	1	862	116.58	200	4.2
G43	69	673	94.705	100	8.4
G44	69	673	94.705	100	8.4
G45	75.2	884	1760.4	300	3.5

G46	15.62	792	561	300	3.15
G47	70.8	915	1728.3	300	3.5
G48	70.8	915	1728.3	300	3.5
G49	29.8	663	785.96	300	3.5
G50	29.8	663	785.96	300	3.5
G51	161	588	307.45	80	9.8
G52	27.7	710	801.32	300	3.5
G53	28.4	666	794.53	300	3.5
G54	28.4	666	794.53	300	3.5

Table C.2 Generator emission coefficients of IEEE 118-bus system

Generator	α_i	β_i	γ_i	ζ_i	λ_i
G1	6.49	-5.554	4.091	2.00E-04	2.857
G2	6.49	-5.554	4.091	2.00E-04	2.857
G3	6.49	-5.554	4.091	2.00E-04	2.857
G4	6.49	-5.554	4.091	2.00E-04	2.857
G5	8.315	-1.895	3.45	5.00E-05	1.102
G6	7.016	-2.461	5.812	5.00E-06	3.611
G7	5.638	-6.047	2.543	5.00E-04	3.333
G8	5.638	-6.047	2.543	5.00E-04	3.333
G9	5.638	-6.047	2.543	5.00E-04	3.333
G10	5.638	-6.047	2.543	5.00E-04	3.333
G11	3.178	-4.522	3.678	5.00E-05	1.817
G12	5.212	-1.951	3.755	5.00E-06	1.551
G13	5.151	-5.555	6.131	1.00E-05	6.667
G14	7.164	-4.356	8.151	8.00E-06	3.611
G15	3.38	-3.55	5.326	2.00E-03	2
G16	3.38	-3.55	5.326	2.00E-03	2
G17	3.38	-3.55	5.326	2.00E-03	2
G18	3.38	-3.55	5.326	2.00E-03	2
G19	3.38	-3.55	5.326	2.00E-03	2
G20	8.081	-2.955	3.687	8.00E-06	5.001
G21	7.631	-2.706	3.968	8.00E-06	1.601
G22	7.971	-2.196	3.382	8.00E-05	2.691
G23	5.151	-5.555	6.131	1.00E-05	6.667
G24	5.151	-5.555	6.131	1.00E-05	6.667
G25	2.985	-4.537	3.118	3.00E-05	3.155
G26	5.793	-5.216	3.665	3.00E-06	3.989
G27	5.151	-5.555	6.131	1.00E-05	6.667
G28	5.212	-1.951	3.755	5.00E-06	1.551
G29	3.178	-4.522	3.678	1.00E-06	1.817
G30	3.965	-3.321	8.615	6.00E-05	1.168
G31	4.586	-5.094	4.258	1.00E-06	8

G32	4.586	-5.094	4.258	1.00E-06	8
G33	4.586	-5.094	4.258	1.00E-06	8
G34	4.586	-5.094	4.258	1.00E-06	8
G35	4.586	-5.094	4.258	1.00E-06	8
G36	4.586	-5.094	4.258	1.00E-06	8
G37	6.816	-4.18	8.183	2.00E-04	1.8
G38	5.151	-5.555	6.131	1.00E-05	6.667
G39	8.081	-2.955	3.687	8.00E-06	5.001
G40	4.156	-3.966	8.183	3.00E-05	1.111
G41	6.518	-5.865	6.578	6.00E-05	5.15
G42	6.518	-5.865	6.578	6.00E-05	5.15
G43	6.518	-5.865	6.578	6.00E-05	5.15
G44	6.518	-5.865	6.578	6.00E-05	5.15
G45	3.178	-4.522	3.678	5.00E-05	1.817
G46	7.971	-2.196	3.382	8.00E-05	2.691
G47	4.586	-5.094	4.258	1.00E-06	8
G48	5.151	-5.555	6.131	1.00E-05	6.667
G49	6.49	-5.554	4.091	2.00E-04	2.857
G50	5.638	-6.047	2.543	5.00E-04	3.333
G51	7.971	-2.196	3.382	8.00E-05	2.691
G52	7.164	-4.356	8.151	8.00E-06	3.611
G53	7.164	-4.356	8.151	1.00E-06	3.611
G54	7.164	-4.356	8.151	1.00E-05	3.611

Table C.3 Generation data of IEEE 118-bus system

Generator	Bus #	$V_{Gi,rating}$ (pu)	$P_{Gi,max}$ (pu)	$P_{Gi,min}$ (pu)	$Q_{Gi,max}$ (pu)	$Q_{Gi,min}$ (pu)
G1	1	1.021	1	0	0.15	-0.05
G2	4	1.048	1	0	3	-3
G3	6	1.042	1	0	0.5	-0.13
G4	8	1.032	1	0	3	-3
G5	10	1.043	5.5	0	2	-1.47
G6	12	1.04	1.85	0	1.2	-0.35
G7	15	1.035	1	0	0.3	-0.1
G8	18	1.037	1	0	0.5	-0.16
G9	19	1.033	1	0	0.24	-0.08
G10	24	1.037	1	0	3	-3
G11	25	1.047	3.2	0	1.4	-0.47
G12	26	1.012	4.14	0	10	-10
G13	27	1.031	1	0	3	-3
G14	31	1.028	1.07	0	3	-3
G15	32	1.031	1	0	0.42	-0.14
G16	34	1.045	1	0	0.24	-0.08

G17	36	1.043	1	0	0.24	-0.08
G18	40	1.032	1	0	3	-3
G19	42	1.03	1	0	3	-3
G20	46	1.025	1.19	0	1	-1
G21	49	1.037	3.04	0	2.1	-0.85
G22	54	1.018	1.48	0	3	-3
G23	55	1.016	1	0	0.23	-0.08
G24	56	1.017	1	0	0.15	-0.08
G25	59	1.033	2.55	0	1.8	-0.6
G26	61	1.033	2.6	0	3	-1
G27	62	1.03	1	0	0.2	-0.2
G28	65	1.005	4.91	0	2	-0.67
G29	66	1.05	4.92	0	2	-0.67
G30 (Slack)	69	1.045	8.052	0	3	-3
G31	70	1.025	1	0	0.32	-0.1
G32	72	1.03	1	0	1	-1
G33	73	1.029	1	0	1	-1
G34	74	1.006	1	0	0.09	-0.06
G35	76	0.99	1	0	0.23	-0.08
G36	77	1.031	1	0	0.7	-0.2
G37	80	1.05	5.77	0	2.8	-1.65
G38	85	1.032	1	0	0.23	-0.08
G39	87	1.048	1.04	0	10	-1
G40	89	1.041	7.07	0	3	-2.1
G41	90	1.032	1	0	3	-3
G42	91	1.035	1	0	1	-1
G43	92	1.034	1	0	0.09	-0.03
G44	99	1.044	1	0	1	-1
G45	100	1.048	3.52	0	1.55	-0.5
G46	103	1.041	1.4	0	0.4	-0.15
G47	104	1.03	1	0	0.23	-0.08
G48	105	1.03	1	0	0.23	-0.08
G49	107	1.024	1	0	2	-2
G50	110	1.03	1	0	0.23	-0.08
G51	111	1.039	1.36	0	10	-1
G52	112	1.024	1	0	10	-1
G53	113	1.045	1	0	2	-1
G54	116	1	1	0	10	-10

Table C.4 Transmission line data of IEEE 118-bus system

Line #	From bus	To bus	Resistance (pu)	Reactance (pu)	Susceptance (pu)	Transformer ratio	Rating (MVA)
1	1	2	0.0303	0.0999	0.0127	1	2.2
2	1	3	0.0129	0.0424	0.00541	1	2.2
3	4	5	0.00176	0.00798	0.00105	1	4.4
4	3	5	0.0241	0.108	0.0142	1	2.2
5	5	6	0.0119	0.054	0.00713	1	2.2
6	6	7	0.00459	0.0208	0.00275	1	2.2
7	8	9	0.00244	0.0305	0.581	1	11
8	8	5	0	0.0267	0	0.985	8.8
9	9	10	0.00258	0.0322	0.615	1	11
10	4	11	0.0209	0.0688	0.00874	1	2.2
11	5	11	0.0203	0.0682	0.00869	1	2.2
12	11	12	0.00595	0.0196	0.00251	1	2.2
13	2	12	0.0187	0.0616	0.00786	1	2.2
14	3	12	0.0484	0.16	0.0203	1	2.2
15	7	12	0.00862	0.034	0.00437	1	2.2
16	11	13	0.02225	0.0731	0.00938	1	2.2
17	12	14	0.0215	0.0707	0.00908	1	2.2
18	13	15	0.0744	0.2444	0.03134	1	2.2
19	14	15	0.0595	0.195	0.0251	1	2.2
20	12	16	0.0212	0.0834	0.0107	1	2.2
21	15	17	0.0132	0.0437	0.0222	1	4.4
22	16	17	0.0454	0.1801	0.0233	1	2.2
23	17	18	0.0123	0.0505	0.00649	1	2.2
24	18	19	0.01119	0.0493	0.00571	1	2.2
25	19	20	0.0252	0.117	0.0149	1	2.2
26	15	19	0.012	0.0394	0.00505	1	2.2
27	20	21	0.0183	0.0849	0.0108	1	2.2
28	21	22	0.0209	0.097	0.0123	1	2.2
29	22	23	0.0342	0.159	0.0202	1	2.2
30	23	24	0.0135	0.0492	0.0249	1	2.2
31	23	25	0.0156	0.08	0.0432	1	4.4
32	26	25	0	0.0382	0	0.96	2.2
33	25	27	0.0318	0.163	0.0882	1	4.4
34	27	28	0.01913	0.0855	0.0108	1	2.2
35	28	29	0.0237	0.0943	0.0119	1	2.2
36	30	17	0	0.0388	0	0.96	6.6
37	8	30	0.00431	0.0504	0.257	1	2.2
38	26	30	0.00799	0.086	0.454	1	6.6
39	17	31	0.0474	0.1563	0.01995	1	2.2
40	29	31	0.0108	0.0331	0.00415	1	2.2
41	23	32	0.0317	0.1153	0.05865	1	2.2
42	31	32	0.0298	0.0985	0.01255	1	2.2
43	27	32	0.0229	0.0755	0.00963	1	2.2
44	15	33	0.038	0.1244	0.01597	1	2.2

45	19	34	0.0752	0.247	0.0316	1	2.2
46	35	36	0.00224	0.0102	0.00134	1	2.2
47	35	37	0.011	0.0497	0.00659	1	2.2
48	33	37	0.0415	0.142	0.0183	1	2.2
49	34	36	0.00871	0.0268	0.00284	1	2.2
50	34	37	0.00256	0.0094	0.00492	1	4.4
51	38	37	0	0.0375	0	0.935	6.6
52	37	39	0.0321	0.106	0.0135	1	2.2
53	37	40	0.0593	0.168	0.021	1	2.2
54	30	38	0.00464	0.054	0.211	1	2.2
55	39	40	0.0184	0.0605	0.00776	1	2.2
56	40	41	0.0145	0.0487	0.00611	1	2.2
57	40	42	0.0555	0.183	0.0233	1	2.2
58	41	42	0.041	0.135	0.0172	1	2.2
59	43	44	0.0608	0.2454	0.03034	1	2.2
60	34	43	0.0413	0.1681	0.02113	1	2.2
61	44	45	0.0224	0.0901	0.0112	1	2.2
62	45	46	0.04	0.1356	0.0166	1	2.2
63	46	47	0.038	0.127	0.0158	1	2.2
64	46	48	0.0601	0.189	0.0236	1	2.2
65	47	49	0.0191	0.0625	0.00802	1	2.2
66	42	49	0.0715	0.323	0.043	1	2.2
67	42	49	0.0715	0.323	0.043	1	2.2
68	45	49	0.0684	0.186	0.0222	1	2.2
69	48	49	0.0179	0.0505	0.00629	1	2.2
70	49	50	0.0267	0.0752	0.00937	1	2.2
71	49	51	0.0486	0.137	0.0171	1	2.2
72	51	52	0.0203	0.0588	0.00698	1	2.2
73	52	53	0.0405	0.1635	0.02029	1	2.2
74	53	54	0.0263	0.122	0.0155	1	2.2
75	49	54	0.073	0.289	0.0369	1	2.2
76	49	54	0.0869	0.291	0.0365	1	2.2
77	54	55	0.0169	0.0707	0.0101	1	2.2
78	54	56	0.00275	0.00955	0.00366	1	2.2
79	55	56	0.00488	0.0151	0.00187	1	2.2
80	56	57	0.0343	0.0966	0.0121	1	2.2
81	50	57	0.0474	0.134	0.0166	1	2.2
82	56	58	0.0343	0.0966	0.0121	1	2.2
83	51	58	0.0255	0.0719	0.00894	1	2.2
84	54	59	0.0503	0.2293	0.0299	1	2.2
85	56	59	0.0825	0.251	0.02845	1	2.2
86	56	59	0.0803	0.239	0.0268	1	2.2
87	55	59	0.04739	0.2158	0.02823	1	2.2
88	59	60	0.0317	0.145	0.0188	1	2.2
89	59	61	0.0328	0.15	0.0194	1	2.2
90	60	61	0.00264	0.0135	0.00728	1	4.4
91	60	62	0.0123	0.0561	0.00734	1	2.2
92	61	62	0.00824	0.0376	0.0049	1	2.2

93	63	59	0	0.0386	0	0.96	4.4
94	63	64	0.00172	0.02	0.108	1	4.4
95	64	61	0	0.0268	0	0.985	2.2
96	38	65	0.00901	0.0986	0.523	1	4.4
97	64	65	0.00269	0.0302	0.19	1	4.4
98	49	66	0.018	0.0919	0.0124	1	4.4
99	49	66	0.018	0.0919	0.0124	1	4.4
100	62	66	0.0482	0.218	0.0289	1	2.2
101	62	67	0.0258	0.117	0.0155	1	2.2
102	65	66	0	0.037	0	0.935	2.2
103	66	67	0.0224	0.1015	0.01341	1	2.2
104	65	68	0.00138	0.016	0.319	1	2.2
105	47	69	0.0844	0.2778	0.03546	1	2.2
106	49	69	0.0985	0.324	0.0414	1	2.2
107	68	69	0	0.037	0	0.935	4.4
108	69	70	0.03	0.127	0.061	1	4.4
109	24	70	0.00221	0.4115	0.05099	1	2.2
110	70	71	0.00882	0.0355	0.00439	1	2.2
111	24	72	0.0488	0.196	0.0244	1	2.2
112	71	72	0.0446	0.18	0.02222	1	2.2
113	71	73	0.00866	0.0454	0.00589	1	2.2
114	70	74	0.0401	0.1323	0.01684	1	2.2
115	70	75	0.0428	0.141	0.018	1	2.2
116	69	75	0.0405	0.122	0.062	1	4.4
117	74	75	0.0123	0.0406	0.00517	1	2.2
118	76	77	0.0444	0.148	0.0184	1	2.2
119	69	77	0.0309	0.101	0.0519	1	2.2
120	75	77	0.0601	0.1999	0.02489	1	2.2
121	77	78	0.00376	0.0124	0.00632	1	2.2
122	78	79	0.00546	0.0244	0.00324	1	2.2
123	77	80	0.017	0.0485	0.0236	1	4.4
124	77	80	0.0294	0.105	0.0114	1	2.2
125	79	80	0.0156	0.0704	0.00935	1	2.2
126	68	81	0.00175	0.0202	0.404	1	2.2
127	81	80	0	0.037	0	0.935	2.2
128	77	82	0.0298	0.0853	0.04087	1	2.2
129	82	83	0.0112	0.03665	0.01898	1	2.2
130	83	84	0.0625	0.132	0.0129	1	2.2
131	83	85	0.043	0.148	0.0174	1	2.2
132	84	85	0.0302	0.0641	0.00617	1	2.2
133	85	86	0.035	0.123	0.0138	1	2.2
134	86	87	0.02828	0.2074	0.02225	1	2.2
135	85	88	0.02	0.102	0.0138	1	2.2
136	85	89	0.0239	0.173	0.0235	1	2.2
137	88	89	0.0139	0.0712	0.00967	1	4.4
138	89	90	0.0518	0.188	0.0264	1	6.6
139	89	90	0.0238	0.0997	0.053	1	6.6
140	90	91	0.0254	0.0836	0.0107	1	6.6

141	89	92	0.0099	0.0505	0.0274	1	2.2
142	89	92	0.0393	0.1581	0.0207	1	2.2
143	91	92	0.0387	0.1272	0.01634	1	2.2
144	92	93	0.0258	0.0848	0.0109	1	2.2
145	92	94	0.0481	0.158	0.0203	1	2.2
146	93	94	0.0223	0.0732	0.00938	1	2.2
147	94	95	0.0132	0.0434	0.00555	1	2.2
148	80	96	0.0356	0.182	0.0247	1	2.2
149	82	96	0.0162	0.053	0.0272	1	2.2
150	94	96	0.0269	0.0869	0.0115	1	2.2
151	80	97	0.0183	0.0934	0.0127	1	2.2
152	80	98	0.0238	0.108	0.0143	1	2.2
153	80	99	0.0454	0.206	0.0273	1	2.2
154	92	100	0.0648	0.295	0.0236	1	2.2
155	94	100	0.0178	0.058	0.0302	1	2.2
156	95	96	0.0171	0.0547	0.00737	1	2.2
157	96	97	0.0173	0.0885	0.012	1	2.2
158	98	100	0.0397	0.179	0.0238	1	2.2
159	99	100	0.018	0.0813	0.0108	1	2.2
160	100	101	0.0277	0.1262	0.0164	1	2.2
161	92	102	0.0123	0.0559	0.00732	1	2.2
162	101	102	0.0246	0.112	0.0147	1	2.2
163	100	103	0.016	0.0525	0.0268	1	4.4
164	100	104	0.0451	0.204	0.02705	1	2.2
165	103	104	0.0466	0.1584	0.02035	1	2.2
166	103	105	0.0535	0.1625	0.0204	1	2.2
167	100	106	0.0605	0.229	0.031	1	2.2
168	104	105	0.00994	0.0378	0.00493	1	2.2
169	105	106	0.014	0.0547	0.00717	1	2.2
170	105	107	0.053	0.183	0.0236	1	2.2
171	105	108	0.0261	0.0703	0.00922	1	2.2
172	106	107	0.053	0.183	0.0236	1	2.2
173	108	109	0.0105	0.0288	0.0038	1	2.2
174	103	110	0.03906	0.1813	0.02305	1	2.2
175	109	110	0.0278	0.0762	0.0101	1	2.2
176	110	111	0.022	0.0755	0.01	1	2.2
177	110	112	0.0247	0.064	0.031	1	2.2
178	17	113	0.00913	0.0301	0.00384	1	2.2
179	32	113	0.0615	0.203	0.0259	1	2.2
180	32	114	0.0135	0.0612	0.00814	1	2.2
181	27	115	0.0164	0.0741	0.00986	1	2.2
182	114	115	0.0023	0.0104	0.00138	1	2.2
183	68	116	0.00034	0.00405	0.082	1	4.4
184	12	117	0.0329	0.14	0.0179	1	2.2
185	75	118	0.0145	0.0481	0.00599	1	2.2
186	76	118	0.0164	0.0544	0.00678	1	2.2

Table C.5 Bus and demand data of IEEE 118-bus system

Bus #	P_{Di} (pu)	Q_{Di} (pu)	$V_{i,max}$ (pu)	$V_{i,min}$ (pu)
1	0.51	0.27	1.06	0.94
2	0.2	0.09	1.06	0.94
3	0.39	0.1	1.06	0.94
4	0.39	0.12	1.06	0.94
5	0	0	1.06	0.94
6	0.52	0.22	1.06	0.94
7	0.19	0.02	1.06	0.94
8	0.28	0	1.06	0.94
9	0	0	1.06	0.94
10	0	0	1.06	0.94
11	0.7	0.23	1.06	0.94
12	0.47	0.1	1.06	0.94
13	0.34	0.16	1.06	0.94
14	0.14	0.01	1.06	0.94
15	0.9	0.3	1.06	0.94
16	0.25	0.1	1.06	0.94
17	0.11	0.03	1.06	0.94
18	0.6	0.34	1.06	0.94
19	0.45	0.25	1.06	0.94
20	0.18	0.03	1.06	0.94
21	0.14	0.08	1.06	0.94
22	0.1	0.05	1.06	0.94
23	0.07	0.03	1.06	0.94
24	0.13	0	1.06	0.94
25	0	0	1.06	0.94
26	0	0	1.06	0.94
27	0.71	0.13	1.06	0.94
28	0.17	0.07	1.06	0.94
29	0.24	0.04	1.06	0.94
30	0	0	1.06	0.94
31	0.43	0.27	1.06	0.94
32	0.59	0.23	1.06	0.94
33	0.23	0.09	1.06	0.94
34	0.59	0.26	1.06	0.94
35	0.33	0.09	1.06	0.94
36	0.31	0.17	1.06	0.94
37	0	0	1.06	0.94
38	0	0	1.06	0.94
39	0.27	0.11	1.06	0.94
40	0.66	0.23	1.06	0.94
41	0.37	0.1	1.06	0.94
42	0.96	0.23	1.06	0.94
43	0.18	0.07	1.06	0.94
44	0.16	0.08	1.06	0.94

45	0.53	0.22	1.06	0.94
46	0.28	0.1	1.06	0.94
47	0.34	0	1.06	0.94
48	0.2	0.11	1.06	0.94
49	0.87	0.3	1.06	0.94
50	0.17	0.04	1.06	0.94
51	0.17	0.08	1.06	0.94
52	0.18	0.05	1.06	0.94
53	0.23	0.11	1.06	0.94
54	1.13	0.32	1.06	0.94
55	0.63	0.22	1.06	0.94
56	0.84	0.18	1.06	0.94
57	0.12	0.03	1.06	0.94
58	0.12	0.03	1.06	0.94
59	2.77	1.13	1.06	0.94
60	0.78	0.03	1.06	0.94
61	0	0	1.06	0.94
62	0.77	0.14	1.06	0.94
63	0	0	1.06	0.94
64	0	0	1.06	0.94
65	0	0	1.06	0.94
66	0.39	0.18	1.06	0.94
67	0.28	0.07	1.06	0.94
68	0	0	1.06	0.94
69	0	0	1.06	0.94
70	0.66	0.2	1.06	0.94
71	0	0	1.06	0.94
72	0.12	0	1.06	0.94
73	0.06	0	1.06	0.94
74	0.68	0.27	1.06	0.94
75	0.47	0.11	1.06	0.94
76	0.68	0.36	1.06	0.94
77	0.61	0.28	1.06	0.94
78	0.71	0.26	1.06	0.94
79	0.39	0.32	1.06	0.94
80	1.3	0.26	1.06	0.94
81	0	0	1.06	0.94
82	0.54	0.27	1.06	0.94
83	0.2	0.1	1.06	0.94
84	0.11	0.07	1.06	0.94
85	0.24	0.15	1.06	0.94
86	0.21	0.1	1.06	0.94
87	0	0	1.06	0.94
88	0.48	0.1	1.06	0.94
89	0	0	1.06	0.94
90	1.63	0.42	1.06	0.94
91	0.1	0	1.06	0.94
92	0.65	0.1	1.06	0.94

93	0.12	0.07	1.06	0.94
94	0.3	0.16	1.06	0.94
95	0.42	0.31	1.06	0.94
96	0.38	0.15	1.06	0.94
97	0.15	0.09	1.06	0.94
98	0.34	0.08	1.06	0.94
99	0.42	0	1.06	0.94
100	0.37	0.18	1.06	0.94
101	0.22	0.15	1.06	0.94
102	0.05	0.03	1.06	0.94
103	0.23	0.16	1.06	0.94
104	0.38	0.25	1.06	0.94
105	0.31	0.26	1.06	0.94
106	0.43	0.16	1.06	0.94
107	0.5	0.12	1.06	0.94
108	0.02	0.01	1.06	0.94
109	0.08	0.03	1.06	0.94
110	0.39	0.3	1.06	0.94
111	0	0	1.06	0.94
112	0.68	0.13	1.06	0.94
113	0.06	0	1.06	0.94
114	0.08	0.03	1.06	0.94
115	0.22	0.07	1.06	0.94
116	1.84	0	1.06	0.94
117	0.2	0.08	1.06	0.94
118	0.33	0.15	1.06	0.94

Reference

- [1] P. Kundur, *Power System Stability and Control*. New York: McGraw-Hill, 1994.
- [2] D. M. Tagare, *Electricity Power Generation: The Changing Dimensions*. New Jersey: IEEE Computer Society Press, 2010.
- [3] N. Jaleeli, L. S. VanSlyck, D. N. Ewart, L. H. Fink, and A. G. Hoffmann, "Understanding automatic generation control," *IEEE Trans. Power Syst.*, vol. 7, no. 3, pp. 1106-1122, Aug. 1992.
- [4] B. H. Chowdhury and S. Rahman, "A review of recent advances in economic dispatch," *IEEE Trans. Power Syst.*, vol. 5, no. 4, pp. 1248-1259, Nov. 1990.
- [5] A. J. Wood and B. F. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. New York: Wiley, 1996.
- [6] P. Kumar, "AGC strategies: A comprehensive review," *Int. J. Power & Energy Syst.*, vol. 16, no. 2, pp. 107-112, May 1996.
- [7] N. Cohn, "Some aspects of tie-line bias control on interconnected power systems," *Amer. Inst. Electr. Eng. Trans.*, vol. 75, no. 3, pp. 1415-1436, Jan. 1956.
- [8] M. Yao, R. R. Shoults, and R. Kelm, "AGC logic based on NERC's new Control Performance Standard and Disturbance Control Standard," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 852-857, May 2000.
- [9] N. Jaleeli and L. S. VanSlyck, "Control performance standards and procedures for interconnected operation," *EPRI Report TR-107813*, vol. 10, Jun. 1997.
- [10] Z. H. Gao, X. L. Teng, and L. Q. Tu, "Hierarchical AGC mode and CPS

- control strategy for interconnected power systems," *Autom. Elect. Power Syst.*, vol. 28, no. 1, pp. 78-81, Jan. 2004.
- [11] Z. H. Gao, X. L. Teng, and X. B. Zhang, "Automatic generation control strategy under control performance standard for interconnected power grids," *Autom. Elect. Power Syst.*, vol. 29, no. 19, pp. 40-44, Oct. 2005.
- [12] A. A. El-Keib, H. Ma, and J. L. Hart, "Economic dispatch in view of the Clean Air Act of 1990," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 972-978, May 1994.
- [13] J. H. Talaq, F. El-Hawary, and M. E. El-Hawary, "A summary of environmental/economic dispatch algorithms," *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp. 1508-1516, Aug. 1994.
- [14] D. Srinivasan and A. G. B. Tettamanzi, "An evolutionary algorithm for evaluation of emission compliance options in view of the Clean Air Act Amendments," *IEEE Trans. Power Syst.*, vol. 12, no. 1, pp. 336-341, Feb. 1997.
- [15] R. Petrovic and B. Kralj, "Economic and environmental power dispatch," *Eur. J. Oper. Res.*, vol. 64, no. 1, pp. 2-11, Jan. 1993.
- [16] S. F. J. Brodsky and R. W. Hahn, "Assessing the influence of power pools on emission constrained economic dispatch," *IEEE Trans. Power Syst.*, vol. 1, no. 1, pp. 57-62, Feb. 1986.
- [17] J. Zahavi and L. Eisenberg, "An application of the economic-environmental power dispatch," *IEEE Trans. Syst. Man Cybern.*, vol. 7, no. 7, pp. 523-530, Jul. 1977.
- [18] J. S. Heslin and B. F. Hobbs, "A multiobjective production costing model for analyzing emissions dispatching and fuel switching," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 836-842, Aug. 1989.
- [19] Y. L. Hu and W. G. Wee, "A hierarchical system for economic dispatch with environmental constraints," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 1076-1082, May 1994.
- [20] G. P. Granelli, M. Montagna, G. L. Pasini, and P. Marannino, "Emission

- constrained dynamic dispatch," *Electr. Power Syst. Res.*, vol. 24, no. 1, pp. 55-64, Jul. 1992.
- [21] R. Ramanathan, "Emission constrained economic dispatch," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1994-2000, Nov. 1994.
- [22] A. A. El-Keib, H. Ma, and J. L. Hart, "Environmentally constrained economic dispatch using the LaGrangian relaxation method," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1723-1729, Nov. 1994.
- [23] T. D. King, M. E. El-Hawary, and F. El-Hawary, "Optimal environmental dispatching of electric power systems via an improved Hopfield neural network model," *IEEE Trans. Power Syst.*, vol. 10, no. 3, pp. 1559-1565, Aug. 1995.
- [24] S. D. Chen and J. F. Chen, "A direct Newton–Raphson economic emission dispatch," *Int. J. Electr. Power Energy Syst.*, vol. 25, no. 5, pp. 411-417, Jun. 2003.
- [25] C. Palanichamy and N. S. Babu, "Analytical solution for combined economic and emissions dispatch," *Electr. Power Syst. Res.*, vol. 78, no. 7, pp. 1129-1137, Jul. 2008.
- [26] J. X. Xu, C. S. Chang, and X. W. Wang, "Constrained multiobjective global optimisation of longitudinal interconnected power system by genetic algorithm," *IEE Proc. Gener. Transm. Distrib.*, vol. 143, no. 5, pp. 435-446, Sep. 1996.
- [27] J. S. Dhillon and D. P. Kothari, "Economic-emission load dispatch using binary successive approximation-based evolutionary search," *IET Gener. Transm. Distrib.*, vol. 3, no. 1, pp. 1-16, Jan. 2009.
- [28] Y. H. Song, G. S. Wang, P. Y. Wang, and A. T. Johns, "Environmental/economic dispatch using fuzzy logic controlled genetic algorithms," *IEE Proc. Gener. Transm. Distrib.*, vol. 144, no. 4, pp. 377-382, Jul. 1997.
- [29] K. P. Wong, B. Fan, C. S. Chang, and A. C. Liew, "Multi-objective generation dispatch using bi-criterion global optimisation," *IEEE Trans.*

- Power Syst.*, vol. 10, no. 4, pp. 1813-1819, Nov. 1995.
- [30] J. F. Chen and S. D. Chen, "Multiobjective power dispatch with line flow constraints using the fast Newton-Raphson method," *IEEE Trans. Energy Convers.*, vol. 12, no. 1, pp. 86-93, Mar. 1997.
- [31] C. A. Coello Coello, D. A. V. Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*. Norwell, MA: Kluwer, 2002.
- [32] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evol. Comput.*, vol. 3, no. 1, pp. 1-16, Mar. 1995.
- [33] D. A. V. Veldhuizen and G. B. Lamont, "Multiobjective evolutionary algorithms: Analyzing the state-of-the-art," *Evol. Comput.*, vol. 8, no. 2, pp. 125-147, Jun. 2000.
- [34] D. F. Jones, S. K. Mirrazavi, and M. Tamiz, "Multi-objective meta-heuristics: An overview of the current state-of-the-art," *Eur. J. Oper. Res.*, vol. 137, no. 1, pp. 1-9, Feb. 2002.
- [35] E. Santacana, G. Rackliffe, Le Tang, and F. Xiaoming, "Getting smart," *IEEE Power Energy Mag.*, vol. 8, no. 2, pp. 41-48, Mar./Apr. 2010.
- [36] N. Jaleeli and L. S. VanSlyck, "Tie-line bias prioritized energy control," *IEEE Trans. Power Syst.*, vol. 10, no. 1, pp. 51-59, Feb. 1995.
- [37] N. Jaleeli and L. S. Vanslyck, "NERC's new control performance standards," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1092-1099, Aug. 1999.
- [38] Ibraheem, P. Kumar, and D. P. Kothari, "Recent philosophies of automatic generation control strategies in power systems," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 346-357, Feb. 2005.
- [39] H. Shayeghi, H. A. Shayanfar, and A. Jalili, "Load frequency control strategies: A state-of-the-art survey for the researcher," *Energy Conv. Manag.*, vol. 50, no. 2, pp. 344-353, Feb. 2009.
- [40] T. Yu, B. Zhou, K. W. Chan, and E. Lu, "Stochastic optimal CPS relaxed

- control methodology for interconnected power systems using Q-learning method," *J. Energy Eng.*, vol. 137, no. 3, pp. 116-129, Sep. 2011.
- [41] The Operation Mode of China Southern Power Grid in 2009 (in Chinese), China Southern Power Grid Co. Ltd., 2009.
- [42] O. I. Elgerd, *Electric Energy Systems Theory: An Introduction*, 2nd ed. New York: McGraw-Hill, 1982.
- [43] T. P. Imthias Ahamed, P. S. Nagendra Rao, and P. S. Sastry, "A reinforcement learning approach to automatic generation control," *Electr. Power Syst. Res.*, vol. 63, no. 1, pp. 9-26, Aug. 2002.
- [44] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge: MIT Press, 1998.
- [45] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237-285, May 1996.
- [46] H. Mine and S. Osaki, *Markovian Decision Processes*. New York: Elsevier, 1970.
- [47] R. Howard, *Dynamic Programming and Markov Decision Processes*. Cambridge, MA: MIT Press, 1960.
- [48] H. Bevrani and T. Hiyama, "On load-frequency regulation with time delays: Design and real-time implementation," *IEEE Trans. Energy Convers.*, vol. 24, no. 1, pp. 292-300, Mar. 2009.
- [49] H. Xia, M. Koyama, G. Leyland, and S. Kraines, "A modularized framework for solving an economic–environmental power generation mix problem," *Int. J. Energy Res.*, vol. 28, no. 9, pp. 769-784, Jul. 2004.
- [50] J. Cai, X. Ma, Q. Li, L. Li, and H. Peng, "A multi-objective chaotic ant swarm optimization for environmental/economic dispatch," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 5, pp. 337-344, Jun. 2010.
- [51] J. Cai, X. Ma, Q. Li, L. Li, and H. Peng, "A multi-objective chaotic particle swarm optimization for environmental/economic dispatch," *Energy Conv. Manag.*, vol. 50, no. 5, pp. 1318-1325, May 2009.
- [52] M. A. Abido, "A niched Pareto genetic algorithm for multiobjective

- environmental/economic dispatch," *Int. J. Electr. Power Energy Syst.*, vol. 25, no. 2, pp. 97-105, Feb. 2003.
- [53] C. M. Huang and Y. C. Huang, "A novel approach to real-time economic emission power dispatch," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 288-294, Feb. 2003.
- [54] M. A. Abido, "A novel multiobjective evolutionary algorithm for environmental/economic power dispatch," *Electr. Power Syst. Res.*, vol. 65, no. 1, pp. 71-81, Apr. 2003.
- [55] M. Mesut, "Economic dispatch with environmental considerations: Tradeoff curves and emission reduction rates," *Electr. Power Syst. Res.*, vol. 71, no. 2, pp. 153-158, Oct. 2004.
- [56] L. H. Wu, Y. N. Wang, X. F. Yuan, and S. W. Zhou, "Environmental/economic power dispatch problem using multi-objective differential evolution algorithm," *Electr. Power Syst. Res.*, vol. 80, no. 9, pp. 1171-1181, Sep. 2010.
- [57] L. Wang and C. Singh, "Environmental/economic power dispatch using a fuzzified multi-objective particle swarm optimization algorithm," *Electr. Power Syst. Res.*, vol. 77, no. 12, pp. 1654-1664, Oct. 2007.
- [58] D. W. Gong, Y. Zhang, and C. L. Qi, "Environmental/economic power dispatch using a hybrid multi-objective optimization algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 6, pp. 607-614, Jul. 2010.
- [59] M. A. Abido, "Environmental/economic power dispatch using multiobjective evolutionary algorithms," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1529-1537, Nov. 2003.
- [60] B. K. Panigrahi, V. R. Pandi, R. Sharma, S. Das, and S. Das, "Multiobjective bacteria foraging algorithm for electrical load dispatch problem," *Energy Conv. Manag.*, vol. 52, no. 2, pp. 1334-1342, Feb. 2011.
- [61] S. Agrawal, B. K. Panigrahi, and M. K. Tiwari, "Multiobjective particle swarm algorithm with fuzzy clustering for electrical power dispatch," *IEEE Trans. Evol. Comput.*, vol. 12, no. 5, pp. 529-541, Oct. 2008.

- [62] M. A. Abido, "Multiobjective particle swarm optimization for environmental/economic dispatch problem," *Electr. Power Syst. Res.*, vol. 79, no. 7, pp. 1105-1113, Jul. 2009.
- [63] L. Wang and C. Singh, "Reserve-constrained multiarea environmental/economic dispatch based on particle swarm optimization with local search," *Eng. Appl. Artif. Intell.*, vol. 22, no. 2, pp. 298-307, Mar. 2009.
- [64] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach," *IEEE Trans. Evol. Comput.*, vol. 3, no. 4, pp. 257-271, Nov. 1999.
- [65] C. A. C. Coello, G. T. Pulido, and M. S. Lechuga, "Handling multiple objectives with particle swarm optimization," *IEEE Trans. Evol. Comput.*, vol. 8, no. 3, pp. 256-279, Jun. 2004.
- [66] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67-82, Apr. 1997.
- [67] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Chichester, U.K.: Wiley, 2001.
- [68] M. Farina and P. Amato, "A fuzzy definition of "optimality" for many-criteria optimization problems," *IEEE Trans. Syst. Man Cybern. Part A-Syst. Hum.*, vol. 34, no. 3, pp. 315-326, May 2004.
- [69] C. Carlsson and R. Fullér, "Fuzzy multiple criteria decision making: Recent developments," *Fuzzy Sets Syst.*, vol. 78, no. 2, pp. 139-153, Mar. 1996.
- [70] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, no. 3-4, pp. 279-292, May 1992.
- [71] Q. Hu and W. Yue, *Markov Decision Processes with Their Applications*. New York: Springer, 2008.
- [72] S. D. Whitehead and L. Lin, "Reinforcement learning of non-Markov decision processes," *Artif. Intell.*, vol. 73, no. 1-2, pp. 271-306, Feb. 1995.

- [73] J. Peng and R. J. Williams, "Incremental multi-step Q-learning," *Mach. Learn.*, vol. 22, no. 1-3, pp. 283-290, Jan. 1996.
- [74] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: Reinforcement learning framework," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 427-435, Feb. 2004.
- [75] G. Hu and C. Wu, "Incremental multi-step R-learning," *J. Beijing Inst. Technol.*, vol. 8, no. 3, pp. 244-250, Sep. 1999.
- [76] S. Mahadevan, "Average reward reinforcement learning: Foundations, algorithms, and empirical results," *Mach. Learn.*, vol. 22, no. 1-3, pp. 159-195, Mar. 1996.
- [77] S. He, Q. H. Wu, and J. R. Saunders, "Group Search Optimizer: An optimization algorithm inspired by animal searching behavior," *IEEE Trans. Evol. Comput.*, vol. 13, no. 5, pp. 973-990, Oct. 2009.
- [78] J. Nash, "Non-cooperative games," *Ann. Math.*, vol. 54, no. 2, pp. 286-295, Sep. 1951.
- [79] J. L. Rodriguez-Amenedo, S. Arnalte, and J. C. Burgos, "Automatic generation control of a wind farm with variable speed wind turbines," *IEEE Trans. Energy Convers.*, vol. 17, no. 2, pp. 279-284, Jun. 2002.
- [80] D. W. Maier and S. L. Larsen, "Standardized SCADA-AGC system effect on system cost, delivery and quality," *IEEE Trans. Power App. Syst.*, vol. PAS-101, no. 1, pp. 185-190, Jan. 1982.
- [81] T. Sasaki and K. Enomoto, "Statistical and dynamic analysis of generation control performance standards," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 476-481, May 2002.
- [82] G. Gross and W. L. Jeong, "Analysis of load frequency control performance assessment criteria," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 520-525, Aug. 2001.
- [83] T. Sasaki and K. Enomoto, "Dynamic analysis of generation control performance standards," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 806-811, Aug. 2002.

- [84] I. Egido, F. Fernandez-Bernal, and L. Rouco, "Evaluation of Automatic Generation Control (AGC) regulators by performance indices using data from real operation," *IET Gener. Transm. Distrib.*, vol. 1, no. 2, pp. 294-302, Mar. 2007.
- [85] H. Naeb-boon, O. Chee-Mun, and R. A. Kramer, "Feasibility of decomposing $A \sim C \sim E \sim 1$ to identify the impact of selected loads on CPS1 and CPS2," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 752-756, Aug. 2002.
- [86] L. R. Chang-Chien, O. Chee-Mun, and R. A. Kramer, "Field tests and refinements of an ACE model," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 898-903, May 2003.
- [87] H. Naeb-boon, O. Chee-Mun, and R. A. Kramer, "Implementation of an ACE1 decomposition method," *IEEE Trans. Power Syst.*, vol. 17, no. 3, pp. 757-761, Aug. 2002.
- [88] R. K. Green, "Transformed automatic generation control," *IEEE Trans. Power Syst.*, vol. 11, no. 4, pp. 1799-1804, Nov. 1996.
- [89] R. J. Mills and W. F. B'Rells, "Automatic generation control-Part I: Process modeling," *IEEE Trans. Power App. Syst.*, vol. PAS-92, no. 2, pp. 710-715, Mar. 1973.
- [90] O. I. Elgerd and C. E. Fosha, "Optimum megawatt-frequency control of multiarea electric energy systems," *IEEE Trans. Power App. Syst.*, vol. PAS-89, no. 4, pp. 556-563, Apr. 1970.
- [91] F. F. Wu and V. S. Dea, "Describing-function analysis of automatic generation control system with governor deadband," *Electr. Power Syst. Res.*, vol. 1, no. 2, pp. 113-116, Apr. 1978.
- [92] A. Khodabakhshian and R. Hooshmand, "A new PID controller design for automatic generation control of hydro power systems," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 5, pp. 375-382, Jun. 2010.
- [93] K. Yamashita and T. Taniguchi, "Optimal observer design for load-frequency control," *Int. J. Electr. Power Energy Syst.*, vol. 8, no. 2, pp. 93-

- 100, Apr. 1986.
- [94] C. E. Fosha and O. I. Elgerd, "The megawatt-frequency control problem: A new approach via optimal control theory," *IEEE Trans. Power App. Syst.*, vol. PAS-89, no. 4, pp. 563-577, Apr. 1970.
- [95] M. S. Calovic, "Automatic generation control: Decentralized area-wise optimal solution," *Electr. Power Syst. Res.*, vol. 7, no. 2, pp. 115-139, Apr. 1984.
- [96] S. Velusami and I. A. Chidambaram, "Decentralized biased dual mode controllers for load frequency control of interconnected power systems considering GDB and GRC non-linearities," *Energy Conv. Manag.*, vol. 48, no. 5, pp. 1691-1702, May 2007.
- [97] K. R. Sudha and R. Vijaya Santhi, "Robust decentralized load frequency control of interconnected power system with generation rate constraint using Type-2 fuzzy approach," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 3, pp. 699-707, Mar. 2011.
- [98] L. Hari, M. L. Kothari, and J. Nanda, "Optimum selection of speed regulation parameters for automatic generation control in discrete mode considering generation rate constraints," *IEE Proc. Gener. Transm. Distrib.*, vol. 138, no. 5, pp. 401-406, Sep. 1991.
- [99] A. Rubaai and V. Udo, "Self-tuning load frequency control: Multilevel adaptive approach," *IEE Proc. Gener. Transm. Distrib.*, vol. 141, no. 4, pp. 285-290, Jul. 1994.
- [100] F. P. Demello, R. J. Mills, and W. F. B'Rells, "Automatic generation control-Part II: Digital control techniques," *IEEE Trans. Power App. Syst.*, vol. PAS-92, no. 2, pp. 716-724, Mar. 1973.
- [101] C. W. Taylor and R. L. Cresap, "Real-time power system simulation for automatic generation control," *IEEE Trans. Power App. Syst.*, vol. 95, no. 1, pp. 375-384, Jan. 1976.
- [102] M. Zribi, M. Al-Rashed, and M. Alrifai, "Adaptive decentralized load frequency control of multi-area power systems," *Int. J. Electr. Power*

- Energy Syst.*, vol. 27, no. 8, pp. 575-583, Oct. 2005.
- [103] C. T. Pan and C. M. Liaw, "An adaptive controller for power system load-frequency control," *IEEE Trans. Power Syst.*, vol. 4, no. 1, pp. 122-128, Feb. 1989.
- [104] R. R. Shoults and J. A. Jativa Ibarra, "Multi-area adaptive LFC developed for a comprehensive AGC simulator," *IEEE Trans. Power Syst.*, vol. 8, no. 2, pp. 541-547, May 1993.
- [105] K. Yamashita and H. Miyagi, "Load frequency self-tuning regulator for interconnected power systems with unknown deterministic load disturbances," *Int. J. Control*, vol. 49, no. 5, pp. 1555-1568, May 1989.
- [106] K. A. Lee, H. Yee, and C. Y. Teo, "Self-tuning algorithm for automatic generation control in an interconnected power system," *Electr. Power Syst. Res.*, vol. 20, no. 2, pp. 157-165, Feb. 1991.
- [107] T. Wen, "Tuning of PID load frequency controller for power systems," *Energy Conv. Manag.*, vol. 50, no. 6, pp. 1465-1472, Jun. 2009.
- [108] H. Bevrani and T. Hiyama, "Robust decentralised PI based LFC design for time delay power systems," *Energy Conv. Manag.*, vol. 49, no. 2, pp. 193-204, Feb. 2008.
- [109] T. Wen and X. Zhan, "Robust analysis and design of load frequency controller for power systems," *Electr. Power Syst. Res.*, vol. 79, no. 5, pp. 846-853, May 2009.
- [110] M. L. Kothari, J. Nanda, D. P. Kothari, and D. Das, "Discrete-mode automatic generation control of a two-area reheat thermal system with new area control error," *IEEE Trans. Power Syst.*, vol. 4, no. 2, pp. 730-738, May 1989.
- [111] A. Kumar, O. P. Malik, and G. S. Hope, "Variable-structure-system control applied to AGC of an interconnected power system," *IEE Proc. Gener. Transm. Distrib.*, vol. 132, no. 1, pp. 23-29, Jan. 1985.
- [112] Z. M. Al-Hamouz and H. N. Al-Duwaish, "A new load frequency variable structure controller using genetic algorithms," *Electr. Power Syst. Res.*,

- vol. 55, no. 1, pp. 1-6, Jul. 2000.
- [113] D. Das, M. L. Kothari, D. P. Kothari, and P. J. Nanda, "Variable structure control strategy to automatic generation control of interconnected reheat thermal system," *IEE Proc. Contr. Theory Appl.*, vol. 138, no. 6, pp. 579-585, Nov. 1991.
- [114] K. Vrdoljak, N. Perić, and I. Petrović, "Sliding mode based load-frequency control in power systems," *Electr. Power Syst. Res.*, vol. 80, no. 5, pp. 514-527, May 2010.
- [115] D. K. Chaturvedi, P. S. Satsangi, and P. K. Kalra, "Load frequency control: A generalised neural network approach," *Int. J. Electr. Power Energy Syst.*, vol. 21, no. 6, pp. 405-415, Aug. 1999.
- [116] H. L. Zeynelgil, A. Demiroren, and N. S. Sengor, "The application of ANN technique to automatic generation control for multi-area power system," *Int. J. Electr. Power Energy Syst.*, vol. 24, no. 5, pp. 345-354, Jun. 2002.
- [117] H. Shayeghi and H. A. Shayanfar, "Application of ANN technique based on μ -synthesis to load frequency control of interconnected power system," *Int. J. Electr. Power Energy Syst.*, vol. 28, no. 7, pp. 503-511, Sep. 2006.
- [118] Ç. Ertuğrul, "Application of fuzzy logic for load frequency control of hydroelectrical power plants," *Energy Conv. Manag.*, vol. 48, no. 4, pp. 1281-1288, Apr. 2007.
- [119] J. Talaq and F. Al-Basri, "Adaptive fuzzy gain scheduling for load frequency control," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 145-150, Feb. 1999.
- [120] İ. Kocaarslan and E. Çam, "Fuzzy logic controller in interconnected electrical power systems for load-frequency control," *Int. J. Electr. Power Energy Syst.*, vol. 27, no. 8, pp. 542-549, Oct. 2005.
- [121] R. Roy, P. Bhatt, and S. P. Ghoshal, "Evolutionary computation based three-area automatic generation control," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5913-5924, Aug. 2010.

- [122] A. Demiroren and H. L. Zeynelgil, "GA application to optimization of AGC in three-area power system after deregulation," *Int. J. Electr. Power Energy Syst.*, vol. 29, no. 3, pp. 230-240, Mar. 2007.
- [123] H. Golpîra, H. Bevrani, and H. Golpîra, "Application of GA optimization for automatic generation control design in an interconnected power system," *Energy Conv. Manag.*, vol. 52, no. 5, pp. 2247-2255, May 2011.
- [124] H. Gozde and M. C. Taplamacioglu, "Automatic generation control application with craziness based particle swarm optimization in a thermal power system," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 1, pp. 8-16, Jan. 2011.
- [125] P. Bhatt, R. Roy, and S. P. Ghoshal, "GA/particle swarm intelligence based optimization of two specific varieties of controller devices applied to two-area multi-units automatic generation control," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 4, pp. 299-310, May 2010.
- [126] H. Shayeghi, A. Jalili, and H. A. Shayanfar, "Multi-stage fuzzy load frequency control using PSO," *Energy Conv. Manag.*, vol. 49, no. 10, pp. 2570-2580, Oct. 2008.
- [127] E. S. Ali and S. M. Abd-Elazim, "Bacteria foraging optimization algorithm based load frequency controller for interconnected power system," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 3, pp. 633-638, Mar. 2011.
- [128] J. Nanda, S. Mishra, and L. C. Saikia, "Maiden application of bacterial foraging-based optimization technique in multiarea automatic generation control," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 602-609, May 2009.
- [129] L. C. Saikia, S. Mishra, N. Sinha, and J. Nanda, "Automatic generation control of a multi area hydrothermal system using reinforced learning neural network controller," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 4, pp. 1101-1108, May 2011.
- [130] F. Daneshfar and H. Bevrani, "Load-frequency control: A GA-based multi-agent reinforcement learning," *IET Gener. Transm. Distrib.*, vol. 4,

- no. 1, pp. 13-26, Jan. 2010.
- [131] S. P. Ghoshal, "Optimizations of PID gains by particle swarm optimizations in fuzzy based automatic generation control," *Electr. Power Syst. Res.*, vol. 72, no. 3, pp. 203-212, Dec. 2004.
- [132] Y. L. Karnavas and D. P. Papadopoulos, "AGC for autonomous power system using combined intelligent techniques," *Electr. Power Syst. Res.*, vol. 62, no. 3, pp. 225-239, Jul. 2002.
- [133] S. P. Ghoshal, "Application of GA/GA-SA based fuzzy automatic generation control of a multi-area thermal generating system," *Electr. Power Syst. Res.*, vol. 70, no. 2, pp. 115-127, Jul. 2004.
- [134] D. Rerkpreedapong, A. Hasanovic, and A. Feliachi, "Robust load frequency control using genetic algorithms and linear matrix inequalities," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 855-861, May 2003.
- [135] X. S. Zhao, F. S. Wen, D. Q. Gan, M. X. Huang, C. W. Yu, and C. Y. Chung, "Determination of AGC capacity requirement and dispatch considering performance penalties," *Electr. Power Syst. Res.*, vol. 70, no. 2, pp. 93-98, Jul. 2004.
- [136] P. Kambale, H. Mukai, J. Spare, and J. Zaborszky, "A reevaluation of the normal operating state control (AGC) of the power system using computer control and system theory-Part III: Tracking the dispatch targets with unit control," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 6, pp. 1903-1912, Jun. 1983.
- [137] R. Bacher and H. P. Van Meeteren, "Real-time optimal power flow in automatic generation control," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1518-1529, Nov. 1988.
- [138] H. Shayeghi, A. Jalili, and H. A. Shayanfar, "A robust mixed H₂/H_∞ based LFC of a deregulated power system including SMES," *Energy Conv. Manag.*, vol. 49, no. 10, pp. 2656-2668, Oct. 2008.
- [139] B. H. Bakken and O. S. Grande, "Automatic generation control in a deregulated power system," *IEEE Trans. Power Syst.*, vol. 13, no. 4, pp.

1401-1406, Nov. 1998.

- [140] P. Bhatt, R. Roy, and S. P. Ghoshal, "Optimized multi area AGC simulation in restructured power systems," *Int. J. Electr. Power Energy Syst.*, vol. 32, no. 4, pp. 311-322, May 2010.
- [141] H. Shayeghi, H. A. Shayanfar, and O. P. Malik, "Robust decentralized neural networks based LFC in a deregulated power system," *Electr. Power Syst. Res.*, vol. 77, no. 3-4, pp. 241-251, Mar. 2007.
- [142] V. Donde, M. A. Pai, and I. A. Hiskens, "Simulation and optimization in an AGC system after deregulation," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 481-489, Aug. 2001.
- [143] P. Bhatt, R. Roy, and S. P. Ghoshal, "Comparative performance evaluation of SMES–SMES, TCPS–SMES and SSSC–SMES controllers in automatic generation control for a two-area hydro–hydro system," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 10, pp. 1585-1597, Dec. 2011.
- [144] S. K. Aditya and D. Das, "Battery energy storage for load frequency control of an interconnected power system," *Electr. Power Syst. Res.*, vol. 58, no. 3, pp. 179-185, Jul. 2001.
- [145] T. Sasaki, T. Kadoya, and K. Enomoto, "Study on load frequency control using redox flow batteries," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 660-667, Feb. 2004.
- [146] H. Asano, K. Yajima, and Y. Kaya, "Influence of photovoltaic power generation on required capacity for load frequency control," *IEEE Trans. Energy Convers.*, vol. 11, no. 1, pp. 188-193, Mar. 1996.
- [147] S. C. Tripathy, R. Balasubramanian, and P. S. Chandramohan Nair, "Adaptive automatic generation control with superconducting magnetic energy storage in power systems," *IEEE Trans. Energy Convers.*, vol. 7, no. 3, pp. 434-441, Sep. 1992.
- [148] R. J. Abraham, D. Das, and A. Patra, "Automatic generation control of an interconnected hydrothermal power system considering superconducting magnetic energy storage," *Int. J. Electr. Power Energy Syst.*, vol. 29, no. 8,

- pp. 571-579, Oct. 2007.
- [149] A. Demiroren and E. Yesil, "Automatic generation control with fuzzy logic controllers in the power system including SMES units," *Int. J. Electr. Power Energy Syst.*, vol. 26, no. 4, pp. 291-305, May 2004.
- [150] S. Pothiya and I. Ngamroo, "Optimal fuzzy logic-based PID controller for load-frequency control including superconducting magnetic energy storage units," *Energy Conv. Manag.*, vol. 49, no. 10, pp. 2833-2838, Oct. 2008.
- [151] K. Sedghisigarchi and A. Feliachi, "Impact of fuel cells on load-frequency control in power distribution systems," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp. 250-256, Mar. 2006.
- [152] D. H. Curtice and T. W. Reddoch, "An assessment of load frequency control impacts caused by small wind turbines," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 1, pp. 162-170, Jan. 1983.
- [153] H. Banakar, C. Luo, and T. O. Boon, "Impacts of wind power minute-to-minute variations on power system operation," *IEEE Trans. Power Syst.*, vol. 23, no. 1, pp. 150-160, Feb. 2008.
- [154] L. Olmos, J. I. de la Fuente, J. L. Z. Macho, R. R. Pecharroman, A. M. Calmarza, and J. Moreno, "New design for the Spanish AGC scheme using an adaptive gain controller," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1528-1537, Aug. 2004.
- [155] I. Egidio, F. Fernandez-Bernal, and L. Rouco, "The Spanish AGC system: Description and analysis," *IEEE Trans. Power Syst.*, vol. 24, no. 1, pp. 271-278, Feb. 2009.
- [156] A. Feliachi and D. Rerkpreedapong, "NERC compliant load frequency control design using fuzzy rules," *Electr. Power Syst. Res.*, vol. 73, no. 2, pp. 101-106, Feb. 2005.
- [157] B. Tyagi and S. C. Srivastava, "A decentralized automatic generation control scheme for competitive electricity markets," *IEEE Trans. Power Syst.*, vol. 21, no. 1, pp. 312-320, Feb. 2006.

- [158] G. A. Chown and B. Wigdorowitz, "A methodology for the redesign of frequency control for AC networks," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1546-1554, Aug. 2004.
- [159] S. H. Hosseini and A. H. Etemadi, "Adaptive neuro-fuzzy inference system based automatic generation control," *Electr. Power Syst. Res.*, vol. 78, no. 7, pp. 1230-1239, Jul. 2008.
- [160] T. H. Mohamed, H. Bevrani, A. A. Hassan, and T. Hiyama, "Decentralized model predictive based load frequency control in an interconnected power system," *Energy Conv. Manag.*, vol. 52, no. 2, pp. 1208-1214, Feb. 2011.
- [161] A. N. Venkat, I. A. Hiskens, J. B. Rawlings, and S. J. Wright, "Distributed MPC strategies with application to power system automatic generation control," *IEEE Trans. Control Syst. Technol.*, vol. 16, no. 6, pp. 1192-1206, Nov. 2008.
- [162] T. Wen, "Unified tuning of PID load frequency controller for power systems via IMC," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 341-350, Feb. 2010.
- [163] J. Nanda, D. P. Kothari, and K. S. Lingamurthy, "Economic-emission load dispatch through goal programming techniques," *IEEE Trans. Energy Convers.*, vol. 3, no. 1, pp. 26-32, Mar. 1988.
- [164] A. Farag, S. Al-Baiyat, and T. C. Cheng, "Economic load dispatch multiobjective optimization procedures using linear programming techniques," *IEEE Trans. Power Syst.*, vol. 10, no. 2, pp. 731-738, May 1995.
- [165] R. Yokoyama, S. H. Bae, T. Morita, and H. Sasaki, "Multiobjective optimal generation dispatch based on probability security criteria," *IEEE Trans. Power Syst.*, vol. 3, no. 1, pp. 317-324, Feb. 1988.
- [166] J. S. Dhillon, S. C. Parti, and D. P. Kothari, "Multiobjective optimal thermal power dispatch," *Int. J. Electr. Power Energy Syst.*, vol. 16, no. 6, pp. 383-389, Dec. 1994.

- [167] J. S. Dhillon, S. C. Parti, and D. P. Kothari, "Stochastic economic emission load dispatch," *Electr. Power Syst. Res.*, vol. 26, no. 3, pp. 179-186, Apr. 1993.
- [168] C. M. Huang, H. T. Yang, and C. L. Huang, "Bi-objective power dispatch using fuzzy satisfaction-maximizing decision approach," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1715-1721, Nov. 1997.
- [169] P. K. Hota, R. Chakrabarti, and P. K. Chattopadhyay, "Economic emission load dispatch through an interactive fuzzy satisfying method," *Electr. Power Syst. Res.*, vol. 54, no. 3, pp. 151-157, Jun. 2000.
- [170] P. C. Chen and C. M. Huang, "Biobjective power dispatch using goal-attainment method and adaptive polynomial networks," *IEEE Trans. Energy Convers.*, vol. 19, no. 4, pp. 741-747, Dec. 2004.
- [171] D. B. Das and C. Patvardhan, "New multi-objective stochastic search technique for economic load dispatch," *IEE Proc. Gener. Transm. Distrib.*, vol. 145, no. 6, pp. 747-752, Nov. 1998.
- [172] M. A. Abido, "Multiobjective evolutionary algorithms for electric power dispatch problem," *IEEE Trans. Evol. Comput.*, vol. 10, no. 3, pp. 315-329, Jun. 2006.
- [173] C. L. Chiang, "Optimal economic emission dispatch of hydrothermal power systems," *Int. J. Electr. Power Energy Syst.*, vol. 29, no. 6, pp. 462-469, Jul. 2007.
- [174] M. Basu, "Dynamic economic emission dispatch using nondominated sorting genetic algorithm-II," *Int. J. Electr. Power Energy Syst.*, vol. 30, no. 2, pp. 140-149, Feb. 2008.
- [175] S. Dhanalakshmi, S. Kannan, K. Mahadevan, and S. Baskar, "Application of modified NSGA-II algorithm to combined economic and emission dispatch problem," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 4, pp. 992-1002, May 2011.
- [176] L. Wang and C. Singh, "Stochastic economic emission load dispatch through a modified particle swarm optimization algorithm," *Electr. Power*

- Syst. Res.*, vol. 78, no. 8, pp. 1466-1476, Aug. 2008.
- [177] D. J. White, "Dynamic programming, Markov chains, and the method of successive approximations," *J. Math. Anal. Appl.*, vol. 6, no. 3, pp. 373-376, Jun. 1963.
- [178] A. G. Barto and S. Mahadevan, "Recent advances in hierarchical reinforcement learning," *Discrete Event Dynamic Systems*, vol. 13, no. 4, pp. 341-379, Oct. 2003.
- [179] G. Weiß, "Distributed reinforcement learning," *Robot. Auton. Syst.*, vol. 15, no. 1-2, pp. 135-142, Jul. 1995.
- [180] J. G. Vlachogiannis and N. D. Hatziargyriou, "Reinforcement learning for reactive power control," *IEEE Trans. Power Syst.*, vol. 19, no. 3, pp. 1317-1325, Aug. 2004.
- [181] H. Song, C. C. Liu, J. Lawarree, and R. W. Dahlgren, "Optimal electricity supply bidding by Markov decision process," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 618-624, May 2000.
- [182] Q. H. Wu and A. C. Pugh, "Reinforcement learning control of unknown dynamic systems," *IEE Proc. Contr. Theory Appl.*, vol. 140, no. 5, pp. 313-322, Sep. 1993.
- [183] C. L. Tomasevicz and S. Asgarpoor, "Optimum maintenance policy using semi-Markov decision processes," *Electr. Power Syst. Res.*, vol. 79, no. 9, pp. 1286-1291, Sep. 2009.
- [184] E. Zitzler, K. Deb, and L. Thiele, "Comparison of multiobjective evolutionary algorithms: Empirical results," *Evol. Comput.*, vol. 8, no. 2, pp. 173-195, Jun. 2000.
- [185] J. N. Morse, "Reducing the size of the nondominated set: Pruning by clustering," *Comput. Oper. Res.*, vol. 7, no. 1-2, pp. 55-66, 1980.
- [186] L. Ben Said, S. Bechikh, and K. Ghedira, "The r-dominance: A new dominance relation for interactive evolutionary multicriteria decision making," *IEEE Trans. Evol. Comput.*, vol. 14, no. 5, pp. 801-818, Oct. 2010.

- [187] R. B. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press, 1991.
- [188] E. Presman and I. Sonin, "The existence and uniqueness of Nash equilibrium point in an m-player game "Shoot Later, Shoot First!"," *Int. J. Game Theory*, vol. 34, no. 2, pp. 185-205, Aug. 2006.
- [189] H. Keiding and B. Peleg, "Correlated equilibria of games with many players," *Int. J. Game Theory*, vol. 29, no. 3, pp. 375-389, Nov. 2000.
- [190] G. De Marco and J. Morgan, "Slightly altruistic equilibria," *J. Optim. Theory Appl.*, vol. 137, no. 2, pp. 347-362, May 2008.
- [191] M. Breton, A. Alj, and A. Haurie, "Sequential Stackelberg equilibria in two-person games," *J. Optim. Theory Appl.*, vol. 59, no. 1, pp. 71-97, Oct. 1988.
- [192] L. R. Chang-Chien, Y. S. Wu, and J. S. Cheng, "Online estimation of system parameters for artificial intelligence applications to load frequency control," *IET Gener. Transm. Distrib.*, vol. 5, no. 8, pp. 895-902, Aug. 2011.
- [193] T. Jaakkola, M. I. Jordan, and S. P. Singh, "On the convergence of stochastic iterative dynamic programming algorithms," *Neural Comput.*, vol. 6, no. 6, pp. 1185-1201, Nov. 1994.
- [194] J. N. Tsitsiklis, "Asynchronous stochastic approximation and Q-learning," *Mach. Learn.*, vol. 16, no. 3, pp. 185-202, Sep. 1994.
- [195] T. Yu, B. Zhou, K. W. Chan, L. Chen, and B. Yang, "Stochastic optimal relaxed automatic generation control in non-Markov environment based on multi-step $Q(\lambda)$ learning," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1272-1282, Aug. 2011.
- [196] A. Datta, S. P. Bhattacharyya, and L. H. Keel, *Linear Control Theory: Structure, Robustness, and Optimization*. Boca Raton: CRC Press, 2009.
- [197] J. Nanda, A. Mangla, and S. Suri, "Some new findings on automatic generation control of an interconnected hydrothermal system with conventional controllers," *IEEE Trans. Energy Convers.*, vol. 21, no. 1, pp.

- 187-194, Mar. 2006.
- [198] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Mach. Learn.*, vol. 3, no. 1, pp. 9-44, Aug. 1988.
- [199] T. Yu, Y. M. Wang, W. J. Ye, B. Zhou, and K. W. Chan, "Stochastic optimal generation command dispatch based on improved hierarchical reinforcement learning approach," *IET Gener. Transm. Distrib.*, vol. 5, no. 8, pp. 789-797, Aug. 2011.
- [200] B. H. Bakken and H. H. Faanes, "Technical and economic aspects of using a long submarine HVDC connection for frequency control," *IEEE Trans. Power Syst.*, vol. 12, no. 3, pp. 1252-1258, Aug. 1997.
- [201] J. Stuller, *Introduction to Signals and Systems*. Toronto, Ont.: Thomson Press, 2007.
- [202] V. Kola, A. Bose, and P. M. Anderson, "Power plant models for operator training simulators," *IEEE Trans. Power Syst.*, vol. 4, no. 2, pp. 559-565, May 1989.
- [203] Z. B. Du, Y. Zhang, L. Liu, X. H. Guan, Y. X. Ni, and F. F. Wu, "Structure-preserved power-frequency slow dynamics simulation of interconnected ac/dc power systems with AGC consideration," *IET Gener. Transm. Distrib.*, vol. 1, no. 6, pp. 920-927, Nov. 2007.
- [204] B. Price and C. Boutilier, "Accelerating reinforcement learning through implicit imitation," *J. Artif. Intell. Res.*, vol. 19, pp. 569-629, Dec. 2003.
- [205] D. McRuer, "Human dynamics in man-machine systems," *Automatica*, vol. 16, no. 3, pp. 237-253, May 1980.
- [206] M. Lopes and J. Santos-Victor, "A developmental roadmap for learning by imitation in robots," *IEEE Trans. Syst. Man Cybern. Part B-Cybern.*, vol. 37, no. 2, pp. 308-321, Apr. 2007.
- [207] P. Tadepalli and D. Ok, "Model-based average reward reinforcement learning," *Artif. Intell.*, vol. 100, no. 1-2, pp. 177-224, Apr. 1998.
- [208] S. P. Coy, B. L. Golden, G. C. Runger, and E. A. Wasil, "Using experimental design to find effective parameter settings for heuristics," *J.*

- Heuristics*, vol. 7, no. 1, pp. 77-97, Jan. 2001.
- [209] A. S. Poznyak and K. Najim, *Learning Automata and Stochastic Optimization*. New York: Springer-Verlag, 1997.
- [210] H. Tuan-Hao, R. I. McKay, D. Essam, and X. H. Nguyen, "On synergistic interactions between evolution, development and layered learning," *IEEE Trans. Evol. Comput.*, vol. 15, no. 3, pp. 287-312, Jun. 2011.
- [211] P. Jong-Bae, L. Ki-Song, S. Joong-Rin, and K. Y. Lee, "A particle swarm optimization for economic dispatch with nonsmooth cost functions," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 34-42, Feb. 2005.
- [212] R. Caponetto, L. Fortuna, S. Fazzino, and M. G. Xibilia, "Chaotic sequences to improve the performance of evolutionary algorithms," *IEEE Trans. Evol. Comput.*, vol. 7, no. 3, pp. 289-304, Jun. 2003.
- [213] F. N. Lee and A. M. Breipohl, "Reserve constrained economic dispatch with prohibited operating zones," *IEEE Trans. Power Syst.*, vol. 8, no. 1, pp. 246-254, Feb. 1993.
- [214] C. J. Barnard and R. M. Sibly, "Producers and scroungers: A general model and its application to captive flocks of house sparrows," *Anim. Behav.*, vol. 29, no. 2, pp. 543-550, May 1981.
- [215] M. Laumanns, L. Thiele, and E. Zitzler, "Running time analysis of multiobjective evolutionary algorithms on pseudo-Boolean functions," *IEEE Trans. Evol. Comput.*, vol. 8, no. 2, pp. 170-182, Apr. 2004.
- [216] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182-197, Apr. 2002.
- [217] M. Farina, K. Deb, and P. Amato, "Dynamic multiobjective optimization problems: Test cases, approximations, and applications," *IEEE Trans. Evol. Comput.*, vol. 8, no. 5, pp. 425-442, Oct. 2004.
- [218] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York: Springer-verlag, 1999.
- [219] M. T. Jensen, "Reducing the run-time complexity of multiobjective EAs:

- The NSGA-II and other algorithms," *IEEE Trans. Evol. Comput.*, vol. 7, no. 5, pp. 503-515, Oct. 2003.
- [220] R. D. Zimmerman, C. E. Murillo-Sánchez, and D. Gan, "MATPOWER: A Matlab power system simulation package", [Online]. Available: <http://www.pserc.cornell.edu/matpower>, 2005.
- [221] E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm," in *Proc. EUROGEN 2001. Evolutionary Methods for Design, Optimization and Control with Applications to Industrial Problems*, Athens, Greece, Sep. 19-21, 2001, pp. 95-100.