1

Introduction

Optimization theories and approaches have been extensively applied to power system planning and operation problems. This is a rather traditional and ongoing research area [1]. With the complication of power systems, the deregulation of the power industry, and the development of smart grids, many new problems have emerged and new methods have been developed. Many optimization theories and approaches have acquired industrial application and introduced technical and economic benefits. The mathematical optimization methods applied in power systems include linear programming, nonlinear programming, mixed integer programming, dynamic programming, artificial intelligence, stochastic programming, etc. This book focuses on the advanced theories and approaches from the perspective of large-scale complex systems, rather than the traditional ones. However, to begin with the fundamentals, we will first review the basic optimization applications in power system planning and operation.

The aims of this chapter are as follows:

1. To present a broad review of mathematical optimization applications to power system planning and operation, which is the foundation for the theories and approaches presented in the subsequent chapters.
2. To explain the basic concepts to those interested in the optimization field, but unfamiliar with power system problems and terminology. It is hoped that this chapter may motivate some people to become involved in the challenging power field.
3. To summarize the results of traditional power system research, to allow the reader to understand the differences among them and the more advanced approaches presented in books, and to encourage new development and further research.

To give the reader a unified mathematical description of different power system optimization problems, the generalized notation used in this book, such as x and u for variables, and f, h, and g for functions, and their power system meanings are

Power System Optimization – Large-scale Complex Systems Approaches, First Edition.
Haoyong Chen, Honwing Ngan and Yongjun Zhang.

explained. Vectors (lower-case) sometimes and matrices (upper-case) usually are in bold face; and matrix transposition is indicated by a superscript T, such as $\mathbf{A}^{\mathrm{T}}$.

The problems discussed include generation, transmission, and distribution expansion planning, optimal operation problems such as hydrothermal unit commitment and dispatch, optimal load flow and volt-ampere reactive (VAR) optimization, and optimization models of electricity markets based on theories of microeconomics.

Numerous important works have appeared on these topics in books and journals all over the world. It is an impossible task to discuss all of them. Since the objective of this chapter is to introduce the basic concepts and methods of power system optimization, we will lay the emphasis of our discussion on research reported by IEEE papers in *IEEE Transactions on Power Systems* and Technical Meetings.

1.1 Power System Optimal Planning

Power system expansion planning is traditionally decomposed into load forecasting, generation planning, and transmission planning. Load forecasting is the basis for power system planning, which provides the basic data for calculation of electric power and energy balance. Although generation planning and transmission planning are essentially indivisible, these two issues have to be decomposed and solved separately and further coordinated due to their different focuses and the difficulty in solving them as a whole.

Traditional power system planning is based on scheme comparison, which selects the recommended scheme from a few of the viable options with some technical and economic criteria. However, because this approach is empirical, the final result is not necessarily optimal. With the fast development of power technologies, the rapidly growing demand for electricity, and the increasingly diversified energy resources used in power generation, the generation mix becomes increasingly complicated. On the other hand, large-scale interconnected systems across different areas have been formed gradually. All these factors have brought difficulties to the economic and technical assessment of power system planning schemes, and traditional planning approaches are difficult to adapt to these challenges. Fortunately, the development of computer science, systems engineering, operational research, and other research areas has provided new means for the optimization of power system planning. Theory and practice in power system optimal planning have made considerable progress in recent years. A number of commercial planning software packages have emerged and their benefits have been affirmed in the power industry.

The objective of power system planning is to determine what schemes are the most beneficial from the overall and long-term perspective. This requires us to choose the best planning scheme from all possible choices. The application of power system optimal planning theories and methods not only can have more accurate and comprehensive technical and economic evaluation, but also can evaluate the impacts of various uncertainties by sensitivity analysis, so that the planning results are produced with a higher referential value.

1.1.1 Generation Expansion Planning

The objective of generation expansion planning is to choose the least expensive expansion scheme (type, number, capacity, and location of generating units), in terms of investment and operation costs, that satisfies certain constraints. The key constraints are electric power and energy balance, which means that the total power and energy produced by all the generating units can meet the requirement of demand. Other technical constraints, such as limitation of resources, also need to be met. Generally, generation expansion is carried out over a planning horizon of many years, which turns into a dynamic optimization problem.

Several key issues should be analyzed quantitatively in generation expansion planning, such as: annual investment flow and operating cost, quantity of primary energy resources used in generation, reliability of electric power supply, etc. The investment cost of building a particular plant in a given year is independent of the other decisions in a given scheme. However, the operating cost is much more complicated, and is related to the generation mix, system load, generating unit outages, transmission network losses, availability and cost of energy from neighboring systems, fuel costs, etc. Some influencing factors are intrinsically random, such as generating unit outages. The fact that units must be added in discrete sizes presents a further complication. Considering all these conditions, the mathematical model of generation expansion planning is large-scale, nonlinear, discrete, and stochastic, which is a very difficult problem to solve.

Generation expansion planning has long been of interest to researchers, and many sophisticated and effective techniques have been developed. The approaches differ in the questions they are intended to answer, the model details, and the optimization methods.

The early work often used linear programming models [2, 3]. The objective functions takes the following form:

$$J = c^t x + d^t u$$

where x denotes the capacities of different types of generating units installed in each year and u specifies the energy produced by each power plant (or plant type). A number of different load levels are considered here. The investment cost c and the operating cost d should be calculated with the method of technological economics. The load levels related to d and u are obtained by dividing estimates of the load duration curves into a number of discrete segments (Figure 1.1). The variables u and x are related through linear constraints so that a plant cannot produce power exceeding its installed capacity. Other constraints limit the capacity of certain types of units and require total capacity to exceed expected load. This formulation is a high-dimensional optimization

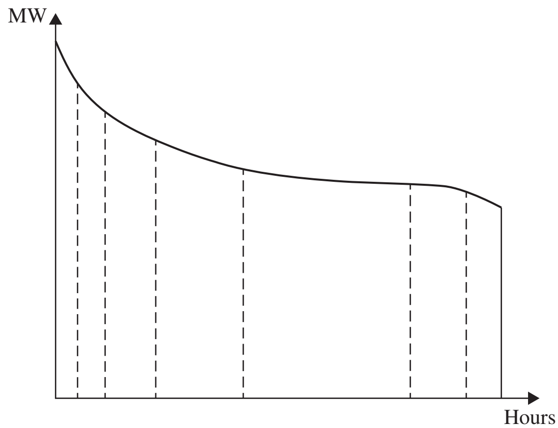

Figure 1.1 Typical load duration curve.

problem. Decomposition techniques such as Dantzig–Wolfe decomposition may be needed to solve it.

A dynamic programming based model of generation planning has been presented by Booth [4, 5]. The method can handle integer variables and nonlinear constraints. The random variables are treated with a probabilistic approach. As a significant innovation, the expected outage rates for various units are considered by modification of the load duration curve. The problem is formulated as: choose $v(t)$ (capacity additions in year t) to minimize

$$J = E\left\{ \sum_{t=1}^{T} f(t, x(t), v(t), \dots) \right\} \tag{1.1}$$

where

$$x(t + 1) = x(t) + v(t)$$

The function $f(\cdot)$ is related to probabilistic load models, fuel models, etc. A variety of technical and economic constraints are considered.

The problem is decomposed into a series of forward dynamic programming problems. A pretreatment is employed to dynamically reduce the dimensionality of the problem. However, the computational burden is still heavy.

A more advanced generation planning model JASP has been proposed by Chen [6], which decomposes the generation planning problem into a high-level power plant investment decision problem and a low-level operation planning problem and solves them by a decomposition–coordination method. Lagrangian relaxation is used to solve the power plant investment decision problem, and a probabilistic production simulation based on the equivalent energy function method is used to solve the operation planning problem. Simulation results show that JASP can not only overcome the "curse of dimensionality" but also find an economical and technically sound generation planning scheme.

1.1.2 Transmission Expansion Planning

Transmission expansion planning is an important part of power system planning, whose task is to determine the optimal power grid structure according to the load growth and generation planning schemes during the planning horizon to meet the requirements of economic and reliable power delivery. In general, transmission planning should answer the following points:

- where to build a new transmission line,
- when to build a new transmission line, and
- what type of transmission lines to build.

Transmission expansion planning is closely related to generation planning. It is based on generation planning, and in turn has some impact on the latter. In generation planning, the influence of geographical distribution of power plants and transmission costs are generally not considered or just considered cursorily, and it is possible that the original generation planning scheme should be modified during transmission planning. Therefore, generation planning and transmission planning should be decomposed and further coordinated so that the whole power system planning can be optimal.

The basic principle of transmission planning is to minimize the power grid investment and operating costs under the premise of ensuring safe and reliable electric power delivery to the load center. Compared with generation planning, transmission planning is more complex. First, the transmission planning should consider the specific network topologies, and each line in the rights of way must be treated as an independent decision variable. Thus the dimension of transmission planning decision variables is higher than that of generation planning. Second, transmission planning should satisfy very complex constraints. Some constraints are related to nonlinear equations, and even related to differential equations. Third, many factors that are either random or difficult to predict are extremely important, such as future load growth in various areas. Therefore, it is difficult to establish a perfect transmission model and even more difficult to solve it.

Transmission planning has two kinds of formulations: static and dynamic. Static transmission planning is concerned only with the planning scheme in a future target year, and it is not necessary to consider the planning scheme transition, which is also known as a level-year plan. Because static transmission planning does not answer the question *when* to build new transmission lines, it is not necessary to consider the time value of capital. With a longer planning period, the planning horizon needs to be divided into several level years and the scheme transition between the level years needs to considered. In this case, we must determine when and where to build new transmission lines. This kind of planning is called long-term or dynamic transmission planning.

An early static formulation was presented in Puntel [7], which attempts to design the optimal network structure for a specified future time. The present network, available rights of way, costs, and future loads and generation levels are assumed known.

The problem is to choose x, the susceptance installed in the rights of way, to minimize

$$J = c^t x + \sum_{k=1}^{M} d_k (P_k / P_{c_k})^{2n} \tag{1.2}$$

The first term $c^t x$ is the investment cost of transmission capacity; the second term is the overload penalty on each of the M transmission lines or transformers. The load P_k on element k is a nonlinear function of x, the given load and generation conditions, even when the network model is linear. There are also inequality constraints on each element of x.

Optimization is performed by first computing a gradient vector $\partial J/\partial x$ through use of an adjoint network. A band around this gradient in the portion of the x space bounded by the limits of additions in each right of way is searched for integer solutions. A cost J is computed for each such x found. The x that minimizes J is the starting point for the next iteration and for the computation of the next $\partial J/\partial x$. The search ends when no improving solution is found.

Once the network has been designed, a sensitivity analysis is performed to determine which contingencies would be the most critical. Each serious outage is tested individually and a composite $\partial J/\partial x$ is computed by summing the gradient of all tests. As in the normal situation, this gradient guides the addition of new lines. The process terminates when as many as possible overloads caused by the contingencies have been eliminated by appropriate network additions.

Garver [8] considered the static optimization problem of designing a network to meet a specific load. The problem is formulated as a power-flow problem. Linear programming is used to find the most direct route from generation to loads: all rights of way can transmit power, but those without transmission lines are penalized to encourage flow through the existing network. A line is added to the right of way with the biggest overload and then a new linear flow is computed. The process terminates when all overloads have been eliminated. The estimated flows on the final network are fairly close to those computed by a standard load flow.

The linear program minimizes an approximation to cost of new facilities, as the penalty term of an overload can be related to the cost of constructing a line in that right of way.

An improved genetic algorithm (GA) approach to optimal multi-stage (dynamic) transmission network planning is presented in a more recent work [9]. The multi-stage planning of a transmission system has to consider not only how to form the network schemes of every stage but also how to coordinate the network schemes of every stage.

Because of the combinatorial nature of the transmission network planning, it has not been well solved by conventional optimization techniques. In the proposed improved GA approach, a fitness function that includes investment and an overload constraint is constructed. The overload is checked by DC load flow. A concise codification model called "redundant binary-coded technique" is proposed. Using this technique, the crossover operation can be executed inside the gene so that the re-combinatorial and search functions of the crossover operator are well utilized. The simulated annealing selector is used to adjust the fitness function in the evolution process. Some improvements are employed to speed up the convergence of the algorithm, such as keeping excellent seeds, mutation in pairs, etc. Based on the proposed model, a computational program has been developed. Three case studies are applied to demonstrate the usefulness and effectiveness of the suggested multi-stage transmission network planning model.

All of these approaches are suboptimal and involve important approximations and simplifying assumptions. The static approaches can handle larger networks, but at the cost of ignoring growth dynamics. The dynamic approaches often suffer from the "curse of dimensionality".

1.1.3 Distribution System Planning

The current development of power systems puts increasing emphasis on distribution systems, which are the lower-voltage networks supplying power from the high-voltage transmission systems to the loads. Distribution system planning has different considerations along with transmission planning, and several special concepts and factors need to be addressed.

Distribution systems tend to be less geographically extensive than transmission systems. However, the network structures are often more complicated and consist of more branches and nodes. On the other hand, the operation of distribution systems may be easier than transmission systems and stability constraints are often not included. Reliability and power quality are usually the main concerns of users and should be taken into consideration in distribution system planning.

The objective of distribution system planning is to determine the optimal construction scheme under the premise of meeting load growth and securing a reliable supply of electricity based on the results of load forecasting during the planning horizon and existing network status, so that the investment and operating costs of the distribution system are minimal. Distribution system planning also has static (one-stage) and dynamic (multi-stage) formulations. In dynamic distribution system planning, the correlations of decision variables among different stages in the planning horizon should be taken into account. The mathematical models are often discrete, nonconvex, nonlinear, and large-scale, and various optimization techniques have been employed.

The early work [10] treats the maximum flow-through capability of the network as the sum of the component capacities and does not explicitly consider whether this capability is actually exploited. This simplification makes the constraints much easier to handle. The planning scheme obtained with this simplification may be optimal for a problem with complex constraints, or it may be used as a bound or goal in optimization subjected to realistic network constraints. Adams and Laughton [10] emphasized the security aspect of the problem by determining which of the possible expansion states at each time period of a planning horizon are feasible, in the sense that they do not produce overloads of circuits or high-voltage transformers. The overall costs (capital and losses) are minimized by determining the proper sequence of states of planning time periods, and dynamic programming is used in optimization.

In a more recent work [11], a GA approach is applied to the optimal multi-stage planning of distribution networks. The objectives for distribution system planning are related to providing the designs and associated implementation plans necessary for an orderly expansion of facilities, minimizing new facility installation costs and operation costs, as well as achieving an acceptable level of reliability. Complex operational constraints such as voltage drop and line thermal limits are considered. GA allows the representation of nonlinearities that are hard to include in mathematical programming methods, and produces multiple solutions that enhance the opportunity for multi-criteria decision-making.

Currently, deregulation in the power system industry and the invention of new-generation technologies have led to innovations in distribution system planning [12]. Distributed generation (DG), with many attractive economical and technical features, in medium- and low-voltage parts of the grid, is one of the alternatives to reinforce distribution systems. In Naderi *et al.* [12], a dynamic distribution system planning model was proposed that considers DG integration into a network as an option to meet the load growth in the planning horizon. An optimal power flow (OPF) is proposed to minimize capital costs for network upgrading, operation and maintenance costs, and the cost of losses for handling the load growth in the planning horizon. A year-dependent decision variable is attached to each investment alternative and results in a dynamic planning scheme, which reduces total planning costs by determining the best timing schedule for investment in network upgrading. A modified genetic algorithm is used to find the optimal topology solution.

1.2 Power System Optimal Operation

1.2.1 Unit Commitment and Hydrothermal Scheduling

Power system optimal operation is a very complex system optimization problem, which is hard to solve as a whole and so is often decomposed into a series of subproblems dealt with separately [13]. For short-term generation scheduling, it is often decomposed into unit commitment, hydrothermal scheduling, power exchange planning, and fuel planning, etc. The purpose of unit commitment and hydrothermal scheduling is to determine which units should be on-line and their generation levels in a power system during a given scheduling horizon (usually a day or a week).

The objective is to minimize total operational costs during the scheduling horizon while satisfying the constraints of power system security and power quality. Unit commitment is a very basic optimal operation problem, whose benefit is generally greater than that of economic dispatch.

Wood [14] describes the basic mathematical models and solution methods of unit commitment and economic dispatch. Unit commitment is a high-dimensional, non-convex, discrete, and nonlinear optimization problem, and it is difficult to find the theoretically optimal solution. However, because it can bring significant economic benefits, various methods have been extensively studied to solve this problem, such as the heuristic, priority list, dynamic programming, integer programming and mixed integer programming, branch-and-bound method, and Lagrangian relaxation method. Artificial intelligence has also be introduced, such as expert systems, artificial neural networks, simulated annealing, genetic algorithms, etc. Sheble and Fahd [15] survey the solution methods and related references of unit commitment.

The primary energy sources used for electric power generation can be broadly classified as renewable and nonrenewable resources. Fossil fuels such as coal, natural gas, oil, and nuclear fuel are nonrenewable resources, which are used for electricity generation in thermal power plants. The most widely used renewable resource for electricity generation is hydro power. Nowadays there are fast developments with other renewable sources such as wind power, solar energy, marine energy, and biomass.

In a thermal plant, electric power is generated as a result of mechanical rotational energy produced by either steam turbines or combustion turbines. From an economic operational point of view, our concern is the relation of fuel cost to the active power generation of the unit, which is an efficiency type model. The fuel cost is equal to the fuel quantity consumed multiplied by the fuel price. The heat rates are often used to express generation efficiency, which include the average heat rate (dividing the fuel quantity by the generation quantity) and incremental heat rate (differentiating the fuel quantity with respect to the generation quantity). The fuel cost curve is often modeled as a quadratic function, and the values of coefficients can be obtained by statistical estimation from experimental data of heat rates. The treatment so far has been static: for the unit commitment problem, dynamics must also be considered. For example, thermal units are shut down and started up at various instants. There are costs related to these operations and time constraints that do not allow plants to be switched in and out frequently (minutes or a few hours).

In a hydro plant, turbines convert the water potential energy into kinetic energy, which in turn is converted into electricity by generators. Hydroelectric installations are classified into two types: conventional and pumped storage. The conversion type is further classified into two classes: storage and run-of-river. The water system modeling problem basically involves a water balance equation at each reservoir, which relates the inflow of water to the reservoir, the water volume in the reservoir, and the rate of usage and spillage. The inflow could be from natural sources (rain, snow, unregulated rivers), or from a reservoir system in which upstream reservoirs release water. In this last case, the dynamics of the reservoir system are important. On the rate of usage and spillage, the question of whether only one plant or a series of cascading plants will use the water must be considered. There are also pumped-storage plants where water is pumped into a higher-level reservoir at time periods when energy from efficient plants can be used for this purpose, and water is discharged through a hydro plant to generate energy at peak periods of consumption where the more efficient plants are at their maximum output.

The most important constraint of unit commitment and hydrothermal scheduling is power balance condition, which is that the sum of the electric power generated by all units must equal the system load plus network loss at every scheduling period. Other constraints include reserve requirements for the system as a whole and for certain specified areas. There are minimum up-time and down-time constraints for each unit. Constraints on hydro energy and power should also be respected.

The solution methods can be classified into heuristic methods, mathematical optimization, artificial intelligence, etc. The earliest ones are heuristic methods such as local optimization and priority list, which are empirical without a strict theoretical foundation, but are often useful in practice. Mathematical optimization methods are those with a strict mathematical basis, among which the most successful ones are dynamic programming and mixed integer programming.

Dynamic programming (DP) is a mathematical method to solve the multi-stage decision optimization. In the enumeration of various possible combinations of states, this method cleverly abandoned solutions that need not be considered. In the UC problem, the entire scheduling horizon is divided into several periods, and each period is a stage of dynamic programming. The states of each stage are all the possible combinations of the unit ON/OFF states. From the initial stage, the cumulative cost (including the fuel cost of generation and start-up cost) to reach each stage is calculated forward, and then, from the state with the smallest cumulative cost in the final stage, a backtracking procedure is carried out, by sequentially recording the state with the total cumulative cost in each stage, and then the optimal UC scheme can be obtained. The problem of dynamic programming is that the computational burden will increase dramatically with the number of units and scheduling periods, which results in the "curse of dimensionality". To overcome this difficulty, different skills are used to limit the number of states. Dynamic programming has been widely used in unit commitment and hydrothermal scheduling, and has been incorporated into some practical scheduling/dispatch software packages.

Mixed integer programming (MIP) includes both integer and noninteger variables, and it can be further divided into linear and nonlinear mixed integer programming according to the type of other variables except for the integer variables. MIP is a very difficult problem, and the commonly used methods include branch-and-bound, Benders' decomposition, generalized Benders' decomposition methods, and so on. In the branch-and-bound method, a branch-and-bound tree is formed and the root node is a relaxation of the original problem. For example, the integer variables are replaced with continuous variables in a certain interval. Its child nodes are also a series of relaxation of the original problem, which can be obtained by fixing the values of some integer variables at the root node. The relaxed problems corresponding to the child nodes have disjoint solution spaces, and the union of these solution spaces is just the solution space of the root node. Each child node is further decomposed into a series of subproblems, and this process is repeated until the leaf nodes of the tree. Because the solution space of a node of the tree is the relaxation of the solution spaces of all its descendant nodes, in a minimization problem, the lower bound of its solution must therefore be the lower bound of the solutions of all its descendant nodes. The key to the branch-and-bound method is that, in the process of calculation, if the solution (or its lower bound) of a node is bigger than a known feasible solution of the original optimization problem, its descendant nodes need no longer be considered. This allows unnecessary calculations to be eliminated.

Benders' decomposition method solves problems in the following form:

$$\max_{x,y} \{c^{\mathrm{T}}x + f(y) \mid \mathbf{A}x + F(y) \leq b,\ x \in R^p,\ y \in S \subseteq R^q\} \tag{1.3}$$

where R^p and R^q are respectively p- and q-dimensional Euclidean spaces; S is any subset of R^q; $\mathbf{A}$ is an $m \times p$ matrix; $f(y)$ is a scalar function and $F(y)$ is an m-component vector function, which are defined on S; and b and c are constant vectors defined on R^m and R^p. Benders' decomposition method decomposes the problem into two subproblems: one is a programming problem defined on S (which may be linear, nonlinear or discrete, etc.), and the other is a linear programming problem defined on R^p. The two subproblems are solved through a multi-step iterative procedure.

MIP solves the mathematical models of unit commitment/hydrothermal scheduling directly, without adding too many restrictions or assumptions. MIP is the mainstream solution method of scheduling problems in the electric power industry.

Power systems are typical large-scale systems, and the optimization and control theories of large-scale systems can be applied. The decomposition and coordination approach of large-scale systems started from Dantzig and Wolfe's decomposition for linear programming [16], and the Lagrangian relaxation (LR) method has been employed in unit commitment. LR is a class of optimization algorithms for solving complex integer and combinatorial optimization problems, which is based on the following ideas. Many difficult integer programming problems are composed of a series of subproblems interrelated by some constraints, which are relatively easy to resolve. Based on this characteristic, the Lagrangian problem is formed by adding a penalty term to the objective function, which equals the sum of products of constraint violation amounts and their respective dual variables. The Lagrangian problem is relatively easy to solve, and for the minimization problem, its optimal value is the lower bound of the optimal value of the primary objective function. The Lagrangian relaxation method will be discussed in detail in Chapter 2.

In the Lagrangian relaxation method for unit commitment, all constraints are divided into two categories: one category includes the systemwide constraints, such as the load balance constraints and spinning reserve constraints; the other category includes the individual unit constraints, such as the generator output power limit constraints, the minimum up-time and minimum down-time constraints, ramp rate constraints, spinning reserve constraints, etc. The systemwide constraints can be written as the penalty term of the objective function, and then the Lagrangian function can be formed. The Lagrangian function can be decomposed into a series of individual unit subproblems, which are generally solved using the dynamic programming method, and the dual problem is often solved by the subgradient method. Application of Lagrangian relaxation in unit commitment began in the 1970s and there is a great deal of literature.

Unit commitment and hydrothermal scheduling based on Lagrangian relaxation have been studied systematically [17–19]. In Guan *et al.* [17], dynamic programming is used to optimize the subproblems and optimization of multipliers is carried out with an adaptive step-size subgradient algorithm. The advantages of the proposed approach include that the generator power outputs do not need to be discretized, and there is a systematic approach to handle ramp rate constraints, as well as an effective initialization procedure. The method is further extended to the hydrothermal scheduling problem in Yan *et al.* [18], where the thermal and hydro power plants are coordinated through Lagrange multipliers, and the hydro power plant subproblems are solved with the priority list method. In Guan *et al.* [19], where the dynamic characteristics of a pumped-storage power station have been considered, the scheduling problem in the whole horizon is solved by dynamic programming on the basis of optimization of single-period operation.

Lagrangian relaxation is a kind of integer and combinatorial optimization algorithm with a mature theoretical foundation and is particularly suitable for solving large-scale system optimization problems. It has the following advantages in unit commitment and hydrothermal scheduling: with the increase of unit number, the computational burden increases almost linearly, which means the dimensional obstacle can be overcome; and the method is very flexible, which means not only that it can successfully solve the unit commitment problem, but also that it can be extended to hydrothermal scheduling problems. However, the algorithm also has some disadvantages: owing to the nonconvexity of the objective function, a duality gap exists and some special measures should be adopted to construct the feasible solution of the primal problem on the basis of the optimal dual solution; and the iterative process of the algorithm may oscillate.

With the development of computer science and artificial intelligence, many new methods have appeared, the most successful of which are the genetic algorithm, simulated annealing, tabu search, expert systems, and so on.

1.2.2 Economic Dispatch

The objective of economic dispatch is to minimize the fuel cost of thermal power plants, assuming that hydro generation has been given. The unit commitment scheme of thermal units is also supposed known. Many power systems today are operated under economic dispatch with calculations made every few minutes. Under normal circumstances, control signals are sent to generating units to adjust their power output in accordance with optimization results. In interconnected systems, optimization results are further adjusted by a load frequency control (automatic generation control, AGC) process, which aims at keeping deviation of frequency and total power interchange with neighboring utilities within preset values. The objective function is the sum of thermal fuel costs as a function of generation power output P. The most basic constraint is the power balance equation. Inequality constraints place limits on P. The mathematical formulation is thus as follows:

$$\min\ f(x)$$

subject to

$$\begin{aligned} h(x) &= 0 \\ \tilde{x} \le x &\le \overline{x} \end{aligned} \tag{1.4}$$

where

$$x = (x_1, \ldots, x_n)^{\mathrm{T}}$$

The solution of this problem is found by observing the Kuhn–Tucker optimality conditions:

$$\begin{gathered} \frac{\partial f(x)}{\partial x} + \lambda \frac{\partial h(x)}{\partial x} = 0 \\ h(x) = 0 \\ x \ \text{ within limits} \end{gathered} \tag{1.5}$$

More recent work deals with more complicated formulations. A particular interest of researchers is to introduce power flow equations directly as equality constraints or as implicit functions. In [20], nodal voltages are used as the state variables of the system. As all other variables can be expressed in terms of nodal voltages, the constraints can be set up based on them for real and reactive generation, voltage magnitudes, flows in transmission lines, etc. Then the economic dispatch problem becomes

$$\min\ f(x) \tag{1.6}$$

subject to

$$\begin{aligned} h(x) &= 0 \quad (k \ \text{ equations, } \ k < n) \\ g(x) &\ge 0 \quad (m \ \text{ inequalities}) \end{aligned} \tag{1.7}$$

where

$$x = (x_1, \ldots, x_n)^{\mathrm{T}}$$

In this formulation, it is necessary that all network equations appear directly as equality constraints or as a part of $g(x)$ in order for the formulation to be complete.

This is the standard nonlinear programming problem. In [20], it is solved by various penalty function methods of constrained optimization.

With the continuing search for alternatives to conventional energy sources, it is necessary to include wind energy conversion system (WECS) generators in the economic dispatch problem. The uncertain nature of the wind speed is represented by the Weibull distribution in Hetzer *et al.* [21]. In addition to the classic economic dispatch factors, factors to account for both overestimation and underestimation of available wind power are also included. The optimization problem can be numerically solved by Lagrangian relaxation, evolution algorithms, simulated annealing (SA), etc. The solution of the economic dispatch problem via the model presented is dependent on the values of many coefficients, such as the scale factor in the Weibull distribution function, the reserve cost for overestimating the wind energy, and the penalty cost for underestimating the wind energy. The level of wind power scheduled from a particular WECS is strongly dependent on the values of the reserve and penalty cost factors associated with the WECS. If the reserve cost coefficient is increased, the scheduled amount of wind power will be reduced, because it becomes more costly to overestimate the amount of wind power available. Conversely, if the penalty cost coefficient is increased, it becomes more costly to underestimate the amount of wind power available, and the system operator has an incentive to increase the scheduled amount of wind power. Economic dispatch considering uncertainty in renewable generation is still an open topic to be further investigated.

1.2.3 Optimal Load Flow

The most rigorous steady-state electric power system network model is provided by the load flow equations, which refers to the problem of solving the network equations. Generally, two groups of data are given for each node: P, Q for load nodes, and P, V for generator nodes, according to the physical quantities that can be easily controlled at those loads. There is a generation node that is treated differently and given by V, σ. The value of σ is set to zero as a reference for the load flow equations and P is left to balance the power of the whole network. This node is called a slack node. Mathematically, the load flow problem can be described as the following n equations:

$$h(x) = 0, \quad x = (x_1, \dots, x_n)^{\mathrm{T}}. \tag{1.8}$$

The x or voltage variables are called the state variables of the system. Once they are known, all other electric quantities, such as reactive generation, slack-bus generation, and power flows in transmission lines, can be directly calculated.

Given a set of loads, different load flow solutions can be obtained by varying other input data. An optimal load flow problem is one that incorporates this exact model in the formulation, which refers to an operating state or load flow solution where some power system quantity is optimized, subject to constraints on the problem variables and on some functions of these variables. Optimal load flow has received more and more attention from power system planning, operation and control areas since the 1960s. At present, the application of advanced control equipment has brought new motivation to optimal load flow research. Furthermore, as the power industry moves into a more competitive and commercial environment, optimal load flow becomes even more important.

Optimal load flow is similar to economic dispatch in form. However, economic dispatch uses formulations of far lower dimension and sophistication. Some relevant variables such as generator voltage magnitudes are not included in optimization. As a result, constraints imposed by considerations of system security are not easily handled by procedures using power balance or other traditional models. The advantages of optimal load flow lie not so much in higher accuracy – more important is its ability to include security constraints in the formulation. A number of problems can be defined by different choices of decision variables, objective functions, and constraints. Some of these problems will be described below together with the techniques used to solve them.

In the approach of Dommel and Tinney [22], two sets of variables are defined: the state variables of the load flow problem, the x variables; and the load flow data control variables, the u variables. The network equations relate the x and u variables. Other load flow quantities that remain fixed, such as power demand at load buses, can be expressed as functions of x and included equality constraints. All inequality constraints not directly on control variables u are called "functional constraints" and are used to penalize the cost function. The problem can be formulated as follows:

$$\min f(x,u) \quad (x = (x_1, \ldots, x_n)^{\mathrm{T}}) \tag{1.9}$$

subject to

$$\begin{aligned} h_1(x) &= 0 \quad (k \text{ equalities}, \ k < n) \\ g(x,u) &\geq 0 \quad (m \text{ functional inequalities}) \\ \tilde{u} \leq u &\leq \overline{u} \quad (n-k \text{ control variables}) \\ h_2(x,u) &= 0 \quad (n-k \text{ network equations}) \end{aligned} \tag{1.10}$$

An $F(x,u)$ function is defined by penalizing $f(x,u)$ with violated functional inequality constraints; n equations $b(x,u)$ can be defined by joining the k equality constraints $h_1(x) = 0$ and the $n-k$ network equations $h_2(x,u) = 0$. The Kuhn–Tucker optimality conditions are as follows:

$$\begin{aligned} &h(x,u) = 0 \\ &\frac{\partial F}{\partial x} + \left(\lambda^t \frac{\partial h}{\partial x}\right)^t = 0 \quad \text{or} \quad \left(\frac{\partial h}{\partial x}\right)^t \lambda = -\frac{\partial F}{\partial x} \\ &\frac{\partial F}{\partial u} + \left(\lambda^t \frac{\partial h}{\partial u}\right)^t = 0 \quad \text{or} \quad \nabla_u F = 0 \ \text{(reduced gradient)} \end{aligned} \tag{1.11}$$

Choosing the feasible values for u, the first set of equations is solved for x. This is the load flow problem. If Newton's method is used with triangular factorization techniques to solve the load flow, the calculation of λ involves a repeat solution with the triangularized Jacobian $\partial h/\partial x$, using the second set of equations. The third group defines the reduced gradient of F with respect to only u variables. This gradient can be used to calculate corrections for u variables

$$\Delta u = \alpha \nabla_u F \tag{1.12}$$

According to the Kuhn–Tucker optimality conditions, violations of inequality constraints on u variables due to the above correction Δu should be prevented by setting u to its limit. At this point, a new iteration is performed, and this process is continued until $\nabla_u F = 0$ for all off-limit u variables.

In a power system, there are strong couplings between real power flows and voltage angles or reactive power flows and voltage magnitudes. However, the P, δ and Q, V variable sets are weakly coupled. Billinton and Sachdeva [23] discussed these coupling effects and suggest a decomposition technique for solving the economic dispatch. The proposed approach decomposes the complete problem into two halves, i.e. the optimum voltage evaluation, and the real power optimization with optimally determined voltages. Both problems are solved by the Fletcher–Powell technique. This decomposition approach is still used as reference for optimal load flow of large-scale power systems.

1.3 Power System Reactive Power Optimization

Reactive power optimization is a kind of large-scale nonlinear optimization problem with multi-variable, multi-restriction, and multi-extreme mathematical characteristics. It has been a hot issue in global electrical fields for years. A great deal of deep research has been carried out and there are thousands of publications on reactive power optimization models and algorithms for power grids with different characteristics.

In the power system planning stage, reactive power optimization means optimal reactive power planning (ORPP). That is an off-line problem concentrating on the optimal allocation of reactive power compensation devices in order to enhance the controllability of power systems. In the power system operation stage, reactive power optimization means optimal reactive power dispatch (ORPD) or optimal reactive power control (ORPC). That is an on-line problem concentrating on the optimal regulation of reactive power and voltage control devices in order to improve the economy and safety of power system operation. On-line problems have very high requirements on the effect of optimization and the calculation speed of solution.

The ORPD problem can be described as an objective function and a set of constraint conditions. The characteristics of the ORPD problem are listed as follows:

1. Multi-objective
2. Number of constraint conditions of various types
3. Nonlinear objective function and constraint conditions

4. Uncertainty of load and operation mode
5. Discrete control variables that cannot be regulated frequently
6. Nonconvexity and multi-extreme
7. Objective function is hard to describe using formula including control variables
8. Solution space lacks connectivity

ORPD encounters two major problems in engineering applications:

1. The discrete control devices cannot be regulated frequently under the change of operation state of power systems at different times, which leads to the time coupling characteristic of ORPD.
2. The control of different regional power networks influences each other when adopting a hierarchical and zoning control strategy, which leads to the space coupling characteristic of ORPD.

The scale of power systems is expanding constantly, which results in higher requirements on a reactive power optimization algorithm. Some characteristics of an algorithm, such as whether it can converge to the optimization solution reliably and rapidly, and whether it can detect and solve infeasible problems, become more and more important. Because the feasible and infeasible subspaces of the ORPD solution space of complicated power networks interlace with each other, solving the ORPD problem becomes very difficult.

The interior point method is one of the most applied mathematical methods for the reactive power optimization problem. It is fast and has good convergence, and its calculation time is insensitive to the scale of the problem. However, how to probe and deal with the infeasible solution in the solution process is one of the obstacles of the interior point method. In addition, solving a high-order correction equation is still a bottleneck in the calculation for the interior point method.

Artificial intelligence algorithms, such as genetic algorithm, particle swarm optimization algorithm, and so on, make searching the global optimal solution in large-scale space possible because of their characteristics of organization, adaptation, self-learning, and parallelism. However, acquiring the theoretical optimal solution usually consumes an amount of computing memory and needs an unacceptably long time to calculate. It is hard for the artificial intelligence algorithms to meet the engineering requirements of real-time voltage and reactive power control of large-scale power systems.

Since optimal reactive power dispatch of large-scale power systems involves very complicated characteristics, such as having multiple objectives, multiple variables, multiple constraints, discreteness, and real-time responsiveness, the present algorithms cannot balance the opposing demands of optimization and computational efficiency. Therefore, in engineering applications, it comes down to the inevitable choice of voltage and reactive power control to reduce the dimensionality of the high-dimensional ORPD problem. There are two methods to reduce the dimensionality, i.e. (1) hierarchical and zoning space decoupling for ORPD of complicated power systems, and (2) time decoupling for reactive power optimization in each time interval. However, the ORPD problem couples in space and time in nature. On the one hand, the results of volt/VAR control of each regional power network interact because of the electricity connection of through tie lines or transformers. On the other hand, the regulating times of volt/VAR control devices are easily too frequent when the aim of reactive power optimization is to seek the minimum network power loss at each profile independently. Therefore, the modeling research of space decoupling and time decoupling for reactive power optimization should be enhanced.

A problem that has received some attention is the optimal control of reactive sources by minimizing reactive generation while observing voltage and network constraints. Nonlinear programming has been used to solve this problem [24, 25].

1.4 Optimization in Electricity Markets

A market is a basic mechanism to realize the optimal allocation of resources. Thus optimization techniques are naturally adopted in various electricity market research topics. Here we focus only on microeconomics-related topics. A general market model has been introduced by Weber and Overbye [26, 27]. The market primarily includes an independent system operator (ISO), a generation company (GenCo), and a consumer, all of which have their own optimization models. The basic models are introduced as follows.

1.4.1 Strategic Participants' Bids

A strategic participant may be a GenCo or a consumer – in other words, the strategic participant controls the strategic bid of a GenCo or consumer to maximize its own profit. The bids generally take one of the two formats shown in Figure 1.2. Most of the literature on market equilibrium analysis uses the continuous bidding format because of the restriction of solution algorithms. However, the block bidding format is often employed in the real electricity market due to the technical features of the power industry.

Suppose there are I GenCos in the market and each GenCo has a strictly convex quadratic generation cost function as

$$\begin{gathered} C_i(P_{Gi}) = \tfrac{1}{2}a_i P_{Gi}^2 + b_i P_{Gi} \\ (\underline{P}_{Gi} \leq P_{Gi} \leq \overline{P}_{Gi};\ \ a_i > 0;\ \ i = 1, 2, \ldots, I) \end{gathered} \tag{1.13}$$

where P_{Gi} is the generation output; a_i and b_i are coefficients; $\underline{P}_{Gi}$ and $\overline{P}_{Gi}$ are respectively the lower and upper limits of P_{Gi}; and I is the number of GenCos. Hence, its

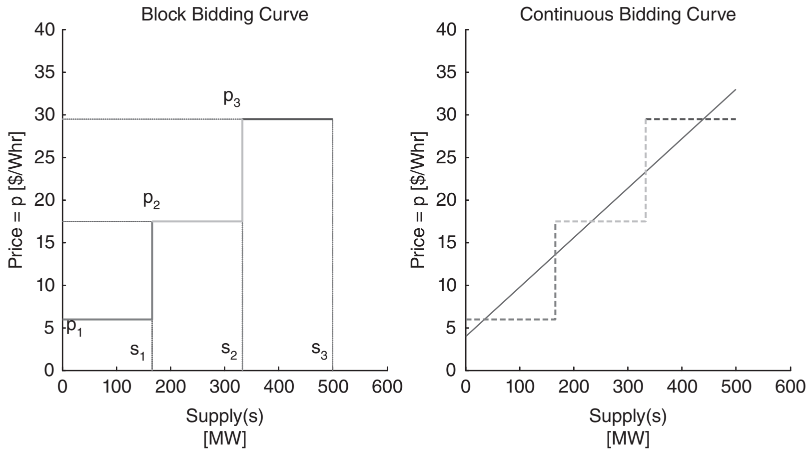

Figure 1.2 Block bidding and continuous bidding curves.

marginal cost can be calculated as

$$\frac{\mathrm{d}C_i(P_{Gi})}{\mathrm{d}P_{Gi}} = a_i P_{Gi} + b_i \tag{1.14}$$

The bids are supposed to be linear functions in the case of the continuous bidding format, so the GenCos are assumed to construct their bids in the form of a linear supply function (LSF) as

$$\widetilde{P}_{Gi} = (p_i - \tilde{b}_i)/\tilde{a}_i \quad (a_i > 0) \tag{1.15}$$

where $\widetilde{P}_{Gi}$ is the submitted quantity, p_i is the bid price for GenCo i, and $\tilde{a}_i$ and $\tilde{b}_i$ are the coefficients (strategic variables) of the LSF. If we suppose that GenCo i chooses $\tilde{a}_i$ and $\tilde{b}_i$, subject to the condition that the two variables have a fixed linear relationship, a strategic parameter k_{Gi} can be used to vary the bid from the true marginal cost function as

$$\widetilde{P}_{Gi} = (p_i/k_{Gi} - b_i)/a_i \quad (k_{Gi} > 0) \tag{1.16}$$

Hence, the strategic parameters of all the GenCos form the GenCos' strategic vector $\mathbf{k}_{Gi}$, in which k_{Gi} is one element.

On the other hand, suppose there are J consumers in the market and the true marginal benefit for each consumer bid is also defined as a linear function as

$$\begin{gathered} P_{Dj} = (p_j - \beta_j)/\alpha_j \\ (\underline{P}_{Dj} \le P_{Dj} \le \overline{P}_{Dj};\ \ \alpha_j > 0;\ \ j = 1, 2, \ldots, J) \end{gathered} \tag{1.17}$$

where P_{Dj} is the quantity demanded; p_j is the locational marginal price (LMP); α_j and β_j are coefficients; $\underline{P}_{Dj}$ and $\overline{P}_{Dj}$ are respectively the lower and upper limits of P_{Dj}; and J is the number of customers.

The benefit can be calculated as

$$B_j(P_{Dj}) = \int p_j \, \mathrm{d}P_{Dj} = 0.5\alpha_j P_{Dj}^2 + \beta_j P_{Dj} \tag{1.18}$$

Also the strategic variable k_{Dj} can similarly be used to construct its bid as follows:

$$\widetilde{P}_{Dj} = (p_j / k_{Dj} - \beta_j)/\alpha_j \quad (k_{Dj} > 0) \tag{1.19}$$

Thus, the consumer's strategic vector $\mathbf{k}_{Di}$ can be composed of all the consumers' strategic variables.

Based on (1.16) and (1.19), the participants determine the strategic vectors and submit the strategic bids to the ISO.

1.4.2 Market Clearing Model

After receiving the bids from all participants, the ISO uses a security-constrained economic dispatch to determine the supplies, the demands, and the LMPs, based on the DC-based optimal power flow. In the market clearing process, the bids are treated as the true costs and benefits for participants, so the optimization problem to maximize a quasi-social welfare, subject to the transmission and generation constraints, can be formulated as follows:

$$\max \ \Gamma(\mathbf{K}_G, \mathbf{K}_D) = \sum_{j=1}^{J} \mathbf{K}_{Dj}\mathbf{B}_j - \sum_{i=1}^{I} \mathbf{K}_{Gi}\mathbf{C}_i \tag{1.20}$$

$$\text{s.t.} \qquad \mathbf{H}\boldsymbol{\theta} = \mathbf{P}_G - \mathbf{P}_D \tag{1.21}$$

$$\underline{F}_l \le F_l \le \overline{F}_l \ (l = 1, 2, \ldots, L) \tag{1.22}$$

$$\underline{P}_{Gi} \le P_{Gi} \le \overline{P}_{Gi} \tag{1.23}$$

$$(i = 1, 2, \ldots, I)$$

where $\mathbf{K}_G$ is the GenCo strategic vector, with $\mathbf{k}_G$ as an element; $\mathbf{K}_D$ is the consumer strategic vector, with $\mathbf{k}_D$ as an element; $\mathbf{B}_j$ is the benefit vector for consumer j, with B_j as an element; $\mathbf{C}_i$ is the cost vector for GenCo i, with C_i as an element; $\mathbf{H}$ is the susceptance matrix; $\boldsymbol{\theta}$ is the vector of bus voltage angles; $\mathbf{P}_G$ is the vector of bus generation outputs; $\mathbf{P}_D$ is the vector of bus demands; F_l is the power flow on line l; $\underline{F}_l$ and $\overline{F}_l$ are the lower and upper flow limits on line l; and L is the number of lines in the system. In a perfectly competitive market, there are no strategic behaviors of GenCos or consumers, and all $\mathbf{k}_G$ and $\mathbf{k}_D$ are equal to 1.

In (1.20), the first equality constraint is the DC power flow equation, the second inequality constraint is the transmission constraint, and the third one is the generation constraint for each GenCo.

For the continuous bidding case, gradient-based approaches such as the interior point method are applicable, and the Lagrange multiplier of the relative power flow constraints in (1.20) are LMPs.

1.4.3 Market Equilibrium Problem

Each rational participant maximizes its profits by choosing and adjusting its strategies based on the market state variation. Therefore, the optimization problem for each participant t (t represents GenCo i or consumer j) could be formulated as the following two-level mathematical program:

$$\max\ R_t(\mathbf{k}_{Gi}, \mathbf{k}_{Dj}) = \sum_{i=1}^{I}(\lambda_{Gi}P'_{Gi} - C_i(P'_{Gi})) + \sum_{j=1}^{J}(B_j(P'_{Dj}) - \lambda_{Dj}P'_{Dj}) \tag{1.24}$$

$$\text{s.t.}\quad \underline{k}_{Gi} \leq k_{Gi} \leq \overline{k}_{Gi} \tag{1.25}$$

$$\underline{k}_{Dj} \leq k_{Dj} \leq \overline{k}_{Dj} \tag{1.26}$$

where λ_{Gi} and λ_{Dj} are LMPs of the GenCo and consumer, respectively, which should be solved through (1.20); P'_{Gi} is the awarded generation for GenCo; P'_{Dj} is the quantity demanded by the consumer; $\underline{k}_{Gi}$ and $\overline{k}_{Gi}$ are the lower and upper limits of the generation strategic variable; and $\underline{k}_{Dj}$ and $\overline{k}_{Dj}$ are the lower and upper limits of the consumer strategic variable.

In a perfectly competitive market, the market equilibrium is called Walrasian equilibrium [28]. However, this is only an ideal case. In reality, the strategic behaviors of market participants should be considered. The participant's profit is dependent not only on its own strategy, but also on the strategies of its opponents. Then this situation should be explained with game theory. The market equilibrium is the point from which each player does not have any incentive to deviate unilaterally (Nash equilibrium), i.e.

$$R_t(\mathbf{k}^*_{Gi}, \mathbf{k}^*_{Dj} \mid \mathbf{k}^*_{-Gi}, \mathbf{k}^*_{-Dj}) \geq R_t(\mathbf{k}'_{Gi}, \mathbf{k}'_{Dj} \mid \mathbf{k}^*_{-Gi}, \mathbf{k}^*_{-Dj}) \tag{1.27}$$

where $\mathbf{k}^*_{Gi}$ and $\mathbf{k}^*_{Dj}$ are strategic vectors of participants in the equilibria; $\mathbf{k}^*_{-Gi}$ and $\mathbf{k}^*_{-Dj}$ are strategic vectors of the opponents of participants in the equilibria; and $\mathbf{k}'_{Gi}$ and $\mathbf{k}'_{Dj}$ are arbitrary strategic vectors of participants.

The left-hand side of (1.27) is the profit of participant t in the equilibria and the right-hand side means the profit in the case that participant t chooses an arbitrary strategic vector with its opponents holding the equilibrium strategies.