

1

Overview of Optimization

Summary

This chapter defines optimization and its basic concepts. It provides examples of various engineering optimization problems.

1.1 Optimization

Engineers are commonly confronted with the tasks of designing and operating systems to meet or surpass specified goals while meeting numerous constraints imposed on the design and operation. Optimization is the organized search for such designs and operating modes. It determines the set of actions or elements that must be implemented to achieve optimized systems. In the simplest case, optimization seeks the maximum or minimum value of an objective function corresponding to variables defined in a feasible range or space. More generally, optimization is the search of the set of variables that produces the best values of one or more objective functions while complying with multiple constraints. A single-objective optimization model embodies several mathematical expressions including an objective function and constraints as follows:

$$\text{Optimize } f(X), \quad X = (x_1, x_2, \dots, x_i, \dots, x_N) \quad (1.1)$$

subject to

$$g_j(X) < b_j, \quad j = 1, 2, \dots, m \quad (1.2)$$

$$x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad i = 1, 2, \dots, N \quad (1.3)$$

in which $f(X)$ = the objective function; X = a set of decision variables x_i that constitutes a possible solution to the optimization problem; x_i = i th decision variable; N = the number of decision variables that determines the dimension

of the optimization problem; $g_j(X)$ = j th constraint; b_j = constant of the j th constraint; m = the total number of constraints; $x_i^{(L)}$ = the lower bound of the i th decision variable; and $x_i^{(U)}$ = the upper bound of the i th decision variable.

1.1.1 Objective Function

The objective function constitutes the goal of an optimization problem. That goal could be maximized or minimized by choosing variables, or decision variables, that satisfy all problem constraints. The desirability of a set of variables as a possible solution to an optimization problem is measured by the value of objective function corresponding to a set of variables.

Some of the algorithms reviewed in this book are explained with optimization problems that involve maximizing the objective function. Others do so with optimization problems that minimize the objective function. It is useful to keep in mind that a maximization (or minimization) problem can be readily converted, if desired, to a minimization (or maximization) problem by multiplying its objective function by -1 .

1.1.2 Decision Variables

The decision variables determine the value of the objective function. In each optimization problem we search for the decision variables that yield the best value of the objective function or optimum.

In some optimization problems the decision variables range between an upper bound and a lower bound. This type of decision variables forms a continuous decision space. For example, choosing adequate proportions of different substances to make a mixture of them involves variables that are part of a continuous decision space in which the proportions can take any value in the range $[0,1]$. On the other hand, there are optimization problems in which the decision variables are discrete. Discrete decision variables refer to variables that take specific values between an upper bound and a lower bound. Integer values are examples of discrete values. For instance, the number of groundwater wells in a groundwater withdrawal problem must be an integer number. Binary variables are of the discrete type also. The typical case is that when taken the value 1 implies choosing one type of action, while taking the value 0 implies that no action is taken. For example, a decision variable equal to 1 could mean building a water treatment plant at a site, while its value equal to 0 means that the plant would not be constructed at that site. Optimization problems involving continuous decision variables are called continuous problems, and those problems defined in terms of discrete decision variables are known as discrete problems. There are, furthermore, optimization problems that may involve discrete and continuous variables. One such example would be an optimization involving the decision of whether or not to build a facility at a certain location and, if so, what its capacity ought to be. The siting variable is of the

binary type (0 or 1), whereas its capacity is a real, continuous variable. This type of optimization problem is said to be of mixed type.

1.1.3 Solutions of an Optimization Problem

Each objective function is expressed in terms of decision variables. When there is only one decision variable, the optimization problem is said to be one-dimensional, while optimization problems with two or more decision variables are called N -dimensional. An N -dimensional optimization problem has solutions that are expressed in terms of one or more sets of solutions in which each solution has N decision variables.

1.1.4 Decision Space

The set of decision variables that satisfy the constraints of an optimization problem is called the feasible decision space. In an N -dimensional problem, each possible solution is an N -vector variable with N elements. Each element of this vector is a decision variable. Optimization algorithms search for a point (i.e., a vector of decision variables) or points (i.e., more than one vector of decision variables) in the decision space that optimizes the objective function.

1.1.5 Constraints or Restrictions

Each optimization problem may have two types of constraints. Some constraints directly restrict the possible value of the decision variables, such as a decision variable x being a positive real number, $x > 0$, or analogous to Equation (1.3). Another form of constraint is written in terms of formulas, such as when two decision variables x_1 and x_2 are restricted to the space $x_1 + x_2 \leq b$ or analogous to Equation (1.2). The goal of an optimization problem is to find an optimal feasible solution that satisfies all the constraints and yields the best value of the objective function among all feasible solutions. Figure 1.1 depicts a constrained two-dimensional decision space with infeasible and feasible spaces.

The set of all feasible solutions constitute the feasible decision space, and the infeasible decision space is made up of all the infeasible decision variables. Evidently, the optimal solution must be in the feasible space.

1.1.6 State Variables

State variables are dependent variables whose values change as the decision variables change their values. State variables are important in engineering problems because they describe the system being modeled and the objective function and constraints are evaluated employing their values. As an example, consider an optimization problem whose objective is to maximize hydropower generation by operating a reservoir. The decision variable is the amount of daily water release passing through turbines. The state variable is the amount

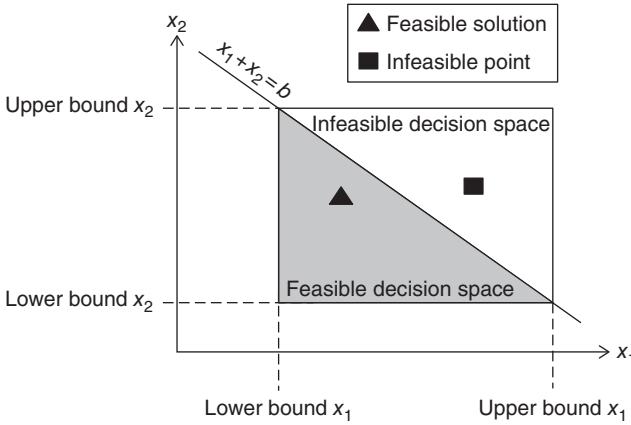


Figure 1.1 Decision space of a constrained two-dimensional optimization problem.

of water stored in the reservoir, which is affected by the water released through turbines according to an equation of water balance that also involves water inflow to the reservoir, evaporation from the reservoir, water diversions or imports to the reservoir, water released from the reservoir bypassing turbines, and other water fluxes that change the amount of reservoir storage.

1.1.7 Local and Global Optima

It has been established that a well-defined optimization problem has a well-defined decision space. Each point of the decision space defines a value of the objective function. A local optimum refers to a solution that has the best objective function in its neighborhood. In a one-dimensional optimization problem, a feasible decision variable X^* is a local optimum of a maximization problem if the following condition holds:

$$f(X^*) \geq f(X), \quad X^* - \varepsilon \leq X \leq X^* + \varepsilon \quad (1.4)$$

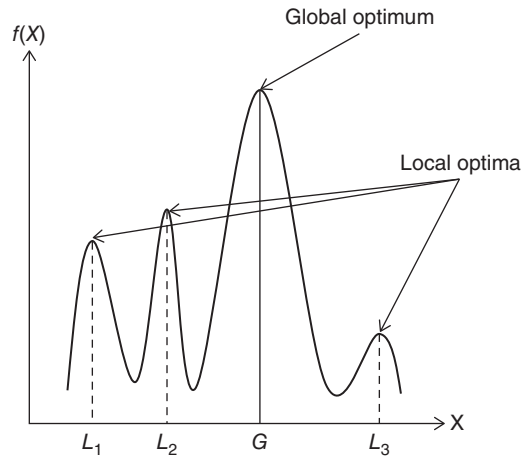
In a minimization problem the local-optimum condition becomes

$$f(X^*) \leq f(X), \quad X^* - \varepsilon \leq X \leq X^* + \varepsilon \quad (1.5)$$

where X^* = a local optimum and ε = limited length in the neighborhood about the local optimum X^* . A local optimum is limited to a neighborhood of the decision space, and it might not be the best solution over the entire decision space.

A global optimum is the best solution in the decision space. Some optimization problems may have more than one—in fact, an infinite number of global optima. These situations arise commonly in linear programming problems. In this case, all the global optima produce the same value of the objective function. There are not decision variables that produce a better objective function

Figure 1.2 Schematic of global and local optimums in a one-dimensional maximizing optimization problem.



value than the global optimum. A one-dimensional optimization problem with decision variable X and objective function $f(X)$ the value X^* is a global optimum of a maximization problem if for any decision variable X the following is true:

$$f(X^*) \geq f(X) \quad (1.6)$$

In a minimization problem we would have

$$f(X^*) \leq f(X) \quad (1.7)$$

Figure 1.2 illustrates global and local optima for a one-dimensional maximization problem.

L_1 , L_2 , and L_3 in Figure 1.2 are local optima, and G denotes the global optimum with the largest value of the objective function. The decision space may be single modal or multimodal. In a single-modal surface, there is only one extreme point, while there are several extremes on the surface of a multimodal problem. In a single-modal problem, there is a single local optimum that is also the global optimum. On the other hand, a multimodal problem may include several local and global optima. However, the decision variables that produce a global optimum must all produce the same value of the global optimum, by definition. Figure 1.3 illustrates the surface of one-dimensional optimization problems with single-modal and multimodal decision spaces in which there is one single optimum.

1.1.8 Near-Optimal Solutions

A near optimum has a very close but inferior value to the global optimum. In some engineering problems, achieving the absolute global optimum is extremely difficult or sometimes impossible because of the innate complexity of the problem or the method employed to solve the problem. Or achieving the

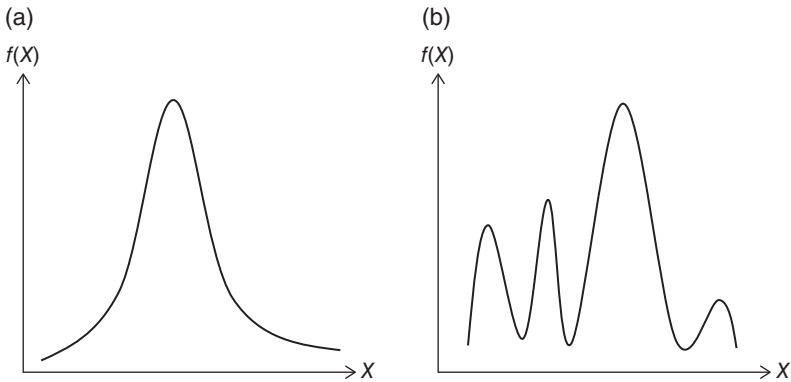


Figure 1.3 Different types of decision spaces: (a) maximization problem with single-modal surface and one global optimum; (b) maximization problem with multimodal surface that has one global optimum.

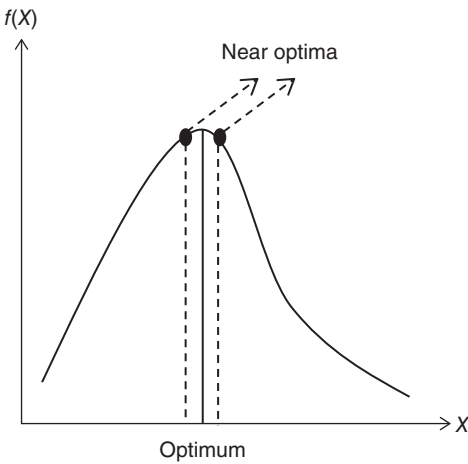


Figure 1.4 Demonstration of near optima in a one-dimensional maximizing optimization problem.

global optimum may be computationally prohibitive. In this situation, a near optimum is calculated and reported as an approximation to the global optimum. Near optima are satisfactory in solving many real-world problems. The proximity of a near optimum to the global optimum depends on the optimization problem being solved and the judgment of the analyst. Figure 1.4 depicts the concept of a near optimum in a maximization problem.

1.1.9 Simulation

Each decision variable of an optimization problem defines an objective function value. The process of evaluating the state variables, which are necessary for

estimation of the objective function, and constraints with any decision variable is known as simulation. A simulation model receives the decision variables as inputs and simulates the system's state variables. Sometimes the simulation model consists of one or more simple mathematical functions and equations. However, most real-world and engineering problems require simulation models with complex procedures that most solve systems of equations and various formulas that approximate physical processes. Simulation is, therefore, the computational imitation of the operation of a real-world process or system over time.

1.2 Examples of the Formulation of Various Engineering Optimization Problems

This section presents examples of the formulation of different types of engineering optimization problems including mechanical design, structural design, electrical engineering optimization, water resources optimization, and calibration of hydrological models.

1.2.1 Mechanical Design

Designing a compound gear train is exemplary of optimal designing. A compound gear train is designed to achieve a particular gear ratio between the driver and driven shafts. The purpose of the gear train design is finding the number of teeth in each gear so that the error between the obtained and required gear ratios is minimized. In practice, the term gear ratio is used interchangeably with velocity ratio. It is defined as the ratio of the angular velocity of the output shaft to that of the input shaft. For a pair of matching gears, the velocity or gear ratio α is calculated as follows:

$$\alpha = \left| \frac{\omega_{out}}{\omega_{in}} \right| = \frac{\theta_{in}}{\theta_{out}} \quad (1.8)$$

in which α = gear ratio; ω_{in} = angular velocity of the input shaft; ω_{out} = angular velocity of the output shaft; θ_{in} = the number of teeth on the input gear; and θ_{out} = the number of teeth on the output gear. The ratio is, thus, inversely proportional to the number of teeth on the input and output gears.

Figure 1.5 shows a compound gear train that is made of four gears. It is desired to produce a gear ratio as close as possible to a required value μ . The objective of the design is to find the number of teeth in each gear so that the error between the obtained and required gear ratios is minimized. Normally, additional considerations such as the number of gear pairs to use and the geometric arrangement of the shafts must be considered in addition to wear. To simplify the problem only the particular configuration shown in Figure 1.5 is considered here.

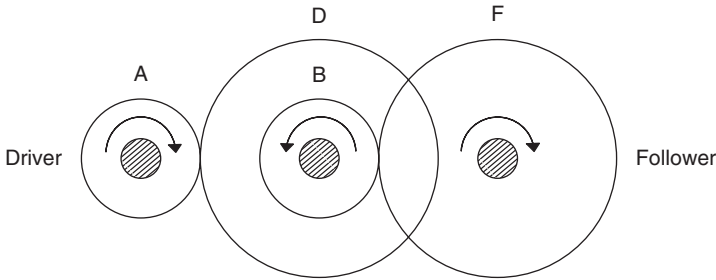


Figure 1.5 Compound gear train made of four gears.

For the system shown in Figure 1.5, the gear ratio is evaluated as follows:

$$\alpha = \frac{\tau_d}{\tau_a} \times \frac{\tau_b}{\tau_f} \quad (1.9)$$

in which τ_d , τ_a , τ_b , and τ_f = the number of teeth on gears D, A, B, and F, respectively.

The number of teeth on each gear constitutes the decision variables:

$$X = (x_1, x_2, x_3, x_4) = (\tau_d, \tau_a, \tau_b, \tau_f) \quad (1.10)$$

Minimizing the square of the difference between the desired gear ratio (μ) and the actual design gear ratio (α) the optimization problem leads to the following optimization problem:

$$\text{Minimize } f(X) = (\mu - \alpha)^2 \quad (1.11)$$

in which

$$\alpha = \frac{x_1}{x_2} \times \frac{x_3}{x_4} \quad (1.12)$$

subject to

$$x^{(L)} \leq x_i \leq x^{(U)}, \quad i = 1, 2, 3, 4 \quad (1.13)$$

where μ = required gear ratio; α = actual gear ratio; $x^{(L)}$ and $x^{(U)}$ = minimum and maximum number of teeth on each gear, respectively. The minimization of the objective function (Equation (1.11)) is with respect to x_1 , x_2 , x_3 , and x_4 . The objective function is nonlinear, and the constraints (Equation (1.13)) are simple bounds on the decision variables. Since the number of teeth is an integer number, this problem has a discrete domain, and the decision variables must take integers values.

1.2.2 Structural Design

Structural optimization problems are created and solved to determine the configurations of structures that satisfy specifications and produce an optimum for a chosen objective function. The main purpose of structural optimization is to minimize the weight of a structure or the vertical deflection of a loaded member. Here, a two-bar truss design model is considered for illustration purposes.

The truss shown in Figure 1.6 is designed to carry a certain load without elastic failure. In addition, the truss is subject to limitations in geometry, area, and stress.

The stresses on nodes A and B are calculated as follows:

$$\sigma_{AC} = \frac{\text{Force} \times \left(\frac{L-L'}{L} \right) \times \sqrt{H^2 + L^2}}{a_1 \times H} \quad (1.14)$$

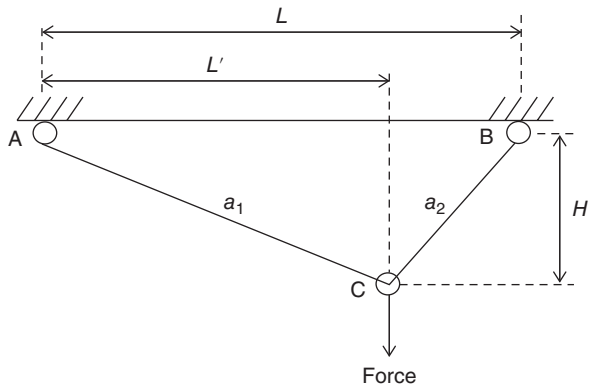
$$\sigma_{BC} = \frac{\text{Force} \times \left(\frac{L'}{L} \right) \times \sqrt{H^2 + (L-L')^2}}{a_2 \times H} \quad (1.15)$$

in which σ_{AC} and σ_{BC} = the stress on node A and B, respectively (N/m^2); *Force* = force on node C (N); *H* = perpendicular distance from AB to point C (m); *L* = length of AB (m); *L'* = length of AC (m); a_1 = cross-sectional area of AC; and a_2 = cross-sectional area of BC (m^2).

In this case, a_1 , a_2 , and *H* are the decision variables of the optimization model:

$$X = (x_1, x_2, x_3) = (a_1, a_2, H) \quad (1.16)$$

Figure 1.6 Schematic of a two-bar truss.



The optimization problem is expressed as follows when the weight of the structure is minimized:

$$\text{Minimize } f(X) = \rho \times \left(\left(a_1 \times \sqrt{H^2 + L^2} \right) + \left(a_2 \times \sqrt{H^2 + (L - L')^2} \right) \right) \quad (1.17)$$

subject to

$$\sigma_{AC}, \sigma_{BC} \leq \sigma_{max} \quad (1.18)$$

$$x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad i = 1, 2, 3 \quad (1.19)$$

in which ρ = the volumetric density of the truss; σ_{max} = the maximum allowable stress; $x_i^{(L)}$ and $x_i^{(U)}$ = the minimum and maximum values of the decision variables, respectively. The minimization of Equation (1.17) is with respect to a_1 , a_2 , and H , which are real-valued variables. The objective function is nonlinear, and so are the constraints.

1.2.3 Electrical Engineering Optimization

Directional overcurrent relays (DOCRs), which protect transmission systems, constitute a classical electrical engineering design problem. DOCRs are part of electrical power systems that isolate faulty lines in the event of failures in the system. DOCRs are logical elements that issue a trip signal to the circuit breaker if a failure occurs within the relay jurisdiction and are placed at both ends of each transmission line. Their coordination is an important aspect of system protection design. The relay coordination problem is to determine the sequence of relay operations for each possible failure location so that the failure section is isolated with sufficient coordination margins and without excessive time delays. The selection of the sequence of relay operations is a function of the power network topology, relay characteristics, and protection philosophy. The DOCR protection scheme consists of two types of settings, namely, current, referred to as plug setting (PS), and time dial setting (TDS), which must be calculated. With the optimization of these settings, an efficient coordination of relays can be achieved, and the faulty transmission line may be isolated, thereby maintaining a continuity of power supply to functional sections of power systems.

The operating time (T) of a DOCR is a nonlinear function of the relay settings including time dial settings (TDS), plug settings (PS), and the fault current (I) seen by the relay. The relay operating time equation for a DOCR is estimated as follows:

$$T = \frac{K_1 \times \xi}{\left(\frac{\gamma}{\psi \times C_{pri-rating}} \right)^{K_2} - K_3} \quad (1.20)$$

in which T = operating time; $K_1, K_2,$ and K_3 = constants that depend upon the specific device being simulated; ξ = time dial settings; ψ = plug settings; γ = faulty current passing through the relay, which is a known value, as it is a system-dependent parameter and continuously measured by monitoring instruments; and $C_{pri-rating}$ = a parameter whose value depends on the number of turns in the current transformer (CT). CT is used to reduce the level of the current so that the relay can withstand it. One “current transformer” is used with each relay, and, thus, $C_{pri-rating}$ is known in the problem.

The TDS and PS of the relays are the decision variables of the optimization model:

$$X = (x_1, x_2, \dots, x_N, x_{N+1}, x_{N+2}, \dots, x_{2N}) = (\xi_1, \xi_2, \dots, \xi_N, \psi_1, \psi_2, \dots, \psi_N) \tag{1.21}$$

where N = number of relays of the system.

The optimization problem is formulated as follows:

$$\text{Minimize } f(X) = \sum_{i=1}^N \sum_{j=1}^M T_{i,j}^{(primary)} \tag{1.22}$$

subject to

$$T^{(backup)} - T^{(primary)} \geq C_T, \quad \text{for all relay pairs} \tag{1.23}$$

$$x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad i = 1, 2, \dots, 2N \tag{1.24}$$

in which M = number of failures; $T_{i,j}^{(primary)}$ = operating time of the primary relay i for a failure j ; $T^{(backup)}$ = operating time of the backup relay; C_T = coordinating time interval; and $x_i^{(L)}$ and $x_i^{(U)}$ = bounds on relay settings. The objective function (Equation (1.22)) is nonlinear, and so are the constraints.

1.2.4 Water Resources Optimization

Flowing water generates energy that can be managed and turned into electricity. This is known as hydroelectric power or hydropower. Dams or reservoirs are the most common type of hydroelectric power plant. Some hydropower dams have several functions such as supplying urban and agriculture water and flood control. This example focuses on hydropower generation exclusively. Figure 1.7 shows a simple schematic of a hydropower dam with its associated water fluxes.

The reservoir storage (S) in each operational time period is calculated as follows:

$$S_{t+1} = I_t + S_t - R_t - Sp_t, \quad t = 1, 2, \dots, N \tag{1.25}$$

$$Sp_t = \begin{cases} 0 & \text{if } I_t + S_t - R_t \leq S_{max} \\ I_t + S_t - R_t - S_{max} & \text{if } I_t + S_t - R_t > S_{max} \end{cases} \tag{1.26}$$

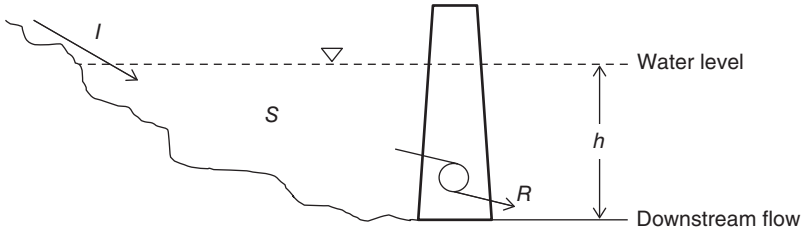


Figure 1.7 Schematic of a hydropower dam.

in which S_{t+1} = the reservoir storage at the end of time period t ; I_t = the volume of input water during time period t ; S_t = the reservoir storage at the start of time period t ; R_t = the volume of release of the reservoir; S_{max} = the reservoir capacity; Sp_t = the volume of overflow that occurs whenever the reservoir storage exceeds the reservoir capacity; and N = the total number of time periods. There is spill or overflow whenever the reservoir storage exceeds the capacity of the reservoir. The storage volume cannot be less than a value like S_{min} given that the floodgate is usually placed higher than the bottom of the dam. This part of the reservoir is usually filled with sediments.

The generated power is a function of water flow and the elevation difference between the hydraulic head at the intake and outlet of the turbine. The generated power in period t is determined as follows:

$$P_t = \eta \times \rho \times g \times h_t \times q_t \quad (1.27)$$

$$q_t = \frac{R_t}{\Delta t \times n_t} \quad (1.28)$$

in which P_t = generated power (W) in period t ; η = efficiency of powerhouse; ρ = density (kg/m^3) ($\sim 1000 \text{ kg}/\text{m}^3$ for water); g = acceleration of gravity ($9.81 \text{ m}/\text{s}^2$); h_t = falling height or effective hydraulic head (m) in time period t ; q_t = water flow in period t (m^3/s); R_t = total volume of release of reservoir in time period t (m^3); n_t = ratio of time when the powerhouse is active; and Δt = length of time period (s). The water level in the reservoir is usually estimated based on the area or volume of water stored in the reservoir by predefined equations.

The volume of water that enters the reservoir in each time period (I_t) is known. The goal of the problem is determining the volume of release in each time period (R_t) so that the total generated power is close to the power plant capacity (PPC) as much as possible. The volume of water release from the reservoir in each time period is the decision variable of the optimization model:

$$X = (x_1, x_2, \dots, x_N) = (R_1, R_2, \dots, R_N) \quad (1.29)$$

where N = total number of decision variables that is equal to the total number of time periods.

The optimization problem minimizes the normalized sum of squared deviations between generated power and power plant capacity and is written as follows:

$$\text{Minimize } f(X) = \sum_{i=1}^N \left(1 - \frac{P_i}{PPC} \right)^2 \quad (1.30)$$

$$P_i = \eta \times \rho \times g \times h_i \times \frac{x_i}{\Delta t \times n_i}, \quad i = 1, 2, \dots, N \quad (1.31)$$

subject to

$$S_{min} \leq S_i \leq S_{max} \quad (1.32)$$

$$0 \leq P_i \leq PPC \quad (1.33)$$

$$x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad i = 1, 2, \dots, N \quad (1.34)$$

in which PPC = the power plant capacity; S_{min} = the minimum storage; and $x_i^{(L)}$ and $x_i^{(U)}$ = the minimum and maximum volume of release in each period i , respectively. The minimum release is governed by different factors such as environmental flows to sustain aquatic life. One of the factors that restrict the maximum release from the reservoir is the capacity of the floodgate.

1.2.5 Calibration of Hydrologic Models

The parameter calibration of hydrologic model is commonly posed as the minimization of a norm of errors between observed and predicted hydrologic values. The routing of floods in river channels is a classic example involving the calibration of hydrologic models. Flood is a natural phenomenon that can cause considerable damage in urban, industrial, and agricultural regions. To prevent those damages it is necessary to implement a hydrologic model to estimate the flood hydrograph at the downstream river reach given the upstream hydrograph. The Muskingum model is a hydrologic model based on the continuity and parameterized storage equations as follows:

Continuity:

$$\frac{dS_t}{dt} = I_t - O_t \quad (1.35)$$

Parameterized storage:

$$S_t = \beta_1 [Y \times I_t + (1 - Y) O_t]^{\beta_2} \quad (1.36)$$

in which S_t , I_t , and O_t = storage, inflow, and outflow in a river reach at time t , respectively; β_1 = storage time constant for a river reach that has a value reasonably close to the travel time of a flood through the river reach; β_2 = exponent for the effect of nonlinearity between accumulated storage and weighted flow;

and Y = weighting factor between 0 and 0.5 for reservoir storage and between 0 and 0.3 for stream channels.

Equation (1.36) is solved for the reach outflow as follows:

$$O_t = \left(\frac{1}{1-Y} \right) \left(\frac{S_t}{\beta_1} \right)^{\frac{1}{\beta_2}} - \left(\frac{Y}{1-Y} \right) I_t \quad (1.37)$$

By combining Equations (1.35) and (1.37), the state equation becomes

$$\frac{\Delta S_t}{\Delta t} = - \left(\frac{1}{1-Y} \right) \left(\frac{S_t}{\beta_1} \right)^{\frac{1}{\beta_2}} + \left(\frac{1}{1-Y} \right) I_t \quad (1.38)$$

$$S_{t+1} = S_t + \Delta S_t \quad (1.39)$$

The routing of a flood hydrograph consists of the following steps:

Step 1: Assume values for the parameters β_1 , β_2 , and Y .

Step 2: Calculate the storage (S_t) with Equation (1.36), with the initial outflow equal to the initial inflow.

Step 3: Calculate the time rate of change of storage volume with Equation (1.38).

Step 4: Estimate the next storage with Equation (1.39).

Step 5: Calculate the next outflow with Equation (1.37).

Step 6: Repeat steps 2–5 for total time steps in the flood routing until reaching a stopping criterion.

The goal of this problem is estimating β_1 , β_2 , and Y so that the sum of the squared differences between observed and predicted outflows is minimized. The parameters β_1 , β_2 , and Y of the Muskingum model are the decision variables:

$$X = (x_1, x_2, x_3) = (\beta_1, \beta_2, Y) \quad (1.40)$$

The optimization problem is formulated as follows:

$$\text{Minimize } SSQ = \sum_{t=1}^M (O_t - \hat{O}_t)^2 \quad (1.41)$$

subject to

$$x_i^{(L)} \leq x_i \leq x_i^{(U)}, \quad i = 1, 2, 3 \quad (1.42)$$

where M = total number of time steps in the flood routing; O_t and \hat{O}_t = observed and routed outflow, respectively, at time t ; and $x_i^{(L)}$ and $x_i^{(U)}$ = minimum and maximum values of parameters of Muskingum model, respectively.

1.3 Conclusion

This chapter introduced foundational concepts of optimization such as the objective function, decision variables, decision space, and constraints. In addition, several examples of formulating engineering optimization problems were presented to illustrate a variety of optimization models.

