# 1

# Optimization

Introduction and Concepts

## 1.1 Optimization and Terminology



Optimization is a procedure for seeking the best choices, and there are essential elements to the procedure. First, there must be criteria for rating choices. This can be alternately stated as a method for determining the value of a metric of goodness, the assessment value of how a person balances desirables and undesirables. One must have this assessable or quantifiable metric of desirability to be able to choose the best. Also, there must be a relation between choices and the evaluation of the outcome. This relation can be either a model or an experiment. The procedure (equation, algorithm, experimental method) for determining the value of the desirability metric is called the objective function (OF). The choices that you can change to improve the OF value are called the decision variables (DVs).

You might want to minimize (costs, expenses, risk, etc.) or maximize (profit, reliability, probability of success, etc.). Optimization will seek DV values that lead to the optimum, either minimum or maximum.

A trial solution (TS) is a particular DV choice. It might not be the optimum. The optimum is denoted as DV$^*$. For simple, idealized applications, one can determine the exact value of DV$^*$; but, for most applications, this is an ideal concept, like the value of $\pi$. You'll get close enough to the true value for a particular need and call the close enough TS the DV$^*$.

There are usually several opposing concepts in the comprehensive statement of an OF. For example, one may wish to minimize cost of food, which leads to the obvious solution—don't consume any. Acknowledging this extreme reveals an opposing concept that needs to be considered as a comprehensive assessment of desirability, which is acquiring adequate nourishment. Further consideration of the situation may also lead one to acknowledge that joy of eating is important and including occasional treats is part of a healthy life. So, the opposing ideals might be to balance cost with nourishment and joy.

Once the equation is obtained, using optimization to determine the optimum is fairly easy. The difficulty in optimization is usually related to obtaining the equation that provides a comprehensive and complete statement of the context.

As another example of opposing ideals, consider how much insulation should be placed in a house. One ideal would be to minimize cost. Calculating the cost of insulation as price per volume times area times thickness results in the model $C = pAt$. If the objective is to determine the thickness that minimizes installed cost, capital, the obvious solution is $t = 0$. Don't install insulation. This extreme, however, might reveal that zero insulation leads to energy losses that are also important. Annual value of energy lost, an expense, might be simply modeled as $E = eA(T_{\text{house}} - T_{\text{outside}})/(r + kt)$, where $k$ is the specific factor for the insulation, $r$ is the insulation capacity of the walls, and $e$ is the unit cost of energy.

*Engineering Optimization: Applications, Methods, and Analysis*, First Edition. R. Russell Rhinehart.

Companion website: www.wiley.com/go/rhinehart/engineeringoptimization

Looking at an $N$-year horizon, the total cost is the initial capital and the $N$ annual expenses. $J = C + NE = pAt + NeA(T_{\text{house}} - T_{\text{outside}})/(r + kt)$. Now the opposing trends are visible. Increasing $t$ increases the first term but decreases the second.

Optimization is a structured, rational search through the DV "space" from an initial TS toward DV*. Although for simple applications the explicit solution to a simple equation can provide the DV* answer, optimization is usually not a magical single leap to the optimum, but an iterative progression. The locus of trial solutions through DV space is termed the path. Optimization desires to find DV* with minimal computational or experimental burden (minimum number of TS evaluations along the path), minimum *a priori* knowledge or human involvement, and maximum assurance that the global optimum has been found and that DV values are close enough to DV*.

Constraints impact both the DV* value and the path to get there. Some constraints are mathematical (such as divide by zero, or square root of a negative number), some are computational (overflow, or subscript out of range), some are physical (composition cannot be >100% or <0%, or keep temperature above freezing point), some are environmental health safety and loss prevention related (keep composition less than lower explosive limit, or keep pressure above vacuum, or don't let tank overflow), some are legal/political (don't violate contract specifications, or stay beyond the 200-mile offshore limit), etc. Some constraints are "hard," meaning there is no tolerance, and absolutely do not violate the constraint. Some are "soft," wherein small violations are permitted, but with a penalty. Hard constraints complicate the TS search sequence. Soft constraints do not limit the TS values, but the user needs to assess the magnitude of the penalty relative to all OF issues.

The canonical statement of an optimization application is

$$\begin{aligned} &\min_{\{\mathbf{DV}\}} J = \text{OF}(\mathbf{DV}) \\ &\text{S.T.}: \quad g(\mathbf{DV}) < 0 \end{aligned} \tag{1.1}$$

If the OF is an equation, for example $y = 3 - x + x^2$, and the DV is the variable $x$, and there are no constraints, the statement becomes

$$\min_{\{x\}} J = 3 - x + x^2 \tag{1.2}$$

For this, an analytical solution is simple. Set $dJ/dx|_{x^*} = 0$, which is $0 = 0 - 1 + 2x^*$, and solve. Here, the optimum value of $x$, $x^* = 1/2$, and with that value of $x^*$, OF* = 2.75.

For applications, however, the OF is usually not a simple equation, but a complicated procedure (often including loops, look-up tables, and internal root finding), which would incorporate design and economics and other sustainability metrics. Although simple examples like the one aforementioned are instructive (and intellectually stimulating as they permit closed-form mathematical analysis), the practical applications are more complicated. This book is about the practical applications, although often it will use simple relations to facilitate reader understanding.

## 1.2 Optimization Concepts and Definitions

We usually optimize for the enterprise. This may be the employing organization, a volunteer team, or humankind; and each is an assembly of humans with diverse views on the desirables. Often, the person doing the optimization has a limited view of the context and the perspective that other customers and stakeholders have. Accordingly, clarity of terminology is essential to ensure that the person doing the optimization is making choices and assumptions that are compatible to the enterprise needs. Here is some essential terminology:

*Objective*: This is the thing or issue that you wish to maximize or minimize—the indicator of goodness or desirability. Usually there are multiple objectives. Perhaps profitability is an objective, but other associated measures of desirability would include resource conservation, safety, sustainability, compliance to contracts, etc. Perhaps health, lifelong joy, bang for the buck, risk, and effort are the objectives. Although those concepts are understood, profitability, health, safety, risk, joy, and effort are vague statements. There are many ways to interpret each of these abstract concepts. The user must provide specifics. Further, it is essential that the objective is comprehensive. If it only represents a portion of the issues that stakeholders will use, the solution will not be quite right.

*Objective function (OF, J)*: This is the procedure to measure desirability (goodness) or undesirability (badness) associated with the thing or issue that you wish to maximize or minimize. If profitability is the objective, then perhaps DCFRR (discounted cash flow rate of return), or PBT (payback time), or NPV (net present value), or LTROA (long-term return on assets) may be the metric used to assess profitability. Each of these balances profit and expenses with investment. Why is there not one measure of profitability? Because, for any one objective there are many possible ways to measure desirability, and depending on the enterprise context, one profitability index may be more appropriate than another. The OF (objective function) is often called a "cost function" representing an understanding that minimizing resource consumption is the only objective, but I prefer to not use that simplistic terminology. Classically, the symbol for the OF is the capital letter $J$. The OF must be concrete, it must provide a value, and it must be quantifiable.

*Objective function value (OF, J)*: This is the value of the objective function, this is what optimization seeks to make as large as possible or as small as possible. OF value examples could include the following: LTROA is 23.7%, or PBT is 1.7 years, or fuel consumption is 42.1 miles/gal.

*Decision variables (DV)*: These are the choices you can make, the values or decisions you can choose to obtain the optimum $J$, or the things that you can change. The DVs are the independent variables that can be changed, and the OF is the response. Some values you cannot change, or so it first seems. However, a second look at the application often will reveal alternate variables that you can change. You need to differentiate between the two and properly identify the degrees of freedom within an application context.

*Optimum values ($DV^*$, $OF^*$)*: These are the DV values that either maximize or minimize the OF. The optimum values are represented as $DV^*$ and $OF^*$.

*Trial solution (TS)*: This is a possible value of the DVs. It might not be the optimum value. It is often the next guess as the DV values are being changed to improve the OF.

*Model*: This is the relation that permits calculating the OF value from a TS of the DV values. The DV may or may not be explicitly shown in the OF. The model is usually not a simple equation. More so it is a procedure. Sometimes experimental results will be used to determine OF values rather than a model.

*Constraints*: These are limits on the DV values or on other associated variables. For example insulation thickness must be nonnegative, $t \geq 0$. However, constraints can be more complicated, for example, structurally, the ceiling might only be able to hold a certain weight, so $wAt <$ limit. Constraints can be on a rate of change, on a future value, on an executable operation, or on any number of aspects of variables. Consider wanting to minimize the number of trips to carry 5 gal of water across the street. The answer is one trip. But if the bucket has a 3 gal capacity, it imposes a constraint on the amount that can be carried in one trip, and the solution is two trips. Constraints can be as influential on the answer as the objective function.

*Method*: This is the procedure used to find DV* values. How will you move from an initial TS toward DV*? How will you specify an initial TS? The answers to such questions are part of the method of optimization.

*Convergence criterion*: Optimization is an iterative sequential procedure to progressively move the TS toward DV*. Probably, it never gets exactly to DV*. When it is close enough, stop the search and claim that DV* has been found. The criterion to determine that the optimizer has found a TS close enough to the optimum to stop the search and call it DV* is termed the convergence criterion.

## 1.3 Examples

Here are four diverse application examples that reveal the meaning of the several definitions given so far.

**Example 1** A one-line function or equation

$$y = 8x - 2x^2 + 4 \tag{1.3}$$

The objective is to find the value of $x$ that maximizes the value of $y$. This can be graphically illustrated in Figure 1.1.

$y$ = OF (often called dependent response)
$x$ = DV (often called independent input)

Using the analytical procedure of setting the derivative of the function to zero and then solving for the DV* value, we easily obtain

$$x_{\text{optimum}} = 2 = x^* \tag{1.4}$$

$$y(x_{\text{optimum}}) = 12 = y^* \tag{1.5}$$

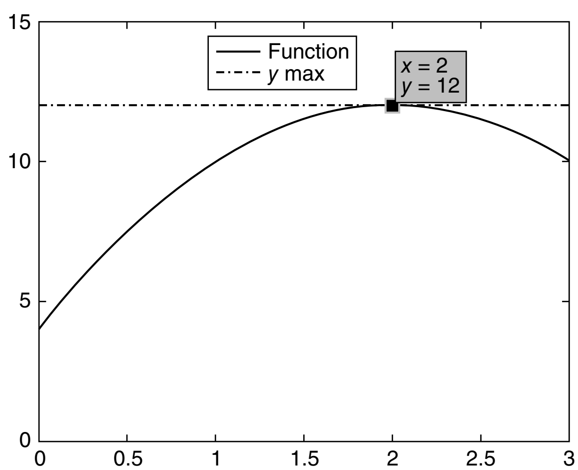



**Figure 1.1** Illustrating a function with a maximum.

However, most important applications are not a simple one-line equation that permits an analytical derivative. Although it is an instructive example, the method is limited to ideal applications.

**Example 2 Economic Optimization**

Find the optimal thickness of insulation in an attic. The initial capital cost is the cost of insulation as price per volume times area times thickness, $C = pAt$, and thickness must be nonnegative, so there is a constraint $t \geq 0$. There may be alternate constraints related to total weight of insulation that the ceiling can support, $\rho At < W$. The annual value of energy lost, an expense, might be simply modeled as $E = eA(T_{\text{house}} - T_{\text{outside}})/(r + kt)$, where $k$ is the specific factor for the insulation, $r$ is the insulation capacity of the walls, and $e$ is the unit cost of energy. Ignoring the time value of money and looking at an $N$-year horizon, the total cost is the initial capital and the $N$ annual expenses. $J = C + NE = pAt + NeA(T_{\text{house}} - T_{\text{outside}})/(r + kt)$. The optimization statement then becomes

$$\begin{aligned} &\min_{\{t\}} J = pAt + \frac{NeA(T_{\text{house}} - T_{\text{outside}})}{r + kt} \\ &\text{S.T.}: \begin{array}{l} t \geq 0 \\ \rho At < W \end{array} \end{aligned} \tag{1.6}$$

Analytically, this admits a solution. Set $dJ/dt = 0$, and solve for insulation thickness.

$$t^* = \frac{1}{k}\left[\sqrt{\frac{p}{kNe(T_{\text{house}} - T_{\text{outside}})}} - r\right] \tag{1.7}$$

One needs to check that the value of $t^*$ is within constraints.

The role of the coefficients in the equation could be questioned. Are they "givens" or are they DVs? The parameters associated with the insulation are $p$ and $k$, but an alternate insulation material choice could permit alternate values. Further, the value for $r$ is related to other construction choices such as ceiling material or roofing material. Is the designer free to change these or not? Finally, the equations reveal that the value of energy lost is related to the house temperature setting. Is an option for minimizing costs to change the house temperature setting? If yes, then perhaps there needs to be a comfort penalty for deviations from the nominally ideal 72°F (22°C). Characteristically, the penalty scales with the square of the deviation from desired, but there needs to be a weighting factor that makes the temperature deviation equivalent to the cost. With such, the optimization statement has evolved to

$$\begin{aligned} &\min_{\{t,\, T_{\text{house}}\}} J = pAt + \frac{NeA(T_{\text{house}} - T_{\text{outside}})}{r + kt} + \lambda(T_{\text{house}} - 72)^2 \\ &\text{S.T.}: \begin{array}{l} t \geq 0 \\ \rho At < W \end{array} \end{aligned} \tag{1.8}$$

**Example 3 Least Squares Regression**

Find best line that goes through data. In Figure 1.2 the data are represented as circles, and a not-good model is indicated by the dashed line. The best model is represented as the solid line. How is the best model chosen? In least squares regression, it is the line that minimizes the sum of squared deviations between data and model-dependent variable (vertical deviations as represented by the vertical lines).

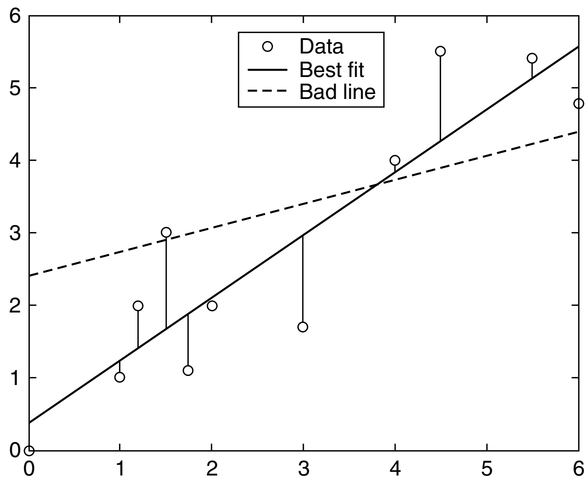



**Figure 1.2** Regression illustration.

In classic regression, we evaluate best as minimizing sum of squared deviation:

$$\mathrm{OF} = \mathrm{SSD} = \sum_{i=1}^{N} \left(y_i - \tilde{y}(x_i)\right)^2 \tag{1.9}$$

where $y_i$ and $x_i$ are data pairs, $N$ is the number of data points, and $\tilde{y}$ is a model value.

The equation of a line is $\tilde{y} = a + bx$, which is linear in both of the coefficients. The objective is to find the values for $a$ and $b$ that minimize the OF. Here, the decision variables are the model coefficients $\{a, b\}$, and the optimization application can be stated as

$$\begin{aligned} &\min_{\{a,b\}} J = \sum_{i=1}^{N} \left(y_i - \tilde{y}(x_i)\right)^2 \\ &\text{S.T.}: \quad \tilde{y}_i = \tilde{y}(x_i) = a + bx_i \end{aligned} \tag{1.10}$$

The statement previously is not the answer; it is a canonical statement of the problem. Any set of model coefficients $\{a, b\}$ are feasible. The set that results in the not best dashed line in Figure 1.2 is a trial solution, a possible DV choice. The optimum DV set that results in the best line in the figure is indicated as $\mathrm{DV}^*$, $\{a^*, b^*\}$. The classical analytical method to solve a least squares regression application with a model that is linear in coefficients is to take the derivative of the OF w.r.t. each model coefficient, set each derivative to zero (the normal equations), and solve the simultaneous linear equations.

Substituting the model into the OF, $J = \sum_{i=1}^{N} \left(y_i - (a + bx_i)\right)^2$, taking the derivatives, and setting them to zero,

$$\frac{\partial J}{\partial a} = 0 = -2\sum_{i=1}^{N} (y_i - a - bx_i) \tag{1.11a}$$

$$\frac{\partial J}{\partial b} = 0 = -2\sum_{i=1}^{N} x_i(y_i - a - bx_i) \tag{1.11b}$$

Rearranging, the normal equations are

$$(N)a + \left(\sum x_i\right)b = \left(\sum y_i\right) \tag{1.12a}$$

$$\left(\sum x_i\right)a + \left(\sum x_i^2\right)b = \left(\sum x_i y_i\right) \tag{1.12b}$$

Since each term in parentheses has known values from the data, the normal equations result in two linear equations in two unknowns, variables $a$ and $b$.

This was a classic least squares approach with a model that is linear in coefficients. A two-coefficient model results in two linear equations. This procedure is scalable. An $M$ coefficient model that is linear in the coefficients results in $M$ normal equations, linear in the $M$ coefficient unknowns, requiring linear algebra techniques to obtain the model coefficient values.

However, nonlinear models are not always amenable to this analytical approach. The derivative might not have a tractable analytical expression, and if it does, the solution to the resulting equations will likely require iterative nonlinear root-finding techniques.

**Example 4 Best Path or Sequence**

Find the best path to visit major baseball parks in the Northeastern United States. Should this be least miles, least time, lowest cost, or most scenic? Start at home, and return to home. One possible sequence is to go to Chicago (C), then Boston (Bo), then New York (N), and so on.

Give a number to each city in sequence. Assign home as points 1 and 8, start and finish. For example, here is one plan, one sequence, one path:

| H | C | Bo | N | Pit | Phi | Bal | H |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

But this sequence might have fewer total miles:

| H | C | Bo | N | Pit | Phi | Bal | H |
|---|---|---|---|---|---|---|---|
| 1 | 7 | 6 | 5 | 2 | 3 | 4 | 8 |

If minimizing distance for the total trip is the objective, using $l_{ij}$ = distance between city $i$ and $j$, the optimization application can be stated as

$$\min_{\text{sequence}\{2,3,\ldots,7\}} J = \sum_{i=1}^{N} l_{ij} \tag{1.13}$$

This is a classic example of the traveling salesman problem (TSP), which is a key element in scheduling planning, sequencing, and logistics. The example reveals several issues. One is that the decision variables are nominal (or category variables), indicated here as integers for the sequence. Accordingly, there is no concept for a derivative, revealing that many applications are not amenable to the classic analytical approach to determining DV*.

The other issue is an additional concept, the constraint. Here the sequence is constrained so that (i) each city number is used, (ii) each is used only once in the sequence, and (iii) the initial and final cities are constrained to $H$.

## 1.4 Terminology Continued

Returning to concepts and nomenclature, constraints are a key issue in optimization.

### 1.4.1 Constraint

These are values or issues that cannot be violated. These may include physical "laws," man-made laws or procedures, infeasible math operations, procedural steps, or many other features. Constraints can be on the DV, OF, or auxiliary (other related, but secondary) variables. Constraints can be on values, rates of change, or transitions. They can be on current action or on future implications of today's action. Constraints often are the most important influence on determining the $DV^*$ values.

Here is an introduction to the diversity of constraints:

- Constraint on transition: When visiting major league ballparks, don't go to NYC immediately after Boston or vice versa because of fan loyalty. When operating a mixer, you must fill and blend material in the mixer before dumping.
- Constraint on rate of change: Don't immediately "floor" the automobile accelerator pedal position, because it startles passengers. Gradually move it from one position to another, no faster than $5^\circ$ of arc per second.
- Constraint on DV: A composition must be between 0 and 100%. A flow rate must be nonnegative.
- Constraint on secondary variables (auxiliary variables): Don't let a downstream tank overflow or run dry. The sum of all compositions must be equal to unity. Keep the fire temperature below the melting point of the nozzle material. An automobile engine temperature cannot exceed $180^\circ$F.
- Constraints may relate to natural laws: Human bodies cannot float in the Earth's atmosphere because of the density difference between mostly water and air. Heat does not naturally flow from cold to hot, because entropy must increase. As close as cellulose is to sugar, and even though grazing animals can digest cellulose, humans cannot live on cellulose.
- Constraints may relate to human laws or procedures: You cannot build your house on your neighbor's property. You cannot freely travel from any one country to any other. Wear your uniform and follow this procedure to clock your work hours.
- Constraints may occur in the future: Today you can make the purchases on your credit card; but the constraint happens in the future when all of your income is paying accumulated interest. You can drive the race car now and pass a pit stop; but the worn tires will fail in half a lap.
- Constraints may be soft or hard: It is OK to violate soft constraints. The speed limit is 60 mph, but enforcement permits you to drive 4 mph higher. Other soft constraint examples, in which mild violation is permissible might include, "Don't yell at your children." Wear white hats from Memorial Day to Labor Day. By contrast, hard constraints may not be violated. "Don't let the mixture composition get into the explosive limits." "Pay your taxes." Don't try to take the logarithm of a negative number.

### 1.4.2 Feasible Solutions

Some ideas, such as perpetual motion, are infeasible. They violate hard constraints, such as the first or second laws of conservation. Feasible solutions are the DV values that do not violate any of the constraints such as those listed previously.

### 1.4.3 Minimize or Maximize

Minimizing seeks the lowest point, or smallest value. Maximizing seeks the highest or largest. But these are mathematically equivalent when the OF is multiplied by negative unity. If you wish to maximize, you could minimize the negative of the OF and find the same DV* value. Turn the graph of Figure 1.3a upside down and compare it to the Figure 1.3b to intuitively confirm this. Chapter 16 will provide mathematical analysis of this and other related useful OF transformations.

This book will focus on minimization. With only one optimization objective, the concepts are less confusing, and there is no loss of generality.

### 1.4.4 Canonical Form of the Optimization Statement

There is a conventional style of presenting the optimization application and associated elements. Equation (1.14) is an example. To begin the statement, in the upper left, state whether the objective is to minimize (min) or maximize (max) the OF. Below the min or max statement explicitly, list the decision variables within braces. Define the objective function, *J*. It might be a simple formula, or something that represents a procedure such as SSD, DCFRR, or the calculation of risk. The objective function might even be its own optimization, for instance, to identify the worst or maximum possible event, and the optimization might be to identify DV values that minimize a worst-case outcome. In designing a product, one might wish to assess the worst possible failure situation and design to minimize the maximum risk. List the constraints in lines that follow the S.T.: (subject to) label. Illustrated here are three types of constraints—one on the decision variables, one on a function of the DVs, and one on a rate of change:

$$\begin{aligned} &\min_{\{\mathrm{DV}\}} J = \text{you define this relation} \\ &\text{S.T.:}\quad P_1 = f(\mathrm{DV}) \le P_{10} \\ &\quad a \le \mathrm{DV}_1 \le b \\ &\quad \text{rate} = \frac{dP_2}{dt} \le \text{rate}_0 \\ &\quad \vdots \end{aligned} \tag{1.14}$$

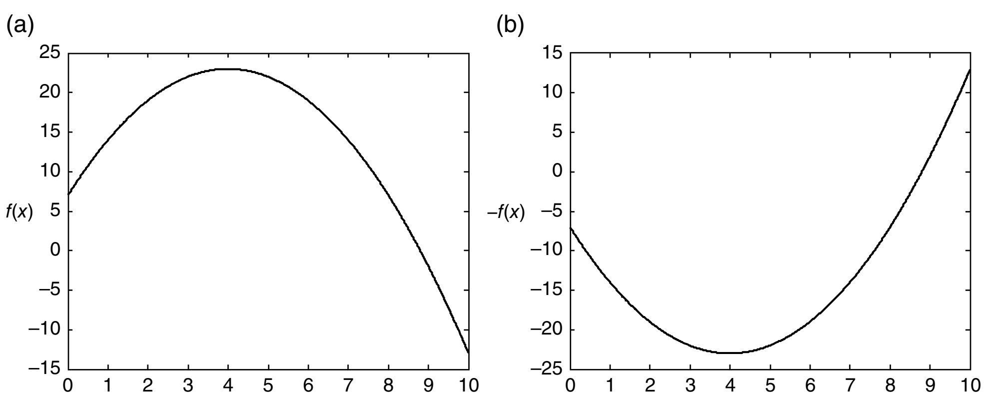



**Figure 1.3** Equivalency of maximizing and minimizing the negative: (a) seeking the maximum and (b) seeking the minimum of the negative of the function.

This canonical form is just a communication tool. It is neither the solution nor the method for finding the solution. It may also be incomplete; and if it is, include the necessary supporting information. For instance, is the OF being evaluated experimentally or by a simulator? In either case the reader may wish to know essential details about the experimental design or the idealizations and methodology of the simulator model.

The optimization math is relatively easy, standard, and widely available. The application problems are typically related to:

- Characterizing the valid and appropriate and comprehensive measure of goodness, how to decide best, the OF
- Identifying the DVs (are they values, timing between values, sequence of operations, which variables?)
- Identifying the constraints ("Oh, I didn't know shortest path required me to stay on the roads!")
- Generating the models so the OF value or constraint conditions can be calculated given the DVs
- Choosing the optimization algorithm that is right for the application
- Choosing thresholds and parameters related to the algorithm initialization, operation, and convergence

Although this book describes many fundamental optimization algorithms, it is dedicated to a focus on the application issues listed previously.

## 1.5 Optimization Procedure

There are eight basic stages in an optimization procedure:

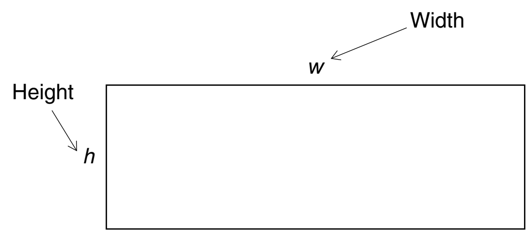
1) *State the OF, or objective function.*
2) *State whether you wish to minimize or maximize.*
3) *State the DVs, or decision variables, and what you are free to change to effect the OF value.*
4) *Use model relationships to relate the DV to all parts of the OF.*
5) *Define constraints.*
6) *State the method used for solving the optimization application. This is the optimization approach, algorithm, procedure, or logic.*
7) *Execute the procedure.*
8) *Reflect on it all.*

A few examples will reveal the stages in developing and solving an optimization application and reveal the often overlooked importance of the eighth.

**Example 5** Minimize the perimeter of a rectangle that provides a desired area by choosing the height and width. Figure 1.4 represents a mental construct of a rectangle. We need the mental construct to derive models that represent the interaction of DVs and OF. The model of the mental construct is not the reality, but it is a mathematical approximation of reality.

1) State the OF, or objective function: $J = \text{perimeter} = 2h + 2w$.
2) State whether you wish to minimize or maximize: Minimize $J = 2h + 2w$.
3) State the DVs, or decision variables, and what you are free to change to effect the OF value. In this case it is just one item, either the height or the width. Given one value, the other is

**Figure 1.4** Analysis of a rectangle.




calculated from the desired area. $A = hw$. I'll choose height as the DV. The basic optimization statement is

$$\min_{\{h\}} J = 2h + 2w \tag{1.15}$$

4) Use model relationships to relate the DV to all parts of the OF. In general we need a model of the device or process to relate state variables to influences. In this case $h$ is explicit in the OF. However, $h$ also affects $w$, which is not explicitly revealed in the OF. We need the complete relation. These models are often termed constitutive relations. From the definition of area in a rectangle,

$$w = \frac{A}{h} \tag{1.16}$$

Inserting it into the OF completely indicates how the choice of $h$-value affects the OF value:

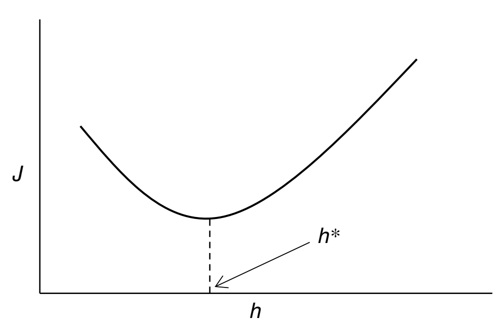
$$\min_{\{h\}} J = 2h + \frac{2A}{h} \tag{1.17}$$

5) Define constraints

$$h > 0 \tag{1.18}$$

6) State the method used for solving the optimization application and for adjusting $h$ to find the value $h^*$ that minimizes the OF. This is the optimization approach, algorithm, method, or logic. The analytical method is one algorithm. Since the OF permits its use, I'll use it. Assume $J$ is a function with one minimum, as illustrated in Figure 1.5.

**Figure 1.5** A function with one minimum.



At the minimum, $dJ/dh|_{h^*} = 0$. Elsewhere, the slope is not zero. So, get the analytical derivative of $J$ w.r.t. $h$, equate it to zero, and solve for the value of $h^*$.

7) Execute the procedure:

$$\left.\frac{dJ}{dh}\right|_{h^*} = 0 = 2 - \frac{2A}{(h^*)^2} \tag{1.19}$$

The solution is

$$h^* = \sqrt{A} \tag{1.20}$$

which also means that $w^* = h^*$ and the rectangle is a square.

8) Reflect on it all (the model, the assumptions, the answer, the method, the acceptance of your solution by others in the enterprise). Evaluate your procedure and results. The first application and solution provides progressive insight. What would you do differently in OF, DV, constraints, model, and method? What idealizations limit validity of result? When others look at your analysis, what might they find incomplete or overlooked? Redo the work with this added depth of insight. This was an ideal initial analysis.

   For instance: Is the item being modeled an ideal rectangle of lines of zero width? Other than mental constructs and intellectual exercises, where do you find such things? An exercise like this might be the idealization of an application to choose a window pane shape to minimize the consumption of frame material. But the frame has thickness, and it must both overlap the window a bit and extend beyond. So, perhaps the volume of frame material, not the perimeter, is the real OF. Also, is it the window pane area or the open area inside the frame that is important? The answer to the analysis above was a square, but how does this fit with your personal experience? Are window panes square? What considerations seem to override the analysis to give an alternate (not $h = w$) solution?

**Example 6** Maximize power delivered to a resistor from a battery with voltage $V$. Figure 1.6 represents a mental construct of a battery within the dashed lines as an ideal battery followed by an internal

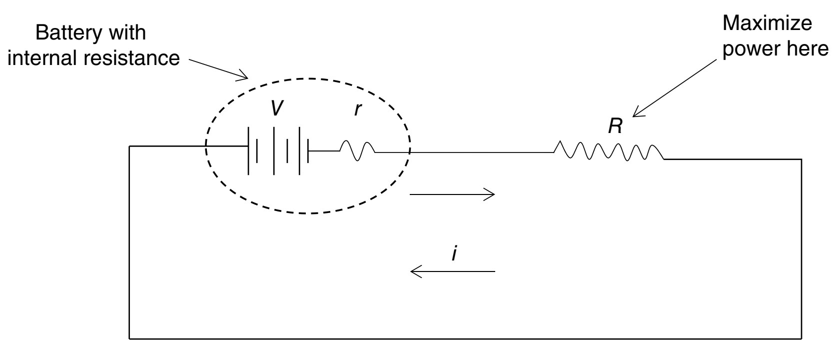



**Figure 1.6** Mental construct of a battery in a resistance circuit.

resistor. We need the mental construct to derive a model that represents the interaction of DVs and OF. The model of the mental construct is not the reality, but it is a mathematical approximation of reality.

1) State the OF, or objective function: $J = \text{Power} = i^2 R$.
2) State whether you wish to minimize or maximize: Maximize $J = i^2 R$.
3) State the DVs, or decision variables, and what you are free to change to effect the OF value. In this case it is just one item, the external resistor resistance, DV = $R$. The basic optimization statement is

$$\max_{\{R\}} J = i^2 R \tag{1.21}$$

4) Use model relationships to relate the DV to all parts of the OF. In general we need a model of the device or process to relate state variables to influences. In this case $R$ is explicit in the OF. However, $R$ also affects $i$, and the model is not explicitly revealed in the OF. We need the complete relation. These models are often termed constitutive relations. From elementary circuit analysis there is a relation between current and resistance:

$$V = i(R + r) \rightarrow i = \frac{V}{R + r} \tag{1.22}$$

Inserting it into the OF completely indicates how the choice of $R$ value affects the OF value:

$$\max_{\{R\}} J = \frac{V^2 R}{(R + r)^2} \tag{1.23}$$

5) Define constraints

$$R > 0 \tag{1.24}$$

6) State the method used for solving the optimization application and for adjusting $R$ to find the value $R^*$ that maximizes OF. This is the optimization approach, algorithm, method, or logic. The analytical method is one algorithm and applicable to this example. Assume $J$ is a function with one maximum as illustrated in Figure 1.7.

At the maximum, $dJ/dR|_{R^*} = 0$. Elsewhere, the slope is not zero. So, get the analytical derivative of $J$ w.r.t. $R$, equate it to zero, and solve for the value of $R$.

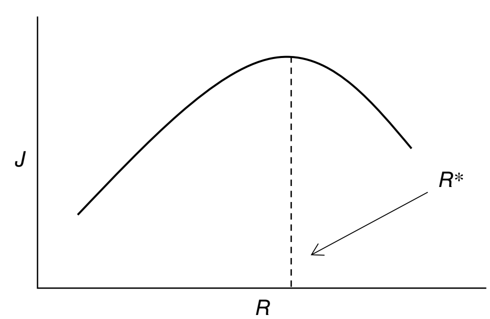



**Figure 1.7** Illustration of a function with a single maximum.

7) Execute the procedure:

$$\left.\frac{dJ}{dR}\right|_{R^*} = 0 = \frac{V^2}{(R+r)^2} - 2\frac{V^2 R}{(R+r)^3} \tag{1.25}$$

The solution is

$$R^* = r \tag{1.26}$$

8) Reflect on it all (the model, the assumptions, the answer, the method, the acceptance of your solution by others in the enterprise). Evaluate your procedure and results. The first application and solution provides progressive insight. What would you do differently in OF, DV, constraints, model, and method? What idealizations limit validity of result? When others look at your analysis, what might they find incomplete or overlooked? Redo the work with this added depth of insight. This was an ideal initial analysis.

   For instance, is the model of the battery correct (is the internal resistance in series with the ideal battery or is it more complicated)? Is the circuit model correct (zero resistance in the wire)? Is there an impedance effect on start-up? Does the battery voltage remain at the initial "$V$" while it is discharging? If you buy a resistor of $R^* = r$ what value will $R$ actually be? Are the "given" $V$ and $r$ values the absolute truth for all circuits to be made? This OF looked at maximizing delivered power, but are there other measures of performance or cost that have been overlooked?

Returning to the importance of stage 8, the simplicity of examples and solutions, which is typical of the classroom teaching style, may provide a pedagogical method that facilitates initial understanding; but it falls short of the comprehensive issues that affect reality and so misguides a student. Optimization is not about the idealized, intellectual mathematical analysis. It is done to improve human choices in design, operations, and procedures, and the application needs to align with the comprehensive reality, not an isolated idealization. Often the step-back reflection on the outcomes that reshapes the statement objective and assumptions is the most important aspect to optimization.

## 1.6 Issues That Shape Optimization Procedures

As mentioned in the prior examples, the classical analytical optimization concept (set the derivative to zero and solve for the DV*) is often not feasible. As a result, a diversity of optimization procedures have been developed to surmount the issues. There are many reasons that make the ideal approach inapplicable. Here is a preview listing of difficulties that real applications present:

- Nonlinearities—The objective function may not have an analytically tractable derivative. If it does, the derivative may require iterative nonlinear root-finding procedures to solve an implicit nonlinear relation.
- Discontinuities—The OF or its derivatives may have discontinuities.
- Discretization—The decision variable might represent integer values (number of queuing lines, number of parallel devices, number of samplings for a delay) or discretized sizes (pipes, resistors,

and shirts come in discrete sizes). Further, the numerical time or space discretization in the model used for the OF will generate steps or ridges (striations) on what might be considered as a smooth OF. Convergence criterion on root-finding techniques used within the model can also cause striations. For these, analytical optimization techniques, which were developed for continuum-valued OFs and DVs, will be confounded.

- Multiple optima—Many objective functions have local minima that would trap the optimization procedure within a local, not the global optimum.
- Flat spots—Some OFs have zero, or effectively zero, response to the DV in regions of saturation or inconsequence, and if the derivative is zero, there is no guidance as to how to improve the DV solution.
- Stochastic response—When OF data is being generated experimentally, there is noise (uncertainty, experimental error) on the OF value. A replicate experiment (an attempt to implement the same DV values) will not produce exactly the same OF value. Here, because of experimental vagaries, moving the TS toward the true DV* value might return a worse, not better OF value, which would indicate that the optimum is in the false direction. When Monte Carlo simulations are used as a surrogate for experiments, the impact is the same.
- Uncertainty—There is uncertainty in the givens. Air pressure, temperature, humidity, and wind velocity continually change. If you design an airplane wing for one set of conditions, what is the outcome at other realizations of the givens? Should the design be to minimize the worst possible outcome over all possible realizations?
- Constraints—Types of constraints were described previously. When constraints are hard (cannot be violated), the optimization procedure needs to be modified to choose an alternate path. If you encounter a high wall blocking your downhill walk, but want to get to the other side, perhaps walk along the wall, not directly downhill.

The useful applications of optimization encounter those sorts of issues. Accordingly, the useful methods for optimization need to work in spite of such issues.

## 1.7 Opposing Trends

If the OF has a monotonic response to the DV, then the optimum is at an extreme. Consider this example:

$$\min_{\{x\}} J = 7e^{-x/10} \tag{1.27}$$

The function decreases monotonically with the $x$-value, as shown in Figure 1.8, and the minimum value for $J$ occurs at $x^* = +\infty$, at an extreme value.

Since a value of infinity is not practical, the solution is to use the largest $x$-value as possible. Some optimization applications legitimately send the DV to a constraint. For example, "What speed should you travel to get there in minimum time?" The answer is drive at the maximum speed possible. However, in my experience, usually such an outcome indicates that the user has not properly included all relevant features in the OF. In a car speed example, these additional considerations may include fuel consumption, wear on the vehicle, and the risks associated with an accident, being arrested, or setting an example for others.

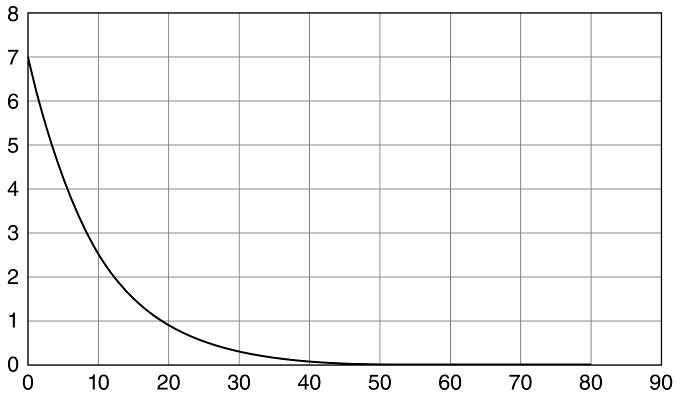



**Figure 1.8** Illustrating a monotonic OF response to a DV.

Usually there are at least two opposing responses affecting the OF, and optimization seeks to find the best balance of the opposing ideals. One response would get better with increasing the DV value, and the other would get worse. This can be illustrated as a product of functions; a simple example building on the OF above is

$$\min_{\{x\}} J = 7xe^{-x/10} \tag{1.28}$$

Here the first function, $7x$, rises with the value of $x$, and the second function, $e^{-x/10}$, diminishes with the value of $x$. The solution is an intermediate value $x^* = 10$, as illustrated in Figure 1.9.

Alternately the opposing functionalities may be additive, as with this example and Figure 1.10 illustration:

$$\min_{\{x\}} J = 7e^{-x/10} - 5e^{-x/5} \tag{1.29}$$

Here the $-5e^{-x/5}$ functionality rises with $x$, and the $7e^{-x/10}$ falls with $x$-value. And $x^* = 10^*\ln(10/7) = 3.566749\ldots$

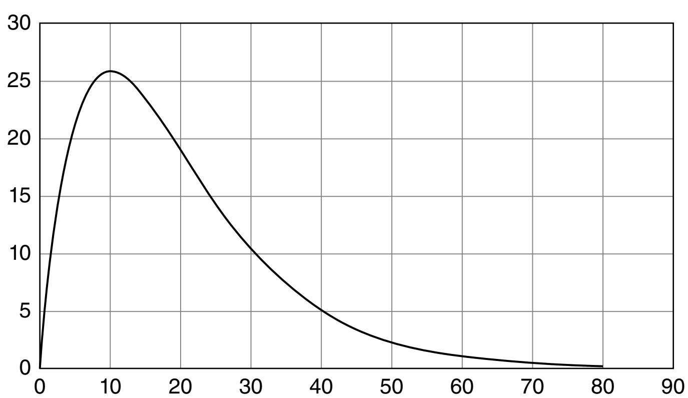



**Figure 1.9** Illustrating a dual OF response to a DV—multiplicative functionality.

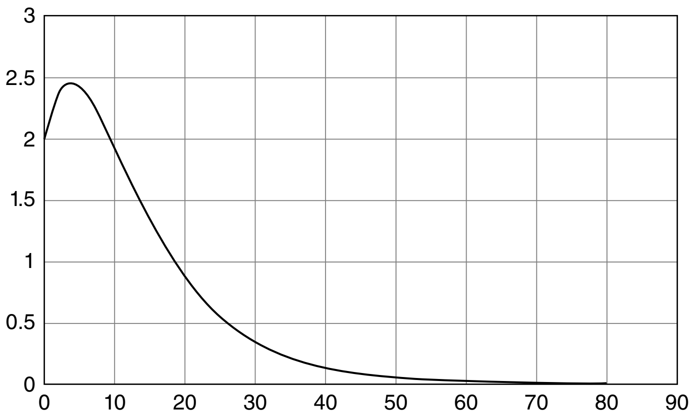



**Figure 1.10** Illustrating a dual OF response to a DV—additive functionality.

In a real application, the opposing influences may be much more complicated functions, such as iterative computational procedures; but for an optimum to exist at a non-extreme $x$-value, at an intermediate $x$-value, the OF must represent a balance of opposing ideals.

Further, one ideal must make larger OF changes than the other when $x$ changes in the $x < x^*$ range, and vice versa in the $x > x^*$. To illustrate this consider the slight perturbation to the application represented by Equation (1.29) in which the $x$-value in the argument of the second term is scaled by 8 instead of 5:

$$\begin{aligned} &\min_{\{x\}} J = 7e^{-x/10} - 5e^{-x/8} \\ &\text{S.T.:} \quad x > 0 \end{aligned} \tag{1.30}$$

The opposing idealities of the functionalities are the same (one increasing with $x$, the other decreasing), but the relative magnitude of the two influences does not change, and the OF is monotonic, as illustrated in Figure 1.11.

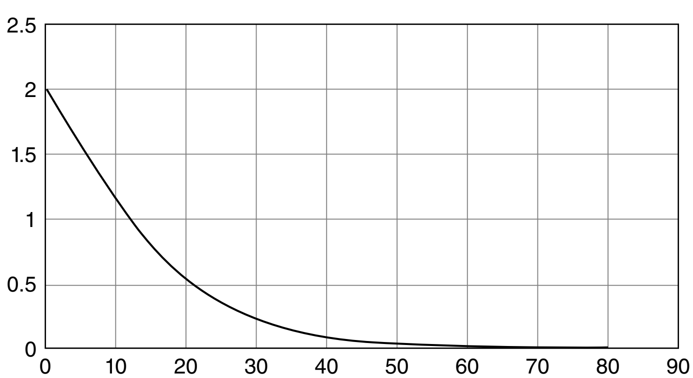



**Figure 1.11** Illustrating a monotonic OF response to a DV—additive functionality.

## 1.8 Uncertainty

A typical optimization application is confounded with multiple aspects of uncertainty. Consider this simple application: "What insulation thickness on a pipe is optimal?" The OF will be related to annualized cost associated with energy losses and a 5-year life of the insulation. Model the cost of insulation as proportional to the volume of material coating a pipe of length $L$ and radius of $r_p$. The insulation outer thickness is $r$. A classic model for the rate of heat transfer per surface area from an insulated pipe (with zero resistance from the metal pipe or internal liquid to pipe convection) is

$$\dot{q}'' = \frac{T_{\text{surrounding}} - T_{\text{internal}}}{\left[\ln\left(r/r_p\right)/2\pi k\right] + (1/2\pi r h)} \tag{1.31}$$

where $k$ represents the thermal conductivity of the insulation and $h$ is the heat transfer coefficient between the insulation surface and the surrounding air.

The objective is to minimize annualized costs over a 5-year period:

$$\min_{\{r\}} J = c_{\text{insulation}}\pi\left(r^2 - {r_p}^2\right)L + 5c_{\text{energy}}Y2\pi r L\frac{T_{\text{surrounding}} - T_{\text{internal}}}{\left[\ln\left(r/r_p\right)/2\pi k\right] + (1/2\pi r h)} \tag{1.32}$$

Here $c$-values represent insulation and energy costs per unit, $Y$ is the number of time units in a year, and $L$ is the piping length.

While the formula might look complicated, and the pipe insulation application might not be familiar to many, it is still a relatively simple application. Input the values for the "givens" ($c_{\text{insulation}}$, $r_p$, $c_{\text{energy}}$, $T_{\text{surrounding}}$, $T_{\text{internal}}$, $k$, $h$, etc.), and it is relatively easy to let a computer determine the optimal value, $r^*$.

If the "givens" remain unchanged, then any computer that solves Equation (1.32) will get the same value for $r^*$ regardless of the day, or who is the data-entering user. However, that answer might not be right. Consider two aspects. One is uncertainty in the givens, and the other is associated with the model.

*Givens:* Will the surrounding temperature remain at the value of $T_{\text{surrounding}}$ for a 5-year period? Will the energy cost remain unchanged from the value of $c_{\text{energy}}$? Might not age and environmental effects change the insulation $k$-value from the installed value? What will the programmer choose to use for the value of $\pi$? Since none of the givens have values that are known with certainty, changing those values to other reasonable values will change the calculated $t^*$ (insulation thickness) $\left(t = r - r_p\right)$ value. For instance, if the optimal value at nominal conditions is $t^* = 2.54$ cm, it might range between 2 and 3 cm for combinations of alternate reasonable values of the givens.

In any case, the same optimization procedure would evaluate the same method of OF calculation. The optimization procedure is not dependent on the values of the givens, but the resulting OF* or DV* outcomes are.

*Model:* Is there no internal resistance to heat transmission (fluid to pipe)? Is the outside heat transfer coefficient, $h$, uniform along the length of the pipe and at the bottom, top, and sides? What is the right model to use to estimate the $h$-value? Alternate models for the process may lead to $t^*$ values that range between 1.7 and 3.5 cm, and compounded with uncertainty in the givens, the conclusion might not be the ideal $t^* = 2.54$, but a more realistic $1.3 < t^* < 3.8$ cm.

The business decision is not based on $t^* = 2.54$ cm. For instance, if the capital supply is ample, then the decision might be to use the 3.8 cm thickness. But if capital is scarce, use the 1.7 thickness;

and if it is found to be inadequate, then add more insulation later. Recognizing uncertainty in the $DV^*$ value is as important as the $DV^*$ value.

## 1.9 Over- and Under-specification in Linear Equations

Over-specified means that there are more independent equations than independent variables, more constraints than variables—in this case the degrees of freedom <0. A straight line can go through any two points, but if you add a third point that is not on the line, the single line cannot simultaneously go through all three points.

Under-specified means that there are fewer equations (constraints) than variables, degrees of freedom >0. For example, if there is only one point, then you could choose the slope of the line and still find a line to go through the point. You are free to choose a variable value. You have a degree of freedom.

Balanced means that there are the same number of equations as variables: DoF = 0.

From a linear equation understanding, consider three equations with three unknowns $\{x, y, z\}$:

$$
\begin{aligned}
a_1 x + b_1 y + c_1 z &= q_1 \\
a_2 x + b_2 y + c_2 z &= q_2 \\
a_3 x + b_3 y + c_3 z &= q_3
\end{aligned}
\tag{1.33}
$$

If the equations are linearly independent, there is a unique solution for the $\{x, y, z\}$ values. The DoF = 0. The three might be considered equations, or specifications, or active constraints, and the $\{x, y, z\}$ set might be considered to be the DVs. The equations can be written in vector-matrix form:

$$
\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix}
\begin{bmatrix} x \\ y \\ z \end{bmatrix}
=
\begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix}
\tag{1.34}
$$

In this familiar representation, the number of equations (which appears to be just one equation when counting the number of equal signs) may not necessarily match the number of unknown variables. Here degrees of freedom = number of DVs − number of specifications (active constraints).

Under-specified: Consider 2 equations and 5 unknowns. Here DOF = 5 − 2 = 3. This means that you can choose any value for three (any three) unknowns and then solve for the others. The solution is not unique. The problem is under-specified. If you can choose which variable and its value, then which is best? This is an optimization. Would your choices be based on the simplest remaining solution, or a conventional value, or something that pleases your boss?

Balanced: Consider 3 equations and 3 unknowns. Here DOF = 0. This means you can find an exact and unique solution.

Over-specified: Consider 4 equations and 3 unknowns. Here DOF = −1. This means that you cannot fit all four conditions simultaneously. The problem is constrained. In this case you could (i) relax (eliminate) one condition and solve for others or (ii) find a best compromise solution with a weighted objective function that includes a penalty for each equation deviation from zero as an OF term. For instance, square each deviation, sum the weighted squares, and seek a DV set of values that minimizes the objective function. (The choice you make is an optimization. Which might be best for an application? How would you assess best?)

## 1.10 Over- and Under-specification in Optimization

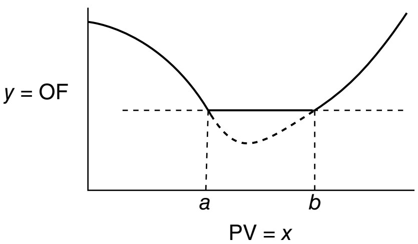



**Figure 1.12** Illustrating an under-specified application.

There are parallels in optimization. Figure 1.12 illustrates an under-specified application, for example:

$$\min_{\{x\}} J = y \begin{cases} = c & \text{if } a \le x \le b \\ = 1 - e^{-(x-d)^2} & \text{otherwise} \end{cases} \tag{1.35}$$

Any $x$ in the $[a, b]$ range is equivalent to any other $x$-value. There is flexibility in the solution. There are extra choices you can make. This may also arise if there are distinct but identical solutions. For instance,

$$\min_{\{x\}} J = \left(x^2 - 9\right)^2 \tag{1.36}$$

Here there are two identical solutions: $x = (3, -3)$. If you can choose one, which is preferred? In this case there are probably reasons other than that expressed in the OF as to why one $x$-value is better than another. Consider resource conservation, future flexibility, reliability, political capital, etc., and include this additional concept into the OF.

Alternately, under-specified may be the result of redundant parameters. Here is an example of a hyper-elastic model of stress versus strain:

$$\sigma = A\left(e^{B\epsilon} - 1\right) \tag{1.37}$$

It appears to have independent model coefficients $A$ and $B$. However, if either $B$ or $\epsilon$ is very small, then $e^{B\epsilon} \cong 1 + B\epsilon$ and the model effectively is equivalent to

$$\sigma \cong AB\epsilon = k\epsilon \tag{1.38}$$

In this view, stress is proportional to a constant, $k$, times strain. The model is dependent on the value of $k$, and it does not matter what the individual values for $A$ and $B$ are, as long as their product is equal to $k$. A contour plot of the least squares solution of

$$\min_{\{A,B\}} J = \sum (\sigma - \tilde{\sigma})^2 \tag{1.39}$$

would show a nearly common (nearly constant valued) minimum (valley) along the curve $B = k/A$, or $DV_1$ proportional to the reciprocal of $DV_2$.

If you notice an under-specified application, multiple $DV^*$ values giving essentially identical $OF^*$ values, then use this finding to trigger reassessment of the models and description.

Figure 1.13 reveals a balanced application with a unique minimum solution.

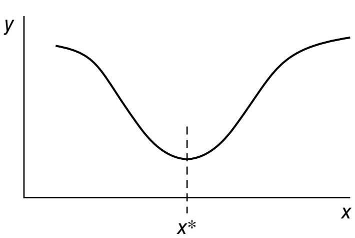



**Figure 1.13** Illustrating a balanced application.

Figure 1.14 reveals an over-specified or constrained application, in which the minimum of $y(x)$ occurs in the constrained $x$-region:

$$\min_{\{x\}} J = y(x) \\ \text{S.T.}: \quad x < a \tag{1.40}$$

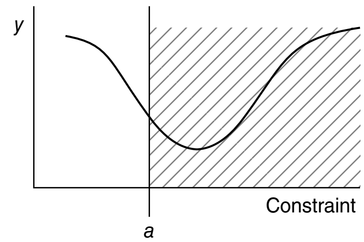



**Figure 1.14** Illustrating an over-specified application.

If over-specified or constrained, you could either (i) use the hard constraint to limit optimizer trial solution values or (ii) add a penalty (soft constraint) for the constraint violation to the OF.

If the solution is on a constraint, and if the constraint defines the DV* value, then the constraint is the critical aspect. Then move your focus from solving the original optimization statement and return focus to the application and do these: (i) Be sure that the constraint is properly specified. (ii) Seek alternate designs (of the process, device, procedure) that would remove the constraint. The real solution may be outside of the mathematical optimization exercise that would seek how to include the constraint in the problem.

## 1.11 Test Functions

Any practical application will not have a simple one-line equation as the OF. However, for instruction and for convenient testing, many people have created simple one-line equations as easy-to-implement, rapidly computed surrogates to reveal fundamental issues. Many of these test functions have become the standard benchmark problems that people use to develop and compare optimizers. The problem is that when these relations constitute the majority of the examples for a novice, it misrepresents the true complexity of optimization.

Appendix E presents a set of applications that are grounded in physical reality illustrating diverse application issues but that are still relatively simple for exploring optimization techniques.

## 1.12 Significant Dates in Optimization

I find that a review of significant dates related to optimization provides revealing issues about why we have certain procedures that seem to dominate our legacy and instructional material. Here are selected dates for publications of algorithms:

- 1669: Newton—Root finding (this is not optimization, but it is a common element in many algorithms) (not published until 1711).
- 1690: Raphson simplified Newton's method for root finding.
- 1740: Simpson extended Newton's method to root finding on the derivative (analytical optimization). Perhaps the optimization application should be named Simpson's method.
- 1807: Legendre presented the least squares method (there were parallel developments by Gauss).
- 1847: Cauchy—Sequential line search on a steepest descent line.
- 1939: Kantorovich—Linear programming (LP) model and an algorithm for solving it.
- 1944: von Neumann and O. Morgenstern solved sequential decision problems by using the idea of dynamic programming (DP). A. Wald (1947) did related research. Another early application of DP is presented by P. Massé (1944) for reservoir management.

- 1948: Dantzig—Linear programming and the simplex tableau, now a common algorithm.
- 1952: Bellman's key publication on the theory of dynamic programming.
- 1952: Hestenes and Stiefel—Conjugate gradient approach.
- 1955: Levenberg—Blending of incremental steepest descent and Newton's methods (seemingly lost in the literature).
- 1956: Frank and Wolfe—Reduced gradient algorithm for constrained optimization.
- 1961: Hooke and Jeeves—Pattern search.
- 1962: Spendley, Hext, and Himsworth—Simplex search.
- 1963—Marquardt's rediscovery and popularizing of Levenberg's approach (now known as Levenberg–Marquardt).
- 1963: Wilson—Sequential quadratic programming.
- 1960s: Genetic algorithms and in general the start of mimetic approaches (algorithms that attempt to mimic nature's way of finding the optimum).
- 1965: Nelder and Mead's heuristic rules for expansion and contraction of SH&H's simplex.
- 1969: Broyden—Quasi-Newton procedures.
- 1974: Lasdon, Fox, and Ratner—Generalized reduced gradient method.
- 1979: Khachaturyan, Semenovskaya, and Vainshtein—Simulated annealing.
- 1980: Beginning of randomized search methods.
- 1989: Glover—Tabu search.
- 1990: Beginning of diverse mimetic algorithms—ant colony and bee colony.
- 1995: Kennedy and Eberhart—Particle swarm.
- 1997: Storn and Price—Differential evolution.
- 2003: Raphael and Smith—Probabilistic Global Search Lausanne.
- 2012: Leapfrogging.

Notable about this timeline is that prior to 1950, prior to widespread computer access, optimization was based on manual calculations. This generally limited applications to linear, quadratic, unconstrained, and well-behaved (mathematically tractable, low dimension) functionalities. And appropriate for these were second-order techniques such as Newton–Raphson.

In the 1950–1960 period, mainframe computers permitted inclusion of algorithms to handle surface aberrations. During this period, direct search (function evaluation only) approaches began to invade what was formerly exclusively optimization based on gradient-based (sensitivity of the OF to the DVs) search algorithms. This period began the rise of the more complicated second-order methods (LM, GRG), direct search (HJ, SHH-NM), and numerical approaches. But, there was limited access to mainframes for day-to-day applications with low economic or nonmilitary impact.

During 1960–1980 mainframes had widening access and began to gain acceptance in process design and optimization, and diverse and innovative ways to use this tool for formerly difficult optimization applications blossomed.

In the 1980s personal desktop computers and widespread convenient access to powerful machines shifted the algorithm focus from the number of function evaluations and single trial solution optimizers to multiplayer methods and their likelihood of finding the global minimum. This also shifted the algorithm development focus from deterministic logical/mathematical algorithms to mimetic stochastic algorithms that mimic nature (genetic algorithms, evolutionary algorithms, particle swarm, ant farm, simulated annealing, etc.).

Today, I sense that methods to solve stochastic applications are rising in importance, even at the day-to-day business level.

Although gradient-based and Newton-type approaches are still important, they were a product of the pre-computer era, and my bias is that direct search approaches are more compatible with computer techniques and solving difficult applications. This book does address archetypical gradient-based approaches but provides greater coverage to direct search approaches. Further, although this book will address classical formulation of optimization applications, it will also rise above the deterministic and quadratic limitations that characterize last-century methods to reveal today's solutions. Finally, as much fun as the mathematical analysis of optimization can be, the application issues have much greater practical importance. So, although this book will reveal some of the underlying analysis, I will limit it to the essential for understanding and emphasize the practical aspects of applications.

The pace of progress has been accelerating. Will it continue, or are we about to the limit of novel optimization ideas that take advantage of the computer? I have to think it will continue.

Nearly all of the techniques aforementioned were developed for deterministic applications. If you repeat the DV values, you get exactly the same OF value as a response. But life is not like that. It is stochastic. Here are three examples: First, if you are seeking to use experimental data as the OF, then experimental variability makes replicate runs not have the exact same response. Second, simulations of what life might impose are also stochastic. For instance, when deciding how much to invest now, for retirement 40 years in the future, one has to forecast inflation rates, interest rates, personal health, career outcomes, etc. There is great uncertainty in these "givens." Any "given" could have this value, or that value, and the value will change from year to year in unpredictable patterns. With the same DV (investment portion of salary) different values for the "givens," different possible situations, different realizations of what is possible, lead to different OF values. Finally, as the third example, I'll mention bootstrapping, a technique for random resampling of data to estimate the uncertainty in models as a result of variability in data. I believe that optimization of stochastic functions (Monte Carlo simulations) is now widely possible, and it is gaining acceptance and appreciation within the business–technical community. It appears that we have only seen the tip of the iceberg of applications and techniques for optimization of stochastic functions.

Also, nearly all of the techniques aforementioned were developed for continuum-valued functions. By contrast, when the model contains either variables that are discretized or conditionals, the OF response to the DV will have ridge/valley/cliff discontinuities. Often a magnified view of the surface is required to visualize these. First, consider discretization striations with a familiar example: a cloth fabric (for a shirt, or curtain, or socks) seems to be a continuum material from afar, but a microscopic view reveals the ridges in the weave or knit. The local striations on a surface will lead many optimizers downhill along the striation, not downhill globally. Alternately, to visualize a sharp valley, consider draping a piece of paper that contains a crease. Many optimizers will tend to jump back and forth across the crease, not follow the slope discontinuity downhill. Again, I believe that there is much opportunity for development of both applications and solutions for non-continuum objective functions.

What tools will be next in the progression of tools for optimization? How might they reshape algorithms? It is a wonder to me how easily water finds the downhill path in a steep valley. It remains difficult for all optimizers, but seems easy for water. Maybe the magic that water uses is that it is composed of individual molecules, and when billions of them are in the downhill side of the valley, the probability of one of them finding the downhill path is high.

All in all, it appears that the near-term future promises much in continued development and analysis of optimization.

## 1.13 Iterative Procedures

The several examples previously have simple equations that permit taking the analytical derivative and then using it to explicitly solve for DV*. However, applications that I've encountered are analytically intractable. The equation that relates OF value to TS value may be too complicated to be confident that the derivative is correctly formulated. Or the OF value might be computed from a black-box simulator that does not reveal internal equations. Or the OF value might be the result of physical experimentation. In these cases, we need another procedure to determine DV*.

There are many that will be revealed in subsequent chapters, but here, I'll introduce a heuristic direct search. The term direct search contrasts procedures that use gradient (derivative) information. A direct search only uses the function evaluation. The term heuristic means that it uses intuitive or tried-and-true human experience, as opposed to a mathematically formulated rule. To understand the heuristic direct search, imagine how a blindfolded tightrope walker would find the bottom of the rope span if he were placed randomly along the rope. The rule is take a step in one direction. If you go downhill, the step was the right direction; so, from this best-so-far spot, make the next step in the same direction. If you go uphill, the step is probably the wrong direction; so, return to the best-so-far spot, and make the next step in the other direction. If the reverse step had the same size, then you return to a prior spot, so when reversing direction, also cut the step size in half. By contrast, if going in the right direction, make the next step a bit larger. Near the minimum the alternating steps are progressively cut in half. Stop, and claim success when the step distance is small enough.

Here is an algorithm outline for the heuristic direct search

1) Initialize
   Choose the initial feasible TS, the base case DV
   Evaluate the OF-value, $OF_{base}$ at the $DV_{base}$
   Choose a step size, $DV_{delta}$
   Choose a convergence threshold for $DV_{delta}$
2) Test a new TS
      Set the new $TS=DV_{base}+DV_{delta}$
      Evaluate the OF-value, $OF_{TS}$ at the TS
      IF $OF_{TS}$ is better than $OF_{base}$ THEN
         Set $OF_{base} = OF_{TS}$
         Set $DV_{base} = TS$
         Set $DV_{delta} = 1.2*DV_{delta}$
      ELSE (meaning that the OF is worse or a constraint is hit)
         Set $DV_{delta} = -0.5*DV_{delta}$
      ENDIF
      IF $|DV_{delta}| <$ convergence threshold EXIT
   Return to a new TS (Go to Point 2)

A range of $DV_{delta}$ attenuation factors seem to work. The 1.2 expansion factor could be from 1.05 to 1.25. If larger than 1.25, it seems to excessively enlarge the step size when going in the right direction, overstepping the minimum, and requiring additional back steps to temper the excess. Alternately, an expansion factor that is closer to unity does not accelerate the progress after a contraction. A value of 1.1–1.2 seems to work well. Similarly, there is a range of good values for the contraction factor. If the new TS is bad, then the contraction should be less than 1/expansion so that the step does not go on the other side of the prior best. However, if the contraction factor is too small, for instance, 0.1, then after a reversal, subsequent steps in the right direction are very small. Contraction values in the 0.5–0.8/expansion range seem good.

In addition to not requiring analytical derivatives, this simple procedure is easy to understand and implement (either in code or intuitively), is robust to hard constraints (or execution errors) and surface discontinuities (such as those related to numerical approximation of DV discretization), and executes rapidly.

Notice that contrasting the analytical method of the prior examples, this procedure does not intelligently jump to the zero derivative spot, where $DV^*$ is calculated from the analytical derivative. This iterative procedure incrementally moves the TS toward the vicinity of $DV^*$ and stops, and claims convergence, when the incremental changes are less than a user-defined convergence value. Like incrementally calculating the next number in the irrational value of $\pi$, this procedure gets progressively closer and stops when the user has defined that the TS is close enough—when $|DV_{\text{delta}}| <$ convergence threshold.

The user's choice of the convergence threshold defines the precision of the solution. Somehow the user must relate the impact of the threshold to the undesirables. If too large a threshold, the solution is not very precise. If too small, the procedure requires excessive iterations.

## 1.14 Takeaway

Optimization is essential in all aspects of your personal and professional life.

Understand the terminology in this chapter. The concepts are essential to doing.

Do, not study: Apply the terminology and canonical form of the statement to everything you can. Practice seeing your reality from the structure of this optimization perspective.

The first solution is probably inadequate. Sketch the procedure and solution, evaluate its elements and outcome with respect to your developing understanding and other's input, erase your solution, re-sketch, evaluate, and erase, sketch, eval., erase, s, e, e, s, e, e, … See. See? Eventually you'll have a comprehensive solution. The solution is not the intellectual exercise that produces a value. It is the right solution as defined by all constituents.

## 1.15 Exercises

**1** Qualitatively discuss issues that would relate to these optimization applications:
- **A** How wide should a hat brim be?
- **B** What is the best temperature for a freezer?
- **C** What is the optimum length for a vacuum cleaner cord?
- **D** Where should you cross the street?
- **E** Should you password protect your cell phone?
- **F** How thick should be a pane of glass for a window?
- **G** What are the criteria to open your umbrella?

**2** State OF, DV, and constraints for the following purposely vague situation statements:
- **A** Increasing the coating thickness on pills inhibits $O_2$ diffusion in, which increases shelf-life.
- **B** The larger the pipe diameter, the smaller the $\Delta P$ required to achieve target flow rate—hence the smaller the pump capital and energy cost and expense.

**C** A set of $y$ versus $x$ data, in general, shows curvature, but it could be that separate linear fits in the large $x$ and small $x$ regions will give as good a fit as a comprehensive cubic model.
**D** If a cab breaks, the driver can't work. If the company has extra cabs, then when one or several are in the shop the drivers can still generate revenue.
**E** There is a temperature control knob in the refrigerator.
**F** How much to eat at each meal, what time to have meals, and how many meals per day?
**G** Boiling campsite water to purify it.
**H** The number of tasks to accept in a job (or courses in a semester).
**I** Layout of a resume.
**J** Length of an electric cord for a home-use vacuum cleaner.
**K** Dimension of the packaging of a multi-pack of soft drink cans (12, 15, 18, 24, 36, … cans in a pack?) (side by side, layered, hexagonally or rectangularly packed?)
**L** Thickness of newspaper.
**M** Mathematical model of a power series type.
**N** Game strategy.
**O** Path to get from house to class.
**P** Method to raise children.
**Q** Orientation of a house.
**R** Type of computer.
**S** Menu for a week of home-prepared meals.
**T** Green-on time of a traffic light at an intersection.
**U** Speed limit on a road.
**V** Number of elevators in a building.
**W** Gasoline contains ethanol.
**X** Passenger jet flying speed.
**Y** Designing a bicycle.
**Z** Driving a bicycle.

3. Use the analytical technique, $dJ/d\text{DV}|_{\text{DV}^*} = 0$, to determine the value of $x$ that maximizes $y = 8x - 2x^2 + 4$.

4. Use the analytical technique, $dJ/d\text{DV}|_{\text{DV}^*} = 0$, to determine the value of $x$ that maximizes $y = \cos(2\pi x/10)$ over the range $0 \le x \le 10$.

5. Use the analytical technique, $dJ/d\text{DV}|_{\text{DV}^*} = 0$, to determine the value of $x$ that maximizes $y = xe^{-x}$.

6. Minimize this function of $D$. You choose numerical values for coefficients ($a > 0$ and $b > 0$), $J = aD^{-4} + bD^{+2}$. Use the analytical technique, $dJ/d\text{DV}|_{\text{DV}^*} = 0$, to solve for $D^*$. Since you can specify coefficient values, there are an infinite number of right answers. How do you know that your $D^*$ is the right value?

7. Choose three $(x, y)$ data pairs, and generate a linear model that best fits the data in a least squares sense. For each of the $i = 1,2,3$ pairs, the equation $y_i = a + bx_i$ is true. Use the analytical technique $dJ/d\text{DV}|_{\text{DV}^*} = 0$ to solve for the $\{a, b\}$ coefficient values. Here, the OF is the sum of squared deviations between model and data. Because there are two decision variables, there are two $dJ/d\text{DV}|_{\text{DV}^*} = 0$ equations, termed the normal equations. They will be linear in coefficients

$\{a, b\}$. Solve them simultaneously. Show each stage of your process. To verify that you have gotten the right answer, compare your answer to any automated method of curve fitting (such as the trend line function in Excel charts).

**8** A person goes to bed at midnight and has to be on the job at 8 : 00 am the next day. Before going to bed she makes a decision about the time, $T$ (hour), for the alarm clock to ring. It takes 30 min to travel to work. The longer she sleeps, the more sleep benefit she gets. Benefit = $1 - e^{-T/3}$. However, the longer she sleeps the greater is the anxiety about, and potential cost associated with, being late to work. Badness = $1/(7.5 - T)$. Graph these to show a qualitative relation of benefit versus hours of sleep and the cost of wake-up time. Additively combining these multiple and competing objectives as a single OF, $J = \left(1 - e^{-T/3}\right) - 1/(8 - T)$. Use the analytical technique, $dJ/d\mathrm{DV}|_{\mathrm{DV}^*} = 0$, to solve for the best wake-up time.

**9** As a batch process ages over several days, it makes a higher yield of desired product, but there is a diminishing returns approach of product made to an asymptotic limit. $M(t)$ is the mass of material produced after a batch time $t$, in days, and a simple diminishing returns equation is $M(t) = k\left(1 - e^{-(t/\tau)}\right)$. When the batch process is terminated, the vessel is emptied to recover product, filled with fresh feed, and a new batch is restarted. The turnaround time for emptying and recharging is $\theta$. The process owner wants to maximize annual production of the desired product. Letting the batch run longer increases the yield per batch, but it means fewer batches per year. How long should each batch run to maximize annual productivity of the vessel? Set up the optimization statement. Clearly indicate OF, DV, and the equations (models) that relate DV to other elements. Use the analytical technique, $dJ/d\mathrm{DV}|_{\mathrm{DV}^*} = 0$, to solve for the best batch duration.

**10** A right circular open tank of height, $h$, and radius, $r$, has volume $\pi r^2 h$ and surface area $\pi r^2 + 2\pi rh$. The cost of a tank is related to the amount of material in the surface of thickness, $t$, density, $\rho$, and cost per unit mass, $c$. Cost = $ct\rho(\pi r^2 + 2\pi rh)$. What dimensions, $r$ and $h$, minimize cost while meeting a specified volume, $V$? State the optimization in standard form. Explain simplifications that you are using. Use the analytical technique, $dJ/d\mathrm{DV}|_{\mathrm{DV}^*} = 0$, to solve the simplified application for $\{r^*, h^*\}$.

**11** Use the analytical technique, $dJ/d\mathrm{DV}|_{\mathrm{DV}^*} = 0$, to show that if $x^*$ is the optimum DV value for $f(x)$, it is also the optimum value for $a + f(x)$.

**12** Implement the heuristic direct search method on any of Exercises 3–11.

**13** Provide an example of each for the following terms from your personal life in the past day, or so, sufficiently and explicitly, but briefly explain so that a reader can clearly understand. Reveal that you understand the terms. But don't reveal anything improper! (i) State an OF and one or more DVs. (ii) State your algorithm for determining the Optimum. (iii) State your stopping criterion (or criteria).

**14** State whether each of these is, or is not, a DV. In case it could be either, clearly explain your reason. Use normal situations: (i) outside temperature, (ii) oven temperature, (iii) fever temperature, (iv) car temperature, and (v) house temperature.

**15** For each model, state whether it is linear or nonlinear in coefficients ($a$, $b$, $c$, etc. that need to be valued in regression modeling) *and* linear or nonlinear in $y(\boldsymbol{x})$ functionality (for finding $\boldsymbol{x}^*$ to optimize $y$). $y = a*\sin(x)$, $y = a*\sin(x + b)$, $y = a + bx_1 + cx_1x_2$, $y = a + bx_1 + cx_2^d$, $y = a + b/x$.

**16** If you were to optimize a written sentence, what would you choose as the DVs, OF, and Constraints? Just list the terms or concepts, not equations.

**17** The volume of a rectangular block of length ($l$), width ($w$), and height ($h$) is $V = lwh$. Its surface area is $S = 2(lw + lh + hw)$. Set up the optimization statement to determine the block dimensions ($l$, $w$, $h$) that meet a particular volume requirement ($V_o$) yet minimize surface area.

**18** Consider optimization of the number of holes in a shoe for the shoelaces and state an OF, the DV, and a constraint.

**19** Consider optimization of a parasol (an umbrella to keep sunlight off a person). Is the handle color an OF, a DV, or a constraint?

**20** Why use optimization algorithms to find the minimum when you could plot the function and see directly where it is?

**21** Here is an equation relating an independent variable, $x$, to the dependent variable, $y$. $y = a + bx + cx^2$. In an optimization application, when would: "$b$" be a DV? "$x$" be a DV?

**22** Reveal that you understand optimization concepts by describing an OF, an associated DV, and a constraint for each of the following. Provide brief and elementary answers.

**A** Consider that you are *designing* a bicycle. State an OF, an associated DV, and a constraint.
**B** Consider that you are *driving* a bicycle. State an OF, an associated DV, and a constraint.
**C** Consider that you are *designing* a better leaf for a tree. State an OF, an associated DV, and a constraint.
**D** Consider that you are *driving* a car. State an OF, an associated DV, and a constraint.

**23** Something decided the thickness of newspaper. It is thicker than tissue paper, but thinner than writing paper. List several of the issues that would guide selecting an optimal thickness value.

**24** Show that in minimizing the sum of squared deviations of data from a common value, $\min_{\{a\}} J = \sum_{i=1}^{N}(x_i - a)^2$, the common value becomes the average, $a = (1/N)\sum x_i$. Use the analytical method to solve the optimization application for the DV value.

**25** True or false? Explain.

**A** ___ The optimizer calculates the OF value.
**B** ___ The convergence threshold affects proximity to the true DV*.
**C** ___ If three optimizer runs from randomized starts find the same DV*, that is the global.
**D** ___ Education should have students work in the knowledge, but not the evaluation level.

**E** ___ The development of the OF models for an application requires engineering skill.
**F** ___ The convergence type must be related to the DV to find DV*.

**26** For two optimization examples, one from your personal life and another from work/research, describe with precision the OF, DV(s), and constraint(s). In this exercise you will describe two optimization applications representing diverse aspects of your experience. For each, state a plausible choice for the OF, DV(s), and constraint(s). Explain why the ones you chose were "right" for your situation. "Describe with precision" means you should provide equations that would quantify how the OF value is calculated, unambiguous statements of quantifying the DV(s) and constraint(s), and units on the OF and DV. For example, "amount of food" as a DV is not as precise as "calorie intake per day." Also, for example, "quietness" of the automobile ride is not a quantifiable OF, but "microphone dB could be." And, for example, "don't place too much in the tank" is not as precise a constraint statement as "the tank contents cannot exceed 55 gal." In a text statement of your objective, you may use value-laden words like good, best, works, smooth, large, fits, pleases, etc. But your OF, DV, and constraint descriptions should not have any subjective, intuitive, or ambiguous meaning.

**27** Return benefit from effort has a diminishing returns response. There is also a threshold effort before there is any return. Consider building a piece of furniture. Effort may start with buying the wood, stain and varnish, and sharpening and aligning the cutting tools. This effort is necessary, but so far there is nothing to sell; there is no benefit to the carpenter. After assembly, light sanding and one coat of finish may make a salable item; but more sanding and another coat of finish makes it an excellent item, increases perceived value to the buyer, and willingness to pay a higher price. Additional carpenter effort beyond excellence moves the item toward perfection, but doesn't increase value over something already excellent. Often an S-shaped relation is used to relate value to effort. The logistic relation is one, $V = a/\left[1 + e^{-b(E-c)}\right]$, where "$E$" represents the time or effort invested and "$V$" the value. You probably have 20 things that you are working on and only 100% effort to give. If you work totally, you don't eat or sleep, which are necessary functions to be able to invest energy in any project. As you personally seek to optimize life outcome, what are DVs, OF, and constraints? For simplicity and clarity, only consider three projects.