# - <br> An Introduction to Model Building 

### 1.1 An Introduction to Modeling

Operations research (often referred to as management science) is simply a scientific approach to decision making that seeks to best design and operate a system, usually under conditions requiring the allocation of scarce resources.

By a system, we mean an organization of interdependent components that work together to accomplish the goal of the system. For example, Ford Motor Company is a system whose goal consists of maximizing the profit that can be earned by producing quality vehicles.

The term operations research was coined during World War II when British military leaders asked scientists and engineers to analyze several military problems such as the deployment of radar and the management of convoy, bombing, antisubmarine, and mining operations.

The scientific approach to decision making usually involves the use of one or more mathematical models. A mathematical model is a mathematical representation of an actual situation that may be used to make better decisions or simply to understand the actual situation better. The following example should clarify many of the key terms used to describe mathematical models.

Eli Daisy produces Wozac in huge batches by heating a chemical mixture in a pressurized container. Each time a batch is processed, a different amount of Wozac is produced. The amount produced is the process yield (measured in pounds). Daisy is interested in understanding the factors that influence the yield of the Wozac production process. Describe a model-building process for this situation.

Solution Daisy is first interested in determining the factors that influence the yield of the process. This would be referred to as a descriptive model, because it describes the behavior of the actual yield as a function of various factors. Daisy might determine (using regression methods discussed in Chapter 24) that the following factors influence yield:

- container volume in liters (V)
- container pressure in milliliters (P)
- container temperature in degrees Celsius (T)
- chemical composition of the processed mixture

If we let $A, B$, and $C$ be percentage of mixture made up of chemicals $A, B$, and $C$, then Daisy might find, for example, that
(1) yield $=300+.8 \mathrm{~V}+.01 \mathrm{P}+.06 \mathrm{~T}+.001 \mathrm{~T} * \mathrm{P}-.01 \mathrm{~T}^{2}-.001 \mathrm{P}^{2}$

$$
+11.7 \mathrm{~A}+9.4 \mathrm{~B}+16.4 \mathrm{C}+19 \mathrm{~A} * \mathrm{~B}+11.4 \mathrm{~A} * \mathrm{C}-9.6 \mathrm{~B} * \mathrm{C}
$$

To determine this relationship, the yield of the process would have to be measured for many different combinations of the previously listed factors. Knowledge of this equation would enable Daisy to describe the yield of the production process once volume, pressure, temperature, and chemical composition were known.

## Prescriptive or Optimization Models

Most of the models discussed in this book will be prescriptive or optimization models. A prescriptive model "prescribes" behavior for an organization that will enable it to best meet its goal(s). The components of a prescriptive model include

- objective function(s)
- decision variables
- constraints

In short, an optimization model seeks to find values of the decision variables that optimize (maximize or minimize) an objective function among the set of all values for the decision variables that satisfy the given constraints.

## The Objective Function

Naturally, Daisy would like to maximize the yield of the process. In most models, there will be a function we wish to maximize or minimize. This function is called the model's objective function. Of course, to maximize the process yield we need to find the values of $\mathrm{V}, \mathrm{P}, \mathrm{T}, \mathrm{A}, \mathrm{B}$, and C that make (1) as large as possible.

In many situations, an organization may have more than one objective. For example, in assigning students to the two high schools in Bloomington, Indiana, the Monroe County School Board stated that the assignment of students involved the following objectives:

- Equalize the number of students at the two high schools.
- Minimize the average distance students travel to school.
- Have a diverse student body at both high schools.

Multiple objective decision-making problems are discussed in Sections 4.14 and 11.13.

## The Decision Variables

The variables whose values are under our control and influence the performance of the system are called decision variables. In our example, V, P, T, A, B, and C are decision variables. Most of this book will be devoted to a discussion of how to determine the value of decision variables that maximize (sometimes minimize) an objective function.

## Constraints

In most situations, only certain values of decision variables are possible. For example, certain volume, pressure, and temperature combinations might be unsafe. Also, A B, and C must be nonnegative numbers that add to 1 . Restrictions on the values of decision variables are called constraints. Suppose the following:

- Volume must be between 1 and 5 liters.

■ Pressure must be between 200 and 400 milliliters.

- Temperature must be between 100 and 200 degrees Celsius.
- Mixture must be made up entirely of A, B, and C.
- For the drug to properly perform, only half the mixture at most can be product A.

These constraints can be expressed mathematically by the following constraints:

$$
\begin{aligned}
\mathrm{V} & \leq 5 \\
\mathrm{~V} & \geq 1 \\
\mathrm{P} & \leq 400 \\
\mathrm{P} & \geq 200 \\
\mathrm{~T} & \leq 200 \\
\mathrm{~T} & \geq 100 \\
\mathrm{~A} & \geq 0 \\
\mathrm{~B} & \geq 0 \\
\mathrm{~A}+\mathrm{B}+\mathrm{C} & =1 \\
\mathrm{~A} & \leq 5
\end{aligned}
$$

## The Complete Optimization Model

After letting $z$ represent the value of the objective function, our entire optimization model may be written as follows:

$$
\begin{aligned}
\text { Maximize } z=300+.8 \mathrm{~V} & +.01 \mathrm{P}+.06 \mathrm{~T}+.001 \mathrm{~T} * \mathrm{P}-.01 \mathrm{~T}^{2}-.001 \mathrm{P}^{2} \\
& +11.7 \mathrm{~A}+9.4 \mathrm{~B}+16.4 \mathrm{C}+19 \mathrm{~A} * \mathrm{~B}+11.4 \mathrm{~A} * \mathrm{C}-9.6 \mathrm{~B} * \mathrm{C}
\end{aligned}
$$

Subject to (s.t.)

$$
\begin{aligned}
\mathrm{V} & \leq 5 \\
\mathrm{~V} & \geq 1 \\
\mathrm{P} & \leq 400 \\
\mathrm{P} & \geq 200 \\
\mathrm{~T} & \leq 200 \\
\mathrm{~T} & \geq 100 \\
\mathrm{~A} & \geq 0 \\
\mathrm{~B} & \geq 0 \\
\mathrm{C} & \geq 0 \\
\mathrm{~A}+\mathrm{B}+\mathrm{C} & =1 \\
\mathrm{~A} & \leq 5
\end{aligned}
$$

Any specification of the decision variables that satisfies all of the model's constraints is said to be in the feasible region. For example, $\mathrm{V}=2, \mathrm{P}=300, \mathrm{~T}=150, \mathrm{~A}=.4, \mathrm{~B}=$ .3 , and $\mathrm{C}=.1$ is in the feasible region. An optimal solution to an optimization model is any point in the feasible region that optimizes (in this case, maximizes) the objective function. Using the LINGO package that comes with this book, it can be determined that the optimal solution to this model is $\mathrm{V}=5, \mathrm{P}=200, \mathrm{~T}=100, \mathrm{~A}=.294, \mathrm{~B}=0, \mathrm{C}=.706$, and $z=183.38$. Thus, a maximum yield of 183.38 pounds can be obtained with a 5 -liter
container, pressure of 200 milliliters, temperature of 100 degrees Celsius, and $29 \% \mathrm{~A}$ and $71 \% \mathrm{C}$. This means no other feasible combination of decision variables can obtain a yield exceeding 183.38 pounds.

## Static and Dynamic Models

A static model is one in which the decision variables do not involve sequences of decisions over multiple periods. A dynamic model is a model in which the decision variables do involve sequences of decisions over multiple periods. Basically, in a static model we solve a "one-shot" problem whose solutions prescribe optimal values of decision variables at all points in time. Example 1 is an example of a static model; the optimal solution will tell Daisy how to maximize yield at all points in time.

For an example of a dynamic model, consider a company (call it Sailco) that must determine how to minimize the cost of meeting (on time) the demand for sailboats during the next year. Clearly Sailco's must determine how many sailboats it will produce during each of the next four quarters. Sailco's decisions involve decisions made over multiple periods, hence a model of Sailco's problem (see Section 3.10) would be a dynamic model.

## Linear and Nonlinear Models

Suppose that whenever decision variables appear in the objective function and in the constraints of an optimization model, the decision variables are always multiplied by constants and added together. Such a model is a linear model. If an optimization model is not linear, then it is a nonlinear model. In the constraints of Example 1, the decision variables are always multiplied by constants and added together. Thus, Example 1's constraints pass the test for a linear model. However, in the objective function for Example 1, the terms $.001 \mathrm{~T}^{*} \mathrm{P},-.01 \mathrm{~T}^{2}, 19 \mathrm{~A} * \mathrm{~B}, 11.4 \mathrm{~A} * \mathrm{C}$, and $-9.6 \mathrm{~B} * \mathrm{C}$ make the model nonlinear. In general, nonlinear models are much harder to solve than linear models. We will discuss linear models in Chapters 2 through 10. Nonlinear models will be discussed in Chapter 11.

## Integer and Noninteger Models

If one or more decision variables must be integer, then we say that an optimization model is an integer model. If all the decision variables are free to assume fractional values, then the optimization model is a noninteger model. Clearly, volume, temperature, pressure, and percentage composition of our inputs may all assume fractional values. Thus, Example 1 is a noninteger model. If the decision variables in a model represent the number of workers starting work during each shift at a fast-food restaurant, then clearly we have an integer model. Integer models are much harder to solve than nonlinear models. They will be discussed in detail in Chapter 9.

## Deterministic and Stochastic Models

Suppose that for any value of the decision variables, the value of the objective function and whether or not the constraints are satisfied is known with certainty. We then have a deterministic model. If this is not the case, then we have a stochastic model. All models in the first 12 chapters will be deterministic models. Stochastic models are covered in Chapters 13, 16, 17, and 19-24.

If we view Example 1 as a deterministic model, then we are making the (unrealistic) assumption that for given values of $\mathrm{V}, \mathrm{P}, \mathrm{T}, \mathrm{A}, \mathrm{B}$, and C , the process yield will always be the same. This is highly unlikely. We can view (1) as a representation of the average yield of the process for given values of the decision variables. Then our objective is to find values of the decision variables that maximize the average yield of the process.

We can often gain useful insights into optimal decisions by using a deterministic model in a situation where a stochastic model is more appropriate. Consider Sailco's problem of minimizing the cost of meeting the demand (on time) for sailboats. The uncertainty about future demand for sailboats implies that for a given production schedule, we do not know whether demand is met on time. This leads us to believe that a stochastic model is needed to model Sailco's situation. We will see in Section 3.10, however, that we can develop a deterministic model for this situation that yields good decisions for Sailco.

### 1.2 The Seven-Step Model-Building Process

When operations research is used to solve an organization's problem, the following sevenstep model-building procedure should be followed:

Step 1: Formulate the Problem The operations researcher first defines the organization's problem. Defining the problem includes specifying the organization's objectives and the parts of the organization that must be studied before the problem can be solved. In Example 1, the problem was to determine how to maximize the yield from a batch of Wozac.

Step 2: Observe the System Next, the operations researcher collects data to estimate the value of parameters that affect the organization's problem. These estimates are used to develop (in step 3) and evaluate (in step 4) a mathematical model of the organization's problem. For example, in Example 1, data would be collected in an attempt to determine how the values of T, P, V, A, B, and C influence process yield.

Step 3: Formulate a Mathematical Model of the Problem In this step, the operations researcher develops a mathematical model of the problem. In this book, we will describe many mathematical techniques that can be used to model systems. For Example 1, our optimization model would be the result of step 3 .

Step 4: Verify the Model and Use the Model for Prediction The operations researcher now tries to determine if the mathematical model developed in step 3 is an accurate representation of reality. For example, to validate our model, we might check and see if (1) accurately represents yield for values of the decision variables that were not used to estimate (1). Even if a model is valid for the current situation, we must be aware of blindly applying it. For example, if the government placed new restrictions on Wozac, then we might have to add new constraints to our model, and the yield of the process [and Equation (1)] might change.

Step 5: Select a Suitable Alternative Given a model and a set of alternatives, the operations researcher now chooses the alternative that best meets the organization's objectives. (There may be more than one!) For instance, our model enabled us to determine that yield was maximized with $\mathrm{V}=5, \mathrm{P}=200, \mathrm{~T}=100, \mathrm{~A}=.294, \mathrm{~B}=0, \mathrm{C}=.706$, and $z=$ 183.38.

Step 6: Present the Results and Conclusion of the Study to the Organization In this step, the operations researcher presents the model and recommendation from step 5 to the decisionmaking individual or group. In some situations, one might present several alternatives and let the organization choose the one that best meets its needs. After presenting the results
of the operations research study, the analyst may find that the organization does not approve of the recommendation. This may result from incorrect definition of the organization's problems or from failure to involve the decision maker from the start of the project. In this case, the operations researcher should return to step 1,2 , or 3 .

Step 7: Implement and Evaluate Recommendations If the organization has accepted the study, then the analyst aids in implementing the recommendations. The system must be constantly monitored (and updated dynamically as the environment changes) to ensure that the recommendations enable the organization to meet its objectives.

In what follows, we discuss three successful management science applications. We will give a detailed (but nonquantitative) description of each application. We will tie our discussion of each application to the seven-step model-building process described in Section 1.2.

### 1.3 CITGO Petroleum

Klingman et al. (1987) applied a variety of management-science techniques to CITGO Petroleum. Their work saved the company an estimated $\$ 70$ million per year. CITGO is an oil-refining and -marketing company that was purchased by Southland Corporation (the owners of the 7-Eleven stores). We will focus on two aspects of the CITGO team's work:

1 a mathematical model to optimize operation of CITGO's refineries, and
2 a mathematical model—supply distribution marketing (SDM) system-that was used to develop an 11-week supply, distribution, and marketing plan for the entire business.

## Optimizing Refinery Operations

Step 1 Klingman et al. wanted to minimize the cost of operating CITGO's refineries.
Step 2 The Lake Charles, Louisiana, refinery was closely observed in an attempt to estimate key relationships such as:

1 How the cost of producing each of CITGO's products (motor fuel, no. 2 fuel oil, turbine fuel, naptha, and several blended motor fuels) depends on the inputs used to produce each product.

2 The amount of energy needed to produce each product. This required the installation of a new metering system.

3 The yield associated with each input-output combination. For example, if 1 gallon of crude oil would yield . 52 gallons of motor fuel, then the yield would equal $52 \%$.

4 To reduce maintenance costs, data were collected on parts inventories and equipment breakdowns. Obtaining accurate data required the installation of a new database-management system and integrated maintenance-information system. A process control system was also installed to accurately monitor the inputs and resources used to manufacture each product.

Step 3 Using linear programming (LP), a model was developed to optimize refinery operations. The model determines the cost-minimizing method for mixing or blending together inputs to produce desired outputs. The model contains constraints that ensure that inputs are blended so that each output is of the desired quality. Blending constraints are discussed in Section 3.8. The model ensures that plant capacities are not exceeded and al-
lows for the fact that each refinery may carry an inventory of each end product. Sections 3.10 and 4.12 discuss inventory constraints.

Step 4 To validate the model, inputs and outputs from the Lake Charles refinery were collected for one month. Given the actual inputs used at the refinery during that month, the actual outputs were compared to those predicted by the model. After extensive changes, the model's predicted outputs were close to the actual outputs.

Step 5 Running the LP yielded a daily strategy for running the refinery. For instance, the model might, say, produce 400,000 gallons of turbine fuel using 300,000 gallons of crude 1 and 200,000 gallons of crude 2.

Steps 6 and 7 Once the database and process control were in place, the model was used to guide day-to-day refinery operations. CITGO estimated that the overall benefits of the refinery system exceeded $\$ 50$ million annually.

## The Supply Distribution Marketing (SDM) System

Step 1 CITGO wanted a mathematical model that could be used to make supply, distribution, and marketing decisions such as:

1 Where should crude oil be purchased?
2 Where should products be sold?
3 What price should be charged for products?
4 How much of each product should be held in inventory?
The goal, of course, was to maximize the profitability associated with these decisions.
Step 2 A database that kept track of sales, inventory, trades, and exchanges of all refined products was installed. Also, regression analysis (see Chapter 24) was used to develop forecasts for wholesale prices and wholesale demand for each CITGO product.

Steps 3 and 5 A minimum-cost network flow model (MCNFM) (see Section 7.4) is used to determine an 11-week supply, marketing, and distribution strategy. The model makes all decisions mentioned in step 1. A typical model run that involved 3,000 equations and 15,000 decision variables required only 30 seconds on an IBM 4381.

Step 4 The forecasting modules are continuously evaluated to ensure that they continue to give accurate forecasts.

Steps 6 and 7 Implementing the SDM required several organizational changes. A new vice-president was appointed to coordinate the operation of the SDM and LP refinery model. The product supply and product scheduling departments were combined to improve communication and information flow.

### 1.4 San Francisco Police Department Scheduling

Taylor and Huxley (1989) developed a police patrol scheduling system (PPSS). All San Francisco (SF) police precincts use PPSS to schedule their officers. It is estimated that PPSS saves the SF police more than $\$ 5$ million annually. Other cities such as Virginia

## Introduction to Linear Programming

Linear programming (LP) is a tool for solving optimization problems. In 1947, George Dantzig developed an efficient method, the simplex algorithm, for solving linear programming problems (also called LP). Since the development of the simplex algorithm, LP has been used to solve optimization problems in industries as diverse as banking, education, forestry, petroleum, and trucking. In a survey of Fortune 500 firms, $85 \%$ of the respondents said they had used linear programming. As a measure of the importance of linear programming in operations research, approximately $70 \%$ of this book will be devoted to linear programming and related optimization techniques.

In Section 3.1, we begin our study of linear programming by describing the general characteristics shared by all linear programming problems. In Sections 3.2 and 3.3, we learn how to solve graphically those linear programming problems that involve only two variables. Solving these simple LPs will give us useful insights for solving more complex LPs. The remainder of the chapter explains how to formulate linear programming models of real-life situations.

### 3.1 What Is a Linear Programming Problem?

In this section, we introduce linear programming and define important terms that are used to describe linear programming problems.

## EXAMPLE $1 \quad$ Glapetto's Woodcarving

Giapetto's Woodcarving, Inc., manufactures two types of wooden toys: soldiers and trains. A soldier sells for $\$ 27$ and uses $\$ 10$ worth of raw materials. Each soldier that is manufactured increases Giapetto's variable labor and overhead costs by $\$ 14$. A train sells for $\$ 21$ and uses $\$ 9$ worth of raw materials. Each train built increases Giapetto's variable labor and overhead costs by $\$ 10$. The manufacture of wooden soldiers and trains requires two types of skilled labor: carpentry and finishing. A soldier requires 2 hours of finishing labor and 1 hour of carpentry labor. A train requires 1 hour of finishing and 1 hour of carpentry labor. Each week, Giapetto can obtain all the needed raw material but only 100 finishing hours and 80 carpentry hours. Demand for trains is unlimited, but at most 40 soldiers are bought each week. Giapetto wants to maximize weekly profit (revenues - costs). Formulate a mathematical model of Giapetto's situation that can be used to maximize Giapetto's weekly profit.

Solution In developing the Giapetto model, we explore characteristics shared by all linear programming problems.

Decision Variables We begin by defining the relevant decision variables. In any linear programming model, the decision variables should completely describe the decisions to be made (in this case, by Giapetto). Clearly, Giapetto must decide how many soldiers and trains should be manufactured each week. With this in mind, we define

$$
\begin{aligned}
& x_{1}=\text { number of soldiers produced each week } \\
& x_{2}=\text { number of trains produced each week }
\end{aligned}
$$

Objective Function In any linear programming problem, the decision maker wants to maximize (usually revenue or profit) or minimize (usually costs) some function of the decision variables. The function to be maximized or minimized is called the objective function. For the Giapetto problem, we note that fixed costs (such as rent and insurance) do not depend on the values of $x_{1}$ and $x_{2}$. Thus, Giapetto can concentrate on maximizing (weekly revenues) - (raw material purchase costs) - (other variable costs).

Giapetto's weekly revenues and costs can be expressed in terms of the decision variables $x_{1}$ and $x_{2}$. It would be foolish for Giapetto to manufacture more soldiers than can be sold, so we assume that all toys produced will be sold. Then

$$
\begin{aligned}
\text { Weekly revenues }= & \text { weekly revenues from soldiers } \\
& + \text { weekly revenues from trains } \\
= & \left(\frac{\text { dollars }}{\text { soldier }}\right)\left(\frac{\text { soldiers }}{\text { week }}\right)+\left(\frac{\text { dollars }}{\text { train }}\right)\left(\frac{\text { trains }}{\text { week }}\right) \\
= & 27 x_{1}+21 x_{2}
\end{aligned}
$$

Also,

$$
\begin{aligned}
\text { Weekly raw material costs } & =10 x_{1}+9 x_{2} \\
\text { Other weekly variable costs } & =14 x_{1}+10 x_{2}
\end{aligned}
$$

Then Giapetto wants to maximize

$$
\left(27 x_{1}+21 x_{2}\right)-\left(10 x_{1}+9 x_{2}\right)-\left(14 x_{1}+10 x_{2}\right)=3 x_{1}+2 x_{2}
$$

Another way to see that Giapetto wants to maximize $3 x_{1}+2 x_{2}$ is to note that
Weekly revenues $=$ weekly contribution to profit from soldiers

- weekly nonfixed costs + weekly contribution to profit from trains

$$
\begin{aligned}
= & \left(\frac{\text { contribution to profit }}{}\right)\left(\frac{\text { soldiers }}{\text { week }}\right) \\
& +\left(\frac{\text { contribution to profit }}{}\right)\left(\frac{\text { trains }}{\text { week }}\right)
\end{aligned}
$$

Also,

$$
\begin{aligned}
& \text { Contribution to profit }=27-10-14=3 \\
& \underline{\text { Contribution to profit }}=21-9-10=2
\end{aligned}
$$

Then, as before, we obtain

$$
\text { Weekly revenues }- \text { weekly nonfixed costs }=3 x_{1}+2 x_{2}
$$

Thus, Giapetto's objective is to choose $x_{1}$ and $x_{2}$ to maximize $3 x_{1}+2 x_{2}$. We use the variable $z$ to denote the objective function value of any LP. Giapetto's objective function is

$$
\begin{equation*}
\text { Maximize } z=3 x_{1}+2 x_{2} \tag{1}
\end{equation*}
$$

(In the future, we will abbreviate "maximize" by max and "minimize" by min.) The coefficient of a variable in the objective function is called the objective function coefficient of the variable. For example, the objective function coefficient for $x_{1}$ is 3 , and the objective function coefficient for $x_{2}$ is 2 . In this example (and in many other problems), the ob-
jective function coefficient for each variable is simply the contribution of the variable to the company's profit.

Constraints As $x_{1}$ and $x_{2}$ increase, Giapetto's objective function grows larger. This means that if Giapetto were free to choose any values for $x_{1}$ and $x_{2}$, the company could make an arbitrarily large profit by choosing $x_{1}$ and $x_{2}$ to be very large. Unfortunately, the values of $x_{1}$ and $x_{2}$ are limited by the following three restrictions (often called constraints):

Constraint 1 Each week, no more than 100 hours of finishing time may be used.
Constraint 2 Each week, no more than 80 hours of carpentry time may be used.
Constraint 3 Because of limited demand, at most 40 soldiers should be produced each week.

The amount of raw material available is assumed to be unlimited, so no restrictions have been placed on this.

The next step in formulating a mathematical model of the Giapetto problem is to express Constraints $1-3$ in terms of the decision variables $x_{1}$ and $x_{2}$. To express Constraint 1 in terms of $x_{1}$ and $x_{2}$, note that

$$
\begin{aligned}
\frac{\text { Total finishing hrs. }}{=} & \left(\frac{\text { finishing hrs. }}{\text { soldier }}\right)\left(\frac{\text { soldiers made }}{\text { week }}\right) \\
& +\left(\frac{\text { finishing hrs. }}{\text { train }}\right)\left(\frac{\text { trains made }}{\text { week }}\right) \\
= & 2\left(x_{1}\right)+1\left(x_{2}\right)=2 x_{1}+x_{2}
\end{aligned}
$$

Now Constraint 1 may be expressed by

$$
\begin{equation*}
2 x_{1}+x_{2} \leq 100 \tag{2}
\end{equation*}
$$

Note that the units of each term in (2) are finishing hours per week. For a constraint to be reasonable, all terms in the constraint must have the same units. Otherwise one is adding apples and oranges, and the constraint won't have any meaning.

To express Constraint 2 in terms of $x_{1}$ and $x_{2}$, note that

$$
\begin{aligned}
& \frac{\text { Total carpentry hrs. }}{}=\left(\frac{\text { carpentry hrs. }}{\text { solider }}\right)\left(\frac{\text { soldiers }}{\text { week }}\right) \\
&+\left(\frac{\text { carpentry hrs. }}{\text { train }}\right)\left(\frac{\text { trains }}{\text { week }}\right) \\
&=1\left(x_{1}\right)+1\left(x_{2}\right)=x_{1}+x_{2}
\end{aligned}
$$

Then Constraint 2 may be written as

$$
\begin{equation*}
x_{1}+x_{2} \leq 80 \tag{3}
\end{equation*}
$$

Again, note that the units of each term in (3) are the same (in this case, carpentry hours per week).

Finally, we express the fact that at most 40 soldiers per week can be sold by limiting the weekly production of soldiers to at most 40 soldiers. This yields the following constraint:

$$
\begin{equation*}
x_{1} \leq 40 \tag{4}
\end{equation*}
$$

Thus (2)-(4) express Constraints $1-3$ in terms of the decision variables; they are called the constraints for the Giapetto linear programming problem. The coefficients of the decision variables in the constraints are called technological coefficients. This is because the technological coefficients often reflect the technology used to produce different products. For example, the technological coefficient of $x_{2}$ in (3) is 1 , indicating that a soldier requires 1 carpentry hour. The number on the right-hand side of each constraint is called
the constraint's right-hand side (or rhs). Often the rhs of a constraint represents the quantity of a resource that is available.

Sign Restrictions To complete the formulation of a linear programming problem, the following question must be answered for each decision variable: Can the decision variable only assume nonnegative values, or is the decision variable allowed to assume both positive and negative values?

If a decision variable $x_{i}$ can only assume nonnegative values, then we add the sign restriction $x_{i} \geq 0$. If a variable $x_{i}$ can assume both positive and negative (or zero) values, then we say that $x_{i}$ is unrestricted in sign (often abbreviated urs). For the Giapetto problem, it is clear that $x_{1} \geq 0$ and $x_{2} \geq 0$. In other problems, however, some variables may be urs. For example, if $x_{i}$ represented a firm's cash balance, then $x_{i}$ could be considered negative if the firm owed more money than it had on hand. In this case, it would be appropriate to classify $x_{i}$ as urs. Other uses of urs variables are discussed in Section 4.12.

Combining the sign restrictions $x_{1} \geq 0$ and $x_{2} \geq 0$ with the objective function (1) and Constraints (2)-(4) yields the following optimization model:

$$
\begin{equation*}
\max z=3 x_{1}+2 x_{2} \quad \text { (Objective function) } \tag{1}
\end{equation*}
$$

subject to (s.t.)

$$
\begin{align*}
2 x_{1}+x_{2} & \leq 100 & & (\text { Finishing constraint }  \tag{2}\\
x_{1}+x_{2} & \leq 80 & & (\text { Carpentry constraint }  \tag{3}\\
x_{1} & \leq 40 & & (\text { Constraint on demand for soldiers) }  \tag{4}\\
x_{1} & \geq 0 & & (\text { Sign restriction })^{\dagger}  \tag{5}\\
x_{2} & \geq 0 & & (\text { Sign restriction }) \tag{6}
\end{align*}
$$

"Subject to" (s.t.) means that the values of the decision variables $x_{1}$ and $x_{2}$ must satisfy all constraints and all sign restrictions.

Before formally defining a linear programming problem, we define the concepts of linear function and linear inequality.

```
DEFINITION - A function }f(\mp@subsup{x}{1}{},\mp@subsup{x}{2}{},\ldots,\mp@subsup{x}{n}{})\mathrm{ of }\mp@subsup{x}{1}{},\mp@subsup{x}{2}{},\ldots,\mp@subsup{x}{n}{}\mathrm{ is a linear function if and only if
    for some set of constants c}\mp@subsup{c}{1}{},\mp@subsup{c}{2}{},\ldots,\mp@subsup{c}{n}{},f(\mp@subsup{x}{1}{},\mp@subsup{x}{2}{},\ldots,\mp@subsup{x}{n}{})=\mp@subsup{c}{1}{}\mp@subsup{x}{1}{}+\mp@subsup{c}{2}{}\mp@subsup{x}{2}{}+\cdots
    c}\mp@subsup{c}{n}{}\mp@subsup{x}{n}{}.\quad
```

For example, $f\left(x_{1}, x_{2}\right)=2 x_{1}+x_{2}$ is a linear function of $x_{1}$ and $x_{2}$, but $f\left(x_{1}, x_{2}\right)=x_{1}^{2} x_{2}$ is not a linear function of $x_{1}$ and $x_{2}$.

DEFINITION - For any linear function $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ and any number $b$, the inequalities $f\left(x_{1}\right.$, $\left.x_{2}, \ldots, x_{n}\right) \leq b$ and $f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \geq b$ are linear inequalities.

Thus, $2 x_{1}+3 x_{2} \leq 3$ and $2 x_{1}+x_{2} \geq 3$ are linear inequalities, but $x_{1}^{2} x_{2} \geq 3$ is not a linear inequality.

[^0]DEFINITION ■ A linear programming problem (LP) is an optimization problem for which we do the following:

1 We attempt to maximize (or minimize) a linear function of the decision variables. The function that is to be maximized or minimized is called the objective function.

2 The values of the decision variables must satisfy a set of constraints. Each constraint must be a linear equation or linear inequality.

3 A sign restriction is associated with each variable. For any variable $x_{i}$, the sign restriction specifies that $x_{i}$ must be either nonnegative ( $x_{i} \geq 0$ ) or unrestricted in sign (urs).

Because Giapetto's objective function is a linear function of $x_{1}$ and $x_{2}$, and all of Giapetto's constraints are linear inequalities, the Giapetto problem is a linear programming problem. Note that the Giapetto problem is typical of a wide class of linear programming problems in which a decision maker's goal is to maximize profit subject to limited resources.

## The Proportionality and Additivity Assumptions

The fact that the objective function for an LP must be a linear function of the decision variables has two implications.

1 The contribution of the objective function from each decision variable is proportional to the value of the decision variable. For example, the contribution to the objective function from making four soldiers $(4 \times 3=\$ 12)$ is exactly four times the contribution to the objective function from making one soldier (\$3).

2 The contribution to the objective function for any variable is independent of the values of the other decision variables. For example, no matter what the value of $x_{2}$, the manufacture of $x_{1}$ soldiers will always contribute $3 x_{1}$ dollars to the objective function.

Analogously, the fact that each LP constraint must be a linear inequality or linear equation has two implications.

1 The contribution of each variable to the left-hand side of each constraint is proportional to the value of the variable. For example, it takes exactly three times as many finishing hours $(2 \times 3=6$ finishing hours) to manufacture three soldiers as it takes to manufacture one soldier ( 2 finishing hours).

2 The contribution of a variable to the left-hand side of each constraint is independent of the values of the variable. For example, no matter what the value of $x_{1}$, the manufacture of $x_{2}$ trains uses $x_{2}$ finishing hours and $x_{2}$ carpentry hours.

The first implication given in each list is called the Proportionality Assumption of Linear Programming. Implication 2 of the first list implies that the value of the objective function is the sum of the contributions from individual variables, and implication 2 of the second list implies that the left-hand side of each constraint is the sum of the contributions from each variable. For this reason, the second implication in each list is called the Additivity Assumption of Linear Programming.

For an LP to be an appropriate representation of a real-life situation, the decision variables must satisfy both the Proportionality and Additivity Assumptions. Two other assumptions must also be satisfied before an LP can appropriately represent a real situation: the Divisibility and Certainty Assumptions.

## The Divisibility Assumption

The Divisibility Assumption requires that each decision variable be allowed to assume fractional values. For example, in the Giapetto problem, the Divisibility Assumption implies that it is acceptable to produce 1.5 soldiers or 1.63 trains. Because Giapetto cannot actually produce a fractional number of trains or soldiers, the Divisibility Assumption is not satisfied in the Giapetto problem. A linear programming problem in which some or all of the variables must be nonnegative integers is called an integer programming problem. The solution of integer programming problems is discussed in Chapter 9.

In many situations where divisibility is not present, rounding off each variable in the optimal LP solution to an integer may yield a reasonable solution. Suppose the optimal solution to an LP stated that an auto company should manufacture $150,000.4$ compact cars during the current year. In this case, you could tell the auto company to manufacture 150,000 or 150,001 compact cars and be fairly confident that this would reasonably approximate an optimal production plan. On the other hand, if the number of missile sites that the United States should use were a variable in an LP and the optimal LP solution said that 0.4 missile sites should be built, it would make a big difference whether we rounded the number of missile sites down to 0 or up to 1 . In this situation, the integer programming methods of Chapter 9 would have to be used, because the number of missile sites is definitely not divisible.

## The Certainty Assumption

The Certainty Assumption is that each parameter (objective function coefficient, righthand side, and technological coefficient) is known with certainty. If we were unsure of the exact amount of carpentry and finishing hours required to build a train, the Certainty Assumption would be violated.

## Feasible Region and Optimal Solution

Two of the most basic concepts associated with a linear programming problem are feasible region and optimal solution. For defining these concepts, we use the term point to mean a specification of the value for each decision variable.

## DEFINITION - The feasible region for an LP is the set of all points that satisfies all the LP's <br> constraints and sign restrictions. ■

For example, in the Giapetto problem, the point $\left(x_{1}=40, x_{2}=20\right)$ is in the feasible region. Note that $x_{1}=40$ and $x_{2}=20$ satisfy the constraints (2)-(4) and the sign restrictions (5)-(6):

```
Constraint (2), \(2 x_{1}+x_{2} \leq 100\), is satisfied, because \(2(40)+20 \leq 100\).
Constraint (3), \(x_{1}+x_{2} \leq 80\), is satisfied, because \(40+20 \leq 80\).
Constraint (4), \(x_{1} \leq 40\), is satisfied, because \(40 \leq 40\).
Restriction (5), \(x_{1} \geq 0\), is satisfied, because \(40 \geq 0\).
Restriction (6), \(x_{2} \geq 0\), is satisfied, because \(20 \geq 0\).
```

On the other hand, the point $\left(x_{1}=15, x_{2}=70\right)$ is not in the feasible region, because even though $x_{1}=15$ and $x_{2}=70$ satisfy (2), (4), (5), and (6), they fail to satisfy (3): $15+70$ is not less than or equal to 80 . Any point that is not in an LP's feasible region is said to be an infeasible point. As another example of an infeasible point, consider ( $x_{1}=40$, $x_{2}=-20$ ). Although this point satisfies all the constraints and the sign restriction (5), it is infeasible because it fails to satisfy the sign restriction (6), $x_{2} \geq 0$. The feasible region for the Giapetto problem is the set of possible production plans that Giapetto must consider in searching for the optimal production plan.

## DEFINITION ■ For a maximization problem, an optimal solution to an LP is a point in the feasible region with the largest objective function value. Similarly, for a minimization problem, an optimal solution is a point in the feasible region with the smallest objective function value.

Most LPs have only one optimal solution. However, some LPs have no optimal solution, and some LPs have an infinite number of solutions (these situations are discussed in Section 3.3). In Section 3.2, we show that the unique optimal solution to the Giapetto problem is $\left(x_{1}=20, x_{2}=60\right)$. This solution yields an objective function value of

$$
z=3 x_{1}+2 x_{2}=3(20)+2(60)=\$ 180
$$

When we say that $\left(x_{1}=20, x_{2}=60\right)$ is the optimal solution to the Giapetto problem, we are saying that no point in the feasible region has an objective function value that exceeds 180. Giapetto can maximize profit by building 20 soldiers and 60 trains each week. If Giapetto were to produce 20 soldiers and 60 trains each week, the weekly profit would be $\$ 180$ less weekly fixed costs. For example, if Giapetto's only fixed cost were rent of \$100 per week, then weekly profit would be $180-100=\$ 80$ per week.

## PROBLEMS

## Group A

1 Farmer Jones must determine how many acres of corn and wheat to plant this year. An acre of wheat yields 25 bushels of wheat and requires 10 hours of labor per week. An acre of corn yields 10 bushels of corn and requires 4 hours of labor per week. All wheat can be sold at $\$ 4$ a bushel, and all corn can be sold at $\$ 3$ a bushel. Seven acres of land and 40 hours per week of labor are available. Government regulations require that at least 30 bushels of corn be produced during the current year. Let $x_{1}=$ number of acres of corn planted, and $x_{2}=$ number of acres of wheat planted. Using these decision variables, formulate an LP whose solution will tell Farmer Jones how to maximize the total revenue from wheat and corn.
2 Answer these questions about Problem 1.
a Is ( $x_{1}=2, x_{2}=3$ ) in the feasible region?
b Is ( $x_{1}=4, x_{2}=3$ ) in the feasible region?
c Is ( $x_{1}=2, x_{2}=-1$ ) in the feasible region?
d Is $\left(x_{1}=3, x_{2}=2\right)$ in the feasible region?
3 Using the variables $x_{1}=$ number of bushels of corn
produced and $x_{2}=$ number of bushels of wheat produced, reformulate Farmer Jones's LP.
4 Truckco manufactures two types of trucks: 1 and 2. Each truck must go through the painting shop and assembly shop. If the painting shop were completely devoted to painting Type 1 trucks, then 800 per day could be painted; if the painting shop were completely devoted to painting Type 2 trucks, then 700 per day could be painted. If the assembly shop were completely devoted to assembling truck 1 engines, then 1,500 per day could be assembled; if the assembly shop were completely devoted to assembling truck 2 engines, then 1,200 per day could be assembled. Each Type 1 truck contributes $\$ 300$ to profit; each Type 2 truck contributes $\$ 500$. Formulate an LP that will maximize Truckco's profit.

## Group B

5 Why don't we allow an LP to have $<$ or $>$ constraints?

### 3.2 The Graphical Solution of Two-Variable Linear Programming Problems

Any LP with only two variables can be solved graphically. We always label the variables $x_{1}$ and $x_{2}$ and the coordinate axes the $x_{1}$ and $x_{2}$ axes. Suppose we want to graph the set of points that satisfies

$$
\begin{equation*}
2 x_{1}+3 x_{2} \leq 6 \tag{7}
\end{equation*}
$$

The same set of points $\left(x_{1}, x_{2}\right)$ satisfies

$$
3 x_{2} \leq 6-2 x_{1}
$$

This last inequality may be rewritten as

$$
\begin{equation*}
x_{2} \leq \frac{1}{3}\left(6-2 x_{1}\right)=2-\frac{2}{3} x_{1} \tag{8}
\end{equation*}
$$

Because moving downward on the graph decreases $x_{2}$ (see Figure 1), the set of points that satisfies (8) and (7) lies on or below the line $x_{2}=2-\frac{2}{3} x_{1}$. This set of points is indicated by darker shading in Figure 1. Note, however, that $x_{2}=2-\frac{2}{3} x_{1}, 3 x_{2}=6-2 x_{1}$, and $2 x_{1}+$ $3 x_{2}=6$ are all the same line. This means that the set of points satisfying (7) lies on or below the line $2 x_{1}+3 x_{2}=6$. Similarly, the set of points satisfying $2 x_{1}+3 x_{2} \geq 6$ lies on or above the line $2 x_{1}+3 x_{2}=6$. (These points are shown by lighter shading in Figure 1.)

Consider a linear inequality constraint of the form $f\left(x_{1}, x_{2}\right) \geq b$ or $f\left(x_{1}, x_{2}\right) \leq b$. In general, it can be shown that in two dimensions, the set of points that satisfies a linear inequality includes the points on the line $f\left(x_{1}, x_{2}\right)=b$, defining the inequality plus all points on one side of the line.

There is an easy way to determine the side of the line for which an inequality such as $f\left(x_{1}, x_{2}\right) \leq b$ or $f\left(x_{1}, x_{2}\right) \geq b$ is satisfied. Just choose any point $P$ that does not satisfy the line $f\left(x_{1}, x_{2}\right)=b$. Determine whether $P$ satisfies the inequality. If it does, then all points on the same side as $P$ of $f\left(x_{1}, x_{2}\right)=b$ will satisfy the inequality. If $P$ does not satisfy the inequality, then all points on the other side of $f\left(x_{1}, x_{2}\right)=b$, which does not contain $P$, will satisfy the inequality. For example, to determine whether $2 x_{1}+3 x_{2} \geq 6$ is satisfied by points above or below the line $2 x_{1}+3 x_{2}=6$, we note that $(0,0)$ does not satisfy $2 x_{1}+3 x_{2} \geq 6$. Because ( 0 , 0 ) is below the line $2 x_{1}+3 x_{2}=6$, the set of points satisfying $2 x_{1}+3 x_{2} \geq 6$ includes the line $2 x_{1}+3 x_{2}=6$ and the points above the line $2 x_{1}+3 x_{2}=6$. This agrees with Figure 1 .

FIGURE 1 Graphing a Linear Inequality


## Finding the Feasible Solution

We illustrate how to solve two-variable LPs graphically by solving the Giapetto problem. To begin, we graphically determine the feasible region for Giapetto's problem. The feasible region for the Giapetto problem is the set of all points $\left(x_{1}, x_{2}\right)$ satisfying

$$
\begin{align*}
2 x_{1}+x_{2} & \leq 100 \quad \text { (Constraints) }  \tag{2}\\
x_{1}+x_{2} & \leq 80  \tag{3}\\
x_{1} & \leq 40  \tag{4}\\
x_{1} & \geq 0 \quad \text { (Sign restrictions) }  \tag{5}\\
x_{2} & \geq 0 \tag{6}
\end{align*}
$$

For a point $\left(x_{1}, x_{2}\right)$ to be in the feasible region, $\left(x_{1}, x_{2}\right)$ must satisfy all the inequalities (2)-(6). Note that the only points satisfying (5) and (6) lie in the first quadrant of the $x_{1}-x_{2}$ plane. This is indicated in Figure 2 by the arrows pointing to the right from the $x_{2}$ axis and upward from the $x_{1}$ axis. Thus, any point that is outside the first quadrant cannot be in the feasible region. This means that the feasible region will be the set of points in the first quadrant that satisfies (2)-(4).

Our method for determining the set of points that satisfies a linear inequality will also identify those that meet (2)-(4). From Figure 2, we see that (2) is satisfied by all points below or on the line $A B\left(A B\right.$ is the line $\left.2 x_{1}+x_{2}=100\right)$. Inequality (3) is satisfied by all points on or below the line $C D\left(C D\right.$ is the line $\left.x_{1}+x_{2}=80\right)$. Finally, (4) is satisfied by all points on or to the left of line $E F\left(E F\right.$ is the line $\left.x_{1}=40\right)$. The side of a line that satisfies an inequality is indicated by the direction of the arrows in Figure 2.

From Figure 2, we see that the set of points in the first quadrant that satisfies (2), (3), and (4) is bounded by the five-sided polygon $D G F E H$. Any point on this polygon or in its interior is in the feasible region. Any other point fails to satisfy at least one of the inequalities $(2)-(6)$. For example, the point $(40,30)$ lies outside $D G F E H$ because it is above the line segment $A B$. Thus $(40,30)$ is infeasible, because it fails to satisfy (2).

FIGURE 2 Graphical Solution of Giapetto Problem


An easy way to find the feasible region is to determine the set of infeasible points. Note that all points above line $A B$ in Figure 2 are infeasible, because they fail to satisfy (2). Similarly, all points above $C D$ are infeasible, because they fail to satisfy (3). Also, all points to the right of the vertical line $E F$ are infeasible, because they fail to satisfy (4). After these points are eliminated from consideration, we are left with the feasible region (DGFEH).

## Finding the Optimal Solution

Having identified the feasible region for the Giapetto problem, we now search for the optimal solution, which will be the point in the feasible region with the largest value of $z=$ $3 x_{1}+2 x_{2}$. To find the optimal solution, we need to graph a line on which all points have the same $z$-value. In a max problem, such a line is called an isoprofit line (in a min problem, an isocost line). To draw an isoprofit line, choose any point in the feasible region and calculate its $z$-value. Let us choose $(20,0)$. For $(20,0), z=3(20)+2(0)=60$. Thus, $(20,0)$ lies on the isoprofit line $z=3 x_{1}+2 x_{2}=60$. Rewriting $3 x_{1}+2 x_{2}=60$ as $x_{2}=$ $30-\frac{3}{2} x_{1}$, we see that the isoprofit line $3 x_{1}+2 x_{2}=60$ has a slope of $-\frac{3}{2}$. Because all isoprofit lines are of the form $3 x_{1}+2 x_{2}=$ constant, all isoprofit lines have the same slope. This means that once we have drawn one isoprofit line, we can find all other isoprofit lines by moving parallel to the isoprofit line we have drawn.

It is now clear how to find the optimal solution to a two-variable LP. After you have drawn a single isoprofit line, generate other isoprofit lines by moving parallel to the drawn isoprofit line in a direction that increases $z$ (for a max problem). After a point, the isoprofit lines will no longer intersect the feasible region. The last isoprofit line intersecting (touching) the feasible region defines the largest $z$-value of any point in the feasible region and indicates the optimal solution to the LP. In our problem, the objective function $z=3 x_{1}+2 x_{2}$ will increase if we move in a direction for which both $x_{1}$ and $x_{2}$ increase. Thus, we construct additional isoprofit lines by moving parallel to $3 x_{1}+2 x_{2}=60$ in a northeast direction (upward and to the right). From Figure 2, we see that the isoprofit line passing through point $G$ is the last isoprofit line to intersect the feasible region. Thus, $G$ is the point in the feasible region with the largest $z$-value and is therefore the optimal solution to the Giapetto problem. Note that point $G$ is where the lines $2 x_{1}+x_{2}=100$ and $x_{1}+x_{2}=80$ intersect. Solving these two equations simultaneously, we find that ( $x_{1}=$ $20, x_{2}=60$ ) is the optimal solution to the Giapetto problem. The optimal value of $z$ may be found by substituting these values of $x_{1}$ and $x_{2}$ into the objective function. Thus, the optimal value of $z$ is $z=3(20)+2(60)=180$.

## Binding and Nonbinding Constraints

Once the optimal solution to an LP has been found, it is useful (see Chapters 5 and 6) to classify each constraint as being a binding constraint or a nonbinding constraint.

[^1]Thus, (2) and (3) are binding constraints.

A constraint is nonbinding if the left-hand side and the right-hand side of the constraint are unequal when the optimal values of the decision variables are substituted into the constraint.

Because $x_{1}=20$ is less than 40,(4) is a nonbinding constraint.

## Convex Sets, Extreme Points, and LP

The feasible region for the Giapetto problem is an example of a convex set.


Figure 3 gives four illustrations of this definition. In Figures 3 a and 3b, each line segment joining two points in $S$ contains only points in $S$. Thus, in both these figures, $S$ is convex. In Figures 3c and 3d, $S$ is not convex. In each figure, points $A$ and $B$ are in $S$, but there are points on the line segment $A B$ that are not contained in $S$. In our study of linear programming, a certain type of point in a convex set (called an extreme point) is of great interest.

DEFINITION ■ For any convex set $S$, a point $P$ in $S$ is an extreme point if each line segment that lies completely in $S$ and contains the point $P$ has $P$ as an endpoint of the line segment.

For example, in Figure 3a, each point on the circumference of the circle is an extreme point of the circle. In Figure 3b, points $A, B, C$, and $D$ are extreme points of $S$. Although point $E$ is on the boundary of $S$ in Figure 3b, $E$ is not an extreme point of $S$. This is because $E$ lies on the line segment $A B$ ( $A B$ lies completely in $S$ ), and $E$ is not an endpoint of the line segment $A B$. Extreme points are sometimes called corner points, because if the set $S$ is a polygon, the extreme points of $S$ will be the vertices, or corners, of the polygon.

The feasible region for the Giapetto problem is a convex set. This is no accident: It can be shown that the feasible region for any LP will be a convex set. From Figure 2, we see that the extreme points of the feasible region are simply points $D, F, E, G$, and $H$. It can be shown that the feasible region for any LP has only a finite number of extreme points. Also note that the optimal solution to the Giapetto problem (point $G$ ) is an extreme point of the feasible region. It can be shown that any LP that has an optimal solution has an extreme point that is optimal. This result is very important, because it reduces the set of points that yield an optimal solution from the entire feasible region (which generally contains an infinite number of points) to the set of extreme points (a finite set).

FIGURE 3 Convex and Nonconvex Sets


d

For the Giapetto problem, it is easy to see why the optimal solution must be an extreme point of the feasible region. We note that $z$ increases as we move isoprofit lines in a northeast direction, so the largest $z$-value in the feasible region must occur at some point $P$ that has no points in the feasible region northeast of $P$. This means that the optimal solution must lie somewhere on the boundary of the feasible region $D G F E H$. The LP must have an extreme point that is optimal, because for any line segment on the boundary of the feasible region, the largest $z$-value on that line segment must be assumed at one of the endpoints of the line segment.

To see this, look at the line segment $F G$ in Figure 2. $F G$ is part of the line $2 x_{1}+$ $x_{2}=100$ and has a slope of -2 . If we move along $F G$ and decrease $x_{1}$ by 1 , then $x_{2}$ will increase by 2 , and the value of $z$ changes as follows: $3 x_{1}$ goes down by $3(1)=3$, and $2 x_{2}$ goes up by $2(2)=4$. Thus, in total, $z$ increases by $4-3=1$. This means that moving along $F G$ in a direction of decreasing $x_{1}$ increases $z$. Thus, the value of $z$ at point $G$ must exceed the value of $z$ at any other point on the line segment $F G$.

A similar argument shows that for any objective function, the maximum value of $z$ on a given line segment must occur at an endpoint of the line segment. Therefore, for any LP, the largest $z$-value in the feasible region must be attained at an endpoint of one of the line segments forming the boundary of the feasible region. In short, one of the extreme points of the feasible region must be optimal. (To test your understanding, show that if Giapetto's objective function were $z=6 x_{1}+x_{2}$, point $F$ would be optimal, whereas if Giapetto's objective function were $z=x_{1}+6 x_{2}$, point $D$ would be optimal.)

Our proof that an LP always has an optimal extreme point depended heavily on the fact that both the objective function and the constraints were linear functions. In Chapter 11, we show that for an optimization problem in which the objective function or some of the constraints are not linear, the optimal solution to the optimization problem may not occur at an extreme point.

## The Graphical Solution of Minimization Problems

## EXAMPLE 2 Dorian Auto

Dorian Auto manufactures luxury cars and trucks. The company believes that its most likely customers are high-income women and men. To reach these groups, Dorian Auto has embarked on an ambitious TV advertising campaign and has decided to purchase 1-minute commercial spots on two types of programs: comedy shows and football games. Each comedy commercial is seen by 7 million high-income women and 2 million highincome men. Each football commercial is seen by 2 million high-income women and 12 million high-income men. A 1-minute comedy ad costs $\$ 50,000$, and a 1-minute football ad costs $\$ 100,000$. Dorian would like the commercials to be seen by at least 28 million high-income women and 24 million high-income men. Use linear programming to determine how Dorian Auto can meet its advertising requirements at minimum cost.

Solution Dorian must decide how many comedy and football ads should be purchased, so the decision variables are
$x_{1}=$ number of 1-minute comedy ads purchased
$x_{2}=$ number of 1-minute football ads purchased

Then Dorian wants to minimize total advertising cost (in thousands of dollars).
Total advertising cost $=$ cost of comedy ads + cost of football ads

$$
\begin{aligned}
& =\left(\frac{\text { cost }}{\text { comedy ad }}\right)\binom{\text { total }}{\text { comedy ads }}+\left(\begin{array}{l}
\text { cost }
\end{array}\right)\binom{\text { total }}{\text { football ads }} \\
& =50 x_{1}+100 x_{2}
\end{aligned}
$$

Thus, Dorian's objective function is

$$
\begin{equation*}
\min z=50 x_{1}+100 x_{2} \tag{9}
\end{equation*}
$$

Dorian faces the following constraints:
Constraint 1 Commercials must reach at least 28 million high-income women.
Constraint 2 Commercials must reach at least 24 million high-income men.
To express Constraints 1 and 2 in terms of $x_{1}$ and $x_{2}$, let HIW stand for high-income women viewers and HIM stand for high-income men viewers (in millions).

$$
\begin{aligned}
\text { HIW } & =\left(\frac{\text { HIW }}{\text { comedy ad }}\right)\binom{\text { total }}{\text { comedy ads }}+\left(\frac{\text { HIW }}{\text { football ad }}\right)\binom{\text { total }}{\text { football ads }} \\
& =7 x_{1}+2 x_{2} \\
\text { HIM } & =\left(\frac{\text { HIM }}{\text { comedy ad }}\right)\binom{\text { total }}{\text { comedy ads }}+\left(\frac{\text { HIM }}{\text { football ad }}\right)\binom{\text { total }}{\text { football ads }} \\
& =2 x_{1}+12 x_{2}
\end{aligned}
$$

Constraint 1 may now be expressed as

$$
\begin{equation*}
7 x_{1}+2 x_{2} \geq 28 \tag{10}
\end{equation*}
$$

and Constraint 2 may be expressed as

$$
\begin{equation*}
2 x_{1}+12 x_{2} \geq 24 \tag{11}
\end{equation*}
$$

The sign restrictions $x_{1} \geq 0$ and $x_{2} \geq 0$ are necessary, so the Dorian LP is given by:

$$
\begin{array}{lrl}
\min z=50 x_{1}+100 x_{2} \\
\text { s.t. } & 7 x_{1}+2 x_{2} & \geq 28 \\
2 x_{1}+12 x_{2} & \geq 24 \\
& x_{1}, x_{2} & \geq 0
\end{array}
$$

This problem is typical of a wide range of LP applications in which a decision maker wants to minimize the cost of meeting a certain set of requirements. To solve this LP graphically, we begin by graphing the feasible region (Figure 4). Note that (10) is satisfied by points on or above the line $A B\left(A B\right.$ is part of the line $\left.7 x_{1}+2 x_{2}=28\right)$ and that

FIGURE 4

(11) is satisfied by the points on or above the line $C D$ ( $C D$ is part of the line $2 x_{1}+$ $12 x_{2}=24$ ). From Figure 4, we see that the only first-quadrant points satisfying both (10) and (11) are the points in the shaded region bounded by the $x_{1}$ axis, $C E B$, and the $x_{2}$ axis.

Like the Giapetto problem, the Dorian problem has a convex feasible region, but the feasible region for Dorian, unlike Giapetto's, contains points for which the value of at least one variable can assume arbitrarily large values. Such a feasible region is called an unbounded feasible region.

Because Dorian wants to minimize total advertising cost, the optimal solution to the problem is the point in the feasible region with the smallest $z$-value. To find the optimal solution, we need to draw an isocost line that intersects the feasible region. An isocost line is any line on which all points have the same $z$-value (or same cost). We arbitrarily choose the isocost line passing through the point $\left(x_{1}=4, x_{2}=4\right)$. For this point, $z=$ $50(4)+100(4)=600$, and we graph the isocost line $z=50 x_{1}+100 x_{2}=600$.

We consider lines parallel to the isocost line $50 x_{1}+100 x_{2}=600$ in the direction of decreasing $z$ (southwest). The last point in the feasible region that intersects an isocost line will be the point in the feasible region having the smallest $z$-value. From Figure 4, we see that point $E$ has the smallest $z$-value of any point in the feasible region; this is the optimal solution to the Dorian problem. Note that point $E$ is where the lines $7 x_{1}+2 x_{2}=$ 28 and $2 x_{1}+12 x_{2}=24$ intersect. Simultaneously solving these equations yields the optimal solution $\left(x_{1}=3.6, x_{2}=1.4\right)$. The optimal $z$-value can then be found by substituting these values of $x_{1}$ and $x_{2}$ into the objective function. Thus, the optimal $z$-value is $z=$ $50(3.6)+100(1.4)=320=\$ 320,000$. Because at point $E$ both the HIW and HIM constraints are satisfied with equality, both constraints are binding.

Does the Dorian model meet the four assumptions of linear programming outlined in Section 3.1?

For the Proportionality Assumption to be valid, each extra comedy commercial must add exactly 7 million HIW and 2 million HIM. This contradicts empirical evidence, which indicates that after a certain point advertising yields diminishing returns. After, say, 500 auto commercials have been aired, most people have probably seen one, so it does little good to air more commercials. Thus, the Proportionality Assumption is violated.

We used the Additivity Assumption to justify writing (total HIW viewers) $=(\mathrm{HIW}$ viewers from comedy ads) + (HIW viewers from football ads). In reality, many of the same people will see a Dorian comedy commercial and a Dorian football commercial. We are double-counting such people, and this creates an inaccurate picture of the total number of people seeing Dorian commercials. The fact that the same person may see more than one type of commercial means that the effectiveness of, say, a comedy commercial depends on the number of football commercials. This violates the Additivity Assumption.

If only 1 -minute commercials are available, then it is unreasonable to say that Dorian should buy 3.6 comedy commercials and 1.4 football commercials, so the Divisibility Assumption is violated, and the Dorian problem should be considered an integer programming problem. In Section 9.3, we show that if the Dorian problem is solved as an integer programming problem, then the minimum cost is attained by choosing $\left(x_{1}=6, x_{2}=1\right)$ or $\left(x_{1}=4, x_{2}=2\right)$. For either solution, the minimum cost is $\$ 400,000$. This is $25 \%$ higher than the cost obtained from the optimal LP solution.

Because there is no way to know with certainty how many viewers are added by each type of commercial, the Certainty Assumption is also violated. Thus, all the assumptions of linear programming seem to be violated by the Dorian Auto problem. Despite these drawbacks, analysts have used similar models to help companies determine their optimal media mix. ${ }^{\dagger}$

[^2]
## PROBLEMS

## Group A

1 Graphically solve Problem 1 of Section 3.1.
2 Graphically solve Problem 4 of Section 3.1.
3 Leary Chemical manufactures three chemicals: A, B, and C . These chemicals are produced via two production processes: 1 and 2. Running process 1 for an hour costs $\$ 4$ and yields 3 units of A, 1 of B, and 1 of C. Running process 2 for an hour costs $\$ 1$ and produces 1 unit of A and 1 of B . To meet customer demands, at least 10 units of $\mathrm{A}, 5$ of B , and 3 of C must be produced daily. Graphically determine a daily production plan that minimizes the cost of meeting Leary Chemical's daily demands.
4 For each of the following, determine the direction in which the objective function increases:
a $z=4 x_{1}-x_{2}$
b $z=-x_{1}+2 x_{2}$
c $z=-x_{1}-3 x_{2}$
5 Furnco manufactures desks and chairs. Each desk uses 4 units of wood, and each chair uses 3 . A desk contributes
$\$ 40$ to profit, and a chair contributes $\$ 25$. Marketing restrictions require that the number of chairs produced be at least twice the number of desks produced. If 20 units of wood are available, formulate an LP to maximize Furnco's profit. Then graphically solve the LP.
6 Farmer Jane owns 45 acres of land. She is going to plant each with wheat or corn. Each acre planted with wheat yields $\$ 200$ profit; each with corn yields $\$ 300$ profit. The labor and fertilizer used for each acre are given in Table 1. One hundred workers and 120 tons of fertilizer are available. Use linear programming to determine how Jane can maximize profits from her land.

TABLE 1

|  | Wheat | Corn |
| :--- | :--- | :--- |
| Labor | 3 workers | 2 workers |
| Fertilizer | 2 tons | 4 tons |

### 3.3 Special Cases

The Giapetto and Dorian problems each had a unique optimal solution. In this section, we encounter three types of LPs that do not have unique optimal solutions.
1 Some LPs have an infinite number of optimal solutions (alternative or multiple optimal solutions).

2 Some LPs have no feasible solutions (infeasible LPs).
3 Some LPs are unbounded: There are points in the feasible region with arbitrarily large (in a max problem) $z$-values.

## Alternative or Multiple Optimal Solutions

## EXAMPLE 3 Alternative Optimal Solutions

An auto company manufactures cars and trucks. Each vehicle must be processed in the paint shop and body assembly shop. If the paint shop were only painting trucks, then 40 per day could be painted. If the paint shop were only painting cars, then 60 per day could be painted. If the body shop were only producing cars, then it could process 50 per day. If the body shop were only producing trucks, then it could process 50 per day. Each truck contributes $\$ 300$ to profit, and each car contributes $\$ 200$ to profit. Use linear programming to determine a daily production schedule that will maximize the company's profits.
Solution The company must decide how many cars and trucks should be produced daily. This leads us to define the following decision variables:

$$
\begin{aligned}
& x_{1}=\text { number of trucks produced daily } \\
& x_{2}=\text { number of cars produced daily }
\end{aligned}
$$

The company's daily profit (in hundreds of dollars) is $3 x_{1}+2 x_{2}$, so the company's objective function may be written as

$$
\begin{equation*}
\max z=3 x_{1}+2 x_{2} \tag{12}
\end{equation*}
$$

The company's two constraints are the following:
Constraint 1 The fraction of the day during which the paint shop is busy is less than or equal to 1 .

Constraint 2 The fraction of the day during which the body shop is busy is less than or equal to 1 .

We have

$$
\begin{aligned}
\text { Fraction of day paint shop works on trucks } & =\left(\frac{\text { fraction of day }}{\text { truck }}\right)\left(\frac{\text { trucks }}{\text { day }}\right) \\
& =\frac{1}{40} x_{1} \\
\text { Fraction of day paint shop works on cars } & =\frac{1}{60} x_{2} \\
\text { Fraction of day body shop works on trucks } & =\frac{1}{50} x_{1} \\
\text { Fraction of day body shop works on cars } & =\frac{1}{50} x_{2}
\end{aligned}
$$

Thus, Constraint 1 may be expressed by

$$
\begin{equation*}
\frac{1}{40} x_{1}+\frac{1}{60} x_{2} \leq 1 \quad \text { (Paint shop constraint) } \tag{13}
\end{equation*}
$$

and Constraint 2 may be expressed by

$$
\begin{equation*}
\frac{1}{50} x_{1}+\frac{1}{50} x_{2} \leq 1 \quad \text { (Body shop constraint) } \tag{14}
\end{equation*}
$$

Because $x_{1} \geq 0$ and $x_{2} \geq 0$ must hold, the relevant LP is

$$
\begin{align*}
& \max z=3 x_{1}+2 x_{2}  \tag{12}\\
& \text { s.t. } \quad \frac{1}{40} x_{1}+\frac{1}{60} x_{2} \leq 1  \tag{13}\\
& \frac{1}{50} x_{1}+\frac{1}{50} x_{2} \leq 1  \tag{14}\\
& x_{1}, x_{2} \geq 0
\end{align*}
$$

The feasible region for this LP is the shaded region in Figure 5 bounded by $A E D F^{\dagger}$
For our isoprofit line, we choose the line passing through the point $(20,0)$. Because $(20,0)$ has a $z$-value of $3(20)+2(0)=60$, this yields the isoprofit line $z=3 x_{1}+$ $2 x_{2}=60$. Examining lines parallel to this isoprofit line in the direction of increasing $z$ (northeast), we find that the last "point" in the feasible region to intersect an isoprofit line is the entire line segment $A E$. This means that any point on the line segment $A E$ is optimal. We can use any point on $A E$ to determine the optimal $z$-value. For example, point $A$, $(40,0)$, gives $z=3(40)=120$.

In summary, the auto company's LP has an infinite number of optimal solutions, or multiple or alternative optimal solutions. This is indicated by the fact that as an isoprofit

[^3]
line leaves the feasible region, it will intersect an entire line segment corresponding to the binding constraint (in this case, $A E$ ).

From our current example, it seems reasonable (and can be shown to be true) that if two points ( $A$ and $E$ here) are optimal, then any point on the line segment joining these two points will also be optimal.

If an alternative optimum occurs, then the decision maker can use a secondary criterion to choose between optimal solutions. The auto company's managers might prefer point $A$ because it would simplify their business (and still allow them to maximize profits) by allowing them to produce only one type of product (trucks).

The technique of goal programming (see Section 4.14) is often used to choose among alternative optimal solutions.

## Infeasible LP

It is possible for an LP's feasible region to be empty (contain no points), resulting in an infeasible LP. Because the optimal solution to an LP is the best point in the feasible region, an infeasible LP has no optimal solution.

## EXAMPLE 4 Infeasible LP

Suppose that auto dealers require that the auto company in Example 3 produce at least 30 trucks and 20 cars. Find the optimal solution to the new LP.

Solution After adding the constraints $x_{1} \geq 30$ and $x_{2} \geq 20$ to the LP of Example 3, we obtain the following LP:

$$
\begin{array}{ll}
\max z= & 3 x_{1}+2 x_{2} \\
\text { s.t. } & \frac{1}{40} x_{1}+\frac{1}{60} x_{2} \leq 1 \\
& \frac{1}{50} x_{1}+\frac{1}{50} x_{2} \leq 1 \tag{16}
\end{array}
$$

FIGURE 6
An Empty Feasible Region (Infeasible LP)


$$
\begin{align*}
x_{1} & \geq 30  \tag{17}\\
x_{2} & \geq 20  \tag{18}\\
x_{1}, x_{2} & \geq 0
\end{align*}
$$

The graph of the feasible region for this LP is Figure 6.
Constraint (15) is satisfied by all points on or below $A B\left(A B\right.$ is $\left.\frac{1}{40} x_{1}+\frac{1}{60} x_{2}=1\right)$.
Constraint (16) is satisfied by all points on or below $C D\left(C D\right.$ is $\left.\frac{1}{50} x_{1}+\frac{1}{50} x_{2}=1\right)$.
Constraint (17) is satisfied by all points on or to the right of $E F\left(E F\right.$ is $\left.x_{1}=30\right)$.
Constraint (18) is satisfied by all points on or above $G H\left(G H\right.$ is $\left.x_{2}=20\right)$.
From Figure 6 it is clear that no point satisfies all of (15)-(18). This means that Example 4 has an empty feasible region and is an infeasible LP.

In Example 4, the LP is infeasible because producing 30 trucks and 20 cars requires more paint shop time than is available.

## Unbounded LP

Our next special LP is an unbounded LP. For a max problem, an unbounded LP occurs if it is possible to find points in the feasible region with arbitrarily large $z$-values, which corresponds to a decision maker earning arbitrarily large revenues or profits. This would indicate that an unbounded optimal solution should not occur in a correctly formulated LP. Thus, if the reader ever solves an LP on the computer and finds that the LP is unbounded, then an error has probably been made in formulating the LP or in inputting the LP into the computer.

For a minimization problem, an LP is unbounded if there are points in the feasible region with arbitrarily small $z$-values. When graphically solving an LP, we can spot an unbounded LP as follows: A max problem is unbounded if, when we move parallel to our original isoprofit line in the direction of increasing $z$, we never entirely leave the feasible region. A minimization problem is unbounded if we never leave the feasible region when moving in the direction of decreasing $z$.

Graphically solve the following LP:

$$
\begin{array}{lr}
\max z=2 x_{1}-x_{2} \\
\text { s.t. } \quad x_{1}-x_{2} & \leq 1 \\
2 x_{1}+x_{2} & \geq 6  \tag{20}\\
x_{1}, x_{2} & \geq 0
\end{array}
$$

Solution
From Figure 7, we see that (19) is satisfied by all points on or above $A B$ ( $A B$ is the line $\left.x_{1}-x_{2}=1\right)$. Also, (20) is satisfied by all points on or above $C D\left(C D\right.$ is $\left.2 x_{1}+x_{2}=6\right)$. Thus, the feasible region for Example 5 is the (shaded) unbounded region in Figure 7, which is bounded only by the $x_{2}$ axis, line segment $D E$, and the part of line $A B$ beginning at $E$. To find the optimal solution, we draw the isoprofit line passing through $(2,0)$. This isoprofit line has $z=2 x_{1}-x_{2}=2(2)-0=4$. The direction of increasing $z$ is to the southeast (this makes $x_{1}$ larger and $x_{2}$ smaller). Moving parallel to $z=2 x_{1}-x_{2}$ in a southeast direction, we see that any isoprofit line we draw will intersect the feasible region. (This is because any isoprofit line is steeper than the line $x_{1}-x_{2}=1$.)

Thus, there are points in the feasible region that have arbitrarily large $z$-values. For example, if we wanted to find a point in the feasible region that had $z \geq 1,000,000$, we could choose any point in the feasible region that is southeast of the isoprofit line $z=1,000,000$.

From the discussion in the last two sections, we see that every LP with two variables must fall into one of the following four cases:

Case 1 The LP has a unique optimal solution.
Case 2 The LP has alternative or multiple optimal solutions: Two or more extreme points are optimal, and the LP will have an infinite number of optimal solutions.
Case 3 The LP is infeasible: The feasible region contains no points.
Case 4 The LP is unbounded: There are points in the feasible region with arbitrarily large $z$-values (max problem) or arbitrarily small $z$-values (min problem).

In Chapter 4, we show that every LP (not just LPs with two variables) must fall into one of Cases 1-4.

FIGURE 7 An Unbounded LP



## The Simplex Algorithm and Goal Programming

In Chapter 3, we saw how to solve two-variable linear programming problems graphically. Unfortunately, most real-life LPs have many variables, so a method is needed to solve LPs with more than two variables. We devote most of this chapter to a discussion of the simplex algorithm, which is used to solve even very large LPs. In many industrial applications, the simplex algorithm is used to solve LPs with thousands of constraints and variables.

In this chapter, we explain how the simplex algorithm can be used to find optimal solutions to LPs. We also detail how two state-of-the-art computer packages (LINDO and LINGO) can be used to solve LPs. Briefly, we also discuss Karmarkar's pioneering approach for solving LPs. We close the chapter with an introduction to goal programming, which enables the decision maker to consider more than one objective function.

### 4.1 How to Convert an LP to Standard Form

We have seen that an LP can have both equality and inequality constraints. It also can have variables that are required to be nonnegative as well as those allowed to be unrestricted in sign (urs). Before the simplex algorithm can be used to solve an LP, the LP must be converted into an equivalent problem in which all constraints are equations and all variables are nonnegative. An LP in this form is said to be in standard form. ${ }^{\dagger}$

To convert an LP into standard form, each inequality constraint must be replaced by an equality constraint. We illustrate this procedure using the following problem.

## EXAMPLE 1 Leather Limited

Leather Limited manufactures two types of belts: the deluxe model and the regular model. Each type requires 1 sq yd of leather. A regular belt requires 1 hour of skilled labor, and a deluxe belt requires 2 hours. Each week, 40 sq yd of leather and 60 hours of skilled labor are available. Each regular belt contributes $\$ 3$ to profit and each deluxe belt, $\$ 4$. If we define

$$
\begin{aligned}
& x_{1}=\text { number of deluxe belts produced weekly } \\
& x_{2}=\text { number of regular belts produced weekly }
\end{aligned}
$$

[^4]the appropriate LP is
\[

$$
\begin{array}{lrl}
\max z=4 x_{1}+3 x_{2} & & \\
\text { s.t. } & & \\
& x_{1}+x_{2} & \leq 40  \tag{2}\\
& & \text { (Leather constraint) } \\
2 x_{1}+x_{2} & \leq 60 & \\
& & \text { (Labor constraint) } \\
& x_{1}, x_{2} & \geq 0
\end{array}
$$
\]

How can we convert (1) and (2) to equality constraints? We define for each $\leq$ constraint a slack variable $s_{i}\left(s_{i}=\right.$ slack variable for $i$ th constraint), which is the amount of the resource unused in the $i$ th constraint. Because $x_{1}+x_{2}$ sq yd of leather are being used, and 40 sq yd are available, we define $s_{1}$ by

$$
s_{1}=40-x_{1}-x_{2} \quad \text { or } \quad x_{1}+x_{2}+s_{1}=40
$$

Similarly, we define $s_{2}$ by

$$
s_{2}=60-2 x_{1}-x_{2} \quad \text { or } \quad 2 x_{1}+x_{2}+s_{2}=60
$$

Observe that a point $\left(x_{1}, x_{2}\right)$ satisfies the $i$ th constraint if and only if $s_{i} \geq 0$. For example, $x_{1}=15, x_{2}=20$ satisfies (1) because $s_{1}=40-15-20=5 \geq 0$.

Intuitively, (1) is satisfied by the point $(15,20)$, because $s_{1}=5 \mathrm{sq}$ yd of leather are unused. Similarly, $(15,20)$ satisfies (2), because $s_{2}=60-2(15)-20=10$ labor hours are unused. Finally, note that the point $x_{1}=x_{2}=25$ fails to satisfy (2), because $s_{2}=$ $60-2(25)-25=-15$ indicates that $(25,25)$ uses more labor than is available.

In summary, to convert (1) to an equality constraint, we replace (1) by $s_{1}=40-$ $x_{1}-x_{2}$ (or $x_{1}+x_{2}+s_{1}=40$ ) and $s_{1} \geq 0$. To convert (2) to an equality constraint, we replace (2) by $s_{2}=60-2 x_{1}-x_{2}$ (or $2 x_{1}+x_{2}+s_{2}=60$ ) and $s_{2} \geq 0$. This converts LP 1 to

$$
\begin{array}{lrl}
\max z=4 x_{1}+3 x_{2} & \\
\text { s.t. } \quad x_{1}+x_{2}+s_{1} & =40  \tag{LP1'}\\
2 x_{1}+x_{2}+s_{2} & =60 \\
& x_{1}, x_{2}, s_{1}, s_{2} & \geq 0
\end{array}
$$

Note that LP $1^{\prime}$ is in standard form. In summary, if constraint $i$ of an $L P$ is $a \leq$ constraint, then we convert it to an equality constraint by adding a slack variable $s_{i}$ to the ith constraint and adding the sign restriction $s_{i} \geq 0$.

To illustrate how a $\geq$ constraint can be converted to an equality constraint, let's consider the diet problem of Section 3.4.


To convert the $i$ th $\geq$ constraint to an equality constraint, we define an excess variable (sometimes called a surplus variable) $e_{i}$. $e_{i}$ will always be the excess variable for the $i$ th
constraint.) We define $e_{i}$ to be the amount by which the $i$ th constraint is oversatisfied. Thus, for the diet problem,

$$
\left.\begin{array}{l}
e_{1}=400 x_{1}+200 x_{2}+150 x_{3}+500 x_{4}-500, \quad \text { or } \\
\\
e_{2}=300 x_{1}+200 x_{2}+150 x_{3}+500 x_{4}-e_{1}-6, \quad \text { or } \quad 3 x_{1}+2 x_{2}-e_{2}=6
\end{array}\right\} \begin{aligned}
& e_{3}=2 x_{1}+2 x_{2}+4 x_{3}+4 x_{4}-10, \quad \text { or } \quad 2 x_{1}+2 x_{2}+4 x_{3}+4 x_{4}-e_{3}=10 \\
& e_{4}=2 x_{1}+4 x_{2}+x_{3}+5 x_{4}-8, \quad \text { or } \quad 2 x_{1}+4 x_{2}+x_{3}+5 x_{4}-e_{4}=8
\end{aligned}
$$

A point $\left(x_{1}, x_{2}, x_{3}, x_{4}\right)$ satisfies the $i$ th $\geq$ constraint if and only if $e_{i}$ is nonnegative. For example, from ( $4^{\prime}$ ), $e_{2} \geq 0$ if and only if $3 x_{1}+2 x_{2} \geq 6$. For a numerical example, consider the point $x_{1}=2, x_{3}=4, x_{2}=x_{4}=0$, which satisfies all four of the diet problem's constraints. For this point,

$$
\begin{aligned}
& e_{1}=400(2)+150(4)-500=900 \geq 0 \\
& e_{2}=3(2)-6=0 \geq 0 \\
& e_{3}=2(2)+4(4)-10=10 \geq 0 \\
& e_{4}=2(2)+4-8=0 \geq 0
\end{aligned}
$$

As another example, consider $x_{1}=x_{2}=1, x_{3}=x_{4}=0$. This point is infeasible; it violates the chocolate, sugar, and fat constraints. The infeasibility of this point is indicated by

$$
\begin{aligned}
& e_{2}=3(1)+2(1)-6=-1<0 \\
& e_{3}=2(1)+2(1)-10=-6<0 \\
& e_{4}=2(1)+4(1)-8=-2<0
\end{aligned}
$$

Thus, to transform the diet problem into standard form, replace (3) by (3'); (4) by (4'); (5) by ( $5^{\prime}$ ); and (6) by ( $6^{\prime}$ ). We must also add the sign restrictions $e_{i} \geq 0(i=1,2,3,4)$. The resulting LP is in standard form and may be written as

$$
\begin{array}{lrlrl}
\min z=50 x_{1}+20 x_{2}+30 x_{3}+80 x_{4} & & \\
\text { s.t. } 400 x_{1}+200 x_{2}+150 x_{3}+500 x_{4}-e_{1} & & =500 \\
& 3 x_{1}+2 x_{2} & & =e_{2} & =6 \\
2 x_{1}+2 x_{2}+4 x_{3}+4 x_{4} & -e_{3} & =10 \\
2 x_{1}+4 x_{2}+4 x_{3}+5 x_{4} & -e_{4} & =8 \\
& x_{i}, e_{i} \geq 0 \quad(i=1,2,3,4) & &
\end{array}
$$

In summary, if the ith constraint of an LP is $a \geq$ constraint, then it can be converted to an equality constraint by subtracting an excess variable $e_{i}$ from the ith constraint and adding the sign restriction $e_{i} \geq 0$.

If an LP has both $\leq$ and $\geq$ constraints, then simply apply the procedures we have described to the individual constraints. As an example, let's convert the short-term financial planning model of Section 3.7 to standard form. Recall that the original LP was

$$
\begin{array}{rlr}
\max z=20 x_{1}+15 x_{2} & \\
\text { s.t. } & x_{1} & \leq 100 \\
x_{2} & \leq 100 \\
50 x_{1}+35 x_{2} & \leq 6,000 \\
20 x_{1}+15 x_{2} & \geq 2,000 \\
x_{1}, x_{2} & \geq 0
\end{array}
$$

Following the procedures described previously, we transform this LP into standard form by adding slack variables $s_{1}, s_{2}$, and $s_{3}$, respectively, to the first three constraints and subtracting an excess variable $e_{4}$ from the fourth constraint. Then we add the sign restrictions $s_{1} \geq 0, s_{2} \geq 0, s_{3} \geq 0$, and $e_{4} \geq 0$. This yields the following LP in standard form:

$$
\begin{aligned}
& \max z=20 x_{1}+15 x_{2} \\
& \text { s.t. } x_{1}+\quad+s_{1} \quad=100 \\
& x_{2}+s_{2}=100 \\
& 50 x_{1}+35 x_{2}+s_{3}=6,000 \\
& 20 x_{1}+15 x_{2} \quad-e_{4}=2,000 \\
& x_{i} \geq 0 \quad(i=1,2) ; \quad s_{i} \geq 0 \quad(i=1,2,3) ; \quad e_{4} \geq 0
\end{aligned}
$$

Of course, we could easily have labeled the excess variable for the fourth constraint $e_{1}$ (because it is the first excess variable). We chose to call it $e_{4}$ rather than $e_{1}$ to indicate that $e_{4}$ is the excess variable for the fourth constraint.

## PROBLEMS

## Group A

1 Convert the Giapetto problem (Example 1 in Chapter 3) to standard form.

2 Convert the Dorian problem (Example 2 in Chapter 3) to standard form.

3 Convert the following LP to standard form:

$$
\begin{aligned}
& \min z=3 x_{1}+x_{2} \\
& \text { s.t. } x_{1} \geq 3 \\
& x_{1}+x_{2} \leq 4 \\
& 2 x_{1}-x_{2}=3 \\
& x_{1}, x_{2} \geq 0
\end{aligned}
$$

### 4.2 Preview of the Simplex Algorithm

Suppose we have converted an LP with $m$ constraints into standard form. Assuming that the standard form contains $n$ variables (labeled for convenience $x_{1}, x_{2}, \ldots, x_{n}$ ), the standard form for such an LP is

$$
\max z=c_{1} x_{1}+c_{2} x_{2}+\cdots+c_{n} x_{n}
$$

(or min)

$$
\begin{array}{cc}
\text { s.t. } & a_{11} x_{1}+a_{12} x_{2}+\cdots+a_{1 n} x_{n}=b_{1} \\
& a_{21} x_{1}+a_{22} x_{2}+\cdots+a_{2 n} x_{n}=b_{2}  \tag{7}\\
\vdots & \vdots \\
& a_{m 1} x_{1}+a_{m 2} x_{2}+\cdots+a_{m n} x_{n}=b_{m} \\
& x_{i} \geq 0 \quad(i=1,2, \ldots, n)
\end{array}
$$

If we define

$$
A=\left[\begin{array}{cccc}
a_{11} & a_{12} & \ldots & a_{1 n} \\
a_{12} & a_{22} & \ldots & a_{2 n} \\
\vdots & \vdots & & \vdots \\
a_{1 n} & a_{m 2} & \cdots & a_{m n}
\end{array}\right]
$$

and

$$
\mathbf{x}=\left[\begin{array}{c}
x_{1} \\
x_{2} \\
\vdots \\
x_{n}
\end{array}\right], \quad \mathbf{b}=\left[\begin{array}{c}
b_{1} \\
b_{2} \\
\vdots \\
b_{m}
\end{array}\right]
$$

the constraints for (7) may be written as the system of equations $A \mathbf{x}=\mathbf{b}$. Before proceeding further with our discussion of the simplex algorithm, we must define the concept of a basic solution to a linear system.

## Basic and Nonbasic Variables

Consider a system $A \mathbf{x}=\mathbf{b}$ of $m$ linear equations in $n$ variables (assume $n \geq m$ ).
A basic solution to $A \mathbf{x}=\mathbf{b}$ is obtained by setting $n-m$ variables equal to 0 and solving for the values of the remaining $m$ variables. This assumes that setting the $n-m$ variables equal to 0 yields unique values for the remaining $m$ variables or, equivalently, the columns for the remaining $m$ variables are linearly independent.

To find a basic solution to $A \mathbf{x}=\mathbf{b}$, we choose a set of $n-m$ variables (the nonbasic variables, or NBV) and set each of these variables equal to 0 . Then we solve for the values of the remaining $n-(n-m)=m$ variables (the basic variables, or BV) that satisfy $A \mathbf{x}=\mathbf{b}$.

Of course, the different choices of nonbasic variables will lead to different basic solutions. To illustrate, we find all the basic solutions to the following system of two equations in three variables:

$$
\begin{align*}
x_{1}+x_{2} & =3 \\
-x_{2}+x_{3} & =-1 \tag{8}
\end{align*}
$$

We begin by choosing a set of $3-2=1$ ( 3 variables, 2 equations) nonbasic variables. For example, if NBV $=\left\{x_{3}\right\}$, then $\mathrm{BV}=\left\{x_{1}, x_{2}\right\}$. We obtain the values of the basic variables by setting $x_{3}=0$ and solving

$$
\begin{aligned}
x_{1}+x_{2} & =3 \\
-x_{2} & =-1
\end{aligned}
$$

We find that $x_{1}=2, x_{2}=1$. Thus, $x_{1}=2, x_{2}=1, x_{3}=0$ is a basic solution to (8). However, if we choose NBV $=\left\{x_{1}\right\}$ and $\mathrm{BV}=\left\{x_{2}, x_{3}\right\}$, we obtain the basic solution $x_{1}=0$, $x_{2}=3, x_{3}=2$. If we choose NBV $=\left\{x_{2}\right\}$, we obtain the basic solution $x_{1}=3, x_{2}=0$, $x_{3}=-1$. The reader should verify these results.

Some sets of $m$ variables do not yield a basic solution. For example, consider the following linear system:

$$
\begin{array}{r}
x_{1}+2 x_{2}+x_{3}=1 \\
2 x_{1}+4 x_{2}+x_{3}=3
\end{array}
$$

If we choose NBV $=\left\{x_{3}\right\}$ and $\mathrm{BV}=\left\{x_{1}, x_{2}\right\}$, the corresponding basic solution would be obtained by solving

$$
\begin{array}{r}
x_{1}+2 x_{2}=1 \\
2 x_{1}+4 x_{2}=3
\end{array}
$$

Because this system has no solution, there is no basic solution corresponding to $\mathrm{BV}=$ $\left\{x_{1}, x_{2}\right\}$.

## Feasible Solutions

A certain subset of the basic solutions to the constraints $A \mathbf{x}=\mathbf{b}$ of an LP plays an important role in the theory of linear programming.

DEFINITION ■ Any basic solution to (7) in which all variables are nonnegative is a basic feasible solution (or bfs).

Thus, for an LP with the constraints given by (8), the basic solutions $x_{1}=2, x_{2}=1$, $x_{3}=0$, and $x_{1}=0, x_{2}=3, x_{3}=2$ are basic feasible solutions, but the basic solution $x_{1}=3, x_{2}=0, x_{3}=-1$ fails to be a basic solution (because $x_{3}<0$ ).

In the rest of this section, we assume that all LPs are in standard form. Recall from Section 3.2 that the feasible region for any LP is a convex set. Let $S$ be the feasible region for an LP in standard form. Recall that a point $P$ is an extreme point of $S$ if all line segments that contain $P$ and are completely contained in $S$ have $P$ as an endpoint. It turns out that the extreme points of an LP's feasible region and the LP's basic feasible solutions are actually one and the same. More formally,

## THEOREM 1

A point in the feasible region of an LP is an extreme point if and only if it is a basic feasible solution to the LP.

See Luenburger (1984) for a proof of Theorem 1.

To illustrate the correspondence between extreme points and basic feasible solutions outlined in Theorem 1, let's look at the Leather Limited example of Section 4.1. Recall that the LP was

$$
\begin{array}{lr}
\max z=4 x_{1}+3 x_{2} & \\
\text { s.t. } \quad x_{1}+x_{2} & \leq 40 \\
2 x_{1}+x_{2} & \leq 60 \\
x_{1}, x_{2} & \geq 0 \tag{2}
\end{array}
$$

By adding slack variables $s_{1}$ and $s_{2}$, respectively, to (1) and (2), we obtain LP 1 in standard form:

$$
\begin{array}{lrl}
\max z=4 x_{1}+3 x_{2} & \\
\text { s.t. } \quad x_{1}+x_{2}+s_{1} & =40 \\
2 x_{1}+x_{2}+s_{2} & =60  \tag{LP1'}\\
& x_{1}, x_{2}, s_{1}, s_{2} & \geq 0
\end{array}
$$

The feasible region for the Leather Limited problem is graphed in Figure 1. Both inequalities are satisfied: (1) by all points below or on the line $A B\left(x_{1}+x_{2}=40\right)$, and (2) by all points on or below the line $C D\left(2 x_{1}+x_{2}=60\right)$. Thus, the feasible region for LP 1 is the shaded region bounded by the quadrilateral $B E C F$. The extreme points of the feasible region are $B=(0,40), C=(30,0), E=(20,20)$, and $F=(0,0)$.

FIGURE 1
Feasible Region for Leather Limited


Table 1 shows the correspondence between the basic feasible solutions to LP $1^{\prime}$ and the extreme points of the feasible region for LP 1. This example should make it clear that the basic feasible solutions to the standard form of an LP correspond in a natural fashion to the LP's extreme points.

In the context of the Leather Limited example, it is easy to show why any bfs is an extreme point. The converse is harder! We now show that for the LL problem, any bfs is an extreme point. Any point in the feasible region for LL may be specified as a fourdimensional column vector with the four elements of the vector denoting $x_{1}, x_{2}, s_{1}$, and $s_{2}$, respectively. Consider the bfs $B$ with $\mathrm{BV}=\left\{x_{2}, s_{2}\right\}$. If $B$ is not an extreme point, then there exists two distinct feasible points $v_{1}$ and $v_{2}$ and non-negative numbers $\sigma_{1}$ and $\sigma_{2}$ satisfying $0<\sigma_{i}<1$ and $\sigma_{1}+\sigma_{2}=1$ such that

$$
\left[\begin{array}{r}
0 \\
40 \\
0 \\
20
\end{array}\right]=\sigma_{1} v_{1}+\sigma_{2} v_{2}
$$

Clearly, both $v_{1}$ and $v_{2}$ must both have $x_{1}=s_{2}=0$. But because $v_{1}$ and $v_{2}$ are both feasible, the values of $x_{2}$ and $s_{2}$ for both $v_{1}$ and $v_{2}$ can be determined by solving $x_{2}=40$ and $x_{2}+s_{2}=60$. These equations have a unique solution (because columns corresponding to basic variables $x_{2}$ and $s_{2}$ are linearly independent). This shows that $v_{1}=v_{2}$, so B is indeed an extreme point.

TABLE 1
Correspondence between Basic Feasible Solutions and Corner Points for Leather Limited

| Basic <br> Variables | Nonbasic <br> Variables | Basic <br> Feasible Solution | Corresponds to <br> Corner Point |
| :--- | :---: | :---: | :--- |
| $x_{1}, x_{2}$ | $s_{1}, s_{2}$ | $s_{1}=s_{2}=0, x_{1}=x_{2}=20$ | $E$ |
| $x_{1}, s_{1}$ | $x_{2}, s_{2}$ | $x_{2}=s_{2}=0, x_{1}=30, s_{1}=10$ | $C$ |
| $x_{1}, s_{2}$ | $x_{2}, s_{1}$ | $x_{2}=s_{1}=0, x_{1}=40, s_{2}=-20$ | Not a bfs because $s_{2}<0$ |
| $x_{2}, s_{1}$ | $x_{1}, s_{2}$ | $x_{1}=s_{2}=0, s_{1}=-20, x_{2}=60$ | Not a bfs because $s_{1}<0$ |
| $x_{2}, s_{2}$ | $x_{1}, s_{1}$ | $x_{1}=s_{1}=0, x_{2}=40, s_{2}=20$ | $B$ |
| $s_{1}, s_{2}$ | $x_{1}, x_{2}$ | $x_{1}=x_{2}=0, s_{1}=40, s_{2}=60$ | $F$ |

We note that more than one set of basic variables may correspond to a given extreme point. If this is the case, then we say the LP is degenerate. See Section 4.11 for a discussion of the impact of degeneracy on the simplex algorithm.

We will soon see that if an LP has an optimal solution, then it has a bfs that is optimal. This is important because any LP has only a finite number of bfs's. Thus we can find the optimal solution to an LP by searching only a finite number of points. Because the feasible region for any LP contains an infinite number of points, this helps us a lot!

Before explaining why any LP that has an optimal solution has an optimal bfs, we need to define the concept of a direction of unboundedness.

### 4.3 Direction of Unboundedness

Consider an LP in standard form with feasible region $S$ and constraints $A \mathbf{x}=\mathbf{b}$ and $\mathbf{x} \geq$ $\mathbf{0}$. Assuming that our LP has $n$ variables, $\mathbf{0}$ represents an $n$-dimensional column vector consisting of all 0 's. A nonzero vector $\mathbf{d}$ is a direction of unboundedness if for all $\mathbf{x} \in S$ and any $c \geq 0, x+c \mathbf{d} \in S$. In short, if we are in the LP's feasible region, then we can move as far as we want in the direction $\mathbf{d}$ and remain in the feasible region. Figure 2 displays the feasible region for the Dorian Auto example (Example 2 of Chapter 3). In standard form, the Dorian example is

$$
\begin{gathered}
\min z=50 x_{1}+100 x_{2} \\
7 x_{1}+2 x_{2}-e_{1}=28 \\
2 x_{1}+12 x_{2}-e_{2}=24 \\
x_{1}, x_{2}, e_{1}, e_{2} \geq 0
\end{gathered}
$$

Looking at Figure 2 it is clear that if we start at any feasible point and move up and to the right at a 45-degree angle, we will remain in the feasible region. This means that

FIGURE 2 Graphical Solution of Dorian Problem


$$
d=\left[\begin{array}{r}
1 \\
1 \\
9 \\
14
\end{array}\right]
$$

is a direction of unboundedness for this LP. It is easy to show (see Problem 6) that $\mathbf{d}$ is a direction of unboundedness if and only if $A \mathbf{d}=0$ and $\mathbf{d} \geq \mathbf{0}$.

The following Representation Theorem [for a proof, see Nash and Sofer (1996)] is the key insight needed to show why any LP with an optimal solution has an optimal bfs.

## THEOREM 2

Consider an LP in standard form, having bfs $\mathbf{b}_{1}, \mathbf{b}_{2}, \ldots, \mathbf{b}_{k}$. Any point $\mathbf{x}$ in the LP's feasible region may be written in the form

$$
\mathbf{x}=\mathbf{d}+\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}
$$

where $\mathbf{d}$ is $\mathbf{0}$ or a direction of unboundedness and $\sum_{i=1}^{i=k} \sigma_{i}=1$ and $\sigma_{i} \geq 0$.

If the LP's feasible region is bounded, then $\mathbf{d}=\mathbf{0}$, and we may write $\mathbf{x}=\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}$, where the $\sigma_{i}$ are nonnegative weights adding to 1 . In this case, we see that any feasible $\mathbf{x}$ may be written as a convex combination of the LP's bfs. We now give two illustrations of Theorem 2.

Consider the Leather Limited example. The feasible region is bounded. To illustrate Theorem 2, we can write the point $G=(20,10)(G$ is not a bfs!) in Figure 3 as a convex combination of the LP's bfs. Note from Figure 3 that point $G$ may be written as $\frac{1}{6} F+\frac{5}{6} H$ [here $\left.H=(24,12)\right]$. Then note that point $H$ may be written as $.6 E+.4 C$. Putting these two relationships together, we may write point $G$ as $\frac{1}{6} F+\frac{5}{6}(.6 E+.4 C)=\frac{1}{6} F+$ $\frac{1}{2} E+\frac{1}{3} C$. This expresses point $G$ as a convex combination of the LP's extreme points.

To illustrate Theorem 2 for an unbounded LP, let's consider Example 2 of Chapter 3 (the Dorian example; see Figure 4) and try to express the point $F=(14,4)$ in the representation given in Theorem 2. Recall that in standard form the constraints for the Dorian example are given by

$$
7 x_{1}+2 x_{2}-e_{1}=28
$$

FIGURE 3 Writing $(20,10)$ as a Convex Combination of bfs

FIGURE 4 Expressing $F=(14,4)$ Using Theorem 2


$$
2 x_{1}+12 x_{2}-e_{2}=24
$$

From Figure 4, we see that to move from bfs $C$ to point $F$ we need to move up and to the right along a line having slope $\frac{4-0}{14-12}=2$. This line corresponds to the direction of unboundedness

$$
\mathbf{d}=\left[\begin{array}{r}
2 \\
4 \\
22 \\
52
\end{array}\right]
$$

Letting

$$
\mathbf{b}_{1}=\left[\begin{array}{r}
12 \\
0 \\
56 \\
0
\end{array}\right] \quad \text { and } \quad \mathbf{x}=\left[\begin{array}{r}
14 \\
4 \\
78 \\
52
\end{array}\right]
$$

we may write $\mathbf{x}=\mathbf{d}+\mathbf{b}_{1}$, which is the desired representation.

### 4.4 Why Does an LP Have an Optimal bis?

Consider an LP with objective function max $\mathbf{c x}$ and constraints $A \mathbf{x}=\mathbf{b}$. Suppose this LP

## THEOREM 3

has an optimal solution. We now sketch a proof of the fact that the LP has an optimal bfs.

If an LP has an optimal solution, then it has an optimal bfs.
Proof Let $\mathbf{x}$ be an optimal solution to our LP. Because $\mathbf{x}$ is feasible, Theorem 2 tells us that we may write $\mathbf{x}=\mathbf{d}+\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}$, where $\mathbf{d}$ is $\mathbf{0}$ or a direction of unboundedness and $\mathbf{b}_{1}, \mathbf{b}_{2}, \ldots, \mathbf{b}_{k}$ are the LP's bfs. Also, $\Sigma_{i=1}^{i=k} \sigma_{i}=1$ and $\sigma_{i} \geq 0$. If $\mathbf{c d}>$
$\mathbf{0}$, then for any $k>0, k \mathbf{d}+\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}$ is feasible, and as $k$ grows larger and larger, the objective function value approaches infinity. This contradicts the fact that the LP has an optimal solution. If $\mathbf{c d}<\mathbf{0}$, then the feasible point $\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}$ has a larger objective function value than $\mathbf{x}$. This contradicts the optimality of $\mathbf{x}$. In short, we have shown that if $\mathbf{x}$ is optimal, then $\mathbf{c d}=0$. Now the objective function value for $\mathbf{x}$ is given by

$$
\mathbf{c x}=\mathbf{c d}+\sum_{i=1}^{i=k} \sigma_{i} \mathbf{c b}_{i}=\sum_{i=1}^{i=k} \sigma_{i} \mathbf{c b}_{i}
$$

Suppose that $\mathbf{b}_{1}$ is the bfs with the largest objective function value. Because $\sum_{i=1}^{i=k}$ $\sigma_{i}=1$ and $\sigma_{i} \geq 0$,

$$
\mathbf{c b}_{1} \geq \mathbf{c x}
$$

Because $\mathbf{x}$ is optimal, this shows that $\mathbf{b}_{1}$ is also optimal, and the LP does indeed have an optimal bfs.

## Adjacent Basic Feasible Solutions

Before describing the simplex algorithm in general terms, we need to define the concept of an adjacent basic feasible solution.

For any LP with $m$ constraints, two basic feasible solutions are said to be adjacent if their sets of basic variables have $m-1$ basic variables in common.

For example, in Figure 3, two basic feasible solutions will be adjacent if they have $2-1=1$ basic variable in common. Thus, the bfs corresponding to point $E$ in Figure 3 is adjacent to the bfs corresponding to point $C$. Point $E$ is not, however, adjacent to bfs $F$. Intuitively, two basic feasible solutions are adjacent if they both lie on the same edge of the boundary of the feasible region.

We now give a general description of how the simplex algorithm solves LPs in a max problem.

Step 1 Find a bfs to the LP. We call this bfs the initial basic feasible solution. In general, the most recent bfs will be called the current bfs, so at the beginning of the problem the initial bfs is the current bfs.

Step 2 Determine if the current bfs is an optimal solution to the LP. If it is not, then find an adjacent bfs that has a larger $z$-value.
Step 3 Return to step 2, using the new bfs as the current bfs.
If an LP in standard form has $m$ constraints and $n$ variables, then there may be a basic solution for each choice of nonbasic variables. From $n$ variables, a set of $n-m$ nonbasic variables (or equivalently, $m$ basic variables) can be chosen in

$$
\binom{n}{m}=\frac{n!}{}
$$

different ways. Thus, an LP can have at most

$$
\binom{n}{m}
$$

basic solutions. Because some basic solutions may not be feasible, an LP can have at most

$$
\binom{n}{m}
$$

basic feasible solutions. If we were to proceed from the current bfs to a better bfs (without ever repeating a bfs), then we would surely find the optimal bfs after examining at most

$$
\binom{n}{m}
$$

basic feasible solutions. This means (assuming that no bfs is repeated) that the simplex algorithm will find the optimal bfs after a finite number of calculations. We return to this discussion in Section 4.11.

In principle, we could enumerate all basic feasible solutions to an LP and find the bfs with the largest $z$-value. The problem with this approach is that even small LPs have a very large number of basic feasible solutions. For example, an LP in standard form that has 20 variables and 10 constraints might have (if each basic solution were feasible) up to

$$
\binom{20}{10}=184,756
$$

basic feasible solutions. Fortunately, vast experience with the simplex algorithm indicates that when this algorithm is applied to an $n$-variable, $m$-constraint LP in standard form, an optimal solution is usually found after examining fewer than $3 m$ basic feasible solutions. Thus, for a 20 -variable, 10 -constraint LP in standard form, the simplex will usually find the optimal solution after examining fewer than $3(10)=30$ basic feasible solutions. Compared with the alternative of examining 184,756 basic solutions, the simplex is quite efficient! ${ }^{\dagger}$

## Geometry of Three-Dimensional LPs

Consider the following LP:

$$
\begin{aligned}
\max z=x_{1}+2 x_{2}+2 x_{3} & \\
\text { s.t. } & \leq 8 \\
2 x_{1}+x_{2} & \leq 8 \\
x_{3} & \leq 10 \\
x_{1}, x_{2}, x_{3} & \geq 0
\end{aligned}
$$

The set of points satisfying a linear inequality in three (or any number of) dimensions is a half-space. For example, the set of points in three dimensions satisfying $2 x_{1}+x_{2} \leq$ 8 is a half-space. Thus, the feasible region for our LP is the intersection of the following five half-spaces: $2 x_{1}+x_{2} \leq 8, x_{3} \leq 10, x_{1} \geq 0, x_{2} \geq 0$, and $x_{3} \geq 0$. The intersection of half-spaces is called a polyhedron. The feasible region for our LP is the prism pictured in Figure 5.

On each face (or facet) of the feasible region, one constraint (or sign restriction) is binding for all points on that face. For example, the constraint $2 x_{1}+x_{2} \leq 8$ is binding for all points on the face $A B C D ; x_{3} \geq 0$ is binding on face $A B F ; x_{3} \leq 10$ is binding on face $D E C ; x_{2} \geq 0$ is binding on face $A D E F ; x_{1} \geq 0$ is binding on face $C B F E$.

Clearly, the corner (or extreme) points of the LP's feasible region are $A, B, C, D, E$, and $F$. In this case, the correspondence between the bfs and corner points is as shown in Table 2.

To illustrate the concept of adjacent basic feasible solutions, note that corner points $A$,

[^5]FIGURE 5 Feasible Region in
Three Dimensions


TABLE 2
Correspondence between bfs and Corner Points

| Basic <br> Variables | Basic Feasible Solution | Corresponds to <br> Corner Point |
| :--- | :---: | :---: |
| $x_{1}, x_{3}$ | $x_{1}=4, x_{3}=10, x_{2}=s_{1}=s_{2}=0$ | $D$ |
| $s_{1}, s_{2}$ | $s_{1}=8, s_{2}=10, x_{1}=x_{2}=x_{3}=0$ | $F$ |
| $s_{1}, x_{3}$ | $s_{1}=8, x_{3}=10, x_{1}=x_{2}=s_{2}=0$ | $E$ |
| $x_{2}, x_{3}$ | $x_{2}=8, x_{3}=10, x_{1}=s_{1}=s_{2}=0$ | $C$ |
| $x_{2}, s_{2}$ | $x_{2}=8, s_{2}=10, x_{1}=x_{3}=s_{1}=0$ | $B$ |
| $x_{1}, s_{2}$ | $x_{1}=4, s_{2}=10, x_{2}=x_{3}=s_{1}=0$ | $A$ |

$E$, and $B$ are adjacent to corner point $F$. Thus, if the simplex algorithm begins at $F$, then we can be sure that the next bfs to be considered will be $A, E$, or $B$.

## PROBLEMS

## Group A

1 For the Giapetto problem (Example 1 in Chapter 3), show how the basic feasible solutions to the LP in standard form correspond to the extreme points of the feasible region.
2 For the Dorian problem (Example 2 in Chapter 3), show how the basic feasible solutions to the LP in standard form correspond to the extreme points of the feasible region.

3 Widgetco produces two products: 1 and 2. Each requires the amounts of raw material and labor, and sells for the price given in Table 3.

Up to 350 units of raw material can be purchased at $\$ 2$ per unit, while up to 400 hours of labor can be purchased at $\$ 1.50$ per hour. To maximize profit, Widgetco must solve the following LP:

$$
\begin{array}{lr}
\max z=2 x_{1}+2.5 x_{2} \\
\text { s.t. } & x_{1}+2 x_{2} \leq 350 \quad \text { (Raw material) }
\end{array}
$$

TABLE 3

|  | Product 1 | Product 2 |
| :--- | :--- | :--- |
| Raw material | 1 unit | 2 units |
| Labor | 2 hours | 1 hour |
| Sales price | $\$ 7$ | $\$ 8$ |

$$
\begin{aligned}
2 x_{1}+x_{2} & \leq 400 \quad \text { (Labor) } \\
x_{1}, x_{2} & \geq 0
\end{aligned}
$$

Here, $x_{i}=$ number of units of product $i$ produced. Demonstrate the correspondence between corner points and basic feasible solutions.
4 For the Leather Limited problem, represent the point $(10,20)$ in the form $\mathbf{c d}+\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}$.
5 For the Dorian problem, represent the point $(10,40)$ in the form $\mathbf{c d}+\sum_{i=1}^{i=k} \sigma_{i} \mathbf{b}_{i}$.

## Group B

6 For an LP in standard form with constraints $A \mathbf{x}=\mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$, show that $\mathbf{d}$ is a direction of unboundedness if and
only if $A \mathbf{d}=0$ and $\mathbf{d} \geq \mathbf{0}$.
7 Recall that Example 5 of Chapter 3 is an unbounded LP. Find a direction of unboundedness along which we can move for which the objective function becomes arbitrarily large.

### 4.5 The Simplex Algorithm

We now describe how the simplex algorithm can be used to solve LPs in which the goal is to maximize the objective function. The solution of minimization problems is discussed in Section 4.4.

The simplex algorithm proceeds as follows:
Step 1 Convert the LP to standard form (see Section 4.1).
Step 2 Obtain a bfs (if possible) from the standard form.
Step 3 Determine whether the current bfs is optimal.
Step 4 If the current bfs is not optimal, then determine which nonbasic variable should become a basic variable and which basic variable should become a nonbasic variable to find a new bfs with a better objective function value.

Step 5 Use EROs to find the new bfs with the better objective function value. Go back to step 3.

In performing the simplex algorithm, write the objective function

$$
z=c_{1} x_{1}+c_{2} x_{2}+\cdots+c_{n} x_{n}
$$

## EXAMPLE 2

in the form
Dakota Furniture Company
$z-c_{1} x_{1}-c_{2} x_{2}-\cdots-c_{n} x_{n}=0$
We call this format the row 0 version of the objective function (row 0 for short).

The Dakota Furniture Company manufactures desks, tables, and chairs. The manufacture of each type of furniture requires lumber and two types of skilled labor: finishing and carpentry. The amount of each resource needed to make each type of furniture is given in Table 4.

Currently, 48 board feet of lumber, 20 finishing hours, and 8 carpentry hours are available. A desk sells for $\$ 60$, a table for $\$ 30$, and a chair for $\$ 20$. Dakota believes that demand for desks and chairs is unlimited, but at most five tables can be sold. Because the available resources have already been purchased, Dakota wants to maximize total revenue.

TABLE 4
Resource Requirements for Dakota Furniture

| Resource | Desk | Table | Chair |
| :--- | :---: | :--- | :--- |
| Lumber (board ft) | 8 | 6 | 1 |
| Finishing hours | 4 | 2 | 1.5 |
| Carpentry hours | 2 | 1.5 | 0.5 |

Defining the decision variables as

$$
\begin{aligned}
& x_{1}=\text { number of desks produced } \\
& x_{2}=\text { number of tables produced } \\
& x_{3}=\text { number of chairs produced }
\end{aligned}
$$

it is easy to see that Dakota should solve the following LP:

$$
\begin{aligned}
& \max z=60 x_{1}+30 x_{2}+20 x_{3} \\
& \text { s.t. } \quad 8 x_{1}+6 x_{2}+x_{3} \leq 48 \quad \text { (Lumber constraint) }
\end{aligned}
$$

$$
\begin{aligned}
4 x_{1}+2 x_{2}+1.5 x_{3} & \leq 20 & & \text { (Finishing constraint) } \\
2 x_{1}+1.5 x_{2}+0.5 x_{3} & \leq 8 & & \text { (Carpentry constraint) } \\
x_{2} & \leq 5 & & \text { (Limitation on table demand) } \\
x_{1}, x_{2}, x_{3} & \geq 0 & &
\end{aligned}
$$

## Convert the LP to Standard Form

We begin the simplex algorithm by converting the constraints of the LP to the standard form discussed in Section 4.1. Then we convert the LP's objective function to the row 0 format. To put the constraints in standard form, we simply add slack variables $s_{1}, s_{2}, s_{3}$, and $s_{4}$, respectively, to the four constraints. We label the constraints row 1 , row 2 , row 3 , and row 4 , and add the sign restrictions $s_{i} \geq 0(i=1,2,3,4)$. Note that the row 0 format for our objective function is

$$
z-60 x_{1}-30 x_{2}-20 x_{3}=0
$$

Putting rows $1-4$ together with row 0 and the sign restrictions yields the equations and basic variables shown in Table 5. A system of linear equations (such as canonical form 0 , shown in Table 5) in which each equation has a variable with a coefficient of 1 in that equation (and a zero coefficient in all other equations) is said to be in canonical form. We will soon see that if the right-hand side of each constraint in a canonical form is nonnegative, a basic feasible solution can be obtained by inspection. ${ }^{\dagger}$

TABLE 5
Canonical Form 0

| Row |  |  | Basic <br> Variable |  |
| :--- | :---: | :--- | :--- | :--- |
| 0 | $z-60 x_{1}-30 x_{2}-20 x_{3}$ |  | $=0$ | $z=0$ |
| 1 | $8 x_{1}+6 x_{2}+x_{3}+s_{1}$ |  | $=48$ | $s_{1}=48$ |
| 2 | $4 x_{1}+2 x_{2}+1.5 x_{3}+s_{2}$ | $=20$ | $s_{2}=20$ |  |
| 3 | $2 x_{1}+1.5 x_{2}+0.5 x_{3}$ | $+s_{3}$ | $=8$ | $s_{3}=8$ |
| 4 | $x_{2}$ |  | $+s_{4}=5$ | $s_{4}=5$ |

[^6]From Section 4.2, we know that the simplex algorithm begins with an initial basic feasible solution and attempts to find better ones. After obtaining a canonical form, we therefore search for the initial bfs. By inspection, we see that if we set $x_{1}=x_{2}=x_{3}=0$, we can solve for the values of $s_{1}, s_{2}, s_{3}$, and $s_{4}$ by setting $s_{i}$ equal to the right-hand side of row $i$.

$$
\mathrm{BV}=\left\{s_{1}, s_{2}, s_{3}, s_{4}\right\} \quad \text { and } \quad \mathrm{NBV}=\left\{x_{1}, x_{2}, x_{3}\right\}
$$

The basic feasible solution for this set of basic variables is $s_{1}=48, s_{2}=20, s_{3}=8$, $s_{4}=5, x_{1}=x_{2}=x_{3}=0$. Observe that each basic variable may be associated with the row of the canonical form in which the basic variable has a coefficient of 1 . Thus, for canonical form $0, s_{1}$ may be thought of as the basic variable for row 1 , as may $s_{2}$ for row $2, s_{3}$ for row 3 , and $s_{4}$ for row 4 .

To perform the simplex algorithm, we also need a basic (although not necessarily nonnegative) variable for row 0 . Because $z$ appears in row 0 with a coefficient of 1 , and $z$ does not appear in any other row, we use $z$ as its basic variable. With this convention, the basic feasible solution for our initial canonical form has

$$
\mathrm{BV}=\left\{z, s_{1}, s_{2}, s_{3}, s_{4}\right\} \quad \text { and } \quad \mathrm{NBV}=\left\{x_{1}, x_{2}, x_{3}\right\}
$$

For this basic feasible solution, $z=0, s_{1}=48, s_{2}=20, s_{3}=8, s_{4}=5, x_{1}=x_{2}=x_{3}=0$.
As this example indicates, a slack variable can be used as a basic variable for an equation if the right-hand side of the constraint is nonnegative.

## Is the Current Basic Feasible Solution Optimal?

Once we have obtained a basic feasible solution, we need to determine whether it is optimal; if the bfs is not optimal, then we try to find a bfs adjacent to the initial bfs with a larger $z$-value. To do this, we try to determine whether there is any way that $z$ can be increased by increasing some nonbasic variable from its current value of zero while holding all other nonbasic variables at their current values of zero. If we solve for $z$ by rearranging row 0 , then we obtain

$$
\begin{equation*}
z=60 x_{1}+30 x_{2}+20 x_{3} \tag{9}
\end{equation*}
$$

For each nonbasic variable, we can use (9) to determine whether increasing a nonbasic variable (and holding all other nonbasic variables at zero) will increase $z$. For example, suppose we increase $x_{1}$ by (holding the other nonbasic variables $x_{2}$ and $x_{3}$ at zero). Then (9) tells us that $z$ will increase by 60 . Similarly, if we choose to increase $x_{2}$ by 1 (holding $x_{1}$ and $x_{3}$ at zero), then (9) tells us that $z$ will increase by 30 . Finally, if we choose to increase $x_{3}$ by 1 (holding $x_{1}$ and $x_{2}$ at zero), then (9) tells us that $z$ will increase by 20. Thus, increasing any of the nonbasic variables will increase $z$. Because a unit increase in $x_{1}$ causes the largest rate of increase in $z$, we choose to increase $x_{1}$ from its current value of zero. If $x_{1}$ is to increase from its current value of zero, then it will have to become a basic variable. For this reason, we call $x_{1}$ the entering variable. Observe that $x_{1}$ has the most negative coefficient in row 0 .

## Determine the Entering Variable

We choose the entering variable (in a max problem) to be the nonbasic variable with the most negative coefficient in row 0 (ties may be broken in an arbitrary fashion). Because each one-unit increase of $x_{1}$ increases $z$ by 60 , we would like to make $x_{1}$ as large as possible. What limits how large we can make $x_{1}$ ? Note that as $x_{1}$ increases, the values of the
current basic variables ( $s_{1}, s_{2}, s_{3}$, and $s_{4}$ ) will change. This means that increasing $x_{1}$ may cause a basic variable to become negative. With this in mind, we look at how increasing $x_{1}$ (while holding $x_{2}=x_{3}=0$ ) changes the values of the current set of basic variables. From row 1, we see that $s_{1}=48-8 x_{1}$ (remember that $x_{2}=x_{3}=0$ ). Because the sign restriction $s_{1} \geq 0$ must be satisfied, we can only increase $x_{1}$ as long as $s_{1} \geq 0$, or $48-$ $8 x_{1} \geq 0$, or $x_{1} \leq \frac{48}{8}=6$. From row $2, s_{2}=20-4 x_{1}$. We can only increase $x_{1}$ as long as $s_{2} \geq 0$, so $x_{1}$ must satisfy $20-4 x_{1} \geq 0$ or $x_{1} \leq \frac{20}{4}=5$. From row $3, s_{3}=8-2 x_{1}$ so $x_{1} \leq \frac{8}{2}=4$. Similarly, we see from row 4 that $s_{4}=5$. Thus, whatever the value of $x_{1}, s_{4}$ will be nonnegative. Summarizing,

$$
\begin{array}{ll}
s_{1} \geq 0 & \text { for }
\end{array} x_{1} \leq \frac{48}{8}=6
$$

This means that to keep all the basic variables nonnegative, the largest that we can make $x_{1}$ is $\min \left\{\frac{48}{8}, \frac{20}{4}, \frac{8}{2}\right\}=4$. If we make $x_{1}>4$, then $s_{3}$ will become negative, and we will no longer have a basic feasible solution. Notice that each row in which the entering variable had a positive coefficient restricted how large the entering variable could become. Also, for any row in which the entering variable had a positive coefficient, the row's basic variable became negative when the entering variable exceeded

> Right-hand side of row

If the entering variable has a nonpositive coefficient in a row (such as $x_{1}$ in row 4), the row's basic variable will remain positive for all values of the entering variable. Using (10), we can quickly compute how large $x_{1}$ can become before a basic variable becomes negative.

$$
\begin{aligned}
& \text { Row } 1 \text { limit on } x_{1}=\frac{48}{8}=6 \\
& \text { Row } 2 \text { limit on } x_{1}=\frac{20}{4}=5 \\
& \text { Row } 3 \text { limit on } x_{1}=\frac{8}{2}=4 \\
& \text { Row } 4 \text { limit on } x_{1}=\text { no limit } \quad \text { (Because coefficient of } x_{1} \text { in row } 4 \text { is nonpositive) }
\end{aligned}
$$

We can state the following rule for determining how large we can make an entering variable.

## The Ratio Test

When entering a variable into the basis, compute the ratio in (10) for every constraint in which the entering variable has a positive coefficient. The constraint with the smallest ratio is called the winner of the ratio test. The smallest ratio is the largest value of the entering variable that will keep all the current basic variables nonnegative. In our example, row 3 was the winner of the ratio test for entering $x_{1}$ into the basis.

## Find a New Basic Feasible Solution: Pivot in the Entering Variable

Returning to our example, we know that the largest we can make $x_{1}$ is 4. For $x_{1}$ to equal 4, it must become a basic variable. Looking at rows $1-4$, we see that if we make $x_{1}$ a basic variable in row 1 , then $x_{1}$ will equal $\frac{48}{8}=6$; in row 2 , $x_{1}$ will equal $\frac{20}{4}=5$; in row $3, x_{1}$ will equal $\frac{8}{2}=4$. Also, because $x_{1}$ does not appear in row $4, x_{1}$ cannot be made a basic variable in row 4 . Thus, if we want to make $x_{1}=4$, then we have to make it a basic variable in row 3 . The fact that row 3 was the winner of the ratio test illustrates the following rule.

## In Which Row Does the Entering Variable Become Basic?

Always make the entering variable a basic variable in a row that wins the ratio test (ties may be broken arbitrarily).

To make $x_{1}$ a basic variable in row 3 , we use elementary row operations to make $x_{1}$ have a coefficient of 1 in row 3 and a coefficient of 0 in all other rows. This procedure is called pivoting on row 3; and row 3 is the pivot row. The final result is that $x_{1}$ replaces $s_{3}$ as the basic variable for row 3. The term in the pivot row that involves the entering basic variable is called the pivot term. Proceeding as we did when we studied the Gauss-Jordan method in Chapter 2, we make $x_{1}$ a basic variable in row 3 by performing the following EROs.
ERO 1 Create a coefficient of 1 for $x_{1}$ in row 3 by multiplying row 3 by $\frac{1}{2}$. The resulting row (marked with a prime to show it is the first iteration) is

$$
\begin{equation*}
x_{1}+0.75 x_{2}+0.25 x_{3}+0.5 s_{3}=4 \tag{row3'}
\end{equation*}
$$

ERO 2 To create a zero coefficient for $x_{1}$ in row 0 , replace row 0 with 60 (row $3^{\prime}$ ) + row 0 .

$$
\begin{equation*}
z+15 x_{2}-5 x_{3}+30 s_{3}=240 \tag{row0'}
\end{equation*}
$$

ERO 3 To create a zero coefficient for $x_{1}$ in row 1 , replace row 1 with -8 (row $3^{\prime}$ ) + row 1 .

$$
\begin{equation*}
-x_{3}+s_{1}-4 s_{3}=16 \tag{row1'}
\end{equation*}
$$

ERO 4 To create a zero coefficient for $x_{1}$ in row 2, replace row 2 with -4 (row $3^{\prime}$ ) + row 2 .

$$
-x_{2}+0.5 x_{3}+s_{2}-2 s_{3}=4
$$

(row 2')
Because $x_{1}$ does not appear in row 4, we don't need to perform an ero to eliminate $x_{1}$ from row 4. Thus, we may write the "new" row 4 (call it row 4' to be consistent with other notation) as

$$
x_{2}+s_{4}=5
$$

(row 4')
Putting rows $0^{\prime}-4^{\prime}$ together, we obtain the canonical form shown in Table 6.
Looking for a basic variable in each row of the current canonical form, we find that

$$
\mathrm{BV}=\left\{z, s_{1}, s_{2}, \mathbf{x}_{1}, s_{4}\right\} \quad \text { and } \quad \mathrm{NBV}=\left\{\mathbf{s}_{3}, x_{2}, x_{3}\right\}
$$

table 6
Canonical Form 1


Thus, canonical form 1 yields the basic feasible solution $z=240, s_{1}=16, s_{2}=4, x_{1}=$ $4, s_{4}=5, x_{2}=x_{3}=s_{3}=0$. We could have predicted that the value of $z$ in canonical form 1 would be 240 from the fact that each unit by which $x_{1}$ is increased increases $z$ by 60. Because $x_{1}$ was increased by 4 units (from $x_{1}=0$ to $x_{1}=4$ ), we would expect that

$$
\begin{aligned}
\text { Canonical form } 1 z \text {-value } & =\text { initial } z \text {-value }+4(60) \\
& =0+240=240
\end{aligned}
$$

In obtaining canonical form 1 from the initial canonical form, we have gone from one bfs to a better (larger $z$-value) bfs. Note that the initial bfs and the improved bfs are adjacent. This follows because the two basic feasible solutions have $4-1=3$ basic variables ( $s_{1}, s_{2}$, and $s_{4}$ ) in common (excluding $z$, which is a basic variable in every canonical form). Thus, we see that in going from one canonical form to the next, we have proceeded from one bfs to a better adjacent bfs. The procedure used to go from one bfs to a better adjacent bfs is called an iteration (or sometimes, a pivot) of the simplex algorithm.

We now try to find a bfs that has a still larger $z$-value. We begin by examining canonical form 1 (Table 6) to see if we can increase $z$ by increasing the value of some nonbasic variable (while holding all other nonbasic variables equal to zero). Rearranging row $0^{\prime}$ to solve for $z$ yields

$$
\begin{equation*}
z=240-15 x_{2}+5 x_{3}-30 s_{3} \tag{11}
\end{equation*}
$$

From (11), we see that increasing the nonbasic variable $x_{2}$ by 1 (while holding $x_{3}=s_{3}=$ 0 ) will decrease $z$ by 15 . We don't want to do that! Increasing the nonbasic variable $s_{3}$ by 1 (holding $x_{2}=x_{3}=0$ ) will decrease $z$ by 30. Again, we don't want to do that. On the other hand, increasing $x_{3}$ by 1 (holding $x_{2}=s_{3}=0$ ) will increase $z$ by 5 . Thus, we choose to enter $x_{3}$ into the basis. Recall that our rule for determining the entering variable is to choose the variable with the most negative coefficient in the current row 0 . Because $x_{3}$ is the only variable with a negative coefficient in row $0^{\prime}$, it should be entered into the basis.

Increasing $x_{3}$ by 1 will increase $z$ by 5 , so it is to our advantage to make $x_{3}$ as large as possible. We can increase $x_{3}$ as long as the current basic variables ( $s_{1}, s_{2}, x_{1}$, and $s_{4}$ ) remain nonnegative. To determine how large $x_{3}$ can be, we must solve for the values of the current basic variables in terms of $x_{3}$ (holding $x_{2}=s_{3}=0$ ). We obtain

$$
\begin{aligned}
& \text { From row } 1^{\prime}: s_{1}=16+x_{3} \\
& \text { From row } 2^{\prime}: s_{2}=4-0.5 x_{3} \\
& \text { From row } 3^{\prime}: x_{1}=4-0.25 x_{3} \\
& \text { From row } 4^{\prime}: s_{4}=5
\end{aligned}
$$

These equations tell us that $s_{1} \geq 0$ and $s_{4} \geq 0$ will hold for all values of $x_{3}$. From row $2^{\prime}$, we see that $s_{2} \geq 0$ will hold if $4-0.5 x_{3} \geq 0$, or $x_{3} \leq \frac{4}{0.5}=8$. From row $3^{\prime}, x_{1} \geq 0$ will hold if $4-0.25 x_{3} \geq 0$, or $x_{3} \leq \frac{4}{0.25}=16$. This shows that the largest we can make $x_{3}$ is $\min \left\{\frac{4}{0.5}, \frac{4}{0.25}\right\}=8$. This fact could also have been discovered by using (10) and the ratio test, as follows:

$$
\begin{aligned}
& \text { Row } 1^{\prime}: \text { no ratio } \\
& \text { Row } 2^{\prime}: \frac{4}{0.5}=8 \\
& \text { Row } 3^{\prime}: \frac{4}{0.25}=16 \\
& \text { Row } 4^{\prime}: \text { no ratio } \quad \quad\left(x_{3} \text { has a nogative coefficient in row } 1\right) \\
&
\end{aligned}
$$

Thus, the smallest ratio occurs in row $2^{\prime}$, and row $2^{\prime}$ wins the ratio test. This means that we should use EROs to make $x_{3}$ a basic variable in row $2^{\prime}$.

TABLE 7
Canonical Form 2

| Row |  |  |  |  | Basic <br> Variable |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $0^{\prime \prime}$ | $z$ | + | $5 x_{2}$ | $+10 s_{2}+10 s_{3}$ | $=280$ | $z=280$ |
| $1^{\prime \prime}$ |  | - | $2 x_{2}$ | $+s_{1}+2 s_{2}-8 s_{3}$ | $=24$ | $s_{1}=24$ |
| $2^{\prime \prime}$ |  | $-2 x_{2}+x_{3}$ | $+2 s_{2}-4 s_{3}$ | $=8$ | $x_{3}=8$ |  |
| $3^{\prime \prime}$ | $x_{1}+1.25 x_{2}$ | $-0.5 s_{2}+1.5 s_{3}$ | $=2$ | $x_{1}=2$ |  |  |
| $4^{\prime \prime}$ |  | $x_{2}$ |  |  |  |  |

ERO 1 Create a coefficient of 1 for $x_{3}$ in row $2^{\prime}$ by replacing row $2^{\prime}$ with 2(row $2^{\prime}$ ):

$$
-2 x_{2}+x_{3}+2 s_{2}-4 s_{3}=8
$$

ERO 2 Create a coefficient of 0 for $x_{3}$ in row $0^{\prime}$ by replacing row $0^{\prime}$ with 5 (row 2$)^{\prime \prime}+$ row 0 ':

$$
z+5 x_{2}+10 s_{2}+10 s_{3}=280
$$

ERO 3 Create a coefficient of 0 for $x_{3}$ in row $1^{\prime}$ by replacing row $1^{\prime}$ with row $2^{\prime \prime}+$ row 1':

$$
-2 x_{2}+s_{1}+2 s_{2}-8 s_{3}=24 \quad \text { (row } 1^{\prime \prime} \text { ) }
$$

ERO 4 Create a coefficient of 0 for $x_{3}$ in row $3^{\prime}$, by replacing row $3^{\prime}$ with $-\frac{1}{4}$ (row $\left.2^{\prime \prime}\right)+3^{\prime}$ :

$$
\left.x_{1}+1.25 x_{2}-0.5 s_{2}+1.5 s_{3}=2 \quad \text { (row } 3^{\prime \prime}\right)
$$

Because $x_{3}$ already has a zero coefficient in row $4^{\prime}$, we may write

$$
x_{2}+s_{4}=5
$$

(row 4")
Combining rows $0^{\prime \prime}-4^{\prime \prime}$ gives the canonical form shown in Table 7 .
Looking for a basic variable in each row of canonical form 2, we find

$$
\mathrm{BV}=\left\{z, s_{1}, x_{3}, x_{1}, s_{4}\right\} \quad \text { and } \quad \mathrm{NBV}=\left\{s_{2}, s_{3}, x_{2}\right\}
$$

Canonical form 2 yields the following bfs: $z=280, s_{1}=24, x_{3}=8, x_{1}=2, s_{4}=5$, $s_{2}=s_{3}=x_{2}=0$. We could have predicted that canonical form 2 would have $z=280$ from the fact that each unit of the entering variable $x_{3}$ increased $z$ by 5 , and we have increased $x_{3}$ by 8 units. Thus,

$$
\text { Canonical form } \begin{aligned}
2 z \text {-value } & =\text { canonical form } 1 z \text {-value }+8(5) \\
& =240+40=280
\end{aligned}
$$

Because the bfs's for canonical forms 1 and 2 have (excluding z) 4-1=3 basic variables in common ( $s_{1}, s_{4}, x_{1}$ ), they are adjacent basic feasible solutions.

Now that the second iteration (or pivot) of the simplex algorithm has been completed, we examine canonical form 2 to see if we can find a better bfs. If we rearrange row $0^{\prime \prime}$ and solve for $z$, we obtain

$$
\begin{equation*}
z=280-5 x_{2}-10 s_{2}-10 s_{3} \tag{12}
\end{equation*}
$$

From (12), we see that increasing $x_{2}$ by 1 (while holding $s_{2}=s_{3}=0$ ) will decrease $z$ by 5 ; increasing $s_{2}$ by 1 (holding $s_{3}=x_{2}=0$ ) will decrease $z$ by 10 ; increasing $s_{3}$ by 1 (holding $x_{2}=s_{2}=0$ ) will decrease $z$ by 10 . Thus, increasing any nonbasic variable will cause $z$ to decrease. This might lead us to believe that our current bfs from canonical form 2 is
an optimal solution. This is indeed correct! To see why, look at (12). We know that any feasible solution to the Dakota Furniture problem must have $x_{2} \geq 0, s_{2} \geq 0$, and $s_{3} \geq 0$, and $-5 x_{2} \leq 0,-10 s_{2} \leq 0$, and $-10 s_{3} \leq 0$. Combining these inequalities with (12), it is clear that any feasible solution must have $z=280+$ terms that are $\leq 0$, and $z \leq 280$. Our current bfs from canonical form 2 has $z=280$, so it must be optimal.

The argument that we just used to show that canonical form 2 is optimal revolved around the fact that each of its nonbasic variables had a nonnegative coefficient in row $0^{\prime \prime}$. This means that we can determine whether a canonical form's bfs is optimal by applying the following simple rule.

## Is a Canonical Form Optimal (Max Problem)?

A canonical form is optimal (for a max problem) if each nonbasic variable has a nonnegative coefficient in the canonical form's row 0 .

REMARKS $\mathbf{1}$ The coefficient of a decision variable in row 0 is often referred to as the variable's reduced cost. Thus, in our optimal canonical form, the reduced costs for $x_{1}$ and $x_{3}$ are 0 , and the reduced cost for $x_{2}$ is 5 . The reduced cost of a nonbasic variable is the amount by which the value of $z$ will decrease if we increase the value of the nonbasic variable by 1 (while all the other nonbasic variables remain equal to 0 ). For example, the reduced cost for the variable "tables" $\left(x_{2}\right)$ in canonical form 2 is 5 . From (12), we see that increasing $x_{2}$ by 1 will reduce $z$ by 5 . Note that because all basic variables (except $z$, of course) must have zero coefficients in row 0 , the reduced cost for a basic variable will always be 0 . In Chapters 5 and 6, we discuss the concept of reduced costs in much greater detail.

These comments are correct only if the values of all the basic variables remain nonnegative after the nonbasic variable is increased by 1 . Increasing $x_{2}$ to 1 leaves $x_{1}, x_{3}$, and $s_{1}$ all nonnegative, so our comments are valid.
2 From canonical form 2, we see that the optimal solution to the Dakota Furniture problem is to manufacture 2 desks $\left(x_{1}=2\right)$ and 8 chairs $\left(x_{3}=8\right)$. Because $x_{2}=0$, no tables should be made. Also, $s_{1}=24$ is reasonable because only $8+8(2)=24$ board feet of lumber are being used. Thus, $48-24=24$ board feet of lumber are not being used. Similarly, $s_{4}=5$ makes sense because, although up to 5 tables could have been produced, 0 tables are actually being produced. Thus, the slack in constraint 4 is $5-0=5$. Because $s_{2}=s_{3}=0$, all available finishing and carpentry hours are being utilized, so the finishing and carpentry constraints are binding.
3 We have chosen the entering variable to be the one with the most negative coefficient in row 0 , but this may not always lead us quickly to the optimal bfs (see Review Problem 11). Actually, even if we choose the variable with the smallest (in absolute value) negative coefficient, the simplex algorithm will eventually find the LP's optimal solution.
4 Although any variable with a negative row 0 coefficient may be chosen to enter the basis, the pivot row must be chosen by the ratio test. To show this formally, suppose that we have chosen to enter $x_{i}$ into the basis, and in the current tableau $x_{i}$ is a basic variable in row $k$. Then row $k$ may be written as

$$
\bar{a}_{k i} x_{i}+\cdots=\bar{b}_{k}
$$

Consider any other constraint (say, row $j$ ) in the canonical form. Row $j$ in the current canonical form may be written as

$$
\bar{a}_{j i} x_{i}+\cdots=\bar{b}_{j}
$$

If we pivot on row $k$, row $k$ becomes

$$
x_{i}+\cdots=\frac{\bar{b}_{k}}{\bar{a}_{k i}}
$$

The new row $j$ after the pivot will be obtained by adding $-\bar{a}_{j i}$ times the last equation to row $j$ of the current canonical form. This yields a new row $j$ of

$$
0 x_{i}+\cdots=\bar{b}_{j}-\frac{\bar{b}_{k} \bar{a}_{j i}}{\bar{a}_{k i}}
$$

We know that after the pivot, each constraint must have a nonnegative right-hand side. Thus, $\bar{a}_{k i}>0$ must hold to ensure that row $k$ has a nonnegative right-hand side after the pivot. Suppose $\bar{a}_{j i}>0$. Then, to ensure that row $j$ will have a nonnegative right-hand side after the pivot, we
must have

$$
\frac{\bar{b}_{j}-\bar{b}_{k} \bar{a}_{j i}}{\bar{a}_{k i}} \geq 0
$$

or (because $\bar{a}_{j i}>0$ )

$$
\frac{\bar{b}_{j}}{\bar{a}_{j i}} \geq \frac{\bar{b}_{k}}{\bar{a}_{k i}}
$$

Thus, row $k$ must be a "winner" of the ratio test to ensure that row $j$ will have a nonnegative righthand side after the pivot is completed.

If $\bar{a}_{j i} \leq 0$, then the right-hand side of row $j$ will surely be nonnegative after the pivot. This follows because

$$
-\frac{\bar{b}_{k} \bar{a}_{j i}}{\bar{a}_{k i}} \geq 0
$$

will now hold.

As promised earlier, we have outlined an algorithm that proceeds from one bfs to a better bfs. The algorithm stops when an optimal solution has been found. The convergence of the simplex algorithm is discussed further in Section 4.11.

## Summary of the Simplex Algorithm for a Max Problem

Step 1 Convert the LP to standard form.
Step 2 Find a basic feasible solution. This is easy if all the constraints are $\leq$ with nonnegative right-hand sides. Then the slack variable $s_{i}$ may be used as the basic variable for row $i$. If no bfs is readily apparent, then use the techniques discussed in Sections 4.12 and 4.13 to find a bfs.

Step 3 If all nonbasic variables have nonnegative coefficients in row 0 , then the current bfs is optimal. If any variables in row 0 have negative coefficients, then choose the variable with the most negative coefficient in row 0 to enter the basis. We call this variable the entering variable.

Step 4 Use EROs to make the entering variable the basic variable in any row that wins the ratio test (ties may be broken arbitrarily). After the EROs have been used to create a new canonical form, return to step 3, using the current canonical form.

When using the simplex algorithm to solve problems, there should never be a constraint with a negative right-hand side (it is okay for row 0 to have a negative right-hand side; see Section 4.6). A constraint with a negative right-hand side is usually the result of an error in the ratio test or in performing one or more EROs. If one (or more) of the constraints has a negative right-hand side, then there is no longer a bfs, and the rules of the simplex algorithm may not lead to a better bfs.

## Representing Simplex Tableaus

Rather than writing each variable in every constraint, we often used a shorthand display called a simplex tableau. For example, the canonical form

$$
\begin{array}{rlr}
z+3 x_{1}+x_{2} & =6 \\
x_{1}+s_{1} & =4
\end{array}
$$

TABLE 8
A Simplex Tableau

| $z$ | $x_{1}$ | $x_{2}$ | $s_{1}$ | $s_{2}$ | rhs | Basic <br> Variable |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 3 | 1 | 0 | 0 | 6 | $z=6$ |
| 0 | 1 | 0 | 1 | 0 | 4 | $s_{1}=4$ |
| 0 | 2 | 1 | 0 | 1 | 3 | $s_{2}=3$ |

$$
2 x_{1}+x_{2} \quad+s_{2}=3
$$

would be written in abbreviated form as shown in Table 8 (rhs $=$ right-hand side). This format makes it very easy to spot basic variables: Just look for columns having a single
entry of 1 and all other entries equal to $0\left(s_{1}\right.$ and $\left.s_{2}\right)$. In our use of simplex tableaus, we will encircle the pivot term and denote the winner of the ratio test by *.

$$
\begin{aligned}
x_{1}-x_{2}+2 x_{3} & \leq 10 \\
x_{1}+x_{2}-x_{3} & \leq 20 \\
x_{1}, x_{2}, x_{3} & \geq 0
\end{aligned}
$$

4 Suppose you want to solve the Dorian problem (Example 2 in Chapter 3) by the simplex algorithm. What difficulty would occur?
5 Use the simplex algorithm to solve the following LP:

$$
\begin{aligned}
\max z=x_{1}+x_{2} & \\
\text { s.t. } \quad 4 x_{1}+x_{2} & \leq 100 \\
x_{1}+x_{2} & \leq 80 \\
x_{1} & \leq 40 \\
x_{1}, x_{2} & \geq 0
\end{aligned}
$$

6 Use the simplex algorithm to solve the following LP:

$$
\begin{array}{ll}
\max z= & x_{1}+x_{2}+x_{3} \\
\text { s.t. } & x_{1}+2 x_{2}+2 x_{3} \leq 20
\end{array}
$$

$$
\begin{aligned}
& \max z=2 x_{1}+3 x_{2} \\
& \text { s.t. } \left.\begin{array}{rl}
x_{1} & +2 x_{2} \leq 6 \\
2 x_{1} & +x_{2} \leq 8 \\
x_{1}, x_{2} & \geq 0
\end{array} \text { } \begin{array}{rl} 
& \leq 8
\end{array}\right)
\end{aligned}
$$

3 Use the simplex algorithm to solve the following problem:

$$
\begin{array}{lr}
\max z=2 x_{1}-x_{2}+x_{3} \\
\text { s.t. } & 3 x_{1}+x_{2}+x_{3} \leq 60 \\
& \\
& 2 x_{1}+x_{2}+2 x_{3} \leq 20 \\
2 x_{1}+2 x_{2}+x_{3} \leq 20 \\
x_{1}, x_{2}, x_{3} & \geq 0
\end{array}
$$

## Group B

7 It has been suggested that at each iteration of the simplex algorithm, the entering variable should be (in a maximization problem) the variable that would bring about the greatest increase in the objective function. Although this usually results in fewer pivots than the rule of entering the most negative row 0 entry, the greatest increase rule is hardly ever used. Why not?

### 4.6 Using the Simplex Algorithm to Solve Minimization Problems

There are two different ways that the simplex algorithm can be used to solve minimization problems. We illustrate these methods by solving the following LP:

$$
\begin{align*}
& \min z=2 x_{1}-3 x_{2} \\
& \text { s.t. } \quad x_{1}+x_{2} \leq 4  \tag{LP2}\\
& x_{1}-x_{2} \leq 6 \\
& \\
& x_{1}, x_{2} \geq 0
\end{align*}
$$

## Method 1

The optimal solution to LP 2 is the point $\left(x_{1}, x_{2}\right)$ in the feasible region for LP 2 that makes $z=2 x_{1}-3 x_{2}$ the smallest. Equivalently, we may say that the optimal solution to LP 2 is the point in the feasible region that makes $-z=-2 x_{1}+3 x_{2}$ the largest. This means that we can find the optimal solution to LP 2 by solving LP $2^{\prime}$ :

$$
\begin{array}{lr}
\max -z=-2 x_{1}+3 x_{2} \\
\text { s.t. } & x_{1}+x_{2} \leq 4 \\
& x_{1}-x_{2} \leq 6  \tag{LP2'}\\
& x_{1}, x_{2} \geq 0
\end{array}
$$

In solving LP $2^{\prime}$, we will use $-z$ as the basic variable for row 0 . After adding slack variables $s_{1}$ and $s_{2}$ to the two constraints, we obtain the initial tableau in Table 9. Because $x_{2}$ is the only variable with a negative coefficient in row 0 , we enter $x_{2}$ into the basis. The ratio test indicates that $x_{2}$ should enter the basis in the first constraint, row 1 . The resulting tableau is shown in Table 10. Because each variable in row 0 has a nonnegative coefficient, this is an optimal tableau. Thus, the optimal solution to LP $2^{\prime}$ is $-z=12, x_{2}=4, s_{2}=10, x_{1}=$ $s_{1}=0$. Then the optimal solution to LP 2 is $z=-12, x_{2}=4, s_{2}=10, x_{1}=s_{1}=0$. Substituting the values of $x_{1}$ and $x_{2}$ into LP 2's objective function, we obtain

TABLE 9
Initial Tableau for LP 2-Method 1

| $-z$ | $x_{1}$ | $x_{2}$ | $s_{1}$ | $s_{2}$ | rhs | Basic <br> Variable | Ratio |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :--- |
| 1 | 2 | -3 | 0 | 0 | 0 | $-z=0$ |  |
| 0 | 1 | $(1)$ | 1 | 0 | 4 | $s_{1}=4$ | $\frac{4}{1}=4^{*}$ |
| 0 | 1 | -1 | 0 | 1 | 6 | $s_{2}=6$ | None |

table 10
Optimal Tableau for LP 2-Method 1

| $-z$ | $x_{1}$ | $x_{2}$ | $s_{1}$ | $s_{2}$ | rhs | Basic <br> Variable |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | 5 | 0 | 3 | 0 | 12 | $-z=12$ |
| 0 | 1 | 1 | 1 | 0 | 4 | $x_{2}=4$ |
| 0 | 2 | 0 | 1 | 1 | 10 | $s_{2}=10$ |

TABLE 11
Initial Tableau for LP 2-Method 2

| $z$ | $x_{1}$ | $x_{2}$ | $s_{1}$ | $s_{2}$ | rhs | Basic <br> Variable | Ratio |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :--- |
| 1 | -2 | 3 | 0 | 0 | 0 | $z=0$ |  |
| 0 | 1 | $(1)$ | 1 | 0 | 4 | $s_{1}=4$ | $\frac{4}{1}=4^{*}$ |
| 0 | 1 | -1 | 0 | 1 | 6 | $s_{2}=6$ | None |

TABLE 12
Optimal Tableau for LP 2-Method 2

| $z$ | $x_{1}$ | $x_{2}$ | $s_{1}$ | $s_{2}$ | rhs | Basic <br> Variable |
| ---: | ---: | ---: | ---: | ---: | ---: | :--- |
| 1 | -5 | 0 | -3 | 0 | -12 | $z=-12$ |
| 0 | 1 | 1 | 1 | 0 | 4 | $x_{2}=4$ |
| 0 | 2 | 0 | 1 | 1 | 10 | $s_{2}=10$ |

$$
z=2 x_{1}-3 x_{2}=2(0)-3(4)=-12
$$

In summary, multiply the objective function for the min problem by -1 and solve the problem as a maximization problem with objective function $-z$. The optimal solution to the max problem will give you the optimal solution to the min problem. Remember that (optimal $z$-value for min problem $)=-($ optimal objective function value $z$ for max problem $)$.

## Method 2

A simple modification of the simplex algorithm can be used to solve min problems directly. Modify Step 3 of the simplex as follows: If all nonbasic variables in row 0 have nonpositive coefficients, then the current bfs is optimal. If any nonbasic variable in row 0 has a positive coefficient, choose the variable with the "most positive" coefficient in row

0 to enter the basis.
This modification of the simplex algorithm works because increasing a nonbasic variable with a positive coefficient in row 0 will decrease $z$. If we use this method to solve LP 2, then our initial tableau will be as shown in Table 11. Because $x_{2}$ has the most positive coefficient in row 0 , we enter $x_{2}$ into the basis. The ratio test says that $x_{2}$
should enter the basis in row
1, resulting in Table 12. Because each variable in row 0 has a nonpositive coefficient, this is an optimal tableau. ${ }^{\dagger}$ Thus, the optimal solution to LP 2 is (as we have already seen) $z=-12, x_{2}=4, s_{2}=$ $10, x_{1}=s_{1}=0$.

[^7]
## Sensitivity Analysis and Duality

Two of the most important topics in linear programming are sensitivity analysis and duality. After studying these important topics, the reader will have an appreciation of the beauty and logic of linear programming and be ready to study advanced linear programming topics such as those discussed in Chapter 10.

In Section 6.1, we illustrate the concept of sensitivity analysis through a graphical example. In Section 6.2, we use our knowledge of matrices to develop some important formulas, which are used in Sections 6.3 and 6.4 to develop the mechanics of sensitivity analysis. The remainder of the chapter presents the important concept of duality. Duality provides many insights into the nature of linear programming, gives us the useful concept of shadow prices, and helps us understand sensitivity analysis. It is a necessary basis for students planning to take advanced topics in linear and nonlinear programming.

### 6.1 A Graphical Introduction to Sensitivity Analysis

Sensitivity analysis is concerned with how changes in an LP's parameters affect the LP's optimal solution.

Reconsider the Giapetto problem of Section 3.1:

$$
\begin{array}{lrl}
\max z=3 x_{1}+2 x_{2} & & \\
\text { s.t. } & 2 x_{1}+x_{2} & \leq 100 \\
& & \text { (Finishing constraint) } \\
& x_{1}+x_{2} & \leq 80 \\
& & \text { (Carpentry constraint) } \\
& x_{1} & \\
& x_{1}, x_{2} & \geq 0
\end{array}
$$

where

$$
\begin{aligned}
& x_{1}=\text { number of soldiers produced per week } \\
& x_{2}=\text { number of trains produced per week }
\end{aligned}
$$

The optimal solution to this problem is $z=180, x_{1}=20, x_{2}=60$ (point $B$ in Figure 1), and it has $x_{1}, x_{2}$, and $s_{3}$ (the slack variable for the demand constraint) as basic variables. How would changes in the problem's objective function coefficients or right-hand sides change this optimal solution?

## Graphical Analysis of the Effect of a Change in an Objective Function Coefficient

If the contribution to profit of a soldier were to increase sufficiently, then it would be optimal for Giapetto to produce more soldiers ( $s_{3}$ would become nonbasic). Similarly, if the

contribution to profit of a soldier were to decrease sufficiently, it would be optimal for Giapetto to produce only trains ( $x_{1}$ would now be nonbasic). We now show how to determine the values of the contribution to profit for soldiers for which the current optimal basis will remain optimal.

Let $c_{1}$ be the contribution to profit by each soldier. For what values of $c_{1}$ does the current basis remain optimal?

At present, $c_{1}=3$, and each isoprofit line has the form $3 x_{1}+2 x_{2}=$ constant, or $x_{2}=-\frac{3 x}{2}+\frac{\text { constant }}{2}$, and each isoprofit line has a slope of $-\frac{3}{2}$. From Figure 1, we see that if a change in $c_{1}$ causes the isoprofit lines to be flatter than the carpentry constraint, then the optimal solution will change from the current optimal solution (point $B$ ) to a new optimal solution (point $A$ ). If the profit for each soldier is $c_{1}$, then the slope of each isoprofit line will be $-\frac{c_{1}}{2}$. Because the slope of the carpentry constraint is -1 , the isoprofit lines will be flatter than the carpentry constraint if $-\frac{c_{1}}{2}>-1$, or $c_{1}<2$, and the current basis will no longer be optimal. The new optimal solution will be $(0,80)$, point $A$ in Figure 1.

If the isoprofit lines are steeper than the finishing constraint, then the optimal solution will change from point $B$ to point $C$. The slope of the finishing constraint is -2 . If $-\frac{c_{1}}{2}<-2$, or $c_{1}>4$, then the current basis is no longer optimal, and point $C(40,20)$ will be optimal. In summary, we have shown that (if all other parameters remain unchanged) the current basis remains optimal for $2 \leq c_{1} \leq 4$, and Giapetto should still manufacture 20 soldiers and 60 trains. Of course, even if $2 \leq c_{1} \leq 4$, Giapetto's profit will change. For instance, if $c_{1}=4$, Giapetto's profit will now be $4(20)+2(60)=\$ 200$ instead of $\$ 180$.

## Graphical Analysis of the Effect of a Change in a Right-Hand Side on the LP's Optimal Solution

A graphical analysis can also be used to determine whether a change in the right-hand side of a constraint will make the current basis no longer optimal. Let $b_{1}$ be the number of available finishing hours. Currently, $b_{1}=100$. For what values of $b_{1}$ does the current
basis remain optimal? From Figure 2, we see that a change in $b_{1}$ shifts the finishing constraint parallel to its current position. The current optimal solution (point $B$ in Figure 2) is where the carpentry and finishing constraints are binding. If we change the value of $b_{1}$, then as long as the point where the finishing and carpentry constraints are binding remains feasible, the optimal solution will still occur where these constraints intersect. From Figure 2, we see that if $b_{1}>120$, then the point where the finishing and carpentry constraints are both binding will lie on the portion of the carpentry constraint below point $D$. Note that at point $D, 2(40)+40=120$ finishing hours are used. In this region, $x_{1}>40$, and the demand constraint for soldiers is not satisfied. Thus, for $b_{1}>120$, the current basis will no longer be optimal. Similarly, if $b_{1}<80$, the carpentry and finishing constraints will be binding at an infeasible point having $x_{1}<0$, and the current basis will no longer be optimal. Note that at point $A, 0+80=80$ finishing hours are used. Thus (if all other parameters remain unchanged), the current basis remains optimal if $80 \leq b_{1} \leq 120$.

Note that although for $80 \leq b_{1} \leq 120$, the current basis remains optimal, the values of the decision variables and the objective function value change. For example, if $80 \leq$ $b_{1} \leq 100$, the optimal solution will change from point $B$ to some other point on the line segment $A B$. Similarly, if $100 \leq b_{1} \leq 120$, then the optimal solution will change from point $B$ to some other point on the line $B D$.

As long as the current basis remains optimal, it is a routine matter to determine how a change in the right-hand side of a constraint changes the values of the decision variables. To illustrate the idea, let $b_{1}=$ number of available finishing hours. If we change $b_{1}$ to $100+\Delta$, we know that the current basis remains optimal for $-20 \leq \Delta \leq 20$. Note that as $b_{1}$ changes (as long as $-20 \leq \Delta \leq 20$ ), the optimal solution to the LP is still the point where the finishing-hour and carpentry-hour constraints are binding. Thus, if $b_{1}=$ $100+\Delta$, we can find the new values of the decision variables by solving

$$
2 x_{1}+x_{2}=100+\Delta \quad \text { and } \quad x_{1}+x_{2}=80
$$




This yields $x_{1}=20+\Delta$ and $x_{2}=60-\Delta$. Thus, an increase in the number of available finishing hours results in an increase in the number of soldiers produced and a decrease in the number of trains produced.

If $b_{2}$ (the number of available carpentry hours) equals $80+\Delta$, it can be shown (see Problem 2) that the current basis remains optimal for $-20 \leq \Delta \leq 20$. If we change the value of $b_{2}$ (keeping $-20 \leq \Delta \leq 20$ ), then the optimal solution to the LP is still the point where the finishing and carpentry constraints are binding. Thus, if $b_{2}=80+\Delta$, the optimal solution to the LP is the solution to

$$
2 x_{1}+x_{2}=100 \quad \text { and } \quad x_{1}+x_{2}=80+\Delta
$$

This yields $x_{1}=20-\Delta$ and $x_{2}=60+2 \Delta$, which shows that an increase in the amount of available carpentry hours decreases the number of soldiers produced and increases the number of trains produced.

Suppose $b_{3}$, the demand for soldiers, is changed to $40+\Delta$. Then it can be shown (see Problem 3) that the current basis remains optimal for $\Delta \geq-20$. For $\Delta$ in this range, the optimal solution to the LP will still occur where the finishing and carpentry constraints are binding. Thus, the optimal solution will be the solution to

$$
2 x_{1}+x_{2}=100 \quad \text { and } \quad x_{1}+x_{2}=80
$$

Of course, this yields $x_{1}=20$ and $x_{2}=60$, which illustrates an important point. In a constraint with positive slack (or positive excess) in an LP's optimal solution, if we change the right-hand side of the constraint to a value in the range where the current basis remains optimal, the optimal solution to the LP is unchanged.

## Shadow Prices

As we will see in Section 6.8, it is often important for managers to determine how a change in a constraint's right-hand side changes the LP's optimal $z$-value. With this in mind, we define the shadow price for the $i$ th constraint of an LP to be the amount by which the optimal $z$-value is improved (improvement means increase in a max problem and decrease in a min problem) if the right-hand side of the $i$ th constraint is increased by 1 . This definition applies only if the change in the right-hand side of Constraint $i$ leaves the current basis optimal.

For any two-variable LP, it is a simple matter to determine each constraint's shadow price. For example, we know that if $100+\Delta$ finishing hours are available (assuming that the current basis remains optimal), then the LP's optimal solution is $x_{1}=20+\Delta$ and $x_{2}=60-\Delta$. Then the optimal $z$-value will equal $3 x_{1}+2 x_{2}=3(20+\Delta)+2(60-\Delta)=$ $180+\Delta$. Thus, as long as the current basis remains optimal, a unit increase in the number of available finishing hours will increase the optimal $z$-value by $\$ 1$. So the shadow price of the first (finishing hour) constraint is $\$ 1$.

For the second (carpentry hour) constraint, we know that if $80+\Delta$ carpentry hours are available (and the current basis remains optimal), then the optimal solution to the LP is $x_{1}=20-\Delta$ and $x_{2}=60+2 \Delta$. Then the new optimal $z$-value is $3 x_{1}+2 x_{2}=$ $3(20-\Delta)+2(60+2 \Delta)=180+\Delta$. Thus, a unit increase in the number of carpentry hours will increase the optimal $z$-value by $\$ 1$ (as long as the current basis remains optimal). So the shadow price of the second (carpentry hour) constraint is $\$ 1$.

We now find the shadow price of the third (demand) constraint. If the right-hand side is $40+\Delta$, then the optimal values of the decision variables remain unchanged, as long as the current basis remains optimal. Then the optimal $z$-value will also remain unchanged, which shows that the shadow price of the third (demand) constraint is $\$ 0$. It turns out that whenever the slack variable or excess variable for a constraint is positive in an LP's optimal solution, the constraint will have a zero shadow price.

Suppose we increase the right-hand side of the $i$ th constraint of an LP by $\Delta b_{i}\left(\Delta b_{i}<0\right.$ means that we are decreasing the right-hand side) and the current basis remains optimal. Then each unit by which Constraint $i$ 's right-hand side is increased will increase the optimal $z$-value (for a max problem) by the shadow price. Thus, the new optimal $z$-value is given by
$($ New optimal $z$-value $)=($ old optimal $z$-value $)+($ Constraint $i$ 's shadow price $) \Delta b_{i}$
For a minimization problem,
$($ New optimal $z$-value $)=($ old optimal $z$-value $)-($ Constraint $i$ 's shadow price $) \Delta b_{i}$
For example, if 95 carpentry hours are available, then $\Delta b_{2}=15$, and the new $z$-value is given by

$$
\text { New optimal } z \text {-value }=180+15(1)=\$ 195
$$

We will continue our discussion of shadow prices in Section 6.8.

## Importance of Sensitivity Analysis

Sensitivity analysis is important for several reasons. In many applications, the values of an LP's parameters may change. For example, the prices at which soldiers and trains are sold may change, as may the availability of carpentry and finishing hours. If a parameter changes, sensitivity analysis often makes it unnecessary to solve the problem again. For example, if the profit contribution of a soldier increased to $\$ 3.50$, we would not have to solve the Giapetto problem again because the current solution remains optimal. Of course, solving the Giapetto problem again would not be much work, but solving an LP with thousands of variables and constraints again would be a chore. A knowledge of sensitivity analysis often enables the analyst to determine from the original solution how changes in an LP's parameters change the optimal solution.

Recall that we may be uncertain about the values of parameters in an LP, for example, the weekly demand for soldiers. With the graphical method, it can be shown that if the weekly demand for soldiers is at least 20, then the optimal solution to the Giapetto problem is still $(20,60)$ (see Problem 3 at the end of this section). Thus, even if Giapetto is uncertain about the demand for soldiers, the company can still be fairly confident that it is optimal to produce 20 soldiers and 60 trains.

Of course, the graphical approach is not useful for sensitivity analysis on an LP with more than two variables. Before learning how to perform sensitivity analysis on an arbitrary LP, we need to use our knowledge of matrices to express simplex tableaus in matrix form. This is the subject of Section 6.2.

## PROBLEMS

## Group A

1 Show that if the contribution to profit for trains is between $\$ 1.50$ and $\$ 3$, the current basis remains optimal. If the contribution to profit for trains is $\$ 2.50$, what would be the new optimal solution?

2 Show that if available carpentry hours remain between 60 and 100 , the current basis remains optimal. If between 60 and 100 carpentry hours are available, then would Giapetto still produce 20 soldiers and 60 trains?

3 Show that if the weekly demand for soldiers is at least 20, the current basis remains optimal, and Giapetto should still produce 20 soldiers and 60 trains.
4 For the Dorian Auto problem (Example 2 in Chapter 3), a Find the range of values of the cost of a comedy ad for which the current basis remains optimal.
b Find the range of values of the cost of a football ad for which the current basis remains optimal.

2 For the following LP, $x_{2}$ and $s_{1}$ are basic variables in the optimal tableau. Use the formulas of this section to determine the optimal tableau.

$$
\begin{aligned}
& \max z=-x_{1}+x_{2} \\
& \text { s.t. } \quad 2 x_{1}+x_{2} \leq 4 \\
& x_{1}+x_{2} \leq 2 \\
& x_{1}, x_{2} \geq 0
\end{aligned}
$$

### 6.3 Sensitivity Analysis

We now explore how changes in an LP's parameters (objective function coefficients, righthand sides, and technological coefficients) change the optimal solution. As described in Section 6.1, the study of how an LP's optimal solution depends on its parameters is called sensitivity analysis. Our discussion focuses on maximization problems and relies heavily on the formulas of Section 6.2. (The modifications for min problems are straightforward; see Problem 8 at the end of this section.)

As in Section 6.2, we let BV be the set of basic variables in the optimal tableau. Given a change (or changes) in an LP, we want to determine whether BV remains optimal. The mechanics of sensitivity analysis hinge on the following important observation. From Chapter 4, we know that a simplex tableau (for a max problem) for a set of basic variables BV is optimal if and only if each constraint has a nonnegative right-hand side and each variable has a nonnegative coefficient in row 0 . This follows, because if each constraint has a nonnegative right-hand side, then BV's basic solution is feasible, and if each variable in row 0 has a nonnegative coefficient, then there can be no basic feasible solution with a higher $z$-value than BV . Our observation implies that whether a tableau is feasible and optimal depends only on the right-hand sides of the constraints and on the coefficients of each variable in row 0 . For example, if an LP has variables $x_{1}, x_{2}, \ldots, x_{6}$, the following partial tableau would be optimal:

$$
\begin{aligned}
z+2 x_{2}+x_{4}+x_{6} & =6 \\
& =1 \\
& =2 \\
& =3
\end{aligned}
$$

This tableau's optimality is not affected by the parts of the tableau that are omitted.
Suppose we have solved an LP and have found that BV is an optimal basis. We can use the following procedure to determine if any change in the LP will cause BV to be no longer optimal.

Step 1 Using the formulas of Section 6.2, determine how changes in the LP's parameters change the right-hand side and row 0 of the optimal tableau (the tableau having BV as the set of basic variables).
Step 2 If each variable in row 0 has a non-negative coefficient and each constraint has a nonnegative right-hand side, then BV is still optimal. Otherwise, BV is no longer optimal.

If BV is no longer optimal, then you can find the new optimal solution by using the Section 6.2 formulas to recreate the entire tableau for BV and then continuing the simplex algorithm with the BV tableau as your starting tableau.

There can be two reasons why a change in an LP's parameters causes BV to be no longer optimal. First, a variable (or variables) in row 0 may have a negative coefficient. In this case, a better (larger $z$-value) bfs can be obtained by pivoting in a nonbasic variable with a negative coefficient in row 0 . If this occurs, we say that BV is now a suboptimal basis. Second, a constraint (or constraints) may now have a negative right-hand side. In this case, at least one member of BV will now be negative and BV will no longer yield a bfs. If this occurs, we say that BV is now an infeasible basis.

We illustrate the mechanics of sensitivity analysis in the Dakota Furniture example. Recall that

$$
\begin{aligned}
& x_{1}=\text { number of desks manufactured } \\
& x_{2}=\text { number of tables manufactured } \\
& x_{3}=\text { number of chairs manufactured }
\end{aligned}
$$

The objective function for the Dakota problem was

$$
\max z=60 x_{1}+30 x_{2}+20 x_{3}
$$

and the initial tableau was

$$
\begin{array}{rlrl}
z-60 x_{1}-30 x_{2}-20 x_{3} & & =0 & \\
8 x_{1}+6 x_{2}+x_{3}+s_{1} & & =48 &  \tag{12}\\
\text { (Lumber constraint) } \\
4 x_{1}+2 x_{2}+1.5 x_{3}+s_{2} & & =20 & \\
\text { (Finishing constraint) } \\
2 x_{1}+1.5 x_{2}+0.5 x_{3} & & s_{3} & =8
\end{array}
$$

The optimal tableau was

$$
\begin{array}{rlrl}
z & +5 x_{2}+10 s_{2}+10 s_{3} & =280 \\
-2 x_{2}+s_{1} & +2 s_{2}-8 s_{3} & =24 \\
-2 x_{2}+x_{3} & +2 s_{2}-4 s_{3} & =8  \tag{13}\\
x_{1}+1.25 x_{2} & & -0.5 s_{2}+1.5 s_{3} & =2
\end{array}
$$

Note that $\mathrm{BV}=\left\{s_{1}, x_{3}, x_{1}\right\}$ and $\mathrm{NBV}=\left\{x_{2}, s_{2}, s_{3}\right\}$. The optimal bfs is $z=280, s_{1}=$ $24, x_{3}=8, x_{1}=2, x_{2}=0, s_{2}=0, s_{3}=0$.

We now discuss how six types of changes in an LP's parameters change the optimal solution:

Change 1 Changing the objective function coefficient of a nonbasic variable
Change 2 Changing the objective function coefficient of a basic variable
Change 3 Changing the right-hand side of a constraint
Change 4 Changing the column of a nonbasic variable
Change 5 Adding a new variable or activity
Change 6 Adding a new constraint (see Section 6.11)

## Changing the Objective Function Coefficient of a Nonbasic Variable

In the Dakota problem, the only nonbasic decision variable is $x_{2}$ (tables). Currently, the objective function coefficient of $x_{2}$ is $c_{2}=30$. How would a change in $c_{2}$ affect the optimal solution to the Dakota problem? More specifically, for what values of $c_{2}$ would $\mathrm{BV}=\left\{s_{1}, x_{3}, x_{1}\right\}$ remain optimal?

Suppose we change the objective function coefficient of $x_{2}$ from 30 to $30+\Delta$. Then $\Delta$ represents the amount by which we have changed $c_{2}$ from its current value. For what values of $\Delta$ will the current set of basic variables (the current basis) remain optimal? We begin by determining how changing $c_{2}$ from 30 to $30+\Delta$ will change the BV tableau. Note that $B^{-1}$ and $\mathbf{b}$ are unchanged, and therefore, from (6), the right-hand side of BV's tableau $\left(B^{-1} \mathbf{b}\right)$ has not changed, so BV is still feasible. Because $x_{2}$ is a nonbasic variable, $\mathbf{c}_{\mathrm{BV}}$ has not changed. From (10), we can see that the only variable whose row 0 coefficient will be
changed by a change in $c_{2}$ is $x_{2}$. Thus, BV will remain optimal if $\bar{c}_{2} \geq 0$, and BV will be suboptimal if $\bar{c}_{2}<0$. In this case, $z$ could be improved by entering $x_{2}$ into the basis.

We have

$$
\mathbf{a}_{2}=\left[\begin{array}{l}
6 \\
2 \\
1.5
\end{array}\right]
$$

and $c_{2}=30+\Delta$. Also, from Section 6.2, we know that $\mathbf{c}_{\mathrm{BV}} B^{-1}=\left[\begin{array}{lll}0 & 10 & 10\end{array}\right]$. Now (10) shows that

$$
\bar{c}_{2}=\left[\begin{array}{lll}
0 & 10 & 10
\end{array}\right]\left[\begin{array}{l}
6 \\
2 \\
1.5
\end{array}\right]-(30+\Delta)=35-30-\Delta=5-\Delta
$$

Thus, $\bar{c}_{2} \geq 0$ holds, and BV will remain optimal, if $5-\Delta \geq 0$, or $\Delta \leq 5$. Similarly, $\bar{c}_{2}<0$ holds if $\Delta>5$, but then BV is no longer optimal. This means that if the price of tables is decreased or increased by $\$ 5$ or less, BV remains optimal. Thus, for $c_{2} \leq 30+$ $5=35$, BV remains optimal.

If BV remains optimal after a change in a nonbasic variable's objective function coefficient, the values of the decision variables and the optimal $z$-value remain unchanged. This is because a change in the objective function coefficient for a nonbasic variable leaves the right-hand side of row 0 and the constraints unchanged. For example, if the price of tables increases to $\$ 33\left(c_{2}=33\right)$, the optimal solution to the Dakota problem remains unchanged (Dakota should still make 2 desks and 8 chairs, and $z=280$ ). On the other hand, if $c_{2}>35$, BV will no longer be optimal, because $\bar{c}_{2}<0$. In this case, we find the new optimal solution by recreating the BV tableau and then using the simplex algorithm. For example, if $c_{2}=40$, we know that the only part of the BV tableau that will change is the coefficient of $x_{2}$ in row 0 . If $c_{2}=40$, then

$$
\bar{c}_{2}=\left[\begin{array}{lll}
0 & 10 & 10
\end{array}\right]\left[\begin{array}{l}
6 \\
2 \\
1.5
\end{array}\right]-40=-5
$$

Now the BV "final" tableau is as shown in Table 2. This is not an optimal tableau (it is suboptimal), and we can increase $z$ by making $x_{2}$ a basic variable in row 3 . The resulting tableau is given in Table 3. This is an optimal tableau. Thus, if $c_{2}=40$, the optimal solution to the Dakota problem changes to $z=288, s_{1}=27.2, x_{3}=11.2, x_{2}=1.6$, $x_{1}=0, s_{2}=0, s_{3}=0$. In this case, the increase in the price of tables has made tables sufficiently more attractive to induce Dakota to manufacture them. Note that after changing a nonbasic variable's objective function coefficient, it may, in general, take more than one pivot to find the new optimal solution.

There is a more insightful way to show that the current basis in the Dakota problem remains optimal as long as the price of tables is decreased or increased by $\$ 5$ or less. From the optimal row 0 in (13), we see that if $c_{2}=30$, then

$$
z=280-10 s_{2}-10 s_{3}-5 x_{2}
$$

This tells us that each table that Dakota manufactures will decrease revenue by $\$ 5$ (in other words, the reduced cost for tables is 5). If we increase the price of tables by more than $\$ 5$, each table would now increase Dakota's revenue. For example, if $c_{2}=36$, each table would increase revenues by $6-5=\$ 1$ and Dakota should manufacture tables. Thus, as before, we see that for $\Delta>5$, the current basis is no longer optimal. This analysis yields another interpretation of the reduced cost of a nonbasic variable: The reduced cost for a nonbasic variable (in a max problem) is the maximum amount by which the

TABLE 2
"Final" (Suboptimal) Dakota Tableau (\$40/Table)

|  |  |  | Basic Variable | Ratio |  |
| ---: | :--- | :--- | :--- | :--- | :--- |
| $z$ | - | $5 x_{2}$ | $+10 s_{2}+10 s_{3}=280$ | $z=280$ |  |
|  | $-2 x_{2}$ | $+s_{1}+2 s_{2}-8 s_{3}=24$ | $s_{1}=24$ | None |  |
|  | $-2 x_{2}+x_{3}$ | $+2 s_{2}-4 s_{3}=8$ | $x_{3}=8$ | None |  |
| $x_{1}+1.25 x_{2}$ | $-0.5 s_{2}+1.5 s_{3}=2$ | $x_{1}=2$ | $1.6^{*}$ |  |  |

table 3
Optimal Dakota Tableau (\$40/Table)

|  |  | Basic Variable |
| :--- | ---: | :--- |
| $z+4 x_{1}$ | $+8 s_{2}+16 s_{3}=288$ | $z=288$ |
| $1.6 x_{1}$ | $+s_{1}+1.2 s_{2}-5.6 s_{3}=27.2$ | $s_{1}=27.2$ |
| $1.6 x_{1}+x_{3}$ | $+1.2 s_{2}-1.6 s_{3}=11.2$ | $x_{3}=11.2$ |
| $0.8 x_{1}+x_{2}$ | $-0.4 s_{2}+1.2 s_{3}=1.6$ | $x_{2}=1.6$ |

variable's objective function coefficient can be increased before the current basis becomes suboptimal, and it becomes optimal for the nonbasic variable to enter the basis.

In summary, if the objective function coefficient for a nonbasic variable $x_{j}$ is changed, the current basis remains optimal if $\bar{c}_{j} \geq 0$. If $\bar{c}_{j}<0$, then the current basis is no longer optimal, and $x_{j}$ will be a basic variable in the new optimal solution.

## Changing the Objective Function Coefficient of a Basic Variable

In the Dakota problem, the decision variables $x_{1}$ (desks) and $x_{3}$ (chairs) are basic variables. We now explain how a change in the objective function coefficient of a basic variable will affect an LP's optimal solution. We begin by analyzing how this change affects the BV tableau. Because we are not changing $B$ (or therefore $B^{-1}$ ) or $\mathbf{b},(6)$ shows that the right-hand side of each constraint will remain unchanged, and $B V$ will remain feasible. Because we are changing $\mathbf{c}_{\mathrm{BV}}$, however, so $\mathbf{c}_{\mathrm{BV}} B^{-1}$ will change. From (10), we see that a change in $\mathbf{c}_{\mathrm{BV}} B^{-1}$ may change more than one coefficient in row 0 . To determine whether BV remains optimal, we must use (10) to recompute row 0 for the BV tableau. If each variable in row 0 still has a nonnegative coefficient, BV remains optimal. Otherwise, BV is now suboptimal. To illustrate the preceding ideas, we analyze how a change in the objective function coefficient for $x_{1}$ (desks) from its current value of $c_{1}=60$ affects the optimal solution to the Dakota problem.

Suppose that $c_{1}$ is changed to $60+\Delta$, changing $\mathbf{c}_{\mathrm{BV}}$ to $\mathbf{c}_{\mathrm{BV}}=\left[\begin{array}{llll}0 & 20 & 60 & +\Delta\end{array}\right]$. To compute the new row 0 , we need to know $B^{-1}$. We could (as in Section 6.2) use the Gauss-Jordan method to compute $B^{-1}$. Recall that this method begins by writing down the $3 \times 6$ matrix $B \mid I_{3}$ :

$$
B \left\lvert\, I_{3}=\left[\begin{array}{lll|lll}
1 & 1 & 8 & 1 & 0 & 0 \\
0 & 1.5 & 4 & 0 & 1 & 0 \\
0 & 0.5 & 2 & 0 & 0 & 1
\end{array}\right]\right.
$$

Then we use EROs to transform the first three columns of $B \mid I_{3}$ to $I_{3}$. At this point, the last three columns of the resulting matrix will be $B^{-1}$.

It turns out that when we solved the Dakota problem by the simplex algorithm, without realizing it, we found $B^{-1}$. To see why this is the case, note that in going from the initial Dakota tableau (12) to the optimal Dakota tableau (13) we performed a series of EROs on the constraints. These EROs transformed the constraint columns corresponding to the initial basis $\left(s_{1}, s_{2}, s_{3}\right)$

$$
\text { from }\left[\begin{array}{ccc}
s_{1} & s_{2} & s_{3} \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{array}\right] \quad \text { to } \quad\left[\begin{array}{ccc}
s_{1} & s_{2} & s_{3} \\
1 & 2 & -8 \\
0 & 2 & -4 \\
0 & -0.5 & 1.5
\end{array}\right]
$$

These same EROs have transformed the columns corresponding to $\mathrm{BV}=\left\{s_{1}, x_{3}, x_{1}\right\}$

$$
\text { from } B=\left[\begin{array}{lll}
s_{1} & x_{3} & x_{1} \\
1 & 1 & 8 \\
0 & 1.5 & 4 \\
0 & 0.5 & 2
\end{array}\right] \quad \text { to } \quad\left[\begin{array}{ccc}
s_{1} & x_{3} & x_{1} \\
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{array}\right]
$$

This means that in solving the Dakota problem by the simplex algorithm, we have used EROs to transform $B$ to $I_{3}$. These same EROs transformed $I_{3}$ into

$$
\left[\begin{array}{ccc}
1 & 2 & -8 \\
0 & 2 & -4 \\
0 & -0.5 & 1.5
\end{array}\right]=B^{-1}
$$

We have discovered an extremely important fact: For any simplex tableau, $B^{-1}$ is the $m \times m$ matrix consisting of the columns in the current tableau that correspond to the initial tableau's set of basic variables (taken in the same order). This means that if the starting basis for an LP consists entirely of slack variables, then $B^{-1}$ for the optimal tableau is simply the columns for the slack variables in the constraints of the optimal tableau. In general, if the starting basic variable for the $i$ th constraint is the artificial variable $a_{i}$, then the $i$ th column of $B^{-1}$ will be the column for $a_{i}$ in the optimal tableau's constraints. Thus, we need not use the Gauss-Jordan method to find the optimal tableau's $B^{-1}$. We have already found $B^{-1}$ by performing the simplex algorithm.

We can now compute what $\mathbf{c}_{\mathrm{BV}} B^{-1}$ will be if $c_{1}=60+\Delta$ :

$$
\begin{align*}
\mathbf{c}_{\mathrm{BV}} B^{-1} & =\left[\begin{array}{lll}
0 & 20 & 60+\Delta
\end{array}\right]\left[\begin{array}{ccc}
1 & 2 & -8 \\
0 & 2 & -4 \\
0 & -0.5 & 1.5
\end{array}\right]  \tag{14}\\
& =\left[\begin{array}{lll}
0 & 10-0.5 \Delta & 10+1.5 \Delta
\end{array}\right]
\end{align*}
$$

Observe that for $\Delta=0$, (14) yields the original $\mathbf{c}_{\mathrm{BV}} B^{-1}$. We can now compute the new row 0 corresponding to $c_{1}=60+\Delta$. After noting that

$$
\mathbf{a}_{1}=\left[\begin{array}{l}
8 \\
4 \\
2
\end{array}\right], \mathbf{a}_{2}=\left[\begin{array}{l}
6 \\
2 \\
1.5
\end{array}\right], \mathbf{a}_{3}=\left[\begin{array}{l}
1 \\
1.5 \\
0.5
\end{array}\right], c_{1}=60+\Delta, c_{2}=30, c_{3}=20
$$

we can use (10) to compute the new row 0 . Because $s_{1}, x_{3}$, and $x_{1}$ are basic variables, their coefficients in row 0 must still be 0 . The coefficient of each nonbasic variable in the new row 0 is as follows:

$$
\bar{c}_{2}=\mathbf{c}_{\mathrm{BV}} B^{-1} \mathbf{a}_{2}-c_{2}=\left[\begin{array}{lll}
0 & 10-0.5 \Delta & 10+1.5 \Delta
\end{array}\right]\left[\begin{array}{l}
6 \\
2 \\
1.5
\end{array}\right]-30=5+1.25 \Delta
$$

$$
\begin{aligned}
& \text { Coefficient of } s_{2} \text { in row } 0=\text { second element of } \mathbf{c}_{\mathrm{BV}} B^{-1}=10-0.5 \Delta \\
& \text { Coefficient of } s_{3} \text { in row } 0=\text { third element of } \mathbf{c}_{\mathrm{BV}} B^{-1}=10+1.5 \Delta
\end{aligned}
$$

Thus, row 0 of the optimal tableau is now

$$
z+(5+1.25 \Delta) x_{2}+(10-0.5 \Delta) s_{2}+(10+1.5 \Delta) s_{3}=?
$$

From the new row 0 , we see that BV will remain optimal if and only if the following hold:

$$
\begin{array}{ll}
5+1.25 \Delta \geq 0 & \left.\quad \text { true iff }^{\dagger} \Delta \geq-4\right) \\
10-0.5 \Delta \geq 0 & \quad \text { (true iff } \Delta \geq 20) \\
10+1.5 \Delta \geq 0 & \quad \text { (true iff } \Delta \geq-(20 / 3))
\end{array}
$$

This means that the current basis remains optimal as long as $\Delta \geq-4, \Delta \leq 20$, and $\Delta \geq$ $-\frac{20}{3}$. From Figure 3, we see that the current basis will remain optimal if and only if $-4 \leq \Delta \leq 20$ : If $c_{1}$ is decreased by $\$ 4$ or less or increased by up to $\$ 20$, the current basis remains optimal. Thus, as long as $56=60-4 \leq c_{1} \leq 60+20=80$, the current basis remains optimal. If $c_{1}<56$ or $c_{1}>80$, the current basis is no longer optimal.

If the current basis remains optimal, then the values of the decision variables don't change because $B^{-1} \mathbf{b}$ remains unchanged. The optimal $z$-value does change, however. To illustrate this, suppose $c_{1}=70$. Because $56 \leq 70 \leq 80$, we know that the current basis remains optimal. Thus, Dakota should still manufacture 2 desks $\left(x_{1}=2\right)$ and 8 chairs $\left(x_{3}=8\right)$. However, changing $c_{1}$ to 70 changes $z$ to $z=70 x_{1}+30 x_{2}+20 x_{3}$. This changes $z$ to $70(2)+20(8)=\$ 300$. Another way to see that $z$ is now $\$ 300$ is to note that we have increased the revenue from each desk by $70-60=\$ 10$. Dakota is making 2 desks, so revenue should increase by $2(10)=\$ 20$, and new revenue $=280+20=\$ 300$.

## When the Current Basis Is No Longer Optimal

Recall that if $c_{1}<56$ or $c_{1}>80$, then the current basis is no longer optimal. Intuitively, if the price of desks is decreased sufficiently (with all other prices held constant), desks will no longer be worth making. Our analysis shows that this occurs if the price of desks is decreased by more than $\$ 4$. The reader should verify (see Problem 2 at the end of this section) that if $c_{1}<56, x_{1}$ is no longer a basic variable in the new optimal solution. On the other hand, if $c_{1}>80$, desks have become profitable enough to make the current basis suboptimal; desks are now so attractive that we want to make more of them. To do this, we must force another variable out of the basis. Suppose $c_{1}=100$. Because $100>$ 80, we know that the current basis is no longer optimal. How can we determine the new optimal solution? Simply create the optimal tableau for $c_{1}=100$ and proceed with the simplex. If $c_{1}=100$, then $\Delta=100-60=40$, and the new row 0 will have

$$
\begin{aligned}
& \bar{c}_{1}=0, \quad \bar{c}_{2}=5+1.25 \Delta=55, \quad \bar{c}_{3}=0, \\
& s_{1} \text { coefficient in row } 0=0 \\
& s_{2} \text { coefficient in row } 0=10-0.5 \Delta=-10 \\
& s_{3} \text { coefficient in row } 0=10+1.5 \Delta=70
\end{aligned}
$$

$$
\text { Right-hand side of row } 0=\mathbf{c}_{\mathrm{BV}} B^{-1} \mathbf{b}=\left[\begin{array}{lll}
0 & -10 & 70
\end{array}\right]\left[\begin{array}{r}
48 \\
20 \\
8
\end{array}\right]=360
$$

From (6), changing $c_{1}$ does not change the constraints in the BV tableau. This means that if $c_{1}=100$, then the BV tableau is as given in Table 4. $\mathrm{BV}=\left\{s_{1}, x_{3}, x_{1}\right\}$ is now subopti-

[^8]
tABLE 4
"Final" (Suboptimal) Tableau If $c_{1}=100$

|  |  |  | Basic Variable | Ratio |
| :--- | :--- | :--- | :--- | :--- |
| $z$ | $+55 x_{2}$ | $-10 s_{2}+70 s_{3}=360$ | $z=360$ |  |
|  | $-2 x_{2}$ | $+s_{1}+2 s_{2}-8 s_{3}=24$ | $s_{1}=24$ | 12 |
| $-2 x_{2}+x_{3}$ | $+2 s_{2}-4 s_{3}=8$ | $x_{3}=8$ | $4^{*}$ |  |
| $x_{1}+1.25 x_{2}$ | $-0.5 s_{2}+1.5 s_{3}=2$ | $x_{1}=2$ | None |  |

table 5
Optimal Dakota Tableau If $c_{1}=100$

|  |  |  | Basic Variable |
| ---: | ---: | ---: | :--- |
| $z$ | $+45 x_{2}+5 x_{3}$ | $+50 s_{3}=400$ | $z=400$ |
|  | $x_{3}+s_{1}$ | $-4 s_{3}=16$ | $s_{1}=16$ |
| $-\quad x_{2}+0.5 x_{3}$ | $+s_{2}-2 s_{3}=4$ | $s_{2}=4$ |  |
| $x_{1}+0.75 x_{2}+0.25 x_{3}$ |  | $+0.5 s_{3}=4$ | $x_{1}=4$ |

mal. To find the new Dakota optimal solution, we enter $s_{2}$ into the basis in row 2 (Table 5). This is an optimal tableau. If $c_{1}=100$, then the new optimal solution to the Dakota problem is $z=400, s_{1}=16, s_{2}=4, x_{1}=4, x_{2}=0, x_{3}=0$. Notice that increasing the profitability of desks has caused Dakota to stop making chairs. The resources that were previously used to make the chairs are now used to make $4-2=2$ extra desks.

In summary, if the objective function coefficient of a basic variable $x_{j}$ is changed, then the current basis remains optimal if the coefficient of every variable in row 0 of the BV tableau remains nonnegative. If any variable in row 0 has a negative coefficient, then the current basis is no longer optimal.

## Interpretation of the Objective Coefficient Ranges Block of the LINDO Output

To obtain a sensitivity report in LINDO, select Yes when asked (after solving LP) whether you want a Range analysis. To obtain a sensitivity report in LINGO, go to Options and select Range (after solving LP). If this does not work, go to Options, choose the General Solver tab, and then go to Dual Computations and select the Ranges and Values option.

In the OBJ COEFFICIENT RANGES block of the LINDO (or LINGO) computer output, we see the amount by which each variable's objective function coefficient may be changed before the current basis becomes suboptimal (assuming all other LP parameters are held constant). Look at the LINDO output for the Dakota problem (Figure 4). For each variable, the CURRENT COEF column gives the current value of the variable's objective function coefficient. For example, the objective function coefficient for DESKS is 60 . The ALLOWABLE INCREASE column gives the maximum amount by which the objective
where $\Delta b_{1}>0$ and $\Delta b_{2}<0$. Let

$$
r_{1}=\frac{\Delta b_{1}}{U_{1}-b_{1}} \quad \text { and } \quad r_{2}=\frac{-\Delta b_{2}}{b_{2}-L_{2}}
$$

Show that if $r_{1}+r_{2} \leq 1$, the current basis remains optimal. (Hint: You must show that

$$
B^{-1}\left[\begin{array}{l}
b_{1}^{\prime} \\
b_{2}^{\prime}
\end{array}\right] \geq\left[\begin{array}{l}
0 \\
0
\end{array}\right]
$$

Use the fact that
$\left[b_{1}^{\prime}, b_{2}^{\prime}\right]=r_{1}\left[U_{1}, b_{2}\right]+r_{2}\left[b_{1}, L_{2}\right]+\left(1-r_{1}-r_{2}\right)\left[b_{1}, b_{2}\right]$ to show this.)

### 6.5 Finding the Dual of an LP

Associated with any LP is another LP, called the dual. Knowing the relation between an LP and its dual is vital to understanding advanced topics in linear and nonlinear programming. This relation is important because it gives us interesting economic insights. Knowledge of duality will also provide additional insights into sensitivity analysis.

In this section, we explain how to find the dual of any LP; in Section 6.6, we discuss the economic interpretation of the dual; and in Sections 6.7-6.10, we discuss the relation that exists between an LP and its dual.

When taking the dual of a given LP, we refer to the given LP as the primal. If the pri$m a l$ is a max problem, then the dual will be a min problem, and vice versa. For convenience, we define the variables for the max problem to be $z, x_{1}, x_{2}, \ldots, x_{n}$ and the variables for the min problem to be $w, y_{1}, y_{2}, \ldots, y_{m}$. We begin by explaining how to find the dual of a max problem in which all variables are required to be nonnegative and all constraints are $\leq$ constraints (called a normal max problem). A normal max problem may be written as

$$
\begin{array}{ll}
\max z= & c_{1} x_{1}+c_{2} x_{2}+\cdots+c_{n} x_{n} \\
\text { s.t. } & a_{11} x_{1}+a_{12} x_{2}+\cdots+a_{1 n} x_{n} \leq b_{1} \\
& a_{21} x_{1}+a_{22} x_{2}+\cdots+a_{2 n} x_{n} \leq b_{2}  \tag{16}\\
& \vdots \quad \vdots \\
& a_{m 1} x_{1}+a_{m 2} x_{2}+\cdots+a_{m n} x_{n} \leq b_{m} \\
& x_{j} \geq 0 \quad(j=1,2, \ldots, n)
\end{array}
$$

The dual of a normal max problem such as (16) is defined to be

$$
\begin{array}{ll}
\min & w= \\
\text { s.t. } & b_{1} y_{1}+b_{2} y_{2}+\cdots+b_{m} x_{m} \\
& a_{11} y_{1}+a_{21} y_{2}+\cdots+a_{m 1} y_{m} \geq c_{1} \\
& a_{12} y_{1}+a_{22} y_{2}+\cdots+a_{m 2} y_{m} \geq c_{2}  \tag{17}\\
& \vdots \\
& a_{1 n} y_{1}+a_{2 n} y_{2}+\cdots+a_{m n} y_{m} \geq c_{n} \\
& y_{i} \geq 0 \quad(i=1,2, \cdots, m)
\end{array}
$$

A min problem such as (17) that has all $\geq$ constraints and all variables nonnegative is called a normal min problem. If the primal is a normal min problem such as (17), then we define the dual of (17) to be (16).

## Finding the Dual of a Normal Max or Min Problem

A tabular approach makes it easy to find the dual of an LP. If the primal is a normal max problem, then it can be read across (Table 14); the dual is found by reading down. Similarly, if the primal is a normal min problem, we find it by reading down; the dual is found

TABLE 14
Finding the Dual of a Normal Max or Min Problem

|  |  | $\max z$ |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\min w$ | $\left(x_{1} \geq 0\right)$ | $\left(x_{2} \geq 0\right)$ | $\cdots$ | $\left(x_{n} \geq 0\right)$ |  |  |
|  | $x_{1}$ | $x_{2}$ |  | $x_{n}$ |  |  |
| $\left(y_{1} \geq 0\right)$ | $y_{1}$ | $a_{11}$ | $a_{12}$ | $\cdots$ | $a_{1 n}$ |  |
| $\left(y_{2} \geq 0\right)$ | $y_{2}$ | $a_{21}$ | $a_{22}$ | $\cdots$ | $a_{2 n}$ |  |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |  | $\vdots$ |  |
| $\left(y_{m} \geq 0\right)$ | $y_{m}$ | $a_{m 1}$ | $a_{m 2}$ | $\cdots$ | $a_{m n}$ |  |
|  |  | $\geq c_{1}$ | $\geq c_{2}$ |  | $\geq b_{2}$ |  |
|  |  |  |  | $\leq b_{m}$ |  |  |

by reading across in the table. We illustrate the use of the table by finding the dual of the Dakota problem and the dual of the diet problems. The Dakota problem is

$$
\begin{array}{lrl}
\max z=60 x_{1}+30 x_{2}+20 x_{3} & & \\
\text { s.t. } & 8 x_{1}+6 x_{2}+x_{3} & \leq 48 \\
& & \text { (Lumber constraint) } \\
4 x_{1}+2 x_{2}+1.5 x_{3} & \leq 20 & \\
& 2 x_{1}+1.5 x_{2}+0.5 x_{3} & \leq 8 \\
& & \text { (Cinishing constraint) } \\
& x_{1}, x_{2}, x_{3} & \geq 0
\end{array}
$$

where

$$
\begin{aligned}
& x_{1}=\text { number of desks manufactured } \\
& x_{2}=\text { number of tables manufactured } \\
& x_{3}=\text { number of chairs manufactured }
\end{aligned}
$$

Using the format of Table 14, we read the Dakota problem across in Table 15. Then, reading down, we find the Dakota dual to be

$$
\begin{aligned}
& \min w=48 y_{1}+20 y_{2}+8 y_{3} \\
& \text { s.t. } \quad 8 y_{1}+4 y_{2}+2 y_{3} \geq 60 \\
& 6 y_{1}+2 y_{2}+1.5 y_{3} \geq 30 \\
& y_{1}+1.5 y_{2}+0.5 y_{3} \geq 20 \\
& y_{1}, y_{2}, y_{3} \geq 0
\end{aligned}
$$

The tabular method of finding the dual makes it clear that the $i$ th dual constraint corresponds to the $i$ th primal variable $x_{i}$. For example, the first dual constraint corresponds to $x_{1}$ (desks), because each number comes from the $x_{1}$ (desk) column of the primal. Simi-
tABLE 15
Finding the Dual of the Dakota Problem

|  | $\max z$ |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| $\min w$ |  | $x_{1}$ | $x_{2}$ | $x_{3}$ |
|  |  |  |  |  |
|  | 8 | 6 | 1 | $\leq 48$ |
| $\left(y_{1} \geq 0\right)$ | $y_{1}$ | 4 | 2 | 1.5 |
| $\left(y_{2} \geq 0\right)$ | $y_{2}$ | 2 | 1.5 | 0.5 |
| $\left(y_{3} \geq 0\right)$ | $y_{3}$ | 2 | $\geq 30$ | $\geq 20$ |
|  |  | $\geq 60$ | $\geq 30$ |  |

larly, the second dual constraint corresponds to $x_{2}$ (tables), and the third dual constraint corresponds to $x_{3}$ (chairs). In a similar fashion, dual variable $y_{i}$ is associated with the $i$ th primal constraint. For example, $y_{1}$ is associated with the first primal constraint (lumber constraint), because each coefficient of $y_{1}$ in the dual comes from the lumber constraint, or the availability of lumber. The importance of these correspondences between the primal and the dual will become clear in Section 6.6.

We now find the dual of the diet problem. Because the diet problem is a min problem, we follow the convention of using $w$ to denote the objective function and $y_{1}, y_{2}, y_{3}$, and $y_{4}$ for the variables. Then the diet problem may be written as

$$
\begin{aligned}
& \min w=50 y_{1}+20 y_{2}+30 y_{3}+80 y_{4} \\
& \text { s.t. } 400 y_{1}+200 y_{2}+150 y_{3}+500 y_{4} \geq 500 \quad \text { (Calorie constraint) } \\
& 3 y_{1}+2 y_{2} \quad \geq 6 \quad \text { (Chocolate constraint) } \\
& 2 y_{1}+2 y_{2}+4 y_{3}+4 y_{4} \geq 10 \quad \text { (Sugar constraint) } \\
& 2 y_{1}+4 y_{2}+y_{3}+5 y_{4} \geq 8 \quad \text { (Fat constraint) } \\
& y_{1}, y_{2}, y_{3}, y_{4} \geq 0 \\
& \text { (Calorie constraint) } \\
& \text { (Sugar constraint) } \\
& \text { (Fat constraint) }
\end{aligned}
$$

where

$$
\begin{aligned}
& y_{1}=\text { number of brownies eaten daily } \\
& y_{2}=\text { number of scoops of chocolate ice cream eaten daily } \\
& y_{3}=\text { bottles of soda drunk daily } \\
& y_{4}=\text { pieces of pineapple cheesecake eaten daily }
\end{aligned}
$$

The primal is a normal min problem, so we can read it down, and read its dual across, in Table 16. We find that the dual of the diet problem is

$$
\begin{aligned}
& \max z=500 x_{1}+6 x_{2}+10 x_{3}+8 x_{4} \\
& \text { s.t. } \quad 400 x_{1}+3 x_{2}+2 x_{3}+2 x_{4} \leq 50 \\
& 200 x_{1}+2 x_{2}+2 x_{3}+4 x_{4} \leq 20 \\
& 150 x_{1}+4 x_{3}+x_{4} \leq 30 \\
& \\
& 500 x_{1} \\
&
\end{aligned}
$$

As in the Dakota problem, we see that the $i$ th dual constraint corresponds to the $i$ th primal variable. For example, the third dual constraint may be thought of as the soda constraint. Also, the $i$ th dual variable corresponds to the $i$ th primal constraint. For example, $x_{3}$ (the third dual variable) may be thought of as the dual sugar variable.

TABLE 16
Finding the Dual of the Diet Problem

|  | $\max z$ |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| $\min w$ | $\left(x_{1} \geq 0\right)$ | $\left(x_{2} \geq 0\right)$ | $\left(x_{3} \geq 0\right)$ | $\left(x_{4} \geq 0\right)$ |  |
|  | $x_{1}$ | $x_{2}$ | $x_{3}$ | $x_{4}$ |  |
| $\left(y_{1} \geq 0\right)$ | $y_{1}$ | 400 | 3 | 2 | 2 |
| $\left(y_{2} \geq 0\right)$ | $y_{2}$ | 200 | 2 | 2 | 4 |
| $\left(y_{3} \geq 0\right)$ | $y_{3}$ | 150 | 0 | 4 | 1 |
| $\left(y_{4} \geq 0\right)$ | $y_{4}$ | 500 | 0 | 4 | $\leq 50$ |
|  |  | $\geq 500$ | $\geq 6$ | $\geq 10$ | $\geq 8$ |

## Finding the Dual of a Nonnormal LP

Unfortunately, many LPs are not normal max or min problems. For example,

$$
\begin{align*}
& \max z=2 x_{1}+x_{2} \\
& \text { s.t. } \quad x_{1}+x_{2}=2 \\
& 2 x_{1}-x_{2} \geq 3  \tag{18}\\
& x_{1}-x_{2} \leq 1 \\
& x_{1}
\end{align*} \quad \geq 0, x_{2} \text { urs }
$$

is not a normal max problem because it has a $\geq$ constraint, an equality constraint, and an unrestricted-in-sign variable. As another example of a nonnormal LP, consider

$$
\begin{align*}
& \min w=2 y_{1}+4 y_{2}+6 y_{3} \\
& \text { s.t. } \quad y_{1}+2 y_{2}+y_{3} \geq 2 \\
& y_{1}-y_{3} \geq 1  \tag{19}\\
& y_{2}+y_{3}=1 \\
& 2 y_{1}+y_{2} \leq 3 \\
& y_{1} \text { urs, } y_{2}, y_{3} \geq 0
\end{align*}
$$

This LP is not a normal min problem because it contains an equality constraint, $\mathrm{a} \leq$ constraint, and an unrestricted-in-sign variable.

Fortunately, an LP can be transformed into normal form (either (16) or (17)). To place a max problem into normal form, we proceed as follows:

Step 1 Multiply each $\geq$ constraint by -1 , converting it into $\mathrm{a} \leq$ constraint. For example, in (18), $2 x_{1}-x_{2} \geq 3$ would be transformed into $-2 x_{1}+x_{2} \leq-3$.

Step 2 Replace each equality constraint by two inequality constraints ( $a \leq$ constraint and $\mathrm{a} \geq$ constraint). Then convert the $\geq$ constraint to $\mathrm{a} \leq$ constraint. For example, in (18), we would replace $x_{1}+x_{2}=2$ by the two inequalities $x_{1}+x_{2} \geq 2$ and $x_{1}+x_{2} \leq 2$. Then we would convert $x_{1}+x_{2} \geq 2$ to $-x_{1}-x_{2} \leq-2$. The net result is that $x_{1}+x_{2}=2$ is replaced by the two inequalities $x_{1}+x_{2} \leq 2$ and $-x_{1}-x_{2} \leq-2$.

Step 3 As in Section 4.14, replace each urs variable $x_{i}$ by $x_{i}=x^{\prime}-x_{i}^{\prime \prime}$, where $x_{i}^{\prime} \geq 0$ and $x_{i}^{\prime \prime} \geq 0$. In (18), we would replace $x_{2}$ by $x_{2}^{\prime}-x_{2}^{\prime \prime}$.

After these transformations are complete, (18) has been transformed into the following (equivalent) LP:

$$
\begin{align*}
& \max z=2 x_{1}+x_{2}^{\prime}-x_{2}^{\prime \prime} \\
& \text { s.t. } \quad x_{1}+x_{2}^{\prime}-x_{2}^{\prime \prime} \leq 2 \\
& -x_{1}-x_{2}^{\prime}+x_{2}^{\prime \prime} \leq-2 \\
& -2 x_{1}+x_{2}^{\prime}-x_{2}^{\prime \prime} \leq-3  \tag{18'}\\
& x_{1}-x_{2}^{\prime}+x_{2}^{\prime \prime} \leq 1 \\
& \\
& x_{1}, x_{2}^{\prime}, x_{2}^{\prime \prime} \geq 0
\end{align*}
$$

Because $\left(18^{\prime}\right)$ is a normal max problem, we could use (16) and (17) to find the dual of (18').

If the primal is not a normal min problem, then we can transform it into a normal min problem as follows:

Step 1 Convert each $\leq$ constraint into a $\geq$ constraint by multiplying through by -1 . For example, in (19), $2 y_{1}+y_{2} \leq 3$ is transformed into $-2 y_{1}-y_{2} \geq-3$.

Step 2 Replace each equality constraint by a $\leq$ constraint and $\mathrm{a} \geq$ constraint. Then transform the $\leq$ constraint into $\mathrm{a} \geq$ constraint. For example, in (19), the constraint $y_{2}+y_{3}=$ 1 is equivalent to $y_{2}+y_{3} \leq 1$ and $y_{2}+y_{3} \geq 1$. Transforming $y_{2}+y_{3} \leq 1$ into $-y_{2}-y_{3} \geq-1$, we see that we can replace the constraint $y_{2}+y_{3}=1$ by the two constraints $y_{2}+y_{3} \geq 1$ and $-y_{2}-y_{3} \geq-1$.

Step 3 Replace any urs variable $y_{i}$ by $y_{i}=\mathrm{y}_{i}^{\prime}-y_{i}^{\prime \prime}$, where $y_{i}^{\prime} \geq 0$ and $\mathrm{y}_{i}^{\prime \prime} \geq 0$. Applying these steps to (19) yields the following standard min problem:

$$
\begin{align*}
\min w=2 y_{1}^{\prime}-2 y_{1}^{\prime \prime}+4 y_{2}+6 y_{3} & \\
\text { s.t. } y_{1}^{\prime}-y_{1}^{\prime \prime}+2 y_{2}+y_{3} & \geq 2 \\
y_{1}^{\prime}-y_{1}^{\prime \prime}-y_{3} & \geq 1 \\
y_{2}+y_{3} & \geq 1  \tag{19'}\\
-y_{2}-6 y_{3} & \geq-1 \\
-2 y_{1}^{\prime}+2 y_{1}^{\prime \prime}-y_{2} & \geq-3 \\
y_{1}^{\prime}, y_{1}^{\prime \prime}, y_{2}, y_{3} & \geq 0
\end{align*}
$$

Because (19') is a normal min problem in standard form, we may use (16) and (17) to find its dual.

We can find the dual of a nonnormal LP without going through the transformations that we have described by using the following rules. ${ }^{\dagger}$

## Finding the Dual of a Nonnormal Max Problem

Step 1 Fill in Table 14 so that the primal can be read across.
Step 2 After making the following changes, the dual can be read down in the usual fashion: (a) If the $i$ th primal constraint is $\mathrm{a} \geq$ constraint, then the corresponding dual variable $y_{i}$ must satisfy $y_{i} \leq 0$. (b) If the $i$ th primal constraint is an equality constraint, then the dual variable $y_{i}$ is now unrestricted in sign. (c) If the $i$ th primal variable is urs, then the $i$ th dual constraint will be an equality constraint.

When this method is applied to (18), the Table 14 format yields Table 17. We note with an asterisk $(*)$ the places where the rules must be used to determine part of the dual. For example, $x_{2}$ urs causes the second dual constraint to be an equality constraint. Also, the first primal constraint being an equality constraint makes $y_{1}$ urs, and the second primal constraint being a $\geq$ constraint makes $y_{2} \leq 0$. Filling in the missing information across from the appropriate asterisk yields Table18. Reading the dual down, we obtain

$$
\begin{aligned}
& \min w=2 y_{1}+3 y_{2}+y_{3} \\
& \text { s.t. } y_{1}+2 y_{2}+y_{3} \geq 2 \\
& y_{1}-y_{2}-y_{3}=1 \\
& y_{1} \text { urs, } y_{2} \leq 0, y_{3} \geq 0
\end{aligned}
$$

In Section 6.8, we give an intuitive explanation of why an equality constraint yields an unrestricted-in-sign dual variable and why $a \geq$ constraint yields a negative dual variable.

We can use the following rules to take the dual of a nonnormal min problem.

[^9]TABLE 17
Finding the Dual of LP (18)

|  |  | $\max z$ |  |
| :--- | :---: | :---: | :---: |
| $\min w$ |  | $\left(x_{1} \geq 0\right)$ | $\left(x_{2} \text { urs }\right)^{*}$ |$]$

tABLE 18
Finding the Dual of LP (18) (Continued)

|  |  | $\max z$ |  |  |
| :--- | :---: | :---: | ---: | :--- |
| $\min w$ |  | $x_{1}$ | $x_{2}$ |  |
|  | 1 | 1 | $=2$ |  |
| $\left(y_{1}\right.$ urs $)$ | $y_{1}$ | 2 | -1 | $\geq 3$ |
| $\left(y_{2} \leq 0\right)$ | $y_{2}$ | 1 | -1 | $\leq 1$ |
| $\left(y_{3} \geq 0\right)$ | $y_{3}$ | $\geq 2$ | $=1$ |  |
|  |  | $\geq 2 r s)$ |  |  |

## Finding the Dual of a Nonnormal Min Problem

Step 1 Write out the primal so it can be read down in Table 14.
Step 2 Except for the following changes, the dual can be read across the table: (a) If the $i$ th primal constraint is a $\leq$ constraint, then the corresponding dual variable $x_{i}$ must satisfy $x_{i} \leq 0$. (b) If the $i$ th primal constraint is an equality constraint, then the corresponding dual variable $x_{i}$ will be urs. (c) If the $i$ th primal variable $y_{i}$ is urs, then the $i$ th dual constraint is an equality constraint.

When this method is applied to (19), we get Table 19. Asterisks $\left(^{*}\right)$ show where the new rules must be used to determine parts of the dual. Because $y_{1}$ is urs, the first dual constraint is an equality. The third primal constraint is an equality, so dual variable $x_{3}$ is urs. Finally, because the fourth primal constraint is a $\leq$ constraint, the fourth dual variable $x_{4}$ must satisfy $x_{4} \leq 0$. We can now complete the table (see Table 20). Reading the dual across, we obtain
table 19
Finding the Dual of LP (19)

|  | $\max z$ |  |  |  |  |  |
| :--- | :---: | ---: | ---: | ---: | ---: | ---: |
| $\min w$ |  | $x_{1}$ | $x_{2}$ | $x_{3}$ | $x_{4}$ |  |
|  | 1 | 1 | 0 | 2 | 2 |  |
| $\left(y_{1} \text { urs }\right)^{*}$ | $y_{1}$ | 2 | 0 | 1 | 1 | $\leq 4$ |
| $\left(y_{2} \geq 0\right)$ | $y_{2}$ | 1 | -1 | 1 | 0 | $\leq 6$ |
| $\left(y_{3} \geq 0\right)$ | $y_{3}$ | $\geq 2$ | $\geq 1$ | $=1^{*}$ | $\leq 3^{*}$ |  |

tABLE 20
Finding the Dual of LP $(19)$ (Continued)

|  |  | $\max z$ |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| $\min w$ |  | $\left(x_{1} \geq 0\right)$ | $\left(x_{2} \geq 0\right)$ | $\left(x_{3} \geq \mathrm{urs}\right)$ | $\left(x_{4} \leq 0\right)$ |
|  | $x_{1}$ | $x_{2}$ | $x_{3}$ | $x_{4}$ |  |
| $\left(y_{1}\right.$ urs $)$ | $y_{1}$ | 1 | 1 | 0 | 2 |
| $\left(y_{2} \geq 0\right)$ | $y_{2}$ | 2 | 0 | 1 | 1 |
| $\left(y_{3} \geq 0\right)$ | $y_{3}$ | 1 | -1 | 1 | $\leq 2$ |
|  |  | $\geq 2$ | $\geq 1$ | $=1$ | $\leq 3$ |

$$
\begin{array}{lrl}
\max z=2 x_{1}+x_{2}+x_{3}+3 x_{4} \\
\text { s.t. } \quad x_{1}+x_{2}+2 x_{4} & =2 \\
2 x_{1}+x_{3}+x_{4} & \leq 4 \\
x_{1}-x_{2}+x_{3} & \leq 6 \\
x_{1}, x_{2} \geq 0, x_{3} \text { urs, } x_{4} & \leq 0
\end{array}
$$

The reader may verify that with these rules, the dual of the dual is always the primal. This is easily seen from the Table 14 format, because when you take the dual of the dual you are changing the LP back to its original position.

## PROBLEMS

## Group A

Find the duals of the following LPs:
$1 \max z=2 x_{1}+x_{2}$
$2 \min w=y_{1}-y_{2}$
s.t. $\quad 2 y_{1}+y_{2} \geq 4$
$y_{1}+y_{2} \geq 1$

$$
y_{1}+2 y_{2} \geq 3
$$

$$
y_{1}, y_{2} \geq 0
$$

$3 \max z=4 x_{1}-x_{2}+2 x_{3}$

$$
\begin{array}{lcl}
\text { s.t. } \quad \begin{aligned}
x_{1}+x_{2} & \leq 5 \\
2 x_{1}+x_{2} & \leq 7 \\
& 2 x_{2}+x_{3}
\end{aligned} \geq 6 \\
& \geq x_{3} & =4 \\
x_{1}+ & \\
x_{1} \geq 0, x_{2}, x_{3} \text { urs }
\end{array}
$$

$4 \min w=4 y_{1}+2 y_{2}-y_{3}$

$$
\begin{array}{ll}
\text { s.t. } & y_{1}+2 y_{2} \leq 6 \\
& y_{1}-y_{2}+2 y_{3}=8 \\
& y_{1}, y_{2} \geq 0, y_{3} \text { urs }
\end{array}
$$

$$
\begin{aligned}
& \text { s.t. } \quad-x_{1}+x_{2} \leq 1 \\
& x_{1}+x_{2} \leq 3 \\
& x_{1}-2 x_{2} \leq 4 \\
& x_{1}, x_{2} \geq 0
\end{aligned}
$$

## Group B

5 This problem shows why the dual variable for an equality constraint should be urs.
a Use the rules given in the text to find the dual of

$$
\begin{aligned}
& \max z=x_{1}+2 x_{2} \\
& \text { s.t. } \begin{aligned}
3 x_{1}+x_{2} & \leq 6 \\
2 x_{1}+x_{2} & =5 \\
x_{1}, x_{2} & \geq 0
\end{aligned}, ~
\end{aligned}
$$

b Now transform the LP in part (a) to the normal form. Using (16) and (17), take the dual of the transformed LP. Use $y_{2}^{\prime}$ and $y_{2}^{\prime \prime}$ as the dual variables for the two primal constraints derived from $2 x_{1}+x_{2}=5$.
c Make the substitution $y_{2}=y_{2}^{\prime}-y_{2}^{\prime \prime}$ in the part (b) answer. Now show that the two duals obtained in parts (a) and (b) are equivalent.

6 This problem shows why a dual variable $y_{i}$ corresponding to a $\geq$ constraint in a max problem must satisfy $y_{i} \leq 0$.
a Using the rules given in the text, find the dual of

$$
\begin{array}{lr}
\max z= & 3 x_{1}+x_{2} \\
\text { s.t. } & x_{1}+x_{2} \leq 1 \\
& -x_{1}+x_{2} \geq 2 \\
& x_{1}, x_{2} \geq 0
\end{array}
$$

b Transform the LP of part (a) into a normal max problem. Now use (16) and (17) to find the dual of the transformed LP. Let $\bar{y}_{2}$ be the dual variable corresponding to the second primal constraint.
c Show that, defining $\bar{y}_{2}=-y_{2}$, the dual in part (a) is equivalent to the dual in part (b).

### 6.6 Economic Interpretation of the Dual Problem

Interpreting the Dual of a Max Problem
The dual of the Dakota problem is

$$
\begin{array}{lrr}
\min w=48 y_{1}+20 y_{2}+8 y_{3} & \\
\text { s.t. } & 8 y_{1}+4 y_{2}+2 y_{3} \geq 60 & \\
6 y_{1}+2 y_{2}+1.5 y_{3} & \geq 30 & \text { (Desk constraint) } \\
& y_{1}+1.5 y_{2}+0.5 y_{3} & \geq 20 \\
& \text { (Table constraint) } \\
y_{1}, y_{2}, y_{3} & \geq 0 &
\end{array}
$$

The first dual constraint is associated with desks, the second with tables, and the third with chairs. Also, $y_{1}$ is associated with lumber, $y_{2}$ with finishing hours, and $y_{3}$ with carpentry hours. The relevant information about the Dakota problem is shown in Table 21.

We are now ready to interpret the Dakota dual (20). Suppose an entrepreneur wants to purchase all of Dakota's resources. Then the entrepreneur must determine the price he or she is willing to pay for a unit of each of Dakota's resources. With this in mind, we define

$$
\begin{aligned}
& y_{1}=\text { price paid for } 1 \text { board } \mathrm{ft} \text { of lumber } \\
& y_{2}=\text { price paid for } 1 \text { finishing hour } \\
& y_{3}=\text { price paid for } 1 \text { carpentry hour }
\end{aligned}
$$

The resource prices $y_{1}, y_{2}$, and $y_{3}$ should be determined by solving the Dakota dual (20). The total price that should be paid for these resources is $48 y_{1}+20 y_{2}+8 y_{3}$. Because the cost of purchasing the resources is to be minimized,

$$
\min w=48 y_{1}+20 y_{2}+8 y_{3}
$$

is the objective function for the Dakota dual.
In setting resource prices, what constraints does the entrepreneur face? Resource prices must be set high enough to induce Dakota to sell. For example, the entrepreneur must offer Dakota at least $\$ 60$ for a combination of resources that includes 8 board feet of lumber, 4 finishing hours, and 2 carpentry hours, because Dakota could, if it desires, use these resources to produce a desk that can be sold for $\$ 60$. The entrepreneur is offering $8 y_{1}+4 y_{2}+2 y_{3}$ for the resources used to produce a desk, so he or she must choose $y_{1}, y_{2}$, and $y_{3}$ to satisfy

$$
8 y_{1}+4 y_{2}+2 y_{3} \geq 60
$$

But this is just the first (or desk) constraint of the Dakota dual. Similar reasoning shows that at least $\$ 30$ must be paid for the resources used to produce a table ( 6 board feet of lumber, 2 finishing hours, and 1.5 carpentry hours). This means that $y_{1}, y_{2}$, and $y_{3}$ must satisfy

$$
6 y_{1}+2 y_{2}+1.5 y_{3} \geq 30
$$

TABLE 21
Relevant Information for Dakota Problem

|  | Resource/Product |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| Resource | Desk | Table | Chair | Amount of <br> Resource <br> Available |
| Lumber (board ft) | 8 | 6 | 1 | 48 |
| Finishing (hours) | 4 | 2 | 1.5 | 20 |
| Carpentry (hours) | 2 | 1.5 | 0.5 | 8 |
| Selling price (\$) | 60 | 30 | 20 |  |

This is the second (or table) constraint of the Dakota dual.
Similarly, the third (or chair) dual constraint,

$$
y_{1}+1.5 y_{2}+0.5 y_{3} \geq 20
$$

states that at least $\$ 20$ (the price of a chair) must be paid for the resources needed to produce a chair ( 1 board foot of lumber, 1.5 finishing hours, and 0.5 carpentry hour). The sign restrictions $y_{1} \geq 0, y_{2} \geq 0$, and $y_{3} \geq 0$ must also hold. Putting everything together, we see that the solution to the dual of the Dakota problem does yield prices for lumber, finishing hours, and carpentry hours. The preceding discussion also shows that the $i$ th dual variable does indeed correspond in a natural way to the $i$ th primal constraint.

In summary, when the primal is a normal max problem, the dual variables are related to the value of the resources available to the decision maker. For this reason, the dual variables are often referred to as resource shadow prices. A more thorough discussion of shadow prices is given in Section 6.8.

## Interpreting the Dual of a Min Problem

To interpret the dual of a min problem, we consider the dual of the diet problem of Section 3.4. In Section 6.5, we found that the diet problem dual was

$$
\begin{array}{lll}
\max z=500 x_{1}+6 x_{2}+10 x_{3}+8 x_{4} & \\
\text { s.t. } & 400 x_{1}+3 x_{2}+2 x_{3}+2 x_{4} \leq 50 & \text { (Brownie constraint) } \\
200 x_{1}+2 x_{2}+2 x_{3}+4 x_{4} \leq 20 & \text { (Ice cream constraint) }  \tag{21}\\
& \begin{aligned}
150 x_{1} & +4 x_{3}+x_{4} \leq 30 \\
500 x_{1} & +4 x_{3}+5 x_{4}
\end{aligned} \leq 80 & \text { (Soda constraint) } \\
& x_{1}, x_{2}, x_{3}, x_{4} \geq 0 & \\
& &
\end{array}
$$

The data for the diet problem are shown in Table 22. To interpret (21), suppose Candice is a "nutrient" salesperson who sells calories, chocolate, sugar, and fat. She wants to ensure that a dieter will meet all of his or her daily requirements by purchasing calories, sugar, fat, and chocolate. Then Candice must determine

$$
\begin{aligned}
& x_{1}=\text { price per calorie to charge dieter } \\
& x_{2}=\text { price per ounce of chocolate to charge dieter } \\
& x_{3}=\text { price per ounce of sugar to charge dieter } \\
& x_{4}=\text { price per ounce of fat to charge dieter }
\end{aligned}
$$

Candice wants to maximize her revenue from selling the dieter the daily ration of required nutrients. Because she will receive $500 x_{1}+6 x_{2}+10 x_{3}+8 x_{4}$ cents in revenue from the dieter, her objective is to

$$
\max z=500 x_{1}+6 x_{2}+10 x_{3}+8 x_{4}
$$

tABLE 22
Relevant Information for Diet Problem

|  | Calories | Chocolate <br> (Ounces) | Sugar <br> (Ounces) | Fat <br> (Ounces) | Price <br> (Cents) |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Brownie | 400 | 3 | 2 | 2 | 50 |
| Ice cream | 200 | 2 | 2 | 4 | 20 |
| Soda | 150 | 0 | 4 | 1 | 30 |
| Cheesecake | 500 | 0 | 4 | 5 | 80 |
| Requirements | 500 | 6 | 10 | 8 |  |

This is the objective function for the dual of the diet problem. But in setting nutrient prices, Candice must set prices low enough so that it will be in the dieter's economic interest to purchase all nutrients from her. For example, by purchasing a brownie for $50 \phi$, the dieter can obtain 400 calories, 3 oz of chocolate, 2 oz of sugar, and 2 oz of fat. So Candice cannot charge more than $50 \notin$ for this combination of nutrients. This leads to the following (brownie) constraint:

$$
400 x_{1}+3 x_{2}+2 x_{3}+2 x_{4} \leq 50
$$

the first constraint in the diet problem dual. Similar reasoning yields the second dual (ice cream) constraint, the third (soda constraint), and the fourth (cheesecake constraint). Again, the sign restrictions $x_{1} \geq 0, x_{2} \geq 0, x_{3} \geq 0$, and $x_{4} \geq 0$ must be satisfied.

Our discussion shows that the optimal value of $x_{i}$ may be interpreted as a price for 1 unit of the nutrient associated with the $i$ th dual constraint. Thus, $x_{1}$ would be the price for 1 calorie, $x_{2}$ would be the price for 1 oz of chocolate, and so on. Again, we see that it is reasonable to associate the $i$ th dual variable $\left(x_{i}\right)$ and the $i$ th primal constraint.

In summary, we have shown that when the primal is a normal max problem or a normal min problem, the dual problem has an intuitive economic interpretation. In Section 6.8 , we explain more about the proper interpretation of the dual variables.

## PROBLEM

## Group A

1 Find the dual of Example 3 in Chapter 3 (an auto company) and give an economic interpretation of the dual problem.

2 Find the dual of Example 2 in Chapter 3 (Dorian Auto) and give an economic interpretation of the dual problem.

### 6.7 The Dual Theorem and Its Consequences

In this section, we discuss one of the most important results in linear programming: the Dual Theorem. In essence, the Dual Theorem states that the primal and dual have equal optimal objective function values (if the problems have optimal solutions). This result is interesting in its own right, but we will see that in proving the Dual Theorem, we gain many important insights into linear programming.

To simplify the exposition, we assume that the primal is a normal max problem with $m$ constraints and $n$ variables. Then the dual problem will be a normal min problem with $m$ variables and $n$ constraints. In this case, the primal and the dual may be written as follows:

$$
\begin{array}{lcc}
\max z= & c_{1} x_{1}+c_{2} x_{2}+\cdots+c_{n} x_{n} \\
\text { s.t. } & a_{11} x_{1}+a_{12} x_{2}+\cdots+a_{1 n} x_{n} \leq b_{1} \\
& a_{21} x_{1}+a_{22} x_{2}+\cdots+a_{2 n} x_{n} \leq b_{2} \\
& \vdots \quad \vdots  \tag{22}\\
& a_{i 1} x_{1}+a_{i 2} x_{2}+\cdots+a_{i n} x_{n} \leq b_{i} \\
& \vdots \quad \vdots & \vdots \\
& a_{m 1} x_{1}+a_{m 2} x_{2}+\cdots+a_{m n} x_{n} \leq b_{m} \\
& x_{j} \geq 0 \quad(j=1,2, \cdots, n)
\end{array}
$$

Primal Problem

From the optimal primal tableau, we find that the optimal dual solution is $z=6, x_{1}=3$, $x_{2}=0, x_{3}=-1$.

## PROBLEMS

## Group A

1 The following questions refer to the Giapetto problem (see Problem 7 of Section 6.3).
a Find the dual of the Giapetto problem.
b Use the optimal tableau of the Giapetto problem to determine the optimal dual solution.
c Verify that the Dual Theorem holds in this instance.
2 Consider the following LP:

$$
\begin{aligned}
& \max z=-2 x_{1}-x_{2}+x_{3} \\
& \text { s.t. } \quad x_{1}+x_{2}+x_{3} \leq 3 \\
& x_{2}+x_{3} \geq 2 \\
& +x_{3}=1 \\
& x_{1} \quad x_{1}, x_{2}, x_{3} \geq 0
\end{aligned}
$$

a Find the dual of this LP.
b After adding a slack variable $s_{1}$, subtracting an excess variable $e_{2}$, and adding artificial variables $a_{2}$ and $a_{3}$, row 0 of the LP's optimal tableau is found to be

$$
z+4 x_{1}+e_{2}+(M-1) a_{2}+(M+2) a_{3}=0
$$

Find the optimal solution to the dual of this LP.
3 For the following LP,

$$
\begin{aligned}
& \max z=-x_{1}+5 x_{2} \\
& \text { s.t. } \quad x_{1}+2 x_{2} \leq 0.5 \\
& -x_{1}+3 x_{2} \leq 0.5 \\
& x_{1}, x_{2} \geq 0
\end{aligned}
$$

row 0 of the optimal tableau is $z+0.4 s_{1}+1.4 s_{2}=$ ? Determine the optimal $z$-value for the given LP.
4 The following questions refer to the Bevco problem of Section 4.10.
a Find the dual of the Bevco problem.
b Use the optimal tableau for the Bevco problem that is given in Section 4.10 to find the optimal solution to the
dual. Verify that the Dual Theorem holds in this instance.
5 Consider the following linear programming problem:

$$
\begin{aligned}
& \max z=4 x_{1}+x_{2} \\
& \text { s.t. } \quad 3 x_{1}+2 x_{2} \leq 6 \\
& 6 x_{1}+3 x_{2} \leq 10 \\
& x_{1}, x_{2} \geq 0
\end{aligned}
$$

Suppose that in solving this problem, row 0 of the optimal tableau is found to be $z+2 x_{2}+s_{2}=\frac{20}{3}$. Use the Dual Theorem to prove that the computations must be incorrect.
6 Show that (for a max problem) if the $i$ th primal constraint is a $\geq$ constraint, then the optimal value of the $i$ th dual variable may be written as (coefficient of $a_{i}$ in optimal row 0 ) - $M$.

## Group B

7 In this problem, we use weak duality to prove Lemma 3.
a Show that Lemma 3 is equivalent to the following: If the dual is feasible, then the primal is bounded. (Hint: Do you remember, from plane geometry, what the contrapositive is?)
b Use weak duality to show the validity of the form of Lemma 3 given in part (a). (Hint: If the dual is feasible, then there must be a dual feasible point having a $w$-value of, say, $w_{o}$. Now use weak duality to show that the primal is bounded.)
8 Following along the lines of Problem 7, use weak duality to prove Lemma 4.
9 Use the information given in Problem 8 of Section 6.3 to determine the dual of the Dorian Auto problem and its optimal solution.

### 6.8 Shadow Prices

We now return to the concept of shadow price that was discussed in Section 6.1. A more formal definition follows.

```
DEFINITION
    The shadow price of the ith constraint is the amount by which the optimal
    z \text { -value is improved (increased in a max problem and decreased in a min problem)}
    if we increase b}\mp@subsup{b}{i}{}\mathrm{ by }1\mathrm{ (from }\mp@subsup{b}{i}{}\mathrm{ to }\mp@subsup{b}{i}{}+1).\mp@subsup{.}{}{\dagger
```

[^10]By using the Dual Theorem, we can easily determine the shadow price of the $i$ th constraint. To illustrate, we find the shadow price of the second constraint (finishing hours) of the Dakota problem. Let $\mathbf{c}_{\mathrm{BV}} B^{-1}=\left[\begin{array}{lll}y_{1} & y_{2} & y_{3}\end{array}\right]=\left[\begin{array}{lll}0 & 10 & 10\end{array}\right]$ be the optimal solution to the dual of the max problem. From the Dual Theorem, we know that

Optimal $z$-value when rhs of constraints are $\left(b_{1}=48, b_{2}=20, b_{3}=8\right)$

$$
\begin{equation*}
=48 y_{1}+20 y_{2}+8 y_{3} \tag{31}
\end{equation*}
$$

What happens to the optimal $z$-value for the Dakota problem if $b_{2}$ (currently 20 finishing hours) is increased by 1 unit (to 21 hours)? We know that changing a right-hand side may cause the current basis to no longer be optimal (see Section 6.3). For the moment, however, we assume that the current basis remains optimal when we increase $b_{2}$ by 1. Then $\mathbf{c}_{\mathrm{BV}}$ and $B^{-1}$ remain unchanged, so the optimal solution to the dual of the Dakota problem remains unchanged.

We next find
Optimal $z$-value when rhs of finishing constraint is $21=48 y_{1}+21 y_{2}+8 y_{3}$
Subtracting (34) from (35) yields

> Change in optimal $z$-value if finishing hours are increased by 1
> $=$ shadow price for finishing constraint 2
> $=y_{2}=10$

This example shows that the shadow price of the ith constraint of a max problem is the optimal value of the ith dual variable. The shadow prices are the dual variables, so we know that the shadow price for $\mathrm{a} \leq$ constraint will be nonnegative; for $\mathrm{a} \geq$ constraint, nonpositive; and for an equality constraint, unrestricted in sign. The examples discussed later in this section give intuitive justifications for these sign conventions.

Similar reasoning can be used to show that if (in a maximization problem) the righthand side of the $i$ th constraint is increased by an amount $\Delta b_{i}$, then (assuming the current basis remains optimal) the new optimal $z$-value may be found from

$$
\begin{equation*}
\text { New optimal } z \text {-value }=\text { old optimal } z \text {-value }+\Delta b_{i}(\text { Constraint } i \text { shadow price }) \tag{37}
\end{equation*}
$$

For a minimization problem, the shadow price of the $i$ th constraint is the amount by which a unit increase in the right-hand side improves, or decreases, the optimal $z$-value (assuming that the current basis remains optimal). It can be shown that the shadow price of the $i$ th constraint of a min problem $=-$ (optimal value of the $i$ th dual variable). If the right-hand side is increased by an amount $\Delta b_{i}$, then (assuming the current basis remains optimal) the new optimal $z$-value may be found from

$$
\text { New optimal } z \text {-value }=\text { old optimal } z \text {-value }-\Delta b_{i}(\text { Constraint } i \text { shadow price })
$$

The following three examples should clarify the shadow price concept.

## EXAMPLE 11 Shadow Prices for Normal Max Problem

For the Dakota problem:
1 Find and interpret the shadow prices
2 If 18 finishing hours were available, what would be Dakota's revenue? (It can be shown by the methods of Section 6.3 that if $16 \leq$ finishing hours $\leq 24$, the current basis remains optimal.)
3 If 9 carpentry hours were available, what would be Dakota's revenue? (For $\frac{20}{3} \leq$ carpentry hours $\leq 10$, the current basis remains optimal.)

4 If 30 board feet of lumber were available, what would be Dakota's revenue? (For $24 \leq$ lumber $\leq \infty$, the current basis remains optimal.)

5 If 30 carpentry hours were available, why couldn't the shadow price for the carpentry constraint be used to determine the new $z$-value?

Solution 1 In Section 6.7, we found the optimal solution to the Dakota dual to be $y_{1}=0, y_{2}=$ $10, y_{3}=10$. Thus, the shadow price for the lumber constraint is 0 ; for the finishing constraint, 10 ; and for the carpentry constraint, 10 . The fact that the lumber constraint has a shadow price of 0 means that increasing the amount of available lumber by 1 board foot (or any amount) will not increase revenue. This is reasonable because we are currently using only 24 of the available 48 board feet of lumber, so adding any more will not do Dakota any good. Dakota's revenue would increase by $\$ 10$ if 1 more finishing hour were available. Similarly, 1 more carpentry hour would increase Dakota's revenue by $\$ 10$. In this problem, the shadow price of the $i$ th constraint may be thought of as the maximum amount that the company would pay for an extra unit of the resource associated with the $i t h$ constraint. For example, an extra carpentry hour would raise revenue by $y_{3}=\$ 10$ (see Example 12 for a max problem in which this interpretation is invalid). Thus, Dakota could pay up to $\$ 10$ for an extra carpentry hour and still be better off. Similarly, the company would be willing to pay nothing ( $\$ 0$ ) for an extra board foot of lumber and up to $\$ 10$ for an extra finishing hour. To answer questions 2-4, we apply (37), using the fact that the old $z$-value $=280$.
$2 y_{2}=10, \Delta b_{2}=18-20=-2$. The current basis is still optimal because $16 \leq 18$ $\leq 24$. Then (37) yields (new revenue) $=280+10(-2)=\$ 260$.
$3 y_{3}=10, \Delta b_{3}=9-8=1$. Because $\frac{20}{3} \leq 9 \leq 10$, the current basis remains optimal. Then (37) yields (new revenue) $=280+10(1)=\$ 290$.
$4 y_{1}=0, \Delta b_{1}=30-48=-18$. Because $24 \leq 30 \leq \infty$, the current basis is still optimal. Then (37) yields (new revenue) $=280+0(-18)=\$ 280$.

5 If $b_{3}=30$, the current basis is no longer optimal, because $30>10$. This means that BV (and therefore $\mathbf{c}_{\mathrm{BV}} B^{-1}$ ) changes, and we cannot use the current set of shadow prices to determine the new revenue level.

## Intuitive Explanation of the Sign of Shadow Prices

We can now give an intuitive explanation of why (in a max problem) the shadow price of $\mathrm{a} \leq$ constraint will always be nonnegative. Consider the following situation: We are given two LP max problems (LP 1 and LP 2) that have the same objective functions. Suppose that every point that is feasible for LP 1 is also feasible for LP 2. This means that LP 2's feasible region contains all the points in LP 1's feasible region and possibly some other points. Then the optimal $z$-value for LP 2 must be at least as large as the optimal $z$-value for LP 1. To see this, suppose that point $x^{\prime}$ (with $z$-value $z^{\prime}$ ) is optimal for LP 1. Because $x^{\prime}$ is also feasible for LP 2 (which has the same objective function as LP 1), LP 2 can attain a $z$-value of $z^{\prime}$ (by using the feasible point $x^{\prime}$ ). It is also possible that by using one of the points feasible for only LP 2 (and not for LP 1), LP 2 might do better than $z^{\prime}$. In short, adding points to the feasible region of a max problem cannot decrease the optimal $z$-value.

We can use this observation to show why $\mathrm{a} \leq$ constraint must have a nonnegative shadow price. For the Dakota problem, if we increase the right-hand side of the carpentry constraint by 1 (from 8 to 9 ), we see that all points that were originally feasible re-
main feasible, and some new points (which use $>8$ and $\leq 9$ carpentry hours) may be feasible. Thus, the optimal $z$-value cannot decrease, and the shadow price for the carpentry constraint must be nonnegative.

The purpose of the following example is to show that (contrary to what many books say) the shadow price of a constraint is not always the maximum price you would be willing to pay for an additional unit of a resource.

## EXAMPLE 12 Shadow Price as a Premium

Leatherco manufactures belts and shoes. A belt requires 2 square yards of leather and 1 hour of skilled labor. A pair of shoes requires 3 sq yd of leather and 2 hours of skilled labor. As many as 25 sq yd of leather and 15 hours of skilled labor can be purchased at a price of $\$ 5 / \mathrm{sq}$ yd of leather and $\$ 10$ /hour of skilled labor. A belt sells for $\$ 23$, and a pair of shoes sells for $\$ 40$. Leatherco wants to maximize profits (revenues - costs). Formulate an LP that can be used to maximize Leatherco's profits. Then find and interpret the shadow prices for this LP.
Solution Define

$$
\begin{aligned}
& x_{1}=\text { number of belts produced } \\
& x_{2}=\text { number of pairs of shoes produced }
\end{aligned}
$$

After noting that

$$
\begin{aligned}
& \text { Cost/belt }=2(5)+1(10)=\$ 20 \\
& \text { Cost/pair of shoes }=3(5)+2(10)=\$ 35
\end{aligned}
$$

we find that Leatherco's objective function is

$$
\max z=(23-20) x_{1}+(40-35) x_{2}=3 x_{1}+5 x_{2}
$$

Leatherco faces the following two constraints:
Constraint 1 Leatherco can use at most 25 sq yd of leather.
Constraint 2 Leatherco can use at most 15 hours of skilled labor.
Constraint 1 is expressed by

$$
2 x_{1}+3 x_{2} \leq 25 \quad \text { (Leather constraint) }
$$

while Constraint 2 is expressed by

$$
x_{1}+2 x_{2} \leq 15 \quad \text { (Skilled-labor constraint) }
$$

After adding the sign restrictions $x_{1} \geq 0$ and $x_{2} \geq 0$, we obtain the following LP:

$$
\begin{array}{lrl}
\max z=3 x_{1}+5 x_{2} & & \\
\text { s.t. } & 2 x_{1}+3 x_{2} & \leq 25 \\
& & \text { (Leather constraint) } \\
& x_{1}+2 x_{2} & \leq 15 \\
& & \text { (Skilled-labor constraint) } \\
x_{1}, x_{2} & \geq 0 &
\end{array}
$$

After adding slack variables $s_{1}$ and $s_{2}$ to the leather and skilled-labor constraints, respectively, we obtain the optimal tableau shown in Table 27. Thus, the optimal solution to Leatherco's problem is $z=40, x_{1}=5, x_{2}=5$. The shadow prices are

$$
\begin{aligned}
& y_{1}=\text { leather shadow price }=\text { coefficient of } s_{1} \text { in optimal row } 0=1 \\
& y_{2}=\text { skilled-labor shadow price }=\operatorname{coefficient} \text { of } s_{2} \text { in optimal row } 0=1
\end{aligned}
$$

TABLE 27
Optimal Tableau for Leatherco

|  | Basic <br> Variable |
| :---: | :---: |
| $z+s_{1}+s_{2}=40$ | $z=40$ |
| $x_{1}+2 s_{1}-3 s_{2}=5$ | $x_{1}=5$ |
| $x_{2}-s_{1}+2 s_{2}=5$ | $x_{2}=5$ |

The meaning of the leather shadow price is that if one more square yard of leather were available, then Leatherco's objective function (profits) would increase by $\$ 1$. Let's look further at what happens if an additional square yard of leather is available. Because $s_{1}$ is nonbasic, the extra square yard of leather will be purchased. Also, because $s_{2}$ is nonbasic, we will still use all available labor. This means that the $\$ 1$ increase in profits includes the cost of purchasing an extra square yard of leather. If the availability of an extra square yard of leather increases profits by $\$ 1$, then it must be increasing revenue by $1+5=\$ 6$. Thus, the maximum amount Leatherco should pay for an extra square yard of leather is $\$ 6$ (not $\$ 1$ ).

Another way to see this is as follows: If we purchase another square yard of leather at the current price of $\$ 5$, profits increase by $y_{1}=\$ 1$. If we purchase another square yard of leather at a price of $\$ 6=\$ 5+\$ 1$, then profits increase by $\$ 1-\$ 1=\$ 0$. Thus, the most Leatherco would be willing to pay for an extra square yard of leather is $\$ 6$.

Similarly, the most Leatherco would be willing to pay for an extra hour of labor is $y_{2}+$ (cost of an extra hour of skilled labor) $=1+10=\$ 11$. In this problem, we see that the shadow price for a resource represents the premium over and above the cost of the resource that Leatherco would be willing to pay for an extra unit of resource.

The two preceding examples show that we must be careful when interpreting the shadow price of $\mathrm{a} \leq$ constraint. Remember that the shadow price for a constraint in a max problem is the amount by which the objective function increases if the right-hand side is increased by 1 .

The following example illustrates the interpretation of the shadow prices of $\geq$ and equality constraints.

## EXAMPLE 13 Shadow Prices for $\geq$ and $=$ Constraints

Steelco has received an order for 100 tons of steel. The order must contain at least 3.5 tons of nickel, at most 3 tons of carbon, and exactly 4 tons of manganese. Steelco receives $\$ 20 /$ ton for the order. To fill the order, Steelco can combine four alloys, whose chemical composition is given in Table 28. Steelco wants to maximize the profit (revenues - costs) obtained from filling the order. Formulate the appropriate LP. Also find and interpret the shadow prices for each constraint.
Solution After we define $x_{i}=$ number of tons of alloy $i$ used to fill the order, Steelco's LP is seen to be

$$
\begin{array}{lrlrl}
\max z=(20-12) x_{1}+(20-10) x_{2}+(20-8) x_{3} & & (20-6) x_{4} \\
\text { s.t. } & 0.06 x_{1}+0.03 x_{2}+0.02 x_{3}+0.01 x_{4} & \geq 3.5 & & \text { (Nickel constraint) } \\
& 0.03 x_{1}+0.02 x_{2}+0.05 x_{3}+0.06 x_{4} & \leq 3 & & \text { (Carbon constraint) } \\
& 0.08 x_{1}+0.03 x_{2}+0.02 x_{3}+0.01 x_{4} & =4 & & \text { (Manganese constraint) } \\
& x_{1}+\quad x_{2}+\quad x_{3}+\quad x_{4} & =100 & & \text { (Order size }=100 \text { tons) } \\
& & &
\end{array}
$$

tABLE 28
Relevant Information for Steelco

|  | Alloy (\%) |  |  |  |
| :--- | ---: | ---: | ---: | ---: |
| Cement | 1 | 2 | 3 | 4 |
| Nickel | 6 | 3 | 2 | 1 |
| Carbon | 3 | 2 | 5 | 6 |
| Manganese | 8 | 3 | 2 | 1 |
| Cost/ton (\$) | 12 | 10 | 8 | 6 |

After adding a slack variable $s_{2}$, subtracting an excess variable $e_{1}$, and adding artificial variables $a_{1}, a_{3}$, and $a_{4}$, the following optimal solution is obtained: $z=1,000, s_{2}=0.25$, $x_{1}=25, x_{2}=62.5, x_{4}=12.5, e_{1}=0, x_{3}=0$. The optimal row 0 is

$$
z+400 e_{1}+(M-400) a_{1}+(M+200) a_{3}+(M+16) a_{4}=1,000
$$

Using (31), (31'), and (31"), we obtain

$$
\begin{aligned}
\text { Shadow price of nickel constraint } & =-\left(\text { coefficient of } e_{1} \text { in optimal row } 0\right) \\
& =-400 \\
\text { Shadow price of carbon constraint } & =\text { coefficient of } s_{2} \text { in optimal row } 0 \\
& =0 \\
\text { Shadow price of manganese constraint } & =\left(\text { coefficient of } a_{3} \text { in optimal row } 0\right)-M \\
& =200 \\
\text { Shadow price of order size constraint } & =\left(\text { coefficient of } a_{4} \text { in optimal row } 0\right)-M \\
& =16
\end{aligned}
$$

By the sensitivity analysis procedures of Section 6.3, it can be shown that the current basis remains optimal if $3.46 \leq b_{1} \leq 3.6$. As long as the nickel requirement is in this range, increasing the nickel requirement by an amount $\Delta b_{1}$ will increase Steelco's profits by $-400 \Delta b_{1}$. For example, increasing the nickel requirement to 3.55 tons $\left(\Delta b_{1}=0.05\right)$ would "increase" (actually decrease) profits by $-400(0.05)=\$ 20$. The nickel constraint has a negative shadow price because increasing the right-hand side of the nickel constraint makes it harder to satisfy the nickel constraint. In fact, an increase in the nickel requirement forces Steelco to use more of the expensive type 1 alloy. This raises costs and lowers profits. As we have already seen, the shadow price of $\mathrm{a} \geq$ constraint (in a max problem) will always be nonpositive, because increasing the right-hand side of a $\geq$ constraint eliminates points from the feasible region. Thus, the optimal $z$-value must decrease or remain unchanged.

By the Section 6.3 sensitivity analysis procedures, for $2.75 \leq b_{2} \leq \infty$, the current basis remains optimal. As stated before, the carbon constraint has a zero shadow price. This means that if we increase Steelco's carbon requirement, Steelco's profit will not change. Intuitively, this is because our present optimal solution contains only $2.75<3$ tons of carbon. Thus, relaxing the carbon requirement won't enable Steelco to reduce costs, so Steelco's profit will remain unchanged.

By the sensitivity analysis procedures, the current basis remains optimal if $3.83 \leq$ $b_{3} \leq 4.07$. The shadow price of the third (manganese) constraint is 200 , so we know that as long as the manganese requirement remains in the given range, increasing it by an amount of $\Delta b_{3}$ will increase profit by $200 \Delta b_{3}$. For example, if the manganese requirement were 4.05 tons $\left(\Delta b_{3}=0.05\right)$, then profits would increase by $(0.05) 200=\$ 10$.

By the sensitivity analysis procedures, the current basis remains optimal if $91.67 \leq$ $b_{4} \leq 103.12$. Because the shadow price of the fourth (order size) constraint is 16 , increasing the order size by $\Delta b_{4}$ tons (with nickel, carbon, and manganese requirements unchanged) would increase profits by $16 \Delta b_{4}$. For example, the profit from a 103 -ton order that required $\geq 3.5$ tons of nickel, $\leq 3$ tons of carbon, and exactly 4 tons of manganese would be $1,000+3(16)=\$ 1,048$.

In this problem, both equality constraints had positive shadow prices. In general, we know that it is possible for an equality constraint's dual variable (and shadow price) to be negative. If this occurs, then the equality constraint will have a negative shadow price. To illustrate this possibility, suppose that Steelco's customer required exactly 4.5 tons of manganese in the order. Because $4.5>4.07$, the current basis is no longer optimal. If we solve Steelco's LP again, it can be shown that the shadow price for the manganese constraint has changed to -54.55 . This means that an increase in the manganese requirement will decrease Steelco's profits.

## Interpretation of the Dual Prices Column of the LINDO Output

For a max problem, LINDO gives the values of the shadow prices in the DUAL PRICES column of the output. The dual price for row $i+1$ on the LINDO output is the shadow price for the $i$ th constraint and the optimal value for the $i$ th dual variable. Thus, in Figure 4, we see that for the Dakota problem,

$$
\begin{aligned}
& y_{1}=\text { shadow price for lumber constraint }=\text { row } 2 \text { dual price }=0 \\
& y_{2}=\text { shadow price for finishing constraint }=\text { row } 3 \text { dual price }=10 \\
& y_{3}=\text { shadow price for carpentry constraint }=\text { row } 4 \text { dual price }=10
\end{aligned}
$$

For a maximization problem, the vector $\mathbf{c}_{\mathrm{BV}} B^{-1}$ (needed for pricing out new activities) is the same as the vector of dual prices given in the LINDO output. For the Dakota problem, we would price out new activities using $\mathbf{c}_{\mathrm{BV}} B^{-1}=\left[\begin{array}{lll}0 & 10 & 10\end{array}\right]$.

For a minimization problem, the entry in the DUAL PRICE column for any constraint is the shadow price. Thus, from the LINDO printout in Figure 6, we find that the shadow prices for the constraints in the diet problem are as follows: calorie $=0$; chocolate $=$ $-2.5 \phi$; sugar $=-7.5 \phi$; and fat $=0$. This implies that

1 Increasing the calorie requirement by 1 will leave the cost of the optimal diet unchanged.

2 Increasing the chocolate requirement by 1 oz will decrease the cost of the optimal diet by $-2.5 \phi$ (that is, increase the cost of the optimal diet by $2.5 \phi$ ).

3 Increasing the sugar requirement by 1 oz will decrease the cost of the optimal diet by $-7.5 \phi$ (that is, increase the cost of the optimal diet by $7.5 \phi$ ).

4 Increasing the fat requirement by 1 oz will leave the cost of the optimal diet unchanged.

The entry in the DUAL PRICE column for any constraint is, however, the negative of the constraint's dual variable. Thus, for the diet problem, we see from Figure 6 that the optimal dual solution to the diet problem is given by $\mathbf{c}_{\mathrm{BV}} B^{-1}=\left[\begin{array}{llll}0 & 2.5 & 7.5 & 0\end{array}\right]$. When


Many important optimization problems can best be analyzed by means of a graphical or network representation. In this chapter, we consider four specific network models-shortest-path problems, maximum-flow problems, CPM-PERT project-scheduling models, and minimum-spanning tree problems-for which efficient solution procedures exist. We also discuss minimum-cost network flow problems (MCNFPs), of which transportation, assignment, transshipment, shortestpath, and maximum-flow problems and the CPM project-scheduling models are all special cases. Finally, we discuss a generalization of the transportation simplex, the network simplex, which can be used to solve MCNFPs. We begin the chapter with some basic terms used to describe graphs and networks.

### 8.1 Basic Definitions

A graph, or network, is defined by two sets of symbols: nodes and arcs. First, we define a set (call it $V$ ) of points, or vertices. The vertices of a graph or network are also called nodes.

We also define a set of $\operatorname{arcs} A$.

An arc consists of an ordered pair of vertices and represents a possible direction of motion that may occur between vertices.

For our purposes, if a network contains an $\operatorname{arc}(j, k)$, then motion is possible from node $j$ to node $k$. Suppose nodes $1,2,3$, and 4 of Figure 1 represent cities, and each arc represents a (one-way) road linking two cities. For this network, $V=\{1,2,3,4\}$ and $A=$ $\{(1,2),(2,3),(3,4),(4,3),(4,1)\}$. For the $\operatorname{arc}(j, k)$, node $j$ is the initial node, and node $k$ is the terminal node. The arc $(j, k)$ is said to go from node $j$ to node $k$. Thus, the arc $(2,3)$ has initial node 2 and terminal node 3 , and it goes from node 2 to node 3 . The arc $(2,3)$ may be thought of as a (one-way) road on which we may travel from city 2 to city 3. In Figure 1, the arcs show that travel is allowed from city 3 to city 4, and from city 4 to city 3 , but that travel between the other cities may be one way only.

Later, we often discuss a group or collection of arcs. The following definitions are convenient ways to describe certain groups or collections of arcs.

DEFINITION ■ A sequence of arcs such that every arc has exactly one vertex in common with the previous arc is called a chain.

FIGURE 1 Example of a Network


DEFINITION $\quad$ A path is a chain in which the terminal node of each arc is identical to the initial node of the next arc.

For example, in Figure 1, $(1,2)-(2,3)-(4,3)$ is a chain but not a path; $(1,2)-(2,3)-$ $(3,4)$ is a chain and a path. The path $(1,2)-(2,3)-(3,4)$ represents a way to travel from node 1 to node 4 .

### 8.2 Shortest-Path Problems

In this section, we assume that each arc in the network has a length associated with it. Suppose we start at a particular node (say, node 1). The problem of finding the shortest path (path of minimum length) from node 1 to any other node in the network is called a shortest-path problem. Examples 1 and 2 are shortest-path problems.

## EXAMPLE 1 Shortest Path

Let us consider the Powerco example (Figure 2). Suppose that when power is sent from plant 1 (node 1 ) to city 1 (node 6 ), it must pass through relay substations (nodes $2-5$ ). For any pair of nodes between which power can be transported, Figure 2 gives the distance (in miles) between the nodes. Thus, substations 2 and 4 are 3 miles apart, and power cannot be sent between substations 4 and 5 . Powerco wants the power sent from plant 1 to city 1 to travel the minimum possible distance, so it must find the shortest path in Figure 2 that joins node 1 to node 6 .

If the cost of shipping power were proportional to the distance the power travels, then knowing the shortest path between plant 1 and city 1 in Figure 2 (and the shortest path between plant $i$ and city $j$ in similar diagrams) would be necessary to determine the shipping costs for the transportation version of the Powerco problem discussed in Chapter 7.

FIGURE 2 Network for Powerco


I have just purchased (at time 0 ) a new car for $\$ 12,000$. The cost of maintaining a car during a year depends on its age at the beginning of the year, as given in Table 1. To avoid the high maintenance costs associated with an older car, I may trade in my car and purchase a new car. The price I receive on a trade-in depends on the age of the car at the time of trade-in (see Table 2). To simplify the computations, we assume that at any time, it costs $\$ 12,000$ to purchase a new car. My goal is to minimize the net cost (purchasing costs + maintenance costs - money received in trade-ins) incurred during the next five years. Formulate this problem as a shortest-path problem.

Solution Our network will have six nodes (1,2,3,4,5, and 6). Node $i$ is the beginning of year $i$. For $i<j$, an arc $(i, j)$ corresponds to purchasing a new car at the beginning of year $i$ and keeping it until the beginning of year $j$. The length of $\operatorname{arc}(i, j)$ (call it $c_{i j}$ ) is the total net cost incurred in owning and operating a car from the beginning of year $i$ to the beginning of year $j$ if a new car is purchased at the beginning of year $i$ and this car is traded in for a new car at the beginning of year $j$. Thus,

$$
\begin{aligned}
c_{i j}= & \text { maintenance cost incurred during years } i, i+1, \ldots, j-1 \\
& + \text { cost of purchasing car at beginning of year } i \\
& - \text { trade-in value received at beginning of year } j
\end{aligned}
$$

Applying this formula to the information in the problem yields (all costs are in thousands)

$$
\begin{array}{ll}
c_{12}=2+12-7=7 & c_{16}=2+4+5+9+12+12-0=44 \\
c_{13}=2+4+12-6=12 & c_{23}=2+12-7=7 \\
c_{14}=2+4+5+12-2=21 & c_{24}=2+4+12-6=12 \\
c_{15}=2+4+5+9+12-1=31 & c_{25}=2+4+5+12-2=21
\end{array}
$$

table 1
Car Maintenance Costs

| Age of Car <br> (Years) | Annual <br> Maintenance <br> Cost (\$) |
| :--- | ---: |
| 0 | 2,000 |
| 1 | 4,000 |
| 2 | 5,000 |
| 3 | 9,000 |
| 4 | 12,000 |

table 2
Car Trade-in Prices

| Age of Car <br> (Years) | Trade-in Price |
| :--- | :---: |
| 1 | 7,000 |
| 2 | 6,000 |
| 3 | 2,000 |
| 4 | 1,000 |
| 5 | 0 |

FIGURE 3 Network for Minimizing Car Costs

$c_{26}=2+4+5+9+12-1=31 \quad c_{45}=2+12-7=7$
$c_{34}=2+12-7=7 \quad c_{46}=2+4+12-6=12$
$c_{35}=2+4+12-6=12 \quad c_{56}=2+12-7=7$
$c_{36}=2+4+5+12-2=21$
We now see that the length of any path from node 1 to node 6 is the net cost incurred during the next five years corresponding to a particular trade-in strategy. For example, suppose I trade in the car at the beginning of year 3 and next trade in the car at the end of year 5 (the beginning of year 6). This strategy corresponds to the path 1-3-6 in Figure 3. The length of this path $\left(c_{13}+c_{36}\right)$ is the total net cost incurred during the next five years if I trade in the car at the beginning of year 3 and at the beginning of year 6 . Thus, the length of the shortest path from node 1 to node 6 in Figure 3 is the minimum net cost that can be incurred in operating a car during the next five years.

## Dijkstra's Algorithm

Assuming that all arc lengths are nonnegative, the following method, known as Dijkstra's algorithm, can be used to find the shortest path from a node (say, node 1) to all other nodes. To begin, we label node 1 with a permanent label of 0 . Then we label each node $i$ that is connected to node 1 by a single arc with a "temporary" label equal to the length of the arc joining node 1 to node $i$. Each other node (except, of course, for node 1 ) will have a temporary label of $\infty$. Choose the node with the smallest temporary label and make this label permanent.

Now suppose that node $i$ has just become the $(k+1)$ th node to be given a permanent label. Then node $i$ is the $k$ th closest node to node 1 . At this point, the temporary label of any node (say, node $i^{\prime}$ ) is the length of the shortest path from node 1 to node $i^{\prime}$ that passes only through nodes contained in the $k-1$ closest nodes to node 1 . For each node $j$ that now has a temporary label and is connected to node $i$ by an arc, we replace node $j$ 's temporary label with

$$
\min \left\{\begin{array}{l}
\text { node } j \prime \text { 's current temporary label } \\
\text { node } i \text { 's permanent label }+ \text { length of } \operatorname{arc}(i, j)
\end{array}\right.
$$

(Here, $\min \{a, b\}$ is the smaller of $a$ and $b$.) The new temporary label for node $j$ is the length of the shortest path from node 1 to node $j$ that passes only through nodes contained in the $k$ closest nodes to node 1 . We now make the smallest temporary label a permanent label. The node with this new permanent label is the $(k+1)$ th closest node to node 1 . Continue this process until all nodes have a permanent label. To find the shortest path from node 1 to node $j$, work backward from node $j$ by finding nodes having labels dif-
fering by exactly the length of the connecting arc. Of course, if we want the shortest path from node 1 to node $j$, we can stop the labeling process as soon as node $j$ receives a permanent label.

To illustrate Dijkstra's algorithm, we find the shortest path from node 1 to node 6 in Figure 2. We begin with the following labels (a $*$ represents a permanent label, and the $i$ th number is the label of the node $i$ ): $\left[\begin{array}{llllll}0^{*} & 4 & 3 & \infty & \infty & \infty\end{array}\right]$. Node 3 now has the smallest temporary label. We therefore make node 3's label permanent and obtain the following labels:

$$
\left[\begin{array}{llllll}
0^{*} & 4 & 3^{*} & \infty & \infty & \infty
\end{array}\right]
$$

We now know that node 3 is the closest node to node 1 . We compute new temporary labels for all nodes that are connected to node 3 by a single arc. In Figure 2 that is node 5.

$$
\text { New node } 5 \text { temporary label }=\min \{\infty, 3+3\}=6
$$

Node 2 now has the smallest temporary label; we now make node 2's label permanent. We now know that node 2 is the second closest node to node 1 . Our new set of labels is

$$
\left[\begin{array}{llllll}
0^{*} & 4^{*} & 3^{*} & \infty & 6 & \infty
\end{array}\right]
$$

Because nodes 4 and 5 are connected to the newly permanently labeled node 2 , we must change the temporary labels of nodes 4 and 5. Node 4's new temporary label is min $\{\infty$, $4+3\}=7$ and node 5 's new temporary label is $\min \{6,4+2\}=6$. Node 5 now has the smallest temporary label, so we make node 5's label permanent. We now know that node 5 is the third closest node to node 1 . Our new labels are

$$
\left[\begin{array}{llllll}
0^{*} & 4^{*} & 3^{*} & 7 & 6^{*} & \infty
\end{array}\right]
$$

Only node 6 is connected to node 5 , so node 6 's temporary label will change to min $\{\infty, 6+2\}=8$. Node 4 now has the smallest temporary label, so we make node 4 's label permanent. We now know that node 4 is the fourth closest node to node 1 . Our new labels are

$$
\left[\begin{array}{llllll}
0^{*} & 4^{*} & 3^{*} & 7^{*} & 6^{*} & 8
\end{array}\right]
$$

Because node 6 is connected to the newly permanently labeled node 4 , we must change node 6 's temporary label to $\min \{8,7+2\}=8$. We can now make node 6 's label permanent. Our final set of labels is $\left[\begin{array}{lllllll}0^{*} & 4^{*} & 3^{*} & 7^{*} & 6^{*} & 8^{*}\end{array}\right]$. We can now work backward and find the shortest path from node 1 to node 6 . The difference between node 6's and node 5 's permanent labels is $2=$ length of arc $(5,6)$, so we go back to node 5 . The difference between node 5 's and node 2 's permanent labels is $2=$ length of arc $(2,5)$, so we may go back to node 2 . Then, of course, we must go back to node 1 . Thus, $1-2-5-6$ is a shortest path (of length 8 ) from node 1 to node 6 . Observe that when we were at node 5 , we could also have worked backward to node 3 and obtained the shortest path 1-3-5-6.

## The Shortest-Path Problem as a Transshipment Problem

Finding the shortest path between node $i$ and node $j$ in a network may be viewed as a transshipment problem. Simply try to minimize the cost of sending one unit from node $i$ to node $j$ (with all other nodes in the network being transshipment points), where the cost of sending one unit from node $k$ to node $k^{\prime}$ is the length of arc ( $k, k^{\prime}$ ) if such an arc exists and is $M$ (a large positive number) if such an arc does not exist. As in Section 7.6, the cost of shipping one unit from a node to itself is zero. Following the method described in Section 7.6, this transshipment problem may be transformed into a balanced transportation problem.


To illustrate the preceding ideas, we formulate the balanced transportation problem associated with finding the shortest path from node 1 to node 6 in Figure 2. We want to send one unit from node 1 to node 6 . Node 1 is a supply point, node 6 is a demand point, and nodes $2,3,4$, and 5 will be transshipment points. Using $s=1$, we obtain the balanced transportation problem shown in Table 3. This transportation problem has two optimal solutions:
$1 z=4+2+2=8, x_{12}=x_{25}=x_{56}=x_{33}=x_{44}=1 \quad($ all other variables equal 0$)$. This solution corresponds to the path $1-2-5-6$.
$2 z=3+3+2=8, x_{13}=x_{35}=x_{56}=x_{22}=x_{44}=1 \quad$ (all other variables equal 0 ). This solution corresponds to the path $1-3-5-6$.

REMARK After formulating a shortest-path problem as a transshipment problem, the problem may be solved easily by using LINGO or a spreadsheet optimizer. See Section 7.1 for details.

## PROBLEMS

## Group A

1 Find the shortest path from node 1 to node 6 in Figure 3.
2 Find the shortest path from node 1 to node 5 in Figure 4.
3 Formulate Problem 2 as a transshipment problem.
4 Use Dijkstra's algorithm to find the shortest path from node 1 to node 4 in Figure 5. Why does Dijkstra's algorithm fail to obtain the correct answer?

## FIGURE 4

## Network for Problem 2



FIGURE 5
Network for Problem 4


5 Suppose it costs $\$ 10,000$ to purchase a new car. The annual operating cost and resale value of a used car are shown in Table 4. Assuming that one now has a new car, determine a replacement policy that minimizes the net costs of owning and operating a car for the next six years.

TABLE 4

| Age of Car <br> (Years) | Resale <br> Value (S) | Operating <br> Cost (\$) |
| :--- | :---: | ---: |
| 1 | 7,000 | 300 (year 1) |
| 2 | 6,000 | 500 (year 2) |
| 3 | 4,000 | 800 (year 3) |
| 4 | 3,000 | 1,200 (year 4) |
| 5 | 2,000 | 1,600 (year 5) |
| 6 | 1,000 | 2,200 (year 6) |

6 It costs $\$ 40$ to buy a telephone from the department store. Assume that I can keep a telephone for at most five years and that the estimated maintenance cost each year of operation is as follows: year $1, \$ 20$; year $2, \$ 30$; year 3 , $\$ 40$; year $4, \$ 60$; year 5, $\$ 70$. I have just purchased a new telephone. Assuming that a telephone has no salvage value, determine how to minimize the total cost of purchasing and operating a telephone for the next six years.
7 At the beginning of year 1 , a new machine must be purchased. The cost of maintaining a machine $i$ years old is given in Table 5.

The cost of purchasing a machine at the beginning of each year is given in Table 6.

There is no trade-in value when a machine is replaced. Your goal is to minimize the total cost (purchase plus maintenance) of having a machine for five years. Determine the years in which a new machine should be purchased.

## Group B

$8^{\dagger}$ A library must build shelving to shelve 200 4-inch high books, 1008 -inch high books, and 8012 -inch high books.

TABLE 5

| Age at Beginning <br> of Year | Maintenance Cost <br> for Next Year (\$) |
| :--- | :---: |
| 0 | 38,000 |
| 1 | 50,000 |
| 2 | 97,000 |
| 3 | 182,000 |
| 4 | 304,000 |

table 6

| Year | Purchase Cost (\$) |
| :--- | :---: |
| 1 | 170,000 |
| 2 | 190,000 |
| 3 | 210,000 |
| 4 | 250,000 |
| 5 | 300,000 |

Each book is 0.5 inch thick. The library has several ways to store the books. For example, an 8 -inch high shelf may be built to store all books of height less than or equal to 8 inches, and a 12 -inch high shelf may be built for the 12 -inch books. Alternatively, a 12 -inch high shelf might be built to store all books. The library believes it costs $\$ 2,300$ to build a shelf and that a cost of $\$ 5$ per square inch is incurred for book storage. (Assume that the area required to store a book is given by height of storage area times book's thickness.)

Formulate and solve a shortest-path problem that could be used to help the library determine how to shelve the books at minimum cost. (Hint: Have nodes 0, 4, 8, and 12, with $c_{i j}$ being the total cost of shelving all books of height $>i$ and $\leq j$ on a single shelf.)

9 A company sells seven types of boxes, ranging in volume from 17 to 33 cubic feet. The demand and size of each box is given in Table 7. The variable cost (in dollars) of producing each box is equal to the box's volume. A fixed cost of $\$ 1,000$ is incurred to produce any of a particular box. If the company desires, demand for a box may be satisfied by a box of larger size. Formulate and solve a shortest-path problem whose solution will minimize the cost of meeting the demand for boxes.

10 Explain how by solving a single transshipment problem you can find the shortest path from node 1 in a network to each other node in the network.
table 7

|  | Box |  |  |  |  |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: | :---: |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |
| Size | 33 | 30 | 26 | 24 | 19 | 18 | 17 |  |
| Demand | 400 | 300 | 500 | 700 | 200 | 400 | 200 |  |

[^11]
### 8.3 Maximum-Flow Problems

Many situations can be modeled by a network in which the arcs may be thought of as having a capacity that limits the quantity of a product that may be shipped through the arc. In these situations, it is often desired to transport the maximum amount of flow from a starting point (called the source) to a terminal point (called the sink). Such problems are
called maximum-flow problems. Several specialized algorithms exist to solve maximumflow problems. In this section, we begin by showing how linear programming can be used to solve a maximum-flow problem. Then we discuss the Ford-Fulkerson (1962) method for solving maximum-flow problems.

## LP Solution of Maximum-Flow Problems

## EXAMPLE 3 Maximum Flow

Sunco Oil wants to ship the maximum possible amount of oil (per hour) via pipeline from node so to node si in Figure 6. On its way from node so to node si, oil must pass through some or all of stations 1, 2, and 3. The various arcs represent pipelines of different diameters. The maximum number of barrels of oil (millions of barrels per hour) that can be pumped through each arc is shown in Table 8. Each number is called an arc capacity. Formulate an LP that can be used to determine the maximum number of barrels of oil per hour that can be sent from so to si.

Solution Node so is called the source node because oil flows out of it but no oil flows into it. Analogously, node si is called the sink node because oil flows into it and no oil flows out of it. For reasons that will soon become clear, we have added an artificial arc $a_{0}$ from the sink to the source. The flow through $a_{0}$ is not actually oil, hence the term artificial arc.

To formulate an LP that will yield the maximum flow from node so to si, we observe that Sunco must determine how much oil (per hour) should be sent through arc (i,j). Thus, we define
$x_{i j}=$ millions of barrels of oil per hour that will pass through arc $(i, j)$ of pipeline
As an example of a possible flow (termed a feasible flow), consider the flow indentified by the numbers in parentheses in Figure 6.

$$
x_{s o, 1}=2, \quad x_{13}=0, \quad x_{12}=2, \quad x_{3, s i}=0, \quad x_{2, s i}=2, \quad x_{s i, s o}=2, \quad x_{s o, 2}=0
$$

FIGURE 6 Network for Sunco Oil

table 8
Arc Capacities for
Sunco Oil

| Arc | Capacity |
| :--- | :---: |
| $(s o, 1)$ | 2 |
| $(s o, 2)$ | 3 |
| $(1,2)$ | 3 |
| $(1,3)$ | 4 |
| $(3, s i)$ | 1 |
| $(2, s i)$ | 2 |

For a flow to be feasible, it must have two characteristics:

$$
\begin{equation*}
0 \leq \text { flow through each arc } \leq \text { arc capacity } \tag{1}
\end{equation*}
$$

and

$$
\begin{equation*}
\text { Flow into node } i=\text { flow out of node } i \tag{2}
\end{equation*}
$$

We assume that no oil gets lost while being pumped through the network, so at each node, a feasible flow must satify (2), the conservation-of-flow constraint. The introduction of the artificial arc $a_{0}$ allows us to write the conservation-of-flow constraint for the source and sink.

If we let $x_{0}$ be the flow through the artificial arc, then conservation of flow implies that $x_{0}=$ total amount of oil entering the sink. Thus, Sunco's goal is to maximize $x_{0}$ subject to (1) and (2):

$$
\begin{aligned}
\max z=x_{0} & & & \\
\text { s.t. } & & & \text { (Arc capacity constraints) } \\
x_{s o, 1} & \leq 2 & & \\
x_{s o, 2} & \leq 3 & & \\
x_{12} & \leq 3 & & \\
x_{2, s i} & \leq 2 & & \\
x_{13} & \leq 4 & & \\
x_{3, s i} & \leq 1 & & \\
x_{0} & =x_{s o, 1}+x_{s o, 2} & & \text { (Node so flow constraint) } \\
x_{s o, 1} & =x_{12}+x_{13} & & \text { (Node } 1 \text { flow constraint) } \\
x_{s o, 2}+x_{12} & =x_{2, s i} & & \text { (Node } 2 \text { flow constraint) } \\
x_{13} & & x_{3, s i} & \\
x_{3, s i}+x_{2, s i} & =x_{0} & & \text { (Node } 3 \text { flow constraint) } \\
x_{i j} & \geq 0 & & \text { (Node } s i \text { flow constraint) }
\end{aligned}
$$

One optimal solution to this LP is $z=3, x_{s o, 1}=2, x_{13}=1, x_{12}=1, x_{s o, 2}=1, x_{3, s i}=$ $1, x_{2, s i}=2, x_{0}=3$. Thus, the maximum possible flow of oil from node so to si is 3 million barrels per hour, with 1 million barrels each sent via the following paths: so $-1-2-s i$, so $-1-3-s i$, and so $-2-s i$.

The linear programming formulation of maximum-flow problems is a special case of the minimum-cost network flow problem (MCNFP) discussed in Section 8.5. A generalization of the transportation simplex (known as the network simplex) can be used to solve MCNFPs.

Before discussing the Ford-Fulkerson method for solving maximum-flow problems, we give two examples for situations in which a maximum-flow problem might arise.

EXAMPLE 4 Airline Maximum-Flow
Fly-by-Night Airlines must determine how many connecting flights daily can be arranged between Juneau, Alaska, and Dallas, Texas. Connecting flights must stop in Seattle and then stop in Los Angeles or Denver. Because of limited landing space, Fly-by-Night is limited to making the number of daily flights between pairs of cities shown in Table 9. Set up a maximum-flow problem whose solution will tell the airline how to maximize the number of connecting flights daily from Juneau to Dallas.
tABLE 9
Arc Capacities for Fly-by-Night Airlines

| Cities | Maximum Number <br> of Daily Flights |
| :--- | :---: |
| Juneau-Seattle $(J, S)$ | 3 |
| Seattle-L.A. $(S, L)$ | 2 |
| Seattle-Denver $(S, D e)$ | 3 |
| L.A.-Dallas $(L, D)$ | 1 |
| Denver-Dallas $(D e, D)$ | 2 |

FIGURE 7 Network for Fly-byNight Airlines


Solution
The appropriate network is given in Figure 7. Here the capacity of arc $(i, j)$ is the maximum number of daily flights between city $i$ and city $j$. The optimal solution to this maximum flow problem is $z=x_{0}=3, x_{J, S}=3, x_{S, L}=1, x_{S, D e}=2, x_{L, D}=1, x_{D e, D}=2$. Thus, Fly-by-Night can send three flights daily connecting Juneau and Dallas. One flight connects via Juneau-Seattle-L.A.-Dallas, and two flights connect via Juneau-Seattle-Denver-Dallas.

## EXAMPLE 5 Matchmaking

Five male and five female entertainers are at a dance. The goal of the matchmaker is to match each woman with a man in a way that maximizes the number of people who are matched with compatible mates. Table 10 describes the compatibility of the entertainers. Draw a network that makes it possible to represent the problem of maximizing the number of compatible pairings as a maximum-flow problem.

Solution Figure 8 is the appropriate network. In Figure 8, there is an arc with capacity 1 joining the source to each man, an arc with capacity 1 joining each pair of compatible mates, and an arc with capacity 1 joining each woman to the sink. The maximum flow in this network is the number of compatible couples that can be created by the matchmaker. For ex-

TABLE 10
Compatibilities for Matching

|  | Loni <br> Anderson | Meryl <br> Streep | Katharine <br> Hephurn | Linda <br> Evans | Victoria <br> Principal |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Kevin Costner | - | C | - | - | - |
| Burt Reynolds | C | - | - | - | - |
| Tom Selleck | C | C | - | - | - |
| Michael Jackson | C | C | - | - | C |
| Tom Cruise | - | - | C | C | C |

Note: C indicates compatibility.

FIGURE 8
Network for Matchmaker

ample, if the matchmaker pairs KC and MS, BR and LA, MJ and VP, and TC and KH, a flow of 4 from source to sink would be obtained. (This turns out to be a maximum flow for the network.)

To see why our network representation correctly models the matchmaker's problem, note that because the arc joining each woman to the sink has a capacity of 1 , conservation of flow ensures that each woman will be matched with at most one man. Similarly, because each arc from the source to a man has a capacity of 1 , each man can be paired with at most one woman. Because arcs do not exist between noncompatible mates, we can be sure that a flow of $k$ units from source to sink represents an assignment of men to women in which $k$ compatible couples are created.

## Solving Maximum-Flow Problems with LINGO

The maximum flow in a network can be found using LINDO, but LINGO greatly lessens the effort needed to communicate the necessary information to the computer. The following LINGO program (in the file Maxflow.lng) can be used to find the maximum flow from source to sink in Figure 6.

```
MODEL:
    1]SETS:
    2]NODES/1..5/;
    3] ARCS (NODES,NODES)/1,2 1,3 2,3 2,4 3,5 4,5 5,1/
    4] : CAP, FLOW;
    5] ENDSETS
    6]MAX=FLOW (5,1);
    7]@FOR(ARCS (I,J) : FLOW (I,J)<CAP (I,J));
    8]@FOR (NODES (I):@SUM (ARCS (J,I) :FLOW (J,I))
    9]=@SUM(ARCS (I,J) : FLOW(I,J)));
    10]DATA:
    11] CAP=2,3,3,4,2,1,1000;
    12] ENDDATA
END
```

If some nodes are identified by numbers, then LINGO will not allow you to identify other nodes with names involving letters. Thus, we have identified node 1 in line 2 with node so in Figure 6 and node 5 in line 2 with node si. Also nodes 1, 2, and 3 in Figure 6 correspond to nodes 2, 3, and 4, respectively, in line 2 of our LINGO program. Thus, line 2 defines the nodes of the flow network. In line 3, we define the arcs of the network by listing them (separated by spaces). For example, 1, 2 represents the arc from the source to node 1 in Figure 6 and 5,1 is the artificial arc. In line 4, we indicate that an arc capacity and a flow are associated with each arc. Line 5 ends the definition of the relevant sets.

In line 6, we indicate that our objective is to maximize the flow through the artificial arc (this equals the flow into the sink). Line 7 specifies the arc capacity constraints; for


### 8.5 Minimum-Cost Network Flow Problems

The transportation, assignment, transshipment, shortest-path, maximum flow, and CPM problems are all special cases of the minimum-cost network flow problem (MCNFP). Any MCNFP can be solved by a generalization of the transportation simplex called the network simplex.

To define an MCNFP, let
$x_{i j}=$ number of units of flow sent from node $i$ to node $j$ through arc $(i, j)$
$b_{i}=$ net supply (outflow - inflow) at node $i$
$c_{i j}=$ cost of transporting 1 unit of flow from node $i$ to node $j$ via arc $(i, j)$
$L_{i j}=$ lower bound on flow through arc $(i, j)$
(if there is no lower bound, let $L_{i j}=0$ )
$U_{i j}=$ upper bound on flow through arc $(i, j)$
(if there is no upper bound, let $U_{i j}=\infty$ )
Then the MCNFP may be written as

$$
\begin{array}{lll}
\min & \sum_{\text {all arcs }} c_{i j} x_{i j} & \\
\text { s.t. } & \sum_{j} x_{i j}-\sum_{k} x_{k i}=b_{i} & \text { (for each node } i \text { in the network) } \\
& L_{i j} \leq x_{i j} \leq U_{i j} & \text { (for each arc in the network) } \tag{9}
\end{array}
$$

Constraints (8) stipulate that the net flow out of node $i$ must equal $b_{i}$. Constraints (8) are referred to as the flow balance equations for the network. Constraints (9) ensure that the flow through each arc satisfies the arc capacity restrictions. In all our previous examples, we have set $L_{i j}=0$.

Let us show that transportation and maximum-flow problems are special cases of the minimum-cost network flow problem.

## Formulating a Transportation Problem as an MCNFP

Consider the transportation problem in Table 28. Nodes 1 and 2 are the two supply points, and nodes 3 and 4 are the two demand points. Then $b_{1}=4, b_{2}=5, b_{3}=-6$, and $b_{4}=$ -3 . The network corresponding to this transportation problem contains arcs $(1,3),(1,4)$, $(2,3)$, and $(2,4)$ (see Figure 45). The LP for this transportation problem may be written as shown in Table 29.

The first two constraints are the supply constraints, and the last two constraints are (after being multiplied by -1 ) the demand constraints. Because this transportation problem

TABLE 28


FIGURE 45
Representation of Transportation Problem as an MCNFP

table 29
MCNFP Representation of Transportation Problem

| $\min z=x_{13}+2 x_{14}+3 x_{23}+4 x_{24}$ |  |  |  |  |  |  |
| ---: | ---: | ---: | ---: | :--- | ---: | :--- |
| $x_{13}$ | $x_{14}$ | $x_{23}$ | $x_{24}$ |  | rhs | Constraint |
| 1 | 1 | 0 | 0 | $=$ | 4 | Node 1 |
| 0 | 0 | 1 | 1 | $=$ | 5 | Node 2 |
| -1 | 0 | -1 | 0 | $=$ | -6 | Node 3 |
| 1 | -1 | 0 | -1 | $=$ | -3 | Node 4 |
|  | All variables non-negative |  |  |  |  |  |

has no arc capacity restrictions, the flow balance equations are the only constraints. We note that if the problem had not been balanced, we could not have formulated the problem as an MCNFP. This is because if total supply exceeded total demand, we would not know with certainty the net outflow at each supply point. Thus, to formulate a transportation (or a transshipment) problem as an MCNFP, it may be necessary to add a dummy point.

## Formulating a Maximum-Flow Problem as an MCNFP

To see how a maximum-flow problem fits into the minimum-cost network flow context, consider the problem of finding the maximum flow from source to sink in the network of Figure 6. After creating an arc $a_{0}$ joining the sink to the source, we have $b_{s o}=b_{1}=b_{2}=$ $b_{3}=b_{s i}=0$. Then the LP constraints for finding the maximum flow in Figure 6 may be written as shown in Table 30.

The first five constraints are the flow balance equations for the nodes of the network, and the last six constraints are the arc capacity constraints. Because there is no upper limit on the flow through the artificial arc, there is no arc capacity constraint for $a_{0}$.

The flow balance equations in any MCNFP have the following important property: Each variable $x_{i j}$ has a coefficient of +1 in the node $i$ flow balance equation, a coefficient of -1 in the node $j$ flow balance equation, and a coefficient of 0 in all other flow balance equations. For example, in a transportation problem, the variable $x_{i j}$ will have a coeffi-
table 30
MCNFP Representation of Maximum-Flow Problem

| $\min z=x_{0}$ |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $\chi_{\text {so, }}$ | $\chi_{s s, 2}$ | $\chi_{13}$ | $\chi_{12}$ | $X_{3, s i}$ | $\chi_{2, s i}$ | $\chi_{0}$ |  | rhs | Constraint |
| 1 | 1 | 0 | 0 | 0 | 0 | -1 | $=$ | 0 | Node so |
| -1 | 0 | 1 | 1 | 0 | 0 | 0 | $=$ | 0 | Node 1 |
| 0 | -1 | 0 | -1 | 0 | 1 | 0 | $=$ | 0 | Node 2 |
| 0 | 0 | -1 | 0 | 1 | 0 | 0 | $=$ | 0 | Node 3 |
| 0 | 0 | 0 | 0 | -1 | -1 | 1 | = | 0 | Node si |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | $\leq$ | 2 | Arc (so, 1) |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | $\leq$ | 3 | Arc (so, 2) |
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | $\leq$ | 4 | $\operatorname{Arc}(1,3)$ |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | $\leq$ | 3 | Arc (1, 2) |
| 0 | 0 | 0 | 0 | 1 | 0 | 0 | $\leq$ | 1 | Arc (3, si) |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | $\leq$ | 2 | Arc (2, si) |
| All variables nonnegative |  |  |  |  |  |  |  |  |  |

cient of +1 in the flow balance equation for supply point $i$, a coefficient of -1 in the flow balance equation for demand point $j$, and a coefficient of 0 in all other flow balance equations. Even if the constraints of an LP do not appear to contain the flow balance equations of a network, clever transformation of an LP's constraints can often show that an LP is equivalent to an MCNFP (see Problem 6 at the end of this section).

An MCNFP can be solved by a generalization of the transportation simplex known as the network simplex algorithm (see Section 8.7). As with the transportation simplex, the pivots in the network simplex involve only additions and subtractions. This fact can be used to prove that if all the $b_{i}$ 's and arc capacities are integers, then in the optimal solution to an MCNFP, all the variables will be integers. Computer codes that use the network simplex can quickly solve even extremely large network problems. For example, MCNFPs with 5,000 nodes and 600,000 arcs have been solved in under 10 minutes. To use a network simplex computer code, the user need only input a list of the network's nodes and arcs, the $c_{i j}$ 's and arc capacity for each arc, and the $b_{i}$ 's for each node. The network simplex is efficient and easy to use, so it is extremely important to formulate an LP, if at all possible, as an MCNFP.

To close this section, we formulate a simple traffic assignment problem as an MCNFP.

## EXAMPLE 7 Traffic MCNFP

Each hour, an average of 900 cars enter the network in Figure 46 at node 1 and seek to travel to node 6. The time it takes a car to traverse each arc is shown in Table 31. In Figure 46 , the number above each arc is the maximum number of cars that can pass by any point on the arc during a one-hour period. Formulate an MCNFP that minimizes the total time required for all cars to travel from node 1 to node 6 .

## Solution Let

$$
x_{i j}=\text { number of cars per hour that traverse the arc from node } i \text { to node } j
$$

Then we want to minimize

$$
z=10 x_{12}+50 x_{13}+70 x_{25}+30 x_{24}+30 x_{56}+30 x_{45}+60 x_{46}+60 x_{35}+10 x_{34}
$$

We are given that $b_{1}=900, b_{2}=b_{3}=b_{4}=b_{5}=0$, and $b_{6}=-900$ (we will not introduce the artificial arc connecting node 6 to node 1). The constraints for this MCNFP are shown in Table 32.

FIGURE 46 Representation of Traffic Example as MCNFP


TABLE 31
Travel Times for Traffic
Example

| Arc | Time <br> (Minutes) |
| :---: | :---: |


| $(1,2)$ | 10 |
| :--- | :--- |
| $(1,3)$ | 50 |
| $(2,5)$ | 70 |
| $(2,4)$ | 30 |
| $(5,6)$ | 30 |
| $(4,5)$ | 30 |
| $(4,6)$ | 60 |
| $(3,5)$ | 60 |
| $(3,4)$ | 10 |


| MCNFP Representation of Traffic Example | $\chi_{12}$ | $\chi_{13}$ | $\chi_{24}$ | $\chi_{25}$ | $\chi_{34}$ | $\chi_{35}$ | $\chi_{45}$ | $\chi_{46}$ | $\chi_{56}$ |  | rhs | Constraint |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $=$ | 900 | Node 1 |
|  | -1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | = | 0 | Node 2 |
|  | 0 | -1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | $=$ | 0 | Node 3 |
|  | 0 | 0 | -1 | 0 | -1 | 0 | 1 | 1 | 0 | = | 0 | Node 4 |
|  | 0 | 0 | 0 | -1 | 0 | -1 | -1 | 0 | 1 | = | 0 | Node 5 |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | -1 | -1 | = | -900 | Node 6 |
|  | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\leq$ | 800 | Arc (1, 2) |
|  | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\leq$ | 600 | Arc (1, 3) |
|  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | $\leq$ | 600 | Arc (2, 4) |
|  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | $\leq$ | 100 | Arc (2, 5) |
|  | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | $\leq$ | 300 | Arc (3, 4) |
|  | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | $\leq$ | 400 | Arc (3, 5) |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | $\leq$ | 600 | Arc (4, 5) |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | $\leq$ | 400 | Arc (4, 6) |
|  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | $\leq$ | 600 | Arc (5, 6) |
| All variables non-negative |  |  |  |  |  |  |  |  |  |  |  |  |

## Solving an MCNFP with LINGO

The following LINGO program (file Traffic.lng) can be used to find the optimal solution to Example 7 (or any MCNFP).

```
MODEL:
    1] SETS:
NODES/1..6/:SUPP;
ARCS (NODES,NODES)/1,2 1,3 2,4 2,5 3,4 3,5 4,5 4,6
:CAP,FLOW,COST;
ENDSETS
MIN=@SUM(ARCS:COST*FLOW);
@FOR(ARCS(I,J):FLOW(I,J)<CAP(I,J));
@FOR(NODES(I):-@SUM(ARCS)(J,I):FLOW(J,I))
+@SUM(ARCS(I,J):FLOW(I,J))=SUPP(I));
10] DATA:
11] COST= 10,50,30,70,10,60,30,60,30;
12] SUPP =900,0,0,0,0,-900;
13] CAP}=800,600,600,100,300,400,600,400,600
    14] ENDDATA
END
```

In line 2, we define the network's nodes and associate a net supply (flow out-flow in) with each node. The supplies data are entered in line 12. In line 3, we define, by listing, the arcs in the network and in line 4 associate a capacity (CAP), a flow (FLOW), and a cost-per-unit-shipped (COST) with each arc. The unit shipping costs data are entered in line 11. Line 6 generates the objective function by summing over all arcs (unit cost for arc)*(flow through arc). Line 7 generates each arc's capacity constraint (arc capacities data are entered in line 13). For each node, lines $8-9$ generate the conservation-of-flow constraint. They imply that for each node I, $-($ flow into node I) + (flow out of node I) $=($ supply of node I). When solved on LINGO, we find that the solution to Example 7 is $z=95,000$ minutes, $x_{12}=700$, $x_{13}=200, x_{24}=600, x_{25}=100, x_{34}=200, x_{45}=400, x_{46}=400, x_{56}=500$.

Our LINGO program can be used to solve any MCNFP. Just input the set of nodes, supplies, arcs, and unit shipping cost; hit GO and you are done!

## PROBLEMS

Note: To formulate a problem as an MCNFP, you should draw the appropriate network and determine the $c_{i j}$ 's, the $b_{i}$ 's, and the arc capacities.

## Group A

1 Formulate the problem of finding the shortest path from node 1 to node 6 in Figure 2 as an MCNFP. (Hint: Think of finding the shortest path as the problem of minimizing the total cost of sending 1 unit of flow from node 1 to node 6.)
2 a Find the dual of the LP that was used to find the length of the critical path for Example 6 of Section 8.4. b Show that the answer in part (a) is an MCNFP.
c Explain why the optimal objective function value for the LP found in part (a) is the longest path in the project network from node 1 to node 6 . Why does this justify our earlier claim that the critical path in a project network is the longest path from the start node to the finish node?
3 Fordco produces cars in Detroit and Dallas. The Detroit plant can produce as many as 6,500 cars, and the Dallas plant can produce as many as 6,000 cars. Producing a car costs $\$ 2,000$ in Detroit and $\$ 1,800$ in Dallas. Cars must be shipped to three cities. City 1 must receive 5,000 cars, city 2 must receive 4,000 cars, and city 3 must receive 3,000
cars. The cost of shipping a car from each plant to each city is given in Table 33. At most, 2,200 cars may be sent from a given plant to a given city. Formulate an MCNFP that can be used to minimize the cost of meeting demand.
4 Each year, Data Corporal produces as many as 400 computers in Boston and 300 computers in Raleigh. Los Angeles customers must receive 400 computers, and 300 computers must be supplied to Austin customers. Producing a computer costs $\$ 800$ in Boston and $\$ 900$ in Raleigh. Computers are transported by plane and may be sent through Chicago. The costs of sending a computer between pairs of cities are shown in Table 34.
a Formulate an MCNFP that can be used to minimize the total (production + distribution) cost of meeting Data Corporal's annual demand.

TABLE 33

|  | To (\$) |  |  |
| :--- | :---: | :---: | :---: |
| From | City $\mathbf{1}$ | City 2 | City 3 3 |
| Detroit | 800 | 600 | 300 |
| Dallas | 500 | 200 | 200 |

TABLE 34

|  | To (\$) |  |  |
| :--- | :---: | :---: | :---: |
| From | Chicago | Austin | Los Angeles |
| Boston | 80 | 220 | 280 |
| Raleigh | 100 | 140 | 170 |
| Chicago | - | 40 | 50 |

b How would you modify the part (a) formulation if at most 200 units could be shipped through Chicago? [Hint: Add an additional node and arc to this part (a) network.]
5 Oilco has oil fields in San Diego and Los Angeles. The San Diego field can produce 500,000 barrels per day, and the Los Angeles field can produce 400,000 barrels per day. Oil is sent from the fields to a refinery, in either Dallas or Houston (assume each refinery has unlimited capacity). To refine 100,000 barrels costs $\$ 700$ at Dallas and $\$ 900$ at Houston. Refined oil is shipped to customers in Chicago and New York. Chicago customers require 400,000 barrels per day, and New York customers require 300,000 barrels per day. The costs of shipping 100,000 barrels of oil (refined or unrefined) between cities are shown in Table 35.
a Formulate an MCNFP that can be used to determine how to minimize the total cost of meeting all demands.
b If each refinery had a capacity of 500,000 barrels per day, how would the part (a) answer be modified?

## Group B

6 Workco must have the following number of workers available during the next three months: month 1,20 ; month 2,16 ; month 3,25 . At the beginning of month 1 , Workco has no workers. It costs Workco $\$ 100$ to hire a worker and $\$ 50$ to fire a worker. Each worker is paid a salary of $\$ 140 /$ month. We will show that the problem of determining a hiring and firing strategy that minimizes the total cost incurred during the next three (or in general, the next $n$ ) months can be formulated as an MCNFP.

$$
\begin{aligned}
& \text { a Let } \\
& x_{i j}=\begin{array}{l}
\text { number of workers hired at beginning of month } i \\
\text { and fired after working till end of month } j-1
\end{array}
\end{aligned}
$$

(if $j=4$, the worker is never fired). Explain why the following LP will yield a minimum-cost hiring and firing strategy:

TABLE 35

|  | To (\$) |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| From | Dallas | Houston | New York | Chicago |
| Los Angeles | 300 | 110 | - | - |
| San Diego | 420 | 100 | - | - |
| Dallas | - | - | 450 | 550 |
| Houston | - | - | 470 | 530 |

$$
\begin{aligned}
& \min z=50\left(x_{12}+x_{13}+x_{23}\right) \\
& +100\left(x_{12}+x_{13}+x_{14}+x_{23}+x_{24}+x_{34}\right) \\
& +140\left(x_{12}+x_{23}+x_{34}\right) \\
& +280\left(x_{13}+x_{24}\right)+420 x_{14} \\
& \text { s.t. (1) } x_{12}+x_{13}+x_{14} \quad-e_{1}=20 \\
& \text { (Month } 1 \text { constraint) } \\
& \text { (2) } x_{13}+x_{14}+x_{23}+x_{24}-e_{2}=16 \\
& \text { (Month } 2 \text { constraint) } \\
& \text { (3) } x_{14}+x_{24}+x_{34} \quad-e_{3}=25 \\
& \text { (Month } 3 \text { constraint) } \\
& x_{i j} \geq 0
\end{aligned}
$$

b To obtain an MCNFP, replace the constraints in part (a) by
i Constraint (1);
ii Constraint (2) - Constraint (1);
iii Constraint (3) - Constraint (2);
iv - (Constraint (3)).
Explain why an LP with Constraints (i)-(iv) is an MCNFP.
c Draw the network corresponding to the MCNFP obtained in answering part (b).
$7^{\dagger}$ Braneast Airlines must determine how many airplanes should serve the Boston-New York-Washington air corridor and which flights to fly. Braneast may fly any of the daily flights shown in Table 36. The fixed cost of operating an airplane is $\$ 800 /$ day. Formulate an MCNFP that can be used to maximize Braneast's daily profits. (Hint: Each node in the network represents a city and a time. In addition to arcs representing flights, we must allow for the possibility that an airplane will stay put for an hour or more. We must ensure that the model includes the fixed cost of operating a plane. To include this cost, the following three arcs might be included in the network: from Boston 7 p.m. to Boston 9 A.m.; from New York 7 p.m. to New York 9 A.m.; and from Washington 7 p.m. to Washington 9 A.m.)
8 Daisymay Van Line moves people between New York, Philadelphia, and Washington, D.C. It takes a van one day to travel between any two of these cities. The company incurs costs of $\$ 1,000$ per day for a van that is fully loaded and traveling, $\$ 800$ per day for an empty van that travels, $\$ 700$ per day for a fully loaded van that stays in a city, and $\$ 400$ per day for an empty van that remains in a city. Each day of the week, the loads described in Table 37 must be shipped. On Monday, for example, two trucks must be sent from Philadelphia to New York (arriving on Tuesday). Also, two trucks must be sent from Philadelphia to Washington on Friday (assume that Friday shipments must arrive on Monday). Formulate an MCNFP that can be used to minimize the cost of meeting weekly requirements. To simplify the formulation, assume that the requirements repeat each week. Then it seems plausible to assume that any of the company's trucks will begin each week in the same city in which it began the previous week.
${ }^{\dagger}$ This problem is based on Glover et al. (1982).

TABLE 36

| Leaves |  |  | Arrives |  |  | Flight <br> Revenue |
| :--- | ---: | :--- | ---: | ---: | ---: | :---: | | Variable Cost |
| :---: |
| of Flight (S) |

TABLE 37

| Trip | Monday | Tuesday | Wednesday | Thursday | Friday |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Phil.-N.Y. | 2 | - | - | - | - |
| Phil.-Wash. | - | 2 | - | - | 2 |
| N.Y.-Phil. | 3 | 2 | - | - | - |
| N.Y.-Wash. | - | - | 2 | 2 | - |
| N.Y.-Phil. | 1 | - | - | - | - |
| Wash.-N.Y. | - | - | 1 | - | 1 |

### 8.6 Minimum Spanning Tree Problems

Suppose that each arc $(i, j)$ in a network has a length associated with it and that arc $(i, j)$ represents a way of connecting node $i$ to node $j$. For example, if each node in a network represents a computer at State University, then arc $(i, j)$ might represent an underground cable that connects computer $i$ with computer $j$. In many applications, we want to determine the set of arcs in a network that connect all nodes such that the sum of the length of the arcs is minimized. Clearly, such a group of arcs should contain no loop. (A loop is often called a closed path or cycle.) For example, in Figure 47, the sequence of arcs $(1,2)-(2,3)-(3,1)$ is a loop.

DEFINITION ■ For a network with $n$ nodes, a spanning tree is a group of $n-1$ arcs that connects all nodes of the network and contains no loops.

FIGURE 47 Illustration of Loop and Minimum Spanning Tree


In Figure 47, there are three spanning trees:
1 Arcs $(1,2)$ and $(2,3)$
2 Arcs $(1,2)$ and $(1,3)$
3 Arcs $(1,3)$ and $(2,3)$
A spanning tree of minimum length in a network is a minimum spanning tree (MST). In Figure 47, the spanning tree consisting of $\operatorname{arcs}(1,3)$ and $(2,3)$ is the unique minimum spanning tree.

The following method (MST algorithm) may be used to find a minimum spanning tree.
Step 1 Begin at any node $i$, and join node $i$ to the node in the network (call it node $j$ ) that is closest to node $i$. The two nodes $i$ and $j$ now form a connected set of nodes $C=$ $\{i, j\}$, and arc $(i, j)$ will be in the minimum spanning tree. The remaining nodes in the network (call them $C^{\prime}$ ) are referred to as the unconnected set of nodes.

Step 2 Now choose a member of $C^{\prime}$ (call it $n$ ) that is closest to some node in $C$. Let $m$ represent the node in $C$ that is closest to $n$. Then the $\operatorname{arc}(m, n)$ will be in the minimum spanning tree. Now update $C$ and $C^{\prime}$. Because $n$ is now connected to $\{i, j\}, C$ now equals $\{i, j, n\}$ and we must eliminate node $n$ from $C^{\prime}$.

Step 3 Repeat this process until a minimum spanning tree is found. Ties for closest node and arc to be included in the minimum spanning tree may be broken arbitrarily.

At each step the algorithm chooses the shortest arc that can be used to expand $C$, so the algorithm is often referred to as a "greedy" algorithm. It is remarkable that the act of being "greedy" at each step of the algorithm can never force us later to follow a "bad arc." In Example 1 of Chapter 9 we will see that for some types of problems, a greedy algorithm may not yield an optimal solution! A justification of the MST algorithm is given in Problem 3 at the end of this section. Example 8 illustrates the algorithm.

## EXAMPLE 8 MST Algorithm

The State University campus has five minicomputers. The distance between each pair of computers (in city blocks) is given in Figure 48. The computers must be interconnected by underground cable. What is the minimum length of cable required? Note that if no arc is drawn connecting a pair of nodes, this means that (because of underground rock formations) no cable can be laid between these two computers.

Solution We want to find the minimum spanning tree for Figure 48.
Iteration 1 Following the MST algorithm, we arbitrarily choose to begin at node 1 . The closest node to node 1 is node 2 . Now $C=\{1,2\}, C^{\prime}=\{3,4,5\}$, and arc $(1,2)$ will be in the minimum spanning tree (see Figure 49a).

Iteration 2 Node 5 is closest (two blocks distant) to $C$. Because node 5 is two blocks from node 1 and from node 2 , we may include either $\operatorname{arc}(2,5)$ or $\operatorname{arc}(1,5)$ in the minimum spanning tree. We arbitrarily choose to include $\operatorname{arc}(2,5)$. Then $C=\{1,2,5\}$ and $C^{\prime}=$ $\{3,4\}$ (see Figure 49b).

Iteration 3 Node 3 is two blocks from node 5, so we may include arc $(5,3)$ in the minimum spanning tree. Now $C=\{1,2,3,5\}$ and $C^{\prime}=4$ (see Figure 49c).

Iteration 4 Node 5 is the closest node to node 4 , so we add arc $(5,4)$ to the minimum spanning tree (see Figure 49d).
We have now obtained the minimum spanning tree consisting of $\operatorname{arcs}(1,2),(2,5),(5,3)$, and $(5,4)$. The length of the minimum spanning tree is $1+2+2+4=9$ blocks.

FIGURE 48 Distances between State University Computers


FIGURE 49 MST Algorithm for Computer Example

a Iteration 1
$C=[1,2]$
$C^{\prime}=[3,4,5]$


C Iteration 3
$\begin{aligned} C & =[1,2,3,5] \\ C^{\prime} & =[4]\end{aligned}$
$C^{\prime}=[4]$

b Iteration 2

$\operatorname{Arcs}(1,2),(2,5),(5,3)$, and $(5,4)$ are the MST
$C=[1,2,5]$
$C^{\prime}=[3,4]$

## PROBLEMS

## Group A

1 The distances (in miles) between the Indiana cities of Gary, Fort Wayne, Evansville, Terre Haute, and South Bend are shown in Table 38. It is necessary to build a state road system that connects all these cities. Assume that for political reasons no road can be built connecting Gary and Fort Wayne, and no road can be built connecting South Bend and Evansville. What is the minimum length of road required?
2 The city of Smalltown consists of five subdivisions. Mayor John Lion wants to build telephone lines to ensure that all the subdivisions can communicate with each other. The distances between the subdivisions are given in Figure 50 . What is the minimum length of telephone line required? Assume that no telephone line can be built between subdivisions 1 and 4.

## Group B

3 In this problem, we explain why the MST algorithm works. Define
$S=$ minimum spanning tree
$C_{t}=$ nodes connected after iteration $t$ of MST algorithm has been completed
$C_{t}^{\prime}=$ nodes not connected after iteration $t$ of MST algorithm has been completed
$A_{t}=$ set of arcs in minimum spanning tree after $t$ iterations of MST algorithm have been completed

TABLE 38

|  | Gary | Fort <br> Wayne | Evansville | Terre <br> Haute | South <br> Bend |
| :--- | ---: | :---: | :---: | :---: | ---: |
| Gary | - | 132 | 217 | 164 | 58 |
| Fort Wayne | 132 | - | 290 | 201 | 79 |
| Evansville | 217 | 290 | - | 113 | 303 |
| Terre Haute | 164 | 201 | 113 | - | 196 |
| South Bend | 58 | 79 | 303 | 196 | - |

FIGURE 50 Network for Problem 2


Suppose the MST algorithm does not yield a minimum spanning tree. Then, for some $t$, it must be the case that all arcs in $A_{t-1}$ are in $S$, but the arc chosen at iteration $t$ (call it $a_{t}$ ) of the MST algorithm is not in $S$. Then $S$ must contain some arc $a_{t}^{\prime}$ that leads from a node in $C_{t-1}$ to a node in $C_{t-1}^{\prime}$ ). Show that by replacing arc $a_{t}^{\prime}$ with arc $a_{t}$, we can obtain a shorter spanning tree than $S$. This contradiction proves that all arcs chosen by the MST algorithm must be in $S$. Thus, the MST algorithm does indeed find a minimum spanning tree.
4 a Three cities are at the vertices of an equilateral triangle of unit length. Flying Lion Airlines needs to supply connecting service between these three cities. What is the minimum length of the two routes needed to supply the connecting service?
b Now suppose Flying Lion Airlines adds a hub at the "center" of the equilateral triangle. Show that the length of the routes needed to connect the three cities has decreased by $13 \%$. (Note: It has been shown that no matter how many "hubs" you add and no matter how many points must be connected, you can never save more than $13 \%$ of the total distance needed to "span" all the original points by adding hubs.) ${ }^{\dagger}$

### 8.7 The Network Simplex Method ${ }^{\ddagger}$

In this section, we describe how the simplex algorithm simplifies for MCNFPs. To simplify our presentation, we assume that for each arc, $L_{i j}=0$. Then the information needed to describe an MCNFP of the form (8)-(9) may be summarized graphically as in Figure 51. We will denote the $c_{i j}$ for each arc by the symbol $\$$, and the other number on each arc will represent the arc's upper bound $\left(U_{i j}\right)$. The $b_{i}$ for any node with nonzero outflow will be listed in parentheses. Thus, Figure 51 represents an MCNFP with $c_{12}=5, c_{25}=2, c_{13}=4, c_{35}=8$,

[^12]

Recall that we defined integer programming problems in our discussion of the Divisibility Assumption in Section 3.1. Simply stated, an integer programming problem (IP) is an LP in which some or all of the variables are required to be non-negative integers. ${ }^{\dagger}$

In this chapter (as for LPs in Chapter 3), we find that many real-life situations may be formulated as IPs. Unfortunately, we will also see that IPs are usually much harder to solve than LPs.

In Section 9.1, we begin with necessary definitions and some introductory comments about IPs. In Section 9.2, we explain how to formulate integer programming models. We also discuss how to solve IPs on the computer with LINDO, LINGO, and Excel Solver. In Sections 9.3-9.8, we discuss other methods used to solve IPs.

### 9.1 Introduction to Integer Programming

An IP in which all variables are required to be integers is called a pure integer programming problem. For example,

$$
\begin{gather*}
\max z=3 x_{1}+2 x_{2} \\
\text { s.t. } \quad x_{1}+x_{2} \leq 6  \tag{1}\\
x_{1}, x_{2} \geq 0, x_{1}, x_{2} \text { integer }
\end{gather*}
$$

is a pure integer programming problem.
An IP in which only some of the variables are required to be integers is called a mixed integer programming problem. For example,

$$
\begin{aligned}
& \max z=3 x_{1}+2 x_{2} \\
& \text { s.t. } \quad x_{1}+x_{2} \leq 6 \\
& x_{1}, x_{2} \geq 0, x_{1} \text { integer }
\end{aligned}
$$

is a mixed integer programming problem ( $x_{2}$ is not required to be an integer).
An integer programming problem in which all the variables must equal 0 or 1 is called a $0-1$ IP. In Section 9.2 , we see that $0-1$ IPs occur in surprisingly many situations. ${ }^{\ddagger}$ The following is an example of a $0-1$ IP:

$$
\begin{align*}
& \max z=x_{1}-x_{2} \\
& \text { s.t. } \quad x_{1}+2 x_{2} \leq 2 \\
& 2 x_{1}-x_{2} \leq 1  \tag{2}\\
& \quad x_{1}, x_{2}=0 \text { or } 1
\end{align*}
$$

Solution procedures especially designed for $0-1$ IPs are discussed in Section 9.7.
${ }^{\dagger}$ A nonlinear integer programming problem is an optimization problem in which either the objective function or the left-hand side of some of the constraints are nonlinear functions and some or all of the variables must be integers. Such problems may be solved with LINGO or Excel Solver.
${ }^{\ddagger}$ Actually, any pure IP can be reformulated as an equivalent 0-1 IP (Section 9.7).

The concept of LP relaxation of an integer programming problem plays a key role in the solution of IPs.

DEFINITION ■ The LP obtained by omitting all integer or $0-1$ constraints on variables is called the LP relaxation of the IP.

For example, the LP relaxation of (1) is

$$
\begin{gather*}
\max z=3 x_{1}+2 x_{2} \\
\text { s.t. } \quad x_{1}+x_{2} \leq 6  \tag{1'}\\
\\
\quad x_{1}, x_{2} \geq 0
\end{gather*}
$$

and the LP relaxation of (2) is

$$
\begin{align*}
& \max z=x_{1}-x_{2} \\
& \text { s.t. } \quad x_{1}+2 x_{2} \leq 2 \\
& 2 x_{1}-x_{2} \leq 1  \tag{2'}\\
& \\
& \\
& x_{1}, x_{2} \geq 0
\end{align*}
$$

Any IP may be viewed as the LP relaxation plus additional constraints (the constraints that state which variables must be integers or be 0 or 1). Hence, the LP relaxation is a less constrained, or more relaxed, version of the IP. This means that the feasible region for any IP must be contained in the feasible region for the corresponding LP relaxation. For any IP that is a max problem, this implies that

$$
\begin{equation*}
\text { Optimal } z \text {-value for LP relaxation } \geq \text { optimal } z \text {-value for IP } \tag{3}
\end{equation*}
$$

This result plays a key role when we discuss the solution of IPs.
To shed more light on the properties of integer programming problems, we consider the following simple IP:

$$
\begin{align*}
& \max z=21 x_{1}+11 x_{2} \\
& \text { s.t. } \quad 7 x_{1}+4 x_{2} \leq 13  \tag{4}\\
& x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{align*}
$$

From Figure 1, we see that the feasible region for this problem consists of the following set of points: $S=\{(0,0),(0,1),(0,2),(0,3),(1,0),(1,1)\}$. Unlike the feasible region for any LP, the one for (4) is not a convex set. By simply computing and comparing the $z$-values for each of the six points in the feasible region, we find the optimal solution to (4) is $z=33, x_{1}=0, x_{2}=3$.

If the feasible region for a pure IP's LP relaxation is bounded, as in (4), then the feasible region for the IP will consist of a finite number of points. In theory, such an IP could be solved (as described in the previous paragraph) by enumerating the $z$-values for each feasible point and determining the feasible point having the largest $z$-value. The problem with this approach is that most actual IPs have feasible regions consisting of billions of feasible points. In such cases, a complete enumeration of all feasible points would require a large amount of computer time. As we explain in Section 9.3, IPs often are solved by cleverly enumerating all the points in the IP's feasible region.

Further study of (4) sheds light on other interesting properties of IPs. Suppose that a naive analyst suggests the following approach for solving an IP: First solve the LP relaxation; then round off (to the nearest integer) each variable that is required to be an integer and that assumes a fractional value in the optimal solution to the LP relaxation.

Applying this approach to (4), we first find the optimal solution to the LP relaxation: $x_{1}=\frac{13}{7}, x_{2}=0$. Rounding this solution yields the solution $x_{1}=2, x_{2}=0$ as a possible

FIGURE 1
Feasible Region for Simple IP (4)

optimal solution to (4). But $x_{1}=2, x_{2}=0$ is infeasible for (4), so it cannot possibly be the optimal solution to (4). Even if we round $x_{1}$ downward (yielding the candidate solution $x_{1}=1, x_{2}=0$ ), we do not obtain the optimal solution $\left(x_{1}=0, x_{2}=3\right.$ is the optimal solution).

For some IPs, it can even turn out that every roundoff of the optimal solution to the LP relaxation is infeasible. To see this, consider the following IP:

$$
\begin{array}{ll}
\max z=4 x_{1}+x_{2} \\
\text { s.t. } \quad 2 x_{1}+x_{2} \leq 5 \\
& 2 x_{1}+3 x_{2}=5 \\
x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{array}
$$

The optimal solution to the LP relaxation for this IP is $z=10, x_{1}=\frac{5}{2}, x_{2}=0$. Rounding off this solution, we obtain either the candidate $x_{1}=2, x_{2}=0$ or the candidate $x_{1}=$ $3, x_{2}=0$. Neither candidate is a feasible solution to the IP.

Recall from Chapter 4 that the simplex algorithm allowed us to solve LPs by going from one basic feasible solution to a better one. Also recall that in most cases, the simplex algorithm examines only a small fraction of all basic feasible solutions before the optimal solution is obtained. This property of the simplex algorithm enables us to solve relatively large LPs by expending a surprisingly small amount of computational effort. Analogously, one would hope that an IP could be solved via an algorithm that proceeded from one feasible integer solution to a better feasible integer solution. Unfortunately, no such algorithm is known.

In summary, even though the feasible region for an IP is a subset of the feasible region for the IP's LP relaxation, the IP is usually much more difficult to solve than the IP's LP relaxation.

### 9.2 Formulating Integer Programming Problems

In this section, we show how practical solutions can be formulated as IPs. After completing this section, the reader should have a good grasp of the art of developing integer programming formulations. We begin with some simple problems and gradually build to more complicated formulations. Our first example is a capital budgeting problem reminiscent of the Star Oil problem of Section 3.6.

Stockco is considering four investments. Investment 1 will yield a net present value (NPV) of $\$ 16,000$; investment 2 , an NPV of $\$ 22,000$; investment 3, an NPV of $\$ 12,000$; and investment 4 , an NPV of $\$ 8,000$. Each investment requires a certain cash outflow at the present time: investment 1, \$5,000; investment 2, \$7,000; investment 3, \$4,000; and investment 4, $\$ 3,000$. Currently, $\$ 14,000$ is available for investment. Formulate an IP whose solution will tell Stockco how to maximize the NPV obtained from investments 1-4.
Solution As in LP formulations, we begin by defining a variable for each decision that Stockco must make. This leads us to define a $0-1$ variable:

$$
x_{j}(j=1,2,3,4)= \begin{cases}1 & \text { if investment } j \text { is made } \\ 0 & \text { otherwise }\end{cases}
$$

For example, $x_{2}=1$ if investment 2 is made, and $x_{2}=0$ if investment 2 is not made.
The NPV obtained by Stockco (in thousands of dollars) is

$$
\begin{equation*}
\text { Total NPV obtained by Stockco }=16 x_{1}+22 x_{2}+12 x_{3}+8 x_{4} \tag{5}
\end{equation*}
$$

To see this, note that if $x_{j}=1$, then (5) includes the NPV of investment $j$, and if $x_{j}=0$, (5) does not include the NPV of investment $j$. This means that whatever combination of investments is undertaken, (5) gives the NPV of that combination of projects. For example, if Stockco invests in investments 1 and 4 , then an NPV of $16,000+8,000=\$ 24,000$ is obtained. This combination of investments corresponds to $x_{1}=x_{4}=1, x_{2}=x_{3}=0$, so (5) indicates that the NPV for this investment combination is $16(1)+22(0)+$ $12(0)+8(1)=\$ 24$ (thousand). This reasoning implies that Stockco's objective function is

$$
\begin{equation*}
\max z=16 x_{1}+22 x_{2}+12 x_{3}+8 x_{4} \tag{6}
\end{equation*}
$$

Stockco faces the constraint that at most $\$ 14,000$ can be invested. By the same reasoning used to develop (5), we can show that

$$
\begin{equation*}
\text { Total amount invested (in thousands of dollars) }=5 x_{1}+7 x_{2}+4 x_{3}+3 x_{4} \tag{7}
\end{equation*}
$$

For example, if $x_{1}=0, x_{2}=x_{3}=x_{4}=1$, then Stockco makes investments 2,3 , and 4. In this case, Stockco must invest $7+4+3=\$ 14$ (thousand). Equation (7) yields a total amount invested of $5(0)+7(1)+4(1)+3(1)=\$ 14$ (thousand). Because at most $\$ 14,000$ can be invested, $x_{1}, x_{2}, x_{3}$, and $x_{4}$ must satisfy

$$
\begin{equation*}
5 x_{1}+7 x_{2}+4 x_{3}+3 x_{4} \leq 14 \tag{8}
\end{equation*}
$$

Combining (6) and (8) with the constraints $x_{j}=0$ or $1(j=1,2,3,4)$ yields the following 0-1 IP:

$$
\begin{aligned}
& \max z=16 x_{1}+22 x_{2}+12 x_{3}+8 x_{4} \\
& \text { s.t. } \quad 5 x_{1}+7 x_{2}+4 x_{3}+3 x_{4} \leq 14 \\
& \quad x_{j}=0 \text { or } 1 \quad(j=1,2,3,4)
\end{aligned}
$$

REMARKS 1 In Section 9.5, we show that the optimal solution to (9) is $x_{1}=0, x_{2}=x_{3}=x_{4}=1, z=$ $\$ 42,000$. Hence, Stockco should make investments 2 , 3, and 4, but not 1 . Investment 1 yields a higher NPV per dollar invested than any of the others (investment 1 yields $\$ 3.20$ per dollar invested, investment $2, \$ 3.14$; investment $3, \$ 3$; and investment $4, \$ 2.67$ ), so it may seem surprising that investment 1 is not undertaken. To see why the optimal solution to (9) does not involve making the "best" investment, note that any investment combination that includes investment 1 cannot use more than $\$ 12,000$. This means that using investment 1 forces Stockco to forgo investing $\$ 2,000$. On the other hand, the optimal investment combination uses all $\$ 14,000$ of the investment budget. This en-
table 1
Weights and Benefits for
Items in Josie's Knapsack

| Item | Weight <br> (Pounds) | Benefit |
| :--- | :---: | :---: |
| 1 | 5 | 16 |
| 2 | 7 | 22 |
| 3 | 4 | 12 |
| 4 | 3 | 8 |

ables the optimal combination to obtain a higher NPV than any combination that includes investment 1. If, as in Chapter 3, fractional investments were allowed, the optimal solution to (9) would be $x_{1}=x_{2}=1, x_{3}=0.50, x_{4}=0, z=\$ 44,000$, and investment 1 would be used. This simple example shows that the choice of modeling a capital budgeting problem as a linear programming or as an integer programming problem can significantly affect the optimal solution to the problem.
2 Any IP, such as (9), that has only one constraint is referred to as a knapsack problem. Suppose that Josie Camper is going on an overnight hike. There are four items Josie is considering taking along on the trip. The weight of each item and the benefit Josie feels she would obtain from each item are listed in Table 1.

Suppose Josie's knapsack can hold up to 14 lb of items. For $j=1,2,3,4$, define

$$
x_{j}=\left\{\begin{array}{l}
1 \text { if Josie takes item } j \text { on the hike } \\
0 \text { otherwise }
\end{array}\right.
$$

Then Josie can maximize the total benefit by solving (9).
In the following example, we show how the Stockco formulation can be modified to handle additional constraints.

## EXAMPLE 2 Capital Budgeting (Continued)

Modify the Stockco formulation to account for each of the following requirements:
1 Stockco can invest in at most two investments.
2 If Stockco invests in investment 2, they must also invest in investment 1.
3 If Stockco invests in investment 2, they cannot invest in investment 4.
Solution 1 Simply add the constraint

$$
\begin{equation*}
x_{1}+x_{2}+x_{3}+x_{4} \leq 2 \tag{10}
\end{equation*}
$$

to (9). Because any choice of three or four investments will have $x_{1}+x_{2}+x_{3}+x_{4} \geq$ 3 , (10) excludes from consideration all investment combinations involving three or more investments. Thus, (10) eliminates from consideration exactly those combinations of investments that do not satisfy the first requirement.

2 In terms of $x_{1}$ and $x_{2}$, this requirement states that if $x_{2}=1$, then $x_{1}$ must also equal 1. If we add the constraint

$$
\begin{equation*}
x_{2} \leq x_{1} \quad \text { or } \quad x_{2}-x_{1} \leq 0 \tag{11}
\end{equation*}
$$

to (9), then we will have taken care of the second requirement. To show that (11) is equivalent to requirement 2 , we consider two possibilities: either $x_{2}=1$ or $x_{2}=0$.

Case $1 x_{2}=1$. If $x_{2}=1$, then the (11) implies that $x_{1} \geq 1$. Because $x_{1}$ must equal 0 or 1 , this implies that $x_{1}=1$, as required by 2 .

Case $2 x_{2}=0$. In this case, (11) reduces to $x_{1} \geq 0$, which allows $x_{1}=0$ or $x_{1}=1$. In short, if $x_{2}=0,(11)$ does not restrict the value of $x_{1}$. This is also consistent with requirement 2 .

In summary, for any value of $x_{2},(11)$ is equivalent to requirement 2 .
3 Simply add the constraint

$$
\begin{equation*}
x_{2}+x_{4} \leq 1 \tag{12}
\end{equation*}
$$

to (9). We now show that for the two cases $x_{2}=1$ and $x_{2}=0$, (12) is equivalent to the third requirement.

Case $1 x_{2}=1$. In this case, we are investing in investment 2, and requirement 3 implies that Stockco cannot invest in investment 4 (that is, $x_{4}$ must equal 0 ). Note that if $x_{2}=1$, then (12) does imply $1+x_{4} \leq 1$, or $x_{4} \leq 0$. Thus, if $x_{2}=1$, then (12) is consistent with requirement 3 .

Case $2 x_{2}=0$. In this case, requirement 3 does not restrict the value of $x_{4}$. Note that if $x_{2}=0$, then (12) reduces to $x_{4} \leq 1$, which also leaves $x_{4}$ free to equal 0 or 1 .

## Fixed-Charge Problems

Example 3 illustrates an important trick that can be used to formulate many location and production problems as IPs.

## EXAMPLE 3 Fixed-Charge IP

Gandhi Cloth Company is capable of manufacturing three types of clothing: shirts, shorts, and pants. The manufacture of each type of clothing requires that Gandhi have the appropriate type of machinery available. The machinery needed to manufacture each type of clothing must be rented at the following rates: shirt machinery, $\$ 200$ per week; shorts machinery, $\$ 150$ per week; pants machinery, $\$ 100$ per week. The manufacture of each type of clothing also requires the amounts of cloth and labor shown in Table 2. Each week, 150 hours of labor and 160 sq yd of cloth are available. The variable unit cost and selling price for each type of clothing are shown in Table 3. Formulate an IP whose solution will maximize Gandhi's weekly profits.

Solution As in LP formulations, we define a decision variable for each decision that Gandhi must make. Clearly, Gandhi must decide how many of each type of clothing should be manufactured each week, so we define
$x_{1}=$ number of shirts produced each week
$x_{2}=$ number of shorts produced each week
$x_{3}=$ number of pants produced each week
table 2
Resource Requirements for Gandhi

| Clothing <br> Type | Labor <br> (Hours) | Cloth <br> (Square Yards) |
| :--- | :---: | :---: |
| Shirt | 3 | 4 |
| Shorts | 2 | 3 |
| Pants | 6 | 4 |

TABLE 3
Revenue and Cost Information for Gandhi

| Clothing <br> Type | Sales <br> Price $(\$)$ | Variable <br> Cost (\$) |
| :--- | :---: | :---: |
| Shirt | 12 | 6 |
| Shorts | 8 | 4 |
| Pants | 15 | 8 |

Note that the cost of renting machinery depends only on the types of clothing produced, not on the amount of each type of clothing. This enables us to express the cost of renting machinery by using the following variables:

$$
\begin{aligned}
& y_{1}= \begin{cases}1 & \text { if any shirts are manufactured } \\
0 & \text { otherwise }\end{cases} \\
& y_{2}= \begin{cases}1 & \text { if any shorts are manufactured } \\
0 & \text { otherwise }\end{cases} \\
& y_{3}= \begin{cases}1 & \text { if any pants are manufactured } \\
0 & \text { otherwise }\end{cases}
\end{aligned}
$$

In short, if $x_{j}>0$, then $y_{j}=1$, and if $x_{j}=0$, then $y_{j}=0$. Thus, Gandhi's weekly profits $=$ (weekly sales revenue) - (weekly variable costs) - (weekly costs of renting machinery).

Also,

$$
\begin{equation*}
\text { Weekly cost of renting machinery }=200 y_{1}+150 y_{2}+100 y_{3} \tag{13}
\end{equation*}
$$

To justify (13), note that it picks up the rental costs only for the machines needed to manufacture those products that Gandhi is actually manufacturing. For example, suppose that shirts and pants are manufactured. Then $y_{1}=y_{3}=1$ and $y_{2}=0$, and the total weekly rental cost will be $200+100=\$ 300$.

Because the cost of renting, say, shirt machinery does not depend on the number of shirts produced, the cost of renting each type of machinery is called a fixed charge. A fixed charge for an activity is a cost that is assessed whenever the activity is undertaken at a nonzero level. The presence of fixed charges will make the formulation of the Gandhi problem much more difficult.

We can now express Gandhi's weekly profits as

$$
\begin{aligned}
\text { Weekly profit }= & \left(12 x_{1}+8 x_{2}+15 x_{3}\right)-\left(6 x_{1}+4 x_{2}+8 x_{3}\right) \\
& -\left(200 y_{1}+150 y_{2}+100 y_{3}\right) \\
= & 6 x_{1}+4 x_{2}+7 x_{3}-200 y_{1}-150 y_{2}-100 y_{3}
\end{aligned}
$$

Thus, Gandhi wants to maximize

$$
z=6 x_{1}+4 x_{2}+7 x_{3}-200 y_{1}-150 y_{2}-100 y_{3}
$$

Because its supply of labor and cloth is limited, Gandhi faces the following two constraints:
Constraint 1 At most, 150 hours of labor can be used each week.
Constraint 2 At most, 160 sq yd of cloth can be used each week.
Constraint 1 is expressed by

$$
\begin{equation*}
3 x_{1}+2 x_{2}+6 x_{3} \leq 150 \quad \text { (Labor constraint) } \tag{14}
\end{equation*}
$$

Constraint 2 is expressed by

$$
\begin{equation*}
4 x_{1}+3 x_{2}+4 x_{3} \leq 160 \quad(\text { Cloth constraint }) \tag{15}
\end{equation*}
$$

Observe that $x_{j}>0$ and $x_{j}$ integer $(j=1,2,3)$ must hold along with $y_{j}=0$ or $1(j=$ $1,2,3$ ). Combining (14) and (15) with these restrictions and the objective function yields the following IP:

$$
\begin{align*}
& \max z=6 x_{1}+4 x_{2}+7 x_{3}-200 y_{1}-150 y_{2}-100 y_{3} \\
& \text { s.t. } 3 x_{1}+2 x_{2}+6 x_{3} \leq 150 \\
& 4 x_{1}+3 x_{2}+4 x_{3} \leq 160  \tag{IP1}\\
& x_{1}, x_{2}, x_{3} \geq 0 ; x_{1}, x_{2}, x_{3} \text { integer } \\
& y_{1}, y_{2}, y_{3}=0 \text { or } 1
\end{align*}
$$

The optimal solution to this problem is found to be $x_{1}=30, x_{3}=10, x_{2}=y_{1}=y_{2}=$ $y_{3}=0$. This cannot be the optimal solution to Gandhi's problem because it indicates that Gandhi can manufacture shirts and pants without incurring the cost of renting the needed machinery. The current formulation is incorrect because the variables $y_{1}, y_{2}$, and $y_{3}$ are not present in the constraints. This means that there is nothing to stop us from setting $y_{1}=y_{2}=y_{3}=0$. Setting $y_{i}=0$ is certainly less costly than setting $y_{i}=1$, so a minimumcost solution to (IP 1) will always set $y_{i}=0$. Somehow we must modify (IP 1) so that whenever $x_{i}>0, y_{i}=1$ must hold. The following trick will accomplish this goal. Let $M_{1}$, $M_{2}$, and $M_{3}$ be three large positive numbers, and add the following constraints to (IP 1):

$$
\begin{align*}
x_{1} & \leq M_{1} y_{1}  \tag{16}\\
x_{2} & \leq M_{2} y_{2}  \tag{17}\\
x_{3} & \leq M_{3} y_{3} \tag{18}
\end{align*}
$$

Adding (16)-(18) to IP 1 will ensure that if $x_{i}>0$, then $y_{i}=1$. To illustrate, let us show that (16) ensures that if $x_{1}>0$, then $y_{1}=1$. If $x_{1}>0$, then $y_{1}$ cannot be 0 . For if $y_{1}=$ 0 , then (16) would imply $x_{1} \leq 0$ or $x_{1}=0$. Thus, if $x_{1}>0, y_{1}=1$ must hold. If any shirts are produced $\left(x_{1}>0\right),(16)$ ensures that $y_{1}=1$, and the objective function will include the cost of the machinery needed to manufacture shirts. Note that if $y_{1}=1$, then (16) becomes $x_{1} \leq M_{1}$, which does not unnecessarily restrict the value of $x_{1}$. If $M_{1}$ were not chosen large, however (say, $M_{1}=10$ ), then (16) would unnecessarily restrict the value of $x_{1}$. In general, $M_{i}$ should be set equal to the maximum value that $x_{i}$ can attain. In the current problem, at most 40 shirts can be produced (if Gandhi produced more than 40 shirts, the company would run out of cloth), so we can safely choose $M_{1}=40$. The reader should verify that we can choose $M_{2}=53$ and $M_{3}=25$.

If $x_{1}=0$, (16) becomes $0 \leq M_{1} y_{1}$. This allows either $y_{1}=0$ or $y_{1}=1$. Because $y_{1}=$ 0 is less costly than $y_{1}=1$, the optimal solution will choose $y_{1}=0$ if $x_{1}=0$. In summary, we have shown that if (16)-(18) are added to (IP 1), then $x_{i}>0$ will imply $y_{i}=1$, and $x_{i}=0$ will imply $y_{i}=0$.

The optimal solution to the Gandhi problem is $z=\$ 75, x_{3}=25, y_{3}=1$. Thus, Gandhi should produce 25 pants each week.

The Gandhi problem is an example of a fixed-charge problem. In a fixed-charge problem, there is a cost associated with performing an activity at a nonzero level that does not depend on the level of the activity. Thus, in the Gandhi problem, if we make any shirts at all (no matter how many we make), we must pay the fixed charge of $\$ 200$ to rent a shirt machine. Problems in which a decision maker must choose where to locate facilities are often fixed-charge problems. The decision maker must choose where to locate various fa-
cilities (such as plants, warehouses, or business offices), and a fixed charge is often associated with building or operating a facility. Example 4 is a typical location problem involving the idea of a fixed charge.

## EXAMPLE 4 The Lockbox Problem

J. C. Nickles receives credit card payments from four regions of the country (West, Midwest, East, and South). The average daily value of payments mailed by customers from each region is as follows: the West, $\$ 70,000$; the Midwest, $\$ 50,000$; the East, $\$ 60,000$; the South, $\$ 40,000$. Nickles must decide where customers should mail their payments. Because Nickles can earn $20 \%$ annual interest by investing these revenues, it would like to receive payments as quickly as possible. Nickles is considering setting up operations to process payments (often referred to as lockboxes) in four different cities: Los Angeles, Chicago, New York, and Atlanta. The average number of days (from time payment is sent) until a check clears and Nickles can deposit the money depends on the city to which the payment is mailed, as shown in Table 4. For example, if a check is mailed from the West to Atlanta, it would take an average of 8 days before Nickles could earn interest on the check. The annual cost of running a lockbox in any city is $\$ 50,000$. Formulate an IP that Nickles can use to minimize the sum of costs due to lost interest and lockbox operations. Assume that each region must send all its money to a single city and that there is no limit on the amount of money that each lockbox can handle.
Solution Nickles must make two types of decisions. First, Nickles must decide where to operate lockboxes. We define, for $j=1,2,3,4$,

$$
y_{j}= \begin{cases}1 & \text { if a lockbox is operated in city } j \\ 0 & \text { otherwise }\end{cases}
$$

Thus, $y_{2}=1$ if a lockbox is operated in Chicago, and $y_{3}=0$ if no lockbox is operated in New York. Second, Nickles must determine where each region of the country should send payments. We define (for $i, j=1,2,3,4$ )

$$
x_{i j}= \begin{cases}1 & \text { if region } i \text { sends payments to city } j \\ 0 & \text { otherwise }\end{cases}
$$

For example, $x_{12}=1$ if the West sends payments to Chicago, and $x_{23}=0$ if the Midwest does not send payments to New York.

Nickles wants to minimize (total annual cost) $=($ annual cost of operating lockboxes $)+$ (annual lost interest cost). To determine how much interest Nickles loses annually, we must determine how much revenue would be lost if payments from region $i$ were sent to region $j$. For example, how much in annual interest would Nickles lose if customers from the West region sent payments to New York? On any given day, 8 days' worth, or $8(70,000)=\$ 560,000$ of West payments will be in the mail and will not be earning in-
table 4
Average Number of Days from Mailing of Payment Until Payment Clears

|  | To |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | City 1 <br> (Los Angeles) | City 2 <br> (Chicago) | City 3 <br> (New York) | City 4 <br> (Atlanta) |
| Region 1 West | 2 | 6 | 8 | 8 |
| Region 2 Midwest | 6 | 2 | 5 | 5 |
| Region 3 East | 8 | 5 | 2 | 5 |
| Region 4 South | 8 | 5 | 5 | 2 |

terest. Because Nickles can earn $20 \%$ annually, each year West funds will result in $0.20(560,000)=\$ 112,000$ in lost interest. Similar calculations for the annual cost of lost interest for each possible assignment of a region to a city yield the results shown in Table 5. The lost interest cost from sending region $i$ 's payments to city $j$ is only incurred if $x_{i j}=1$, so Nickles's annual lost interest costs (in thousands) are

$$
\begin{aligned}
\text { Annual lost interest costs }= & 28 x_{11}+84 x_{12}+112 x_{13}+112 x_{14} \\
& +60 x_{21}+20 x_{22}+50 x_{23}+50 x_{24} \\
& +96 x_{31}+60 x_{32}+24 x_{33}+60 x_{34} \\
& +64 x_{41}+40 x_{42}+40 x_{43}+16 x_{44}
\end{aligned}
$$

The cost of operating a lockbox in city $i$ is incurred if and only if $y_{i}=1$, so the annual lockbox operating costs (in thousands) are given by

$$
\text { Total annual lockbox operating cost }=50 y_{1}+50 y_{2}+50 y_{3}+50 y_{4}
$$

Thus, Nickles's objective function may be written as

$$
\begin{align*}
\min z= & 28 x_{11}+84 x_{12}+112 x_{13}+112 x_{14} \\
& +60 x_{21}+20 x_{22}+50 x_{23}+50 x_{24} \\
& +96 x_{31}+60 x_{32}+24 x_{33}+60 x_{34}  \tag{19}\\
& +64 x_{41}+40 x_{42}+40 x_{43}+16 x_{44} \\
& +50 y_{1}+50 y_{2}+50 y_{3}+50 y_{4}
\end{align*}
$$

Nickles faces two types of constraints.
Type 1 Constraint Each region must send its payments to a single city.
Type 2 Constraint If a region is assigned to send its payments to a city, that city must have a lockbox.

## TABLE 5

Calculation of Annual Lost Interest

| Assignment | Annual Lost <br> Interest Cost (\$) |
| :--- | :---: |
| West to L.A. | $0.20(70,000) 2=28,000$ |
| West to Chicago | $0.20(70,000) 6=84,000$ |
| West to N.Y. | $0.20(70,000) 8=112,000$ |
| West to Atlanta | $0.20(70,000) 8=112,000$ |
| Midwest to L.A. | $0.20(50,000) 6=60,000$ |
| Midwest to Chicago | $0.20(50,000) 2=20,000$ |
| Midwest to N.Y. | $0.20(50,000) 5=50,000$ |
| Midwest to Atlanta | $0.20(50,000) 5=50,000$ |
| East to L.A. | $0.20(60,000) 8=96,000$ |
| East to Chicago | $0.20(60,000) 5=60,000$ |
| East to N.Y. | $0.20(60,000) 2=24,000$ |
| East to Atlanta | $0.20(60,000) 5=60,000$ |
| South to L.A. | $0.20(40,000) 8=64,000$ |
| South to Chicago | $0.20(40,000) 5=40,000$ |
| South to N.Y | $0.20(40,000) 5=40,000$ |
| South to Atlanta | $0.20(40,000) 2=16,000$ |

The type 1 constraints state that for region $i(i=1,2,3,4)$ exactly one of $x_{i 1}, x_{i 2}, x_{i 3}$, and $x_{i 4}$ must equal 1 and the others must equal 0 . This can be accomplished by including the following four constraints:

$$
\begin{array}{ll}
x_{11}+x_{12}+x_{13}+x_{14}=1 & \text { (West region constraint) } \\
x_{21}+x_{22}+x_{23}+x_{24}=1 & \text { (Midwest region constraint) } \\
x_{31}+x_{32}+x_{33}+x_{34}=1 & \text { (East region constraint) } \\
x_{41}+x_{42}+x_{43}+x_{44}=1 & \text { (South region constraint) } \tag{23}
\end{array}
$$

The type 2 constraints state that if

$$
\begin{equation*}
x_{i j}=1 \quad \text { (that is, customers in region } i \text { send payments to city } j \text { ) } \tag{24}
\end{equation*}
$$

then $y_{j}$ must equal 1 . For example, suppose $x_{12}=1$. Then there must be a lockbox at city 2 , so $y_{2}=1$ must hold. This can be ensured by adding 16 constraints of the form

$$
\begin{equation*}
x_{i j} \leq y_{j} \quad(i=1,2,3,4 ; j=1,2,3,4) \tag{25}
\end{equation*}
$$

If $x_{i j}=1$, then (25) ensures that $y_{j}=1$, as desired. Also, if $x_{1 j}=x_{2 j}=x_{3 j}=x_{4 j}=0$, then (25) allows $y_{j}=0$ or $y_{j}=1$. As in the fixed-charge example, the act of minimizing costs will result in $y_{j}=0$. In summary, the constraints in (25) ensure that Nickles pays for a lockbox at city $i$ if it uses a lockbox at city $i$.

Combining (19)-(23) with the $4(4)=16$ constraints in (25) and the $0-1$ restrictions on the variables yields the following formulation:

$$
\begin{aligned}
\min z= & 28 x_{11}+84 x_{12}+112 x_{13}+112 x_{14}+60 x_{21}+20 x_{22}+50 x_{23}+50 x_{24} \\
& +96 x_{31}+60 x_{32}+24 x_{33}+60 x_{34}+64 x_{41}+40 x_{42}+40 x_{43}+16 x_{44} \\
& +50 y_{1}+50 y_{2}+50 y_{3}+50 y_{4} \\
\text { s.t. } \quad & x_{11}+x_{12}+x_{13}+x_{14}=1 \quad \text { (West region constraint) } \\
& x_{21}+x_{22}+x_{23}+x_{24}=1 \quad \text { (Midwest region constraint) } \\
& x_{31}+x_{32}+x_{33}+x_{34}=1 \quad \text { (East region constraint) } \\
& x_{41}+x_{42}+x_{43}+x_{44}=1 \quad \text { (South region constraint) } \\
& x_{11} \leq y_{1}, x_{21} \leq y_{1}, x_{31} \leq y_{1}, x_{41} \leq y_{1}, x_{12} \leq y_{2}, x_{22} \leq y_{2}, x_{32} \leq y_{2}, x_{42} \leq y_{2} \\
& x_{13} \leq y_{3}, x_{23} \leq y_{3}, x_{33} \leq y_{3}, x_{43} \leq y_{3}, x_{14} \leq y_{4}, x_{24} \leq y_{4}, x_{34} \leq y_{4}, x_{44} \leq y_{4}
\end{aligned}
$$

All $x_{i j}$ and $y_{j}=0$ or 1
The optimal solution is $z=242, y_{1}=1, y_{3}=1, x_{11}=1, x_{23}=1, x_{33}=1, x_{43}=1$. Thus, Nickles should have a lockbox operation in Los Angeles and New York. West customers should send payments to Los Angeles, and all other customers should send payments to New York.

There is an alternative way of modeling the Type 2 constraints. Instead of the 16 constraints of the form $x_{i j} \leq y_{j}$, we may include the following four constraints:

$$
\begin{array}{ll}
x_{11}+x_{21}+x_{31}+x_{41} \leq 4 y_{1} & \\
x_{12}+x_{22}+x_{32}+x_{42} \leq 4 y_{2} & \\
\text { (Los Angeles constraint) } \\
x_{13}+x_{23}+x_{33}+x_{43} \leq 4 y_{3} & \\
\text { (New York constraint) } \\
x_{14}+x_{24}+x_{34}+x_{44} \leq 4 y_{4} & \\
\text { (Atlanta constraint) }
\end{array}
$$

For the given city, each constraint ensures that if the lockbox is used, then Nickles must pay for it. For example, consider $x_{14}+x_{24}+x_{34}+x_{44} \leq 4 y_{4}$. The lockbox in Atlanta is used if $x_{14}=1, x_{24}=1, x_{34}=1$, or $x_{44}=1$. If any of these variables equals 1 , then the Atlanta constraint ensures that $y_{4}=1$, and Nickles must pay for the lockbox. If all these variables are 0 , then the act of minimizing costs will cause $y_{4}=0$, and the cost of the At-
lanta lockbox will not be incurred. Why does the right-hand side of each constraint equal 4? This ensures that for each city, it is possible to send money from all four regions to the city. In Section 9.3, we discuss which of the two alternative formulations of the lockbox problem is easier for a computer to solve. The answer may surprise you!

## Set-Covering Problems

The following example is typical of an important class of IPs known as set-covering problems.

## EXAMPLE $5 \quad$ Facility-Location Set-Covering Problem

There are six cities (cities 1-6) in Kilroy County. The county must determine where to build fire stations. The county wants to build the minimum number of fire stations needed to ensure that at least one fire station is within 15 minutes (driving time) of each city. The times (in minutes) required to drive between the cities in Kilroy County are shown in Table 6. Formulate an IP that will tell Kilroy how many fire stations should be built and where they should be located.

Solution For each city, Kilroy must determine whether to build a fire station there. We define the $0-1$ variables $x_{1}, x_{2}, x_{3}, x_{4}, x_{5}$, and $x_{6}$ by

$$
x_{i}= \begin{cases}1 & \text { if a fire station is built in city } i \\ 0 & \text { otherwise }\end{cases}
$$

Then the total number of fire stations that are built is given by $x_{1}+x_{2}+x_{3}+x_{4}+$ $x_{5}+x_{6}$, and Kilroy's objective function is to minimize

$$
z=x_{1}+x_{2}+x_{3}+x_{4}+x_{5}+x_{6}
$$

What are Kilroy's constraints? Kilroy must ensure that there is a fire station within 15 minutes of each city. Table 7 indicates which locations can reach the city in 15 minutes or less. To ensure that at least one fire station is within 15 minutes of city 1 , we add the constraint

$$
x_{1}+x_{2} \geq 1 \quad \text { (City } 1 \text { constraint) }
$$

This constraint ensures that $x_{1}=x_{2}=0$ is impossible, so at least one fire station will be built within 15 minutes of city 1 . Similarly the constraint

$$
\left.x_{1}+x_{2}+x_{6} \geq 1 \quad \text { (City } 2 \text { constraint }\right)
$$

ensures that at least one fire station will be located within 15 minutes of city 2 . In a similar fashion, we obtain constraints for cities 3-6. Combining these six constraints with the
table 6
Time Required to Travel between Cities in Kilroy County

|  | To |  |  |  |  |  |
| :--- | :---: | ---: | ---: | :---: | :---: | :---: |
| From | City $\mathbf{1}$ | City 2 | City 3 | City 4 | City 5 | City 6 |
| City 1 | 0 | 10 | 20 | 30 | 30 | 20 |
| City 2 | 10 | 0 | 25 | 35 | 20 | 10 |
| City 3 | 20 | 25 | 0 | 15 | 30 | 20 |
| City 4 | 30 | 35 | 15 | 0 | 15 | 25 |
| City 5 | 30 | 20 | 30 | 15 | 0 | 14 |
| City 6 | 20 | 10 | 20 | 25 | 14 | 0 |

TABLE 7
Cities within 15 Minutes of
Given City

| City | Within $\mathbf{1 5}$ Minutes |
| :--- | :---: |
| 1 | 1,2 |
| 2 | $1,2,6$ |
| 3 | 3,4 |
| 4 | $3,4,5$ |
| 5 | $4,5,6$ |
| 6 | $2,5,6$ |

objective function (and with the fact that each variable must equal 0 or 1 ), we obtain the following $0-1 \mathrm{IP}$ :

$$
\begin{aligned}
& \min z=x_{1}+x_{2}+x_{3}+x_{4}+x_{5}+x_{6} \\
& \text { s.t. } x_{1}+x_{2} \quad \geq 1 \quad \text { (City } 1 \text { constraint) } \\
& x_{1}+x_{2} \quad+x_{6} \geq 1 \quad \text { (City } 2 \text { constraint) } \\
& x_{3}+x_{4} \quad \geq 1 \quad \text { (City } 3 \text { constraint) } \\
& x_{3}+x_{4}+x_{5} \quad \geq 1 \quad \text { (City } 4 \text { constraint) } \\
& x_{4}+x_{5}+x_{6} \geq 1 \quad \text { (City } 5 \text { constraint) } \\
& +x_{5}+x_{6} \geq 1 \quad \text { (City } 6 \text { constraint) } \\
& x_{i}=0 \text { or } 1 \quad(i=1,2,3,4,5,6)
\end{aligned}
$$

One optimal solution to this IP is $z=2, x_{2}=x_{4}=1, x_{1}=x_{3}=x_{5}=x_{6}=0$. Thus, Kilroy County can build two fire stations: one in city 2 and one in city 4 .

As noted, Example 5 represents a class of IPs known as set-covering problems. In a set-covering problem, each member of a given set (call it set 1 ) must be "covered" by an acceptable member of some set (call it set 2). The objective in a set-covering problem is to minimize the number of elements in set 2 that are required to cover all the elements in set 1 . In Example 5, set 1 is the cities in Kilroy County, and set 2 is the set of fire stations. The station in city 2 covers cities 1 , 2 , and 6 , and the station in city 4 covers cities 3,4 , and 5 . Set-covering problems have many applications in areas such as airline crew scheduling, political districting, airline scheduling, and truck routing.

## Either-Or Constraints

The following situation commonly occurs in mathematical programming problems. We are given two constraints of the form

$$
\begin{align*}
f\left(x_{1}, x_{2}, \ldots, x_{n}\right) & \leq 0  \tag{26}\\
g\left(x_{1}, x_{2}, \ldots, x_{n}\right) & \leq 0 \tag{27}
\end{align*}
$$

We want to ensure that at least one of (26) and (27) is satisfied, often called either-or constraints. Adding the two constraints $\left(26^{\prime}\right)$ and $\left(27^{\prime}\right)$ to the formulation will ensure that at least one of (26) and (27) is satisfied:

$$
\begin{align*}
& f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq M y \\
& g\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq M(1-y) \tag{27'}
\end{align*}
$$

In (26') and (27'), $y$ is a $0-1$ variable, and $M$ is a number chosen large enough to ensure that $f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq M$ and $g\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq M$ are satisfied for all values of $x_{1}, x_{2}, \ldots, x_{n}$ that satisfy the other constraints in the problem.

Let us show that the inclusion of constraints (26') and (27') is equivalent to at least one of (26) and (27) being satisfied. Either $y=0$ or $y=1$. If $y=0$, then ( $26^{\prime}$ ) and (27') become $f \leq 0$ and $g \leq M$. Thus, if $y=0$, then (26) (and possibly (27)) must be satisfied. Similarly, if $y=1$, then (26') and (27') become $f \leq M$ and $g \leq 0$. Thus, if $y=1$, then (27) (and possibly (26)) must be satisfied. Therefore, whether $y=0$ or $y=1$, (26') and (27') ensure that at least one of (26) and (27) is satisfied.

The following example illustrates the use of either-or constraints.

## EXAMPLE 6 Ether-Or Constraint

Dorian Auto is considering manufacturing three types of autos: compact, midsize, and large. The resources required for, and the profits yielded by, each type of car are shown in Table 8. Currently, 6,000 tons of steel and 60,000 hours of labor are available. For production of a type of car to be economically feasible, at least 1,000 cars of that type must be produced. Formulate an IP to maximize Dorian's profit.

Solution Because Dorian must determine how many cars of each type should be built, we define
$x_{1}=$ number of compact cars produced
$x_{2}=$ number of midsize cars produced
$x_{3}=$ number of large cars produced

Then contribution to profit (in thousands of dollars) is $2 x_{1}+3 x_{2}+4 x_{3}$, and Dorian's objective function is

$$
\max z=2 x_{1}+3 x_{2}+4 x_{3}
$$

We know that if any cars of a given type are produced, then at least 1,000 cars of that type must be produced. Thus, for $i=1,2,3$, we must have $x_{i} \leq 0$ or $x_{i} \geq 1,000$. Steel and labor are limited, so Dorian must satisfy the following five constraints:

Constraint $1 \quad x_{1} \leq 0$ or $x_{1} \geq 1,000$.
Constraint $2 x_{2} \leq 0$ or $x_{2} \geq 1,000$.
Constraint $3 \quad x_{3} \leq 0$ or $x_{3} \geq 1,000$.
Constraint 4 The cars produced can use at most 6,000 tons of steel.
Constraint 5 The cars produced can use at most 60,000 hours of labor.

TABLE 8
Resources and Profits for Three Types of Cars

|  | Car Type |  |  |
| :--- | :--- | :--- | :--- |
| Resource | Compact | Midsize | Large |
| Steel required | 1.5 tons | 3 tons | 5 tons |
| Labor required | 30 hours | 25 hours | 40 hours |
| Profit yielded (\$) | 2,000 | 3,000 | 4,000 |

From our previous discussion, we see that if we define $f\left(x_{1}, x_{2}, x_{3}\right)=x_{1}$ and $g\left(x_{1}, x_{2}\right.$, $\left.x_{3}\right)=1,000-x_{1}$, we can replace Constraint 1 by the following pair of constraints:

$$
\begin{gathered}
x_{1} \leq M_{1} y_{1} \\
1,000-x_{1} \leq M_{1}\left(1-y_{1}\right) \\
y_{1}=0 \text { or } 1
\end{gathered}
$$

To ensure that both $x_{1}$ and $1,000-x_{1}$ will never exceed $M_{1}$, it suffices to choose $M_{1}$ large enough so that $M_{1}$ exceeds 1,000 and $x_{1}$ is always less than $M_{1}$. Building $\frac{60,000}{30}=2,000$ compacts would use all available labor (and still leave some steel), so at most 2,000 compacts can be built. Thus, we may choose $M_{1}=2,000$. Similarly, Constraint 2 may be replaced by the following pair of constraints:

$$
\begin{gathered}
x_{2} \leq M_{2} y_{2} \\
1,000-x_{2} \leq M_{2}\left(1-y_{2}\right) \\
y_{2}=0 \text { or } 1
\end{gathered}
$$

You should verify that $M_{2}=2,000$ is satisfactory. Similarly, Constraint 3 may be replaced by

$$
\begin{gathered}
x_{3} \leq M_{3} y_{3} \\
1,000-x_{3} \leq M_{3}\left(1-y_{3}\right) \\
y_{3}=0 \text { or } 1
\end{gathered}
$$

Again, you should verify that $M_{3}=1,200$ is satisfactory. Constraint 4 is a straightforward resource constraint that reduces to

$$
1.5 x_{1}+3 x_{2}+5 x_{3} \leq 6,000 \quad(\text { Steel constraint })
$$

Constraint 5 is a straightforward resource usage constraint that reduces to

$$
30 x_{1}+25 x_{2}+40 x_{3} \leq 60,000 \quad(\text { Labor constraint })
$$

After noting that $x_{i} \geq 0$ and that $x_{i}$ must be an integer, we obtain the following IP:

$$
\begin{aligned}
\max z=2 x_{1}+3 x_{2} & +4 x_{3} \\
\text { s.t. } \quad x_{1} & \leq 2,000 y_{1} \\
1,000-x_{1} & \leq 2,000\left(1-y_{1}\right) \\
x_{2} & \leq 2,000 y_{2} \\
1,000-x_{2} & \leq 2,000\left(1-y_{2}\right) \\
x_{3} & \leq 1,200 y_{3} \\
1,000-x_{3} & \leq 1,200\left(1-y_{3}\right) \\
1.5 x_{1}+3 x_{2}+5 x_{3} & \leq 6,000 \quad \text { (Steel constraint) } \\
30 x_{1}+25 x_{2}+40 x_{3} & \leq 60,000 \quad \text { (Labor constraint) } \\
x_{1}, x_{2}, x_{3} & \geq 0 ; x_{1}, x_{2}, x_{3} \text { integer } \\
y_{1}, y_{2}, y_{3} & =0 \text { or } 1
\end{aligned}
$$

The optimal solution to the IP is $z=6,000, x_{2}=2,000, y_{2}=1, y_{1}=y_{3}=x_{1}=x_{3}=0$. Thus, Dorian should produce 2,000 midsize cars. If Dorian had not been required to manufacture at least 1,000 cars of each type, then the optimal solution would have been to produce 570 compacts and 1,715 midsize cars.

## If-Then Constraints

In many applications, the following situation occurs: We want to ensure that if a constraint $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)>0$ is satisfied, then the constraint $g\left(x_{1}, x_{2}, \ldots, x_{n}\right) \geq 0$ must be satisfied, while if $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)>0$ is not satisfied, then $g\left(x_{1}, x_{2}, \ldots, x_{n}\right) \geq 0$ may or may not be satisfied. In short, we want to ensure that $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)>0$ implies $g\left(x_{1}, x_{2}\right.$, $\left.\ldots, x_{n}\right) \geq 0$.

To ensure this, we include the following constraints in the formulation:

$$
\begin{gather*}
-g\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq M y  \tag{28}\\
f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq M(1-y)  \tag{29}\\
y=0 \text { or } 1
\end{gather*}
$$

As usual, $M$ is a large positive number. ( $M$ must be chosen large enough so that $f \leq M$ and $-g \leq M$ hold for all values of $x_{1}, x_{2}, \ldots, x_{n}$ that satisfy the other constraints in the problem.) Observe that if $f>0$, then (29) can be satisfied only if $y=0$. Then (28) implies $-g \leq 0$, or $g \geq 0$, which is the desired result. Thus, if $f>0$, then (28) and (29) ensure that $g \geq 0$. Also, if $f>0$ is not satisfied, then (29) allows $y=0$ or $y=1$. By choosing $y=1,(28)$ is automatically satisfied. Thus, if $f>0$ is not satisfied, then the values of $x_{1}, x_{2}, \ldots, x_{n}$ are unrestricted and $g<0$ or $g \geq 0$ are both possible.

To illustrate the use of this idea, suppose we add the following constraint to the Nickles lockbox problem: If customers in region 1 send their payments to city 1 , then no other customers may send their payments to city 1 . Mathematically, this restriction may be expressed by

$$
\begin{equation*}
\text { If } x_{11}=1, \quad \text { then } \quad x_{21}=x_{31}=x_{41}=0 \tag{30}
\end{equation*}
$$

Because all $x_{i j}$ must equal 0 or $1,(30)$ may be written as
If $x_{11}>0, \quad$ then $\quad x_{21}+x_{31}+x_{41} \leq 0, \quad$ or $\quad-x_{21}-x_{31}-x_{41} \geq 0$
If we define $f=x_{11}$ and $g=-x_{21}-x_{31}-x_{41}$, we can use (28) and (29) to express (30') [and therefore (30)] by the following two constraints:

$$
\begin{gathered}
x_{21}+x_{31}+x_{41} \leq M y \\
x_{11} \leq M(1-y) \\
y=0 \text { or } 1
\end{gathered}
$$

Because $-g$ and $f$ can never exceed 3, we can choose $M=3$ and add the following constraints to the original lockbox formulation:

$$
\begin{gathered}
x_{21}+x_{31}+x_{41} \leq 3 y \\
x_{11} \leq 3(1-y) \\
y=0 \text { or } 1
\end{gathered}
$$

## Integer Programming and Piecewise Linear Functions ${ }^{\dagger}$

The next example shows how $0-1$ variables can be used to model optimization problems involving piecewise linear functions. A piecewise linear function consists of several straight-line segments. The piecewise linear function in Figure 2 is made of four straightline segments. The points where the slope of the piecewise linear function changes (or the range of definition of the function ends) are called the break points of the function. Thus, $0,10,30,40$, and 50 are the break points of the function pictured in Figure 2.

[^13]

To illustrate why piecewise linear functions can occur in applications, suppose we manufacture gasoline from oil. In purchasing oil from our supplier, we receive a quantity discount. The first 500 gallons of oil purchased cost $25 \phi$ per gallon; the next 500 gallons cost $20 \phi$ per gallon; and the next 500 gallons cost $15 \phi$ per gallon. At most, 1,500 gallons of oil can be purchased. Let $x$ be the number of gallons of oil purchased and $c(x)$ be the cost (in cents) of purchasing $x$ gallons of oil. For $x \leq 0, c(x)=0$. Then for $0 \leq x \leq 500$, $c(x)=25 x$. For $500 \leq x \leq 1,000, c(x)=$ (cost of purchasing first 500 gallons at $25 \notin$ per gallon) + (cost of purchasing next $x-500$ gallons at $20 ¢$ per gallon $)=25(500)+$ $20(x-500)=20 x+2,500$. For $1,000 \leq x \leq 1,500, c(x)=$ (cost of purchasing first $1,000$ gallons $)+$ (cost of purchasing next $x-1,000$ gallons at $15 ¢$ per gallon $)=$ $c(1,000)+15(x-1,000)=7,500+15 x$. Thus, $c(x)$ has break points $0,500,1,000$, and 1,500 and is graphed in Figure 3.

A piecewise linear function is not a linear function, so one might think that linear programming could not be used to solve optimization problems involving these functions. By using $0-1$ variables, however, piecewise linear functions can be represented in linear form. Suppose that a piecewise linear function $f(x)$ has break points $b_{1}, b_{2}, \ldots, b_{n}$. For some $k$ $(k=1,2, \ldots, n-1), b_{k} \leq x \leq b_{k+1}$. Then, for some number $z_{k}\left(0 \leq z_{k} \leq 1\right), x$ may be written as

$$
x=z_{k} b_{k}+\left(1-z_{k}\right) b_{k+1}
$$

Because $f(x)$ is linear for $b_{k} \leq x \leq b_{k+1}$, we may write

$$
f(x)=z_{k} f\left(b_{k}\right)+\left(1-z_{k}\right) f\left(b_{k+1}\right)
$$

To illustrate the idea, take $x=800$ in our oil example. Then we have $b_{2}=500 \leq 800 \leq$ $1,000=b_{3}$, and we may write

$$
\begin{aligned}
x & =\frac{2}{5}(500)+\frac{3}{5}(1,000) \\
f(x) & =f(800)=\frac{2}{5} f(500)+\frac{3}{5} f(1,000) \\
& =\frac{2}{5}(12,500)+\frac{3}{5}(22,500)=18,500
\end{aligned}
$$

We are now ready to describe the method used to express a piecewise linear function via linear constraints and $0-1$ variables:

Step 1 Wherever $f(x)$ occurs in the optimization problem, replace $f(x)$ by $z_{1} f\left(b_{1}\right)+$ $z_{2} f\left(b_{2}\right)+\cdots+z_{n} f\left(b_{n}\right)$.

Step 2 Add the following constraints to the problem:

$$
\begin{gathered}
z_{1} \leq y_{1}, z_{2} \leq y_{1}+y_{2}, z_{3} \leq y_{2}+y_{3}, \ldots, z_{n-1} \leq y_{n-2}+y_{n-1}, z_{n} \leq y_{n-1} \\
y_{1}+y_{2}+\cdots+y_{n-1}=1 \\
z_{1}+z_{2}+\cdots+z_{n}=1 \\
x=z_{1} b_{1}+z_{2} b_{2}+\cdots+z_{n} b_{n} \\
y_{i}=0 \text { or } 1 \quad(i=1,2, \ldots, n-1) ; \quad z_{i} \geq 0 \quad(i=1,2, \ldots, n)
\end{gathered}
$$

## EXAMPLE 7 IP with Piecewise Linear Functions

Euing Gas produces two types of gasoline (gas 1 and gas 2) from two types of oil (oil 1 and oil 2). Each gallon of gas 1 must contain at least 50 percent oil 1, and each gallon of gas 2 must contain at least 60 percent oil 1 . Each gallon of gas 1 can be sold for $12 \phi$, and each gallon of gas 2 can be sold for $14 \not \subset$. Currently, 500 gallons of oil 1 and 1,000 gallons of oil 2 are available. As many as 1,500 more gallons of oil 1 can be purchased at the following prices: first 500 gallons, $25 \phi$ per gallon; next 500 gallons, $20 \phi$ per gallon; next 500 gallons, $15 \phi$ per gallon. Formulate an IP that will maximize Euing's profits (revenues - purchasing costs).

Solution Except for the fact that the cost of purchasing additional oil 1 is a piecewise linear function, this is a straightforward blending problem. With this in mind, we define

$$
\begin{aligned}
x & =\text { amount of oil } 1 \text { purchased } \\
x_{i j} & =\text { amount of oil } i \text { used to produce gas } j \quad(i, j=1,2)
\end{aligned}
$$

Then (in cents)
Total revenue - cost of purchasing oil $1=12\left(x_{11}+x_{21}\right)+14\left(x_{12}+x_{22}\right)-c(x)$
As we have seen previously,

$$
c(x)= \begin{cases}25 x & (0 \leq x \leq 500) \\ 20 x+2,500 & (500 \leq x \leq 1,000) \\ 15 x+7,500 & (1,000 \leq x \leq 1,500)\end{cases}
$$

Thus, Euing's objective function is to maximize

$$
z=12 x_{11}+12 x_{21}+14 x_{12}+14 x_{22}-c(x)
$$

Euing faces the following constraints:
Constraint 1 Euing can use at most $x+500$ gallons of oil 1 .
Constraint 2 Euing can use at most 1,000 gallons of oil 2.
Constraint 3 The oil mixed to make gas 1 must be at least $50 \%$ oil 1 .

Constraint 4 The oil mixed to make gas 2 must be at least $60 \%$ oil 1 .
Constraint 1 yields

$$
x_{11}+x_{12} \leq x+500
$$

Constraint 2 yields

$$
x_{21}+x_{22} \leq 1,000
$$

Constraint 3 yields

$$
\frac{x_{11}}{x_{11}+x_{21}} \geq 0.5 \quad \text { or } \quad 0.5 x_{11}-0.5 x_{21} \geq 0
$$

Constraint 4 yields

$$
\frac{x_{12}}{x_{12}+x_{22}} \geq 0.6 \quad \text { or } \quad 0.4 x_{12}-0.6 x_{22} \geq 0
$$

Also all variables must be nonnegative. Thus, Euing Gas must solve the following optimization problem:

$$
\begin{array}{rlrl}
\max z=12 x_{11}+12 x_{21}+14 x_{12}+14 x_{22} & -c(x) \\
\text { s.t. } & x_{11}+x_{12} & \leq x+500 \\
& & x_{21} & \\
& & \geq 0 \\
0.5 x_{11}-0.5 x_{21} \\
& & \geq 0.4 x_{12}-0.6 x_{22} & \geq 0 \\
& x_{i j} \geq 0,0 \leq x \leq 1,500
\end{array}
$$

Because $c(x)$ is a piecewise linear function, the objective function is not a linear function of $x$, and this optimization is not an LP. By using the method described earlier, however, we can transform this problem into an IP. After recalling that the break points for $c(x)$ are $0,500,1,000$, and 1,500 , we proceed as follows:
Step 1 Replace $c(x)$ by $c(x)=z_{1} c(0)+z_{2} c(500)+z_{3} c(1,000)+z_{4} c(1,500)$.
Step 2 Add the following constraints:

$$
\begin{aligned}
x & =0 z_{1}+500 z_{2}+1,000 z_{3}+1,500 z_{4} \\
z_{1} & \leq y_{1}, z_{2} \leq y_{1}+y_{2}, z_{3} \leq y_{2}+y_{3}, z_{4} \leq y_{3} \\
z_{1} & +z_{2}+z_{3}+z_{4}=1, \quad y_{1}+y_{2}+y_{3}=1 \\
y_{i} & =0 \text { or } 1(i=1,2,3) ; z_{i} \geq 0(i=1,2,3,4)
\end{aligned}
$$

Our new formulation is the following IP:

$$
\begin{align*}
& \max z=12 x_{11}+12 x_{21}+14 x_{12}+14 x_{22}-z_{1} c(0)-z_{2} c(500) \\
& -z_{3} c(1,000)-z_{4} c(1,500) \\
& \begin{array}{lllll}
\text { s.t. } & x_{11} & +\quad x_{12} & & \leq x+500 \\
& x_{21} & + & x_{22} & \leq 1,000
\end{array} \\
& 0.5 x_{11}-0.5 x_{21} \quad \geq 0 \\
& 0.4 x_{12}-0.6 x_{22} \geq 0 \\
& x=0 z_{1}+500 z_{2}+1,000 z_{3}+1,500 z_{4}  \tag{31}\\
& z_{1} \leq y_{1}  \tag{32}\\
& z_{2} \leq y_{1}+y_{2}  \tag{33}\\
& z_{3} \leq y_{2}+y_{3} \tag{34}
\end{align*}
$$

$$
\begin{align*}
& z_{4} \leq y_{3}  \tag{35}\\
& y_{1}+y_{2}+y_{3}=1  \tag{36}\\
& z_{1}+z_{2}+z_{3}+z_{4}=1  \tag{37}\\
& y_{i}=0 \text { or } 1 \quad(i=1,2,3) ; z_{i} \geq 0 \quad(i=1,2,3,4) \\
& x_{i j} \geq 0
\end{align*}
$$

To see why this formulation works, observe that because $y_{1}+y_{2}+y_{3}=1$ and $y_{i}=0$ or 1 , exactly one of the $y_{i}$ 's will equal 1 , and the others will equal 0 . Now, (32)-(37) imply that if $y_{i}=1$, then $z_{i}$ and $z_{i+1}$ may be positive, but all the other $z_{i}$ 's must equal 0 . For instance, if $y_{2}=1$, then $y_{1}=y_{3}=0$. Then (32)-(35) become $z_{1} \leq 0, z_{2} \leq 1, z_{3} \leq 1$, and $z_{4} \leq 0$. These constraints force $z_{1}=z_{4}=0$ and allow $z_{2}$ and $z_{3}$ to be any nonnegative number less than or equal to 1 . We can now show that (31)-(37) correctly represent the piecewise linear function $c(x)$. Choose any value of $x$, say $x=800$. Note that $b_{2}=500 \leq$ $800 \leq 1,000=b_{3}$. For $x=800$, what values do our constraints assign to $y_{1}, y_{2}$, and $y_{3}$ ? The value $y_{1}=1$ is impossible, because if $y_{1}=1$, then $y_{2}=y_{3}=0$. Then (34)-(35) force $z_{3}=z_{4}=0$. Then (31) reduces to $800=x=500 z_{2}$, which cannot be satisfied by $z_{2} \leq 1$. Similarly, $y_{3}=1$ is impossible. If we try $y_{2}=1$ (32) and (35) force $z_{1}=z_{4}=0$. Then (33) and (34) imply $z_{2} \leq 1$ and $z_{3} \leq 1$. Now (31) becomes $800=x=500 z_{2}+1,000 z_{3}$. Because $z_{2}+z_{3}=1$, we obtain $z_{2}=\frac{2}{5}$ and $z_{3}=\frac{3}{5}$. Now the objective function reduces to

$$
12 x_{11}+12 x_{21}+14 x_{21}+14 x_{22}-\frac{2 c(500)}{5}-\frac{3 c(1,000)}{5}
$$

Because

$$
c(800)=\frac{2 c(500)}{5}+\frac{3 c(1,000)}{5}
$$

our objective function yields the correct value of Euing's profits!
The optimal solution to Euing's problem is $z=12,500, x=1,000, x_{12}=1,500$, $x_{22}=1,000, y_{3}=z_{3}=1$. Thus, Euing should purchase 1,000 gallons of oil 1 and produce 2,500 gallons of gas 2 .

In general, constraints of the form (31)-(37) ensure that if $b_{i} \leq x \leq b_{i+1}$, then $y_{i}=1$ and only $z_{i}$ and $z_{i+1}$ can be positive. Because $c(x)$ is linear for $b_{i} \leq x \leq b_{i+1}$, the objective function will assign the correct value to $c(x)$.

If a piecewise linear function $f(x)$ involved in a formulation has the property that the slope of $f(x)$ becomes less favorable to the decision maker as $x$ increases, then the tedious IP formulation we have just described is unnecessary.

## EXAMPLE 8 Media Selection with Piecewise Linear Functions

Dorian Auto has a $\$ 20,000$ advertising budget. Dorian can purchase full-page ads in two magazines: Inside Jocks (IJ) and Family Square (FS). An exposure occurs when a person reads a Dorian Auto ad for the first time. The number of exposures generated by each ad in IJ is as follows: ads $1-6,10,000$ exposures; ads $7-10,3,000$ exposures; ads $11-15,2,500$ exposures; ads $16+, 0$ exposures. For example, 8 ads in IJ would generate $6(10,000)+2(3,000)=66,000$ exposures. The number of exposures generated by each ad in FS is as follows: ads $1-4,8,000$ exposures; ads $5-12,6,000$ exposures; ads 13-15, 2,000 exposures; ads $16+, 0$ exposures. Thus, 13 ads in FS would generate $4(8,000)+$
$8(6,000)+1(2,000)=82,000$ exposures. Each full-page ad in either magazine costs $\$ 1,000$. Assume there is no overlap in the readership of the two magazines. Formulate an IP to maximize the number of exposures that Dorian can obtain with limited advertising funds.

Solution If we define

$$
\begin{aligned}
& x_{1}=\text { number of IJ ads yielding } 10,000 \text { exposures } \\
& x_{2}=\text { number of IJ ads yielding } 3,000 \text { exposures } \\
& x_{3}=\text { number of IJ ads yielding } 2,500 \text { exposures } \\
& y_{1}=\text { number of FS ads yielding } 8,000 \text { exposures } \\
& y_{2}=\text { number of FS ads yielding } 6,000 \text { exposures } \\
& y_{3}=\text { number of FS ads yielding } 2,000 \text { exposures }
\end{aligned}
$$

then the total number of exposures (in thousands) is given by

$$
10 x_{1}+3 x_{2}+2.5 x_{3}+8 y_{1}+6 y_{2}+2 y_{3}
$$

Thus, Dorian wants to maximize

$$
z=10 x_{1}+3 x_{2}+2.5 x_{3}+8 y_{1}+6 y_{2}+2 y_{3}
$$

Because the total amount spent (in thousands) is just the toal number of ads placed in both magazines, Dorian's budget constraint may be written as

$$
x_{1}+x_{2}+x_{3}+y_{1}+y_{2}+y_{3} \leq 20
$$

The statement of the problem implies that $x_{1} \leq 6, x_{2} \leq 4, x_{3} \leq 5, y_{1} \leq 4, y_{2} \leq 8$, and $y_{3} \leq 3$ all must hold. Adding the sign restrictions on each variable and noting that each variable must be an integer, we obtain the following IP:

$$
\begin{aligned}
& \max z=10 x_{1}+3 x_{2}+2.5 x_{3}+8 y_{1}+6 y_{2}+2 y_{3} \\
& \text { s.t. } x_{1}+x_{2}+x_{3}+y_{1}+y_{2}+y_{3} \leq 20 \\
& x_{1} \quad \leq 6 \\
& x_{2} \quad \leq 4 \\
& x_{3} \quad \leq 5 \\
& y_{1} \quad \leq 4 \\
& y_{2} \leq 8 \\
& y_{3} \leq 3 \\
& x_{i}, y_{i} \text { integer }(i=1,2,3) \\
& x_{i}, y_{i} \geq 0 \quad(i=1,2,3)
\end{aligned}
$$

Observe that the statement of the problem implies that $x_{2}$ cannot be positive unless $x_{1}$ assumes its maximum value of 6 . Similarly, $x_{3}$ cannot be positive unless $x_{2}$ assumes its maximum value of 4 . Because $x_{1}$ ads generate more exposures than $x_{2}$ ads, however, the act of maximizing ensures that $x_{2}$ will be positive only if $x_{1}$ has been made as large as possible. Similarly, because $x_{3}$ ads generate fewer exposures than $x_{2}$ ads, $x_{3}$ will be positive only if $x_{2}$ assumes its maximum value. (Also, $y_{2}$ will be positive only if $y_{1}=4$, and $y_{3}$ will be positive only if $y_{2}=8$.)

The optimal solution to Dorian's IP is $z=146,000, x_{1}=6, x_{2}=2, y_{1}=4, y_{2}=8$, $x_{3}=0, y_{3}=0$. Thus, Dorian will place $x_{1}+x_{2}=8$ ads in IJ and $y_{1}+y_{2}=12$ ads in FS.

In Example 8, additional advertising in a magazine yielded diminishing returns. This ensured that $x_{i}\left(y_{i}\right)$ would be positive only if $x_{i-1}\left(y_{i-1}\right)$ assumed its maximum value. If additional advertising generated increasing returns, then this formulation would not yield the correct solution. For example, suppose that the number of exposures generated by each IJ ad was as follows: ads $1-6,2,500$ exposures; ads $7-10,3,000$ exposures; ads $11-15,10,000$ exposures. Suppose also that the number of exposures generated by each FS is as follows: ads $1-4,2,000$ exposures; ads 5-12, 6,000 exposures; ads 13-15, 8,000 exposures.

If we define

$$
\begin{aligned}
& x_{1}=\text { number of IJ ads generating 2,500 exposures } \\
& x_{2}=\text { number of IJ ads generating } 3,000 \text { exposures } \\
& x_{3}=\text { number of IJ ads generating } 10,000 \text { exposures } \\
& y_{1}=\text { number of FS ads generating } 2,000 \text { exposures } \\
& y_{2}=\text { number of FS ads generating } 6,000 \text { exposures } \\
& y_{3}=\text { number of FS ads generating } 8,000 \text { exposures }
\end{aligned}
$$

the reasoning used in the previous example would lead to the following formulation:

$$
\begin{aligned}
& \max z=2.5 x_{1}+3 x_{2}+10 x_{3}+2 y_{1}+6 y_{2}+8 y_{3} \\
& \text { s.t. } x_{1}+x_{2}+x_{3}+y_{1}+y_{2}+y_{3} \leq 20 \\
& x_{1} \quad \leq 6 \\
& x_{2} \quad \leq 4 \\
& x_{3} \quad \leq 5 \\
& y_{1} \quad \leq 4 \\
& y_{2} \leq 8 \\
& y_{3} \leq 3 \\
& x_{i}, y_{i} \text { integer } \quad(i=1,2,3) \\
& x_{i}, y_{i} \leq 0 \quad(i=1,2,3)
\end{aligned}
$$

The optimal solution to this IP is $x_{3}=5, y_{3}=3, y_{2}=8, x_{2}=4, x_{1}=0, y_{1}=0$, which cannot be correct. According to this solution, $x_{1}+x_{2}+x_{3}=9$ ads should be placed in IJ. If 9 ads were placed in IJ, however, then it must be that $x_{1}=6$ and $x_{2}=3$. Therefore, we see that the type of formulation used in the Dorian Auto example is correct only if the piecewise linear objective function has a less favorable slope for larger values of $x$. In our second example, the effectiveness of an ad increased as the number of ads in a magazine increased, and the act of maximizing will not ensure that $x_{i}$ can be positive only if $x_{i-1}$ assumes its maximum value. In this case, the approach used in the Euing Gas example would yield a correct formulation (see Problem 8).

## Solving IPs with LINDO

LINDO can be used to solve pure or mixed IPs. In addition to the optimal solution, the LINDO output for an IP gives shadow prices and reduced costs. Unfortunately, the shadow prices and reduced costs refer to subproblems generated during the branch-andbound solution - not to the IP. Unlike linear programming, there is no well-developed theory of sensitivity analysis for integer programming. The reader interested in a discussion of sensitivity analysis for IPs should consult Williams (1985).

TABLE 54

| Product | Demand | Unit Profit <br> Contribution (\$) | Fixed Charge (\$) |
| :--- | :---: | :---: | :---: |
| 1 | 40 | 2 | 30 |
| 2 | 60 | 5 | 40 |
| 3 | 65 | 6 | 50 |
| 4 | 70 | 7 | 60 |

TABLE 55

|  | Product |  |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Resource Usage | 1 | 2 | 3 | 4 |
| 1 | 1 | 2 | 3.5 | 4 |
| 2 | 5 | 6 | 7 | 9 |
| 3 | 3 | 4 | 5 | 6 |

### 9.3 The Branch-and-Bound Method for Solving Pure Integer Programming Problems

In practice, most IPs are solved by using the technique of branch-and-bound. Branch-andbound methods find the optimal solution to an IP by efficiently enumerating the points in a subproblem's feasible region. Before explaining how branch-and-bound works, we need to make the following elementary but important observation: If you solve the LP relaxation of a pure IP and obtain a solution in which all variables are integers, then the optimal solution to the LP relaxation is also the optimal solution to the IP.

To see why this observation is true, consider the following IP:

$$
\begin{gathered}
\max z=3 x_{1}+2 x_{2} \\
\text { s.t. } \quad 2 x_{1}+x_{2} \leq 6 \\
x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{gathered}
$$

The optimal solution to the LP relaxation of this pure IP is $x_{1}=0, x_{2}=6, z=12$. Because this solution gives integer values to all variables, the preceding observation implies that $x_{1}=0, x_{2}=6, z=12$ is also the optimal solution to the IP. Observe that the feasible region for the IP is a subset of the points in the LP relaxation's feasible region (see Figure 10). Thus, the optimal $z$-value for the IP cannot be larger than the optimal $z$-value for the LP relaxation. This means that the optimal $z$-value for the IP must be $\leq 12$. But the point $x_{1}=0, x_{2}=6, z=12$ is feasible for the IP and has $z=12$. Thus, $x_{1}=0$, $x_{2}=6, z=12$ must be optimal for the IP.

FIGURE 10 Feasible Region for an IP and Its LP Relaxation


The Telfa Corporation manufactures tables and chairs. A table requires 1 hour of labor and 9 square board feet of wood, and a chair requires 1 hour of labor and 5 square board feet of wood. Currently, 6 hours of labor and 45 square board feet of wood are available. Each table contributes $\$ 8$ to profit, and each chair contributes $\$ 5$ to profit. Formulate and solve an IP to maximize Telfa's profit.

## Solution Let

$$
\begin{aligned}
& x_{1}=\text { number of tables manufactured } \\
& x_{2}=\text { number of chairs manufactured }
\end{aligned}
$$

Because $x_{1}$ and $x_{2}$ must be integers, Telfa wants to solve the following IP:

$$
\begin{array}{lcl}
\max z= & 8 x_{1}+5 x_{2} & \\
\text { s.t. } & x_{1}+x_{2} \leq 6 & (\text { Labor constraint }) \\
& 9 x_{1}+5 x_{2} \leq 45 & (\text { Wood constraint }) \\
& x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{array}
$$

The branch-and-bound method begins by solving the LP relaxation of the IP. If all the decision variables assume integer values in the optimal solution to the LP relaxation, then the optimal solution to the LP relaxation will be the optimal solution to the IP. We call the LP relaxation subproblem 1. Unfortunately, the optimal solution to the LP relaxation is $z=\frac{165}{4}, x_{1}=\frac{15}{4}, x_{2}=\frac{9}{4}$ (see Figure 11). From Section 9.1, we know that (optimal $z$-value for IP $) \leq$ (optimal $z$-value for LP relaxation). This implies that the optimal $z$-value for the IP cannot exceed $\frac{165}{4}$. Thus, the optimal $z$-value for the LP relaxation is an upper bound for Telfa's profit.

Our next step is to partition the feasible region for the LP relaxation in an attempt to find out more about the location of the IP's optimal solution. We arbitrarily choose a variable that is fractional in the optimal solution to the LP relaxation-say, $x_{1}$. Now observe that every point in the feasible region for the IP must have either $x_{1} \leq 3$ or $x_{1} \geq 4$. (Why can't a feasible solution to the IP have $3<x_{1}<4$ ?) With this in mind, we "branch" on the variable $x_{1}$ and create the following two additional subproblems:

FIGURE 11 Feasible Region for Telfa Problem


FIGURE 12 Feasible Region for Subproblems 2 and 3 of Telfa Problem

Subproblem 2 Subproblem $1+$ Constraint $x_{1} \geq 4$.
Subproblem 3 Subproblem $1+$ Constraint $x_{1} \leq 3$.
Observe that neither subproblem 2 nor subproblem 3 includes any points with $x_{1}=\frac{15}{4}$. This means that the optimal solution to the LP relaxation cannot recur when we solve subproblem 2 or subproblem 3.

From Figure 12, we see that every point in the feasible region for the Telfa IP is included in the feasible region for subproblem 2 or subproblem 3. Also, the feasible regions for subproblems 2 and 3 have no points in common. Because subproblems 2 and 3 were created by adding constraints involving $x_{1}$, we say that subproblems 2 and 3 were created by branching on $x_{1}$.

We now choose any subproblem that has not yet been solved as an LP. We arbitrarily choose to solve subproblem 2. From Figure 12, we see that the optimal solution to subproblem 2 is $z=41, x_{1}=4, x_{2}=\frac{9}{5}$ (point $C$ ). Our accomplishments to date are summarized in Figure 13.

A display of all subproblems that have been created is called a tree. Each subproblem is referred to as a node of the tree, and each line connecting two nodes of the tree is called an arc. The constraints associated with any node of the tree are the constraints for the LP relaxation plus the constraints associated with the arcs leading from subproblem 1 to the node. The label $t$ indicates the chronological order in which the subproblems are solved.


The optimal solution to subproblem 2 did not yield an all-integer solution, so we choose to use subproblem 2 to create two new subproblems. We choose a fractionalvalued variable in the optimal solution to subproblem 2 and then branch on that variable. Because $x_{2}$ is the only fractional variable in the optimal solution to subproblem 2, we branch on $x_{2}$. We partition the feasible region for subproblem 2 into those points having $x_{2} \geq 2$ and $x_{2} \leq 1$. This creates the following two subproblems:

Subproblem 4 Subproblem $1+$ Constraints $x_{1} \geq 4$ and $x_{2} \geq 2=$ subproblem $2+$ Constraint $x_{2} \geq 2$.

Subproblem 5 Subproblem $1+$ Constraints $x_{1} \geq 4$ and $x_{2} \leq 1=$ subproblem $2+$ Constraint $x_{2} \leq 1$.

The feasible regions for subproblems 4 and 5 are displayed in Figure 14. The set of unsolved subproblems consists of subproblems 3,4 , and 5 . We now choose a subproblem to solve. For reasons that are discussed later, we choose to solve the most recently created subproblem. (This is called the LIFO, or last-in-first-out, rule.) The LIFO rule implies that we should next solve subproblem 4 or subproblem 5 . We arbitrarily choose to solve subproblem 4. From Figure 14 we see that subproblem 4 is infeasible. Thus, subproblem 4 cannot yield the optimal solution to the IP. To indicate this fact, we place an $\times$ by subproblem 4 (see Figure 15). Because any branches emanating from subproblem 4 will yield no useful information, it is fruitless to create them. When further branching on a subproblem cannot yield any useful information, we say that the subproblem (or node) is fathomed. Our results to date are displayed in Figure 15.

Now the only unsolved subproblems are subproblems 3 and 5. The LIFO rule implies that subproblem 5 should be solved next. From Figure 14, we see that the optimal solution to subproblem 5 is point $I$ in Figure 14: $z=\frac{365}{9}, x_{1}=\frac{40}{9}, x_{2}=1$. This solution does not yield any immediately useful information, so we choose to partition subproblem 5's feasible region by branching on the fractional-valued variable $x_{1}$. This yields two new subproblems (see Figure 16).

Subproblem 6 Subproblem $5+$ Constraint $x_{1} \geq 5$.
Subproblem 7 Subproblem $5+$ Constraint $x_{1} \leq 4$.

FIGURE 14 Feasible Regions for Subproblems 4 and 5 of Telfa Problem


Together, subproblems 6 and 7 include all integer points that were included in the feasible region for subproblem 5. Also, no point having $x_{1}=\frac{40}{9}$ can be in the feasible region for subproblem 6 or subproblem 7. Thus, the optimal solution to subproblem 5 will not recur when we solve subproblems 6 and 7. Our tree now looks as shown in Figure 17.

Subproblems 3, 6, and 7 are now unsolved. The LIFO rule implies that we next solve subproblem 6 or subproblem 7. We arbitrarily choose to solve subproblem 7. From Figure 16, we see that the optimal solution to subproblem 7 is point $H: z=37, x_{1}=4$, $x_{2}=1$. Both $x_{1}$ and $x_{2}$ assume integer values, so this solution is feasible for the original IP. We now know that subproblem 7 yields a feasible integer solution with $z=37$. We also know that subproblem 7 cannot yield a feasible integer solution having $z>37$. Thus, further branching on subproblem 7 will yield no new information about the optimal solution to the IP, and subproblem has been fathomed. The tree to date is pictured in Figure 18.

FIGURE 17
Telfa Subproblems 1, 2, 4, and 5 Solved


FIGURE 19 Branch-and-Bound Tree After Six Subproblems Have Been Solved

FIGURE 20 Final Branch-and-Bound Tree for Telfa Problem


A solution obtained by solving a subproblem in which all variables have integer values is a candidate solution. Because the candidate solution may be optimal, we must keep a candidate solution until a better feasible solution to the IP (if any exists) is found. We have a feasible solution to the original IP with $z=37$, so we may conclude that the optimal $z$-value for the IP $\geq 37$. Thus, the $z$-value for the candidate solution is a lower bound on the optimal $z$-value for the original IP. We note this by placing the notation $\mathrm{LB}=37$ in the box corresponding to the next solved subproblem (see Figure 19).

The only remaining unsolved subproblems are 6 and 3. Following the LIFO rule, we next solve subproblem 6. From Figure 16, we find that the optimal solution to subproblem 6 is point $A: z=40, x_{1}=5, x_{2}=0$. All decision variables have integer values, so this is a candidate solution. Its $z$-value of 40 is larger than the $z$-value of the best previous candidate (candidate 7 with $z=37$ ). Thus, subproblem 7 cannot yield the optimal solution of the IP (we denote this fact by placing an $\times$ by subproblem 7 ). We also update our LB to 40 . Our progress to date is summarized in Figure 20.

Subproblem 3 is the only remaining unsolved problem. From Figure 12, we find that the optimal solution to subproblem 3 is point $F: z=39, x_{1}=x_{2}=3$. Subproblem 3 cannot yield a $z$-value exceeding the current lower bound of 40 , so it cannot yield the optimal solution to the original IP. Therefore, we place an $\times$ by it in Figure 20. From Figure 20, we see that there are no remaining unsolved subproblems, and that only subproblem 6 can yield the optimal solution to the IP. Thus, the optimal solution to the IP is for Telfa to manufacture 5 tables and 0 chairs. This solution will contribute $\$ 40$ to profits.

In using the branch-and-bound method to solve the Telfa problem, we have implicitly enumerated all points in the IP's feasible region. Eventually, all such points (except for the optimal solution) are eliminated from consideration, and the branch-and-bound procedure is complete. To show that the branch-and-bound procedure actually does consider all points in the IP's feasible region, we examine several possible solutions to the Telfa problem and show how the procedure found these points to be nonoptimal. For example, how do we know that $x_{1}=2, x_{2}=3$ is not optimal? This point is in the feasible region for subproblem 3, and we know that all points in the feasible region for subproblem 3 have $z \leq 39$. Thus, our analysis of subproblem 3 shows that $x_{1}=2, x_{2}=3$ cannot beat $z=40$ and cannot be optimal. As another example, why isn't $x_{1}=4, x_{2}=2$ optimal? Following the branches of the tree, we find that $x_{1}=4, x_{2}=2$ is associated with subproblem 4. Because no point associated with subproblem 4 is feasible, $x_{1}=4, x_{2}=2$ must fail to satisfy the constraints for the original IP and thus cannot be optimal for the Telfa problem. In a similar fashion, the branch-and-bound analysis has eliminated all points $x_{1}, x_{2}$ (except for the optimal solution) from consideration.

For the simple Telfa problem, the use of the branch-and-bound method may seem like using a cannon to kill a fly, but for an IP in which the feasible region contains a large number of integer points, the procedure can be very efficient for eliminating nonoptimal points from consideration. For example, suppose we are applying the branch-and-bound method and our current LB $=42$. Suppose we solve a subproblem that contains 1 million feasible points for the IP. If the optimal solution to this subproblem has $z<42$, then we have eliminated 1 million nonoptimal points by solving a single LP!

The key aspects of the branch-and-bound method for solving pure IPs (mixed IPs are considered in the next section) may be summarized as follows:

Step 1 If it is unnecessary to branch on a subproblem, then it is fathomed. The following three situations result in a subproblem being fathomed: (1) The subproblem is infeasible; (2) the subproblem yields an optimal solution in which all variables have integer values; and (3) the optimal $z$-value for the subproblem does not exceed (in a max problem) the current LB.

Step 2 A subproblem may be eliminated from consideration in the following situations: (1) The subproblem is infeasible (in the Telfa problem, subproblem 4 was eliminated for this reason); (2) the LB (representing the $z$-value of the best candidate to date) is at least as large as the $z$-value for the subproblem (in the Telfa problem, subproblems 3 and 7 were eliminated for this reason).

Recall that in solving the Telfa problem by the branch-and-bound procedure, many seemingly arbitrary choices were made. For example, when $x_{1}$ and $x_{2}$ were both fractional in the optimal solution to subproblem 1, how did we determine the branching variable? Or how did we determine which subproblem should next be solved? The manner in which these questions are answered can result in trees that differ greatly in size and in the computer time required to find an optimal solution. Through experience and ingenuity, practitioners of the procedure have developed guidelines on how to make the necessary decisions.

Two general approaches are commonly used to determine which subproblems should be solved next. The most widely used is the LIFO rule, which chooses to solve the most recently created subproblem. ${ }^{\dagger}$ LIFO leads us down one side of the branch-and-bound tree (as in the Telfa problem) and quickly finds a candidate solution. Then we backtrack our way up to the top of the other side of the tree. For this reason, the LIFO approach is often called backtracking.

The second commonly used method is jumptracking. When branching on a node, the jumptracking approach solves all the problems created by the branching. Then it branches again on the node with the best $z$-value. Jumptracking often jumps from one side of the tree to the other. It usually creates more subproblems and requires more computer storage than backtracking. The idea behind jumptracking is that moving toward the subproblems with good $z$-values should lead us more quickly to the best $z$-value.

If two or more variables are fractional in a subproblem's optimal solution, then on which variable should we branch? Branching on the fractional-valued variable that has the greatest economic importance is often the best strategy. In the Nickles example, suppose the optimal solution to a subproblem had $y_{1}$ and $x_{12}$ fractional. Our rule would say to branch on $y_{1}$ because $y_{1}$ represents the decision to operate (or not operate) a lockbox in city 1 , and this is presumably a more important decision than whether region 1 payments should be sent to city 2 . When more than one variable is fractional in a subproblem solution, many computer codes will branch on the lowest-numbered fractional variable. Thus, if an integer programming computer code requires that variables be numbered, they should be numbered in order of their economic importance $(1=$ most important $)$.

## REMARKS

1 For some IP's, the optimal solution to the LP relaxation will also be the optimal solution to the IP. Suppose the constraints of the IP are written as $A \mathbf{x}=\mathbf{b}$. If the determinant of every square submatrix of $A$ is $+1,-1$, or 0 , we say that the matrix $A$ is unimodular. If $A$ is unimodular and each element of $\mathbf{b}$ is an integer, then the optimal solution to the LP relaxation will assign all variables integer values [see Shapiro (1979) for a proof] and will therefore be the optimal solution to the IP. It can be shown that the constraint matrix of any MCNFP is unimodular. Thus, as was discussed in Chapter 8, any MCNFP in which each node's net outflow and each arc's capacity are integers will have an integer-valued solution.
2 As a general rule, the more an IP looks like an MCNFP, the easier the problem is to solve by branch-and-bound methods. Thus, in formulating an IP, it is good to choose a formulation in which as many variables as possible have coefficients of $+1,-1$, and 0 . To illustrate this idea, recall that the formulation of the Nickles (lockbox) problem given in Section 9.2 contained 16 constraints of the following form:
Formulation 1

$$
\begin{equation*}
x_{i j} \leq y_{j}(i=1,2,3,4 ; j=1,2,3,4) \tag{25}
\end{equation*}
$$

[^14]As we have already seen in Section 9.2, if the 16 constraints in (25) are replaced by the following 4 constraints, then an equivalent formulation results:

## Formulation 2

$$
\begin{aligned}
& x_{11}+x_{21}+x_{31}+x_{41} \leq 4 y_{1} \\
& x_{12}+x_{22}+x_{32}+x_{42} \leq 4 y_{2} \\
& x_{13}+x_{23}+x_{33}+x_{43} \leq 4 y_{3} \\
& x_{14}+x_{24}+x_{34}+x_{44} \leq 4 y_{4}
\end{aligned}
$$

Because formulation 2 has $16-4=12$ fewer constraints than formulation 1, one might think that formulation 2 would require less computer time to find the optimal solution. This turns out to be untrue. To see why, recall that the branch-and-bound method begins by solving the LP relaxation of the IP. The feasible region of the LP relaxation of formulation 2 contains many more noninteger points than the feasible region of formulation 1. For example, the point $y_{1}=y_{2}=y_{3}=y_{4}$ $=\frac{1}{4}, x_{11}=x_{22}=x_{33}=x_{44}=1$ (all other $x_{i j}$ 's equal 0 ) is in the feasible region for the LP relaxation of formulation 2, but not for formulation 1. The branch-and-bound method must eliminate all noninteger points before obtaining the optimal solution to the IP, so it seems reasonable that formulation 2 will require more computer time than formulation 1. Indeed, when the LINDO package was used to find the optimal solution to formulation 1, the LP relaxation yielded the optimal solution. But 17 subproblems were solved before the optimal solution was found for formulation 2. Note that formulation 2 contains the terms $4 y_{1}, 4 y_{2}, 4 y_{3}$, and $4 y_{4}$. These terms "disturb" the network-like structure of the lockbox problem and cause the branch-and-bound method to be less efficient.
3 When solving an IP in the real world, we are usually happy with a near-optimal solution. For example, suppose that we are solving a lockbox problem and the LP relaxation yields a cost of $\$ 200,000$. This means that the optimal solution to the lockbox IP will certainly have a cost of at least $\$ 200,000$. If we find a candidate solution during the course of the branch-and-bound procedure that has a cost of, say, $\$ 205,000$, why bother to continue with the branch-and-bound procedure? Even if we found the optimal solution to the IP, it could not save more than $\$ 5,000$ in costs over the candidate solution with $z=205,000$. It might even cost more than $\$ 5,000$ in computer time to find the optimal lockbox solution. For this reason, the branch-and-bound procedure is often terminated when a candidate solution is found with a $z$-value close to the $z$-value of the LP relaxation.
4 Subproblems for branch-and-bound problems are often solved using some variant of the dual simplex algorithm. To illustrate this, we return to the Telfa example. The optimal tableau for the LP relaxation of the Telfa problem is

$$
\begin{aligned}
z \quad+1.25 s_{1}+0.75 s_{2} & =41.25 \\
x_{2}+2.25 s_{1}-0.25 s_{2} & =2.25 \\
x_{1}-1.25 s_{1}+0.25 s_{2} & =3.75
\end{aligned}
$$

After solving the LP relaxation, we solved subproblem 2, which is just subproblem 1 plus the constraint $x_{1} \geq 4$. Recall that the dual simplex is an efficient method for finding the new optimal solution to an LP when we know the optimal tableau and a new constraint is added to the LP. We have added the constraint $x_{1} \geq 4$ (which may be written as $x_{1}-e_{3}=4$ ). To utilize the dual simplex, we must eliminate the basic variable $x_{1}$ from this constraint and use $e_{3}$ as a basic variable for $x_{1}-$ $e_{3}=4$. Adding - (second row of optimal tableau) to the constraint $x_{1}-e_{3}=4$, we obtain the constraint $1.25 s_{1}-0.25 s_{2}-e_{3}=0.25$. Multiplying this constraint through by -1 , we obtain $-1.25 s_{1}+0.25 s_{2}+e_{3}=-0.25$. After adding this constraint to subproblem 1's optimal tableau, we obtain the tableau in Table 56. The dual simplex method states that we should enter a variable from row 3 into the basis. Because $s_{1}$ is the only variable with a negative coefficient in row $3, s_{1}$ will enter the basis in row 3. After the pivot, we obtain the (optimal) tableau in Table 57. Thus, the optimal solution to subproblem 2 is $z=41, x_{2}=1.8, x_{1}=4, s_{1}=0.20$.
table 56 Initial Tableau for Solving Subproblem 2 by Dual Simplex

|  |  |  | Basic Variable |
| ---: | :--- | :--- | :--- |
| $z \quad+1.25 s_{1}+0.75 s_{2}$ | $=41.25$ | $z=41.25$ |  |
| $x_{2}+2.25 s_{1}-0.25 s_{2}$ | $=2.25$ | $x_{2}=2.25$ |  |
| $x_{1}-1.25 s_{1}+0.25 s_{2}$ | $=3.75$ | $x_{1}=3.75$ |  |
|  | $-1.25 s_{1}+0.25 s_{2}+e_{3}$ | $=-0.25$ | $e_{3}=-0.25$ |

TABLE 57
Optimal Tableau for Solving Subproblem 2 by Dual Simplex

|  |  |  | Basic Variable |
| ---: | ---: | ---: | :--- |
| $z$ | $+\quad s_{2}+\quad e_{3}=41$ | $z=41$ |  |
|  | $x_{2}+0.20 s_{2}+1.8 e_{3}=1.8$ |  | $x_{2}=1.8$ |
| $x_{1}$ | $-e_{3}=4$ |  | $x_{1}=4$ |
|  | $s_{1}-0.20 s_{2}-0.80 e_{3}=0.20$ | $s_{1}=0.20$ |  |

5 In Problem 8, we show that if we create two subproblems by adding the constraints $x_{k} \leq i$ and $x_{k} \geq i+1$, then the optimal solution to the first subproblem will have $x_{k}=i$ and the optimal solution to the second subproblem will have $x_{k}=i+1$. This observation is very helpful when we graphically solve subproblems. For example, we know the optimal solution to subproblem 5 of Example 9 will have $x_{2}=1$. Then we can find the value of $x_{1}$ that solves subproblem 5 by choosing $x_{1}$ to be the largest integer satisfying all constraints when $x_{2}=1$.

## Solver Tolerance Option for Solving IPs

When solving integer programming problems with the Excel Solver, you may go to Options and set a tolerance. A tolerance value of, say, .20 , causes the Excel Solver to stop when a feasible solution is found that has an objective function value within $20 \%$ of the optimal $z$-value for the problem's LP relaxation. For instance, in Example 9, the optimal $z$-value for the LP relaxation was 41.25 . With a tolerance of .20 , the Solver would stop whenever a feasible integer solution is found with a $z$-value exceeding $(1-.2)(41.25)=33$. Thus, if we solved Example 9 with the Excel Solver and found a feasible integer solution having $z=35$, then the Solver would stop because this solution would be within $20 \%$ of the LP relaxation bound.

Why set a nonzero tolerance? For many large IP problems, it might take a long time (weeks or months!) to find an optimal solution. It might take much less time to find a near-optimal solution (say, within $5 \%$ of the optimal LP relaxation). In this case, we would be much better off with a near-optimal solution, and use of the tolerance option might be appropriate.

## PROBLEMS

## Group A

Use branch-and-bound to solve the following IPs:

1

2 The Dorian Auto example of Section 3.2.
$3 \quad \max z=2 x_{1}+3 x_{2}$

$$
\text { s.t. } \quad \begin{aligned}
x_{1}+2 x_{2} & \leq 10 \\
3 x_{1}+4 x_{2} & \leq 25 \\
x_{1}, x_{2} & \geq 0 ; x_{1}, x_{2} \text { integer }
\end{aligned}
$$

$4 \quad \max z=4 x_{1}+3 x_{2}$

$$
\text { s.t. } 4 x_{1}+9 x_{2} \leq 26
$$

$$
8 x_{1}+5 x_{2} \leq 17
$$

$$
x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
$$

$$
\begin{aligned}
& \max z=5 x_{1}+2 x_{2} \\
& \text { s.t. } \quad 3 x_{1}+x_{2} \leq 12 \\
& x_{1}+x_{2} \leq 5 \\
& x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{aligned}
$$

5

$$
\begin{array}{lr}
\max z=4 x_{1}+5 x_{2} \\
\text { s.t. } & x_{1}+4 x_{2} \geq 5 \\
3 x_{1}+2 x_{2} & \geq 7 \\
& x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{array}
$$

$6 \quad \max z=4 x_{1}+5 x_{2}$

$$
\begin{aligned}
\text { s.t. } \quad 3 x_{1}+2 x_{2} & \leq 10 \\
x_{1}+4 x_{2} & \leq 11 \\
3 x_{1}+3 x_{2} & \leq 13 \\
x_{1}, x_{2} & \geq 0 ; x_{1}, x_{2} \text { integer }
\end{aligned}
$$

7 Use the branch-and-bound method to find the optimal solution to the following IP:

$$
\begin{aligned}
& \max z=7 x_{1}+3 x_{2} \\
& \text { s.t. } \quad 2 x_{1}+x_{2} \leq 9 \\
& 3 x_{1}+2 x_{2} \leq 13 \\
& x_{1}, x_{2} \geq 0 ; x_{1}, x_{2} \text { integer }
\end{aligned}
$$

## Nonlinear Programming

In previous chapters, we have studied linear programming problems. For an LP, our goal was to maximize or minimize a linear function subject to linear constraints. But in many interesting maximization and minimization problems, the objective function may not be a linear function, or some of the constraints may not be linear constraints. Such an optimization problem is called a nonlinear programming problem (NLP). In this chapter, we discuss techniques used to solve NLPs.

We begin with a review of material from differential calculus, which will be needed for our study of nonlinear programming.

### 11.1 Review of Differential Calculus

## Limits

The idea of a limit is one of the most basic ideas in calculus.

DEFINITION ■ The equation

$$
\lim _{x \rightarrow a} f(x)=c
$$

means that as $x$ gets closer to $a$ (but not equal to $a$ ), the value of $f(x)$ gets arbitrarily close to $c$.

It is also possible that $\lim _{x \rightarrow a} f(x)$ may not exist.

## EXAMPLE 1 Limits

1 Show that $\lim _{x \rightarrow 2} x^{2}-2 x=2^{2}-2(2)=0$.
2 Show that $\lim _{x \rightarrow 0} \frac{1}{x}$ does not exist.
Solution 1 To verify this result, evaluate $x^{2}-2 x$ for values of $x$ close to, but not equal to, 2 .
2 To verify this result, observe that as $x$ gets near $0, \frac{1}{x}$ becomes either a very large positive number or a very large negative number. Thus, as $x$ approaches $0, \frac{1}{x}$ will not approach any single number.

## Continuity

DEFINITION - A function $f(x)$ is continuous at a point $a$ if

$$
\lim _{x \rightarrow a} f(x)=f(a)
$$

If $f(x)$ is not continuous at $x=a$, we say that $f(x)$ is discontinuous (or has a discontinuity) at $a$.

## EXAMPLE 2 Continuous Functions

Bakeco orders sugar from Sugarco. The per-pound purchase price of the sugar depends on the size of the order (see Table 1). Let

$$
\begin{aligned}
x & =\text { number of pounds of sugar purchased by Bakeco } \\
f(x) & =\text { cost of ordering } x \text { pounds of sugar }
\end{aligned}
$$

Then

$$
\begin{aligned}
& f(x)=25 x \text { for } 0 \leq x<100 \\
& f(x)=20 x \text { for } 100 \leq x \leq 200 \\
& f(x)=15 x \text { for } x>200
\end{aligned}
$$

For all values of $x$, determine if $x$ is continuous or discontinuous.
From Figure 1, it is clear that

$$
\lim _{x \rightarrow 100} f(x) \quad \text { and } \quad \lim _{x \rightarrow 200} f(x)
$$

do not exist. Thus, $f(x)$ is discontinuous at $x=100$ and $x=200$ and is continuous for all other values of $x$ satisfying $x \geq 0$.

TABLE 1
Price of Sugar Paid by Bakeco

| Size of Order | Price per Pound (c) |
| ---: | :---: |
| $0 \leq x<100$ | 25 |
| $100 \leq x \leq 200$ | 20 |
| $x>200$ | 15 |

FIGURE 1 Cost of Purchasing Sugar for Bakeco

table 2
Rules for Finding the Derivative of a Function

| Function | Derivative of Function |
| :--- | :--- |
| $a$ | 0 |
| $x$ | 1 |
| $a f(x)$ | $a f^{\prime}(x)$ |
| $f(x)+g(x)$ | $f^{\prime}(x)+g^{\prime}(x)$ |
| $x^{n}$ | $n x^{n-1}$ |
| $e^{x}$ | $e^{x}$ |
| $a^{x}$ | $a^{x} \ln a$ |
| $\ln x$ | $\frac{1}{x}$ |
| $[f(x)]^{n}$ | $n[f(x)]^{n-1} f^{\prime}(x)$ |
| $e^{f(x)}$ | $e^{f(x)} f^{\prime}(x)$ |
| $a^{f(x)}$ | $a^{f(x)} f^{\prime}(x) \ln a$ |
| $\ln f(x)$ | $\frac{f^{\prime}(x)}{f x}$ |
| $f(x) g(x)$ | $f(x) g^{\prime}(x)+f^{\prime}(x) g(x)$ |
| $\frac{f(x)}{g(x)}$ | $\frac{g(x) f^{\prime}(x)-f(x) g^{\prime}(x)}{g(x)^{2}}$ |

## Differentiation

DEFINITION - The derivative of a function $f(x)$ at $x=a\left[\right.$ written $\left.f^{\prime}(a)\right]$ is defined to be

$$
\lim _{\Delta x \rightarrow 0} \frac{f(a+\Delta x)-f(a)}{\Delta x}
$$

If this limit does not exist, then $f(x)$ has no derivative at $x=a$.
We may think of $f^{\prime}(a)$ as the slope of $f(x)$ at $x=a$. Thus, if we begin at $x=a$ and increase $x$ by a small amount $\Delta$ ( $\Delta$ may be positive or negative), then $f(x)$ will increase by an amount approximately equal to $\Delta f^{\prime}(a)$. If $f^{\prime}(a)>0$, then $f(x)$ is increasing at $x=$ $a$, whereas if $f^{\prime}(a)<0$, then $f(x)$ is decreasing at $x=a$. The derivatives of many functions can be found via application of the rules in Table 2 ( $a$ represents an arbitrary constant). Example 3 illustrates the use and interpretation of the derivative.

## EXAMPLE 3 Product Profitability

If a company charges a price $p$ for a product, then it can sell $3 e^{-p}$ thousand units of the product. Then, $f(p)=3,000 p e^{-p}$ is the company's revenue if it charges a price $p$.

1 For what values of $p$ is $f(p)$ decreasing? For what values of $p$ is $f(p)$ increasing?
2 Suppose the current price is $\$ 4$ and the company increases the price by $5 \phi$. By approximately how much would the company's revenue change?

Solution We have

$$
f^{\prime}(p)=-3,000 p e^{-p}+3,000 e^{-p}=3,000 e^{-p}(1-p)
$$

1 For $p<1, f^{\prime}(p)>0$ and $f(p)$ is increasing, whereas for $p>1, f^{\prime}(p)<0$ and $f(p)$ is decreasing.

2 Using the interpretation of $f^{\prime}(4)$ as the slope of $f(p)$ at $p=4$ (with $\Delta p=0.05$ ), we see that the company's revenue would increase by approximately

$$
0.05\left(3,000 e^{-4}\right)(1-4)=-8.24
$$

In actuality, of course, the company's revenue would increase by

$$
\begin{aligned}
f(4.05)-f(4) & =3,000(4.05) e^{-4.05}-3,000(4) e^{-4} \\
& =211.68-219.79=-8.11
\end{aligned}
$$

## Higher Derivatives

We define $f^{(2)}(a)=f^{\prime \prime}(a)$ to be the derivative of the function $f^{\prime}(x)$ at $x=a$. Similarly, we can define (if it exists) $f^{(n)}(a)$ to be the derivative of $f^{(n-1)}(x)$ at $x=a$. Thus, for Example 3,

$$
f^{\prime \prime}(p)=3,000 e^{-p}(-1)-3,000 e^{-p}(1-p)
$$

## Taylor Series Expansion

In the Taylor series expansion of a function $f(x)$, given that $f^{(n+1)}(x)$ exists for every point on the interval $[a, b]$, we can write for any $h$ satisfying $0 \leq h \leq b-a$,

$$
\begin{equation*}
f(a+h)=f(a)+\sum_{i=1}^{i=n} \frac{f^{(i)}(a)}{i!} h^{i}+\frac{f^{(n+1)}(p)}{(n+1)!} h^{n+1} \tag{1}
\end{equation*}
$$

where (1) will hold for some number $p$ between $a$ and $a+h$. Equation (1) is the $\boldsymbol{n}$ thorder Taylor series expansion of $f(x)$ about $a$.

## EXAMPLE 4 Taylor Series Expansion

Find the first-order Taylor series expansion of $e^{-x}$ about $x=0$.
Solution Because $f^{\prime}(x)=-e^{-x}$ and $f^{\prime \prime}(x)=e^{-x}$, we know that (1) will hold on any interval $[0, b]$. Also, $f(0)=1, f^{\prime}(0)=-1$, and $f^{\prime \prime}(x)=e^{-x}$. Then (1) yields the following first-order Taylor series expansion for $e^{-x}$ about $x=0$ :

$$
e^{-h}=f(h)=1-h+\frac{h^{2} e^{-p}}{2}
$$

This equation holds for some $p$ between 0 and $h$.

## Partial Derivatives

We now consider a function $f$ of $n>1$ variables $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$, using the notation $f\left(x_{1}\right.$, $x_{2}, \ldots, x_{n}$ ) to denote such a function.

DEFINITION ■ The partial derivative of $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ with respect to the variable $x_{i}$ is written $\frac{\partial f}{\partial x_{i}}$, where

$$
\frac{\partial f}{\partial x_{i}}=\lim _{\Delta x_{i} \rightarrow 0} \frac{f\left(x_{1}, \ldots, x_{i}+\Delta x_{i}, \ldots, x_{n}\right)-f\left(x_{1}, \ldots, x_{i}, \ldots, x_{n}\right)}{\Delta x_{i}}
$$

Intuitively, if $x_{i}$ is increased by $\Delta$ (and all other variables are held constant), then for small values of $\Delta$, the value of $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ will increase by approximately $\Delta \frac{\partial f}{\partial x_{i}}$. We find $\frac{\partial f}{\partial x_{i}}$ by treating all variables other than $x_{i}$ as constants and finding the derivatives of $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$. More generally, suppose that for each $i$, we increase $x_{i}$ by a small amount $\Delta x_{i}$. Then the value of $f$ will increase by approximately

$$
\sum_{i=1}^{i=n} \frac{\partial f}{\partial x_{i}} \Delta x_{i}
$$

## EXAMPLE 5 When Is a Function Increasing?

The demand $f(p, a)=30,000 p^{-2} a^{1 / 6}$ for a product depends on $p=$ product price (in dollars) and $a=$ dollars spent advertising the product. Is demand an increasing or decreasing function of price? Is demand an increasing or decreasing function of advertising expenditure? If $p=10$ and $a=1,000,000$, then by how much (approximately) will a $\$ 1$ cut in price increase demand?

Solution

$$
\begin{aligned}
& \frac{\partial f}{\partial p}=30,000\left(-2 p^{-3}\right) a^{1 / 6}=-60,000 p^{-3} a^{1 / 6}<0 \\
& \frac{\partial f}{\partial a}=30,000 p^{-2}\left(\frac{a^{-5 / 6}}{6}\right)=5,000 p^{-2} a^{-5 / 6}>0
\end{aligned}
$$

Thus, an increase in price (with advertising held constant) will decrease demand, while an increase in advertising (with price held constant) will increase demand. Because

$$
\frac{\partial f}{\partial p}(10,1,000,000)=-60,000\left(\frac{1}{1,000}\right)(1,000,000)^{1 / 6}=-600
$$

a $\$ 1$ price cut will increase demand by approximately $(-1)(-600)$, or 600 units.

We will also use second-order partial derivatives extensively. We use the notation $\frac{\partial^{2}}{\partial x_{i} \partial x_{j}}$ to denote a second-order partial derivative. To find $\frac{\partial^{2}}{\partial x_{i} \partial x_{j}}$, we first find $\frac{\partial f}{\partial x_{i}}$ and then take its partial derivative with respect to $x_{j}$. If the second-order partials exist and are everywhere continuous, then

$$
\frac{\partial^{2} f}{\partial x_{i} \partial x_{j}}=\frac{\partial^{2} f}{\partial x_{j} \partial x_{i}}
$$

## EXAMPLE 6 Second-Order Partial Derivatives

For $f(p, a)=30,000 p^{-2} a^{1 / 6}$, find all second-order partial derivatives.

Solution

$$
\begin{aligned}
\frac{\partial^{2} f}{\partial p^{2}} & =-60,000\left(-3 p^{-4}\right) a^{1 / 6}=\frac{180,000 a^{1 / 6}}{p^{4}} \\
\frac{\partial^{2} f}{\partial a^{2}} & =5,000 p^{-2}\left(\frac{-5 a^{-11 / 6}}{6}\right)=-\frac{25,000 p^{-2} a^{-11 / 6}}{6} \\
\frac{\partial^{2} f}{\partial a \partial p} & =5,000\left(-2 p^{-3}\right) a^{-5 / 6}=-10,000 p^{-3} a^{-5 / 6} \\
\frac{\partial^{2} f}{\partial p \partial a} & =-60,000 p^{-3}\left(\frac{a^{-5 / 6}}{6}\right)=-10,000 p^{-3} a^{-5 / 6}
\end{aligned}
$$

Observe that for $p \neq 0$ and $\mathrm{a} \neq 0$,

$$
\frac{\partial^{2} f}{\partial a \partial p}=\frac{\partial^{2} f}{\partial p \partial a}
$$

## PROBLEMS

## Group A

1 Find $\lim _{h \rightarrow 0} \frac{3 h+h^{2}}{h}$.
2 It costs Sugarco $25 \phi / \mathrm{lb}$ to purchase the first 100 lb of sugar, $20 \phi / \mathrm{lb}$ to purchase the next 100 lb , and $15 \phi$ to buy each additional pound. Let $f(x)$ be the cost of purchasing $x$ pounds of sugar. Is $f(x)$ continuous at all points? Are there any points where $f(x)$ has no derivative?
3 Find $f^{\prime}(x)$ for each of the following functions:
a $x e^{-x}$
b $\frac{x^{2}}{x^{2}+1}$
c $e^{3 x}$
d $(3 x+2)^{-2}$
e $\ln x^{3}$
4 Find all first- and second-order partial derivatives for $f\left(x_{1}, x_{2}\right)=x_{1}^{2} e^{x_{2}}$.
5 Find the second-order Taylor series expansion of $\ln x$ about $x=1$.

## Group B

6 Let $q=f(p)$ be the demand for a product when the price is $p$. For a given price $p$, the price elasticity $E$ of the product is defined by

$$
E=\frac{\text { percentage change in demand }}{\text { percentage change in price }}
$$

If the change in price $(\Delta p)$ is small, this formula reduces to

$$
E=\frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}}=\left(\frac{p}{q}\right)\left(\frac{d q}{d p}\right)
$$

a Would you expect $f(p)$ to be positive or negative?
b Show that if $E<-1$, a small decrease in price will increase the firm's total revenue (in this case, we say that demand is elastic).
c Show that if $-1<E<0$, a small price decrease will decrease total revenue (in this case, we say demand is inelastic).
7 Suppose that if $x$ dollars are spent on advertising during a given year, $k\left(1-e^{-c x}\right)$ customers will purchase a product $(c>0)$.
a As $x$ grows large, the number of customers purchasing the product approaches a limit. Find this limit.
b Can you give an interpretation for $k$ ?
c Show that the sales response from a dollar of advertising is proportional to the number of potential customers who are not purchasing the product at present.
8 Let the total cost of producing $x$ units, $c(x)$, be given by $c(x)=k x^{1-b}(0<b<1)$. This cost curve is called the learning or experience cost curve.
a Show that the cost of producing a unit is a decreasing function of the number of units that have been produced.
b Suppose that each time the number of units produced is doubled, the per-unit product cost drops to $r \%$ of its previous value (because workers learn how to perform their jobs better). Show that $r=100\left(2^{-b}\right)$.
9 If a company has $m$ hours of machine time and $w$ hours of labor, it can produce $3 m^{1 / 3} w^{2 / 3}$ units of a product. Currently, the company has 216 hours of machine time and 1,000 hours of labor. An extra hour of machine time costs $\$ 100$, and an extra hour of labor costs $\$ 50$. If the company has $\$ 100$ to invest in purchasing additional labor and machine time, would it be better off buying 1 hour of machine time or 2 hours of labor?

### 11.2 Introductory Concepts

DEFINITION - A general nonlinear programming problem (NLP) can be expressed as follows:
Find the values of decision variables $x_{1}, x_{2}, \ldots, x_{n}$ that

$$
\begin{array}{lc}
\max & (\text { or } \min ) z=f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \\
\text { s.t. } & g_{1}\left(x_{1}, x_{2}, \ldots, x_{n}\right) \quad(\leq,=, \text { or } \geq) b_{1} \\
\text { s.t. } & g_{2}\left(x_{1}, x_{2}, \ldots, x_{n}\right) \quad(\leq,=, \text { or } \geq) b_{2}  \tag{2}\\
& \vdots \\
& g_{m}\left(x_{1}, x_{2}, \ldots, x_{n}\right) \quad(\leq,=, \text { or } \geq) b_{m}
\end{array}
$$

As in linear programming, $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is the NLP's objective function, and $g_{1}\left(x_{1}, x_{2}, \ldots, x_{n}\right)(\leq,=$, or $\geq) b_{1}, \ldots, g_{m}\left(x_{1}, x_{2}, \ldots, x_{n}\right)(\leq,=$, or $\geq) b_{m}$ are the NLP's constraints. An NLP with no constraints is an unconstrained NLP.

The set of all points $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ such that $x_{i}$ is a real number is $R^{n}$. Thus, $R^{1}$ is the set of all real numbers. The following subsets of $R^{1}$ (called intervals) will be of particular interest:

$$
\begin{aligned}
{[a, b] } & =\text { all } x \text { satisfying } a \leq x \leq b \\
{[a, b) } & =\text { all } x \text { satisfying } a \leq x<b \\
(a, b] & =\text { all } x \text { satisfying } a<x \leq b \\
(a, b) & =\text { all } x \text { satisfying } a<x<b \\
{[a, \infty) } & =\text { all } x \text { satisfying } x \geq a \\
(-\infty, b] & =\text { all } x \text { satisfying } x \leq b
\end{aligned}
$$

The following definitions are analogous to the corresponding definitions for LPs given in Section 3.1.

## DEFINITION

The feasible region for NLP (2) is the set of points $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ that satisfy the $m$ constraints in (2). A point in the feasible region is a feasible point, and a point that is not in the feasible region is an infeasible point.

Suppose (2) is a maximization problem.
DEFINITION - Any point $\bar{x}$ in the feasible region for which $f(\bar{x}) \geq f(x)$ holds for all points $x$ in the feasible region is an optimal solution to the NLP. [For a minimization problem, $\bar{x}$ is the optimal solution if $f(\bar{x}) \leq f(x)$ for all feasible $x$.]

Of course, if $f, g_{1}, g_{2}, \ldots, g_{m}$ are all linear functions, then (2) is a linear programming problem and may be solved by the simplex algorithm.

## Examples of NLPs

## EXAMPLE $7 \quad$ Profit Maximization

It costs a company $c$ dollars per unit to manufacture a product. If the company charges $p$ dollars per unit for the product, customers demand $D(p)$ units. To maximize profits, what price should the firm charge?

Solution The firm's decision variable is $p$. Since the firm's profit is $(p-c) D(p)$, the firm wants to solve the following unconstrained maximization problem: $\max (p-c) D(p)$.

## EXAMPLE 8 Production Maximization

If $K$ units of capital and $L$ units of labor are used, a company can produce $K L$ units of a manufactured good. Capital can be purchased at $\$ 4 /$ unit and labor can be purchased at $\$ 1 /$ unit. A total of $\$ 8$ is available to purchase capital and labor. How can the firm maximize the quantity of the good that can be manufactured?

Solution Let $K=$ units of capital purchased and $L=$ units of labor purchased. Then $K$ and $L$ must satisfy $4 K+L \leq 8, K \geq 0$, and $L \geq 0$. Thus, the firm wants to solve the following constrained maximization problem:

$$
\begin{aligned}
& \max z=K L \\
& \text { s.t. } \quad 4 K+L \leq 8 \\
& K, L \geq 0
\end{aligned}
$$

## Solving NLPs with LINGO

LINGO may be used to solve NLPs on a PC. Figure 2 (file Cap.lng) contains the LINGO formulation and output for Example 8. From the Value column, we see that LINGO has found the solution $K=1$ and $L=4$, which has an objective function value of 4 . As we shall soon see, this is indeed the optimal solution to Example 8. However, in general, there is no guarantee that the solution found by LINGO is an optimal solution. Throughout this chapter, we will detail the circumstances in which you can be sure that LINGO will find the optimal solution to an NLP.

Note that the ${ }^{\wedge}$ symbol is used to indicate raising to a power and $*$ indicates multiplication. LINGO has several built-in functions including

- $\operatorname{ABS}(\mathrm{X})=$ absolute value of X
- $\operatorname{EXP}(\mathrm{X})=e^{x}$

■ $\operatorname{LOG}(\mathrm{X})=$ natural logarithm of X
In Sections 11.9 and 11.10, we will discuss the Price column of the LINGO output. We will not discuss the Reduced Cost column.

## Differences Between NLPs and LPs

Recall from Chapter 3 that the feasible region for any LP is a convex set (that is, if $A$ and $B$ are feasible for an LP, then the entire line segment joining $A$ and $B$ is also feasible). Also recall that if an LP has an optimal solution, then there is an extreme point of the feasible region that is optimal. We will soon see, however, that even if the feasible region for an NLP is a convex set, the optimal solution (unlike the optimal solution for an LP) need not be an extreme point of the NLP's feasible region. The previous example illustrates this idea. Figure 3 shows graphically the feasible region (bounded by triangle $A B C$ ) for the example and the isoprofit curves $K L=1, K L=2$, and $K L=4$. We see that the optimal solution to the example occurs where an isoprofit curve is tangent to the boundary of the feasible region. Thus, the optimal solution to the example is $z=4, K=1, L=4$ (point $D)$. Of course, point $D$ is not an extreme point of the NLP's feasible region. For this ex-

FIGURE 2


FIGURE 3 An NLP Whose Optimal

Solution Is Not an Extreme Point


FIGURE 4 An NLP Whose Optimal Solution Is Not on Boundary of Feasible Region

ample (and many other NLPs with linear constraints), the optimal solution fails to be an extreme point of the feasible region because the isoprofit curves are not straight lines. In fact, the optimal solution for an NLP may not be on the boundary of the feasible region. For example, consider the following NLP:

$$
\begin{array}{ll}
\max z= & f(x) \\
\text { s.t. } & 0 \leq x \leq 1
\end{array}
$$

where $f(x)$ is pictured in Figure 4. The optimal solution for this NLP is $z=1, x=\frac{1}{2}$. Of course, $x=\frac{1}{2}$ is not on the boundary of the feasible region.

## Local Extremum

For any NLP (maximization), a feasible point $x=\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a local maximum if for sufficiently small $\epsilon$, any feasible point $x^{\prime}=\left(x_{1}^{\prime}, x_{2}^{\prime}, \ldots, x_{n}^{\prime}\right)$ having $\left|x_{i}-x_{i}^{\prime}\right|<\epsilon(i=1,2, \ldots, n)$ satisfies $f(x) \geq f\left(x^{\prime}\right)$.

In short, a point $x$ is a local maximum if $f(x) \geq f\left(x^{\prime}\right)$ for all feasible $x^{\prime}$ that are close to $x$. Analogously, for a minimization problem, a point $x$ is a local minimum if $f(x) \leq f\left(x^{\prime}\right)$ holds for all feasible $x^{\prime}$ that are close to $x$. A point that is a local maximum or a local minimum is called a local, or relative, extremum.

For an LP (max problem), any local maximum is an optimal solution to the LP. (Why?) For a general NLP, however, this may not be true. For example, consider the following NLP:

$$
\begin{aligned}
& \max z=f(x) \\
& \text { s.t. } \quad 0 \leq x \leq 10
\end{aligned}
$$

where $f(x)$ is given in Figure 5. Points $A, B$, and $C$ are all local maxima, but point $C$ is the unique optimal solution to the NLP.

Unlike an LP, an NLP may not satisfy the Proportionality and Additivity assumptions. For instance, in Example 8, increasing $L$ by 1 will increase $z$ by $K$. Thus, the effect on $z$ of increasing $L$ by 1 depends on $K$. This means that the example does not satisfy the Additivity Assumption.
FIGURE 5
A Local Maximum May Not Be the Optimal Solution to an NLP


The NLP

$$
\begin{aligned}
& \max z=x^{1 / 3}+y^{1 / 3} \\
& \text { s.t. } \quad x+y=1 \\
& \\
& x, y \geq 0
\end{aligned}
$$

does not satisfy the Proportionality Assumption, because doubling the value of $x$ does not double the contribution of $x$ to the objective function.

## More Examples of NLP Formulations

We now give three more examples of nonlinear programming formulations.

## EXAMPLE 9 Ofico NLP

Oilco produces three types of gasoline: regular, unleaded, and premium. All three are produced by combining lead and crude oil brought in from Alaska and Texas. The required sulphur content, octane levels, minimum daily demand (in gallons), and sales price per gallon of each type of gasoline are given in Table 3. The crude brought in from Alaska is made by blending two types of crude: Alaska1 and Alaska2. The Alaska crude is blended in Alaska and shipped via pipeline to Oilco's Texas refinery. At most, 10,000 gallons of crude per day can be shipped from Alaska. The sulphur content, octane level, daily maximum amount available (in gallons) and purchase cost (per gallon) for each type of Alaska crude, Texas crude, and lead are given in Table 4. Of course, unleaded gasoline can contain no lead. Formulate an NLP to help Oilco maximize the daily profit obtained from selling gasoline. ${ }^{\dagger}$
Solution After defining the following decision variables:

$$
\begin{aligned}
R & =\text { gallons of regular gasoline produced daily } \\
U & =\text { gallons of unleaded gasoline produced daily } \\
P & =\text { gallons of premium gasoline produced daily } \\
A 1 & =\text { gallons of Alaskal crude purchased daily } \\
A 2 & =\text { gallons of Alaska } 2 \text { crude purchased daily } \\
T & =\text { gallons of Texas crude purchased daily } \\
L & =\text { gallons of lead purchased daily }
\end{aligned}
$$

[^15]
### 11.3 Convex and Concave Functions

Convex and concave functions play an extremely important role in the study of nonlinear programming problems.

Let $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ be a function that is defined for all points $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ in a convex set $S .^{\dagger}$

```
DEFINITION ■ A function f(x, , x , .., 柱) is a convex function on a convex set S if for any
    x ^ { \prime } \in S \text { and } x ^ { \prime \prime } \in S
        f[cx'}+(1-c)\mp@subsup{x}{}{\prime\prime}]\leqcf(\mp@subsup{x}{}{\prime})+(1-c)f(\mp@subsup{x}{}{\prime\prime}
holds for \(0 \leq c \leq 1\).
DEFINITION A function \(f\left(x_{1}, x_{2}, \ldots, x_{n}\right)\) is a concave function on a convex set \(S\) if for any \(x^{\prime} \in S\) and \(x^{\prime \prime} \in S\)
\[
f\left[c x^{\prime}+(1-c) x^{\prime \prime}\right] \geq c f\left(x^{\prime}\right)+(1-c) f\left(x^{\prime \prime}\right)
\]
holds for \(0 \leq c \leq 1\).
```

From (3) and (4), we see that $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a convex function if and only if $-f\left(x_{1}\right.$, $x_{2}, \ldots, x_{n}$ ) is a concave function, and conversely.

To gain some insights into these definitions, let $f(x)$ be a function of a single variable. From Figure 15 and inequality (3), we find that $f(x)$ is convex if and only if the line segment joining any two points on the curve $y=f(x)$ is never below the curve $y=f(x)$. Similarly, Figure 16 and inequality (4) show that $f(x)$ is a concave function if and only if the straight line joining any two points on the curve $y=f(x)$ is never above the curve $y=f(x)$.

FIGURE 15
A Convex Function


Point $A=\left(x^{\prime}, f\left(x^{\prime}\right)\right)$
Point $D=\left(x^{\prime \prime}, f\left(x^{\prime \prime}\right)\right)$
Point $C=\left(c x^{\prime}+(1-c) x^{\prime \prime}, c f\left(x^{\prime}\right)+(1-c) f\left(x^{\prime \prime}\right)\right)$
Point $B=\left(c x^{\prime}+(1-c) x^{\prime \prime}, f\left(c x^{\prime}+(1-c) x^{\prime \prime}\right)\right)$
From figure: $f\left(c x^{\prime}+(1-c) x^{\prime \prime}\right) \leq c f\left(x^{\prime}\right)+(1-c) f\left(x^{\prime \prime}\right)$

FIGURE 16
A Concave Function


Point $A=\left(x^{\prime}, f\left(x^{\prime}\right)\right)$
Point $D=\left(x^{\prime \prime}, f\left(x^{\prime \prime}\right)\right)$
Point $C=\left(c x^{\prime}+(1-c) x^{\prime \prime}, f\left(c x^{\prime}+(1-c) x^{\prime \prime}\right)\right)$
Point $B=\left(c x^{\prime}+(1-c) x^{\prime \prime}, c f\left(x^{\prime}\right)+(1-c) f\left(x^{\prime \prime}\right)\right)$
From figure: $f\left(c x^{\prime}+(1-c) x^{\prime \prime}\right) \geq c f\left(x^{\prime}\right)+(1-c) f\left(x^{\prime \prime}\right)$

[^16]For $x \geq 0, f(x)=x^{2}$ and $f(x)=e^{x}$ are convex functions and $f(x)=x^{1 / 2}$ is a concave function. These facts are evident from Figure 17.

a Convex

b Convex

FIGURE 17
Examples of Convex and Concave Functions

## EXAMPLE 13


c Concave

## Sum of Convex Functions

It can be shown (see Problem 12 at the end of this section) that the sum of two convex functions is convex and the sum of two concave functions is concave. Thus, $f(x)=x^{2}+$ $e^{x}$ is a convex function.

## EXAMPLE 14 Neither Convex nor Concave Function

Because the line segment $A B$ lies below $y=f(x)$ and the line segment $B C$ lies above $y=$ $f(x), f(x)$ as pictured in Figure 18 is not a convex or a concave function.


## EXAMPLE 15

Both Convex and Concave Linear Function
A linear function of the form $f(x)=a x+b$ is both a convex and a concave function. This follows from

$$
\begin{aligned}
f\left[c x^{\prime}+(1-c) x^{\prime \prime}\right] & =a\left[c x^{\prime}+(1-c) x^{\prime \prime}\right]+b \\
& =c\left(a x^{\prime}+b\right)+(1-c)\left(a x^{\prime \prime}+b\right) \\
& =c f\left(x^{\prime}\right)+(1-c) f\left(x^{\prime \prime}\right)
\end{aligned}
$$

Both (3) and (4) hold with equality, so $f(x)=a x+b$ is both a convex and a concave function.

Before discussing how to determine whether a given function is convex or concave, we prove a result that illustrates the importance of convex and concave functions.

## THEOREM 1

Consider NLP (2) and assume it is a maximization problem. Suppose the feasible region $S$ for NLP (2) is a convex set. If $f(x)$ is concave on $S$, then any local maximum for NLP (2) is an optimal solution to this NLP.

Proof If Theorem 1 is false, then there must be a local maximum $\bar{x}$ that is not an optimal solution to NLP (2). Let $S$ be the feasible region for NLP (2) (we have assumed that $S$ is a convex set). Then, for some $x \in S, f(x)>f(\bar{x})$. The inequality (4) implies that for any $c$ satisfying $0<c<1$,

$$
\begin{aligned}
f[c \bar{x}+(1-c) x] & \geq c f(\bar{x})+(1-c) f(x) \\
& >c f(\bar{x})+(1-c) f(\bar{x}) \quad[\text { from } f(x)>f(\bar{x})] \\
& =f(\bar{x})
\end{aligned}
$$

Now observe that for $c$ arbitrarily near $1, c \bar{x}+(1-c) x$ is feasible (because $S$ is convex) and is near $\bar{x}$. Thus, $\bar{x}$ cannot be a local maximum. This contradiction proves Theorem 1.

Similar reasoning can be used to prove Theorem $1^{\prime}$ (see Problem 11 at the end of this section).

## THEOREM $\mathbf{1}^{\prime}$

Consider NLP (2) and assume it is a minimization problem. Suppose the feasible region $S$ for NLP (2) is a convex set. If $f(x)$ is convex on $S$, then any local minimum for NLP (2) is an optimal solution to this NLP.

Theorems 1 and $1^{\prime}$ demonstrate that if we are maximizing a concave function (or minimizing a convex function) over a convex feasible region $S$, then any local maximum (or local minimum) will solve NLP (2). As we solve NLPs, we will repeatedly apply Theorems 1 and $1^{\prime}$.

We now explain how to determine if a function $f(x)$ of a single variable is convex or concave. Recall that if $f(x)$ is a convex function of a single variable, the line joining any two points on $y=f(x)$ is never below the curve $y=f(x)$. From Figures 9 and 10, we see that $f(x)$ convex implies that the slope of $f(x)$ must be nondecreasing for all values of $x$.

## THEOREM 2

Suppose $f^{\prime \prime}(x)$ exists for all $x$ in a convex set $S$. Then $f(x)$ is a convex function on $S$ if and only if $f^{\prime \prime}(x) \geq 0$ for all $x$ in $S$.

Because $f(x)$ is convex if and only if $-f(x)$ is concave, Theorem $2^{\prime}$ must also be true.

## THEOREM $\mathbf{2}^{\prime}$

Suppose $f^{\prime \prime}(x)$ exists for all $x$ in a convex set $S$. Then $f(x)$ is a concave function on $S$ if and only if $f^{\prime \prime}(x) \leq 0$ for all $x$ in $S$.

## EXAMPLE 16 Determining If a Function Is Convex or Concave

1 Show that $f(x)=x^{2}$ is a convex function on $S=R^{1}$.
2 Show that $f(x)=e^{x}$ is a convex function on $S=R^{1}$.
3 Show that $f(x)=x^{1 / 2}$ is a concave function on $S=(0, \infty)$.
4 Show that $f(x)=a x+b$ is both a convex and a concave function on $S=R^{1}$.
Solution $1 f^{\prime \prime}(x)=2 \geq 0$, so $f(x)$ is convex on $S=R^{1}$.
$2 f^{\prime \prime}(x)=e^{x} \geq 0$, so $f(x)$ is convex on $S=R^{1}$.
$3 f^{\prime \prime}(x)=-x^{-3 / 2} / 4 \leq 0$, so $f(x)$ is a concave function on $S(0, \infty)$.
$4 f^{\prime \prime}(x)=0$, so $f(x)$ is both convex and concave on $S=R^{1}$.

How can we determine whether a function $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ of $n$ variables is convex or concave on a set $S \subset R^{n}$ ? We assume that $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ has continous second-order partial derivatives. Before stating the criterion used to determine whether $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is convex or concave, we require three definitions.

The Hessian of $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is the $n \times n$ matrix whose $i j$ th entry is

$$
\frac{\partial^{2} f}{\partial x_{i} \partial x_{j}}
$$

We let $H\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ denote the value of the Hessian at $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$. For example, if $f\left(x_{1}, x_{2}\right)=x_{1}^{3}+2 x_{1} x_{2}+x_{2}^{2}$, then

$$
H\left(x_{1}, x_{2}\right)=\left[\begin{array}{ll}
6 x_{1} & 2 \\
2 & 2
\end{array}\right]
$$

An $\boldsymbol{i}$ th principal minor of an $n \times n$ matrix is the determinant of any $i \times i$ matrix obtained by deleting $n-i$ rows and the corresponding $n-i$ columns of the matrix.

Thus, for the matrix

$$
\left[\begin{array}{ll}
-2 & -1 \\
-1 & -4
\end{array}\right]
$$

the first principal minors are -2 and -4 , and the second principal minor is $-2(-4)-$ $(-1)(-1)=7$. For any matrix, the first principal minors are just the diagonal entries of the matrix.

The $\boldsymbol{k}$ th leading principal minor of an $n \times n$ matrix is the determinant of the $k \times k$ matrix obtained by deleting the last $n-k$ rows and columns of the matrix.

We let $H_{k}\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ be the $k$ th leading principal minor of the Hessian matrix evaluated at the point $\left(x_{1}, x_{2}, \ldots, x_{n}\right)$. Thus, if $f\left(x_{1}, x_{2}\right)=x_{1}^{3}+2 x_{1} x_{2}+x_{2}^{2}$, then $H_{1}\left(x_{1}, x_{2}\right)=6 x_{1}$, and $H_{2}\left(x_{1}, x_{2}\right)=6 x_{1}(2)-2(2)=12 x_{1}-4$.

By applying Theorems 3 and $3^{\prime}$ (stated below, without proof), the Hessian matrix can be used to determine whether $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a convex or a concave (or neither) function on a convex set $S \subset R^{n}$. [See Bazaraa and Shetty pages 91-93 (1993) for proof of Theorems 3 and $3^{\prime}$.]

## THEOREM 3

Suppose $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ has continuous second-order partial derivatives for each point $x=\left(x_{1}, x_{2}, \ldots, x_{n}\right) \in S$. Then $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a convex function on $S$ if and only if for each $x \in S$, all principal minors of $H$ are nonnegative.

## EXAMPLE 17 Using the Hessian to Ascertain Convexity or Concavity 1

Show that $f\left(x_{1}, x_{2}\right)=x_{1}^{2}+2 x_{1} x_{2}+x_{2}^{2}$ is a convex function on $S=R^{2}$.
Solution We find that

$$
H\left(x_{1}, x_{2}\right)=\left[\begin{array}{ll}
2 & 2 \\
2 & 2
\end{array}\right]
$$

The first principal minors of the Hessian are the diagonal entries (both equal $2 \geq 0$ ). The second principal minor is $2(2)-2(2)=0 \geq 0$. For any point, all principal minors of $H$ are nonnegative, so Theorem 3 shows that $f\left(x_{1}, x_{2}\right)$ is a convex function on $R^{2}$.

## THEOREM $\mathbf{3}^{\prime}$

Suppose $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ has continuous second-order partial derivatives for each point $x=\left(x_{1}, x_{2}, \ldots, x_{n}\right) \in S$. Then $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a concave function on $S$ if and only if for each $x \in S$ and $k=1,2, \ldots, n$, all nonzero principal minors have the same sign as $(-1)^{k}$.

## EXAMPLE 18 Using the Hessian to Ascertain Convexity or Concavity 2

Show that $f\left(x_{1}, x_{2}\right)=-x_{1}^{2}-x_{1} x_{2}-2 x_{2}^{2}$ is a concave function on $R^{2}$.
Solution We find that

$$
H\left(x_{1}, x_{2}\right)=\left[\begin{array}{ll}
-2 & -1 \\
-1 & -4
\end{array}\right]
$$

The first principal minors are the diagonal entries of the Hessian ( -2 and -4 ). These are both nonpositive. The second principal minor is the determinant of $H\left(x_{1}, x_{2}\right)$ and equals $-2(-4)-(-1)(-1)=7>0$. Thus, $f\left(x_{1}, x_{2}\right)$ is a concave function on $R^{2}$.

## EXAMPLE 19 Using the Hessian to Ascertain Convexity or Concavity 3

Show that for $S=R^{2}, f\left(x_{1}, x_{2}\right)=x_{1}^{2}-3 x_{1} x_{2}+2 x_{2}^{2}$ is not a convex or a concave function.
Solution We have

$$
H\left(x_{1}, x_{2}\right)=\left[\begin{array}{rr}
2 & -3 \\
-3 & 4
\end{array}\right]
$$

The first principal minors of the Hessian are 2 and 4. Because both the first principal minors are positive, $f\left(x_{1}, x_{2}\right)$ cannot be concave. The second principal minor is $2(4)-$ $(-3)(-3)=-1<0$. Thus, $f\left(x_{1}, x_{2}\right)$ cannot be convex. Together, these facts show that $f\left(x_{1}, x_{2}\right)$ cannot be a convex or a concave function.

## EXAMPLE 20 Using the Hessian to Ascertain Convexity or Concavity 4

Show that for $S=R^{3}, f\left(x_{1}, x_{2}, x_{3}\right)=x_{1}^{2}+x_{2}^{2}+2 x_{3}^{2}-x_{1} x_{2}-x_{2} x_{3}-x_{1} x_{3}$ is a convex function.

Solution The Hessian is given by

$$
H\left(x_{1}, x_{2}, x_{3}\right)=\left[\begin{array}{rrr}
2 & -1 & -1 \\
-1 & 2 & -1 \\
-1 & -1 & 4
\end{array}\right]
$$

By deleting rows (and columns) 1 and 2 of Hessian, we obtain the first-order principal minor $4>0$. By deleting rows (and columns) 1 and 3 of Hessian, we obtain the firstorder principal minor $2>0$. By deleting rows (and columns) 2 and 3 of Hessian, we obtain the first-order principal minor $2>0$.

By deleting row 1 and column 1 of Hessian, we find the second-order principal minor

$$
\operatorname{det}\left[\begin{array}{rr}
2 & -1 \\
-1 & 4
\end{array}\right]=7>0
$$

By deleting row 2 and column 2 of Hessian, we find the second-order principal minor

$$
\operatorname{det}\left[\begin{array}{rr}
2 & -1 \\
-1 & 4
\end{array}\right]=7>0
$$

By deleting row 3 and column 3 of Hessian, we find the second-order principal minor

$$
\operatorname{det}\left[\begin{array}{rr}
2 & -1 \\
-1 & 2
\end{array}\right]=3>0
$$

The third-order principal minor is simply the determinant of the Hessian itself. Expanding by row 1 cofactors we find the third-order principal minor

$$
\begin{aligned}
& 2[(2)(4)-(-1)(-1)]-(-1)[(-1)(4)-(-1)(-1)] \\
& +(-1)[(-1)(-1)-(-1)(2)]=14-5-3=6>0
\end{aligned}
$$

Because for all $\left(x_{1}, x_{2}, x_{3}\right)$ all principal minors of the Hessian are nonnegative, we have shown that $f\left(x_{1}, x_{2}, x_{3}\right)$ is a convex function on $R^{3}$.

## PROBLEMS

## Group A

On the given set $S$, determine whether each function is convex, concave, or neither.

$$
\begin{array}{ll}
1 & f(x)=x^{3} ; S=[0, \infty) \\
2 & f(x)=x^{3} ; S=R^{1} \\
3 & f(x)=\frac{1}{x} ; S=(0, \infty) \\
4 & f(x)=x^{a}(0 \leq a \leq 1) ; S=(0, \infty) \\
5 & f(x)=\ln x ; S=(0, \infty) \\
6 & f\left(x_{1}, x_{2}\right)=x_{1}^{3}+3 x_{1} x_{2}+x_{2}^{2} ; S=R^{2} \\
7 & f\left(x_{1}, x_{2}\right)=x_{1}^{2}+x_{2}^{2} ; S=R^{2} \\
8 & f\left(x_{1}, x_{2}\right)=-x_{1}^{2}-x_{1} x_{2}-2 x_{2}^{2} ; S=R^{2} \\
9 & f\left(x_{1}, x_{2}, x_{3}\right)=-x_{1}^{2}-x_{2}^{2}-2 x_{3}^{2}+.5 x_{1} x_{2} ; S=R^{3}
\end{array}
$$

10 For what values of $a, b$, and $c$ will $a x_{1}^{2}+b x_{1} x_{2}+c x_{2}^{2}$ be a convex function on $R^{2}$ ? A concave function on $R^{2}$ ?

## Group B

11 Prove Theorem 1'.
12 Show that if $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ and $g\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ are convex functions on a convex set $S$, then $h\left(x_{1}, x_{2}, \ldots, x_{n}\right)=$ $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)=g\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a convex function on $S$.
13 If $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a convex function on a convex set $S$, show that for $c \geq 0, g\left(x, x_{2}, \ldots, x_{n}\right)=c f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a convex function on $S$, and for $c \leq 0, g\left(x_{1}, x_{2}, \ldots, x_{n}\right)=$ $c f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is a concave function on $S$.
14 Show that if $y=f(x)$ is a concave function on $R^{1}$, then $z=\frac{1}{f(x)}$ is a convex function [assume that $f(x)>0$ ].
15 A function $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is quasi-concave on a convex set $S \subset R^{n}$ if $x^{\prime} \in S, x^{\prime \prime} \in S$, and $0 \leq c \leq 1$ implies

$$
f\left[c x^{\prime}+(1-c) x^{\prime \prime}\right] \geq \min \left[f\left(x^{\prime}\right), f\left(x^{\prime \prime}\right)\right]
$$

Show that if $f$ is concave on $R^{1}$, then $f$ is quasi-concave. Which of the functions in Figure 19 is quasi-concave? Is a quasi-concave function necessarily a concave function?
16 From Problem 12, it follows that the sum of concave

## FIGURE 19


functions is concave. Is the sum of quasi-concave functions necessarily quasi-concave?
17 Suppose a function's Hessian has both positive and negative entries on its diagonal. Show that the function is neither concave nor convex.

18 Show that if $f(x)$ is a non-negative, increasing concave function, then $\ln [f(x)]$ is also a concave function.
19 Show that if a function $f\left(x_{1}, x_{2}, \ldots, x_{n}\right)$ is quasiconcave on a convex set $S$, then for any number $a$ the set $S_{a}=$ all points satisfying $f\left(x_{1}, x_{2}, \ldots, x_{n}\right) \geq a$ is a convex set.
20 Show that Theorem 1 is untrue if $f$ is a quasi-concave function.
21 Suppose the constraints of an NLP are of the form $g_{i}\left(x_{1}, x_{2}, \ldots, x_{n}\right) \leq b_{i}(i=1,2, \ldots m)$. Show that if each of the $g_{i}$ is a convex function, then the NLP's feasible region is convex.

## Group C

22 If $f\left(x_{1}, x_{2}\right)$ is a concave function on $R^{2}$, show that for any number $a$, the set of $\left(x_{1}, x_{2}\right)$ satisfying $f\left(x_{1}, x_{2}\right) \geq a$ is a convex set.
23 Let $\mathbf{Z}$ be a $N(0,1)$ random variable, and let $F(x)$ be the cumulative distribution function for $Z$. Show that on $S=$ $(-\infty, 0], F(x)$ is an increasing convex function, and on $S=$ $[0, \infty), F(x)$ is an increasing concave function.
24 Recall the Dakota LP discussed in Chapter 6. Let $v(L$, $\mathrm{FH}, \mathrm{CH}$ ) be the maximum revenue that can be earned when $L$ sq board ft of lumber, FH finishing hours, and CH carpentry hours are available.
a Show that $v(L, F H, C H)$ is a concave function.
b Explain why this result shows that the value of each additional available unit of a resource must be a nonincreasing function of the amount of the resource that is available.

c

### 11.4 Solving NLPs with One Variable

In this section, we explain how to solve the NLP

$$
\begin{array}{ll}
\max (\text { or } \min ) & f(x) \\
\text { s.t. } & x \in[a, b] \tag{5}
\end{array}
$$

[If $b=\infty$, then the feasible region for NLP (5) is $x \geq a$, and if $a=-\infty$, then the feasible region for (5) is $x \leq b$.]

To find the optimal solution to (5), we find all local maxima (or minima). A point that is a local maximum or a local minimum for (5) is called a local extremum. Then the optimal solution to (5) is the local maximum (or minimum) having the largest (or smallest) value of $f(x)$. Of course, if $a=-\infty$ or $b=\infty$, then (5) may have no optimal solution (see Figure 20).

There are three types of points for which (5) can have a local maximum or minimum (these points are often called extremum candidates):

Case 1 Points where $a<x<b$, and $f^{\prime}(x)=0$ [called a stationary point of $f(x)$ ].
Case 2 Points where $f^{\prime}(x)$ does not exist.
Case 3 Endpoints $a$ and $b$ of the interval $[a, b]$.

## Case 1. Points Where $a<x<b$ and $f^{\prime}(x)=0$

Suppose $a<x<b$, and $f^{\prime}\left(x_{0}\right)$ exists. If $x_{0}$ is a local maximum or a local minimum, then $f^{\prime}\left(x_{0}\right)=0$. To see this, look at Figures 21a and 21b. From Figure 21a, we see that if $f^{\prime}\left(x_{0}\right)$ $>0$, then there are points $x_{1}$ and $x_{2}$ near $x_{0}$ where $f\left(x_{1}\right)<f\left(x_{0}\right)$ and $f\left(x_{2}\right)>f\left(x_{0}\right)$. Thus, if $f^{\prime}\left(x_{0}\right)>0, x_{0}$ cannot be a local maximum or a local minimum. Similarly, Figure 21 b shows that if $f^{\prime}\left(x_{0}\right)<0$, then $x_{0}$ cannot be a local maximum or a local minimum. From Figures 21c and 21d, however, we see $f^{\prime}\left(x_{0}\right)=0$, then $x_{0}$ may be a local maximum or a local minimum. Unfortunately, Figure 21e shows that $f^{\prime}\left(x_{0}\right)$ can equal zero without $x_{0}$ being a local maximum or a local minimum. From Figure 21c, we see that if $f^{\prime}(x)$ changes from positive to negative as we pass through $x_{0}$, then $x_{0}$ is a local maximum. Thus, if $f^{\prime \prime}\left(x_{0}\right)<0, x_{0}$ is a local maximum. Similarly, from Figure 21d, we see that if $f^{\prime}(x)$ changes from negative to positive as we pass through $x_{0}, x_{0}$ is a local minimum. Thus, if $f^{\prime \prime}\left(x_{0}\right)>$ $0, x_{0}$ is a local minimum.

a $\max f(x)$
s.t. $x \in(-\infty, b]$

b $\max f(x)$
s.t. $x \in[a, \infty)$

a $f^{\prime}\left(x_{0}\right)>0$
$f\left(x_{1}\right)<f\left(x_{0}\right)$
$f\left(x_{2}\right)>f\left(x_{0}\right)$
$x_{0}$ not a local extremum


C $f^{\prime}\left(x_{0}\right)=0$

$$
\text { For } x<x_{0}, f^{\prime}(x)>0
$$

For $x>x_{0}, f^{\prime}(x)<0$
$x_{0}$ is a local maximum

b $f^{\prime}\left(x_{0}\right)<0$
$f\left(x_{1}\right)>f\left(x_{0}\right)$
$f\left(x_{2}\right)<f\left(x_{0}\right)$
$x_{0}$ not a local extremum

d $f^{\prime}\left(x_{0}\right)=0$
For $x<x_{0}, f^{\prime}(x)<0$
For $x>x_{0}, f^{\prime}(x)>0$
$x_{0}$ is a local maximum

e $x_{0}=0$ not a local maximum or a local minimum but $f^{\prime}\left(x_{0}\right)=0$

FIGURE 21 How to Determine Whether $x_{0}$ Is a Local Maximum or a Local Minimum When $f^{\prime}\left(x_{0}\right)$ Exists

## THEOREM 4

If $f^{\prime}\left(x_{0}\right)=0$ and $f^{\prime \prime}\left(x_{0}\right)<0$, then $x_{0}$ is a local maximum. If $f^{\prime}\left(x_{0}\right)=0$ and $f^{\prime \prime}\left(x_{0}\right)$ $>0$, then $x_{0}$ is a local minimum.

What happens if $f^{\prime}\left(x_{0}\right)=0$ and $f^{\prime \prime}\left(x_{0}\right)=0$ (this is the case in Figure 21e)? In this case, we determine whether $x_{0}$ is a local maximum or a local minimum by applying Theorem 5.

## THEOREM 5

If $f^{\prime}\left(x_{0}\right)=0$, and
1 If the first nonvanishing (nonzero) derivative at $x_{0}$ is an odd-order derivative [ $f^{(3)}\left(x_{0}\right), f^{(5)}\left(x_{0}\right)$, and so on], then $x_{0}$ is not a local maximum or a local minimum.

2 If the first nonvanishing derivative at $x_{0}$ is positive and an even-order derivative, then $x_{0}$ is a local minimum.

3 If the first nonvanishing derivative at $x_{0}$ is negative and an even-order derivative, then $x_{0}$ is a local maximum.

We omit the proofs of Theorems 4 and 5. [They follow in a straightforward fashion by applying the definition of a local maximum and a local minimum to the Taylor series expansion of $f(x)$ about $x_{0}$.] Theorem 4 is a special case of Theorem 5 . We ask you to prove Theorems 4 and 5 in Problems 16 and 17.

## Case 2. Points Where $f^{\prime}(x)$ Does Not Exist

If $f(x)$ does not have a derivative at $x_{0}, x_{0}$ may be a local maximum, a local minimum, or neither (see Figure 22). In this case, we determine whether $x_{0}$ is a local maximum or a local minimum by checking values of $f(x)$ at points $x_{1}<x_{0}$ and $x_{2}>x_{0}$ near $x_{0}$. The four possible cases that can occur are summarized in Table 9.


| TABLE 9 |  |  |
| :--- | :--- | ---: |
| How to Determine Whether a Point Where $f^{\prime}(x)$ Does Not Exist Is a Local Maximum |  |  |
| or a Local Minimum |  |  |
| Relationship Between |  | $x_{0}$ |
| $f\left(x_{0}\right), f\left(x_{1}\right)$, and $f\left(x_{2}\right)$ | Not local extremum | 16 a |
| $f\left(x_{0}\right)>f\left(x_{1}\right) ; f\left(x_{0}\right)<f\left(x_{2}\right)$ | Not local extremum | 16 b |
| $f\left(x_{0}\right)<f\left(x_{1}\right) ; f\left(x_{0}\right)>f\left(x_{2}\right)$ | Local maximum | 16 c |
| $f\left(x_{0}\right) \geq f\left(x_{1}\right) ; f\left(x_{0}\right) \geq f\left(x_{2}\right)$ | Local minimum | 16 d |
| $f\left(x_{0}\right) \leq f\left(x_{1}\right) ; f\left(x_{0}\right) \leq f\left(x_{2}\right)$ |  |  |

## Case 3. Endpoints $a$ and $b$ of $[a, b]$

From Figure 23, we see that
If $f^{\prime}(a)>0$, then $a$ is a local minimum.
If $f^{\prime}(a)<0$, then $a$ is a local maximum.
If $f^{\prime}(b)>0$, then $b$ is a local maximum.
If $f^{\prime}(b)<0$, then $b$ is a local minimum.
If $f^{\prime}(a)=0$ or $f^{\prime}(b)=0$, draw a sketch like Figure 22 to determine whether $a$ or $b$ is a local extremum.

The following examples illustrate how these ideas can be applied to solve NLPs of the form (5).


## EXAMPLE 21 Profit Maximization by Monopolist

It costs a monopolist $\$ 5 /$ unit to produce a product. If he produces $x$ units of the product, then each can be sold for $10-x$ dollars $(0 \leq x \leq 10)$. To maximize profit, how much should the monopolist produce?
Solution Let $P(x)$ be the monopolist's profit if he produces $x$ units. Then

$$
P(x)=x(10-x)-5 x=5 x-x^{2} \quad(0 \leq x \leq 10)
$$

Thus, the monopolist wants to solve the following NLP:

$$
\begin{array}{ll}
\max & P(x) \\
\text { s.t. } \quad 0 \leq x \leq 10
\end{array}
$$

We now classify all extremum candidates:
Case $1 P^{\prime}(x)=5-2 x$, so $P^{\prime}(2.5)=0$. Because $P^{\prime \prime}(x)=-2, x=2.5$ is a local maximum yielding a profit of $P(2.5)=6.25$.
Case $2 P^{\prime}(x)$ exists for all points in $[0,10]$, so there are no Case 2 candidates.
Case $3 a=0$ has $P^{\prime}(0)=5>0$, so $a=0$ is a local minimum; $b=10$ has $P^{\prime}(10)=$ $-15<0$, so $b=10$ is a local minimum.

Thus, $x=2.5$ is the only local maximum. This means that the monopolist's profits are maximized by choosing $x=2.5$.

Observe that $P^{\prime \prime}(x)=-2$ for all values of $x$. This shows that $P(x)$ is a concave function. Any local maximum for $P(x)$ must be the optimal solution to the NLP. Thus, Theorem 1 implies that once we have determined that $x=2.5$ is a local maximum, we know that it is the optimal solution to the NLP.

## EXAMPLE 22

Finding Clobal Maximum When Endpoint Is a Maximum
Let

$$
\begin{array}{lll}
f(x)=2-(x-1)^{2} & \text { for } & 0 \leq x<3 \\
f(x)=-3+(x-4)^{2} & \text { for } & 3 \leq x \leq 6
\end{array}
$$

Find

$$
\begin{array}{ll}
\max & f(x) \\
\text { s.t. } & 0 \leq x \leq 6
\end{array}
$$

Solution
Case 1 For $0 \leq x<3, f^{\prime}(x)=-2(x-1)$ and $f^{\prime \prime}(x)=-2$. For $3<x \leq 6, f^{\prime}(x)=$ $2(x-4)$ and $f^{\prime \prime}(x)=2$. Thus, $f^{\prime}(1)=f^{\prime}(4)=0$. Because $f^{\prime \prime}(1)<0, x=1$ is a local maximum. Because $f^{\prime \prime}(4)>0, x=4$ is a local minimum.

Case 2 From Figure 24, we see that $f(x)$ has no derivative at $x=3$ (for $x$ slightly less than 3 , $f^{\prime}(x)$ is near -4 , and for $x$ slightly bigger than $3, f^{\prime}(x)$ is near -2 ). Because $f(2.9)=$ $-1.61, f(3)=-2$, and $f(3.1)=-2.19, x=3$ is not a local extremum.
Case 3 Because $f^{\prime}(0)=2>0, x=0$ is a local minimum. Because $f^{\prime}(6)=4>0, x=$ 6 is a local maximum.

Thus, on $[0,6], f(x)$ has a local maximum for $x=1$ and $x=6$. Because $f(1)=2$ and $f(6)=1$, we find that the optimal solution to the NLP occurs for $x=1$.


## Decision Making under Uncertainty

We have all had to make important decisions where we were uncertain about factors that were relevant to the decisions. In this chapter, we study situations in which decisions are made in an uncertain environment.

The following model encompasses several aspects of making a decision in the absence of certainty. The decision maker first chooses an action $a_{i}$ from a set $A=\left\{a_{1}, a_{2}, \ldots, a_{k}\right\}$ of available actions. Then the state of the world is observed; with probability $p_{j}$, the state of the world is observed to be $s_{j} \in S=\left\{s_{1}, s_{2}, \ldots, s_{n}\right\}$. If action $a_{i}$ is chosen and the state of the world is $s_{j}$, the decision maker receives a reward $r_{i j}$. We refer to this model as the state-of-the-world decision-making model.

This chapter presents the basic theory of decision making under uncertainty: the widely used Von Neumann-Morgenstern utility model, and the use of decision trees for making decisions at different points in time. We close by looking at decision making with multiple objectives.

### 13.1 Decision Criteria

In this section, we consider four decision criteria that can be used to make decisions under uncertainty.

## EXAMPLE 1

## Newspaper Vendor

News vendor Phyllis Pauley sells newspapers at the corner of Kirkwood Avenue and Indiana Street, and each day she must determine how many newspapers to order. Phyllis pays the company $20 \phi$ for each paper and sells the papers for $25 \phi$ each. Newspapers that are unsold at the end of the day are worthless. Phyllis knows that each day she can sell between 6 and 10 papers, with each possibility being equally likely. Show how this problem fits into the state-of-the-world model.

Solution In this example, the members of $S=\{6,7,8,9,10\}$ are the possible values of the daily demand for newspapers. We are given that $p_{6}=p_{7}=p_{8}=p_{9}=p_{10}=\frac{1}{5}$. Phyllis must choose an action (the number of papers to order each day) from $A=\{6,7,8,9,10\}$.

If Phyllis purchases $i$ papers and $j$ papers are demanded, then $i$ papers are purchased at a cost of $20 i \phi$, and $\min (i, j)$ papers are sold for $25 \phi$ each. ${ }^{\dagger}$ Thus, if Phyllis purchases $i$ papers and $j$ papers are demanded, she earns a net profit of $r_{i j}$, where

$$
\begin{array}{ll}
r_{i j}=25 i-20 i=5 i & (i \leq j) \\
r_{i j}=25 j-20 i & (i \geq j)
\end{array}
$$

The values of $r_{i j}$ are tabulated in Table 1.
${ }^{\dagger} \min (i, j)$ is the smaller of $i$ and $j$.

TABLE 1
Rewards for News Vendor

|  | Papers Demanded |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Papers <br> Ordered | 6 | 7 | 8 | 9 | 10 |
| 6 | $30 \phi$ | $30 \phi$ | $30 \phi$ | $30 \phi$ | $30 \phi$ |
| 7 | $10 \phi$ | $35 \phi$ | $35 \phi$ | $35 \phi$ | $35 \phi$ |
| 8 | $-10 \phi$ | $15 \phi$ | $40 \phi$ | $40 \phi$ | $40 \phi$ |
| 9 | $-30 \phi$ | $-5 \phi$ | $20 \phi$ | $45 \phi$ | $45 \phi$ |
| 10 | $-50 \phi$ | $-25 \phi$ | $0 \phi$ | $25 \phi$ | $50 \phi$ |

## Dominated Actions

Why did we not consider the possibility that Phyllis would order $1,2,3,4,5$, or more than 10 papers? Answering this question involves the idea of a dominated action.

An action $a_{i}$ is dominated by an action $a_{i^{\prime}}$ if for all $s_{j} \in S, r_{i j} \leq r_{i^{\prime} j}$, and for some state $s_{j^{\prime}}, r_{i j^{\prime}}<r_{i^{\prime} j^{\prime}}$.

If action $a_{i}$ is dominated, then in no state of the world is $a_{i}$ better than $a_{i^{\prime}}$, and in at least one state of the world $a_{i}$ is inferior to $a_{i^{\prime}}$. Thus, if action $a_{i}$ is dominated, there is no reason to choose $a_{i}$ ( $a_{i^{\prime}}$ would be a better choice).

If Phyllis orders $i$ papers ( $i=1,2,3,4,5$ ), she will earn (for all states of the world) a profit of $5 i \phi$. From the table of rewards, we see that, for $i=1,2,3,4,5$, ordering 6 papers dominates ordering $i$ papers $\left(j^{\prime}=6,7,8,9\right.$, or 10 will do). Similarly, the reader should check that ordering $i$ papers $(i>11)$ is dominated by ordering 10 papers (see Problem 3 at the end of this section). A quick check shows that none of the actions in $A=$ $\{6,7,8,9,10\}$ are dominated. Thus, Phyllis should indeed choose her action from $A=$ $\{6,7,8,9,10\}$.

We now discuss four criteria that can be used to choose an action.

## The Maximin Criterion

For each action, determine the worst outcome (smallest reward). The maximin criterion chooses the action with the "best" worst outcome.

DEFINITION ■ The maximin criterion chooses the action $a_{i}$ with the largest value of $\min _{j \in S} r_{i j}$.

For Example 1, we obtain the results in Table 2. Thus, the maximin criterion recommends ordering 6 papers. This ensures that Phyllis will, no matter what the state of the world, earn a profit of at least $30 \notin$. The maximin criterion is concerned with making the worst possible outcome as pleasant as possible. Unfortunately, choosing a decision to mitigate the worst case may prevent the decision maker from taking advantage of good fortune. For example, if Phyllis follows the maximin criterion, she will never make less than $30 \phi$, but she will never make more than $30 \phi$.

TABLE 2
Computation of Maximin Decision for News Vendor

| Papers <br> Ordered | Worst State <br> of the World | Reward in Worst State <br> of the World |
| :---: | :---: | :---: |
| 6 | $6,7,8,9,10$ | $30 \phi$ |
| 7 | 6 | $10 \phi$ |
| 8 | 6 | $-10 \phi$ |
| 9 | 6 | $-30 \phi$ |
| 10 | 6 | $-50 \phi$ |

## The Maximax Criterion

For each action, determine the best outcome (largest reward). The maximax criterion chooses the action with the "best" best outcome.

DEFINITION ■ The maximax criterion chooses the action $a_{i}$ with the largest value of $\max _{j \in S} r_{i j}$.

For Example 1, we obtain the results in Table 3. Thus, the maximax criterion would recommend ordering 10 papers. In the best state (when 10 papers are demanded), this yields a profit of $50 \phi$. Of course, making a decision according to the maximax criterion leaves Phyllis open to the disastrous possibility that only 6 papers will be demanded, in which case she loses 50 ¢ .

## Minimax Regret

The minimax regret criterion (developed by L. J. Savage) uses the concept of opportunity cost to arrive at a decision. For each possible state of the world $s_{j}$, find an action $i^{*}(j)$ that maximizes $r_{i j}$. That is, $i^{*}(j)$ is the best possible action to choose if the state of the world is actually $s_{j}$. Then for any action $a_{i}$ and state $s_{j}$, the opportunity loss or regret for $a_{i}$ in $s_{j}$ is $r_{i^{*}(j), j}-r_{i j}$. For example, if $j=7$ papers are demanded, the best decision is to order $i^{*}(7)=7$ papers, yielding a profit of $r_{77}=7(25)-7(20)=35 \phi$. Suppose we chose to order $i=6$ papers. Since $r_{67}=6(25)-6(20)=30 \phi$, the opportunity loss or regret for $i=6$ and $j=7$ is $35-30=5 \phi$. Thus, if we order 6 papers and 7 papers are demanded, in hindsight we realize that by making the optimal choice (ordering 7 papers) for the actual state of the world ( 7 papers demanded), we would have done $5 \phi$ better than we did by ordering 6 papers. Table 4 shows the opportunity cost or regret matrix for Example 1.

| Papers Ordered | State Yielding Best Outcome | Best Outcome |
| :---: | :---: | :---: |
| 6 | 6, 7, 8, 9, 10 | $30 ¢$ |
| 7 | 7, 8, 9, 10 | $35 ¢$ |
| 8 | 8, 9, 10 | $40 ¢$ |
| 9 | 9, 10 | 45¢ |
| 10 | 10 | 50¢ |

The minimax regret criterion chooses an action by applying the minimax criterion to the regret matrix. In other words, the minimax regret criterion attempts to avoid disappointment over what might have been. From the regret matrix in Table 4, we obtain the minimax regret decision in Table 5. Thus, the minimax regret criterion recommends ordering 6 or 7 papers.

## The Expected Value Criterion

The expected value criterion chooses the action that yields the largest expected reward. For Example 1, the expected value criterion would recommend ordering 6 or 7 papers (see Table 6).

The decision-making criteria discussed in this section may seem reasonable, but many people make decisions without using any of them. A more comprehensive model of individual decision making, the Von Neumann-Morgenstern utility model, is discussed in Section 13.2.

TABLE 4
Regret Matrix for News Vendor

| Papers Ordered | Papers Demanded |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 6 | 7 | 8 | 9 | 10 |
| 6 | $30-30=0 ¢$ | $35-30=5 ¢$ | $40-30=10 ¢$ | $45-30=15 ¢$ | $50-30=20 ¢$ |
| 7 | $30-10=20 ¢$ | $35-35=0 ¢$ | $40-35=5 \phi$ | $45-35=10 ¢$ | $50-35=15 \phi$ |
| 8 | $30+10=40 ¢$ | $35-15=20 ¢$ | $40-40=0 ¢$ | $45-40=5 \phi$ | $50-40=10 ¢$ |
| 9 | $30+30=60 ¢$ | $35+5=40 ¢$ | $40-20=20 ¢$ | $45-45=0 ¢$ | $50-45=5 \phi$ |
| 10 | $30+50=80 ¢$ | $35+25=60 ¢$ | $40-0=40 ¢$ | $45-25=20 ¢$ | $50-50=0 ¢$ |


| Papers Ordered | Maximum Regret |
| :---: | :---: |
| 6 | 20¢ |
| 7 | 20¢ |
| 8 | 40¢ |
| 9 | 60¢ |
| 10 | 80¢ |

TABLE 6
Computation of Expected Value Decision for News Vendor

| Papers Ordered | Expected Reward |
| :--- | :---: |
| 6 | $\frac{1}{5}(30+30+30+30+30)=30 ¢$ |
| 7 | $\frac{1}{5}(10+35+35+35+35)=30 ф$ |
| 8 | $\frac{1}{5}(-10+15+40+40+40)=25 \phi$ |
| 9 | $\frac{1}{5}(-30-5+20+45+45)=15 \phi$ |
| 10 | $\frac{1}{5}(-50-25+0+25+50)=0 \phi$ |

## PROBLEMS

## Group A

1 Pizza King and Noble Greek are two competing restaurants. Each must determine simultaneously whether to undertake small, medium, or large advertising campaigns. Pizza King believes that it is equally likely that Noble Greek will undertake a small, a medium, or a large advertising campaign. Given the actions chosen by each restaurant, Pizza King's profits are as shown in Table 7. For the maximin, maximax, and minimax regret criteria, determine Pizza King's choice of advertising campaign.

TABLE 7

|  | Noble Greek Chooses |  |  |
| :--- | :---: | :---: | :---: |
| Pizza King | Small | Medium | Large |
| Chooses | $\$ 6,000$ | $\$ 5,000$ | $\$ 2,000$ |
| Small | $\$ 5,000$ | $\$ 6,000$ | $\$ 1,000$ |
| Medium | $\$ 9,000$ | $\$ 6,000$ | $\$ 0$ |

2 Sodaco is considering producing a new product: Chocovan soda. Sodaco estimates that the annual demand for Chocovan, D (in thousands of cases), has the following mass function: $P(\mathbf{D}=30)=.30, P(\mathbf{D}=50)=.40$, $P(\mathbf{D}=80)=.30$. Each case of Chocovan sells for $\$ 5$ and incurs a variable cost of $\$ 3$. It costs $\$ 800,000$ to build a plant to produce Chocovan. Assume that if $\$ 1$ is received every year (forever), this is equivalent to receiving $\$ 10$ at
the present time. Considering the reward for each action and state of the world to be in terms of net present value, use each decision criterion of this section to determine whether Sodaco should build the plant.

3 For Example 1, show that ordering 11 or more papers is dominated by ordering 10 papers.

## Group B

4 Suppose that Pizza King and Noble Greek stop advertising but must determine the price they will charge for each pizza sold. Pizza King believes that Noble Greek's price is a random variable $\mathbf{D}$ having the following mass function: $P(\mathbf{D}=\$ 6)=.25, P(\mathbf{D}=\$ 8)=.50, P(\mathbf{D}=$ $\$ 10)=.25$. If Pizza King charges a price $p_{1}$ and Noble Greek charges a price $p_{2}$, Pizza King will sell $100+$ $25\left(p_{2}-p_{1}\right)$ pizzas. It costs Pizza King $\$ 4$ to make a pizza. Pizza King is considering charging $\$ 5, \$ 6, \$ 7, \$ 8$, or $\$ 9$ for a pizza. Use each decision criterion of this section to determine the price that Pizza King should charge.
5 Alden Construction is bidding against Forbes Construction for a project. Alden believes that Forbes's bid is a random variable $\mathbf{B}$ with the following mass function: $P(\mathbf{B}=$ $\$ 6,000)=.40, P(\mathbf{B}=\$ 8,000)=.30, P(\mathbf{B}=\$ 11,000)=.30$. It will cost Alden $\$ 6,000$ to complete the project. Use each of the decision criteria of this section to determine Alden's bid. Assume that in case of a tie, Alden wins the bidding. (Hint: Let $p=$ Alden's bid. For $p \leq 6,000,6,000<$ $p \leq 8,000,8,000<p \leq 11,000$, and $p>11,000$, determine Alden's profit in terms of Alden's bid and Forbes's bid.)

### 13.2 Utility Theory

We now show how the Von Neumann-Morgenstern concept of a utility function can be used as an aid to decision making under uncertainty.

Consider a situation in which a person will receive, for $i=1,2, \ldots, n$, a reward $r_{i}$ with probability $p_{i}$. This is denoted as the lottery $\left(p_{1}, r_{1} ; p_{2}, r_{2} ; \ldots ; p_{n}, r_{n}\right)$. A lottery is often represented by a tree in which each branch stands for a possible outcome of the lottery, and the number on each branch represents the probability that the outcome will occur. Thus, the lottery $\left(\frac{1}{4}, \$ 500 ; \frac{3}{4}, \$ 0\right)$ could be denoted by


Suppose we are asked to choose between two lotteries ( $L_{1}$ and $L_{2}$ ). With certainty, lottery $L_{1}$ yields $\$ 10,000$ :

$$
L_{1} \xrightarrow{1} \$ 10,000
$$

Lottery $L_{2}$ consists of tossing a coin. If heads comes up, we receive $\$ 30,000$, and if tails comes up, we receive $\$ 0$ :

$L_{1}$ yields an expected reward of $\$ 10,000$, and $L_{2}$ yields an expected reward of $\left(\frac{1}{2}\right)(30,000)+$ $\left(\frac{1}{2}\right)(0)=\$ 15,000$. Although $L_{2}$ has a larger expected value than $L_{1}$, most people prefer $L_{1}$ to $L_{2}$ because $L_{1}$ offers the certainty of a relatively large payoff, whereas $L_{2}$ yields a substantial $\left(\frac{1}{2}\right)$ chance of earning a reward of $\$ 0$. In short, most people prefer $L_{1}$ to $L_{2}$ because $L_{1}$ involves less risk (or uncertainty) than $L_{2}$.

Our goal is to determine a method that a person can use to choose between lotteries. Suppose he or she must choose to play $L_{1}$ or $L_{2}$ but not both. We write $L_{1} \mathrm{p} L_{2}$ if the person prefers $L_{1}$. We write $L_{1} 1 L_{2}$ if he or she is indifferent between choosing $L_{1}$ and $L_{2}$. If $L_{1} \mathrm{i} L_{2}$, we say that $L_{1}$ and $L_{2}$ are equivalent lotteries. Finally, we write $L_{2} \mathrm{p} L_{1}$ if the decision maker prefers $L_{2}$.

Suppose we ask a decision maker to rank the following lotteries:


The Von Neumann-Morgenstern approach to ranking these lotteries is as follows. Begin by identifying the most favorable $(\$ 30,000)$ and the least favorable $(-\$ 10,000)$ outcomes that can occur. For all other possible outcomes ( $r_{1}=\$ 10,000, r_{2}=\$ 500$, and $r_{3}=\$ 0$ ), the decision maker is asked to determine a probability $p_{i}$ such that he or she is indifferent between two lotteries:


Suppose that for $r_{1}=\$ 10,000$, the decision maker is indifferent between

and for $r_{2}=\$ 500$, indifferent between

$$
\begin{array}{lll}
1  \tag{2}\\
\hline
\end{array} \$ 500 \quad \text { and } \quad \begin{array}{ll}
\frac{.62}{} & \$ 30,000 \\
.38 & -\$ 10,000
\end{array}
$$

and for $r_{3}=\$ 0$, indifferent between


Using (1)-(3), the decision maker can construct lotteries $L_{1}^{\prime}, L_{2}^{\prime}, L_{3}^{\prime}$, and $L_{4}^{\prime}$ such that $L_{i}^{\prime} L_{i}$ and each $L_{i}^{\prime}$ involves only the best $(\$ 30,000)$ and the worst $(-\$ 10,000)$ possible outcomes. Thus, from (1), we find that $L_{1} L_{1}^{\prime}$, where


From (3), we find that $L_{2} \mathrm{i} L_{2}^{\prime \prime}$, where

$L_{2}^{\prime \prime}$ is a compound lottery in which with probability .50 we receive $\$ 30,000$ and with probability .50 we play a lottery yielding a .60 chance at $\$ 30,000$ and a .40 chance at $-\$ 10,000$. More formally, a lottery $L$ is a compound lottery if for some $i$, there is a probability $p_{i}$ that the decision maker's reward is to play another lottery $L^{\prime}$. The following is an example of a compound lottery:


Thus, with probability $.50, L$ yields a reward of $-\$ 4$, and with probability $.50, L$ causes us to play $L^{\prime}$. If a lottery is not a compound lottery, it is a simple lottery.

Returning to our discussion of $L_{2}^{\prime \prime}$, we observe that $L_{2}^{\prime \prime}$ is a lottery that yields a $.50+$ $.50(.60)=.80$ chance at $\$ 30,000$ and a $.40(.50)=.20$ chance at $-\$ 10,000$. Thus, $L_{2} \mathrm{i} L_{2}^{\prime \prime} \mathrm{i} L_{2}^{\prime}$, where


Similarly, using (3), we find that $L_{3} L_{3}^{\prime}$, where


Using (2), we find that the decision maker is indifferent between $L_{4}$ and $L_{4}^{\prime \prime}$, where


In actuality, however, $L_{4}^{\prime \prime}$ yields a $.98(.62)=.6076$ chance at $\$ 30,000$ and a $.02+.38(.98)=$ .3924 chance at $-\$ 10,000$. Thus, $L_{4} \mathrm{i} L_{4}^{\prime \prime} \mathrm{i} L_{4}^{\prime}$, where


Since $L_{i} \mathrm{i} L_{i}^{\prime}$, we may rank $L_{1}, L_{2}, L_{3}$, and $L_{4}$ by ranking $L_{1}^{\prime}, L_{2}^{\prime}, L_{3}^{\prime}$, and $L_{4}^{\prime}$. Consider two lotteries whose only possible outcomes are $\$ 30,000$ (the most favorable outcome) and $-\$ 10,000$ (the least favorable outcome). If he or she is given a choice between two lotteries of this type, the decision maker simply chooses the lottery with the larger chance of receiving the most favorable outcome. Applying this idea to $L_{1}^{\prime}$ through $L_{4}^{\prime}$ yields $L_{1}^{\prime} \mathrm{p} L_{2}^{\prime} \mathrm{p} L_{4}^{\prime} \mathrm{p} L_{3}^{\prime}$. Since $L_{i} L_{i}^{\prime}$, we may conclude that $L_{1} \mathrm{p} L_{2} \mathrm{p} L_{4} \mathrm{p} L_{3}$.

We now give a more formal description of the process that we have used to rank $L_{1}$, $L_{2}, L_{3}$, and $L_{4}$. The utility of the reward $r_{i}$, written $u\left(r_{i}\right)$, is the number $q_{i}$ such that the decision maker is indifferent between the following two lotteries:


This definition forces $u$ (least favorable outcome) $=0$ and $u$ (most favorable outcome) $=$ 1. For our possible payoffs of $\$ 30,000,-\$ 10,000, \$ 0, \$ 500$, and $\$ 10,000$, we first find that $u(\$ 30,000)=1$ and $u(-\$ 10,000)=0$. Then $(1)-(3)$ yield $u(\$ 10,000)=.90$, $u(\$ 500)=.62$, and $u(\$ 0)=.60$. The specification of $u\left(r_{i}\right)$ for all rewards $r_{i}$ is called the decision maker's utility function.

For a given lottery $L=\left(p_{1}, r_{1} ; p_{2}, r_{2} ; \ldots ; p_{n}, r_{n}\right)$, define the expected utility of the lottery $L$, written $E(U$ for $L)$, by

$$
E(U \text { for } L)=\sum_{i=1}^{i=n} p_{i} u\left(r_{i}\right)
$$

Thus, in our example

$$
\begin{aligned}
& E\left(U \text { for } L_{1}\right)=1(.90)=.90 \\
& E\left(U \text { for } L_{2}\right)=.50(1)+.50(.60)=.80 \\
& E\left(U \text { for } L_{3}\right)=1(.60)=.60 \\
& E\left(U \text { for } L_{4}\right)=.02(0)+.98(.62)=.6076
\end{aligned}
$$

Recall that we found that $L_{i} L_{i}^{\prime}$, where $L_{i}^{\prime}$ yielded an $E\left(U\right.$ for $\left.L_{i}\right)$ chance at $\$ 30,000$ and a $1-E\left(U\right.$ for $\left.L_{i}\right)$ chance at $-\$ 10,000$. Thus, in choosing between lotteries $L_{1}^{\prime}, L_{2}^{\prime}, L_{3}^{\prime}$, and $L_{4}^{\prime}$ (or equivalently, $L_{1}, L_{2}, L_{3}$, and $L_{4}$ ), we simply chose the lottery with the largest
expected utility. Given two lotteries $L_{1}$ and $L_{2}$, we may choose between them via the expected utility criteria:

$$
\begin{array}{lll}
L_{1} \mathrm{p} L_{2} & \text { if and only if } & E\left(U \text { for } L_{1}\right)>E\left(U \text { for } L_{2}\right) \\
L_{2} \mathrm{p} L_{1} & \text { if and only if } & E\left(U \text { for } L_{2}\right)>E\left(U \text { for } L_{1}\right) \\
L_{1} \mathrm{i} L_{2} & \text { if and only if } & E\left(U \text { for } L_{2}\right)=E\left(U \text { for } L_{1}\right)
\end{array}
$$

## Von Neumann-Morgenstern Axioms

Von Neumann and Morgenstern proved that if a person's preferences satisfy the following axioms, then he or she should choose between lotteries by using the expected utility criterion.

## Axiom 1: Complete Ordering Axiom

For any two rewards $r_{1}$ and $r_{2}$, one of the following must be true: The decision maker (1) prefers $r_{1}$ to $r_{2}$, (2) prefers $r_{2}$ to $r_{1}$, or (3) is indifferent between $r_{1}$ and $r_{2}$. Also, if the person prefers $r_{1}$ to $r_{2}$ and $r_{2}$ to $r_{3}$, then he or she must prefer $r_{1}$ to $r_{3}$ (transitivity of preferences).

In our discussion, we used the Complete Ordering Axiom to determine the most and least favorable outcomes.

## Axiom 2: Continuity Axiom

If the decision maker prefers $r_{1}$ to $r_{2}$ and $r_{2}$ to $r_{3}$, then for some $c(0<c<1), L_{1} \mathrm{i} L_{2}$, where


In our informal discussion, we used the Continuity Axiom when we found, for example, that $L_{3} \mathrm{i} L_{3}^{\prime}$, where

$$
L_{3}-1 \quad \$ 0 \quad L_{3}^{\prime}-\begin{aligned}
& \frac{.60}{} \begin{array}{l}
\text { } \\
.40 \\
\end{array}-\$ 10,000
\end{aligned}
$$

## Axiom 3: Independence Axiom

Suppose the decision maker is indifferent between rewards $r_{1}$ and $r_{2}$. Let $r_{3}$ be any other reward. Then for any $c(0<c<1), L_{1} \mathrm{i} L_{2}$, where

$L_{1}$ and $L_{2}$ differ only in that $L_{1}$ has a probability $c$ of yielding a reward $r_{1}$, whereas $L_{2}$ has a probability $c$ of yielding a reward $r_{2}$. Thus, the Independence Axiom implies that the decision maker views a chance $c$ at $r_{1}$ and a chance $c$ at $r_{2}$ to be of identical value, and this view holds for all values of $c$ and $r_{3}$. We applied the Independence Axiom when we used (3) to claim that $L_{2} 1 L_{2}^{\prime \prime}$, where


## Axiom 4: Unequal Probability Axiom

Suppose the decision maker prefers reward $r_{1}$ to reward $r_{2}$. If two lotteries have only $r_{1}$ and $r_{2}$ as their possible outcomes, he or she will prefer the lottery with the higher probability of obtaining $r_{1}$.

We used the Unequal Probability Axiom when we concluded, for example, that $L_{1}^{\prime}$ was preferred to $L_{2}^{\prime}$ (because $L_{1}^{\prime}$ had a .90 chance at $\$ 30,000$ and $L_{2}^{\prime}$ had only a .80 chance at $\$ 30,000$ ).

## Axiom 5: Compound Lottery Axiom

Suppose that when all possible outcomes are considered, a compound lottery $L$ yields (for $i=1,2, \ldots, n)$ a probability $p_{i}$ of receiving a reward $r_{i}$. Then $L \mathrm{i} L^{\prime}$, where $L^{\prime}$ is the simple lottery $\left(p_{1}, r_{1} ; p_{2}, r_{2} ; \ldots ; p_{n}, r_{n}\right)$.

For example, consider the following compound lottery:

$L$ yields a $.50+.50(.40)=.70$ chance at $-\$ 4$ and a $.50(.60)=.30$ chance at $\$ 6$. Thus, $L i L^{\prime \prime}$, where


In our informal discussion, we used the Compound Lottery Axiom when, for example, we stated that the compound equivalent of $L_{2}\left(L_{2}^{\prime \prime}\right)$

was equivalent to the following simple lottery:


## Why We May Assume $u$ (Worst Outcome) $=0$ and $u$ (Best Outcome) $=1$

Up to now, we have assumed that $u$ (least favorable outcome) $=0$ and $u$ (most favorable outcome) $=1$. Even if a decision maker's utility function does not have these values, we can transform his or her utility function (without changing the preferences among lotteries) into a utility function having $u$ (least favorable outcome) $=0$ and $u$ (most favorable outcome) $=1$.

## LEMMA 1

Given a utility function $u(x)$, define for any $a>0$ and any $b$ the function $v(x)=$ $a u(x)+b$. Given any two lotteries $L_{1}$ and $L_{2}$, it will be the case that

1 A decision maker using $u(x)$ as his or her utility function will have $L_{1} \mathrm{p} L_{2}$ if and only if a decision maker using $v(x)$ as his or her utility function will have $L_{1} \mathrm{p} L_{2}$.

2 A decision maker using $u(x)$ as his or her utility function will have $L_{1} \mathrm{i} L_{2}$ if and only if a decision maker using $v(x)$ as his or her utility function will have $L_{1} \mathrm{i} L_{2}$.

## Proof Let

$$
\begin{aligned}
& L_{1}=\left(p_{1}, r_{1} ; p_{2}, r_{2} ; \ldots ; p_{n}, r_{n}\right) \\
& L_{2}=\left(p_{1}^{\prime}, r_{1}^{\prime} ; p_{2}^{\prime}, r_{2}^{\prime} ; \ldots ; p_{m}^{\prime}, r_{m}^{\prime}\right)
\end{aligned}
$$

Suppose the decision maker using $u(x)$ prefers $L_{1}$ to $L_{2}$. Then by the expected utility criterion, we know that

$$
\begin{equation*}
\sum_{i=1}^{i=n} p_{i} u\left(r_{i}\right)>\sum_{i=1}^{i=m} p_{i}^{\prime} u\left(r_{i}^{\prime}\right) \tag{4}
\end{equation*}
$$

Now the $v(x)$ decision maker will have $L_{1} \mathrm{p} L_{2}$ if

$$
\begin{equation*}
\sum_{i=1}^{i=n} p_{i}\left[a u\left(r_{i}\right)+b\right]>\sum_{i=1}^{i=m} p_{i}^{\prime}\left[a u\left(r_{i}^{\prime}\right)+b\right] \tag{5}
\end{equation*}
$$

Since

$$
\sum_{i=1}^{i=n} p_{i}=\sum_{i=1}^{i=m} p_{i}^{\prime}=1
$$

(5) simplifies to

$$
\begin{equation*}
a \sum_{i=1}^{i=n} p_{i} u\left(r_{i}\right)+b>a \sum_{i=1}^{i=m} p_{i}^{\prime} u\left(r_{i}^{\prime}\right)+b \tag{6}
\end{equation*}
$$

Since $a>0$, (6) follows from (4). Thus, if the $u(x)$ decision maker has $L_{1} \mathrm{p} L_{2}$, the $v(x)$ decision maker has $L_{1} \mathrm{p} L_{2}$. Similarly, if (6) holds, then (4) will hold. Thus, if the $v(x)$ decision maker has $L_{1} \mathrm{p} L_{2}$, the $u(x)$ decision maker will also have $L_{1} \mathrm{p} L_{2}$. A similar argument can be used to prove part (2) of Lemma 1.

Using Lemma 1, we can show that without changing how an individual ranks lotteries, we can transform the decision maker's utility function into one having $u$ (least favorable outcome $)=0$ and $u$ (most favorable outcome $)=1$. To illustrate, let's reconsider ranking lotteries $L_{1}-L_{4}$. Suppose our decision maker's utility function had $u(-\$ 10,000)=-5$ and $u(\$ 30,000)=10$. Define $v(x)=a u(x)+b$. Choose $a$ and $b$ so that $v(\$ 30,000)=$ $10 a+b=1$ and $v(-\$ 10,000)=-5 a+b=0$. Then $a=\frac{1}{15}$ and $b=\frac{1}{3}$. Then by Lemma 1 , the utility function $v(x)=\frac{u(x)}{15}+\frac{1}{3}$ will yield the same ranking of lotteries as does $u(x)$, and we will have constructed $v(x)$ so that $v(\$ 30,000)=1$ and $v(-\$ 10,000)=0$. Thus, we see that without loss of generality, we may assume that $u$ (least favorable outcome) $=$ 0 and $u$ (most favorable outcome $)=1$.

## Estimating an Individual's Utility Function

How might we estimate an individual's (call her Jill) utility function? We begin by assuming that the least favorable outcome (say, $-\$ 10,000$ ) has a utility of 0 and that the most favorable outcome (say, $\$ 30,000$ ) has a utility of 1 . Next we define a number $x_{1 / 2}$ having $u\left(x_{1 / 2}\right)=\frac{1}{2}$. To determine $x_{1 / 2}$, ask Jill for the number (call it $x_{1 / 2}$ ) that makes her indifferent between


Since Jill is indifferent between the two lotteries, they must have the same expected utility. Thus, $u\left(x_{1 / 2}\right)=\left(\frac{1}{2}\right)(1)+\left(\frac{1}{2}\right)(0)=\frac{1}{2}$.

This procedure yields a point $x_{1 / 2}$ having $u\left(x_{1 / 2}\right)=\frac{1}{2}$. Suppose Jill states that $x_{1 / 2}=$ $-\$ 3,400$. Using $x_{1 / 2}$ and the least favorable outcome ( $-\$ 10,000$ ) as possible outcomes, we can construct a lottery that can be used to determine the point $x_{1 / 4}$ having a utility of $\frac{1}{4}$ (that is, $u\left(x_{1 / 4}\right)=\frac{1}{4}$ ). Point $x_{1 / 4}$ must be such that Jill is indifferent between


Then $u\left(x_{1 / 4}\right)=\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{2}\right)(0)=\frac{1}{4}$. Thus, $x_{1 / 4}$ will satisfy $u\left(x_{1 / 4}\right)=\frac{1}{4}$. Suppose Jill states that $x_{1 / 4}=-\$ 8,000$. This gives us another point on Jill's utility function.

Jill can now use the $x_{1 / 2}$ and $\$ 30,000$ outcomes to construct a lottery that will yield a value $x_{3 / 4}$ satisfying $u\left(x_{3 / 4}\right)=\frac{3}{4}$. (How?) Suppose that $x_{3 / 4}=\$ 8,000$. Similarly, outcomes of $x_{1 / 4}$ and $-\$ 10,000$ can be used to construct a lottery that will yield a value $x_{1 / 8}$ satisfying $u\left(x_{1 / 8}\right)=\frac{1}{8}$. Now Jill's utility function can be approximated by drawing a curve (smooth, we hope) joining the points

$$
(-\$ 10,000,0),\left(x_{1 / 8}, 1 / 8\right),\left(x_{1 / 4}, 1 / 4\right), \ldots,(\$ 30,000,1)
$$

The result is shown in Figure 1. Unfortunately, if a decision maker's preferences violate any of the preceding axioms (such as transitivity), this procedure may not yield a smooth curve. If it does not yield a relatively smooth curve, more sophisticated procedures for assessing utility functions must be used (see Keeney and Raiffa (1976)).

FIGURE 1

## Jill's Utility Function



## Relation between an Individual's Utility Function and His or Her Attitude toward Risk

A decision maker's utility function contains information about his or her attitude toward risk. To discuss this information, we need to define the concepts of a lottery's certainty equivalent and risk premium.

DEFINITION ■ The certainty equivalent of a lottery $L$, written $C E(L)$, is the number $C E(L)$ such that the decision maker is indifferent between the lottery $L$ and receiving a certain payoff of $C E(L)$.

For example, we saw earlier that Jill was indifferent between


Thus, $C E(L)=-\$ 3,400$.

DEFINITION ■ The risk premium of a lottery $L$, written $R P(L)$, is given by $R P(L)=E V(L)$ $C E(L)$, where $E V(L)$ is the expected value of the lottery's outcomes.

For example, if

then $E V(L)=\left(\frac{1}{2}\right)(\$ 30,000)+\left(\frac{1}{2}\right)(-\$ 10,000)=\$ 10,000$. We have already seen that $C E(L)=$ $-\$ 3,400$. Thus, $R P(L)=10,000-(-3,400)=\$ 13,400$; Jill values $L$ at $\$ 13,400$ less than its expected value, because she does not like the large degree of uncertainty that is associated with the reward yielded by $L$.

Let a nondegenerate lottery be any lottery in which more than one outcome can occur. With respect to attitude toward risk, a decision maker is

1 Risk-averse if and only if for any nondegenerate lottery $L, R P(L)>0$
2 Risk-neutral if and only if for any nondegenerate lottery $L, R P(L)=0$
3 Risk-seeking if and only if for any nondegenerate lottery $L, R P(L)<0$
An individual's attitude toward risk depends on the concavity (or convexity) of his or her utility function.

```
A function }u(x)\mathrm{ is said to be strictly concave (or strictly convex) if for any
two points on the curve }y=u(x)\mathrm{ , the line segment joining those two points
lies entirely (with the exception of its endpoints) below (or above) the curve
y=u(x).
```

If $u(x)$ is differentiable, then $u(x)$ will be strictly concave if and only if $u^{\prime \prime}(x)<0$ for all $x$ and $u(x)$ will be strictly convex if and only if $u^{\prime \prime}(x)>0$ for all $x$. It can easily be shown that a decision maker with a utility function $u(x)$ is

1 Risk-averse if and only if $u(x)$ is strictly concave
2 Risk-neutral if and only if $u(x)$ is a linear function (if $u(x)$ is both convex and concave)
3 Risk-seeking if and only if $u(x)$ is strictly convex
To illustrate these definitions, we show that a decision maker with a concave utility function $u(x)$ exhibits risk-averse behavior (has $R P(L)>0$ ). Consider a binary lottery $L$ (a lottery with only two possible outcomes):


Suppose $u(x)$ is strictly concave. Then, from Figure 2, we see that

$$
E(U \text { for } L)=p u\left(x_{1}\right)+(1-p) u\left(x_{2}\right)=y \text {-coordinate of point } 1
$$

Since $C E(L)$ is the value $x^{*}$ having $u\left(x^{*}\right)=E(U$ for $L$, Figure 2 shows that $C E(L)<$ $E V(L)$, so $R P(L)>0$. This follows because the strict concavity of $u(x)$ implies that the line segment joining the points $\left(x_{1}, u\left(x_{1}\right)\right)$ and $\left(x_{2}, u\left(x_{2}\right)\right)$ lies below the curve $u(x)$.

We can also give an algebraic proof that $u(x)$ strictly concave implies that $R P(L)=$ $E V(L)-C E(L)>0$. Recall that for

$E V(L)=p x_{1}+(1-p) x_{2}$. Now the strict concavity of $u(x)$ implies that $u\left[p x_{1}+(1-\right.$ p) $\left.x_{2}\right]>p u\left(x_{1}\right)+(1-p) u\left(x_{2}\right)=E(U$ for $L)$. Thus, the decision maker prefers $p x_{1}+$ $(1-p) x_{2}=E V(L)$ with certainty to the prospect of playing $L$. The certainty equivalent

FIGURE 2
Why a Concave Utility Function Implies Risk-Averse Behavior

of $L$ must be less than $p x_{1}+(1-p) x_{2}=E V(L)$. This implies that $R P(L)=E V(L)-$ $C E(L)>0$, and the decision maker exhibits risk-averse behavior. In Problem 4 at the end of this section, the reader will be asked to show that if $u(x)$ is strictly convex, the decision maker exhibits risk-seeking behavior.

If the decision maker is risk-neutral (that is, $u(x)=a x+b$ ), he or she chooses among lotteries via the expected reward criterion of Section 13.1 (see Problem 5 at the end of this section). Thus, when ranking lotteries, a risk-neutral decision maker considers only the expected value (and not the risk) of the lotteries.

Example 2 illustrates the concepts of risk premium, certainty equivalent, and risk aversion.

## EXAMPLE 2 Joan's Assets

Joan's utility function for her asset position $x$ is given by $u(x)=x^{1 / 2}$. Currently, Joan's assets consist of $\$ 10,000$ in cash and a $\$ 90,000$ home. During a given year, there is a . 001 chance that Joan's home will be destroyed by fire or other causes. How much would Joan be willing to pay for an insurance policy that would replace her home if it were destroyed?

Solution Let $x=$ annual insurance premium. Then Joan must choose between the following lotteries:


Joan will prefer $L_{1}$ to $L_{2}$ if $L_{1}$ 's expected utility exceeds $L_{2}$ 's expected utility. Thus, $L_{1} \mathrm{p} L_{2}$ if and only if

$$
\begin{aligned}
(100,000-x)^{1 / 2} & >.001(10,000)^{1 / 2}+.999(100,000)^{1 / 2} \\
& >.10+315.91154 \\
& >316.01154
\end{aligned}
$$

Squaring both sides of the last inequality we find that $L_{1} \mathrm{p} L_{2}$ if and only if

$$
\begin{aligned}
100,000-x & >(316.01154)^{2} \\
x & <\$ 136.71
\end{aligned}
$$

Thus, Joan would pay up to $\$ 136.71$ for insurance. Of course, if $p=\$ 136.71, L_{1} \mathrm{i} L_{2}$.
Let's compute the risk premium for $L_{2}$ :

$$
E V\left(L_{2}\right)=.001(10,000)+.999(100,000)=\$ 99,910
$$

(an expected loss of $100,000-99,910=\$ 90$ ). Since $E\left(U\right.$ for $\left.L_{2}\right)=316.01154$, we can find $C E\left(L_{2}\right)$ from the relation $u\left(C E\left(L_{2}\right)\right)=316.01154$, or $\left[C E\left(L_{2}\right)\right]^{1 / 2}=316.01154$. Thus, $C E\left(L_{2}\right)=(316.01154)^{2}=\$ 99,863.29$, and

$$
R P\left(L_{2}\right)=E V\left(L_{2}\right)-C E\left(L_{2}\right)=99,910-99,863.29=\$ 46.71
$$

Therefore, Joan is willing to pay for annual home insurance $\$ 46.71$ more than the expected loss of $\$ 90$. (Recall that Joan was willing to pay up to $90+46.71=\$ 136.71$ to avoid the risk involved in her home being destroyed.) Joan exhibits risk-averse behavior $\left(R P\left(L_{2}\right)>0\right)$. Since

$$
u^{\prime \prime}(x)=\frac{-x^{-3 / 2}}{4}<0
$$

$u(x)$ is strictly concave, and $R P(L)>0$ would hold for any nondegenerate lottery.

In reality, many people exhibit both risk-seeking behavior (they purchase lottery tickets, go to Las Vegas) and risk-averse behavior (they buy home insurance). A person whose utility function contains both convex and concave segments may exhibit both risk-averse and risk-seeking behavior. Consider a decision maker whose utility function $u(x)$ for change in current asset position is given in Figure 3. If forced to choose between

what would this person do?
From Figure 3, we find that $u(0)=.20, u(2,500)=.50$, and $u(-300)=.18$. Thus, $E(U$ for $\left.L_{1}\right)=.20$ and $E\left(U\right.$ for $\left.L_{2}\right)=.10(.50)+.90(.18)=.212$. Thus, $L_{2} \mathrm{p} L_{1}$. This means that $L_{2}$ has a certainty equivalent of at least $\$ 0$. Since $E V\left(L_{2}\right)=-\$ 20$, this implies that $R P\left(L_{2}\right)=$ $E V\left(L_{2}\right)-C E\left(L_{2}\right)<0$. The decision maker exhibits risk-seeking behavior in this situation, because for changes in asset position between $\$ 0$ and $\$ 2,500, u(x)$ is a convex function.

Now suppose the decision maker can, for $\$ 200$, insure himself against a loss of $\$ 2,000$, which occurs with probability .08 . Then he must choose between


From Figure 3, $u(-200)=.19, u(0)=.20$, and $u(-2,000)=0$. Thus, $E\left(U\right.$ for $\left.L_{3}\right)=.19$ and $E\left(U\right.$ for $\left.L_{4}\right)=.80(0)+.92(.20)=.184$, and $L_{3} \mathrm{p} L_{4}$. This shows that $C E\left(L_{4}\right)<$ $-\$ 200$. Since $E V\left(L_{4}\right)=.08(-2,000)+.92(0)=-\$ 160, R P\left(L_{4}\right)=E V\left(L_{4}\right)-C E\left(L_{4}\right)>$ 0 , and the decision maker is exhibiting risk-averse behavior, because $u(x)$ is concave for $-2,000<x<0$. Thus, if his utility function has both convex and concave segments, a person can exhibit both risk-seeking and risk-averse behavior.

FIGURE 3 A Utility Function That Exhibits Both Risk-Seeking and Risk-Averse Behavior


## Exponential Utility

Classes of "ready-made" utility functions have been developed. One important class is called exponential utility and has been used in many financial investment analyses. An exponential utility function has only one adjustable numerical parameter, and there are straightforward ways to discover the most appropriate value of this parameter for a particular individual or company. So the advantage of using an exponential utility function is that it is relatively easy to assess. The drawback is that exponential utility functions do not capture all types of attitudes toward risk. Nevertheless, their ease of use has made them popular.

An exponential utility function has the following form:

$$
U(x)=1-e^{-x / R}
$$

Here, $x$ is a monetary value (a payoff if positive, a cost if negative), $U(x)$ is the utility of this value, and $R>0$ is an adjustable parameter called the risk tolerance. Basically, the risk tolerance measures how much risk the decision maker will tolerate. The larger the value of $R$, the less risk averse the decision maker is. That is, a person with a large value of $R$ is more willing to take risks than a person with a small value of $R$.

To assess a person's (or company's) exponential utility function, we need only assess the value of $R$. There are a couple of tips for doing this. First, it has been shown that the risk tolerance is approximately equal to that dollar amount $R$ such that the decision maker is indifferent between the following two options:

- Option 1: Obtain no payoff at all
- Option 2: Obtain a payoff of $R$ dollars or a loss of $R / 2$ dollars, depending on the flip of a fair coin

For example, if I am indifferent between a bet where I win $\$ 1,000$ or lose $\$ 500$, with probability 0.5 each, and not betting at all, then my $R$ is approximately $\$ 1,000$. From this criterion it certainly makes intuitive sense that a wealthier person (or company) ought to have a larger value of $R$. This has been found in practice.

A second tip for finding $R$ is based on empirical evidence found by Ronald Howard, a prominent decision analyst. Through his consulting experience with several large companies, he discovered tentative relationships between risk tolerance and several financial
variables-net sales, net income, and equity. (See Howard (1992).) Specifically, he found that $R$ was approximately $6.4 \%$ of net sales, $124 \%$ of net income, and $15.7 \%$ of equity for the companies he studied. For example, according to this prescription, a company with net sales of $\$ 30$ million should have a risk tolerance of approximately $\$ 1.92$ million. Howard admits that these percentages are only guidelines. However, they do indicate that larger and more profitable companies tend to have larger values of $R$, which means that they are more willing to take risks involving given dollar amounts.

## PROBLEMS

## Group A

1 Suppose my utility function for asset position $x$ is given by $u(x)=\ln x$.
a Am I risk-averse, risk-neutral, or risk-seeking?
b I now have $\$ 20,000$ and am considering the following two lotteries:

$$
\begin{array}{ll}
L_{1}: & \text { With probability } 1 \text {, I lose } \$ 1,000 . \\
L_{2}: & \text { With probability } .9 \text { I I gain } \$ 0 . \\
& \text { With probability } .1, \text { I lose } \$ 10,000 .
\end{array}
$$

Determine which lottery I prefer and the risk premium of $L_{2}$.
2 Answer Problem 1 for a utility function $u(x)=x^{2}$.
3 Answer Problem 1 for a utility function $u(x)=2 x+1$.
4 Show that a decision maker who has a strictly convex utility function will exhibit risk-seeking behavior.
5 Show that a decision maker who has a linear utility function will rank two lotteries according to their expected value.
6 A decision maker has a utility function for monetary gains $x$ given by $u(x)=(x+10,000)^{1 / 2}$.
a Show that the person is indifferent between the status quo and
$L$ : With probability $\frac{1}{3}$, he or she gains $\$ 80,000$
With probability $\frac{2}{3}$, he or she loses $\$ 10,000$
b If there is a $10 \%$ chance that a painting valued at $\$ 10,000$ will be stolen during the next year, what is the most (per year) that the decision maker would be willing to pay for insurance covering the loss of the painting?
7 Patty is trying to determine which of two courses to take. If she takes the operations research course, she believes that she has a $10 \%$ chance of receiving an A, a $40 \%$ chance for a B, and a $50 \%$ chance for a C. If Patty takes a statistics course, she has a $70 \%$ chance for a B, a $25 \%$ chance for a
C, and a $5 \%$ chance for a D. Patty is indifferent between


She is also indifferent between


If Patty wants to take the course that maximizes the expected utility of her final grade, which course should she take?

8 We are going to invest $\$ 1,000$ for a period of 6 months. Two potential investments are available: T-bills and gold. If the $\$ 1,000$ is invested in T-bills, we are certain to end the 6 -month period with $\$ 1,296$. If we invest in gold, there is a $\frac{3}{4}$ chance that we will end the 6 -month period with $\$ 400$ and a $\frac{1}{4}$ chance that we will end the 6 -month period with $\$ 10,000$. If we end up with $x$ dollars, our utility function is given by $u(x)=x^{1 / 2}$. Should we invest in gold or T-bills?
9 We now have $\$ 5,000$ in assets and are given a choice between investment 1 and investment 2 . With investment 1 , $80 \%$ of the time we increase our asset position by $\$ 295,000$, and $20 \%$ of the time we increase our asset position by $\$ 95,000$. With investment $2,50 \%$ of the time we increase our asset position by $\$ 595,000$, and $50 \%$ of the time we increase our asset position by $\$ 5,000$. Our utility function for final asset position $x$ is $u(x)$. We are given the following values for $u(x): ~ u(0)=0, u(640,000)=.80, u(810,000)=$ $.90, u(0)=0, u(90,000)=.30, u(1,000,000)=1$, $u(490,000)=.7$.
a Are we risk-averse, risk-seeking, or risk-neutral? Explain.
b Will we prefer investment 1 or investment 2?
10 My current income is $\$ 40,000$. I believe that I owe $\$ 8,000$ in taxes. For $\$ 500$, I can hire a CPA to review my tax return; there is a $20 \%$ chance that she will save me $\$ 4,000$ in taxes. My utility function for (disposable income) $=$ (current income) - (taxes) - (payment to accountant) is given by $\sqrt{x}$ where $x$ is disposable income. Should I hire the CPA?

## Group B

$11^{\dagger}$ (The Allais Paradox) Suppose we are offered a choice between the following two lotteries:
$L_{1}$ : With probability 1 , we receive $\$ 1$ million.
$L_{2}$ : With probability .10 , we receive $\$ 5$ million. With probability .89 , we receive $\$ 1$ million. With probability .01 , we receive $\$ 0$.
Which lottery do we prefer? Now consider the following two lotteries:
${ }^{\dagger}$ Based on Allais (1953).
$L_{3}$ : With probability .11, we receive $\$ 1$ million. With probability .89 , we receive $\$ 0$.
$L_{4}$ : With probability .10 , we receive $\$ 5$ million. With probability .90 , we receive $\$ 0$.
Which lottery do we prefer? Suppose (like most people), we prefer $L_{1}$ to $L_{2}$. Show that $L_{3}$ must have a larger expected utility than $L_{4}$.
12 (The St. Petersburg Paradox) Let $L$ represent the following lottery. I toss a coin until it comes up heads. If the first heads is obtained on the $n$th toss of the coin, I receive a payoff of $\$ 2^{n}$.
a If I were a risk-neutral decision maker, what would be the certainty equivalent of $L$ ? Is this reasonable?
b If a decision maker's utility function for increasing wealth by $x$ dollars is given by $u(x)=\log _{2}(x)$, what would be the certainty equivalent of $L$ ?
13 Joe is a risk-averse decision maker. Which of the following lotteries will he prefer?

$$
\begin{array}{ll}
L_{1}: & \text { With probability } .10, \text { Joe loses } \$ 100 . \\
& \text { With probability } .90, \text { Joe receives } \$ 0 \\
L_{2}: & \text { With probability } .10 \text {, Joe loses } \$ 190 . \\
& \text { With probability } .90 \text {, Joe receives } \$ 10 .
\end{array}
$$

$14^{\dagger}$ (The Ellsberg Paradox) An urn contains 90 balls. It is known that 30 are red and that each of the other 60 is either yellow or black. One ball will be drawn at random from the urn. Consider the following four options:
Option 1 We receive $\$ 1,000$ if a red ball is drawn.
Option 2 We receive $\$ 1,000$ if a yellow ball is drawn.
Option 3 We receive $\$ 1,000$ if a yellow or black ball is drawn.
Option 4 We receive $\$ 1,000$ if a red or black ball is drawn.
a Explain why most people prefer option 1 over option 2 and also prefer option 3 over option 4.
b If we prefer option 1 to option 2, explain why we should also prefer option 4 over option 3.
15 Although the Von Neumann-Morgenstern axioms seem plausible, there are many reasonable situations in which people appear to violate these axioms. For example, suppose
${ }^{\dagger}$ Based on Ellsberg (1961).
tABLE 8

|  | Starting <br> Salary | Location | Opportunity for <br> Advancement |
| :--- | :---: | :---: | :---: |
| Job 1 | E | S | G |
| Job 2 | G | E | S |
| Job 3 | S | G | E |

a recent college graduate must choose between three job offers on the basis of starting salary, location of job, and opportunity for advancement. Given two job offers that are satisfactory with regard to all three attributes, the graduate will decide between two job offers by choosing the one that is superior on at least two of the three attributes. Suppose he or she has three job offers and has rated each one as shown in Table $8(\mathrm{E}=$ excellent, $\mathrm{G}=$ good, and $\mathrm{S}=$ satisfactory). Show that the graduate's preferences among these jobs violate the Complete Ordering Axiom.

## Group C

16 Suppose my utility function for my asset position is $u(x)=x^{1 / 2}$. I have $\$ 10,000$ at present. Consider the following lottery:
$L$ : With probability $\frac{1}{2}, L$ yields a payoff of $\$ 1,025$. With probability $\frac{1}{2}, L$ yields a payoff of $-\$ 199$.
a If I don't have the right to play $L$, find an equation that when solved would yield the amount I would be willing to pay for the right to play $L$. This is called the buying price of lottery $L$.
b If I have the right to play $L$, what is the least I would accept from somebody who wanted to buy the right to play $L$ ? (After someone else buys $L$, I can't play $L$.) This is called the selling price of lottery $L$.
c Answer part (b) for the case that I have $\$ 1,000$.
d Suppose that my utility function for my asset position is $u(x)=1-e^{-x}$. Show that for all possible asset positions, the buying price of $L$ and the selling price of $L$ will remain the same. Show that for all asset positions, the buying price of $L$ will equal the selling price of $L$.

### 13.3 Flaws in Expected Maximization of Utility: Prospect Theory and Framing Effects

The axioms underlying expected maximization of utility (EMU) seem reasonable, but in practice people's decisions often deviate from the predictions of EMU. Psychologists Tversky and Kahneman ${ }^{\ddagger}$ (1981) developed prospect theory and framing effects for values to try and explain why people deviate from the predictions of EMU.
${ }^{\dagger}$ In 2002, Kahneman received the Nobel Prize for Economics, in large part honoring his work with Tversky. Tversky was not awarded the prize because he died in 1996 (Nobel Prizes are not given posthumously).

## Prospect Theory

Here is one example of a decision that cannot be explained by EMU. Ask a person to choose between lottery 1 and lottery 2 :

| Lottery 1: | $\$ 30$ for certain |
| :--- | :--- |
| Lottery 2: | $80 \%$ chance at $\$ 45$ and $20 \%$ chance at $\$ 0$ |

Most people prefer lottery 1 to lottery 2 . Next ask the same person to choose between lottery 3 and lottery 4 :

Lottery 3: $20 \%$ chance at $\$ 45$ and $80 \%$ chance at $\$ 0$
Lottery 4: $25 \%$ at $\$ 30$ and $75 \%$ chance at $\$ 0$
Most people choose lottery 3 over lottery 4 . Now let $u(0)=0$ and $u(45)=1$. A decision maker following EMU will choose lottery 1 over lottery 2 if and only if $u(30)>.8$. A decision maker following EMU will choose lottery 3 over lottery 4 if and only if $.2>$ $.25 u(30)$ or $u(30)<.8$. This implies that a believer in EMU cannot choose lottery 1 over lottery 2 and lottery 3 over lottery 4 . Thus, for this situation, the choices of most people contradict EMU. Tversky and Kahneman developed prospect theory to explain the decision-making paradox we have just described. Prospect theory assumes that we do not treat probabilities as they are given in a decision-making problem. Instead, the decision maker treats a probability $p$ for an event as a "distorted" probability $\Pi(p)$. A $\Pi(p)$ function that seems to explain many paradoxes is shown in Figure 4.

The shape of the $\Pi(p)$ function in the figure implies that individuals are more sensitive to changes in probability when the probability of an event is small (near 0 ) or large (near 1). The equation we used to construct our $\Pi(p)$ curve is $\Pi(p)=1.89799 \mathrm{p}-3.55995 p^{2}+$ $2.662549 p^{3}$. How does prospect theory explain our paradox? From the values of $\Pi(p)$ given in Figure 5, we can compare the expected "prospects" of lottery 1 versus lottery 2 and lottery 3 versus lottery 4 .

| Prospect for lottery 1: | $u(30)$ |
| :--- | :--- |
| Prospect for lottery 2: | .602 |
| Prospect for lottery 3: | .258 |
| Prospect for lottery 4: | $.293 u(30)$. |

Thus, lottery 1 is preferred to lottery 2 if $u(30)>.602$, while lottery 3 is preferred to lottery 4 if $.258>.293 u(30)$ or $u(30)<.258 / .293=.88$. Our paradox evaporates, because for many people, $u(30)$ will be between .602 and .88 !


|  | C | D |
| ---: | ---: | ---: |
| 15 | 0.1 | 0.156803 |
| 16 | 0.15 | 0.213497 |
| 17 | 0.2 | 0.258382 |
| 18 | 0.25 | 0.293455 |
| 19 | 0.3 | 0.320713 |
| 20 | 0.35 | 0.342153 |
| 21 | 0.4 | 0.359771 |
| 22 | 0.45 | 0.375565 |
| 23 | 0.5 | 0.391531 |
| 24 | 0.55 | 0.409667 |
| 25 | 0.6 | 0.431969 |
| 26 | 0.65 | 0.460434 |
| 27 | 0.7 | 0.497059 |
| 28 | 0.75 | 0.543841 |
| 29 | 0.8 | 0.602778 |
| 30 | 0.85 | 0.675865 |
| 31 | 0.9 | 0.7651 |
| 32 | 0.95 | 0.872479 |
| 33 | 1 | 1 |

## Framing

The idea of framing is based on the fact that people often set their utility function from the standpoint of a frame or status quo from which they view the current situation. Most people's utility functions treat a loss of a given value as being more serious than a gain of an identical value. This is reflected in the utility function shown in Figure 6, which is convex for losses and concave for gains.

To see how framing can explain the failure of EMU, consider the following problem that Tversky and Kahneman gave to a group of students. The US is preparing for the outbreak of a disease that is expected to kill 600 people. Two alternative programs have been proposed:

Program I: 200 people are saved.
Program II: With probability $\frac{1}{3}, 600$ people are saved.
Most students preferred program I, probably because with program II there is a large risk of saving nobody. Since the programs are phrased in terms of lives saved, most people take the frame or reference point for this problem to be no lives saved or 600 people dead. Since the effect of each program is expressed in gains, and the utility function is concave for gains, we find that $\left.u(200)=u\left(\left(\frac{2}{3}\right) 0+\left(\frac{1}{3}\right) 600\right)\right)>\left(\frac{1}{3}\right) u(600)+\left(\frac{2}{3}\right) u(0)=\left(\frac{1}{3}\right) u(600)$. This implies, of course, that the person chooses program I over program II.

FIGURE 6
Utility Function for Framing

Next, Tversky and Kahneman rephrased the problem as follows:

$$
\begin{array}{ll}
\text { Program I: } & 400 \text { people die. } \\
\text { Program II: } & \text { With probability } \frac{2}{3}, 600 \text { people die. }
\end{array}
$$

Now most people choose program II. Note that both program I's are identical, as are both program II's. Why do most people choose program II for the second phrasing of the alternatives? The second phrasing shifts most people's reference points from "No lives saved" (in first phrasing) to "Nobody dies." The outcomes are expressed as losses (deaths), so the convexity of the utility curve for losses implies that

$$
\left(\frac{2}{3}\right) u(-600)=\left(\frac{2}{3}\right) u(-600)+\left(\frac{1}{3}\right) u(0)>u\left(\left(\frac{2}{3}\right)(-600)+\frac{1}{3}(0)\right)=u(-400)
$$

This implies, of course, that the person chooses program II over program I.

## PROBLEMS

## Group A

1 Explain how prospect theory and/or framing explains the Allais Paradox. (See Problem 11 of Section 13.2.)
2 Suppose a decision maker has a utility function $u(x)=x^{1 / 3}$. We flip a fair coin and receive $\$ 10$ for heads and $\$ 0$ for tails.
a Using expected utility theory, determine the certainty equivalent of this lottery.
b Using $\Pi(p)=1.89799 p-3.55995 p^{2}+$ $2.662549 p^{3}$, use prospect theory to determine the certainty equivalent of the lottery.
c Intuively explain why your answer in part (b) is smaller than your answer in part (a).
d What implications does this problem have for the method used in Section 13.2 to estimate a person's utility function?
3 You are given a choice between lottery 1 and lottery 2. You are also given a choice between lottery 3 and lottery 4 .

Lottery 1: A sure gain of $\$ 240$
Lottery 2: $\quad 25 \%$ chance to gain $\$ 1,000$ and $75 \%$ chance to gain nothing
Lottery 3: A sure loss of $\$ 750$
Lottery 4: A $75 \%$ chance to lose $\$ 1,000$ and a $25 \%$ chance of losing nothing
$84 \%$ of all people prefer lottery 1 over lottery 2 , and $87 \%$ choose lottery 4 over lottery 3 .
a Explain why the choice of lottery 1 over lottery 2 and lottery 4 over lottery 3 contradicts expected utility maximization. (Hint: Compare lottery $1+$ lottery 4 to lottery $2+$ lottery 3 .)
b Can you explain this anomalous behavior?
4 Tversky and Kahneman asked 72 respondents to choose between lottery 1 and lottery 2 and lottery 3 and lottery 4.

Lottery 1: A . 001 chance at winning $\$ 5,000$ and a .999 chance of winning $\$ 0$
Lottery 2: A sure gain of $\$ 5$
Lottery 3: A . 001 chance of losing $\$ 5,000$ and a .999 chance of losing $\$ 0$
Lottery 4: A sure loss of $\$ 5$
More than $75 \%$ of all participants preferred lottery 1 to lottery 2 and lottery 4 to lottery 3 .
a Which choices would be made by a risk-averse decision maker?
b Which choices would be made by a risk-seeking decision maker?
c How does the observed behavior of the participants contradict expected utility maximization?
d How does prospect theory resolve the contradiction?

### 13.4 Decision Trees

Often, people must make a series of decisions at different points in time. Then decision trees can be used to determine optimal decisions. A decision tree enables a decision maker to decompose a large complex decision problem into several smaller problems.

## EXAMPLE $3 \quad$ Colaco Marketing

Colaco currently has assets of $\$ 150,000$ and wants to decide whether to market a new chocolate-flavored soda, Chocola. Colaco has three alternatives:

Alternative 1 Test market Chocola locally, then utilize the results of the market study to determine whether or not to market Chocola nationally.

Alternative 2 Immediately (without test marketing) market Chocola nationally.
Alternative 3 Immediately (without test marketing) decide not to market Chocola nationally.

In the absence of a market study, Colaco believes that Chocola has a $55 \%$ chance of being a national success and a $45 \%$ chance of being a national failure. If Chocola is a national success, Colaco's asset position will increase by $\$ 300,000$, and if Chocola is a national failure, Colaco's asset position will decrease by $\$ 100,000$.

If Colaco performs a market study (at a cost of $\$ 30,000$ ), there is a $60 \%$ chance that the study will yield favorable results (referred to as a local success) and a $40 \%$ chance that the study will yield unfavorable results (referred to as a local failure). If a local success is observed, there is an $85 \%$ chance that Chocola will be a national success. If a local failure is observed, there is only a $10 \%$ chance that Chocola will be a national success. If Colaco is risk-neutral (wants to maximize its expected final asset position), what strategy should the company follow?

Solution To draw a decision tree that represents Colaco's problem, we begin at the present and proceed toward future events and decisions. The decision tree in Figure 7 is constructed with two kinds of forks: decision forks (denoted by $\square$ ) and event forks (denoted by $\bigcirc$ ).

A decision fork represents a point in time when Colaco has to make a decision. Each branch emanating from a decision fork represents a possible decision. An example of a decision fork occurs when Colaco must determine whether or not to test market Chocola.

Test market Chocola

Don't test market Chocola
An event fork is drawn when outside forces determine which of several random events will occur. Each branch of an event fork represents a possible outcome, and the number on each branch represents the probability that the event will occur. For example, if Colaco decides to test market Chocola, the company faces the following event fork when observing the results of the test market study:


A branch of a decision tree is a terminal branch if no forks emanate from the branch. Thus, the branches indicating National success and National failure are terminal branches of Colaco's decision tree. Since we are maximizing expected final asset position at each terminal branch, we must enter the final asset position that will result if the path leading to the given terminal branch occurs. For example, the terminal branch National failure that follows Local failure leads to a final asset position of $150,000-30,000-100,000=\$ 20,000$. If we were maximizing expected revenues, we would enter revenues on each terminal branch.

To determine the decisions that will maximize Colaco's expected final asset position, we work backward (sometimes called "folding back the tree") from right to left. ${ }^{\dagger}$ At each

[^17]FIGURE 7
Colaco's Decision Tree (Risk-Neutral)

event fork, we calculate the expected final asset position and enter it in $\bigcirc$. At each decision fork, we denote by $\|$ the decision that maximizes the expected final asset position and enter the expected final asset position associated with that decision in $\square$. We continue working backward in this fashion until we reach the beginning of the tree. Then the optimal sequence of decisions can be obtained by following the $\|$.

We begin by determining the expected final asset positions for the following three event forks:

1 Market nationally after Local success. Here we have an expected final asset position of $.85(420,000)+.15(20,000)=\$ 360,000$.

2 Market nationally after Local failure. Here we have an expected final asset position of $.10(420,000)+.90(20,000)=\$ 60,000$.

3 Market nationally after Don't test market. Here we have an expected final asset position of $.55(450,000)+.45(50,000)=\$ 270,000$.

We may now evaluate three decision forks:
1 Decision after Local success. Market nationally yields a larger expected final asset position than Don't market nationally, so we $\|$ Market nationally and enter an expected final asset position of $\$ 360,000$.

2 Decision after Local failure. Don't market nationally yields a larger expected final asset position than Market nationally, so we $\|$ Don't market nationally and enter an expected final asset position of $\$ 120,000$.

3 Decision for Don't test market. Market nationally yields a larger expected final asset position than Don't market nationally, so we $\|$ Market nationally and enter an expected final asset position of $\$ 270,000$.

We now must evaluate the event fork emanating from the Test market decision. This event fork yields an expected final asset position of $.60(360,000)+.40(120,000)=$ $\$ 264,000$, which is entered in $\bigcirc$.

All that remains is to determine the correct decision at the decision fork Test market versus Don't test market. We have found that Test market yields an expected final asset position of $\$ 264,000$, and Don't test market yields an expected final asset position of $\$ 270,000$. Thus, we $\|$ Don't test market and enter $\$ 270,000$ in

We have now reached the beginning of the tree and have found that Colaco's optimal decision is Don't test market and then Market nationally. This strategy will yield an expected final asset position of $\$ 270,000$. Observe that the decision tree also tells us that if we had test marketed and then acted optimally (Market nationally after Local success and Don't market nationally after Local failure), we would have obtained an expected final asset position of $\$ 264,000$.

## Incorporating Risk Aversion into Decision Tree Analysis

Note that Colaco's optimal strategy yields a .45 chance that the company will end up with a relatively small final asset position of $\$ 50,000$. On the other hand, the strategy of test marketing and acting optimally on the results of the test market study yields only a $(.60)(.15)=.09$ chance that Colaco's asset position will be below $\$ 100,000$. (Why?) Thus, if Colaco is a risk-averse decision maker, the strategy of immediately marketing nationally may not reflect the company's preference.

To illustrate how risk aversion may be incorporated into decision tree analysis, suppose that Colaco has the risk-averse utility function $u(x)$ in Figure $8(x=$ final asset position). (How do we know that this utility function exhibits risk aversion?) To determine Colaco's optimal decisions (that is, the decisions that maximize expected utility), simply replace each final asset position $x_{0}$ with its utility $u\left(x_{0}\right)$. Then at each event fork, compute the expected utility of Colaco's final asset position, and at each decision fork, choose the branch having the largest expected utility.

FIGURE 8
Colaco's Utility Function


We find from Figure 8 that $u(\$ 450,000)=1, u(\$ 420,000)=.99, u(\$ 150,000)=.48$, $u(\$ 120,000)=.40, u(\$ 50,000)=.19$, and $u(\$ 20,000)=0$. Substituting these values into the decision tree of Figure 7 yields the decision tree in Figure 9. We compute the expected utility at the following three event forks:

1 Market nationally after Local success. Here we have an expected utility of .85(.99) + $.15(0)=.8415$.

2 Market nationally after Local failure. Here we have an expected utility of .10(.99) + $.90(0)=.099$.

3 Market nationally after Don't test market. Here we have an expected utility of .55(1) + $.45(.19)=.6355$.

We may now evaluate three decision forks:
1 Decision after Local success. Market nationally yields a larger expected utility than Don't market nationally, so for this fork we || Market nationally and enter an expected utility of .8415 .

2 Decision after Local failure. Don't market nationally yields a larger expected utility than Market nationally, so for this fork we \| Don't market nationally and enter an expected utility of .40 .

3 Decision for Don't test market. Market nationally yields a larger expected utility than Don't market nationally, so for this fork we || Market nationally and enter an expected utility of . 6355 .

We now must evaluate the event fork emanating from the Test market decision. This event fork yields an expected utility of $.60(.8415)+.40(.40)=.6649$, which is entered in $\bigcirc$. All that remains is to determine the correct decision at the decision fork Test market versus Don't test market. We know that Test market yields an expected utility of .6649,

and Don't test market yields an expected utility of .6355, so we \| Test market and enter an expected utility of . 6649 in $\square$.

We have now reached the beginning of the tree and have found that Colaco's optimal decision is to begin by test marketing. If a local success is observed, then Colaco should market Chocola nationally; if a local failure is observed, then Colaco should not market Chocola nationally. This optimal strategy yields only a $.60(.15)=.09$ chance that Colaco will have a final asset position of less than $\$ 100,000$. This reflects the risk-averse nature of the utility function in Figure 8. Also, we see from Figure 8 that $u(\$ 226,000)=.665$. Since Colaco views the current situation as having an expected utility of .6649 , this means that the company considers the current situation equivalent to a certain asset position of $\$ 226,000$. Thus, if somebody offered to pay more than $226,000-150,000=\$ 76,000$ to buy the rights to Chocola, Colaco should take the offer. This is because receiving more than $\$ 76,000$ for the rights to Chocola would bring Colaco's asset position to more than $150,000+76,000=\$ 226,000$, and this situation has a higher expected utility than .665 .

## Expected Value of Sample Information

Decision trees can be used to measure the value of sample or test market information. To illustrate how this is done, we again assume that Colaco is risk-neutral. What is the value of the information that would be obtained by test marketing Chocola?

We begin by determining Colaco's expected final asset position if the company acts optimally and the test market study is costless. We call this expected final asset position Colaco's expected value with sample information (EVWSI). From Figure 7, we see that if we Test market and then act optimally, we will now have an expected final asset position of $264,000+30,000=\$ 294,000$. Since $\$ 294,000$ is larger than the expected asset position of the Don't test market branch $(\$ 270,000)$, we find that EVWSI $=\$ 294,000$.

We next determine the largest expected final asset position that Colaco would obtain if the test market study were not available. We call this the expected value with original information (EVWOI). From the Don't test market branch of Figure 7, we find EVWOI $=$ $\$ 270,000$. Now the expected value of the test market information, referred to as expected value of sample information (EVSI), is defined to be EVSI = EVWSI - EVWOI.

In the Colaco example, EVSI is the most that Colaco can pay for the test market information and still be at least as well off as without the test market information. Thus, for the Colaco example, EVSI $=294,000-270,000=\$ 24,000$. Since the cost of the test market study $(\$ 30,000)$ exceeds EVSI, Colaco should not (as we already know) conduct the test market study.

## Expected Value of Perfect Information

We can modify the analysis used to determine EVSI to find the value of perfect information. By perfect information we mean that all uncertain events that can affect Colaco's final asset position still occur with the given probabilities (so there is still a .55 chance of Chocola being a national success and a .45 chance that Chocola will be a national failure), but Colaco finds out whether Chocola is a national success or a national failure before making the decision to market Chocola nationally or not. This information can then be used to determine Colaco's optimal marketing strategy. Thus, expected value with perfect information (EVWPI) is found by drawing a decision tree in which the decision maker has perfect information about which state has occurred before making a decision. Then the expected value of perfect information (EVPI) is given by EVPI = EVWPI EVWOI.

FIGURE 10 Expected Value with Perfect Information (EVWPI) for Colaco


For the Colaco example, we find from Figure 10 that EVWPI $=\$ 315,000$. Then EVPI $=$ $315,000-270,000=\$ 45,000$. Thus, a perfect (one that was always correct) test marketing study would be worth $\$ 45,000$. EVPI is a useful upper bound on the value of sample or test market information; that is, no sample or test market information (no matter how good) can be worth more than $\$ 45,000$.

## EXAMPLE 4 Art Dealer

An art dealer's client is willing to buy the painting Sunplant at $\$ 50,000$. The dealer can buy the painting today for $\$ 40,000$ or can wait a day and buy the painting tomorrow (if it has not been sold) for $\$ 30,000$. The dealer may also wait another day and buy the painting (if it is still available) for $\$ 26,000$. At the end of the third day, the painting will no longer be available for sale. Each day, there is a .60 probability that the painting will be sold. What strategy maximizes the dealer's expected profit?

Solution
The decision tree for this example is given in Figure 11. The key to drawing this decision tree is that each day, the dealer must choose between buying the painting and waiting another day. Of course, waiting might mean that the dealer may never be able to buy the painting. As we see from the decision tree, the dealer should buy the painting on the first day.

FIGURE 11 Decision Tree for Example 4


## PROBLEMS

## Group A

1 Oilco must determine whether or not to drill for oil in the South China Sea. It costs $\$ 100,000$, and if oil is found, the value is estimated to be $\$ 600,000$. At present, Oilco believes there is a $45 \%$ chance that the field contains oil. Before drilling, Oilco can hire (for $\$ 10,000$ ) a geologist to obtain more information about the likelihood that the field will contain oil. There is a $50 \%$ chance that the geologist will issue a favorable report and a $50 \%$ chance of an unfavorable report. Given a favorable report, there is an
$80 \%$ chance that the field contains oil. Given an unfavorable report, there is a $10 \%$ chance that the field contains oil. Determine Oilco's optimal course of action. Also determine EVSI and EVPI.

2 The decision sciences department is trying to determine which of two copying machines to purchase. Both machines will satisfy the department's needs for the next ten years. Machine 1 costs $\$ 2,000$ and has a maintenance agreement,

## Game Theory

In previous chapters, we have encountered many situations in which a single decision maker chooses an optimal decision without reference to the effect that the decision has on other decision makers (and without reference to the effect that the decisions of others have on him or her). In many business situations, however, two or more decision makers simultaneously choose an action, and the action chosen by each player affects the rewards earned by the other players. For example, each fast-food company must determine an advertising and pricing policy for its product, and each company's decision will affect the revenues and profits of other fast-food companies.

Game theory is useful for making decisions in cases where two or more decision makers have conflicting interests. Most of our study of game theory deals with situations where there are only two decision makers (or players), but we briefly study $n$-person (where $n>2$ ) game theory also. We begin our study of game theory with a discussion of two-player games in which the players have no common interest.

### 14.1 Two-Person Zero-Sum and Constant-Sum Games: Saddle Points Characteristics of Two-Person Zero-Sum Games

1 There are two players (called the row player and the column player).
2 The row player must choose 1 of $m$ strategies. Simultaneously, the column player must choose 1 of $n$ strategies.

3 If the row player chooses his $i$ th strategy and the column player chooses his $j$ th strategy, then the row player receives a reward of $a_{i j}$ and the column player loses an amount $a_{i j}$. Thus, we may think of the row player's reward of $a_{i j}$ as coming from the column player.

Such a game is called a two-person zero-sum game, which is represented by the matrix in Table 1 (the game's reward matrix). As previously stated, $a_{i j}$ is the row player's

TABLE 1
Example of Two-Person Zero-Sum Game

| Row Player's <br> Strategy | Column Player's Strategy |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
|  | Column 1 | Column 2 | $\cdots$ | Column $n$ |
|  | $a_{11}$ | $a_{12}$ | $\cdots$ | $a_{1 n}$ |
| Row 2 | $a_{21}$ | $a_{22}$ | $\cdots$ | $a_{2 n}$ |
| $\vdots$ | $\vdots$ | $\vdots$ |  | $\vdots$ |
| Row $m$ | $a_{m 1}$ | $a_{m 2}$ | $\cdots$ | $a_{m n}$ |

reward (and the column player's loss) if the row player chooses his $i$ th strategy and the column player chooses his $j$ th column strategy.

For example, in the two-person zero-sum game in Table 2, the row player would receive two units (and the column player would lose two units) if the row player chose his second strategy and the column player chose his first strategy.

A two-person zero-sum game has the property that for any choice of strategies, the sum of the rewards to the players is zero. In a zero-sum game, every dollar that one player wins comes out of the other player's pocket, so the two players have totally conflicting interests. Thus, cooperation between the two players would not occur.

John von Neumann and Oskar Morgenstern developed a theory of how two-person zero-sum games should be played, based on the following assumption.

## Basic Assumption of Two-Person Zero-Sum Game Theory

Each player chooses a strategy that enables him to do the best he can, given that his opponent knows the strategy he is following. Let's use this assumption to determine how the row and column players should play the two-person zero-sum game in Table 3.

How should the row player play this game? If he chooses row 1 , then the assumption implies that the column player will choose column 1 or column 2 and hold the row player to a reward of four units (the smallest number in row 1 of the game matrix). Similarly, if the row player chooses row 2 , then the assumption implies that the column player will choose column 3 and hold the row player's reward to one unit (the smallest or minimum number in the second row of the game matrix). If the row player chooses row 3 , then he will be held to the smallest number in the third row (5). Thus, the assumption implies that the row player should choose the row having the largest minimum. Because max $(4,1,5)=5$, the row player should choose row 3. By choosing row 3, the row player can ensure that he will win at least max (row minimum) $=$ five units.

From the column player's viewpoint, if he chooses column 1, then the row player will choose the strategy that makes the column player's losses as large as possible (and the row player's winnings as large as possible). Thus, if the column player chooses column 1, then the row player will choose row 3 (because the largest number in the first column is the 6 in the third row). Similarly, if the column player chooses column 2, then the row player will again choose row 3 , because $5=\max (4,3,5)$. Finally, if the column player chooses column 3, the row player will choose row 1 , causing the column player to lose $10=\max$

TABLE 2

| 1 | 2 | 3 | -1 |
| ---: | ---: | ---: | ---: |
| 2 | 1 | -2 | 0 |

table 3
A Game with a Saddle Point

| Row Player's <br> Strategy | Column Player's Strategy |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
|  | Column 1 | Column 2 | Column 3 | Row <br> Minimum |
| Row 1 | 4 | 4 | 10 | 4 |
| Row 2 | 2 | 3 | 1 | 1 |
| Row 3 | 6 | 5 | 7 | 5 |
| Column <br> Maximum | 6 | 5 | 10 |  |

$(10,1,7)$ units. Thus, the column player can hold his losses to min $($ column maximum $)=$ $\min (6,5,10)=5$ by choosing column 2 .

We have shown that the row player can ensure that he will win at least five units and the column player can hold the row player's winnings to at most five units. Thus, the only rational outcome of this game is for the row player to win exactly five units; the row player cannot expect to win more than five units, because the column player (by choosing column 2) can hold the row player's winnings to five units.

The game matrix we have just analyzed has the property of satisfying the saddle point condition:

$$
\begin{equation*}
\max _{\substack{\text { all } \\ \text { rows }}}(\text { row minimum })=\min _{\substack{\text { all } \\ \text { columns }}}(\text { column maximum }) \tag{1}
\end{equation*}
$$

Any two-person zero-sum game satisfying (1) is said to have a saddle point. If a twoperson zero-sum game has a saddle point, then the row player should choose any strategy (row) attaining the maximum on the left side of (1). The column player should choose any strategy (column) attaining the minimum on the right side of (1). Thus, for the game we have just analyzed, a saddle point occurred where the row player chose row 3 and the column player chose column 2 . The row player could make sure of receiving a reward of at least five units (by choosing the optimal strategy of row 3), and the column player could ensure that the row player would receive a reward of at most five units (by choosing the optimal strategy of column 2). If a game has a saddle point, then we call the common value of both sides of (1) the value $(v)$ of the game to the row player. Thus, this game has a value of 5 .

An easy way to spot a saddle point is to observe that the reward for a saddle point must be the smallest number in its row and the largest number in its column (see Problem 4 at the end of this section). Thus, like the center point of a horse's saddle, a saddle point for a two-person zero-sum game is a local minimum in one direction (looking across the row) and a local maximum in another direction (looking up and down the column).

A saddle point can also be thought of as an equilibrium point in that neither player can benefit from a unilateral change in strategy. For example, if the row player were to change from the optimal strategy of row 3 (to either row 1 or row 2), his reward would decrease, while if the column player changed from his optimal strategy of column 2 (to either column 1 or column 3 ), the row player's reward (and the column player's losses) would increase. Thus, a saddle point is stable in that neither player has an incentive to move away from it.

Many two-person zero-sum games do not have saddle points. For example, the game in Table 4 does not have a saddle point, because

$$
\max (\text { row minimum })=-1<\min (\text { column maximum })=+1
$$

In Sections 14.2 and 14.3, we explain how to find the value and the optimal strategies for two-person zero-sum games that do not have saddle points.
table 4
A Game with No Saddle Point

| Row Player's <br> Strategy | Column Player's Strategy |  |  |
| :--- | :---: | :---: | :---: |
| Column 1 | Column 2 | Row Minimum |  |
| Row 1 | -1 | +1 | -1 |
| Row 2 | +1 | -1 | -1 |
| Column | +1 | +1 |  |
| Maximum |  |  |  |

## Two-Person Constant-Sum Games

Even if a two-person game is not zero-sum, two players can still be in total conflict. To illustrate this, we now consider two-person constant-sum games.

A two-person constant-sum game is a two-player game in which, for any choice of both player's strategies, the row player's reward and the column player's reward add up to a constant value $c$.

Of course, a two-person zero-sum game is just a two-person constant-sum game with $c=0$. A two-person constant-sum game maintains the feature that the row and column players are in total conflict, because a unit increase in the row player's reward will always result in a unit decrease in the column player's reward. In general, the optimal strategies and value for a two-person constant-sum game may be found by the same methods used to find the optimal strategies and value for a two-person zero-sum game.

## EXAMPLE 1 Constant Sum TV Game

During the 8 to 9 P.M. time slot, two networks are vying for an audience of 100 million viewers. The networks must simultaneously announce the type of show they will air in that time slot. The possible choices for each network and the number of network 1 viewers (in millions) for each choice are shown in Table 5. For example, if both networks choose a western, the matrix indicates that 35 million people will watch network 1 and $100-35=65$ million people will watch network 2 . Thus, we have a two-person constant-sum game with $c=100$ (million). Does this game have a saddle point? What is the value of the game to network 1 ?

Solution Looking at the row minima, we find that by choosing a soap opera, network 1 can be sure of at least $\max (15,45,14)=45$ million viewers. Looking at the column maxima, we find that by choosing a western, network 2 can hold network 1 to at most min $(45,58$, 70) $=45$ million viewers. Because

$$
\max (\text { row minimum })=\min (\text { column maximum })=45
$$

we find that Equation (1) is satisfied. Thus, network 1's choosing a soap opera and network 2's choosing a western yield a saddle point; neither side will do better if it unilaterally changes strategy (check this). Thus, the value of the game to network 1 is 45 million viewers, and the value of the game to network 2 is $100-45=55$ million viewers. The optimal strategy for network 1 is to choose a soap opera, and the optimal strategy for network 2 is to choose a western.
tABLE 5
A Constant-Sum Game

|  | Network 2 |  |  | Row <br> Network 1 |
| :--- | :---: | :---: | :---: | :---: |
|  | Western | Soap Opera | Comedy | Minimum |
| Western | 35 | 15 | 60 | 15 |
| Soap Opera | 45 | 58 | 50 | 45 |
| Comedy | 38 | 14 | 70 | 14 |
| Column | 45 | 58 | 70 |  |
| Maximum |  |  |  |  |

## PROBLEMS

## Group A

1 Find the value and optimal strategy for the game in Table 6.
2 Find the value and the optimal strategies for the twoperson zero-sum game in Table 7.

## Group B

3 Mad Max wants to travel from New York to Dallas by the shortest possible route. He may travel over the routes shown in Table 8. Unfortunately, the Wicked Witch can block one road leading out of Atlanta and one road leading out of Nashville. Mad Max will not know which roads have been blocked until he arrives at Atlanta or Nashville. Should Mad Max start toward Atlanta or Nashville? Which routes should the Wicked Witch block?

## TABLE 6

## $2 \quad 2$

13

## Group C

4 Explain why the reward for a saddle point must be the smallest number in its row and the largest number in its column. Suppose a reward is the smallest in its row and the largest in its column. Must that reward yield a saddle point? (Hint: Think about the idea of weak duality discussed in Chapter 6.)
tABLE 8

| Route | Length of Route <br> (Miles) |
| :--- | :---: |
| New York-Atlanta | 800 |
| New York-Nashville | 900 |
| Nashville-St. Louis | 400 |
| Nashville-New Orleans | 200 |
| Atlanta-St. Louis | 300 |
| Atlanta-New Orleans | 600 |
| St. Louis-Dallas | 500 |
| New Orleans-Dallas | 300 |

## TABLE 7

| 4 | 5 | 5 | 8 |
| :--- | :--- | :--- | :--- |
| 6 | 7 | 6 | 9 |
| 5 | 7 | 5 | 4 |
| 6 | 6 | 5 | 5 |

### 14.2 Two-Person Zero-Sum Games: Randomized Strategies, Domination, and Graphical Solution

In the previous section, we found that not all two-person zero-sum games have saddle points. We now discuss how to find the value and optimal strategies for a two-person zerosum game that does not have a saddle point. We begin with the simple game of Odds and Evens.

## EXAMPLE 2 Odds and Evens

Two players (called Odd and Even) simultaneously choose the number of fingers (1 or 2) to put out. If the sum of the fingers put out by both players is odd, then Odd wins $\$ 1$ from Even. If the sum of the fingers is even, then Even wins $\$ 1$ from Odd. We consider the row player to be Odd and the column player to be Even. Determine whether this game has a saddle point.

Solution This is a zero-sum game, with the reward matrix shown in Table 9. Because max (row minimum $)=-1$ and $\min ($ column maximum $)=+1$, Equation (1) is not satisfied, and this game has no saddle point. All we know is that Odd can be sure of a reward of at least
table 9
Reward Matrix for Odds and Evens

| Row Player <br> (Odd) | Column Player (Even) |  |  |
| :--- | :---: | :---: | :---: |
|  | 1 Finger | 2 Fingers | Row Minimum |
| 1 Finger | -1 | +1 | -1 |
| 2 Fingers | +1 | -1 | -1 |
| Column | +1 | +1 |  |
| Maximum |  |  |  |

-1 , and Even can hold Odd to a reward of at most +1 . Thus, it is unclear how to determine the value of the game and the optimal strategies. Observe that for any choice of strategies by both players, there is a player who can benefit by unilaterally changing her strategy. For example, if both players put out one finger, then Odd could have increased her reward from -1 to +1 by putting out two fingers. Thus, no choice of strategies by the player is stable. We now determine optimal strategies and the value for this game.

## Randomized or Mixed Strategies

To progress further with the analysis of Example 2 (and other games without saddle points), we must expand the set of allowable strategies for each player to include randomized strategies. Until now, we have assumed that each time a player plays a game, the player will choose the same strategy. Why not allow each player to select a probability of playing each strategy? For Example 2, we might define

$$
\begin{aligned}
& x_{1}=\text { probability that Odd puts out one finger } \\
& x_{2}=\text { probability that Odd puts out two fingers } \\
& y_{1}=\text { probability that Even puts out one finger } \\
& y_{2}=\text { probability that Even puts out two fingers }
\end{aligned}
$$

If $x_{1} \geq 0, x_{2} \geq 0$, and $x_{1}+x_{2}=1$, then $\left(x_{1}, x_{2}\right)$ is a randomized, or mixed, strategy for Odd. For example, the mixed strategy $\left(\frac{1}{2}, \frac{1}{2}\right)$ could be realized by Odd if she tossed a coin before each play of the game and put out one finger for heads and two fingers for tails. Similarly, if $y_{1} \geq 0, y_{2} \geq 0$, and $y_{1}+y_{2}=1$, then $\left(y_{1}, y_{2}\right)$ is a mixed strategy for Even.

Any mixed strategy $\left(x_{1}, x_{2}, \ldots, x_{m}\right)$ for the row player is a pure strategy if any of the $x_{i}$ equals 1 . Similarly, any mixed strategy $\left(y_{1}, y_{2}, \ldots, y_{n}\right)$ for the column player is a pure strategy if any of the $y_{i}$ equals 1 . A pure strategy is a special case of a mixed strategy in which a player always chooses the same action. Recall from Section 14.1 that the game in Table 10 had a value of 5 (corresponding to a saddle point), so the row player's optimal strategy could be represented as the pure strategy $(0,0,1)$, and the column player's optimal strategy could be represented as the pure strategy $(0,1,0)$.

We continue to assume that both players will play two-person zero-sum games in accordance with the basic assumption of Section 14.1. In the context of randomized strate-
gies, the assumption (from the standpoint of Odd) may be stated as follows: Odd should choose $x_{1}$ and $x_{2}$ to maximize her expected reward under the assumption that Even knows the value of $x_{1}$ and $x_{2}$.

It is important to realize that even though we assume that Even knows the values of $x_{1}$ and $x_{2}$, on a particular play of the game, she is not assumed to know Odd's actual strategy choice until the instant the game is played.

## Graphical Solution of Odds and Evens

## Finding Odd's Optimal Strategy

With this version of the basic assumption, we can determine the optimal strategy for Odd. Because $x_{1}+x_{2}=1$, we know that $x_{2}=1-x_{1}$. Thus, any mixed strategy may be written as $\left(x_{1}, 1-x_{1}\right)$, and it suffices to determine the value of $x_{1}$. Suppose Odd chooses a particular mixed strategy $\left(x_{1}, 1-x_{1}\right)$. What is Odd's expected reward against each of Even's strategies? If Even puts out one finger, then Odd will receive a reward of -1 with probability $x_{1}$ and a reward of +1 with probability $x_{2}=1-x_{1}$. Thus, if Even puts out one finger and Odd chooses the mixed strategy $\left(x_{1}, 1-x_{1}\right)$, then Odd's expected reward is

$$
(-1) x_{1}+(+1)\left(1-x_{1}\right)=1-2 x_{1}
$$

As a function of $x_{1}$, this expected reward is drawn as line segment $A C$ in Figure 1. Similarly, if Even puts out two fingers and Odd chooses the mixed strategy ( $x_{1}, 1-x_{1}$ ), Odd's expected reward is

$$
(+1)\left(x_{1}\right)+(-1)\left(1-x_{1}\right)=2 x_{1}-1
$$

which is line segment $D E$ in Figure 1.
Suppose Odd chooses the mixed strategy $\left(x_{1}, 1-x_{1}\right)$. Because Even is assumed to know the value of $x_{1}$, for any value of $x_{1}$ Even will choose the strategy (putting out one or two fingers) that yields a smaller expected reward for Odd. From Figure 1, we see that, as a function of $x_{1}$, Odd's expected reward will be given by the $y$-coordinate in $D B C$. Odd wants to maximize her expected reward, so she should choose the value of $x_{1}$ corresponding to point $B$. Point $B$ occurs where the line segments $A C$ and $D E$ intersect, or

FIGURE 1

where $1-2 x_{1}=2 x_{1}-1$. Solving this equation, we obtain $x_{1}=\frac{1}{2}$. Thus, Odd should choose the mixed strategy $\left(\frac{1}{2}, \frac{1}{2}\right)$. The reader should verify that against each of Even's strategies, $\left(\frac{1}{2}, \frac{1}{2}\right)$ yields an expected reward of zero. Thus, zero is a floor on Odd's expected reward, because by choosing the mixed strategy $\left(\frac{1}{2}, \frac{1}{2}\right)$, Odd can be sure that (for any choice of Even's strategy) her expected reward will always be at least zero.

## Finding Even's Optimal Strategy

We now consider how Even should choose a mixed strategy ( $y_{1}, y_{2}$ ). Again, because $y_{2}=1-y_{1}$, we may ask how Even should choose a mixed strategy $\left(y_{1}, 1-y_{1}\right)$. The basic assumption implies that Even should choose $y_{1}$ to minimize her expected losses (or, equivalently, minimize Odd's expected reward) under the assumption that Odd knows the value of $y_{1}$. Suppose Even chooses the mixed strategy $\left(y_{1}, 1-y_{1}\right)$. What will Odd do? If Odd puts out one finger, then her expected reward is

$$
(-1) y_{1}+(+1)\left(1-y_{1}\right)=1-2 y_{1}
$$

which is line segment $A C$ in Figure 2. If Odd puts out two fingers, then her expected reward is

$$
(+1)\left(y_{1}\right)+(-1)\left(1-y_{1}\right)=2 y_{1}-1
$$

which is line segment $D E$ in Figure 2. Because Odd is assumed to know the value of $y_{1}$, she will put out the number of fingers corresponding to $\max \left(1-2 y_{1}, 2 y_{1}-1\right)$. Thus, for a given value of $y_{1}$, Odd's expected reward (and Even's expected loss) will be given by the $y$-coordinate on the piecewise linear curve $A B E$.

Now Even chooses the mixed strategy $\left(y_{1}, 1-y_{1}\right)$ that will make Odd's expected reward as small as possible. Thus, Even should choose the value of $y_{1}$ corresponding to the lowest point on $A B E$ (point $B$ ). Point $B$ is where the line segments $A C$ and $D E$ intersect, or where $1-2 y_{1}=2 y_{1}-1$, or $y_{1}=\frac{1}{2}$. The basic assumption implies that Even should choose the mixed strategy $\left(\frac{1}{2}, \frac{1}{2}\right)$. For this mixed strategy, Even's expected loss (and Odd's expected reward) is zero. We say that zero is a ceiling on Even's expected loss

FIGURE 2 Choosing Even's Strategy


TABLE 11
How to Make a Nonoptimal Strategy Pay the Price

| Odd's Mixed <br> Strategy | Even Can <br> Choose | Odd's Expected Reward <br> (Even's expected losses) |
| :--- | :---: | :---: |
| $x_{1}<\frac{1}{2}$ | 2 fingers | $<0$ (on $B D$ in Figure 1) |
| $x_{1}>\frac{1}{2}$ | 1 finger | $<0($ on $B C$ in Figure 1) |
| Even's Mixed | Odd Can <br> Choose | Odd's Expected Reward <br> (Even's expected losses) |
| Strategy | 1 finger | $>0$ (on $A B$ in Figure 2) |
| $y_{1}<\frac{1}{2}$ | 2 fingers | $>0($ on $B E$ in Figure 2) |
| $y_{1}>\frac{1}{2}$ |  |  |

(or Odd's expected reward), because by choosing the mixed strategy $\left(\frac{1}{2}, \frac{1}{2}\right)$, Even can ensure that her expected loss (for any choice of strategies by Odd) will not exceed zero.

## More on the Idea of Value and Optimal Strategies

For the game of Odds and Evens, the row player's floor and the column player's ceiling are equal. This is not a coincidence. When each player is allowed to choose mixed strategies, the row player's floor will always equal the column player's ceiling. In Section 14.3, we use the Dual Theorem of Chapter 6 to prove this interesting result. We call the common value of the floor and ceiling the value of the game to the row player. Any mixed strategy for the row player that guarantees that the row player gets an expected reward at least equal to the value of the game is an optimal strategy for the row player. Similarly, any mixed strategy for the column player that guarantees that the column player's expected loss is no more than the value of the game is an optimal strategy for the column player. Thus, for Example 2, we have shown that the value of the game is zero, the row player's optimal strategy is $\left(\frac{1}{2}, \frac{1}{2}\right)$, and the column player's optimal strategy is $\left(\frac{1}{2}, \frac{1}{2}\right)$.

Example 2 illustrates that by allowing mixed strategies, we have enabled each player to find an optimal strategy in that if the row player departs from her optimal strategy, the column player may have a strategy that reduces the row player's expected reward below the value of the game, and if the column player departs from her optimal strategy, the row player may have a strategy that increases her expected reward above the value of the game. Table 11 illustrates this idea for the game of Odds and Evens.

For example, suppose that Odd chooses a nonoptimal mixed strategy with $x_{1}<\frac{1}{2}$. Then, by choosing two fingers, Even ensures that Odd's expected reward can be read from BD in Figure 1. This means that if Odd chooses a mixed strategy having $x_{1}<\frac{1}{2}$, then her expected reward can be negative (less than the value of the game).

To close this section, we find the value and optimal strategies for a more complicated game.

## EXAMPLE 3 Coin Toss Came with Bluffing

A fair coin is tossed, and the result is shown to player 1. Player 1 must then decide whether to pass or bet. If player 1 passes, then he must pay player $2 \$ 1$. If player 1 bets, then player 2 (who does not know the result of the coin toss) may either fold or call the bet. If player 2 folds, then she pays player $1 \$ 1$. If player 2 calls and the coin comes up heads, then she pays player $1 \$ 2$; if player 2 calls and the coin comes up tails, then player 1 must pay her
\$2. Formulate this as a two-person zero-sum game. Then graphically determine the value of the game and each player's optimal strategy.

Solution Player 1's strategies may be represented as follows: PP, pass on heads and pass on tails; $\mathbf{P B}$, pass on heads and bet on tails; BP, bet on heads and pass on tails; and $\mathbf{B B}$, bet on heads and bet on tails. Player 2 simply has the two strategies call and fold. For each choice of strategies, player 1's expected reward is as shown in Table 12.

To illustrate these computations, suppose player 1 chooses BP and player 2 calls. Then with probability $\frac{1}{2}$, heads is tossed. Then player 1 bets, is called, and wins $\$ 2$ from player 2. With probability $\frac{1}{2}$, tails is tossed. In this case, player 1 passes and pays player $2 \$ 1$. Thus, if player 1 chooses BP and player 2 calls, then player 1's expected reward is $\left(\frac{1}{2}\right)(2)+$ $\left(\frac{1}{2}\right)(-1)=\$ 0.50$. For each line in Table 12, the first term in the expectation corresponds to heads being tossed, and the second term corresponds to tails being tossed.

Example 3 may be described as the two-person zero-sum game represented by the reward matrix in Table 13. Because max (row minimum) $=0<\min ($ column maximum $)=$ $\frac{1}{2}$, this game does not have a saddle point. Observe that player 1 would be unwise ever to choose the strategy PP, because (for each of player 2's strategies) player 1 could do better than $\mathbf{P P}$ by choosing either $\mathbf{B P}$ or $\mathbf{B B}$. In general, a strategy $i$ for a given player is dominated by a strategy $i^{\prime}$ if, for each of the other player's possible strategies, the given player does at least as well with strategy $i^{\prime}$ as he or she does with strategy $i$, and if for at least one of the other player's strategies, strategy $i^{\prime}$ is superior to strategy $i$. A player may eliminate all dominated strategies from consideration. We have just shown that for player 1, BP

TABLE 12
Computation of Reward Matrix for Example 3

| Player 1's Expected Reward |  |  |
| :--- | :--- | :--- |
| PP vs. call | $\left(\frac{1}{2}\right)(-1)+\left(\frac{1}{2}\right)(-1)$ | $=-\$ 1$ |
| PP vs. fold | $\left(\frac{1}{2}\right)(-1)+\left(\frac{1}{2}\right)(-1)$ | $=-\$ 1$ |
| PB vs. call | $\left(\frac{1}{2}\right)(-1)+\left(\frac{1}{2}\right)(-2)$ | $=-\$ 1.50$ |
| PB vs. fold | $\left(\frac{1}{2}\right)(-1)+\left(\frac{1}{2}\right)(1)$ | $=\$ 0$ |
| BP vs. call | $\left(\frac{1}{2}\right)(2)+\left(\frac{1}{2}\right)(-1)$ | $=\$ 0.50$ |
| BP vs. fold | $\left(\frac{1}{2}\right)(1)+\left(\frac{1}{2}\right)(-1)$ | $=\$ 0$ |
| BB vs. call | $\left(\frac{1}{2}\right)(2)+\left(\frac{1}{2}\right)(-2)$ | $=\$ 0$ |
| BB vs. fold | $\left(\frac{1}{2}\right)(1)+\left(\frac{1}{2}\right)(1)$ | $=\$ 1$ |

TABLE 13
Reward Matrix for Example 3

|  | Player 2 |  |  |
| :--- | ---: | ---: | ---: |
| Player 1 | Call | Fold | Row Minimum |
| PP | -1 | -1 | -1 |
| PB | $-\frac{3}{2}$ | 0 | $-\frac{3}{2}$ |
| BP | $\frac{1}{2}$ | 0 | 0 |
| BB | 0 | 1 | 0 |
| Column | $\frac{1}{2}$ | 1 |  |
| Maximum |  |  |  |

or BB dominates PP. Similarly, the reader should be able to show that player 1's $\mathbf{P B}$ strategy is dominated by $\mathbf{B P}$ or $\mathbf{B B}$. After eliminating the dominated strategies $\mathbf{P P}$ and $\mathbf{P B}$, we are left with the game matrix shown in Table 14.

As with Odds and Evens, this game has no saddle point, and we proceed with a graphical solution. Let

$$
\begin{aligned}
& x_{1}=\text { probability that player } 1 \text { chooses } \mathbf{B P} \\
& x_{2}=1-x_{1}=\text { probability that player } 1 \text { chooses } \mathbf{B B} \\
& y_{1}=\text { probability that player } 2 \text { chooses call } \\
& y_{2}=1-y_{1}=\text { probability that player } 2 \text { chooses fold }
\end{aligned}
$$

To determine the optimal strategy for player 1 , observe that for any value of $x_{1}$, her expected reward against calling is

$$
\left(\frac{1}{2}\right)\left(x_{1}\right)+0\left(1-x_{1}\right)=\frac{x_{1}}{2}
$$

which is line segment $A B$ in Figure 3. Against folding, player 1's expected reward is

$$
0\left(x_{1}\right)+1\left(1-x_{1}\right)=1-x_{1}
$$

which is line segment $C D$ in Figure 3. Player 2 is assumed to know the value of $x_{1}$, so player 1's expected reward (as a function of $x_{1}$ ) is given by the piecewise linear curve $A E D$
tABLE 14
Reward Matrix for Example 3 After Dominated
Strategies Have Been Eliminated

|  | Player 2 |  |  |
| :--- | :---: | :---: | :---: |
| Player 1 | Call | Fold | Row Minimum |
| BP | $\frac{1}{2}$ | 0 | 0 |
| BB | 0 | 1 | 0 |
| Column | $\frac{1}{2}$ | 1 |  |
| Maximum |  |  |  |

FIGURE 3 How Player 1 Chooses Optimal Strategy in Example 3


FIGURE 4 How Player 2 Chooses Optimal Strategy in Example 3

in Figure 3. Thus, to maximize her expected reward, player 1 should choose the value of $x_{1}$ corresponding to point $E$, which solves $x_{1} / 2=1-x_{1}$, or $x_{1}=\frac{2}{3}$. Then $x_{2}=1-\frac{2}{3}=$ $\frac{1}{3}$, and player 1's expected reward against either of player 2's strategies is $\frac{x_{1}}{2}$ (or $1-$ $\left.x_{1}\right)=\frac{1}{3}$.

How should player 2 choose $y_{1}$ ? (Remember, $y_{2}=1-y_{1}$.) For a given value of $y_{1}$, suppose player 1 chooses $\mathbf{B P}$. Then her expected reward is

$$
\left(\frac{1}{2}\right)\left(y_{1}\right)+0\left(1-y_{1}\right)=\frac{y_{1}}{2}
$$

which is line segment $A B$ in Figure 4. For a given value of $y_{1}$, suppose player 1 chooses BB. Then her expected reward is

$$
0\left(y_{1}\right)+1\left(1-y_{1}\right)=1-y_{1}
$$

which is line segment $C D$ in Figure 4. Thus, for a given value of $y_{1}$, player 1 will choose a strategy that causes his expected reward to be given by the piecewise linear curve CEB in Figure 4. Knowing this, player 2 should choose the value of $y_{1}$ corresponding to point $E$ in Figure 4. The value of $y_{1}$ at point $E$ is the solution to $\frac{y_{1}}{2}=1-y_{1}$, or $y_{1}=\frac{2}{3}$ (and $y_{2}=\frac{1}{3}$ ). You should check that no matter what player 1 does, player 2's mixed strategy $\left(\frac{2}{3}, \frac{1}{3}\right)$ ensures that player 1 earns an expected reward of $\frac{1}{3}$.

In summary, the value of the game is $\frac{1}{3}$ to player 1 ; the optimal mixed strategy for player 1 is $\left(\frac{2}{3}, \frac{1}{3}\right)$; and the optimal strategy for player 2 is also $\left(\frac{2}{3}, \frac{1}{3}\right)$.

REMARKS 1 Observe that player 1 should bet $\frac{1}{3}$ of the time that she has a losing coin. Thus, our simple model indicates that player 1's optimal strategy includes bluffing.
2 In Problem 4 at the end of this section, it will be shown that if player 1 deviates from her optimal strategy, player 2 can hold her to an expected reward that is less than the value $\left(\frac{1}{3}\right)$ of the game. Similarly, Problem 5 will show that if player 2 deviates from her optimal strategy, player 1 can earn an expected reward in excess of the value $\left(\frac{1}{3}\right)$ of the game.
3 Although we have only applied the graphical method to games in which each player (after dominated strategies have been eliminated) has only two strategies, the graphical approach can be used to solve two-person zero-sum games in which only one player has two strategies (games in which the reward matrix is $2 \times n$ or $m \times 2$ ). We choose, however, to solve all non- $2 \times 2$ two-person games by the linear programming method outlined in the next section.


### 17.1 What Is a Stochastic Process?

Suppose we observe some characteristic of a system at discrete points in time (labeled 0 , $1,2, \ldots)$. Let $\mathbf{X}_{t}$ be the value of the system characteristic at time $t$. In most situations, $\mathbf{X}_{t}$ is not known with certainty before time $t$ and may be viewed as a random variable. A discrete-time stochastic process is simply a description of the relation between the random variables $\mathbf{X}_{0}, \mathbf{X}_{1}, \mathbf{X}_{2}, \ldots$ Some examples of discrete-time stochastic processes follow.

## EXAMPLE 1 The Cambler's Ruin

At time 0 , I have $\$ 2$. At times $1,2, \ldots$, I play a game in which I bet $\$ 1$. With probability $p$, I win the game, and with probability $1-p$, I lose the game. My goal is to increase my capital to $\$ 4$, and as soon as I do, the game is over. The game is also over if my capital is reduced to $\$ 0$. If we define $\mathbf{X}_{t}$ to be my capital position after the time $t$ game (if any) is played, then $\mathbf{X}_{0}, \mathbf{X}_{1}, \ldots, \mathbf{X}_{t}$ may be viewed as a discrete-time stochastic process. Note that $\mathbf{X}_{0}=2$ is a known constant, but $\mathbf{X}_{1}$ and later $\mathbf{X}_{t}$ 's are random. For example, with probability $p, \mathbf{X}_{1}=3$, and with probability $1-p, \mathbf{X}_{1}=1$. Note that if $\mathbf{X}_{t}=4$, then $\mathbf{X}_{t+1}$ and all later $\mathbf{X}_{t}$ 's will also equal 4. Similarly, if $\mathbf{X}_{t}=0$, then $\mathbf{X}_{t+1}$ and all later $\mathbf{X}_{t}$ 's will also equal 0 . For obvious reasons, this type of situation is called a gambler's ruin problem.

EXAMPLE $2 \quad$ Choosing Balls from an Urn
An urn contains two unpainted balls at present. We choose a ball at random and flip a coin. If the chosen ball is unpainted and the coin comes up heads, we paint the chosen unpainted ball red; if the chosen ball is unpainted and the coin comes up tails, we paint the chosen unpainted ball black. If the ball has already been painted, then (whether heads or tails has been tossed) we change the color of the ball (from red to black or from black to red). To model this situation as a stochastic process, we define time $t$ to be the time af-
ter the coin has been flipped for the $t$ th time and the chosen ball has been painted. The state at any time may be described by the vector $\left[\begin{array}{lll}u & r & b\end{array}\right]$, where $u$ is the number of unpainted balls in the urn, $r$ is the number of red balls in the urn, and $b$ is the number of black balls in the urn. We are given that $\mathbf{X}_{0}=\left[\begin{array}{ccc}2 & 0 & 0\end{array}\right]$. After the first coin toss, one ball will have been painted either red or black, and the state will be either $\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$ or $\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$. Hence, we can be sure that $\mathbf{X}_{1}=\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$ or $\mathbf{X}_{1}=\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$. Clearly, there must be some sort of relation between the $\mathbf{X}_{t}$ 's. For example, if $\mathbf{X}_{t}=\left[\begin{array}{lll}0 & 2 & 0\end{array}\right]$, we can be sure that $\mathbf{X}_{t+1}$ will be $\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]$.

## EXAMPLE 3 CSL Computer Stock

Let $\mathbf{X}_{0}$ be the price of a share of CSL Computer stock at the beginning of the current trading day. Also, let $\mathbf{X}_{t}$ be the price of a share of CSL stock at the beginning of the $t$ th trading day in the future. Clearly, knowing the values of $\mathbf{X}_{0}, \mathbf{X}_{1}, \ldots, \mathbf{X}_{t}$ tells us something about the probability distribution of $\mathbf{X}_{t+1}$; the question is, what does the past (stock prices up to time $t$ ) tell us about $\mathbf{X}_{t+1}$ ? The answer to this question is of critical importance in finance. (See Section 17.2 for more details.)

We close this section with a brief discussion of continuous-time stochastic processes. A continuous-time stochastic process is simply a stochastic process in which the state of the system can be viewed at any time, not just at discrete instants in time. For example, the number of people in a supermarket $t$ minutes after the store opens for business may be viewed as a continuous-time stochastic process. (Models involving continuous-time stochastic processes are studied in Chapter 20.) Since the price of a share of stock can be observed at any time (not just the beginning of each trading day), it may be viewed as a continuous-time stochastic process. Viewing the price of a share of stock as a continuoustime stochastic process has led to many important results in the theory of finance, including the famous Black-Scholes option pricing formula.

### 17.2 What Is a Markov Chain?

One special type of discrete-time stochastic process is called a Markov chain. To simplify our exposition, we assume that at any time, the discrete-time stochastic process can be in one of a finite number of states labeled $1,2, \ldots, s$.

DEFINITION A discrete-time stochastic process is a Markov chain if, for $t=0,1,2, \ldots$ and all states,

$$
\begin{align*}
P\left(\mathbf{X}_{t+1}\right. & \left.=i_{t+1} \mid \mathbf{X}_{t}=i_{t}, \mathbf{X}_{t-1}=i_{t-1}, \ldots, \mathbf{X}_{1}=i_{1}, \mathbf{X}_{0}=i_{0}\right) \\
& =P\left(\mathbf{X}_{t+1}=i_{t+1} \mid \mathbf{X}_{t}=i_{t}\right) \tag{1}
\end{align*}
$$

Essentially, (1) says that the probability distribution of the state at time $t+1$ depends on the state at time $t\left(i_{t}\right)$ and does not depend on the states the chain passed through on the way to $i_{t}$ at time $t$.

In our study of Markov chains, we make the further assumption that for all states $i$ and $j$ and all $t, P\left(\mathbf{X}_{t+1}=j \mid \mathbf{X}_{t}=i\right)$ is independent of $t$. This assumption allows us to write

$$
\begin{equation*}
P\left(\mathbf{X}_{t+1}=j \mid \mathbf{X}_{t}=i\right)=p_{i j} \tag{2}
\end{equation*}
$$

where $p_{i j}$ is the probability that given the system is in state $i$ at time $t$, it will be in a state $j$ at time $t+1$. If the system moves from state $i$ during one period to state $j$ during the next period, we say that a transition from $i$ to $j$ has occurred. The $p_{i j}$ 's are often referred to as the transition probabilities for the Markov chain.

Equation (2) implies that the probability law relating the next period's state to the current state does not change (or remains stationary) over time. For this reason, (2) is often called the Stationarity Assumption. Any Markov chain that satisfies (2) is called a stationary Markov chain.

Our study of Markov chains also requires us to define $q_{i}$ to be the probability that the chain is in state $i$ at time 0 ; in other words, $P\left(\mathbf{X}_{0}=i\right)=q_{i}$. We call the vector $\mathbf{q}=\left[q_{1}\right.$ $q_{2} \quad \cdots \quad q_{s}$ ] the initial probability distribution for the Markov chain. In most applications, the transition probabilities are displayed as an $s \times s$ transition probability matrix $P$. The transition probability matrix $P$ may be written as

$$
P=\left[\begin{array}{cccc}
p_{11} & p_{12} & \cdots & p_{1 s} \\
p_{21} & p_{22} & \cdots & p_{2 s} \\
\vdots & \vdots & & \vdots \\
p_{s 1} & p_{s 2} & \cdots & p_{s s}
\end{array}\right]
$$

Given that the state at time $t$ is $i$, the process must be somewhere at time $t+1$. This means that for each $i$,

$$
\begin{array}{r}
\sum_{j=1}^{j=s} P\left(\mathbf{X}_{t+1}=j \mid P\left(\mathbf{X}_{t}=i\right)\right)=1 \\
\sum_{j=1}^{j=s} p_{i j}=1
\end{array}
$$

We also know that each entry in the $P$ matrix must be nonnegative. Hence, all entries in the transition probability matrix are nonnegative, and the entries in each row must sum to 1 .

## EXAMPLE 1 The Gambler's Ruin (Continued)

Find the transition matrix for Example 1.
Solution Since the amount of money I have after $t+1$ plays of the game depends on the past history of the game only through the amount of money I have after $t$ plays, we definitely have a Markov chain. Since the rules of the game don't change over time, we also have a stationary Markov chain. The transition matrix is as follows (state $i$ means that we have $i$ dollars):

$P=$| State |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: |
| 0 |  |  |  |  |
| 1 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |\(\left[\begin{array}{ccccc}\$ 0 \& \$ 1 \& \$ 2 \& \$ 3 \& \$ 4 <br>

1 \& 0 \& 0 \& 0 \& 0 <br>
1-p \& 0 \& p \& 0 \& 0 <br>
0 \& 1-p \& 0 \& p \& 0 <br>
0 \& 0 \& 1-p \& 0 \& p <br>
0 \& 0 \& 0 \& 0 \& 1\end{array}\right]\)

If the state is $\$ 0$ or $\$ 4$, I don't play the game anymore, so the state cannot change; hence, $p_{00}=p_{44}=1$. For all other states, we know that with probability $p$, the next period's state will exceed the current state by 1 , and with probability $1-p$, the next period's state will be 1 less than the current state.

FIGURE 1 Graphical Representation of Transition Matrix for Gambler's Ruin


A transition matrix may be represented by a graph in which each node represents a state and $\operatorname{arc}(i, j)$ represents the transition probability $p_{i j}$. Figure 1 gives a graphical representation of Example 1's transition probability matrix.

## EXAMPLE 2 Choosing Balls (Continued)

Find the transition matrix for Example 2.
Solution Since the state of the urn after the next coin toss only depends on the past history of the process through the state of the urn after the current coin toss, we have a Markov chain. Since the rules don't change over time, we have a stationary Markov chain. The transition matrix for Example 2 is as follows:

State
$\left.P=\begin{array}{lll}{[0} & 1 & 1\end{array}\right]\left[\begin{array}{lll}{[0} & 2 & 0\end{array}\right]\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]\left[\begin{array}{lll}0 & 2 & 0\end{array}\right]\left[\begin{array}{lll}0 & 0 & 2\end{array}\right]\left[\begin{array}{llllllll}2 & 0 & 0\end{array}\right]\left[\begin{array}{lllll}1 & 1 & 0\end{array}\right]\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$

To illustrate the determination of the transition matrix, we determine the $\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$ row of this transition matrix. If the current state is $\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$, then one of the events shown in Table 1 must occur. Thus, the next state will be $\left[\begin{array}{ccc}1 & 0 & 1\end{array}\right]$ with probability $\frac{1}{2},\left[\begin{array}{ll}0 & 2\end{array}\right.$ $0]$ with probability $\frac{1}{4}$, and $\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]$ with probability $\frac{1}{4}$. Figure 2 gives a graphical representation of this transition matrix.

TABLE 1
Computations of Transition Probabilities If Current State Is $\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$

| Event | Probability | New State |  |
| :--- | :---: | :---: | :---: |
| Flip heads and choose unpainted ball | $\frac{1}{4}$ | $\left[\begin{array}{lll}0 & 2 & 0\end{array}\right]$ |  |
| Choose red ball | $\frac{1}{2}$ | $\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$ |  |
| Flip tails and choose unpainted ball | $\frac{1}{4}$ | $\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]$ |  |



In recent years, students of finance have devoted much effort to answering the question of whether the daily price of a stock share can be described by a Markov chain. Suppose the daily price of a stock share (such as CSL Computer stock) can be described by a Markov chain. What does that tell us? Simply that the probability distribution of tomorrow's price for one share of CSL stock depends only on today's price of CSL stock, not on the past prices of CSL stock. If the price of a stock share can be described by a Markov chain, the "chartists" who attempt to predict future stock prices on the basis of the patterns followed by past stock prices are barking up the wrong tree. For example, suppose the daily price of a share of CSL stock follows a Markov chain, and today's price for a share of CSL stock is $\$ 50$. Then to predict tomorrow's price of a share of CSL stock, it does not matter whether the price has increased or decreased during each of the last 30 days. In either situation (or any other situation that might have led to today's $\$ 50$ price), a prediction of tomorrow's stock price should be based only on the fact that today's price of CSL stock is $\$ 50$. At this time, the consensus is that for most stocks the daily price of the stock can be described as a Markov chain. This idea is often referred to as the efficient market hypothesis.

## PROBLEMS

## Group A

1 In Smalltown, $90 \%$ of all sunny days are followed by sunny days, and $80 \%$ of all cloudy days are followed by cloudy days. Use this information to model Smalltown's weather as a Markov chain.
2 Consider an inventory system in which the sequence of events during each period is as follows. (1) We observe the inventory level (call it $i$ ) at the beginning of the period.
(2) If $i \leq 1,4-i$ units are ordered. If $i \geq 2,0$ units are ordered. Delivery of all ordered units is immediate. (3) With probability $\frac{1}{3}, 0$ units are demanded during the period; with probability $\frac{1}{3}, 1$ unit is demanded during the period; and with probability $\frac{1}{3}, 2$ units are demanded during the period. (4) We observe the inventory level at the beginning of the next period.

Define a period's state to be the period's beginning inventory level. Determine the transition matrix that could be used to model this inventory system as a Markov chain.
3 A company has two machines. During any day, each machine that is working at the beginning of the day has a $\frac{1}{3}$ chance of breaking down. If a machine breaks down during the day, it is sent to a repair facility and will be working two days after it breaks down. (Thus, if a machine breaks down during day 3 , it will be working at the beginning of day 5 .) Letting the state of the system be the number of machines working at the beginning of the day, formulate a transition probability matrix for this situation.

## Group B

4 Referring to Problem 1, suppose that tomorrow's Smalltown weather depends on the last two days of

Smalltown weather, as follows: (1) If the last two days have been sunny, then $95 \%$ of the time, tomorrow will be sunny. (2) If yesterday was cloudy and today is sunny, then $70 \%$ of the time, tomorrow will be sunny. (3) If yesterday was sunny and today is cloudy, then $60 \%$ of the time, tomorrow will be cloudy. (4) If the last two days have been cloudy, then $80 \%$ of the time, tomorrow will be cloudy.

Using this information, model Smalltown's weather as a Markov chain. If tomorrow's weather depended on the last three days of Smalltown weather, how many states will be needed to model Smalltown's weather as a Markov chain? (Note: The approach used in this problem can be used to model a discrete-time stochastic process as a Markov chain even if $\mathbf{X}_{t+1}$ depends on states prior to $\mathbf{X}_{t}$, such as $\mathbf{X}_{t-1}$ in the current example.)

5 Let $\mathbf{X}_{t}$ be the location of your token on the Monopoly board after $t$ dice rolls. Can $\mathbf{X}_{t}$ be modeled as a Markov chain? If not, how can we modify the definition of the state at time $t$ so that $\mathbf{X}_{0}, \mathbf{X}_{1}, \ldots, \mathbf{X}_{t}, \ldots$ would be a Markov chain? (Hint: How does a player go to Jail? In this problem, assume that players who are sent to Jail stay there until they roll doubles or until they have spent three turns there, whichever comes first.)
6 In Problem 3, suppose a machine that breaks down returns to service three days later (for instance, a machine that breaks down during day 3 would be back in working order at the beginning of day 6). Determine a transition probability matrix for this situation.

## 17.3 n-Step Transition Probabilities

Suppose we are studying a Markov chain with a known transition probability matrix $P$. (Since all chains that we will deal with are stationary, we will not bother to label our Markov chains as stationary.) A question of interest is: If a Markov chain is in state $i$ at time $m$, what is the probability that $n$ periods later the Markov chain will be in state $j$ ? Since we are dealing with a stationary Markov chain, this probability will be independent of $m$, so we may write

$$
P\left(\mathbf{X}_{m+n}=j \mid \mathbf{X}_{m}=i\right)=P\left(\mathbf{X}_{n}=j \mid \mathbf{X}_{0}=i\right)=P_{i j}(n)
$$

where $P_{i j}(n)$ is called the $\boldsymbol{n}$-step probability of a transition from state $i$ to state $j$.
Clearly, $P_{i j}(1)=p_{i j}$. To determine $P_{i j}(2)$, note that if the system is now in state $i$, then for the system to end up in state $j$ two periods from now, we must go from state $i$ to some state $k$ and then go from state $k$ to state $j$ (see Figure 3). This reasoning allows us to write

$$
\begin{aligned}
P_{i j}(2)= & \sum_{k=1}^{k=s}(\text { probability of transition from } i \text { to } k) \\
& \times(\text { probability of transition from } k \text { to } j)
\end{aligned}
$$

Using the definition of $P$, the transition probability matrix, we rewrite the last equation as

$$
\begin{equation*}
P_{i j}(2)=\sum_{k=1}^{k=s} p_{i k} p_{k j} \tag{3}
\end{equation*}
$$

The right-hand side of (3) is just the scalar product of row $i$ of the $P$ matrix with column $j$ of the $P$ matrix. Hence, $P_{i j}(2)$ is the $i j$ th element of the matrix $P^{2}$. By extending this reasoning, it can be shown that for $n>1$,

$$
\begin{equation*}
P_{i j}(n)=i j \text { th element of } P^{n} \tag{4}
\end{equation*}
$$

Of course, for $n=0, P_{i j}(0)=P\left(\mathbf{X}_{0}=j \mid \mathbf{X}_{0}=i\right)$, so we must write

$$
P_{i j}(0)= \begin{cases}1 & \text { if } j=i \\ 0 & \text { if } j \neq i\end{cases}
$$

We illustrate the use of Equation (4) in Example 4.

FIGURE 3 $P_{i j}(2)=p_{i 1} p_{1 j}+$ $p_{i 2} p_{2 j}+\cdots+p_{i s} p_{s j} \quad$ Time 0


Suppose the entire cola industry produces only two colas. Given that a person last purchased cola 1 , there is a $90 \%$ chance that her next purchase will be cola 1 . Given that a person last purchased cola 2 , there is an $80 \%$ chance that her next purchase will be cola 2 .

1 If a person is currently a cola 2 purchaser, what is the probability that she will purchase cola 1 two purchases from now?

2 If a person is currently a cola 1 purchaser, what is the probability that she will purchase cola 1 three purchases from now?

Solution
We view each person's purchases as a Markov chain with the state at any given time being the type of cola the person last purchased. Hence, each person's cola purchases may be represented by a two-state Markov chain, where

$$
\text { State } 1=\text { person has last purchased cola } 1
$$

State $2=$ person has last purchased cola 2
If we define $\mathbf{X}_{n}$ to be the type of cola purchased by a person on her $n$th future cola purchase (present cola purchase $=\mathbf{X}_{0}$ ), then $\mathbf{X}_{0}, \mathbf{X}_{1}, \ldots$ may be described as the Markov chain with the following transition matrix:

## Cola 1 Cola 2

$$
P=\begin{aligned}
& \text { Cola } 1 \\
& \text { Cola } 2
\end{aligned}\left[\begin{array}{cc}
.90 & .10 \\
.20 & .80
\end{array}\right]
$$

We can now answer questions 1 and 2 .
1 We seek $P\left(\mathbf{X}_{2}=1 \mid \mathbf{X}_{0}=2\right)=P_{21}(2)=$ element 21 of $P^{2}$ :

$$
P^{2}=\left[\begin{array}{ll}
.90 & .10 \\
.20 & .80
\end{array}\right]\left[\begin{array}{ll}
.90 & .10 \\
.20 & .80
\end{array}\right]=\left[\begin{array}{ll}
.83 & .17 \\
.34 & .66
\end{array}\right]
$$

Hence, $P_{21}(2)=.34$. This means that the probability is .34 that two purchases in the future a cola 2 drinker will purchase cola 1 . By using basic probability theory, we may obtain this answer in a different way (see Figure 4). Note that $P_{21}(2)=$ (probability that next purchase is cola 1 and second purchase is cola 1 ) + (probability that next purchase is cola 2 and second purchase is cola 1$)=p_{21} p_{11}+p_{22} p_{21}=(.20)(.90)+(.80)(.20)=.34$.
2 We seek $P_{11}(3)=$ element 11 of $P^{3}$ :

$$
P^{3}=P\left(P^{2}\right)=\left[\begin{array}{ll}
.90 & .10 \\
.20 & .80
\end{array}\right]\left[\begin{array}{ll}
.83 & .17 \\
.34 & .66
\end{array}\right]=\left[\begin{array}{ll}
.781 & .219 \\
.438 & .562
\end{array}\right]
$$

Therefore, $P_{11}(3)=.781$.

FIGURE 4 Probability That Two Periods from Now, a Cola 2 Purchaser Will Purchase Cola 1 Is $.20(.90)+$ $.80(.20)=.34$



In many situations, we do not know the state of the Markov chain at time 0 . As defined in Section 17.2, let $q_{i}$ be the probability that the chain is in state $i$ at time 0 . Then we can determine the probability that the system is in state $i$ at time $n$ by using the following reasoning (see Figure 5).

$$
\begin{align*}
& \text { Probability of being in state } j \text { at time } n \\
& \qquad \begin{array}{l}
=\sum_{i=1}^{i=s}(\text { probability that state is originally } i) \\
\quad \times(\text { probability of going from } i \text { to } j \text { in } n \text { transitions }) \\
= \\
=\sum_{i=1}^{i=s} q_{i} P_{i j}(n) \\
= \\
\mathbf{q}\left(\text { column } j \text { of } P^{n}\right)
\end{array}
\end{align*}
$$

where $\mathbf{q}=\left[\begin{array}{llll}q_{1} & q_{2} & \cdots & q_{s}\end{array}\right]$.
To illustrate the use of (5), we answer the following question: Suppose $60 \%$ of all people now drink cola 1, and $40 \%$ now drink cola 2 . Three purchases from now, what fraction of all purchasers will be drinking cola 1 ? Since $\mathbf{q}=\left[\begin{array}{ll}.60 & .40\end{array}\right]$ and $\mathbf{q}\left(\right.$ column 1 of $\left.P^{3}\right)=$ probability that three purchases from now a person drinks cola 1 , the desired probability is

$$
\left[\begin{array}{ll}
.60 & .40
\end{array}\right]\left[\begin{array}{l}
.781 \\
.438
\end{array}\right]=.6438
$$

Hence, three purchases from now, $64 \%$ of all purchasers will be purchasing cola 1 .
To illustrate the behavior of the $n$-step transition probabilities for large values of $n$, we have computed several of the $n$-step transition probabilities for the Cola example in Table 2.

| TABLE 2 <br> $n-S t e p ~ T r a n s i t i o n ~ P r o b a b i l i t i e s ~ f o r ~ C o l a ~ D r i n k e r s ~$ |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: |
| $n$ $P_{11}(n)$ $P_{12}(n)$ $P_{21}(n)$ $P_{22}(n)$ <br> 1 .90 .10 .20 .80 <br> 2 .83 .17 .34 .66 <br> 3 .78 .22 .44 .56 <br> 4 .75 .25 .51 .49 <br> 5 .72 .28 .56 .44 <br> 10 .68 .32 .65 .35 <br> 20 .67 .33 .67 .33 <br> 30 .67 .33 .67 .33 <br> 40 .67 .33 .67 .33 |  |  |  |  |

For large $n$, both $P_{11}(n)$ and $P_{21}(n)$ are nearly constant and approach .67. This means that for large $n$, no matter what the initial state, there is a .67 chance that a person will be a cola 1 purchaser. Similarly, we see that for large $n$, both $P_{12}(n)$ and $P_{22}(n)$ are nearly constant and approach .33. This means that for large $n$, no matter what the initial state, there is a .33 chance that a person will be a cola 2 purchaser. In Section 5.5, we make a thorough study of this settling down of the $n$-step transition probabilities.

REMARK We can easily multiply matrices on a spreadsheet using the MMULT command, as discussed in Section 13.7.

## PROBLEMS

## Group A

1 Each American family is classified as living in an urban, rural, or suburban location. During a given year, $15 \%$ of all urban families move to a suburban location, and $5 \%$ move to a rural location; also, $6 \%$ of all suburban families move to an urban location, and $4 \%$ move to a rural location; finally, $4 \%$ of all rural families move to an urban location, and $6 \%$ move to a suburban location.
a If a family now lives in an urban location, what is the probability that it will live in an urban area two years from now? A suburban area? A rural area?
b Suppose that at present, $40 \%$ of all families live in an urban area, $35 \%$ live in a suburban area, and $25 \%$ live in a rural area. Two years from now, what percentage of American families will live in an urban area?
c What problems might occur if this model were used to predict the future population distribution of the United States?

2 The following questions refer to Example 1.
a After playing the game twice, what is the probability that I will have $\$ 3$ ? How about $\$ 2$ ?
b After playing the game three times, what is the probability that I will have $\$ 2$ ?
3 In Example 2, determine the following $n$-step transition probabilities:
a After two balls are painted, what is the probability that the state is $\left[\begin{array}{lll}0 & 2 & 0\end{array}\right]$ ?
b After three balls are painted, what is the probability that the state is $\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]$ ? (Draw a diagram like Figure 4.)

### 17.4 Classification of States in a Markov Chain

In Section 17.3, we mentioned the fact that after many transitions, the $n$-step transition probabilities tend to settle down. Before we can discuss this in more detail, we need to study how mathematicians classify the states of a Markov chain. We use the following transition matrix to illustrate most of the following definitions (see Figure 6).

$$
P=\left[\begin{array}{ccccc}
.4 & .6 & 0 & 0 & 0 \\
.5 & .5 & 0 & 0 & 0 \\
0 & 0 & .3 & .7 & 0 \\
0 & 0 & .5 & .4 & .1 \\
0 & 0 & 0 & .8 & .2
\end{array}\right]
$$ in $i$ and ends in $j$, such that each transition in the sequence has a positive probability of occurring.

A state $j$ is reachable from state $i$ if there is a path leading from $i$ to $j$.

FIGURE 6 Graphical Representation of Transition Matrix


Two states $i$ and $j$ are said to communicate if $j$ is reachable from $i$, and $i$ is reachable from $j$.

For the transition probability matrix $P$ represented in Figure 6, state 5 is reachable from state 3 (via the path $3-4-5$ ), but state 5 is not reachable from state 1 (there is no path from 1 to 5 in Figure 6). Also, states 1 and 2 communicate (we can go from 1 to 2 and from 2 to 1).

DEFINITION A A set of states $S$ in a Markov chain is a closed set if no state outside of $S$ is reachable from any state in $S$.

From the Markov chain with transition matrix $P$ in Figure $6, S_{1}=\{1,2\}$ and $S_{2}=\{3$, $4,5\}$ are both closed sets. Observe that once we enter a closed set, we can never leave the closed set (in Figure 6, no arc begins in $S_{1}$ and ends in $S_{2}$ or begins in $S_{2}$ and ends in $S_{1}$ ).

DEFINITION ■ A state $i$ is an absorbing state if $p_{i i}=1$.

Whenever we enter an absorbing state, we never leave the state. In Example 1, the gambler's ruin, states 0 and 4 are absorbing states. Of course, an absorbing state is a closed set containing only one state.

DEFINITION ■ A state $i$ is a transient state if there exists a state $j$ that is reachable from $i$, but the state $i$ is not reachable from state $j$.

In other words, a state $i$ is transient if there is a way to leave state $i$ that never returns to state $i$. In the gambler's ruin example, states 1,2 , and 3 are transient states. For example (see Figure 1), from state 2, it is possible to go along the path $2-3-4$, but there is no way to return to state 2 from state 4. Similarly, in Example 2, $\left[\begin{array}{lll}2 & 0 & 0\end{array}\right]$, $\left[\begin{array}{lll}1 & 1 & 0\end{array}\right]$, and $\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$ are all transient states (in Figure 2, there is a path from $\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$ to $\left[\begin{array}{ll}0 & 0\end{array}\right.$ 2], but once both balls are painted, there is no way to return to $\left.\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]\right)$.

After a large number of periods, the probability of being in any transient state $i$ is zero. Each time we enter a transient state $i$, there is a positive probability that we will leave $i$ forever and end up in the state $j$ described in the definition of a transient state. Thus, eventually we are sure to enter state $j$ (and then we will never return to state $i$ ). To illustrate, in Example 2, suppose we are in the transient state $\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$. With probability 1, the unpainted ball will eventually be painted, and we will never reenter state $\left[\begin{array}{lll}1 & 0 & 1\end{array}\right]$ (see Figure 2 ).

DEFINITION - If a state is not transient, it is called a recurrent state.

In Example 1, states 0 and 4 are recurrent states (and also absorbing states), and in Example 2, $\left[\begin{array}{lll}0 & 2 & 0\end{array}\right]$, $\left[\begin{array}{lll}0 & 0 & 2\end{array}\right]$, and $\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]$ are recurrent states. For the transition matrix $P$ in Figure 6, all states are recurrent.

DEFINITION ■ A state $i$ is periodic with period $k>1$ if $k$ is the smallest number such that all paths leading from state $i$ back to state $i$ have a length that is a multiple of $k$. If a recurrent state is not periodic, it is referred to as aperiodic.

For the Markov chain with transition matrix

$$
Q=\left[\begin{array}{lll}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{array}\right]
$$

each state has period 3. For example, if we begin in state 1 , the only way to return to state 1 is to follow the path $1-2-3-1$ for some number of times (say, $m$ ). (See Figure 7.) Hence, any return to state 1 will take $3 m$ transitions, so state 1 has period 3 . Wherever we are, we are sure to return three periods later.

DEFINITION If all states in a chain are recurrent, aperiodic, and communicate with each other, the chain is said to be ergodic.

The gambler's ruin example is not an ergodic chain, because (for example) states 3 and 4 do not communicate. Example 2 is also not an ergodic chain, because (for example) [2 $0 \quad 0]$ and $\left[\begin{array}{lll}0 & 1 & 1\end{array}\right]$ do not communicate. Example 4, the cola example, is an ergodic Markov chain. Of the following three Markov chains, $P_{1}$ and $P_{3}$ are ergodic, and $P_{2}$ is not ergodic.

Nonergodic

$$
P_{1}=\left[\begin{array}{ccc}
\frac{1}{3} & \frac{2}{3} & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} \\
0 & \frac{1}{4} & \frac{3}{4}
\end{array}\right] \quad \text { Ergodic }
$$

$$
P_{2}=\left[\begin{array}{cccc}
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
0 & 0 & \frac{2}{3} & \frac{1}{3} \\
0 & 0 & \frac{1}{4} & \frac{3}{4}
\end{array}\right.
$$

FIGURE 7
A Periodic Markov
Chain $k=3$


$$
P_{3}=\left[\begin{array}{ccc}
\frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\
\frac{2}{3} & \frac{1}{3} & 0 \\
0 & \frac{2}{3} & \frac{1}{3}
\end{array}\right] \quad \text { Ergodic }
$$

$P_{2}$ is not ergodic because there are two closed classes of states (class $1=\{1,2\}$ and class $2=\{3,4\}$ ), and the states in different classes do not communicate with each other.

After the next two sections, the importance of the concepts introduced in this section will become clear.

## PROBLEMS

## Group A

1 In Example 1, what is the period of states 1 and 3?
2 Is the Markov chain of Section 17.3, Problem 1, an ergodic Markov chain?

3 Consider the following transition matrix:

$$
P=\left[\begin{array}{cccccc}
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 \\
\frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{3} & 0 & 0 & 0 & \frac{2}{3}
\end{array}\right]
$$

a Which states are transient?
b Which states are recurrent?
c Identify all closed sets of states.
d Is this chain ergodic?
4 For each of the following chains, determine whether the Markov chain is ergodic. Also, for each chain, determine the recurrent, transient, and absorbing states.

$$
P_{1}=\left[\begin{array}{rrr}
0 & .8 & .2 \\
.3 & .7 & 0 \\
.4 & .5 & .1
\end{array}\right] \quad P_{2}=\left[\begin{array}{rrrr}
.2 & .8 & 0 & 0 \\
0 & 0 & .9 & .1 \\
.4 & .5 & .1 & 0 \\
0 & 0 & 0 & 1
\end{array}\right]
$$

5 Fifty-four players (including Gabe Kaplan and James Garner) participated in the 1980 World Series of Poker. Each player began with $\$ 10,000$. Play continued until one player had won everybody else's money. If the World Series of Poker were to be modeled as a Markov chain, how many absorbing states would the chain have?

6 Which of the following chains is ergodic?

$$
P_{1}=\left[\begin{array}{rrr}
.4 & 0 & .6 \\
.3 & .3 & .4 \\
0 & .5 & .5
\end{array}\right] \quad P_{2}=\left[\begin{array}{cccc}
.7 & 0 & 0 & .3 \\
.2 & .2 & .4 & .2 \\
.6 & .1 & .1 & .2 \\
.2 & 0 & 0 & .8
\end{array}\right]
$$

### 17.5 Steady-State Probabilities and Mean First Passage Times

In our discussion of the cola example (Example 4), we found that after a long time, the probability that a person's next cola purchase would be cola 1 approached .67 and .33 that it would be cola 2 (see Table 2). These probabilities did not depend on whether the person was initially a cola 1 or a cola 2 drinker. In this section, we discuss the important concept of steadystate probabilities, which can be used to describe the long-run behavior of a Markov chain.

The following result is vital to an understanding of steady-state probabilities and the long-run behavior of Markov chains.

THEOREM 1
Let $P$ be the transition matrix for an $s$-state ergodic chain. ${ }^{\dagger}$ Then there exists a vector $\pi=\left[\begin{array}{llll}\pi_{1} & \pi_{2} & \cdots & \pi_{s}\end{array}\right]$ such that

[^18]

## Deterministic Dynamic Programming

Dynamic programming is a technique that can be used to solve many optimization problems. In most applications, dynamic programming obtains solutions by working backward from the end of a problem toward the beginning, thus breaking up a large, unwieldy problem into a series of smaller, more tractable problems.

We introduce the idea of working backward by solving two well-known puzzles and then show how dynamic programming can be used to solve network, inventory, and resourceallocation problems. We close the chapter by showing how to use spreadsheets to solve dynamic programming problems.

### 18.1 Two Puzzles ${ }^{\dagger}$

In this section, we show how working backward can make a seemingly difficult problem almost trivial to solve.

## EXAMPLE 1 Match Puzzile

Suppose there are 30 matches on a table. I begin by picking up 1, 2, or 3 matches. Then my opponent must pick up 1,2 , or 3 matches. We continue in this fashion until the last match is picked up. The player who picks up the last match is the loser. How can I (the first player) be sure of winning the game?

Solution If I can ensure that it will be my opponent's turn when 1 match remains, I will certainly win. Working backward one step, if I can ensure that it will be my opponent's turn when 5 matches remain, I will win. The reason for this is that no matter what he does when 5 matches remain, I can make sure that when he has his next turn, only 1 match will remain. For example, suppose it is my opponent's turn when 5 matches remain. If my opponent picks up 2 matches, I will pick up 2 matches, leaving him with 1 match and sure defeat. Similarly, if I can force my opponent to play when $5,9,13,17,21,25$, or 29 matches remain, I am sure of victory. Thus, I cannot lose if I pick up $30-29=1$ match on my first turn. Then I simply make sure that my opponent will always be left with $29,25,21,17,13,9$, or 5 matches on his turn. Notice that we have solved this puzzle by working backward from the end of the problem toward the beginning. Try solving this problem without working backward!

## EXAMPLE 2 MHIk

I have a $9-\mathrm{oz}$ cup and a $4-\mathrm{oz}$ cup. My mother has ordered me to bring home exactly 6 oz of milk. How can I accomplish this goal?

[^19]TABLE 1

| Moves in the Cup-and-Milk Problem |  |
| :--- | :---: |
| No. of Ounces <br> in $9-0 z$ Cup | No. of Ounces <br> in $4-0 z$ |
| 6 | 0 |
| 6 | 4 |
| 9 | 1 |
| 0 | 1 |
| 1 | 0 |
| 1 | 4 |
| 5 | 0 |
| 5 | 4 |
| 9 | 0 |
| 0 | 0 |

Solution By starting near the end of the problem, I cleverly realize that the problem can easily be solved if I can somehow get 1 oz of milk into the 4 -oz cup. Then I can fill the 9 -oz cup and empty 3 oz from the $9-\mathrm{oz}$ cup into the partially filled $4-\mathrm{oz}$ cup. At this point, I will be left with 6 oz of milk. After I have this flash of insight, the solution to the problem may easily be described as in Table 1 (the initial situation is written last, and the final situation is written first).

## PROBLEMS

## Group A

1 Suppose there are 40 matches on a table. I begin by picking up $1,2,3$, or 4 matches. Then my opponent must pick up $1,2,3$, or 4 matches. We continue until the last match is picked up. The player who picks up the last match is the loser. Can I be sure of victory? If so, how?

2 Three players have played three rounds of a gambling game. Each round has one loser and two winners. The losing player must pay each winner the amount of money that the winning player had at the beginning of the round. At the end of the three rounds each player has $\$ 10$. You are told that each player has won one round. By working backward, determine the original stakes of the three players. [Note: If the answer turns out to be (for example) 5, 15, 10, don't worry about which player had which stake; we can't really tell which player ends up with how much, but we can determine the numerical values of the original stakes.]

## Group B

3 We have 21 coins and are told that one is heavier than any of the other coins. How many weighings on a balance will it take to find the heaviest coin? (Hint: If the heaviest coin is in a group of three coins, we can find it in one weighing. Then work backward to two weighings, and so on.)
4 Given a $7-$ oz cup and a $3-$ oz cup, explain how we can return from a well with 5 oz of water.

### 18.2 A Network Problem

Many applications of dynamic programming reduce to finding the shortest (or longest) path that joins two points in a given network. The following example illustrates how dynamic programming (working backward) can be used to find the shortest path in a network.

Joe Cougar lives in New York City, but he plans to drive to Los Angeles to seek fame and fortune. Joe's funds are limited, so he has decided to spend each night on his trip at a friend's house. Joe has friends in Columbus, Nashville, Louisville, Kansas City, Omaha, Dallas, San Antonio, and Denver. Joe knows that after one day's drive he can reach Columbus, Nashville, or Louisville. After two days of driving, he can reach Kansas City, Omaha, or Dallas. After three days of driving, he can reach San Antonio or Denver. Finally, after four days of driving, he can reach Los Angeles. To minimize the number of miles traveled, where should Joe spend each night of the trip? The actual road mileages between cities are given in Figure 1.

Solution Joe needs to know the shortest path between New York and Los Angeles in Figure 1. We will find it by working backward. We have classified all the cities that Joe can be in at the beginning of the $n$th day of his trip as stage $n$ cities. For example, because Joe can only be in San Antonio or Denver at the beginning of the fourth day (day 1 begins when Joe leaves New York), we classify San Antonio and Denver as stage 4 cities. The reason for classifying cities according to stages will become apparent later.

The idea of working backward implies that we should begin by solving an easy problem that will eventually help us to solve a complex problem. Hence, we begin by finding the shortest path to Los Angeles from each city in which there is only one day of driving left (stage 4 cities). Then we use this information to find the shortest path to Los Angeles from each city for which only two days of driving remain (stage 3 cities). With this information in hand, we are able to find the shortest path to Los Angeles from each city that is three days distant (stage 2 cities). Finally, we find the shortest path to Los Angeles from each city (there is only one: New York) that is four days away.

To simplify the exposition, we use the numbers $1,2, \ldots, 10$ given in Figure 1 to label the 10 cities. We also define $c_{i j}$ to be the road mileage between city $i$ and city $j$. For example, $c_{35}=580$ is the road mileage between Nashville and Kansas City. We let $f_{t}(i)$ be the length of the shortest path from city $i$ to Los Angeles, given that city $i$ is a stage $t$ city. ${ }^{\dagger}$

## Stage 4 Computations

We first determine the shortest path to Los Angeles from each stage 4 city. Since there is only one path from each stage 4 city to Los Angeles, we immediately see that $f_{4}(8)=$ 1,030 , the shortest path from Denver to Los Angeles simply being the only path from Denver to Los Angeles. Similarly, $f_{4}(9)=1,390$, the shortest (and only) path from San Antonio to Los Angeles.

## Stage 3 Computations

We now work backward one stage (to stage 3 cities) and find the shortest path to Los Angeles from each stage 3 city. For example, to determine $f_{3}(5)$, we note that the shortest path from city 5 to Los Angeles must be one of the following:

Path 1 Go from city 5 to city 8 and then take the shortest path from city 8 to city 10 .
Path 2 Go from city 5 to city 9 and then take the shortest path from city 9 to city 10 .
The length of path 1 may be written as $c_{58}+f_{4}(8)$, and the length of path 2 may be written as $c_{59}+f_{4}(9)$. Hence, the shortest distance from city 5 to city 10 may be written as

[^20]FIGURE 1 Joe's Trip Across the United States


$$
f_{3}(5)=\min \left\{\begin{array}{l}
c_{58}+f_{4}(8)=610+1,030=1,640^{*} \\
c_{59}+f_{4}(9)=790+1,390=2,180
\end{array}\right.
$$

[the $*$ indicates the choice of arc that attains the $f_{3}(5)$ ]. Thus, we have shown that the shortest path from city 5 to city 10 is the path $5-8-10$. Note that to obtain this result, we made use of our knowledge of $f_{4}(8)$ and $f_{4}(9)$.

Similarly, to find $f_{3}(6)$, we note that the shortest path to Los Angeles from city 6 must begin by going to city 8 or to city 9 . This leads us to the following equation:

$$
f_{3}(6)=\min \left\{\begin{array}{l}
c_{68}+f_{4}(8)=540+1,030=1,570^{*} \\
c_{69}+f_{4}(9)=940+1,390=2,330
\end{array}\right.
$$

Thus, $f_{3}(6)=1,570$, and the shortest path from city 6 to city 10 is the path $6-8-10$.
To find $f_{3}(7)$, we note that

$$
f_{3}(7)=\min \left\{\begin{array}{l}
c_{78}+f_{4}(8)=790+1,030=1,820 \\
c_{79}+f_{4}(9)=270+1,390=1,660^{*}
\end{array}\right.
$$

Therefore, $f_{3}(7)=1,660$, and the shortest path from city 7 to city 10 is the path $7-9-10$.

## Stage 2 Computations

Given our knowledge of $f_{3}(5), f_{3}(6)$, and $f_{3}(7)$, it is now easy to work backward one more stage and compute $f_{2}(2), f_{2}(3)$, and $f_{2}(4)$ and thus the shortest paths to Los Angeles from city 2 , city 3 , and city 4 . To illustrate how this is done, we find the shortest path (and its length) from city 2 to city 10 . The shortest path from city 2 to city 10 must begin by going from city 2 to city 5 , city 6 , or city 7 . Once this shortest path gets to city 5 , city 6 , or city 7, then it must follow a shortest path from that city to Los Angeles. This reasoning shows that the shortest path from city 2 to city 10 must be one of the following:

Path 1 Go from city 2 to city 5 . Then follow a shortest path from city 5 to city 10 . A path of this type has a total length of $c_{25}+f_{3}(5)$.

Path 2 Go from city 2 to city 6 . Then follow a shortest path from city 6 to city 10 . A path of this type has a total length of $c_{26}+f_{3}(6)$.

Path 3 Go from city 2 to city 7 . Then follow a shortest path from city 7 to city 10 . This path has a total length of $c_{27}+f_{3}(7)$. We may now conclude that

$$
f_{2}(2)=\min \left\{\begin{array}{l}
c_{25}+f_{3}(5)=680+1,640=2,320^{*} \\
c_{26}+f_{3}(6)=790+1,570=2,360 \\
c_{27}+f_{3}(7)=1,050+1,660=2,710
\end{array}\right.
$$

Thus, $f_{2}(2)=2,320$, and the shortest path from city 2 to city 10 is to go from city 2 to city 5 and then follow the shortest path from city 5 to city $10(5-8-10)$.

Similarly,

$$
f_{2}(3)=\min \left\{\begin{array}{l}
c_{35}+f_{3}(5)=580+1,640=2,220^{*} \\
c_{36}+f_{3}(6)=760+1,570=2,330 \\
c_{37}+f_{3}(7)=660+1,660=2,320
\end{array}\right.
$$

Thus, $f_{2}(3)=2,220$, and the shortest path from city 3 to city 10 consists of arc $3-5$ and the shortest path from city 5 to city $10(5-8-10)$.

In similar fashion,

$$
f_{2}(4)=\min \left\{\begin{array}{l}
c_{45}+f_{3}(5)=510+1,640=2,150^{*} \\
c_{46}+f_{3}(6)=700+1,570=2,270 \\
c_{47}+f_{3}(7)=830+1,660=2,490
\end{array}\right.
$$

Thus, $f_{2}(4)=2,150$, and the shortest path from city 4 to city 10 consists of arc $4-5$ and the shortest path from city 5 to city $10(5-8-10)$.

## Stage 1 Computations

We can now use our knowledge of $f_{2}(2), f_{2}(3)$, and $f_{2}(4)$ to work backward one more stage to find $f_{1}(1)$ and the shortest path from city 1 to city 10 . Note that the shortest path from city 1 to city 10 must begin by going to city 2 , city 3 , or city 4 . This means that the shortest path from city 1 to city 10 must be one of the following:
Path 1 Go from city 1 to city 2 and then follow a shortest path from city 2 to city 10 . The length of such a path is $c_{12}+f_{2}(2)$.
Path 2 Go from city 1 to city 3 and then follow a shortest path from city 3 to city 10 . The length of such a path is $c_{13}+f_{2}(3)$.
Path 3 Go from city 1 to city 4 and then follow a shortest path from city 4 to city 10 . The length of such a path is $c_{14}+f_{2}(4)$. It now follows that

$$
f_{1}(1)=\min \left\{\begin{array}{l}
c_{12}+f_{2}(2)=550+2,320=2,870^{*} \\
c_{13}+f_{2}(3)=900+2,220=3,120 \\
c_{14}+f_{2}(4)=770+2,150=2,920
\end{array}\right.
$$

## Determination of the Optimal Path

Thus, $f_{1}(1)=2,870$, and the shortest path from city 1 to city 10 goes from city 1 to city 2 and then follows the shortest path from city 2 to city 10 . Checking back to the $f_{2}(2)$ calculations, we see that the shortest path from city 2 to city 10 is $2-5-8-10$. Translating the numerical labels into real cities, we see that the shortest path from New York to Los An-
geles passes through New York, Columbus, Kansas City, Denver, and Los Angeles. This path has a length of $f_{1}(1)=2,870$ miles.

## Computational Efficiency of Dynamic Programming

For Example 3, it would have been an easy matter to determine the shortest path from New York to Los Angeles by enumerating all the possible paths [after all, there are only $3(3)(2)=18$ paths]. Thus, in this problem, the use of dynamic programming did not really serve much purpose. For larger networks, however, dynamic programming is much more efficient for determining a shortest path than the explicit enumeration of all paths. To see this, consider the network in Figure 2. In this network, it is possible to travel from any node in stage $k$ to any node in stage $k+1$. Let the distance between node $i$ and node $j$ be $c_{i j}$. Suppose we want to determine the shortest path from node 1 to node 27 . One way to solve this problem is explicit enumeration of all paths. There are $5^{5}$ possible paths from node 1 to node 27. It takes five additions to determine the length of each path. Thus, explicitly enumerating the length of all paths requires $5^{5}(5)=5^{6}=15,625$ additions.

Suppose we use dynamic programming to determine the shortest path from node 1 to node 27 . Let $f_{t}(i)$ be the length of the shortest path from node $i$ to node 27 , given that node $i$ is in stage $t$. To determine the shortest path from node 1 to node 27, we begin by finding $f_{6}(22), f_{6}(23), f_{6}(24), f_{6}(25)$, and $f_{6}(26)$. This does not require any additions. Then we find $f_{5}(17), f_{5}(18), f_{5}(19), f_{5}(20), f_{5}(21)$. For example, to find $f_{5}(21)$ we use the following equation:

$$
f_{5}(21)=\min _{j}\left\{c_{21, j}+f_{6}(j)\right\} \quad(j=22,23,24,25,26)
$$

Determining $f_{5}(21)$ in this manner requires five additions. Thus, the calculation of all the $f_{5}(\cdot)$ 's requires $5(5)=25$ additions. Similarly, the calculation of all the $f_{4}(\cdot)$ 's requires 25 additions, and the calculation of all the $f_{3}(\cdot)$ 's requires 25 additions. The determination of all the $f_{2}(\cdot)$ 's also requires 25 additions, and the determination of $f_{1}(1)$ requires 5 additions. Thus, in total, dynamic programming requires $4(25)+5=105$ additions to find

FIGURE 2
Illustration of Computational Efficiency of Dynamic Programming

the shortest path from node 1 to node 27 . Because explicit enumeration requires 15,625 additions, we see that dynamic programming requires only 0.007 times as many additions as explicit enumeration. For larger networks, the computational savings effected by dynamic programming are even more dramatic.

Besides additions, determination of the shortest path in a network requires comparisons between the lengths of paths. If explicit enumeration is used, then $5^{5}-1=3,124$ comparisons must be made (that is, compare the length of the first two paths, then compare the length of the third path with the shortest of the first two paths, and so on). If dynamic programming is used, then for $t=2,3,4,5$, determination of each $f_{t}(i)$ requires $5-1=4$ comparisons. Then to compute $f_{1}(1), 5-1=4$ comparisons are required. Thus, to find the shortest path from node 1 to node 27 , dynamic programming requires a total of $20(5-1)+4=84$ comparisons. Again, dynamic programming comes out far superior to explicit enumeration.

## Characteristics of Dynamic Programming Applications

We close this section with a discussion of the characteristics of Example 3 that are common to most applications of dynamic programming.

## Characteristic 1

The problem can be divided into stages with a decision required at each stage. In Example 3, stage $t$ consisted of those cities where Joe could be at the beginning of day $t$ of his trip. As we will see, in many dynamic programming problems, the stage is the amount of time that has elapsed since the beginning of the problem. We note that in some situations, decisions are not required at every stage (see Section 18.5).

## Characteristic 2

Each stage has a number of states associated with it. By a state, we mean the information that is needed at any stage to make an optimal decision. In Example 3, the state at stage $t$ is simply the city where Joe is at the beginning of day $t$. For example, in stage 3, the possible states are Kansas City, Omaha, and Dallas. Note that to make the correct decision at any stage, Joe doesn't need to know how he got to his current location. For example, if Joe is in Kansas City, then his remaining decisions don't depend on how he goes to Kansas City; his future decisions just depend on the fact that he is now in Kansas City.

## Characteristic 3

The decision chosen at any stage describes how the state at the current stage is transformed into the state at the next stage. In Example 3, Joe's decision at any stage is simply the next city to visit. This determines the state at the next stage in an obvious fashion. In many problems, however, a decision does not determine the next stage's state with certainty; instead, the current decision only determines the probability distribution of the state at the next stage.

## Characteristic 4

Given the current state, the optimal decision for each of the remaining stages must not depend on previously reached states or previously chosen decisions. This idea is known as the principle of optimality. In the context of Example 3, the principle of optimality
reduces to the following: Suppose the shortest path (call it $R$ ) from city 1 to city 10 is known to pass through city $i$. Then the portion of $R$ that goes from city $i$ to city 10 must be a shortest path from city $i$ to city 10 . If this were not the case, then we could create a path from city 1 to city 10 that was shorter than $R$ by appending a shortest path from city $i$ to city 10 to the portion of $R$ leading from city 1 to city $i$. This would create a path from city 1 to city 10 that is shorter than $R$, thereby contradicting the fact that $R$ is a shortest path from city 1 to city 10 . For example, if the shortest path from city 1 to city 10 is known to pass through city 2 , then the shortest path from city 1 to city 10 must include a shortest path from city 2 to city $10(2-5-8-10)$. This follows because any path from city 1 to city 10 that passes through city 2 and does not contain a shortest path from city 2 to city 10 will have a length of $c_{12}+$ [something bigger than $\left.f_{2}(2)\right]$. Of course, such a path cannot be a shortest path from city 1 to city 10 .

## Characteristic 5

If the states for the problem have been classified into one of $T$ stages, there must be a recursion that relates the cost or reward earned during stages $t, t+1, \ldots, T$ to the cost or reward earned from stages $t+1, t+2, \ldots, T$. In essence, the recursion formalizes the working-backward procedure. In Example 3, our recursion could have been written as

$$
f_{t}(i)=\min _{j}\left\{c_{i j}+f_{t+1}(j)\right\}
$$

where $j$ must be a stage $t+1$ city and $f_{5}(10)=0$.
We can now describe how to make optimal decisions. Let's assume that the initial state during stage 1 is $i_{1}$. To use the recursion, we begin by finding the optimal decision for each state associated with the last stage. Then we use the recursion described in characteristic 5 to determine $f_{T-1}(\cdot)$ (along with the optimal decision) for every stage $T-1$ state. Then we use the recursion to determine $f_{T-2}(\cdot)$ (along with the optimal decision) for every stage $T-2$ state. We continue in this fashion until we have computed $f_{1}\left(i_{1}\right)$ and the optimal decision when we are in stage 1 and state $i_{1}$. Then our optimal decision in stage 1 is chosen from the set of decisions attaining $f_{1}\left(i_{1}\right)$. Choosing this decision at stage 1 will lead us to some stage 2 state (call it state $i_{2}$ ) at stage 2 . Then at stage 2 , we choose any decision attaining $f_{2}\left(i_{2}\right)$. We continue in this fashion until a decision has been chosen for each stage.

In the rest of this chapter, we discuss many applications of dynamic programming. The presentation will seem easier if the reader attempts to determine how each problem fits into the network context introduced in Example 3. In the next section, we begin by studying how dynamic programming can be used to solve inventory problems.

## PROBLEMS

## Group A

1 Find the shortest path from node 1 to node 10 in the network shown in Figure 3. Also, find the shortest path from node 3 to node 10 .
2 A sales representative lives in Bloomington and must be in Indianapolis next Thursday. On each of the days Monday, Tuesday, and Wednesday, he can sell his wares in Indianapolis, Bloomington, or Chicago. From past experience, he believes that he can earn $\$ 12$ from spending a day in Indianapolis, \$16 from spending a day in Bloomington, and $\$ 17$ from spending a day in Chicago. Where should he spend the first three days

FIGURE 3


TABLE 2

|  | To |  |  |
| :--- | :---: | :---: | :---: |
| From | Indianapolis | Bloomington | Chicago |
| Indianapolis | - | 5 | 2 |
| Bloomington | 5 | - | 7 |
| Chicago | 2 | 7 | - |

and nights of the week to maximize his sales income less travel costs? Travel costs are shown in Table 2.

## Group B

3 I must drive from Bloomington to Cleveland. Several paths are available (see Figure 4). The number on each arc is the length of time it takes to drive between the two cities. For example, it takes 3 hours to drive from Bloomington to


Cincinnati. By working backward, determine the shortest path (in terms of time) from Bloomington to Cleveland. [Hint: Work backward and don't worry about stages-only about states.]

### 18.3 An Inventory Problem

In this section, we illustrate how dynamic programming can be used to solve an inventory problem with the following characteristics:
1 Time is broken up into periods, the present period being period 1, the next period 2, and the final period $T$. At the beginning of period 1 , the demand during each period is known.

2 At the beginning of each period, the firm must determine how many units should be produced. Production capacity during each period is limited.

3 Each period's demand must be met on time from inventory or current production. During any period in which production takes place, a fixed cost of production as well as a variable per-unit cost is incurred.

4 The firm has limited storage capacity. This is reflected by a limit on end-of-period inventory. A per-unit holding cost is incurred on each period's ending inventory.
5 The firm's goal is to minimize the total cost of meeting on time the demands for periods $1,2, \ldots, T$.

In this model, the firm's inventory position is reviewed at the end of each period (say, at the end of each month), and then the production decision is made. Such a model is called a periodic review model. This model is in contrast to the continuous review models in which the firm knows its inventory position at all times and may place an order or begin production at any time.

If we exclude the setup cost for producing any units, the inventory problem just described is similar to the Sailco inventory problem that we solved by linear programming in Section 3.10. Here, we illustrate how dynamic programming can be used to determine a production schedule that minimizes the total cost incurred in an inventory problem that meets the preceding description.

A company knows that the demand for its product during each of the next four months will be as follows: month 1,1 unit; month 2,3 units; month 3,2 units; month 4,4 units. At the beginning of each month, the company must determine how many units should be produced during the current month. During a month in which any units are produced, a setup cost of $\$ 3$ is incurred. In addition, there is a variable cost of $\$ 1$ for every unit produced. At the end of each month, a holding cost of $50 \phi$ per unit on hand is incurred. Capacity limitations allow a maximum of 5 units to be produced during each month. The size of the company's warehouse restricts the ending inventory for each month to 4 units at most. The company wants to determine a production schedule that will meet all demands on time and will minimize the sum of production and holding costs during the four months. Assume that 0 units are on hand at the beginning of the first month.

Solution Recall from Section 3.10 that we can ensure that all demands are met on time by restricting each month's ending inventory to be nonnegative. To use dynamic programming to solve this problem, we need to identify the appropriate state, stage, and decision. The stage should be defined so that when one stage remains, the problem will be trivial to solve. If we are at the beginning of month 4 , then the firm would meet demand at minimum cost by simply producing just enough units to ensure that (month 4 production) + (month 3 ending inventory) $=$ (month 4 demand). Thus, when one month remains, the firm's problem is easy to solve. Hence, we let time represent the stage. In most dynamic programming problems, the stage has something to do with time.

At each stage (or month), the company must decide how many units to produce. To make this decision, the company need only know the inventory level at the beginning of the current month (or the end of the previous month). Therefore, we let the state at any stage be the beginning inventory level.

Before writing a recursive relation that can be used to "build up" the optimal production schedule, we must define $f_{t}(i)$ to be the minimum cost of meeting demands for months $t, t+1, \ldots, 4$ if $i$ units are on hand at the beginning of month $t$. We define $c(x)$ to be the cost of producing $x$ units during a period. Then $c(0)=0$, and for $x>0, c(x)=3+x$. Because of the limited storage capacity and the fact that all demand must be met on time, the possible states during each period are $0,1,2,3$, and 4 . Thus, we begin by determining $f_{4}(0), f_{4}(1), f_{4}(2), f_{4}(3)$, and $f_{4}(4)$. Then we use this information to determine $f_{3}(0)$, $f_{3}(1), f_{3}(2), f_{3}(3)$, and $f_{3}(4)$. Then we determine $f_{2}(0), f_{2}(1), f_{2}(2), f_{2}(3)$, and $f_{2}(4)$. Finally, we determine $f_{1}(0)$. Then we determine an optimal production level for each month. We define $x_{t}(i)$ to be a production level during month $t$ that minimizes the total cost during months $t, t+1, \ldots, 4$ if $i$ units are on hand at the beginning of month $t$. We now begin to work backward.

## Month 4 Computations

During month 4, the firm will produce just enough units to ensure that the month 4 demand of 4 units is met. This yields

$$
\begin{aligned}
& f_{4}(0)=\text { cost of producing } 4-0 \text { units }=c(4)=3+4=\$ 7 \text { and } x_{4}(0)=4-0=4 \\
& f_{4}(1)=\text { cost of producing } 4-1 \text { units }=c(3)=3+3=\$ 6 \text { and } x_{4}(1)=4-1=3 \\
& f_{4}(2)=\text { cost of producing } 4-2 \text { units }=c(2)=3+2=\$ 5 \text { and } x_{4}(2)=4-2=2 \\
& f_{4}(3)=\text { cost of producing } 4-3 \text { units }=c(1)=3+1=\$ 4 \text { and } x_{4}(3)=4-3=1 \\
& f_{4}(4)=\text { cost of producing } 4-4 \text { units }=c(0)=\$ 0 \quad \text { and } \quad x_{4}(4)=4-4=0
\end{aligned}
$$

## Month 3 Computations

How can we now determine $f_{3}(\mathrm{i})$ for $i=0,1,2,3,4$ ? The cost $f_{3}(i)$ is the minimum cost incurred during months 3 and 4 if the inventory at the beginning of month 3 is $i$. For each possible production level $x$ during month 3 , the total cost during months 3 and 4 is

$$
\begin{equation*}
\left(\frac{1}{2}\right)(i+x-2)+c(x)+f_{4}(i+x-2) \tag{1}
\end{equation*}
$$

This follows because if $x$ units are produced during month 3 , the ending inventory for month 3 will be $i+x-2$. Then the month 3 holding cost will be $\left(\frac{1}{2}\right)(i+x-2)$, and the month 3 production cost will be $c(x)$. Then we enter month 4 with $i+x-2$ units on hand. Since we proceed optimally from this point onward (remember the principle of optimality), the cost for month 4 will be $f_{4}(i+x-2)$. We want to choose the month 3 production level to minimize (1), so we write

$$
\begin{equation*}
f_{3}(i)=\min _{x}\left\{\left(\frac{1}{2}\right)(i+x-2)+c(x)+f_{4}(i+x-2)\right\} \tag{2}
\end{equation*}
$$

In (2), $x$ must be a member of $\{0,1,2,3,4,5\}$, and $x$ must satisfy $4 \geq i+x-2 \geq 0$. This reflects the fact that the current month's demand must be met $(i+x-2 \geq 0)$, and ending inventory cannot exceed the capacity of $4(i+x-2 \leq 4)$. Recall that $x_{3}(i)$ is any value of $x$ attaining $f_{3}(i)$. The computations for $f_{3}(0), f_{3}(1), f_{3}(2), f_{3}(3)$, and $f_{3}(4)$ are given in Table 3 .

## Month 2 Computations

We can now determine $f_{2}(i)$, the minimum cost incurred during months 2,3 , and 4 given that at the beginning of month 2 , the on-hand inventory is $i$ units. Suppose that month 2 production $=x$. Because month 2 demand is 3 units, a holding cost of $\left(\frac{1}{2}\right)(i+x-3)$ is

## table 3

Computations for $f_{3}(i)$

| $i$ | $x$ | $\left(\frac{1}{2}\right)(i+x-2)+c(x)$ | $f_{4}(i+x-2)$ | Total Cost <br> Months 3,4 | $f_{3}(i)$ <br> $x_{3}(i)$ |
| :--- | :--- | :---: | :---: | :---: | :---: |
| 0 | 2 | $0+5=5$ | 7 | $5+7=12^{*}$ | $f_{3}(0)=12$ |
| 0 | 3 | $\frac{1}{2}+6=\frac{13}{2}$ | 6 | $\frac{13}{2}+6=\frac{25}{2}$ | $x_{3}(0)=2$ |
| 0 | 4 | $1+7=8$ | 5 | $8+5=13$ |  |
| 0 | 5 | $\frac{3}{2}+8=\frac{19}{2}$ | 4 | $\frac{19}{2}+4=\frac{27}{2}$ |  |
| 1 | 1 | $0+4=4$ | 7 | $4+7=11$ | $f_{3}(1)=10$ |
| 1 | 2 | $\frac{1}{2}+5=\frac{11}{2}$ | 6 | $\frac{11}{2}+6=\frac{23}{2}$ | $x_{3}(1)=5$ |
| 1 | 3 | $1+6=7$ | 5 | $7+5=12$ |  |
| 1 | 4 | $\frac{3}{2}+7=\frac{17}{2}$ | 4 | $\frac{17}{2}+4=\frac{25}{2}$ |  |
| 1 | 5 | $2+8=10$ | 0 | $10+0=10^{*}$ |  |
| 2 | 0 | $0+0=0$ | 7 | $0+7=7 *$ | $f_{3}(2)=7$ |
| 2 | 1 | $\frac{1}{2}+4=\frac{9}{2}$ | 6 | $\frac{9}{2}+6=\frac{21}{2}$ | $x_{3}(2)=0$ |
| 2 | 2 | $1+5=6$ | 5 | $6+5=11$ |  |
| 2 | 3 | $\frac{3}{2}+6=\frac{15}{2}$ | 4 | $\frac{15}{2}+4=\frac{23}{2}$ |  |
| 2 | 4 | $2+7=9$ | 0 | $9+0=9$ |  |
| 3 | 0 | $\frac{1}{2}+0=\frac{1}{2}$ | 6 | $\frac{1}{2}+6=\frac{13}{2} *$ | $f_{3}(3)=\frac{13}{2}$ |
| 3 | 1 | $1+4=5$ | 5 | $5+5=10$ | $x_{3}(3)=0$ |
| 3 | 2 | $\frac{3}{2}+5=\frac{13}{2}$ | 4 | $\frac{13}{2}+4=\frac{21}{2}$ |  |
| 3 | 3 | $2+6=8$ | 0 | $8+0=8$ |  |
| 4 | 0 | $1+0=1$ | 5 | $1+5=6^{*}$ | $f_{3}(4)=6$ |
| 4 | 1 | $\frac{3}{2}+4=\frac{11}{2}$ | 4 | $\frac{11}{2}+4=\frac{19}{2}$ | $x_{3}(4)=0$ |
| 4 | 2 | $2+5=7$ | 0 | $7+0=7$ |  |

incurred at the end of month 2. Thus, the total cost incurred during month 2 is $\left(\frac{1}{2}\right)(i+$ $x-3)+c(x)$. During months 3 and 4, we follow an optimal policy. Since month 3 begins with an inventory of $i+x-3$, the cost incurred during months 3 and 4 is $f_{3}(i+$ $x-3$ ). In analogy to (2), we now write

$$
\begin{equation*}
f_{2}(i)=\min _{x}\left\{\left(\frac{1}{2}\right)(i+x-3)+c(x)+f_{3}(i+x-3)\right\} \tag{3}
\end{equation*}
$$

where $x$ must be a member of $\{0,1,2,3,4,5\}$ and $x$ must also satisfy $0 \leq i+x-3 \leq$ 4. The computations for $f_{2}(0), f_{2}(1), f_{2}(2), f_{2}(3)$, and $f_{2}(4)$ are given in Table 4.

## Month 1 Computations

The reader should now be able to show that the $f_{1}(i)$ 's can be determined via the following recursive relation:

$$
\begin{equation*}
f_{1}(i)=\min _{x}\left\{\left(\frac{1}{2}\right)(i+x-1)+c(x)+f_{2}(i+x-1)\right\} \tag{4}
\end{equation*}
$$

where $x$ must be a member of $\{0,1,2,3,4,5\}$ and $x$ must satisfy $0 \leq i+x-1 \leq 4$. Since the inventory at the beginning of month 1 is 0 units, we actually need only determine $f_{1}(0)$ and $x_{1}(0)$. To give the reader more practice, however, the computations for $f_{1}(1), f_{1}(2), f_{1}(3)$, and $f_{1}(4)$ are given in Table 5.

## Determination of the Optimal Production Schedule

We can now determine a production schedule that minimizes the total cost of meeting the demand for all four months on time. Since our initial inventory is 0 units, the minimum cost for the four months will be $f_{1}(0)=\$ 20$. To attain $f_{1}(0)$, we must produce $x_{1}(0)=1$

| $i$ | $x$ | $\left(\frac{1}{2}\right)(i+x-3)+c(x)$ | $f_{3}(i+x-3)$ | Total Cost Months 2-4 | $\begin{aligned} & f_{2}(i) \\ & x_{2}(i) \\ & \hline \end{aligned}$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | 3 | $0+6=6$ | 12 | $6+12=18$ | $f_{2}(0)=16$ |
| 0 | 4 | $\frac{1}{2}+7=\frac{15}{2}$ | 10 | $\frac{15}{2}+10=\frac{35}{2}$ | $x_{2}(0)=5$ |
| 0 | 5 | $1+8=9$ | 7 | $9+7=16^{*}$ |  |
| 1 | 2 | $0+5=5$ | 12 | $5+12=17$ | $f_{2}(1)=15$ |
| 1 | 3 | $\frac{1}{2}+6=\frac{13}{2}$ | 10 | $\frac{13}{2}+10=\frac{33}{2}$ | $x_{2}(1)=4$ |
| 1 | 4 | $1+7=8$ | 7 | $8+7=15^{*}$ |  |
| 1 | 5 | $\frac{3}{2}+8=\frac{19}{2}$ | $\frac{13}{2}$ | $\frac{19}{2}+\frac{13}{2}=16$ |  |
| 2 | 1 | $0+4=4$ | 12 | $4+12=16$ | $f_{2}(2)=14$ |
| 2 | 2 | $\frac{1}{2}+5=\frac{11}{2}$ | 10 | $\frac{11}{2}+10=\frac{31}{2} *$ | $x_{2}(2)=3$ |
| 2 | 3 | $1+6=7$ | 7 | $7+7=14 *$ |  |
| 2 | 4 | $\frac{3}{2}+7=\frac{17}{2}$ | $\frac{13}{2}$ | $\frac{17}{2}+\frac{13}{2}=15$ |  |
| 2 | 5 | $2+8=10$ | 6 | $10+6=16$ |  |
| 3 | 0 | $0+0=0$ | 12 | $0+12=12$ * | $f_{2}(3)=12$ |
| 3 | 1 | $\frac{1}{2}+4=\frac{9}{2}$ | 10 | $\frac{9}{2}+10=\frac{29}{2}$ | $x_{2}(3)=0$ |
| 3 | 2 | $1+5=6$ | 7 | $6+7=13$ |  |
| 3 | 3 | $\frac{3}{2}+6=\frac{15}{2}$ | $\frac{13}{2}$ | $\frac{15}{2}+\frac{13}{2}=14$ |  |
| 3 | 4 | $2+7=9$ | 6 | $9+6=15$ |  |
| 4 | 0 | $\frac{1}{2}+0=\frac{1}{2}$ | 10 | $\frac{1}{2}+10=\frac{21}{2} *$ | $f_{2}(4)=\frac{21}{2}$ |
| 4 | 1 | $1+4=5$ | 7 | $5+7=12$ | $x_{2}(4)=0$ |
| 4 | 2 | $\frac{3}{2}+5=\frac{13}{2}$ | $\frac{13}{2}$ | $\frac{13}{2}+\frac{13}{2}=13$ |  |
| 4 | 3 | $2+6=8$ | 6 | $8+6=14$ |  |

TABLE 5
Computations for $f_{1}(i)$

| $i$ | $x$ | $\left(\frac{1}{2}\right)(i+x-1)+c(x)$ | $f_{2}(i+x-1)$ | Total Cost | $f_{1}(i)$ <br> $x_{1}(i)$ |
| :--- | :--- | :---: | :---: | :---: | :---: |
| 0 | 1 | $0+4=4$ | 16 | $4+16=20 *$ | $f_{1}(0)=20$ |
| 0 | 2 | $\frac{1}{2}+5=\frac{11}{2}$ | 15 | $\frac{11}{2}+15=\frac{41}{2}$ | $x_{1}(0)=1$ |
| 0 | 3 | $1+6=7$ | 14 | $7+14=21$ |  |
| 0 | 4 | $\frac{3}{2}+7=\frac{17}{2}$ | 12 | $\frac{17}{2}+12=\frac{41}{2}$ |  |
| 0 | 5 | $2+8=10$ | $\frac{21}{2}$ | $10+\frac{21}{2}=\frac{41}{2}$ |  |
| 1 | 0 | $0+0=0$ | 16 | $0+16=16 *$ | $f_{1}(1)=16$ |
| 1 | 1 | $\frac{1}{2}+4=\frac{9}{2}$ | 15 | $\frac{9}{2}+15=\frac{39}{2}$ | $x_{1}(1)=0$ |
| 1 | 2 | $1+5=6$ | 14 | 20 |  |
| 1 | 3 | $\frac{3}{2}+6=\frac{15}{2}$ | 12 | $\frac{15}{2}+12=\frac{39}{2}$ |  |
| 1 | 4 | $2+7=9$ | $\frac{21}{2}$ | $9+\frac{21}{2}=\frac{39}{2}$ |  |
| 2 | 0 | $\frac{1}{2}+0=\frac{1}{2}$ | 15 | $\frac{1}{2}+15=\frac{31}{2} *$ | $f_{1}(2)=\frac{31}{2}$ |
| 2 | 1 | $1+4=5$ | 14 | $5+14=19$ | $x_{1}(2)=0$ |
| 2 | 2 | $\frac{3}{2}+5=\frac{13}{2}$ | 12 | $\frac{13}{2}+12=\frac{37}{2}$ |  |
| 2 | 3 | $2+6=8$ | $\frac{21}{2}$ | $8+\frac{21}{2}=\frac{37}{2}$ |  |
| 3 | 0 | $1+0=1$ | 14 | $1+14=15 *$ | $f_{1}(3)=15$ |
| 3 | 1 | $\frac{3}{2}+4=\frac{11}{2}$ | 12 | $\frac{11}{2}+12=\frac{35}{2}$ | $x_{1}(3)=0$ |
| 3 | 2 | $2+5=7$ | $\frac{21}{2}$ | $7+\frac{21}{2}=\frac{35}{2}$ |  |
| 4 | 0 | $\frac{3}{2}+0=\frac{3}{2}$ | 12 | $\frac{3}{2}+12=\frac{27}{2} *$ | $f_{1}(4)=\frac{27}{2}$ |
| 4 | 1 | $2+4=6$ | $\frac{21}{2}$ | $6+\frac{21}{2}=\frac{33}{2}$ | $x_{1}(4)=0$ |

unit during month 1 . Then the inventory at the beginning of month 2 will be $0+1-$ $1=0$. Thus, in month 2 , we should produce $x_{2}(0)=5$ units. Then at the beginning of month 3 , our beginning inventory will be $0+5-3=2$. Hence, during month 3 , we need to produce $x_{3}(2)=0$ units. Then month 4 will begin with $2-2+0=0$ units on hand. Thus, $x_{4}(0)=4$ units should be produced during month 4 . In summary, the optimal production schedule incurs a total cost of $\$ 20$ and produces 1 unit during month 1 , 5 units during month 2,0 units during month 3 , and 4 units during month 4 .

Note that finding the solution to Example 4 is equivalent to finding the shortest route joining the node $(1,0)$ to the node $(5,0)$ in Figure 5. Each node in Figure 5 corresponds to a state, and each column of nodes corresponds to all the possible states associated with a given stage. For example, if we are at node $(2,3)$, then we are at the beginning of month 2 , and the inventory at the beginning of month 2 is 3 units. Each arc in the network represents the way in which a decision (how much to produce during the current month) transforms the current state into next month's state. For example, the arc joining nodes ( 1 , 0 ) and ( 2,2 ) (call it arc 1 ) corresponds to producing 3 units during month 1 . To see this, note that if 3 units are produced during month 1 , then we begin month 2 with $0+3-$ $1=2$ units. The length of each arc is simply the sum of production and inventory costs during the current period, given the current state and the decision associated with the chosen arc. For example, the cost associated with arc 1 would be $6+\left(\frac{1}{2}\right) 2=7$. Note that some nodes in adjacent stages are not joined by an arc. For example, node $(2,4)$ is not joined to node $(3,0)$. The reason for this is that if we begin month 2 with 4 units, then at the beginning of month 3 , we will have at least $4-3=1$ unit on hand. Also note that we have drawn arcs joining all month 4 states to the node ( 5,0 ), since having a positive inventory at the end of month 4 would clearly be suboptimal.

FIGURE 5 Network Representation of Inventory Example


Returning to Example 4, the minimum-cost production schedule corresponds to the shortest path joining $(1,0)$ and $(5,0)$. As we have already seen, this would be the path corresponding to production levels of 1, 5, 0, and 4. In Figure 5, this would correspond to the path beginning at $(1,0)$, then going to $(2,0+1-1)=(2,0)$, then to $(3,0+$ $5-3)=(3,2)$, then to $(4,2+0-2)=(4,0)$, and finally to $(5,0+4-4)=(5,0)$. Thus, our optimal production schedule corresponds to the path $(1,0)-(2,0)-(3,2)-(4$, $0)-(5,0)$ in Figure 5.

## PROBLEMS

## Group A

1 In Example 4, determine the optimal production schedule if the initial inventory is 3 units.

2 An electronics firm has a contract to deliver the following number of radios during the next three months; month 1, 200 radios; month 2, 300 radios; month 3, 300 radios. For each radio produced during months 1 and 2, a $\$ 10$ variable cost is incurred; for each radio produced during month 3, a $\$ 12$ variable cost is incurred. The inventory cost is $\$ 1.50$ for each radio in stock at the end of a month. The cost of setting up for production during a month is $\$ 250$.

Radios made during a month may be used to meet demand for that month or any future month. Assume that production during each month must be a multiple of 100 . Given that the initial inventory level is 0 units, use dynamic programming to determine an optimal production schedule.
3 In Figure 5, determine the production level and cost associated with each of the following arcs:
a $(2,3)-(3,1)$
b $(4,2)-(5,0)$

### 18.4 Resource-Allocation Problems

Resource-allocation problems, in which limited resources must be allocated among several activities, are often solved by dynamic programming. Recall that we have solved such problems by linear programming (for instance, the Giapetto problem). To use linear programming to do resource allocation, three assumptions must be made:

Assumption 1 The amount of a resource assigned to an activity may be any nonnegative number.


[^0]:    ${ }^{\dagger}$ The sign restrictions do constrain the values of the decision variables, but we choose to consider the sign restrictions as being separate from the constraints. The reason for this will become apparent when we study the simplex algorithm in Chapter 4.

[^1]:    A constraint is binding if the left-hand side and the right-hand side of the constraint are equal when the optimal values of the decision variables are substituted into the constraint.

[^2]:    ${ }^{\dagger}$ Lilien and Kotler (1983).

[^3]:    ${ }^{\dagger}$ Constraint (13) is satisfied by all points on or below $A B\left(A B\right.$ is $\left.\frac{1}{40} x_{1}+\frac{1}{60} x_{2}=1\right)$, and (14) is satisfied by all points on or below $C D\left(C D\right.$ is $\left.\frac{1}{50} x_{1}+\frac{1}{50} x_{2}=1\right)$.

[^4]:    ${ }^{\dagger}$ Throughout the first part of the chapter we assume that all variables must be nonnegative $(\geq 0)$. The conversion of urs variables to nonnegative variables is discussed in Section 4.12.

[^5]:    ${ }^{\dagger}$ In solving many LPs with 50 variables and $m \leq 50$ constraints, Chvàtal (1983) found that the simplex algorithm examined an average of $2 m$ basic feasible solutions before finding an LP's optimal solution.

[^6]:    ${ }^{\dagger}$ If a canonical form with nonnegative right-hand sides is not readily available, however, then the techniques described in Sections 4.12 and 4.13 can be used to find a canonical form and a basic feasible solution.

[^7]:    ${ }^{\dagger}$ To see that this tableau is optimal, note that from row $0, z=-12+5 x_{1}+3 s_{1}$. Because $x_{1} \geq 0$ and $s_{1} \geq 0$, this shows that $z \geq-12$. Thus, the current bfs (which has $z=-12$ ) must be optimal.

[^8]:    $\dagger$ "If and only if"

[^9]:    ${ }^{\dagger}$ In Problems 5 and 6 at the end of this section, we show that these rules are consistent with taking the dual of the transformed LP via (16) and (17).

[^10]:    ${ }^{\dagger}$ This assumes that after the right-hand side of Constraint $i$ has been changed to $b_{i}+1$, the current basis remains optimal.

[^11]:    ${ }^{\dagger}$ Based on Ravindran (1971).

[^12]:    ${ }^{\dagger}$ Based on Peterson (1990).
    ${ }^{\ddagger}$ This section covers topics that may be omitted with no loss of continuity.

[^13]:    ${ }^{\dagger}$ This section covers topics that may be omitted with no loss of continuity.

[^14]:    ${ }^{\dagger}$ For two subproblems created at the same time, many sophisticated methods have been developed to determine which one should be solved first. See Taha (1975) for details.
    ${ }^{\ddagger}$ The determinant of a matrix is defined in Section 2.6.

[^15]:    ${ }^{\dagger}$ Based on Haverly (1978).

[^16]:    ${ }^{\dagger}$ Recall from Chapter 3 that a set $S$ is convex if $x^{\prime} \in S$ and $x^{\prime \prime} \in S$ imply that all points on the line segment joining $x^{\prime}$ and $x^{\prime \prime}$ are members of $S$. This ensures that $c x^{\prime}+(1-c) x^{\prime \prime}$ will be a member of $S$.

[^17]:    ${ }^{\dagger}$ See Chapters 17 and 18 for an explanation of working backward (often called dynamic programming).

[^18]:    ${ }^{\dagger}$ To see why Theorem 1 fails to hold for a nonergodic chain, see Problems 11 and 12 at the end of this section. For a proof of this theorem, see Isaacson and Madsen (1976, Chapter 3).

[^19]:    ${ }^{\dagger}$ This section covers topics that may be omitted with no loss of continuity.

[^20]:    ${ }^{\dagger}$ In this example, keeping track of the stages is unnecessary; to be consistent with later examples, however, we do keep track.

