Optimization

Part 4

- Root finding and optimization are related, both involve guessing and searching for a point on a function.
- Fundamental difference is:
  - Root finding is searching for zeros of a function or functions
  - Optimization is finding the minimum or the maximum of a function of several variables.

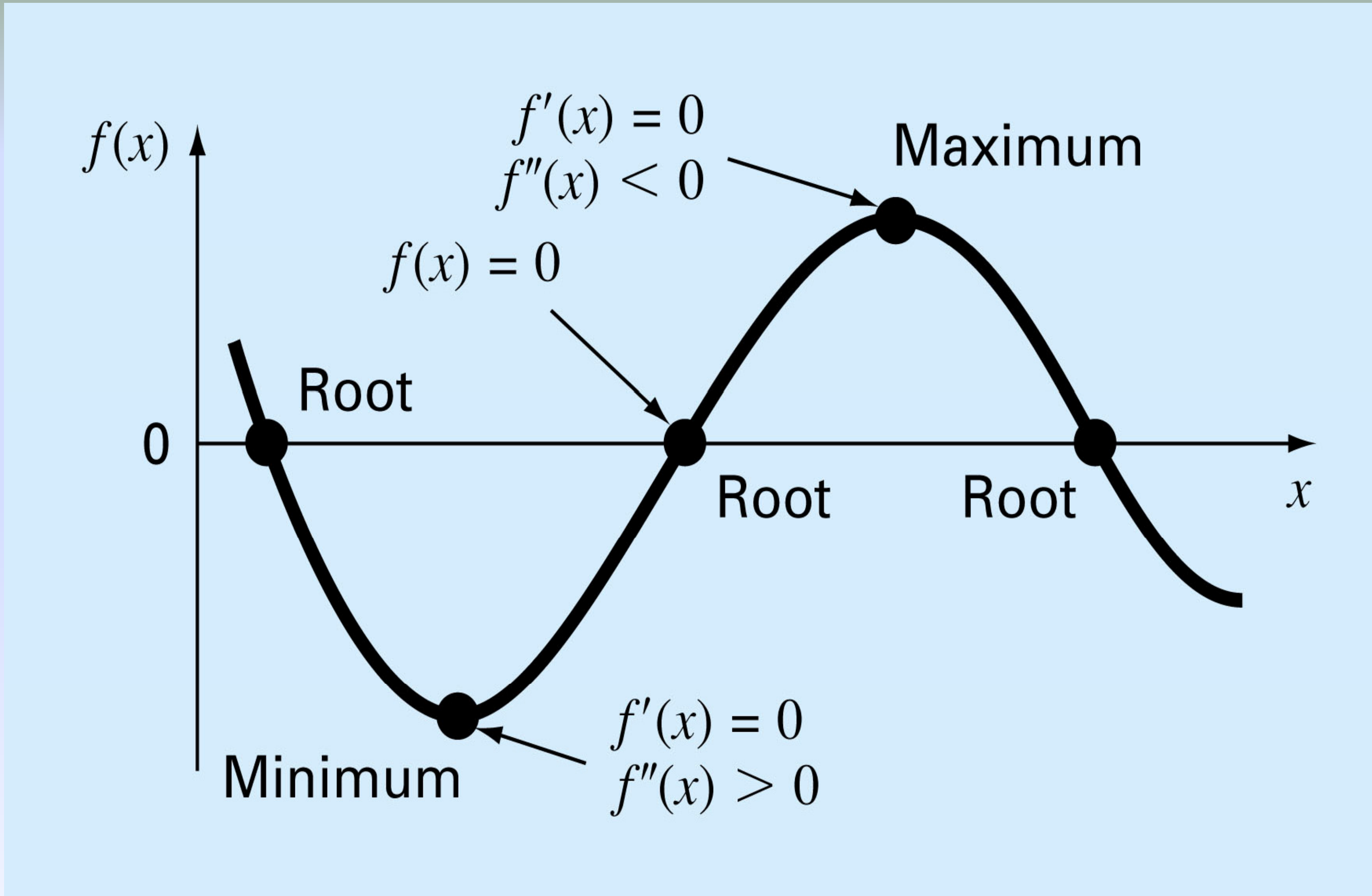


FIGURE PT4.1
A function of a single variable illustrating the difference between roots and optima.

Mathematical Background

- An *optimization* or *mathematical programming* problem generally be stated as:

  Find x, which minimizes or maximizes $f(x)$ subject to

$$d_i(x) \le a_i \quad i = 1,2,\ldots,m^*$$
$$e_i(x) = b_i \quad i = 1,2,\ldots,p^*$$

Where x is an n-dimensional *design vector*, $f(x)$ is the *objective function*, $d_i(x)$ are *inequality constraints*, $e_i(x)$ are *equality constraints*, and a_i and b_i are constants

- Optimization problems can be classified on the basis of the form of *f(x)*:
  - If *f(x)* and the constraints are linear, we have linear programming.
  - If *f(x)* is quadratic and the constraints are linear, we have quadratic programming.
  - If *f(x)* is not linear or quadratic and/or the constraints are nonlinear, we have nonlinear programming.
- When equations(*) are included, we have a *constrained optimization* problem; otherwise, it is *unconstrained optimization* problem.

One-Dimensional Unconstrained Optimization

Chapter 13

- In *multimodal* functions, both local and global optima can occur. In almost all cases, we are interested in finding the absolute highest or lowest value of a function.

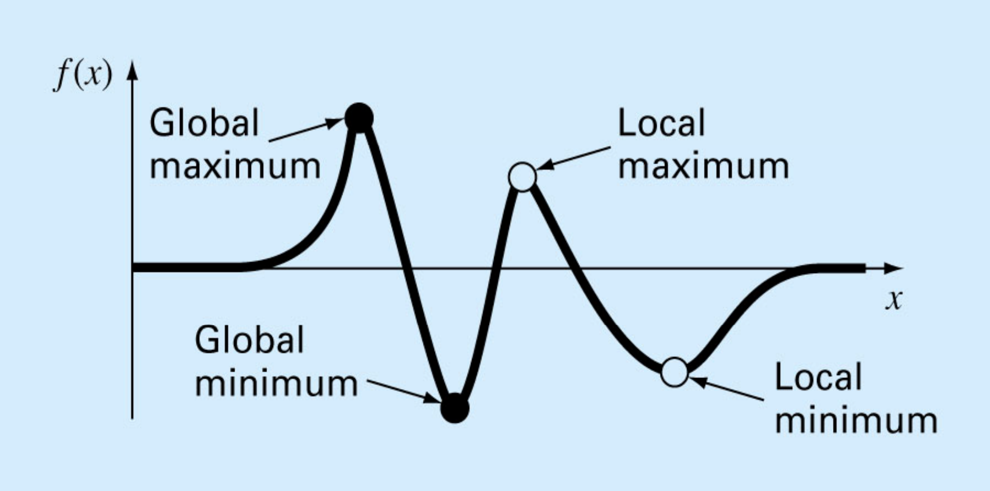


FIGURE 13.1
A function that asymptotically approaches zero at plus and minus ∞ and has two maximum and two minimum points in the vicinity of the origin. The two points to the right are local optima, whereas the two to the left are global.

How do we distinguish global optimum from a local one?

- By graphing to gain insight into the behavior of the function.
- Using randomly generated starting guesses and picking the largest of the optima as global.
- Perturbing the starting point to see if the routine returns a better point or the same local minimum.

Golden-Section Search

- A *unimodal* function has a single maximum or a minimum in the a given interval. For a *unimodal* function:
  - First pick two points that will bracket your extremum [x_l, x_u].
  - Pick an additional third point within this interval to determine whether a maximum occurred.
  - Then pick a fourth point to determine whether the maximum has occurred within the first three or last three points.
  - The key is making this approach efficient by choosing intermediate points wisely thus minimizing the function evaluations by replacing the old values with new values.

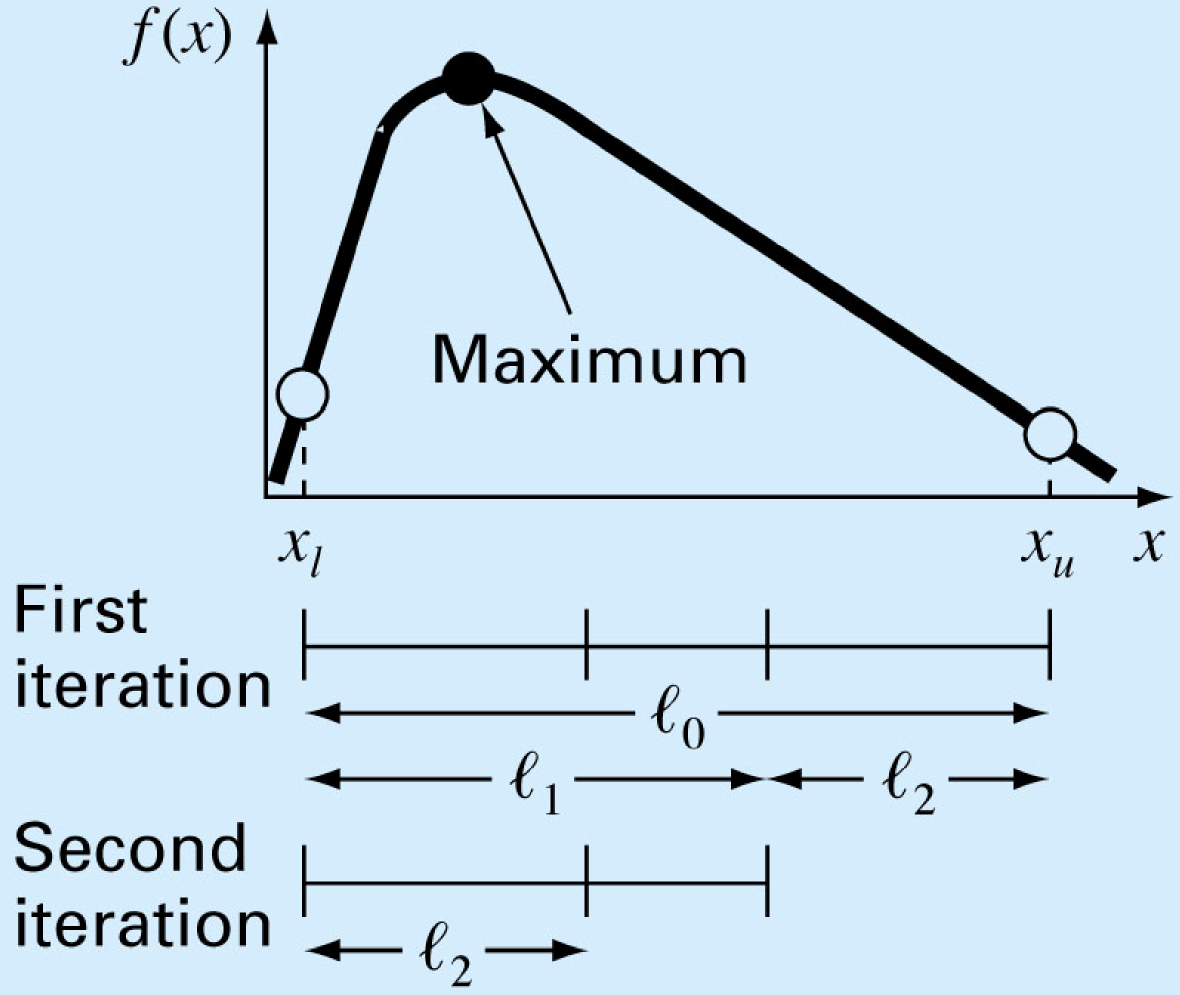
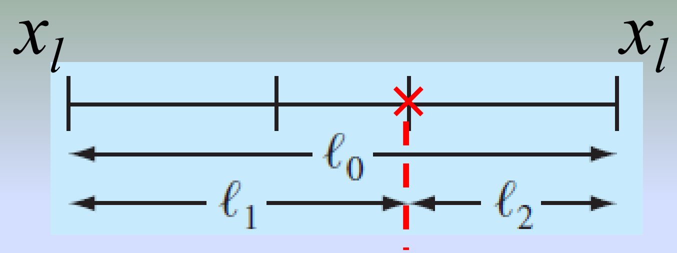


FIGURE 13.2
The initial step of the golden-section search algorithm involves choosing two interior points according to the golden ratio.

$$l_0 = l_1 + l_2$$

$$\frac{l_1}{l_0} = \frac{l_2}{l_1}$$



- The first condition specifies that the sum of the two sub lengths l_1 and l_2 must equal the original interval length.

- The second say that the ratio of the lengths must be equal

$$\frac{l_1}{l_1+l_2}=\frac{l_2}{l_1} \qquad R=\frac{l_2}{l_1} \qquad 1+R=\frac{1}{R} \qquad R^2+R-1=0 \qquad R=\frac{-1\pm\sqrt{1-4(-1)}}{2}=0.61803;\ -1.61803$$

$$\frac{l_1}{l_1+l_2}=\frac{l_2}{l_1} \qquad S=\frac{l_1}{l_2} \qquad \frac{1}{1+\frac{1}{S}}=\frac{1}{S} \qquad S^2-S-1=0 \qquad S=\frac{1\pm\sqrt{1-4(-1)}}{2}=1.61803;\ -0.61803$$

Golden Ratio: $\frac{\sqrt{5}\pm 1}{2}$

R=(small)/(big)=**0.61803** or S=(big)/(small)=**1.61803**

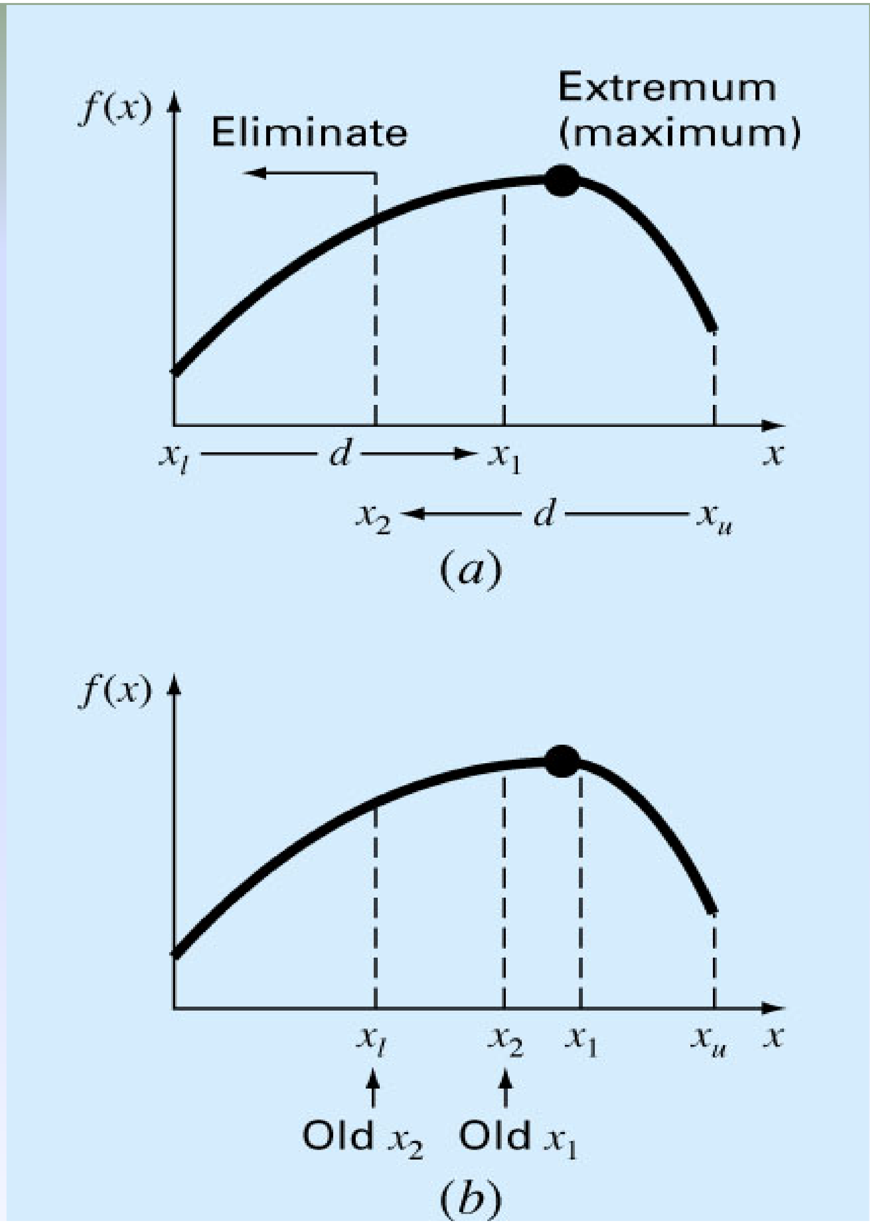


FIGURE 13.4

(*a*) The initial step of the golden-section search algorithm involves choosing two interior points according to the golden ratio. (*b*) The second step involves defining a new interval that includes the optimum.

Golden-Section Search Method:

The method starts with two initial guesses, x_l and x_u, that bracket one local extremum of $f(x)$:

- Next two interior points x_1 and x_2 are chosen according to the golden ratio

$$d = \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

$$x_1 = x_l + d$$

$$x_2 = x_u - d$$

- The function is evaluated at these two interior points.

Two results can occur:

- If $f(x_1)>f(x_2)$ then the domain of x to the left of x_2 from x_l to x_2, can be eliminated because it does not contain the maximum. Then, x_2 becomes the new x_l for the next round.

- If $f(x_2)>f(x_1)$, then the domain of x to the right of x_1 from x_l to x_2, would have been eliminated. In this case, x_1 becomes the new x_u for the next round.

New $x_{1\ (i)}$'s determined as before

$$x_1 = x_l + \frac{\sqrt{5}-1}{2}(x_u - x_l)$$

- The real benefit from the use of golden ratio is because the original x_1 and x_2 were chosen using golden ratio, we do not need to recalculate all the function values for the next iteration.

HOMEWORK:

See the Pseudocode for Golden-Section search algorithm in page 362. Write your MATLAB codes to test this algorithm with Example 13.1 in page 360.

362

FIGURE 13.5
Algorithm for the golden-section search.

360

EXAMPLE 13.1

Parabolic Interpolation

Parabolic interpolation takes advantage of the fact that a second-order polynomial often provides a good approximation to the shape of $f(x)$ near an optimum.

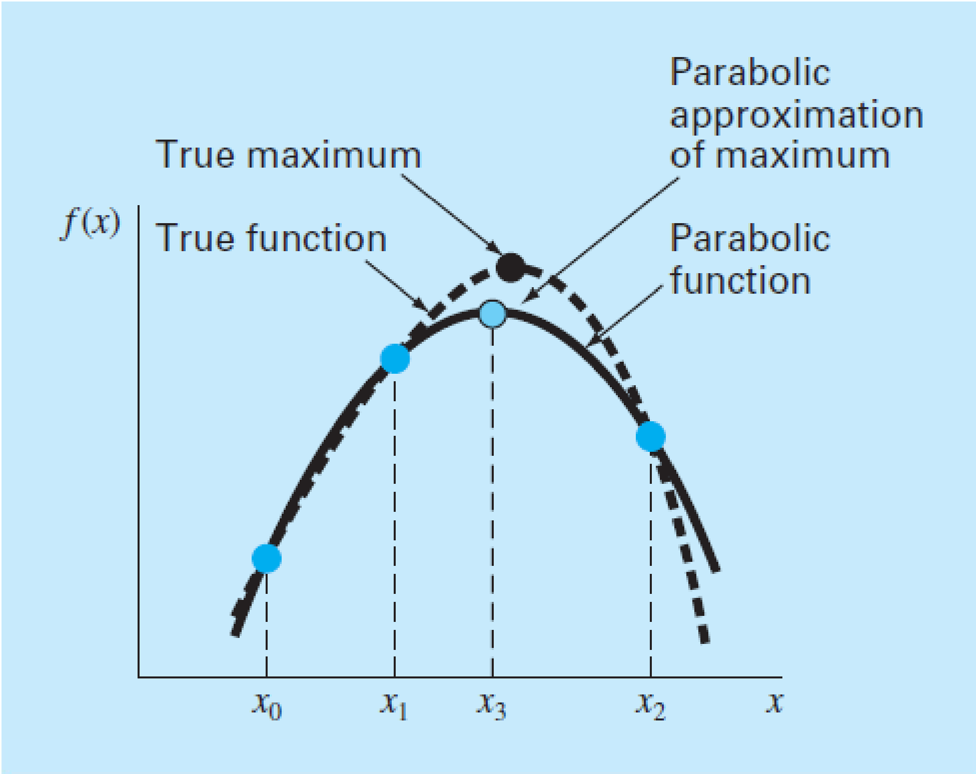
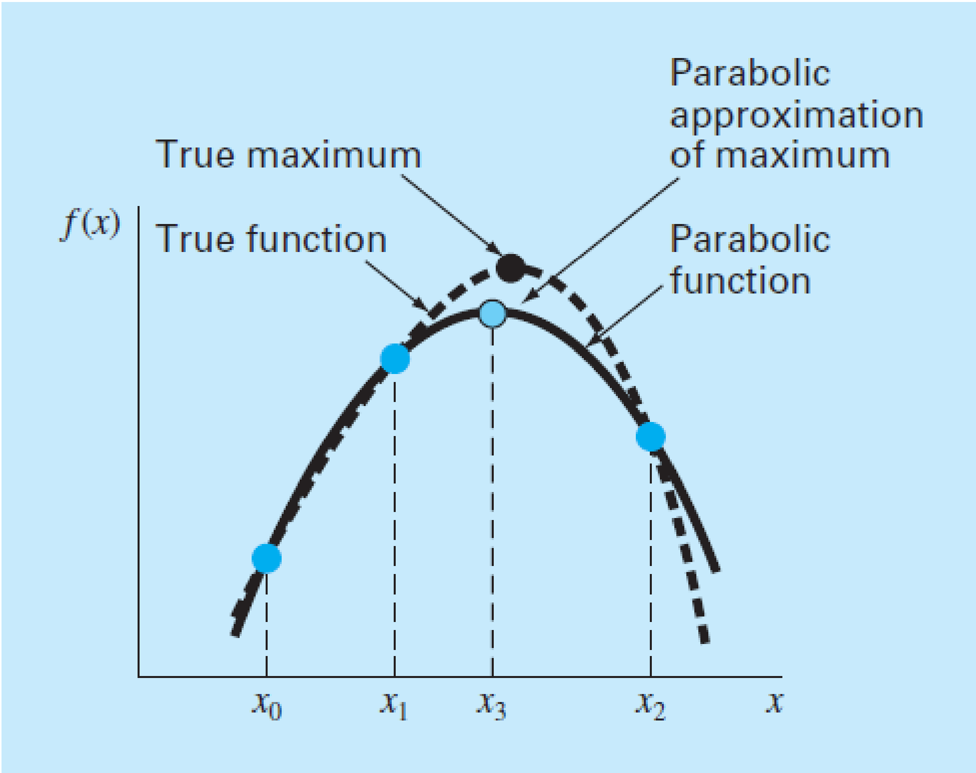


FIGURE 13.6
Graphical description of parabolic interpolation.

Parabolic Interpolation

Just as there is only one straight line connecting two points, *there is only one quadratic polynomial or parabola connecting three points*. Thus, if we have three points that jointly bracket an optimum, we can fit a parabola to the points.



$$x_3 = \frac{f(x_0)(x_1^2 - x_2^2) + f(x_1)(x_2^2 - x_0^2) + f(x_2)(x_0^2 - x_1^2)}{2f(x_0)(x_1 - x_2) + 2f(x_1)(x_2 - x_0) + 2f(x_2)(x_0 - x_1)}$$

where x_0, x_1, and x_2 are the initial guesses, and x_3 is the value of x that corresponds to the maximum value of the parabolic fit to the guesses.

Newton's Method

- A similar approach to Newton- Raphson method can be used to find an optimum of $f(x)$ by defining a new function $g(x)=f'(x)$. Thus because the same optimal value x^* satisfies both

  $f'(x^*)=g(x^*)=0$

  We can use the following as a technique to the extremum of $f(x)$.

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}$$

Homework

- Modify the *m-file* written for Newton- Raphson method (*p2_NewtonRaphson_basic_algorithm.m*) for the Newton's method and test your code for Example 13.3 page 365.

MATLAB: Finding the minimum of a single variable function

MATLAB Function: fminbnd

```
>> help fminbnd
fminbnd - Find minimum of single-variable function on fixed interval

    This MATLAB function returns a value x that is a local minimizer of the function
    that is described in fun in the interval x1 < x < x2.

    x = fminbnd(fun,x1,x2)
    x = fminbnd(fun,x1,x2,options)
    [x,fval] = fminbnd(...)
    [x,fval,exitflag] = fminbnd(...)
    [x,fval,exitflag,output] = fminbnd(...)

    Reference page for fminbnd

    See also fminsearch, function_handle, fzero, optimset

    Other functions named fminbnd
        optim/fminbnd
```

MATLAB: Finding the minimum of a multivariable function

MATLAB Function: fminsearch

```
>> help fminsearch
fminsearch - Find minimum of unconstrained multivariable function using derivative-free method

    This MATLAB function starts at the point x0 and returns a value x that is a
    local minimizer of the function described in fun.

    x = fminsearch(fun,x0)
    x = fminsearch(fun,x0,options)
    [x,fval] = fminsearch(...)
    [x,fval,exitflag] = fminsearch(...)
    [x,fval,exitflag,output] = fminsearch(...)

    Reference page for fminsearch

    See also fminbnd, function_handle, optimset
```

```
>> f=@(x) 2+x(1)-x(2)+2*x(1)^2+2*x(1)*x(2)+x(2)^2;
>> [x,fval]=fminsearch(f,[-0.5,0.5])

x =
    -1.0000    1.5000
fval =
     0.7500
```

Multidimensional Unconstrained Optimization

Chapter 14

- Techniques to find minimum and maximum of a function of several variables are described.
- These techniques are classified as:
  - That require derivative evaluation
    - *Gradient* or descent (or *ascent*) methods
  - That do not require derivative evaluation
    - *Non-gradient* or *direct* methods.

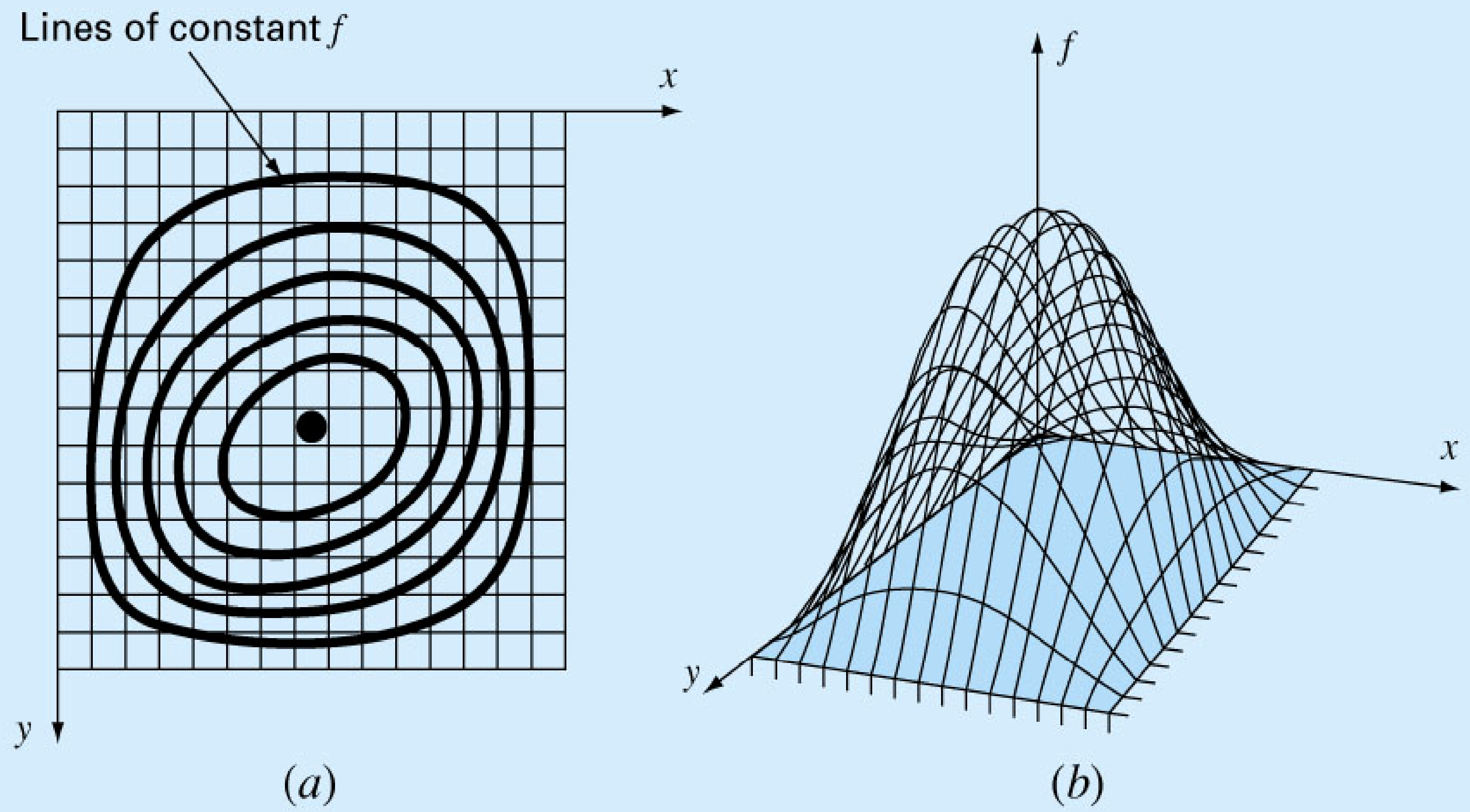


FIGURE 14.1 The most tangible way to visualize two-dimensional searches is in the context of ascending a mountain (maximization) or descending into a valley (minimization).
(a) A 2-D topographic map that corresponds to the 3-D mountain in (b).

DIRECT METHODS

Random Search

- Based on evaluation of the function randomly at selected values of the independent variables.
- If a sufficient number of samples are conducted, the optimum will be eventually located.
- Example: maximum of a function

$$f(x, y)=y-x-2x^2-2xy-y^2$$

can be found using a random number generator.

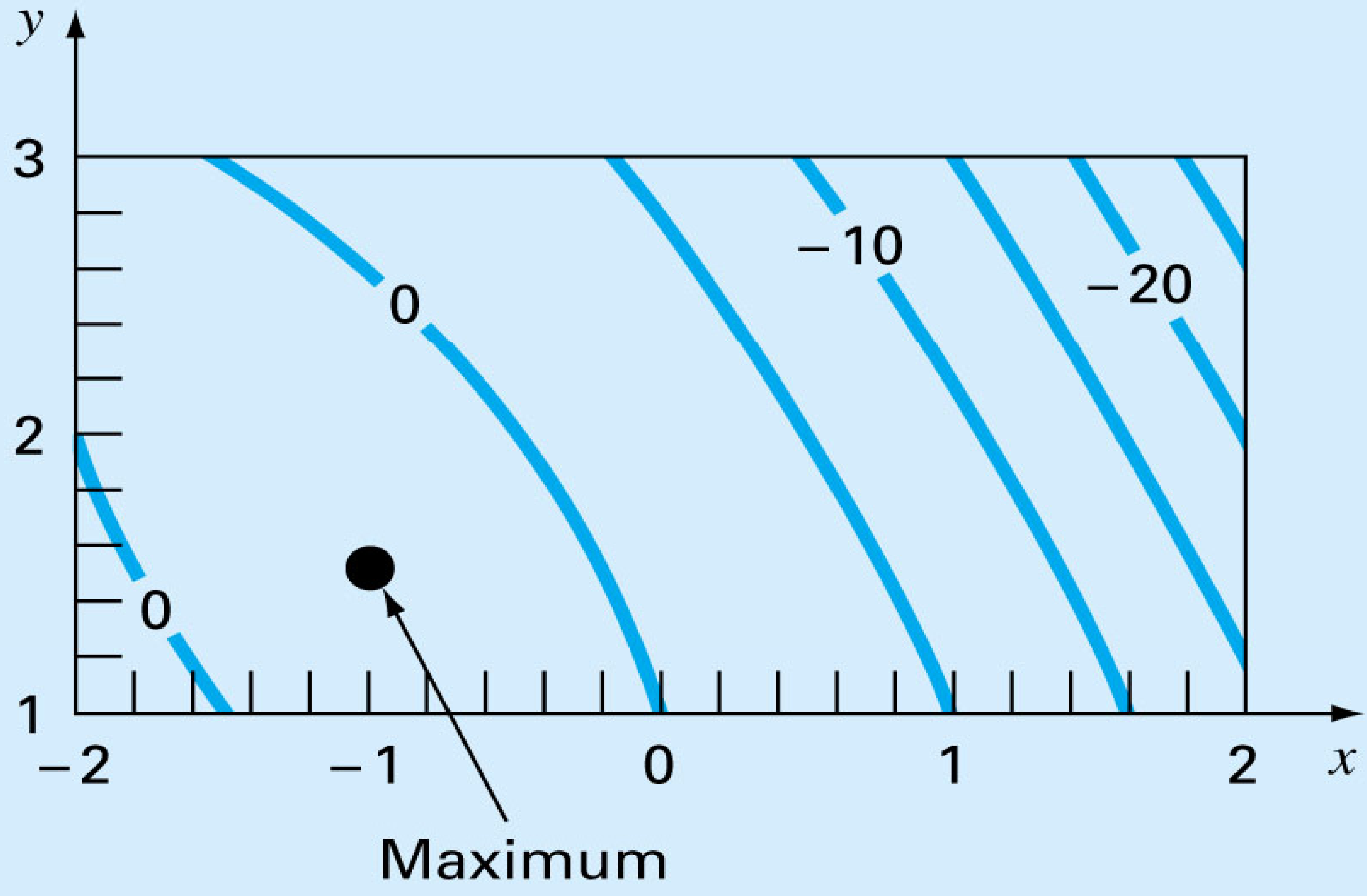


FIGURE 14.2
Equation (E14.1.1) showing the maximum at $x = -1$ and $y = 1.5$.

Advantages/

- Works even for discontinuous and nondifferentiable functions.
- Always finds the global optimum rather than the global minimum.

Disadvantages/

- As the number of independent variables grows, the task can become onerous.
- Not efficient, it does not account for the behavior of underlying function.

Univariate and Pattern Searches

- More efficient than random search and still doesn’t require derivative evaluation.
- The basic strategy is:
  - Change one variable at a time while the other variables are held constant.
  - Thus problem is reduced to a sequence of one-dimensional searches that can be solved by variety of methods.
  - The search becomes less efficient as you approach the maximum.

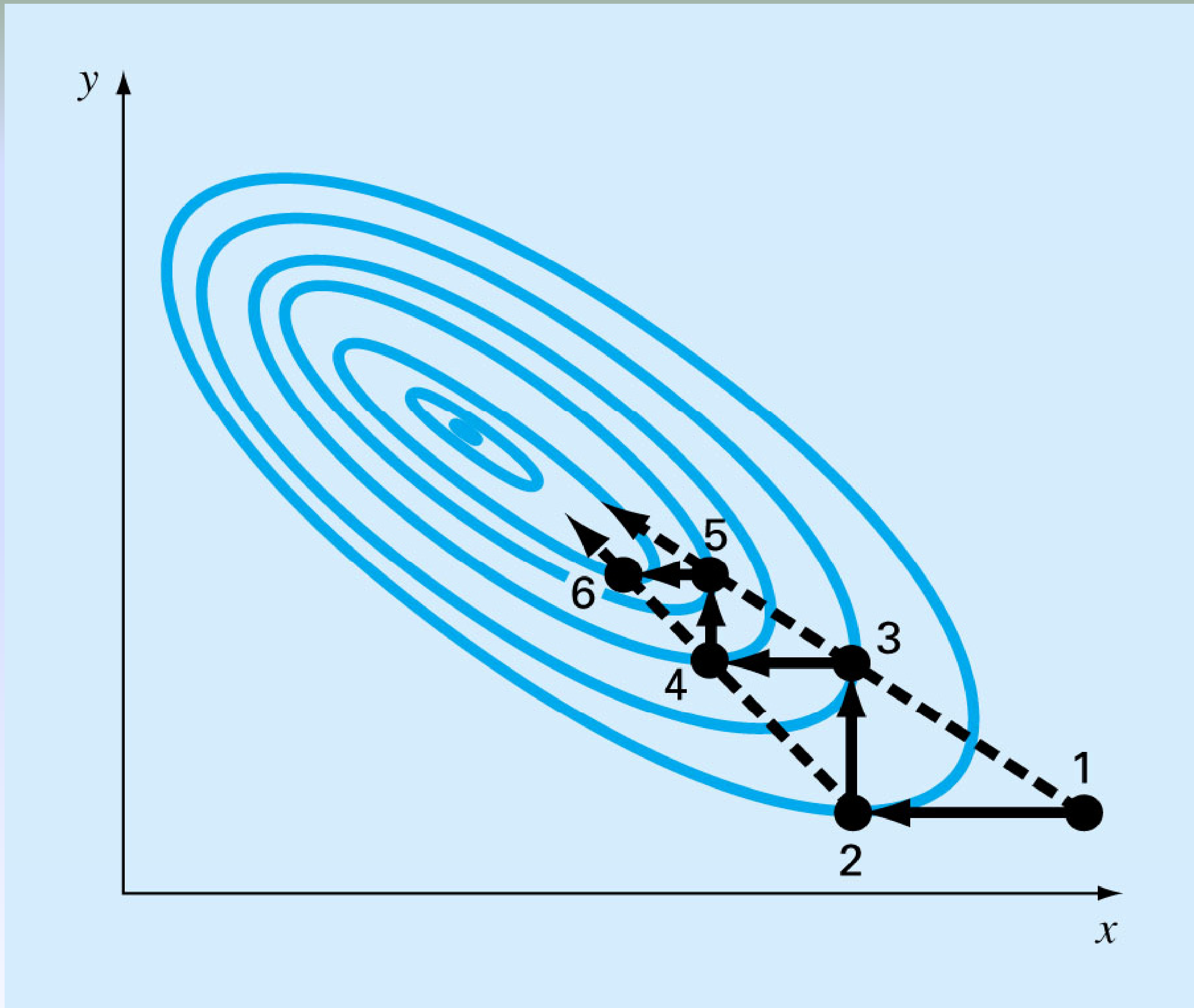


FIGURE 14.3
A graphical depiction of how a univariate search is conducted.

- *Pattern directions* can be used to shoot directly along the ridge towards maximum.

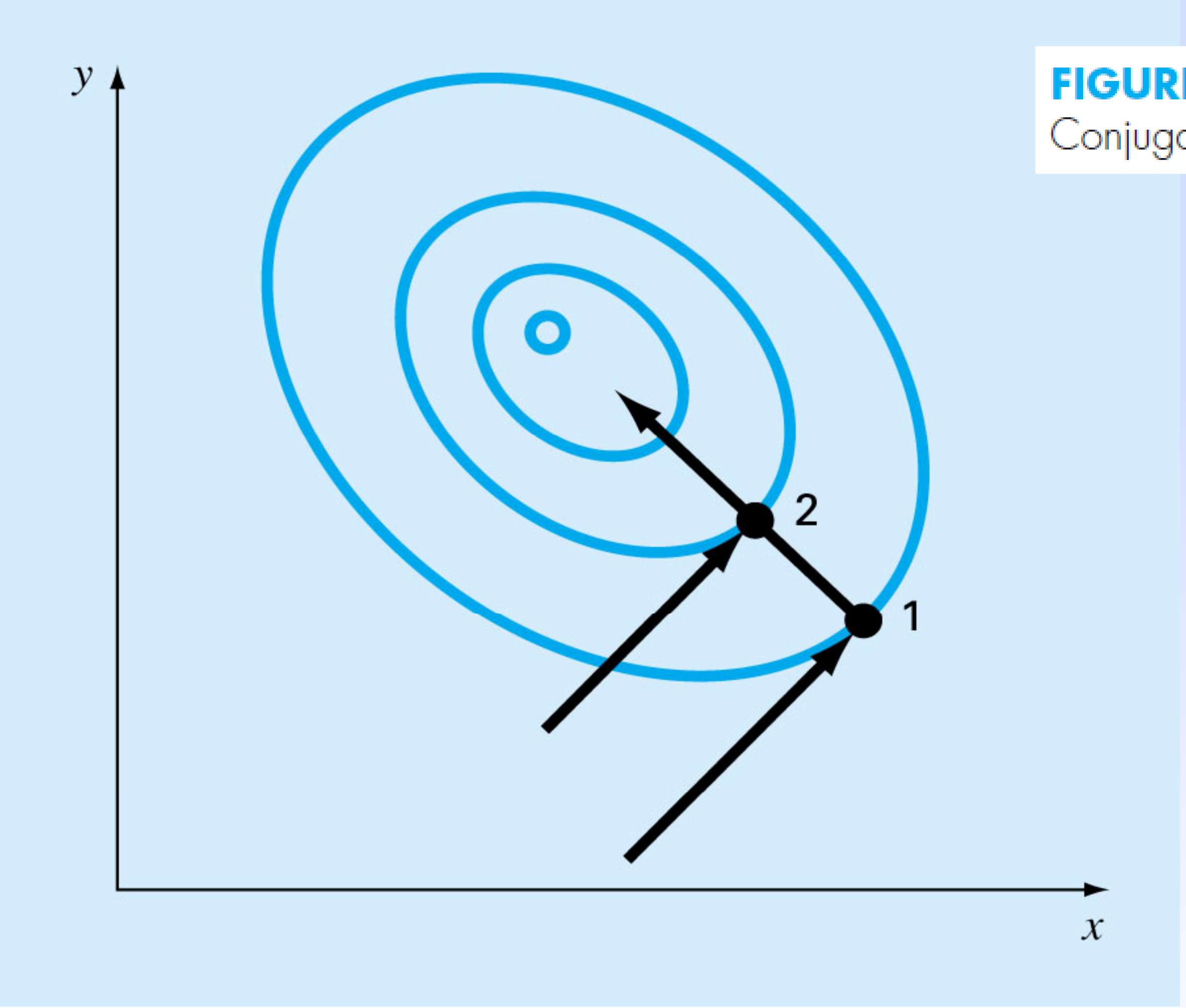


FIGURE 14.4
Conjugate directions.

- Best known algorithm, *Powell's method*, is based on the observation that if points 1 and 2 are obtained by one-dimensional searches in the same direction but from different starting points, then, the line formed by 1 and 2 will be directed toward the maximum. Such lines are called *conjugate directions*.

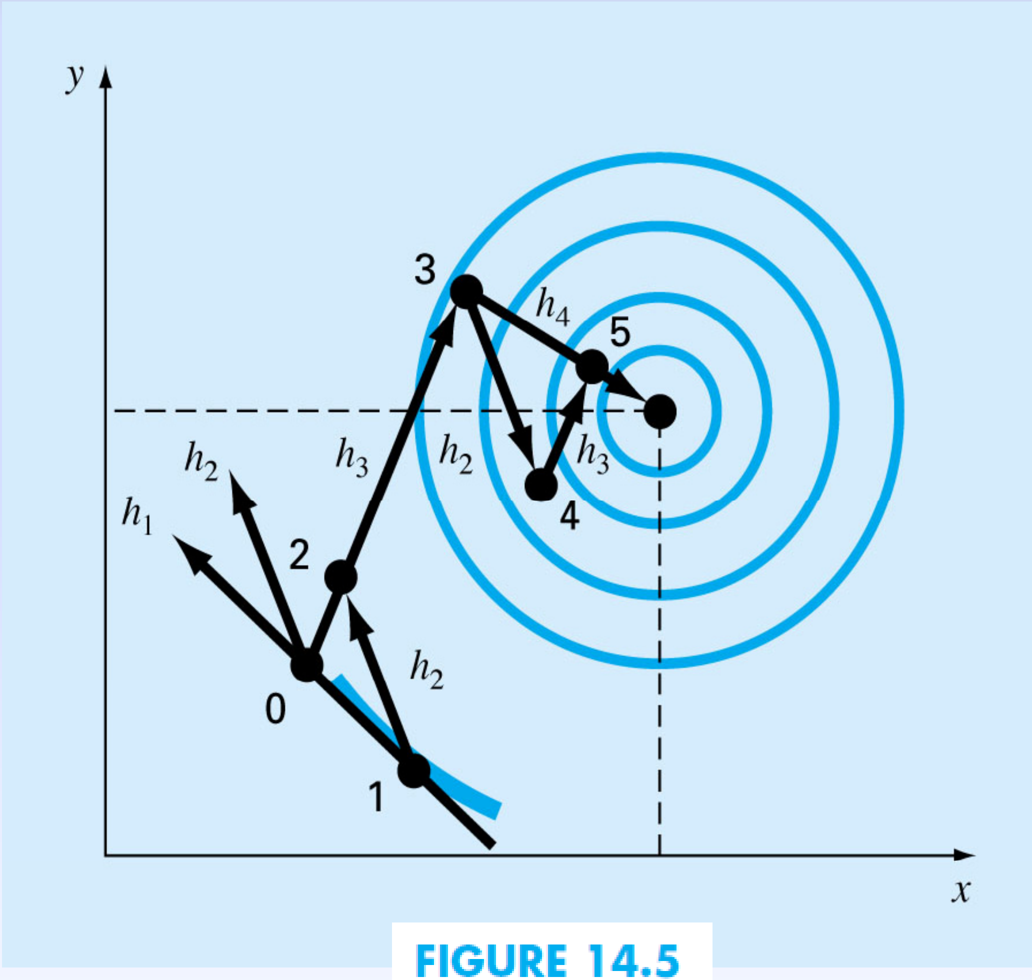


FIGURE 14.5
Powell's method.

GRADIENT METHODS
Gradients and Hessians

The Gradient:

- If *f(x,y)* is a two dimensional function, the *gradient* vector tells us
  - What direction is the steepest ascend?
  - How much we will gain by taking that step?

$$\nabla f = \frac{\partial f}{\partial x}\mathrm{i} + \frac{\partial f}{\partial y}\mathrm{j} \qquad \text{or } \textit{del f}$$

Directional derivative of *f(x,y)* at point *x=a* and *y=b*

•For n dimensions:

$$\nabla f(x) = \begin{Bmatrix} \dfrac{\partial f}{\partial x_1}(\mathrm{x}) \\ \dfrac{\partial f}{\partial x_2}(\mathrm{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\mathrm{x}) \end{Bmatrix}$$

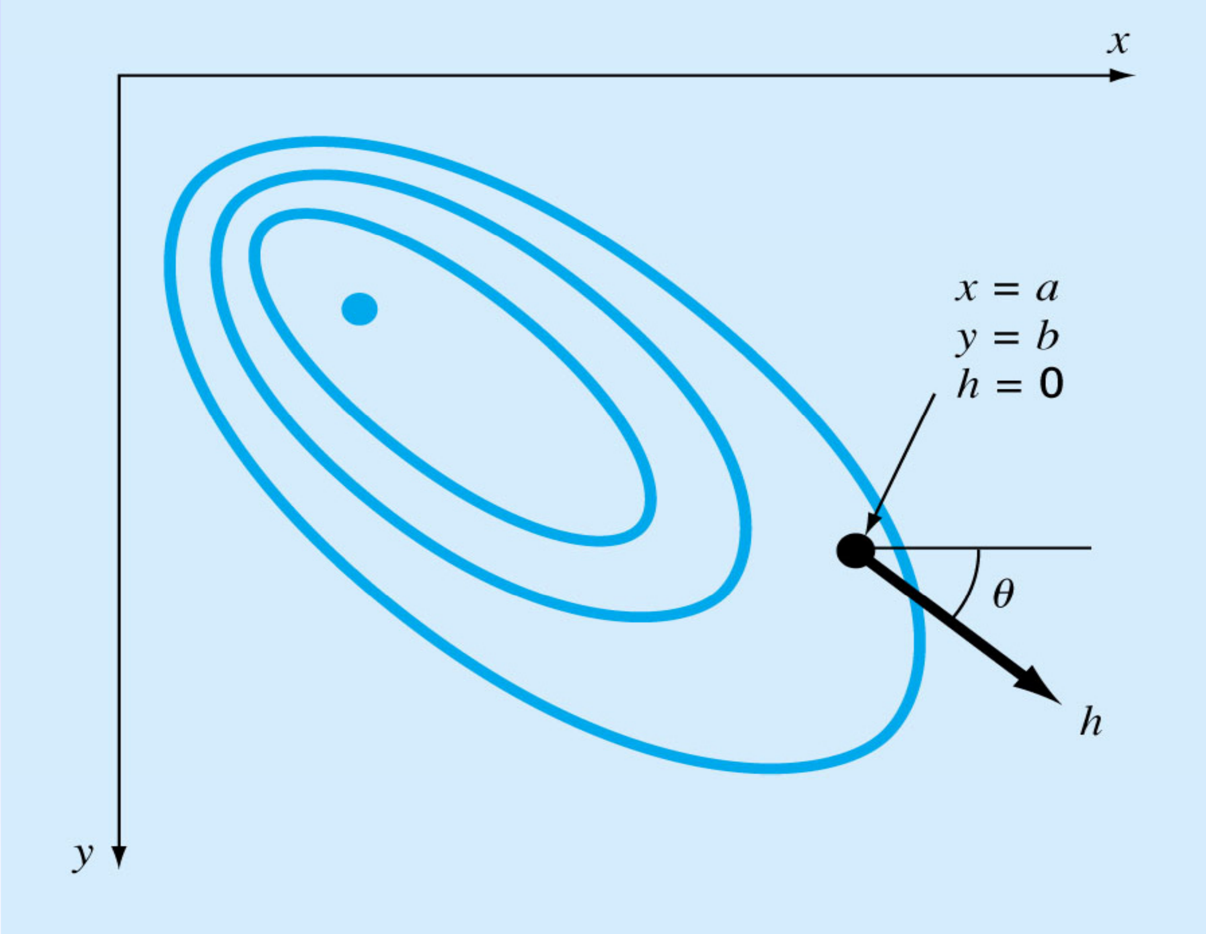
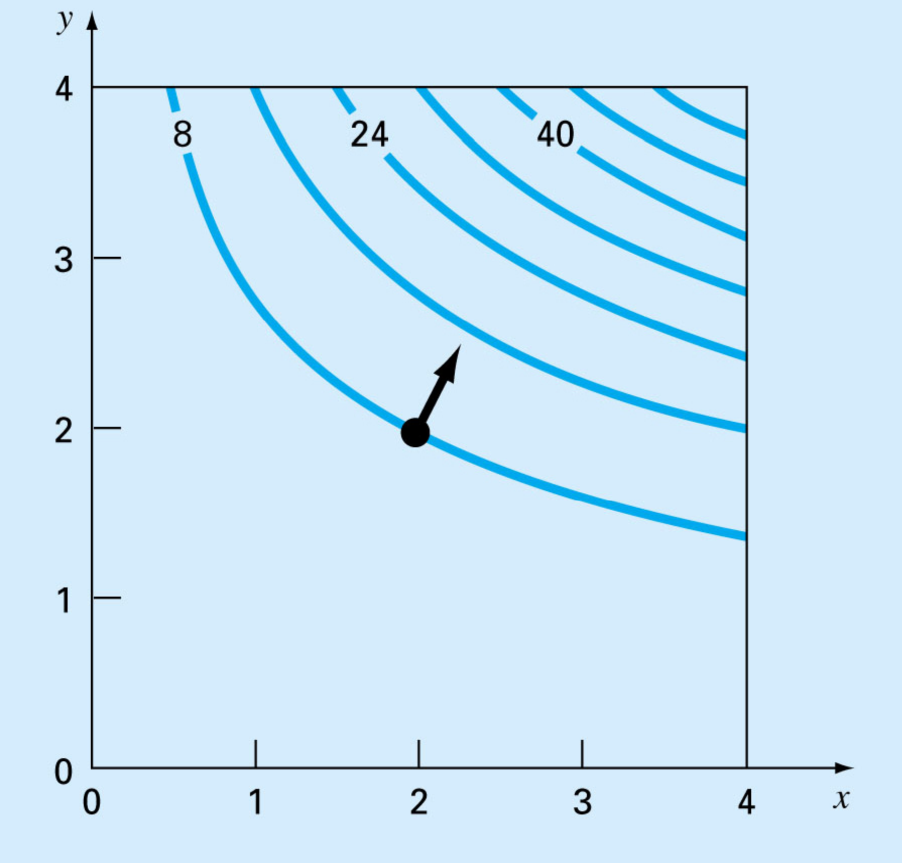


FIGURE 14.6
The directional gradient is defined along an axis h that forms an angle θ with the x axis.

The Hessian:

- For one dimensional functions both first and second derivatives valuable information for searching out optima.
  - First derivative provides (a) the steepest trajectory of the function and (b) tells us that we have reached the maximum.
  - Second derivative tells us that whether we are a maximum or minimum.
- For two dimensional functions whether a maximum or a minimum occurs involves not only the partial derivatives w.r.t. x and y but also the second partials w.r.t. x and y.

FIGURE 14.7
The arrow follows the direction of steepest ascent calculated with the gradient.



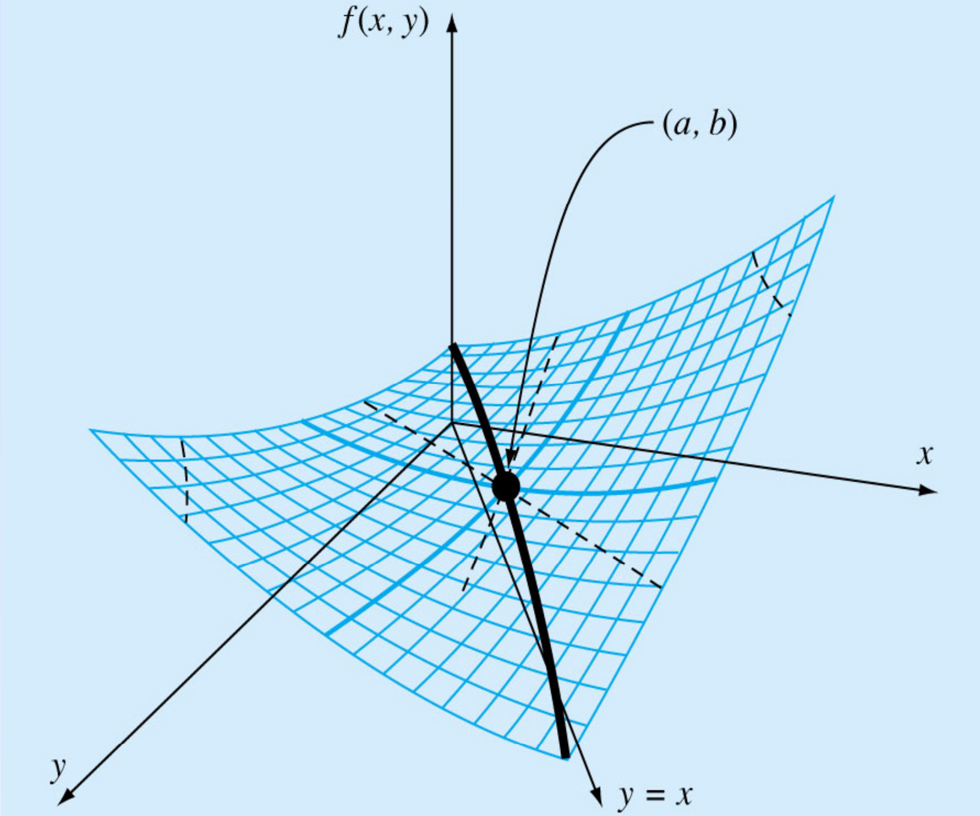


FIGURE 14.8
A saddle point ($x = a$ and $y = b$). Notice that when the curve is viewed along the x and y directions, the function appears to go through a minimum (positive second derivative), whereas when viewed along an axis $x = y$, it is concave downward (negative second derivative).

- Assuming that the partial derivatives are continuous at and near the point being evaluated

$$|H| = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2$$

$$\text{If } |\mathrm{H}| \succ 0 \text{ and } \frac{\partial^2 f}{\partial x^2} \succ 0, \text{ then f(x, y) has a local minimum}$$

$$\text{If } |\mathrm{H}| \succ 0 \text{ and } \frac{\partial^2 f}{\partial x^2} \prec 0, \text{ then f(x, y) has a local minimum}$$

$$\text{If } |\mathrm{H}| \prec 0, \text{ then f(x, y) has a saddle point}$$

The quantity [**H**] is equal to the determinant of a matrix made up of second derivatives

The Steepest Ascend Method

- Start at an initial point (x_o, y_o), determine the direction of steepest ascend, that is, the gradient. Then search along the direction of the gradient, h_o, until we find maximum. Process is then repeated.

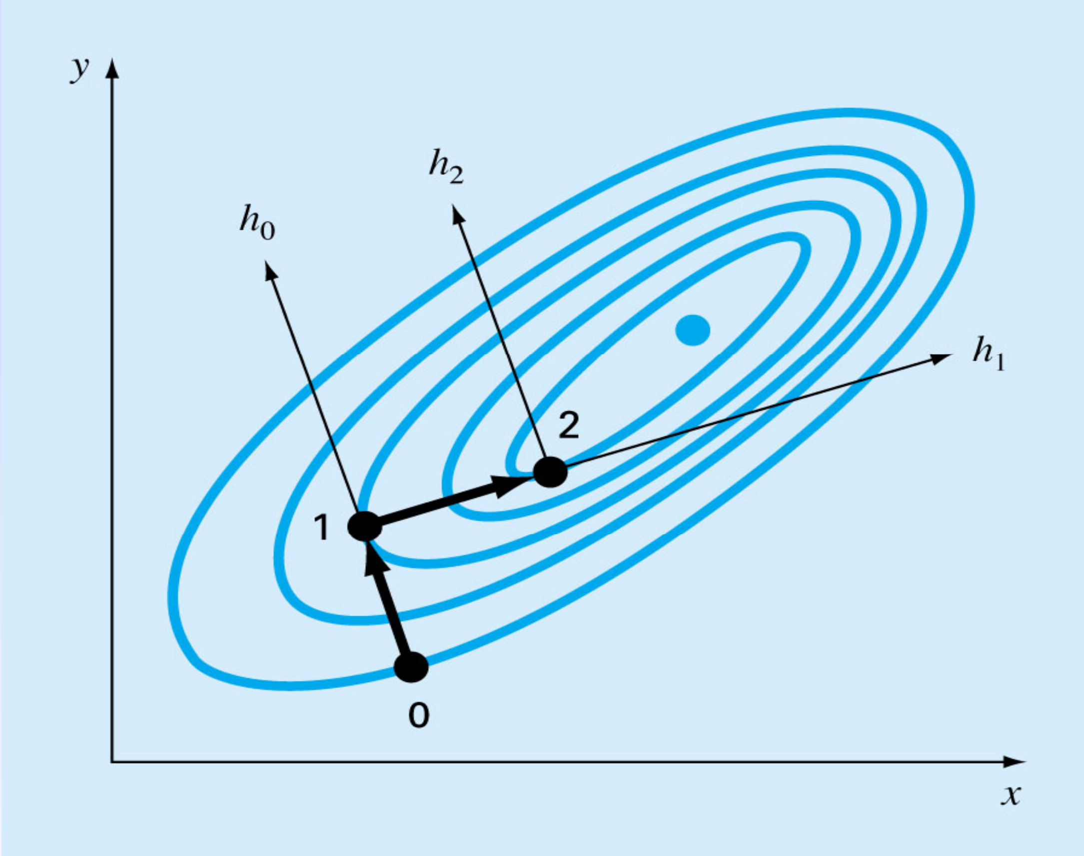


FIGURE 14.9
A graphical depiction of the method of steepest ascent.

- The problem has two parts
  - Determining the "best direction" and
  - Determining the "best value" along that search direction.
- Steepest ascent method uses the gradient approach as its choice for the "best" direction.
- To transform a function of *x* and *y* into a function of *h* along the gradient section:

$$x = x_o + \frac{\partial f}{\partial x} h$$

$$y = y_o + \frac{\partial f}{\partial y} h$$

h is distance along the h axis

- If x_o=1 and y_o=2

$$\nabla f = 3i + 4j$$

$$x = 1 + 3h$$

$$y = 2 + 4h$$

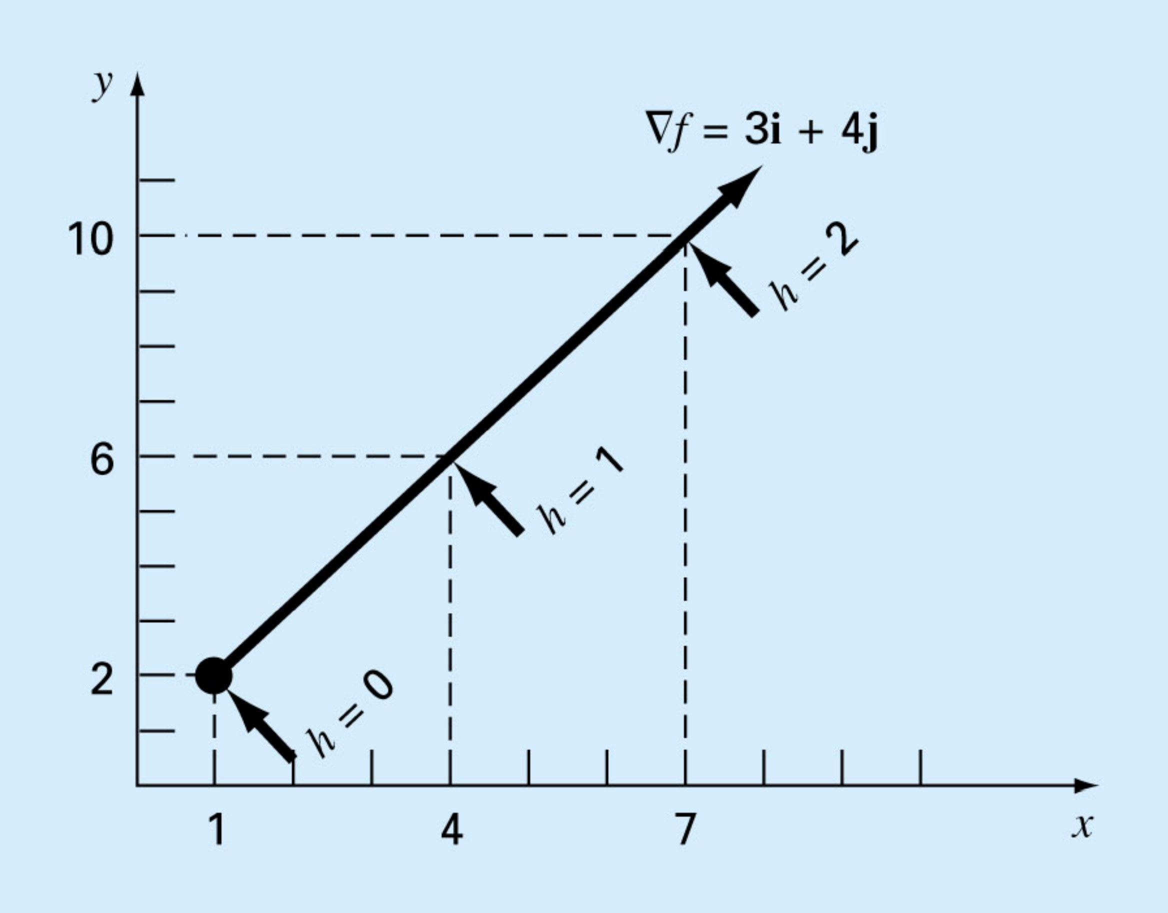


FIGURE 14.10
The relationship between an arbitrary direction *h* and *x* and *y* coordinates.

Constrained Optimization
Chapter 15

LINEAR PROGRAMMING

- An optimization approach that deals with meeting a desired objective such as maximizing profit or minimizing cost in presence of constraints such as limited resources
- Mathematical functions representing both the objective and the constraints are linear.

Standard Form:

- Basic linear programming problem consists of two major parts:
  - The objective function
  - A set of constraints
- For maximization problem, the objective function is generally expressed as

$$\text{Maximize } Z = c_1 x_1 + c_2 x_2 + \cdots + c_n x_n$$

c_j= payoff of each unit of the *j*th activity that is undertaken

x_j= magnitude of the *j*th activity

Z= total payoff due to the total number of activities

- The constraints can be represented generally as

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \leq b_i$$

- Where a_{ij}=amount of the ith resource that is consumed for each unit of the jth activity and b_i=amount of the ith resource that is available
- The general second type of constraint specifies that all activities must have a positive value, x_i>0 .
- Together, the objective function and the constraints specify the linear programming problem.

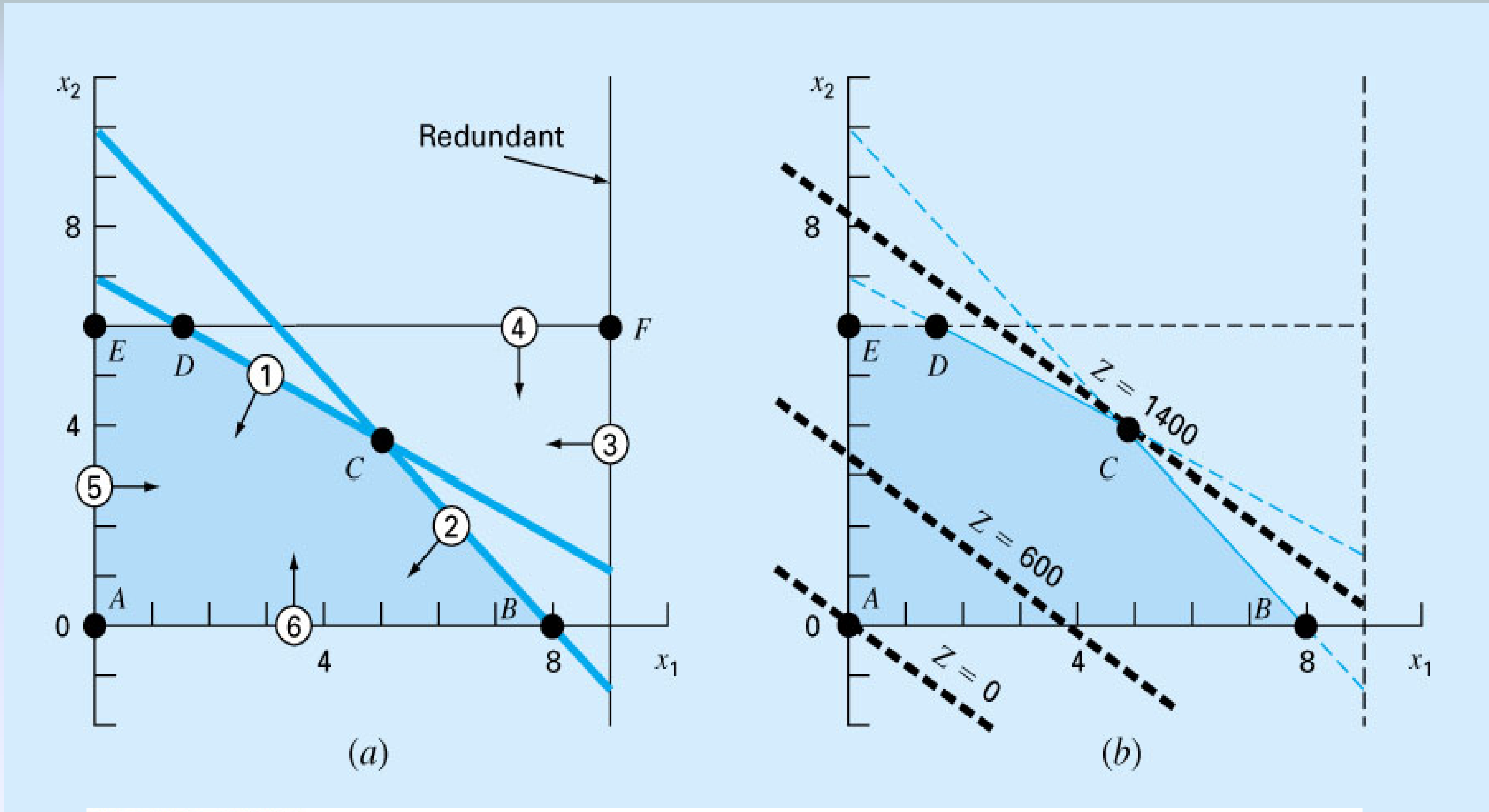


FIGURE 15.1

Graphical solution of a linear programming problem. (*a*) The constraints define a feasible solution space. (*b*) The objective function can be increased until it reaches the highest value that obeys all constraints. Graphically, the function moves up and to the right until it touches the feasible space at a single optimal point.

Possible outcomes that can be generally obtained in a linear programming problem/

1. *Unique solution*. The maximum objective function intersects a single point.
2. *Alternate solutions*. Problem has an infinite number of optima corresponding to a line segment.
3. *No feasible solution*.
4. *Unbounded problems*. Problem is under-constrained and therefore open-ended.

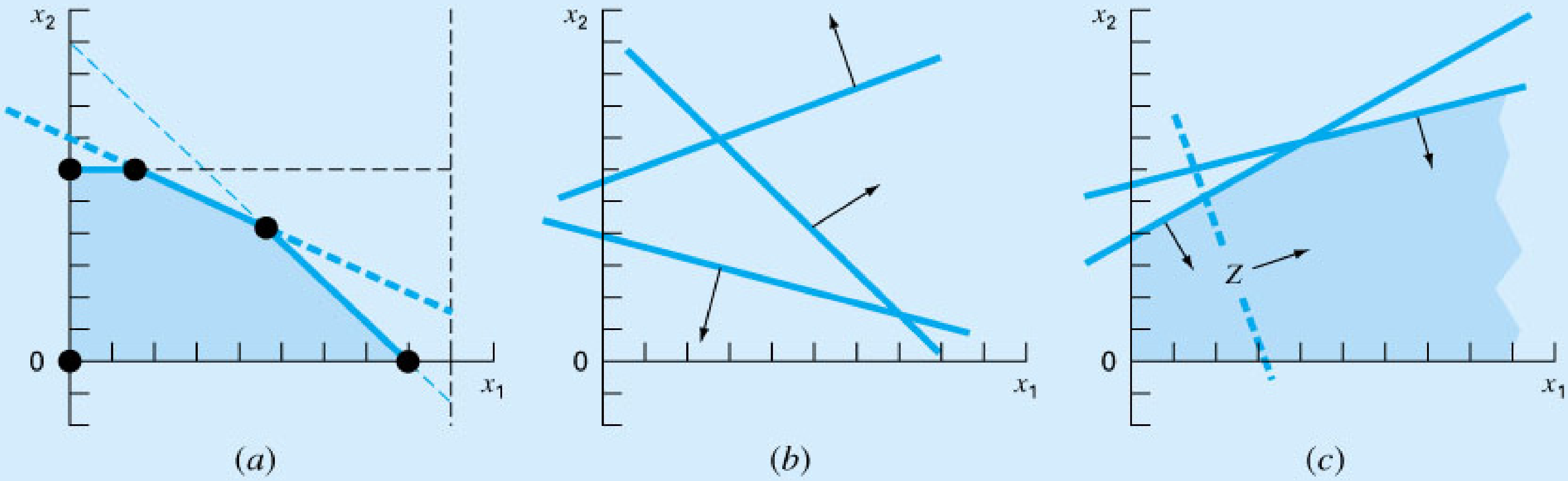


FIGURE 15.2
Aside from a single optimal solution (for example, Fig. 15.1*b*), there are three other possible outcomes of a linear programming problem: (*a*) alternative optima, (*b*) no feasible solution, and (*c*) an unbounded result.

The Simplex Method/

- Assumes that the optimal solution will be an extreme point.
- The approach must discern whether during problem solution an extreme point occurs.
- To do this, the constraint equations are reformulated as equalities by introducing slack variables.

- A slack variable measures how much of a constrained resource is available, e.g.,

  $7x_1 + 11\, x_2 \leq 77$

  If we define a slack variable S_1 as the amount of raw gas that is not used for a particular production level (x_1, x_2) and add it to the left side of the constraint, it makes the relationship exact.

  $7x_1 + 11\, x_2 + S_1 = 77$

- If slack variable is positive, it means that we have some slack that is we have some surplus that is not being used.
- If it is negative, it tells us that we have exceeded the constraint.
- If it is zero, we have exactly met the constraint. We have used up all the allowable resource.

Maximize

$$Z = 150x_1 + 175x_2$$

$$\begin{aligned}
7x_1 + 11x_2 + S_1 &= 77 \\
10x_1 + 8x_2 + S_2 &= 80 \\
x_1 + S_3 &= 9 \\
x_2 + S_4 &= 6
\end{aligned}$$

$$x_1, x_2, S_1, S_2, S_3, S_4 \geq 0$$

- We now have a system of linear algebraic equations.
- For even moderately sized problems, the approach can involve solving a great number of equations. For *m* equations and *n* unknowns, the number of simultaneous equations to be solved are:

$$C_m^n = \frac{n!}{m!(n-m)!}$$

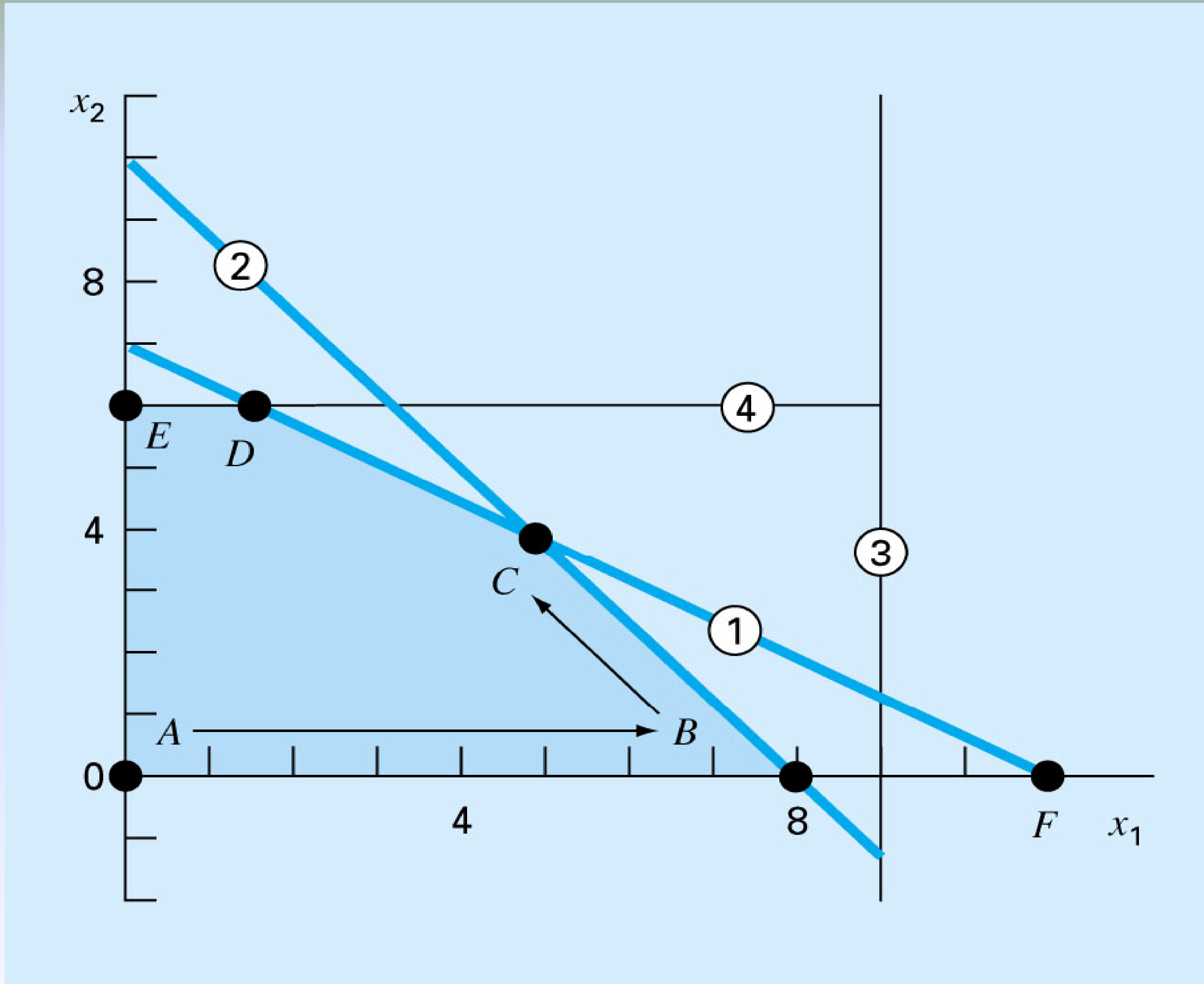


FIGURE 15.3
Graphical depiction of how the simplex method successively moves through feasible basic solutions to arrive at the optimum in an efficient manner.